

**Crystallographic studies on Echinomycin
bisintercalation on DNA and an alternating
D-alanyl,L-homoalanyl PNA.**

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

vorgelegt von
Jose Antonio Cuesta Seijo
aus Gijón, Spanien

Göttingen, 2005

D7

Referent: Prof. G. M. Sheldrick

Korreferent: Prof. Jörg Magull

Tag der mündlichen Prüfung: 28.06.05

ACKNOWLEDGEMENTS

I am specially grateful to Prof. George Sheldrick for the scientifically satisfying and productive time in Germany. He has been an infinite source of knowledge and an example over this three years. Many thanks also for the warm and human treatment, so uncommon in the competitive scientific world nowadays. The relaxed work atmosphere in the group can never be sufficiently appreciated.

Thanks to Isabel Usón and Regine Herbst Irmer, without them all the disorders and the terrifying twinnings would still be in my head and not in this thesis.

My gratitude to Prof. Ulf Diederichsen and Markus, Jinhua, Katrin and Björn for providing samples for the projects and for their infinite patience with the apparently never ending and never working crystallization trials.

I am in debt with Ehmke Pohl and Manfred Weiss for help with critical synchrotron data collections.

I would also like to thank all members of my research group, past and present, both for sharing their knowledge with me and for the warmth welcome they gave me in a foreign country. Also for the invaluable practical help in moments of scientific need.

Special thanks to Marianna, Tine, Johanna, Tim, Stephan and Madhu for help with the corrections of this manuscript.

Warmhearted thankfulness to the spanish mafia, Roberto, Markus, Maria, Nuria, Alberto, Victor, Miguel, Santi and Alvaro for bringing Spain with them to this small place in central Europe and so many hours of joy.

I am indebted to my family in Spain for the continued support, also from a distance.

More than just thanks to Tine for her tenderness, understanding and so many small and big things that turned Göttingen into a home for me. She makes all the difference.

ABBREVIATIONS

A	Adenine
ADP	Anisotropic displacement parameters
Ala	Alanine
C	Cytosine
CC	Correlation coefficient
CCD	Coupled charge device
CSD	Cambridge structural database
Cys	Cysteine
Dac	D-alanyl cysteine
Dag	D-alanyl guanine
DANO	Difference anomalous
DESY	Deutsches elektronen synchrotron
Dly	D-lysine
DNA	Deoxyribonucleic acid
EMBL	European molecular biology laboratory
fom	Figure of merit
G	Guanine
HPLC	High performance liquid chromatography
Lhc	L-homoalanyl cytosine
Lhg	L-homoalanyl guanine
Lys	Lysine
Me	Methyl
MAD	Multiwavelength anomalous diffraction
MES	2-Morpholinoethanesulphonic acid
Mme	Monomethyl ether
MPD	2-Methyl,2-pentenediol
MIR	Multiple isomorphous replacement
MPE	Mean phase error
mRNA	Messenger ribonucleic acid
NDB	Nucleic acid database
NMR	Nuclear magnetic resonance

PCR	Polymerase chain reaction
PDB	Protein databank
PEG	Polyethylene glycol
PNA	Peptide nucleic acid
RMSD	Root mean square deviation
RNA	Ribonucleic acid
SAD	Single wavelength anomalous diffraction
Ser	Serine
SF	Structure factors
SIR	Single isomorphous replacement
T	Thymine
Tris	Tris(hydroxymethyl)-aminomethane
tRNA	Transfer ribonucleic acid
U	Uracil
Val	Valine
wMPE	Weighted mean phase error

MOLECULAR GRAPHICS

The three-dimensional molecular graphics in this work have been produced with the programs XFIT (McRee, 1999), DINO (A. Philippsen, <http://www.dino3d.org>) and PYMOL (DeLano, 2003).

Some of the figures have been rendered with Raster3D (Merrit & Bacon, 1997)).

TABLE OF CONTENTS

I	THEORETICAL BACKGROUND	1
I.1	THE THREE-DIMENSIONAL STRUCTURE OF DUPLEX DNA	1
I.1.1	<i>Chemical structure</i>	1
I.1.2	<i>Long range three-dimensional order</i>	2
I.1.3	<i>Base pairing schemes</i>	3
I.1.4	<i>Description of the conformation of DNA</i>	5
	Torsion angles	5
	Basepairs and helical geometrical parameters.....	6
I.2	MACROMOLECULAR CRYSTALLIZATION	8
I.3	X-RAY DIFFRACTION BY MACROMOLECULES	11
I.4	METHODS FOR OBTAINING INITIAL PHASES	13
I.4.1	<i>Molecular replacement</i>	13
	The rotation function	13
	The translation function.....	14
I.4.2	<i>Isomorphous replacement</i>	14
I.4.3	<i>Anomalous diffraction</i>	15
	SAD phasing.....	16
	MAD phasing	17
I.4.4	<i>Direct methods</i>	18
I.5	STRUCTURE REFINEMENT	20
II	STRUCTURES OF ECHINOMYCIN IN COMPLEX WITH DOUBLE STRANDED DNAS	23
II.1	INTRODUCTION	23
II.1.1	<i>Chemical identity and biological effects</i>	23
II.1.2	<i>Aims of the project</i>	25
II.1.3	<i>Known effects on the DNA</i>	26
	Much to learn about it, much to learn from it.....	29
II.2	STRUCTURE OF ECHINOMYCIN-DNA COMPLEXES PERFORMING ONLY HOOGSTEEEN BASE PAIRING OUTSIDE THE INTERCALATION SITE	30
II.2.1	<i>Crystallization and data collection</i>	30
	The GE1 crystal	30
	The GE2 crystal	31
	On the cryoprotection protocol of GE2	32
	The AE1 crystal	33
	The CE crystal	33
	Crystal and drop properties.....	34
II.2.2	<i>Structure solution and refinement</i>	36
II.2.3	<i>Structure of the complexes</i>	37
	Crystal structure of GE1	37
	Structure of Echinomycin.....	38
	Structure of the DNA.....	40
	Interactions between Echinomycin and DNA	44
	Crystal structure of GE2	47
	Crystal structure of AE1	48
	Crystal packing in the P6 ₃ 22 crystals	49

Crystal structure of CE	50
Comparison of the structures	52
II.3 STRUCTURE OF AN ECHINOMYCIN-DNA COMPLEX PERFORMING BOTH HOOGSTEEEN AND WATSON-CRICK BASEPAIRS OUTSIDE THE INTERCALATION SITE	55
II.3.1 <i>Crystallization and data collection</i>	55
The AE2 crystal	55
II.3.2 <i>Structure solution and refinement of the AE2 crystal</i>	58
II.3.3 <i>The nature of the atoms in the metal site</i>	61
II.3.4 <i>Crystal structure of AE2</i>	66
Overall description of the structure	66
Base pairing scheme in AE2.....	67
Crystal packing	70
II.3.5 <i>Echinomycin structure in AE2</i>	72
II.3.6 <i>DNA structure in AE2</i>	73
II.3.7 <i>Discussion, biological implications and future prospects</i>	76
III CRYSTAL STRUCTURE OF AN ALTERNATING D-ALANYL, L-HOMOALANYL PNA	79
III.1 1. INTRODUCTION	79
III.1.1 <i>PNAs as nucleic acid analogs</i>	79
III.1.2 <i>2-aminoethyl-glycine PNAs</i>	79
III.1.3 <i>Alanyl PNAS</i>	81
III.1.4 <i>Aims of the project</i>	81
III.2 CRYSTALLIZATION ATTEMPTS ON ALANYL PNAS	82
Alanyl GGCGCC PNA	82
Alanyl GGCGCC-Lys PNA	83
Alanyl Lys-GGCGCC-Lys PNA	84
III.3 CRYSTAL STRUCTURE OF AN ALTERNATING D-ALANYL, L-HOMOALANYL PNA	84
III.3.1 <i>The PNA4c crystal form</i>	85
Data collection	87
Structure solution and refinement.....	87
Description of the structure of PNA4c	90
Structure analysis.....	92
Crystal packing	98
III.3.2 <i>Other crystal forms</i>	98
III.4 DISCUSSION AND FUTURE PERSPECTIVES	100
IV SUMMARY	103
IV.1 STRUCTURES OF ECHINOMYCIN IN COMPLEX WITH DOUBLE STRANDED DNAs.....	103
IV.2 CRYSTAL STRUCTURE OF AN ALTERNATING D-ALANYL, L-HOMOALANYL PNA	104
REFERENCES	105

I THEORETICAL BACKGROUND

I.1 The three-dimensional structure of duplex DNA

I.1.1 Chemical structure.

Most DNAs in biologically functional forms can be approximately described by the well known Watson-Crick model. Each chain of the DNA is a polymer composed of deoxyribonucleotides linked by phosphodiester bonds. The phosphate unit is linked to the 5' carbon of the riboses in a nucleotide and to the 3' carbon of the next nucleotide in the chain so that by consensus the chain runs in the 5'→3' direction of the riboses. The backbone is made of alternating phosphates and deoxyriboses while the nucleobases act as side chains linked to the C1' carbon of the sugars.

The nucleobases carry the genetic information and also link two DNA chains together to form a double helix. Each nucleotide of a chain forms hydrogen bonds to complementary nucleotides in the other chain, guanine normally binding to cytosine and adenine to thymine, with both chains running antiparallel to each other. Each basepair of this double polymer is rotated relative to the direction of the chains compared with the contiguous ones and overall the molecule adopts the form of a right handed double helix (*Figure I-1*). The bases are located inside and the phosphate groups outside, two grooves can be defined which run along the helix: a minor groove at the side of the basepairs where the phosphates and riboses are and a major groove at the opposite side of the basepairs.

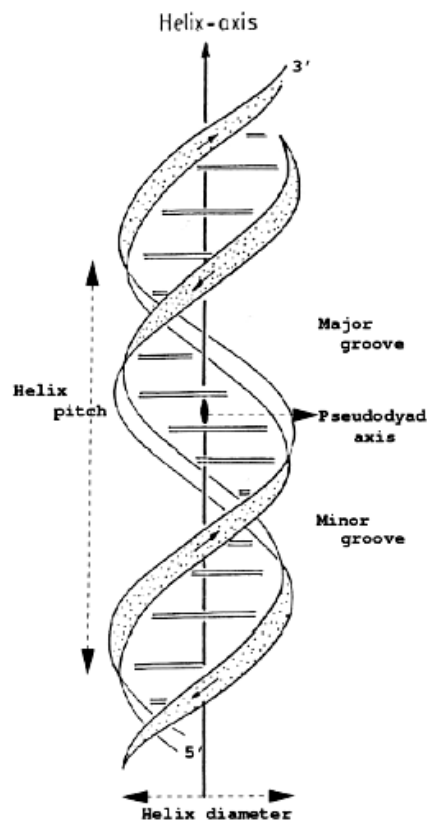


Figure I-1. Schematic view of the antiparallel double helix of DNA.

I.1.2 Long range three-dimensional order

DNA chains have lengths extending to millions of nucleotides and simplified models have been developed to describe their long range shape and properties. One such model is the worm-like chain (Bloomfield *et al.*, 2000). The DNA is described as a rod of limited elasticity, approximately linear in short stretches but with the capacity of bending slightly at each nucleotide step. In the long range, small accumulated distortions result in changes of direction, rolls and loops. An office elastic rubber band is a good analogy. The persistence length is defined as the length that an average polymer extends before the starting direction is inverted. This lies for duplex DNA in the range of 500\AA (Bloomfield *et al.*, 2000). Most proteins have a well defined tertiary structure as do RNAs with tRNAs as an example, DNAs normally lack this. The consequence is that most DNA molecules in the range of 100 bases do not have a definite structure although they can adopt one by interacting with other molecules, as exemplified by the nucleosome core particle (Richmond & Davey, 2003).

Crystallographic studies require molecular units with a well defined three-dimensional structure so that a crystal packing can arise and for this reason they are normally limited to

short DNAs stretches of less than 20 nucleotides in length (Neidle, 1999) if the structure is only or predominantly DNA, much longer segments can be studied if they are stabilized by interaction with proteins.

I.1.3 Base pairing schemes

The nucleobases in a nucleic acid molecule undergo relatively regular and unspecific base stacking interactions with adjacent nucleobases of the same strand. The nitrogen and oxygen atoms of the bases offer a variety of hydrogen bond donor and acceptor sites (*Figure I-2*) that originate the interactions with the second strand in duplex DNA and sometimes with a third or a fourth chain. A large variety of possible base pairing modes arises, including Watson-Crick, reverse Watson-Crick, Hoogsteen, reverse Hoogsteen, and wobble pairs with the only common point that the bases involved are approximately coplanar. Also non canonical basepairs, not involving C-G or A-T basepairs are found and some bases are involved in several binding schemes simultaneously. These base pairing schemes are probably the most important feature for describing the structure of a nucleic acid.

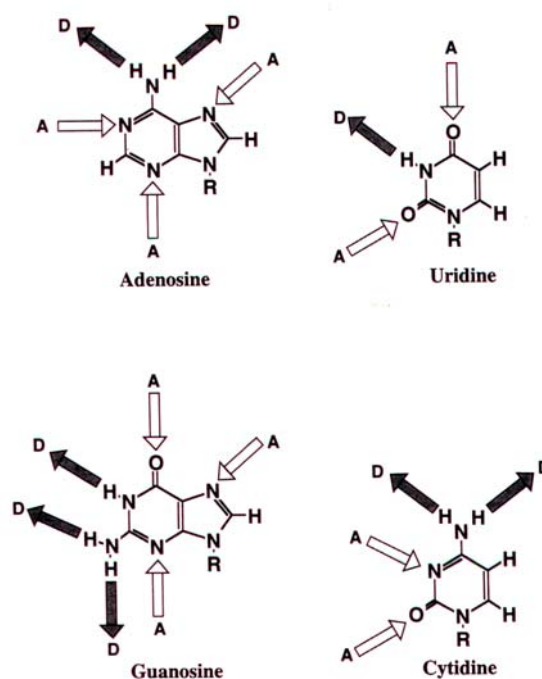


Figure I-2. Hydrogen bonding sites in standard nucleobases. Dark arrows indicate hydrogen donor sites, white arrows indicate acceptor sites. Thymine has the same scheme as uridine.

By far the most common of all hydrogen bonding schemes is the Watson-Crick base pairing (*Figure I-3*). This base complementarity is responsible for the transmission of genetic information, with guanine binding to cytosine by three and adenine to thymine by two hydrogen bonds. Other base pairing schemes are biologically less important and arise mostly in RNAs, which have a less regular structure.

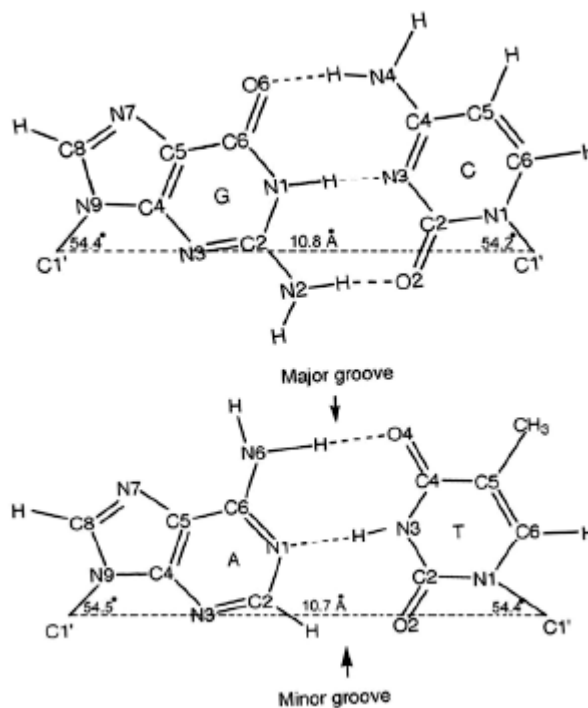


Figure I-3. The canonical Watson-Crick basepairs of DNA with their hydrogen bonds. The standard numbering scheme for nucleobases is also indicated in the figure.

The Hoogsteen base pairing scheme (*Figure I-4*) was observed for the first time in the crystal structure of a complex of adenine and thymine bases methylated in the position linking to the sugars in the nucleosides (Hoogsteen, 1959). In this mode of binding the purines are rotated 180° around the glycosidic bond and form two hydrogen bonds to the pyrimidines through N7 and O6 or N6. Protonation of cytosine at N3 is a prerequisite for C-G Hoogsteen basepairs. Its occurrence in oligonucleotide structures was first detected in the crystal structure of a hexamer complexed with the antibiotic Triostin A (Wang *et al.*, 1984) and was later also found in other distorted DNA structures. More recently, crystallographic studies have found Hoogsteen basepairs within undistorted B-DNA (Aishima *et al.*, 2002) and in the complex structure of a human polymerase specialized in replication through DNA lesions (Nair *et al.*, 2004). It probably plays a small but significant role in the biological functions of DNA molecules.

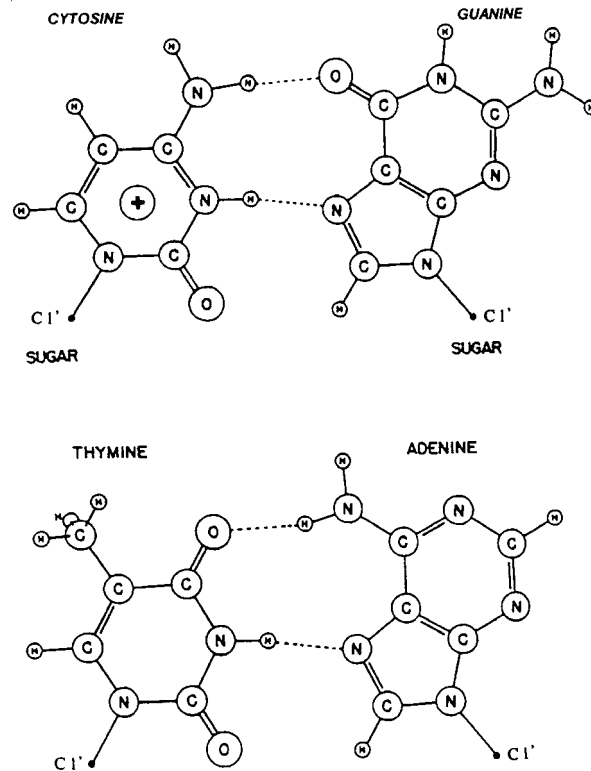


Figure I-4. Hydrogen bonding and relative base orientation in Hoogsteen basepairs.

I.1.4 Description of the conformation of DNA

Torsion angles

Several torsion angles provide for a considerable conformational flexibility within the phosphate-ribose backbone (Figure I-5).

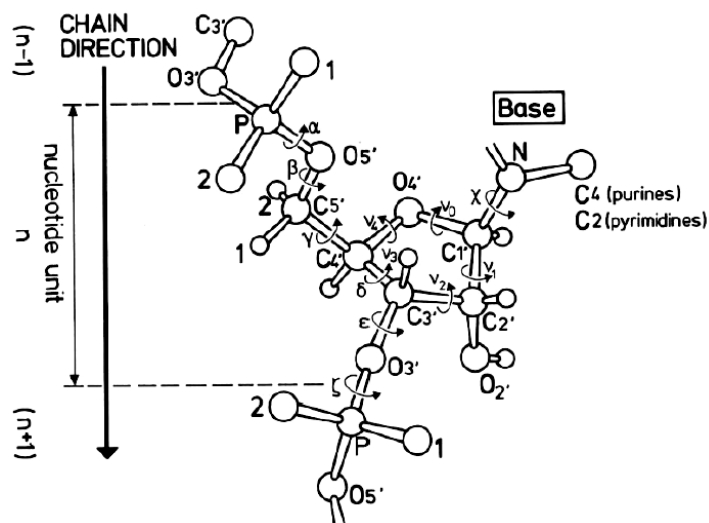


Figure I-5. Torsion angles in the DNA backbone.

The glycosidic torsion angle χ is defined as going from O4' in the riboses to C2 of pyrimidines or C4 of purines. It is preferentially in the *anti* conformation ($\chi=180^\circ\pm 90^\circ$) but changes to the *syn* conformation for the purines involved in Hoogsteen base pairing and shows a significant natural range of variation. It defines the relative angles of the base plane and the ribose plane.

The five torsion angles corresponding to the ribose (v_0 to v_4) are strongly correlated and can be substituted by a pseudorotation angle P and a maximum torsion angle Φ_m . A more intuitive representation of the sugar conformation can be achieved by indicating which atom of the ring is out of the plane of the other four, the side of the plane in which that atom lies is indicated by *endo*, on the side of the base, or *exo*, on the opposite side. C2'-*endo* is typical of B-DNA while C3'-*endo* is the most common sugar conformation in A-form RNAs and DNAs.

Basepairs and helical geometrical parameters

The relative position and orientation of both nucleobases in a basepair and of consecutive basepairs are important landmarks in every nucleic acid structure. Several different descriptions for such parameters have been used through time, and attention must be paid to the reference frame in which they are defined. In the literature different geometrical features can be found with the same name and vice versa and different values for such parameters have been reported depending on the chosen reference frame. In this work the parameter set derived from the *Tsukuba Workshop on Nucleic Acid Structure and Interactions* of 1999, recommended by the NDB (Berman *et al.*, 1992) and implemented in the program 3DNA (Lu & Olson, 2003) will be used.

16 parameters can be defined referring to rotations and translations between both bases in a basepair, two consecutive basepairs or the local helical axis (*Figure I-6*). The propeller value deviates significantly from the ideal coplanarity in most nucleotide structures and the values of twist and rise are also not expected to lie anywhere near 0° .

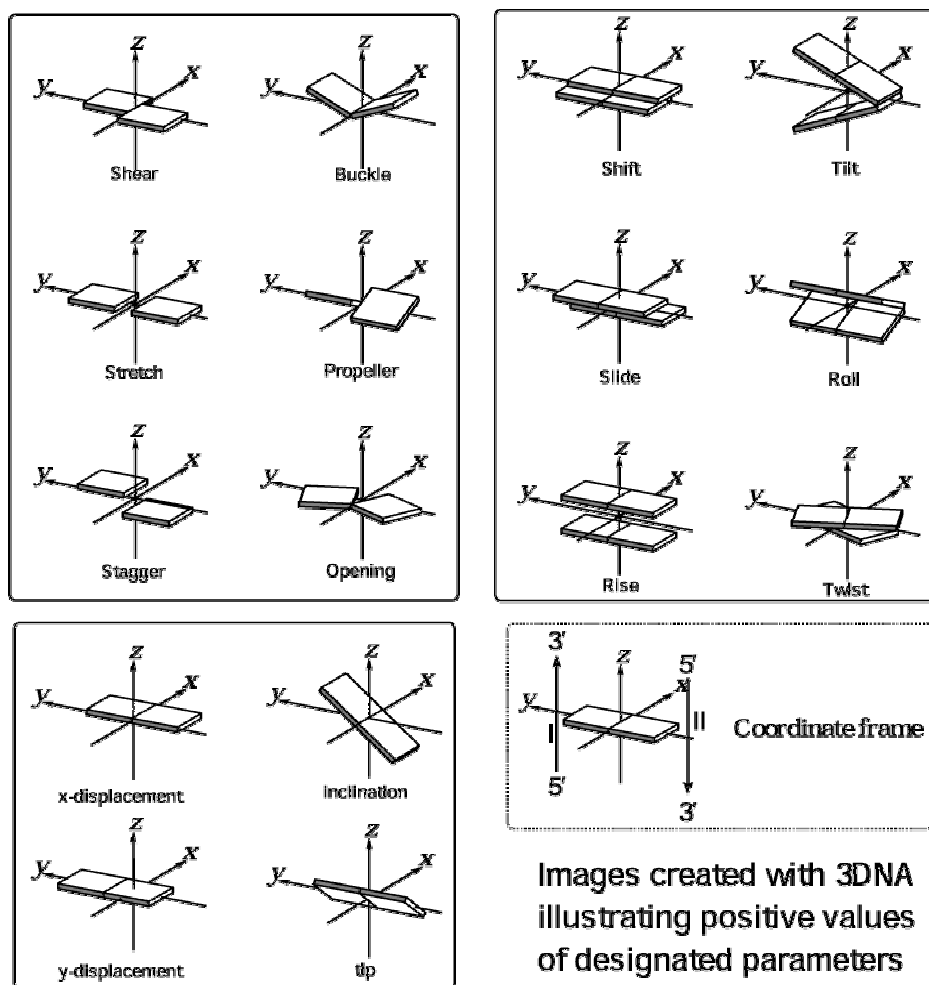


Figure I-6. Standard reference frame for the description of nucleic acid base-pair geometry as recommended in the NDB (Berman *et al.*, 1992).

These geometrical parameters are optimized for the structures of complementary Watson-Crick basepairs. Some of the parameter are strongly correlated, for example a positive buckle followed by a negative one will result in a too large apparent basepair rise.

As useful as this scheme is for standard basepairs, in the case of Hoogsteen basepairs two conditions concur to render most of those parameters unusable at least for comparison with standard values. First the coordinate system used in the definition of the geometrical parameters changes relative to the Watson-Crick frame since the imaginary line between C6 of pyrimidines and C8 of purines is affected by purine rotation into the *syn* conformation. Second the basepairs can be considered as being strongly distorted since they perform a different hydrogen bonding scheme.

I.2 Macromolecular crystallization

Crystals are a prerequisite for any monocrystal X-ray diffraction experiment. In the case of macromolecules, obtaining suitable crystals is often the rate limiting step in the full process of X-ray crystal structure determination. In some cases crystals are never obtained.

Three major techniques are commonly employed for the crystallization of proteins and nucleic acids:

a) Dialysis:

The sample containing the macromolecule is placed inside a dialysis cell. The cell is placed inside a solution containing the crystallization agents, which then diffuse through the membrane into the dialysis cell reducing the solubility of the macromolecule. The technique is slow and prone to errors, nowadays it is rarely used.

b) Batch crystallization:

A solution containing the macromolecule and the crystallization agents is prepared so that the molecule of interest becomes slightly supersaturated. The solution is isolated in a vessel and left undisturbed for a given period of time. Crystals grow until the solubility limit of the molecule is reached, at which point the process stops.

This is a method of choice for the crystallization of small molecules, but tends to work better when large amounts of the molecule are available, which is rarely the case with macromolecules. In recent times the microbatch technique is becoming increasingly popular. In the microbatch method, small batch solutions are prepared and covered with a layer of oil, frequently paraffin oil. The oil isolates each drop preventing evaporation and mixing. Using this technique drops with volumes in the submicroliter range can be prepared where successful crystallization is possible. This technique is specially amenable to automatization with robotic systems.

c) Vapour diffusion:

Two stock solutions, one containing the molecule of interest and another containing crystallization agents (salts, organic molecules, buffer, etc.) are prepared independently. A drop is made by mixing the two solutions in a given proportion so that the macromolecule is still undersaturated, drop sizes are typically in the range of microliters. The drop is either suspended (hanging drop) or set in a plastic well (sitting drop) over a reservoir of the solution containing the crystallization agents and the whole system is isolated. A scheme of the

hanging drop method is shown in *Figure I-7*. With time, water (and other volatiles) diffuse through the vapor phase into the reservoir solution, which has a lower vapor pressure. Thus, the concentrations in the drop increase until supersaturation is reached, at this point crystal growth can start. In vapor diffusion techniques every drop effectively samples a big range of concentrations over time and very small amounts of sample are necessary.

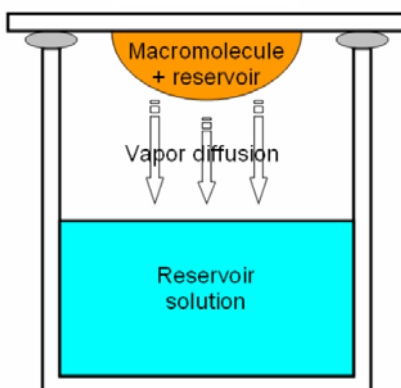


Figure I-7. Hanging drop setup with a cover slide. The concentrations in the drop increase over time. The relative size of the drop is greatly exaggerated.

The crystallization process is sometimes informally described as a slow, ordered precipitation. The molecules must interact in an ordered manner to produce a crystal lattice. In order to obtain crystals, a solution must be obtained in which the macromolecules or complexes of interest are supersaturated. A typical solubility diagram and the evolution of some crystallization experiments is shown in *Figure I-8*. At low degrees of supersaturation the system is typically metastable: it is supersaturated but no precipitation or crystallization takes place. If a crystal nucleus is already present, then the molecules will incorporate into it and the crystal will grow; the degree of supersaturation is reduced in the process. A certain degree of supersaturation is normally necessary to initiate nucleation, but a low supersaturation is necessary to obtain slow, ordered crystal growth. If the degree of supersaturation is too high, precipitation will occur.

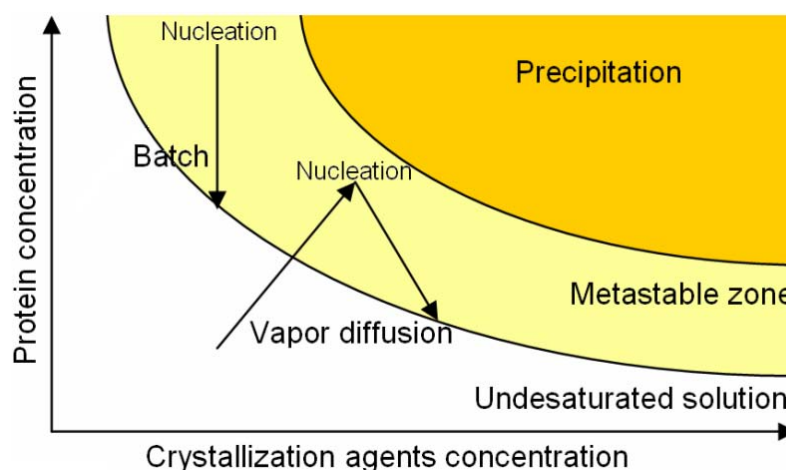


Figure I-8. Schematic solubility diagram of most macromolecules. In the yellow and orange areas the molecule of interest is supersaturated, either crystal growth or precipitation must occur. The evolution of the concentrations in a batch crystallization and in a vapour diffusion experiment are shown.

A large number of parameters control the outcome of a crystallization experiment. The most important parameters are the concentration of the macromolecule, salt types and concentrations, the presence of organic molecules, the pH and the temperature. Although the principles governing crystal growth are well known, the behavior of most macromolecules and the fact that all parameters are correlated results in a process of great complexity. In practice, a screening is made in which hundreds of conditions are tested. The conditions are chosen to sample many different crystallization agents and as much parameter space as possible. Once the first crystals are obtained, a process of optimization begins in which all parameters are systematically varied, this process being basically one of trial and error.

I.3 X-ray diffraction by macromolecules

Molecules are composed of atoms and have a well defined electron density. The electrons of the molecule can interact with electromagnetic waves in general and with X-ray radiation in particular, producing scattering. For most crystallographic applications only elastic (Thompson) scattering is considered, where there is no exchange of energy with the molecules. If exchange of energy takes place, the scattering is termed inelastic (Compton scattering), this is used in anomalous scattering techniques (I.4.2).

The scattering from an atom depends on its electron density distribution. It is described by the atomic scattering factor f_a :

$$f_a = \int_0^{\infty} \lambda r \rho_a(r) \frac{\sin\left(2\pi r \frac{2\sin\theta}{\lambda}\right)}{\sin\theta} dr$$

where λ is the wavelength of the X-rays, ρ_a is the electron density in a given position and θ is half of the diffraction angle.

The atomic scattering factor must be corrected in order to account for atomic and molecular vibrations:

$$f'_a = f_a e^{\left(-B \frac{\sin^2\theta}{\lambda^2}\right)}, \text{ where } B = 8\pi^2 u^2$$

where B is the atomic displacement parameter and u is the square mean displacement of the atom due to vibration. If the anisotropic nature of the vibrations is taken into account it is necessary to introduce a more complex correction factor with several (usually six) atomic displacement parameters.

In a crystal diffraction is only observed when constructive interference between the different unit cells takes place. The diffracted rays can thus be assigned Miller indices h, k and l in reference to the family of planes of the crystal that originates the diffraction. The diffracted rays are characterized by structure factors F:

$$F_{hkl} = \sum_{i=1}^n f'_a e^{[2\pi i(hx_i + ky_i + lz_i)]}$$

where the sum runs over the atoms in the unit cell, and

$$F_{hkl} = |F_{hkl}|e^{i\alpha_{hkl}}$$

$|F_{hkl}|$ is the amplitude of the diffracted wave and α_{hkl} is its phase, F is a complex number.

If the amplitudes (or intensities) and the phases of the reflections (diffracted beams) are known, an electron density map can be calculated:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{[-2\pi i(hx+ky+lz)+i\alpha_{hkl}]}$$

In a X-ray diffraction experiment only intensities are measured. The amplitude of the reflections is known but the phases are lost. Phases can be calculated once a model of the molecule is available but an electron density map is necessary in order to build an initial model. This is known as the crystallographic phase problem, methods for obtaining a initial set of experimental phases are described in the next section.

