

# **Crystallographic study on oligonucleotide coiled-coils**

**Thesis submitted for the Degree of Doctor of Philosophy**

**Daniela De Luchi**

**Barcelona, 2008**

**Departament d'Enginyeria Química**  
**Escola Tècnica Superior d'Enginyeria Industrial de Barcelona**  
**Universitat Politècnica de Catalunya**

# **Crystallographic study on oligonucleotide coiled-coils**

**Memoria presentada por Daniela De Luchi  
para acceder al Grado de Doctor en Ciencias.**

**Trabajo realizado en el Departamento  
d'Enginyeria Química de la ETSEIB-UPC,  
dirigido por el Dr. Juan A. Subirana Torrent y  
codirigido por la Dra. J. Lourdes Campos.**

**Barcelona, Junio 2008**

# ACTA DE QUALIFICACIÓ DE LA TESI DOCTORAL

Reunit el tribunal integrat pels sota signants per jutjar la tesi doctoral:

Títol de la tesi: .....

Autor de la tesi: .....

Acorda atorgar la qualificació de:

- No apte
- Aprovat
- Notable
- Excel·lent
- Excel·lent Cum Laude

Barcelona, ..... de/d'..... de .....

El President

El Secretari

.....  
(nom i cognoms)

.....  
(nom i cognoms)

El vocal

El vocal

El vocal

.....  
(nom i cognoms)

.....  
(nom i cognoms)

.....  
(nom i cognoms)

*Ai miei genitori e a Martina*

## Agradecimientos

En primer lugar quiero darle las gracias a mi Director de tesis, Prof. Juan A. Subirana, por haberme dado la posibilidad de realizar una tesis doctoral, por haber sido una fuente de conocimientos, experiencia, recursos y continuas ideas. Le quiero agradecer la ayuda en la realización y redacción de este trabajo, en sus varias fases, y especialmente por la ayuda en la interpretación de los diagramas de difracción aquí presentados.

Quiero darle mis más sinceros agradecimientos a la Dra. Valentina Tereshko, en primer lugar por la ayuda profesional, los consejos y el apoyo y, no menos importante, por haber hecho tan agradable mi estancia en Chicago.

A la Dra. Lourdes Campos, por su amabilidad, la infinita paciencia y la continua disponibilidad y ayuda en más de una ocasión.

A la Dra. Isabel Usón, por el continuo interés y disponibilidad; también le quiero dar las gracias al Prof. G. M. Sheldrick por hacer posibles mis estancias en Goettingen, donde la formación de los estudiantes resulta verdaderamente importante.

Le quiero dar las gracias a mi hermana Martina por el continuo apoyo y las muchas imágenes que me ha ayudado a realizar y que ahora están en esta tesis (*grazie* Architetto Marty!).

Por las correcciones de este manuscrito un “*grazie*” especial a la Dra. Marianna Biadene (¡y sin olvidar las muchas cenas juntas!).

Le quiero dar las gracias a la Dra. Nuria Valls por su amistad y por haberme ayudado y escuchado tantas veces, dentro y fuera del trabajo; también quiero agradecerle a la Dra. Carme Cáceres su disponibilidad y su continuo interés.

A todos los compañeros del departamento, por hacer más agradables las horas pasadas juntos: a las ya doctoras Montse Vera y Meritxell Palau, a Sebastià Gestí, Emma Botines, Laura y Elena, Elsa, Gina y Mireya, a las Dras. Nuria Saperas, Lourdes Urpí, Lourdes Franco, Maria Teresa Casas y a todos los profesores, doctorandos y proyectistas que han estado y están en el departamento.

A las compañeras de piso, Rosa, Dominique y (actualmente) a M<sup>a</sup>. Carmen y Antonia, gracias por haberme hecho sentir como en casa desde el primer día y por acogerme cada vez que vuelvo.

Un “gracias” especial a Jordi, por todo lo que hemos compartido, los viajes, las excursiones y mucho más.

Un immenso grazie ai miei genitori, che nonostante la distanza mi sono stati vicini in tutti questi anni. Grazie per aver appoggiato, cercato di capire o solo accettato le mie scelte, non ce l'avrei fatta senza di voi!

Esta tesis ha sido realizada gracias a una beca del proyecto europeo HPRN-CT-2000-00009 y a la beca pre-doctoral AP2003-2309 otorgada por el Ministerio de Educación y Ciencia de España.

## Abstract

The crystallographic study of the coiled-coils generated by DNA oligonucleotides is the main subject of this thesis.

When the straight axis of a simple helix (minor coil) follows itself a helical path, then the structure is called a coiled-coil (major coil). The parameters that define a DNA coiled-coil are: the inclination  $\beta$  of the oligonucleotides axis with respect to the major coil axis; the number  $N$  of oligonucleotides per turn; the kink angle  $\theta$  and the torsion angle  $\tau$  between consecutive oligonucleotides.

Previous works show that the DNA sequences d(AT)<sub>6</sub> (Campos *et al.*, 2005) and d(AT)<sub>5</sub> generate coiled-coils with very different geometrical characteristics. In order to better understand the properties of these structures, fourteen oligonucleotides with sticky-ended sequences have been crystallized. The presence of a sticky end determines the coiled-coil properties. The sequences studied are (CG)<sub>n</sub>(AT)<sub>m</sub> and (AT)<sub>m</sub>(CG)<sub>n</sub>, and some other very similar to those. The majority of them have  $n = 1$  and  $m > 1$ , so that the sticky end is usually represented by the sequence d(CG).

The geometrical characteristics of the coiled-coils have been studied, and the relation between the aforementioned parameters ( $\beta$ ,  $N$ ,  $\theta$  and  $\tau$ ) has been calculated.

Due to the intrinsically difficult crystallization of such sequences, in several cases only poor diffracting crystals have been obtained and the determination of their atomic structures has not been possible. Despite this, the structure of the sequence dCG(AT)<sub>5</sub> could be determined at 3.1 Å resolution showing unambiguously that the (AT)<sub>5</sub> fragment generates a double helix with Hoogsteen base pairs. It is not clear whether the Hoogsteen base pairing influences or not the geometry of the super-coils and only some hypothesis could be formulated. Recently, a crystal of the short fragment dCG(AT)<sub>2</sub> has been obtained and its diffraction pattern has been measured up to 2.6 Å resolution. At the present time, the structure has not been solved yet and only some preliminary considerations are shown in this work.

As a complementary study, the melting temperatures ( $T_m$ ) of AT-rich oligonucleotides have been determined, and a simple equation for their prediction is shown (see Appendix III.1).

Finally, the structure of the complex of d(U<sup>Br</sup>AGG) with an anthraquinone derivative, previously solved in our laboratory, has been refined (see Appendix III.3).

## List of abbreviations

A	Adenine
C	Cytosine
CC	Correlation coefficient
CCD	Coupled charge device
CSD	Cambridge structural database
DNA	Deoxyribonucleic acid
G	Guanine
HPLC	High performance liquid chromatography
MAD	Multiwavelength anomalous diffraction
MIR	Multiple isomorphous replacement
MPD	2-Methyl-2-pentenediol
MR	Molecular Replacement
NDB	Nucleic acid database
NMR	Nuclear magnetic resonance
PDB	Protein databank
PEG	Polyethylene glycol
RMSD	Root mean square deviation
RNA	Ribonucleic acid
R-WC	Reverse Watson-Crick
SAD	Single wavelength anomalous diffraction
SF	Structure factor
SIR	Single isomorphous replacement
SR	Synchrotron radiation
T	Thymine
TMAO	Trimethylamine <i>n</i> -oxide
U	Uracil
W-C	Watson-Crick

## Table of contents

I.	Theoretical background .....	1
I.1	The structure of the DNA double helix .....	1
I.1.1	Chemical structure .....	2
I.1.2	DNA conformation .....	6
Torsion angles .....	6	
Base pairs, dimer step and helical parameters .....	7	
I.2	X-Ray Diffraction and Macromolecular Crystallography .....	9
I.2.1	Macromolecular crystallization .....	9
Crystallization techniques .....	11	
I.2.2	Crystals and symmetry .....	13
I.2.3	X-Ray sources .....	17
I.2.4	Detectors .....	19
I.2.5	Principles of X-ray diffraction .....	21
I.2.6	The Patterson function .....	25
I.2.7	Data collection .....	26
I.2.8	The Phase Problem .....	29
Molecular Replacement .....	29	
Isomorphous replacement .....	31	
Anomalous scattering .....	32	
Direct methods .....	34	
I.2.9	Structure refinement .....	35
I.2.10	Validation and deposition .....	37
I.2.11	Introduction to fiber diffraction .....	38
Diffraction by helical molecules .....	40	
II.	Crystallographic study on oligonucleotide coiled-coils .....	43
II.1	INTRODUCTION: AT-rich DNA sequences .....	43
d(ATATATATATAT), a dodecamer (Campos <i>et al.</i> , 2005) .....	48	
d(ATATATATAT), a decamer .....	49	
II.2	Determination of the parameters of a coiled-coil .....	51
II.3	Aim of the project .....	52
II.4	Geometry of the coiled-coil .....	53
II.4.1	Calculation of the geometrical parameters of the coiled-coil .....	56

Examples.....	57
The Dodecamer d(CGATATATATAT), [CG(AT)5] .....	58
The Decamer d(CGATATATAT), [CG(AT)4] .....	59
II.5 DODECAMERS .....	62
II.5.1 Structure of the DNA coiled-coil formed by d(CGATATATATAT).....	63
Introduction .....	63
Crystallization.....	64
Data collection and structure determination .....	64
Structure description.....	66
II.5.2 d(CGCGATATATAT) d[(CG) <sub>2</sub> (AT) <sub>4</sub> ].....	70
The D18B2 crystal.....	70
The D17A3 crystal .....	73
II.5.3 d(ATATATATATCG) d[(AT) <sub>5</sub> CG].....	74
II.5.4 d(ATATATATATGC) d[(AT) <sub>5</sub> GC].....	77
II.5.5 d(GCATATATATAT) d[GC (AT) <sub>5</sub> ].....	83
The D34D3 crystal .....	83
The D28C21 crystal .....	86
The D28C1 crystal .....	87
The D34D6 crystal .....	88
II.5.6 d(CGATATGCATAT) d[CG(AT) <sub>2</sub> GC(AT) <sub>2</sub> ].....	90
II.5.7 Discussion of dodecamer structures.....	96
II.6 DECAMERS .....	98
II.6.1 d(CGATATATAT) d[CG(AT) <sub>4</sub> ].....	99
Introduction .....	99
The P10A4 crystal.....	101
Crystallization and Data Collection.....	101
The Diffraction and the Unit Cell.....	102
The P8C2, the P8D41, the P10B1 and the P9C3 crystals.....	104
The P8C2 crystal .....	104
The P8D41 crystal.....	105
The P10B1 crystal.....	106
The P9C3 crystal .....	107
The P8C63 crystal .....	108
Crystallization and Data Collection.....	108

The Diffraction and the Unit Cell.....	108
Packing features of CGATATATAT.....	110
Summary of the CG(AT) <sub>4</sub> structure .....	118
II.6.2    d(ATATATATCG) d[(AT) <sub>4</sub> CG].....	120
II.6.3    d(ATATATATATT) d(AT) <sub>5</sub> T .....	126
II.7    OCTAMERS .....	128
II.7.1    d(CGTATATA) d[CG(TA) <sub>3</sub> ].....	129
II.7.2    d(CGATATAT) d[CG(AT) <sub>3</sub> ] and d(ATATATCG) d[(AT) <sub>3</sub> CG].....	132
d[CG(AT) <sub>3</sub> ].....	132
d[(AT) <sub>3</sub> CG].....	133
II.8    HEXAMER .....	135
II.8.1    d(CGATAT), preliminary considerations .....	135
II.9    SUMMARY AND CONCLUSIONS .....	137
Sequences studied .....	137
Conclusions .....	137
III.    Appendix .....	140
III.1    The influence of size on the thermal stability of oligonucleotides: the case of AT sequences. ....	140
III.2    Structure of the DNA Coiled-coil formed by d(CGATATATATAT).....	143
III.3    An inverted anthraquinone-DNA crystal structure (article in preparation). ....	147
IV.    Bibliography.....	154

# I. THEORETICAL BACKGROUND

## I.1 The structure of the DNA double helix

DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) are the most common nucleic acids; they are present in all cells and viruses and their main role is to carry the genetic information.

DNA was first isolated, in 1869, by the Swiss physician Friedrich Miescher. In 1919, this discovery was followed by Phoebus Levene's identification of the base, sugar and phosphate nucleotide unit. Levene suggested that DNA consisted of a string of nucleotide units linked together through the phosphate groups. However, Levene thought the chain was short and the bases repeated in a fixed order. In 1937, William Astbury produced the first X-ray diffraction patterns that showed that DNA had a regular structure. Frederick Griffith, in 1928, with his experiments with bacteria, provided the first clear suggestion that DNA carried the genetic information. The role of DNA in heredity was confirmed in 1952, when DNA was proven to be the genetic material of the T2 phage.

In 1953, based on X-ray fiber diffraction images taken by Rosalind Franklin, James D. Watson and Francis Crick published in the journal *Nature* what is now accepted as the first accurate model of DNA structure. In the same issue, an article on DNA structure of Maurice Wilkins and his colleagues was also published. In 1962, after Franklin's death, Watson, Crick and Wilkins jointly received the Nobel Prize in Physiology or Medicine. However, debate continues on who should receive credit for the discovery, as the Watson and Crick article in *Nature* was based on Franklin's data without either acknowledgment or her knowledge.

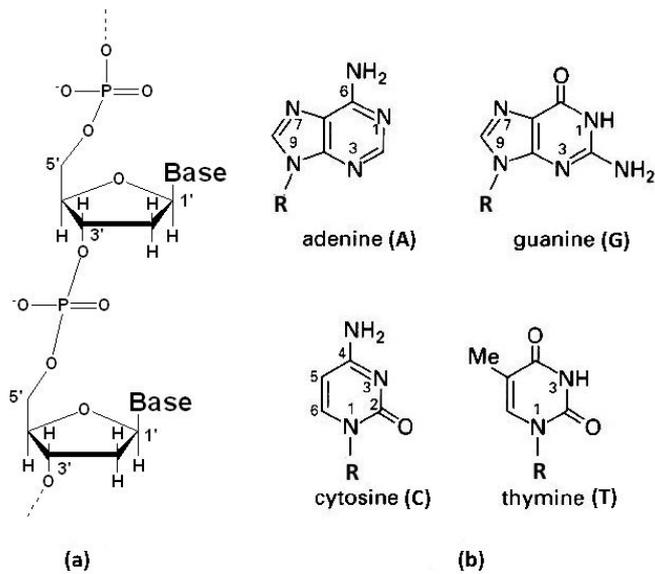
The method of fiber diffraction can only give an average structure of DNA. The understanding of the influence of sequence on the DNA structure became possible only when short DNA fragments with known sequence became available thanks to organic chemistry. It was then in 1980 that in the group of Richard E. Dickerson the structure of the dodecamer d(CGCGAATTCGCG) was determined confirming the validity of the model derived twenty years before from fiber diffraction data.

### **I.1.1 Chemical structure**

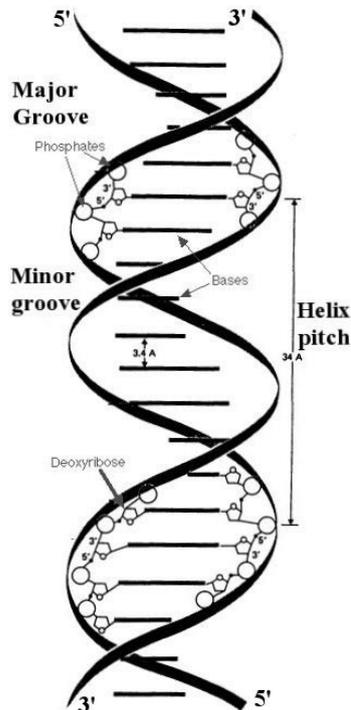
Nucleic acids are polymers of simple units called nucleotides. Nucleotides are made up of three components: base, sugar and phosphate.

The sugar is the ribose for RNA and the deoxyribose for DNA. There are four standard bases: two purines (Pur or R) and two pyrimidines (Pyr or Y). The purines are adenine (Ade or A) and guanine (Gua or G); the pyrimidines are cytosine (Cyt or C) and thymine (Thy or T) in the DNA and cytosine and uracil (Ura or U) in the RNA. The nucleotides are linked by phosphodiester bonds between the 5' carbon of the sugar in one nucleotide and the 3' carbon in the next sugar along the chain (the chain runs in the 5'→3' direction of the riboses alternating phosphates and sugars). The nucleobases are linked to the 1' carbon of the sugars. The chemical structures of a single phosphodiesteric chain of DNA and of the four standard DNA nucleobases are shown in Figure I.1.1.

The nucleobases form hydrogen bonds with the complementary base, linking two DNA chains together to form a double helix; guanine normally binds to cytosine and adenine to thymine. Both chains run antiparallel to each other. The bases are located in the inside of the helix and the phosphates in the outside, the path taken by the two backbones form a major (wider) groove and a minor (narrower) groove (see Figure I.1.2).



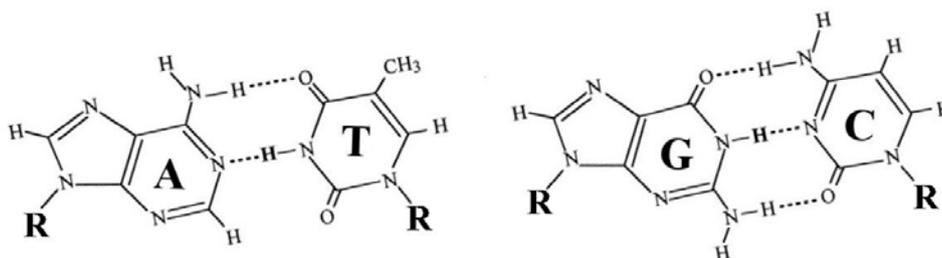
**Figure I.1.1.** (a) Chemical structure of a single DNA chain: the phosphates are linked to the 5' carbon of one deoxyribose and to the 3' carbon of the next ribose. The bases (A, G, C and T) are linked to the 1' carbon of the sugar. (b) Standard DNA nucleobases. They offer a variety of hydrogen bond donor and acceptor sites that originate the interactions with the second strand in duplex DNA and sometimes with a third or a fourth chain. Uracil (not shown) has the same donor-acceptor scheme of thymine.



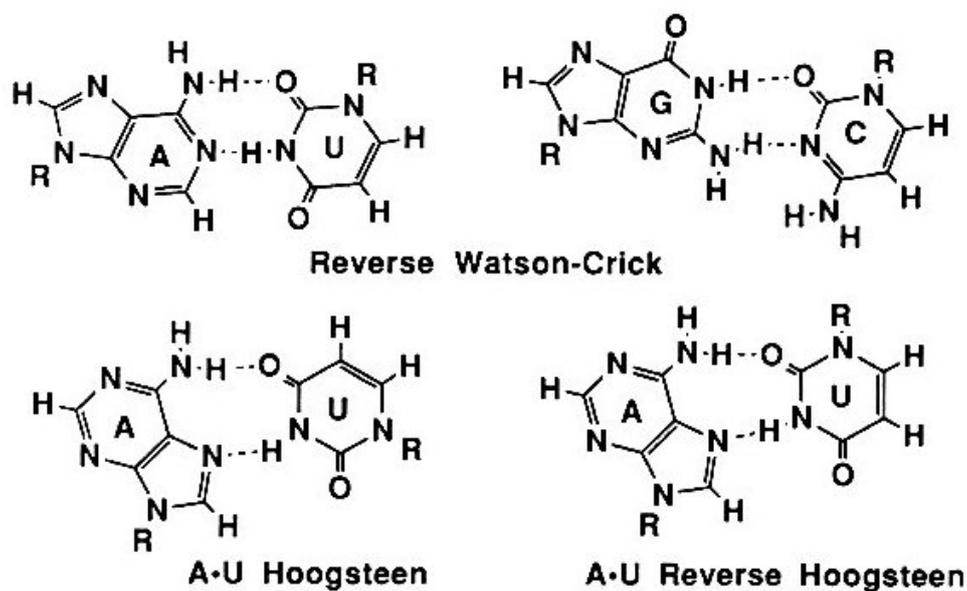
**Figure I.1.2.** Schematic view of an antiparallel DNA double helix. The helical pitch of about 34 Å is indicated. The major and minor grooves generated by the double helix are shown.

Nucleic acids commonly form helical secondary structures using two or more strands. The most common DNA conformation *in vivo* is represented by the B-form. The B-DNA can be described by the well known Watson - Crick Model of the double helix, derived in 1953 from the X-ray diffraction pattern of a DNA fiber. The base of this model is the specific recognition between a purine and a pyrimidine base: adenine with thymine (uracil in the RNA) and guanine with cytosine. Standard Watson-Crick hydrogen bonds are shown in Figure I.1.3. These combinations lead to virtually identical base geometries. This identity was the basis of the realization that it is possible to build a regular double helix with an arbitrary sequence and it was also the basis for understanding the replication of the genetic code. As shown in Figure I.1.3, A-T pairs are maintained by two hydrogen bonds, while C-G pairs present three hydrogen bonds.

Reverse Watson-Crick base pairing is also possible. This binding mode often occurs in parallel-stranded DNA. In this case, only two hydrogen bonds are found between guanine and cytosine (Figure I.1.4). The most important pairing alternative to the Watson-Crick is probably the Hoogsteen scheme (see Figure I.1.4). It 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 N6 (AT base pair) or O6 (CG base pair). Protonation of cytosine is a prerequisite for CG Hoogsteen base-pairs to occur.

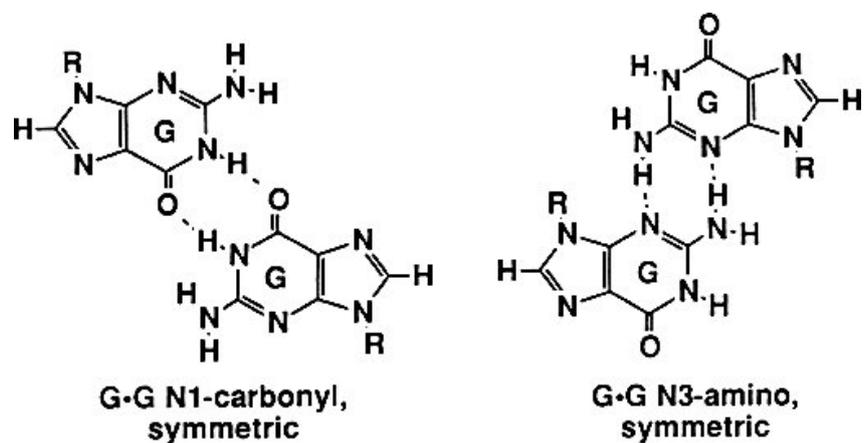


**Figure I.1.3.** The canonical Watson-Crick basepairs, stabilized by two H-bonds in the AT base pair and by three H-bonds in the CG base pair.



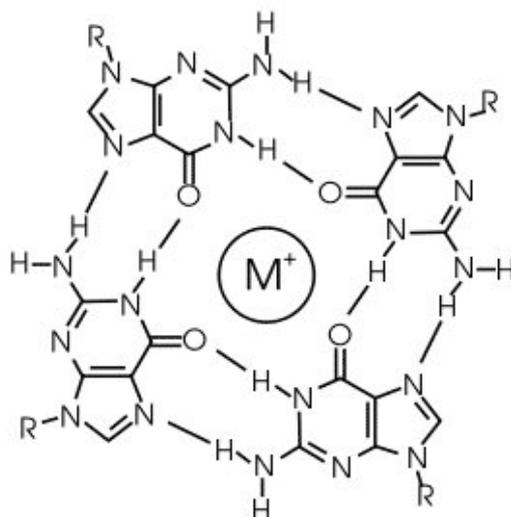
**Figure I.1.4.** Reverse Watson-Crick, Hoogsteen and Reverse Hoogsteen hydrogen bonds. CG Hoogsteen pairing (not shown) is also possible and commonly found in nucleic acids triple helices; protonation of cytosine at N3 is necessary for it to occur.

Homo-purinic R:R and homo-pyrimidinic Y:Y base pairings are also possible, in particular guanine-guanine base pairs are easily found in G-rich sequences (see Figure I.1.5).



**Figure I.1.5.** Homo-purinic guanine-guanine symmetric hydrogen bonds.

Guanines rich sequences tend to form four-stranded (or quadruplex) structures called G tetrads, in which the guanines interact through their Watson-Crick and Hoogsteen edges (Figure I.1.6).



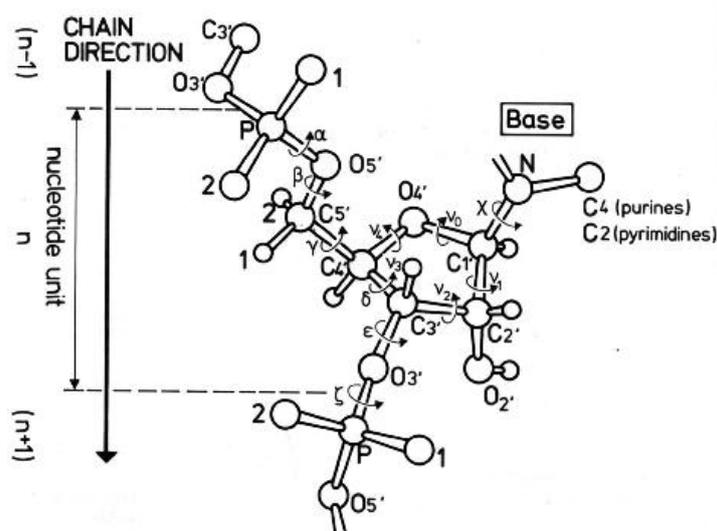
**Figure I.1.6.** G-quadruplex hydrogen bonding scheme: the tetrads are normally stabilized by a monovalent cation (as  $K^+$  or  $Na^+$ ) in the centre of the structure. The Hoogsteen hydrogen bondings between adjacent guanines are shown

## I.1.2 DNA conformation

### Torsion angles

Nucleic acids chains are highly flexible due to the torsion angles in the sugar-phosphate backbone (Figure I.1.7). The backbone torsion angles are as follows:  $\alpha$  (P-O5'),  $\beta$  (O5'-C5'),  $\gamma$  (C5'-C4'),  $\delta$  (C4'-C3'),  $\epsilon$  (C3'-O3') and  $\zeta$  (O3'-P). Since the sugar forms a ring, the intracyclic torsion angles of the sugar rings,  $\nu_0$ - $\nu_4$ , are dependent on one another. Deviations from planarity forces one of the ring atoms out of the plane, the net effect is termed the "sugar puckering". The face of the sugar ring that is toward the glycosidic bond is termed the "endo" face, the face that is away is termed the "exo" face. C2'-endo is typical of B-DNA while C3'-endo is the most common sugar conformation in A-form RNAs and DNAs.

The glycosidic bond goes from C1' to N9 for purines and from C1' to N1 for pyrimidines. The permitted angles of rotation  $\chi$  exist in two regions: *anti* and *syn*. The *anti* conformation corresponds to  $\chi = 180^\circ \pm 90^\circ$ , the *syn* conformation corresponds to  $\chi = 0^\circ \pm 90^\circ$ . The latter value is typical for the purines involved in Hoogsteen base pairing.

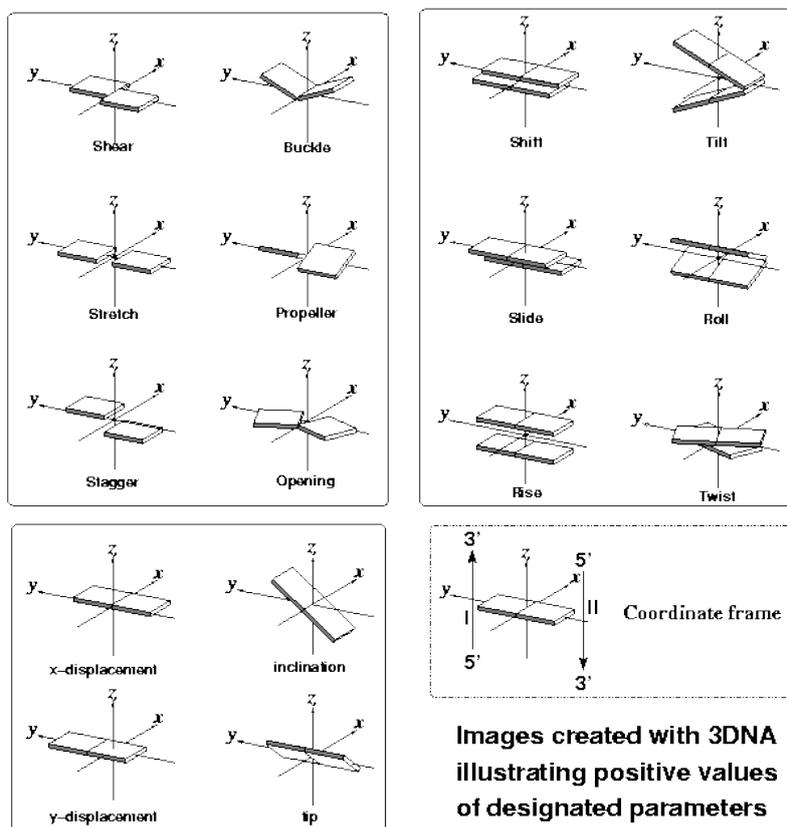


**Figure I.1.7.** Torsion angles in the nucleic acid backbone.

### Base pairs, dimer step and helical parameters

A common point of reference is needed to describe the three-dimensional arrangements of bases and base pairs in nucleic acid structures. The program 3DNA (Lu & Olson, 2003) used in this work implements the parameter sets derived from the *Tsukuba Workshop on Nucleic Acid Structure and Interactions* held in 1999 and recommended by the NDB (Berman *et al.*, 1992). There are two sets of local parameters commonly in use in nucleic acid conformational analysis: step parameters which show the stacking geometry between neighbor base-pairs, and helical parameters which demonstrate the position and orientation of a base-pair relative to the helical axis (Figure I.1.8). The values of local vs helical rise and twist from these two sets of parameters can be quite different in DNAs which deviate significantly from B-form DNA.

Although very useful for standard Watson-Crick hydrogen bonds, in the case of Hoogsteen basepairs, the purine rotation into the *syn* conformation and the different hydrogen bonding scheme makes most of the aforementioned parameters unusable at least for comparison with standard values.



**Figure I.1.8.** Standard reference frame for the description of nucleic acid base-pair geometry as recommended in the NDB (Berman *et al.*, 1992).

## **I.2 X-Ray Diffraction and Macromolecular Crystallography**

### **I.2.1 Macromolecular crystallization**

Obtaining crystals in many cases remains a trial and error process and frequently represents the rate limiting step in the determination of macromolecular structures. Crystals are grown by slow, controlled decrease of the solubility of the macromolecule, usually dissolved in an aqueous solution. The basic requirements for a crystallization experiment are the purity and homogeneity of the sample.

All the oligonucleotides studied in this thesis have been synthesized on an automatic synthesizer by the phosphoramidite method and purified by gel filtration and reverse phase HPLC at the Institut Pasteur of Paris.

A large number of variables have an influence on the macromolecule crystallization and there is no or very limited a priori information about which one must be modified. Many variables can influence the crystallization, among them: pH, ionic strength, DNA (protein) concentration, nature and concentration of the precipitant (MPD, PEG, spermine, etc), temperature, nature and concentration of ions and additives.

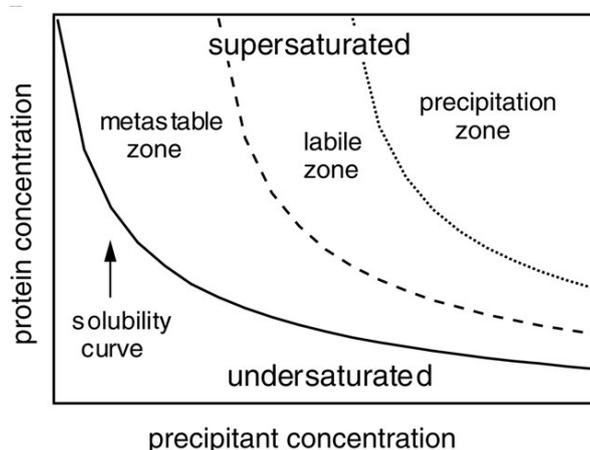
A typical macromolecular phase diagram is shown in Figure I.2.1. Crystals dissolve in the undersaturated region, where the concentration is below the macromolecule solubility, and grow in the supersaturated region. A relatively large supersaturation is required to overcome the activation energy barrier which exists when forming the crystal (Asherie, 2004). If the supersaturation is too large, then disordered structures, such as aggregates or precipitates, may form.

There are three stages of crystallization common to all systems: nucleation, growth and cessation of growth.

Nucleation is the process by which molecules or non-crystalline aggregates which are free in solution come together in such a way to produce a thermodynamically stable aggregate with a repeating lattice, the first semblance of the solid state. The nuclei can be formed in the labile zone of the supersaturation area of the phase diagram. The degree to which nucleation occurs is determined by the degree of the supersaturation of the solute.

Crystal growth generally starts at solute concentration sufficient for nucleation to occur, and continues at concentration below the nucleation threshold. After the formation of the first nuclei the concentration of the solute is slightly decreased, the nucleation stops and the crystals formed can grow to a bigger size without the competition of new nuclei.

Cessation of growth can occur for different reasons. The most obvious is the decrease in concentration of the crystallizing solute to the point where the solid and solution phases reach exchange equilibrium.



**Figure I.2.1.** A schematic phase diagram showing the solubility of a macromolecule in solution as a function of the precipitant concentration (Mc Pherson, 1999).

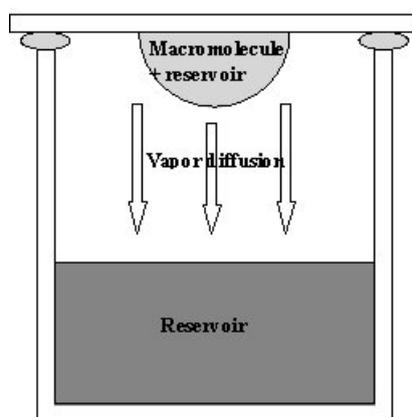
## Crystallization techniques

**Vapor diffusion** is the most common method of crystallization. A schematic representation of a hanging drop is shown in Figure I.2.2. A small amount of oligonucleotide solution is mixed with the precipitant (typical volumes range from 1 to 10  $\mu\text{L}$ ) and placed on a siliconised cover slip. The drop is then suspended and sealed over the well solution. The difference in precipitant concentration between the drop and the well solution is the driving force which causes water to evaporate from the drop until the concentration of the precipitant in the drop equals that of the well solution. The sitting drop is a variation of the method which allows bigger volumes to equilibrate.

**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.

**Batch method** In the batch method, concentrated protein is mixed with concentrate precipitant solution to produce a final concentration which is supersaturated in terms of the solute macromolecule and therefore leads to crystallization. This can be done with up to ml amounts of solution and typically results in larger crystals due to the larger volumes of solute present and the lower chance of impurities diffusing to the face of the crystal. This technique is by far the most expensive in terms of consumption of the solute macromolecule.

**Liquid-liquid diffusion** The protein and precipitant solutions are layered on top of each other allowing a slow equilibration. Nucleation and crystal growth generally occurs at the interface between the two layers, at which both the concentrations are at their highest values.



**Figure I.2.2.** Vapor diffusion from a hanging drop. The macromolecular concentration in the drop increases over time.

DNA (and protein) crystals are characterized by relatively high water content, typically ranging from 30% to 80% of the volume. The typical volume per base pair,  $V_{bp}$ , varies from about  $1300 \text{ \AA}^3$  per base pair, for medium/high resolution structures, to about  $2000 \text{ \AA}^3$  per base pair, for low resolution structures. Real crystals are mosaics of many submicroscopic arrays in rough alignment with each other; this phenomenon is much more pronounced in macromolecular crystals than in crystals of rigid organic or inorganic molecules.

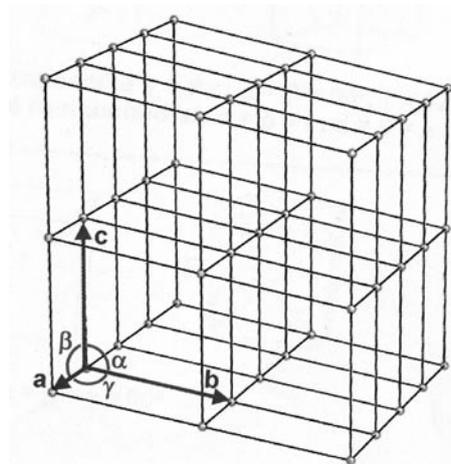
## I.2.2 Crystals and symmetry

Crystals are made up of identical parallelepiped-shaped blocks called unit cells that constitute a three dimensional translation lattice (Figure I.2.3). The cell is defined by the vectors **a**, **b** and **c**; they define the length *a*, *b*, *c*, and the angles  $\alpha$ ,  $\beta$ ,  $\gamma$  which characterize the unit cell. The volume *V* of the unit cell can be calculated as follows:

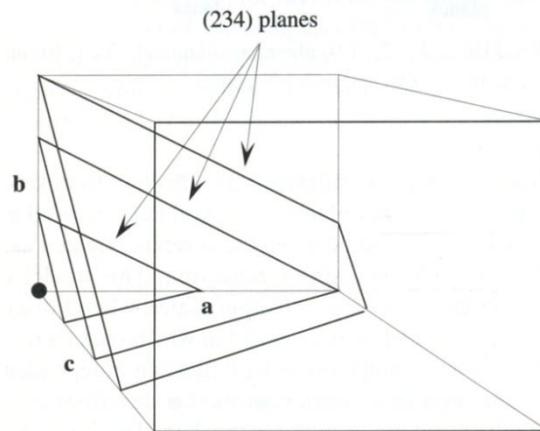
$$V = abc[1 - \cos^2 \alpha - \cos^2 \beta - \cos^2 \gamma + 2 \cos \alpha \cdot \cos \beta \cdot \cos \gamma]^{1/2}.$$

The unit cell is the smallest unit that can generate the entire crystal by translation operations alone. The content of the unit cell is obtained by repetition of a single object through the symmetry elements. This part of the unit cell is called the asymmetric unit. Within the cell there can be several symmetry related asymmetric units with identical contents, but in general in different orientations. The cell is always chosen so that the symmetry elements are positioned in accord with volume A of the *International Tables for Crystallography*.

To define the planes in the crystal the Miller indices (*h*, *k*, *l*) have been introduced. The *h*, *k*, *l* terms define parallel planes with intercepts *a/h*, *b/k*, *c/l* on the three *a*, *b*, *c* axes of the unit cell with *h*, *k*, *l* small integer numbers. For example, the (234) planes, shown in Figure I.2.4, cut the unit cell edges **a** into two parts, **b** into three parts and **c** into four parts.



**Figure I.2.3.** Diagram of the lattice created by the translation of the unit cell; the vectors **a**, **b** and **c** and the angles  $\alpha$ ,  $\beta$  and  $\gamma$  are indicated.



**Figure I.2.4.** The intersection of three (234) planes with a unit cell. Note that the (234) planes cut the unit cell edges **a** into two parts, **b** into three parts and **c** into four parts (Rhodes, 1993).

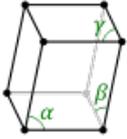
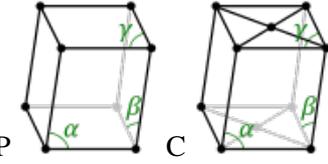
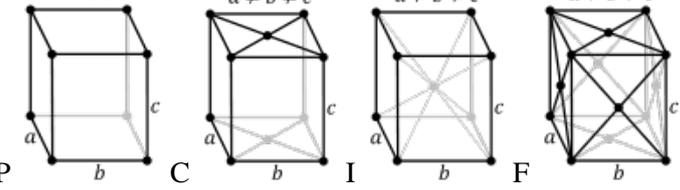
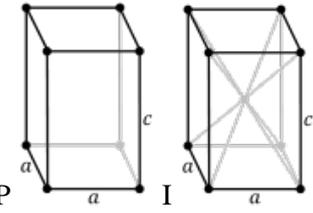
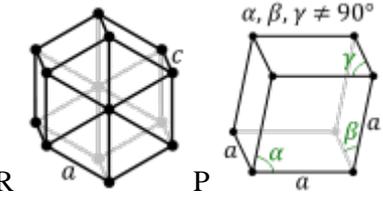
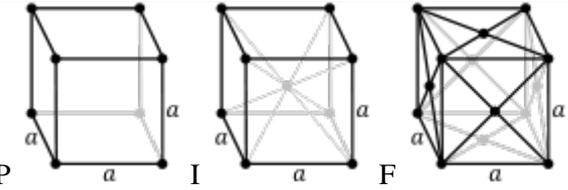
The symmetry of each finite object such as a molecule can be described by a self-consistent set of symmetry operations called a point group. The point group is thus the name given to the collection of symmetry elements of a finite object. There are 32 classes of point groups, given by the combination of the following symmetry elements:

- Mirror plane, which does not occur in crystals of proteins and DNA because they are chiral molecules.
- Rotation axis, characterized by a rotation about one axis of  $360^\circ/N$  where  $N$  can only be 1, 2, 3, 4 or 6.
- Inversion point, which does not occur in crystals of proteins because they are chiral molecules.

By analysis of the rotational symmetry, crystals can be divided into seven crystal systems with defined characteristics and parameters.

There are fourteen Bravais Lattices which are categories of translation lattices: they can be primitive (*P*), body centred (*I*), face centred (*F*) and *C* lattices in the case of Monoclinic and orthorhombic systems having a position on the (001) face. The seven crystal systems and the fourteen Bravais Lattices are shown in Table I.1.

**Table I.1.** The seven crystal systems and the fourteen Bravais Lattices.

The seven crystal systems	The fourteen Bravais Lattices
Triclinic $a \neq b \neq c$ $\alpha \neq \beta \neq \gamma$	$\alpha, \beta, \gamma \neq 90^\circ$ 
Monoclinic $a \neq b \neq c$ $\alpha = \gamma = 90^\circ \neq \beta$	$\alpha \neq 90^\circ$ $\beta, \gamma = 90^\circ$ $\alpha \neq 90^\circ$ $\beta, \gamma = 90^\circ$ 
Orthorhombic $a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	$a \neq b \neq c$ $a \neq b \neq c$ $a \neq b \neq c$ $a \neq b \neq c$ 
Tetragonal $a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	$a \neq c$ $a \neq c$ 
Hexagonal $a = b \neq c$ $\alpha = \beta = 90^\circ, \gamma = 120^\circ$ Rhombohedral $a = b = c$ $\alpha = \beta = \gamma$	$a \neq c$ $\alpha, \beta, \gamma \neq 90^\circ$ 
Cubic $a = b = c$ $\alpha = \beta = \gamma = 90^\circ$	

Other symmetry elements are:

- Glide plane. Obtained by a combination of a mirror plane and a translation, it is not possible for chiral crystals.
- Screw axis. A rotation is combined with a translation parallel to the rotation axis. The molecule is shifted by a fraction of unit cell and rotated.

The combination of the 32 point groups with the Bravais Lattice and the screw axis and glide plane symmetry operations gives rise to 230 space groups of which only 65 are possible for chiral molecules.

The diffraction pattern of a crystal exhibits the same crystal symmetry but with an additional centre of symmetry, in the absence of anomalous scattering. The diffraction pattern symmetries are grouped in 11 Laue classes. The presence of symmetry elements like screw axes can be detected since they give rise to systematic absences in the diffraction pattern. The space group can often, but not always, be found unambiguously considering the Bravais Lattice, the Laue symmetry and the systematic absences.

A special case of symmetry is non-crystallographic symmetry (NCS) through which the molecules within one asymmetric unit are related by appropriate operations.

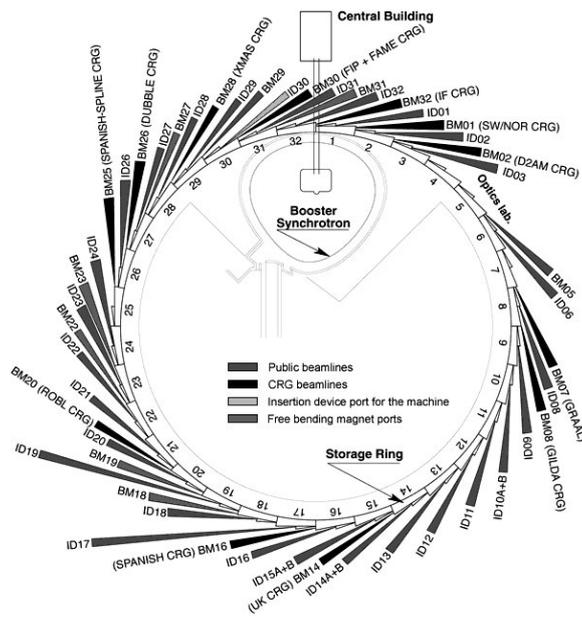
### I.2.3 X-Ray sources

X-rays in the useful range for crystallography can be produced in two different ways. In conventional generators radiation is produced by bombarding a metal target with electrons, while for synchrotron radiation X-rays are obtained by accelerating particles in storage rings.

**Laboratory sources** X-rays are obtained by bombarding a metal target (most commonly copper or molybdenum) with electrons produced by a heated filament and accelerated by an electric field. A high-energy electron collides with and displaces an electron from a low-lying orbital in a target metal atom. Then, an electron from higher orbital drops into the resulting vacancy and emits its excess of energy as an X-ray photon. The most common metals used to produce X-rays are copper ( $K_{\alpha}$  at 1.54 Å) and molybdenum ( $K_{\alpha}$  at 0.71 Å).

**Synchrotron radiation (SR)** X-rays are obtained by accelerating particles (electrons or positrons) that circulate at velocities near the speed of light in a storage ring. Particle storage rings are the most powerful X-ray sources; a schematic representation of the ESRF of Grenoble is shown in Figure I.2.5. The storage ring is not circular but consists of straight and curved parts. Electrons or positrons circulate in bunches at velocities near the speed of light. Radiation is produced either at the curved parts (Bending Magnet) or with the help of Insertion Devices like Wigglers or Undulators.

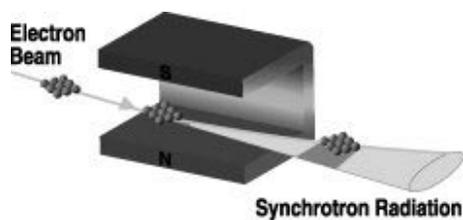
Apart from the high intensity, an additional advantage of using synchrotron radiation is the tunability of the wavelength and potentially its pulse length for time resolved experiments (Laue diffraction). The tunability allows the use of shorter wavelengths to minimise absorption as well as to perform multiple wavelength anomalous dispersion (MAD) experiments selecting the wavelength at the characteristic energy absorption edge.



**Figure I.2.5.** Scheme of the European Synchrotron Radiation Facility, ESRF, of Grenoble, France.

To produce radiation, electrons or positrons are first accelerated in a LINAC (linear accelerator); the particles are then injected in bunches directly into the storage ring or into a so called booster in which they are brought to the desired speed before being redirected in the storage ring. In the storage ring the bunches of particles travel in the vacuum focused horizontally and vertically by quadrupole magnets and bent by bending magnets.

The emission of radiation occurs when the particles deviate in their trajectory and bending magnets (BM) are the first source of synchrotron radiation, as shown in Figure I.2.6.



**Figure I.2.6.** Synchrotron radiation produced at a bending magnet (BM).

As it is shown in Figure I.2.7, the insertion device (ID) is comprised of rows of magnets with alternating polarity; Insertion Devices are installed in a straight section of the electron orbit.

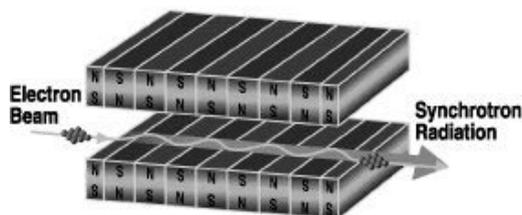


Figure I.2.7. Synchrotron radiation produced at an insertion device (ID).

## I.2.4 Detectors

The development of detectors for X-ray crystallography has been very fast in the last few years, stimulated by the increasing use of synchrotron radiation which requires not only good quality but also very fast detectors. The most important performance characteristics of a detector are its sensitivity, dynamic range, resolving power and readout time.

Sensitivity is most often quantified by the Detective Quantum Efficiency (DQE). DQE is the ratio of the sensitivity of the detector to an ideal photon-counting detector (Naday *et al.*, 1994; Stanton *et al.*, 1993). DQE depends on the intensities of the Bragg reflection. All analog detectors exhibit higher DQE for strong reflections.

Dynamic range quantifies the largest signal that can be recorded before saturation of the detector relative to the smallest signal that can be distinguished above the noise floor. To record the entire Bragg diffraction pattern it is necessary for the dynamic range of the detector to be at least equal to the range of reflection intensities of the sample of interest. A large dynamic range is useful because it makes less critical to select the optimal exposure time for the experiment.

The resolving power of the detector quantifies the ability to resolve very close spots, allowing achieving (near) atomic resolution data for samples with large unit cells without changing the  $2\theta$  position of the detector.

The readout downtime is the time during which the X-ray beam must be shuttered off in order to read out the integrated signal in the detector.

