Crystallographic studies on DNA in complex with antibiotics and effects of anisotropic scaling on structure solution

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

vorgelegt von
Roland Pfoh
aus Bad Hersfeld

Göttingen, 2008

D 7

Acknowledgements

Abbreviations

A	adenine
ALA	L-alanine
BESSY	Berliner Elektronenspeicherring-Gesellschaft für Synchrotron-Strahlung
C	cytosine
CC	correlation coefficient
CCP4	Collaborative computational project number 4
dm	density modification
DESY	Deutsches Elektronen Synchrotron
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
ESI	electrospray ionization
ESI(+)	positive ion electrospray ionization
ESI(−)	negative ion electrospray ionization
et al.	et alii
G	guanine
HPLC	high performance liquid chromatography
M	molar
MAD	multi-wavelength anomalous dispersion
MES	2-morpholino-ethanesulfonic acid
MIR	multiple isomorphous replacement
MVA	L-N-methyl-valine
NCY	L-N-methyl-cysteine
NMR	nuclear magnetic resonance
QX	quinoxaline
rms	root mean square
SLS	Swiss light source
SAD	single-wavelength anomalous dispersion
SER	D-serine
SIR	single isomorphous replacement
SIRAS	single isomorphous replacement with anomalous scattering
sp.	species
T	thymine
PEG	polyethylene glycol
PDB	protein data bank
PDB ID	identification code for entries in the protein data bank
v/v	substance volume per solvent volume
w/v	substance weight per solvent volume
wMPE	weighted mean phase error

Molecular graphics

The three-dimensional molecular graphics in this work were produced with the program PYMOL (*DeLano Scientific LLC,* South San Francisco CA, USA).

Abstract

Crystallographic studies on DNA in complex with antibiotics and effects of anisotropic scaling on structure solution

This work investigates the structure of complexes between DNA and antibiotics discussing the crystal structures of d(AACCGGTT)-trioxacarcin, d(CGTACG)-trioxacarcin and d(ACGTACGT)-echinomycin. Both trioxacarcin A and echinomycin are natural compounds produced by streptomyces. Trioxacarcin A intercalates between the DNA base pairs and additionally forms a covalent bond to N7 of a guanine base. In the first complex the non-terminal thymine is flipped out from the duplex whereas in the second one trioxacarcin induces the formation of a DNA hexaplex. Echinomycin on the other hand binds to duplex DNA by bisintercalation of its two quinoxaline residues. The crystal structure shows contacts between DNA bases and bivalent manganese ions which might be related with the DNA unwinding induced by echinomycin. All three crystal structures were solved by experimental phasing using MAD, SAD or SIRAS. A more methods-oriented part of the work investigates the effect of anisotropic scaling on structure solution. With the example of a D,L-alternating peptide it is shown that anisotropic scaling can improve the electron density maps dramatically in the case of severe anisotropic diffraction.

Abstract (deutsche Übersetzung)

Kristallstrukturanalyse von DNA in Komplexen mit Antibiotika und Effekte anisotroper Skalierung auf die Strukturlösung

Die Arbeit befasst sich mit der Strukturaufklärung von Komplexen zwischen DNA und Antibiotika, wobei die Kristallstrukturen der drei Komplexe d(AACCGGTT)-Trioxacarcin, d(CGTACG)-Trioxacarcin und d(ACGTACGT)-Echinomycin beschrieben werden. Sowohl bei Trioxacarcin A als auch bei Echinomycin handelt es sich um Naturstoffe, die von Streptomyceten hergestellt werden. Trioxacarcin A interkaliert zwischen den Basenpaaren der DNA und bindet außerdem kovalent zu N7 einer Guanin-Base. Im erstgenannten Komplex verursacht Trioxacarcin das Herausdrehen der nicht-endständigen Thymin-Base aus dem gebildeten Duplex, während bei dem Oligonukleotid d(CGTACG) die Bildung eines Hexaplexes induziert wird. Echinomycin bindet sich dagegen durch Bisinterkalation an DNA, wofür zwei Quinoxalin-Gruppen eingesetzt werden. Die Kristallstruktur zeigt eine Wechselwirkung der DNA-Basen mit Mangan(II)-Ionen, die wahrscheinlich durch die starke Verzerrung der DNA durch Echinomycin ermöglicht wird. Alle drei Kristallstrukturen wurden durch experimentelle Phasenbestimmung gelöst, wobei die Methoden MAD, SAD und SIRAS angewendet wurden. Ein methodischer Teil der Arbeit befasst sich mit dem Effekt der anisotropen Skalierung auf die Strukturlösung von Makromolekülen. Es wird am Beispiel einer Peptid-Struktur gezeigt, dass bei stark anisotrop streuenden Kristallen die Qualität der Elektronendichte-Karten mit Hilfe einer entsprechenden Skalierung deutlich verbessert werden kann.

Table of contents

Introduction

Bacteria develop and synthesize compounds that can interact with DNA in order to disturb biochemical pathways of their enemies. The producer itself has often highly specified protection mechanisms for its own DNA. Most antibiotics in nature are synthesized by streptomyces, which are mainly found in the soil and belong to the group of actinobacteria. The two DNA intercalators investigated in this work – trioxacarcin A and echinomycin – also originate from them.

Only a small amount of the DNA in the genome of different organisms including human is used for the encoding of structural proteins, the biological role of the rest is still unknown in most cases. The study of interactions between antibiotics and DNA might increase the knowledge about biochemical functions of DNA. Additionally, these compounds are interesting for pharmaceutical applications. Many natural compounds and their derivatives that interact with DNA are used to treat cancer and viral diseases. Detailed structural information about complexes between antibiotics and DNA makes it possible to obtain suitable drugs via specific chemical modifications. This is of course also true for synthetic inorganic compounds like *cis*-platin.

Macromolecular crystallography provides reliable structural information. Unfortunately, it is not possible to obtain suitable crystals from every DNA-drug complex because of the high flexibility of long DNA strands. Self complementary oligonucleotides containing six to twelve bases are usually used for initial crystallization experiments. If the binding type and the induced distortion of the DNA are not known from previous crystal structures, the phases have to be determined experimentally for structure solution. In most such cases it is necessary to crystallize the complex with a suitable heavy atom.

Trioxacarcin A shows anti-cancer and anti-malaria activity and is known to bind to double stranded DNA. It alkylates a guanine base and is able to extract this base at higher temperatures, leaving abasic DNA. Abasic DNA is very unstable and can lead to strand breakages with fatal consequences to the organism. It was possible to crystallize the drug with two different self-complementary oligonucleotides and determine two different crystal structures.

Echinomycin has been a leading candidate for anti-tumor agents and is being investigated in various clinical studies. It binds to duplex DNA by bisintercalation preferentially around CG steps of DNA. Aim of the project was the discovery of new crystal forms of the known DNA-echinomycin complex with different metal ions. Crystal structures with self-complementary oligonucleotides have been already determined, one of them revealed a metal ion at the bisintercalation site.

In a third, more method oriented project the effect of anisotropic scaling on structure solution is tested. Anisotropic diffraction affects a high amount of crystals and makes experimental phasing more difficult. Three different approaches of anisotropic scaling including local scaling were carried out with an unpublished protein and a D,L-alternating peptide.

1. Crystallographic background

1.1 STRUCTURE SOLUTION OF MACROMOLECULES

The crystallographic phase problem

Electron density $\rho(\mathbf{x})$ for every point $\mathbf{x} = (x, y, z)$ in the unit cell with volume V_C is given by the Fourier summation

$$\rho(\mathbf{x}) = \frac{1}{V_C} \cdot \sum_{\mathbf{h}} \mathbf{F}(\mathbf{h}) \cdot \exp(-i2\pi \cdot \mathbf{h}^T \mathbf{x}) \tag{1.1}$$

where $\mathbf{h}$ is the vector of the reciprocal lattice indices h, k, l and $\mathbf{h}^T$ the transposed matrix of it. The complex structure factor $\mathbf{F}(\mathbf{h})$ consists of an amplitude $F(\mathbf{h})$ and a phase $\varphi(\mathbf{h})$:

$$\mathbf{F}(\mathbf{h}) = F(\mathbf{h}) \cdot \exp(i \cdot \varphi(\mathbf{h})) \tag{1.2}$$

In a diffraction experiment only intensities $I(\mathbf{h})$ of reflections $\mathbf{h}$ can be measured, which are related to the structure factor amplitudes $F(\mathbf{h})$ ($I \sim F^2$). The phases $\varphi(\mathbf{h})$ are not directly obtainable. This is known as the crystallographic phase problem.

Once a structural model with atomic coordinates $\mathbf{x}$ for atom j is available, the structure factor can be calculated by a summation over every atom in the unit cell:

$$\mathbf{F}(\mathbf{h}) = \sum_{j} f_j \cdot \exp(i2\pi \cdot \mathbf{h}^T \mathbf{x}_j) \tag{1.3}$$

where f_j is the atomic scattering factor for atom j including the atomic vibration u.

Different methods can be used for initial structure solution of macromolecules. These methods are described below. Small molecule structures are often solved by direct methods, which use statistical relations between amplitudes and phases to solve the phase problem directly. Direct methods require atomic resolution (better than 1.2 Å) and therefore cannot be applied to macromolecular structures in most cases.

Anomalous Scattering

Anomalous scattering occurs when the wavelength of the radiation corresponds to an absorption edge of an atom. The scattered radiation undergoes a phase and amplitude shift because of interactions with the atom. This effect has to be taken into account for the calculation of the atomic scattering factor f_0^{anom}, which is now dependant on the wavelength λ:

$$f_0^{\text{anom}}(\lambda) = f_0 + f'(\lambda) + if''(\lambda) \tag{1.4}$$

In this equation f_0 is the atomic scattering factor without anomalous scattering, f' is the real part (amplitude shift) and f'' the imaginary part (phase shift) of the anomalous scattering contribution.

Since the structure factor $\mathbf{F(h)}$ is a sum of the scattering factors of all atoms in the unit cell, every reflection is influenced by an anomalous scatterer. In presence of an anomalous scattering atom, Friedel's law is no longer valid:

$$\mathbf{F(h)} \neq \mathbf{F(-h)} \tag{1.5}$$

Macromolecules contain mainly hydrogen, carbon, nitrogen and oxygen. The effects of anomalous scattering are very small because the absorption edges of these elements are too far away from the wavelengths used in X-ray crystallography. For heavier elements like sulfur it is possible to measure anomalous differences accurately enough to solve macromolecular structures. The magnitude of the anomalous scattering contributions is specific for each element (Rhodes, 2000). The phase of the anomalous scattering contribution depends on the position of the heavy atom in the unit cell, which can be determined by Patterson methods. Thus, the members of a Friedel pair can be used to estimate the phase of a reflection.

Experimental phasing

The starting phases φ_T of the total macromolecular structure can be obtained with the calculated phase of the heavy atom substructure φ_{HA}:

$$\varphi_T = \varphi_{HA} + \alpha \tag{1.6}$$

α describes the phase shift and can be estimated from the experimental data.

Isomorphous replacement

Isomorphous replacement was historically very important, because the first protein structures were solved with this method. It does not use anomalous scattering. It depends on the fact that the presence of heavy elements like mercury or uranium effects the intensity of reflections. For *single isomorphous replacement* (SIR) a derivative crystal containing the heavy element is measured in addition to the native crystal. In the case of *multiple isomorphous replacement* (MIR) two or more derivatives are measured and each of them has to contain a heavy element on a different site in the unit cell. For a successful experiment the used crystals have to be isomorphous, which practically means that the unit cell lengths should not differ by more than about 1 Å from each other.

For SIR, the following approximation can be used:

$$\left| F_{\text{derivative}} \right| - \left| F_{\text{native}} \right| = \frac{2f'}{f_0} \cdot \left| F_{\text{HA}} \right| \cdot \cos \alpha \tag{1.7}$$

For reflections with large normalized isomorphous differences α will be close to 0° or 180°. Although differences between amplitudes of the derivative and the native crystal are larger than differences of a Friedel pair, isomorphous replacement suffers a lot from non-isomorphism.

Multi-wavelength anomalous dispersion (MAD)

A MAD experiment is probably the most reliable experimental phasing method for macromolecules because it does not suffer from non-isomorphism and produces a high amount of phase information. For a Se-MAD experiment methionine is replaced by seleno-methionine to introduce an anomalous scattering atom to the protein. The disadvantage of Se-MAD is the time consuming preparation of proteins. Structures containing DNA can be solved by replacing thymine with bromo-uracil. MAD produces about twice as much phase information compared to *single-wavelength anomalous dispersion* (SAD). This is important especially for crystals with weak diffraction power. MAD experiments have to be carried out at synchrotrons because at least two different wavelengths are necessary, which have to be adapted to the anomalous scattering atom. One dataset has to be collected at the peak wavelength, the other one at the wavelength of the inflection point of *f*". Before each MAD experiment these points have to be determined with a fluorescence scan from the crystal, Figure 1.1 shows such a scan at the absorption edge of bromine. Sometimes data are also collected at a third wavelength with higher energy. In any case, the peak dataset should be collected first and the inflection dataset last. Only this way *radiation damage* will support the experiment but will harm it if it is carried out in a different order.
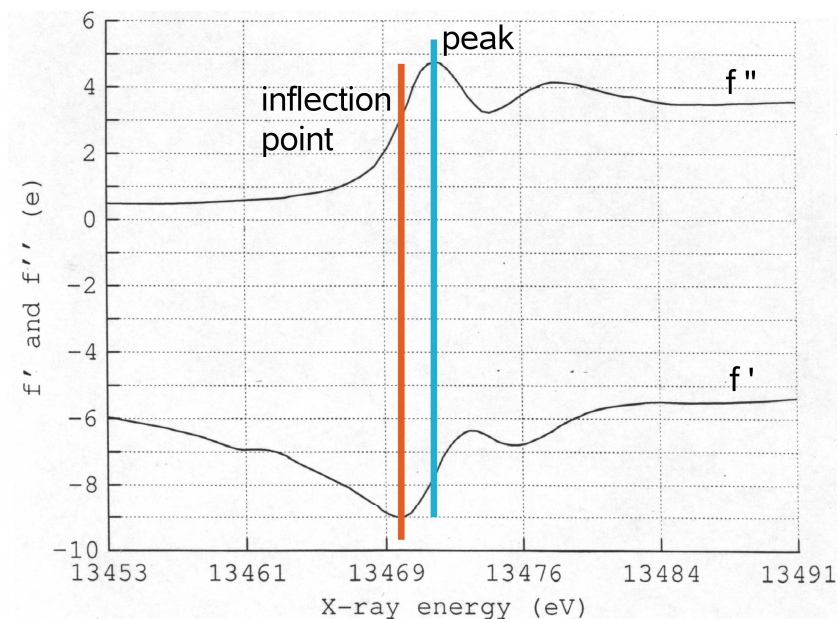
Figure 1.1. Fluorescence scan at the absorption edge of bromine. The scan belongs to dataset d(AACCGG[^{5Br}U]T) which is discussed in chapter 2.

Karle (1980) and Hendrickson *et al.* (1985) showed that the measured intensities are given by

$$|F(+\mathbf{h})|^2 = |F_\mathrm{T}|^2 + a|F_\mathrm{HA}|^2 + b|F_\mathrm{T}| \cdot |F_\mathrm{HA}| \cdot \cos\alpha + c|F_\mathrm{T}| \cdot |F_\mathrm{HA}| \cdot \sin\alpha \tag{1.8}$$

and

$$\left|F(-\mathbf{h})\right|^2 = \left|F_{\mathrm{T}}\right|^2 + a\left|F_{\mathrm{HA}}\right|^2 + b\left|F_{\mathrm{T}}\right|\cdot\left|F_{\mathrm{HA}}\right|\cdot\cos\alpha - c\left|F_{\mathrm{T}}\right|\cdot\left|F_{\mathrm{HA}}\right|\cdot\sin\alpha \qquad (1.9)$$

where $a = (f''^2 + f'^2)/2$, $b = 2f'/f_0$ and $c = 2f''/f_0$.

If data for two wavelengths are available, the phase problem can be solved because there are four observations [$F(\pm\mathbf{h})$ at peak and inflection point] an three unknowns (F_{T}, F_{HA} and α). The best results are obtained if the differences in f' and the sum of f'' are both large, therefore the wavelengths at peak and inflection point are used.

Single isomorphous replacement with anomalous scattering (SIRAS)

A good alternative to MIR is a SIRAS experiment, because only one heavy atom derivative is required and the obtained electron density maps are of high quality, almost comparable to the ones from MAD. (In practice MAD data are better than SIRAS data because of isomorphism.) It is possible to collect the derivative dataset on a home source, if the heavy atom shows significant anomalous scattering at the used wavelength. Additionally, a native dataset without the heavy atom has to be collected.

Single-wavelength anomalous dispersion (SAD)

In the last few years SAD experiments have become more and more successful, probably because it became possible to collect high quality data with highly sensible detectors. SAD experiments can be performed at home sources, because only one wavelength is necessary. It has to lie in the appropriate range of the anomalous scatterer. Measurements at home sources have the advantage that high redundant data can be collected. This is often not possible at synchrotrons because the crystal can only be rotated around a fixed axis at most beam lines. The quality of electron density maps obtained by SAD is usually not as high as compared to MAD or SIRAS maps.

SAD can be regarded as a special case of MAD with data from only one wavelength instead of two (or more). Assuming that

$$\left|F_{\mathrm{T}}\right| = 0.5\cdot\left[\left|F(+\mathbf{h})\right| + \left|F(-\mathbf{h})\right|\right] \qquad (1.10)$$

a subtraction of the MAD equations (1.8) and (1.9) gives

$$\left|F(+\mathbf{h})\right| - \left|F(-\mathbf{h})\right| = c\cdot\left|F_{\mathrm{HA}}\right|\cdot\sin\alpha \qquad (1.11)$$

Here, for reflections with large normalized isomorphous differences α will be close to 90° or 270°.

Molecular replacement

At the moment the most widely used method for macromolecular crystal structure solution is molecular replacement. It is a trial-and-error method which uses an already existing model that is related to an appreciable extent with the new structure. For each molecule three rotation angles and three translations have to be defined. The search strategies have to be computationally tractable (Evans & McCoy, 2008). The reason of the increasing success rate of molecular replacement is that in the last few years more models of different protein and DNA motives became available and computers became

faster. Additionally molecular replacement requires less experimental effort than the other methods and therefore is usually tried first. The disadvantage of molecular replacement is model bias, which becomes more severe at low resolutions. Model bias implies that the model of the solved structure will look to some extent like the old model, even if it is a bit different in reality. Related to this is the problem that errors in the parent models might sum up in the new structures. An extensive use of molecular replacement might actually lead to a decreasing quality of the obtained structures.

Summary

During 2006 over 67% of the crystal structures deposited in the PDB were solved using molecular replacement (Long *et al.*, 2007). The two most important methods for solving structures with experimental phases are SAD and MAD. 9.6% of the structures deposited during 2006 were solved by SAD, 9.4% by MAD. Only 1.8% of the structures were solved by MIR. Sometimes there is no way round experimental phasing, especially when the structure contains a high amount of unknown motives. Table 1.1 summarizes the different methods for structure solution with their different limitations.

Table 1.1. Summary of methods for structure solution.

method	comment
direct methods	resolution < 1.2 Å required.
molecular replacement	good search model required, works also for twinned crystals, only native dataset required, problem: model bias.
SAD	anomalous scatterer required, possible at home source, data need a high accuracy, density modification can better compensate for weak phase information when the resolution is high (< 1.7 Å) or the crystals have a high solvent content.
MAD	anomalous scatterer required, synchrotron required, works also at low resolutions, can produce high quality maps.
isomorphous replacement	different heavy atom derivatives required, map quality depends on number of derivatives, problem: non-isomorphism.
SIRAS	derivative crystal with anomalous scatterer (heavy atom) and native crystal required, possible at home source, can produce high quality maps.

1.2 DATA SCALING

After data collection from a single crystal the indexed reflections have to be integrated and scaled. These procedures depend on the used detector. Most modern integration programs for area detectors come along with a scaling program that was designed for them, for example SAINT with SADABS (Bruker AXS, Madison WI, USA), XDS with CORRECT (Kabsch, 1993), in HKL2000 DENZO with SCALEPACK (Otwinowski &

Minor, 1997), and in CCP4 MOSFLM with SCALA (Collaborative computational project number 4, 1994). The data have to be scaled because symmetry-related reflections were measured on different images. Each image is taken at a different orientation of the crystal and the cryosolution around it. This causes differences in absorption and also in the number of unit cells the X-rays are passing. In some synchrotrons there is a decay of beam intensity with time. At modern synchrotrons with high beam intensity the crystal is seriously damaged during data collection, which results in diffraction decay with collection time.

For X-ray diffraction, the scattering factor f_0 of a spherical atom depends on the atom type and $(\sin\theta)/\lambda$, where θ is the wave angle with the reflection plane and λ the wavelength of the X-ray beam. By Bragg's law $(\sin\theta)/\lambda$ is given by $(1/2d)$ for first order reflections, where d is the interplanar spacing. The scattering factor decreases as $(\sin\theta)/\lambda$ increases, which is due to the specific volume of the electron cloud that diffracts the X-rays (Stout & Jensen, 1989). This results in a diffraction fall-off for higher resolutions.

Thermal motion, a vibration of the atom about its rest point, is also affecting the scattering power. Its magnitude depends on temperature, atom mass and the chemical environment of the atom. Thermal motion spreads the electron cloud over a larger volume, this further pronounces the fall-off of the scattering power with increasing $(\sin\theta)/\lambda$. A scattering factor f including the effect of thermal motion can be calculated by the expression

$$f = f_0 \cdot e^{-B(\sin^2\theta)/\lambda^2} \tag{1.12}$$

where B is related to the mean-square amplitude of the atomic vibration u:

$$B = 8\pi^2 \cdot \overline{u^2} \tag{1.13}$$

The dimension of the B value is Å^2.

Wilson (1942) proposed a statistical approximation of the B value and a conversion factor that places the observed intensities I_{rel} on an absolute scale. Such a factor can be obtained by relating the *average* observed intensities $<I_{rel}>$ with theoretically calculated *average* intensities $<I_{abs}>$:

$$\left\langle I_{rel} \right\rangle = C \cdot \left\langle I_{abs} \right\rangle \tag{1.14}$$

The average intensity depends mainly on the overall cell content, the position of the atoms plays a minor role. Assuming a random distribution of N atoms in the unit cell, the average intensity is given by

$$\left\langle I_{abs} \right\rangle = \sum_{i=1}^{N} f_i^2 \tag{1.15}$$

Combining equation (1.12) with (1.15) and assuming that B has the same value for all atoms results in

$$\langle I_{abs} \rangle = e^{-2B(\sin^2 \theta)/\lambda^2} \cdot \sum_{i=1}^{N} (f_0^2)_i \qquad (1.16)$$

With this information, equation (1.14) can be brought into the following form:

$$\ln \frac{\langle I_{rel} \rangle}{\sum_{i=1}^{N} (f_0^2)_i} = \ln C - 2B(\sin^2 \theta)/\lambda^2 \qquad (1.17)$$

The variation of the intensity with $(\sin \theta)/\lambda$ is avoided by dividing reciprocal space into concentric shells with assumed constant intensity. If the left side of equation (1.5) is evaluated for each of the shells and these values are plotted against $(\sin^2 \theta)/\lambda^2$, a Wilson plot is obtained (Wilson, 1942). The B value can be derived from the slope and C from the intercept at $(\sin^2 \theta)/\lambda^2 = 0$. Corresponding to equation (1.14) scale constant $k = C^{-0.5}$ can be used to convert relative amplitudes into absolute ones.

Anisotropic data scaling

In reality, the thermal motion of atoms mentioned above is not isotropic, it depends on the direction. An anisotropic motion can be described by a symmetric tensor:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix} \qquad (1.18)$$

$\boldsymbol{\beta}$ is dimensionless, because it contains the reciprocal cell dimensions. In the following form the anisotropic parameters b_{ij} are expressed on the same scale as the isotropic B value (Sheriff & Hendrickson, 1987):

$$\boldsymbol{\beta} = \frac{1}{4} \begin{pmatrix} a*^2 b_{11} & a*b*b_{12} & a*c*b_{13} \\ a*b*b_{21} & b*^2 b_{22} & b*c*b_{23} \\ a*c*b_{31} & b*c*b_{32} & c*^2 b_{33} \end{pmatrix} \qquad (1.19)$$

where $a*$, $b*$ and $c*$ are the reciprocal lattice edges.