I.4 Methods for obtaining initial phases

I.4.1 Molecular replacement

If the structure of a similar molecule is available, the molecular replacement technique can be used to obtain an initial model. The molecule for which a model is available is called the search model, it can also be the molecule of interest if the structure of a different crystal form, an NMR model or a model from any other source are available.

The problem of solving the structure is then reduced to finding six variables, three rotational and three translational, which describe the transformation of the search model into the target structure. The correct solution can be identified by comparison of the observed and calculated structure factors or the corresponding Patterson maps. A full six dimensional search is (still) beyond the capabilities of modern computers. A possible solution is to divide the problem into two smaller tasks, performing separated rotational and translational searches. As both have only three parameters, an exhaustive search in a suitable grid is possible for each part.

The rotation function

The Patterson map of a crystal contains all interatomic vectors present in the structure. These interatomic vectors can be divided into two classes: intramolecular atomic vectors, between atoms of the same molecule, and intermolecular atomic vectors, originated between atoms of different molecules. The first class is composed predominantly of short vectors, whereas in the second class long vectors are the most frequent.

A rotation function can be defined as:

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

where P is the value of the Patterson map at point \mathbf{u} and \mathbf{C} is the rotation matrix that describes the rotation performed on the search model. The integration is performed over a volume U that contains predominantly intermolecular atomic vectors. The matrix \mathbf{C} is a function of three variables, for example the eulerian rotation angles. The value of the rotation function is maximal (peak overlap can cause it not to be) for the correct rotation, since the calculated Patterson map corresponds to the observed one except for the intermolecular vectors.

The translation function

Once the orientation of the search model is known the problem is reduced to finding the three translational parameters that will place the search model in the unit cell. Since search models are not identical to the target ones and the rotations are often inaccurately determined, the translation search can be difficult due to the presence of noise. Several translation functions have been proposed, one is the T function (Crowther & Blow, 1967):

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

where \mathbf{t} represents the translation vector and the integration is made over the entire Patterson maps. This function has a maximum when both Patterson maps superimpose, that is, when the rotation matrix and the translation vector are correctly determined.

Once the rotations and translations transforming the search model into the target model are determined, an initial model is available. It can be used to calculate an electron density map, which can then be improved, or input directly into the refinement process.

I.4.2 Isomorphous replacement

Isomorphous replacement was the technique used to solve the first protein structure, that of myoglobin (Kendrew *et al.*, 1958).

The Isomorphous Replacement method uses the difference between the diffraction intensities of various crystals to solve the phase problem. A native crystal and one or more derivative crystals containing one or more heavy atoms (obtained by cocrystallization or heavy atom soaking of native crystals) are necessary. In the ideal case those crystals are isomorphous, meaning that the only difference between them is the presence of the heavy atoms in the derivative, all the other atoms are in the same positions in all crystals.

In the first step the positions of the heavy atoms have to be determined. The relation between the structure factors of the native protein, \mathbf{F}_P , the heavy atoms, \mathbf{F}_H and the derivative \mathbf{F}_{PH} is as follows:

$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$$

where \mathbf{F}_H tends to be small in comparison with \mathbf{F}_P and \mathbf{F}_{PH} . A Patterson synthesis with coefficients $|\mathbf{F}_{HP}^2 - \mathbf{F}_P^2|$, the isomorphous difference Patterson, is dominated by the

contribution of the heavy atoms to the structure factors. Since the number of heavy atoms is small, their positions in the crystal can be deduced from that Patterson map.

A graphic solution in the case of Single Isomorphous Replacement, SIR, where only one heavy atom derivative is used, is illustrated in *Figure I-9*. In the SIR case, an ambiguity between two phases results, the centroid phase can be accepted as an approximation to the solution. In the multiple isomorphous replacement technique, MIR, where several heavy atom derivatives are used, the phases can be determined unambiguously except for the experimental errors.

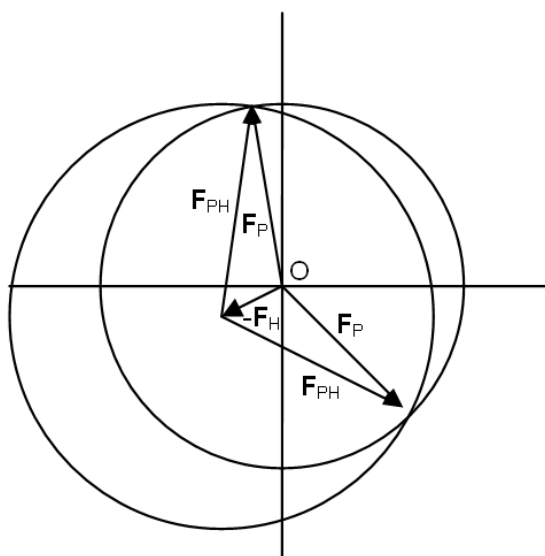


Figure I-9. Argand diagram for SIR. Only F_H and the modulus of F_P and F_{PH} , the radius of the circles, are known. The intersection of the circles determines the two possible values of the phase.

I.4.3 Anomalous diffraction

If an X-ray photon has an energy equal or higher than that of the absorption edge of an atom in the crystal, it can get absorbed by the atom. The photon may be reemitted with a phase shift, Friedel's law breaks down and anomalous diffraction occurs. The scattering factor of an atom for which anomalous dispersion occurs is:

$$f_{\text{total}}(\theta, \lambda) = f(\theta) + f'(\lambda) + if''(\lambda)$$

where the imaginary part is responsible for the phase shift. The situation is illustrated graphically in *Figure I-10*.

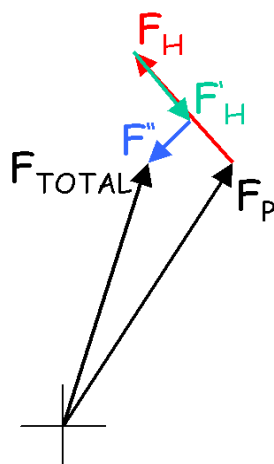


Figure I-10. Argand diagram for one reflection with anomalous scattering.

f'' has appreciable values only in the vicinity of the absorption edge and f''' only at energies equal or higher than the absorption edge. Only elements in the fourth or lower periods of the periodic table have absorption edges at wavelengths achievable at synchrotron beamlines used in macromolecular crystallography (*Figure I-11*). Amenable heavy atoms may be naturally present in the structure or added by a heavy atom soak.

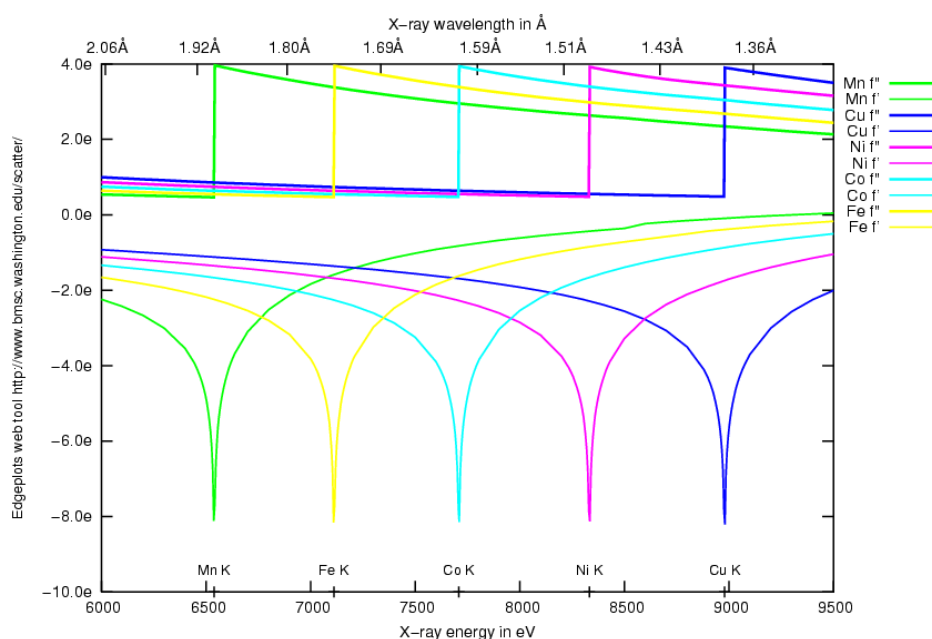


Figure I-11. Theoretical absorption edges and values of f' and f'' for some transition metals.

SAD phasing

The SAD method used data collected at a single wavelength where anomalous diffraction is significant. The Bijvoet differences, $\Delta\mathbf{F} = |\mathbf{F}_+| - |\mathbf{F}_-|$ are used to locate the positions of the

anomalous scatterers in the cell by inspection of the anomalous difference Patterson map, by direct methods on the anomalous differences or by other means. The origin of the anomalous differences is illustrated in *Figure I-12*. The value of $|\Delta\mathbf{F}|$ is proportional to f'' . It is not possible to derive exact native phases with this method, but an estimation can be made when $|\Delta\mathbf{F}|$ is large. Centroid phases can be derived in a similar way as in the SIR case. The programs SHELXC/D/E use a different paradigm, where phase shifts of 90° or 270° between the heavy atom phases and the total phases are derived from the sign of $\Delta\mathbf{F}$ and weights are assigned as a function of $|\Delta\mathbf{F}|$. One single crystal is used in the procedure and thus lack of isomorphism is not a problem. The resulting phases are usually of poor quality and need to be improved by density modification to render an interpretable electron density map.

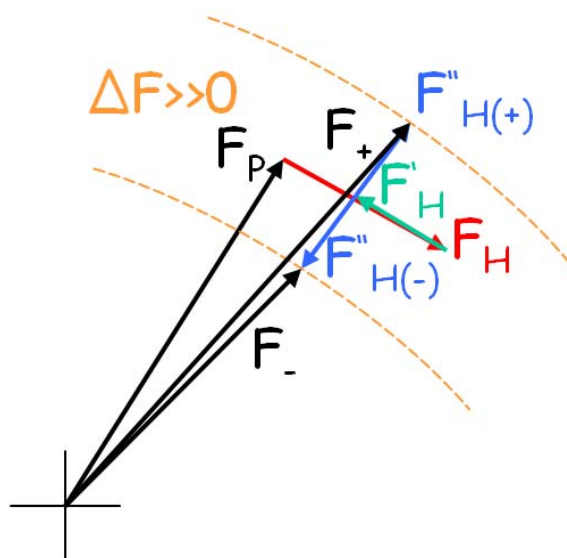


Figure I-12. Argand diagram illustrating the origin of the Bijvoet differences. \mathbf{F}_+ and its component vectors are actually mirrored about the horizontal (real) axis. In this case a phase shift of approximately 90° between \mathbf{F}_P and \mathbf{F}_H can be derived.

MAD phasing

Data at several wavelengths are collected in a MAD experiment, preferably all using the same crystal. Normally datasets are collected at the absorption peak, at the inflection point and at a high energy remote or a low energy remote wavelengths, far from the absorption edge; but at least mathematically, two datasets suffice. Both anomalous differences ($\Delta\mathbf{F}$) and dispersive differences are available, and an exact estimation (limited by the experimental errors) of the phases of the structure factors is possible. The dispersive differences originate in

the different values of f' at different wavelengths and are maximum between the peak and the inflection point.

1.4.4 Direct methods

The term direct methods is usually used for a procedure in which a structure is solved by a search for all atoms in the structure without any detailed knowledge of it. Direct methods are routinely used to solve small molecule structures in a few seconds, but become less powerful as the number of atoms increases. A necessary condition to solve a structure by direct methods (with some exceptions) is the availability of complete diffraction data to a resolution of at least 1.2 Å. This is probably related to the fact that only at that resolution the atomic peaks are resolved in the electron density map. For direct methods calculations the structure factors are normalized in resolution shells. The normalized structure factors, designated \mathbf{E}_{hkl} , correspond to the structure factors of a molecule with point atoms.

It is possible to derive relationships between the phases of different reflections from very simple assumptions, like that the electron density is never negative or that the atoms are randomly distributed in the asymmetric unit. The most important of those relationships is the triple phase relation:

$$\varphi_{\mathbf{h}} = \varphi_{\mathbf{h}'} + \varphi_{\mathbf{h}-\mathbf{h}'}$$

where \mathbf{h} stands for the miller indices of the reflections, h , k and l .

The tangent formula (Karle & Hauptmann, 1956), or other formulas derived from it are usually used in direct methods programs to refine the estimated phases:

$$\tan \varphi_{\mathbf{h}} = \frac{\sum_{\mathbf{h}'} |E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'}| \sin(\varphi_{\mathbf{h}'} + \varphi_{\mathbf{h}-\mathbf{h}'})}{\sum_{\mathbf{h}'} |E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'}| \cos(\varphi_{\mathbf{h}'} + \varphi_{\mathbf{h}-\mathbf{h}'})}$$

Also other relationships exist. Typically, random sets of phases are calculated and refined using the tangent or other formulas to achieve a stable solution. The resulting phases are used to calculate an electron density map which is peak-searched to produce a list of atomic positions. This philosophy, termed conventional direct methods, typically renders solutions for molecules with less than 100 atoms, but rarely succeeds where the number of atoms is greater than 200.

A different philosophy towards direct methods is that of dual space recycling, initiated by the program Shake and Bake (Miller *et al.*, 1993). A starting set of atoms (random or better) is generated and the corresponding structure factors calculated. Those phases are refined in reciprocal space and used to calculate an electron density map, which is peak searched to generate an optimized list of atoms in real space. With the improved list of atoms, new structure factors are calculated and the process is repeated over several cycles (*Figure I-13*).

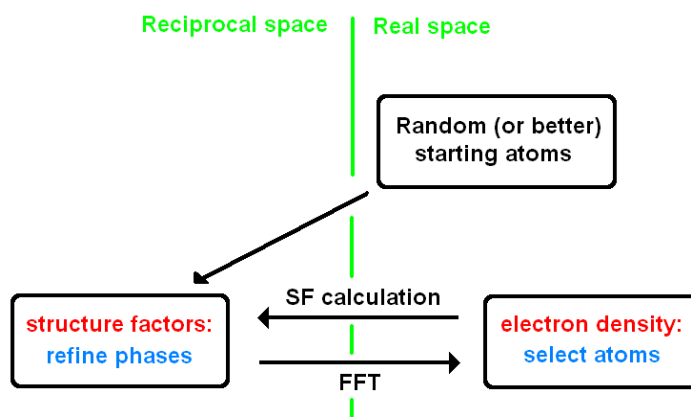


Figure I-13. The Shake and Bake procedure as implemented in SHELXD.

Only the strongest reflections of each resolution shell are used in the calculation. Correlation coefficients can be used to identify the best solution over many tries. Structures with more than 1000 atoms can be solved with this method, which is also amenable to obtain the heavy atom substructure in SIR, MIR, SAD and MAD.

1.5 Structure refinement

Initial models, either obtained by molecular replacement or built in an electron density map, are only approximate descriptions of the structure. In order to obtain a precise molecular structure, it is necessary to refine the model against the experimental intensities or amplitudes of the structure factors. Both the spatial coordinates and the displacement parameters of the atoms are refined. During the refinement, the structural parameters are varied to optimize the agreement between the calculated and the observed structure factors. Several mathematical formalisms exist, the most important of which are maximum likelihood refinement and least squares refinement. In the SHELXL program, used for the refinements in this thesis, the function that is actually minimized is:

$$Q = \sum_{hkl} w_{hkl} \left(|F_o|^2 - |F_c|^2 \right)^2$$

For computational speed and mathematical stability reasons, a conjugate gradient least squares procedure is normally used with macromolecules.

Attention must be paid to the data to parameter ratio during the refinement. The refinement is only stable if the number of data is much higher than the number of parameters, that is, the minimization problem has to be overdetermined. The number of data collected from macromolecular crystals is strongly limited by the resolution limit of the crystals. If too many parameters are fitted, they will adopt unreasonable values to compensate for experimental errors, inaccuracies of the physical model of the molecules, etc. If anisotropic displacement parameters are modeled for the atoms, the number of parameters per atom increases from four (for an isotropic model) to nine. Restraints like ideal bond distances and angles, coplanarity of the atoms in aromatic groups or a smooth change of the ADPs, must be included in the refinement. In SHELXL, restraints are treated as extra observations and assigned a suitable weight, thus improving the effective data to parameter ratio. The introduction of riding hydrogens, whose position can be deduced from that of the heavier atoms, does not introduce any new parameters in the refinement.

The agreement between the model and the experimental data is normally measured with the crystallographic R factor:

$$R = \frac{\sum_{hkl} ||F_o| - |F_c||}{\sum_{hkl} |F_o|}$$

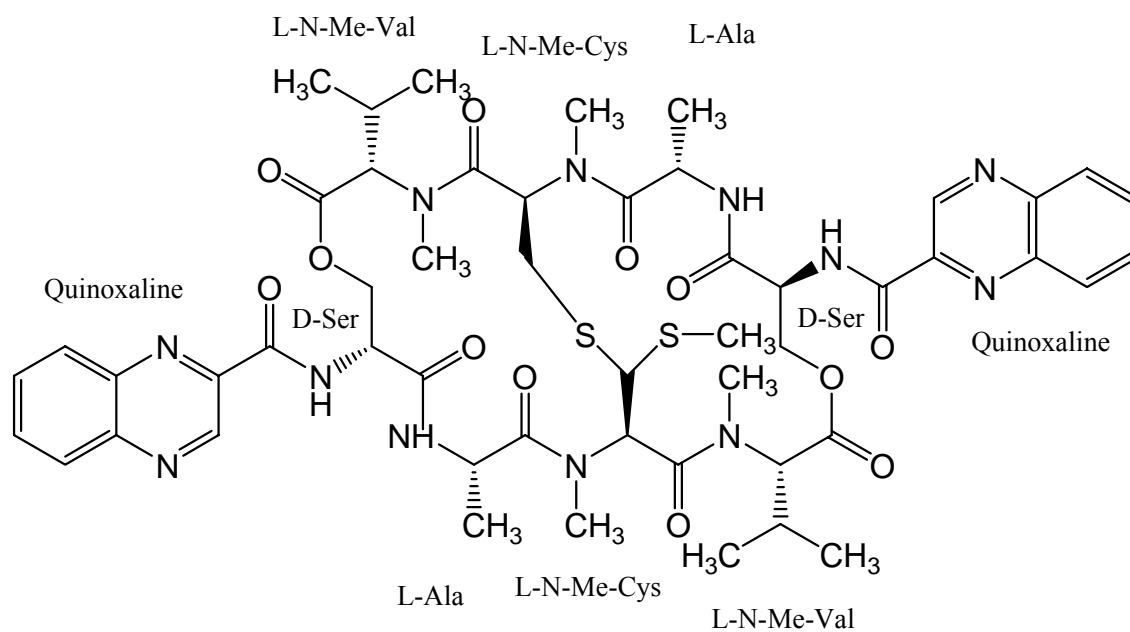
An additional R factor is employed in macromolecules to monitor the degree of overfitting of the data: The free R-factor (Brünger, 1992). A small percentage of reflections are omitted from the refinement process and their agreement with the refined data (defined as for the conventional R factor) monitored. A strong deviation between R and R_{free} indicates overfitting of the model.

II STRUCTURES OF ECHINOMYCIN IN COMPLEX WITH DOUBLE STRANDED DNAs

II.1 Introduction

II.1.1 Chemical identity and biological effects

Echinomycin is a depsipeptide antibiotic from streptomycetes. Its chemical structure is shown in *Figure II-1*. It has a rigid bicyclic backbone composed mostly of modified or chirally inverted aminoacids. The sequence is a repeat of four aminoacids connected through two ester linkages and a thioacetal bridge. The nitrogens of the D-serines bind quinoxaline bases in such an orientation that double intercalation into DNA enclosing two basepairs is possible.



Echinomycin

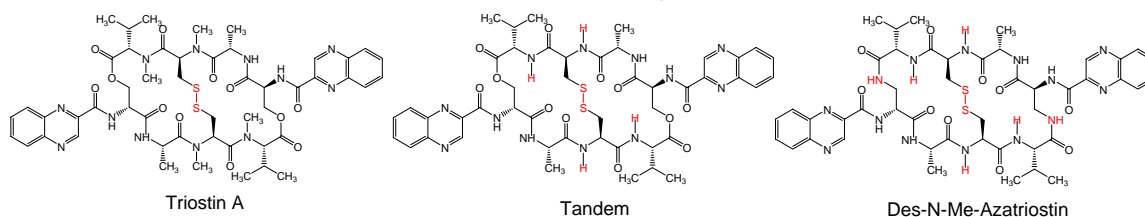


Figure II-1. Chemical formulae of some quinoxaline antibiotics. The differences with Echinomycin are highlighted in red.

Echinomycin was the first bisintercalation discovered (Waring and Wakelin, 1974) and stands today as the canonical representative of the quinoxaline antibiotics that bind to DNA by bisintercalation. Other members of the family are shown in *Figure II-1*.

Echinomycin binds to double stranded DNA both *in vitro* (Waring & Wakelin, 1974; Wakelin & Waring, 1976; Waring, 1993) and *in vivo* (May *et al.*, 2004) and interferes with replication and transcription (Ward *et al.*, 1965; Sato *et al.*, 1967) as well as with nucleosome structure (Leslie & Fox, 2002) and chromatin decondensation (May *et al.*, 2004). Echinomycin and many chemical modifications of it have been and are currently undergoing clinical trials as anticancer agents (Park *et al.* 2004).

The family of antibiotics has common structural features. Triostin A (Sheldrick *et al.*, 1995) differs from Echinomycin only in the bridge, Equinomycin 2QN (Sheldrick *et al.* 1995) has the same backbone but different bases. Echinomycin itself (already hinted in Ughetto *et al.*, 1985) has a rectangular, rigid, relatively planar backbone with the bases protruding away from opposite corners to the same side of the backbone. As a result, both chromophores are preoriented so that they may lay parallel to each other, perpendicular to the plane of the backbone and approximately 10 Å apart, an energetically favorable arrangement for bisintercalation so that no void space is left towards the two enclosed basepairs or to the minor groove.

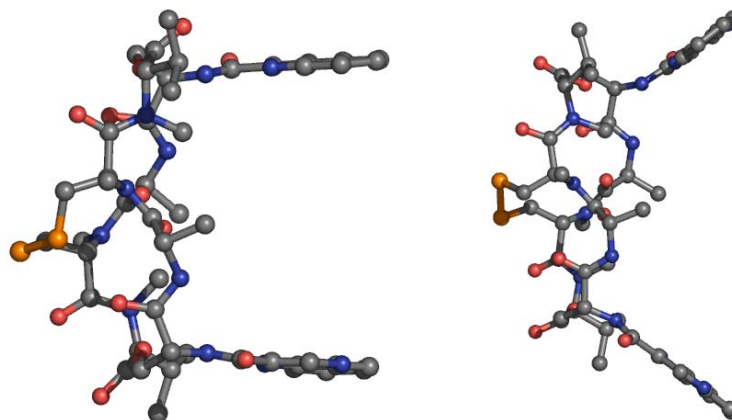


Figure II-2. Triostin A in the structure of the complex with CGTACG (left) and crystallized alone (right). The bases are more open in the latest and undergo interactions with their symmetry equivalents emulating partial intercalation in the crystal.

The preorientation of the bases is still far from being rigid as shown in the DNA bonded and standalone crystal structures of Triostin A (Sheldrick *et al.*, 1995; Wang *et al.*, 1984) (*Figure II-2*) and can even strongly deviate from the parallel conformation if that leads to

extra hydrogen bonds like in Echinomycin 2QN. The structure of TANDEM (Hossain *et al.*, 1982) (*Figure II-3*) is somewhat different with the valyl amides involved in intramolecular hydrogen bonds with the alanyl carbonyls bringing the quinoxalines out of position. A rearrangement of the structure must still be possible since TANDEM readily bisintercalates duplex DNA (Low, Olsen & Waring 1984).

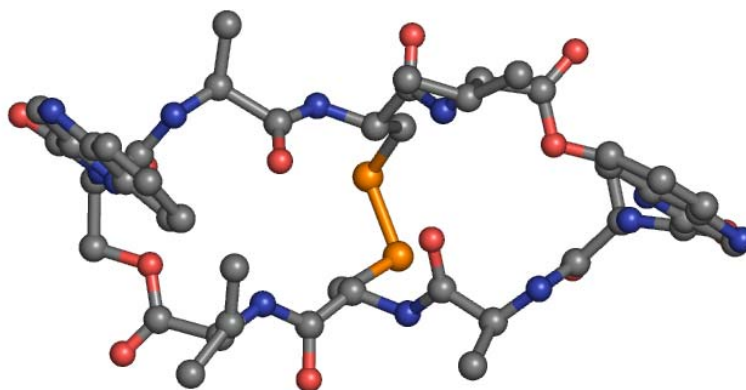


Figure II-3. Crystal structure of TANDEM.

II.1.2 Aims of the project

The crystallographic study on the structure and interactions of Echinomycin-DNA complexes here presented is part of a collaboration with Prof. Ulf Diederchsen from the Institute of Organic Chemistry of the University of Göttingen. Chemically modified Quinoxaline antibiotics should be tailored for new modes of DNA recognition taking advantage of the preorganized structure of the family. Not only bisintercalation but also intercalation at abasic positions and binding in the major groove of duplex DNA without intercalation can conceivably be induced, with potential for recognition of DNA sequences and defects and the consequent diagnostic and therapeutic applications.

Cocrystallization trials of modified azatriostins with short complementary DNAs are already ongoing but they are still not in a sufficiently advanced state as to include them in this report.

Interest on the work arises not only from its role in the formerly cited project but also as a takeover on a series of crystal structures dating from the decade of 1980. Two structural determinations of complexes of Triostin A with a duplex DNA hexamer and with a duplex DNA octamer (Wang *et al.*, 1984 and Waring & Wakelin, 1986) showed for the first time Hoogsteen basepairs in DNAs, sparking a burst of research activity on the causes and effects of this mode of base pairing. A further article (Ughetto *et al.*, 1985) reports qualitatively and

semiquantitatively on the structure of a DNA-Echinomycin complex but the coordinates of this structure were neither published nor deposited and are thus unavailable to the scientific community. Although there was considerable research activity on this kind of interaction and although different Echinomycin related molecules have found potential application in anticancer therapy, no further crystallographic studies were published ever since.

It is thus also aim of this work to take advantage of the progress of the last 20 years in crystallographic methods, increasing the too scarce range of structures and sequences of this family of compounds for which detailed structures of the interaction are available. Specially low temperature synchrotron data collection and refinement techniques have improved, making possible to obtain a precise structure of the original complex of Echinomycin with (CGTACG)₂ as well as new complexes with different DNA sequences.

Echinomycin alone has resisted all crystallization attempts for more that 20 years. During this project it has been attempted to crystallize Echinomycin by vapour diffusion methods from mixtures of water with different alcohols and organic solvents and also by using modern, protein oriented, crystallization screens from Hampton Research and Jena Bioscience, but no crystals were obtained. Success was met upon cocrystallization with short DNA duplexes rendering the structures that will be presented further on in this work.

II.1.3 Known effects on the DNA

Echinomycin was shown to be a bisintercalator to DNA (Waring & Wakelin, 1974) even before the correct chemical formula was available (Dell *et al.*, 1975).

Subsequently, footprinting studies shed light on the sequence selectivity of Echinomycin. In the footprinting technique, bound DNA is left to react partially with endonucleases or some other chemical probe. The resulting mixtures are then separated by electrophoresis resulting in a series of bands for DNA fragments of different length. The absence of fragments of given sizes is indicative of DNA protection by ligand binding.

Echinomycin was found to bind mostly around CG steps with a preference for AT basepairs in the surrounding sites (Low, Drew & Waring, 1984), although those are not the only binding sequences. Triostin A has the same preferential CG binding site, but TANDEM binds better around TA steps showing a role of the peptidic backbone in sequence recognition.

When the crystallographic structures of the complex with Triostin A became available the concept that DNA could only perform Watson-Crick base pairing broke down. There was indeed bisintercalation around the CG steps, whether they were terminal or not, but the

surrounding bases were paired in the Hoogsteen conformation instead of the usual Watson-Crick. This was true for both kind of basepairs, C-G and A-T (Quigley *et al.*, 1986) with the purines rotated to the *syn* conformation and hydrogen bonding through N7 rather than N1. This was the first ever observation of base pairing other than Watson-Crick in a presumably functional DNA, what since then has been shown to happen in other systems (Nair *et al.*, 2004).

The crystal structures of the complexes showed that hydrogen bonds formed between the guanines inside the bisintercalation site and the alanines of Echinomycin and Triostin A are the reason for the observed sequence selectivity. There is a high shape complementarity between the peptidic backbone of the antibiotics and the minor groove of the DNA leading to extensive Van der Waals interactions that are probably responsible for most of the binding energy. Hoogsteen base pairing places the C1' carbons of the sugars only some 8.6 Å apart instead of 10.5 Å for Watson-Crick pairing, leading to a better fit of the shape of both molecules. It was reasoned that this could be the reason for the change to the Hoogsteen conformation. All three structures showed the same interaction, and it was assumed that this was the general binding mode for Echinomycin and Triostin A. This was also in agreement with the preference for A-T basepairs surrounding the intercalation site. The formation of a G-C Hoogsteen basepair requires protonation of the Cytosine at N3 and involves the loss of a hydrogen bond relative to the Watson-Crick mode.

Further studies were made to confirm this effect but the result was quite the opposite. Footprinting studies found that the hypersensitivity effect upon Echinomycin binding on adenines is similar to that caused by other intercalators from which Hoogsteen base pairing was not suspected (Jeppesen & Nielsen, 1988). Portugal *et al.* (1988) extended the result to the guanines: the expected protection of guanine N7 by Hoogsteen base pairing was not found nor were any pH effects noticeable as expected for the needed protonation of the cytosines. Unwinding of the bases surrounding the binding site was proposed to explain the hyperreactivity, in some cases extending several basepairs away from the bisintercalation site. It was also found that the presence of 7-deaza-2'-deoxy-adenosine, incapable of making Hoogsteen basepairs, did not alter the binding properties contiguous to CG sites (McLean *et al.*, 1989). A detailed study by Sayers & Waring (1993) also failed to detect any effect from the substitution of guanines for 7-deaza-2'-deoxy-guanines. It must be noted that the two last chemical substitutions prevent only the formation of a second hydrogen bond with the complementary base, not the rotation of the purines into the *syn* conformation itself.

The structural problem of the presence or absence of Hoogsteen basepairs in the complexes of DNA with quinoxaline antibiotics has also been tackled by NMR. Even if this has not always led to a three-dimensional structure, the Hoogsteen versus Watson-Crick question can be elucidated through the *syn* or *anti* conformation of the purines. Gilbert & Feigon (1991) showed that the complex with (ACGTACGT)₂ at 274 K has all four Hoogsteen basepairs flanking the bisintercalation sites in accordance with the crystal structure of the complex with Triostin A. At higher temperatures the two central basepairs are lost or in equilibrium with the Watson-Crick mode. For the complex with (TCGATCGA)₂ all basepairs are Watson-Crick. In the complex between Echinomycin and (ACGTATACGT)₂ the terminal basepairs were Hoogsteen but the internal ones were now standard Watson-Crick although destabilized relative to free DNA (Gilbert & Feigon, 1992). The complex with (ACGT)₂ also shows Hoogsteen base pairing but not the one with (TCGA)₂ (Gao & Patel, 1988). NMR studies on Triostin A (Address & Feigon, 1994) and two other related Quinomycin antibiotics (Searle, 1994; Chen & Patel, 1995) suggest that Hoogsteen basepairs are only favored if the bases are terminal, but not any more when they are internal to the oligonucleotides. The mass of NMR evidence thus suggests that the Hoogsteen basepairs are possible but not necessarily favored under normal biological conditions. Furthermore, the NMR studies indicate that the structural changes in the DNA extend some bases away from the CG step even if the Hoogsteen binding mode does not, in accordance with the footprinting evidence.

Molecular dynamics studies have been made for all four possible tetramers (Gallego *et al.*, 1993; Gallego *et al.*, 1994) and a set of hexamers including modified nucleobases (Gallego *et al.* 1994b). The stacking interactions are responsible for approximately 50% of the interaction energy and the dipolar moments of the external basepairs seem to modulate the sequence preferences in the surrounding bases. (ACGT)₂ and (GCCG)₂ have a most stable arrangement in the Hoogsteen conformation while (TCGA)₂ and (CCGG)₂ would be destabilized and adopt the Watson-Crick conformation. The studies remain within the range of short sequences and thus the influence of the surrounding basepairs on the structure of the DNA itself has not been taken into account.

Atomic force microscopy has also been applied to the problem by studying the contour length of DNA fragments complexed and uncomplexed with Echinomycin (Tseng *et al.*, 2004). The results of those direct observation confirm the existence of other binding sites apart of CG steps and show doubtlessly that the presence of the 2-amino group of the purines at least in one chain is a necessary condition for Echinomycin binding.

