Computergestützte Untersuchungen zur Dynamik und Wasserbindung des Ω-loops in TEM β-Lactamasen

Von der Fakultät Geo- und Biowissenschaften
der Universität Stuttgart zur Erlangung der Würde eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

Vorgelegt von
Fabian Alexander Bös
aus Freiburg im Breisgau

Institut für Technische Biochemie
der Universität Stuttgart

2007

Hauptberichter: Prof. Dr. Jürgen Pleiss
Mitberichter: Prof. Dr. Rolf D. Schmid

Tag der mündlichen Prüfung: 11.12.2007

Teile dieser Arbeit wurden bereits veröffentlicht / Parts of this thesis have previously been published:

Meinen Eltern gewidmet

Inhaltsverzeichnis

Zusammenfassung

Der massive Einsatz von β-Lactam-Antibiotika in der Medizin und Landwirtschaft hat bei Mikroorganismen zu einem verstärkten Auftreten von Resistenzen gegenüber diesen Substanzen geführt. Die Enzyme der TEM β-Lactamasen stellen in gramnegativen Bakterien die Hauptursache für die Resistenz gegenüber β-Lactam-Antibiotika dar. Sie gehören zu den Klasse A β-Lactamasen (Serin-β-Lactamasen) und hydrolysieren effizient und irreversibel die Amid-Bindung des β-Lactam-Rings, wodurch das Antibiotikum in eine inaktive Form überführt wird. Der Ω-loop ist ein in allen Klasse A β-Lactamasen konserviertes Strukturelement und positioniert die für den Reaktionsmechanismus essentielle Aminosäure Glu166. Der Ω-loop befindet sich an der Oberfläche des Enzyms und ist durch eine Salzbrücke zwischen den Aminosäuren Arg164 und Asp179 in seiner Form fixiert. Der Grad seiner Flexibilität ist jedoch auf Grund teils widersprüchlicher Ergebnisse bisheriger Untersuchungen unklar. Da bekannt ist, dass Wasser einen großen Einfluss auf die Struktur, Funktion und Dynamik von Proteinen hat, wurde in dieser Arbeit mit computergestützten Methoden der Zusammenhang zwischen der Wasserbindung und der Dynamik des Ω-loops untersucht. Zudem wurde im Rahmen dieser Arbeit an der Automatisierung von multiplen molekulardynamischen Simulationen in Cluster- und Gridumgebungen sowie an neuen Methoden zur Analyse der Proteindynamik mitgearbeitet.

Die Kristallwassermoleküle aus 49 hochaufgelösten Röntgenkristallstrukturen der β-Lactamase-Familien TEM, SHV und CTX-M wurden einer Cluster-Analyse unterzogen, um Aussagen über konservierte Wassermoleküle in Klasse A β-Lactamasen treffen zu können. Es wurden insgesamt 13 Wassermoleküle identifiziert, welche in mehr als 90% der 49 Strukturen konserviert waren; darunter auch zwei zu 100% konservierte Wassermoleküle. Sechs der insgesamt 13 konservierten Wassermoleküle waren am Ω-loop lokali-

siert, drei weitere konservierte Wassermoleküle waren mit anderen loop-Elementen in der Proteinstruktur assoziiert. Die am Ω-loop lokalisierten Wassermoleküle befanden sich in einer Aushöhlung, welche der Ω-loop mit den gegenüberliegenden Aminosäuren des Proteinkerns formt. Diese konservierten Wassermoleküle ermöglichen die Ausbildung eines Wasserstoffbrückennetzwerks zwischen den Aminosäuren des Ω-loops und denen des Proteinkerns und stabilisieren somit den Ω-loop. Weiterhin konnte gezeigt werden, dass sich die evolutionäre Beziehung zwischen den drei β-Lactamase-Familien interessanterweise auch in der unterschiedlichen Anzahl der Wassermoleküle widerspiegelt, die zwischen den einzelnen Familien konserviert sind.

Um die Dynamik und Flexibilität des Ω-loops näher zu charakterisieren und den Einfluss der konservierten Wassermoleküle darauf zu untersuchen, wurden multiple molekulardynamische Simulationen der TEM β-Lactamase in Wasser als explizitem Lösungsmittel durchgeführt. Die Durchführung mehrerer Simulationen ermöglichte eine effizientere Durchmusterung des Konformationsraumes und erlaubte es, die Reproduzierbarkeit und Signifikanz der Ergebnisse besser beurteilen zu können. Es zeigte sich, dass die TEM β-Lactamase ein hoch geordnetes und rigides Protein ist, dass nur in loop-Bereichen eine erhöhte Flexibilität aufweist. Im Falle des Ω-loops stellte sich heraus, dass sich dieser in zwei Bereiche mit jeweils unterschiedlichen Flexibilitäten einteilen lässt. Während der Hauptteil des Ω-loops eine geringe Flexibilität aufweist, sind die Aminosäuren 173 bis 177 durch eine erhöhte Flexibilität gekennzeichnet. Dieses kurze Teilstück des Ω-loops zeigte eine deutliche Bewegung, die einem Öffnen und Schließen der Aushöhlung am Ω-loop entspricht. Durch die Kombination mehrerer Simulationen konnte auch gezeigt werden, dass die Ergebnisse früherer Arbeiten über die Flexibilität des Ω-loops keinesfalls im Widerspruch zueinander stehen, sondern jeweils nur unterschiedliche Teilschritte der Ω-loop Bewegung beschreiben, deren Gesamtdauer über die einer einzelnen Simulation hinausgeht. Für vier der sechs in der Kristallstrukturanalyse identifizierten konservierten Wassermoleküle konnte in den multiplen Simulationen reproduzierbar gezeigt werden, dass sie Wasserbrücken zwischen den Aminosäuren des Ω-loops und den gegenüberliegenden Aminosäuren des Proteinkerns ausbilden. Die Aminosäuren 173 bis 177 sind jedoch nicht an der Ausbildung dieser Wasserbrücken beteiligt, was ihre erhöhte Flexibilität erklärt. In Kontrollsimulationen, die ohne das ursprüngliche Kristallwasser gestartet wurden, erfolgte innerhalb der ersten 200 ps eine rasche Solvatisierung der Aushöhlung am Ω-loop und damit die Ausbildung der stabilisierenden Wasserbrücken. Weiterhin konnte eine

sich spontan öffnende Lücke in der Ω-loop-Aushöhlung beobachtet werden, an der auch die Seitenkette von Arg164 beteiligt ist. Damit konnte eine mögliche Erklärung für die Konformationsänderung des Ω-loops vorgeschlagen werden, welche durch die häufig in ESBL-Varianten auftretende Arg164Ser Substitution verursacht wird.

Multiple molekulardynamische Simulationen von Proteinen wie der TEM β-Lactamase sind mit einem sehr hohen Rechenaufwand verbunden, der nur durch die Nutzung von Cluster- und Grid-Umgebungen gedeckt werden kann. Zudem stellt die Durchführung multipler Simulationen eine logistische Herausforderung in Bezug auf die Vorbereitung, Durchführung und Auswertung der zahlreichen Simulationen dar, die nur noch sehr schwer von Hand zu bewältigen ist. Im Rahmen dieser Arbeit wurde daher in Zusammenarbeit mit dem Höchstleistungsrechenzentrum der Universität Stuttgart der gesamte Ablauf einer multiplen molekulardynamischen Simulation in ein System zur automatisierten Durchführung wissenschaftlicher Arbeitsabläufe implementiert und erfolgreich in einer Grid-Umgebung getestet. Es zeigte sich, dass gegenüber der manuellen Durchführung der einzelnen Simulationen eine enorme Zeitersparnis möglich ist. Als weiterer Vorteil stellte sich heraus, dass die Verwendung eines solchen Systems zur Fehlerminimierung bei der Durchführung multipler molekulardynamischer Simulationen beiträgt, da eine interaktive Erstellung aller notwendigen Eingabedateien durch den Benutzer entfällt. Zusätzlich können die einmal erstellten Arbeitsabläufe wieder verwendet und mit anderen Nutzern ausgetauscht werden.

Eine direkte und zeitabhängige Analyse der Proteindynamik, die über das bloße Betrachten von Atomen im dreidimensionalen Raum hinausgeht, ist bisher nicht möglich. In Zusammenarbeit mit dem Institut für Visualisierung und Interaktive Systeme der Universität Stuttgart wurde daher eine neuartige Methode zur haptischen Interaktion und Analyse von molekulardynamischen Simulationen entwickelt. Hierzu wurde ein haptisches Eingabegerät, dass Bewegungen in sechs Freiheitsgraden ermöglicht und über einen stiftähnlichen Griff bedient wird, mit einem Visualisierungsprogramm für molekulardynamische Simulationen gekoppelt. Dieser Ansatz ermöglicht durch das Zuweisen von Oberflächeneigenschaften, wie z.B. glatt oder rau, an die betrachtete Proteinstruktur die Übermittlung zusätzliche Informationen. Zudem ist es möglich, über das haptische Eingabegerät direkt mit der Proteinstruktur zu interagieren und somit die Proteindynamik im Verlauf einer molekulardynamischen Simulation interaktiv und zeitabhängig

zu analysieren. Im Gegensatz zur herkömmlichen Darstellung der Atomflexibilität als skalare Größe ohne Berücksichtigung der Bewegungsrichtung übermittelt der haptische Ansatz dem Benutzer diese zusätzliche Information.

Summary

Due to the massive use of β-lactam antibiotics in medicine and livestock farming, the bacterial resistance against these compounds has considerably increased. Although there are quite a few different mechanisms for antibiotic resistance, the production of TEM β-lactamases is the main course for resistance against β-lactam antibiotics in gram-negative bacteria. TEM β-lactamases belong to the class A serine-β-lactamases and efficiently hydrolyze the amide bond of the four-membered β-lactam-ring, thus rendering the antibiotic ineffective. The Ω-loop is a conserved structural element in all class A β-lactamases and is directly involved in the catalytic reaction of the enzymes because it positions the essential amino acid Glu166. It is located at the surface of the enzyme and it's conformation is anchored by a conserved salt bridge formed between Arg164 and Asp179. However, the flexibility of the Ω-loop is still under debate, as previous studies led to contradictory results. Because water plays an important role in protein structure, function and dynamics, in this work the interplay between water binding and the dynamics at the Ω-loop was investigated using computational methods. Further, in this work tools for efficiently performing multiple molecular dynamics simulations in cluster and grid environments as well as novel methods for analyzing protein dynamics have been co-developed.

Crystal water molecules from 49 high-resolution X-ray structures of the TEM, SHV, and CTX-M class A β-lactamase families were systematically analyzed using a clustering-based approach to identify conserved water molecules in class A β-lactamases. Overall, 13 water molecules were found to be conserved in at least 90% of the 49 structures, including two water positions which were found to be conserved in all structures. From the 13 conserved water molecules, six are located inside a cavity lined by the Ω-loop and residues of the protein core. Forming a dense hydrogen bond network between

those residues, this layer of conserved water molecules thus stabilize the Ω-loop. Three additional conserved water molecules were associated with other loop regions in the protein structure. Further, the evolutionary relationship between the three families derived from the number of conserved water molecules is similar to the relationship derived from phylogenetic analysis.

Multiple molecular dynamics simulations of the TEM β-lactamase in explicit water were performed and analyzed to investigate the dynamics of the Ω-loop and its interplay with bound water molecules. Using multiple simulations allows a more efficient sampling of the proteins conformational space and was used to estimate the reproducibility and significance of the results. The analysis of protein flexibility showed that the TEM-1 β-lactamase is a very compact and rigid protein with highly ordered secondary structure elements and slightly less ordered loop regions. However, while the main part of Ω-loop is rigid, residues 173 to 177 exhibited a considerably higher flexibility. These residues perform a motion that resembles an opening and closing movement of the Ω-loop cavity. By using multiple molecular dynamics simulations the results of two previous studies, which at first seemed to be contradictory, could be explained: They both described parts of a motion, which occurs on a time scale that was not accessible in a single simulation. Stabilizing water bridges, mediated by four of the six conserved water molecules identified in the crystal structure analysis, were repeatedly shown to occur inside the Ω-loop cavity, explaining the heterogeneous distribution of flexibility at the Ω-loop. While the main part is stabilized towards the protein core by the water bridges, the residues 173 to 177 lack these interactions and thus show an increased flexibility. The cavity was shown to be solvated in simulations starting without the crystal water within the first 200 ps. Further, some water molecules were observed to leave the Ω-loop cavity through a dynamically formed opening in which also the side chain of Arg164 is involved. This led to the proposal of a possible explanation for the conformational change of the Ω-loop, which is induced by the Arg164Ser substitution commonly found in extended-spectrum β-lactamases.

Multiple molecular dynamics simulations of proteins such as the TEM β-Lactamase require extensive computational power and run on large PC clusters or in a grid environment. The tasks of preparing, running and analyzing of multiple simulations can no longer be performed interactively and therefore have to be automated. In cooperation

with the High Performance Computing Center Stuttgart at the University of Stuttgart, the procedure of a multiple molecular dynamics simulation experiment was implemented in a system for scientific workflows and successfully tested in a grid environment. Compared to manually performing multiple molecular dynamics simulations, a significant amount of time could be saved by using such a system. The base for common errors, such as misspelling in input files was also minimized. Beneficial is also the reusability of the once designed procedures and the possibility to share them with other users.

A direct time-dependent analysis of protein dynamics other than looking at individual atoms or group of atoms in 3D space has not been possible up to now. In cooperation with the Visualization and Interactive Systems Group at the University of Stuttgart a novel method for haptic interaction and analysis of molecular dynamics trajectories was therefore developed. A haptic device, which is operated via a pencil-like handhold and that allows movements in six degrees of freedom, was integrated into a molecular dynamics trajectory viewer. This approach allows transmitting of additional information by mapping surface properties such as roughness or smoothness onto the protein structure. Additionally, it is now possible to directly interact with the protein during the course of the molecular dynamics simulation and to interactively investigate the protein dynamics. Tracking of motion allows the time-dependent analysis of anisotropic protein dynamics, something that is not covered in standard analysis of molecular dynamics trajectories.

1 Einleitung

1.1 β-Lactam-Antibiotika

1.1.1 Strukturen und Klassifikation

Eine der bedeutendsten Entwicklungen im Bereich der medizinischen Versorgung hat ihren Ursprung in der Beobachtung, dass die Anwesenheit eines Pilzes des Stammes *Penicillium* das Wachstum einer Bakterienkultur beeinträchtigen kann. Diesen Effekt entdeckte Alexander Fleming 1928 bei einer Staphylokokken-Kultur (Fleming, 1929). Jedoch wurde die antibakterielle Eigenschaft des Penicillins und damit sein therapeutisches Potential erst Jahre später von Florey und Chain an mit Streptokokken infizierten Mäusen gezeigt (Chain et al., 1940). Der erfolgreiche Einsatz von Penicillin in der klinischen Praxis legte den Grundstein für die Behandlung bakterieller Infektionen und weckte das Interesse an dieser Substanzklasse, was in den nachfolgenden zwei Jahrzehnten zur Entdeckung weiterer Antibiotika, darunter auch den Cephalosporinen (Florey, 1955), führte. Das Spektrum der klinisch genutzten β-Lactam-Antibiotika beschränkte sich anfangs auf die zwei Präparate Penicillin G und Penicillin V, erfuhr jedoch später mit der Entwicklung halbsynthetischer Penicilline und Cepahlosporine eine massive Erweiterung. Mengen- sowie wertmäßig gehören die β-Lactam-Antibiotika heute zu den wichtigsten Antibiotika (Demain und Elander, 1999).

β-Lactam-Antibiotika gehören zur Klasse der Aminosäure- und Peptidantibiotika und besitzen einen für sie charakteristischen vieratomigen β-Lactam-Ring, der als isolierter Ring oder in Verbindung mit anderen Ringen als bizyklische Struktur vorliegen kann (Bryan und Godfrey, 1991). Dieses Strukturmerkmal besitzen auch die β-Lactamase-

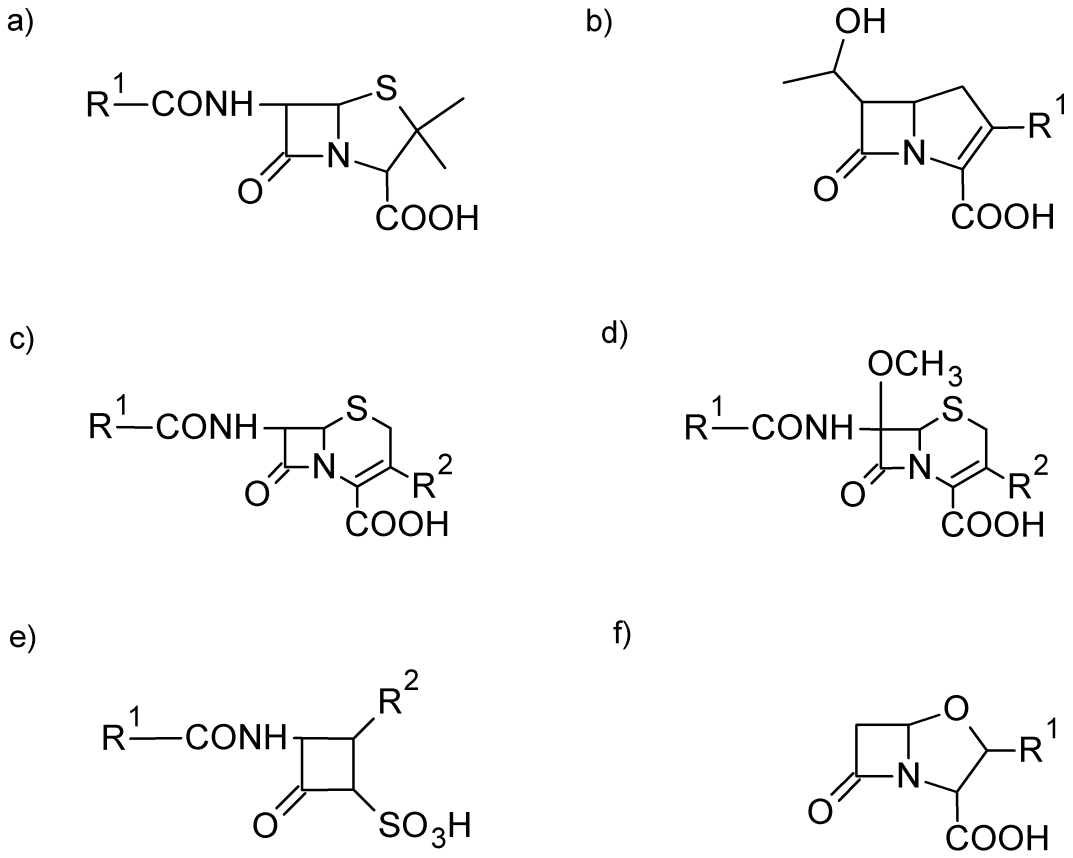


Abbildung 1: Grundgerüst einiger Vertreter der β-Lactam-Antibiotika: a) Penicilline, b) Carbapeneme, c) Cephalosporine, d) Cephamycine, e) Monobactame, f) Clavulansäure (β-Lactamase-Inhibitor). Nähere Erläuterungen siehe Text. Verändert nach Essack (2001).

Inhibitoren, die an dieser Stelle aber nur kurz erwähnt werden sollen, da sie nicht antibiotisch wirksam sind.

Historisch gesehen sind die "klassischen" Hauptvertreter der β-Lactam-Antibiotika die Penicilline und Cephalosporine (Abbildung 1). Penicilline besitzen einen an den β-Lactam-Ring ankondensierten Thiazolidin-Ring und verfügen über eine variable Acyl-Seitenkette. Die Biosynthese der Penicilline erfolgt in *Penicillium chrysogenum* durch enzymatische Verknüpfung von L-α-Aminoadipinsäure, L-Valin und L-Cystein zu einem Tripeptid (Banko et al., 1987), welches durch einen weiteren enzymatischen Schritt in Isopenicillin N umgewandelt wird (O'Sullivan et al., 1979). Die Zugabe von Seitenkettenvorstufen (z.B. Phenylessigsäure) in das Kulturmedium bewirkt einen durch das Enzym

Transacetylase katalysierten Austausch der Aminoadipinsäure gegen die zugegebene Seitenkettenvorstufe. Dies führt im Fall von Phenylessigsäure zu Penicillin G oder je nach zugegebener Vorstufe zu einem anderen biosynthetischen Penicillin-Derivat mit entsprechender Seitenkette. Cephalosporine enthalten einen an den β-Lactam-Ring ankondensierten Dihydrothiazin-Ring (Newton und Abraham, 1955). Die Biosynthese in *Acremonium chrysogenum* verläuft analog der von Penicillin über Isopenicillin N und führt dann über eine Epimerisierung der Seitenkette und enzymkatalysierter oxidativer Ringerweiterung zu Cephalosporin C (Demain, 1963). Im Gegensatz zu *P. chrysogenum* besitzt *A. chrysogenum* keine Transacetylase, so dass ein Zusatz von Seitenkettenvorstufen nicht zu biosynthetischen Cephalosporin-Derivaten führt. Als Ausgangsbasis für die Herstellung halbsynthetischer Penicilline und Cephalosporine dienen 6β-Aminopenicillansäure (6-APA) (Batchelor et al., 1959) bzw. 7β-Aminocephalosporansäure (7-ACA) (Morin et al., 1962), welche mittels enzymatischer oder chemischer Hydrolyse aus Penicillin G bzw. Cephalosporin C gewonnen werden. Die chemische Addition von verschiedenen Acyl-Seitenketten an 6-APA oder 7-ACA erfolgt anschließend meist durch eine Schotten-Baumann-Reaktion (Rolinson und Sutherland, 1973).

Neben den Penicillinen und Cephalosporinen sowie ihren Derivaten umfasst die Klasse der β-Lactam-Antibiotika noch die Cephamycine, Carbapeneme, Monobactame und Monocarbame (Livermore und Williams, 1996). Bei den Cephamycinen ersetzt eine Methoxy-Gruppe den Wasserstoff an der Position C7 des Cephalosporin-Grundgerüsts. Klinische Anwendung finden aber hauptsächlich halbsynthetische Derivate wie Cefoxitin, die ein breites Wirkungsspektrum besitzen (Miller, 1983). Carbapeneme wie z.B. Imipenem und Meropenem ähneln von ihrer Struktur her den Penicillinen, besitzen jedoch im Fünfring an Stelle des Schwefels ein Kohlenstoffatom und eine Doppelbindung zwischen C2 und C3. Carbapeneme haben das breiteste Aktivitätsspektrum aller β-Lactam-Antibiotika, sowohl gegenüber gramnegativen als auch grampositiven Arten, und zeichnen sich durch eine hohe Stabilität gegenüber β-Lactamasen aus (Hawkey, 1997). Die Struktur der Monobactame und Monocarbame besteht aus einem einzelnen β-Lactam-Ring. Das Grundgerüst der Monobactame und Monocarbame besitzt eine nur schwache antibiotische Wirkung und wurde daher mit verschiedenen Seitenketten substituiert (Livermore und Williams, 1996). Aztreonam, das einzige klinisch verwendete Monobactam, wirkt hauptsächlich gegen gramnegative Bakterien (Bonner und Sykes, 1984).

1.1.2 Wirkungsweise und Zielstrukturen

Die Zielstrukturen der β-Lactam-Antibiotika sind die Transpeptidasen, die am Aufbau der bakteriellen Zellwand beteiligt sind (Spratt, 1994). Eine intakte Zellwand ist die Voraussetzung, um den hohen osmotischen Druck zu kompensieren, der durch den Konzentrationsunterschied von gelösten Stoffen zwischen dem Cytoplasma des Organismus und der Umgebung entsteht (Livermore und Williams, 1996). Um überlebensfähig zu sein, muss die Zellwand während des gesamten bakteriellen Zellzyklus intakt bleiben und unterliegt daher einem stetigen Auf-, Ab- und Umbau (Jacoby, 1997).

Trotz ihrer Unterschiede im Zellwandaufbau besitzen sowohl gramnegative wie auch grampositive Bakterien eine Schicht aus dreidimensional vernetzten Peptidoglykansträngen. Das Peptidoglykan besteht aus den Grundbausteinen N-Acetylglucosamin und N-Acetylmuraminsäure, die alternierend angeordnet und untereinander β-(1,4)-glykosidisch zu langen, unverzweigten Polysaccharid-Filamenten verbunden sind. Die Querverknüpfung der Polysaccharid-Filamente erfolgt durch Transpeptidasen an Hand von kurzen Peptidseitenketten, die an den Muraminsäure-Monomeren abzweigen. Sie werden durch Isopeptidbindungen entweder direkt oder über Peptidbrücken, wie z.B. Pentaglycin, verknüpft.

β-Lactam-Antibiotika werden von den Transpeptidasen kovalent gebunden und beeinträchtigen damit den finalen Schritt der Peptidoglykan-Biosynthese (Danziger und Pendland, 1995). Daher werden die Transpeptidasen auch Penicillin-Binde-Proteine (PBP) genannt. Bakterien besitzen mehrere PBP, welche auf Grund ihres Molekulargewichtes in hoch- und niedermolekulare PBP eingeteilt werden. Die essentiellen hochmolekularen PBP sind Multidomänenproteine und werden in bifunktionale (Transpeptidase- und Transglycosidase-Aktivität) und monofunktionale PBP (nur Transpeptidase-Aktivität) unterschieden. Die nicht essentiellen niedermolekularen PBP sind monofunktional und übernehmen mit Carboxypeptidase-Aktivität die Regulierung der Peptidoglykanvernetzung (Goffin und Ghuysen, 2002; Morlot et al., 2004; Macheboeuf et al., 2005).

Fast alle PBP besitzen einen konservierten Serinrest im aktiven Zentrum und arbeiten nach einem Acylierungs-Deacylierungs-Mechanismus, bei dem vorübergehend ein Acylenzym gebildet wird. Auf Grund der sterischen Ähnlichkeit mit den natürlichen

Substraten der PBP, den Peptidseitenketten der Muraminsäure-Monomere, binden die β-Lactam-Antibiotika ebenfalls kovalent an den Serinrest. Der daraus resultierende Serinester zwischen β-Lactam-Antibiotikum und PBP wird aber im Gegensatz zu einem Serinester mit einem natürlichen Substrat nicht oder nur sehr langsam wieder hydrolysiert. Dadurch entstehen stabile Acyl-Enzyme und den so inhibierten PBP ist es nicht mehr möglich, an der Zellwand-Biosynthese mitzuwirken (Matagne et al., 1998). Kann die Quervernetzung der Polysaccharid-Filamente nicht mehr erfolgen, ist die Zellwand nicht mehr in der Lage dem osmotischen Druck standzuhalten und es kommt zur Bakteriolyse (McDowell und Reed, 1989).

1.1.3 Resistenzmechanismen

Resistenzen gegenüber β-Lactam-Antibiotika treten sowohl in grampositiven als auch in gramnegativen Bakterien auf. Ein Bakterienstamm wird dann als resistent bezeichnet, wenn bei einer Konzentration des Antibiotikums, die normalerweise ausreicht um einen typischen Vertreter dieser Art im Wachstum zu hemmen, die gewünschte antibiotische Wirkung nicht mehr erzielt werden kann. Die Ursache hierfür sind eine Vielzahl von verschiedenen Resistenzmechanismen, die auch in Kombination miteinander auftreten können. Während in gramnegativen Bakterien die Spaltung der β-Lactam-Antibiotika durch β-Lactamasen den bedeutsamsten Resistenzmechanismus darstellt (siehe 1.1.4), überwiegt bei grampositiven Bakterien die Resistenz durch Veränderungen an den Zielstrukturen, den PBP.

Veränderte PBP zeigen keine oder nur geringe Affinität gegenüber β-Lactam-Antibiotika und können daher von diesen nicht mehr oder nur in geringerem Umfang inaktiviert werden. Dieser Mechanismus tritt sowohl bei gramnegativen als auch bei grampositiven Bakterien auf, wobei er bei letzteren eine besonders große Rolle spielt. Im Vergleich zu β-Lactam-sensitiven Stämmen besitzen Methicillin-resistente *Staphylococcus aureus* ein chromosomales Genfragment (mecA), welches für ein zusätzliches PBP (PBP2a) kodiert (Hartman und Tomasz, 1984). PBP2a besitzt eine nur sehr geringe Affinität gegenüber β-Lactam-Antibiotika (Hartman und Tomasz, 1986) und übernimmt die Transpeptidase-Aktivität der β-Lactam-sensitiven PBP (Pinho et al., 2001). Weitere Beispiele für β-Lactam-insensitive PBP sind PBP2x aus *Streptococcus pneumoniae* (Grebe und Haken-

beck, 1996) und PBP5 aus *Enterococcus faecium* (Sifaoui et al., 2001).

Einen weiteren Resistenzmechanismus stellt die Verringerung der Antibiotika-Konzentration am Wirkort dar. Hydrophile Substanzen wie die β-Lactam-Antibiotika werden in gramnegativen Bakterien über Kanäle, sogenannte Porine, in der äußeren Membran transportiert. Eine veränderte Struktur oder eine geringere Anzahl dieser Porine verringert die Membranpermeabilität für manche β-Lactam-Antibiotika. Dies schränkt den Transport in den periplasmatischen Raum und damit zu den PBP an der inneren Zellmembran ein (Nikaido, 2000). Ein Mangel an Porinen als Ursache für die Resistenz gegenüber β-Lactam-Antibiotika wurde besonders häufig bei *Klebsiella pneumoniae* und *Enterobacter spp.* beobachtet (Poole, 2002). Die aktive Entfernung der β-Lactam-Antibiotika aus der Zelle erfolgt mit Hilfe von Effluxpumpen. Bisher sind fünf Familien von bakteriellen Effluxpumpen beschrieben worden: die *major facilitator superfamily* (MFS), die *ATP-binding cassette* (ABC) Familie, die *resistance-nodulation-division* (RND) Familie, die *multidrug and toxic compound extrusion* (MATE) Familie und die *small multidrug resistance* (SMR) Familie (Bambeke et al., 2000; Li und Nikaido, 2004). Effluxpumpen finden sich in gramnegativen als auch in grampositiven Bakterien und können spezifisch oder unspezifisch für die zu transportierenden Substrate sein. Sie nutzen allesamt die protonenmotorische Kraft als Energiequelle für den Transport. Die Ausnahme sind Pumpen der ABC Familie, die ihre Energie durch ATP-Hydrolyse gewinnen (Webber und Piddock, 2003). Die Mitglieder der RND Familie, welche fast ausschließlich chromosomal vorliegen, sind unter gramnegativen Bakterien weit verbreitet und im Hinblick auf die Resistenz gegenüber β-Lactam-Antibiotika am bedeutsamsten (Poole, 2004a).

1.1.4 Resistenz durch β-Lactamasen

In gramnegativen Bakterien stellen die β-Lactamasen (EC 3.5.2.6) die Hauptursache für eine Resistenz gegenüber β-Lactam-Antibiotika dar. Sie sind jedoch auch in grampositiven Organismen zu finden, spielen hier aber nur eine untergeordnete Rolle. β-Lactamasen katalysieren effizient und irreversibel die Hydrolyse der Amid-Bindung des β-Lactam-Rings und überführen somit das Antibiotikum in eine inaktive Form (Majiduddin et al., 2002). Die Entstehung von β-Lactamasen wird als Anpassung an den evolutionären Druck gesehen, den β-Lactam-Antibiotika-produzierende Mikroorganismen auf ihre Um-

welt ausüben (Medeiros, 1997a). Interessanterweise weisen manche β-Lactamasen eine strukturelle Homologie mit den PBP auf und es wird vermutet, das sie evolutionär aus PBP hervorgegangen sind, welche die Fähigkeit entwickelt haben, den Serinester mit dem β-Lactam-Antibiotikum zu hydrolysieren (Massova und Mobashery, 1998; Medeiros, 1997b).

Obwohl β-Lactamasen schon vor der klinischen Einführung von Penicillin beschrieben wurden (Abraham und Chain, 1940), hatte der massive und weit verbreitete Einsatz von β-Lactam-Antibiotika in der Medizin und Landwirtschaft einen nicht unerheblichen Einfluss auf ihre rapide Verbreitung und Evolution (Davies, 1994). Bis heute sind über 530 β-Lactamasen bekannt (Babic et al., 2006). β-Lactamasen sind bei gramnegativen Bakterien im Periplasma lokalisiert, bei Grampositiven werden sie hingegen in das umgebende Medium ausgeschieden oder sind an die Cytoplasmamembran gebunden (Nielsen und Lampen, 1982). Die Gene, welche für β-Lactamasen kodieren, liegen chromosomal oder auf Plasmiden vor. Eine Expression erfolgt konstitutiv oder nach Induktion (Fisher et al., 2005). Viele β-Lactamase-Gene befinden sich auf Transposons und ermöglichen damit eine Weitergabe zwischen verschiedenen Plasmiden und Organismen (Livermore, 1995).

Zwei Klassifikations-Schemata für β-Lactamasen finden heutzutage ihre Anwendung. Die Bush-Jacoby-Medeiros-Klassifikation nutzt das Substrat- und Inhibitorprofil der β-Lactamasen, d. h. es wird das Spektrum der hydrolysierten β-Lactam-Antibiotika sowie eine mögliche Hemmung durch Inhibitoren wie z.B. Clavulansäure für die Einteilung berücksichtigt (Bush et al., 1995). Die Klassifikation erfolgt in vier Hauptgruppen sowie diversen Untergruppen. β-Lactamasen der Gruppe 1 hydrolysieren Cephalosporine, werden aber nicht ausreichend von Clavulansäure gehemmt. Die der Gruppe 2 hydrolysieren sowohl Penicilline als auch Cephalosporine, werden aber alle durch Clavulansäure gehemmt. Mitglieder der Gruppe 3 sind Metallo-β-Lactamasen und hydrolysieren Carbapeneme, sind resistent gegenüber Clavulansäure und werden durch Ethylendiamintetraessigsäure (EDTA) inhibiert. Gruppe 4 schließlich umfasst verschiedene weitere β-Lactamasen. Die Klassifizierung über das Substrat- und Inhibitorprofil sieht sich jedoch mit dem Problem konfrontiert, dass selbst schon einzelne Substitutionen in der Aminosäuresequenz einer β-Lactamase das Substratspektrum und die Empfindlichkeit gegenüber Inhibitoren erheblich beeinflussen können. Dies führt dann zwangsläufig zu

einer Zuordnung in eine andere Gruppe (Livermore, 1995). Hiervon ist die Klassifikation nach Ambler, welche auf Ähnlichkeiten der Aminosäuresequenz beruht, nicht betroffen (Ambler, 1980). Die ursprünglich vorgeschlagene Klassifikation wurde nachträglich erweitert (Jaurin und Grundstrom, 1981; Ouellette et al., 1987) und unterscheidet zwischen den vier Klassen A bis D. Die Klassen A, C und D umfassen Serin-β-Lactamasen, welche unter Verwendung eines katalytisch aktiven Serinrestes und der Bildung eines Acyl-Enzym-Zwischenproduktes die Hydrolyse der Amid-Bindung durchführen (Matagne et al., 1998). Enzyme der Klasse B sind Metallo-β-Lactamasen und benötigen ein oder zwei Zinkatome für die Hydrolyse der β-Lactam-Antibiotika (Walsh et al., 2005).

1.2 TEM β-Lactamasen

1.2.1 Verbreitung und Nomenklatur

Die TEM β-Lactamasen gehören nach der Ambler-Klassifikation zu den Klasse A β-Lactamasen, in der Klassifikation nach Bush-Jacoby-Medeiros werden sie den Gruppen 2b, 2be und 2br zugeteilt. Sie sind die Hauptursache für die Resistenz gegenüber β-Lactam-Antibiotika in gramnegativen Bakterien (Bradford, 2001). Die erste plasmidkodierte TEM β-Lactamase (TEM-1)[1] wurde 1965 in Griechenland bei Enterobakterien beschrieben (Datta und Kontomichalou, 1965). In nur wenigen Jahren erfolgte die weltweite Verbreitung und die Übertragung in andere Bakterienspezies: Das Enzym wurde 1969 bei *P. aeruginosa* und 1973 bei *Vibrio cholerae* gefunden und hatte sich 1974 in die Gattungen *Haemophilus* und *Neisseria* ausgebreitet (Matthew, 1979). Die rasche Ausbreitung wurde nicht zuletzt durch das Vorhandensein des Gens blaTEM-1 auf den Transposons Tn2 und Tn3 begünstigt (Huang et al., 1996). Heutzutage findet man TEM β-Lactamasen in einer Vielzahl pathogener Mikroorganismen (Bradford, 2001).

Seit der Entdeckung der ersten TEM β-Lactamase sind bis heute 157 Varianten des Enzyms beschrieben worden[2], die man unter dem Begriff der TEM β-Lactamase Familie zu-

[1] Die Abkürzung TEM steht im Übrigen für „Temoneira“, der Name des Patienten, aus dem der Erreger isoliert wurde.

[2] siehe http://www.lahey.org/Studies/temtable.asp - Stand April 2007

sammenfasst. Die einzelnen Mitglieder der TEM β-Lactamase Familie besitzen eine sehr hohe Sequenzidentität und unterscheiden sich jeweils in nur ein bis sieben Aminosäure-Substitutionen, welche aber teilweise drastische Auswirkungen auf die biochemischen Eigenschaften des Enzyms haben (siehe 1.2.4). Die Mitglieder der TEM β-Lactamase Familie werden durch eine angehängte Nummer voneinander unterschieden. Indikation für eine neue Nummer ist eine neue, bisher noch nicht bekannte Aminosäuresequenz (Bush und Jacoby, 1997). Um die Mitglieder der TEM β-Lactamase Familie besser mit anderen Klasse A Familien vergleichen zu können, wurde eine Standardnummerierung für die Aminosäuren eingeführt (Ambler et al., 1991): Essentielle und konservierte Aminosäuren haben eine fixe Positionsnummer (z.B. Ser70), was zur Folge hat, dass bei TEM β-Lactamasen manche Positionsnummern nicht besetzt sind (z.B. ist die auf die Position 238 nachfolgende Aminosäure die mit der Nummer 240). Alle Positionsangaben in dieser Arbeit beziehen sich auf diese Standardnummerierung nach Ambler et al. (1991).

1.2.2 Struktureller Aufbau

TEM β-Lactamasen sind globuläre, monomere Proteine mit einem durchschnittlichen Molekulargewicht von 29 kDa und zwei strukturell unterscheidbaren Domänen (siehe Abbildung 2) (Jelsch et al., 1992). Die α/β-Domäne besteht aus fünf β-Faltblattsträngen und drei gegen die Lösungsmittelgrenze gepackten α-Helices. Gegenüber der α/β-Domäne liegt die *all*-α-Domäne, welche sich aus acht α-Helices zusammensetzt. Durch zwei Gelenkregionen (*hinge regions*), in denen elektrostatische Wechselwirkungen und Wasserstoffbrücken größere Konformationsänderungen verhindern, sind beide Domänen miteinander verbunden. Das aktive Zentrum befindet sich, dem Lösungsmittel zugänglich, an der Grenze der beiden Domänen (Matagne et al., 1998; Diaz et al., 2003).

