UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL ATOMIC ENERGY AGENCY
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/916 - 26

**SEVENTH COLLEGE ON BIOPHYSICS:**
***Structure and Function of Biopolymers: Experimental and Theoretical Techniques.***
***4 - 29 March 1996***

# *Molecular Dynamics of DNA*

Richard LAVERY
Laboratoire de Biochimie Théorique, CNRS
Institut de Biologie Physico-Chimique
13 Rue Pierre et Marie Curie
Paris 75005
FRANCE

# 7th College on Biophysics, Structure & Function of Biopolymers
# Experimental and Theoretical Techniques

## ICTP, Trieste 1996

## MOLECULAR DYNAMICS OF DNA

**Richard Lavery**
**(richard@ibpc.fr)**

**Laboratoire de Biochimie Théorique, CNRS**
**Institut de Biologie Physico-Chimique**
**13 rue Pierre et Marie Curie, Paris 75005**
**France**

### Lecture I

Simulation methods for modelling biopolymers in general and nucleic acids in particular. Force fields, coordinate systems, energy optimisation and dynamics.

### Lecture II

Results of simulations. Base sequence effects.
Conformational analysis.

### Lecture III

Deformations in DNA. Curvature. Base pair opening.
Nanomanipulation. Protein-nucleic acid interactions.

# MODELING NUCLEIC ACIDS: FINE STRUCTURE, FLEXIBILITY AND CONFORMATIONAL TRANSITIONS

RICHARD LAVERY
**Laboratoire de Biochimie Théorique, CNRS**
**Institut de Biologie Physico-Chimique, Paris**

## INTRODUCTION

It is not an exaggeration to say that nucleic acids are at the very heart of life. Nucleic acids are the storage medium for all genetic information and, via protein synthesis, direct the functioning of our cells. With the appearance of the human genome project, nucleic acids are very much in the news and available sequence information is now increasing at an exponential rate. Despite these facts our knowledge of the structure and dynamics of nucleic acids and how they perform their biological roles at an atomic level of detail is still very limited. The beautiful double helix of DNA was proposed 40 years ago (Watson & Crick, 1953), but the first single crystal of a DNA oligomer clearly showing base sequence effects dates from only 12 years ago (Drew et al, 1981; Dickerson & Drew, 1981). The picture has also been complicated by the discovery of many other fascinating structural forms of DNA including left-handed and parallel-stranded double helices, triple and quadruple helices and cruciforms, and this list will surely continue to grow. Since the first crystal structure of a transfer RNA we have also seen that nucleic acids can fold into compact forms just as complex as the folding of globular proteins and that, in the case of the ribozymes, they even share with proteins the ability to catalyse chemical reactions (Uhlenbeck, 1991).

During a meeting which addressed the problem of unusual DNA structures (Wells & Harvey, 1988), Richard Dickerson stated "To me, the most important single question that a molecular biologist can ask about DNA structure today is: How does base sequence influence helix geometry? ... The model of DNA as a straight line in a genetic map has been of enormous practical utility. But DNA is a real molecule, with a real structure, and behaviour that follows from that structure" (Dickerson, 1988). This question is still very far from being answered. Although much useful information has come from crystal structures, the number of oligomers studied increases rather slowly. The elongated form of the double helix also hinders the application of NMR to the study of nucleic acids, the information gathered being insufficient to define a detailed structure without input from other sources.

It thus appears that molecular modeling, if it is of a sufficiently high quality, can make a useful contribution to understanding nucleic acid structure and dynamics. However, achieving the necessary level of quality is not easy. Experimentally nucleic acids are rather fragile macromolecules, sensitive to

their environment of solvent and counterions, able to undergo major structural transitions between different allomorphic forms as well as transient deformations such as base pair opening. They can be deformed in many ways, through interactions with drugs or proteins, by supercoiling or by bending. Successful modeling should be able to reproduce all of these phenomena.

If the appropriate techniques can be developed there is no shortage of important problems to be addressed. Knowing how base sequence influences the static and dynamic structure of DNA would lead us to an understanding of how target sites are recognized by proteins, drugs or toxic agents. Understanding DNA deformation, and notably bending, would help in explaining the packaging of DNA within cells as well as elements of genetic control. Studying unusual DNA structures could lead to new artificial way of controlling gene expression (Hélène & Toulmé, 1990; Strobel & Dervan, 1990) with major clinical consequences. The list could be continued.

For the last 8 years or so our group has been involved in attempting to answer some of these problems with the help of molecular modeling. Rather than using existing general-purpose modeling techniques, we have, in most cases, attempted to develop new approaches which are specifically adapted to studying nucleic acids. This choice has influenced the direction taken by our research and the viewpoint that we have developed concerning nucleic acid behaviour. In consequence, the present chapter on modeling nucleic acids will not be an impersonal review of the field. it is rather a personal overview of the main results we have obtained and how we look at possible future developments. This naturally does not exclude discussions of much important work carried out in other laboratories, but no attempt is made to provide an exhaustive summary. It should also be added that almost all the results discussed concern DNA, although the modeling techniques used can equally well be applied to RNA. This choice is again influenced by our own studies, but it should be noted that much exciting work is now underway in the RNA field (Westhof et al, 1990; Michel & Westhof, 1990).

The organisation of the chapter begins with a discussion of the combinatorial aspects of base sequences which is aimed at defining more clearly how to approach the study of sequence effects on nucleic acid structure. We will then take a large scale look at the different structural classes that can be formed by nucleic acid helices. This is followed by a section on how to describe such helices in molecular detail. Obtaining such descriptions turns out to be vital, since only a quantitative structural analysis can enable different DNA conformations to be compared, thus bringing sequence dependent effects to light. The techniques used for the computer modeling of DNA are then discussed, pointing out both those features which work well and the areas where development is still needed. With this knowledge in hand, we pass to theoretical results on sequence effects, firstly concerning the optimal energy conformations of both regular and irregular sequences and then aspects of flexibility and conformational transitions. The final section is a

discussion of possible developments in theory and its application to nucleic acids which may be expected in the coming years.

## BASE SEQUENCES

In order to get an idea of how to approach the possible effects of base sequence on nucleic acid structure, it is interesting to start by considering the number of sequences which may be formed from a given number of bases. If we begin with dinucleotide sequences, table 1 shows that although there are $4^2$=16 ways of building a dinucleotide sequence, there are, in fact, only 10 unique dinucleotides. This is because 6 pairs of dinucleotides are simply the "Watson" and "Crick" strands of the same sequences. These degenerate pairs are shown in lower case in table 1. (Note that the choice to conserve, for example, the notation TG rather than CA is arbitrary. We have chosen notations containing purines in preference to pyrimidines and guanine in preference to adenine - following the order of bases G, A, C, T used to construct table 1). The remaining 4 dinucleotides GC, CG, AT and TA are not degenerate because they show inversion symmetry between their two strands.

In the case of trinucleotide sequences, the situation is simpler since inversion symmetry is ruled out for an odd number of bases (symmetry, in this case, would imply that the central base pair of the sequence would have to be formed from two identical bases). Every trinucleotide sequence therefore has a degenerate partner and the number of unique sequences is $4^3/2$=32. The trinucleotide sequences are also shown in table 1, divided into 2 matrices depending on the central base of the sequence. We have used the same rules described for dinucleotides for naming each sequence.

Table 1 contains one further step in complexity, leading us to tetranucleotide sequences. In this case there are $4^4$=256 possible combinations, which actually give rise to 136 unique sequences. These sequences are shown in upper case in the table - grouped according to their central dinucleotide. How the number of unique sequences arises can be understood on the basis of the central dinucleotides. If the dinucleotide has a degenerate partner (see above), each tetranucleotide formed around this dinucleotide will also have a degnerate tetranucleotide. These cases concern 6 dinucleotide sequences and consequently lead to 6x16=96 degenerate tetranucleotides. With the 4 remaining dinucleotide sequences, GC, CG, AT and TA, inversion symmetry leads to 6 degeneracies within the associated tetranucleotides (shown in lower cases in table 1). We thus have to remove a total of 96+4x6=120 degeneracies from the total of 256 combinations, leading to 136 unique tetranucleotide sequences. Finally, following the rule given above, we can rapidly calculate that their are 512 ($4^5/2$) unique pentanucleotides, but we will not list these sequences here.

Where does this exercise in combinatorial algebra lead us. Firstly, we can note that in the simplest approximation to DNA fine structure we would

Table 1. Unique sequences as a function of sequence length (degenerate combinations are indicated by lowercase letters)

| | G | A | C | T | ... 10 Dinucleotides |
|---|---|---|---|---|---|
| G | GG | GA | GC | GT | |
| A | AG | AA | ac | AT | |
| C | CG | ca | cc | ct | |
| T | TG | TA | tc | tt | |

| (G) | G | A | C | T | ... 32 Trinucleotides |
|---|---|---|---|---|---|
| G | GGG | GGA | GGC | GGT | |
| A | AGG | AGA | AGC | AGT | |
| C | CGG | CGA | CGC | CGT | |
| T | TGG | TGA | TGC | TGT | |

| (A) | G | A | C | T |
|---|---|---|---|---|
| G | GAG | GAA | GAC | GAT |
| A | AAG | AAA | AAC | AAT |
| C | CAG | CAA | CAC | CAT |
| T | TAG | TAA | TAC | TAT |

| (GG) | G | A | C | T | ... 136 Tetranucleotides |
|---|---|---|---|---|---|
| G | GGGG | GGGA | GGGC | GGGT | |
| A | AGGG | AGGA | AGGC | AGGT | |
| C | CGGG | CGGA | CGGC | CGGT | |
| T | TGGG | TGGA | TGGC | TGGT | |

| (GA) | G | A | C | T |
|---|---|---|---|---|
| G | GGAG | GGAA | GGAC | GGAT |
| A | AGAG | AGAA | AGAC | AGAT |
| C | CGAG | CGAA | CGAC | CGAT |
| T | TGAG | TGAA | TGAC | TGAT |

| (GT) | G | A | C | T |
|---|---|---|---|---|
| G | GGTG | GGTA | GGTC | GGTT |
| A | AGTG | AGTA | AGTC | AGTT |
| C | CGTG | CGTA | CGTC | CGTT |
| T | TGTG | TGTA | TGTC | TGTT |

| (AA) | G | A | C | T |
|---|---|---|---|---|
| G | GAAG | GAAA | GAAC | GAAT |
| A | AAAG | AAAA | AAAC | AAAT |
| C | CAAG | CAAA | CAAC | CAAT |
| T | TAAG | TAAA | TAAC | TAAT |

| (AG) | G | A | C | T |
|---|---|---|---|---|
| G | GAGG | GAGA | GAGC | GAGT |
| A | AAGG | AAGA | AAGC | AAGT |
| C | CAGG | CAGA | CAGC | CAGT |
| T | TAGG | TAGA | TAGC | TAGT |

| (TG) | G | A | C | T |
|---|---|---|---|---|
| G | GTGG | GTGA | GTGC | GTGT |
| A | ATGG | ATGA | ATGC | ATGT |
| C | CTGG | CTGA | CTGC | CTGT |
| T | TTGG | TTGA | TTGC | TTGT |

| (GC) | G | A | C | T |
|---|---|---|---|---|
| G | GGCG | GGCA | GGCC | ggct |
| A | AGCG | AGCA | AGCC | AGCT |
| C | CGCG | CGCA | cgcc | cgct |
| T | tgcg | TGCA | tgcc | tgct |

| (CG) | G | A | C | T |
|---|---|---|---|---|
| G | GCGG | GCGA | GCGC | gcgt |
| A | ACGG | ACGA | ACGC | ACGT |
| C | CCGG | CCGA | ccgc | ccgt |
| T | tcgg | TCGA | tcgc | tcgt |

| (AT) | G | A | C | T |
|---|---|---|---|---|
| G | GATG | GATA | GATC | gatt |
| A | AATG | AATA | AATC | AATT |
| C | CATG | CATA | catc | catt |
| T | tatg | TATA | tatc | tatt |

| (TA) | G | A | C | T |
|---|---|---|---|---|
| G | GTAG | GTAA | GTAC | gtat |
| A | ATAG | ATAA | ATAC | ATAT |
| C | CTAG | CTAA | ctac | ctat |
| T | ttag | TTAA | ttac | ttat |

assume that inter-base pair parameters such as twist and rise can be determined by the dinucleotide sequence alone. In this case there would to 10 cases to study. In a next nearest neighbour approximation, which is certainly more realistic (Calladine, 1982; Yanagi et al, 1991), we would have to pass to tetranucleotide sequences and, thus, to 136 different cases. Similarly for intra-base pair parameters such as propeller or buckle, the nearest neighbour approach would lead us to consider 32 different trinucleotides, whereas the next nearest neighbour approximation would require consideration of the 512 unique pentanucleotides. These numbers put the sequence dependent structure problem in perspective and also underline the that fact that the few existing crystallographic structures of DNA oligomers only give us a first glimpse of possible sequence effects. The problem is thus a daunting one from either an experimental or a theoretical point of view.

What set of DNA oligomers would be best for looking at sequence effects. One choice would be to build the elements under study, for example tetranucleotides, into a longer oligomer XXXXαβγδXXXX in order to maintain a proper double helical environment and to avoid fraying of the ends of the tetranucleotide. Since we would be functioning under a next nearest neighbour approximation (looking only at the properties of the central 'βγ' dinucleotide), the nature of the ends 'XXXX' should be unimportant and we would have 136 different oligomers (containing the 136 unique tetranucleotides) to study. In fact one can do better than this by taking advantage of repeating sequences.

If we return to the 10 unique dinucleotides, table 2 shows that if we make use of oligomers with either mononucleotide or dinucleotide repeating sequences, then we can look at the 10 dinucleotides using only 6 oligomers. Since these oligomers have repeating sequences, we can also profit during modeling by imposing helical symmetry. This has the double advantage of reducing the number of independent variables representing the conformation and of avoiding structural end-effects - note that we can also make use of the central energy approximation described in the theory section, to eliminate energetic end-effects and to speed up the calculations.

Following a similar procedure, we can obtain information on the 32 unique trinucleotide sequences with only 10 further oligomers. Note that, in this case, we have taken advantage of the fact 8 trinucleotides (underlined in table 1) are already contained in the oligomers used for studying dinucleotides. Now lets take one further step to tetranucleotides. We can again profit from the fact that certain tetranucleotides have already been treated within di- or tri-nucleotide sequences (these cases have respectively double and single underlines in table 1). The remaining cases can be investigated using a total of 33 repetitive oligomers shown in table 2.

Overall we can see that 6+10+33=49 symmetric oligomers are sufficient to study all unique sequences up to the tetranucleotide level, many fewer than the 136 simple "box" calculations described above. An additional advantage of this technique (which can also be extended to longer sequences such as

pentanucleotides) is that successively longer sequences can be treated in separate projects, the results of each stage contributing to the following stage. Studying the 10 dinucleotides, thus also gives information on 8 trinucleotides and 10 tetranucleotides, while studying the remaining trinucleotides already gives us information on a further 30 tetranucleotides.

Table 2. Regular oligomers contining all unique sequences of 2, 3 and 4 bases.

