

***In vitro* and *in silico* Models for *in vivo* Events:**

- A. Identification of Factors Impeding the Production of a Single-chain Antibody Fragment in *Escherichia coli* by Comparing *in vivo* and *in vitro* Expression**
  
- B. An Assay for the Prediction of Substrate Specificities of Metallo- $\beta$ -lactamases Based on Molecular Dynamics Simulations**

Bei der Fakultät Geo- und Biowissenschaften der Universität Stuttgart eingereichte  
Dissertation zur Erlangung des Grades eines  
Doktors der Naturwissenschaften (Dr. rer. nat.)

vorgelegt von  
Peter Ölschläger  
aus Remshalden

1. Referent: Prof. Dr. Rolf D. Schmid  
2. Referent: Prof. Dr. Christoph Syldatk

Tag der mündlichen Prüfung: 14. August, 2002



## Danksagung

Mein besonderer Dank gilt Prof. Dr. Rolf D. Schmid für die Überlassung und Betreuung der beiden interessanten Themen und für die optimalen Arbeitsbedingungen sowohl bezüglich experimenteller als auch rechnergestützter Arbeiten.

Dr. Jutta Schmitt und Dr. Stefan Lange danke ich für die sehr gute und ausdauernde Betreuung von Thema A.

Bei P.D. Dr. Jürgen Pleiss bedanke ich mich für die kompetente Betreuung von Thema B.

Srividhya Srikant-Iyer danke ich für die Überlassung des Plasmids pETK(+).

Großer Dank gebührt allen Mitgliedern des Instituts für Technische Biochemie für permanente fachliche und moralische Unterstützung. Besonders hervorheben möchte ich Volker Nödinger, der mir bei der DNA- und Proteinsequenzierung sowie in allen labortechnischen Angelegenheiten stets mit Rat und Tat zur Seite stand, und Markus Fischer und Florian Barth, die mich kompetent und geduldig bei der Bewältigung von Hardware- und Software-Problemen unterstützt haben.

Dr. Martin Siemann-Herzberg, Institut für Bioverfahrenstechnik der Universität Stuttgart, möchte ich für die Betreuung von Thema A sowie ihm und Sabine Arnold, ebenfalls Institut für Bioverfahrenstechnik, für interessante Diskussionen danken.

Dr. Josef Altenbuchner, Institut für Industrielle Genetik der Universität Stuttgart, danke ich für das Plasmid pTST101.

Bei Dr. Annett Burzlaff, Institut für Zellbiologie und Immunologie der Universität Stuttgart, bedanke ich mich für die mikroskopischen Aufnahmen.

Ich möchte mich bedanken bei Sandra Baumann and Kai Scharenweber, Institut für Bioverfahrenstechnik der Universität Stuttgart für die Durchführung der zellfreien gekoppelten Transkriptions-/Translations-Experimente.

Achim Hauck und Susanne Zibek danke ich für die technische Unterstützung bei der Durchführung von Fed-batch-Kultivierungen.

Ich bedanke mich bei Petra Vogelsang, Thomas Kauf und Peter Trodler für die wissenschaftliche Mitarbeit im Labor sowie bei Matthias Dietrich und Daniela Dürschnabel für die wissenschaftliche Mitarbeit beim molecular modeling.

Mein größter Dank gilt meinen Eltern Ilse und Heinz Ölschläger, die mir immer und absolut bedingungslos in allen Bereichen des Lebens zur Seite stehen und somit die Durchführung dieser Arbeit möglich gemacht haben.



## Zusammenfassung

In Teil A wurde der Atrazin-spezifische scFv (single-chain variables Antikörperfragment) K411B, der in der Herbizid-Analytik eingesetzt werden kann, entweder im Zytoplasma oder im Periplasma von *Escherichia coli* BL21(DE3) exprimiert. Für die periplasmatische Produktion, wurde der scFv N-terminal mit einem pelB leader fusioniert, wohingegen die nicht-fusionierte Variante zu zytoplasmatischer Expression führte. Der Expressionslevel unterschied sich deutlich: der Grad an Proteinansammlung des scFv mit leader war 2,3 mal höher als der des Proteins ohne leader. Um diesen Sachverhalt weiter zu untersuchen, wurden die jeweiligen Translationsprofile durch gekoppelte *in vitro* Transkription-/Translations-Versuche erstellt und ergaben entsprechende Ergebnisse. Periplasmatische Expression ergab nur 10% richtig gefalteten scFv. Derselbe Anteil wurde erhalten, wenn der scFv *in vitro* exprimiert wurde. Dies zeigte, daß die oxidierende Umgebung des Periplasmas die Proteinfaltung nicht positiv beeinflusste. Folglich untermauerten die *in vitro* erhaltenen Daten die Beobachtungen *in vivo* und legten nahe, daß die Diskrepanz in den Expressionslevels die Folge unterschiedlicher Translationseffizienzen war.

Verstärkt grün fluoreszierendes Protein (EGFP) wurde C-terminal an den scFv fusioniert, um ein „online monitoring“ während des Produktionsprozesses zu erlauben. Das erzeugte Fusionsprotein konnte in einem Fluoreszenz-Immunoassay, fluorophor-linked immunosorbent assay (FLISA) genannt, eingesetzt werden, um die Funktionalität des Fusionsproteins zu messen. Die scFv-Domäne, die an das immobilisierte Hapten bindet, kann durch Fluoreszenzmessung der EGFP-Domäne detektiert werden. Das zeitraubende Binden von sekundären Antikörpern und die Enzymreaktion, die für ELISAs (enzyme-linked immunosorbent assays) notwendig sind, war nicht erforderlich. Deshalb ist die Assay-Zeit von 1,5 Stunden für den FLISA viel kürzer als die eines vergleichbaren ELISAs, der ungefähr 5 Stunden in Anspruch nimmt. Die Menge der löslichen Fraktion von Extrakten von *E. coli* Zellen, die das Fusionsprotein exprimierten und die normalisierte Fluoreszenz zeigten eine lineare Korrelation mit  $R^2 > 0.99$ . Der FLISA wurde benutzt, um die Menge an funktionellem, in *E. coli* exprimiertem scFv-EGFP Fusionsprotein zu bestimmen. Das Fusionsprotein konnte nicht im Periplasma, aber im Zytoplasma exprimiert werden und zeigte ein ähnliches Expressionsprofil wie der scFv. *In vitro* ergab die Expression mit und ohne pelB leader das gleiche Produktionsprofil wie der scFv. Das deutete darauf hin, das weder die Translationseffizienz noch die Löslichkeit sondern andere Faktoren die periplasmatische Expression des Fusionsproteins verhinderten.

In einer Hochzelldichte-Kultivierung wurden 25 g (2 g pro liter) scFv produziert. Da der größte Teil des rekombinanten Proteins unlöslich war, wurde eine Refolding-Methode entwickelt, die den scFv nahezu quantitativ in die aktive Form brachte. Alternativ konnte aktiver scFv aus dem Kulturüberstand durch immobilisierte Metallionen-Affinitäts-Chromatographie (IMAC) zur

Homogenität aufgereinigt werden. In der *s*-Triazin-Analytik eingesetzt, zeigte der FLISA mit scFv-EGFP Fusionsprotein bessere Ergebnisse als der ELISA mit scFv.

In Teil B wurde mit Hilfe von molekulardynamischen Simulationen ein Assay zur Vorhersage von Substratspezifitäten der IMP-1 Metallo- $\beta$ -Lactamase und Varianten gegenüber Cephalosporin-Antibiotika entwickelt. Um eine verlässliche Modeling-Methode für das zwei Zinkatome enthaltende aktive Zentrum zu etablieren, wurde ein rein ungebundener Ansatz und ein kationischer Dummyatom-Ansatz ausprobiert. Letzterer gab das Protein besser wieder. Das Enzym im Komplex mit einem Mercapto-carboxylat-Inhibitor, so wie es kristallisiert worden war, konnte mit dieser Methode bei 300 K gut simuliert werden. Das aktive Zentrum, insbesondere die Koordination der beiden Zinkatome blieb stabil. Anschließend wurde das Modell erfolgreich auf das freie Enzym und das Enzym im Komplex mit einem  $\beta$ -Lactam in einer Zwischenstufe angewandt. In dieser Struktur ist das  $\beta$ -Lactam bereits hydrolysiert und zu einem Zink mit dem Carboxylat, zum anderen Zink mit dem anionischen Stickstoff, die durch die Hydrolyse der Amidbindung entstanden sind, koordiniert. Cephalothin, ein Substrat das gut umgesetzt wird, blieb in dieser Zwischenstufe bei 300 K über einen Zeitraum von 1,2 ns stabil gebunden. Die Zink-Zink-Abstände des freien Enzyms und des Enzyms im Komplex mit der  $\beta$ -Lactam-Zwischenstufe unterschieden sich deutlich and deuten darauf hin, daß während der katalytischen Umsetzung des Substrats eine Bewegung der Zinks stattfindet.

Um IMP-6, eine Mutante, die sich in einer Aminosäure von IMP-1 unterscheidet, und andere  $\beta$ -Lactam-Substrate zu untersuchen, wurden die Moleküle entsprechend Cephalothin jeweils in IMP-1 und IMP-6 gedockt. Für diese Enzym/Substrat-Kombinationen waren Literaturdaten ( $k_{cat}/K_M$ -Werte) vorhanden, und diese wiesen deutliche Unterschiede auf. Da nicht alle Kombinationen bei 300 K in der Zwischenstufe stabil blieben, wurden die molekulardynamischen Simulationen bei 100 K durchgeführt. Nachdem mehrere Simulationen für jede Kombination durchgeführt worden waren, stellten sich drei Parameter heraus, die zu experimentellen Daten korreliert waren: 1. In Kombinationen mit sehr ineffizienter Hydrolyse verlor der anionische Stickstoff Kontakt zum entsprechenden Zink, 2. wurde eine Deformation des Winkels zwischen dem anionischen Stickstoff, dem Dummy-Atom (ein Teil des kationischen Dummy-Atom-Ansatzes) und dem Zink bei Kombinationen mit schlechter Umsetzung beobachtet, und 3. erfolgte eine Vergrößerung des Zink-Zink-Abstandes in ineffizienten Enzym/Substrat-Kombinationen. Der genannte Winkel und der genannte Abstand waren für jede Kombination voneinander abhängig und wenn sie in 2D-Diagrammen gegeneinander aufgetragen wurden, ergab sich eine Gruppierung in effiziente, mittelmäßig aktive und ineffiziente Metallo- $\beta$ -Lactamase/ $\beta$ -Lactam-Kombinationen.

## Summary

In Part A the atrazine-specific scFv (single-chain variable antibody fragment) K411B, that can be applied in herbicide analysis, was expressed either in the cytoplasm or the periplasm of *Escherichia coli* BL21(DE3). For periplasmic production, the scFv was N-terminally fused to the pelB leader, whereas the unfused variant resulted in cytoplasmic expression. The expression level differed significantly: the extent of protein accumulation of the scFv with leader was 2.3 times higher than that of the protein without leader. To further investigate this, the respective translation profiles were generated by coupled *in vitro* transcription/translation assays and gave according results. Periplasmic expression resulted in only 10% correctly folded scFv. The same percentage was obtained when the scFv was expressed *in vitro*, indicating that the oxidizing environment of the periplasm did not increase proper folding. Thus, the data obtained *in vitro* confirmed the findings observed *in vivo* and suggested that the discrepancy in expression levels was due to different translation efficiencies.

Enhanced green fluorescent protein (EGFP) was fused C-terminally to the scFv to allow online monitoring during the production process. The resulting fusion protein could be used in a novel fluorescence immunoassay referred to as fluorophor-linked immunosorbent assay (FLISA) in order to measure functionality of the fusion protein. The scFv domain, which binds to the immobilized hapten, can be detected by measuring the fluorescence of the EGFP domain. The time-consuming binding of secondary antibodies and enzyme reaction, necessary for ELISAs (enzyme-linked immunosorbent assays), is not required. Consequently, the assay time of 1.5 hours needed to complete the FLISA is much shorter than that of comparable ELISAs, which require about 5 hours. The amount of the soluble fraction of cell extracts from *Escherichia coli* expressing the fusion protein and the normalized fluorescence signal showed a linear correlation with  $R^2 > 0.99$ . The FLISA was used to determine the amount of functional scFv-EGFP fusion protein expressed in *E. coli*. The fusion protein could not be expressed in the periplasm but in the cytoplasm and showed a similar expression profile as the scFv. *In vitro*, the expression with and without the pelB leader rendered the same production profile as for the scFv. This indicated that neither the translation efficiency nor the solubility but other factors impeded periplasmic expression of the fusion protein.

In a high-cell-density cultivation, 25 grams (2 grams per liter) of scFv were produced. As the major fraction of recombinant protein was insoluble, a refolding method was developed yielding active scFv in nearly quantitative yields. Alternatively, active scFv could be purified from the culture supernatant to homogeneity by immobilized metal affinity

chromatography. When applied to *s*-triazine analysis, the FLISA performed with scFv-EGFP fusion protein showed better results than the respective ELISA using scFv.

In Part B an assay based on molecular dynamics simulations for the prediction of substrate specificities of the IMP-1 metallo- $\beta$ -lactamase and variants towards cephalosporin antibiotics was developed. To establish a reliable modeling method for the active site, containing two zinc atoms, a purely nonbonded approach and a cationic dummy atom approach were tested. The latter was shown to give a better representation of the protein. The enzyme in complex with a mercaptocarboxylate inhibitor as it had been crystallized could be simulated well with this procedure at 300 K. The active site, i.e. the coordination of the two zinc atoms remained stable. Subsequently, the model was successfully extended to the free enzyme and to the enzyme in complex with a  $\beta$ -lactam bound as an intermediate. In this structure the  $\beta$ -lactam is already hydrolyzed and coordinated to one zinc via the carboxylate, to the other zinc via the anionic nitrogen resulting from the amide bond cleavage. Cephalothin, a substrate that is known to be converted well, remained stable in the intermediate structure at 300 K for 1.2 ns. The zinc-zinc distances in the free enzyme and the enzyme in complex with the  $\beta$ -lactam intermediate differed significantly and suggest that upon catalytic conversion of the antibiotic a movement of the zinc atoms occurs.

To investigate IMP-6, a mutant differing in one amino acid from IMP-1, and other  $\beta$ -lactam substrates, the molecules were docked into IMP-1 and IMP-6, respectively, according to cephalothin. For these enzyme/substrate combinations literature data ( $k_{cat}/K_M$  values) were available and these showed significant differences. As not all combinations remained stable in the intermediate structure at 300 K, molecular dynamics simulations were performed at 100 K. After several simulations had been carried out for each combination, three parameters proved to be correlated to the experimental data: 1. For combinations with very inefficient hydrolysis, the anionic nitrogen lost contact to the respective zinc, 2. a deformation of the angle between the anionic nitrogen, the dummy atom (a component of the cationic dummy atom approach) and the zinc atom was observed for combinations with poor conversion and 3. an increase of the zinc-zinc distance occurred in inefficient enzyme/substrate combinations. The mentioned angle and distance were related in each combination and when they were plotted versus each other in a 2D chart, a grouping of inefficient, middle-rate and efficient metallo- $\beta$ -lactamase/ $\beta$ -lactam combinations was observed.



## Abbreviations

% (m/v)	mass percent
% (v/v)	volume percent
% B/B <sub>0</sub>	normalized absorbance
°C	degree Celsius
µg	microgramm
µl	microliter
A	absorbance (dimensionless)
Å	Ångström (=10 <sup>-10</sup> m)
A <sub>0</sub>	absorbance without hapten
A <sub>excess</sub>	absorbance at excess of hapten
Amp <sup>R</sup>	ampicillin resistance
APS	ammonium persulfate
BCA	bicinchoninic acid
<i>bla</i>	β-lactamase gene
bp	base pairs
BPB	bromphenol blue
BSA	bovine serum albumine
CAZ	ceftazidime
CEF	cephalothin
CIAP	calf intestinal alkaline phosphatase
CTX	cefotaxime
d	demineralized
DCC	dicyclohexylcarbodiimide
dd	double demineralized
ddNTP	dideoxynucleoside-5'-triphosphate
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside-5'-triphosphate
DTE	dithioerythritol
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced green fluorescent protein
Fab	antigen binding fragment
Fc	crystalline fragment
fs	femtosecond
Fv	variable fragment
g	gramm
h	hour
IgG	immunoglobulin G
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-β-D-thiogalactoside

---

kb	kilobases
$k_{\text{cat}}$	half-maximal turnover rate
kDa	kilo Dalton
$K_M$	Michaelis-Menten constant
LB	Luria-Bertani
LMW	low molecular weight
LOR	cephaloridine
M	mole per liter
mg	milligramm
min	minute
ml	milliliter
mM	millimole per liter
mmol	millimole
mol	mole
NHS	N-hydroxy-succinimid
nm	nanometer
ns	nanosecond
OD	optical density = absorbance at 578 nm
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffer with sodium chloride
PCR	polymerase chain reaction
PEG	polyethylenglycol
pO <sub>2</sub>	oxygen partial pressure
PVDF	polyvinylidene fluoride
$R_f$	retention factor
rpm	rotations per minute
RT	room temperature
s	second
scFv	single-chain variable fragment
SDS	sodium dodecyl sulfate
SRP	signal recognition particle
TAE	Tris/acetate/EDTA buffer
TBE	Tris/boric acid/EDTA buffer
TE	Tris/EDTA buffer
TES	trace element solution
TEMED	N,N,N',N'-tetramethyl-ethylendiamine
TMB	3,3',5,5'-tetramethylbenzidine
Tris	tris-(hydroxymethyl)-aminomethane
TS	Tris/sodium chloride buffer
TSS	transformation storage solution
UV	ultraviolet
v	volume

Additionally, the common codes for nucleotides, amino acids and customary abbreviations for restriction enzymes, polymerases etc. were used.

---

<b>A 1</b>	<b>INTRODUCTION</b>	<b>15</b>
A 1.1	ANTIBODIES AND ANTIBODY FRAGMENTS	15
A 1.1.1	<i>The Immune System and Immunoglobulins</i>	15
A 1.1.2	<i>Antibodies</i>	15
A 1.1.3	<i>Antibody Fragments</i>	18
A 1.1.4	<i>The scFv K411B and its Application in s-Triazine Analysis</i>	19
A 1.2	EXPRESSION OF RECOMBINANT PROTEINS IN <i>ESCHERICHIA COLI</i>	20
A 1.3	EXPRESSION OF RECOMBINANT PROTEINS <i>IN VITRO</i> BY CELL-FREE COUPLED TRANSCRIPTION/TRANSLATION	22
A 1.4	LIMITING FACTORS IN THE PRODUCTION OF RECOMBINANT PROTEINS IN <i>ESCHERICHIA COLI</i> 23	
A 1.4.1	<i>Transcription</i>	23
A 1.4.2	<i>Translation</i>	24
A 1.4.3	<i>Translocation</i>	24
A 1.4.4	<i>Protein Folding</i>	25
A 1.4.5	<i>Cultivation Strategies</i>	27
A 1.4.6	<i>Downstream Processing</i>	28
A 1.5	DETERMINATION OF THE AMOUNT OF TOTAL AND FUNCTIONAL RECOMBINANT PROTEIN	28
A 1.5.1	<i>Antibody-immobilized Competitive ELISA</i>	29
A 1.5.2	<i>Hapten-immobilized Competitive ELISA</i>	30
A 1.6	FLUOLABELING OF PROTEINS BY FUSION TO GREEN FLUORESCENT PROTEIN	30
A 1.7	AIM OF THIS WORK	31
<b>A 2</b>	<b>MATERIALS AND METHODS</b>	<b>32</b>
A 2.1	MATERIALS	32
A 2.1.1	<i>Chemicals and Enzymes</i>	32
A 2.1.2	<i>Instruments</i>	34
A 2.1.3	<i>Consumables</i>	36
A 2.1.4	<i>Microorganisms</i>	36
A 2.1.5	<i>Plasmids</i>	37
A 2.1.6	<i>Synthetic Oligonucleotides</i>	37
A 2.1.6.1	Primers for DNA Sequencing	37
A 2.1.6.2	Primers for Introduction of a His <sub>6</sub> -tag at the C-terminus of the scFv-EGFP fusion protein	37
A 2.1.6.3	Primers for QuikChange PCRs	38

A 2.2	METHODS.....	39
A 2.2.1	<i>Molecular-Genetic Methods</i> .....	39
A 2.2.1.1	Isolation of Plasmid DNA from <i>Escherichia coli</i> with QIAprep Spin Miniprep Kit and QIAGEN Plasmid Midi Kit .....	39
A 2.2.1.2	Extraction and Purification of Plasmid DNA from <i>Escherichia coli</i> for Fast Tests.....	39
A 2.2.1.3	Precipitation of Plasmid DNA with Ethanol .....	39
A 2.2.1.4	Precipitation of Plasmid DNA with Isopropanol.....	40
A 2.2.1.5	Agarose Gel Electrophoresis of DNA .....	40
A 2.2.1.6	Isolation of DNA from Agarose Gels with the QIAquick Gel Extraction Kit.....	40
A 2.2.1.7	Enzymatic DNA modifications .....	40
A 2.2.1.8	Automated DNA Sequencing .....	41
A 2.2.1.9	Polymerase Chain Reaction (PCR).....	42
A 2.2.1.10	QuikChange PCR for Site-directed Mutagenesis .....	42
A 2.2.2	<i>Microbiological Methods</i> .....	43
A 2.2.2.1	Growth Media for <i>Escherichia coli</i> .....	43
A 2.2.2.2	Transformation of <i>Escherichia coli</i> by Heat Shock.....	44
A 2.2.2.3	Expression of Recombinant Proteins in <i>Escherichia coli</i> .....	45
A 2.2.3	<i>Organic Synthesis</i> .....	47
A 2.2.3.1	Synthesis of 2,4-Dichloro-6-ethylamino-s-triazine .....	47
A 2.2.3.2	Synthesis of C <sub>6</sub> -Simazine .....	47
A 2.2.4	<i>Protein-biochemical Methods</i> .....	48
A 2.2.4.1	Cell Disruption and Sample Preparation .....	48
A 2.2.4.2	Quantification of Protein .....	49
A 2.2.4.3	SDS-PAGE.....	49
A 2.2.4.3	N-terminal Sequencing of Proteins.....	51
A 2.2.4.5	Immunological Detection of Proteins (Western Blot) .....	52
A 2.2.4.6	Immunoassays in Microtiter Plates.....	52
A 2.2.4.7	Concentration of Protein.....	56
A 2.2.4.8	Purification of Polyhistidine-tagged proteins by Immobilized Metal Affinity Chromatography .....	56
A 2.2.4.9	Refolding of Insoluble Protein .....	56
A 2.2.4.10	Fluorescence Measurement .....	57
A 2.2.5	<i>Microscopy</i> .....	57
<b>A 3</b>	<b>RESULTS</b> .....	<b>58</b>
A 3.1	CLONING PROCEDURES.....	58
A 3.1.1	<i>Construction of Plasmids for the Expression of the Atrazine-specific ScFv K411B</i> .....	58

A 3.1.1.1	Plasmids for the Expression of the ScFv K411B under Control of the T7 Promoter	58
A 3.1.1.2	Plasmids for the Expression of the ScFv K411B under Control of the <i>rhaBAD</i> Promoter	59
A 3.1.2	Construction of Plasmids for the Expression of the scFv-EGFP Fusion Protein	59
A 3.2	Expression of the scFv K411B and the ScFv-EGFP Fusion Protein in <i>E. coli</i> BL21(DE3) in Shake Flask Cultures	62
A 3.2.2	Quantification of Total Recombinant Protein by Densitometric Evaluation	63
A 3.2.3	Quantification of Functional Recombinant Protein	63
A 3.2.3.1	Comparison of Soluble and Insoluble Fractions	63
A 3.2.3.2	Determination of Active ScFv Using Non-competitive Hapten-immobilized ELISA	64
A 3.2.3.3	Determination of Active ScFv-EGFP Fusion Protein Using Fluorescence Measurement and Non-competitive Hapten-immobilized FLISA	68
A 3.2.4	Microscopic Investigation of <i>E. coli</i> Cells Expressing the ScFv with and without <i>pelB</i> Leader and the ScFv-EGFP Fusion Protein without <i>pelB</i> Leader	71
A 3.3	EXPRESSION OF THE SCFV K411B AND THE SCFV-EGFP FUSION PROTEIN IN CELL-FREE COUPLED TRANSCRIPTION/TRANSLATION	72
A 3.4	COMPARISON OF THE <i>IN VIVO</i> AND <i>IN VITRO</i> RESULTS OF THE EXPRESSION OF SCFV AND SCFV-EGFP FUSION PROTEIN WITH AND WITHOUT PELB LEADER SEQUENCE	74
A 3.5	EXPRESSION OF SCFV WITH PELB LEADER SEQUENCE AND SCFV-EGFP FUSION PROTEIN WITHOUT PELB LEADER SEQUENCE IN <i>E. COLI</i> BL21(DE3) PLYSS IN 2 L FED-BATCH CULTIVATION	76
A 3.6	EXPRESSION OF SCFV WITH PELB LEADER SEQUENCE IN <i>E. COLI</i> BW3110 IN 13 L FED-BATCH CULTIVATION	79
A 3.7	DOWNSTREAM PROCESSING OF RECOMBINANT PROTEIN	81
A 3.7.1	Refolding of Recombinant Protein	81
A 3.7.1.1	Refolding of the scFv K411B	81
A 3.7.1.2	Refolding of the scFv-EGFP Fusion Protein	83
A 3.7.2	One-step Purification of Recombinant Protein by Immobilized Metal Affinity Chromatography	84
A 3.7.2.1	Purification of the ScFv K411B	84
A 3.7.2.2	Purification of the ScFv-EGFP Fusion Protein	85
A 3.7.3	Characterization of the Refolded ScFv K411B	86
A 3.8	APPLICATION OF THE SCFV K411B AND THE SCFV-EGFP FUSION PROTEIN IN S-TRIAZINE ANALYSIS	86
A 3.8.1	Competitive Antibody-fragment-immobilized ELISA	86

A 3.8.2	Competitive Hapten-immobilized ELISA.....	88
A 3.8.3	Competitive Hapten-immobilized FLISA.....	89
A 3.8.4	Cross Reactivities of the Atrazine-specific ScFv and the ScFv-EGFP Fusion Protein Towards other s-Triazines.....	89
<b>A 4</b>	<b>DISCUSSION .....</b>	<b>91</b>
A 4.1	QUANTIFICATION OF FUNCTIONAL HETEROLOGOUS PROTEIN.....	91
A 4.1.1	QUANTIFICATION OF FUNCTIONAL SCFV AND SCFV-EGFP FUSION PROTEIN USING HAPTEN-IMMOBILIZED ELISA AND FLISA.....	91
A 4.2	EXPRESSION ANALYSIS OF THE SCFV K411B AND THE SCFV-EGFP FUSION PROTEIN IN <i>E. COLI</i> .....	95
A 4.2.1	Analysis of Expression by Comparison to <i>in vitro</i> Results.....	97
A 4.2.1.1	Accumulation of heterologous protein .....	98
A 4.2.1.2	Solubility and functionality .....	99
A 4.2.1.3	Conclusions .....	100
A 4.2.2	Expression of the ScFv with <i>pelB</i> Leader Sequence in Fed-batch Cultivations.....	101
A 4.3	REFOLDING AND PURIFICATION OF THE SCFV K411B AND THE SCFV-EGFP FUSION PROTEIN.....	102
A 4.	APPLICATION OF THE SCFV K411B AND THE SCFV-EGFP FUSION PROTEIN IN S-TRIAZINE ANALYSIS.....	103
4.5	CONCLUSIONS.....	105
<b>A 5</b>	<b>REFERENCES.....</b>	<b>106</b>
<b>B 1</b>	<b>INTRODUCTION.....</b>	<b>115</b>
B 1.1	ANTIBIOTICS.....	115
B 1.1.1	$\beta$ -Lactam Antibiotics.....	117
B 1.2	ANTIBIOTIC RESISTANCE.....	120
B 1.2.1	Resistance towards $\beta$ -Lactams.....	120
B 1.3	$\beta$ -LACTAMASES.....	121
B 1.3.1	Serine- $\beta$ -lactamases .....	121
B 1.3.2	Metallo- $\beta$ -lactamases.....	122
B 1.3.2.1	Structure of Metallo- $\beta$ -lactamases .....	123
B 1.3.2.2	Mechanism of $\beta$ -Lactam Hydrolysis .....	125
B 1.4	MOLECULAR MODELING.....	127
B 1.4.1	Molecular Dynamics Simulations of Metallo-proteins.....	128
B 1.4.1.1	Molecular Dynamics Simulations of Metallo- $\beta$ -lactamases.....	129

---

B 1.5	AIM OF THIS WORK.....	130
<b>B 2</b>	<b>MATERIALS AND METHODS.....</b>	<b>131</b>
B 2.1	MATERIALS.....	131
<i>B 2.1.1</i>	<i>Hardware</i> .....	<i>131</i>
<i>B 2.1.2</i>	<i>Software</i> .....	<i>131</i>
B 2.2	METHODS.....	132
<i>B 2.2.1</i>	<i>System Setup for MD Simulations</i> .....	<i>132</i>
B 2.2.1.1	Definition of the Inhibitor, the $\beta$ -Lactams and the Hydroxide .....	132
B 2.2.1.2	Docking of the $\beta$ -Lactams and the Hydroxide.....	134
B 2.2.1.3	Definition of the Active Site.....	135
B 2.2.1.4	Protein Preparation .....	136
B 2.2.1.5	Solvation and Neutralization of the Protein.....	137
<i>B 2.2.2</i>	<i>Minimization</i> .....	<i>137</i>
<i>B 2.2.3</i>	<i>Heating</i> .....	<i>139</i>
<i>B 2.2.4</i>	<i>Equilibration</i> .....	<i>140</i>
<i>B 2.2.5</i>	<i>Unconstrained MD Simulation</i> .....	<i>141</i>
<i>B 2.2.6</i>	<i>Analysis of MD Simulations</i> .....	<i>142</i>
B 2.2.6.1	Visual Analysis.....	142
B 2.2.6.2	Energy Analysis.....	142
B 2.2.6.3	Root Mean Square (RMS) Deviation .....	143
B 2.2.6.4	Measurement of Geometric Parameters Over the Course of Simulations .....	143
<b>B 3</b>	<b>RESULTS.....</b>	<b>144</b>
B 3.1	DEVELOPMENT OF THE MODEL FOR IMP-1 IN COMPLEX WITH THE MERCAPTOCARBOXYLATE INHIBITOR.....	144
<i>B 3.1.1</i>	<i>The Purely Nonbonded Approach</i> .....	<i>146</i>
<i>B 3.1.2</i>	<i>The Cationic Dummy Atom Approach</i> .....	<i>150</i>
B 3.1.2.1	The Cationic Dummy Atom Approach With Protonated Inhibitor .....	151
B 3.1.2.2	The Cationic Dummy Atom Approach With Deprotonated Inhibitor .....	153
<i>B 3.1.4</i>	<i>Evaluation of the Different Models of IMP-1 in Complex with the Mercaptocarboxylate Inhibitor</i> .....	<i>155</i>
B 3.2	APPLICATION OF THE MODEL TO THE FREE IMP-1 ENZYME.....	157
B 3.3	APPLICATION OF THE MODEL TO IMP-1 IN COMPLEX WITH CEPHALOTHIN.....	160
B 3.4	DEVELOPMENT OF AN ASSAY FOR SUBSTRATE SPECIFICITY OF METALLO- $\beta$ - LACTAMASES.....	163
<i>B 3.4.1</i>	<i>MD simulations of IMP-1 and IMP-6 with Cephalothin and Cephaloridine at 300 K..</i>	<i>164</i>

---

<i>B 3.4.2 MD Simulations of IMP-1 and IMP-6 with Cephalothin, Cephaloridine, Cefotaxime and Ceftazidime at 100 K</i> .....	165
<i>B 3.4.3 Analysis of the 100 K MD Simulations</i> .....	169
B 3.4.3.1 Contact Between the Anionic Nitrogen and the Dummy Atom of Zn2.....	169
B 3.4.3.2 Zn1-Zn2 Distance and Angle Between the Anionic Nitrogen, the Dummy Atom and Zn2.....	169
<b>B 4 DISCUSSION</b> .....	<b>175</b>
B 4.1 MOLECULAR DYNAMICS SIMULATIONS OF IMP-1 IN COMPLEX WITH THE MERCAPTOCARBOXYLATE INHIBITOR.....	175
<i>B 4.1.1 Purely Nonbonded Model versus Cationic Dummy Atom Model</i> .....	175
<i>B 4.1.2 Protonation State of the Inhibitor</i> .....	176
<i>B 4.1.3 Movement of the Lid and the Lip</i> .....	177
B 4.2 MD SIMULATIONS WITH IMP-1 AS FREE ENZYME AND IN COMPLEX WITH CEPHALOTHIN.....	177
B 4.3 ASSAY FOR SUBSTRATE SPECIFICITY OF METALLO- $\beta$ -LACTAMASES.....	179
<i>B 4.3.1 Anionic Nitrogen-Dummy Distance</i> .....	180
<i>B 4.3.2 Zn1-Zn2 Distance and Angle Between the Anionic Nitrogen, the Dummy Atom and Zn2</i> .....	180
<i>B 4.3.3 Variance of the Assay</i> .....	181
B 4.4 OUTLOOK.....	182
<b>B 5 REFERENCES</b> .....	<b>183</b>
<b>B 6 ATTACHMENT</b> .....	<b>190</b>
<b>LEBENS LAUF</b> .....	<b>194</b>



## A 1 Introduction

### A 1.1 Antibodies and Antibody Fragments

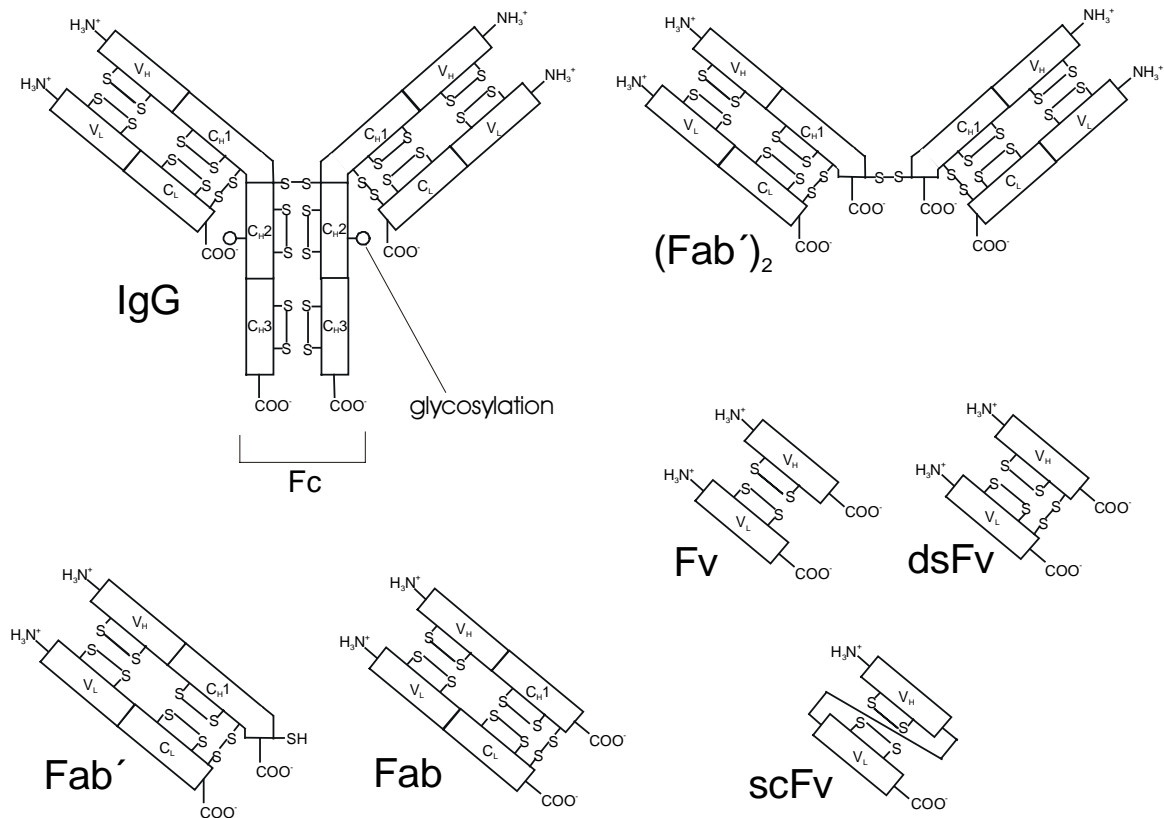
#### A 1.1.1 The Immune System and Immunoglobulins

The immune system of vertebrates is a complex system of molecules and cells designed to protect the animals against microorganisms by distinguishing between „own“ and „foreign“. Antibodies are part of the humoral immune response, are produced by plasma cells and can either inactivate the foreign particle, the antigen, or label it for further processing by the complement system or macrophages. T-cells are part of the cellular immune response. Killer T-cells kill cells that present antigens on their surface and helper T-cells stimulate the humoral immune response by supporting the differentiation of B-cells into plasma-cells. In both cases, the antigen, which is presented by the target cell with a MHC (major histocompatibility complex) protein, is recognized via T-cell receptors [1]. Additionally, cytokines (interleukines and interferones) play an important role in the activation of B- and T-cells [2]. Antibodies, T-cell receptors, MHC proteins and some other receptors belong to the group of immunoglobulins, whose evolutionary relationship and structural appearance is manifested in one or more immunoglobulin domains [3]. Human immunoglobulins can be divided into 5 classes according to their function and physico-chemical properties: IgG, IgM, IgA, IgE and IgD.

#### A 1.1.2 Antibodies

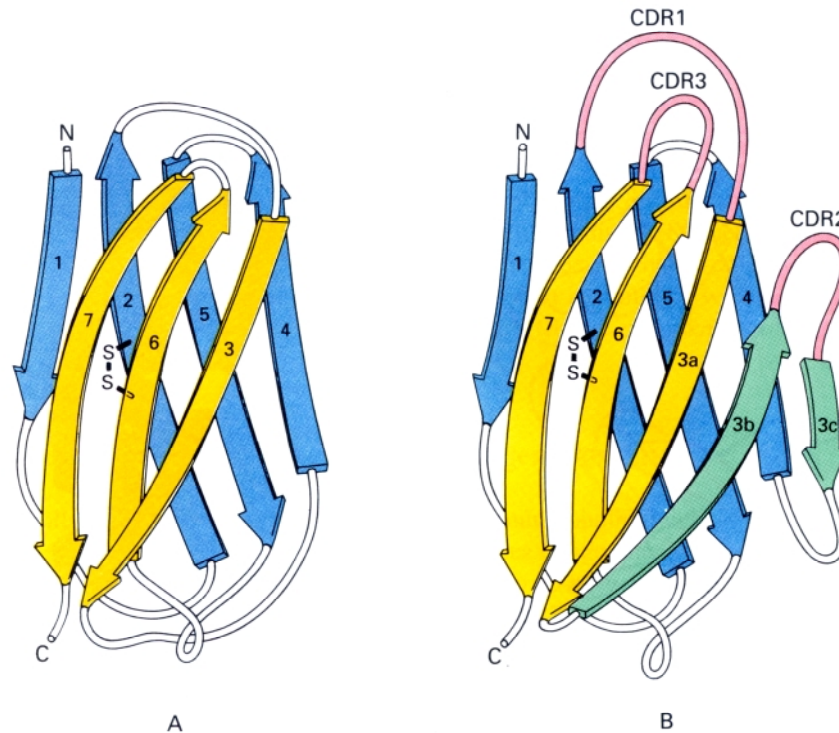
IgGs are the most abundant antibodies and are described in the following. *In vivo* they have a half-life of 20 days and 10 to 15 grams are found in 1 liter of plasma [2]. The IgG molecule consists of two identical light chains and two identical heavy chains, which have a tremendous variation due to alternate splicing and somatic mutation. Each chain has a variable domain ( $V_H$  and  $V_L$ ), which is important for antigen binding. Additionally, the light chains have one constant domain ( $C_L$ ), while the heavy chains have three constant domains ( $C_{H1}$ ,  $C_{H2}$  and  $C_{H3}$ ). By disulfide bridges and noncovalent interactions, the four chains are arranged in a way that results in a Y-shaped molecule. Proteolytic digestion of this molecule with papain results in two antigen binding fragments (Fab) and one crystalline fragment (Fc), called like that because it can be crystallized easily. In the Fab units of the IgG, the variable

domains of one light and one heavy chain are associated as well as the neighboring constant domains. In the Fc unit, the two C<sub>H</sub>2 domains and the two C<sub>H</sub>3 domains are associated, respectively (Figure A 1).



**Figure A 1.** Structure of an IgG and antibody fragments. (Fab')<sub>2</sub> can be obtained by digestion with pepsin, Fab' by consequent reduction. Fab results from digestion with papain. Fv is usually expressed as a recombinant protein. DsFv and scFv are stabilized by a genetically introduced disulfide bond or a peptide linker, respectively.

These domains are structurally related and are called immunoglobulin domains. They consist of two sandwich-like arranged  $\beta$ -sheets, each consisting of 3 or 4 antiparallel  $\beta$ -strands in constant domains and 4 or 5 antiparallel  $\beta$ -strands in variable domains (Figure A 2). The structure of the antiparallel  $\beta$ -strands is fixed by a disulfide bond between two essential cysteines and stabilized by hydrophobic interactions. In the variable domains, three loops consisting of the complementarity determining regions (CDRs) are of great importance for specific antigen binding. All parts of the IgG molecule have specific functions. The Fab fragments are responsible for antigen binding. Fc binds to a Fc receptor of neutrophil and T-killer cells and the C-terminal constant domains bind to the Fc receptor of macrophages. The N-terminal, glycosylated constant domains of Fc bind to a component of the complement system.



**Figure A 2.** Immunoglobulin domain of a constant domain (A) and a variable domain (B). Blue: N-terminal  $\beta$ -sheet, yellow: C-terminal  $\beta$ -sheet, green: extra  $\beta$ -strands in variable domains, red: complementarity determining regions (CDRs).

Antibodies with binding specificity for a certain antigen can be obtained by immunization of test animals. After some weeks a mixture of antibodies with binding specificities for different epitopes (determinants which are recognized by the antibodies) on the target antigen can be found in the serum of these animals. Each type of these antibodies is expressed by one B-cell clone. Consequently, as the mixture contains the antibodies expressed by different B-cell clones, it is referred to as a polyclonal antibody. The hybridoma technique developed by Kohler and Milstein [4] allows the production of monoclonal antibodies specific for only one epitope. B-cells are isolated after immunization from the spleen and fused with tumor cells. By selection and proliferation of a single hybridoma cell expressing antibodies with the desired binding specificity, monoclonal antibodies can be produced in theoretically unlimited amounts.

Many antibodies have been genetically engineered to reduce immunogenicity in humans resulting in chimeric molecules consisting of mouse variable regions and human constant regions [5] as well as humanized antibodies in which just the CDRs are of rodent origin [6].

### A 1.1.3 Antibody Fragments

For many applications such as ELISA (enzyme-linked immunosorbent assay) or *in situ* labeling techniques, the antigen binding activity is of interest, but not the effector activities of the constant regions. Therefore, often antibody fragments consisting only of parts of the parental monoclonal antibody, which confer the characteristic binding activity are prepared (Figure 1). Small antibody fragments have advantages compared to whole immunoglobulins for some clinical applications, such as good penetration of solid tumors and rapid clearance [7, 8]. Additionally, they can be produced conveniently in various expression systems, such as plants, bacteria [9], yeast cells [10-13], by phage display libraries [14] and eucaryotic and procaryotic *in vitro* translation systems [15, 16].

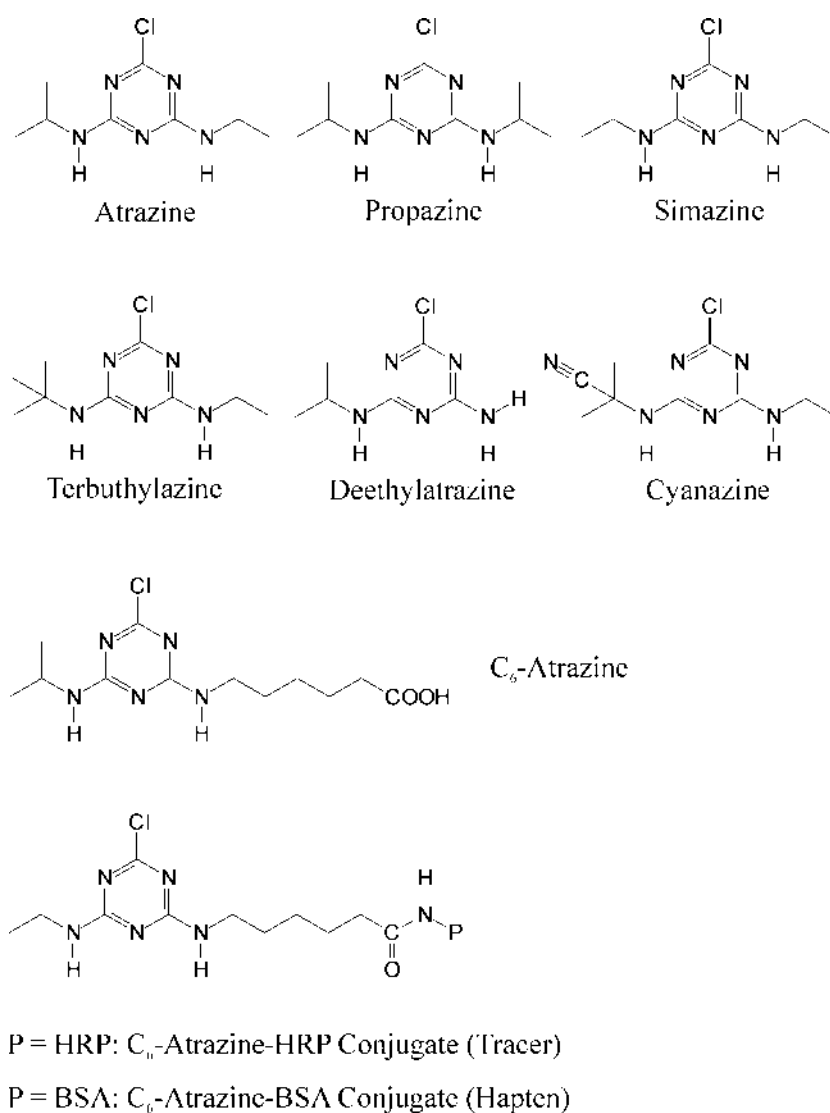
Digestion of IgG with pepsin yields the  $F(ab')_2$  fragment, consisting of the two Fab' fragments, which are linked by a disulfide bridge. Mild reduction of this disulfide bond releases the two Fab' fragments [17], which correspond to the Fab fragment mentioned above, except that the heavy chain is slightly longer due to the different cleavage site of pepsin in comparison to papain. Apart from digestion of IgG with papain, the Fab fragment can also be produced by heterologous expression, for example in *E. coli*. The respective light chain and heavy chain are expressed separately the correct assembly inside the host cell is a frequent problem. Shibui and colleagues only obtained 90 mg/l of functional Fab fragment in a fed-batch culture, although the accumulation of the two chains, each expressed under control of its own tac promoter on the same plasmid and translocated to the periplasm, were much higher (2.88 and 1.28 g/l, respectively). The major part accumulated as periplasmic inclusion bodies and had to be refolded and combined *in vitro* [18]. Due to these problems, Fab fragments are preferably expressed in yeast cells and secreted into the medium [10-13].

The smallest antibody fragment that retains an intact antigen binding site is the variable fragment (Fv) that consists only of the two variable domains,  $V_H$  and  $V_L$ . However, this fragment is unstable, as these domains are free to dissociate. Two strategies have been adopted to overcome this. The first is to link the domains with a peptide to generate a single-chain Fv (scFv). The second is to introduce cysteines at the interface between the  $V_H$  and the  $V_L$  domains, resulting in a Fv stabilized by disulfide bond formation (dsFv) [17]. The first approach is more commonly used because it also reduces the number of genes to be expressed to one and a linker consisting of  $(Gly_4 Ser)_3$  has proven to work well [19]. In addition, by phage display techniques large numbers of scFvs can be screened for antigen binding [14]. Furthermore, scFvs have been fused to a variety of other compounds for different purposes,

such as toxins or drugs for medical applications [7], fluorophors for labeling techniques [20] and other scFvs, resulting in bispecific and bivalent diabodies [21].

#### **A 1.1.4 The scFv K411B and its Application in *s*-Triazine Analysis**

The atrazine-specific monoclonal antibody K4E7 was raised against 4-chloro-6-(ethylamino)-1,3,5-triazine-2-(6-aminohexanecarboxylic acid) using the hybridoma technique [22]. An scFv derivative of this parental monoclonal antibody was constructed and expressed with phages using a phagemid vector and its binding activity was characterized by hapten-immobilized ELISA [23]. Using this assay, the *s*-triazine atrazine can be detected in water samples with a detection limit of 1.5 µg/l. The use of this herbicide is prohibited in the European Economic Community (EEC) since 1991 and the acceptable limit set by the EEC Drinking Water Directive (DWD) in ground water is 0.1 µg/l. For a standard application in *s*-triazine analysis, this ELISA format would have to be further optimized. However, probably due to illegal use and its long persistence, concentrations exceeding the limit were detected recently in surface water in Portugal (2.7 µg/l) [24], in the Llobregat and the Ter river in northeastern Spain [25] and in ground and rain water in the Province of Limburg in the Netherlands (0.2 µg/l) [26]. The scFv K411B was shown to cross react with some other *s*-triazines, especially with propazine (~150%) and to some extent with cyanazine, deethylatrazine, simazine and terbuthylazine (~20%) (Figure A 3) [23]. The cross reactivity pattern was in accordance with the paternal antibody K4E7.



**Figure A 3.** Different *s*-triazines that bind to the scFv K411B. In C<sub>6</sub>-atrazine, hexanoic acid was introduced. Via this linker the hapten can be linked to amino groups of proteins, such as HRP for the tracer conjugate or BSA for the immobilization of the hapten.

## A 1.2 Expression of Recombinant Proteins in *Escherichia coli*

Systematically, the bacterium *Escherichia coli* belongs to the group of Gram-negative, facultative anaerobic rods. Under anaerobic conditions it obtains energy by fermentation and secretes organic acids with formic acid being the characteristic compound. As representative members of this group such as *E. coli*, *Salmonella* and *Shigella* live in the intestines (greek: *enteron*) they are also grouped in the family Enterobacteriaceae. This family is characterized by high mobility by peritrichous flagellation, presence of hemins and the ability to grow under aerobic as well as anaerobic conditions. Due to its long persistence also outside of the

intestines *E. coli* is used as an indicator organism for faecal contamination of drinking water [27].

Modesty with regard to nutrients, a generation time of only 20 minutes in complex medium at 37°C and the good genetic accessibility have rendered *E. coli* a popular organism for the expression of recombinant proteins, DNA amplification and cloning procedures [28]. However, many factors may limit the efficient expression of heterologous proteins in *E. coli*. Recombinant proteins can either be expressed in the cytoplasm or in the periplasm, the space between the inner and outer cell membrane, which is believed to favor disulfide bond formation due to its oxidizing environment. In *E. coli* efficient secretion of proteins into the periplasm can be accomplished via the Sec-dependent pathway. The N-terminal signal sequence of the nascent protein binds to the chaperone SecB, which binds to SecA. The whole complex binds to a channel formed by SecY and SecE and the protein is secreted into the periplasm [1]. Although the periplasm is analogous to the endoplasmic reticulum of eucaryotes to some extent—periplasmic chaperones and folding catalysts are thought to assist proper protein folding [29]—the compartmentation and its potential for posttranslational modifications is far less sophisticated than in eucaryotic cells. This can impose problems when expressing heterologous proteins in *E. coli*, especially when they require posttranslational modifications to exert their proper action.

In eucaryotic cells, posttranslational modifications like removal of leader sequences, glycosylation, chaperone- and folding-catalyst-assisted protein folding and assembly of subunits take place in certain cell compartments, the endoplasmic reticulum and the Golgi apparatus, which are separated from the cytoplasm by membranes. Proteins to be processed in this membrane system, have an N-terminal signal peptide similar to those in *E. coli*, which is recognized by the SRP (signal recognition particle) already during the biosynthesis of the protein by the ribosome. The SRP binds to a specific receptor in the endoplasmic reticulum and enables the secretion of the protein into the endoplasmic reticulum through a channel, which also functions as a ribosome receptor. This membrane system also provides an efficient pathway for secretion of proteins by exocytosis and integration of membrane proteins into the phospholipid bilayer. Also proteins synthesized in the cytoplasm may have signal peptides, which direct them to mitochondria, chloroplasts, peroxisomes or the nucleus, and which may need to be removed at the destination.

Some posttranslational modifications typical for many eucaryotic proteins, such as glycosylation cannot be performed in *E. coli*. Consequently, many attempts to express

eucaryotic proteins functionally in *E. coli* have failed and often toxic effects are observed in trying to do so. Antibodies consisting of four subunits, which are connected by disulfide bridges and have to be glycosylated are an example for such proteins. They must be produced in eucaryotic cells. On the other hand, engineered antibody fragments have been expressed functionally and in high yields in *E. coli*. Secretion into the medium does hardly occur in *E. coli*, so cytoplasmic or periplasmic extracts have to be prepared and the product must be isolated from these extracts. Additionally, the recombinant protein has to be refolded when it is expressed in an insoluble form as inclusion bodies. Anyway, as *E. coli* can be cultivated to cell dry masses of up to 150 grammes per liter in about two days and up to 30% of the total cell protein can consist of recombinant protein, it is a very convenient and effective host for the large scale production of recombinant proteins. For example Strandberg and Enfors obtained 19.2 g/l proteinA- $\beta$ -galactosidase fusion [30]. Martineau *et al.* reported on the functional expression of 3.1 g/l of a genetically engineered scFv [31]. Of course these are proteins which are expressed extremely well and the expression level of each protein depends on many factors, such as plasmid stability, the promoter system and transcription, translation, translocation inside the cell, protein stability and the behaviour of the expression strain under extreme conditions.

### **A 1.3 Expression of Recombinant Proteins *in vitro* by Cell-free Coupled Transcription/Translation**

Alternatively, proteins can be expressed *in vitro* by cell-free coupled transcription/translation. Therefore, the plasmid with the gene coding for the recombinant protein under control of the bacteriophage T7 promoter, T7 RNA polymerase, NTPs, a bacterial cell extract system, *E. coli*-tRNA and amino acids together with substances for selection and energy supply are mixed in one batch to exert transcription of the gene into mRNA as well as translation into the polypeptide [32-34]. The amount of expressed recombinant protein can be estimated by the incorporation of  $^{14}\text{C}$ -leucine during the cell-free translation, SDS-PAGE of the resulting mixture and autoradiographic evaluation [35, 36].

The productivities in such *in vitro* experiments are far below those of *in vivo* expressions. Even for proteins that are well expressed *in vitro*, such as the EGFP (30  $\mu\text{g/ml}$ ) [34], the calculated specific *in vitro* translation rate (incorporated amino acids/second) of only 0.1/s is far below the rate generally estimated for *in vivo* systems (10 to 20 amino acids/second) [37]. Consequently, such experiments are not useful for producing large



quantities of recombinant protein, but rather for testing mutants. Ryabova *et al.* have used cell-free translation to test the effect of sequence changes on recombinant antibody binding [38]. *In vitro* systems are also a means to express proteins, which are not or very weakly expressed *in vivo*. Merk *et al.* successfully expressed two scFvs, one with the original and one with the reversed arrangement of the variable domains of the light and heavy chains, *in vitro*, although in *E. coli* the scFv with the reversed arrangement could not be expressed functionally [16]. In both reports [16, 38] folding catalysts and chaperones were used to increase the amount of active protein.

## **A 1.4 Limiting Factors in the Production of Recombinant Proteins in *Escherichia coli***

Some eucaryotic proteins, which need sophisticated posttranslational modifications, which cannot be achieved in the more primitive procaryotic cell as shown in A 1.2, simply cannot be expressed in *E. coli*. However, for a number of recombinant proteins *E. coli* is a convenient expression system, but still many factors can impede the efficient expression of functional proteins. These can occur on the level of transcription, translation, translocation or protein folding. Additionally, the behaviour of the host cells under extreme conditions such as high-cell-density cultivation and downstream processing have to be taken into account. These items are discussed in the following.

### **A 1.4.1 Transcription**

Transcription comprises the transfer of the information encoded in the DNA sequence to the short-lived mRNA, which in turn transports this information to the ribosomes, where the actual protein synthesis takes place. The number of transcripts that reach the ribosomes depends on the number of DNA molecules carrying the respective gene, the transcription rate and the stability of the mRNA. The use of stable high-copy plasmids encoding the gene of interest and the coupling of translation directly to transcription in procaryotes reduces this problem to the transcription rate, which depends very much on the promoter system and the RNA polymerase. Promoters have also been optimized for the recombinant expression in *E. coli*. A very powerful expression system is the bacteriophage T7 promoter in combination with the T7 RNA polymerase [39]. It is so efficient that mRNA is synthesized much faster than it can be translated by the typical set of ribosomes in *E. coli*. Consequently, when using a stable high copy plasmid and a strong promoter transcription is not a limiting factor.

### A 1.4.2 Translation

Translation is a highly complex process, in the course of which the nucleotide sequence encoded in the mRNA is translated into an amino acid sequence, which determines the structure and function of the protein. Translation is performed by the ribosomes, which in procaryotes consist of three different rRNA species and fifty five proteins. Also aminoacyl-tRNAs and various translation factors are necessary for efficient translation. The gene and its upstream region need certain features to make translation possible. The start codon AUG coding for formylmethionine (in rare cases GUG for valine or UUG for leucine) initiates translation. Additionally, a purine-rich sequence, the Shine Dalgarno sequence, some nucleotides upstream of the start codon, which binds via hydrogen bonds to a complementary sequence at the 3'-end of the 16S-rRNA is necessary for efficient translation initiation [40]. This sequence varies among different species and its distance to the start codon, the start codon itself and also the first codons downstream of the start codon (+2 codon) determine the efficiency of translation initiation. Ringquist *et al.* found that a distance of eight nucleotides between the Shine Dalgarno sequence and the start codon and an AAA +2 codon, led to the highest expression levels in *E. coli* [41]. Cooperative effects of the initiation codon and its flanking regions [42] and the codon bias at the +2 codon [43] have been investigated in more detail. It has been assumed that a positive effect on gene expression by adenine at the first position of the +2 codon may be associated with the decoding tRNA and that evolution has favored codons at the +2 position that give high translation initiation [43]. If mRNA is present in excess, secondary structures which influence translation initiation may form [35].

In conclusion, translation can very easily become a bottle-neck for high-level expression of recombinant proteins. Additionally, it can be very difficult to find out determinants which are responsible for bad translation initiation, because cooperative effects occur and factors determining translation efficiency are not yet completely understood.

### A 1.4.3 Translocation

The mechanism of translocation of proteins across the cell membrane in *E. coli* mediated by leader sequences has been described above (A 1.2). The leader sequences of *ompA* and *pelB*, an outer membrane protein and a pectate lyase proved to be very efficient and are removed once the protein has entered the periplasm. Although leader sequences like these are necessary for a Sec-dependent translocation, they are not sufficient. Translocation in *E. coli* is incompletely understood [44]. Sec-independent translocation can also occur in *E. coli* [45].

The pros and cons of periplasmic or cytoplasmic expression of recombinant proteins in general and of antibody fragments in special in *E. coli* have been intensively discussed [29, 31]. On the one hand, secretion of scFv into the periplasm seems to be a very promising strategy, because the oxidizing environment allows the efficient formation of disulfide bonds. Cytoplasmatic expression frequently leads to the formation of inclusion bodies due to the aggregation of malfolded protein [46]. On the other hand, toxic effects and formation of periplasmic aggregates may counteract the high level expression of active protein in the periplasm [29].

If the translation level is too high, the efficiency of secretion can decrease, depending on the protein. Therefore, the optimal translational level has to be determined empirically for each heterologous protein. Simmons and Yansura observed that for the 14 kDa protein NT3 higher translation rates were favorable, while for the 8 kDa protein RANTES a lower translation level led to higher yields of secreted protein [47]. They suggest that too high translation levels might overwhelm the secretory apparatus for the translocation of proteins.