**Image Plate** The heart of the Image Plate is a storage phosphor screen. When the storage phosphor is exposed to X-rays, secondary electrons are trapped in color centers, whose number is proportional to the X-ray energy. After the exposure, these metastable centers can be excited by a red laser to release visible photons in a process known as photostimulation or bleaching. To completely erase the remaining, unbleached centers after the plate is read, it is exposed to an intense, broadband light source for some tens of seconds. This process brings the phosphor back to the ground state. The biggest advantage of this scheme is that it allows a relatively large active area (up to 345 mm diameter). Perhaps, their biggest disadvantage is their relatively long readout time since it typically takes several tens of microseconds to bleach each pixel on the image plate (Nave, 1999). The total readout time for the entire plate is typically on the order of 1 to 2 minutes. This long readout time is a serious disadvantage in experiments at synchrotron beamlines. The other principle disadvantage of the Image Plate is its relatively low sensitivity.

**Charge Couple Device** CCD-based detectors were developed in order to address the long readout times and the low sensitivity of the Image Plates. A schematic representation of a CCD detector is given in Figure I.2.8. The X-rays excite a scintillator screen to produce visible photons, a fiber optic taper transfers the light photons to the CCD chip, in which the photons induce a charge generation. The charges are then transferred and detected. The CCD reads out the photon almost instantaneously and immediately after is ready for a new exposure.

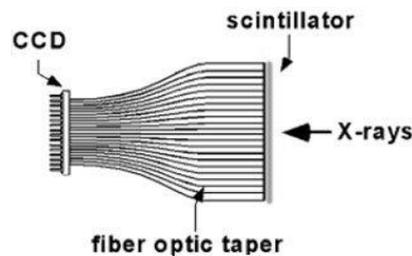


Figure I.2.8. X-rays excite a phosphor screen, producing visible photons which are focussed onto a CCD imager using a fiber optic taper.

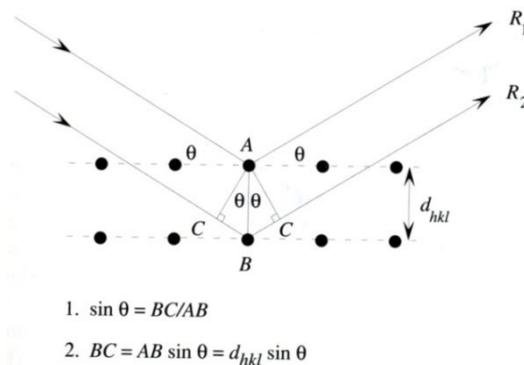
## I.2.5 Principles of X-ray diffraction

X-rays are electromagnetic waves characterized by a wavelength  $\lambda$  in the range of 0.1-100 Å. The interaction between the traveling waves and the electrons in the crystal gives rise to scattering. Two kinds of scattering take place: Thomson, also called elastic or coherent scattering (no exchange of energy with the molecules takes place), and Compton or inelastic scattering. In X-ray crystallography, coherent scattering gives rise to diffraction.

In 1913, W. L. Bragg and his son showed that diffraction could be regarded as if it were reflection from sets of equivalent, parallel planes of atoms in a crystal. The planes are designated by a set of three numbers called lattice or Miller indices,  $hkl$ . The index  $h$  gives the number of parts into which the set of planes cut the edge  $a$  of each cell; the indexes  $k$  and  $l$  respectively give the number of parts into which the set of planes cut the edges  $b$  and  $c$  (Figure I.2.4). Each set of parallel planes is treated as an independent diffractor and produces a single reflection. W. L. Bragg showed that a set of parallel planes with indexes  $hkl$  and interplanar spacing  $d_{hkl}$  produces a diffracted beam only when the angle of incidence  $\theta$  of the X-rays of wavelength  $\lambda$  meets the following condition (the so called Bragg's law):

$$\lambda = 2d_{hkl} \sin \theta \quad (\text{I.1})$$

The geometric construction and the equations in Figure I.2.9 show the necessary conditions for producing a strong diffracted ray.

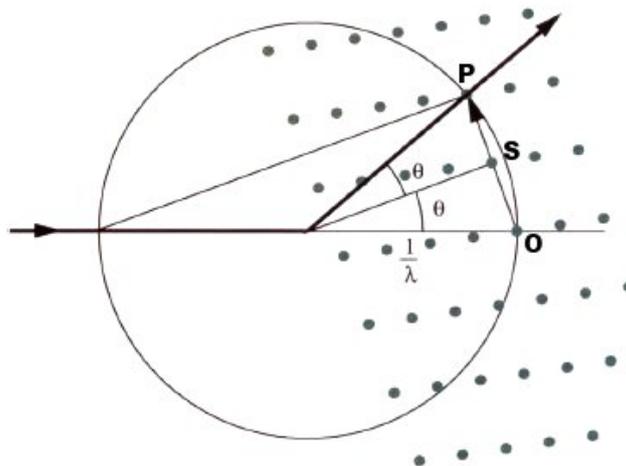


**Figure I.2.9.** Conditions that produce strong diffracted rays. If the additional distance traveled by the more deeply penetrating ray  $R_2$  is an integer multiple of  $\lambda$ , then rays  $R_1$  and  $R_2$  interfere constructively (Rhodes, 1993).

If this difference in path length for rays reflected from successive planes is equal to an integral number of wavelengths (that is, if  $\lambda = 2d_{hkl} \sin \theta$ ), then the rays reflected from successive planes emerge from the crystal in phase with each other, interfering constructively to produce a strong diffracted beam.

The  $hkl$  planes can be described through a vector  $\mathbf{S}$  normal to the  $hkl$  plane and of length  $1/d_{hkl}$ . The points at the end of these vectors form the reciprocal lattice. The reciprocal lattice is spatially linked to the crystal because of the way the lattice points are defined, so if the crystal is rotated, the reciprocal lattice rotates with it. Each reciprocal lattice point must be arranged with respect to the X-ray beam in order to satisfy Bragg's law and produce a reflection from the crystal.

The Bragg's law of diffraction is illustrated in three dimensions by the Ewald sphere (shown in Figure I.2.10). The radiation of wavelength  $\lambda$  is represented by a sphere of radius  $1/\lambda$ . The crystal is represented by the reciprocal lattice with its origin at the point O on the Ewald sphere where the beam leaves it. If the reciprocal lattice point P lies on the surface of the Ewald sphere, the length of the vector  $\mathbf{S}$ , perpendicular to the reflecting plane  $hkl$ , is  $2 \sin \vartheta / \lambda = 1/d$ , that is the Bragg's law.



**Figure I.2.10.** The Ewald construction. When a reciprocal lattice point with indices  $hkl$  lies on the surface of the Ewald sphere, the interference condition for that particular reflection is fulfilled (Dauter, 1999).

X-rays are significantly scattered basically only by electrons. The amplitude of scattering for an atom is known as the *atomic scattering factor* and is described by the following equation:

$$f = \int_V \rho(\mathbf{r}) \exp[2\pi i \mathbf{r} \cdot \mathbf{S}] d\mathbf{r} \quad (\text{I.2})$$

where  $\rho(\mathbf{r})$  is the electron density of an atom, at position  $\mathbf{r}$ , and  $\mathbf{S}$  is the scattering vector. The atomic scattering factor depends on the length of  $\mathbf{S}$  (since  $\mathbf{S} = \frac{2\sin\theta}{\lambda}$ ) but is independent of the direction of the vector  $\mathbf{S}$ . The bigger the angle  $\theta$  (thus higher the resolution) the smaller is the scattering factor.

Since the atoms scattering the X-rays are not fixed in their position, but vibrate around an equilibrium position their scattering factor is affected. The motion is dependent on the temperature. The scattering factor diminishes because of thermal vibration especially at high diffraction angles. In order to account for atomic and molecular vibrations, the atomic scattering factor must be corrected as follows:

$$f' = f e^{\left(-B \frac{\sin^2\theta}{\lambda^2}\right)} \quad (\text{I.3})$$

where  $B = 8\pi^2 u^2$ . For structure with low resolution only the isotropic temperature factor can be refined while for high (near to atomic) resolution structures the anisotropic B factor can be refined (with several, usually six, atomic displacement parameters).

The X-ray radiation scattered by one unit cell is known as the *structure factor* and symbolized by  $F$  or  $F(hkl)$ . It is the Fourier transform of the scattering density (electrons in the molecule) sampled at the reciprocal lattice point  $hkl$ . The intensity of the scattered radiation is proportional to the square of the amplitude,  $|F|^2$ . The structure factor is represented by:

$$F(hkl) = |F(hkl)| e^{-i\alpha(hkl)} \quad (\text{I.4})$$

with  $|F(hkl)|$  representing the amplitude of the scattered wave, and  $\alpha(hkl)$  its phase relative to the origin of the unit cell.

$F(hkl)$  can also be written as the sum of contributions from each volume element of electron density in the unit cell:

$$\mathbf{F}(hkl) = \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(x, y, z) \exp[\pi i (hx + ky + lz)] dx dy dz \quad (\text{I.5})$$

The structure factor is the Fourier transform of the electron density and vice versa, therefore the electron density can be written as follows:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| e^{[-2\pi i(hx+ky+lz)+i\alpha(hkl)]} \quad (\text{I.6})$$

While structure amplitudes are directly obtained from measured reflection intensities, the phases are lost. This is known as the crystallographic phase problem, methods for obtaining an initial set of experimental phases will be described in paragraph I.2.8.

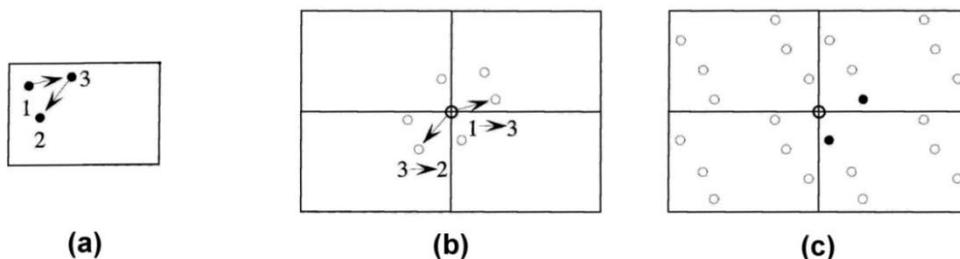
## I.2.6 The Patterson function

If the Fourier transform, used to calculate the electron density map, is written with all the phase angles  $\alpha = 0$ , the so called Patterson function is obtained:

$$P(uvw) = \frac{1}{V} \sum_{hkl} |F(hkl)|^2 \cos[2\pi(hu + kv + lw)] \quad (\text{I.7})$$

where  $u$ ,  $v$ ,  $w$  are the coordinates of the Patterson cell.  $P(\mathbf{u})$  (where  $\mathbf{u} = uvw$ ) can be considered as the convolution of the electron density with itself. The result is that the high values of  $P$  happens at positions  $\mathbf{u}$  corresponding to an interatomic distance vector. A simple example of construction of Patterson map is shown in Figure I.2.11 (Rhodes, 1993).

The number of peaks in a Patterson map is  $N^2$ , reduced to  $N \times (N - 1)$  because  $N$  of them are located at the origin. The Patterson map represents an important tool for structure determination as it is the essence of the Molecular Replacement, and it is the basis for finding the heavy atoms positions in the Isomorphous Replacement or Multiple Anomalous Dispersion Methods.

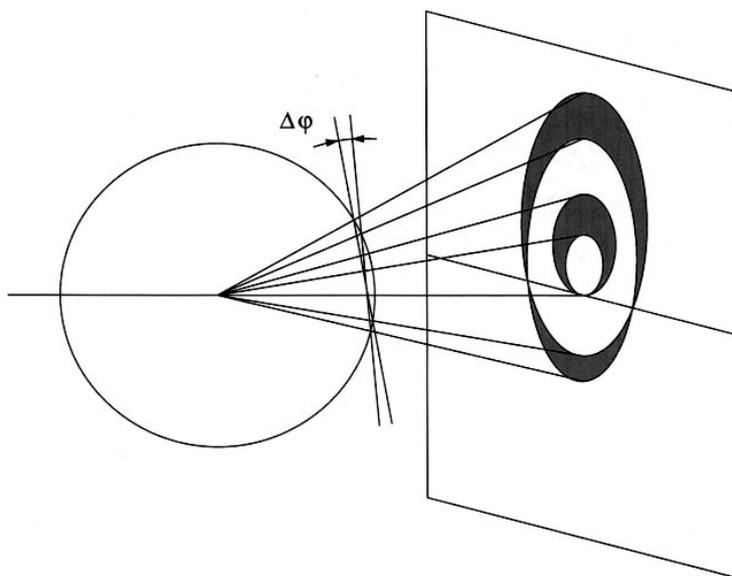


**Figure I.2.11.** Construction of a Patterson map. (a) Structure of unit cell containing three atoms. Two of the six interatomic vectors are shown. (b) Patterson map is constructed by moving all interatomic vectors to the origin. Patterson “atoms” (peaks in the contour map) occur at the head of each vector. (c) Complete Patterson map, containing all peaks from (b) in all unit cells. Peak at origin results from self-vectors. Image of original structure is present (origin and two darkened peaks) among other peaks. (Rhodes, 1993).

## I.2.7 Data collection

Diffraction from a crystal is obtained when Bragg's law is fulfilled. Ewald sphere illustrates Bragg's law of diffraction in three dimensions. When a reciprocal lattice point lies at the surface of the Ewald sphere, the interference condition for that particular reflection is fulfilled and it gives rise to a diffracted beam. When the crystal is not rotated during the X-ray exposure the diffraction pattern (called a "still" photograph) will consist of spots arranged in a set of concentric ellipses. As shown in Figure I.2.12, in the rotation method, the start and end orientations of the diffracting plane form two intersecting ellipses with all reflections recorded between them in the form of a lune.

Due to the crystal mosaicity (crystals are composed of small blocks slightly misoriented with respect to one another) and the beam divergence (the incident radiation is not directed precisely along one line), the diffraction corresponding to a particular reflection is spread over a range of crystal rotation, some reflections come into the diffracting position during one exposure and finish during the next (Dauter, 1999).



**Figure I.2.12.** The rotation method and the "lunes". When the crystal is rotated, reflections from the same plane in the reciprocal lattice form a lune, limited by two ellipses corresponding to the start and end positions (Dauter, 1999).

Several variables must be taken into account for a successful data collection, among them: the rotation range, the crystal-to-detector distance, the blind region, the total rotation range.

**Rotation range** The maximum permitted rotation range to avoid overlap of neighboring lunes can be estimated with this formula:  $\Delta\varphi = 180d/\pi a - \eta$ , where  $\eta$  depends on the mosaicity and beam divergence,  $d$  is the high-resolution limit and  $a$  is the length of the primitive unit cell dimension along the direction of the X-ray beam. The best orientation of the crystal is with the longest unit cell axis along the spindle axis of crystal rotation, in this way the longest edge will never lie parallel to the beam and reflections will not overlap.

**Crystal-to-detector distance** The distance should be adjusted to match the maximum resolution of the diffraction. If one unit cell dimension is so large that setting the distance to maximum resolution leads to overlap of reflections, it is better to sacrifice the resolution and set the distance so that reflection profiles separate.

**Blind region** The reciprocal lattice points lying close to the rotation axis will never cross the Ewald sphere, and will never diffract. This part of the reciprocal lattice, on both sides of the spindle axis, is called the “blind region”. It is narrow at low resolution and wide at high resolution. If the crystal is triclinic there is no way to avoid loss of completeness due to the blind region, but if the crystal has symmetry axes, it is possible at least to collect symmetry equivalent reflections to those in the blind region.

**Total rotation range** The total rotation range affects the completeness of the data set. Due to limited available time at the synchrotron, it is often impossible to collect  $180^\circ$  or  $360^\circ$ ; it is thus important to quickly index the data and, on the basis of the symmetry and orientation of the crystal, decide the total rotation range necessary for the data collection.

The goal of data collection is a set of consistently measured, indexed intensities. Several computer programs have been developed in order to process diffraction data.

Although it is not a proper statistical quantifier, the data quality is usually judged by the global  $R_{merge}$  factor, it is given by:  $R_{merge} = \sum[I^i - \langle I \rangle] / \sum \langle I \rangle$  where  $I^i$  is the individual intensity measurement and  $\langle I \rangle$  is the average intensity for this reflection (Dauter, 1999). This value is highly influenced by the redundancy of the data and is always higher for data in high symmetry space group than those in low symmetry. A good quality indicator is the ration of intensities to their uncertainties,  $I/\sigma$ , the accepted resolution limit is where the  $I/\sigma$  falls below about 2.0. Well scaled data should have their  $\chi^2$  as close as possible to 1, its value is given by:

$$\chi^2 = \sum \frac{(I - \langle I \rangle)^2}{(\sigma^2 \cdot N / (N - 1))} \quad (I.8)$$

## I.2.8 The Phase Problem

The main problem in crystallography after having recorded good data is the lack of phase information. The possible ways to obtain phases are:

- Molecular Replacement (MR): if a similar structure is known, it can be used to calculate initial phases.
- Multiple Isomorphous Replacement (MIR) or Single Isomorphous Replacement (SIR).
- Anomalous Diffraction (MAD, SAD).
- Direct Methods.

### Molecular Replacement

Molecular Replacement (MR) exploits the existence of a known model structure to solve the phase problem of an unknown structure. The first requirement is the similarity between the unknown and the known structure. Placement of the molecule in the target unit cell requires its proper orientation and precise position, which involves rotation and translation.

The principles of the MR method are based on the Patterson function of a crystal structure. The Patterson function represents a vector map in which interatomic distance vectors are represented by peaks of positive density. When the interatomic distances are between atoms inside the molecule they are called self-Patterson vectors while when between atoms belonging to different molecules in the unit cell they are called cross-Patterson vectors. The distinction between self and cross vectors is fundamental to MR since similar or identical molecules will give similar or equal Patterson map respectively, apart from a rotational and a translational term. This is because the self-vector give a representation of the molecule itself, while cross-vectors are useful in finding the position of the model in the unit cell because they are related to intermolecular atomic distances.

The principle of separating the Patterson vectors into these two groups can be used for orientation and translation determination.

**The rotation function** was first proposed by Rossman and Blow in 1962. If the model and the new molecule are similar, and if they are oriented in the same way in unit cells of the same dimensions and symmetry, they should give very similar Patterson maps. The rotation function  $R$ , used to find the right orientation, can be defined as:

$$R(\mathbf{C}) = \int_U P_{obs}(\mathbf{u})P_{calc}(\mathbf{Cu})d\mathbf{u} \quad (\text{I.9})$$

where  $P_{obs}$  and  $P_{calc}$  are the Patterson functions for the observed data and the model respectively,  $\mathbf{C}$  is the rotation matrix that describes the rotation performed on the search model. When  $R(\mathbf{C})$  is maximised a potential solution is found since the two Patterson functions superimpose their maxima. When the rotation function has found its maximum the model has to be positioned in real space in the unit cell.

**The translation function**  $T$ , as proposed by Crowther and Blow in 1967 is given by:

$$T(\mathbf{t}) = \int_U P_{obs}(\mathbf{u})P_{calc}(\mathbf{u}, \mathbf{t})d\mathbf{u} \quad (\text{I.10})$$

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

The program package AMoRe (Automated Molecular Replacement) (Navaza, 1994) is divided into four parts, the first to prepare the data for the calculating (SORTING and TABLING), the second to perform the rotation function (ROTING), giving the results sorted according to the  $R$ -factor.

$$R = \frac{\sum_{hkl} |F_{obs} - k|F_{calc}|}{\sum_{hkl} |F_{obs}|} \quad (\text{I.11})$$

The best solutions are then input to the translation function (TRAINING) which gives a list of peaks with both  $R$ -factor and correlation coefficient ( $CC$ ).

$$CC = \frac{\sum_{hkl} (|F_{obs}|^2 - \overline{|F_{obs}|^2})^2 \times \sum_{hkl} (|F_{calc}|^2 - \overline{|F_{calc}|^2})^2}{\sqrt{\left[ \sum_{hkl} (|F_{obs}|^2 - \overline{|F_{obs}|^2})^2 \times \sum_{hkl} (|F_{calc}|^2 - \overline{|F_{calc}|^2})^2 \right]}} \quad (\text{I.12})$$

When a solution is found, the last part of the package performs a rigid body fitting refinement (FITING). The rotation and translation are applied to the model which is then used for the phase calculation. It should be kept in mind that the initial phases are biased from the starting model, bias usually being removed during successive cycles of refinement and model building.

In the case of the DNA, the knowledge of the general helix type allows the use of an idealized helix as a search model in MR. The most powerful clue in solving the structures of B-DNA oligomers in various space groups has been the occurrence of the strong reflections at about 3.3 Å resolution. Such reflections are a consequence of the stacking of the base pairs along the B-DNA helix and their position reveals the direction of the helical axis (Dickerson *et al.*, 1998).

### Isomorphous replacement

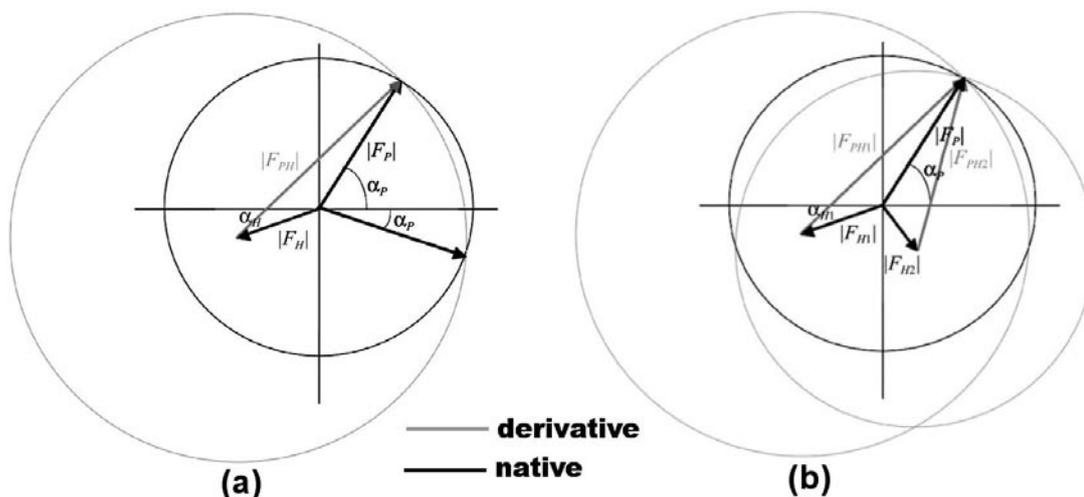
The Isomorphous Replacement uses the difference between the diffraction intensities of various crystals to solve the phase problem. This difference is generated by the presence of heavy atoms in the crystal. The introduction of heavy atoms is obtained either by soaking or by co-crystallization. Ideally the introduction of heavy atoms does not alter the crystal and the native and the derivatives should be isomorphous.

Based on the difference of magnitude between native and derivative structure factors, this method requires at least two data sets. The structure factors of the native,  $\mathbf{F}_P$ , and of the derivative,  $\mathbf{F}_{PH}$ , are related by the following equation:

$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H \quad (\text{I.13})$$

$\mathbf{F}_H$  is the structure factor of the heavy atoms; it tends to be small in comparison with  $\mathbf{F}_P$  and  $\mathbf{F}_{PH}$ . The first step is the determination of the heavy atom positions, calculating the isomorphous difference Patterson map. Since the number of heavy atoms is small, their position can be deduced from the Patterson map.

The use of Single Isomorphous Replacement, where only one heavy atom derivative is used, leads to a phase ambiguity, clearly illustrated by the Harker construction shown in Figure I.2.13 (a). In Multiple Isomorphous Replacement (MIR), where more than one heavy atom derivative is used, the phases can be determined unambiguously except for the experimental error. The graphical solution for MIR is shown in Figure I.2.13 (b).



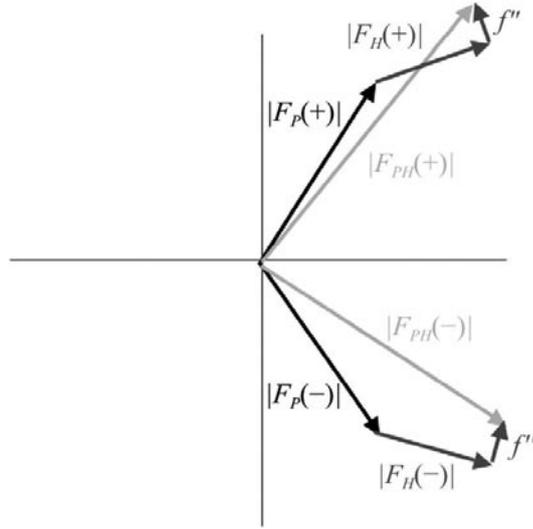
**Figure I.2.13.** Harker diagrams for SIR (a) and MIR (b).

### Anomalous scattering

The possibility of getting the phases using Multiwavelength Anomalous Dispersion (MAD) or Single Anomalous Dispersion (SAD) experiments rely on the presence of anomalously scattering atoms in the crystal. Their presence allows exploitation of the so called “break-down” of Friedel’s law (see Figure I.2.14):

$$|F_{hkl}| \neq |F_{-h-k-l}| \quad (\text{I.14})$$

In the absence of anomalous scatterers the structure factors vectors  $\mathbf{F}_{(hkl)}$  and  $\mathbf{F}_{(-hkl)}$  have the same magnitude.



**Figure I.2.14.** Breakdown of Friedel's law:  $|F_{hkl}| \neq |F_{-h-k-l}|$

The anomalous scattering occurs at the absorption edge when the X ray photon energy is sufficient to promote an electron from the inner shell. The scattering factor of an atom for which anomalous dispersion occurs is:

$$f_{anom.} = f^0(\theta) + f'(\lambda) + if''(\lambda) \quad (I.15)$$

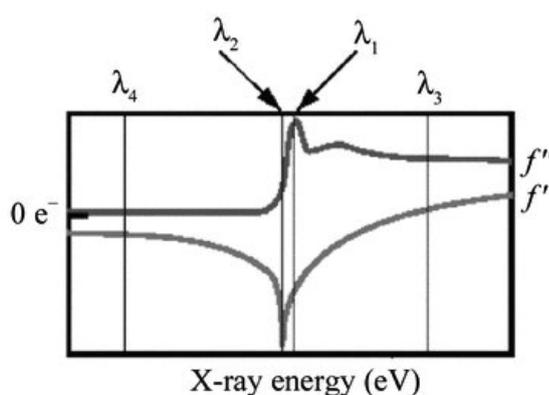
The normal scattering component  $f^0$  is independent of the wavelength (around the absorption edge) and decreases as the scattering angle  $2\theta$  increases. Both  $f'$  and  $f''$  depend on the incident wavelength, but are independent of the scattering angle. Because of this dependence on the wavelength  $\lambda$ , Hendrickson proposed to separate the contribution to the scattering in two components one due to normal and the other one due to anomalous scattering. The total scattering is given by the contribution of all the scattering components at the wavelength  $\lambda$ .

$$\begin{aligned} |\lambda F_T|^2 = & |F|^2 + a(\lambda)|F_A|^2 + b(\lambda)|F||F_A| \cos[\varphi - \varphi_A] \\ & + c(\lambda)|F||F_A| \sin[\varphi - \varphi_A] \end{aligned} \quad (I.16)$$

The total non-anomalous scattering factor is a sum of the structure factor of the protein (or DNA),  $F$ , and of the structure factors of the anomalous scatterers ignoring

their anomalous effect,  $F_A$ ;  $a(\lambda) = (f'^2 + f''^2)/f^{\circ 2}$ ;  $b(\lambda) = 2(f'/f^\circ)$  and  $c(\lambda) = 2(f''/f^\circ)$ .

The availability of tunable synchrotron beam lines allows the users to vary the two components  $f'$  and  $f''$  of the anomalous scattering factor by varying the wavelength of the incoming beam close to the absorption edge characteristic for each element. To optimize the dispersive signal  $f'$ , the energy (the wavelength,  $\lambda_2$ ) is moved to the edge inflection point, while for the anomalous difference  $f''$ ,  $\lambda_1$  is moved right above the edge. A third wavelength  $\lambda_3$ , so called remote, is usually collected at higher energy (shorter wavelength) also for collecting higher resolution data (Figure I.2.15).



**Figure I.2.15.** Typical absorption curve for anomalous scatterer, [www.bmsc.washington.edu](http://www.bmsc.washington.edu).

## Direct methods

Direct methods are routinely used to solve small molecules structures, but become less powerful as the number of atoms increases. This method is in fact limited to structures with no more than 600 atoms and with a resolution of at least 1.2 Å. Their application in solving macromolecular structures is not usually possible either for computer time requirement or for the low data-to-parameter ration due to the limited resolution.

A successful application of direct methods is represented by the program *Shake and Bake* (Miller *et al.*, 1993), where structure factors are calculated from a starting model based on random atoms. Their phases are refined and an electron density map is generated. The highest peaks are kept and new atoms are generated for the next cycles.

## I.2.9 Structure refinement

Once the initial phases or an initial model are available they have to be refined against the experimental data and the final structure has to be built. The aim of refinement is to find a model in which the position of the atoms gives calculated structure factors ( $F_{calc}$ ) as close as possible to the observed structure factors ( $F_{obs}$ ). Most macromolecular crystals do not diffract to atomic resolution thus the ratio of observation to parameters is low and allows the refinement of only the positional x, y, z atomic parameter and a single isotropic temperature factor. Since the X-ray data from crystals are often limited, stereochemical information such as bond length, bond angles, planar groups and contact distances (Olson *et al.*, 2001) can be used as additional sources of information (Jensen, 1997) in the form of constraints or restraints. This allows the increase of data-to-parameter ratio. A *constraint* is a fixed value for a certain parameter; a *restraint* is a secondary condition imposed upon the parameters, such as the condition that all bond lengths and bond angles are within a specific range of values (Rhodes, 1993).

The progress of refinement is monitored by comparing the measured structure-factor amplitudes  $|F_{obs}|$  with amplitudes  $|F_{calc}|$  from the current model. The most widely measure of convergence uses  $R$  and  $R_{free}$  factors, usually defined as a percentage:

$$R = \frac{\sum_{hkl} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl} |F_{obs}|} \quad (\text{I.17})$$

The use of  $R_{free}$  was proposed by Brünger (1992) as a tool to check whether the refinement is correct or if the improvement of the  $R$  factor is only an artifact due to overfitting of the data. The idea is to exclude a fraction of the data (typically from 5% to 10%) from the refinement process as a cross validation tool. At any stage of the refinement,  $R_{free}$  measures how well the current atomic model predicts a subset of the

measured intensities that were not included in the refinement, whereas  $R$  measures how well the current model predicts the entire data set that produced the model. A strong deviation between  $R$  and  $R_{free}$  indicates overfitting of the model.

In crystallographic refinement two functions are commonly used. They are the least-square residual and the maximum likelihood.

The least-square refinement has been used for many years (Konnert, 1976; Konnert, *et al.*, 1980); the function to be minimized is:

$$Q = \sum_{hkl} w(hkl) (|F_{obs}(hkl)| - |F_{calc}(hkl)|)^2 \quad (\text{I.18})$$

The summation is over all crystallographically independent reflections and  $w$  is the weight given to an observation.

The main limitation is the possibility of getting trapped in local minima, when the model is not very good or not complete. To overcome this problem different approaches have been chosen. The program CNS (Brünger *et al.*, 1998) includes a molecular dynamics algorithm which exploits a simulated annealing technique. In the simulation the temperature is increased and the atoms are allowed to move freely from their original position, the temperature is then slowly cooled down allowing the structure to rearrange and eventually find a global minimum (Brünger *et al.*, 1998).

In the maximum-likelihood method, implemented in the program REFMAC (Murshudov *et al.*, 1997), given the model, the probability function that a set of data would be observed is calculated:

$$L = \prod_i P(\bar{F}_{i,obs}; \bar{F}_{i,calc}) \quad (\text{I.19})$$

where  $P(\bar{F}_{i,obs}; \bar{F}_{i,calc})$  is the probability distribution of the structure factor  $\bar{F}_{i,obs}$  given the model structure factor  $\bar{F}_{i,calc}$  (Pannu, *et al.* 1996).

**TLS refinement** (Howlin *et al.*, 1993; Schomaker *et al.*, 1993; Murshudov *et al.*, 1999). Normally four parameters per atom are refined for macromolecular structures:

the three co-ordinates and a temperature factor. The temperature factor is a measure of the mean displacement of an atom. Since data sets from proteins or DNA are generally not highly overdetermined, the temperature factor is considered as isotropic; only for very high resolution data it can be refined anisotropically. In 1968, Schomaker & Trueblood described a parameterization that allows the description of anisotropic motion with many fewer parameters than an independent anisotropic  $B$  factor for each atom. This parameterization is called TLS (translation, libration and screw). In this system the motion of a group of atoms is described by three matrices. The explicit assumption of TLS- $B$ -factors is that the group of atoms moves as a rigid group. In the TLS formalism, 20 parameters are used to describe the motion of the entire group of atoms. Since the anisotropic  $B$ -factor of one atom requires six parameters, any TLS group composed of more than three atoms results in a decrease in the total number of parameters. The TLS refinement has been implemented in the program REFMAC 5 (Murshudov *et al.*, 1997).

### **I.2.10 Validation and deposition**

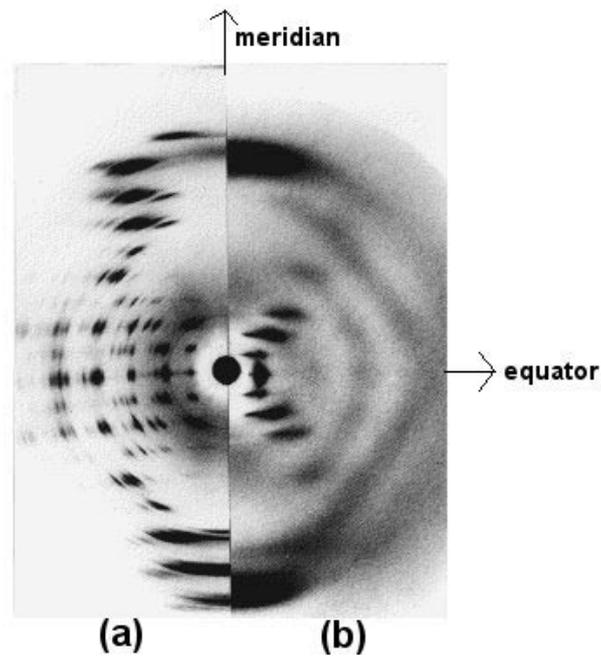
Structure deposition at the Nucleic Acid Data Base (NDB) or at the Protein Data Bank (PDB) is the final and fundamental step once a structure is refined. The importance of structure validation before deposition is of course fundamental. In this work the program 3DNA (Lu & Olson, 2003) has been used. This program implements the parameter set derived from the *Tsukuba Workshop on Nucleic Acid Structure and Interactions* of 1999 and recommended by the NDB (Berman *et al.*, 1992). It must be taken into account that the geometrical parameters calculated by the program 3DNA are optimized for the structures of complementary Watson-Crick basepairs, the presence of Hoogsteen hydrogen bonds, makes some of the parameters calculated by the 3DNA unusable at least for comparison with standard values.

## I.2.11 Introduction to fiber diffraction

Although in this work no fiber diffraction experiment has been performed, many of the crystals obtained were characterized by a fibrous nature. It appears therefore relevant to briefly introduce the concepts of fiber diffraction.

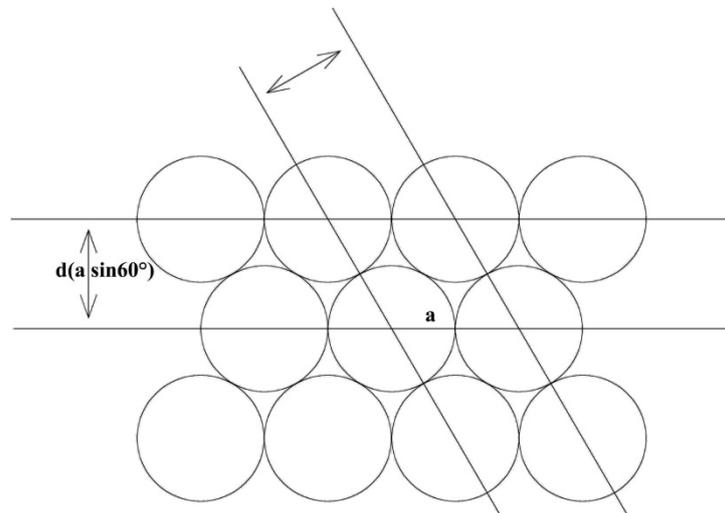
Many biological macromolecules are of a fibrous nature. Sometimes the orientation is intrinsic, but often the molecules can be oriented into fibers when isolated from the cells. The oriented fiber is placed in a collimated X-ray beam at right angles to the beam and the fiber diffraction pattern is recorded on a film placed a few centimeters away from the fiber. The direction parallel to the fiber axis and through the center of a fiber diffraction pattern is referred to as the meridian, and the direction perpendicular to this is called the equator (see Figure I.2.16).

Fibers are usually composed of long, chain-like molecules, packed together with their axes parallel, or nearly parallel, to the fiber axis. The degree of order within fibers may vary considerably. A famous example is given by the fiber diffraction obtained from A and B DNA forms, shown in Figure I.2.16. In the cases like the A form of DNA (Figure I.2.16), the molecules are regularly arranged so that they form crystalline regions, but the different crystalline regions within a fiber are randomly oriented about the fiber axis. The diffraction patterns from such fibers are similar to single crystal rotation photographs, with all Bragg reflections registered at one time and with the appearance of arcs due to the disorder. In other cases, the degree of order is much lower. If molecules are randomly displaced relative to each other in the direction of fiber axis, discrete spots are only observed along the equator, and the higher layer lines have a continuous distribution of intensity along them, this is the case of B-DNA, shown in Figure I.2.16 (b).



**Figure I.2.16.** Fiber diffractions of A DNA (a) and B DNA (b). The meridian and the equator are shown.

Discrete spots along the meridian indicate periodicity along the fiber axis. The diffraction along the equator gives information about the structure in projection down the fiber axis, thus, discrete spots along the equator are related with the lateral distance between molecules, an example of hexagonal arrangement is given in Figure I.2.17.

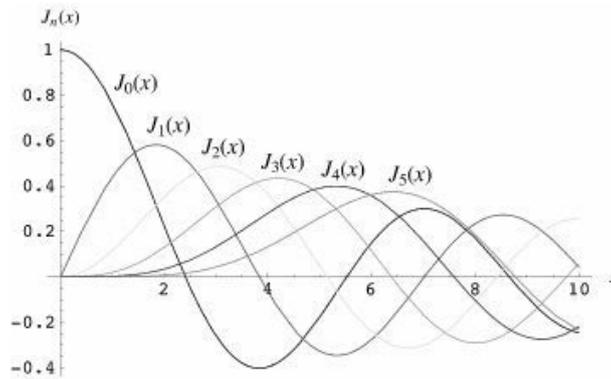


**Figure I.2.17.** Projection down the fiber axis of a hexagonal arrangement of helical molecules, from the fiber diffraction pattern, the distance between helices can be measured.

## Diffraction by helical molecules

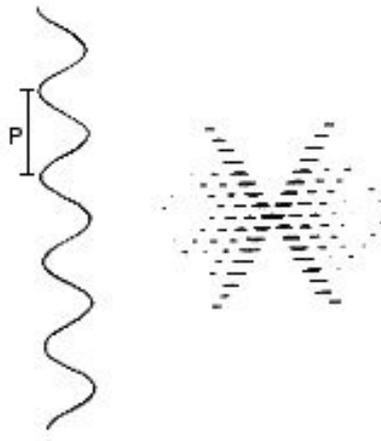
The theory of diffraction by helical molecules was first developed in 1952 by Cochran, Crick and Vand and by Stokes (unpublished).

Crick showed that the diffraction from a helix occurs along a series of equidistant layer lines. The intensity along the layer lines is continuous and can be calculated via a “Fourier-Bessel Transform”. Bessel functions enter the equation because a cylindrical coordinate system is used. The variation of  $J_n(x)$  with  $n$  and  $x$  is shown in Figure I.2.18. Only the zero order Bessel function,  $J_0(x)$ , has a non-zero value at  $x = 0$ ; the value of  $J_n(x)$  for the first maximum decreases as  $n$  increases.



**Figure I.2.18.** Bessel functions.

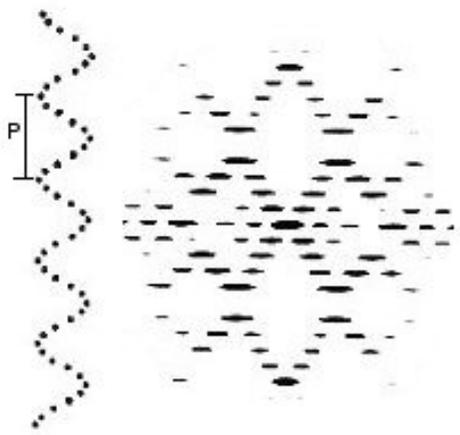
Crick showed that for a continuous helix the order of Bessel function  $n$ , occurring on a certain layer line, is the same as the layer line number  $l$ . A continuous helix and its diffraction pattern are shown in Figure I.2.19. Because the order of Bessel function increases with layer line number, the position of the first strong peak moves further away from the meridional direction, generating the characteristic “helix cross”.



**Figure I.2.19.** A continuous helix and its diffraction pattern. The pitch  $P$  of the helix is shown.

The position of the first strong peak is also inversely proportional to the radius of the helix. The spacing between the layer-lines is inversely proportional to the pitch ( $P$ ) of the helix.

A discontinuous helix can be considered as a set of scattering points, equally spaced along a helix, with  $p$  the vertical distance between such points and  $\omega$  the turn angle between them. The pitch  $P$  can be directly derived from the first two values. In projection onto the helical axis (see Figure I.2.20), the structure has a regular repeat period  $p$ . This gives rise to meridional reflections on the diffraction patterns. The cross-like pattern seen in the diffraction of continuous helices is still visible and, in addition, it is repeated at each meridional reflection. This gives rise to characteristic diamond-shaped regions above and below the centre of the pattern, as shown in Figure I.2.20. For a simple helix which repeats in one turn the spacing between layer lines is given by  $1/P$ . The distance of the first meridional layer line along the meridian is given by  $1/p$ .



**Figure I.2.20.** A discontinuous helix of ten residues in one turn and its diffraction pattern.

The B-DNA is a simple helix which repeats in one turn. With its 10 base pairs per turn, the average angle  $\omega$  is of about  $36^\circ$ . The spacing between the bases corresponds to  $3.4 \text{ \AA}$  ( $p = 3.4 \text{ \AA}$ ) and the pitch  $P$  is of about  $34 \text{ \AA}$ . In frozen crystals like those studied here the value of  $p$  is usually smaller ( $3.2 \text{ \AA}$ - $3.35 \text{ \AA}$ )

## II. CRYSTALLOGRAPHIC STUDY ON OLIGONUCLEOTIDE COILED-COILS

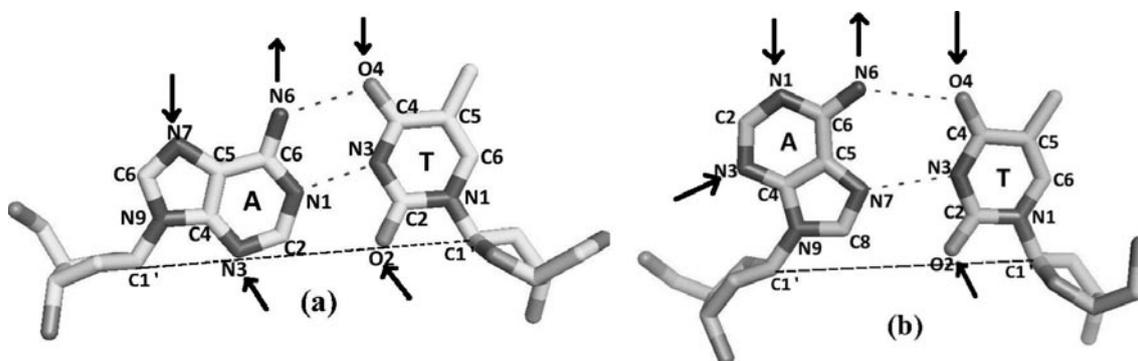
### II.1 INTRODUCTION: AT-rich DNA sequences

Analysis of the human genome sequence has confirmed the presence of extensive noncoding regions (Lander *et al.*, 2001). Such regions are also present in practically all eukaryotic genomes, but their biological role is unclear. Interestingly, in most cases, they are rich in AT base pairs. Centromeres and pericentromeric regions (Choo *et al.*, 1997), introns (Lander *et al.*, 2001), scaffold-associated regions or matrix attachment regions (Liebich *et al.*, 2002), gene desert (Lander *et al.*, 2001; Nobrega *et al.*, 2003), complex genes (Nobrega *et al.*, 2003), and some satellites (Sainz *et al.*, 1989) are all very rich in AT base pairs.

In contrast with the abundance of AT base pairs in extensive regions of the genome, there is no structure available for any protein interacting with a DNA fragment that only contains AT base pairs. There are also few structural studies of oligonucleotides with such sequences. An overview of the structure of thirty-three all-AT duplexes has been recently reported (Campos *et al.*, 2006). Before this work, only a few alternating structures had been determined: d(ATAT) (Viswamitra *et al.*, 1982), d(ATATAT) (Abrescia *et al.*, 2002) and d(AT)<sub>6</sub> (Campos *et al.*, 2005). While d(ATAT) shows Watson-Crick base pairs, although in a nonstandard conformation, d(ATATAT) crystallizes as a duplex structure with Hoogsteen base pairs. On the other hand, d(AT)<sub>6</sub> forms a coiled-coil probably in Hoogsteen conformation.

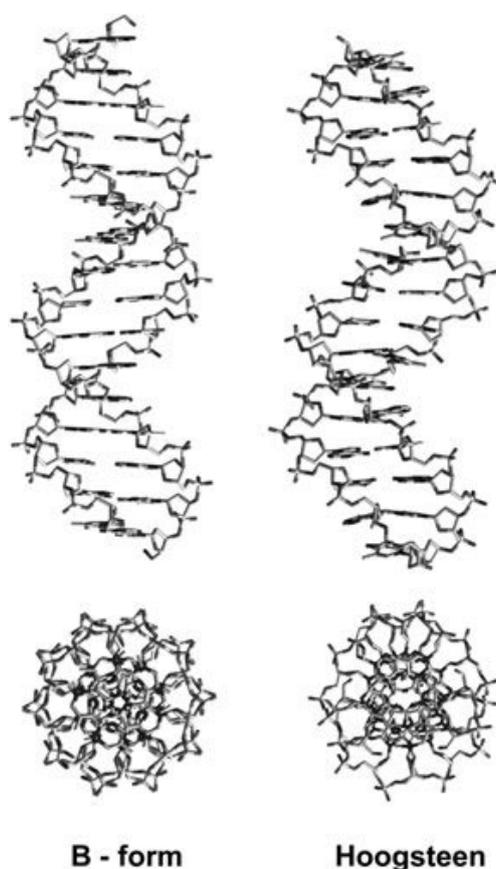
Hoogsteen base pairs have been known for more than 40 years. The Hoogsteen base pairing scheme was observed for the first time in the crystal structure of a complex of adenine and thymine bases methylated in the position linking to the sugars in the nucleosides (Hoogsteen, 1959). As shown in Figure II.1.1, 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 N6, or O6 for C·G base pairs. Protonation of cytosine at N3 is a prerequisite for C·G Hoogsteen basepairs. The characteristic change to the *syn* conformation of the purine base adenine (whereas thymines are in the normal *anti* conformation) is also found in Z-form DNA, where the guanine undergoes a similar rotation.

Hoogsteen base pairs were postulated for U(A·U) triple helices (Felsenfeld *et al.*, 1957). Such interactions have also been found in chemically modified nucleic acids (Hakoshima *et al.*, 1981; Isaksson *et al.*, 2001). Isolated base pairs have been reported in some protein/DNA complexes (Patikoglou *et al.*, 1999; Nair *et al.*, 2004) and occasionally in RNA (Leontis & Westhof, 1998). A·T Hoogsteen and standard Watson-Crick hydrogen bondings are shown in Figure II.1.1.



**Figure II.1.1.** Comparison of A·T Watson-Crick (a) and Hoogsteen (b) base pairs in duplex conformation. The minor groove is facing downwards; hydrogen bond acceptors and donors are indicated by arrows; hydrogen bonds and C1'-C1' distances are shown in dashed lines. The minor groove is narrower in the Hoogsteen case and has lost a hydrogen bond acceptor atom (adenine N3). The major groove has a similar appearance, but in the Hoogsteen conformation an additional external N3 atom is present. The conformation of the glycosidic angle is *syn* in the Hoogsteen adenine base and *anti* in all other cases.

Although the overall appearance of the Hoogsteen duplex is very similar to the standard B-form DNA, as shown in Figure II.1.2, there are important differences. The minor groove of Hoogsteen DNA is narrower due to the shorter C1'-C1' distances (Figure II.1.1). The accessibility of the grooves toward interaction with solvent and proteins is different in the two forms. The N3 atom of adenine, which lies in the minor groove in B-form DNA, is now moved to the major groove. As a result, the minor groove becomes less electronegative with only one hydrogen bond acceptor represented by the O2 of thymine. This characteristic of the minor groove, together with its narrowness, makes it an appropriate target for interactions with hydrophobic groups. In the Hoogsteen duplex, the helical axis is found at the edge of the base pairs, approximately at the midpoint of the hydrogen bond between adenine N6 and thymine O4. The major groove is in the center of the helix, and the phosphates are externally located. This situation is also found in A-form DNA.



