

STRUCTURE DETERMINATION OF PEPTIDES WITH ANTIMICROBIAL ACTION

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Gábor Bunkóczi
aus Debrecen, Ungarn

Göttingen, 2004

D 7

Referent: Prof. G. M. Sheldrick, PhD

Korreferent: Prof. Dr. R. Ficner

Tag der mündlichen Prüfung: 29.04.2004

ACKNOWLEDGEMENT

I am indebted to Prof. George M. Sheldrick for providing an excellent scientific atmosphere and superior working conditions both on the instrumental and the theoretic level; I was often dependent on his wide knowledge, which he never hesitated to share with me when he saw the necessity. I would explicitly like to thank him for being such a good teacher; his seminars treated nearly all aspects of crystallography and were essential in developing my understanding of the method. I hope he can assist students in the future with as much commitment as he helped my work and wish him good health and less involvement in administration so he could better concentrate on scientific problems that concern him.

I am also obliged to Dr. László Vértesy and Matthias Schiell from Aventis Pharma Deutschland for believing in the importance of X-ray crystallographic investigations, for good advice and for supplying samples. I hope the cooperation between the two groups will continue smoothly after the retirement of Dr. Vértesy and produce as many astonishing results as before.

During my work, I was continuously supported not only in the lab but also at home by my wife, Judit É. Debreczeni, whose profound knowledge, skill, experience and kindness I highly esteem. I would explicitly like to thank her for reading the whole manuscript and drawing figures.

I am grateful to Madhumati Sevvana, Christine Schlicker, Drs. Regine Herbst-Irmer, Tim Grüne, Stephan Rühl and Trixie Wagner for finding and correcting numerous mistakes in the forthcoming thesis. I thank for the whole group, especially for Drs. Thomas R. Schneider and Isabel Usón for an informal working atmosphere and for discussions. I am grateful to Helmut Dehnhardt for his excellent technical assistance and for his free and easy manner. I would also like to express my thanks to everyone who supported me in the acquisition of the German language, especially to Stephan Müller.

At last, but not at least, I thank my parents for giving me free choice in selecting my future profession, for their constant support and tolerance.

LIST OF ABBREVIATIONS

AA	–	amino acid
Ac	–	acetyl
Aib	–	α -aminoisobutyric acid
AIDS	–	acquired immunodeficiency syndrome
Ala	–	alanine
Asn	–	asparagine
Asp	–	aspartate
Asx	–	asparagine or aspartate
BESSY	–	Berliner Elektronenspeicherring
CC	–	correlation coefficient
CSD	–	Cambridge Structural Database
Dab	–	2,3-diaminobutyric acid
Dec	–	decenoic acid
DNA	–	deoxyribonucleic acid
Dpg	–	3,5-dihydroxyphenylglycine
EF	–	elongation factor
Et	–	ethyl
fMet	–	formyl methionine
GlcNAc	–	N-acetylglucosamine
Gln	–	glutamine
Glu	–	glutamate
Gly	–	glycine
HAc	–	acetic acid
HIV	–	human immunodeficiency virus
Hyp	–	4-hydroxyproline
IF	–	initiation factor
Ita	–	Δ^3 -isotetradecenoic acid
Iva	–	isovaline
Kyn	–	kynurenine
Leu	–	leucine
Lys	–	lysine

Map	–	β -methyl-aspartate
Me	–	methyl
Me-Glu	–	3-methylglutamate
Mpg	–	4-monohydroxyphenylglycine
mRNA	–	messenger RNA
MurNAc	–	N-acetylmuramic acid
NaAc	–	sodium acetate
NADPH	–	nicotinamide adenine dinucleotide phosphate
NCS	–	non-crystallographic symmetry
Orn	–	ornithine
PBP	–	penicillin-binding protein
PDB	–	Protein Databank
PEG	–	polyethylene glycol
Phe	–	phenylalanine
Pheol	–	phenylalaninol
Pip	–	pipecolinic acid
Pro	–	proline
r.m.s.d.	–	residual mean-square deviation
RNA	–	ribonucleic acid
rRNA	–	ribosomal RNA
Ser	–	serine
sp	–	species (sing.)
spp	–	species (plur.)
tRNA	–	transfer RNA
Trp	–	tryptophan
UDP	–	uridine diphosphate
UTP	–	uridine triphosphate
Val	–	valine

MOLECULAR GRAPHICS

Three-dimensional drawings of molecules were generated by MolScript (Kraulis, 1991) and rendered with Raster3D (Merritt & Bacon, 1997). Electron density figures were generated with BobScript (Esnouf, 1999) and rendered with Raster3D or generated with PyMOL (DeLano, 2003) and rendered with PyMOL's built-in renderer. Molecular surfaces were also created using PyMOL. Figure 5.16 was drawn using XtalView (McRee, 1999) and rendered with Raster3D.

TABLE OF CONTENTS

I. INTRODUCTION.....	1
THE OCCURRENCE OF DISEASES.....	1
<i>Bacteria</i>	1
<i>Viruses</i>	3
<i>Spread of diseases</i>	4
ANTIBIOTICS.....	4
<i>The bacterial cell wall</i>	6
<i>Bacterial protein synthesis</i>	10
<i>Other biosynthetic reactions targeted by antibiotics</i>	12
<i>Antibiotic resistance</i>	12
ANTIVIRAL AGENTS.....	17
<i>Major targets in antiviral action</i>	17
<i>Antivirals against HIV</i>	18
<i>HIV drug resistance</i>	22
II. METHODS IN CRYSTALLOGRAPHY.....	25
INTRODUCTION.....	25
DATA COLLECTION AND PROCESSING.....	26
<i>Indexing</i>	27
<i>Orientation and cell parameter refinement</i>	31
<i>Intensity measurement</i>	31
<i>Scaling and post-refinement</i>	35
SOLUTION OF THE PHASE PROBLEM.....	37
<i>Direct methods</i>	37
<i>Molecular replacement</i>	40
MODEL BUILDING.....	42
STRUCTURE REFINEMENT.....	43
<i>Least-squares minimisation</i>	44
<i>Maximum likelihood refinement</i>	44
<i>Restrained refinement</i>	45
<i>Common problems</i>	46
III. CRYSTAL STRUCTURES OF CEPHAIBOLS.....	47
INTRODUCTION.....	47
STRUCTURE DETERMINATION.....	49
<i>Crystallisation</i>	49
<i>Measurement</i>	50
<i>Structure solution</i>	52
<i>Model building and refinement</i>	52
STRUCTURE DESCRIPTION.....	53
<i>Amino acid sequence</i>	53
<i>Backbone conformation</i>	54
<i>Crystal packing</i>	56
BIOLOGICAL IMPLICATIONS.....	59
<i>Membrane channels</i>	59
<i>Sequence variation and association properties</i>	59
<i>Implications for mechanism of action</i>	60

IV. CRYSTAL STRUCTURE OF FEGLYMYCIN.....	61
INTRODUCTION.....	61
STRUCTURE DETERMINATION	63
<i>Crystallisation</i>	63
<i>Measurement</i>	63
<i>Structure solution</i>	66
<i>Model building and refinement</i>	68
STRUCTURE DESCRIPTION.....	71
<i>Amino acid residues</i>	71
<i>Dimeric structure</i>	71
<i>Crystal packing</i>	78
BIOLOGICAL IMPLICATIONS	81
<i>Channel formation</i>	81
<i>Membrane processes</i>	82
<i>Interaction surface</i>	82
V. CRYSTAL STRUCTURE OF LIPOPEPTIDE ANTIBIOTICS.....	85
INTRODUCTION.....	85
STRUCTURE DETERMINATION	89
<i>Crystallisation</i>	89
<i>Measurement</i>	91
<i>Structure solution</i>	97
<i>Model building and refinement</i>	98
STRUCTURE DESCRIPTION.....	99
<i>Amino acid residues and sequence</i>	99
<i>Conformation</i>	100
<i>Crystal packing</i>	105
BIOLOGICAL IMPLICATIONS	110
<i>Ca²⁺-ions and biological action</i>	110
<i>Physiological role of association</i>	111
<i>Implications for antibacterial action</i>	112
VI. SUMMARY.....	115
VII. REFERENCES.....	117

I. INTRODUCTION

THE OCCURRENCE OF DISEASES

The habitat of all living beings on Earth is the biosphere, which has a finite size and therefore interactions among the creatures are inherently present. One form of these interactions is symbiosis – the association of organisms with one another. Based on the type of association, symbiosis can further be divided into:

- (a) **mutualism**, in which both members of the association benefit.
- (b) **commensalism**, where there is no apparent benefit or harm to either member.
- (c) **parasitism**, in which one member (the parasite) lives at the expense of the other (the host).

A disease is a manifestation of a host-parasite interaction, in which from the medical point of view the host is human and the parasite is a pathogen. In most of the cases, the pathogen is either a virus or a bacterium.

Bacteria

In humans, the presence of bacteria is a general phenomenon, as they are regularly found on all body surfaces; only the blood and the internal tissues are germ-free. Our normal intestinal flora contains approximately 450 aerobic and anaerobic bacterial species in such high concentrations (10^{10} - 10^{12} bacteria per ml stool) that the number of symbiotic bacteria exceeds the number of human cells (Gracey, 1981; Lebek & Cottier, 1992). Not all of this huge amount of bacteria can cause disease, though. A large percentage of bacteria found in the intestine is actually benevolent: they take up nutrients from the stool that humans cannot normally digest and convert them to more easily absorbable ones or to essential compounds that we cannot produce (Coates, 1975; Metges, 2000; Hooper *et al.*, 2002). Another class of symbiotic bacteria includes species that are potential pathogens; these normally live in a commensal or parasitic relationship without producing a disease when the immune system, tissue resistance or anatomical barriers of the host are intact. Such microorganisms also include frequent causes of hospital infections like certain strains of *Escherichia coli* (Ochoa & Cleary, 2003) or *Enterococcus faecalis* (Kayser, 2003). However, there are some bacteria that

do not associate with their host except in the case of disease, like *Bacillus anthracis* or *Vibrio cholerae*.

Bacterial infection

If a bacterium, whether or not a constituent of the normal human flora, breaches the body surfaces and breaks into the sterile areas, an infection occurs. Although the infection normally implies colonisation, multiplication, invasion or persistence of a pathogen within a host, it is not equivalent to a disease, which refers to an infection that causes significant overt damage to the host.

There are several mechanisms in the body that provide protection against bacterial attacks. The first line of defence is undoubtedly the skin, which is a formidable barrier against bacterial penetration. In case a bacterium found its way through the skin, the immune system is responsible for destroying the intruder. Although these mechanisms function reliably in most of the cases, several bacteria have developed ways of overcoming or evading them (Dersch, 2003).

To cross the skin, bacteria can exploit a wound (e.g. vector bite, burn), and some microorganisms are capable of going through ubiquitous small skin abrasions or of direct penetration. However, most of the infections occur not on the skin, but through internal surfaces, like the respiratory, alimentary or urogenital tracts. Initially, only a small number of bacteria are deposited on these mucous surfaces. To survive, these have to compete with commensals that normally inhabit these surfaces, penetrate the mucus that covers them and adhere to epithelial cells (colonisation). Infectious organisms have developed specific means of attachment to host tissues, without which their colonies would normally be eliminated by normal host motility (Marra & Isberg, 1996). The key players in this process are the bacterial adhesins, proteins that interact with specific host cell receptors (Klemm & Schembri, 2000; Hauck, 2002). The inherent specificity in the adherence is responsible for apparent preference of particular bacteria to certain tissues or species and provides explanation for natural immunity of certain races within a species.

When a solid colony has already been established at the appropriate portal of entry, bacteria penetrate into the tissues by going through or between the epithelial cells or by destroying them (invasion). When in the tissue, bacteria have to withstand the attacks of the immune system and the specific immune response. Invasins, proteins taking part in the offensive, can effectively degrade the intercellular matrix to promote spread and act as non-specific toxins to induce lysis of tissue cells in the immediate vicinity of bacterial growth.

Finally, to be pathogenic, bacteria have to damage the tissues, which is normally achieved either by the production of a toxin or by lysing the host cell.

A bacterial infection does not necessarily imply the development of a disease. Some bacteria rarely cause disease if they infect, while certain microbes usually do. The ability of a bacterium to cause disease is generally referred to as its virulence or degree of pathogenicity. The manifesting virulence is the product of all the factors involved in the infection process (Smith, 1995). Research on bacterial pathogens aims to identify the genetic basis of these factors and to determine their function.

Viruses

Viruses are tiny particles that are built up of their genetic information (in the form of either RNA or DNA), a protein shell that protects the nucleic acids (nucleocapsid) and in enveloped viruses an additional lipid membrane. They extensively vary in their shape and complexity. Unlike bacteria, viruses are unable to function outside a host cell and scientists are divided whether or not a virus should be considered a living being.

All viruses have a common type of action. First, they recognize and then attach to the appropriate host cell by means of their protein shell and inject their genetic material into the cytosol. The injected genetic instructions recruit cell enzymes and begin to produce and assemble viral proteins. The assembled virus particles leave the host by either lysing and destroying it or by protruding from the cell and breaking away with a piece of lipid membrane.

Once inside the host cell, the virus does not necessarily start its own reproduction immediately. Some viruses, such as herpes and HIV, integrate their genetic material into the host cell's chromosomes, silently wait and allow the host cell to reproduce. When an environmental or predefined genetic signal arrives, the sleeping viral genes awake and take over the host's protein synthetic machinery.

Viruses exhibit remarkable host cell specificity. Upon approach towards the host cell, viruses require a specific receptor molecule that is complementary to the coat protein. This dependency on the presence of a certain protein places the main restriction on the type of cells a virus can infect (Baranowski *et al.*, 2001).

Spread of diseases

Although markedly different in structure and function, there are common mechanisms for transmitting a bacterial or a viral infection. In general, pathogens are transmitted either directly (e.g. through physical contact or body fluids) or with an appropriate carrier (contaminated food, water or insects). Preventing the transmission is probably the most important way of combatting a disease.

ANTIBIOTICS

Antibiotics were originally compounds isolated from soil microorganisms that were selectively toxic to other microbes. The group *Streptomyces* were found to be especially active in synthesising them and investigation of this family led to the discovery of several classes of antimicrobial compounds (e.g. streptomycin, tetracycline and erythromycin). The exact role of these compounds in the survival and economy of the producing organisms has not yet been clarified, though (Williams *et al.*, 1989). As the public need for antibiotics increased and bacterial resistance became widespread, semisynthetic antibiotics (e.g. non-hydrolysable penicillins) were also introduced into medical practice and now there exist several classes of completely synthetic compounds with antimicrobial properties (e.g. fluoroquinolones). Some commercially available antibiotics are listed in Table 1.1.

All antibiotics possess a common principle in their mechanism of action: they achieve selective toxicity against bacteria by inhibiting an important step in their biosynthetic machinery that is not present or markedly different in eukaryotes. Such targets emerge by divergent evolutionary strategies followed by prokaryotes and eukaryotes, and may be divided into two classes. On one side, bacteria have introduced new structural substances that were unnatural and thus resistant against common eukaryotic degradation mechanisms. On the other side, certain key actors in the biosynthetic machinery of eukaryotes, although bearing a common ancestor with that of prokaryotes, have undergone substantial alteration in the course of evolution or have been replaced by more advanced variants. Major targets of antibiotics include bacterial cell-wall and protein synthesis (Figure 1.1).

Antibiotic	Antimicrobial spectrum	Mode of action
β-lactams		
Penicillin G	Gram+	Inhibit transpeptidation in cell-wall synthesis by covalent modification of the enzyme
Ampicillin	Gram+ Gram−	
Cefotaxime (cephalosporin)		
Imipenem (carbapenem)		
Aztreonam (monobactam)	Gram−	
Glycopeptides		
Vancomycin	Gram+	Inhibit transpeptidation in cell-wall synthesis by binding the substrate
Teicoplanin		
Aminoglycosides		
Gentamycin	Gram+ aerobic Gram−	Inhibit protein synthesis by blocking the 70S initiation complex
Kanamycin		
Tetracyclines		
Tetracycline	Gram+ Gram−	Inhibit protein synthesis by blocking tRNA binding sites on the ribosome
Doxicycline		
Oxazolidinones		
Linezolid	Gram+	Inhibit initiation of protein synthesis by binding the 50S ribosomal subunit
Quinolones		
Ciprofloxacin	Gram+ Gram−	Inhibit transcription by preventing the action of topoisomerases/gyrases
MLS-antibiotics		
Erythromycin	Gram+ cocci Gram− anaerobes	Inhibit transpeptidase activity of the ribosome
Lincomycin		
Chloramphenicol	Gram+, Gram−	
Rifampin	Gram+ Gram−	Inhibit transcription by binding to RNA polymerase
Sulfonamides		
Trimethoprim/Sulfamethoxazole	Gram+ aerobes Gram− aerobes	Inhibit folic acid synthesis

Table 1.1

Antimicrobial spectrum and action mechanism of selected members of commercially available antibiotic classes.

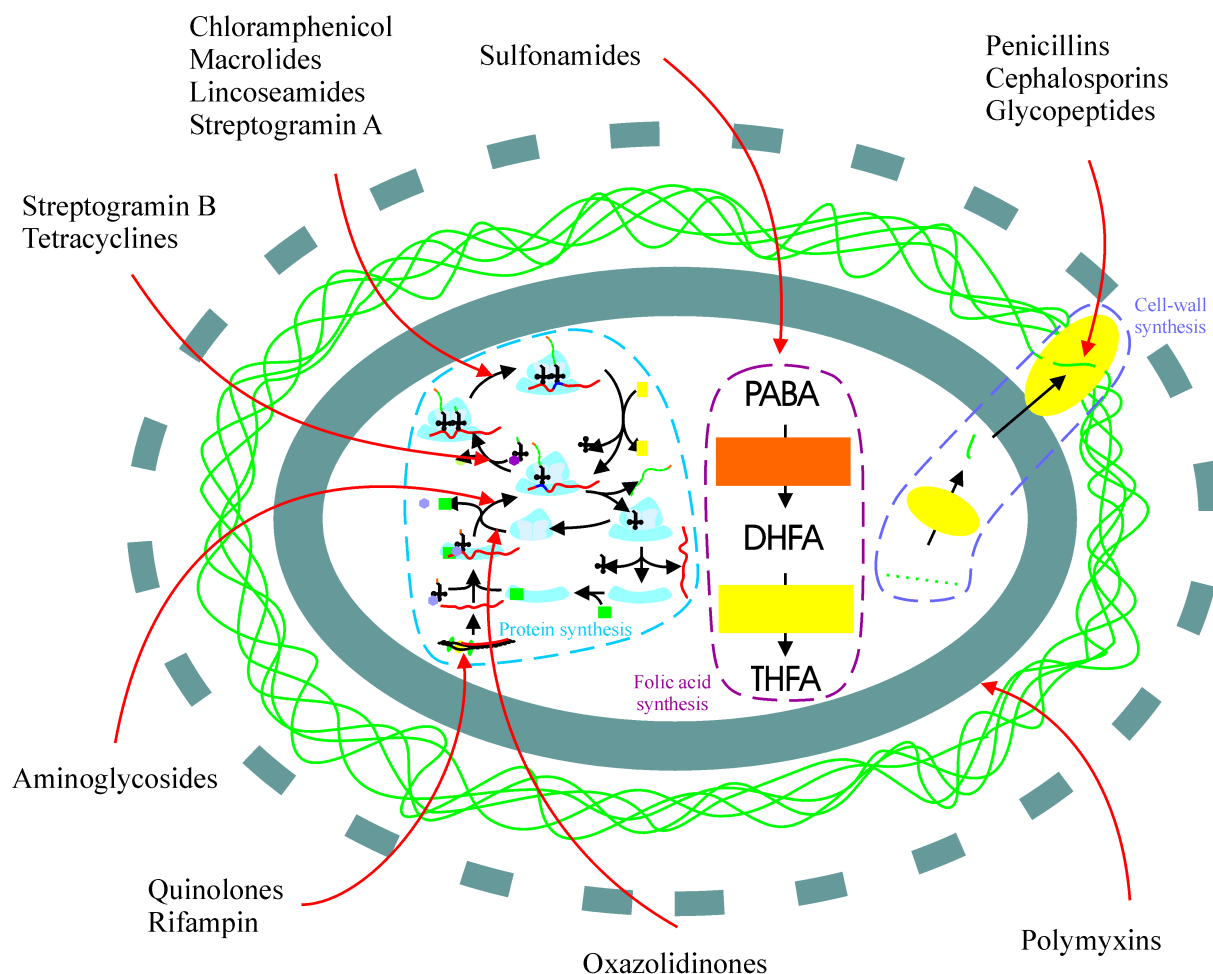


Figure 1.1

Biochemical reactions in bacteria targeted by antibiotic classes. Major targets are the protein, the cell-wall and the folic acid biosynthesis, but antibiotics having different action mechanism (e.g. permeabilisation of the cell membrane) are also known. PABA: *para*-aminobenzoic acid, DHFA: dihydrofolic acid, THFA: tetrahydrofolic acid. See Figures 1.2 and 1.3 for a more detailed reaction scheme of cell-wall biosynthesis and Figure 1.4 for that of protein synthesis.

The bacterial cell wall

The material of the bacterial cell wall, peptidoglycan or murein, is a polymer of unusual chemical structure that provides strength and rigidity to bacterial cells (Gale *et al.*, 1981). The novel features of peptidoglycan, namely the presence of D-amino acids and γ -peptide bonds, are responsible for its chemical and biological resistance, but also for its vulnerability (Neuhaus & Hammes, 1981). Disruption of the peptidoglycan layer results in cell lysis and subsequent death due to the high internal osmotic pressure present in bacterial cells.

As peptidoglycan is an essential substance for bacteria in maintaining cell integrity, it is not surprising that most naturally occurring antibiotics inhibit enzymes taking part in its biosynthesis (Waxmann & Strominger, 1983; Barna & Williams, 1984; Frère & Joris, 1985; Reynolds, 1989).

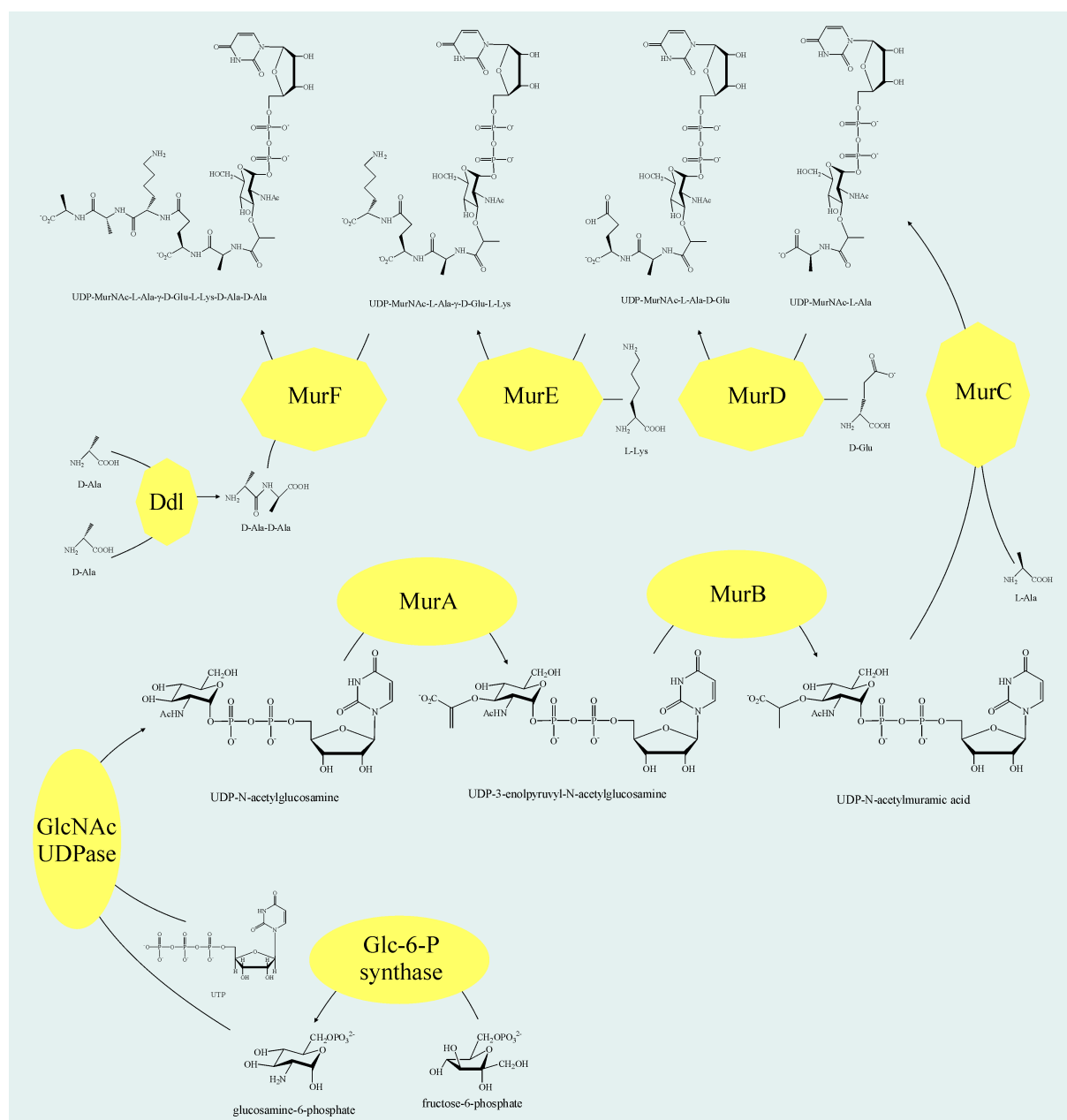


Figure 1.2

Biosynthesis of the bacterial cell wall I: reactions in the cytosol. Participating enzymes and compounds are designated according to conventional nomenclature or after the corresponding gene. The cell-wall pentapeptide corresponds to that of Gram-positive bacteria; in Gram-negative organisms L-Lys is replaced by *meso*-diaminopimelate.

The functional murein has a layered structure built up by glycan strands, which are cross-linked through peptide chains. The glycan strands consist of alternating residues of N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc) linked by β -1,4-glycosidic bonds. Peptide chains are connected to the D-lactyl group of MurNAc and join neighbouring glycan strands by means of a 9-residue sequence. Biosynthesis of murein proceeds in two stages. First, the GlcNAc–MurNAc–L-Ala– γ -D-Glu–X–D-Ala–D-Ala monomer (where X is *meso*-diaminopimelate in Gram-negative and L-Lys in Gram-positive bacteria) is assembled in the cytosol, then transported across the cell membrane and added to the continuously growing peptidoglycan layer (Schleifer & Kandler, 1972; Kamiryo & Matsubashi, 1972).

Biosynthesis of the monomer utilises the primary metabolite fructose-6-phosphate, which is converted into glucosamine-6-phosphate by the corresponding synthetase, acetylated at the amino group and linked to a UDP carrier (Figure 1.2). The resulting UDP–GlcNAc is further converted by UDP–GlcNAc enolpyruvyl transferase (encoded by the MurA gene) followed by the reduction of the enol to a lactyl group by the NADPH-dependent MurB-reductase and yields UDP-MurNAc.

The sequence of the pentapeptide moiety is not DNA-encoded and the synthesis follows a non-ribosomal pathway directed by the specificity of amino acid ligases that catalyse subsequent steps. The reaction starts with the MurC enzyme that adds an L-Ala residue to the lactyl group of UDP-MurNAc, and proceeds through the addition of D-Glu (by the MurD enzyme) to the C-terminal and then *meso*-diaminopimelate/L-Lys (MurE enzyme) to the γ -carboxylate of D-Glu. The two D-Ala residues that complete the pentapeptide are first linked by a D-Ala–D-Ala ligase (ddlA or ddlB gene) and then connected to the N-terminus of UDP-MurNAc-tripeptide by the MurF enzyme. The required D-amino acids are either synthesised from the corresponding L-amino acid or from an α -keto acid. The assembled precursor is then linked to an undecaprenyl lipid carrier by phospho-MurNAc-pentapeptide translocase (encoded on the MraY gene) and an additional GlcNAc is attached to the 4-OH group of MurNAc by MurG transferase (Figure 1.3). The complete lipid-II precursor is finally translocated across the cell membrane by an unknown mechanism and processed by penicillin-binding proteins (PBPs), which first add the precursor to the growing glycan strands (glycosyltransferase activity) and then cross-link the peptides to yield interstrand peptide brides (transpeptidase activity).

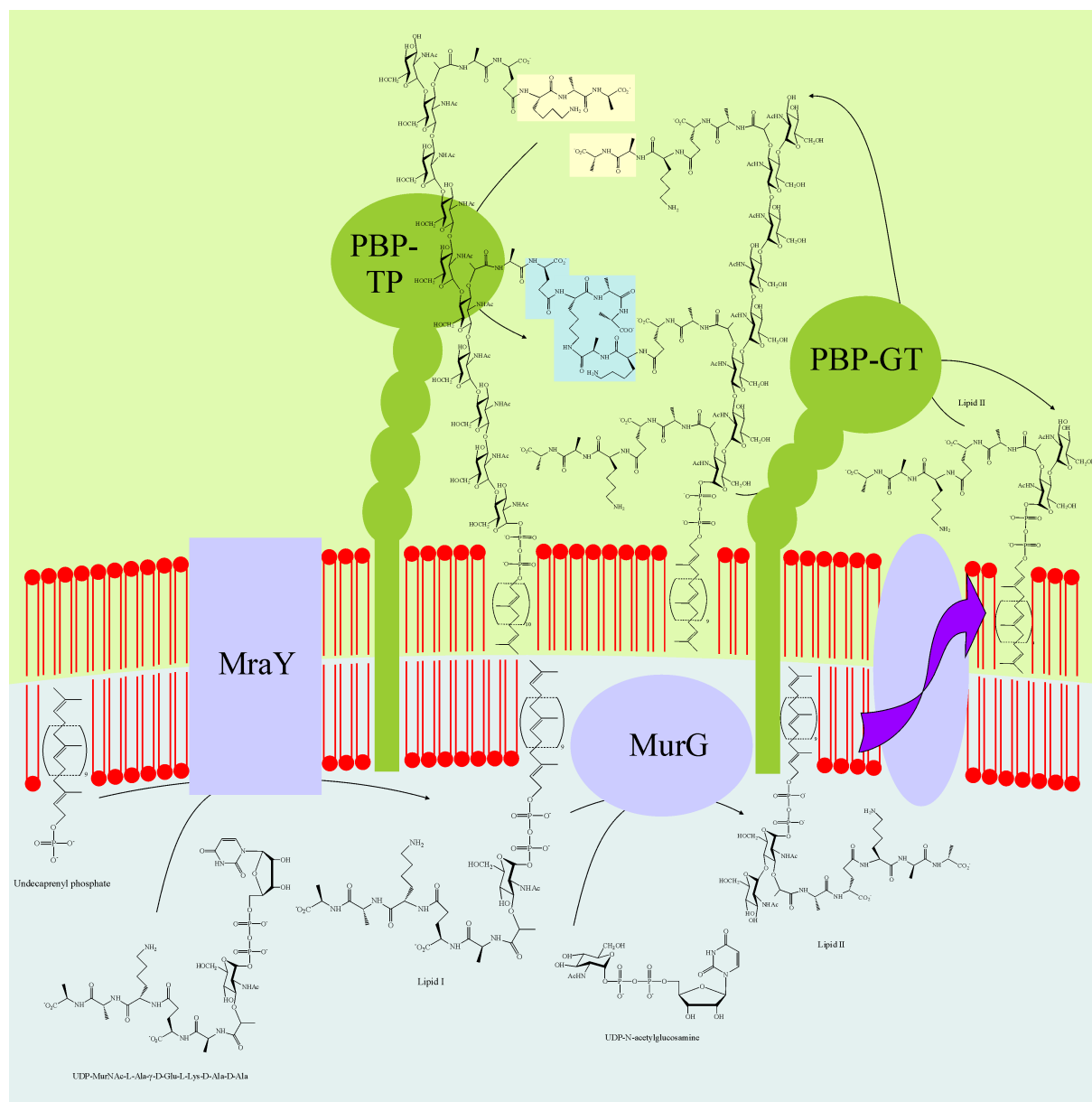


Figure 1.3

Biosynthesis of the bacterial cell wall II: reactions involving the cell membrane. PBP: penicillin-binding protein, GT: glycosyltransferase, TP: transpeptidase. The remaining enzymes are named after the encoding gene. The "flippase" enzyme has not been identified yet. The cell-wall building block corresponds to that of Gram-positive bacteria; in Gram-negative organisms L-Lys is replaced by *meso*-diaminopimelate.

Although murein biosynthesis consists of several steps, antibiotics used in current medical practice are limited in their action to transpeptidase inhibitors. The most important group is indisputably the one of β -lactams, as this group is present on the market not only with the most variants, but also with the highest quantity. Based on their chemical structure,

β -lactams may be classified into penicillin and ampicillin derivatives, cephalosporines, carbapenems and monobactams, which all use the same principle for antimicrobial action (a four-member lactam ring that can react with and thus inactivate the transpeptidase domain of the PBP) and differ only in pharmacokinetic parameters and antimicrobial spectrum. The second main group comprises glycopeptide antibiotics, which inhibit transpeptidation by binding the D-Ala–D-Ala terminus of the pentapeptide moiety, therefore preventing approach and subsequent action of the transpeptidase enzyme (Chatterjee & Perkins, 1966; Reynolds, 1989). Novel agents that target distinct steps of peptidoglycan synthesis are in development (Bugg & Walsh, 1992).

Bacterial protein synthesis

The ribosomal protein synthesis machinery found in bacteria is notably different from the one in eukaryotes and many antibiotics function by inhibiting one or more essential reactions in it. Synthesis of a protein starts generally with transcription of genetic information encoded in the DNA into messenger RNA (Figure 1.4). This process implies the action of topoisomerases that untwist the DNA to make it accessible for the RNA polymerase enzyme, which creates a complementary RNA-copy of the DNA. The completed mRNA chain then binds an inactive 30S ribosome subunit and positions the starting AUG codon to the partial P-site of the ribosome. A tRNA molecule with a formylated methionine carried by initiation factor IF-2 can now bind to the P-site. Upon binding, the IF-3 initiation factor preventing the 30S subunit from premature association with the 50S dissociates, and the 50S subunit joins in. After this step the ribosome is operational and begins to synthesise the encoded protein by adding the amino acids encoded by the codons and carried by the complementary tRNA and then repositioning itself one codon further (translocation). The aminoacyl-tRNA is carried by elongation factor (EF) Tu and translocation is initiated by EF-G. When a stop codon is reached, the ribosome releases the polypeptide chain, detaches from the mRNA and dissociates into subunits (Noller, 1991).

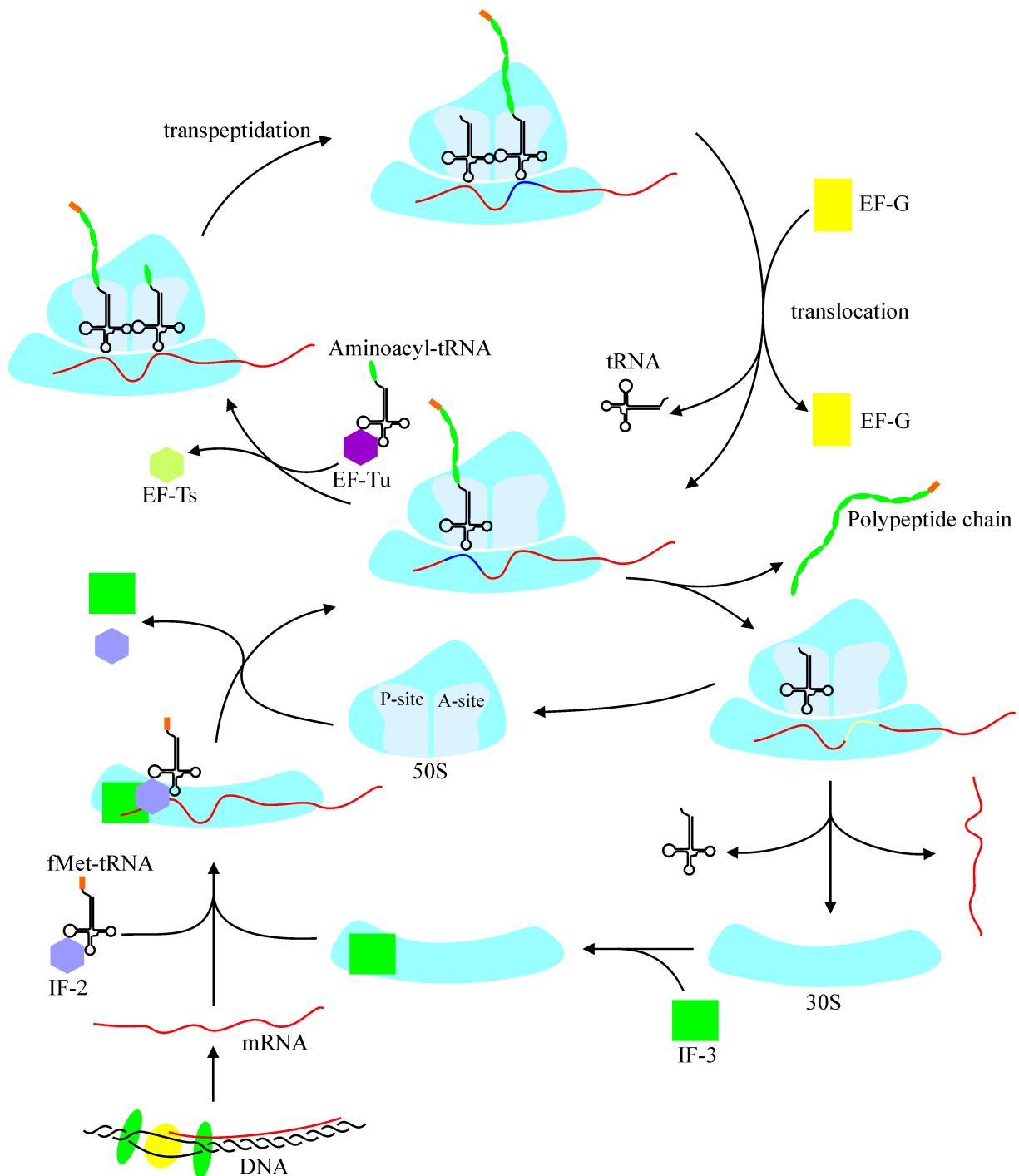


Figure 1.4

Steps of protein synthesis in bacteria. The process starts with transcription of the gene into an mRNA and ends with a polypeptide chain and the intact mRNA. This can be reused and employed for the synthesis of the same protein till it gets degraded.

Antibiotics that interfere with bacterial protein synthesis include (Figure 1.1):

- (a) **quinolones** prevent the function of topoisomerase IV and DNA-gyrase and thus inhibit transcription (Emmerson & Jones, 2003).
- (b) **rifampin** inhibits the RNA-polymerase (Furesz, 1970).
- (c) **oxazolidinones** prevent initiation of the ribosome (Norrby, 2001; Rehm, 2002).
- (d) **aminoglycosides** bind to the 16S rRNA of the 30S ribosomal subunit and block the initiation complex (Kotra *et al.*, 2000).
- (e) **tetracyclines** interact with aminoacyl-tRNA binding sites present on the 30S subunit and reduce their binding affinity (Durckheimer, 1975).
- (f) **MLS antibiotics** (macrolides, lincosamides and streptogramins) bind the 23S rRNA of the 50S subunit and block peptidyltransferase activity (Di Giambattista *et al.*, 1989; Cocito *et al.*, 1997).
- (g) **chloramphenicol** also inhibits the peptidyltransferase activity (Yunis, 1988).

However, as human cells also contain organelles that have bacterial origin (e.g. mitochondria) and where the bacterial protein synthetic machinery is partially active, these antibiotics may damage those as well. As murein synthesis inhibitors act on the cell surface and do not depend on transport across the membrane for antimicrobial action, they are not expected to cause side effects as frequently as protein biosynthesis inhibitors.

Other biosynthetic reactions targeted by antibiotics

Less common antibiotics in medical practice include the group of sulfonamides (Seydel, 1968), which inhibit the synthesis of tetrahydrofolic acid, an essential cofactor in nucleotide and amino acid biosynthesis, and polymyxins (Hermesen *et al.*, 2003), which interact with the cell membrane and cause leakage of small molecules (Figure 1.1).

Antibiotic resistance

The increased use of antibiotics in medical practice constitutes an evolutionary force that drives bacteria to develop resistance mechanisms against these antimicrobial compounds (Levy, 1998). As most of the antibiotics have natural origin or mimic those isolated from natural sources, resistance mechanisms naturally exist in antibiotic-producing microorganisms such as *Streptomyces* or in soil bacteria, against which these substances were originally directed (Cundliffe, 1984). The genes for withstanding antibiotics are therefore inherently

present in nature, but it is the appearance of resistance genes in clinical isolates of infectious bacteria that gives rise to considerable concern (Neu, 1992).

Major resistance mechanisms

Bacteria have developed four general means of withstanding the exposure to antibiotics (Hayes & Wolf, 1990):

- (a) **inactivation or reduced metabolic activation of the antibiotic.** This mechanism is responsible for low-level resistance against β -lactams and high-level resistances against aminoglycosides, lincosamides, streptogramins and chloramphenicol (Table 1.2).
- (b) **prevention of access to the target or active removal of the antibiotic from the cytosol (drug sequestration or efflux).** Examples for this mechanism include the natural resistance of certain Gram-positive bacteria to β -lactam antibiotics, as these possess an extraordinarily thick peptidoglycan layer that the antibiotic cannot penetrate, and resistances against macrolides, chloramphenicol and tetracyclines (Table 1.2).
- (c) **alteration of antibiotic target site or substitution of target enzyme (function duplication).** This is the mechanism of high-level resistance against β -lactams, which have low affinity towards certain types of PBPs, against glycopeptides, where a small modification of the cell-wall disaccharide-pentapeptide involves a 1000-fold reduction of binding affinity and against several protein synthesis inhibitors, where a mutation in the ribosome or in the inhibited enzyme renders them inactive (Table 1.2).
- (d) **increased target concentration or repair of damaged target.** Overexpressing the inhibited enzyme is the means of numerous low-level resistances, while repair is generally responsible for resistances against DNA-methylating agents.