Die vier wichtigsten Strukturelemente für den katalytischen Mechanismus der TEM β-Lactamase sind hochkonserviert (Matagne et al., 1998) und liegen in direkter Umgebung des aktiven Zentrums. Das erste Element (Ser70–X–X–Lys) befindet sich am N-terminalen Ende der ersten Helix der *all*-α-Domäne und umfasst das katalytisch aktive Serin sowie ein Lysin, dessen Seitenkette zum aktiven Zentrum hin gerichtet ist. Das zweite Element (Ser130–Asp–Asn, „SDN-loop“) liegt auf einem kurzen loop in der *all*-α-Domäne und begrenzt das aktive Zentrum zu dieser Seite hin. Die Seitenketten von

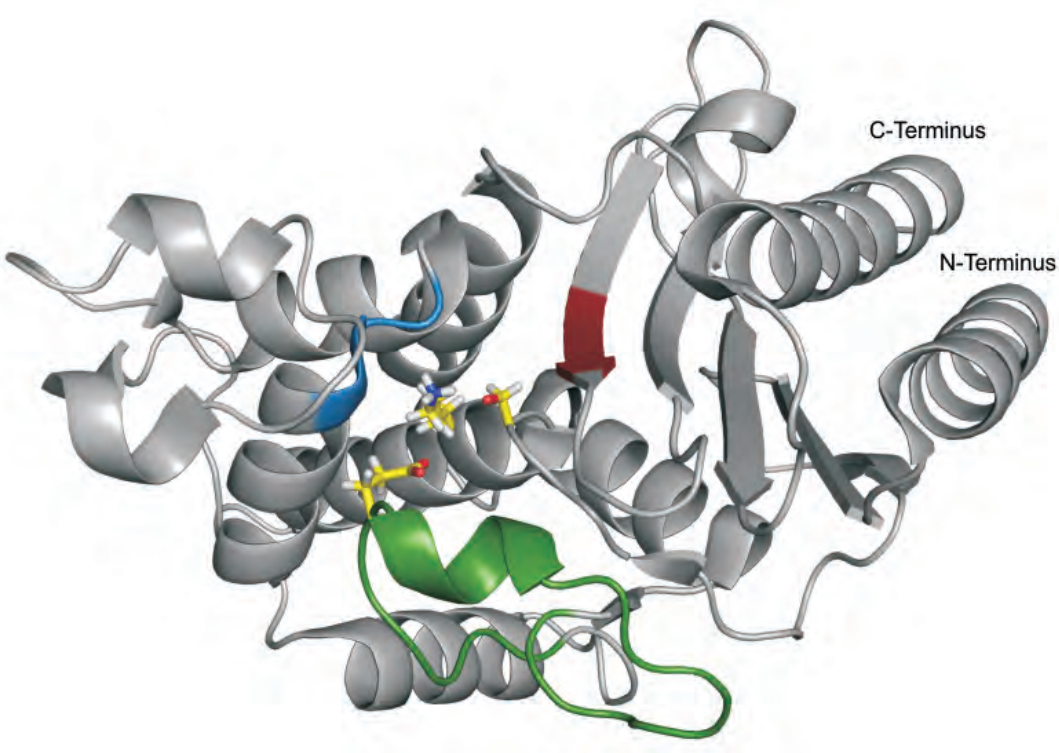


Abbildung 2: TEM-1 β-Lactamase in der *cartoon* Darstellung. Die katalytisch wichtigen Aminosäuren Ser70, Lys73 und Glu166 sind gelb dargestellt. Der Ω-loop ist grün, der „SDN-loop“ blau und die „KTG-Triade“ rot eingefärbt.

Ser130 und Asn132 sind zum aktiven Zentrum hin orientiert, während die Seitenkette von Asp131 im Inneren des Enzyms verankert ist. Auf dem innersten β-Faltblattstrang der α/β-Domäne befindet sich das dritte Element (Lys234–Thr/Ser–Gly, „KTG-Triade“), welches das aktive Zentrum zur anderen Seite hin begrenzt. Das vierte Element umfasst die Aminosäuren 164 bis 179 und wird auf Grund seiner Form als Ω-loop bezeichnet.

Der Ω-loop ist als strukturelles Element nicht nur in den TEM β-Lactamasen, sondern auch in allen anderen β-Lactamasen der Klasse A konserviert (Vakulenko et al., 1995, 1999; Majiduddin und Palzkill, 2003) und wird durch eine Salzbrücke zwischen den Aminosäuren Arg164 und Asp179 in seiner Form fixiert. Der Grad seiner Flexibilität ist Gegenstand aktueller Diskussion: Einerseits bildet der Ω-loop nur wenige Kontakte mit dem Kern des Proteins und es wurde ihm daher eine erhöhte Flexibilität zugesprochen (Massova und Mobashery, 1998). Dies wurde durch eine molekulardynamische Untersu-

chung teilweise unterstützt, in der ein Teilbereich des Ω-loops als flexibel identifiziert wurde (Roccatano et al., 2005). Jedoch konnte in einer weiteren Simulation für den Ω-loop keine Abweichung von der Kristallstruktur beobachtet werden (Diaz et al., 2003). In fast allen Röntgenkristallstrukturen besitzt der Ω-loop zudem immer die gleiche Konformation und in einer Untersuchung mittels Kernresonanzspektroskopie konnte keine erhöhte Flexibilität festgestellt werden (Savard und Gagne, 2006).

1.2.3 Reaktionsmechanismus und die Funktion des Ω-loops

Der Reaktionsmechanismus der TEM β-Lactamasen verwendet das aktive Ser70 als Nukleophil für die Spaltung des β-Lactam-Rings und lässt sich in einen Acylierungs- und Deacylierungsschritt unterteilen (Matagne et al., 1998). Die Reaktion beginnt mit der Bildung eines nicht kovalenten Enzym-Substrat-Komplexes und führt über einen tetrahydralen Acylierungs-Übergangszustand zu einem Acyl-Enzym-Zwischenprodukt, in welchem das Substrat in Form eines Esters vorübergehend kovalent an das katalytische Serin gebunden ist. Anschließend erfolgt ein Angriff durch ein hydrolytisches Wasser auf das Acyl-Enzym-Zwischenprodukt und im weiteren Verlauf der Reaktion die Bildung eines tetrahydralen Deacylierungs-Übergangszustandes. Dessen Zerfall führt letztendlich zu einem hydrolysierten Produkt und zu einem regenerierten Enzym. Sowohl während des Acylierungs- als auch während des Deacylierungs-Übergangszustandes ist der Carbonylsauerstoff des β-Lactam-Rings negativ geladen. Er wird, ähnlich wie bei den Serin-Proteasen, über ein sogenanntes *oxyanion hole* stabilisiert. Dieses wird bei den TEM β-Lactamasen durch die Amidgruppen der Aminosäuren Ser70 und Ala237 gebildet (Page und Laws, 2000).

Dieser Reaktionsmechanismus benötigt sowohl im Acylierungs- wie auch im Deacylierungsschritt eine katalytische Base: Zum einen muss im Acylierungsschritt das Ser70 aktiviert werden, um die Amid-Bindung des β-Lactam-Rings nukleophil angreifen zu können, zum anderen benötigt das Wasser im Deacylierungsschritt eine Aktivierung, um die Hydrolyse des Esters durchführen zu können. Untersuchungen mittels ortsgerichteter Mutagenese und Strukturbestimmungen des Enzyms mit gebundenen Übergangszustand-Analoga identifizierten das konservierte Glu166 auf dem Ω-loop als katalytische Base für den Deacylierungsschritt (Adachi et al., 1991; Strynadka et al., 1992). Uneinigkeit

herrscht indes über die Antwort auf die Frage, welche Aminosäure als katalytische Base für den Acylierungsschritt fungiert. Diskutiert werden sowohl eine indirekte Aktivierung von Ser70 durch Glu166 über ein koordiniertes Wassermolekül (Lamotte-Brasseur et al., 1991), als auch eine direkte Aktivierung durch Lys73, dass sich in unmittelbarer Nähe des Ser70 befindet (Jelsch et al., 1993). Dieses müsste allerdings hierzu in einem deprotonierten Zustand vorliegen. Beide Theorien haben Unterstützung durch Strukturuntersuchungen (Minasov et al., 2002), Computersimulationen (Hermann et al., 2003, 2005) sowie Studien über den pKa Wert von Lys73 erhalten (Golemi-Kotra et al., 2004), wobei die Mehrzahl der Ergebnisse auf Glu166 als katalytische Base hindeut. Dies würde bedeuten, dass es sich bei der Katalyse durch TEM β-Lactamasen um einen symmetrischen Reaktionsmechanismus handelt. Mit der Positionierung der katalytischen Base Glu166 besitzt der Ω-loop somit eine wichtige Funktion für die Aktivität und den Reaktionsmechanismus der TEM β-Lactamasen.

1.2.4 Erweitertes Substratspektrum

Wie schon erwähnt, sind seit der Entdeckung der TEM-1 β-Lactamase 157 Varianten beschrieben worden. Substitutionen wurden bisher an 37 der insgesamt 263 Aminosäurepositionen beschrieben, wobei maximal sieben Substitutionen in einer einzelnen TEM-Variante auftraten. TEM-Varianten können sich funktionell von TEM-1 durch ein erweitertes Substratspektrum (*extended-spectrum beta-lactamase*, ESBL) oder eine Resistenz gegenüber β-Lactamase-Inhibitoren (*inhibitor-resistent*, IRT) unterscheiden (Bradford, 2001). Der ESBL Phänotyp zeichnet sich im Vergleich zu TEM-1 durch ein erweitertes Substratspektrum aus, was hauptsächlich auf eine räumliche Vergrößerung des aktiven Zentrums zurückzuführen ist. Diese TEM β-Lactamasen sind in der Lage, auch Cephalosporine der dritten und vierten Generation mit ihren sperrigen R1 Seitenketten zu binden und zu hydrolysieren. TEM β-Lactamasen mit ESBL Phänotyp traten kurz nach der Einführung von Ceftazidim und Cefotaxim in die klinische Praxis auf (Knothe et al., 1983). Nach wie vor stellen *K. pneumoniae* und *E. coli* mit ESBL TEM β-Lactamasen eine ernsthafte Gefährdung für Patienten auf Intensivstationen dar (Paterson et al., 2001).

Eine der am häufigsten auftretenden Substitutionen die zu einem ESBL Phänotyp führt,

ist der Austausch von Arginin durch Serin oder Histidin an der Position 164. Es wurde spekuliert, dass mit dem Wegfall der elektrostatischen Wechselwirkungen zwischen Arg164 und Asp179 die Fixierung des Ω-loops aufgehoben wird und sich dadurch die Flexibilität des Ω-loops erhöht (Vakulenko et al., 1999). Tatsächlich zeigt die Kristallstruktur der TEM-64 β-Lactamase, in der Arg164 durch ein Serin substitutiert ist, eine Konformationsänderung des Ω-loops für die Aminosäuren 167 bis 175. Dies hat eine Erweiterung des aktiven Zentrums zur Folge, so dass dieses mehr Raum für die sperrigen Seitengruppen der Dritt- und Viertgenerations-Cephalosporine bietet (Wang et al., 2002).

2 Ergebnisse und Diskussion

2.1 Konservierte Wassermoleküle stabilisieren den Ω-loop in Klasse A β-Lactamasen

(siehe Manuskript: *Conserved water molecules stabilize the Ω-loop in class A β-lactamases*, Seite 43)

Der Ω-loop ist ein in Klasse A β-Lactamasen hochkonserviertes strukturelles Element, dessen Aufgabe es ist, die katalytisch essentielle Aminosäure Glu166 zu positionieren (Vakulenko et al., 1995, 1999; Majiduddin und Palzkill, 2003). Da er nur wenige Kontakte mit dem Kern des Proteins ausbildet, wurde ihm eine erhöhte Flexibilität zugesprochen (Massova und Mobashery, 1998), was aber in Untersuchungen mittels Kernresonanzspektroskopie nicht bestätigt werden konnte (Savard und Gagne, 2006). Da bekannt ist, dass Wasser einen großen Einfluss auf die Struktur, Dynamik und Funktion von Proteinen hat (Mattos, 2002; Despa, 2005; Prabhu und Sharp, 2006; Helms, 2007), wurden konservierte Wassermoleküle in den Strukturen von verwandten Klasse A β-Lactamase-Familien identifiziert und ihr Einfluss auf die Stabilität des Ω-loop untersucht.

Die Kristallwassermoleküle aus insgesamt 49 hochaufgelösten Röntgenkristallstrukturen der β-Lactamase-Familien TEM (25 Strukturen), SHV (10 Strukturen) und CTX-M (14 Strukturen) wurden nach einer Überlagerung der Proteine auf die Koordinaten der C_α-Atome einer Cluster-Analyse unterzogen. Diese Methode ermöglicht es, Positionen von Wassermolekülen zu identifizieren, die in der Mehrzahl der untersuchten Strukturen besetzt sind. Es ist bekannt, dass sowohl die Auflösung als auch die Kristallisationsbedingungen einen Einfluss auf die Anzahl der Wassermoleküle in einer Röntgenkristall-

struktur haben. Um möglichst verlässliche Aussagen über konservierte Wassermoleküle in Klasse A β-Lactamasen treffen zu können, wurden daher alle verfügbaren Strukturen von TEM, SHV und CTX-M β-Lactamasen verwendet, die eine Auflösung von 2.2 Å oder höher besaßen. Insgesamt wurden 13 Wassermoleküle identifiziert, welche in mehr als 90% der 49 Strukturen konserviert waren; darunter auch zwei zu 100% konservierte Wassermoleküle. Für jedes der 13 konservierten Wassermoleküle wurden mögliche Wasserstoffbrücken zu benachbarten Aminosäuren und Wassermolekülen berechnet.

Es zeigte sich, dass bis auf vier konservierte Wassermoleküle an der Proteinoberfläche alle anderen konservierten Wassermoleküle mit verschiedenen loop-Elementen der Proteinstruktur assoziiert sind. Sechs der insgesamt 13 konservierten Wassermoleküle sind am Ω-loop lokalisiert. Der Ω-loop, gebildet durch die Aminosäuren 164 bis 179, formt mit den gegenüberliegenden Aminosäuren 65 bis 69 des Proteinkerns eine Aushöhlung, in der sich die sechs konservierten Wassermoleküle befinden. Diese Wassermoleküle bilden ein Wasserstoffbrückennetzwerk zwischen den Aminosäuren des Ω-loops und denen des Proteinkerns, wobei überwiegend Atome des Proteinrückgrades in die Ausbildung der Wasserstoffbrücken involviert sind. Dies bietet eine Erklärung für den Widerspruch, warum der Ω-loop eine geringe Flexibilität aufweist, obwohl er nur wenige Kontakte zum Rest des Proteins besitzt. Aus der großen Anzahl an konservierten Wassermolekülen in der Ω-loop-Aushöhlung und der von ihnen vermittelten Interaktionen mit dem Protein schließen wir, dass diese Ansammlung von konservierten Wassermolekülen den Ω-loop stabilisiert und mit dem Rest des Proteins verbindet. Diese Annahme ist in Übereinstimmung mit früheren Studien an anderen Proteinen, die eine Beteiligung von konservierten Wassermolekülen an der Positionierung und Stabilisierung von loop-Strukturen festgestellt haben (Loris et al., 1994; Ogata und Wodak, 2002). Das bevorzugte Auftreten von konservierten Wassermolekülen in der Nähe von Aminosäuren, die nicht Teil von α-Helices und β-Faltblattsträngen sind, wurde zudem durch eine statistische Analyse verschiedener hochaufgelöster Proteinstrukturen bestätigt. Dabei bilden konservierte Wassermoleküle genau die Wasserstoffbrücken zu Atomen des Proteinrückgrades aus, die ansonsten durch die benachbarten Aminosäuren in α-Helices und β-Faltblattsträngen abgesättigt sind (Park und Saven, 2005).

Eine in vohergehenden Arbeiten (Loris et al., 1994; Bottoms et al., 2006) festgestellte Korrelation zwischen dem Auftreten von konservierten Wassermolekülen in der Umge-

bung von konservierten Aminosäuren konnte bei den untersuchten Klasse A β-Lactamase-Strukturen generell nicht bestätigt werden. Es zeigte sich, dass selbst wenn Atome von Seitenketten in die Ausbildung einer Wasserstoffbrücke zu einem konservierten Wassermolekül beteiligt sind, eine Konservierung der Aminosäure nicht zwingend vorliegen muss. Die Aufrechterhaltung eines Wasserstoffbrückennetzwerkes ist somit trotz einer Substitution der Seitenkette möglich.

Bottoms et al. (2006) schlug vor, dass konservierte Wassermoleküle analog zu konservierten Aminosäuren in einem Mulisequenzalignement betrachtet werden können. Als Konsequenz aus dieser Analogie folgern wir, dass es auch möglich sein sollte, phylogenetische Untersuchungen auf Basis konservierter Wassermoleküle durchführen zu können. Die drei in dieser Arbeit untersuchten Familien der Klasse A β-Lactamasen besitzen zwar eine sehr ähnliche Struktur, zeigen aber eine nur mittelmäßige Sequenzidentität. Sequenzbasierte phylogenetische Studien an Klasse A β-Lactamasen haben ergeben, dass sich die Enzyme der CTX-M Familie evolutionär schon sehr früh von denen der TEM bzw. SHV Familie entfernt haben (Hall und Barlow, 2004). Diese evolutionäre Beziehung spiegelt sich interessanterweise auch in der Anzahl der Wassermoleküle wieder, die in den Strukturen der TEM und SHV Familien hochkonserviert sind, aber in den Strukturen der CTX-M Familie fehlen. Auch in Bezug auf die Anzahl der Wassermoleküle, die jeweils in nur einer Familie konserviert sind, besitzen die Strukturen der CTX-M β-Lactamasen, verglichen mit denen der TEM und SHV β-Lactamasen, die weitaus höchste Anzahl.

2.2 Untersuchungen zur Dynamik und Wasserbindung des Ω-loops in einer TEM β-Lactamase mittels multiplen molekulardynamischen Simulationen.

(siehe Manuskript: *Multiple molecular dynamics simulations of TEM β-lactamase: Dynamics and water binding of the Ω-loop*, Seite 69)

Um die Dynamik und Flexibilität des Ω-loops näher zu charakterisieren und den Einfluss konservierter Wassermoleküle darauf zu verstehen, wurden multiple molekulardy-

namische Simulationen der TEM β-Lactamase in Wasser als explizitem Lösungsmittel durchgeführt. Ausgehend von einer hochaufgelösten Röntgenkristallstruktur wurden zwei Systeme, eines mit und eines ohne das ursprüngliche Kristallwasser, erstellt. Für beide Systeme wurden jeweils 10 Simulationen von je 5 ns Dauer und mit unterschiedlichen Verteilungen der Anfangsgeschwindigkeiten berechnet. Dieser Ansatz der multiplen molekulardynamischen Simulationen wurde verwendet, da er eine effizientere Durchmusterung des Konformationsraumes gewährleistet, als eine einzelne Simulation mit entsprechend langer Simulationsdauer. Weiterhin erlaubt dieser Ansatz, die Reproduzierbarkeit und Signifikanz der Ergebnisse besser beurteilen zu können (Carlson et al., 1996; Daggett, 2000; Gorfe et al., 2002; Likic et al., 2005; Legge et al., 2006).

Generell konnte im Hinblick auf die Flexibilität und Dynamik sowie die Wasserbindungseigenschaften des Proteins kein signifikanter Unterschied zwischen den beiden Systemen festgestellt werden. Daraus folgt, dass sich die Systeme nach 5 ns Simulationszeit soweit relaxiert haben, dass die ursprünglichen Startbedingungen keinen Einfluss mehr haben. Wie aus dem Verlauf der mittleren Abweichung (*root-mean-square deviation*, RMSD) von der Kristallstruktur und den 2D-RMSD Darstellungen ersichtlich wurde, verhielt sich das Protein in allen Simulationen stabil. Mit der Ausnahme von wenigen, solvent-exponierten loop-Bereichen kam es zu keinen größeren Konformationsänderungen im Vergleich zur Startstruktur. Die Flexibilität und Dynamik des Proteins wurde durch die Berechnungen der mittleren Atomfluktuationen (*root mean square fluctuation*, RMSF) und des Ordnungsparameters S^2 für jede Aminosäure analysiert. Globale und konzertierte Bewegungen von mehreren Atomen wurden durch eine Hauptkomponentenanalyse (*essential dynamics*) charakterisiert. Alle drei Analysen weisen auf ein hoch geordnetes und rigides Protein hin, welches nur in den loop-Bereichen eine erhöhte Flexibilität zeigt. Dies deckt sich mit den Erkenntnissen aus zwei theoretischen Arbeiten (Diaz et al., 2003; Roccatano et al., 2005) und der Untersuchung der TEM-1 β-Lactamase mittels Kernresonanzspektroskopie (Savard und Gagne, 2006). Die in dieser Arbeit berechneten S^2-Werte stimmten qualitativ, für manche Aminosäuren sogar quantitativ, mit den experimentellen Werten überein.

Es stellte sich heraus, dass sich der Ω-loop in zwei Teilbereiche mit jeweils unterschiedlichen Flexibilitäten gliedert. Während die Aminosäuren 164 bis 172 sowie 178 und 179 durch ähnlich niedrige RMSF- bzw. hohe S^2-Werte wie im Inneren des Proteins gekenn-

zeichnet sind, weisen die Aminosäuren 173 bis 177 eine deutlich erhöhte Flexibilität auf. Distanzmessungen zwischen diesen Aminosäuren und den gegenüberliegenden Aminosäuren im Kern des Proteins bestätigten die Beweglichkeit dieses Teilstücks des Ω-loops. In vorhergehenden Arbeiten wurde die Flexibilität des Ω-loops unterschiedlich beurteilt: Während Diaz et al. (2003) keine besondere Abweichung von der Kristallstruktur feststellen konnte, beobachtete Roccatano et al. (2005) ein Umklappen dieses Ω-loop Teilstücks hin zum Kern des Proteins. Durch die Kombination mehrerer Simulationen konnte gezeigt werden, dass die Ergebnisse früherer Arbeiten keinesfalls im Widerspruch zueinander stehen, sondern jeweils nur unterschiedliche Teilschritte einer Bewegung beschreiben, deren Gesamtdauer über die einer einzelnen Simulation hinausgeht.

Wie an Hand von Kristallstrukturanalysen gezeigt wurde, formt der Ω-loop mit den gegenüberliegenden Aminosäuren 65 bis 69 eine Aushöhlung, in der sich sechs konservierte Wassermoleküle befinden (siehe Manuskript „*Conserved water molecules stabilize the Ω-loop in class A β-lactamases*“, Seite 43). Für vier dieser sechs konservierten Wassermoleküle konnte in den multiplen Simulationen beider Systeme reproduzierbar gezeigt werden, dass sie Wasserbrücken zwischen den Aminosäuren des Ω-loops und Wasserbrücken zwischen den Aminosäuren des Ω-loops und den gegenüberliegenden Aminosäuren des Proteinkerns ausbilden. Die seitens des Ω-loops in die Wasserbrücken involvierten Aminosäuren waren dabei ausschließlich in den rigiden Bereichen des Ω-loops zu finden. Die Mehrzahl der Wasserbrücken war zwischen 80-90% der analysierten Simulationszeit ausgebildet und die daran beteiligten Wassermoleküle zeigten meist nur eine geringe Austauschrate. Dies ist in Übereinstimmung mit experimentellen Arbeiten an anderen Proteinen, die zeigten dass Wassermoleküle im Inneren des Proteins mit den Atomen des Proteinrückgrades interagieren (Ernst et al., 1995) und dass deren Verweildauer im Vergleich zu Wassermolekülen an der Proteinoberfläche sehr viel länger ist (Otting et al., 1991). In allen Simulationen, die ohne das ursprüngliche Kristallwasser gestartet wurden, erfolgte innerhalb der ersten 200 ps eine rasche Solvatisierung der Ω-loop-Aushöhlung und damit die Ausbildung der Wasserbrücken. Die hohe Verweildauer der Wassermoleküle in der Ω-loop-Aushöhlung, die von diesen Wassermolekülen vermittelten Wasserbrücken und die rasche Solvatisierung der Ω-loop-Aushöhlung unterstützen die Theorie, dass der Ω-loop durch eine Ansammlung von konservierten Wassermolekülen stabilisiert und mit dem Rest des Proteins verbunden wird. Experimentell wurde bereits an einem anderen Protein gezeigt, dass Wassermoleküle, die im Inneren des Proteins gebunden

sind, zur Stabilität des Proteins beitragen (Takano et al., 2003). Die fehlende Stabilisierung durch Wasserbrücken würde damit auch die erhöhte Flexibilität des von den Aminosäuren 173 bis 177 gebildeten Teilbereich des Ω-loops erklären.

Weiterhin wurde beobachtet, dass ein Austausch der Wassermoleküle in der Ω-loop-Aushöhlung mit dem umgebenden Lösungsmittel auch durch eine sich spontan öffnende Lücke zwischen den Seitenketten der Aminosäuren Arg164, Ile173, Asp176 und Arg178 stattfinden kann. Dies war bisher aus der Betrachtung der Kristallstrukturen nicht ersichtlich und gibt einen Hinweis auf die mögliche Wirkungsweise der Mutation Arg164Ser/His, welche sich häufig in ESBL-Varianten der TEM β-Lactamasen findet. Es ist bekannt, dass mit dem Verlust der Salzbrücke Arg164–Asp179 eine Konformationsänderungen des Ω-loops einhergeht. So zeigt die Kristallstruktur der TEM-64 β-Lactamase, in der Arg164 durch ein Serin substitutiert ist, eine Konformationsänderung des Ω-loops für die Aminosäuren 167 bis 175 (Wang et al., 2002). Eine Erklärung für diese Konformationsänderung wäre, dass der Austausch der Arginin-Seitenkette mit einer Seitenkette geringerer Größe eine häufigere, wenn nicht gar permanente Ausbildung der oben genannten Lücke verursacht. Dadurch wäre ein Austreten der Wassermoleküle aus der Ω-loop-Aushöhlung erleichtert, was mit einem Verlust der stabilisierenden Wirkung dieser Wassermoleküle einhergehen würde.

2.3 Automatisierung von multiplen molekulardynamischen Simulationen in Cluster- und Grid-Umgebungen

(siehe Veröffentlichungen: *A management system for complex parameter studies and experiments in grid computing*, Seite 101 und *GriCoL: A language for scientific grid*, Seite 107)

Molekulardynamische Simulationen von Proteinen und den sie umgebenden Lösungsmittelsystemen sind sehr rechenintensiv, da die zu untersuchenden Systeme eine große Anzahl von Atomen umfassen (Größenordnung 10^4-10^5 Atome) und sehr langsam relaxie-

ren (10^6-10^7 Iterationsschritte, entsprechend 1-10 ns). Von jedem System müssen zudem mehrere Simulationen mit einer unterschiedlichen Verteilung der Anfangsgeschwindigkeiten durchgeführt werden, um ein effizientes Sampling des Konformationsraumes zu gewährleisten und um die Reproduzierbarkeit und Signifikanz der Ergebnisse beurteilen zu können (*multiple molecular dynamics*, MMD). Dies erhöht die benötigte Rechenzeit weiter und stellt eine logistische Herausforderung in Bezug auf die Vorbereitung, Durchführung und Auswertung der zahlreichen Simulationen dar, die nur noch sehr schwer von Hand zu bewältigen ist. Die benötigte Rechenzeit wird heutzutage entweder von lokal betriebenen Rechnerverbänden (Cluster), die die Rechenkraft einzelner Computer bündeln, oder von Grid-Umgebungen bereitgestellt. Eine Grid-Umgebung ist eine Infrastruktur, die eine integrierte, gemeinschaftliche Verwendung von meist geographisch getrennten Ressourcen, wie zum Beispiel einzelnen Clustern, erlaubt (Foster und Kesselman, 1999). Jedoch stellen Cluster und Grid-Umgebungen meist keine auf spezielle Anwendungen zugeschnittenen Werkzeuge zur Verfügung, so dass eine effiziente Gestaltung und automatisierte Durchführung von Arbeitsabläufen nur durch erheblichen Programmieraufwand möglich ist.

Das am Höchstleistungsrechenzentrum Stuttgart (HRLS) entwickelte *Science Experimental Grid Laboratory* (SEGL) ist ein System zur Durchführung von Parameterstudien und automatisierten Arbeitsabläufen mit einer Anbindung an Cluster- und Grid-Umgebungen (Currle-Linde et al., 2004, siehe Seite 101 und Currle-Linde et al., 2005). Es verwendet die durch eine grafische Benutzeroberfläche unterstützte Grid-Sprache GriCol (Currle-Linde et al., 2006, siehe Seite 107) und ermöglicht den modularen Aufbau komplexer Simulationsexperimente und Arbeitsabläufe (*control flows*) aus einzelnen Funktionsbausteinen (*blocks*). Ein *control flow* ist jeweils aus mehreren *solver blocks*, die für die Ausführung eines Programms verantwortlich sind, und *control blocks*, die Kontrollstrukturen wie Entscheidungen oder Schleifen zur Verfügung stellen, aufgebaut. Jedem *solver-* und *control-block* liegt dabei ein individuell konfigurierbarer Datenfluss (*data flow*) zu Grunde, der über verschiedene Module gesteuert werden kann.

Der gesamte Ablauf einer MMD-Simulation wurde erfolgreich in SEGL implementiert und getestet. Die einzelnen Arbeitsschritte wie Vorbereitung, Durchführung und Auswertung der Simulationen wurden in einem *control flow* abgebildet und die dazu notwendigen *solver-* und *control blocks* erstellt und konfiguriert. Zur Entfernung energetisch

ungünstiger Startbedingungen, wie deformierte Bindungslängen, Bindungswinkel oder van-der-Waals Interaktionen, wurde zuerst ein *solver block* verwendet, der eine Energieminimierung durchführt. Anschließend wurde das Proteinrückgrat mit einem harmonischen Potential fixiert, um das System an die Simulationsparameter wie Druck und Temperatur anzupassen. Die Kraftkonstanten für das Potential wurden von anfänglich 10 kcal/mol in vier Schritten auf Null reduziert. Während in Currle-Linde et al., 2004 (siehe Seite 101) hierzu noch vier *solver blocks* zum Einsatz kamen, wurden diese vier Schritte in Currle-Linde et al., 2006 (siehe Seite 107) unter Verwendung eines Parametermoduls in einen einzigen *solver block* integriert. Daraufhin folgten weitere Simulationsabschnitte, um das System in einen Gleichgewichtszustand zu bringen sowie die Produktionsphase, in der die Daten für nachfolgende Auswertungen generiert werden. Für jeden Typ von Auswertung wurde ein eigener *solver block* verwendet. Auswertungen welche die Daten mehrerer Simulationen miteinander kombinieren, wurden mit einem *control block* realisiert. Dieser stellt sicher, dass erst alle Simulationen vollständig abgeschlossen sind, bevor die Auswertung gestartet wird.

Der so erstellte *control flow* wurde mit der Topologie und den initialen Koordinaten einer TEM β-Lactamase auf einer Cluster-Umgebung gestartet und erfolgreich durchlaufen. Das Starten und Überwachen der einzelnen Berechnungen wurde dabei vollständig und ohne Interaktion des Benutzers von SEGL übernommen. Alle bei diesem Ablauf erzeugten Daten wurden in einer Datenbank gespeichert. Es zeigte sich, dass die Erstellung und automatisierte Durchführung von MMD-Simulationen mit Hilfe von SEGL gegenüber der manuellen Durchführung einzelner Simulationen eine enorme Zeitersparnis ermöglichte. Ein weiterer Vorteil bei der Verwendung von SEGL ist die Fehlerminimierung, da die Erstellung aller notwendigen Eingabedateien und die Programmaufrufe durch SEGL erfolgen. Zudem bietet SEGL die Möglichkeit, funktionsfähige Abläufe mit anderen Benutzern auszutauschen. Die Ablage aller Dateien und Simulationsparameter in einer auf das jeweilige Simulationsexperiment zugeschnittenen objekt-orientierten Datenbank stellt sicher, dass auch zu einem späteren Zeitpunkt alle Daten geordnet zur Verfügung stehen.

2.4 Verwendung eines haptischen Eingabegerätes zur Untersuchung der Proteindynamik

(siehe Veröffentlichung: *Time-based haptic analysis of protein dynamics*, Seite 115)

Neben der Struktur stellen die Flexibilität und Dynamik wichtige Determinanten für die Funktionsweise von Proteinen dar. Eine integrierte Betrachtung von Proteinstruktur, -dynamik und -funktion ist daher unumgänglich. Die Methode der molekulardynamischen Simulation (MD-Simulation) ist eines der wichtigsten theoretischen Werkzeuge zur Untersuchung der Proteindynamik. MD-Simulationen ermöglichen es, detaillierte Aussagen über die Struktur, Dynamik und Thermodynamik biologischer Moleküle und ihrer Komplexe zu machen. Bei der Untersuchung der Flexibilität müssen die unterschiedlichen Zeitskalen berücksichtigt werden, auf denen sich Proteindynamik abspielen kann: Schnelle, unkorrelierte Bewegungen einzelner Atome und Aminosäuren sind zu unterscheiden von langsamen, korrelierten Bewegungen ganzer Proteinbereiche oder -domänen.

Die Flexibilität einzelner Atome ist in MD-Simulationen über die isotrope Berechnung der mittleren quadratischen Abweichung von ihren Gleichgewichtspositionen zugänglich und wird als *root mean square fluctuation* (RMSF) bezeichnet. Dieser Wert ist mit dem aus Röntgenkristallstrukturen erhaltenen B-Faktor vergleichbar, welcher ein Maß für die Unordnung in einem Proteinkristall und damit auch ein Maß für die Flexibilität einzelner Atome darstellt. Die Visualisierung der aus MD-Simulationen berechneten Proteinflexibilität erfolgt klassischerweise durch Auftragen der RMSF-Werte gegen die Atom- oder Aminosäurenummer, durch ein farbkodiertes Aufblenden auf die Proteinstruktur oder durch das Anzeigen korrelierter Bewegungen mit Hilfe von Pfeilen. All dies sind jedoch indirekte und nicht zeitabhängige Betrachtungsweisen der Proteinflexibilität, da sie erst aus MD-Simulationen berechnet werden müssen, und zeitlich gemittelte Eigenschaften darstellen. Eine direkte, zeitabhängige Analyse der Proteinflexibilität, die über das bloße Betrachten von Atomen im dreidimensionalen Raum hinausgeht, ist bisher nicht möglich.

In Zusammenarbeit mit dem Institut für Visualisierung und Interaktive Systeme wur-

de daher eine neuartige Methode zur haptischen Interaktion und Analyse von MD-Simulationen entwickelt. Hierzu wurde ein haptisches Eingabegerät (PHANTOM[1]) mit dem Visualisierungsprogramm VMD (Humphrey et al., 1996) gekoppelt. Das PHANTOM ist ein Eingabegerät, dass Bewegungen in sechs Freiheitsgraden ermöglicht und über einen stiftähnlichen Griff bedient wird. Dies erlaubt es dem Benutzer, direkt mit dem Protein zu interagieren, sozusagen das Protein während des Abspielens der Simulation zu „erfühlen“ und sich an einzelnen Atomen „festzuhalten“. Beim Abspielen einer Trajektorie und gleichzeitigem Festhalten an einem Atom bewegt sich dann die Hand des Benutzers in Übereinstimmung mit den anisotropen Bewegungen, die das Atom während der MD-Simulation durchläuft. Um die Implementierung und die Machbarkeit einer haptischen Analyse von Trajektorien zu validieren, wurde eine MD-Simulation einer TEM β-Lactamase durchgeführt. Ausgehend von einer hochaufgelösten Röntgenkristallstruktur (Minasov et al., 2002) wurde eine 3 ns lange Trajektorie erzeugt und zunächst mit herkömmlichen Methoden analysiert. Im Verlauf der MD-Simulation erwies sich das Protein als stabil und zeigte eine Abweichung von nur 0.9–1.1 Å von der Kristallstruktur. Die Berechnung des RMSF bzw. der B-Faktoren für alle Atome ergab eine typische Verteilung von mehr und weniger flexiblen Bereichen. Das Innere des Proteins ist dabei charakterisiert durch Atome mit geringer Flexibilität, während lösungsmittelexponierte Atome an der Proteinoberfläche eine höhere Flexibilität aufweisen. Diese Ergebnisse konnten mit der haptischen Betrachtung der Simulation nachvollzogen und bestätigt werden. Es waren haptisch deutliche Unterschiede wahrnehmbar, je nachdem ob zum Beispiel ein Atom aus dem Inneren des Proteins oder an der lösungsmittelexponierten Oberfläche ausgewählt wurde.

Bisherige Arbeiten auf dem Gebiet der haptischen Analyse von (bio-)molekularen Strukturen beschränken sich entweder auf die haptische Darstellung von Eigenschaften (Křenek, 2000, 2003) und Konformationsübergängen (Křenek, 2001) kleiner chemischer Moleküle oder auf erste Ansätze im Bereich des haptischen Dockings (Maciejewski et al., 2005). Ein weiteres Einsatzgebiet für das haptische Eingabegerät in Kombination mit VMD stellt die bereits erfolgreich praktizierte, interaktiv gesteuerte MD-Simulation dar (Stone et al., 2001). Hierbei wird während der MD-Simulation mit dem Protein interagiert, und es werden gezielt Simulationsparameter verändert. Um die Auswirkungen der Benutzerinteraktion auf die Simulation zeitnah zu berechnen und darzustellen ist je-

[1] siehe http://www.sensable.com

doch ein enormer Rechenaufwand notwendig. Eine nachträgliche Analyse der Trajektorie mit dem haptischen Eingabegerät war aber bisher auch bei der interaktiv gesteuerten MD-Simulation nicht möglich.

Mit dieser Arbeit werden zwei neue Methoden der haptische Analyse von Trajektorien und Proteinstrukturen in VMD zur Verfügung gestellt. Erstens ist es durch das Zuweisen von Oberflächeneigenschaften, wie z.B. glatt oder rau, an die betrachtete Proteinstruktur möglich, zusätzliche Informationen zu übermitteln. Ein Beispiel wäre die Einfärbung des Proteins nach Hydrophobie und die Kodierung der Ladung über die Oberflächeneigenschaften. Zweitens ist es nun möglich, die Proteinflexibilität in MD-Simulationen direkt und zeitabhängig zu analysieren. Im Gegensatz zur herkömmlichen Darstellung der Atomflexibilität als skalare Größe ohne Berücksichtigung der Bewegungsrichtung, übermittelt der haptische Ansatz dem Benutzer diese zusätzliche Information. Durch Nutzen einer Filterfunktion können schnelle, unkorrelierte Bewegungen geglättet und nur langsame Bewegungen auf größeren Zeitskalen betrachtet werden. Dadurch sind auch Einblicke in die Bewegung ganzer Proteinbereiche wie Sekundärstrukturelemente oder Proteindomänen möglich. Jedoch gestaltet sich die absolute, quantitative Wahrnehmung der Flexibilität noch als schwierig. Ebenso ist unklar, wieviele verschiedene Größenordnungen von Flexibilität ein Benutzer unterscheiden kann. Um diese Frage zu klären sind weitere Nutzerstudien und Untersuchungen notwendig.

3 Darstellung der Ergebnisse als Publikationsmanuskripte und Publikationen in englischer Sprache

1. Conserved water molecules stabilize the Ω-loop in class A β-lactamases. Seite 43

2. Multiple molecular dynamics simulations of TEM β-lactamase: Dynamics and water binding of the Ω-loop. Seite 69

3. A management system for complex parameter studies and experiments in grid computing. Seite 101

4. GriCoL: A language for scientific grid. Seite 107

5. Time-based haptic analysis of protein dynamics. Seite 115

3.1 Conserved water molecules stabilize the Ω-loop in class A β-lactamases

3.1.1 Abstract

A set of 49 high-resolution (≤ 2.2 Å) structures of the TEM, SHV, and CTX-M class A β-lactamase families was systematically analyzed to investigate the role of conserved water molecules in the stabilization of the Ω-loop. Overall, 13 water molecules were found to be conserved in at least 45 structures, including two water positions which were found to be conserved in all structures. From the 13 conserved water molecules, six are located at the Ω-loop, forming a dense cluster with hydrogen bonds to residues at the Ω-loop as well as to the rest of the protein. This layer of conserved water molecules, packed between the Ω-loop and the rest of the protein acts as structural glue and explains why the Ω-loop exhibits only a low degree of flexibility in experimental studies, despite having few contacts with the rest of protein. A correlation between conserved water molecules and conserved protein residues could in general not be detected, with the exception of the conserved water molecules at the Ω-loop. Further, the evolutionary relationship between the three families derived from the number of conserved water molecules is similar to the relationship derived from phylogenetic analysis.