If a great number of thermal vibrations have the same direction inside a crystal, the whole diffraction becomes anisotropic. In this case the reciprocal lattice does not have the shape of a sphere any more but of an ellipsoid. This can be directly visible in diffraction images, when the maximum resolution depends on the direction. Figure 1.2 shows an anisotropic diffraction pattern of a crystal containing DNA. In one direction the crystal diffracts to more than 1.3 Å resolution, whereas in another direction to less than 1.5 Å resolution.
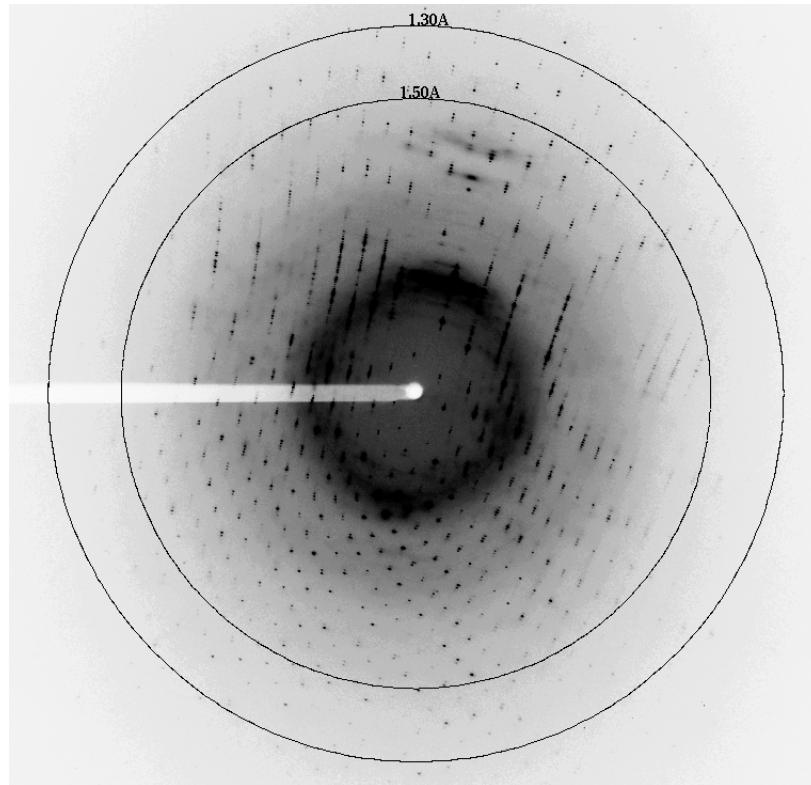
Figure 1.2. Anisotropic diffraction image of a DNA-drug complex (belonging to SLS dataset discussed in chapter 4).

If the resolution is high (better than about 1.5 Å) the individual atoms can be refined anisotropically and any overall anisotropy will be included. In this case, the expression for the scattering factor f expands to

$$f = f_o \cdot e^{-2\mathbf{h}^{\mathrm{T}}\boldsymbol{\beta}\mathbf{h}} \tag{1.20}$$

with

$$\mathbf{h}^{\mathrm{T}}\boldsymbol{\beta}\mathbf{h} = \frac{1}{4}(a*^2 h^2 b_{11} + 2a*b* hk b_{12} + 2a*c* hl b_{13} + b*^2 k^2 b_{22} + 2b*c* kl b_{23} + c*^2 l^2 b_{33}) \tag{1.21}$$

In most cases, macromolecules cannot be refined anisotropically, because the ratio between observations and parameters would become too small. Here, overall anisotropy can be taken into account by a TLS refinement, which treats a group of atoms as rigid body and refines its movement (Schomaker & Trueblood, 1968). Three tensors **T**, **L** and **S** are employed for this method. **T** and **L** are used for the description of translation and libration vibrations and **S** for the correlation between them, and then covers the case of a screw motion. The advantage of TLS is that a suitable description of anisotropy is obtained with only a few parameters required for refinement.

If anisotropy is taken into account during structure refinement, the aim is to bring the calculated structure factor F_c closer to the observed data by adjusting the model. In principle it should make no difference if the overall anisotropy is included in the model or the observed data are corrected for it before refinement. The first alternative has the advantage that the observed data are kept more *original* and the effect of an anisotropic correction can be evaluated by looking at the R and R_{free} values (see below). On the other hand it might be necessary to take anisotropy into account when a model is not yet available, for example when structure solution is difficult. In the following, three methods for anisotropic data scaling are described. The different methods have been tested as part of this work; the results are given in chapter 5.

In the case of anisotropic diffraction the resolution limit will also be anisotropic. This means that an anisotropic dataset should have incomplete higher resolution shells if all data available are included. Anisotropic scaling works only for real data, it cannot transform noise into data.

Anisotropic Scaling by Sheriff & Hendrickson

The method proposed by Sheriff & Hendrickson (1987) treats anisotropic diffraction like thermal motion. Therefore the anisotropy depends on the diffraction direction and the resolution. If the content of the asymmetric unit is known, the anisotropy can be expressed as

$$I_{an1}(\mathbf{h}) = I_{obs}(\mathbf{h}) \cdot \exp(-2\mathbf{h}^{\mathrm{T}}\langle\boldsymbol{\beta}\rangle\mathbf{h}) \qquad (1.22)$$

where $I_{obs}(\mathbf{h})$ is the observed intensity and $I_{an1}(\mathbf{h})$ the anisotropically scaled intensity.

The overall anisotropic temperature parameter <$\boldsymbol{\beta}$> contains also the isotropic component and refers to the whole unit cell.

Anisotropic Scaling by Shakked

In the method from Shakked (1983) a normalization is used to describe anisotropic diffraction. The decrease of intensity with resolution is not included. The observed intensities are multiplied by the following anisotropic scaling term:

$$I_{an2}(\mathbf{h}) = I_{obs}(\mathbf{h}) \cdot A(\mathbf{h}) \qquad (1.23)$$

where

$$A(\mathbf{h}) = \frac{\sum_{i,j=1}^{3} h_i \cdot a*_i \cdot h_j \cdot a*_j \cdot A_{ij}}{\sum_{i,j=1}^{3} (h_i \cdot a*_i) \cdot (h_j \cdot a*_j)} \qquad (1.24)$$

A_{ij} are the components of the symmetric scaling tensor **A**, h_i and a_i^* are the Miller indices and reciprocal unit cell vectors, respectively.

Local scaling

Local scaling proposed by Matthews & Czerwinski (1975) takes systematic errors into account that cause consistently larger or smaller structure factors in different regions of reciprocal space (e.g. absorption, misalignment, anisotropic diffraction).

An overall scale factor is determined for the whole dataset and therefore has a higher statistical certainty. A local scale factor on the other hand is determined for reflections in a local region, for example in a sphere in reciprocal space with a certain radius. Assuming the reflections inside this region have a constant systematic error it can be eliminated by multiplying the amplitudes of these reflections with the local scale factor.

The local scale factor K should minimize the summation over the reflections in the local area:

$$\min = \sum_{\mathbf{h}} w \cdot \left(K^{-1} - \frac{F_2(\mathbf{h})}{F_1(\mathbf{h})} \right) \tag{1.25}$$

Here, F_1 is the structure factor amplitude of data which is assumed to be error free and F_2 is the structure factor amplitude of data with a certain error that is assumed to be constant in the local scaling area. w is a weighting factor.

This leads to

$$Q = \frac{\sum F_1 F_2}{\sum F_1^2} \tag{1.26}$$

and

$$\sigma(Q) = \sqrt{\frac{\sum (QF_1 - F_2)^2}{(n-1) \cdot \sum F_1^2}} \tag{1.27}$$

with $Q = 1/K$ and $\sigma(Q) \approx \sigma(K)$.

The average fluctuation of the local scale factor $<\Delta K>$ is given by the rms value of $(K-1)$ for all local scale factors. The average systematic variation of the local scale factor is

$$\left\langle \Delta K_s \right\rangle = \sqrt{< \Delta K >^2 - \sigma(K)^2} \tag{1.28}$$

By averaging over all $\sigma(K)$ values one can obtain an estimate of the average random error $<\sigma(K)>$.

This local scale factor is determined by a smaller quantity of reflections than the overall one and has a higher statistical uncertainty. A critical point is the choice of the local area. For an optimal application the average systematic variation in the local scale

factor $<\Delta K_s>$ should exceed the average error of its estimation $<\sigma(K)>$ to the greatest possible amount.

1.3 STRUCTURE REFINEMENT

The structure model as well as the experimental data contain always errors. Therefore the calculated structure factor amplitudes F_{calc} differ from the observed structure factor amplitudes F_{obs}. The phases can be calculated more precisely if the model is adequate. This includes a proper solvent model for macromolecules, precise coordinates of atoms and a correct description of their thermal motion. During the refinement of parameters a target function is minimized or maximized. This can be done by the methods of least squares (Cruickshank, 1970):

$$\sum_{\mathbf{h}} w_{\mathbf{h}} (|F_{obs}(\mathbf{h})|^2 - |F_{calc}(\mathbf{h})|^2)^2 = \min \qquad (1.29)$$

where $w_{\mathbf{h}}$ is a weighting factor that takes into account that some reflections have been measured more precisely than others. The least squares refinement is sensitive to large errors in the starting model and therefore works best at resolutions better than 1.7 Å.

For incomplete models at lower resolutions maximum likelihood refinements are useful. Here, a probability function is maximized:

$$\prod_{\mathbf{h}} P(|F_{obs}(\mathbf{h})|; |F_{calc}(\mathbf{h})|) = \max \qquad (1.30)$$

where P is the probability distribution (Pannu & Read, 1996).

Both methods require additional chemical information in the form of constraints and restraints, which increase the data to parameter ratio. A constraint is an exact mathematical condition and reduces the number of parameters. Examples are constraints for special positions, rigid groups, riding hydrogen atoms, fixed occupancies and free variables [used in SHELXL (Sheldrick & Schneider, 1997)]. A restraint on the other hand has a standard deviation and increases the number of observations. Examples are restraints for distances and angles, planarity and chiral volumes, antibumping (prevents non bonded atoms coming too close to each other), non-crystallographic symmetry, isotropic and anisotropic thermal motions.

Validation criteria

The fit between atomic model and observed data can be described by the R value:

$$R = \frac{\sum_{\mathbf{h}} \left\| F_{obs}(\mathbf{h}) \right| - k \left| F_{calc}(\mathbf{h}) \right\|}{\sum_{\mathbf{h}} |F_{obs}(\mathbf{h})|} \qquad (1.31)$$

where k is a scale factor.

Especially at lower resolutions it is possible to overfit or misfit the diffraction data, because also incorrect models can be refined to good R values. To avoid this, Brünger

(1992) introduced the R_{free} value. It is calculated from a test set of reflections that is omitted in the modeling and refinement process. With this test set an overfitted model will give a worse R_{free} value than a model with a correct physical meaning.

The R_{free} value should lie between 3% and 5% over R_{work}, which is calculated without the reflections from the test set. If R_{free} differs more than 5% from R_{work}, the model is probably overfitted. A reason for this can be that too many water molecules were put in, for example. If R_{free} differs less than 3% from R_{work}, this could indicate non-crystallographic symmetry or twinning. For the test set usually 5% of the reflections from the original dataset are used.

1.4 DNA GEOMETRY

A right-handed double helical structure of DNA was postulated by Watson & Crick (1953) with the use of an X-ray fiber photograph taken by Rosalind Franklin. The Watson-Crick base pairing scheme is shown in Figure 1.3.
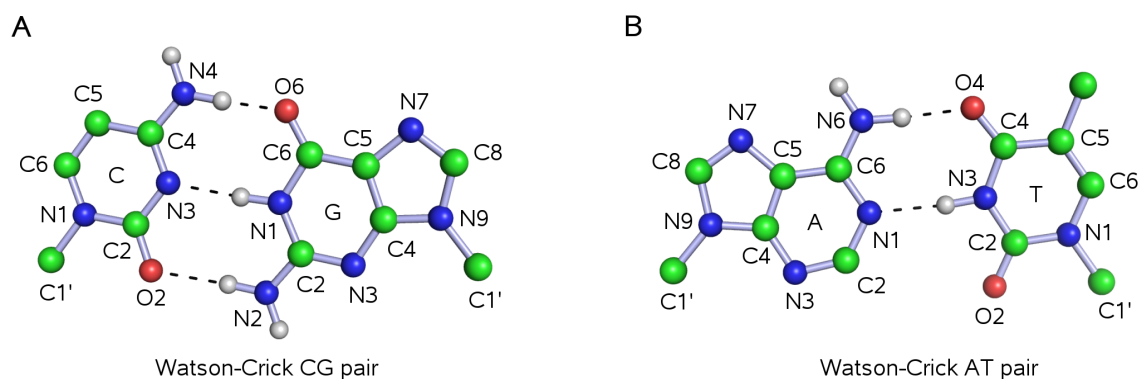
Figure 1.3. The Watson-Crick base pairing scheme for cytosine and guanine (**A**) and adenine and thymine (**B**). Hydrogen bonds are indicated by dashed lines.

Cytosine is forming three hydrogen bonds with guanine whereas adenine is forming two hydrogen bonds with thymine. For some time this was regarded as being the only biologically relevant structure. An alternative base pairing scheme described by Hoogsteen (1959) is often observed between adenine and thymine bases (Figure 1.4).

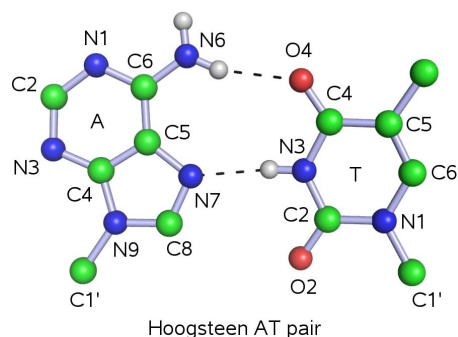**Figure 1.4.** The Hoogsteen base pairing scheme for adenine and thymine. Hydrogen bonds are indicated by dashed lines.

To form it, adenine rotates by about 180° around the glycosyl bond between C1' and N9. Now N7 of adenine is involved in hydrogen bonding instead of N1. Today it is established that the DNA molecule can convert between various other structures depending on the sequence and the environment (Bansal, 1999). Nearly every letter of the alphabet has been used to describe the different forms of DNA (Ghosh & Bansal, 2003).

B-DNA shown in Figure 1.5A resembles strongly the original model from Watson & Crick. It has a helical twist of 36.5° per base pair. The right handed helix geometry can be realized by following the blue line in Figure 1.5A, which connects the phosphorus atoms belonging to one strand. The major groove is wide whereas the minor groove is narrow. B-DNA is still regarded as the most important biological form, followed by A-DNA. A left-handed double helix called Z-DNA was discovered by Wang *et al.* (1979), shown in Figure 1.5B. It is often observed for sequences with alternating cytosine (pyrimidine) and thymine (purine) bases ([CG]$_n$). The repeating unit of Z-DNA is a dinucleotide, resulting in a characteristic zigzag backbone indicated by a blue line. The minor groove is narrow and deep whereas the major groove is flat.
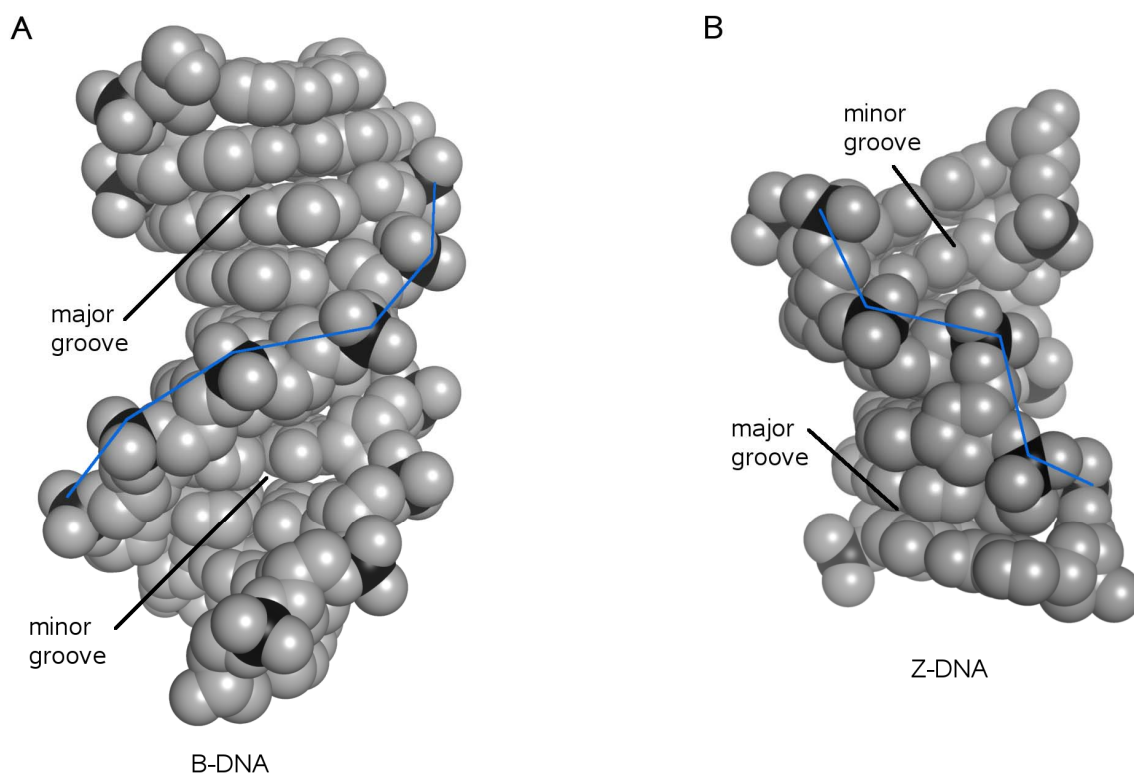
Figure 1.5. Space filling view of B-DNA (**A**) (PDB ID 1en3) and Z-DNA (**B**) (PDB ID 390d). Phosphorus atoms are shown in black. The phosphorus atoms belonging to one strand are connected by a blue line. B-DNA is a right-handed double-helix, Z-DNA a left-handed double-helix.

The conformational variability of DNA molecules is due to the backbone consisting of seven bonds with rotational freedom. Additionally the sugars contain five bonds with pseudo-rotational freedom. Figure 1.6 shows the six backbone torsion angles α, β, γ, δ, ε, ζ and the glycosyl torsion angle χ in the nucleotide unit of DNA. The ranges and mean values for different conformational types of DNA have been determined by Schneider *et*

al. (1997) by analyzing available crystal structures of oligonucleotides. Table 1.2 compares the backbone and glycosyl torsion angles of B- and Z-DNA. The differences of the two forms are most obvious in the α torsion angle, the γ torsion angle for A/G (purines) and the glycosyl torsion angle for A/G.
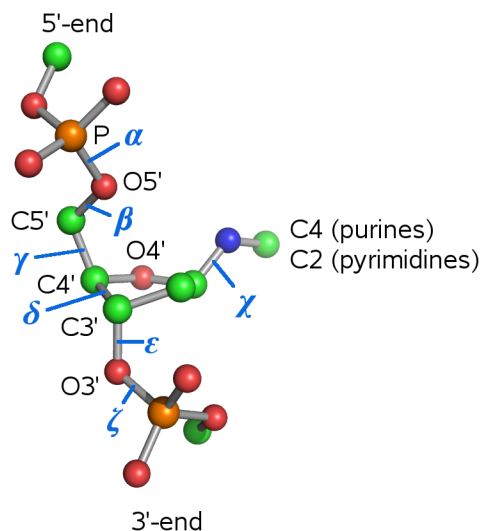
Figure 1.6. The backbone torsion angles α, β, γ, δ, ε, ζ and the glycosyl torsion angle χ in the nucleotide unit of DNA.

Table 1.2. Backbone and glycosyl torsion angle ranges of B- and Z-DNA.

	Angle range, degrees						
	α	β	γ	δ	ε	ζ	χ
B-DNA	270–330	130–200	20–80	70–180	160–270	150–300	200–300
Z-DNA (A/G)	40–100	150–250	160–210	80–160	180–300	280–340 40–100	50–90
(T/C)	150–250	150–250	20–90	80–160	180–300	40–100	180–220

In the case of Z-DNA values for purines (A/G) and pyrimidines (T/C) have to be distinguished.

The DNA structures described in the next chapters adopt B- and Z-geometry.

2. Guanine robbery – Trioxacarcin A bound to d(AACCGGTT)

2.1 INTRODUCTION

Trioxacarcins were first isolated in 1981 from the marine-derived micro-organism *Streptomyces bottropensis* DO-45 (Tomita *et al.*, 1981; Tamaoki *et al.*, 1981). They are cytotoxic against various cancer cell lines, active against Gram-positive and Gram-negative bacteria and exhibit anti-malaria activity (Maskey *et al.*, 2004a). Trioxacarcin A (Figure 2.1) contains a complex ring system that is attached to sugars at both ends, at the 4- and 13-positions, and causes it to exhibit intensive green fluorescence in solution; in powder form it is yellow.
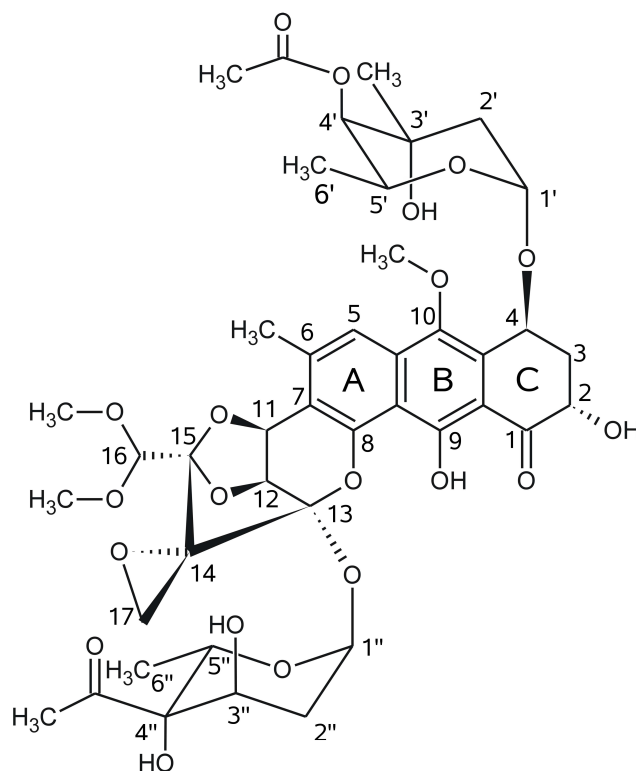
Figure 2.1 Schematic diagram of trioxacarcin A.

Nucleophilic attack of N7 of guanine opens the 'epoxide (**1**)' to form a covalent bond to the 'guanine (**2**)' in a DNA molecule (Figure 2.2). This alkylation is favored when a thymine is located on the 3'-side of guanine and does not take place when the guanine is terminal (Fitzner *et al.*, 2008). Cleavage of this 'trioxacarcin-DNA complex (**3**)' at 373 K results in the natural product 'gutingimycin (**4**)' (Maskey *et al.*, 2004a; Maskey *et al.*, 2004b), named after the ancient name for the city of Göttingen, leaving an abasic DNA. Presumably this cleavage takes place under milder conditions *in vivo*.

Anthracyclines resemble trioxacarcin A in that they also contain a planar aromatic ring system with one ore more sugars attached to it, although they can only intercalate but not

bind covalently to DNA. Some of them are used for the treatment of various cancer types, e.g. daunomycin (= daunorubicin) against leukemia. In DNA-daunomycin complexes the positively charged amino sugar of daunomycin is positioned in the minor groove of DNA (Wang *et al.*, 1987). The anthracycline nogalamycin has bulky sugar residues at both ends of the molecule that interact with both grooves of DNA (Williams *et al.*, 1990). Pluramycin antibiotics (Hansen & Hurley, 1996) such as hedamycin (Hansen *et al.*, 1995) and altromycin B (Sun *et al.*, 1993), and also psorospermin (Hansen *et al.*, 1996), are even more similar to trioxacarcin A because they contain in addition one or more epoxides and so can both intercalate and alkylate DNA; like trioxacarcin they bind covalently to the guanine N7. However no crystal structures have been reported of such covalent antibiotic-DNA complexes.
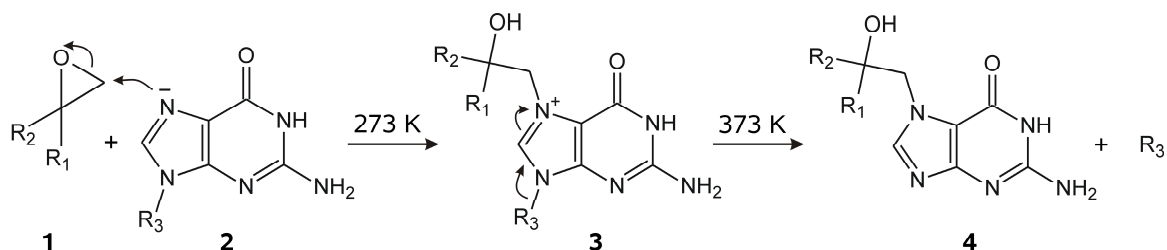

Figure 2.2 Proposed mechanism for the reaction of trioxacarcin A with DNA. Only the epoxide group of trioxacarcin A (**1**) is shown, the rest of the molecule is described by R_1 and R_2. R_3 stands for the abasic DNA.