**Figure II.1.2.** Comparison of ideal B DNA with Hoogsteen DNA (Abrescia *et al.*, 2002).

An alternating sequence such as d(AT)<sub>3</sub> has two types of base steps: AT and TA. In B-form DNA the latter base step differs considerably in the  $\omega$  twist parameter (Yuan *et al.*, 1992), whereas in the Hoogsteen case the twist values are similar in both cases (Abrescia *et al.*, 2002).

Thus far there is no evidence that long stretches of Hoogsteen DNA double helices may occur in vivo. However, there are a number of functions which are poorly understood in which Hoogsteen DNA may play a role.

A unique feature of d(AT)<sub>3</sub> crystal is the presence of extrahelical terminal adenines and thymines located in the minor groove (Abrescia *et al.*, 2002). This behavior may favor a more hydrophobic environment and thus stabilize Hoogsteen base pairs.

The behavior of the oligonucleotide d(ATATAT) was also studied by NMR (Abrescia *et al.*, 2004), in order to determine its structure in solution, and its melting temperature ( $T_m$ ) has been determined by UV spectroscopy (De Luchi *et al.*, 2003). The NMR data showed that standard Watson-Crick duplexes are formed in solution. The UV spectra and melting curve did not show either any feature that might indicate the presence of Hoogsteen DNA in solution. The published results of the study on the melting stability of d(AT)<sub>3</sub> and other AT-rich DNA sequences are reported in Appendix III.1 (De Luchi *et al.*, 2003). The melting behavior of thirteen AT-rich oligonucleotides has been studied and a simple equation useful to predict the  $T_m$  of oligonucleotides of any composition and size has been derived. The melting temperature of 100% adenines and thymines sequences can be predicted as follows:

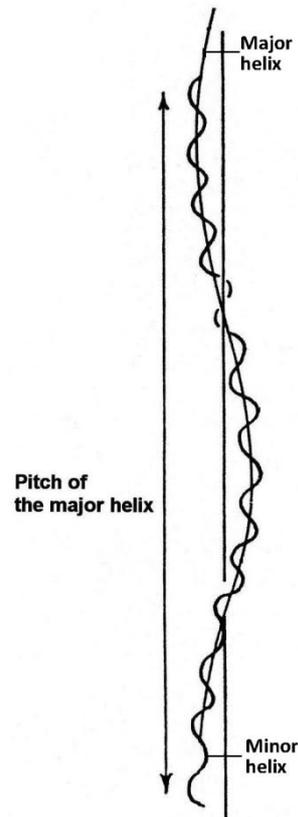
$$T_m^{-1} = 2.806 \cdot 10^{-3} + 5.247 \cdot 10^{-3} n^{-1}$$

(II.1)

where  $T_m$  is the melting temperature and  $n$  is the number of base pairs in the oligonucleotide.

In the case of d(AT)<sub>3</sub>, the formation of A·T Hoogsteen base pairs appears to be cooperative; in fact such base pairing has never been detected in mixed sequence oligonucleotide crystals. In order to confirm this hypothesis, longer (AT)<sub>n</sub> sequences have been crystallized. While d(AT)<sub>4</sub> did not give crystals, the sequences d(AT)<sub>5</sub> and d(AT)<sub>6</sub> generated crystals suitable for X-ray diffraction experiments.

Instead of standard crystalline structure, the  $d(AT)_5$  and  $d(AT)_6$  sequences formed very stable coiled-coils. Coiled-coils are generated when the straight axis of a simple helix follows itself a helical path. The helical path which follows the axis is called the major helix and the helix which is formed around this axis is called the minor helix (Figure II.1.3). The X-ray diffraction of a coiled-coil is characterized by the splitting of the layer lines of the simple helix into a series of closely spaced layer lines (Fraser *et al.*, 1964).



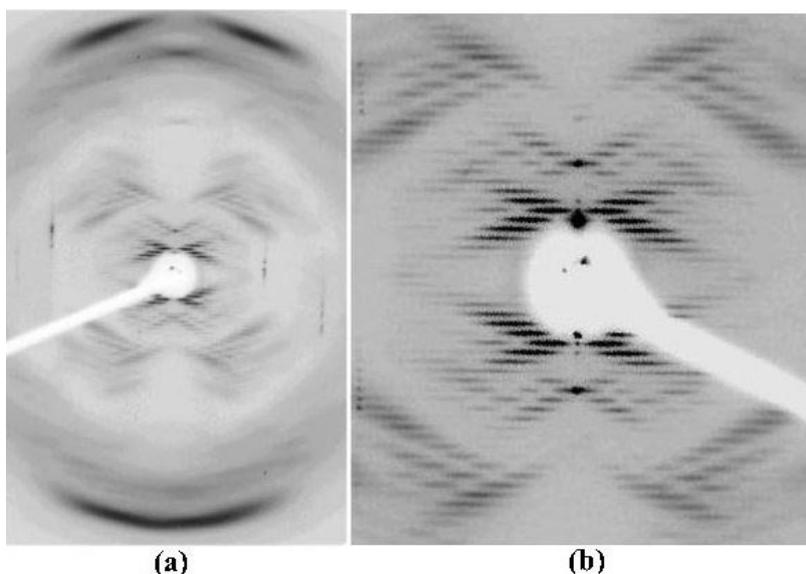
**Figure II.1.3.** Schematic representation of a coiled coil: the minor helix follows a helical path and generates a superhelix (Wilson, 1996).

The characteristics of  $d(AT)_6$  and  $d(AT)_5$  structures are summarized below.

**$d(ATATATATATAT)$ , a dodecamer (Campos *et al.*, 2005)**

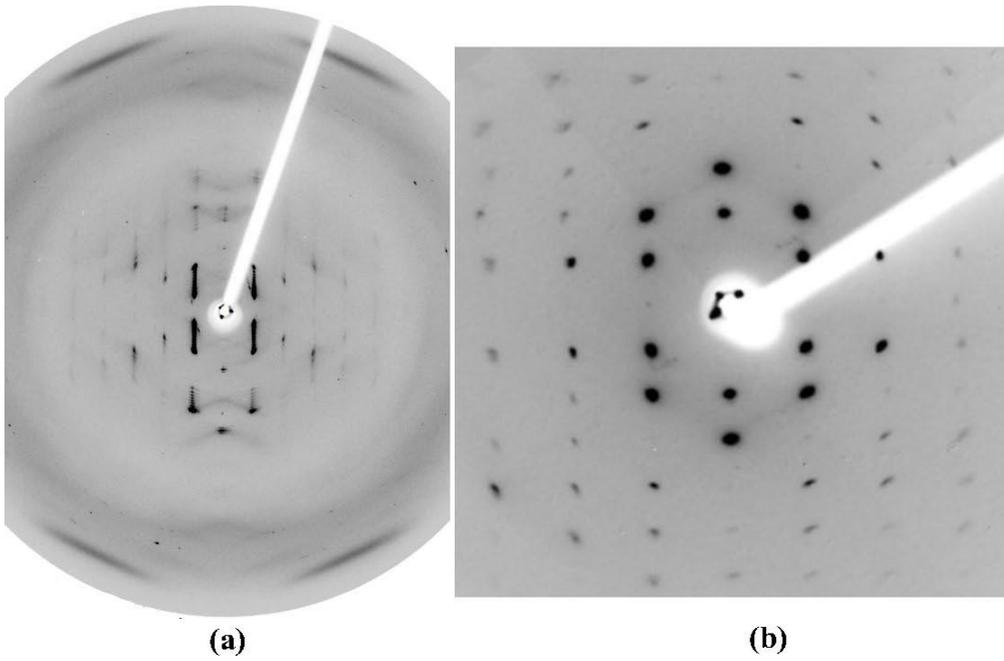
Typical diffraction patterns of  $d(AT)_6$  crystals are shown in Figure II.1.4. The diffraction can be interpreted as due to a pseudocontinuous DNA coiled-coil formed by consecutive duplexes. The diffraction patterns show the coexistence of Bragg reflections and layer lines with continuous diffraction. Crystalline regions give rise to Bragg spots; partially disordered regions in which the molecules are randomly displaced by vertical and rotational movement give rise to the continuous diffraction.

Diffraction data could only be obtained with a limited resolution of 5 Å; this is probably due to a combination of factors: high solvent content, screw disorder and multiple possible alignments (Figure II.1.5). The isomorphous structure of  $dCG(AT)_5$  determined at 3.1 Å resolution will be described in greater details in paragraph II.5.1.



**Figure II.1.4.**  $(AT)_6$  meridional diffraction patterns ( $3^\circ$ ) showing the coexistence of Bragg spots and layer lines. (a) The long  $c^*$  axis of the unit cell is approximately vertical. Meridional Bragg reflections (006 and 0012) are clearly apparent. The prominent stacking reflections at 3.25 Å deviate an angle  $\beta$  of about  $19^\circ$  from a meridional orientation. (b) Enlarged view of a meridional oscillation pattern ( $5^\circ$ ) (Campos *et al.*, 2005).





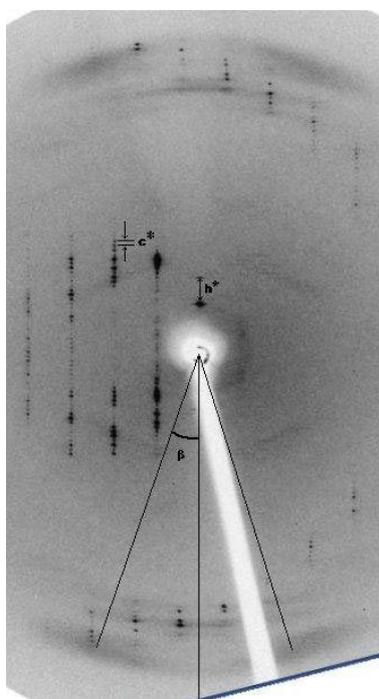
**Figure II.1.6.** (a) Meridional oscillation pattern of a crystal obtained from the sequence  $d(AT)_5$ . The individual duplexes are inclined an angle  $\beta$  of about  $20^\circ$  with respect to the major coil axis. (b) Enlarged view of the equatorial oscillation pattern.

## II.2 Determination of the parameters of a coiled-coil

From a coiled coil diffraction pattern which the  $l$  axis is vertical, the following parameters can be measured:

- The  $\beta$  inclination of the duplexes with respect to the major coil axis;
- The height  $h$  occupied by one duplex in the unit cell, and from the latter value, taking into account the  $\beta$  inclination of the duplex, the length  $l$  of the duplex can be calculated ( $l = h/\cos \beta$ ).
- The number  $N$  of duplexes in one unit cell, which corresponds to a whole number of turns of the coiled coil, usually one. It is calculated as the ratio  $c/h$ .
- The approximate average rise of the oligonucleotides in an individual duplex can be determined from the spacing of the broad stacking reflection.

The interpretation of a diffraction pattern of a crystal of  $\text{CG(AT)}_5$  is shown as an example in Figure II.2.1.



**Figure II.2.1.** Interpretation of a diffraction pattern of a coiled-coil. The inclination  $\beta$  of the duplexes and the height  $h$  occupied by the repeating unit along the  $c$  axis can be measured. The unit cell parameters ( $c^*$  is shown) can be determined if the symmetry of the crystal is known. The distance between spots on the  $00l$  axis corresponds to the height  $h$ .

## II.3 Aim of the project

Previous works show that the DNA sequences  $d(AT)_6$  (Campos *et al.*, 2005) and  $d(AT)_5$  (unpublished results) generate coiled-coils with very different geometrical characteristics. The aim of this project is to study the properties of the coiled-coils.

1. With this purpose, fourteen oligonucleotides with sticky ended sequences have been crystallized. The presence of the sticky end determines the coiled-coil properties. The sequences that have been studied are  $(CG)_n(AT)_m$  and  $(AT)_m(CG)_n$ , plus some other very similar to those. The majority of them have  $n = 1$  and  $m > 1$ , so that the sticky end is usually represented by CG.
2. The geometrical characteristics of the coiled-coils have been studied.
3. It has been tried to determine whether the Hoogsteen base pairing influences or not the geometry of the super-coils.
4. As a complementary study, the melting temperatures of AT-rich oligonucleotides have been measured; the results are reported in Appendix III.1.

The sequences studied in this thesis are summarized in Table II.1.

**Table II.1** List of the oligonucleotide sequences crystallized in this work.

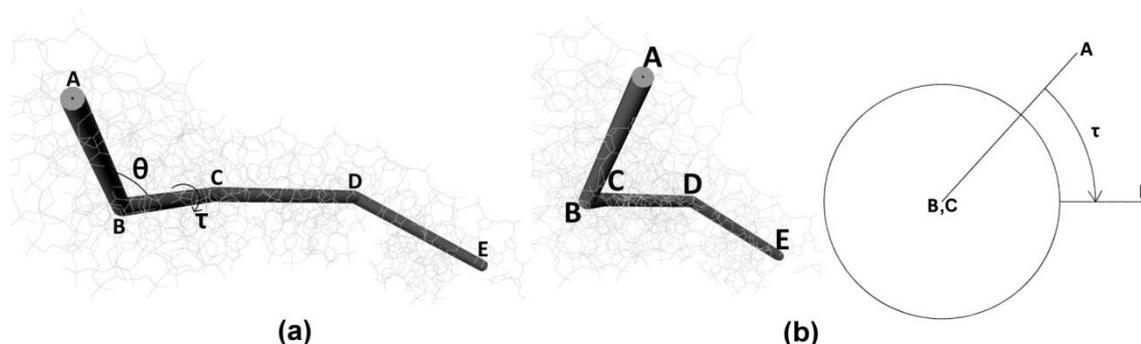
Dodecamers	Decamers	Octamers
$d(CGATATATATAT): CG(AT)_5$	$d(CGATATATAT): CG(AT)_4$	$d(CGTATATA): CG(TA)_3$
$d(CGCGATATATAT): (CG)_2(AT)_4$	$d(ATATATATCG): (AT)_4CG$	$d(CGATATAT): CG(AT)_3$
$d(ATATATATATCG): (AT)_5CG$	$d(ATATATATATT): (AT)_5T$	$d(ATATATCG): (AT)_3CG$
$d(ATATATATATGC): (AT)_5GC$		
$d(GCATATATATAT): GC(AT)_5$		
$d(CGATATGCATAT): CG(AT)_2GC(AT)_2$		
$d(CGCGCGATATAT): (CG)_3(AT)_3^*$		
Hexamers	Tetramers	
$d(CGATAT): CG(AT)_2$	$d(CGAT)^*$	

\*No crystals suitable for X-ray experiments have been obtained from the sequences  $(CG)_3(AT)_3$  and  $d(CGAT)$ .

## II.4 Geometry of the coiled-coil

As said in paragraph II.1, when the straight axis of a simple helix follows itself a helical path, then the structure is called a coiled-coil (Figure II.1.3). The supercoil conformation of DNA is well known from studies of circular molecules. The DNA in the nucleosome also forms a supercoiled structure. DNA oligonucleotides may also associate by following a helical path in some protein/DNA complexes (Bunting *et al.*, 2003). However, prior to the study published in 2005 by Campos *et al.*, no studies were available on isolated, regular, continuous coiled-coil DNA molecules. Ropes formed by supercoiled  $\alpha$ -helices, though, are a classical biophysical model. In particular, the model has been used in the study of keratin (Fraser *et al.*, 1964).

In this work, studies on fourteen sticky ended oligonucleotides are presented. They form pseudo-continuous DNA helices. In about half of the cases, the individual duplexes axes are inclined a few degrees with respect to the major coil axis and thus generate a coiled-coil structure. The geometrical characteristics of the coiled-coils (also called superhelices) have been studied. In fact the coiled-coils generated by oligonucleotides with sticky ends may be considered as kinked coils, whose geometries are defined by the two angles  $\theta$  and  $\tau$ . The kink angle  $\theta$  is formed by two consecutive duplexes, whereas  $\tau$  is the torsion angle which relates three consecutive duplexes (see Figure II.4.1). This simplification assumes that the duplexes are perfectly straight. Small intrinsic bends in the duplexes will in fact be incorporated in the angle  $\theta$ . If the structure is approximately continuous,  $\tau$  is directly related with  $\omega$ , the twist value of the individual base pairs.



**Figure II.4.1.** (a) Lateral view of a DNA coiled-coil, the kink angle  $\theta$  and the torsion angle  $\tau$  are shown. The torsion angle  $\tau$  represents the rotation of  $\overline{AB}$  with respect to  $\overline{CD}$  about the connective fragment  $\overline{BC}$ . (b) Projections of the coil along  $\overline{BC}$ . The  $\tau$  angle is shown.

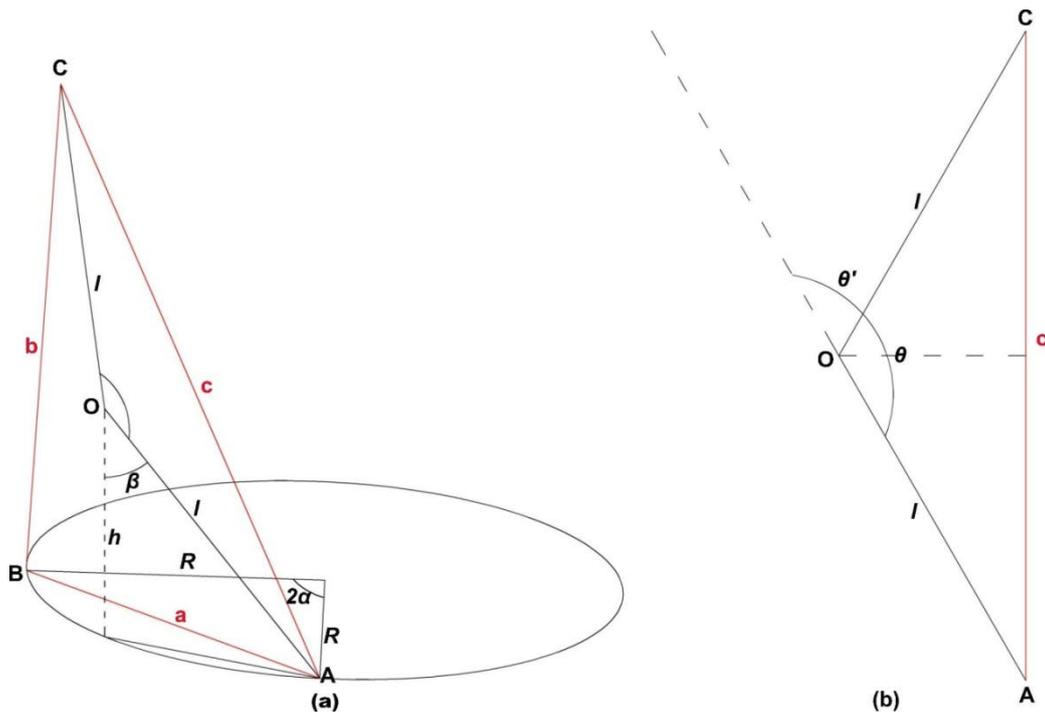
Although the geometry of a coiled-coil is determined by the kink angle  $\theta$  and the torsion angle  $\tau$ , from the experimental diffraction patterns the following parameters can be measured (see paragraph : the inclination  $\beta$  of the oligonucleotides, the length  $h$  of the repeating unit and the number  $N$  of duplexes per turn.

A schematic view of two consecutive duplexes, part of a coiled coil, is shown in Figure II.4.2:

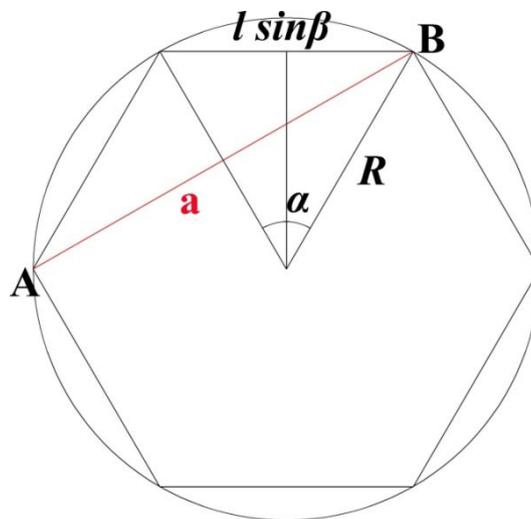
- $h$  is the length occupied by the repeating unit along the major axis of the coiled-coil;
- $\beta$  represents the inclination of the duplex axis with respect to the major coil axis;
- $l$  is the length of one duplex;
- $\alpha$  is the turn angle, related to the number of duplex per turn,  $N$ ;
- $R$  is the radius of the cylinder described by the coiled-coil;
- $a$  is the projection of two consecutive fragments along the major helix axis;
- $b$  is projection of two consecutive fragments onto the coiled-coil axis;
- $\theta$  is the kink angle between two consecutive fragments;
- $\theta'$  is  $\theta$  supplementary angle.

A coiled-coil describes a cylinder with a radius  $R$ . The number  $N$  of duplexes per turn determines the turn angle  $\alpha$  (with  $\alpha = 360^\circ/N$ ). In the simple case of a coiled-coil with six duplexes per turn,  $\alpha$  is equal to  $60^\circ$ . The projection of such a helix along the major coil axis is shown in Figure II.4.3. The turn angle  $\alpha$  is always equal to or smaller than the torsion angle  $\tau$ .

These geometrical considerations are equally valid for more complex situations, when the duplexes form a coil with a non-integer value of  $N$ , for example twenty four duplexes in five turns. In the cases studied here, such a situation does not appear to be present.



**Figure II.4.2.** (a) Perspective view of two consecutive duplexes, part of a coiled coil.  $l$  is the length of one duplex;  $\beta$  is the inclination of one duplex axis with respect to the coiled-coil axis;  $R$  is the radius of the cylinder described by the coiled-coil;  $\alpha$  is the turn angle; (b) the kink angle  $\theta$  between two consecutive oligonucleotides is shown;  $\theta'$  is its supplementary angle.



**Figure II.4.3.** Projection along the major axis of a coiled-coil with six oligonucleotides per turn. The turn angle  $\alpha$  of  $60^\circ$  is shown.

## II.4.1 Calculation of the geometrical parameters of the coiled-coil

From the experimental diffraction patterns, the inclination  $\beta$  of each duplex, as well as the number  $N$  of duplexes per turn of the coiled-coil can be obtained; their values are given in the tables in which the experimental results are reported. The handedness of the coil cannot be calculated directly from these values. From  $\beta$  and  $N$ , the values of  $R$ ,  $\theta$  and  $\tau$  can be calculated as follows.

- **Radius  $R$ .** The turn angle  $\alpha$  is given by the following equation (see Figure II.4.3):

$$\alpha = \frac{360}{N} \quad (\text{II.2})$$

The relation between  $\alpha$  and the  $\beta$  inclination is given by (see Figure II.4.2 a):

$$l \sin \beta / 2 = R \sin(\alpha/2) \quad (\text{II.3})$$

Rearranging the two equations, the radius  $R$  can be calculated as follows:

$$R = l \frac{\sin \beta}{2 \sin \frac{180}{N}} = l \frac{\sin \beta}{2 \sin \frac{\alpha}{2}} \quad (\text{II.4})$$

- **Kink angle  $\theta$**  The sides of the triangle  $A\hat{B}C$  shown in Figure II.4.2 (a) are given by:

$$\mathbf{a} = 2R \sin \alpha = l \frac{\sin \beta}{\sin \frac{\alpha}{2}} \left( 2 \sin \frac{\alpha}{2} \cos \frac{\alpha}{2} \right) = 2l \sin \beta \cos \frac{\alpha}{2} \quad (\text{II.5})$$

$$\mathbf{b} = 2l \cos \beta \quad (\text{II.6})$$

Thus, from the Pythagorean Theorem,  $\mathbf{c}$  is given by:

$$\mathbf{c} = (\mathbf{a}^2 + \mathbf{b}^2)^{1/2} = 2l \left( \sin^2 \beta \cos^2 \frac{\alpha}{2} + \cos^2 \beta \right)^{1/2} \quad (\text{II.7})$$

And, from the relation between  $\mathbf{c}$  and  $\theta$  shown in Figure II.4.2 (b),  $\mathbf{c}$  is also given by:

$$\mathbf{c} = 2l \sin \frac{\theta}{2} \quad (\text{II.8})$$

Thus the kink angle  $\theta$  can be written as a function of the turn angle  $\alpha$  and the  $\beta$  inclination angle:

$$\sin \frac{\theta}{2} = \left( \sin^2 \beta \cos^2 \frac{\alpha}{2} + \cos^2 \beta \right)^{1/2} \quad (\text{II.9})$$

$$\sin^2 \frac{\theta}{2} = 1 - \sin^2 \beta + \sin^2 \beta \cos^2 \frac{\alpha}{2} = 1 + \sin^2 \beta \left( \cos^2 \frac{\alpha}{2} - 1 \right) \quad (\text{II.10})$$

$$\sin^2 \frac{\theta}{2} = 1 - \sin^2 \beta \sin^2 \frac{\alpha}{2} \quad (\text{II.11})$$

Or, rearranging the equation:

$$\cos \frac{\theta}{2} = \sin \beta \sin \frac{\alpha}{2} \quad (\text{II.12})$$

- **Torsion angle  $\tau$**  The value of  $\tau$  can be calculated from  $\alpha$  and  $\beta$  by the following formula (Van Meerssche & Feneau-Dupont, 1984):

$$\cos \tau = \frac{2 \cos^2 \frac{\alpha}{2}}{1 - \sin^2 \frac{\alpha}{2} \sin^2 \beta} - 1 \quad (\text{II.13})$$

The value of  $\tau$  will be either positive or negative, depending on the handedness of the supercoil.

It is interesting to consider two extreme cases. When  $\beta = 90^\circ$  we are faced with a flat polygon, with  $\tau = 0^\circ$  and  $\theta = 180^\circ - \alpha$ . At the other extreme, when  $\beta = 0^\circ$  a straight coiled-coil is generated, that has been named a HASO structure, where HASO stays for Helical Arrangements of Stacked Oligonucleotides (Campos *et al.*, 2006). In the latter case  $\theta = 180^\circ$  and  $\tau$  and  $\alpha$  have the same value.

### Examples

The calculations given above allow a straightforward explanation of the large difference between the geometries of the coiled-coils observed for CG(AT)<sub>5</sub> and CG(AT)<sub>4</sub>, which will be described later. Both should have a similar value of angle  $\theta$ , since the CG base pairs which define the kink are the same. On the other hand the  $\tau$  value should be quite different, since there is a difference of two base pairs in the length of oligonucleotide duplexes.

*The Dodecamer d(CGATATATATAT), [CG(AT)5]*

As shown in paragraph II.5.1, the molecular structure of this dodecamer could be determined from the X-ray results which showed that it is a right handed coil. From the experimental pattern it could be derived that:  $\beta = 20^\circ$ ,  $N = 6$  and  $\alpha = 60^\circ$ , thus, from equation II.12,  $\theta = 160.3^\circ$ :

$$\cos \frac{\theta}{2} = \sin 20^\circ \sin 30^\circ$$

$$\theta = 160.3^\circ$$

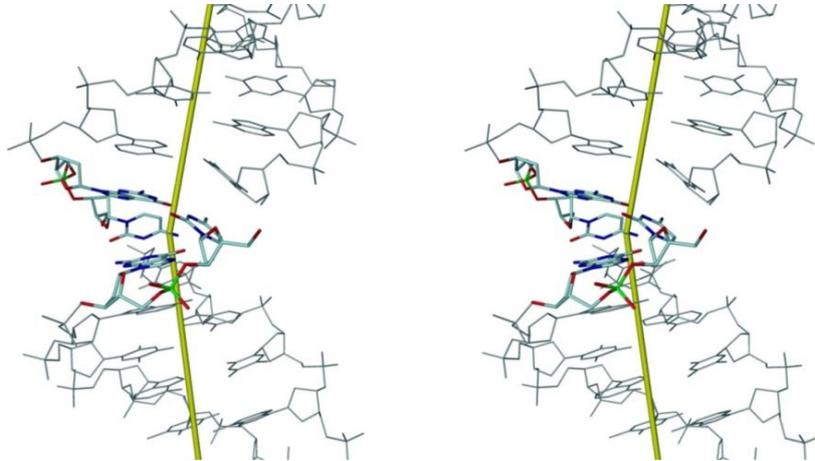
$$\theta' = 19.7^\circ$$

And from equation II.13,  $\tau$  is about  $57^\circ$ :

$$\cos \tau = \frac{2 \cos^2 \frac{60^\circ}{2}}{1 - \sin^2 \frac{60^\circ}{2} \sin^2 20^\circ} - 1 = 0.545$$

$$\tau = 56.96^\circ \approx 57^\circ$$

In this case:  $\theta' \approx \beta$  and  $\tau \approx \alpha$ . The coiled-coil has a radius  $R$  of about  $12.5 \text{ \AA}$ , given by equation II.4 ( $R = 36.75 \text{ \AA} (\sin 20^\circ / 2 \sin 30^\circ) = 12.57 \text{ \AA}$ ). The stereo view of the kink generated by the sticky end CG is shown in Figure II.4.4.



**Figure II.4.4.** Stereo view of the coiled-coil generated by the sequence  $d(CGATATATATAT)$ , the sticky ends (cyan) are shown. The discontinuity in the coil generates a kink of about  $160^\circ$  (see text for further details); the AT part of the molecule is shown in grey; the axes of the individual duplexes are in yellow (De Luchi *et al.*, 2006).

*The Decamer d(CGATATATAT), [CG(AT)4]*

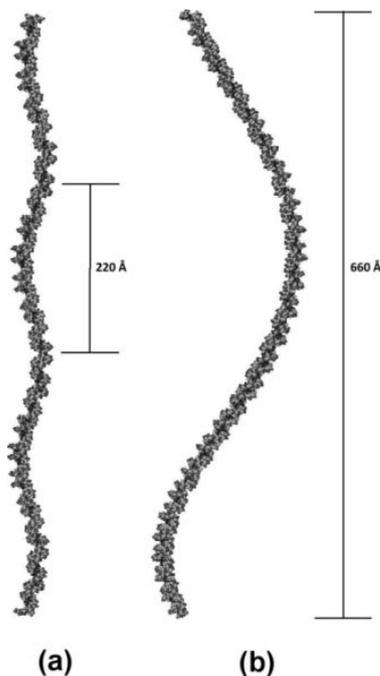
The torsion angle  $\tau$  is directly related to the twist value  $\omega$  of the individual base pairs. With ten base pairs instead of twelve and being the average twist value per base pair  $\langle\omega\rangle$  of about  $35^\circ$ ,  $\tau$  is about  $-13^\circ$  ( $\tau_{CGAT4} = \tau_{CGAT\%} - 2 \times \langle\omega\rangle$ ). A negative torsion angle is characteristic of left-handed coils. From the diffraction pattern, it is known that:  $\beta = 28^\circ$ ,  $N = 23$  and  $\alpha = 360^\circ/23 = 15.65^\circ$ , thus, from equation II.12:

$$\cos \frac{\theta}{2} = \sin 28^\circ \sin 7.83^\circ$$
$$\theta = 172.7^\circ$$

in this case the value of  $\theta'$ , of about  $7.3^\circ$ , is significantly different from the  $\beta$  value; with a kink angle  $\theta$  of about  $180^\circ$ , the coiled-coil is very smooth; the torsion angle  $\tau$  is very similar to  $\alpha$  ( $-\tau = 13.8^\circ \approx \alpha$ ). The radius  $R$  of the coiled-coil is of about  $47 \text{ \AA}$  ( $R = 27.5 \text{ \AA} (\sin 28^\circ / 2 \sin 7.83^\circ) = 47.3 \text{ \AA}$ ).

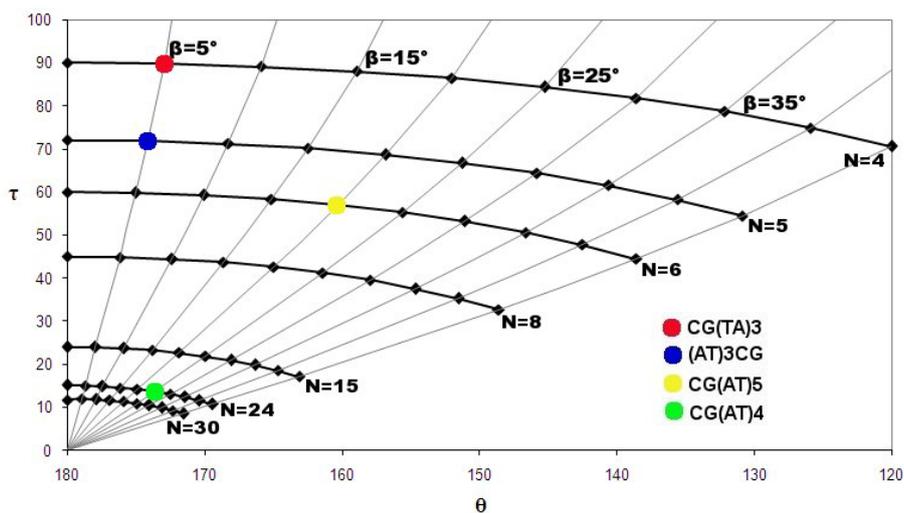
The difference in the  $\theta$  values found in the cases aforementioned should be attributed to changes in the intrinsic curvature of the duplexes. Packing interactions may also have an influence.

It should be noted that, in the case of  $CG(AT)_4$ , more complex coil structures, such as 24 residues in 5 turns, have been excluded. The diffraction from such coils would not coincide with the observed patterns. The two coils generated by  $CG(AT)_4$  and  $CG(AT)_5$  are shown in Figure II.4.5.

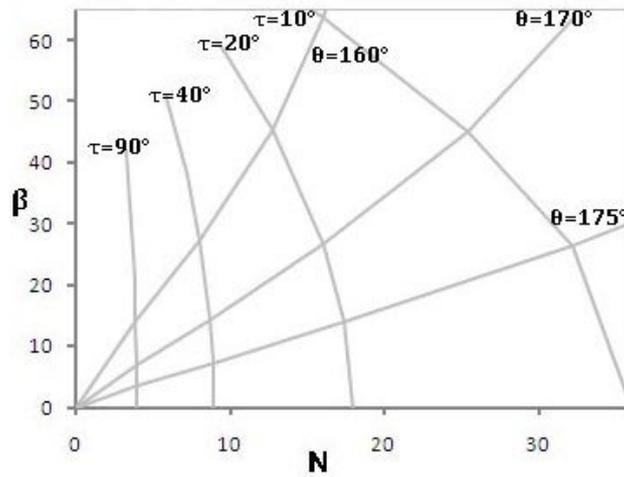


**Figure II.4.5.** (a) Coiled-coil generated by the dodecamer CG(AT)<sub>5</sub>, with a pitch of about 220 Å and six oligonucleotides per turn. (b) Coiled-coil generated by the decamer CG(AT)<sub>4</sub>, with a pitch of about 660 Å and twenty-three oligonucleotides per turn.

Values of  $N$  and  $\beta$  have been plotted as a function of  $\theta$  and  $\tau$  and vice versa in Figure II.4.6 and Figure II.4.7. In Figure II.4.6 the values of  $\theta$  and  $\tau$  have been calculated for  $N = 4, 5, 6, 15, 23, 30$  and  $\beta = 5^\circ$  to  $45^\circ$ . The geometrical characteristics of the coiled-coils generated by the sequences highlighted in Figure II.4.6 are summarized in Table II.2.



**Figure II.4.6.** Values of  $N$  and  $\beta$  as a function of  $\theta$  and  $\tau$ . Four examples are highlighted: d(CGATATATA), in red, d(ATATATCG), in blue, d(CGATATATATAT) in yellow and d(CGATATATAT) in green.



**Figure II.4.7.** Values of  $\theta$  and  $\tau$  as a function of  $N$  and  $\beta$ .

**Table II.2.** Geometrical characteristics of the coiled-coils generated by  $CG(TA)_3$ ,  $(AT)_3CG$ ,  $CG(AT)_4$  and  $CG(AT)_5$ .

Sequence	$\beta$ (°)	$N$	$\alpha$ (°) <sup>(1)</sup>	$\cos \frac{\theta}{2}$ <sup>(2)</sup>	$\theta$ (°)	$\theta'$ (°)	$\cos \tau$ <sup>(3)</sup>	$\tau$ (°)	$R$ <sup>(4)</sup>
$CG(TA)_3$	5	4	90	0.062	172.93	7.06	0.0038	89.8	1.61
$(AT)_3CG$	5	5	72	0.051	174.12	5.87	0.3126	71.8	1.93
$CG(AT)_4$	28	23	15.65	0.064	172.7	7.3	0.971	-13.8	47.3
$CG(AT)_5$	20	6	60	0.171	160.3	19.7	0.545	57	12.57

- (1) Equation II.2.  
(2) Equation II.12.  
(3) Equation II.13.  
(4) Equation II.4.

## II.5 DODECAMERS

The following dodecamers have been crystallized:

- **d(CGATATATATAT)**, see paragraph II.5.1.
- **d(CGCGATATATAT)**, see paragraph II.5.2.
- **d(ATATATATATCG)**, see paragraph II.5.3.
- **d(ATATATATATGC)**, see paragraph II.5.4.
- **d(GCATATATATAT)**, see paragraph II.5.5.
- **d(CGATATGCATAT)**, see paragraph II.5.6.
- **d(CGCGCGATATAT)**.

The following sequences generated isomorphous crystals: d(CGATATATATAT), d(CGCGATATATAT), d(ATATATATATCG). They are isomorphous with the (AT)<sub>6</sub> structure (Campos *et al.*, 2005).

No crystals suitable for X-ray diffraction have been obtained from sequence d(CGCGCGATATAT).



## Crystallization

Crystals obtained under different conditions diffracted to a maximum of 3 Å resolution. Long, flexible needles were grown at 13 °C. The hanging-drop vapor diffusion technique was used with 2-methyl-2,4-pentanediol (MPD) as a precipitant. The drop was equilibrated against a reservoir solution at 31% (vol/vol) MPD. The crystals were very sensitive to temperature, at 13 °C they presented rounded edges. Due to their high flexibility it was not possible to freeze them. By changing temperature to 4 °C, they became more rigid and suitable for manipulation. The best crystals have been obtained from the following conditions: 0.5 mM DNA duplex, 1.5 mM Spermine Tetrahydrochloride, 50 mM KCl, 50 mM NaCacodylate pH 6.5, 16.7 mM Thymidine and, alternatively, 15 mM TMAO (Trimethylamine-N-oxide) or 15 mM Ethylene Carbonate. The latter additives were not essential to obtain crystals, but they improved the resolution. They were used as possible hydrophobic residues suitable to enter the minor groove (Abrescia *et al.*, 2002). Different crystals gave practically identical unit cells.

## Data collection and structure determination

Diffraction data were measured with cryocooling at 110 K at the European Synchrotron Radiation Facility in the Spanish beamline BM16, processed and scaled using HKL2000 (Otwinoski & Minor, 1997) in the P6<sub>1</sub>22 space group. The structure was solved by molecular replacement with the program AMoRe (Navaza, 1994) using the best data set processed at 3 Å resolution. The fragment (AT)<sub>5</sub> has been used as a search model. The models with a Hoogsteen and a Watson-Crick base pairing were constructed from PDB entries 1GQU and 1DN9, respectively. The reasonable solution was found only for the Hoogsteen model. In this solution, the fragment (AT)<sub>5</sub> was aligned along the crystal *c* axis and had a molecular dyad which coincided with the crystallographic 2-fold axis. The extra CG bases have been manually built, in order to maintain a pseudo-continuous coil. To test the possibility of alternative pairing in the AT region, the Watson-Crick model was superimposed on the central decamer. Both models were refined using the CNS package (Brünger *et al.*, 1998). The refinement was carried for hexamer pGATATA/pTATAT+C that composes the asymmetric unit.

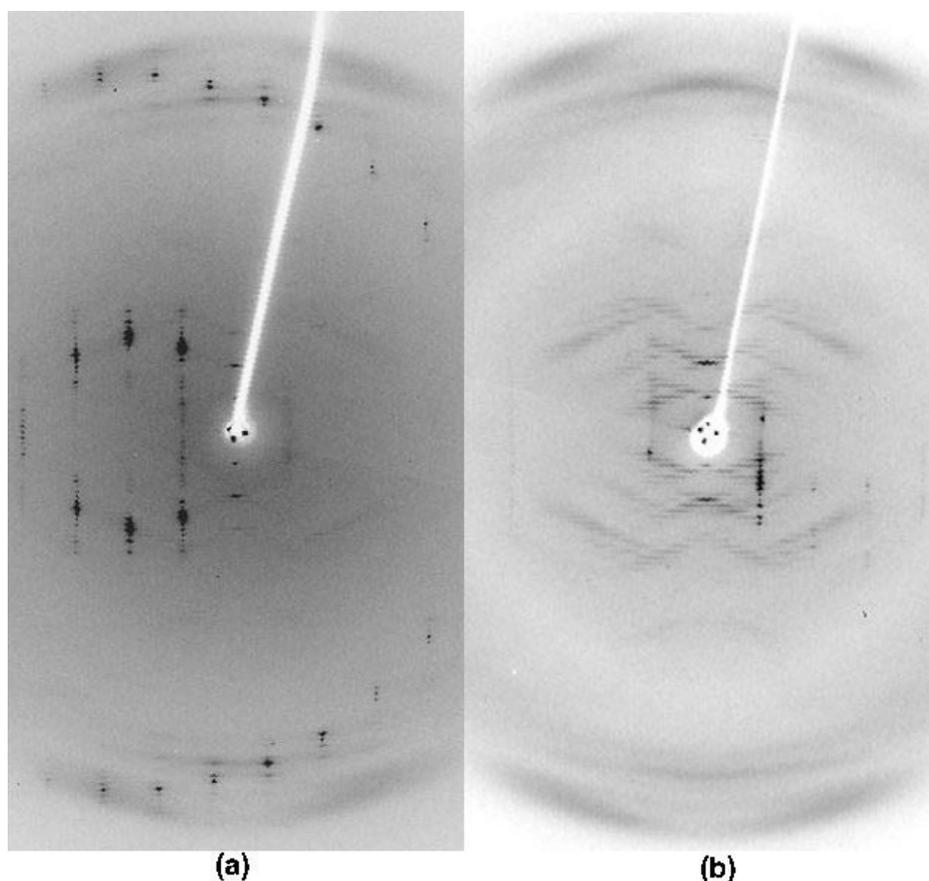
At all stages of the refinement the restraints were applied to maintain the geometry of the hydrogen bonds in Hoogsteen and Watson-Crick base pairs. Initially, each model was refined as a rigid body. After that, each base pair, each nucleotide and each nucleotide, phosphate and sugar were respectively treated as a rigid body. Simulated annealing was then carried out, yielding to an R-factor 10% lower for a Hoogsteen model than for a Watson-Crick model. At this stage, the electron density maps were calculated and both models were compared. In the AT region, the geometry of the Watson-Crick model was completely distorted, while the Hoogsteen model displayed the good hydrogen bonding and stacking arrangement. The Hoogsteen model was accepted for AT pairs in the further refinement. The region of the C·G base pairs showed a very poor electron density map which indicated high disorder. The C·G base pairs were maintained in the Watson-Crick conformation, but it cannot be excluded that they could also be in the Hoogsteen form. The entire dodecamer strand was then reconstructed from the refined hexamer model and the refinement concluded in program Refmac5 (Murshudov *et al.*, 1997). TLS parameters for one group were refined. The data collection and refinement statistics are given in Table II.3.

**Table II.3.** Data collection and refinement statistics

<b>Data collection</b>	
$\lambda$ , Å	0.9794
Cell parameters, Å	$a = b = 26.541$ $c = 220.527$
Space group	P6 <sub>1</sub> 22
Resolution range, Å	23.00-3.1
Completeness (%)	98.82
Overall redundancy	9.8
R <sub>symm</sub>	0.059
<b>Refinement statistics</b>	
N <sup>o</sup> DNA atoms	243
N <sup>o</sup> of reflections	984
Asymmetric unit content	One dodecamer (single strand)
R <sub>work</sub> , %	33.1
R <sub>free</sub> , % (10%)	34.6
rms bond lengths, Å	0.008
rms bond angles, °	2.040

## Structure description

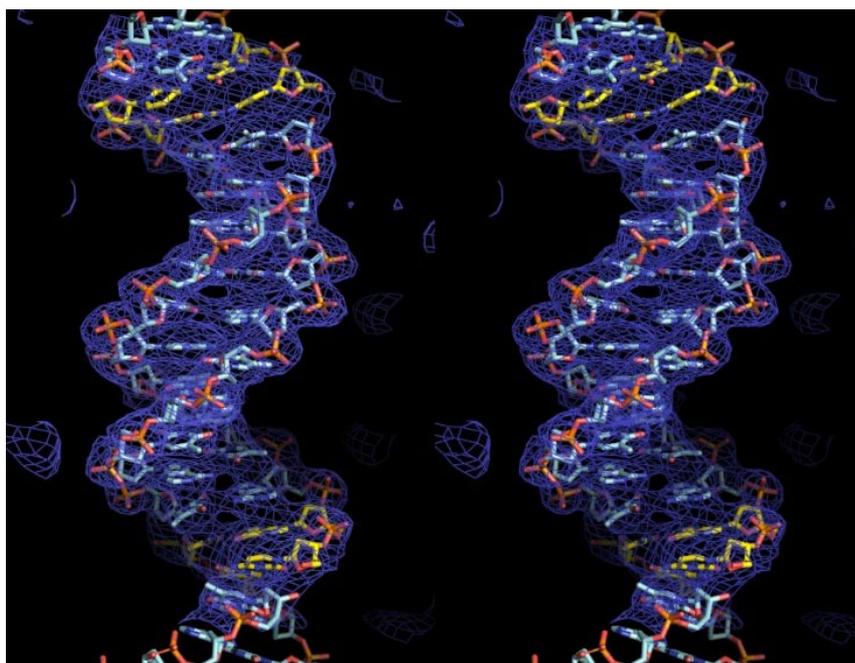
Typical diffraction patterns given by the sequence d(CGATATATATAT) are shown in Figure II.5.2. In some cases d(CGATATATATAT) crystals showed very few Bragg spots, only continuous diffraction on layer lines, as previously observed in d(ATATATATATAT) (Campos *et al.*, 2005). The presence of continuous diffraction indicates screw disorder, with parallel orientation of the diffracting entities. Slight variations in the crystallization conditions resulted in crystals which show little screw disorder and Bragg spots which extend up to 3 Å. Examples of both types of diffraction are given in Figure II.5.2. Although the resolution was much better than in the previous study of d(AT)<sub>6</sub>, the data had high mosaicity as it is apparent in Figure II.5.2.



**Figure II.5.2.** Oscillation patterns obtained from two different crystals with the  $c$  axis approximately perpendicular to the X-ray beam. (a) Crystalline pattern; high mosaicity in the upper layer lines may be appreciated ( $0.5^\circ$  oscillation). (b) Fiber-like pattern showing few Bragg spots and clear layer-lines which indicate strong screw disorder ( $5^\circ$  oscillation). The off-meridional stacking reflections indicate that in both cases the duplexes are inclined about  $25^\circ$  ( $\beta$ ) with respect to the  $c$  axis.

X-ray structure determination of d(CGATATATATAT) showed unambiguously that the AT base pairs had the Hoogsteen conformation. The electron density map is shown in Figure II.5.3. In particular the overall shape of the molecule demonstrates a narrow minor groove and the helical axis displaced towards one side of the base pairs, as expected for a Hoogsteen duplex (Abrescia *et al.*, 2004). The major groove is rather deep, as it is clearly apparent in Figure II.5.3.

The conformational parameters of the structure are given in Table II.4. The coiled coils in the crystal do not show any obvious lateral interactions. It appears that coiling is an intrinsic feature of these oligonucleotides, not due to crystal packing, which explains the strong tendency to screw disorder (see Figure II.5.2). The volume per base pair in the crystal is  $1871 \text{ \AA}^3$ , significantly greater than the value usually found in conventional oligonucleotide crystals (around  $1300 \text{ \AA}^3$ ). The high solvent content leads to high mosaicity in the crystals and limits the resolution of the diffraction data.



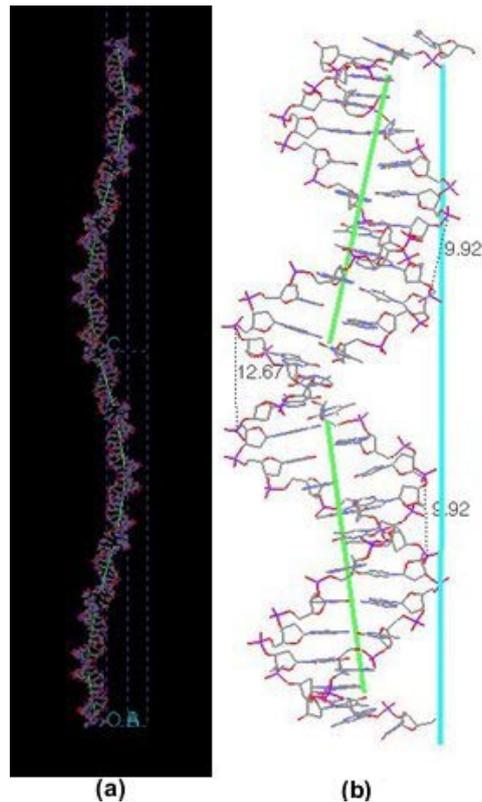
**Figure II.5.3.** Stereo view of the electron density of one duplex and its two cohesive ends. The C-G base pairs are shown in yellow. The Hoogsteen duplex formed by ten A-T base pairs is shown in cyan; it is practically straight. The narrow minor groove at the center and the deep major groove at both sides can be appreciated.