Keywords

Conserved water molecules; β-lactamases, Ω-loop, protein families, TEM, SHV, CTX-M, conserved hydrogen bond network, evolution

Abbreviations

CWM, conserved water molecule

3.1.2 Introduction

β-lactamases (EC 3.5.2.6) hydrolyze the β-lactam ring and therefore are the main cause for resistance against β-lactam antibiotics. A sequence based classification scheme was initially proposed by Ambler (Ambler, 1980) and soon thereafter extended by others (Jaurin and Grundstrom, 1981; Ouellette et al., 1987). Class A β-lactamases include the plasmid-borne TEM, SHV, and CTX-M β-lactamases which are most commonly found in gram negative bacteria. The TEM and SHV β-lactamase families are closely related (sequence identity of 67%), while the CTX-M family is more distant (sequence identity of 40% to TEM and SHV) (Tzouvelekis et al., 2000). The members of the TEM and SHV families differ by only one to seven amino acids, while the members of the CTX-M family are more diverse (80% sequence identity) (Jacoby, 2006). Because of their importance in antibiotic resistance, the enzymes of class A β-lactamases are well studied and characterized (Massova and Mobashery, 1998; Matagne et al., 1998; Bradford, 2001; Poole, 2004; Fisher et al., 2005; Babic et al., 2006), and a vast amount of crystallographic data is available. While sequence similarity between the protein families is moderate, the structures of all TEM, SHV, and CTX-M β-lactamases show a remarkably high similarity. The enzymes are composed of two domains that are closely packed together: An all-α-domain consisting of 8 α-helices, and an α-/β-domain, consisting of 3 α-helices and 5 β-strands (Jelsch et al., 1992). The active site cavity is part of the interface between the two domains and is limited by the Ω-loop. The Ω-loop (residues 161 to 179) is a conserved structural element in all class A β-lactamases and is directly involved in the catalytic reaction of the enzymes because it positions the general base Glu166. Its conformation is anchored by a highly conserved salt bridge formed between Arg164 and Asp179 (Vakulenko et al., 1995, 1999; Majiduddin and Palzkill, 2003). However, the long Ω-loop has only few contacts with the rest of the protein and therefore was speculated to be a flexible element (Massova and Mobashery, 1998). Although a recent molecular dynamics study (Roccatano et al., 2005) attributes some flexibility to a part of the Ω-loop, the degree of flexibility associated with the Ω-loop is under debate, as in nearly all X-ray structures of the class A β-lactamases the Ω-loop has the same conformation. From a recently published NMR study of the TEM-1 β-lactamase it is evident that the protein is one of the most ordered proteins studied by high resolution NMR to date, and there was no indication of increased flexibility of the Ω-loop (Savard and Gagne,

2006).

One reason for the low flexibility of the Ω-loop in the absence of protein-protein interaction could be stabilization by water molecules. Water is known to have a crucial role in protein structure, flexibility and activity (Mattos, 2002; Despa, 2005; Prabhu and Sharp, 2006; Helms, 2007). Water molecules bind via hydrogen bonds to the side chain and backbone atoms of proteins (Park and Saven, 2005), to polar atoms of substrates and ligands (Lu et al., 2007), as well as to other water molecules. This enables water to mediate protein-protein and protein-ligand contacts, and to take part in enzyme catalysis. X-ray crystallography has long been used to analyze water at protein surfaces, since crystal structures determined at high resolution provide a detailed picture of protein hydration (Nakasako, 2004). Comparative studies of crystal structures showed that there are water binding sites on the surface and the interior of a protein which are occupied by a water molecule in different crystal structures. Cluster analysis and density based approaches have been successfully used for the identification and analysis of these conserved water molecules (CWM) in multiple crystal structures of a protein or in crystal structures of a protein family (Sanschagrin and Kuhn, 1998; Bottoms et al., 2002; Ogata and Wodak, 2002; Mustata and Briggs, 2004; Bottoms et al., 2006).

For class A β-lactamases crystallographic data supports the conservation of at least one water molecule in the active site which is known to be essential to the enzymatic reaction (Matagne et al., 1998). But so far, little is known about conserved water molecules outside of the active site. In the present study we systematically analyzed high resolution crystal structures of the evolutionary related TEM, SHV, and CTX-M class A β-lactamase families to determine conserved water molecules and investigate their role in protein structure and especially the stabilization of the Ω-loop.

3.1.3 Material & methods

Data preparation

Crystal structures of the TEM, SHV, and CTX-M β-lactamase families were retrieved from the Protein Data Bank (PDB) (Berman et al., 2000). β-lactamase variants with

core disrupting mutations and β-lactamases bound to inhibitory proteins were excluded. As it was found that the amount of crystal water depends on the atomic resolution of the crystal structure (Carugo and Bordo, 1999), only structures with an atomic resolution of 2.2 Å or higher were included. PDB entries containing more than one β-lactamase were pre-processed to extract the coordinates of single proteins including water molecules within a radius of 6 Å around the protein. This pre-processing was necessary for most of the CTX-M crystal structures. Thus, 25 protein structures of the TEM β-lactamase family (PDB entries 1AXB, 1BT5, 1BTL, 1ERM, 1ERO, 1ERQ, 1ESU, 1FQG, 1JVJ, 1JWP, 1JWV, 1JWZ, 1LHY, 1LI0, 1LI9, 1M40, 1NXY, 1NY0, 1NYM, 1NYY, 1TEM, 1XPB, 1YT4, 1ZG4, 1ZG6), 10 protein structures of the SHV β-lactamase family (PDB entries 1N9B, 1ONG, 1Q2P, 1RCJ, 1SHV, 1TDG, 1TDL, 1VM1, 2A3U, 2A49), and 14 protein structures of the CTX-M β-lactamase family (PDB entries 1YLJ, 1YLP, 1YLT, 1YLW, 1YLY, 1YLZ, 1YM1, 1YMS, 1YMX) were analyzed. A detailed table of all 49 structures with their original references, information on atomic resolution, free R-factor, number of water molecules, mutations, and active side occupancy is available as supplementary material.

Identification of conserved water molecules

All structures including the crystal water molecules were superimposed by multiple-structure fitting of the C_α-atoms using the McLachlan algorithm (Mclachlan, 1982) as implemented in the program ProFit[1]. The fitting zones were initially determined by a Needleman-Wunsch sequence alignment and then iteratively adapted during the fitting process. Subsequently, a complete-linkage hierarchical cluster analysis of the water molecules in all 49 superimposed structures was performed using the program WatCH (Sanschagrin and Kuhn, 1998). The clustering process groups together water molecules from different structures that physically overlap and represent one specific water position. The default cut-off value of 2.4 Å for the maximum inter-water distance was used. With this cut-off value, water molecules with a center-to-center distance of 2.4 Å will overlap by 50% if an approximate effective radius of 1.6 Å is assumed. It ensures that clusters do not contain several water molecules from the same crystal structure. The

[1]ProFit, version 2.5.4, Martin, A.C.R., http://www.bioinf.org.uk/software/profit

cluster analysis results in water molecule positions with a conservation ranging from 0.02% (found in 1 out of 49 structures) to 10% (found in all structures). Water positions found to be occupied in at least 45 out of 49 analyzed structures (conservation > 90%) were labeled conserved water molecules (CWMs).

Visualization of results and identification of hydrogen bonding partner

Visualization of the results and preparation of the figures was done using PyMOL (DeLano, 2002). As representative structures of the TEM, SHV, and CTX-M families the PDB entries 1M40, 1TDL, and 1YLT, respectively, were chosen. Hydrogen bonding partners of the CWMs were identified by determining protein nitrogen and oxygen atoms as well as other CWMs within a distance of 4.0 Å in the representative structures. The conservation of hydrogen bonding partners was derived from the multiple sequence alignment by T-COFFEE (Notredame et al., 2000) of all 49 proteins (provided as supplementary material). To identify crystal contacts mediated by CWMs, all structures were analyzed for hydrogen bonds between CWMs and atoms of symmetry-related proteins.

3.1.4 Results

Identification of conserved water molecules in class A β-lactamase crystal structures

49 crystal structures from three class A β-lactamase families with a resolution of at least 2.2 Å were superimposed and analyzed for conserved water molecules (CWMs): 25 structures from the TEM family, 10 structures from the SHV family, and 14 structures from the CTX-M family. In 32 of the 49 structures an inhibitor or substrate molecule is bound to the active site. The number of crystal water molecules in the structures of the TEM and SHV families ranges from 63 to 566 and 81 to 414, with an average of 271 and 280, respectively. The structures of the CTX-M family contain more crystal water molecules: 304 to 489 with an average of 408. Subsequent cluster analysis resulted in a total of 1993 water molecule positions. Only 13 of them were found in at least 45 out

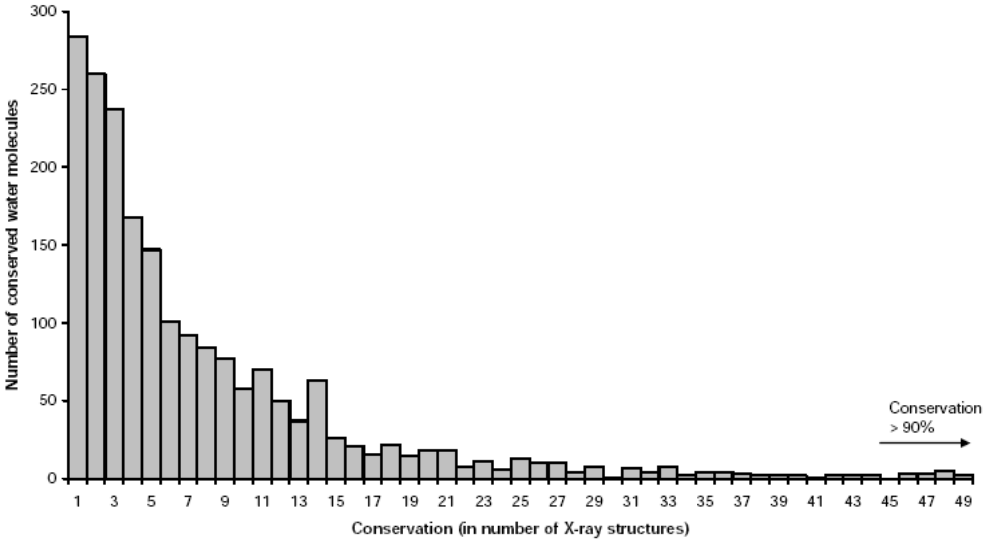


Figure 1: The number of water molecules (y-axis) found to be conserved in a given number of crystal structures (x-axis).

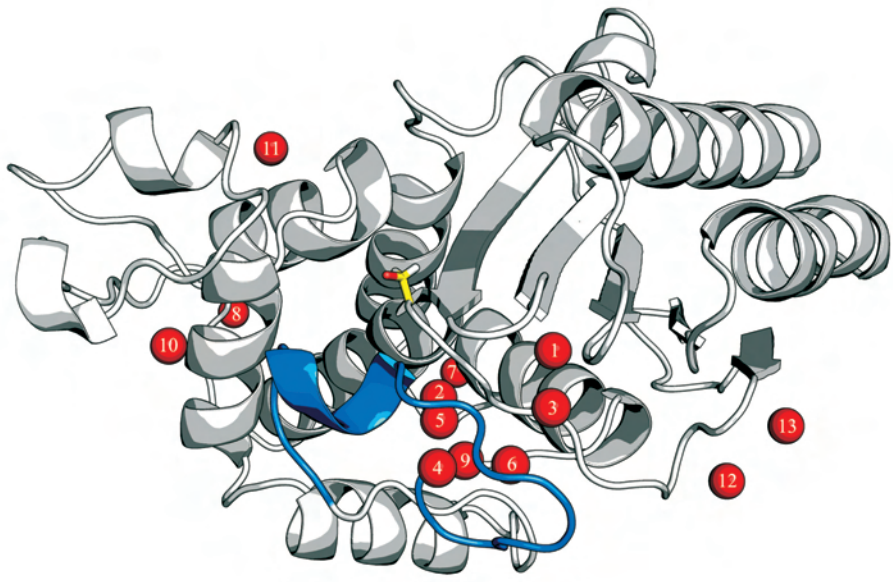


Figure 2: Water molecules conserved in more than 90% of all analyzed structures, shown together with a representative structure of the TEM β-lactamases (in grey). The Ω-loop is highlighted in blue, the active Ser70 is shown as yellow sticks, and the CWMs are shown as red spheres numbered according to table 1.

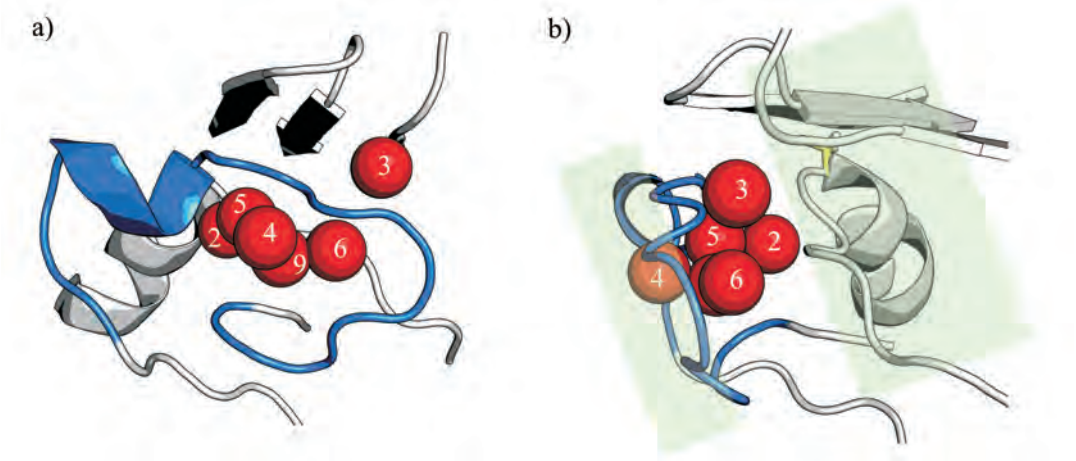


Figure 3: The Ω-loop (residues 164 to 179, in blue) and the rest of the protein (residues 65 to 70, in grey) form the two outer layers of a sandwich-like structure (highlighted in light green) with a layer of six highly conserved water molecules between them serving as structural glue. a) Front and b) side view. The numbering of the conserved water molecules is according to table 1.

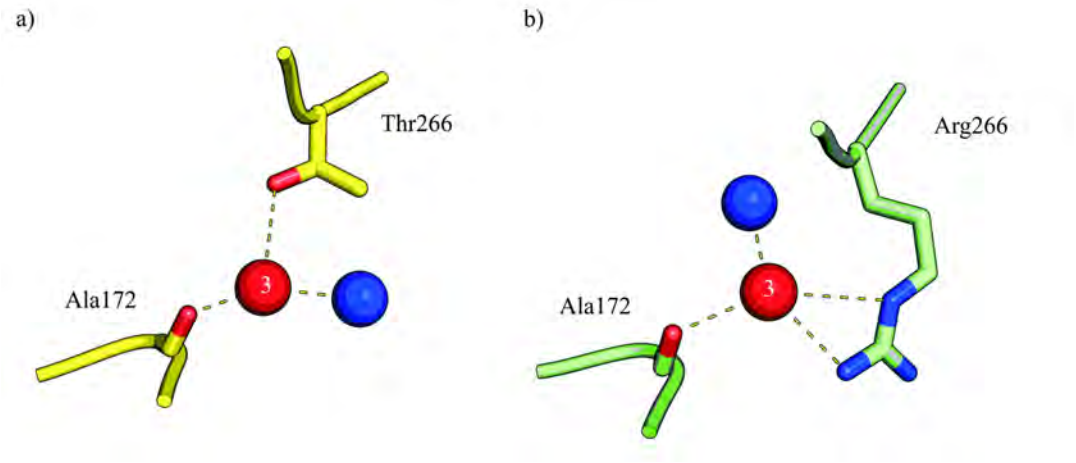


Figure 4: CWM3 (red sphere) forms hydrogen bonds with the carbonyl atom of Ala172 on the Ω-loop, with the side chain atom of the protein core (Thr266 in TEM and CTX-M, Arg266 in SHV), and with a water molecule (blue sphere). a) In TEM and CTX-M this water molecule is located at a position at which in SHV an arginine side chain is located. b) In contrast, in SHV this water molecule is located at a position at which in TEM and CTX-M a threonine side chain is located.

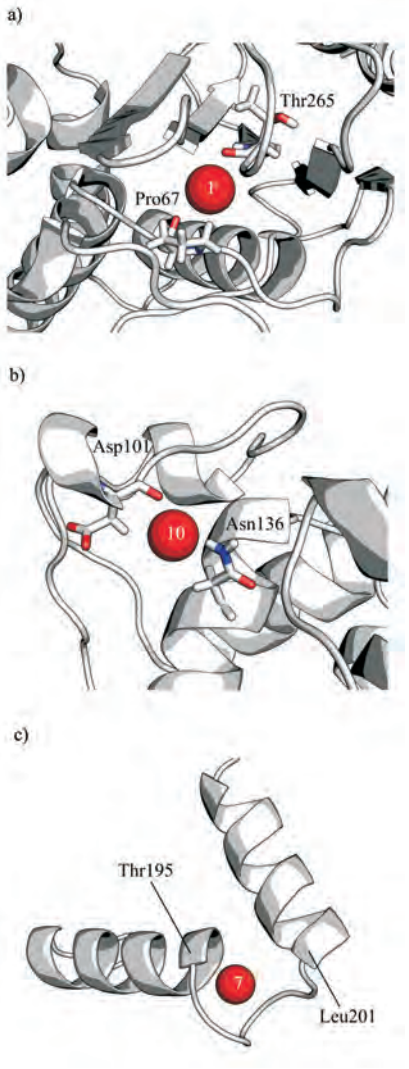


Figure 5: Loop-stabilizing conserved water molecules: a) CWM1 is buried in the inside of the protein and forms hydrogen bonds to the main chain atoms of residue 67 and 265. b) CWM10 forms hydrogen bonds to the carbonyl atom of Asp101 and the side chain nitrogen atom of Asn136. Both residues are 100% conserved throughout the three families. c) CWM7 is situated in a small loop comprising residues 195 to 201. The loop region forms a turn of more than 90 degrees between two helices.

of 49 analyzed structures (figure 1) and were labeled conserved water molecule CWM1 to CWM13 (table 1), including two water positions which were found to be conserved in all analyzed structures.

CWMs are found everywhere in the protein structure and can be assigned to two groups (figure 2): Group 1 consists of six CWMs interacting with the Ω-loop. Group 2 comprises the remaining seven CWMs, which are located in the interior of the protein and at the surface of the enzyme. None of the CWMs of group 1 is involved in forming crystal contacts between symmetry-related proteins. Even those CWMs of group 2 that are located at the protein surface mediate crystal contacts in only a minor fraction of the analyzed structures.

Conserved water stabilizing the Ω-loop (group 1)

Six CWMs interact with the Ω-loop and are located in a narrow, tunnel–shaped cavity lined by the Ω-loop (residues 164 to 179) and residues 65 to 69, which form a loop near the catalytic Ser70 and Lys73. The CWMs 2, 3, 4, 5, 6, and 9 form a hydrogen bond network between themselves, the residues of the Ω-loop, and the loop near the catalytic Ser70 (figure 3). The structure of TEM-64 (PDB entry 1JWZ) is the only structure that does not contain any of the conserved Ω-loop water molecules besides CWM2.

CWM2 is one of the two water molecules that are conserved in all structures. It forms four hydrogen bonds to backbone atoms: Two with residues of the Ω-loop (Leu169, Asp179) and two with residues of the protein core (Met68, Met/Cys69), as well as an additional hydrogen bond to the adjacent CWM5. CWM5 forms four hydrogen bonds to residues of the Ω-loop: Three to backbone atoms (Leu169, Glu171, and Ala172) and one to side chain atoms (Arg164). Two additional hydrogen bonds are formed to the adjacent CWM4 and CWM9. CWM4 forms four hydrogen bonds to residues of the Ω-loop: Two to backbone atoms (Ala172, Ile173) and two to side chain atoms (Arg164, Asp176). An additional hydrogen bond is formed to the adjacent CWM9. In the CTX-M family, the side chain of Thr171 is also a hydrogen bond partner. CWM9 forms three hydrogen bonds to residues of the Ω-loop: To backbone atoms of Ala172, Ile173, and Arg178 in the TEM family and to backbone atoms of Asp176 and Arg178 in the SHV

and CTX-M families. In the latter families, also side chain atoms of Arg164 are hydrogen bond partners. Additional hydrogen bonds are formed to the adjacent CWM4, 5, and 6. CWM6 forms two hydrogen bonds to residues of the Ω-loop: Two to backbone atoms (Ile/Leu173, Asp176) and in the SHV family one additional hydrogen bond to the side chain atoms of Arg266. CWM3 forms three hydrogen bonds: With the carbonyl atom of Ala172 on the Ω-loop, with the side chain atom of the protein core (Thr266 in TEM and CTX-M, Arg266 in SHV), and with a water molecule. In TEM and CTX-M this water molecule is located at a position at which in SHV an arginine side chain (figure 4a) is located. This water molecule is conserved inside TEM and CTX-M to 52% and 100%, respectively. In contrast, in SHV this water molecule is located at a position where in TEM and CTX-M a threonine side chain is located (figure 4b), and is conserved to 100% inside this family.

Conserved water molecules not associated with the Ω-loop (group 2)

Of the remaining seven CWMs, three are located in the protein interior and are involved in the stabilization of other loop regions. CWM1 is buried in the inside of the protein and conserved in all structures (figure 5a). It forms hydrogen bonds to backbone atoms (Pro67 in all families, Thr/Leu/Phe265 in TEM/SHV/CTX-M, and Arg244 in TEM/SHV) and in CTX-M to side chain atoms of Asn245. It bridges the loop of residues 65 to 69 with the second β-strand in the α/β domain. CWM10 forms two hydrogen bonds: One to the carbonyl of Asp101 and one to the side chain of Asn136 (figure 5b), and it links the long loop of residue 91 to 119 to the helix on which Asn136 is located. Asp101 and Asn136 are completely conserved in the three families. CWM7 is located inside a short loop region between two perpendicular helices (figure 5c). This turn is stabilized by hydrogen bonds to the backbone atoms of the loop residues 195 to 201.

Four CWMs are located at the surface of the enzyme. CWM8 forms one hydrogen bond to the backbone nitrogen of Leu91 in all families. Leu91 is the only conserved amino acid in the loop region between residues 87 and 99. The structures of the CTX-M family have more hydrogen bonding partners, as their loop conformation is distinct from the structures of the TEM and SHV families. In eight structures of the SHV family CWM8 is involved in a crystal contact to Arg222 of a symmetry-related protein. CWM11 is

situated at the bottom of a small dent at the protein surface which is formed by the residues 124 and 125. Hydrogen bonds are primarily formed to backbone atoms, but also to side chain atoms of these residues. An additional hydrogen bond is formed to a water molecule which is conserved in 44 of 49 structures and also occupies the dent. In two TEM structures CWM11 forms a crystal contact to Glu104. CWM12 is located at the N-terminal helix and forms primarily hydrogen bonds to backbone atoms, but also to side chain atoms of residues 61 to 64. CWM12 forms no crystal contacts. CWM13 is located 4.8 Å away from CWM12 and forms one hydrogen bond to the backbone atom of Arg61. Crystal contacts are formed to His96 and Tyr97 in 21 structures of the TEM family.

3.1.5 Discussion

Identification of conserved water molecules

Conserved water molecules were identified in a set of 49 high-resolution crystal structures of three class A β-lactamase families. Despite being diverse in sequence, all class A β-lactamases have a similar structure which allowed complete linkage cluster analysis to be applied after superposition of all structures into a common reference frame. As compared to the method of visual inspection, computational methods have the advantage of speed and objectivity when studying conserved water molecules in crystal structures of related proteins (Bottoms et al., 2006). A total of 13 water molecules were conserved in more than 90% of the structures, including two water molecules present in all structures. About the same number of conserved waters was found in previous studies of different proteins (Sanschagrin and Kuhn, 1998; Ogata and Wodak, 2002; Mustata and Briggs, 2004; Bottoms et al., 2006), despite the fact that those studies differ in the method applied, the number of structures, and their resolution. In general, the number of water molecules in crystallographic models increases with resolution (Carugo and Bordo, 1999), while crystallization conditions have only a moderate effect (Mattos, 2002). For a reliable analysis of CWMs, it is therefore desirable to investigate a large number of highly-resolved structures derived from proteins crystallized under different conditions. Is has been shown that artifacts due to crystal contacts can be reduced by comparing structures

Table 1: Water molecules with more than 90% conservation in all crystal structures. Listed are the group, overall conservation, the conservation in the TEM, SHV, and CTX-M families, and hydrogen bonding partners. Abbreviations: Gr.: group, Cons.: conservation.

CWM	Gr.	Cons.	Cons. in family:			Hydrogen bonds in family:		
			TEM	SHV	CTX-M	TEM	SHV	CTX-M
1	2	49	25	10	14	N,O Pro67 N,O Thr265 O Arg244	N,O Pro67 N,O Thr265 O Arg244	N,O Pro67 N,O Phe265 OD1,ND2 Asn245
2	1	49	25	10	14	N Met68 N Met69 O Leu169 O Asp179	N Met68 N Met69 O Leu169 O Asp179	N Met68 N Cys69 O Leu169 O Asp179
3	1	48	24	10	14	O Ala172 OG1 Thr266	O Ala172 NE,NH2 Arg266	O Ala172 O Ile173 O Pro174 OG1 Thr266
4	1	48	24	10	14	NH2 Arg164 N Ala172 N Ile173 OD2 Asp176	NH1,NH2 Arg164 N Ala172 N Leu173 OD2 Asp176	NH1,NH2 Arg164 OG1 Thr171 N Ala172 N Ile173 OD2 Asp176
5	1	48	24	10	14	NH2 Arg164 O Leu169 N Glu171 N Ala172	NH2 Arg164 O Leu169 N Glu171 N Ala172	NH2 Arg164 O Leu169 N Thr171 N Ala172
6	1	48	24	10	14	N,O Ile173 O Asp176	O Leu173 N,O Asp176 NH1,NH2 Arg266	N,O Ile173 N, O Asp176
7	2	48	24	10	14	O Lys192 O Leu193 N,O Gly196 N Leu198 N,O Leu199 NE Arg204	O Lys192 O Leu193 N Leu194 N Thr195 N Ser196 N Gln197 N Arg198 N Leu199 NE2 Gln204	O Gln192 O Leu193 O Gly196 N,O His197 N,O Ala198 N,O Leu199
8	2	47	23	10	14	N Leu91	N Leu91 OE1 Glu92	O Lys88 N,O Gln89 N Leu90 N Leu91
9	1	47	24	9	14	N Ala172 N Ile173 O Arg178	NH2 Arg164 O Asp176 O Arg178	NH2 Arg164 O Asp176 O Arg178
10	2	47	25	8	14	O Asp101 ND2 Asn136	O Asp101 ND2 Asn136	O Asp101 ND2 Asn136
11	2	47	22	10	14	OG Ser124 N Ala125	O Ala124 N Ala125 OG1 Thr128	O Ala124 N Ala125
12	2	46	23	9	14	O Arg61 N Pro62 N Glu63 N Glu64	N Ala62 N Asp63 N Glu64	N Gly62 N,OD2 Asp63 N Glu64
13	2	46	23	9	14	N Arg61	N Arg61	N Arg61

from different space groups (Loris et al., 1999; Bottoms et al., 2006). The 49 β-lactamase structures investigated here have been crystallized in two different space groups ($P2_12_12_1$ and $P2_1$). Thus, an influence of crystal contacts to conservation can be excluded for most CWMs with the exception of CWM8, CWM11, and CWM13 at the enzyme surface, which form crystal contacts in some structures.

Stability of the Ω-loop

Besides the four CWMs that are located at the protein surface, all other CWMs are associated with loops. Nearly the half of all CWMs (6 out of 13) are present in the catalytically relevant Ω-loop. From the presence of this large number of highly conserved water molecules we conclude that it is this conserved water cluster that stabilizes the Ω-loop and links it to the protein. The occurrence of CWMs especially in loop regions is in accordance with the results of previous studies on other proteins that have shown the involvement of CWMs in positioning loops to β-strands (Loris et al., 1994), stabilization of hairpin structures (Loris et al., 1994), and stabilization of a twisted β-turns (Ogata and Wodak, 2002). This is further supported by a statistical analysis of high resolution protein structures which concluded that well-resolved, internal water molecules preferentially reside near residues that are not part of an β-helix or a β-strand and help to satisfy the intramolecular hydrogen bond needs of backbone atoms (Park and Saven, 2005). Indeed, in our study the large majority of all hydrogen bonds formed by CWMs at the Ω-loop involve backbone atoms. The Ω-loop and the rest of the protein form the two outer layers of a sandwich-like structure with a CWM layer between them serving as structural glue. The contradiction of the Ω-loop - not being flexible despite having few contacts with the rest of protein - can now be resolved by taking water molecules into account. A stabilization of protein interfaces by water-mediated hydrogen bonds has been reported for many protein-protein complexes (Mustata and Briggs, 2004, and references therein). Acting as structural glue by hydrating main chain atoms of turns, loops, and coils and thus stabilizing loops seems to be a key structural role of CWMs in β-lactamases.

One of the Ω-loop stabilizing waters (CWM4) is hydrogen bonded to the N_ε of the Arg164 side chain and carboxyl oxygen of Asp176. We therefore speculate that this interaction

can have an additional stabilizing effect on the Ω-loop in the TEM, SHV, and CTX-M wild type β-lactamases, as it links together the side chains of the opposed residues. It has been shown that upon disruption of the Arg164-Asp179 salt bridge via substitutions by uncharged amino acids, the structure of the Ω-loop breaks down and more bulky substrates can bind to the active site (Vakulenko et al., 1999), which leads to extended spectrum β-lactamases activity of these mutants. Indeed, a conformational change of the Ω-loop in the region between residues 167 to 175 is seen in the crystal structure of variant TEM 64, where the salt bridge is disrupted by an Arg164Ser substitution. In this structure, a part of the Ω-loop moves up to 4.5 Å and thus the active site is enlarged (Wang et al., 2002). Interestingly, the crystal structure of TEM-64 is the only structure in our analysis that contains none of the water molecule CWM3, 4, 5, and 6. We attribute this to the additional change in backbone conformation of residues 171 to 175 which displaces the four CWMs.

Conserved waters and conserved amino acids

Previously it was concluded that residues forming a conserved water binding site are generally conserved (Loris et al., 1994; Bottoms et al., 2006). In class A β-lactamases the situation is more ambiguous. Most hydrogen bonds of the CWMs are formed to main chain atoms, thus side chains could be variable as long as substitutions do not disturb the local backbone geometry. Nevertheless, there seems to be no correlation between CWMs and residue conservation. Nearly the same number of main chain interactions to residues conserved in all families and to residues only conserved inside each family were found. Often, the same conserved water molecules can interact with conserved as well as with non-conserved residues. However, the situation at the Ω-loop is different. Most of the residues that interact with the CWMs are conserved throughout all families, with the exception of residues 69, 171, and 173. A high conservation at the amino acid sequence level is compulsory in protein-water interaction when side chain atoms are involved. CWM4 at the Ω-loop and CWM10 are hydrogen bonded to the side chains of Asp176 and Asn136, respectively, which are conserved in all analyzed structures. Nevertheless, we also observed the substitution of residues which are involved in side chain interactions to a CWM. The Ω-loop-associated CWM3 is the center of a hydrogen bond network to residues 172, 266, and a water molecule. While the orientation of residue 266 depends on

the type of side chain, the structure of this hydrogen bond network is conserved: upon replacing Thr266 by arginine, the side chain and the water molecule change place, while the position of CWM3 and of residue 172 is maintained. The role of conserved water molecules in maintaining hydrogen bond networks has been demonstrated previously for protein-protein interfaces, protein-ligand, and protein-cofactor binding (Xu et al., 1997; Bottoms et al., 2002; Sharrow et al., 2005). The hydrogen bond network of CWM3 is an example of an intra-protein network which is stable upon side chain substitution between protein families.

Evolutionary aspects of conserved water molecules

It has been proposed to treat the conservation of protein-bound water in analogy to the conservation of amino acid position in a multiple sequence alignment (Bottoms et al., 2006). Likewise, in analogy to sequence motifs, which often denote structurally or functionally important residues and are derived from multiple sequence alignments, motifs of CWMs could be defined from comparative crystal structure analysis. In the case of the class A β-lactamases, the cluster of Ω-loop associated CWMs would justify for such a CWM motif, as it has the function of stabilizing the Ω-loop and it is found in all of the analyzed structures of the class A β-lactamase families except the structure with the changed Ω-loop conformation.

As a consequence of the analogy between conserved water and conserved residues phylogenetic analysis based on conserved waters can be performed. The three class A β-lactamase families analyzed in this work have similar structure, but only a moderate sequence similarity. Phylogenetic analysis has shown that the CTX-M family diverged earlier in evolution than the SHV and TEM families (Hall and Barlow, 2004). A similar evolutionary relationship is observable when looking at the number of water molecules which are more conserved between the TEM and SHV families than to the CTX-M family. The structures of the TEM and SHV families have five conserved water molecules in common which are not present in the structures of the CTX-M family. Between the CTX-M and TEM families, two water molecules are highly conserved which are not found in the SHV family, and between the SHV and CTX-M families only one is conserved. The more distant relationship of CTX-M is further supported by analyzing the

differences between the families. Water molecules conserved more than 90% in a specific family but not found in the other families indicate how far the families have diverged from each other. The TEM, SHV, and CTX-M families have 5, 7, and 45 uniquely conserved water molecules, respectively. They are located mostly at the surface of the protein, and their occurrence in only one family reflects changes of the local environment in that particular region of the protein structure. We hope that future comparative studies on other protein families will provide further insight into the evolutionary aspects of water conservation.

Acknowledgments

Support by the Deutsche Forschungsgemeinschaft (DFG) within the priority program "Directed Evolution to Optimize and Understand Molecular Biocatalysts" (SPP1170) is gratefully acknowledged.

3.1.6 References

Ambler, R. P., The structure of beta-lactamases. Philos Trans R Soc Lond B Biol Sci 289 (1036), 321–331, 1980.

Babic, M., Hujer, A. M., Bonomo, R. A., What's new in antibiotic resistance? Focus on beta-lactamases. Drug Resist Updat 9 (3), 142–156, 2006.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E., The Protein Data Bank. Nucleic Acids Res 28 (1), 235–242, 2000.

Bottoms, C. A., Smith, P. E., Tanner, J. J., A structurally conserved water molecule in Rossmann dinucleotide-binding domains. Protein Sci 11 (9), 2125–2137, 2002.

Bottoms, C. A., White, T. A., Tanner, J. J., Exploring structurally conserved solvent sites in protein families. Proteins 64 (2), 404–421, 2006.

Bradford, P. A., Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. Clin Microbiol Rev 14 (4), 933–951, 2001.

Carugo, O., Bordo, D., How many water molecules can be detected by protein crystallography? Acta Crystallogr D Biol Crystallogr 55, 479–483, 1999.

DeLano, W., Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography 40, 2002.

Despa, F., Biological water: Its vital role in macromolecular structure and function. Ann NY Acad Sci 1066 (1), 1–11, 2005.

Fisher, J. F., Meroueh, S. O., Mobashery, S., Bacterial resistance to beta-lactam antibiotics: Compelling opportunism, compelling opportunity. Chem Rev 105 (2), 395–424, 2005.

Hall, B. G., Barlow, M., Evolution of the serine beta-lactamases: past, present and future. Drug Resist Updat 7 (2), 111–123, 2004.

Helms, V., Protein dynamics tightly connected to the dynamics of surrounding and internal water molecules. ChemPhysChem 8 (1), 23–33, 2007.

Jacoby, G. A., Beta-lactamase nomenclature. Antimicrob Agents Chemother 50 (4), 1123–1129, 2006.

Jaurin, B., Grundstrom, T., Ampc cephalosporinase of *Escherichia coli* K-12 has a different evolutionary origin from that of beta-lactamases of the penicillinase type. Proc Natl Acad Sci U S A 78 (8), 4897–4901, 1981.

Jelsch, C., Lenfant, F., Masson, J. M., Samama, J. P., Beta-lactamase TEM1 of *E. coli* crystal structure determination at 2.5 Å resolution. FEBS Letters 299 (2), 135–142, 1992.

Loris, R., Langhorst, U., De Vos, S., Decanniere, K., Bouckaert, J., Maes, D., Tran-

sue, T. R., Steyaert, J., Conserved water molecules in a large family of microbial ribonucleases. Proteins 36 (1), 117–134, 1999.

Loris, R., Stas, P. P. G., Wyns, L., Conserved waters in legume lectin crystal-structures - the importance of bound water for the sequence-structure relationship within the legume lectin family. J Biol Chem 269 (43), 26722–26733, 1994.

Lu, Y., Wang, R., Yang, C. Y., Wang, S., Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. J Chem Inf Model 47 (2), 668–75, 2007.

Majiduddin, F. K., Palzkill, T., An analysis of why highly similar enzymes evolve differently. Genetics 163 (2), 457–466, 2003.

Massova, I., Mobashery, S., Kinship and diversification of bacterial penicillin-binding proteins and beta-lactamases. Antimicrob Agents Chemother 42 (1), 1–17, 1998.

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

Mattos, C., Protein-water interactions in a dynamic world. Trends Biochem Sci 27 (4), 203–208, 2002.

Mclachlan, A. D., Rapid comparison of protein structures. Acta Crystallogr A 38 (Nov), 871–873, 1982.

Mustata, G., Briggs, J. M., Cluster analysis of water molecules in alanine racemase and their putative structural role. Protein Eng 17 (3), 223–234, 2004.

Nakasako, M., Water-protein interactions from high-resolution protein crystallography. Philos Trans R Soc Lond B Biol Sci 359 (1448), 1191–1204, 2004.

Notredame, C., Higgins, D. G., Heringa, J., T-coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302 (1), 205–217, 2000.

Ogata, K., Wodak, S. J., Conserved water molecules in MHC class-1 molecules and their putative structural and functional roles. Protein Eng 15 (8), 697–705, 2002.

Ouellette, M., Bissonnette, L., Roy, P. H., Precise insertion of antibiotic-resistance determinants into Tn21-like transposons - nucleotide-sequence of the Oxa-1 beta-lactamase gene. Proc Natl Acad Sci U S A 84 (21), 7378–7382, 1987.

Park, S., Saven, J. G., Statistical and molecular dynamics studies of buried waters in globular proteins. Proteins 60 (3), 450–463, 2005.

Poole, K., Resistance to beta-lactam antibiotics. Cell Mol Life Sci 61 (17), 2200–2223, 2004.

Prabhu, N., Sharp, K., Protein-solvent interactions. Chem Rev 106 (5), 1616–1623, 2006.

Roccatano, D., Sbardella, G., Aschi, M., Amicosante, G., Bossa, C., Di Nola, A., Mazza, F., Dynamical aspects of TEM-1 beta-lactamase probed by molecular dynamics. J Comput Aided Mol Des 19 (5), 329–340, 2005.

Sanschagrin, P. C., Kuhn, L. A., Cluster analysis of consensus water sites in thrombin and trypsin shows conservation between serine proteases and contributions to ligand specificity. Protein Sci 7 (10), 2054–64, 1998.

Savard, P. Y., Gagne, S. M., Backbone dynamics of TEM-1 determined by NMR: Evidence for a highly ordered protein. Biochemistry 45 (38), 11414–11424, 2006.

Sharrow, S. D., Edmonds, K. A., Goodman, M. A., Novotny, M. V., Stone, M. J., Thermodynamic consequences of disrupting a water-mediated hydrogen bond network in a protein:pheromone complex. Protein Sci 14 (1), 249–256, 2005.

Tzouvelekis, L. S., Tzelepi, E., Tassios, P. T., Legakis, N. J., CTX-M-type beta-lactamases: an emerging group of extended-spectrum enzymes. Int J Antimicrob Agents 14 (2), 137–142, 2000.

Vakulenko, S. B., Taibi-Tronche, P., Toth, M., Massova, I., Lerner, S. A., Mobashery,

S., 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), 23052–23060, 1999.