2.2 MATERIALS AND METHODS

Trioxacarcin A was obtained from the marine *Streptomycin* sp. isolate B8652 by fermentation (Maskey *et al.*, 2002). Oligonucleotides were purchased already purified by HPLC from biomers.net GmbH and used without further purification. Crystals were grown at 40 °C in hanging drops by vapor diffusion; it was found by experiment that the higher than usual temperature produced better quality single crystals. The solution in the reservoir contained 1.55 M tri-ammonium citrate (pH 7.0) and 30% v/v DMSO. The DNA-drug solution contained 2.5 mM DNA (single-strand concentration), 2.8 mM trioxacarcin A and 25% v/v methanol. The DNA-drug solution was prepared by mixing trioxacarcin A stock solution (containing 5.0 mM trioxacarcin A and 50% v/v methanol) with a 5.6 mM DNA solution in a 1:1 ratio at room temperature and incubation for three days at 4 °C. The hanging drops prepared from 1 µl DNA-drug solution and 2 µl reservoir solution were equilibrated against 500 µl reservoir solution. Yellow plates grew within 24 h to a size of $0.2 \times 0.1 \times 0.05$ mm (Figure 2.3).
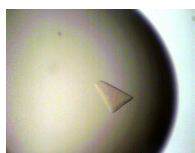

Figure 2.3. Yellow crystal plate of the trioxacarcin-d(AACCGGTT) complex.

For data collection at 100 K in a nitrogen gas stream the crystals were transferred to a cryosolution containing 1.55 M tri-ammonium citrate (pH 7.0), 30% v/v DMSO and 15%

v/v glycerol. Three different oligonucleotides were used to obtain crystals as described above: native d(AACCGGTT) and brominated d(AACCGG[5BrU]T) and d(AACCGGT[5BrU]). In the brominated oligonucleotides thymidine is replaced by 5-bromo-deoxyuridine.

The DNA-drug complex crystallized in space group P4$_1$22 with unit-cell $a = b = 37.60$ Å and $c = 91.60$ Å. Data were collected at beam line 14.2 at BESSY, Berlin with a MAR-165 CCD detector. The crystals containing brominated oligonucleotides were used for two Br-MAD experiments. From both crystals peak and inflection datasets were collected, 180 frames for each dataset with φ-rotation of 1° per frame (see Table 1 for crystallographic details). In order to avoid radiation damage, a 0.53 mm aluminum filter was used to decrease the intensity of the direct beam. In addition, a native dataset was collected to 1.78 Å resolution from a crystal containing the native oligonucleotide.

Table 2.1. Data collection, phasing and refinement statistics

	Native	d(AACCGG[5BrU]T)		d(AACCGGT[5BrU])	
Crystal data					
Space group	P4$_1$22	P4$_1$22		P4$_1$22	
a, Å	37.60	37.60		37.56	
c, Å	91.62	91.21		90.97	
		peak	inflection	peak	inflection
Diffraction data					
Wavelength, Å	0.92039	0.92032	0.92047	0.92033	0.92042
Resolution limit, Å	1.67	2.18	2.18	2.39	2.39
Total reflections	103552	45885	46392	36333	29272
Unique reflections	8225	3789	3790	2917	2756
Completeness, %					
Overall	99.0	89.7	98.7	98.8	93.4
Outermost resolution shell	94.7	89.3	98.5	90.7	59.4
$I/\sigma(I)$					
Overall	27.81	13.2	14.6	13.8	14.8
Outermost resolution shell	2.92	4.5	4.2	2.7	1.5
Phasing					
Resolution, Å		2.6			
pseudo-free CC after dm		0.61			
Refinement					
Reflections used	6488				
Resolution, Å	1.78				
R_{work}	22.0				
R_{free}	26.5				
rms deviation					
Bond length, Å	0.016				
Bond angles, °	2.5				
Average B factor (all atoms), Å^2	53.16				

The datasets were integrated with HKL2000 (Otwinowski & Minor, 1997) and the space group determinated by XPREP (Bruker AXS, Madison WI, USA). SHELXD (Schneider & Sheldrick, 2002) was used for substructure solution by searching for two bromine atoms with a resolution cutoff at 2.6 Å. Substructure solution succeeded only with the d(AACCGG[5BrU]T)-derivative, probably because the crystals diffracted better.

SHELXE (Sheldrick, 2002) was employed for phasing and density modification. In the experimental map the position of thymine could be deduced from the bromine positions from the substructure solution and the anomalous maps calculated using SHELXE. It was also possible to recognize the drug in the experimental map (Figure 2.4A). The asymmetric unit consists of one complete duplex with two intercalated trioxacarcins. The graphics program COOT (Emsley & Cowtan, 2004) was used for manual model building. The structure was refined isotropically with TLS constraints against the native dataset with REFMAC (Murshudov *et al.*, 1997) to R_{work} = 22.0% and R_{free} = 26.5% (Figure 2.4B). Helical parameters were calculated with 3DNA (Lu & Olson, 2003). The final coordinates have been deposited in the PDB (PDB ID 3c2j) together with the structure factors. The work in this chapter has already been published (Pfoh *et al.*, 2008).
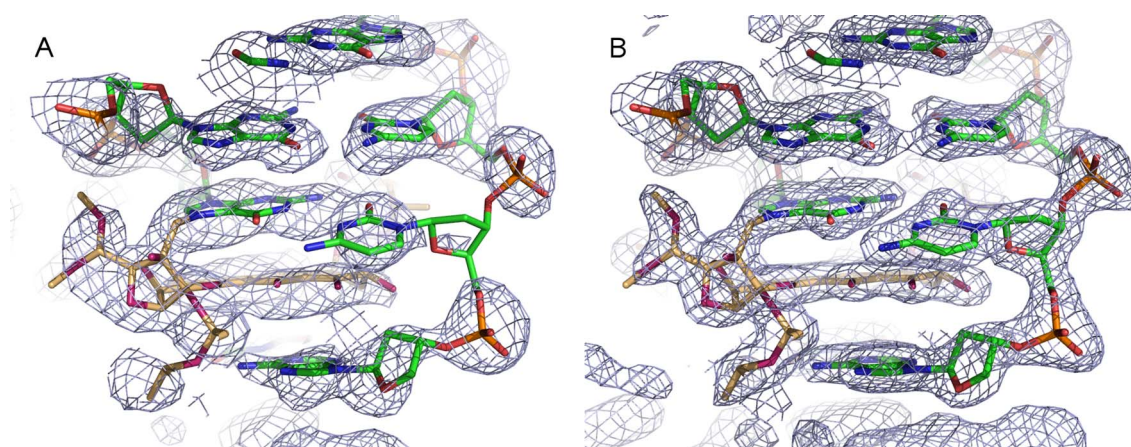

Figure 2.4 Experimental map (**A**) and map after final refinement (**B**), both contoured at a 1σ level. The carbon atoms of the trioxacarcin are shown in light orange.

2.3 RESULTS AND DISCUSSION

Overall structure

As predicted in the study of the sequence specificity (Fitzner *et al.*, 2008), trioxacarcin A binds covalently to d(AACCGGTT) by alkylating the N7 position of the guanine that is followed by a thymine. Unexpectedly this thymine is flipped out of the duplex DNA and the adenine originally paired with it now forms a base-pair with the following thymine in the sequence, a so-called register-shift. Trioxacarcin A intercalates at the 3'-side of the alkylated guanine (Figure 2.5). The aromatic rings A and B of the drug are involved in stacking interactions with the DNA. As found for nogalamycin (Williams *et al.*, 1990), trioxacarcin A interacts with both grooves of the DNA (Figure 2.6); the 4-sugar is positioned in the minor groove, the 13-sugar in the major groove.

Since the duplex does not lie on a two-fold axis, the two self-complementary strands are crystallographically independent, with some small structural differences caused by interactions of the residues A(1), A(101), T(7) and T(107), which are no longer base-paired within the duplex, with different symmetry related residues in the crystal. A(1) forms a Hoogsteen base-pair with a symmetry equivalent of T(7). In contrast, T(107) lies close to a symmetry equivalent of itself. Residue A(101) does not appear to make specific contacts to neighboring molecules and is highly disordered, whereas N1 of A(1) accepts

an hydrogen bond (2.7 Å) from a symmetry equivalent of 4"-OH of trioxacarcin(109). The general conformations of both strands are fairly similar except for residues A(1) and A(101) that are positioned at the 5'-terminus of the oligonucleotide and do not base-pair within the duplex.
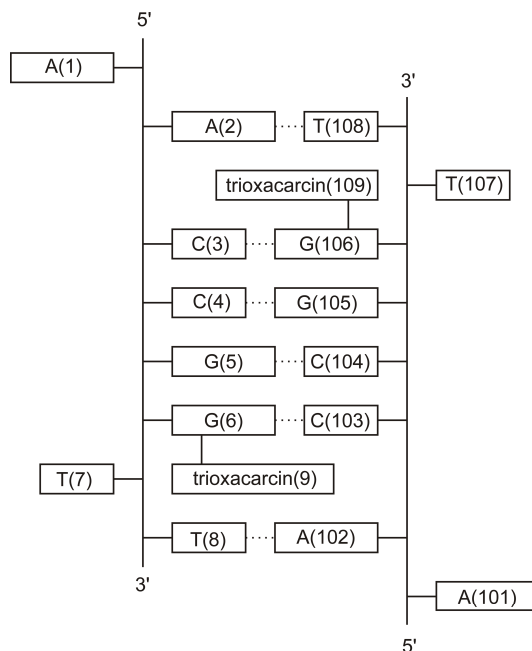
Figure 2.5 Simplified view of the DNA-trioxacarcin duplex with residue names.
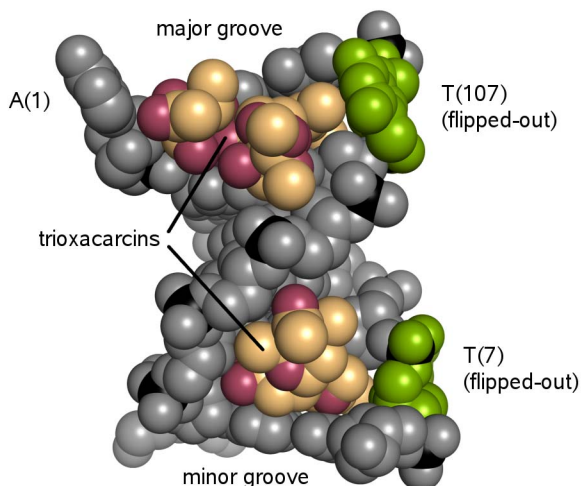
Figure 2.6 Space-filling view of the DNA-trioxacarcin duplex. Trioxacarcin is shown in red (oxygen atoms) and light orange (carbon atoms); the DNA is shown in light gray except for the phosphorus atoms that are shown in black and the two residues containing the flipped-out thymines in green. On the bottom is the minor groove visible enclosing the 4-sugar, on the bottom the major groove containing the 13-sugar (left side) and the two 16-methoxy groups (right-side). On the extreme upper left side the terminal A(1) can be seen.

The DNA-trioxacarcin duplex shows a distorted B-DNA geometry with Watson-Crick base-pairing. In Table 2, the sugar-phosphate and glycosyl torsion angles are compared

with the usual ranges and mean values for B-DNA (Schneider *et al.*, 1997). The flipping-out of T(7) is most evident in its ζ torsion angle of 82° [74º for T(107)] that differs by about 180° from the standard value. The other torsion angles all lie in ranges typically observed in B-DNA structures.

Table 1.2. Backbone and glycosyl torsion angles

Residue	Angle, degrees						
	α	β	γ	δ	ε	ζ	χ
A2 (next to intercalation)	282	176	60	155	254	280	209
A102	308	180	50	155	251	270	224
C3 (next to intercalation)	288	170	46	150	206	192	276
C103	287	157	52	147	218	188	274
C4	308	149	53	133	168	265	240
C104	298	139	56	134	183	238	242
G5	307	191	47	151	180	262	262
G105	311	181	48	149	181	268	259
G6 (next to intercalation)	293	183	51	144	238	264	273
G106	303	185	46	141	233	267	270
T7 (flipped-out)	302	159	61	147	251	82	252
T107	281	184	69	161	254	74	258
T8 (next to intercalation)	276	172	51	143	-	-	232
T108	236	184	81	130	-	-	221
B-DNA (range)	270–330	130–200	20–80	70–180	160–270	150–300	200–300
B-DNA (mean)	298	176 (I) 146 (II)	48	128 (I) 144 (II)	184 (I) 246 (II)	265 (I) 174 (II)	258 (A/G) 241 (T/C) 271 (II)

Torsion angles are defined as α: O3'-P-O5'-C5', β: P-O5'-C5'-C4', γ: O5'-C5'-C4'-C3', δ: C5'-C4'-C3'-O3', ε: C4'-C3'-O3'-P, ζ: C3'-O3'-P-O5', χ(Purines): O4'-C1'-N9-C4, χ(Pyrimidines): O4'-C1'-N1-C2.For the mean values BI and BII conformations are distinguished, in the case of the glycosyl torsion angle χ also purines and pyrimidines for the BI conformation.

Antibiotic-DNA interactions

In addition to the covalent bond, trioxacarcin A forms direct and water mediated hydrogen bonds with the DNA (Figure 2.7). There is an hydrogen bond between the 2-OH of trioxacarcin and O4' of the deoxyribose attached to the cytosine opposite to the alkylated guanine [2.8 Å for trioxacarcin(9) and 2.7 Å for trioxacarcin(109)]. In the minor groove, N2 of the alkylated guanine donates an hydrogen bond to 3'-OH of the 4-sugar of trioxacarcin (3.1 Å for both trioxacarcins). There is also a water mediated hydrogen bond between N3 of the alkylated guanine and 3'-OH. The 4-sugar interacts only with the alkylated guanine. In the major groove, the 1"-oxygen of the 13-sugar of trioxacarcin(109) accepts an hydrogen bond (3.2 Å) from N6 of residue A(2). In the case of trioxacarcin(9) and A(102), the corresponding hydrogen bond is mediated by a water molecule, giving rise to a small difference between the two strands. The 13-sugar is also involved in an internal hydrogen bond [3.4 Å for trioxacarcin(9) and 3.1 Å for trioxacarcin(109)] between 3"-OH and 14-OH (Figure 2.7), stabilizing the OH-group formed by the nucleophilic attack on the epoxide ring. The 13-sugar of the trioxacarcin is

held parallel to the trioxacarcin chromophore by this internal hydrogen bond, whereas the 4-sugar is orientated perpendicular to it.
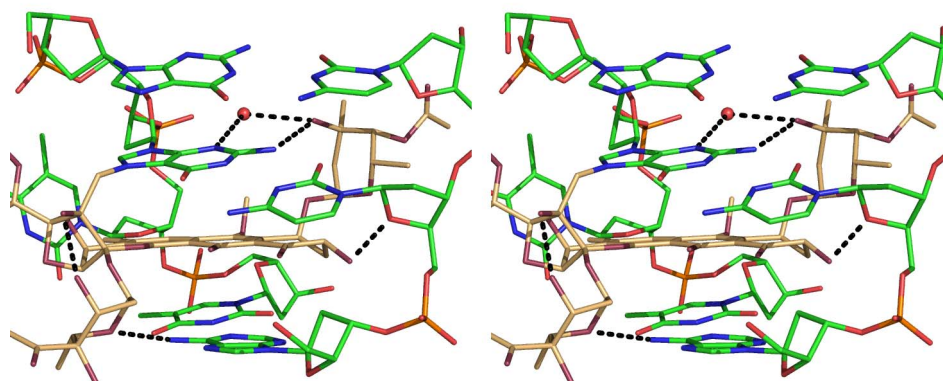

Figure 2.7 Stereo view of the DNA-trioxacarcin duplex. Base-pair G(105)-C(4) is at the top, followed by base-pair G(106)-C(3). Trioxacarcin A is bound to G(106) and intercalates on its 3'-side between basepair G(106)-C(3) and basepair T(108)-A(2), which is at the bottom. The carbon atoms of the trioxacarcin are shown in light orange. The flipped-out thymine is visible behind the trioxacarcin on the left. The red sphere represents a water molecule. Hydrogen bonds are indicated with dashed lines.

For the formation of gutingimycin, the guanine-sugar bond between N9 and C1' (Figure 2.8) has to break and the antibiotic has to leave the DNA duplex taking the guanine with it. Assuming that the guanine-sugar bond breaks whilst the double strand is still intact, gutingimycin would then only be attached to the abasic DNA by two hydrogen bonds because one hydrogen bond involved a symmetry equivalent that would not be relevant in solution and the other two hydrogen bonds that held it in place are between the 4-sugar of the trioxacarcin and the guanine.
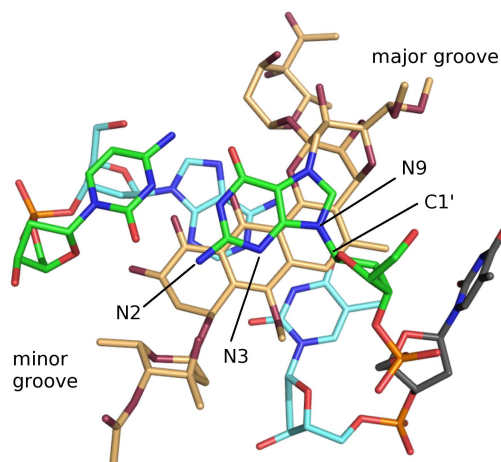

Figure 2.8 View down the DNA helix axis. The carbon atoms of the upper base-pair G(106)-C(3) are drawn in green, the ones of the lower base-pair T(108)-A(2) in turquoise. The trioxacarcin is positioned between the two base-pairs, with its carbon atoms colored in light orange. The carbon atoms of the flipped-out base T(107) are drawn in gray.

One trioxacarcin is involved in stacking interactions with G(106) and A(2) (Figure 2.8), the other with G(6) and A(102). Residues T(108) and T(8) interact slightly with the antibiotics but C(3) and C(103) do not. In the case of daunomycin or nogalamycin the long axis of the aglycone is nearly perpendicular to that of the base-pairs (Wang *et al.*, 1987, Williams *et al.*, 1990), but this is not the case for trioxacarcin, which is constrained by the alkylation site. The aromatic rings A and B of the trioxacarcins lie below the atoms N9, N3 and C1' of residues G(106) and G(6). The long axis of the trioxacarcin aglycone is nearly parallel to that of the alkylated guanine (Figure 2.8) and runs close to the sugar-phosphate backbone of this guanine. This orientation brings the 10-methoxy groups of the trioxacarcins approximately into the positions where the deoxyriboses of residues T(107) and T(7) would lie if they were not flipped out. The flipped-out thymines are positioned near the 6-methyl groups of the trioxacarcins, the distance between C2 of T(7) and 6-methyl of trioxacarcin(9) is 3.3 Å [3.6 Å for the corresponding atoms of T(107) and trioxacarcin(109)].

Distortion of the DNA

The following analysis is based on standard nomenclature of nucleic acid structure parameters (Olson *et al.*, 2001). The base-pairs of the DNA-trioxacarcin duplex are distorted in several different ways. At the intercalation site they are not planar; T(108)-A(2) is buckled by 9° and G(106)-C(3) by –10° [T(8)-A(102) by –6° and G(6)-C(103) by 13°]. This effect is well known for intercalators and explained by the need to maximize van der Waals contacts. The helical twist angles from base-pair T(108)-A(2) to base-pair A(102)-T(8) are 48°, 35°, 40°, 37° and 45°; for B-DNA 37° is typical (Olson *et al.*, 2001). Non-covalent intercalators such as daunomycin or nogalamycin usually unwind DNA (reducing these angles), whereas the intercalation of trioxacarcin combined with the flipping-out of a base in one of the two strands leads to an increased helical twist and an opening of the base-pair (rotation of the bases relative to each other in the plane of the base-pair) for T(108)-A(2) of 10° [8° for T(8)-A(102)] towards the major groove. The angle between base-pairs T(108)-A(2) and G(106)-C(3) (tilt) is –12° and opens towards the strand with the flipped-out thymine T(107) [corresponding values for T(8)-A(102)/G(6)-C(103)], also visible in Fig. 2.7. Another distortion induced by the trioxacarcin is a displacement along the long axis of base-pairs T(108)-A(2) and G(106)-C(3) (slide) by 2.0 Å [2.2 Å for T(8)-A(102)/G(6)-C(103)], also visible in Figure 2.8. All these distortions are within the ranges observed for both complexed and uncomplexed double-stranded DNA structures.

Crystal packing

The duplexes form infinite columns perpendicular to the *c* axis of the crystals, as shown in Figure 2.9. The columns running parallel to each other are lying in planes perpendicular to the *c* axis, forming a criss-crossed pattern with the columns in the neighboring planes. The columns themselves are held together by terminal base pair A(2)-T(108) stacking on T(8)-A(102) of a symmetry related duplex, visualized in Figure 2.10. The Hoogsteen base pair formed by the flipped out residues T(7) and A(1) stacks on a symmetry equivalent pair. This connects the columns running perpendicular towards each others (colored in yellow and gray in Figure 2.10 corresponding to Figure 2.9).
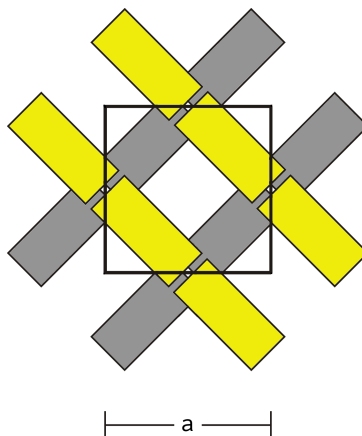
Figure 2.9 Schematic diagram of the crystal packing in the unit cell with view along the *c* axis.
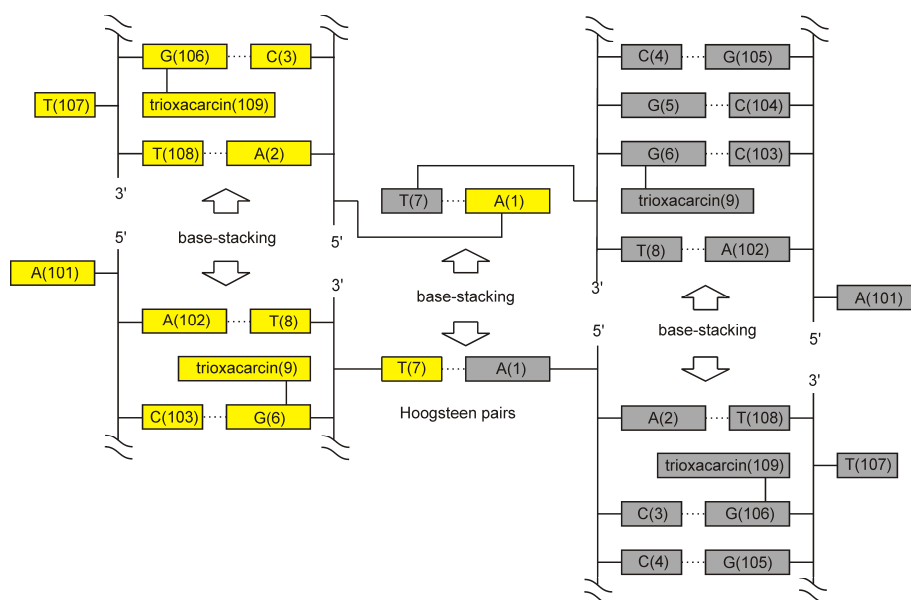
Figure 2.10 Crystallographic contacts between symmetry related molecules. The column colored in yellow is running perpendicular to the one in gray. The colors correspond to Figure 2.9.