(a) Dinucleotide sequences: 6 oligomers

| | | |
|---|---|---|
| CGCG.. | TATA.. | TGTG.. |
| GAGA.. | GGGG.. | AAAA.. |

(b) Trinucleotide sequences: 10 oligomers

| | | |
|---|---|---|
| GGAGGA... | GGCGGC... | GGTGGT... |
| AACAAC... | AATAAT... | AAGAAG... |
| CGACGA... | TGATGA... | AGCAGC... |
| AGTAGT... | | |

(c) Tetranucleotide sequences: 33 oligomers

| | | |
|---|---|---|
| GGGAGGGA.... | GGGCGGGC.... | GGGTGGGT.... |
| AAAGAAAG.... | AAACAAAC.... | AAATAAAT.... |
| GGCCGGCC.... | GGAAGGAA.... | GGTTGGTT.... |
| AATTAATT.... | GGACGGAC.... | GGCAGGCA.... |
| GGTAGGTA.... | GGATGGAT.... | GGTCGGTC.... |
| GGCTGGCT.... | AAGCAAGC.... | AACGAACG.... |
| AACTAACT.... | AATCAATC.... | AAGTAAGT.... |
| AATGAATG.... | GAGTGAGT.... | GAGCGAGC.... |
| GCGTGCGT.... | AGATAGAT.... | ACAGACAG.... |
| ACATACAT.... | GCTAGCTA.... | GCATGCAT.... |
| GATCGATC.... | GACTGACT.... | GTACGTAC.... |

A last remark can be made concerning the next nearest neighbour level approximation, that is, tetranucleotide sequences for inter-base pair parameters and pentanucleotide sequences for intra-base pair parameters. If it is assumed that the next nearest neighbour effect is less subtle than that of the nearest neighbours, it might also be assumed that only differences between purines and pyrimidines can be distinguished within the outer base pairs. If this were true, the 136 tetranucleotides and the 512 pentanucleotides could be reduced to respectively 36 and 128 cases. This would represent an important simplification, but whether such an approximation is justified remains to be seen.

# DESCRIBING NUCLEIC ACID STRUCTURE

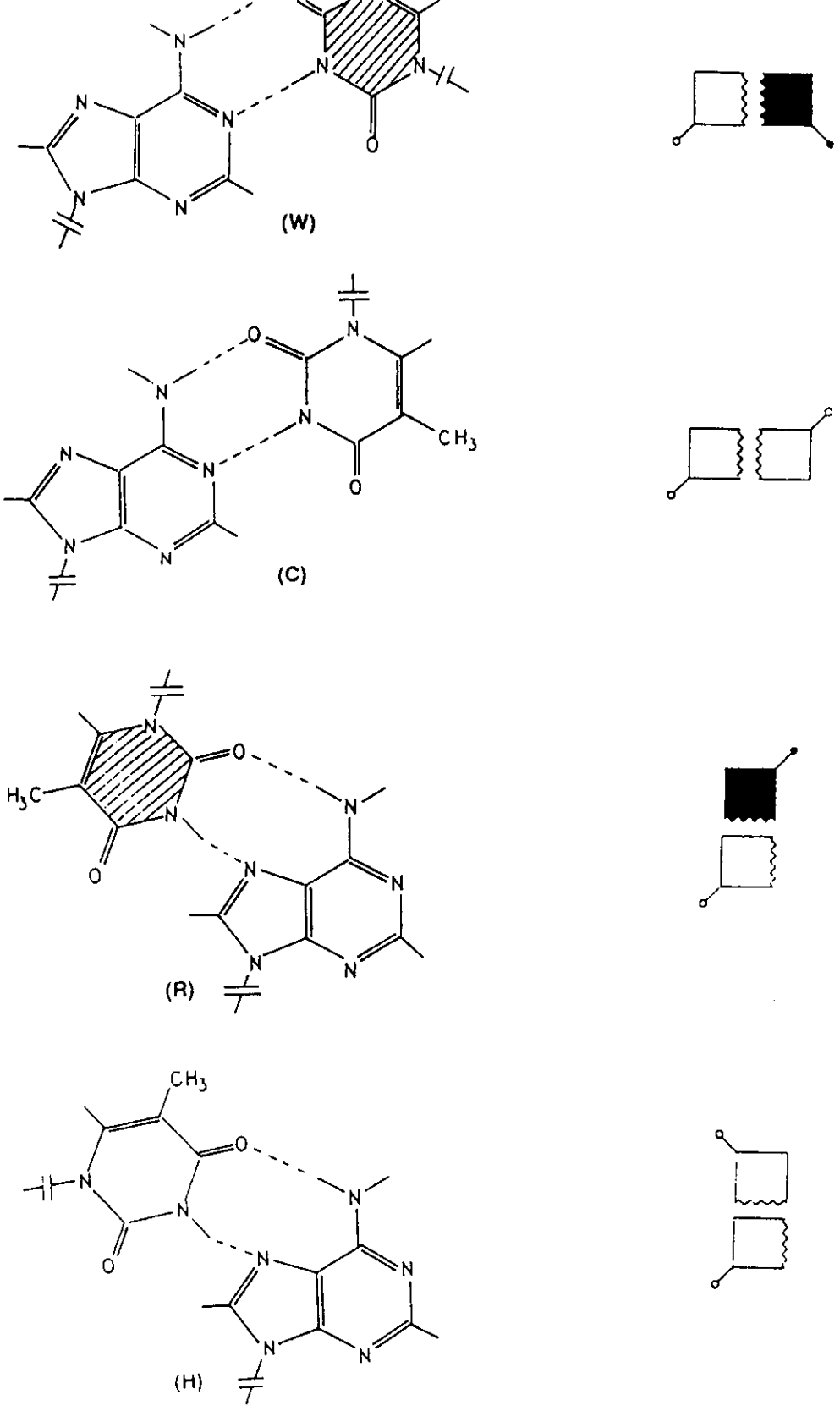
## a) Structural Classes

The second level of description that we will discuss concerns the way nucleotide strands can be put together to form helical structures. Due to the rapid increase in the variety of known nucleic acid architectures, it seems to be useful to make such a classification in a clear and homogeneous way (Lavery et al, 1992). At this level, we will ignore all fine conformational details, making a description based only on how many strands interact, the orientation of these strands with respect to one another (parallel or anti-parallel) and the type of the base-base hydrogen bonding interactions involved.

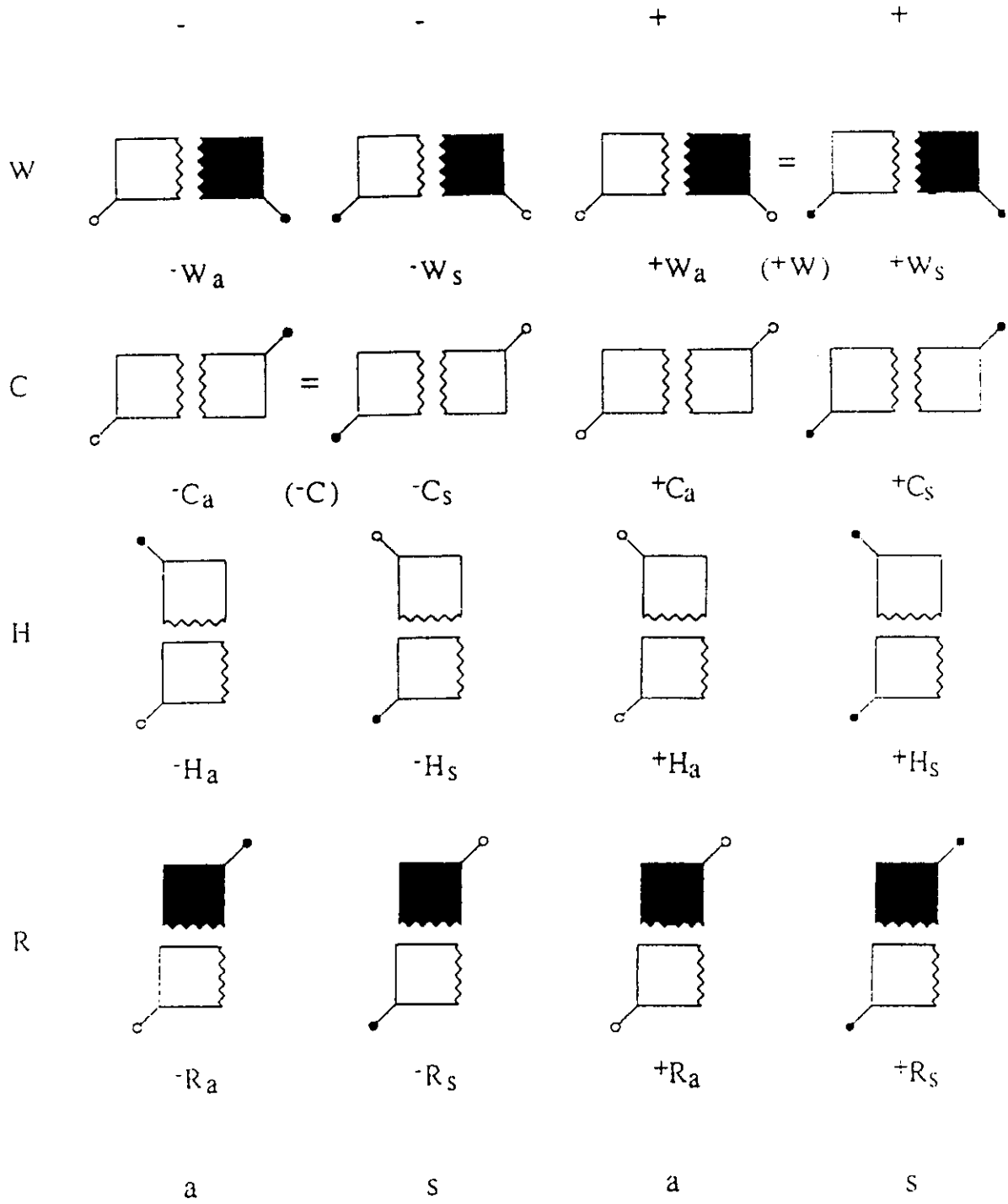
The starting point for our classification is the observation that each nucleic acid base has two unique faces (Rose et al, 1980; Harvey 1983). The definition of one such face can be made using the right-hand rule, the fingers of the right hand pointing around the shortest distance from the glycosidic bond to the Watson-Crick pairing edge of the base. The direction of the thumb then indicates a unique face which will be conventionally coloured white, the opposing face being black. Figure 1 shows standard and stylized views of four types of base pairing: Watson-Crick (W), reversed Watson-Crick (C), Hoogsteen (H) and reversed Hoogsteen (R). Note that the bases in W and R base pairs both show one black and one white face, while in C and H base pairs both bases show the same colour.

In order to represent a nucleotide, we extend our simplified diagrams by adding a circle to the end of the line representing the glycosidic bond: An open circle corresponding to the 5'->3' direction of the phosphodiester strand pointing upwards (type 'a') and a filled circle corresponding to the 5'->3' direction pointing downwards (type 's'). As you may have guessed, the letters 'a' and 's' were chosen to recall the anti and syn conformations around the glycosidic bond, but do not generally imply that such conformations must be associated with the corresponding nucleotides. We will return to this point shortly in connection with Z-DNA.

When we combine two nucleotides, each of which can be in one of two possible states, to make a duplex using each of the 4 pairing schemes shown in figure 1, it is possible to make a total of 2x2x4=16 combinations. In fact, as figure 2 shows, there are only 14 unique classes due to 2 degeneracies created by the pseudo-dyad symmetry of W and R pairing (note each base pair in figure 2 is oriented so that the left hand base or lower base shows its white face). Each structural family can be defined by a concise notation consisting of a letter to specify the base pairing (W, C, H or R), a prefix indicating whether the strand directions are parallel (+) or anti-parallel (-) and a suffix specifying whether the left-hand (or lower) nucleotide is of type 'a' or type 's' (this index can be dropped in the case of the degenerate pairs $^{+}W_a$ / $^{+}W_s$ and $^{-}C_a$ / $^{-}C_s$). It is possible to deduce the type of the second nucleotide using the rule: When




(1) Atomic and diagrammatic representations of four types of base pairing. W: Watson-Crick, C: Reversed Watson-Crick, R: Reversed Hoogsteen, H: Hoogsteen.




(2) Diagrammatic representations of double stranded nucleic acid helices.

the base pair faces are of a single colour, parallel strands imply nucleotides of the same type and anti-parallel strands imply nucleotides of different types. The opposite is true when two colours appear on the base pair face.

The family $^{-}W_a$ in fact corresponds to the most common form of duplex DNA in the B or A conformations with anti-parallel strands, Watson-Crick hydrogen bonding and a-type nucleotides which are indeed in anti conformations. The family $^{-}W_s$ corresponds to Z-DNA. This representation of the Z form makes it clear that base pairs have to be turned over in passing from the B to the Z conformation (Wang et al, 1979; Harvey, 1983), since to align the strand directions between the first two families in figure 2 it is necessary to invert the $^{-}W_s$ diagram around a horizontal axis leading to a base pair with a black face on the left and a white face on the right. If we interpret the nucleotide types 'a' and 's' as leading to anti and syn conformations, our notation would imply that Z-DNA should contain only syn nucleotides. In reality, the difficulty of forcing pyrimidines into the syn form leads Z-compatible purine-pyrimidine base sequences to adopt alternating anti and syn conformations. This however cannot change overall strand directions and thus the rotation of the pyrimidine sugar along with the base pair results in the contorted 'Z' backbone pathway characteristic of this type of duplex DNA. We must thus re-emphasize that the 'a' and 's' notation only relates base orientation to the strand direction and does not refer to the state of the glycosidic angle. Apart from the special case of Z-DNA, assuming a=anti and s=syn is likely to work and indeed this assumption forms the basis of a rule for predicting strand directions recently proposed (Westhof, 1992).

Returning to the structural families, an alternative way to solve the difficulty of forming syn nucleotides is to change the stereochemistry at C1'. This takes us from natural β-nucleotides to α-nucleotides and results in diminishing steric hindrance between C5' and the base in the case of a syn conformation. It is thus not surprising that an all α-nucleotide duplex belonging to the family $^{-}W_s$ has indeed been observed (Morvan et al, 1987). The final family which can be made with Watson-Crick base pairs, $^{+}W$, has also been observed recently in parallel stranded duplexes where the second strand is again composed of α-nucleotides which easily accept the syn conformation (Sun et al, 1988). Of the reversed Watson-Crick duplexes only the $^{+}C_a$ family is currently known for poly(dC).poly($dC^{+}$) duplexes under acidic conditions (Arnott et al, 1976). It is also found in the parallel stranded structure proposed by Pattabiraman on the basis of molecular modeling (Pattabiraman, 1986) and subsequently observed in parallel stranded AT sequences (Ramsing & Jovin, 1988).

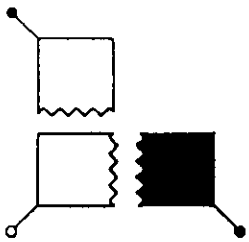
Since normal duplex DNA generally prefers W or C type base pairing, no Hoogsteen or reversed Hoogsteen pairings are seen unless the former possibilities are excluded. This is the case in triple helices where the Watson-Crick hydrogen bonding possibilities are absorbed in the underlying duplex or in the case of certain chemical modification (Rich et al, 1961; Fukui & Ikehara,

1979). In order to form triple helices (or 'triplexes') it is necessary to combine two double helix families from figure 2. Due to base pairing possibilities, this requires the presence of a purine which can bind to two other bases via Watson-Crick (or reversed Watson-Crick) pairing at the same time as Hoogsteen (or reversed Hoogsteen) pairing. If we assume that the left hand base of each Watson-Crick family is a purine we can then build a total of 16 possible triple helices (figure 3) using in turn each of the Hoogsteen and reversed Hoogsteen families compatible with a given Watson-Crick family. It should be noted that there are no other possibilities with Watson-Crick base pairs, since if we had assumed the right hand base to be a purine we would have simply obtained the same families in a different orientation. It should also be noted that, once a triple helix has been formed, the degenerate ($^{+}W_a$, $^{+}W_S$) and ($^{-}C_a$, $^{-}C_S$) families become distinct since only one of their constituent bases carries the third strand.