Much to learn about it, much to learn from it

All the research sparked after that first detection of a Hoogsteen basepair in DNA has contributed much to our knowledge of DNA structure, its flexibility and its interactions with ligands, specially on the recognition of specific sequences by small molecule drugs. The availability of more three-dimensional detailed structures of such complexes can only provide new insights on recognition and modulation of DNA structure and can be of great help to the design of new drugs and molecules for the clinical use.

II.2 Structure of Echinomycin-DNA complexes performing only Hoogsteen base pairing outside the intercalation site

II.2.1 Crystallization and data collection

Oligonucleotides with the self-complementary sequences GCGTACGC, ACGTACGT and CGTACG were purchased already HPLC purified from Carl-Roth GmbH and were used in all experiments without further purification. Lyophilized Echinomycin from Sigma-Aldrich (code E-4392) was used for all crystallization experiments without further purification. Echinomycin is not water soluble so it was dissolved in methanol. The DNAs were dissolved in water and then mixed with the methanolic Echinomycin solutions. Stock solutions for all three cases had 50% water and 50% methanol and remained clear without signs of precipitation. For the Echinomycin₂-(GCGTACGC)₂ and Echinomycin₂-(CGTACG)₂ complexes a solution with 4 mg/ml DNA and a 1.05 Echinomycin:DNA molar ratio were used. For the Echinomycin₂-(ACGTACGT)₂ complex the Echinomycin:DNA molar ratio was set to 1.1 and stock solutions with DNA concentrations of 4 mg/ml and 0.5 mg/ml were used.

The GE1 crystal

The crystal diffracting to the highest resolution was obtained by the hanging drop method. 2 μ l of the GCGTACGC-Echinomycin stock solution were mixed with 2 μ l of a reservoir solution consisting of 1.1 M Li₂SO₄, 0.05 M MgCl₂, 0.1 M NaAc buffer at pH 4.5 and 0.02 M spermine tetrachloride and the mixture drop was suspended over the reservoir solution. A 0.85 mm long and 0.08 mm thick hexagonal needle grew at 293 K (*Figure II-4*).

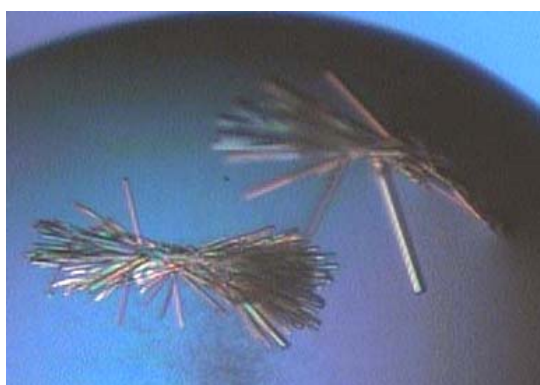


Figure II-4. Microscope picture of typical hexagonal needles and needle clusters resulting from the lithium sulfate based conditions.

For cryoprotection a cryosalt (Rubinson *et al.*, 2000) was chosen as the crystal had a relatively high salt content. The crystal was soaked in a solution consisting of 7 M Li₂SO₄, 0.025 M MgCl₂ and 0.05 M MES buffer at pH 6 for short time and frozen by plunging it into liquid nitrogen. Data collection was carried out at 100 K at beamline X13 at EMBL/DESY, Hamburg. Integration was performed with Denzo (Otwinowski & Minor, 1997) and scaling with SADABS (Bruker Nonius, 2002) to a resolution of 1.1 Å. The crystal belongs to space group P6₃22 as the crystal of the complex of Triostin A with the same DNA (Wang *et al.* 1986) but the unit cell is somewhat smaller. From now on this crystal will be referred to as “GE1”. Unit cell and data collection statistics are listed in *Table II-1*.

Crystal	GE1	GE2	CE	AE
Sequence	(GCGTACGC) ₂	(GCGTACGC) ₂	(CGTACG) ₂	(ACGTACGT) ₂
Space group	P6 ₃ 22	P6 ₃ 22	C2	P6 ₃ 22
Unit cell (Å)	a = b = 39.37, c = 79.73	a = b = 39.18, c = 79.89	a = 29.51, b = 39.37, c = 79.73, β = 114.99°	a = b = 39.93, c = 80.10
Wavelength (Å)	0.8126	0.9000	0.9000	0.9000
Resolution (Å)	1.10 (1.20)	1.26 (1.40)	1.40 (1.50)	1.50 (1.60)
Total reflections	253373	164952	42417	139625
Unique reflections	15165	10337	10857	6499
Completeness (%)	97.2 (95.7)	99.5 (99.4)	97.8 (94.3)	99.0 (98.4)
R _{int} (%)	4.93 (24.25)	4.95 (35.32)	5.46 (27.5)	4.24 (42.25)
I/σ(I)	30.01 (11.47)	28.25 (6.70)	12.72 (4.87)	39.11 (9.03)
Data*/restraints/ parameters	14680/2074/ 2979	9312/5120/ 3121	9768/6395/ 4286	5928/5061/ 3065
R (%)	14.59	17.89	18.29	19.76
R _{free} (%)	16.81	22.09	23.50	23.74

Table II-1. Dataset and refinement statistics for crystals GE1, GE2, AE1 and CE. In parenthesis the values for the high resolution shells, whose lower limits are shown in parenthesis in the resolution line. *The number of data listed is the number of reflection used in the refinement, that is, the working set, the rest to the total number of reflections is the test set.

The GE2 crystal

A second crystal of the same complex was obtained from radically different conditions, with the high salt concentration substituted by a diluted organic molecule. 1.3 µl of stock solution were mixed with 2.6 µl of a reservoir solution consisting of 9% PEG 550

monomethylether, 0.05 M MgCl₂, 0.1 M sodium acetate buffer at pH 4.5 and 0.01 M spermine tetrachloride. A 0.2 mm wide and 0.1 mm thick hexagonal prism grew in the hanging drop at 293 K (*Figure II-5*).

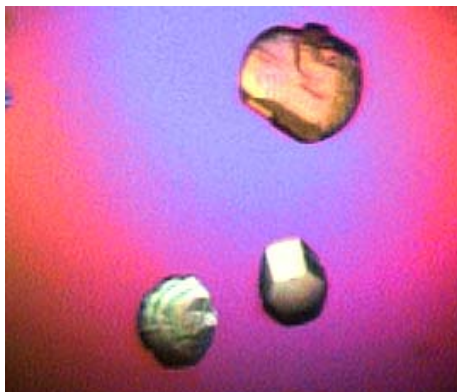


Figure II-5. The hexagonal prism from the isopropanol based conditions. The crystal habit is different from GE1, where needles were the typical result.

No cryoprotectant was used, instead the crystal was taken out of the drop in a nylon loop, the droplet in it let to evaporate in the air for approximately five seconds and frozen directly in the 100 K nitrogen stream of the diffractometer. Data collection was carried at 100K at the Protein Structure Factory beamline BL14.1 at BESSY, Berlin, on a MAR CCD detector and integrated and scaled with XDS (Kabsch, 1993) to a resolution of 1.26 Å. This crystal will from now on be referred to as “GE2”. The unit cell and space group were similar to those of GE1 and it can only be considered to be the same crystal form, but taking into account the very different crystallization conditions it will be regarded as a different crystal for the sake of the discussion. Cell and data collection statistics for GE2 are listed in *Table II-1*.

On the cryoprotection protocol of GE2

The 9% concentration of PEG 550 monomethylether in the starting droplet is far too small for effective cryoprotection. The final concentration must have been in the range of at least 30%, as is typically necessary with glycerol. The droplet taken in a small nylon loop can have a very small volume and consequently a big surface/volume ratio allowing for fast concentration rates before freezing. More typical cryoprotection schemes with extra glycerol or PEG were tried, but the diffraction from those experiments was of poorer quality. It can be speculated that in the latest case the crystals spend longer times in concentrated solutions of alcohols or other agents with possible damaging effects. Although the empirical cryoprotection method used here overcomes that, it suffers of lack of reproducibility and too

long a time (the right one depending on the size of the droplet) can lead to unbearably high concentrations.

The AE1 crystal

For the complex between ACGTACGT and Echinomycin the stock solution with a concentration of 0.5 mg/ml of DNA was used. 20 μ l of it were mixed in a hanging drop with 1 μ l of a reservoir solution consisting of 32% PEG 200, 6% PEG 3350, 0.05 M MgCl₂, 0.1 M MES buffer at pH 6 and 0.02 M spermine tetrachloride and suspended over the reservoir. A 0.6 mm long and 0.1 mm thick hexagonal bar shaped crystal grew after two weeks at 293 K (*Figure II-6*) and was frozen directly in liquid nitrogen without further cryoprotection. The 20:1 ratio of stock solution to reservoir solution is very infrequent and leads to extreme concentration changes during the development of the drop.

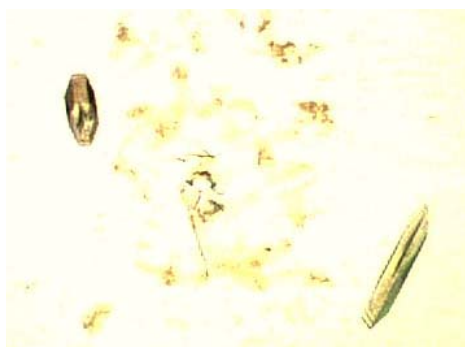


Figure II-6. Hexagonal needles of the AE1 type of crystal

Crystals could be grown with much better reproducibility in more standard drops, from the 4 mg/ml stock solution with 2:1 or 3:1 volume ratios, but they were of smaller size. Data collection was carried at BL14.1, BESSY, at 100 K. The images were integrated and scaled with XDS to a resolution of 1.5 Å. This crystal will from now on be referred to as “AE1”. The unit cell and space group were similar to those of GE1 and GE2 and are listed together with the data collection statistics in *Table II-1*.

The CE crystal

1 μ l of the CGTACG-Echinomycin stock solution was mixed with 4 μ l of a reservoir solution consisting of 30% MPD, 0.04 M MgCl₂, 0.1 M MES buffer at pH 6 and 0.02 M spermine tetrachloride. After 1 month at 285 K, a pyramidal crystal of dimensions 0.2 x 0.4 x 0.6 mm was obtained (*Figure II-7*).

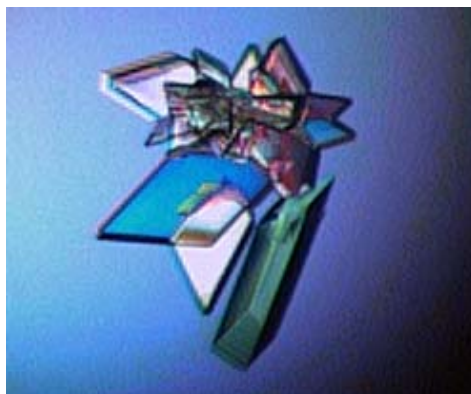


Figure II-7. Crystals from a drop similar to that of CE. The pyramidal shape was likely a diagonal cut through one crystal with rhombic habit like in the picture.

The solution already contained enough cryoprotectant and the crystal was frozen directly in liquid nitrogen. Data collection was carried at 100 K at beamline BL14.1, BESSY and the data were integrated and scaled with XDS to a resolution of 1.4 Å. An *in house* dataset from a similar crystal was assigned to spacegroup F222 as for the complex with Triostin A (Wang *et al.*, 1984) but the synchrotron dataset to 1.4 Å fitted better spacegroup C2 with an R_{int} of 3.5% for 9475 merged reflections while for F222 it was 8.9% for 14587 merged reflections. The rest of the work was carried also in space group C2. This crystal will from now on be referred to as “CG”. The unit cell, space group and data collection statistics are reported in *Table II-1*.

Crystal and drop properties

Diffraction from all those crystals was typically very anisotropic (*Figure II-8*) probably caused by the base stacking being in the same direction(s) all though the crystals.

This caused problems with the integration of the data as the strongest reflections would often show up at resolutions of 3.4 Å (the base stacking distance) and with fiber diffraction background. Some of those reflections are also not present in the datasets as they were overloaded even in the low resolution runs (*Figure II-9*) and some of them have overestimated backgrounds and correspondingly too low intensities. The scaling also suffers from the anisotropy as including the reflections with the highest resolution in the best diffraction direction implies including lots of extremely weak reflections in other directions in reciprocal space. These phenomenons could be the reason while the refinements converged to relatively high R factors after refinement.

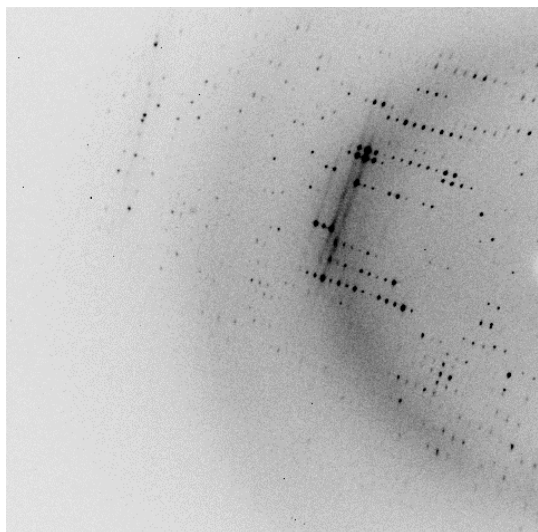


Figure II-8. One image from an in-house data collection on a crystal of the CE form. The maximum intensity is found at 3.4 Å with band-like background and a second maximum can be seen at approximately 1.7 Å in the same direction.

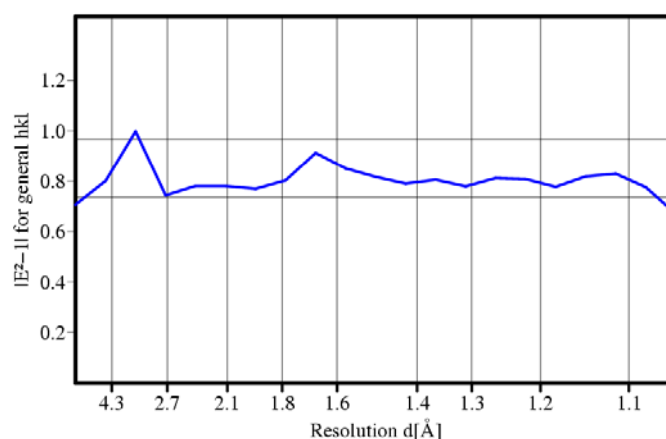


Figure II-9. $|E^2-1|$ plot from SADABS for GE1. The data are normalized in resolution shells. The peaks at 3.4 Å, 1.7 Å and 1.13 Å (fractions of 3.4 Å) indicate an anomalous intensity distribution with some very strong reflections.

Many of the hanging drops developed precipitates seconds after being mixed only to become clear again in the course of some minutes. This probably reflects that Echinomycin and DNA are not or not fully complexed in the water-methanol mixture: fast loss of methanol by evaporation during the process of setting up the drops led to Echinomycin precipitation while further loss of methanol and complex formation redissolved it. The reservoirs were in all cases of 1 ml and did not contain methanol so that final alcohol concentrations in the drops

after vapour equilibration were under 1% in all cases and the crystals grew in a aqueous environment.

II.2.2 Structure solution and refinement.

The structure of GE1 was solved by molecular replacement with EPMR (Kissinger *et al.*, 1999). Diffraction data to 4 Å were used in the search with the Triostin A complex with the same DNA oligomer (Wang *et al.*, 1986) as a model. EPMR performs a six-dimensional search modifying rotations and translations simultaneously by means of a genetic algorithm to maximize the correlation coefficient (on intensities) between the search model and the dataset. a solution was found in the evolutionary search with a correlation coefficient of 85.1% and an R factor of 25.1%. The thioacetal bridge was then built manually with ideal angles and distances. The high resolution of the data would also have allowed a solution by direct methods, a test run with SHELXD (Sheldrick *et al.*, 2001) with default parameters yielded a solution were many of the nucleobases were readily identified.

A GE1 model stripped of waters and disorders was directly used as the starting model for refinement of GE2.

Initial models for AE1 and CE were also obtained by molecular replacement with EPMR using conveniently edited parts of the GE1 structure as search models. For AE1 a search using data to 3.5 Å yielded a solution with a correlation coefficient of 70.1% and an R factor of 42.4%. In the case of CE two copies of the model were searched sequentially using data to 3 Å. The correlation coefficients were 41.0% and 74.7% for the first and second chains respectively with R factors of 52.9% and 34.7% respectively.

All four structures were refined using SHELXL-97 (Sheldrick and Schneider, 1997) with standard protein (Engh & Huber, 1991) and DNA (Parkinson, 1996) 1-2 and 1-3 distances. Restraints for Echinomycin were taken from crystals of small molecules in the CSD (Allen, 2002) as necessary. Given the high resolution of the data and the anisotropic character of the crystals a fully anisotropic restrained refinement with isotropicity restraints for the solvent molecules was chosen. The change from isotropic to anisotropic models was in all cases accompanied by a substantial decrease in both R and R_{free} . In SHELXL refinement restraints are treated as extra observations, that is, as data, and helped in holding the data to parameter ratio in reasonable values (*Table II-1*). The behavior of R_{free} and the physical sense of the resulting models confirmed the adequateness of this model.

II.2.3 Structure of the complexes

The structure of all four crystals shows overall similar features apart from the necessary differences arising from the different DNA sequences. The structure of the complex in crystal GE1, the one with the highest resolution, will be described first and taken as an example for the discussion of the molecular conformations and interactions of Echinomycin with DNA. The other crystal forms will be discussed briefly afterwards.

Crystal structure of GE1

The main molecular unit in the crystal consists of a complex with two DNA chains and two Echinomycins. Echinomycin binds in the minor groove around both CG steps (*Figure II-10 a*) in a bisintercalative manner with the bases protruding to the major groove of the DNA duplex (*Figure II-10 c*). The asymmetric unit consists of only one DNA chain and one Echinomycin molecule (shown in black in *Figure II-10 b*) and the rest of the complex is generated by a crystallographic two-fold axis, both Echinomycins and both DNA chains are thus identical. The self-complementary DNA chains are arranged in an antiparallel manner like in B-DNA. The model was refined to a resolution of 1.1 Å with a final R factor of 14.66% and a R_{free} of 16.81%. 80 waters and a chlorine were included in the solvent model.

The coordinates for this structure are deposited in the PDB with code *1pfe*.

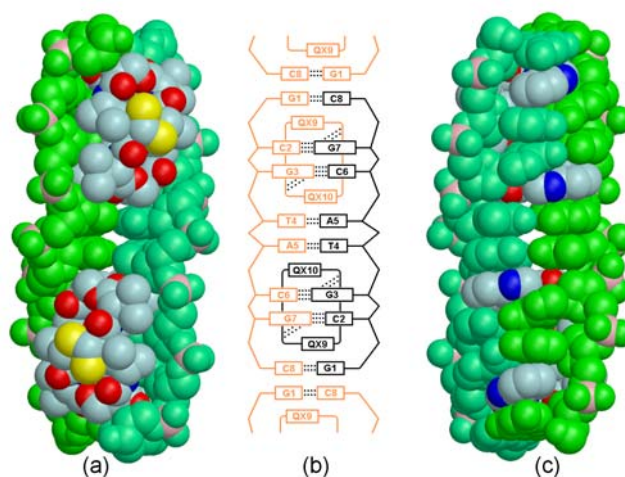


Figure II-10. (a) & (c): Space filling views of the complex in the GE1 crystal. DNAs are green (light and dark for the two different chains) with phosphorus atoms in pink, Echinomycins are colored in gray (carbons), blue (nitrogens), red (oxygens) and yellow (sulfur). Only the main conformation is shown. (b): Schematic view showing the numbering of the bases, QX stands for quinoxaline. The asymmetric unit is shown in black, symmetry equivalents in orange. The dotted lines represent the hydrogen bonds.

Structure of Echinomycin

The depsipeptide backbone of Echinomycin takes the approximate form of a cradle, with the thioacetal bridge at the bottom providing extra rigidity. The backbone itself roughly defines a slightly twisted, elongated rectangle with the 2-carbonyl-quinoxaline groups protruding from opposite corners to the upper side of the cradle so that the overall molecular shape is that of a C (*Figure II-11*).

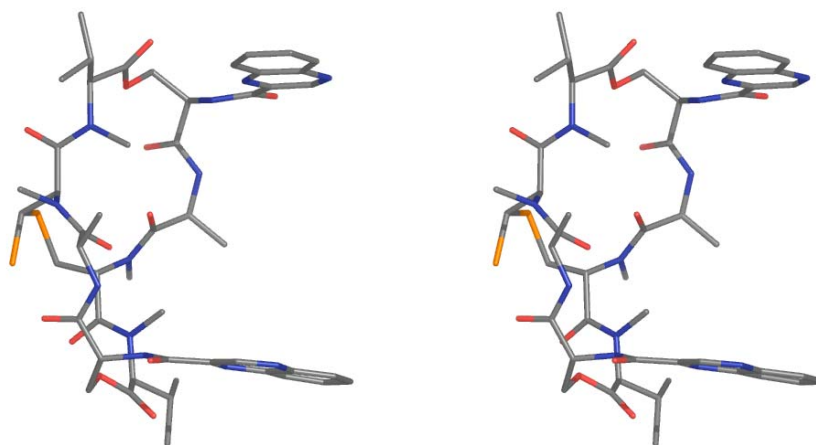


Figure II-11. Stereo view of Echinomycin from the GE1 crystal. Only one conformation is shown.

All peptide bonds are trans, including the ones linking to the quinoxalines, as well as the ester bonds.

If the backbone is considered as a rectangle, one can speak of two sides of the plane, one where the quinoxalines are, which is binding to the DNA; and the other where the thioacetal bridge is. The nitrogens of the serines are facing the DNA side as they bridge to the quinoxalines. In the alanines, the side chains and their nitrogens and oxygens also face that side of the depsipeptide plane. The alanines' oxygens and nitrogens lie in a diagonal line through the rectangle and all four atoms are involved in hydrogen bonding with the DNA. The β -carbons are also in contact with the DNA. The side chains and the methylated nitrogens of the valines also lie in the quinoxaline side. The carbonyl oxygen of the valines is also in the same side but directly faces the quinoxalines themselves at a slightly too long distance to make a hydrogen bond to their N1 atoms.

The side chain of the serines is part of the depsipeptide backbone and its γ -oxygens point slightly to the side of the rectangle where the side chains of the N-methyl-cysteines form the thioacetal bridge are. The carbonyl oxygens and the methylated nitrogens of the cysteines and

the serine carbonyls lie in that side too and thus away from the DNA. The distance between the α -carbons of the N-methyl-cysteines is 3.69 Å, shorter than in the disulfide bridge of the Triostin A structures (Sheldrick *et al.*, 1995; Wang *et al.*, 1984; Wang *et al.*, 1986) so that the cycle looks thinner but the overall structure remains the same.

The 2-carbonyl-quinoxaline units are responsible for bisintercalation in DNA. The oxygens of the carbonyl groups deviate are minimally displaced from the plane of the quinoxalines by 0.16 Å and 0.13 Å. The two bases must ideally be parallel to each other and 10.2 Å apart (three times 3.4 Å, a typical Van der Waals distance for DNA bases and intercalators) to bisintercalate in DNA. Protruding from opposite corners of the backbone rectangle, the bases are rotated by 96.9° relative to each other around their normal axes so that both point approximately to the axis of the DNA. They deviate from the parallel position by only 8.34°, being slightly open. Measuring the distance between two non parallel plane groups is a non trivial question. Considering the distance between any given atom of one quinoxaline to the closest atom of the other, this distance has a minimum of 9.79 Å and a maximum of 10.37 Å (necessarily overestimated by atoms lying out of the projection of the other base). For many atom pairs this distance is less than 10 Å and thus a bit too short for ideal bisintercalation.

The chemical structure of Echinomycin is symmetrical except for the thioacetal bridge and it approximately maintains this symmetry in the crystals. The electron density clearly showed disorder for the bridge but not for the backbone. The density in the thioacetal region suggested not mobility of the atoms, but a two-fold rotation of the whole antibiotic about its symmetry axis. This would let the other atoms in nearly the same positions of their equivalents leading to correlations and thus to instabilities during the crystallographic refinement. Consequently was a second conformation modeled only for the bridge, all other atoms were modeled in a single orientation (*Figure II-12a*). Small deviations from the ideal symmetry can easily be absorbed in the anisotropic displacement parameters. *Figure II-12b* shows a superposition of both possibilities taken from the CE crystal where that kind of disorder is not present, most atoms lie in virtually the same position. The occupancies for both orientations refined to 56% and 44% respectively. As a result, the position of the non-bridge atoms in the refined structure corresponds roughly to the mean of both possible orientations making impossible an assessment of potential asymmetries from this model.

The methyl group in the thioacetal group is connected to a sulphur modeled in two positions and would itself have some torsional mobility, enhanced by the long sulphur-carbon

distances. Its electron density is spread over a large volume and could not be detected in the electron density maps. It was omitted from the model of crystal GE1.

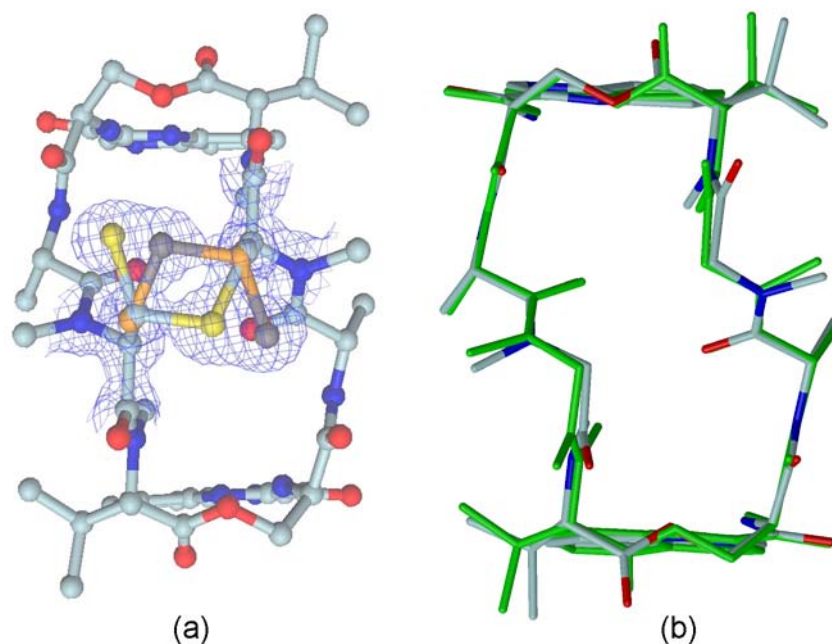


Figure II-12. (a) Electron density and modeled disorder in the thioacetal bridge of GE1. Sulfurs are yellow (first conformation) and dark grey (second conformation). The density is from the $2F_o - F_c$ map contoured at the 1σ level. (b) Two-fold symmetry of the backbone of Echinomycin in the CE structure. The green version has been rotated 180° and superimposed in its own original image (in colors) after removing the bridge. The RMSD of the coordinates is 0.20 \AA .

Structure of the DNA

Binding of Echinomycin has dramatical effects on the structure of the DNA. The most obvious of all is the switch of some basepairs from Watson-Crick base pairing to Hoogsteen base pairing. Basepairs G1-C8, T4-A5, A5-T4 and C8-G1 adopt the Hoogsteen mode, that is, all basepairs external to the bisintercalation site (*Figure II-13*). External and adjacent is a more precise description, since NMR studies suggest that this effect would not be transmitted to the following basepairs.

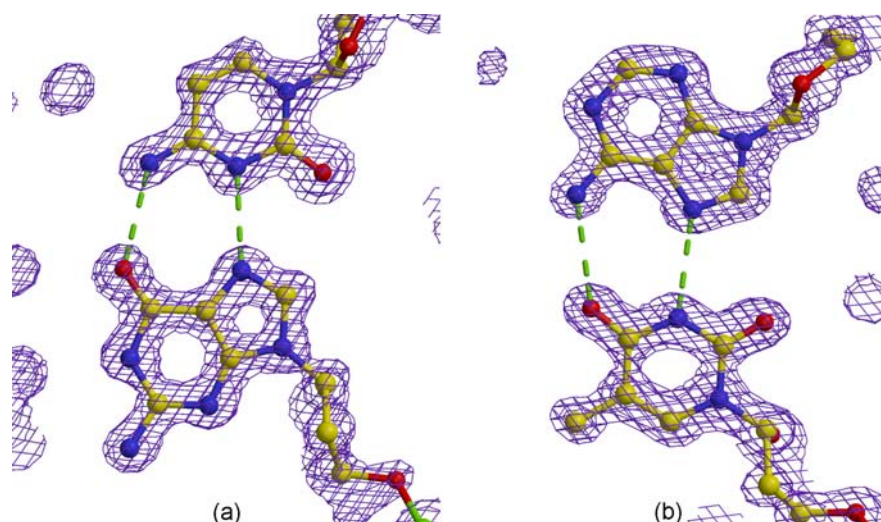


Figure II-13. Hoogsteen basepairs in the GE1 structure. Electron density from $2F_o - F_c$ maps contoured at the 2σ level. (a) G1-C8 basepair. (b) T4-A5 basepair. The A5-T4 and C8-G1 basepairs are symmetry equivalents of the ones shown.

In Hoogsteen basepairs the purines are flipped into the *syn* conformation and bind to the pyrimidines by their Hoogsteen side making only 2 hydrogen bonds. In this mode N6 and N7 of the adenines make hydrogen bonds to O4 and N3 of the thymines respectively while O6 and N7 of the guanines are H-bonded to N4 and N3 of the cytosines. This last N3 atom must be protonated as it becomes the hydrogen bond donor. The pK_a of N3 of cytosines is 4.6 (Windholz, 1983). The crystal was grown at pH 4.5 but protonation was maintained through the cryoprotection step at pH 6.0. Crystals did also readily grow at pH 6.0, not any more at pH 7.0, reflecting a shift to higher values at this pK_a . Wang *et al.*, (1986) reported the same for the complex with Triostin A with crystals growing at pHs up to 6.5 and a shift in the pK_a value of protonated cytosines that are part of stable structures has been noticed before (Hartman & Rich, 1965). Extra stabilization effects accompanying the protonation shift must influence the position of the equilibrium, possibly the shortening of the C1'-C1' distance associated with the Hoogsteen base pairing (Table II-2), thus allowing for numerous extra Van der Waals contacts between the DNA and the antibiotic. Binding can occur also without the shift to the Hoogsteen mode as shown by NMR structures and footprinting studies (many of the latest carried at pH 7). The basepairs internal to the bisintercalation site maintain the classical Watson-Crick binding mode. Inter-base hydrogen bonding distances for this crystal are listed in Table II-3, all are within normal ranges.

Basepair	G1-C8	C2-G7	G3-C6	T4-A5	A5-T4	G6-C3	C7-G2	G8-C1
C1'-C1' distance (Å)	8.2	10.5	10.6	8.2	8.2	10.6	10.5	8.2
Type	HG	WC	WC	HG	HG	WC	WC	HG

Table II-2. Distances between the C1' atoms of the sugars in the GE1 crystal.

Atoms	Distances (Å)			Atoms	Distances (Å)	Basepair type
	GE1	GE2	CE			
G1_O6...C108_N4	2.85	2.85	-	A1_N6...T108_O4	2.92/2.89	HG
G1_N7...C108_N3	2.73	2.68	-	A1_N7...T108_N3	2.77/2.86	
C2_O2...G107_N2	2.84	2.77	2.80	C2_O2...G107_N2	2.86	WC
C2_N3...G107_N1	2.89	2.87	2.88	C2_N3...G107_N1	2.92	
C2_N4...G107_O6	2.86	2.77	2.83	C2_N4...G107_O6	2.82	
G3_N1...C106_N3	2.93	2.90	2.94	G3_N1...C106_N3	2.84	WC
G3_N2...C106_O2	2.80	2.77	2.83	G3_N2...C106_O2	2.76	
G3_O6...C106_N4	2.97	2.86	2.87	G3_O6...C106_N4	2.93	
T4_N3...A105_N7	2.85	2.86	2.89	T4_N3...A105_N7	2.85	HG
T4_O4...A105_N6	2.95	2.78	2.90	T4_O4...A105_N6	2.89	
A5_N6...T104_O4	2.95	2.78	2.94	A5_N6...T104_O4	2.89	HG
A5_N7...T104_N3	2.85	2.86	2.79	A5_N7...T104_N3	2.85	
C6_O2...G103_N2	2.80	2.77	2.78	C6_O2...G103_N2	2.76	WC
C6_N3...G103_N1	2.93	2.90	2.95	C6_N3...G103_N1	2.84	
C6_N4...G103_O6	2.97	2.86	2.98	C6_N4...G103_O6	2.93	
G7_N1...C102_N3	2.89	2.87	2.84	G7_N1...C102_N3	2.92	WC
G7_N2...C102_O2	2.84	2.77	2.84	G7_N2...C102_O2	2.86	
G7_O6...C102_N4	2.86	2.77	2.83	G7_O6...C102_N4	2.82	
C8_N4...G101_O6	2.85	2.85	-	T8_N3...A101_N7	2.77/2.86	HG
C8_N3...G101_N7	2.73	2.68	-	T8_O4...A101_N6	2.92/2.89	

Table II-3. Hydrogen bonding distances in the basepairs. For GE1, GE2 and AE1 100 has been added to the residue numbers of symmetry-equivalent bases. WC stands for Watson-Crick and HG for Hoogsteen base pairing. In the case of AE1 both conformations for the terminal basepairs are included and both have the same bonding scheme.