Comparison with related DNA-antibiotic complexes

Structures of anthracyclines intercalated in DNA without the formation of covalent bonds invariably show the drug intercalated at the 5'-side of a guanine, usually between C and G or T and G. NMR studies of hedamycin-DNA (Hansen *et al.*, 1095; Pavlopoulos *et al.*, 1996; Owen *et al.*, 2002), altromycin-DNA (Sun *et al.*, 1993) and psorospermin-DNA (Hansen *et al.*, 1996) complexes, in which a covalent bond is formed by nucleophilic attack of the guanine N7 on an epoxide as in trioxacarcin, showed a similar intercalation site to that observed for the anthracyclines, namely on the 5'-side of the alkylated guanine, without a flipped-out nucleobase. The crystal structure of trioxacarcin bound to d(AACCGGTT) reported here reveals a quite different intercalation on the 3'-side of guanine combined with a flipped-out thymine on the same side. This result is consistent with sequence selectivity studies (Fitzner *et al.*, 2008) that report a preferred adduct

formation with 5'-AATTTGTAATT or 5'-AATTAGTAATT compared to 5'-AATTTGAAATT or 5'-AATTAGAAATT, showing that only a variation on the 3'-side of the alkylated guanine influences drug binding. In contrast, in the pluramycin and related complexes it is always the base on the 5'-side of the alkylated guanine that influences the sequence specificity, consistent with the observed intercalation site.

Since trioxacarcin A does not react with a guanine positioned at the 3'-terminus of a DNA-oligonucleotide (Fitzner *et al.*, 2008), it appears that docking of the antibiotic, guided by the hydrogen bonds discussed above and the location of the bulky sugars in the minor and major grooves, with preliminary non-covalent binding to the DNA-helix is a prerequisite for the reaction of the trioxacarcin A with a guanine base. The detailed structure of this non-covalent complex before guanine attack is not yet known. It has to be taken into account that trioxacarcin A would have to thread into the DNA backbone, either by transient melting or by unwinding of the helix (Williams *et al.*, 1990; Collier *et al.*, 1984).

The flipping-out of bases plays an important role in certain DNA-protein interactions; for example in the DNA repair enzyme uracil-DNA glycosylase (UDG) the discrimination between uracil and thymine is initiated by thermally induced opening of TA and UA base-pairs (Parker *et al.*, 2007). Base-pair dynamics may also be connected with the sequence selectivity of trioxacarcin A. An unresolved question is how streptomyces protect their own DNA from the drug.

3. Trioxacarcin bound to d(CGTACG) – Guanine binding antibiotic induces formation of a DNA hexaplex

3.1 INTRODUCTION

The structure of trioxacarcin bound to d(CGTACG) is a more specific case because the oligonucleotide is affected by the combined interactions of two trioxacarcins. In the structure discussed in the previous chapter the two trioxacarcins bound to the oligonucleotide were not influencing each other.

3.2 MATERIALS AND METHODS

Trioxacarcin A was obtained from marine *Streptomyces* sp. isolate B8652 by fermentation (Maskey *et al.*, 2002). Oligonucleotides were purchased already purified by HPLC from Carl Roth GmbH and used without further purification.

Crystallization
The DNA-drug solution used for crystallization contained 2.5 mM d(CGTACG) (single-strand concentration), 2.8 mM trioxacarcin A and 25% v/v methanol. It was prepared by mixing trioxacarcin A stock solution (containing 5.6 mM trioxacarcin A and 50% v/v methanol) with a 5.0 mM DNA solution in a 1 to 1 ratio at room temperature and incubation for three days at 4 °C.

The initial crystallization condition containing 30% v/v PEG 400, 0.1 M sodium acetate pH 4.5, and 0.2 M calcium acetate was found with a commercial sparse matrix screen at 20 °C (Emerald BioSystems, Wizard I, condition 44). After optimization of this condition red needles with a length of 100 μm were obtained (Figure 3.1), which diffracted to about 3.2 Å resolution at beam line X12 (DESY). Since trioxacarcin A is yellow in powder form, the red color of the crystals was not expected. The crystallization drop was green fluorescent at the beginning and changed to red after one day; the red crystals appeared after three days. Crystallization at 40 °C led to an improvement in crystal quality, the resolution limit was increased to 2.1 Å at the same beam line. The reproducibility was not very good at 40 °C, which could be improved by addition of L-cysteine to the drop and a volatile alcohol (e.g. ethanol, *n*-propanol, 1,3-propanediol) to the reservoir. These substances were found with a commercial additive screen (Hampton Research, additive screen with 96 conditions). Table 3.1 gives detailed information for the crystallization of the crystals used for structure solution and refinement. It has to be added that the sodium acetate buffer is certainly affected by the amount of calcium acetate in the crystallization drop. The measured pH for the crystallization condition Native-SLS is 5.5 instead of 5.0.

For data collection at 100 K the crystals were flash cooled in liquid nitrogen. In most cases the mother liquor around the crystals was sufficient for cryoprotection, only for long collection times greater than a day glycerol had to be added (U-derivative in Table 3.1).

Figure 3.1. Red needle-shaped crystals with a length of 100 μm of the trioxacarcin-d(CGTACG) complex.

Table 3.1. Crystallization and soaking conditions

	Native-SLS	Native-DESY	Co-MAD	U-derivative
Method	Hanging-drop	Hanging-drop	Hanging-drop, cocrystallization	Hanging-drop, UO$_2$-soaking
Content of reservoir solution	23% v/v PEG 400, 0.18 M calcium acetate, 0.09 M sodium acetate pH 5.0, 3% v/v ethanol	23% v/v PEG 400, 0.18 M calcium acetate, 0.09 M sodium acetate pH 5.0, 4% v/v 1,3-propanediol	30% v/v PEG 400, 0.40 M calcium acetate, 0.1 M sodium acetate pH 5.0	23% v/v PEG 400, 0.18 M calcium acetate, 0.09 M sodium acetate pH 5.0, 4% v/v *n*-propanol
Reservoir volume, μl	500	500	500	500
Temperature, °C	40	40	20	40
Additives for drop	0.1 M L-cysteine	0.1 M L-cysteine	0.1 M cobalt chloride	0.1 M L-cysteine
Drop solution, μl DNA/additive/reservoir	2.0/0.2/1.0	2.0/0.2/1.0	3.0/0.6/3.0	2.0/0.2/1.0
Content of drop at stage of setting up	1.6 mM DNA, 1.8 mM trioxacarcin A, 7% v/v PEG 400, 0.06 M calcium acetate, 0.03 M sodium acetate pH 5.0, 0.9% v/v ethanol, 6 mM L-cysteine	1.6 mM DNA, 1.8 mM trioxacarcin A, 7% v/v PEG 400, 0.06 M calcium acetate, 0.03 M sodium acetate pH 5.0, 1.3% v/v 1,3-propanediol, 6 mM L-cysteine	1.1 mM DNA, 1.3 mM trioxacarcin A, 14% v/v PEG 400, 0.18 M calcium acetate, 0.05 M sodium acetate pH 5.0, 17 mM cobalt chloride	1.6 mM DNA, 1.8 mM trioxacarcin A, 7% v/v PEG 400, 0.06 M calcium acetate, 0.03 M sodium acetate pH 5.0, 1.3% v/v *n*-propanol, 6 mM L-cysteine
Soaking solution				0.2 M uranyl nitrate, 25% v/v PEG 400, 0.2 M calcium acetate, 0.1 M sodium acetate pH 5.0
Back-soaking solution				10% v/v glycerol 25% v/v PEG 400, 0.2 M calcium acetate, 0.1 M sodium acetate pH 5.0
Cryo solution	mother liquor	mother liquor	mother liquor	Back-soaking solution

Structure solution

The U-derivative dataset (Table 3.2) was integrated by SAINT (Bruker AXS, Madison WI, USA), all other datasets were integrated with HKL2000 (Otwinowski & Minor, 1997), space group determination and dataset preparation was done by XPREP (Bruker AXS, Madison WI, USA).

Table 3.2. Data collection, phasing and refinement statistics

	Native-SLS	Native-DESY	Co-MAD		U-derivative
Crystal data					
Space group	$P6_322$	$P6_322$	$P6_322$		$P6_322$
a, Å	49.71	49.29	49.62		49.77
c, Å	39.86	39.83	40.37		39.74
Beamline	PXII (SLS)	X12 (DESY)	14.1 (BESSY)		Home-source
Collection date	25.02.07	01.11.06	04.11.06		14.–16.11.06
			Peak	Inflection	
Diffraction data					
Wavelength, Å	0.99987	0.89818	1.60518	1.60599	1.54178
Resolution limit, Å	1.78	2.10	2.80	2.80	2.77
Total reflections	60423	34486	8883	9093	11168
Unique reflections	3091	1904	849	848	862
Completeness, %					
Overall	99.6	99.6	97.8	97.8	98.9
Outermost resolution shell	97.6	98.3	98.7	100.0	90.5
Redundancy					
Overall	19.48	18.11	10.23	10.49	12.81
Outermost resolution shell	10.39	7.53	9.55	10.73	3.93
$I/\sigma(I)$					
Overall	25.15	20.94	18.76	18.49	12.51
Outermost resolution shell	4.08	3.63	2.29	2.11	2.38
R_{int}					
Overall	6.19	7.72	6.19	6.63	13.72
Outermost resolution shell	39.55	41.56	44.31	54.09	39.54
R_{sigma}					
Overall	2.41	2.62	3.29	3.25	7.30
Outermost resolution shell	13.43	28.96	64.32	60.56	46.76
Phasing					
Resolution, Å			3.2		3.1
pseudo-free CC after dm			0.65		0.63
Refinement					
Reflections used	2476				
Resolution, Å	1.89				
R_{work}	19.86				
R_{free}	24.44				
rms deviation					
Bond length, Å	0.014				
Bond angles, °	2.6				
Average B factor, Å² (all atoms)	38.87				

For the first attempt to solve the structure a native dataset to 2.1 Å and a cobalt-MAD dataset to 2.8 Å resolution were used (Table 3.2). It is known that cobalt(II) ions bind to terminal guanines of oligonucleotides. With a resolution cutoff at 3.2 Å SHELXD (Schneider & Sheldrick, 2002) found five possible cobalt sites, of which only the first one turned out to be a true site. SHELXE (Sheldrick, 2002) produced the best map (Figure 3.2A) with a phase extension to 1.2 Å and 100 cycles of density modification. It was not possible to trace this map manually because of a weighted mean phase error of 71° and the fact that it was not possible to locate trioxacarcin. Probably, the diffraction power of the crystal containing the cobalt derivative was not high enough. These crystals were grown at 20 °C, it was not possible to grow crystals containing cobalt at 40 °C.
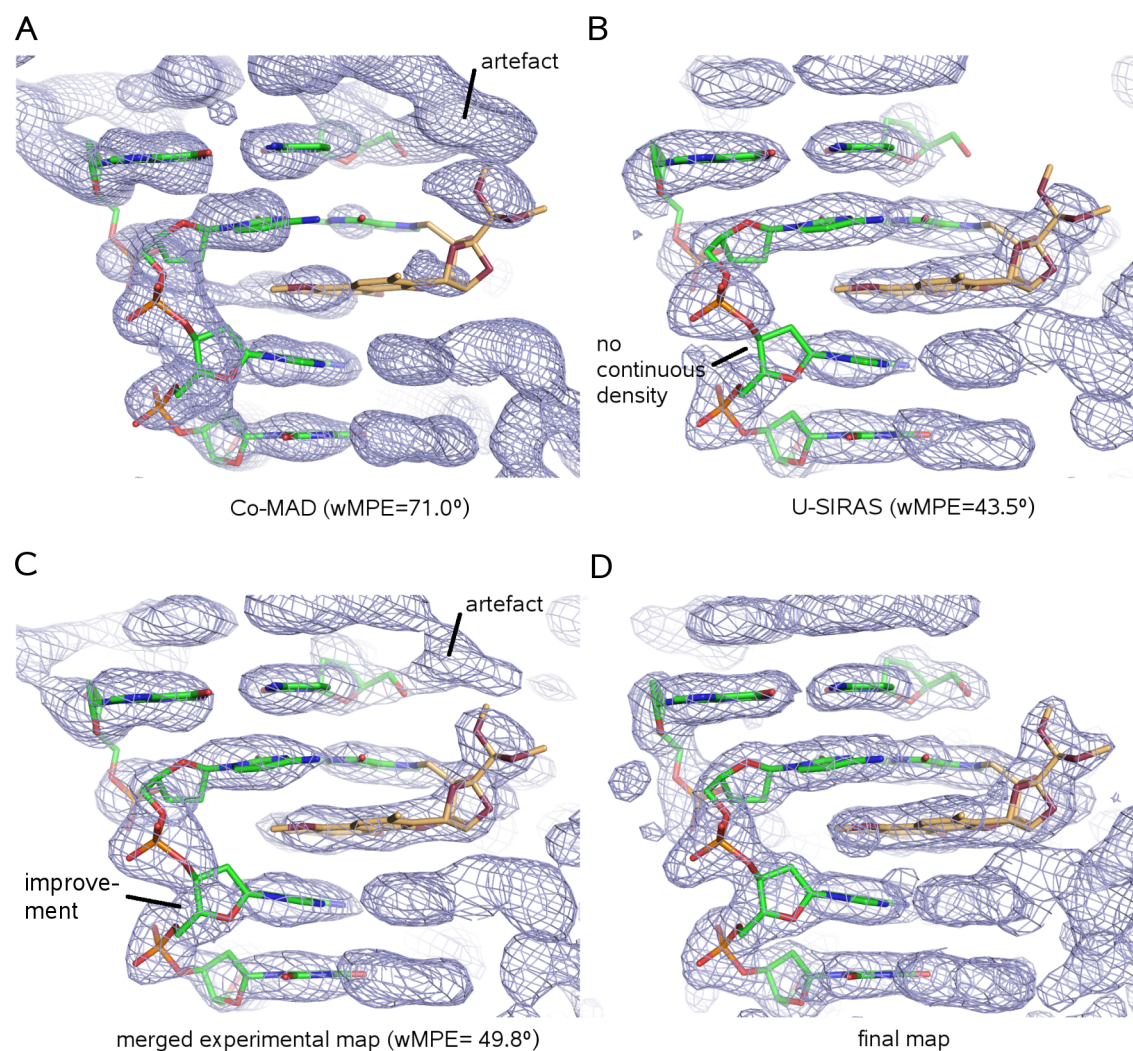

Figure 3.2. Experimental map from cobalt-MAD (**A**), experimental map from uranium-SIRAS (**B**), merged experimental map (**C**) and map after final refinement (**D**), all contoured at a 1 σ level. The carbon atoms of the trioxacarcin are shown in light orange. The measurement artefact in (**A**) is still visible in (**C**).

The second trial was performed with a crystal soaked for 10 s in a 0.2 M uranyl-nitrate solution, back-soaked in cryosolution, and measured on a home-source with a Smart 6000 detector (see Table 3.1 for details of the soaking solutions and Table 3.2 for data

statistics). The uranyl-derivative dataset was combined with the native one from DESY to create a SIRAS dataset. A single uranium site on the 3-fold axis (0.667, 0.333, 0.066) was found with a resolution cutoff of 3.1 Å with SHELXD. The electron density map (Figure 3.2B) produced by SHELXE contained lower phase errors (wMPE = 43.5°) than the previous one and showed density matching the drug.

The experimental maps from the Co-MAD and U-SIRAS experiments were merged together in reciprocal space, which produced an interpretable map (Figure 3.2C). This showed that both maps contained correct information since they were obtained with different experiments. Although the merged map still contained some errors from the weak Co-MAD data (close to the upper base pair in Figure 3.2A and 3.2C is a misleading aretefact of electron density) it also shows an improvement in some regions compared to the U-SIRAS map in Figure 3.2B (the electron density of the backbone from the three lower bases on the left is more continuous in the merged map). The weighted mean phase errors of all experimental maps were calculated with the final refined native model.

Model building and refinement

The manual model building was difficult because at the beginning only the alkylated guanine base and the drug could be assigned to the experimental density without doubt. The strand direction could be worked out because the distance between sugar and phosphate group is usually a bit shorter on the 3'-side of the sugar.

The structure was refined isotropically with TLS constraints against the native dataset from SLS with REFMAC (Murshudov *et al.*, 1997) to R_{work} = 19.9% and R_{free} = 24.4%. The final map is visible in Figure 3.2D and the refinement statistics are listed in Table 3.2.

3.3 RESULTS AND DISCUSSION

Overall structure

Trioxacarcin binds covalently to d(CGTACG) by alkylating the N7 position of G(2) and intercalates at the 3'-side of the guanine (Figure 3.3). Two of the self-complementary strands form a duplex which is lying on a 2-fold axis; the two strands are therefore crystallographically related and have identical conformations.

Interestingly, there are no continuous stacking interactions within the duplex. The stacking is interrupted between trioxacarcin(7) and T(3), resulting in a kink in the DNA duplex. The 2-fold axis is running through the kink, with two calcium ions positioned very close to it (Figure 3.4A). The interruption of stacking results in two columns of bases running in different directions. This is indicated in Figure 3.4B by two blue arrows. The kink is transforming a part of the minor groove into a cavity.

The main part of the duplex exhibits Z-geometry with Watson-Crick base-pairing [base pairs C(1)-G(6#1) and G(2)-C(5#1)], the part in the kink region involving residues T(3) and A(4) shows B-geometry. In Figure 3.4B the change between the two backbone conformations is visible, the B- and Z-form is indicated with different colors for the residues. The arrangement of bases in the duplex can be understood best by comparing Figure 3.4B with Figure 3.3, where the residues are in the same orientation. In Table 3.3, the sugar-phosphate and glycosyl torsion angles are compared with the observed ranges for B- and Z-DNA (Schneider *et al.*, 1997).
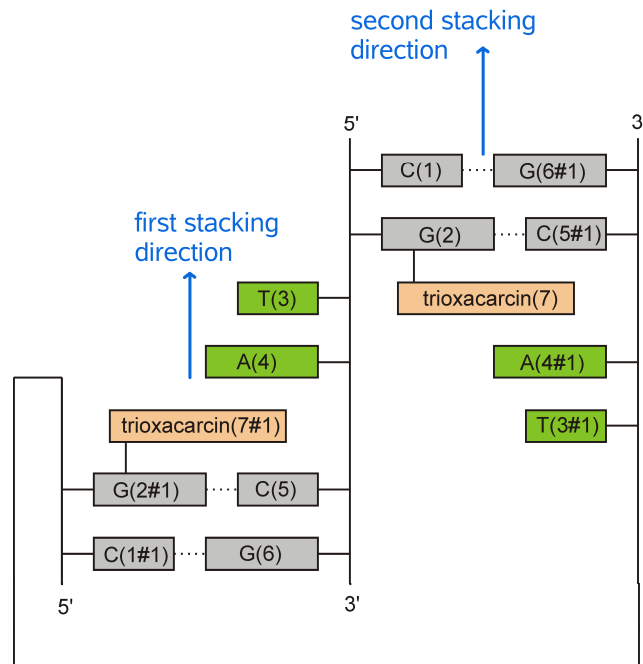
Figure 3.3. Simplified view of the DNA-trioxacarcin duplex with residue names. The interruption of base stacking inside the duplex is illustrated by two separated columns. The two stacking directions are given by blue arrows. Hydrogen bonding is indicated by dotted lines. The residues that do not form base pairs within the duplex are drawn in green.
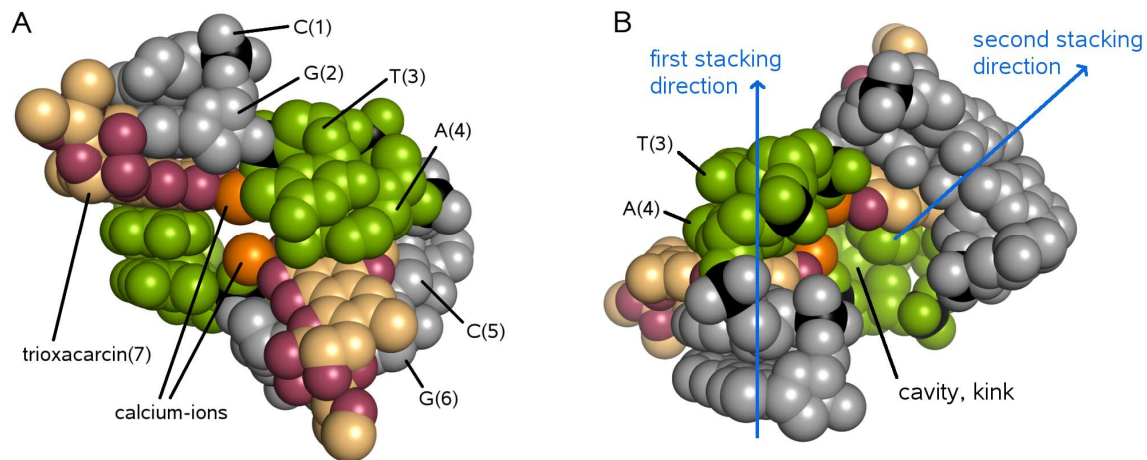
Figure 3.4. Space-filling views of the DNA-trioxacarcin duplex. Trioxacarcin is shown in red (oxygen atoms) and light orange (carbon atoms); the DNA is shown in light gray (Z-DNA geometry) and green (B-DNA geometry, bases not pairing within the duplex). The colors for bases correspond to the ones in Figure 3.3. In (**A**) the interrupted stacking between residue G(2) and T(3) is visible. (**B**) shows the cavity to which the minor groove is transformed. Also the change from right to left handed helix geometry is visible by looking at the phosphor atoms shown in black. The two columns of stacked bases running in different directions are indicated by blue arrows.

Table 3.3. Backbone and glycosyl torsion angles

Residue	Angle, degrees						
	α	β	γ	δ	ε	ζ	χ
C1	-	-	168	146	276	86	206
G2	68	180	169	91	233	164	68
T3	305	190	40	144	186	282	247
A4	280	171	45	110	231	297	69
C5	178	192	47	142	257	96	204
G6	65	190	176	90	-	-	65
B-DNA	270–330	130–200	20–80	70–180	160–270	150–300	200–300
Z-DNA (A/G)	40–100	150–250	160–210	80–160	180–300	280–340 40–100	50–90
(T/C)	150–250	150–250	20–90	80–160	180–300	40–100	180–220

Torsion angles are defined as α: O3'-P-O5'-C5', β: P-O5'-C5'-C4', γ: O5'-C5'-C4'-C3', δ: C5'-C4'-C3'-O3', ε: C4'-C3'-O3'-P, ζ: C3'-O3'-P-O5', χ(Purines): O4'-C1'-N9-C4, χ(Pyrimidines): O4'-C1'-N1-C2. In the case of Z-DNA values for purines (A/G) and pyrimidines (T/C) have to be distinguished.

Obviously, base pairs T(3)-A(4#1) and A(4)-T(3#1) (#1 indicates a symmetry related residue belonging to the same duplex) have been ripped apart because of the presence of the two trioxacarcins in the complex, which are located close to each other. T(3) and A(4) are now forming two Hoogsteen base pairs with each other's symmetry equivalents belonging to a different duplex. A third duplex is binding to the four remaining bases. This leads to the formation of a hexaplex including three duplexes (Figure 3.5) which is also stabilized by stacking interactions. The hexaplex has point symmetry 32 and lies on the positions (1/3, 2/3, 3/4) and (2/3, 1/3, 1/4) in the unit cell. It has an unusual shape with a channel along its 3-fold axis and three cavities on its sides (Figure 3.6A). The terminal base pairs of the duplexes are not involved in inter-duplex contacts, only the adenines and thymines are forming contacts. Two halves of two different duplexes have a continuous stacking within the hexaplex, indicated by a blue arrow in Figure 3.6B. The complicated three-dimensional arrangement of bases in the hexaplex can be understood best by comparing Figure 3.6B with Figure 3.5, where all residues are shown in corresponding colors.