Ways of acquiring resistance

There are several means of sharing genetic information in the microbial world, some of which are routinely used in the laboratory. Bacteria exploit all existing ways of genetic transfer to obtain the necessary genes, as these are essential for survival. In general, the exchange of genetic information can take place by transformation (exchange of DNA), transduction (bacteriophage) or conjugation by plasmids (extrachromosomal DNA), from which the latter is probably the most common (Schaberg *et al.*, 1981; Milatovic & Braveny, 1987). In addition to conjugative plasmids, bacteria possess transposons that are able to move

Antibiotic group	Resistance class	Mechanism
β -lactams	inactivation	Hydrolysis by β -lactamases
	target alteration	Reduced affinity PBPs for β -lactams
	sequestration	Suppression of porin expression
Glycopeptides	target alteration	Replacement of D-Ala–D-Ala with D-Ala–D-Lac or D-Ala–D-Ser
Aminoglycosides	inactivation	Antibiotic acetylation, phosphorylation or adenylation
	sequestration	Solid cell wall
Tetracyclines	target alteration	Ribosome conformation change
	efflux	Active removal from cytosol
Quinolones	target alteration	Mutations in topoisomerase/gyrase
MLS-antibiotics	inactivation	Antibiotic acetylation, methylation or hydrolysis
	target alteration	Methylation of an adenine in 23S rRNA of 50S ribosomal subunit
	efflux	Active removal from cytosol
Chloramphenicol	inactivation	Antibiotic acetylation
Sulfonamides	function duplication	Acquisition of a new enzyme not susceptible to the drug

Table 1.2

Resistance mechanisms against antibiotics.

between transmissible plasmids and chromosomes. These mechanisms also facilitate the spread of resistance genes among different species (Levy, 1992).

Spread of resistance genes in the community

It is now generally accepted that the spread of antibiotic resistance was initiated and driven by widespread and imprudent use of antibiotics (Neu, 1992; Levy, 2002). As a result, the resistance genes are spread over the world and it is only a question of time when they are transferred into pathogens. The following factors are believed to have an impact on the dissemination of resistance genes (Klare *et al.*, 2003):

- (a) **Use of antibiotics as growth promoters in animal husbandry.** As antibiotics employed for this purpose are relatives of those commonly used in human medicine and the dose they are given is well below the lethal concentration, bacteria that inhabit the gastrointestinal tract were placed among optimal conditions for developing resistance. As a result, the species *Enterococcus faecalis* and *Enterococcus faecium* have acquired broad-spectrum resistance against antibiotics. Although these bacteria

possess a low intrinsic pathogenicity and virulence potential, meat products manufactured from animals treated with antibiotics are often contaminated and consumption resulted in widespread occurrence of these bacteria in the human population, which facilitates the transfer of these genes to pathogens with higher virulence (Wegener *et al.*, 1999; Witte *et al.*, 2000).

- (b) **Improper use of antibiotics in human medicine.** Antibiotics are not freely available in civilised countries, but this obviously does not seem to guarantee prudent use. It has been demonstrated several times that medical doctors prescribe antibiotics even when a viral infection has been diagnosed, and therefore the control over antibiotic circulation is only apparent. Patients often do not follow the directions received and stop taking the drug after they feel better, which may result in the emergence of resistant bacterial strains (Amidi *et al.*, 1975).
- (c) **Long persistence of antibiotics in the environment.** There are relatively few mechanisms that can inactivate antibiotics and after being excreted by individuals or animals, they continue their existence in municipal waters and exert their selective pressure further (Levy, 2001).
- (d) **Increased use of antimicrobial agents (biocides) in everyday life.** Biocides have been introduced relatively recently into household use in forms ranging from floor cleaners, dishwashing detergents and antimicrobial soaps. Although these agents are not structurally related to antibiotics, there are cases reported where resistance to one of these agents resulted in a cross-resistance against antibiotics used in medical praxis. It is currently debated in the literature whether or not their increasing use is responsible for the development of antibiotic resistance (Russel, 2003).
- (e) **Spread of resistant microorganisms by individuals.** We unwarily contaminate our environment with bacteria present on our skin that can then be picked up from the infected surfaces by others. Increased circulation of goods contributes to worldwide spread of resistance (Miller *et al.*, 1996).

Strategies in overcoming bacterial resistance

Since resistant strains exist against a growing number of commercially available antibiotics, the question of controlling bacterial infections is raised with more emphasis than ever before. It now seems certain that the development of new antibiotics is not sustainable on the longer run and a different attitude towards the microbial world is needed in finding a stable solution. In the meantime, however, treatments should be developed to handle existing

health hazards (Levy, 2002). Novel trends in combatting bacterial infections can be summarised as follows.

- (a) **Probiotics.** Resistant strains could be replaced by viable, harmless microbes that can successfully compete with pathogens. Bacterial strains should be selected that constitute no risk to human health and be administered to patients. The method can also be successfully employed in preventing infection by establishing and maintaining a competitive flora (Gorbach, 1997; Bengmark, 2002).
- (b) **Blocking resistance mechanisms.** The molecular basis of antibiotic resistance can be determined and new drugs targeting the resistance mechanisms or ones not susceptible to those can be developed. This approach can restore the efficacy of whole families of antibiotics (Nelson & Levy, 1999).
- (c) **New drugs.** There are several essential steps in the biosynthetic apparatus of bacteria that can be targeted. This approach is essential to treat infections that are not susceptible to existing antibiotics (Bugg & Walsh, 1992; Di Guilmi *et al.*, 2002).
- (d) **Implementation of a prudent antibiotic use in the community.** Antibiotic resistance is chiefly supported by improper use of antibiotics and work is needed in educating people about the appropriate use of these important drugs. This movement is currently being organised by the Alliance for the Prudent Use of Antibiotics (APUA, www.apua.org).
- (e) **Vaccine development.**

In spite of every effort, a quick solution in overcoming antibiotic resistance is not expected. The resistance genes are spread over the world and it is not certain whether antibiotic resistance can be made reversible. Resistance mechanisms that do not affect biological fitness of bacteria is likely to remain a problem for a prolonged period (Andersson, 2003).

ANTIVIRAL AGENTS

The established way of preventing and handling viral infections is undoubtedly the use of vaccines. Vaccines are cheap, safe, often provide lifelong protection and their administration usually does not require expert personnel. Certain viruses, however, exhibit quick alteration of their surface proteins so that they are no longer recognised by the immune system as antigens, and therefore the production of an efficient vaccine is hindered (Stratov *et al.*, 2004). In these cases, antiviral chemotherapy is put into action.

Although the first antiviral drug has been approved as early as 1966 (Davies *et al.*, 1964), new antiviral agents are still inferior even to early antibiotics when the inhibitory efficiency and spectrum of use are compared. Despite considerable development that occurred in this field, antiviral chemotherapy is usually limited to partial inhibition of virus particles and has not yet achieved complete eradication of the infection alone. Significant repression of virus count can generally be achieved and the remainder of viruses is either eliminated or placed back under control by the immune system.

Developers of antiviral agents face unique difficulties in the course of this procedure. As free viruses do not possess operational biosynthetic machinery, inhibitory agents can influence them to a lesser extent only until they infect a cell. After infection, however, viral biosynthetic reactions are so highly integrated into the cellular facility that it proves demanding to inhibit one without interfering with the other. On the other hand, the desired high specificity implies a relatively narrow antiviral profile. An additional complication may be the relatively high variability of viral genotypes. A suitable mutation may confer resistance against the antiviral agent before actually administering the drug and prolonged exposure would select the resistant variant even if its biological fitness were impaired.

Major targets in antiviral action

Although the details are virus-specific, all viruses infect via a common mechanism that offers the following points of action for antiviral agents:

- (a) Cell-free virus.
- (b) Attachment and receptor binding.
- (c) Membrane fusion (enveloped viruses)/injection of genetic material (naked viruses).
- (d) Transcription and integration of genetic material.

- (e) Translation of genetic material.
- (f) Processing of expressed proteins and virus assembly.

Antivirals against HIV

The human immunodeficiency virus (HIV) belongs to the enveloped retrovirus family and was identified as the cause of AIDS in 1984. Globally, AIDS is now the fourth leading cause of death (Omobosola & Henry, 2003). Driven by the spread of AIDS and the growing number of HIV-infections, several antiviral agents have been developed and introduced into the market (Beach, 1998). Although current drug targets are confined to HIV reverse transcriptase and HIV protease, new drugs in clinical trials offer novel targets and are expected to ease the treatment of resistant strains (Gulick, 2003).

Proteins involved in HIV infection

The free virus is drifting in the blood stream (Figure 1.5) and recognises vulnerable CD4⁺ T-cells by a region on the cell surface that is complementary in the charge distribution to viral envelope glycoproteins (Moulard *et al.*, 2000). This region also helps in orienting the virus in the vicinity of a CD4 cellular receptor, which binds the *gp120* viral protein with high affinity (Kwong *et al.*, 1998; Rizzuto *et al.*, 1998; Wyatt & Sodroski, 1998). After receptor binding, *gp120* undergoes a conformational change that activates a surface on *gp120* to bind the cellular co-receptors CCR5 and CXCR4 (Alkhatib *et al.*, 1996; Dragic *et al.*, 1996; Berger, 1997). This event initiates another conformational rearrangement and exposes the *gp41* viral glycoprotein, which in turn promotes fusion of the cellular and viral membranes (Chan *et al.*, 1997). The virus entry is thus completed and the virion is disassembled to functional parts (uncoating).

A large nucleoprotein complex is then introduced into the cytoplasm, which contains two copies of the viral RNA and three viral enzymes essential to promote the infection. These three enzymes comprise the HIV reverse transcriptase, the HIV integrase and the HIV protease, all being promising targets of antiretroviral chemotherapy. The reverse transcriptase consists of two subunits and catalyses the RNA-directed DNA synthesis followed by degradation of the RNA-portion of the DNA-RNA hybrid and then the synthesis of a complementary DNA-chain in order to build up a double helix (Jacobo-Molina *et al.*, 1993; Katz & Skalka, 1994; Frankel & Young, 1998). When the synthesis of the viral DNA is ready, it is transferred to the nucleus where it is integrated into host DNA catalysed by the integrase enzyme (Pluymers *et al.*, 2001). Initially, the expression of viral proteins proceeds with a low

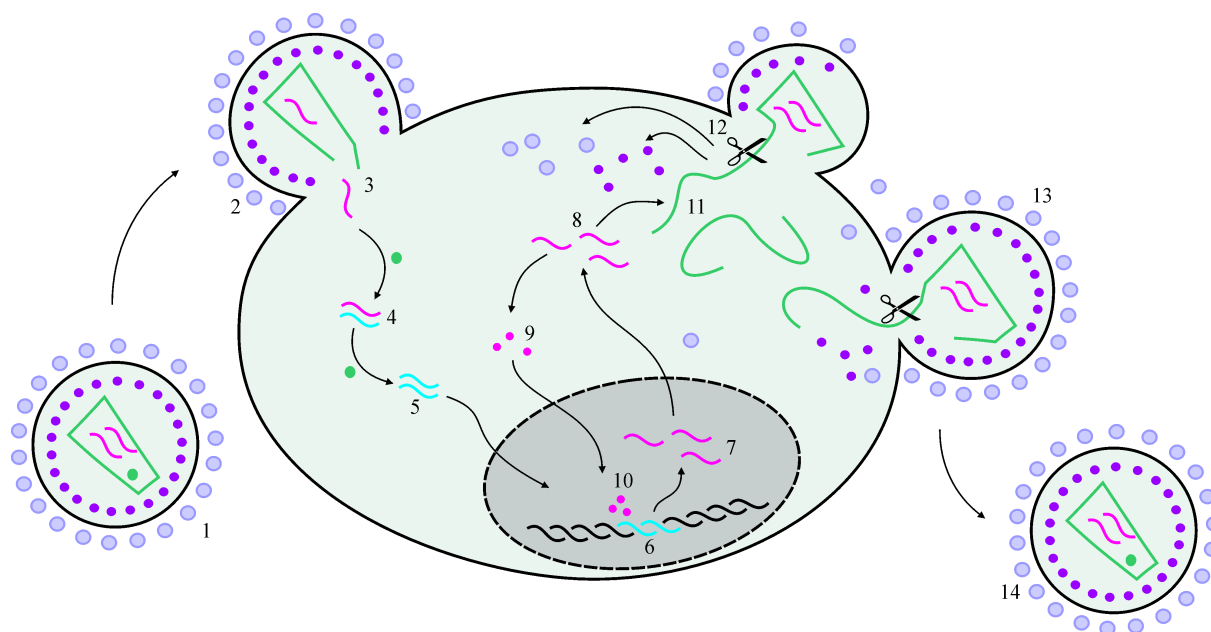


Figure 1.5

Course of HIV infection. 1: cell-free virus, 2: attachment/entry, 3: uncoating, 4: reverse transcription – RNA/DNA hybrid, 5: double-stranded DNA, 6: nuclear import/integration, 7: transcription of viral genome, 8: viral genomic RNA, 9: translation of viral regulatory proteins, 10: enhanced transcription of viral genome, 11: translation of viral polyproteins, 12: processing of viral polyproteins by HIV protease, formation of matrix, capsid and nucleocapsid proteins, 13: assembly/maturation, 14: assembled infectious virion.

rate, though, and therefore viral regulatory proteins are expressed first that enhance expression and initiate translation of the viral polyprotein complexes (Cullen, 1995). The expressed precursor proteins join the transcribed genomic RNA and form an inactive virion, which is post-processed by the HIV protease by cleaving the polyproteins at specific sites (Nagy *et al.*, 1994). The particle is then surrounded by a modified cell membrane and leaves the cell.

Reverse transcriptase inhibitors

This group of antiretroviral drugs has been launched by the introduction of zidovudine (Figure 1.6) in 1987 and still comprises the most important drugs for HIV suppression (Table 1.3). Nowadays several members are available and based on the chemical structure three distinct groups can be distinguished: nucleoside analogues, nucleotide analogues and nonnucleoside inhibitors (Beach, 1998). Albeit chemically different, these drugs inhibit the replication of the virus, therefore they prevent new infections, but do not influence virus

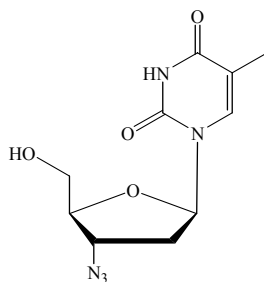


Figure 1.6

The structure of the first anti-HIV drug, zidovudine. The structure emulates a nucleoside, and is active as the triphosphate derivative. The drug is converted into the active form intracellularly by thymidine kinase. As there is no 3'-phosphate, incorporation of zidovudine results in chain termination. Although zidovudine has an appreciably higher affinity towards the HIV reverse transcriptase than towards human DNA-polymerase, it is toxic to cells where DNA-synthesis occurs.

production in already infected cells. Nucleoside and nucleotide analogues act by being a substrate of the reverse transcriptase enzyme, but cause chain termination upon incorporation since they lack the 3'-OH group. Nonnucleoside inhibitors consist of a structurally diverse group and exert their inhibitory action by binding various areas on the enzyme (Frank, 2002).

Protease inhibitors

Introduced in 1996, protease inhibitors are important constituents of highly active antiretroviral therapy since they inhibit a complementary step in the virus life cycle, namely the virus assembly (Table 1.3). Their inhibitory action causes release of immature, non-infectious virions and thus protects uninfected cells (Kitchen *et al.*, 1995). Protease inhibitors are transition-state analogues of the substrate of HIV protease (Figure 1.7), which cleaves preferentially between a phenylalanine/tyrosine and a proline residue (Wlodawer, 1994).

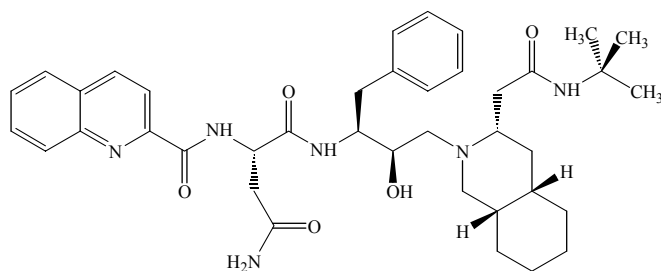


Figure 1.7

Structure of the HIV protease inhibitor saquinavir. The inhibitor emulates the structure of the transition-state intermediate for HIV protease, which cleaves preferentially between a phenylalanine/tyrosine and a proline residue.

Name	Trade name	Introduction
Nucleoside/nucleotide analogue reverse transcriptase inhibitors		
Zidovudine	Retrovir	1987
Didanosine	Videx	1991
Zalcitabin	Hivid	1992
Stavudine	Zerit	1994
Lamivudine	Epivir	1995
Abacavir	Ziagen	1999
Tenofovir	Viread	2002
Nonnucleoside reverse transcriptase inhibitors		
Nevirapine	Viramune	1996
Delavirdine	Rescriptor	1997
Efavirenz	Sustiva	1998
Protease inhibitors		
Saquinavir	Invirase	1996
Ritonavir	Norvir	1996
Indinavir	Crixivan	1996
Nelfinavir	Viracept	1997
Amprenavir	Agenerase	1999
Lopinavir	Kaletra	2000

Table 1.3

Antiretroviral agents currently used in chemotherapy against HIV.

Promising novel drugs

Currently, there are several HIV-inhibitors in various clinical phase trials (Pani *et al.*, 2002). Carrageenan, a large polysaccharide isolated from seaweed, is able to bind the virus envelope through its charged sulfate groups and may be used in preventing virus transmission. There are several compounds that target the viral *gp120* protein and hinder its binding to the CD4 receptors or to the CCR5/CXCR4 co-receptors. Functional membrane fusion inhibitors are about to be introduced into medical praxis. New compounds that act at the previously untargeted HIV integrase are also actively developed.

HIV drug resistance

The HIV reverse transcriptase is not a very specific enzyme and the fidelity of the process is relatively low when compared to other DNA or RNA-polymerases. On average, when creating a reverse transcript, one mutation is introduced (Mansky & Temin, 1995). As the replication cycle does not take longer than a day, each possible single mutant and several of the double mutants are generated within 24 hours taking into account the number of infected cells and length of the HIV genome. However, it was found that the wild-type virus possesses a higher degree of biological fitness and the mutant viruses are suppressed since their replication rate is slower (Frank, 2002; Omobosola & Henry, 2003). On the other hand, a suitable mutation may confer resistance to certain inhibitor molecules, and when antiretroviral chemotherapy is initiated, the resistant strain is selected.

Strategies in antiretroviral chemotherapy

Most of the inhibitors present in clinical practice require multiple mutations until resistance is developed (Balint, 2001); moreover, in certain cases a mutation induced by one inhibitor makes the virus susceptible again to another inhibitor, against which it has already acquired resistance (Larder, 1995). Nevertheless, the development of resistance is usually fast and takes between a week and a year when the drug is given in monotherapy (Beach, 1998), depending on the number of mutations that have to accumulate. For that reason, it is now generally accepted that anti-HIV chemotherapy should be conducted by employing a minimum of three drugs in parallel, at least two of which are reverse transcriptase inhibitors (Frank, 2002). When the virus is not allowed to replicate, new mutations do not occur and thus resistance cannot develop. Adherence to drug regimen is also an important factor; studies suggest that omitting as low as 5% of the medicament doses increases considerably the risk of losing control over the virus (Paterson *et al.*, 2000). As pill-taking imposes an enormous burden on patients, suboptimal adherence is not uncommon and an often cause of failure. Assessment of adherence is an important factor in deciding when to initiate the treatment (Frank, 2002; Omobosola & Henry, 2003).

Facing resistance

If resistance develops, one option may be to remove the given drug from the regimen and wait for the wild-type, sensitive virus to reemerge (Omobosola & Henry, 2003). However, some mutations do not significantly affect biological fitness of the virus and the mutation will persist. As certain mutations induce cross-resistance to several other inhibitors,

the emergence of a resistant strain complicates the treatment substantially. There is also a growing incidence of transmitting resistant viral strains, which exacerbates the future prospects of HIV therapy (Little *et al.*, 1999).

The development of resistance during antiretroviral therapy will sooner or later occur. Since more inhibitors are used in parallel, the resistance against a certain drug may not become manifested, but the chances of developing resistance against the remaining drugs are enhanced. For that reason, resistance testing is an important supplement to the drug regimen. In the case of virologic failure, a resistance test is performed to determine which drugs have retained their activity (Sebastian & Faruki, 2004).

A further complication of HIV is the potency of causing latent infection. The reservoir of HIV-infected cells in the body is not supposed to be cleared out in a short time and the therapy needs to be continued for decades. For this reason, HIV is considered incurable and sustainability of anti-HIV chemotherapy is dependent upon the continuous development of new antiviral agents.

II. METHODS IN CRYSTALLOGRAPHY

INTRODUCTION

The micro world has attracted the attention of mankind for a long time, but the actual deepness of examination was limited by the level of development. First investigations were most likely to be performed by eye, and later were supported by lens in form of a magnifying glass or microscope. However, as the size of the resolvable details is in the order of the wavelength of radiation used for the examination, the atomic structure of molecules cannot be seen directly through an optical device with visible light, although visible light can be used in spectroscopic techniques to obtain geometric information of the molecule. It turns out that average interatomic distances lie in the wavelength range of soft X-ray radiation, which would therefore be suitable for such an investigation, but the refractive index of nearly all materials for X-rays is very near to one, which implies that a refocused image of the object cannot be obtained.

Diffraction techniques differ from conventional microscopy in that they detect the primary diffraction image and "refocusing" is performed mathematically. X-ray diffraction can therefore be applied to determine the structure of molecules. However, as the interaction of X-rays with matter is relatively weak, the presence of a vast number of molecules is necessary to get diffraction above the level of detection. This can be achieved by using an amorphous solid or in gas phase, where the molecules are in all possible orientations, or with a crystal, where a strict ordering is present. In the following only the crystal will be discussed in detail.

Let us now consider the scattering that occurs when a single molecule is irradiated by monochromatic radiation. The molecule is supposed to consist of atoms i with position vectors \mathbf{r}_i from a suitable origin, let the wavelength be denoted by λ , and define the scattering vector \mathbf{S} as $\mathbf{S}=\mathbf{s}-\mathbf{s}_0$, where \mathbf{s}_0 and \mathbf{s} give the direction of the primary X-ray beam and any diffraction direction, respectively, and their length is by definition $1/\lambda$. Then the function $\mathbf{F}(\mathbf{S})$ can be calculated as:

$$\mathbf{F}(\mathbf{S}) = \sum_i f_i(|\mathbf{S}|) \exp(2\pi i \mathbf{r}_i \mathbf{S}) \quad (1)$$

where $f_i(|\mathbf{S}|)$ is the scattering factor of atom i , and depends only on the length of \mathbf{S} , which can be given as $\sin \Theta/\lambda$, where 2Θ is the scattering angle. The measurable scattered intensity in direction \mathbf{s} would then be proportional to $|\mathbf{F}(\mathbf{S})|^2$. However, this does not correspond to any of the scenes mentioned above, as in the first case the measured intensity distribution is a sum of scattering from all molecules, each in an individual orientation, while in case of a crystal, the convolution of scattering from one molecule with the scattering from the lattice can be measured, and can also be described as $\mathbf{F}(\mathbf{S})$ sampled on a grid determined by lattice parameters. The originally continuous $\mathbf{F}(\mathbf{S})$ is thus converted into a discrete function, which has the same value as $\mathbf{F}(\mathbf{S})$ if \mathbf{S} can be expressed as

$$\mathbf{S} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \quad (2),$$

where \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* are basis vectors of the reciprocal lattice and h , k and l are integer numbers, and zero elsewhere. The reciprocal unit cell is defined as

$$\mathbf{a}\mathbf{a}^* = 1 \text{ and } \mathbf{a}\mathbf{b}^* = \mathbf{a}\mathbf{c}^* = 0 \quad (3),$$

where \mathbf{a} , \mathbf{b} and \mathbf{c} constitute the direct-space unit cell, and analogously for all three vectors. The scattered intensity in direction \mathbf{s} is proportional to $|\mathbf{F}(\mathbf{S})|^2$, which can therefore easily be measured; phases of certain structure factor combinations can also be measured, but only with great difficulty and therefore it is not practised. Thus, half of the information necessary to "refocus" the diffraction image is lost in the measurement stage: to reconstruct $\mathbf{F}(\mathbf{S})$ as a complex number, both the amplitude and the phase are necessary. This is the crystallographic phase problem, and since it requires simultaneous phase acquisition for all reflections, it appears to be difficult to solve.

The process of X-ray crystallographic structure determination of a molecule consists of several steps: (i) crystallisation, (ii) measurement, (iii) solution of the phase problem, (iv) generation of the atomic model (model building) and (v) fitting the atomic model to the measured data (refinement).

DATA COLLECTION AND PROCESSING

Biological crystallography, in accordance to its scientific character, utilizes primary data that originate usually from a measurement. The measured data can then be analysed using various methods, but the precision of the end result is strongly dependent on the accuracy of the starting point. The measurement is therefore a crucial step in the structure determination process.

The raw data measured with an area detector consist of images that display the intensity distribution in the measured directions. This distribution is composed of two parts:

- (a) Reflections occur in directions for which the Bragg-conditions hold (i.e. \mathbf{S} can be expressed as a linear combination of the three reciprocal lattice vectors with integral coefficients) and $|\mathbf{F}(\mathbf{S})|$ is not too small.
- (b) Background scattering coming from the air, the crystal support, etc.

These two can easily be discriminated because Bragg-scattering is a discrete function, while background scattering is continuous. The information present on these images can be divided into two classes: the positions of the Bragg-reflections and their intensities.

Indexing

The crystal is usually mounted in a random orientation and in the general case, nothing is known about it. In order to be able to predict the positions of Bragg reflections, one should first determine the unit cell dimensions of the crystal and the orientation of these with respect to the laboratory coordinate system. This is equivalent to assigning indices h, k, l (or \mathbf{h} in vector notation) to all reflections and this process is thus called indexing.

To establish the lattice periodicity, the diffraction images are peak-searched to determine observed positions for the reflections. The centroids of spots are determined in two dimensions (if thin sliced data are available, also in the third, rotation direction), and converted from the laboratory coordinate system to crystal-fixed coordinates. It is crucial to obtain as accurate coordinates as possible, and it is also important to exclude any contamination originating from background scattering or electronic noise; usually only strong spots are included in the reflection list and spurious reflections, consisting of a single pixel or being too close to other spots are filtered out. These observed scattering vectors \mathbf{s}_i can then be transformed into a crystal-fixed coordinate system. Transformation from the laboratory coordinate system requires precise knowledge of the direct beam coordinates and detector distortion parameters, which should therefore be input adequately (Figure 2.1).

Given the availability of a sufficient number of spots representative of the lattice periodicity ($\mathbf{x}_i = \mathbf{s}_i - \mathbf{s}_0$ vectors, where $i=1, \dots, n$), indexing can be performed using several algorithms. The task is to determine a matrix \mathbf{A} for that

$$\mathbf{x}_i = \mathbf{A}\mathbf{h}_i \tag{4}$$

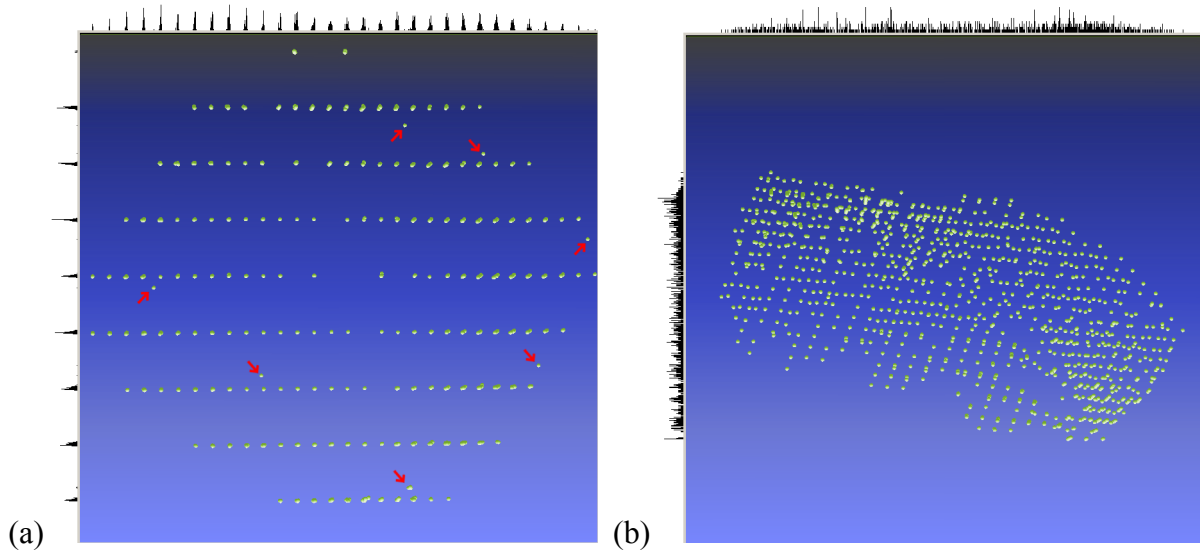


Figure 2.1

Observed reflection centroids displayed in crystal-fixed coordinate system. In (a) the lattice is oriented so that the view is along one real-space axis. The reflections form lines and cluster at certain positions (see projections on top and left). Few noise reflections (marked with red arrows) are easily identified, but they do not affect the projection seriously and their effect is marginal. Imprecise detector parameters would lead to curved lines and broad clusters and would complicate indexing. In (b) the same lattice is shown, but in a random orientation.

is fulfilled with the components of \mathbf{h}_i being close to integer numbers for most of the reflections. The matrix \mathbf{A} describes the unit cell vectors and their orientation with respect to the crystal coordinate system and can be written:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad (5)$$

the components of which are related to the reciprocal unit cell as

$$|\mathbf{a}^*|^2 = a_{11}^2 + a_{12}^2 + a_{13}^2 \quad |\mathbf{b}^*||\mathbf{c}^*|\cos\alpha^* = a_{12}a_{13} + a_{22}a_{23} + a_{32}a_{33} \quad (6)$$

and similarly for the remaining unit cell constants.

The Sparks algorithm

The three shortest non-coplanar vectors are chosen from the list of vectors and the matrix \mathbf{X} is composed from them (Sparks, 1976). Rewriting equation (4) and rearranging by taking inverses the following is obtained:

$$\mathbf{A}^{-1} = \mathbf{H}\mathbf{X}^{-1} \quad (7)$$

which can also be written as:

$$\mathbf{a}_1 = \mathbf{g}_1 \mathbf{X}^{-1} \quad \mathbf{a}_2 = \mathbf{g}_2 \mathbf{X}^{-1} \quad \mathbf{a}_3 = \mathbf{g}_3 \mathbf{X}^{-1} \quad (8)$$

where $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3$ are row vectors composed of integer numbers. The algorithm proceeds by generating integer triples starting with small numbers and solves one of equations (8) for $\mathbf{a}_1, \mathbf{a}_2$ or \mathbf{a}_3 . A solution is accepted if for most of the observed reciprocal vectors holds

$$\mathbf{a}_1 \mathbf{x}_i = h_i \quad i=1, \dots, n \quad (9),$$

where h_i is an integer, and analogously for \mathbf{a}_2 and \mathbf{a}_3 . The shortest non-coplanar vectors are selected from the generated list and cell reduction is performed. This algorithm was designed to find the unit cell from a small number of reflections (ca. 25) and then checking the axial directions to assure that not a real-space subcell has been selected. The method is very quick, but sensitive to noise reflections or reflections from satellite crystals; the three selected vectors have to belong to the same lattice. An extension of the method uses difference vectors calculated from the reflection list as:

$$\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j \quad i, j=1, \dots, n \quad (10)$$

which have a higher chance of being elementary unit cell vectors and also difference vector clusters with many nearly identical members assure that the vector is likely to correspond to a single lattice.

This algorithm can be modified to index overlapped lattices. However, the chance for selecting vectors that do not belong to one lattice is much higher in the presence of a twin component. This is overcome by difference-vector cluster analysis and by increasing the number of trials.

The Kabsch algorithm

Kabsch developed a method that depends on the availability of many reflections, as is usually the case for biological samples and area detectors (Kabsch, 1993). This algorithm starts with the generation of low-resolution difference vectors, which are analysed on a three-dimensional histogram. The histogram is peak-searched and a list of difference-vector clusters are obtained with length \mathbf{v}_i^* and population f_i ($i=1, \dots, m$). In the next step, the set of non-coplanar $\mathbf{b}_1^*, \mathbf{b}_2^*, \mathbf{b}_3^*$ vectors are selected from the list that maximise the function Q (defined as follows) and taken as the reciprocal basis vectors.

$$Q(\mathbf{b}_1^*, \mathbf{b}_2^*, \mathbf{b}_3^*) = \sum_{i=1}^m f_i \exp \left(- \sum_{k=1}^3 \left\{ \left[\max(|\xi_k^i - h_k^i| - \varepsilon, 0) / \varepsilon \right]^2 + \left[\max(|h_k^i| - \delta, 0) \right]^2 \right\} \right) \quad (11),$$

where ε and δ are suitably chosen constants, $\xi_k^i = \mathbf{v}_i^* \mathbf{b}_k$, h_k^i is the nearest integer to ξ_k^i and \mathbf{b}_k ($k=1, \dots, 3$) is the corresponding real-space axis as defined in (3). This function penalises a

basis vector set if there are many \mathbf{v}_i^* difference vector clusters with several members that deviate in indices from integer numbers by more than ε , or if these low-resolution difference vectors happen to have indices higher than δ (prevents the selection of too short basis vectors that may occur in the presence of a twin lattice). This basis set is then refined using difference vector clusters \mathbf{v}_i^* and the reduced cell is determined. Indexing of observed vectors \mathbf{x}_i is performed by a local indexing algorithm that indexes vectors with respect to reflections with previously determined indices. This algorithm performs well in most cases, but requires many reflections. It is not sensitive to errors in detector distortion parameters and may index twin lattices in favourable cases.

The Duisenberg algorithm

This has been specially devised for problem cases, i.e. twin lattices (Duisenberg, 1992). Three reflections are chosen randomly from the reflection list and the normal of their plane is calculated. A projection of all reflections is then calculated onto this direction. If the three reflections belong to the same lattice, their plane normal coincides with a direct-lattice vector and a projection onto this direction would result in clustering at positions corresponding to integer numbers. "Alien" reflections do not obey this rule and land somewhere between the clusters and are discarded (Figure 2.1). The periodicity of the clusters is determined and stored together with the number of fitting reflections. If the triplet contains "alien" reflections (noise or contamination from other twin domains), a very short period or only a few fitting reflections are obtained and therefore can be filtered out. This search is performed with a large number (>1000) of triplets and the direct axes found are combined to index as many reflections as possible in all three dimensions within a predefined deviation from integer values. Non-fitting reflections can be used in a following run to determine lattice parameters for them as well. This algorithm is comparatively slow, but gives appropriate results with few and also with many reflections.

The Fourier algorithm

In this method each direction of a hemisphere is considered on a suitable grid (Steller *et al.*, 1997). Projections are generated onto the selected direction, and the periodicity is established with a Fourier-transform (the position of largest maximum after the origin peak gives the periodicity and its height corresponds to the number of fitting vectors). Large peaks are therefore likely to indicate potential direct-lattice vectors, which are refined and stored, and at the end three linearly independent vectors are selected and cell reduction is performed. This algorithm is relatively fast but requires many reflections, as for the Fourier-

transformation to give sharp peaks several lattice plane intersections are needed. However, a small number of "alien" reflections is tolerated well and the algorithm could in principle be used to index twin lattices with a slight modification in the cell selection routine. Another improvement possibility is to perform a three-dimensional Fourier-transformation and determine all three cell dimensions in one step.

The cell_now algorithm

This program (Bruker Nonius, 2002) generates random direct-space vectors \mathbf{d} between pre-defined limits in length and then for each vector the function

$$Q = \sum_{i=1}^n w_i (T_i - M_i)^2 \quad \text{with} \quad w_i = \frac{I_i^{\frac{1}{2}} P^2}{P^2 + (T_i - M_i)^2} \quad (12)$$

is minimised, where I_i is the intensity of reflection i , $T_i = \mathbf{d} \cdot \mathbf{x}_i$ and M_i is the nearest integer. The weights w_i downweight non-fitting reflections that deviate more than P^2 from integer indices (which is similar to ε in Kabsch's formula) and thus minimise the effects of "alien" reflections or reflections from another twin domain. After locating the first domain, the program rotates the same unit cell and tries to dock it using non-fitting reflections. Since the cell is only reoriented and not redetermined, minor domains can also be located in this way. The algorithm is relatively slow, but is insensitive to high numbers of noise reflections or the presence of multiple twin domains.

Orientation and cell parameter refinement

For an accurate intensity measurement of Bragg peaks, the predicted spot positions should coincide with the observed ones preferably within one pixel. Indexing algorithms are not able to achieve such high precision since they disregard errors arising from imprecise alignment. Taking the measurement geometry into account, most of these errors can be corrected for given the availability of an indexed list of reflections. This procedure optimises a limited number of parameters (typically the unit cell dimensions, the orientation matrix, direct beam coordinates and detector distortion parameters) that affect observed or predicted reflection centroids and minimises the deviations generally with a least-squares procedure.

Intensity measurement

The raw data collected in a diffraction experiment consist of images with showing a characteristic pattern of spots. However, most of the information present on these images do not get used in subsequent procedures, and can be discarded. The only important information

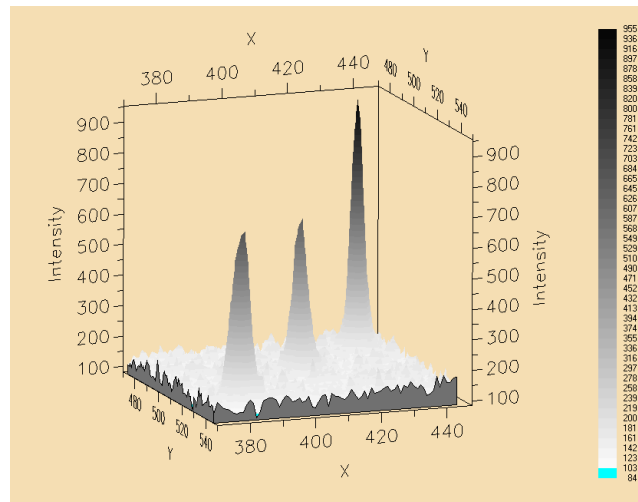


Figure 2.2

The intensity of reflections is proportional to the volume under their contours after the estimated background has been subtracted. This may be complicated when the background noise is high. Reflection profiles are also slightly different because of the spatial resolution of the detector.

that has to be extracted from the images with as high precision as possible is the intensity of the individual Bragg peaks. Since after this step the information content of the frames (from the crystallographic point of view) is converted into a different form that requires much less storage space without the loss of information, this process is often referred to as data reduction.

The intensity of a reflection is given by the volume under the peak contours in an area detector (Figure 2.2). However, to obtain a precise estimate of the reflection intensity, the effect of background and Poissonian noise has to be taken into account, which is not a trivial procedure considering that the intensities are inherently weak. In the case of area detectors, the observed intensity distribution in a suitably chosen area around a Bragg-peak can be written as:

$$I_i = I_i^{peak} + I_i^{bkg} \quad (13)$$

where I_i , I_i^{peak} , I_i^{bkg} are pixel counts that are observed, attributed to Bragg-scattering or diffuse scattering with unspecified origin, respectively, and i is a pixel within the measurement box. If the pixel i is far from the reflection centroid, the contribution of Bragg-intensity is negligible, so the observed intensity is accounted for by the background. The measurement box can therefore be divided into background (B) and peak (P) regions, with n and m pixels,

respectively. As the background is believed to be a smoothly varying function, it can be estimated as (Leslie, 1999):

$$I_i^{bkg} = ax_i + by_i + c \quad (14),$$

where a , b and c are constants, x_i and y_i are the coordinates of pixel i relative to the peak centroid. For thin-slice data collection, a third term dependent on the rotation angle can also be included (Duisenberg *et al.*, 2003). This background plane can be determined by least-squares fitting using pixels from the background region. An alternative technique in determining the background is to read consecutive frames spanning a defined angular range, average pixel values that are near to the value of the lowest (Kabsch, 2001) and then smooth them with a suitable function.

There are several factors that have to be considered when giving an estimate for the reflection intensity, I^{peak} . As the simplest approach, this can be evaluated by

$$I^{peak} = \sum_{i \in P} (I_i - I_i^{bkg}) \quad (15),$$

where I_i^{bkg} is estimated by either of the methods described above (summation integration). This is referred to as the BPB (background-peak-background) method, and has the outstanding property that if I_i^{peak} is negligible in the background region, the estimate is safe with regard to systematic error (Diamond, 1969). However, since the standard uncertainty associated with this intensity estimate can be given as

$$\sigma_{I^{peak}}^2 = G \left[\sum_{i \in P} I_i + \left(\frac{m}{n} \right)^2 \sum_{i \in B} I_i \right] \approx G \left(I^{peak} + \frac{m+n}{n} I^{bkg} \right) \quad (16),$$

where

$$I^{bkg} = \sum_{i \in P} I_i - I^{peak} \approx \frac{m}{n} \sum_{i \in B} I_i \quad (17)$$

and G is the detector gain (number of X-ray photons/pixel count), the background governs the standard deviation when the intensity of the Bragg-peak is low. The division of the measurement box into peak and background areas is also somewhat arbitrary, but to avoid systematic errors the peak region should not be selected too small. On the other hand, too large a peak region increases the standard deviation and decreases signal-to-noise for weak reflections. There are several approaches for dividing the measurement box into peak and background regions, e.g. the one that maximises $I/\sigma(I)$ (Lehmann & Larsen, 1979) or to predict the peak area based on physical factors like crystal size or beam divergence (Duisenberg *et al.*, 2003).

An improvement in the precision of weak intensities can be achieved with profile fitting that estimates the intensity as (Otwinowski & Minor, 1997)

$$I^{peak} = \frac{\sum_{i \in P} \frac{p_i(I_i - I_i^{bkg})}{\sigma_i^2}}{\sum_{i \in P} \frac{p_i^2}{\sigma_i^2}} \quad (18),$$

where σ_i^2 is the variance of pixel i and p_i is the value of the predicted profile. If the intensity of the peak compared to the background is low, the standard uncertainty of the profile-fitted intensity can be given as

$$\sigma_{I^{peak}}^2 \approx GI_{average} \frac{\left(\sum_{i \in P} p_i\right)^2}{\sum_{i \in P} p_i^2} \quad (19)$$

with all the pixels in the peak region having approximately the same value ($I_{average}$). The expected gain in signal-to-noise ratio can then be estimated (Leslie, 1999)

$$\frac{\sigma_{I_{summed}^{peak}}^2}{\sigma_{I_{profile}^{peak}}^2} = \frac{m \sum_{i \in P} p_i^2}{\left(\sum_{i \in P} p_i\right)^2} \quad (20).$$

Since the value of the right-hand side, which depends on the shape of the standard profile, is typically 2, the increase in signal-to-noise ratio for very weak reflections is about 40%. It can also be demonstrated that for strong reflections the profile-fitting procedure reduces to summation integration. However, to avoid systematic errors in intensity estimation, the predicted profile p_i has to be a good estimate of the observed reflection profile (Diamond, 1969). To ascertain a proper match, profiles are not modelled by any theoretical function, but learnt from the observed profile of strong reflections. The profile changes across the face of the detector, owing to changes in incidence angle and certain geometric factors. For that reason, the profile of a reflection is usually estimated as a weighted average of profiles from strong reflections being spatially close. An alternative approach is the transformation of pixels into a local coordinate system described by Kabsch (2001). Defining the basis vector set as

$$\begin{aligned} \mathbf{a} &= \mathbf{n} \times \mathbf{s} / |\mathbf{n} \times \mathbf{s}| \\ \mathbf{b} &= \mathbf{s} \times \mathbf{a} / |\mathbf{s} \times \mathbf{a}| \\ \mathbf{c} &= (\mathbf{s} + \mathbf{s}_0) / |\mathbf{s} + \mathbf{s}_0| \end{aligned} \quad (21)$$

where \mathbf{n} is any unit vector perpendicular to the incident beam vector \mathbf{s}_0 , the spots are treated as if they had followed the shortest path through the Ewald sphere. This transformation substantially reduces variation in reflection profile, and yields well-conserved three-dimensional model profiles applicable to all reflections. Three-dimensional profile fitting eliminates an additional problem, namely that the profile of partially recorded reflections can significantly differ from that of fully recorded reflections and depend on the degree of partiality (Leslie, 1999).