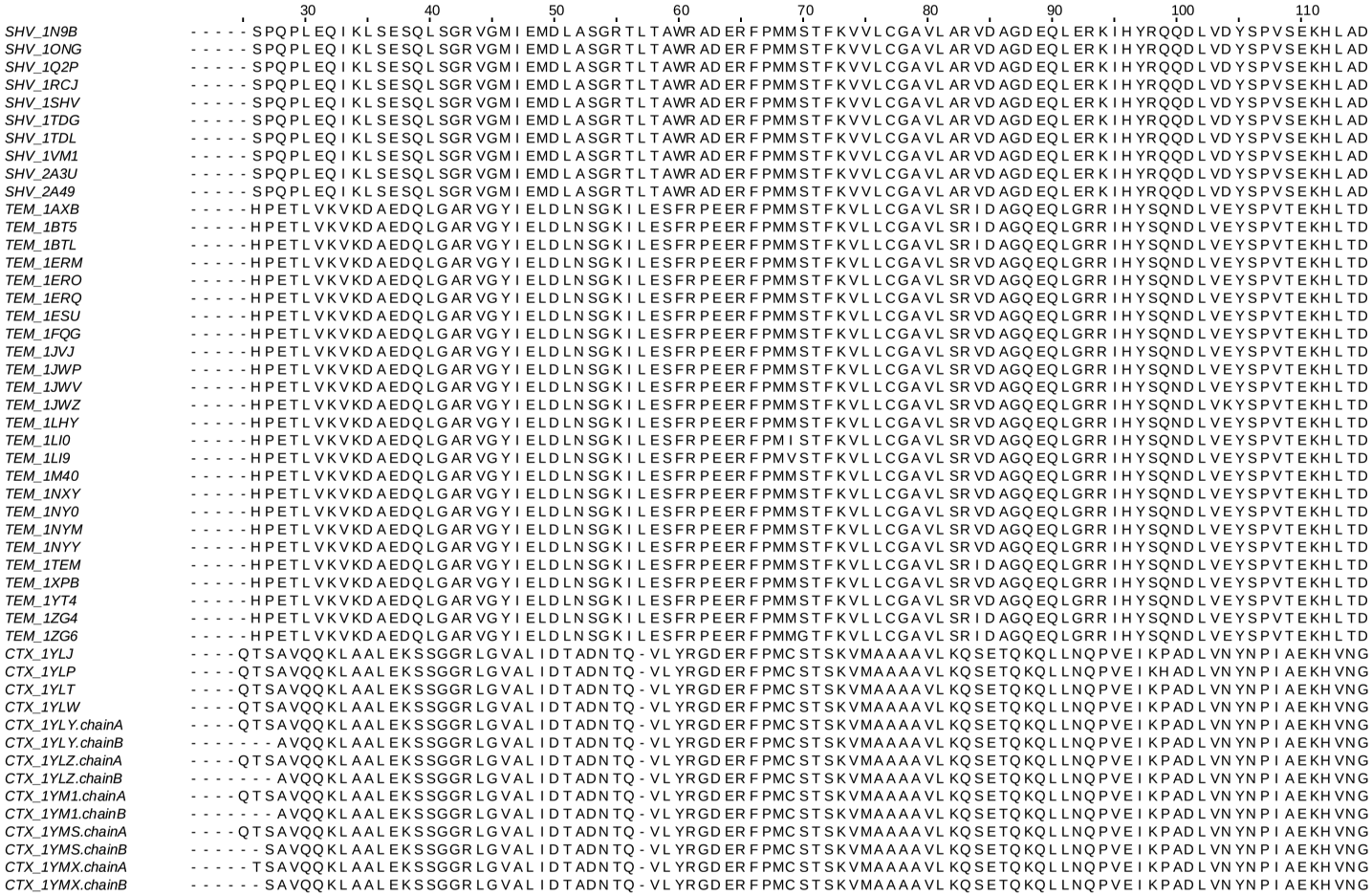
Vakulenko, S. B., Toth, M., Taibi, P., Mobashery, S., Lerner, S. A., Effects of Asp-179 mutations in Tem(Puc19) beta-lactamase on susceptibility to beta-lactams. Antimicrob Agents Chemother 39 (8), 1878–1880, 1995.

Wang, X., Minasov, G., Shoichet, B. K., Noncovalent interaction energies in covalent complexes: TEM-1 beta-lactamase and beta-lactams. Proteins 47 (1), 86–96, 2002.

Xu, D., Tsai, C. J., Nussinov, R., Hydrogen bonds and salt bridges across protein-protein interfaces. Protein Eng 10 (9), 999–1012, 1997.

3.1.7 Supplementary material

Details for all 49 structures (table 2–4) with their PDB identifier, information on atomic resolution, free R-factor, number of water molecules, mutations, and active side occupancy as well as the multiple sequence alignment of all 49 proteins used to derive residue conservation is provided as supplementary material.


30 40 50 60 70 80 90 100 110
SHV_1N9B -----SPQPLEQIKLSESQLSGRVGMIEMDLASGRTLTAWRADERFPMMSTFKVVLCGAVLARVDAGDEQLERKIHYRQQDLVDYSPVSEKHLAD
SHV_1ONG -----SPQPLEQIKLSESQLSGRVGMIEMDLASGRTLTAWRADERFPMMSTFKVVLCGAVLARVDAGDEQLERKIHYRQQDLVDYSPVSEKHLAD
SHV_1Q2P -----SPQPLEQIKLSESQLSGRVGMIEMDLASGRTLTAWRADERFPMMSTFKVVLCGAVLARVDAGDEQLERKIHYRQQDLVDYSPVSEKHLAD
SHV_1RCJ -----SPQPLEQIKLSESQLSGRVGMIEMDLASGRTLTAWRADERFPMMSTFKVVLCGAVLARVDAGDEQLERKIHYRQQDLVDYSPVSEKHLAD
SHV_1SHV -----SPQPLEQIKLSESQLSGRVGMIEMDLASGRTLTAWRADERFPMMSTFKVVLCGAVLARVDAGDEQLERKIHYRQQDLVDYSPVSEKHLAD
SHV_1TDG -----SPQPLEQIKLSESQLSGRVGMIEMDLASGRTLTAWRADERFPMMSTFKVVLCGAVLARVDAGDEQLERKIHYRQQDLVDYSPVSEKHLAD
SHV_1TDL -----SPQPLEQIKLSESQLSGRVGMIEMDLASGRTLTAWRADERFPMMSTFKVVLCGAVLARVDAGDEQLERKIHYRQQDLVDYSPVSEKHLAD
SHV_1VM1 -----SPQPLEQIKLSESQLSGRVGMIEMDLASGRTLTAWRADERFPMMSTFKVVLCGAVLARVDAGDEQLERKIHYRQQDLVDYSPVSEKHLAD
SHV_2A3U -----SPQPLEQIKLSESQLSGRVGMIEMDLASGRTLTAWRADERFPMMSTFKVVLCGAVLARVDAGDEQLERKIHYRQQDLVDYSPVSEKHLAD
SHV_2A49 -----SPQPLEQIKLSESQLSGRVGMIEMDLASGRTLTAWRADERFPMMSTFKVVLCGAVLARVDAGDEQLERKIHYRQQDLVDYSPVSEKHLAD
TEM_1AXB -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1BT5 -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1BTL -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1ERM -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1ERO -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1ERQ -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1ESU -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1FQG -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1JVJ -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1JWP -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1JWV -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1JWZ -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVKYSPVTEKHLTD
TEM_1LHY -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1LI0 -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMISTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1LI9 -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMVSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1M40 -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1NXY -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1NY0 -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1NYM -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1NYY -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1TEM -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1XPB -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1YT4 -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1ZG4 -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
TEM_1ZG6 -----HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMGTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD
CTX_1YLJ ----QTSAVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG
CTX_1YLP ----QTSAVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKHADLVNYNPIAEKHVNG
CTX_1YLT ----QTSAVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG
CTX_1YLW ----QTSAVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG
CTX_1YLY.chainA ----QTSAVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG
CTX_1YLY.chainB -------AVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG
CTX_1YLZ.chainA ----QTSAVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG
CTX_1YLZ.chainB -------AVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG
CTX_1YM1.chainA ----QTSAVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG
CTX_1YM1.chainB -------AVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG
CTX_1YMS.chainA ----QTSAVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG
CTX_1YMS.chainB ------SAVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG
CTX_1YMX.chainA -----TSAVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG
CTX_1YMX.chainB ------SAVQQKLAALEKSSGGRLGVALIDTADNTQ-VLYRGDERFPMCSTSKVMAAAAVLKQSETQKQLLNQPVEIKPADLVNYNPIAEKHVNG

120
130
140
150
160
170
180
190
200
SHV_1N9B
SHV_1ONG
SHV_1Q2P
SHV_1RCJ
SHV_1SHV
SHV_1TDG
SHV_1TDL
SHV_1VM1
SHV_2A3U
SHV_2A49
TEM_1AXB
TEM_1BT5
TEM_1BTL
TEM_1ERM
TEM_1ERO
TEM_1ERQ
TEM_1ESU
TEM_1FQG
TEM_1JVJ
TEM_1JWP
TEM_1JWV
TEM_1JWZ
TEM_1LHY
TEM_1LI0
TEM_1LI9
TEM_1M40
TEM_1NXY
TEM_1NY0
TEM_1NYM
TEM_1NYY
TEM_1TEM
TEM_1XPB
TEM_1YT4
TEM_1ZG4
TEM_1ZG6
CTX_1YLJ
CTX_1YLP
CTX_1YLT
CTX_1YLW
CTX_1YLY.chainA
CTX_1YLY.chainB
CTX_1YLZ.chainA
CTX_1YLZ.chainB
CTX_1YM1.chainA
CTX_1YM1.chainB
CTX_1YMS.chainA
CTX_1YMS.chainB
CTX_1YMX.chainA
CTX_1YMX.chainB

220
230
240
250
260
270
280
290
300
SHV_1N9B MVDDRVAGPLIRSVLPAGWFIADKTGAS-ERGARGIVALLGP-NNKAERIVVIYLRDTPASMAERNQQIAGIGAALIEHWQR
SHV_1ONG MVDDRVAGPLIRSVLPAGWFIADKTGAG-ERGARGIVALLGP-NNKAERIVVIYLRDTPASMAERNQQIAGIGAALIEHWQR
SHV_1Q2P MVDDRVAGPLIRSVLPAGWFIADKTGAG-ERGARGIVALLGP-NNKAERIVVIYLRDTPASMAERNQQIAGIGAALIEHWQR
SHV_1RCJ MVDDRVAGPLIRSVLPAGWFIADKTGAG-ERGARGIVALLGP-NNKAERIVVIYLRDTPASMAERNQQIAGIGAALIEHWQR
SHV_1SHV MVDDRVAGPLIRSVLPAGWFIADKTGAG-ERGARGIVALLGP-NNKAERIVVIYLRDTPASMAERNQQIAGIGAALIEHWQR
SHV_1TDG MVDDRVAGPLIRSVLPAGWFIADKTGAG-ERGARGIVALLGP-NNKAERIVVIYLRDTPASMAERNQQIAGIGAALIEHWQR
SHV_1TDL MVDDRVAGPLIRSVLPAGWFIADKTGAG-ERGARGIVALLGP-NNKAERIVVIYLRDTPASMAERNQQIAGIGAALIEHWQR
SHV_1VM1 MVDDRVAGPLIRSVLPAGWFIADKTGAG-ERGARGIVALLGP-NNKAERIVVIYLRDTPASMAERNQQIAGIGAALIEHWQR
SHV_2A3U MVDDRVAGPLIRSVLPAGWFIADKTGAG-ERGARGIVALLGP-NNKAERIVVIYLRDTPASMAERNQQIAGIGAALIEHWQR
SHV_2A49 MVDDRVAGPLIRSVLPAGWFIADKTGAG-ERGARGIVALLGP-NNKAERIVVIYLRDTPASMAERNQQIAGIGAALIEHWQR
TEM_1AXB MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1BT5 MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1BTL MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1ERM MEA-KVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1ERO MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1ERQ MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1ESU MEADKVAGPLLRSALPAGWFIADKAGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1FQG MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1JVJ MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1JWP MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1JWV MEADKVAGPLLRSALPAGWFIADKSGAA-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1JWZ MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1LHY MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSSGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1LI0 MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1LI9 MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1M40 MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1NXY MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1NY0 MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1NYM MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1NYY MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1TEM MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1XPB MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1YT4 MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1ZG4 MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
TEM_1ZG6 MEADKVAGPLLRSALPAGWFIADKSGAG-ERGSRGIIAALGP-DGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW--
CTX_1YLJ LKGNTTGAASIRAGLPTSWTAGDKTGSG-DYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YLP LKGNTTGAASIRAGLPTSWTVGDKTGSG-GYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YLT LKGNTTGAASIRAGLPTSWTVGDKTGSG-DYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YLW LKGNTTGAASIRAGLPTSWTAGDKTGSG-GYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YLY.chainA LKGNTTGAASIRAGLPTSWTAGDKTGSG-DYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YLY.chainB LKGNTTGAASIRAGLPTSWTAGDKTGSG-DYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YLZ.chainA LKGNTTGAASIRAGLPTSWTVGDKTGSG-DYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YLZ.chainB LKGNTTGAASIRAGLPTSWTAGDKTGSG-DYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YM1.chainA LKGNTTGAASIRAGLPTSWTAGDKTGSG-DYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YM1.chainB LKGNTTGAASIRAGLPTSWTAGDKTGSG-DYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YMS.chainA LKGNTTGAASIRAGLPTSWTAGDKTGSG-DYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YMS.chainB LKGNTTGAASIRAGLPTSWTAGDKTGSG-DYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YMX.chainA LKGNTTGAASIRAGLPTSWTAGDKTGSG-DYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--
CTX_1YMX.chainB LKGNTTGAASIRAGLPTSWTAGDKTGSG-DYGTTNDIAVIWP-QGRAPLVLVTYFTQPQQNAESRRDVLASAARIIAEGL--

Table 2: Crystal structures of the TEM β-lactamases sorted by atomic resolution. All structures have the space group $P2_12_12_1$. Abbreviations: R-Free: free R-factor, Res.: Resolution, Wat.: Number of water molecules in the crystal structure

PDB-ID	R-Free	Res. [Å]	Active site	Variant, mutation	Wat.
1M40	0.112	0.85	Inhibitor	M182T	566
1NYM	0.148	1.20	Inhibitor	M182T	436
1YT4	0.223	1.40		S130G	326
1LI9	0.189	1.52		TEM-34,M69V	396
1ZG4	0.240	1.55		TEM-1	194
1NXY	0.196	1.60	Inhibitor	M182T	447
1LI0	0.217	1.61		TEM-32,M69I,M182T	328
1ERM	n/a	1.70	Inhibitor	TEM-1	137
1FQG	n/a	1.70	Substrate	E166N	150
1JVJ	0.193	1.73	Inhibitor	N132A	418
1JWP	0.220	1.75		M182T	274
1NY0	0.198	1.75	Inhibitor	M182T	367
1BT5	0.238	1.80	Inhibitor	V82I,A182V	347
1BTL	n/a	1.80		TEM-1	199
1JWZ	0.189	1.80	Inhibitor	TEM-64,E104K,M182T,R164S	331
1JWV	0.203	1.85	Inhibitor	G238A	417
1ERQ	n/a	1.90	Inhibitor	TEM-1	121
1NYY	0.230	1.90	Inhibitor	M182T	240
1XPB	0.158	1.90		TEM-1	135
1TEM	0.202	1.95	Inhibitor	V84I,A184V	240
1AXB	0.206	2.00	Inhibitor	V84I,A184V	127
1ESU	n/a	2.00		S235A	137
1LHY	0.212	2.00		TEM-30,R241S	224
1ERO	n/a	2.10	Inhibitor	TEM-1	63
1ZG6	0.275	2.10		S70G	157

Table 3: Crystal structures of the SHV β-lactamases sorted by atomic resolution. All structures have the space group $P2_12_12_1$. Abbreviations: R-Free: free R-factor, Res.: Resolution, Wat.: Number of water molecules in the crystal structure

PDB-ID	R-Free	Res. [Å]	Active site	Variant, mutation	Wat.
1N9B	0.150	0.90		SHV-2, G238S	367
1ONG	0.186	1.10	Inhibitor	SHV-1	296
2A3U	0.194	1.34	Inhibitor	SHV-1, E166A	264
2A49	0.194	1.34	Inhibitor	SHV-1, E166A	316
1RCJ	0.171	1.63	Inhibitor	SHV-1, E166A	300
1TDG	0.176	1.80	Inhibitor	SHV-1, S130G	396
1TDL	0.181	1.80		SHV-1, S130G	414
1SHV	0.254	1.98		SHV-1	81
1Q2P	0.189	2.00	Inhibitor	SHV-1	193
1VM1	0.257	2.02	Inhibitor	SHV-1	175

Table 4: Crystal structures of the CTX-M β-lactamases sorted by atomic resolution. Multiple chains extracted from a single PDB entry are denoted as "A" or "B" after the PDB entry. All structures have the space group $P2_12_12_1$, except those denoted as "A" or "B", which have the space group $P2_1$. Abbreviations: R-Free: free R-factor, Res.: Resolution, Wat.: Number of water molecules in the crystal structure

PDB-ID	R-Free	Res. [Å]	Active site	Variant, mutation	Wat.
1YLJ	0.129	0.98		CTX-M-9a	489
1YLT	0.157	1.10		CTX-M-14, A231V	427
1YM1.A	0.140	1.12	Inhibitor	CTX-M-9a	421
1YM1.B	0.140	1.12	Inhibitor	CTX-M-9a	450
1YLP	0.158	1.20		CTX-M-27, D240G	320
1YLY.A	0.168	1.25	Inhibitor	CTX-M-9	441
1YLY.B	0.168	1.25	Inhibitor	CTX-M-9	447
1YLZ.A	0.172	1.35	Inhibitor	CTX-M-14	423
1YLZ.B	0.172	1.35	Inhibitor	CTX-M-14	461
1YMS.A	0.193	1.60	Inhibitor	CTX-M-9	382
1YMS.B	0.193	1.60	Inhibitor	CTX-M-9	435
1YMX.A	0.197	1.70	Substrate	CTX-M-9	304
1YMX.B	0.197	1.70	Substrate	CTX-M-9	338
1YLW	0.204	1.74		CTX-M-16	374

3.2 Multiple molecular dynamics simulations of TEM β-lactamase: Dynamics and water binding of the Ω-loop

3.2.1 Abstract

The Ω-loop of TEM β-lactamase is involved in substrate recognition and catalysis. The effect of bound water molecules to its structure and dynamics was investigated by performing 20 simulations of 5 ns using different initial conditions. Although no conformational changes occurred and the root mean square deviation from the X-ray structure was only 1 Å, multiple MD simulations considerably improved the conformational sampling. The high rigidity of the lactamase which was determined by calculating the root mean square fluctuations and the generalized order parameter, S^2, was in agreement with experimental observations. The residues in secondary structure elements are highly ordered, while loop regions are more flexible. Interestingly, the Ω-loop is rigid with order parameters similar to secondary structure elements, but the tip of the loop (residue 173 to 177) has a considerably higher flexibility. The rigidity of the Ω-loop is mediated by stabilizing water molecules inside a cavity between the Ω-loop and the protein core. In contrast, the tip of the Ω-loop lacks these interactions. Hydration of the cavity and exchange of water with the bulk solvent occurs via two pathways: the flexible tip which serves as a door to the cavity and a temporary channel involving the side chain of Arg164.

Keywords

TEM, β-lactamase, Ω-loop, flexibility, multiple molecular dynamics simulations, water bridges, conserved water molecules

Abbreviations

MMD, multiple molecular dynamics; RMSD, root mean square deviation; RMSF, root mean square fluctuation

3.2.2 Introduction

Plasmid-encoded TEM β-lactamases are the main cause for resistance against β-lactam antibiotics in gram-negative bacteria (Bradford, 2001). They efficiently hydrolyze the β-lactam ring and therefore render the antibiotic inactive. First isolated in 1965 from *E. coli* (Datta and Kontomichalou, 1965), the TEM β-lactamases within a few years have spread worldwide and are now found in many pathogenic organisms (Poole, 2004). TEM β-lactamases (EC 3.5.2.6) belong to the class A β-lactamases and are globular enzymes with an molecular weight of 29 kDa. Their structure is composed of two domains that are closely packed together: An all-α-domain consisting of 8 α-helices, and an α-/β-domain, consisting of 3 α-helices and 5 β-strands (Jelsch et al., 1992). The active site cavity, with the catalytic residues Ser70, Lys73, and Lys234, is located at the interface between the two domains and is limited by the Ω-loop. The Ω-loop (residues 161 to 179) is a conserved structural element in all class A β-lactamases and mediates their catalytic function because it positions the catalytic Glu166 and is involved in substrate binding. Its conformation is stabilized by a highly conserved salt bridge formed between Arg164 and Asp179 (Vakulenko et al., 1995, 1999; Majiduddin and Palzkill, 2003).

Because of their importance in antibiotic resistance, the structure and catalytic function of TEM β-lactamases are well studied and characterized (Matagne et al., 1998; Fisher et al., 2005), but only a few studies addressed the dynamical features of the TEM β-lactamases and the catalytically relevant Ω-loop. Computational analysis of the TEM β-lactamases using molecular dynamics (MD) simulations (Diaz et al., 2003; Roccatano et al., 2005) showed a stable enzyme that does not exhibit large structural rearrangements. This was experimentally confirmed by a recently published NMR study of the TEM β-lactamase from which it is evident that the protein is one of the most ordered proteins (Savard and Gagne, 2006). However, the flexibility of the Ω-loop is

still under debate: From structural analysis it was found that the Ω-loop has only few contacts with the rest of the protein and therefore was speculated to be a flexible element (Massova and Mobashery, 1998). This is supported by the molecular dynamics study of Roccatano et al. (2005), in which a flap-like motion of the Ω-loop was observed. In contrast, in the MD simulation of Diaz et al. (2003) as well as in the NMR study of Savard and Gagne (2006) no indication of an increased flexibility of the Ω-loop in the picosecond-to-nanosecond time scale was found. In a comparative study of 49 high resolution structures of class A β-lactamases, we identified 6 conserved water molecules located in a narrow, tunnel-shaped cavity lined by the Ω-loop and residues 65 to 69 of the protein core. By mediating hydrogen bonds between the Ω-loop and the protein core, these conserved water molecules provide an explanation for the stability and low flexibility of the Ω-loop (see manuscript "Conserved water molecules stabilize the Ω-loop in class A β-lactamases" at page 43).

In this study, the dynamics of the Ω-loop and its interplay with bound water molecules was investigated using multiple molecular dynamics (MMD) simulations of the TEM β-lactamase. To investigate the influence of initial water positions given in the crystallographic structure, two protein-water systems were studied: One with the crystal water included and the other one starting without the crystal water molecules. By using multiple simulations for each system, we were able to estimate the reproducibility of our results and the significance of differences between the two systems.

3.2.3 Material & methods

MMD simulations of TEM β-lactamase

A highly resolved crystal structure of TEM β-lactamase (PDB entry 1M40, resolution 0.85 Å) (Minasov et al., 2002) was used as a starting point for multiple molecular dynamics (MMD) simulations. All non-protein atoms were deleted manually and the positions of crystal water molecules were saved into a separate file. Protonation states of titratable groups were determined using PDB2PQR (Dolinsky et al., 2004), retaining existing hydrogen atoms in the crystal structure. Two different lactamase-water systems were

prepared using the all-atom ff03 force field (Duan et al., 2003): One including the crystal water molecules (referred to as MMD_{wat}) and one without the crystal water molecules (referred to as MMD_{nowat}). Each system was immersed in a 10 Å layer truncated octahedron periodic water box of TIP3P waters. The total amount of water molecules was 10302 for the MMD_{wat} system and 8168 for the MMD_{nowat} system. The net charge of -7 for both systems was neutralized by adding seven sodium counterions. The total system size was 34974 and 28572 atoms for the MMD_{wat} and MMD_{nowat} system, respectively.

The PMEMD program of AMBER 8.0 was used for energy minimization and MD simulations (Case et al., 2004). For each system, 10 independent MD simulations under periodic boundary conditions were performed using a different initial velocity distribution for each simulation. The SHAKE algorithm (Ryckaert et al., 1977) was applied to all bonds containing hydrogen atoms, and a time step of 1 fs was used. The Berendsen method was used to couple the systems to a constant temperature of 300 K and a pressure of 1 bar (Berendsen et al., 1984). The electrostatic interaction was evaluated by the particle-mesh Ewald method (Darden et al., 1993) and Lennard-Jones interactions were evaluated using a 8.0 Å cutoff. The systems were subjected to 50 steps of energy minimization by steepest descent to remove repulsive van der Waals contacts, followed by 150 steps of energy minimization by conjugate gradient. In the first 150 ps of the subsequent MD simulations, the backbone and side chain atoms were restrained using a harmonic potential with a gradually decreasing force constant from 5 to 0.1 kcal/mol for the backbone atoms and 1.0 to 0.1 kcal/mol for the side chains atoms. This was followed by an unrestrained equilibration phase of 1 ns and a production run of 4 ns in which snapshots were saved every 250 fs.

Analysis protein structure and dynamics

Analysis was done using the PTRAJ program of AMBER 8.0 unless stated otherwise and all protein figures were created by VMD (Stone et al., 2001). For results obtained on a per-residue basis, always the average values and standard deviations were calculated for the 10 MMD_{wat} or the 10 MMD_{nowat} simulations. The stability of the simulations was analyzed by calculating of the root-mean-square deviation of the backbone atoms between each conformer and the initial conformation (RMSD) and between all conform-

ers (2D-RMSD). Root mean-square fluctuations (RMSF) of the backbone atoms were calculated relative to the average structure of the last 2 ns, and all coordinate frames were first superimposed onto the average structure to remove overall translation and rotation. Order parameters S^2 were calculated for the 10 MD trajectories of each system applying a model-free approach of Lipari and Szabo (Lipari and Szabo, 1982) to the normalized time-correlation functions of the N-H internuclear vectors over the last 2 ns after overall rotation was removed. For each residue, S^2 was avergaged over the 10 simulations. The order parameters of residues for which time-correlation functions did not reach a plateau value in more than six simulations were treated as "not determined". A principal component analysis of the individual as well as combined trajectories, consisting of the concatenated last 4 ns of the 10 MMD_{wat} or the 10 MMD_{nowat} simulations, was performed using programs from the GROMACS software package (Spoel et al., 2005). Eigenvectors and eigenvalues were obtained from the diagonalization of the covariance matrices of the C_α atoms, and the principal components were generated by projecting the trajectories on the respective eigenvectors (Amadei et al., 1993). The cosine content of the principal components was calculated to estimate if they describe random diffusion (Hess, 2000). The overlap between the covariance matrix of each trajectory and that of the combined trajectory was calculated to estimate the similarity of the sampled conformational space between the single and the combined trajectory (Hess, 2002).

Identification of highly occupied water molecule positions and water bridges

For the determination of highly occupied water molecule positions in MMD simulations and in multiple crystal structures of the TEM β-lactamase, a complete-linkage hierarchical cluster analysis of water molecules was applied using the program WatCH (Sanschagrin and Kuhn, 1998). The clustering process groups together water molecules in different frames of the MMD trajectories or from different crystal structures that physically overlap and represent one specific water position. The default cut-off value of 2.4 Å for the maximum inter-water distance was used. With this cut-off value, water molecules with a center-to-center distance of 2.4 Å will overlap by 50% if an approximate effective radius of 1.6 Å is assumed. It ensures that clusters do not contain several water molecules from the same snapshot or crystal structure. After superimposition onto a reference frame, the last nanosecond of each trajectory was analyzed, including only the

250 closest water molecules to the protein. Water molecule positions occupied in at least 90% of the snapshots were extracted.

Crystal structures of the TEM β-lactamase family were retrieved from the Protein Data Bank (PDB) (Berman et al., 2000). β-lactamase variants with core disrupting mutations and β-lactamases bound to inhibitory proteins were excluded. Because it was found that the amount of crystal water depends on the atomic resolution of the crystal structure (Carugo and Bordo, 1999), only structures with an atomic resolution of at least 2.2 Å were included. A total of 25 protein structures of the TEM β-lactamase family (PDB entries 1AXB, 1BT5, 1BTL, 1ERM, 1ERO, 1ERQ, 1ESU, 1FQG, 1JVJ, 1JWP, 1JWV, 1JWZ, 1LHY, 1LI0, 1LI9, 1M40, 1NXY, 1NY0, 1NYM, 1NYY, 1TEM, 1XPB, 1YT4, 1ZG4, 1ZG6) were analyzed, containing between 63 and 566 (average: 271) crystal water molecules. All structures including the crystal water molecules were superimposed by multiple-structure fitting of the C_α atoms using the McLachlan algorithm (Mclachlan, 1982) as implemented in the program ProFit[2]. The fitting zones were initially determined by a Needleman-Wunsch sequence alignment and then iteratively adapted during the fitting process. Water positions found to be occupied in at least 23 out of 25 analyzed structures (conservation $> 90\%$) were finally compared to the highly occupied water molecule positions obtained from the analysis of the MMD_{wat} and MMD_{nowat} simulations.

Water bridges, defined as a set of two or more non-water polar atoms of distinct residues that interact with a particular water molecule, were identified and characterized during the last nanosecond of each trajectory using the software ANKUSH (Sanjeev, 2004). Only water molecules within a distance of 4.0 Å to any non-water polar atom of the protein were considered. For all water bridges that were formed for at least 30% of the analyzed time occupancies and the number of unique water molecules involved in a particular water bridge were averaged and the exchange rate calculated.

[2]ProFit, version 2.5.4, Martin, A.C.R., http://www.bioinf.org.uk/software/profit

3.2.4 Results

Multiple molecular dynamics (MMD) simulations of TEM β-lactamase, comprising a total simulation time of 100 ns, were performed. Two different lactamase-water systems were studied: One including the crystal water molecules (referred to as MMD_{wat}) and one without the crystal water molecules (referred to as MMD_{nowat}). For each system, 10 independent MD simulations with a length of 5 ns each and a different distribution of the initial velocity were performed. Their overall stability and the sampling was analyzed. The dynamical features and the water binding properties of the Ω-loop were investigated in detail for the 10 MMD_{wat} simulations. In the final section the influence of ommitting the crystal water at the beginning of the MMD_{nowat} simulations was assessed.

Stability of the simulations

For each simulation, the protein stability during the MMD simulations was assessed by monitoring the root mean square deviation (RMSD) with respect to the minimized starting structure. All 20 simulations were stable and showed only a small deviation in the protein backbone conformation (figure 1). Over the last 1 ns of the trajectory the average $RMSD_{wat}$ and $RMSD_{nowat}$ was 0.95 ± 0.06 Å and 0.96 ± 0.10 Å, respectively. 2D-RMSD plots of each simulation revealed that in most simulations the protein backbone keeps its initial conformation throughout the complete simulation, whereas in six simulations the protein backbone undergoes small conformational changes. Snapshots were extracted from those regions of the 2D-RMSD plots where a conformational change was observed. Superimposition of the snapshots showed that the conformational changes occurred mainly in the hinge region between the all-α and the α/β domain (residues 216 to 229), a short part of the Ω-loop (residues 173 to 177), other solvent exposed loops, or the C- and N-terminal amino acids. Simulations with changes in protein backbone conformation were present in both MMD_{wat} and MMD_{nowat} simulations. $RMSD_{wat}$ graphs, $RMSD_{nowat}$ graphs, 2D-$RMSD_{wat}$ plots, and 2D-$RMSD_{nowat}$ plots for all 20 simulations are provided as supplementary material.

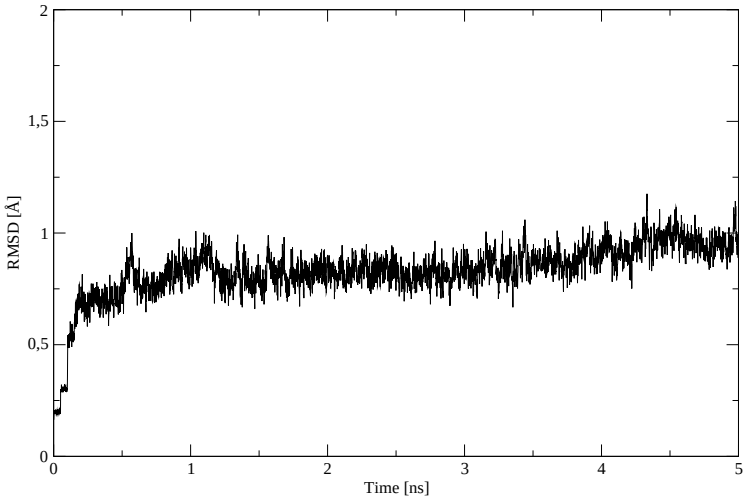


Figure 1: The root mean square deviation of the protein backbone from the minimized starting structure during the MD2$_{wat}$ simulation is typical for all 20 MMD$_{wat}$ and MMD$_{nowat}$ simulations.

Sampling of conformational space

To estimate the sampling of conformational space, an essential dynamics analysis was performed for the individual trajectories as well as for combined trajectories, consisting of the concatenated last 4 ns of the 10 MMD$_{wat}$ and MMD$_{nowat}$ simulations. The average overlap values of the individual simulations with the combined trajectories were 0.57 $\pm$ 0.01 and 0.56 $\pm$ 0.01 for the MMD$_{wat}$ and MMD$_{nowat}$ trajectories, respectively. This shows that the sampling of conformational space by a single 5 ns simulation is limited. Even though each individual MD simulation deviates only 1 Å from the initial starting structure, they contribute differently to the two major motions represented by the eigenvectors 1 and 2, as the projection of the individual simulations onto the first and second eigenvectors of the combined trajectories demonstrates (for MMD$_{wat}$ see figure 2 and for MMD$_{nowat}$ see supplementary material).

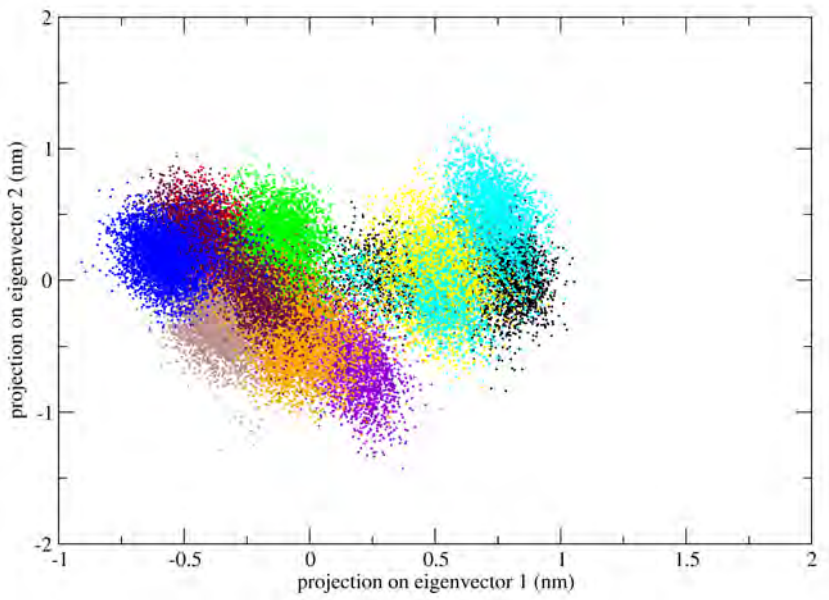


Figure 2: 2D projection of the MMD_{wat} simulations onto the first and second eigenvector of the combined trajectory. Individual MD simulations are distinguished by colors.

Dynamical features of the TEM β-lactamase

The backbone flexibility of the TEM beta-lactamase, as indicated by the residue-averaged root mean square fluctuations (RMSF) calculated over the last 1 ns time window (figure 3 *A*), is characterized by low overall RMSF values ($RMSF_{wat} = 0.53 \pm 0.14$ Å). Residues located in secondary structure elements show less flexibility (α-helices: $RMSF_{wat} = 0.47 \pm 0.10$ Å; β-sheets: $RMSF_{wat} = 0.39 \pm 0.06$ Å) than residues located in solvent exposed loops and turns ($RMSF_{wat} = 0.59 \pm 0.14$ Å). While the majority of the Ω-loop residues have RMSF values similar to residues located in secondary structure elements ($RMSF_{wat} = 0.43 \pm 0.05$ Å), the residues 173 to 177 have much higher RMSF values ($RMSF_{wat} = 0.69 \pm 0.15$ Å) and thus are referred to as flexible tip of the Ω-loop.

The essential dynamics analysis of the combined MMD_{wat} trajectory shows that the largest correlated motion is a subtle movement of two domains relative to each other, which is best described as contracting and stretching motion, and a flap-like motion of the flexible tip. The principal component associated with this motion had a cosine content of 0.52, indicating that it partly describes random diffusion due to insufficient

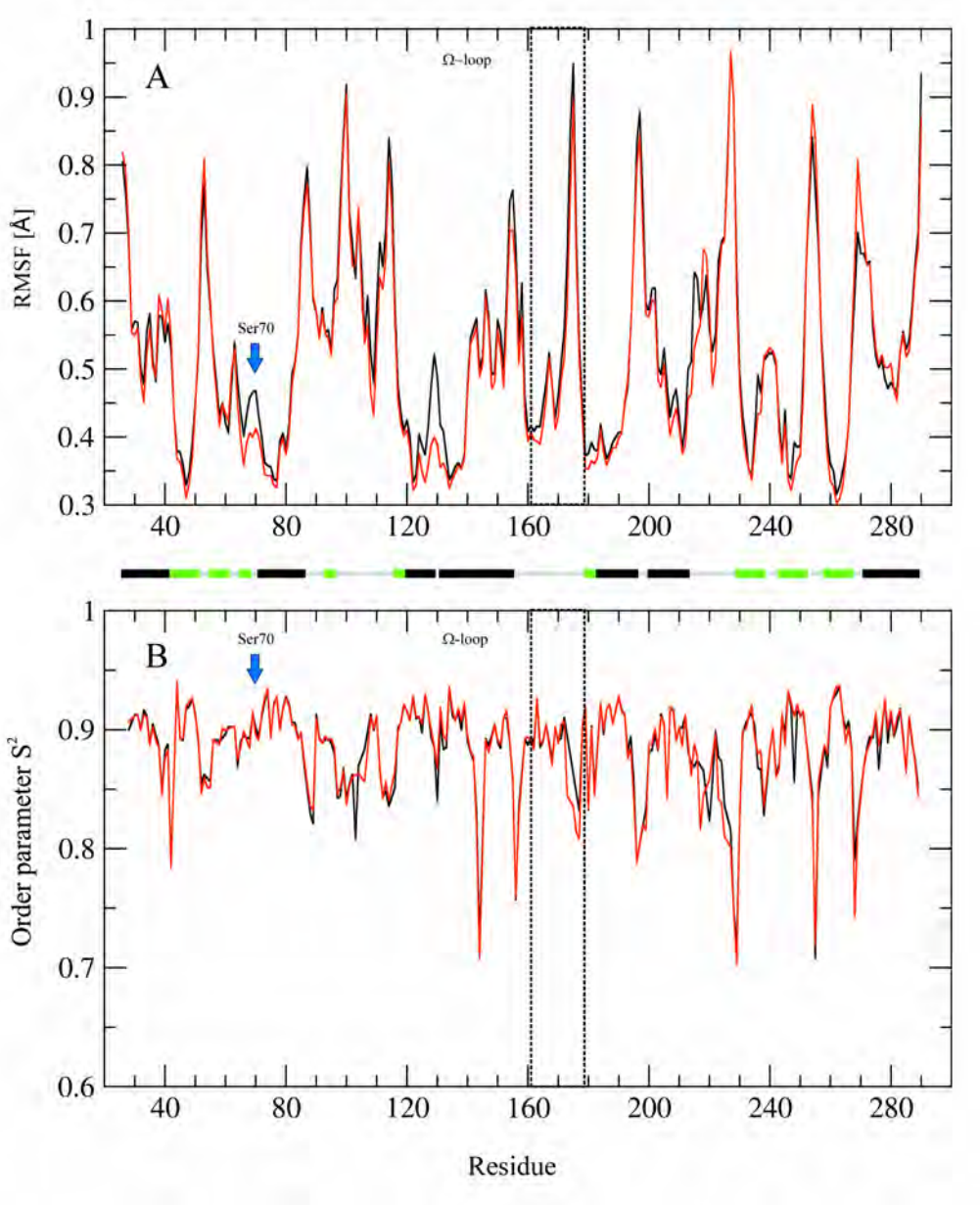


Figure 3: Dynamical features of the TEM β-lactamase: (*A*) RMSF and (*B*) order parameter S^2 of the MMD_{wat} (red) and MMD_{nowat} (black) simulations. The catalytic Ser70 and the residues forming the Ω-loop are marked and the secondary structure is annotated (helices: black; sheets: green; loops: grey). Residues for which no order parameter could be determined are missing in the graph.

sampling. However, subsequent principal components with cosine contents close to zero also participate the Ω-loop motion.

The flexibility of the protein backbone was additionally characterized by calculating the order parameter S^2 from the time-correlation functions of the N-H bond vectors (figure 3 *B*). Consistent with the results obtained by the RMSF calculations, the residues with a high order parameter are mostly located in secondary structure elements of the protein (for α-helices and β-sheets: $S^2_{wat} = 0.90 \pm 0.02$), while residues in loops have a lower order parameter ($S^2_{wat} = 0.86 \pm 0.05$). It should be noted that the time-correlation function of eight residues in MMD_{wat} did not reach a plateau value in seven or more simulations and therefore no order parameter could be determined. All catalytic relevant residues (Ser70, Lys73, and Lys234) are highly ordered ($S^2_{wat} > 0.90$), with the exception of Ser130 ($S^2_{wat} = 0.87$). Consistent with the results of the RMSF analysis, the Ω-loop is rigid with order parameters similar to residues located in secondary structure elements ($S^2_{wat} = 0.90 \pm 0.02$), while the residues in the tip of the Ω-loop show remarkably lower order parameters ($S^2_{wat} = 0.83 \pm 0.02$).