#### **A 1.4.4 Protein Folding**

Protein folding is not completely understood so far. It is generally assumed that the native state of a protein is determined by its amino acid sequence and corresponds to the global minimum in its energy landscape. However, a protein can be trapped in a misfolded state which corresponds to a local energy minimum. The main reason for the gigantic number of unfolded conformations is that often the energy differences between different rotamers (conformational isomers resulting from rotation of  $\sigma$ -bonds between heavy atoms) are small, and thus there are comparable occupancies of many different orientations of the protein backbone and sidechains [48]. Such misfolding events are more likely to occur when a protein is taken from its original environment, in which it was designed to fold correctly, to another environment with different chemical and physical properties, e.g. pH, viscosity or redox-potential.

Additionally, some proteins need helper proteins for correct folding even in the original organism without helper proteins. Helper proteins do not change the energy landscape of a protein but help the protein in finding its global energy minimum. They can be grouped into folding catalysts and molecular chaperones. The former increase the recovery yields of active proteins by accelerating rate-limiting steps along the folding pathway, such as disulfide and prolyl *cis-trans* isomerization. The latter appear to favor on-pathway folding by

transiently interacting with folding intermediates and suppressing their aggregation [49]. Many members of these classes have homologous counter-parts in eucaryotes and in procaryotes. For example the cytosolic *E. coli* thioredoxin and the endoplasmic PDI of eucaryotes both catalyze disulfide isomerization. Peptidyl prolyl *cis-trans* isomerases (PPIases) are present in the cytosol of *E. coli* and eucaryotic cells. Accordingly, eucaryotic heat shock protein 60 (Hsp60) corresponds to the *E. coli* heat shock protein GroEL and eucaryotic Hsp70 to *E. coli* DnaK, respectively [50, 51]. Although heat shock proteins are abundant proteins in unstressed cells, their transcription is transiently upregulated under stress conditions [49]. It has been reported that co-expression of molecular chaperones and folding catalysts can assist proper folding of recombinant proteins in *E. coli* [52].

As mentioned above, heterologous proteins can be expressed either in the periplasm or the cytoplasm. Due to the different physico-chemical properties and the different helper protein equipment of these compartments, in many cases the question whether translocation is favorable or not is directly coupled to the protein folding problem. In the cytoplasm of *E. coli*, correct folding may be achieved by co-expression of the *E. coli* chaperone teams DnaK-DnaJ-GrpE and GroEL-GroES [53]. The first is believed to stabilize proteins, also secretory proteins, in their unfolded state and therefore prevent premature folding, while the second forms multimeric barrel-shaped complexes, which can accommodate proteins from ~15 to 60 kDa in their interior and assist their proper folding [51]. Disulfide bonds do not form in the cytosol of *E. coli* because of reducing components like glutathione and thioredoxins. Bessette *et al.* genetically engineered *E. coli* strains in a way, that some of these components were rendered into oxidizing components and others into reducing disulfide isomerases. The result was a 20-fold increase in the yield of active proteins with multiple disulfide bonds in the cytoplasm, which was even higher than the yield of these proteins when directed to the oxidizing periplasm [54]. On the other hand, heterologous disulfide-bond-containing proteins fold correctly in the cytoplasm even without disulfide bonds. Tavladoraki *et al.* were able to show that a wild type scFv folded correctly in the reducing environment of the cytoplasm without the formation of disulfide bridges but with free sulfhydryl groups, as shown by a gel shift assay [55]. In other cases, scFvs have been engineered to fold correctly without the formation of disulfide bonds [31].

It is not clear, if periplasmic chaperones exist. The two proteins SecD and SKP might be involved in late translocation steps on the periplasmic side [29]. However, folding catalysts have been co-expressed successfully in the periplasm to improve correct folding. Due to the

oxidizing environment of the periplasm, formation of disulfide bonds is favored there and correct protein folding is assisted by protein disulfide isomerases (PDIs), which are potent in reducing wrong disulfide bonds and allowing them to form correctly. Accordingly, the overexpression of the *E. coli* disulfide isomerase has been reported to improve the folding of a T-cell receptor fragment in the periplasm [56]. Mammalian PDIs can also fulfil this task. Co-expression of rat PDI increased the yield of heterologous, secreted bovine pancreatic trypsin inhibitor [57] and co-expression of human PDI improved folding of an antibody Fab' fragment in the periplasm of *E. coli* [58].

The effects of PDIs and chaperones have also been investigated in the production of scFvs by cell-free translation [38]. PDI was also tested in folding of a Fab fragment *in vitro* [59] and co-expression of the *E. coli* SKP chaperone was shown to be beneficial for phage display of an scFv [60].

Improvement of the folding of heterologous proteins in the presence of chaperones and folding catalysts can occur, although in some cases with limited success. It has to be kept in mind, that many of these helper proteins need to be present at stoichiometric amounts. If they are co-expressed, this means that the maximally possible amount of the target protein decreases, because recombinant protein cannot exceed a certain percentage of the total cell protein. Using them in cell-free protein synthesis or for *in vitro* refolding results in high expenditure.

An alternative to the production of functional protein *in vivo* is refolding of insoluble recombinant protein *in vitro* after expression (A 1.4.6). This can even be advantageous, because the insoluble recombinant protein can be separated from soluble cell components prior to refolding.

### **A 1.4.5 Cultivation Strategies**

For the production of heterologous proteins in *E. coli* in larger scales it is preferable to use robust strains, which are fast-growing and can reach high cell densities. For example Korz *et al.* obtained a dry cell mass of 148 grammes per liter with the strain TG1 after 44 hours using an exponential feeding strategy [61]. Wilms *et al.* report on the cultivation of the strain BW3110 to 100 grammes per liter also with an exponential feed [62]. For such high-cell-density cultivations usually defined synthetic media are used. High-level expression of genes from the start of cultivation can lead to plasmid instability and growth inhibition by the gene products. This can be avoided by employing inducible expression systems. However, most of

the commonly used promoters in *E. coli* are negatively and not tightly regulated [62] In addition, due to the relatively high basal expression the impeding events mentioned above have to be faced. Especially if the expressed protein is toxic, the induction should not be done until high cell densities are reached and untight regulation is of a big hindrance. Wilms *et al.* circumvented this problem by using a tightly regulated expression system consisting of the positively regulated *E. coli rhaBAD* promoter, which is induced by L-rhamnose, and the rhamnose-deficient strain BW3110. High-level secretion of heterologous protein can interfere with the normal function of the Sec pathway causing cell toxicity [54]. Therefore, cytoplasmic expression is favorable and if periplasmic expression is necessary, induction should be done at a late point in the cultivation.

#### **A 1.4.6 Downstream Processing**

Downstream processing includes all steps involved in the processing of a recombinant protein following the actual production in the bioprocess. Such steps can be refolding, purification and chemical modification. Obviously, refolding is only necessary for proteins produced as insoluble aggregates. Efforts have been made to develop refolding protocols, but actually they have to be optimized more or less empirically for each protein [63, 64]. Refolding of insoluble scFv is also possible [46]. Often, the refolding procedure results in an enriched or purified protein. To what extent a protein has to be purified depends on its application. Technically applied proteins often do not require high purity. In biocatalytic applications frequently whole cell extracts or even whole cells are used [65]. For protein assays such as ELISA sometimes also crude cell extracts can be used, although the use of purified protein is preferable to diminish matrix effects. For proteins and peptides used in medical therapy the purity requirements are very high. Chemical modifications of recombinant proteins can be enzymatic fragmentations, e.g. to get rid of an affinity tag after purification or to yield Fab fragments from monoclonal antibodies. Monoclonal antibodies have also been modified by lipidation and carbohydrate amination resulting in increased organ uptake and decreased immunogenicity [66].

#### **A 1.5 Determination of the Amount of Total and Functional Recombinant Protein**

To determine the efficiency of an expression system for a certain heterologous protein, the amounts of total and functional recombinant protein need to be monitored. The relative

amounts of total recombinant protein, prepared from total cell extracts, can be quantified after separation of the contained proteins by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and subsequent densitometric evaluation. The concentration of total protein can be determined easily with commercial protein assay kits or densitometrically from the gel. Methods using purification procedures and subsequent protein determination or Western Blotting for quantitative analyses can be useful but require a certain optimization of the assay in order to obtain reproducible and reliable values.

For the quantification of active enzymes often assays are available, which use the conversion of a suitable substrate that results in a color reaction or a pH shift. For the determination of active scFvs the binding of the protein needs to be monitored. For this purpose enzyme-linked immunosorbent assays are suitable. Other possible methods are measurement of surface plasmon resonance or immobilization on and subsequent elution from affinity columns followed by protein determination. With ELISA, active protein can be captured very specifically from complex mixtures and due to the signal amplification by the enzyme reaction also small amounts of active scFv can be detected. Many different ELISA formats have been reported. For the detection of relatively large antigens, such as proteins, the most commonly used technique is the sandwich ELISA [67], where two different antibodies bind to two different epitopes of the antigen. In contrast, for the detection of haptens (small molecules which can act as epitopes but are incapable by themselves of eliciting an antibody response) such as *s*-triazine herbicides, competitive ELISAs have proven suitable, where the analyte competes with tracer or immobilized hapten. These assays involve the use of monoclonal antibodies [22, 68, 69], as well as the use of antibody fragments, such as Fab fragments [10] and single-chain fragments [23, 70]. Either the antibody or the antibody-fragment can be immobilized (A 1.5.1) or the hapten (A 1.5.2).

### **A 1.5.1 Antibody-immobilized Competitive ELISA**

In such competitive assay formats, the antibody or antibody fragment is immobilized via a capture antibody in microtiter plates. The analyte competes with a tracer, which consists of a conjugate of a derivative of the analyte and horseradish peroxidase (HRP), which converts tetramethylbenzidine (TMB) into a blue dye. The enzyme reaction is stopped upon the addition of acid and the blue dye is converted into a yellow one, which can be measured photometrically at 450 nm [22, 69].

### **A 1.5.2 Hapten-immobilized Competitive ELISA**

Alternatively, the hapten can be immobilized [23, 68-70]. In that case the immobilized hapten-ovalbumin or -BSA conjugate (Figure A 3) competes with the dissolved analyte for the antibody in solution. Bound antibody is then labeled with an anti-mouse-IgG-HRP conjugate and detection is carried out as described above. The use of a biotinylated antibody and a streptavidin-HRP conjugate is also possible [69]. Hapten-immobilized ELISAs are suitable for the quantification and characterization of antibodies or antibody fragments. In that case either no analyte or analyte at a constant concentration is added for competition. Blocking of the coated microtiter plates may be necessary to reduce unspecific binding. When tagged single-chain fragments are used, an additional incubation step involving an anti-Etag antibody [23] or anti-polyhistidine-tag antibody (this work) for labeling the tagged scFv might become necessary due to insufficient recognition of the scFv by anti-mouse-IgG.

### **A 1.6 Fluolabeling of Proteins by Fusion to Green Fluorescent Protein**

Many proteins have been fused to green fluorescent protein (GFP) or variants for different purposes. Waldo and coworkers fused GFP to the C-terminus of different proteins and found out that productive folding of the upstream protein domains expressed alone is correlated to the fluorescence of *E. coli* cells expressing the respective GFP fusions. Based on this observation, they used the fluorescent indicator of protein folding to evolve proteins that are normally prone to aggregation during expression in *E. coli* into correctly folding and functional proteins [71]. Fusions of recombinant protein to GFP have been used to quantitatively monitor the heterologous protein production [72] and protein solubility [73]. Additionally, GFP fusions have been used to monitor localization of the recombinant protein in *E. coli* cells and purification procedures [74, 75]. Fusions of scFvs to GFP are interesting tools for labeling antigens. Griep *et al.* have successfully constructed fusion proteins of scFvs directed against the lipopolysaccharide of the bacterium *Ralstonia solanacearum* and GFP to label these bacteria for flow cytometry and fluorescence microscopy [76]. Similar fusion proteins have been used to label E6 protein of a human papillomavirus [77] and hepatitis B surface antigen [20] in COS cells. Hink and coworkers studied the structural dynamics of a fusion protein of GFP and an scFv, which were separated from each other by a flexible hinge, in more detail. They concluded that both domains were allowed to move without mutual interference and that both fusion partners behaved independently, in a fashion identical to the



parent protein [78]. Arai *et al.* demonstrated, that fluorescence resonance energy transfer (FRET) can be observed, when two GFP variants that act as fluorescence donor and acceptor and have been fused to the light and the heavy chain of a Fv, respectively, come close upon association of the two chains resulting in the functional Fv [79].

## A 1.7 Aim of this Work

The expression of heterologous proteins in *E. coli* is often not achieved in the desired way. Either expression levels are low or the expression does not yield functional protein. Although one of the best known and most commonly used host organisms for the expression of recombinant proteins, in many cases it is not at all clear which factors impede the efficient heterologous expression in *E. coli*. The aim of this work was to identify such factors. This information could be useful in finding and eliminating bottle-necks in the protein production when using *E. coli*. If improvement of the existing system does not seem feasible it might help in selecting an alternative expression system.

Studies comparing *in vivo* and *in vitro* expression of scFvs have shown that certain scFvs can be well expressed *in vitro* but only in small quantities *in vivo* [16]. By comparing the expression levels and ratios of active and inactive recombinant protein in both systems, it should be possible to identify some factors which limit the functional expression of scFvs *in vivo*. Such drawbacks are due to either the intrinsic properties of the protein, the mRNA, the gene-encoding DNA, the physico-chemical properties of the environment or physiological effects. Some factors (transcription efficiency, translation efficiency and spontaneous protein folding) occur both *in vitro* and *in vivo*. Other factors, like translocational events or toxic effects only appear *in vivo*. Therefore, by comparing *in vitro* and *in vivo* expression it should be possible to assign impeding factors to either the former or the latter group.

For such a comparative approach, different constructs of the atrazine-specific scFv K411B [23] should be created and investigated. The expression of the scFv with an N-terminal pelB leader sequence for secretion of the scFv into the periplasm and the respective scFv without leader sequence should be compared in *E. coli* and *in vitro*. Additionally, the influence of a C-terminal fusion of the EGFP gene to the scFv should be investigated. For the determination of functional scFv and scFv-EGFP fusion protein suitable assays had to be developed. Constructs, that resulted in good expression levels, should be used for the production of the scFv and protocols should be developed in order to obtain pure and active protein.

## A 2 Materials and Methods

### A 2.1 Materials

If compounds are indicated in molarity, mass or volume percentage without any further explanation, this refers to their amount in aqueous solution.

#### A 2.1.1 Chemicals and Enzymes

Table A 2.1

Company	Chemicals and Enzymes
Amersham Pharmacia Biotech AB, Uppsala, Sweden	PD-10 Sephadex columns, LMW Electrophoresis Calibration Kit, Sephadex™ G-50 Superfine
Amresco, Solon, USA	PAGE-PLUS CONCENTRATE
BioRad Laboratories, Richmond, USA	BioRad protein assay kit, SDS-PAGE standard (low molecular weight)
Clontech, Palo Alto, USA	TALON™ resins
DIFCO-Laboratories, Detroit, USA	tryptone
Fluka Chemie, Buchs, Switzerland	agar, ampicillin (sodium salt), 6-aminohexanoic acid, ammonium chloride, ammonium hydroxide solution, ammonium peroxy disulfate, ammonium sulfate, L-arginine, boric acid, bromphenol blue, tert-butylmethylether, calcium chloride dihydrate, carbenicillin disodium salt, 4-chloro-1-naphthol, chloramphenicol, Coomassie Brilliant Blue R-250, cyanuric chloride, di-ammonium hydrogen citrate, di-ammonium hydrogen phosphate, 1,2-dimethoxyethane, dimethylsulfoxide, dioxane, di-potassium hydrogen phosphate, di-sodium hydrogen phosphate dihydrate, ethanol, ethanolamine, ethidiumbromide, ethylacetate, ethylenediaminetetraacetic acid (EDTA), ferric chloride hexahydrate, formamide, formic acid, D (+) –glucose monohydrate, L-gluthathion oxidized, glycine, hydrochloric acid, imidazole, kanamycin sulfate, lysozyme from hen egg white, 2-mercaptoethanol, magnesium chloride hexahydrate, magnesium sulfate anhydrous, magnesium sulfate heptahydrate, potassium carbonate, silica gel,

	sodium acetate anhydrous, sodium carbonate, sodium chloride, sodium dihydrogen phosphate monohydrate, sodium dithionite, sodium hydrogen carbonate, sodium hydroxide, sodium sulfate anhydrous, sodium sulfate decahydrate, D(+)-sucrose, sulfuric acid, N,N,N',N'-tetramethylethylene diamine, toluol, trichloroacetic acid, yeast extract, zinc sulfate heptahydrate
Gibco BRL GmbH, Eggenstein, Germany	agarose, standard for agarose gels (1 kb, 1 kb plus ladder), urea
Kodak, New Haven, USA	ammonium persulfate
Merck, Darmstadt, Germany	1,4-dithiothreitol (DTT), cobaltous(II) chloride hexahydrate, manganese(II) sulfate monohydrate
Pierce, Rockford, USA	BCA <sup>TM</sup> Protein Assay Kit
QIAGEN, Hilden, Germany	QIAquick Gel Extraction Kit QIAprep Spin Miniprep Kit QIAGEN Plasmid Midi Kit
Riedel-de-Haen, Seelze, Germany	acetic acid, acetone, atrazine, cyanazine, deethylatrazine, glycerol, n-hexane, methanol, propazine, simazine, terbutylatrazine, tris-(hydroxymethyl)-aminomethane, Tween 20
Roche, Mannheim, Germany	complete <sup>TM</sup> , T4 DNA ligase, restriction endonucleases
Roth, Karlsruhe, Germany	ampicillin disodium salt, Rotiphorese® Gel 30 (30% (m/v) acrylamide in water with 0.8% (m/v) bisacrylamide)
Sigma-Aldrich, Taufkirchen, Germany	albumine bovine fraction V, 2-aminoethanol, casein, citric acid dihydrate, copper(II) sulfate pentahydrate, N, N'-dicyclohexylcarbodiimide (DCC), 1,4- dithioerythritol (DTE), goat-anti-mouse-IgG HRP conjugate, guanidine hydrochloride, hydrogen peroxide (30% solution in water), N-hydroxysuccinimide (NHS), 3,3',5,5'-tetramethylbenzidine (TMB), thiamine hydrochloride
Southern Biotechnology Associates, Inc. Birmingham, USA	FluoromountG embedding medium
Stratagene, Heidelberg, Germany	QuikChange Site-Directed Mutagenesis Kit

## A 2.1.2 Instruments

**Table A 2.2**

<b>Purpose</b>	<b>Instrument</b>	<b>Company</b>
absorbance measurement	Ultrospec 3000 UV/Vis	Amersham Pharmacia Biotech AB, Freiburg, Germany
agarose gel electrophoresis	DNA Sub Cell™, Mini Sub DNA Cell™, Mini Sub Cell GT™, Power Pac 300	BioRad, München, Germany
	Video Copy Processor P66E	Mitsubishi, Cambridge, USA
	BWM 9X Monitor	Javelin Electronics, Schaumburg, USA
	UV-lamp table	MWG-Biotech, Ebersberg, Germany
centrifugation	Centrifuges 5417 C, 5417 R, 5810 R Universal 30 F	Eppendorf, Hamburg, Germany
	Sorvall® RC-5B (rotors: SLA 3000, SA 600)	Hettich, Tuttlingen, Germany Du Pont Instruments, Leipzig, Germany
DNA concentration	Concentrator 5301	Eppendorf, Hamburg, Germany
DNA sequencing	377 DNA Sequencer	Applied Biosystems, Weiterstadt, Germany
fed-batch cultivation	Labfors 3 l bioreactor	Infors AG, Bottmingen, Switzerland
	3 l bioreactor 30 l bioreactor	Bioengineering, Wald, Switzerland
incubation	WTE incubators shaking incubators	Binder, Tuttlingen, Germany Infors AG, Bottmingen, Switzerland
	Thermomixer 5436/Comfort Heidolph REAX 3 orbital shaker water bath B3, DC 10 cryostat RMS, RM6	Eppendorf, Hamburg, Germany Heidolph, Kelheim, Germany Haake, Karlsruhe, Germany MGW Lauda, Hannoversch-Münden, Germany
microscopy	Leica DMIRBE microscope	Leica Microsystems, Wetzlar, Germany
	HCX PL Apo 63x/1.32-0.6 oil immersion objective Polychrome 2	TILL Photonics, München, Germany

	dichroic mirror: F53-009 bandpass filter 535/50 nm Orca C4742-95 camera  Openlab 2	AHF Analysentechnik, Tübingen, Germany Hamamatsu Photonics K.K., Hamamatsu, Japan Improvision, Coventry, England
microtiter plate reading	FLUOstar	BMG LabTechnologies, Offenburg, Germany
microtiter plate washing	Titertek Microplate Washer M96V Flow, Titertek Microplate Stacker	ICN, Costa Mesa, USA
PAGE	Minigel-Twin G42 Power Pac 3000, Power Pac 300Model 583 Gel Dryer	Biometra, Göttingen, Germany BioRad, München, Germany
PCR	Robocycle gradient 40  Mastercycler Gradient	Stratagene, Amsterdam, The Netherlands Eppendorf, Hamburg, Germany
pH measurement	Digital pH Meter pH500	WTW, Weilheim, Germany
pipetting	0.5-10, 2-20, 10-100, 100-1000, 500-2500, 500-5000 µl pipets, 30-300 µl multi-channel pipet, multi-stepper pipet	Eppendorf, Hamburg, Germany
protein blotting	Trans-Blot SD, semidry transfer cell	BioRad, München, Germany
protein sequencing	491 Procise Protein Sequencer 785 Programmable Absorbance Detector 140 C Microgradient System	Applied Biosystems, Weiterstadt, Germany
rotary evaporator	Rotavapor R-134, Waterbath B-480	Büchi, Flawil, Switzerland
scales	Basic MC1 Research RC 210 D	Sartorius, Göttingen, Germany
sonification	Sonifier W-250	Branson, Dietzenbach, Germany
ultrafiltration	amicon 8050, 80200	Millipore, Bedford, USA

### A 2.1.3 Consumables

**Table A 2.3**

Consumables	Company
pipet tips, 15, 50 ml reaction tubes	Greiner, Nuertingen, Germany
pipet tips, 0.5, 1.5, 2.0 ml reaction tubes	Eppendorf, Hamburg, Germany
sterile filters	Millipore, Molsheim, France
ultrafiltration membranes	Millipore, Bedford, USA
filter paper	BioRad, Hercules, USA
Sequi-Blot™ PVDF membrane	
Zeta-Probe® GT Genomic Tested Blotting Membrane	
microtiter plates MaxiSorp	Nunc, Roskilde, Denmark
microtiter plates FluoroNunc MaxiSorp	

### A 2.1.4 Microorganisms

**Table A 2.4**

Organism	Genotype/Features	Reference/Company
<i>Escherichia coli</i> DH5 $\alpha$	<i>deoR</i> , <i>endA1</i> , <i>gyrA96</i> , <i>hsdR17</i> ( $r_k$ - $m_k$ +), <i>recA1</i> , <i>relA1</i> , <i>supE44</i> , <i>thi-1</i> , $\Delta$ ( <i>lacZYA</i> - <i>argFV169</i> ), $\Phi$ 80 <i>lacZ</i> $\Delta$ M15, F	[80] Clontech, Heidelberg, Germany
<i>Escherichia coli</i> BL21(DE3)	B, F <sup>-</sup> , <i>dcm</i> , <i>ompT</i> , <i>hsdS</i> ( $r_B$ - $m_B$ -), <i>gal</i> (DE3)	Novagen, Madison, USA
<i>Escherichia coli</i> BL21(DE3) pLysS	B, F <sup>-</sup> , <i>dcm</i> , <i>ompT</i> , <i>hsdS</i> ( $r_B$ - $m_B$ -), <i>gal</i> (DE3), [pLysS Camr]a]	Novagen, Madison, USA
<i>Escherichia coli</i> BW3110	Derivative of <i>E. coli</i> W3110 [81] with deficiency in L-rhamnose metabolism ( <i>rhaB</i> )	[62]

## A 2.1.5 Plasmids

**Table A 2.5**

Plasmid	Promoter	Resistance	Features	Reference/Company
pETK(+)	T7	Amp <sup>R</sup>	pelB leader, scFv K411B gene, His <sub>6</sub> -tag,	[82] [83]
pEGFP	<i>lac</i>	Amp <sup>R</sup>	EGFP gene	Clontech, Palo Alto, USA
pTST101	T7, <i>rhaBAD</i>	Amp <sup>R</sup>	MalE-GFP gene	[84]

## A 2.1.6 Synthetic Oligonucleotides

### A 2.1.6.1 Primers for DNA Sequencing

**Table A 2.6**

Name	Binding Site	Sequence
T7_pp_f	T7 promoter primer	5'-TAATACGACTCACTATAGG-3'
T7_pp_r	T7 terminator primer	5'-GCTAGTTATTGCTCAGCG-3'
C-Fusi	~100 bp upstream of 3'-end of K411B-scFv gene	5'-CCCTCACCATTAATCCTGTGG-3'

### A 2.1.6.2 Primers for Introduction of a His<sub>6</sub>-tag at the C-terminus of the scFv-EGFP fusion protein

**Table A 2.7**

Name	Binding Site / Feature	Sequence
Sall F	<i>Sall</i> restriction site in scFv- EGFP gene	5'-GGAATTCGTCGACATGGTGA-3'
NotI R	3'-terminus of EGFP gene / sequence coding for His <sub>6</sub> -tag	5'-GATACGCGCCGCGTGGTGGTGGTG GTGGTGCTTGTACAGCTCGTCCATG-3'

## A 2.1.6.3 Primers for QuikChange PCRs

Table A 2.8.

Name	Purpose	Sequence (mutations in bold, insertions underlined, deletions:   )
Stop-codon F	integration of stop-codon 3' of the scFv-EGFP gene	5'-CACCACCACT <b>TGAG</b> CGGCCGC-3'
pet1 F	removal of T 5' of the His <sub>6</sub> -tag in the scFv-EGFP gene	5'-GACGAGCTGTACAAG   AAAGCGGCCGCACTC-3'
Shiny F	removal of AT 5' of the start codon in pJOEK(+)	5'-GAAGGAGATATACAT   ATGAAATACCTGCTGC-3'
Dalgy F	removal of AC 5' of the start codon in pETK(-) and pETKE(-)	5'-GAAGGAGATATACAT   ATGGTCCAGGTGAAAC-3'
negACive F	insertion of AC 5' of the start codon in pETK(+)	5'-GAAGGAGATATACAT <b>AC</b> ATGAAATACCTGCTGC-3'
Lys1 F	insertion of AAA 3' of the start codon in pJOEK(-)-AT	5'-GGAGATATACATATG <b>AAA</b> GTCCATGTGAAACTG-3'
Lys2 F	insertion of AAA 3' of the start codon in pJOEK(-)-AT+AAA	5'-GATATACATATGAAA <b>AAA</b> GTCCAGGTGAAACTG-3'

*Note.* Only forward primers are shown. Reverse primers were indicated with „R“ instead of „F“ and were complementary to the respective forward primers.



## A 2.2 Methods

### A 2.2.1 Molecular-Genetic Methods

#### A 2.2.1.1 Isolation of Plasmid DNA from *Escherichia coli* with QIAprep Spin Miniprep Kit and QIAGEN Plasmid Midi Kit

Plasmid DNA prepared using these kits is suitable for restriction enzyme digestion, ligation, transformation and sequencing. They are based on the lysis of cells under basic conditions [85] and utilize silica-gel membrane technology for purification.

After recovering bacteria from overnight cultures by centrifugation, the plasmid isolation was done according to the instructions of the manufacturer.

#### A 2.2.1.2 Extraction and Purification of Plasmid DNA from *Escherichia coli* for Fast Tests

For purposes such as analytical restriction enzyme digestions, where purity and yield were not crucial, a quick method based on the alkaline lysis [85] was used.

Bacteria were recovered from 2 ml of overnight culture by centrifugation for 1 min at 20,800 g and resuspended in 200  $\mu$ l of ice-cold solution 1 (100 mM Tris-HCl, pH 7.5, 10 mM EDTA, 400  $\mu$ g/ml RNase I) by vigorous vortexing. Cells were then lysed by the addition of 200  $\mu$ l freshly prepared solution 2 (1 M NaOH, 5.3% (m/v) SDS) and vortexing. After addition of 200  $\mu$ l ice-cold solution 3 (3 M  $K^+$ , 5 M  $Ac^-$ ) and vortexing, precipitated proteins and genomic DNA were removed by centrifugation (5 min, 20,800 g, 4°C). The plasmid DNA in the supernatant was precipitated with isopropanol, dried and then resuspended in 20  $\mu$ l TE buffer (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA).

#### A 2.2.1.3 Precipitation of Plasmid DNA with Ethanol

One volume of DNA solution was mixed with 1/10 volume of 3 M NaAc, pH 4.8 and 2.5 volumes of ethanol (100%, -20°C), vortexed and kept at -20 or -70°C for 20 min to allow precipitation. After centrifugation (20,800 g, 30 min 4°C) and removal of the supernatant, the pellet was washed with 70% ethanol, dried and resuspended in TE buffer.

**A 2.2.1.4 Precipitation of Plasmid DNA with Isopropanol**

One volume of DNA solution was mixed with 0.7 volumes of isopropanol, incubated at RT for 15 min and centrifuged (20,800 g, 15 min, RT). The supernatant was removed, the pellet washed with 70% ethanol, dried and resuspended in TE buffer.

**A 2.2.1.5 Agarose Gel Electrophoresis of DNA**

Agarose gel electrophoresis was carried out as described [86]. 1% (m/v) agarose gels were prepared by resuspending the correct amount of powdered agarose in 1x TAE buffer (40 mM Tris-HCl, 20 mM acetic acid, 2 mM EDTA, pH 8.3), heating in a microwave until the agarose dissolved, adding ethidium bromide to a final concentration of 0.5 µg/ml and pouring the warm agarose solution into the plastic tray supplied with the electrophoresis apparatus after sealing the open ends with plastic tape and inserting a plastic comb. After the gel was completely set, the comb and the plastic tape were carefully removed and the gel mounted in the electrophoresis tank, which was filled with 1x TAE buffer in a way, that the gel was covered to a depth of about a mm. The samples of DNA were mixed with 6x DNA loading buffer (30% (v/v) glycerine, 0.2% (m/v) BPB, 25 mM EDTA, pH 7.5) and loaded into the slots of the submerged gel. As a standard 7.5 µl of kb ladder (Gibco, prepared as suggested by the manufacturer) were loaded. After running the gel for 20 to 30 min at 120 V, the gel was examined by ultraviolet light and photographed for documentation. To isolate DNA by gel extraction, DNA-containing bands were excised.

**A 2.2.1.6 Isolation of DNA from Agarose Gels with the QIAquick Gel Extraction Kit**

The extraction of DNA from agarose gels was carried out according to the instructions of the manufacturer. The purification step is based on the specific binding of the DNA to a silica gel matrix.

**A 2.2.1.7 Enzymatic DNA modifications**

Restriction enzymes bind specifically to and cleave double-stranded DNA at specific sites within or adjacent to a particular sequence known as the recognition sequence [86]. DNA digestions with restriction enzymes were performed under the reaction conditions specific for each enzyme.

To avoid religation of vectors cut only with one enzyme, the 5'-phosphate groups were removed by treatment with CIAP. To transform sticky ends into blunt ends, either the 5'-

overhang was filled in with dNTPs and the Klenow fragment of DNA polymerase I, or single-stranded overhangs of DNA were degraded with Mung Bean Nuclease.

Ligations were performed with T4-DNA ligase at RT for one hour.

All enzymatic DNA modifications were performed according to the manufacturers' recommendations. If necessary, enzymes were inactivated by heat, buffer ingredients were removed by dialysis or the desired DNA fragments were purified by agarose gel electrophoresis followed by gel extraction.

### A 2.2.1.8 Automated DNA Sequencing

Automated DNA sequencing based on the enzymatic chain terminator technique, developed by Sanger *et al.* [87] was done using a 377 DNA Sequencer (Applied Biosystems) and a Terminator Ready Reaction Mix with Ampli-Taq-Fs (Perkin Elmer). Sequencing reactions were performed with four different fluorescent labels identifying the four ddNTPs, instead of the usual radioactive labels. These fluorophors were excited with two argon lasers at 488 and 514 nm, respectively, when the respective bands passed the lasers during the electrophoresis. The specific emissions were detected and the data collected for analysis [88, 89]. The thermal cycling mixture was composed as follows:

Terminator Ready Reaction Mix	4 $\mu$ l
sequencing primer	20 pmol
DNA	300–500 ng
ddH <sub>2</sub> O	ad 20 $\mu$ l

**Table A 2.9.** Thermal cycling conditions for the sequencing reactions.

Step	Denaturation	Annealing	Extension	Cycle Number
1	95°C, 4 min			
2	95°C, 40 s	55°C, 30 s	60°C, 4 min	25
3			72°C, 4 min	1

The products were purified by ion-exchange chromatography using Sephadex<sup>TM</sup> G-50 Superfine (Amersham Pharmacia) in microtiter plates according to the instructions of the manufacturer. The eluates were dried at 80°C, resuspended in 3  $\mu$ l of loading buffer (80% (v/v) 25 mM EDTA, pH 8.0, 20% (v/v) formamide), denatured for 2 min at 95°C and cooled

to RT. 0.7  $\mu$ l of that mixture were loaded onto the gel and the electrophoresis was performed as suggested by the manufacturer. A 5.25% (m/v) polyacrylamide gel was used, consisting of

urea	18.0 g
PAGE-PLUS CONCENTRATE (Amresco)	6.6 ml
10x TBE buffer (0.9 M Tris, 0.9 M boric acid, 20 mM EDTA)	5.0 ml
ddH <sub>2</sub> O	ad 50 ml
10% (m/v) APS solution	250 $\mu$ l
TEMED	25 $\mu$ l

### A 2.2.1.9 Polymerase Chain Reaction (PCR)

PCRs were carried out in order to amplify specific DNA fragments [90]. Standard PCRs were pipetted in 0.5 ml PCR reaction tubes as follows:

10x polymerase buffer	5 $\mu$ l
dNTP-mix (each 2.5 mM)	4 $\mu$ l
forward primer	100 pmol
reverse primer	100 pmol
template DNA	$\pm$ 100 ng
Taq DNA-polymerase	1-2 U
25 mM MgCl <sub>2</sub>	4 $\mu$ l
ad ddH <sub>2</sub> O	50 $\mu$ l

**Table A 2.10.** PCR conditions.

Step	Denaturation	Annealing	Extension	Cycle Number
1	95°C, 4 min			
2	95°C, 1 min	53°C, 1:30 min	72°C, 1:30 min	25
3			72°C, 8 min	1

### A 2.2.1.10 QuikChange PCR for Site-directed Mutagenesis

Point mutations were introduced into the genes using the QuikChange Site-Directed Mutagenesis Kit (Stratagene), according to the manufacturer's instructions. Depending on the sequences of the primers, the annealing temperature was varied between 50 and 65°C.

Following the PCR, the template DNA was digested with *DpnI*. 15 µl of the resulting solution were dialyzed against ddH<sub>2</sub>O for 15 min and transformed into *E. coli* DH5α.

## A 2.2.2 Microbiological Methods

### A 2.2.2.1 Growth Media for *Escherichia coli*

#### A 2.2.2.1.1 *LB Medium [91]*

tryptone	10 g/l
yeast extract	5 g/l
NaCl	5 g/l
agar (optional for agar plates)	15 g/l

If required, 100 mg/l ampicillin or carbenicillin or 34 mg/l chloramphenicol (in ethanol) were added. Stock solutions (100 mg/ml and 34 mg/ml) were filter sterilized.

#### A 2.2.2.1.2 *Synthetic Medium [62] for shake flask and fed-batch cultivation*

To prepare the salt solution for 1 l of synthetic medium, the following salts were dissolved in 900 ml dH<sub>2</sub>O and autoclaved.

Na <sub>2</sub> SO <sub>4</sub>	2.0 g/l
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	2.68 g/l
NH <sub>4</sub> Cl	0.5 g/l
K <sub>2</sub> HPO <sub>4</sub>	14.6 g/l
NaH <sub>2</sub> PO <sub>4</sub> *H <sub>2</sub> O	3.6 g/l
(NH <sub>4</sub> ) <sub>2</sub> -H-citrate	1.0 g/l

For shake flask cultivations 100 ml 10% (m/v) glucose solution, 2 ml 1 M MgSO<sub>4</sub> solution, 3 ml trace element solution (each autoclaved separately) and 1 ml 1% (m/v) thiamine solution (filter sterilized) were added. The batch media of the 2 l and the 15 l fed-batch cultivations contained 1.5 and 2.5% (m/v) glucose, respectively. For selection, antibiotics were added as described above. One liter trace element solution (TES) contained the following compounds:

---

CaCl <sub>2</sub> *2H <sub>2</sub> O	0.5 g/l
ZnSO <sub>4</sub> *7H <sub>2</sub> O	0.18 g/l
MnSO <sub>4</sub> *H <sub>2</sub> O	0.1 g/l
Na <sub>2</sub> EDTA	20.1 g/l
FeCl <sub>3</sub> *6H <sub>2</sub> O	16.7 g/l
CuSO <sub>4</sub> *5H <sub>2</sub> O	0.16 g/l
CoCl <sub>2</sub> *6H <sub>2</sub> O	0.18 g/l

For feeding of the cells in 2 l fed-batch cultivation, salts for 800 ml of synthetic medium (4/5 of the amounts listed above) were dissolved in 600 ml dH<sub>2</sub>O, autoclaved and mixed with an autoclaved solution of 74.8 g glucose monohydrate (68 g glucose) in 200 ml dH<sub>2</sub>O.

The feeding solution for 15 l fed-batch cultivation was partitioned into two batches to avoid precipitation of the ingredients:

Feed 1: 2750 g glucose monohydrate (2500 g glucose) were resuspended in 3.5 l dH<sub>2</sub>O and dissolved by autoclaving. 98.5 g MgSO<sub>4</sub>\*7H<sub>2</sub>O were dissolved in 150 ml dH<sub>2</sub>O and autoclaved. 2.5 g thiamine were dissolved in 50 ml dH<sub>2</sub>O and filter sterilized. All solutions were mixed with 500 ml TES in a 5 l feeding flask.

Feed 2: 396 g (NH<sub>4</sub>)<sub>2</sub>HPO<sub>4</sub> were dissolved in 1 l dH<sub>2</sub>O, adjusted to pH 7.0 with concentrated phosphoric acid and autoclaved.

#### **A 2.2.2.2 Transformation of *Escherichia coli* by Heat Shock**

To produce competent cells, 50 ml of LB medium were inoculated 1:100 with an overnight culture and incubated with shaking at 37°C. The cells were collected at OD<sub>578</sub> of 0.3-0.7 by centrifugation (10 min, 3,020 g, 4°C), resuspended in 2 ml TSS (10% (m/v) PEG 6000, 5% (v/v) DMSO, 50 mM MgSO<sub>4</sub> in LB medium) and incubated on ice for 5 min. 200 µl of cell suspension were mixed with at least 5 ng plasmid DNA and incubated for 20 min on ice. The cells were subjected to heat shock at 42°C for 45 sec, followed by the addition of 800 µl LB medium and incubated for one hour with shaking at 37°C, before the cells were spread on selective agar plates and incubated overnight at 37°C.

### **A 2.2.2.3 Expression of Recombinant Proteins in *Escherichia coli***

Recombinant proteins were either expressed under control of the T7 promoter in *E. coli* BL21(DE3) or *E. coli* BL21(DE3) pLysS or under control of the *rhaBAD* promoter in *E. coli* BW3110.

#### **A 2.2.2.3.1 Expression under Control of the T7 Promoter**

The bacteriophage T7 RNA polymerase/promoter system is very efficient for the overexpression of genes in *E. coli* [92]. The T7 promoter is derived from the bacteriophage T7 and has a strong transcription initiation signal and a highly efficient ribosome binding site from the T7 major capsid protein. For expression of the target DNA, which is cloned downstream of the T7 promoter, a source of the very selective and active T7 RNA polymerase, must be provided by the host cell [39]. In host cells like *E. coli* BL21(DE3) and BL(DE3) pLysS the T7 RNA polymerase gene is encoded in the genomic DNA under control of the *lacUV5* promoter and its expression can be induced with IPTG [93]. The latter strain has an additional plasmid encoding a T7 lysozyme and conferring chloramphenicol resistance, which inactivates T7 RNA polymerase in small amounts to prevent background expression. The scFv K411B gene was cloned into pET20b(+), a vector carrying the T7 promoter, a pelB leader sequence for periplasmic expression and a poly-histidine tag for easy purification, resulting in the plasmid pETK(+) [82].

##### **A 2.2.2.3.1.1 Expression in shake flasks**

50 or 100 ml of LB or synthetic medium were inoculated 1:100 with overnight cultures of *E. coli* BL21(DE3) or BL21(DE3) pLysS transformed with the pET20b(+) derivative for expression of the heterologous gene and incubated at 30°C with shaking. BL21(DE3) with plasmids were selected by carbenicillin (100 mg/l) and BL21(DE3) pLysS additionally with chloramphenicol (34 mg/l). Induction was done with 1 mM IPTG at an OD<sub>578</sub> of 1. After 2-4 hours cells were harvested by centrifugation.

##### **A 2.2.2.3.1.2 Expression in 2 l fed-batch cultivation**

Fed-batch cultivation of *E. coli* BL21(DE3) pLysS containing the respective plasmid, was performed in a 3 l bioreactor (Bioengineering) at 30°C. After inoculation of 1.2 l of synthetic medium containing carbenicillin and chloramphenicol with 20 ml of a shake flask culture of

$OD_{578}=5.7$  ( $c_x=1.17$  g/l dry biomass) and cultivation of the cells to an  $OD_{578}$  of 18, 0.8 of feed medium as described in 2.1.1.2 were added at an exponential feed rate of  $0.1\text{ h}^{-1}$ , resulting in a final cultivation volume of about 1.6 l (sample taking and evaporation regarded). During the cultivation, pH and  $pO_2$  were monitored via electrodes. The pH was adjusted automatically to 7.0 by the addition of 38% ammonium hydroxide solution and the  $pO_2$  was kept between 50 and 90% by regulating the stirrer speed and the air flow. Induction was done with 1 mM IPTG at  $OD_{578}=17$ .

#### **A 2.2.2.3.2 Expression under Control of the rhaBAD Promoter [62]**

The *rhaBAD* promoter is a very tightly regulated *E. coli* promoter. However, in wild type strains big amounts of expensive L-rhamnose are necessary due to its metabolization. Wilms and coworkers inhibited the consumption by inactivating the *rhaB* gene of *E. coli* W3110, responsible for the first irreversible step in rhamnose catabolism, resulting in the strain BW3110. The vector pJOE3078 [84] was used for expression. Upstream of the *rhaBAD* promoter, there is a T7 promoter also allowing expression with T7 RNA polymerase/promoter systems.

##### **A 2.2.2.3.2.1 Expression in shake flasks**

Shake flask expressions using *E. coli* BW3110 containing the pJOE 3078 derivative were performed as described for the T7 expression system except that cells were induced with 0.2% (m/v) L-rhamnose for 6-8 hours.

##### **A 2.2.2.3.2.2 Expression in 13 l fed-batch cultivation**

Fed-batch cultivation of *E. coli* BW3110/pJOEK(+)-AC, was performed in a 30 l bioreactor (Bioengineering) at 30°C in synthetic medium as described previously [62]. After inoculation of 8 l of synthetic medium (A 2.2.2.1.2) with 400 ml of a shake flask culture of  $OD_{578}=3.7$  ( $c_x=1.11$  g/l dry biomass) and cultivation of cells to an  $OD_{578}$  of 25, 4.2 l of feed 1 and 1 l of feed 2 (A 2.2.2.1.2) were added at an exponential feed rate of  $0.1\text{ h}^{-1}$ , resulting in a final cultivation volume of about 12 l (sample taking and evaporation regarded). pH and  $pO_2$  were regulated as described for the 2 l cultivation. Induction was done with 0.2% (m/v) L-rhamnose at  $OD_{578}=140$ .



### A 2.2.3 Organic Synthesis

#### A 2.2.3.1 Synthesis of 2,4-Dichloro-6-ethylamino-s-triazine

10 g cyanuric chloride (54 mmol) were dissolved in 200 ml 1,2-dimethoxyethane and cooled down to  $-10^{\circ}\text{C}$ . 7.2 ml 70% (v/v) ethylamine (110 mmol) were mixed with 50 ml 1,2-dimethoxyethane, cooled to  $4^{\circ}\text{C}$  and added dropwise with vigorous agitation while the temperature was kept below  $0^{\circ}\text{C}$ . After 1 hour, the cold mixture was filtrated and the precipitate washed with 1,2-dimethoxyethane. The filtrate was concentrated in a rotary evaporator and purified by column chromatography (silica gel 60, n-hexane/ethylacetate/acetic acid (85:13:2)).

yield of product (M=181.06): 23.3% (2.28 g)

melting point:  $102-104^{\circ}\text{C}$ ,  $R_f$  with toluol/acetone (85:15): 0.65

$^1\text{H-NMR}$ :  $\delta=1.26$  (t, 3H,  $\text{CH}_3\text{-CH}_2\text{-NH-}$ ), 3.55 (m, 2H,  $\text{CH}_3\text{-CH}_2\text{-NH-}$ ), 6.47 (m, 1H,  $\text{CH}_3\text{-CH}_2\text{-NH-}$ )

$^{13}\text{C-NMR}$ :  $\delta=14.38$  ( $\text{CH}_3\text{-CH}_2\text{-NH-}$ ), 36.57 ( $\text{CH}_3\text{-CH}_2\text{-NH-}$ ), 165.03-171.03 (C of s-triazine ring)

#### A 2.2.3.2 Synthesis of $\text{C}_6$ -Simazine (4-Chloro-6-(iso-ethylamino)-s-triazine-2-6-aminohexane-carboxylic acid) [94]

1.81 g 2,4-dichloro-6-ethylamino-s-triazine (10 mmol) were dissolved in 20 ml tert-butylmethylether. 2.62 g 6-aminohexanoic acid (20 mmol) and 5.52 g potassium carbonate (40 mmol) were dissolved in 20 ml  $\text{dH}_2\text{O}$  and added to the organic solution with vigorous agitation. After 6 hours of reaction time at  $30^{\circ}\text{C}$  the phases were separated. The aqueous phase was acidified carefully with formic acid and the precipitate was filtrated, washed with  $\text{dH}_2\text{O}$ , suspended in 40 ml  $\text{dH}_2\text{O}$  and resolved with 1 M sodium hydroxide. The solution was filtrated and acidified with formic acid, the precipitate was washed with  $\text{dH}_2\text{O}$  and air-dried.

Yield of product (M=287.8): 15.7% (453 mg)

melting point:  $141-143^{\circ}\text{C}$ .

$^1\text{H-NMR}$ :  $\delta=1.18$  (t, 3H,  $\text{CH}_3\text{-CH}_2\text{-NH-}$ ), 1.39 (m, 2H,  $\text{-NH-(CH}_2\text{)}_2\text{-CH}_2\text{-(CH}_2\text{)}_2\text{-COOH}$ ), 1.63 (m, 4H,  $\text{-NH-CH}_2\text{-CH}_2\text{-CH}_2\text{-CH}_2\text{-COOH}$ ), 2.67 (t, 2H,  $\text{-NH-(CH}_2\text{)}_4\text{-}$ )

$\text{CH}_2\text{-COOH}$ ) 3.38 (m, 4H,  $\text{CH}_3\text{-CH}_2\text{-NH-}$  and  $\text{-NH-CH}_2\text{-(CH}_2\text{)}_4\text{-COOH}$ ), 6.42-6.73 (m, 2H,  $\text{CH}_3\text{-CH}_2\text{-NH-}$  and  $\text{-NH-(CH}_2\text{)}_5\text{-COOH}$ )  
 $^{13}\text{C-NMR}$ :  $\delta$ =14.46 ( $\text{CH}_3\text{-CH}_2\text{-NH-}$ ), 24.43 ( $\text{-NH-(CH}_2\text{)}_2\text{-CH}_2\text{-(CH}_2\text{)}_2\text{-COOH}$ ), 26.24 ( $\text{-NH-CH}_2\text{-CH}_2\text{-CH}_2\text{-CH}_2\text{-COOH}$ ), 29.09 ( $\text{-NH-(CH}_2\text{)}_4\text{-CH}_2\text{-COOH}$ ), 33.99 ( $\text{-NH-CH}_2\text{-(CH}_2\text{)}_4\text{-COOH}$ ), 35.58 ( $\text{CH}_3\text{-CH}_2\text{-NH-}$ ), 165.47-167.87 (C of *s*-triazine ring), 175.57 ( $\text{-NH-(CH}_2\text{)}_5\text{-COOH}$ )

## A 2.2.4 Protein-biochemical Methods

### A 2.2.4.1 Cell Disruption and Sample Preparation

Cells were harvested by centrifugation (30 min, 3,000-5,000 g, 4°C). The resulting supernatant (extracellular fraction) contained soluble proteins that were released by extracellular secretion and cell lysis.

In order to separate proteins in the periplasm, the cytoplasm and insoluble protein, cell extracts of the respective fractions were prepared as suggested by Novagen [39]: Periplasmic fractions were prepared by osmotic shock followed by centrifugation. The resulting supernatant was the soluble periplasmic fraction. The pellet after centrifugation, containing the protoplasts and insoluble proteins from the periplasm, was used to prepare the soluble fraction of the cytoplasm and the insoluble fraction with insoluble proteins of the periplasm and the cytoplasm. The preparation of the cytoplasmic fraction included the use of lysozyme, the detergent Triton X-100, EDTA and sonification. Equivalent amounts of these samples were used for SDS-PAGE, ELISA and FLISA.

In order to prepare soluble and insoluble fractions without any further fractionation, cells were only resuspended in 80 mM PBS buffer (80 mM phosphate buffer, 0.85% (m/v) sodium chloride, pH 7.2), sonified (2x 1 min, output 3, 30% duty cycle) and centrifuged (20,000 g, 4°C, 30 min). For downstream processing, i.e. refolding of insoluble protein and purification, lysozyme, detergents and EDTA were thought to be disturbing. Therefore cells were resuspended in 80 mM PBS buffer and lysed by prolonged sonification (3-5x 1 min, output 3, 30% duty cycle).

### A 2.2.4.2 Quantification of Protein

#### A 2.2.4.2.1 Protein Assays Based on Protein-specific Colour Reactions

The concentrations of proteins in different samples were determined using the BCA Protein Assay Kit (Pierce) or the BioRad Protein Assay Kit (BioRad) according to the instructions of the manufacturers. The former involves the reaction of proteins with Cu(II)-ions in alkaline medium resulting in Cu(I)-ions, which form a complex with BCA with an absorbance maximum at 562 nm. The latter is based on the observation that the absorbance maximum for an acidic solution of Coomassie Brilliant Blue G-250 shifts from 465 nm to 595 nm when binding to protein occurs. Bradford used this principle in a protein assay [95]. This assay is less susceptible to complexing reagents and can also be used for samples containing imidazole, guanidine or arginine in low concentrations. For both assays BSA was used as a standard.

#### A 2.2.4.2.2 Densitometric Determination of Protein Concentration

Different concentrations of the sample with the protein of interest and the LMW standard from the LMW Electrophoresis Calibration Kit (Table 2.11 in 2.10.3) were separated by SDS-PAGE. After Coomassie Brilliant Blue staining and drying, the gel was scanned and the bands were analyzed densitometrically using the ScionImage software (<http://www.scioncorp.com>). This method is very useful when the protein has not been purified. However, attention must be paid at protein amounts higher than 800 ng per slot, as saturation of the density may occur.

### A 2.2.4.3 SDS-PAGE

For analytical and preparative purposes proteins were separated under denaturing conditions according to Laemmli [96] in 12.5% (m/v) SDS-polyacrylamide gels. The 4% stacking gel was prepared with

Rotiphorese® Gel 30	0.53 ml
4x Upper Tris (12.11 g Tris, 0.8 g SDS, ddH <sub>2</sub> O ad 200 ml, pH 6.8)	1.00 ml
dH <sub>2</sub> O	2.47 ml
10% (m/v) APS	40 µl
TEMED	4 µl

The 12.5% resolving gel consisted of

Rotiphorese® Gel 30	3.33 ml
4x Lower Tris (36.46 g Tris, 0.8 g SDS, ddH <sub>2</sub> O ad 200 ml, pH 8.8)	2.00 ml
ddH <sub>2</sub> O	2.67 ml
10% (m/v) APS	40 µl
TEMED	4 µl

Protein samples containing about 5 µg protein (0.5 to 15 µl of samples) were mixed with at least one volume SDS loading buffer (100 mM Tris, 200 mM DTT, 4% (m/v) SDS, 0.2% (m/v) BPB, 20% (v/v) glycerole), denatured for 5 min at 95°C, cooled to room temperature and loaded into the slots of the stacking gel.

Two standards were used as references: a low molecular weight standard from BioRad for qualitative and the standard from the LMW Electrophoresis Calibration Kit (Amersham Pharmacia Biotech) for quantitative analysis. Some of the proteins in the latter are engineered to have even molecular masses. The BioRad LMW was used as suggested by the manufacturer. Concerning the Amersham Pharmacia Biotech LMW, the contents of one vial were dissolved in 200 µl SDS loading buffer and 200 µl ddH<sub>2</sub>O. Table 2.11 shows the protein composition of both standards including the molecular masses and for the Amersham Pharmacia Biotech standard also the concentration of each protein.

**Table A 2.11**

Protein	BioRad LMW	LMW Electrophoresis Calibration Kit	
	Molecular Mass	Molecular Mass	µg Protein/µl
phosphorylase B	97.4 kDa	94.0 kDa	160.0
serum albumin	66.2 kDa	67.0 kDa	207.5
ovalbumin	45.0 kDa	43.0 kDa	367.5
carbonic anhydrase	31.0 kDa	30.0 kDa	207.5
trypsin inhibitor	21.5 kDa	20.1 kDa	220.0
lysozyme	14.4 kDa	14.4 kDa	302.5

Gels were run with electrode buffer (0.3% (m/v) Tris, 1.44% (m/v) glycine, 0.1% (m/v) SDS, pH 8.3) for 10 min at 10 mA per gel followed by 25 mA per gel until the BPB band reached

the bottom of the resolving gel. Gels were stained for at least one hour with Coomassie Brilliant Blue solution (0.1% (m/v) Coomassie Brilliant Blue R250, 30% (v/v) methanol, 10% (v/v) acetic acid) and destained in 30% (v/v) methanol, 10% (v/v) acetic acid. For storage, gels were dried at 80°C under vacuum for 2 h.

#### **A 2.2.4.3 N-terminal Sequencing of Proteins**

Proteins were sequenced N-terminally according to the method described by Edman [97]. The N-terminal amino acids were released from the peptide chain using phenylthioisocyanate and the released phenylthiohydantoin derivatives were analyzed by HPLC due to their different retention times.

The sample containing the protein of interest was separated by SDS-PAGE and electroblotted onto polyvinylidene fluoride (PVDF) as described [98]. Three filter papers each of the size of the gel were incubated for about 5 min in one of the following buffers, respectively:

anode buffer 1: 0.3 M Tris, 10% (v/v) methanol, pH 10.4,

anode buffer 2: 25 mM Tris, 10% (v/v) methanol, pH 10.4,

cathode buffer: 25 mM Tris, 40 mM 6-aminohexanoic acid, 10% (v/v) methanol, pH 9.4.

The gel was equilibrated in cathode buffer. The components of the sandwich were arranged on the anode platform of the semidry transfer cell (BioRad) in the following order (from bottom to top), preventing any bubbles between the layers.

1. three filter papers incubated in anode buffer 1
2. three filter papers incubated in anode buffer 2
3. a PVDF membrane cut to the size of the gel and wetted with 100% methanol
4. the SDS-polyacrylamid gel equilibrated in cathode buffer
5. three filter papers incubated in cathode buffer

After positioning the cathode on top, the negatively charged proteins were transferred onto the PVDF membrane at 15 V for 45 min. The membrane was soaked in staining solution (0.025% (m/v) Coomassie Brilliant Blue R-250, 40% (v/v) methanol) and destained in 50% (v/v) methanol, until distinct bands were visible. The membrane was rinsed with dH<sub>2</sub>O, the

protein of interest was excised and applied to the protein sequencer or, alternatively, stored at  $-20^{\circ}\text{C}$ .

#### **A 2.2.4.5 Immunological Detection of Proteins (Western Blot)**

Proteins separated by SDS-PAGE were transferred onto Zeta-Probe® membrane using a semidry transfer cell (BioRad). Therefore the gel and a Zeta-Probe® membrane of the size of the gel were incubated for 15 min, 8 filter papers of the same size were soaked in transfer buffer (20% (v/v) methanol, 142 mM glycine, 25 mM Tris-HCl, pH 8.3). The sandwich was arranged on the anode platform of the semidry transfer cell in the following order (from bottom to top).

1. four soaked filter papers
2. the Zeta-Blot® membrane
3. the gel
4. four soaked filter papers

Blotting was performed at 15 V for 45 min. The membrane was air-dried and blocked in blocking solution (1% (m/v) casein in TS buffer (0.85% (m/v) sodium chloride, 10 mM Tris-HCl, pH 7.5)). Then the membrane was incubated overnight under shaking at  $4^{\circ}\text{C}$  with 0.1% (v/v) mouse anti-polyhistidine monoclonal IgG solution in blocking buffer, washed twice for 20 min in 0.3% (v/v) Tween 20 in blocking buffer followed by three 10 min washing steps in TS buffer. Subsequently the membrane was incubated for one hour under shaking at RT in 0.05% (v/v) goat anti-mouse polyclonal antibody-HRP conjugate in blocking buffer, followed by three 10 min washes in TS. Bound antibody-HRP conjugate was detected by oxidation of 4-chloro-1-naphthol into a violet precipitate by the HRP. This was done by incubating the membrane in staining solution (0.05% (m/v) 4-chloro-1-naphthol dissolved in 16.7% (v/v) methanol, 15 mM di-sodium hydrogen phosphate, 12.5 mM citric acid, pH 6.0, 0.5% (v/v) hydrogen peroxide) until distinct bands were visible. Finally the membrane was rinsed with  $\text{dH}_2\text{O}$  and air-dried.

#### **A 2.2.4.6 Immunoassays in Microtiter Plates**

The described ELISAs (2.2.4.6.2) and the FLISA (2.2.4.6.3) can be used for the characterization and quantification of antibodies or antibody fragments as well as for the

determination of *s*-triazines. Quantities given for individual incubation steps always refer to one cavity. If not mentioned otherwise, incubation of the respective solutions was always done for 1 hour at RT with orbital shaking. Washing steps were done between all incubation steps with 8 mM PBST (1:10 dilution of 80 mM PBS buffer (2.10.1) containing 0.05 % (v/v) Tween 20) with three washes each (Titertek Microplate Washer M96V and Titertek Microplate Stacker, ICN).

#### ***A 2.2.4.6.1 Synthesis of C<sub>6</sub>-s-Triazine Protein Conjugates***

##### ***A 2.2.4.6.1.1 Synthesis of C<sub>6</sub>-Atrazine Horseradish Peroxidase (HRP) Conjugate***

The coupling of C<sub>6</sub>-triazines to amino groups of HRP was done according to a method described previously [99]. Thereby an activated ester is formed by the carboxylic acid function of C<sub>6</sub>-atrazine and the hydroxyl group of N-hydroxysuccinimide (NHS) using dicyclohexylcarbodiimide (DCC) as a catalyst and water acceptor. The activated ester bond can then be converted in an amide bond between the carboxylic acid function of C<sub>6</sub>-atrazine and an amino group on the surface of a protein.