Scaling and post-refinement

Integrated intensities should be corrected for variations in incident beam intensity, diffraction volume, crystal decomposition, absorption and several other factors. This is done by minimising the deviation of symmetry or Friedel-related reflections by a correction dependent on these physical factors. The corrected intensity can be written as

$$I_c = I_o S(n) P(\mathbf{u}) R(n, \Theta) \quad (22),$$

where I_o and I_c are the uncorrected and corrected intensity, respectively, n is the frame number, \mathbf{u} represents direction cosines relative to reciprocal unit cell axes and Θ is the Bragg-angle. The first term $S(n)$ assigns a scale factor for each frame and corrects for changes in incident beam intensity and in diffraction volume. $P(\mathbf{u})$ describes anisotropic absorption by the crystal and the support and can be estimated as

$$P(\mathbf{u}) = a_0 + \sum_l \sum_{m=-l}^l a_{lm} Y_l^m(\mathbf{u}) \quad (23),$$

where a_0 and a_{lm} are coefficients and $Y_l^m(\mathbf{u})$ denotes an l -order spherical harmonic. The third term $R(n, \Theta)$ corrects for decomposition manifesting as an increase in B -values and can be written as

$$R(n, \Theta) = \exp\left(B(n) \frac{\sin^2 \Theta}{\lambda^2}\right) \quad (24),$$

where $B(n)$ gives the variation of temperature factors as a function of frame number, and since it is correlated with scale factors, a smoothly varying function should be used. These equations are valid for fully recorded reflections or for three-dimensional profile fit. The partials can either be added to form a fully recorded reflection or each partial can be used to give an estimate of the total intensity (Rossmann & van Beek, 1999)

$$I_o^n = \frac{I_n}{\zeta_n} \quad (25),$$

where I_n is the intensity measured on frame n and ζ_n is the partiality, which is defined as the volume portion of the reflection that swept through the Ewald sphere on that frame. The function minimised can be written as

$$\Psi = \sum_{\mathbf{h}} \sum_{i=1}^{n_{\mathbf{h}}} w_{\mathbf{h}}^i (I_c^i(\mathbf{h}) - \langle I(\mathbf{h}) \rangle) \quad (26),$$

where the inner summation goes for all $n_{\mathbf{h}}$ observations of the unique reflection \mathbf{h} , $\langle I(\mathbf{h}) \rangle$ is the mean intensity that can be taken as the median of all $I_c^i(\mathbf{h})$ values or as a weighted average. After the scaling procedure, the agreement of symmetry-related reflections is indicated by

$$R_{\text{int}} = \frac{\sum_{\mathbf{h}} \sum_{i=1}^{n_{\mathbf{h}}} |I_c^i(\mathbf{h}) - \langle I(\mathbf{h}) \rangle|}{\sum_{\mathbf{h}} \sum_{i=1}^{n_{\mathbf{h}}} I_c^i(\mathbf{h})} \quad (27),$$

where only those unique reflections are included in the summation that have more than two symmetry equivalents.

When all the reflections are put on the same scale, outliers (large, but sporadic fluctuations in the data) can be identified and rejected. In the last step, standard uncertainties estimated in the integration stage have to be corrected to explain the observed variance of symmetry equivalent reflections. This is usually achieved by modifying them as given in

$$\sigma_{\text{abs}}^2(\mathbf{h}) = K \left[\sigma_{\text{raw}}^2(\mathbf{h}) + (g \langle I(\mathbf{h}) \rangle)^2 \right] \quad (28),$$

where K and g are determined such, that the quantity

$$\chi^2 = \sum_{\mathbf{h}} \frac{1}{n_{\mathbf{h}} - 1} \sum_{i=1}^{n_{\mathbf{h}}} \frac{(I_c^i(\mathbf{h}) - \langle I(\mathbf{h}) \rangle)^2}{\sigma_c^2(\mathbf{h})} \quad (29)$$

is near to one and does not show any variation neither as a function of resolution nor as a function of intensity.

After the measurement one can obtain the most accurate unit cell by determining the lattice constants that best fit all observed reflection positions. In the case of a three-dimensional profile fit, reflection centroids are available and this procedure is identical to initial cell refinement apart from the fact that the values of geometric parameters are better defined. For two-dimensional profile fit procedures, centroid positions can be determined from the estimated partiality, and refinement of partiality therefore allows the determination of mosaicity and lattice parameters (Rossmann & van Beek, 1999).

SOLUTION OF THE PHASE PROBLEM

After data processing, a list of structure factors $|\mathbf{F}(\mathbf{S})|$ is available (expressed as $|\mathbf{F}_h|$ and the corresponding unit cell). However, to calculate an electron density map, the phase φ_h of the complex number

$$\mathbf{F}_h = |\mathbf{F}_h| \exp i\varphi_h \quad (30)$$

is necessary (i is the imaginary unit). There are several methods available for obtaining approximate values for the phases; an average phase error of 60° is enough to calculate an interpretable electron density map, or at least to initiate phase improvement.

Direct methods

At a sufficiently high resolution, the number of unique reflections is much higher than the number of coordinates for atoms present in the asymmetric unit, i.e. the phase problem is overdetermined. Direct methods exploit this fact and are able to provide phase estimates for most of the reflections but only if data extend to high (1.20 Å or better) resolution.

The tangent formula

Unlike the amplitudes, which are well defined apart from a global scale factor, the phases depend on the choice of the origin. It is easily conceivable that an origin shift \mathbf{o} changes the phase values according to

$$\varphi'_h = \varphi_h - 2\pi \mathbf{h} \cdot \mathbf{o} \quad (31).$$

However, it can be proved that the product of n structure factors is invariant under origin translation if

$$\mathbf{h}_1 + \mathbf{h}_2 + \dots + \mathbf{h}_n = \mathbf{0} \quad (32).$$

The phases for certain classes of these structure factor combinations (called structure invariants) can be estimated from the reflection intensities under given assumptions. One such assumption is that the electron density is not negative [if $\mathbf{F}(000)$ is included in the summation]; while another approach is to assume the presence of randomly distributed resolved point atoms, which is in practice more effective. To fit this second assumption, the observed structure factors are modified in order to remove the effects of thermal motion and normalised structure factors are introduced as

$$|\mathbf{E}_h|^2 = \frac{|\mathbf{F}_h|^2 / \varepsilon}{\langle |\mathbf{F}_h|^2 / \varepsilon \rangle} \quad (33),$$

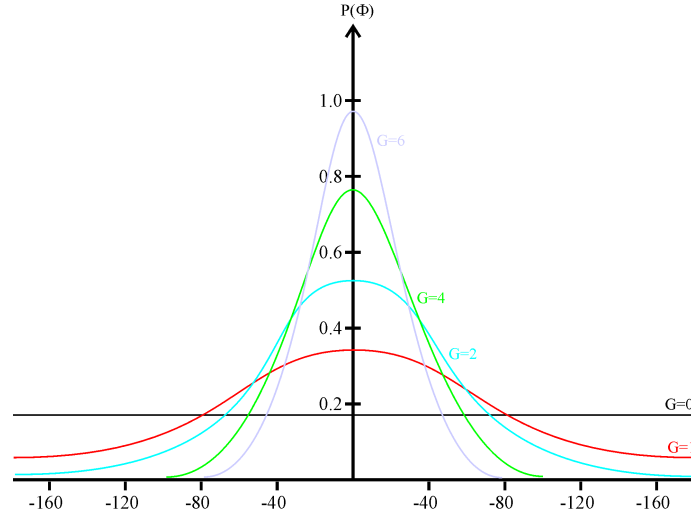


Figure 2.3

The Cochran-distribution for $\Phi_{\mathbf{h},\mathbf{k}}$. G is defined as $2N^{1/2}|\mathbf{E}_{\mathbf{h}}\mathbf{E}_{\mathbf{k}}\mathbf{E}_{-\mathbf{h}-\mathbf{k}}|$.

where ε is the statistical weight of reflection \mathbf{h} , and the average is taken in resolution shells. It can be shown that the phase $\varphi_{\mathbf{h}}$ given the values of $|\mathbf{E}_{\mathbf{h}}|$, $|\mathbf{E}_{\mathbf{k}}|$, $|\mathbf{E}_{-\mathbf{h}-\mathbf{k}}|$, $\varphi_{\mathbf{k}}$ and $\varphi_{-\mathbf{h}-\mathbf{k}}$ follows a von Mises-type distribution (Figure 2.3) given as (Cochran, 1955)

$$P(\varphi_{\mathbf{h}}) = \frac{1}{2\pi I_0(2N^{1/2}|\mathbf{E}_{\mathbf{h}}\mathbf{E}_{\mathbf{k}}\mathbf{E}_{-\mathbf{h}-\mathbf{k}}|)} \exp\left[2N^{1/2}|\mathbf{E}_{\mathbf{h}}\mathbf{E}_{\mathbf{k}}\mathbf{E}_{-\mathbf{h}-\mathbf{k}}|\cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})\right] \quad (34).$$

The phase distribution of the triplet structure invariant ($\Phi_{\mathbf{h},\mathbf{k}} = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}$) is centred on zero, and the greater the value of $N^{1/2}|\mathbf{E}_{\mathbf{h}}\mathbf{E}_{\mathbf{k}}\mathbf{E}_{-\mathbf{h}-\mathbf{k}}|$, the more reliable is this phase value (Figure 2.4).

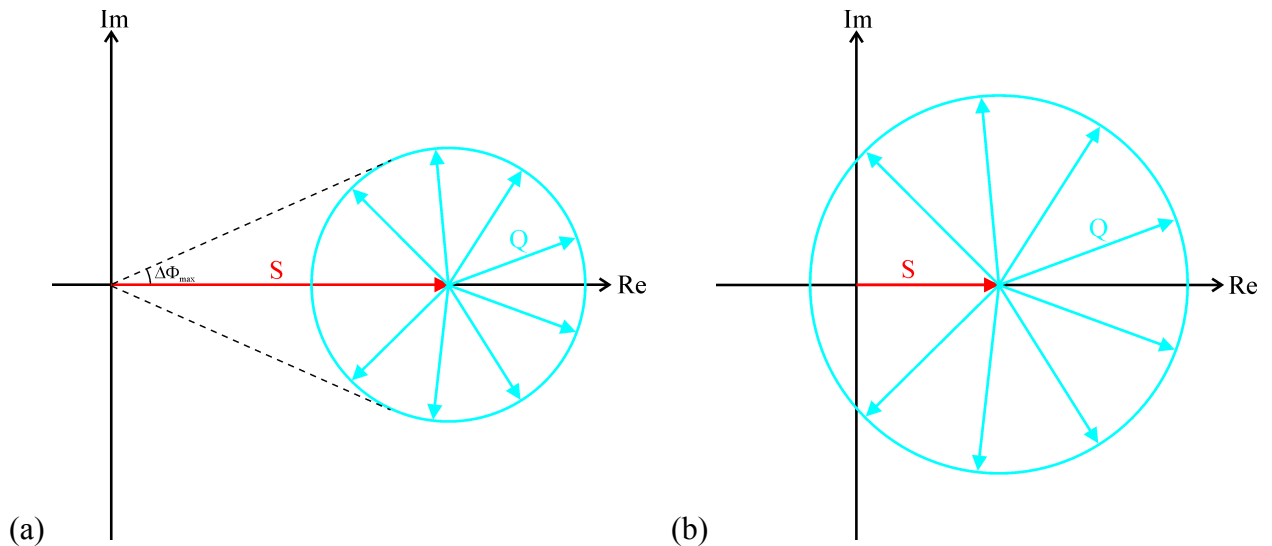


Figure 2.4

The strength of a triple-phase relation displayed graphically. (a) If $S = k(|\mathbf{E}_{\mathbf{h}}|^2 + |\mathbf{E}_{\mathbf{k}}|^2 + |\mathbf{E}_{-\mathbf{h}-\mathbf{k}}|^2 - 2)$ is large (k is a constant, S is a real number), then the resulting phase can be estimated reliably. (b) If S is small, then the component \mathbf{Q} , whose value is unpredictable if no prior information about the interatomic vectors is available, dominates and no phase estimate is possible. $\mathbf{E}_{\mathbf{h}}\mathbf{E}_{\mathbf{k}}\mathbf{E}_{-\mathbf{h}-\mathbf{k}} \approx S + \mathbf{Q}$.

However, usually several triple-phase relations can be formed for reflection \mathbf{h} and these can be combined to give a better estimate for $\varphi_{\mathbf{h}}$ (Karle & Hauptman, 1956):

$$\tan \varphi_{\mathbf{h}} = - \frac{\sum_{\mathbf{k}} |\mathbf{E}_{\mathbf{k}} \mathbf{E}_{-\mathbf{h}-\mathbf{k}}| \sin(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})}{\sum_{\mathbf{k}} |\mathbf{E}_{\mathbf{k}} \mathbf{E}_{-\mathbf{h}-\mathbf{k}}| \cos(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})} \quad (35).$$

This equation (the tangent formula) is the basis of most conventional direct phasing procedures. The general strategy is to select reflections with large $|\mathbf{E}_{\mathbf{h}}|$ values, assign random phases to them and iteratively determine phases in turn for each of the selected reflections until convergence. A figure-of-merit is then calculated for each trial and the process is repeated several times (100-5000). Standard figures-of-merit comprise a combination of several factors like negative quartet phase relations and R_{α} (the agreement between the observed and calculated reliability parameter of triple-phase relations). The best phase set is then selected, and an electron density map is calculated (with only the phased strong reflections or after a phase extension to weaker reflections) that is peak-searched. This procedure can routinely solve small-molecule structures up to 100 unique atoms, but tends to lose enantiomorph discrimination and drifts towards a pseudo-centrosymmetric solution as the number of atoms increase. A relatively complete dataset to at least 1.2 Å is a prerequisite.

Dual-space recycling

This modification of traditional direct phasing procedures can extend the limitation on the size of structures that can be solved by an order of magnitude, although the availability of high-resolution data is still a prerequisite (Sheldrick *et al.*, 2001). This procedure starts from random atoms and calculates a phase set from them. This phase set can then be refined by the tangent formula or by the minimal function:

$$R(\Phi) = \frac{\sum_{\mathbf{h}, \mathbf{k}} 2N^{-\frac{1}{2}} |\mathbf{E}_{\mathbf{h}} \mathbf{E}_{\mathbf{k}} \mathbf{E}_{-\mathbf{h}-\mathbf{k}}| \left[\cos \Phi_{\mathbf{h}, \mathbf{k}} - \frac{I_1(2N^{-\frac{1}{2}} |\mathbf{E}_{\mathbf{h}} \mathbf{E}_{\mathbf{k}} \mathbf{E}_{-\mathbf{h}-\mathbf{k}}|)}{I_0(2N^{-\frac{1}{2}} |\mathbf{E}_{\mathbf{h}} \mathbf{E}_{\mathbf{k}} \mathbf{E}_{-\mathbf{h}-\mathbf{k}}|)} \right]^2}{\sum_{\mathbf{h}, \mathbf{k}} 2N^{-\frac{1}{2}} |\mathbf{E}_{\mathbf{h}} \mathbf{E}_{\mathbf{k}} \mathbf{E}_{-\mathbf{h}-\mathbf{k}}|} \quad (36).$$

The resulting phases are used to calculate an electron density map, which is then peak-searched to obtain new atom positions. This list of atoms can be manipulated and the new atoms are used to calculate a new phase set. This procedure is repeated several times, and at the end a figure-of-merit is calculated, which is either the value of the minimal function or the correlation coefficient between $|\mathbf{E}_{\text{obs}}|$ and $|\mathbf{E}_{\text{calc}}|$ (Fujinaga and Read, 1987).

To increase efficiency, starting atoms can be selected so that they are consistent with the Patterson-function (Patterson seeding). In this procedure, a random peak is selected among strong Patterson-peaks; this potential two-atom vector is placed in the unit cell at random positions and Patterson deconvolution is performed with the most promising one. The resulting atom set is used as starting atoms. An increase in search efficiency can also be achieved by the application of the random omit technique in the real space cycle, in which about 30% of the peaks are discarded at random. If figures-of-merit indicate a potential solution, a few cycles of peaklist-optimisation can also be performed, which is a modification of the dual-space recycling that tests each atom found in the real-space cycle and eliminates it if an increase in the correlation coefficient can be observed. However, as this algorithm is computationally very expensive, this should only be performed for promising tries.

Dual-space recycling is a great advance over traditional direct methods, and the largest structure solved so far contains more than 2000 atoms. The availability of atomic resolution data is a requirement, although there are cases reported where lower resolution data proved to be sufficient for structure solution (Usón *et al.*, 1999). Experience suggests that the presence of heavier atoms significantly increases the success rate and thereby extends limitations on the size of the structure, and probably also on the resolution limit.

Molecular replacement

This method can be regarded as the opposite of direct methods since it does not require high resolution data, but the availability of a model structure that is similar to the unknown structure. The whole procedure is based on the Patterson-function:

$$P(\mathbf{u}) = \frac{1}{V} \sum_{\mathbf{h}} |\mathbf{F}_{\mathbf{h}}|^2 \cos(2\pi \mathbf{h} \mathbf{u}) \quad (37),$$

which can be calculated directly from the measured data (V is the volume of the unit cell). It can be demonstrated that the Patterson-function is a vector map, i.e. peaks occur in the Patterson-map at positions that correspond to interatomic vectors. As the number of peaks in the Patterson-function is n^2-n+1 (where n is the number of atoms in the unit cell), heavy overlap of peaks can be expected even for moderately sized structures.

Assume that a model structure homologous to the one in the asymmetric unit is available. In this case, the phase problem can be overcome by transferring the phases calculated from the model to the observed reflections. However, this would require accurately orienting and positioning the model in the unit cell, the content of which is unknown.

Thus, before phases could be obtained, the orientation and the position of the model have to be established (Rossmann & Arnold, 1993).

The rotation search

It is common practice to separate the inherently six-dimensional problem into two three-dimensional ones. This can be achieved by realising that Patterson peaks arising from interatomic vectors among atoms of one molecule (self-Patterson vectors) do not depend on the position of the molecule but only on its orientation. If the Patterson-function could be divided into peaks from self-Patterson vectors and cross-Patterson vectors (vectors between atoms of neighbouring molecules), the observed self-Patterson map and the one calculated from the model in the same orientation would be identical. Although in general such a separation cannot be achieved for the observed Patterson-function, a good estimation can be given by assuming that self-Patterson vectors are short, while cross-Patterson vectors are long. The rotation function, which is the overlap between the observed and calculated Patterson-functions is defined as

$$R(\mathbf{C}) = \int_U P_{\text{obs}}(\mathbf{u}) P_{\text{calc}}(\mathbf{Cu}) d\mathbf{u} \quad (38),$$

where \mathbf{C} is a rotation matrix and the integration is performed for volume U that contains predominantly self-Patterson vectors. This volume is usually chosen as a sphere with radius r that is about 70-80% of the maximum distance from the centre of mass. For a globular protein molecule this radius would include approximately the same percentage of all self-Patterson vectors and only a small percentage of cross-Patterson vectors that only increase the noise. Thus, by selecting the integration sphere too large, the noise from cross-Patterson vectors increases and would at some point conceal the solution. On the other hand, too small a radius would decrease the discrimination of solutions from noise peaks, so an optimum in r should be found. This requires a certain degree of experimenting with the parameter, especially for elongated molecules. However, if the Patterson-function is such that no separation of self-Patterson and cross-Patterson vectors can be achieved (e.g. tight packing), the procedure is likely to fail.

The translation search

When the orientation of the molecule has been established, the position can be deduced by moving the molecule on a suitable grid and calculating R_1 [equation (43)] or the correlation coefficient. The correlation coefficient is more reliable as it is independent of the scale factor

between calculated and measured intensities. An alternative approach is the use of the translation function

$$T(\mathbf{m}) = \int_V P_{\text{obs}}(\mathbf{u}) P_{\text{calc}}(\mathbf{u}, \mathbf{m}) d\mathbf{u} \quad (39),$$

where \mathbf{m} is the position vector of the centre-of-mass from the crystal origin and V is the volume of the asymmetric unit. In the space group P1 no translation search is needed for the first molecule, as the crystal origin can be chosen arbitrarily.

Six-dimensional search

Although computationally demanding, a six-dimensional search can also be performed (Kissinger *et al.*, 1999). This is likely to give better results when the above assumptions do not hold, i.e. for elongated molecules or when the solvent content is low, as it does not require the separation of self- and cross-Patterson peaks. Another condition under which this method may prove superior is the presence of non-crystallographic symmetry with high degree; in that case the rotation search is also likely to fail, since the observed Patterson-function is the overlap of several copies of the one calculated from the model and the maximum value of the rotation function is inversely correlated with the degree of non-crystallographic symmetry, and thus may not discriminate well from noise peaks.

MODEL BUILDING

After structure solution, approximate phases are available to calculate an electron density map. The distribution of electron density has to be interpreted and modelled by atoms or molecules that are arranged in a chemically sensible way. However, for an *ab initio* direct method structure solution, atomic positions are written out and they only have to be assigned to appropriate atoms. On the other hand, if the structure is solved with molecular replacement, the homologous structure used for solution can be taken as an initial model, and model building is reduced to correcting the differing parts.

Model building can be problematic if the starting phases are not good enough to obtain an interpretable electron density map (e.g. for isomorphous replacement and anomalous scattering techniques). In this case, density modification can be used. In this procedure, an electron density map is calculated from the observed structure factors and phases, and it is modified according to prior knowledge. The modified map is transformed back and new structure factor amplitudes and phases are obtained. The new phases are then combined with the old ones, while the amplitudes are discarded (although they may be used for calculating

weights in the phase recombination stage). Prior knowledge of the electron density is the flatness of the solvent region (solvent flattening), the magnitude distribution of electron density in the protein region (histogram matching) or the identity of electron density map in certain regions of the unit cell (NCS-averaging). Even in the case of molecular replacement, solvent flattening can prove useful, since it can reduce model bias inherently present in the method.

STRUCTURE REFINEMENT

When an interpretation of the electron density in form of an atomic model is available, this model is usually fitted to the measured data by suitable refinement procedures. The number of parameters to be determined depends on the resolution: for high resolution it is common to fit nine parameters for an atom (coordinates + anisotropic vibrations), for medium resolution four parameters (coordinates + isotropic vibration), while for low resolution structures it is usual to group the temperature factors and to refine atomic coordinates only as torsion angles, i.e. less than four parameters for an atom. A general formula for the calculation of structure factors can be given as (Cruickshank, 1970)

$$\mathbf{F}(\mathbf{S}) = \sum_i f_i^0(|\mathbf{S}|) q(\mathbf{S}) \exp(2\pi i \mathbf{r}_i \mathbf{S}) \quad (40),$$

where f_i^0 is the scattering factor of atom i not subject to vibration, \mathbf{r}_i is the position vector of atom i and $q(\mathbf{S})$ describes the effect of atomic vibrations. This latter is approximated by

$$q(\mathbf{S}) = \exp(-2\pi^2 U_i |\mathbf{S}|^2) \quad (41)$$

in the medium resolution case, where U_i is the mean square amplitude of vibration in any direction, while in the high-resolution case

$$q(\mathbf{S}) = \exp(-2\pi^2 \mathbf{S}^+ \mathbf{U}_i \mathbf{S}) \quad (42),$$

where \mathbf{U}_i is a symmetric tensor and \mathbf{S}^+ is the transpose of vector \mathbf{S} . The agreement of the model with the observed data is commonly described by the crystallographic R -factor

$$R_1 = \frac{\sum_{\mathbf{h}} \|\mathbf{F}_o(\mathbf{h}) - \mathbf{F}_c(\mathbf{h})\|}{\sum_{\mathbf{h}} |\mathbf{F}_o(\mathbf{h})|} \quad (43),$$

where the summation is performed usually for $|\mathbf{F}_o|$ -values above a certain threshold.

Least-squares minimisation

The best fit of the model to the observed intensities can be obtained by minimising the quantity

$$M_1 = \sum_{F > 4\sigma(F)} w_h \left(|\mathbf{F}_o(\mathbf{h})| - |\mathbf{F}_c(\mathbf{h})| \right)^2 \quad (44)$$

or

$$M_2 = \sum_h w_h \left(|\mathbf{F}_o(\mathbf{h})|^2 - |\mathbf{F}_c(\mathbf{h})|^2 \right)^2 \quad (45).$$

Equation (44) is usually employed in the structure refinement of macromolecules, while for small molecular refinement equation (45) is dominant. Advantage of (44) is the smaller number of data used in the procedure; the cutoff has to be used since for weak or negative reflections the estimation of $\sigma(|\mathbf{F}_h|)$ from the formula

$$\sigma(|\mathbf{F}|) = \frac{\sigma(|\mathbf{F}|^2)}{2|\mathbf{F}|} \quad (46)$$

is not valid any more, and the weights usually depend on the associated standard deviation [e.g. $w_h = 1/\sigma^2(|\mathbf{F}_h|)$]. On the other hand, $\sigma(|\mathbf{F}_h|^2)$ is an experimental quantity, and therefore no assumptions are involved in the derivation. Thus, using equation (45), all data can be used in the refinement, which is beneficial in pseudo-symmetric structures where low-intensity reflections prove to be important. In the latter case, weights are usually given in the form

$$w = \left[\sigma^2(|\mathbf{F}_o|^2) + (gP)^2 \right] \quad (47),$$

where $P = \frac{1}{3}|\mathbf{F}_o|^2 + \frac{2}{3}|\mathbf{F}_c|^2$, and the best value of g can be found automatically. The solution proceeds by evaluating first and second derivatives and using a full-matrix inversion or by conjugate gradient minimisation. The full-matrix approach has the advantage that standard uncertainties are also obtained (equal to diagonal elements of the inverse normal matrix), but the required storage and the CPU-time required for matrix inversion scale with the second and third power of the number of parameters, respectively. The conjugate gradient optimisation scales with the first power of the number of parameters, and is numerically more stable than matrix inversion.

Maximum likelihood refinement

The least-squares technique can be regarded as a special case of maximum likelihood. However, the assumptions under which maximum likelihood reduces to least squares are only

poorly justified in the case of refinement of atomic models for protein structures (Pannu & Read, 1996). These assumptions can be summarised as follows:

- (a) the deviation between $|\mathbf{F}_o|$ and $|\mathbf{F}_c|$ can be given as a Gaussian function.
- (b) the mean deviation is zero.
- (c) the standard deviation of the Gaussian is independent of the parameters of the atomic model.

The distribution of the true structure factor is best approximated by a Gaussian centred on $D\mathbf{F}_c$ (Read, 1990), where

$$D = \exp \left[-\pi^2 \langle \Delta \mathbf{r} \rangle^2 \left(\frac{\sin \Theta}{\lambda} \right)^2 \right] \quad (48)$$

and $\Delta \mathbf{r}$ is the coordinate error, i.e. the distribution is Gaussian in the complex plane. Therefore, the distribution $|\mathbf{F}_o| - |\mathbf{F}_c|$ will follow a Rice-distribution, and in the general case, especially in the initial stages, a maximum likelihood target is more appropriate in refinement. However, the procedure involves a huge increase in computational demand or many approximations to make it more efficient.

Restrained refinement

In macromolecular refinement the number of observations is comparable to that of parameters. However, this ratio can be improved by introducing *a priori* knowledge in form of restraints. A restraint is treated as an additional "observation" with a standard deviation and is included in the target function that gets minimised. Common restraints include:

- (a) bond length restraints. Distortion of bond lengths requires considerable energy and the resolution is often not good enough to resolve atoms, so these can be restrained to have values determined using high-resolution small-molecular structures with small standard deviation.
- (b) bond angle restraints. In this case the energy demand of distortion is significantly less, but it may still be justified to restrain these values, but with a higher standard deviation.
- (c) planarity restraints. These are usually employed for chemical moieties that are believed to be planar under normal circumstances.

- (d) anti-bumping. The assumption that two non-bonded atoms should not be nearer than the sum of their van der Waals distances is usually valid. This restraint is special from the point that in one direction the deviation should not be penalised.
- (e) anisotropic displacement parameter restraints. Anisotropic temperature factor refinement increases the number of parameters considerably, so the inclusion of these restraints is often necessary.

Even when high-resolution data are available, for a protein structure the use of restraints is recommended.

Common problems

In protein crystallography the main problem is usually constituted by the low data-to-parameter ratio in structure refinement, and the tendency to overfit the data by using too many parameters, which in compensation for other factors assume physically unrealistic values. To determine whether the addition of a set of parameters results in a better description of electron density, a given percentage of the reflections is usually excluded from the refinement and used for the calculation of R_{free} (Brünger, 1992). This is defined in an analogous way to R_1 [equation (43)], but the summation goes for reflections not used in the refinement. R_{free} reflections are selected randomly or in thin shells when non-crystallographic symmetry is present. Although this is a good global guide for describing the course of the refinement, it is not sensitive to small changes, e.g. modelling discrete disorder for a sidechain.

Another important issue is the description of disordered solvent always present in protein crystals (Moews & Kretsinger, 1975). There is indication that Babinet's principle that is widely used in solvent modelling is not satisfactory and especially for crystals with high solvent content, the value of standard agreement indicators is often too high.

III. CRYSTAL STRUCTURES OF CEPHAIBOLS

INTRODUCTION

Cephaibols comprise a group of peptaibol antibiotics and have been isolated from the soil fungus *Acremonium tubakii*, DSM 127741. Apart from their considerable antibacterial potency, it has been established that cephaibols possess pronounced anthelmintic action and activity against ectoparasites (Schiell *et al.*, 2001). Other surprising biological properties, like induction of pigment formation in *Phoma destructiva* and potential neuroleptic effects have also been reported (Ritzau *et al.*, 1997; Schlegel *et al.*, 2002).

The amino acid sequence of cephaibols shows marginal variation among group members (Figure 3.1), the main difference being some methyl groups at positions 5, 8 and 12. However, as there is at least a 10-fold variation in the antibacterial potency, the presence or absence of these methyl groups seems to have a substantial effect upon antimicrobial action. Unlike other types of antibiotics, peptaibols frequently occur naturally as microheterogeneous mixtures; this peculiarity has long been established and attributed to their non-ribosomal

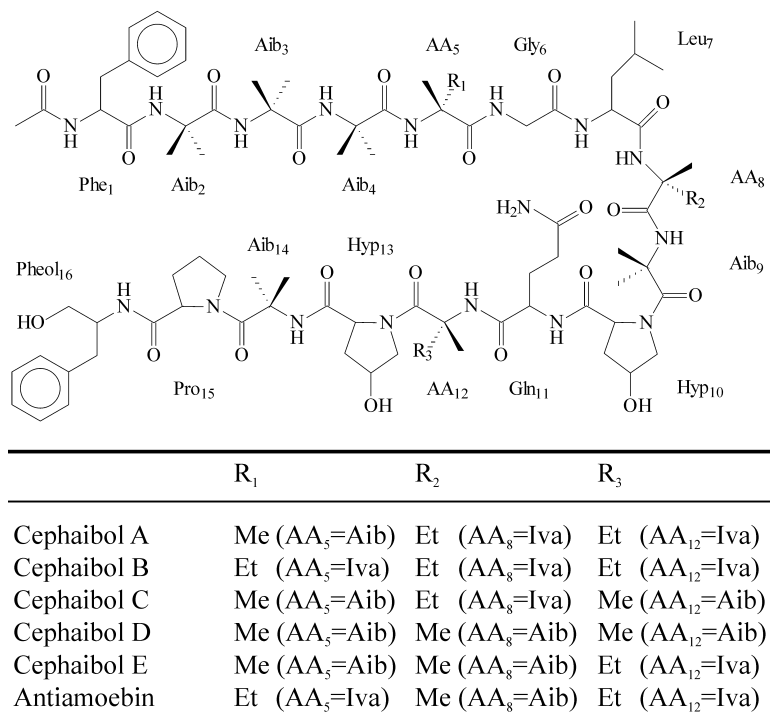


Figure 3.1

Amino acid sequence of members of the cephaibol series and comparison with antiamoebin.

peptide synthesis, in which Aib is frequently replaced by other α,α -dialkylated amino acids (Leclerc *et al.*, 1998).

Three-dimensional structures of various peptaibols have been determined by X-ray diffraction for alamethicin (Fox & Richards, 1982), [Leu1]zervamicin (Karle *et al.*, 1991), antiamoebin (Karle *et al.*, 1998; Snook *et al.*, 1998), trichotoxin (Chugh *et al.*, 2002) and ampullosporin (Kronen *et al.*, 2003). The NMR solution structures of chrysospermin C (Anders *et al.*, 2000), antiamoebin (Galbraith *et al.*, 2003) and zervamicin IIb (Balashova *et al.*, 2000; Shenkarev *et al.*, 2002) have also been reported. These molecules seem to adopt a common structural motif that consists of a long helical section that has a variable bend at a proline or a hydroxyproline residue near to the middle of the helix. Since the action of peptaibols is believed to derive from their interaction with biological phospholipid membranes, in which the formation of ion channels results in increased ion permeability (Menestrina *et al.*, 1986), the length should be an important factor in determining the potency. Alamethicin with its 20 residues and moderate bending angle (about 30°) is more than 33 Å long and therefore able to span biological membranes. Conductivity measurements with several artificial and naturally occurring membrane systems did show the formation of highly voltage sensitive and weakly cation selective membrane channels (Gordon & Haydon, 1972; Hanke & Boheim, 1980; Latorre & Alvarez, 1981). The 18-residue trichotoxin and the 19-residue chrysospermin C possess similar structures, although they are shorter and their bending angles are markedly different (10° and 38° , respectively). [Leu1]zervamicin consists of only 16 residues and its bending angle varies among the existing structures in the range 30 – 45° . Consequently, it is significantly shorter than alamethicin, only 29 Å long, but conductivity measurements still prove the formation of membrane channels (Balaram *et al.*, 1992). Moreover, it was found that in all crystal forms [Leu1]zervamicin molecules aggregated in a similar fashion to form water channels and suggested a gating mechanism for cation transport. The structure of antiamoebin has been determined independently in methanol (Snook *et al.*, 1998) and in a partially membrane-mimetic environment (Karle *et al.*, 1998). Although different crystal forms were obtained, the peptide conformation and even the molecular packing show strong similarity between the two structures. The 16-residue peptide, in comparison with [Leu1]zervamicin, possesses a much sharper bend nearly 55° , but only marginally shorter than [Leu1]zervamicin. Initial membrane conductivity measurements failed to give evidence of membrane channel formation, and although later largely voltage-insensitive channel formation was observed in specific membrane media (Duclohier *et al.*,

1998), a hypothesis for an ion carrier mechanism was put forward. Ampullosporin consists of only 15 residues and a bending angle of about 60° and therefore with a length of 20 Å it is the shortest of all peptaibols with known three-dimensional structure.

Cephaibols, although similar to antiamoebin, were available as pure compounds that have marginal structural difference but wide variation in antibacterial potency and provided an opportunity to investigate the effect of small structural changes on backbone conformation and association properties.

STRUCTURE DETERMINATION

Crystallisation

Quantities of cephaibols were ranging from 200 mg (cephaibol A) to 5 mg (cephaibol D and E). Although crystallisation procedures for several peptaibol antibiotics had been available (Fox & Richards, 1982; Karle *et al.*, 1991; Karle *et al.*, 1998; Snook *et al.*, 1998), these were not feasible because of either the high amount of antibiotic required (e.g. antibiotic concentration up to 200 mg/ml) or the danger of sample decomposition in a step (e.g. boiling). To render the crystallisation of all cephaibol samples possible, the vapour diffusion method was selected as the only one compatible with the limited availability of certain members and the expected high proportion of organic solvents. Initially, the published conditions suitably modified to hanging drop technique were tested, but no diffraction quality crystals were obtained. As the next step, screening was performed with cephaibol A, which was available in the largest amount. It was expected that a well-established crystallisation condition might also function for all of the different samples and demand only certain degree of optimisation, and screening was therefore conducted with adequate caution. Due to the low solubility of the peptide in water, dedicated screens with high organic solvent content available from Jena Bioscience were employed. Crystals were obtained from several conditions, but they had the same morphology, and for that reason, the condition yielding nearly diffraction quality crystals was selected and optimised. Eventually, high-quality crystals diffracting to atomic resolution were obtained by mixing 4 µl of 5 mg/ml cephaibol A in ethanol:water=1:1 with 2 µl of 0.1 M NaAc/HAc pH=4.6 and 35% ethanol. Cephaibol B crystals were obtained from slightly different conditions (33% ethanol vs. 35 % ethanol). Cephaibol C required comparatively extensive alteration and was grown by mixing 2 µl of 40 mg/ml cephaibol C in ethanol:water=1:1 with 2 µl of 0.1 M NaAc/HAc pH=4.6 and 37% ethanol. Initial crystals

from cephaibol E were obtained using a procedure similar to cephaibol C. Cephaibol D gave spherulites from these conditions, but microcrystals could be obtained by mixing equal volumes of ethyleneglycol:ethanol=55:45 with 20 mg/ml cephaibol D in ethanol:water=1:1. Crystal growth took from several days (cephaibol A, B and C) to several months (cephaibol D and E). Crystal dimensions varied from 0.7·0.4·0.1 mm³ (for cephaibol A, B and C) down to 0.1·0.05·0.01 mm³ (cephaibol D and E).

Measurement

Before measurement, a suitable crystal was selected and soaked in a cryoprotectant solution consisting of the reservoir solution and 20% ethyleneglycol, mounted with an adequately sized loop and shock-frozen in a cold nitrogen stream. Diffraction images were collected on a Bruker rotating anode with Osmic focusing mirrors and a SMART6000 4K CCD detector. Built-in functionalities of the program Proteum were used for indexing and cell refinement; integration was performed with SAINT and scaling with SADABS. XPREP was employed for space group determination and for the calculation of statistics (Bruker Nonius, 2002). Unit cell parameters and intensity statistics are shown in Table 3.1.

	Cephaibol A	Cephaibol B	Cephaibol C	Cephaibol D	Cephaibol E
Space group	P2 ₁ 2 ₁ 2	P2 ₁	P2 ₁ 2 ₁ 2 ₁		
Unit cell parameters	a=30.534 Å b=37.787 Å c=9.115 Å	a=32.146 Å b=9.126 Å c=37.982 Å β=111.36°	a=9.002 Å b=28.619 Å c=41.100 Å	a=9.275 Å b=29.590 Å c=39.724 Å α=68.80° β=83.30° γ=80.98°	a=9.250 Å b=29.250 Å c=39.381 Å α=69.03° β=85.45° γ=81.65°
No. of reflections	67303	73436	76788		
No. of unique reflections	7025	16045	8776		
Resolution (Å)	0.95 (1.05-0.95)	0.89 (1.00-0.89)	0.89 (1.00-0.89)		
Completeness	98.6% (96.6%)	97.0% (90.0%)	99.3% (97.6%)		
Redundancy	9.45 (4.66)	4.44 (2.16)	8.69 (4.36)		
<I/σ(I)>	29.21 (10.67)	23.29 (10.37)	50.44 (26.51)		
R _{int}	0.0431 (0.1156)	0.0437 (0.0786)	0.0306 (0.0349)		

Table 3.1

Data collection statistics. For cephaibol D and E the primitive unit cells are shown. For cephaibol D, this cell can be converted to a C-centred monoclinic cell with cell dimensions a=58.448 Å, b=9.275 Å, c=40.041 Å, β=112.64°. Values in parentheses refer to the last resolution shell.

$$R_{\text{int}} = \sum |I - \langle I \rangle| / \sum I.$$

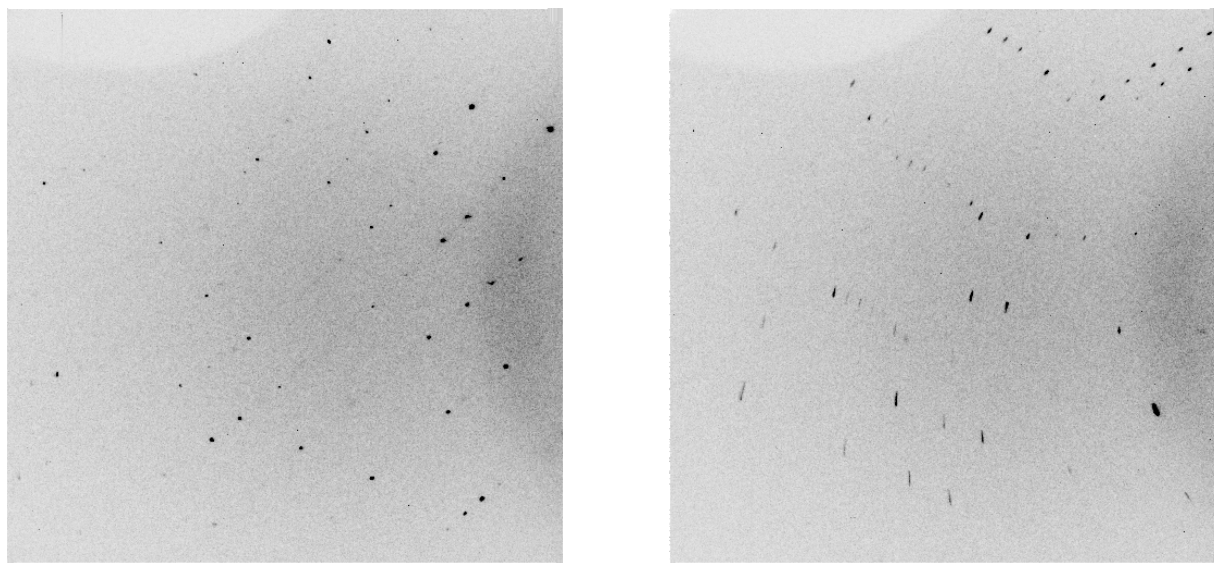


Figure 3.2

Diffraction images of cephaibol B showing high anisotropic mosaicity. However, as both the R_{int} and the refinement R -factors are similar to those of cephaibol A and C, it can be regarded that data processing software can handle such property when data are collected in an appropriate way. These images were taken at $2\Theta=80^\circ$ and with an ω -difference of approximately 90° . The shadow in the upper left corner is caused by the nozzle of the cryostream.

It was astonishing to realise that despite the high degree of structural homology and nearly identical crystallisation conditions, the crystals were not isomorphous and differed in unit cell parameters as well as in space group. Crystallising homologous samples in a similar way therefore does not necessarily yield the same crystal form, and crystal packing may be extremely sensitive to small structural differences even if they are not involved in crystal contacts. This observation also implies that as tiny difference as a single atom may prevent a sample from crystallisation, but on the other hand, puts forward the theory that homologous substances are likely to crystallise under similar conditions and justifies the development of crystallisation screens specialised on protein classes.