Rings A, B and C of the drug are involved into stacking interactions with the DNA. The two sugars of trioxacarcin A are not present in the crystal structure. Mass spectrometry performed with dissolved crystals showed that they have been cleaved off, probably after the alkylation of guanine.
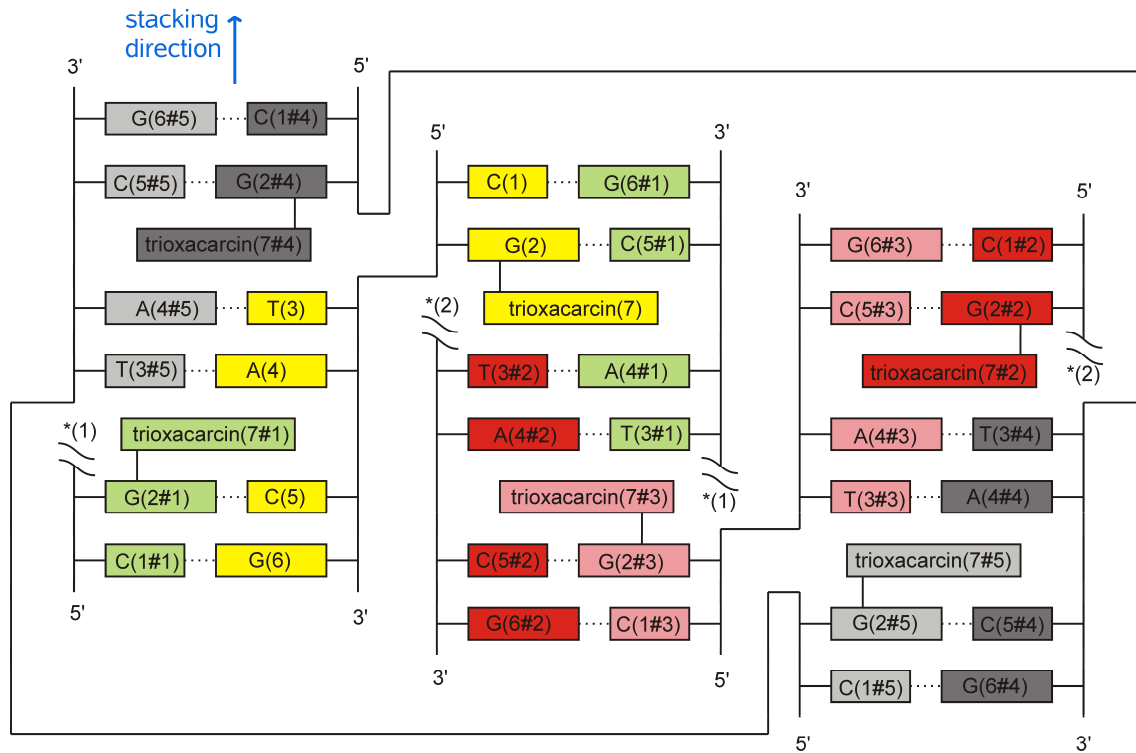
Figure 3.5. Simplified view of the DNA hexaplex formed by three duplexes. Duplex1 is shown in yellow/green, duplex2 in red/pink and duplex3 in light/dark gray. The # sign indicates symmetry related residues, G(2#1) means "symmetry mate number 1 of guanine residue 2". One of the three directions of base stacking is indicated by a blue arrow.
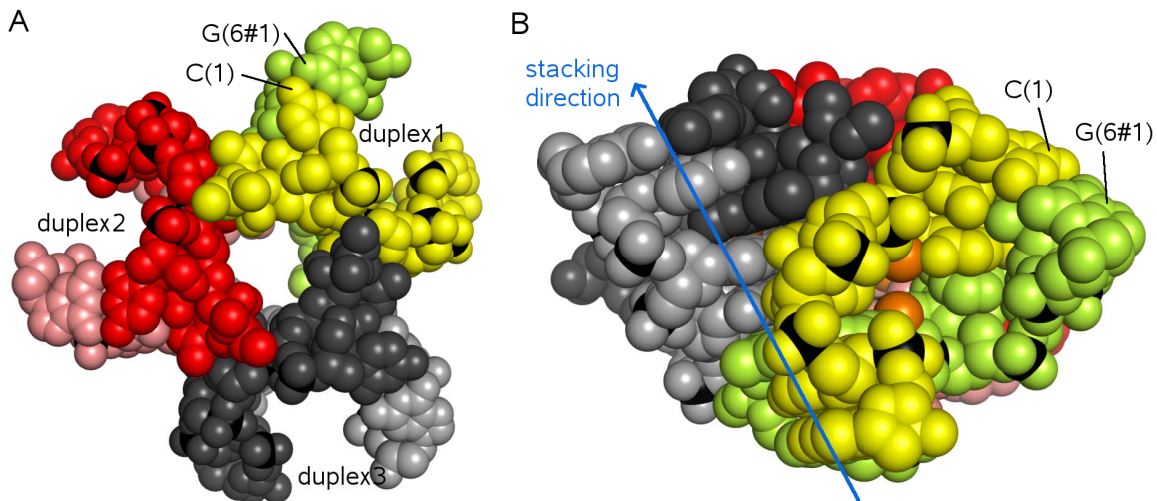
Figure 3.6. Space-filling view of the DNA-trioxacarcin hexaplex along the 3-fold axis (**A**) and perpendicular to it (**B**). One duplex is shown in yellow/green, the second in red/pink and the third in light/dark gray. These colors correspond to Figure 3.5. The phosphorus atoms are shown in black, the calcium ions in orange. One of the three directions of base stacking is indicated by a blue arrow, also corresponding to Figure 3.5.

Antibiotic-DNA interactions

Beside the covalent bond, there is an interaction between trioxacarcin and the oligo-nucleotide mediated by a calcium ion involving the 1- and 2-oxygens from trioxacarcin as well as O2P from residue T(3) (Figure 3.7). The symmetry-related calcium ion is also involved in this complex. The two calcium ions are bridged by two water molecules. These water molecules are forming hydrogen bonds to N1 of residue A(4) and to its symmetry equivalent A(4#1). The calcium-complex is positioned in the center of the DNA-duplex and fills the gap from the flipped-out residues T(3) and A(4).
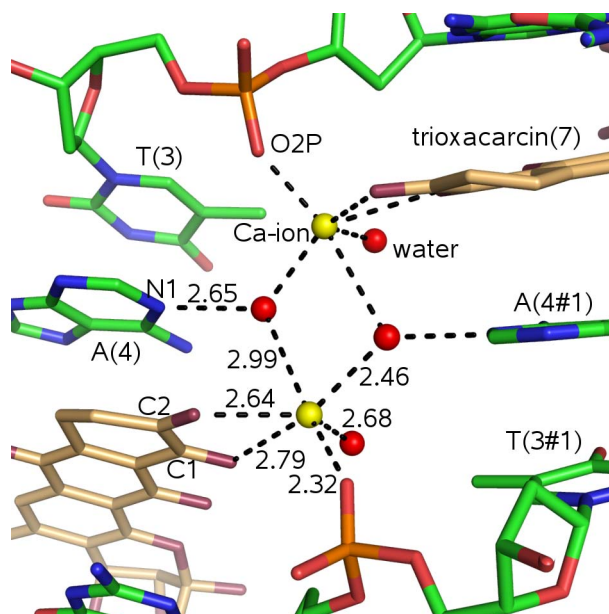

Figure 3.7. View of the calcium sites in the trioxacarcin-DNA complex. The bond lengths are given in Å. The calcium ions are shown as yellow spheres, the water molecules as red spheres.

Because of the lack of the sugar moieties the drug does not have any direct hydrogen bonds with the DNA.

The chromophoric ring system of trioxacarcin is involved in stacking interactions with G(2) and A(4#1) and slightly with C(5#1). There is also strong interaction with T(3#2), which is belonging to a different duplex and is positioned under ring C of trioxacarcin (Figure 3.8). This interaction stabilizes the hexaplex shown in Figure 3.5 and 3.6. In contrast to the previous trioxacarcin-d(AACCGGTT) complex, the long axis of the trioxacarcin aglycon is perpendicular to base pair G(2)-A(4#1) and nearly parallel to base-pair A(4#1)-T(3#2).
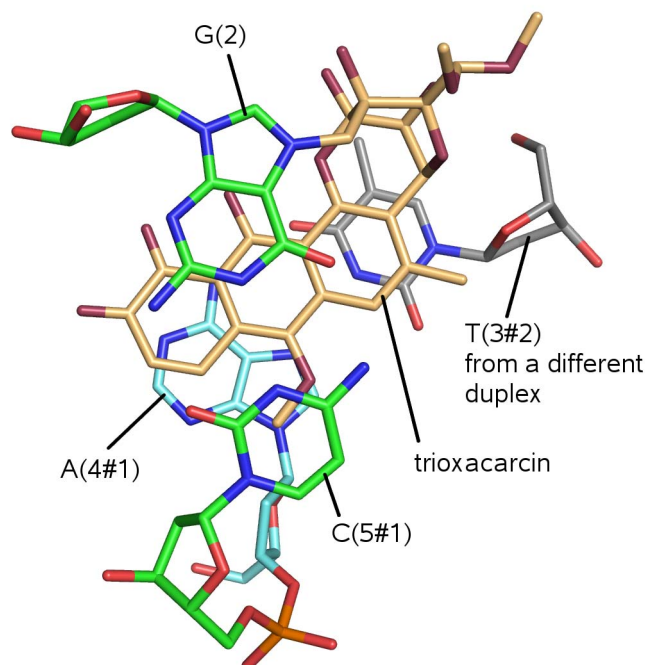
Figure 3.8. View down the DNA helix axis. The carbon atoms of the upper base-pair G(2)-C(5#1) are drawn in green, the ones of the lower base-pair A(4#1)-(3#2) in turquoise and gray. The trioxacarcin is positioned between the two base-pairs, with its carbon atoms colored in light orange.

Sugar cleavage

The red color of the crystals indicates an absorption shift in the chromophoric ring system of the antibiotic due to a chemical reaction. Trioxacarcin A contains two sugars. Those could not be found in the electron density maps of the drug-DNA complex, whereas every other part of the structure could be modeled.

DNA-drug stock solution of green fluorescent color and dissolved red crystals were analyzed by *electrospray ionization mass spectrometry* (Fenn *et al.*, 1989). Both *negative ion electrospray ionization* [ESI(–)] and *positive ion electrospray ionization* [ESI(+)] were used. With the green fluorescent stock solution a molecular weight of 2667 g/mol was determined with ESI(–), with the dissolved red crystals 2328 g/mol with ESI(–) and ESI(+). ESI(+) was done in the second case in order to confirm the previous result, since the amount of sample was very small and close to the detection limit of the spectrometer. The difference of 339 g/mol indicates a cleavage of both trioxacarcin sugars. For all cases the values were derived by deconvolution, which means all the differently charged species that were detected were calculated for a state of neutral charge.

The calculated molecular weight of trioxacarcin A ($C_{42}H_{52}O_{20}$) is 876 g/mol, the one of a $[d(CGTACG)]^{5-}$ anion ($C_{58}H_{69}O_{34}N_{23}P_5$) is 1786 g/mol and the one for a $[trioxacarcin-d(CGTACG)]^{4-}$ anion ($C_{100}H_{122}O_{54}N_{23}P_5$) is 2663 g/mol (only main isotopes were used for this calculation). The difference of 4 g/mol between the value derived by deconvolution (2667 g/mol) and the calculated mass can be explained with the fourfold negative charge of the drug-DNA complex. The oligonucleotide has a charge of –5, but after the reaction with trioxacarcin A an additional positive charge is formed and an additional hydrogen atom is bound (see reaction mechanism in Figure 2.2).

The observed color change in the crystallization drop together with the results from mass spectrometry suggest that the sugars were still attached to the covalent drug-DNA complex in the stock solution but cleaved from it due to the conditions in the crystallization drop. This might have made crystallization easier, since the deglycosylated compound is probably less soluble. In order to find out whether trioxacarcin loses its sugars because of the attachment to DNA, pure trioxacarcin A was added to the same crystallization condition without DNA at 20 °C. With DNA the drop color changed from green to red after 24 h, without DNA after 2 h. This indicates that the sugars of trioxacarcin are even a bit more protected when the antibiotic is in complex with DNA.

Unfortunately, the chemical nature of the deglycosylated compound could not derived so far. Figure 3.9 shows the difference density at the two deglycosylated positions of the drug. For a hydrolysis one would expect to find an OH-group at the aglycon, but the 4-oxygen is not visible in the difference density (Figure 3.9A).
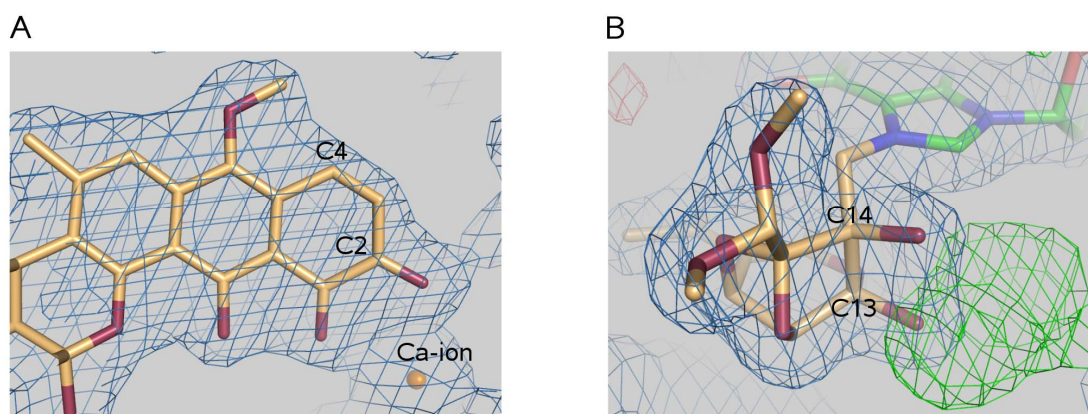

Figure 3.9. Difference density at the deglycosylated sites at C4 (**A**) and C13 (**B**). The blue mesh represents the electron density after final refinement, contoured at a 1 σ level. The green mesh represents positive difference density at a 3 σ level (indicating that electrons are missing in the model), the red mesh represents negative difference density at a –3 σ level (indicating that the model contains too many electrons, fairly visible here).

It was not possible to interpret the difference density close to C13 completely (Figure 3.9B). The 13- and 14-oxygens seem to be still present, so for refinement the 13-sugar was replaced by an OH-group. During refinement it turned out that the chiral volume restraint on C2 is not appropriate, so this carbon seems to be sp^2-hybridized. Therefore the formation of a 1,2-anthraquinone derivative was assumed (Figure 3.10), which is also in correspondence with the red color of the crystals. The calculated difference between the original complex and the assumed model is 394 g/mol, according to the mass spectra this difference should be 55 g/mol more. This means that there are still three or four non-hydrogen atoms missing in the model, which fits approximately to the volume of positive difference density around C13.
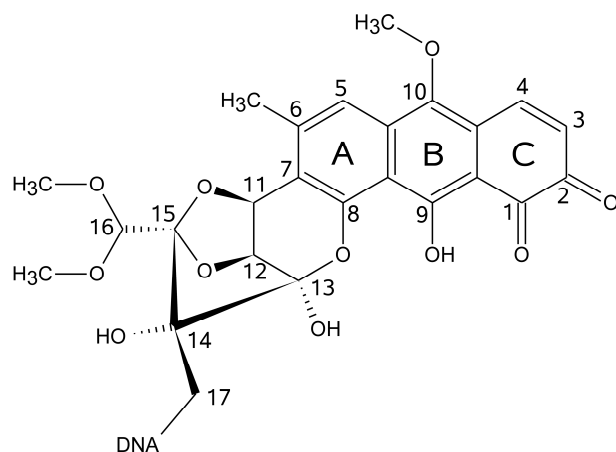
Figure 3.10. Schematic diagram of the assumed deglycosylated model used for refinement.

In order to see if there would be space left for the trioxacarcin sugars in the hexaplex, the model of gutingimycin with sugars was superimposed on the deglycosylated compound (Figure 3.11).
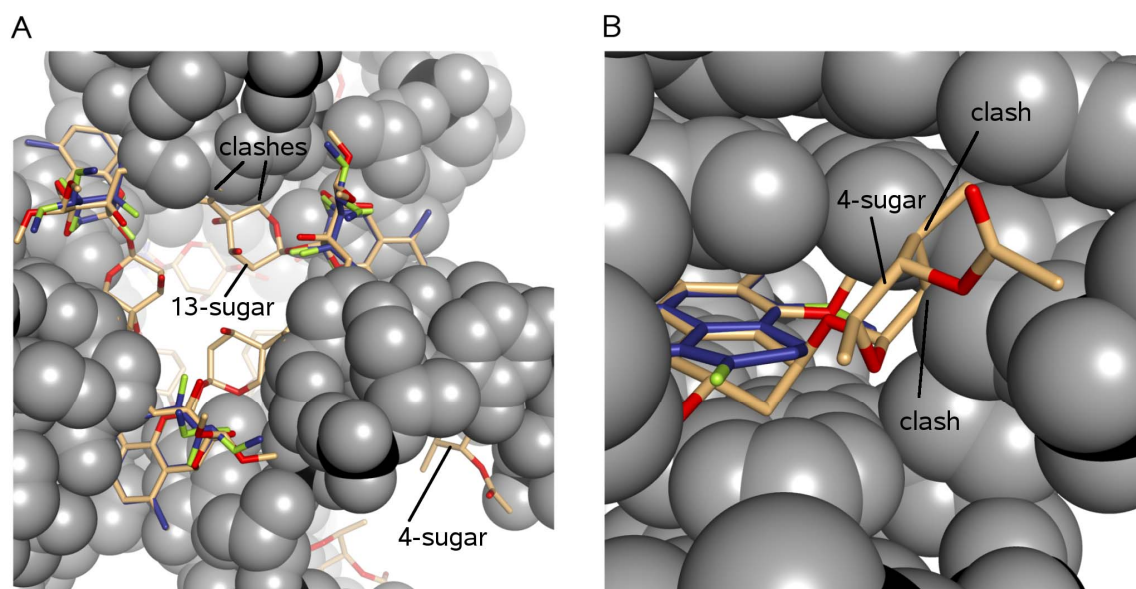
Figure 3.11. Superposition of the deglycosylated compound and gutingimycin. Gutingimycin is shown in light orange (carbon atoms) and red (oxygen atoms); the deglycosylated trioxacarcin bound to DNA is shown in blue (carbon atoms) and yellow (oxygen atoms). The clashes between the sugars of the drug and DNA are indicated. In (**A**) the 13-sugar is visible in the central hole of the hexaplex close to the three fold axis, in (**B**) the 4-sugar in the minor groove.

According to this superposition, the 13-sugar would sit in the channel of the hexaplex (Figure 3.11A), together with five symmetry equivalents. The DNA would clash with C6" and the 4"-acethyl group (for nomenclature of trioxacarcin A see Figure 2.1). The 4-sugar would lie in the minor groove of the duplex (Figure 3.11B), in this case 3'-OH and C2' would clash with residue C(5#1). It has to be taken into account that the trioxacarcin sugars are able to rotate around two bonds, so they could probably avoid the clashes. With the results from the superposition the hypotheses can be formed that the original

complex containing the full trioxacarcin has a similar geometry to the observed one and that the observed hexaplex is also present in solution. If the sugars were cleaved from a duplex in solution, the formation of a channel inside the hexaplex big enough to contain six sugars (Figure 3.6A) would be unlikely.

In the structure of the uranyl-derivative two uranyl ions are positioned inside the channel of the hexaplex on the three-fold axis. Thus, the cleavage of the 13-sugar helped to solve the crystal structure.

Crystal packing

The crystals contain linear solvent channels of different size and shape along the *c* axis (Figure 3.12A). The biggest of these channels is positioned along the 6-fold screw axis and has a star shaped profile. Another channel that is formed by the holes in the hexaplex (Figure 3.6A) is running along the 3-fold axis and enabled the successful incorporation of uranyl ions for structure solution already mentioned above.

From the view perpendicular to the *c* axis (Figure 3.12B) the continuous stacking interactions inside the crystal becomes visible. The crystal is held together by a stacking interaction between base pairs C(1)-G(6#1) and G(6#8)-C(1#9), which is the only contact between the DNA-hexaplexes. Each hexaplex is in contact with six others.
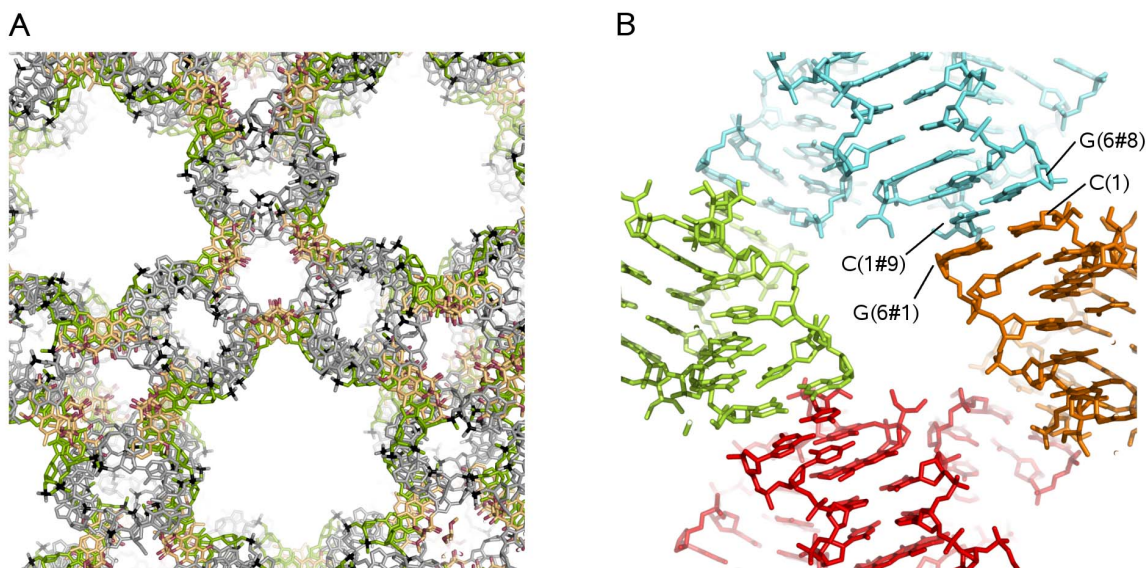
Figure 3.12. Crystal packing. (**A**) View along the *c* axis. Trioxacarcin is shown in red (oxygen atoms) and light orange (carbon atoms); the DNA is shown in light gray [residues C(1), G(2), C(5), G(6)] and green [residues T(3) and A(4)] except for the phosphorus atoms that are shown in black. (**B**) View perpendicular to the *c* axis. Each hexamer is shown in a different color.

Comparison with the previous structure and gutingimycin

As already mentioned, the main difference between the d(AACCGG*TT)- and the d(CG*TACG)-duplex (the alkylated guanine is G*) is the fact that in the first one the two bound drugs can be regarded as independent from each other whereas in the second one they cannot. The reason for this not only lies in the different distances between the two trioxacarcins in the two structures. The main reason is that the disturbance of the DNA

caused by the drug is more prominent on the 3'-side of the alkylated guanine than on the 5'-side (on the 3'-side a thymine is flipped out in both structures). In the d(AACCGG*TT)-duplex these disturbances are directed to the termini whereas in the d(CG*TACG)-duplex they are both directed to the centre and combining each other.

   Another big difference in the two structures is the orientation of the alkylated guanine towards the drug (Figure 3.13). Figure 3.13A shows trioxacarcin bound to guanine in the d(AACCGGTT)-duplex, Figure 3.13B shows the same residues in the d(CGTACG)-duplex. N2 of this guanine lies above ring C in both structures. In the d(AACCGGTT)-duplex N3 is positioned above ring B and N9 above ring A, whereas in the d(CGTACG)-duplex N1 is above ring B and O6 above ring A. Figure 3.13C shows the small molecule crystal structure of gutingimycin crystallized without DNA (Maskey *et al.*, 2004b). In order to compare the different structures the chromophoric ring system of the trioxacarcin has been brought in a similar orientation in Figure 3.13. It is obvious that in gutingimycin and the trioxacarcin-d(CGTACG) complex the guanines have the same orientation towards the drug, whereas in the trioxacarcin-d(AACCGGTT) complex the guanine base is rotated by about 180º. Before the determination of the trioxacarcin-d(CGTACG) complex it was assumed that the guanine base bound to trioxacarcin would rotate after liberation from the trioxacarcin-d(AACCGGTT) complex to give the conformation observed in the crystal structure of the natural compound gutingimycin (Pfoh *et al.*, 2008). With the second structure discussed in this chapter the question arises if there are maybe two stable forms of the cleavage product gutingimycin with two different orientations of the guanine base. This would give a hint to the biological more relevant DNA binding mode of trioxacarcin A.
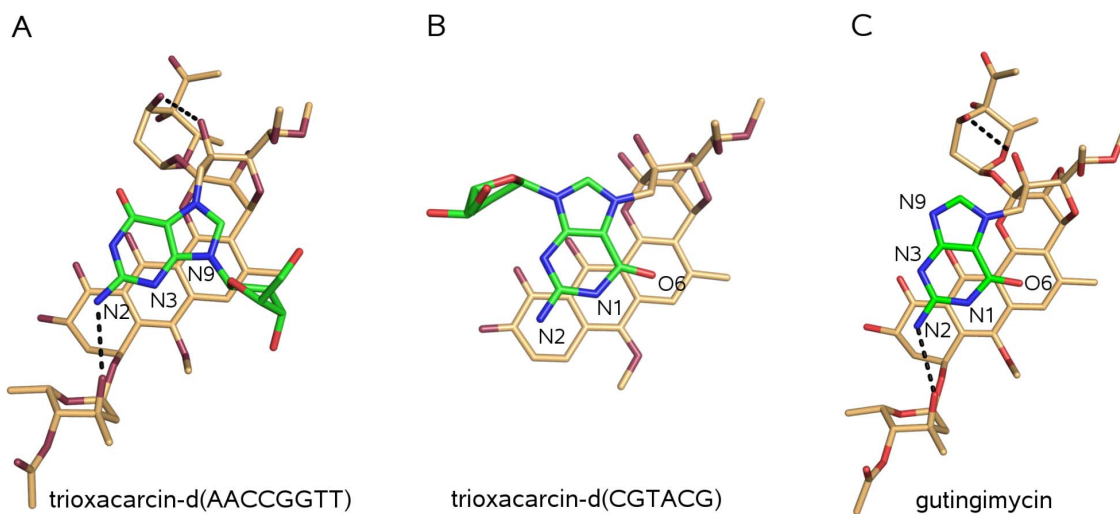
Figure 3.13. Comparison between the orientation of guanine towards trioxacarcin. (**A**) Detail from the trioxacarcin-d(AACCGGTT) complex. (**B**) Detail from the trioxacarcin-d(CGTACG) complex. (**C**) Small molecule crystal structure of gutingimycin crystallized without DNA. Internal hydrogen bonds are shown in dashed lines. The guanine base has the same orientation towards the chromophoric ring system in (**B**) and (**C**), whereas in (**A**) it is rotated about 180°.