C<sub>6</sub>-atrazine (4-chloro-6-(iso-propylamino)-*s*-triazine-2-6-aminohexane-carboxylic acid) was synthesized according to Baemner [100]. 1 mg C<sub>6</sub>-atrazine (3.2 μmol), 1.7 mg NHS (15 μmol) and 6.2 mg DCC (30 μmol) were mixed in 130 μl dioxane in a 1.5 ml reaction tube and incubated with mild shaking for 18 hours at RT. The batch was centrifuged and the supernatant, containing the activated ester, was added dropwise to a solution of 2 mg HRP in 3 ml 0.13 M NaHCO<sub>3</sub> and incubated for 3 hours in the dark. Subsequently, the conjugate was purified by gel filtration using PD-10 Sephadex columns (Amersham Pharmacia Biotech) and 80 mM PBS buffer. Collected 0.5 ml fractions were measured photometrically at 403 nm (absorbance maximum of HRP), 260 nm (absorbance maximum of atrazine) and 220 nm (absorbance maximum of HRP and atrazine). Those with the highest ratio of A<sub>220</sub> to A<sub>260</sub> were used as tracer.

##### ***A 2.2.4.6.1.2 Synthesis of C<sub>6</sub>-s-triazine bovine serum albumine (BSA) conjugates***

Conjugates of C<sub>6</sub>-atrazine and C<sub>6</sub>-simazine with BSA were synthesized in analogy to the method described in A 2.10.6.1.1 except that BSA was used instead of HRP and protein could not be monitored in the course of the purification at 403 nm. C<sub>6</sub>-simazine was synthesized as

described in A 2.9. Fractions containing the most conjugate and the least free atrazine were used for precoating microtiter plates for hapten-immobilized ELISA and FLISA.

#### **A 2.2.4.6.2 Enzyme-linked Immunosorbent Assays (ELISAs)**

ELISAs were performed in 96 well Maxisorp microtiter plates (NuncImmuno™, MaxiSorp™, Nunc). Extinctions were measured at 450 nm using the FLUOstar, (BMG LabTechnologies).

##### **A 2.2.4.6.2.1 Antibody-fragment-immobilized ELISA**

Antibody-fragment-immobilized ELISA was performed similar as described previously [22]. Microtiter plates were precoated with the capture antibody (200 µl of 0.1% (v/v) rabbit anti-mouse antibody solution (Sigma-Aldrich) in 50 mM sodium carbonate buffer pH 9.6) overnight at 4°C. After washing, the plates were either used directly or stored at -20°C. 200 µl of an appropriate scFv dilution in 80 mM PBS buffer were applied to each well. A variant of the above format contained one additional incubation step: between the precoating of the microtiter plates with rabbit anti-mouse antibody and the incubation with the scFv, each well was incubated with 200 µl 0.04 % (v/v) mouse anti-polyhistidine antibody solution in 80 mM PBS buffer. For the detection of *s*-triazines in a competitive immunoassay, 200 µl of different dilutions of the respective *s*-triazine and 50 µl of 0.05% (v/v) atrazine-HRP conjugate solution in 80 mM PBS buffer were added. After a final washing step, 200 µl of substrate solution (25 ml 0.1 M sodium acetate, pH 5.5, 400 µl 0.6% (m/v) 3,3',5,5'-tetramethylbenzidine (TMB), 100 µl 1% (v/v) hydrogen peroxide) were added and incubated until a distinct blue dye was visible. The color reaction was stopped with 1 M sulfuric acid and the absorbance was measured at 450 nm.

##### **A 2.2.4.6.2.2 Hapten-immobilized ELISA**

Alternatively, a hapten-immobilized ELISA was performed similar to a method described by Giersch and coworkers [69]. Microtiter plates were coated with 0.1 % (v/v) C<sub>6</sub>-atrazine-BSA conjugate solution in 50 mM sodium carbonate buffer pH 9.6 overnight at 4°C. The microtiter plates were blocked with 300 µl 1 % (w/v) casein in 80 mM PBS buffer to avoid unspecific binding of secondary antibodies. After incubation with 100 µl of the *s*-triazine standard solution or water when characterizing scFv and 100 µl of the appropriate scFv dilution in 80



mM PBS buffer, the microtiter plate was incubated with 200  $\mu$ l 0.04% (v/v) mouse anti-polyhistidine IgG solution in 80 mM PBS buffer. Omission of this step resulted in very poor labeling of the scFv. Plates were then incubated with 0.02% (v/v) goat anti-mouse IgG horseradish peroxidase conjugate solution in 80 mM PBS buffer. After a final washing step, the color reaction was carried out as described in 2.10.6.2.1.

#### ***A 2.2.4.6.3 Fluorophor-Linked Immunosorbent Assay (FLISA)***

For the fluorophor-linked immunosorbent assay, black microtiter plates (FluoroNunc<sup>TM</sup>, MaxiSorp<sup>TM</sup>, Nunc) were coated as described for the hapten-immobilized ELISA with either the C<sub>6</sub>-atrazine-HRP or the C<sub>6</sub>-simazine-HRP conjugate. For the determination of *s*-triazines, plates were incubated with 100  $\mu$ l per well of different *s*-triazine dilutions in water and 100  $\mu$ l of different dilutions of samples containing the scFv-EGFP fusion protein in 80 mM PBS buffer. For the characterization of the fusion protein, no *s*-triazine was used. After washing and the addition of 200  $\mu$ l 80 mM PBS per well, the remaining fluorescence was measured (FLUOstar, BMG LabTechnologies, excitation at 485 nm, emission at 538 nm).

#### ***A 2.2.4.6.4 Evaluation of ELISAs and FLISAs***

The normalization of absorbance and fluorescence signals as well as the determination of IC<sub>50</sub>-values, detection limits and cross-reactivities were done as described previously [69].

##### ***A 2.2.4.6.4.1 Normalization of ELISA and FLISA Signals***

Usually between 4 and 6 cavities of a microtiter plate were measured with the same sample and the averages of the measured absorbances were used for further evaluation. Normalization was done by using the equation

$$\% B / B_0 = \frac{(A - A_{\text{excess}})}{(A_0 - A_{\text{excess}})} \cdot 100,$$

with %B/B<sub>0</sub> being the normalized signal, A the measured absorbance or fluorescence, A<sub>0</sub> the signal obtained without *s*-triazine and A<sub>excess</sub> with an excess of *s*-triazine, respectively.

#### **A 2.2.4.6.4.2 Determination of $IC_{50}$ -values**

The  $IC_{50}$ -value describes the center of the working range of an assay, at which 50% inhibition are observed. To obtain  $IC_{50}$ -values for the investigated herbicides, calibration curves were made using normalized values measured with different concentrations of the respective *s*-triazine with the concentration plotted in a logarithmic scale on the x-axis. The obtained curves were fitted sigmoidally with the Origin 5.0 software and the concentration for 50% inhibition were determined.

#### **A 2.2.4.6.4.3 Determination of Cross Reactivities**

The cross reactivities of cyanazine, terbutylazine, propazine, deethylatrazine and simazine towards the scFv K411B with atrazine as a reference were examined. The cross reactivity in % (%CR) is defined by

$$\%CR = \frac{IC_{50}(\text{atrazine})}{IC_{50}(\textit{s-triazine})} \cdot 100$$

#### **A 2.2.4.7 Concentration of Protein**

Samples with low concentrations of protein were concentrated using Amicon stirring cells and Amicon YM10 filters (Millipore, cut-off: 10 kDa) according to the instructions of the manufacturer.

#### **A 2.2.4.8 Purification of Polyhistidine-tagged proteins by Immobilized Metal Affinity Chromatography (IMAC)**

Polyhistidine-tagged proteins were purified under native conditions by IMAC using TALON<sup>TM</sup> Metal Affinity Resins (Clontech) according to the manufacturer's recommendations at pH 7.0 using imidazole for elution. However, for washing steps "equilibration buffer" without imidazole was used, as concentrations of 5 mM imidazole in the "washing buffer" resulted in partial elution of the protein.

#### **A 2.2.4.9 Refolding of Insoluble Protein**

To obtain active scFv from inclusion bodies, refolding was done in a similar way as reported [63]. 17 g wet cells, derived from the fed-batch cultivation were washed, resuspended 10-fold

concentrated in lysis buffer and disrupted by sonification. After centrifugation (30,000 g, 4 °C, 30 min) the pellet of cell debris and insoluble proteins was resuspended and stirred gently for 10 hours at 20°C in 80 ml of denaturation buffer (8 M guanidine hydrochloride, 2 mM EDTA, 150 mM dithioerythritol (DTE), 0.1 M tris-(hydroxymethyl)-aminomethane hydrochloride (Tris-HCl), pH 8.0) in order to solubilize the protein. After centrifugation as above, the supernatant was diluted 1:25 in renaturation buffer (0.5 M L-arginine, 2 mM ethylenediaminetetraacetate (EDTA), 16 mM oxidized glutathione, 0.1 M Tris-HCl pH 8.0), stirred gently for 72 hours at 10°C, concentrated 50-fold as described in 2.10.7 and centrifuged (20,000 g, 4 °C, 5 min) to obtain the soluble fraction of the concentrate.

#### **A 2.2.4.10 Fluorescence Measurement**

Fluorescence emitted by the EGFP domain of the scFv-EGFP fusion protein was measured in black microtiter plates (FluoroNunc™, MaxiSorp™, Nunc) during the incubation period for the FLISA (FLUOstar, BMG LabTechnologies, excitation at 485 nm, emission at 538 nm).

#### **A 2.2.5 Microscopy**

Light microscopy and fluorescence microscopy were performed using a Leica DMIRBE microscope (Leica Microsystems, Wetzlar, Germany) and an HCX PL Apo 63x/1.32-0.6 oil immersion objective. Fluorescence setup: excitation: 488 nm with Polychrome 2 (TILL Photonics, München, Germany); dichroic mirror: F53-009 (AHF Analysentechnik, Tübingen, Germany), emission: bandpass filter 535/50 nm (AHF Analysentechnik). Images were taken with a Hamamatsu Orca C4742-95 camera (Hamamatsu Photonics K.K., Hamamatsu, Japan) and processed with Openlab 2 (Improvision, Coventry, England).

Cells were harvested by centrifugation 2 hours after induction and cultivation at 30°C, washed in 80 mM PBS buffer and resuspended 5-times concentrated in 2% (m/v) paraformaldehyde in 80 mM PBS. After a 20-minute incubation period at 37°C, the cells were centrifuged, resuspended 10 times concentrated in FluoromountG embedding medium (Southern Biotechnology Associates) and analyzed by microscopy.

## A 3 Results

### A 3.1 Cloning Procedures

#### A 3.1.1 Construction of Plasmids for the Expression of the Atrazine-specific ScFv K411B

##### A 3.1.1.1 Plasmids for the Expression of the ScFv K411B under Control of the T7 Promoter

The plasmid pETK(+), containing the gene encoding the scFv K411B with a pelB leader at the N-terminus for periplasmic expression and a His<sub>6</sub>-tag at the C-terminus was used as a starting point [82]. It is a derivative of the pET20b(+) vector (Novagen) designed for the high-level expression of heterologous proteins under control of the T7 promoter. The scFv K411B gene consists of the heavy-chain-encoding region at the 5'-end and the light-chain-encoding region at the 3'-end, both separated by a sequence coding for a (Gly<sub>4</sub>-Ser)<sub>3</sub> linker [23]. To obtain pETK(-), in which the sequence coding for the pelB leader sequence was deleted to allow cytoplasmic expression, pETK(+) was digested with *Nde*I and *Nco*I, followed by Klenow fill-in and blunt-end religation of the 4369 bp fragment. In order to test whether the insertion of two extra nucleotides upstream of the start codon (inserted as a consequence of the cloning procedure) had an effect on translation efficiency, the original number of nucleotides between the Shine Dalgarno sequence and the start codon was restored by removing these two additional nucleotides with QuikChange PCR using the primers Dalgy F/R, resulting in the plasmid pETK(-)-AC. The restored sequence between the Shine Dalgarno sequence and the start codon is supposed to give optimal translation efficiency [41]. For comparative purposes, two extra nucleotides were inserted into the plasmid pETK(+) at the 5'-side of the start codon by QuikChangePCR using the primers negACive F/R, resulting in pETK(+)+AC. As expression levels of pETK(-)-AC were low and a +2 codon consisting of AAA is supposed to be beneficial for translation initiation [41], an additional AAA codon was inserted at the +2 codon by QuikChange PCR using the primers Lys1 F/R, resulting in the plasmid pETK(-)-AC+AAA. Still not having obtained good expression levels, a further AAA codon was introduced at the same position using the primers Lys2 F/R, giving the plasmid pETK(-)-AC+AAAAAA. This was done in order to prevent the formation of a stable

secondary structure of the mRNA as suggested by Sabine Arnold, Institute of Biochemical Engineering, University of Stuttgart, based on molecular modeling studies (personal communication). The results of all cloning procedures were verified by DNA sequencing. The coding regions and the translation initiation regions of all used plasmids are shown in Figure A 4 and Table A 3.1.

#### **A 3.1.1.2 Plasmids for the Expression of the ScFv K411B under Control of the *rhaBAD* Promoter**

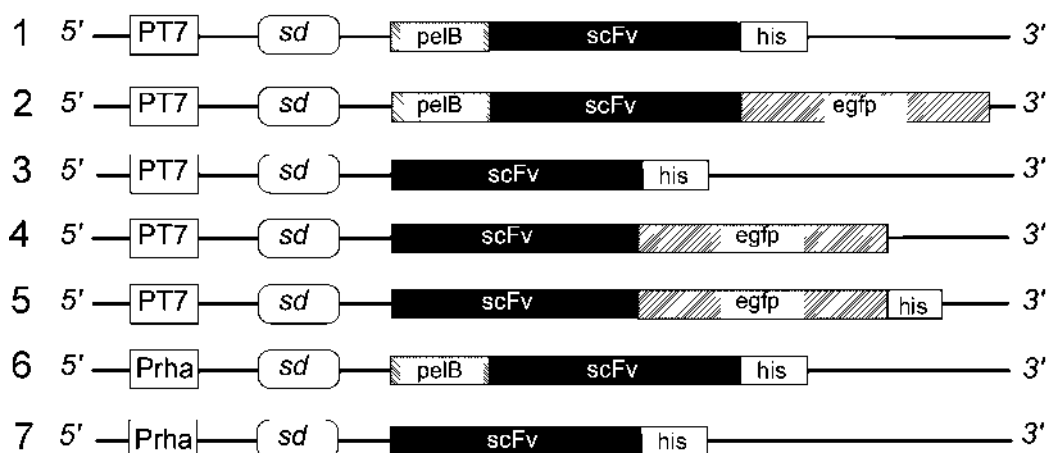
For expression under control of the *rhaBAD* promoter, the genes coding for the scFv with and without the pelB leader sequence were cloned into the vector pTST101 [84] after excision of the MalE-EGFP gene. This vector has an additional T7 promoter upstream of the *rhaBAD* promoter. The pTST101 was digested with *HindIII* and pETK(+) with *Bpu1102I*. After Klenow fill-in both fragments were digested with *NdeI*. The 4583 bp fragment after digestion of pTST101 and the 927 bp fragment after digestion of pETK(+) were ligated resulting in pJOEK(+). This also resulted in the insertion of two extra nucleotides A and T upstream of the start codon, which were removed by QuikChange PCR using the primers Shiny F/R, giving pJOEK(+)-AT. pJOEK(-) without the pelB leader sequence for cytoplasmic expression and pJOEK(-)-AC with the corrected sequence 5' of the start codon were created from pJOEK(+)-AT according to the creation of pETK(-) and pETK(-)-AC from pETK(+) using the respective cloning procedure and primers. Further details are shown in Figure A 4 and Table A 3.1.

#### **A 3.1.2 Construction of Plasmids for the Expression of the scFv-EGFP Fusion Protein**

The scFv-EGFP fusion protein was expressed under control of the T7 promoter using pET20b(+) derivatives. To generate pETKE(+) and pETKE(-), pETK(+) and pETK(-) were digested with *SalI*, pEGFP with *NcoI*. After Klenow fill-in, the linearized plasmids were digested with *NotI*; the resulting 724 bp pEGFP fragment was ligated with the linearized plasmids pETK(+) and pETK(-), respectively. The resulting plasmids pETKE(+) and pETKE(-) contain the gene coding for the fusion protein consisting of the scFv gene followed directly in frame by the EGFP gene, with and without the pelB leader sequence at the 5'-terminus. To restore the original number of nucleotides between the Shine Dalgarno sequence and the start codon in pETKE(-), two additional nucleotides were removed by QuikChange

PCR using the primers Dalgy F/R, resulting in pETKE(-)-AC. Again, a control plasmid was created by inserting two extra bases 5' of the start codon in pETKE(+) by QuikChange PCR using the primers negACive F/R, resulting in pETKE(+)+AC.

In order to fuse a His<sub>6</sub>-tag-coding sequence to the 3'-end of the scFv-EGFP gene, the EGFP-encoding sequence was amplified with the primer *SalI* F binding at the *SalI* restriction site (the site at which the scFv gene and the EGFP gene had been fused) and the primer *NotI* R including the 3'-end of the EGFP gene, the His<sub>6</sub>-tag sequence and a *NotI* restriction site. Digestion of the PCR product and pETKE(-) with *SalI* and *NotI* and ligation of the digested PCR product and the 4420 bp vector fragment after digestion of pETKE(-) did not yield ligands. Assuming that the hardly ligatable *SalI*-cut sticky ends were responsible for that, they were transformed into blunt ends. In order to achieve this without introducing a frameshift, the vector was digested with *SalI* followed by treatment with Mung Bean Nuclease and the PCR product was also cut with *SalI* followed by Klenow fill. Subsequent digestion of both fragments with *NotI*, ligation and transformation resulted in cells containing the expected construct. The stop codon was integrated in a second step by QuikChange PCR using the primers Stop codon F/R. DNA sequencing revealed, that the His<sub>6</sub>-tag was not in frame and a stop codon had formed upstream of the His<sub>6</sub>-tag. Therefore a T was removed using QuikChange PCR and the primers pet 1 F/R, finally resulting in the plasmid pETKEHS(-) coding for the scFv-EGFP fusion protein with a C-terminal His<sub>6</sub>-tag. The additional nucleotides 5' of the start-codon were again removed using QuikChange PCR and the primers Dalgy F/R, resulting in pETKEHS(-)-AC. The constructs are depicted in Figure A 4 and the translation initiation regions are shown in Table A 3.1.



**Figure A 4.** 1 to 7: the different types of constructs used in this work. Further details are shown in Table 3.1. PT7: T7 promoter, Prha: *rhaBAD* promoter, *sd*: Shine Dalgarno sequence, pelB: pelB leader sequence, his: His<sub>6</sub>-tag.

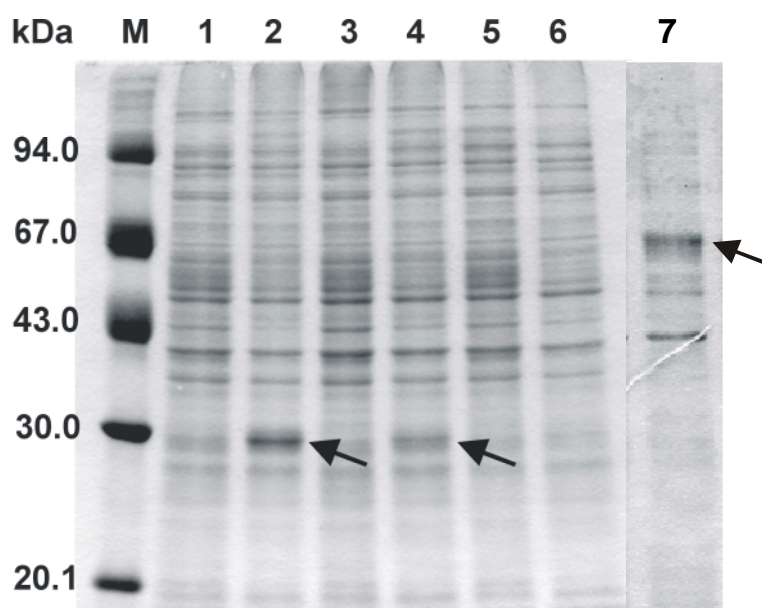
**Table A 3.1.** Features of the used constructs. *sd*: Shine Dalgarno sequence, tic: translation initiation codon.

Plasmid Name	Size (bp)	Construct Type	Translation Initiation Sequence	
			<i>sd</i>	<u>tic</u> -encoding-gene
pETK(+)	4433	1	5′-AGAAGGAGAUUAUACAUAUGAAAUAC-3′	
pETK(+)+AC	4435	1	5′-AGAAGGAGAUUAUACAUAUGAAAUAC-3′	
pETKE(+)	5148	2	5′-AGAAGGAGAUUAUACAUAUGAAAUAC-3′	
pETKE(+)+AC	5150	2	5′-AGAAGGAGAUUAUACAUAUGAAAUAC-3′	
pETK(-)	4369	3	5′-AGAAGGAGAUUAUACAUAUGGUCCAG-3′	
pETK(-)-AC	4367	3	5′-AGAAGGAGAUUAUACAUAUGGUCCAG-3′	
pETK(-)-AC+AAA	4370	3	5′-AGAAGGAGAUUAUACAUAUGAAAGUCCAG-3′	
pETK(-)-AC+AAAAAA	4373	3	5′-AGAAGGAGAUUAUACAUAUGAAAAAGUCCAG-3′	
pETKE(-)	5084	4	5′-AGAAGGAGAUUAUACAUAUGGUCCAG-3′	
pETKE(-)-AC	5082	4	5′-AGAAGGAGAUUAUACAUAUGGUCCAG-3′	
pETKEHS(-)	5120	5	5′-AGAAGGAGAUUAUACAUAUGGUCCAG-3′	
pETKEHS(-)-AC	5118	5	5′-AGAAGGAGAUUAUACAUAUGGUCCAG-3′	
pJOEK(+)	5512	6	5′-AGAAGGAGAUUAUACAUAUGAAAUAC-3′	
pJOEK(+)-AT	5510	6	5′-AGAAGGAGAUUAUACAUAUGAAAUAC-3′	
pJOEK(-)	5446	7	5′-AGAAGGAGAUUAUACAUAUGGUCCAG-3′	
pJOEK(-)-AC	5444	7	5′-AGAAGGAGAUUAUACAUAUGGUCCAG-3′	

### A 3.2 Expression of the scFv K411B and the ScFv-EGFP Fusion Protein in *E. coli* BL21(DE3) in Shake Flask Cultures

The genes coding for the atrazine-specific antibody fragment scFv K411B with and without pelB leader sequence and the fusion protein of the respective scFv and a C-terminal EGFP without pelB leader sequence were expressed in *E. coli* BL21(DE3) cells in shake flask cultures. The scFv contains two disulfide bonds and it was of interest to investigate whether periplasmic expression favored correct disulfide bond formation and folding. *E. coli* BL21(DE3) cells transformed with pETKE(+) coding for the fusion protein with pelB leader sequence were not viable and the protein could not be expressed in these cells. Induced cells transformed with the other plasmids were harvested and cell extracts were prepared to investigate the accumulation of total and of functional recombinant protein in different cell compartments.

To check whether the recombinant proteins were expressed, total cell extracts were separated by SDS-PAGE and analyzed after Coomassie Brilliant Blue Staining (Figure A 5).



**Figure A 5.** SDS-PAGE of total cell extracts of *E. coli* BL21(DE3) cells. M: Molecular weight marker, the molecular mass (kDa) of the proteins is indicated on the left. Lanes 1 and 2: BL21(DE3)/pETK(+), lanes 3 and 4: BL21(DE3)/pETK(-), lanes 5 and 6: BL21(DE3)/pET20b(+), lane 7: BL21(DE3)/pETKE(-). Lanes 1, 3 and 5: uninduced, lanes 2, 4, 6 and 7: induced. Arrows indicate the scFv K411B and the scFv-EGFP fusion protein, respectively.



For pETK(+) (lane 2), a clear band of 29 kDa, which corresponds to the calculated mass of the mature scFv, was obtained after expression of scFv with pelB leader sequence. For pETK(-) (lane 4), a band of identical size was observed which resulted from the expression of the scFv without pelB leader. This indicates that the pelB leader sequence of the scFv precursor had been cut off and translocation to the periplasm had occurred (lane 2). For pETKE(-) a clear band at about 55 kDa was observed corresponding to the scFv-EGFP fusion protein (calculated mass: 55.2 kDa). In the controls without induction (lanes 1 and 3) and with the plasmid pET20(+) without any encoding sequence (lanes 5 and 6) no distinct bands were visible at the respective positions.

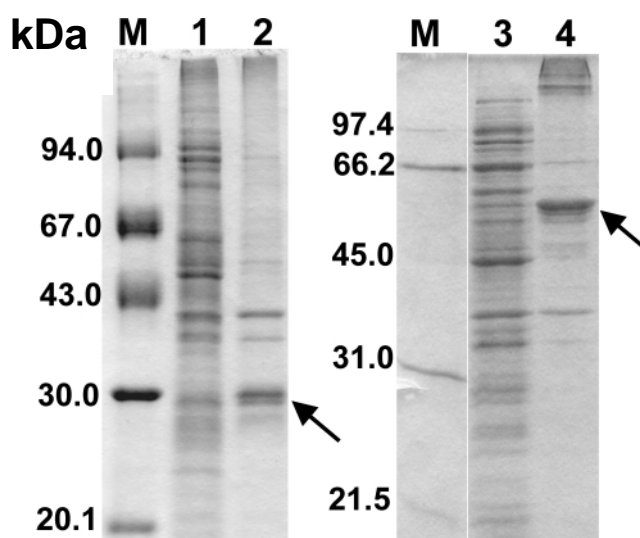
### **A 3.2.2 Quantification of Total Recombinant Protein by Densitometric Evaluation**

Densitometric evaluation of the Coomassie-Brilliant-Blue-stained SDS-polyacrylamide gel shown in Figure A 5 was used for a rough estimation of the protein content in lane 2 and lane 4 and indicated a 2.3 times greater extent of product accumulation for the scFv expressed in the periplasm compared to scFv expressed in the cytoplasm. The band corresponding to the scFv in lane 2 was estimated to contribute about 12% to the total cell protein. Accordingly, the band in lane 4 corresponding to the scFv indicated ~5% recombinant protein compared to the the total cell protein. A similar value was estimated for the scFv-EGFP fusion protein in lane 7. Thus, it was shown that after periplasmic expression more heterologous protein had accumulated than after cytoplasmic expression.

### **A 3.2.3 Quantification of Functional Recombinant Protein**

#### **A 3.2.3.1 Comparison of Soluble and Insoluble Fractions**

When a protein is soluble, it usually has attained its native state and consequently also its functionality. Therefore the amount of functional protein can be estimated by comparing bands of the respective protein after SDS-PAGE in soluble and insoluble fractions. However, in complex mixtures, solubilities can only be roughly estimated by this method. Soluble and insoluble fractions of cells expressing the scFv in the periplasm and cells expressing the scFv-EGFP fusion protein in the cytoplasm were prepared using sonification and separated by SDS-PAGE (Figure A 6).



**Figure A 6.** SDS-PAGE of soluble and insoluble fractions of BL21(DE3) cells expressing the scFv in the periplasm (lanes 1 and 2) and the scFv-EGFP in the cytoplasm (lanes 3 and 4). M: molecular weight marker, the molecular weights are indicated on the left of the gels. lanes 1 and 3: soluble fractions, lanes 2 and 4: insoluble fractions. Arrows indicate the respective recombinant proteins.

It was estimated that between 10 and 20% of both proteins were soluble. It seemed that direction into the periplasm did not increase proper folding and that the percentage of soluble protein was the same after periplasmic and cytoplasmic expression.

### A 3.2.3.2 Determination of Active ScFv Using Non-competitive Hapten-immobilized ELISA

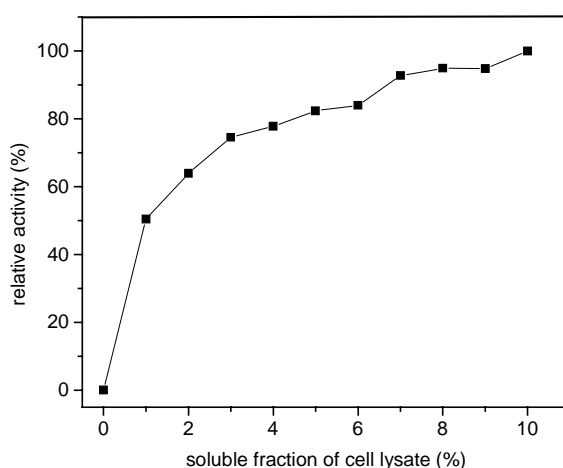
#### A 3.2.3.2.1 Development of the Hapten-immobilized ELISA

Due to the limited accuracy of estimation of solubilities from SDS-PAGE, it was preferred to use an activity assay to determine functional recombinant scFv. Antibody-fragment-immobilized ELISA was not found to be suitable for the quantification of active scFv, because the absorbance signals were very low also after enzyme reaction periods of at least 1 hour. Using undiluted cell extracts from shake flask expressions, absorbances of only 0.05 above background were achieved. The background was defined by the signal obtained in wells without scFv. Also the application of anti-polyhistidine IgG as an additional capture antibody increased absorbances only to about 0.2 above background.

In contrast, non-competitive hapten-immobilized ELISA proved to be practicable for the determination of active scFv. To investigate if this assay format could also be performed without anti poly-histidine IgG for labeling of bound scFv, this incubation step was left out,

but only absorbances of less than 0.05 above background were achieved when using scFv in crude cell extracts. In contrast, using 0.7  $\mu\text{g/ml}$  anti-polyhistidine IgG, absorbances of more than 2 above background were obtained. This clearly indicates that the scFv is very poorly recognized by the goat anti mouse antibody HRP conjugate. However, the IgG also bound unspecifically to the microtiter plate wells. Thus, without blocking absorbances of up to 2 in wells without scFv were obtained. Application of 1% casein in 80 mM PBS buffer as blocking reagent could lower this background to an absorbance of 0.2. Therefore, blocking (300  $\mu\text{l}$  1% (m/v) casein in 80 mM PBS) and incubation with 0.7  $\mu\text{g/ml}$  anti poly-histidine IgG was used in the following analysis of functional scFv by hapten-immobilized ELISA.

In Figure A 7 different dilutions of a soluble fraction prepared by sonification of *E. coli* cells expressing scFv in the periplasm were applied to hapten-immobilized ELISA. Unfortunately, the amount of scFv from cell extracts applied to hapten-immobilized ELISA plotted against the ELISA signals did not result in a linear correlation.

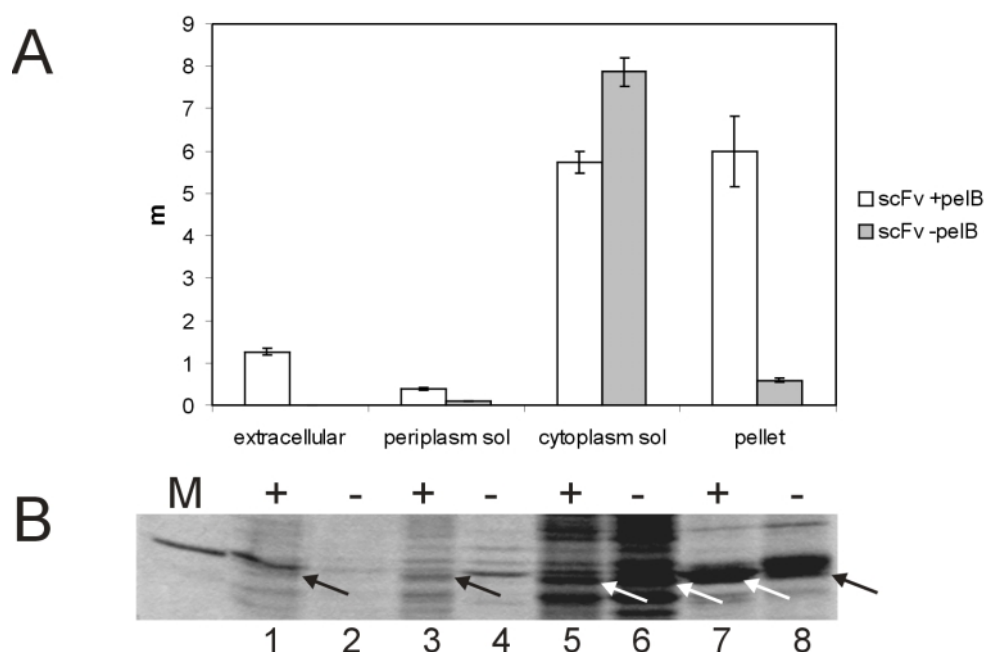


**Figure A 7.** Cell extract of *E. coli* cells expressing scFv were prepared by sonification. The soluble fraction was diluted in 80 mM PBS buffer and applied to hapten-immobilized ELISA.

Consequently, it was not possible to deduce absolute values of active scFv directly from ELISA data, but rather to make qualitative comparative observations. However, when applying different dilutions and taking the slope in the linear range, an estimation of the amount of active scFv could be made.

### A 3.2.3.2.2 Application of the Hapten-immobilized ELISA for ScFv Analysis

The established activity assay was now applied to the examination of the scFv expression. To investigate the amounts of functional protein within different fractions, the following samples were prepared as recommended by Novagen and applied to ELISA: the extracellular fraction (culture supernatant), the soluble fraction of the periplasmic extract, the soluble fraction of the cytoplasmic extract, prepared from protoplasts centrifuged after periplasmic extraction, and the resuspended pellet, which contained insoluble protein of both cell compartments and the medium.. In this case, in order to obtain the different cell fractions. As mentioned above, different dilutions were applied to ELISA and the slopes in the linear range were used to estimate the amount of active scFv (Figure A 8A). An SDS polyacrylamide gel of the respective samples is shown in Figure A 8B.



**Figure A 8.** Distribution of active recombinant protein in different cell fractions.

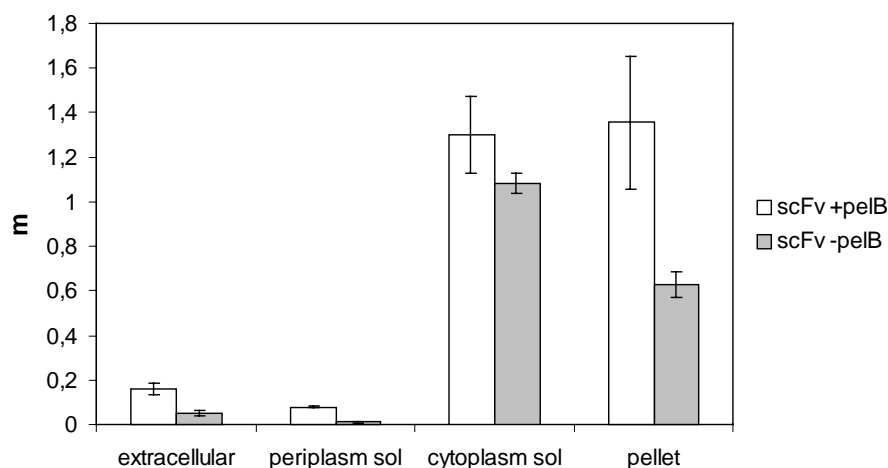
**A.** Binding of scFv was determined by hapten-immobilized ELISA. To account for the non-linear relationship of ELISA signals and applied protein mentioned above, different sample dilutions were used. m: slope (corrected according to the respective dilutions) in the linear range of the respective curves.

**B.** SDS-PAGE of the different fractions. M: carbonic anhydrase (31.0 kDa), lanes 1 and 2: culture supernatant, lanes 3 and 4: soluble periplasmic fraction, lanes 5 and 6: soluble cytoplasmic fraction, lanes 7 and 8: insoluble fraction, + : scFv with pelB, -: scFv without pelB. The bands at 29 kDa corresponding to the scFv are indicated by arrows.

In the medium and the periplasmic fraction from the expression of scFv with pelB leader bands are visible at 29 kDa after SDS-PAGE (Figure A 8B, lanes 1 and 3), which do not appear in cytoplasmic expression (lanes 2 and 4).

Interestingly, when using the sample preparation procedure as recommended by Novagen for the preparation of periplasmic, cytoplasmic and insoluble fractions, different results were obtained than when simply sonifying cells as shown in Figure A 6. Now in the soluble cytoplasmic fraction a stronger band for cytoplasmic expression (lane 6) was visible than for periplasmic expression (lane 5), indicating a solubility of the scFv expressed in the cytoplasm of about 50%. For periplasmic expression the main fraction was found in the pellet (lane 7) indicating that the solubility was about 10-20% as previously shown in Figure A 6.

When summing up the activities estimated by ELISA (Figure A 8A) in all fractions, the total activity in the strain expressing scFv in the periplasm (pETK(+)) was 1.6 times higher than in the strain expressing scFv in the cytoplasm (pETK(-)). Compared to the total expression level, which was 2.3 times higher for pETK(+), this is lower than expected. Either the activity was lower for pETK(+) or higher for pETK(-) than expected. The distribution of the activities in the different fractions differed significantly. For pETK(+), clear signals were obtained in the extracellular fraction (9% relative activity with the sum of all fractions being 100%), the periplasmic fraction (3%), the cytoplasmic fraction (43%) and the pellet (45%). For pETK(-), no activity was detectable in the medium and almost none in the periplasmic fraction (1%). The relative activity was significantly higher in the cytoplasm (93%) and much lower in the pellet (6%) compared to the strain that expressed scFv with pelB leader. As the cytoplasm of *E. coli* does not favor disulfide bond formation, it is likely that the scFv in the cytoplasmic fraction and the pellet obtained activity by spontaneous refolding and oxidation during sample preparation and during resuspension in lysis buffer, respectively. To investigate if the amount of soluble protein increased with time, all samples were incubated at room temperature for three days and remeasured. The results are shown in Figure A 9.



**Figure A 9.** Measurements were done according to Figure A 8A, but after incubation of the samples at 20°C for three days.

The absorbances and the resulting slopes obtained in both measurements differed significantly. This shows the low reproducibility of the hapten-immobilized ELISA if it is not performed under exactly the same conditions due to the many incubation steps. Additionally, still intact cells in the samples might have continued growing and affected the results. This could be responsible e.g. for the signal, that now appeared in the extracellular fraction for pETK(-). However, an interesting observations was made: The activity in the pellet fraction from the periplasmic expression of scFv remained the same compared to the respective soluble fraction of the cytoplasm. In contrast, for cytoplasmic expression the pellet fraction gained activity in comparison to the soluble fraction of the cytoplasm (35% relative activity compared to 6% in the measurement directly after sample preparation). Additionally, it was significantly higher than the activity in the extracellular fraction after periplasmic expression, whereas it was clearly lower than this in the first measurement, excluding a measuring error.

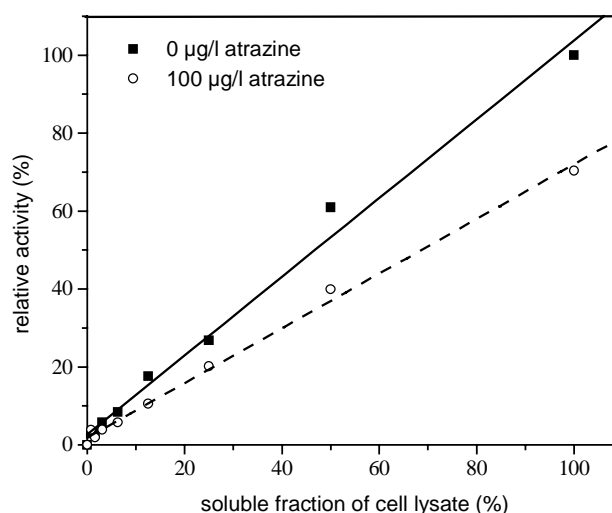
### A 3.2.3.3 Determination of Active ScFv-EGFP Fusion Protein Using Fluorescence Measurement and Non-competitive Hapten-immobilized FLISA

#### A 3.2.3.3.1 Development of the Non-competitive Hapten-immobilized FLISA

A fusion protein of the scFv K411B and EGFP was created in order to use the fluorescence as a means for online monitoring of the production process. However, it should be tested if the fusion protein could be used for a hapten-immobilized fluorescence immunoassay. This assay was named fluorophor linked immunosorbent assay (FLISA) The basic idea of the assay is to

measure the fluorescence of the EGFP and by that determine the amount or the activity of an antibody fragment. The procedure to perform the assay is described in detail in the Materials and Methods part (A 2.6.10.3) and shown schematically in comparison to the hapten-immobilized ELISA described above in the Discussion. In comparison to the 5 hours necessary to perform the hapten-immobilized ELISA, this assay required merely one incubation step and therefore the assay time was only 1.5 hours.

When applying different dilutions of extracts of *E. coli* BL21(DE3) cells expressing scFv-EGFP to the FLISA, a good linear correlation of the amount of fusion protein and the observed fluorescence signal was observed either without atrazine for competition or with different constant atrazine concentrations. Figure A 10 exemplarily shows a representative curve obtained by FLISA without and with 100 µg/l atrazine.



**Figure A 10.** Hapten-immobilized FLISA with different amounts of scFv-EGFP. 100 µl water or 100 µl of an atrazine solution (100 µg/l (m/v) atrazine dissolved in water) and different dilutions of the soluble fraction of a cell lysate of BL21(DE3)/pETKE(-), expressing the scFv-EGFP fusion protein were added to each well of the microtiter plate and the assay was performed as described in Material and Methods. The normalized fluorescence signals were plotted against the content of lysate and linear regressions were fitted with the Origin 5.0 software. The observed correlation without atrazine can be expressed as  $y = 2.72847 + 1.01047 x$  ( $R^2 = 0.9952$ ) and with 100 µg/l atrazine as  $y = 1.84673 + 0.70178 x$  ( $R^2 = 0.9978$ ).

At an atrazine concentration of 100 µg/ml the slope decreased and at concentrations higher than 10,000 µg/l, the slope approached 0 (data not shown). This confirms that the fluorescence depends on the specific but not the unspecific binding of the fusion protein to the

microtiter plate. These results show that the FLISA can be applied to determine bifunctional fusion protein in complex mixtures such as cell extracts.

#### ***A 3.2.3.3.2 Application of the Non-competitive Hapten-immobilized FLISA and Fluorescence Measurement***

The FLISA was applied to examine different cell fractions of *E. coli* BL21(DE3) cells expressing the scFv-EGFP fusion protein without pelB leader (pETKE(-)). Fluorescence was also measured in these samples to investigate if fluorescence correlated with the FLISA signal and therefore was caused by the fusion protein.

Induced cells were harvested and fractionated according to the method from Novagen. Then the fluorescence of the prepared samples was measured directly and the bifunctional activity was monitored using the FLISA. The results are shown in Table A 3.2.

**Table A 3.2.** Fluorescence and FLISA signals obtained from fractionated *E. coli* BL21(DE3) cells expressing the scFv-EGFP fusion protein without pelB leader.

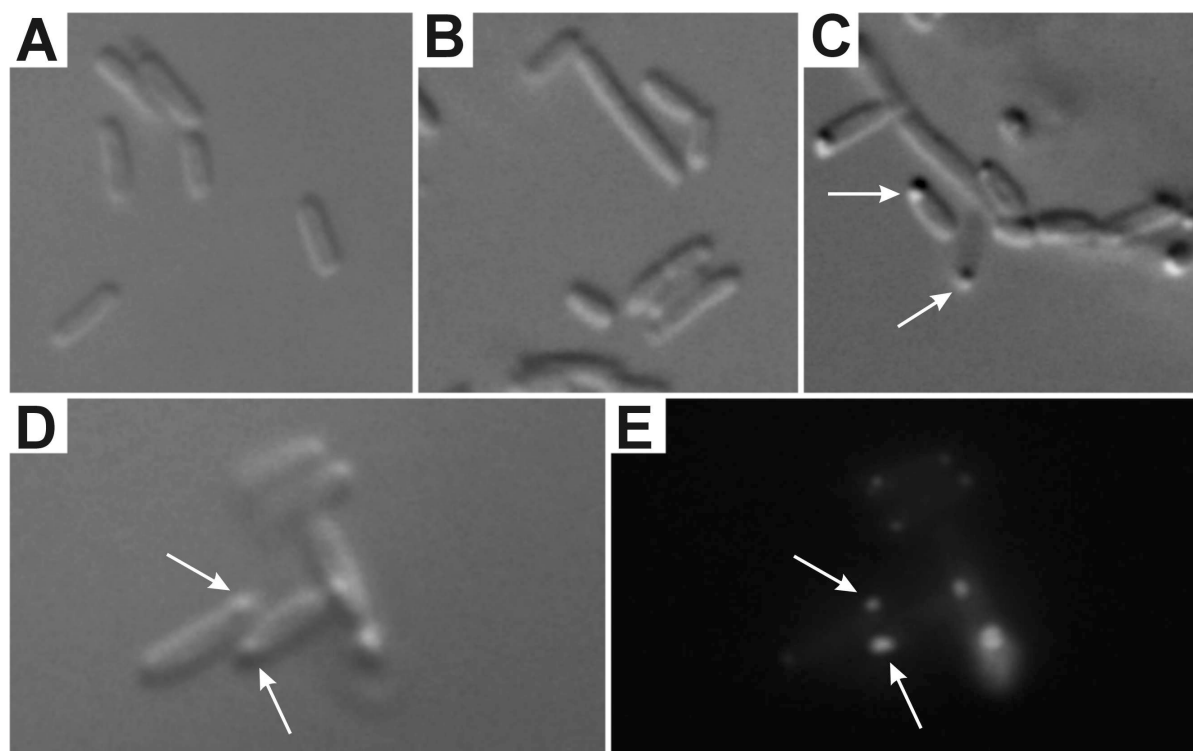
<b>Cell Fraction</b>	<b>Absolute Fluorescence (rfu)</b>	<b>Relative Fluorescence (%)</b>	<b>Absolute FLISA Binding (rfu)</b>	<b>Relative FLISA Binding (%)</b>
extracellular	48460	40	0	0
periplasm sol	5724	5	0	
cytoplasm sol	50232	42	296	69
pellet	16157	13	133	31

Results obtained with cells expressing the scFv-EGFP fusion protein without pelB leader are similar to those with cells expressing the scFv without pelB measured by ELISA (Figure A 8 and A 9). Binding activity of the scFv-domain, measured by FLISA, was only found in the soluble fraction of the cytoplasm (69%) and the pellet (31%). Fluorescence could also be detected in the medium and the periplasmic fraction, but as no binding activity could be observed, the signals were certainly not emitted by the bifunctional fusion protein. When only considering the fluorescence in the cytoplasm and the pellet, they contributed 76% (cytoplasm) and 24% (pellet) and closely corresponded to the values obtained FLISA. However, FLISA was more suitable to determine bifunctional fusion protein.



### A 3.2.4 Microscopic Investigation of *E. coli* Cells Expressing the ScFv with and without pelB Leader and the ScFv-EGFP Fusion Protein without pelB Leader

*E. coli* BL21(DE3) cells expressing the respective proteins were examined by light microscopy and the cells expressing the scFv-EGFP fusion protein additionally by fluorescence microscopy. The obtained images are shown in Figure A 11.



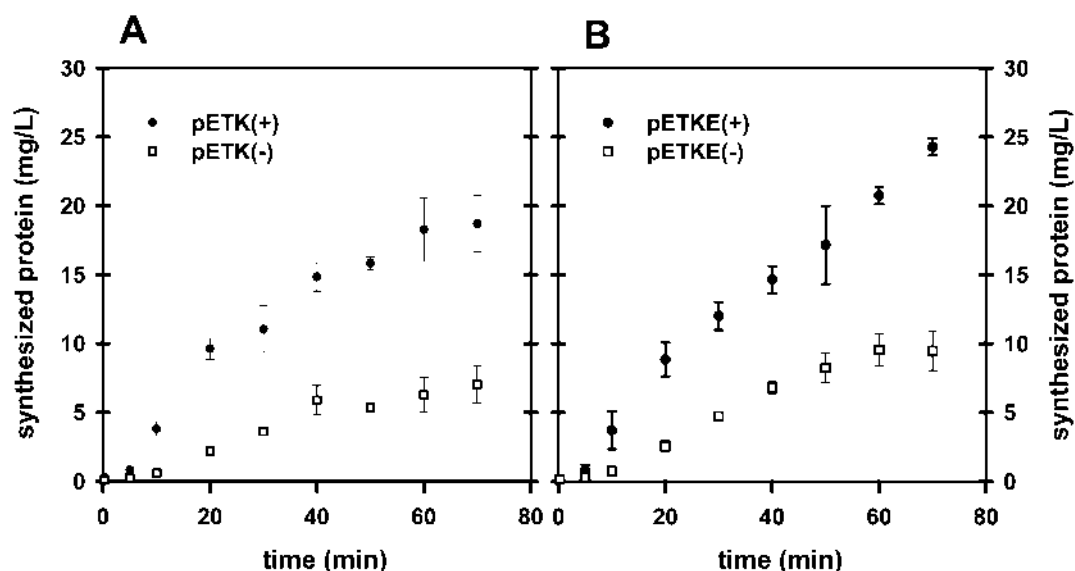
**Figure A 11.** Images of *E. coli* BL21(DE3) cells expressing scFv K411B constructs. A: Control, BL21(DE3)/pET20b(+), B: BL21(DE3)/pETK(+), C: BL21(DE3)/pETK(-), D and E: BL21(DE3)/pETKE(-). D and E depict the same cells. Arrows indicate the knobs which are discussed in the text.

In light microscopy (Fig. A 12A-D), the phenotypes of BL21(DE3) cells expressing the scFv with pelB (pETK(+)) and BL21(DE3) cells transformed with the empty vector pET20b(+) were alike. Periplasmic inclusion bodies were not observed (Fig. A 11B), indicating that insoluble processed scFv was evenly distributed in the periplasm. For cells expressing the scFv without pelB (pETK(-)), however, knobs were visible at one or either end of the cylindrically-shaped cells (Fig. A 11C). These knobs probably consist of cytoplasmic inclusion bodies of malfolded recombinant protein. They contribute ~5% to the total volume of the cells. From SDS-PAGE (Figure A 5) it was estimated, that for cytoplasmic expression

the heterologous protein contributed about the same percentage to the total protein. This implies that inside the living cell indeed most of the recombinant protein is insoluble and not functional. This is in good agreement with the results from SDS-PAGE of soluble and insoluble fractions after sonification of *E. coli* cells expressing the fusion protein, where about 90% of the recombinant protein were insoluble (Figure A 6). On the other hand, activities were high and protein bands after SDS-PAGE relatively strong after the extraction procedure suggested by Novagen (Figure A 8). After the microscopic examination it can be stated that this effect must be due to events occurring after cell disruption and the *in vivo* value for the solubility of the proteins expressed in the cytoplasm can be determined as no more than 10% in intact cells.

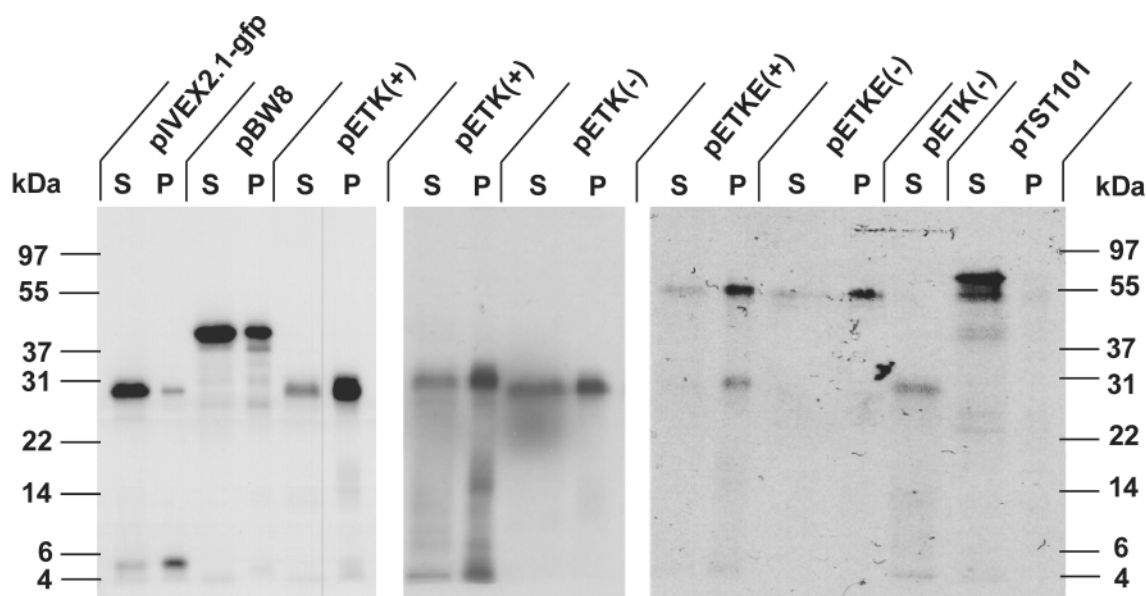
### **A 3.3 Expression of the ScFv K411B and the ScFv-EGFP Fusion Protein in Cell-free Coupled Transcription/Translation**

Cell-free coupled transcription/translation (*in vitro*) expression experiments were performed by the cooperation partners Martin Siemann-Herzberg and coworkers at the Institute of Biochemical Engineering, University of Stuttgart. The T7 promoter/RNA polymerase expression system was used. *In vitro* protein synthesis was performed for 60 min. At intervals of 10 min the incorporation of  $^{14}\text{C}$ -leucine into the synthesized protein was measured autoradiographically and thus the concentration of accumulated protein was determined. In Figure A 12 typical curves monitoring the production of protein in such a system are shown exemplarily for the constructs pETK(+), pETK(-), pETKE(+) and pETKE(-). For the constructs with pelB leader sequence higher final protein concentrations were obtained than for the respective constructs without pelB leader (19 mg/l for pETK(+)) and 24 mg/l for pETKE(+) compared to 7 mg/l for pETK(-) and 10 mg/l for pETKE(-).



**Figure A 12.** Protein expression kinetics of cell-free coupled transcription/translation. **A** pETK(+) and pETK(-), coding for the scFv K411B with and without pelB leader. **B** pETKE(+) and pETKE(-), coding for the scFv K411B-EGFP fusion protein with and without pelB leader. Experiments were done twice, standard deviations are indicated.

For these constructs also the soluble and insoluble fractions after the cell-free protein synthesis were prepared by centrifugation. Figure A 13 shows the soluble and insoluble fractions of pETK(+), pETK(-), pETKE(+) and pETKE(-) together with the control plasmids pIVEX2.1-gfp, pBW8 and pTST101 after SDS-PAGE.



**Figure A 13.** Qualitative analysis of cell-free protein synthesis. Samples were loaded onto SDS-polyacrylamide gels after 60 min of coupled *in vitro* transcription/translation synthesis and autoradiographed. The molecular masses (in kDa) of the proteins are indicated on the left and the right. S (supernatant) and P (pellet) indicate the soluble and insoluble fractions of each batch reaction. pIVEX2.1-gfp, pBW8 and pTST101 are control plasmids.

For all the constructs encoding the scFv or the scFv-EGFP fusion protein the major fraction of protein was found in the pellet fraction and only ~10% in the soluble fraction. For the proteins encoded by the control plasmids a higher percentage of protein was found in the soluble fraction. Using the same procedure a number of other plasmids were investigated. The results are summarized in Table A 3.3. For a detailed description of the respective plasmids see Figure A 4 and Table A 3.1.F

**Table A 3.3.** Plasmids expressed *in vitro* with the final concentration of accumulated protein and percentage of the soluble fraction.

Plasmid	Encoded Gene	Final Protein Concentration (µg/ml)	Soluble Fraction (%)
pETK(+)	scFv with leader	19	8
pETK(-)	scFv without leader	7	11
pETK(-)-AC	scFv without leader	3	n.d.*
pETK(-)-AC +AAA	scFv without leader	6	n.d.*
pETK(-)-AC +AAAAAA	scFv without leader	15	n.d.*
pJOEK(-)-AT	scFv without leader	2 **	n.d.*
pETKE(+)	scFv-EGFP with leader	24	12
pETKE(+)+AC	scFv-EGFP with leader	22	n.d.*
pETKE(-)	scFv-EGFP without leader	10	10
pETKE(-)-AC	scFv-EGFP without leader	10	n.d.*

\* not determined

\*\* For pJOEK(-)-AT the T7 promoter upstream of the *rhaBAD* promoter was used for cell-free expression.

### A 3.4 Comparison of the *in vivo* and *in vitro* Results of the Expression of ScFv and ScFv-EGFP Fusion Protein with and without pelB Leader Sequence

Four proteins encoded by the four plasmids pETK(+), pETK(-), pETKE(+) and pETKE(-) should be expressed *in vivo* and *in vitro*: the scFv K411B with and without pel leader sequence and the respective scFv-EGFP fusion protein with and without pelB leader sequence. ScFv-EGFP with pelB leader sequence could not be expressed in *E. coli*

BL21(DE3) cells, although it could successfully be expressed *in vitro*. The other three proteins were compared with regard to their total expression level and to the percentage of soluble and insoluble protein.

As mentioned above, observations made concerning the solubilities were contradictory for the proteins expressed in the cytoplasm. From microscopy of intact cells it was estimated that ~10% of the recombinant protein were soluble. In contrast, after the cell extraction procedure by Novagen this protein seemed to be partially refolded and contributed high relative activities compared to the activity measured in all cell fractions (~90% for the scFv (Figure A 8A) and ~70% for the scFv-EGFP fusion protein (Table A 3.2)). It seems possible, that protein from inclusion bodies had been refolded during the sample preparation procedure. As this information is interesting with regard to production, the respective values are also included in a comparison of *in vivo* and *in vitro* results listed in Table A 3.4. For cytoplasmic scFv it was observed, that ~50% of soluble protein gave ~90% of the activity, whereas for periplasmic expression 10% soluble protein contributed 43% of the activity (Figure A 8). Therefore it was estimated that the 70% relative activity of the scFv-EGFP fusion protein were due to ~25% soluble fusion protein.

In Table A 3.4 the data collected as described above are summarized. Expression levels and percentages of soluble protein *in vitro* were taken from Table A 3.3. Expression levels *in vivo* are taken from SDS-PAGE (Figure A 5). Percentage of soluble protein in intact cells were observed in microscopy (Figure A 11) and percentages of soluble protein in cell extracts are the values estimated from SDS-PAGE, ELISA and FLISA data after sample preparation according to Novagen (Figure A 8, Table A 3.2).

The data obtained for the expression levels of those proteins that could be expressed in *E. coli* BL21(DE3) are in correlation to the expression levels of the respective proteins *in vitro*. The percentages of soluble recombinant protein estimated in the cells expressing cytoplasmic scFv and scFv-EGFP in microscopy, that of scFv expressed in the periplasm in cell extracts and those of all four proteins expressed *in vitro* are in accordance. The higher percentages of scFv and scFv-EGFP in the soluble cytoplasmic fraction after sample preparation following cytoplasmic expression do not fit into the scheme and will be discussed.

**Table A 3.4.** Total expression levels and percentage of soluble heterologous protein *in vitro* and *in vivo* (BL21(DE3)).

Plasmid	pETK(+)	pETK(-)	pETKE(+)	pETKE(-)
Protein	scFv with pelB leader	scFv without pelB leader	scFv-EGFP with pelB leader	scFv-EGFP without pelB leader
expression level <i>in vitro</i> (µg/ml)	19	7	24	10
percentage of soluble fraction <i>in vitro</i> (%)	8	11	12	10
expression level <i>in vivo</i> (% of total protein)	12	5	n.d.*	5
percentage of soluble fraction in intact cells (%)	n.d.*	~10	n.d.*	~10
percentage of soluble fraction in cell extracts (%)	~10	~50	n.d.*	~25

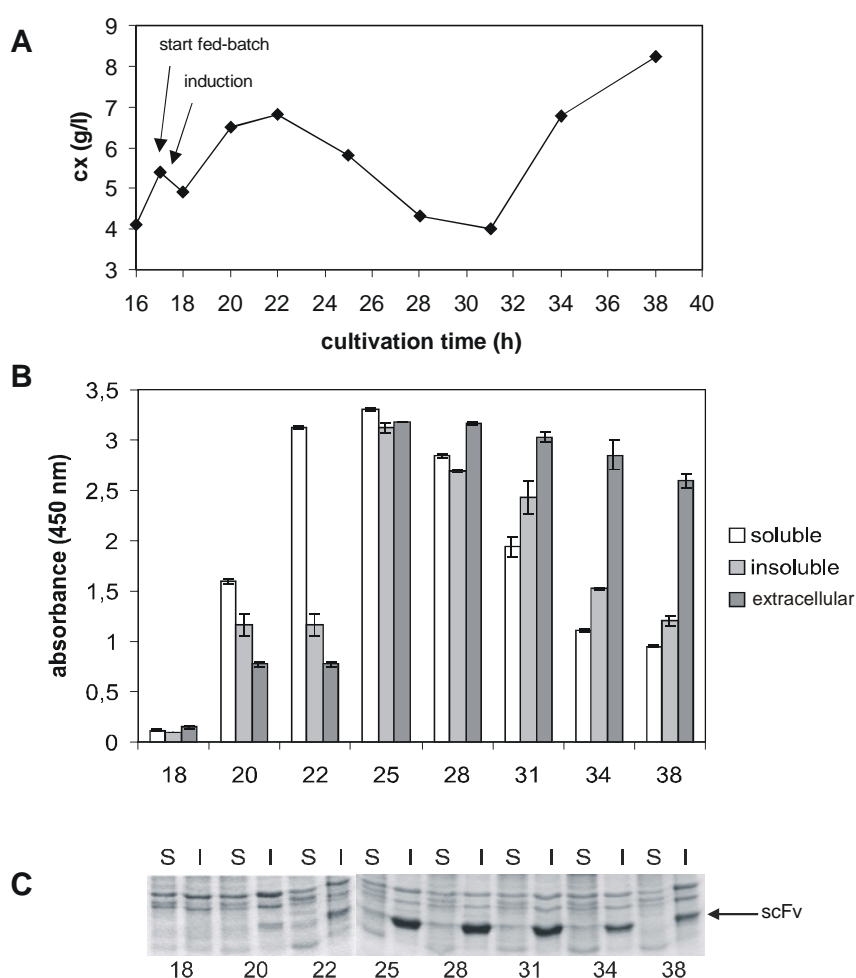
\* not determined.