Full datasets were collected for cephaibol A, B and C and processed without any major difficulty. The crystals gave strong diffraction and the only resolution-limiting condition proved to be the maximum detector 2Θ -angle. Cephaibol B exhibited significant anisotropy of mosaicity, but this property was routinely handled by SAINT (Figure 3.2) and its effect is present neither in data quality nor in structure refinement. Frames for indexing were collected for cephaibol E, but as the diffraction limit was only 1.8 \AA and the crystal was non-merohedrally twinned, no dataset was collected. It is worth noting that cephaibol E had a

markedly different cell from cephaibol C in spite of the fact that cephaibol E could integrate into the crystal lattice of cephaibol C and was therefore expected to yield the same crystal form. Cephaibol D yielded crystals comparable in size and diffraction limit to those of cephaibol E, although the crystal characterised was proved to be single. It is noteworthy that the primitive cells of cephaibol D and E are very similar. The metric symmetry for cephaibol D was determined as C-centred monoclinic, but it is not known whether the Laue-symmetry corresponds to that or not. Cell dimensions of cephaibol E also allow conversion to a C-centred monoclinic cell, but the distortion in angles is significant; however, this may be a consequence of the reduced precision in cell determination due to non-merohedral twinning.

Structure solution

The asymmetric unit of cephaibols was sufficiently small and the resolution was sufficiently high to obtain structure solution without complication. SHELXD (Sheldrick *et al.*, 2001) was employed to find the peptide atoms (FIND 120 for cephaibol A and C, which have one molecule in the asymmetric unit and FIND 240 for cephaibol B, where there are two molecules in the asymmetric unit), and the solvent molecules (three cycles of peaklist optimisation by allowing 10-15% more atoms than expected from the structural formula). Cephaibol A had the highest success rate (9 out of 300 tries), followed by cephaibol B (5 out of 500) and then cephaibol C (3 out of 400). One try took about 70 s for cephaibol A and B and about 25 s for cephaibol C on a Pentium III 600 MHz processor. The solution rate was therefore roughly equal, although the asymmetric unit of cephaibol B is nearly twice as large.

Model building and refinement

After successful solution, several cycles of temperature factor refinement were performed using the program SHELXL (Sheldrick & Schneider, 1997) and peaks in the electron density have been assigned to corresponding atoms with the program XP. When virtually all peptide atoms had been found, the refinement was carried on with SHELXL using suitable bond length, bond angle, chiral volume and planarity restraints. Throughout the refinement, $2mF_o - DF_c$ and $F_o - F_c$ type maps were displayed with the program XtalView (McRee, 1999), which was also used for manual editing of the structure and identifying disorder components. Water molecules were added with the program SHELXWAT and by hand. All non-hydrogen atoms were refined anisotropically with rigid bond, similarity, and

for solvent molecules, approximately isotropic restraints; hydrogen atoms were included in later stages of refinement. Refinement details are shown in Table 3.2.

	Cephaibol A	Cephaibol B	Cephaibol C
Resolution range (Å)	37.79-0.95	35.37-0.89	28.62-0.89
$R_1[F > 4\sigma(F)]$ /all data]	0.0835/0.0872	0.0709/0.0746	0.0712/0.0719
$R_{\text{free}} [F > 4\sigma(F)]$ /all data]	0.1082/0.1120	0.0756/0.0793	0.0758/0.0770
No. of parameters	1429	3121	1294
No. of restraints	2016	4633	1519
No. of peptide atoms	119	240	118
No. of non-hydrogen atoms	133	268	135
Solvent content	21.9%	20.0%	23.0%
R.m.s.d. from ideal geometry			
Bond length (Å)	0.017	0.021	0.018
Angle distances (Å)	0.035	0.040	0.028
Common planes (Å)	0.416	0.505	0.499

Table 3.2

Model refinement statistics.

STRUCTURE DESCRIPTION

Amino acid sequence

The overall structures of the molecules were found to be very similar. All three cephaibols assume a helical conformation that is sharply bent at Hyp10. The bending angle is nearly 55° and seems not to vary much. Before the structure determination it was unclear whether the peptides contained D- or L-isovaline, as the chemical structures had been established with NMR-spectroscopy and mass spectrometry. The maps unambiguously showed that all isovalines are present as D-isovaline (assuming that the common amino acids are present as L-enantiomers).

Backbone conformation

Cephaibol A

Cephaibol A crystallises in the space group $P2_12_12$ with one molecule in the asymmetric unit. The whole molecule is well defined apart from two residues (Phe16 and Pro15) that each exhibit two distinct conformations. This 1:1 disorder implies the flipping of the proline ring into the other envelope conformation and the consequent motion of Phe16. Apart from this fact, the structure shows considerable homology to the previously determined peptaibol structures in terms of backbone conformation, secondary structure and intramolecular hydrogen bonding pattern. Accordingly, in spite of the high Aib content, residues 1 to 9 form an α -helix, residues 10-12 a short 3_{10} -helix, while the conformation of residues 12-16 should be considered as two overlapping β -turns. The bend of the helix occurs at Hyp10, which is identical to that in antiamoebin and [Leu1]zervamicin, and the bending angle is comparable to that of antiamoebin.

Cephaibol B

Cephaibol B crystallises in the monoclinic space group $P2_1$, but with very similar cell parameters to cephaibol A and therefore has two molecules in the asymmetric unit. Although the overall structures of the two molecules are similar, different regions were found to be discretely disordered. In the first molecule (molecule I), Hyp10, where the molecule is bent, shows a 9:1 disorder for the OH group being in axial and equatorial position, respectively. The backbone of molecule I from Iva12 to Phe16 inclusive also adopts two different conformations. This disorder may arise from the interaction with the second molecule (molecule II), in which the interacting region (from Ac0 to Aib3) is also disordered, allowing different hydrogen bonds between the two conformations. It is also noteworthy that in molecule II the OH group of Phe16 adopts two conformations, but the sum of occupancies turns out to be slightly higher than unity. This fact seems to be accounted for by mass spectrometric evidence that together with cephaibol B there is another compound present that has a sequence identical to that of the principal constituent, but with an aldehyde function instead of the alcohol group at Phe16 (Schiell *et al.*, 2001). As a geminal diol, an aldehyde group would convert into two OH groups, and increase the occupancy for both disorder components. It should be emphasised that according to the electron density map, only a minority of the molecules ends in phenylalaninal and the rest have phenylalaninol with a disordered OH as C-terminus. However, the presence of the aldehyde seems to be significant.

Cephaibol C

Cephaibol C crystallises in the orthorhombic system with cell edges very similar to those of cephaibol A, but in the space group $P2_12_12_1$. This difference in symmetry involves different interactions amongst the molecules and consequently different packing. The structure and packing of cephaibol C was found to be very similar to that of [Leu1]zervamicin, which crystallises in the space group $P2_1$, and contains no disorder. The electron density map showed the presence of some cephaibol E contamination, which manifested itself as Aib12 being partially Iva. This contamination originates from the retention times of cephaibol C and E being nearly identical, making these two components very difficult to separate (Schiell *et al.*, 2001).

Comparison of the structures

To search for rigid fragments, the program ESCET (Schneider, 2002) was employed with coordinate estimated standard deviations generated from temperature factors using Cruickshank's DPI equation (Cruickshank, 1999) plus linear B-factor scaling. In total four different molecules were compared simultaneously and in a pairwise manner (side chain atoms were not included). On average, the molecules are about 20% identical within the experimental error (this number is a slight underestimation as the precision of the structures is relatively high), but in special cases the similarity reaches higher levels. Cephaibol A and C were found to share about 60% of identical structure, while cephaibol A also has about 60% in common with one of the cephaibol B molecules. Interestingly, the two cephaibol B molecules were found to be less similar, showing only 30% identity.

To identify any apparent motion, rigid atoms found by ESCET were fitted together with the program LSQKAB (Kabsch, 1976), and the whole molecules were transformed using the matrix resulting from this superposition. The molecules were then compared using computer graphics. In the case of cephaibol A and cephaibol C, the two structures fit very well onto each other from their N-terminus till position 12, where the sequence differs (Figure 3.3a). For cephaibol A and cephaibol B, the centres assume nearly identical conformation, only the termini seem to be flexible, and the sequence difference at position 5 seems not to be responsible for the conformational change (Figure 3.3b). Simultaneous superposition of all models reveals that despite the low structural identity nearly the whole N-terminal helix is rigid, while the C-terminus exhibits conformational diversity (Figure 3.3c). This finding can be confirmed by the analysis of difference-distance matrices that give no indication of a rigid,

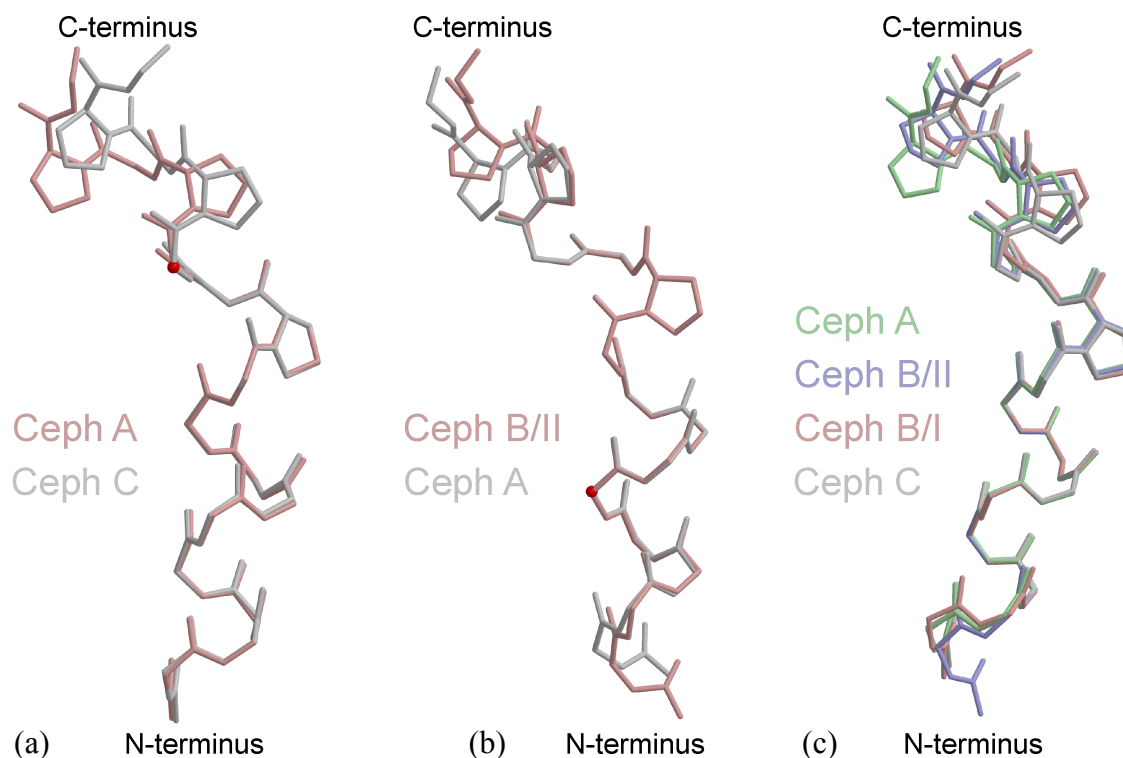


Figure 3.3

Conformation differences among cephaibol A, B and C. (a) More than half of cephaibol A is virtually identical with cephaibol C and the deviation starts at position 12 (marked with the red ball), where the sequence differs. (b) Cephaibol B fits very well on cephaibol A and position 5 (the difference in sequence) seems not to be responsible for the different conformations. (c) Simultaneous superposition of cephaibols reveals a rigid N-terminal helix, which includes two of three positions where the sequence may differ.

but smaller sized C-terminal helix and a flexible hinge region. The observed extensive disorder of C-termini also suggests that this region can assume multiple conformations.

Crystal packing

In the case of cephaibol C, nearly the same arrangement was found as for antiamoebin and [Leu1]zervamicin, i.e. two V-shaped molecules that are related by a translation along the crystallographic **a** axis face a third one and form a channel that looks like an X when viewed from the side (Figure 3.4). Strong intermolecular hydrogen bonds connect the three molecules, especially between $\text{O}(\text{Gly6})\text{--}\text{N}^{\epsilon}(\text{Gln11})'$, $\text{O}^{\delta}(\text{Hyp10})\text{--}\text{N}^{\epsilon}(\text{Gln11})'$, $\text{O}^{\delta}(\text{Hyp10})\text{--}\text{O}(\text{Leu7})'$, $\text{O}(\text{Leu7})\text{--}\text{O}^{\delta}(\text{Hyp10})''$, $\text{N}^{\epsilon}(\text{Gln11})\text{--}\text{O}^{\delta}(\text{Hyp10})''$ and $\text{N}^{\epsilon}(\text{Gln11})\text{--}\text{O}(\text{Gly6})''$ (the apostrophes ' and '' denote different symmetry equivalent molecules). This arrangement is believed to be representative of the functional membrane channel (Karle *et al.*, 1991).

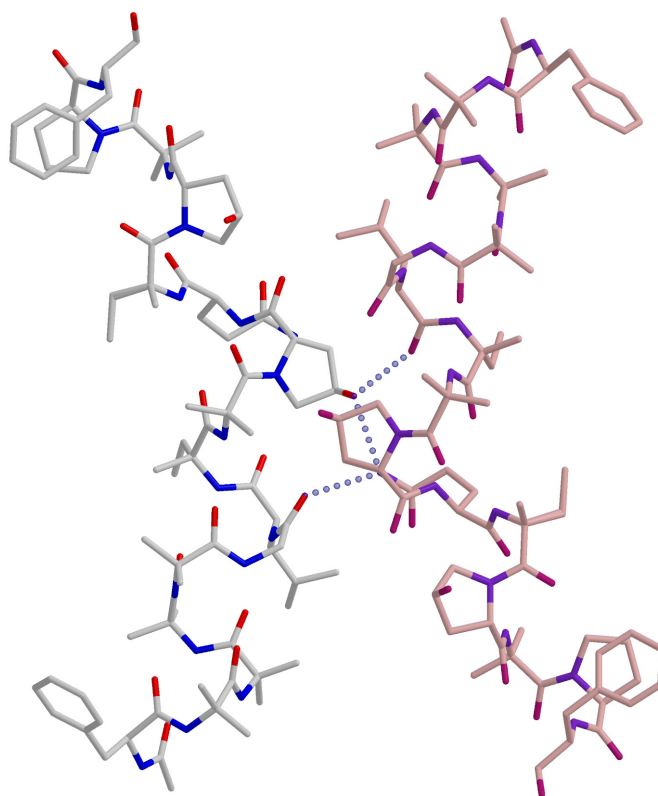


Figure 3.4

Three cephaibol C molecules adopt a channel-like arrangement (the left molecule represents the asymmetric unit, the right molecule is a symmetry equivalent that is designated by one apostrophe in the text and the third molecule, which is directly behind the right one, is designated by two apostrophes; hydrogen bonds that involve the third molecule are not shown).

Hydrogen bonds between $\text{N}(\text{Phe1})\text{--O}(\text{Pro15})''$, $\text{N}(\text{Aib2})\text{--OH}(\text{Pheol16})''$ and $\text{O}^\delta(\text{Hyp13})\text{--O}^\epsilon(\text{Gln11})''$ connect to other symmetry related molecules and seem to be important in stabilising the crystal lattice.

Surprisingly, cephaibol A and cephaibol B adopt different packing motifs to cephaibol C. The main hydrogen bonds connect two termini and the centre of one molecule to the terminus of another. The molecules form a layer-like zigzag structure (Figure 3.5). There is no sign of any channel formation similar to that observed in the case of cephaibol C, and Hyp10, which was involved in most of the intermolecular interactions, is now pointing towards a water-filled cavity. In the case of cephaibol A the following hydrogen bonds are observed between the molecules: $\text{N}(\text{Phe1})\text{--OH}(\text{Pheol16A})'$ is the head-to-tail interaction, where A refers to one of two distinct conformations with higher occupancy, while the

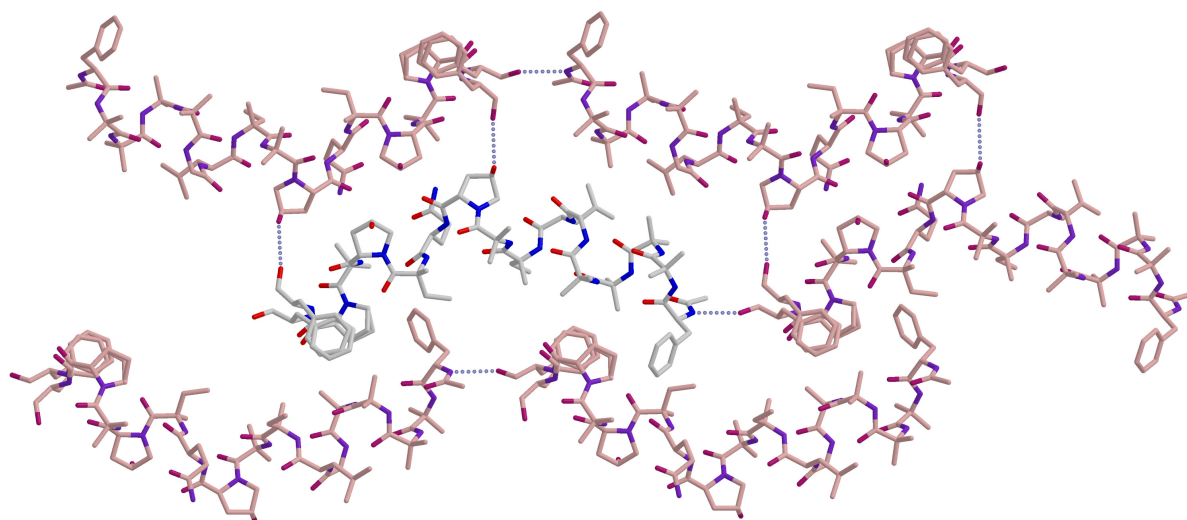


Figure 3.5

Cephaibol A and B (cephaibol A is shown) form a zigzag structure that is not indicative of membrane channel formation. Hydrogen bonds involving Gln11 are omitted, because they would involve molecules that are behind the layer that is shown.

$O^{\delta}(\text{Hyp10})-\text{OH}(\text{Pheol16B})''$, $N^{\epsilon}(\text{Gln11})-O^{\delta}(\text{Hyp13})'''$ and $O^{\epsilon}(\text{Gln11})-O^{\delta}(\text{Hyp13})'''$ interactions connect the centre of the molecule with the C-termini of three molecules. This pattern also accounts for the disorder observed at the C-terminus, since $\text{OH}(\text{Pheol16})$ is able to participate in hydrogen bonds in both conformations, although the first is marginally more favourable. The two molecules in the cephaibol B crystal, although having the same packing arrangement as cephaibol A, interact at the termini by means of more hydrogen bonds: $N(\text{Phe1/I})-O(\text{Pro15/II})'$, $N(\text{Phe1/I})-\text{OH}(\text{Pheol16B/II})'$, $N(\text{Aib2/I})-\text{OH}(\text{Pheol16B/II})'$ and $N(\text{Aib2/I})-\text{OH}(\text{Pheol16A/II})'$ at the N-terminus and $O(\text{Pro15A/I})-N(\text{Phe1A/II})$, $\text{OH}(\text{Pheol16A/I})-N(\text{Phe1A/II})$, $\text{OH}(\text{Pheol16A/I})-N(\text{Aib2A/II})$, $\text{OH}(\text{Pheol16B/I})-N(\text{Phe1B/II})$ and $\text{OH}(\text{Pheol16B/I})-N(\text{Aib2B/II})$ at the C-terminus (/I and /II refers to molecule I and II, respectively). This higher number of hydrogen bonds seems to be a consequence of the increased flexibility at the termini that enables them to interact in various ways. Both disorder components take part in roughly the same number of hydrogen bonds and therefore have nearly equal occupancies. However, there is a reduction in the number of intermolecular hydrogen bonds involving the centre; only the $N^{\epsilon}(\text{Gln11/I})-O^{\delta}(\text{Hyp13B/II})''$ interaction can be detected.

BIOLOGICAL IMPLICATIONS

Membrane channels

Several water and other solvent (mostly ethanol) molecules have been located during the refinement process. For cephaibol A and B, the solvent molecules seem to be equally distributed along the peptides, although some concentration is observed near the bend in the helix. On the contrary, for cephaibol C the solvent molecules seem to occupy exclusively the cavity between the two legs of the X. This is in accordance with the [Leu1]zervamicin structure with the difference that the cephaibol C helix is not amphipatic, and suggests that even this channel would allow the transport of water molecules and ions when immersed in a biological membrane.

Sequence variation and association properties

Although closely related in chemical structure, the potency of cephaibol antibiotics shows considerable variation among the family members. Nevertheless, the assumption that variations in antibacterial action arises from significant differences in the three-dimensional structure can be rejected since their structures are very similar even at high resolution. However, comparison of intermolecular interactions reveals large differences in association properties. It seems that as the molecules become more flexible i.e. the amount of disorder in the peptide backbone increases, the number of interactions involving the bend region decreases, so that a new packing arrangement appears when changing from cephaibol C to cephaibol A. For cephaibol B, the centre takes part in only a handful of interactions, while the interactions involving the termini get stronger. It can be assumed that membrane-channel-forming potency decreases in this order. It is interesting to note that the activity of cephaibols increase in the same order. Since cephaibols (as all peptaibol antibiotics) occur naturally as microheterogeneous mixtures, intermolecular interactions involving different members could lead to a large number of associates with different properties and account for their wide variety of action. In the case of gramicidin, where microheterogeneity also occurs, it was concluded that a small proportion of slightly different molecules might promote membrane integration and conversion into the conducting form and microheterogeneity was therefore not an undesired by-product of insufficient enzyme specificity, but had important biological role (Burkhart, Gassman *et al.*, 1998), which might be the reason for microheterogeneity of peptaibols as well.

Implications for mechanism of action

The mechanism of action of peptaibol antibiotics has not been unambiguously identified yet, but it seems to be widely accepted that antibiotic molecules constitute channels and thereby increase the permeability of biological membranes. Although conductivity measurements consistently prove the formation of ion channels, alternative techniques give contradicting results. For example, Aib residues in trichogin GA IV were substituted with a nitroxide spin label and the extent of insertion was assessed by the burial of the spin label using ESR and fluorescence quenching techniques (Peggion *et al.*, 2003; Milov *et al.*, 2003). Both approaches led to the conclusion that the helical axis was oriented in the plane of the bilayer and therefore no membrane channel formation was observed. More recently, using fluorescence and light-scattering measurements, the same authors concluded that trichogin was most likely to constitute pores in biological membranes (Stella *et al.*, 2004).

The crystal structures of cephaibols suggest two possible action mechanisms. Cephaibol C is likely to constitute membrane channels, while cephaibol A and B probably act as ion carriers. A crystal structure in a lipid bilayer (e.g. cubic lipid phase, bicelles) may prove useful in establishing the mode of interaction with the membrane, as well as structure determination of cephaibol D and E that would complete the systematic investigation of the cephaibol series.

It seems therefore likely that peptaibols exhibit a variety of interactions with the cell membrane. In addition, their biological action is not limited to antibacterial potency, as peptaibols show other innovative phenomena like the induction of hypothermia and potential neuroleptic effects. The investigation of this secondary metabolite class therefore cannot be regarded as near to completion. These peptides seem to be occurring in many more species than previously thought, and it can be assumed that they have important biological functions (Degenkolb *et al.*, 2003).

IV. CRYSTAL STRUCTURE OF FEGLYMYCIN

INTRODUCTION

Feglymycin is a naturally occurring peptide isolated from *Streptomyces* sp DSM 11171. The chemical structure of the peptide was elucidated using NMR-spectroscopy and mass spectrometry. Feglymycin was found to strongly inhibit the formation of HIV syncytia *in vitro*, but the molecular mechanism of antiviral action has not been investigated yet. A weak antibacterial activity against Gram-positive bacteria was also reported (Vértesy *et al.*, 1999).

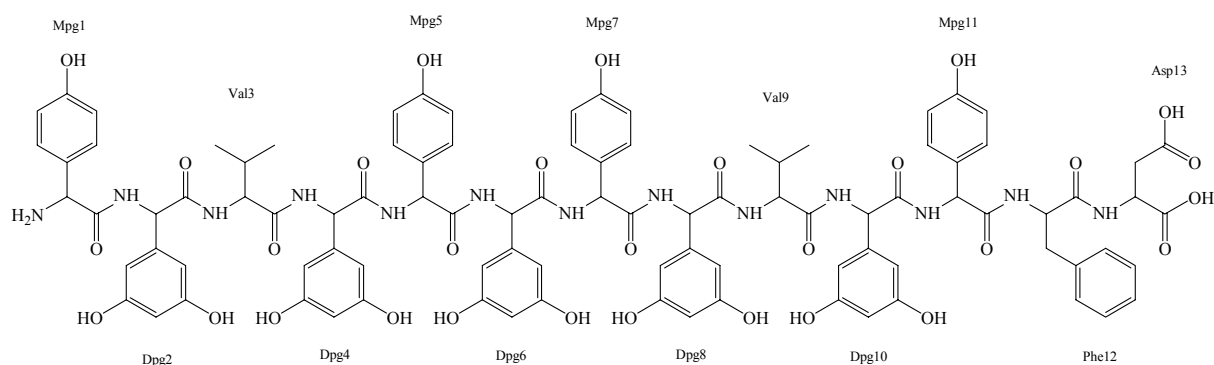
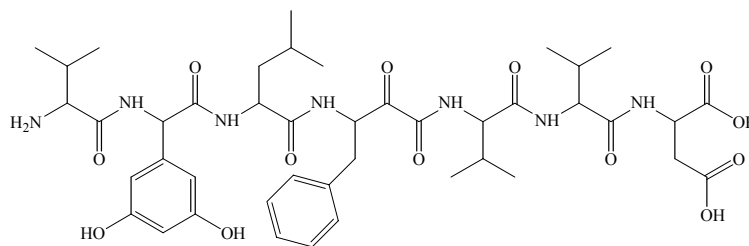


Figure 4.1

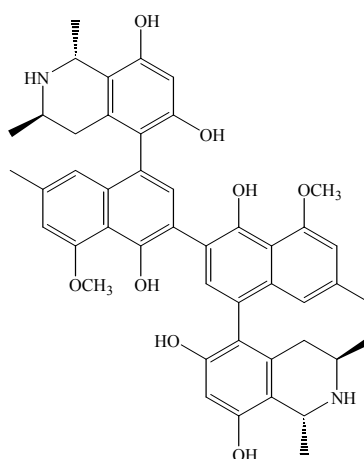
Chemical structure of feglymycin.

Feglymycin contains a high percentage of unusual amino acids: out of the 13 residues in the sequence, nine are either 4-hydroxyphenylglycine or 3,5-dihydroxyphenylglycine (Figure 4.1). Based on its chemical structure and biological activity, feglymycin constitutes a new group among antibiotics. Structural relatives include complestatin (Kaneko *et al.*, 1989), the complestatin-related chloropeptins (Tanaka *et al.*, 1997) and certain retrovirus protease inhibitors (RPI-856 A, B, C and D; Asano *et al.*, 1994). These substances are all peptides that vary in length and contain at least one of the unusual hydroxyphenylglycine residues (Figure 4.2). Additional similar properties are that they are all produced by diverse *Streptomyces* sp and inhibit various steps of the replication of HIV. Complestatin and chloropeptins were found to prevent the binding of the *gp120* viral protein to CD4 cellular receptor, while RPI-856 A, B, C and D inhibit the HIV protease with varying affinity. Moreover, the hydroxyphenyl moiety seems to have additional biologically interesting activities. This group



Chemical structure of RPI-856 A and B, which differ in the chirality of the β -methine carbon in the 3-amino-2-oxo-4-phenylbutyric acid residue.

readily undergoes oxidation and yields the corresponding quinone, which is a nucleoside analogue and therefore a potential substrate for DNA and RNA polymerases. A large percentage of anti-HIV agents isolated from natural sources also contains a 1,3-hydroxyphenyl group or a similar structural element in a condensed ring system. These compounds were shown to inhibit the HIV reverse transcriptase with varying efficiency, and in some cases additional antiviral activity has also been reported (Ng *et al.*, 1997; Vlietnick *et al.*, 1998; Matthée *et al.*, 1999). For example, michellamine (Figure 4.3), a compound isolated from a rain forest plant, was initially identified as a potent inhibitor of the HIV reverse transcriptase (Boyd *et al.*, 1994). In later studies, however, it also was found to inhibit cellular fusion and syncytium formation (McMahon *et al.*, 1995). It seems therefore plausible that feglymycin might also possess multiple action mechanisms as an anti-HIV agent.



Chemical structure of michellamine B.

STRUCTURE DETERMINATION

Crystallisation

Feglymycin is a white powder that is moderately soluble in water and well soluble in alcohols. As feglymycin was available only in a limited amount (77 mg) and crystallisation conditions had not yet been established, it seemed reasonable to crystallise it with the hanging drop method, in which the substance used per crystallisation trial is negligible in comparison with batch methods commonly employed to crystallise small molecules. An adequate amount of the peptide was dissolved in ethanol:water=1:2 mixture to a concentration of 20 mg/ml and 2 μ l aliquots were employed in crystallisation trials at room temperature with the Hampton Crystal Screen I kit (Jancarik & Kim, 1991). Crystals appeared in two conditions:

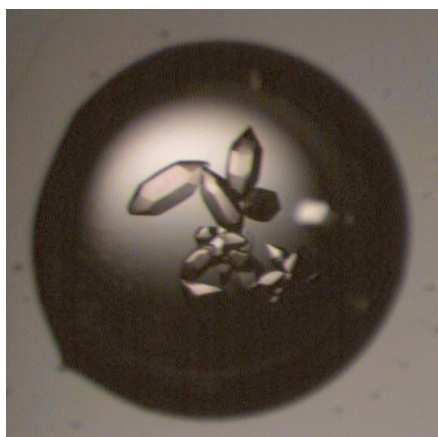
- (a) Condition 8: 0.1 M Na-cacodylate/cacodylic acid pH=6.5, 30% isopropanol, 0.2 M sodium citrate. Small needles were obtained.
- (b) Condition 36: 0.1 M Tris/Tris HCl pH=8.5, 8 % PEG8000. Crystals having hexagonal bipyramidal morphology were grown.

These two conditions were optimised independently, and diffraction quality crystals (with dimensions up to 0.5·0.3·0.2 mm³) were obtained by:

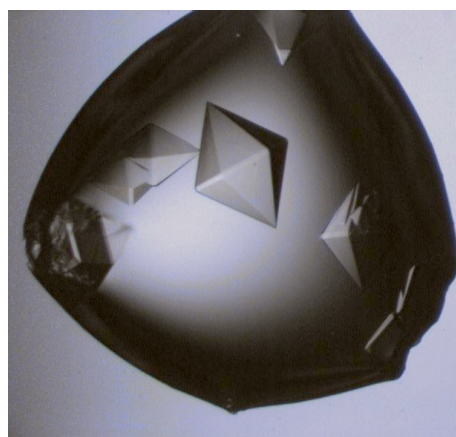
- (a) mixing 2 μ l of 40 mg/ml feglymycin in ethanol:water=1:4 with 2 μ l of either 0.1 M Na-cacodylate/cacodylic acid pH=6.5, 30% isopropanol and 0.25 M Na-citrate or 0.25 M Na-citrate/citric acid pH=6.5 and 30% isopropanol at 4 °C. Crystals appeared within a month (crystal form **II**, Figure 4.4).
- (b) mixing 4 μ l of 10 mg/ml feglymycin in ethanol:water=1:4 with 2 μ l of 0.1 M Tris/Tris HCl pH=8.4 and 4% PEG8000 at room temperature. Crystal growth took about two weeks (crystal form **I**, Figure 4.4).

Measurement

For measurement, a crystal was taken out from the drop with a suitably sized loop and transferred into a droplet of the cryoprotectant solution containing 0.2 M Tris/Tris HCl, 4% PEG8000 and 25% glycerol for **I** and 0.3 M Na-citrate/citric acid pH=6.5, 30% isopropanol and 20% 1,2-propanediol for **II**. After a short soaking time, crystals were taken out with the loop and shock frozen in a cold nitrogen stream operating at 100 K (when crystallisation was performed at room temperature) or by dipping the loop into liquid nitrogen (for crystals grown at 4 °C).



Crystal form I



Crystal form II

Figure 4.4

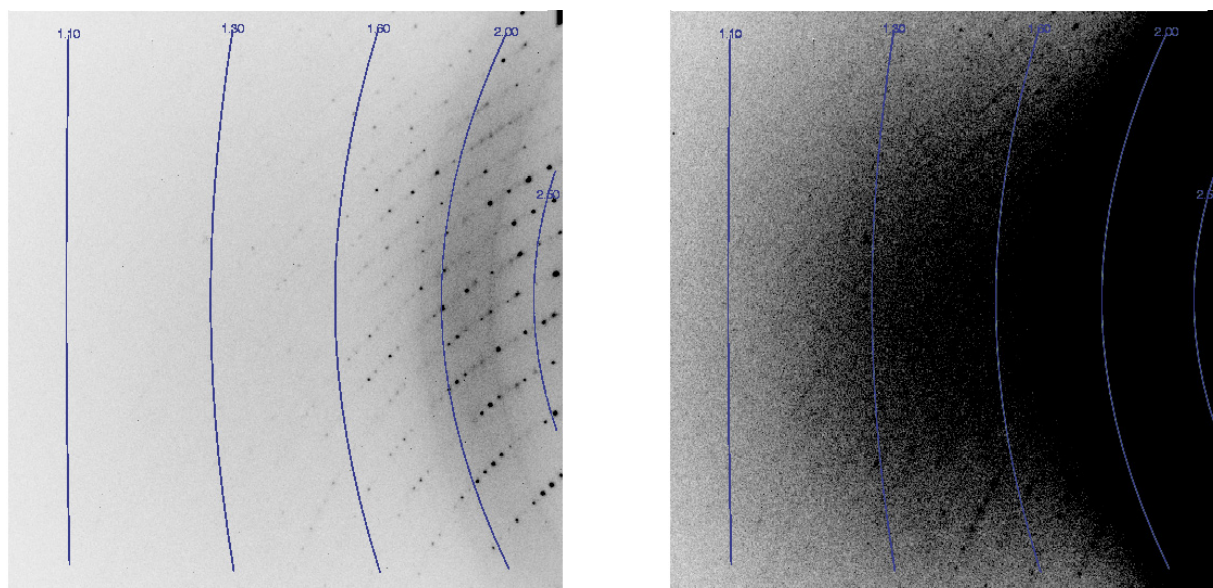
Crystal morphologies of feglymycin.

Crystals were then characterised by collecting datasets on a Bruker rotating anode equipped with Osmic focusing mirrors and a SMART6000 4K CCD detector. Images were indexed with the program Proteum and data were processed using SAINT and SADABS (Bruker Nonius, 2002). XPREP was employed for space group determination and for merging. Data collection statistics are summarised in Table 4.1.

	Crystal form I		Crystal form II	
	6/m	6/mmm	6/m	6/mmm
Unit cell parameters (Å)	a=b=83.544, c=35.734		a=b=60.302, c=83.748	
$\langle E^2 - 1 \rangle$	0.730		0.960	
Space group	P6 ₅	P6 ₅ 22	P6 ₄	P6 ₄ 22
No. of reflections	578199	606418	484869	484869
No. of unique reflections	57985	34003	34149	18476
Resolution (Å)	1.10 (1.20-1.10)	1.06 (1.15-1.06)	1.40 (1.50-1.40)	1.40 (1.50-1.40)
Completeness	100% (100%)	99.7% (98.7%)	99.8% (99.1%)	99.8% (99.2%)
Redundancy	9.97 (4.64)	17.78 (8.11)	14.17 (10.95)	26.20 (20.71)
$\langle I/\sigma(I) \rangle$	25.31 (3.06)	31.67 (3.62)	21.50 (3.49)	29.18 (4.79)
R_{int}	0.0408 (0.4482)	0.0488 (0.5682)	0.0634 (0.6665)	0.0640 (0.6811)

Table 4.1

Data collection statistics for feglymycin crystal forms. Statistics are included for the real and also for the apparent space group. $R_{\text{int}} = \sum |I - \langle I \rangle| / \sum I$.

**Figure 4.5**

Diffraction from crystal form **I** (same frame, different contour levels). Based on this image (frame no. 1, which actually has a low scale factor) a resolution cutoff about 1.4 Å may be justified, but a low-background detector, high redundancy and careful data collections gives intensities to 1.1 Å.

Crystal form **I** was found to have a hexagonal unit cell with dimensions $a=b=83.54$ Å $c=35.73$ Å and examination of the systematic absences together with the Laue-symmetry suggested that the space group was either $P6_122$ or $P6_522$. A dataset consisting of three 180° ω -scans at $2\Theta=64^\circ$ for the high-resolution data (1.05-2.0 Å) and three 180° ω -scans at $2\Theta=30^\circ$ for the low-resolution shells (to 1.55 Å) was collected. After processing, it was indicated by the statistics that the crystal, although weakly, diffracted to atomic resolution, as in the highest resolution shell (1.15-1.06 Å) the average intensity was marginally higher (3.62σ) than the limit of statistical significance (3σ). This was probably a joint benefit of the comparatively high redundancy, the thin-sliced data collection strategy and the highly sensitive low-background detector as on the frames very few reflections above 1.2 Å were discernible (Figure 4.5).

Crystal form **II** also belongs to the hexagonal system with cell dimensions $a=b=60.30$ Å $c=83.75$ Å and systematic absences indicated $P6_222$ or $P6_422$ as possible space groups. The crystal diffracted to about 1.8 Å on the home source and therefore a higher resolution dataset on the BL1 beamline at BESSY was also collected. To avoid overloads in the low-resolution shells, the data were collected in two runs with considerably shorter exposure times for the low-resolution frames. The data extends to 1.4 Å and was integrated with HKL2000

and scaled with SADABS. It is noteworthy that datasets of this crystal had an unusual $|E^2-1|$ value (0.960) more common for centrosymmetric space groups (0.968).

Structure solution

Crystal form I

Structure solution of antibiotic molecules is usually complicated by the presence of several copies of the molecule in the asymmetric unit, which becomes too big for traditional direct methods even if atomic resolution data were available, and a markedly different structure from everyday proteins, which decrease the efficiency of molecular replacement and the traditional heavy atom derivatisation, since the required interaction surfaces are missing. For that reason, a currently topical method for derivatising the crystal with non-specific anomalous scatterers, namely an iodide soak was performed first by soaking a crystal in about 0.5 M KI, but deterioration of diffraction quality (maximum resolution worse than 3 Å) and no significant anomalous signal in the dataset indicated that this method is not applicable for this crystal.

As atomic resolution data were available, *ab initio* solution using direct methods was the method of choice. However, the absence of heavy atoms and the relatively large unit cell that could accommodate up to 1000 atoms did not suggest a quick solution, as the largest equal-atom structure solved so far with SHELXD (Sheldrick *et al.*, 2001) contained only a few more than 700 atoms and the process took about six months CPU time on a Pentium 200 processor. Two different strategies were employed in parallel:

- (a) search for a considerable portion of the atoms (FIND 600) and then perform peaklist optimisation (PLOP 800 1000).
- (b) find only a small percentage of the atoms (FIND 200) and do several cycles of peaklist optimisation (PLOP 300 450 675 1000) for the promising ones (TEST 25 5).

Strategy (a) was slower, it required about one hour for a try, but it was expected that the efficiency of the dual-space recycling in the search compensated for the increased computational demand. In strategy (b), which required only 9 minutes/try, the increased speed was exploited to test a higher number of starting atom sets when searching for the solution.

After one week CPU-time on a 2 GHz Intel Xeon processor, SHELXD run (b) gave a solution, although it could not be discriminated from non-solutions based on the standard figure-of-merit *final correlation coefficient*, which should be above 65 in order to consider the structure to be solved with high probability. However, both the CC and the CCweak

figures-of-merit showed a good discrimination of the solution from the non-solutions. Computer graphics examination of the solution with the program XP revealed a chemically sensible peptide structure with the expected amino acids and confirmed the success. Later the presence of perfect merohedral twinning was recognised that may account for the low correlation coefficient value. As the screw axis was undoubtedly identified previously as being a $6_1/6_5$, obtaining the twin law was straightforward. However, direct solution with the twin law and the lower symmetry space group was unsuccessful, probably because of the size of the structure and the lower resolution (merging in 6/m reduced the redundancy and justified a cutoff at 1.1 Å).

Crystal form II

As the data extended to only sub-atomic resolution and the structure of crystal form **I** was available at that time, molecular replacement was tried. The program EPMR (Kissinger *et al.*, 1999) was employed with a feglymycin dimer as the search model and data between 10 and 3.2 Å were used. EPMR was selected as the search model was relatively small and division of the six-dimensional search for a rotational and a translational one is no longer valid in such a case. Assuming 50% solvent content found at crystal form **I**, two dimers were expected in the asymmetric unit. As the hand of the space group is not determined and the search model is chiral, solution was attempted in $P6_222$ and $P6_422$. In $P6_222$ no promising results were obtained, while in $P6_422$ EPMR found the first dimer (dimer 1) with a high

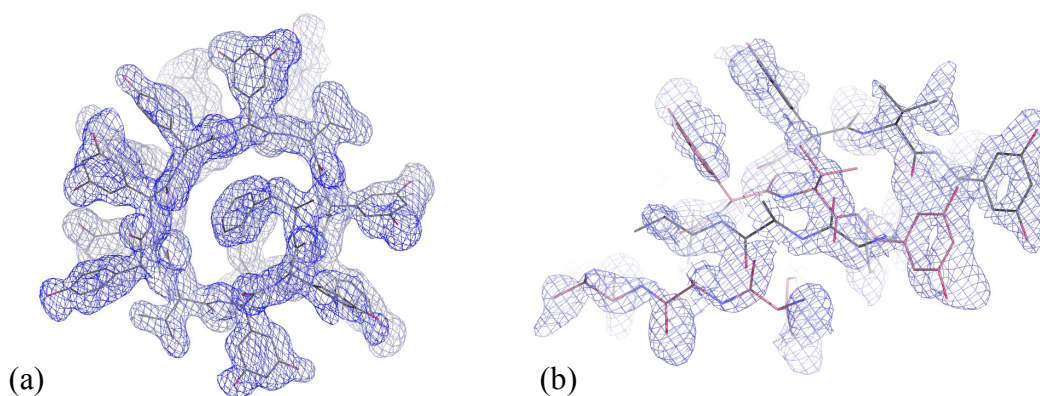


Figure 4.6

Electron density from SHELXE for crystal form **II** in $P6_422$. (a) Dimer 1 is clearly seen in the density and even the side chains are recognisable, which were not included in the starting model. (b) In contrast, electron density on dimer 2 is not directly interpretable and is accounted for by two overlapped molecules. The model is taken from the final refinement, but symmetry equivalents were generated according to $P6_422$ symmetry.

success rate and good discrimination, but addition of the second molecule (dimer 2) did not improve the figures-of-merit. Examining the packing of the two dimers revealed that while dimer 1 packed in a chemically sensible way, dimer 2 had several close contacts with itself. As the resolution of the data is superior to normal proteins and the solvent content is not too low, SHELXE (Sheldrick, 2002) was expected to produce an excellent map even when starting from a partial solution. For that purpose, dimer 2 was removed from the model, and dimer 1 was used as a starting seed. The obtained map was by no means clearer than the one calculated from the model using SigmaA-weights (Figure 4.6). This suggested the presence of perfect merohedral twinning, and based on packing violations, the same twin law was found as for crystal form **I**. It is interesting to note that the characteristic indication for twinning, namely the $|E^2-1|$ value in that case is actually too high and is not typical for a twinned structure. EPMR was then re-run in the lower symmetry space group, but could locate two dimers corresponding to dimer 1 in P6₄22. In another try, this tetramer was held fixed and the missing two dimers were searched as one tetramer. This approach led to a solution with respectable statistics.