**Table II.4.** Conformational parameters of Hoogsteen DNA\*.

Angles										
Atom	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\chi$	Twist	C1'-C1' Å	Rise, Å
<b>A3</b>	-	-	23.4	134.9	-149.4	-117.3	63.8	27.50	8.2	3.61
<b>T4</b>	-70.0	143.1	72.4	105.5	-168.2	-92.3	-143.5	37.18	8.6	3.05
<b>A5</b>	-51.0	177.1	36.6	125.1	-167.0	-103.9	55.7	34.05	8.2	3.38
<b>T6</b>	-40.0	170.8	30.8	131.2	-69.4	-104.2	-113.6	42.16	8.1	3.33
<b>A7</b>	-62.4	175.1	59.3	124.9	-177.7	-90.0	46.5	27.11	8.0	3.68
<b>T8</b>	-57.2	-168.0	34.1	158.6	-141.2	-144.1	-83.8	42.16	8.0	3.33
<b>A9</b>	-78.9	145.7	65.4	118.4	149.1	-119.2	67.2	34.05	8.1	3.38
<b>T10</b>	-160.9	-88.7	104.4	107.4	100.7	-55.2	-104.4	37.18	8.2	3.05
<b>A11</b>	-146.1	-90.3	67.9	138.4	147.8	-96.7	87.2	27.50	8.6	3.61
<b>T12</b>	-87.1	-133.7	71.3	141.6	-	-	-78.9	-	8.2	-

\* The values have been calculated with the 3DNA program, based on C1'-C1' vectors (Lu & Olson, 2003).

The overall structure is shown in Figure II.5.4. It is a right handed coiled-coil, that shows kinks at the position of the CG base pairs, where the phosphodiester chain is interrupted (see Figure II.5.1). The kinks result in a strong compression on the major groove and opening of the minor groove, as it is apparent in Figure II.5.4, in agreement with classic studies (Dickerson *et al.*, 1983). The absence of two phosphate residues facilitates this distortion. The central AT decamer forms a straight duplex, as it is evident in Figure II.5.3. The terminal A·T base pairs form a large angle (equivalent to roll). The CG dimer sequence is compressed between both terminal A·T pairs. Since the structure is practically isomorphous with d(AT)<sub>6</sub>, the origin of the kinks should be attributed to the discontinuity of the phosphodiester chain, rather than to the presence of a short CG stretch.



**Figure II.5.4.** View of two turns of the coiled-coil (a) and detail of two consecutive duplexes (b). The axis of the duplexes is shown in green, whereas the axis of the coiled-coil is shown in cyan. The molecules are projected onto the plane formed by the duplex axes (green), so that the widening of the minor groove in the kink region may be easily appreciated.

It is interesting to note that decamer d(CGACGATCGT) also crystallizes as a continuous duplex with its sticky ends paired (Qiu *et al.*, 1997), but as a standard-Watson-Crick straight double helix. The results obtained with the dodecamer d[CG(AT)<sub>2</sub>GC(AT)<sub>2</sub>] (see paragraph II.5.6) also show straight double helices in the B-form. Since the starting sequence CGA is the same in the case of CG(AT)<sub>5</sub>, it is tempting to speculate that the coiled-coil conformation requires Hoogsteen base pairs.

The coiled-coil is a stable, rigid structure which represents a new conformation of DNA, as part of the polymorphism found in AT sequences reviewed elsewhere (Abrescia *et al.*, 2004). Such sequences are very abundant in non-coding regions of the genome (Abrescia *et al.*, 2004), but their structure and function are not known.

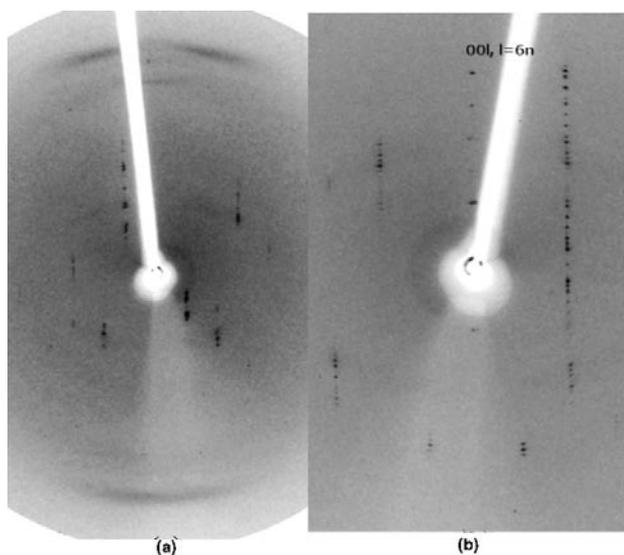
## II.5.2 d(CGCGATATATAT) d[(CG)<sub>2</sub>(AT)<sub>4</sub>]

The sequence (CG)<sub>2</sub>(AT)<sub>4</sub> has been crystallized at 13°C using the hanging-drop vapor diffusion technique and 2-methyl-2,4-pentanediol as a precipitant. Due to the high nucleation rate, instead of big single crystals, several small crystals, together with precipitate and thin needles, have been obtained. Only two crystals could be frozen and their diffraction pattern collected. From now on they will be referred to as “D18B2” and “D17A3”.

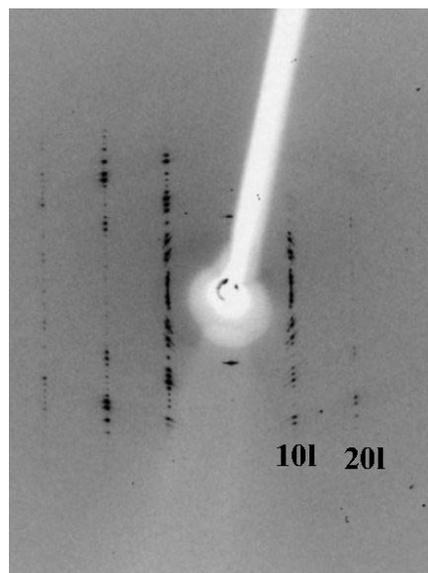
### The D18B2 crystal

The D18B2 crystal was obtained in the following conditions: 0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 1.0 mM Spermine Tetrahydrochloride, 2.5 mM MgCl<sub>2</sub>, 11.1 mM Thymidine and 1 μl of suspension of 4-aminophenylsulfon.

Typical diffraction patterns of D18B2 crystal are shown in Figure II.5.5 and Figure II.5.6. The D18B2 crystal is isomorphous to crystals generated by the sequences d[CG(AT)<sub>5</sub>], d[(AT)<sub>5</sub>CG] and d(AT)<sub>6</sub>. The sequence (CG)<sub>2</sub>(AT)<sub>4</sub> generates a coiled-coil, where the duplexes axes are inclined an angle  $\beta$  of about 11° with respect to the major coil axis. Unlike d[CG(AT)<sub>5</sub>] and d[(AT)<sub>5</sub>CG] crystals, in this case, no streaks have been detected in the diffraction patterns, suggesting the absence of the screw disorder found in the crystals of the isomorphous sequences.



**Figure II.5.5.** Oscillation patterns (2° and 3°) obtained from crystal D18B2, the long  $c^*$  axis of the unit cell is approximately vertical. (a) The stacking reflections at 3.25 Å deviate about 11° ( $\beta$ ) from a meridional orientation. (b) Meridional 00l Bragg reflections (with  $l = 6n$ ) are evident.



**Figure II.5.6.** Oscillation pattern ( $3^\circ$ ) of crystal D18B2; Bragg reflections in the equatorial region (with Miller indices 101, 201, 301) are shown. No streaks are visible.

The data set has a maximum resolution of about  $5.5 \text{ \AA}$ . It could be indexed in  $P6_1$  space group with the following unit cell:  $a = b = 26.37 \text{ \AA}$  and  $c = 227.5 \text{ \AA}$ . Dataset statistics are shown in Table II.5. Due to the low resolution of the data set, it has not yet been possible to solve the structure.

**Table II.5.** Dataset statistics for crystal D18B2.

Crystal	D18B2
DNA sequence	d(CGCGATATATAT)
Space group	$P6_1$
Unit cell ( $\text{\AA}$ )	$a = b = 26.37, c = 227.5$
Resolution ( $\text{\AA}$ )	5.5
Wavelength ( $\text{\AA}$ )	0.977
Unique reflections	782
Completeness (%)	96.7
$R_{\text{int}}$ (%)	7.3
$I/\sigma$	11.7

Given the sequence of the oligonucleotide, the DNA duplexes are organized in a staggered fashion and thus give rise to a continuous DNA double helix with nicks in both strands, as shown in Figure II.5.7. Nicks in either strand can be located only at distances of four base pairs.



**Figure II.5.7.** Model of the organization of the oligonucleotides in the crystal.

The unit cell parameters correspond to a cell that contains stacks of six dodecamers along the  $c$  direction (equivalent to 72 base pairs per unit cell) with their positions related by the helical  $P6_1$  symmetry. The volume per base pair is of about  $1900 \text{ \AA}^3$ , which indicates a large amount of solvent in the crystal structure. The stacking reflections at about  $3.25 \text{ \AA}$  are found at both sides of the meridian, indicating that the DNA duplexes are inclined an angle  $\beta$  of about  $11^\circ$  from the meridional or  $c$  direction (Figure II.5.5 a). The absence of layer lines with continuous diffraction is probably due to the presence of four CG base pairs instead of only two, as in  $CG(AT)_5$  and  $(AT)_5CG$ . The CG base pairs probably stabilize the structure.

The height  $h$  occupied by one duplex along the  $c$  direction of the unit cell corresponds to about  $37.9 \text{ \AA}$  ( $h = c/N = 227.5/6$ ); the length  $l$  of one dodecamer, taking into account its  $\beta$  inclination of about  $11^\circ$  is  $38.6 \text{ \AA}$  ( $l = h/\cos \beta$ ). The projection of one duplex onto the equatorial plane is thus of  $7.4 \text{ \AA}$  (given by:  $l \times \sin \beta$ ). The geometrical parameters of the coiled-coil generated by  $(CG)_2(AT)_4$  are summarized in Table II.6. The kink angle  $\theta$  between the two straight parts of the duplex (the  $(AT)_4$  and the  $(CG)_2$  part) is about  $11^\circ$ . The dimensions of the coiled coil are as follows: pitch ( $P$ )  $227.5 \text{ \AA}$ ; radius ( $R$ )  $7.4 \text{ \AA}$ ; inclination ( $\beta$ )  $11^\circ$ ; kink angle ( $\theta$ )  $11^\circ$ ; number  $N$  of dodecamers per turn, six.

**Table II.6.** Geometrical parameters of the coiled-coil generated by  $(CG)_2(AT)_4$ .

Crystal	P ( $\text{\AA}$ )	$N$	$l$ ( $\text{\AA}$ )	Rise ( $\text{\AA}$ )	$\beta$ ( $^\circ$ )	$R$ ( $\text{\AA}$ )	$\theta$ ( $^\circ$ )	$\alpha$ ( $^\circ$ )	$\tau$ ( $^\circ$ )
D18B2	227.5	6	38.6	3.22	11	7.4	169	60	59

$P$  is the pitch of the coiled-coil.

$N$  is the number of oligonucleotides per turn.

$l$  is the length of one oligonucleotide (taking into account its inclination  $\beta$ ).

$\beta$  is the inclination of the axis of the minor helix with respect to the major coil.

$R$  is the radius of the cylinder described by the coiled-coil (eq. II.4).

$\theta$  is the kink angle between two consecutive oligonucleotides (eq. II.12).

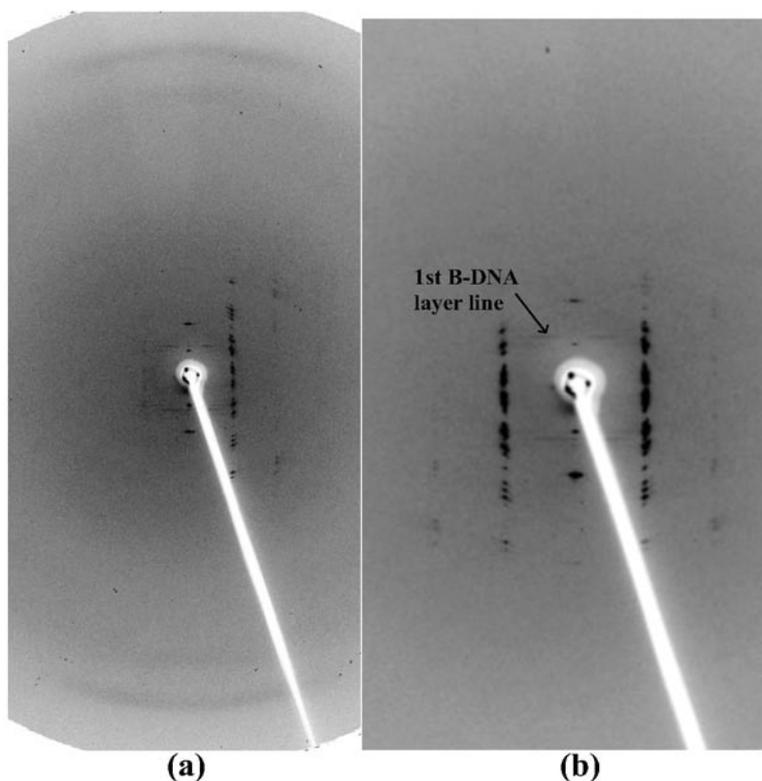
$\alpha$  is the turn angle (eq. II.2).

$\tau$  is the torsion angle (eq. II.13).

## The D17A3 crystal

The D17A3 crystal was obtained in the following conditions: 0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 1.0 mM Spermine Tetrahydrochloride and 5 mM BaCl<sub>2</sub>.

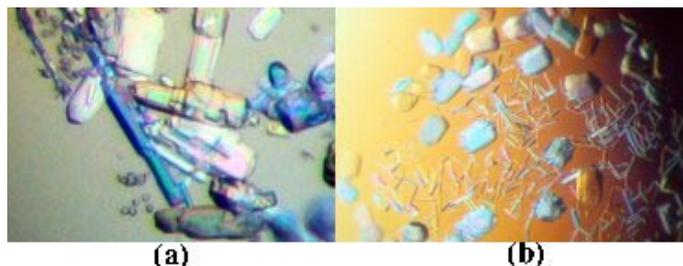
Typical diffraction patterns are shown in Figure II.5.8. The structure is different from the one described above for the D18B2 crystal. The DNA duplexes axes are almost parallel to the major coil axis. In fact, the stacking reflections at about 3.2 Å practically do not deviate from the meridional direction and appear like an arc rather than a sharp straight reflection. The first layer line of B form DNA is visible at about 32 Å resolution (Figure II.5.8 b). Therefore crystal D17A3 forms a standard continuous double helix with a repeat of about 10 base pairs per turn ( $32 \text{ \AA} / 3.2 \text{ \AA}$ ) and an average twist per base pair,  $\langle \omega \rangle$ , of 36°. The absence of one phosphate in the phosphodiesteric chain does not alter the structure and the helix is practically continuous.



**Figure II.5.8.** Oscillation patterns (5° and 15°) of the P17A3 crystal. (a) The stacking reflections appear like an arc, the inclination of the DNA duplexes is thus very small. (b) A few streaks are evident, but rather than a splitting of the layer line typical of coiled-coils, they appear to be the typical layer lines of the B DNA, only the 1<sup>st</sup> layer line is indicated.

### II.5.3 d(ATATATATATCG) d[(AT)<sub>5</sub>CG]

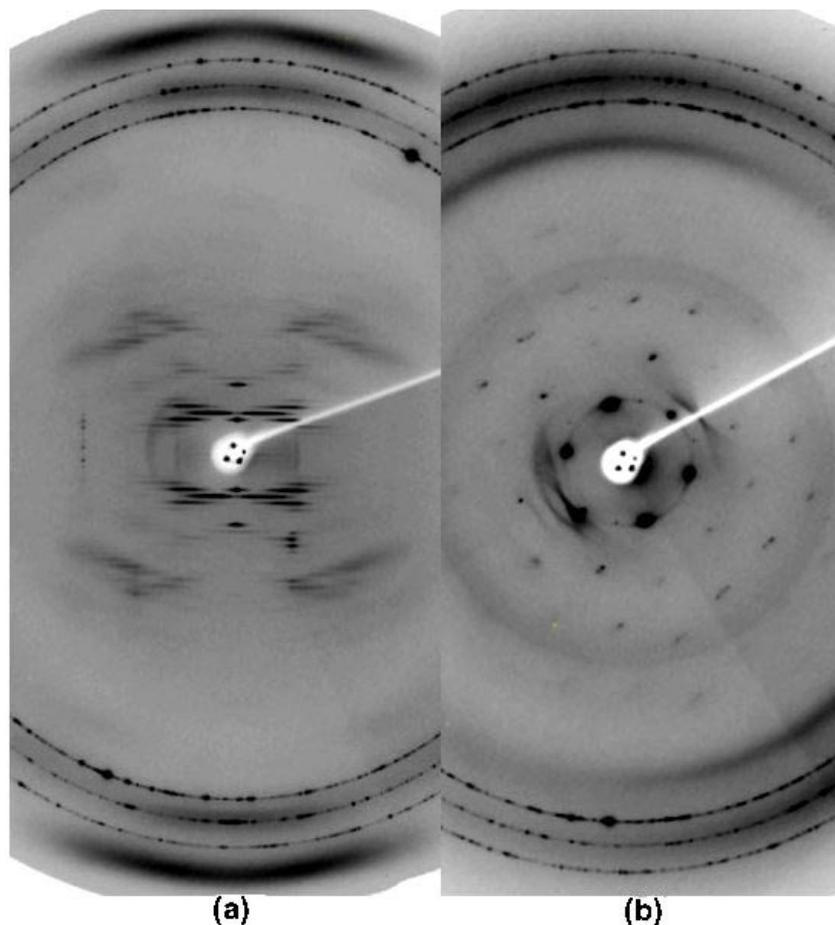
The sequence (AT)<sub>5</sub>CG was crystallized at 16 °C using the hanging-drop vapor diffusion technique and 2-methyl-2,4-pentanediol as a precipitant. Crystals (Figure II.5.9) have been obtained in the following conditions: 0.5 mM DNA duplex, 25 mM NaCacodylate pH 6.5, 1.5 mM Spermine Tetrahydrochloride, 50 mM KCl and 1 mM CoCl<sub>2</sub>. Only one crystal has been tested, from now on it will be referred to as “D34A5”.



**Figure II.5.9.** (a, b) Microscopic photographs of (AT)<sub>5</sub>CG crystals.

Typical diffraction patterns of crystal D34A5 are shown in Figure II.5.10. The diffraction pattern shows the coexistence of Bragg reflections and layer lines with continuous diffraction. The Bragg spots have a maximum resolution of about 6 Å. The data set could be indexed in a hexagonal space group with the following unit cell: *a* and *b* parameters about 26.3 Å and *c* parameter about 217 Å.

The sequence (AT)<sub>5</sub>CG generates a coiled-coil with six duplexes per turn (i.e., 72 base pairs along the *c* direction of the unit cell). The volume per base pair is thus of about 1805 Å<sup>3</sup> which indicates a large amount of solvent in the crystal ( $V_{bp} = abc \sin \gamma / 72$ ). The continuous layer lines appear at spacings that correspond to the 217 Å-repeat. The stacking reflections, at about 3.2 Å, are found at both sides of the meridian, indicating that the DNA duplexes are inclined an angle  $\beta$  of about 15° from the meridional or *c* direction. Crystalline regions give rise to the Bragg spots, on the other hand partially disordered regions give rise to the continuous diffraction. The dimensions of the molecules appear to be the same in both regions, because the *c* spacings of the Bragg reflections and of the continuous layer lines are identical.



**Figure II.5.10.** Oscillation patterns ( $15^\circ$ ) of crystal D34A5. (a) The long  $c^*$  axis of the unit cell is approximately vertical. Meridional Bragg reflections (006 and 0012) are evident. The prominent stacking reflections at about  $3.2 \text{ \AA}$  resolution deviate about  $15^\circ$  ( $\beta$ ) from the meridional direction. (b) Oscillation pattern of the same crystal rotated  $90^\circ$  with respect to (a), it approximately corresponds to the equatorial region of the crystal. A hexagonal symmetry is evident. Due to the orientation of the crystal, in this region most of the reflections are overlapped.

Given the sequence of the oligonucleotide, the DNA duplexes are organized in a staggered fashion and thus give rise to a continuous DNA double helix with nicks in both strands. As shown in Figure II.5.11, due to the terminal of guanine and cytosine, the pairing scheme is unique.



**Figure II.5.11.** Model of organization of the oligonucleotides in the crystal. They form duplexes with sticky ends that generate infinitely long molecules with staggered nicks in both strands.

The continuous layer lines are organized in groups that emanate from the meridional region. A prominent group is centered on the seventh layer line, which corresponds to  $31 \text{ \AA}$  ( $c/7$ ), equivalent to the first layer line of a DNA duplex, taking into account its  $\beta$  inclination of  $15^\circ$ .

Scattering of the layer lines is centered around the 7<sup>th</sup>, 14<sup>th</sup>, 21<sup>st</sup>, etc., layer lines, which correspond to the first three layer lines of the original scattering of a continuous DNA duplex (see Figure II.5.10). The 21<sup>st</sup> layer line region is significantly weaker than the rest. The coiled-coil structure results in splitting of the layer lines of the original double helix.

Both the Bragg reflections and the continuous layer lines can be explained by the same coiled-coiled model.

The height  $h$  occupied by a dodecamer along the  $c$  direction of the cell corresponds to about  $36.2 \text{ \AA}$  ( $h = c/6 \approx 36.2 \text{ \AA}$ ). Its approximate length  $l$  is of about  $36.8 \text{ \AA}$  ( $l = h/\cos \beta = 37.4 \text{ \AA}$ ) and its projection onto the equatorial plane is of about  $7 \text{ \AA}$  ( $36.8 \times \sin 11^\circ = 7.0 \text{ \AA}$ ), that is an estimation of the radius  $R$  of the coiled-coil. The geometrical parameters of the coiled-coil are summarized in Table II.7.

**Table II.7.** Geometrical parameters of the coiled-coil generated by the sequence  $(AT)_5CG$ .

Crystal	P ( $\text{\AA}$ )	$N$	$l$ ( $\text{\AA}$ )	Rise ( $\text{\AA}$ )	$\beta$ ( $^\circ$ )	$R$ ( $\text{\AA}$ )	$\theta$ ( $^\circ$ )	$\alpha$ ( $^\circ$ )	$\tau$ ( $^\circ$ )
D34A2	217	6	37.4	3.2	15	9.7	165	60	58.3

P is the pitch of the coiled-coil.

$N$  is the number of duplexes per turn.

$l$  is the length of one duplex (taking into account its inclination  $\beta$ ).

$\beta$  is the inclination of the axis of the minor helix with respect to the major coil.

$R$  is the radius of the cylinder described by the coiled-coil (eq. II.4).

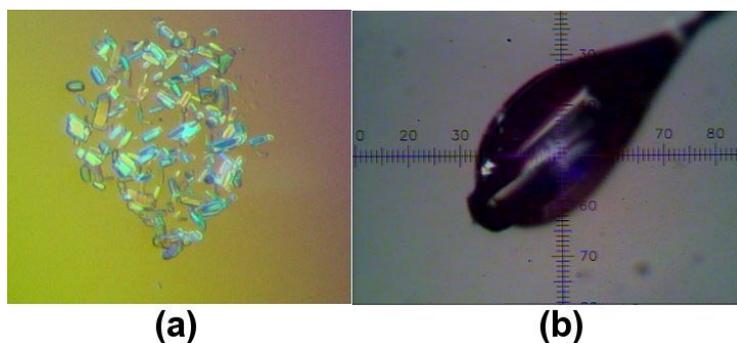
$\theta$  is the kink angle between two consecutive duplexes (eq. II.12).

$\alpha$  is the turn angle (eq. II.2).

$\tau$  is the torsion angle (eq. II.13).

## II.5.4 d(ATATATATATGC) d[(AT)<sub>5</sub>GC]

The oligonucleotide d(ATATATATATGC) was crystallized at 13 °C and 17 °C using the hanging-drop vapor diffusion technique and 2-methyl-2,4-pentanediol as a precipitant. Several crystals were obtained in very similar conditions: 0.5 mM DNA duplex, 25-50 mM NaCacodylate pH 6.5, 1.0-2.0 mM Spermine Tetrahydrochloride and 20 mM KCl, with or without the addition of 16.7 mM Thymidine. The presence of Thymidine did not improve the diffraction pattern of the crystals obtained. The best diffracting crystal (Figure II.5.12 b) was obtained at 13 °C in the following conditions: 0.5 mM DNA duplex, 25 mM NaCacodylate pH 6.5, 1.0 mM Spermine Tetrahydrochloride, 50 mM KCl and MPD 30%. This crystal will from now on be referred to as “D31B1”.

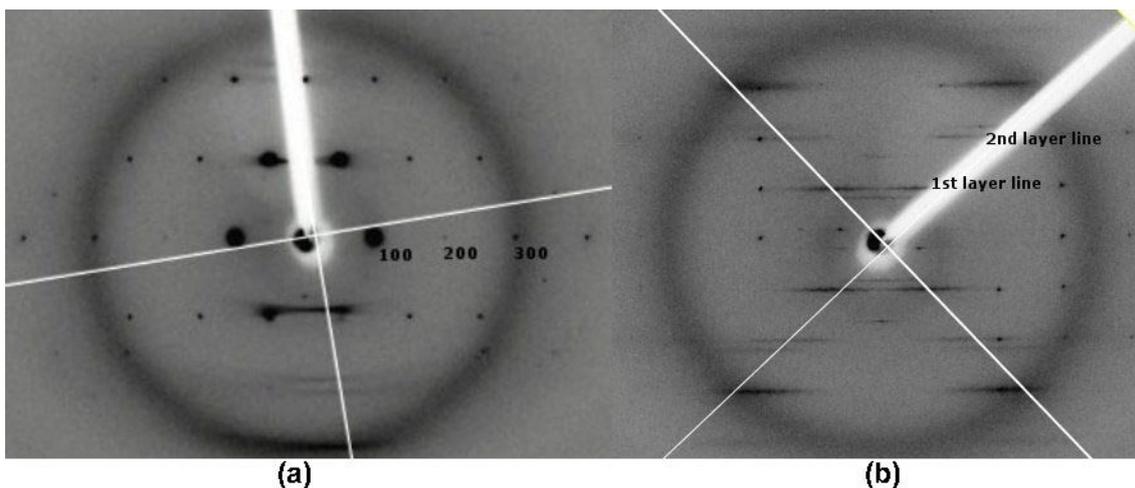


**Figure II.5.12.** (a) (AT)<sub>5</sub>GC crystals. (b) D31B1 crystal during data collection at the BM16 beamline at the ESRF, Grenoble.

Typical diffraction patterns of crystal D31B1 are shown in Figure II.5.13, Figure II.5.14 and Figure II.5.15. The diffraction patterns show the coexistence of Bragg reflections and layer lines with continuous diffraction. The Bragg spots approximately lay on the layer lines. Due to several factors, as low resolution (5.5 Å), intrinsic symmetry of the structure, presence of a small pseudo-cell, only a limited number of spots is visible. The combination of all these factors prevented the automatic determination of the unit cell which has instead been manually determined.

The stacking reflections, at about 3.30 Å resolution along the meridional axis (Figure II.5.17) indicate that this sequence forms a straight helix.

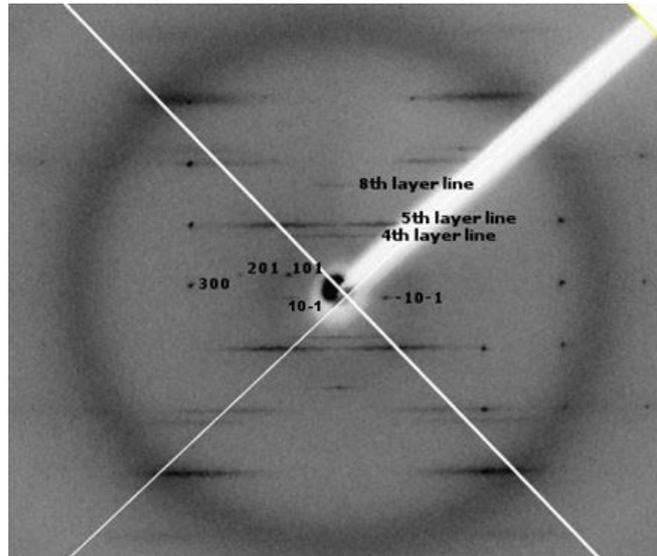
Apparently, the diffraction patterns could be indexed as a pseudo-hexagonal unit cell with  $a = b = 25.3$  Å and  $c = 29.8$  Å. These parameters could be obtained indexing the diffraction patterns as shown in Figure II.5.13.



**Figure II.5.13.** (a) Oscillation pattern ( $12^\circ$ ) of the equatorial region of the crystal. Assuming a pseudo-hexagonal symmetry,  $a$  and  $b$  are equal to  $25.3 \text{ \AA}$  ( $a = b = d_{100}/\sin 60$ , where  $d_{100} = 21.9 \text{ \AA}$ ). (b) Oscillation pattern ( $12^\circ$ ) of the same crystal approximately rotated  $90^\circ$  with respect to (a). In the figure some prominent layer lines which apparently correspond to the original diffraction of the DNA duplex are indicated. However, they appear at spacings which are multiple of  $29.8 \text{ \AA}$ , which correspond to a duplex of nine base pairs per turn (instead of ten) with an average rise of  $3.3 \text{ \AA}$ . See text for further discussion.

The equatorial plane of the crystal is shown in Figure II.5.13 (a), while in Figure II.5.13 (b) the meridional plane is shown. The spacing between the spots lying on the layer lines along the  $c$  direction corresponds to the  $c$  parameter equal to  $29.8 \text{ \AA}$ . According to this result, the double helix, which has an average rise of  $3.3 \text{ \AA}$ , appears to have nine base pairs per turn, instead of the ten base pairs per turn typical of the B-form DNA.

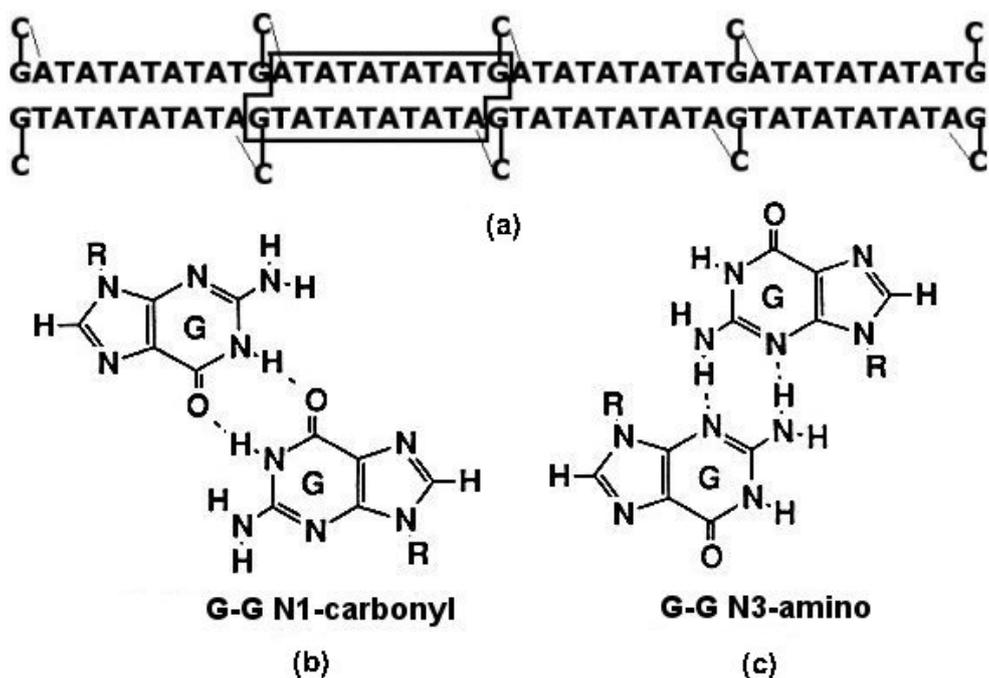
The indexing of the diffraction pattern shown in Figure II.5.13 only apparently fits with the data. In the equatorial region there are spots that cannot be explained by the unit cell aforementioned. The new indexing, shown in Figure II.5.14, corresponds to a hypothetical trigonal unit cell, with  $a = b = 44.8 \text{ \AA}$  and  $c \approx 147 \text{ \AA}$ . The first layer line in Figure II.5.13 (b) now becomes the fifth layer line, the presence of the fourth and eighth layer lines indicates that the repeating unit consists of a eleven-mer fragment of DNA with a length  $h$  of  $36.8 \text{ \AA}$  (this distance corresponds to the spacing between the 4<sup>th</sup> and 8<sup>th</sup> layer lines).



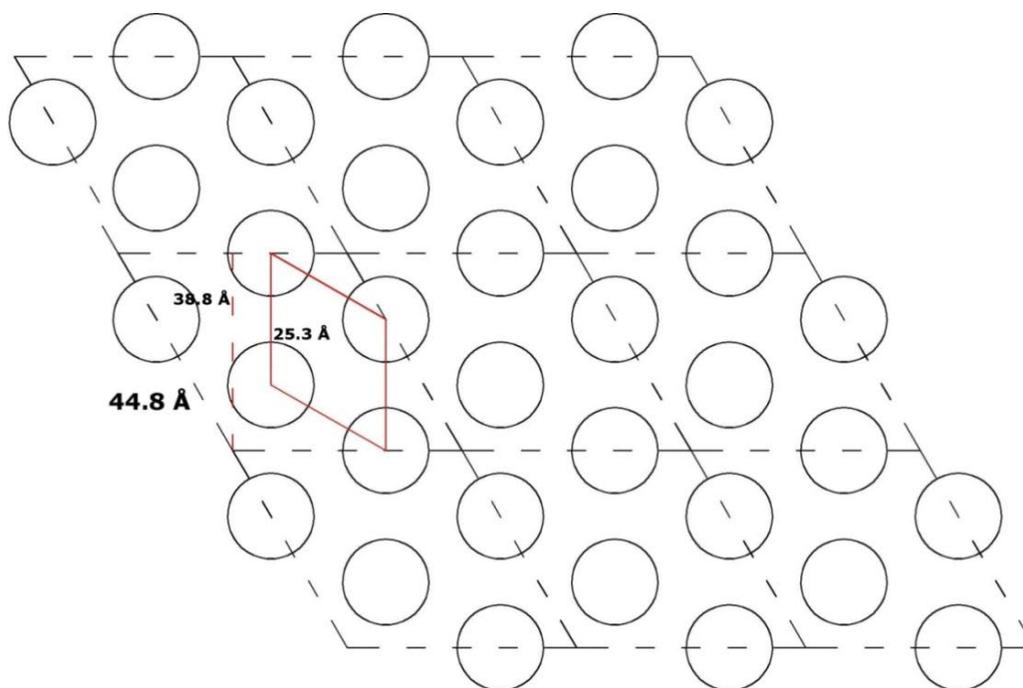
**Figure II.5.14.** Indexing of the meridional diffraction pattern. Although reflection 100 does not appear in this diffraction pattern, the presence of reflections 101, 10-1 and -10-1 allows the determination of the unit cell parameters.  $a$  corresponds to about  $45 \text{ \AA}$  ( $a = d_{100}/\sin 60 = 44.8 \text{ \AA}$ ). The spacing between the layer-lines corresponds to the  $c$  dimension of about  $147 \text{ \AA}$ , thus there are four duplexes per unit cell along the  $c$  direction ( $c/h = 147 \text{ \AA}/36.8 \text{ \AA}$ ).

A model that could explain this diffraction pattern corresponds to an eleven base pairs duplex with a Guanine-Guanine base-pair interaction and an extra-helical terminal Cytosine, as shown in Figure II.5.15. Terminal Cytosines are often disordered and cannot be detected in the electron density (Liu *et al.*, 1998; Abrescia *et al.*, 1999; Liu *et al.*, 1999; Valls *et al.*, 2004). The terminal Cytosine might also interact with an A-T base-pair in the major groove (Abrescia *et al.*, 1999; Pous *et al.*, 2008).

The unit cell sides ( $a = b$ ) of about  $45 \text{ \AA}$  correspond to a distance of  $26 \text{ \AA}$  between the axes of neighbor duplexes. This distance between columns of duplexes corresponds to a large volume per base pair of about  $1900 \text{ \AA}^3$ . Such a high hydration in the absence of strong interactions explains the low resolution of this crystal. The projection along the  $c$  direction of nine unit cells is shown in Figure II.5.16; in each unit cell there are three columns of four eleven-mer duplexes ( $c/h = 147 \text{ \AA}/36.8 \text{ \AA} \cong 4$  where  $h$  corresponds to the length of a duplex with eleven base pairs and an average rise of  $3.3 \text{ \AA}$ ).



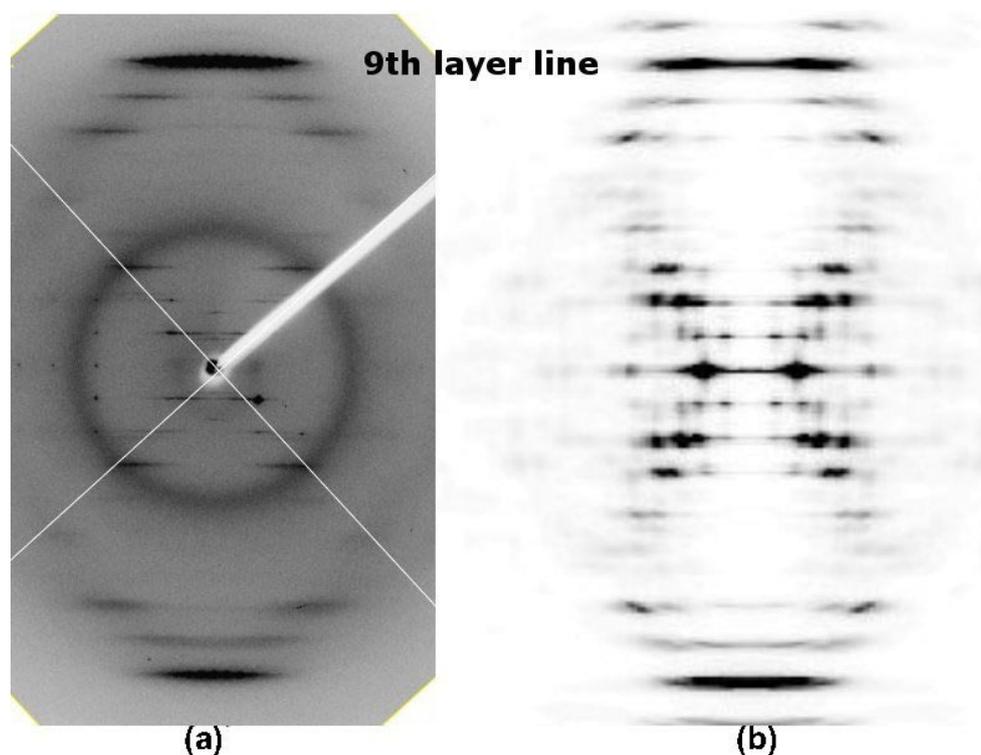
**Figure II.5.15.** (a) Model of the arrangement of eleven-mer fragments. This model was used to calculate the diffraction patterns shown in Figure II.5.17 and Figure II.5.18. (b and c). Possible symmetric G-G base pair interactions: (b) N1-O6 symmetric hydrogen bonds; (c) N2-N3 symmetric hydrogen bonds.



**Figure II.5.16.** Projection of nine unit cells in trigonal space group. In red, the small pseudocell with  $a = b = 25.3 \text{ \AA}$  containing the projection of one duplex; in black dashed lines, the bigger unit cell ( $a = b = 44.8 \text{ \AA}$ ), with three columns of duplexes. The red dashed line corresponds to the Bragg distance  $d_{100}$  of  $38.8 \text{ \AA}$ , measured from the diffraction pattern (see Figure II.5.14).

The continuous diffraction, calculated with CERIUS-2 (Accelrys, San Diego), of the eleven-mer model (Figure II.5.15) is shown in Figure II.5.17 (b). The model was built assuming the G-G pairing, with an extra Cytosine interacting with the neighbor AT base pair. The hypothesis of an eleven-mer repeating unit is confirmed by the comparison between the calculated and the experimental diffractions. In particular there is a strong agreement between the calculated and the experimental stacking diffractions at 3.3 Å. The stacking layer line corresponds to the 9<sup>th</sup> layer line of a nine-mer DNA fragment.

The simulated crystalline diffraction of the same model, calculate with CERIUS-2, is shown in Figure II.5.18. The model should be improved, in order to give a better agreement with the experimental diffraction, but for our purposes it is enough to see that *hkl* reflections with  $l = 5$ , are evident in both the calculated and the experimental patterns (see also Figure II.5.14).



**Figure II.5.17.** (a) Oscillation pattern (12°) obtained from the D31B1 crystal. The  $c^*$  axis is approximately vertical. The prominent stacking reflection at about 3.34 Å along the meridional axis indicates that this sequence forms a straight continuous helix. (b) Calculated fiber diffraction obtained with CERIUS-2 (equivalent to 180° oscillation). Note that the strong 9<sup>th</sup> layer line recorded in the experimental pattern appears also in the calculated diffraction.



## II.5.5 d(GCATATATATAT) d[GC (AT)<sub>5</sub>]

The sequence d(GCATATATATAT) was chosen as a variation of d(CGATATATATAT) [CG(AT)<sub>5</sub>]. Several crystals have been obtained, but, due to the low resolution of the data and the high mosaicity of the crystals, their structures have not been solved yet and only one data set could be indexed. The characteristics of the diffracting crystals are summarized in Table II.8.

The oligonucleotide GC(AT)<sub>5</sub> was crystallized at 4°C using the hanging-drop vapor diffusion technique and 2-methyl-2,4-pentanediol as a precipitant. Long needles were obtained in several conditions: 0.50 mM DNA duplex, 25-50 mM NaCacodylate pH 6.5 and 1-1.5 mM Spermine Tetrahydrochloride, with or without salt (20 mM NaCl, 50 mM KCl, 10 mM MgCl<sub>2</sub>, 1 mM CoCl<sub>2</sub> or 1 mM NiCl<sub>2</sub>). The crystals will from now referred to as “D34D3”, “D28C21”, “D28C1” and “D34D6” crystals. Only the D34D3 data set could be indexed; the crystal was obtained in the following conditions: 0.50 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 1.5 mM Spermine Tetrahydrochloride, 10 mM MgCl<sub>2</sub> and 1 mM CoCl<sub>2</sub>.

### The D34D3 crystal

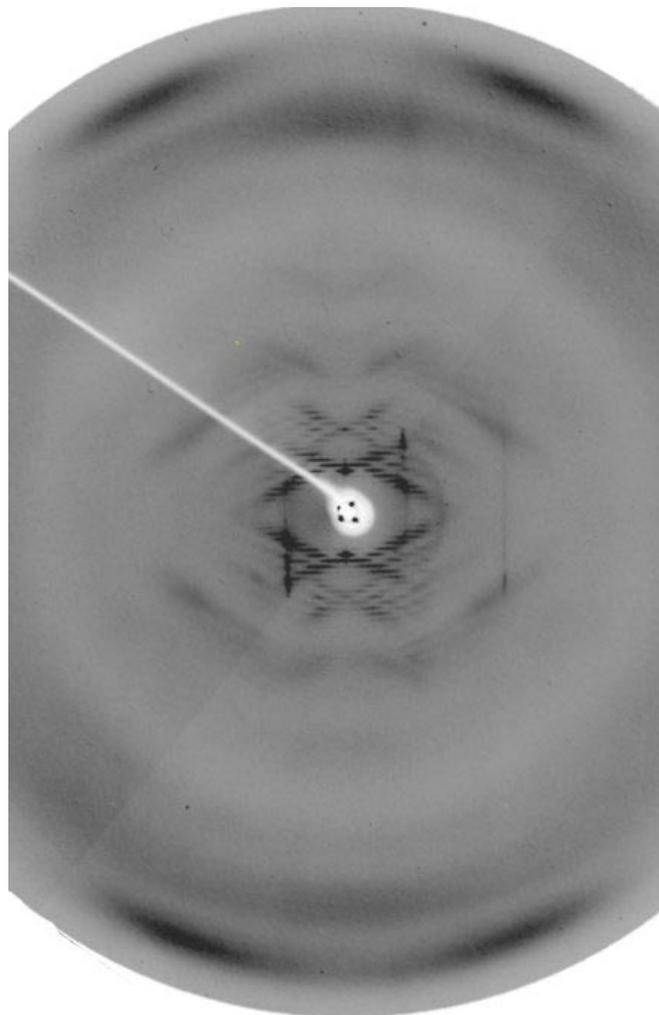
Typical diffraction patterns of D34D3 crystal are shown in Figure II.5.19 and Figure II.5.20. The diffraction patterns show the coexistence of Bragg reflections and layer lines with continuous diffraction. The Bragg spots have a maximum resolution of 11 Å, and they could be indexed in the space group C222<sub>1</sub> with the following unit cell parameters:  $a = 28.6 \text{ \AA}$ ,  $b = 47.6 \text{ \AA}$  and  $c = 479 \text{ \AA}$ . Since the ratio between the  $b$  and  $a$  parameters of the unit cell ( $b/a$ ) approximately corresponds to the root square of three ( $28.62 \times \sqrt{3} = 49.6$ ), the unit cell is likely to be trigonal or pseudohexagonal, in which case  $a$  and  $b$  unit cell parameters would correspond to 27.8 Å. The diffraction patterns confirm the possibility of a pseudo-hexagonal symmetry, as shown in Figure II.5.20.

As shown in Figure II.5.20 (a), the spacing of the continuous fiber diffraction corresponds to half of the Bragg spacing. The  $c$  parameter of about 480 Å corresponds to a cell that contains stacks of fourteen dodecamers corresponding to two turns of the coiled-coil.

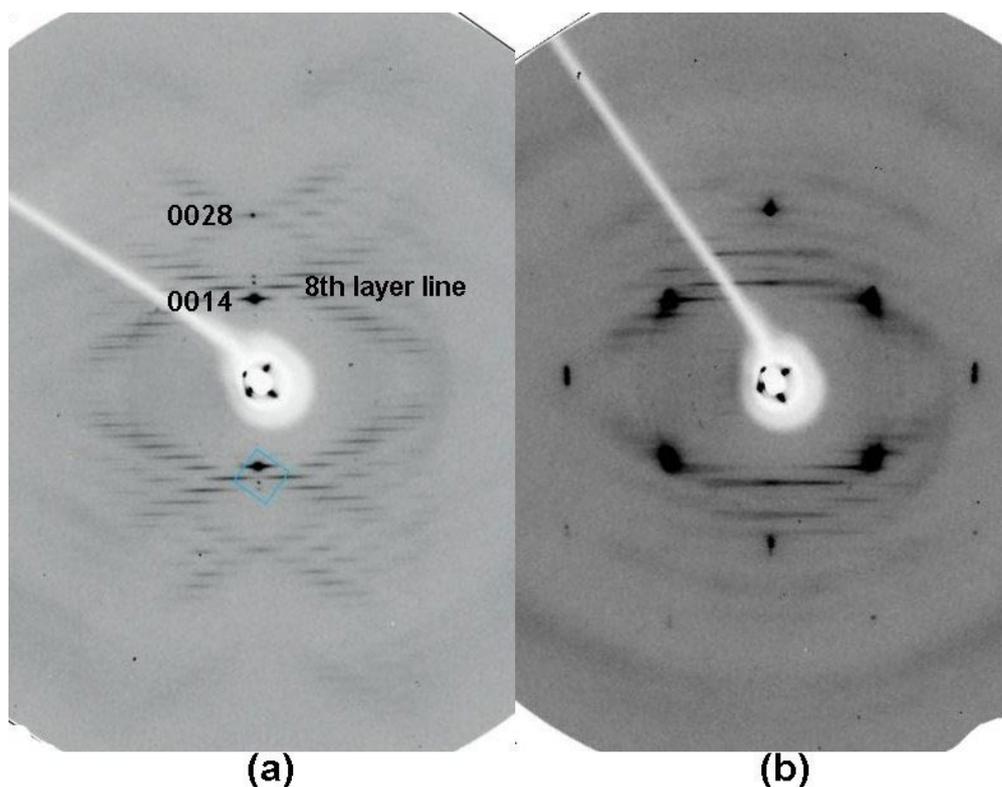
As shown in Figure II.5.19, the strong DNA stacking reflections are found at both sides of the meridian, indicating that the DNA duplexes are inclined about  $24^\circ$  from the meridional or  $c$  direction.

The height  $h$  occupied by one oligonucleotide along the  $c$  direction of the unit cell is approximately  $34.2 \text{ \AA}$  ( $c/14$ ).

Thus, considering the  $\beta$  inclination of the duplexes of about  $24^\circ$ , the length  $l$  of one DNA duplex is of  $37.5 \text{ \AA}$  ( $l = h/\cos \beta = 37.5 \text{ \AA}$ ); the average rise of the DNA thus is of  $3.12 \text{ \AA}$  ( $37.5 \text{ \AA}/12 = 3.12 \text{ \AA}$ ).



**Figure II.5.19.** Oscillation pattern ( $15^\circ$ ) obtained from D34D3 crystal. The long  $c^*$  axis of the unit cell is vertical. The prominent stacking reflections deviate about  $24^\circ$  ( $\beta$ ) from the meridional orientation. Bragg spots with Miller indices  $10l$  are also evident.



**Figure II.5.20.** Oscillation patterns ( $5^\circ$ ) of crystal D34D3. (a) An enlarged view of the meridional diffraction pattern is shown: layer lines spacing corresponds to half of Bragg spots spacings. The 8<sup>th</sup> layer line of the B-DNA and the Bragg spots with indices 0014 and 0028 are indicated. (b) The equatorial region of the diffraction pattern: a pseudo-hexagonal symmetry is visible.