### A 3.5 Expression of scFv with pelB Leader Sequence and scFv-EGFP Fusion Protein without pelB Leader Sequence in *E. coli* BL21(DE3) pLysS in 2 l Fed-batch Cultivation

To investigate how cells expressing the scFv and the scFv-EGFP fusion protein behave in a larger scale, two 2 l fed-batch cultivations with *E. coli* BL21(DE3) pLysS cells carrying selected constructs were performed as described in the Materials and Methods part (A 2.8.3.1.2). *E. coli* BL21(DE3) pLysS produces lysozyme for tight regulation of the T7 promoter. For expression of the scFv, the plasmid pETK(+) was chosen, because it gave the best expression level *in vivo* (Figure A 5) as well as *in vitro* (Table A 3.3). For expression of the fusion protein the plasmid pETKEHS(-)-AC lacking the pelB leader sequence was picked out, because periplasmic expression was not possible. Additionally, the very closely related plasmid pETKE(-)-AC gave best expression rates for the scFv-EGFP fusion protein without pelB *in vitro* (together with pETKE(-), Table A 3.3). pETKEHS(-)-AC only differed from pETKE(-)-AC in the His<sub>6</sub>-tag-coding region at the 3'-end, which was not believed to influence the expression level. One aim of the fed-batch cultivation was to produce a big

amount of fusion protein with His<sub>6</sub>-tag for purification with immobilized metal affinity chromatography (IMAC). This was intended for further characterization of the fusion protein.

The dry cell mass  $c_x$  during the cultivation of BL21(DE3) pLysS/pETK(+) is shown in Figure A 14A. At the end of the cultivation 8 g/l were obtained. At several time points after induction, samples were taken, cells lysed by sonification and soluble and insoluble fractions prepared. The occurrence of active scFv in these fractions and the culture supernatant over the course of the cultivation was monitored by non-competitive hapten-immobilized ELISA and is shown in Figure A 14B. SDS-PAGE of the soluble and the insoluble samples is shown in Figure A 14C.



**Figure A 14A.** Dry cell mass  $c_x$  in gramms per liter over the course of the cultivation of *E. coli* BL21(DE3) pLysS/pETK(+).

**B.** ELISA data obtained with the extracellular fraction, the soluble and the insoluble fraction of cell extracts directly before and after induction. The cultivation times in hours are indicated.

**C.** SDS-PAGE of the soluble (S) and the insoluble (I) fractions of cell extracts used in B at certain cultivation times in hours.

In the first few hours after induction, the major part of active scFv was found in the soluble fraction (Figure A 14, white bars). With some hours retardation, the amount of active scFv also increased in the resuspended pellet (Figure A 14, light grey bars) and in the extracellular fraction (Figure A 14, dark grey bars). At the same time the biomass decreased, possibly as a consequence of the overproduction of recombinant protein as well as cell lysis, which is also indicated by the ELISA signal in the medium. Cell lysis might be in part the result of the expression of lysozyme in pLysS cells. Also the overproduction of mRNA by the strong T7 promoter might impose a heavy physiological load on the cells, thus decreasing cell growth. SDS-PAGE (Figure A 15C) showed, that the maximum amount of scFv was achieved in the sample from 7 hours after induction (25 h cultivation time). This was also the sample with the highest total activity. The scFv contributed about 15% to the total cell protein as determined densitometrically and was found mainly in the pellet. 140 mg of scFv could be harvested from 1 liter of the final cultivation broth, as determined by a protein assay of the refolded scFv. Considering that at the time with the maximal accumulation of protein the scFv concentration was about four times higher than in the final broth (determined densitometrically by comparison of the insoluble fractions at 25 and 38 hours cultivation time in Figure A 14C), the production of more than half a gramm of scFv per liter can be achieved by this procedure.

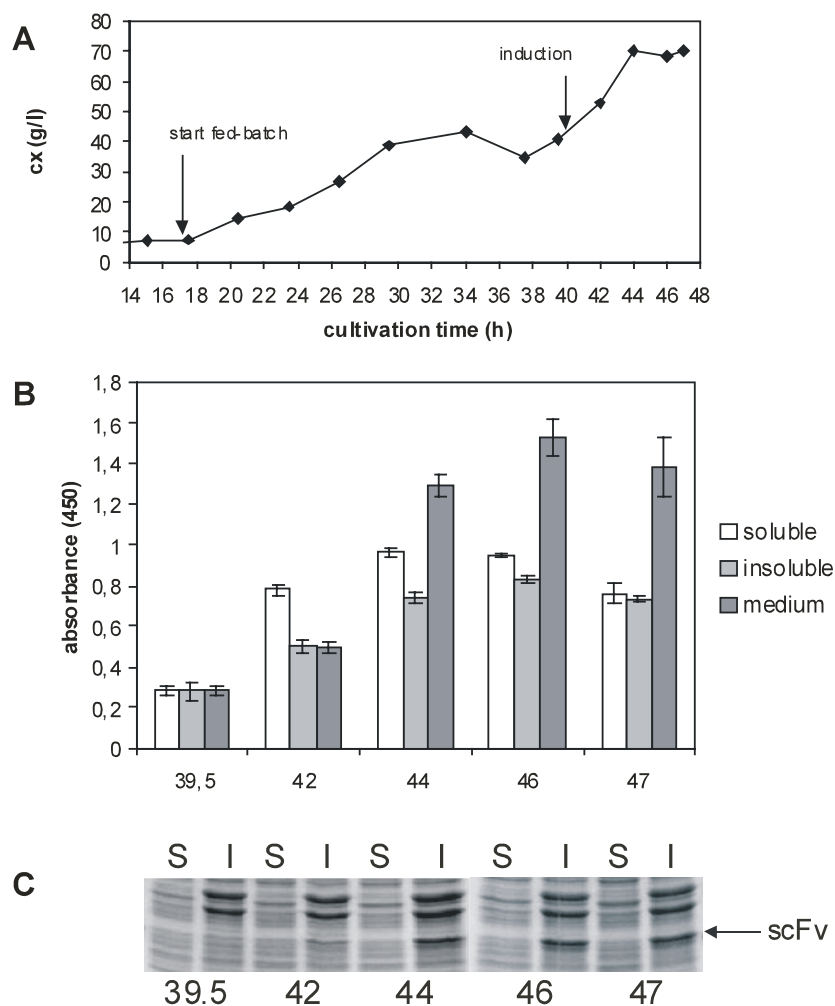
The cultivation of *E. coli* BL21(DE3) pLysS/pETKEHS(-)-AC was mainly performed for preparatory purposes to produce scFv-EGFP fusion protein and the activities were not monitored over the course of the cultivation in different fractions as for the scFv. On SDS-PAGE a distinct band was seen at 55 kDa in the insoluble fraction, contributing about 80-90% to the total heterologous protein. However, the band was significantly weaker than in the cultivation for periplasmic expression of the scFv. Bifunctional activity of the scFv-EGFP fusion protein was measured in the soluble fraction as well as in the resuspended pellet. These observations indicate a lower expression level for this protein without pelB leader than for the scFv with pelB leader.



### A 3.6 Expression of ScFv with pelB Leader Sequence in *E. coli* BW3110 in 13 l Fed-batch Cultivation

Problems during the bioprocess as observed in the 2 l cultivation (decrease in cell growth after induction) might be due to the over-production of mRNA, the production of lysozyme or the not very tightly regulated T7 promoter. For large scale production *E. coli* BL21(DE3) pLysS cells are not suitable

Therefore an alternative expression system designed for high-cell-density-cultivation was chosen, the rhamnose-deficient *E. coli* strain BW3113 in combination with the positively regulated *rhaBAD* promoter [62]. As the scFv with pelB leader sequence showed the highest expression level, this protein was chosen for the production in high-cell-density-cultivation of *E. coli* BW3110 under control of the positively regulated *rhaBAD* promoter. The gene encoding the respective protein was subcloned into the pTST101 vector after excision of the MalE-EGFP gene as described in A 3.1.1.2, resulting in the plasmid pJOEK(+)-AT. A 13 l fed-batch cultivation was performed according to Wilms *et al* [62]. Dry cell mass  $c_x$  during the cultivation of *E. coli* BW3110/pJOEK(+)-AT, ELISA data obtained as described for the 2 l fed-batch cultivations and an SDS-PAGE of the soluble and insoluble fractions are shown in Figure A 15. The cultivation yielded a final dry cell mass of 70 g per liter or about 0.9 kg in the total cultivation after 48 hours (Figure A 15A). In the final sample, the band on SDS-PAGE corresponding to the scFv was evaluated densitometrically and the concentration of scFv was observed to be about 2 g per liter or 3% of the total protein. This is five times less than the percentage of accumulated scFv observed upon expression in *E. coli* BL21(DE3) pLysS under control of the T7 promoter in the 2 l fed-batch cultivation. However, low expression levels were also observed for the pJOE plasmid derivatives when used *in vitro* (e.g. only 2  $\mu$ g/ml for pJOEK(-)-AT in comparison to 7  $\mu$ g/ml for pETK(-)-AC encoding the respective protein in the pET20b(+) derivative). Considering that in addition to the use of a different strain and promoter system the dry cell mass at the point of induction was about 8 times higher than for the 2 l fed-batch cultivation, the pattern of the ELISA data obtained from different cell fractions is similar in both cultivations (Figure A 15B and also see Figure A 14B). Again, in the first few hours the major activity was found in the soluble fraction (white bars), although the major portion of scFv was found in the insoluble fraction as shown by SDS-PAGE (Fig. A 15C).



**Figure A 15A.** Dry cell mass in grams per liter over the course of the cultivation.  
**B.** ELISA data obtained with the medium (culture supernatant), the soluble and the insoluble fraction of cell extracts directly before and after induction.  
**C.** SDS-PAGE of the soluble and the insoluble fractions of cell extracts used in B.

Activity then gradually increased in the insoluble fraction resulting in a similar final signal compared to the soluble fraction. The highest signal was obtained in the extracellular fraction 4 hours after induction. This was probably again due to cell lysis, which is not surprising because the high cell density required agitation velocities of more than 1,000 rpm to keep the oxygen partial pressure at a constant value of about 50%. Finally, by this procedure about 25 g of mainly insoluble scFv were produced.

## A 3.7 Downstream Processing of Recombinant Protein

### A 3.7.1 Refolding of Recombinant Protein

As demonstrated, expression of the scFv K411B and the scFv-EGFP fusion protein with the constructs giving the highest expression levels, i.e. those with the pelB leader, resulted mainly in the accumulation of recombinant protein as insoluble protein. Therefore, the insoluble proteins were refolded into active protein.

To achieve this, a method described by Buchner *et al.* for the efficient refolding for single-chain immunotoxins [63] was used and modified. This procedure in principle consists of two steps: the denaturation of the single-chain protein under reducing (DTE) and denaturing (guanidine) conditions and renaturation by rapid dilution under oxidizing (oxidized glutathion) and labilizing (L-arginine, prevents the aggregation of misfolded protein) conditions. This method was modified in two points:

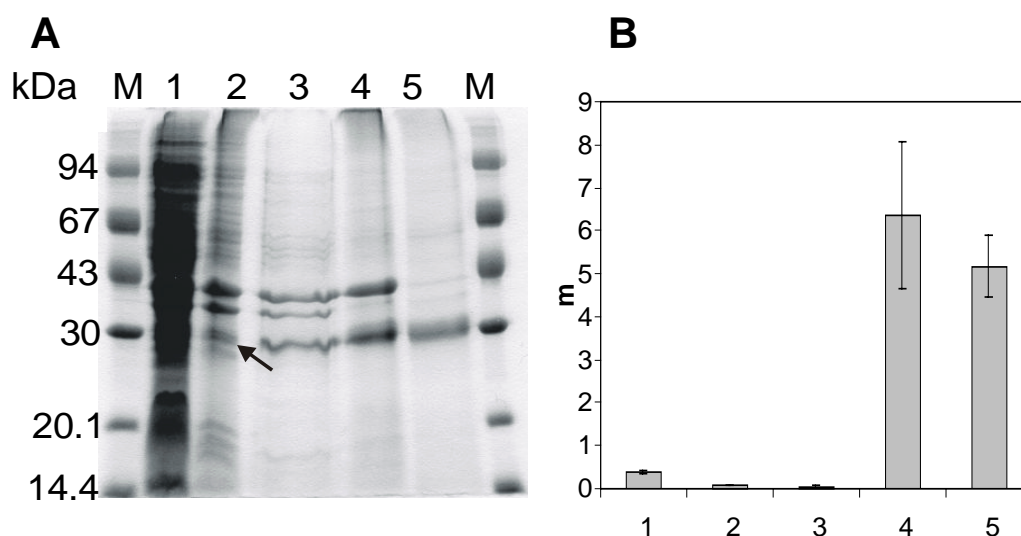
1. The denaturing buffer contained guanidine at 8 M instead of 6 M, which proved to be necessary to solubilize the scFv.
2. Dilution into refolding buffer was done 25-fold instead of 100-fold in order to save reagents and time necessary for the subsequent concentration.

The resulting refolding solution was concentrated by ultrafiltration (10 kDa cut-off) to the original volume and the filtrate could successfully be reused after addition of oxidized glutathion.

#### A 3.7.1.1 Refolding of the scFv K411B

Although significant amounts of active scFv were detected by ELISA in the culture supernatant as well as in all fractions of cell extracts of *E. coli* BL21(DE3) pLysS/pETK(+) cells obtained from the 2 l fed-batch cultivation, about 90 % of the overexpressed protein accumulated as inactive protein. This was harvested from the cultivation broth after the bioreaction by cell disruption and centrifugation. As shown by SDS-PAGE (Figure A 16A) the major part of the insoluble scFv (lane 2, arrow) was solubilized in the denaturing buffer (lane 3). Two heavier proteins were also present in the soluble fraction after denaturation. They always appeared in the insoluble fraction of *E. coli* cells, also before induction (see Figures A 5, A 14C, A 15C). After renaturation (lane 4), only two main bands remained, one

corresponding to the scFv at about 29 kDa and a heavier one at about 43 kDa. By centrifugation at 20,000 g for 5 min, the heavier protein was removed leaving almost homogeneous scFv (lane 5). Protein determination in this sample with Bio-Rad Protein Assay resulted in 1.70 mg/ml (m/v) total protein and the purity of the scFv was estimated to 85% from the gel. Consequently, about 170 mg soluble scFv of 85% purity could be obtained by this refolding method from 1 l cultivation broth. Hapten-immobilized ELISA of the different samples at different dilutions were performed and the obtained slopes in the linear range are shown in Figure A 16B.



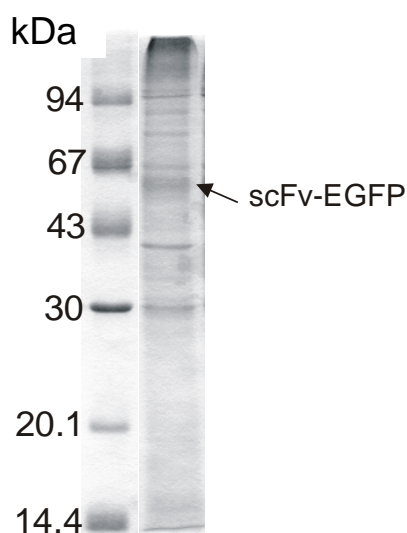
**Figure A 16A.** Coomassie Brilliant Blue stained polyacrylamide gel of different samples during the refolding procedure. Lane 1: total cell lysate, lane 2: insoluble fraction of the cell lysate, the arrow indicating the scFv, lane 3: soluble fraction after denaturation, lane 4: total renaturation preparation, lane 5: soluble fraction of the renaturation preparation, M: molecular weight marker, molecular masses in kDa are indicated on the left.

**B.** Data obtained from the respective samples in hapten-immobilized ELISA. Different dilutions were applied and the slope in the linear range is depicted.

They revealed, that activity after refolding of the insoluble scFv was 13-fold higher than in the initial total cell lysate, 65-fold higher than in the insoluble fraction of the cell lysate, which was submitted to refolding. The 13-fold increase in activity for refolded scFv in comparison to the total cell lysate is in good agreement with the estimation that initially about 90% of the scFv were insoluble and inactive. Therefore, the precipitated scFv in the insoluble pellet fraction could be solubilized and refolded nearly quantitatively by the method described above.

### A 3.7.1.2 Refolding of the scFv-EGFP Fusion Protein

For refolding of the scFv-EGFP fusion protein obtained from the 2 l fed-batch cultivation of BL21(DE3) pLysS/pETKEHS(-)-AC the same procedure was applied. It resulted in an increase of bifunctional activity by a factor of about two and was therefore much less efficient than for the scFv (factor 13). No purification was achieved using this method but only an enrichment to about 20% of fusion protein, as shown by SDS-PAGE (Figure A 17).



**Figure A 17.** SDS-PAGE of refolded scFv-EGFP fusion proteins. Molecular weight marker in kDa on the left, a sample after refolding on the right. The arrow indicates the fusion protein at ~55 kDa.

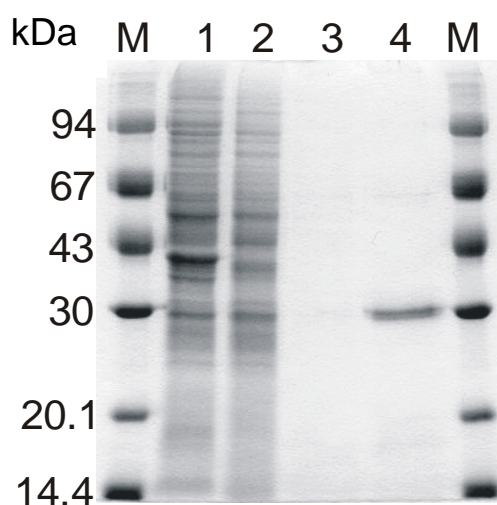
In addition, after the refolding bands were visible after SDS-PAGE at about 30 kDa. These might have resulted from partial degradation of the fusion protein during the refolding procedure. When stored at 4°C activity decreased much faster than for the scFv. After one month FLISA could be performed with the refolded scFv-EGFP fusion protein with good results, although the protein was not purified to homogeneity, but after three months the signals approximated background signal (data not shown). The reason for the instability might have been proteolytic degradation. However, also the addition of complete<sup>TM</sup> protease inhibitor did not tremendously increase the stability indicating an intrinsic instability of the protein.

### A 3.7.2 One-step Purification of Recombinant Protein by Immobilized Metal Affinity Chromatography (IMAC)

In IMAC the ability of certain amino acids to bind coordinatively to metal ions is used to bind proteins specifically to a resin in which the metal ions are immobilized. In this work the TALON resin (Clontech) containing  $\text{Co}^{2+}$  ions and the C-terminal  $\text{His}_6$ -tag of the recombinant protein, consisting of 6 consecutive histidines was used. Elution of bound protein under native conditions was done with 150 mM imidazole, which competitively displaced the histidine side chains bound to the metal ions. Washing steps were done without imidazole, as 5 mM imidazole (as suggested by the manufacturer) resulted in elution of the scFv.

#### A 3.7.2.1 Purification of the ScFv K411B

ScFv was purified from the 10-fold concentrated medium of the 2 l fed-batch cultivation of *E. coli* BL21(DE3) pLysS/pETK(+). In Figure A 18 SDS-PAGE of different samples during the purification procedure is shown.



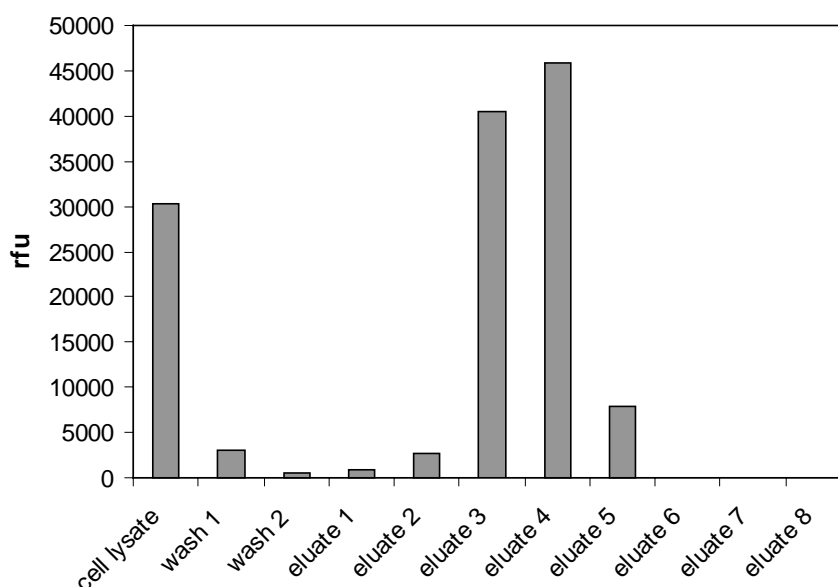
**Fig. A 18.** Coomassie Brilliant Blue stained polyacrylamide gel of different samples during the purification procedure using TALON resin. Lane 1: 10-fold concentrated medium from the fed batch cultivation, lane 2: flow through, lane 3: first wash, lane 4: elution fraction with the scFv, M: low molecular weight marker. Molecular massed in kDa are indicated on the left.

Homogeneous scFv was obtained from this one step purification procedure. 10-fold concentrated culture supernatant (Figure A 18, lane 1) from the fed batch cultivation was applied to the column (lane 2: flow through), washed (lane 3) and eluted with 150 mM imidazole at pH 7.0, yielding the scFv in high purity (lane 4). The concentration was 110  $\mu\text{g/ml}$  as determined with the Bio-Rad Protein Assay Kit. Submission of different dilutions of

the samples to hapten-immobilized ELISA revealed, that the relative activity after purification was 45% in comparison to the 10-fold concentrated medium. The protein concentration in the 10-fold concentrated medium was 11 mg/ml. Accordingly, the specific activity of the purified scFv measured by ELISA was 45 times higher than in the 10-fold concentrated medium. This procedure showed that scFv can be purified very specifically out of big volumes such as culture supernatants.

### A 3.7.2.2 Purification of the ScFv-EGFP Fusion Protein

The soluble fraction of a cell extract of BL21(DE3)/pETKEHS(+)-AC expressing His<sub>6</sub>-tagged scFv-EGFP fusion protein from a shake flask cultivation was applied to IMAC. Washing steps as well as fractions from the elution were monitored by measuring the fluorescence at 488 nm excitation and 538 nm emission. The obtained data are shown in Figure A 19.



**Figure A 19.** Fluorescence measurements of the cell lysate applied to IMAC, the washing steps and the fractions collected during elution with 150 mM imidazole. rfu: relative fluorescence units.

There was an accumulation of fluorescence in three fractions of the elution (eluate 3-5 with 40,530, 46,790 and 7,818 rfu, respectively), which was significantly higher than in the other fractions (<3,000 rfu) and also higher than in the cell lysate (30,216 rfu). However, when checked by SDS-PAGE, these fractions contained a heterogeneous mixture of proteins and no distinct band was visible at 55 kDa corresponding to the fusion protein. In FLISA no bifunctional activity could be observed in these fractions, although this was the case in the

cell lysate applied to the IMAC. Obviously, the fluorescence signals in the mentioned fractions were either due to unspecific fluorescence or due to only a part of the fusion protein after fragmentation during the purification procedure. As a conclusion, bifunctional scFv-EGFP fusion protein could not be purified by this method.

### **A 3.7.3 Characterization of the Refolded ScFv K411B**

The scFv was found to be very stable and could still be used for ELISA after one year, when stored at 4°C. Refolded scFv was electro-blotted and sequenced N-terminally (16 amino acids clearly determined). The sequence corresponded to the sequence of the mature protein, indicating that the signal peptide had been cut off and that translocation into the periplasm was efficient. The scFv was further characterized with regard to its binding specificity by ELISA.

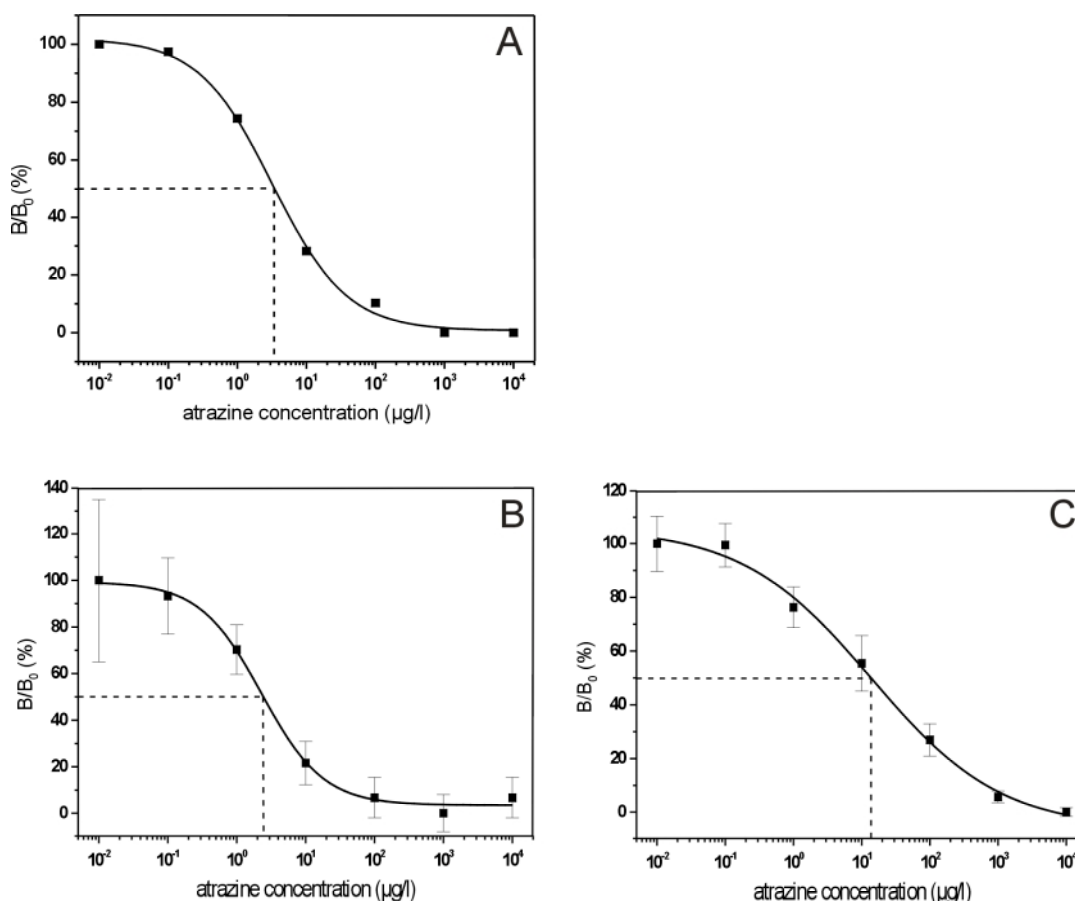
## **A 3.8 Application of the scFv K411B and the scFv-EGFP Fusion Protein in *s*-Triazine Analysis**

Above it has been shown that the non-competitive hapten-immobilized ELISA and FLISA are suitable to characterize and determine amounts of heterologous scFv and scFv-EGFP fusion protein (A 3.2.3.2). However, the recombinant proteins can also be used to determine *s*-triazines in competitive hapten-immobilized assay formats. For ELISA also immobilization of the scFv is thinkable. Both recombinant proteins were tested for their application in such competitive assays.

### **A 3.8.1 Competitive Antibody-fragment-immobilized ELISA**

In the antibody-fragment-immobilized ELISA the scFv is immobilized and the analyte and a labeled tracer compete for scFv molecules. Bound tracer, a C<sub>6</sub>-atrazine-HRP conjugate produces a signal in an enzymatic reaction after the other substances have been washed off. The scFv was immobilized either via a rabbit anti-mouse capture antibody or additionally with a secondary mouse anti-polyhistidine IgG. The ELISA was performed either with scFv directly obtained from crude cell extracts of cultures expressing the protein or with scFv refolded as described in A 3.7.1. In Figure A 20 typical calibration curves for atrazine are shown using scFv from crude cell extracts and refolded scFv with and without mouse anti-polyhistidine IgG.





**Figure A 20.** Antibody-fragment-immobilized ELISA. Atrazine was applied in different concentrations and the absorbances obtained at 450 nm were normalized as described in Materials and Methods. The sigmoidal curves were fitted with the Origin 5.0 software. The  $IC_{50}$  values were determined graphically at 50% inhibition.

**A.** ScFv from crude cell extracts, without mouse anti-polyhistidine IgG.  $A_{0.01} - A_{\text{excess}} = 0.04$ .

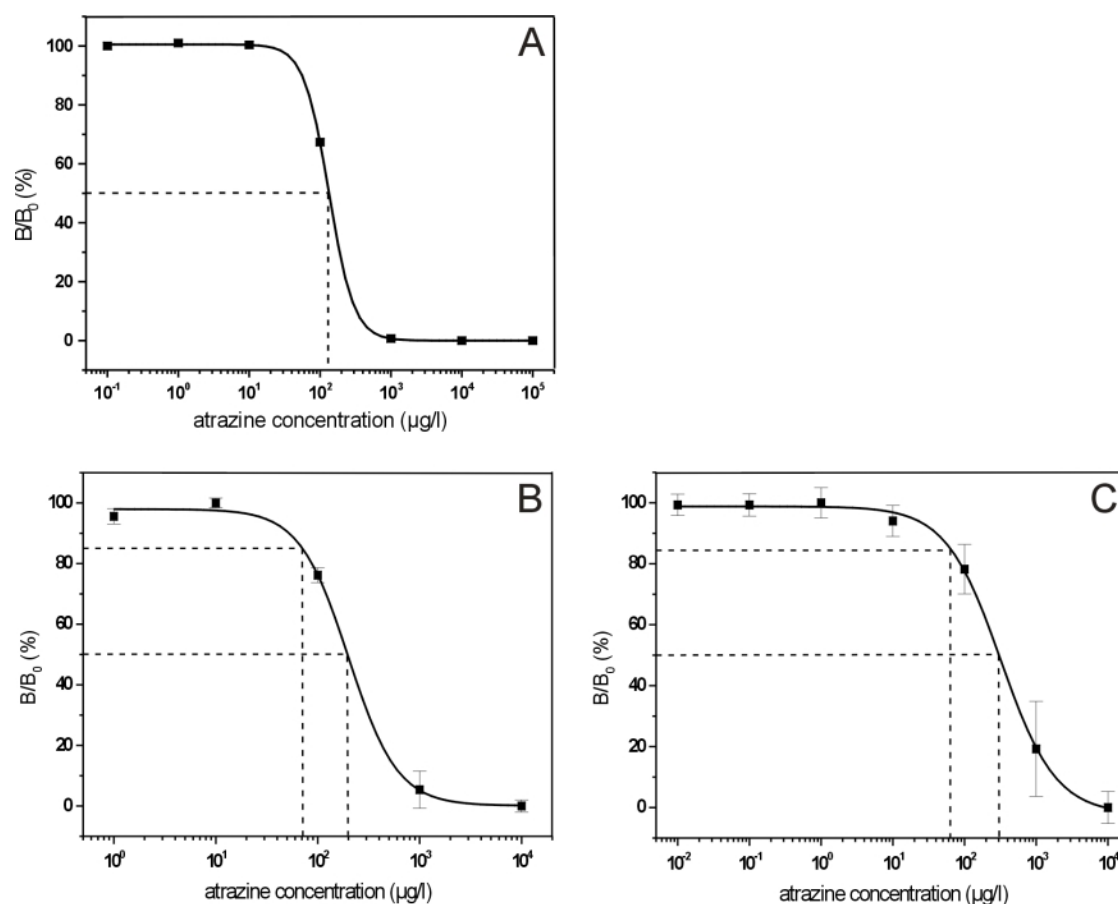
**B.** Refolded scFv diluted 1:200, without mouse anti-polyhistidine IgG.  $A_{0.01} - A_{\text{excess}} = 0.02$ .

**C.** Refolded scFv diluted 1:200, with mouse anti-polyhistidine IgG.  $A_{0.01} - A_{\text{excess}} = 0.09$ .

Obviously, the measured absorbances were very low and this is very likely also the reason for the big error bars. The lowest  $IC_{50}$ -value was obtained with refolded scFv and without mouse anti-polyhistidine IgG with 2.2  $\mu\text{g/l}$  atrazine (Fig. A 20B), followed by crude cell extract without mouse anti-polyhistidine IgG (3.2  $\mu\text{g/l}$ , Fig. A 20A). The application of mouse anti-polyhistidine IgG together with refolded scFv led to a higher  $IC_{50}$ -value (12  $\mu\text{g/l}$ , Fig. A 20C), but on the other hand to higher absorbances. Detection limits were not determined as the standard deviations were very big (Figure A 20 B and C) or because only one measurement was made (Figure A 20A).

### A 3.8.2 Competitive Hapten-immobilized ELISA

This assay format corresponds to hapten-immobilized ELISA for the characterization of scFv with the exception that here dissolved analyte competes with the immobilized hapten. C<sub>6</sub>-atrazine-BSA conjugate was used as the hapten. It was performed with (Fig. A 21A and A 21C) and without (Fig. A 21B) mouse anti-polyhistidine IgG. When using mouse anti-polyhistidine IgG, both crude cell extract (Fig. A 21A) and refolded scFv (Fig. A 21C) were tested. The calibration curves obtained for atrazine are shown in Figure A 21.



**Figure A 21.** Hapten-immobilized ELISA. The  $IC_{50}$  values and the detection limits were determined graphically at 50 and 15% inhibition, respectively.

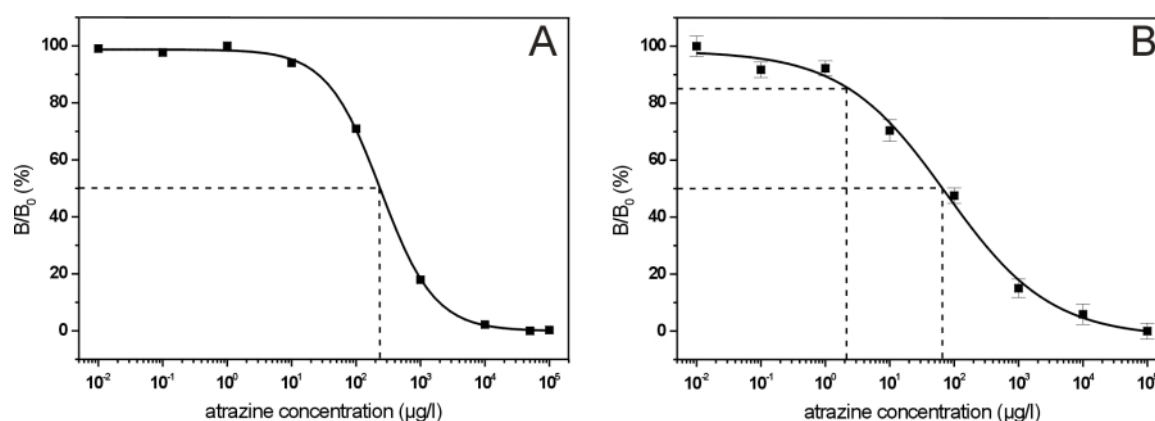
- A.** Calibration curve for atrazine with mouse anti-polyhistidine IgG using scFv from crude cell extract.  $A_{0.1} - A_{\text{excess}} = 0.3$
- B.** As A, but using refolded scFv diluted 1:200 and without the use of mouse anti-polyhistidine IgG.  $A_1 - A_{\text{excess}} = 0.04$
- C.** As B, but with the use of mouse anti-polyhistidine IgG.  $A_1 - A_{\text{excess}} = 2.3$ .

Again the assay with mouse anti-polyhistidine IgG showed clearly higher signals (Fig. A 21A and C) than without (Fig. A 21B). The  $IC_{50}$ -values (120, 200 and 300  $\mu\text{g/l}$  for Fig. A 21A, B and C, respectively) and also the detection limits defined as the atrazine concentration at 15%

inhibition (70 and 65  $\mu\text{g/l}$  for Fig. A 21B and C) were very high for all these assay formats. For Figure A 21A only one measurement was made and no detection limit determined.

### A 3.8.3 Competitive Hapten-immobilized FLISA

Hapten-immobilized FLISA with  $C_6$ -atrazine BSA conjugate being the hapten was performed with scFv-EGFP from crude cell extracts to obtain a calibration curve for atrazine (Fig. A 22A). This was a single measurement and no detection limit was determined.



**Figure A 22.** Standard curves obtained for atrazine with hapten-immobilized FLISA.

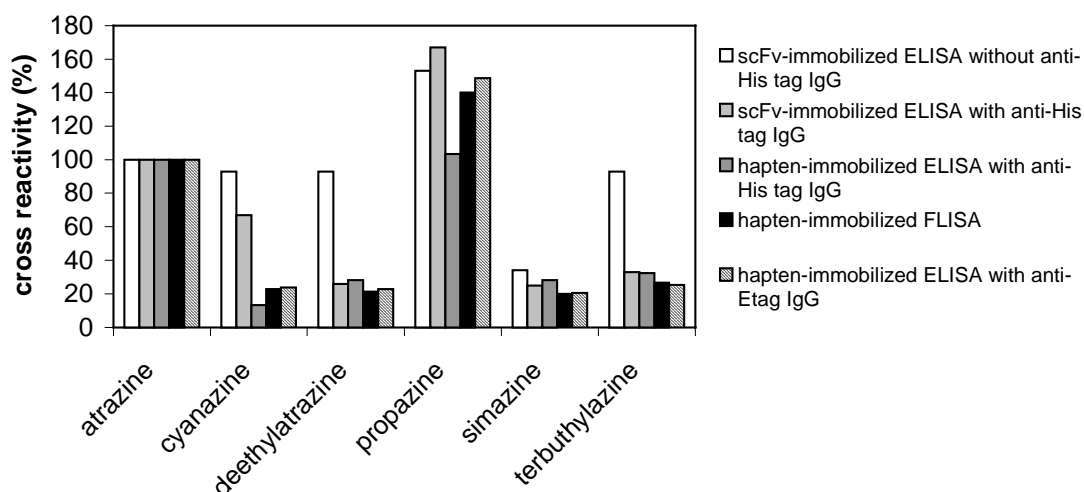
**A.** Using scFv-EGFP from a crude cell extract.

**B.** Using refolded scFv-EGFP, diluted 1:5.

In Figure A 22B the calibration curve using refolded scFv-EGFP is shown. In this case the IC<sub>50</sub>-value was lower than when using crude cell extracts (65  $\mu\text{g/l}$  (Figure A 22B) compared to 220  $\mu\text{g/l}$  (Figure A 22A)). With refolded scFv-EGFP, the detection limit was found to be 2  $\mu\text{g/l}$ .

### A 3.8.4 Cross Reactivities of the Atrazine-specific ScFv and the ScFv-EGFP Fusion Protein Towards other *s*-Triazines

Calibration curves for five other *s*-triazines (cyanazine, deethylatrazine, propazine, simazine and terbuthylazine) were prepared and cross reactivities were determined. The results obtained in the different assay formats were compared with literature data obtained with the corresponding scFv expressed with phages and are summarized in Figure A 23.



**Figure A 23.** Comparison of cross reactivities towards different *s*-triazines in different immunoassays.

The striped columns on the right show literature data obtained with the respective scFv K411B expressed with phages [23].

The concordance of the FLISA with the literature data is much better than that of the tested ELISA formats. Hapten-immobilized ELISA did not result in satisfactory data and is not represented in Figure A 24. Obviously the fusion of the EGFP domain to the scFv did not influence its binding specificity.

## **A 4 Discussion**

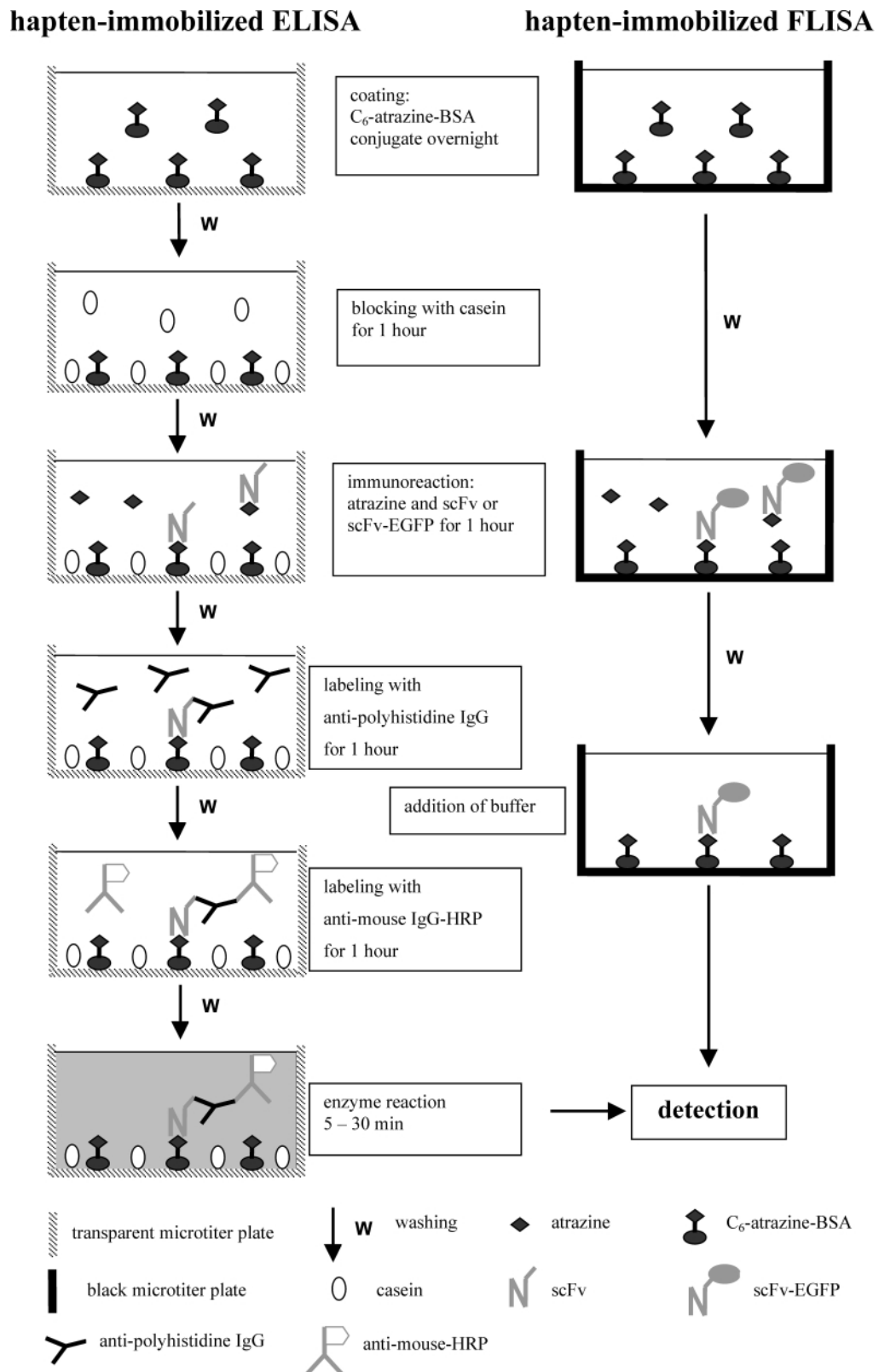
### **A 4.1 Quantification of Functional Heterologous Protein**

The aim of this work was to investigate the expression of recombinant antibody fragments with respect to the total expression level but also with respect to functionality of heterologous protein. The comparison of soluble and insoluble fractions is a way to determine the percentage of soluble protein which is usually correlated to the percentage of functional protein. However, this method was not suitable for shake flask cultivations, as the amounts of soluble heterologous protein were usually too small to allow a reliable determination. Additionally, this was made more difficult by the variety of other proteins present in cell extracts. When heterologous protein is expressed in the form of inclusion bodies, the amount of this insoluble protein can be roughly estimated by light microscopy and in the case of the scFv-EGFP fusion protein also by fluorescence microscopy. However, those methods are rather a rough estimation than a quantification. Therefore assays specific for functionality were applied for the detection of small amounts of functional recombinant protein in heterogeneous mixtures like cell extracts.

#### **A 4.1.1 Quantification of Functional ScFv and ScFv-EGFP Fusion Protein Using Hapten-immobilized ELISA and FLISA**

Non-competitive hapten-immobilized ELISA and FLISA were applied successfully to quantify functional recombinant protein. The scFv and the scFv-EGFP fusion protein investigated here were only model proteins and these assays can in principle be applied for any binding protein and any fusion of a binding protein and a fluorescent protein.

The hapten is immobilized in the microtiter plates. Applied scFv or scFv-EGFP binds to the hapten, whereas other proteins are washed off. These assays can also be used as competitive assays for the determination of *s*-triazines (A 3.8.2 and A 3.8.3). However, for quantification of functional scFv and scFv-EGFP fusion protein they can be performed without competition. The procedures to perform the two competitive ELISA and FLISA are shown schematically in Figure A 24. Non-competitive assays were performed accordingly with the exception that no herbicide was present during the immunoreaction.



**Figure A 24.** Schematic representation of the hapten-immobilizes ELISA and FLISA. Here the competitive formats are depicted. The non-competitive assay are performed accordingly with the exception that no herbicide is added in the immunoreaction.

In principle, the FLISA is a variant of the described ELISA. The main difference is the labeling technique of bound antibody fragment. Whereas in the hapten-immobilized ELISA bound scFv is labeled with anti-polyhistidine IgG, which subsequently is labeled with a goat anti-mouse antibody-HRP conjugate performing an enzyme reaction, in the FLISA bound scFv is simply labeled with the C-terminal EGFP domain, whose fluorescence can be measured.

Hapten immobilized ELISA can in principle be used for determining active scFv as well as active fusion protein. However, for the determination of the fusion protein the FLISA was preferred. The hapten-immobilized FLISA had substantial advantages over hapten-immobilized ELISA formats. The most striking difference is the time that is required for the assay. The ELISA takes at least 5 hours when starting with hapten-coated microtiter plates (which can be stored at  $-20^{\circ}\text{C}$ ); this involves a total of 4 hours of various incubation steps plus the time required for pipetting, washing, the enzyme reaction and absorbance measurement. FLISA, however, takes only 1.5 hours. This includes a 1-hour incubation period, one washing step, pipetting and fluorescence measurement. The minimal time requirements are crucial for applications such as medium and high throughput screening with crude cell extracts. Such screenings are required for the optimization of affinity or folding behavior of scFvs by approaches involving directed evolution and for monitoring production processes and downstream processes such as refolding, which usually take several days. In such cases, an assay time of only 1.5 hours can be considered as online monitoring. Apart from saving time, it should be stressed that the reduction to a single incubation step also helps avoid additional expenditures for secondary and labeled antibodies, as well as the use of toxic and acidic substances (TMB and sulfuric acid). Waldo and coworkers used the fusion of GFP to different proteins for a rapid protein-folding assay [71]. Although that was a more rapid *in vivo* assay, they only determined solubility but not specific activity of the tested proteins.

However, FLISA requires the protein of interest to be fused to EGFP. This can be achieved easily once the gene and the expression system are available. In order to conveniently clone other scFvs into the expression vector, suitable restriction sites could be introduced. In cases where the scFv but not the fusion protein is of interest after production and characterization, it is possible to introduce an enterokinase cleavage site between the scFv and EGFP domains in order to enzymatically remove EGFP. Such a procedure has been described for the separation of the two domains of a fusion protein of GFP and organophosphorus hydrolase [74].

Another advantage of the FLISA, which is of particular importance in determining the amount of functional scFv-EGFP fusion protein, is the a linear correlation between the scFv-EGFP fusion protein applied to the assay and the measured fluorescence signal. In contrast, the hapten-immobilized ELISA showed saturation already at low concentrations. The reasons for the strong correlation in the FLISA might be that each scFv domain is connected to exactly one EGFP domain and that the affinity of the scFv domain to the C<sub>6</sub>-atrazine-BSA conjugate is very high. In the ELISA on the other hand, association and dissociation of anti-polyhistidine IgGs and polyclonal anti-mouse-antibody-HRP conjugates represent equilibrium reactions and therefore depend on the concentrations of the binding partners, leading to a saturation at high levels of bound scFv. Furthermore, the enzyme reaction is not only limited by the enzyme concentration, but also by the diffusion of the substrate and the product. However, experiments carried out with different dilutions of scFv and different incubation periods during the enzyme reaction suggest, that the enzyme reaction was not the limiting step, but rather the binding of the secondary antibody or the antibody-HRP conjugate. Insufficient binding of the scFv is not very likely, because the respective scFv-EGFP fusion protein, which showed very similar binding properties as the scFv, did not show such a saturation even at high concentrations. In other ELISA formats, such as the sandwich ELISA, the amplification of the signal by secondary antibodies and the subsequent enzymatic staining reaction are necessary to increase the sensitivity towards the analyte. In this competitive immunoassay, however, this is not necessary and for the quantification of bound protein even more of a hindrance.

It should be stressed that the atrazine-specific K411B scFv domain and the EGFP domain used in the FLISA presented in this work are just one example. It would certainly be desirable to show that FLISA can be used equally well with other fusion proteins that consist of a binding protein and a fluorescent protein domain. The fusion protein of a lipopolysaccharide-specific scFv and a GFP mutant that was shown to function well in flow cytometry and immunofluorescent cell staining of *Ralstonia solanacearum* [76] could be a good candidate. Casey and coworkers reported on the staining of COS7-cells transfected with a hepatitis B surface antigen with an EGFP-scFv fusion protein [20]. For protein engineering and online monitoring of the production process of such fusion proteins, the FLISA could be the method of choice.

Although the hapten-immobilized ELISA did not show a linear correlation between the applied scFv and the measured absorbance, a rough estimation of the amount of active scFv



could be made by applying different scFv dilutions and taking the slope in the linear range. However, this increases the number of wells necessary for the measurements and also the time required for the evaluation of data. Also, the decision about where to consider a curve of a limited number of points linear or not might be quite arbitrary.

## **A 4.2 Expression Analysis of the ScFv K411B and the ScFv-EGFP Fusion Protein in *E. coli***

The intension of this work was to express the scFv K411B gene, coding for an atrazine-specific antibody fragment, that can be applied for the detection of the herbicide in ground and drinking water [23], heterologously in *E. coli* with (pETK(+)) and without (pETK(-)) the pelB leader sequence and with (pETKE(-)) and without (pETK(-)) the C-terminal fusion of the EGFP at high levels. In doing so, it was observed that the yields of total and functional protein differed for the respective constructs. To further elucidate this finding, the yields of heterologous protein encoded by the different constructs and the proportion of soluble and insoluble scFv and scFv-EGFP fusion protein were also determined *in vitro*.

*In vivo*, the total expression level and the expression level of functional protein for scFv with and without pelB leader sequence were determined. The total amount of scFv accumulating after expression of the scFv precursor was 2.3 times, the total activity 1.6 times higher than after expression of mature scFv when summing up the activities measured in different fractions after cell extraction. The major portion of the precursor was translocated into the periplasmic space and processed to mature protein as shown by N-terminal sequencing, followed either by partial cell lysis or secretion of scFv into the culture supernatant. Consequently, translocation into the periplasm did not seem to be a limiting factor in this case. Simmons and Yansura have reported that for some proteins too high translational levels can impede secretion of heterologous protein [47]. Most of the protein in the periplasm formed insoluble periplasmic protein that was found in the insoluble fraction after centrifugation. However, no periplasmic inclusion bodies were visible in light microscopy. This is somehow surprising, because formation of periplasmic inclusion bodies has been reported [101], but it can be explained with the fact that the protein is secreted at various locations of the inner cell membrane and forms insoluble protein prior to aggregating together with a multitude of other molecules to inclusion bodies which are visible in light microscopy. A relatively small fraction of active scFv was present in the soluble fraction of the cytoplasmic extract, whereas the majority of the protein was found in the pellet. This

protein gave a signal in ELISA, which was surprising, because ELISA data represent only active protein. Probably this observation was due to spontaneous refolding.

No translocation into the periplasm occurred in cells expressing scFv and scFv-EGFP fusion protein without leader (pETK(-) and pETKE(-)). In contrast to scFv with leader (pETK(+)), the amount of active protein in the soluble fraction of the cytoplasmic extract was significantly higher than in the resuspended pellet. For the scFv and the scFv-EGFP fusion protein without leader (pETK(-) and pETKE(-)), the soluble cytoplasmic fraction constituted ~90% and ~70% of the total activity, respectively. As the cytoplasm of *E. coli* does not favor disulfide bond formation, it is likely that the scFv and the scFv-EGFP in the cytoplasmic fraction and the pellet obtained activity by spontaneous refolding and oxidation during sample preparation and during resuspension in lysis buffer, respectively. Using microscopy, in intact cells expressing the proteins without leader, inclusion bodies contributing ~5% to the total cell protein were observed. This is in good agreement with the percentage determined for the total recombinant protein by SDS-PAGE (Figure A 5, lane 4) and supports the possibility of spontaneous refolding after cell disruption. Although periplasmic aggregates were also present during the preparation of the cytoplasmic extract of cells expressing the scFv in the periplasm and additionally the total expression level of this protein was higher, the soluble cytoplasmic fraction showed only low relative activity (43%). The observations suggest that cytoplasmic aggregates can be renatured more easily in lysis buffer than periplasmic aggregates. This might be due to the lack of disulfide bond formation in cytoplasmic expression and erroneous disulfide bond formation in periplasmic expression. To test this assumption, the activity in the fractions from the expression of scFv with and without pelB leader was remeasured after incubation for three days at 20°C. Whereas the relative activities in the soluble cytoplasmic fraction and in the resuspended pellet remained almost the same for periplasmic expression (45% and 47% after three days compared to 43% and 45% right after sample preparation), they had changed significantly for cytoplasmic expression (61% and 35% compared to 92% and 6%), possibly due to further spontaneous refolding of the cytoplasmic inclusion bodies. This effect was especially observed when using the sample preparation procedure as suggested by Novagen. Here, the lysis buffer contained Triton X-100, that might have favored refolding. The malformed scFv from periplasmic aggregates could not be further renatured in lysis buffer, but with good yields following a procedure similar to that reported by Buchner *et al.* [63], which includes reduction and re-oxidation (A

3.7.1.1). Once the scFv had obtained its native state, it proved to be very stable and still showed good functionality after one year of storage at 4°C.

#### **A 4.2.1 Analysis of Expression by Comparison to *in vitro* Results**

Many factors can influence the *in vivo* expression of recombinant protein. However, some of these effects can only be seen *in vivo*, while others occur *in vivo* and *in vitro*. The former were defined as “physiological effects” (complex posttranslational events such as translocation inside the cell, processing, assisted protein folding, protein degradation and also toxic effects), which only occur in the living cell, but not or to a much lower extent in cell extracts. In contrast, the latter were considered as “intrinsic effects” (transcription and translation efficiency and spontaneous protein folding), which depend on intrinsic properties of the nucleotide sequences of the DNA and mRNA and on the amino acid sequence.

*In vivo* protein folding is so far not completely understood, but certainly depends on intrinsic properties of the protein, which is stipulated by the amino acid sequence but also on “physiological effects”. These include the redox potential of the environment and the presence of folding catalysts which influence the formation of disulfide bridges, and chaperones which may assist the proper folding of certain proteins. In the expression system used here, “intrinsic effects” are reduced to the translation efficiency and the spontaneous protein folding problem because derivatives of the vector plasmid pET20b(+) are high-copy plasmids and the T7-promoter expression system is so efficient that mRNA is synthesized much faster than it can be translated by the typical set of ribosomes in *E. coli*. However, this overproduction of mRNA may lead to secondary structures, which are of minor importance in natural prokaryotic expression systems in which transcription and translation are directly coupled.

In the following discussion, it was attempted to assign the experimentally observed data to either “physiological” or “intrinsic effects”. As stated earlier, data obtained *in vitro* correspond to “intrinsic effects” and data obtained *in vivo* are the result of “intrinsic” and “physiological effects”. The relative expression level and the solubilities of heterologous protein were investigated. The measurements, in particular in the *in vivo* system, were not very precise. Therefore, a classification of the observed data as high (>60%), moderate (60-15%), or low (<15%), with 100% representing the maximum amount of protein accumulation (*in vivo* ~10% of total cell protein, *in vitro* 24 µg ml<sup>-1</sup> protein) and 100% of soluble/functional protein, was preferred. This classification might seem arbitrary. However, values within one class are at least twice as high as the values that were put in the next lower class.

#### A 4.2.1.1 Accumulation of heterologous protein

Table A 4.1 summarizes the *in vitro* and *in vivo* results for the extent of accumulated scFv and scFv-EGFP fusion protein with and without pelB leader sequence, respectively.

**Table A 4.1.** Expression levels of recombinant proteins in *in vitro* and *in vivo* experiments.

	scFv with leader	scFv without leader	scFv-EGFP with leader	scFv-EGFP without leader
<i>in vitro</i>	high	moderate	high	moderate
<i>in vivo</i>	high	moderate	n.d. *	moderate
effect	intrinsic	intrinsic	physiological	intrinsic

\* not determined.

The excellent concordance of the *in vitro* and *in vivo* data for the scFv with and without leader and scFv-EGFP without leader suggest that the different extents of accumulation observed *in vivo* reflect “intrinsic effects”. Due to the efficient promoter and the fact that the amino acid sequence influences the structure, but not the expression level of a protein, here they can be assigned specifically to the nucleotide sequence of the mRNA and translation efficiency. By comparing the amounts of accumulated recombinant protein for pETK(+), and pETKE(+) with pETK(-) and pETK(-)-AC as well as with pETKE(-) and pETKE(-)-AC, it was shown that the two extra nucleotides directly upstream of the start codon had no influence on translation efficiency. This is in contrast to observations made by Ringquist *et al.*, who observed highest expression levels for spacings between the Shine Dalgarno sequence and the start codon corresponding to that in the construct without the two additional nucleotides [41].

A plausible explanation for the low expression levels of the genes without the leader sequence might be the presence of the codon GUC at the +2 position (the codon directly downstream of the start codon). In the constructs containing the pelB leader sequence, the +2 codon is AAA, which is the most common codon at this position in *E. coli* and leads to higher gene expression levels than most other codons [42, 102]. The codon usage in this region seems to be important for the initiation of translation [43]. Furthermore, the secondary structure of the mRNA of the constructs without the pelB leader sequence might have an impact on the recognition of the mRNA by the ribosome and therefore suppress the negative effect of the two additional nucleotides.

Severe problems were encountered with the *in vivo* expression of the scFv-EGFP fusion protein with leader sequence. As it could be expressed at high levels *in vitro* ( $24 \mu\text{g ml}^{-1}$ ), this problem can be attributed clearly to “physiological effects”. Such problems have been discussed elsewhere [20, 76] and also toxic effects of other fusion proteins with EGFP have been described [75].