Model building and refinement

Both structures were refined against F_o^2 with no intensity cutoff using SHELXL, which can handle twinning and is generally recommended for high-resolution structures. Detwinned intensity data output by SHELXL (Sheldrick & Schneider, 1997) were used to calculate and inspect $2mF_o-DF_c$ and F_o-F_c type maps using the program XtalView (McRee, 1999) for manipulation of atoms. Throughout the refinement, bond length, bond angle and chiral volume restraints were imposed, which were taken from the Engh and Huber library (Engh & Huber, 1991) for aspartate, phenylalanine and valine or generated from the Cambridge Structural Database (Allen, 2002) for 4-hydroxyphenylglycine and 3,5-dihydroxyphenylglycine. For the two enantiomers of 4-hydroxyphenylglycine present in the structure, the same restraints were employed except for the sign of the chiral volume restraint, which was inverted. Planarity restraints were applied to peptide bonds and aromatic side chains. Solvent atoms were added using SHELXWAT and also manually. All non-hydrogen atoms were refined anisotropically with suitable rigid bond and similarity restraints. Approximately isotropic restraints were also employed for solvent water molecules. Hydrogen atoms were included in the later stages of refinement. Refinement statistics are listed in Table 4.2.

Crystal form I

Refinement was initiated in the space group $P6_522$ (chirality determined by assigning the common amino acids to have L-configuration found by HPLC on chiral medium). After two rounds of B-value refinement, all peptide atoms could be located. The structure consisted of three peptide chains in one and a half dimers, a crystallographic twofold axis generating the second half of the dimer. The R -values were 31% and 33% for R_1 and R_{free} , respectively, which are suspiciously higher values than expected for a complete structure. It was realised at this stage that although a dimer nearly obeyed twofold symmetry, the side chain of Phe12 residue did not and would always collide with its symmetry equivalent. This finding suggests that this twofold axis cannot be crystallographic as it would always imply the presence of disorder, the space group is therefore only $P6_5$ that is twinned by a twofold rotation perpendicular to c and contains three dimers in the asymmetric unit and a non-crystallographic twofold axis. The nearly perfect non-crystallographic twofold also offered an explanation to the severe non-planarity of the benzene rings that was significant at such a high resolution. Refinement in $P6_5$ gave lower R -values ($R_1=27.03\%$ vs. 31.00%) and solved the

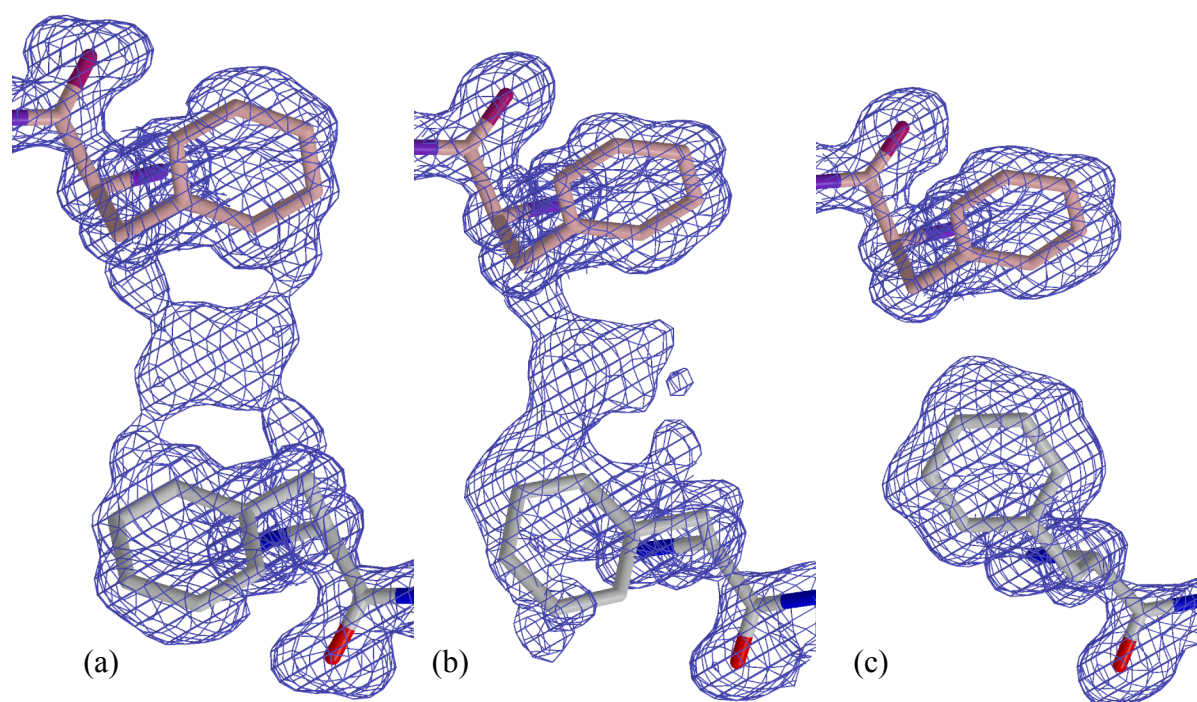


Figure 4.7

View of the sidechain of Phe12 and its symmetry equivalent along a crystallographic twofold axis. (a) In $P6_522$ there is density in between, but fitting the atoms would involve a disorder, $R_1=31.0\%$. (b) After the first refinement in $P6_5$, the density is polarised towards one of the residues (no longer symmetry equivalents), $R_1=27.0\%$. (c) After adjustment, clear density appears on the sidechains, $R_1=26.9\%$.

problem of the side chain conformation for the aforementioned phenylalanine residue (Figure 4.7). Comparison of K -values ($K = \langle F_o \rangle / \langle F_c \rangle$) as a function of reflection intensity as output by SHELXL showed a more common, flat behaviour for the refinement in P6₅, while when refining in P6₅22 without twinning resulted in a K -value significantly higher for the weakest reflections than the average. Refinement in P6₅ without the twin law resulted in a significant increase in R -values. Despite the usual $|E^2 - 1|$ value, the assumption of twinning was considered to be justified and refinement was carried on in P6₅. The final model contains 78 residues, several PEG units and 77 water molecules, and with 1026 non-hydrogen atoms it is by far the largest equal-atom structure solved with *ab initio* direct methods so far.

Crystal form II

As the presence of twinning had already been established in the structure solution stage, refinement was initiated taking the twin law into account. The availability of the complete structure from crystal form I simplified model building to completing the solvent structure since the conformation of feglymycin in the two crystal forms was found to coincide nearly exactly. In the electron density maps the occurrence of significant radiation damage can be seen, especially at the aspartate residues, and zero-dose extrapolation improves the map in these regions. The final model contains 104 residues, 4 isopropanol and 106 water molecules.

	Crystal form I	Crystal form II
Resolution range (Å)	72.35-1.10	28.37-1.40
R_1 [$F > 4\sigma(F)$ /all data]	0.1434/0.1689	0.1590/0.1764
R_{free} [$F > 4\sigma(F)$ /all data]	0.1674/0.1950	0.1957/0.2193
No. of parameters	9811	11147
No. of restraints	13414	14594
No. of peptide atoms	828	1104
No. of non-H atoms	1026	1221
Solvent content	51.2%	52.6%
Twin ratio	0.494	0.500
R.m.s.d. from ideal geometry		
Bond lengths (Å)	0.022	0.017
Angle distances (Å)	0.045	0.037
Common plane (Å)	0.290	0.293

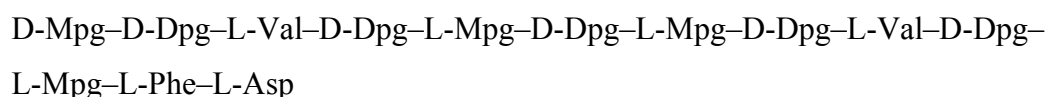
Table 4.2

Refinement statistics.

STRUCTURE DESCRIPTION

Amino acid residues

As feglymycin contains unusual amino acids and the chemical structure was established with NMR and MS techniques, nothing was known about the chirality except for the fact that the common amino acids aspartate, phenylalanine and valine were identified to have L-configuration. The X-ray structure revealed that all the 3,5-dihydroxyphenylglycine (Dpg) residues have D-configuration, the first 4-monohydroxyphenylglycine (Mpg) residue possesses D-configuration, while the other three Mpg residues have L-configuration. The sequence therefore can be written as:



Feglymycin thus consists of alternating D,L-amino acids with certain deviations from this regularity near the termini. This pattern of amino acids with alternating chirality is not unusual among secondary metabolites and the extensively studied antibiotic gramicidin isolated from *Bacillus brevis* exhibits the same alternation with the difference that for gramicidin this pattern runs along the sequence without exceptions (Sarges & Witkop, 1956).

Dimeric structure

Feglymycin was found to form dimers in both crystal forms. The dimers are of the right-handed double-stranded antiparallel double β -helix type and contain about 9 residues/turn (Figure 4.8). This finding is consistent with all known X-ray structures of gramicidin, where dimer formation is also a characteristic feature. However, in the case of gramicidin, the helix rise is 5.6 residues/turn for the uncomplexed structure (Langs, 1988) and 7.2 residues/turn when crystallised as a Cs^+ -complex (Burkhart, Li *et al.*, 1998). An additional difference is the presence of two phenylalanine sidechains inside the feglymycin channel, while the gramicidin pore is empty (Figure 4.9). Nevertheless, exhibiting such extensive similarities, it is probably justified to classify feglymycin as a structural analogue of gramicidin.

Dimer formation

The feglymycin dimer is a very compact molecule that is stabilised by a huge number of secondary interactions. The formation of the dimer can be visualised as follows. In the first

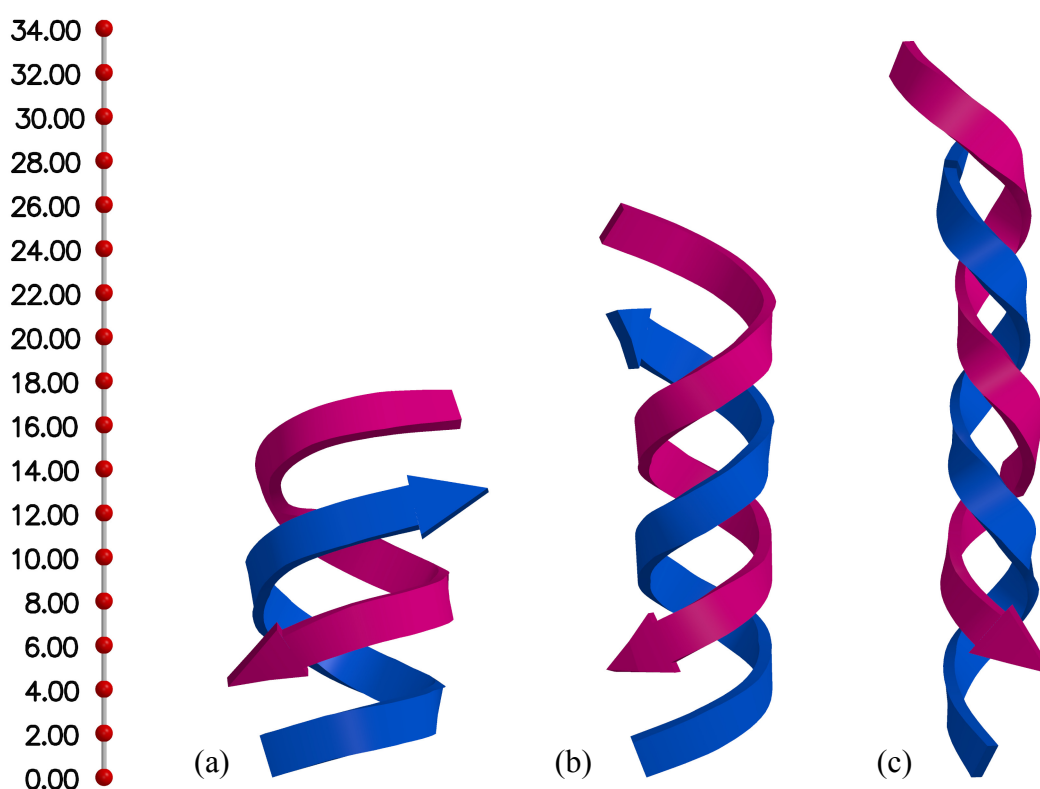


Figure 4.8

Schematic representation of the size and structure of feglymycin (a), gramicidin- Cs^+ (b) and uncomplexed gramicidin (c) dimers. (a) and (b) are right-handed helices, while (c) is left-handed. The absolute scale is shown by the ruler on the left in Å units.

step, two molecules of feglymycin associate in an antiparallel fashion to form a two-stranded β -sheet. This β -sheet is unusual by the fact that all sidechains are pointing to the same side of the strand because of the regular alternate chiral arrangement of the residues and has a considerable sterical tension originating from close contacts between sidechain atoms, which exerts a force on the backbone and tries to bend it. The backbone gives way to the pressure and assumes a helical conformation. In the last step, a conformational rearrangement of the sidechains to form interstrand hydrogen bonds further stabilises the dimer.

Dimer symmetry

In case of gramicidin, the dimer both in the uncomplexed and the Cs^+ -complexed form was found to exhibit noncrystallographic symmetry with the mainchain atoms being nearly exactly related by a twofold axis and some minor deviations at certain sidechains. As a homologue, feglymycin was also expected to show twofold symmetry.

To get an unbiased comparison of monomers in a dimer, rigid sets of atoms were identified with the program ESCET (Schneider, 2002) among all molecules in crystal form **I**.

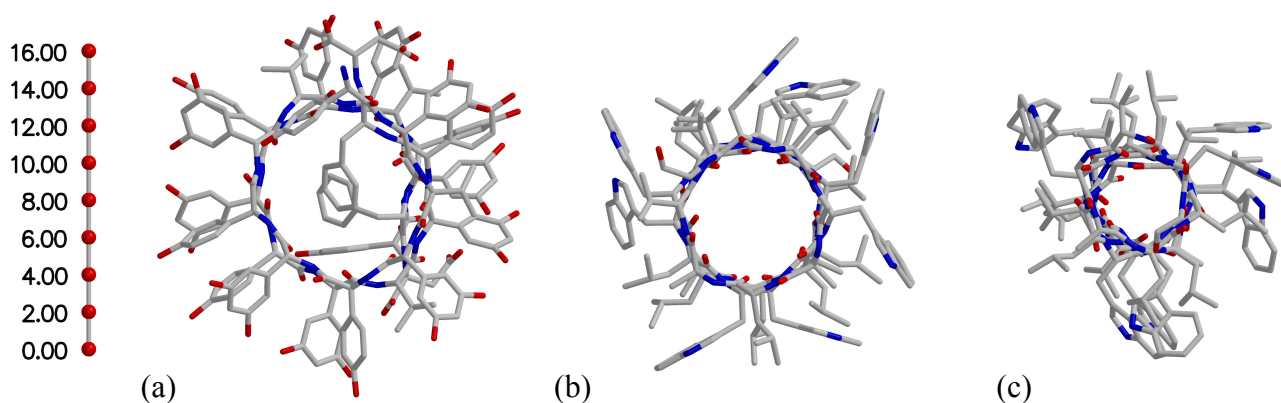


Figure 4.9

View along the helix axis for feglymycin (a), gramicidin-Cs⁺ (b) and for the uncomplexed gramicidin (c) dimer. Each helix has a different turn: (a) belongs to the $\beta^{9.0}$ -family, (b) to the $\beta^{7.2}$ -family and (c) to the $\beta^{5.6}$ -family and accordingly differ in pore diameter. The ruler on the left is in Å units.

As crystal form **II** was solved using molecular replacement with a dimer from crystal form **I**, it was found that all four dimers in crystal form **II** closely resembled the structure of the model and were omitted from the analysis. The rigid atoms in the six monomers present in crystal form **I** were then fitted onto each other and compared using XtalView (Figure 4.10).

The six monomers were found to be significantly different within the error of the analysis. Based on their conformation, the monomers can be divided into two classes (A and B), the main difference being the mainchain conformation for residues 12 and 13. A dimer is built up by one monomer having class A and one exhibiting class B conformation and this implies that the feglymycin dimer is less symmetric than the one in gramicidin. This

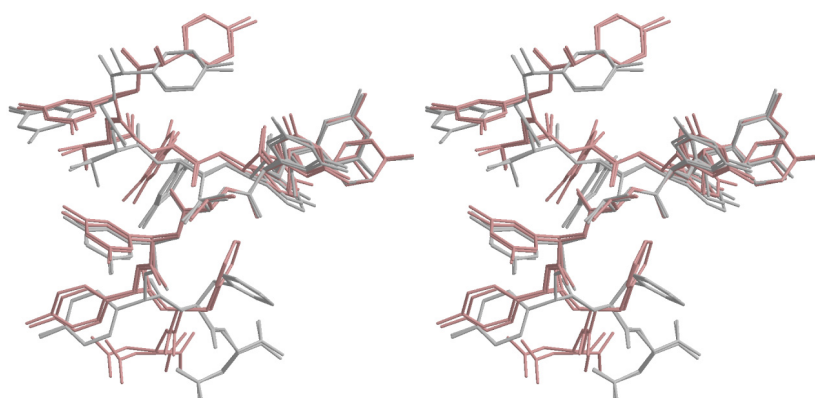


Figure 4.10

Stereo view of the fit of feglymycin monomers reveals two conformation classes (grey and redbrown). The N-terminus is on the top, the C-terminus and the neighbouring Phe12 is near the bottom. A fit of four monomers is displayed.

difference may arise from the fact that feglymycin is not as regular concerning the chirality alternation as gramicidin, and the finding that the deviation occurs in a region where the regular alternation is broken seems to confirm the assumption. It is interesting to note that at the other terminus, where one residue violates the regularity, there is also a conformational diversity observed. The conformational flexibility exhibited by this residue may help in adaptation to the environment and therefore be important in stabilising the crystal lattice, which is supported by the finding that this residue is fundamentally involved in interdimer interactions.

Secondary interactions

The helical dimer present in the structure is stabilised by a number of hydrogen bonds. As the structure is fairly complex, for a sufficiently clear description a notation is introduced at this point that will be used throughout this chapter. Atoms are designated following PDB notation for conventional residues; for unconventional residues the numbering is shown in Figure 4.11. Independent dimers in the asymmetric unit will be distinguished by roman numbers. Monomers are denoted by capital letters A and B that refer to their conformation class, followed by the number of the dimer. If this number is omitted, the statement is meant generally and applies to all dimers. Apostrophes indicate symmetry equivalents; for those generated by a common symmetry operator the same notation is used. Thus, the hydrogen bonds can be classified as:

- (a) "primary" hydrogen bonds between mainchain atoms of the two antiparallel chains. Interactions belonging to this category join the carbonyl oxygen and peptide nitrogen atoms of even-numbered residues of one chain with the even-numbered residues of the

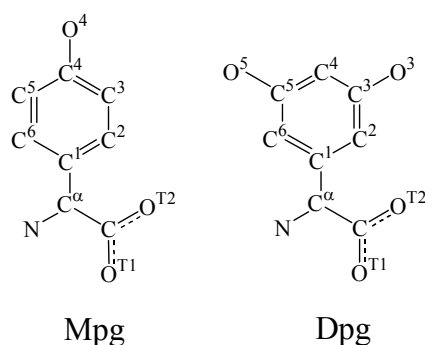


Figure 4.11

Atom names for 4-monohydroxyphenylglycine (Mpg) and 3,5-dihydroxyphenylglycine (Dpg).

other in an antiparallel fashion (Figure 4.12). Residues Dpg2-Phe12 are involved and as there are two interactions per residue, altogether 12 such hydrogen bonds are formed. An additional hydrogen bond $O^{\delta 1}(\text{Asp13-A})-\text{N}(\text{Mpg1-B})$ provides further strength to the dimer. However, there is no interaction between $\text{N}(\text{Mpg1-A})-O^{\delta 1}(\text{Asp13-B})$, i.e. between the other termini, which is a consequence of the displacement of Asp13 in the class B chains. These hydrogen bonds are conserved in all independent dimers.

- (b) "secondary" hydrogen bonds between mainchain atoms that are corollaries of the helical structure. On the other hand, these are interactions that stabilise the double helix. There are considerably less hydrogen bonds that belong to this class, as only the odd numbered residues between Val9-Asp13 participate (Figure 4.12). These involve 5 hydrogen bonds, namely $O(\text{Val9-A})-\text{N}(\text{Asp13-B})$, $\text{N}(\text{Mpg11-A})-\text{O}(\text{Mpg11-B})$, $\text{O}(\text{Mpg11-A})-\text{N}(\text{Mpg11-B})$, $\text{N}(\text{Asp13-A})-\text{O}(\text{Val9-B})$ and $O^{\text{T1}}(\text{Asp13-A})-\text{N}(\text{Val9-B})$. It is noteworthy that the $\text{N}(\text{Val9-A})-\text{O}^{\text{T1}}(\text{Asp13-B})$ is not present since Asp13-B is further away. Two additional interactions between $\text{O}(\text{Val9-A})-\text{O}^{\text{T1}}(\text{Asp13-B})$ and the opposite direction $\text{O}^{\text{T1}}(\text{Asp13-A})-\text{O}(\text{Val9-B})$ strengthen the interaction further, which suggest that the termini are protonated. Since the bond distance in the first interaction is 3.6 Å, while in the second it is 3.3 Å, these interactions are rather weak. The helical

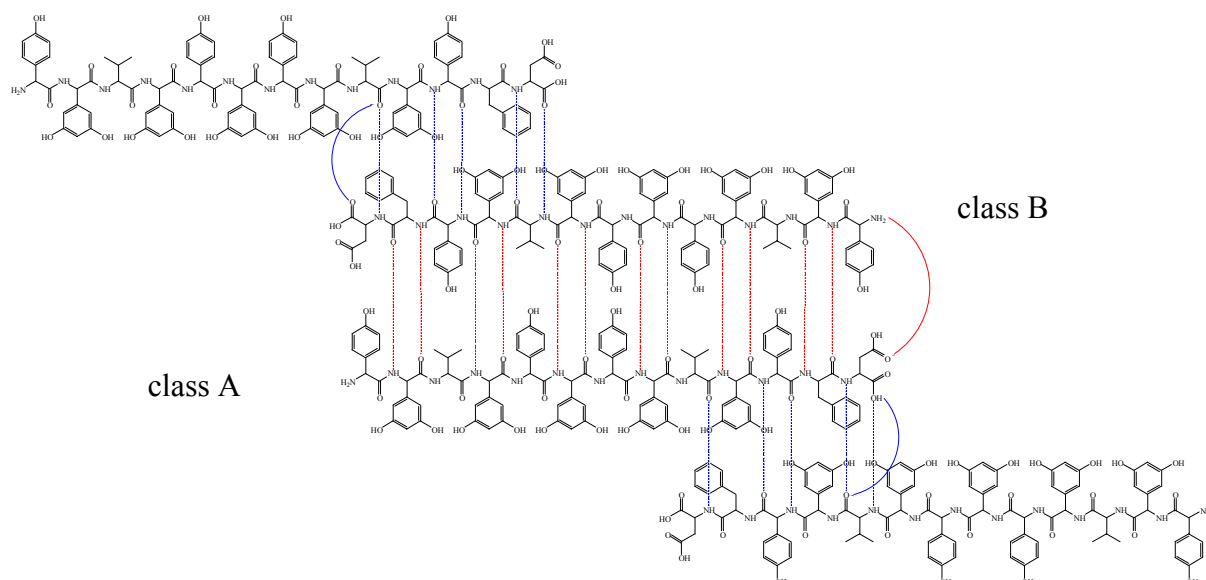


Figure 4.12

Hydrogen bonds within a feglymycin dimer. "Primary" hydrogen bonds (red) connect the two antiparallel chains, while "secondary" hydrogen bonds (blue) secure the helical structure. The "offset" of the helix is thus eight residues.

structure is not stabilised by as many interactions as those that join the two chains, and can be expected to rearrange under favourable conditions. Such a process has already been described for gramicidin that forms a left-handed double helix when uncomplexed, but undergoes substantial conformational movement when a suitable metal atom is added to the solution and converts into a right-handed double helix (Burkhart *et al.*, 1999). Little variation in these hydrogen bonds among independent dimers in the asymmetric unit can be observed, namely the $O(\text{Val9-A})-O^{\text{T1}}(\text{Asp13-B})$ can be replaced by $O(\text{Val9-A})-O^{\delta^1}(\text{Asp13-B})$.

(c) hydrogen bonds that involve hydroxyphenyl moieties. Feglymycin possesses therefore two shells of hydrogen bonds: one is constituted by mainchain atoms and the other comprises hydrogen bonds between sidechain hydroxyls that are further apart from the helix axis (Figure 4.13). This multi-shell structure may provide unprecedented rigidity to the feglymycin dimer. The ordering of these bonds produces long-range hydrogen-bonded chains on the surface of the molecules, and gives directionality to all Dpg residues. Four such chains can be discerned:

- (i) $O^3(\text{Dpg2-A})-O^{\delta^1, \delta^2}(\text{Asp13-B})O^{\delta^2}-O^3(\text{Dpg10-A})O^5-O^5(\text{Dpg4-B})$,
- (ii) $O^3(\text{Dpg4-A})-O^5(\text{Dpg10-B})O^3-\text{HOH}-O^3(\text{Dpg2-B})-O^{\delta^2}(\text{Asp13-A})$,
- (iii) $O^3(\text{Dpg6-A})-O^5(\text{Dpg8-B})O^3-\text{HOH}-O^{\text{T}}(\text{Asp13-A})$ and
- (iv) $O^3(\text{Dpg8-A})-O^3(\text{Dpg6-B})$.

As atoms involved in these interactions are on the surface and most susceptible to packing interactions, the hydrogen-bonding pattern shows considerable variation

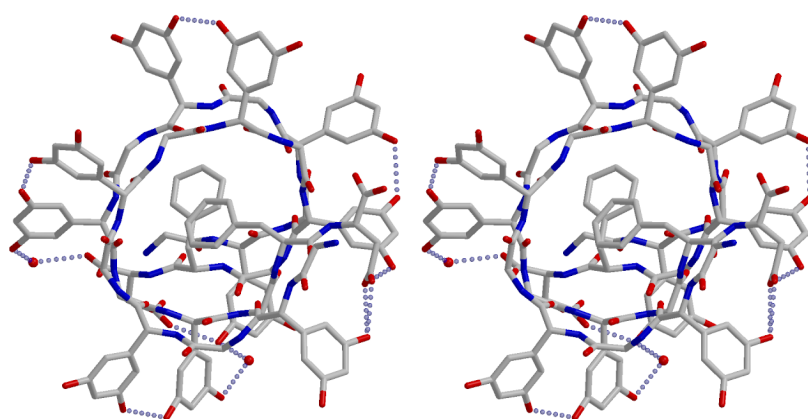


Figure 4.13

Stereo view showing the outer hydrogen-bonding shell of the feglymycin dimer. Sidechains not participating in these interactions have been truncated for clarity.

among the independent dimers. These include the absence of equivalents to water molecules, and break chain (ii) into two sub-chains and terminate chain (iii). The hydroxyphenyl moiety of Mpg1 in certain cases coils towards the backbone and participates in hydrogen bonds $O^4(\text{Mpg1-A})\cdots N(\text{Mpg7-A})$ and $O^4(\text{Mpg1-A})\cdots O(\text{Mpg7-A})$.

Apart from the hydrogen bonds, hydrophobic interactions also take part in stabilising the structure. Straight after the inner hydrogen-bonding sphere there is a thick shell filled with aromatic residues that all run approximately parallel to each other and compose a 30° -angle with the helix axis. In addition, the benzene ring of Phe12-A stacks on the one of Phe12-B. This interaction is within the helix (and thus inside the inner hydrogen-bonding shell) and therefore the least sensitive to environmental impacts.

Helical structure

The feglymycin helix belongs to the antiparallel double-stranded double β -helix family (also known as $\pi\pi$ -helix). Peptides consisting of alternating D,L-amino acids are known to form such helices. An early work studying possible conformations of poly(D-L)-peptides (Colonna-Cesari *et al.*, 1977) postulated the existence of several families based on the helical turn, which is then dependent on the "offset" in residues participating in the "secondary"-type hydrogen bonds, i.e. the length of the sequence that is not overlapped by helix formation (Figure 4.12). Based on theoretical considerations, they concluded that the first member of this series has a minimum of 5.6 residues/turn, followed by the next member with 7.2 residues, and then 9.0, 10.7 and 12.6 residues/turn. The helix is stabilised as the pore grows since the number of close contacts with opposite residues decreases, but for a finite size helix the number of hydrogen bonds resulting from helix formation also decreases and the optimal turn is somewhere near 9.0. These molecules can also form parallel helices, but these become unstable when the helix diameter increases, the widest pore occurs at around 7.2 residues/turn.

Gramicidin was shown to form two types of these helices: one of them is an antiparallel $\beta^{5.6}$ -helix (uncomplexed), while the other one is an antiparallel $\beta^{7.2}$ -helix (Cs^+ -complex). The first one has a left-handed screw, while the other one coils to the right. The average diameter of the $\beta^{5.6}$ -helix is about 5.5 Å, which allows passage of ions with ion radius not greater than 1 Å. On the other hand, the $\beta^{7.2}$ -helix has a helix diameter around 8 Å and pore dimensions allowing particles as large as 4.5 Å in diameter to cross the channel (Figure 4.9). These numbers are vital to understand the function, as gramicidin is a membrane channel

peptide that causes the leakage of biological membranes. The exact conducting form is either the $\beta^{7.2}$ -helix described above or a head-to-head single-stranded helix found by NMR investigations conducted in micelles and lipid bilayers (Lomize *et al.*, 1992; Ketchum *et al.*, 1996; Townsley *et al.*, 2001). The $\beta^{5.6}$ -helix is thought to be important for membrane insertion, and synthetic modifications of gramicidin that do not convert to the $\beta^{7.2}$ -helix after membrane integration are poor ion-conductors. The gramicidin pore can allow Cs^+ -ions to pass (diameter=3.6 Å), which also supports the assumption that it is the $\beta^{7.2}$ -form and not the $\beta^{5.6}$ -form that is responsible for membrane conduction.

On the other hand, feglymycin corresponds to a $\beta^{9.0}$ -helix with the exact helical turn varying from 8.8 (chain B) to 9.2 (chain A) residues/turn. In this case, the channel diameter is 9.5-12 Å and the pore size is between 6.5-9.0 Å. The channel is wide enough to accommodate a benzene ring since there are two phenylalanine sidechains in the pore. In both of the crystal structures, one of which crystallised from 4% PEG8000 (virtually from water), the other one from 30% isopropanol corresponding to an apolar environment, the conformations of the peptides were found to coincide very closely. It is therefore likely that feglymycin retains the $\beta^{9.0}$ -helix when inserts into the membrane (if at all) and does not readily undergo such substantial conformational rearrangement as found for gramicidin.

Crystal packing

In both of the crystal forms, feglymycin composes infinite helical chains of dimers enforced by the 6_5 or the 6_4 crystallographic screw axis. These chains are so symmetrical that, not regarding certain atoms, a 180° -rotation perpendicular to **c** transforms the helix to itself. This property may be responsible for merohedral twinning present in these crystals. Although a dimer is fairly symmetrical, the two chains (class A and B) that build it up differ in conformation at the termini, which provides directionality to the dimers. Reversing the directionality would involve 180° rotations around the $\text{C}^\alpha\text{--C}^\beta$ bond for two phenylalanine sidechains. This not likely to happen spontaneously as to remove the steric hindrance would require the folding out of the helix. The side of the dimer containing the N-terminus of the class A chain will be referred to as "head" and the other one as "tail".

Crystal form I

The asymmetric unit of crystal form **I** contains three dimers that build up one helical chain and all three theoretically possible fittings, namely head-to-head, head-to-tail and tail-to-tail are present amongst the molecules. One of the dimers has arbitrarily been chosen as

dimer I, and the other two dimers are numbered sequentially as they follow each other in the helix. In the apparent higher symmetry space group (P6₅22) dimer I would be a symmetry equivalent of dimer III and dimer II-chain A of dimer II-chain B.

Dimer I is connected to a symmetry equivalent of dimer III via a head-to-head interaction. Hydrogen bonds between mainchain amide and carbonyl atoms, namely N(Mpg1-A/I)–O(Mpg5-A/III'), O(Mpg1-A/I)–N(Mpg5-A/III'), N(Val3-A/I)–O(Val3-A/III'), O(Val3-A/I)–N(Val3-A/III'), N(Mpg5-A/I)–O(Mpg1-A/III') and O(Mpg5-A/I)–N(Mpg1-A/III') constitute the dimer interface, which is nearly a perfect continuation of the helix-forming hydrogen bonds. Hydrophobic interactions between the aromatic rings of Dpg2-A/I–Dpg4-A/III' and Dpg4-A/I–Dpg2-A/III' provide further strength. At the other end, dimer I is linked to dimer II by a tail-to-tail interface that consists of hydrogen bonds N(Mpg1-B/I)–O(Val3-B/II), O(Mpg1-B/I)–N(Val3-B/II), O⁴(Mpg1-B/I)–O(Mpg5-B/II), N(Val3-B/I)–O(Mpg1-B/II), O(Val3-B/I)–N(Mpg1-B/II) and O(Mpg5-B/I)–O⁴(Mpg1-B/II), and aromatic interactions between sidechains at Mpg1-B/I–Mpg1-B/II and Dpg2-B/I–Dpg2-B/II. The mainchain atoms do not participate as much as observed at the head-to-head interface but the N-terminal hydroxyphenyl moiety also takes part in the interaction. To complete the palette, there is a head-to-tail interaction between dimer II and III, constituted by hydrogen bonds N(Mpg1-A/II)–O(Mpg3-A/III), O(Mpg1-A/II)–N(Mpg3-A/III), O⁴(Mpg-A/II)–O(Mpg5-A/III), O(Val3-A/II)–N(Mpg1-A/III) and N(Val3-A/II)–O(Mpg1-A/III), and the aromatic stacking between Dpg2-A/II–Dpg2-B/III. This interface is similar to the tail-to-tail interaction except for the participation of the hydroxylphenyl group. The extensive involvement and the diversity of interfaces between the dimers account for the conformational variety of Mpg1. As this residue does not follow the chirality alternation, it is not held tight and flexible enough to adapt to the environment.

It is instructive to compare the different interfaces regarding the empty space they enclose. In the head-to-head interface the terminal Mpg residues are displaced towards the wall of the helix by a hydrogen bond, and as a result, they open up and give access to the hollow space inside the feglymycin dimer, which is big enough to accommodate an 8-residue-long polyethyleneglycol chain (Figure 4.14a). In contrast, the hydrophobic interaction between these amino acids observed in the tail-to-tail dimer attracts them towards the helix centre and blocks the entry to the dimer interior (Figure 4.14b). The head-to-tail interface occupies a position in between.

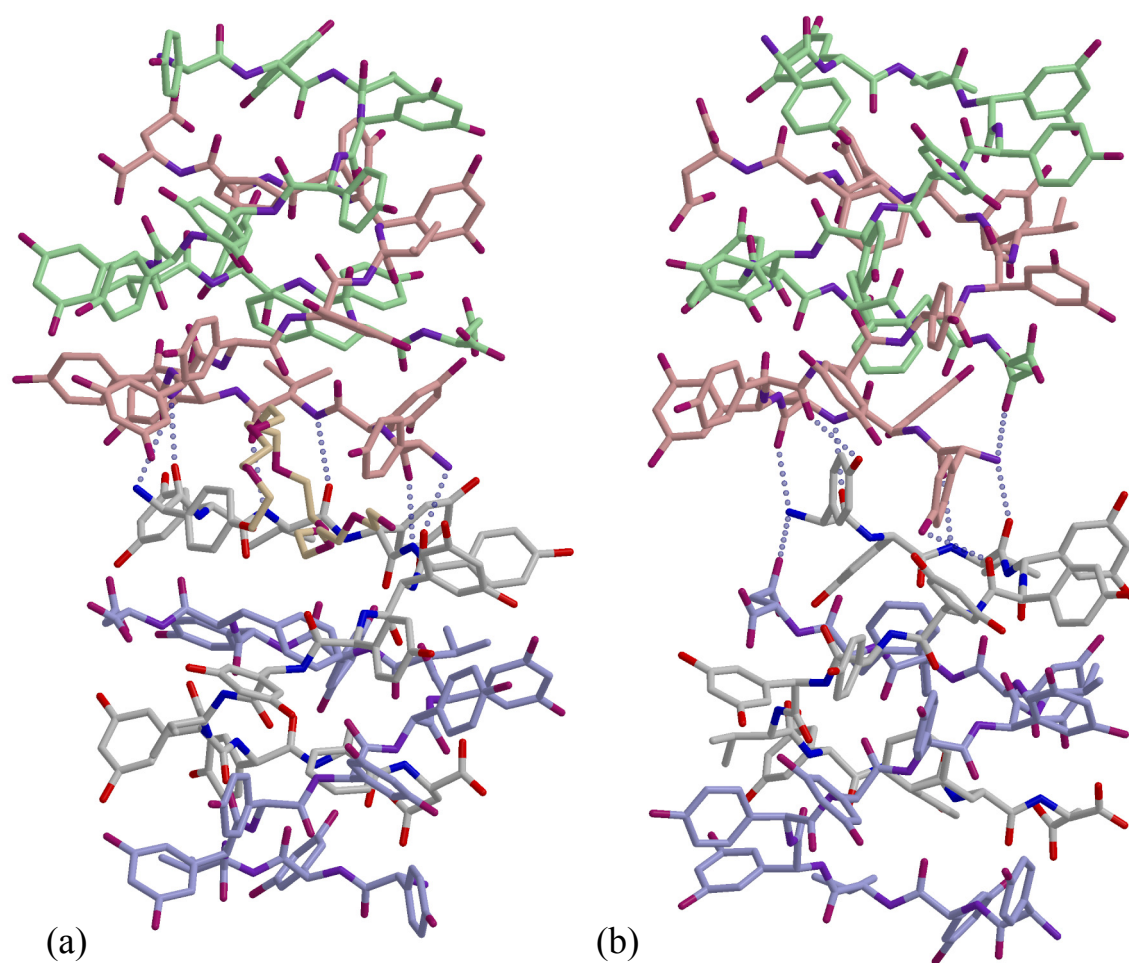


Figure 4.14

Interdimer contacts in the feglymycin crystal. The tetramer shown consists of two dimers (grey-blue and redbrown-green). (a) The head-to-head interface, where the phenyl rings of terminal Mpg residues open up and give access to the molecular interior. The cavity is big enough to accommodate a polyethyleneglycol chain (light brown). (b) The tail-to-tail interface, in which these phenyl rings interact with each other strongly and close the entry to the internal hollow.

On the other hand, very few interactions can be observed between the helices that build up the crystal. These are limited to some hydrogen bonds between $O^5(\text{Dpg2-A/I})-O^5(\text{Dpg4-A/III}')$, $O^5(\text{Dpg2-B/I})-O^5(\text{Dpg2-A/II}')$, $O^5(\text{Dpg2-B/I})-O^5(\text{Dpg2-B/III}')$, $O^5(\text{Dpg2-B/II})-O^5(\text{Dpg2-B/III}')$, $O^5(\text{Dpg2-B/II})-O^5(\text{Dpg2-A/II}')$ and $O^5(\text{Dpg4-B/I})-O^5(\text{Dpg4-B/III}''')$. There are also big holes in the crystal filled with disordered water, and in other regions the helices are concentrated.

Crystal form II

This crystal form contains four dimers that build up two helices generated by the 6_4 axis. The two helices are not related by crystallographic symmetry, but closely resemble each other. One helix is built up by dimer I and II, while the other one by dimer III and IV in a nearly identical fashion. There is no direct interaction between the crystallographically independent helices, only among symmetry-related ones. This is likely to be the reason for the fragility of this crystal form compared to the other one. The twofold axis introduced by twinning holds for one helix, but not for the second, and introduces an apparent disorder (Figure 4.6). As the two helices are almost identical, only one of them is presented.

The two independent dimers in the helix associate first to a tetramer via a tail-to-tail interaction nearly identical to that observed in crystal form I. The only difference is the appearance of an additional hydrogen bond between $O^4(\text{Mpg1-B/I})-O^4(\text{Mpg1-B/II})$. These tetramers then associate via head-to-head interfaces as described above. The helices interact with symmetry related ones via four hydrogen bonds between $O^5(\text{Dpg2-B/I})-O^5(\text{Dpg2-B/I}')$, $O^5(\text{Dpg2-B/I})-O^5(\text{Dpg2-B/II}')$, $O^5(\text{Dpg2-B/II})-O^5(\text{Dpg2-B/II}')$, $O^5(\text{Dpg2-B/II})-O^5(\text{Dpg2-B/I}')$, which are also analogous to interactions in crystal form I. Comparison of the interfaces in terms of empty space they surround yields the same result with the exception that in this crystal form the head-to-head cavity is occupied by two isopropanol molecules.

BIOLOGICAL IMPLICATIONS

Feglymycin is a unique antibiotic in terms of sequence and antimicrobial spectrum. The structure determination revealed that it closely resembled the membrane ion channel peptide gramicidin. Feglymycin forms dimers in aqueous and also in predominantly apolar environments, and the dimer is very likely to be the biologically active unit. Under favourable conditions the dimers can associate to tetramers, but it needs still to be determined whether this is relevant to the function.

Channel formation

The width of biological membranes varies among the species, but it has been shown that a minimum size of about 30 Å is necessary to span bacterial membranes. This value is only approximate and compounds lying slightly under this limit have also been shown to form ion channels under certain conditions (Duclohier *et al.*, 1998). The gramicidin dimer in the

Cs⁺-complex, which is believed to be relevant to the membrane conducting form, is about 25 Å in length if one considers the C^α-atoms only or 28 Å if all the atoms are regarded and is long enough to span through the membrane. Other physico-chemical properties like hydrophobic surfaces essential for membrane integration and a hole formed by the structure that allows passage of ions smaller than a certain size also support this assumption.