Scattering at the layer lines is centered on the 8<sup>th</sup> and 14<sup>th</sup> layer lines. They correspond to the first two layer lines of the original scattering of a continuous DNA duplex with about ten base pairs per turn. The spacing of the continuous diffraction is about 240 Å, i.e. half of the Bragg spacing along the  $c$  direction (479 Å). The first DNA layer line (or the 8<sup>th</sup> layer line of the dodecamer) is at about 29.9 Å resolution ( $240 \text{ Å}/8 = 29.9 \text{ Å}$ ), which is approximately the length of 10 base pairs of the dodecamer (taking into account the inclination  $\beta$  of the oligonucleotides:  $l \times \cos \beta = 29.2 \text{ Å}$ , where  $l$  is the length of a duplex of ten base pairs with an average rise of 3.2 Å).

The diffraction pattern can be interpreted as due to a mosaic structure: crystalline regions that give rise to the Bragg spots and partially disordered regions in which the molecules are randomly displaced by vertical and rotational movement. The DNA duplexes are organized in a staggered fashion and thus give rise to a continuous double helix with nicks in both strands (Figure II.5.21).

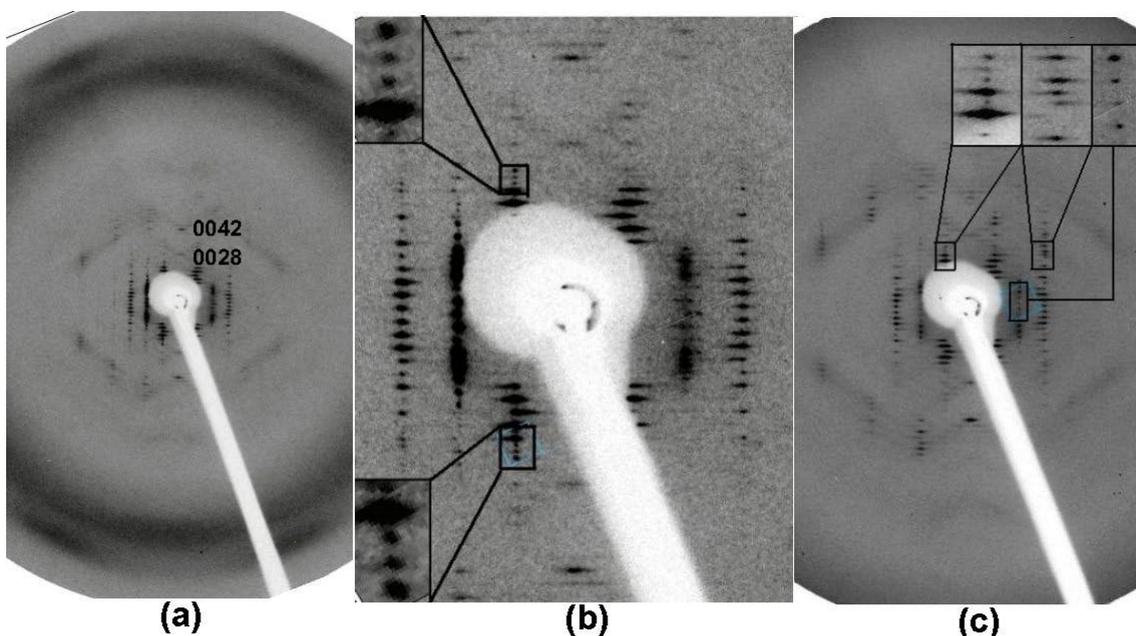


**Figure II.5.21.** A model of organization of the oligonucleotides in the crystal. They form duplexes with sticky ends that generate infinitely long molecules, with staggered nicks in both strands.

Data sets from other crystals have been collected (see Table II.8 for crystallization conditions): all of them show the coexistence of Bragg spots and continuous layer lines. Details relative to each of them are given in the paragraphs below.

### The D28C21 crystal

Typical diffraction patterns of D28C21 crystal are shown in Figure II.5.22. The sequence generates a coiled-coil with 14 oligonucleotides per turn. Layer lines and Bragg spots have the same spacings of about 450-480 Å. The inclination  $\beta$  of the duplexes axis is of about 25° with respect to the major coil axis.

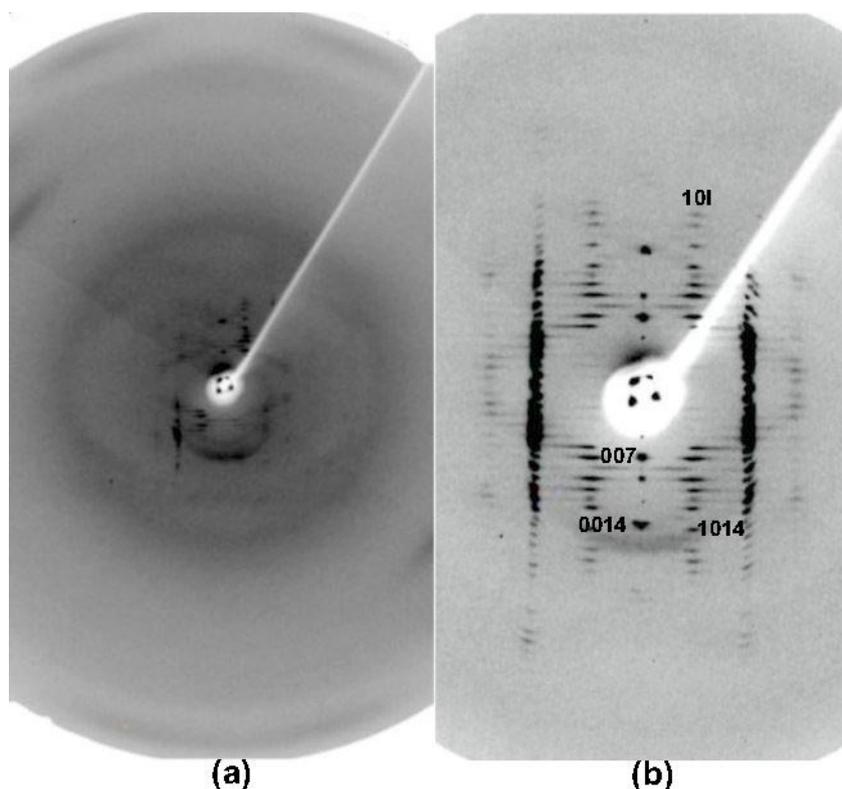


**Figure II.5.22.** Oscillation patterns (3°) obtained from crystal D28C21. (a) The long  $c^*$  axis of the unit cell is approximately vertical. Meridional Bragg reflections (0028 and 0042) are shown. The stacking reflections at 3.25 Å deviate about 25° ( $\beta$ ) from a meridional orientation. (b and c) The spacing of the layer lines corresponds to spacing of the Bragg reflections; due to the strong fiber background, the spacing distances had to be manually measured with the function MEASURE CELL implemented in the program MOSFLM (Leslie, 1992).

## The D28C1 crystal

The diffraction pattern of D28C1 crystal appears slightly different with respect to the diffraction of crystals D28C21 and D34D3. Apparently, the diffraction pattern could be indexed as shown in Figure II.5.23 (b): Bragg spots with indices 007, 0014 and 1014 are shown. Reflection 0014 is not exactly at the same level of reflection 1014: it appears in between the hypothetical 1012 and 1014 reflections. Therefore its Miller index, instead of 0014, is 0027, and the coiled-coil has 6.75 oligonucleotides per turn (i.e. 27 duplexes in 4 turns).

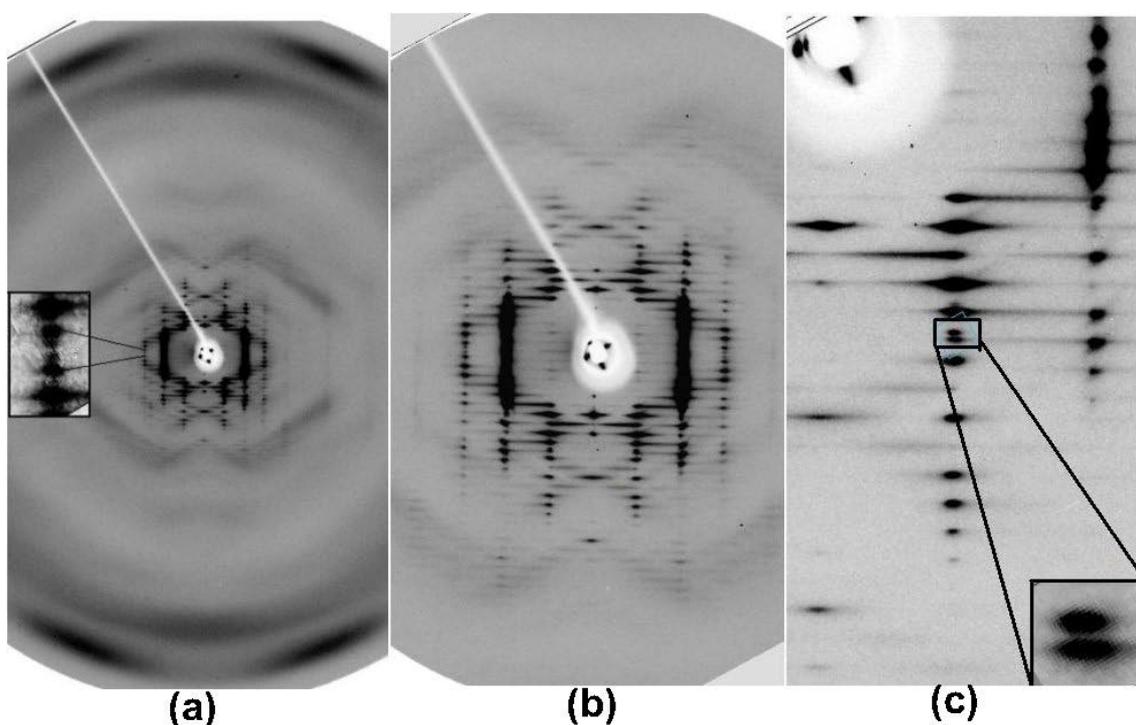
From the meridional diffraction pattern, the value of the  $c$  parameter of the unit cell can be determined: if the indexing shown in Figure II.5.23 (b) is correct, the  $c$  parameter corresponds to about 260 Å. If, as said before, the true index of reflection 0014 is 0027, the  $c$  dimension of the unit cell must be approximately doubled to 500 Å, the latter value would be in agreement with the  $c$  values found for the other crystals of this sequence.



**Figure II.5.23.** Oscillation patterns (a, 5°; b, 15°) of crystal D28C1. (a) Approximately meridional diffraction pattern. The stacking reflections at 3.28 Å deviate about 20° ( $\beta$ ) from a meridional orientation. (b) Enlarged view of the meridional diffraction: Bragg reflections with hypothetical Miller indices 007 and 0014 are shown; also Bragg reflections in the equatorial region (Miller indices 101) are visible.

## The D34D6 crystal

Typical diffraction patterns of PD34D6 crystal are shown in Figure II.5.24. In Figure II.5.24 (a), the long  $c^*$  axis of the unit cell is approximately vertical. The stacking reflections at  $3.25 \text{ \AA}$  deviate about  $24^\circ$  from the meridional orientation. Apparently the  $c$  parameter, manually measured, is of about  $240 \text{ \AA}$ , but the presence of “double spots” suggests a bigger  $c$  parameter of about  $900 \text{ \AA}$  (Figure II.5.24 c). Unfortunately, the presence of strong fiber diffraction does not allow an accurate determination of the unit cell parameters.



**Figure II.5.24.** Oscillation patterns (a,  $15^\circ$ ; b and c,  $5^\circ$ ) obtained from crystal D34D6. (c) Enlarged view showing the presence of “double spots”.

**Table II.8.** Summary of the characteristics of GC(AT)5 crystals, for all of them a pseudo-hexagonal symmetry is assumed. The maximum resolution is of about 11 Å. See text for further details.

Crystal	Unit cell (Å)	$\beta$ (°)	$l$ (Å)	Rise (Å)	Streaks	$N$	Crystallization conditions
D34D3	$a = b = 27.8$ $c = 479$	24°	37.4	3.12	Yes	7	0.5 mM DNA, 50 mM NaCacodylate pH 6.5, 10 mM MgCl <sub>2</sub> , 1 mM CoCl <sub>2</sub> , 1.5 mM Spermine and 28% MPD.
D28C21	$a$ and $b$ not determined $c \approx 480$	25°	37.8	3.15	Yes	14	0.5 mM DNA, 25 mM NaCacodylate pH 6.5, 50 mM KCl, 1 mM Spermine and 28% MPD.
D28C1	$a = b \approx 54$ $c \approx 260$ or 500	22°	-	3.25	Yes	7 or 6.75	0.5 mM DNA, 25 mM NaCacodylate pH 6.5, 20 mM NaCl, 1 mM Spermine and 30% MPD.
D34D6	$c$ apparently is about 240 Å, the presence of “double spots” clearly suggests a bigger $c$ dimension of about 900 Å. $a$ and $b$ are about 50 Å	24°	-	3.25	Yes	-	0.5 mM DNA, 50 mM NaCacodylate pH 6.5, 50 mM KCl, 1 mM NiCl <sub>2</sub> , 1.5 mM Spermine and 30% MPD.

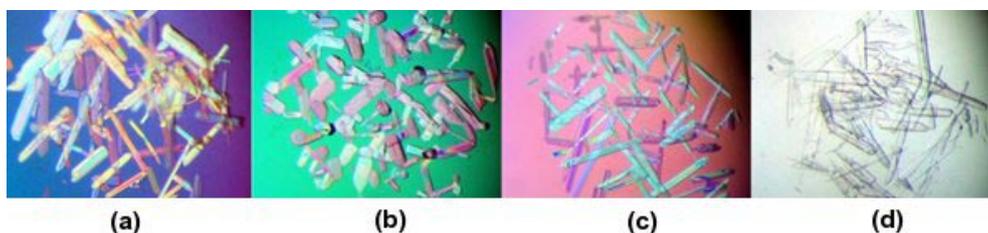
## II.5.6 d(CGATATGCATAT) d[CG(AT)<sub>2</sub>GC(AT)<sub>2</sub>]

The sequence d(CGATATGCATAT) was chosen in order to better understand the features of the structure of d[CG(AT)<sub>5</sub>] (De Luchi *et al.*, 2006). Our results indicate that the DNA duplexes are organized in a staggered fashion and thus give rise to a continuous DNA double helix with nicks in both strands. The base pairing, shown in Figure II.5.25 appears to be unique.



**Figure II.5.25.** Model of the organization of the oligonucleotides in the crystal. They form duplexes with sticky ends that generate infinitely long molecules, with staggered nicks in both strands.

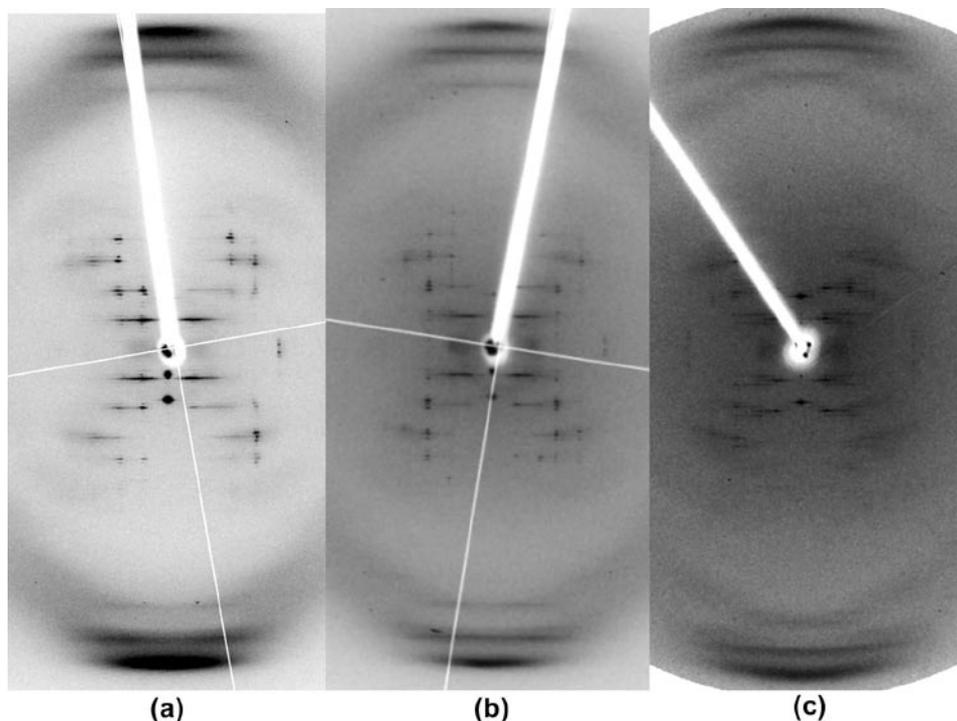
The sequence d(CGATATGCATAT) was crystallized at 17 °C by using the hanging-drop vapor diffusion technique and 2-methyl-2,4-pentanediol as a precipitant. Long needles were obtained from several solution conditions (Figure II.5.26): 0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 0.5-1.5 mM Spermine Tetrahydrochloride, with or without additives (Thymidine, TMAO, caprolactam, MgCl<sub>2</sub>, BaCl<sub>2</sub>, CoCl<sub>2</sub>, CaCl<sub>2</sub>). The addition of divalent cations seems to be essential for a good diffraction, while the addition of the other additives (Thymidine, TMAO and Caprolactam) did not have any appreciable influence on the diffraction pattern of the crystals obtained. Seven crystals have been obtained (D50A5, D59D4, D50A6, D59D6, D59D6-1, D60A2 and D60A3); details relative to the each of them are reported in Table II.9.



**Figure II.5.26.** Microscope photographs of crystals of d(CGATATGCATAT), all of them appeared as long needles. The following additives have been used: (a) D59D6 crystal: 5 mM CaCl<sub>2</sub>, TMAO and Thymidine; (b) D59D4 crystal: 2.5 mM CaCl<sub>2</sub>; (c) D50A6 crystal: 5 mM CaCl<sub>2</sub>; (d) D60A3 crystal: BaCl<sub>2</sub> and Caprolactam.

Typical diffraction patterns are shown in Figure II.5.27. The diffraction patterns show the coexistence of Bragg reflections and layer lines with continuous diffraction. The Bragg spots have a maximum resolution of 7.0 Å. The data set could be indexed on a hexagonal unit cell with the following parameters:  $a = b = 46.5 \text{ \AA}$  and  $c = 230.5 \text{ \AA}$ . The unit cell parameters correspond to a cell which contains three stacks of six dodecamers, equivalent to 216 base pairs per unit cell. The volume per base pair is thus  $1998 \text{ \AA}^3$ , which is in agreement with previous results. The volume per base pair indicates the presence of a large amount of solvent, which could explain the low resolution observed in the diffraction patterns. The continuous layer lines appear at spacings that correspond to the  $230.5 \text{ \AA}$ -repeat. The same model can explain both the Bragg spots and the continuous layer lines.

The strong stacking reflections are slightly off-meridional and not sharp, as shown in Figure II.5.27. This is evident in Figure II.5.27 (c), where the stacking appears like an arc more than like a sharp streak. In the same frame, the splitting of the layer lines, a typical feature of the coiled-coils, is also evident. The DNA duplexes are thus inclined from the meridional or  $c$  direction by an angle  $\beta$  of  $0^\circ$  to  $5^\circ$  degrees.



**Figure II.5.27.** Oscillation patterns of three different crystals of d(CGATATGCATAT). (a) D60A2 crystal ( $3^\circ$ ); (b) D59D6 crystal ( $5^\circ$ ) and (c) D59D4 crystal ( $15^\circ$ ).

When the duplexes axes are parallel (i.e.  $\beta = 0^\circ$ ), the height  $h$  occupied by one dodecamer along the  $c$  direction corresponds to  $38.4 \text{ \AA}$  ( $c/6 = 38.41 \text{ \AA}$ ), with an average rise of  $3.2 \text{ \AA}$  ( $h/12 = 3.2 \text{ \AA}$ ). The dodecamers are organized end-to-end and build a continuous coil with nicks corresponding to the sticky ends, as shown in Figure II.5.25. Neighbor duplexes are practically coaxial. The overall rotation of one duplex with respect to its neighbor in a column is called  $\Omega$ . The value of  $\Omega$  can be exactly determined once the number  $N$  of duplexes stacked in one unit cell is known.  $N\Omega$  must be an exact multiple of  $360^\circ$ :

$$N\Omega = 360m \quad (\text{II.14})$$

with  $m$  a whole number. In the case of the sequence  $\text{CG(AT)}_2\text{GC(AT)}_2$ ,  $N = 6$ ,  $m = 7$  and  $\Omega = 420^\circ$ .  $\Omega$  is also related to the individual base pair twist angle by the relation:

$$\Omega = n\langle\omega\rangle \quad (\text{II.15})$$

where  $n$  is the number of base pairs in a duplex and  $\langle\omega\rangle$  is the average twist of its base steps. In one unit cell there are stacks of 72 base pairs ( $N \times 12$ ). The average twist value for each base pair is thus of  $35^\circ$  ( $\langle\omega\rangle = \Omega/72$ ); consequently there are 10.3 base pairs per turn ( $360^\circ/\langle\omega\rangle$ ).

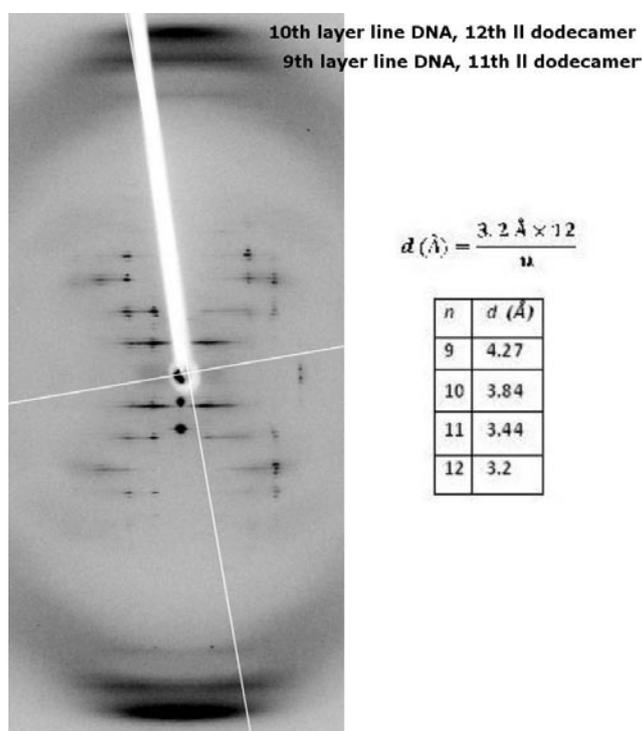
The average DNA twist in solution is 10.4 base pairs per turn, therefore it could be concluded that the absence of one phosphate does not change the organization of the double helix. The sequence  $\text{d(CGATATGCATAT)}$  generates a practically continuous double helix of standard B-form DNA. On the other hand, the average twist value  $\langle\omega\rangle$  of  $35^\circ$  is slightly smaller than the  $35.9^\circ$  value reported in the literature for mixed CG/AT sequences (Gorin *et al.*, 1995). This observation indicates that although the overall structure of the DNA is not affected by the absence of a phosphate, the individual twist values suffer a small change.

When  $\beta = 5^\circ$  (Figure II.5.27 (c)), all previous calculations are practically still correct. Due to the  $\beta$  inclination of helical axis with respect to the major coil axis, some parameters are slightly different. The crystal D59D4 has been indexed in the following hexagonal unit cell:  $a = b = 47.61 \text{ \AA}$  and  $c = 227.96 \text{ \AA}$ . The height  $h$  occupied by one dodecamer along the  $c$  direction is of about  $38 \text{ \AA}$  ( $c/6 = 37.99 \text{ \AA}$ ). Therefore, the length  $l$  of a dodecamer corresponds to  $38.14 \text{ \AA}$  ( $l = h/\cos\beta$ ). Its projection onto the equatorial plane is of  $3.32 \text{ \AA}$  ( $l \times \sin\beta$ ). The radius of the coiled-coil can be estimated from the latter value.

In B-form DNA, the 9<sup>th</sup> layer line is generally found at 3.55 Å; on the other hand, in the meridional diffraction pattern shown in Figure II.5.28, the 9<sup>th</sup> layer line corresponds to 3.44 Å. This is probably due to the fact that the B-form-DNA 9<sup>th</sup> layer line coincides with the dodecamer 11<sup>th</sup> layer line, whose meridional spacing  $d$  corresponds to 3.44 Å ( $d(\lambda) = \frac{3.2 \text{ Å} \times 12}{11}$ ).

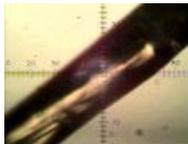
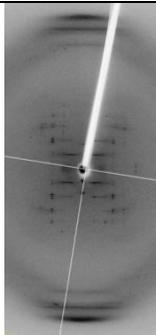
Although the sequence d(CGATATGCATAT) was chosen just to improve the crystallization of d[CG(AT)<sub>5</sub>], a different structure, probably due to the presence of the central GC bases, was obtained. The four CG base pairs form Watson-Crick hydrogen bonds and therefore force the flanking ATs to also pair through Watson-Crick bonding instead of Hoogsteen, as seen in d[CG(AT)<sub>5</sub>] (De Luchi *et al.*, 2006). Probably thanks to the coherence in the H-bonds throughout the whole molecule, the bends formed by the two terminal bases are less pronounced and the resulting structure is a standard B-DNA instead of a coiled-coil.

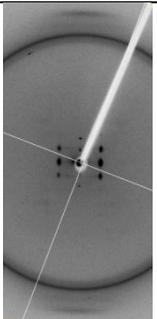
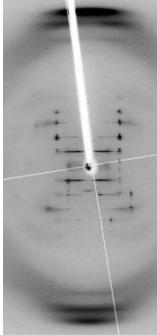
The characteristics of the diffracting crystals are summarized in Table II.9



**Figure II.5.28.** Oscillation pattern (3°) obtained from crystal D60A2, which has  $\beta = 0^\circ$ . The 9<sup>th</sup> layer line of the B-form DNA, found at 3.44 Å instead of 3.55 Å, corresponds to the 11<sup>th</sup> layer line of the dodecamer d(CGATATGCATAT). In the table on the right, the meridional spacings  $d$  for the dodecamer have been calculated as follows:  $d(\lambda) = \frac{3.2 \text{ Å} \times 12}{n}$ , where  $n$  is the layer line number.

**Table II.9.** Summary of the characteristics of the diffracting crystals of sequence d(CGATATGCATAT).

Crystal	Unit cell (Å)	Streaks	N	l (Å)	$\beta$	Rise (Å)	Comments	Crystallization conditions
D50A5	$c \approx 230$ (for 00l, $l=6n$ ) $a$ and $b$ not determined	Yes	6	38.3	$0^\circ$	3.22	-	0.5 mM DNA, 50 mM NaCac pH 6.5, 10 mM $MgCl_2$ , 1 mM $CoCl_2$ , 1 mM Spermine and MPD 37%.
D59D4	$a=b=47.61$ $c=227.96$ P6(1) (for 00l, $l=6n$ )	Yes	6	37.9	$5^\circ$	3.16	See Figure II.5.27	0.5 mM DNA, 50 mM NaCac pH 6.5, 5 mM $CaCl_2$ , 1.5 mM Spermine and MPD 28%.
D50A6	---	Diffuse	--	---	$0^\circ$	3.21	Big xtal, but very poor diffraction 	0.5 mM DNA, 50 mM NaCac pH 6.5, 5 mM $CaCl_2$ , 1.0 mM Spermine and MPD 37%.
D59D6	$a=b=46.5$ $c=230.5$ P6(1) (for 00l, $l=6n$ )	Yes	6	38.41	$0^\circ$	3.2	 This is a standard B-DNA	0.5 mM DNA, 50 mM NaCac pH 6.5, 5 mM $CaCl_2$ , TMAO, Thymidine, 1.0 mM Spermine and MPD 28%.

Crystal	Unit cell (Å)	Streaks	N	l (Å)	$\beta$	Rise (Å)	Comments	Crystallization conditions
D59D6-1	$c=230$ $a$ and $b$ not determined	Diffuse diffraction	-	38.3	$0^\circ$	3.17	 Diffuse diffraction, liquid crystal.	0.5 mM DNA, 50 mM NaCac pH 6.5, 5 mM $\text{CaCl}_2$ , TMAO, Thymidine, 1.0 mM Spermine and MPD 28%.
D60A2	$a=b=46.5$ $c=230.5$ P6(1) (for 00l, $l=6n$ )	yes	6	38.41	$0^\circ$	3.2	 Standard B-form DNA	0.4 mM DNA, 50 mM NaCac pH 6.5, 5 mM $\text{BaCl}_2$ , 75 mM Caprolactam, 1.0 mM Spermine and MPD 26%.
D60A3	-			38.3	$0^\circ$	3.18	Very weak diffraction	0.4 mM DNA, 50 mM NaCac pH 6.5, 5 mM $\text{BaCl}_2$ , 100 mM Caprolactam, 1.0 mM Spermine and MPD 26%.

## II.5.7 Discussion of dodecamer structures

The dodecamers studied in this work tend to form coiled-coils in the Hoogsteen conformation, with some exceptions due to the sequence of the DNA oligonucleotides.

In the structure generated by the sequence  $(AT)_5GC$  the terminal cytosine is flipped out from the duplex (paragraph II.5.4). This is the only case studied with cytosine in the 3' end of the sequence. Cytosines at the 5' end form stable base pairs in the sequences studied in this work.

When the alternating AT duplex is interrupted by CG base pairs, a standard B-form DNA is generated, as in  $CG(AT)_2GC(AT)_2$  (paragraph II.5.6).

Apparently, the Hoogsteen hydrogen bonding is a prerequisite for the formation of coiled-coiled structures. The following sequences are characterized by the same sticky end and the same starting sequence:  $d(CGATATATATAT)$  (De Luchi *et al.*, 2006),  $d(CGACGATCGT)$  (Qiu *et al.*, 1997) and  $d(CGATATGCATAT)$  (see paragraph II.5.6); nevertheless, only the sequence  $CG(AT)_5$  gave rise to a coiled-coil. While the sequences  $d(CGACGATCGT)$  and  $d(CGATATGCATAT)$  generate standard Watson-Crick B-DNAs, the sequence  $CG(AT)_5$  forms a Hoogsteen superhelix. It is thus reasonable to conclude that the discriminating factor in order to generate a coiled-coil is the possibility of generating Hoogsteen hydrogen bonds instead of Watson-Crick ones in the straight duplex regions.

Another striking feature of the dodecamers studied here, is the apparently small value of the DNA rise: several cases show a rise smaller than  $3.2 \text{ \AA}$ , corresponding to a dodecamer length  $l$  shorter than  $38 \text{ \AA}$ . This fact might suggest a distortion in the kink corresponding to the sticky end.

A summary of the characteristics of the dodecamer structures is shown in Table II.10.

**Table II.10.** Characteristics of the dodecamers crystallized in this thesis.

Sequence	Unit cell (Å)	$l$ (Å)	Rise (Å)	$\beta$ (°)	$N$	$P$ (Å)	$R$ (Å)	$\theta$ (°)	$\alpha$ (°)	Streaks*	Structure
<b>CG(AT)<sub>5</sub></b>	$a = b = 26.54$ $c = 220.53$	39.11	3.26	20	6	220.53	11.7	160.4	60	Yes	Coiled-coil
<b>(CG)<sub>2</sub>(AT)<sub>4</sub></b>	$a = b = 26.37$ $c = 227.5$	38.6	3.22	11	6	227.5	7.4	169	60	No	Coiled-coil
<b>(AT)<sub>5</sub>CG</b>	$a = b = 26.3$ $c = 217$	37.4	3.2	15	6	217	9.7	165	60	Yes	Coiled-coil
<b>(AT)<sub>5</sub>GC</b>	$a = b = 44.8$ $c = 147$	36.8	3.34	0	4	147	-	180	-	Yes	Flipped out cytosine
<b>GC(AT)<sub>5</sub></b>	D34D3: $a = b = 27.8$ $c = 479$	37.5	3.12	24	7	239.5	17.4	159.4	51.4	Yes	Coiled-coil
	D28C21: $a$ and $b$ not determined $c = 480$	37.8	3.15	25	14	480	8.7	160	25.7	Yes	
	D28C1: $a = b \approx 54$ $c \approx 260$ or $500$	-	3.25	22	7 or 6.75	-	-	-	-	Yes	
	D34D6: $c \approx 900$ , $a = b \approx 50$	-	3.25	24	-	-	-	-	-	Yes	
<b>CG(AT)<sub>2</sub>GC(AT)<sub>2</sub></b>	$a = b = 46.5$ $c = 230.5$	38.4	3.2	0-5	6	230.5	-	180	60	Yes	Continuous B-form
<b>(CG)<sub>3</sub>(AT)<sub>3</sub></b>	-	No crystals have been obtained									

\*The intensity of the streaks varies significantly in different crystals of the same sequence.

**P** is the pitch of the coil.

**N** is the number of duplexes per turn.

**l** is the length of one duplex.

**$\beta$**  is the inclination of the axis of the minor helix with respect to the major coil.

**R** is the radius of the cylinder described by the coiled-coil (equation II.4).

**$\theta$**  is the kink angle between two consecutive duplexes (equation II.12);

**$\alpha$**  is the turn angle (equation II.2).

## II.6 DECAMERS

The following decamers have been crystallized:

- **d(CGATATATAT)**, see paragraph II.6.1.
- **d(ATATATATCG)**, see paragraph II.6.2.
- **d(ATATATATATT)**, see paragraph II.6.3.

The sequence d(ATATATATATT), has been included in the decamers section because of its big *c* dimension.

## II.6.1 d(CGATATATAT) d[CG(AT)<sub>4</sub>]

### Introduction

The sequence d(CGATATATAT) was chosen in the effort to better understand the structure of d(ATATATATAT) [d(AT)<sub>5</sub>] (see paragraph II.1).

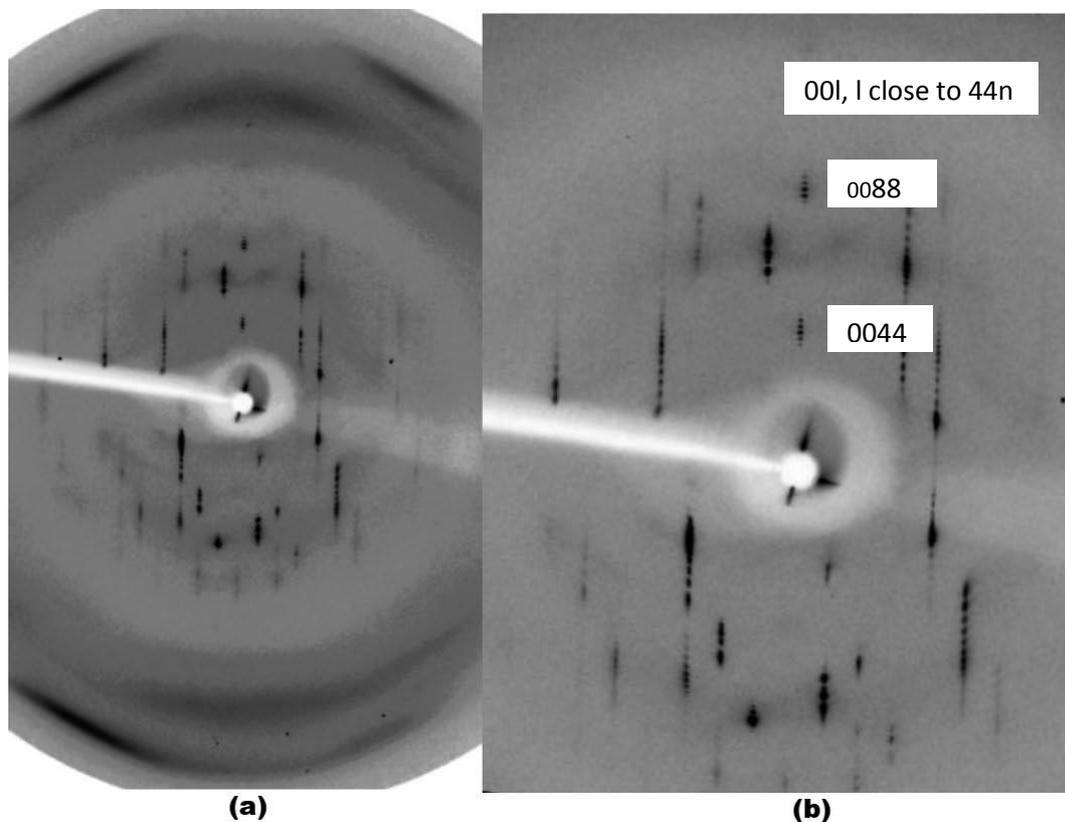
The oligonucleotide was crystallized at 16°. The hanging-drop vapor diffusion technique was used with 2-methyl-2,4-pentanediol as a precipitant. Crystals were obtained from very similar crystallization conditions: 0.3-0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 1 mM Spermine Tetrahydrochloride and 20-50 mM KCl. The concentration of the precipitant was gradually increased, but once the critical precipitant concentration threshold was reached, the crystals grew very fast leading to a high mosaicity and a low resolution of the diffraction pattern.

Diffraction data were measured with cryocooling at 110 K at the beamline BM16 of the European Synchrotron Radiation Facility in Grenoble. Several crystals generated by CG(AT)<sub>4</sub> have been studied, they will be referred to as: “P10A4”, “P8C2”, “P8D41”, “P10B1”, “P9C3” and “P8C63”; each of them is described in the paragraphs below and a summary is presented in Table II.12. The packing features will be discussed in the section “Packing features of CGAT4” (see page 110).

A typical diffraction pattern of the crystal P10B1 is shown in Figure II.6.1. Unlike most of the other sticky-ended sequences studied in this thesis, no streaks have been observed proving a high degree of order in the CG(AT)<sub>4</sub> crystals. Additionally, the well defined sharp stacking reflections confirmed they are not mosaic structures in which crystalline regions coexist with molecules randomly displaced by vertical and rotational movement. However, no crystal with sufficient resolution in order to determine unambiguously its molecular structure was obtained.

Although the majority of the data sets could not be automatically indexed, the unit cell parameters could be manually determined and possible space groups have been suggested. All collected data sets present very similar characteristics. The unit cell presents an unusually large *c* parameter corresponding to about 650-700 Å. The unit cell thus contains stacks of 22-24 decamers.

The sequence  $CG(AT)_4$  gave rise to coiled-coiled structures, where the duplexes axes are inclined with respect to the major coil axis by  $25^\circ$ - $32^\circ$  ( $\beta$ ), depending on the crystal; in the case of crystal P10B1, shown in Figure II.6.1,  $\beta = 28^\circ$ .



**Figure II.6.1.** Oscillation pattern ( $15^\circ$ ) obtained from the crystal P10B1. (a) The long  $c^*$  axis of the unit cell is approximately vertical. The stacking reflections at about  $3.2 \text{ \AA}$  deviate about  $28^\circ$  ( $\beta$ ) from a meridional orientation. (b) Enlarged view of the center of the diffraction. Meridional Bragg reflections close to  $0044n$  reflections (with  $n$  a whole number) are shown.

Given the sequence of the oligonucleotide, the DNA duplexes are organized in a staggered fashion that gives rise to a continuous DNA double helix with nicks in both strands, a model of the organization of the oligonucleotides is shown in Figure II.6.2.



**Figure II.6.2.** A model of organization of the oligonucleotides in the crystal. They form duplexes with sticky ends that generate infinitely long molecules, with staggered nicks in both strands. Due to the presence of the sticky end  $d(CG)$ , the pairing scheme appears to be unique.

Due to the extremely large  $c$  parameter of the unit cell, we encountered the “phi-overlap” problem, which becomes extremely serious when the very long  $c$  axis gets close to being parallel to the beam. In order to avoid this problem, the detector had to be moved as far back as possible and the crystal had to be oriented with the long cell axis roughly (but not perfectly) aligned with the rotation axis of the goniometer. If the long axis and the beam are perfectly aligned, there will be an uncollected region along  $c^*$ . It would also be useful to have a large detector and an X-ray beam with a diameter and a divergence as small as possible.

The characteristics of the crystals of the sequence d(CGATATATAT) are summarized in Table II.12. Similar results were obtained in all cases, except in the P8C63 crystal which presents a larger unit cell.

## **The P10A4 crystal**

### *Crystallization and Data Collection*

A hanging drop of 10  $\mu$ l was prepared with the following composition: 0.33 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 50 mM KCl, 1 mM Spermine Tetrachloride and 10% of MPD. The drop was incubated at 18 °C over 800  $\mu$ l of a reservoir consisting of an aqueous solution with 15% of MPD. Over more than two weeks, the MPD concentration of the reservoir was increased from the 15% to 33%. At this percentage of precipitant, crystals appeared suddenly and grew very rapidly. The crystal, with slightly curved faces and some irregularities, had a trapezoidal shape with the longest dimension of 100  $\mu$ m, the other edges could not be measured. Due to the high sensitivity to the temperature, the crystal had to be handled with extreme care frozen at 4 °C. This crystal will from now on be referred to as “P10A4”.

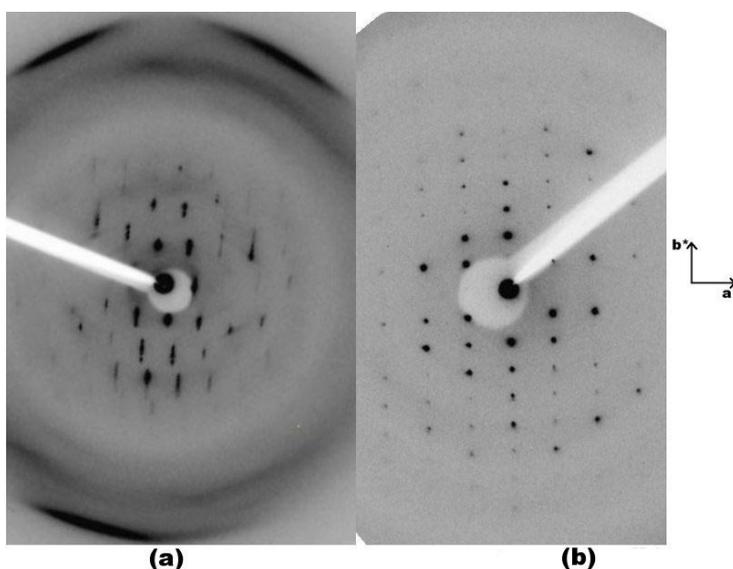
The P10A4 crystal was mounted in a nylon loop at 4 °C and flash frozen in liquid nitrogen. No extra cryoprotectant was used. Data collection was carried on at 100 K at beamline BM16 at the ESRF, Grenoble. Hundred eighty degrees (180°) of oscillation data with a rotation of 3° per image were collected, with a detector distance of 350 cm; one diffraction pattern was also collected at the detector distance of 150 cm, in order to record the stacking reflections of DNA (at about 3.25 Å resolution).

Typical diffraction patterns are shown in Figure II.6.3. The stacking reflections, found at both sides of the meridian, indicate that the DNA duplexes are inclined from the meridional or  $c$  direction of about  $25^\circ$  ( $\beta$ ).

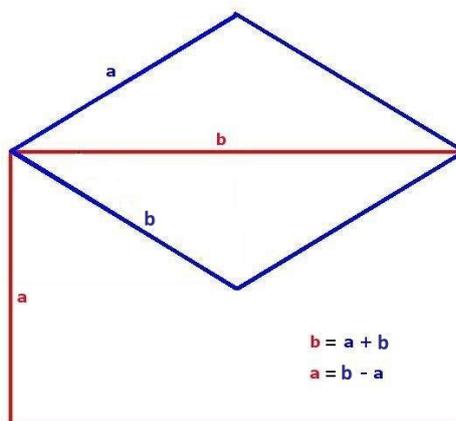
Due to the orientation of the crystal, only  $60^\circ$  have been taken into account during the integration, the remained  $120^\circ$  correspond to a region of the crystal very close to the “equatorial” region of the diffraction, for this reason most of the reflections overlap, as shown in Figure II.6.3 (b) and their indexing and integration is not possible. A striking feature is the absence of  $h00$  reflections.

### *The Diffraction and the Unit Cell*

The diffraction patterns showed anisotropy and very high mosaicity, as shown in Figure II.6.3 (a). The data have been indexed in the monoclinic system with the following unit cell:  $a = 30.84$ ,  $b = 51.45$ ,  $c = 646.31$  and  $\beta = 90.88^\circ$ . The ratio between the  $b$  and  $a$  parameters of the unit cell is equal to 1.67 ( $b/a = 1.67$ ), which is very close to the square root of three ( $\sqrt{3} = 1.73$ ). In the case of the crystal P10A4, the fact that  $a$  multiplied by  $\sqrt{3}$  is only approximately equal to  $b$ , indicates a pseudo-hexagonal symmetry. The conversion of the monoclinic unit cell into a pseudo-hexagonal cell is shown in Figure II.6.4.

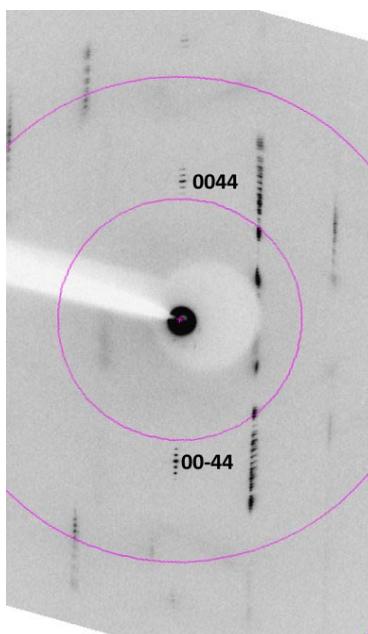


**Figure II.6.3.** Oscillation patterns ( $3^\circ$ ) obtained from the P10A4 crystal. (a) The long  $c^*$  axis is approximately vertical. The prominent stacking reflections at about  $3.25 \text{ \AA}$  deviate about  $25^\circ$  ( $\beta$ ) from a meridional orientation. (b) Oscillation pattern close to the equatorial region. The high inclination of the crystal respect to the rotation axes causes the overlapping of many reflections and does not allow their integration.



**Figure II.6.4.** The monoclinic cell parameters are  $a = 30.84 \text{ \AA}$ ,  $b = 51.45 \text{ \AA}$ ,  $c = 646.31 \text{ \AA}$  and  $\beta = 90.88^\circ$ ; the monoclinic unit cell (red) can be easily converted in a pseudo-hexagonal unit cell (blue) with  $a$  and  $b$  approximately equal to  $30 \text{ \AA}$ . In this case the ratio between  $a$  and  $b$  is not exactly equal to the square root of three ( $\sqrt{3}$ ), in which case it would strongly suggest a hexagonal symmetry.

The  $c$  parameter of  $646.31 \text{ \AA}$  corresponds to a cell containing stacks of 22 decamers ( $c/h$ , with  $h$  equal to  $29.4 \text{ \AA}$ );  $0044n$  reflections, with  $n$  a whole number, are evident along the meridian, as shown in Figure II.6.5. The length of the repeating unit along the  $c$  direction is thus given by the Bragg distance  $d_{0044}$ , which is equal to  $14.7 \text{ \AA}$  and corresponds to half a decamer considering its  $\beta$  inclination of  $25^\circ$ .



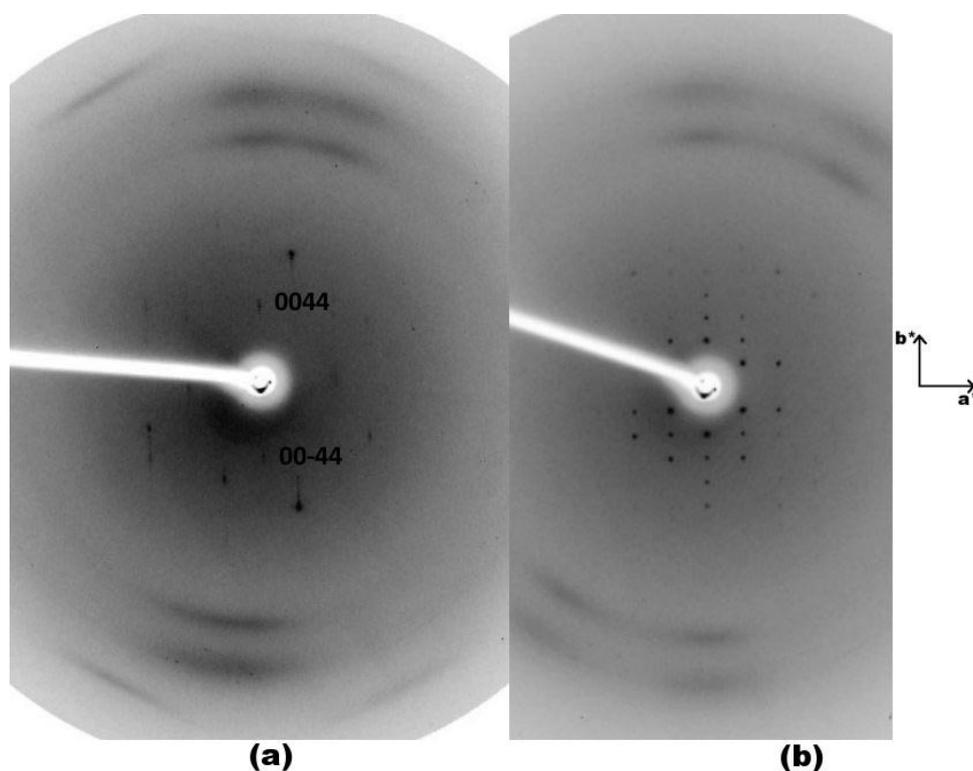
**Figure II.6.5.** Oscillation pattern ( $3^\circ$ ) of P10A4 crystal. The  $c^*$  axis is approximately vertical; the meridional Bragg reflections (with Miller indices close to 0044) are shown.