   The different orientation of the guanine base in the two DNA-drug complexes is due to the change from B- to Z-DNA, because here the glycosyl torsion angle changes by about 180º. The orientation of the trioxacarcin-sugars is almost identical in gutingimycin

and the trioxacarcin-d(AACCGGTT) complex since they are held in place by the same internal hydrogen bonds (Figure 3.13A and 3.13C).

In spite of the differences mentioned above some similarities can also be found in the drug-DNA complexes. Figure 3.14 compares the intercalation sites of the trioxacarcins in both complexes in a simplified manner and shows also the extracted general motive that holds for both structures. In both cases trioxacarcin binds to a guanine followed by a thymine and intercalates on the 3'-side of the guanine. The thymine is flipped out in contrast to the adenine originally belonging to it, which is in stacking interactions with the drug. This motif has not been reported for any other DNA intercalator yet.

It seems to be the case that the flipping-out of thymine combined with the immobility of adenine leads to some instability in the DNA because a proper base pairing on the 3'-side of the alkylated guanine is not possible anymore. In the trioxacarcin-d(AACCGGTT) complex this disturbance is compensated by a register shift [see upper left side of Figure 3.14, A(102) pairs with T(8) instead of T(7)], which is possible in an oligonucleotide. In a longer DNA double strand a register shift would transfer the disturbance to a different position but not compensate it. In the trioxacarcin-d(CGTACG) complex the disturbance from the first trioxacarcin is compensated by a second trioxacarcin; and also by interaction with a different duplex (see upper right side of Figure 3.14).
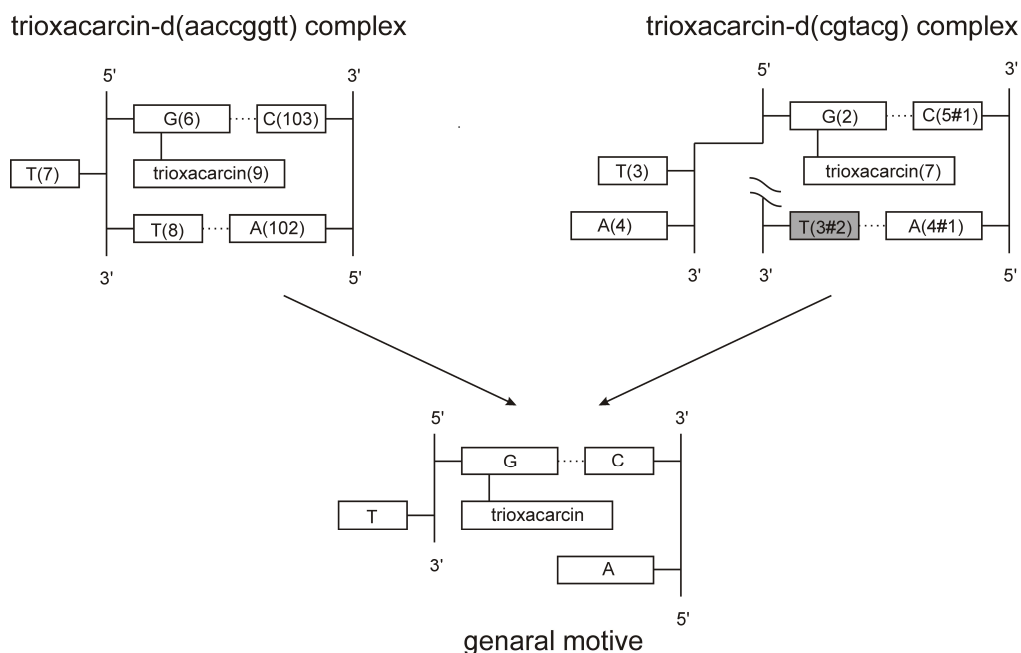
Figure 3.14. Comparison between the intercalation sites in the two different structures. The residue in gray belongs to a different duplex. The general motive holds for both structures.

Biological relevance

Since the duplexes are connected to a hexaplex by *non-terminal* flipped-out bases, this interaction is not restricted to small oligonucleotides but could also happen with an infinite DNA strand. In Figure 3.15 two possible transitions from double stranded DNA to a hexaplex with six trioxacarcin A molecules are suggested. The proposed formations

require repeats of a short sequence like d(CGTACG) with a small number of bases between the repeated unit. For the proposal A in Figure 3.15 the sequence has to be repeated three times, for proposal B six times. So-called tandem repeats with a variable number of repeat units can be observed in the genome of vertebrates (Jeffreys *et al.*, 1985). The oligonucleotide with sequence d(GCGGGAGC) is a shortened part of a repeat unit found in the human genome. Kondo *et al.* (2004, 2006) found by X-ray analyses that it adopts a base-intercalated (zipper-like) duplex, which can form multiplexes. A quadruplex (PDB ID 1v3n), a hexaplex (PDB ID 2fza) and an octaplex (PDB ID 1v3p) have been observed so far.
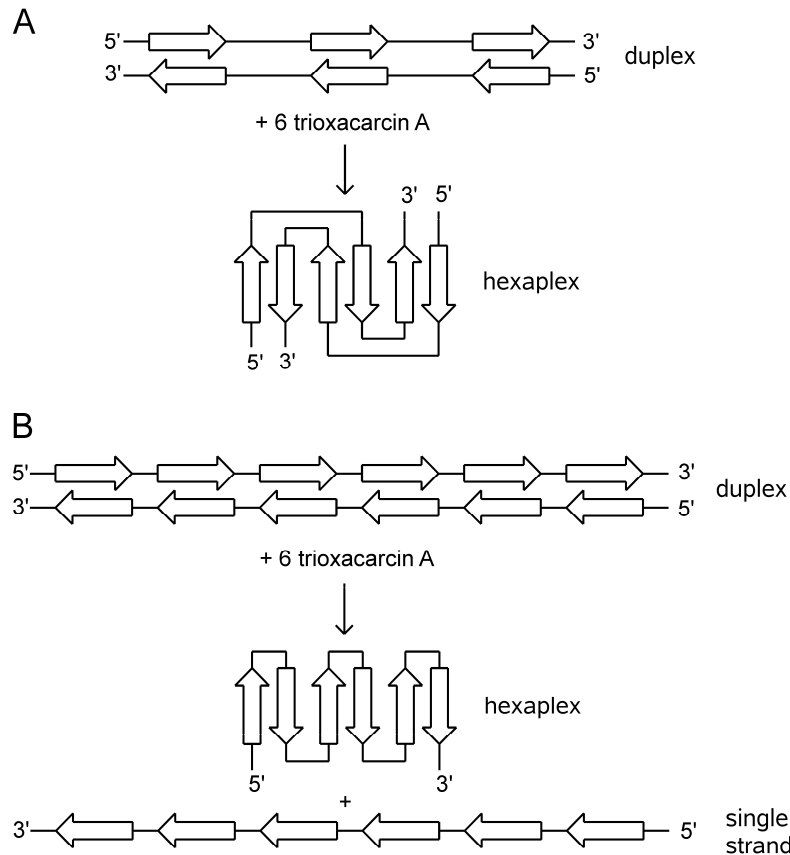
Figure 3.15. Schematic diagram of two possible formations of a hexaplex from double stranded DNA and six trioxacarcin A molecules *in vivo*. The arrows represent small repetitive units. Corresponding to the observed hexaplex in this work, an arrow would stand for d(CGTACG). In (**A**) the hexaplex consists of two DNA strands, in (**B**) it is formed only with one strand.

Two different crystal structures with trioxacarcin bound to DNA have been described. They show large differences but also have some common features. A difficult question is if they are representative of a drug binding *in vivo* and which of the two structures is closer to reality. Compared to an oligonucleotide the DNA in a living cell is very different: it is much longer, it is interacting with proteins, small molecules and ions at the same time. In some parts its double strands are separated to form junctions and loops, in other parts it gets cut and repaired. A sequence selectivity study or a crystal structure determined with an oligonucleotide might therefore not show the most preferable binding

site of an antibiotic. For detecting it a footprinting study employing endonucleases and separation by gel electrophoresis would be the appropriate method. This was beyond the scope of this work but could be a future perspective.

4. Interactions of an echinomycin-DNA complex with manganese(II) ions

4.1 INTRODUCTION

Echinomycin is found in streptomyces and belongs to the group of quinoxaline antibiotics that bind to DNA by bisintercalation (Waring & Wakelin, 1974). It consists of two identical depsipeptides containing D-serine, L-N-methyl-valine, L-N-methyl-cysteine and L-alanine with a quinoxaline base attached to D-Serine (Dell *et al.*, 1975). The two peptide strands are connected by two ester linkages and a thioacetal bridge (Figure 4.1).
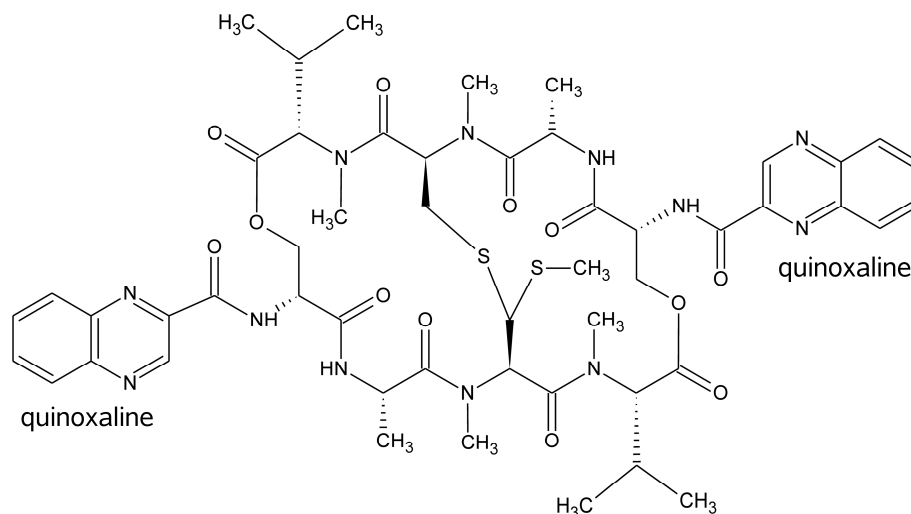
Figure 4.1. Schematic diagram of echinomycin.

Footprinting studies indicated that echinomycin binds mostly around 5'-GC with a preference for AT basepairs at the surrounding sites (Low *et al.*, 1984). At the moment crystal structures of the drug in complex with d(GCGTACGC), d(CGTACG) and d(ACGTACGT) are available. In all of them the two quinoxaline bases are intercalating around the 5'-GC sites and the depsipeptide backbone is positioned in the minor groove. In most of the structures all basepairs next to the 5'-GC site are in the Hoogsteen mode (Cuesta-Seijo & Sheldrick, 2005), only in one structure with d(ACGTACGT) some of these basepairs are in Watson-Crick mode (Cuesta-Seijo *et al.*, 2006). This is consistent with NMR experiments performed with different oligonucleotides in which both base-pairing modes are found for bases flanking the bisintercalation site (Gilbert & Feigon, 1991).

In the crystal structure presented here echinomycin is in complex with d(ACGTACGT) and crystallized in space group $P4_12_12$. There are two known structures with the same sequence, one in $P6_322$ (PDB ID 1xvn) and the other one in $P4_22_12$ containing two duplexes in the asymmetric unit and an unexpected metal bound to N7 of a guanine (PDB ID 2adw). The latter one is the structure mentioned above containing Watson-Crick basepairs next to the bisintercalation site (Cuesta-Seijo *et al.*, 2006). It was found that the metal site in this structure is occupied by both zinc(II) and nickel(II) ions,

but that it is probably also capable of binding different transition metal cations based on their availability. This assumption was the initial motivation for crystallization trials of the same complex with different metal ions, in case of the structure presented here with manganese(II). Another aim was to further investigate the flexibility of the base pairing for the bases next to the bisintercalation site.

4.2 MATERIALS AND METHODS

Echinomycin was purchased in lyophilized form from Sigma-Aldrich, d(ACGTACGT) from Carl Roth GmbH purified by HPLC. Both substances were used without further purification.

Crystallization

Crystals were grown at 20 °C in hanging drops by vapor diffusion. The reservoir solution contained 24% v/v PEG 200, 6% w/v PEG 3350, 16 mM manganese(II) chloride, 20 mM spermine tetrachloride and 0.1 M MES buffer (2-morpholino-ethanesulfonic acid) pH 6.0. DNA-drug solution contained 0.21 mM DNA (single-strand concentration), 0.25 mM echinomycin and 50% v/v methanol. The DNA-drug solution was prepared by mixing echinomycin dissolved in methanol with aqueous oligonucleotide solution in a 1 to 1 ratio at room temperature and incubation for six days at 4 °C. Hanging drops prepared from 20 µl DNA-drug solution and 1 µl reservoir solution were equilibrated against 500 µl reservoir solution. After one week spherulites appeared in the crystallization drop, after two months colorless, bar-shaped tetragonal crystals had grown to a size of $0.1 \times 0.1 \times 0.3$ mm. For data collection at 100 K the crystals were flash cooled in liquid nitrogen. The mother liquor around the crystals proved to be sufficient for cryoprotection.

Structure solution and refinement

All datasets used for this project are listed in Table 4.1. The in-house dataset home2 was integrated by SAINT (Bruker AXS, Madison WI, USA), all other datasets were integrated by XDS (Kabsch, 1993), data scaling was performed with SADABS (Bruker AXS, Madison WI, USA), space group determination and dataset preparation was done by XPREP (Bruker AXS, Madison WI, USA).

The structure was solved with a highly redundant SAD dataset measured in-house on a three-circle diffractometer using the anomalous dispersion of manganese. With a resolution cutoff of 2.1 Å SHELXD (Schneider & Sheldrick, 2002) found three manganese sites with a relative occupancy of 1.0 : 0.4 : 0.1, of which the first two turned out to be real. With a phase extension to 1.0 Å and 200 cycles of density modification SHELXE (Sheldrick, 2002) produced an easily interpretable map (Figure 4.2A).

For refinement, a high resolution dataset collected at SLS (Switzerland), which contained a lot of overloaded reflections, was merged with an in-house dataset from a one-circle diffractometer without overloads. The merged dataset is almost complete to 1.25 Å resolution, beyond that resolution some reflections are missing; the completeness in the last resolution shell from 1.20 – 1.10 Å is 76.5%.

The structure was refined anisotropically against merged data with SHELXL (Sheldrick & Schneider, 1997) to $R_1 = 16.0\%$ and $R_{free} = 20.3\%$ (Figure 4.2B).

Table 4.1. Data collection, phasing and refinement statistics

	SLS	home1	home2
Crystal data			
Space group	P4$_1$2$_1$2	P4$_1$2$_1$2	P4$_1$2$_1$2
a, Å	26.54	26.43	26.46
c, Å	162.32	162.27	162.44
Diffraction data			
Beamline	PXII (SLS)	Home-source (mar345)	Home-source (Smart 6000)
Collection date	23.05.08	22–23.12.04 by J. A. Cuesta-Seijo	21–24.12.05
Wavelength, Å	1.00000	1.54178	1.54178
Resolution limit, Å	1.1	2.23	1.65
Total reflections	130500	37857	175783
Unique reflections	25106	3325	7727
Completeness, %			
Overall	92.2	99.5	97.7
Outermost resolution shell	76.5	99.5	89.4
Redundancy			
Overall	5.20	11.39	22.75
Outermost resolution shell	2.17	11.59	9.31
$I/\sigma(I)$			
Overall	24.06	23.93	41.42
Outermost resolution shell	4.44	12.85	9.75
R_{int}			
Overall	3.28	7.42	4.78
Outermost resolution shell	20.19	21.53	19.02
R_{sigma}			
Overall	2.84	3.16	1.86
Outermost resolution shell	22.63	7.81	10.40
Data merging (sls-home1)			
R_{int}(sls-home1), %	15.38		
Completeness, %			
Overall	94.5		
Outermost resolution shell	76.5		
to 1.25 Å resolution	99.7		
Phasing			
Resolution, Å			2.1
pseudo-free CC after dm, %			70.13
Refinement			
Reflections used	22411		
Resolution, Å	1.1		
$R_1(F_O > 4\sigma\,F_O)$, %	15.97		
$R_{free}(F_O > 4\sigma\,F_O)$, %	20.34		
rms deviation			
DFIX, Å	0.016		
DANG, Å	0.033		

A

B
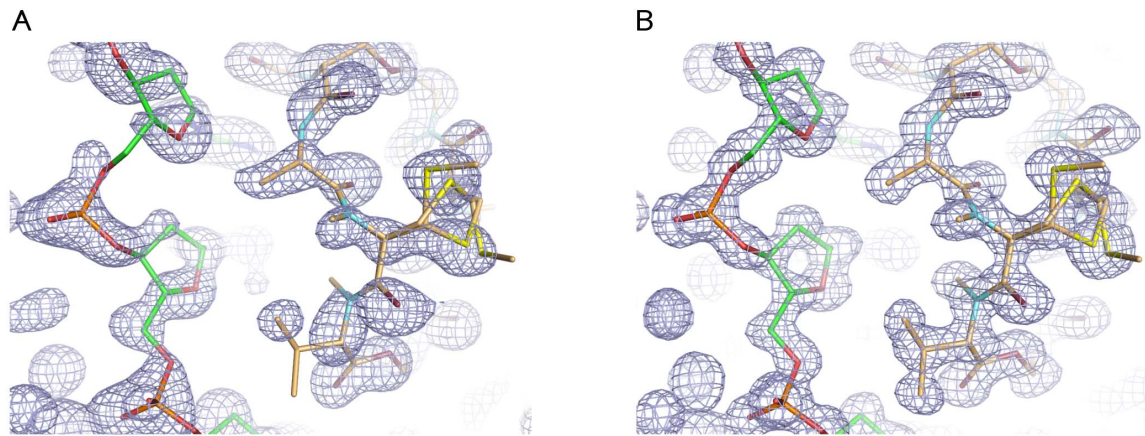
Figure 4.2. Experimental map (**A**) and map after final refinement (**B**), both contoured at a 1 σ level. The carbon atoms of echinomycin are shown in light orange.

4.3 RESULTS AND DISCUSSION

Overall structure

The asymmetric unit consists of a DNA duplex with two echinomycin molecules bisintercalating around both CG steps (Figure 4.3). The sugar-phosphate backbone belonging to residues A(101) and C(102) is slightly disordered, resulting in two different orientations of the terminal OH-group. The occupancy of the two conformations refined to 52% to 48%.
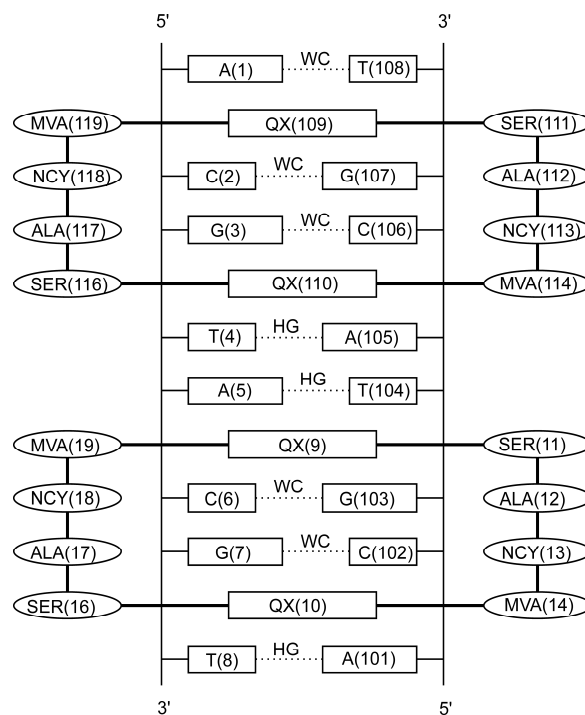
Figure 4.3. Simplified view of the DNA-echinomycin complex with residue names. WC stands for Watson-Crick and HG for Hoogsteen base pairing.

Besides the intercalation of the quinoxaline bases between the AT and CG base pairs the depsipeptide backbone positioned in the minor groove forms four direct hydrogen bonds with the two enclosed guanine bases (Figure 4.4). The carbonyl oxygen of alanine accepts a hydrogen bond from N2 of guanine (3.15–3.21 Å, range for four hydrogen bonds) and the alanine nitrogen donates to N3 of guanine (3.01–3.09 Å, range for four hydrogen bonds). The distances are very similar for both echinomycins.
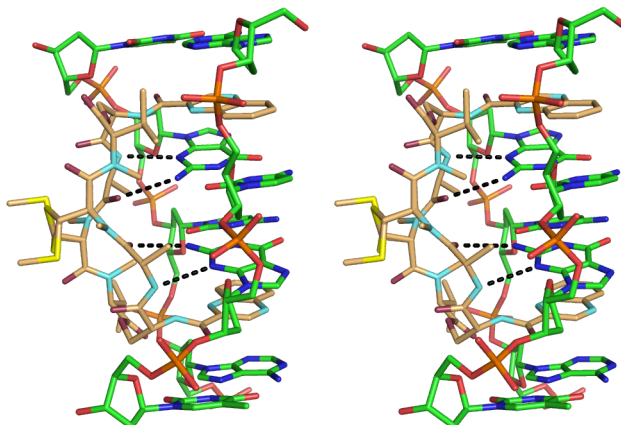

Figure 4.4. Stereo view of the bisintercalation site with echinomycin1. The carbon atoms of echinomycin are shown in light orange. Hydrogen bonds are indicated with dashed lines.

In accordance with all previous crystal structures the bases enclosed by echinomycin [base pairs C(2)-G(107), G(3)-C(106), C(6)-G(103) and G(7)-C(102)] show Watson-Crick pairing, which is also indicated in Figure 4.3. Most bases flanking the bisintercalation site perform Hoogsteen pairing except for the terminal base pair A(1)-T(108) performing Watson-Crick pairing. This result supports the conclusion from Cuesta-Seijo *et al.* (2006) that for bases flanking the bisintercalation site Hoogsteen conformation is almost equal in energy to Watson-Crick conformation. In Figure 4.5 base stacking at the intercalation site is compared for the different base pairing modes. In Figure 4.5A the upper AT basepair is in Watson-Crick mode, in Figure 4.5B it is in Hoogsteen mode. In both cases the adenine base, which has to rotate about 180° around the glycosilic bond to form a Hoogsteen pair, is involved in stronger stacking interactions with the intercalated quinoxaline base compared to thymine.

As observed in previous structures there are two conformations of the thioacetal bridge because it is not obeying the two-fold symmetry of the rest of the echinomycin molecule. The antibiotic was modeled with alternative conformations for the bridge including C_α, C_β, S_γ and C_δ of the N-methyl-cysteines and full occupancies for the remaining atoms. The occupancies for the two conformations after final refinement are 49%/51% (echinomycin1, residue 9–19) and 62%/38% (echinomycin2, residue 109–119). The S-methyl groups could be detected in the difference map and were included in the final model. Their orientation is visible in Figure 4.4.