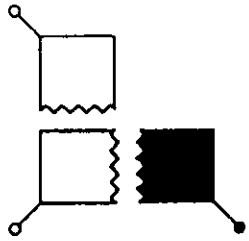
As shown in figure 3, it is possible to give each of these triple helices a unique notation corresponding to the two constituent double helices. The first family, built from a $^{-}W_a$ Watson-Crick duplex and a $^{-}H_a$ Hoogsteen duplex, thus becomes $^{-}W_a{}^{-}H_a$. In fact, the nucleotide type indicated for the second base pair can be dropped since it must be identical to that of the first pair - remember that the nucleotide type refers to the left-hand or lower nucleotides for the duplexes in figure 2 and is thus necessarily the same for any pair of duplexes used to form a triplex. The first triplex family can thus be uniquely defined as $^{-}W_a{}^{-}H$. For completeness, we also note that it is possible to construct triple helices starting from reversed Watson-Crick base pairs, leading again to 16 unique possibilities. Table 3 lists the type of each nucleotide in all the double stranded and triple stranded structures we have defined above.

The best known triple helix made by adding a Hoogsteen bonded thymidine strand to an poly(dA).poly(dT) double helix (Arnott & Selsing, 1974) corresponds to the family $^{-}W_a{}^{+}H$ since all nucleotides are of a-type and the Hoogsteen bound poly(dT) strand is parallel to the adenosine strand of the duplex. An identical triple helix family $CGC^{+}$ can also be formed under acidic conditions by adding a protonated cytosine strand to a poly(dG).poly(dC) duplex, again using Hoogsteen hydrogen bonding. The only other way to form an all a-type triple helix starting from a Watson-Crick duplex is the family $^{-}W_a{}^{-}R$ (see table 3) which has indeed been experimentally observed for AAT and CGG triple helices (Broitman et al, 1987; Pilch et al 1991) where the two purine strands form an anti-parallel reversed Hoogsteen duplex. A related family containing s-type nucleotides in the third strand, $^{-}W_a{}^{+}R$, is also known to exist when α-thymidine nucleotides, which can easily adopt the syn conformation, are built into the third strand of an TAT triplex (Sun et al, 1991).

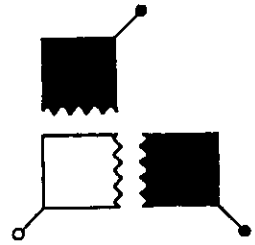
Forming the triple helices shown in the second row of figure 3 seems unlikely since the only known form of the $^{-}W_S$ duplex is Z-DNA in which the major groove face of the base pairs is sterically hindered. In contrast, starting from a parallel strand Watson-Crick duplex $^{+}W$ (formed using an α-nucleotide



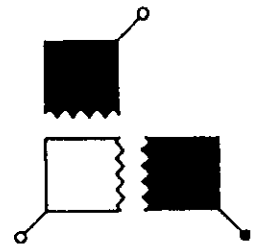
$^{-}W_a{}^{-}H$



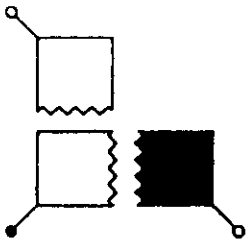
$^{-}W_a{}^{+}H$



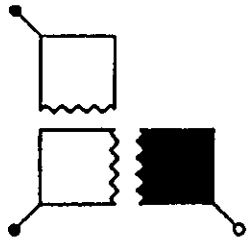
$^{-}W_a{}^{-}R$



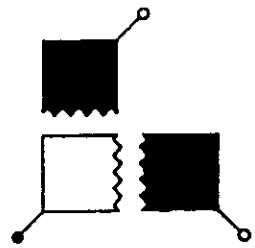
$^{-}W_a{}^{+}R$



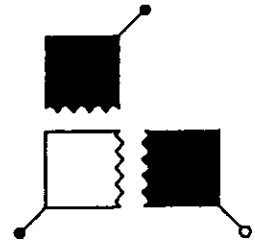
$^{-}W_s{}^{-}H$



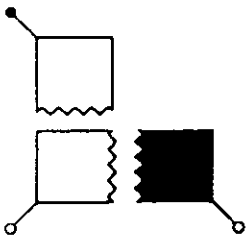
$^{-}W_s{}^{+}H$



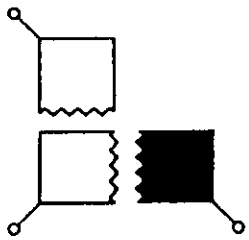
$^{-}W_s{}^{-}R$



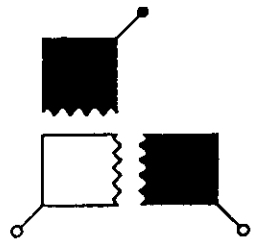
$^{-}W_s{}^{+}R$



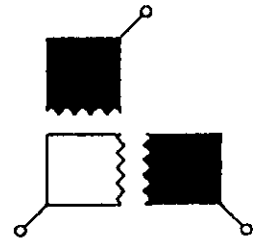
$^{+}W_a{}^{-}H$



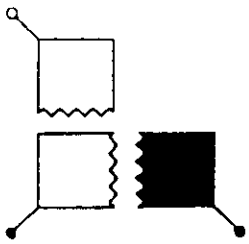
$^{+}W_a{}^{+}H$



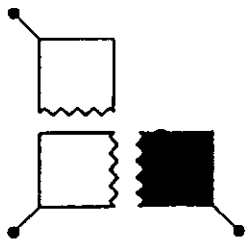
$^{+}W_a{}^{-}R$



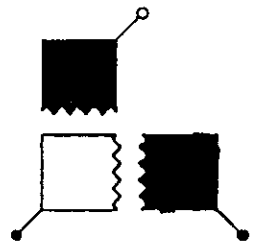
$^{+}W_a{}^{+}R$



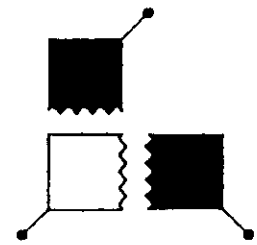
$^{+}W_s{}^{-}H$



$^{+}W_s{}^{+}H$



$^{+}W_s{}^{-}R$



$^{+}W_s{}^{+}R$

(3) Diagrammatic representations of triple stranded nucleic acid helices.

pyrimidine strand) it may be possible to form triple helices belonging to the families $^{+}W_a{}^{+}H$ or $^{+}W_a{}^{-}R$ which only require syn conformations in the Watson-Crick bound pyrimidine strand (table 3). Without using α-nucleotides, but starting from a parallel stranded reversed Watson-Crick duplex $^{+}C_a$ (Ramsing & Jovin, 1988), it may also be possible to build triple helices of the families $^{+}C_a{}^{+}H$ or $^{+}C_a{}^{-}R$ where all strands contain only a-type nucleotides. Forming any of these 4 latter triplexes however depends on the ability of the underlying duplex to adopt a conformation which sufficiently exposes the major groove face to accept a third strand. Results from the Jovin's group suggest that an internal H-DNA triple helix formed under negative supercoiling stress by a plasmid containing a parallel-stranded $(dA)_{15}.(dT)_{15}$ insert in fact undergoes rearrangement to form an AAT triplex belonging to the family $^{-}W_a{}^{-}R$ (Klysik et al, 1991).

Table 3. Classification of double and triple stranded nucleic acid families. Nucelotide types (a/s) are shown and known structural families are underlined

| Double strand | | Triple strand (W) | | Triple strand (C) | |
|---|---|---|---|---|---|
| $^{-}W_a$ | <u>aa</u> | $^{-}W_a{}^{-}H$ | aas | $^{-}C_a{}^{-}H$ | ass |
| $^{-}W_s$ | <u>ss</u> | $^{-}W_a{}^{+}H$ | <u>aaa</u> | $^{-}C_a{}^{+}H$ | asa |
| $^{+}W$ | <u>as</u> | $^{-}W_a{}^{-}R$ | <u>aaa</u> | $^{-}C_a{}^{-}R$ | asa |
| | | $^{-}W_a{}^{+}R$ | <u>aas</u> | $^{-}C_a{}^{+}R$ | ass |
| $^{-}C$ | as | $^{-}W_s{}^{-}H$ | ssa | $^{-}C_s{}^{-}H$ | saa |
| | | $^{-}W_s{}^{+}H$ | sss | $^{-}C_s{}^{+}H$ | sas |
| $^{+}Ca$ | <u>aa</u> | $^{-}W_s{}^{-}R$ | sss | $^{-}Cs{}^{-}R$ | sas |
| $^{+}C_s$ | ss | $^{-}W_s{}^{+}R$ | ssa | $^{-}Cs{}^{+}R$ | saa |
| $^{-}H_a$ | as | $^{+}W_a{}^{-}H$ | ass | $^{+}C_a{}^{-}H$ | aas |
| $^{-}H_s$ | sa | $^{+}W_a{}^{+}H$ | asa | $^{+}C_a{}^{+}H$ | aaa |
| $^{+}H_a$ | <u>aa</u> | $^{+}W_a{}^{-}R$ | asa | $^{+}C_a{}^{-}R$ | aaa |
| $^{+}H_s$ | ss | $^{+}W_a{}^{+}R$ | ass | $^{+}C_a{}^{+}R$ | aas |
| $^{-}R_a$ | <u>aa</u> | $^{+}W_s{}^{-}H$ | saa | $^{+}C_s{}^{-}H$ | ssa |
| $^{-}R_s$ | ss | $^{+}W_s{}^{+}H$ | sas | $^{+}C_s{}^{+}H$ | sss |
| $^{+}R_a$ | <u>as</u> | $^{+}W_s{}^{-}R$ | sas | $^{+}C_s{}^{-}R$ | sss |
| $^{+}R_s$ | sa | $^{+}W_s{}^{+}R$ | saa | $^{+}C_s{}^{+}R$ | ssa |

It is possible to continue this classification to deal with quadruple stranded structures (Guschlbauer et al, 1990) such as those formed by poly(dG) or poly(dI) gels (denoted $(^{+}H_a)^2$). It should also be remarked that the classification described above applies equally well to deoxyribo- and to

ribonucleic acids. Standard duplex RNA, like A-DNA, falls in the $^{-}W_a$ family, however more complex forms of RNA exhibit, at least locally, other structural families. This is notably the case for transfer RNA's.

In summary, this characterization describes nucleic acid structure at a level roughly equivalent to the secondary structures of proteins. Nucleic acids are however seen to have a much richer variety of forms at this level.

**b) Helicoidal Parameters**

Increasing interest in the fine details of macromolecular conformation has led to a search for quantitative methods for describing such structures. This need has been very clearly felt in the case of nucleic acids, notably since sequence dependent curvature has been shown to play an important biological role. However the problem of correctly describing irregular conformations and, in particular, of deriving an appropriate global axis is much more complicated than it may seem at first sight. Interest in this problem has been stimulated by an increasing number of numerical simulations carried out on biological macromolecules. Such studies also require appropriate techniques for quantitatively analyzing the molecular conformations which result. Indeed, without such data, it becomes very difficult to progress in analyzing the energetic factors which give rise to the conformations observed.

We have thus tried to develop a rigourous algorithm for dealing with such problems (Lavery & Sklenar, 1988, 1989, 1990). Our starting point was to look for a way to extend regular helical geometry to the description of irregular systems so that the notion of a helical axis could be conserved, although this axis would in general be a space curve rather than a straight line. This approach distinguishes our algorithm from a so-called local view of irregular polymers, where the concept of an axis is lost altogether. In the case of a perfect helix, such as that described by the fibre coordinates of double stranded DNA (Arnott, 1970), the axis is a rigourously defined straight line. Every monomer has the same relative position and orientation with respect to its local axis segment and each monomer in the structure can then be reached from the preceding monomer by a fixed rotation around the axis coupled with a fixed translation along the axis. These properties mean that, for a regular system, the helical axis can always be located by simple vectorial algebra (Rosenberg et al, 1976).

Once we move away from exact helical symmetry we are confronted with the question whether an appropriate "axis" can still be found. If the deviations from perfect symmetry are small we can still try to locate a straight line axis which best fits the structure studied. However, as the deformations become more important, this solution is clearly not tenable and if we still want an "axis" then we must accept that it must be curved. The difficulty of defining such an object has led many people to abandon the notion of an axis altogether and rather to limit themselves to defining the position of each monomer with

respect to the preceding monomer using an appropriate set of 3 translations and 3 rotations (Fratini et al, 1982; von Kitzing & Diekmann, 1987; Battacharya & Bansal, 1988; Soumpasis & Tung, 1988). This choice, if correctly formulated, leads to an independent set of local parameters which fully define the molecular conformation. However, local parameters (which are not in fact helicoidal parameters in any real sense) are of limited use if we want to define the overall shape of the molecule and, in particular, its curvature.

Our approach involves a least squares optimisation procedure based on a function which mathematically describes departures from ideal helical symmetry. How do we define such departures from helical symmetry? Consider, firstly, the requirement that the axis should be as straight as possible. If we consider the overall axis to be made up of segments, one segment for each monomer within the structure, then we should ensure that these segments are, as far as possible, aligned and that the lateral displacement between successive segments is as small as possible. If we define successive segments i-1 and i by the vectors $U_{i-1}$ and $U_i$, passing through the points $P_{i-1}$ and $P_i$, these criteria can be formulated as two simple positive terms,

$$a_i = (U_i - U_{i-1})^2$$

which will become zero if the vectors are aligned and,

$$b_i = [S_i - \langle U_i \rangle (\langle U_i \rangle^T S_i)]^2$$

(where $\langle U_i \rangle = (U_i + U_{i-1}) / \_U_i + U_{i-1}\_$, $S_i = P_i - P_{i-1}$ and $\langle U_i \rangle^T S_i$ indicates the scalar product between these two vectors)

which is the square of the displacement between the P points measured perpendicularly to the mean vector $\langle U_i \rangle$. In order to treat a polymer consisting of N monomers we need only sum these terms over all the junctions between successive axis segments. This leads to the terms,

$$A = \Sigma_{i=2,N}\, a_i \qquad \text{and} \qquad B = \Sigma_{i=2,N}\, b_i$$

In order to deal with the orientation of successive monomers with respect to their individual axis segments, we firstly define the position of each monomer i by a point $E_i$ and an axis system $J_i,K_i,L_i$ fixed to the atoms of the monomer. For nucleic acids, the reference point E is placed at the position where the helical axis cuts the plane of the base in the B conformation deduced from fibre diffraction (Arnott et al, 1980). The L vector is chosen perpendicular to the base plane (coming upwards from the white face - see preceding section), the J vector is the pseudo-dyad vector pointing into the major groove and the K vector points roughly along the long axis of the base pair. With this system, the helicoidal parameters describing the base position with respect to

the helical axis are all zero for the canonical B conformation. Note that choices based on specific base atoms, base pair axes or on inertial axes will lead to apparent irregularity even in perfectly symmetrical helices. It is also remarked however that it may be necessary to least squares fit (McLachlan, 1979) standard bases to the actual base coordinates in the case of the distorted structures which can result from molecular dynamics calculations.