In the GE1 structure, and equally in GE2, the N3 proton of the Hoogsteen cytosines could not be detected in the electron density maps at a significant level (3σ in the F_o-F_c map) and was not included in the model.

A conformational analysis of the DNA illustrates the adaptations accompanying bisintercalation. The Hoogsteen base pairing changes the frame of reference for defining basepair orientations and renders many usual helical parameters useless, at least for comparison with normal values, approximations will be taken as necessary. Also the bases inside the intercalation sites show considerable buckling so that the bases are not parallel (not even approximately parallel) to each other. The rise, distance between basepairs, increases to values between 5.30 Å at C1' and 6.66 Å at N2 of guanine G7 around the terminal quinoxaline and between 4.94 Å at C1' of guanine G3 and 6.66 Å at O6 around the internal quinoxaline in the complex. Those are values measured at the sugar end and at the hydrogen bonded end of the bases, the mean value is slightly under 6 Å, significantly less than the 6.8 Å expected for ideal bisintercalation. The complex has a certain degree of bending but the minimum distance from C1' of C8, laying approximately on the hypothetical axis of the complex, to the basepair at the opposite end is 38.07 Å, corresponding to a mean of 3.46 Å for each of the 11 base steps. This value is a bit higher than the standard 3.4 Å, regions of appreciable compression in the complex come together with others where the DNA structure relaxes and opens. Contacts to adjacent bases (or quinoxalines) no longer than 3.5 Å can be measured in all cases but are often the minimum distances between non parallel bases. The cited relaxation must then be attributed to those cases.

The helix also gets significantly unwound upon binding as is necessary to accommodate the rise values concomitant to intercalation. Again definition of an helix axis against which to refer other angles is difficult since Hoogsteen basepairs change the reference frame. An approximate twist value can be derived by considering the pseudo-torsion angle defined by the C1' atoms in two consecutive basepairs. This way a twist angle of 54.56° can be measured between basepairs G1-C8 and T4-A5 while for the central basepairs the twist angle is 23.64°. This way an unwinding angle of 53.74° is measured between the first and fourth basepairs relative to the 36.1° twist per basepair of B-DNA. The 12.46° unwinding of the central basepair hints to an extension of the unwinding effect beyond the intercalation site although its environment is far from that of a free DNA duplex. Extrapolating this values, a similar unwinding angle per Echinomycin molecule results as the 66° reported for Triostin A in complex with (GCGTACGC)₂ (Wang *et al.*, 1986), and higher than the 45°-55° reported for Echinomycin binding in solution (Lee & Waring, 1978).

Some disorder was observed for the phosphate of C2, it was modeled with two conformations as well as the sugars linked to it, but the density for the adjacent bases was clear and only one conformation was used for them.

The glycosidic torsion angles χ are not strongly affected by binding (*Table II-4*) and can be compared with those of B-DNA except of course those of G1 and A5 that are rotated into the *syn* conformation to form Hoogsteen basepairs. The sugar puckering takes a significant part of the adaptation work carried by the DNA and shows a range of different values (*Table II-4*).

Base	G1	C2	G3	T4	A5	C6	G7	C8
χ	74.7	-108.8	-108.5	-95.3	68.7	-109.6	-106.2	-100.2
Amp	25.5	44.0	40.4	30.6	43.6	39.5	40.1	34.6
P	161.6	45.8	131.9	164.5	97.0	61.9	115.6	173.9
Pucker	C2'-endo	C4'-exo	C1'-exo	C2'-endo	O4'-endo	C4'-exo	C1'-exo	C2'-endo

Table II-4. Angles in the sugars of GE1. χ is the glycosidic angle, Amp the amplitude of the pseudorotation, P the pseudorotation angle and Pucker the sugar pucker angle.

Interactions between Echinomycin and DNA

Echinomycin binds in the minor groove of the duplex DNA around all CG steps although it protrudes into the major groove. Its quinoxaline groups incorporate themselves partially in the base stacking of the DNA originating the rise in the surrounding bases and the DNA unwinding. This stacking takes place mostly with the purine bases external to the intercalation site, in the internal side the quinoxalines lie mostly over the space in the middle of the basepairs.

The interaction can be divided into three kinds of contacts: stacking interactions, Van der Waals contacts and hydrogen bonds. Most of the binding energy comes probably from the stacking interaction although they are far from maximized. Strong stacking takes place only to the external side of the Echinomycin while the internal bases are not parallel to the quinoxalines, leading to a smaller number of contacts than possible. The strong buckling of the internal basepairs means that in order to avoid too close contacts at some points the distance increases in others well over 3.5 Å between the quinoxalines and the DNA bases and to over 4 Å between the DNA bases themselves. This buckling is in the order of 23.5° for

GE1 (and similar in the other crystals) and could be deemed unavoidable on the light of the limited space between the quinoxalines, smaller than 10 Å. In all the series of structures of complexes between Triostin A or Echinomycin and DNA the antibiotics have a relatively rigid structure while the DNA is more flexible and adapts its conformation. Without pushing the quinoxalines apart there is no space between them for two basepairs in a parallel orientation, while pushing the quinoxalines apart would then make them tilt out of the plane of the other bases and extend the loose fit also to the external basepairs. Close contacts between Echinomycin and the DNA are shown in *Figure II-14*.

Some of the energy lost in the base stacking is recovered by a close fit between the Echinomycin backbone and the minor groove of the DNA. The change to the Hoogsteen basepair configuration brings the C1' carbons of the DNA 2 Å closer than in the Watson crick mode making the minor groove thinner. In this way a great number of Van der Waals contacts arises (*Figure II-14*), it is possible that this be the reason for the DNA to adopt Hoogsteen basepairing. The side chains of the alanines even protrude between the sugars of nucleotides C2-G3 and C6-G7 probably contributing also to the buckling of those basepairs. All these contacts take place in the quinoxaline face of the Echinomycin backbone. The thioacetal bridge is located behind the backbone, protruding away from the DNA. Since it has no strong interactions, it also has a small contribution to the energy of the complex originating the two-fold disorder of Echinomycin. All parts of Echinomycin interacting with the DNA are symmetrically arranged.

There are no contacts closer than 3 Å in this crystal structure in disagreement with previous studies (Ughetto *et al.*, 1985) where such non-hydrogen-bond close contacts were reported both internal to the Echinomycin and between Triostin A and the DNA. Those structures were refined to much lower resolutions (2.2 Å) with less sophisticated software and during their refinement it was also necessary to constrain the hydrogen bonding geometry which can introduce stress. The authors pleaded for caution in interpreting individual details of those structures as they were clashing with the limits of the available methods. No constraints or restraints were imposed in the GE1 hydrogen bonding and together with the much higher resolution it should provide a more reliable picture.

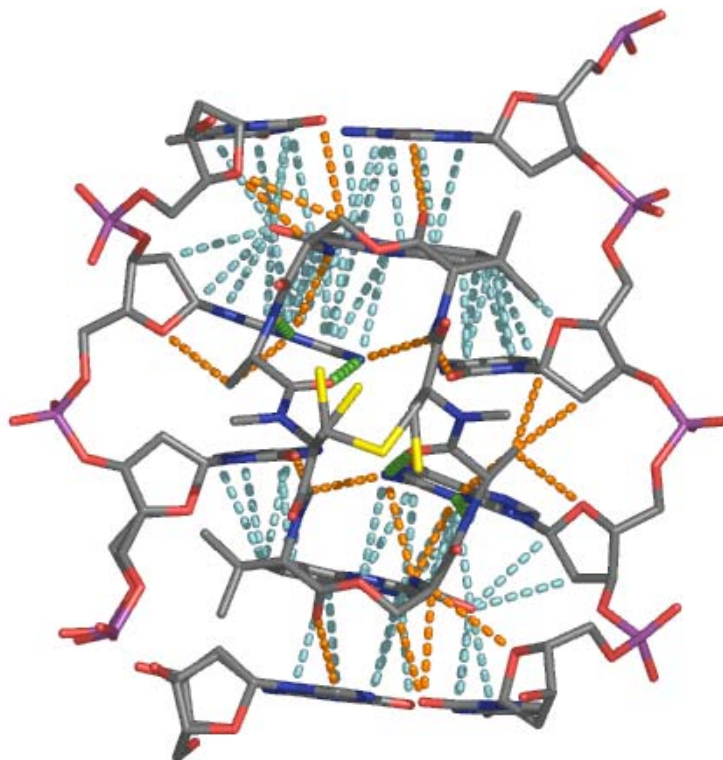


Figure II-14. Contacts closer than 3.5 Å between Echinomycin and the DNA. Only half of the complex is shown, the other half containing the second Echinomycin is symmetry equivalent to this. Contacts made by the quinoxalines are shown in blue (not only base stacking), contacts made by the backbone are in orange and hydrogen bonds in green. The disordered thioacetal bridge has no Van der Waals contacts.

The third kind of interactions is hydrogen bonds between the Echinomycin and the DNA. Four such bonds are found nearly symmetrically arranged between the nitrogens of the alanines and N3 of the non-terminal guanines and between the carbonyl groups of the same alanines and N2 of the same guanines (*Table II-5*). These hydrogen bonds are responsible for the recognition of the CG intercalation step: One guanine is recognized in each DNA chain, thus defining the CG step. In the old crystal structures, also in the Triostin A complex with CGTACG (Wang *et al.*, 1984), only three hydrogen bonds were identified, the fourth one had distances of 3.6 Å and 4.1 Å. This fourth hydrogen bond was found here with distances in the order of 3.15 Å.

Crystal	Donor	Acceptor	Distance (Å)
GE1	Ala-12 N	G-7 N3	3.05
GE2	Ala-12 N	G-7 N3	3.03
AE	Ala-12 N	G-7 N3	3.12
CE	Ala-12 N	G-3 N3	2.95
CE	Ala-112 N	G-103 N3	2.99
GE1	Ala-17 N	G-3 N3	3.09
GE2	Ala-17 N	G-3 N3	3.01
AE	Ala-17 N	G-3 N3	2.93
CE	Ala-17 N	G-107 N3	2.98
CE	Ala-117 N	G-7 N3	2.96
GE1	G-7 N2	Ala-12 O	3.13
GE2	G-7 N2	Ala-12 O	3.22
AE	G-7 N2	Ala-12 O	3.16
CE	G-3 N2	Ala-12 O	3.18
CE	G-103 N2	Ala-112 O	3.15
GE1	G-3 N2	Ala-17 O	3.16
GE2	G-3 N2	Ala-17 O	3.08
AE	G-3 N2	Ala-17 O	3.19
CE	G-107 N2	Ala-17 O	3.14
CE	G-7 N2	Ala-117 O	3.17

Table II-5. Hydrogen bond lengths between the Echinomycins and the DNA.

Crystal structure of GE2

The DNA sequence and the space group of GE2 are the same as for GE1 and the cell and structure are also very similar, GE2 can be considered to be the same crystal form as GE1. It is included as a different one here for the purpose of discussion only. The fact that both structures are so similar is remarkable taking into account the different crystallization conditions. GE1 is in an environment of very high ionic strength, with 1.1 M Li₂SO₄ as the main precipitant and 7 M concentration of it during the cryoprotection step where the molecules still can undergo limited conformational changes. In GE2 the pH is lower, the ionic strength low and the main precipitant agent is PEG 550 mme, presumably in aprox. 30%

concentration upon freezing. This reflects the structural rigidity of the main motif in the crystals, the 2:2 DNA:Echinomycin complex.

In this crystal it was necessary to model not only the phosphate of C2 but also that of A5 and the adjacent sugars in a double conformation. Again the density of the bases was clear and only one conformation needed there.

The double conformation in the thioacetal bridge reflecting overall two-fold disorder in the Echinomycin, it was also observed for this crystal and modeled as for GE1 with occupancies refining to 51% and 49%. In this case the methyl group of the bridge could be observed in the difference map and was included in the final model for both conformations.

A double conformation was also detected and modeled for N-Me-Valine 14 and the corresponding ester linkage (not in the other valine). The valines hardly interact with the DNA (*Figure II-14*) and are the main source of chemical variability between this kind of antibiotics.

The proton in N3 of Cytosine C8 necessary for Hoogsteen base pairing was also not observed in the difference density even at this lower pH and was not included in the final model.

One magnesium ion and 56 water molecules were included in the solvent model.

Crystal structure of AE1

Again this structure is very similar to GE1 except for the obvious change in the DNA sequence. The terminal A-T bases (only one of each in the asymmetric unit) in both 5' and 3' ends also undergo Hoogsteen pairing but were observed to have massive disorder and the electron density was difficult to interpret. The solution was to test parallel refinements of all possible binding schemes. The model with Hoogsteen base pairing for two conformations explained the density with normal B factors and caused a drop of 2% in both R factor and R_{free} . It involves a coupled movement of 2 Å of the bases in the base pairing direction in both cases, if it assumed that it is a concerted movement, the base stacking is preserved as are the hydrogen bonding distances. The disorder was modeled as two conformations for the whole nucleotides plus the adjacent sugars. Interestingly, this happens without necessity of carrying the disorder to the Echinomycin they stack on. The involved quinoxaline shows no disorder nor higher than normal B factors and it stacks mostly to the adenines in both conformations as is usually observed.

Sugar puckers and pseudorotation angles are similar as those for GE1 for the central four bases but radically different for the terminal ones. This can probably be attributed to the disorder modeled and those values should probably not be taken as a reliable reference.

The disorder in the thioacetal bridge was also observed and modeled similarly to GE1 with refined occupancies of 54% and 46%, as in GE1 the methyl groups could not be observed and were not included in the final model.

One magnesium ion and 41 water molecules were included in the solvent model.

Crystal packing in the P6₃22 crystals

The duplexes in these crystals form infinite columns along the crystallographic *c* axis with the terminal bases stacking on their symmetry equivalents (G1 on C8* and C8 on G1*). Two kinds of solvent channels are present along the *c* axis with approximate diameters of 30 Å and 10 Å (*Figure II-15*). The disulfide bridges of Echinomycin face the big ones while the phosphates of the DNA backbone face the small channels. The thioacetal bridges interact only with the solvent.

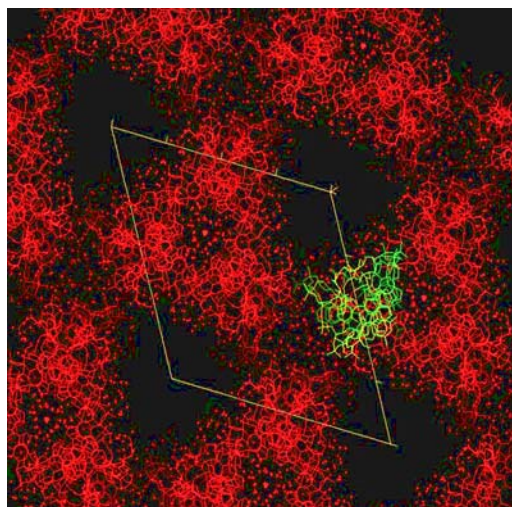


Figure II-15. Crystal packing in the GE1 crystal, one DNA duplex is shown in green.

$[\text{Mg}(\text{OH}_2)_6]^{2+}$ complexes are found binding to the phosphates of G4 from three different columns in the small solvent channels of GE2 and AE. The magnesium ions were not found in the GE1 structure, they were probably displaced by the high lithium concentration. Since this is a relatively important crystal contact and all three crystals are of the same form, it can be speculated that the magnesium complexes took part in crystal growth and were only displaced in the cryoprotection step when the lithium concentration rose from 1.1 M to 7 M. There are also two hydrogen bonds from N1 and N2 of guanine G1 (which are free since G1

is not Watson-Crick basepaired) to the phosphate of C8 or T8 in a symmetry equivalent. Another hydrogen bond between N4 of C2 and the phosphate of G7 in the same symmetry equivalent as before also contributes to the packing.

Crystal structure of CE

The asymmetric unit of this crystal form contains a full duplex with two DNA hexamers and two Echinomycin molecules. The Echinomycins are intercalated around both CG steps as expected and the central AT basepairs are in the Hoogsteen conformation while the CG basepairs are all four Watson-Crick (*Figure II-16*). Overall, the structure resembles the biologically relevant unit of GE1 after omission of the terminal basepairs.

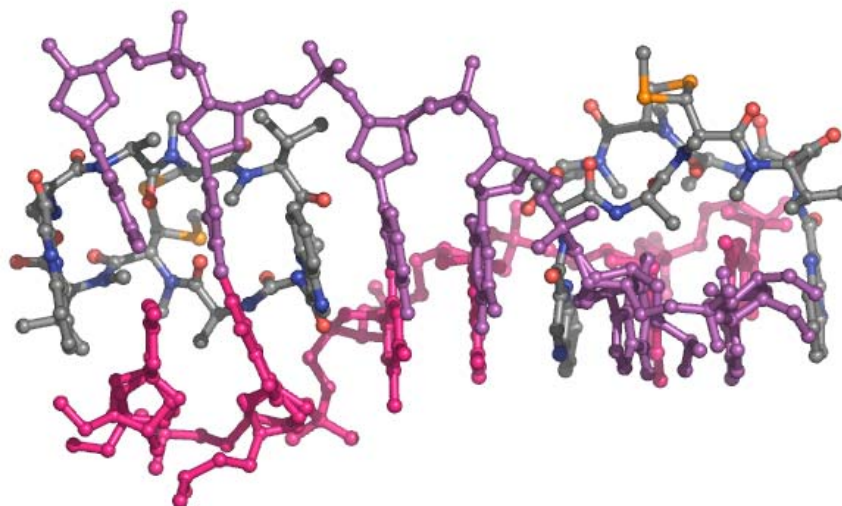


Figure II-16. Asymmetric unit of crystal CE. DNA in pink and violet, Echinomycins in colors. The bases enclosed by the Echinomycin are buckled.

The structure was modeled and solved in spacegroup C2 but it nearly obeys the higher symmetry of F222. In F222 the asymmetric unit would be one DNA chain and one Echinomycin molecule with the second of each kind generated by a crystallographic two-fold axis. The difference between the two scenarios could only be detected when synchrotron, high resolution data were available. The two-fold axis is broken in the C2 structure only by the first two nucleotides of each chain and also partially by the solvent model with slightly different hydration patterns in both chains. The other 14 residues, four in the DNA and ten in the Echinomycins, closely follow the two-fold axis. The RMSD for the superposition of those 14 residues from both halves of the complex is only 0.11 Å, while all 16 residues it is 0.82 Å.

Refinement in the higher symmetry space group F222 was also tested and proved to be possible with similar residuals as for the C2 refinement, the asymmetries would get modeled as disorder in F222. Even in C2 it was necessary to model two conformations for the backbone of C2 and G3, but not for the nucleobases.

In this crystal form both Echinomycins were modeled in only one orientation with clear electron density (*Figure II-17*). The thioacetal bridges are now involved in Van der Waals contacts at 3.8 Å to symmetry equivalent Echinomycins and the second orientation would make short contacts with the consequent energetic penalty. Also the methyl groups of the bridges could be observed and were included in the final model, for both Echinomycins they are oriented towards the center of the complex.

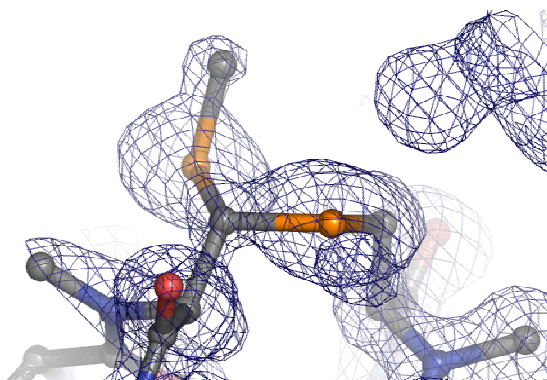


Figure II-17. Electron density from the $2F_o - F_c$ map contoured at the 1.5σ level in the region of the thioacetal bridge of one Echinomycin in the CE crystal. The density of the sulfurs is higher than for any other atom. The density in the upper-right corner corresponds to a symmetry equivalent molecule.

The hydrogen bonding scheme is fully equivalent to that of the octamers both for the base pairing (*Table II-3*) and for the recognition of the guanines by Echinomycin (*Table II-5*). All four hydrogen bonds are observed between the alanines and the guanines in both halves of the asymmetric unit in disagreement with the results from the constrained refinement in Ughetto *et al.* (1985).

The superposition of one Echinomycin from this structure on itself is shown in *Figure II-12b*. The Echinomycins show no disorder in this crystal form and consequently possible asymmetries are not promediated. The thioacetal bridges were removed from the models and remaining atoms superimposed on their equivalents after a 180° rotation prior to the superposition. The results are very similar for both Echinomycins, for which independent superpositions were made, with RMSDs of 0.20 Å in both cases. The differences concentrate in the valine side chains (and to a lesser extent in the ester linkages) and are smaller in the

backbone. This degree of internal symmetry is remarkable since the DNA does not offer a symmetric environment, in this crystal form one of the quinoxalines is terminal and external to the DNA while the other is internal and thus strongly constrained, still they occupy spatially equivalent positions.

The buckle angles in the basepairs internal to the intercalation site oscillate between 19.75° and 22.04° , not very different to those found in the GE1 crystal. The glycosidic torsion angles deviate slightly and the sugar puckers significantly from those of GE1 and this structure should not be taken as a detailed reference for the conformation of the complex in solution since the lack of the terminal bases external to the bisintercalation site provides the DNA with an extra, unnatural flexibility.

The complexes are found again forming infinite columns in the crystal, in this case with quinoxaline-quinoxaline stacking. There are two families of columns crossing each other at an angle without direct hydrogen bonding but with extensive Van der Waals contacts to each other (*Figure II-18*). There are also two $[\text{Mg}(\text{OH}_2)_6]^{2+}$ complexes (only three water molecules in the asymmetric unit) linking together guanine 3 and quinoxaline 9 both within each complex and to a symmetry equivalent in a different column as well as guanine 103 and quinoxaline 109 to their respective equivalents in the same way.

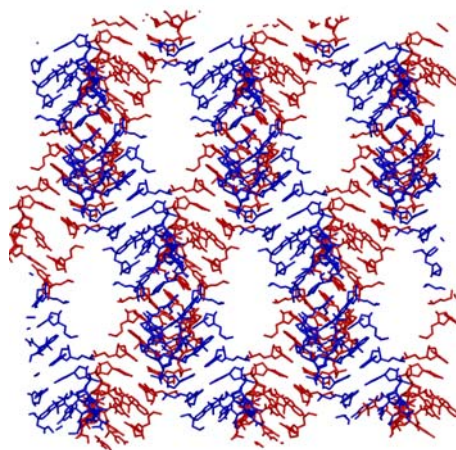


Figure II-18. Criss-crossed infinite columns in the CE crystal.

Comparison of the structures

All four crystals show the same basic structure up to a surprising degree of fidelity taking into account the different DNA sequences and the different spacegroups. Superpositions on the asymmetric unit of the octamers taking GE1 as a reference and only the main conformations in the disordered residues give RMSDs of 0.60 \AA for GE2 and 0.83 \AA for AE.

A superposition of the common residues for the biological units (duplexes) in crystals GE1 and CE give an RMSD of only 1.36 Å with the biggest differences concentrating in the thioacetal bridges, which are inverted in CE relative to the main conformation of GE1, and the terminal quinoxalines which undergo crystal contacts in CE1. That different sequences of DNA and solvent environments result in such closely related structures speaks for a strong control of the DNA structure by Echinomycin.

If the superposition is carried out separately for the DNA and the Echinomycin parts (*Figure II-19*) it is evident that almost all of the (small) differences are concentrated on the side of the DNA. All Echinomycins are nearly identical while the DNAs, specially the terminal bases involved in crystal packing are not. The terminal bases of CE are not involved in the packing themselves but, as a difference to the octamers, they are not conformationally constrained by an extra base external to the Echinomycin.

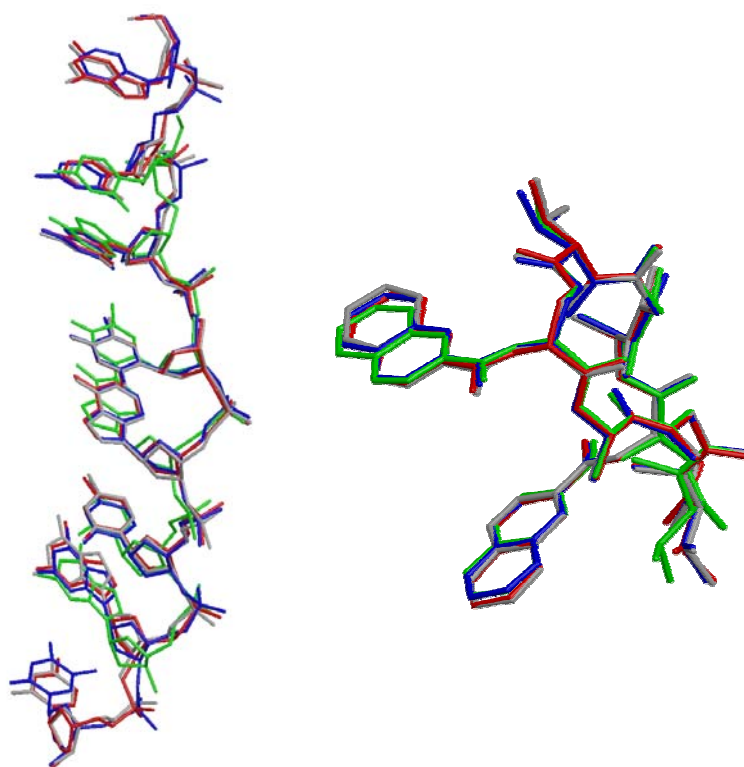


Figure II-19. Superposition of the DNA parts alone (left) and the Echinomycins alone (right) for crystals GE1 (red), GE2 (white), AE1 (blue) and CE (green). For CE only one Echinomycin and one DNA chain is shown.

A comparison has also been carried out relative to the complexes with Triostin A from 1984-1986. Coordinates are only available for the Triostin A complexes. For the GCGTACGC complex with Triostin A compared with GE1 no convincing results could be

obtained when superposing all of the asymmetric unit, but restraining the superposition for the nucleobases and quinoxalines only as shown in *Figure II-20* gives an RMSD of 0.50 Å. The position and conformation of Echinomycin do not correspond exactly to that of the Triostin as expected for the shorter bridge in the latest. The positions of the sugars are very different in both cases with sugars 3, 6 and 7 in the Triostin A complex far from the B-DNA like conformation observed for the Echinomycin complexes. This can probably be attributed to problems with the refinement at the time and the lower resolution of 2.2 Å, the authors pleaded for caution on interpreting individual details of that structure. The buckling of the nucleobases internal to the intercalation site is also bigger for the Echinomycin complex.

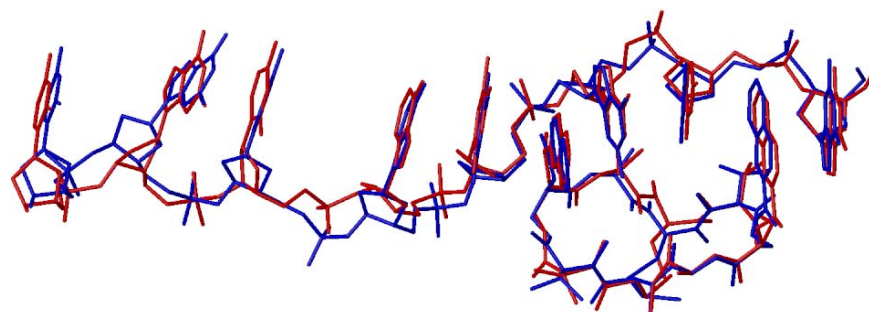


Figure II-20. Resulting orientations and structures after superposition of the bases and quinoxalines in the asymmetric units of GE1 (blue) and the Triostin A complex from Wang *et al.* (1986).

A superposition of the bases and quinoxalines for the CGTACG complexes with Triostin A (in F222) and Echinomycin (in C2) gives an RMSD of only 0.17 Å and 0.18 Å for both possibilities in the C2 duplex and this time also the sugars superimpose satisfactorily confirming the similarity of the binding and the structures of Triostin A and Echinomycin.

II.3 Structure of an Echinomycin-DNA complex performing both Hoogsteen and Watson-Crick basepairs outside the intercalation site.

II.3.1 Crystallization and data collection

The same DNA oligonucleotide (ACGTACGT) and Echinomycin samples were used as for AE1 crystal (*Pag. 33*). The water:methanol ratio in the stock solution was 1:1 and the molar ratio of Echinomycin to DNA was 1.1 with a DNA concentration of 0.5 mg/ml.

The AE2 crystal

A hanging drop was prepared by mixing 25 μ l of the DNA-Echinomycin stock solution with 1 μ l of reservoir solution consisting of 32% PEG 200, 6% PEG 3350, MES buffer 0.1 M at pH 6 and spermine tetrachloride 0.02 M. No $MgCl_2$ was included in the drop. The hanging drop was incubated over 1 ml of the reservoir at 293 K. Over more than one month a needle shaped crystal grew in the drop. The crystal had slightly curved faces and was 1.25 mm long and 0.05 mm wide (*Figure II-21*). This crystal will from now on be referred to as “AE2”.



Figure II-21. A crystal of the AE2 form. The long sides of the needle were always slightly curved and the crystals would only grow weeks after microcrystals or precipitate had developed in the drops.

The main difference between the drop setups for this kind of crystal and the AE1 kind is the absence of $MgCl_2$. Crystals of the AE2 kind grew over a range of concentrations of all the other agents in the mother liquor but would only grow in total absence of magnesium salts while crystals of the AE1 kind grew both with and without magnesium in the mother liquor. AE1 like crystals grew sometimes in setups for the AE2 form, but the opposite was not true.

Also, crystals of the AE1 type could also grow from more concentrated stocks of DNA-Echinomycin with DNA concentrations of 4 mg/ml if the drop ratio was reduced to 2:1

or 3:1, while the AE2 type of crystals only grew from the diluted stock solution and drop ratios of at least 10:1. Interestingly, the final concentrations in the drops are expected to be very similar in both scenarios and the AE2 type crystals would only start growing approximately two weeks after setting up the drops, when the concentrations of all compounds are expected to be the equilibrium ones or very close to it.

The crystal was mounted in a nylon loop and flash frozen in liquid nitrogen. No extra cryoprotectant was used. Data collection was carried on at 100 K at beamline BL14.1, BESSY. 120° of oscillation data with a rotation of 1° per image were collected in a high resolution dataset and another 120° of data in a low resolution pass. The images were integrated and scaled with XDS to a resolution of 1.6 Å. The crystal belongs to spacegroup P4₂2₁2 with unit cell a = b = 80.904 Å, c = 48.194 Å. The data collection statistics for AE2 are listed in *Table II-6*.

Crystal	AE2
DNA sequence	(ACGTACGT) ₂
Space group	P4 ₂ 2 ₁ 2
Unit cell (Å)	a = b = 80.904, c = 48.194
Wavelength (Å)	0.9000
Resolution (Å)	1.60 (1.70)
Total reflections	269046
Unique reflections	21134
Completeness (%)	97.6 (96.3)
Redundancy	12.42 (10.07)
R _{int} (%)	7.16 (39.41)
I/σ(I)	18.70 (5.33)
Data*/restraints/parameters	19031/15202/10153
R (%)	17.96
R _{free} (%)	22.15

Table II-6. Dataset and refinement statistics for crystal AE2. In parenthesis are the values for the high resolution shell, its lower limit are shown in parenthesis in the resolution line.

*The number of data listed is the number of reflections used in the refinement, that is, the working set, the rest to the total number of unique reflections is the test set.

The length of the needle shape of the crystal that was used for data collection must be taken into consideration. Protein crystals and equally DNA crystals are typically mounted in

nylon loops and frozen directly in them, either in liquid nitrogen or in the 100 K nitrogen gas stream of a cryostat. A nylon loop of the length of this crystal would be 1.3 mm long and approximately 0.5 mm wide, taking a considerable volume of drop liquid with it as well as the crystal. This water based amorphous glass produces background scattering in the diffraction images increasing both the background level of the images and the standard uncertainties of the measured background intensities. Since the background intensities have to be subtracted from the measured intensities, the background adds its own standard deviation to that of the integrated spot intensities, thus the presence of liquid in the loop reduces the $I/\sigma(I)$ ratio of the measured intensities and the effective diffraction limit of the crystals.