The oligonucleotide d(ACGTACGT) without echinomycin crystallizes in A-DNA form like most DNA octamers (Wilcock *et al.*, 1996). With two echinomycin molecules bound to it, the geometry resembles slightly B-DNA, but not every backbone torsion angle lies in the typically observed region (Table 4.2). The DNA helix is highly unwound

with a flat minor groove resembling Z-DNA, although the DNA-drug complex still has a right-handed helix geometry.
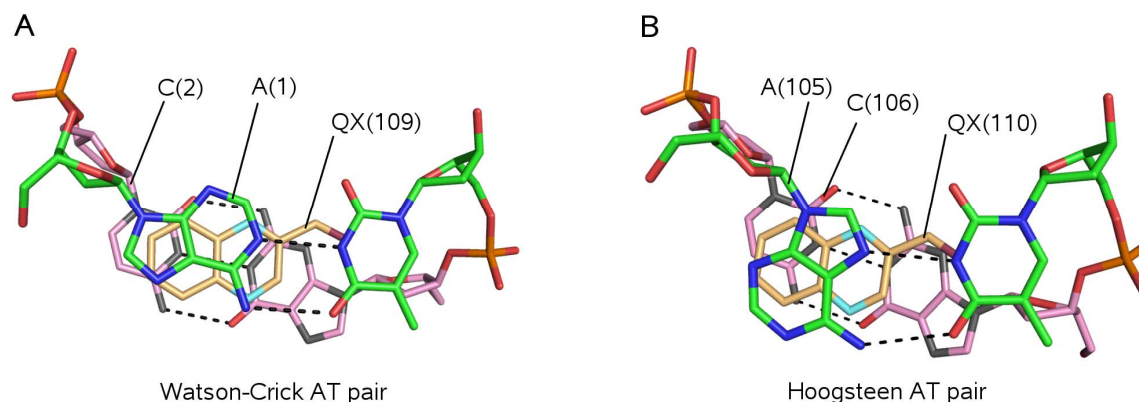
A

C(2) A(1)

QX(109)

Watson-Crick AT pair

B

A(105) C(106)

QX(110)

Hoogsteen AT pair

Figure 4.5. View down the DNA helix. The carbon atoms of the upper AT pair are shown in green, the ones of the lower CG pair in pink. The quinoxaline base is positioned between the two pairs, with its carbon atoms colored in light orange.

Table 4.2. Backbone and glycosyl torsion angles

Residue	α	β	γ	δ	ε	ζ	χ
			Angle, degrees				
A1	-	-	46	150	210	288	252
A101 (HG)	-	-	313 (188)	119 (83)	223 (203)	311(292)	63 (66)
C2	135	158	188	88	237	294	245
C102	180 (185)	128 (166)	186	96	224	285	251
G3	261	66	161	143	267	201	254
G103	256	88	147	140	271	190	257
T4 (HG)	288	166	49	140	181	270	269
T104 (HG)	290	163	51	139	181	271	266
A5 (HG)	287	176	49	89	212	291	71
A105 (HG)	299	168	43	100	215	292	71
C6	172	144	180	81	191	277	261
C106	173	138	179	78	180	279	257
G7	297	162	55	79	175	272	261
G107	153	203	187	129	233	191	264
T8 (HG)	307	166	56	87	-	-	219
T108	317	141	51	135	-	-	252
B-DNA	270–330	130–200	20–80	70–180	160–270	150–300	200–300
Z-DNA (A/G)	40–100	150–250	160–210	80–160	180–300	280–340 40–100	50–90
(T/C)	150–250	150–250	20–90	80–160	180–300	40–100	180–220

Torsion angles are defined as α: O3'-P-O5'-C5', β: P-O5'-C5'-C4', γ: O5'-C5'-C4'-C3', δ: C5'-C4'-C3'-O3', ε: C4'-C3'-O3'-P, ζ: C3'-O3'-P-O5', χ(Purines): O4'-C1'-N9-C4, χ(Pyrimidines): O4'-C1'-N1-C2. Torsion angles in brackets describe a second conformation. In the case of Z-DNA values for purines (A/G) and pyrimidines (T/C) have to be distinguished. HG stands for Hoogsteen base pairing

Interactions with manganese(II) ions

The metal site described by Cuesta-Seijo *et al*. (2006) can be confirmed in this work. Two guanine bases are in complex with manganese ions, in particular atom N7 of residue G(107) and atom N7 of residue G(3). In the first case, the phosphate group of a symmetry related molecule and four water molecules are also involved in the complex, which has a nearly perfect octahedral geometry (Figure 4.6). The distance between manganese and N7 of residue G(107) is 2.31 Å.
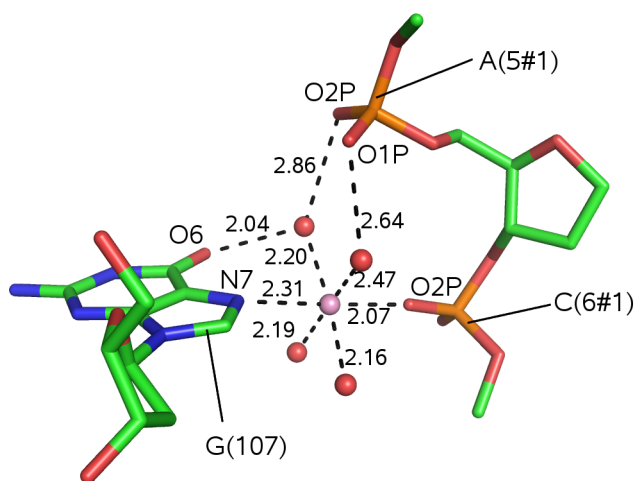

Figure 4.6. Guanine base in complex with a manganese(II) ion. The manganese ion is shown as pink sphere, water molecules as red spheres. Distances are given in Å.

The site of the second manganese ion is not fully occupied. The ion lies close to a symmetry equivalent of itself, therefore the occupancy was set to 50%, which fits approximately to the occupancy calculated by SHELXD (see above). The distance to N7 of residue G(3) is 2.38 Å. Water molecules surrounding the ion do not appear clearly in the difference map, probably because they are involved in some disorder due to the half occupied metal side, and were not modeled. The ratio in the crystallization drop between manganese ions and DNA-echinomycin duplexes was 8 to 1.

It was suggested by Gao *et al*. (1993) that direct binding between cobalt(II) and N7 of intra-helical guanine is not possible in A- and B-form DNA because there would not be enough room for the hydration waters of the metal ion in the deep major groove. This assumption was expanded to other bivalent transition metals (Ni^{2+}, Zn^{2+}) and confirmed by several crystallographic studies (Abrescia *et al*., 2001; Labiuk *et al*., 2003). In these studies interactions with transition metals where only observed for terminal or flipped-out bases but not for intra-helical bases. Mg^{2+} usually binds to N7 of guanine in B-DNA through a water molecule from its coordination sphere but the bivalent transition metals like Mn^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+} and Zn^{2+} are softer lewis acids and have therefore a higher affinity for nitrogen.

The situation is completely different for Z-DNA where the major groove is flat and the bases are more exposed. Here Gao *et al*. (1993) observed direct binding between cobalt(II) and N7 of intra-helical guanine with an averaged bond distance of 2.3 Å.

In order to get an impression how frequently manganese ions interact directly with DNA nucleobases a search in the PDB for structures obtained by X-ray analysis (44,415

structures available) was performed on 25.07.2008. The search for structures containing Mn^{2+} ions (1145 structures available) and DNA (2686 structures available, allowing presence of protein and RNA) resulted in 78 hits. These structures were analyzed manually and 19 of them showed a direct contact between manganese(II) and a non-terminal intra-helical nucleobase. Those structures were exclusively DNA-protein complexes, in which the DNA is significantly bent by the protein, which supports the original assumption from Gao *et al.* (1993) mentioned above. 13 of them contained the nucleosome core particle (e. g. PDB ID 2nzd), three contained a polymerase (e. g. PDB ID 3C2L), two contained a transposase (e. g. PDB ID 2vju) and one an integration host factor protein (PDB ID 2iie). The nucleosome core particle with manganese ions is shown in Figure 4.7. The particle contains a DNA duplex with 145 base pairs (Ong *et al.*, 2007), which is wrapped around a histone protein octamer. It has to be added that these statistics do not take into account how many crystallization trials with manganese were actually performed.
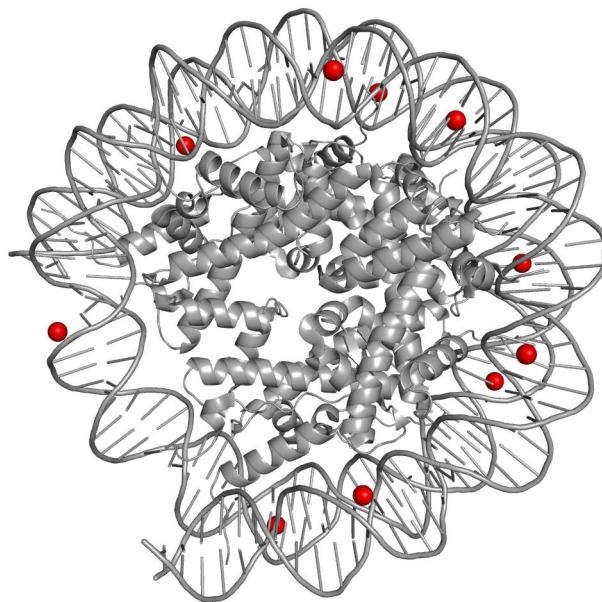
Figure 4.7. The nucleosome core particle (PDB ID 2nzd). Manganese ions are shown as red spheres.

The two guanine bases complexing manganese are both enclosed by echinomycin2, a position where the DNA helix is highly unwound. Interestingly also two bases enclosed by echinomycin1 are involved in some untypical interactions (Figure 4.8). N7 of residue G(103) accepts a hydrogen bond (2.77 Å) from the terminal OH-group of a symmetry equivalent of residue A(1) and N4 of residue C(102) donates a hydrogen bond (2.98 Å) to the phosphate backbone belonging to a symmetry equivalent of residue T(8). In a previous structure (PDB ID 2adw) a molecule of the buffer MES is present at this site, forming a hydrogen bond with N7 of guanine. Packing interactions involving non-terminal bases belonging to duplex DNA are not very common and it seems likely that these effects are supported by echinomycin.
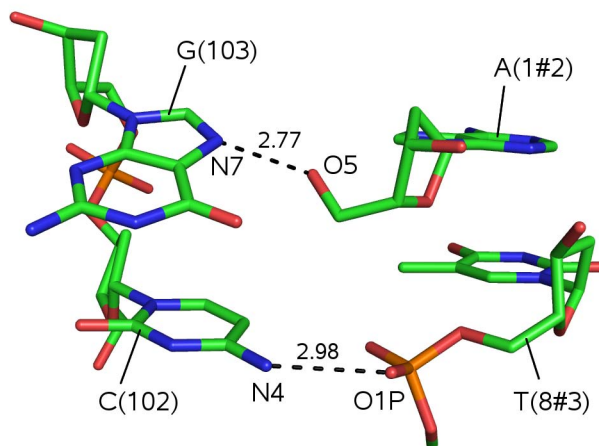
Figure 4.8. Packing interactions between bases enclosed by echinomycin and residues belonging to symmetry related molecules. Distances are given in Å.

DNA unwinding

Figure 4.9 illustrates the unwinding in the DNA-echinomycin complex (A) in comparison with B-DNA (B).
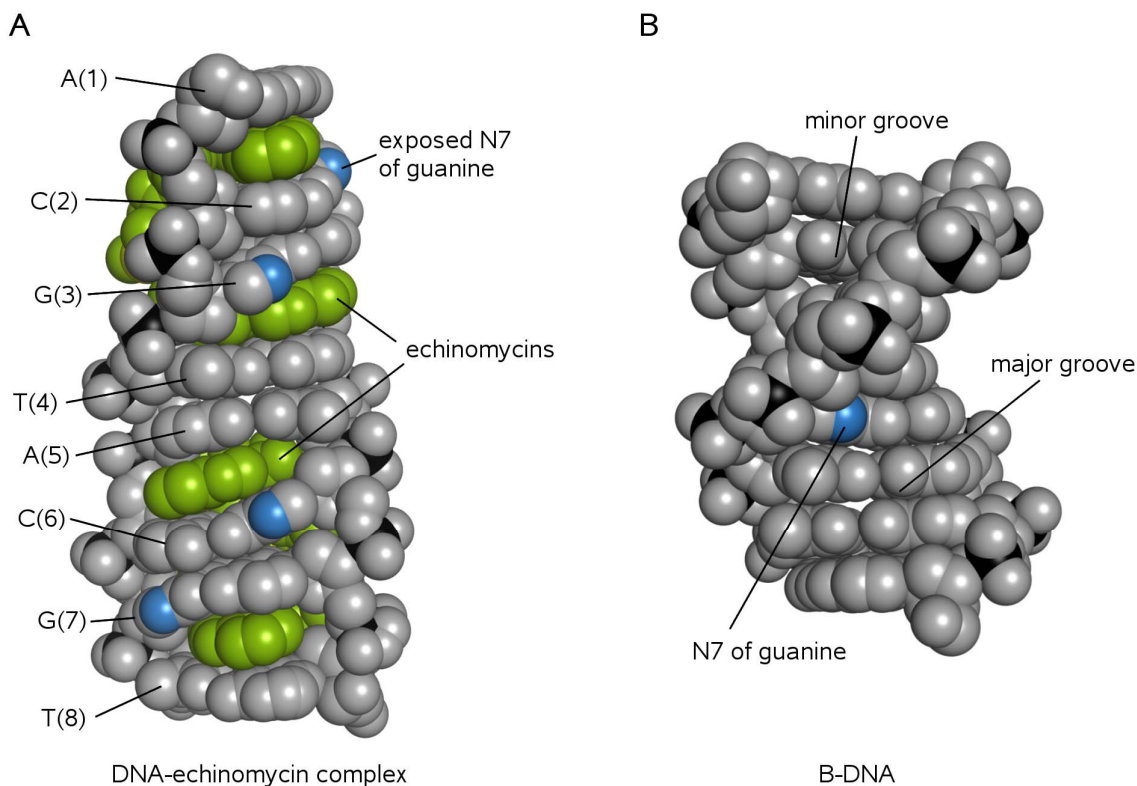

Figure 4.9. Space-filling model of the DNA-echinomycin complex (**A**) and B-DNA (**B**). Echinomycin is shown in green, phosphorus atoms in black, N7 of guanines in blue.

The B-DNA structure (Chiu & Dickerson, 2000; PDB ID 1en3) of decamer d(CCAACGTTGG) was shortened to an octamer by omitting the terminal residues in order to compare the overall helical twist of the two structures. The B-DNA structure is twisted from the top to the bottom base pair by more than 360°, whereas the DNA-drug complex is twisted from A(1) to T(8) only by about 120°. Some base pairs are actually oriented parallel to each other in this structure. The major groove in B-DNA is wide and deep, in the DNA-drug complex it is not recognizable any more. Some parts of it seem to be on the verge of obtaining a left-handed helix geometry, which can be recognized by looking at the two phosphate groups at the upper left side. In both structures N7 atoms belonging to non-terminal guanine bases are highlighted in blue. In the DNA-drug complex the two upper N7 atoms are interacting directly with manganese and the third one from above is accepting a hydrogen bond from a OH-group, only the one belonging to guanine G(7) is in contact with a water molecule. It is recognizable that all four N7 atoms are very exposed in this structure, whereas in B-DNA the N7 of guanine is more hidden at the side of the major groove, mainly by the phosphate group but also by the sugar and the base from the cytosine residue positioned above it.

Crystal packing

The DNA duplexes form infinite columns along the c-axis of the crystal (Figure 4.10). The manganese ions are involved in crystallographic contacts between the columns.
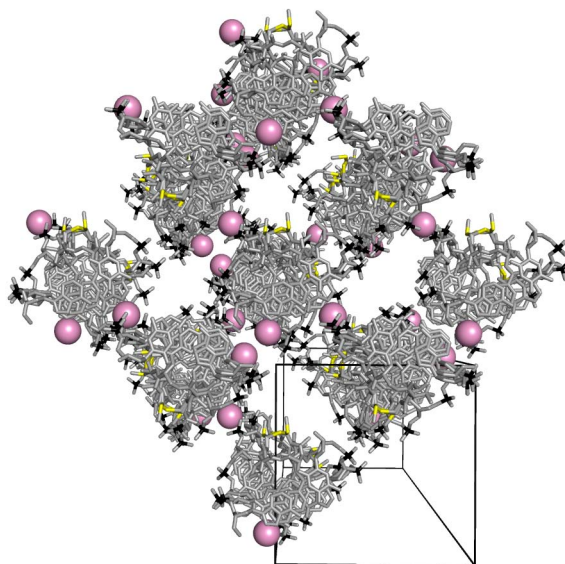
Figure 4.10. View along the *c* axis of the crystal. Manganese ions are shown as pink spheres, sulfur atoms of echinomycin are colored in yellow, phosphorus atoms in black. The black box represents the unit cell.

Besides the packing interactions already described above there are some more involving mostly terminal bases and the sugar-phosphate backbone. Their big number probably explains the strong diffraction of the crystals. The columns itself are stabilized by base stacking between the duplexes with T(8) lying over A(1#4) and A(101) over T(108#4). [A(1#4) means symmetry equivalent number 4 of residue A(1), this notation is

used to distinguish different symmetry equivalents of the DNA-drug complex.] The second stacking interaction is also supported by a hydrogen bond (2.57 Å) between the disordered terminal O5' of residue A(101) (conformation b) and the phosphate group of residue T(108#4). Additionally, N6 of residue A(101) donates a hydrogen bond (2.86 Å) to O1P of residue G(7#3) and N1 of residue A(101) accepts a hydrogen bond (2.47 Å) from O2P of residue G(7#3).

5. Effects of anisotropic scaling on structure solution

5.1 INTRODUCTION

In modern macromolecular crystallography a high amount of automatization is desired to speed up the process of structure determination. For proteins it has become common practice to use an autotracing program for initial model building (Perrakis *et al.*, 1999). The programs locate typical protein motives like α-helices or β-sheets in the experimental electron density and try to fit the correct amino acids into it. The sequence of amino acids in the protein is usually known and can be read into the autotracing program. If the experimental electron density map has a high quality, the major part of a protein can be built automatically and only critical parts have to be built manually. Similar programs are being developed for DNA and RNA at the moment. However, the success of automated and manual model building depends on the experimental density maps.

The quality of experimental maps can be improved by different techniques, for example density modification. Here, the basic idea is to make a chemical sensible modification of the experimental electron density. Improved estimates for the phases are then obtained by an inverse Fourier transformation. One of the simplest ideas for a physical sensible modification is to truncate negative electron density to zero. Another approach is to detect protein and solvent regions in the unit cell and to truncate the density in the solvent region to zero. (Crystals of macromolecules contain a high amount of non-ordered solvent, usually between 30% and 70%.) Therefore, density modification works best if the solvent content in a crystal is high.

For density modification the quality and the resolution of the native dataset are important and also how well the native datasets fits to the derivative datasets used for phase determination. It can actually happen that the heavy atom substructure is determined correctly but that it is not possible to obtain an interpretable map because of problems in the native dataset. Such problems can have different reasons; one of them is anisotropic diffraction. In this study the effect of anisotropic scaling of native data on phasing, density modification and autotracing is tested. Three different methods for anisotropic scaling proposed by Sheriff & Hendrickson (1987), Shakked (1983) and Matthews & Czerwinski (1975) are compared. A detailed description of these methods is given in chapter 1.2 (Data scaling).

5.2 MATERIALS AND METHODS

General procedure

Phasing, density modification and autotracing were tested with SHELXQ (version 2006/4), which is closely related to SHELXE (Sheldrick, 2002) but not distributed yet. The program was run in two different modes, depending on the origin of the phase information. If the substructure was determined with SHELXD (Schneider & Sheldrick, 2002), the program was run in normal mode, which means that a hkl-file containing native data, another hkl-file containing FA and α values and a res-file containing the heavy atom positions were read in. If some phase information from a different source

were available, it was read in as phi-file in phs-format, together with a hkl-file containing native data and an ins-file containing cell and symmetry information.

The native datasets were anisotropically scaled by the program XPREP (test version), with the three methods described above. Each dataset was tested with different input parameters for SHELXQ to ensure a statistical variation of the results. A detailed list of these parameters is given below for the different test structures. SHELXQ was run on a grid system involving 22 computers in the group of Isabel Usón (Institute of Molecular Biology of Barcelona, Spain). The log files produced by SHELXQ containing mean phase errors and validation criteria for the autotracing were analyzed with perl scripts.

Test structure 1

Test structure 1 refers to a protein which has not been published yet. The final model contains 592 amino acids, the space group is C2 (a = 135.74 Å; b = 51.39 Å, c = 108.44 Å; β = 110.94°). Data collection, initial structure solution and structure refinement was done by D. Monferrer. The starting phases for SHELXQ were generated by the program SHARP (Bricogne et al., 2003) with data from a MIRAS experiment using mercury and xenon derivatives. Four different datasets were compared:

- original   native synchrotron dataset collected to 1.93 Å resolution
- an1   original dataset anisotropically scaled with the method from Sheriff & Hendrickson
- an2   original dataset anisotropically scaled with the method from Shakked
- an3   original dataset anisotropically scaled with the local scaling method

The fcf-files for the calculation of the mean phase error with SHELXQ were generated by a refinement with SHELXL against dataset *original* (the resulting fcf-file was also used for datasets *an1*, *an2* and *an3*). For the evaluation of the results from autotracing the coordinate files from these refinements were used (ent-file in pdb-format).

Table 5.1 gives a discription of the input parameters used in SHELXQ. The program was run with the following command:

shelxq dataset.phi –mvar-m –avar-a –svar-s –vvar-v –evar-e –wvar-w –u0.2 –t0.8 –x2.

The used values for the expressions in blue are given in Table 5.2. Either 100 cycles of density modification without autotracing or 10 cycles of density modification with 9 cycles of autotracing were tried. The solvent content was varied from 30% to 60% in steps of 10%. Three different values were used for the *low density reduction factor* (0.0, 0.25 or 0.5). The phases were extended to five different resolutions up to 1.0 Å. Also the original resolution was used (-e parameter not set). The weight for retained experimental phases was set to 0.0, 0.25 or 0.5. To save computing time the correct substructure was used, so that no invertion of the substructure was necessary. Each of the four datasets was tested with all combinations of the parameter setting. This results in 1728 jobs for SHELXQ, 216 with autotracing and 216 without autotracing for each datset.

Table 5.1. Description of input parameters for SHELXQ

parameter	description
m	Cycles of density modification
a	Iterations autotracing
s	Solvent fraction
v	Low density reduction factor
e	Phase extension
w	Weight for retained experimental phases
u	CC threshold to accept alpha helix or tri-peptide fragment
t	Mean atomic density threshold (sigma) to accept peptide chain
x1	Reads in a fcf-file for calculation of the mean phase error
x2	Reads in a fcf-file for calculation of the mean phase error and an ent-file for an evaluation of the autotracing

Table 5.2. Values of input parameters for SHELXQ (*test structure1*).

var-m, var-a	var-s	var-v	var-e	var-w
100, -	0.3	0	-	0
10, 9	0.4	0.25	1.9	0.25
	0.5	0.5	1.7	0.5
	0.6		1.5	
			1.2	
			1.0	

Test structure 2

Test structure 2 is an alternating D,L-peptide (Alexopoulus *et al.*, 2004; PDB ID 1uno), which crystallized in space group $P4_32_12$ (*a*, *b* = 27.999 Å; *c* = 78.932 Å). The asymmetric unit contains two molecules that form a right-handed antiparallel *β*-helix. The structure has been solved by SIRAS using an iodine-soaked crystal, the iodine substructure was determined by SHELXD. The following datasets were compared:

- original   native synchrotron dataset collected to 1.20 Å resolution
- an1   original dataset anisotropically scaled with the method from Sheriff & Hendrickson
- an2   original dataset anisotropically scaled with the method from Shakked
- an3   original dataset anisotropically scaled with the local scaling method

The fcf-files for the calculation of the mean phase error with SHELXQ were generated analogue to *test structure 1*. SHELXQ was run with the following command, values for the expressions written in blue color are given in Table 5.3):

shelxq dataset yy –mvar-m –svar-s –vvar-v –evar-e –wvar-w –u0.2 –t0.8 –x1

The FA and α values are read from the file yy.hkl, the iodide positions from yy.res. No autotracing was tried with these datasets, because the structure does not contain any typical elements like α-helices. This results in 288 jobs for SHELXQ, 72 for each dataset.