Point $P_i$ is now fixed on the axis vector $U_i$, without loss of generality, by assuming that $P_i$-$E_i$ is perpendicular to $U_i$. The differences in the rotational position of the axis vectors U with respect to successive monomer reference systems is then given by the term,

$$c_i = \Sigma_{Xi} (U_i^T X_i - U_{i-1}^T X_{i-1})^2$$

where $X_i$ is respectively, $J_i$, $K_i$ and $L_i$. Similarly, the differences in the translational position of the P points is given by the term,

$$d_i = \Sigma_{Xi} [(P_i-E_i)^T X_i - (P_{i-1}-E_{i-1})^T X_{i-1}]^2$$

As before we can now treat a polymer made up from N monomers by summing these terms over all monomer pairs to yield,

$$C = \Sigma_{i=2,N} c_i$$
$$D = \Sigma_{i=2,N} d_i$$

We can now form our function F which will measure the irregularity of any chosen polymer conformation:

$$F(h) = W (A + C) + B + D$$

Each of the four terms we have defined will be zero for a perfectly regular polymer and will be positive for irregular conformations. The variables of the function, denoted by the letter h, are simply two translations and two rotations for each monomer i which place the axis system $U_i$,$P_i$ in space with respect to the fixed monomer reference system. In order to obtain a balanced weighting between the rotational terms of the function (A,C) and the translational terms (B,D) it is necessary to multiply the rotation angles contained within by the former terms by the average distance separating successive monomers. This implies that A and C should be multiplied by the square of this distance (denoted as W above). The appropriate value of W will thus depend on the type of polymer being treated. For nucleic acids we have taken W=10.

Having defined this function we can now use a minimiser to find the optimal values of the variables h which lead to a minimal value of F(h). For this purpose we employ a conjugate gradient algorithm since the analytic first derivatives of F(h) can easily be formulated (Lavery & Sklenar, 1988). The

minimal value of F(h) found for any given polymer is already a guide to its degree of structural irregularity. It should also be noted that the irregularity of each pair of adjacent monomers within the structure can also be measured since, after optimisation, we can easily calculate a value $f_i$ for each such pair where,

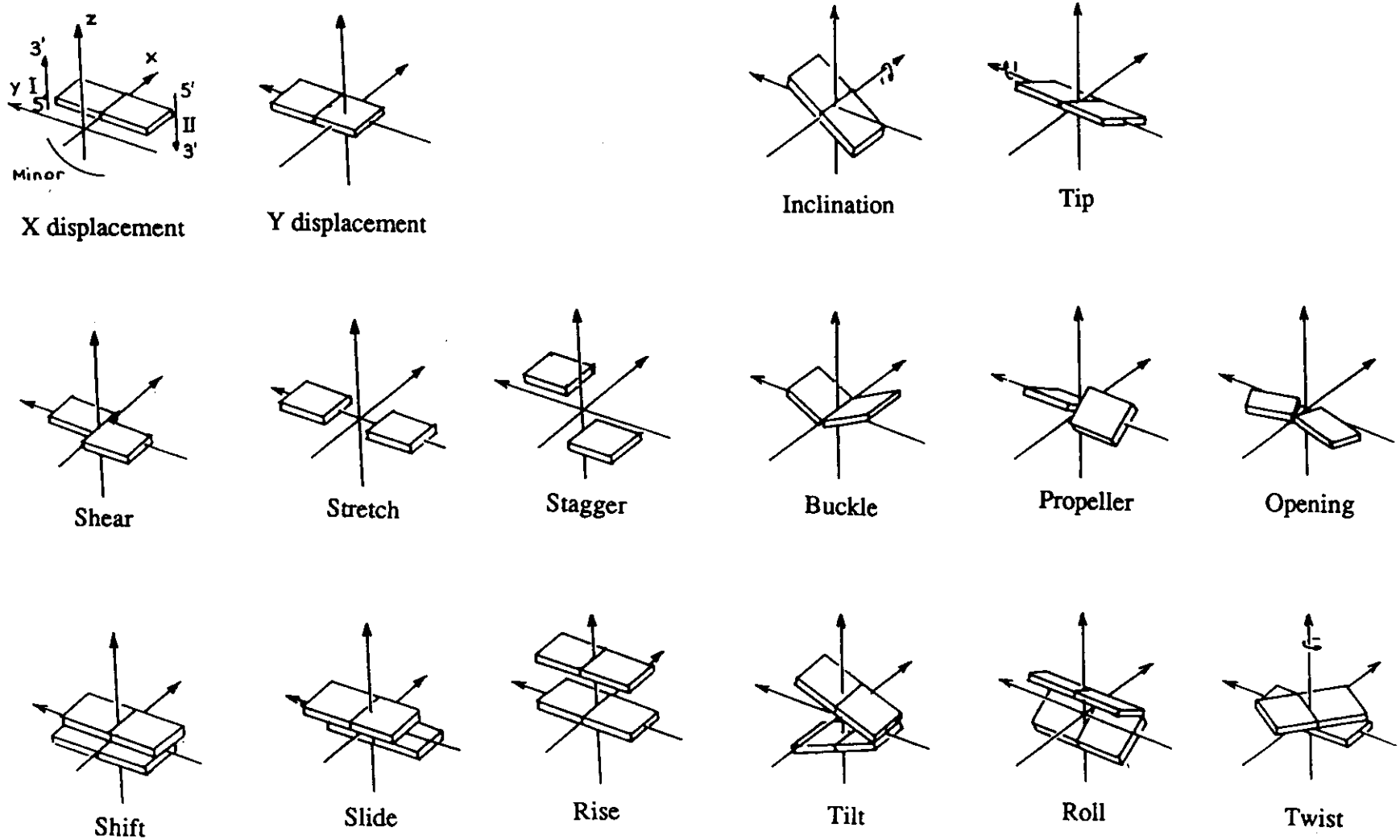
$$f_i = W\,(a_i + c_i) + b_i + d_i$$

This value is termed the dimeric irregularity function, or DIF, and can be used to locate the zones where helical symmetry is most disturbed.

It is finally important to remark that the axis definition we obtain by this method, consisting of the ensemble of U vectors and P points, is a truly global description of the polymer conformation since the optimisation is performed for a function which not only considers each type of departure from helical symmetry, but also simultaneously considers the position of all monomers making up the polymer. The overall space curve which can be derived from this data (by cubic spline fitting) is thus in a very real sense an optimal and global guide to the conformation of the polymer concerned.

Having minimized the Curves function we are in a position to calculate the helicoidal parameters which define the macromolecular conformation in question. In general six parameters (3 translations and 3 rotations) are always sufficient to define the position of any object in space with respect to a chosen reference. Thus, in the local parameter approach to polymers, one calculates six values which position each monomer with respect to the preceding monomer. For our global analysis, we must take into account the definition of the overall axis. This can be done with 4 parameters at each axis junction (axis X and Y displacement, axis inclination and axis tip). The monomers can then be positioned with respect to their local axis segment with 4 parameters (X displacement, Y displacement, inclination and tip) and with respect to the preceding monomer by rise and twist.

Although no other parameters are necessary to completely define the Curves solution it is often useful to deduce other dependent parameters which can facilitate reading overall macromolecular conformation. For single stranded polymers it is interesting to calculate 2 overall translations and 2 overall rotations between successive monomers. These parameters are termed shift, slide, tilt and roll respectively. Along with rise and twist they lead to a full set of 6 inter-monomer parameters. For double stranded systems such as duplex DNA, the single strand values described can be used to generate a set of 4 mean base pair-axis parameters and a set of 6 mean inter-base pair parameters (which all carry the same names as for single stranded systems). However, it is also interesting to define a set of 6 inter-strand parameters (which, in the case of DNA, describe the internal base pair geometry). The latter set again contains 3 translations (shear, stretch and stagger) and 3 rotations (buckle, propeller twist and opening). All these parameters, shown in




(4) Helicoidal parameters defined in the Cambridge convention.

figure 4, obey the sign and name conventions set out during the 1988 EMBO workshop in Cambridge (Dickerson et al, 1989).

It is finally remarked that the computer program "Curves" we have developed specifically for the nucleic acids also evaluates the local parameters between successive bases and base pairs so that comparison with other algorithms may be made easily. However, as we have discussed in length in a previous publication (Lavery & Sklenar, 1989), local parameters can be very misleading for interpreting macromolecular conformation, particularly curvature, and, in our opinion, cannot replace the evaluation of an overall axis and true global parameters. These arguments will not be repeated here. Recent versions of the Curves program have been extended to treat 3- and 4-stranded complexes, abasic sites, dangling ends and structures showing strong dinucleotide effects such as Z-DNA. The most recent development includes an option to measure the width and depth of helical grooves (Boutonnet et al, 1992; Stofer & Lavery, 1992).

One particularly interesting application of the Curves algorithm is the analysis of molecular dynamics simulations for biological macromolecules. In such cases our algorithm can be used very effectively to interpret the mass of conformation data which result from such studies. This analysis firstly allows the evolution of macromolecular conformation to be described by rigourously defined independent parameters. The variation of these parameters in time has been made easy to visualize with the help of the "Dials and Windows" technique developed by the group of David Beveridge (Ravishankar et al, 1989; Swaminathan et al, 1990). This approach allows both helicoidal and dihedral parameters to be represented graphically in a very compact way and also offers interesting possibilities for studying differential parameter changes which would greatly help in locating and characterizing conformational transitions. In addition, animations of the axis and ribbon diagrams resulting from a Curves analysis allow both local and global conformational changes to be seen much more easily. In particular, for nucleic acids, base pair deformations, axis bending and groove width can easily be monitored, while for proteins secondary structure deformations and domain motions are made very clear. Further developments, still underway, allow Curves parameters to be used as the basis for measuring the persistence length of DNA. This work has led to interesting refinements of the notions of static and dynamic bending within DNA oligomers (Prevost et al, 1992).

## TECHNIQUES FOR MODELING NUCLEIC ACIDS

### a) Force fields

The basic empirical force fields used to model nucleic acids are generally not very different from those employed for other types of biological macromolecule, consisting of pairwise additive non-bonded terms

(electrostatic interactions, repulsion and dispersion) complemented by harmonic bond length and angle penalties and sinusoidal dihedral barriers. The sugar-phosphate backbone requires some special consideration especially through the use of anomeric (or 'gauche') dihedral terms which appear to be necessary for correctly reproducing both phosphodiester and sugar puckering conformations (see, for example, Schlick et al, 1987; Olson & Sussman, 1982; Hayes et al, 1977). Sets of parameters have been developed by several groups. In our studies, we use the rather conventional formulation (Lavery et al, 1986a, 1986c) shown below, which excludes bond length stretching terms, bond lengths being frozen in our modeling approach.

$$E= \Sigma\, QiQj/\varepsilon(R)Rij + \Sigma\,( -Aij/Rij^6 + Bij/Rij^{12} )$$

$$+ \Sigma\,[\, Cos\tau\,( -Aij^{HB}/Rij^6 + Bij^{HB}/Rij^{12} )$$

$$+ (1-Cos\tau)\,( -Aij/Rij^6 + Bij/Rij^{12} )\,]$$

$$+ \Sigma\, Vs/2\,( 1 \pm Cos\, Ns\tau s ) + \Sigma\, Va\,( \sigma a-\sigma a^{\circ})^2$$

The first term of this formula is the electrostatic energy, calculated as a sum of interactions between atomic monopoles Qi damped by a dielectric function $\varepsilon(R)$ which is described below. The monopoles we employ are calculated by the Hückel-Del Re method, specially reparameterised to obtain the best possible fit with SCF *ab initio* electrostatic potential and field distributions around the nucleic acid subunits (Lavery et al, 1984). The next three terms represent the Lennard-Jones or dispersion-repulsion energy calculated with a 6-12 dependence and using, in part, the parameter set developed by (Zhurkin et al, 1980). Hydrogen bonds are dealt with by the latter two of these terms, which take into account angular dependence, fitted to quantum chemical calculations, by mixing together two sets of A,B parameters using a cosine function of the angle $\tau$ formed by the vectors X-H and H-Y for a bond X-H...Y. All of these terms are summed over pairs of atoms separated by at least three chemical bonds. The last two terms represent the distortion energy associated with torsion angles $\tau s$ (including anomeric effects) and valence angles $\sigma a$. Polarization energies are excluded here. No energy cut-offs are used and the internal energies of the bases are not calculated.

The dielectric function $\varepsilon(R)$ we use is based on a model of the dielectric damping of the electrostatic interaction between two charges in a polar solvent (Hingerty et al, 1985). We have reformulated this function as shown below so that it is possible to vary both the plateau value of the dielectric reached at long distance (D) and the slope of the sigmoidal segment of the function (S).

$$\varepsilon(R)= D - (D-1)/2\,[\, (RS)^2 + 2RS + 2\,]\, exp\,(-RS)$$

The values used presently are D=78 and S=0.16, this slope having been found most appropriate for modeling B-DNA in aqueous solution (Lavery, 1988a, 1988b). This damping is combined with a reduction of all phosphate net charges to -0.5e in order to mimic the effect of counter-ions. However, it should be noted that when using $\varepsilon(R)$, the DNA conformation is only weakly dependent on these charges. Our own work and that carried out in other laboratories (Fritsch & Westhof, 1990; Jayaram et al, 1990) suggest that this simple electrostatic model is quite successful in reproducing DNA conformation in solution and in describing DNA-ion interactions, without being computationally time consuming. Tests against Poisson-Boltzmann calculations (Friedman & Honig, 1992) however suggest that a slope of S=0.356, which reproduces Hingerty's original distance dependence, is more appropriate (see also the discussion of Mazur & Jernigan, 1991).

It must be stressed that using a distance dependent dielectric is a very poor substitute for true Poisson-Boltzmann calculations on a well defined dielectric model (Gilson & Honig, 1988; Harvey 1989; Davis & McCammon, 1990; Bashford, 1991; Moult, 1992). Treating the electrostatic term of the force field for nucleic acids is a particularly difficult problem. Nucleic acids are highly charged molecules are stabilized by the presence of both water and of counterions. It is possible that a complete Poisson-Boltzmann treatment may be sufficient to deal which such environmental effects and several groups are trying to include this approach within molecular mechanics calculations in a numerically efficient way (Gilson & Honig, 1991; Juffer et al, 1991; Sharp, 1991; Zauhar, 1991; Sklenar et al, 1990). It is also possible however that interactions between the nucleic acids and, at least, first shell hydration waters may still have to be treated explicitly (Westhof & Beveridge, 1990).

Molecular dynamics calculations in water are another way to approach the problem of modeling an aqueous environment. Experience with such simulations is progressing, but the calculations remain very expensive and, although the time period simulated rarely exceeds a fraction of a nanosecond, this period is sufficient for oligomeric helices to be destroyed unless constraints on hydrogen bonding or the terminal bases are imposed (Beveridge et al, 1992; Brahms et al, 1992). Such effects may be due to the force field employed, but may also be influenced by problems of equilibration. Even in properly stabilized systems, other problems exist, such as the very small number of counterions employed, generally limited to one cation per phosphate group to ensure electroneutrality, and the low mobility of these ions during the simulation. Despite increasing computer speed, it thus seems unlikely that dynamic simulations in water will be able to be carried out in sufficient numbers to really help in unravelling the mechanism of base sequence effects.