In this case a nylon loop of approximately half the length of the crystal was used (*Figure II-22*). The volume of the droplet taken together with the crystal is in this way smaller. Furthermore, half of the crystal lies completely out of the droplet. This part of the crystal is free from external cryoprotectant and consequently relatively free of background scattering. The lack of an external pool of cryoprotectant can lead to drying of the crystal and a deterioration of its diffraction properties but this can be prevented through quick handling during the freezing procedure. A bigger diffraction limit was observed for this part of the crystal and only that part was irradiated during the synchrotron data collection.

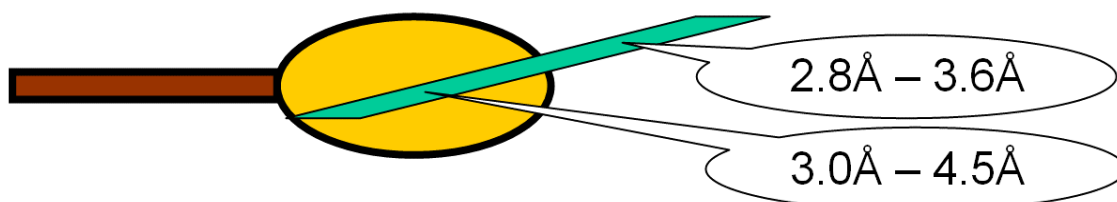


Figure II-22. Schematic diagram of the mounting of the AE2 crystal in the nylon loop. Only the part of it which is inside the loop is actually surrounded by vitrified cryoprotectant. In house tests on a rotating anode diffractometer showed a better (anisotropic) diffraction limit for the part of the crystal lying outside the loop, which was approximately 0.6 mm long. The numbers in the figure correspond to the diffraction limits in the best and the worse directions in reciprocal space in the test images.

The synchrotron data collection was intended to serve as a high resolution native dataset. A molecular replacement solution of the phase problem could then be attempted or heavy atom derivatives prepared. An analysis of the highly redundant dataset (*Table II-7*) showed the presence of anomalous scattering in the crystal with the $I/\sigma(I)$ ratio for the Bijvoet differences indicating that useful phase information could probably be obtained at least up to

2.5 Å resolution. The nature of the anomalous scatterer or scatterers was unknown at the time since no heavier atoms than chlorine were included in the crystallization setup.

Resolution	#Data	#Theory	%Complete	Redundancy	Mean I	Mean I/s	R(int)	Rsigma				
Inf - 4.54	1061	1091	97.3	16.44	199.6	57.32	0.0455	0.0193				
4.54 - 3.56	1065	1079	98.7	19.89	124.2	56.08	0.0535	0.0154				
3.56 - 3.09	1073	1088	98.6	20.13	187.3	45.83	0.0590	0.0209				
3.09 - 2.79	1095	1115	98.2	15.79	46.4	32.53	0.0866	0.0232				
2.79 - 2.58	1097	1110	98.8	11.18	22.5	25.45	0.0844	0.0331				
2.58 - 2.42	1086	1101	98.6	11.25	18.8	22.68	0.0941	0.0366				
2.42 - 2.29	1131	1159	97.6	11.15	13.8	18.79	0.1185	0.0454				
2.29 - 2.18	1164	1191	97.7	11.22	10.3	15.66	0.1490	0.0577				
2.18 - 2.09	1139	1159	98.3	11.31	8.2	13.59	0.1773	0.0690				
2.09 - 2.01	1206	1231	98.0	11.25	6.4	10.84	0.2270	0.0872				
2.01 - 1.94	1201	1225	98.0	11.29	5.0	9.48	0.2634	0.1072				
1.94 - 1.88	1190	1229	96.8	11.21	4.5	8.83	0.2885	0.1184				
1.88 - 1.83	1101	1136	96.9	11.20	4.7	8.29	0.2883	0.1161				
1.83 - 1.78	1226	1254	97.8	11.25	3.3	6.52	0.3987	0.1573				
1.78 - 1.74	1083	1114	97.2	11.27	3.1	6.30	0.3909	0.1678				
1.74 - 1.70	1186	1228	96.6	11.17	3.4	6.50	0.3718	0.1573				
1.70 - 1.66	1300	1336	97.3	11.03	4.1	6.21	0.3446	0.1359				
1.66 - 1.62	1426	1474	96.7	9.57	2.8	4.66	0.4557	0.2122				
1.62 - 1.60	304	336	90.5	7.54	2.1	3.42	0.5960	0.3724				

1.70 - 1.60	3335	3463	96.3	10.07	3.3	5.33	0.3941	0.1784				
Inf - 1.60	21134	21656	97.6	12.42	34.2	18.70	0.0716	0.0295				
Merged [A], lowest resolution = 28.93 Angstroms, 9567 outliers downweighted												

Estimation of SAD delta-F values												
Local scaling employed with radius of 0.2081 reciprocal Angstroms												
Anomalous signal/noise ratios (1.0 is random). The first line is based on												
input sigmas, the second on variances of F+ and F- (if not already averaged):												
Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.0 - 2.6 - 2.4 - 2.2 - 2.0 - 1.8 - 1.6 A	1.56	2.30	1.95	1.97	1.88	1.84	1.59	1.41	1.33	1.23	1.17	1.08
	1.67	2.16	1.85	1.82	1.56	1.34	1.35	1.32	1.21	1.12	1.08	1.03
68.9 Neighbors used on average for F+/F- local scaling												
Rint(anom) = 0.0500 before and 0.0503 after local scaling												
Current dataset contains 15967 SAD delta(F)												

Table II-7. Extract from XPREP (Bruker Nonius, 2002) output for the reflection data of crystal AE2. The data had been scaled but not merged with XDS. The two upper sections of the output show the overall statistics of the dataset. The third part shows the estimated $I/\sigma(I)$ values for the anomalous differences $|F_+|-|F_-|$. A typical cutoff value for phasing is where the anomalous $I/\sigma(I)$ falls under 1.3.

II.3.2 Structure solution and refinement of the AE2 crystal

The unexpected presence of anomalous scattering in the crystal was used to solve the phase problem. The program SHELXD was used to find a heavy atom substructure responsible for the anomalous scattering. Successful runs would always find one heavy atom in the asymmetric unit. It must be noted that SHELXD cannot determine the nature of the heavy atom or atoms present. One solution with a correlation coefficients CC_{all} of 33.90% for the strong data (used in the calculation) and CC_{weak} of 19.45% (for the data not used in the calculation) was used as input for SHELXE (Sheldrick *et al.*, 2001) to perform phase calculations and density modification. The enantiomer of the heavy atom substructure was

also considered in SHELXE since the heavy atom substructure is chiral in this spacegroup. 500 cycles of density modification with SHELXE with an estimated solvent content of 45% and using all data yielded a contrast of 0.70 (0.43 for the enantiomer), a connectivity of 0.94 (0.90 for the enantiomer) and a pseudo free correlation coefficient of 86.73% (81.87% for the enantiomer). The resulting electron density map (*Figure II-23*) showed interpretable electron density and was used for model building. Once the final refined model was available the weighted mean phase error of this experimental density map was calculated with SHELXPRO (Bruker Nonius, 2002) (*Figure II-24*). It was 39.9°.

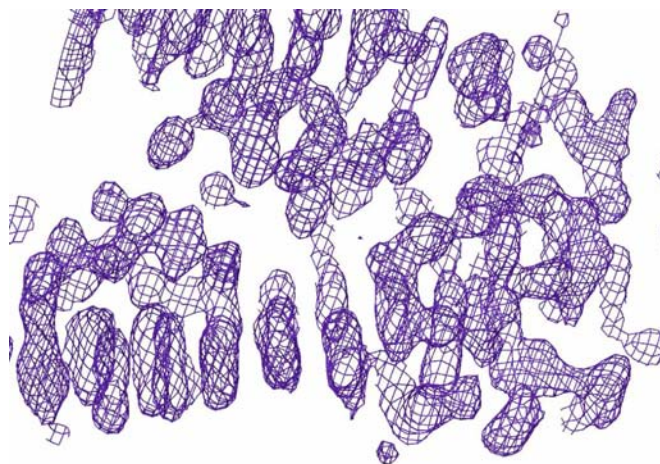


Figure II-23. Experimental electron density map of the AE2 crystal to 2.3 Å contoured at the 1 σ level used for model building. The density of one Echinomycin in side view can be identified in the lower-left quadrant, the bicyclic backbone of another in the right half. The density for all nucleobases was also clear. The connectivity of the map degraded if phases to higher resolution were included.

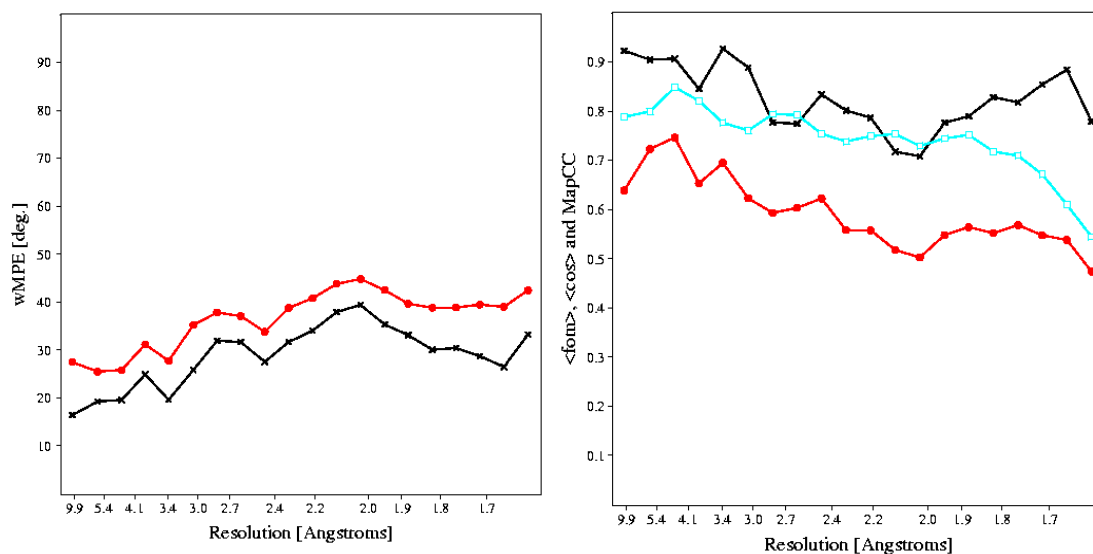


Figure II-24. Statistics of the experimental map for AE2 compared to the final refined model. Left: mean phase error, f.o.m. weighted (red) and F.f.o.m. weighted (black). Right: mean f.o.m. (blue), mean cosine of the phases errors (red) and map correlation coefficient (black).

A starting model was built by hand in the experimental electron density with XFIT (McRee, 1999). The asymmetric unit contains two DNA duplexes and two Echinomycin molecules in this crystal form.

Restrained full anisotropic refinement with riding hydrogens was carried out with SHELXL as for the AE1 crystal. The resulting data to parameter ratio of 2.07 and the resolution of 1.6 Å are relatively low for the introduction of an anisotropic model for the atomic displacement parameters but the anisotropic character of the diffraction intensities recommended testing it and it was retained thereafter. The change to the anisotropic model was justified by a decrease of the crystallographic residuals upon its implementation. The R factor decreased from 27.77% to 24.58% and R_{free} from 31.63% to 28.52% at that stage with greater improvements for the high resolution shells (*Table II-8*). The data to parameter ratio improves to a value of 3.37 during the anisotropic refinement (only the working set of reflections included) if the extra restraints are counted as observations.

The final model has $R = 17.96\%$ and $R_{\text{free}} = 22.15\%$ and includes 1784 atoms with 166 waters, one metal site and a MES molecule in the asymmetric unit.

<u>Isotropic model</u>											
Res. (Å)	1.60	1.67	1.74	1.82	1.92	2.04	2.20	2.44	2.79	3.57	inf
R1	0.466	0.434	0.383	0.338	0.340	0.309	0.283	0.269	0.238	0.251	
<u>Anisotropic model</u>											
Res. (Å)	1.60	1.67	1.74	1.82	1.92	2.04	2.20	2.44	2.79	3.57	inf
R1	0.308	0.267	0.270	0.253	0.263	0.255	0.232	0.225	0.216	0.240	

Table II-8. R factor in resolution shells before and after the introduction of the anisotropic model. The residual drops to 2/3 of its isotropic value at high resolution.

II.3.3 The nature of the atoms in the metal site

The nature of the anomalous diffracting atom or atoms was unknown through most of the refinement since no atom heavier than chlorine was included in the crystallization setup and chlorine was not expected to incorporate itself in the crystal in a well ordered position. The electron density showed an octahedrally coordinated site with four water molecules and N7 of Guanines G7 and G207 as ligands. The distances to the ligands are in the order of 2.0-2.2 Å and the ligand-metal-ligand angles are all close to 90°. A search for similar geometries in the CSD showed that at least Mn²⁺, Co²⁺, Ni²⁺, Cu²⁺ and Zn²⁺ are capable of forming such complexes, with nucleobase nitrogens as ligands and with similar geometries. A nickel cation was modeled as a first attempt and kept in the model for most of the refinement but tests showed that all of those transition metals could be included in the refinement without significantly affecting the geometry, the B factors or the crystallographic residuals.

It was decided to tackle the problem with anomalous scattering methods. The crystal used for the data collection of AE2 was not available at the time so two other of the same kind, also grown in the absence of magnesium and of any intentionally included heavy atoms were selected and mounted and frozen as for AE2.

The first one was grown at 293 K using the same stock solutions as for AE2 from a hanging drop with 20 µl of DNA-Echinomycin stock solution and 1 µl of reservoir solution consisting of 30% PEG 200, 0.1 M MES buffer at pH 6 and 0.02 M spermine tetrachloride. No PEG 3350 was included for that particular drop. The crystal was a 0.6 mm long and 0.15 mm wide needle and was used only for fluorescence scans.

The second crystal was grown at 283 K using the same stock solutions as for AE2, the hanging drop was composed of 20 µl of DNA-Echinomycin stock solution with a DNA concentration of 0.5 mg/ml as for AE2 and 1 µl of reservoir solution consisting of 30% PEG

200, 5% PEG 3350, 0.1 M MES buffer at pH 6 and 0.02 M spermine tetrachloride. The crystal was 1 mm long, 0.07 mm wide and was used for data collection at several wavelengths.

Laboratory datasets indicated that they were of the same crystal form as AE2 and it was expected that the same anomalous scatterers would be present for all three crystals.

Data from these crystals were collected at beamline X31 at EMBL/DESY, Hamburg.

Five fluorescence scans were performed around the absorption edges of zinc (energies from 9559 eV to 9759 eV were scanned), copper (from 8879 eV to 9079 eV), nickel (from 8232 eV to 8432 eV), cobalt (from 7609 eV to 7809 eV) and iron (from 7012 eV to 7220 eV). The hardware of the beamline did not allow to reach the manganese absorption edge. The data were very noisy (*Figure II-25*). Later it was found out that the crystal was somewhat damaged, but the fluorescence scans allowed the observation of the presence or the absence of an absorption edge as indicated by both, an absorption peak and a different baseline level at the high and low energy sides of the scans.

No peaks could be observed for iron, cobalt or copper. A weak but clear peak was found for the nickel absorption edge and a stronger one for zinc. It must be noted that the higher fluorescence counts for the zinc peak can be a result of a different hardware setup and beam characteristics at the shorter wavelength and does not necessarily indicate a higher proportion of zinc in the crystal.

Datasets with a total oscillation angle of 180° , 1° per image, were measured at wavelengths 1.00 Å, 1.362 Å, 1.47 Å and 1.52 Å, so that datasets at both the high energy and low energy side of the absorption edges of zinc, copper and nickel were available for comparison. The experimental parameters were adjusted in each dataset in order to get similar counting statistics for the strong reflections. All four wavelengths measured lay at the high energy side of the absorption peaks of cobalt and manganese so no discrimination between them is possible based on the anomalous diffraction present in the datasets. The f' values at those wavelengths for the heavy atoms expected to be present in the crystal are listed in *Table II-9* (Manfred S. Weiss, personal communication).

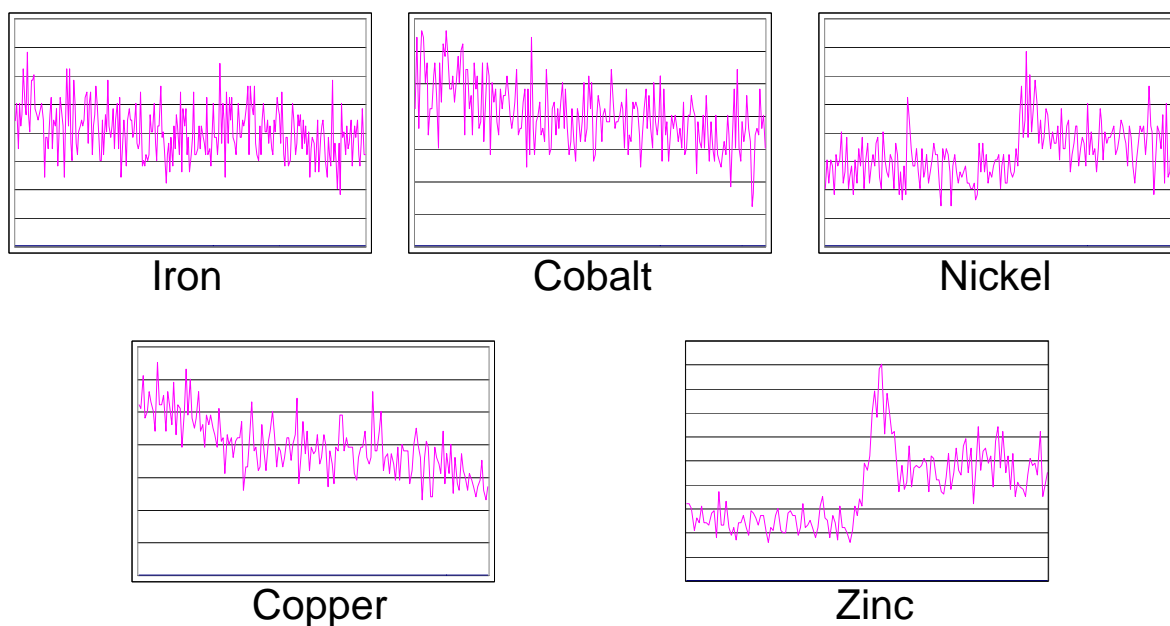


Figure II-25. Fluorescence scans for the absorption edges of iron, cobalt, nickel, copper and zinc. The fluorescence counts are in the vertical axis with arbitrary units, the wavelength in the horizontal axis with a non constant energy step and shorter wavelengths to the right. The low energy (left) and high energy (right) limits for each scan are indicated in the text.

$\lambda(\text{\AA})$	$f'(\text{Zn})$	$f'(\text{Cu})$	$f'(\text{Ni})$	$f'(\text{Co})$	$f'(\text{S})$	$f'(\text{P})$
1.0000	2.55	2.29	2.02	1.78	0.24	0.19
1.3620	0.54	3.80	3.37	2.98	0.44	0.34
1.4700	0.62	0.54	3.82	3.33	0.51	0.40
1.5200	0.66	0.58	0.50	3.53	0.54	0.42

Table II-9. Estimated f' values (in electrons) for the anomalous diffraction of the atoms possibly present in the AE2 crystal. The 1.00 \AA wavelength was chosen far from the absorption edge of Zn upon considerations of data quality.

The diffraction limit was 2.8 \AA at the best direction of the best dataset. Since only the anomalous differences were of interest it was decided to apply a common resolution cutoff to all datasets avoiding the negative effect of the weakest reflections on profile fitting and scaling.

All four datasets were integrated with Denzo with a resolution cutoff of 4 \AA and scaled with SADABS. The diffraction images showed anisotropic diffraction, anisotropic diffraction profiles and the presence of small satellites. As a result the quality of the datasets was

suboptimal, especially for extracting anomalous differences, with R_{int} values (from SADABS) of 8.47% at 1.00 Å, 9.66% at 1.362 Å, 10.00% at 1.47 Å and 9.95% at 1.52 Å.

The anomalous differences from those datasets were combined with phases φ_c calculated from the final model of AE2 to obtain anomalous Fourier maps. First the standard procedure implemented in the program FFT (Collaborative, 1994) was used. It makes a Fourier synthesis with amplitudes $F=|F_{(+)}|-|F_{(-)}|$ and phases $\varphi=\varphi_c -90^\circ$. The resulting electron density peaks in the anomalous maps had peak heights of 5.13 σ at 1.00 Å with the peak as the top of the peaklist, 2.72 σ at 1.362 Å with the peak not ranking between the 50 highest of the map, and 3.34 σ at 1.47 Å (fifth highest peak of the map). In the 1.52 Å datasets there was no peak at the metal position up to a level of 2 σ . Even if the signal is marginal, it gets enhanced above the absorption edges of nickel (the dataset at 1.47 Å) and zinc (1.00 Å) in good agreement with the fluorescence scans.

A new facility in SHELXE (local test version) was also used to calculate anomalous Fourier maps on those four datasets. First, XPREP uses the sign of the anomalous differences to estimate a phase shift between the final model phases and the heavy atom phases (see introduction) of either 90° or 270° for the acentric reflections. Then several cycles of density modification are run with SHELXE starting from the phases of the final refined model in order to obtain improved weights. The phase shift is applied backwards from those calculated phases to obtain phases for the heavy atom substructure. Five cycles of density modification with a solvent fraction of 0.4 and an artificial B-value of 30 Å² were run for all four datasets (Table II-10).

$\lambda(\text{Å})$	Electron density at the metal site (σ)	Electron density at the peak (σ)	Peak - metal site distance (Å)	Highest other peak in the map (σ)
1.00	10.84	10.9	0.13	3.9
1.362	4.90	5.1	0.48	4.2
1.47	6.08	6.8	0.77	4.9
1.52	1.03	-	-	4.5

Table II-10. Statistics for the peak at the metal site in the anomalous maps from SHELXE.

The statistics are better than for the calculation with FFT but the results are qualitatively similar. All three significant peaks are now above the second peak of the map. The improvement in map quality can probably be attributed to the weighting scheme although solvent flattening could also have improved the phases for the very low resolution reflections.

Even in the anomalous map from SHELXE there are no peaks at the position of the phosphorus and sulphur atoms as would be expected for better datasets. A reference level for the estimation of the anomalous electron density at the peak is also lacking and only a qualitative comparison of the peak heights is possible. Assuming a metal site composed of 50% nickel and 50% zinc, the added f'' values from both atoms are $4.6 e^-$ at 1.00 \AA , $4.0 e^-$ at 1.362 \AA , $4.7 e^-$ at 1.47 \AA , $1.2 e^-$ at 1.52 \AA . The relatively higher peak height (in standard deviation units) at 1.00 \AA can be attributed to the higher quality of that dataset at a much shorter wavelength than the others, reducing the noise in the map. The absence of a peak in the 1.52 \AA datasets discards the presence of significant amounts of cobalt or lighter transition elements at the site.

It was decided that the quality of the datasets does not allow for a more precise interpretation of the occupancies. A higher zinc occupancy would also explain the higher sigma level for the 1.00 \AA dataset relative to the 1.47 \AA one, but more datasets (and more synchrotron time) would be necessary to confirm that. A metal site with fixed 50% occupancies for both, nickel and zinc was adopted for the final refinement cycles.

It was not attempted to calculate double difference anomalous maps, or *difference DANO maps* (Than *et al.*, 2005), since it was estimated that the noise of the diffraction data was too high and it would only get enhanced by taking double differences of the intensities.

Metal additives were also tested as additives in crystallization experiments based on the conditions of growth of AE2. MnCl_2 as an additive would only produce a different crystal form while FeCl_3 , CoCl_2 and NiCl_2 inhibited crystal growth at the concentrations tested (1 mM and 0.1 M). With ZnCl_2 as an additive in the conditions of growth of AE2 modified only in the drop volumes, with drops composed of $15 \mu\text{l}$ of stock solution + $1 \mu\text{l}$ of reservoir solution in this case, crystals of the AE2 type were obtained at 0.5 mM Zn^{2+} concentration but a different crystal form was obtained at 1 mM Zn^{2+} concentration.

Copper as an additive was more extensively tested in a series of crystallization experiments based on the crystallization conditions of AE2 but scaled down in size to $500 \mu\text{l}$ of reservoir and drops composed of $10 \mu\text{l}$ of the DNA-Echinomycin stock solution and $0.5 \mu\text{l}$ of reservoir solution. The concentration of CuCl_2 as an additive (in the reservoir) was varied from 0 to 1 mM (*Figure II-26*). It can be seen that Cu^{2+} promotes crystal growth at moderate concentrations but inhibits it at higher ones. The failure to produce crystals must not necessarily be interpreted as a negative result since it was also frequently observed in drops without any additive. A high concentration of copper favors nucleation with the number of

crystals increasing and the crystal size decreasing correspondingly until crystal growth is inhibited at 0.8 mM concentration.

The crystals obtained with copper as an additive are of the same crystal form as AE2. It can only be interpreted that copper was not found in the AE2 type crystals because it was not present in the crystallization cocktail, however, it can also be incorporated in the crystal lattice, presumably in the same position as nickel and zinc. The inhibition of crystal growth at high concentrations of copper suggests that it interacts with the complex and displaces nickel and zinc from it. It also opens the possibility that other metals might also get incorporated in the lattice, the negative results in which they failed to promote crystal growth may be statistical artifacts or the effect of an excessive concentration, probably the case with nickel.

Additional experimental evidence would be necessary, but it can be speculated that the metal binding site in the AE2 crystal is a promiscuous one capable of binding several transition metal cations based on availability and not on specificity of binding.

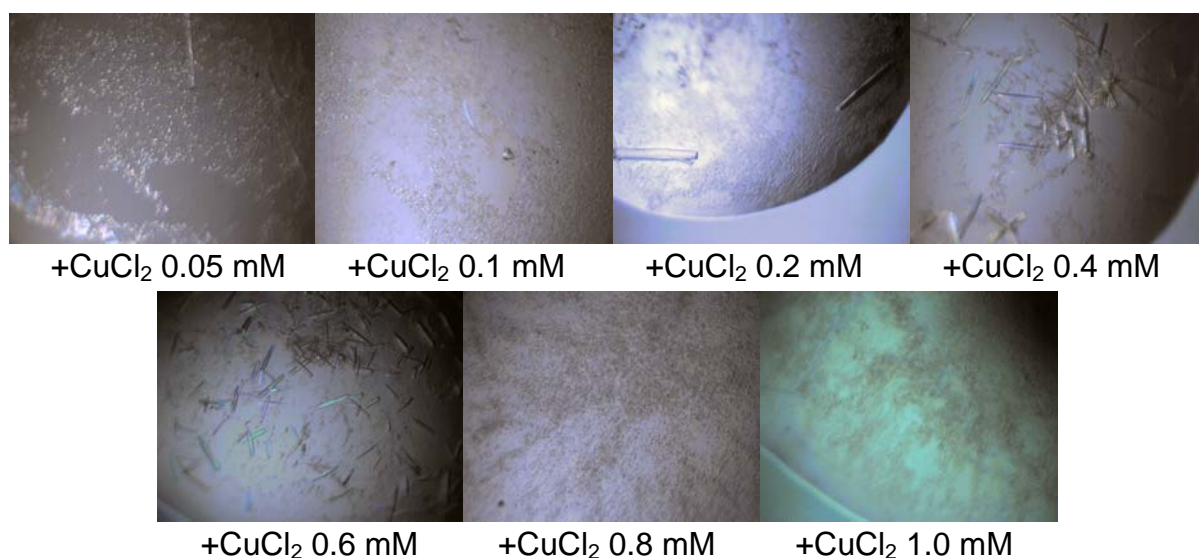


Figure II-26. Microscope photographs of the drops with CuCl_2 as additive. The scale is the same in all pictures.

II.3.4 Crystal structure of AE2

Overall description of the structure.

The asymmetric unit in the AE2 crystal is composed of four independent DNA chains with the selfcomplementary sequence ACGTACGT forming two antiparallel duplexes with four Echinomycin molecules bisintercalating all four CG steps from the minor groove. The

four DNA molecules are here numbered as: A1 to T8 for the first strand, A101 to T108 for the strand antiparallel to the former and A201 to T208 and A301 to A308 for the other duplex. Echinomycin residues are numbered from 9 to 19 and 109 to 119 for the ones in the first duplex and from 209 to 219 and 309 to 319 for those binding to the second DNA duplex (*Figure II-27*).

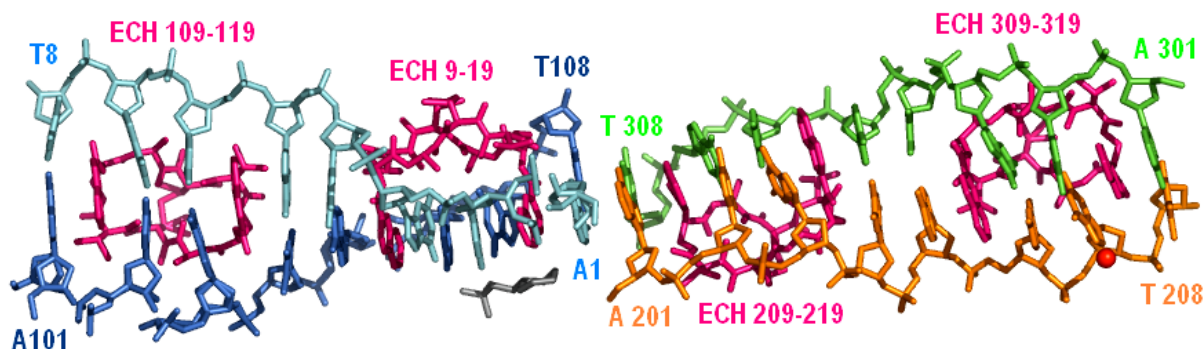


Figure II-27. Asymmetric unit of the AE2 crystal. Water molecules have been omitted for clarity. The four independent DNA chains are shown in cyan, blue, green and orange, the Echinomycins in pink, the red ball is the metal site and the MES molecule is represented in grey.

Several copies of all important molecules are present in the asymmetric unit of AE2 allowing for an insight into the structural flexibility of the complex that was absent from all other crystal structures analyzed earlier in this chapter. It should also be kept in mind that crystals of the AE1 form grew sometimes in the same drops as those of the AE2 form, effectively representing a third independent copy of the complex of Echinomycin with DNA duplex with this sequence.

A metal cation forms an octahedral complex with four water molecules and N7 of G207 and N7 of G7 of a symmetry equivalent duplex, linking the two DNA duplexes together. The nitrogens of the bases occupy axial positions in the complex with distances to the metal atom of 2.22 Å for N7 of G207 and 2.10 Å for N7 of G7* of the symmetry equivalent. This metal site was modeled as 50% nickel / 50% zinc with the occupancies adding to one and the positions and atomic displacement parameters of both metals constrained to be equal.

Also 86 waters and a MES molecule are included in the model of the asymmetric unit.

Base pairing scheme in AE2

This is the first crystal structure of a complex between a quinoxaline antibiotic and duplex DNA where the bases surrounding the intercalation site show Watson-Crick base pairing. All

eight basepairs internal to the 4 bisintercalation sites are Watson-Crick paired as usual but only five of the eight external basepairs are in the Hoogsteen mode. Adenines 5, 101 and 105 are all in the *original* (from B-DNA) *anti* conformations and involved in standard Watson-Crick basepairs with Thymines 104, 8 and 4 respectively (*Figure II-28*).

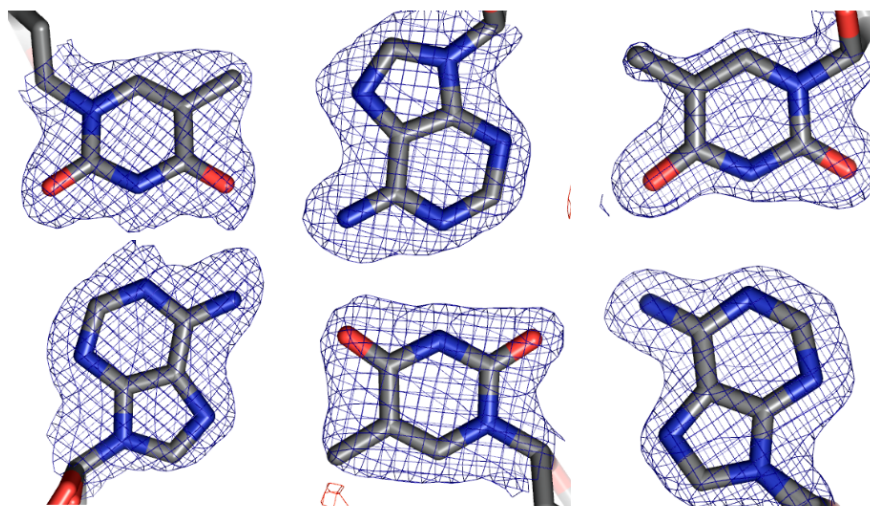


Figure II-28. Electron density for basepairs T4-A105 (left), A5-T104 (centre) and T8-A101 (right). The blue density corresponds to the $2F_o-F_c$ map contoured at the 1.5σ level. The red patches of electron density are part of the difference Fourier map at the -3σ level. There was no difference density at $+3 \sigma$ in the pictured areas.