#### A 4.2.1.2 Solubility and functionality

The classifications of the solubilities are shown in Table A 4.2. For the *in vivo* experiments they refer to the ratio of active to inactive protein after sample preparation, which might not reflect the situation in the living cell because spontaneous refolding might have occurred after cell disruption. However, whether a protein can be refolded spontaneously after isolation from the cell may depend on its processing inside the cell.

**Table A 4.2.** Solubility of recombinant proteins in *in vitro* and *in vivo* experiments.

	scFv with leader	scFv without leader	scFv-EGFP with leader	scFv-EGFP without leader
<i>in vitro</i>	low	low	low	low
<i>in vivo</i>	low	moderate	n.d. *	moderate
effect	intrinsic	physiological	-	physiological

\* not determined.

For the scFv with leader sequence, the *in vivo* data correspond to the *in vitro* data and only approximately 10% of recombinant protein were soluble. Consequently and according to the classification, the low percentage of soluble protein observed in periplasmic expression reflects an intrinsic effect. Too rapid gene expression or volume limitations can be excluded, because *in vitro* other proteins have been expressed at much higher levels and with approximately 100% of soluble protein. In cell-free coupled transcription/translation no efficient oxidation and formation of disulfide bonds is possible, which is necessary for most scFvs to fold correctly. Directly after the experiment the insoluble protein was separated by centrifugation, not allowing any further spontaneous refolding. *In vivo* expression of the scFv into the oxidizing environment of the periplasm did not increase the percentage of soluble protein. Here oxidation might have occurred, but in an erroneous manner. For cytoplasmic

expression, quite high amounts of active scFv (~50%) and scFv-EGFP (~25%) were observed after sample preparation based on comparison of relative activities in soluble and insoluble fractions. Since *in vitro* expression yielded only 10% of soluble protein, it is reasonable to assume that *in vivo* initially also the majority of the recombinant protein was misfolded in the non-oxidizing cytoplasm, which was undermined by microscopic observations. However, in contrast to periplasmic expression, the heterologous protein seemed to be kept in a state, which could be more easily refolded to the native state during sample preparation. This included the use of the detergent Triton X-100, incubation with lysozyme under shaking and sonification. During these steps a part of the protein might have been refolded and oxidized to its native state. The fact that this procedure did not result in a comparable increase of activity for periplasmic aggregates, although the total amount of scFv was higher, we assign to “physiological effects”.

#### **A 4.2.1.3 Conclusions**

It was shown that the quantitative comparison of *in vitro* and *in vivo* expression of an scFv and an scFv-EGFP fusion protein allows the identification of factors that impede the high-level protein production in *E. coli*. Once identified, these factors can be partially eliminated. Here, translation efficiency which is influenced by the mRNA sequence was a limiting factor. The optimization of the nucleotide sequence with regard to the formation of secondary mRNA structures is under further investigation. The formation of insoluble protein could be assigned to intrinsic properties of the protein, although spontaneous refolding into the native state seemed to be favored after expression in the cytoplasm. However, in order to directly obtain all scFv expressed heterologously in *E. coli* in a soluble form, protein engineering or the use of genetically engineered strains as described by Besette *et al.* [54] might be necessary. The failure to express the scFv-EGFP fusion protein fused to the pelB leader *in vivo* was likely due to physiological effects, because it was expressed well *in vitro*. With this in mind, *in vitro* production or selection of other expression systems might prove useful.

Many recombinant proteins are expressed very poorly in *E. coli* or in an inactive form. On the one hand, a combination of *in vitro* and *in vivo* expression can help to identify factors impeding the expression. On the other hand, for some constructs the *in vitro* assay could adequately reproduce the *in vivo* expression levels. Therefore, *in vitro* assays might be practicable tools to screen for good expression vectors before going into the *in vivo* expression system.

### A 4.2.2 Expression of the ScFv with pelB Leader Sequence in Fed-batch Cultivations

In order to produce scFv for characterization in a 2 l cultivation *E. coli* BL21(DE3) pLysS cells carrying the plasmid pETK(+) were used. The T7-promoter was induced at a dry cell mass of about 5 g/l and a final dry cell mass of about 8 g/l was reached. The maximal accumulation of recombinant scFv was about 15% of the total cell protein. In a second fed-batch cultivation (13 l final volume) *E. coli* BW3110 cells carrying the plasmid pJOEK(+)-AT were used. The *rhaBAD* promoter was induced at a dry cell mass of more than 40 g/l, the final dry cell mass was 70 g/l and the maximal percentage of scFv compared to the total cell protein was 3%. In spite of all these differences between both systems, two points with striking similarity were observed: The pattern of active scFv in different cell fractions and the medium monitored by ELISA were corresponding and the percentage of soluble protein in both cases did not exceed 10%, although different cell densities and expression levels were obtained.

In the 2 l cultivation the dry cell mass decreased significantly at the time points with maximal accumulation of recombinant scFv. This might be due to the production of heterologous protein, especially in the periplasm, which might have toxic effects. An alternative explanation would be a toxic effect caused by the accumulation of big amounts of mRNA, as the T7 promoter is very strong.

Unfortunately, the expression level obtained with the *rhaBAD* expression system was only about one fifth of that in the respective T7 expression system. This might be due to the different plasmids used rather than due to the different promoters, because low expression levels with the pJOEK derivatives were also observed in cell-free transcription/translation when using the T7 promoter located upstream of the *rhaBAD* promoter on this vector (Table A 3.3). The sequences from the Shine Dalgarno sequence to the stop codon of the scFv gene were exactly the same in both vectors. However the sequences between the respective promoter regions and the gene-encoding sequences differed as well in number as in sequence. The same is true for the sequence between the stop codon and the respective terminator sequences. These different sequences might have an effect on the formation of secondary structures of the mRNAs and therefore influence translation initiation. This finding still has to be investigated in more detail and such effects are of general interest for the optimization of expression levels of heterologous proteins.

Anyway, in spite of the bad expression level, in the 13 l fed-batch cultivation the production of scFv could be enhanced to 2 g/l compared to 0.5 g/l in the 2 l fed-batch cultivation. This was due to the higher cell density.

### **A 4.3 Refolding and Purification of the ScFv K411B and the ScFv-EGFP Fusion Protein**

Both refolding and purification worked very well for the scFv. After refolding the activity measured in ELISA could be increased by a factor of 13 compared to the initial total cell extract. Assuming that about 90% of the scFv had been misfolded, a 10-fold increase in activity would have been expected upon quantitative refolding. According to this and to SDS-PAGE it can be stated, that the insoluble scFv from periplasmic expression could be refolded nearly quantitatively and also the dilution factor of 25 in rapid dilution was sufficient. Purification by immobilized metal affinity chromatography (IMAC) yielded homogeneous scFv at a concentration of 110 µg/ml.

Refolding of the scFv-EGFP fusion protein was successful to some extent, however no purification was achieved by this procedure, but only an enrichment to about 20%. Additionally, the scFv-EGFP was quite unstable and lost activity within the first few days after refolding. Purification by IMAC under native conditions was not possible at all. Fractions with increased fluorescence did not contain, but in those no full length fusion protein, indicating that it had been degraded during the procedure. In the reports of Casey *et al.*, Griep *et al.* and Feilmeier *et al.* [20, 75, 76] also instability of fusion proteins was reported. The fusion protein might have been degraded resulting in a fluorophore without binding activity. The instability could have been caused by the movement of the two globular domains, which might strain the linking amino acids mechanically. On the other hand it is possible that the measured fluorescence was completely unspecific, as many proteins were seen in fractions with increased fluorescence (data not shown). High fluorescence was also measured in the extracellular and periplasmic fraction of cells expressing the fusion protein, but no binding activity together with fluorescence could be detected (Table A 3.2).



#### **A 4.4 Application of the ScFv K411B and the ScFv-EGFP Fusion Protein in *s*-Triazine Analysis**

Although both the hapten-immobilized ELISA and FLISA proved to be very practical for the characterization and at least the FLISA also for the quantification of functional heterologous protein, problems were encountered when using these assays for the analysis of *s*-triazines. These were mainly due to the high IC<sub>50</sub>-values and detection limits.

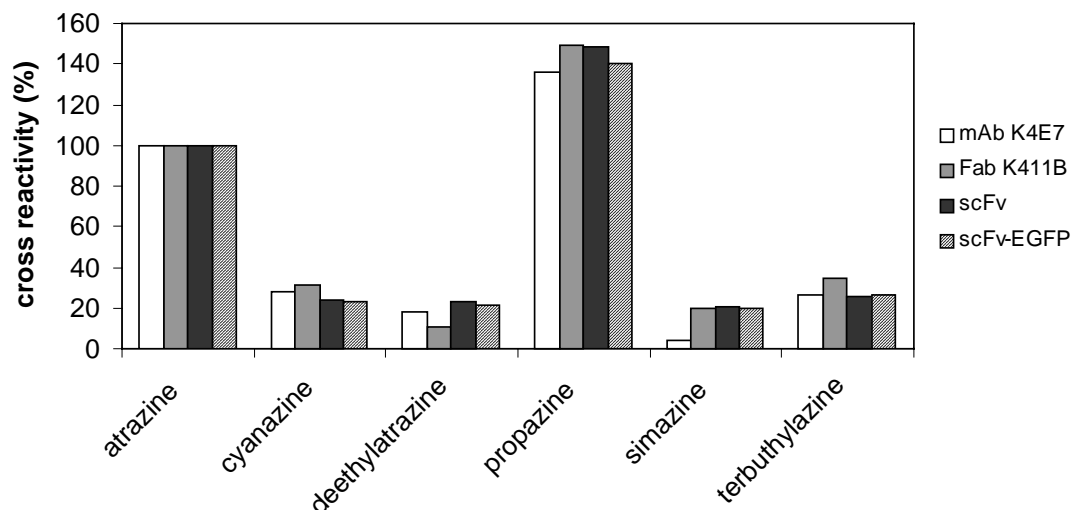
The antibody-fragment-immobilized ELISAs were quite good with respect to the IC<sub>50</sub>-values (2.3 µg/l when using refolded scFv), but due to the low absorbances, the standard deviations were very high and therefore the detection limit could not be defined at 15% inhibition. The reason for the low absorbances was probably the bad recognition of the scFv by the polyclonal rabbit anti-mouse antibody. Assays with the respective Fab fragment, which has more epitopes, showed much higher absorbances. The application of the mouse anti-polyhistidine IgG was intended to improve the recognition of the scFv via the His<sub>6</sub>-tag. However, this antibody did not bind well and therefore had to be applied at relatively high concentrations. This resulted in the movement of the IC<sub>50</sub>-value to even higher concentrations. The absorbances were slightly higher, and also the error bars remained very big. A possibility to improve the sensitivity of this ELISA format could be the application of a more specific capture antibody, either a more specific antibody against the scFv or a better binding antibody against another tag such as the anti-Etag antibody in combination with an Etagged scFv [23]. However, the Etag-specific antibodies are expensive. Also the use of a biotinylated scFv and immobilized streptavidin for capturing the scFv might be practicable.

The hapten-immobilized ELISAs on the other hand showed high signals and smaller error bars but very high IC<sub>50</sub>-values, the lowest being 120 µg/ml atrazine in the hapten-immobilized ELISA with cell extract and mouse anti-polyhistidine IgG. Kramer *et al.* obtained a better value (25 µg/ml [23]) possibly due to more efficient labeling of the scFv using an Etag. Omitting this additional antibody did not decrease the IC<sub>50</sub>-values either, but the absorbances. Again, the bad results are very likely due to the poor recognition of the scFv by the goat anti-mouse antibody-HRP conjugate or the mouse anti-polyhistidine IgG, respectively. Also in this case an alternative would also be the application of other labeling techniques such as more specific antibodies or the biotin-streptavidine system. Giersch *et al.* lowered the IC<sub>50</sub>-value of a terbuthylazine-specific hapten-immobilized ELISA about 50% by using a biotinylated monoclonal antibody and streptavidin-HRP conjugate instead of the terbuthylazine-specific antibody and anti-mouse-IgG-HRP conjugate for labeling [69].

Anyway, if this would be possible here, the detection limit would still be too high for an application in *s*-triazine analysis. Engineering of the scFv and the use of other haptens represent another way to lower the detection limit. Kusharyoto obtained an eight times higher sensitivity of a Fab-fragment-immobilized atrazine-specific ELISA by using an engineered Fab fragment with increased affinity towards atrazine in combination with a tracer with lower affinity [103]. A similar approach might also work for a hapten-immobilized ELISA. Of course it could also be applied in the improvement of the scFv-immobilized ELISA once the labeling problem will be solved.

In the FLISA, the IC<sub>50</sub>-value and the detection limit for atrazine were in the same order of magnitude as those previously shown for hapten-immobilized ELISA when using an Etagged scFv and an anti-Etag antibody-HRP conjugate for labeling [23]. This in turn supports the conclusion that the problem with the hapten-immobilized ELISA in this work was really due to inefficient labeling. Anyway, the detection limit of 2 µg atrazine per liter obtained by this method is still not applicable for the investigation of ground and drinking water. Further improvement could be achieved by protein engineering and the use of other haptens as mentioned above. In this work, C<sub>6</sub>-simazine BSA conjugate was used as an alternative hapten. Due to the lower affinity of simazine to the scFv K411B, it was expected to decrease the IC<sub>50</sub>-values and consequently the detection limits, but no such effect could be observed. Although the scFv has a lower affinity to simazine than to atrazine (see Figure A 24), this could not be verified for C<sub>6</sub>-simazine and C<sub>6</sub>-atrazine, respectively. This phenomenon has been observed before [103] and is probably due to the high affinity of the scFv to the spacer rather than to the *s*-triazine molecule. The labeling of the scFv with EGFP proved to be very efficient and did not change the binding characteristics of the scFv. In addition to a reduction of the assay time, the FLISA resulted in the best concordance with the hapten-immobilized ELISA using the scFv expressed with phages [23], regarding the IC<sub>50</sub>-value and the detection limit for atrazine as well as the cross reactivities towards the other investigated *s*-triazines.

In Figure A 25 the cross reactivities obtained with the scFv-EGFP fusion protein are compared to the reported values of the parental monoclonal antibody K4E7 [22], the respective Fab fragment K411B [10] and the scFv K411B expressed with phages [23]. The monoclonal antibody and the Fab fragment were characterized with antibody-(fragment)-immobilized ELISA, the scFv with a hapten-immobilized ELISA and an anti-Etag IgG for labeling.



**Figure A 25.** Cross reactivities of the parental monoclonal antibody (m Ab) K4E7, the Fab fragment K411B, the scFv K411B and the respective scFv-EGFP towards six *s*-triazines.

The results obtained with the FLISA are in good concordance with the literature data and indicate that the binding specificity was not altered by the fusion of EGFP to the scFv.

## A 4.5 Conclusions

The expression of a scFv and a scFv-EGFP fusion protein was studied by comparison of *in vivo* and *in vitro* experiments. Constructs with favorable expression levels were used to produce both proteins in fed-batch cultivations. The scFv was either isolated from soluble fractions by immobilized metal affinity chromatography (IMAC) or from the insoluble fraction by an efficient refolding method. Both procedures yielded protein in a purity of 85% and higher. The scFv-EGFP fusion protein could neither be purified nor refolded efficiently. However both proteins could be applied successfully in hapten-immobilized ELISA and FLISA, a novel fluorescence immunoassay, also without purification.

## A 5 References

1. Stryer, L. (1995) *Biochemie*, 4 edn, Spektrum Akademischer Verlag, Heidelberg.
2. Klinke, R. & Silbernagl, S. (1996) *Lehrbuch der Physiologie*, 2 edn, Thieme, Stuttgart.
3. Janeway, C. A. & Travers, P. (1997) *Immunologie*, 2 edn, Spektrum Akademischer Verlag, Heidelberg.
4. Kohler, G. & Milstein, C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity, *Nature*. 256, 495-7.
5. Boulianne, G. L., Hozumi, N. & Shulman, M. J. (1984) Production of functional chimaeric mouse/human antibody, *Nature*. 312, 643-6.
6. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. (1986) Replacing the complementarity-determining regions in a human antibody with those from a mouse, *Nature*. 321, 522-5.
7. Huston, J. S., McCartney, J., Tai, M. S., Mottola-Hartshorn, C., Jin, D., Warren, F., Keck, P. & Oppermann, H. (1993) Medical applications of single-chain antibodies, *Int Rev Immunol*. 10, 195-217.
8. Huston, J. S., George, A. J., Adams, G. P., Stafford, W. F., Jamar, F., Tai, M. S., McCartney, J. E., Oppermann, H., Heelan, B. T., Peters, A. M., Houston, L. L., Bookman, M. A., Wolf, E. J. & Weiner, L. M. (1996) Single-chain Fv radioimmunotargeting, *Q J Nucl Med*. 40, 320-33.
9. Le Gall, F., Bove, J. M. & Garnier, M. (1998) Engineering of a single-chain variable-fragment (scFv) antibody specific for the stolbur phytoplasma (Mollicute) and its expression in escherichia coli and tobacco plants, *Appl Environ Microbiol*. 64, 4566-72.
10. Lange, S., Schmitt, J. & Schmid, R. D. (2001) High-yield expression of the recombinant, atrazine-specific Fab fragment K411B by the methylotrophic yeast *Pichia pastoris*, *J Immunol Methods*. 255, 103-14.
11. Takahashi, K., Yuuki, T., Takai, T., Ra, C., Okumura, K., Yokota, T. & Okumura, Y. (2000) Production of humanized Fab fragment against human high affinity IgE receptor in *Pichia pastoris*, *Biosci Biotechnol Biochem*. 64, 2138-44.
12. Edqvist, J., Keranen, S., Penttila, M., Straby, K. B. & Knowles, J. K. (1991) Production of functional IgM Fab fragments by *Saccharomyces cerevisiae*, *J Biotechnol*. 20, 291-300.

13. Horwitz, A. H., Chang, C. P., Better, M., Hellstrom, K. E. & Robinson, R. R. (1988) Secretion of functional antibody and Fab fragment from yeast cells, *Proc Natl Acad Sci U S A.* 85, 8678-82.
14. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains, *Nature.* 348, 552-4.
15. Nicholls, P. J., Johnson, V. G., Andrew, S. M., Hoogenboom, H. R., Raus, J. C. & Youle, R. J. (1993) Characterization of single-chain antibody (sFv)-toxin fusion proteins produced in vitro in rabbit reticulocyte lysate, *J Biol Chem.* 268, 5302-8.
16. Merk, H., Stiege, W., Tsumoto, K., Kumagai, I. & Erdmann, V. A. (1999) Cell-free expression of two single-chain monoclonal antibodies against lysozyme: effect of domain arrangement on the expression, *J Biochem (Tokyo).* 125, 328-33.
17. Verma, R., Boleti, E. & George, A. J. (1998) Antibody engineering: comparison of bacterial, yeast, insect and mammalian expression systems, *J Immunol Methods.* 216, 165-81.
18. Shibui, T., Munakata, K., Matsumoto, R., Ohta, K., Matsushima, R., Morimoto, Y. & Nagahari, K. (1993) High-level production and secretion of a mouse-human chimeric Fab fragment with specificity to human carcino embryonic antigen in *Escherichia coli*, *Appl Microbiol Biotechnol.* 38, 770-5.
19. Pantoliano, M. W., Bird, R. E., Johnson, S., Asel, E. D., Dodd, S. W., Wood, J. F. & Hardman, K. D. (1991) Conformational stability, folding, and ligand-binding affinity of single-chain Fv immunoglobulin fragments expressed in *Escherichia coli*, *Biochemistry.* 30, 10117-25.
20. Casey, J. L., Coley, A. M., Tilley, L. M. & Foley, M. (2000) Green fluorescent antibodies: novel in vitro tools, *Protein Eng.* 13, 445-52.
21. Holliger, P., Prospero, T. & Winter, G. (1993) "Diabodies": small bivalent and bispecific antibody fragments, *Proc Natl Acad Sci U S A.* 90, 6444-8.
22. Giersch, T. (1993a) A New Monoclonal Antibody for the Sensitive Detection of Atrazine with Immunoassay in Microtiter Plate and Dipstick Format, *J. Agric. Food Chem.* 41, 1006-1011.
23. Kramer, K. & Hock, B. (1996) Recombinant Single-chain Antibodies against s-Triazines, *Food & Agricultural Immunology.* 8, 97-109.

24. Tauler, R., de Azevedo, D. A., Lacorte, S., Cespedes, R., Viana, P. & Barcelo, D. (2001) Organic pollutants in surface waters from Portugal using chemometric interpretation, *Environ Technol.* 22, 1043-54.
25. Quintana, J., Marti, I. & Ventura, F. (2001) Monitoring of pesticides in drinking and related waters in NE Spain with a multiresidue SPE-GC-MS method including an estimation of the uncertainty of the analytical results, *J Chromatogr A.* 938, 3-13.
26. Van Maanan, J. M., De Vaan, M. A., Veldstra, A. W. & Hendrix, W. P. (2001) Pesticides and nitrate in groundwater and rainwater in the Province of Limburg in The Netherlands, *Environ. Monit. Assess.* 72, 95-114.
27. Schlegel, H. G. (1992) *Allgemeine Mikrobiologie*, 7 edn, Thieme, Stuttgart.
28. Dellweg, H., Schmid, R. D. & Trommer, W. E. (1992) *Roempp Lexikon Biotechnologie*, Thieme, Stuttgart.
29. Wulfing, C. & Pluckthun, A. (1994) Protein folding in the periplasm of *Escherichia coli*, *Mol Microbiol.* 12, 685-92.
30. Strandberg, L. & Enfors, S. O. (1991) Factors influencing inclusion body formation in the production of a fused protein in *Escherichia coli*, *Appl Environ Microbiol.* 57, 1669-74.
31. Martineau, P., Jones, P. & Winter, G. (1998) Expression of an antibody fragment at high levels in the bacterial cytoplasm, *J Mol Biol.* 280, 117-27.
32. Pratt, J. M. (1984) *Coupled Transcription-Translation in Prokaryotic Cell-free Systems*, IRL Press, Oxford, England.
33. Schindler, P. T., Macherhammer, F., Arnold, S., Reuss, M. & Siemann, M. (1999) Investigation of translation dynamics under cell-free protein biosynthesis conditions using high-resolution two-dimensional gel electrophoresis, *Electrophoresis.* 20, 806-12.
34. Schindler, P. T., Baumann, S., Reuss, M. & Siemann, M. (2000) In vitro coupled transcription translation: effects of modification in lysate preparation on protein composition and biosynthesis activity, *Electrophoresis.* 21, 2606-9.
35. Arnold, S., Siemann, M., Scharnweber, K., Werner, M., Baumann, S., Reuss, M. (2001) Kinetic modeling and simulation of in vitro transcription by phage T7n RNA polymerase, *Biotechnol. Bioeng.* 72, 548-561.
36. Katanaev, V. L., Spirin, A. S., Reuss, M. & Siemann, M. (1996) Formation of bacteriophage MS2 infectious units in a cell-free translation system, *FEBS Lett.* 397, 143-8.

37. Bremer, H. & Dennis, P. P. (1996) Modulation of Chemical Composition and Other Parameters of the Cell By Growth Rate. in *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Microbiology* pp. 1553-1569, American Society for Microbiology, Washington DC.
38. Ryabova, L. A., Desplancq, D., Spirin, A. S. & Pluckthun, A. (1997) Functional antibody production using cell-free translation: effects of protein disulfide isomerase and chaperones, *Nat Biotechnol.* 15, 79-84.
39. Novagen. (2000) pET System Manual (TB055)
40. Shine, J. & Dalgarno, L. (1975) Determinant of cistron specificity in bacterial ribosomes, *Nature.* 254, 34-8.
41. Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D. & Gold, L. (1992) Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site, *Mol Microbiol.* 6, 1219-29.
42. Stenstrom, C. M., Holmgren, E. & Isaksson, L. A. (2001a) Cooperative effects by the initiation codon and its flanking regions on translation initiation, *Gene.* 273, 259-65.
43. Stenstrom, C. M., Jin, H., Major, L. L., Tate, W. P. & Isaksson, L. A. (2001b) Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli*, *Gene.* 263, 273-84.
44. Wickner, W., Driessen, A. J. & Hartl, F. U. (1991) The enzymology of protein translocation across the *Escherichia coli* plasma membrane, *Annu Rev Biochem.* 60, 101-24.
45. Pugsley, A. P. (1993) The complete general secretory pathway in gram-negative bacteria, *Microbiol Rev.* 57, 50-108.
46. Sanchez, L., Ayala, M., Freyre, F., Pedroso, I., Bell, H., Falcon, V. & Gavilondo, J. V. (1999) High cytoplasmic expression in *E. coli*, purification, and in vitro refolding of a single chain Fv antibody fragment against the hepatitis B surface antigen, *J Biotechnol.* 72, 13-20.
47. Simmons, L. C. & Yansura, D. G. (1996) Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*, *Nat Biotechnol.* 14, 629-34.
48. Noelting, B. (1999) *Protein Folding Kinetics*, Springer, Berlin.
49. Baneyx, F. (1999) In Vivo Folding of Recombinant Proteins in *Escherichia coli* in *Manual of Industrial Microbiology and Biotechnology* (Demain, A. L. & Davies, J. E., eds) pp. 830, American Society for Microbiology Press, Washington D.C.

50. Gething, M. J. & Sambrook, J. (1992) Protein folding in the cell, *Nature*. 355, 33-45.
51. Hartl, F. U. (1996) Molecular chaperones in cellular protein folding, *Nature*. 381, 571-9.
52. Georgiou, G. & Valax, P. (1996) Expression of correctly folded proteins in *Escherichia coli*, *Curr Opin Biotechnol*. 7, 190-7.
53. Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. & Yura, T. (1998) Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*, *Appl Environ Microbiol*. 64, 1694-9.
54. Bessette, P. H., Aslund, F., Beckwith, J. & Georgiou, G. (1999) Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm, *Proc Natl Acad Sci U S A*. 96, 13703-8.
55. Tavladoraki, P., Girotti, A., Donini, M., Arias, F. J., Mancini, C., Morea, V., Chiaraluce, R., Consalvi, V. & Benvenuto, E. (1999) A single-chain antibody fragment is functionally expressed in the cytoplasm of both *Escherichia coli* and transgenic plants, *Eur J Biochem*. 262, 617-24.
56. Wulfing, C. & Pluckthun, A. (1994) Correctly folded T-cell receptor fragments in the periplasm of *Escherichia coli*. Influence of folding catalysts, *J Mol Biol*. 242, 655-69.
57. Ostermeier, M., De Sutter, K. & Georgiou, G. (1996) Eukaryotic protein disulfide isomerase complements *Escherichia coli* dsbA mutants and increases the yield of a heterologous secreted protein with disulfide bonds, *J Biol Chem*. 271, 10616-22.
58. Humphreys, D. P., Weir, N., Lawson, A., Mountain, A. & Lund, P. A. (1996) Co-expression of human protein disulphide isomerase (PDI) can increase the yield of an antibody Fab' fragment expressed in *Escherichia coli*, *FEBS Lett*. 380, 194-7.
59. Lilie, H., McLaughlin, S., Freedman, R. & Buchner, J. (1994) Influence of protein disulfide isomerase (PDI) on antibody folding in vitro, *J Biol Chem*. 269, 14290-6.
60. Hayhurst, A. & Harris, W. J. (1999) *Escherichia coli* skp chaperone coexpression improves solubility and phage display of single-chain antibody fragments, *Protein Expr Purif*. 15, 336-43.
61. Korz, D. J., Rinas, U., Hellmuth, K., Sanders, E. A. & Deckwer, W. D. (1995) Simple fed-batch technique for high cell density cultivation of *Escherichia coli*, *J Biotechnol*. 39, 59-65.



62. Wilms, B., Hauck, A., Reuss, M., Syldatk, C., Mattes, R., Siemann, M. & Altenbuchner, J. (2001) High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the *Escherichia coli* rhaBAD promoter, *Biotechnol Bioeng.* 73, 95-103.
63. Buchner, J., Pastan, I. & Brinkmann, U. (1992) A method for increasing the yield of properly folded recombinant fusion proteins: single-chain immunotoxins from renaturation of bacterial inclusion bodies, *Anal Biochem.* 205, 263-70.
64. Lilie, H., Schwarz, E. & Rudolph, R. (1998) Advances in refolding of proteins produced in *E. coli*, *Curr Opin Biotechnol.* 9, 497-501.
65. Li, T. & Rosazza, J. P. (2000) Biocatalytic synthesis of vanillin, *Appl Environ Microbiol.* 66, 684-7.
66. Malfroy, B. (2000). Chemical Modifications of Monoclonal Antibodies for Intracellular Immunotherapy. Paper presented at the *GAb 2000, An International Symposium on Downstream Processing of Genetically Engineered Antibodies and Related Molecules*, Barcelona, Spain.
67. Hefle, S. L., Jeannot, E. & Taylor, S. L. (2001) Development of a sandwich enzyme-linked immunosorbent assay for the detection of egg residues in processed foodst, *J Food Prot.* 64, 1812-6.
68. Reimer, G. J., Gee, S. J. & Hammcock, B. C. (1998) Comparison of a Time-Resolved Fluorescence Immunoassay and an Enzyme-Linked Immunosorbent Assay for the Analysis of Atrazine in Water, *J. Agric. Food Chem.* 46, 3353-3358.
69. Giersch, T., Kramer, K. & Hock, B. (1993b) Optimization of a monoclonal antibody-based enzyme immunoassay for the detection of terbuthylazine, *The Science of the Total Environment.* 132, 435-448.
70. Strachan, G., Grant, S. D., Learmonth, D., Longstaff, M., Porter, A. J. & Harris, W. J. (1998) Binding characteristics of anti-atrazine monoclonal antibodies and their fragments synthesised in bacteria and plants, *Biosens Bioelectron.* 13, 665-73.
71. Waldo, G. S., Standish, B. M., Berendzen, J. & Terwilliger, T. C. (1999) Rapid protein-folding assay using green fluorescent protein [see comments], *Nat Biotechnol.* 17, 691-5.
72. Albano, C. R., Randers-Eichhorn, L., Bentley, W. E. & Rao, G. (1998) Green fluorescent protein as a real time quantitative reporter of heterologous protein production, *Biotechnol Prog.* 14, 351-4.

73. Cha, H. J., Wu, C. F., Valdes, J. J., Rao, G. & Bentley, W. E. (2000) Observations of green fluorescent protein as a fusion partner in genetically engineered *Escherichia coli*: monitoring protein expression and solubility, *Biotechnol Bioeng.* 67, 565-74.
74. Wu, C. F., Cha, H. J., Rao, G., Valdes, J. J. & Bentley, W. E. (2000) A green fluorescent protein fusion strategy for monitoring the expression, cellular location, and separation of biologically active organophosphorus hydrolase, *Appl Microbiol Biotechnol.* 54, 78-83.
75. Feilmeier, B. J., Iseminger, G., Schroeder, D., Webber, H. & Phillips, G. J. (2000) Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*, *J Bacteriol.* 182, 4068-76.
76. Griep, R. A., van Twisk, C., van der Wolf, J. M. & Schots, A. (1999) Fluobodies: green fluorescent single-chain Fv fusion proteins, *J Immunol Methods.* 230, 121-30.
77. Schwalbach, G., Sibler, A. P., Choulier, L., Deryckere, F. & Weiss, E. (2000) Production of fluorescent single-chain antibody fragments in *Escherichia coli*, *Protein Expr Purif.* 18, 121-32.
78. Hink, M. A., Griep, R. A., Borst, J. W., van Hoek, A., Eppink, M. H., Schots, A. & Visser, A. J. (2000) Structural dynamics of green fluorescent protein alone and fused with a single chain Fv protein, *J Biol Chem.* 275, 17556-60.
79. Arai, R., Ueda, H., Tsumoto, K., Mahoney, W. C., Kumagai, I. & Nagamune, T. (2000) Fluorolabeling of antibody variable domains with green fluorescent protein variants: application to an energy transfer-based homogeneous immunoassay, *Protein Eng.* 13, 369-76.
80. Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids, *J Mol Biol.* 166, 557-80.
81. Hill, C. W. & Harnish, B. W. (1981) Inversions between ribosomal RNA genes of *Escherichia coli*, *Proc Natl Acad Sci U S A.* 78, 7069-72.
82. Srividya, K. (1999) *Cloning and functional expression of atrazine specific recombinant single chains in Excherichia coli*, Indian Institute of Technology, Madras.
83. Oelschlaeger, P., Srikant-Iyer, S., Lange, S., Schmitt, J. & Schmid, R. D. Fluorophor-Linked Immunosorbent Assay (FLISA): a Time- and Cost-Saving Method for the Characterization of Antibody Fragments Using a Fusion Protein of a Single-Chain Antibody Fragment (scFv) and Enhanced Green Fluorescent Protein (EGFP), *Anal Biochem, in print.*

84. Stumpp, T., Wilms, B. & Altenbuchner, J. (2000) Ein neues, L-Rhamnose-induzierbares Expressionssystem fuer *Escherichia coli*, *BIOspektrum*. 6, 33-36.
85. Birnboim, H. C. & Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA, *Nucleic Acids Res.* 7, 1513-23.
86. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*, 2 edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
87. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors, *Proc Natl Acad Sci U S A.* 74, 5463-7.
88. Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A. & Baumeister, K. (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides, *Science.* 238, 336-41.
89. Freeman, M., Baehler, C. & Spotts, S. (1990) Automated laser-fluorescence sequencing, *Biotechnology (N Y).* 8, 147-8.
90. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction, *Cold Spring Harb Symp Quant Biol.* 51, 263-73.
91. Luria, S., Adams, J. & Ting, R. (1960) Transduction of lactose utilizing ability among strains of *E. coli* and properties of phage particles, *Virology.* 12, 348-390.
92. Tabor, S. & Richardson, C. C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes, *Proc Natl Acad Sci U S A.* 82, 1074-8.
93. Studier, F. W. & Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, *J Mol Biol.* 189, 113-30.
94. Goodrow, M. H., Harrison, R. O. & Hammock, B. D. (1990) Hapten Synthesis, Antibody Development, and Competitive Inhibition Enzyme Immunoassay for *s*-Triazine Herbicides, *J. Agric. Food Chem.* 38, 990-996.
95. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal Biochem.* 72, 248-54.
96. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature.* 227, 680-5.

97. Edman, P. (1950) Method for the determination of the amino acid sequence in peptides, *Acta Chem Scand.* 4, 289-298.
98. Matsudaira, P. (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes, *J Biol Chem.* 262, 10035-8.
99. Wittmann, C. & Hock, B. (1989) Improved enzyme immunoassay for the analysis of s-triazines in water samples, *Food and Agricultural Immunology.* 1, 211.
100. Bäumner, A. J. (1997) *Entwicklung eines Immunsensors zum Nachweis von Pestiziden: ein Einmaltest mit Liposomen und amperometrischer Detektion*, University of Stuttgart, Stuttgart.
101. Kipriyanov, S. M., Dubel, S., Breitling, F., Kontermann, R. E., Heymann, S. & Little, M. (1995) Bacterial expression and refolding of single-chain Fv fragments with C-terminal cysteines, *Cell Biophys.* 26, 187-204.
102. Gren, E. J. (1984) Recognition of messenger RNA during translational initiation in *Escherichia coli*, *Biochimie.* 66, 1-29.
103. Kusharyoto, W. (2001) *Strukturelle Untersuchungen an Atrazin- und 2,4-Dichlorphenoxyessigsäure-spezifischen Antikörperfragmenten mittels molecular modelling und ortsspezifischer Mutagenese*, University of Stuttgart, Stuttgart.

## B 1 Introduction

### B 1.1 Antibiotics

In 1928 Alexander Fleming discovered a fungal infection of *Penicillium notatum* on a culture of *Staphylococcus aureus* that inhibited the growth of the Gram-positive bacterium. Ten years later penicillin G (benzylpenicillin), one of the responsible antibiotics (greek: *anti bios* = against life) produced by the fungus could be isolated and characterized. In 1941 a patient was treated successfully with the antibiotic and the production in an industrial scale was started in 1945. In 1947 Streptomycin produced by *Streptomyces griseus* was discovered. In contrast to benzylpenicillin, this new antibiotic was also active against Gram-negative bacteria. In the following years numerous other antibiotics were discovered with Actinomycetes being the most important source [1, 2].

Antibiotics can either just inhibit the growth of other microorganisms (bacteriostatics, fungistatics) or actually kill them (bactericides, fungicides). They are biologically active already at very low concentrations and with high specificity. Being secondary metabolites, their natural function is not yet elucidated. Although the genetic apparatus required for the synthesis of antibiotics poses a tremendous load on the organisms, it has been conserved in the course of evolution. Therefore, it seems likely that they confer some advantages to the producer, e.g. the inhibition of competitors for the same substrates. However, such antagonistic relationships can hardly be detected in the natural environment [3].

Antibiotics can be classified according to their chemical structure. In Table B 1.1 some of these classes with examples are shown.

**Table B 1.1.** Classification of antibiotics according to their chemical structure [1].

<b>Class</b>	<b>Chemical Structure</b>	<b>Examples (Application)</b>
carbohydrate antibiotics	amino glycosides	streptomycin (medicine) kasuagamycin (rice-fungicide)
macrocyclic lactones	macrolide antibiotics polyene antibiotics ansamycine	erythromycin (medicine) pimaricin (cheese production) rifamycin (tuberculosis)
chinones and related antibiotics	tetracyclins anthracyclins	tetracyclin, chlortetracyclin (medicine, food conservation) doxorubicin (tumor therapy)
amino acid and peptide antibiotics	amino acid derivatives $\beta$ -lactam antibiotics peptide antibiotics chromopeptides glycopeptides	cyclosporin (transplantation medicine) phosphinothricin (pest management) penicillins, cephalosporins (medicine) bacitracin (medicine) virginiamycin (animal feeding) actinomycin (tumor therapy) bleomycin (tumor therapy), vancomycin (medicine), avoparcin (cattle husbandry)
N-containing heterocycles	nucleoside antibiotics	polyoxine, blasticidin S (fungicides in pest management)
O-containing heterocycles	polyether antibiotics	monensin (poultry feeding)
alicyclic antibiotics	cycloalcanes derivatives	cycloheximid (leaf fungicide)
aromatic antibiotics	benzol derivatives	chloramphenicol (medicine) griseofulvin (fungicide)

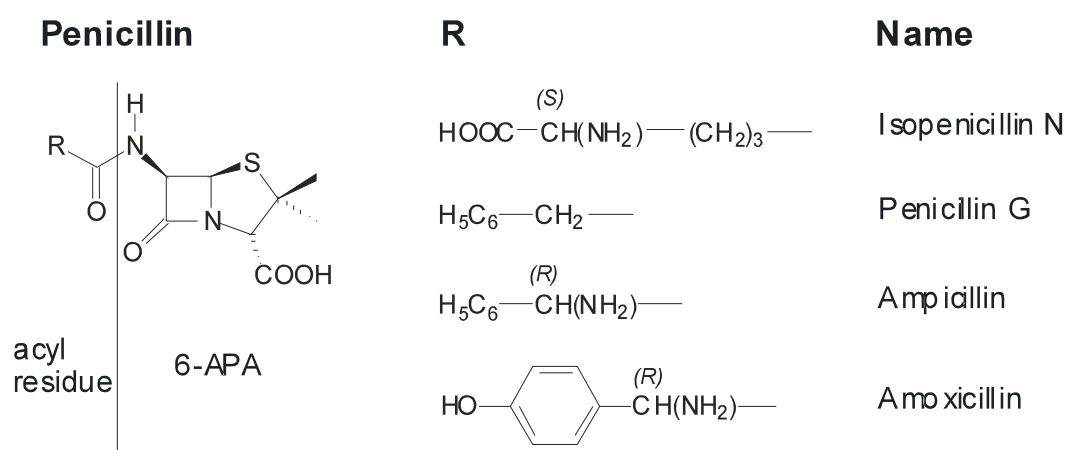
These compounds can exert their antibiotic activity by interacting with

1. the biosynthesis and function of the genetic material,
2. the biosynthesis of cell constituents,
3. the biosynthesis and function of proteins,
4. the biosynthesis and function of cell membranes
5. and the biosynthesis of the cell wall.

### B 1.1.1 $\beta$ -Lactam Antibiotics

$\beta$ -Lactam antibiotics have a cytostatic effect on bacteria by binding irreversibly to a transpeptidase involved in cell wall peptidoglycan polymerization [4, 5]. They are the most important antibiotics due to their high effectiveness, low toxicity and the possibility to chemically and enzymatically modify them. Amongst the  $\beta$ -lactams, cephalosporins and penicillins are the most commonly used antimicrobial agents with a world-wide production of 30,000 tons corresponding to about 1 billion US dollars [1].

Penicillin derivatives (Penems) consist of 6 $\beta$ -amino-penicillanic acid (6-APA), comprising a  $\beta$ -lactam ring and a thiazolidin ring, and a variable acyl residue (Figure B 1).

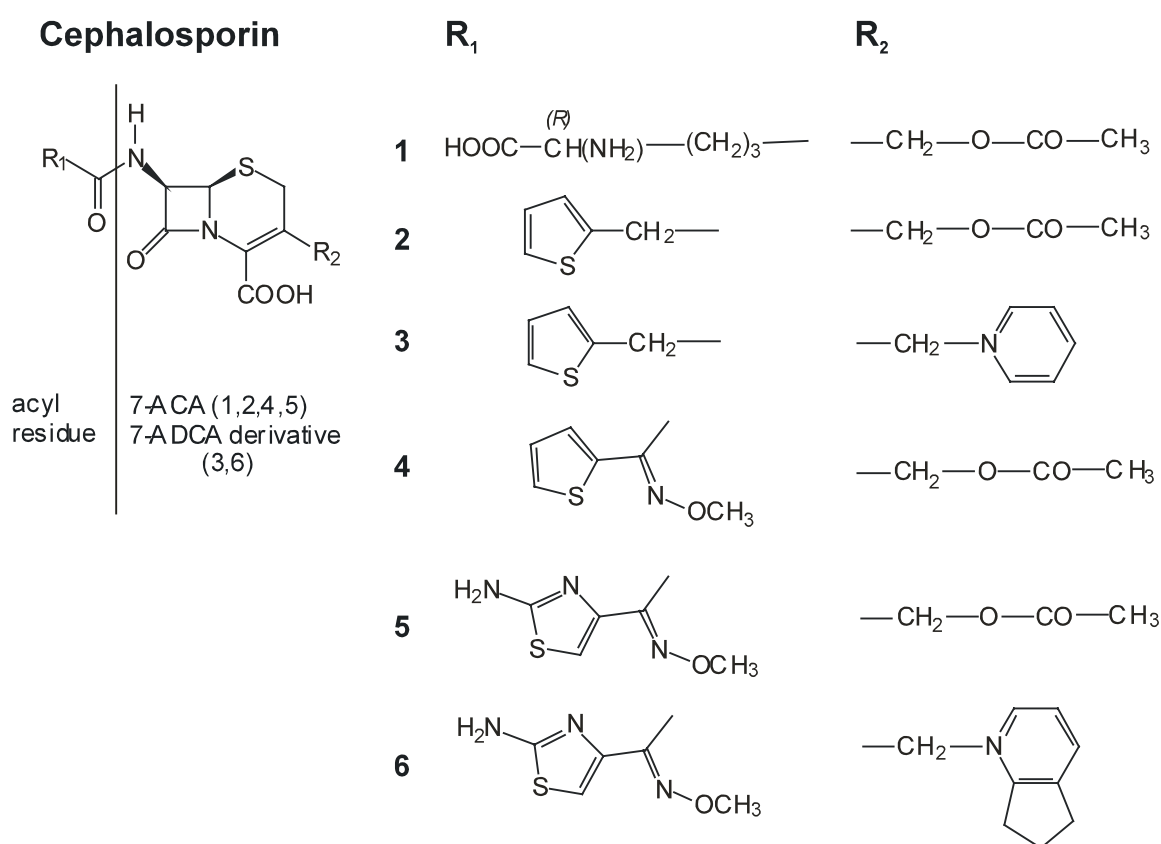


**Figure B 1.** On the left the chemical structure of penicillin, consisting of 6 $\beta$ -aminopenicillanic acid (6-APA) and an acyl residue is shown, on the right some derivatives.

They are used in human and veterinary medicine, but also in molecular biology. *Penicillium chrysogenum* produces isopenicillin N as a primary product, which is converted into penicillin G (benzylpenicillin) when feeding phenylacetic acid. Penicillin G is unstable under acidic conditions and only acts on Gram-positive bacteria. The penems ampicillin and carbenicillin

are more stable and also active against some Gram-negative bacteria like *Escherichia coli*, *Salmonella* and *Shigella*. They are useful tools for the selection of clones transformed with plasmids containing a  $\beta$ -lactamase gene. Amoxicillin has the same *in vitro* activity as ampicillin, but is more effective due to better resorbance and excretion. A variety of other penems have been synthesized starting from 6-APA.

Cephalosporin derivatives (Cephems) are built up by 7 $\beta$ -amino-cephalosporanic acid (7-ACA) or 7 $\beta$ -amino-desacetoxy-cephalosporanic acid (7-ADCA) derivatives ( $\beta$ -lactam ring and a dihydrothiazin ring) and an acyl residue (Figure B 2).



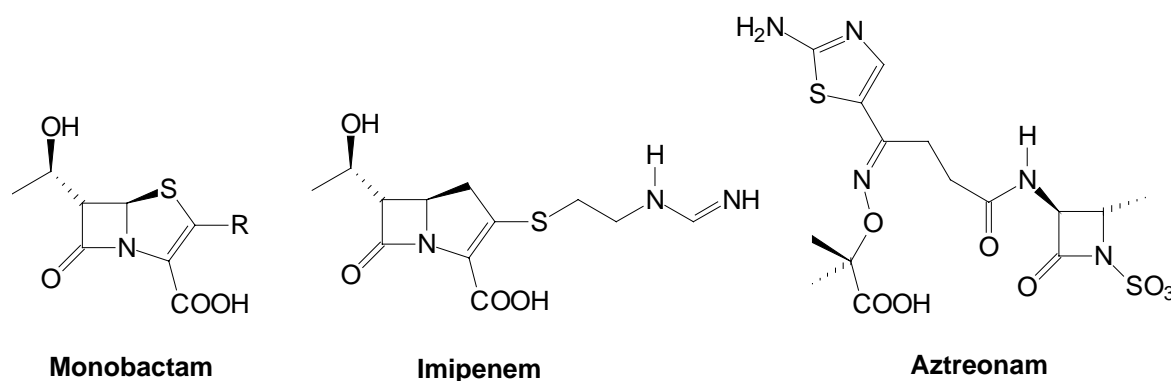
**Figure B 2.** The chemical structure of some cephalosporins. 1: the naturally occurring cephalosporin C, 2 and 3: cephalothin and cefaloridine, two 1<sup>st</sup> generation cephems, 4: cefuroxime, a 2<sup>nd</sup> generation cephem, 5: cefotaxime: a 3<sup>rd</sup> generation cephem, 6: Cefpirome, a 4<sup>th</sup> generation cephem.

The first cephalosporin to be isolated in 1953 from *Agremonium chrysogenum* was cephalosporin C. In the biosynthesis, the five-ring of 6-APA can be converted into the six-ring of 7-ADCA via a ring-extending reaction. This reaction can also be performed chemically. Alternatively, 7-ACA is produced from cephalosporin C isolated from *A.*



*chrysogenum* cultures. Starting from 7-ACA or 7-ADCA, a number of semi-synthetic cepems, varying in mainly two positions, have been synthesized. According to their specificity towards bacteria they have been classified into four generations [6]. Whereas first generation cepems are only active against bacteria, that do not produce cephalosporinases, those of the second generation are not inactivated by some  $\beta$ -lactamases, e.g. that of *Bacteroides*, and are effective against *Haemophilus* species and Enterobacteriaceae. However, many have to be applied parenterally. Third generation cephalosporins can be applied orally and have broad-spectrum activity against many Enterobacteriaceae and especially *Pseudomonas* and *Actinobacter*. Even higher effectiveness against Enterobacteriaceae and *Pseudomonas aeruginosa* is achieved with fourth generation cepems. Today cephalosporins contribute 30% to the total antibiotics market.

Besides penicillins and cephalosporins a number of other  $\beta$ -lactam compounds with antibiotic activity have been isolated and synthesized. Monobactams, carbapenems and monocycles are shown in Figure B 3.



**Figure B 3.** The chemical structure of monobactams, imipenem (a carbapenem) and aztreonam (a monocyclic  $\beta$ -lactam).

## B 1.2 Antibiotic Resistance

Most of the inhibitory effects conferred by antibiotics are due to their binding to proteins or rRNA molecules involved in vital functions, e.g. transcription (rifamycin), translation (streptomycin, erythromycin, tetracyclin, chloramphenicol) or cell wall biosynthesis ( $\beta$ -lactam antibiotics). There are some possibilities how microorganisms can protect themselves against antibiotics:

1. Alteration of the permeability of the cell membrane.
2. Removal of the antibiotic out of the cell with specific or unspecific transport systems.
3. Modification of the target of the antibiotic. If the target is a protein, resistance may be conferred by mutation of the protein to an insensitive form or by expression of insensitive proteins that can substitute for the sensitive protein.
4. Chemical inactivation of the antibiotic, e.g. acetylation of chloramphenicol [7], acetylation, adenylation and phosphorylation of aminoglycoside antibiotics [8] and hydrolysis of  $\beta$ -lactams.

Due to the short generation time of microorganisms, they are capable of responding rapidly to selective pressure and become resistant to new antibiotics. They can achieve this by evolutive modification of the target or antibiotic-modifying enzymes. Another possibility is the gene transfer via plasmids or transposons providing the organism with genes encoding for new mechanisms of resistance.

### B 1.2.1 Resistance towards $\beta$ -Lactams

Decreased permeability, modification of the target and hydrolysis of the antibiotic are the most important causes for resistance to  $\beta$ -lactams.

Permeability of the cell membrane for a certain antibiotic depends mainly on its hydrophobicity and steric properties. In intrinsically resistant microorganisms, e.g. pneumococci, often a qualitatively and quantitatively different expression pattern of penicillin binding proteins (PBPs) is observed. The same is true for cloxacillin-resistant *Bacillus megaterium* strains. Here the expression of the sensitive PBP2 is downregulated and that of PBP3 upregulated [9].

Three types of  $\beta$ -lactam hydrolyzing enzymes are known:  $\beta$ -lactamases (hydrolysis of the amide bond in the  $\beta$ -lactam ring), aminohydrolases (cleavage of the amide bond in the side chain at the  $\beta$ -lactam nucleus) and cephalosporin acetyl esterases (hydrolysis of the ester residue of 7-ACA-containing cephalosporins). Clinically, the formation of  $\beta$ -lactamases is the most important reason for antibiotic resistance [9]. These proteins are encoded on plasmids or chromosomes, are expressed constitutively or after induction and are secreted into the space between cell membrane and cell wall in Gram-positive strains or into the periplasmic space in Gram-negative bacteria [10]. Following the hydrolysis of the  $\beta$ -lactam ring, the antibiotic is believed to be further degraded.

### B 1.3 $\beta$ -Lactamases

$\beta$ -Lactamases [EC 3.5.2.6] can be grouped into four classes based on their primary sequence homology [11]. Classes A, C and D (serine- $\beta$ -lactamases) use a serine-dependent mechanism for the hydrolysis of the amide bond proceeding through an acyl-enzyme intermediate [12]. In contrast, class B enzymes (metallo- $\beta$ -lactamases) contain either one or two zinc(II) atoms. A hydroxide coordinated to one or two zincs is believed to be the nucleophilic agent [13, 14].

#### B 1.3.1 Serine- $\beta$ -lactamases

The class A enzymes are largely plasmid encoded, occur widely distributed among Gram-positive and Gram-negative bacteria and can hydrolyze both penicillins and cephalosporins. The clinically most important and best investigated  $\beta$ -lactamases of this group are TEM, first isolated in 1965, shortly after the introduction of ampicillin for clinical use [15], and the SHV enzyme with 68% sequence identity [16]. They are widespread among Enterobacteriaceae, such as *Escherichia coli* and *Salmonella paratyphi* and other Gram-negative species like *Haemophilus influenzae*, *Neisseria gonorrhoeae* and *Pseudomonas aeruginosa*. For these serine- $\beta$ -lactamases protein evolution under selective pressure has been demonstrated [17]. Selective pressure for mutations can either be imposed by the use of extended-spectrum  $\beta$ -lactams (ceftazidime, cefotaxime, aztreonam), the application of inhibitors or a combination of both. For the TEM and SHV enzymes this resulted in the evolution of  $\beta$ -lactamases that could cleave the extended-spectrum drug or escape inactivation by the inhibitor. Both types of mutations have been identified in clinical isolates.

Class C  $\beta$ -lactamases are mostly chromosomally encoded. However, since the first discovery of plasmid-borne *ampC* genes encoding class C  $\beta$ -lactamases in 1988, they have been detected in many regions of the world and contribute to the spread of multidrug resistant *Klebsiella pneumoniae*, *Escherichia coli* and other Enterobacteriaceae [18].

Enzymes of class D are the least understood. Bonnet *et al.* recently reported on the class D  $\beta$ -lactamase OXA-23 in *Proteus mirabilis*, isolated from a hospital in France and capable of hydrolyzing imipenem [19]. The class D OXA-10  $\beta$ -lactamase has been crystallized and some conclusions about the mechanism were made [20].

A number of inhibitors for serine- $\beta$ -lactamases are available. Clavulanic acid is a naturally occurring secondary metabolite produced by *Streptomyces clavuligerus* and its biosynthesis has been reviewed [21]. It is a  $\beta$ -lactam, structurally very similar to penicillins, and the acyl-enzyme intermediate binds tightly to the protein. Additionally, a number of clavulanic acid derivatives and other inhibitors such as sulbactams, tazobactams and boronate inhibitors have been synthesized [22-25].

### **B 1.3.2 Metallo- $\beta$ -lactamases**

Class B or metallo- $\beta$ -lactamases have long been considered as mere biochemical curiosities compared to the widespread serine- $\beta$ -lactamases. Metallo- $\beta$ -lactamase activity was first discovered in *Bacillus cereus* in 1966 [26]. In 1982 a metallo- $\beta$ -lactamase produced by *Stenotrophomonas maltophilia* was found. However, major concern was not raised until 1993, when it was found that the resistance to carbapenems of an increasing number of clinically noxious strains is caused by the synthesis of zinc(II)-containing  $\beta$ -lactamases [27]. To date metallo- $\beta$ -lactamases have been found in pathogenic species of *Bacteroides*, *Pseudomonas*, *Stenotrophomonas*, *Aeromonas*, *Chryseobacterium*, *Serratia*, *Klebsiella*, *Shigella* [28] and more recently in *Acinetobacter* [29, 30], in *Citrobacter youngae* [31], in *Alcaligenes xylosoxidans* [32], in *Chryseobacterium meningosepticum* [33] and *Caulobacter crescentus* [34]. Due to their broad spectrum of substrates (penicillins, cephalosporins, carbapenems) and their insensitiveness towards serine- $\beta$ -lactamase inhibitors, metallo- $\beta$ -lactamases have become a serious problem in nosocomial strains in Japan [35], China [31] and Italy [29]. Inhibitors, that act on metallo- $\beta$ -lactamases are mainly electron donors which coordinate to one or two zinc(II) atoms of the enzyme, such as biphenyl tetrazoles [36], thiol compounds [28, 37], thioesters, that are hydrolyzed to thiol compounds [38, 39], or succinic acids [40].

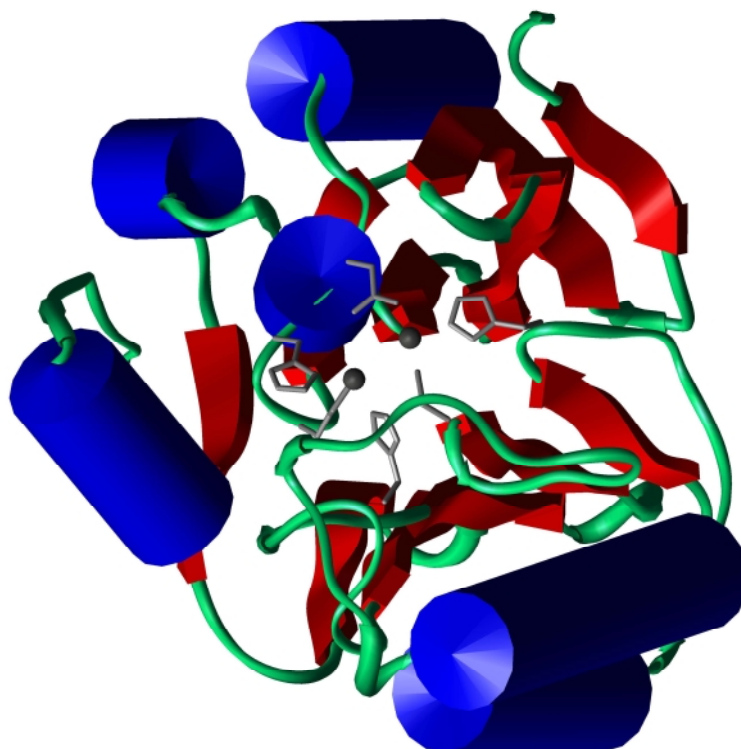
As shown for the serine- $\beta$ -lactamases, also metallo- $\beta$ -lactamases remain a challenge for chemotherapy due to their potential to improve  $\beta$ -lactam hydrolysis and to escape inhibitor binding by mutation under selective pressure. Some variants of the IMP-1 metallo- $\beta$ -lactamase originally isolated from *Pseudomonas aeruginosa* have been isolated: IMP-2 to IMP-10 with the number of amino acid substitutions ranging from one (IMP-6 [41]) up to 26 (IMP-2 [29]) compared to IMP-1 and with different substrate specificities. It has been assumed that IMP-1 might have evolved from IMP-3, because due to only one amino acid substitution it is capable of more efficient  $\beta$ -lactam hydrolysis [42].

### B 1.3.2.1 Structure of Metallo- $\beta$ -lactamases

In 1995 the crystal structure of the *Bacillus cereus* metallo- $\beta$ -lactamase was published [43]. The authors found that the protein revealed a new type of protein fold: a  $\beta\beta$  sandwich with helices on each external face. The  $\alpha\beta\beta\alpha$  fold is now known as the metallo- $\beta$ -lactamase fold. According to an internal molecular symmetry it might be the result of gene duplication. Based on database searches and protein sequence alignments, it has been found that at least 23 protein families have this fold in common and together form the metallo- $\beta$ -lactamase fold superfamily. Other families are e.g. the glyoxalase family, the alkylsulfatase family, the Rec2 family and the phosphodiesterase family. Two motifs are characteristic for this superfamily: 1. a  $\beta$ -strand with a conserved terminal aspartate and 2. a HxHxDH motif, with the third histidine being replaced by an arginine in metallo- $\beta$ -lactamases. In the mononuclear *B. cereus* metallo- $\beta$ -lactamase, the first two histidines participate in zinc coordination while the aspartate projects close to the active site [44].

Additionally to the mononuclear *Bacillus cereus* enzyme, three binuclear metallo- $\beta$ -lactamases from *Bacteroides fragilis* [45], *Stenotrophomonas maltophilia* [46] and from *Pseudomonas aeruginosa* IMP-1 [28] have been crystallized. In these structures the aspartate from the HxHxDH motif (in metallo- $\beta$ -lactamases HxHxDR) coordinates to the second zinc. Whereas all four enzymes have two metal binding sites, only the *B. fragilis*, the *S. maltophilia* and IMP-1 metallo- $\beta$ -lactamases bind two zincs with high affinity, thus enabling more efficient catalysis. the *B. cereus* enzyme can also bind a second zinc, but with lower affinity. As it can function with one zinc as well as with two zincs, it has been considered an intermediate between mono- and binuclear  $\beta$ -lactamases [47]. Concha *et al.* have assumed

that this polymorphism may represent a rapidly evolving adaptive mechanism [28]. In Figure B 4 the structure of the IMP-1 metallo- $\beta$ -lactamase is shown.