In contrast, feglymycin has a shorter sequence and a wider hole; therefore the dimer is only 18 Å long measuring between the farthest C^α-atoms or 20 Å when all atoms are included. This is significantly less than required to span a biological membrane (Figure 4.8). Moreover, the channel in the feglymycin dimer is blocked by four amino acid sidechains (Figure 4.9). Two of them, namely the phenyl rings of Mpg1-A and Mpg1-B are shown to possess the necessary flexibility to open and close, but even minor displacements of the Phe12-A and Phe12-B sidechains are not feasible because of extensive steric hindrance.

Membrane processes

It is likely that feglymycin can easily integrate into the membrane as it has hydrophobic residues on its surface, and might function as an ion carrier rather than a channel between the two sides. There are ion carriers known that have alternating D,L-amino acids in their sequence, like valinomycin (Duax *et al.*, 1996). The transport is usually highly selective for a given anion or a cation, which is K⁺ in the case of valinomycin (Tosteson *et al.*, 1967). However, such selectivity is not expected for feglymycin, where the internal cavity can accommodate nearly any particles under a certain size. Nevertheless, as feglymycin can integrate into the membrane, it may also be able to penetrate it, and interact with enzymes in the cytosol. This mechanism may be important in the HIV-inhibitory activity and suggests a scheme for the transport of feglymycin into the infected cell.

Interaction surface

Concerning its structure, feglymycin is a perfectly tailored molecule for a specific interaction. First, the structure is considerably rigid, which is a conspicuous feature of virtually all natural inhibitor molecules. Secondly, the molecular organisation of the dimer utilises all Dpg residues in stabilising the structure, while the Mpg residues are involved in these interactions to a much lesser extent. On the other hand, Mpg residues are ideal in constituting a very well defined interaction surface: rotation around the C^α–C^β bond does not affect the position of its hydroxyl, which is therefore stationary. A well-defined hydrogen-

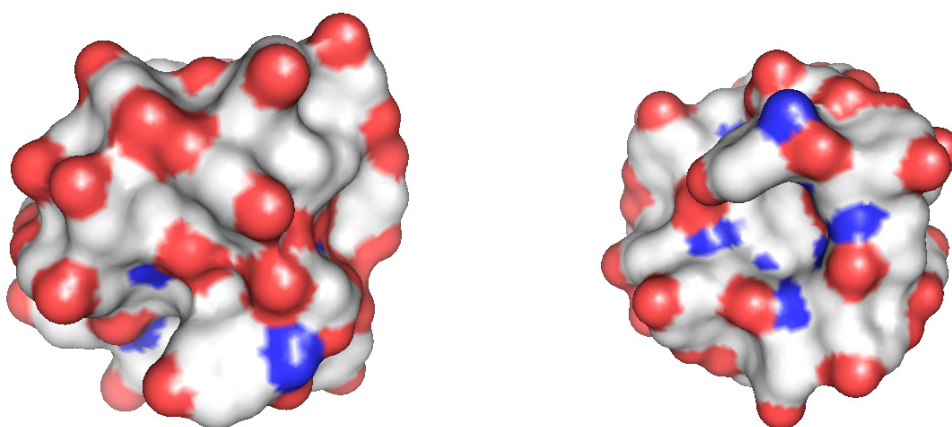


Figure 4.15

The surface of the feglymycin dimer showing potential interactions (hydrogen bonds and hydrophobic recesses). The orientation in (a) corresponds to Figure 4.8 and in (b) to that on Figure 4.9 and shows that the feglymycin channel is blocked. Carbon atoms are grey, nitrogen atoms are blue and oxygen atoms are red.

bond pattern is important in molecular recognition: hydrogen bonds provide the required specificity, even if the affinity is a result of an extensive number of non-specific, hydrophobic interactions. Moreover, hydrophobic patches are also present on the feglymycin surface: the valine residues constitute recesses on the molecular exterior at certain places that might be important in connecting feglymycin to its target (Figure 4.15). As feglymycin retains its conformation in both polar and apolar environments, it is also conceivable that it exerts inhibitory activity within the membrane, i.e. on the membrane-embedded surface of a membrane-bound or transmembrane protein. In the light of the foregoing evidence it seems therefore likely that feglymycin is not a membrane agent, but a strong enzyme inhibitor acting on a wide range of targets.

V. CRYSTAL STRUCTURE OF LIPOPEPTIDE ANTIBIOTICS

INTRODUCTION

Lipopeptide antibiotics constitute a class of antibacterial agents that are highly active against multiresistant bacteria. Amphomycin, the first member of the series, was discovered more than 50 years ago (Heinemann *et al.*, 1953), and was followed by the isolation of crystallomycin (Lomakina & Brazhnikova, 1959), aspartocin (Shay *et al.*, 1960), glumamycin (Shibata *et al.*, 1962), laspartomycin (Naganawa *et al.*, 1968), tsushimycin (Shoji *et al.*, 1968; Shoji & Otsuka, 1969) and other antibiotics belonging to the same group. However, probably because of their hemolytic properties and resulting toxicity, interest in this antibiotic class has waned, although the antibiotic daptomycin discovered by Eli Lilly progressed in the 1980s to clinical phase I and phase II trials. As reversible skeletal muscle toxicity was observed at high doses (Tally *et al.*, 1999) and treatment failures were experienced in patients with *Staphylococcus aureus* endocarditis (Garrison *et al.*, 1989), the development was suspended. More recently, the increased occurrence of infections caused by resistant Gram-positive bacteria led to renewed interest in daptomycin and in 1997 Cubist Pharmaceuticals licensed the drug from Lilly and reinitiated clinical trials. Daptomycin is currently being approved as an intravenous formulation (under the trade name Cidecin) in the USA, and introduction to the European market is intended in the near future (Rotschafer *et al.*, 1988; Bronson & Barrett, 2001; Strahilevitz & Rubinstein, 2002; Linden, 2002; Woodford, 2003; Abbanat *et al.*, 2003).

The chemical structure of amphomycins was first investigated in the 1960s. Performing chemical degradation studies on the antibiotic glumamycin, Inoue and Fujino (Inoue, 1962a; Inoue, 1962b; Inoue, 1962c; Fujino, 1965) determined the amino acid composition and the structure of the fatty acid component and proposed two possible sequences for the peptide core. Investigating the antibiotic amphomycin, Bodanszky *et al.* (1973) arrived at an alternative sequence, and revised the one reported for glumamycin by proving the identity of the two antibiotics. More recently, Vértesy *et al.* (2000) deduced the chemical structure of the related friulimicins by two independent spectroscopical methods and confirmed the correctness of one of the sequences suggested by Fujino (1965). According to this, amphomycin antibiotics possess a peptide framework consisting of a ten-membered

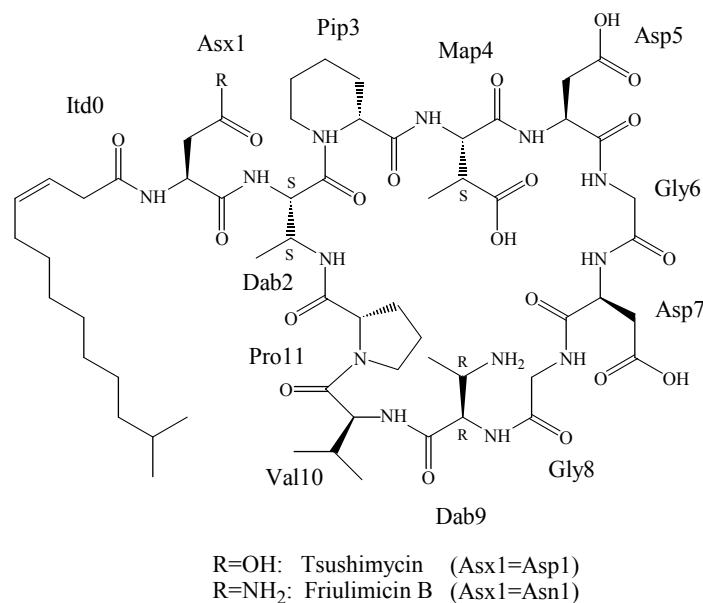


Figure 5.1

Chemical structure of tsushimycin and friulimicin B according to Vértésy *et al.* (2000).

cyclopeptide ring and an exocyclic amino acid, the N-terminus of which is acylated by a fatty acid residue (Figure 5.1). Differences among the group members arise principally in the structure of the fatty acid substituent. On the other hand, daptomycin (Figure 5.2) was shown to have a markedly different structure that shows homology to the general amphomycin sequence only in consisting a cyclic decapeptide core, which has aspartates at positions 4 and 6 relative to the attachment point of exocyclic amino acids, and the presence of a fatty acid residue (Debono *et al.*, 1988).

Amphomycin was found to be active against Gram-positive bacteria, but the exact mechanism of action has not been conclusively determined. Early studies indicated that amphomycin was a specific inhibitor of bacterial cell-wall synthesis, acting at the level of phospho-N-acetylmuramoyl-pentapeptide-transferase (MraY), which catalyses the transfer of the UDP-MurNAc-pentapeptide to the undecaprenyl carrier (Tanaka *et al.*, 1977; Tanaka *et al.*, 1979; Tanaka *et al.*, 1982). It was also found that amphomycin inhibited the synthesis of dolichol-linked saccharides in eukaryotes, which is essential in the glycosylation of glycoproteins that contain mannose and GlcNAc, although higher concentrations were necessary to achieve an effect similar to that on the bacterial enzyme (Kang *et al.*, 1978). It was proposed that amphomycin acts by binding the substrate dolichylmonophosphate with high affinity and in a Ca²⁺-dependent manner (Banerjee, 1989), as no binding was observed in the absence of Ca²⁺. Since dolichylphosphate is similar in structure to the undecaprenyl

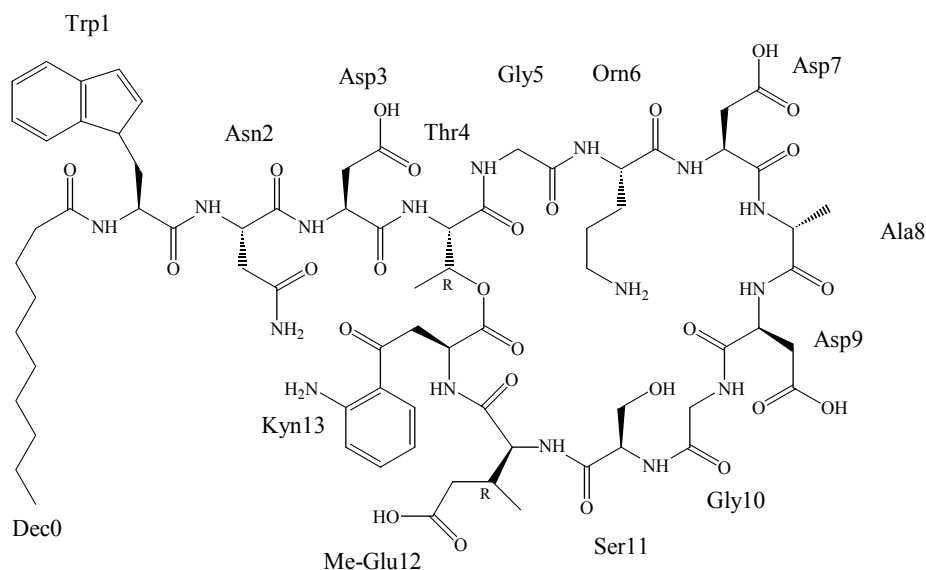


Figure 5.2

Chemical structure of daptomycin.

carrier in prokaryotes, it may be concluded that bactericidal action of amphomycin is achieved by complexing this compound and thereby preventing murein biosynthesis. The amphomycin-relative daptomycin was first reported to act by inhibiting peptidoglycan biosynthesis at the stage of formation of the nucleotide-linked sugar-peptide precursors (Allen *et al.*, 1987). However, it was pointed out by Canepari *et al.* (1990) that daptomycin does not enter the cell and suggested the synthesis of lipoteichoic acids occurring in the cytoplasmic membrane as the site of action. Teichoic acids (Figure 5.3) are present on the outer surface of Gram-positive bacteria and are involved in the circulation of divalent cations (Lambert *et al.*, 1977). Lilly scientists published results presenting evidence that daptomycin dissipates the membrane potential without affecting the chemical gradient across the membrane and proposed this as being the basis of antimicrobial action (Alborn *et al.*, 1991).

It is instructive to compare lipopeptides with polymyxin antibiotics. Similarly to lipopeptides, polymyxins are also cyclopeptide antibiotics, but consist of a seven-membered peptide ring, exocyclic amino acids and a fatty acid tail (Vogler *et al.*, 1965). In contrast to lipopeptides, polymyxins are polycationic species as most of the amino acids have amino groups on their sidechains. Polymyxins therefore readily bind anionic species like phosphatidylglycerols that results in a rapid permeability change of the cell membrane, release of cellular materials and subsequent cell death (El Mashak & Tocanne, 1980; Yin *et al.*, 2003).

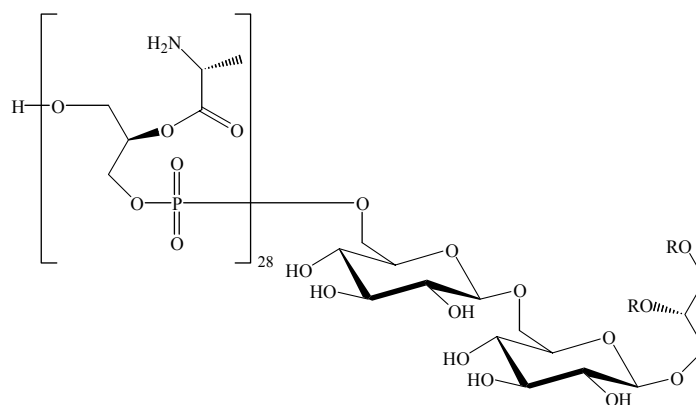


Figure 5.3

Structure of lipoteichoic acid of *Staphylococcus aureus*. R represents a long alkyl chain.

Lipopeptide antibiotics possess superior antibacterial potency, as they are rapidly bactericidal not only against multiresistant *Staphylococcus aureus*, but also against *Enterococcus* spp, a property that none of the currently licensed antibiotics can claim if administered alone. Resistance to lipopeptide antibiotics is rare; daptomycin, the only member being evaluated in clinical trials is active against 99% of tested clinical isolates; in addition, the distribution of minimum inhibitory concentration values were unimodal, and therefore the existence of a defined resistance mechanism is unlikely (Barry *et al.*, 2001). The level of Ca^{2+} ions in the medium proved to be important when tested against less susceptible isolates, as in trials on Ca^{2+} -supplemented media corresponding to physiological Ca^{2+} -levels minimum inhibitory concentration values were found to be two- to fourfold lower. Lipopeptides have limited stability under physiological conditions; spontaneous chemical reactions readily occur because of the presence of the Asp-Gly sequence pattern, which occurs more than once in certain members, and results in the loss of activity (Geiger & Clarke, 1987). This property provides a mechanism for self-clearance from the environment and may contribute to preventing the development of resistance. These antibiotics have not been employed as food additives in animal husbandry, so resistance could not have developed in enterococci and spread in the society. Given the superior antibacterial activity primarily against problematic microbes and inherent ways for keeping development of antibacterial resistance at a low level, lipopeptides are perfect antibiotics for reinforcing or replacing glycopeptides as a last line of defence against deadly bacterial infections.

STRUCTURE DETERMINATION

Crystallisation

Three different samples belonging to the group of lipopeptide antibiotics were kindly provided by Aventis Pharma Deutschland in development-scale quantities: A-1437 B (probably identical with tsushimycin, 93% pure) and A-1437 D (friulimicin) as Na⁺-salt (85% pure) and as Ca₂Cl₂-salt (93% pure). The slight impurity is a consequence of isomerisation reactions at aspartate residues and since it proved excessively cost-intensive, no further purification was attempted.

Lipopeptide antibiotics are known to form crystals, and in the purification protocol of friulimicin crystallisation was included as the final step. It was shown that only the active antibiotic crystallises under the established conditions and thereby can be separated from the isomerised, inactive substance. As starting material, the Na⁺-salt of friulimicin is used, which is dissolved in water, pH-adjusted to 6.4 and converted to the Ca²⁺-form by the addition of CaCl₂. Crystallisation is initiated by slowly increasing the ethanol content of the solution.

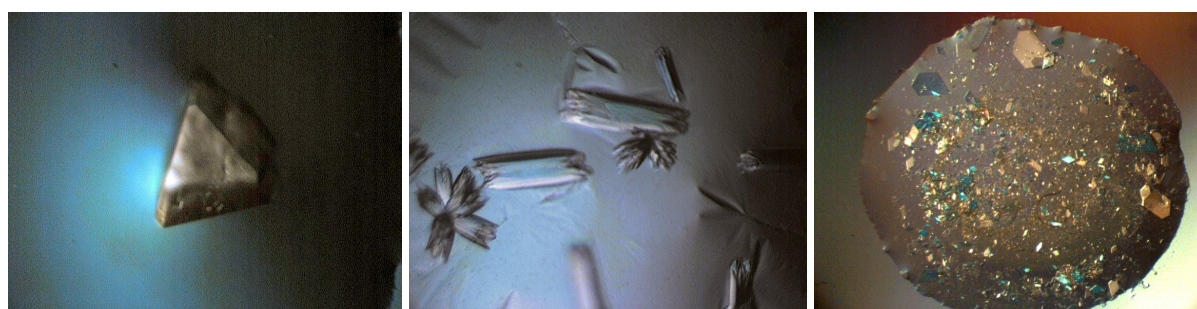
As sufficient amounts from all samples were available, a slightly modified procedure was tried in a batch-crystallisation setup. Since slow addition requires adequate instrumentation, which was not available, 250 µl of an aqueous solution containing 40 mg/ml Na⁺-friulimicin was simply mixed with an equal volume of the crystallisation condition in reservoirs of tissue culture plates. The crystallisation condition consisted of 0.1 M Na-cacodylate/cacodylic acid pH=6.0 or 0.1 M NaAc/HAc pH=5.4, 0.2 M CaCl₂ and 30-50% acetonitrile or ethanol. Large needle crystals appeared within a day. However, picking crystals out of the reservoir proved complicated and quality of the crystals was not sufficient and therefore this approach was discontinued.

As a method of choice, the hanging drop method was selected, and the foregoing procedure was modified accordingly. Unfortunately, no improvement in crystal quality was observed and for this reason, screening was performed with all available kits (Hampton Crystal Screen I, Crystal Screen II, JBScreen 7, 8 and 9). Crystals were obtained from nearly all conditions that contained CaCl₂ and an organic solvent. The most promising crystals were grown from 0.1 M NaAc/HAc pH=4.6, 0.2 M CaCl₂ and 30% isopropanol. It is noteworthy that according to crystallisation experiments performed at Aventis Pharma, isopropanol proved not suitable as an organic solvent since it led to amorphous precipitation. However, replacement of isopropanol in the foregoing condition with the recommended ethanol or

acetonitrile yielded no crystals. Since it was of higher purity and calcium was essential for crystallisation, the Ca_2Cl_2 -form was used in further experiments. After optimisation, well-formed block crystals were obtained from 0.1 M NaAc/HAc pH=4.6, 0.2 M CaCl_2 , 30% isopropanol and 10% glycerol that diffracted to 3-3.5 Å resolution on a rotating anode (crystal form **I**, Figure 5.4).

The crystals were not very stable at room temperature and began to decompose after one week. This observation hinted that the stability of the substance is not sufficient at room temperature and diffraction limit may be improved by performing crystallisation experiments in the cold room. As crystallisation conditions vary with the temperature, first a coarse grid screen was performed and new conditions were then optimised. Block crystals having a different habit than those obtained at room temperature grew in the course of several days from 0.1 M NaAc/HAc pH=4.6, 0.14 M CaCl_2 and 14% isopropanol. These crystals diffracted much better (1.8 Å on a rotating anode, atomic resolution at BESSY), but proved to be of another crystal modification and were non-merohedrally twinned (crystal form **II**, Figure 5.4).

To promote the growth of crystals belonging to crystal form **I**, a step-wise cooling procedure was selected. A coarse grid-screen was prepared and the temperature was held constant at 12 °C in a Hampton incubator. Crystals resembling in shape the ones grown at room temperature were obtained from 0.1 M NaAc/HAc pH=4.6, 30% isopropanol, 10% glycerol and 0.10 M CaCl_2 . Based on this finding a modified coarse screen followed by optimisation was performed at 4 °C. Crystals identical in morphology to that of crystal form **I** were obtained from 0.1 M NaAc/HAc pH=4.6, 38% isopropanol, and 0.06 M CaCl_2 within a week. It must be noted that crystal formation was extensively sensitive to Ca^{2+} -concentration, as no crystals grew outside the range of 0.05-0.07 M CaCl_2 . These crystals were very fragile and to avoid mechanical stress upon freezing, crystallisation conditions were modified to



Friulimicin crystal form I

Friulimicin crystal form II

Tsushimycin

Figure 5.4

Crystal morphologies of friulimicin and tsushimycin.

include the cryoprotectant to 0.1 M NaAc/HAc pH=4.6, 32% isopropanol, 0.30 M sucrose and 0.06 M CaCl₂. These crystals indeed belonged to crystal form **I** and diffracted to 2.5 Å in-house and to 1.6 Å at BESSY.

Examination of the crystallisation procedure received from Aventis Pharma revealed one more possibility for the improvement of crystal quality. It was found that the decomposition rate depends on the pH, being the fastest at the isoelectric point (pH=4.4) and decreasing steeply towards neutral pH. As the crystallisation was not excessively sensitive to alteration of the pH, the buffer was changed and comparable quality crystals were grown from Na-cacodylate/cacodylic acid pH=6.0, 40% isopropanol and 0.08 M CaCl₂. Unfortunately, no significant improvement in diffraction limit was observed.

Independently of friulimicin, crystallisation experiments were started with tsushimycin as well. Although the difference between friulimicin and tsushimycin are marginal, there is a considerable difference in their toxicity and antimicrobial potency. It was also observed when working with peptaibols that one-atom modifications can substantially alter crystallisation properties and therefore tsushimycin crystals might be the means of overcoming the difficulties posed by friulimicin. As no crystals were obtained from any of the conditions used for friulimicin, the procedure from Aventis Pharma was used as a starting point. Small crystals were obtained from 0.1 M NaAc/HAc pH=4.6, 30% ethanol and 0.4 M CaCl₂. It was recognised early that crystallisation demanded a lower pH, since the isoelectric point of tsushimycin is at about pH=3.6, where acetate has little buffer capacity. To retain reproducibility, the pH was set by mixing 1.0 M NaAc/HAc pH=4.0 with different amounts of 1.0 M HAc. Thin plate-like crystals were grown from 0.1 M NaAc/HAc pH=4.0, 0.12 M HAc, 38% ethanol and 0.80 M CaCl₂ (Figure 5.4). These crystals were grown at room temperature and were sufficiently stable, starting to decompose after 6 months. To support cryoprotection, later trials were supplemented with 0.40 M 1,6-hexanediol.

Measurement

Friulimicin crystal form I

Since crystals were grown in the cold room and sucrose had been included in the condition provided sufficient cryoprotection, a suitable crystal was selected, taken out with an adequately sized loop and dipped into a Dewar vessel full of liquid nitrogen. The crystal was characterised by collecting datasets on a Bruker rotating anode, Osmic focusing mirrors and Bruker SMART6000 4K CCD detector. Diffraction was observed to 2.5 Å resolution.

However, problems arose in unit cell determination as different indexing algorithms implemented in Proteum gave different answers that all fitted the crystal lattice within experimental error. Difference vector-based algorithms yielded a primitive trigonal cell with cell dimensions $a=b=73.4$ Å and $c=105.7$ Å, while the Fourier-algorithm gave a rhombohedral cell with $a=b=c=63.2$ Å and $\alpha=\beta=\gamma=71^\circ$ (corresponding to $a=b=73.6$ Å and $c=140.1$ Å in hexagonal setting). It turned out later that the crystal had an unusual intensity distribution and strong reflections constituted a sub-lattice; this property cause algorithms employing Fourier-transformation to give the wrong answer. However, the program cell_now indexes the lattice routinely and selects automatically the right primitive trigonal cell.

After the ambiguity with the unit cell had been resolved, datasets were collected in order to extract the anomalous signal originating from the incorporated Ca^{2+} -ions. Three φ -scans and five ω -scans were performed, which provided two approximately equivalent datasets in terms of data redundancy and precision. Integration was performed using SAINT and scaling with SADABS. The correlation coefficient between the signed anomalous differences calculated independently from the two types of scans was calculated and then all runs were scaled together in a subsequent SADABS run and anomalous differences were derived from the complete dataset with XPREP (Bruker Nonius, 2002).

Not regarding high apparent mosaicity, crystal quality seemed satisfactory and a strong anomalous signal was expected, but it was indicated by the statistics that the signal extends to only 4.0 Å resolution. It was realised at this point that the unusual intensity distribution ($|E^2-1|=1.215$) might be responsible for this result: with more than 50% of the reflections below 2σ , the crystal behaves like a centred lattice but as the pseudo-"systematically absent" reflections do not follow a regular pattern, they cannot be filtered out. This has two consequences on data processing and statistics:

- (a) Reflection profiles. In SAINT, all reflections contribute to the determination of standard reflection shapes, which in this case is suffering from the overwhelming number of "missing" reflections. A solution might be the exclusion of weak reflections from this procedure (or not to use profile fitting at all, e. g. as in the program EVALCCD) or the determination of a real-space sub-lattice, which can be used to solve the phase problem and then convert the results to the original cell. This latter solution is based on the assumption that pure translational symmetry is present in the unit cell, and selecting a sub-cell would only convert the non-crystallographic symmetry into crystallographic one, and intensity distribution would then follow

regular Wilson-statistics. When a solution is achieved, it can be extended to the original cell and thus overcome the phase problem.

- (b) Distorted statistics. As there is a high number of reflections with low intensity, anomalous differences calculated from them are also weak and unreliable. As the programs print an average value for each resolution shell, this value is artificially low, since it includes all reflections, half of which do not contribute to the intensity and only increase the standard deviation. However, the weak reflections are not used in anomalous substructure solution, since direct method programs select only the strongest 10% of the data, therefore a resolution cutoff at a significantly higher resolution may be justified.

Derivatisation of the crystal with Sr^{2+} or Ba^{2+} , which are also natural substrates of friulimicin proved unsuccessful, as no crystals were obtained. However, co-crystallisation with iodide yielded crystals comparable in quality to the original ones. It still remains to be established whether these crystals contain any iodide or not.

A higher resolution synchrotron dataset has also been collected at the BL1 beamline at BESSY. Indexing was performed using XDS (Kabsch, 2001), which utilises a similar indexing algorithm implemented in *cell_now* and therefore had no difficulty in finding the unit cell. XDS was also used for integration and scaling, since crystal mosaicity was relatively high and while processing with other packages was suffering from this property, the integration algorithm in XDS could handle such a case routinely. Data were collected to 1.6 Å resolution, which would allow phase extension and refinement if starting phases could be obtained (Table 5.1).

This crystal thus resisted all attempts to solve the structure so far. However, recent availability of novel data processing programs, a structural model for the related tsushimycin and promising iodide-derivatised crystals may help in overcoming the phase problem and determine the structure. The low resolution of anomalous data is surely a hindrance as a large number of Ca^{2+} -ions should be located in a moderately large unit cell, i.e. the data-to-parameter ratio is too low. Search algorithms based on the Patterson-function may perform better in these cases, as direct methods decrease the number of reflections further by selecting the strongest 10% of the data and may not be optimal for such a problem.

	Friulimicin crystal form I		Tsushimycin
Beamline	In-house	BL1@BESSY	BL1@BESSY
Unit cell parameters	a=b=73.426 Å, c=105.699 Å	a=b=73.175 Å, c=106.508 Å	a=33.491 Å, b=36.386 Å, c=37.518 Å $\alpha=65.64^\circ$, $\beta=68.35^\circ$, $\gamma=69.90^\circ$
$\langle E^2-1 \rangle$	1.215	0.932	0.741
Space group	P3 ₁ 21 or P3 ₂ 21		P1
No. of reflections	449050	402457	355145
No. of unique reflections	13029	41712	75901
Resolution (Å)	2.40 (2.50-2.40)	1.63 (1.63-1.75)	1.00 (1.10-1.00)
Completeness	96.8% (80.5%)	99.7% (99.0%)	96.2% (92.1%)
Redundancy	33.37 (3.25)	9.62 (6.97)	4.50 (3.62)
$\langle I/\sigma(I) \rangle$	21.85 (1.74)	9.67 (3.05)	9.49 (3.35)
R_{int}	0.0813 (0.7327)	0.1432 (0.6760)	0.1039 (0.4990)

Table 5.1

Data collection statistics. The significant difference in the $\langle |E^2-1| \rangle$ value between the two friulimicin datasets indicates either a deviation from perfect translational symmetry at higher resolution or the failure of profile fitting because of an extensive number of weak reflections in the dataset collected at BESSY. $R_{\text{int}} = \Sigma |I - \langle I \rangle| / \Sigma I$.

Friulimicin crystal form II

Not risking possible adverse effects, crystallisation trays with crystals were not taken out from the cold room. A crystal was selected and transferred from the drop into cryoprotectant solution consisting of 0.1 M NaAc/HAc pH=4.6, 14% isopropanol, 0.14 M CaCl₂ and 0.62 M sucrose. After a short soaking time, the crystal was taken out with a loop and dipped into a vessel filled with liquid nitrogen and mounted on a Bruker rotating anode, equipped with Osmic focusing mirrors and Bruker SMART6000 4K CCD diffractometer. The crystal diffracted relatively well, as spots were observed to 1.8 Å resolution. Unit cell determination posed an enormous problem, though. For that purpose, several frames were collected at a comparatively longer distance to promote separation of spots and were thresholded by Proteum. The file containing the spot positions was examined with RLATT and possible unit cell vectors were searched. The spot distribution suggested the presence of a cell vector with 28 Å length, but nothing else could be concluded from the layers. Vector search was then performed with cell_now, using a relatively wide interval (20-220 Å) and a fine grid (99999). This resulted in a long list of vectors, which were

interpreted by hand. As a unit cell vector around 28 Å seemed likely, vectors were searched that made a special angle (90° or 120°) with this, appeared twice in the list and could be consistently divided into two sets that were related by a single rotation. A vector with a length of 93 Å and another 84 Å long one were found and the initial unit cell thus established were $a=93.3$ Å, $b=28.5$ Å, $c=84.0$ Å, $\alpha=\gamma=90^\circ$ and $\beta=123.5^\circ$. This cell was then docked by `cell_now`; three domains were located with increasing figure-of-merit and a sudden drop at the fourth domain. It was therefore concluded that three domains were present. Examination of reciprocal lattice with `RLATT` revealed not indexed rows of reflections exactly between indexed ones, which hinted that one unit cell vector was twice as long. The final unit cell at that point corresponded therefore to $a=93.1$ Å, $b=28.5$ Å, $c=168.3$ Å, $\alpha=\gamma=90^\circ$ and $\beta=123.5^\circ$. Integration was performed with all three domains using `SAINT`, and the resulting data were scaled with `TWINABS`. Examination with `XPREP` revealed that reflections corresponding to a C-centred lattice were significantly weaker, and integration was repeated using C-centred lattice setting. A substantial improvement in integration statistics was observed, which confirmed the assumption. Moreover, a C-centred cell with aforementioned parameters could be converted into a C-centred orthorhombic cell with cell parameters $a=28.5$ Å, $b=93.1$ Å and $c=140.4$ Å. This cell was located with `cell_now`, and the twin law was determined that corresponded to a 180° rotation about the two diagonals of the bc plane. Unfortunately, this operation does not create a closed group, i.e. rotations about free diagonals of domains would generate an infinite number of cells. In all cases the presence of only three domains could be proved, but it is not known whether only these or other domains are also present that disappear in the noise (Figure 5.5).

This crystal has also been measured at the BL1 beamline at BESSY; diffraction maxima could be observed to atomic resolution. A dataset was collected at a relatively long distance and processing was attempted using `XDS`. `XDS` could automatically locate a unit cell corresponding to the primitive setting of the orthorhombic C-cell, and thus confirmed the correctness of the indexing procedure. The dataset has not been processed yet, but the recent availability of programs handling overlapped lattices (Duisenberg *et al.*, 2003) now provides a possibility of solving this crystal structure as well. A dataset extending to atomic resolution could be detwinned accurately enough to provide a starting point for direct methods, which might even solve such a large asymmetric unit considering the high number of Ca^{2+} -ions present in the structure.

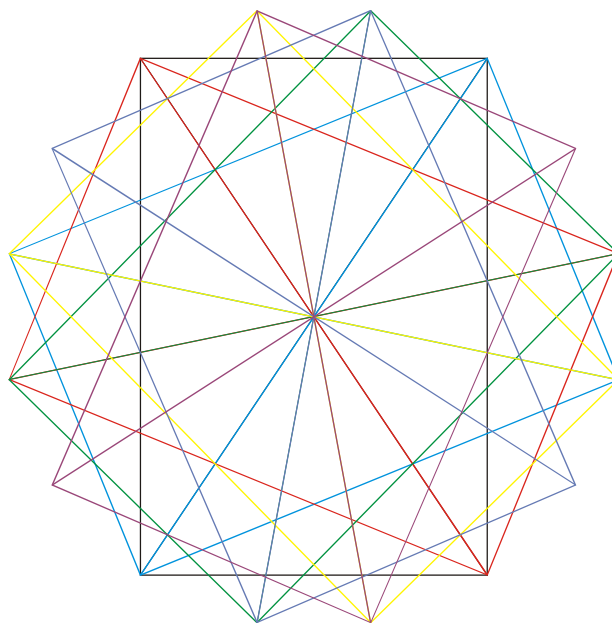


Figure 5.5

Twinning in friulimicin crystal form II with a twofold rotation about the diagonal of the *bc* plane. Only seven possible domains are shown, but theoretically the operation would generate an infinite series.

Tsushimycin

Although thin and small, tsushimycin crystals were relatively resistant to mechanical impacts. As crystals grew at room temperature, a suitable one was selected, soaked in cryoprotectant containing 0.1 M NaAc/HAc pH=4.0, 0.05 M HAc, 34% ethanol, 0.6 M 1,6-hexanediol and 0.8 M CaCl₂ when necessary, and mounted on our rotating anode. Despite the small crystal size, diffraction was observed to comparatively high resolution, as even at 1.6 Å resolution some spots could be seen. However, the crystal exhibited anisotropic mosaicity, i.e. in certain directions a good-looking diffraction pattern was observed, while in other directions streaks rather than spots (Figure 5.6). Annealing of crystals was tried, but no convincing improvement was observed. Previous experience taught that appropriate integration algorithms could easily handle anisotropic mosaicity, and therefore such a crystal was taken to BESSY and datasets were collected on the BL1 beamline using comparatively thin slicing. The unit cell and space group had already been established in-house of being the lowest possible symmetry, which made a special data collection strategy necessary. It also turned out that the crystal was significantly sensitive to radiation damage and decomposed in the beam after collecting 90° of rotation range. For these reasons, a new crystal was mounted on a goniometer head equipped with an arc, the beam size was reduced to 0.1·0.1 mm² to

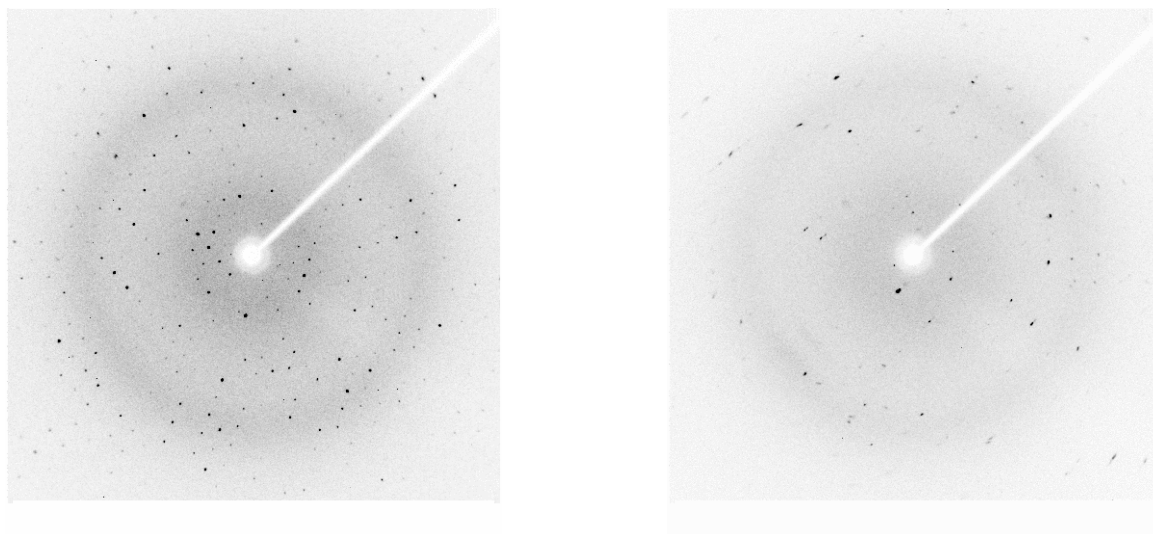


Figure 5.6

Diffraction frames from tsushimycin showing anisotropic mosaicity. In one orientation highly regular spots are observed, while after 90° rotation, the spots exhibit an elongated profile. However, modern integration programs can easily handle such property.

spare diffracting volume, and five runs were collected that each covered a rotation range of 90° and differed in crystal orientation. Between the runs, the crystal was re-centred to irradiate non-damaged parts. In the sixth run, low-resolution reflections were collected with short exposure times to avoid overloads. Processing was performed with XDS, separate runs were scaled together using XSCALE. Statistics generated by XPREP indicated that the data extended to atomic resolution with an $I/\sigma(I)=3.3$ in the highest resolution shell (Table 5.1).

Structure solution

The cell of tsushimycin is comparatively small, but considering the low symmetry, the size of the asymmetric unit corresponds to that of an average protein crystal. As a comparison, the P1 form of lysozyme crystallises in a unit cell with $a=26.65 \text{ \AA}$, $b=30.80 \text{ \AA}$, $c=33.63 \text{ \AA}$, $\alpha=88.3^\circ$, $\beta=107.4^\circ$ and $\gamma=112.2^\circ$, which is about 35% smaller in volume than the one of tsushimycin ($V=24279 \text{ \AA}^3$ vs. $V=37708 \text{ \AA}^3$). This crystal form of lysozyme could easily be solved using *ab initio* direct methods based on data available to 0.95 \AA ; the presence of several sulfur and particularly chlorine atoms together with P1 symmetry support the process. As tsushimycin binds Ca^{2+} -ions and a dataset comparable to that of lysozyme in terms of resolution was available, it was expected that structure solution would work as plainly as for lysozyme.

Previous experience suggested that *ab initio* direct method solution works best if a small fragment (preferably the heavy atoms) is located first with dual space recycling and then the structure is expanded using peaklist optimisation. Thus, a SHELXD (Sheldrick *et al.*, 2001) job corresponding to this strategy was set up that instructed the program to find 30 atoms in the dual-space recycling (FIND 30), reject tries if the correlation coefficient is below a certain threshold (TEST 15 5) and then expand the structure with peaklist optimisation from 30 to 1300 atoms in ten steps. An immediate solution has been achieved: all tries out of the performed nine led to the right solution (and therefore no bimodal distribution has been obtained). As one try required about 2 minutes of CPU times on a 2.8 GHz Intel Xeon, this corresponds to a solution rate of about 30 solutions/hour, which is much higher than those observed for cephaibols (about 0.5 solution/hour) not to mention that for feglymycin (0.016 solution/hour) although tsushimycin is two- to tenfold larger. The presence of heavier atoms therefore assists substantially the structure solution and hints that under certain circumstances structures containing a few heavier elements besides the protein could be solved even if data do not extend to atomic resolution.

Model building and refinement

The structure provided by SHELXD was incomplete and about 20% of atoms were missing. First, peaks were assigned corresponding to peptide atoms according to the sequence reported by Vértessy *et al.* (2000) using the program XP. To build up the missing part of the structure, a map was created using SHELXE (Sheldrick, 2002) starting from the direct method solution and specifying a relatively low solvent content (25%). A very clear map was obtained and missing atoms were added manually using the program XtalView (McRee, 1999). The structure was refined against all F_o^2 with no intensity cutoff using SHELXL (Sheldrick & Schneider, 1997). Throughout the refinement, bond length, angle distance, chiral volume and planarity restraints were imposed on appropriate atoms. As most of the residues are not commonly found in proteins, bond length and bond angle distances were generated using small molecules from the CSD (Allen, 2002). When no suitable fragment was found (e.g. for isotetradecenoic acid), restraints were generated using a table compiled by Allen *et al.* (1992) listing common bond lengths found in organic compounds. Solvent atoms were added using SHELXWAT and manually. All non-hydrogen atoms were refined anisotropically with suitable rigid bond and similarity restrains. For solvent waters, approximately isotropic restraints were also employed. Hydrogen atoms were included in the later stages of

Resolution range (Å)	26.02-1.00
R_1 [$F > 4\sigma(F)$ /all data]	0.1318/0.1464
R_{free} [$F > 4\sigma(F)$ /all data]	0.1640/0.1791
No. of parameters	11407
No. of restraints	12261
No. of peptide atoms	1069
No. of non-H atoms	1283
Solvent content	40.0%
R.m.s.d. from ideal geometry	
Bond lengths (Å)	0.014
Angle distances (Å)	0.033
Common planes (Å)	0.430

Table 5.2

Summary of final refinement for tsushimycin.

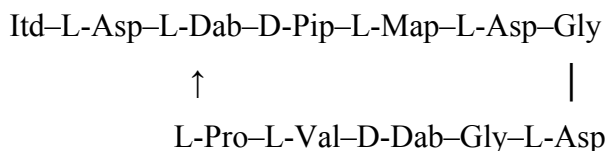
refinement. About 5% of the reflections (selected in thin shells) were omitted from the refinement and reserved for the calculation of R_{free} . Refinement statistics are shown in Table 5.2.

STRUCTURE DESCRIPTION

Amino acid residues and sequence

The absolute configuration of tsushimycin was assigned based on the configuration of chiral centres established by Bodanszky *et al.* (1969). Accordingly, aspartate, β -methyl-aspartate, proline and valine residues are present as L-enantiomers, the pipecolinic acid as D-enantiomer, while from 2,3-diaminobutyric acid two diastereomers are present: Dab2 is the L-threo-form, while Dab10 is D-erythro. This was in good agreement with the X-ray structure, although the configuration of C^β (Dab10) could not be determined unambiguously, as the scattering properties of carbon and nitrogen are very similar, and it proves to be problematic to distinguish these elements even at such high resolution. However, characteristics in the chemical environment allowed identification of N^γ (Dab10), since it was found to be closely situated to two charged aspartate sidechains (Figures 5.12 and 5.13).

On the other hand, the high resolution allowed unambiguous determination of atom connectivity. The measurement confirms the correctness of one of the two structures proposed originally by Fujino (1965), which has recently been supported by spectroscopic evidence (Vértesy *et al.*, 2000; Kong & Carter, 2003). The sequence of tsushimycin can therefore be written as:



The high number of aspartates common in Ca^{2+} -binding proteins is noteworthy, together with the Asp-Gly pattern that occurs twice in the peptide ring. This leads to decreased stability under physiological conditions with half-lives in the order of one week. The structure of the fatty acid sidechain could also be confirmed together with the geometry of the double bonds, which is present as the *cis*-form.