All three cited basepairs are in the same duplex, actually there is only one Hoogsteen basepair in it. Significantly, the corresponding basepairs in the other duplex are all in the Hoogsteen mode. This is the result of intrinsic flexibility in the structure and not a crystal artifact. The B values are in the same range for corresponding basepairs in both duplexes irrespective of whether they are in the Watson-Crick or in the Hoogsteen configuration. Basepairs T8-A101 and T208-A301 are terminal and involved in crystal contacts. They are also contiguous to bases G7 and G207 which undergo binding to the metal site with potential distorting effects on the geometry. Basepairs T4-A105 and A5-T104 though are in the centre of the duplexes, protected from any potential conformational end effects, and are not involved in any crystal packing contacts. The molecular geometry in that area is most likely representative of the complexes in solution for that particular kind of sequence. Both binding modes are represented without apparent conformational disorder in the structure and are probably available to DNAs between two contiguous Echinomycin binding sites.

All hydrogen bonded distances corresponding to the base pairing in the complexes are listed in *Table II-11*. The abnormally short values in the range of 2.5 \AA correspond to the

most disordered area of the structure, for residues A305, C306 and G307, with higher than average B values in the order of 50 Å². There is no significant residual density in the difference map in this area and it was not necessary to model double conformations, but the geometry of that part of the model should be interpreted with caution.

Atoms	Distance (Å)	Type	Atoms	Distance (Å)	Type
A1_N6...T108_O4	2.76	HG	A201_N6...T308_O4	3.15	HG
A1_N7...T108_N3	3.06		A201_N7...T308_N3	2.85	
C2_O2...G107_N2	3.15	WC	C202_O2...G307_N2	2.79	WC
C2_N3...G107_N1	3.03		C202_N3...G307_N1	2.84	
C2_N4...G107_O6	3.00		C202_N4...G307_O6	2.51	
G3_N1...C106_N3	2.79	WC	G203_N1...C306_N3	2.80	WC
G3_N2...C106_O2	2.88		G203_N2...C306_O2	2.88	
G3_O6...C106_N4	2.67		G203_O6...C306_N4	2.49	
T4_N3...A105_N1	2.69	WC	T204_N3...A305_N7	2.83	HG
T4_O4...A105_N6	3.03		T204_O4...A305_N6	3.06	
A5_N1...T104_N3	2.89	WC	A205_N6...T304_O4	2.94	HG
A5_N6...T104_O4	2.86		A205_N7...T304_N3	2.90	
C6_O2...G103_N2	2.80	WC	C206_O2...G303_N2	2.89	WC
C6_N3...G103_N1	2.85		C206_N3...G303_N1	2.86	
C6_N4...G103_O6	2.72		C206_N4...G303_O6	2.77	
G7_N1...C102_N3	2.99	WC	G207_N1...C302_N3	2.95	WC
G7_N2...C102_O2	2.93		G207_N2...C302_O2	2.98	
G7_O6...C102_N4	2.96		G207_O6...C302_N4	2.88	
T8_N3...A101_N1	2.90	WC	T208_N3...A301_N7	2.89	HG
T8_O4...A101_N6	2.86		T208_O4...A301_N6	2.94	

Table II-11. Hydrogen bonding distances for the base pairing in AE2. WC stands for Watson-Crick and HG for Hoogsteen base pairing.

The hydrogen bonding between the Echinomycins and the guanines resembles that of the other crystals (*Table II-12*), again all four hydrogen bonds are found for each Echinomycin. It should be noted that the orientation of Echinomycin 309-319 is inverted relative to the others and thus has different hydrogen bonding partners in the table. For this Echinomycin, as in GE1, the bridge has been modeled disordered with occupancies of 64% and 36% reflecting the presence of both possible orientations in the crystal. The inversion only represents that the main conformation, that with the highest occupancy, is reversed relative to the highest occupied orientation of the other copies of the antibiotic in the asymmetric unit and not a change in the nature of the binding.

Donor	Acceptor	Distance (Å)
Ala-12 N	G-107 N3	2.96Å
G-107 N2	Ala-12 O	3.00Å
Ala-17 N	G-3 N3	2.84Å
G-3 N2	Ala-17 O	3.14Å
Ala-112 N	G-7 N3	2.89Å
G-7 N2	Ala-112 O	3.05Å
Ala-117 N	G-103 N3	2.93Å
G-103 N2	Ala-117 O	3.26Å
Ala-212 N	G-307 N3	3.04Å
G-307 N2	Ala-212 O	3.26Å
Ala-217 N	G-203 N3	3.05Å
G-203 N2	Ala-217 O	3.13Å
Ala-312 N	G-303 N3	3.06Å
G-303 N2	Ala-312 O	3.12Å
Ala-317 N	G-207 N3	2.95Å
G-207 N2	Ala-317 O	3.10Å

Table II-12. Echinomycin-DNA hydrogen bonding distances in AE2.

Crystal packing

The asymmetric unit of AE2 contains two duplexes of DNA and four Echinomycins. Both duplexes stack on each other with the nucleobase of A201 stacking on that of A1 and the bases of T308 and T8 also stacking on each other. In total 16 basepairs stack in the asymmetric unit (*Figure II-27*), however, the two duplexes do not form a pseudocontinuous double helix since the individual chains run in opposed senses at the site.

At one end of the column of 16 basepairs the basepair formed by T208 and A301 stacks on their symmetry equivalents in a similar way: thymine on thymine and adenine on adenine with inversion of the sense at the point of stacking; the total of basepairs stacked on each other is subsequently 32. The basepair A101-T8 does not stack on other nucleobases and thus this crystal lack the infinite columns observed in GE1, GE2, AE and CE.

The basepair T208-A301 is also involved in a lateral contact with basepair T8*-A101* of a symmetry equivalent complex. N6 of A301 has a hydrogen bond at 2.96 Å distance to O4 of T8* and O4 of T208 has a longer one at 3.37 Å to N6 of A101*.

Base G207 N7 is involved in binding to the metal site with N7 of G7* as the ligand in the opposite position of the coordination sphere of the metal (*Figure II-29*). The geometry of those bases does not seem to be strongly distorted by binding to the metal site. This G7* belongs to the same symmetry equivalent duplex that binds to basepair T208-A301. The two complexes depart from each other at an angle after this basepair.

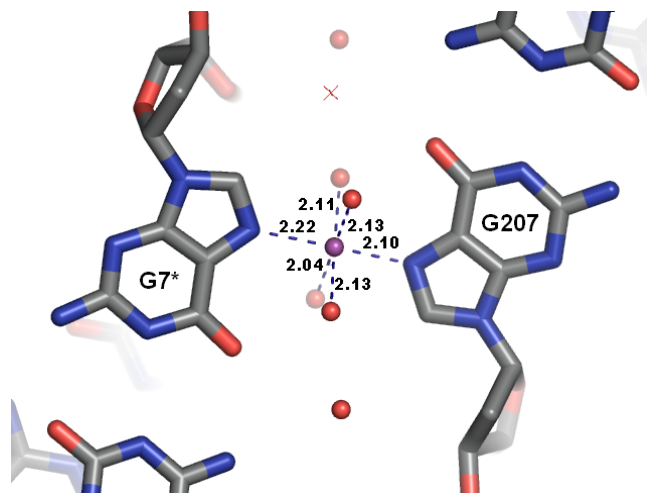


Figure II-29. The metal site has a nearly perfect octahedral geometry. It links two different DNA duplexes laterally.

Nucleotides 202-206 and 303-307 are not involved in crystal packing contacts.

N-methyl-cysteine 218 makes Van der Waals contacts to N-methyl-valine 114* and the backbone of T8* of one and the same symmetry equivalent duplex which is also involved in packing contacts to basepair A201-T308. T308 O4 makes a short hydrogen bond at 2.59 Å to the carbonyl oxygen of N-methyl-cysteine 113* while A201 is involved in four Van der Waals contacts to the same Echinomycin as T308 and to G7* in that complex.

The other DNA duplex comprising A1-T8 and A101-T108 makes a bigger number of crystal packing contacts but the part contacting the duplex A201-T208 and A301-T308 does not. Nucleotides A1-T5 and A104-T108 have no crystal contacts except water mediated or through the MES molecule.

The Echinomycin molecule associated with the step C102-G103 makes four crystal contacts to one symmetry related DNA-Echinomycin complex with the bridge sulphur of N-methyl-cysteine 118* at 3.49 Å of T308* O3' and the carbonyl oxygen of N-methyl-cysteine 113 at 2.59 Å of the same atom. Ser 211 O has a long contact to A 201* O5' while N-methyl-cysteine 218* O is in Van der Waals contact to both, N-methyl-valine 114 CG2 and C6 C5'.

G7 N7 is involved in complex formation in the metal site as mentioned before.

Lastly, A101 is stacked on one ester linkage of an Echinomycin and makes a hydrogen bond from N6 to T208* O2 of that complex. It also makes a Van der Waals contact to the sugar of that nucleotide. T8 on its side stacks on the sugar of A301* and has O4 forming a hydrogen bond to A306* N6 of a different symmetry equivalent as A101.

From a bigger perspective the structure has two kinds of continuous solvent channels along the c axis, the bigger of them with an approximate cross section 55 Å long and 35 Å wide (*Figure II-30*).

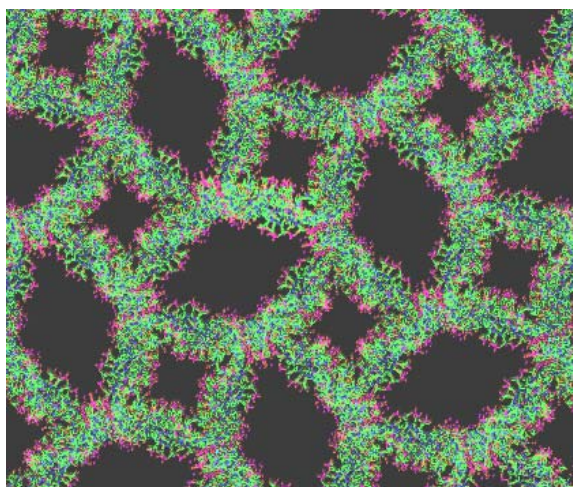


Figure II-30. The crystal packing forms two types of solvent channels along the c axis. All phosphates (in pink) are oriented towards the solvent channels.

II.3.5 Echinomycin structure in AE2

The structure of all four Echinomycins in this crystals closely resembles that of GE1 or AE1. All four echinomycins in AE2 were individually superposed on that of AE1 after removing the second conformations. The resulting RMSDs are 0.34 Å, 0.53 Å, 0.21 Å and 0.16 Å (*Figure II-31*). The conformation of the backbone is virtually the same in all cases and the quinoxalines are all in similar positions. The biggest differences are again found at the ester bonds and the valine side chains.

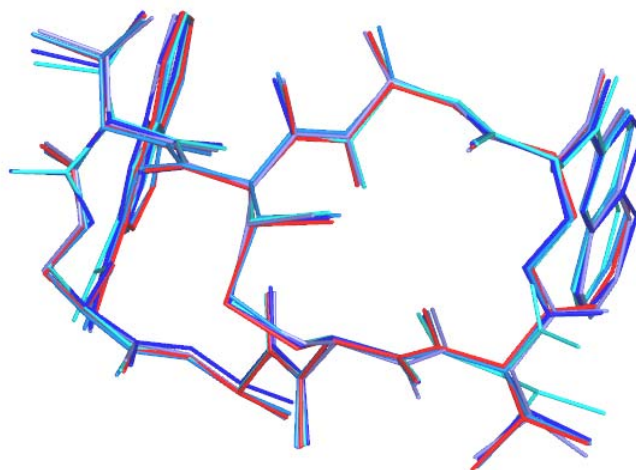


Figure II-31. Superposition of all four echinomycins in the AE2 crystal (in different blue tones) on that of AE1. Only the main conformations were considered.

As for GE1, some of the Echinomycins have two orientations available for binding to the DNA and as for GE1, this was modeled as disorder only in the thioacetal bridge. In Echinomycin 9-19 the occupancies for both orientations refined to values of 81% and 19%. There is some process favoring one of them but the antibiotic undergoes no crystal contacts. This must be attributed to the conformation of the DNA acting at a distance, possibly transmitted to the bridge through the solvent.

Echinomycins 109-119 and Echinomycin 209-219 both undergo several crystal packing contacts, some involving the N-methyl cysteines. Still, the former shows both possible conformations while the latest does not. Echinomycin 309-319 is also present in both orientations and is not involved in contacts external to its complex. The methyl group in the bridge could not be located in the electron density and was omitted from the model of all Echinomycins. Whichever the forces that control the selection of one or the other orientation of Echinomycin relative to the DNA, they are of a subtle nature.

II.3.6 DNA structure in AE2

The reference structure for analyzing the DNA conformation in AE2 must necessarily be AE1 with the same DNA sequence and able of growing in the same conditions, even the same drops as AE2, but diffracting to a higher resolution. AE1 has already been described as overall similar to GE1 and GE2 before in this chapter.

A superposition of the first DNA duplex (residues 1-108) of AE2 on the crystallographic duplex of AE1, taking only the main conformations into account, gives an RMSD of 2.15 Å.

For the second duplex, comprising residues 201-308 the RMSD is of 1.39 Å, smaller as now all basepair have the same configuration as in AE1, but still significant. The superpositions are shown in *Figure II-32*.

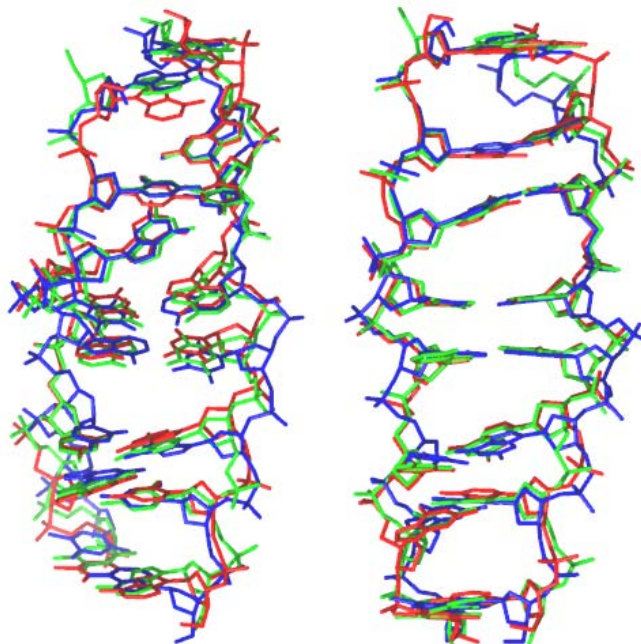


Figure II-32. Superposition of the DNA duplexes of AE2 on that of AE1. The duplex comprising residues 1-108, the one with the Watson-Crick basepairs outside the bisintercalation site, is in blue. The second duplex from AE2 is in green and that of AE, the reference, in red. The orientation is similar in both images, with adenine A1 up-left.

The in AE2 DNA adopts only roughly the same conformation as in AE1. The position of the bases is similar in the direction of the helix axis but not so much perpendicular to it. The match is also better internal to the intercalation site than external to it, with big differences in the terminal basepairs (all except one involved in packing contacts, also in AE1). The terminal basepair again, as in AE1, shows freedom to slide perpendicularly to the helix axis without disrupting the complex. Basepairs A1-T108 and A201-T308 in AE2 lie further away from the quinoxaline as in AE1, specially the thymines, but they still incur in some too close contacts to the quinoxaline at distances of 2.99 Å and 3.05 Å respectively.

Interestingly, the bases G7 and G207 involved in complexing to the metal do not show more distortion relative to AE1 than the others.

The Watson-Crick basepairs outside the bisintercalation sites, with adenines A5, A101 and A105 in the *anti* conformation, do not significantly affect the position of the bases (naturally it affects the orientation), but displace the sugars and the phosphates further away from the helix axis (*Table II-13*). It has already been mentioned that adenine A101 is also

involved in several crystal packing contacts but A5 and A105 are not and should not be strongly distorted. Although basepairs T204-A305 and A205-T304 are in the canonical Hoogsteen conformation, they exhibit high B factors, 45.69 Å² in average for the base of T304 and 45.80 Å² for the base of A305. This could reflect lack of stabilization of the Hoogsteen conformation and the potential ability to change it.

Basepair	C1'-C1'distance (Å)	Type	Basepair	C1'-C1'distance (Å)	Type
A1-T108	8.88	HG	A201-T308	8.08	HG
C2-G107	10.65	WC	C202-G307	10.47	WC
G3-C106	10.61	WC	G203-C306	10.74	WC
T4-A105	10.17	WC	T204-A305	8.24	HG
A5-T104	10.47	WC	A205-T304	8.31	HG
C6-G103	10.54	WC	C206-G303	10.53	WC
G7-T102	10.65	WC	G207-T302	10.49	WC
T8-A101	10.50	WC	T208-A301	8.32	HG

Table II-13. C1'-C1' distances through the basepairs in AE2. In red the three basepairs external to the bisintercalation site showing the Watson-Crick base pairing scheme. HG stands for Hoogsteen base pairing, WC for Watson-Crick base pairing.

All basepairs internal to the bisintercalation sites show buckling angles similar to those of AE1 or GE1, in the range of 20°-25°. The AT basepairs have buckling angles close to 0° except for T8-A101 where it is 14.87°. It is not casual that this is the only basepair in GE1, GE2, CE, AE1 or AE2 external to the bisintercalation sites with a buckle angle greater than 7°. It is the only basepair not involved in base stacking at both sides, it stacks to an ester linkage and a sugar in the terminal face of the bases. It has thus greater conformational flexibility than the other basepairs. This basepair also suffers of compression from the quinoxalines, with a minimum distance of 3.00 Å, this is alleviated by buckling it away from the intercalator. It can be speculated that DNA in solution may be capable of adopting a similar solution to the problem of lack of space around the quinoxalines, the deformation from ideal planarity can be progressively absorbed by subsequent basepairs. No other basepairs of these crystal structures can adopt such a solution since the next basepairs, belonging to different duplexes, also stack on a second quinoxaline.

As already mentioned for GE1, the presence of Hoogsteen basepairs disrupts the standard definition of the twist angle based on the orientation of the bases for a Watson-Crick basepair.

Approximate twist values can be derived by using pseudo torsion angles involving the four C1' atoms of every two consecutive basepairs. The unwinding angles caused by the Echinomycins derived in that way are 48.67° between basepairs A1-T108 and T4-A105, 56.42° between A5-T104 and T8-A101, 62.20° between A201-T308 and T204-A305 and 63.18° between A205-T304 and T208-A301. Again the total unwinding angles would have to be increased by approximately 5° for the following basepairs at both sides as derived from half the unwinding for the central basepairs of the duplexes.

The presence of only Watson-Crick basepairs around Echinomycin 109-119 offers the possibility of measuring the unwinding angle caused by that Echinomycin using the standard definition of the twist angle as recommended by the NDB (Olson *et al.*, 2001). The program 3DNA (Lu & Olson, 2003) measures twist values of 28.94° between T4-A105 and A5-A104, 27.17° for the A5-C6 step, 10.44° for C6-G7 and 15.14° between G7-C102 and T8-A101. Adding the twist values and comparing to a standard value of 36.1° for B-DNA an unwinding angle of 55.55° results for the four basepair bisintercalation site, plus 3.54° extra at both sides extrapolating from the T4-A5 step for an estimated total of 62.63°.

The sugar puckering angles show a predominance of C2'-endo and closely related conformations (17 deoxyriboses) but C4'-exo and C3'-endo are also frequent (8 riboses).

Most glycosidic angles χ are clustered around an angle of -100° except those of the five Hoogsteen alanines, which have χ angles between +71.7° and +76.7°.

II.3.7 Discussion, biological implications and future perspectives

The nature of the heavy atoms present in the metal binding site of AE2 can be considered elucidated, but the question of the origin of those heavy atoms remains speculative. The behavior of the crystals of the AE2 form, growing only after a drop with an very high amount of DNA stock solution had been concentrated by vapour diffusion, suggests that an impurity present in that stock solution is responsible for it, and that the crystals would only start growing after a critical concentration of the metal cation has been achieved.

AE2 is the first crystal structure of a complex between a quinoxaline antibiotic and duplex DNA where some of the basepairs external to the bisintercalation site show Watson-Crick base pairing. NMR studies on the same DNA sequence showed Hoogsteen base pairing at 274 K and a tendency to change to Watson-Crick at higher temperatures. It is very interesting that the results from crystallography and X-ray diffraction start to converge in this area of knowledge. Both techniques suggest that a certain degree of flexibility exists in those

complexes, although the NMR results show Watson-Crick pairing more frequently. Even if other techniques for analyzing DNA conformation in solutions cannot find evidence of Hoogsteen base pairing, it should not be disregarded that it may be present, if not in molecules free in solution, maybe in complexes with proteins, the environment in which the biological effects of the binding of quinoxaline antibiotics take form. Most footprinting experiments have been carried out at neutral pH, opening the possibility that pH dependent behavior has been overseen. The results presented here show also that the Hoogsteen basepairs detected in the decade of 1980 were not an artifact of X-ray crystallography, with tight packing allowing also for the Watson-Crick mode. Crystal packing artifacts are only expected to produce conformations that, maybe not being the one of lower energy, are at least available to the molecules without a severe energetic penalty, and are thus representative of possible states of the molecules.

Two new crystal forms have been obtained with Zn^{2+} and Mn^{2+} as additives based on the conditions for growth of AE2 like crystals. Manganese promoted the growth of crystals belonging to space group $P4_12_12$ or $P4_32_12$, while zinc at 1 mM concentrations promoted the growth of crystals belonging to spacegroup $I4_1$. Both crystal forms were obtained shortly before this work was finished. Only laboratory datasets are available and synchrotron data collection is planned. Once these structures are solved, they will add to the data presented here and to our understanding of the conformational flexibility the complexes of DNA with quinoxaline antibiotics.

An issue worth mentioning is that all crystallographic studies on quinoxaline antibiotics complexed to duplex DNA have been carried so far on DNA hexamers or octamers, with two bisintercalation sites very close to each other. Consequently, the results are plagued by end effects on the DNA conformation and by both binding sites influencing each other as well as from crystal packing. Although the short sequences simplify the process of obtaining diffraction quality crystals, they deviate somehow from the important biological question, namely, which is the conformation of the duplexes in solution, in a non distorted DNA that is not influenced by end effects and at non saturation concentrations of the antibiotic, where binding sites are expected to be typically more than two basepairs away from each other. It would be interesting to perform crystallographic studies on longer DNA molecules, with only one isolated echinomycin binding site in the center of the nucleic acid. That could help to obtain a more precise picture of the distortions effected on the DNA by this kind of antibiotics, a much needed information for assisting in the process of drug design.

III CRYSTAL STRUCTURE OF AN ALTERNATING D-ALANYL, L-HOMOALANYL PNA

III.1 1. Introduction

III.1.1 PNAs as nucleic acid analogs.

PNA stands for Peptide-Nucleic-Acid, synthetic molecules that mimic more or less closely the structure of natural nucleic acids. They are polymeric structures with a peptide based backbone. The position of the side chains in the polymer is occupied by the natural nucleobases of nucleic acids, Adenine, Cytosine, Guanine, Thymine and Uracil, or alternatively chemically modified nucleobases. The word acid in the name is just an analogy to the molecules that they are expected to mimic, many are not acids.

Some PNAs can bind to DNAs and RNAs, in those cases uses in therapy are possible. With or without the capability of binding to natural nucleic acids, they can serve as a model for nucleic acids. The presence identical nucleobases in the molecule allows PNAs to adopt the same hydrogen bonded binding modes as nucleic acids. They also have the same kind of base stacking interactions. Different PNAs have different backbones, the possible synthetic variations are infinite. Most importantly, PNAs lack the phosphates of the backbone of nucleic acids, this way the electrostatic effects, which are omnipresent in nucleic acids, get decoupled of all the other conditionants of their structure. Furthermore, the geometrical constraints and steric impediments caused by the ribose backbones are absent in PNAs. Those features make PNAs useful model molecules for testing our understanding of the physical forces modeling natural nucleic acids, for developing methods to predict their structure and for creating new ways of influencing their function.

Even a possible role of PNAs in the origin of life has been proposed (Böhler *et al.*, 1995), as they constitute an alternative non-phosphodiester genetic material.

III.1.22-aminoethyl-glycine PNAs

A whole new field of research that is now penetrating all scientific fields related to nucleic acids got started with a publication describing a synthetic molecule able of recognizing DNAs with sequence selectivity (Nielsen *et al.*, 1991). This settled the polymeric

structure of what has been the base of most PNAs to date (*Figure III-1*), an artificial bipeptidic polymeric backbone with nucleobases attached to the side chains.

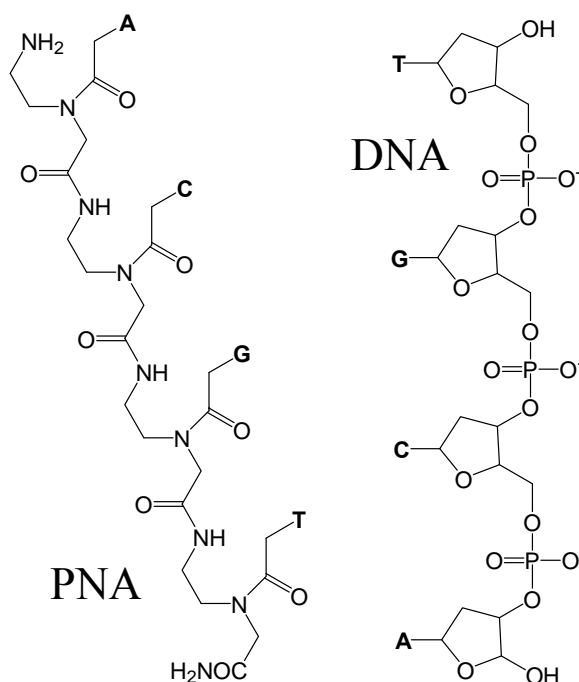


Figure III-1. Chemical structure of 2-aminoethyl-glycine PNAs compared to that of DNA.

The number of bonds in-between the nucleobases and their type are similar in both cases.

This kind of PNAs combine the ability of binding to nucleic acids with unique chemical properties: They are non ionic, achiral and resistant to peptidases and nucleases (Ray & Nordén, 2000).

This kind of PNAs binds to DNAs and RNAs forming a variety of structures. PNAs form duplex structures with themselves (Rasmussen & Sandholm, 1997) and with natural nucleic acids. The binding to DNAs and RNAs is associated with an increase in duplex stability relative to equivalent DNA or RNA duplexes and with increased sequence selectivity (Ray & Nordén, 2000). PNAs also form triplexes and quadruplexes with natural nucleic acids, mostly adapting to the host structure.

The high binding selectivity, the ability to displace natural nucleic acids from their duplexes and the resistance to enzymatic hydrolysis turn this kind of molecules into promising therapeutic agents. PNAs can interact with a particular sequence in a particular gene, possibly detecting mutations in it. Thus they can theoretically interfere *in vivo* with the transcription of a gene (antigene strategy) or alternatively with translation from the mRNA (antisense strategy); in both cases inhibiting the synthesis of selected proteins in the cell. Widespread use of those techniques has been hampered by the difficulty of delivering PNAs

into the cell or the cell nucleus (Ray & Nordén, 2000). It should be possible to overcome this hurdle in the process, then a bright future will unfold for the therapeutic use of PNAs.

The binding properties of 2-aminoethy-glycine PNAs turn them into a powerful tool for molecular biology, with potential and proven uses in PCR, Southern hybridization and in nucleic acid purification among others (Ray & Nordén, 2000). They are also a powerful diagnostic tool with applications in single basepair mutation analysis and as a DNA biosensor with a high sequence specificity (Ray & Nordén, 2000).

III.1.3 Alanyl PNAs

Another family of PNAs are alanyl PNAs. They have a polyalanine backbone with the nucleobases attached to the β -carbons of the side chains. The alanines are alternating D and L-aminoacids, if the peptide adopts a β -strand like structure all bases are located at the same side (Figure III-2).

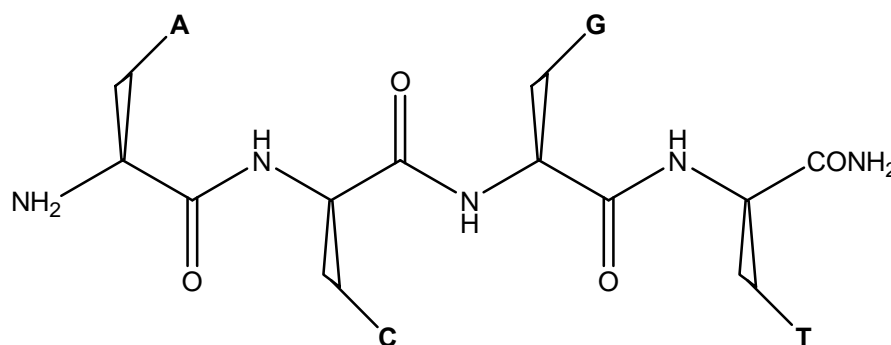


Figure III-2. Chemical structure of an alternating D-L alanyl PNA.

The distance between nucleobases is the same as between α -carbons, 3.6 Å (Dietrich *et al.*, 2001), only slightly longer than in nucleic acids. Alanyl PNAs are thus able of adopting the ideal 3.4 Å base stacking distance within the frame of an β -strand and can form straight antiparallel duplexes, whose structures resemble that of a ladder (Diederichsen, 1997). A rigid backbone is obtained with very limited nucleobase conformational freedom, thus the geometrical conditionants of DNAs are strongly simplified (Dietrich *et al.*, 2001).

III.1.4 Aims of the project

No experimental three-dimensional structure of an alanyl PNA has been determined to date. The short term objective of the project in collaboration with Prof. Ulf Diederichsen from the Institute of Organic Chemistry of the University of Göttingen was to obtain the first single

crystal structure from this family of model compounds in order to study their conformation in detail.

The structural details should aid a the long term objective of the project: using the stable duplex structure of alanyl-PNAs as a tunable scaffold where synthetic metalloenzyme analogs can be built. Alanyl PNAs offer a rigid structure where some bases on both strands can be modified, for example with histidines in the place of the nucleobases, to build metallic binding centers. Surrounding bases offer an enzyme-like apolar environment while modifications in the backbones would allow for the precise setup and fine tuning of the geometry of the metal binding centre (*Figure III-3*).

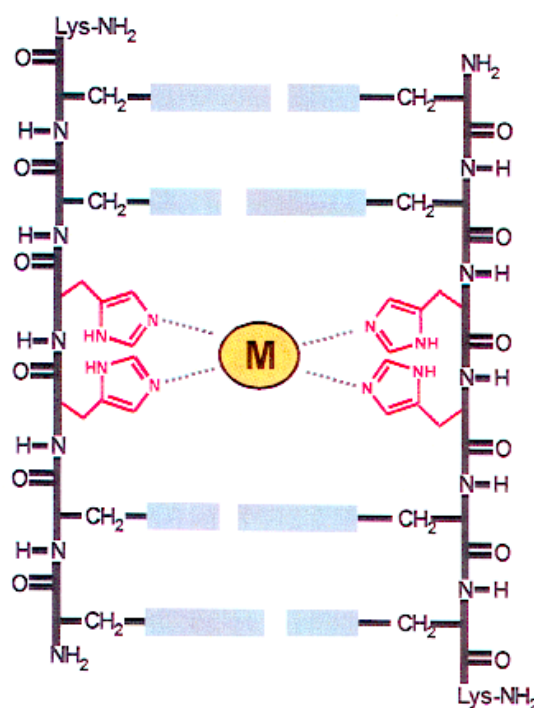


Figure III-3. Schematic example of an alanyl-PNA as scaffold for the design of a metallic binding centre. The geometrical variations are nearly infinite. Specially the distance between both strand can be varied at will. Diagram courtesy of Prof. U. Diederichsen.

III.2 Crystallization attempts on alanyl PNAs

Alanyl GGCGCC PNA

The first crystallization attempt was carried out on an alanyl PNA with sequence GGCGCC where underlined residues are D-alanine based and non underlined residues are based on L-alanine (*Figure III-4*). This molecule will be named PNA1 from now on.