To further analyze the higher flexibility of the residues in the tip of the Ω-loop, the distance between the center of mass of the C_α atoms of these residues and the center of mass of the C_α atoms of the oppositely located residues 65 and 66 of the protein core was measured over the complete length of the trajectories at 1 ps intervals (table 1). The average distance in MMD_{wat} simulations (8.4 ± 0.6 Å) was similar to the initial distance in the crystal structure (8.2 Å). However, when taking into account the minimum and maximum values for this distance ($MMD^{min}_{wat} = 6.3$ Å, $MMD^{max}_{wat} = 11$ Å), it becomes evident that this particular part of the Ω-loop is indeed flexible and can undergo considerable conformational changes. As these residues contribute to the entrance to the Ω-loop cavity, increasing the distance to the protein core can be referred to as an opening movement and decreasing likewise as a closing movement (figure 4). However, during most of the 5150 snapshots analyzed for each trajectory, the distance is within 1 Å of the distance of the crystal structure. If the distance changes more than 1 Å with respect to the crystal structure (shorter than 7.2 Å or larger than 9.2 Å), then in most simulations an opening is observed. With the exception of $MD1_{wat}$ and $MD1_{nowat}$, it seems that opening and closing are mutually exclusive in an individual simulation within the time scale of 5 ns.

Table 1: Distance (in Å) between the center of mass of the C_α atoms of residues 173 to 177 and the center of mass of the C_α atoms of residues 65 and 66 measured over the complete length of the trajectories. Listed are the mean values and the standard deviations, the minimum and maximum values, and the number of snapshots (out of 5150) in which the distance is 1 Å larger or smaller than the distance in the crystal structure.

Simulation	Mean ± SD.	Min.	Max.	<= 7.2 Å	>= 9.2 Å
in MMD_{wat} simulations:					
MD1	8.0 ± 0.5	6.5	10.0	268	64
MD2	8.7 ± 0.4	7.1	10.4	1	782
MD3	7.7 ± 0.4	6.7	9.8	313	1
MD4	7.6 ± 0.4	6.3	9.1	674	0
MD5	8.8 ± 0.5	7.4	11.0	0	1252
MD6	8.6 ± 0.4	7.3	10.0	0	435
MD7	8.5 ± 0.4	7.2	10.4	0	327
MD8	8.5 ± 0.4	7.1	10.0	1	275
MD9	8.5 ± 0.4	7.4	10.0	0	314
MD10	8.7 ± 0.4	7.6	10.6	0	741
in MMD_{nowat} simulations:					
MD1	8.0 ± 0.6	6.8	10.7	100	385
MD2	8.7 ± 0.5	7.5	10.8	0	809
MD3	7.8 ± 0.5	5.9	9.7	385	34
MD4	8.5 ± 0.5	6.9	10.4	22	377
MD5	7.6 ± 0.5	6.2	9.8	635	13
MD6	8.4 ± 0.4	7.4	10.7	0	225
MD7	8.3 ± 0.3	6.8	9.9	9	43
MD8	8.3 ± 0.4	6.7	10.2	9	176
MD9	8.4 ± 0.4	7.4	9.8	0	139
MD10	8.3 ± 0.3	7.0	9.7	1	79

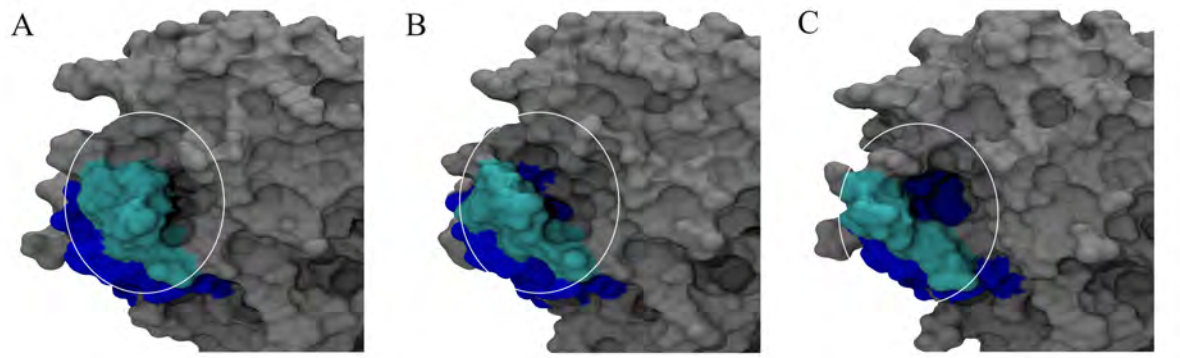


Figure 4: Two selected conformers of the simulation MD1$_{nowat}$ where the tip moved towards the protein core (A) or away from the protein core (C), thus decreasing or increasing the Ω-loop cavity (indicated by a white circle) in comparison to the X-ray structure (B) with an intermediate position of the tip.

Highly occupied water molecule positions and water bridges at the Ω-loop

Highly conserved water molecules were determined from the MMD$_{wat}$ trajectories for the last nanosecond of the simulations, which resulted in 59 distinct water positions that were conserved in more than 90% of the analyzed trajectory frames. To compare with experimental data, the clustering was also applied to 25 TEM β-lactamase crystal structures, resulting in 29 water molecule positions having a conservation of at least 90%. The majority of the conserved water molecules in the crystal structures as well as in the MMD$_{wat}$ simulations are located at the surface of the protein, bound via hydrogen bonds to a single protein residue or protein residues close-by in sequence. Comparing the results of the crystal structure and MMD$_{wat}$ analysis resulted in 19 water molecules conserved in the crystal structures and in the MMD$_{wat}$ simulations. They include three water molecules which are situated in the narrow, tunnel-shaped Ω-loop cavity. However, another three water molecule positions at the entrance of the Ω-loop cavity, which were detected to be conserved in the crystal structures, were not found in the MMD$_{wat}$ simulations.

To further characterize the water molecules located in the Ω-loop cavity, water bridges (WB) were calculated for the last nanosecond of the MMD$_{wat}$ simulations. This allows to identify the residues interacting with the conserved water molecules and to derive

temporal information like occupancies (in percent of analyzed simulation time) and exchange rates (number of different water molecules involved in forming a certain water bridge during the analyzed simulation time). In general, it was found that water bridges with high occupancy correspond to locations of conserved water molecules detected in the cluster analysis. Four water bridges could be distinguished inside the Ω-loop cavity. Three of them are among the highest occupied water bridges detected and form hydrogen bonds exclusively to main chain atoms of the adjacent residues. WB1 connects residues Met68 and Met69 of the protein core to residues Leu169 and Asp179 of the Ω-loop. It has a high occupancy ($MMD_{wat} = 88.6 \pm 11.1$) and a very low exchange rate ($MMD_{wat} = 1.3 \pm 0.5$ ns^{-1}). WB2 interconnects residues Leu169 and Ala172 of the Ω-loop with an occupancy ($MMD_{wat} = 86.8 \pm 29.4$) and an exchange rate ($MMD_{wat} = 1.4 \pm 0.5$ ns^{-1}) similar to WB1. WB3 resides between residues Ile173 and Asp176 at the entrance of the Ω-loop cavity and has a lower occupancy ($MMD_{wat} = 80.7 \pm 9.7$) and a considerably higher exchange rate ($MMD_{wat} = 14.4 \pm 3.4$ ns^{-1}). The fourth water bridge, WB4, connects the carbonyl atom of Ala172 on the Ω-loop to the side chain of the opposed Thr266 of the protein core. In contrast to the other water bridges inside the Ω-loop, WB4 was only present in four MMD_{wat} simulations, hence its low occupancy when averaged over all simulations (30.2 ± 39.0). The exchange rate of WB4 is 6.1 ± 6.3 ns^{-1}.

Exchange with bulk solvent

During the simulations, an exchange of water molecules between the Ω-loop cavity and the bulk solvent occurs at an average exchange rate of about 6 ns^{-1}. The paths of the initial crystal water molecules in the Ω-loop cavity during the course of the MMD_{wat} simulations were investigated. There are two possible routes for water molecules to leave the Ω-loop cavity: through the entrance of the Ω-loop cavity, which opens to the bulk solvent, and by a second route via residues Arg164, Ile173, Asp176, and Arg178. This second route was not obvious from visual inspection of the starting crystal structure because it is only temporarily formed. Exchange via this route gives an explanation for the higher exchange rate of WB3. The ratio between crystal water molecules leaving the Ω-loop cavity through the entrance and crystal water molecules leaving through the temporarily formed opening is approximately 3:2. However, no preference could be

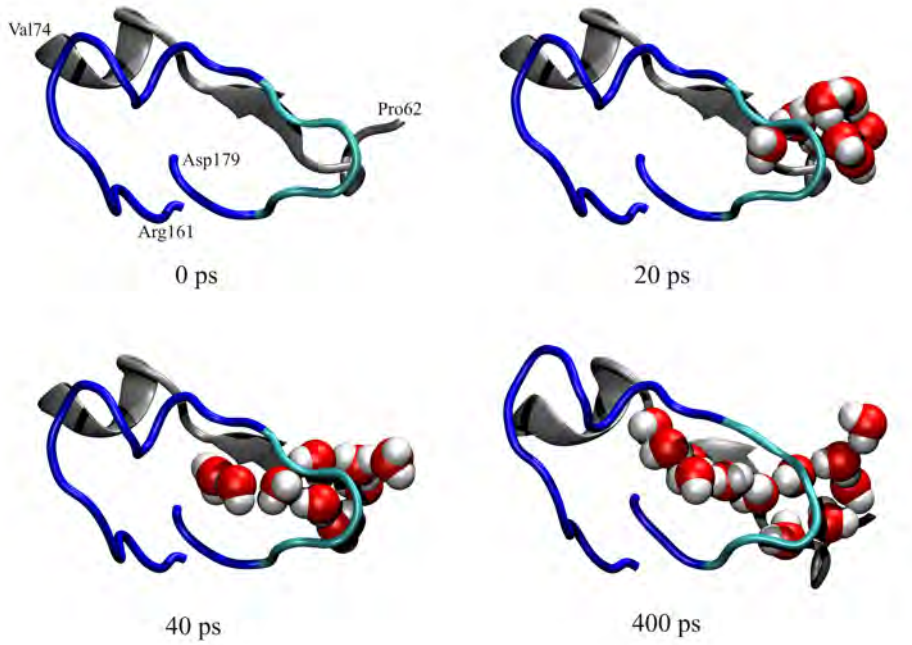


Figure 5: Gradual solvation of the Ω-loop cavity consisting of the Ω-loop (residues 161 to 179, dark blue) and residues of the protein core (residues 62 to 74, grey) at four different time steps of simulation $MD1_{nowat}$. The first water molecules enter the cavity at 20 ps near the flexible tip (light blue).

detected, which crystal water molecule in the Ω-loop cavity leaves by which route.

Hydration of the Ω-loop cavity in MMD_{nowat} simulations

In the MMD_{nowat} simulations, the crystal water molecules were excluded from the initial system preparation and therefore the cavity at the Ω-loop initially did not contain any water molecules. To gain insight into the process of Ω-loop solvation, all MMD_{nowat} trajectories were visually analyzed. It was observed that the Ω-loop cavity is solvated at an early stage of the simulations and that the solvation process is similar in all simulations: At approximately 20 ps, the first water molecule enters the opening of the Ω-loop cavity formed by residues 65 and 66 of the protein core and the flexible tip (residues 174 to 176). Soon thereafter, further water molecules follow (figure 5).

After 400 ps, the cavity at the Ω-loop is fully solvated and all conserved water molecule positions are occupied, and the water bridges are established. From now on, water bridges in the Ω-loop cavity are identical in the MMD_{nowat} and MMD_{wat} simulations.

Comparison of the MMD_{nowat} and the MMD_{wat} simulations

The comparison between the MMD_{nowat} and MMD_{wat} simulations of the TEM β-lactamase showed that the RMSF and S^2 values were nearly identical (figure 3, *A* and *B*). They exhibit the same distribution between rigid secondary structure elements (α-helices: $RMSF_{nowat}$ = 0.48 ± 0.10 Å and S^2_{nowat} = 0.90 ± 0.02; β-sheets: $RMSF_{nowat}$ = 0.41 ± 0.06 Å and S^2_{nowat} = 0.90 ± 0.02), the more flexible loops ($RMSF_{nowat}$ = 0.61 ± 0.14 Å and S^2_{nowat} = 0.86 ± 0.04), and the Ω-loop with its rigid part ($RMSF_{nowat}$ = 0.45 ± 0.05 Å and S^2_{nowat} 0.90 ± 0.01) and flexibile tip ($RMSF_{nowat}$ = 0.74 ± 0.16 Å and S^2_{nowat} = 0.83 ± 0.02). Taking the standard deviation into account, even the discrepancy between the RMSF values around residues 70 and 130 were not significant (Ser70: $RMSF_{nowat}$ = 0.47 ± 0.11 Å and $RMSF_{wat}$ = 0.41 ± 0.06 Å; Ser130: $RMSF_{nowat}$ = 0.49 ± 0.09 Å and $RMSF_{wat}$ = 0.39 ± 0.04 Å). The essential dynamics analysis of the MMD_{nowat} simulations showed the same subtle domain motions in the first and second principal components and the flap-like motion of the Ω-loop in the third principal component. Similar to the MMD_{wat} simulations, the principal components associated with the domain movements had considerable cosine contents (0.32 and 0.29), in contrast to the principal component associated with the Ω-loop motion, which had almost no cosine content. The distance measurements between the tip and the protein core yielded similar results as compared to the MMD_{wat} simulations (average distance: 8.2 ± 0.5 Å; MMD^{min}_{nowat} = 5.9 Å; MMD^{max}_{nowat} = 10.9 Å). Concerning the water binding properties of the TEM β-lactamase, the results of the cluster analysis and of the water bridge calculations were in good agreement between the MMD_{nowat} and MMD_{wat} simulations. From the 29 water molecule positions found to be conserved in the crystal structures, 18 are also conserved in the MMD_{nowat} simulations, including those three buried in the Ω-loop cavity. Compared to MMD_{wat} simulations, the water bridges WB1 and WB2 have slightly higher occupancies and are associated with smaller standard deviations (WB1 = 92.7 ± 3.9; WB2 = 96.2 ± 4.8), whereas WB3 has similar values (81.2 ± 6.6). The exchange rates for all three water bridges did not differ between the two systems

(WB1 = 1.2 ± 0.4 ns^{-1}; WB2 = 1.7 ± 1.1 ns^{-1}; WB3 = 15.3 ± 3.7 ns^{-1}). WB4 was present in seven MMD_{nowat} simulations, had also a low occupancy when averaged over all simulations (59.4 ± 42.0), and a similar exchange rate (9.2 ± 6.4 ns^{-1}). From the similar water binding properties and the reproducible dynamical behavior it is evident that there is no influence of the missing initial crystal waters on the MD simulations. Regardless of the initial starting conditions, within 1 ns the water binding patterns of TEM β-lactamase have been established and hence it is not surprising that the dynamics of the protein and of the Ω-loop are virtually the same in both systems.

3.2.5 Discussion

Sampling by multiple MD simulations

The conformational flexibility of the TEM β-lactamase and its interaction with tightly bound water molecules has been investigated using multiple molecular dynamics (MMD) simulations. A total of 20 simulations, each with a length of 5 ns and a different distribution of the initial velocities, were performed. The approach of performing multiple simulations provides several advantages. First, the sampling of the protein's conformational space is improved as compared to a single MD simulation, as shown by the overlap between the individual simulations and the combined trajectory. It has been shown that repeating MD simulations with slightly different initial conditions results in different values for the properties of interest (Elofsson and Nilsson, 1993). The reason for this observation is the insufficient sampling a protein's conformational space during a single MD simulation (Hess, 2002; Straub and Thirumalai, 1993; Caves et al., 1998). To overcome this limitation, it was suggested and has been repeatedly confirmed that the approach of performing multiple short-time MD simulations improves the sampling of conformational space and is more efficient than performing a single long MD simulation, in which the system might get trapped in a particular region of conformational space for a long time (Carlson et al., 1996; Daggett, 2000; Gorfe et al., 2002; Likic et al., 2005; Legge et al., 2006).

The second advantage offered by MMD simulations is the subsequent statistical analysis

of the results which provides information about their reproducibility. Similar to good laboratory practice in experimental setups, MD simulations have to be performed as multiple, independent experiments. Thus, for each property derived from the trajectory distributions can be assigned and analyzed. Deriving average values and standard deviations is crucial to interpret the value of a calculated property and the significance of an observed difference between two different systems. While this basically holds for all calculated or measured properties, it was most evident for the calculated root mean square fluctuation (RMSF) that showed the highest variation between individual simulations. Taking an extreme example, the RMSF values for Met129 ranged from 0.36 to 0.49 Å (0.40 ± 0.04 Å) and from 0.39 to 0.72 Å (0.52 ± 0.12 Å) in the 10 individual MMD_{wat} and MMD_{nowat} simulations, respectively. Thus, the difference between the MMD_{wat} and MMD_{nowat} simulations is within the standard deviation. However, if only one MD simulation of each system would have been performed, a comparison could have yielded a whole range of outcomes: $\text{RMSF}_{wat} < \text{RMSF}_{nowat}$, $\text{RMSF}_{wat} = \text{RMSF}_{nowat}$, or $\text{RMSF}_{wat} > \text{RMSF}_{nowat}$.

Dynamical behavior of the TEM β-lactamase

All MMD simulations performed in this work have shown that the TEM β-lactamase is stable during the time scale of the simulation. It only deviates slightly from the initial crystal structure and does not undergo significant conformational rearrangements. The dynamical behavior and the flexibility of the protein were probed by analyzing the RMSF and S^2 order parameters. From both properties it is evident that residues located in secondary structure elements have a low flexibility and are highly ordered, in contrast to residues located in loop regions, which are more flexible and less ordered. This is in agreement with a previous short 1 ns MD simulation of the TEM β-lactamase, in which only small structural changes occurred during the simulation (Diaz et al., 2003). However, a more recent 5 ns MD simulation using a different force field and water model describes a motion of one domain relative to the other by essential dynamics analysis (Roccatano et al., 2005). We acknowledge the observation of similar, subtle motions in the MMD simulations, although the associated principal components showed a considerable cosine content in the combined trajectories. Consequently, even with the MMD simulations performed in this work, the sampling of the conformational space is

not enough to satisfactorily describe these motions.

It was shown that the order parameter S^2 for the backbone amide N-H bond vectors from nuclear magnetic resonance (NMR) can be reproduced by MD simulations for residues exhibiting motions on a fast time scale (Philippopoulos et al., 1997). In a recent NMR study of the TEM β-lactamase, an average generalized order parameter S^2 of 0.90 ± 0.02 was determined for the complete protein, thus giving experimental evidence for a highly ordered protein (Savard and Gagne, 2006). S^2 calculated from our MMD simulations showed agreement with the experimental data. The highly ordered N- and C-terminal residues as well as the high order of active site residues are reproduced by the MMD simulations. With the agreement between the calculated and experimental S^2, our study further strengthens the picture of the TEM β-lactamase and its overall dynamical behavior as a very compact and rigid protein. Since all other members of the class A β-lactamases have a very similar structure, it may be speculated that this overall rigidity is a common property of the protein family. While the main part of the Ω-loop is rigid, with S^2 comparable to residues in the protein core, the tip of the Ω-loop (residues 173 to 177) has a considerably higher flexibility. Indicated by the essential dynamics analysis and the distance measurements, the tip of the Ω-loop is able to move closer to as well as further away from the protein core. Therefore, the MMD simulations can consistently explain results from previous MD simulations, which have been thought to be contradictory: While Diaz et al. (2003) observed only slight deviations from the crystal structure, in the MD simulation of Roccatano et al. (2005) the tip of the Ω-loop moved closer to the core of the enzyme. Both events, the unchanged Ω-loop tip conformation and the closing movement of the tip, have been observed in different individual MD simulations of this work. In addition, it was demonstrated that the tip of the Ω-loop performs an opening movement in most MD simulations, thus underlining the importance of multiple simulations. However, a repeated opening-closing motion could not be observed in any of the individual simulation, which indicates that this motion occurs on a longer time scale. This is supported by the NMR study of Savard and Gagne (2006), who observed the presence of motions of the Ω-loop at the microsecond-to-millisecond time scale. The high flexibility of the Ω-loop tip residues is also supported by the NMR analysis, which demonstrates that S^2 drops from 0.91 ± 0.04 for the rigid part of the Ω-loop to 0.84 ± 0.01 for the flexible tip region (Savard and Gagne, 2006).

The role of water in stabilizing the Ω-loop

Water is known to have a crucial role in protein structure, dynamics and activity (Mattos, 2002; Despa, 2005; Prabhu and Sharp, 2006; Helms, 2007) and therefore the interaction of water molecules with the residues of the Ω-loop has been investigated in detail. The analysis of a large number of highly-resolved TEM β-lactamase structures derived from proteins crystallized under different conditions led to the finding that 6 highly conserved water molecules are associated with the Ω-loop. However, in the MMD simulations only three water molecules were identified to be conserved in the simulations. These three water molecules, buried in the Ω-loop cavity, form water bridges between residues of the Ω-loop and between residues of the Ω-loop and the protein core. The three conserved water molecules not detected in the MMD simulations are located at the entrance of the Ω-loop cavity and close to the flexible tip of the Ω-loop. They cannot form well defined water bridges to the protein, as suitable residues are too far away. Being close to the bulk solvent, their rapid fluctuation prevents the detection of these water molecules as conserved water molecules in the clustering analysis, although visual inspection clearly demonstrates the existence of water molecules at the entrance of the Ω-loop cavity throughout the simulation.

The high residence time of the water molecules inside the Ω-loop cavity, the high conservation in multiple crystal structures, the rapid occupation of their positions in the MMD_{nowat} simulations, and formation of highly occupied water bridges led us to the conclusion that these water molecules are involved in stabilizing the Ω-loop conformation and in packing the Ω-loop to the protein core. This conclusion is supported by the results of previous studies on crystallographic data of other proteins that have shown the involvement of conserved water molecules in packing loops to β-strands (Loris et al., 1994), stabilization of hairpin structures (Loris et al., 1994), and stabilization of twisted β-turns (Ogata and Wodak, 2002). It is further supported by a statistical analysis of high-resolution protein structures which concluded that well-resolved, internal water molecules preferentially reside near residues that are not part of an α-helix or a β-strand, and help to saturate intramolecular hydrogen bond donors and acceptors of the protein backbone (Park and Saven, 2005). Indeed, in our MMD simulations the hydrogen bonds of the water bridges are exclusively formed with backbone atoms. It was shown experi-

mentally that water can directly interact with the protein backbone and side chains in the protein interior (Ernst et al., 1995) and that the residence times of buried water molecules are much longer than those of water molecules at the protein surface (Otting et al., 1991). A differential scanning calorimetry study has indicated that buried water molecules contribute to protein stability (Takano et al., 2003). Therefore, the water molecules packed between the Ω-loop and the protein core provide an explanation for the rigidity of the Ω-loop and the flexibility of the tip. While the main part of the Ω-loop is stabilized towards the protein core by water molecules, which act as structural glue, the tip of the Ω-loop lacks these interactions and thus shows an increased flexibility.

The observation of some water molecules leaving the Ω-loop cavity through a temporarily formed opening lined by the side chains of Arg164, Ile173, Asp176, and Arg178 gives rise to a new view on the effect of the Arg164Ser/His substitution that is often observed in extended-spectrum β-lactamases (ESBL). Arg164 is part of a salt bridge (Arg164-Asp179) that anchors the Ω-loop. It has been shown that substituting Arg164 by uncharged amino acids disrups the salt bridge and the structure of the Ω-loop breaks down. As a consequence, more bulky substrates can bind to the active site (Vakulenko et al., 1999), which leads to extended spectrum β-lactamases activity of these mutants. A conformational change of the Ω-loop in the region between residues 167 to 175 is seen in the crystal structure of variant TEM 64 (Wang et al., 2002), in which the salt bridge is disrupted by an Arg164Ser substitution and almost all conserved water molecules at the Ω-loop are missing. Taking into account our observation that conserved water molecules help to stabilize the Ω-loop, a possible explanation for the effect of the Arg164Ser or Arg164His substitution would be that the smaller side chains promote the opening of the temporary water channel and this leads to an increase in the exchange rate and thus to a destabilization of the Ω-loop.

Acknowledgments

Support by the Deutsche Forschungsgemeinschaft (DFG) within the priority program "Directed Evolution to Optimize and Understand Molecular Biocatalysts" (SPP1170) is gratefully acknowledged. The authors thank Michael Lerner for providing the Python script for the order parameter calculation.

3.2.6 References

Amadei, A., Linssen, A. B., Berendsen, H. J., Essential dynamics of proteins. Proteins 17 (4), 412–425, 1993.

Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., Haak, J. R., Molecular-dynamics with coupling to an external bath. J Chem Phys 81 (8), 3684–3690, 1984.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E., The Protein Data Bank. Nucleic Acids Res 28 (1), 235–242, 2000.

Bradford, P. A., Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. Clin Microbiol Rev 14 (4), 933–951, 2001.

Carlson, M. L., Regan, R. M., Gibson, Q. H., Distal cavity fluctuations in myoglobin: protein motion and ligand diffusion. Biochemistry 35 (4), 1125–1136, 1996.

Carugo, O., Bordo, D., How many water molecules can be detected by protein crystallography? Acta Crystallogr D Biol Crystallogr 55, 479–483, 1999.

Case, D., Darden, T., Cheatham III, T., Simmerling, C., Wang, J., Duke, R., Luo, R., Merz, K., Wang, B., Pearlman, D., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J., Ross, W., Kollman, P., Amber 8. University of California, San Francisco, 2004.

Caves, L. S., Evanseck, J. D., Karplus, M., Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. Protein Sci 7 (3), 649–666, 1998.

Daggett, V., Long timescale simulations. Curr Opin Struct Biol 10 (2), 160–164, 2000.

Darden, T., York, D., Pedersen, L., Particle mesh Ewald: An N · log(N) method for Ewald sums in large systems. J Chem Phys 98, 10089–10092, 1993.

Datta, N., Kontomichalou, P., Penicillinase synthesis controlled by infectious R factors in *Enterobacteriaceae*. Nature 208 (5007), 239–241, 1965.

Despa, F., Biological water: Its vital role in macromolecular structure and function. Ann NY Acad Sci 1066 (1), 1–11, 2005.

Diaz, N., Sordo, T. L., Jr, K. M. M., Suarez, D., Insights into the acylation mechanism of class A beta-lactamases from molecular dynamics simulations of the TEM-1 enzyme complexed with benzylpenicillin. J Am Chem Soc 125 (3), 672–684, 2003.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., Baker, N. A., PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res 32, W665–W667, 2004.

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G. M., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. M., Kollman, P., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24 (16), 1999–2012, 2003.

Elofsson, A., Nilsson, L., How consistent are molecular dynamics simulations? Comparing structure and dynamics in reduced and oxidized *Escherichia coli* thioredoxin. J Mol Biol 233 (4), 766–780, 1993.

Ernst, J. A., Clubb, R. T., Zhou, H. X., Gronenborn, A. M., Clore, G. M., Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. Science 267 (5205), 1813–1817, 1995.

Fisher, J. F., Meroueh, S. O., Mobashery, S., Bacterial resistance to beta-lactam antibiotics: Compelling opportunism, compelling opportunity. Chem Rev 105 (2), 395–424, 2005.

Gorfe, A. A., Ferrara, P., Caflisch, A., Marti, D. N., Bosshard, H. R., Jelesarov, I., Calculation of protein ionization equilibria with conformational sampling: pK(a) of a model leucine zipper, GCN4 and barnase. Proteins 46 (1), 41–60, 2002.

Helms, V., Protein dynamics tightly connected to the dynamics of surrounding and internal water molecules. ChemPhysChem 8 (1), 23–33, 2007.

Hess, Similarities between principal components of protein dynamics and random diffusion. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 62 (6 Pt B), 8438–8448, 2000.

Hess, B., Convergence of sampling in protein simulations. Phys Rev E Stat Nonlin Soft Matter Phys 65 (3 Pt 1), 031910, 2002.

Jelsch, C., Lenfant, F., Masson, J. M., Samama, J. P., Beta-lactamase TEM1 of *E. coli* crystal structure determination at 2.5 Å resolution. FEBS Letters 299 (2), 135–142, 1992.

Legge, F. S., Budi, A., Treutlein, H., Yarovsky, I., Protein flexibility: multiple molecular dynamics simulations of insulin chain b. Biophys Chem 119 (2), 146–157, 2006.

Likic, V. A., Gooley, P. R., Speed, T. P., Strehler, E. E., A statistical approach to the interpretation of molecular dynamics simulations of calmodulin equilibrium dynamics. Protein Sci 14 (12), 2955–2963, 2005.

Lipari, G., Szabo, A., Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J Am Chem Soc 104 (17), 4546–4559, 1982.

Loris, R., Stas, P. P. G., Wyns, L., Conserved waters in legume lectin crystal-structures - the importance of bound water for the sequence-structure relationship within the legume lectin family. J Biol Chem 269 (43), 26722–26733, 1994.

Majiduddin, F. K., Palzkill, T., An analysis of why highly similar enzymes evolve differently. Genetics 163 (2), 457–466, 2003.

Massova, I., Mobashery, S., Kinship and diversification of bacterial penicillin-binding proteins and beta-lactamases. Antimicrob Agents Chemother 42 (1), 1–17, 1998.

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

Mattos, C., Protein-water interactions in a dynamic world. Trends Biochem Sci 27 (4), 203–208, 2002.

Mclachlan, A. D., Rapid comparison of protein structures. Acta Crystallogr A 38 (Nov), 871–873, 1982.

Minasov, G., Wang, X., Shoichet, B. K., An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation. J Am Chem Soc 124 (19), 5333–40, 2002.

Ogata, K., Wodak, S. J., Conserved water molecules in MHC class-1 molecules and their putative structural and functional roles. Protein Eng 15 (8), 697–705, 2002.

Otting, G., Liepinsh, E., Wüthrich, K., Protein hydration in aqueous solution. Science 254 (5034), 974–980, 1991.

Park, S., Saven, J. G., Statistical and molecular dynamics studies of buried waters in globular proteins. Proteins 60 (3), 450–463, 2005.

Philippopoulos, M., Mandel, A. M., Palmer, A. G., Lim, C., Accuracy and precision of NMR relaxation experiments and MD simulations for characterizing protein dynamics. Proteins 28 (4), 481–493, 1997.

Poole, K., Resistance to beta-lactam antibiotics. Cell Mol Life Sci 61 (17), 2200–2223, 2004.

Prabhu, N., Sharp, K., Protein-solvent interactions. Chem Rev 106 (5), 1616–1623, 2006.

Roccatano, D., Sbardella, G., Aschi, M., Amicosante, G., Bossa, C., Di Nola, A., Mazza, F., Dynamical aspects of TEM-1 beta-lactamase probed by molecular dynamics. J Comput Aided Mol Des 19 (5), 329–340, 2005.

Ryckaert, J. P., Ciccotti, G., Berendsen, H. J. C., Numerical-integration of cartesian

equations of motion of a system with constraints - molecular-dynamics of n-alkanes. J Comput Phys 23 (3), 327–341, 1977.

Sanjeev, B., Ankush. Indian Institute of Science., 2004.

Sanschagrin, P. C., Kuhn, L. A., Cluster analysis of consensus water sites in thrombin and trypsin shows conservation between serine proteases and contributions to ligand specificity. Protein Sci 7 (10), 2054–64, 1998.

Savard, P. Y., Gagne, S. M., Backbone dynamics of TEM-1 determined by NMR: Evidence for a highly ordered protein. Biochemistry 45 (38), 11414–11424, 2006.

Spoel, D. V. D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., Berendsen, H. J. C., Gromacs: fast, flexible, and free. J Comput Chem 26 (16), 1701–1718, 2005.

Stone, J. E., Gullingsrud, J., Schulten, K., A system for interactive molecular dynamics simulation. In: I3D '01: Proceedings of the 2001 symposium on Interactive 3D graphics, ACM Press, New York, NY, USA, 191–194, 2001.

Straub, J. E., Thirumalai, D., Theoretical probes of conformational fluctuations in S-peptide and RNase A/3'-UMP enzyme product complex. Proteins 15 (4), 360–373, 1993.

Takano, K., Yamagata, Y., Yutani, K., Buried water molecules contribute to the conformational stability of a protein. Protein Eng 16 (1), 5–9, 2003.

Vakulenko, S. B., Taibi-Tronche, P., Toth, M., Massova, I., Lerner, S. A., Mobashery, S., 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), 23052–23060, 1999.

Vakulenko, S. B., Toth, M., Taibi, P., Mobashery, S., Lerner, S. A., Effects of Asp-179 mutations in Tem(Puc19) beta-lactamase on susceptibility to beta-lactams. Antimicrob Agents Chemother 39 (8), 1878–1880, 1995.

Wang, X., Minasov, G., Shoichet, B. K., Noncovalent interaction energies in covalent complexes: TEM-1 beta-lactamase and beta-lactams. Proteins 47 (1), 86–96, 2002.

3.2.7 Supplementary material

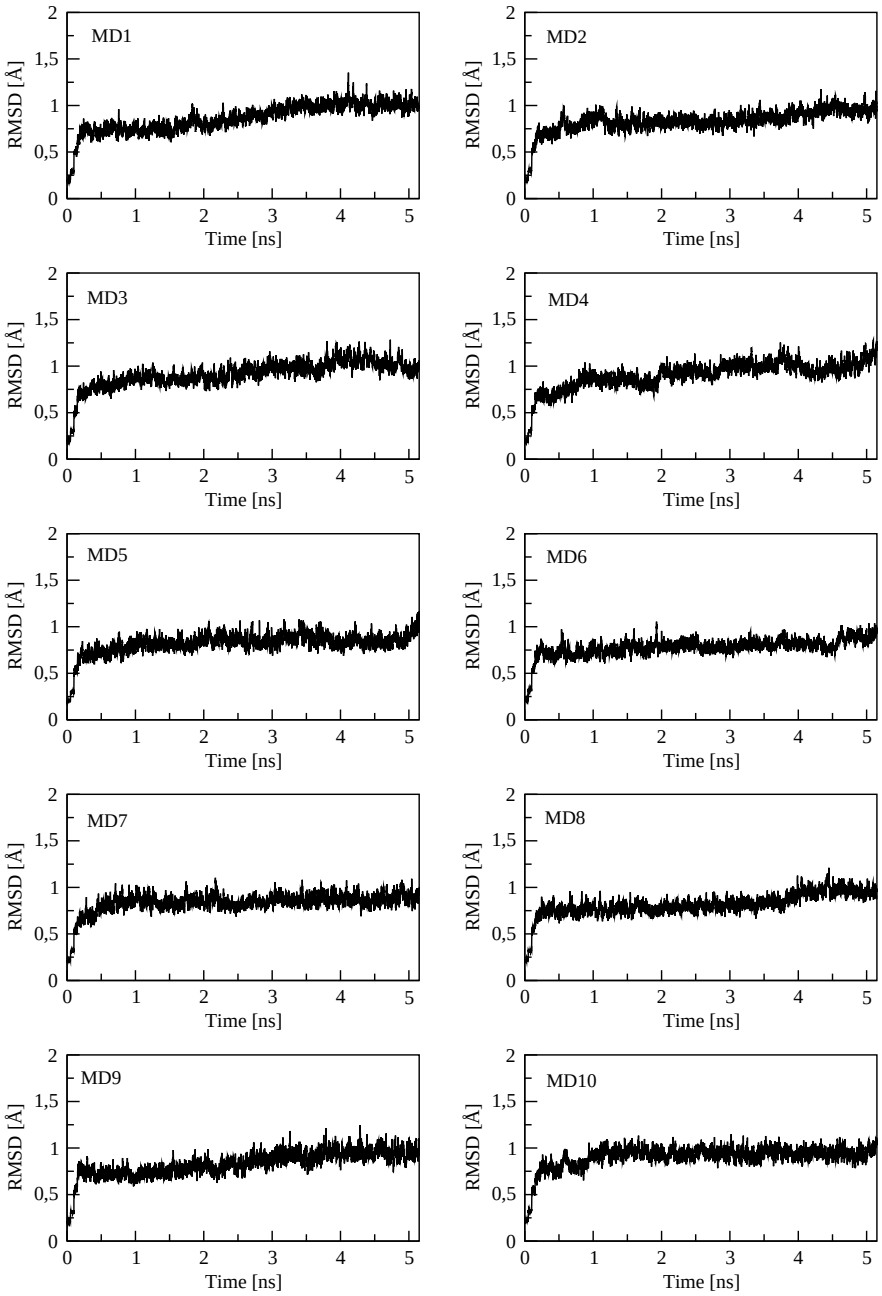


Figure 6: Root mean square deviation of the protein backbone from the minimized starting structure during the 10 MMD_{wat} simulations

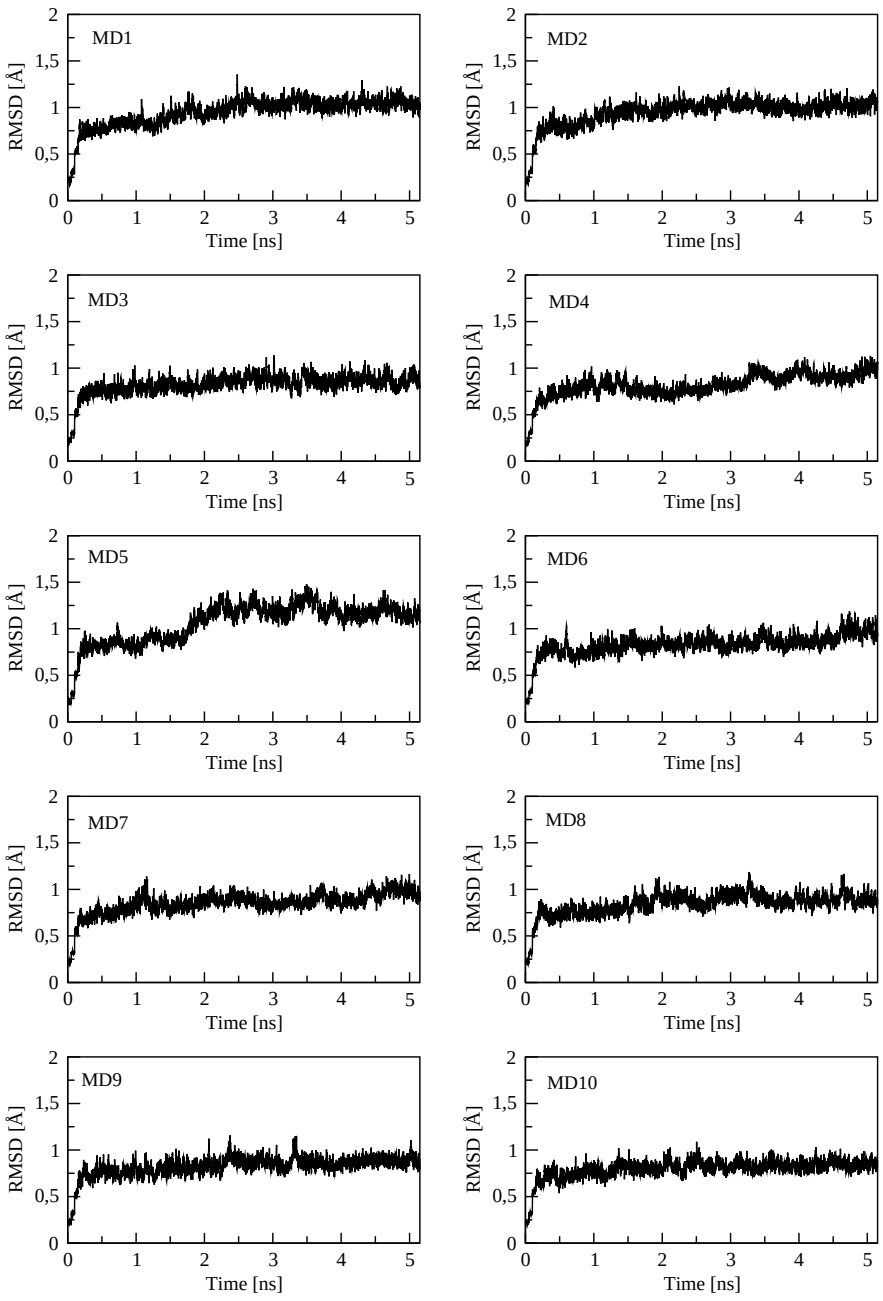


Figure 7: Root mean square deviation of the protein backbone from the minimized starting structure during the 10 MMD_{nowat} simulations