## Internal and Helicoidal Coordinates

The limits imposed by molecular dynamics simulations led us to look at other ways to model nucleic acids. We were, and are, principally interested in three features of nucleic acid behaviour: firstly, base sequence effects, which, as already discussed, require studying 10's, if not 100's, of different oligomeric sequences, secondly, conformational transitions, which often involve very large movements and typically occur on long time scales and, thirdly, the mechanisms underlying the recognition of target sites on nucleic acids by drugs or proteins. All these problems required the consideration of very many large molecular systems and consequently calculations have to be made very quickly. Energy minimisation seemed a possible route, since when extended to defining energy hypersurfaces, it can be a very powerful tool for looking at structural deformations. However, the usefulness of minimisation in macromolecular systems is generally hindered by severe problems connected with local minima. The complexity of the problem can easily be appreciated when we note that a single turn of double helical DNA represents roughly 2000 cartesian variables. We thus set out to adapt our representation of DNA to reduce these problems, firstly, by reducing the number of independent variables and, secondly, by choosing variables adapted to representing the movements occurring within nucleic acids.

The choice of variables used to model macromolecular structures has considerable consequences for the usefulness of the resulting algorithm. In the case of the nucleic acids, helicoidal variables are obviously of great interest. Over several years we have thus developed and refined a method which enables such variables to be exploited in conjunction with an internal variable model of nucleotide flexibility. This has several advantages for the types of modeling we were interested in undertaking:

- It allows both regular and irregular nucleic acid segments to be constructed easily, including unusual structures such as bends, junctions, loops, mis-matched pairs and triple helices.

- Chosen conformational features may be "frozen" while all other variables are energy relaxed. This, in turn, opens the way for calculating optimal energy pathways involved in a whole range of structural deformations including, for example, transitions between the different allomorphic forms of DNA.

- A very considerable reduction can be achieved in the number of variables necessary to model a given segment of nucleic acid compared to the classical cartesian coordinate molecular mechanics (roughly 10 times less). In addition, since these variables are much better adapted to describing the

nature of the flexibility of the molecular system studied, energy optimisation can be achieved faster and larger changes in conformation can be made.

- Helical symmetry can easily be imposed by grouping together symmetry equivalent sets of variables, leading to further reductions in the number of independent variables representing the system under study. As an example, a turn of DNA constrained to mononucleotide symmetry can be represented by roughly 20 variables, a reduction of 100 times compared to molecular mechanics.

The modeling algorithm we have developed on this theoretical basis is termed "Jumna" (Junction Minimisation of Nucleic Acids). Many features have been added to this algorithm since its conception (Lavery, 1988a, 1988b) and it has also drawn on experience gained with the earlier approaches SIR (Sklenar et al, 1986; Lavery et al, 1986a, 1986b) and Cinflex (Lavery et al, 1986c). The present version of the Jumna code (Version 7.0) contains roughly 12,000 lines of Fortran and provides an extensive set of tools for studying nucleic acids and their complexes.

Jumna begins by breaking a DNA oligomer down into 3'-monophosphate nucleotide units (with the exception of the 3'-terminii which are nucleosides). These units are then positioned in space with respect to a common helical axis system using a set of 6 helicoidal variables. The internal flexibility of each nucleotide is represented by single bond rotations at the glycosidic link and within the phosphodiester backbone and valence angles within the sugar ring. All other valence angles and all bond lengths are taken to be fixed. (The exo-valence angles of the sugar rings do in fact vary in a coupled way with the endo-valence angles to maintain a symmetric position of the sugar ring substituents, but they are not independent variables).

The independent variables associated with each nucleotide are consequently 3 translations and 3 rotations, which position the nucleotide with respect to the helical axis system, the glycosidic dihedral, 3 valence angles and 2 dihedrals within the sugar moiety and 2 backbone dihedrals $\varepsilon$ (C4'-C3'-O3'-P) and $\zeta$ (C3'-O3'-P-O5'). Other sugar and backbone variables are dependent and are determined by the closure conditions which involve the C4'-O1' bond length within each sugar ring, the inter-nucleotide O5'-C5' bond and the valence angles P-O5'-C5' and O5'-C5'-C4'. These constraints are imposed via harmonic energy penalty terms. The corresponding force constants are adjusted to satisfy closure distances to within 0.02Å and closure angles to within 1°.

Breaking down the nucleic acid into nucleotide units has several notable advantages. Firstly, the "junctions" between consecutive nucleotides do not have to be closed for the starting structure and, consequently, it is not necessary to find the right nucleotide internal conformations corresponding to a chosen set of helical parameters before beginning energy optimisation. This

greatly simplifies the investigation of unusual and irregular nucleic acid structures for which no prior conformational information exists. Secondly, "junctions" between nucleotides can open during optimisation allowing for passage between conformational states which would otherwise be separated by important energy barriers. Thirdly, it is not necessary to develop and solve the complicated equations describing the "closure" of the constrained system as has been attempted in a limited way by other authors (see e.g. Tumanyan & Esipova, 1975; Zhurkin et al, 1978).

The conjugate gradient minimisation algorithm that we use requires analytically calculated derivatives of the conformational energy with respect to all independent variables. These energy derivatives are obtained by first calculating the force on each atom of the DNA fragment generated by the energy functional already discussed. The total derivative with respect to any given rotational variable can then be obtained by calculating the vector products of the atomic forces on the moving atoms with vectors linking these atoms to the appropriate centre of rotation and summing the components of these products in the direction of the rotation axis. In the case of helical translation variables the derivatives are simpler, being just the sum of the components of the atomic forces on the moving atoms in the direction of the translation. One should note however that this procedure is made somewhat more difficult by the fact that changing any independent variable results in changing the dependent backbone torsions (C4'-C5', C5'-O5' and O5'-P) for at least one junction and can also result in changes in the dependent sugar variables. Since these torsions also contribute to the conformational energy of the DNA fragment, their derivatives must also be calculated and (after multiplication by the derivative relating the dependent to the independent variable change) added to those generated by the independent variable itself.

It is remarked that the helicoidal parameters used to position each nucleotide and independent commuting variables which obey the Cambridge convention. However, they are only the true helicoidal parameters describing the oligomer in question when symmetry constraints are in action. In the case of irregular oligomers, the optimal axis describing the nucleic acid fragment will not be the Z-cartesian axis as in Jumna and consequently the optimal helicoidal parameter description will also be different. This description can be obtained after energy optimisation using the Curves algorithm discussed previously.

## Symmetry and Other Constraints

Jumna now contains a wide variety of options for modeling special situations. One of the most extensive choices that can be made involves helical symmetry. By regrouping sets of symmetry equivalent variables it is possible to impose symmetries with unit cells containing any number of nucleotides, with or without symmetry relationships between different strands in the

complex. It is also possible to split an oligomer into different zones where different symmetries are imposed. This allows helically symmetric ends to be added to an otherwise free oligomeric segment and also enables junctions between sequences corresponding to different symmetries to be generated. These options apply to the full range of structures treated by Jumna, including 1- to 4- stranded helices including both parallel and antiparallel strand polarities and strands of different lengths.

It is also remarked that when symmetry is imposed on a DNA oligomer, it is possible to avoid end-effects and to gain time by calculating only the energy of the central repeating unit ($E_{MM}$, where M=(N+1)/2 for an oligomer formed from N repeating units) plus the interaction energy of this unit with all the other units within the oligomer ($E_{MI}$). This enables us to calculate the quantity $E_M$, as defined below, which corresponds to the energy of one repeating unit within the environment of an effectively infinite DNA.

$$E_M = 1/2\, E_{MM} + 1/2\, \Sigma_{i=1,N}\, E_{MI}$$

or, alternatively, by noting that $E_{M-I,M} = E_{M,M+I}$,

$$E_M = \Sigma_{i=M,N}\, E_{MI}$$

The total energy of the oligomer can be obtained by simply summing the $E_M$ values (M=1,N) for all the repeating units in each oligomer. If the oligomer used is sufficiently long so that its terminal repeating units make negligible contributions to $E_M$, the energy of N repeating units within an infinitely long nucleic acid simply becomes $N.E_M$.

Apart from helical symmetry, many other constraints can be imposed using Jumna. These include inter-atomic distance and dihedral angle constraints used to fit conformations to NMR data. Other options allow helicoidal variables to be controlled, for example, to twist, stretch or bend an oligomer or to provoke conformational transitions. It is also possible to directly control sugar phase and amplitude which, as we shall see shortly, is of great interest for investigating DNA fine structure.

**Adiabatic mapping**

Jumna has recently been extended for the purpose of carrying out adiabatic mapping in 1 or 2 dimensions. Once the variables (or constraints) to be scanned, their starting values and the step intervals are defined, Jumna can carry out successive minimisations fully automatically, reading in the structure of the nearest neighbouring conformation already obtained, changing the appropriate variables or constraints, performing the new minimisation and writing out the energetic and conformational details of the resulting structure. This has enabled us to look at energy surfaces containing many hundreds of

individually minimized conformations and has greatly extended our understanding of the mechanics of nucleic acids. Happily, this development was programmed at the same time as our passage to 30 Mflop workstations. This allowed maps of 200 points to be calculated in roughly 6-8 hrs for DNA dodecamers. Graphic tools were also generated to analyse the energetics and structural details revealed by these maps and several examples will be shown in the following sections.

## Dynamics and Large Scale Models

Before leaving the discussion of modeling techniques applicable to modeling nucleic acids there are two fields where much theoretical work is still being done. The first of these areas concerns dynamics. As mentioned above although molecular dynamics and Monte Carlo simulations has been applied to nucleic acids both *in vacuo* and with a surrounding shell of explicit water molecules (Goodfellow & Williams, 1992; Beveridge et al, 1992), these simulations are far from being routine. Interesting results have been obtained concerning the behaviour of different force fields (Srinivasan et al, 1990; Swaminathan et al, 1991), the positioning of hydrating water molecules and counterions (Subramanian et al, 1988; Jayaram et al, 1990; Beveridge et al, 1992), hydrogen bonding (Fritsch & Westhof, 1991) and the role of dynamics underlying NOE signals (Withka et al, 1990, 1991, 1992). However, the short time scale of these simulations and the sensitivity of DNA to its environment remain major constraints. Interesting work is underway using Monte Carlo simulations based on internal coordinate representations of DNA which may turn out to be better adapted to simulating larger scale and slower movements (Zhurkin et al, 1990, 1991).

A second area of study which is very promising involves the development of simplified models of the nucleic acids which allow very large scale systems to be simulated. The largest systems to have been treated in atomic detail are probably dynamic studies of $tRNA^{Phe}$ (Harvey et al, 1985), a 4-strand junction model (von Kitzing et al, 1990) and a DNA nodule model (Sprous & Harvey, 1992). These systems involve up to roughly 100 nucleotides. In order to study the behaviour of larger systems, of which DNA plasmids are a good example, some detail has to be sacrificed. Large scale models may also be of great interest for studying the dynamics of intermediate size systems by offering routes for simplifying the mechanics of the system. Work in this field began with phenomenological models of sequence induced DNA curvature (for a recent review, see Trifonov, 1991). More detailed models have allowed the mechanics and dynamics of plasmid supercoiling to be simulated (Tan & Harvey, 1989, 1990) and recent work from Tamar Schlick and Wilma Olson has made very elegant progress in this field (Schlick & Olson, 1992).

# STUDYING BASE SEQUENCE EFFECTS

## Regular Sequences

### *Energy minima: B-DNA*

Our first systematic attempt to understand base sequence effects on DNA conformation and flexibility involved extensive minimisation studies with the six oligomers listed below (Poncin et al, 1992a). These studies turned out to be very important in forming our current views on DNA fine structure and flexibility.

| | |
|---|---|
| CG: | CGCGCGCGCGCGCG |
| TA: | TATATATATATATA |
| TG: | TGTGTGTGTGTGTG |
| GA: | GAGAGAGAGAGAGA |
| GG: | GGGGGGGGGGGGGG |
| AA: | AAAAAAAAAAAAAA |

As explained earlier, these symmetric sequences (referred to from now on by the two-letter codes listed above) contain the 10 unique dinucleotides and should thus give some insight into inter-base pair parameters at the nearest neighbour level. As a bonus they also contain 8 trinucleotides (CGC, TAT, TGT, GTG, GAG, AGA, GGG, AAA) and 10 tetranucleotides (CGCG, GCGC, TATA, ATAT, TGTG, GTGT, GAGA, AGAG, GGGG, AAAA). Our calculations were carried out imposing either mononucleotide or dinucleotide symmetry, according to the sequence in question, for respectively 15 or 14 nucleotide duplexes. Dyad symmetry was also imposed where appropriate - namely for the CG and TA oligomers. The central energy approach, described earlier, was used in all cases to avoid end-effects and to speed up calculations.

Perhaps the most important point to make about these studies is that we attempted to localize all energy minima in the region of conformational space that could be called B-DNA. The formulation of Jumna leads to a reduction in the number of degrees of freedom describing DNA, which, coupled with the use of helicoidal variables, diminishes problems associated with getting trapped in local minima. Nevertheless, since Minimisation algorithms cannot climb energy barriers it is clearly necessary to multiply the starting points used. This involved carrying out initial minimisations for each oligomer starting from B-DNA fibre coordinates (Arnott & Hukins, 1973; Arnott et al 1980) and continuing until all energy gradients were virtually zero. We then used the final conformation of each oligomer as the starting conformation for all the other sequences. These minimisations either led to conformations which had already been seen or resulted in new lower energy states. When such states occurred, the corresponding conformations again served as starting

points for all the other sequences. As a further check we subjected each stable structure to stretching-compression and twisting-untwisting deformations to verify its range of stability and to attempt to cross energy barriers leading to further energy minimal conformations. By this laborious procedure, involving many hundreds of minimisations, we could finally be virtually certain that we had located all the energy minima of each oligomer belonging to the B family.

The first surprising result of this study was that, with the exception of the homopolymeric oligomers GG and AA, each sequence was found to be capable of adopting a small number of rather different conformations with very similar energies. We located respectively 3, 4, 5 and 6 conformations for the dinucleotide sequences CG, TA, GA and TG (see table 4). The full geometrical descriptions of these conformations are given in an earlier publication (Poncin et al, 1992a). However, the results shown in figure 5 suffice to illustrate the important range of structural parameters associated with the stable conformations of each of the dinucleotide sequences.

Table 4. Energies of the minimal energy sub-state conformations for the 6 oligomers studied (Kcal/mol)

| Sequence | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $(CG)_{14}$ | -821.3 | -818.7 | -816.0 | | | |
| $(TA)_{14}$ | -562.1 | -559.7 | -556.3 | -553.1 | | |
| $(TG)_{14}$ | -692.7 | -692.2 | -684.8 | -691.5 | -695.1 | -687.0 |
| $(GA)_{14}$ | -686.9 | -678.8 | -676.1 | -675.8 | -670.6 | |
| $(AA)_{15}$ | -603.1 | | | | | |
| $(GG)_{15}$ | -873.8 | | | | | |

An insight into how these conformational "sub-states" could be characterized came from looking closely at sugar puckering. As figure 6 shows the sugars associated with the 20 energy minima located fell into two puckering families, C2'-endo and O1'-endo. Within the former family a subdivision into two groups having either high phase and low amplitude or lower phase and high amplitude can be seen. We denote these three types of puckering by 'E' (O1'-endo, East on the pseudorotation ring), 'S' (C2'-endo high phase, South on the pseudorotation ring) and 'X' (C2'-endo low phase, close to C1'-eXo). The range of phase angles associated with the S group of puckers is however dependent on whether pyrimidine or purine nucleotides are involved.