Conformation

In the asymmetric unit, twelve independent molecules are present. Each molecule binds a Ca^{2+} -ion and association of the molecules is mediated by additional Ca^{2+} -ions. Peptide atoms are well defined in the electron density and only a handful number of discretely disordered sidechains were found. On the other hand, most of the fatty acid residues are continuously disordered, although there are molecules whose long alkyl chains are clearly visible. Crystal packing is tight, but there are large cavities among the molecules, filled with water or with disordered alkyl chains.

Overall structure

The molecules were found to form a globule with the Ca^{2+} -ion in the middle and the peptide chain covering it. The backbone conformation resembles that of K^+ -bound valinomycin, i.e. a low-order Lissajoux-curve, taking into account that the cyclopeptide core in valinomycin consists of 12 residues and therefore it comprises three "waves" (Duax *et al.*, 1996), while tsushimycin contains only two (Figure 5.7). Valinomycin exhibits a regular alteration of amino acid chirality according to the pattern $-(\text{D-L-L-D})-$, but the presence of such a pattern cannot be demonstrated in tsushimycin, although the Gly residues would allow some degree of freedom in assigning chiralities in order to obey this regular alternation.

Conformations of independent molecules were compared by identifying rigid atoms with ESCET (Schneider, 2002), fitting these rigid atoms on each other, transforming the

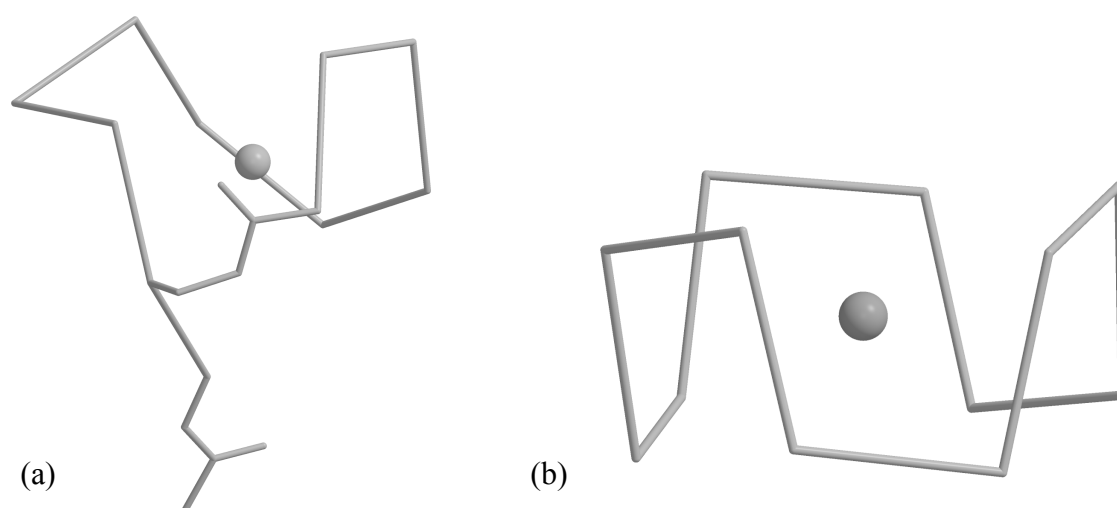


Figure 5.7

Backbone traces of (a) tsushimycin and (b) valinomycin showing homologous backbone conformations. The long fatty acid sidechain of tsushimycin has been truncated for clarity.

whole molecule with the determined operation and evaluating the results graphically. Although a high degree of conformational diversity was expected, surprisingly good agreement was found among independent molecules: the cyclopeptide ring adopts a nearly identical conformation in all 12 cases, and even the floppy sidechains are in a highly homologous arrangement (Figure 5.8). This finding suggests that the structure of the antibiotic with a bound Ca^{2+} is fairly well defined. Closer examination allowed classification of molecules into two classes, based on the sidechain conformation of Asp1.

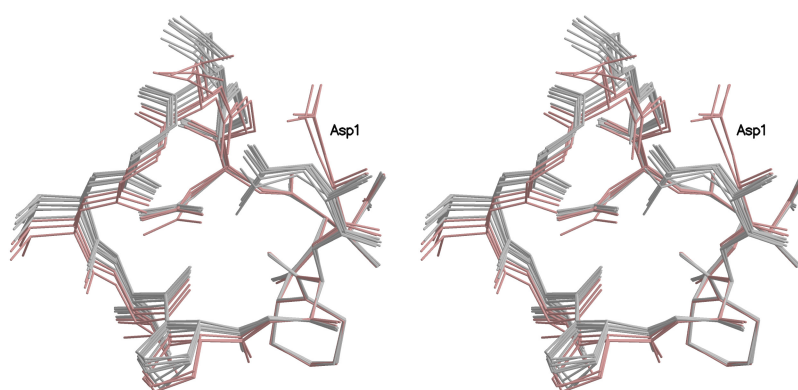


Figure 5.8

Stereo view of the superposition of all 12 independent tsushimycin molecules. Molecules belonging to different conformation classes are distinguished by their colour (class I molecules are grey, class II molecules are redbrown). The aspartate sidechain (Asp1) decisive in the assignment of conformation class is marked with a label.

Calcium binding

The Ca^{2+} -ion present in the central cavity of the molecules is coordinated by several oxygen atoms. Four of these are backbone carbonyl oxygens, one of them is from a water molecule, and one or two carboxyl oxygens from aspartate sidechains complete the coordination shell. The aforementioned two conformation classes also differ in the number of oxygen atoms binding to the Ca^{2+} -ion, since the sidechain of Asp1 also participates in Ca^{2+} -coordination in one class (class I), while it is not sufficiently near in the remaining one (class II). It is surprising that in spite of the presence of several aspartate sidechains, mainchain carbonyl atoms play a decisive role in Ca^{2+} -binding. The arrangement of these corresponds to that in valinomycin, although in valinomycin there are six such oxygen atoms, while in tsushimycin there are only four, corresponding to two coordinating mainchain carbonyl oxygens in each "wave" of the Lissajoux-curve.

In molecules belonging to class I, seven oxygen atoms coordinate the Ca^{2+} -ion with pentagonal bipyramidal geometry (Figure 5.9). Calcium-oxygen distances vary in the range 2.28-2.48 Å, with the water molecule being the most distant, while oxygen atoms from the antibiotic being nearer to the Ca^{2+} -ion. Equatorial positions are occupied by backbone carbonyl oxygens O(Dab2), O(Gly6), O(Gly8), O(Va10) and the water molecule; the aspartate carboxyl oxygens $\text{O}^\delta(\text{Asp1})$ and $\text{O}^\delta(\text{Asp5})$ are in axial positions, probably because of the larger sterical demand of the carboxyl group. Axial oxygens are found to be nearer to

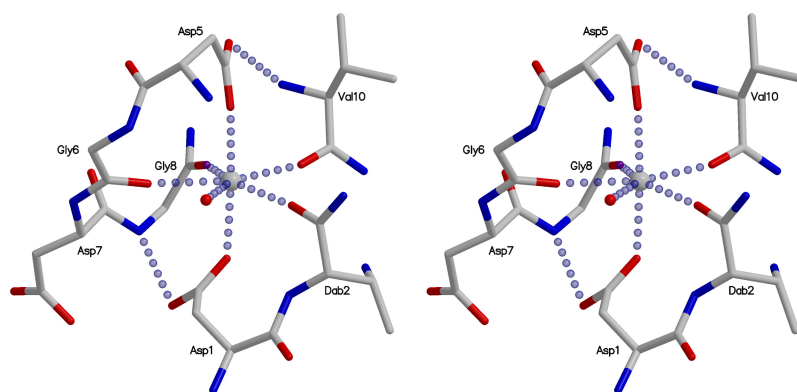


Figure 5.9

Stereo view of calcium binding in class I molecules. The Ca^{2+} -ion is coordinated by seven oxygen atoms according to pentagonal bipyramidal geometry. Sidechains of Asp1 and Asp5 are in axial positions, while the water molecule and mainchain carbonyl oxygens are equatorial. Residues not participating in the interaction are not shown.

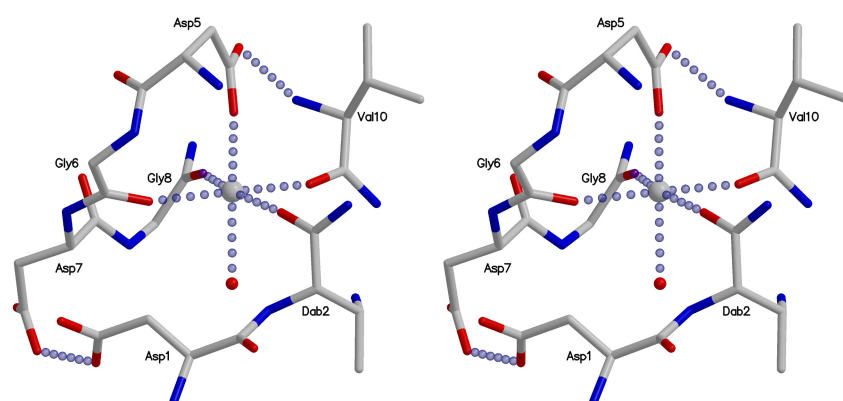


Figure 5.10

Stereo view of calcium binding found in class II molecules. The Ca^{2+} -ion is now surrounded by six oxygen atoms according to octahedral geometry. The sidechain of Asp1 is displaced and cannot participate in the binding, but is substituted by a water molecule. Residues not taking part in the interaction have been omitted for clarity.

the Ca^{2+} -ion (2.36 and 2.28 Å for Asp1 and Asp5, respectively) than equatorial ones, which suggests that these groups are deprotonated and negatively charged at the pH of the crystallisation. Bond angles between axial and equatorial oxygens are in the range 83-98°, while those between equatorial groups vary between 65-79°, with that between the water molecule and neighbouring carbonyl oxygens being the smallest, in accordance with the water oxygen being farther away.

On the other hand, class II molecules bind the Ca^{2+} according to a regular octahedral geometry (Figure 5.10). The carboxyl oxygen of Asp1 is displaced and does not participate in Ca^{2+} -binding. However, it is substituted by a water molecule, while the equatorial water molecule present in class I Ca^{2+} -coordination is missing. Remaining coordinating atoms keep their positions, although a slight closure of the peptide chain can be observed, as distances from the Ca^{2+} -ion are now 2.26 Å (for the carboxyl oxygen from Asp5) or between 2.30-2.36 Å (carbonyl oxygens). The water molecule with a binding distance of 2.53 Å is significantly more distant than any of the other coordinating atoms. There are slight changes in the bond angles: the one between "equatorial" substituents and "axial" ones is between 86-97°, while angles between "equatorial" atoms vary in the range 80-106°, the largest being between O(Dab2) and O(Gly6), where the equatorial water in class I molecules binds. It may also be speculated that the water molecule in this case is in fact a chloride to balance the charges, but this is not likely, since refined as water, temperature factors of the atom in question lies in the range observed for neighbouring water molecules.

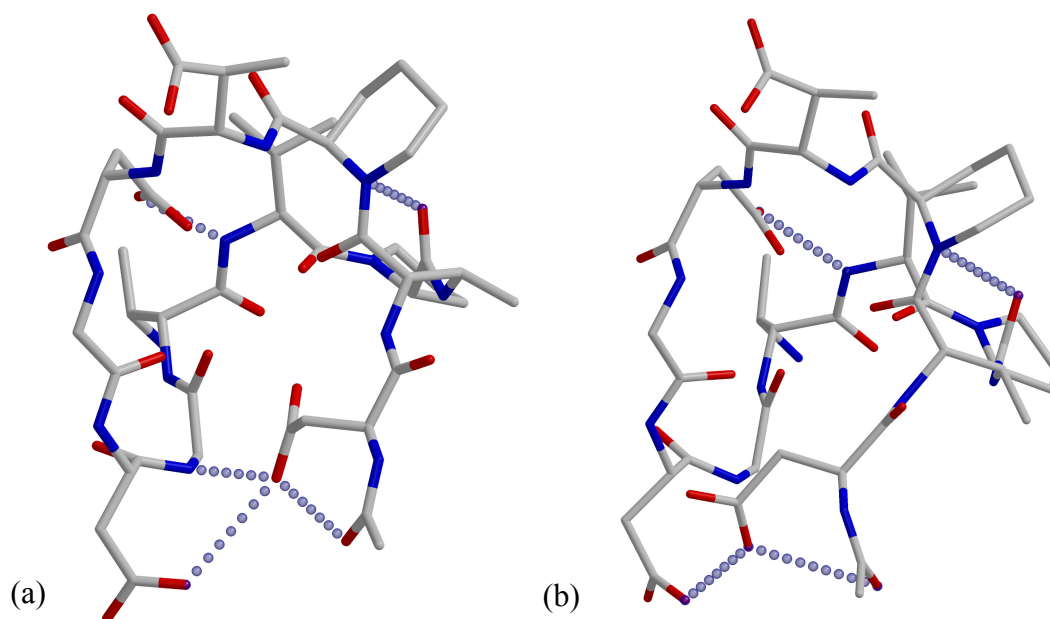


Figure 5.11

Secondary interactions that do not involve the central Ca^{2+} -ion in class I (a) and class II (b) molecules. The sidechain of Asp1 (bottom of the picture) can be stabilised in two distinct conformations. The Ca^{2+} -ion has been omitted to achieve a clearer view.

The two binding modes may also be informative about the process of entry or exit of the Ca^{2+} -ion, which may happen by three mechanisms. The displacement of the sidechain of Asp1 opens up the binding pocket, so Ca^{2+} can enter or leave, and a similar process involving the sidechain of Asp5 is also possible. An entry path between O(Dab2) and O(Gly6) can also be considered, as in this case a pre-formed octahedral binding formation would attract the ion and the gap between the oxygen atoms is large enough to accommodate a water molecule; an exit via this mechanism is unlikely, though.

Secondary interactions

Apart from interactions involved in coordination of Ca^{2+} , only a few stabilising hydrogen bonds can be seen. In class I molecules, strong hydrogen bonds fix carboxyl groups of Asp1 and Asp5 to amide nitrogens of Gly8 and Val10, respectively, via carboxyl oxygens not participating in Ca^{2+} -coordination. Sub-van der Waals distances likely to correspond to weak hydrogen bonds can be found between $\text{O}^{\delta}(\text{Asp1})\text{--O}^{\delta}(\text{Asp7})$ and $\text{O}^{\delta}(\text{Asp1})\text{--O}(\text{ItD0})$ with about 3.3 Å distance between non-hydrogen atoms, while the one between $\text{N}(\text{Pip3})\text{--O}(\text{Pro11})$ is more likely to be a dipole-dipole interaction, although the binding distance is similar (Figure 5.11). In class II molecules, $\text{O}^{\delta}(\text{Asp5})$ is 2.9 Å away from $\text{N}(\text{Val10})$, but no interaction between $\text{O}^{\delta}(\text{Asp1})$ and $\text{N}(\text{Gly8})$ can be observed. However, $\text{O}^{\delta}(\text{Asp1})$ is now

connected to O^δ(Asp7) via a stronger interaction than in class I molecules, as the binding distance is reduced to 2.8 Å from the previous 3.3 Å. Interactions between O^δ(Asp1)–O(It0) and N(Pip3)–O(Pro11) are also present in class II molecules with bond lengths similar to class I molecules. The fact that the sidechain of Asp1 can be fixed in two conformations supports the entry and exit-mechanisms in that this moiety acts as the gate. The long alkyl chain of It0 is pointing away and does not interact with hydrophobic residues on the peptide core.

Crystal packing

The packing of tsushimycin molecules in the unit cell is not homogeneous. There are large regions that seem to be empty, while in other parts molecules concentrate. It is interesting to note that there seems to be polar and apolar cavities in the crystal, and the molecules act as separators between them. Considering also the shape of the apolar cavity, which resembles a sphere, one might describe the unit cell as a micelle, with hydrophobic sidechains pointing inwards and polar peptide heads being on the surface. This micelle is slightly distorted, as in one direction there are channels between neighbouring hollows and therefore it is better described as a string of pearls. As the unit cell has similar cell dimensions in all three directions, the micelle is nearly isometrical.

Interaction surfaces

Although the packing is tight, only a limited number of specific interactions between the molecules can be seen. There seem to be three well-defined regions in the sequence, with which these molecules interact, although there are considerable differences in their extent. The first and largest region involves three residues and seems to be decisive in determining whether a molecule belongs to conformation class I or class II, while the remaining two surfaces consist of one residue. As there are 12 independent molecules in the unit cell, but intermolecular interactions seem to be highly homologous, a relative notation will be used in the following to distinguish residues, i.e. molecules will be numbered chronologically as they appear in the text.

For class I molecules, O^δ(Asp1/1), O(It0/1) and O^δ(Asp7/1) coordinate a Ca²⁺-ion according to octahedral geometry (Figure 5.12). The octahedron is completed by O^δ(Map4/2) from a second molecule and two additional water molecules. Oxygen atoms from antibiotic molecules are in "equatorial" positions, while the waters are "axial". Bond lengths vary between 2.28-2.39 Å; the peptide atoms were found to occupy a closer position, while the waters are further away. The variation in bond angles is not as extensive as the one found for

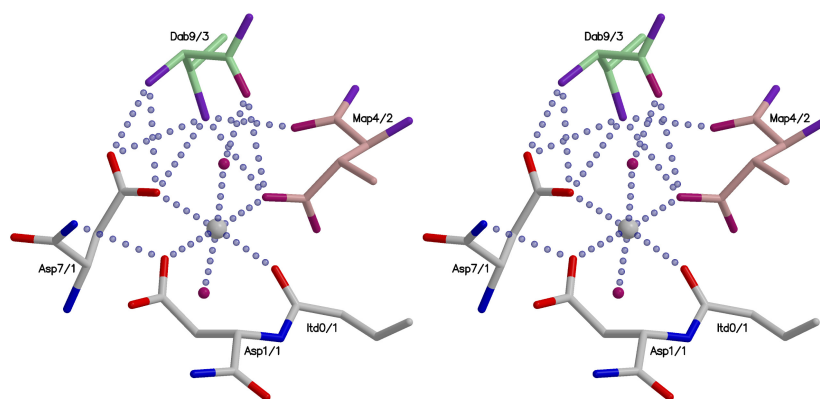


Figure 5.12

Stereo view of the Ca^{2+} -mediated intermolecular association surface in conformation class I. Although the secondary interactions are not too strong, their number may compensate and provide relatively high stability to the interface.

the central Ca^{2+} , and therefore the octahedron is nearly regular. These interactions are supplemented by hydrogen bonds and salt bridges between the O^δ atoms of Asp7/1 and either $\text{N}^\gamma(\text{Dab9/3})$ or $\text{N}(\text{Dab9/3})$ from a third molecule, the $\text{O}^\delta(\text{Map4/2})-\text{N}^\gamma(\text{Dab9/3})$, the $\text{O}(\text{Map4/2})-\text{N}^\gamma(\text{Dab9/3})$ and the $\text{O}^\delta(\text{Map4/2})-\text{O}(\text{Dab9/3})$ hydrogen bonds. One of the axial water molecules is kept in position by a hydrogen bond to $\text{O}(\text{Dab9/3})$. Interactions involving $\text{N}^\gamma(\text{Dab9})$ proved helpful in assigning the chemical element and thereby the chirality of $\text{C}^\beta(\text{Dab9})$, as $\text{C}^\gamma(\text{Dab9})$ is not likely to be involved in such interactions. The exact strength of these interactions vary slightly among independent molecules, but it is consistent that $\text{O}(\text{Map4/2})-\text{N}^\gamma(\text{Dab9/3})$ and one of the two $\text{O}^\delta(\text{Asp7/1})-\text{N}(\text{Dab9/3})$ are the stronger ones with distances around 2.9 Å. The remaining interactions are between 3.2–3.3 Å in bond length, but as they have a salt bridge character, where the distance-dependence does not show such a sharp cutoff, there may still be relatively strong connections between the molecules.

The interface in class II molecules looks similar, apart from the fact that the sidechain of Asp1 now occupies an "axial" position in the octahedron, while one of the waters is displaced to become "equatorial" (Figure 5.13). However, there seems to be no bound Ca^{2+} in the centre of the octahedron, only a water molecule. This can be concluded from the bond distances, which are in the order of 2.5–2.8 Å in this case. Moreover, assigning the chemical element as Ca^{2+} , the occupancy refines to 40%, which nearly exactly corresponds to a fully occupied water molecule. Remaining connections involved in the interface are analogous to those in class I molecules. The "equatorial" water molecule is also hydrogen-bonded to $\text{N}(\text{Gly8/1})$,

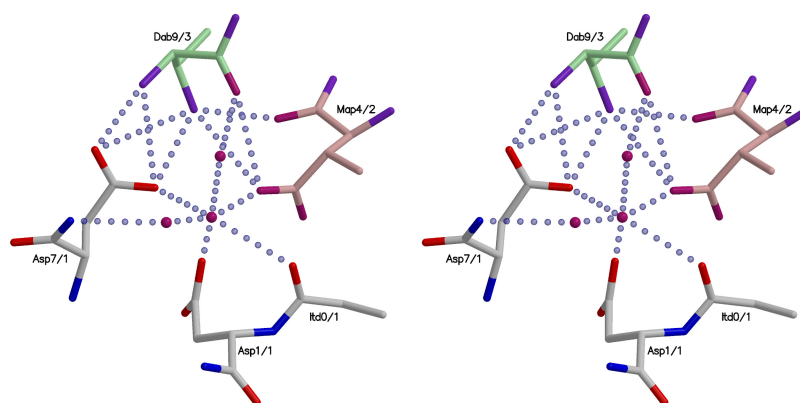


Figure 5.13

Intermolecular connection in molecules from conformation class II. The octahedrally coordinated atom is likely to be a water molecule based on bond distances and temperature factors. This interface is therefore less stable than the one found in class I molecules, but may still be strong enough to keep the molecules in a dimer.

which binds $O^\delta(\text{Asp1})$ in class I molecules. Since distances from the central water molecule are longer, the octahedron is larger and antibiotic molecules are marginally further apart.

As molecules associate in a highly homologous way, Map4 and Dab9 from molecule 1 are also involved in intermolecular interactions. These amino acids take part in the same type of interface, and play analogous roles to what has been described above for Map4/2 and Dab9/3. Therefore, each interface connects three molecules, and each molecule participates in three such interfaces.

Association

Careful examination of molecular packing reveals a higher degree of order among the molecules. Although each molecule participates in three interfaces, molecule 1 and molecule 3 are connected twice by acting symmetrically in two interfaces, i.e. in one interface, molecule 1 binds the Ca^{2+} and molecule 3 participates via $N^\gamma(\text{Dab9})$, while in the other one molecule 3 coordinates the Ca^{2+} and molecule 1 takes part with its $N^\gamma(\text{Dab9})$. This results in a dimer in which the monomers are related by a non-crystallographic twofold axis (Figure 5.14). These dimers are then connected to other dimers via sidechains of Map4 residues present in both chains. A dimer therefore is connected to four other dimers, as it donates two and accepts two Map4 carboxyl groups, leading to heavily interconnected crystal packing. The non-crystallographic twofold axis is oriented in a way that both Itd0 residues point to the same direction, therefore separates the polar and the apolar surface of the dimer. The long alkyl

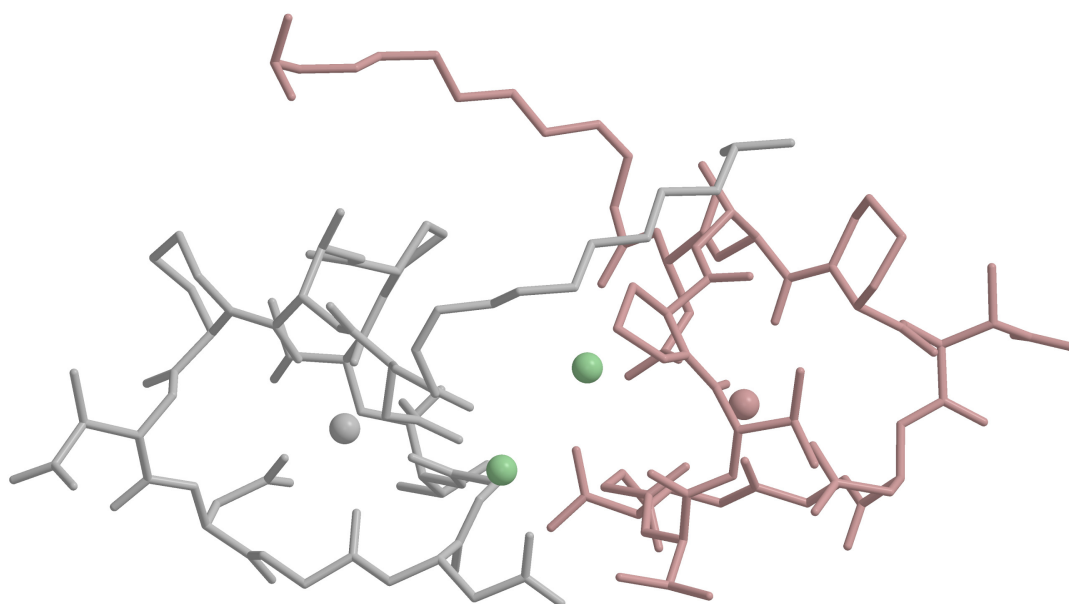


Figure 5.14

Side view of a tsushimycin dimer. It is noteworthy that all aspartate sidechains point to the bottom and constitute a hydrophilic surface there, while the opposite side of the molecule is predominantly hydrophobic. The green Ca^{2+} -ions are involved in dimer interaction and occupy a slightly off-centre position. The opening of the intermolecular tunnel is therefore wide enough to allow the binding of a long polymer chain. The residues on the two side extremes of the dimer are the β -methyl-aspartates (Map4) that also take part in the intermolecular interface and link dimers like a joint.

chains can then interact with each other or with hydrophobic surfaces on the other molecule thereby stabilising the structure.

Although missing the Ca^{2+} -ion in their interface, class II molecules can also participate in dimer formation, since mixed class I–class II dimers were found to have a structure nearly identical to pure class I–class I dimers. No class II–class II dimers were found, though. This may arise from the reduced strength of the class II interface or may simply be accounted for by statistics as there are nine class I and three class II molecules in the cell and the probability of pairing two class II molecules with each other is low.

Binding clefts

Dimerisation results in the formation of a tunnel between monomers. The two sides of the tunnel is composed of hydrophobic surfaces of the two molecules, the bottom is made up of the binding interfaces, while the top can be opened or closed by the long alkyl sidechains (Figure 5.15). The tunnel is about 9 Å long and 5–8 Å wide, non-polar in the middle and

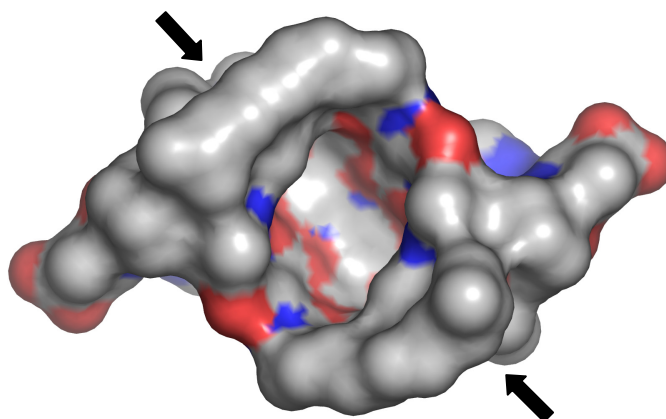


Figure 5.15

Top view of a tsushimycin dimer showing the intermolecular tunnel and an opening at its centre, which would provide space for sidechains connected to the lipoteichoic acid polymer bound in the binding cleft. The two sides of the tunnel are marked with arrows.

charged at the two ends because of incorporated Ca^{2+} -ions. The formation of the cleft requires therefore the presence of Ca^{2+} -ions, and can be anchored to the cell membrane with the long alkyl sidechains.

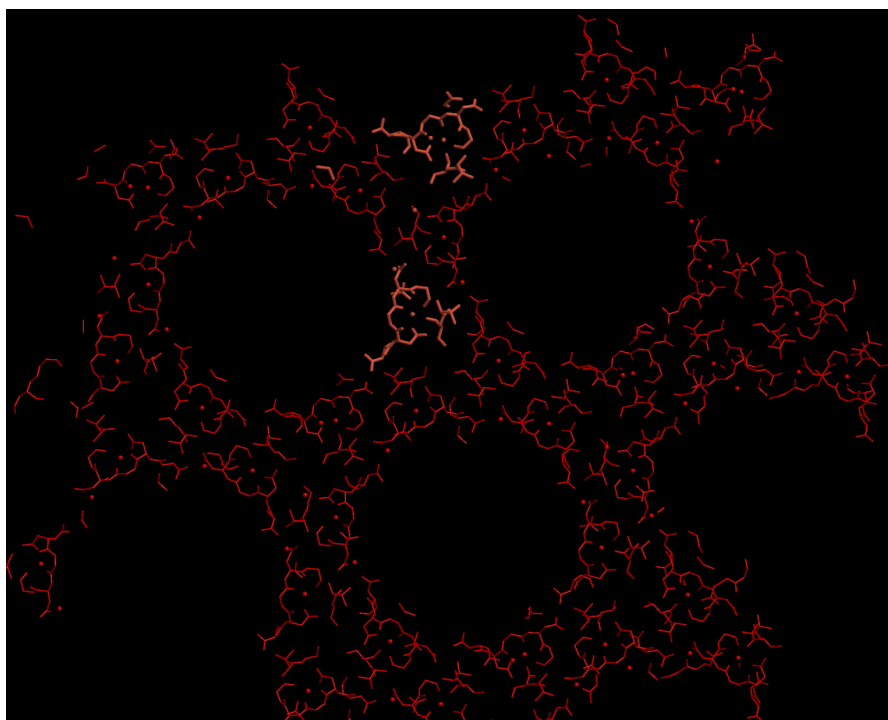


Figure 5.16

Packing in tsushimycin crystals. The molecules constitute spheres, which are hydrophobic inside and hydrophilic outside and may be regarded as small micelles.

Micelles

The dimers in the unit cell were found to be oriented in a way that their long fatty acid sidechains point into a large cavity. Apart from visible parts of these alkyl chains, no electron density can be found in this hollow, which therefore can be regarded as empty. Charged groups of molecules are all pointing outside and constitute a polar surface. This arrangement resembles the structure of a putative micelle, which is elongated in one dimension because of slightly different unit cell dimensions (Figure 5.16). Each unit cell consists of a micelle, which is connected to other ones through structural waters that are well defined in the density; there seems to be about three layers of waters between neighbouring conglomerates, but a single Ca^{2+} -ion was also identified as a link between two adjoining micelle surfaces.

BIOLOGICAL IMPLICATIONS

Lipopeptide antibiotics constitute a highly promising group of drugs against resistant bacteria, but little is known about their exact biological action and even the chemical structure has not been unambiguously established. High resolution X-ray crystal structure determination of the prominent member tsushimycin confirmed the chemical structure proposed by Vértessy *et al.* (2000), and may also give hints in clarifying the mechanism of antimicrobial potency.

Ca^{2+} -ions and biological action

Structure stabilisation

The conformational uniformity of the antibiotic is strongly dependent on the presence of Ca^{2+} -ions. It was found that apart from those involving the central Ca^{2+} -ion, there are very few secondary interactions that stabilise the conformation observed in the crystal structure. The central Ca^{2+} -ion plays therefore a decisive role in determining the conformation of the molecule. Addition of CaCl_2 to the antibiotic induces a significant decrease in solubility of the antibiotic; since most of the carboxyl groups remain on the antibiotic surface even in Ca^{2+} -bound form, the reduction in solubility may be a consequence of a remarkable drop in conformational diversity. It is also noteworthy that in spite of the presence of several aspartate residues, which are conserved even in distant analogues, the Ca^{2+} is coordinated primarily by mainchain carbonyl oxygens, although two aspartate residues also take part in the binding.

It is therefore plausible to assume that the antibiotic has no well defined conformation in the absence of Ca^{2+} , and consequently cannot exert its inhibitory effect. Upon addition of Ca^{2+} the active structure is stabilised by supplementary interactions and the conformational diversity exhibited without Ca^{2+} is reduced and becomes remarkably uniform.

Dimerisation and toxicity

The dimers found in the crystal structure are also strongly dependent on the presence of Ca^{2+} -ions, as each intermolecular interface involves the presence of such an ion. Dimerisation also occurs if one of the Ca^{2+} -ions is missing, as demonstrated by the existence of mixed class I–class II dimers, which miss one Ca^{2+} -ion; however, it is not clear, whether class II–class II dimers not involving Ca^{2+} ions in dimer interactions can also exist.

It is instructive to compare tsushimycin with friulimicin, which differ only in one position in the sequence: instead of an aspartate, friulimicin possesses asparagine as the first residue (Figure 5.1). This small difference results in a significant decrease in toxicity without affecting antibacterial potency. This aspartate residue coordinates two Ca^{2+} -ions in tsushimycin class I molecules; however, in friulimicin it can coordinate only one, as one of its oxygen atoms is replaced by nitrogen. It may be inferred that the binding of Ca^{2+} in the connecting interface is therefore partially responsible for toxicity. However, as there is no structure available for friulimicin, it cannot be decided whether dimerisation of antibiotic molecules is also affected by the replacement of Asp1 with Asn1.

Physiological role of association

Dimerisation

As all the molecules present in the crystal structure of tsushimycin were found to compose dimers, it can be assumed that this unit may be relevant to biological action. The dimers are best described as amphipatic molecules, since one of their sides is predominantly hydrophilic because of the presence of charged carboxyl groups, while the opposite side exposes mostly hydrophobic surfaces. This property may be important in interacting with the bacterial cell membrane. In addition, the two monomers enclose free space between them in form of a tunnel that is also open towards the apolar side. Alternatively, this arrangement allows the antibiotic dimer to join its substrate and anchor it to the cell membrane with its long fatty acid sidechains.

The binding cleft is constituted by mostly hydrophobic amino acids, although the Ca^{2+} -ions in intermolecular connection points are located in a highly polar and charged

environment. The distance between the Ca^{2+} -ions stabilising the dimer is about 9.2 Å, which corresponds well to the length of the repeating unit in lipoteichoic acids, which follows a similar frequency alteration in hydrophobicity: the highly charged phosphate groups are linked via a glycerol that is acylated on the middle OH-group by a D-Ala residue (Figure 5.3). The distance between the oxygen atoms of neighbouring phosphate groups is about 8.4 Å if the chain is in fully extended conformation. The opening on the tsushimycin dimer is exactly where the D-Ala residue would be if the phosphate groups were located at the Ca^{2+} -ions. Binding of tsushimycin to lipoteichoic acids would therefore involve the polymer glycerol phosphate chain to run between tsushimycin monomers, with phosphate groups anchored to Ca^{2+} -ions and D-Ala residues pointing out through the hole on the antibiotic dimer (Figure 5.15). Although both tsushimycin and lipoteichoic acids are polyanionic species, Ca^{2+} -ions present in the interface may provide sufficient positive charge for both of them to bind. Lipoteichoic acids are known to interact with divalent cations, and therefore may well correspond to the site of action of lipopeptide antibiotics.

Micelles

In aqueous solutions lipopeptides are shown to associate to micelles in a Ca^{2+} -dependent manner. In the absence of Ca^{2+} , particles with hydrodynamic radius around 66 nm were found that correspond to micelle sizes up to several million Da. On the other hand, the Ca^{2+} -salt exhibits in solution a narrow distribution centred at 3.15 nm and is equivalent to a micelle with 55 kDa size constituted by 30-40 antibiotic molecules. Micelles found in the tsushimycin crystal are significantly smaller than this as they consist of only 12 molecules, but one has to take into consideration that they were grown in highly alcoholic solution that may affect micelle sizes. This structure can therefore also be regarded as a representative of a functional micelle.

Implications for antibacterial action

The exact mechanism of action of lipopeptide antibiotics has not been unambiguously identified yet, although several possibilities were proposed. Crystal structure investigations were initiated in order to clarify the underlying molecular basis of antimicrobial potency. Although no conclusive decisions can be drawn from the crystal structure, a number of important observations can be made.

Based on the crystal structure, tsushimycin is likely to act as a dimer. Since dimers are associated to expose highly polar surfaces to the solvent, it is unlikely that they penetrate the

bacterial cell membrane albeit the formation of inverse micelles cannot be excluded. Thus, the periplasmic space can be proposed as their primary site of action, which limits possible mechanisms to:

- (a) Inhibition of phospho-N-acetylmuramoyl-pentapeptide-transferase (MraY), which is a transmembrane protein and may allosterically be regulated from the extracellular side of the membrane.
- (b) Interaction with lipoteichoic acids, which are on the cell surface and are therefore accessible to the antibiotic.
- (c) Dissipation of the membrane potential, but probably not with an ion carrier mechanism.

The toxic effect of these antibiotics was predominantly responsible for the cancellation of clinical trials with daptomycin. Although dosing regimens provide a solution in minimising side effects, reduction of toxicity is an important issue in antibiotic development (Debono *et al.*, 1988). The structure of friulimicins, which are considerably less toxic, would therefore be also of interest in the design of novel lipopeptide antibiotics. The recent availability of the crystal structure of tsushimycin may provide a quick way of overcoming the phase problem and determine the structure of friulimicin as well.

VI. SUMMARY

Peptide antibiotics belonging to three different families were investigated using X-ray crystallography. The peptaibol antibiotics cephaibol A, B, C, D and E differ only marginally in chemical structure, although this is associated with a variation more than an order of magnitude in the antibacterial potency. No diffraction quality crystals were obtained from cephaibol D and cephaibol E and therefore only unit cell dimensions were established. Cephaibol A, B and C gave well-diffracting crystals and their structures were determined at ca. 0.9 Å resolution. They adopt a helical conformation with a sharp bend at the central hydroxyproline. Superposition of the models shows that the N-terminal helix is rigid and the C-terminus is flexible, but the differences in three-dimensional structure are negligible. However, there are differences in the hydrogen-bonding pattern for the three structures, and only in the case of cephaibol C does the packing emulate the formation of a membrane channel believed to be important for their biological function.

Feglymycin is a unique peptide in terms of chemical structure and antimicrobial action. The chirality of amino acids alternates regularly, except for at the termini, where there are deviations from this pattern. Two crystal modifications were obtained: one form grew from aqueous solvent, while the other one from a highly alcoholic environment. In both crystal forms the peptide was found to form asymmetric dimers in form of an antiparallel double-stranded double $\beta^{9.0}$ -helix and resemble the membrane channel peptide gramicidin. However, the feglymycin channel is blocked by two phenylalanine sidechains and the dimer is not as long as gramicidin and it is probably too short to span a biological membrane. The dimeric structure is stabilised by numerous hydrogen bonds and therefore it is considerably rigid. Based on the model, it is more likely that feglymycin is a wide-spectrum enzyme inhibitor than a membrane agent.

Friulimicin and tsushimycin belong to the group of lipopeptide antibiotics, and are highly active against multiresistant bacteria. Two crystal forms were obtained from friulimicin, but in one case the crystal diffracted to moderate resolution only, while in the other case it proved to be non-merohedrally twinned and therefore in neither case could the phase problem be solved. Tsushimycin gave one crystal modification and was solved using *ab initio* direct methods. The structure contains 12 molecules; each of these binds a Ca^{2+} -ion that seems to be important in stabilising the three-dimensional structure. The molecules

associate to dimers via interactions mediated by additional Ca^{2+} -ions. The crystal structure confirms the sequence assignment proposed by Vértessy *et al.* (2000). The packing emulates the structure of a micelle and the molecules act as separators between polar and non-polar cavities. The mechanism of antimicrobial action cannot be unambiguously identified from the crystal structure, but it seems likely that the antibiotic cannot enter the bacterial cell and acts on its surface.

VII. REFERENCES

- Abbanat, D., Macielag, M. & Bush, K. (2003). Novel antibacterial agents for the treatment of serious Gram-positive infections. *Expert Opin. Investig. Drugs* **12**, 379-399.
- Alborn, W. E. Jr., Allen, N. E. & Preston, D. A. (1991). Daptomycin disrupts membrane potential in growing *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **35**, 2282-2287.
- Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M. & Berger, E. A. (1996). CC CKR5: a RANTES, MIP-1 α , MIP-1 β receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* **272**, 1955-1958.
- Allen, F. H. (2002). The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Cryst.* **B58**, 380-388.
- Allen, F. H., Kennard, O., Watson, D. G., Brammer, L., Orpen, A. G. & Taylor, R. (1992). Typical interatomic distances: organic compounds. *International Tables for Crystallography*, Vol. C, edited by A. J. C. Wilson, pp. 685-706. Kluwer Academic Publishers, Dordrecht.
- Allen, N. E., Hobbs, J. N. Jr., Alborn, W. E. Jr. (1987). Inhibition of peptidoglycan biosynthesis in Gram-positive bacteria by LY146032. *Antimicrob. Agents Chemother.* **31**, 1093-1099.
- Amidi, S., Solter, S., Rashidian, B., Zokajan, A.-R. & Razmjolan, F. (1975). Antibiotic use and abuse among physicians in private practice in Shiraz, Iran. *Med. Care* **13**, 341-345.
- Anders, R., Ohlenschlager, O., Soskic, V., Wenschuh, H., Heise, B. & Brown, L. R. (2000). The NMR solution structure of the ion channel peptaibol chrysospermin C bound to dodecylphosphocholine micelles. *Eur. J. Biochem.* **267**, 1784-1794.
- Andersson, D. I. (2003). Persistence of antibiotic resistant bacteria. *Curr. Opin. Microbiol.* **6**, 452-456.
- Asano, T., Matsuoka, K., Hida, T., Kobayashi, M., Kitamura, Y., Hayakawa, T., Iinuma, S., Kakinuma, A. & Kato, K. (1994). Novel retrovirus protease inhibitors, RPI-856 A, B, C and D, produced by *Streptomyces* sp. AL-322. *J. Antibiot.* **47**, 557-565.
- Balaram, P., Krishna, K., Sukumar, M., Mellor, I. R. & Sansom, M. S. (1992). The properties of ion channels formed by zervamicins. *Eur. Biophys. J.* **21**, 117-128.
- Balashova, T. A., Shenkarev, Z. O., Tagaev, A. A., Ovchinnikova, T. V., Raap, J. & Arseniev, A. S. (2000). NMR structure of the channel-former zervamicin IIb in isotropic solvents. *FEBS Lett.* **466**, 333-336.