Table 5.3. Values of input parameters for SHELXQ (*test structure2*).

var-m	var-s	var-v	var-e	var-w
300	0.3	0	-	0
	0.4	0.25	1.0	0.25
	0.5	0.5		0.5
	0.6			

5.3 RESULTS AND DISCUSSION

Test structure 1

According to the different scaling methods, the intensities were multiplied by the following scaling terms in XPREP:

Sheriff & Hendrickson (1987): $\exp (0.00014\ h^2 + 0.00038\ k^2 + 0.00020\ l^2 + 0.00009\ hl)$

Shakked (1983): $0.00009\ h^2 + 0.00042\ k^2 + 0.00007\ l^2 + 0.00007\ hl$

Matthews & Czerwinski (1975):   radius = $0.0848\ \text{Å}^{-1}$,
821 neighbors used on average.

In Table 5.4 the effect of the different scaling methods on the weighted mean phase error (wMPE) and the pseudo-free correlation coefficient after 100 cycles of density modification without autotracing is shown.

Table 5.4. Results for *test structure 1* without autotracing

dataset	wMPE, °		Pseudo free CC, %	
	mean	best	mean	best
original	60.4	53.2	71.9	77.9
an1	59.8	52.4	70.0	76.6
an2	59.9	52.6	70.2	76.5
an3	59.7	52.3	69.8	76.2

For each validation criteria mean values and the top values are given. The anisotropically scaled datasets gave nearly identical results, wMPE is slightly lower compared with the original dataset. The correlation coefficient is slightly lower for the anisotropically scaled datasets.

In Table 5.5 the effect of the different scaling methods on autotracing is analyzed. The weighted mean phase error (wMPE), the number of traced residues (which can be traced correctly or incorrectly) and the percentage of the model that could be traced correctly is given for each datasets. Only jobs with autotracing are included, for the calculation values from the autotracing cycles 4–9 were used.

Table 5.5. Effect on autotracing for *test structure 1*

dataset	wMPE, °		Number of traced residues		Percentage of final model traced	
	mean	best	mean	most	mean	most
original	54.2	49.5	371.3	441	61.0	72.4
an1	53.9	48.4	376.6	458	62.0	75.4
an2	54.1	48.5	375.5	451	61.8	74.4
an3	53.5	48.3	380.0	440	62.5	72.1

The anisotropically scaled datasets produced slightly better values than the original one. With dataset *an1* it was possible to trace 75.4% of the final model, with the original one 72.4% could be traced. Datasets *an1* led to better top values (458 residues traced in the best case, 17 more than with original dataset), *an3* to better mean values (380.0 residues traced on average, 8.7 more than with original dataset).

Test structure 2

Anisotropic diffraction is more severe in the second test structure, which gets obvious by comparing the local scaling terms given below with the ones for the previous structure.

Sheriff & Hendrickson (1987): $\exp (0.00221\ h^2 + 0.00221\ k^2 + 0.000321\ l^2)$

Shakked (1983): $0.00197\ h^2 + 0.00197\ k^2 + 0.000071\ l^2$

Matthews & Czerwinski (1975): radius = 0.1965 Å^{-1}, 1790 neighbors used on average.

Table 5.6 shows the mean and best values of the mean phase error and the pseudo-free correlation coefficient for the different datasets. The best mean phase error for the original dataset is 64°, whereas all anisotropically scaled datasets give top results below 20°. As observed in the previous test structure, the original dataset gives a higher mean value for the correlation coefficient, although in this case all anisotropically scaled datasets produce better top values above 80%.

Table 5.6. Results for *test structure 2*

dataset	wMPE, °		pseudo free CC, %	
	mean	best	mean	best
original	78.6	64.0	72.0	77.8
an1	55.5	18.6	68.8	80.9
an2	51.8	19.3	69.5	81.4
an3	57.6	18.7	67.8	80.1

Amongst the anisotropically scaled datasets, *an2* gives the highest correlation coefficients and the lowest mean value for wMPE.

Figure 5.1 draws the correlation coefficient against the mean phase error, each dataset is shown in a different color.
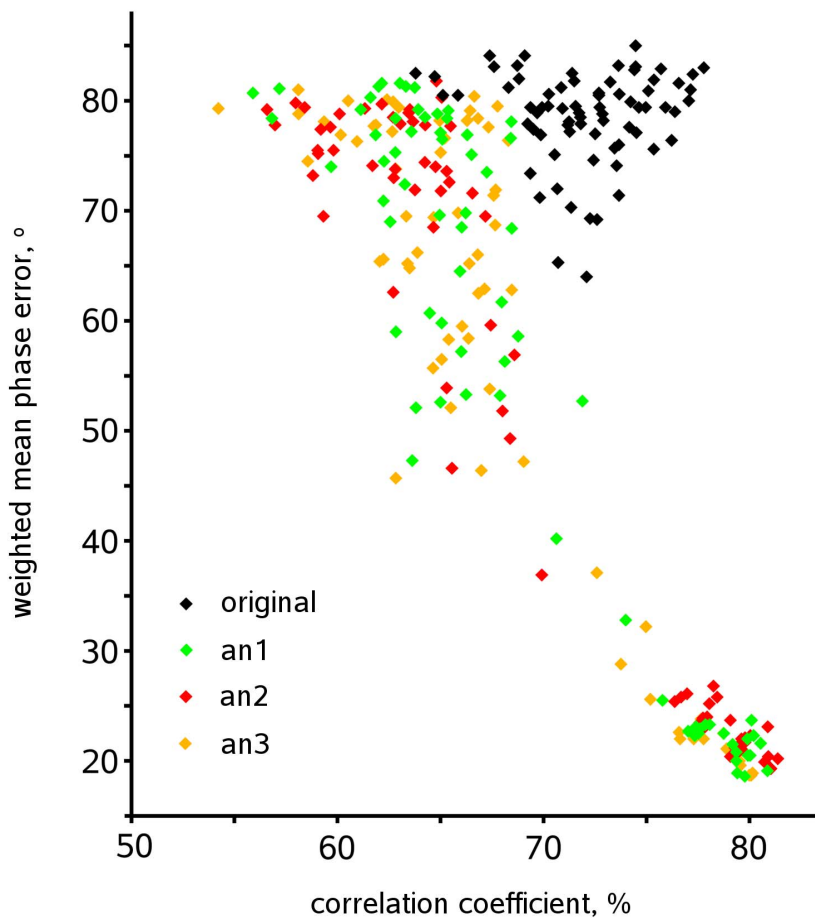

Figure 5.1. Correlation coefficient drawn against wMPE for *test structure 2*. Only the anisotropically scaled datasets produce mean phase errors below 30°.

It is visible that a lot of trials produced mean phase errors above 45° for the anisotropically scaled datasets with relatively low correlation coefficients. Almost no

trials with mean phase error between 30° and 45° can be observed, but some below 30° having a high correlation coefficient. Out of 72 trials for each dataset 21 resulted in a wMPE below 30° for dataset *an1*, 29 for *an2*, 17 for *an3* and none for *original*. Trials that resulted in mean phase errors below 30° were carried out almost exclusively with a solvent content of 50% or 60%. Thus, the solvent content seems to be the most critical parameter in this case.

Figure 5.2 compares the experimental electron density map obtained with the original dataset (A) with the experimental map obtained with an anisotropically scaled dataset (B). Both maps were obtained with the same input parameters for SHELXQ. For this structure anisotropic scaling has a dramatic effect on the quality of the experimental maps. Without anisotropic scaling the map is very difficult to interpret because the peptide backbone is hardly detectable and very few side chains are visible. With anisotropic scaling a high quality map can be obtained that shows holes in nearly every aromatic ring.
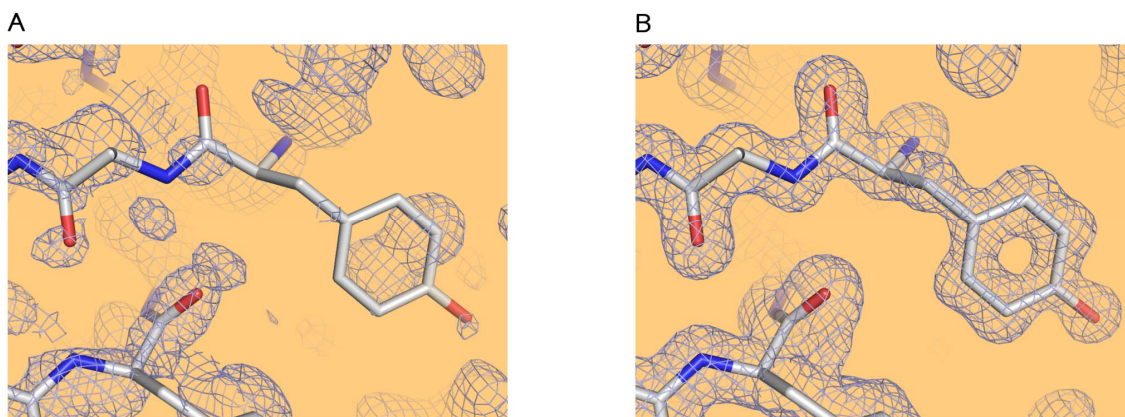

Figure 5.2. Experimental maps obtained with original (**A**) and anisotropically scaled datasets (**B**). The map on the left side has a weighted mean phase error of 69.2°, the one on the right side 18.9°.

Conclusions

The different scaling methods give slightly different results, but the number of tests in this study does not allow a general conclusion which of the three is the best one. It might actually be the case that the optimal scaling strategy is different for each structure. It could be shown that anisotropic scaling makes structure solution easier and electron density maps more interpretable, depending on the amount of anisotropy. In cases with very high anisotropy, the effect of anisotropic scaling can be dramatic as for the second test structure.

Summary

To shed more light onto the stereo chemical requirements for the covalent bond formation and subsequent elimination of a nucleobase, the structure of trioxacarcin A covalently bound to d(AACCGGTT) was determined to 1.78 Å resolution by X-ray diffraction. Experimental phases were obtained by MAD experiments using isomorphous crystals containing brominated nucleotides. The crystal structure reveals an unexpected base flip-out at the intercalation site combined with intercalation, alkylation and register-shift. It provides insight into the mechanism of the formation of gutingimycin with the abstraction of a guanine base from DNA and adds to our understanding of DNA manipulation by antibiotics.

Another structure of trioxacarcin A bound to d(CGTACG) was determined to 1.89 Å resolution. In both cases crystallization at 40 °C improved the crystal quality significantly compared to 20 °C, which suggests using this temperature routinely for DNA crystallization. Experimental phases were obtained by a MAD experiment using a cobalt derivative and a SIRAS experiment from a crystal soaked in uranyl nitrate. The structure factors from both experimental methods were merged in reciprocal space. In this case the crystal structure reveals the formation of an unexpected hexaplex due to deformation of the classical DNA duplex by two trioxacarcin molecules. One of the main difference to the previous structure is the orientation of the guanine base to the chromophoric ring system of the drug, which corresponds to the one observed in the cleavage product gutingimycin.

A new crystal form of echinomycin in complex with d(ACGTACGT) was obtained with manganese(II) ions, the structure could be determined to 1.2 Å resolution. The structure presented here is similar to its previous analogues. In this work many crystallographic contacts with bases in the major groove involving also bivalent manganese ions were observed which might be related with the DNA unwinding induced by echinomycin. This might be an explanation for the biological activity of echinomycin.

Three different methods of anisotropic scaling were tested on an unpublished protein and a D,L-alternating peptide. In the first case autotracing produced slightly better results for the anisotropically scaled datasets. In the second case anisotropic diffraction was more severe and anisotropic scaling improved the electron density maps dramatically. It could be shown that the quality of the native dataset affects structure solution during density modification.

References

Abrescia N.G.A., Huynh-Dinh T., Subirana J.A. (2002) Nickel-guanine interactions in DNA: crystal structure of nickel-d[CGTGTACACG]$_2$. *J. Biol. Inorg. Chem.*, **7**, 195–199.

Alexopoulos E., Küsel A., Sheldrick G.M., Diederichsen U., Usón I. (2004) Solution and structure of an alternating D,L-peptide. *Acta Crystallogr.*, **D60**, 1971–1980.

Bansal M. (1999) DNA structure: Yet another avatar? *Curr. Sci.* **76**, 1178–1181.

Bieniossek C., Schütz P., Bumann M., Limacher A., Usón I., Baumann U. (2006) The crystal structure of the carboxy-terminal domain of human translation initiation factor elF5. *J. Mol. Biol.*, **360**, 457–465.

Bricogne G., Vornrhein C., Flensburg C., Schiltz M., Paciorek W. (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr.*, **D59**, 2023–2030.

Brünger A.T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.

Chiu T.K., Dickerson R.E. (2000) 1 Å crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium. *J. Mol. Biol.*, **301**, 915–945.

Collaborative computational project number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Collier D.A., Neidle S., Brown J.R. (1984) Molecular models for the interaction of the anti-tumor drug nogalamycin with DNA. *Biochem. Pharmacol.*, **33**, 2877–2880.

Cruickshank D.W.J. (1970) Least squares refinement of atomic parameters. *Crystallographic Computing*, Munksgaard, Copenhagen, 187–196.

Cuesta-Seijo J.A., Sheldrick G.M. (2005) Structures of complexes between echinomycin and duplex DNA. *Acta Crystallogr.*, **D61**, 442–448.

Cuesta-Seijo J.A., Weiss M.S., Sheldrick G.M. (2006) Serendipitious SAD phasing of an echinomycin-(ACGTACGT)$_2$ bisintercalation complex. *Acta Crystallogr.*, **D62**, 417–424.

Dell A., Williams D.H., Morris H.R., Smith G.A., Fenney J., Roberts G.C.K. (1975) Structure revision of the antibiotic echinomycin. *J. Am. Chem. Soc.*, **97**, 2497–2502.

Emsley P., Cowtan K. (2004) Coot: model building tools for molecular graphics. *Acta Crystallogr.*, **D60**, 2126–2132.

Evans P., McCoy A. (2008) An introduction to molecular replacement. *Acta Crystallogr.*, **D64**, 1–10.

Fenn J.B., Mann M., Meng C.K., Wong S.F., Whitehouse C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science*, **246**, 64–71.

Fitzner A., Frauendorf H., Laatsch H., Diederichsen U. (2008) Formation of gutingimycin: analytical investigations of trioxacarcin A mediated alkylation of dsDNA. *Anal. Bioanal. Chem.*, **390**, 1139–1147.

Gao Y.-G., Sriram M., Wang A.H.-J. (1993) Crystallographic studies of metal ion-DNA interactions: different binding modes of cobalt(II), copper(II) and barium(II) to N7 of guanines in Z-DNA and a drug-DNA complex. *Nucleic Acids Res.*, **21**, 4093–4101.

Ghosh A., Bansal M. (2003) A glossary of DNA structures from A to Z. *Acta Crystallogr.*, **D59**, 620–626.

Hansen M.R., Hurley L.H. (1996) Pluramycins. old drugs having modern friends in structural biology. *Acc. Chem. Res.*, **29**, 249–258.

Hansen M., Lee S.-J., Cassady J.M., Hurley L.H. (1996) Molecular details of the structure of a psorospermin-DNA covalent/intercalation complex and associated DNA sequence selectivity. *J. Am. Chem. Soc.*, **118**, 5553–5561.

Hansen M., Yun S., Hurley L.H. (1995) Hedamycin intercalates the DNA helix and, through carbohydrate-mediated recognition in the minor groove, directs N7-alkylation of guanine in the major groove in a sequence-specific manner. *Chemistry & Biology*, **2**, 229–240.

Hendrickson W.A., Smith J.L., Sheriff S. (1985) Direct phase determination based on anomalous scattering. *Methods in Enzym.*, **115**, 41–54.

Hoogsteen K. (1959) The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Crystallogr.*, **12**, 822–823.

Jeffreys A.J., Wilson V., Thein S.W. (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature*, **314**, 67–73.

Kabsch W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.*, **26**, 795–800.

Karle J. (1980) Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. *Int. J. Quant. Chem.*, **7**, 357–367.

Kondo J., Adachi W., Umeda S., Sunami T., Takénaka A. (2004) Crystal structures of a DNA octaplex with *I*-motif of G-quartets and its splitting into two quadruplexes suggests a folding mechanism of eight tandem repeats. *Nucleic Acids Res.*, **32**, 2541–2549.

Kondo J., Ciengshin T., Juan E.C., Sato Y., Mitomi K., Shimizu S. Takénaka A. (2006) Crystal structure of d(gcGXGAgc) with X=G: a mutation at X is possible to occur in a base-intercalated duplex for multiplex formation. *Nucleosides Nucleotides Nucleic Acids*, **25**, 693–704.

Labiuk S.L., Delbaere L.T.J., Lee J.S. (2003) Cobalt(II), nickel(II) and zinc(II) do not bind to intra-helical N(7) guanine positions in the B-form crystal structure of d(GGCGCC). *J. Biol. Inorg. Chem.*, **8**, 715–720.

Long F., Vagin A.A., Young P., Murshudov G.N. (2007) BALBES: a molecular replacement pipeline. *Acta Crystallogr.*, **D64**, 125–132.

Low C.M.L., Drew H.R., Waring M.J. (1984) Sequence-specific binding of echinomycin to DNA: evidence for conformational changes affecting flanking sequences. *Nucleic Acids Res.*, **12**, 4865–4879.

Lu X.-J., Olson W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.

Maskey R.P., Helmke E., Laatsch H. (2002) Parimycin: isolation and structure elucidation of a novel cytotoxic 2,3-dihydroquinzarin analogue of γ–indomycinone from a marine streptomycete isolate. *J. Antibiot.*, **55**, 1031–1035.

Maskey R.P., Helmke E., Kayser O., Fiebig H.H., Maier A., Busche A., Laatsch H. (2004a) Anti-cancer and antibacterial trioxacarcins with high anti-malaria activity from a marine streptomycete and their absolute stereochemistry. *J. Antibiot.*, **57**, 771–779.

Maskey R.P., Sevvana M., Usón I., Helmke E., Laatsch H. (2004b) Gutingimycin: a highly complex metabolite from a marine streptomycete. *Angew. Chem. Int. Ed.*, **43**, 1281–1383.

Matthews B.W., Czerwinski E.W. (1975) Local scaling: a method to reduce systematic errors in isomorphous replacement and anomalous scattering measurements. *Acta Crystallogr.*, **A31**, 480–487.

Murshudov G.N., Vagin A.A., Dodson E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.*, **D53**, 240–255.

Olson W.K., Bansal M., Burley S.K., Dickerson R.E., Gerstein M., Harvey S.C., Heinemann U., Lu X.-J., Neidle S., Shakked Z. *et al.* (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **313** 229–237.

Ong M.S., Richmond T.J., Davey C.A. (2007) DNA stretching and extreme kinking in the nucleosome core. *J. Mol. Biol.*, **368**, 1067–1074.

Otwinowski Z., Minor W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.

Owen E.A., Burley G.A., Carver J.A., Wickham G., Keniry M.A. (2002) Structural investigation of the hedamycin d(ACCGGT) complex by NMR and restrained molecular dynamics. *Biochem. Biophys. Res. Comm.*, **290**, 1602–1608.

Pannu N.S., Read R.J. (1996) Improved refinement through maximum likelihood. *Acta Crystallogr.*, **A52**, 659–668.

Parker J.B., Bianchet M.A., Krosky D.J., Friedman J.I., Amzel L.M., Stivers J.T. (2007) Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature*, **449**, 433–437.

Pavlopoulos S., Bicknell W., Craik D.J., Wickham G. (1996) Structural characterization of the 1:1 adduct formed between the antitumor antibiotic hedamycin and the oligonucleotide duplex d(CACGTG)$_2$ by 2D NMR spectroscopy. *Biochemistry*, **35**, 9314–9324.

Perrakis A., Morris R., Lamzin V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.*, **6**, 458–463.

Pfoh R., Laatsch H., Sheldrick G.M. (2008) Crystal structure of trioxacarcin A covalently bound to DNA. *Nucleic Acids Res.*, **36**, 3508–3514.

Rhodes G. *Crystallography made crystal clear.* Academic Press, San Diego, 2000.

Schneider B., Neidle S., Bermann H.M. (1997) Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers*, **42**, 113–124.

Schneider T.R., Sheldrick G.M. (2002) Substructure solution with SHELXD. *Acta Crystallogr.*, **D58**, 1772–1779.

Schomaker V., Trueblood K. (1968) On the rigid-body motion of molecules in crystals. *Acta Crystallogr.*, **B24**, 63–76.

Shakked Z. (1983) Anisotropic scaling of three-dimensional intensity data. *Acta Crystallogr.*, **A39**, 278–279.

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.

Sheriff S., Hendrickson W.A. (1987) Description of overall anisotropy in diffraction from macromolecular crystals. *Acta Crystallogr.*, **D43**, 118–121.

Stout G.H., Jensen L.H. *X-ray structure determination*. John Wiley & Sons, New York, 1989.

Sun D., Hansen M., Clement J.J., Hurley L.H. (1993) Structure of the altromycin B (N7-guanine)-DNA adduct. A proposed prototypic DNA adduct structure for the pluramycin antitumor antibiotics. *Biochemistry*, **32**, 8068–8074.

Tamaoki T., Shirahata K., Iida T., Tomita F. (1981) Trioxacarcins, novel antitumor antibiotics. II. Isolation, physico-chemical properties and modes of action. *J. Antibiot.*, **34**, 1525–1530.

Tomita F., Tamaoki T., Morimoto M., Fujimoto K. (1981) Trioxacarcins, novel antitumor antibiotics. I. Producing organism, fermentation and biological activities. *J. Antibiot.*, **34**, 1519–1524.

Wang A.H.-J., Quigley G.J., Kolpak F.J., Crawford J.L., van Boom J.H., van der Marel G.A., Rich A. (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, **282**, 680–686.

Wang A.H.-J., Ughetto G., Quigley G.J., Rich A. (1987) Interactions between an anthracycline antibiotic and DNA: molecular structure of daunomycin complexed to d(CpGpTpApCpG) at 1.2 Å resolution. *Biochemistry*, **26**, 1152–1163.

Waring M.J., Wakelin L.P.G. (1974) Echinomycin: a bifunctional intercalating antibiotic. *Nature*, **252**, 653–657.

Watson J.D., Crick F.H.C. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.

Wilcock D.J., Adams A., Cardin C.J., Wakelin L.P.G. (1996) Structure of the DNA octanucleotide d(ACGTACGT)$_2$. *Acta Crystallogr.*, **D52**, 481–485.

Williams L.D., Egli M., Gao Q., Bash P., Marel G.A., Boom J.H., Rich A., Frederick C.A. (1990) Structure of nogalamycin bound to a DNA hexamer. *Proc. Natl. Acad. Sci. USA*, **87**, 2225–2229.

Wilson A.J.C. (1942) Determination of absolute from relative X-ray intensity data. *Nature*, **150**, 150–151.

List of publications and meetings

Publications

Pfoh R., Laatsch H., Sheldrick G.M. (2008) Crystal structure of trioxacarcin A covalently bound to DNA. *Nucleic Acids Res.*, **36**, 3508–3514.

Tietze L.F., Schuster H.J., Hampel S.M., Rühl S., Pfoh R. (2008) Enantio- and diastereoselective synthesis of duocarmycine-based prodrugs for a selective treatment of cancer by epoxide opening. *Chem. Eur. J.*, **14**, 895–901.

Mösch-Zanetti N.C., Sachse A., Pfoh R., Vidovic D., Magull J. (2005) Rhenium oxo compounds containing η^2-pyrazolate ligands. *Dalton Trans.*, **12**, 2124–2129.

Meetings

01/2007   *"CCP4 Study Weekend 2007 – Molecular replacement"* in Reading, UK

09/2007   Talk at the *"Heart of Europe bio-Crystallography Meeting"* on *"Trioxacarcins and their interactions with DNA"* in Bedlewo, Poland

06/2008   Poster presentation at the *"American Crystallographic Association Meeting"* on *"Crystal structure of Trioxacarcin A covalently bound to DNA"* in Knoxville, USA

Curriculum Vitae

Persönliche Daten

Name	Roland Pfoh
Geburtsdatum	19.08.1978
Geburtsort	Bad Hersfeld
Staatsangehörigkeit	deutsch
Familienstand	ledig

Schulbildung

06/1998	Abitur an der Werratalschule in Heringen/Werra

Hochschulstudium

10/2000–05/2005	Chemiestudium an der Georg-August-Universität in Göttingen
09/2004–02/2005	Diplomarbeit an der Georg-August-Universität in Göttingen im Arbeitskreis von Prof. George M. Sheldrick zum Thema „Kristallstrukturbestimmung organischer Verbindungen mit besonderer Beachtung der absoluten Struktur".
08/2005–10/2008	Dissertation an der Georg-August-Universität in Göttingen im Arbeitskreis von Prof. George M. Sheldrick zum Thema „Crystallographic studies on DNA in complex with antibiotics and effects of anisotropic scaling on structure solution".

Göttingen, den 17.09.2008