If we now classify the conformations listed in table 4 in terms of these notations, it is found that each sub-state has a distinct combination of sugar puckers (table 5). We were thus led to the conclusion that sugar puckering was playing a major role in characterizing the stable minima found. Comparing tables 4 and 5 shows that the most stable conformations are obtained with C2'-

endo sugar puckers as expected in B-DNA. However, O1'-endo sugars can be introduced for pyrimidine nucleotides with only slight decreases in stability and, in one case, even lead to a more stable conformation ($TG_5$). The observation that O1'-endo puckers are limited to pyrimidines is in line with both experimental results (Metzler et al, 1990; Searle & Wakelin, 1990; Privé et al, 1991; Schmitz et al, 1991) and other modeling studies (Zhurkin et al, 1990). For sequences containing both cytosine and thymine it is found that cytidines can change to O1'-endo puckers somewhat more easily than thymidines. It should be noted however that no conformation contains homo-pyrimidine strands with uniquely O1'-endo puckers. Conformations with O1'-endo puckers always have such sugars sandwiched between two C2'-endo sugars. S and X category C2'-endo puckers which occur for both pyrimidines and purines are also seen (with the exception $TA_1$) to occur in alternation within each strand of the dinucleotide sequences.

Table 5. Sub-state conformations classified according to their sugar puckering (S: low amplitude C2'-endo, X: high amplitude C2'-endo, E: O1'-endo, s: intermediate between S and X forms)

| Sequence | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $(CG)_{14}$ | XS | SX | SE | | | |
| | SX | XS | ES | | | |
| $(TA)_{14}$ | ss | XS | SX | SE | | |
| | ss | SX | XS | ES | | |
| $(TG)_{14}$ | SX | XX | XS | SE | SE | SX |
| | XS | SS | SX | ES | XS | ES |
| $(GA)_{14}$ | SX | SS | XS | SE | XE | |
| | XS | XE | SE | XS | SS | |
| $(AA)_{15}$ | ss | | | | | |
| | ss | | | | | |
| $(GG)_{15}$ | ss | | | | | |
| | ss | | | | | |

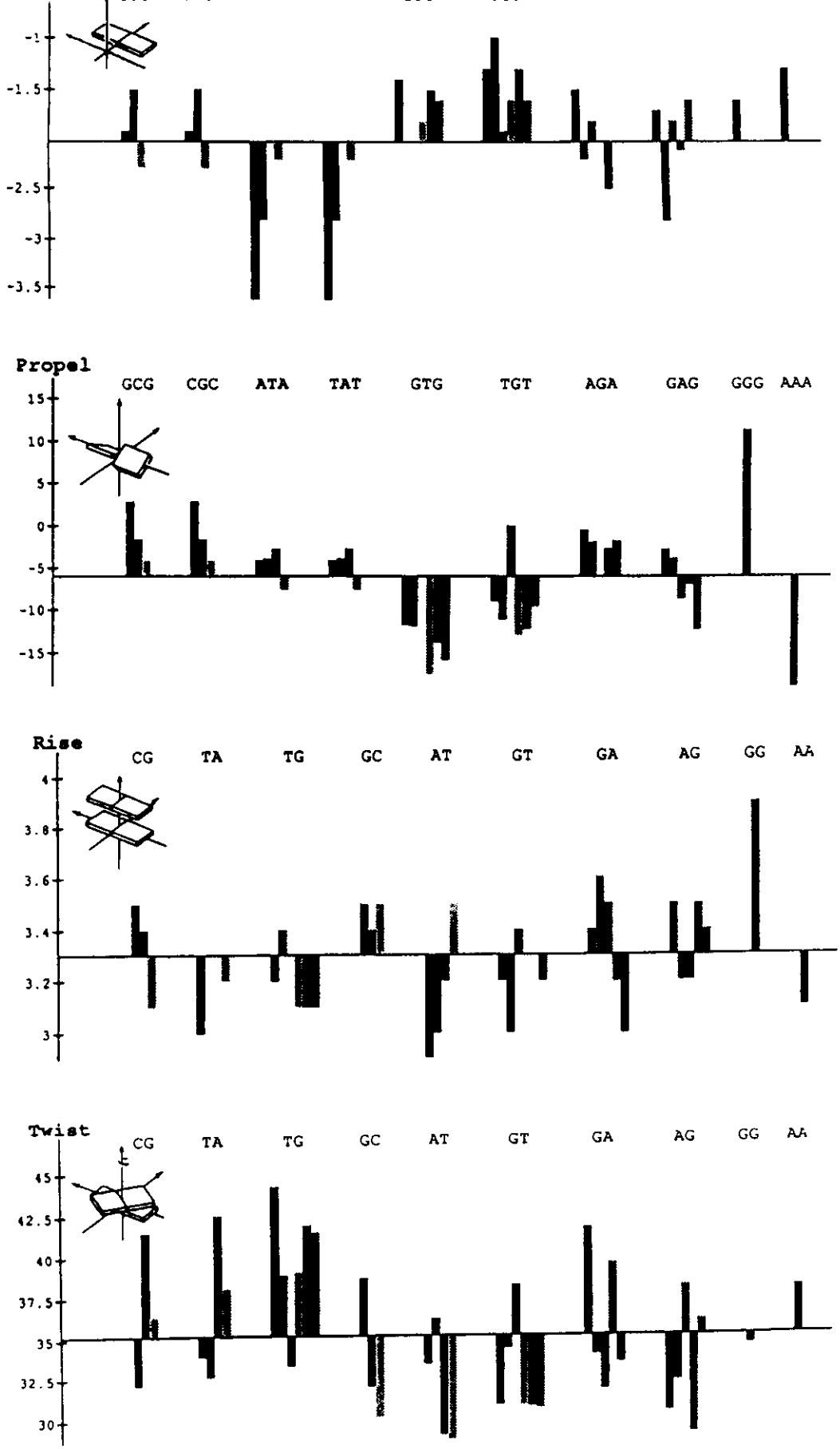
Notice that the sugar puckers of the two strands at a given dinucleotide level may be identical (eg. $CG_1$ SX:SX, writing both junctions in the 5' -> 3' sense) or mixed (eg. $TG_2$ SX:XS or $TG_5$ XS:ES) but only three conformations involve two C2'-endo sugars attached to the same base pair ($TG_2$, $GA_2$ and $GA_5$). The SE:ES combination is clearly forbidden since this would imply the

existence of base pairs S-S and E-E, the latter of which would have two O1'-endo puckers, whereas such puckers are only observed for pyrimidines. Special cases occur for the homopolymers which are forced by the imposed symmetry to adopt uniform puckering. These sugars (indicates by a lowercase 's') are found to lie between the S and X families. Similar puckers are also found for the conformation $TA_1$. It is remarked that more recent studies of poly(dA).poly(dT) show that when this homopolymer is not constrained to respect mononucleotide symmetry, its most stable form actually exhibits dinucleotide symmetry with an alternation of S and X sugar puckers in each strand, although, in this case, such an effect cannot be caused by the base sequence alone.

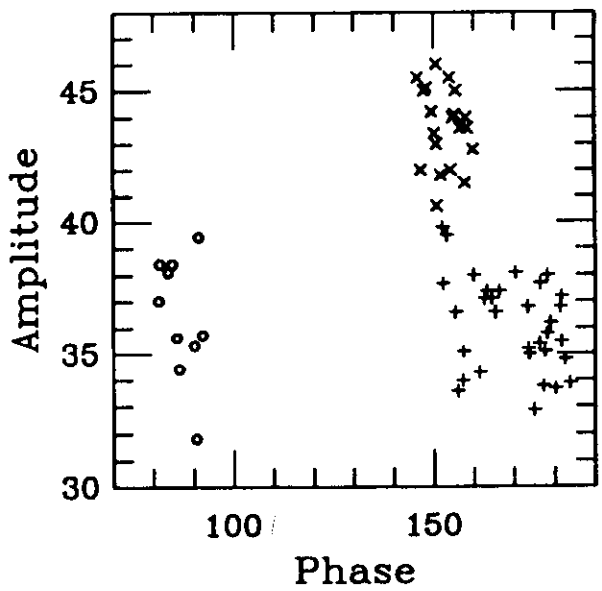
Returning to the conformational parameters describing the different sub-states, we find wide variations in all parameters even if the mean values (corresponding to the horizontal axes in figure 5) are close to those of the canonical B-form. To take one example, twist angles vary from 29° to 44°, a range of 15°. In alternating sequences, high twists are always followed by low twists and vice versa as has been found in both theoretical (Hao & Olson, 1989) and NMR studies (Schmitz et al, 1991). Moreover, although certain dinucleotides show a tendency for above or below average twists, most show important variations between their possible sub-state conformations. Similar remarks apply to other helicoidal parameters.

Backbone dihedrals fall into 3 groups in terms of their variability amongst the sub-states: $\alpha$ and $\gamma$ vary very little, $\beta$ and $\varepsilon$ have a range of about 20° and $\zeta$ is very variable with a range of roughly 50°. These variations are also in line with experimental findings (Saenger, 1984; Roongta et al, 1990; Kaluarachchi et al, 1991) and lie well within the ranges defined by energy minimisation on double stranded dinucleoside monophosphates (Srinivasan & Olson, 1987). Sugar puckering has an important influence on backbone dihedrals. A phosphodiester junction bounded by an O1'-endo sugar on the 3' or on the 5' side leads to greatly reduced variability in all the backbone dihedrals and, in some case, to distinct changes in the mean values. This is also true for the glycosidic dihedral $\chi$ which is found to be more flexible when associated with X family sugars. $\chi$ is also closely coupled to the phase angle of adjoining sugar, becoming more negative as the phase decreases. This feature was also clearly noted in crystallographic studies (Grzeskowiak et al, 1991; Fratini et al, 1982).

For dinucleotide steps where both strands adopt the same sugar puckers we can go further and predict the ranges of a certain number of helicoidal parameters including twist and slide. In these cases, twist is clearly seen to be linked to the 3' sugar of each dinucleotide junction, decreased phase in this sugar leading to smaller twist angles. It may be remarked that recent NMR studies also find qualitative coupling between sugar pucker and twist within alternating base sequences (see, for example, Gronenborn et al, 1984; Schmitz et al, 1990; Gochin & James, 1990). Earlier theoretical studies have also noted




(5) Histograms of selected helicoidal parameters for the B-DNA sub-state conformations. (The horizontal axis is placed at the sequence averaged value of the corresponding parameter. Pale shading indicates conformations containing O1'-endo sugars. Translational parameters in Å, rotational parameters in degrees).




(6) Sugar phase versus amplitude for the 20 sub-state conformations of the 6 regular oligomers (+: 'S' high phase C2'-endo, x: 'X' low phase C2'-endo, o: 'E' low amplitude O1'-endo)

the same coupling, O1'-endo (E) sugars leading to reduced twist (Kollman et al, 1982; Zhurkin et al, 1990).

Apart from the impact of sugar puckers on the sub-state conformations, we have also found a number of strong correlations which are independent of base sequence and thus seem to reflect the underlying mechanics of the double helix. These correlations include pairs of helicoidal parameters such as propeller-inclination and twist-slide, pairs of backbone dihedrals such as $\zeta$-$\varepsilon$ and mixed parameters such as $\zeta$-twist.

To conclude, these studies brought to light the existence of conformational sub-states, characterized by sugar puckering, for all the alternating sequences studied. Only the 2 homopolymeric sequences led to a single stable conformation. The conformations of the different sub-states of a given sequence show important variations, but their stabilities are very similar. Sugar puckers are found to fix, to a large extent, the conformation of the intervening phosphodiester backbone and a number of the structural characteristics of the corresponding base pair step (notably, twist and slide) regardless of the base sequence involved.

### *Mapping the Sugar Plane*

Perhaps the single most important result of the studies I have described is the central role which appears to be played by sugar puckering. This finding has recently led us to look beyond individual minimal energy conformations, in an attempt to get a more global view of the conformational space occupied by the DNA double helix. Since 4 of the 6 oligomeric sequences we have studied contain only two symmetry-distinct sugars (CG, TA, GG, AA), it is possible to use adiabatic mapping to look at the energy plane defined by the phase angles of these sugars. We have seen that sugar puckering characterizes the sub-states of the B-allomorph and is linked to a variety of helical parameters in both the B and A forms, it should thus define the most interesting "slice" through the DNA energy hypersurface.

In the case of B-DNA, this study should enable us to answer important questions concerning the sub-state conformations described above, namely, what energy barriers separate the different sub-states and what are the transition pathways which link them. This information will then determine whether a given oligomeric fragment of DNA would be likely to exist in a single conformational state (naturally undergoing thermal fluctuations at ambient temperatures) or, in contrast, would be in equilibrium between several distinct conformational forms.

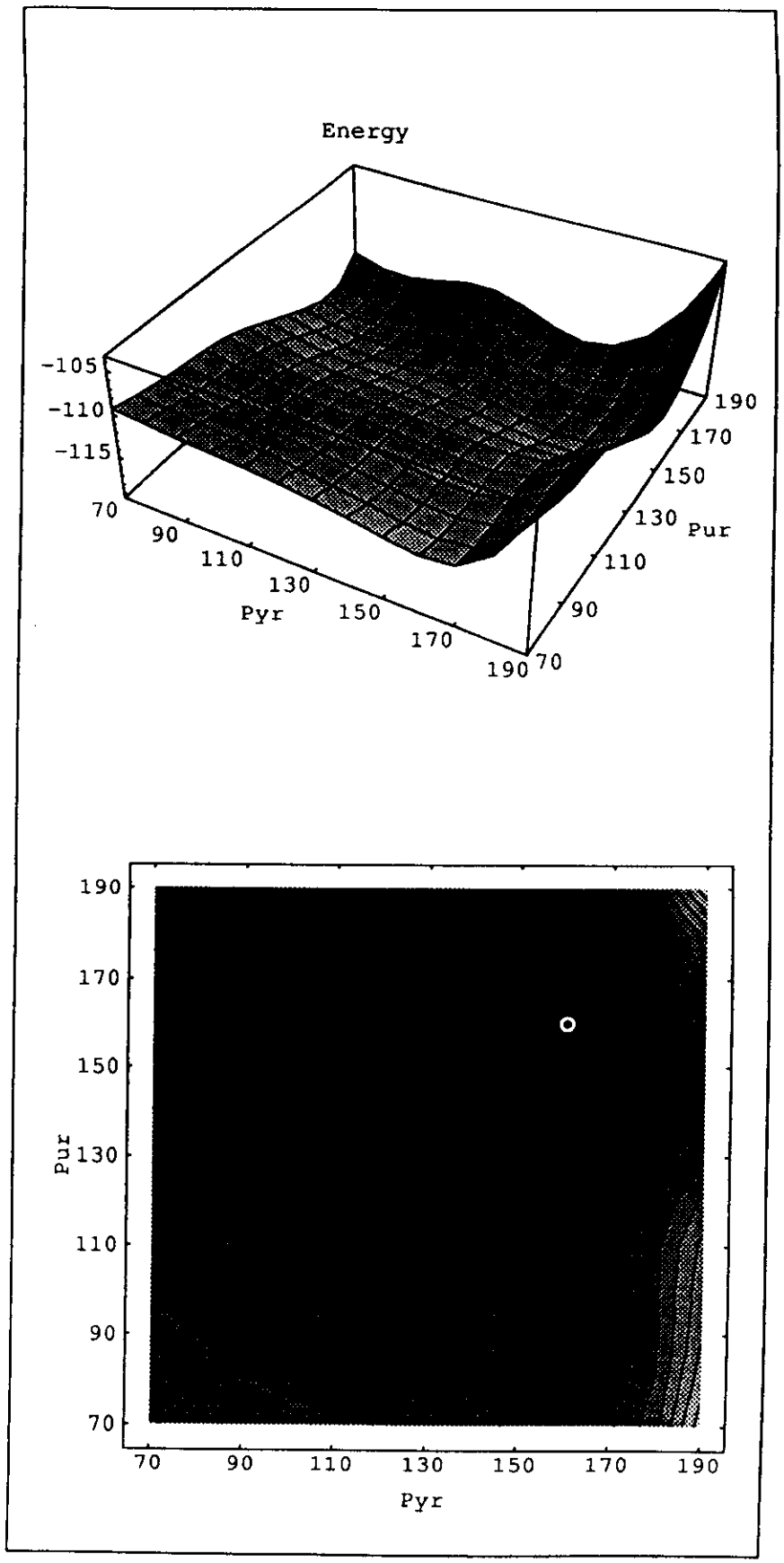
Although this work has only just begun, it has already led to some interesting findings. Firstly, all the stable conformations found by energy minimisation can be located in the energy plane defined by the sugar puckers. Secondly, although maps have been generated using different starting points we have never been able to generate more than one energy surface for a given

base sequence. We have even been able to build up maps covering both the B and A conformational regions (*vide infra*) as jigsaw puzzles, different fragments of the map being calculated separately from different initial conformations. These fragments have always been found to fit together perfectly as pieces of the same energy plane. This implies, at least for the 4 regular sequences under study, that once the sugar puckers are fixed we also fix a unique DNA conformation. Since, within the Jumna model, double helices constrained to homonomous dinucleotide symmetry (CG, TA) or heteronomous mononucleotide symmetry (GG, AA) are represented by respectively 28 and 30 variables, it is a remarkable finding that these structures can apparently be simplified to a 2 variable description, namely their sugar phases.