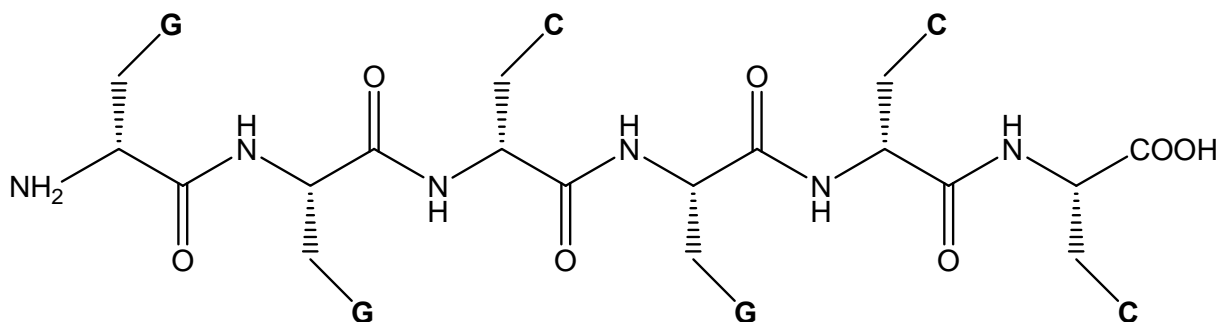


Figure III-4. Chemical structure of PNA1.

PNA1 has a selfcomplementary sequence, it was expected to form an antiparallel duplex with Watson-Crick base pairing.

Unfortunately PNA1 turned out not to be water soluble. An attempt to dissolve it in a phosphate buffer at pH 7 was made and the supernatant, supposedly saturated with PNA1, was used for hanging drop crystallization trials with Hampton Screens I and II (Hampton Research Corp.). No PNA crystals were obtained.

Alanyl GGCGCC-Lys PNA

In an attempt to improve the solubility, a terminal L-lysine was added to the formula of PNA1 (Figure III-5). This molecule will be named PNA2 from now on

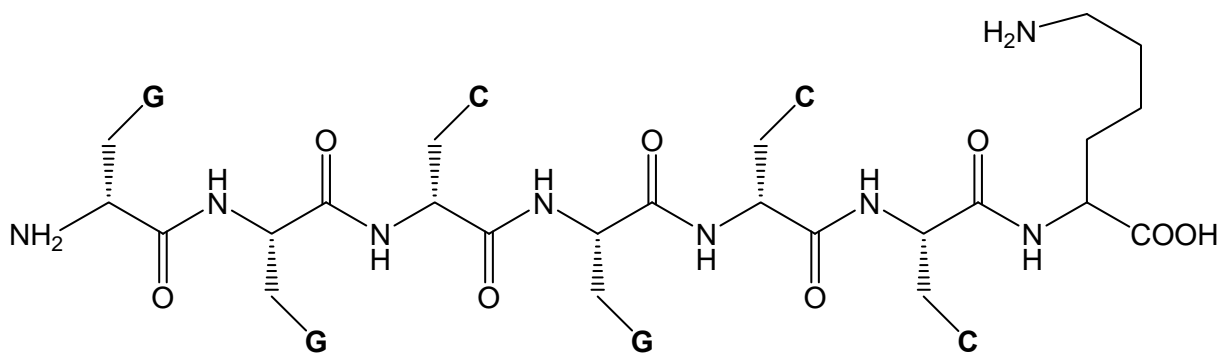


Figure III-5. Chemical structure of PNA2.

The solubility was still poor, estimated to be lower than 0.1 mg/ml. The sample was incubated with 5 ml of phosphate buffer and the supernatant was used for crystallization trials. To improve the total amount of PNA per drop, sitting drops were used with drop ratios up to 40 μ l of PNA stock solution to 5 μ l of reservoir solution. Hampton Screens I and II and the Hampton Matrix Screen were tested at several temperatures.

All crystallization attempts were unsuccessful, probably due to the too poor solubility of the molecule.

Alanyl Lys-GGCGCC-Lys PNA

A second lysine was added at the N-terminus of PNA2 to improve the solubility. The lysine is expected to be protonated at neutral pH, thus the molecule would carry three positive charges. L-lysine was used in the N-terminus, D-lysine in the C-terminus to keep the D-L alternation of the aminoacids. The carboxylic group at the C-terminus was substituted by an amide group (*Figure III-6*). This molecule will be named PNA3 from now on.

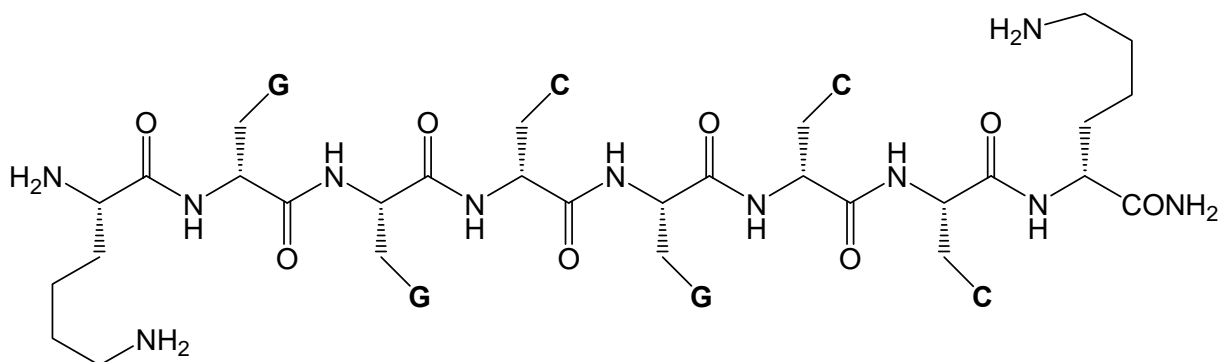


Figure III-6. Chemical structure of PNA3.

The solubility of PNA3 at pH 7 was estimated to be at least 50 $\mu\text{g/ml}$. The improvement relative to PNA2 is too large to be attributed only to the extra lysine, thus the removal of the terminal carboxylic group contribute to the effect. The molecule carries now only positive charges at neutral pH and electrostatic repulsion might prevent aggregation.

Due to the limited amount of sample (2 mg initially) no crystallization experiments were performed at high concentration of PNA. Several hanging drop screens were prepared with PNA3 concentrations ranging from 0.3 mg/ml to 2 mg/ml at 277 K and 293 K. Hampton Screens I and II, Hampton Matrix Screen, Hampton INDEX HT screen, Hampton Salt RX screen and Emerald Cryoscreens I and II (Emerald BioSystems) were tried. A relatively high proportion of drops containing precipitates, specially at 277 K, indicated the adequateness of the concentration range used for screening. The presence of counter ions for the cationic amines of PNA3 in most screening solutions is probably responsible for the drop in solubility. No diffracting crystals of PNA3 could be obtained to date.

III.3 Crystal structure of an alternating D-alanyl, L-homoalanyl PNA

A variant of alanyl PNAs with a different backbone was also used for crystallization trials. Still consisting of alternating D and L aminoacids, the alanyl skeleton was changed to

alternating alanyl and homoalanyl aminoacids, with one extra carbon before the nucleobase in every second residue. The selfcomplementary sequence was kept as GGCGCC (the chirality is inverted relative to PNA1). D-lysine was added at the N-terminus and L-lysine at the C-terminus to achieve solubility in water. The molecule is water soluble to at least 70 mg/ml. The exact chemical formula of this PNA is shown in *Figure III-7*. This molecule will be named PNA4 from now on.

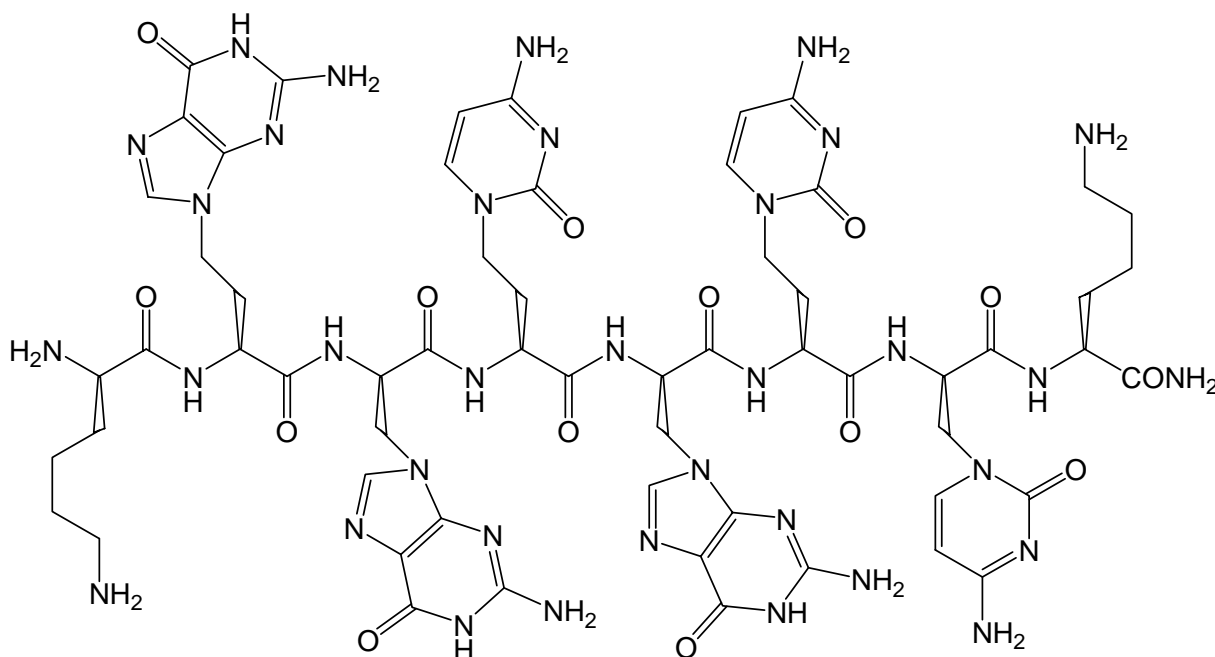


Figure III-7. Chemical structure of PNA4.

PNA4 was expected to form an antiparallel duplex with Watson-Crick base pairing. Through homologation the distance between the strands was expected to increase by approximately 1 Å relative to pure alanyl PNAs.

Three different crystal forms were obtained with PNA4. The structure of the first two in chronological order could not be solved so far. They will be analyzed later, the crystals will be named PNA4a and PNA4b from now on. The third crystal form obtained in chronological order rendered a solution for the structure of PNA4. This crystal form will be named PNA4c from now on and will be analyzed first.

III.3.1 The PNA4c crystal form

The crystal was obtained by the hanging drop method. 4 µl of a PNA4 solution in water with 2 mg/ml concentration was mixed with 4 µl of reservoir solution consisting of 2.3 M

1,6-hexanediol, 0.2 M MgCl₂ and 0.1 M Tris-HCl buffer at pH 8.5. The hanging drop was incubated over 1 ml of the same reservoir solution at 283 K.

A 0.15 mm long and 0.1 mm thick prismatic crystal grew within a cluster of similar crystals (*Figure III-8*). It was separated from the cluster with a stainless steel acupuncture needle, mounted in a nylon loop and flash frozen in liquid nitrogen without addition of extra cryoprotectant.



Figure III-8. Cluster of crystals of the PNA4c crystal form.

It was difficult to manipulate the crystals of this crystal form. The concentration of 1,6-hexanediol was often close to the saturation value and after a couple of minutes of exposure to air at 293 K, crystals of hexanediol would start growing in the drop making mounting of the PNA4 crystals impossible. Much before crystals of 1,6-hexanediol nucleated, the concentration in the surface of the drop had risen sufficiently as to negatively affect the diffraction quality of the crystals, still most crystals grew in clusters and manipulation prior to mounting was necessary. An attempt was made to slow down the concentration process by transferring the drops to a cold room at 277 K and allowing them to equilibrate at that temperature prior to mounting, but the solubility of hexanediol is consequently lower. The nucleation and growth of 1,6-hexanediol crystals would start in the cold room after several seconds of exposure of the drop to the air. Finally, manipulation of the drops in the cold room shortly after their transfer to 277 K turned out to be the solution, the lower temperature of the air would slow the evaporation rate of water while the not precooled drops allowed for the necessary time to manipulate the crystals before nucleation of 1,6-hexanediol began. The concentration of 1,6-hexanediol at the surface of the drops (the droplet of mother liquor finally present in the loop comes from the surface of the drops) was this way low enough as not to negatively affect the diffraction properties of the crystals. Resolution limits amenable for direct methods could be obtained with this method.

Data collection

Data collection was carried out at beamline X10SA of the Swiss Light Source, Switzerland, at 100 K. A dataset obtained at $\lambda=0.7999$ Å was integrated and scaled with XDS to a resolution of 1.0 Å. The crystal belonged to spacegroup $P2_1$ with unit cell dimensions $a = 26.308$ Å, $b = 30.733$ Å, $c = 33.550$ Å, $\beta = 99.53^\circ$. Data collection statistics are listed in *Table III-1*.

Crystal	PNA4c
Space group	$P2_1$
Unit cell	$a=26.308$ Å, $b=30.733$ Å, $c=33.550$ Å, $\beta=99.53^\circ$
Wavelength (Å)	0.7999
Resolution (Å)	1.00 (1.09)
Total reflections	375007
Unique reflections	28573
Completeness (%)	99.5 (98.7)
Redundancy	13.06 (11.58)
R_{int} (%)	6.09 (41.33)
$I/\sigma(I)$	20.99 (5.16)
Data*/restraints/parameters	27136/6252/4966
R (%)	15.03
R_{free} (%)	20.19

Table III-1. Dataset and refinement statistics for crystal PNA4c. In parenthesis the values for the high resolution shell, its lower limit in parenthesis in the resolution line. *The number of data listed is the number of reflections used in the refinement, that is, the working set, the rest to the total number of unique reflections is the test set.

Structure solution and refinement

With 1.0 Å resolution data available, it was possible to attempt the solution of the structure by direct methods with a search for all atoms from the native dataset. This was successful already before the entire dataset was available. A run of the program SHELXD using data to 1.05 Å resolution with commands FIND 438 (the expected number of atoms in two duplexes of the PNA4) and PLOP 460,480,500 to make an expansion of the structure rendered a solution in try number 109 with correlation coefficients after the search phase CC_{all} of 47.84% for strong data (used in the calculation) and CC_{weak} of 17.61% for the data not used

in the calculation. Those values were significantly higher than those for the other tries. After the PLOP phase the correlation coefficient improved to 61.39% for a solution containing only 381 atoms (*Figure III-9*), too many atoms for a solution containing a duplex and too few atoms for two duplexes. Later it was found out that although the crystal has four copies of PNA4 in the asymmetric unit, most lysines are disordered. Many nucleobases could be readily identified at this stage but the model was far from complete. It was not attempted to refine the parameters of the direct methods search.

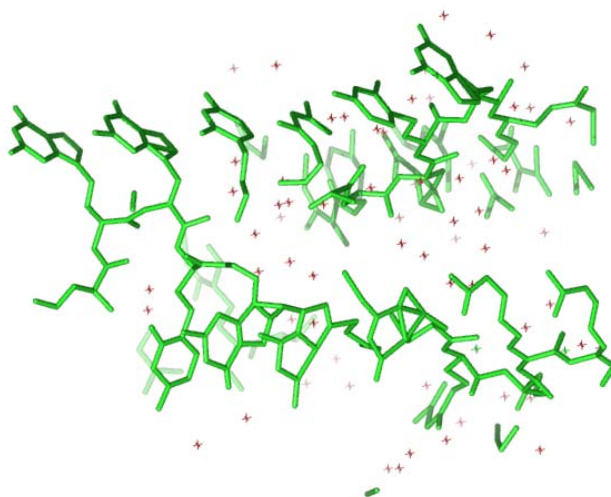


Figure III-9. Solution from SHELXD. The structure is correct but incomplete. The model is an equal atom model, the pink crosses represent isolated atoms in the solution.

This starting set of atoms was used as input for SHELXE to perform phase calculation and density modification. 30 cycles of density modification with an estimated solvent content of 40% were done. Inversion of the SHELXD solution was also performed at this stage since the solution obtained by direct methods corresponded to the enantiomer of PNA4. SHELXE produced a map with estimated correlation coefficients in the order of 97%, a contrast of 0.511, a connectivity of 0.870 and a pseudo-free correlation coefficient of 88.25%. Once the final model was available the weighted mean phase error of this experimental density map could be calculated with SHELXPRO (*Figure III-10*), which was only 18.75°. The still model bias free electron density map was of excellent quality and showed nearly every atom present in the final model (*Figure III-11*).

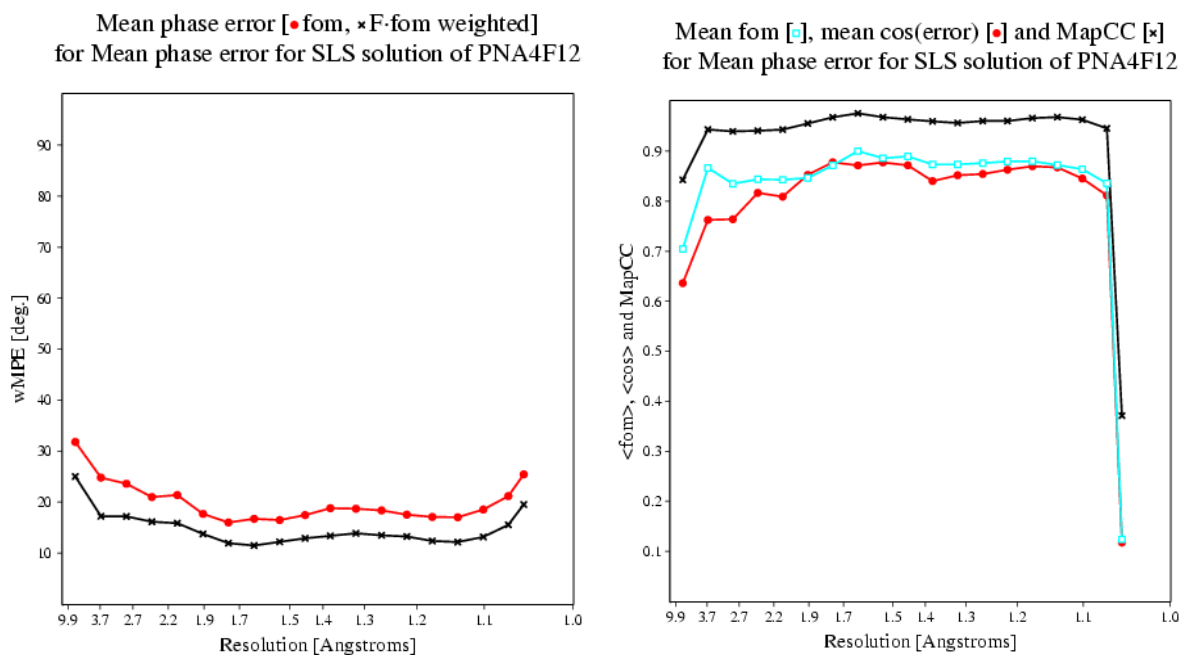


Figure III-10. Statistics of the experimental map for PNA4c compared to the final refined model. Left: mean phase error, f.o.m. weighted (red) and F·f.o.m. weighted (black). Right: mean f.o.m. (blue), mean cosine of the phases errors (red) and map correlation coefficient (black).

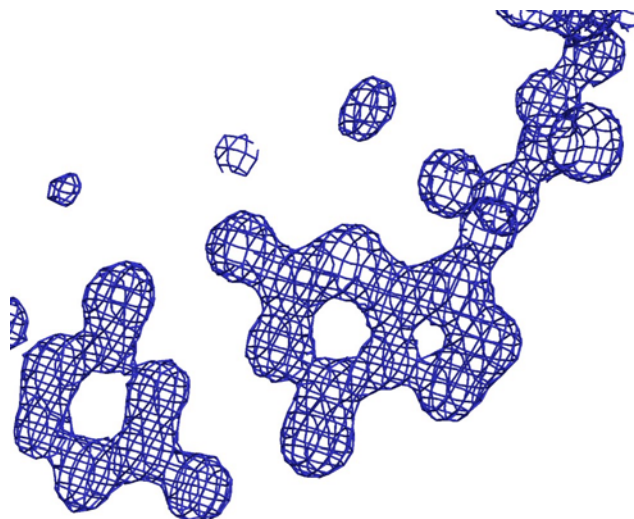


Figure III-11. Part of the experimental electron density map of PNA4c. The map is contoured at the 2σ level. Even some water molecules are visible in the map.

A starting model was built manually in the electron density with the program XFIT.

Restrained full anisotropic refinement was performed with SHELXL to a resolution of 1.00 Å. Standard protein (Engh & Huber, 1991) and DNA (Parkinson, 1996) 1-2 and 1-3 distance restraints were used for the peptidic backbone and the nucleobases respectively. No

reference value was provided to the refinement for the connections between the protein like and the nucleobase parts. Instead, same distance and same angle (1-3 distances in SHELXL) restraints were generated on a residue-type basis. The values were restrained to be the same in all residues of the same kind, for example the eight L-homoalanyl-cytosines in the asymmetric unit, but the overall values were free to vary. The asymmetric unit contains four copies of the PNA4 molecule, with some parts of some lysines disordered and not included in the final model. The final model includes a total of 399 atoms excluding hydrogens and second conformations, plus 140 water molecules. Only 10.2% of the asymmetric unit is empty of atoms in the final model according to PROCHECK (Laskowski *et al.*, 1993).

Description of the structure of PNA4c

The expected structure of PNA4 was an extended antiparallel duplex, mostly linear in nature. It was expected that individual strands would adopt a β -strand like backbone resulting in little or no helicity. All basepairs should be standard Watson-Crick. However, the structure found in the PNA4c crystal deviates significantly from the expected model.

In PNA4c each individual chain has a 90° bend in the backbone every second residue, so that only four nucleobases are oriented in the same direction. Hydrogen bonds within the backbone stabilize this orientation. The four nucleobases are the first, second, fifth and sixth of each chain, with the fifth base stacking on the second. The two central nucleobases project away from the backbone in the opposite direction and their Watson-Crick edges point approximately 90° apart relative to the Watson-Crick edges of the other four (*Figure III-12*). Most lysines are disordered, parts of five of them were not observed in the electron density and were not included in the model. None of the monomers has both lysines ordered.

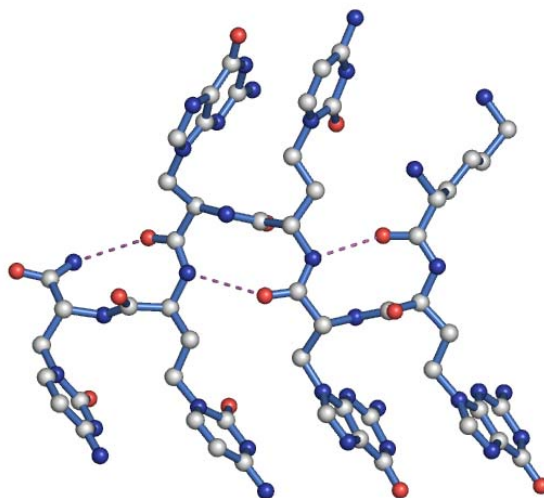


Figure III-12. One monomer of PNA4c. All four adopt similar conformations. The chain runs from right to left. The hydrogen bonds within the monomer backbone are drawn in magenta with dashed lines.

The asymmetric unit contains four monomers and all guanines are Watson-Crick paired to cytosines and vice versa. The four stacked bases of each monomer form an antiparallel duplex with the identical four of another monomer while the two central basepairs bind to the equivalents of yet a third monomer in an antiparallel way. The base pairing of all four monomers defines a tetramer with the approximate shape of a square cage with open top and bottom. Each monomer makes a side of the cage (*Figure III-13*). The numbering scheme used and the base pairing in the tetramer are shown schematically in *Figure III-14*.

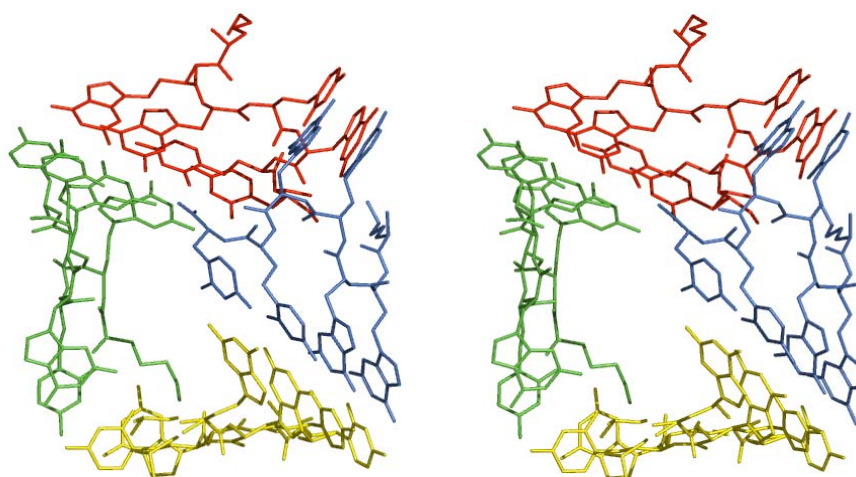


Figure III-13. Stereo picture of the tetramer structure of PNA4c. Each monomer is colored differently. The solvent has been omitted for clarity.

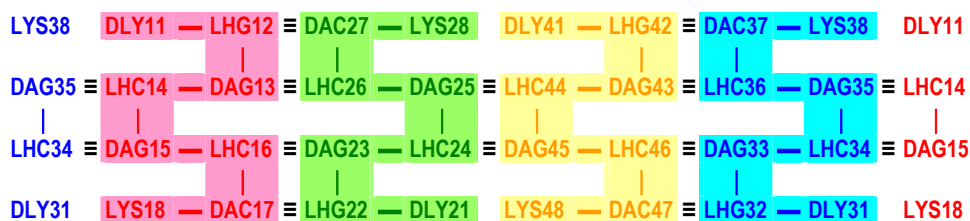


Figure III-14. Base pairing and residue numbering in the tetramer of PNA4c. The monomers are colored as in *Figure III-13*. The left and right sides of the figure are the same residues, the tetramer forms a closed ring. The base pairing is indicated by triple lines. D stands for D aminoacids, L for L aminoacids, A for alanyl, H for homoalanyl. DLY and LYS are the terminal lysines. Parts of many lysines are not present in the final model.

Structure analysis

The structure is specially well defined in the central part of each monomer. All atoms in the residues containing the nucleobases are clearly identified in the density and have B values mostly under 25 \AA^2 . There are no disorders modeled for this part of the structure. The six central residues of the monomers will be analyzed first.

The backbone of all monomers has a similar conformation, with RMSD for the six central residues relative to the first monomer of 0.51 \AA for the second monomer, 0.72 \AA for the third and 0.59 \AA for the fourth monomer in the asymmetric unit. The conformation of the monomers is characterized by a bend in the direction of the chain of approximately 90° at each α -carbon, hence the main chain follows a zigzag path. In terms of torsion angles this corresponds to Φ angles around -60° for the L-aminoacids and around 100° for the D-aminoacids and Ψ angles cluster around 130° for the L-aminoacids and around -15° for the D-aminoacids. The two central residues of each monomer have a different, less constrained environment, but their backbone torsion angles deviate very little from the values of the other residues (*Table III-2*). The Ψ angle for Dag27 is out of the trend of the other monomers. All terminal lysines were disordered and only some atoms from them were included in the final model. Accordingly the torsion angles in which the lysines are involved are less reliable than the rest.

Residue	Angle (°)	Lhg2	Dag3	Lhc4	Dag5	Lhc6	Dag7
+10	Φ	-57.61	105.83	-53.92	90.68	-67.29	91.59
+20	Φ	-56.06	102.95	-63.52	93.20	-72.51	109.64
+30	Φ	-64.79	100.20	-64.71	86.34	-55.52	111.07
+40	Φ	-55.15	104.42	-56.15	93.23	-60.57	101.90
+10	Ψ	128.67	-15.71	127.69	-9.16	137.80	-18.34
+20	Ψ	118.04	-27.16	125.36	-0.65	134.60	-154.78
+30	Ψ	123.38	-26.34	131.96	2.80	127.87	-0.45
+40	Ψ	132.64	-23.18	130.56	-13.22	132.37	-23.86

Table III-2. Torsion angles for the backbones of PNA4c. The real residue numbers are obtained by adding the numbers in the first column in the residue names.

The zigzagging backbone allows the formation of intramolecular hydrogen bonds within the main chain, these hydrogen bonds have a regular structure. The backbone oxygen of the D-aminoacids forms a hydrogen bond with the nitrogen of the third next (always L) aminoacid where available (*Table III-3*) (*Figure III-12*). As for the torsion angles, Lys28 N is out of the overall trend and does not form the hydrogen bond to Dag25 O. This is the only exception found to the regular hydrogen bonding, although the O-N distances vary between monomers and residues. Last, the side chain of Dly11 bends towards the main chain so that NZ forms a long hydrogen bond at 3.24 Å to Lhg12 O. The other two ordered lysines in the model do not form similar bonds.

Donor	Lhc14 N	Lhc24 N	Lhc34 N	Lhc44 N
Acceptor	Dly11 O	Dly21 O	Dly31 O	Dly41 O
Distance (Å)	3.05	2.89	2.96	2.90
Donor	Lhc16 N	Lhc26 N	Lhc36 N	Lhc46 N
Acceptor	Dag13 O	Dag23 O	Dag33 O	Dag43 O
Distance (Å)	2.97	3.06	3.28	3.04
Donor	Lys18 N	Lys28 N	Lys38 N	Lys48 N
Acceptor	Dag15 O	Dag25 O	Dag35 O	Dag45 O
Distance (Å)	3.14	5.40	3.14	3.18

Table III-3. Backbone hydrogen bonding in PNA4c.

The base pairing in the structure is surprisingly regular since no two different basepairs from a monomer have the same environment (meaning the same surrounding bases). All 36 hydrogen bonds lie within the range 2.75 Å - 2.94 Å (*Table III-4*) and the Watson-Crick edges of the bases are thus nearly parallel to each other in the basepairs.

Distance(Å) \ Basepair	N2-O2	N1-N3	O6-N4
Lhg12-Dac27	2.80	2.85	2.83
Dag13-Lhc26	2.93	2.92	2.81
Dag23-Lhc16	2.85	2.90	2.81
Lhg22-Dac27	2.80	2.81	2.84
Dag35-Lhc14	2.84	2.88	2.87
Dag15-Lhc34	2.85	2.93	2.87
Dag45-Lhc24	2.86	2.92	2.87
Dag25-Lhc44	2.82	2.92	2.94
Lhg42-Dac37	2.75	2.89	2.92
Dag43-Lhc36	2.79	2.93	2.89
Dag33-Lhc46	2.88	2.92	2.86
Lhg32-Dac47	2.80	2.89	2.84

Table III-4. Hydrogen bonding distances for the base pairing in PNA4c. Guanines are always listed first. All basepairs are standard Watson-Crick.

Since the base pairing in PNA4c is always Watson-Crick, standard nucleic acid geometrical parameters can be used for the description of basepair geometry. Only the nucleobases are considered in the definitions.

All of the internal basepair geometrical parameters have values close to ideal (zero for all parameters), indicating the absence of stress in this structure. The mean values for all 12 basepairs calculated with 3DNA (with standard deviations in parenthesis) are: Shear: -0.01 Å (0.23 Å), Stretch: -0.15 Å (0.03 Å), Stagger: -0.05 Å (0.14 Å), Buckle: 0.99° (6.66°), Propeller: -0.31° (6.36°), Opening: -0.93° (0.94°).

The stretch angle deviates systematically from zero. This effect is also present in high resolution nucleic acid crystal structures (Olson *et al.*, 2001). The ideal values defined in the program 3DNA (Lu & Olson, 2003) result in longer hydrogen bonds than observed in real

nucleic acid structures, as is the case here. The stretch values found here are similar to those of A-DNA and B-DNA.

The local base-pair step parameters are listed in *Table III-5*. They characterize the relative positions of possible chemical groups inserted in the position of the bases that would form a metallic binding site. There are four basepair stacked segments in the structure.

Basepair step	Shift(Å)	Slide(Å)	Rise(Å)	Tilt(°)	Roll(°)	Twist(°)
12-27/13-26	-0.68	3.56	3.55	2.48	4.62	14.05
13-26/16-23	-0.04	2.09	2.83	-1.19	2.60	9.78
16-23/17-22	0.33	3.82	3.54	-0.36	1.25	10.72
14-35/15-34	0.48	3.10	3.39	-1.52	4.77	17.20
24-45/25-44	-0.27	3.22	3.49	1.47	3.44	16.60
32-47/33-46	-0.15	3.87	3.56	2.45	1.50	11.81
33-46/36-43	-0.07	2.69	2.92	2.04	4.28	11.25
36-43/37-42	-0.02	4.27	3.76	-1.01	1.82	20.23

Table III-5. Basepair step parameters for PNA4c.