## The P8C2, the P8D41, the P10B1 and the P9C3 crystals

The P8C2, the P8D41, the P10B1 and the P9C3 crystals are practically isomorphous with the P10A4 crystal (see previous paragraph). Typical diffraction patterns are shown in Figure II.6.6, Figure II.6.7, Figure II.6.8 and Figure II.6.9.

### *The P8C2 crystal*

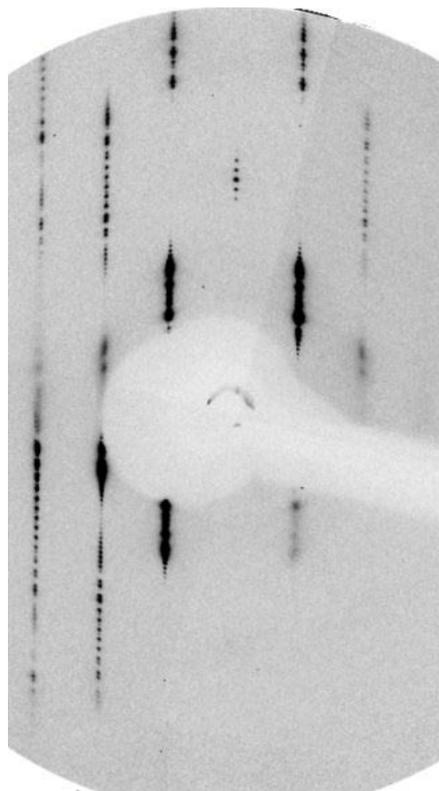
The crystal P8C2 was obtained in the following conditions: 0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 20 mM KCl and 1 mM Spermine Tetrahydrochloride. Typical diffraction patterns are shown in Figure II.6.6.



**Figure II.6.6.** Oscillation patterns ( $5^\circ$ ) of the P8C2 crystal. (a) The long  $c^*$  axis is approximately vertical; the stacking reflections at about  $3.25 \text{ \AA}$  found at both sides of the meridian indicate a  $\beta$  inclination of about  $25^\circ$  of the duplexes with respect to the coiled-coil axis (b) Approximately equatorial region of the crystal, the  $b^*$  and the  $a^*$  axes are shown.

### *The P8D41 crystal*

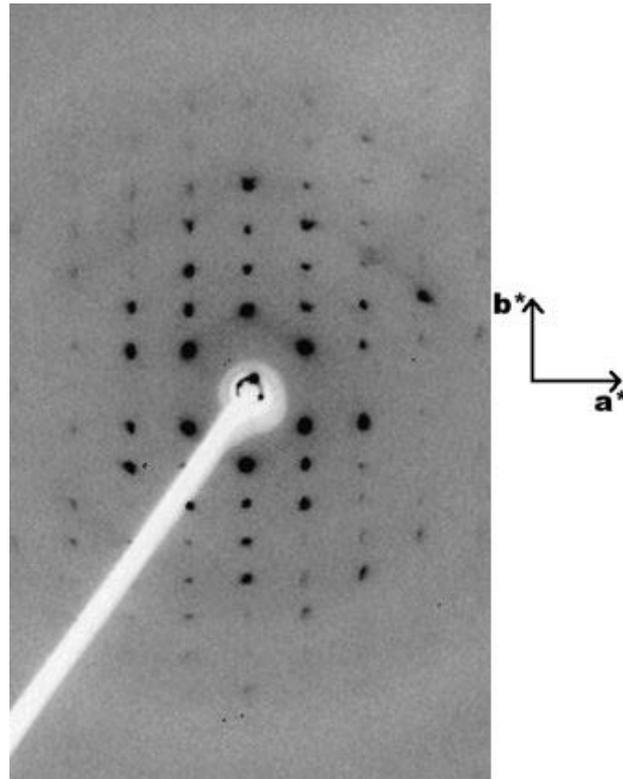
The P8D41 crystal was obtained in the following conditions: 0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 35 mM KCl and 1 mM Spermine Tetrahydrochloride. The data set has been indexed in the following orthorhombic unit cell:  $a = 30.5 \text{ \AA}$ ,  $b = 49.4 \text{ \AA}$ ,  $c = 661 \text{ \AA}$  and  $\beta = 90.6^\circ$ , with 23 duplexes per turn. Typical diffraction pattern is shown in Figure II.6.7.



**Figure II.6.7.** Meridional oscillation pattern ( $5^\circ$ ) of crystal P8D41 Bragg spots with Miller indices close to 0046 are evident, indicating that there are 23 duplexes per turn.

***The P10B1 crystal***

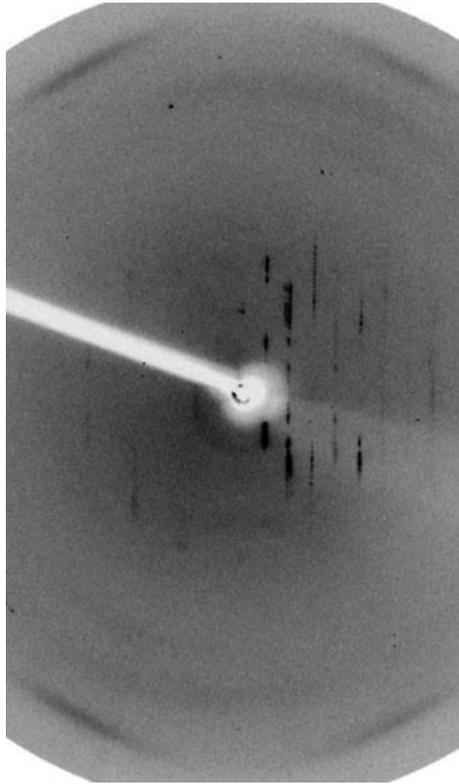
The P10B1 crystal was obtained in the following conditions: 0.4 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 50 mM KCl and 1 mM Spermine Tetrahydrochloride. The meridional diffraction pattern of crystal P10B1 is shown in Figure II.6.1. Equatorial diffraction is shown in the figure below. Note the absence of  $h00$  reflections, which are also absent in the other cases (see Figure II.6.3 b and Figure II.6.6).



**Figure II.6.8.** Oscillation pattern ( $5^\circ$ ) of the equatorial region of crystal P10B1.

*The P9C3 crystal*

The P9C3 crystal was obtained in the following conditions: 0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 20 mM NaCl and 1 mM Spermine Tetrahydrochloride. The data set could not be automatically indexed. The  $\beta$  inclination of the duplexes axes is of about  $30^\circ$ . A typical diffraction pattern is shown in Figure II.6.9.



**Figure II.6.9.** Meridional oscillation pattern ( $5^\circ$ ) of crystal PD9C3; the  $\beta$  inclination of the duplexes with respect to the major coil axis is of about  $30^\circ$ .

## The P8C63 crystal

### *Crystallization and Data Collection*

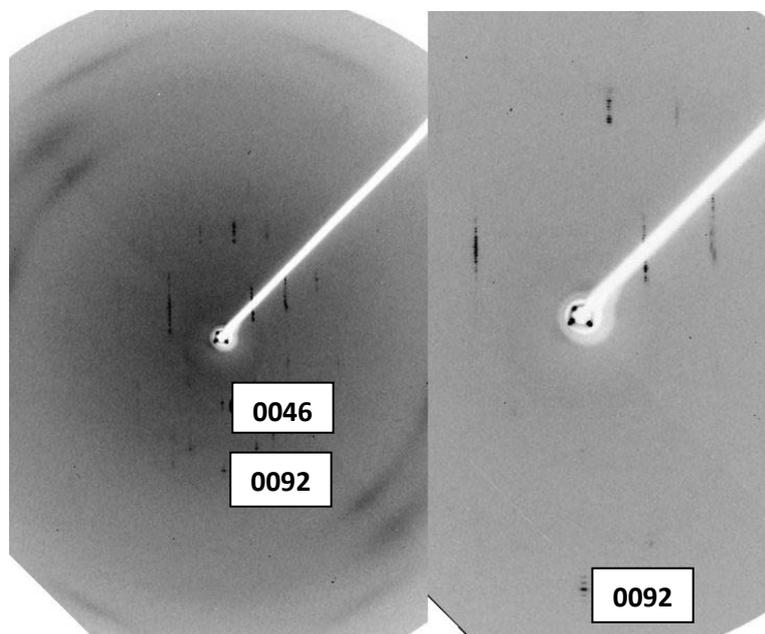
A hanging drop was set up in the following conditions: 0.5 mM DNA, 50 mM NaCacodylate pH 6.5, 20 mM KCl, 1 mM Spermine Tetrachloride and 5%MPD. The hanging drop was incubated at 18 °C over 800 µl of the reservoir consisting of an aqueous solution of MPD 25%. The drop developed precipitate seconds after being mixed only to become clear in some minutes. In a few days, the MPD concentration of the reservoir was increased from the 25% to 35%, at this percentage of precipitant, several needles and small crystals appeared, but no amorphous precipitate was present anymore. In order to eliminate the needles and the small crystals, the MPD concentration was reduced and the temperature was cyclically changed: 37 °C (30 minutes) → 20 °C (1 hour) → 16 °C (days) → 14 °C (days). This cycle was repeated twice and eventually the temperature was maintained at 16 °C. Several small crystals were obtained which presented a “cubic” shape, with one edge of 50 µm, and the diagonal of 60 µm. The crystals have been frozen at 4 °C. This crystal will from now on be referred to as “P8C63”. 180° of oscillation data with a rotation of 1.5° per image were collected at the beamline BM16 of the ESRF in Grenoble.

### *The Diffraction and the Unit Cell*

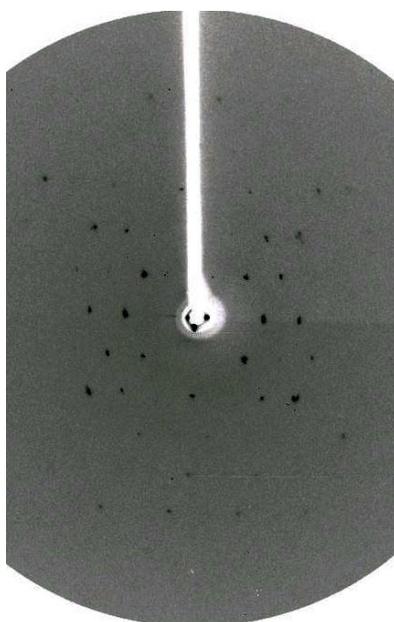
Typical diffraction patterns are shown in Figure II.6.10 and Figure II.6.11. The stacking reflections at about 3.25 Å, found at both sides of the meridian, indicate that the DNA duplexes are inclined about 32° ( $\beta$ ) from the meridional or  $c$  direction (Figure II.6.10).

The data set has been processed in three space groups: Orthorhombic, Monoclinic and Hexagonal (see Table II.11, page 116). The Orthorhombic unit cell ( $a = 55.5$  Å,  $b = 95.6$  Å and  $c = 645.37$  Å,  $C222_1$  space group) is approximately four times bigger than P10A4 cell, but shows some contraction (in fact  $a_{P8C63}/2 \times a_{P10A4} = 0.90$  and  $b_{P8C63}/2 \times b_{P10A4} = 0.93$ ). The volume per base pair of 1985 Å<sup>3</sup> is therefore significantly smaller than in the P10A4 case (see Figure II.6.12). It appears that the various cycles of cooling and heating of the crystals described above have produced a more compact structure. The  $c$  parameter corresponds to a unit cell with stacks of 23 oligonucleotides.

As in P10A4 case, the relation between  $a$  and  $b$  parameters is described by the following equation:  $a \times \sqrt{3} \approx b$ , suggesting a pseudo-hexagonal symmetry (see Figure II.6.4).



**Figure II.6.10.** Oscillation patterns ( $1.5^\circ$ ) of crystal P8C63. The stacking reflections at about  $3.25 \text{ \AA}$  deviate approximately  $32^\circ$  from a meridional orientation. Meridional reflections with Miller indices close to 0046 are indicated.

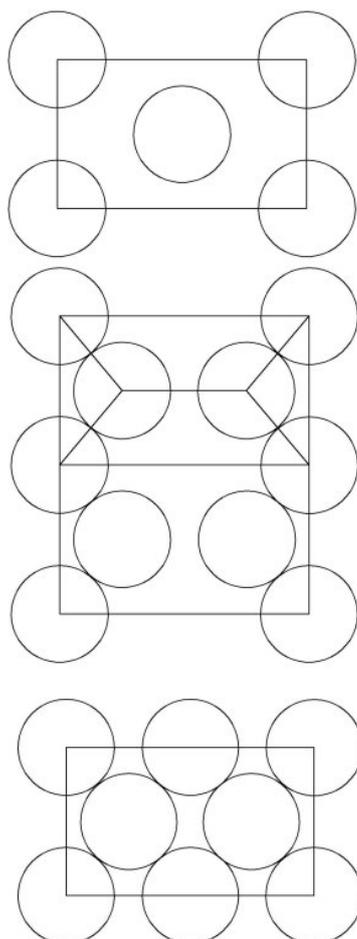


**Figure II.6.11.** Oscillation pattern ( $1.5^\circ$ ) in a region close to the equator. Due to the inclination of the crystal, the view shown has a substantial deviation away from the true equator. A pseudo-hexagonal symmetry is clearly recognizable.

## Packing features of CGATATATAT

The dodecamers that form coiled coils tend to pack in a simple pseudo-hexagonal cell. They also present a high tendency to show screw disorder as revealed by streaks perpendicular to the  $c^*$  axis of the crystal. On the other hand, the decamer CGAT4 does not show streaks and appears to be highly crystalline as demonstrated by well defined Bragg spots. Also the CGAT4 coils are packed in unit cells which contain two or more coils. In this section the packing behavior in such unit cells will be analyzed.

Most of the diffraction patterns, with the exception of P8C63 crystal, could be indexed in a monoclinic unit cell which is quasi orthorhombic. Two, three or four coils could be accommodated in such unit cell; the projection of the cell along the  $c$  direction with the corresponding volume per base pair ( $V_{bp}$ ) is shown in Figure II.6.12.



Two oligonucleotides per unit cell:

$$V_{bp} = \frac{30.8 \text{ \AA} \times 51.4 \text{ \AA} \times 3.3 \text{ \AA} \times \cos 25^\circ}{2} = 2374 \text{ \AA}^3$$

Three oligonucleotides per unit cell:

$$V_{bp} = \frac{30.8 \text{ \AA} \times 51.4 \text{ \AA} \times 3.3 \text{ \AA} \times \cos 25^\circ}{3} = 1583 \text{ \AA}^3$$

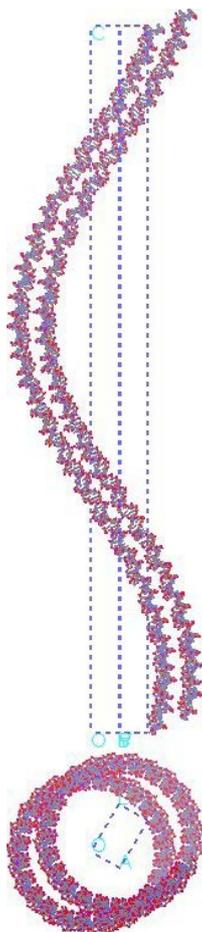
Four oligonucleotides per unit cell:

$$V_{bp} = \frac{30.8 \text{ \AA} \times 51.4 \text{ \AA} \times 3.3 \text{ \AA} \times \cos 25^\circ}{4} = 1187 \text{ \AA}^3$$

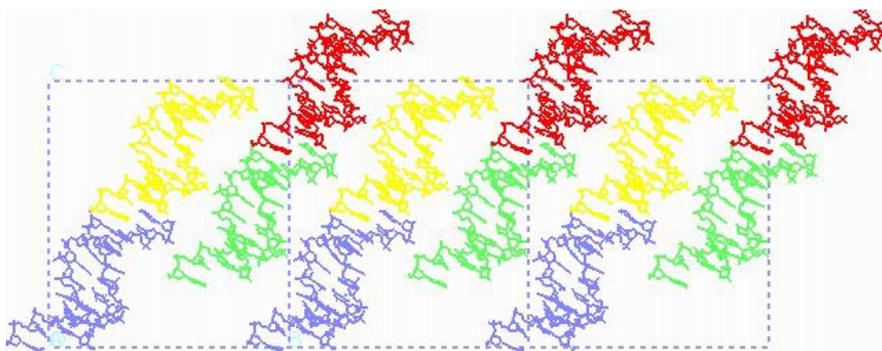
**Figure II.6.12.** Projections along the  $c$  direction of the P10A4 monoclinic unit cell. The correspondent values of volume per base pair ( $V_{bp}$ ) have been calculated.

The presence of four coils could be immediately excluded since it would require a high packing density which is unusual in oligonucleotide crystals. A model with three coils appears to be incompatible with the distribution of  $00l$  reflections, which are only found with even  $l$ . The logical conclusion is to propose a model with two coils in the unit cell as shown in Figure II.6.13. The model has been built with 24 oligonucleotide duplexes in a helical turn of the coil.

A single coil would give rise to  $00l$  reflections with  $l$  equal to a multiple of twenty four ( $l = 24n$ , with  $n$  a whole number). However the experimental diffraction indicated that  $00l$  reflections appeared only in the neighborhood of  $l = 48n$ . In order to achieve this result, the two coils in the unit cell must be displaced by exactly half duplex in the vertical direction, as shown in Figure II.6.14.



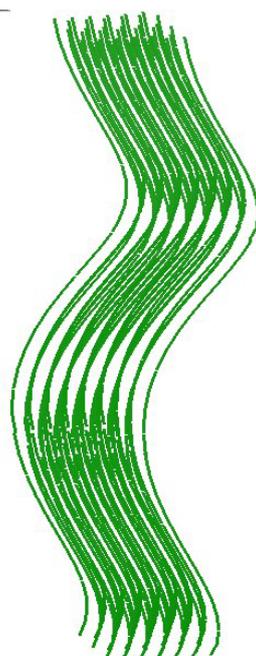
**Figure II.6.13.** Model of two CGATATATAT coils crystallized in the P1 unit cell. Each coil has twenty four duplexes per turn. The duplexes interact through their terminal CG bases. The unit cell is shown at an angle in order to demonstrate that there are no clashes among both coils.



**Figure II.6.14.** A two duplex region of six neighbor coils is shown. Neighbor duplexes are displaced by one half of their length in the vertical direction. As a result the  $00l$  reflections are only visible when  $l = 48n$ , with  $n$  a whole number. Each individual duplex is presented in a different color.

The repeating unit along the  $c^*$  axis is then half a duplex, so that now only  $00l$  reflections with  $l = 48n$  should be expected, as observed in the experimental diffraction. A simplified view of a group of such duplexes is presented in Figure II.6.15.

Another feature of the diffraction patterns presented above is the presence of a few  $00l$  reflections close to the main meridional reflections. Such reflections always have an even Miller index  $l$ , which might indicate the presence of a screw axis in the  $c^*$  direction. The presence of these few spots also indicates a slight geometrical distortion/displacement of the coils.

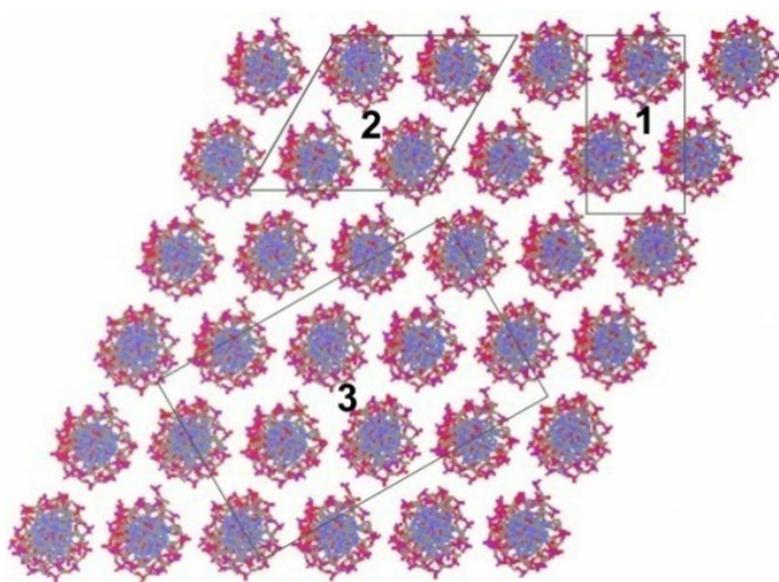


**Figure II.6.15.** Schematic representation of a bundle of coiled coils.

A model in the monoclinic  $P2_1$  space group has been built, but the calculated diffraction patterns showed important disagreement with the experimental observations. Therefore a model in a  $P1$  unit cell has been built, with the following cell dimensions  $a = 30.84 \text{ \AA}$ ,  $b = 51.45 \text{ \AA}$ ,  $c = 646.31 \text{ \AA}$ ,  $\beta = 90.88^\circ$  and  $\alpha = \gamma = 90^\circ$  and two independent coils in the unit cell.

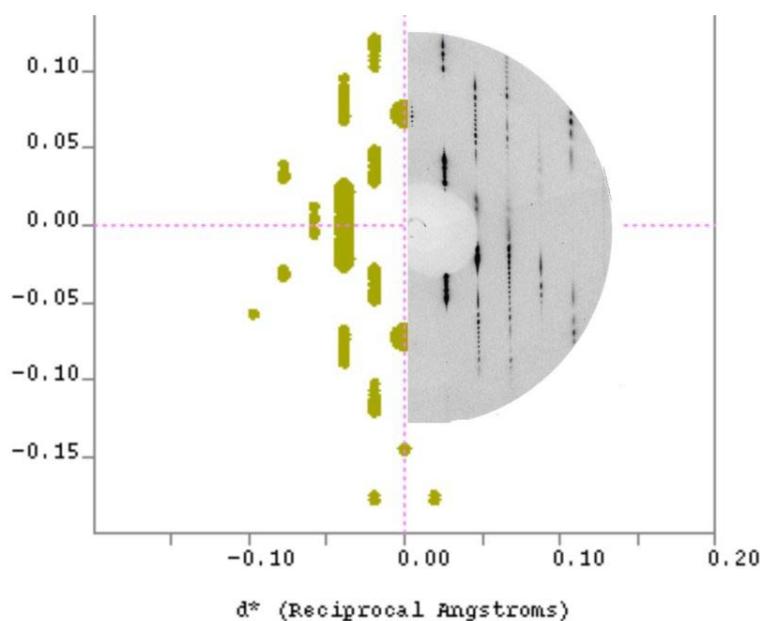
The staggered conformation shown in Figure II.6.14 is likely to be maintained by interactions between neighbor coils. The intermolecular forces which originate such a mutual orientation of the two neighbor coils are not obvious. It can be only suggested that they are due to electrostatic effects which involve the counterions present in the crystallization buffer.

The distance between neighbor coils slightly varies in different regions, as it can be appreciated in Figure II.6.13. A section amongst a group of coils is shown in Figure II.6.16. Although the distance variation could generate local distortions, no serious clashes are found among neighbor duplexes and the phosphate-phosphate distances are always larger than  $8 \text{ \AA}$ .

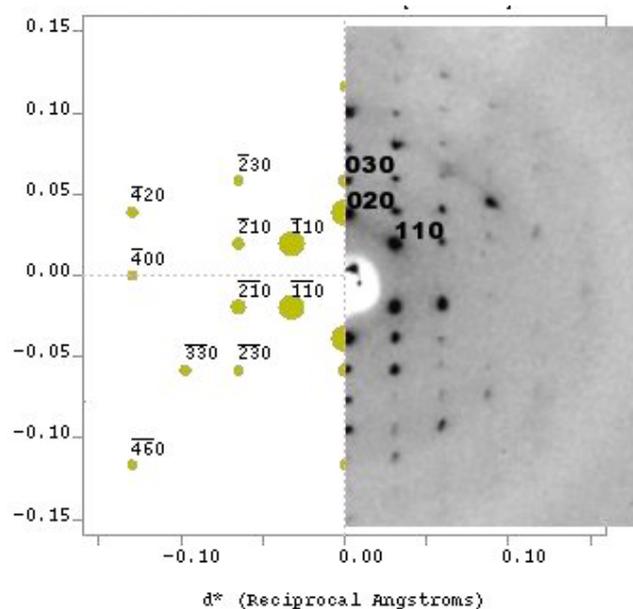


**Figure II.6.16.** Section through a region of the P8C63 pseudo-hexagonal  $P2_1$  crystal described in the text. Neighbor coils do not show contacts among themselves. A practically identical result is obtained from the simpler  $P1$  unit cell shown in Figure II.6.13. Different unit cells are indicated in the figure as described in the text. Cell 1 is the  $P1$  cell of most crystals. Cell 2 is the  $P2_1$  cell assumed for P8C63. Cell 3 is the larger cell also possible for P8C63.

The diffraction given by the model in the P1 unit cell and shown in Figure II.6.13 has been simulated with the program Cerius 2 (Accelrys, San Diego). Comparisons with the experimental patterns are given in Figure II.6.17 and Figure II.6.18. The overall agreement is quite satisfactory, in particular in the  $100$  zone. Diffraction in the  $001$  zone showed a more apparent disagreement, but it should be noted that diffraction in this zone is comparatively rather weak, only the  $110$  and  $020$  spots are strong. Small changes in the translational and rotational relative positions of the two coils in the unit cell give rise to noticeable changes in this zone.



**Figure II.6.17.** Comparison of the simulated diffraction in the  $100$  zone of the P1 cell with an oscillation pattern ( $5^\circ$ ) obtained from the P8D41 crystal in which the same zone was approximately recorded.



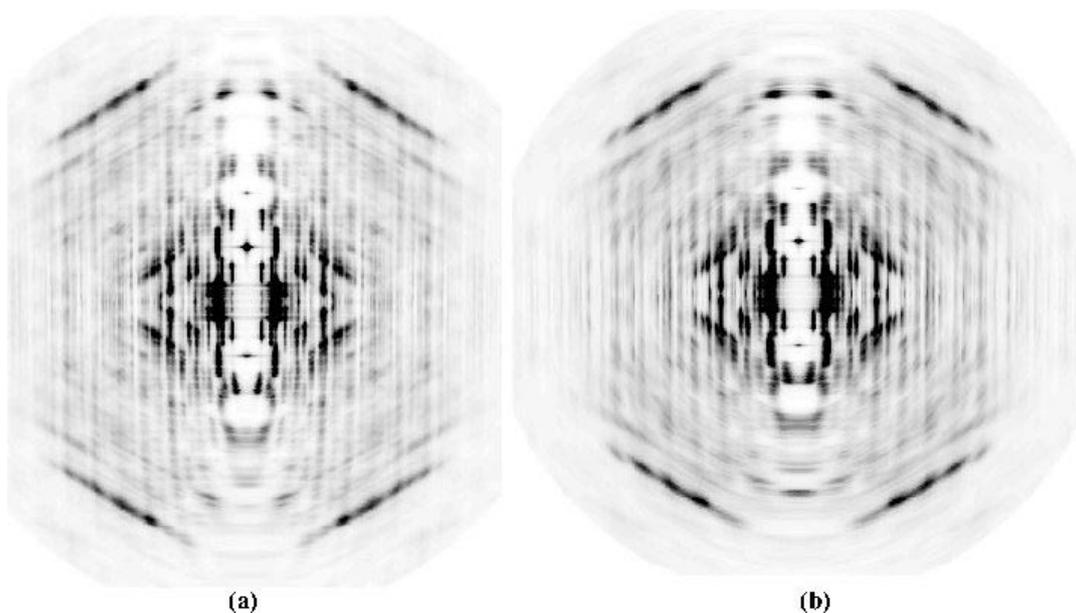
**Figure II.6.18.** Comparison of the simulated diffraction in the 001 zone of the P1 cell with an oscillation pattern ( $3^\circ$ ) close to this region. The comparison is only approximate due to the high inclination of the crystal, so that  $hkl$  reflections with  $l$  small are superimposed to the  $hk0$  reflections. A striking feature of diffraction in this region is the absence/low intensity of  $h00$  reflections in both the simulated and the experimental pattern. See text for further details.

In the case of the P8C63 crystal a larger unit cell was found. The available data have been processed in three different space groups; the results are shown in Table II.11. The  $P6_3$  possibility could be immediately excluded, since, due to the size of the unit cell, it should accommodate four coils, which is not compatible with a hexagonal symmetry. Although the best processing statistics have been obtained in the  $C222_1$  space group, a centered space group did not allow placing the coils in a pseudo-hexagonal arrangement as it was expected. Although another orthorhombic space group could not be excluded, it was not possible to find a model which could give an adequate simulated diffraction.

**Table II.11.** Dataset statistics for crystal P8C63. In parenthesis are the values for the high resolution shell, its lower limits are shown in parenthesis in the resolution line.

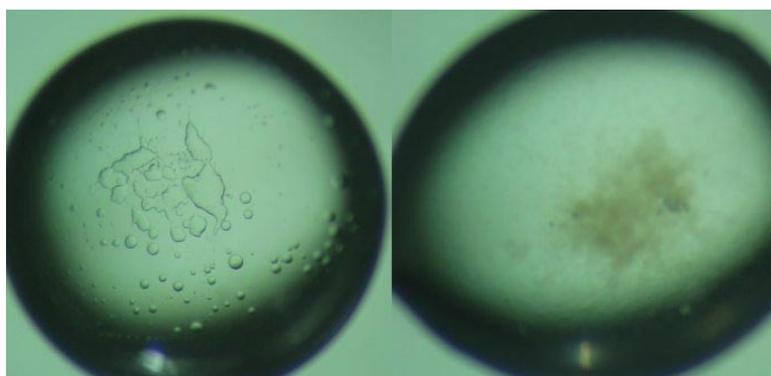
Crystal	P8C63		
Sequence	d(CGATATATAT)		
Wavelength (Å)	0.9794		
Detector distance (cm)	450		
Oscillation range(°)	1.5		
$\beta$ (°)	32		
Rise (Å)	3.25		
N	23		
Streaks	No		
Space group	<b>C222<sub>1</sub></b>	<b>C2</b>	<b>P6<sub>3</sub></b>
Unit cell	$a = 58.921$ $b = 98.454$ $c = 645.37$	$a = 57.489$ $b \approx 645$ $c =$ $56.217$ $\beta \approx 120^\circ$	$a = b = 57.38$ $c = 645$
Resolution	8 (8.10)	10 (10.11)	8 (8.10)
Total Reflections	3374	8754	7035
Unique reflections	2125	4142	1131
$R_{\text{int}}$ (%)	8.63 (15.74)	9.0 (29)	21 (38)
$I/\sigma(I)$	5.25 (2.86)	2.68 (1.94)	3.64 (1.37)
Completeness	91.2 (93.0)	59.2 (73.1)	62.4 (66.7)

Eventually, a model in the  $P2_1$  space group was used, with a unit cell similar to the  $P6_3$  one:  $a$  and  $b$  equal to  $57.4 \text{ \AA}$ ,  $c$  equal to  $658 \text{ \AA}$ , and  $\gamma$  to  $120^\circ$ . With respect to the processing, the  $c$  value is larger ( $658 \text{ \AA}$  instead of  $645 \text{ \AA}$  reported in Table II.11); in fact the model presents twenty four duplexes per turn instead of twenty three. The asymmetric unit contained the two independent coils found in the smaller unit cell (see Figure II.6.13). The packing of such a unit cell is shown in Figure II.6.15. A section through the model is represented in Figure II.6.16, where the  $P2_1$  unit cell is indicated, as well as a possible orthorhombic unit cell. The calculated fiber diffraction pattern for this model is practically identical to the diffraction given by the model in the smaller  $P1$  cell and shown in Figure II.6.19.



**Figure II.6.19.** Comparison of the calculated fiber diffraction patterns of the models for the small P1 cell (a) and the larger P21 cell (b) assumed for the P8C63 crystal.

In the effort of understand the structure generated by this oligonucleotide, the d(CGATATATAT) fragment has been studied by Dr C. Gonzales by NMR. This part of the work is still in a preliminary stage, but a different result was obtained in a solution with pH below 5. Crystallization trials have been set up at pH 4 and 5. Only amorphous precipitate and phase separation have been obtained (Figure II.6.20).



**Figure II.6.20.** Phase separation and amorphous precipitate obtained at pH 4.5 (a) and 5.0 (b).

### Summary of the CG(AT)<sub>4</sub> structure

- The sequence CG(AT)<sub>4</sub> crystallizes with monovalent cations.
- It generates coiled-coils with 22-24 oligonucleotide duplexes per turn.
- It generates a coiled coil with an inclination  $\beta$  of about 30°.
- The diffraction patterns are characterized by sharp stacking reflections and the absence of streaks: it has higher crystallinity than d(AT)<sub>5</sub>, which usually presents streaks in its diffraction patterns.
- The maximum resolution of about 5 Å is not sufficient to determine the molecular structure of the crystals.
- In analogy with the structure of CG(AT)<sub>5</sub> (see paragraph II.5.1), the hydrogen bonds in the (AT)<sub>4</sub> part of the structure might be in Hoogsteen conformation, however, there is no evidence of this hypothesis.

The characteristics of CG(AT)<sub>4</sub> crystals are summarized in Table II.12.

**Table II.12.** Characteristics of CG(AT)<sub>4</sub> crystals, not all the diffracting crystals are shown.

Crystal	Unit cell (Å)	Space group	Osc. range	$\beta$	<i>N</i>	Streaks	Crystallization conditions
<b>P8D41*</b>	$a = 30.5$ $b = 49.4$ $c = 661$ $\beta = 90.6^\circ$	P1	5°	25°-28° approx	23	No	0.5 mM DNA, 50 mM NaCac. pH 6.5, 35mM KCl, 1 mM Spermine and MPD 28%
<b>P10A4*</b>	$a = 30.8$ $b = 51.45$ $c = 646.3$ $\beta = 90.8^\circ$	P1	3°	25°	22	No	0.5 mM DNA, 50 mM NaCac. pH 6.5, 50mM KCl, 1mM Spermine and MPD 33%
<b>P8C63</b>	$a = 55.5$ $b = 95.6$ $c = 645.4$ $\alpha = \beta = \gamma = 90^\circ$	C222 <sub>1</sub> P2 <sub>1</sub>	1.5°	32°	23	No	0.5 mM DNA, 50 mM NaCac. pH 6.5, 20mM KCl, 1mM Spermine and MPD 25%

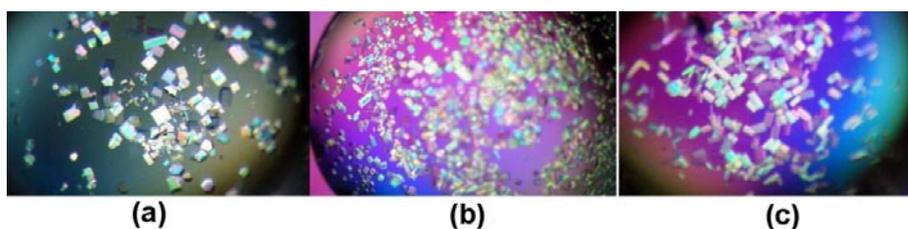
\*Crystals P8D41 and P10A4 are practically isomorphous.

## II.6.2 d(ATATATATCG) d[(AT)<sub>4</sub>CG]

The sequence d(AT)<sub>4</sub>CG was chosen as a variation of d(CGATATATAT), to see if the change of the sticky end position could generate a different structure.

The oligonucleotide was crystallized at 13 °C. The hanging-drop vapor diffusion technique was used with 2-methyl-2,4-pentanediol as a precipitant. The crystals (Figure II.6.21) were obtained in very similar conditions: 0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 1-1.5 mM Spermine tetrahydrochloride, 1 mM CoCl<sub>2</sub> and 10 mM MgCl<sub>2</sub>, with or without 16.7 mM Thymidine (the Thymidine had no appreciable influence on the diffraction of the crystals obtained). The crystals will be referred to as “D35A1-2”, “D35A1-3”, “D35A1-4”, “D35A4” and “D35A5”.

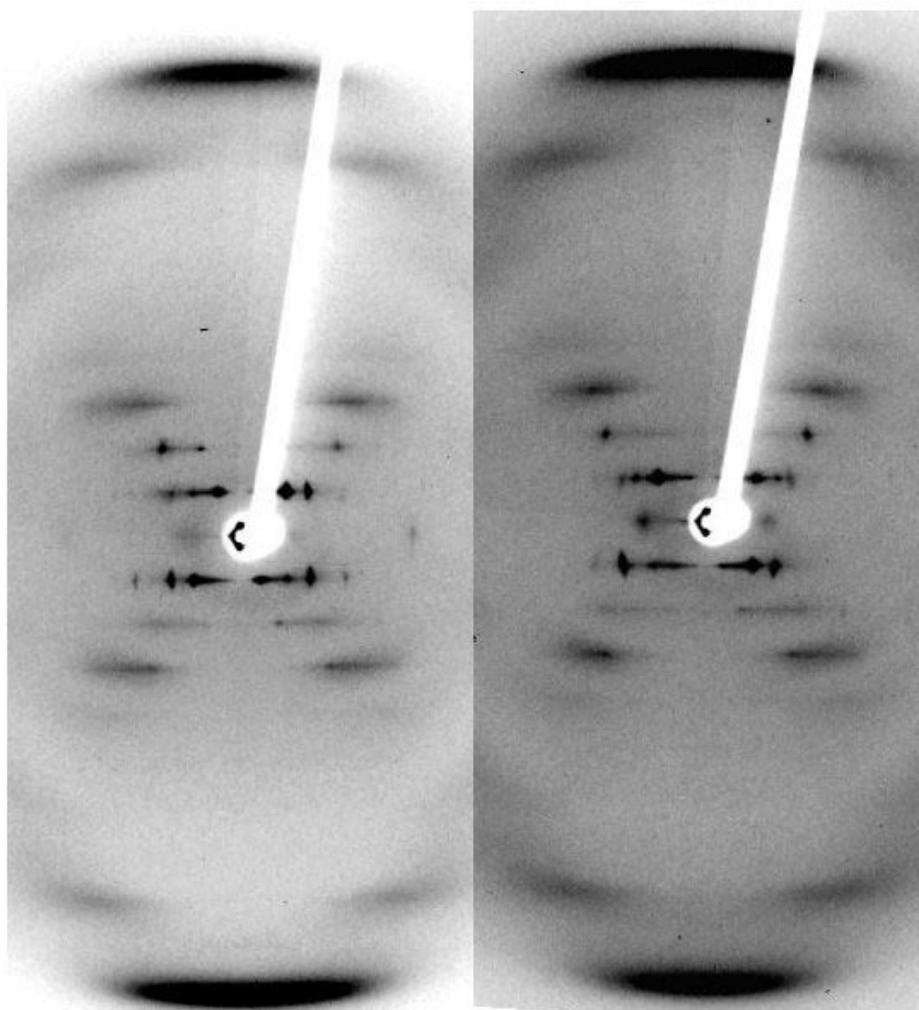
All the crystals gave rise to fiber-like diffraction patterns with very few or no Bragg reflections. Typical diffraction patterns are shown in Figure II.6.22 and Figure II.6.23. The crystals used for the data collection grew in the same drop, but were frozen at different times: D35A1-2 crystal was frozen four weeks after the drop was set up (Figure II.6.22); D35A1-3 and D35A1-4 crystals were frozen four months after the drop was set up (Figure II.6.23 and Figure II.6.24). The d(AT)<sub>4</sub>CG sequence generates continuous coils, in the form of a standard B-DNA or of a coiled-coil.



**Figure II.6.21.** Microscope photographs of d(ATATATATCG) crystals. They have been obtained in the following conditions: (a) D35A1 crystal: 0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 10 mM MgCl<sub>2</sub>, 1 mM CoCl<sub>2</sub> and 1.0 mM Spermine tetrahydrochloride. (b) D35A4 crystal: 0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 10 mM MgCl<sub>2</sub>, 1 mM CoCl<sub>2</sub> and 1.5 mM Spermine tetrahydrochloride. (c) D35A5 crystal: 0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 10 mM MgCl<sub>2</sub>, 1 mM CoCl<sub>2</sub>, 1.0 mM Spermine tetrahydrochloride and 16.7 mM Thymidine.

The diffraction patterns shown in Figure II.6.22 are typical of standard B-form DNA. The stacking reflections at approximately 3.25 Å resolution are found along the meridian, indicating that the continuous coil forms infinite straight columns of standard B-form DNA.

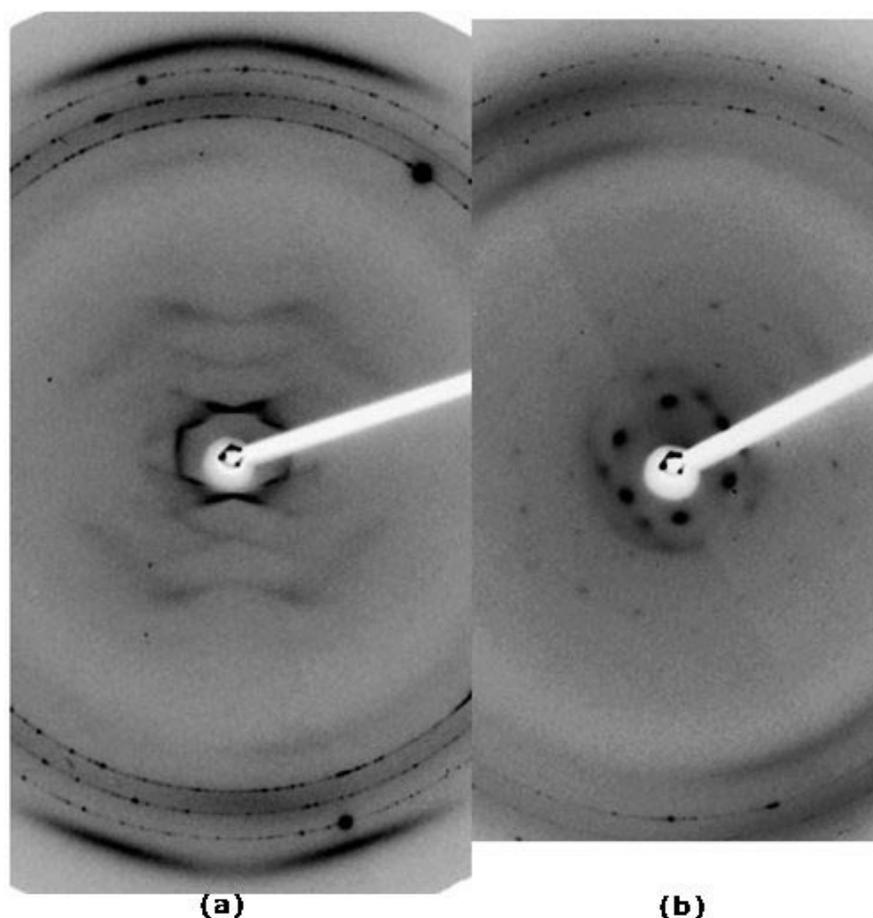
In the diffraction patterns shown in Figure II.6.23 and Figure II.6.24, the stacking reflections at 3.2 Å are found at both sides of the meridian, indicating that the DNA duplexes are inclined a  $\beta$  angle of 12° to 25° (depending on the crystal) from the meridional or  $c$  direction.



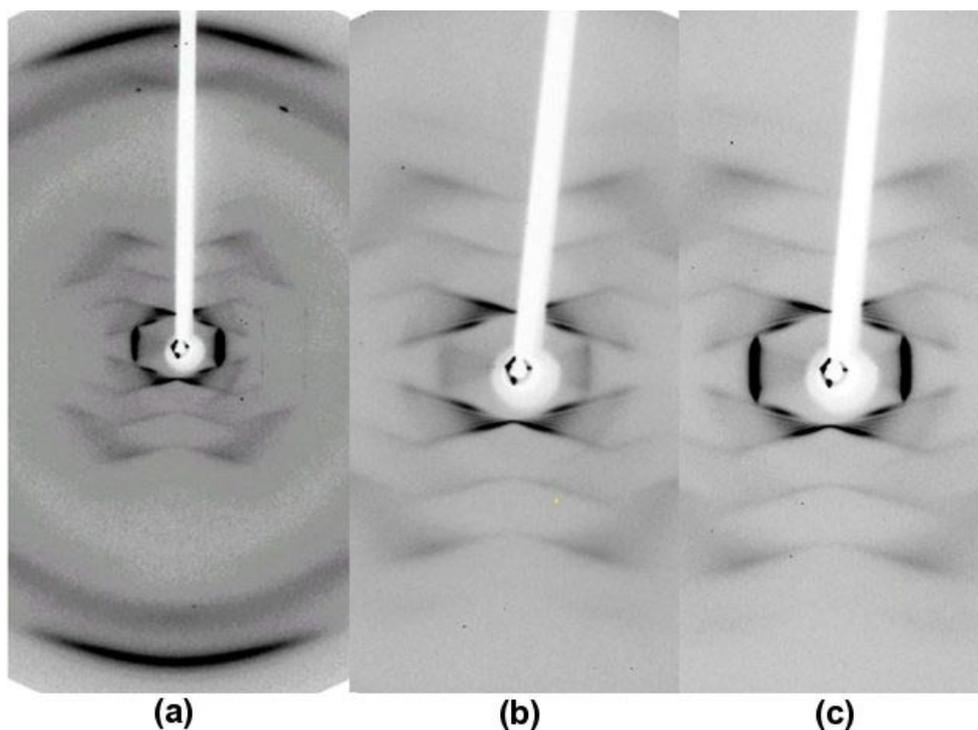
**Figure II.6.22.** Oscillation patterns (3°) of crystal D35A1-2 (frozen four weeks after the drop was set up). The two diffraction patterns are rotated 90° one respect to the other. This is a fiber-like diffraction of a B-form DNA, with a maximum intensity at about 3.2 Å resolution, corresponding to the stacking reflection. The stacking reflections occur at the 10<sup>th</sup> DNA layer line and correspond to a pseudo-continuous DNA helix of ten stacked base pairs and a pitch  $P$  of about 32 Å (B-form DNA).

The data set could not be indexed, but a pseudo-hexagonal symmetry could be recognized from the diffraction patterns (see Figure II.6.23 (b) and Figure II.6.24). Bragg reflections of  $10l$  order appeared only every  $60^\circ$ , suggesting a trigonal or pseudo-hexagonal symmetry.  $a$  and  $b$  unit cell parameters have been measured and correspond to about  $27 \text{ \AA}$  ( $a = b = d_{100}/\sin 60$ ).

The enlarged view of the meridional oscillation pattern shown in Figure II.6.25, reveals the existence of layer lines, whose spacing of about  $920 \text{ \AA}$  corresponds to the unit cell  $c$  parameter.

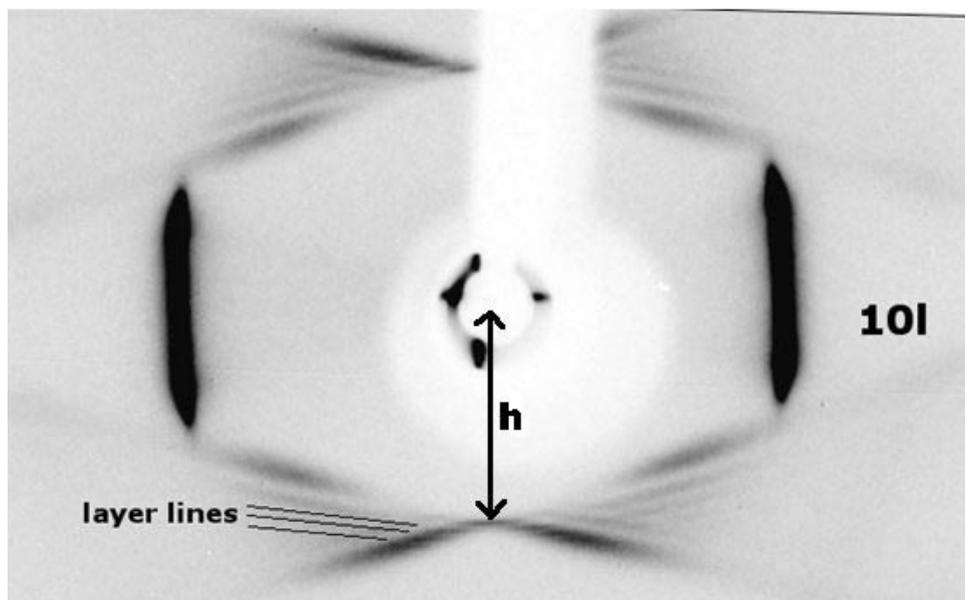


**Figure II.6.23.** Oscillation patterns ( $5^\circ$ ) of crystal D35A1-3 (frozen four months after the drop was set up). This is a fiber-like diffraction of a coiled-coil. The stacking reflections at  $3.25 \text{ \AA}$  found at both sides of the meridian, indicate that the DNA duplexes are inclined about  $17^\circ$  ( $\beta$ ) from the meridional or  $c$  direction. The two diffraction patterns are rotated of  $90^\circ$  one respect to the other: (a) approximate fiber meridional projection; (b) approximate fiber equatorial projection. The latter projection is visible due to the high inclination (of about  $70^\circ$ ) of the rotation axis with respect to the meridional or  $c^*$  axis. In (b) a pseudo-hexagonal symmetry can be recognized; it reflects the packing of the DNA columns projected onto the equatorial plane.