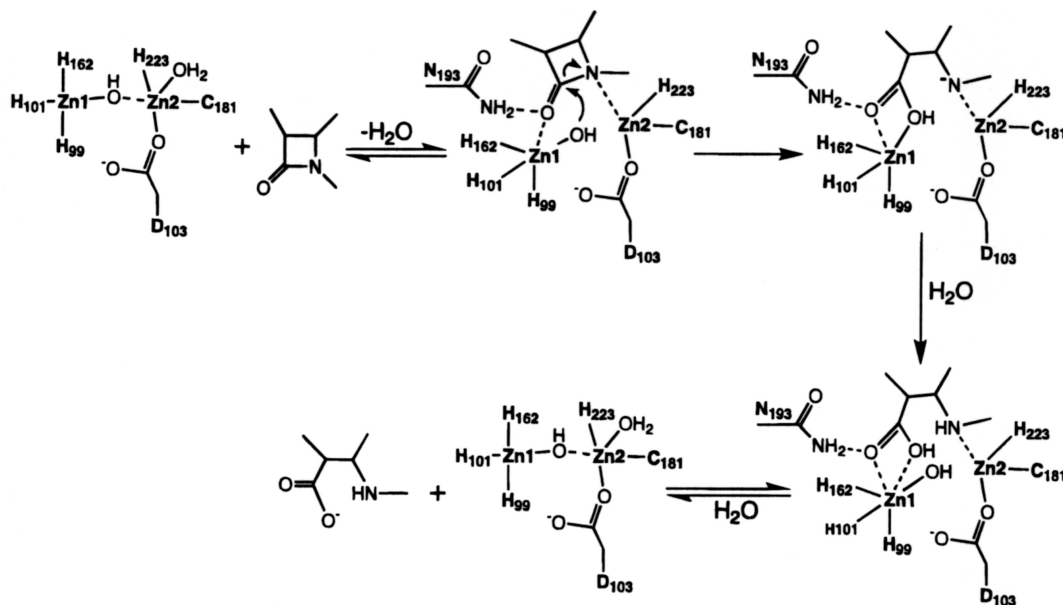
**Figure B 4.** Structure of the IMP-1 metallo- $\beta$ -lactamase.  $\alpha$ -helices are represented as blue cylinders,  $\beta$ -strands as red bands and other regions of the protein in green. The active site residues are depicted as grey sticks and the zincs in dark grey and CPK.

In the free enzyme one zinc (Zn1) is coordinated by three histidines and a hydroxide ion, which is assumed to be the nucleophilic agent for the cleavage of the  $\beta$ -lactam ring. The other zinc (Zn2) is coordinated by the hydroxide as well and additionally by the conserved aspartic acid, a cysteine, a histidine and a water molecule. Zn1 has a tetrahedral and Zn2 a trigonal bipyramidal coordination. The active site corresponds to those of the *B. fragilis* and the *B. cereus* enzymes. In the *S. maltophilia* enzyme other amino acids are involved in zinc coordination.

#### B 1.3.2.2 Mechanism of $\beta$ -Lactam Hydrolysis

The mechanism of the *Bacteroides fragilis* metallo- $\beta$ -lactamase, which is probably the same in IMP-1, has been studied based on the photometric measurement of nitrocefin hydrolysis

[13]. This cephalosporin is a chromophor and its absorbance maximum shifts upon hydrolysis: the substrate, the product and also an intermediate appear at different wavelengths. Thus it was shown that a relatively stable anionic intermediate binding to both zincs occurs (Figure B 5).



**Figure B 5.** Proposed mechanism of *B. fragilis* metallo- $\beta$ -lactamase catalysis [13]

Wang *et al.* proposed that the hydroxide attacks the carbonyl C of the amide bond while the substrate is stabilized by asparagine 193. After cleavage of the amide bond, an anionic intermediate is stabilized by coordination of the carboxyl group to Zn1 and coordination of the anionic N to Zn2.

The mechanism for the mononuclear *Bacillus cereus* metallo- $\beta$ -lactamase is similar to that of the binuclear enzymes. The carbonyl C of the amide bond is also attacked by a hydroxide, but the C-N cleavage occurs at or after the rate-limiting step and requires the protonation of the leaving lactam nitrogen [48]. In contrast, in the binuclear enzymes, Zn2 stabilizes the intermediate structure resulting in an acceleration of the C-N bond cleavage by a factor of 375 compared to the mononuclear enzyme. Therefore, a later event, the release of the intermediate initiated by protonation of the nitrogen, becomes the rate-limiting step. The  $k_{cat}/K_M$  values of the binuclear enzymes are generally 10 times higher. Therefore, the use of a second zinc is considered an evolutionary adaptation to selective pressure [49].

This possibility was further investigated by an artificial „backward evolution“ of the binuclear *B. fragilis* enzyme to a mononuclear enzyme corresponding to the *B. cereus*

metallo- $\beta$ -lactamase. Therefore a cysteine distant from the second zinc site was mutated to arginine as in the mononuclear enzyme and indeed the  $k_{cat}$  value for nitrocefin hydrolysis decreased. Removal of one equivalent of zinc from the mutant further decreased the  $k_{cat}$  and it closely approximated that of the mononuclear enzyme [49]. This is a further example of how a point mutation can significantly affect the efficiency of  $\beta$ -lactam hydrolysis.

Facing the problem of antibiotic resistance is essential to learn more about the mechanism of metallo- $\beta$ -lactamases and the way they have evolved and may evolve in the future under the selective pressure of antibiotic and inhibitor application. Therefore several mutation analyses have been carried out with metallo- $\beta$ -lactamases. The essential results of these are the following:

1. The type and the arrangement of the amino acids ligating to the zinc atoms are essential for efficient hydrolysis. If due to mutations no zinc binds to the second zinc site, activity may still be observed but decreases significantly [49-51].
2. A protonated lysine (K161 in IMP-1) interacts with the deprotonated carboxylate at the five- or six-ring of the  $\beta$ -lactam and has been shown to be essential for enzymatic activity [52].
3. A peripheral asparagine (N167 in IMP-1) that has been assumed to be involved in substrate binding [13, 28, 45] is not essential for  $\beta$ -lactam hydrolysis. For the second generation cephalosporins cefuroxime and cefoxitin and the carbapenem imipenem,  $k_{cat}/K_M$  values were slightly decreased when asparagine was mutated to alanine or aspartate [52]. For ampicillin, nitrocefin, cefotaxime and cefaloridine even an increase of  $k_{cat}/K_M$  was reported for the alanine mutant [51].
4. IMP-1 with a serine close to Zn2 (S196) converted cephalosporins with a pyridinium residue at the dihydrothiazine ring (cephaloridine and ceftazidime), imipenem and penicillins with significantly higher  $k_{cat}/K_M$  values than respective glycine mutants (IMP-3 and IMP-6) [42].

A further method that can be applied to learn more about the metallo- $\beta$ -lactamase mechanism and substrate specificities is computer-aided molecular modeling.

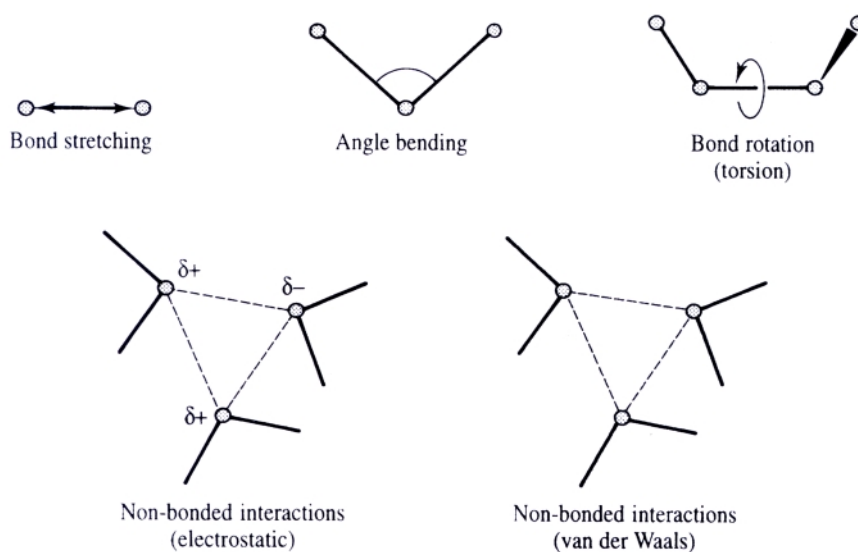


## B 1.4 Molecular Modeling

Molecular modeling is concerned with ways to mimic the behaviour of molecules and molecular systems [53]. Due to the complexity of molecular systems it is usually associated with computer techniques. Molecular systems can either be calculated with quantum-mechanical (QM) methods, with molecular-mechanical (MM) approaches or with a combination of both (QM/MM methods).

QM calculations are based on the Schroedinger equation and deal with the electrons in the system. They are more accurate but very time-consuming and at the present only feasible for small systems, but not for proteins or DNA molecules. However, they will probably gain importance with increasing computer power.

MM methods ignore the electronic motions and calculate the energy of a system as a function of the nuclear positions only. The force fields that are used in such systems are deduced from empirical contemplations and regard mainly the laws of Newton for electrostatic and van der Waals interactions and classical mechanical terms for bond and angle deformations and rotations about single bonds. Cut offs have to be set for long range interactions and approximations have to be made in order to avoid too large complexity and make the calculation practicable. A molecular mechanic force field is good, when it can mimic molecular systems with high fidelity and when the parameters developed from data on small molecules can be used to study larger molecules. MM methods are usually the method of choice for the modeling of biopolymers such as proteins. The contributions to a molecular mechanics force field are shown in Figure B 6.



**Figure B 6.** Schematic representation of the four key contributions to a molecular mechanics force field: bond stretching, angle bending and torsional terms and non-bonded interactions.

Simulations carried out with such force fields are also called molecular dynamics (MD) simulations. In cases where non-conventional bond types are to be modeled in a molecule, such as transition states in enzymatic reactions or coordinative bonds in proteins containing transition metals, MM methods may become insufficient.

### B 1.4.1 Molecular Dynamics Simulations of Metallo-proteins

Using force fields that are typically used for molecular dynamics (MD) simulations, such as the AMBER or the CHARMM force field, either bonded or nonbonded interactions can be calculated, but no coordinative bonds. This problem might be addressed adequately in the future using quantum-mechanical approaches. Such methods are being developed. They use a quantum-mechanical representation of the active site containing the metal center and a molecular-mechanical representation of the surrounding protein. An example for such simulations are Car-Parinello simulations named after the initiators [54] and have been applied e.g. for the simulation of a copper-containing galactose oxidase [55]. However, at the present these methods are not practicable for the simulation of big molecules over long periods.

Many metallo-proteins such as iron-containing P450 [56] and copper-containing plastocyanin [57] have been investigated using molecular dynamics simulations. Zinc-containing proteins are of increasing biological interest and a number of them have been

simulated, e.g. a carboxypeptidase A and a carbonic anhydrase with a nonbonded approach [58]. A metallo-proteinase [59], a farnesyltransferase [60] and a phosphotriesterase [61] have been modeled using a promising method called „cationic dummy atom approach“. In this procedure the zinc's  $4s4p^3$  vacant orbitals, which accommodate the lone-pair electrons of zinc coordinates [62] are mimicked by four dummy atoms which are located tetrahedrally around the zinc atom and carry a charge of +0.5 each. These dummy atoms are able to keep the ligands in the correct orientation, allowing a stable simulation of the tetrahedral zinc coordination. This procedure has been applied successfully for MD simulations of mononuclear [60] as well as binuclear zinc-containing proteins [61].

#### **B 1.4.1.1 Molecular Dynamics Simulations of Metallo- $\beta$ -lactamases**

The mononuclear *Bacillus cereus* enzyme has been simulated with a bonded approach [14, 63]. The enzyme was simulated either with a hydroxide or a water coordinated to the zinc. The zinc environment was also investigated with QM and QM/MM methods. It was shown that both configurations seem to be possible, but the binding of a hydroxide resulted in a much less fluctuating active site region [14]. Simulations with the respective protein in complex with benzylpenicillin suggested that according to the mechanism for mononuclear enzymes mentioned in B 1.3.2.2 the  $\beta$ -lactam hydrolysis functions via a one-step mechanism and that histidine 210 (corresponding to the histidine coordinating to Zn2 in binuclear metallo- $\beta$ -lactamases) acts as a proton donor.

The binuclear enzyme from *Bacteroides fragilis* was simulated using a mixed bonded and nonbonded approach [64]. The amino acids ligating to the zinc atoms were regarded as covalently bonded whereas the bridging hydroxide and the water coordinating to Zn2 were kept in place only by nonbonded interactions. In a simulation with benzylpenicillin, loop conformation changes and a decrease in flexibility of the protein as a whole were observed. In another study of the *B. fragilis* enzyme the bridging hydroxide was explicitly bonded to Zn2 in MD simulations [65]. Using QM calculations, the authors showed that the breaking of the Zn1-OH-Zn2 bridge was favored when the active site aspartate was protonated. This resulted in an increase of the Zn1-Zn2 distance to  $\sim 4.5$ - $5$  Å compared to  $3.5$  Å with the intact bridge.

## B 1.5 Aim of this Work

The aim of this work was to establish an *in silico* assay for the determination of substrate specificities of metallo- $\beta$ -lactamases. In order to achieve this, molecular dynamics simulations with the respective enzymes in complex with  $\beta$ -lactams should be carried out. The substrates should be docked in an intermediate structure, because the stability of this structure was assumed to be correlated to catalytic activity. The analysis of such an intermediate structure could give new insights into the mechanism of hydrolysis and the way the protein interacts with the substrate and the zinc atoms. Additionally, in these simulations parameters should be found that are a measure for the efficiency of  $\beta$ -lactam hydrolysis.

To develop the assay, experimental data that describe  $k_{cat}/K_M$  values for certain metallo- $\beta$ -lactamase variant/ $\beta$ -lactam antibiotic combinations should be taken as a training set to find parameters in the simulations that correlate to the experimental observations. It was of interest to investigate, why different substrates are converted at different rates, but also why different mutants have a different capability of hydrolyzing  $\beta$ -lactams. If parameters indicating efficient or inefficient enzymatic activity would be found, they could be applied to make predictions for enzyme/antibiotic combinations, that have not yet been investigated experimentally.

## B 2 Materials and Methods

### B 2.1 Materials

#### B 2.1.1 Hardware

**Table B 2.1.** Hardware components used in this work.

<b>Instrument</b>	<b>Operating System</b>	<b>Purpose</b>
PC	Windows 2000	analysis of data
Silicon Graphics Octane 2	UNIX	visualization file preparation MD simulations trajectory analysis
PC cluster	LINUX	MD simulations trajectory analysis

#### B 2.1.2 Software

**Table B 2.2.** Software used in this work.

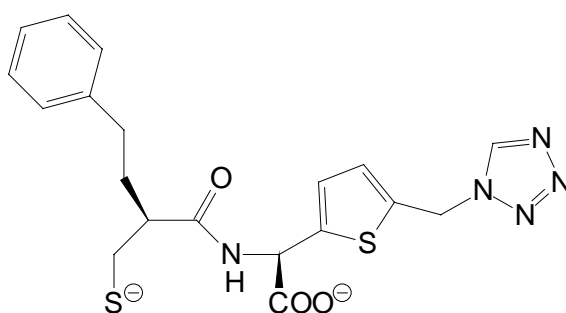
<b>Software</b>	<b>Purpose</b>
Nedit	text editing
InsightII MolBuilder AMPAC/MOPAC	visualization building of molecules and docking of substrates into proteins minimization of molecules
Gaussian 98	calculation of electrostatic potentials (ESPs)
AMBER 6.0 software package: RESP xLEap sander ptraj protonate ambpdb	fitting of electrostatic potentials (ESPs) file preparation for MD simulations minimizations, MD simulations trajectory analysis calculation of protonation state of titratable groups conversion of amber output files into PDB files
VMD viewer	visualization of trajectories
Origin 5.0	data processing and visualization
EXCEL	data processing and visualization

## B 2.2 Methods

### B 2.2.1 System Setup for MD Simulations

#### B 2.2.1.1 Definition of the Inhibitor, the $\beta$ -Lactams and the Hydroxide

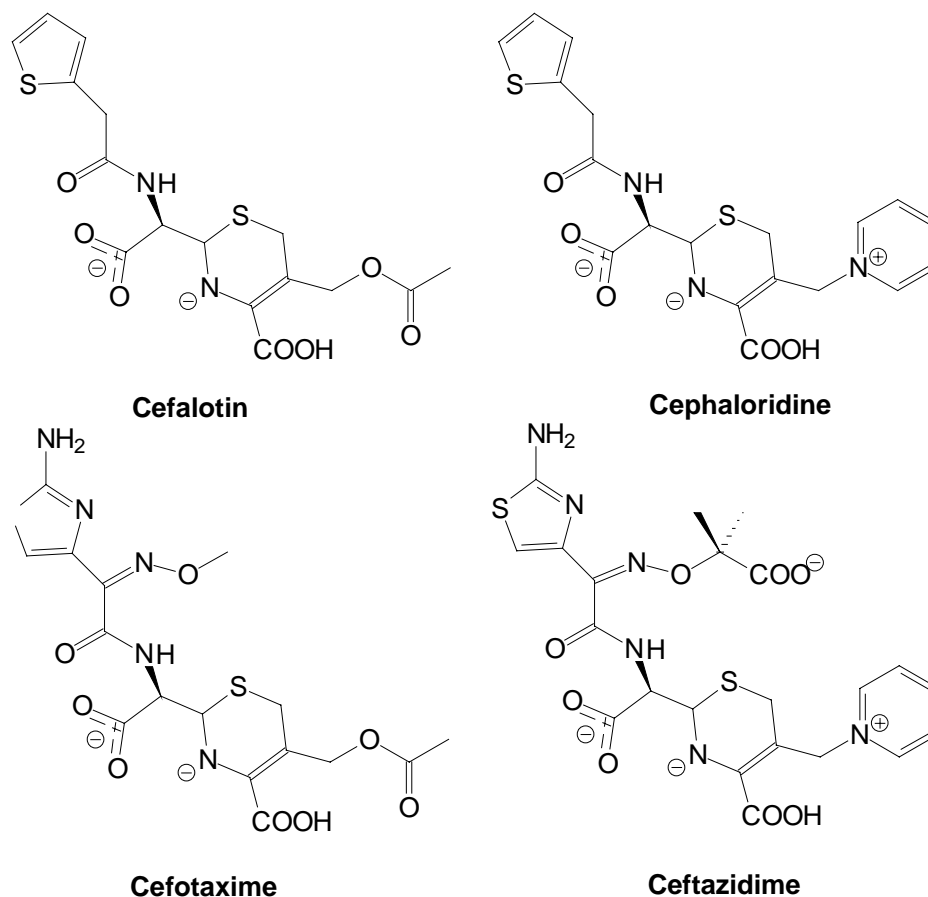
The mercaptocarboxylate inhibitor (MCI) and the cephalosporins cephalothin (CEF), cephaloridine (LOR), cefotaxime (CTX) and ceftazidime (CAZ) were built with the MolBuilder modul of the Insight program. The inhibitor was built with a deprotonated thiol group and either a protonated or deprotonated carboxyl group (Figure B 7).



**Figure B 7.** Mercaptocarboxylate inhibitor as built for MD simulations. Here the variant with the deprotonated carboxyl group is shown.

The cephalosporins were built as hydrolyzed intermediates (Figure B 8). The nitrogen resulting from the  $\beta$ -lactam bond cleavage was anionic. The carboxyl group resulting from the cleavage was treated as protonated in the purely nonbonded model and as deprotonated in the cationic dummy atom method. This was in accordance with the method, that recommends deprotonation of ligands in the first coordination shell of the zinc(II) atoms [60]. Accordingly, the carboxyl group at the six-ring was protonated in this model (second coordination shell of zinc) whereas in the purely nonbonded model this was deprotonated. It is assumed that the carboxylate group interacts with the protonated lysine 161. However, when it was deprotonated in the cationic dummy atom method, it coordinated to Zn<sup>2+</sup>. Additionally, the anionic intermediate then had a high negative charge (e.g. -3 for cephalothin). Therefore, in this case, the carboxylate at the six-ring was regarded as a second coordination shell ligand and as a proton acceptor compensating for the deprotonation of the carboxyl group resulting from lactam hydrolysis. The amino group at the thiazole ring of cefotaxime and ceftazidime was not protonated as this residue is located in a hydrophobic pocket. Additionally the

thiazole ring withdraws electron density from substituents. The carboxyl group of ceftazidime in the oxime side chain was deprotonated, as this residue was docked in a way that the carboxylate was on the surface of the protein and faced the solvent.



**Figure B 8.** The hydrolyzed cephalosporins as they were used for the cationic dummy atom method. For the purely nonbonded model the protonation of the carboxyl groups resulting from lactam cleavage and at the six-ring was reversed.

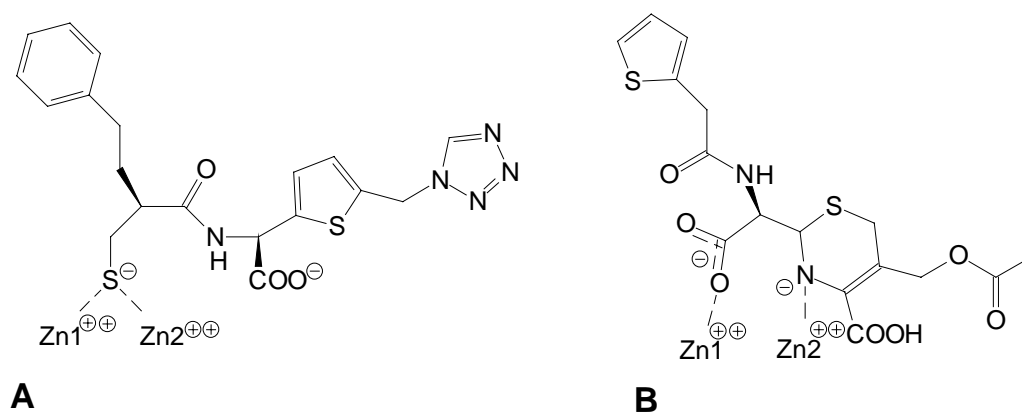
The structures were minimized semi-empirically with AMPAC/MOPAC and the Gaussian Z matrix output was used to create the input file for Gaussian 98. After quantum-mechanical calculation of the electrostatic potentials (ESPs) with Gaussian 98, the ESPs were fitted with the RESP software as suggested in the manuals: equivalent heavy atoms were set as equal in the first fit. These values were fixed and equivalent hydrogens were set as equal in the second fit.

PDB files of the minimized molecules were read into xLEap and bonds were assigned to atoms at a distance of up to 2.0 Å. The ESPs and an adequate atom type were assigned to each atom. For the thiophenyl ring two new atom types were defined: ST for the sulfur atom and CP for the carbon atoms in 2-position to the sulfur. The van der Waals parameters of S

and CA were used for these new types. Bond, angle and dihedral parameters for new conformations were taken over from the most closely related structures. For bonds involving ST and CP, the bond lengths from the AMPAC/MOPAC-minimized structures were used and the force constants were adapted according to similar bond types. The force field modification file is attached. The hydroxide molecule was defined as reported [60] (O of type OH, charge:  $-1.2049$ , H of type HO, charge  $+0.2049$ ). The molecules were then saved in a supplementary library each under a three-letter residue name.

### B 2.2.1.2 Docking of the $\beta$ -Lactams and the Hydroxide

The inhibitor was deleted from the PDB file and the hydrolyzed cephalosporins were docked with the Insight software according to the mechanism proposed by Wang *et al.* [13]: For the nonbonded approach the hydroxide oxygen of the carboxyl group resulting from  $\beta$ -lactam cleavage, for the cationic dummy atom model one of the two respective carboxylate oxygens was oriented towards Zn1. The anionic nitrogen was docked to Zn2. The thiophenyl- and thiazole-ring in R<sub>1</sub> were oriented in a similar way as the phenyl-containing residue of the inhibitor and the acetoxy group and pyridinium ring in R<sub>2</sub> according to the (5-methylene-4-tetrazole)-2-thiophenyl residue of the inhibitor. In Figure B 9 the binding mode of cephalothin is shown exemplarily in comparison to the inhibitor.



**Figure B 9A.** Schematic representation of the binding mode of the mercaptocarboxylate inhibitor with regard to the zinc atoms. **B.** Binding mode of the cephalothin intermediate.

The methoxy group of cefotaxime and the (carboxy-dimethyl)-methoxy group of ceftazidime were oriented towards the solvent-accessible region of their binding pocket. The structures with the docked substrates were saved as PDB files, the residue name of the new fragment was modified to the respective name in the supplementary library. The atom names of the



substrate had to be modified in a way, that xLEap recognized the molecule. Consequently the new residue was identified and the corresponding parameters assigned by xLEap upon reading in the PDB file when the supplementary library had been loaded before. The hydroxide molecule was docked by deleting the inhibitor except the sulfur atom and the neighbouring carbon atom and changing them to OH and HO and the residue name accordingly.

### B 2.2.1.3 Definition of the Active Site

#### B 2.2.1.3.1 *The Purely Nonbonded Model*

Several values have been suggested for the van der Waals parameters of the zinc(II) atoms in nonbonded models ( $\sigma = 0.69 \text{ \AA}$ ,  $\epsilon = 0.014$  [66],  $\sigma = 1.45 \text{ \AA}$ ,  $\epsilon = 0.025 \text{ kcal/mol}$  [67],  $\sigma = 1.95 \text{ \AA}$  and  $\epsilon = 0.25 \text{ kcal/mol}$  [58]). The values from Stote and Karplus were preferred for this work. Although developed for the CHARMM force field, they are very close to those of the chemically and physically similar copper atom in the AMBER force field ( $\sigma = 2.20 \text{ \AA}$  and  $\epsilon = 0.20 \text{ kcal/mol}$ ). Copper is in the 11<sup>th</sup> group and has an atomic mass 63.55, whereas zinc is in the 12<sup>th</sup> group and has the atomic mass 65.39. The formal charge of the zincs was +2. The histidines ligating to zinc were neutral, the aspartate, the cysteine and the thiol of the inhibitor were deprotonated.

#### B 2.2.1.3.2 *The Cationic Dummy Atom Model*

Zinc with dummies was built in Insight, read into xLEap and parametrized according to Pang *et al.* [60]: The central zinc atom was defined as ZD with an atomic mass of 61.38, no charge, a van der Waals radius of  $3.1 \text{ \AA}$  and an  $\epsilon$  of  $10^{-6} \text{ kcal/mol}$ . The tetrahedrally arranged dummy atoms were defined as DZ with an atomic mass of 1.0 and a +0.5 charge. Both van der Waals parameters were 0. The dummy atoms were at a distance of  $0.9 \text{ \AA}$  from the central atom. Force field parameters are listed in the force field parameter file. The structure was geometrically optimized in a vacuum minimization and saved as a PDB file and an entry in the supplementary library with the residue name ZDU.

In accordance with the cationic dummy atom method, titratable groups ligating to the zinc were all deprotonated resulting in negatively charged aspartate, cysteinate and histidinate. In the absence of zinc, histidine is usually neutral, but based on ab initio calculations it has been shown that the proton dissociation energy of imidazole when coordinated to  $\text{Zn}^{2+}$  is

dramatically reduced, suggesting the existence of imidazolate as zinc ligand in proteins [68]. Histidinate is not included in the AMBER topology files. Therefore it was defined as suggested by Pang *et al.* [60] and added to the supplementary library as HIT. The second coordination shell titratable group aspartic acid 170 was protonated.

#### **B 2.2.1.4 Protein Preparation**

The crystal structure of the IMP-1 metallo- $\beta$ -lactamase is deposited in the Protein Data Bank (PDB entry 1DD6 [28]). This file consists of two polypeptide chains. Chain A including the crystal water, zinc atoms and inhibitor was taken for the modeling. Chain B, its zincs, inhibitor and crystal waters were deleted. To determine the protonation state of titratable amino acids distant from the active site, the zinc(II) atoms were deleted and the protonate program was run on the apo protein at pH 7.0. The amino acids were protonated as suggested by the program except those in the active site. Histidine 19 was changed to the  $\epsilon$ -protonated tautomer (HIE), thus allowing hydrogen bonds to the backbones of valine 36 or isoleucine 47.

For the purely nonbonded model the ligating amino acids aspartic acid 81 and cysteine 158 were treated as deprotonated, histidines 77, 79, 139 and 197 as neutral. Therefore cysteine 155 had to be changed to CYM. Histidine 79 was changed to the  $\epsilon$ -protonated tautomer (HIE) so that the nitrogen oriented towards the zinc was deprotonated.

The cationic dummy atom procedure includes the deprotonation of titratable groups in the first coordination sphere of zinc and protonation of those in the second coordination sphere. According to this, aspartic acid 81 and cysteine 158 were deprotonated. In order to additionally deprotonate histidines 77, 79, 139 and 197, they were changed to HIT in the pdb file. Aspartic acid 170 (second coordination sphere) was interpreted as a proton acceptor for the proton from histine 79 and protonated. To introduce the zinc-dummy molecules instead of the Zn atoms into the protein, the Zn atoms (ZN residues) were renamed to ZD and the residue to ZDU. xLEap then automatically added the missing dummies when the pdb files were loaded.

Additionally to IMP-1, a S196G mutant (IMP-6) was simulated. This mutant could easily be constructed by deleting the side chain atoms of serine 196 and changing the residue name to GLY. xLEap then automatically added the missing hydrogen upon loading the file in the programm.

### B 2.2.1.5 Solvation and Neutralization of the Protein

Using the xLEap software, the structures with bound cephalosporins and the free enzyme (with the hydroxide as a bridging ligand between the zincs) were solvated with a shell of TIP3P water 2.0 Å around the protein with a minimal distance of 0.4 Å to allow water enter into empty places resulting from removal of the inhibitor. Then all structures were solvated with a truncated octaeder of TIP3P water 10.0 Å around the protein and neutralized. The proteins prepared for the nonbonded model were positively charged and had to be neutralized with one or two Cl<sup>-</sup>, depending on the molecules docked; those for the cationic dummy atom model were negatively charged and were neutralized with one or two Na<sup>+</sup> counter ions. Finally the system was saved as a topology file (*filename.parm*) and a coordinate file (*filenamd.xyz*).

### B 2.2.2 Minimization

Minimization is done in order to remove high-energetic bond deformations and van der Waals interactions and takes the molecular system into a local energy minimum, from which the MD simulation can be started.

Minimizations were performed with the sander software using a batch file script like the following (for IMP-1 in complex with cephalothin (CEF)):

```
cat << eof1 > IMP1_CEF_min.in
# Minimization of IMP-1 with explicit solvent
&cntrl
    imin=1, ntp=50, ntb=1, igb=1, maxcyc=1000, ncyc=500, ntr=1,
&end
# Group definition
    500
RES 1 2
END
# Group definiton
    500
RES 219
END
END
eof1
```

The first and last line indicate the beginning and the name as well as the end of the input file.

The input file itself describes the minimization procedure:

```
imin=1:          perform a minimization
ntp=50:          print the energy information to the output file every 50 steps
```

`ntb=1:` apply a periodic boundary with constant volume  
`igb=1:` use the generalized Born model and not the PME (Particle Mesh Ewald) method. These are methods of calculating long-range Coulomb interactions. The Born model needs more memory. The PME method is an approximation.  
`maxcyc=1000:` minimize for at the maximum 1000 cycles  
`ncyc=500:` after 500 cycles change the method of minimization from steepest descent to conjugate gradient  
`ntr=1:` apply restraints to the defined residues

The restraint residues were 1, 2 and 219, the two zincs and the substrate and a force constant of 500 was applied. A high value for the force constant was chosen because the docked structure was only hypothetical and had to be stabilized until the system was geometrically optimized. For the minimization the Born model was used for the calculation of long-range Coulomb interactions (`igb=1`). This procedure is more accurate but requires more calculation time and memory. However, the initial minimization is crucial and there disadvantages are accepted.

The command to start the minimization with this script was:

```
sander -O -i IMP1_CEF_min.in -o IMP1_CEF_min.out -p IMP1_CEF.parm -c
IMP1_CEF.xzy -ref IMP1_CEF.xzy -r IMP1_CEN_min.xyz
```

`sander:` start the sander program  
`-O:` overwrite already existing files with the same name  
`-i, -o, -p, -c, -ref, -r:` indicate that an input, output, topology, coordinate, reference or restart file follows

The scripts for the minimization, the heating and equilibration steps and the MD simulations were usually written in one batch script to have them performed subsequently.

### B 2.2.3 Heating

One or two separate heating periods were used for simulations at 100 K and 4 for simulations at 300 K with increasing time constants (`tautp`) for heat bath coupling (`ntt=1`). The volume

was kept constant ( $ntb=1$ ). To investigate reproducibility, several simulations of the same system were initiated with different starting velocities, which were assigned at 10 K at the first heating step by a random number generator. This is dependent on a random number generator seed ( $IG=x$ ). For the following steps, the velocities of the respective previous coordinate files were read in. A typical script for an initial heating step from 10 K to 50 K is

```
cat << eof2 > IMPl_CEF_heat1.in
# Heating to 50 K with constant volume
&cntrl
    ntx=1,  irest=0,  ntp=50,  ntwr=1000,  iwrap=0,  ntwx=250,  ntf=2,
    ntc=2,  tol=0.00000001,  ntb=1,  igb=0,  ntr=1,
    nstlim=1000,  dt=0.001,
    temp0=50.0,  tempi=10,  IG=30,  ntt=1,  tautp=0.2,
&end
# Group definition
    100
RES    1      2
END
# Group definiton
    100
RES    219
END
END
eof2
```

meaning

$ntx=1$ :	velocities are not read in from another coordinate file
$irest=0$ :	no restart calculation
$ntp=50$ :	print the energy information to the output file every 50 steps
$ntwr=1000$ :	write a restart file every 1000 steps
$iwrap=0$ :	molecules are not translated back to the primary box
$ntwx=250$ :	write the coordinates to the tractectory (.crd) file every 250 steps
$ntf=2$ :	bond interactions involving hydrogens are omitted
$ntc=2$ :	bonds involving hydrogens are constrained (SHAKE)
$tol=0.00000001$ :	relative geometrical tolerance for coordinate resetting in SHAKE is $10^{-8}$ Å
$ntb=1$ :	apply a periodic boundary with constant volume
$igb=0$ :	the PME (Particle Mesh Ewald Method) is used
$ntr=1$ :	apply restraints to the defined residues
$nstlim=1000$ :	perform 1000 MD-steps

---

dt=0.001:	use 1 fs time step
temp0=50.0:	keep the system at 50 K
tempi=10.0:	the initial temperature is 10 K
IG=30:	the seed for the random number generator
ntt=1:	the temperature is kept constant using the Berendsen coupling algorithm [69]
tautp=0.2:	the time constant for heat bath coupling for the solute is 0.2 ps

A much smaller force constant (100) could be used for the zincs and the substrate during the heating period. The sander command was similar to the one for the minimization. Now the restart file from the minimization was used as coordinate and reference file (-c, -ref) and additionally a trajectory output file was created (-x IMP1\_CEF\_heat1.crd).

For the following heating steps the tautp value was gradually increased and in order to read in velocities from the previous restart file, ntx was put to 7 and irst to 1.

#### **B 2.2.4 Equilibration**

In the equilibration, the constraints were gradually released after the system had reached its final temperature. In the first period, the force constant of the constraints in the last heating period was taken over and then halved in each following equilibration period until the system was finally let free. Now the pressure was kept constant instead of the volume (ntb=2, ntp=1). The time constant for heat bath coupling for the solute and the pressure relaxation time were gradually increased until the final values equal to those used in the unconstrained simulation were reached. An input file for a first equilibration is shown below:

```

cat << eof6 > IMPl_CEF_eq1.in
# Initial equilibration with constant pressure
&cntrl
    ntx=7,  irest=1,  ntp=50,  ntwr=1000,  iwrap=0,  ntwx=250,  ntf=2,
    ntc=2,  tol=0.00000001,  ntb=2,  igb=0,  ntr=1,
    nstlim=5000,  dt=0.001,
    temp0=300.0,  tempi=300.0,  ntt=1,  tautp=0.5,
    ntp=1,  pres0=1.0,  taup=0.7,
&end
# Group definition
    100
RES    1    2
END
# Group definiton
    100
RES    219
END
END
eof6

```

New variables are:

ntx=7:	formatted coordinates and velocities are read in
irest=1:	restart calculation
ntb=2:	use a periodic boundary with constant pressure
ntp=1:	MD with isotropic position scaling
pres=1.0:	reference pressure (in bar) at which the system is maintained
taup=0.7:	pressure relaxation time (in ps)

The input command was as for heating periods with the respectively altered file names.

### B 2.2.5 Unconstrained MD Simulation

The input file for an unconstrained MD simulation corresponded to the equilibration input files with the exception, that no constraints were imposed, that tautp values were between 0.8 and 1.0, that taup values were between 1.0 and 1.2 and that the number of simulation steps was much higher (between 100,000 equaling 100 ps and 1,000,000 equaling 1 ns with a time step of 1 fs).

The most important parameters for heating, equilibration and MD simulation are summarized in Table B 2.3.

**Table B 2.3.**

<b>Period</b>	<b>Parameters</b>
Heating	ntf=2, ntc=2 (SHAKE applied to hydrogen bonds) ntr=1 (restraints are applied) ntb=1 (periodic boundary with constant volume)
Equilibration	ntf=2, ntc=2 (SHAKE applied to hydrogen bonds) ntr=1 (restraints are applied) ntb=2 (periodic boundary with constant pressure)
MD simulation	ntf=2, ntc=2 (SHAKE applied to hydrogen bonds) ntb=2 (periodic boundary with constant pressure)

The simulation of the free enzyme was carried out with SHAKE applied to all atoms (ntf=3, ntc=3), as coordinate resetting could not be achieved otherwise.

## **B 2.2.6 Analysis of MD Simulations**

### **B 2.2.6.1 Visual Analysis**

In order to obtain insight into the structure of the last simulation step, a pdb file was created with the ambpdb program from the topology and the restart coordinate file. The waters were deleted and the structure visualized in Insight.

Average structures over a certain number of trajectory frames (typically over 10 ps at the end of the simulation time) were created with the ptraj program. Distances and angles of interest were measured in Insight. Superimpositions were made via the backbone atoms.

Trajectories were analyzed visually using the VMD viewer.

### **B 2.2.6.2 Energy Analysis**

The potential energy of the total system gives information about the energetic equilibration of a molecular system. This value (named EPTOT in the output file) was extracted from the output file with a perl script over the course of the simulation. Trajectory analysis was started after the total potential energy was equilibrated.



**B 2.2.6.3 Root Mean Square (RMS) Deviation**

The backbone RMS value of each structure during the simulation compared to the first structure analyzed gives information about the geometric equilibration of the protein. If the RMS value remains at a constant value (typically between 1.0 and 1.5 Å), the system is considered equilibrated. However, it has to be kept in mind, that significant changes in the RMS values due to conformational changes may occur after long periods of saturation.

The RMS values of the backbone atoms were analyzed with the ptraj program with the first analyzed structure as a reference.

**B 2.2.6.4 Measurement of Geometric Parameters Over the Course of Simulations**

Geometric parameters like distances and angles were measured over the course of the simulation using the ptraj program.

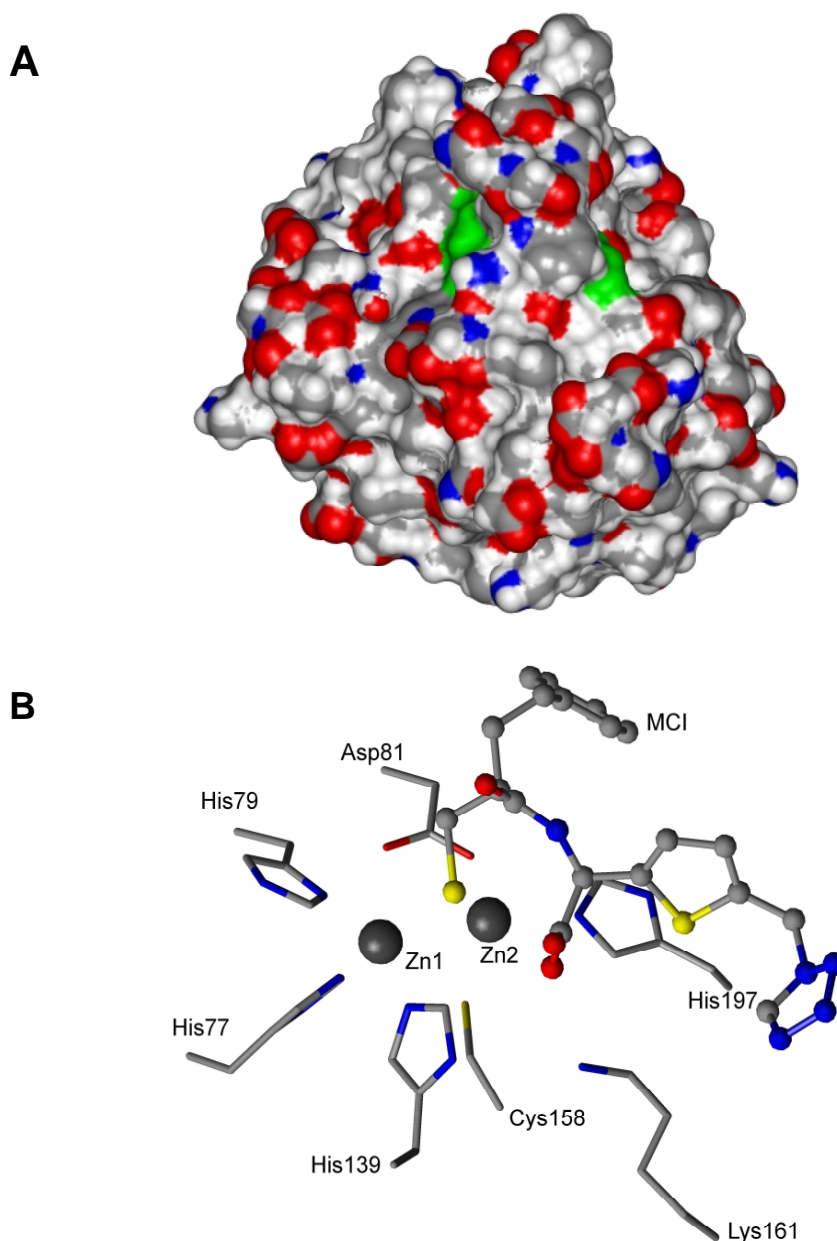
## B 3 Results

The idea of this work was to model metallo- $\beta$ -lactamases in complex with  $\beta$ -lactam antibiotics in order to learn more about the mechanism, the substrate specificity and also the effect of mutations. A model had to be established that gives a good representation of the protein. No crystal structures of metallo- $\beta$ -lactamases in complex with antibiotics are available. Therefore a crystal structure of the IMP-1 metallo- $\beta$ -lactamase from *Pseudomonas aeruginosa* in complex with a mercaptocarboxylate inhibitor [28] was selected. As this inhibitor is structurally similar to cephalosporins, these could be docked according to the inhibitor. To develop a reliable procedure for molecular dynamics (MD) simulations, the protein in complex with the inhibitor was used. Then the method was extended to the enzyme in complex with substrates. They were docked in an intermediate structure that is close to the transition state with the highest energy. The activation energy necessary to obtain the transition state is related to the conversion rate: the higher the activation energy, the smaller the conversion rate. This is also true for an intermediate close to the transition state: if its energy is high and it is instable, the conversion rate is low. Therefore it was expected to see different geometries or movements when  $\beta$ -lactams that show different stability towards the enzyme were docked or when enzymes with different substrate specificities were used.

### B 3.1 Development of the Model for IMP-1 in Complex with the Mercaptocarboxylate Inhibitor

A model had to be developed that allowed stable simulation of the protein and at the same time flexibility in the active site. The latter is important in order to investigate movements that are crucial for the mechanism. For the treatment of the zinc atoms the nonbonded approach was chosen, because covalent binding of the zincs would have led to high rigidity. The primary goal was to simulate the enzyme as it had been crystallized in complex with a mercaptocarboxylate inhibitor to 2.0 Å [28] without any significant conformational changes.

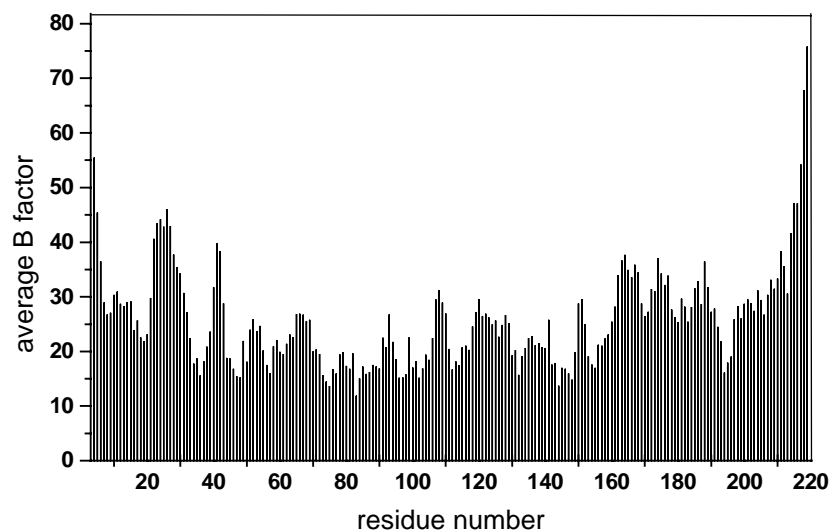
As mentioned in the introduction, metallo- $\beta$ -lactamases have an  $\alpha\beta\beta\alpha$  fold as was observed in crystal structures [28, 43, 45]. The active site is open on two sides towards the solvent and covered by a lid (Figure B 10A). The inhibitor, colored in green, is bound in the the enzyme. The active site is well organized (Figure B 10B).



**Figure B 10A.** A Connolly-surface representation of the IMP-1 enzyme in complex with the mercaptocarboxylate inhibitor (MCI). Grey: C, red: O, blue: N, white: H, green: MCI. **B.** Active site of IMP-1 in complex with MCI. Coloring of atoms as above. Yellow: S. Amino acids are displayed as sticks and are labeled at the C $\alpha$  atom, the inhibitor as balls and sticks, the zincs as CPK in dark grey.

The two zinc(II) atoms are at a distance of 3.6 Å, the distances between the ligands and the zincs are between 2.1 and 2.4 Å with the sulfur atom of the inhibitor being a bridging ligand between both metal atoms (Figure B 10B). The zincs have the coordination number 4 and the ligand-zinc-ligand angles are all close to the ideal tetrahedral angle of 109.5° with values ranging from 100° to 125°. Protonated lysine 161 is important for binding of the docked

molecules and interacts with their carboxylate. In Figure B 11 the average B factors of the backbone atoms in the crystal are shown for each residue.

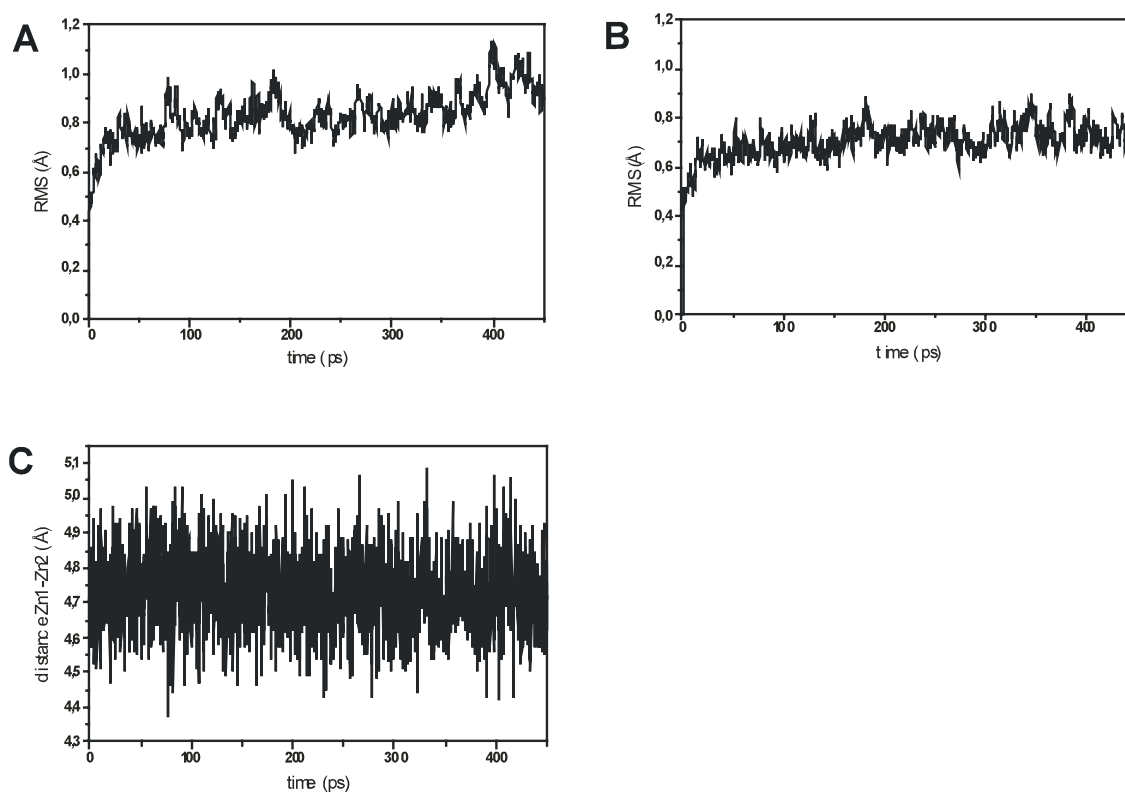


**Figure B 11.** B factors of the backbone atoms in average for each residue.

The highest B factors were observed in the N- and C-terminus and additionally in the region from residues 22 to 32, corresponding to the lid that covers the active site. The first method tested for MD simulations was a purely nonbonded approach.

### **B 3.1.1 The Purely Nonbonded Approach**

In the purely nonbonded approach the zinc(II) atoms were treated as having a formal charge of +2, a van der Waals radius  $\sigma$  of 1.95 Å and an  $\epsilon$ -value of  $10^{-6}$  [58]. A 450 ps simulation was carried out with this model at 300 K with the molecules solvated in a periodic water box. The RMS values of the protein backbone and the distance between the two zinc-atoms over the course of the simulation are shown in Figure B 12. As the N-, the C-terminus and the lid region had the highest B factors and were expected to be very mobile, resulting in high RMS values, the RMS values were additionally determined for residues 50-200. This region includes all the active site residues but excludes the regions with very high B factors.



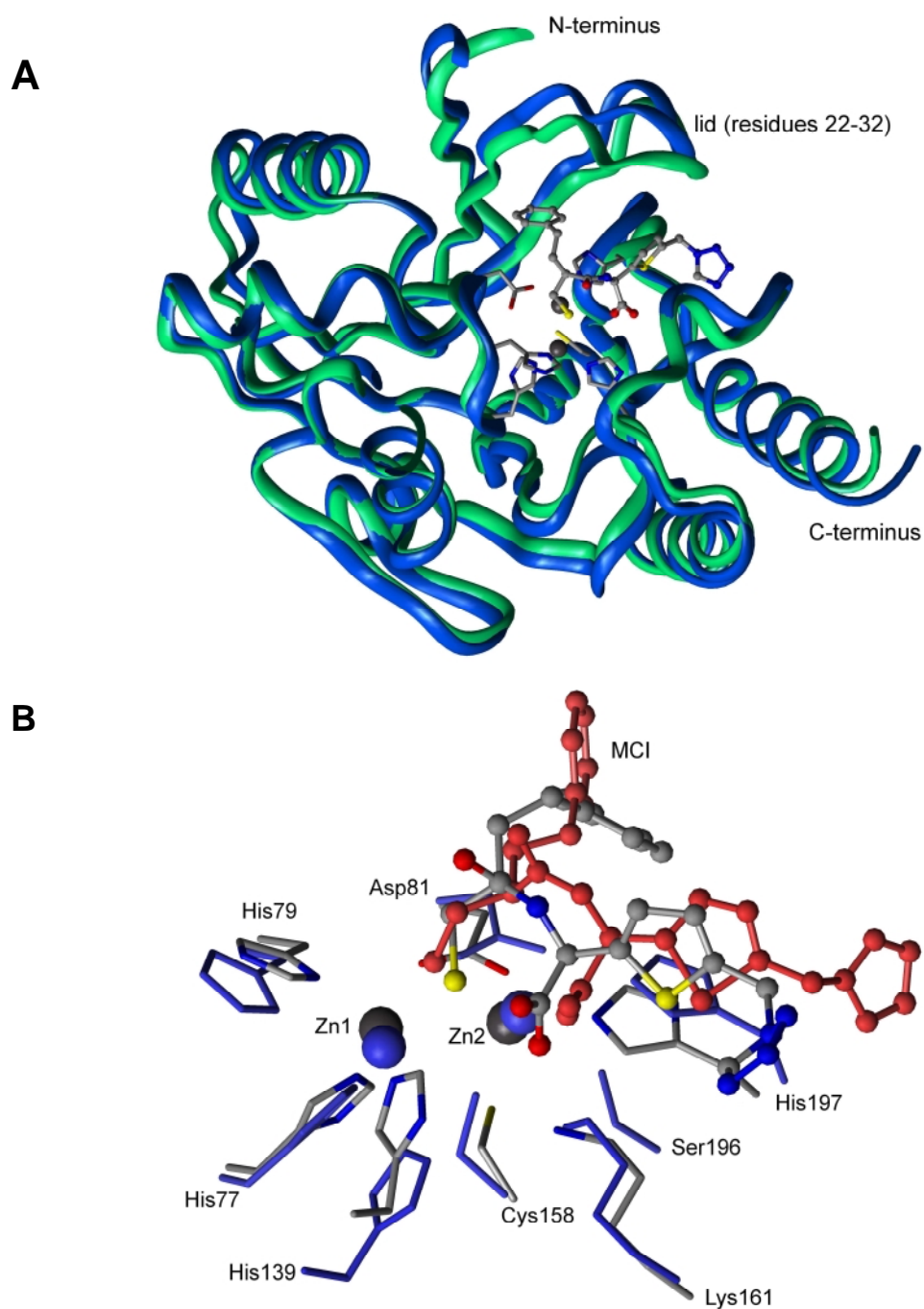
**Figure B 12A.** Backbone RMS values of the whole protein.

**B.** Backbone RMS values of the region from residue 50 to 200.

**C.** Zinc-zinc distance over the course of the purely nonbonded simulation of IMP-1 in complex with the mercaptocarboxylate inhibitor.

The total system was not yet geometrically equilibrated as can be seen from Figure B 12A (RMS values of 1.0 Å, still increasing). However, the part of the enzyme from residue 50-200 was geometrically equilibrated after 450 ps (RMS values 0.7 Å and constant, Figure B 12B). The distances and angles in the active side remained stable. In Figure B 12C this is shown exemplarily for the zinc-zinc distance. The other distances and angles were also constant (data not shown).

An average structure was made over the last 10 ps of the simulation and overlaid via the backbone with the crystal structure (Figure B 13).



**Figure B 13A.** The modeled structure and the crystal structure overlaid via the backbone atoms. The backbones are displayed as ribbons. green: crystal structure, blue: modeled structure. The active site residues of the crystal structure are displayed.

**B.** The active sites of the crystal structure and the modeled structure after the purely nonbonded simulation. Crystal structure: colored by atom.

Light grey: C, red: O, blue: N, yellow: S, dark grey: zincs, modeled structure: amino acids and zincs in blue, inhibitor in red. Amino acids are displayed as sticks and labeled at the C $\alpha$  atom, MCI: mercaptocarboxylate inhibitor as balls and sticks, zincs as CPK.

The average structure obtained from the nonbonded model deviated from the crystal structure.

In Figure B 13A the biggest deviations between the two backbones are in the termini and the

lid. This was expected, because without these regions the RMS values were much lower than when they were included. When analyzing the trajectory visually, a back and forth motion was visible. The other regions were comparable concerning the backbones. However, the side chains in the active site deviated significantly from the crystal structure, although they remained stable during the MD simulation. Distances and angles were measured and gave the following results:

1. The zinc-zinc distance had increased by 1.1 Å (from 3.6 Å to 4.7 Å).
2. The ligand-zinc distances had increased by 0.7 Å in average.
3. The ligand-zinc-ligand angles had changed by up to 47° from the original values and showed a strong deformation of the tetrahedral angle of 109.5°.
4. Additional ligands had entered into the first coordination sphere of the zinc(II) atoms. The sulfur atom of cysteinate 158 now functioned as a bridging ligand between the two zincs, resulting in a coordination number of 5 for Zn1. The second oxygen atom of aspartate 81, an oxygen atom of the inhibitor carboxylate and the oxygen from serine 196 had entered into the first coordination sphere of Zn2, changing its coordination number to 7 (Figure B 13B).

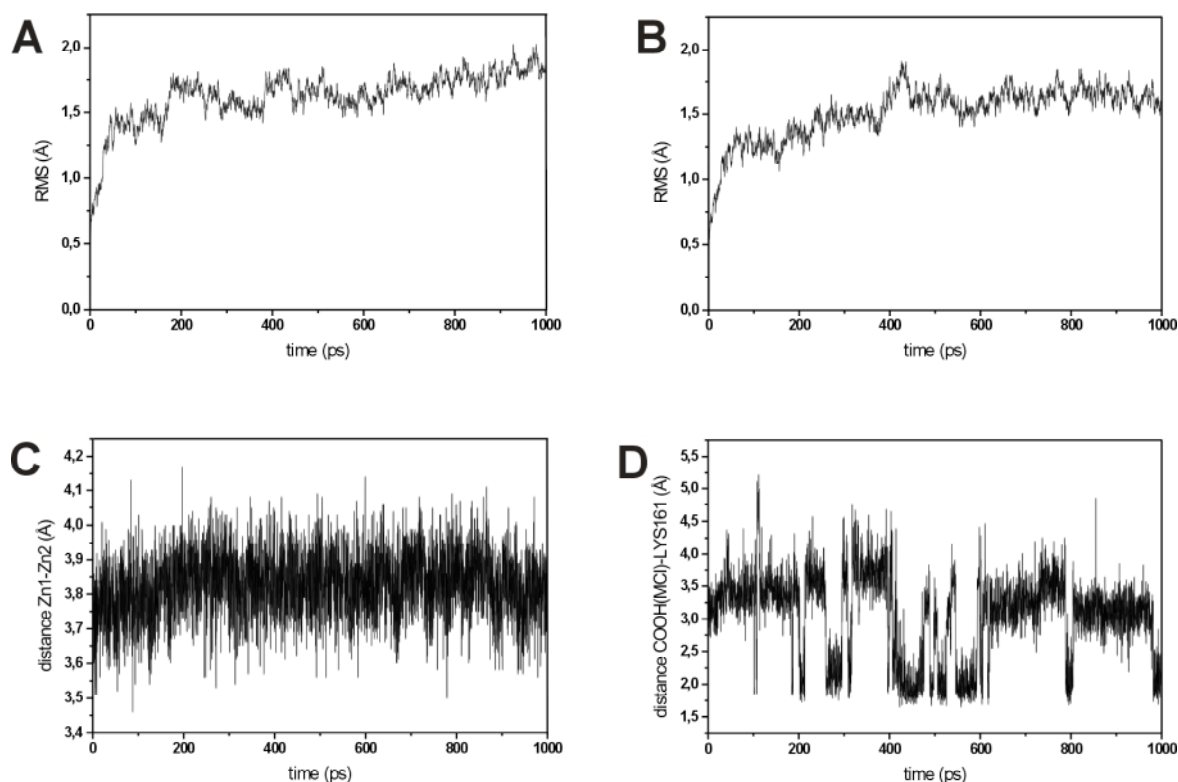
Details on distances and angles in the active site will be summarized in B 3.1.4. In conclusion, the active site took in a stable structure that did not correspond to the crystal structure and was therefore not suitable for further investigation.

### **B 3.1.2 The Cationic Dummy Atom Approach**

In the cationic dummy atom approach, the zinc atoms were surrounded by four dummy atoms carrying the charge of +0.5 each and mimicking vacant  $4s4p^3$  orbitals. All ligands of the first coordination shell of the zincs were deprotonated and those of the second coordination shell protonated. It was not clear how to treat the carboxyl group of the inhibitor. On the one hand protonation would be in accordance with the cationic dummy atom model [60] as it can be considered as a second coordination shell ligand (distance to Zn1: 4.43 Å, to Zn2: 3.57 Å). On the other hand, the carboxylate is known to interact with protonated lysine 161. Therefore both methods were tested.

### B 3.1.2.1 The Cationic Dummy Atom Approach With Protonated Inhibitor

A 1 ns MD simulation of the IMP-1 metallo- $\beta$ -lactamase was carried out with the inhibitor treated as protonated. The RMS values for the whole protein and for the region from residue 50-200, the zinc-zinc distance and the lysine 161-inhibitor carboxylate distance over the course of the simulation are shown in Figure B 14.