The values of slide are specially significant since it is consistently high. This indicates that the bases are not directly stacked on each other, the degree of stacking is indeed surprisingly low in this structure. If the bases are considered as polygons between punctual atoms they tend not to lie in the projection of their neighbors, still, there is significant stacking if their Van der Waals volumes and whole basepairs are taken into account. The high slide values are only possible if the bases are not perpendicular to the backbones (the line between α -carbons) so that the 3.7 Å - 3.8 Å distances between α -carbons can translate into a lower rise per basepair step. The mean rise is 3.55 Å for the six steps where the bases are adjacent in the sequence and 3.38 Å overall. The alanyl-homoalanyl alternance also contributes to the high slide values. Interestingly, the basepair steps where the residues are not contiguous in the sequence integrate well in the overall basepair step parameters although with smaller slide values. The smaller rise parameters for these basepair steps are mostly an artifact resulting from buckle angles of opposite signs in the involved basepairs. Visual inspection confirms that the rise per basepair is more constant than the values in the table suggest, and close to the mean value over the whole structure.

Tilt and roll have relatively small values, but twist does not. The average twist angle is 13.95°, always with right handed basepair steps. As a result the structure has a significant

degree of helicity. The imaginary resulting helix would have 25.8 bases per helix turn and a helix pitch of 87.23 Å.

The regularity of the backbone and the base parameters result in the tetramer structure being composed of six similar units. Every couple of residues that are contiguous in the sequence and their basepair partners define a *sandwich* of basepairs in which the β -carbons are projected away from each other (*Figure III-15*). This structural unit is very similar in all cases, and accommodates GC-GC, CG-GC and CG-CG steps equally well. The torsion angles that would correspond to the glycosidic torsion angle χ of the nucleotides divide this structural unit in two families. The central two residues of each monomer have *gauche*⁺ torsion angles in this position (all eight in two *sandwich* units) with the carbonyls projecting towards the helical minor groove while all others are *gauche*⁻ with the carbonyls projecting into the major groove.

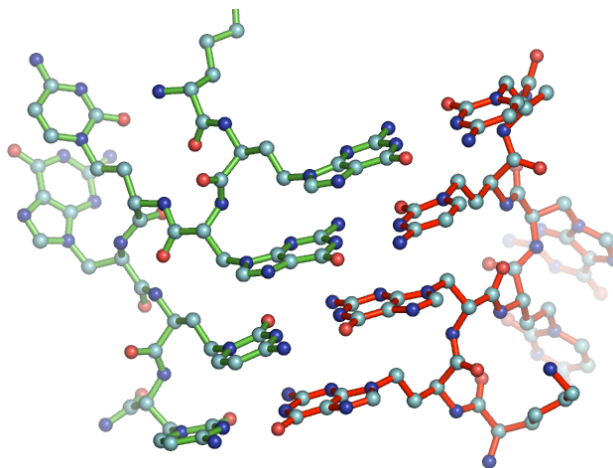


Figure III-15. Repetitive *sandwich* units in PNA4c. The lower one is formed by Lhg12-Dac27 and Dag13-Lhc26 (the lowest two basepairs), the upper one by Lhc16-Dag23 and Dac17-Lhg22 (immediately above the first). The bases projecting backwards in the figure are involved in similar units (except for the torsion angle linking to the bases) but their binding partners have been omitted for clarity.

The distances between α -carbons are important parameters for the design of enzyme analogs. For alanyl-like PNAs they can be varied by adjusting the lengths of the linkers between the backbone and the nucleobases. The values for these distances are listed in *Table III-6*.

α -carbons	12-27	13-26	14-35	15-34	16-23	17-22
Distance (Å)	12.72	12.88	11.61	11.83	12.86	12.46
α -carbons	32-47	33-46	24-45	25-44	36-43	37-42
Distance (Å)	12.63	12.74	11.85	11.81	12.60	12.24

Table III-6. Distances between α -carbons within basepairs.

The values are more flexible than expected. Specially the distances for the two central basepairs of each monomer, which are in a different environment as the others, are nearly 1 Å shorter. Since the guanine N9 – cytosine N1 distances are fairly constant, all 8.9 Å or 9.0 Å, this reflects flexibility in the side chains, probably provided by the linkers. The two groups of α -carbon - α -carbon distances are the same as the two families of pseudo-glycosidic torsion angles.

Only three of the eight lysines are defined in the electron density. All three are N-terminus lysines, the other five have been partially omitted from the model, only atoms defined in the difference electron density maps were included. It is satisfying to find that the lysines are mostly disordered with no definite structure since they were included in the PNA only to improve the solubility. It is to be expected that the lysines have little or no effect on the conformation of the central six residues of each monomer.

The N-terminal amine groups are expected to be protonated at pH 8.5. Actually the amine nitrogen of Dly21 forms a hydrogen bond at 2.89 Å distance to Dac17* N4 of a symmetry equivalent and Dly31 N forms a hydrogen bond at 2.96 Å to Dac47* N4 of another symmetry equivalent. In both cases the lysine nitrogen must act as a hydrogen bond acceptor and the lysines thus cannot be protonated. This caused problems during the refinement where initially they were modeled with three riding hydrogens each. In the case of Dly31, the removal of those hydrogens alone (the position of two hydrogens is not geometrically determined and thus not amenable to the riding model for hydrogens) was followed by an improvement of 0.5% in both the R factor and R_{free} ; there was no space for the extra hydrogen and other atoms were displaced as a result. This eliminated as well strong difference density from the F_o-F_c map. Whether Dly11 and Dly41 are protonated can not be determined, Dly11 was modeled as protonated while the hydrogens of Dly41 are omitted from the model due to disorder of the surrounding atoms.

Crystal packing

The tetramers are stacked on each other so that infinite, wide solvent channels are formed along the crystallographic *a* axis (*Figure III-16*). The tetramers do not lie directly on top of each other as the alternance of red and yellow in *Figure III-16* indicates, but diagonally to each other in different layers. If the tetramers are considered as filled cubes, the crystal packing would resemble a three-dimensional checkers board, where only every second cube is occupied and the cubes (the tetramers) meet only in the vertices. The nucleobases are in the corners of each tetramer, this arrangement allowing for intensive inter-tetramer base stacking. This results in some, but not all, infinite columns of basepairs; some nucleobases are stacked on other groups or face disordered solvent on the outside of the tetramer. Some hydrogen bond lateral contacts between nucleobases and some lysine hydrogen bonds also help to form the crystal lattice. The exact situation is probably too complex to be detailed here.

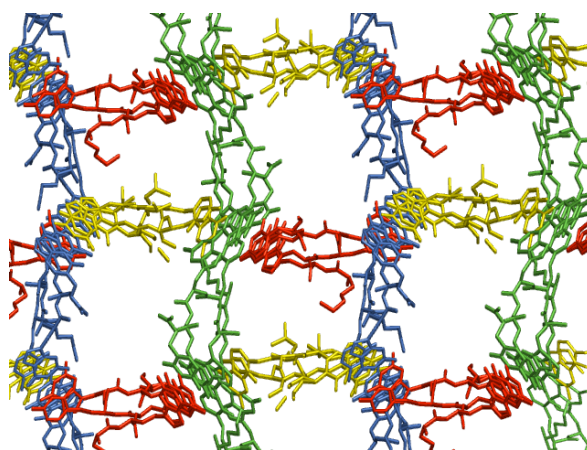


Figure III-16. Crystal packing in PNA4c. The crystallographic *a* axis runs perpendicular of the paper. The solvent has been omitted for clarity.

III.3.2 Other crystal forms

A different crystal form, PNA4a, was also obtained. Typical crystallization conditions were hanging drops formed by a PNA4 solution with 2 mg/ml concentration and a reservoir solution consisting of 3.2 M hexanediol, 0.2 M MgCl₂ and 0.1 M Tris-HCl buffer at pH 8.5. The temperature yielding the best crystals was 283 K but crystals were also obtained at 277 K and 293 K. PNA4a crystals grew as needles, typically in clusters (*Figure III-17*).

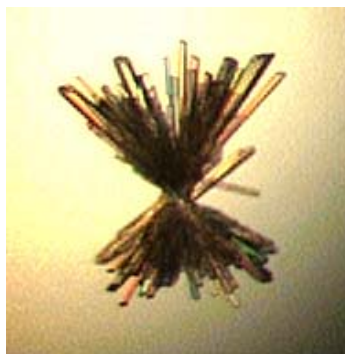


Figure III-17. Cluster of needles of PNA4a crystals.

The crystals of this form have an apparent spacegroup $P4_22_12$ with cell $a = b = 39.65 \text{ \AA}$, $c = 53.91 \text{ \AA}$ but the systematic absences are not clear. A local program (George M. Sheldrick) was used to divide the dataset in parity groups, depending on whether the h , k and l indexes of the reflections are odd or even. The mean $I/\sigma(I)$ value was calculated for each group (*Table III-7*).

Parity group (hkl)	eee	oee	oeo	ooo	eeo	oeo	ooo	ooo
$I/\sigma(I)$	8.941	5.419	4.518	0.134	0.206	4.593	5.585	9.226
Domains	A+B	A	B	-	-	B	A	A+B

Table III-7. Mean $I/\sigma(I)$ values for the parity groups of reflections of PNA4a indexed in $P4_22_12$. Even reflection indexes are represented as e , odd indexes as o . Two domains with different systematic absences explain the intensity distribution (see text). The domain letters illustrate the domains for which reflections are present.

The two axes of equal length and the different intensities for the parity groups suggested the possibility of twinning with a non primitive lattice centering and exchange of the axes as the twin law. Spacegroup $C222_1$ with a cell $a = 39.65 \text{ \AA}$, $b = 53.91 \text{ \AA}$, $c = 39.65 \text{ \AA}$ for domain A and exchange of the a and c axis for domain B (thus producing an a centred cell and the non standard spacegroup $A2_122$) would explain the intensity distribution through the different parity groups. Both twin components would then have different conditions for the absence of reflections as shown in *Table III-7*.

Datasets up to 1.5 \AA resolution are available and work is currently in progress towards a solution of the structure of PNA4a based on that twinning scheme.

Crystals of the PNA4b form were frequently obtained in conditions similar to those for PNA4a. In some drops all 3 crystal forms, PNA4a, PNA4b and PNA4c grew simultaneously. Crystals of PNA4b had a rhombic crystal habit (*Figure III-18*).

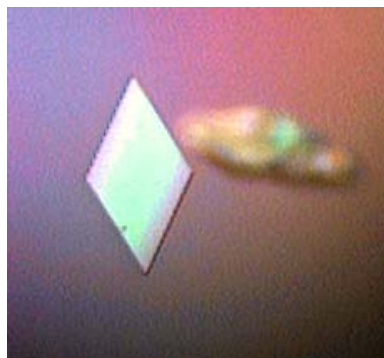


Figure III-18. A crystal of the PNA4b form.

The crystals have an apparent spacegroup $C222_1$ with unit cell dimensions $a = 57.03 \text{ \AA}$, $b = 108.42 \text{ \AA}$, $c = 82.32 \text{ \AA}$. Again the systematic absences were difficult to assign. The unit cell obeys the equation

$$c \cos\beta \approx -a/2$$

a well known twin generating relationship (Rudolph *et al.*, 2004). The crystals are likely twinned belonging to spacegroup $P2_1$ with unit cell $a = 57 \text{ \AA}$, $b = 82 \text{ \AA}$, $c = 61 \text{ \AA}$, $\beta = 117^\circ$. Datasets are available to a resolution of 2.1 \AA and work is currently in progress towards a solution of the structure.

III.4 Discussion and future perspectives

The crystal structure of PNA4c is a tetramer and not a dimer as expected. This could be a crystallization artifact given the high concentrations employed and indeed the melting curve of PNA4 (at much lower concentrations) has two inflexion points, likely corresponding to a dimer and a tetramer transition (*Figure III-19*).

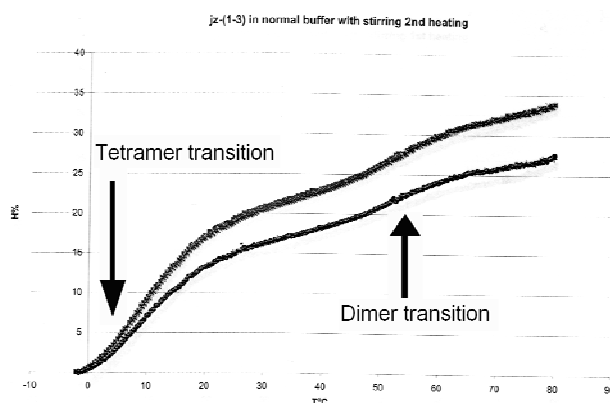


Figure III-19. Melting curve of PNA4 in a phosphate buffer at pH 7 (Courtesy of Jinhua Zhang). Both a heating and a cooling phase are shown.

At the present stage it can only be speculated whether the dimer structure would correspond to half of the tetramer in PNA4c, with only four basepairs, or if it would be different, possibly performing all six basepairs. The dimer to tetramer transition although, allows for the formation of extra hydrogen bonds within the backbones without reducing the number of hydrogen bonds in the basepairing. An energetic penalty would arise from the reduced base stacking in the tetramer, this is compensated by extensive intertetramer stacking in the crystal.

The utility of the structure as a scaffold would be the same as a dimer or as a tetramer, but a minimum number of basepairs must be present to provide a stable structure after the modified bases have been implemented in the molecule. With the structure of PNA4c and assuming that two bases in each chain have to be modified, the result would be only two basepairs surrounding the synthetic enzymatic site, likely more would be necessary to provide a stable scaffold.

The regular backbone of the monomers (and the dimer) suggests the possibility that a longer monomer would continue the base stacking in the same way. To test this theory the dimer formed by residues 12-17 and 22-27 was selected. Two copies were made and residues 12, 13, 26 and 27 of the second copy were superimposed on residues 16, 17, 23 and 22 of the first copy respectively. Since the superimposed nucleobases are different, only the backbone atoms and the β -carbons, 20 atoms in total, were used for the superposition. The resulting RMSD is 0.55 Å, only atoms directly linking to lysines deviate substantially from their equivalents (*Figure III-20*).

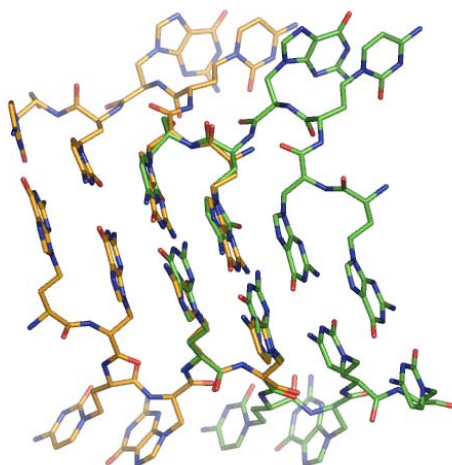


Figure III-20. Superposition of a dimer from PNA4c on itself. The first copy is represented with green carbons, the second with orange carbons.

A continuous hydrogen bond pattern arises in the backbone for the residues whose bases are not paired in the duplex. This pattern closely resembles the one present in the tetramer, which also gets extended. The nucleobases are of different nature in both superimposed copies, but they are oriented in the same way and any possible combination of guanine-cytosine basepairs would be integrated in the model without stress and without disrupting the Watson-Crick base pairing.

A hexamer of stacked basepairs arises in this way which closely mimics the regular structure of four basepairs observed in PNA4c. Even the right handed helicity is preserved and it seems likely that a regular duplex of any desired length could be constructed in this way. This would provide the necessary number of basepairs to obtain a stable scaffold independently of the functionalities that are built into the molecules. The bases not involved in duplex formation as shown in *Figure III-20* could likely be removed from the molecule or other chemical entities may be placed in their positions.

The solution of the structure of the other two crystal forms, PNA4a and PNA4b may in the future resolve the issue of whether the structure observed in PNA4c is the only one available to this molecules or if a family of conformations is possible, possibly with a fully extended duplex between them.

IV SUMMARY

IV.1 Structures of echinomycin in complex with double stranded DNAs

The structure of complexes between the antibiotic echinomycin and different short DNA duplexes have been studied by single crystal X-ray diffraction methods.

Echinomycin is a DNA bisintercalator with applications as an anticancer agent. The structures of its complexes with duplex DNA with sequences (GCGTACGC)₂, (ACGTACGT)₂ and (CGTACG)₂ have been solved by molecular replacement with the structure of a Triostin A complex as the starting model. All three complexes contain two echinomycin molecules bisintercalating around both CG steps of the DNA. For the complex with CGTACG an ambiguity in the assignment of the space group was discovered.

With all three DNAs it has been found that the structure of the complexes closely resembles that of the original Triostin A complexes, specially the DNA basepairs immediately external to the bisintercalation site are in the Hoogsteen configuration, with the purines rotated into the *syn* conformation. Regular hydrogen bonding from the alanines of Echinomycin to both guanines within the bisintercalation sites has been observed in contrast to the crystal structures of the complexes with Triostin A. The possibility of Echinomycin binding to the DNA in two different orientations has been observed for the first time.

Another crystal form has been obtained for the complex of Echinomycin with (ACGTACGT)₂ and its structure has been solved by means of single wavelength anomalous diffraction. In this crystal form some of the A-T basepairs external to the bisintercalation site are paired in the standard Watson-Crick mode, providing the first experimental three-dimensional structure of that kind of interaction. Other basepairs external to the bisintercalation site adopt the Hoogsteen basepairing mode, illustrating the conformational flexibility of the interaction.

IV.2 Crystal structure of an alternating D-alanyl, L-homoalanyl PNA

PNAs are DNA analogues with value in therapy and as a DNA model. It is tried to use them as an scaffold for function oriented chemical synthesis of enzyme analogs.

The structures of several alanyl PNAs have been studied by crystallographic methods.

Diffracting crystals of pure alanyl PNAs could not be obtained but crystallization attempts were successful for an alternating D-alanyl, L-homoalanyl PNA with six built-in nucleobases of sequence GGCGCC, enabling this PNA to form antiparallel duplexes with itself.

The structure of the crystal was solved by direct methods. It was found that the monomers fold in a zigzagging path with intramolecular hydrogen bonds so that four nucleobases are oriented to one side of the backbone and two to the other side. Four of these monomers associate to form a tetrameric structure held together by Watson-Crick base pairing. Although the structure is different to the initially expected one, it looks likely that no major modifications are needed to obtain the desired functionality.

REFERENCES

- Adress, K. J. & Feigon, J.** (1994). NMR investigation of Hoogsteen base pairing in quinoxaline antibiotic – DNA complexes: comparison of 2:1 Equinomycin, triostin A and [N-MeCys³,N-MeCys⁷] TANDEM complexes with DNA oligonucleotides. *Nucleic Acids Res.*, **22**(24), 5485-5491.
- Aishima, J., Rossitza, K. G., Noah, J. E., Gan, H. H., Schilick, T. & Wolberger, C.** (2002). A Hoogsteen base pair embedded in undistorted B-DNA. *Nucleic Acids Res.*, **30**(23), 5244-5252.
- Allen, F. H.** (2002). The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Cryst.*, **B58**, 380-388.
- Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A. & Demeny, T.** (1992). The Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids. *Biophys. J.*, **63**, 751-759.
- Bloomfield, V. A., Crothers, D. M. & Tinoco, I. Jr.** (2000). Nucleic acids, structures, properties and functions. University science books.
- Böhler, C., Nielsen, P. E. & Orgel, L. E.** (1995). Template switching between PNA and RNA oligonucleotides. *Nature*, **376**, 578-581.
- Brünger A. T.** (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.
- Bruker Nonius** (2002). XPREP, SADABS computer programs.
- Chan, H. & Patel, D. J.** (1995). Solution structure of a Quinomycin Bisintercalator-DNA Complex. *J. Mol. Biol.*, **246**, 164-179.
- Collaborative Computational Project, Number 4.** (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst.*, **D50**, 760-763.
- Crowther, R. A. & Blow, D. M.** (1967). A method of positioning a known molecule in an unknown crystal structure. *Acta Cryst.*, **23**, 544-548.
- DeLano, W. L.** (2003). The PyMOL molecular graphics system. *DeLano Scientific LLC*, San Carlos, CA, USA.
- 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.
- Diederichsen, U.** (1997). Alanyl PNA: Evidence for Linear Band Structures Based on Guanine-Cytosine Base Pairs. *Angew. Chem. Int. Ed. Engl.*, **36**(17), 1886-1889.
- Dietrich, B., Hupp, T. A. & Engels, B.** (2001). Binding Properties in Protein Nucleic Acids. *Proceedings of the 2001 NIC Symposium*.
- Engh, R. A. & Huber, R.** (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.*, **A47**, 392-400.
- Gallego, J., Ortiz, R. & Gago, F.** (1993). A Molecular Dynamics Study of the Bis-Intercalation Complexes of Echinomycin with d(ACGT)₂ and d(TCGA)₂: Rationale for Sequence-Specific Hoogsteen Base Pairing. *J. Med. Chem.*, **36**, 1548-1561.

- Gallego, J., Luque, F. J., Orozco, M. & Gago, F.** (1994). Binding of Echinomycin to d(GCGC)₂ and d(CCGG)₂: Distinct Stacking Interactions Dictate the Sequence-Dependent Formation of Hoogsteen Base Pairs. *J. Biomol. Struct. Dyn.*, **12**(1), 111-129.
- Gallego, J., Luque, F. J., Orozco, M., Burgos, C., Alvarez-Builla, J., Rodrigo, M. M. & Gago, F.** (1994). DNA Sequence-Specific Reading by Echinomycin: Role of Hydrogen bonding and Stacking Interactions. *J. Med. Chem.*, **37**, 1602-1609.
- Gao, X. & Patel, D. J.** (1988). NMR Studies of Echinomycin Bisintercalation Complexes with d(A1-C2-G3-T4) and d(T1-C2-G3-A4) Duplexes in Aqueous Solution: Sequence-Dependent Formation of Hoogsteen A1•T4 and Watson-Crick T1•A4 Base Pairs Flanking the Bisintercalation Site. *Biochemistry*, **27**, 1744-1751.
- Gilbert, D. E. & Feigon, J.** (1991). The DNA Sequence at Echinomycin Binding Sites Determines the Structural Changes Induced by Drug Binding: NMR Studies of Echinomycin Binding to [d(ACGTACGT)]₂ and [d(TCGATCGA)]₂. *Biochemistry*, **20**, 2483-2494.
- Gilbert, D. E. & Feigon, J.** (1992). Proton NMR study of the [d(ACGTATACGT)]₂-echinomycin complex: conformational changes between echinomycin binding sites. *Nucleic Acids Res.*, **20**(10), 2411-2420.
- Hartman, K. A. Jr. & Rich, A.** (1965). The tautomeric form of helical polyribocytidylic acid. *J. Am. Chem. Soc.*, **87**, 2033-2039.
- Hoogsteen, K.** (1959). The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Cryst.*, **12**, 822-823.
- Hossain, M. B., van der Helm, D., Olsen, R. K., Jones, P. G., Sheldrick, G. M., Egert, E., Kennard, O., Waring, M. J., & Viswamitra, M. A.** (1982). Crystal and Molecular Structure of the Quinoxaline Antibiotic Analogue TANDEM (Des-N-tetramethyltrioestin A). *J. Am. Chem. Soc.*, **104**, 3401-3408.
- Jeppesen, C. & Nielsen, P. E.** (1988). Detection of intercalation-induced changes in DNA structure by reaction with diethyl pyrocarbonate of potassium permanganate, evidence against the induction of Hoogsteen base pairing by echinomycin. *FEBS Letters*, **231**(1), 172-176.
- 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. & Hauptmann, M.** (1956). A theory of phase determination for the four types of non centrosymmetric space groups 1P222, 2P22, 3P₁2, 3P₂2. *Acta Cryst.*, **9**, 635-651.
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C.** (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, **181**, 662-666.
- Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B.** (1999). Rapid automated molecular replacement by evolutionary search. *Acta Cryst.* **D55**, 484-491.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M.** (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283-291.
- Lee, J. S. & Waring, M. J.** (1978). Bifunctional intercalation and sequence specificity in the binding of quinomycin and triostin antibiotics to deoxyribonucleic acid. *Biochem. J.*, **173**, 115-128.
- Leslie, K. D. & Fox, K. R.** (2002). Interaction of Hoechst 33258 and echinomycin with nucleosomal DNA fragments containing isolated ligand binding sites. *Biochemistry*, **41**, 3484-3497.
- 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.
- Low, C. M. L., Olsen, R. K. & Waring, M. J.** (1984). Sequence preferences in the binding to DNA of triostin A and TANDEM as reported by DNAase I footprinting. *FEBS Lett.*, **176**, 414-420.

- 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**(17), 5108-5121.
- May, L.G., Madine, M. A. & Waring, M. J.** (2004). Echinomycin inhibits chromosomal DNA replication and embryonic development in vertebrates. *Nucleic Acids Res.*, **32**(1), 65-72.
- McLean, M. J., Seela, F. & Waring, M. J.** (1989). Echinomycin-induced hypersensitivity to osmium tetroxide of DNA fragments incapable of forming Hoogsteen base pairs. *Proc. Natl. Acad. Sci. USA*, **86**, 9687-9691.
- McRee, D. E.** (1999). XtalView/Xfit – a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.*, **125**, 156-165.
- Merrit, E. A. & Bacon, D. J.** (1997). Raster3D : photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.
- Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptmann, H. A.** (1993). On the application of the minimal principle to solve unknown structures. *Science*, **259**, 1430-1433.
- Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L. & Aggarwal, A. K.** (2004). Replication by human DNA polymerase- α occurs by Hoogsteen base-pairing. *Nature*, **430**, 377-380.
- Neidle, S.** (1999). Oxford Handbook of Nucleic Acid Structure. Oxford University Press. Edited by Stephen Neidle.
- Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O.** (1991). Sequence selective recognition of DNA by strand displacement with a thymine substituted polyamide. *Science*, **254**, 1497-1500.
- Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. L., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C. S., Westhof, E., Wolberger, C. & Berman, H. M.** (2001). A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, **313**(1), 229-237.
- Otwinowski, Z. & Minor, W.** (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.*, **276**, 307-326.
- Park, J. Y., Park, S. J., Shim, K. Y., Lee, K. J., Kim, Y.-B., Kim, Y. H. & Kim, S. K.** (2004). Echinomycin and a novel analogue induce apoptosis of HT-29 cells via the activation of MAP kinases pathway. *Pharm. Res.*, **50**, 201-207.
- Portugal, J., Fox, K. R., McLean, M. J., Richenberg, J. L. & Waring, M. J.** (1988). Diethyl pyrocarbonate can detect a modified DNA structure by the binding of quinoxaline antibiotics. *Nucleic Acids Res.*, **16**(9), 3655-3670.
- Parkinson, G.** (1996). New Parameters for the Refinement of Nucleic Acid-Containing Structures. *Acta Cryst.*, **D52**, 57-64.
- Quigley, G. J., Ughetto, G., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J. & Rich, A.** (1986). Non-Watson-Crick G·C and A·T Base Pairs in a DNA-Antibiotic Complex. *Science*, **232**, 1255-1258.
- Rasmussen, H. & Sandholm, J.** (1997). Crystal structure of a peptide nucleic acid (PNA) duplex at 1.7Å resolution. *Nature Struct. Biol.*, **4**(2), 98-101.
- Ray, A. & Nordén, B.** (2000). Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future. *FASEB J.*, **14**, 1401-1060.
- Richmond, T. J. & Davey, C. A.** (2003). The structure of DNA in the nucleosome core. *Nature*, **423**, 145-150.

- Rubinson, K. A., Ladner, J. E., Tordova, M. & Gilliland, G. L.** (2000). Cryosalts: suppression of ice formation in macromolecular crystallography. *Acta Cryst.*, **D56**, 996-1001.
- Rudolph, M. G., Wingren, C., Crowley, M. P., Chien, Y-H. & Wilson, I. A.** (2004). Combined pseudo-merohedral twinning, non-crystallographic symmetry and pseudo-translation in a monoclinic crystal form of the $\gamma\delta$ T-cell ligand T10. *Acta Cryst.*, **D60**, 656-664.
- Sato, K., Shiratori, O. & Katagiri, K. J.** (1967). The mode of action of quinoxaline antibiotics. Interaction of Quinomycin A with deoxyribonucleic acid. *J. Antibiotics*, **20**, 270-276.
- Sayers, E. W. & Waring, M. J.** (1993). Footprinting Titration Studies on the Binding of Echinomycin to DNA Incapable of Forming Hoogsteen Base Pairs. *Biochemistry*, **32**, 9094-9107.
- Searle, M. S.** (1994). Binding of Quinomycin antibiotic UK-65,662 to DNA: ^1H -n.m.r. studies of drug-induced changes in DNA conformation in complexes with $\text{d}(\text{ACGT})_2$ and $(\text{GACGTC})_2$. *Biochem. J.*, **304**, 967-979.
- Sheldrick, G. M., Heine, A., Schmidt-Bäse, K., & Pohl, E.** (1995). Structures of Quinoxaline antibiotics. *Acta Cryst.*, **B51**, 987-999.
- Sheldrick, G. M. & Schneider, T. R.** (1997). SHELXL: High Resolution Refinement. *Methods Enzymol.*, **277**, 319-343.
- Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, M. & Usón, I.** (2001). Ab initio phasing. *International Tables for Macromolecular Crystallography*, Vol. F, edited by E. Arnold & M. Rossmann, 333-345. Dordrecht: Kluwer Academic Publishers.
- Than, M. E., Henrich, S., Bourenkov, G. P., Bartunik, H. D., Huber, R. & Bode, W.** (2005). The endoproteinase furin contains two essential Ca^{2+} ions stabilizing its N-terminus and the unique S1 specificity pocket. *Acta Cryst.*, **D61**, 505-512.
- Tseng, T. D., Ge, H., Wang, X., Edwardson, J. M., Waring, M. J., Fitzgerald, W. J. & Henderson, R. M.** (2005). Atomic Force Microscopy Study of the Structural Effects Induced by Echinomycin Binding to DNA. *J. Mol. Biol.*, **245**, 745-758.
- Ughetto, G., Wang, A. H.-J., Quigley, G. J., van der Marel, G. A., van Boom, J. H. & Rich, A.** (1985). A comparison of the structure of echinomycin and triostin A complexed to a DNA fragment. *Nucleic Acids Res.*, **13**, 2305-2323.
- Wakelin, L. P. G. & Waring, M. J.** (1976). The Binding of Echinomycin to Deoxyribonucleic Acid. *Biochem. J.*, **157**, 721-740.
- Wang, A. H.-J., Ughetto, G., Quigley, G. J., Hakoshima, T., van der Marel, G. A., van Boom, J. H. & Rich, A.** (1984). The Molecular Structure of a DNA-Triostin A Complex. *Science*, **225**, 1115-1121.
- Wang, A. H.-J., Ughetto, G., Quigley & Rich, A.** (1986). Interactions of Quinoxaline Antibiotic and DNA: The Molecular Structure of a Triostin A – $\text{d}(\text{GCGTACGC})$ Complex. *J. Biomol. Struct. Dyn.*, **4**(3), 319-342.
- Ward, D. C., Reich, E. & Goldberg, I. H.** (1965). Base specificity in the interaction of polynucleotides with antibiotic drugs. *Science*, **149**, 1259-1263.
- Waring, M. J. & Wakelin, L. P. G.** (1974). Echinomycin: a bifunctional intercalating antibiotic. *Nature*, **252**, 653-657.
- Waring, M. J.** (1993). *Molecular aspects of Anticancer Drug-DNA Interactions*. Edited by S. Neidle & M. J. Waring, Vol. 1, 213-242. London: Macmillan.
- Windholz, M.** (1983). Editor. *The Merck Index*. 10th ed. Rahway, NJ, USA: Merck & Co. Inc.

PUBLIKATIONEN IN FACHZEITSCHRIFTEN

Structures of complexes between echinomycin and duplex DNA. **Jose A. Cuesta-Seijo** and George M. Sheldrick. Acta Cryst. 2005, **D61**, 442-448.

LEBENS LAUF

PERSÖNLICHE DATEN

Name Jose Antonio Cuesta Seijo
Geburtsdatum 16.11.1975
Geburstort Gijón, Spanien
Staatangehörigkeit spanisch
Familienstand ledig

SCHULBILDUNG

Bis 1993 Colegio Ursulinas de Jesús, Gijón, Spanien

HOCHSCHULSTUDIUM

1993-1999 Diplom Chemie. Spezialisierung in Organischer Chemie
Fakultät für Chemie der Universität Oviedo, Asturias, Spanien

1999-2000 Diplomarbeit in der Abteilung für Phisikalische und Anorganische
Chemie der Universität Oviedo, Spanien im Arbeitskreis von Prof.
Santiago García-Granda: „Protokolle für die Kristallisierung und
Sammlung von Röntgendaten bis zu Atomare Auflösung von
Trypsin-Inhibitor Komplexen“

PROMOTION

2002-2005 Dissertation am Lehrstuhl für Strukturchemie der Georg-August
Universität Göttingen im Arbeitskreis von Prof. George M. Sheldrick:
„Crystallographic studies of Echinomycin-DNA complexes and an
alternating D-alanyl, L-homoalanyl PNA“

Göttingen, den 24.05.2005

Jose Antonio Cuesta Seijo