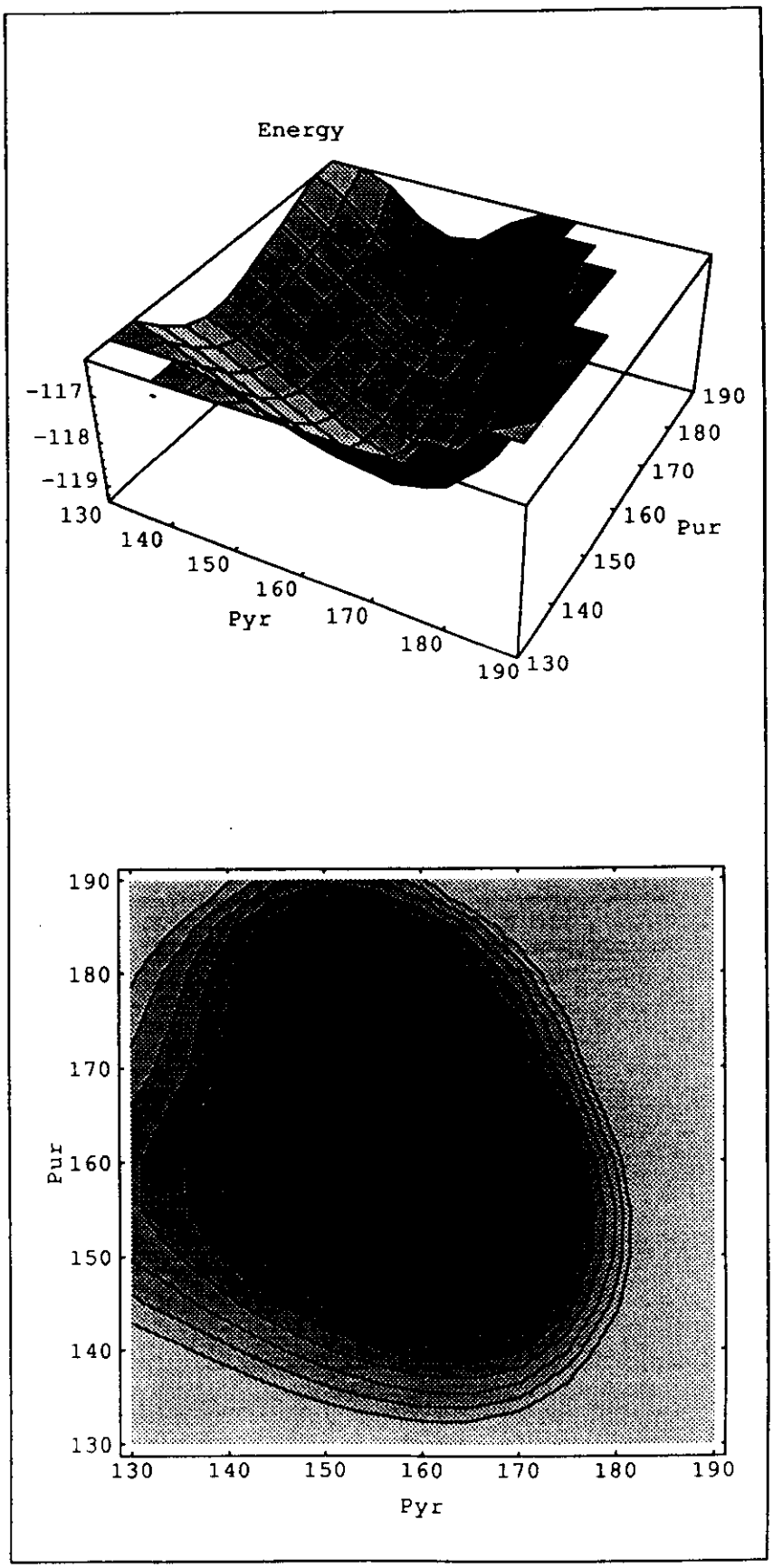
Another surprising result of the adiabatic mapping was that although the conformational sub-states found for B-DNA could be detected in the sugar plane maps, they were in fact generally separated from one another by only very small energy barriers. It is first worth noting that for a very wide variety of sugar puckers the energy of the oligomers studied varied by roughly 2-3 Kcal/mol per base pair. Significant energy penalties were only found for phases above 180°, pyrimidine sugars being particularly resistant to adopting high phase angles. An example of this can be seen in figure 7 for the CG oligomer. The map of the south and east puckering zones clearly shows the principal minimum corresponding to the SX:SX conformation given in table 5. One can also distinguish the SE:ES conformation which is separated by a barrier of roughly 2 Kcal/mol (for the CG.GC repeating unit, that is, roughly 1 Kcal/mol per base pair). Locating the third minimum listed in table 5 (XS:XS) requires a zoom-in view of the south pucker zone (figure 8). This minimum in fact turns out to be little more than a shoulder on the main SX:XS well. It is nevertheless a meaningful point on the energy surface since it characterizes the clearly asymmetric nature of the main well.

These studies are far from complete, but it seems we can already conclude that certain sub-states we have found by minimisation are separated by only negligible energy barriers. The barriers leading from O1'-endo puckering are larger than those separating S and X forms of C2'-endo puckers, but it is possible that at room temperature such oligomers could be in equilibrium between all these sub-states. However, it should be added that such an equilibrium is likely to involve conformational fluctuations which will follow the low energy valleys seen on the sugar plane map. The sub-state conformations can thus be viewed as useful sign posts along these pathways.

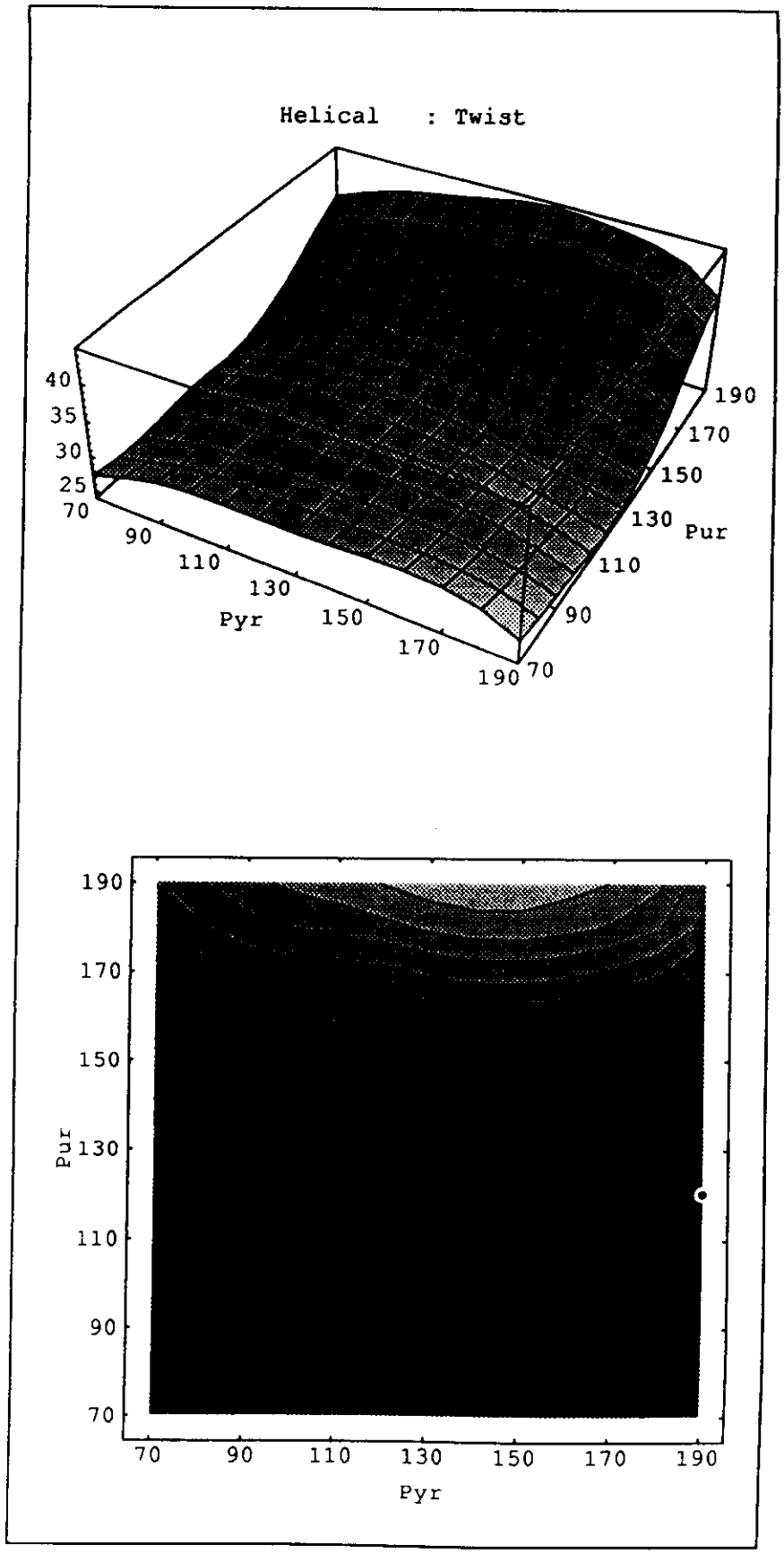
A last remark concerning the sugar plane maps is that they are also helping us to understand the mechanics of the double helix by refining the data we had already collected on correlations between different structural parameters during our energy minimisation studies. Two examples of this are given in figure 9 where we can see Xdisp and twist as a function of sugar



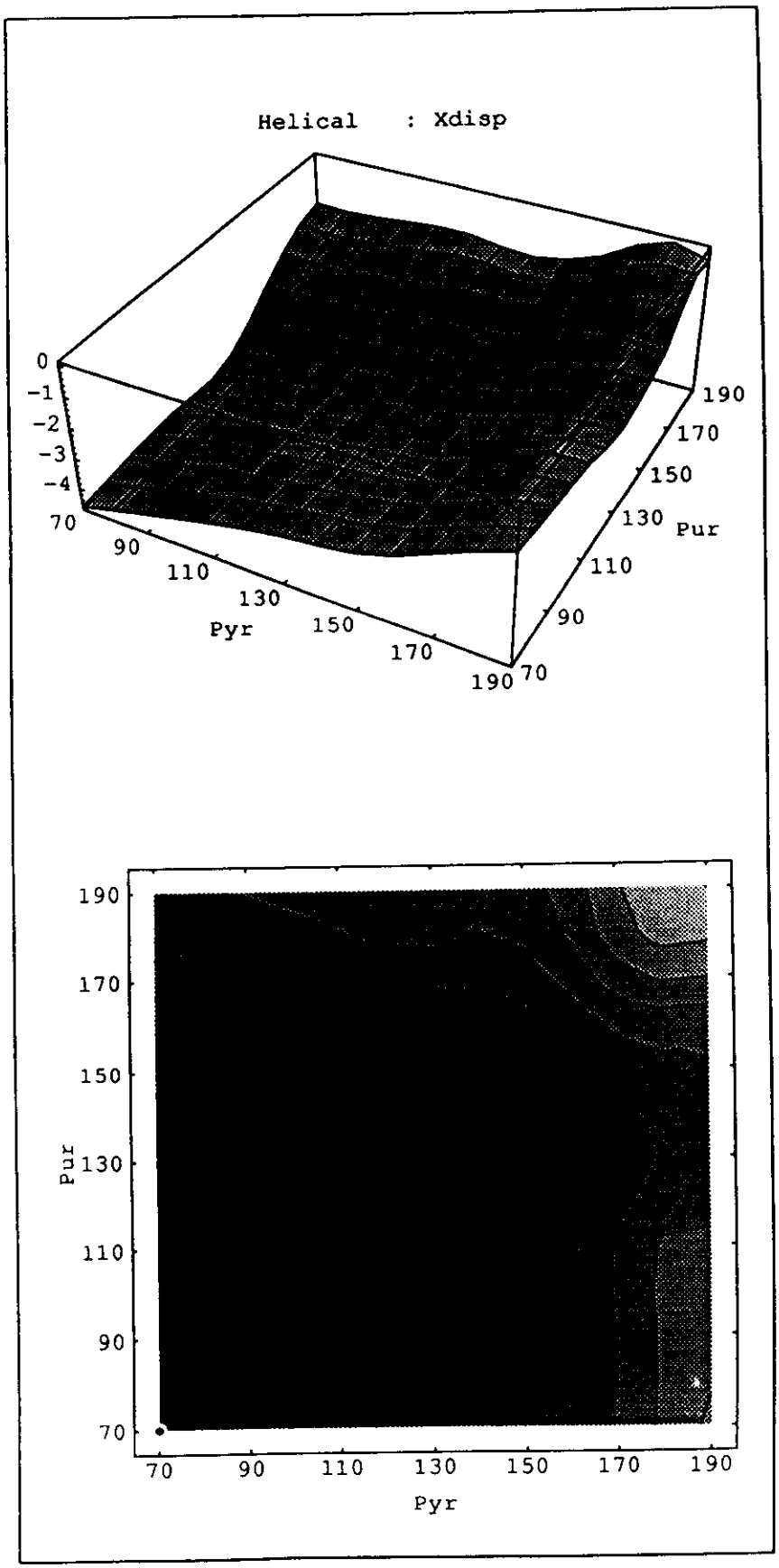

(7) Energy surface and contour map for the CG alternating sequence as a function of its sugar phase angles (Energy in Kcal/mol, phase in degrees).




(8) Energy surface and contour map for the CG alternating sequence as a function of its sugar phase angles: a detailed view of the southern region (Energy in Kcal/mol, phase in degrees).