**Figure II.6.24.** Oscillation patterns ( $15^\circ$ ) of crystal D35A1-4; starting angles: (a)  $0^\circ$ , (b)  $60^\circ$  and (c)  $120^\circ$ ; the  $c^*$  axes is well oriented along the spindle axes, strong 10l reflections appear every  $60^\circ$ , suggesting a hexagonal symmetry.  $a$  and  $b$  unit cell parameters correspond to about  $27 \text{ \AA}$  ( $a = d_{100}/\sin 60^\circ$ ,  $d_{100}$  corresponds to about  $23.5 \text{ \AA}$ ).

The height  $h$  occupied by one duplex in the unit cell along the  $c$  direction can be measured from the first meridional spot in the diffraction pattern (see Figure II.6.25); it corresponds to about  $31 \text{ \AA}$ . The length  $l$  of the DNA duplex corresponds to  $32.1 \text{ \AA}$  ( $l = h/\cos \beta$ ). Therefore there are 30 oligonucleotides per unit cell along the  $c$  direction ( $c/h = 923 \text{ \AA}/31 \text{ \AA}$ ).



**Figure II.6.25.** Enlarged view of the oscillation pattern reported in Figure II.6.24 (c). The spacing between layer lines of about 920 Å corresponds to the unit cell  $c$  parameter. The height  $h$  occupied by one duplex along the  $c$  direction (measured from the first meridional spot) corresponds to about 31 Å. There are 30 duplexes per unit cell along the  $c$  direction ( $c/h$ ). Diffraction with 101 Miller indices is also evident.

Given the sequence of the oligonucleotide, the DNA duplexes are organized in a staggered fashion and thus give rise to a continuous DNA double helix with nicks in both strands, as shown in Figure II.6.26.

The duplexes axes can be parallel to the continuous coil axis or can deviate  $12^\circ$  to  $25^\circ$  from it. With these data it is not possible to formulate any hypothesis to explain the existence of the two different structures. In Table II.13 a summary of all diffracting crystals is reported.



**Figure II.6.26.** Model of organization of the oligonucleotides in the crystal. They form duplexes with sticky ends that generate infinitely long molecules, with staggered nicks in both strands.

**Table II.13.** Characteristics of the diffracting crystals of the sequence (AT)4CG. No Bragg reflections have been found in any of these diffraction patterns.

Crystal	Unit cell	Osc. range	$\beta$	Rise (Å)	Comments	Crystallization conditions
D34B3-2 D34B3-4	-	+15°	0°	3.25	Pseudo-continuous coil, probably B-form.	0.5 mM DNA, 50 mM NaCac pH6.5, 10 mM MgCl <sub>2</sub> , 1 mM CoCl <sub>2</sub> and 1.0 mM Spermine.
D35A1-1 D35A1-2	-	+15°	0°	3.2	Pseudo-continuous coil, probably B-form, crystal frozen four weeks after drop D35A1 was set up.	0.5 mM DNA, 50 mM NaCac pH6.5, 10 mM MgCl <sub>2</sub> , 1 mM CoCl <sub>2</sub> and 1.0 mM Spermine.
D35A1-3	Trigonal $a = b \approx 27 \text{ \AA}$ $c \approx 920 \text{ \AA}$	+5°	17°	3.25	Coiled-coil, crystal frozen four months after drop D35A1 was set up. Thirty duplexes per unit cell.	0.5 mM DNA, 50 mM NaCac pH6.5, 10 mM MgCl <sub>2</sub> , 1 mM CoCl <sub>2</sub> and 1.0 mM Spermine.
D35A1-4	-	+15°	12°	3.25	Coiled-coil, crystal frozen four months after drop D35A1 was set up.	0.5 mM DNA, 50 mM NaCac pH6.5, 10 mM MgCl <sub>2</sub> , 1 mM CoCl <sub>2</sub> and 1.0 mM Spermine.
D35A5-1 D35A5-2	-	+5°	25°	3.25	Coiled-coil, diffuse diffraction, only stacking visible.	0.5 mM DNA, 50 mM NaCac pH6.5, 10 mM MgCl <sub>2</sub> , 1 mM CoCl <sub>2</sub> , 1.0 mM Spermine and 16.7 mM Thymidine.

### II.6.3 d(ATATATATATT) d(AT)<sub>5</sub>T

The sequence (AT)<sub>5</sub>T was chosen to see if the extra thymine could be found in the neighbor groove of the (AT)<sub>5</sub> part, as previously found for the sequence (AT)<sub>3</sub> (Abrescia *et al.*, 2004). The sequence (AT)<sub>5</sub>T is included in the decamers section because of the big dimension of the unit cell *c* parameter.

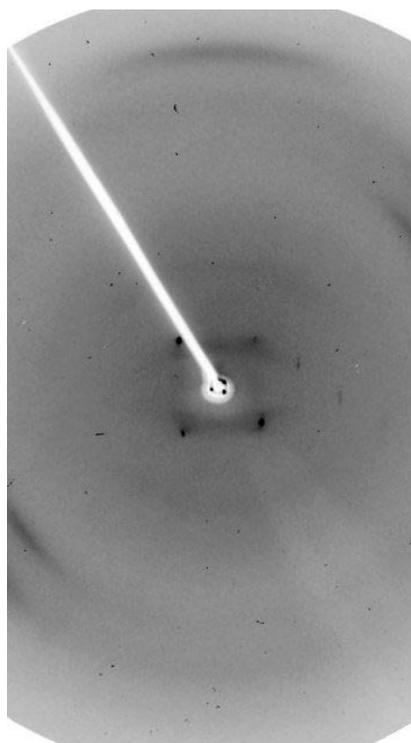
All the crystals obtained from the sequence (AT)<sub>5</sub>T had curved sides and, often, a gelatinous consistency (see Figure II.6.27).



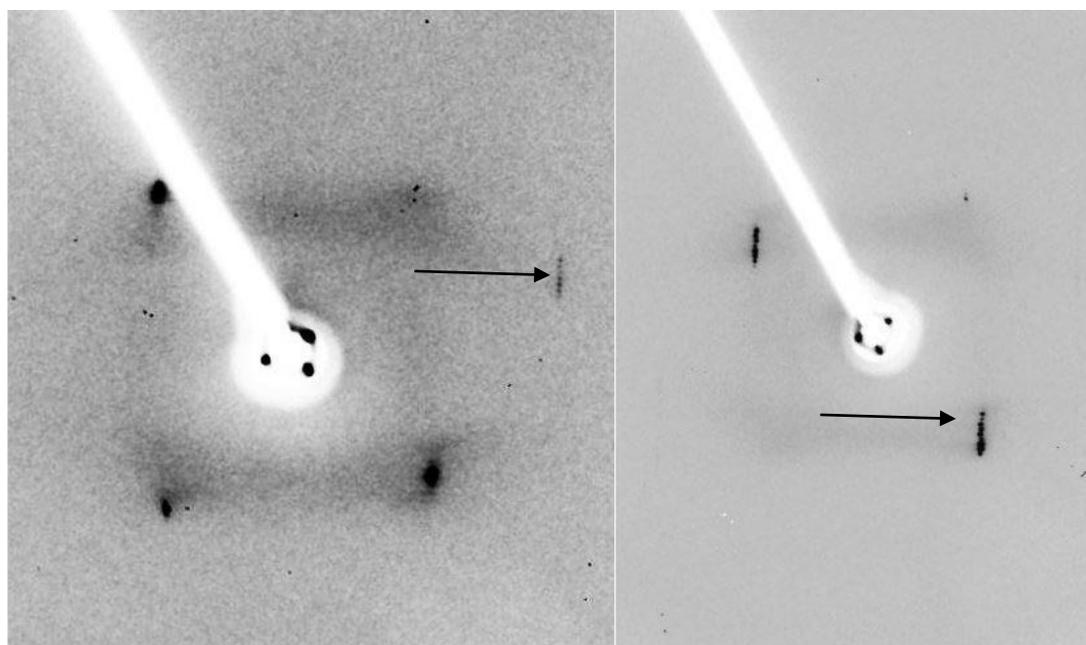
**Figure II.6.27.** Three microscopic photographs of crystals of the sequence (AT)<sub>5</sub>T. All of them presented round edges and were often gelatinous.

Only one crystal could be tested in an X-ray diffraction experiment. The crystal was obtained in the following conditions: 0.5 mM DNA duplex, 25 mM NaCacodylate pH 6.5, 10 mM MgCl<sub>2</sub>, 16.7 mM Thymidine, 4-aminophenylsulphone and 1.5 mM Spermine Tetrahydrochloride.

A typical diffraction pattern is shown in Figure II.6.28. The sequence probably generates a pseudo-continuous coil. The stacking reflection at about 3.25 Å has the form of an arc, suggesting a slight  $\beta$  inclination of about 5° of the duplexes with respect to the continuous coil axis. The *c* parameter of the unit cell has been manually measured and is about 545 Å (see Figure II.6.29).



**Figure II.6.28.** Oscillation pattern ( $5^\circ$ ) obtained from a d(ATATATATATT) crystal. The stacking reflections at about  $3.25 \text{ \AA}$  have the form of an arc; the sequence probably forms a pseudo-continuous coil.



**Figure II.6.29.** Two enlarged views of  $(AT)_5T$  diffraction patterns. The Bragg spots (indicated by the arrows) allow an estimation of the  $c$  parameters of the unit cell, which corresponds to about  $545 \text{ \AA}$ .

## II.7 OCTAMERS

The following octamers have been crystallized:

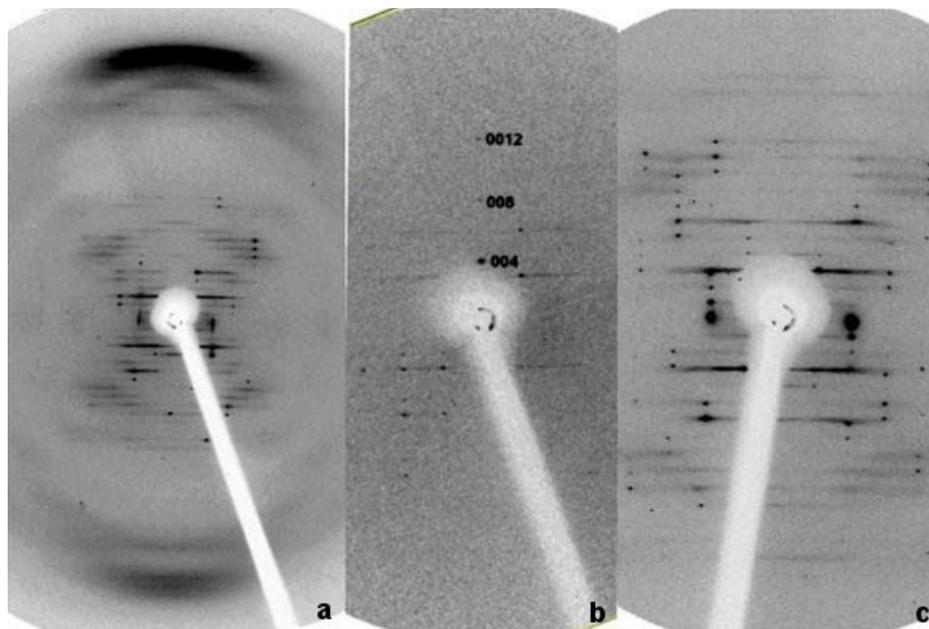
- **d(CGTATATA)**, see paragraph II.7.1.
- **d(CGATATAT)**, see paragraph II.7.2.
- **d(ATATATCG)**, see paragraph II.7.2.

The sequences d(CGTATATA) and d(CGATATAT) are isomorphous B-forms DNA, while d(ATATATCG) is probably Hoogsteen DNA.

## II.7.1 d(CGTATATA) d[CG(TA)<sub>3</sub>]

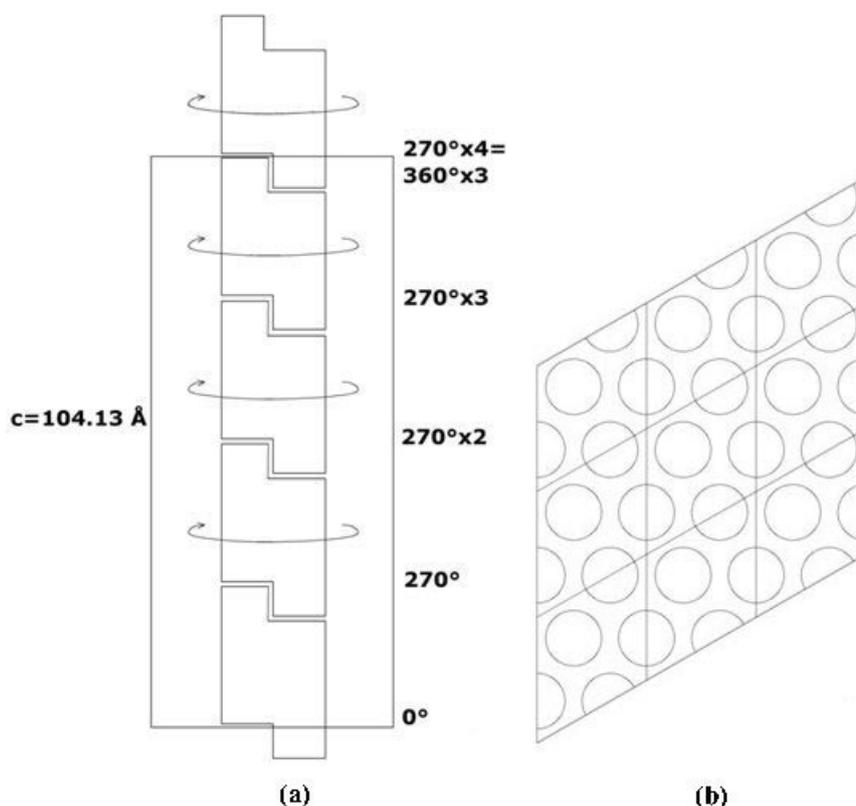
The sequence CG(TA)<sub>3</sub> was crystallized at 20° using the hanging-drop vapor diffusion technique and 2-methyl-2,4-pentanediol as precipitant. Crystals were obtained in the following conditions: 0.4-0.5 mM DNA duplex, 20 mM NaCacodylate pH 6.5 and 1 mM Spermine Tetrahydrochloride, with the following additives: 21 mM Leucinamide, 16.7 mM Thymidine, 4-aminophenylsulfone. Several big needles appeared only one day after the drop was set up. The temperature was first increased to 37 °C in order to melt the crystals and then gradually lowered to 20 °C→16 °C→13 °C→4 °C. The quality of the crystals could not be significantly improved. The crystals diffracted to a maximum resolution of about 6.5 Å.

Typical diffraction patterns are shown in Figure II.7.1. The diffraction patterns show the coexistence of Bragg reflections and layer lines with continuous diffraction. In spite of the rather low resolution of the data set the pattern could be indexed in the P3 space group with the following unit cell:  $a = b = 46.06$  and  $c = 104.13$ . The spacings of the Bragg spots and of the continuous diffraction are the same.



**Figure II.7.1.** Oscillation patterns (10°) of a crystal of CG(TA)<sub>3</sub>. (a) The long  $c^*$  axis of the unit cell is approximately vertical. The prominent stacking reflections at about 3.25 Å deviate about 5° ( $\beta$ ) from the meridional orientation. (b) Meridional Bragg reflections with Miller indices 004, 008 and 0012 are shown. (c) Bragg reflections in the equatorial region with Miller index 101 are evident. The spacing of the Bragg spots coincides with the spacing of the continuous diffraction.

The unit cell parameters correspond to a cell with three columns of four duplexes, i.e. 96 base pairs. The average volume per base pair is thus of about  $2000 \text{ \AA}^3$ .  $a$  and  $b$  parameters of the unit cell correspond to a distance of  $26 \text{ \AA}$  between the axes of neighbor duplexes (Figure II.7.2 b). The high hydration in the absence of strong interactions between columns in the crystal explains the low resolution of the data sets.



**Figure II.7.2.** (a) Schematic view of a column of the pseudocontinuous duplex. In the unit cell there are stacks of four duplexes; each duplex is rotated an angle  $\Omega$  of  $270^\circ$  with respect to its neighbors in the column. (b) Projection on the  $xy$  plane of nine unit cells.

The structure generated by the sequence  $CG(TA)_3$  is a pseudo-continuous coiled-coil where the duplexes are inclined an angle  $\beta$  of about  $5^\circ$  with respect to the major coil axis. Due to the low value of  $\beta$ , neighbor duplexes are approximately coaxial. The continuity of the coiled-coil is guaranteed by the sticky ends (see Figure II.7.3).



**Figure II.7.3.** Model of the organization of the oligonucleotides in the crystal. They form duplexes with sticky ends that generate infinitely long molecules, with staggered nicks in both strands.

The height  $h$  occupied by one octamer along the  $c$  direction is of about 26 Å ( $c/4 = 26.03$  Å). Therefore, the length  $l$  of a duplex corresponds to 26.13 Å ( $l = h/\cos \beta = 26.13$  Å).

The octamers are organized end-to-end to build a continuous coil. The structure is formed by a duplex with four base pairs and a distortion due to the two terminal base pairs, which generate nicks in both strands. Kinks ( $\theta$ ) of about  $173^\circ$  are generated ( $\cos \frac{\theta}{2} = \sin \beta \sin \frac{180^\circ}{N}$ , where  $N = 4$ ), see also Table II.14.

As said before in paragraph II.5.6, the overall rotation  $\Omega$  of one duplex with respect to its neighbor in a column is related to the twist value  $\omega$  of the individual base pairs. The value of  $\Omega$  can be exactly determined once the number  $N$  of duplexes stacked in one unit cell is known. The product between  $N$  and  $\Omega$  must be an exact multiple of  $360^\circ$ , so that the following relationship must be obeyed, with  $m$  a whole number:

$$N\Omega = 360m.$$

The helical arrangement that is obtained with these geometrical requirements has  $N$  individual duplexes in  $m$  turns, equivalent to a number of duplexes per turn of the overall helix given by the ratio  $360/\Omega$ .

In the structure generated by the sequence CG(TA)<sub>3</sub>, there are four ( $N$ ) duplexes (i.e. 32 base pairs) in three ( $m$ ) turns; thus  $\Omega$  is  $270^\circ$  ( $\Omega = 360^\circ m/N$ ). The average twist value  $\langle \omega \rangle$  for one octamer is  $33.75^\circ$  ( $\langle \omega \rangle = \Omega/8$ ), which corresponds to a helix of 10.7 base pairs per turn. Typically the DNA in solution has 10.4 base pairs per turn. It can be therefore concluded that the sequence d(CGTATATA) generates a pseudo-continuous helix where the absence of a phosphate has only a small influence on the overall structure of the pseudo continuous coil.

The discontinuity created by the overhanging bases guanine and cytosine, is probably responsible of a small change in the twist value of the base pairs. According to the values calculated by Gorin *et al.* in 1995, the average twist value of the (TA)<sub>3</sub> part of the octamer is  $35^\circ$ . Thus, in the CG discontinuity the average twist value  $\langle \omega \rangle$  should be  $30^\circ$  (smaller than the  $37^\circ$  found by Gorin *et al.* for terminal CG base step):

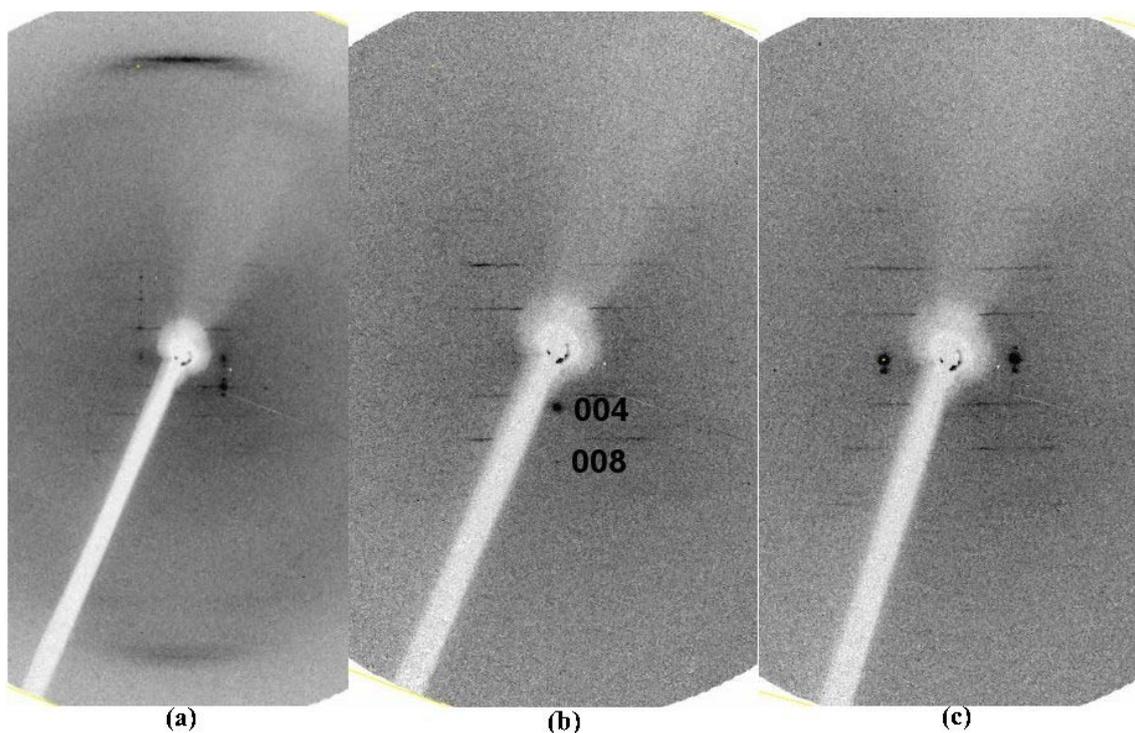
$$\langle \omega_{CG} \rangle = \frac{(\langle \omega \rangle \times 8 - \langle \omega_{(TA)_3} \rangle \times 6)}{2} = 30^\circ$$

## II.7.2 d(CGATATAT) d[CG(AT)<sub>3</sub>] and d(ATATATCG) d[(AT)<sub>3</sub>CG]

### d[CG(AT)<sub>3</sub>]

The sequence CG(AT)<sub>3</sub> has been crystallized using the hanging drop vapor diffusion technique and MPD as a precipitant; the crystals have been obtained in the following conditions: 0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 1 mM Netropsine, 10 mM MgCl<sub>2</sub> and 1 mM Spermine Tetrahydrochloride.

Typical diffraction patterns are shown in Figure II.7.4. The crystal structure of the sequence CG(AT)<sub>3</sub> is isomorphous with the structure of the sequence CG(TA)<sub>3</sub>. The unit cell parameters (manually determined) correspond to a trigonal unit cell with *a* and *b* parameters of about 46 Å and the *c* parameter of about 104 Å. The *c* value corresponds to a unit cell with stacks of four octamers along the *c* direction. The sequence CG(AT)<sub>3</sub> generates crystals of straight pseudo-continuous coils (i.e.  $\beta = 0^\circ$ ).



**Figure II.7.4.** Oscillation patterns (5°) of d(CGATATAT) crystals. The unit cell parameters, in a trigonal or pseudohexagonal space group, have been manually determined and are as follows: *a* and *b* about 46 Å and *c* about 104 Å. In (b) the Bragg spots with Miller indices 004 and 008 are shown. (c) Bragg spots with Miller indices 101 are evident.

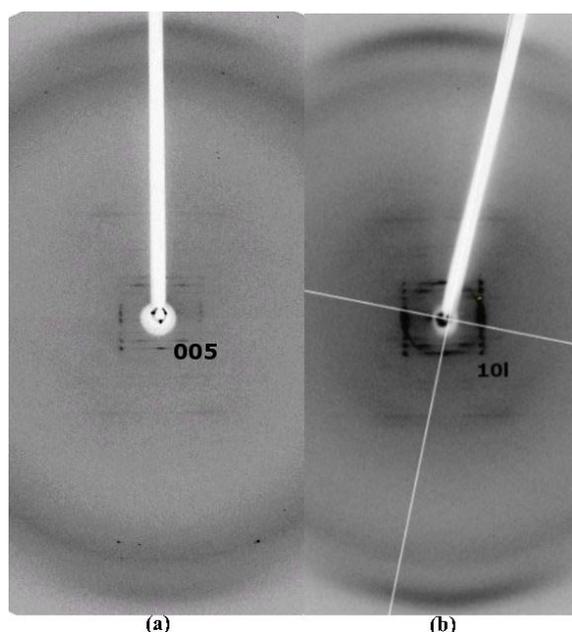
## $d[(AT)_3CG]$

The sequence  $(AT)_3CG$  has been crystallized using the hanging drop vapor diffusion technique and MPD as a precipitant; the crystals have been obtained in the following conditions: 0.5 mM DNA duplex, 50 mM NaCacodylate pH 6.5, 50 mM KCl, 1 mM  $CoCl_2$  and 1 mM Spermine Tetrahydrochloride.

Typical diffraction patterns obtained from crystals of  $d(ATATATCG)$  are shown in Figure II.7.5. Despite the poor diffraction, the unit cell has been manually determined; it corresponds to a trigonal or pseudo-hexagonal cell with  $a$  and  $b$  parameters of about 46 Å and  $c$  parameter apparently of about 130 Å. The  $c$  parameter corresponds to a unit cell with a stack of five duplexes along the  $c$  direction. The overall rotation  $\Omega$  of one duplex with respect to its neighbor in a column is thus  $288^\circ$  ( $\Omega = \frac{360^\circ m}{N}$ , with  $N = 5$  and  $m = 4$ ). The average twist value per base pair  $\langle\omega\rangle$  is thus  $36^\circ$  ( $\Omega/8$ ); the latter value is in agreement with those previously calculated for standard B-form DNA (Gorin *et al.*, 1995).

The typical DNA streaks are not clearly visible, in particular the second layer line is extremely weak, for this reason it can be speculated on a possible discontinuity that the change in hydrogen bonds type (from Watson-Crick for the  $d(CG)$  to Hoogsteen for the AT part) can create.

The characteristics of the octamers studied in this work are summarized in Table II.14.



**Figure II.7.5.** Oscillation patterns ( $15^\circ$ ) obtained from crystals of  $d(ATATATCG)$ . The stacking reflections are found at about 3.25 Å resolution.

**Table II.14.** Characteristics of the octamers studied in this work.

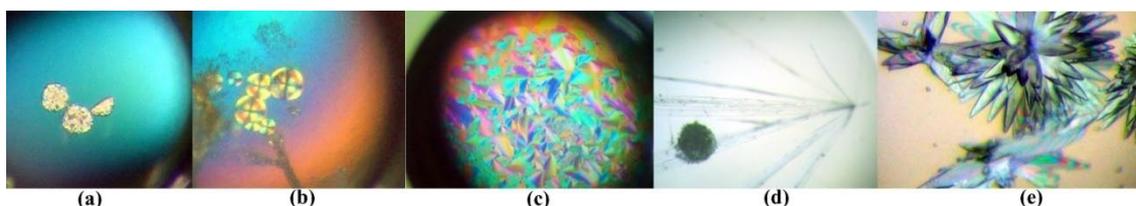
Sequence	Unit cell (Å)	$\beta$	Rise (Å)	$l$	$N$	Resolution	Comments	Crystallization conditions
<b>CG(TA)<sub>3</sub></b>	$a = b = 46.06$ $c = 104.13$	$\approx 5^\circ$	3.25	26.13	4	6.5	Probably standard B-form.	0.5 mM DNA, 20 mM NaCacodylate pH 6.5, 20 mM KCl, 16.7 mM Thymidine, 1 $\mu$ l 4-aminophenilsulfon, 1 mM Spermine tetrahydrochloride and MPD 20%.
<b>CG(AT)<sub>3</sub></b>	$a = b \approx 46$ $c \approx 104$	$0^\circ$	3.24	26.9	4	16	Mainly streaks; isomorphous with CG(TA) <sub>3</sub> .	0.5 mM DNA, 50 mM NaCacodylate pH 6.5, 10 mM MgCl <sub>2</sub> , 1 mM Netropsine, 1 mM Spermine tetrachloride and MPD 32%.
<b>(AT)<sub>3</sub>CG</b>	$a = b \approx 46$ $c \approx 130$	$\approx 5^\circ$	3.23	26	5	16	Probably Hoogsteen.	0.5 mM DNA, 50 mM NaCacodylate pH 6.5, 1 mM CoCl <sub>2</sub> , 50 mM KCl, 1 mM Spermine tetrachloride and MPD 30%.

## II.8 HEXAMER

### II.8.1 d(CGATAT), preliminary considerations

The sequence CG(AT)<sub>2</sub> has been crystallized using the hanging drop vapor diffusion technique and PEG 4000 as a precipitant. The hanging drops were prepared by mixing 1 µl of the DNA stock solution for a final concentration of 0.3 mM with 1 µl of 31 conditions of the Hampton Research crystallization screen for DNA, Natrix (HR2-116). A birefringent precipitate was obtained in the following conditions: 80 mM Mg Acetate, 50 mM NaCacodylate pH 6.5, 30% PEG 4000. Drops with increasing Spermine concentration (0 mM to 6 mM) were set up. The hanging drops were incubated at 13 °C over 800 µl of a reservoir consisting of an aqueous solution of 30% PEG 4000. After a few weeks two-dimensional plates appeared in those drops with a Spermine concentration higher than 2 mM, while at lower concentration only precipitate or spherulites have been obtained. More than one month after the drops were set up, instead of the plates and the precipitate, needles started to appear: extremely thin needles at 2 mM Spermine concentration, clusters of big needles at higher Spermine concentrations (see Figure II.8.1). The needles had to be separated in order to obtain a single crystal suitable for the data collection. This crystal will from now on be referred to as “P52D3”.

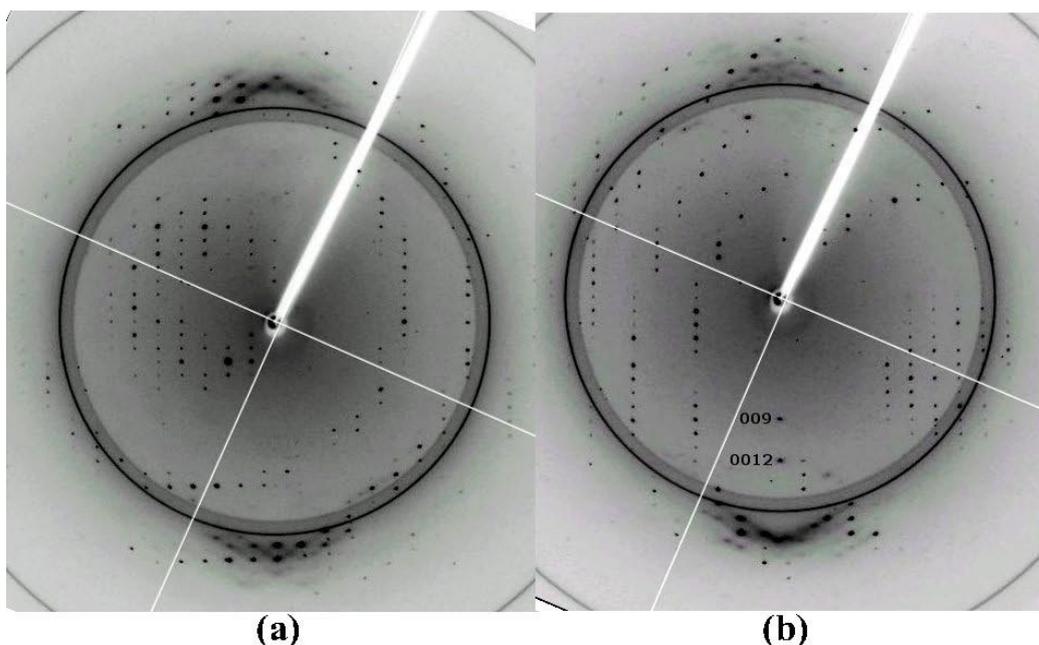
The Spermine/DNA ratio is much higher than for the other sequences studied in this thesis, for instance, in the CGAT5 case Spermine was three times more concentrated than DNA, while in the P52D3 case Spermine concentration is ten to seventeen times higher than DNA concentration. Besides, PEG has been used as a precipitant instead of MPD and the salt concentration (80 mM Mg Acetate) is relatively high.



**Figure II.8.1.** Microscope photographs of crystals of CG(AT)<sub>2</sub> obtained at Spermine different concentrations: (a, b) 0-1 mM Spermine; (c) 2-3 mM Spermine after one month the drop was set up; (d) 2 mM Spermine; (e) 4-6 mM Spermine.

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 BM16, ESRF, Grenoble. 180° of oscillation data with a rotation of 2° per image were collected in a high resolution data set and another 180° of data in a low resolution pass. The images were integrated and scaled with the HKL package to a resolution of 2.6 Å. The crystal belongs to a hexagonal space group, with the following unit cell:  $a = b = 38.44$  Å and  $c = 57.00$  Å.

Typical diffraction patterns are shown in Figure II.8.2, the stacking diffraction at about 3.2 Å is slightly off meridian indicating that the duplexes form a coiled-coil and their axes are slightly inclined with respect to the major axis. The curious shape of the stacking reflection might also indicate a peculiar arrangement of the junction. The data indicates that the crystal is formed by infinite parallel columns along the  $c$  dimension.



**Figure II.8.2.** Oscillation patterns (2°) of crystal D52D3, the maximum resolution is about 2.6 Å. (b) Bragg reflections with Miller indices 009 and 0012 are shown.

## II.9 SUMMARY AND CONCLUSIONS

### Sequences studied

- All the sequences crystallized in this work are characterized by the presence of sticky ends, with the exception of  $d(AT)_5T$ .
- The overhanging sequences are either CG or GC.
- The overhang is either at the 5' or at the 3' end of the sequence.
- The central part of the sequences is represented by alternating fragments,  $(TA)_n$  or  $(AT)_n$ .
- The following sequences have been studied:

Overhanging sequence	DNA sequences	
CG	$CG(TA)_n$	$n=3$
	$CG(AT)_n$	$n=2, 3, 4, 5$
	$(AT)_nCG$	$n=3, 4, 5$
	$CGCG(AT)_n$	$n=4$
GC	$GC(AT)_n$	$n=5$
	$(AT)_nGC$	$n=5$

- The characteristics of the structures studied in this thesis are summarized in Table II.15.

### Conclusions

1. Practically all the sequences crystallized in this thesis are characterized by a high nucleation rate as well as a high crystal growth rate, with the exception of  $CG(AT)_2$  whose crystals grow within months. As a result the crystals have high mosaicity and often a fibrous structure.
2. The duplexes with sticky ends usually form infinite pseudo-continuous coiled-coils with staggered nicks.
3. Only the structure of the sequence  $CG(AT)_5$  could be determined at 3.1 Å resolution. It generates a right handed coiled coil with six duplexes per turn. The  $(AT)_5$  fragment is in the Hoogsteen conformation. The sticky end CG is assumed

to be in a standard Watson-Crick conformation, but that region appears disordered. The kink that gives rise to the coiled coil is attributed to the discontinuity of the phosphodiester chain.

4. All the sequences studied here pack with a hexagonal or pseudo-hexagonal symmetry. The geometry of the coiled-coils is determined by the angle  $\theta$  between consecutive duplexes and the torsion angle  $\tau$ . The latter is equivalent to the usual twist parameter  $\omega$ . The majority of the sequences generate coiled-coils, with different  $\beta$  inclination and number  $N$  of residues per turn:

	$N$	$\beta$
<b>Dodecamers</b>	6; 6,75; 7; 14	11°; 20°; 25°
<b>Decamers</b>	30-22-24	15°-25°-32°
<b>Octamers</b>	4-5	0°-5°

5. Depending on the value of  $\tau$  the coiled-coils may be either right handed, as in the case of  $\text{CG(AT)}_5$ , or left handed, as in the case of  $\text{CG(AT)}_4$ .
6. It is not clear whether the hydrogen bonding mode has an influence on the geometry of the coils or not, but it seems that a discontinuity in the hydrogen bonding is necessary for a coiled-coil to form. We suggest that the central alternating AT region forms Hoogsteen base pairs. However this conformation has only been firmly determined for  $\text{d(ATATAT)}$  and  $\text{d(CGATATATATAT)}$ .
7. Special cases:
  - In the case of the sequence  $\text{CGATATGCATAT}$  the central CG base pairs force the flanking AT to form standard Watson-Crick hydrogen bonds, thus a standard B-form DNA is generated (see also Qiu *et al.*, 1997).
  - In the case of  $(\text{AT})_5\text{GC}$ , the flipped out cytosine probably interacts with a neighbor AT base pair, the kink thus results less pronounced. The duplex appears to adopt a standard B-form.
  - The packing of the sequence  $\text{CG(AT)}_4$  varies depending on the precipitant concentration.

**Table II.15.** Summary of the characteristics of the sequences crystallized in this thesis. The coiled-coils are probably Hoogsteen in most cases. In brackets are the numbers of oligonucleotides per turn.

Sequence	n=2	n=3	n=4	n=5
<b>AT(AT)<sub>n</sub></b>	Abrescia <i>et al.</i> , 2002	Not studied	Coiled-coil $c = 523$ (18) (work in progress)	Campos <i>et al.</i> , 2005 Coiled-coil $c = 216$ (6)
<b>CG(AT)<sub>n</sub></b>	$c = 57.00$ (3)	Continuous coil $c \approx 104$ (4)	Coiled-coil $c \approx 645$ (22 – 24) Various types	Coiled-coil $c = 220.53$ (6)
<b>CG(TA)<sub>n</sub></b>	Not studied	Coiled-coil $c = 104.12$ (4)	Not studied	Not studied
<b>GC(AT)<sub>n</sub></b>	Not studied	Not studied	Not studied	Coiled-coil $c = 479$ (7 – 14)
<b>(AT)<sub>n</sub>CG</b>	Not studied	Coiled-coil $c = 130$ (5)	I. B-form. Fiber diffraction II. Coiled-coil $c = 923$ (30)	Coiled-coil $c = 217$ (6)
<b>(AT)<sub>n</sub>GC</b>	Not studied	Not studied	Not studied	Flipped out cytosine. $c = 147$ (4)
<b>CGCG(AT)<sub>n</sub></b>	Not studied	Not studied	Coiled-coil $c = 227.5$ (6)	Not studied
<b>CG(AT)<sub>n</sub>GC(AT)<sub>n</sub></b>	Not studied	B-form $c = 230.5$ (6)	Not studied	Not studied

## **III. APPENDIX**

### **III.1 The influence of size on the thermal stability of oligonucleotides: the case of AT sequences.**

Notes & Tips

## The influence of size on the thermal stability of oligonucleotides: the case of AT sequences

Daniela De Luchi,<sup>a</sup> Catherine Gouyette,<sup>b</sup> and Juan A. Subirana<sup>a,\*</sup>

<sup>a</sup> *Department d'Enginyeria Química, ETSEIB, Universitat Politècnica de Catalunya, Av. Diagonal 647, Barcelona E-08028, Spain*

<sup>b</sup> *Unité de Chimie Organique, Institut Pasteur, 28 rue du Dr. Roux, Paris 75724, France*

Received 23 May 2003

Since the early studies of Marmur and Doty [1], the melting temperature of DNA and oligonucleotides has been used for characterization. In our laboratory we are studying the structure of short AT-rich oligonucleotides by X-ray diffraction [2]. It is often necessary to know their thermal stability. Also we need to determine whether they show any sign of changes in conformation as a function of sequence and temperature. The theory of the melting transitions of oligonucleotides has been analyzed in detail by Marky and Breslauer [3]. However the theory is complex and it is not easy to use in a straightforward manner. In this paper we present a simple approach that allows a rapid determination of the influence of the various factors that determine the melting behavior of short oligonucleotides, in particular the influence of size.

### Materials and methods

Oligonucleotides were synthesized on an automatic synthesizer by the phosphoramidite method and purified by gel filtration and reverse-phase HPLC. Samples were prepared for melting experiments by diluting appropriate aliquots from a concentrated stock solution of the oligonucleotides with 300  $\mu$ L of the melting buffer (2 mM  $\text{NaH}_2\text{PO}_4$ , 6 mM  $\text{Na}_2\text{HPO}_4$ , pH 7.0, 1 M NaCl), for a final absorbance at 260 nm between 0.3 and 0.6 optical density in 1-mm-path-length cells at 20 °C. Thus the duplex concentration was about 50  $\mu$ M for hexamers and 25  $\mu$ M for dodecamers.

Melting curves were obtained at 260-nm wavelength in a Varian Cary 100 spectrophotometer fitted with a

thermoelectrically controlled sample holder. The heating rate was fixed at 0.5 °C  $\text{min}^{-1}$  and data were collected at 0.5 °C intervals up to 80 °C. Prior to the melting experiments, the samples were degassed and heated to 80 °C and then slowly annealed to the starting temperature. Water condensation on the cuvette exterior in the low-temperature region was avoided by flushing with a stream of dry nitrogen (4 L  $\text{min}^{-1}$ ). However, due to the high humidity in Barcelona and the poor design of the commercial instrument chamber, it was often found impossible to work below 10 °C. The absorbance versus temperature curves were determined for sample and reference. Then the reference curve was subtracted from the sample curve. The resulting curves were smoothed and the maximum of the first derivative was taken to identify the melting temperature  $T_m$ .

### Results

#### UV spectra

The UV spectra were similar in all the samples that we have studied (results not shown), with absorption maxima in the range 260–263 nm depending on size and sequence. In alternating AT oligonucleotides we found upon denaturation a blueshift of the maximum in the range 0.6–1.5 nm. Nonalternating oligonucleotides did not show such a shift. No bimodal melting curves were observed. The values of  $T_m$  that we have found are given in Table 1.

#### Influence of concentration

The process of oligonucleotide melting is an equilibrium process in which duplex structures and single

\* Corresponding author. Fax: +34-934010978.  
E-mail address: [juan.a.subirana@upc.es](mailto:juan.a.subirana@upc.es) (J.A. Subirana).

Table 1  
Melting temperatures of oligonucleotides in 1 M NaCl

Sequence	$T_m$
TATATA	0.92 <sup>a</sup>
(CG)TATATA	10.05
AATATATT	19.50
ATATATAT	16.50
(CG)ATATATAT	29.02
GTATATAC	35.02
GATATATC	32.02
AATAATTATT	37.01
TATATATATA	30.00
ATATATATAT	31.50
ATATATATAT(T)	35.00
CTATATATAG	38.00
AATTAATTAATT	44.01
ATATATATATAT	40.51

<sup>a</sup> Obtained from the value of TATATATATA extrapolated by using Eq. (2).

strands coexist as a function of temperature. At high temperature only single strands are present. At low temperature different structures may coexist: perfect duplexes, hairpins, frayed ends, long duplex structures formed by several single strands, etc. Concentration will have an influence on the type of structures found at each temperature: high concentrations will favor duplex formation. This effect has been studied by various authors for particular sequences [4–7]. The results obtained depend strongly on the solvent and the size and sequence of the oligonucleotides studied. In the range of size and concentration used in our study the decrease in melting temperature observed is about 5–10°C upon a 10-fold decrease of concentration. The larger the oligonucleotide, the smaller is the effect. No influence of concentration is expected for genomic DNA fragments. We have carried out our experiments for comparative purposes at similar concentrations in all cases (see Materials and methods).

#### Influence of ionic strength

When the data of Marmur and Doty [1] are plotted as a function of ionic strength, it is found that  $T_m$  increases in a linear fashion as a function of the logarithm of the ionic strength. An increase of 14.7°C in melting temperature was detected upon a 10-fold increase of salt concentration. A similar behavior has been found in oligonucleotides [7] and polynucleotides [8], although the equivalent increase in temperature differs in each case, values between 12 and 22°C have been reported. At high salt concentration (above 0.5 M) the effect of ionic strength diminishes. Thus Marmur and Doty [1] find a smaller increase of  $T_m$  of about 11.7°C between 0.1 and 1.0 M salt. The latter value is used to correct the values reported by other investigators in the figures presented below.

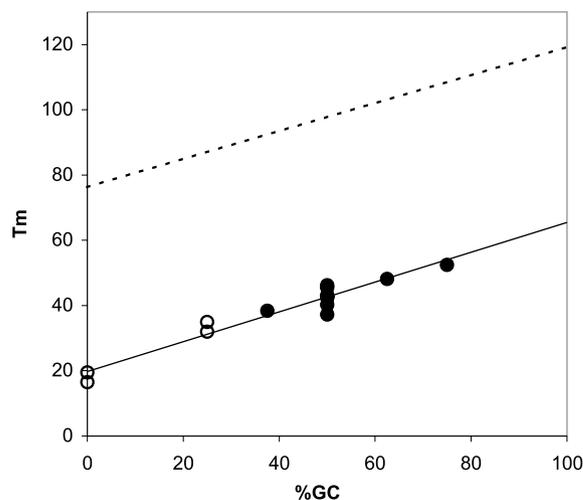
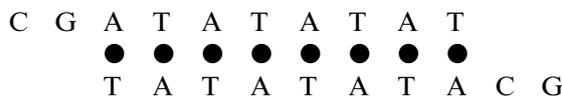


Fig. 1. Dependence of the melting temperatures,  $T_m$ , on DNA composition; dashed line, bacterial DNAs [1]; solid line, octamer (empty points are taken from Table 1; full points are taken from [10]). The data correspond to 1 M NaCl concentration. The difference in salt concentration has been corrected for bacterial DNA.

#### Influence of composition and sequence

The overall composition (CG%) of the oligonucleotides has a practically identical influence on  $T_m$ , as it is found in bacterial DNAs, as shown in Fig. 1. The increase in  $T_m$  for 1% increase in CG content is 0.43°C for bacterial DNA and 0.46°C for the octamers shown in Fig. 1. On the other hand the influence of sequence is noticeable, as it is apparent from the values given in Table 1 and those presented by other authors [9,10]. For example a difference of 7°C is found between AATAATTATT and TATATATATA, as shown in Table 1. The first sequence has two TA steps, while the second one has five. The end sequences also have an effect on  $T_m$ . These observations are in agreement with those of Movileanu et al [11] who observed that poly(dA)·poly(dT) melted at a temperature higher by 4.3°C than that of poly d(AT). It is clear that the very mobile TA step [12] decreases the temperature of AT-rich sequences, but a quantitative estimate cannot be given due to the additional influence of the end sequences.

A puzzling observation is the influence of unpaired ends on  $T_m$ . The extra terminal T in ATATATATATT increases  $T_m$  by 3.5°C when compared with ATATATATAT. It appears that the extra base at the end has a capping effect which increases the  $T_m$  of the decamer. Much more striking is the effect of adding an extra CG sequence at the 5' end of the oligonucleotide. In the case of CGTATATA,  $T_m$  increases 9.1°C and in CGATATATAT the increase is 12.5°C. The expected structure of the latter duplex is



It appears that the additional CG ends may allow pairing of different duplexes and thus increase substantially the melting temperature. These are additional examples of the strong effect of terminal sequences on  $T_m$ .

### Influence of size

The main purpose of this paper is to ascertain the influence of size on  $T_m$  for AT oligonucleotides. If we assume that the gain in entropy  $\Delta S_0$  for a molecule when going from duplex to single strand is a constant that does not depend on either size or sequence, the melting temperature can be given by:

$$T_m = \frac{n\Delta H}{n\Delta S + \Delta S_0}, \quad (1)$$

where  $\Delta H$  and  $\Delta S$  are the average values for melting enthalpy and entropy of a single A·T pair and  $n$  is the number of basepairs in the oligonucleotide. The equation can be transformed to:

$$T_m^{-1} = T_{m\infty}^{-1} + kn^{-1}, \quad (2)$$

where  $T_{m\infty}$  is the melting temperature of a polymer and  $k$  is a constant equal to  $(\Delta S_0/\Delta H)$ . The  $T_m$  values obtained by us and some other results from the literature are plotted in this way in Fig. 2. Values of  $T_{m\infty} = 356.38$  K and  $k = 5.247 \cdot 10^{-3} \text{ K}^{-1}$  are found for oligonucleotides devoid of C·G base pairs. In a similar way, an equation can be derived from literature values [4,6,9,10] for oligonucleotides with different compositions. Thus the equations to apply for 0 and 100% CG oligonucleotides would be

$$0\%CG, \quad T_m^{-1} = 2.806 \cdot 10^{-3} + 5.247 \cdot 10^{-3}n^{-1} \quad (3a)$$

and

$$100\%CG, \quad T_m^{-1} = 2.518 \cdot 10^{-3} + 4.225 \cdot 10^{-3}n^{-1}. \quad (3b)$$

Similar equations can be derived by interpolation for oligonucleotides with intermediate compositions.

### Conclusion

Our analysis of the influence of different factors has allowed us to derive the Eqs. (3a) and (3b) which should be useful to predict the melting temperature  $T_m$  of oligonucleotides of any composition and size. The prediction is only approximate, since it is clear that sequence and end effects are important. Such effects have been discussed in detail by SantaLucia [13]. Furthermore there is an influence of concentration which increases as the size of the oligonucleotide decreases.

Finally we should add that Scheffler et al. [14,15] studied long d(AT) sequences, which have a complex behavior due to the formation of hairpins and mismatches. Our results are in general agreement with those reported by the latter authors, although we did not find a biphasic transition for TATATATATA (Fig. 6 in [15]).