**Figure B 14A.** RMS values of the backbone of the whole protein over the course of the simulation with the cationic dummy atom method and protonated inhibitor.

**B.** RMS values of the backbone of residues 50-200.

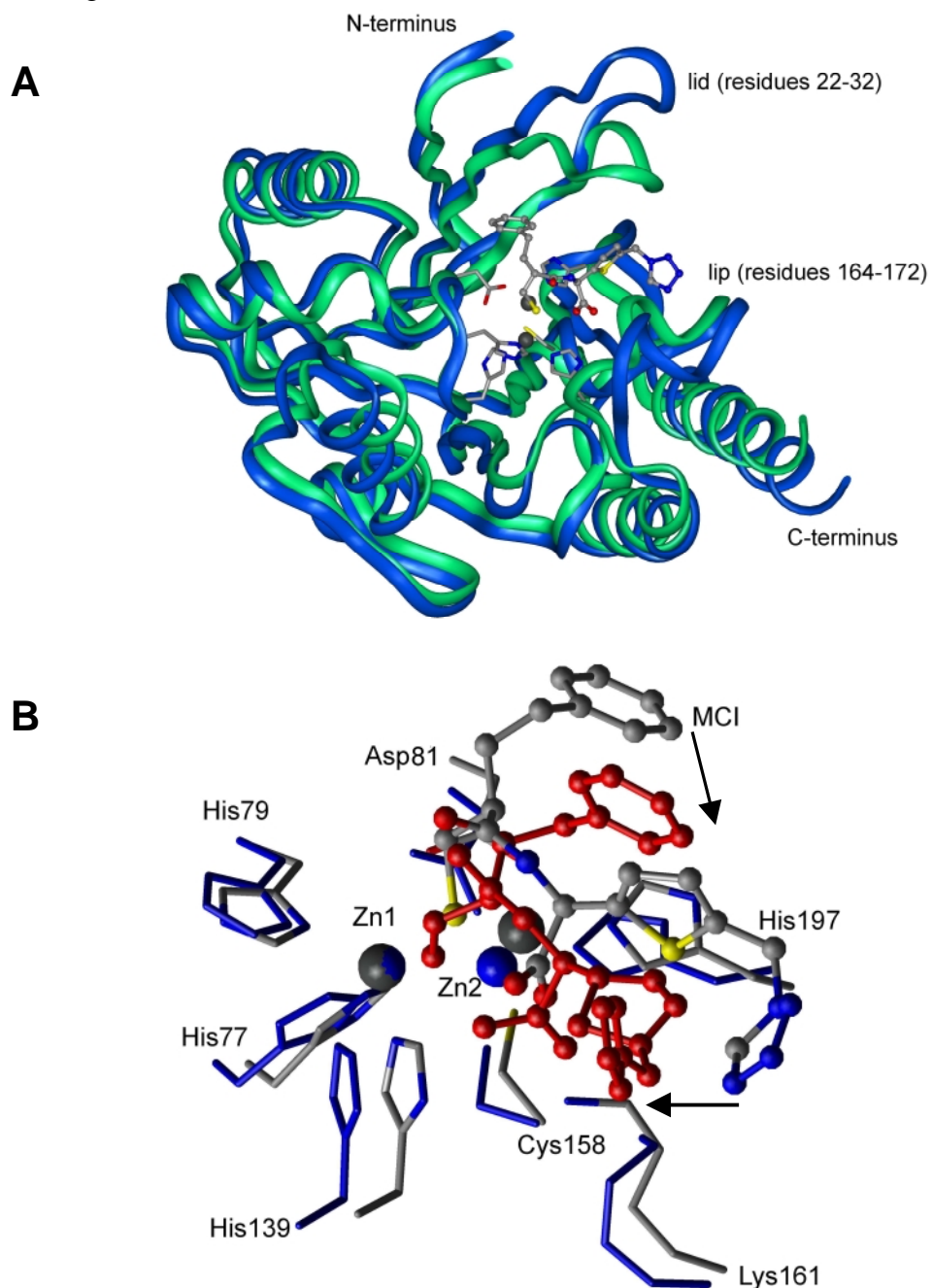
**C.** Zinc-zinc distance during the simulation.

**D.** Distance between the carbonyl oxygen of the carboxyl group of the inhibitor and the hydrogen of lysine 161 that was closest to the oxygen at the end of the simulation.

As in the simulation using the purely nonbonded model, the RMS values of the whole protein still increased after 1 ns (1.8 Å, Figure B 14A), but the region from residue 50 to 200 seemed to be equilibrated. Here the RMS values were higher than in the purely nonbonded simulation (1.5 Å, Figure B 14B) compared to 0.7 Å (Figure B 12B)). The zinc-zinc distance was at about 3.8 Å (Figure B 14C) throughout the simulation and therefore close to the crystal structure (3.6 Å). The distance between the carboxyl group oxygen of the inhibitor and a lysine 161 hydrogen alternated between 2 and 3-4 Å, due to exchange of the carbonyl oxygen with the hydroxyl oxygen and the 3 hydrogens bound to the side chain nitrogen of lysine 161.



An average structure of the last 10 ps of the simulation was made and overlaid with the crystal structure (Figure B 15).



**Figure B 15A.** The structure modeled with the cationic dummy atom method and protonated inhibitor and the crystal structure overlaid via the backbone atoms. The backbones are displayed as ribbons. green: crystal structure, blue: modeled structure. The active site residues of the crystal structure are displayed.

**B.** The active sites of the crystal structure and the modeled structure after the simulation. Crystal structure: colored by atom.

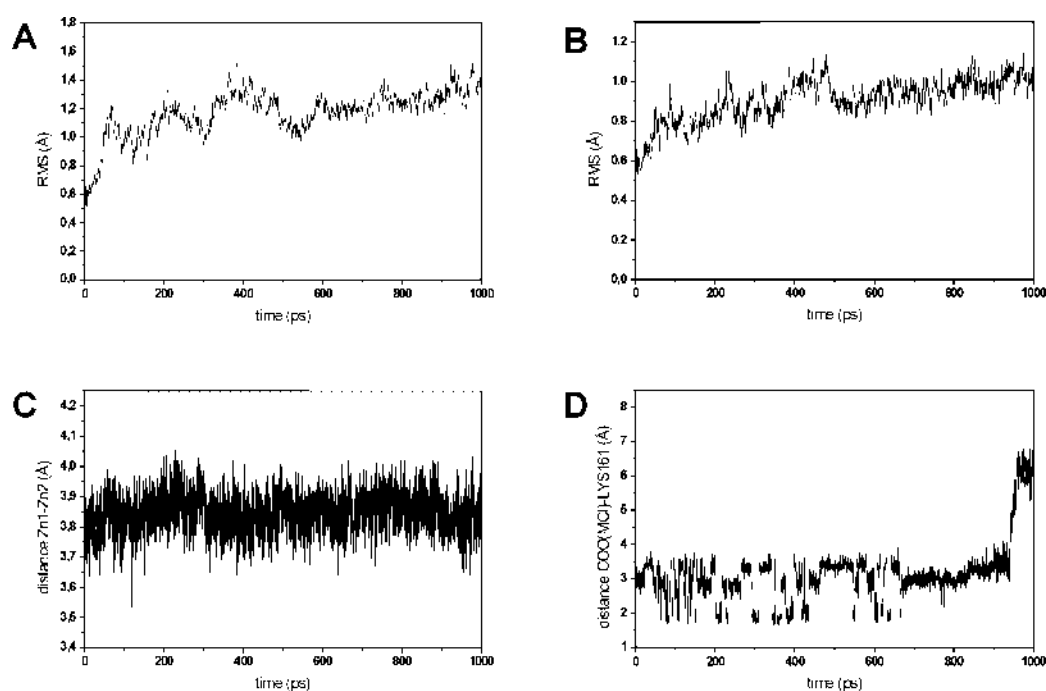
Light grey: C, red: O, blue: N, yellow: S, dark grey: zincs, modeled structure: amino acids and zincs in blue, inhibitor in red. Amino acids are displayed as sticks and labeled at the C $\alpha$  atom, MCI: mercaptocarboxylate inhibitor as balls and sticks, zincs as CPK.

Movements of the inhibitor are indicated with arrows.

Again a big movement in the lid (residues 22-32) was observed. Additionally, the loop consisting of residues 164-172 had moved. This loop was called “lip” in this work and is located opposite of the lid with the active site between them. Interestingly, this region also showed higher average B factors of the backbone atoms in the crystal structure (Figure B 11). The side chains of the active site amino acids were at slightly different positions, but the coordination of the zincs remained stable throughout the simulation. Details are summarized in B 3.1.4. The inhibitor side chains had moved out of their original position. Visual inspection of the trajectory showed that the opening of the lid was accompanied by a movement of the phenyl ring of the inhibitor. However, the thiolate remained coordinated to both zincs.

### B 3.1.2.2 The Cationic Dummy Atom Approach With Deprotonated Inhibitor

A 1 ns MD simulation of the IMP-1 metallo- $\beta$ -lactamase was carried out with the inhibitor treated as deprotonated. The RMS values of the total protein, of residue 50-200, the zinc-zinc distance and the distance between the carboxylate of the inhibitor and lysine 161 over the course of the simulation are shown in Figure B 16.



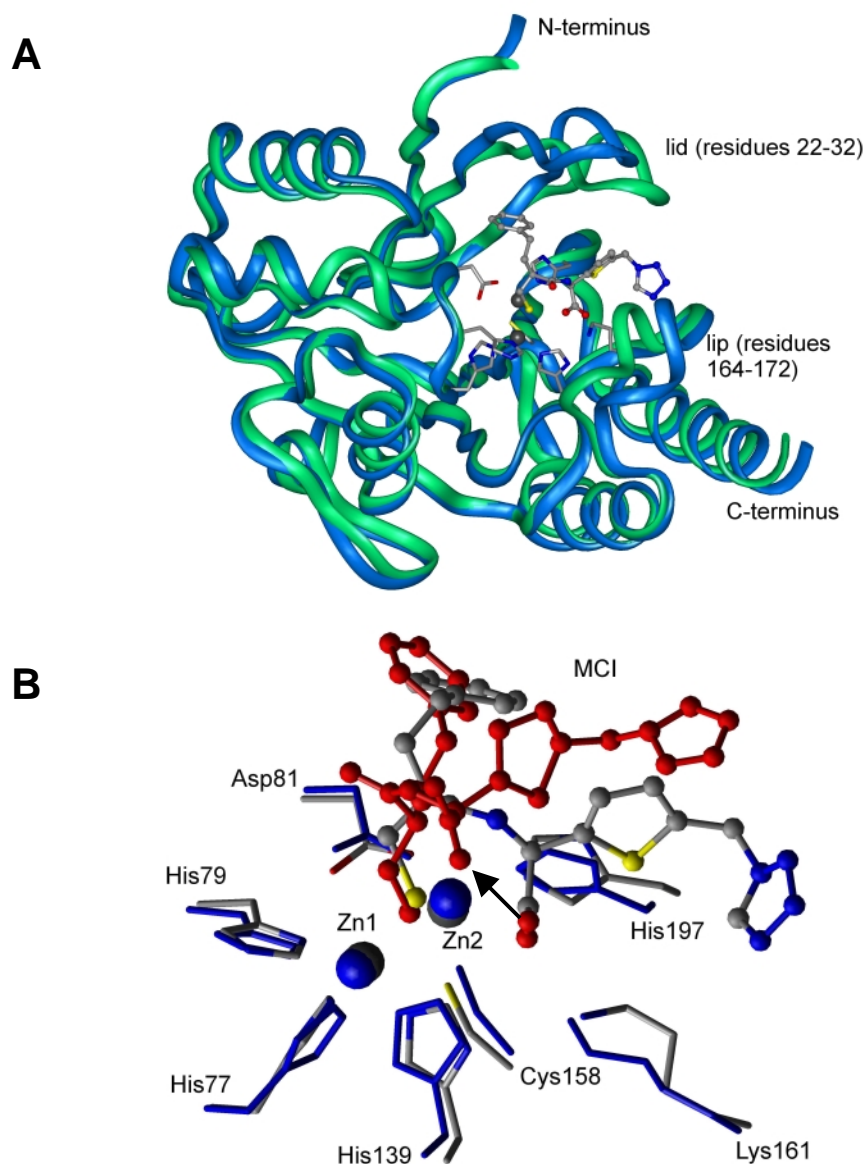
**Figure B 16A.** RMS values of the backbone of the whole protein over the course of the simulation with the cationic dummy atom method and deprotonated inhibitor.

**B.** RMS values of the backbone of residues 50-200.

**C.** Zinc-zinc distance during the simulation.

**D.** Distance between the oxygen of the carboxylate group of the inhibitor and the hydrogen of lysine 161 that was closest to the oxygen at the end of the simulation.

The RMS values of residues 50-200 were much lower than in the simulation with protonated inhibitor (1.0 Å compared to 1.5 Å). Again the values of the whole protein were clearly higher (1.3 Å). The active site was stable and the zinc-zinc distance remained at 3.8 Å. The distance between the inhibitor carboxylate and lysine 161 remained between 1.8 and 2.5 Å until it suddenly increased after 950 ps. An average structure made over the last 10 ps of the 1 ns simulation was overlaid with the crystal structure and is shown in Figure B 17.



**Figure B 17A.** The modeled structure using the cationic dummy atom approach and deprotonated inhibitor and the crystal structure overlaid via the backbone atoms. The backbone is displayed as ribbon. green: crystal structure with the active site colored by atom, blue: modeled structure.

**B.** The active sites of the crystal structure and the modeled structure after the cationic dummy atom simulation. Crystal structure: colored by atom, light grey: C, red: O, blue: N, yellow: S, dark grey: zincs, modeled structure: amino acids and zincs in blue, inhibitor in red. Amino acids are displayed as sticks and labeled at the C $\alpha$  atom, MCI: mercaptocarboxylate inhibitor as balls and sticks (carboxylate motion indicated by the arrow), zincs as CPK.

Apart from the lid and the lip, the modeled structure was in excellent accordance with the crystal structure. All the amino acid side chains in the active site and the phenyl ring of the inhibitor remained at their original positions. However, the (5-methylene-4-tetrazole)-2-thiophenyl residue of the inhibitor had moved. The inhibitor carboxylate had lost contact to lysine 161 and was oriented towards the side chain amide of asparagine 167 (2.3 Å distance). This motion occurred at the end of the simulation. To test whether this structure was stable, the simulation was extended for further 300 ps. The carboxylate remained in the new position and also interacted with the backbone of asparagine 167 and the side chain of tryptophane 28.

### B 3.1.4 Evaluation of the Different Models of IMP-1 in Complex with the Mercaptocarboxylate Inhibitor

The results obtained with the three models were compared to decide which one best represented the crystal structure. In the crystal structure and the average structures, distances between Zn1 and Zn2, between the ligating amino acids and the zinc atoms and between the inhibitor carboxylate/carboxyl group and lysine 161 were measured and compared. The results are summarized in Table B 3.1.

**Table B 3.1.** Distances in Å measured in the active site of the respective structures.

Distance	Structure			
	Crystal Structure 1DD6 Chain A	Purely Nonbonded Model	Cationic Dummy Atom Model, MCI protonated	Cationic Dummy Atom Model, MCI deprotonated
Zn1-Zn2	3.6	4.69	3.81	3.81
His77-Zn1	2.19	2.91	1.99	2.01
His79-Zn1	2.15	2.98	2.01	2.01
His139-Zn1	2.16	2.86	1.99	2.00
MCI-S-Zn1	2.24	2.95	2.11	2.11
Asp81-Zn2	2.09	2.70	1.90	1.92
Cys158-Zn2	2.33	2.89	2.08	2.07
His197-Zn2	2.26	2.96	2.00	2.02
MCI-S-Zn2	2.4	2.95	2.13	2.11
MCI-COO-Lys161-N	1.8	2.87	2.09	6.42

Accordingly the angles between ligands and the zinc atoms in the active site were measured and are summarized in Table B 3.2.

**Table B 3.2.** Angles measured in the active site of the crystal structure and the modeled structures.

Angle	Structure			
	Crystal Structure 1DD6 Chain A	Purely Nonbonded Model	Cationic Dummy Atom Model, MCI protonated	Cationic Dummy Atom Model, MCI deprotonated
His77-Zn1-His79	108°	67°	108°	110°
His77-Zn1-His139	108°	86°	110°	109°
His77-Zn1-MCI-S	125°	91°	109°	114°
His79-Zn1-His139	108°	122°	104°	107°
His79-Zn1-MCI-S	107°	92°	120°	116°
His139-Zn1-MCI-S	100°	141°	105°	100°
Asp81-Zn2-Cys158	113°	160°	114°	112°
Asp81-Zn2-His197	100°	72°	103°	110°
Asp81-Zn2-MCI-S	105°	106°	114°	105°
Cys158-Zn2-His197	107°	124°	114°	107°
Cys158-Zn2-MCI-S	119°	75°	105°	105°
His197-Zn2-MCI-S	112°	125°	119°	118°

Whereas for both cationic dummy atom models the distances and angles approximated closely to those in the crystal structure, they deviated significantly in the purely nonbonded model. The distances in the cationic dummy atom model were slightly decreased compared to the crystal structure. The angles were close to the ideal tetrahedral angle of 109.5° and the original tetrahedral coordination of the zinc(II) atoms could be maintained throughout both simulations. For the purely nonbonded approach this was not the case. The angle deformation was accompanied by the entrance of additional amino acid side chains into the first coordination sphere of the zincs, resulting in higher coordination numbers. The zinc coordination numbers for the crystal structure and the modeled structures are summarized in Table B 3.3.

**Table B 3.3.** Coordination numbers of Zn1 and Zn2 in the crystal structure and the modelled structures.

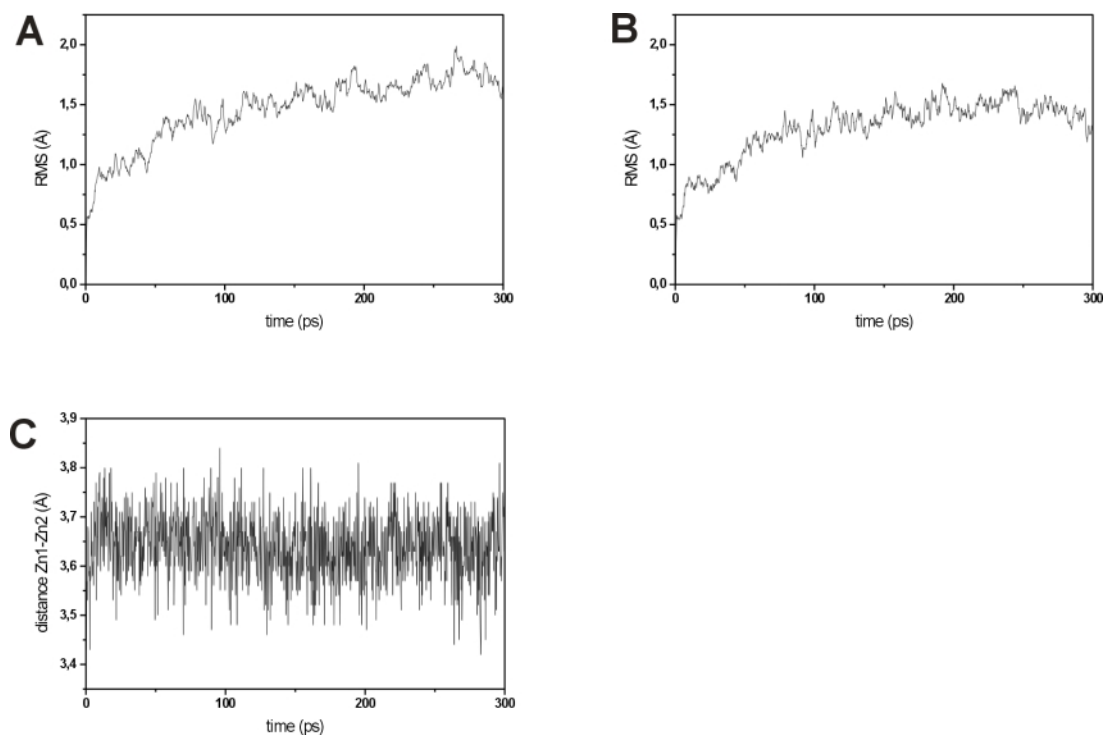
Coordinated Atom	Structure			
	Crystal Structure 1DD6 Chain A	Purely Nonbonded Model	Cationic Dummy Atom Model, MCI protonated	Cationic Dummy Atom Model, MCI deprotonated
Zn1	4	5	4	4
Zn2	4	7	4	4

The coordination number for Zn1 had increased to 5, for Zn2 even to 7 in the nonbonded model.

All values measured in the cationic dummy atom model with the protonated inhibitor closely approximated those in the crystal structure and described the active site residues best.

### B 3.2 Application of the Model to the Free IMP-1 Enzyme

To test if the cationic dummy atom model could also be applied to the modeling of the IMP-1 metallo- $\beta$ -lactamase as a free enzyme, the inhibitor was deleted and a hydroxide was introduced as a bridging ligand between the two zinc(II) atoms. A 300 ps simulation was performed at 300 K. The RMS values and the zinc-zinc distances monitored over the course of the simulation are shown in Figure B 18.

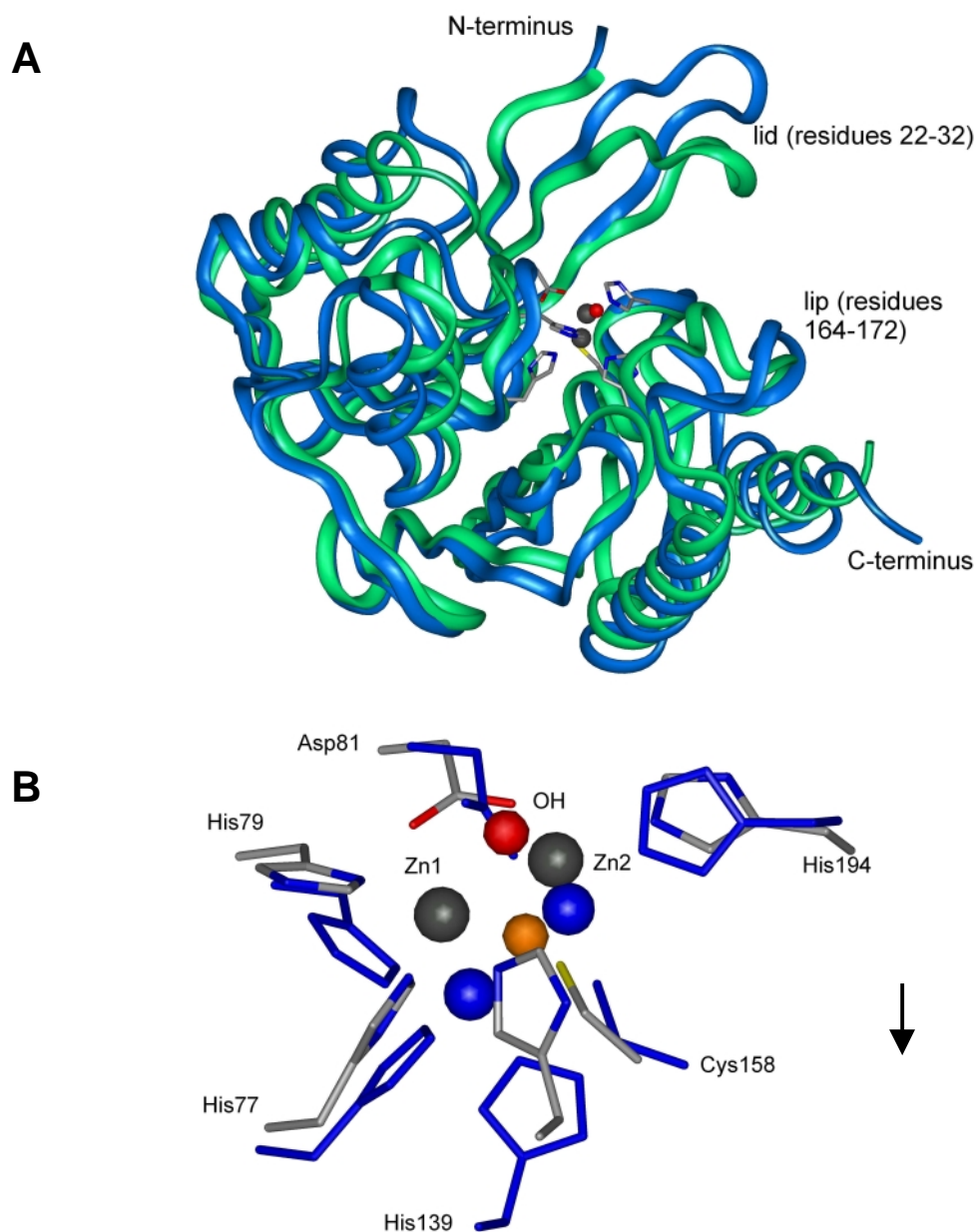


**Figure B 18A.** RMS values of the backbone of the free IMP-1 enzyme over the course of a 300 ps simulation.

**B.** RMS values of the backbone atoms of residues 50-200.

**C.** Zinc-zinc distance over the MD simulation.

Like in the simulations of IMP-1 in complex with the inhibitor, the RMS values of the total protein were higher than in the region from residue 50-200 (1.7 Å compared to 1.4 Å). The active site was stable over the 300 ps and the zinc-zinc distance was 3.6 Å, slightly smaller than for the metallo- $\beta$ -lactamase in complex with the inhibitor. An average structure was made at the end of the simulation and overlaid with the starting structure, which corresponded to the crystal structure with the exception, that the inhibitor was replaced by a hydroxide with the hydroxide oxygen taking on the position of the inhibitor sulfur (Figure B 19).



**Figure B 19A.** The modeled structure of the free IMP-1 enzyme and the crystal structure overlaid via the backbone atoms. The backbones are displayed as ribbons. green: crystal structure, blue: modeled structure. The active site residues of the starting structure are displayed.

**B.** The active sites of the crystal structure and the modeled structure after the simulation. Crystal structure: colored by atom. Light grey: C, red: O, blue: N, yellow: S, dark grey: zincs. Modeled structure: amino acids and zincs in blue, hydroxide oxygen in orange. Amino acids are displayed as sticks and labeled at the C $\alpha$  atom, OH: hydroxide, zincs as CPK. The arrow indicates the movement of the active site residues.

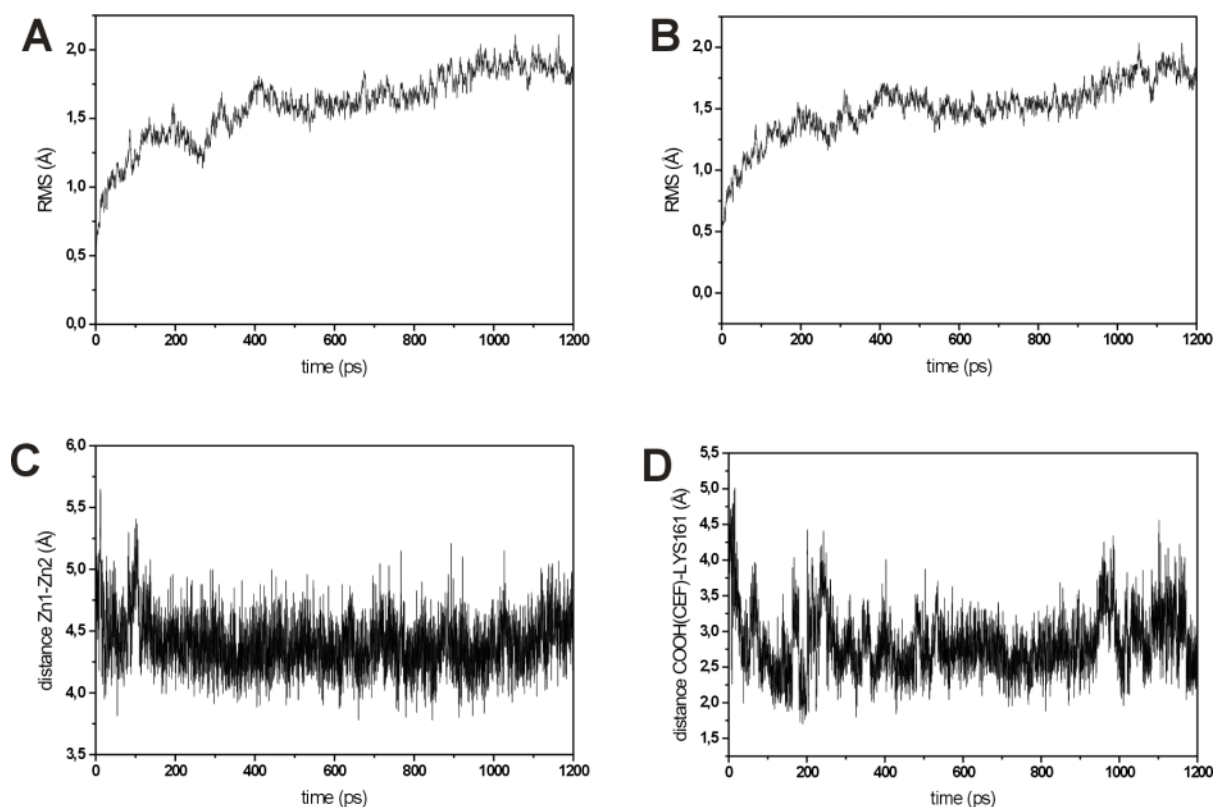
Some conformational changes were observed in the modeled free enzyme compared to the starting structure. These were observed mainly in the lid, the lip and the termini, but also in other regions all over the protein. The active site had moved compared to the overlaid backbones, but a stable tetrahedral coordination of the zincs was kept as for the simulations



with inhibitor. The ligand-zinc distances were all between 1.9 and 2.1 Å and the ligand-zinc-ligand angles were between 100 and 125°. The zinc-zinc distance did not increase as with the inhibitor but remained at 3.6 Å.

### **B 3.3 Application of the Model to IMP-1 in Complex with Cephalothin**

It was investigated if a  $\beta$ -lactam antibiotic could be simulated in an intermediate structure. Therefore, cephalothin was selected. It is a cephalosporin and is known to be converted well by the IMP-1 metallo- $\beta$ -lactamase [42] and the intermediate was assumed to be quite stable. Its chemical structure is shown in the Materials and Methods part (B 2.2.1.1). It was docked into the active site of the enzyme as an intermediate as described in Materials and Methods. This intermediate structure corresponds to a mechanism that has been suggested previously [13] based on kinetic investigations of the very similar *Bacteroides fragilis* metallo- $\beta$ -lactamase. In contrast to that report the carboxyl group resulting from hydrolysis of the amide bond and located close to Zn1 was deprotonated and the carboxyl group on the six ring of cephalothin was protonated. The simulations with IMP-1 in complex with the inhibitor showed that protonation of the carboxylate did not interfere with the interaction between the  $\beta$ -lactam and lysine 161, which was assumed to be crucial for efficient hydrolysis. A 1.2 ns MD simulation was performed of this structure and the RMS values, the zinc-zinc distances and the distances between the cephalothin carboxyl group and lysine 161 were monitored during the simulation (Figure B 20).



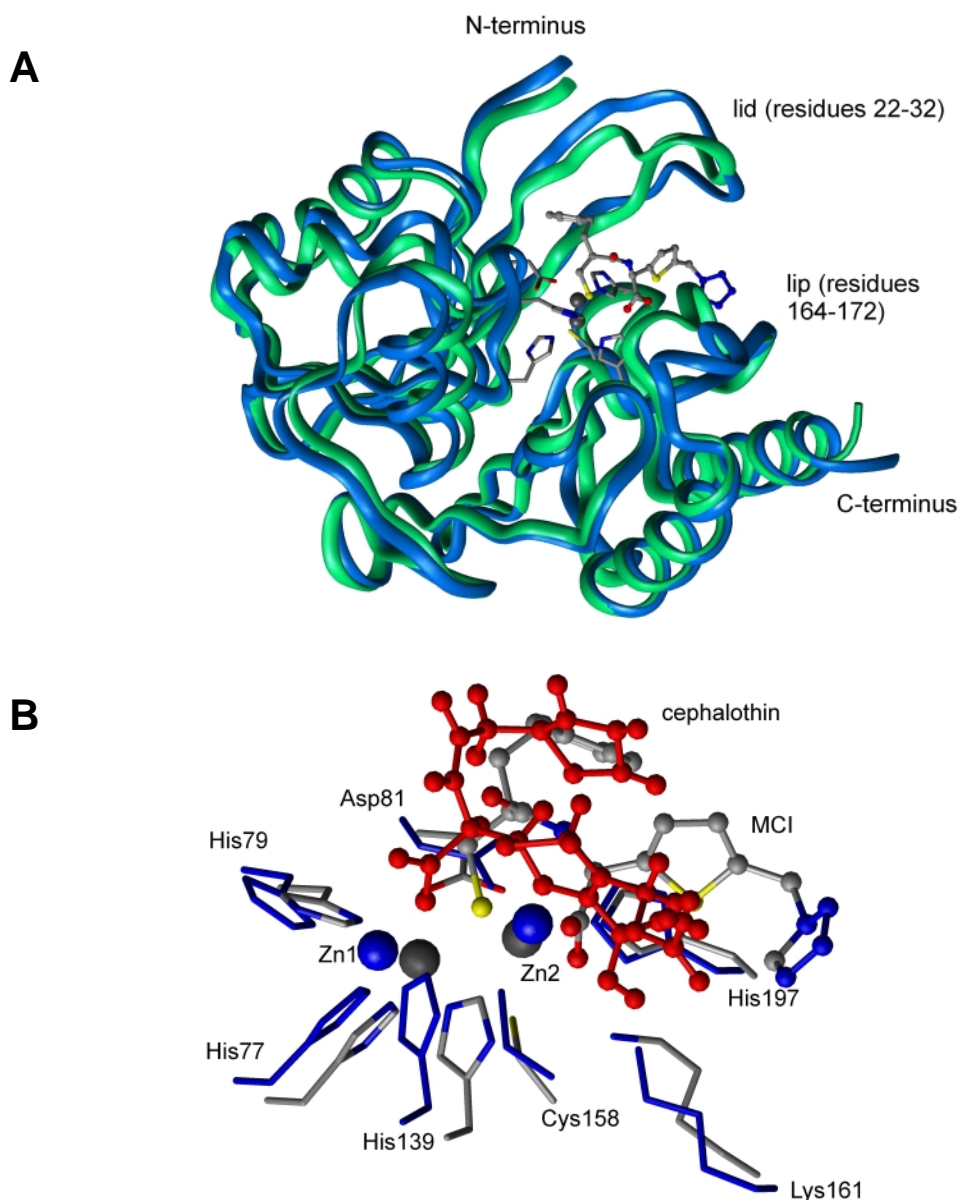
**Figure B 20A.** RMS values of the backbone of the IMP-1 enzyme in complex with the cephalothin intermediate over the course of the 1.2 ns simulation.

**B.** RMS values of the backbone of residues 50-200.

**C.** Zinc-zinc distance over the MD simulation.

**D.** Distance between the protonated carboxyl group of cephalothin and lysine 161.

In this case the RMS values of residues 50-200 approximately corresponded to the RMS values of the whole protein, indicating that they were to a great extent due to motion in the region from residues 50-200. Both the zinc-zinc distance and the distance between the carboxyl group of cephalothin and lysine 161 decreased in the first 200 ps and then remained at values of about 4.4 and 3 Å. An average structure of the last 10 ps was made and overlaid with the crystal structure of IMP-1 in complex with the mercaptocarboxylate inhibitor. The binding mode of the substrate was completely different from that of the inhibitor. There was no more bridging ligand between the two zinc(II) atoms, but each one had a separate fourth ligand: Zn1 the oxygen from the carboxylate resulting from the amide cleavage, Zn2 the anionic nitrogen resulting from the amide cleavage. Additionally, the side chains of cephalothin differed significantly from those of the mercaptocarboxylate inhibitor.



**Figure B 21A.** The modeled structure of the IMP-1 enzyme in complex with the cephalothin intermediate and the crystal structure overlaid via the backbone atoms. The backbones are displayed as ribbons. Green: crystal structure, blue: modeled structure. The active site residues of the crystal structure are displayed.

**B.** The active sites of the crystal structure and the modeled structure after the simulation. Crystal structure: colored by atom. Light grey: C, red: O, blue: N, yellow: S, dark grey: zincs, MCI: inhibitor in balls and sticks. Modeled structure: amino acids and zincs in blue, cephalothin in red as balls and sticks. Amino acids are displayed as sticks and labeled at the C $\alpha$  atom, zincs as CPK.

Taking all the differences mentioned above into account, the two structures are surprisingly similar. The ligand-zinc distances in the modeled structure of IMP-1 in complex with the cephalothin intermediate were between 1.9 and 2.2 Å and the ligand-zinc-ligand angles were all between 100 and 115° indicating a good tetrahedral coordination. The most striking

difference observed was an increased distance between the two zincs (4.6 Å), probably due to the missing bridging ligand. This was mainly the result of a movement of Zn1, whereas Zn2 as well as its amino acid ligands stayed in place. The carboxyl group of the cephalothin intermediate stayed oriented towards lysine 161 throughout the simulation although it was protonated (also see Figure B 20D). Asparagine 167, a residue that has been assumed to be important for stabilizing the substrate in the active site [13] did not interact with the substrate. In the average structure at the end of the 1.2 ns simulation, the carboxylate resulting from the amide bond hydrolysis of the substrate, which is assumed to interact with the asparagine, and the amino acid were at a distance of more than 8 Å.

### B 3.4 Development of an Assay for Substrate Specificity of Metallo- $\beta$ -lactamases

The IMP-1 enzyme and variants in complex with other  $\beta$ -lactam antibiotics were investigated assuming that parameters could be found that correlate with experimental data. To develop such an assay a number of enzyme-substrate combinations were considered, for which  $k_{cat}/K_M$  values had been determined experimentally. It was important to take data that had been collected under the same conditions by the same group. This was the case for a study by Iyobe *et al.* [42]. Of these data, mutants and substrates were selected in a way that big differences existed between different mutants as well as between different substrates. Additionally, the substrates should have a similar chemical structure so that they could be docked accordingly and variable docking could be excluded to influence the simulation results. The two metallo- $\beta$ -lactamase variants IMP-1 and IMP-6 (a S196G mutant of IMP-1) and the four cephalosporins cephalothin (CEF), cephaloridine (LOR), cefotaxime (CTX) and ceftazidime (CAZ) met these requirements (the structures of these substrates are shown in Materials and Methods, Figure B 8; also see Figure B 2 in the introduction). The  $k_{cat}/K_M$  values of the respective combinations are summarized in Table B 3.4.

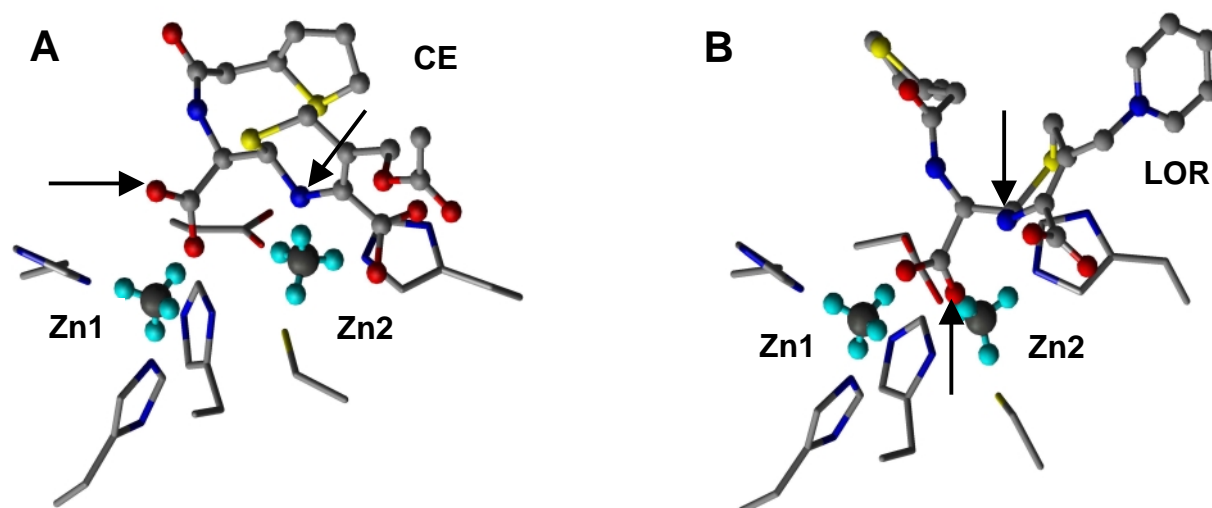
**Table B 3.4.**  $k_{cat}/K_M$  values for different metallo- $\beta$ -lactamase/cephalosporin combinations.

Enzyme	$k_{cat}/K_M$ ( $\mu\text{M}^{-1} \text{s}^{-1}$ )			
	CEF	CTX	LOR	CAZ
IMP-1	32.5	16.1	8.6	0.35
IMP-6	39.2	18.3	1.25	0.031

Two substrates, CEF and CTX are converted well by both variants; LOR and CAZ are less efficiently hydrolyzed. The former have a acetoxy group as residue 2, whereas the latter have a positively charged pyridinium ring in the respective position. In CTX and CAZ residue 1 is more bulky than in CEF and LOR. For LOR and CAZ different  $k_{cat}/K_M$  values had been measured with the two enzymes. The  $k_{cat}/K_M$  values of all combinations ranged from 0.031 to 39.2  $\mu\text{M}^{-1} \text{s}^{-1}$  (a factor of more than 1,000).

### B 3.4.1 MD simulations of IMP-1 and IMP-6 with Cephalothin and Cephaloridine at 300 K

The IMP-1 metallo- $\beta$ -lactamase in complex with cephalothin had been modeled successfully at 300 K as shown in B 3.3. To investigate if this was also possible for the IMP-6 enzyme and enzymes in complex with substrates resulting in lower  $k_{cat}/K_M$  values, cephaloridine was docked into IMP-1 accordingly and the respective IMP-6 mutants were generated. All four structures were heated to 100 K and simulated without constraints for 100 ps for an initial equilibration of the system. Then the systems were heated to 300 K and simulated without constraints at 300 K. For IMP-1 in complex with cephalothin and cephaloridine as well as IMP-6 in complex with cephalothin the intermediate structure of the substrate remained stable at 300 K, whereas for IMP-6 in complex with cephaloridine the anionic nitrogen of the substrate lost contact to Zn2. In Figure B 22 the modeled structures of IMP-1/CEF and IMP-6/LOR are compared.



**Figure B 22A.** Active site of IMP-1 in complex with the cephalothin intermediate after the 1.2 ns simulation.

**B.** Active site of IMP-6 in complex with the cephaloridine intermediate after 100 ps.

Atoms are colored by atom: grey: C, red: O, blue: N, yellow: S, dark grey: zincs, cyan: dummy atoms. The amino acids are displayed as sticks, the substrates and the zincs as balls and sticks. The zincs are displayed with their dummy atoms to clarify the coordination. For the names and sequence numbers of the residues see Figure B 10B which shows the same perspective.

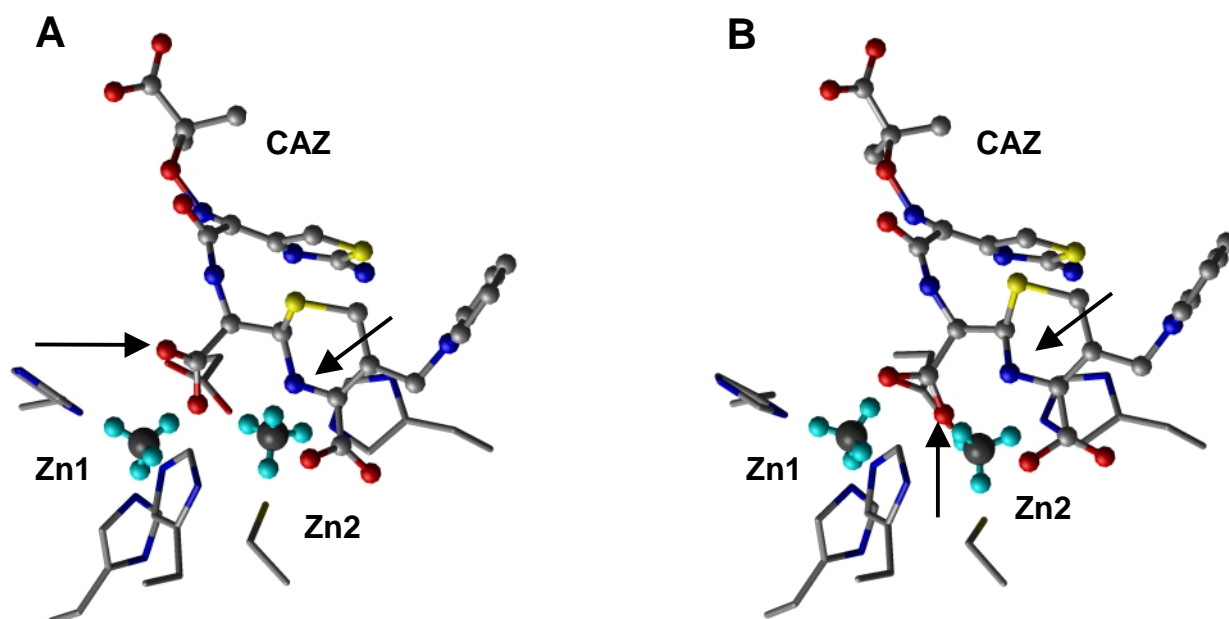
The arrows indicate the second oxygen of the carboxylate group that remains free in IMP-1/CEF and replaces the anionic nitrogen in IMP-6/LOR as well as the anionic nitrogen.

In the structure of IMP-1 in complex with cephalothin (Figure B 22A) one carboxylate oxygen was coordinated to one dummy atom of the Zn 1 and the anionic nitrogen of the  $\beta$ -

lactam intermediate was coordinated to a dummy atom of Zn2. The structures of IMP-1/LOR and IMP-6/CEF corresponded to IMP-1/CEF and the nitrogen remained coordinated to Zn2. In the structure of IMP-6 with cephaloridine (Figure B 22B), however, the anionic nitrogen originally coordinated to the zinc-dummy molecule 2 was displaced by the free oxygen of the carboxylate resulting from amide bond cleavage. This observation is in agreement with experimental data: enzymes that showed high  $k_{cat}/K_M$  values for the conversion of the  $\beta$ -lactam (IMP-1/CEF, IMP-6/CEF and IMP-1/LOR) could be modeled successfully with the cephalosporin docked as an intermediate, whereas in the one that showed a low  $k_{cat}/K_M$  value (IMP-6/LOR) the intermediate structure broke down. However, the intention was to model all combinations in the intermediate structure and combinations with CAZ had even lower  $k_{cat}/K_M$  values than IMP-6/LOR. Assuming that the intermediate structures for all combinations might be stable at lower temperatures, they were simulated at 100 K.

### **B 3.4.2 MD Simulations of IMP-1 and IMP-6 with Cephalothin, Cephaloridine, Cefotaxime and Ceftazidime at 100 K**

Additionally to the two cephalosporins investigated above, cefotaxime and ceftazidime that were less efficiently hydrolyzed in experiment were included resulting in eight enzyme/substrate combinations as shown in Table B 3.4. Simulations were carried out at 100 K but besides that under the same conditions as the 300 K simulations. After 200 ps the structures were averaged over 10 ps and compared. In seven structures the intermediate structure was still intact. For IMP-6/CAZ (the combination with the lowest  $k_{cat}/K_M$  value in experiment) the anionic nitrogen was replaced by the carboxylate oxygen after 30 ps. An average structure at 20 ps before the replacement and one at 40 ps after the replacement are shown in Figure B 23.



**Figure B 23A.** Active site of IMP-6 in complex with the ceftazidime intermediate after 20 ps of simulation at 100 K.

**B.** Active site of the structure after 40 ps.

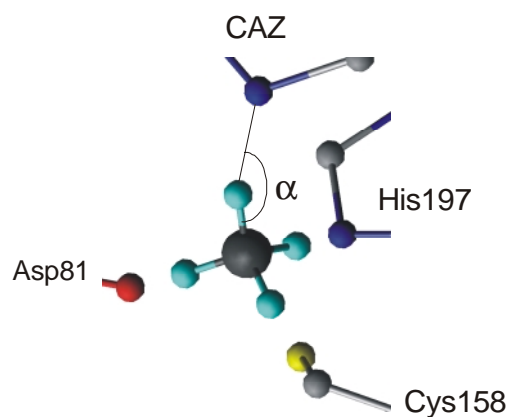
Atoms are colored by atom: grey: C, red: O, blue: N, yellow: S, dark grey: zincs, cyan: dummy atoms. The amino acids are displayed as sticks, the substrates and the zincs as balls and sticks. For the names and sequence numbers of the residues see Figure B 10B.

The arrows indicate the second oxygen of the carboxylate group that was free before the replacement and bound to Zn2 after the replacement as well as the anionic nitrogen.

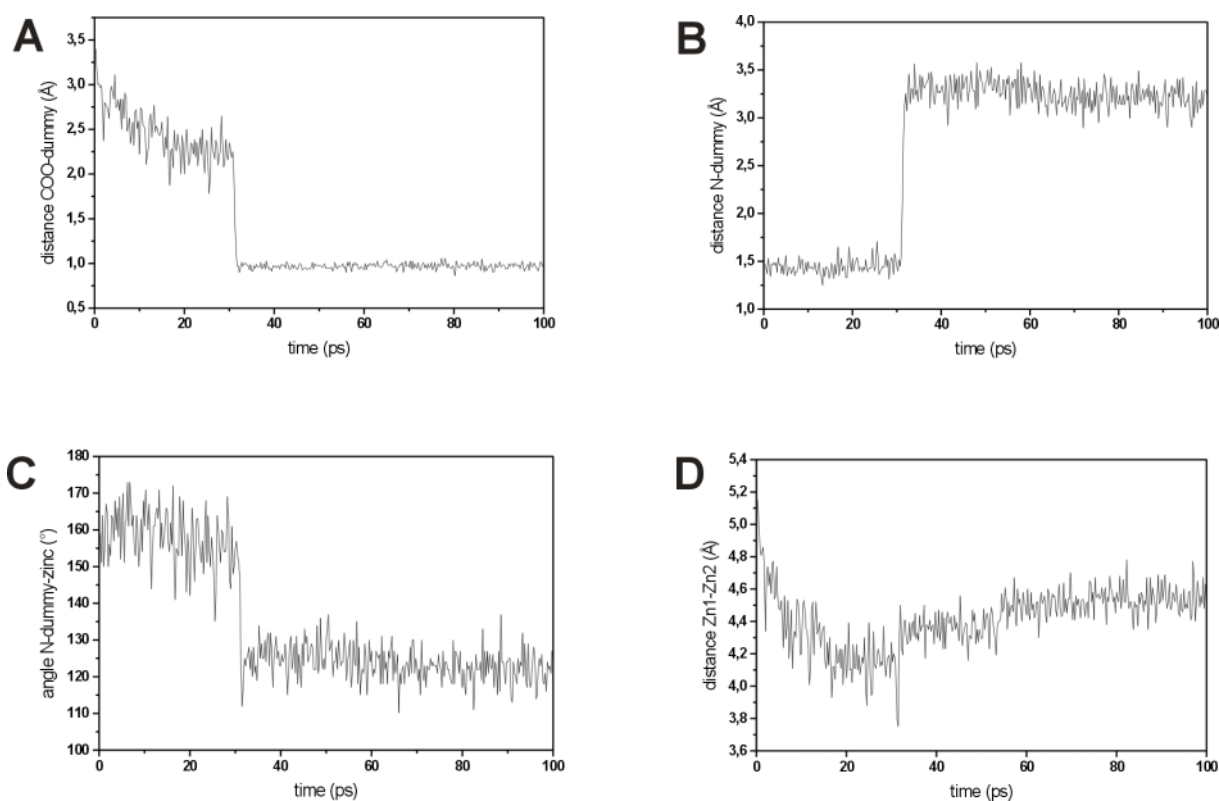
IMP-6/CAZ had a  $k_{cat}/K_M$  value of  $0.031 \mu\text{M}^{-1} \text{s}^{-1}$  in experiment. At 100 K an according replacement had happened as for IMP-6/LOR at 300 K (experiment:  $k_{cat}/K_M = 1.25 \mu\text{M}^{-1} \text{s}^{-1}$ ). However, IMP-6/LOR was stable at 100 K. Lowering the temperature to 100 K helped to make the model less sensitive and allowed to model more combinations with relatively low  $k_{cat}/K_M$  values in the intermediate structure. The only exception was IMP-6/CAZ with an extremely low  $k_{cat}/K_M$  value.

In order to investigate in more detail, what happened before and during the replacement of the nitrogen by the oxygen, the distances between the anionic nitrogen of the substrate and the dummy atom, as well as that between the carboxylate oxygen of the substrate and the dummy atom were measured over the course of the first 100 ps of the simulation. Additionally, a deformation of the angle anionic nitrogen-dummy-Zn2 and a change of the Zn1-Zn2 distance were observed in the trajectory. Figure B 24 illustrates the angle in which the deformation was observed. These values were also monitored. The results of the measurements are shown in Figure B 25.





**Figure B 24.** Scheme of Zn2 with its dummy atoms, the surrounding ligands and the angle measured.



**Figure B 25A.** The distance between the oxygen of the carboxylate group of the  $\beta$ -lactam resulting from amide bond cleavage to the dummy atom of Zn2.

**B.** Distance between the nitrogen resulting from amide bond cleavage of the substrate and the dummy atom of Zn2.

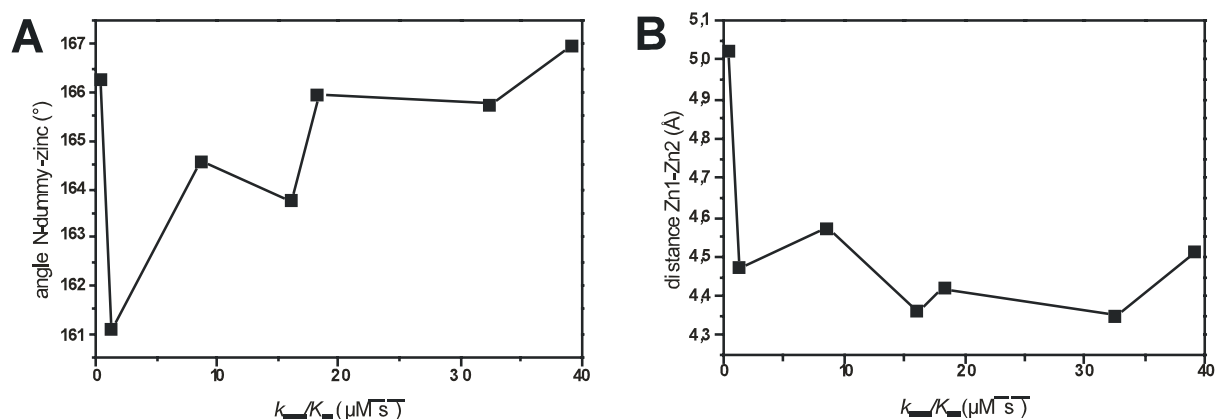
**C.** Angle between the anionic nitrogen, the dummy atom and Zn2.

**D.** Distance between the two zincs.

From Figures B 25A and B it is obvious that the exit of the anionic nitrogen from and the entrance of the carboxylate oxygen into the zinc coordination position took place simultaneously. The carboxylate oxygen gradually approximated the dummy in the first 30 ps

(distance decreased from 3.0 to 2.2 Å). Then it replaced the nitrogen. The oxygen was then at a distance of 1.0 Å to the dummy atom, whereas the distance nitrogen-dummy atom before was 1.5, indicating that the oxygen was bound more tightly. The measured angle decreased until the replacement and then dropped to a constant 120° (Figure B 25C). The zinc-zinc distance decreased to 4.2 Å, lower than the equilibrated distance in IMP-1/CEF (Figure B 18C) and after the replacement increased to 4.5 Å (Figure B 25D).

The measured angle deviated from the value that would have been expected for a stable coordination. If the ligand would be perfectly tetrahedrally coordinated, it would be in one line with the dummy atom and therefore the angle would be 180°. Here, however, it decreased to about 150°, before the nitrogen lost contact to the dummy. It seemed that such an angle was too small to keep the ligand in place. Also the decrease in the zinc-zinc distance seemed to be related to the conformational changes in some way. These two parameters were investigated for the 7 enzyme/substrate combinations in which the intermediate structure was stable. For each simulation, after 200 ps this angle and the zinc-zinc distance were measured and averaged over 10 ps. The results are summarized in Figure B 26.



**Figure B 26A.** Angles measured between the anionic nitrogen, the dummy atom and Zn2 for the seven enzyme/substrate combinations after 200 ps of simulation plotted against the experimentally determined  $k_{cat}/K_M$  values.

**B.** Distances between Zn1 and Zn2 in the respective structures plotted against the  $k_{cat}/K_M$  values.

The data points are linked in order to make the relatedness clear.

The measured angles were roughly correlated with the  $k_{cat}/K_M$  values except the value for IMP-1/CAZ, which was very high. Interestingly, this combination also showed an outlying zinc-zinc distance, whereas it was almost equal for the other combinations (4.3-4.6 Å). Also IMP-1/LOR resulted in a relatively high angle and had a relatively high zinc-zinc distance.

Thus, the N-dummy-zinc angle and the zinc-zinc distances seemed to be related and to be important parameters. When the zinc-zinc distance was higher, a higher angle was possible.

Large molecular systems such as proteins are very complex and it depends on many factors in which conformation the system ends after an MD simulation. Therefore only one simulation of each structure may not be representative. To take this into account, further simulations were performed of the respective structures with different starting conditions. These were varied by randomly assigning starting velocities to the atoms of the system at the beginning of the heating periods. Additionally, some of the 200 ps simulations were extended to 1 ns or more. All simulations are summarized in Table B 3.5. For the long simulations, two measurements were made, one at 200 ps and one at the end. In one simulation of IMP-1/CEF a shift was observed after 550 ps (Figure B 27). Therefore another measurement was made before this shift at 400 ps.

### **B 3.4.3 Analysis of the 100 K MD Simulations**

#### **B 3.4.3.1 Contact Between the Anionic Nitrogen and the Dummy Atom of Zn2**

In some simulations, the anionic substrate nitrogen had lost contact to the dummy atom of Zn2 and the intermediate structure had broken down. This was the case for the IMP-6/CAZ simulation mentioned above and a second simulation with the same combination. In one of three measurements of the IMP-1/CAZ combination this also happened, which is not surprising because IMP-1/CAZ also is an inefficient combination for hydrolysis ( $k_{cat}/K_M$  value of  $0.35 \mu\text{M}^{-1} \text{s}^{-1}$ ). Also in one measurement of IMP-1/LOR the nitrogen lost contact. Experimentally a  $k_{cat}/K_M$  value of  $8.6 \mu\text{M}^{-1} \text{s}^{-1}$  was measured for this combination which is significantly lower than that of the good substrates CEF and CTX. All simulations, in which the intermediate had broken down, were not considered in the further analysis.