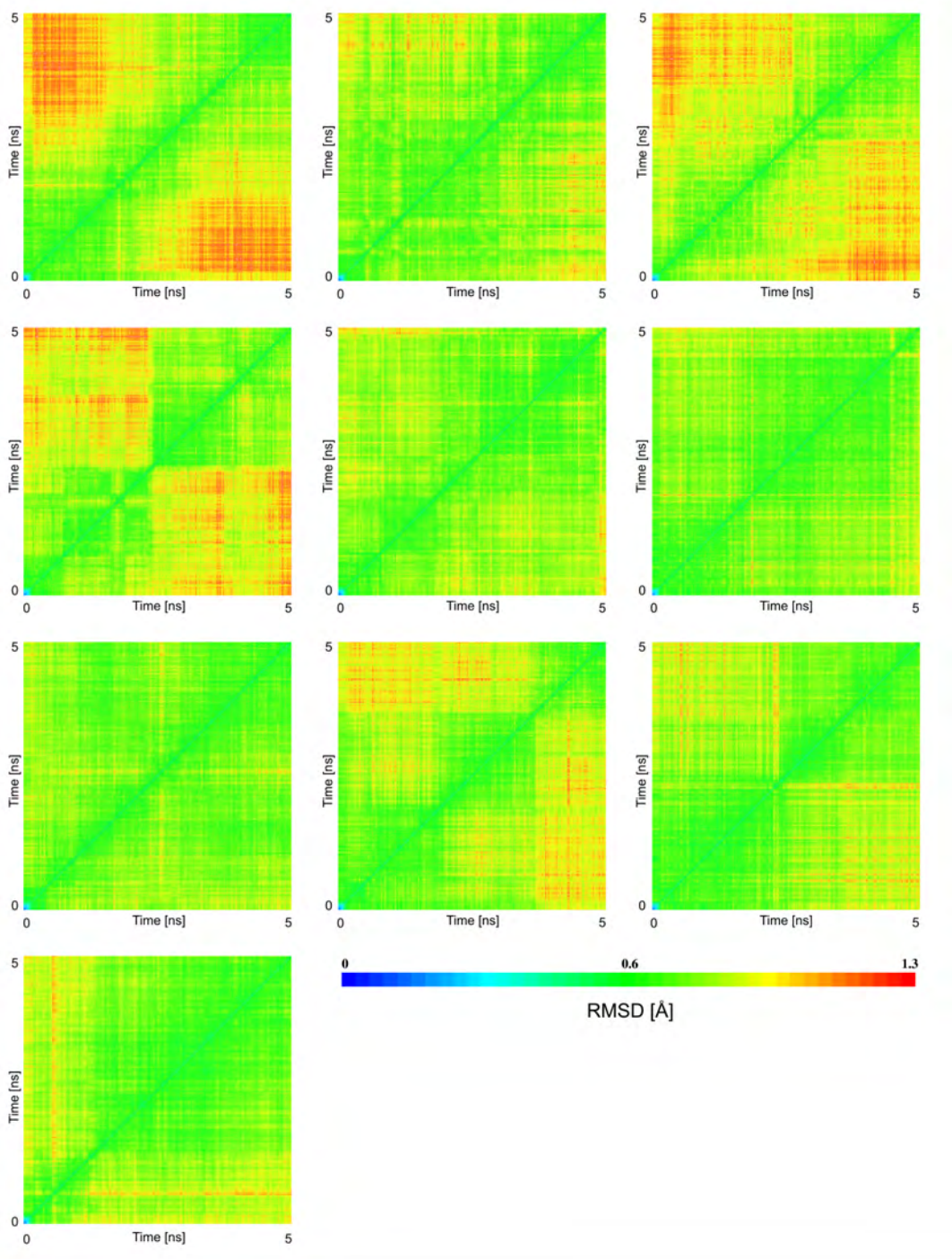


Figure 8: 2D-RMSD plots of the 10 MMD_{wat} simulations

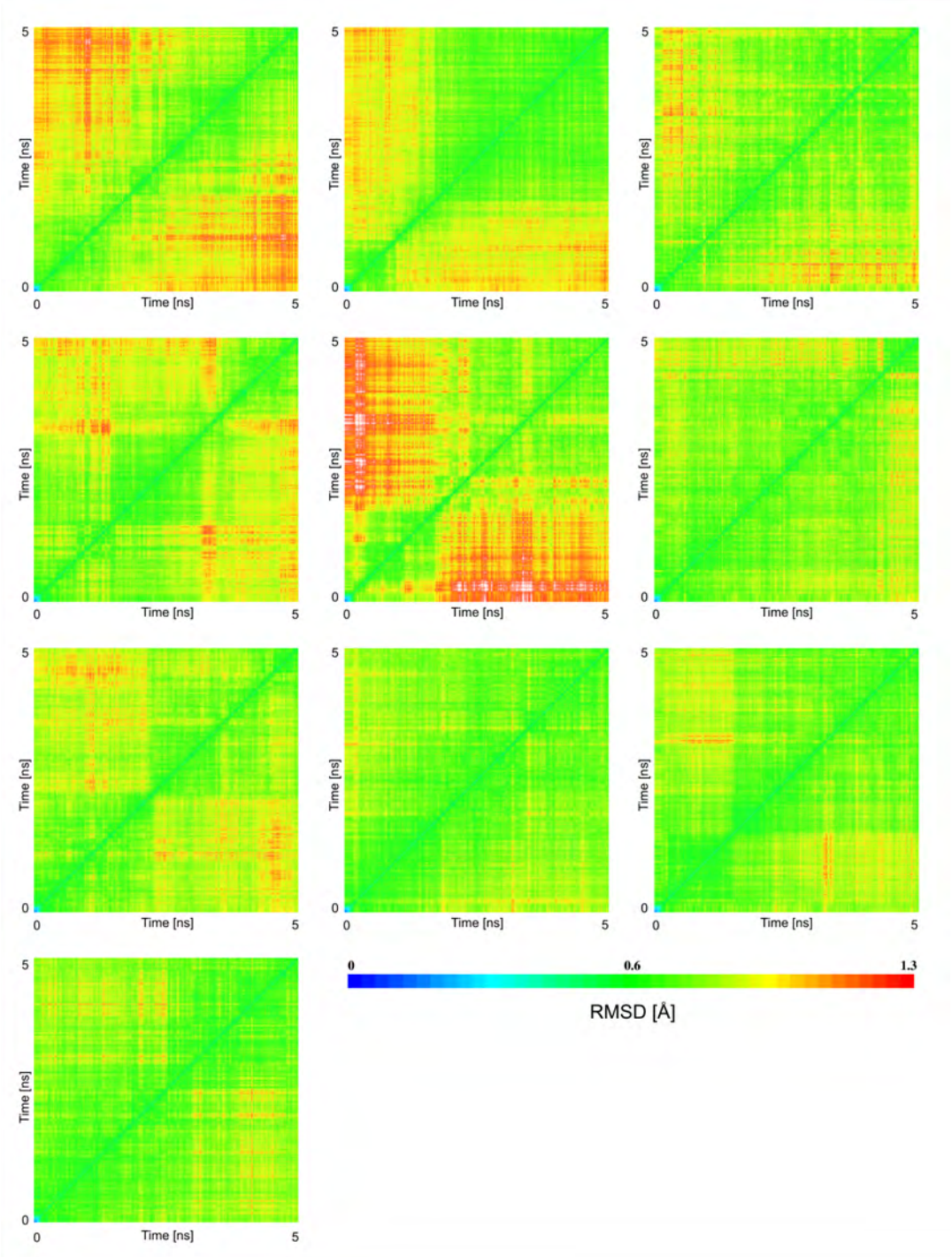


Figure 9: 2D-RMSD plots of the 10 MMD_{nowat} simulations

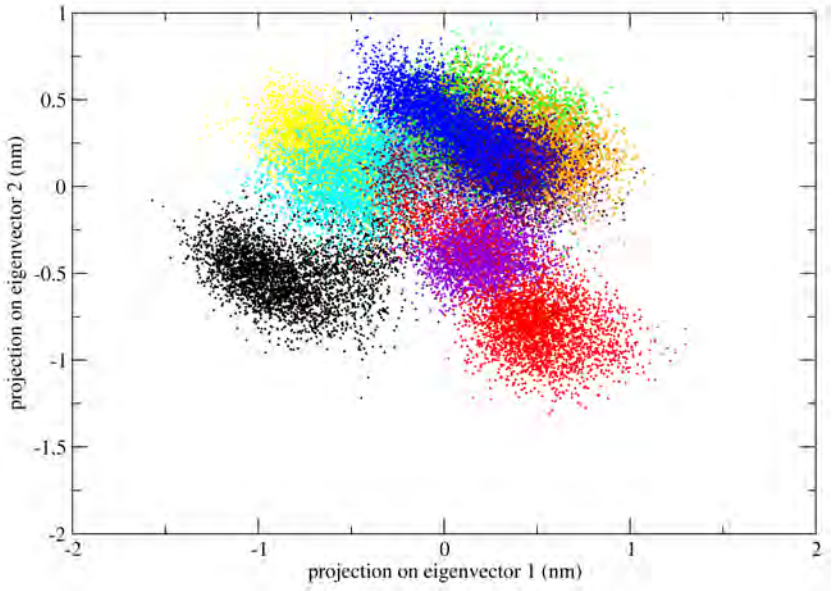


Figure 10: 2D projection of the MMD_{nowat} simulations onto the first and second eigenvector of the combined trajectory. Individual MD simulations are distinguished by colors.

Referierte Publikation, erschienen in: Proceedings of the 16th IASTED International Conference on Parallel and Distributed Computing and Systems, Acta Press, 34 – 39, 2004. ISBN:0-88986-421-7

A MANAGEMENT SYSTEM FOR COMPLEX PARAMETER STUDIES AND EXPERIMENTS IN GRID COMPUTING

Natalia Currle-Linde[1], Fabian Boes[2], Peggy Lindner[1], Juergen Pleiss[2], and Michael M. Resch[1]
[1]High Performance Computing Center Stuttgart (HLRS)
University of Stuttgart
Allmandring 30, 70550 Stuttgart, Germany
{linde, lindner, resch}@hlrs.de

[2]Institute of Technical Biochemistry
University of Stuttgart
Allmanding 31, 70569 Stuttgart, Germany
{fabian.boes, juergen.pleiss}@itb.uni-stuttgart.de

Abstract — Although some known parameter study tools [1] have successfully been used to solve standard parameter study problems, they are not yet sufficient to handle the current generation of complex parameter investigations. Furthermore, time constraints typically require that multiparametric tasks are executed on a large number of parallel computational resources. Grid Computing potentially helps by enabling access to vast computational resources. But in practice the application of multiple Grid resources to complex tasks leads to considerable programming effort for the scientist. One way to reduce both the work and the cost associated with performing parametric investigations of complex scientific experiments could be to automate the creation of complex modules for the computation and the dynamic control of the study, for example through the use of the results from previous computational stages. In this paper we propose a concept for the design and implementation of an automated parametric modeling system for producing complex dynamically-controlled parameter studies without the use of any specific programming language in the process of parameterization, computation and the control of the experiment.

Keywords — Distributed Computing, Parameter Studies, Grid Computing.

I. MOTIVATION AND BACKGROUND

Parameter sweep tasks are typically easy to parallelize. Thus, for this type of task, a parallelized distributed computational model is very appropriate. Until quite recently, thorough parameter studies were limited by the availability of adequate computer resources [1], [7]. The advent of global Grid technology makes it possible to integrate resources from distributed scientific centers with those of one's own environment, creating a technical basis for complex parametric investigations [7].

The automatic generation of parameter studies is suitable only for those cases with simple information processing. For these cases, the complete information for processing of user-(task) information consists of a succession of previous and next wizards which, in turn, determine the processing sequence. This model can only be applied to single-level parametric modeling; it is not suitable for the description of complex processes.

Complex processes, as a rule, result in several levels of parameterization, repeated processing and data archiving, conclusions and branches during the processing, as well as synchronization of parallel branches and processes.

In these cases the parameterization of data is an extremely difficult and work-intensive process. Moreover, users are very sensitive to the level of the automation for application preparation [7], [13]. They must be able to define a fine grained logical execution process, identify the position in the input area of the parameters, to be changed in the course of the experiment, as well as formulate the parameterization rules. All other remaining details of the parameter study generation should be hidden from the user. We introduce here a parallel distributed processing model for the implementation of complex parameter studies which provides new aspects for handling user processes.

Furthermore, the lack of well-organized data storage for complex parameter studies is a problem for the scientist [7]. This makes it extremely difficult to handle hundreds, or even thousands, of experimental data volumes. A database is the ideal means for the storage of input data, intermediate results and output data of the experiment, as well as for the analysis of results. Fast search capability, storage reliability and the availability of information for the parameter study application accommodate our demands for fast and precise information selection at the start of an experiment.

Modern parametric studies, which require computer resources from diverse locations, are supposed to fit perfectly the idea of Grid Computing [2]. Grid environments are very dynamic - some resources are made available for users, others, however, are occupied with the execution of other processes. Therefore, the

parameterstudy management system must be capable of adapting to the resources currently available in the network. The introduction of a resource broker mechanism for dynamic redirection of the jobs to a parameter study dispatcher during the execution of the experiment enables the achievement of a result within a minimum time.

The organization of this paper is as follows: in chapter II we first refer to the related work in the field of parameter sweep support tools and show how our system differs from these projects. In the following two chapters (III and IV) we describe the architecture and implementation of our own system. Chapter V shows a first Use Case of the experiment system in the field of bio-molecular dynamics. Finally we give a summary and an overview of the ongoing work.

II. Related work

For parameter studies there are currently a variety of different approaches (e.g. Nimrod, ILab, AppLeS/APST). Nimrod [3], [4] is a tool that can be used to manage the execution of parametric studies across distributed computers. It supports the creation of parameter sweeps based on a simple declarative language. Therefore, Nimrod can only be applied for single-level parametric modeling. With Nimrod/G [5] and Nimrod/O [6] the Grid services provided by Globus [12] are used for job launching, resource brokering and scheduling to allow the efficient execution of the parameter studies within a certain budgetary requirement.

Compared with Nimrod, ILab [7], [8], [9] allows the generation of multi-parametric models and adds workflow management. With the help of a sophisticated GUI, the user can integrate several steps and dependencies within a parameter study task. ILab, developed at NASA, supports the execution of parameter sweep jobs at the IPG (NASA's Information Power Grid). Although ILab provides the possibility to realise complex parameter studies, the complexity is limited by the CAD (Computer Assisted Design) screen which does not support many nested levels.

Triana [11] is an open-source problem solving environment that abstracts the complexities of composing distributed workflows. It provides a "pluggable software architecture" to be used for the dynamic orchestration of applications from a group of predefined commodity software modules.

Beside the above mentioned environments, tools like Condor [14], UNICORE [16] or AppLeS (Application-Level scheduler) [10] can be used to launch pre-existing parameter studies within distributed resources.

Within all the aforementioned tools, real dynamic parameterizations are not supported which are necessary for a number of applications, especially in the field of optimization. In these cases the result of the simulation has to fulfill certain criteria. For instance, some value or average value has to be larger or smaller than a given threshold. For these applications an iterative approach is required. The result of a simulation will be compared to the desired criteria. If the criteria are met the simulation is finished. If not, the application automatically has to update its input files and has to restart the simulation cycle. The purpose of our work presented here is specifically to close that gap and provide a tool that gives some support for dynamic parameterization.

III. System Architecture and Implementation

Figure 1 shows the system architecture of the experiment management system. It consists of three main components: the User Workstation (Client), the Parameter Study Server and an object-oriented database (OODB). The system operates according to a Client-Server-Model in which the Parameter Study Server interacts with remote target computers using a Grid Middleware Service. The software follows the Java 2 Platform Enterprise Edition (J2EE) architecture and we chose the JBOSS [18] implementation to realize it. The software is written in Java and runs on Windows, as well as on UNIX platforms. To integrate the OODB the Java Data Objects (JDO) implementation of FastObjects [17] is used.

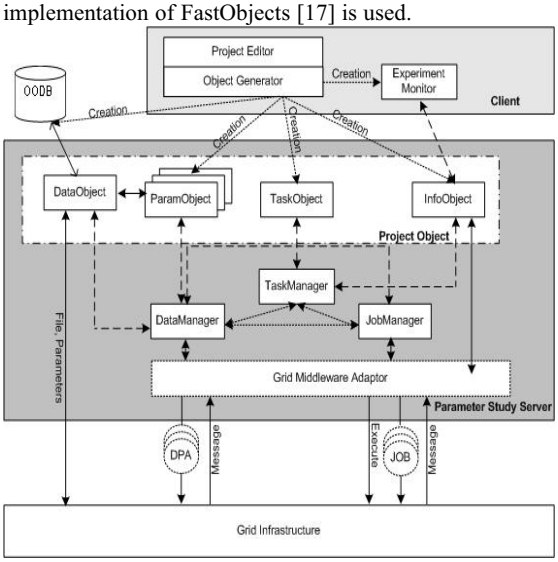


Figure 1. System architecture

The client on the user's workstation is composed of three subsystems: the ProjectEditor, the ObjectGenerator and the ExperimentMonitor. The ProjectEditor is the graphical user interface which is used to design the parameter study. It contains module icons for the specification of the experiment and interfaces for editing the project specific OODB. With the help of nested interfaces, the user is able to define the settings for each module individually.

The Object Generator initializes the creation of all further program objects for each parameter study project: the individual ParamObjects (containing the parameter set and its rules for modification), a TaskObject (containing

the task definition), a DataObject (containing information about the OODB) and an InfoObject are generated. These are all Java objects, created and administered on the server side. Furthermore, an ExperimentMonitor object will be generated for visualization and for the actual control of the complete process. The ExperimentMonitor interacts with the InfoObject. The InfoObject collects all the information about the status of the computer resources in the network, passes this to the TaskManager (see below) and accumulates status information about the running tasks.

The server also consists of three subsystems: the TaskManager, the JobManager and the DataManager. The TaskManager is the central component of the Parameter Study Server. It chooses the necessary computer resources for each computation stage after receiving information about their availability from the InfoObject. The InfoObject collects this information from the underlying infrastructure using existing services like the Globus resource broker. The TaskManager then informs the DataManager and the JobManager about the chosen resources and starts/stops each branch and blocks. It also takes care of synchronizing the events and controlling the message exchange between the parallel task branches. On the user's request, the TaskManager can block the program flow. It informs the InfoObject about the current status of the jobs and the individual processes. The JobManager and the DataManager work closely together to control the interaction of the task and the Grid resources. They synchronize the preparation of the data as well as the initiation of the jobs. The DataManager controls the parameterization, the transport of data and the parameter files into the working directories of the remote target computers as well as their exchange and the archiving of the result files and the parameters in the experiment's database. The preparation of the actual input files for the execution of the jobs is done with the help of the Data Preparation Agent (DPA). The JobManager generates jobs, places them in the queue and observes their execution.

The automatic creation of the project specific OODB is done according to the structure designed by the user. The database collects all relevant information for the realization of the experiment such as input data for the parameter study, parameterization rules, etc. A detailed description for designing a project parameter study is discussed in chapter IV.

To realize a smooth interaction of the Parameter Study Server with currently existing Grid middleware, the DataManager and the JobManager do not communicate directly with the Grid resources, but rather via the Grid Middleware Adaptors. These adaptors are used to establish the communication to existing Grid middleware services like job execution and monitoring services. Currently, e.g. Globus [12] and UNICORE [15] offer these services.

At the moment a UNICORE adaptor is under development. In the current implementation we have adapted the Arcon command line client to work as a UNICORE interface within the adaptor. With the help of the Arcon library, developed at Fujitsu Laboratories of Europe for the UNICORE project, the TaskManager is able to create an Abstract Job Object (AJO) [16] which is necessary to run jobs on UNICORE sites. The AJO contains the job description. The UNICORE adaptor submits this AJO to a specified UNICORE site and processes the outcome of the UNICORE job. The required information to establish connections to UNICORE servers is stored within the InfoObject.

IV. PARAMETER MODELING FROM THE USER'S VIEW

This chapter explains the process of designing a parameter study using the proposed system. Figure 2 shows an example of a task flow of such an experiment as it will appear in the ProjectEditor. The graphical description of the application flow has two purposes: firstly, collecting all information for the creation of the Java objects described in the previous chapter and, secondly, to be used for the visualization of the current experiment in the ExperimentMonitor. For instance, the current point of execution is emphasized in a specific color within a running experiment.

For designing the experiment, there are various facilities provided for denoting the program blocks, paths, branching, events, conditions, messages etc. The modules can be linked themselves, providing the direction of the data flow and the sequence of the computation processes. Each module has its own ID and can contain different properties, consisting of parameters and signals.

To specify these properties, the user can allocate for example file names, arrays of dates, individual parameters, their values etc. Signals for modules describe initial and subsequent starts or permit further steps after the execution of the corresponding operations. A complex experiment system can contain various levels of nested blocks.

Within the control flow (see Figure 2) the user defines the rules and conditions for the execution of the modular computation process (logical schema of the experiment). Two different kinds of blocks are used: computational blocks and control blocks. Each computational block can represent a simple parameter study. On this block level, the manipulation of data will be handled. The control block is used to define branches, conditions or synchronization points.

In the example shown in Figure 2 the whole experiment is split into i-branches, which can be executed in parallel due to the logic of the experiment. Each branch consists of several program blocks interconnected by arrowed lines, which indicate their sequence of execution and/or possibility of parallel execution. The diagram also displays the control modules and the synchronization process. Using the information contained in each module, the system can begin to branch, collect and synchronize the processes as well as exchange messages between the processes during the execution of the experiment.

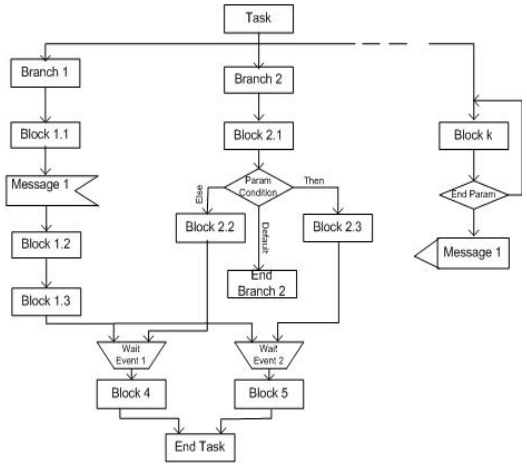


Figure 2. Sample Task Flow (control flow)

However, the definition of a specific part of a program can be done within each module in a very fine grained way as shown in Figure 3: the experimental Block 1.2 consists of computing modules, parameter modules, data replacement modules and the corresponding elements of the experiment OODB. These are connected to each other with arrowed lines showing the direction of data transfer between modules and the sequence of execution during the computation process. In this example, not only simple, but also nested formation cycles of sets of dependant parameters are represented. The transition from one program block to the next can also be carried out. This allows an optimal simplified experimental design process.

Before starting the execution of the whole experiment the user can do a simulated run to verify the correct chain of the logical process of the experiment. In this case an automatic or stepwise tracking of the experiment can be done within the ControlFlow window of the ProjectEditor. The actual active module, block or process is displayed with colored pulsing points. If necessary the user can now adjust the course of the experiment.

The execution of the designed experiment will start by activating the TaskManager. The TaskManager has complete control of the experiment's actions. It first chooses the computer resources currently available in the network, and then activates all the parallel processes of the experiment's first stage. Afterwards, the TaskManager starts the DataManager and JobManager. The DataManager first transfers input data to the working directory of the corresponding target computer, and second activates the ParamObjects for generating the first parameter set. The generated parameter set is linked with replacement processes and then delivered to the corresponding target computer. After the replacement of the specified parameters they are all ready for the first stage of computation. Concurrently to these processes, the JobManager prepares all the jobs for the first stage. After receiving a status message regarding the availability of all parameterized input data, the first jobs will be started.

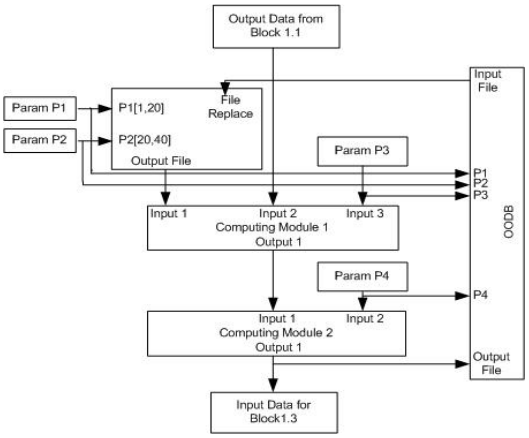


Figure 3. Block 1.2 (data flow)

After the preparation of this value for the first parameter set, the DataManager initializes the preparation of input files, which contain the values of the second parameter set. The JobManager prepares a job set for the processing of the second set of the parameter. After receiving a notification about the availability of data resources, the JobManager sends the job set to an appropriate remote computer. The preparation of further stages is accomplished in the same manner. After receiving a notification of the completion of the first stage of computation and the execution of the transfer operation to the next one, the TaskManager analyses the current status of computer resources, chooses the most suitable target computer for the next stage and starts the parallel process for this. In all stages, the output file is archived immediately after being received by the experiment's database. The control of all processes of the next respective information processing levels always takes place according to the pattern described above.

After starting the ExperimentMonitor on their workstation, the user receives continuously updated status information regarding the experiment's progress.

V. USE CASE – MOLECULAR DYNAMICS SIMULATION OF PROTEINS

Molecular dynamics (MD) simulation is one of the principal tools in the theoretical study of biological molecules. This computational method calculates the time dependent behavior of a molecular system. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids. These methods are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their interaction with substrates, ligands, and inhibitors.

The goal of a project currently under development at the Institute of Technical Biochemistry at the University of

Stuttgart is to establish a general, generic molecular model that describes the substrate specificity of enzymes and predicts short- and long-range effects of mutations on structure, dynamics, and biochemical properties of the protein. It is based on a large number of long time scale MD simulations of enzyme variants with covalently bound substrate intermediates, solvated in a water box.

A typical MD simulation is divided into three phases. After the energy minimization, in which the system is relaxed to a local energy minimum, the equilibration phase follows. During this phase the system approaches an equilibrium. The duration of this phase can vary greatly. Its duration cannot be predicted and has a high sensitivity to initial conditions. Reaching the equilibrium, however, is a prerequisite to starting the production phase, during which equilibrium properties of the system are assembled.

To analyze the behaviour of an enzyme-substrate complex and to study the effect of mutations to substrate specificity, multiple MD simulations are carried out by systematically varying initial velocities, exchanging substrates and individual amino acids. While most projects on MD simulation are still managed by hand, large-scale parameter studies may involve up to a thousand MD simulations. Therefore this is worthwhile to be controlled by a workflow system. The proposed experiment management system provides all the necessary functionalities to design complex MD parameter studies combined with control of job execution in a distributed computing environment. Furthermore, the system will help users to run experiments automatically according to predefined criteria.

Figure 4 shows the example usage of the ProjectEditor GUI to generate a complete MD simulation task. Energy minimization and MD simulations will be performed using the *Sander* program of the *AMBER 7.0* software package. Starting from a user provided molecular system including the enzyme, the substrate and the surrounding solvent, X simulations will be generated. X is the number of simulations which can be executed at the same time and therefore can be parallelized. Every simulation consists of the same molecular system, but uses a different initial velocity distribution for the simulation in the parameter file. The workflow follows the aforementioned scheme for MD simulations. First the system will be relaxed to a local energy minimum (energy minimization phase). In the following phases some results from the previous phases are used and therefore they have to be run in sequential order.

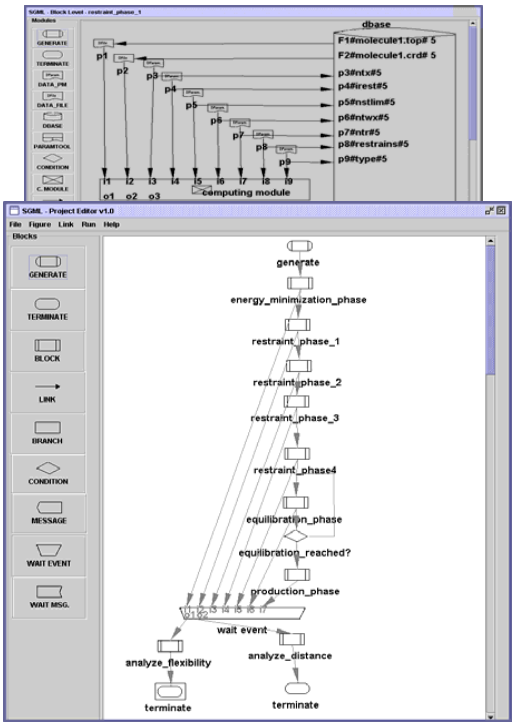


Figure 4: Screenshot of bio-molecular dynamics experiment

After finishing the first phase, a restraint phase follows in which the system is adapted to the simulation temperature using positional restraints on the enzyme atoms. The restraint phase consists of several simulation runs with a decreasing restraint force parameter. As an example, details for the restraint phase 1 are shown in Figure 4 as well. In this case, 9 Parameter Modules are used on the DataFlow level to create the parameterized input files and to collect all the information for the X simulation runs. Within the OODB interface, the user enters all settings.

In the equilibration phase, the enzyme is allowed to move freely and should reach an equilibrium. 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 the simulations will be compared to each other and the task is terminated.

In the final version, multiple simulations of each enzyme-substrate combination will be performed with ten different initial velocity distributions. After heating the protein-substrate-solvent systems and subsequent equilibration, a trajectory of at least 2 ns will be generated (50 processor days for each simulation). For analysis, the geometry of the substrate binding site and the flexibility will be systematically compared between the complexes. Biologists expect that changes in the geometry of the active site and the distances between the catalytic residues and the substrate will correlate with the experimentally determined activity. It will be essential to determine a set of parameters, which are indicative of catalytic activity.

To generate the model, a total of up to of 3000 (30 variants x 10 substrates x 10 velocity distributions) MD simulations must be set up, performed and analyzed, utilizing more than ten different programs. Each simulation will typically produce a trajectory output file with a size of several gigabytes, so that data storage, management and analysis become a serious challenge. These tasks can no longer be performed interactively and therefore have to be automated.

VI. Conclusion

This paper presented the concept and description of a first implementation of a management system for design of complex and hierarchical parameter studies which offers an efficient way to execute scientific experiments. The proposed parameter study tools define both the means to design a complex parameter study and the means for the control of the experiment in a distributed computer network.

Currently, the graphical user interface for the creation of the experiment objects is already implemented. The automatic creation of the individual experiment OODB and their integration into the JBoss framework has been done. The workflow creation has been tested with an application framework from the area of bioinformatics. For test purposes a simple ssh adaptor was developed to start jobs on remote resources. An adaptor for UNICORE is currently under development. The implementation of a Globus adaptor to integrate resource broker services will be done in the future.

References

[1] de Vivo, A., Yarrow, M., McCann, K.: A comparison of parameter study creation and job submission tools; Technical report NAS-01-002, NASA Ames Research Center, Moffet Filed, CA, 2000.

[2] Foster, I.: The Grid: A New Infrastructure for 21st Century Science; Physics Today, 55(2):42-47, 2002.

[3] Abramson, D., Giddy, J. and Kotler, L.: High Performance Parametric Modeling with Nimrod/G: Killer Application for the Global Grid?; International Parallel and Distributed Processing Symposium (IPDPS), pp 520-528, Cancun, Mexico, May 2000.

[4] Buyya, R., Branson, K., Giddy, J., and Abramson, D.: The Virtual Laboratory: Enabling Molecular Modeling for Drug Design on the World Wide Grid; The Journal of Concurrency and Computation: Practice and Experience (CCPE), Volume 15, Issue 1, Pages: 1-25, Wiley Press, USA, January 2003.

[5] Buyya, R., Abramson, D., Giddy, J.: Nimrod, G.: An architecture for a Resource Management and Scheduling System in a Global Computational Grid; The 4th International Conference on High-Performance Computing in Asia-Pacific Region (HPC Asia 2000); Beijing / China, May 2002.

[6] Abramson, D, Lewis, A. and Peachy, T.: Nimrod/O: A Tool for Automatic Design Optimization, The 4th International Conference on Algorithms & Architectures for Parallel Processing (ICA3PP 2000), Hong Kong, 11 - 13 December 2000.

[7] Yarrow, M., McCann, K., Biswas, R., van der Wijngaart, R.: An Advanced User Interface Approach for Complex Parameter Study Process Specification on the Information Power Grid; Proceedings of the 1st Workshop on Grid Computing (GRID 2002); Bangalore / India, December 2000.

[8] Yarrow, M., McCann, K. M., Tejnil, E., and deVivo, A.: Production-Level Distributed Parametric Study Capabilities for the Grid; Grid Computing - GRID 2001 Workshop Proceedings, Denver, CO, November 2001.

[9] McCann, K. M., Yarrow, M., deVivo, A., Mehrotra P.: ScyFlow: An Environment for the Visual Specification and Execution of Scientific Workflows; GGF10 Workshop on Workflow in Grid Systems, Berlin, 2004.

[10] Casanova, H., Obertelli, G., Berman, F., Wolski, R.: The AppLeS Parameter Sweep Template: User-Level Middleware for the Grid; Proceedings of the Super Computing (SC 2002) Conference, Dallas / USA, 2002.

[11] Taylor, I., Shields M., Wang I., and Philp, R.: Distributed P2P Computing within Triana: A Galaxy Visualization Test Case; IPDPS 2003 Conference, Nice, 2003.

[12] Foster, I., Kesselman C.: The Globus Project: A Status Report; Proc. IPPS/SPDP '98 Heterogeneous Computing Workshop, pp. 4-18, 1998.Describes the status of the Globus system as of early 1998.

[13] Burq, S., Melnikoff, S., Branson, K., and Buyya, R.: Visual Parametric Modeler for Rapid Composition of Parameter-Sweep Applications for Processing on Global Grids; Java in Computational Science Workshop, Proceedings of the 3rd International Conference on Computational Science (ICCS 2003), Springer Verlag Publications (LNCS Series), June 2 - 4, 2003, Melbourne, Australia.

[14] Thain, D., Tannenbaum, T., and Livny, M.: Condor and the Grid; in Fran Berman, Anthony J.G. Hey, Geoffrey Fox, editors, Grid Computing: Making The Global Infrastructure a Reality, John Wiley, 2003. ISBN: 0-470-85319-0.

[15] Erwin, D.: UNICORE – A Grid Computing Environment; Grid Computing Environments 2001 - Special Issue of Concurrency and Computation: Practice and Experience.

[16] Erwin, D. (Ed.): 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.

[17] FastObject webpage: http://www.fastobjects.com

[18] JBoss webpage: http://www.jboss.org/products/jboss

GriCoL: A Language for Scientific Grids

Natalia Currle-Linde, Panagiotis Adamidis, Michael Resch
High Performance Computing Center (HLRS)
University of Stuttgart
Nobelstrasse 19, 70550 Stuttgart, Germany
(linde, adamidis, resch)@hlrs.de

Fabian Bös, Jürgen Pleiss
Institute of Technical Biochemistry
University of Stuttgart
Allmandring 31, 70569 Stuttgart, Germany
(fabian.boes, juergen.pleiss)@itb.uni-stuttgart.de

Abstract

The development of the Grid has opened new possibilities for scientists and engineers to execute large-scale modeling experiments. This has stimulated the generation and development of tools for the creation and management of complex computing experiments in the Grid. Among these, tools for the automation of the programming of experiments play a significant role. In this paper we present GriCoL, which we propose as a simple and efficient language for the description of complex Grid experiments. We also describe the environment within which GriCoL works, namely the Science Experimental Grid Laboratory (SEGL) system for designing and executing complex experiments. As an illustration, the paper includes the description of a biochemical experiment using Molecular Dynamics simulations.

1. Introduction

Today, the solution of many large-scale problems in fields such as physics, biology, economics and management needs enormous computer resources, which can often be provided only by the Grid [1]. The description and execution of such complex experiments is not a trivial task and demands considerable effort from the creator of the experiment because existing workflow systems do not offer a sufficiently high organizational level to specify and execute complex modeling experiments. The reason for this can be explained as follows:

- Existing instruments for the description and execution of complex experiments work with specific modules tailored to a corresponding application domain and not with standard adjustable mathematical modules. Therefore, experiment-specific programs must be designed every time a new experiment is created.
- An important characteristic of a language for the description of complex experiments, which requires improvement, is the possibility to generate an executable program of maximum efficiency. This implies that the language must be organized in such a way as to ensure maximal parallelization of all processing, both between and within language elements. In other words, the system must at the inter-element layer provide possibilities of parallel execution on all branches of the experiment as well as pipelined processing of data in the nodes of the experiment program which are connected in sequence.
- A program for a complex experiment requires the description not only of the logic of the experiment but also of the control of complex data flows during the process of its execution. Practically all existing workflow systems have only primitive mechanisms of data flow control which take little account of the dynamic structure of the Grid.
- Existing systems commonly do not use any universal instruments which provide a centralized information space (a database) for the experiment. This means these systems have problems when describing the interaction between different nodes of the experiment.
- Although workflow systems generally have graphical tools for the description of modeling experiments,

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

these are not sophisticated enough to provide the level of precision or intuitive clarity that the user needs for the description of complex experiments.

For the design of Grid Concurrent Language (GriCoL), a language for describing complex modeling experiments, we aimed to overcome the above limitations of existing systems, which are described in the second section. In particular, GriCoL was conceived on the principle of a two-level description of an experiment program. This enables simple descriptions of the logic of execution as well as of the data flow of highly complex experiments (consisting of several hundred components). Consequently, GriCoL was envisaged to offer scientists the possibility to use all types of parallelism in an experiment program. A more detailed description of these and other important features of GriCoL is given in the third section. The fourth section presents the architecture of SEGL [2], the system within which GriCoL is utilized for designing and executing complex modeling experiments. The fifth section presents the realization of a biochemical experiment in the field of Molecular Dynamics (MD) simulations.

2. Related Work

There are many workflow management systems such as Kepler [3], Triana [3], Pegasus [3]. A detailed survey of existing systems along with a taxonomy which can be used for classifying and characterising workflow systems is provided in [3]. A fundamental difference between approaches is the workflow structure. Many systems, such as Pegasus and GridFlow [3], represent their workflows as a Directed Acyclic Graph (DAG) which allows a workflow structure of sequence, parallelism and choice. But they are limited with respect to the power of the model; iteration is not part of the DAG structure. Workflow-based systems such as GSFL [4], and BPEL4WS [4] have solved these problems but are too complicated to be mastered by an average user. With these tools, even for experienced users, it is difficult to describe non-trivial workflow processes involving data and computing resources.

Another fundamental difference is the use of either the data flow or control flow programming paradigm. Most existing workflow systems, including Kepler and Triana, use the data flow paradigm. This means that control flow such as iteration is difficult to design and maintain and such systems cannot carry out dynamic, self-steering iterations as in an application such as a parameter study. GriCoL overcomes these deficiencies by supporting multiple levels of description of the experiment program, with control flow at the top level and data flow nested within this.

There are other workflows, such as SOMA [5], which are tailored only to a corresponding application domain. SOMA is a grid workflow system for small molecule property calculations and aims to improve the efficiency in designing molecular structures with desired properties. However, it is a highly specific tool for its application domain and currently focuses on molecular structure and property calculation workflows.

There are some efforts in implementing tools for parameter investigation studies e.g. Nimrod [6] and Ilab [6]. These tools are able to generate parameter sweeps and jobs, running them in a Grid and collecting the data. ILab also allows the calculation of multi-parametric models in independent separate tasks in a complicated workflow for multiple stages. However, none of these tools is able to perform the task dynamically by generating new parameter sets by an automated optimization strategy as is needed for handling complex parameter problems. In addition to the above mentioned environments, tools like Condor [6], UNICORE [7] or AppLeS [6] (Application-Level Scheduler) can be used to launch pre-existing parameter studies using distributed resources. These, however, give no special support for dynamic parameter studies.