### References

- [1] J. Marmur, P. Doty, Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature, *J. Mol. Biol.* 5 (1962) 109–118.
- [2] N.G.A. Abrescia, A. Thompson, T. Huynh-Dinh, J.A. Subirana, Crystal structure of an antiparallel DNA fragment with Hoogsteen base pairing, *Proc. Natl. Acad. Sci. USA* 99 (2002) 2806–2811.
- [3] L.A. Marky, K.J. Breslauer, Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves, *Biopolymers* 26 (1987) 1601–1620.
- [4] M. Petersheim, D.H. Turner, Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp, *Biochemistry* 22 (1983) 256–263.
- [5] S.M. Freier, D. Deprisco Albergo, D.H. Turner, Solvent effects on the dynamics of (dG–dC)<sub>3</sub>, *Biopolymers* 22 (1983) 1107–1131.
- [6] L.A. Marky, N.R. Kallenbach, K.A. McDonough, N. Seeman, K.J. Breslauer, The melting behavior of a DNA junction structure: a calorimetric and spectroscopic study, *Biopolymers* 26 (1987) 1621–1634.
- [7] A.P. Williams, C.E. Longfellow, S.M. Freier, R. Kierzek, D.H. Turner, Laser temperature-jump, spectroscopic, and thermodynamic study of salt effects on duplex formation by dGCATGC, *Biochemistry* 28 (1989) 4283–4291.
- [8] R.B. Inman, R.L. Baldwin, Helix-random coil transitions in DNA homopolymer pairs, *J. Mol. Biol.* 8 (1964) 452–469.
- [9] M.J. Doktycz, M.D. Morris, S.J. Dormady, K.L. Beattie, K.B. Jacobson, Optical melting of 128 octamer DNA duplexes, *J. Biol. Chem.* 270 (1995) 8439–8445.
- [10] P. Wu, S. Nakano, N. Sugimoto, Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation, *Eur. J. Biochem.* 269 (1995) 2821–2830.

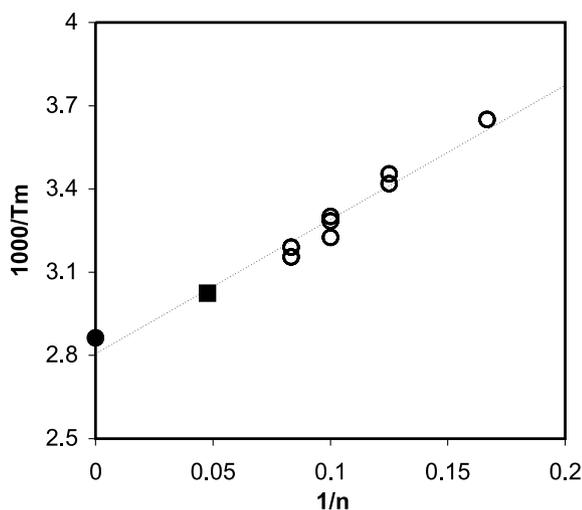


Fig. 2. Dependence of the melting temperature,  $T_m$ , on size  $n$  for oligonucleotides with 100% AT basepairs; empty points are taken from Table 1; the full circle is taken from [16]; the full square is taken from [15]. The difference in salt concentration has been corrected in the latter cases.

- [11] L. Movileanu, J.M. Benevides, G.J. Thomas Jr., Determination of base and backbone contributions to the thermodynamics of premelting and melting transitions in *B* DNA, *Nucleic Acids Res.* 30 (2002) 3767–3777.
- [12] J.A. Subirana, T. Faria, Influence of sequence on the conformation of the B-DNA helix, *Biophys. J.* 73 (1997) 333–338.
- [13] J. SantaLucia Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc. Natl. Acad. Sci. USA* 95 (1998) 1460–1465.
- [14] I.E. Scheffler, E.L. Elson, R.L. Baldwin, Helix formation by dAT oligomers I. Hairpin and straight-chain helices, *J. Mol. Biol.* 36 (1968) 291–304.
- [15] I.E. Scheffler, E.L. Elson, R.L. Baldwin, Helix formation by dAT oligomers II. Analysis of the helix-coil transitions of linear and circular oligomers, *J. Mol. Biol.* 48 (1970) 145–171.
- [16] K. Tanaka, Y. Yamada, M. Shionoya, Formation of silver(I)-mediated DNA duplex and triplex through on alternative base pair of pyridine nucleobases, *J. Am. Chem. Soc.* 124 (2002) 8802–8803.

**III.2 Structure of the DNA Coiled-coil formed by  
d(CGATATATATAT)**

DOI: 10.1002/cbic.200500449

## Structure of the DNA Coiled Coil Formed by d(CGATATATATAT)

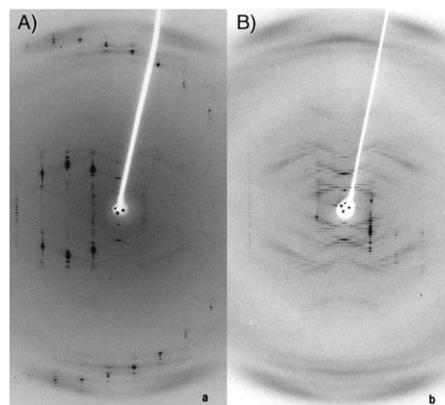
Daniela De Luchi,<sup>[a]</sup> Valentina Tereshko,<sup>[b]</sup>  
Catherine Gouyette,<sup>[c]</sup> and Juan A. Subirana<sup>\*[a]</sup>

AT-rich sequences of DNA are highly polymorphic, as we have reviewed elsewhere<sup>[1]</sup> and are abundant in noncoding regions of the genome.<sup>[1,2]</sup> As part of our studies on these sequences, we have determined the molecular structure of the coiled coil obtained from duplex d(CGATATATATAT) by single-crystal X-ray diffraction. We recently described the structure of the coiled coil formed by duplex d(ATATATATATAT).<sup>[2]</sup> Due to the presence of sticky ends it forms a continuous double helix. However diffraction data could only be obtained with a limited resolution of 5 Å; this is probably due to a combination of factors: high solvent content, screw disorder, and possible multiple alignments. In order to eliminate the latter factor, we crystallized d(CGATATATATAT), which can only form stacked infinite continuous duplexes in a single way, as shown in Figure 1. The crystals were practically isomorphous to those previously described.<sup>[2]</sup> We determined the 3.1 Å X-ray structure and found that this dodecamer forms a continuous coiled coil with Hoogsteen base pairs between A and T.



**Figure 1.** Scheme of base pairing of the dodecamer used in this study. Infinite continuous duplexes are formed that have missing phosphates (vertical lines) in both phosphodiester chains, staggered by two base pairs. Dyad axes are indicated by ellipses.

In some cases, the crystals showed very few Bragg spots, only continuous X-ray fiber diffraction on layer lines, as previously observed in d(ATATATATATAT).<sup>[2]</sup> The presence of continuous diffraction indicates screw disorder, with parallel orientation of the diffracting entities. Slight variations in the crystallization conditions resulted in crystals that show little screw disorder and Bragg spots that extend for up to 3 Å. Examples of both types of diffraction are given in Figure 2. Although the resolution was much better than in our previous study of d(AT)<sub>6</sub>, the data had high mosaicity, as is apparent in Figure 2.



**Figure 2.** Oscillation patterns obtained from two different crystals with the *c*-axis approximately perpendicular to the X-ray beam. A) Crystalline pattern, high mosaicity in the upper layer lines may be appreciated (0.5° oscillation). B) Fiber-like pattern showing few Bragg spots and clear layer lines, which indicate strong screw disorder (5° oscillation). The off-meridional stacking reflections indicate that, in both cases, the duplexes are inclined by about 20° with respect to the *c*-axis.

Nevertheless, our study showed unambiguously that the AT base pairs had the Hoogsteen conformation. The CG base pairs were not clearly defined and were assumed to form Watson–Crick hydrogen bonds. The electron density map is shown in Figure 3. All features confirm the Hoogsteen conformation of the duplex. In particular, the overall shape of the molecule demonstrates a narrow minor groove and the helical axis displaced towards one side of the base pairs, as expected for a Hoogsteen duplex.<sup>[1]</sup>

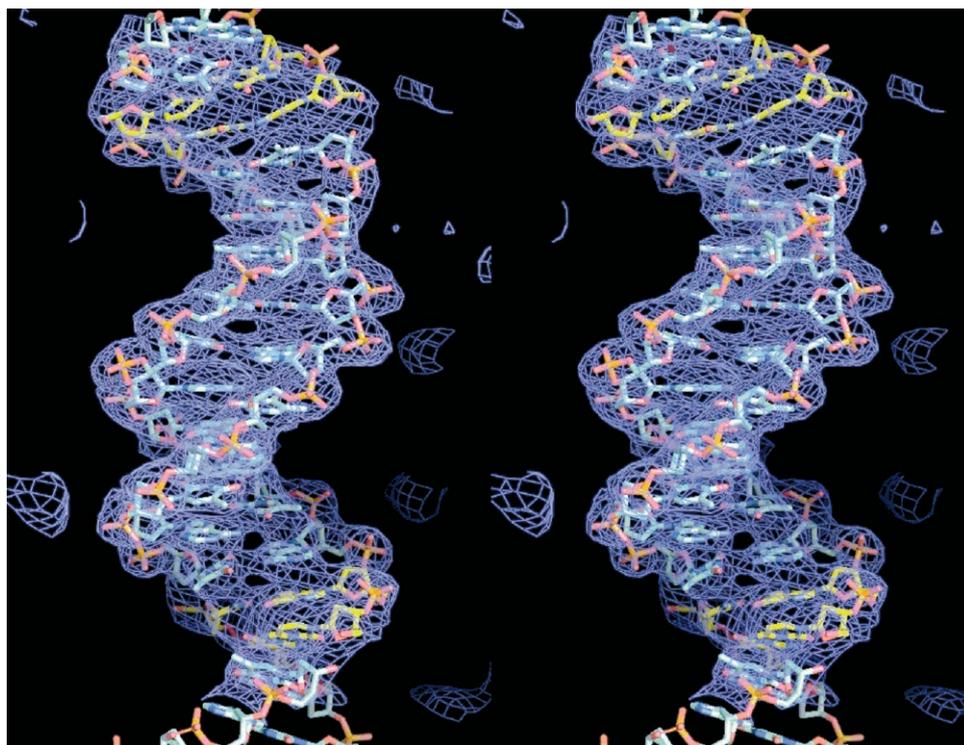
The coiled coils in the crystal do not show any obvious lateral interactions. It appears that coiling is an intrinsic feature of these oligonucleotides and not due to crystal packing; this explains the strong tendency to screw disorder (Figure 2). The volume per base pair in the crystal is 1868 Å<sup>3</sup>, significantly greater than the value usually found in conventional oligonucleotide crystals (around 1300 Å<sup>3</sup>). The high solvent content leads to high mosaicity in the crystals and limits the resolution of the diffraction data. Furthermore the comparatively high *R* factors (Table 1) are related to such features of our crystals. The large volume per base pair indicates that the Hoogsteen structure is stable under highly hydrated conditions, although we have not detected Hoogsteen DNA in solution.<sup>[1]</sup>

The overall structure is shown in Figure 4. The coil has a pitch of 220.5 Å, with six dodecamers per turn. Its average radius is 11.7 Å. The coiled coil shows kinks at the position of the CG base pairs, where the phosphodiester chain is interrupted (Figure 1). The kinks result in a strong compression on the major-groove side and opening of the minor groove, as is apparent in Figure 4, in agreement with the classic studies of Dickerson et al.<sup>[3]</sup> The absence of two phosphate residues facilitates this distortion. The central AT decamer forms a straight duplex, as is apparent in Figure 3. The A·T base pairs at both sides of the CG region form a large angle (equivalent to roll) and compress the CG dimer sequence. Since the structure is practically isomorphous with d(AT)<sub>6</sub>, the origin of the kinks

[a] D. De Luchi, Prof. J. A. Subirana  
Departament d'Enginyeria Química, Universitat Politècnica de Catalunya  
Av. Diagonal 647, 08028 Barcelona (Spain)  
Fax: (+34)934-010-978  
E-mail: juan.a.subirana@upc.edu

[b] Dr. V. Tereshko  
Department of Biochemistry and Molecular Biology, Chicago University  
920 East 58th Street, Chicago, IL 60637 (USA)

[c] C. Gouyette  
Institut Pasteur  
28 rue du Docteur Roux, 75724 Paris Cedex 15 (France)



**Figure 3.** Stereo view of the electron density of one duplex and its two cohesive ends. The C-G base pairs are shown in yellow. The practically straight Hoogsteen duplex formed by ten A-T base pairs is shown in cyan. The narrow minor groove at the center and the deep major groove at both sides can be appreciated.

**Table 1.** Crystallographic and structural data.

Cell parameters [Å]	$a = b = 26.54$ ; $c = 220.53$
space group	$P6_122$
asymmetric unit	One dodecamer (single strand)
volume/base pair	$1868 \text{ \AA}^3$
$R_{\text{work}}/R_{\text{free}}$ [%]	32.2/34.2
AT decamer duplex <sup>[a]</sup>	
C1'–C1' [Å]	$8.26_{0.19}$
twist [°]	$34.1_{6.0}$
rise [Å]	$3.36_{0.27}$
glycosidic angle $\chi^\circ$ (A)	$67_{16}$
glycosidic angle $\chi^\circ$ (T)	$-99_{23}$

[a] The values have been calculated with the 3 DNA program, based on C1'–C1' vectors.<sup>[9]</sup> Standard deviations are given as subindexes.

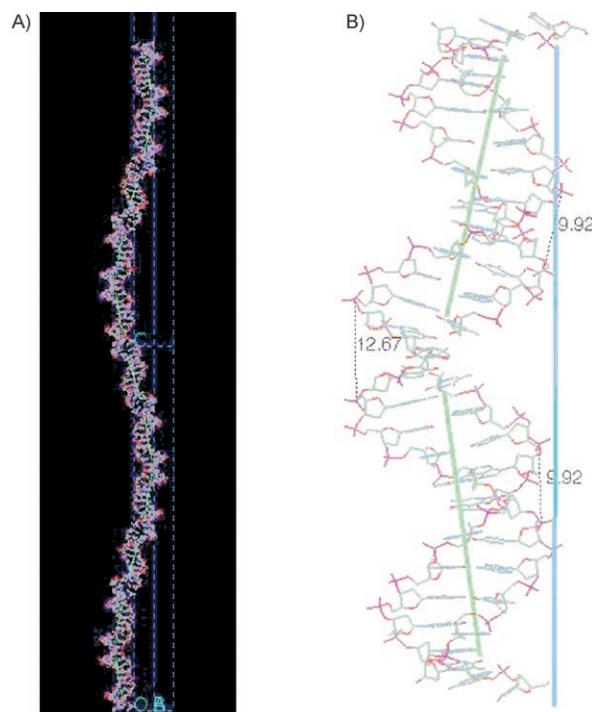
should be attributed to the discontinuity of the phosphodiester chain, rather than to the presence of a short CG stretch. It is interesting to note that d(CGACGATCGT) also crystallizes as a continuous duplex with its sticky ends paired,<sup>[4]</sup> but as a B-form straight double helix. Since the starting sequence, CGA, is the same in our case, it is tempting to speculate that the coiled-coil conformation requires Hoogsteen base pairs. Further work with other sequences is required in order to analyze this possibility. In principle, coiled coils should not require Hoogsteen-type pairing.

The coiled coil is a stable, rigid structure that represents a new conformation of DNA, as part of the polymorphism found in AT sequences.<sup>[1]</sup> Such sequences are very abundant in non-

coding regions of the genome,<sup>[2]</sup> but their structure and function are not known. We are presently studying other sequences in order to determine under what conditions DNA coiled coils may be formed.

## Experimental Section

Crystals obtained under different conditions diffracted to a maximum of 3 Å resolution. Long, flexible needles were grown at 13 °C by using the hanging-drop vapor-diffusion technique with 2-methylpentane-2,4-diol (MPD) as a precipitant. The drop was equilibrated against a reservoir solution at 31% (v/v) MPD. Due to their high flexibility, it was not possible to freeze the needles for X-ray data collection; however, on decreasing the temperature to 4 °C, they became more rigid and suitable for manipulation. The best crystals were obtained from the following solution: duplex (0.5 mM), spermine (1.5 mM), KCl (50 mM),



**Figure 4.** A) View of two turns of the coiled coil and B) detail of two consecutive duplexes. The axis of the duplexes is shown in green, whereas the axis of the coiled coil is shown in cyan. The molecules are projected onto the plane formed by the duplex axes (green), so that the widening of the minor groove in the kink region may be easily appreciated.

sodium cacodylate (50 mM) at pH 6.5, trimethylamine-*N*-oxide (15 mM). The latter additive was not essential for obtaining crystals, but improved the resolution.

Diffraction data were measured with cryocooling at 110 K at beamline BM16 (ESRF), processed, and scaled with HKL2000.<sup>[5]</sup> The structure was solved by molecular replacement with AMoRe.<sup>[6]</sup> The fragment (AT)<sub>5</sub> has been used as search model, with either Hoogsteen or Watson–Crick base pairing constructed from PDB entries 1GQU and IDN9, respectively. A reasonable solution was only found for the Hoogsteen model, with the (AT)<sub>5</sub> fragment aligned along the crystal *c*-axis. The molecular dyad coincided with the crystallographic twofold axis. The extra CG bases were manually built. To test the possibility of alternative pairing in the AT region, the Watson–Crick model was superimposed on the central decamer. Both models were refined by using the CNS package.<sup>[7]</sup> At all stages of refinement, restraints were applied to maintain the geometry of the hydrogen bonds in Hoogsteen and Watson–Crick base pairs. Initially, each model was refined as a rigid body. After that, each base pair, each nucleotide, and each base, phosphate and sugar were, respectively, treated as a rigid body. Simulated annealing was then carried out, yielding *R* factors 10% lower for a Hoogsteen model than for a Watson–Crick model. At this stage, the electron density maps were calculated, and both models were compared. In the AT region, the geometry of the Watson–Crick model was completely distorted, while the Hoogsteen model displayed adequate hydrogen bonds and stacking arrangement. The region of the C-G base pairs showed a very poor electron-density map that indicated high disorder. The C-G base pairs were maintained in the Watson–Crick conformation, but it cannot be excluded that they could also be in the Hoogsteen form. We concluded the refinement with Refmac5.<sup>[8]</sup> Given the low resolution of our data, only rigid-body refinement of B factors was performed by using the translation, libration, and screw-rotation (TLS) parameters of the whole oligonucleotide, as described in Refmac.<sup>[8]</sup>

Atomic coordinates have been deposited in the PDB (code 2AF1)

## Acknowledgements

We thank the BM16 staff of ESRF (Grenoble) for assistance in data collection. We are also thankful to Drs. L. Campos and N. Valls for help throughout this work. This work has been supported by grants BIO2002-00317 from the Ministerio de Ciencia y Tecnología and 2001 SGR 00250 from the Generalitat de Catalunya.

**Keywords:** coiled coils · DNA · oligonucleotides · X-ray crystallography

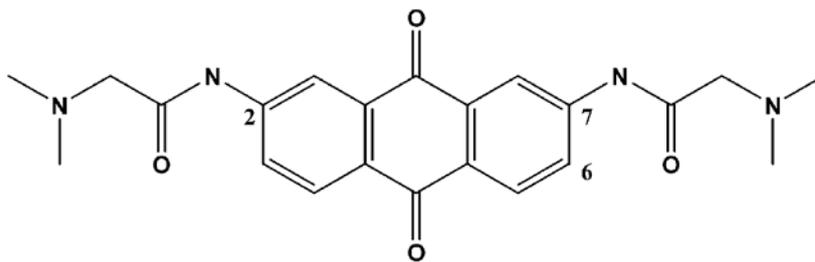
- [1] N. G. A. Abrescia, C. González, C. Gouyette, J. A. Subirana, *Biochemistry* **2004**, *43*, 4092–4100.
- [2] J. L. Campos, L. Urpí, T. Sanmartín, C. Gouyette, J. A. Subirana, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3663–3666.
- [3] R. E. Dickerson, M. L. Kopka, P. Pjura, *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 7099–7103.
- [4] H. Qiu, J. C. Dewan, N. C. Seeman, *J. Mol. Biol.* **1997**, *267*, 881–898.
- [5] Z. Otwinoski, W. Minor, *Methods Enzymol.* **1997**, *276*, 307–326.
- [6] J. Navaza, *Acta Crystallogr. Sect. A Found Crystallogr.* **1994**, *50*, 157–163.
- [7] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse, J. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1998**, *54*, 905–921.
- [8] G. N. Murshudov, A. A. Vagin; E. J. Dodson, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1997**, *53*, 240–255.
- [9] X. J. Lu, W. K. Olson, *Nucleic Acids Res.* **2003**, *31*, 5108–5121.

Received: November 1, 2005

Published online on February 17, 2006

### **III.3 An inverted anthraquinone-DNA crystal structure (article in preparation)**

In this communication we present the structure of an anthraquinone drug (Figure III.3.1) crystallized in association with the tetranucleotide dU<sup>Br</sup>-dA-dG-dG.



**Figure III.3.1.** Chemical structure of the 2,7-Bis[2-(dimethylamino)propionamido]anthracene-9,10-dione.

The crystal was obtained at 37 °C in the following conditions: 0.8 mM DNA, 22 mM NaCacodylate pH 7.0, 15 mM NaCl, 0.8 mM MgCl<sub>2</sub>. Diffraction data were measured with cryocooling at 110 K at the European Synchrotron Radiation Facility in the Spanish beamline BM16. Data from different crystals were measured: crystal 1 (TAGG) was collected at  $\lambda = 0.907 \text{ \AA}$  and diffracted to a maximum resolution of 2.4  $\text{\AA}$ . Crystal 2 was collected up to 3.0  $\text{\AA}$  at three different wavelengths in order to use MAD with Bromine as heavy atom. Wavelengths were recorded from a fluorescence spectrum measured from the crystal. The data were integrated and scaled with the HKL suite (Otwinowsky *et al.*, 1997). The space group turned out to be tetragonal P4<sub>3</sub>2<sub>1</sub>2 with  $a = b = 29.7 \text{ \AA}$  and  $c = 60.88 \text{ \AA}$ . Initial phases were obtained by multiple-wavelength anomalous dispersion (MAD) with the program CNS (Brünger *et al.*, 1998). Two Bromines were detected. The model was manually built and the anthraquinones were inserted in the electron density. The model was then refined using the MAD phases with the CNS package (Brünger *et al.*, 1998). Refinement was concluded with Refmac5 (Murshudov *et al.*, 1997) using data up to 2.8  $\text{\AA}$ . Rigid body refinement of B factors was performed using the TLS parameters of two groups. Dataset and refinement statistics are shown in Table III.3.1.

The oligonucleotide dU<sup>Br</sup>-dA-dG-dG has a partial sequence of human telomere repeat motifs. It had been previously studied by NMR (Kettani *et al.*, 1997) and formed a guanine tetrad. So we thought that it would be a good model in order to study the eventual interaction of anthraquinone drugs with telomere sequences. Instead we found a complex structure in which several anthraquinone drugs are stacked. The structure

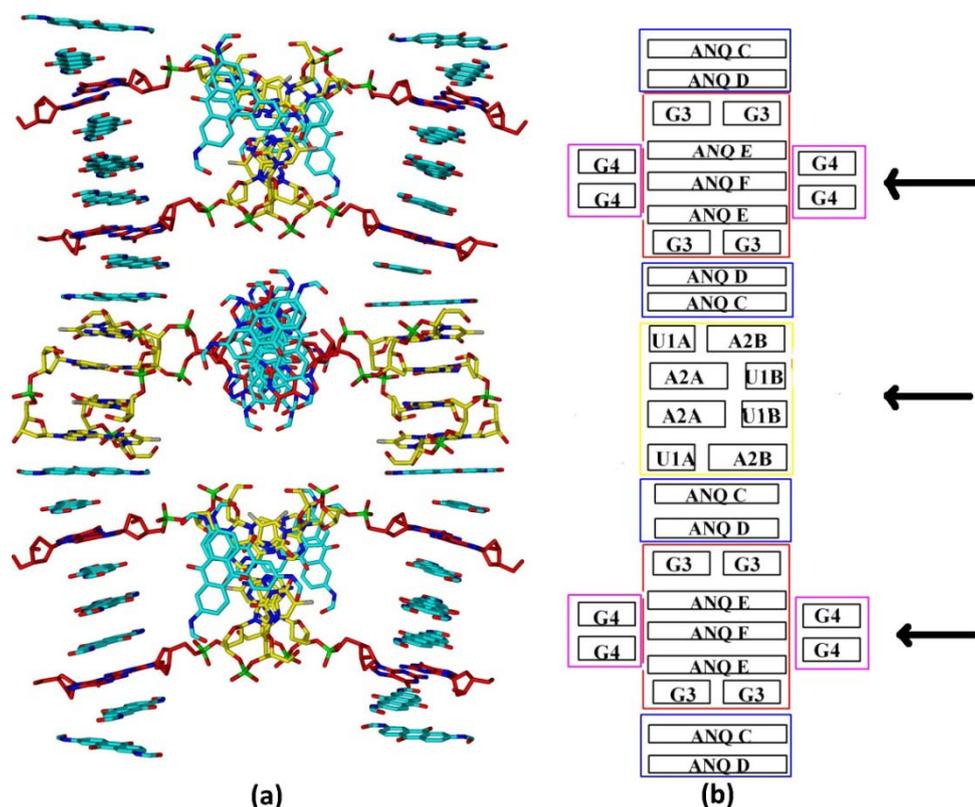
looks like a crystal of anthraquinone in which several base pairs are intercalated, as shown in Figure III.3.2.

The asymmetric unit of this structure contains three and one half anthraquinone molecules and two oligonucleotide chains. The DNA bases form guanine-guanine, adenine-thymine standard and adenine-thymine reverse Watson-Crick base pairs. They are intercalated among the anthraquinone residues. The column of stacked anthraquinone/base pairs is established by several dyad axes as indicated in Figure III.3.2.

Guanine 4 forms an independent group of stacked bases which lie outside of the stacked columns, as indicated in Figure III.3.2 b. They interact with the A-T base pairs in neighbor stacked columns as shown in Figure III.3.3.

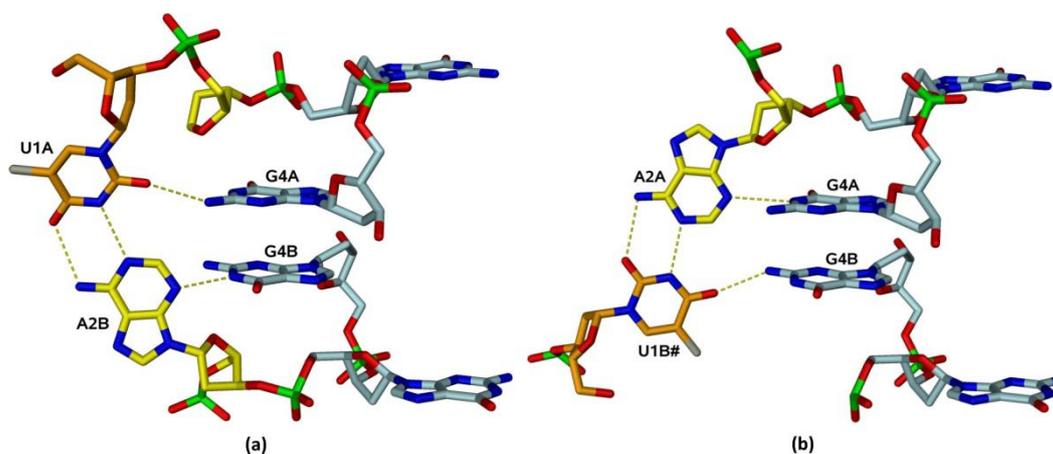
**Table III.3.1.** Crystallographic and refinement statistics. In parenthesis are shown the values for the high resolution shell.

<b>Space group</b>	P4 <sub>3</sub> 2 <sub>1</sub> 2			
<b>Unit cell (Å)</b>				
<i>a = b</i>	29.7	30.67	30.66	30.66
<i>c</i>	60.88	61.78	61.77	61.78
<b>Data collection</b>	Native	Peak	Inflection	Remote
<b>Wavelength (Å)</b>	0.907	0.921	0.920	0.861
<b>Resolution (Å)</b>	2.4	3.0	3.0	3.0
<b>Reflections</b>				
<b>Total</b>	1258	1128	1099	1093
<b>Unique</b>	757	718	700	700
<b>Average redundancy</b>	0.99 (0.98)	0.99 (1)	0.96 (0.74)	0.95 (0.73)
<b>Completeness (%)</b>	97.8 (99.3)	98.7 (100)	73.8 (95.9)	73.1 (95.2)
<b>R<sub>sym</sub> (%)</b>	3.5 (28.4)	6.9 (13.7)	7 (15.4)	6.6 (17)
<b>Average I/σ (I)</b>	19.38 (4.55)	11.61 (5)	12.62 (6.52)	12.37 (6.08)
<b>Refinement</b>				
<b>Resolution range</b>	28-2.8			
<b>Final R/R<sub>free</sub> (%)</b>	23.2/25.9			



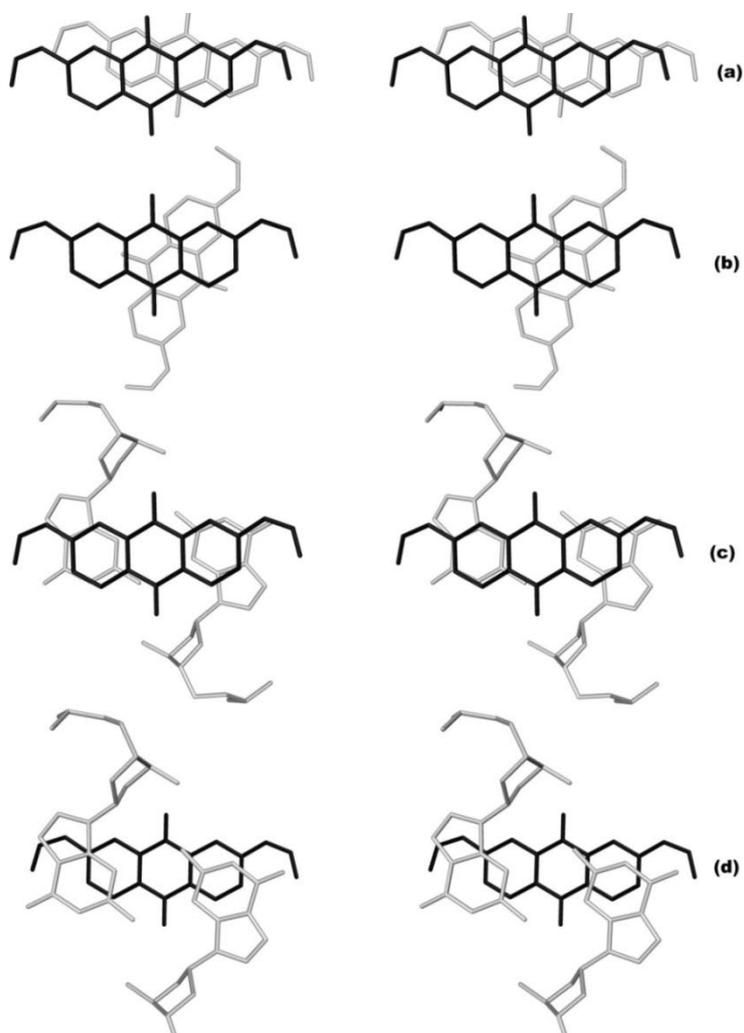
**Figure III.3.2.** (a) Perpendicular columns in the crystal; guanines 4 at the crossing points are omitted. (b) Schematic representation of one column. The dyad axes are shown.

This work demonstrates that short oligonucleotides may adopt quite unexpected shapes. This structure contains several unusual features which are not found in larger oligonucleotides. They have been mentioned above and are shown in Figure III.3.2, Figure III.3.3 and Figure III.3.4.



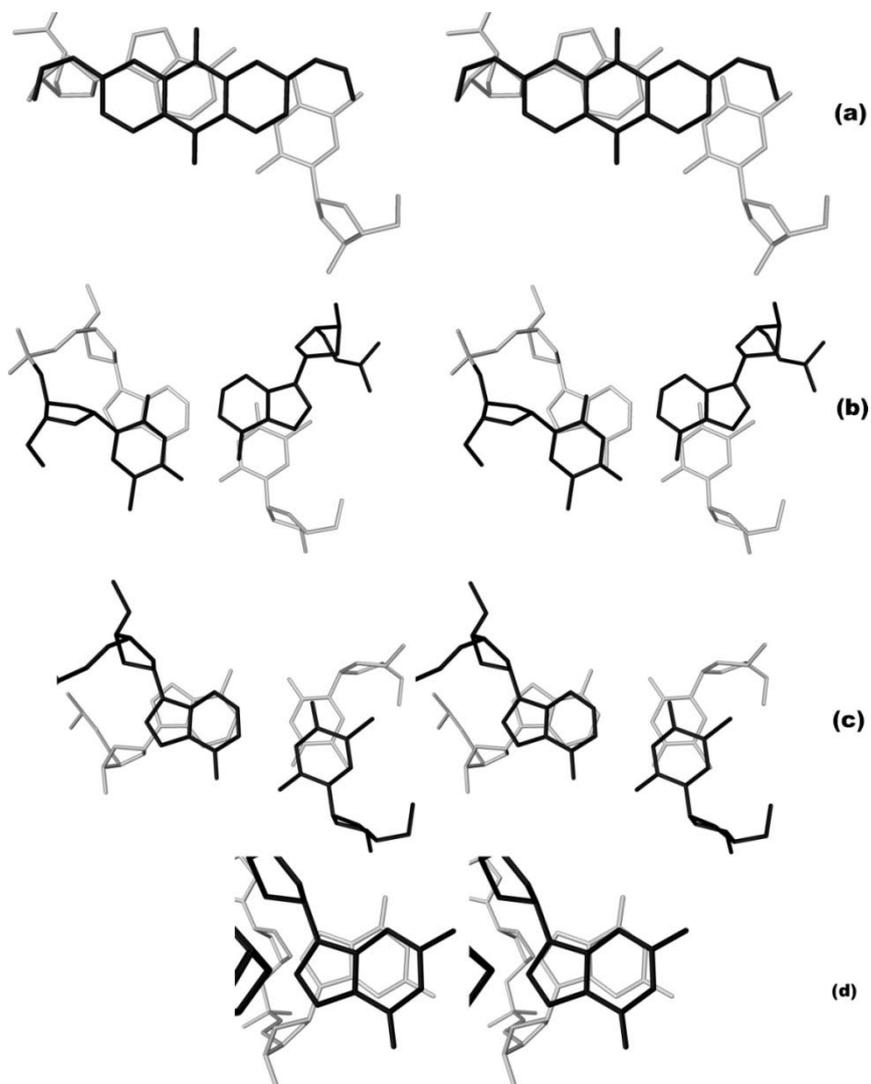
**Figure III.3.3.** (a) Standard A-T Watson-Crick hydrogen bonds. (b) Reverse Watson-Crick hydrogen bonds. Guanines 4 stabilize the A-T base pairs at the crossing points.

The structure also shows several unexpected stacking features, as shown in Figure III.3.4.



**Figure III.3.4.** Stereo views of stacking interactions in the crystal: (a) anthraquinones E and F; (b) anthraquinones C and D; (c) anthraquinone E and guanines 3; (d) guanines 3 and anthraquinone D.

Additional stacking patterns are shown in Figure III.3.5.



**Figure III.3.5.** Stereo views of stacking interactions in the crystal: (a) anthraquinone C/A-T (Watson-Crick base pair); (b) A-T (W-C)/A-T (R-WC); (c) A-Ts (R-WC); (d) guanines 4 stacking interaction.

In particular the drug/drug stacking features are completely different in two regions, as shown in Figure III.3.4 a and b. Stacking on the guanine-guanine base pairs is also different at both sides. At one side there is very good stacking (Figure III.3.4 c), whereas at the opposite side stacking is not so complete (Figure III.3.4 d). These features are complementary to those described in previous studies and should be taken into account for a better understanding of drug/DNA stacking features. They demonstrate a large variety of interactions which cannot be easily understood.

We have compared our structure with that found in some anthraquinone derivative crystals extracted from the Cambridge Structural Data Base (CSD). The structure is strongly influenced by the size and polarity of the substituents attached to the anthraquinone rings. Anthraquinones with small lateral substituents usually pack as infinite stacked columns, usually slanted (Kingsford *et al.*, 1995; Il'in *et al.*, 1975; Janczak, 1995; Fu *et al.*, 1998), similar to the situation represented in Figure III.3.4 a. In some cases the inclination is very large (Popova *et al.*, 1975), so that ring 1 stacks on ring 3. In one case (de Ruyck *et al.*, 2006), although the substituents are small (two OH groups), the anthraquinone molecules are stacked in an approximate perpendicular fashion, as in Figure III.3.4 b. Cases of staggered stacking are also found (de Abreu, 2002). Finally, in the presence of bulky substituents no stacking is found (Agbandje *et al.*, 1992). In summary, the variability of stacking features present in our structure (Figure III.3.4) is reminiscent of what is found in crystals of anthraquinone derivatives.

## IV. BIBLIOGRAPHY

**Abrescia, N. G. A., Gonzáles, C., Gouyette, C. & Subirana, J. A.** (2004). X-Ray and NMR studies of the DNA oligomer d(ATATAT): Hoogsteen base pairing in duplex DNA. *Biochemistry*, **43**, 4092-4100.

**Abrescia, N. G. A., Thompson, A., Huynh-Dinh, T. & Subirana, J. A.** (2002). Crystal structure of an antiparallel DNA fragment with Hoogsteen base pairing. *Proc. Natl. Acad. Sci. USA*, **99**, 2806-2811.

**Abrescia, N. G., Malinina, L., Fernandez, L. G., Huynh-Dinh, T., Neidle, S. & Subirana J. A.** (1999). Structure of the oligonucleotide d(CGTATATACG) as a site-specific complex with nickel ions. *Nucleic Acids Res.*, **27**, 1593-1599.

**Abrescia, N. G., Malinina, L. S. & Subirana J. A.** (1999). Stacking interaction of guanine with netropsine in the minor groove of d(CGTATATACG)<sub>2</sub>. *Journal of Mol. Biol.*, **294(3)**, 657-666.

**Agbandje, M., Jenkins, T. C., McKenna, R., Reszka, A. P. & Neidle S.** (1992). Anthracene-9,10-diones as potential anticancer agents. Synthesis, DNA-binding, and biological studies on a series of 2,6-disubstituted derivatives. *J. Med. Chem.*, **35**, 1418 – 1429.

**Amemiya, Y.** (1997). X-ray Storage-Phosphor Imaging-Plate Detectors : High-Sensitivity X-ray Area detector. *Methods in Enzymol.*, **276**, 233-243.

**Asherie, N.** (2004). Protein crystallization and phase diagrams. *Methods*, **34**, 266-272.

**Baikalov, I. & Dickerson, R. E.,** (1998) Molecular Replacement using DNA helical symmetry, *Acta Cryst. D***54**, 324-333.

**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.

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

**Brünger, A. T., Adams, P. D. & Rice, L. M.** (1998) New Applications of simulated annealing in crystallographic refinement. *Direct Methods for Solving Macromolecular Structures*, 143-157.

**Brunger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L.** (1998). Crystallography & NMR system (CNS): A new software system for macromolecular structure determination. *Acta Cryst.*, **D54**, 905-921.

**Bunting, K. A., Roe, S. M., Headley, A., Brown, T., Savva, R. & Pearl, L.H.** (2003). Crystal structure of the *Escherichia coli dcm* very-short-patch DNA repair endonuclease bound to its reaction product-site in a DNA superhelix. *Nucleic Acids Research*, **31**(6), 1633-1639.

**Campos, J. L., Urpí, L., Sanmartín, T., Gouyette, C. & Subirana, J. A.** (2005). DNA coiled-coils. *Proc. Natl. Acad. Sci. USA*, **102**, 3663-3666.

**Campos, L., Valls, N., Urpi, L., Gouyette, C., Sanmartin, T., Richter, M., Alechaga, E., Santaolalla, A., Baldini, R., Creixell, M., Ciurans, R., Skokan, P., Pous, J. & Subirana, J. A.** (2006). Overview of the structure of all-AT oligonucleotides: organization in helices and packing interactions, *Biophysical Journal*, **91**, 892-903.

**Cerius2** (Acelrys Inc, San Diego, CA).

**Choo, K. H. A.** (1997). The Centromere, Oxford Univ. Press, Oxford.

**Cochran, W., Crick, F. H. & Vand, V.** (1952). The structure of synthetic polypeptides. I. The transform of atoms on a helix. *Acta Cryst.* **5**, 581-586.

- 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.
- Dauter, Z.** (1999). Data-collection strategies. *Acta Cryst.*, **D55**, 1703-1717.
- de Abreu, F. C., de O. Lopes, A., Alves Pereira, M., De Simone, C. A. & Goulart M. O. F.** (2002). Nitrooxyquinones: synthesis, X-ray diffraction and electrochemical studies. *Tetrahedron Lett.*, **43**, 8153-8157.
- De Luchi, D., Gouyette, C. & Subirana, J. A.** (2003). The influence of size on the thermal stability of oligonucleotides: the case of AT sequences. *Analytical Biochemistry*, **322**, 279-282.
- De Luchi, D., Tereshko, V., Gouyette, C. & Subirana, J. A.** (2006). Structure of the DNA coiled-coil formed by d(CGATATATATAT). *ChemBioChem*, **7**, 585-587.
- DeLano, W. L.** (2003). The PyMol molecular graphic system. *DeLano Scientific LLC*, San Carlos, CA, USA.
- Dickerson, R. E., Kopka, M. L. & P Pjura** (1983) A random-walk model for helix bending in B-DNA. *Proc Natl Acad Sci U S A.*, **80**(23), 7099–7103.
- DINO** (A. Philippsen, <http://www.dino3d.org>).
- Fraser, R. D. B., MacRae, T. P. & Miller, A.** (1964). A coiled-coil model of alpha-keratin structure. *J. Mol. Biol.*, **10**, 147-156.
- Fraser, R. D. B., MacRae, T. P. & Miller, A.** (1964). The Fourier transform of the coiled-coil model for  $\alpha$ -keratin. *Acta Cryst.*, **17**, 813-816.
- Fu, Y. & Brock, C. P** (1998). Temperature Dependence of the Rigid-Body Motion of Anthraquinone. *Acta Cryst.* **B54** (3), 308–315.

**Gorin, A. A., Zhurkin, V. B. & Olson, W. K.** (1995). B-DNA twisting correlated with base-pair morphology. *J. Mol. Biol.*, **247**, 34-48.

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

**Howlin, B., Butler, S.A., Moss, D.S., Harris, G.W. & Driessen, H.P.C.** (1993). TLSANL: TLS parameter analysis program for segmented anisotropic refinement of macromolecular structures. *J. Appl. Cryst.*, **26**, 622-624.

**Il'in, S. G., Chetkina, L. A. & Golder, G. A.** (1975). *Kristallografiya*, **20**, 1051.

**International Tables for X-ray Crystallography** (1992). Vol C, edited by A. J. C. Wilson. Dordrecht: Kluwer Academic Publishers.

**Janczak, J.** (1995). 2-Aminoanthraquinone. *Acta Cryst. C* **51** (7), 1381–1382.

**Jensen, H. L.** (1997) Refinement and reliability of macromolecular models based on X-ray diffraction data. *Methods in Enzymol.* **366**, 277-353.

**Jérôme de Ruyck, Julien Preat, Eric A. Perpète, Denis Jacquemin, Johan Wouters** (2006) 2,6-Dihydroxyanthraquinone: an isomer of the well known alizarin dye. *Acta Cryst. E* **62** (10), o4503–o4505.

**Kingsford-Adaboh, R. & Kashino, S.** (1995). Disordered Structure of 2-Methylantraquinone. *Acta Cryst. C* **51** (10), 2094–2096.

**Konnert, J. H. & Hendrickson, W. A.** (1980). A restrained-parameter thermal-factor refinement procedure. *Acta Cryst.*, **A36**, 344-350.

**Konnert, J. H.** (1976). A restrained-parameter structure-factor least-squares refinement procedure for large asymmetric units. *Acta Cryst.*, **A32**, 614-617.

**Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitz Hugh, W., et al.** (2001). Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.

- Leontis N. B. & Westhof E.** (1998). A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. *J. Mol. Biol.*, **283**, 571-583.
- Leslie, A. G. W.** (1992). Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 and ESF-EAMCB *Newsletter on Protein Crystallography*, **26**.
- Liebich, I., Bode, J., Reuter, I. & Wingender, E.** (2002). Evaluation of sequence motifs found in scaffold/matrix-attached regions (S/MARs). *Nucleic Acids Res.*, **30**, 3433-3442.
- Liu, J., Malinina, L. & Subirana, J. A.** (1998). The structure of the most studied DNA fragment changes under the influence of ions: a new packing of d(CGCGAATTCGCG). *FEBS letters*, **438** (3), 211-214.
- Liu, J. & Subirana, J. A.** (1998). Structure of d(CGCGAATTCGCG) in the presence of Ca<sup>2+</sup> ions. *J. Biol. Chem.*, **274** (35), 24749-24752.
- 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.
- M. F Perutz** (1990) How W. L. Bragg invented X-ray analysis *Acta Cryst.*, **A46** (8), 633-643.
- Mc. Pherson, A.** (1999). Crystallization of biological macromolecules. CSHL Press, Cold Spring Harbor.
- 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.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J.** (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.*, **D53**, 240-255.

- Murshudov, G. N., Vagin, A. A. & Dodson, E. J.** (1997). Refinement of Macromolecular structures by Maximum likelihood method. *Acta Cryst.*, **D53**, 240-255.
- Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J.** (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Cryst.*, **D55**, 247-255.
- Naday, I., Westbrook, E. M., Westbrook, M. L., Travis, D. J., Stanton, M., Phillips, W. C. & Xie, J.** (1994). Characterization and data collection on a direct-coupled CCD x-ray detector. *Nucl. Instrum. and Meth.* **A348**, 635-640.
- Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L. & Aggarwal, A. K.** (2004). Replication by human DNA polymerase- $\alpha$  occurs by Hoogsteen base-pairing. *Nature*, **430**, 377-380.
- Navaza, J.** (1994). AMoRe: an automated package for molecular replacement. *Acta Cryst.*, **A50**, 157-163.
- Navaza, J.** (1999). Implementation of molecular replacement in AMoRe. *Acta Cryst.*, **D57**, 1593-1599.
- Nave, C.** (1999). Matching X-ray source, optics, and detectors to protein crystallography requirements. *Acta. Cryst.* **D55**, 1663-1668.
- Needle, S.** (1999). Oxford Handbook of Nucleic Acid Structure. Oxford University Press. Edited by Stephen Neidle.
- Nobrega, M. A., Ovcharenko, I., Afzal V. & Rubin, E. M.** (2003). Scanning human gene deserts for long-range enhancers. *Science*, **302**, 413.
- 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.

**Popova, E. G., Chetkina, L. A. & Dzyabchenko, A. V.** (1975). *Kristallografiya*, **20**, 931.

**Pous, J., Urpì, L., Subirana, J. A., Gouyette, C., Navaza, J. & Campos J. L.** (2008). Stabilization by extra-helical Thymines of a DNA duplex with Hoogsteen base pairs. *JACS*, **130**, 6755-6760.

**Qiu, H., Dewan, J. C. & Seeman, N. C.** (1997). The Crystal Structure of d-CGACGATCGT. *J. Mol. Biol.*, **267**, 881-898.

**Rhodes, G.**, (2006) Crystallography made crystal clear, Elsevier.

**Sáinz, J., Azorín, F. & Cornudella, J.** (1989). Detection and molecular cloning of highly repeated DNA in the sea cucumber sperm. *Gene*, **80**, 57-64.

**Schomaker V. & Trueblood, K.N.** (1968). Efficient anisotropic refinement of Macromolecular structures using FFT. *Acta Cryst.*, **B24**, 63-76.

**Stanton, M., Phillips, W. C., O'Mara, D., Naday, I. & Westbrook, E. M.** (1993). Area detector design II: application to a modular CCD-based detector for x-ray crystallography. *Nucl. Instrum. and Meth.* **A325**, 558-567.

**Suwalsky, M., Traub, W., Shmueli, U. & J. A. Subirana** (1969). An X-ray study of the interaction of DNA with Spermine. *J. Mol. Biol.* **42**, 363-373.

**Valls, N., Wright, G., Steiner, R. A., Murshudov, G. N. & Subirana J. A.** (2004). DNA variability in five crystals structures of d(CGCAATTGCG). *Acta Cryst.*, **D60**, 680-685.

**Van Meerssche, M. & Feneau-Dupont, J.** (1984). Introduction à la CRISTALLOGRAPHIE et à la CHIMIE STRUCTURALE. Editions Peeters, Paris.

**Viswamitra, M. A., Shakked, Z., Sheldrick, J. G. M., Salisbury, S. A. & Kennard, O.** (1982). Structure of the deoxytetranucleotide d-pApTpApT and a sequence-dependent model for poly(dA-dT). *Biopolymers*, **21**, 513-533.

**Watson J.D. and Crick F.H.C.** (1953). A Structure for Deoxyribose Nucleic Acid. *Nature*, **171**, 737-738.

**Westbrook, E. M. & Naday, I.** (1997) CCD Area Detectors. *Methods in Enzymol*, **276**, 244-268.

**Wilkins M.H.F., A.R. Stokes A.R. & Wilson, H.R.** (1953). Molecular Structure of Deoxypentose Nucleic Acids, *Nature*, **171**, 738-740.

**Wilson, H. R.** (1990). Diffraction of X-rays of proteins, nucleic acid and viruses. Edwald Arnold Publishers. London.

**Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E.** (1980). Crystal structure analysis of a complete turn of B-DNA. *Nature*, **287**, 755–758.

The three-dimensional molecular graphics in this work have been produced with the programs Cerius2 (Accelrys, San Diego), DINO (A. Philippsen, <http://www.dino3d.org>) and PYMOL (DeLano, 2003).