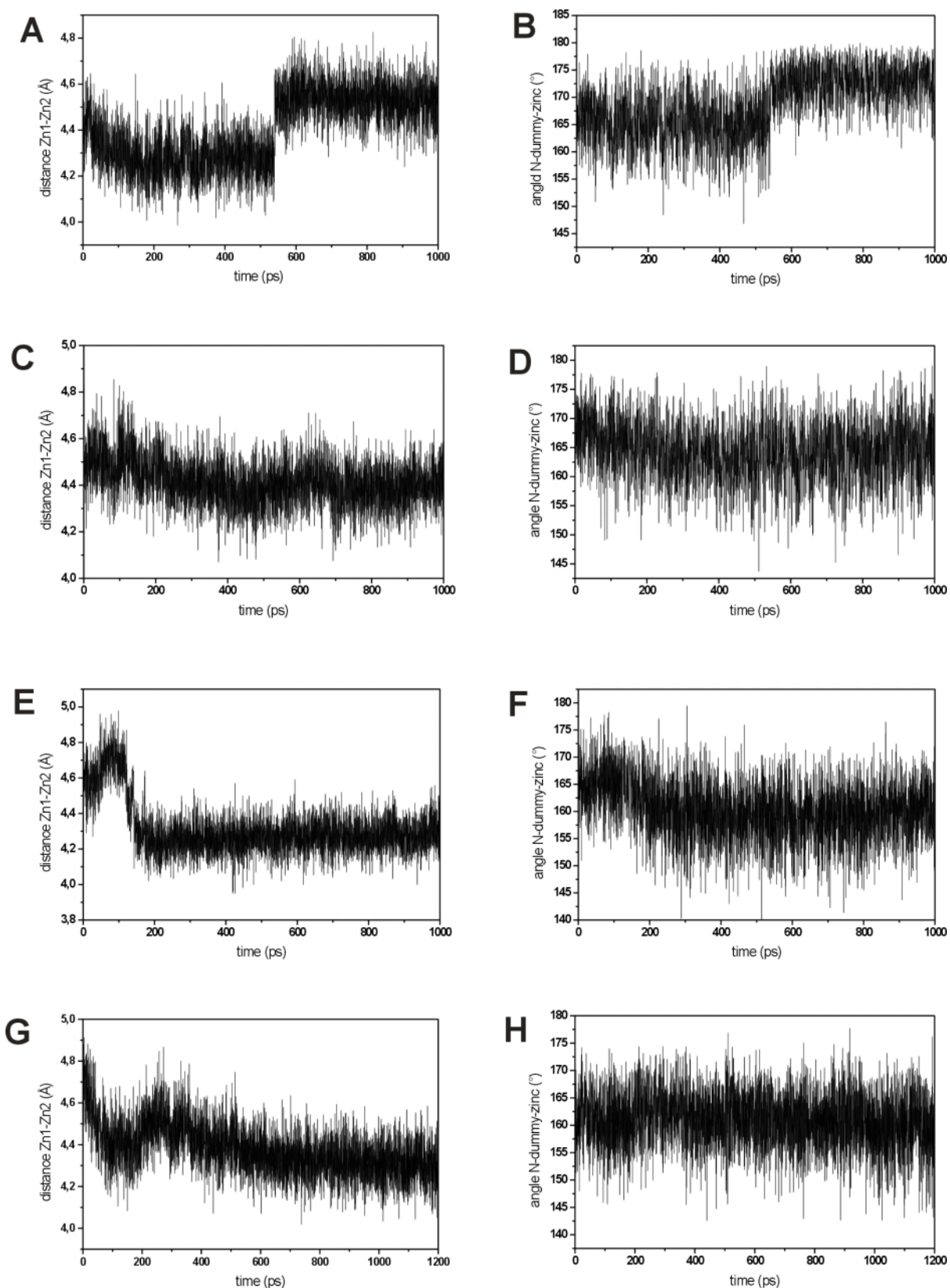
#### **B 3.4.3.2 Zn1-Zn2 Distance and Angle Between the Anionic Nitrogen, the Dummy Atom and Zn2**

As mentioned above, the Zn1-Zn2 distance and the N-dummy-Zn2 angle were related and were therefore analyzed together. For the long simulations of IMP-1/CEF, IMP-6/CEF, IMP-1/LOR and IMP-6/LOR both parameters were monitored over the course of the simulations and are shown in Figure B 27.

**Table B 3.5.** Summary of the performed simulations.

<b>Enzyme/Substrate</b>	<b>Simulation Time (ps)</b>	<b>N still bound to Zn2 ?</b>	<b>Zinc-zinc Distance (Å)</b>	<b>N-dummy-zinc Angle (°)</b>
IMP-1/CEF	200	yes	4.35	165.8
IMP-1/CEF*	400	yes	4.30	163.2
IMP-1/CEF*	1000	yes	4.51	171.4
IMP-1/CEF	250	yes	4.55	161.2
IMP-1/CEF	200	yes	5.02	174.2
IMP-6/CEF	200	yes	4.51	167.0
IMP-6/CEF*	1000	yes	4.41	165.6
IMP-6/CEF	200	yes	4.47	164.1
IMP-6/CEF	200	yes	5.84	172.3
IMP-1/LOR	200	yes	4.57	164.6
IMP-1/LOR*	1000	yes	4.30	158.8
IMP-1/LOR	200	no	4.56	147.5
IMP-1/LOR	470	yes	4.17	157.1
IMP-6/LOR	200	yes	4.47	161.1
IMP-6/LOR*	1200	yes	4.28	159.0
IMP-6/LOR	260	yes	4.78	169.4
IMP-6/LOR	220	yes	4.11	158.3
IMP-1/CTX	200	yes	4.36	163.7
IMP-1/CTX*	870	yes	4.38	163.1
IMP-1/CTX	200	yes	4.96	172.0
IMP-1/CTX	200	yes	4.13	161.5
IMP-6/CTX	200	yes	4.42	165.9
IMP-6/CTX	200	yes	5.12	167.7
IMP-6/CTX	200	yes	4.32	169.8
IMP-1/CAZ	200	yes	5.02	166.3
IMP-1/CAZ*	1100	yes	5.10	168.51
IMP-1/CAZ	260	yes	4.42	161.4
IMP-1/CAZ	200	no	4.50	115.8
IMP-6/CAZ	245	no	4.35	120.4
IMP-6/CAZ	240	no	4.84	116.0

\* Simulation continued from the previous simulation listed in the table.

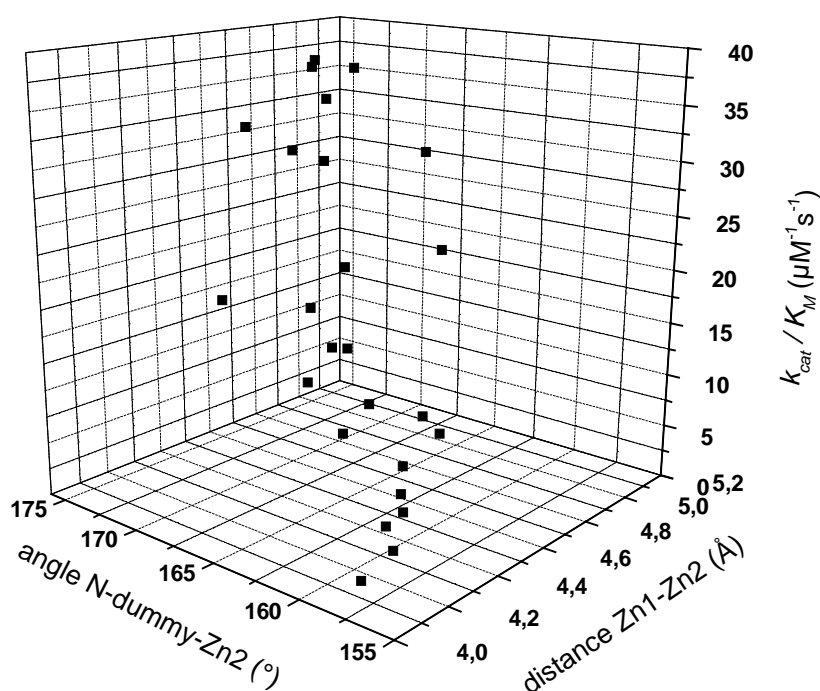


**Figure B 27.** Zinc-zinc distances (A, C, E, G) and N-dummy-zinc angles (B, D, F, H) of four simulations. A, B. IMP-1/CEF. C, D. IMP-6/CEF. E, F. IMP-1/LOR. G, H. IMP-6/LOR. The parameters were measured after 200 ps and at the end of the simulation, for IMP-1/CEF additionally after 400 ps of simulation.

From Figure B 27 A and B (IMP-1/CEF) it is obvious that the zinc-zinc distance and the measured angle are coupled. After 550 ps both values jump to a higher value. Also for the other combinations the two parameters are in concordance. This can be seen very clearly in Figure B 27 E and F (IMP-1/LOR): After an initial increase of both parameters, they decrease concordantly later on. This observation further supports that the two parameters depend on each other and have to be considered both.

For all simulations the Zn1-Zn2 distance was between 4.1 and 5.1 Å except one simulation of IMP-6/CEF. Here the distance was 5.84. Quantum-mechanical calculations of the *Bacteroides fragilis* metallo- $\beta$ -lactamase without bridging ligand between the two zincs resulted in a zinc-zinc distance of 4.5-5 Å [65]. Cephalosporin intermediates might keep the zincs closer together than 4.5 Å, but distances significantly higher than 5 Å have neither been observed experimentally nor with quantum-mechanical calculations. Therefore this value was considered an outlier.

All the other data were plotted in a 3D chart with the experimentally determined  $k_{cat}/K_M$  value, the Zn1-Zn2 distance and the N-dummy-Zn2 angle indicated on one axis, each. The chart is shown in Figure B 28.

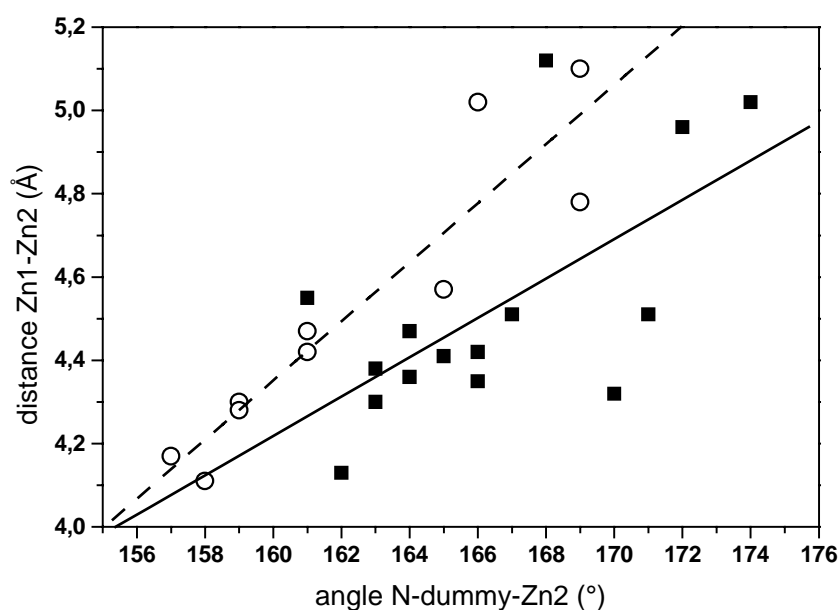


**Figure B 28.** 3D representation of the data from the 100 K simulations.

All data points were approximately found in one plane, although with a considerable variance. To check if there were significant differences between good and bad enzyme/substrate combinations (high-efficient and low-efficient hydrolysis), the data were put into two groups:

1. Those with  $k_{cat}/K_M$  values higher than  $10 \mu\text{M}^{-1}\text{s}^{-1}$  (IMP-1/CEF, IMP-6/CEF, IMP-1/CTX, IMP-6/CTX) and
2. those with  $k_{cat}/K_M$  values smaller than  $10 \mu\text{M}^{-1}\text{s}^{-1}$  (IMP-1/LOR, IMP-6/LOR, IMP-1/CAZ).

The two data sets were plotted in a 2D chart with the the Zn1-Zn2 distance and the angle on one axis, respectively. Then a linear regression was calculated. The results are shown in Figure B 29.



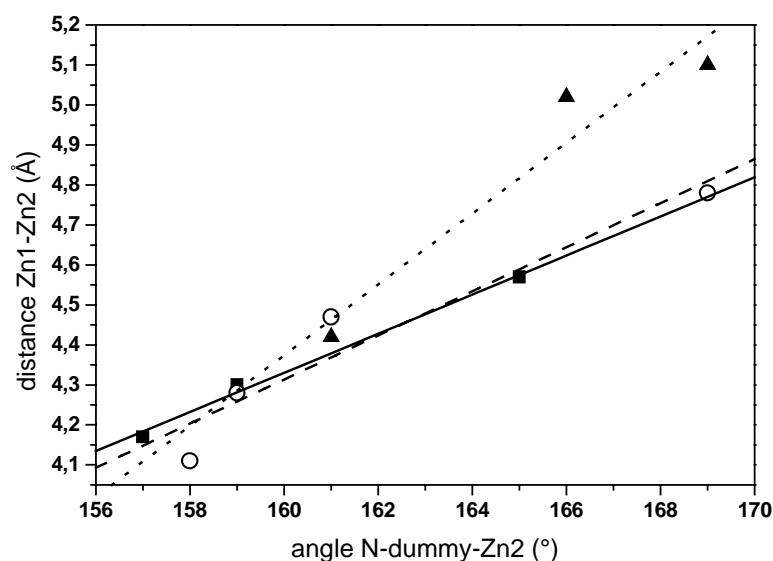
**Figure B 29.** 2D representation of the data points for the different enzyme/substrate combinations separated by high (■ and solid regression line) and low (○ and dashed regression line) hydrolytic efficiency.

The linear regression equations were

$$y = -3.3 \pm 2.6 + (4.7 \pm 1.6) x \quad \text{for the high-efficiency combinations and}$$

$$y = -7.0 \pm 1.6 + (7.1 \pm 1.0) x \quad \text{for low efficiency combinations.}$$

A difference in the two groups is visible. The linear regression straight line for less efficient combinations is steeper, but has a smaller y axis intersection. To see if a further separation could be observed, the low-efficiency enzyme-substrate combinations were separated for the three combinations IMP-1/LOR, IMP-6/LOR and IMP-1/CAZ. The data were again plotted as in Figure B 29 and linear regression equations were calculated (Figure B 30).



**Figure B 30.** 2D representation of the data points for the different enzyme/substrate combinations: IMP-1/LOR (■ and solid regression line), IMP-6/LOR (○ and dashed regression line), IMP-1/CAZ (▲ and dotted regression line).

The linear regression equations were

$$y = -3.5 \pm 6.4 + (4.9 \pm 4.0) x \quad \text{for IMP-1/LOR}$$

$$y = -4.5 \pm 1.9 + (5.5 \pm 1.2) x \quad \text{for IMP-6/LOR}$$

$$y = -9.8 \pm 4.1 + (8.9 \pm 2.5) x \quad \text{for IMP-1/CAZ.}$$

The same tendency as above was observed: combinations that hydrolyse  $\beta$ -lactams less efficiently result in a steeper linear regression straight line and in a lower y axis intersection. With respect to the plane on which the data points are located in a 3D chart, this means that it has a torsion.

Consequently, with this modeling procedure it was possible to observe differences between different enzyme/substrate combinations that were correlated to experimental data.



## B 4 Discussion

### B 4.1 Molecular Dynamics Simulations of IMP-1 in Complex with the Mercaptocarboxylate Inhibitor

#### B 4.1.1 Purely Nonbonded Model versus Cationic Dummy Atom Model

Two nonbonded approaches were used to model the IMP-1 metallo- $\beta$ -lactamase as it had been crystallized in complex with an inhibitor [28]: a purely nonbonded approach [58] and a cationic dummy atom approach [59-61].

In the presence of four ligands tetrahedral coordination would be the expected structure for  $Zn^{2+}$  as a transition metal with  $d^{10}$  configuration and one empty  $4s$  and three empty  $4p$  orbitals, which can form four tetrahedrally arranged  $4s4p^3$  hybrid orbitals and accommodate electron density from ligands. With slight deviations, this architecture was also observed in the protein environment.

With a purely nonbonded approach it was impossible to keep the tetrahedral coordination in an MD simulation. The electrostatic interactions are not directed, thus allowing any other potential ligand close to the  $Zn^{2+}$  enter into its first coordination sphere. This method might work, if there are only four ligands available, but this is usually not the case in a complex protein environment, where there are typically other amino acids with sufficient flexibility and the suitable electrostatic potential to function as additional ligands. The entry of additional ligands, glutamic acid or aspartic acid, into the first coordination sphere has been reported with the AMBER 95 and also the CHARMM force field [60, 61, 58]. In the nonbonded model used in this work, additionally to acidic groups (the second oxygen of aspartate 81 and the carboxylate of the inhibitor), cysteinate 158 coordinated not only to Zn2, but also to Zn1 and therefore acted as a bridging ligand between the two zincs like the mercaptocarboxylate thiolate in the modeled structure (Figure B 13B).

Using the cationic dummy atom approach [60], this problem could be circumvented. The tetrahedral coordination of the zinc(II) atoms was kept throughout the two 1 ns MD simulations. Aspartic acid 81 stayed coordinated to Zn2 with only one oxygen and cysteine 158 coordinated only to Zn2, but not to Zn1 (Figures B 15B and B 17B). An explanation for this observation might be the more directed electrostatic interactions between ligands and transition metal atom in this method. There is a smaller force attracting other potential ligands

close to the zinc atom, because each dummy atom only has a +0.5 charge and each dummy is already „occupied“ by a negatively charged ligand. It should be mentioned, that it was really necessary to deprotonate the ligands for this model. Simulations with neutral histidines were also performed, but in the course of the simulation they lost contact to the dummies (data not shown).

Another severe problem with the nonbonded model was the strong electrostatic repulsion of the two zinc(II) atoms, which resulted in an increase of their distance by 1.1 Å in the 450 ps MD simulation (Figure B 12C). An according observation was made in an MD simulation of a phosphotriesterase also containing a binuclear zinc metal center when using a classical nonbonded approach [61]. This finding probably arises from the fact, that the very close zinc atoms had a charge of +2. This does not describe nature very well, as in the complex a delocalization of the electron density from the ligands into the vacant metal orbitals takes place, therefore reducing the positive charge there. The cationic dummy atom approach mimics this effect to some extent by delocalizing the positive charge from the center of the metal atom towards the four ligands by 0.9 Å, respectively. As seen from Table B 3.1 and B 3.2, the geometry still slightly differed from the crystal structure in the distances (+0.2 Å for the zinc-zinc distance and -0.2 Å for the ligand-zinc distances in average) and the angles (15° deviation at the maximum). However, taking into account the resolution of 2.0 Å of the crystal structure this model can still be considered a very good representation of the IMP-1 metallo-β-lactamase.

#### **B 4.1.2 Protonation State of the Inhibitor**

In nonbonded models the electrostatic interactions and therefore the protonation state of the modeled molecules may have a big impact on the results of the simulation. According to the cationic dummy atom method [60] the thiol group of the inhibitor was a first coordination shell ligand and therefore deprotonated. For the carboxyl group, both protonation states were possible and both possibilities were investigated. Also when protonated, the carboxyl group remained oriented towards lysine161. Therefore deprotonation did not seem to be necessary for this interaction. In the model with the protonated carboxyl group the protein deviated more from the crystal structure than that with the carboxylate. This could be seen from overlaying the crystal structure and the modeled structures (Figure B 15B compared to Figure B 17B) as well as by the RMS values (1.5 Å for residues 50-200 (Figure B 14B) compared to 1.0 (Figure B 16B)). However, it is not clear if this effect is caused by the protonated carboxyl group.

In the model with the respective group treated as deprotonated a better accordance of the modeled structure with the crystal structure was achieved with respect to the zinc atoms and the ligating amino acids. However, the more negatively charged carboxylate now interacted with the amide group of asparagine 167 resulting in a delocalization of the inhibitor at the end of the 1 ns simulation. Therefore, in this case the deprotonation had a mischievous effect and the model with the protonated carboxyl group gave a better over-all representation of the protein.

### **B 4.1.3 Movement of the Lid and the Lip**

In all performed simulations significant movements of the lid (residues 22-32) and the lip (residues 164-172) were observed. These regions are loop structures that cover the active site. They are probably important for substrate binding. To allow a substrate to enter into the active site, they need to move. Consequently, it is not surprising that they showed high mobility during the MD simulations. Also the B factors observed in the crystal structure [Concha, 2000 #29] are higher in these regions than in the rest of the protein (apart from the N- and C-terminus). Mainly the lid and lip movement were shown to be responsible for the high and still increasing RMS values after 1 ns. The relaxation time for such movements may be much longer. Tai *et al.* reported on a 10 ns MD simulation of Mouse Acetylcholinesterase and did not achieve equilibration until 4 ns [70]. In this work, the movement of the lid was not of interest, but rather to obtain a stable coordination of the zincs in the active site. Although the lid and the side chains of the inhibitor moved, the coordination remained stable in the cationic dummy atom models. When excluding the lid and the termini of the protein, the RMS values reached saturation after 1 ns. When measuring distances and angles in the active site that were used for the assay in this work, it was found that they fluctuated about a constant average value as is shown for the zinc-zinc distance e.g. in Figure B 14C. The average values were significant.

## **B 4.2 MD Simulations with IMP-1 as Free Enzyme and in Complex with Cephalothin**

A 300 ps MD simulation with the free enzyme was performed. The sulfur atom of the inhibitor, coordinated to both zincs as a bridging ligand, was substituted by a hydroxide oxygen. This hydroxide is believed to be the nucleophilic agent that attacks the  $\beta$ -lactam bond

[13]. Although some conformational changes occurred, the tetrahedral coordination of the zincs remained stable during the simulation. The zinc-zinc distance was 3.6 Å and constant throughout the 300 ps period. This value is close to the zinc-zinc distance of 3.5 Å in the structurally very similar *Bacteroides fragilis* metallo-β-lactamase that was crystallized as a free enzyme [45]. The IMP-1 metallo-β-lactamase was also crystallized as a free enzyme, but an acetate molecule took in the role of the bridging ligand and not a hydroxide (PDB entry 1DDK [28]). The zinc-zinc distance was 3.3 Å. However, it has to be mentioned that in both free enzyme structures Zn2 is not coordinated tetrahedrally but trigonal-bipyramidally. In addition to the hydroxide, aspartate 81, cysteine 158 and histidine 197, a water molecule is coordinated to this zinc atom. Zn2 also showed a trigonal bipyramidal coordination. In spite of this difference concerning the coordination of Zn2, the model simulated in this work gives a stable representation of the active site of the free enzyme.

Also the IMP-1 enzyme in complex with a cephalothin intermediate could be simulated successfully. This intermediate corresponds to the structure that forms before the protonation of the nitrogen resulting from the amide bond hydrolysis [13]. The protonation and subsequent release of the product are the rate-limiting steps. The intermediate was modified in two points: the carboxyl group resulting from amide bond hydrolysis was deprotonated and the carboxylate at the six-ring of cephalothin was protonated (see Figure B 5 and B 9B). The former was necessary to keep the oxygen coordinated to Zn1, the second to prevent replacement of the original ligands by this group. Both modifications are in agreement with the model [60], that includes deprotonation of first coordination shell ligands and protonation of second coordination shell ligands. Additionally, the simulation with the deprotonated inhibitor has shown, that deprotonation of the carboxyl group oriented towards lysine 161 does not necessarily stabilize the docked molecule. The zinc-zinc distance in this structure was significantly increased (4.6 Å compared to 3.6 Å in the crystal structure with the inhibitor). As the zincs remained at distances close to the crystal structure in the cationic dummy atom models with the inhibitor and with hydroxide, it can be assumed that in this case the increase is the result of the different binding mode of the substrate intermediate and not the result of bad modeling as observed in the purely nonbonded simulation. The fact that the intermediate structure could be modeled at 300 K over 1.2 ns indicates that this is actually a relevant structure and supports the mechanism of hydrolysis proposed by Wang *et al* [13].

Both structures, IMP-1 as a free enzyme and in complex with a  $\beta$ -lactam intermediate occur in the mechanism of hydrolysis as described [13] (see Figure B 5). However, a big difference in the zinc-zinc distance was observed. The small distance in the simulated structure of the free enzyme has also been observed experimentally in the crystal structure of the *B. fragilis* enzyme [45]. No crystal structure of the enzyme in complex with a substrate intermediate is available. However, quantum-mechanical calculations have been performed with the *B. fragilis* enzyme [65]. It was observed, that when aspartate 103, corresponding to aspartate 81 in IMP-1, was protonated, the Zn1-OH-Zn2 bridge broke and the Zn1-Zn2 distance increased to 4.5-5 Å. This is in good agreement with the results of the MD simulation of IMP-1 in complex with the cephalothin intermediate. Here also the Zn1-OH-Zn2 bridge does not exist and an increased Zn1-Zn2 distance of 4.6 Å after 1.2 ns was observed. Consequently, in the course of the hydrolysis of a  $\beta$ -lactam a change of the Zn1-Zn2 distance would occur.

Another point that was investigated was if asparagine 167 interacted with the bound substrate. In the simulation with the deprotonated inhibitor, the inhibitor carboxylate interacted with the asparagine 167 side chain. In the report on the IMP-1 crystal structure [28] it was found that the carboxyl group of the inhibitor interacted with the main chain amide nitrogen of asparagine 167 and the carbonyl oxygen of the inhibitor with the side chain of asparagine 167 through a water molecule. It has also been assumed that substrates are stabilized through an interaction of the amide carbonyl oxygen and the asparagine 167 side chain amide [13]. However, no such interaction could be observed in the MD simulation at 300 K. This is in accordance with experimental data: Haruta *et al.* found that an Asn167Ala mutation had no effect on the efficiency of hydrolysis for cephalothin [52]. Materon and Palzkill even reported, that the same mutation showed more efficient hydrolysis of some  $\beta$ -lactam substrates (ampicillin, nitrocefin, cefotaxime, cephaloridine) and concluded that asparagine 167 is not critical for the binding and turnover of those substrates [51].

### **B 4.3 Assay for Substrate Specificity of Metallo- $\beta$ -lactamases**

Only selected enzyme/substrate combinations remained stable at 300 K in the intermediate structure. For most combinations, i.e. those with low  $k_{cat}/K_M$  values the anionic nitrogen resulting from amide bond hydrolysis lost contact and was substituted by the second carboxylate oxygen. Therefore, MD simulations were carried out at 100 K. This does not exactly represent nature, but it was still possible to compare different enzyme/substrate complexes because they were all treated equally and simulated at 100 K. Several simulations

were carried out for each enzyme/substrate combination and the N (anionic nitrogen of the substrate intermediate)-dummy distance, the Zn1-Zn2 distance and the N-dummy-Zn2 angle, were monitored at the end of the simulations.

### B 4.3.1 Anionic Nitrogen-Dummy Distance

This parameter was correlated to experimental data: For IMP-6/CAZ with the lowest  $k_{cat}/K_M$  value of all investigated combinations, the N reproducibly lost contact to the Zn2-dummy. Consequently this parameter could be used to investigate if an unknown substrate would be converted well or very poorly (or not at all). An important factor for the substitution of the N by the carboxylate O might be the charge of both atoms. They are listed in Table B 3.6.

**Table B 3.6.** Charges of the anionic nitrogen and the carboxylate oxygen of the modeled cephalosporins

Atom	Charges of the Anionic N and Carboxylate O of			
	CEF	CTX	LOR	CAZ
N	-0.4984	-0.4473	-0.4709	-0.3341
O	-0.8016	-0.8160	-0.7660	-0.7691

The N in CAZ had clearly the lowest charge and the O was almost as negative as in the other substrates. Therefore it could more easily displace the nitrogen. Its higher affinity to the dummy atom was also clear from the fact that it bound with a distance of only 1.0 Å, whereas the N was at a distance of 1.5 Å (Figure B 25A and B). Nevertheless, IMP-1 could be modeled in the intermediate structure with the same substrate. In two out of three simulations the intermediate remained intact, in IMP-6/LOR it did not stay intact in a single simulation. This indicates that differences between two metallo-β-lactams could be observed.

### B 4.3.2 Zn1-Zn2 Distance and Angle Between the Anionic Nitrogen, the Dummy Atom and Zn2

Both parameters seemed to be important and somehow correlated to the efficiency of β-lactam hydrolysis. The values of all simulations (except one outlier) were located on a plane when plotting the data in a 3D chart with the experimentally observed  $k_{cat}/K_M$  values, the zinc-zinc distance and the N-dummy-zinc angle indicated on the three axes. There was a certain

variance, but when projecting the data on a 2D chart, significant differences could be observed between combinations with efficient and inefficient hydrolysis: The slopes and y axis intersections of linear regression straight lines were correlated to experimentally determined  $k_{cat}/K_M$  values. Differences were observed for one enzyme in combination with two different substrates (IMP-1/LOR and IMP-1/CAZ) as well as for the two enzymes in complex with the same substrate (IMP-1/LOR and IMP-6/LOR). However, the latter was smaller.

It is interesting that the zinc-zinc distance and the measured angle are correlated. In a perfect tetrahedral coordination of zinc, the ligands would be arranged in a way, that between them an angle of  $109.5^\circ$  would exist. This is exactly the arrangement of the dummy atoms. In the cationic dummy atom approach the ligands are oriented towards the dummies. If no steric or electrostatic interactions disturb this arrangement, the zinc, the dummy atom and the ligating atom are in one line resulting in a  $180^\circ$  angle between the three atoms. If this ideal arrangement is disturbed, the angle is decreased, as observed in the case of the IMP-1 metallo- $\beta$ -lactamase when complexed to a substrate intermediate. However, the zinc atoms are somewhat mobile and can „follow“, decreasing the deformation of the angle. Such an event could directly be observed in the 1 ns simulation of IMP-1/CEF: the zinc-zinc distance and the angle increased simultaneously. Therefore both factors needed to be regarded.

### **B 4.3.3 Variance of the Assay**

The data points obtained in the different simulations were not located exactly on one plane in the 3D representation. Neither can this be expected. Proteins are very complex systems and small differences at the beginning of the simulation can significantly influence the outcome of the simulation, in this case also the measured parameters. For some combinations in one simulation the N lost contact to the dummy atom, in others it did not (IMP-1/LOR, IMP-1/CAY). However, this variance is due to the system and the method of MD simulations. And this also makes sense. MD simulations are carried out to mimic proteins and to learn more about their structure and function. In nature each protein can behave in a different manner from another protein with exactly the same amino acid sequence. The macroscopic effect of a protein is the cooperation of a tremendous number of single molecules. With respect to MD simulations of single proteins this means that they will never give exactly the same result when they are carried out under different conditions, but if a considerable amount of

simulations is performed, the average outcome may give a good representation of the natural system. To further validate and refine the assay, more simulations will have to be carried out.

## **B 4.4 Outlook**

By MD simulations, some parameters were found that are important for substrate specificity of different metallo- $\beta$ -lactamases. In turn, the analysis of these parameters can be used as an *in silico* assay to test new enzyme/substrate combinations for hydrolysis. The interesting thing about this assay is that it works both for different substrates and for different mutants. Therefore, it could be applied in planning the synthesis of new  $\beta$ -lactam antibiotics as well as in investigating potential mutants of existing metallo- $\beta$ -lactamases for their capability of inactivating antibiotics. In the last few decades since the discovery of the first antimicrobial agent, a race has evolved between the creation of new antibiotics and inhibitors on the one hand and microbes escaping this selective pressure with a variety of mechanisms on the other hand. The most efficient mechanism is the mutation of antibiotics-inactivating enzymes. An *in silico* assay as presented in this work might be a useful tool to predict the evolution of new and more efficient metallo- $\beta$ -lactamases before they actually evolve. Then humans would be one step ahead of the microbes and might even suppress such a frequently fatal evolution.



## B 5 References

1. Schmid, R. (2002) *Taschenatlas der Biotechnologie und Gentechnik*, 1st edn, Wiley-VCH, Weinheim.
2. Dellweg, H., Schmid, R. & Trommer, W. (1992) *Biotechnologie*, 1st edn, Thieme, Stuttgart.
3. Schlegel, H. (1992) *Allgemeine Mikrobiologie*, 7 edn, Thieme, Stuttgart.
4. Lee, W., McDonough, M. A., Kotra, L., Li, Z. H., Silvaggi, N. R., Takeda, Y., Kelly, J. A. & Mobashery, S. (2001) A 1.2-Å snapshot of the final step of bacterial cell wall biosynthesis, *Proc Natl Acad Sci U S A*. 98, 1427-31.
5. Green, D. W. (2002) The bacterial cell wall as a source of antibacterial targets, *Expert Opin Ther Targets*. 6, 1-19.
6. Graefe, U. (1992) *Biochemie der Antibiotika, Struktur, Biosynthese, Wirkungsmechanismus*, Spektrum Akademischer Verlag, Heidelberg.
7. Irvin, J. E. & Ingram, J. M. (1980) Chloramphenicol-resistant variants of *Pseudomonas aeruginosa* defective in amino acid transport, *Can J Biochem*. 58, 1165-71.
8. Umezawa, H. & Kondo, S. (1982) Mechanism of Resistance to Aminoglycoside Antibiotics in *Aminoglycoside Antibiotics* (Umezawa, H. & Kondo, H., eds) pp. 267-292, Springer, New York/Heidelberg/Berlin/Tokyo.
9. Georgopapadakou, N. & Sykes, R. (1983) Bacterial Enzymes Interaction with  $\beta$ -Lactam Antibiotics in *Antibiotics Containing the  $\beta$ -Lactam Structure* (Demain, A. & Solomon, N., eds) pp. 1-78, Springer, New York/Heidelberg/Berlin/Tokyo.
10. Cundliffe, E. (1984) Self defence in antibiotic-producing organisms, *Br Med Bull*. 40, 61-7.
11. Ambler, R. (1980) The structure of  $\beta$ -lactamases, *Philos Trans Roy Soc Lond B Biol Sci*. 289, 321-331.
12. Strynadka, N. C., Adachi, H., Jensen, S. E., Johns, K., Sielecki, A., Betzel, C., Sutoh, K. & James, M. N. (1992) Molecular structure of the acyl-enzyme intermediate in beta-lactam hydrolysis at 1.7 Å resolution, *Nature*. 359, 700-5.
13. Wang, Z., Fast, W. & Benkovic, S. J. (1999) On the mechanism of the metallo-beta-lactamase from *Bacteroides fragilis*, *Biochemistry*. 38, 10013-23.
14. Suárez, D. & Merz, K. (2001) Molecular Dynamics Simulations of the Mononuclear Zinc- $\beta$ -lactamase from *Bacillus Cereus*, *J Am Chem Soc*. 123, 3759-70.

15. Datta, N. & Kontomichalou, P. (1965) Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae, *Nature*. 208, 239-41.
16. Du Bois, S. K., Marriott, M. S. & Amyes, S. G. (1995) TEM- and SHV-derived extended-spectrum beta-lactamases: relationship between selection, structure and function, *J Antimicrob Chemother*. 35, 7-22.
17. Petrosino, J., Cantu, C., 3rd & Palzkill, T. (1998) beta-Lactamases: protein evolution in real time, *Trends Microbiol*. 6, 323-7.
18. Bauernfeind, A., Schneider, I., Jungwirth, R., Sahly, H. & Ullmann, U. (1999) A novel type of AmpC beta-lactamase, ACC-1, produced by a Klebsiella pneumoniae strain causing nosocomial pneumonia, *Antimicrob Agents Chemother*. 43, 1924-31.
19. Bonnet, R., Marchandin, H., Chanal, C., Sirot, D., Labia, R., De Champs, C., Jumas-Bilak, E. & Sirot, J. (2002) Chromosome-encoded class D beta-lactamase OXA-23 in *Proteus mirabilis*, *Antimicrob Agents Chemother*. 46, 2004-6.
20. Maveyraud, L., Golemi-Kotra, D., Ishiwata, A., Meroueh, O., Mobashery, S. & Samama, J. P. (2002) High-resolution X-ray structure of an acyl-enzyme species for the class D OXA-10 beta-lactamase, *J Am Chem Soc*. 124, 2461-5.
21. Liras, P. & Rodriguez-Garcia, A. (2000) Clavulanic acid, a beta-lactamase inhibitor: biosynthesis and molecular genetics, *Appl Microbiol Biotechnol*. 54, 467-75.
22. Heinze-Krauss, I., Angehrn, P., Charnas, R. L., Gubernator, K., Gutknecht, E. M., Hubschwerlen, C., Kania, M., Oefner, C., Page, M. G., Sogabe, S., Specklin, J. L. & Winkler, F. (1998) Structure-based design of beta-lactamase inhibitors. 1. Synthesis and evaluation of bridged monobactams, *J Med Chem*. 41, 3961-71.
23. Hubschwerlen, C., Angehrn, P., Gubernator, K., Page, M. G. & Specklin, J. L. (1998) Structure-based design of beta-lactamase inhibitors. 2. Synthesis and evaluation of bridged sulfactams and oxamazins, *J Med Chem*. 41, 3972-5.
24. Ness, S., Martin, R., Kindler, A. M., Paetzel, M., Gold, M., Jensen, S. E., Jones, J. B. & Strynadka, N. C. (2000) Structure-based design guides the improved efficacy of deacylation transition state analogue inhibitors of TEM-1 beta-Lactamase(), *Biochemistry*. 39, 5312-21.
25. Kaur, K., Lan, M. J. K. & Pratt, R. F. (2001) Mechanism of inhibition of the class C beta-lactamase of *Enterobacter cloacae* P99 by cyclic acyl phosph(on)ates: Rescue by return, *Journal of the American Chemical Society*. 123, 10436-10443.

26. Sabath, L. D. & Abraham, E. P. (1966) Zinc as a cofactor for cephalosporinase from *Bacillus cereus* 569, *Biochem J.* *98*, 11C-3C.
27. Payne, D. J. (1993) Metallo-beta-lactamases--a new therapeutic challenge, *J Med Microbiol.* *39*, 93-9.
28. Concha, N. O., Janson, C. A., Rowling, P., Pearson, S., Cheever, C. A., Clarke, B. P., Lewis, C., Galleni, M., Frere, J. M., Payne, D. J., Bateson, J. H. & Abdel-Meguid, S. S. (2000) Crystal structure of the IMP-1 metallo beta-lactamase from *Pseudomonas aeruginosa* and its complex with a mercaptocarboxylate inhibitor: binding determinants of a potent, broad-spectrum inhibitor, *Biochemistry.* *39*, 4288-98.
29. Riccio, M. L., Franceschini, N., Boschi, L., Caravelli, B., Cornaglia, G., Fontana, R., Amicosante, G. & Rossolini, G. M. (2000) Characterization of the metallo-beta-lactamase determinant of *Acinetobacter baumannii* AC-54/97 reveals the existence of bla(IMP) allelic variants carried by gene cassettes of different phylogeny, *Antimicrob Agents Chemother.* *44*, 1229-35.
30. Chu, Y. W., Afzal-Shah, M., Houang, E. T., Palepou, M. I., Lyon, D. J., Woodford, N. & Livermore, D. M. (2001) IMP-4, a novel metallo-beta-lactamase from nosocomial *Acinetobacter* spp. collected in Hong Kong between 1994 and 1998, *Antimicrob Agents Chemother.* *45*, 710-4.
31. Hawkey, P. M., Xiong, J., Ye, H., Li, H. & M'Zali, F. H. (2001) Occurrence of a new metallo-beta-lactamase IMP-4 carried on a conjugative plasmid in *Citrobacter youngae* from the People's Republic of China, *FEMS Microbiol Lett.* *194*, 53-7.
32. Iyobe, S., Kusadokoro, H., Takahashi, A., Yomoda, S., Okubo, T., Nakamura, A. & O'Hara, K. (2002) Detection of a variant metallo-beta-lactamase, IMP-10, from two unrelated strains of *Pseudomonas aeruginosa* and an *Alcaligenes xylosoxidans* strain, *Antimicrob Agents Chemother.* *46*, 2014-6.
33. Vessillier, S., Docquier, J. D., Rival, S., Frere, J. M., Galleni, M., Amicosante, G., Rossolini, G. M. & Franceschini, N. (2002) Overproduction and biochemical characterization of the *Chryseobacterium meningosepticum* BlaB metallo-beta-lactamase, *Antimicrob Agents Chemother.* *46*, 1921-7.
34. Docquier, J. D., Pantanella, F., Giuliani, F., Thaller, M. C., Amicosante, G., Galleni, M., Frere, J. M., Bush, K. & Rossolini, G. M. (2002) CAU-1, a subclass B3 metallo-beta-lactamase of low substrate affinity encoded by an ortholog present in the *Caulobacter crescentus* chromosome, *Antimicrob Agents Chemother.* *46*, 1823-30.

35. Laraki, N., Franceschini, N., Rossolini, G. M., Santucci, P., Meunier, C., de Pauw, E., Amicosante, G., Frere, J. M. & Galleni, M. (1999) Biochemical characterization of the *Pseudomonas aeruginosa* 101/1477 metallo-beta-lactamase IMP-1 produced by *Escherichia coli*, *Antimicrob Agents Chemother.* *43*, 902-6.
36. Toney, J. H., Fitzgerald, P. M., Grover-Sharma, N., Olson, S. H., May, W. J., Sundelof, J. G., Vanderwall, D. E., Cleary, K. A., Grant, S. K., Wu, J. K., Kozarich, J. W., Pompliano, D. L. & Hammond, G. G. (1998) Antibiotic sensitization using biphenyl tetrazoles as potent inhibitors of *Bacteroides fragilis* metallo-beta-lactamase, *Chem Biol.* *5*, 185-96.
37. Huntley, J. J., Scrofani, S. D., Osborne, M. J., Wright, P. E. & Dyson, H. J. (2000) Dynamics of the metallo-beta-lactamase from *Bacteroides fragilis* in the presence and absence of a tight-binding inhibitor, *Biochemistry.* *39*, 13356-64.
38. Hammond, G. G., Huber, J. L., Greenlee, M. L., Laub, J. B., Young, K., Silver, L. L., Balkovec, J. M., Pryor, K. D., Wu, J. K., Leiting, B., Pompliano, D. L. & Toney, J. H. (1999) Inhibition of IMP-1 metallo-beta-lactamase and sensitization of IMP-1-producing bacteria by thioester derivatives(dagger), *FEMS Microbiol Lett.* *179*, 289-96.
39. Mollard, C., Moali, C., Papamicael, C., Damblon, C., Vessilier, S., Amicosante, G., Schofield, C. J., Galleni, M., Frere, J. M. & Roberts, G. C. (2001) Thiomandelic acid, a broad spectrum inhibitor of zinc beta-lactamases: kinetic and spectroscopic studies, *J Biol Chem.* *276*, 45015-23.
40. Toney, J. H., Hammond, G. G., Fitzgerald, P. M., Sharma, N., Balkovec, J. M., Rouen, G. P., Olson, S. H., Hammond, M. L., Greenlee, M. L. & Gao, Y. D. (2001) Succinic acids as potent inhibitors of plasmid-borne IMP-1 metallo-beta-lactamase, *J Biol Chem.* *276*, 31913-8.
41. Yano, H., Kuga, A., Okamoto, R., Kitasato, H., Kobayashi, T. & Inoue, M. (2001) Plasmid-encoded metallo-beta-lactamase (IMP-6) conferring resistance to carbapenems, especially meropenem, *Antimicrob Agents Chemother.* *45*, 1343-8.
42. Iyobe, S., Kusadokoro, H., Ozaki, J., Matsumura, N., Minami, S., Haruta, S., Sawai, T. & O'Hara, K. (2000) Amino acid substitutions in a variant of IMP-1 metallo-beta-lactamase, *Antimicrob Agents Chemother.* *44*, 2023-7.
43. Carfi, A., Pares, S., Duee, E., Galleni, M., Duez, C., Frere, J. M. & Dideberg, O. (1995) The 3-D structure of a zinc metallo-beta-lactamase from *Bacillus cereus* reveals a new type of protein fold, *Embo J.* *14*, 4914-21.

44. Aravind, L. (1999) An evolutionary classification of the metallo-beta-lactamase fold proteins, *In Silico Biol.* 1, 69-91.
45. Concha, N. O., Rasmussen, B. A., Bush, K. & Herzberg, O. (1996) Crystal structure of the wide-spectrum binuclear zinc beta-lactamase from *Bacteroides fragilis*, *Structure.* 4, 823-36.
46. Ullah, J. H., Walsh, T. R., Taylor, I. A., Emery, D. C., Verma, C. S., Gamblin, S. J. & Spencer, J. (1998) The crystal structure of the L1 metallo-beta-lactamase from *Stenotrophomonas maltophilia* at 1.7 Å resolution, *J Mol Biol.* 284, 125-36.
47. Fabiane, S. M., Sohi, M. K., Wan, T., Payne, D. J., Bateson, J. H., Mitchell, T. & Sutton, B. J. (1998) Crystal structure of the zinc-dependent beta-lactamase from *Bacillus cereus* at 1.9 Å resolution: binuclear active site with features of a mononuclear enzyme, *Biochemistry.* 37, 12404-11.
48. Bicknell, R. & Waley, S. G. (1985) Cryoenzymology of *Bacillus cereus* beta-lactamase II, *Biochemistry.* 24, 6876-87.
49. Fast, W., Wang, Z. & Benkovic, S. J. (2001) Familial mutations and zinc stoichiometry determine the rate-limiting step of nitrocefin hydrolysis by metallo-beta-lactamase from *Bacteroides fragilis*, *Biochemistry.* 40, 1640-50.
50. Haruta, S., Yamaguchi, H., Yamamoto, E. T., Eriguchi, Y., Nukaga, M., O'Hara, K. & Sawai, T. (2000) Functional analysis of the active site of a metallo-beta-lactamase proliferating in Japan, *Antimicrob Agents Chemother.* 44, 2304-9.
51. Materon, I. C. & Palzkill, T. (2001) Identification of residues critical for metallo-beta-lactamase function by codon randomization and selection, *Protein Sci.* 10, 2556-65.
52. Haruta, S., Yamamoto, E. T., Eriguchi, Y. & Sawai, T. (2001) Characterization of the active-site residues asparagine 167 and lysine 161 of the IMP-1 metallo beta-lactamase, *FEMS Microbiol Lett.* 197, 85-9.
53. Leach, A. (2001) *Molecular Modelling*, 2nd edn, Prentice Hall, Essex.
54. Car, R. & Parrinello, M. (1988) Structural, dynamical, and electronic properties of amorphous silicon: An ab initio molecular dynamics study, *Physical Review Letters.* 60, 204-207.
55. Rothlisberger, U., Carloni, P., Doclo, K. & Parrinello, M. (2000) A comparative study of galactose oxidase and active site analogs based on QM/MM Car-Parrinello simulations, *J Biol Inorg Chem.* 5, 236-50.

56. Kuhn, B., Jacobsen, W., Christians, U., Benet, L. Z. & Kollman, P. A. (2001) Metabolism of sirolimus and its derivative everolimus by cytochrome P450 3A4: insights from docking, molecular dynamics, and quantum chemical calculations, *J Med Chem.* *44*, 2027-34.
57. Arcangeli, C., Bizzarri, A. R. & Cannistraro, S. (2001) Molecular dynamics simulation and essential dynamics study of mutated plastocyanin: structural, dynamical and functional effects of a disulfide bridge insertion at the protein surface, *Biophys Chem.* *92*, 183-99.
58. Stote, R. H. & Karplus, M. (1995) Zinc binding in proteins and solution: a simple but accurate nonbonded representation, *Proteins.* *23*, 12-31.
59. Pang, Y. P. (1999) Novel zinc protein molecular dynamics simulations: Steps toward antiangiogenesis for cancer treatment, *Journal of Molecular Modeling.* *5*, 196-202.
60. Pang, Y. P., Xu, K., Yazal, J. E. & Prendergas, F. G. (2000) Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach, *Protein Sci.* *9*, 1857-65.
61. Pang, Y. P. (2001) Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method, *Proteins.* *45*, 183-9.
62. Roe, R. & Pang, Y. (1999) Zinc's exclusive tetrahedral coordination governed by its electronic structure, *J Mol Model.* *5*, 134-40.
63. Díaz, N., Suárez, D. & Merz, K. (2001) Molecular Dynamics Simulations of the Mononuclear Zinc- $\beta$ -lactamase from *Bacillus cereus* Complexed with Benzylpenicillin and a Quantum Chemical Study of the Reaction Mechanism, *J Am Chem Soc.* *123*, 9867-79.
64. Salsbury, F. R., Jr., Crowley, M. F. & Brooks, C. L., 3rd. (2001) Modeling of the metallo-beta-lactamase from *B. fragilis*: structural and dynamic effects of inhibitor binding, *Proteins.* *44*, 448-59.
65. Suarez, D., Brothers, E. N. & Merz, K. M., Jr. (2002) Insights into the Structure and Dynamics of the Dinuclear Zinc beta-Lactamase Site from *Bacteroides fragilis*, *Biochemistry.* *41*, 6615-30.
66. Terp, G. E., Christensen, I. T. & Jorgensen, F. S. (2000) Structural differences of matrix metalloproteinases. Homology modeling and energy minimization of enzyme-substrate complexes, *Journal of Biomolecular Structure & Dynamics.* *17*, 933-946.

67. Wasserman, Z. R. & Hodge, C. N. (1996) Fitting an inhibitor into the active site of thermolysin: a molecular dynamics case study, *Proteins*. *24*, 227-37.
68. Yazal, J. & Pang, Y. (1999) Ab Initio Calculations of Proton Dissociation Energies of Zinc Ligands: Hypothesis of Imidazolate as Zinc Ligand in Proteins, *J Phys Chem B*. *103*, 8773-9.
69. Berendsen, H., Postma, J. & VanGunsteren, W. (1984) Molecular-dynamics with coupling to an external bath, *J Chem Phys*. *81*, 3684-3690.
70. Tai, K., Shen, T., Borjesson, U., Philippopoulos, M., McCammon, J. A. (2001) Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase, *Biophys J*. *81*, 715-24.

**B 6 Attachment**

#Force field modifications for IMPl

## MASS

ZN	65.38
IM	35.45
ST	32.06
ZD	61.38
DZ	1.00
CP	12.01

## BOND

C*-C*	450.000	1.375
C*-HA	367.000	1.080
CT-NA	337.000	1.475
NA-NA	427.000	1.381
NA-CQ	427.000	1.381
CQ-HA	367.000	1.080
N3-C*	367.000	1.471
C*-C	469.000	1.409
OS-C	450.000	1.364
CT-NC	337.000	1.475
CC-CC	512.000	1.375
C*-S	227.000	1.810
CC-N3	367.000	1.471
CC-C	469.000	1.409
C-N2	490.000	1.335
C-H2	367.000	1.080
C*-ST	227.000	1.685
C-C	469.000	1.409
CV-HA	367.000	1.080
CV-ST	227.000	1.685
CR-N2	424.000	1.383
CR-ST	227.000	1.685
DZ-ZD	540.000	0.900
CP-CP	410.000	1.440
CP-C*	500.000	1.390
CP-HA	367.000	1.080

## ANGLE

CA-CT-N3	80.000	111.200
CA-CT-C	63.000	111.100
CT-C-OH	70.000	117.000
N3-CT-H1	50.000	109.500
N3-CT-S	50.000	114.700
O-C-OH	80.000	126.000
N3-CT-HC	50.000	109.500
CA-CT-H1	50.000	109.500
C*-CT-N	63.000	110.100
CT-C*-CP	70.000	120.000
C*-C*-CT	70.000	120.000
CT-C*-ST	70.000	120.000
C-CT-C*	63.000	111.100
C*-CT-H1	50.000	109.500
CP-C*-HA	30.000	128.000
C*-CP-HA	30.000	126.000
CP-CP-HA	30.000	122.000
C*-CP-CP	70.000	120.000
C*-ST-C*	70.000	120.000



CP-C*-ST	70.000	120.000
C*-CT-NA	63.000	110.100
CT-NA-NA	70.000	121.200
CT-NA-CQ	70.000	121.200
NA-NA-NA	70.000	121.200
NA-CQ-HA	35.000	115.450
NA-CQ-NA	70.000	129.100
CQ-NA-NA	70.000	121.200
NA-CT-HC	50.000	109.500
N3-C*-C*	70.000	125.000
N3-C*-C	70.000	125.000
CT-N3-C*	50.000	109.500
S-CT-C*	50.000	114.700
C*-C-O2	70.000	117.000
C*-C*-C	63.000	120.000
CT-C*-CT	70.000	125.000
ST-C*-HA	30.000	120.000
C*-CT-NC	63.000	110.100
CT-NC-CA	70.000	120.000
NC-CT-H1	50.000	109.500
NC-CA-HA	35.000	120.000
NC-CA-CA	70.000	120.000
CA-NC-CA	70.000	120.000
N3-CC-CC	70.000	125.000
CC-CC-C	63.000	120.000
C*-S-CT	62.000	98.900
CT-N3-CC	50.000	109.500
CT-CC-CC	70.000	125.000
CT-C*-S	70.000	125.000
C*-C*-S	70.000	125.000
S-CT-HC	50.000	109.500
CT-N2-CT	50.000	123.000
CC-C-O2	70.000	117.000
C-CC-N3	70.000	120.000
CT-N2-C	50.000	123.000
N2-C-H2	35.000	120.000
N2-C-N*	70.000	120.000
C-N2-H	35.000	120.000
N*-C-H2	35.000	120.000
H-N*-H	35.000	120.000
N-C-C	70.000	120.000
C-C-N*	70.000	120.000
C-C-CC	70.000	120.000
C-C-O	80.000	120.000
C-N*-OS	70.000	120.000
C-CC-NB	70.000	120.000
C-CC-CV	70.000	120.000
N*-C-CC	70.000	120.000
N*-OS-CT	60.000	109.500
OS-CT-C	50.000	109.500
NB-CR-N2	70.000	120.000
NB-CR-ST	70.000	120.000
NB-CC-CV	70.000	120.000
CC-CV-HA	35.000	120.000
CC-CV-ST	70.000	120.000
CV-ST-CR	70.000	120.000
CR-N2-H	35.000	120.000
N2-CR-ST	70.000	120.000
ST-CV-HA	35.000	120.000
DZ-ZD-DZ	55.000	109.500
DZ-DZ-DZ	55.000	60.000

DZ-DZ-ZD	55.000	35.250
C*-C-O	70.000	120.000
C*-C-OH	70.000	120.000

## DIHE

HA-C*-CP-HA	4	20.600	0.000	2.000
HA-CP-CP-HA	4	20.600	0.000	2.000
ST-C*-CP-HA	4	20.600	180.000	2.000
NA-NA-CT-HC	6	0.000	0.000	2.000
NA-NA-CQ-HA	4	6.000	180.000	2.000
NA-CQ-NA-NA	4	6.000	0.000	2.000
CQ-NA-CT-HC	6	0.000	180.000	2.000
CQ-NA-NA-NA	4	6.000	0.000	2.000
NA-NA-NA-NA	4	6.000	0.000	2.000
CT-C*-CP-HA	4	20.600	0.000	2.000
CT-NA-NA-NA	4	6.000	180.000	2.000
CT-NA-CQ-HA	4	6.000	0.000	2.000
CT-NA-CQ-NA	4	6.000	180.000	2.000
C*-CP-CP-HA	4	20.600	180.000	2.000
CP-CP-C*-HA	4	20.600	180.000	2.000
C*-CT-NA-NA	6	0.000	180.000	2.000
C*-CT-NA-CQ	6	0.000	0.000	2.000
CP-CP-C*-ST	4	20.600	0.000	2.000
CP-C*-ST-C*	4	14.500	0.000	2.000
CP-CP-C*-CT	4	20.600	180.000	2.000
C*-CP-CP-C*	4	20.600	0.000	2.000
C*-ST-C*-CT	4	14.500	180.000	2.000
CA-NC-CT-H1	6	0.000	0.000	2.000
CT-C*-C*-C	4	14.500	180.000	2.000
C*-ST-C*-HA	4	14.500	180.000	2.000
C*-C*-C-O2	4	14.500	180.000	2.000
C*-CT-NC-CA	6	0.000	0.000	2.000
N3-C*-C*-CT	4	14.500	180.000	2.000
N3-C*-C-O2	4	14.500	180.000	2.000
CT-N3-C*-C*	6	0.000	0.000	2.000
CT-N3-C*-C	6	0.000	180.000	2.000
H2-C-N2-H	4	5.800	180.000	2.000
N*-C-N2-H	4	5.800	0.000	2.000
O2-C-CC-N3	4	14.500	180.000	2.000
O2-C-CC-N3	4	14.500	0.000	2.000
CT-N2-C-H2	4	5.800	0.000	2.000
CT-N2-C-N*	4	5.800	180.000	2.000
CT-S-C*-C*	4	14.500	180.000	2.000
S-C*-C*-N3	4	14.500	180.000	2.000
S-C*-C*-C	4	14.500	0.000	2.000
CT-CC-CC-N3	4	14.500	0.000	2.000
CT-C*-C*-C	4	14.500	180.000	2.000
CT-C*-S-CT	4	14.500	180.000	2.000
CT-N3-CC-C	6	0.000	180.000	2.000
CC-CC-N3-CT	6	0.000	0.000	2.000
CC-CC-C-O2	4	14.500	180.000	2.000
CC-CC-C-O2	4	14.500	0.000	2.000
N-C-C-N*	4	14.500	180.000	2.000
N-C-C-N*	4	14.500	0.000	2.000
C-C-CC-CV	4	14.500	0.000	2.000
C-C-CC-NB	4	14.500	180.000	2.000
N*-C-C-O	4	14.500	0.000	2.000
N*-C-CC-NB	4	14.500	180.000	2.000
O-C-C-CC	4	14.500	180.000	2.000
CV-ST-CR-NB	4	14.500	0.000	2.000
CV-ST-CR-N2	4	14.500	180.000	2.000

---

CC-CV-ST-CR	4	14.500	0.000	2.000
NB-CR-N2-H	4	9.600	180.000	2.000
ST-CR-N2-H	4	9.600	180.000	2.000
CR-ST-CV-HA	4	14.500	180.000	2.000
N*-C-CC-CV	4	14.500	180.000	2.000
N-C-C-CC	4	14.500	180.000	2.000
N3-C*-C-O	4	14.500	180.000	2.000
N3-C*-C-OH	4	14.500	0.000	2.000
C*-C*-C-O	4	14.500	0.000	2.000
C*-C*-C-OH	4	14.500	180.000	2.000

## NONBON

ZN	1.950	0.250
IM	2.470	0.100
ST	2.000	0.250
ZD	3.100	0.000001
DZ	0.000	0.000
CP	1.908	0.086

## Lebenslauf

**Name:** Peter Ölschläger  
**Geboren:** 31.10.1970 in Stuttgart

### Studium:

10.1991 Universität Hohenheim  
 Beginn des Grundstudiums Biologie/Diplom  
 ab 10.1996 Parallelstudium von Biologie und Chemie LAG und Biologie/Diplom  
 31.05.1999 Abschluß des Studiums: Diplom in Biologie  
 1. Staatsexamen Biologie und Chemie  
 Diplomarbeit: „Aufreinigung und biochemische Charakterisierung einer Esterase aus *Streptomyces antibioticus* Tü4“  
 (Genetik, durchgeführt am Institut für Technische Biochemie der Universität Stuttgart unter Anleitung von Prof. Dr. Rolf D. Schmid)  
 seit 06.1999 Anfertigung der Dissertation am Institut für Technische Biochemie der Universität Stuttgart unter Anleitung von Prof. Dr. Rolf D. Schmid

### Konferenzbeiträge:

Peter Oelschlaeger, Srividhya Srikant-Iyer, Stefan Lange, Jutta Schmitt, Rolf D. Schmid  
 Expression, Refolding, Purification and Application of the Atrazine-specific Single-chain Variable-fragment (scFv) Antibody K411B.  
*Poster-Präsentation, GAb2000, International Symposium on Downstream Processing of Genetically Engineered Antibodies, Oktober 2000, Barcelona.*

Peter Oelschlaeger, Rolf D. Schmid, Juergen Pleiss  
 Molecular Dynamics Simulations of a Metallo- $\beta$ -lactamase in Complex with a Mercaptocarboxylate Inhibitor and in Complex with a  $\beta$ -Lactam.  
*Poster-Präsentation, 16. Darmstädter Molecular-Modelling-Workshop, Mai 2002, prämiert*

### Publikationen:

Peter Oelschlaeger, Srividhya Srikant-Iyer, Stefan Lange, Jutta Schmitt, Rolf D. Schmid  
 Fluorophor-linked Immunosorbent Assay (FLISA): a Time- and Cost-saving Method for the Characterization of Antibody Fragments Using a Fusion Protein of a Single-chain Antibody Fragment (scFv) and Enhanced Green Fluorescent Protein (EGFP)  
*Analytical Biochemistry, in print*

Peter Oelschlaeger, Stefan Lange, Jutta Schmitt, Martin Siemann-Herzberg, Matthias Reuss, Rolf D. Schmid  
 Identification of factors impeding the production of a single-chain antibody fragment in *Escherichia coli* by comparing *in vitro* and *in vivo* expression  
*eingereicht bei Applied Microbiology and Biotechnology*

**Eidesstattliche Erklärung:**

Hiermit versichere ich, daß ich die vorliegende Arbeit nur unter Zuhilfenahme der angegebenen Hilfsmittel selbständig verfaßt habe.

Stuttgart, 24. Juni, 2002

Peter Ölschläger