Complex parameter studies can be facilitated by allowing the system to dynamically select parameter sets on the basis of previous intermediate results. This dynamic parameterization capability requires an iterative, self-steering approach. Possible strategies for the dynamic selection of parameter sets include genetic algorithms, gradient-based searches in the parameter space, and linear and nonlinear optimization techniques. An effective tool requires support of the creation of applications of any degree of complexity, including unlimited levels of parameterization, iterative processing, data archiving, logical branching, and the synchronization of parallel branches and processes.

3. Grid Concurrent Language

The philosophy of GriCoL is based on a known fact: despite the wide variety of complex application tasks across different fields of science and technology, the set of components (e.g gradient searches, genetic algorithms) within these tasks is very limited. These components, once they have been created in standard form, can be repeatedly reused for modeling complex systems. If one of the desired components is not available, it is much simpler to implement this component and add it to the language than to generate a new complex application. Also, existing languages for Grid computing suffer from an insufficiently high abstraction level which makes the programming of complex experiments a difficult and painstaking job. Therefore, this paper proposes the following solution: a component-based language for describing complex modeling experiments with a sufficient level of abstraction so that the scientist does not require a knowledge of the Grid or of parallel

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

programming.

3.1 Common properties and principle of organization

GriCoL is a universal language for programming complex computer- and data-intensive tasks without being tied to a specific application domain.

GriCoL is a graphical-based language with mixed type and is based on a component-structure model. The main elements of this language are blocks and modules, which have a defined internal structure and interact with each other through a defined set of interfaces. In addition, language elements can have structured dialog windows through which additional operators can be written into these elements.

The language is of an entirely parallel nature. It can implement parallel processing of many data sets at all levels, i.e. inside simple language elements (modules); at the level of more complex language structures (blocks) and for the entire experiment.

In general, the possibility of parallel execution of operations in all nodes of the experiment program is unlimited. (It is limited only by the logic of the experiment's execution.).

In order to utilize the capacities of supercomputer applications and to enable interaction with other language elements and structures, we make use of the principle of wrapping of functionality in components. Practically all GriCoL language elements have been designed on this principle.

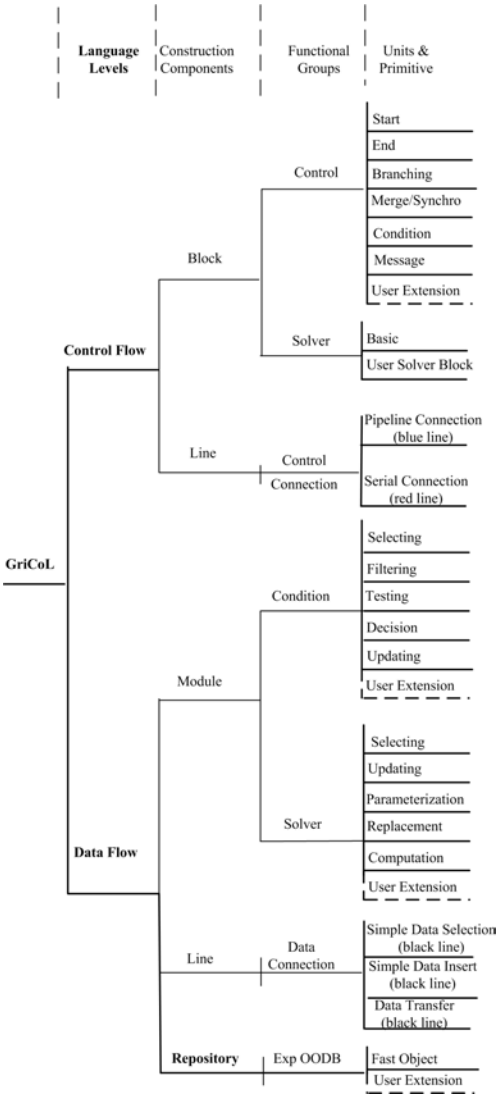
An additional property of the language is its extensibility through the use of components. With the help of the dialog program Units Library Assistant, it is possible to add new functional modules to the language library. Program codes for the parallel machine are wrapped in the standard language capsule. This wrapping principle allows GriCoL to incorporate legacy code.

Another important property of the language is that it is multi-tiered. The multilayer principle (the sub-division into control flow and data flow) as well as the graphical context of the language makes it much easier than with a Grid workflow system for the user to understand and describe the experiment.

3.2 Structure and components

GriCoL combines the component model of programming with the methodology of parallel programming. The language represents a multi-tiered model of organization. This enables the user when describing the experiment to concentrate primarily on the logic of the experiment program and subsequently on the description of the individual parts of the program. Now we describe the internal organization of this language. The elements of GriCoL are shown in Figure 1.

Figure 1. Structure of GriCoL



Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

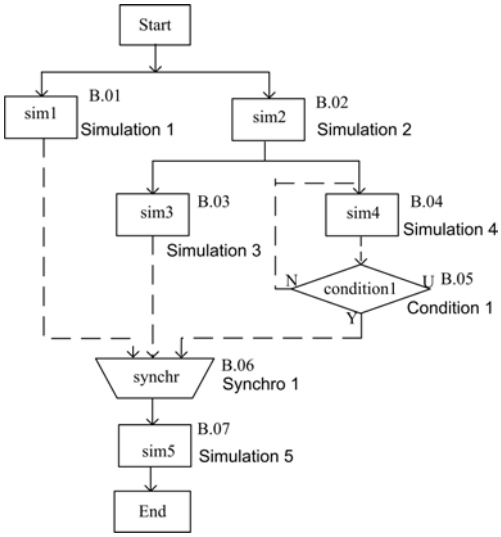
GriCoL has a two-layer model for the description of the scientific experiment and an additional sub-layer (Repository) for the description of the database required for the creation of the experiment. The top level of the experiment program is the control flow level, which describes the logical sequence of execution. The main elements of this level are blocks: control blocks and solver blocks. A solver block is the program object which performs some complete operation. The standard example of a solver block can be a simple parameter sweep. The control block is the program object which allows the changing of the sequence of the execution according to a specified criterion. The lower level, the data flow level, provides a detailed description of components at the top level, the control flow level. The main elements of the data flow level are program modules and database sections. The sublayer provides a common description of the database and a section for making additions to the database if necessary.

The elements of the language have graphical notation and are represented by icons (for modules and blocks) or as connection lines.

Control Flow Level

Figure 2 illustrates the above mentioned for an experiment at the control flow level.

Figure 2. Example of an experiment program. Control flow level



As can be seen from this figure the language components make it possible to generate multilayer dynamic-control experiment programs with branches. GriCoL offers the user a complete range of control mechanisms on experiment processes: parallelization, testing of conditions and branching, synchronization and fusion, as well as exchange of messages and signals.

Solver blocks represent the nodes of data processing. Control blocks are either nodes of data analysis or nodes for the synchronization of data computation processes. They evaluate results and then choose a path for further experiment development. Another important language element on the control flow level is the connection line. Connection lines indicate the sequence of execution of blocks in the experiment and, together with control blocks, describe the logic of execution of the experiment program. There are two mechanisms of interaction between blocks which are described with the help of connection lines (either red-solid or blue-dashed). If the connection line is blue in colour, the procedure is as follows: each time the computation of an individual data set has been finished, i.e. after completion of a program run within a block, control is transferred to the next block. This process is repeated until all program runs in the block have been completed. That means a pipelined operation on the set of runs. If the connection line is red in colour, control is not passed to the next block before all runs in the previous block have been finished. That means a barrier on the set of runs.

Experiment operations always begin with the start block and finish with the end block. In the example (see Figure 2), the start-block begins a parallel operation in solver blocks B.01 and B.02. After execution of B.02, processes begin in solver blocks B.03 and B.04. Each data set which has been computed in B.04 is evaluated in control block B.05. The data sets meeting the criterion are selected for further computation. These operations are repeated until all data sets from the output of B.04 have been evaluated. The data sets selected in this way (in our example these of input I3 of B.06) are synchronized by a merge/synchronize block with the corresponding data sets of the other inputs of B.06 (I1 and I2). The final computation takes place in solver block B.07.

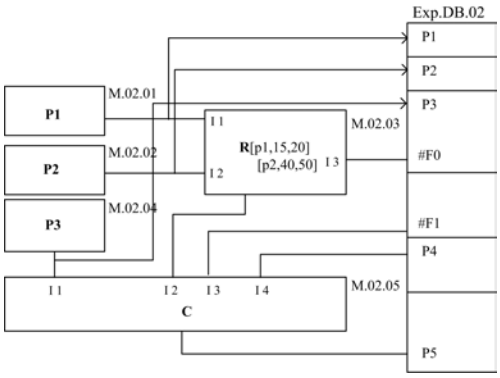
Data Flow Level

A detail of programming on the data flow level is represented on Figure 3.

A typical example of a solver block program is a modeling program (or a program fragment) which cyclically computes a large number of input data sets. At this level, the user can describe the manipulation of data in a very fine grained way. The solver block consists of computation (C), replacement (R), parameterization (P) modules and a data base (Exp.DB). These are connected to each other with connection lines showing the data transfer between modules

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

Figure 3. Example of a solver block program. Data flow level



and the sequence of execution during the computation process.

Each module is a Java object, which has a standard structure and consists of several sections. For example: each computation module (C) consists of four sections. The first section organizes the preparation of input data. The second generates the job and controls its execution. The third initializes and controls the record of the result in the experiment data base. The fourth section controls the execution of module operation. It also informs the main program of the block about the manipulation of certain sets of data and when execution within a block is complete.

After a block is started, the parameterization module (P) and replacement module (R) wait for the request from the corresponding inputs of the computation module (C). After that, they generate a set of input data according to rules specified by the user, either as mathematical formulae or a list of parameter values. In this example three variants of parameterization are represented:

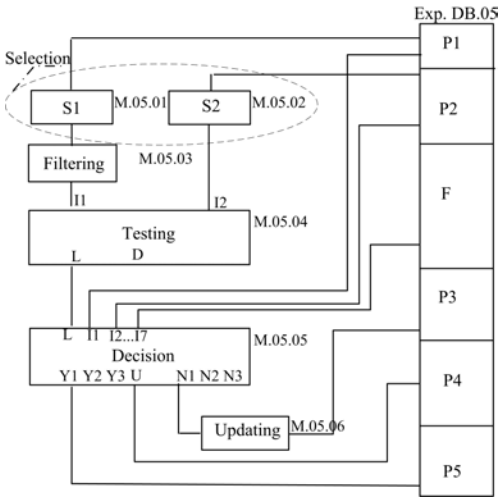
(a) Direct transmission of the parameter values with the job. In this case, parameterization module (P3) transfers the generated parameter value to the computation module (C) upon its request. The computation module generates the job, including converting parameter values into corresponding job parameters. This method can be used if the parameterized value is a number, symbol or combination of both.

(b) Parameterized objects are large arrays of information (DB.02-P4 in Figure 3) which are kept in the experiment database. These parameters are copied directly from the experiment database to the corresponding file server and then written with the same array name with the index of the number of the stage. In this case, attributes of the job are sent to the file server as references (an array of data).

(c) If it is important, then the preparation of the data is moved outside of the main program. This allows the creation of a more universal computation module. Furthermore, it allows scaling, i.e. avoiding limitations in the size, position, type and number of the parameterized objects used in a module. In these cases the replacement module is used. During the preparation of the next set of input data, new parameter values P1 und P2 are generated. The generated parameter set is linked with replacement processes and then delivered to the corresponding fileserver, where the replacement process is executed.

A typical control block program carries out an iterative analysis of the data sets from previous steps of the experiment program and selects either the direction for the further development of the experiment or examines whether the input data sets are ready for further computation, and subsequently synchronizes their further processing. An example of such a program can be seen in Figure 4. The condition

Figure 4. Example of a control block program (condition block). Data flow level



block shown in this figure consists of two selection modules, a filtering, testing, decision and updating module, as well as a database. A testing module is a functional program which analyses input data sets on the basis of certain criteria. A decision module redirects data for further com-

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

putation depending on the results obtained by the testing module. The selection and update modules are for the interaction between modules (computation, testing, etc.) and the database.

As we can see, the main elements are modules. In principle all simple control blocks can be mono-programs, which consist of units rather than modules, as for example the synchronization block in Figure 2.

Using a small number of modules, it is possible to create a large number of different program blocks for experiments in different fields of science and engineering.

For the creation of a block program it is necessary to indicate all nodes (modules) of computation data and to show the data transfer between modules. Data transfer is indicated with the help of connection lines which connect inputs and outputs of modules. Interaction between modules of a block is based on the request and reply principle. The block contains a module which takes the role of request module (in our cases this is computation module M.02.05 in Figure 3 and decision module M.05.05 in Figure 4). The other modules of the block make replies. Requests go from bottom to top (i.e. from the generating/request module to the reply module), and replies go from top to bottom (i.e. from the reply module to the request module). In each reply module processing continues until the list of parameter inputs is exhausted.

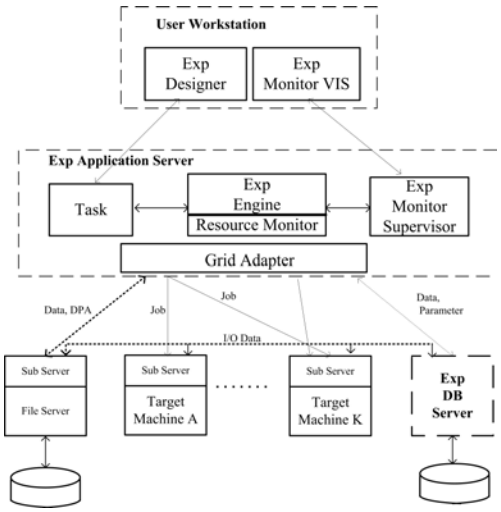
4. Science Experimental Grid Laboratory (SEGL) architecture

SEGL is a problem solving environment enabling the automated creation, start and monitoring of complex experiments and supports its effective execution on the Grid. SEGL is the environment within which the GriCoL language is utilized.

Figure 5 shows the system architecture of SEGL. It consists of three main components: the User Workstation (Client), the ExpApplicationServer (Server) and the ExpDBServer (OODB). The system operates according to a Client-Server-Model in which the ExpApplicationServer interacts with remote target computers using a Grid Middleware Service such as UNICORE [7] and SSH. The implementation is based on the Java 2 Platform Enterprise Edition (J2EE) specification and JBOSS Application Server. The database used is an Object Oriented Database (OODB) with a library tailored to the application domain of the experiment.

The two key parts of SEGL are: Experiment Designer (ExpDesigner), used for the design of experiments by working with elements of GriCoL, and the runtime system (ExpEngine). From the user's perspective, complex experiment scenarios are realized in Experiment Designer using GriCoL to represent the experiment. The technical mapping from this user perspective to the underlying infrastructure is carried out via the use of control flow and data flow. The runtime System of SEGL (ExpEngine) chooses the necessary computer resources, organizes and controls the sequence of execution according to the task flow and the condition of the experiment program. The user sets up initial conditions and parameters using GriCoL. Then SEGL monitors and steers the experiment, informing the user of the current status. SEGL employs a dynamic parameterization capability which requires an iterative, self-steering approach. The communication between the Experiment Designer and the runtime system supports the monitoring of execution of the experiment.

Figure 5. Architecture of SEGL



5. Use Case: Molecular Dynamics simulation of proteins

Molecular dynamics (MD) simulation is one of the principal tools in the theoretical study of biological molecules. This computational method calculates the time dependent behaviour of a molecular system. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids. These methods are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their interaction with substrates, ligands, and inhibitors.

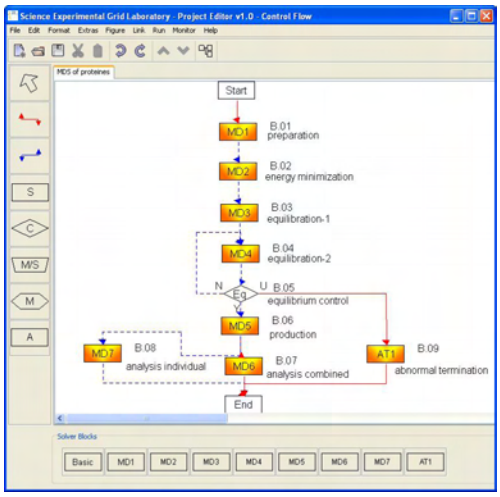
A common task for a computational biologist is to investigate the determinants of substrate specificity of an en-

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

zyme. On one hand, the same naturally occurring enzyme converts some substrates better than others. One the other hand, often mutations are found, in nature or by laboratory experiments, which change the substrate specificity, sometimes in a dramatic way. To understand these effects, multiple MD simulations are performed consisting of different enzyme-substrate combinations. The ultimate goal is to establish a general, generic molecular model that describes the substrate specificity of an enzyme and predicts short- and long-range effects of mutations on structure, dynamics, and biochemical properties of the protein.

While most projects on MD simulation are still managed by hand, large-scale MD simulation studies may involve up to thousands of MD simulations. Each simulation will typically produce a trajectory output file with a size of several gigabytes, so that data storage, management and analysis become a serious challenge. These tasks can no longer be performed interactively and therefore have to be automated. Therefore it is worthwhile to use an experiment management system that provides a language (GriCoL) that is able to describe all the necessary functionalities to design complex MD parameter studies. The experiment management system must be combined with the control of job execution in a distributed computing environment. Figure 6 shows

Figure 6. Control flow



the schematic setup of a large-scale MD simulation study. Starting from user provided structures of the enzyme, enzyme variants (a total of 30) and substrates (a total of 10), in the first step the preparation solver block (B.01) is used to generate all possible enzyme substrate combinations (a total of 300). This is accomplished by using the select module in the data flow of the preparation solver block, which builds the Cartesian product of all enzyme variants and substrates. Afterwards for all combinations the molecular system topology is built. These topologies describe the system under investigation for the MD simulation program. For better statistical analysis and sampling of the proteins' conformational space, each system must be simulated 10 times, using a different staring condition each time. Here the selection module of the GriCoL language is used in the data flow of the preparation solver block to generate automatically for every system ten new systems with a different starting condition, resulting in 3000 topologies. All 3000 topology files are stored in the experiment database and are used in all subsequent simulations. That means, all following MD solver blocks run their tasks for each of the 3000 topologies. The dashed connection lines in figure 6 (e.g. between B.01 and B.02) indicate that as soon as a particular simulation task in B.01 has finished, it can be passed on to the next bock B.02. A solid line between two blocks (e.g. between B.06 and B.07) means that all tasks have to be finished in the preceding block before the control flow proceeds to the next block.

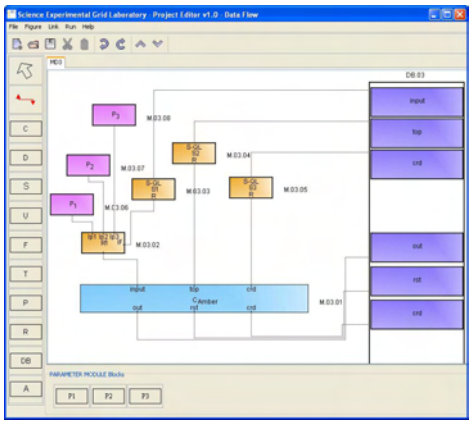
Starting from the 3000 topologies generated in the preparation solver block (B.01), the control flow is as follows: The energy minimization solver block (B.02) is used to remove possible initial bad interactions between atoms. The equilibration solver block (B.03) now starts an equilibration run for each of the 3000 energy minimized systems to slowly adapt them to the simulation environment like temperature and pressure. Several short simulation runs are performed with changing parameters that control the behavior of the simulation, e.g. the restraining of atomic positions or the coupling to a temperature bath. Here, GriCoL's parameterization and replacement modules come into play.

The data flow of B.03 is shown in detail in figure 7. The computation module (C-Amber) for the simulation program needs the system topology, the coordinates of the system to start from and an input file. Whereas the system topology and the starting coordinates are taken out of the experiment database via the selection module (S2 and S3), the input files are created for each simulation run individually. Therefore three parameterization modules (P1 - P3) provide the values for the replacement module (R1) that puts these values into an input file skeleton taken from the experiment database (via selection module S1). The resulting output from the simulation is put back into the experiment database.

In this use case, three subsequent simulation steps with changing parameters are required in the equilibration solver block B.03. This is followed by, at first, one more equilibration run (B.04) that has no changes in simulation parameters. During the equilibration run, which usually needs

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

Figure 7. Data flow of Block B.03



days to weeks, strongly depending on the numbers of CPUs available and the size of the system, the system should reach equilibrium. 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 multiple system properties at frequent intervals. The equilibrium control block (B.05) is used for this. Once the conditions for equilibrium are met, the equilibration phase for this system will be terminated and the production solver block (B.06) is started for this particular system. Systems that have not reached equilibrium yet are subjected to another round of the equilibration solver block (B.04). During the production solver block (B.06), which performs a MD simulation with a predefined amount of simulation steps, equilibrium properties of the system are assembled. Afterwards the trajectories from the production run are subjected to different analysis tools. While some analysis tools are run for each individual trajectory (B.08), some tools need all trajectories for their analysis (B.07).

The benefits of using an experiment management system like this are obvious. Beside the time saved for setting up, submitting, and monitoring the thousands of jobs, the base for common errors like misspelling in input files is also minimized. Beneficial is also the reusability of the computation modules: once defined for a specific program, it can be reused for other control flows and can be shared among colleagues requiring that program. The equilibration control helps to minimize simulation overhead as simulations which have already reached equilibrium are submitted to the production phase while others that have not are simulated further. The storage of simulation results in the experiment database enables the scientist to later retrieve and compare the results easily.

6. Conclusion

This paper presented the concept and description of GriCoL, a language for describing complex tasks, such as parametric investigations of scientific experiments, using the resources of a Grid. GriCoL provides a user-friendly, efficient, extensible, multilayered language capable of using the possibilities of high performance computer applications and building more complex and dynamic constructions out of these.

References

[1] Foster, I., Kesselman, C.: The Grid: Blueprint for a Future Computing Infrastructure, Morgan Kaufmann Publishers, USA, 1999.

[2] Currle-Linde, N., Küster, U., Resch, M., Risio, B.: Science Experimental Grid Laboratory (SEGL) Dynamical Parameter Study in Distributed Systems, ParCo 2005, Malaga, Spain, 2005.

[3] Yu, J., Buyya, R.: A Taxonomy of Workflow Management Systems for Grid Computing, Journal of Grid Computing, Volume 3, Numbers 3-4, pp. 171-200, September 2005.

[4] Tony, A., Curbera, F., Dholakia, H., Goland, Y., Klein, J., Leymann, F., Liu, K., Roller, D., Smith, D., Thatte, S., Trickovic, I., Weerawarana, S.: Specification: Business Process Execution Language for Web Services Version 1.1, May 05, 2003: http://www.106.ibm.com/developerworks/library/ws-bpel.

[5] Lehtovuori, P., Nyrönen, T.: SOMA - Workflow for Small Molecule Property Calculations on a Multiplatform Computing Grid,Computing Grid, J. Chem. Inf. Model., 2006, 46(2).

[6] de Vivo, A., Yarrow, M. McCann, K.: Acomparison of parameter study creation and job submossion tools; Technical report NAS-01002, NASA Ames Research Center, Moffet Filed, CA,2000.

[7] Erwin, D. (Ed.): Joint Project Report for the BMBF Project UNICORE Plus, Grant Number: 01 IR 001 A-D, Duration: January 2000 - December 2002.

Proceedings of the Second IEEE International
Conference on e-Science and Grid Computing (e-Science'06)
0-7695-2734-5/06 $20.00 © 2006 IEEE

Time-Based Haptic Analysis of Protein Dynamics

Katrin Bidmon Guido Reina
Visualization and Interactive Systems Group (VIS)
University of Stuttgart, Germany
{Katrin.Bidmon | Guido.Reina}@vis.uni-stuttgart.de

Fabian Bös Jürgen Pleiss
Institute of Technical Biochemistry (ITB)
University of Stuttgart, Germany
{Fabian.Boes | Juergen.Pleiss}@itb.uni-stuttgart.de

Thomas Ertl
Visualization and Interactive Systems Group (VIS)
University of Stuttgart, Germany
Thomas.Ertl@vis.uni-stuttgart.de

Abstract

We present a new approach for evaluating a protein's simulated trajectory via time-based haptic feedback. Molecular-scale trajectories are highly dynamic and thus pose new demands on haptic interaction. Traditionally, the flexibility of each atom within the protein is represented by a scalar, not taking into account the direction of movement. Using our haptic approach for tracking the atoms gives the user this additional information and a more intuitive feedback than numbers or even plain graphics. We combine this with touchable protein and atom surfaces and data-driven surface properties as well as stereo representation. Having conceived this tool in close collaboration between experts from the field of biochemistry and computer scientists ensures the appropriateness of our approach.

1. Introduction

Proteins are macromolecules with highly complex structure and dynamics. It is now evident that protein flexibility and correlated motions are essential for protein function. Therefore, considerable interest in studying protein dynamics in relation to protein function [1] has arisen. Of all possible shapes of a molecule the most interesting ones are those which are local minima in terms of potential energy. These local minima are called conformations. Hence, conformational behaviour of a molecule is the process of transitions between individual conformations of the molecule around the energy minimum.

Molecular dynamics (MD) simulation is one of the principal tools in the theoretical study of biological molecules. Based on Newton's equations of motion, this computational method calculates the time-dependent behavior of a molecular system. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins. This method is now routinely used to investigate the structure, dynamics and thermodynamics of proteins and their interaction with substrates, ligands, and inhibitors.

The flexibility of a protein can be derived from the trajectory of a MD simulation by calculating the isotropic root mean square fluctuation (RMSF) of individual atoms after removing translational and rotational movements. Experimental data of protein flexibility is available through the B-factors obtained by X-ray crystallography. The B-factor is a measurement of disorder in the protein crystal; a high B-factor means that a certain atom is highly mobile. The RMSF obtained from MD simulations can be directly related to the experimentally determined B-factors. Typically, the flexibility of atoms at the protein surface and thus in contact with the surrounding solvent is higher than the flexibility of those buried in the densely packed core of the protein.

When examining this flexibility, one must take into consideration that protein dynamics take place on different time scales: fast uncorrelated motions of single atoms or amino acids are distinct from slow correlated motions between different parts of the protein or protein domains. Both types of anisotropic motions must be analyzed and characterized in order to understand the impact of protein flexibility on protein function. Our approach aims to support the user in this task. The macroscopic motion of the protein as a whole is subtracted from the simulation to refer to the protein's local coordinate system.

Visualization of protein flexibility data from MD simulations usually involves plotting the flexibility data against the atom number (Fig. 2), static pictures of the protein structure with the degree of flexibility color–coded and mapped

Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)
0-7695-2738-8/07 $20.00 © 2007 IEEE

onto the individual atoms (Fig. 3), or representing the global correlated motions as arrows in a static picture. A direct time-dependent analysis of the flexibility other than looking at individual atoms or group of atoms in 3D space has not been possible up to now.

Therefore, we present a method for haptic interaction with the protein to give extended insight into the anisotropic motion of the protein. We chose to couple a PHANTOM [23] haptic device to VMD [9], an open source tool widely used for molecular visualization, also capable of handling molecular dynamics trajectories. To do so, we used the SensAble OpenHaptics toolkit [10] whose HLAPI allows the programmer to directly set a virtual proxy position, as well as surface constraints, as defined in [24].

So, the user can touch the protein in various representations together with different surface characteristics and select the atom for flexibility exploration. Locking the instrument tip to the selected atom, the user's hand is moved in accordance with the atom's conformational path. The user gets direct haptic feedback and thus a better insight into the anisotropic motion.

The paper is structured as follows: in Sec. 2 we summarize related work, Sec. 3 describes the data we use as a base of our work. Sec. 4 details the haptic interaction methods with the protein and atom surfaces detailed in Sec. 4.1 and the time-based haptics in Sec. 4.2. The results are given in Sec. 5 and we conclude our paper along with some comments on future work in Sec. 6.

2. Related Work

Beside VMD, there are various tools for protein visualization such as PyMol [7], Amira [2], MoilView [27], MD Display [4], and Chimera [22], but VMD is widely used and fits our requisites best. All of these tools offer a wide range of representations, some of them structure-based, others based on the calculated surface of the proteins. For some representations, e.g. the solvent accessible surface, it is sensible to implement the option for exploring the protein surface haptically. Hence, there are several publications about haptic interaction with such surfaces, as detailed below. In principle, VMD does support haptic devices, but neither collision detection between cursor and molecule nor intermolecule collision detection without resorting to a simulation backend.

A prototype Molecular Visualiser (MV) application, based on Web3D standards and extended by support of haptic interaction, such as frictional surfaces, was developed by Davies et al. [6]. To combine haptics with 3D molecular visualization, a Reachin Display was used. This tool is made for teaching and research alike, but might suffer from performance bottlenecks for large molecules, due to the use of relatively inefficient data formats and platforms.

Sankaranarayanan et al. [25] created multi-modality enhancements of tangible models by superimposing graphical information on top of existing physical models and by providing haptic feedback for the superimposed structures using a PHANTOM device. In our application scenario, however, no physical models are available, so we did not look into employing mixed-reality techniques.

In [19], Mark et al. integrated force feedback into a real-time virtual environment. Particular problems such as decoupling of the haptic servo loop from the main application loop for high-quality forces, benefits including haptic-textured surfaces, and device independence are addressed. Most of these issues are meanwhile solved by the OpenHaptics toolkit.

A lot of work in the field of haptics was done by Křenek et al.: in [15] they focus on mapping quantities of chemical meaning to force fields. The special case of haptically mapping the required energy for switching between different conformations was studied in [14]. In contrast to that, we do not interact with the simulation itself, but only observe the complex system of a protein in solvent. A closely related work was [11] where the haptic rendering of molecular flexibility was investigated, but bond flexibility was used in the model instead of real pre-computed conformational paths. Moreover, binding the haptics frequency to that of graphics limited feedback by computational power. In ensuing work [12, 21] improvements on conformational paths and offline calculation were made to improve the model's interactivity. But as in the earlier work, not the properties of a special trajectory were focused as in our work, but the switching between conformations.

Lundin et al. [17, 16] presented a proxy-based approach to volume haptics instead of the iso-value based approaches. So they were able to add properties like friction and stiffness to the material, similarly to our approach. Likewise, volume haptics is the subject of [18], where docking two proteins was the focus and haptic rendering was combined with stereoscopic vision–with only moderate success. In contrast to most of these papers, our approach aims at highlighting the dynamic properties of individual atoms in a single protein instead of investigating the interaction between several macromolecules.

3. Data

TEM β-lactamases are enzymes that hydrolyze lactam antibiotics and thus transfer antibiotic resistance to pathogenic organisms. They are plasmid-based, monomeric enzymes with a length of 263 amino acids and a molecular weight of approximately 28 kDa. Figure 1 shows the protein in cartoon representation. TEM β-lactamases and their different phenotypes as well as their role in in antibiotic resistance has been reviewed in detail in [3].

Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)
0-7695-2738-8/07 $20.00 © 2007 IEEE

Figure 1. Crystal structure of TEM-1 β-lactamase in cartoon representation. α-helices are colored in red, β-sheets in yellow and interconnecting loop regions in green.

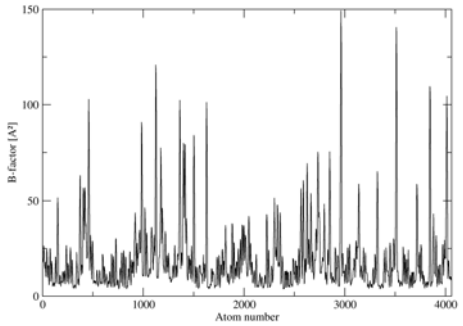


Figure 2. B-factors per atom of TEM-1 β-lactamase.

To understand the function of this protein on a molecular basis and investigate the impact of protein flexibility thereon, we have performed a 3 ns MD simulation of TEM-1 β-lactamase.

A highly resolved structure (resolution 0.85 Å) [20] was used as a starting point. The protein was immersed in a 10 Å layer truncated octahedron periodic water box of TIP3P waters. The system was subjected to 50 steps of energy minimization by steepest descent to remove initial close van der Waals contacts, followed by 150 steps of energy minimization by conjugate gradient. After energy minimization, simulation was performed with periodic boundary conditions at 300 K. The SHAKE algorithm was applied to all bonds containing hydrogen atoms, and a time step of 1 fs was used. The Berendsen method was used to couple the systems to constant temperature and pressure. The electrostatic interactions were evaluated by the particle-mesh Ewald method. The Lennard-Jones interactions were evaluated using a 8.0 Å cutoff. For the first 60 ps of the MD simulation the backbone atoms were restrained using a harmonic potential with a gradually decreasing force constant from 10 to 1 kcal/mol. The structure was further equilibrated by non-restrained MD simulation for 2.0 ns followed by a production run of 1.0 ns, in which snapshots were saved for later analysis every 250 fs. Simulation and data analysis were performed using the AMBER7 software package [5].

The resulting trajectory was stable and the backbone atoms deviated by only 0.9 – 1.1 Å from the experimental X-ray structure. To assess the protein flexibility, the root mean square fluctuations (in terms of B-factors) for all protein atoms were calculated. All coordinate frames from last nanosecond of the trajectory were superimposed on the average conformation to remove overall translation and rotation. In Fig. 2 the per-atom B-factors derived from the MD simulation are plotted. From the plot it becomes evident that there are regions in the protein that are more flexible than others. When mapped onto the protein structure, it can be seen that regions of low flexibility are mostly located in secondary structure elements and in the core of the protein. Opposite to that, loops and regions at the protein surface show a much higher flexibility. Figure 3 shows only a part of the protein and focusses on the active site and the adjacent Ω-loop. The protein is shown in surface representation and the B-factors per atom are color-coded. In the active site, where the chemical reaction takes place, the atoms are less flexible than on the Ω-loop, which partly shields the active site towards the solvent.

4. Haptics

We have extended VMD for our prototypical implementation. VMD already supports haptic devices, however the forces generated by the user are used as input for a simulation back-end (NAMD), enabling the user to influence the simulation parameters. Feedback is the result of forces generated by the simulation influencing the selection of atoms the user is grabbing. No collision feedback is given with the standard tools. A powerful cluster or SMP machine is needed to ensure that the simulation can be updated in real time. We wanted the user to benefit from haptic feedback even when the visualization is decoupled from the simulation, and trajectory files are loaded off-line. Additionally the user should be able to feel the surface of the different graphical representations of a molecule. Therefore we modified the rendering subsystem of VMD to interface

Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)
0-7695-2738-8/07 $20.00 © 2007 IEEE

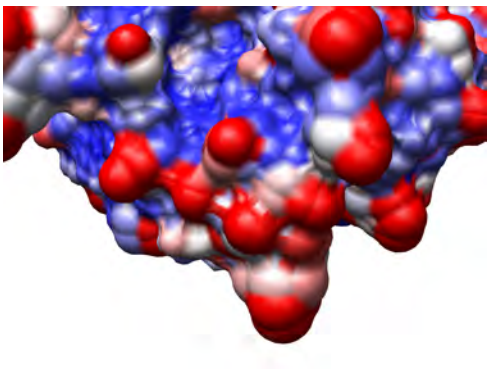


Figure 3. Active site and Ω-loop of TEM-1 β-lactamase. B-factors per atom are color-coded and mapped onto the molecular surface. Blue = low B-factor, Red = high B-factor.

with the SenSable OpenHaptics toolkit directly. To locate the tip of the haptic device, we added a virtual representation – a small sphere – of the PHANTOM position to the visualization. Since it has been shown that haptic interaction is much more effective when the visualization is stereo-enabled [8], we make use of stereo rendering. Fortunately, VMD does support stereo rendering out of the box, however we had to implement a stencil stereo mode which is column-interleaved to be able to use our SeeReal autostereoscopic LCD [26].

4.1. Surface Interaction

For the navigation of the 3D space around a molecule and the selection of specific atoms it is very helpful that the user feels the molecular surface. Even with depth perception through stereo visualization, the actual distance from the cursor to the molecule surface is difficult to estimate precisely. However, if the user gets additional feedback from touch events, it makes the localization of the 3D cursor in relation to the atoms much easier and adds a certain consistency between the visualization and the user's input, as tactile feedback is known from daily life.

To give the user access to additional properties of a protein, we experimented with different surface properties as can be defined with the OpenHaptics toolkit. When setting the friction based on the type of the atom that contributed to the generation of a certain triangle, the user can, for example, visualize another attribute on the surface and still find out where the atom boundaries on the surface are because of the perceived roughness changes. The charge of amino acids can also be mapped for a defined charge of the cursor: if the charges are inverted, the surface is sticky (high friction), while a collision event with the same polarity will result in a force applied along the surface normal to repel the cursor.

4.2. Time-based Feedback

The main contribution of our work is the time-dependent force feedback caused by the motion of selected atoms in the displayed molecule. Using the button on the PHANTOM stylus, the user can attach himself to the position of a chosen atom, which means that for static datasets the PHANTOM will just be locked in place. When trajectory playback is activated, the user's hand will be dragged around in space according to the motion the selected atom performs over time. Even at the highest playback speed, this motion is still a bit discontinuous, as the different discrete positions are just handed over to the PHANTOM device during the redraw of the scene. To achieve a more pleasant experience, the haptic rendering needs to be much faster than the visualization [13], so we added interpolated position reporting at a fixed rate, regardless of the animation update speed. In practice 100Hz seemed adequate, as no discrete positional jumps could be felt in the datasets we have tested. We are currently using linear interpolation, but higher order interpolation could be easily implemented as well. A drawback of this smoother movement – which also made us avoid higher order interpolation for now – is that even drastic directional changes are not perceived very clearly anymore. To overcome this limitation, we augmented the feedback magnitude by scaling the movement by a fixed factor (currently 4.0). This scaling is calculated relatively to the position p_g of an atom when it is first grabbed, that means that all movement of this atom occurs in a subspace surrounding this position and is perceived relatively to it. To emphasize that movement, we applied the scaling also relatively to the position p_g, effectively scaling the traversed subspace. The applied scaling is unaffected by zooming, as it is calculated in the local molecule's coordinate system. The user can feel no discrepancy in distances since it is very difficult to mentally map the motion on screen absolutely to the motion in the real world, however the perception of small and smooth movements is facilitated.

5. Results

Integration of the PHANTOM haptic device into the molecular viewer VMD allows the user to accomplish two important kinds of analysis tasks in the field of molecular modelling / simulation.

First, mapping molecular properties onto the surface and coupling them to haptic feedback makes them directly fee-

Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)
0-7695-2738-8/07 $20.00 © 2007 IEEE

lable for the user. It is possible to intuitively explore the active site and get a sensoric impression of e.g. the atomic charge distribution.

Second, tracking of motion allows the analysis of anisotropic protein dynamics at different levels, something that is not covered in standard analysis of MD trajectories. The usage of single-atom tracking with the PHANTOM is very intuitive. The user can directly feel motion in all three dimensions without perceptual issues as there would be for stereoscopic vision. Haptic feedback also does not suffer from occlusion or cluttering like visualization. The visualization would need selection-sensitive transparency to make the tracking of the focus (selection) in the context (protein) easier, while we can just use the visualization as context while the user feels what the focus is doing.

It is quite simple to distinguish areas of high motion from those with low activity when grabbing the atoms. The difference between the relatively rigid atoms of the active site and the more flexible atoms that constitute to the Ω-loop becomes evident to the user. However, the perception of the absolute quantity of motion is difficult. It is also unclear yet how many different movement magnitudes the user can absolutely tell apart.

To filter out high-frequency movements and just pass the large changes to the user, we can make use of the smoothing option of VMD which also directly affects our haptic rendering. Using trajectory smoothing therefore allows the user to filter out the fast uncorrelated motions of individual atoms and amino acids and directly focus on slow correlated motions between different parts of the protein or protein domains.

6. Conclusion and Future Work

We have presented a method for time-based haptic rendering of molecular dynamics using a wide-spread application and commodity hardware that allows the user to compare the flexibility of various parts of a protein. We combined this with haptic rendering of protein attributes as surface properties to allow for better insight through a higher dimensionality in output. This may involve molecular attributes like atomic charges or hydrophobicity or any other property that can be calculated and mapped onto a per-atom or per-residue basis.