- Balint, G. A.** (2001). Antiretroviral therapeutic possibilities for human immunodeficiency virus/acquired immunodeficiency syndrome. *Pharmacol. Ther.* **89**, 17-27.
- Banerjee, D. K.** (1989). Amphomycin inhibits mannosylphosphoryldolichol synthesis by forming a complex with dolichylmonophosphate. *J. Biol. Chem.* **264**, 2024-2028.
- Baranowski, E., Ruiz-Jarabo, C. M. & Domingo, E.** (2001). Evolution of cell recognition by viruses. *Science* **292**, 1102-1105.
- Barna, J. C. J. & Williams, D. H.** (1984). The structure and mode of action of glycopeptide antibiotics of the vancomycin group. *Annu. Rev. Microbiol.* **38**, 339-357.
- Barry, A. L., Fuchs, P. C. & Brown, S. D.** (2001). *In vitro* activities of daptomycin against 2,789 clinical isolates from 11 North American medical centers. *Antimicrob. Agents Chemother.* **45**, 1919-1922.
- Beach, J. W.** (1998). Chemotherapeutic agents for human immunodeficiency virus infection: mechanism of action, pharmacokinetics, metabolism, and adverse reactions. *Clin. Ther.* **20**, 2-25.
- Bengmark, S.** (2002). Gut microbial ecology in critical illness: is there a role for prebiotics, probiotics, and synbiotics? *Curr. Opin. Crit. Care* **8**, 145-151.
- Berger, E. A.** (1997). HIV entry and tropism: the chemokine receptor connection. *AIDS* **11**, A3-A16.
- Bodanszky, M., Chaturvedi, N. C., Scozzie, J. A., Griffith, R. K. & Bodanszky, A.** (1969). Constituents of amphomycin. *Antimicrob. Agents Chemother.* **9**, 135-138.
- Bodanszky, M., Sigler, G. F. & Bodanszky, A.** (1973). Structure of the peptide antibiotic amphomycin. *J. Am. Chem. Soc.* **95**, 2352-2357.
- Boyd, M. R., Hallock, Y. F., Cardellina, J. H. 2nd., Manfredi, K. P., Blunt, J. W., McMahon, J. B., Buckheit, R. W. Jr., Bringmann, G., Schaffer, M., Cragg, G. M., Thomas, D. W. & Jato, J. G.** (1994). Anti-HIV michellamines from *Ancistrocladus korupensis*. *J. Med. Chem.* **37**, 1740-1745.
- Bronson, J. J. & Barrett, J. F.** (2001). Quinolone, everninomycin, glycylcycline, carbapenem, lipopeptide and cephem antibacterials in clinical development. *Curr. Med. Chem.* **8**, 1775-1793.
- Bruker Nonius** (2002). Cell_now, Proteum, RLATT, SAINT, SADABS, SMART, TWINABS, XPREP and XP computer programs.
- Brünger, A. T.** (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.
- Bugg, T. D. H. & Walsh, C. T.** (1992). Intracellular steps of bacterial cell wall peptidoglycan biosynthesis: antibiotics, and antibiotic resistance. *Nat. Prod. Rep.* **9**, 199-215.

- Burkhart, B. M., Gassman, R. M., Langs, D. A., Pangborn, W. A. & Duax, W. L.** (1998). Heterodimer formation and crystal nucleation of gramicidin D. *Biophys. J.* **75**, 2135-2146.
- Burkhart, B. M., Li, N., Langs, D. A., Pangborn, W. A. & Duax, W. L.** (1998). The conducting form of gramicidin A is a right-handed double-stranded double helix. *Proc. Natl. Acad. Sci. USA* **95**, 12950-12955.
- Burkhart, B. M., Gassman, R. M., Langs, D. A., Pangborn, W. A., Duax, W. L. & Pletnev, V.** (1999). Gramicidin D conformation, dynamics and membrane ion transport. *Biopolymers* **51**, 129-144.
- Canepari, P., Boaretti, M., Del Mar LLeó, M. & Satta, G.** (1990). Lipoteichoic acid as a new target for activity of antibiotics: mode of action of daptomycin (LY146032). *Antimicrob. Agents Chemother.* **34**, 1220-1226.
- Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S.** (1997). Core structure of *gp41* from the HIV envelope glycoprotein. *Cell* **89**, 263-273.
- Chatterjee, A. N. & Perkins, H. R.** (1966). Compounds formed between nucleotides related to the biosynthesis of bacterial cell wall and vancomycin. *Biochem. Biophys. Res. Commun.* **24**, 489-494.
- Chugh, J. K., Brückner, H. & Wallace, B. A.** (2002). Model for a helical bundle channel based on the high-resolution crystal structure of trichotoxin_A50E. *Biochemistry* **41**, 12934-12941.
- Coates, M. E.** (1975). The influence of the gut microflora on the nutrition of its host. *Bibl. Nutr. Dieta* **22**, 101-108.
- Cochran, W.** (1955). Relations between the phases of structure factors. *Acta Cryst.* **8**, 473-478.
- Cocito, C., Di Giambattista, M., Nyssen, E. & Vannuffel, P.** (1997). Inhibition of protein synthesis by streptogramins and related antibiotics. *J. Antimicrob. Chemother.* **39**, A7-A13.
- Colonna-Cesari, F., Premilat, S., Heitz, F., Spach, G. & Lotz, B.** (1977). Helical structures of poly(D-L-peptides). A conformational energy analysis. *Macromolecules* **10**, 1284-1288.
- Cruickshank, D. W. J.** (1970). Least-squares refinement of atomic parameters. *Crystallographic computing*, edited by F. R. Ahmed, S. R. Hall & C. P. Huber, pp. 187-196. Munksgaard, Copenhagen.
- Cruickshank, D. W. J.** (1999). Remarks about protein structure precision. *Acta Cryst.* **D55**, 583-601.
- Cullen, B. R.** (1995). Regulation of HIV gene expression. *AIDS* **9**, A19-A32.

- Cundliffe, E. (1984). Self defence in antibiotic-producing organisms. *Br. Med. Bull.* **40**, 61-67.
- Davies, W. L., Grunert, R. R., Haff, R. F., McGahen, J. W., Neumayer, E. M., Paulshock, M., Watts, J. C., Wood, T. R., Hermann, E. C. & Hoffmann, C. E. (1964). Antiviral activity of 1-adamantanamine (amantadine). *Science* **144**, 862-863.
- Debono, M., Abbot, B. J., Molloy, R. M., Fukuda, D. S., Hunt, A. H., Daupert, V. M., Counter, F. T., Ott, J. L., Carrell, C. B., Howard, L. C., Boeck, L. D. & Hamill, R. L. (1988). Enzymatic and chemical modifications of lipopeptide antibiotic A21978C: the synthesis and evaluation of daptomycin (LY146032). *J. Antibiot.* **41**, 1093-1105.
- Degenkolb, T., Berg, A., Gams, W., Schlegel, B. & Gräfe, U. (2003). The occurrence of peptaibols and structurally related peptaibiotics in fungi and their mass spectrometric identification via diagnostic fragment ions. *J. Peptide Sci.* **9**, 666-678.
- DeLano, W. L. (2003). The PyMOL molecular graphics system. *DeLano Scientific LLC*, San Carlos, CA, USA.
- Dersch, P. (2003). Molecular and cellular mechanisms of bacterial entry into host cells. *Contrib. Microbiol.* **10**, 183-209.
- Diamond, R. (1969). Profile analysis in single crystal diffractometry. *Acta Cryst.* **A25**, 43-55.
- Di Giambattista, M., Chinali, G. & Cocito, C. (1989). The molecular basis of the inhibitory activities of type A and type B synergimycins and related antibiotics on ribosomes. *J. Antimicrob. Chemother.* **24**, 485-507.
- Di Guilmi, A. M., Dessen, A., Dideberg, O. & Vernet, T. (2002). Bifunctional penicillin-binding proteins: focus on the glycosyltransferase domain and its specific inhibitor moenomycin. *Curr. Pharm. Biotechnol.* **3**, 63-75.
- Dragic, T., Litwin, V., Allaway, G. P., Martin, S. R., Huang, Y., Nagashima, K. A., Cayan, C., Maddon, P. J., Koup, R. A., Moore, J. P. & Paxton, W. A. (1996). HIV-1 entry into CD4⁺ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**, 667-673.
- Duax, W. L., Griffin, J. F., Langs, D. A., Smith, G. D., Grochulski, P., Pletnev, V. & Ivanov, V. (1996). Molecular structure and mechanisms of action of cyclic and linear ion transport antibiotics. *Biopolymers* **40**, 141-155.
- Duclohier, H., Snook, C. F. & Wallace, B. A. (1998). Antiamoebin can function as a carrier or as a pore-forming peptaibol. *Biochim. Biophys. Acta* **1415**, 255-260.
- Duisenberg, A. J. M. (1992). Indexing in single-crystal diffractometry with an obstinate list of reflections. *J. Appl. Cryst.* **25**, 92-96.
- Duisenberg, A. J. M., Kroon-Batenburg, L. M. J. & Schreurs, A. M. M. (2003). An intensity evaluation method: EVAL-14. *J. Appl. Cryst.* **36**, 220-229.

- Durckheimer, W.** (1975). Tetracyclines: chemistry, biochemistry, and structure-activity relations. *Angew. Chem. Int. Ed. Engl.* **14**, 721-734.
- El Mashak, E. M. & Tocanne, J. F.** (1980). Polymyxin B-phosphatidylglycerol interactions. A monolayer (π , ΔV) study. *Biochim. Biophys. Acta* **596**, 165-179.
- Emmerson, A. M. & Jones, A. M.** (2003). The quinolones: decades of development and use. *J. Antimicrob. Chemother.* **51**, S13-S20.
- Engh, R. A. & Huber, R.** (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.* **A47**, 392-400.
- Esnouf, R. M.** (1999). Further additions to Molscript version 1.4, including reading and contouring of electron-density maps. *Acta Cryst.* **D55**, 938-940.
- Fox, R. O. & Richards, F. M.** (1982). A voltage-gated ion channel model inferred from the crystal structure of alamethicin at 1.5-Å resolution. *Nature* **300**, 325-330.
- Frank, I.** (2002). Antivirals against HIV-1. *Clin. Lab. Med.* **22**, 741-757.
- Frankel, A. D. & Young, J. A.** (1998). HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **67**, 1-25.
- Frère, J. M. & Joris, B.** (1985). Penicillin-sensitive enzymes in peptidoglycan biosynthesis. *Crit. Rev. Microbiol.* **11**, 299-396.
- Fujinaga, M. & Read, R. J.** (1987). Experiences with a new translation-function program. *J. Appl. Cryst.* **20**, 517-521.
- Fujino, M.** (1965). On glumamycin, a new antibiotic. VI. An approach to the amino acid sequence. *Bull. Chem. Soc. Jpn.* **38**, 517-522.
- Furesz, S.** (1970). Chemical and biological properties of rifampicin. *Antibiot. Chemother.* **16**, 316-351.
- Galbraith, T. P., Harris, R., Driscoll, P. C. & Wallace, B. A.** (2003). Solution NMR studies of antiameobin, a membrane channel-forming polypeptide. *Biophys. J.* **84**, 185-194.
- Gale, E. F., Cundliffe, E., Reynolds, P. E., Richmond, M. H. & Waring, M. J.** (1981). *The molecular basis of antibiotic action*, 2. ed. Wiley-Interscience, New York.
- Garrison, M. W., Rotschafer, J. C. & Crossley, K. B.** (1989). Suboptimal effect of daptomycin in the treatment of bacteremias. *South Med. J.* **82**, 1414-1415.
- Geiger, T. & Clarke, S.** (1987). Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. *J. Biol. Chem.* **262**, 785-794.
- Gorbach, S. L.** (1997). Treating diarrhoea. *BMJ* **314**, 1776-1777.
- Gordon, L. G. & Haydon, D. A.** (1972). The unit conductance channel of alamethicin. *Biochim. Biophys. Acta* **255**, 1014-1018.

- Gracey, M. S.** (1981). Nutrition, bacteria and the gut. *Br. Med. Bull.* **37**, 71-75.
- Gulick, R. M.** (2003). New antiretroviral drugs. *Clin. Microbiol. Infect.* **9**, 186-193.
- Hanke, W. & Boheim, G.** (1980). The lowest conductance state of the alamethicin pore. *Biochim. Biophys. Acta* **596**, 456-462.
- Hauck, C. R.** (2002). Cell adhesion receptors - signaling capacity and exploitation by bacterial pathogens. *Med. Microbiol. Immunol.* **191**, 55-62.
- Hayes, J. D. & Wolf, C. R.** (1990). Molecular mechanisms of drug resistance. *Biochem. J.* **272**, 281-295.
- Heinemann, B., Kaplan, M. A., Muir, R. D. & Hooper, I. R.** (1953). Amphomycin, a new antibiotic. *Antibiot. Chemother.* **3**, 1239-1242.
- Hermesen, E. D., Sullivan, C. J. & Rotschafer, J. C.** (2003). Polymyxins: pharmacology, pharmacokinetics, pharmacodynamics, and clinical applications. *Infect. Dis. Clin. North Am.* **17**, 545-562.
- Hooper, L. V., Midtvedt, T. & Gordon, J. I.** (2002). How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu. Rev. Nutr.* **22**, 283-307.
- Inoue, M.** (1962a). On glumamaycin, a new antibiotic. II. Isolation and identification of amino acids constituting glumamaycin. *Bull. Chem. Soc. Jpn.* **35**, 1249-1254.
- Inoue, M.** (1962b). On glumamaycin, a new antibiotic. III. Fatty acid, a constituent of the antibiotic. *Bull. Chem. Soc. Jpn.* **35**, 1255-1257.
- Inoue, M.** (1962c). On glumamaycin, a new antibiotic. IV. The amino acid moiety. *Bull. Chem. Soc. Jpn.* **35**, 1556-1559.
- Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D. Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H. & Arnold, E.** (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. USA* **90**, 6320-6324.
- Jancarik, J. & Kim, S. H.** (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* **24**, 409-411.
- Kabsch, W.** (1976). A solution for the best rotation to relate two sets of vectors. *Acta Cryst.* **A32**, 922-923.
- Kabsch, W.** (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795-800.
- Kabsch, W.** (2001). Integration, scaling, space-group assignment and post refinement. *International Tables for Crystallography*, Vol. F, edited by E. Arnold & M. G. Rossmann, pp. 218-225. Kluwer Academic Publishers, Dordrecht.

- Kamiryo, T. & Matsushashi, M.** (1972). The biosynthesis of the cross-linking peptides in the cell wall peptidoglycan of *Staphylococcus aureus*. *J. Biol. Chem.* **247**, 6306-6311.
- Kaneko, I., Kamoshida, K. & Takahashi, S.** (1989). Complestatin, a potent anti-complement substance produced by *Streptomyces lavendulae*. I. Fermentation, isolation and biological characterization. *J. Antibiot.* **42**, 236-241.
- Kang, M. S., Spencer, J. P. & Elbein, A. D.** (1978). Amphomycin inhibition of mannose and GlcNAc incorporation into lipid-linked saccharides. *J. Biol. Chem.* **253**, 8860-8866.
- Karle, I. L., Flippen-Anderson, J. L., Agarwalla, S. & Balaram, P.** (1991). Crystal structure of [Leu1]zervamicin, a membrane ion-channel peptide: Implications for gating mechanisms. *Proc. Natl. Acad. Sci. USA* **88**, 5307-5311.
- Karle, I. L., Perozzo, M. A., Mishra, V. K. & Balaram, P.** (1998). Crystal structure of the channel-forming polypeptide antiamoebin in a membrane-mimetic environment. *Proc. Natl. Acad. Sci. USA* **95**, 5501-5504.
- Karle, J. & Hauptman, H.** (1956). A theory of phase determination for the four types of non-centrosymmetric space groups 1P222, 2P22, 3P₁2, 3P₂2. *Acta Cryst.* **9**, 635-651.
- Katz, R. A. & Skalka, A. M.** (1994). The retroviral enzymes. *Annu. Rev. Biochem.* **63**, 133-173.
- Kayser, F. H.** (2003). Safety aspects of enterococci from the medical point of view. *Int. J. Food Microbiol.* **88**, 255-262.
- Ketchum, R. R., Lee, K. C., Huo, S. & Cross, T. A.** (1996). Macromolecular structural elucidation with solid-state NMR-derived orientational constraints. *J. Biomol. NMR* **8**, 1-14.
- Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B.** (1999). Rapid automated molecular replacement by evolutionary search. *Acta Cryst.* **D55**, 484-491.
- Kitchen, V. S., Skinner, C., Ariyoshi, K., Lane, E. A., Duncan, I. B., Burckhardt, J., Burger, H. U., Bragman, K., Pinching, A. J. & Weber, J. N.** (1995). Safety and activity of saquinavir in HIV infection. *Lancet* **345**, 952-955.
- Klare, I., Konstabel, C., Badstübner, D., Werner, G. & Witte, W.** (2003). Occurrence and spread of antibiotic resistances in *Enterococcus faecium*. *Int. J. Food Microbiol.* **88**, 269-290.
- Klemm, P. & Schembri, M. A.** (2000). Bacterial adhesins: function and structure. *Int. J. Med. Microbiol.* **290**, 27-35.
- Kong, F. & Carter, G. T.** (2003). Structure determination of glycinocins A to D, further evidence for the cyclic structure of the amphomycin antibiotics. *J. Antibiot.* **56**, 557-564.
- Kotra, L. P., Haddad, J. & Mobashery, S.** (2000). Aminoglycosides: perspectives on mechanisms of action and resistance and strategies to counter resistance. *Antimicrob. Agents Chemother.* **44**, 3249-3256.

- Kraulis, P. J.** (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.
- Kronen, M., Görls, H., Nguyen, H.-H., Reissmann, S., Bohl, M., Sühnel, J. & Gräfe, U.** (2003). Crystal structure and conformational analysis of ampullosporin A. *J. Peptide Sci.* **9**, 729-744.
- Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A.** (1998). Structure of an HIV *gp120* envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648-659.
- Lambert, P. A., Hancock, I. C. & Baddiley, J.** (1977). Occurrence and function of membrane teichoic acids. *Biochim. Biophys. Acta* **472**, 1-12.
- Langs, D. A.** (1988). Three-dimensional structure at 0.86 Å of the uncomplexed form of the transmembrane ion channel peptide gramicidin A. *Science* **241**, 188-191.
- Larder, B. A.** (1995). Viral resistance and the selection of antiretroviral combinations. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **10**, S228-S233.
- Latorre, R. & Alvarez, O.** (1981). Voltage-dependent channels in planar lipid bilayer membranes. *Physiol. Rev.* **61**, 77-150.
- Lebek, G. & Cottier, H.** (1992). Notes on the bacterial content of the gut. *Curr. Stud. Hematol. Blood Transfus.* **59**, 1-18.
- Leclerc, G., Rebuffat, S., Goulard, C. & Bodo, B.** (1998). Directed biosynthesis of peptaibol antibiotics in two *Trichoderma* strains. I. Fermentation and isolation. *J. Antibiot.* **51**, 170-183.
- Lehmann, M. S. & Larsen, F. K.** (1979). A method for location of the peaks in step-scan measured Bragg reflexions. *Acta Cryst.* **A30**, 580-584.
- Leslie, A. G. W.** (1999). Integration of macromolecular diffraction data. *Acta Cryst.* **D55**, 1696-1702.
- Levy, S. B.** (1992). *The antibiotic paradox: how miracle drugs are destroying the miracle*. Plenum Publishing, New York.
- Levy, S. B.** (1998). The challenge of antibiotic resistance. *Scientific American* **278**, 46-53.
- Levy, S. B.** (2001). Antibiotic resistance: consequences of inaction. *Clin. Infect. Dis.* **33**, S124-S129.
- Levy, S. B.** (2002). Factors impacting on the problem of antibiotic resistance. *J. Antibiot. Chemother.* **49**, 25-30.
- Linden, P. K.** (2002). Treatment options for vancomycin-resistant enterococcal infections. *Drugs* **62**, 425-441.

- Little, S. J., Daar, E. S., D'Aquila, R. T., Keiser, P. H., Connick, E., Whitcomb, J. M., Hellmann, N. S., Petropoulos, C. J., Sutton, L., Pitt, J. A., Rosenberg, E. S., Koup, R. A., Walker, B. D. & Richman, D. D. (1999). Reduced antiretroviral susceptibility among patients with primary HIV infection. *JAMA* **282**, 1142-1149.
- Lomakina, N. N. & Brazhnikova, M. G. (1959). The composition of crystallomycin. *Biokhimiia* **24**, 425-431.
- Lomize, A. L., Orekhov, V. Yu. & Arseniev, A. S. (1992). Refinement of the spatial structure of the gramicidin A transmembrane ion-channel. *Bioorg. Khim.* **18**, 182-200.
- Mansky, L. M. & Temin, H. M. (1995). Lower *in vivo* mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* **69**, 5087-5094.
- Marra, A. & Isberg, R. R. (1996). Common entry mechanisms. Bacterial pathogenesis. *Curr. Biol.* **6**, 1084-1086.
- Matthée, G., Wright, A. D. & König, G. M. (1999). HIV reverse transcriptase inhibitors of natural origin. *Planta Med.* **65**, 493-506.
- McMahon, J. B., Currens, M. J., Gulakowski, R. J., Buckheit, R. W. Jr., Lackman-Smith, C., Hallock, Y. F. & Boyd, M. R. (1995). Michellamine B, a novel plant alkaloid, inhibits human immunodeficiency virus-induced cell killing by at least two distinct mechanisms. *Antimicrob. Agents Chemother.* **39**, 484-488.
- McRee, D. E. (1999). XtalView – a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156-165.
- Menestrina, G., Voges, K. P., Jung, G. & Boheim, G. (1986). Voltage-dependent channel formation by rods of helical polypeptides. *J. Membrane Biol.* **93**, 111-132.
- Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.
- Metges C. C. (2000). Contribution of microbial amino acids to amino acid homeostasis of the host. *J. Nutr.* **130**, 1857S-1864S.
- Milatovic, D. & Braveny, I. (1987). Development of resistance during antibiotic therapy. *Eur. J. Clin. Microbiol.* **6**, 234-244.
- Miller, Y. W., Eady, E. A., Lacey, R. W., Cove, J. H., Joanes, D. N. & Cunliffe, W. J. (1996). Sequential antibiotic therapy for acne promotes the carriage of resistant staphylococci on the skin of contacts. *J. Antimicrob. Chemother.* **38**, 829-837.
- Milov, A. D., Tsvetkov, Yu. D., Formaggio, F., Crisma, M., Toniolo, C. & Raap, J. (2003). Self-assembling and membrane modifying properties of a lipopeptaibol studied by CW-ESR and PELDOR spectroscopies. *J. Peptide Sci.* **9**, 690-700.

- Moews, P. C. & Kretsinger, R. H.** (1975). Refinement of the structure of carp muscle calcium-binding parvalbumin by model building and difference Fourier analysis. *J. Mol. Biol.* **91**, 201-225.
- Moulard, M., Lortat-Jacob, H., Mondor, I., Roca, G., Wyatt, R., Sodroski, J., Zhao, L., Olson, W., Kwong, P. D. & Sattentau, Q. J.** (2000). Selective interactions of polyanions with basic surfaces on human immunodeficiency virus type 1 *gp120*. *J. Virol.* **74**, 1948-1960.
- Naganawa, H., Hamada, M., Maeda, K., Okami, Y., Takeuchi, T. & Umezawa, H.** (1968). Laspartomycin, a new antistaphylococcal peptide. *J. Antibiot.* **21**, 55-62.
- Nagy, K., Young, M., Baboonian, C., Merson, J., Whittle, P. & Oroszlan, S.** (1994). Antiviral activity of human immunodeficiency virus type 1 protease inhibitors in a single cycle of infection: evidence for a role of protease in the early phase. *J. Virol.* **68**, 757-765.
- Nelson, M. L. & Levy, S. B.** (1999). Reversal of tetracycline resistance mediated by different bacterial tetracycline resistance determinants by an inhibitor of the Tet(B) antiport protein. *Antimicrob. Agents Chemother.* **43**, 1719-1724.
- Neu, H. C.** (1992). The crisis in antibiotic resistance. *Science* **257**, 1064-1073.
- Neuhaus, F. C. & Hammes, W. P.** (1981). Inhibition of cell wall biosynthesis by analogues and alanine. *Pharmacol. Ther.* **14**, 265-319.
- Ng, T. B., Huang, B., Fong, W. P. & Yeung, H. W.** (1997). Anti-human immunodeficiency virus (anti-HIV) natural products with special emphasis on HIV reverse transcriptase inhibitors. *Life Sci.* **61**, 933-949.
- Noller, H. F.** (1991). Ribosomal RNA and translation. *Annu. Rev. Biochem.* **60**, 191-227.
- Norrby, R.** (2001). Linezolid – a review of the first oxazolidinone. *Expert Opin. Pharmacother.* **2**, 293-302.
- Ochoa, T. J. & Cleary, T. G.** (2003). Epidemiology and spectrum of disease of *Escherichia coli* O157. *Curr. Opin. Infect. Dis.* **16**, 259-263.
- Omobosola, A. & Henry, K.** (2003). Current trends in the treatment of HIV infection, 2003. *Minn. Med.* **86**, 39-44.
- Otwinowski, Z. & Minor, W.** (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
- Pani, A., Loi, A. G., Mura, M., Marceddu, T., La Colla, P. & Marongiu, M. E.** (2002). Targeting HIV: old and new players. *Curr. Drug Targets Infect. Dis.* **2**, 17-32.
- Pannu, N. S. & Read, R. J.** (1996). Improved refinement through maximum likelihood. *Acta Cryst.* **A52**, 659-668.

- Paterson, D. L., Swindells, S., Mohr, J., Brester, M., Vergis, E. M., Squier, C., Wagener, M. M. & Singh, N.** (2000). Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. *Ann. Intern. Med.* **133**, 21-30.
- Peggion, C., Formaggio, F., Crisma, M., Epand, R. F., Epand, R. M. & Toniolo, C.** (2003). Trichogin: a paradigm for lipopeptaibols. *J. Peptide Sci.* **9**, 679-689.
- Pluymers, W., De Clercq, E. & Debyser, Z.** (2001). HIV-1 integration as a target for antiretroviral therapy: a review. *Curr. Drug Targets Infect. Dis.* **1**, 133-149.
- Read, R. J.** (1990). Structure-factor probabilities for related structures. *Acta Cryst.* **A46**, 900-912.
- Rehm, S. J.** (2002). Two new treatment options for infections due to drug-resistant Gram-positive cocci. *Cleve. Clin. J. Med.* **69**, 397-401.
- Reynolds, P. E.** (1989). Structure, biochemistry and mechanism of action of glycopeptide antibiotics. *Eur. J. Clin. Microbiol. Infect. Dis.* **8**, 943-950.
- Ritzau, M., Heinze, S., Dornberger, K., Berg, A., Fleck, W., Schlegel, B., Härtl, A. & Gräfe, U.** (1997). Ampullosporin, a new peptaibol-type antibiotic from *Sepedonium ampullosporum* HKI-0053 with neuroleptic activity in mice. *J. Antibiot.* **50**, 722-728.
- Rizzuto, C. D., Wyatt, R., Hernandez-Ramos, N., Sun, Y., Kwong, P. D., Hendrickson, W. A. & Sodroski, J.** (1998). A conserved HIV *gp120* glycoprotein structure involved in chemokine receptor binding. *Science* **280**, 1949-1953.
- Rossmann, M. G. & Arnold, E.** (1993). Patterson and molecular-replacement techniques. *International Tables for Crystallography*, Vol. B, edited by U. Shmueli, pp. 235-263. Kluwer Academic Publishers, Dordrecht.
- Rossmann, M. G. & van Beek, C. G.** (1999). Data processing. *Acta Cryst.* **D55**, 1631-1640.
- Rotschafer, J. C., Garrison, M. W. & Rodvold, K. A.** (1988). Therapeutic update on glycopeptide and lipopeptide antibiotics. *Pharmacotherapy* **8**, 211-219.
- Russel, A. D.** (2003). Biocide use and antibiotic resistance: the relevance of laboratory findings to clinical and environmental situations. *Lancet Infect. Dis.* **3**, 794-803.
- Sarges, R. & Witkop, B.** (1956). Gramicidin A. V. The structure of valine- and isoleucine-gramicidin A. *J. Am. Chem. Soc.* **87**, 2011-2020.
- Schaberg, D. R., Rubens, C. E., Alford, R. H., Farrar, W. E., Schaffner W. & McGee, Z. A.** (1981). Evolution of antimicrobial resistance and nosocomial infection. Lessons from the Vanderbilt experience. *Am. J. Med.* **70**, 445-448.
- Schiell, M., Hofmann, J., Kurz, M., Schmidt, F. R., Vértessy, L., Vogel, M., Wink, J. & Seibert, G.** (2001). Cephaibols, new peptaibol antibiotics with anthelmintic properties from *Acremonium tubaki* DSM 12774. *J. Antibiot.* **54**, 220-233.

- Schlegel, B., Härtl, A., Berg, A., Kronen, M., Gräfe, U., Vértesy, L., Wink, J. & Brückner, H. (2002). Peptaibols as inducers of fungal pigment formation and correlation to hypothermia in mice. Poster No. 15 at the workshop *Peptaibols: biosynthesis, structural diversity, bioactivity and mode of action*. 9-11 October 2002, Jena, Germany.
- Schleifer, K. H. & Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.* **36**, 407-477.
- Schneider, T. R. (2002). A genetic algorithm for the identification of conformationally invariant regions in protein molecules. *Acta Cryst.* **D58**, 195-208.
- Sebastian, J. & Faruki, H. (2004). Update on HIV resistance and resistance testing. *Med. Res. Rev.* **24**, 115-125.
- Seydel, J. K. (1968). Sulfonamides, structure-activity relationship, and mode of action. Structural problems of the antibacterial action of 4-aminobenzoic acid (PABA) antagonists. *J. Pharm. Sci.* **57**, 1455-1478.
- Shay, A. J., Adam, J., Martin, J. H., Hausmann, W. K., Shu, P. & Bohonos, N. (1960). Aspartocin. I. Production, isolation, and characteristics. *Antibiot. Annu.* **7**, 194-198.
- Sheldrick, G. M. (2002). Macromolecular phasing with SHELXE. *Z. Kristallogr.* **217**, 644-650.
- Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: high resolution refinement. *Methods Enzymol.* **277**, 319-343.
- Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, M. & Usón, I. (2001). *Ab initio* phasing. *International Tables for Crystallography*, Vol. F, edited by E. Arnold & M. G. Rossmann, pp. 333-351. Kluwer Academic Publishers, Dordrecht.
- Shenkarev, Z. O., Balashova, T. A., Efremov, R. G., Yakimenko, Z. A., Ovchinnikova, T. V., Raap, J. & Arseniev, A. S. (2002). Spatial structure of zervamicin IIb bound to DPC micelles: implications for voltage-gating. *Biophys. J.* **82**, 762-771.
- Shibata, M., Kanzaki, T., Nakazawa, K., Inoue, M., Hitomi, H., Mizuno, K., Fujino, M. & Miyake, A. (1962). On glumamycin, a new antibiotic. *J. Antibiot.* **15**, 1-6.
- Shoji, J. & Otsuka, H. (1969). Studies on tsushimycin. II. The structures of constituent fatty acids. *J. Antibiot.* **22**, 473-479.
- Shoji, J., Kozuki, S., Okamoto, S., Sakazaki, R. & Otsuka, H. (1968). Studies on tsushimycin. Isolation and characterization of an acidic acylpeptide containing a new fatty acid. *J. Antibiot.* **21**, 439-443.
- Smith, H. (1995). The revival of interest in mechanisms of bacterial pathogenicity. *Biol. Rev.* **70**, 277-316.
- Snook, C. F., Woolley, G. A., Oliva, G., Pattabhi, V., Wood, S. P., Blundell, T. L. & Wallace, B. A. (1998). The structure and function of antiamoebin I, a proline-rich membrane-active polypeptide. *Structure* **6**, 783-792.

- Sparks, R. A.** (1976). Trends in minicomputer hardware and software. Part I. *Crystallographic computing techniques*, edited by F. R. Ahmed, K. Huml & B. Sedláček, pp.452-467. Munksgaard, Copenhagen.
- Stella, L., Mazzuca, C., Venanzi, M., Palleschi, A., Didone, M., Formaggio, F., Toniolo, C. & Pispisa, B.** (2004). Aggregation and water-membrane partition as major determinants of the activity of the antibiotic peptide trichogin GA IV. *Biophys. J.* **86**, 936-945.
- Steller, I., Bolotovskiy, R. & Rossmann, M. G.** (1997). An algorithm for automatic indexing of oscillation images using Fourier analysis. *J. Appl. Cryst.* **30**, 1036-1040.
- Strahilevitz, J. & Rubinstein, E.** (2002). Novel agents for resistant Gram-positive infections – a review. *Int. J. Infect. Dis.* **6**, S38–S46.
- Stratov, I., DeRose, R., Purcell, D. F. & Kent, S. J.** (2004). Vaccines and vaccine strategies against HIV. *Curr. Drug Targets* **5**, 71-88.
- Tally, F. P., Zeckel, M., Wasilewski, M. M., Carini, C., Berman, C. L., Drusano, G. L. & Oleson, F. B.** (1999). Daptomycin: a novel agent for Gram-positive infections. *Expert Opin. Investig. Drugs* **8**, 1223-1238.
- Tanaka, H., Iwai, Y., Ōiwa, R., Shinohara, S., Shimizu, S., Oka, T. & Ōmura, S.** (1977). Studies on bacterial cell wall inhibitors. II. Inhibition of peptidoglycan synthesis *in vivo* and *in vitro* by amphomycin. *Biochim. Biophys. Acta* **497**, 633-640.
- Tanaka, H., Ōiwa, R., Matsukura, S. & Ōmura, S.** (1979). Amphomycin inhibits phospho-N-acetylmuramyl-pentapeptide translocase in peptidoglycan synthesis of *Bacillus*. *Biochem. Biophys. Res. Commun.* **86**, 902-908.
- Tanaka, H., Ōiwa, R., Matsukura, S., Inokoshi, J. & Ōmura, S.** (1982). Studies on bacterial cell wall inhibitors. X. Properties of phospho-N-acetylmuramoyl-pentapeptide-transferase in peptidoglycan synthesis of *Bacillus megaterium* and its inhibition by amphomycin. *J. Antibiot.* **35**, 1216-1221.
- Tanaka, H., Matsuzaki, K., Nakashima, H., Ogino, T., Matsumoto, A., Ikeda, H., Woodruff, H. B. & Ōmura, S.** (1997). Chloropeptins, new anti-HIV antibiotics inhibiting gp120-CD4 binding from *Streptomyces* sp. I. Taxonomy, fermentation, isolation, and physico-chemical properties and biological activities. *J. Antibiot.* **50**, 58-65.
- Tosteson, D. C., Cook, P., Andreoli, T. & Tieffenberg, M.** (1967). The effect of valinomycin on potassium and sodium permeability of HK and LK sheep red cells. *J. Gen. Physiol.* **50**, 2513-2525.
- Townsley, L. E., Tucker, W. A., Sham, S. & Hinton, J. F.** (2001). Structures of gramicidins A, B, and C incorporated into sodium dodecyl sulfate micelles. *Biochemistry* **40**, 11676-11686.

- Usón, I., Sheldrick, G. M., de La Fortelle, E., Bricogne, G., Di Marco, S., Priestle, J. P., Grütter, M. G. & Mittl, P. R. E. (1999). The 1.2 Å crystal structure of hirustasin reveals the intrinsic flexibility of a family of highly disulphide-bridged inhibitors. *Structure* **7**, 55-63.
- Vértesy, L., Aretz, W., Knauf, M., Markus, A., Vogel, M. & Wink, J. (1999). Feglymycin, a novel inhibitor of the replication of the human immunodeficiency virus. Fermentation, isolation and structure elucidation. *J. Antibiot.* **52**, 374-382.
- Vértesy, L., Ehlers, E., Kogler, H., Kurz, M., Meiwes, J., Seibert, G., Vogel, M. & Hammann, P. (2000). Friulimicins: Novel lipopeptide antibiotics with peptidoglycan synthesis inhibiting activity from *Actinoplanes friuliensis* sp. nov. *J. Antibiot.* **53**, 816-827.
- Vlietnick, A. J., De Bruyne, T., Apers, S. & Pieters, L. A. (1998). Plant-derived leading compounds for chemotherapy of human immunodeficiency virus (HIV) infection. *Planta Med.* **64**, 97-109.
- Vogler, K., Studer, R. O., Lanz, P., Lergier, W. & Böhni, E. (1965). Synthesen in der Polymyxin-Reihe. 9. Synthese von Polymyxin B₁. *Helv. Chim. Acta* **48**, 1161-1177.
- Waxman, D. J. & Strominger, J. L. (1983). Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics. *Annu. Rev. Biochem.* **52**, 825-869.
- Wegener, H. C., Aarestrup, F. M., Jensen, L. B., Hammerum, A. M. & Bager, F. (1999). Use of antimicrobial growth promoters in food animals and *Enterococcus faecium* resistance to therapeutic antimicrobial drugs in Europe. *Emerg. Infect. Dis.* **5**, 329-335.
- Williams, D. H., Stone, M. J., Hauck, P. R. & Rahman, S. K. (1989). Why are secondary metabolites (natural products) biosynthesized? *J. Nat. Prod.* **52**, 1189-1208.
- Witte, W., Tschäpe, H., Klare, I. & Werner, G. (2000). Antibiotics in animal feed. *Acta Vet. Scand. Suppl.* **93**, 37-44.
- Wlodawer, A. (1994). Rational drug design: the proteinase inhibitors. *Pharmacotherapy* **14**, S9-S20.
- Woodford, N. (2003). Novel agents for the treatment of resistant Gram-positive infections. *Expert Opin. Investig. Drugs* **12**, 117-137.
- Wyatt, R. & Sodroski, J. (1998). The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* **280**, 1884-1888.
- Yin, N., Marshall, R. L., Matheson, S. & Savage, P. B. (2003). Synthesis of lipid A derivatives and their interactions with polymyxin B and polymyxin B nonapeptide. *J. Am. Chem. Soc.* **125**, 2426-2435.
- Yunis, A. A. (1988). Chloramphenicol: relation of structure to activity and toxicity. *Annu. Rev. Pharmacol. Toxicol.* **28**, 83-100.

PUBLIKATIONEN

VERÖFFENTLICHUNGEN IN FACHZEITSCHRIFTEN

- [1] Neculai, A.-M., Cummins, C. C., Neculai, D., Roesky, H. W., Bunkóczi, G., Walfort, B. & Stalke, D. (2003). Elucidation of a Sc(I) complex by DFT calculations and reactivity studies. *Inorg. Chem.* **42**, 8803-8810.
- [2] Bunkóczi, G., Schiell, M., Vértesy, L. & Sheldrick, G. M. (2003). Crystal structures of cephaibols. *J. Peptide Sci.* **9**, 745-752.
- [3] Lehmann, C., Debreczeni, J. É., Bunkóczi, G., Dauter, M., Dauter, Z., Vértesy, L. & Sheldrick, G. M. (2003). Structures of four crystal forms of decaplanin. *Helv. Chim. Acta* **86**, 1478-1487.
- [4] Fitjer, L., Gerke, R., Weiser, J., Bunkóczi, G. & Debreczeni, J. É. (2003). Helical primary structures of four-membered rings: (M)-trispiro[3.0.0.3.2.2]tridecane. *Tetrahedron* **59**, 4443-4449.
- [5] Mernyák, E., Wölfling, J., Bunkóczi, G., Luo, L., Schneider, T. R. & Schneider, Gy. (2003). Stereoselective synthesis of the two trans-(16-hydroxymethyl)-3-methoxy-13 α -estra-1,3,5(10)-trien-17-ol isomers. *Coll. Czech. Chem. Commun.* **68**, 1141-1148.
- [6] Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). In-house measurement of the sulfur anomalous signal and its use for phasing. *Acta Cryst.* **D59**, 688-696.
- [7] Debreczeni, J. É., Bunkóczi, G., Girmann, B. & Sheldrick, G. M. (2003). In-house phase determination of the lima bean trypsin inhibitor: a low-resolution sulfur-SAD case. *Acta Cryst.* **D59**, 393-395.
- [8] Neculai, D., Neculai, A.-M., Roesky, H. W., Magull, J. & Bunkóczi, G. (2002). Synthesis and structure of a new fluorinated beta-ketoiminato ligand and its lithium derivative. *J. Fluorine Chem.* **118**, 131-134.
- [9] Wölfling, J., Frank, É., Mernyák, E., Bunkóczi, G., Cuesta Seijo, J. A. & Schneider, Gy. (2002). Synthesis of novel halogen-containing D-homoestrone and 13 α -D-homoestrone derivatives by Lewis acid-induced intramolecular Prins reaction. *Tetrahedron* **58**, 6851-6861.
- [10] Lehmann, C., Bunkóczi, G., Vértesy, L. & Sheldrick, G. M. (2002). Structures of glycopeptide antibiotics with peptides that model bacterial cell-wall precursors. *J. Mol. Biol.* **318**, 723-732.

POSTERPRÄSENTATIONEN

- [1] **Bunkóczi, G., Debreczeni, J. É., Sevvana, M. & Sheldrick, G. M.** (2003). Search for a test crystal. *Strategies in Macromolecular Structure Determination at 3rd Generation Synchrotrons*, June 17-25, Grenoble, France.
- [2] **Bunkóczi, G., Debreczeni, J. É., Sevvana, M., Vučković, M. & Sheldrick, G. M.** (2003). In-house phasing with iodide soaks. *Workshop on Phasing with Soft X-rays*, February 25-27, Brixen, Italy.
- [3] **Bunkóczi, G., Schiell, M., Vértesy, L. & Sheldrick, G. M.** (2002). Crystal Structures of cephaibols. *Peptaibols: Biosynthesis, Structural Diversity, Bioactivity and Mode of Action*, October 9-11, Jena, Germany.
- [4] **Bunkóczi, G., Debreczeni, J. É., Kärcher, J. & Sheldrick, G. M.** (2002). In-house phasing with iodide soaks. *XIX. Congress and General Assembly of the International Union of Crystallography*, August 6-15, Geneva, Switzerland.
- [5] **Bunkóczi, G., Lehmann, C., Vértesy, L. & Sheldrick, G. M.** (2001). High affinity binding to glycopeptide antibiotics. *Sixth International School on the Crystallography of Biological Macromolecules*, May 13-17, Como, Italy.

LEBENS LAUF

PERSÖNLICHE DATEN

Name	Gábor Bunkóczy
Geburtsdatum	01. Oktober 1976
Geburtsort	Debrecen, Ungarn
Staatsangehörigkeit	ungarisch
Familienstand	verheiratet

SCHULBILDUNG

1991-1995	Fazekas Mihály Gimnázium, Budapest, mit Abschluss Abitur
-----------	--

HOCHSCHULSTUDIUM

1995-2000	Eötvös Lorand Universität Budapest
1999-2000	Diplomarbeit am Lehrstuhl für Theoretische Chemie im Arbeitskreis von Prof. G. Náray-Szabó: „Röntgenkristallographische Untersuchung von Kalmodulin-Bisindolantagonist Komplexen”

PROMOTION

2000-2004	Dissertation am Lehrstuhl für Strukturchemie der Georg-August-Universität Göttingen im Arbeitskreis von Prof. G.M. Sheldrick: „Structure determination of peptides with antibacterial action”
-----------	--

Göttingen, den 16.03.2004

Gábor Bunkóczy