The integration of haptic devices into molecular dynamics visualization tools is by far not limited to time-dependent analysis of protein flexibility. A future development may involve the usage of the haptic device as a manual molecular docking tool combined with force feedback based on the charges of the involved molecules. By changing the size and shape of the haptic cursor into that of a small chemical ligand, it is possible to directly test whether a certain molecule fits into the active site of a protein. By incorporating the calculation of protein-ligand interaction energies it is further possible to render attractive or repulsive forces with the haptic device between the protein and compound.

References

[1] P. Agarwal, A. Geist, and A. Gorin. Protein dynamics and enzymatic catalysis: Investigating the peptidyl-prolyl cis-trans isomerization activity of cyclophilin a. *Biochemistry*, 43(33):10605–10618, 2004.

[2] Amira – advanced 3d visualization and volume modeling. http://www.amiravis.com/.

[3] P. Bradford. Extended-spectrum β-lactamases in the 21st century: Characterization, epidemiology, and detection of this important resistance threat. *Clinical Microbiology Reviews*, 14(4):933–951, 2001.

[4] T. Callahan, E. Swanson, and T. Lybrand. Md display: An interactive graphics program for visualization of molecular dynamics trajectories. *Journal of Molecular Graphics*, 14:39–41, 1996.

[5] D. Case, D. Pearlman, J. Caldwell, T. Cheatham III, J. Wang, W. Ross, C. Simmerling, T. Darden, K. Merz, R. Stanton, A. Cheng, J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R. Radmer, Y. Duan, J. Pitera, I. Massova, G. Seibel, U. Singh, P. Weiner, and P. Kollman. Amber 7. University of California, San Francisco, 2002.

[6] R. Davies, N. John, J. MacDonald, and K. Hughes. Visualization of molecular quantum dynamics: a molecular visualization tool with integrated web3d and haptics. In *Proceedings of Web3D '05*, pages 143–150, 2005.

[7] W. DeLano. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, 40, 2002. http://pymol.sourceforge.net/.

[8] M. Fujimoto and Y. Ishibashi. The effect of stereoscopic viewing of a virtual space on a networked game using haptic media. In *Proceedings of ACE '04*, pages 317–320, 2004.

[9] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[10] B. Itkowitz, J. Handley, and W. Zhu. The openhaptics toolkit: A library for adding 3d touch navigation and haptics to graphics applications. In *Proceedings of WHC '05*, pages 590–591, Washington, DC, USA, 2005. IEEE Computer Society.

[11] A. Křenek. Haptic rendering of molecular flexibility. In *Proceedings of PURS2000*, pages 19–26, 2000.

[12] A. Křenek. Haptic rendering of molecular conformations. In *Proceedings of Eurohaptics 2001*, pages 142–145, 2001.

[13] A. Křenek. Haptic rendering of complex force fields. In *EGVE '03: Proceedings of the workshop on Virtual environments 2003*, pages 231–239, 2003.

[14] A. Křenek and Z. Kabeláč. Studying conformational behaviour with phantom device(s). In *Proceedings of PURS'99*, pages 3–6, 1999.

[15] A. Křenek, M. Černohorský, and Z. Kabeláč. Haptic visualization of molecular data. In V. Skala, editor, *WSCG'99 Conference Proceedings*, 1999.

Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)
0-7695-2738-8/07 $20.00 © 2007 IEEE

[16] K. Lundin, B. G., and A. Ynnerman. General proxy-based haptics for volume visualization. In *Proceedings of WHC '05*, pages 557–560, Washington, DC, USA, 2005. IEEE Computer Society.

[17] K. Lundin, A. Ynnerman, and B. Gudmundsson. Proxy-based haptic feedback from volumetric density data. In *Proceedings of the Eurohaptic Conference*, pages 104–109, 2002.

[18] R. Maciejewski, S. Choi, D. Ebert, and H. Tan. Multi-modal perceptualization of volumetric data and its application to molecular docking. In *Proceedings of WHC '05*, pages 511–514, Washington, DC, USA, 2005. IEEE Computer Society.

[19] W. Mark, S. Randolph, M. Finch, J. Van Verth, and I. Taylor, R.M. Adding force feedback to graphics systems: issues and solutions. In *Proceedings of SIGGRAPH '96*, pages 447–452, 1996.

[20] G. Minasov, X. Wang, and B. Shoichet. An ultrahigh resolution structure of tem-1 beta-lactamase suggests a role for glu166 as the general base in acylation. *Journal of the American Chemical Society*, 124(19):5333–5340, 2002.

[21] I. Peterlik and A. Křenek. Haptically driven travelling through conformational space. In *Proceedings of WHC '05*, pages 342–347, 2005.

[22] E. Pettersen, T. Goddard, C. Huang, G. Couch, D. Greenblatt, E. Meng, and T. Ferrin. Ucsf chimera - a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.

[23] The PHANTOM Desktop device home page. http://www.sensable.com/haptic-phantom-desktop.htm.

[24] D. C. Ruspini, K. Kolarov, and O. Khatib. The haptic display of complex graphical environments. In *Computer Graphics (SIGGRAPH 97 Conference Proceedings)*, pages 345–352, 1997.

[25] G. Sankaranarayanan, S. Weghorst, M. Sanner, A. Gillet, and A. Olson. Role of haptics in teaching structural molecular biology. In *Proceedings of HAPTICS'03*, page 363, 2003.

[26] (: SeeReal Technologies. http://www.seereal.com/.

[27] C. Simmerling, R. Elber, and J. Zhang. Moil-view - a program for visualization of structure and dynamics of biomolecules and sto- a program for computing stochastic paths. *Modelling of Biomolecular Structure and Mechanisms*, pages 241–265, 1995.

Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)
0-7695-2738-8/07 $20.00 © 2007 IEEE

Gesamtliteraturverzeichnis

Abraham, E. P., Chain, E., An enzyme from bacteria able to destroy penicillin. Nature 146 (4), 837–838, 1940.

Abramson, D., Giddy, J., Kotler, L., High performance parametric modeling with Nimrod/G: Killer application for the global grid? In: IPDPS '00: Proceedings of the 14th International Symposium on Parallel and Distributed Processing, 520–528, 2000a.

Abramson, D., Lewis, A., Peachy, T., Nimrod/O: A tool for automatic design optimization. In: The 4th International Conference on Algorithms & Architectures for Parallel Processing (ICA3PP 2000), 2000b.

Adachi, H., Ohta, T., Matsuzawa, H., Site-directed mutants, at position 166, of RTEM-1 beta-lactamase that form a stable acyl-enzyme intermediate with penicillin. J Biol Chem 266 (5), 3186–3191, 1991.

Agarwal, P., Geist, A., Gorin, A., Protein dynamics and enzymatic catalysis: Investigating the peptidyl-prolyl cis-trans isomerization activity of cyclophilin A. Biochemistry 43 (33), 10605–10618, 2004.

Amadei, A., Linssen, A. B., Berendsen, H. J., Essential dynamics of proteins. Proteins 17 (4), 412–425, 1993.

Ambler, R. P., The structure of beta-lactamases. Philos Trans R Soc Lond B Biol Sci 289 (1036), 321–331, 1980.

Ambler, R. P., Coulson, A. F., Frère, J. M., Ghuysen, J. M., Joris, B., Forsman, M.,

Levesque, R. C., Tiraby, G., Waley, S. G., A standard numbering scheme for the class A beta-lactamases. Biochem J 276 (Pt 1), 269–270, 1991.

Babic, M., Hujer, A. M., Bonomo, R. A., What's new in antibiotic resistance? Focus on beta-lactamases. Drug Resist Updat 9 (3), 142–156, 2006.

Bambeke, F. V., Balzi, E., Tulkens, P. M., Antibiotic efflux pumps. Biochem Pharmacol 60 (4), 457–470, 2000.

Banko, G., Demain, A., Wolfe, S., δ-(l-α-aminoadipyl)-L-cysteinyl-D-valine synthetase. A multifunctional enzyme with broad substrate specificity for the synthesis of penicillin and cephalosporin precursors. J Am Chem Soc 109, 2858–2860, 1987.

Batchelor, F., Doyle, F., Nayler, J., Rolinson, G., Synthesis of penicillin: 6-aminopenicillanic acid in penicillin fermentations. Nature 183 (4656), 257–258, 1959.

Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., Haak, J. R., Molecular-dynamics with coupling to an external bath. J Chem Phys 81 (8), 3684–3690, 1984.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E., The Protein Data Bank. Nucleic Acids Res 28 (1), 235–242, 2000.

Bonner, D. P., Sykes, R. B., Structure activity relationships among the monobactams. J Antimicrob Chemother 14 (4), 313–327, 1984.

Bottoms, C. A., Smith, P. E., Tanner, J. J., A structurally conserved water molecule in Rossmann dinucleotide-binding domains. Protein Sci 11 (9), 2125–2137, 2002.

Bottoms, C. A., White, T. A., Tanner, J. J., Exploring structurally conserved solvent sites in protein families. Proteins 64 (2), 404–421, 2006.

Bradford, P. A., Extended-spectrum beta-lactamases in the 21st century: characteriza-

tion, epidemiology, and detection of this important resistance threat. Clin Microbiol Rev 14 (4), 933–951, 2001.

Bryan, L., Godfrey, A., 1991. Antibiotics in Laboratory Medicine, 3rd ed. Williams and Wilkins, Baltimore, Ch. Beta-lactam antibiotics, mode of action and bacterial resistance, 599–664.

Burq, S., Melnikoff, S. J., Branson, K., Buyya, R., Visual parameteric modeler for rapid composition of parameter-sweep applications for processing on global grids. In: International Conference on Computational Science, 739–749, 2003.

Bush, K., Jacoby, G., Nomenclature of TEM beta-lactamases. J Antimicrob Chemother 39 (1), 1–3, 1997.

Bush, K., Jacoby, G. A., Medeiros, A. A., A functional classification scheme for beta-lactamases and its correlation with molecular structure. Antimicrob Agents Chemother 39 (6), 1211–1233, 1995.

Buyya, R., Abramson, D., Giddy, J., Nimrod/g: an architecture for a resource management and scheduling system in a global computational grid. In: High Performance Computing in the Asia-Pacific Region, 2000. Proceedings. The Fourth International Conference/Exhibition on, Vol. 1, 283–289vol.1, 2000.

Buyya, R., Branson, K., Giddy, J., Abramson, D., The virtual laboratory: enabling molecular modeling for drug design on the world wide grid. Concurr Comput Pract Exper 15, 1–25, 2003.

Callahan, T., Swanson, E., Lybrand, T., MD Display: An interactive graphics program for visualization of molecular dynamics trajectories. J Mol Graph Model 14, 39–41, 1996.

Carlson, M. L., Regan, R. M., Gibson, Q. H., Distal cavity fluctuations in myoglobin: protein motion and ligand diffusion. Biochemistry 35 (4), 1125–1136, 1996.

Carugo, O., Bordo, D., How many water molecules can be detected by protein crystal-

lography? Acta Crystallogr D Biol Crystallogr 55, 479–483, 1999.

Casanova, H., Berman, F., Obertelli, G., Wolski, R., The AppLeS parameter sweep template: User-level middleware for the grid. In: Supercomputing, ACM/IEEE 2000 Conference, 60–60, 2000.

Case, D., Darden, T., Cheatham III, T., Simmerling, C., Wang, J., Duke, R., Luo, R., Merz, K., Wang, B., Pearlman, D., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J., Ross, W., Kollman, P., Amber 8. University of California, San Francisco.

Case, D., Pearlman, D., Caldwell, J., Cheatham III, T., Wang, J., Ross, W., Simmerling, C., Darden, T., Merz, K., Stanton, R., Cheng, A., Vincent, J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R., Duan, Y., Pitera, J., Massova, I., Seibel, G., Singh, U., Weiner, P., Kollman, P., Amber 7. University of California, San Francisco.

Caves, L. S., Evanseck, J. D., Karplus, M., Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. Protein Sci 7 (3), 649–666, 1998.

Chain, E., Florey, H. W., Gardner, A. D., Heatley, N. G., Jennings, M. A., Orr-Ewing, J., Sanders, A. G., Penicillin as a chemotherapeutic agent. Lancetii, 226–228, 1940.

Currle-Linde, N., Adamidis, P., Resch, M., Bös, F., Pleiss, J., Gricol: A language for scientific grids. In: e-Science and Grid Computing, 2006. e-Science '06. Second IEEE International Conference on, 62–62, 2006.

Currle-Linde, N., Bös, F., Lindner, P., Pleiss, J., Resch, M., A management system for complex parameter studies and experiments in grid computing. In: Parallel and Distributed Computing and Systems (PDCS), Cambridge, MA, USA, 34–39, 2004.

Currle-Linde, N., Risio, B., Küster, U., Resch, M., Science experimental grid laboratory (SEGL) dynamical parameter study in distributed systems. In: Parallel Computing: Current & Future Issues of High-End Computing, PARCO 2005, 2005.

Daggett, V., Long timescale simulations. Curr Opin Struct Biol 10 (2), 160–164, 2000.

Danziger, L. H., Pendland, S. L., Bacterial resistance to beta-lactam antibiotics. Am J Health Syst Pharm 52 (6 Suppl 2), S3–S8, 1995.

Darden, T., York, D., Pedersen, L., Particle mesh Ewald: An N · log(N) method for Ewald sums in large systems. J Chem Phys 98, 10089–10092, 1993.

Datta, N., Kontomichalou, P., Penicillinase synthesis controlled by infectious R factors in *Enterobacteriaceae*. Nature 208 (5007), 239–241, 1965.

Davies, J., Inactivation of antibiotics and the dissemination of resistance genes. Science 264 (5157), 375–382, 1994.

Davies, R., John, N., MacDonald, J., Hughes, K., Visualization of molecular quantum dynamics: a molecular visualization tool with integrated Web3D and haptics. In: Proceedings of Web3D '05, 143–150, 2005.

DeLano, W., Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography 40, 2002.

Demain, A., Biosynthesis of cephalosporin C and its relation to penicillin formation. Trans NY Acad Sci 25, 731–740, 1963.

Demain, A. L., Elander, R. P., The beta-lactam antibiotics: past, present, and future. Antonie Van Leeuwenhoek 75 (1-2), 5–19, 1999.

Despa, F., Biological water: Its vital role in macromolecular structure and function. Ann NY Acad Sci 1066 (1), 1–11, 2005.

DeVivo, A., Yarrow, M., McCann, K., A comparison of parameter study creation and job submission tools. Technical Report NAS-01-002, NASA Ames Research Center, Moffet Field, CA.

Diaz, N., Sordo, T. L., Jr, K. M. M., Suarez, D., Insights into the acylation mechanism of class A beta-lactamases from molecular dynamics simulations of the TEM-1 enzyme complexed with benzylpenicillin. J Am Chem Soc 125 (3), 672–684, 2003.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., Baker, N. A., PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res 32, W665–W667, 2004.

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G. M., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. M., Kollman, P., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24 (16), 1999–2012, 2003.

Elofsson, A., Nilsson, L., How consistent are molecular dynamics simulations? Comparing structure and dynamics in reduced and oxidized *Escherichia coli* thioredoxin. J Mol Biol 233 (4), 766–780, 1993.

Ernst, J. A., Clubb, R. T., Zhou, H. X., Gronenborn, A. M., Clore, G. M., Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. Science 267 (5205), 1813–1817, 1995.

Erwin, D., Joint project report for the BMBF project UNICORE Plus, grant number: 01 IR 001 A-D, Duration: January 2000 to December 2002.

Erwin, D. W., Unicore - a grid computing environment. Concurr Comput Pract Exper 14 (13-15), 1395–1410, 2002.

Essack, S. Y., The development of beta-lactam antibiotics in response to the evolution of beta-lactamases. Pharm Res 18 (10), 1391–1399, 2001.

Fisher, J. F., Meroueh, S. O., Mobashery, S., Bacterial resistance to beta-lactam antibiotics: Compelling opportunism, compelling opportunity. Chem Rev 105 (2), 395–424, 2005.

Fleming, A., On the antibacterial action of cultures of a *Penicillium*, with special reference to their use in the isolation of *B. influenzae*. British Journal of Experimental Pathology 10 (1), 226–236, 1929.

Florey, H. W., Antibiotic products of a versatile fungus. Ann Intern Med 43 (3), 480–490,

1955.

Foster, I., Kesselman, C., The globus project: A status report. In: HCW '98: Proceedings of the Seventh Heterogeneous Computing Workshop, IEEE Computer Society, Washington, DC, USA, 4–18, 1998.

Foster, I., Kesselman, C. (Eds.), 1999. The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann Publishers.

Fujimoto, M., Ishibashi, Y., The effect of stereoscopic viewing of a virtual space on a networked game using haptic media. In: Proceedings of ACE '04, 317–320, 2004.

Goffin, C., Ghuysen, J.-M., Biochemistry and comparative genomics of SxxK superfamily acyltransferases offer a clue to the mycobacterial paradox: presence of penicillin-susceptible target proteins versus lack of efficiency of penicillin as therapeutic agent. Microbiol Mol Biol Rev 66 (4), 702–38, table of contents, 2002.

Golemi-Kotra, D., Meroueh, S. O., Kim, C., Vakulenko, S. B., Bulychev, A., Stemmler, A. J., Stemmler, T. L., Mobashery, S., The importance of a critical protonation state and the fate of the catalytic steps in class a beta-lactamases and penicillin-binding proteins. J Biol Chem 279 (33), 34665–34673, 2004.

Gorfe, A. A., Ferrara, P., Caflisch, A., Marti, D. N., Bosshard, H. R., Jelesarov, I., Calculation of protein ionization equilibria with conformational sampling: pK(a) of a model leucine zipper, GCN4 and barnase. Proteins 46 (1), 41–60, 2002.

Grebe, T., Hakenbeck, R., Penicillin-binding proteins 2b and 2x of *Streptococcus pneumoniae* are primary resistance determinants for different classes of beta-lactam antibiotics. Antimicrob Agents Chemother 40 (4), 829–834, 1996.

Hall, B. G., Barlow, M., Evolution of the serine beta-lactamases: past, present and future. Drug Resist Updat 7 (2), 111–123, 2004.

Hartman, B. J., Tomasz, A., Low-affinity penicillin-binding protein associated with beta-lactam resistance in *Staphylococcus aureus*. J Bacteriol 158 (2), 513–516, 1984.

Hartman, B. J., Tomasz, A., Expression of methicillin resistance in heterogeneous strains of *Staphylococcus aureus*. Antimicrob Agents Chemother 29 (1), 85–92, 1986.

Hawkey, P. M., Resistance to carbapenems. J Med Microbiol 46 (6), 451–454, 1997.

Helms, V., Protein dynamics tightly connected to the dynamics of surrounding and internal water molecules. ChemPhysChem 8 (1), 23–33, 2007.

Hermann, J. C., Hensen, C., Ridder, L., Mulholland, A. J., Höltje, H.-D., Mechanisms of antibiotic resistance: QM/MM modeling of the acylation reaction of a class A beta-lactamase with benzylpenicillin. J Am Chem Soc 127 (12), 4454–4465, 2005.

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 beta-lactamases through QM/MM modeling. J Am Chem Soc 125 (32), 9590–9591, 2003.

Hess, Similarities between principal components of protein dynamics and random diffusion. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 62 (6 Pt B), 8438–8448, 2000.

Hess, B., Convergence of sampling in protein simulations. Phys Rev E Stat Nonlin Soft Matter Phys 65 (3 Pt 1), 031910, 2002.

Huang, W. Z., Petrosino, J., Hirsch, M., Shenkin, P. S., Palzkill, T., Amino acid sequence determinants of beta-lactamase structure and activity. J Mol Biol 258 (4), 688–703, 1996.

Humphrey, W., Dalke, A., Schulten, K., VMD – Visual Molecular Dynamics. J Mol Graph Model 14, 33–38, 1996.

Itkowitz, B., Handley, J., Zhu, W., The OpenHaptics toolkit: A library for adding 3D touch navigation and haptics to graphics applications. In: Proceedings of WHC '05, IEEE Computer Society, Washington, DC, USA, 590–591, 2005.

Jacoby, G. A., Extended-spectrum beta-lactamases and other enzymes providing resis-

tance to oxyimino-beta-lactams. Infect Dis Clin North Am 11 (4), 875–887, 1997.

Jacoby, G. A., Beta-lactamase nomenclature. Antimicrob Agents Chemother 50 (4), 1123–1129, 2006.

Jaurin, B., Grundstrom, T., Ampc cephalosporinase of *Escherichia coli* K-12 has a different evolutionary origin from that of beta-lactamases of the penicillinase type. Proc Natl Acad Sci U S A 78 (8), 4897–4901, 1981.

Jelsch, C., Lenfant, F., Masson, J. M., Samama, J. P., Beta-lactamase TEM1 of *E. coli* crystal structure determination at 2.5 Å resolution. FEBS Letters 299 (2), 135–142, 1992.

Jelsch, C., Mourey, L., Masson, J. M., Samama, J. P., Crystal structure of *Escherichia coli* TEM1 beta-lactamase at 1.8 Å resolution. Proteins 16 (4), 364–83, 1993.

Knothe, H., Shah, P., Krcmery, V., Antal, M., Mitsuhashi, S., Transferable resistance to cefotaxime, cefoxitin, cefamandole and cefuroxime in clinical isolates of *Klebsiella pneumoniae* and *Serratia marcescens*. Infection 11 (6), 315–317, 1983.

Křenek, A., Haptic rendering of molecular flexibility. In: Proceedings of PURS2000, 19–26, 2000.

Křenek, A., Haptic rendering of molecular conformations. In: Proceedings of Eurohaptics 2001, 142–145, 2001.

Křenek, A., Haptic rendering of complex force fields. In: EGVE '03: Proceedings of the workshop on Virtual environments 2003, 231–239, 2003.

Křenek, A., Kabeláč, Z., Studying conformational behaviour with PHANToM device(s). In: Proceedings of PURS'99, 3–6, 1999.

Křenek, A., Černohorský, M., Kabeláč, Z., Haptic visualization of molecular data. In: Skala, V. (Ed.), WSCG'99 Conference Proceedings, 1999.

Lamotte-Brasseur, J., Dive, G., Dideberg, O., Charlier, P., Frère, J. M., Ghuysen, J. M.,

Mechanism of acyl transfer by the class A serine beta-lactamase of *Streptomyces albus* G. Biochem J 279 (Pt 1), 213–221, 1991.

Legge, F. S., Budi, A., Treutlein, H., Yarovsky, I., Protein flexibility: multiple molecular dynamics simulations of insulin chain b. Biophys Chem 119 (2), 146–157, 2006.

Lehtovuori, P. T., Nyrönen, T. H., Soma – workflow for small molecule property calculations on a multiplatform computing grid. J Chem Inf Model 46 (2), 620–625, 2006.

Li, X.-Z., Nikaido, H., Efflux-mediated drug resistance in bacteria. Drugs 64 (2), 159–204, 2004.

Likic, V. A., Gooley, P. R., Speed, T. P., Strehler, E. E., A statistical approach to the interpretation of molecular dynamics simulations of calmodulin equilibrium dynamics. Protein Sci 14 (12), 2955–2963, 2005.

Lipari, G., Szabo, A., Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J Am Chem Soc 104 (17), 4546–4559, 1982.

Livermore, D., Williams, J., 1996. Antibiotics in Laboratory Medicine, 4th ed. Williams and Wilkins, Baltimore, Ch. Mode of action and mechanisms of bacterial resistance, 502–577.

Livermore, D. M., Beta-lactamases in laboratory and clinical resistance. Clin Microbiol Rev 8 (4), 557–584, 1995.

Loris, R., Langhorst, U., De Vos, S., Decanniere, K., Bouckaert, J., Maes, D., Transue, T. R., Steyaert, J., Conserved water molecules in a large family of microbial ribonucleases. Proteins 36 (1), 117–134, 1999.

Loris, R., Stas, P. P. G., Wyns, L., Conserved waters in legume lectin crystal-structures - the importance of bound water for the sequence-structure relationship within the legume lectin family. J Biol Chem 269 (43), 26722–26733, 1994.

Lu, Y., Wang, R., Yang, C. Y., Wang, S., Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. J Chem Inf Model 47 (2), 668–75, 2007.

Lundin, K., G., B., Ynnerman, A., General proxy-based haptics for volume visualization. In: Proceedings of WHC '05, IEEE Computer Society, Washington, DC, USA, 557–560, 2005.

Lundin, K., Ynnerman, A., Gudmundsson, B., Proxy-based haptic feedback from volumetric density data. In: Proceedings of the Eurohaptic Conference, 104–109, 2002.

Macheboeuf, P., Guilmi, A. M. D., Job, V., Vernet, T., Dideberg, O., Dessen, A., Active site restructuring regulates ligand recognition in class A penicillin-binding proteins. Proc Natl Acad Sci U S A 102 (3), 577–582, 2005.

Maciejewski, R., Choi, S., Ebert, D., Tan, H., Multi-modal perceptualization of volumetric data and its application to molecular docking. In: Proceedings of WHC '05, IEEE Computer Society, Washington, DC, USA, 511–514, 2005.

Majiduddin, F. K., Materon, I. C., Palzkill, T. G., Molecular analysis of beta-lactamase structure and function. Int J Med Microbiol 292 (2), 127–137, 2002.

Majiduddin, F. K., Palzkill, T., An analysis of why highly similar enzymes evolve differently. Genetics 163 (2), 457–466, 2003.

Mark, W., Randolph, S., Finch, M., Van Verth, J., Taylor, R.M., I., Adding force feedback to graphics systems: issues and solutions. In: Proceedings of SIGGRAPH '96, 447–452, 1996.

Massova, I., Mobashery, S., Kinship and diversification of bacterial penicillin-binding proteins and beta-lactamases. Antimicrob Agents Chemother 42 (1), 1–17, 1998.

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

Matthew, M., Plasmid-mediated beta-lactamases of gram-negative bacteria: properties and distribution. J Antimicrob Chemother 5 (4), 349–358, 1979.

Mattos, C., Protein-water interactions in a dynamic world. Trends Biochem Sci 27 (4), 203–208, 2002.

McCann, K. M., Yarrow, M., DeVivo, A., Mehrotra, P., Scyflow: an environment for the visual specification and execution of scientific workflows: Research articles. Concurr Comput Pract Exper 18 (10), 1155–1167, 2006.

McDowell, T. D., Reed, K. E., Mechanism of penicillin killing in the absence of bacterial lysis. Antimicrob Agents Chemother 33 (10), 1680–1685, 1989.

Mclachlan, A. D., Rapid comparison of protein structures. Acta Crystallogr A 38 (Nov), 871–873, 1982.

Medeiros, A. A., Beta-lactamases: quality and resistance. Clin Microbiol Infect 3 Suppl 4, S2–S9, 1997a.

Medeiros, A. A., Evolution and dissemination of beta-lactamases accelerated by generations of beta-lactam antibiotics. Clin Infect Dis 24 Suppl 1, S19–S45, 1997b.

Miller, A., 1983. Antibiotics Containing the Beta-Lactam Structure 11. Springer–Verlag, Heidelberg, Ch. In vitro and in vivo laboratory evaluation of b-lactam antibiotics, 79–118.

Minasov, G., Wang, X., Shoichet, B. K., An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation. J Am Chem Soc 124 (19), 5333–40, 2002.

Morin, R. B., Jackson, B. G., Flynn, E. H., Roeske, R. W., Chemistry of Cephalosporin Antibiotics. I. 7-Aminocephalosporanic Acid from Cephalosporin C. J Am Chem Soc 84 (17), 3400–3401, 1962.

Morlot, C., Noirclerc-Savoye, M., Zapun, A., Dideberg, O., Vernet, T., The D,D-

carboxypeptidase PBP3 organizes the division process of *Streptococcus pneumoniae*. Mol Microbiol 51 (6), 1641–1648, 2004.

Mustata, G., Briggs, J. M., Cluster analysis of water molecules in alanine racemase and their putative structural role. Protein Eng 17 (3), 223–234, 2004.

Nakasako, M., Water-protein interactions from high-resolution protein crystallography. Philos Trans R Soc Lond B Biol Sci 359 (1448), 1191–1204, 2004.

Newton, G., Abraham, E., Cephalosporin C, a new antibiotic containing sulphur and D-alpha-aminoadipic acid. Nature 175 (4456), 548, 1955.

Nielsen, J. B., Lampen, J. O., Membrane-bound penicillinases in gram-positive bacteria. J Biol Chem 257 (8), 4490–4495, 1982.

Nikaido, H., Crossing the envelope: how cephalosporins reach their targets. Clin Microbiol Infect 6 Suppl 3, 22–26, 2000.

Notredame, C., Higgins, D. G., Heringa, J., T-coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302 (1), 205–217, 2000.

Ogata, K., Wodak, S. J., Conserved water molecules in MHC class-1 molecules and their putative structural and functional roles. Protein Eng 15 (8), 697–705, 2002.

O'Sullivan, J., Bleaney, R. C., Huddleston, J. A., Abraham, E. P., Incorporation of 3H from delta-(L-alpha-amino (4,5-3H)adipyl)-L-cysteinyl-D-(4,4-3H)valine into isopenicillin N. Biochem J 184 (2), 421–426, 1979.

Otting, G., Liepinsh, E., Wüthrich, K., Protein hydration in aqueous solution. Science 254 (5034), 974–980, 1991.

Ouellette, M., Bissonnette, L., Roy, P. H., Precise insertion of antibiotic-resistance determinants into Tn21-like transposons - nucleotide-sequence of the Oxa-1 beta-lactamase gene. Proc Natl Acad Sci U S A 84 (21), 7378–7382, 1987.

Page, M. I., Laws, A. P., The chemical reactivity of β-lactams, β-sultams and β-phospholactams. Tetrahedron 56, 5631–5638, 2000.

Park, S., Saven, J. G., Statistical and molecular dynamics studies of buried waters in globular proteins. Proteins 60 (3), 450–463, 2005.

Paterson, D. L., Ko, W. C., Gottberg, A. V., Casellas, J. M., Mulazimoglu, L., Klugman, K. P., Bonomo, R. A., Rice, L. B., McCormack, J. G., Yu, V. L., Outcome of cephalosporin treatment for serious infections due to apparently susceptible organisms producing extended-spectrum beta-lactamases: implications for the clinical microbiology laboratory. J Clin Microbiol 39 (6), 2206–2212, 2001.

Peterlik, I., Křenek, A., Haptically driven travelling through conformational space. In: Proceedings of WHC '05, 342–347, 2005.

Pettersen, E., Goddard, T., Huang, C., Couch, G., Greenblatt, D., Meng, E., Ferrin, T., UCSF Chimera - a visualization system for exploratory research and analysis. J Comput Chem 25 (13), 1605–1612, 2004.

Philippopoulos, M., Mandel, A. M., Palmer, A. G., Lim, C., Accuracy and precision of NMR relaxation experiments and MD simulations for characterizing protein dynamics. Proteins 28 (4), 481–493, 1997.

Pinho, M. G., Filipe, S. R., de Lencastre, H., Tomasz, A., Complementation of the essential peptidoglycan transpeptidase function of penicillin-binding protein 2 (PBP2) by the drug resistance protein PBP2A in *Staphylococcus aureus*. J Bacteriol 183 (22), 6525–6531, 2001.

Poole, K., Outer membranes and efflux: the path to multidrug resistance in gram-negative bacteria. Curr Pharm Biotechnol 3 (2), 77–98, 2002.

Poole, K., Efflux-mediated multiresistance in gram-negative bacteria. Clin Microbiol Infect 10 (1), 12–26, 2004a.

Poole, K., Resistance to beta-lactam antibiotics. Cell Mol Life Sci 61 (17), 2200–2223,

2004b.

Prabhu, N., Sharp, K., Protein-solvent interactions. Chem Rev 106 (5), 1616–1623, 2006.

Roccatano, D., Sbardella, G., Aschi, M., Amicosante, G., Bossa, C., Di Nola, A., Mazza, F., Dynamical aspects of TEM-1 beta-lactamase probed by molecular dynamics. J Comput Aided Mol Des 19 (5), 329–340, 2005.

Rolinson, G. N., Sutherland, R., Semisynthetic penicillins. Adv Pharmacol Chemother 11, 151–220, 1973.

Ruspini, D. C., Kolarov, K., Khatib, O., The haptic display of complex graphical environments. In: Computer Graphics (SIGGRAPH 97 Conference Proceedings), 345–352, 1997.

Ryckaert, J. P., Ciccotti, G., Berendsen, H. J. C., Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes. J Comput Phys 23 (3), 327–341, 1977.

Sanjeev, B., Ankush. Indian Institute of Science.

Sankaranarayanan, G., Weghorst, S., Sanner, M., Gillet, A., Olson, A., Role of haptics in teaching structural molecular biology. In: Proceedings of HAPTICS'03, 363, 2003.

Sanschagrin, P. C., Kuhn, L. A., Cluster analysis of consensus water sites in thrombin and trypsin shows conservation between serine proteases and contributions to ligand specificity. Protein Sci 7 (10), 2054–64, 1998.

Savard, P. Y., Gagne, S. M., Backbone dynamics of TEM-1 determined by NMR: Evidence for a highly ordered protein. Biochemistry 45 (38), 11414–11424, 2006.

Sharrow, S. D., Edmonds, K. A., Goodman, M. A., Novotny, M. V., Stone, M. J., Thermodynamic consequences of disrupting a water-mediated hydrogen bond network in a protein:pheromone complex. Protein Sci 14 (1), 249–256, 2005.

Sifaoui, F., Arthur, M., Rice, L., Gutmann, L., Role of penicillin-binding protein 5 in ex-

pression of ampicillin resistance and peptidoglycan structure in *Enterococcus faecium*. Antimicrob Agents Chemother 45 (9), 2594–2597, 2001.

Simmerling, C., Elber, R., Zhang, J., MOIL-View – a program for visualization of structure and dynamics of biomolecules and STO – a program for computing stochastic paths. Modelling of Biomolecular Structure and Mechanisms, 241–265, 1995.

Spoel, D. V. D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., Berendsen, H. J. C., Gromacs: fast, flexible, and free. J Comput Chem 26 (16), 1701–1718, 2005.

Spratt, B. G., Resistance to antibiotics mediated by target alterations. Science 264 (5157), 388–393, 1994.

Stone, J. E., Gullingsrud, J., Schulten, K., A system for interactive molecular dynamics simulation. In: I3D '01: Proceedings of the 2001 symposium on Interactive 3D graphics, ACM Press, New York, NY, USA, 191–194, 2001.

Straub, J. E., Thirumalai, D., Theoretical probes of conformational fluctuations in S-peptide and RNase A/3'-UMP enzyme product complex. Proteins 15 (4), 360–373, 1993.

Strynadka, N. C., Adachi, H., Jensen, S. E., Johns, K., Sielecki, A., Betzel, C., Sutoh, K., James, M. N., Molecular structure of the acyl-enzyme intermediate in beta-lactam hydrolysis at 1.7 Å resolution. Nature 359 (6397), 700–5, 1992.

Takano, K., Yamagata, Y., Yutani, K., Buried water molecules contribute to the conformational stability of a protein. Protein Eng 16 (1), 5–9, 2003.

Taylor, I., Shields, M., Wang, I., Philp, R., Distributed P2P computing within Triana: A galaxy visualization test case. In: IPDPS '03: Proceedings of the 17th International Symposium on Parallel and Distributed Processing, IEEE Computer Society, Washington, DC, USA, 16.1, 2003.

Thain, D., Tannenbaum, T., Livny, M., 2002. Condor and the Grid. In: Berman, F.,

Fox, G., Hey, T. (Eds.), Grid Computing: Making the Global Infrastructure a Reality. John Wiley & Sons Inc.

Tony, A., Curbera, F., Dholakia, H., Goland, Y., Klein, J., Leymann, F., Liu, K., Roller, D., Smith, D., Thatte, S., Trickovic, I., Weerawarana, S., Business process execution language for web services version (BPEL4WS) 1.1.

Tzouvelekis, L. S., Tzelepi, E., Tassios, P. T., Legakis, N. J., CTX-M-type beta-lactamases: an emerging group of extended-spectrum enzymes. Int J Antimicrob Agents 14 (2), 137–142, 2000.

Vakulenko, S. B., Taibi-Tronche, P., Toth, M., Massova, I., Lerner, S. A., Mobashery, S., 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), 23052–23060, 1999.

Vakulenko, S. B., Toth, M., Taibi, P., Mobashery, S., Lerner, S. A., Effects of Asp-179 mutations in Tem(Puc19) beta-lactamase on susceptibility to beta-lactams. Antimicrob Agents Chemother 39 (8), 1878–1880, 1995.

Walsh, T. R., Toleman, M. A., Poirel, L., Nordmann, P., Metallo-beta-lactamases: the quiet before the storm? Clin Microbiol Rev 18 (2), 306–325, 2005.

Wang, X., Minasov, G., Shoichet, B. K., Noncovalent interaction energies in covalent complexes: TEM-1 beta-lactamase and beta-lactams. Proteins 47 (1), 86–96, 2002.

Webber, M. A., Piddock, L. J. V., The importance of efflux pumps in bacterial antibiotic resistance. J Antimicrob Chemother 51 (1), 9–11, 2003.

Xu, D., Tsai, C. J., Nussinov, R., Hydrogen bonds and salt bridges across protein-protein interfaces. Protein Eng 10 (9), 999–1012, 1997.

Yarrow, M., McCann, K. M., Biswas, R., der Wijngaart, R. F. V., An advanced user interface approach for complex parameter study process specification on the informa-

tion power grid. In: GRID '00: Proceedings of the First IEEE/ACM International Workshop on Grid Computing, Springer-Verlag, London, UK, 146–157, 2000.

Yarrow, M., McCann, K. M., Tejnil, E., Vivo, A. D., Production-level distributed parametric study capabilities for the grid. In: GRID '01: Proceedings of the Second International Workshop on Grid Computing, Springer-Verlag, London, UK, 166–176, 2001.

Yu, J., Buyya, R., A taxonomy of workflow management systems for grid computing. J Grid Comput 3 (3-4), 171–200, 2005.

Danksagungen

Herrn Prof. Dr. Rolf D. Schmid danke ich für die Möglichkeit, diese Arbeit an seinem Institut durchführen zu können sowie für die vortrefflichen Arbeitsbedingungen an seinem Institut. Bei Herrn Prof. Dr. Jürgen Pleiss möchte ich mich für die Überlassung des Themas, die fachkundige Betreuung der Arbeit und die Bereitstellung eines optimalen Arbeitsplatzes bedanken. Neben seinen Anregungen und Ideen gab er mir stets auch die Möglichkeit, meine eigenen Vorstellungen umzusetzen. Besonders bedanken möchte ich mich auch dafür, dass er mir die Möglichkeit gegeben hat, im Rahmen meiner Promotion zweimal als Dozent nach Vietnam reisen zu können und dieses wundervolle Land und seine Menschen kennen lernen zu dürfen.

Den Mitarbeitern des Instituts für Technische Biochemie bin ich für die stets offene und freundliche Zusammenarbeit sehr dankbar. Allen meinen Kollegen aus der Abteilung Bioinformatik möchte ich an dieser Stelle einen besonderen Dank für das äußerst angenehme Arbeitsklima aussprechen. Viele unserer gemeinsamen Unternehmungen in den letzen vier Jahren sind zu bleibenden Erinnerungen geworden, insbesondere die immer sehr geselligen Sitzungen des Ältestenrates.

Ich danke Dr. Natalia Currle-Linde vom Höchstleistungsrechenzentrum der Universität Stuttgart sowie Katrin Bidmon, Guido Reina und Sebastian Grottel vom Institut für Visualisierung und Interaktive Systeme an der Universität Stuttgart für die interessante und erfolgreiche Zusammenarbeit. Für die Programmierung allerlei nützlicher Tools danke ich meiner zeitweiligen wissenschaftlichen Hilfskraft Hendrik Nell.

Der Deutschen Forschungsgemeinschaft danke ich für die Finanzierung meiner Arbeit im Rahmen des Schwerpunktprojektes 1170 (Directed Evolution to Optimize and Understand Molecular Biocatalysts).

Meiner Freundin Anja-Cathrin danke ich für ihr Verständnis für die vielen vor dem Computer verbrachten Nächte, aber ganz besonders für ihre liebevolle Unterstützung und ihre Geduld während der letzen Jahre.

Meinen Eltern danke ich für die immerwährende Unterstützung und dafür, dass sie mir das Studium ermöglicht und immer an mich geglaubt haben.

Erklärung

Ich erkläre hiermit, daß ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Stuttgart, den 11. Oktober 2007