(9) Helical parameters as a function of sugar phase for the CG alternating sequence: (a) Twist, (b) Xdisp (Xdisp in Å, twist and phase in degrees).


Helical : Xdisp
0
-1
-2
-3
-4
70
90
110
130
150
170
190
Pyr
70
90
110
130
150
170
190
Pur
190
170
150
130
110
90
70
Pur
70
90
110
130
150
170
190
Pyr

pucker - again for the CG oligomer. The correlations between these variables and the sugars are very striking.

## DNA FLEXIBILITY

Understanding the biological role of DNA requires an understanding of the influence of base sequence not only on the stable conformation of the double helix, but also on its flexibility. This is clearly the case for processes involving the recognition of DNA by proteins and drug molecules, where "indirect read-out" of local conformation and local deformability can both play important roles. We have begun an investigation of base sequence effects on flexibility by studying global twisting-untwisting and stretching-compression deformations of the regular oligomers discussed earlier. In this section we will also look at two specific local deformations, one involving base movements and the other involving backbone conformational changes.

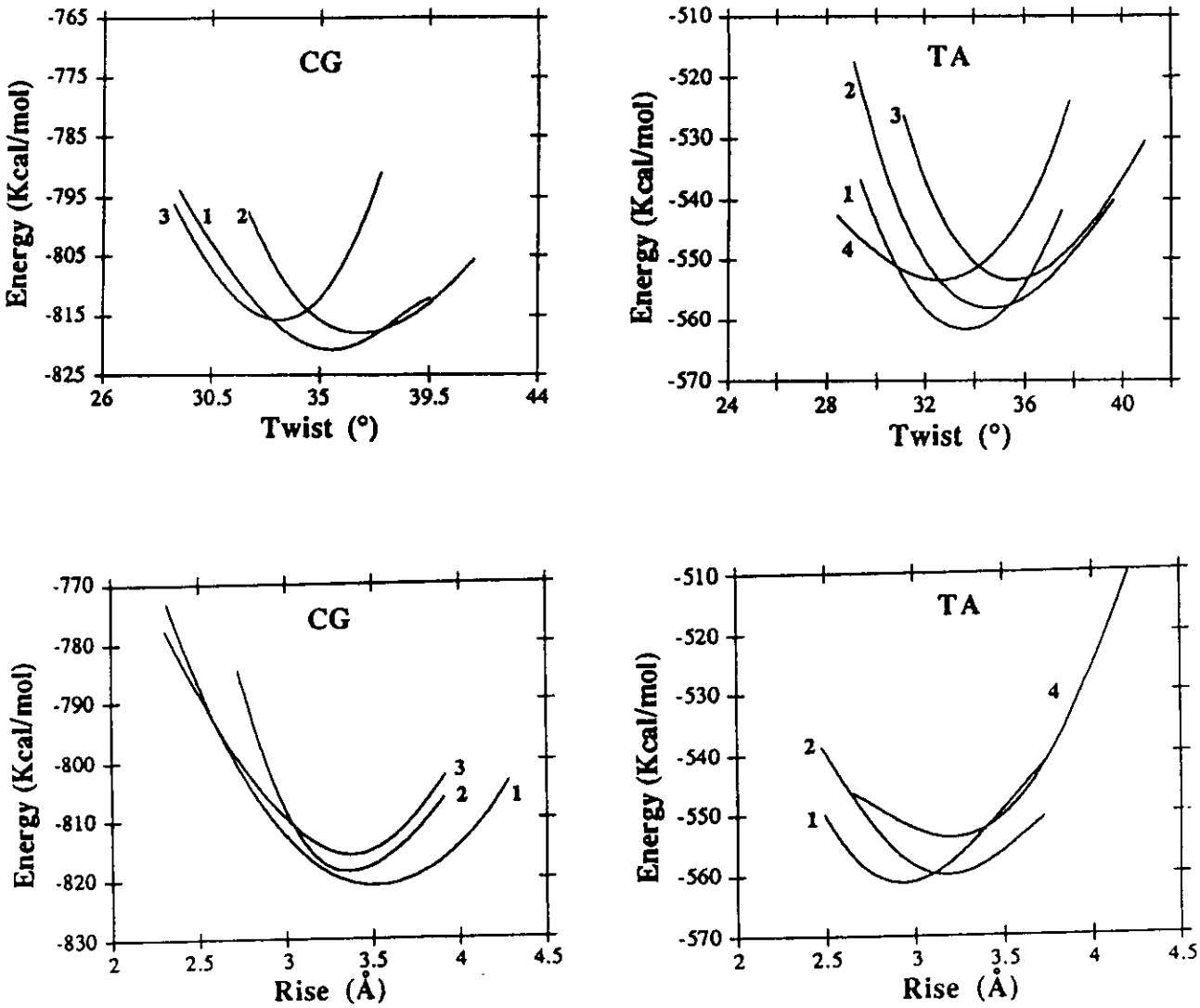
### Twisting and Stretching the Double Helix

In order to study the deformability of the stable conformations of the six regular oligomers discussed earlier, we have used constraint options within Jumna to vary both the total rise and the total twist of the conformations in question. We have considered both the A- and B-forms of each sequence. The deformation imposed covers a range of roughly 15° in twist and 2Å in rise, per base pair step. Note that, in the alternating oligomers, the two distinct base pair steps are free to vary independently.

These global helix deformations led to roughly quadratic energy curves centred around each minimal energy conformation. Some examples are given in figure 10, other data may be found in an earlier publication (Poncin et al, 1992b). A first point which can be made from these results is that such deformations can easily change the relative stability of the B-DNA conformational sub-states. Taking $(CG)_n$ as an example, although the form $CG_1$ is most stable in the relaxed oligomer, overwinding leads $CG_2$ to become more stable, while underwinding favours $CG_3$. Such changes occur more frequently in response to twisting than to stretching as the sub-state conformations differ more in terms of their optimal twist angles than in terms of rise.

Using these deformation curves we can compare the relative flexibilities of A and B form DNA as a function of base sequence. In order to generate quantitative values, we have followed earlier authors (Zhurkin et al, 1982) in calculating the half-width of the deformation curves at 0.5RT (0.3 Kcal/mol) above the minimum. The overall ranges found are as follows:

| | | |
|---|---|---|
| B-DNA: | Twist 1°-4° | Rise 0.1Å-0.4Å |
| A-DNA: | Twist 2° | Rise 0.3Å-0.7Å |




(10) Twisting and stretching deformation energy curves for the B-DNA oligomers: CG and TA alternating sequences. (Curves are numbered following the order of the sub-states given in tables 4 and 6).

A-DNA and B-DNA thus have similar flexibilities with respect to the deformations we have imposed, although the average twist flexibility of A-DNA is slightly lower than that of B-DNA (by 0.1°) and its average rise flexibility rather higher (by 0.2Å). These results are in good agreement with theoretical calculations by (Zhurkin et al, 1982, 1990 and references therein) and experimental results (Millar et al, 1982), but differ considerably from earlier results from our laboratory which included explicit counterions and led to very rigid A-DNA conformations (Lavery et al 1986a). These numbers are also in disagreement modeling which ignored the presence of the phosphodiester backbones (Sarai et al, 1989).

A-DNA shows little sequence dependence effects beyond a weaker resistance to stretching for the $(GG)_n$ and $(CG)_n$ oligomers which, as mentioned earlier, both exhibit large inter-base pair rise. For B-DNA sequence effects are much clearer and we can make a rough classification of the flexibility of the oligomers as follows.

| | |
|---|---|
| Twist: | GG > GA > CG > TA = AA > TG |
| Rise: | GG = GA = CG = TA > TG = AA |

This order however hides the relative flexibilities of the two junctions within each of the alternating oligomers. To obtain a comparison of individual junctions we have thus calculated the ratio of the relative changes in twist or rise for each dinucleotide during deformation. These results indicate that TA junctions are more flexible in twist than AT junctions, in line with hydrogen exchange experiments (Leroy et al, 1988) and earlier theoretical results involving ligand binding (Zakrzewska, 1992), and that GA junctions are also more flexible in twist than AG junctions. Data on torsional flexibility (Hogan et al, 1983; Hard et al, 1989; Zhurkin et al, 1990) would appear to support the similar results we obtain for AA and AT sequences (Hard et al, 1989), but are not clear concerning GG versus AA (Millar et al, 1982; Hogan et al, 1983). In terms of rise, the differences are less marked, but pur-pyr steps appear more flexible than pyr-pur steps.

It is found that, with few exceptions, different B-DNA sub-state conformations of the same oligomer have very similar flexibilities. Thus, the considerable changes in optimal helicoidal and backbone parameters seen for these conformers appear to have little influence on their reaction to external stress. The only clear exceptions to this rule are $TA_4$ and $GA_5$ (in twist) and $TG_3$ (in twist and stretch). These effects may be explained by the fact that all these conformers have the lowest stacking energies of the corresponding oligomers. $TA_3$ also has a low stacking, but is presumably less flexible because of its unusually large twist in the optimal conformation.

How does the double helix respond to imposed stress? The results averaged over sequence are summarized in table 6. For B-DNA, little change is

seen in the inter-base pair parameters, apart from rise and twist when these parameters correspond to the deformation imposed. Twist deformation of B-DNA shows that Xdisp is directly linked to twist and varies to almost the same extent independent of sequence or sub-state conformation. Inclination changes principally with overwinding becoming more negative in all cases. Propeller shows the opposite behaviour by varying more on underwinding and becoming more positive. There is little sequence effect for either of these parameters and only minor sub-state effects are seen. Rise deformation of B-DNA again causes changes in Xdisp, inclination and propeller. Xdisp changes are rather uniform, but with a tendency to diminish more under compression than to increase upon stretching. Inclination shows a uniform inverse correlation to rise deformation. Propeller generally becomes rapidly more positive under stretching and only slightly more negative upon compression. Rise shows clear differences between the junctions within the alternating sequences, the pur-pyr steps being more flexible than pyr-pur steps in all cases. Both types of deformation also lead to considerable changes in the backbone dihedrals (with the exception of $\alpha$, in the case of twist deformation, and $\varepsilon$). The glycosidic angle also changes by roughly 10°, as does the sugar phase.

For A-DNA, inter-base pair parameters again vary little, with the notable exception of roll which is affected by both types of deformation. The major changes again occur in Xdisp, inclination and propeller, although the latter parameter is only involved under twisting stress. Xdisp and propeller show similar behaviour in the A and B allomorphs. Inclination, on the other hand, is reversed, decreasing with overwinding in B and increasing with overwinding in A. Sequence effects on Xdisp and propeller are minor. Backbone dihedrals vary only slightly and generally less than in B-DNA. Only $\alpha$ and $\chi$ show changes of 10° or more, and this, only in the case of twist deformation. It can also be remarked that sugar phase changes considerably less for A-DNA than for B-DNA, in line with the much smaller scatter of phase angles observed in the optimal conformations of the A form oligomers. This effect has also been noted in earlier modeling (Zhurkin et al, 1982).

By looking at the wider range of conformations generated by twisting and stretching deformation it is possible to see a number of structural correlations which are not always visible on the basis of the energy minima alone. Correlations have been found between pairs of helicoidal parameters, between pairs of backbone parameters and for mixed helicoidal-backbone pairs that hold for both the A and B allomorphs. In the first category we can note inclination and rise which show an overall correlation coefficient of C= -0.81 (note that when a junction parameter is compared to a base pair parameter, as in this case, the latter parameter is averaged over the two base pairs bordering the junction in question). Inclination is also coupled to propeller and Xdisp although less strongly. Coupling is also found between slide and twist (C= 0.65). Several backbone dihedral angle correlations are common to both

allomorphs such as that shown between α and ε. A very strong correlation is also seen between the sugar phase and the glycosidic angle χ. Phase is also seen to be coupled to both ε and ζ (on the 3'-side of the sugar).

Table 6. Total changes in helicoidal and backbone parameters of B-DNA and A-DNA as a result of deformation (values are base sequence averages).

Axis-Base pair and Intra-Base pair:

| | B-DNA | | A-DNA | |
|---|---|---|---|---|
| | Twist | Rise | Twist | Rise |
| X disp (Å) | 3.2 | 2.8 | 1.1 | 2.1 |
| Y disp (Å) | 0.2 | 0.2 | 0.2 | 0.1 |
| Inclin (°) | 14 | 33 | 22 | 23 |
| Tip (°) | 2 | 3 | 6 | 3 |
| Buckle (°) | 5 | 6 | 8 | 2 |
| Propeller (°) | 14 | 15 | 25 | 5 |
| Inter-Base pair: | | | | |
| Shift (Å) | 0.3 | 0.1 | 0.2 | 0.1 |
| Slide (Å) | 0.3 | 0.3 | 0.3 | 0.2 |
| Rise (Å) | 0.3 | 1.2 | 0.9 | 1.1 |
| Tilt (°) | 2 | 2 | 2 | 1 |
| Roll (°) | 3 | 4 | 14 | 8 |
| Twist (°) | 8 | 3 | 9 | 2 |
| Backbone and sugar pucker: | | | | |
| α (°) | 4 | 8 | 10 | 4 |
| β (°) | 9 | 6 | 6 | 4 |
| γ (°) | 10 | 6 | 5 | 5 |
| δ (°) | 9 | 5 | 5 | 2 |
| ε (°) | 3 | 3 | 5 | 3 |
| ξ (°) | 10 | 7 | 3 | 2 |
| χ (°) | 8 | 12 | 14 | 4 |
| Phase (°) | 15 | 9 | 6 | 3 |
| Amplitude (°) | 3 | 2 | 5 | 2 |

If we limit ourselves to B-DNA deformations alone both inclination versus Xdisp and slide versus twist correlations are seen and it is found that slide is more constrained at low twist. There is also an interesting correlation

between sugar phase and Ydisp which indicates that O1'-endo sugars require a lateral displacement of the bases and also that sugars in the high phase sub-group of C2'-endo (termed 'S') lead to more flexibility in Ydisp than to the low phase C2'-endo 'X' sugars. For A-DNA, propeller and twist are correlated, as are inclination and twist, and propeller and the glycosidic angle $\chi$. Twist is also strongly coupled to the backbone dihedral $\alpha$ which was seen found to be the most flexible dihedral under twist deformation for A-DNA. $\alpha$ is also found to be coupled to inclination and to propeller.

Lastly, it is perhaps interesting to note how the grooves of the double helix react to imposed deformations. The results most relevant for biological interactions probably concern the twist deformation of B-DNA. In this case, the minor groove width changes much more than that of the major groove and narrows more with overwinding than it opens with underwinding. Sequence effects are minor, but AT rich oligomers are more susceptible to groove narrowing. With the exception of $(GG)_n$, the major groove uniformly closes with underwinding, but is almost unaffected by overwinding. The minor groove becoming deeper upon overwinding and the major groove shallower. These effects are again largest in AT rich sequences. It should be noted from these results that whereas the sum of the groove depths remains roughly constant, this is not true of groove widths and it is perfectly possible for one groove to become narrower without the opposing groove opening up. Rise deformation of B-DNA leads to globally similar effects on groove geometry, stretching playing the same role as overwinding. For A-DNA relatively little change is seen under deformation. Under rise deformation, the deep and narrow major groove behaves very much like that of B-DNA. Twist deformation however leads to different effects, overwinding provoking major groove narrowing for all sequences and underwinding resulting in a reduction in groove depth. It is remarked that groove geometry has also been the subject of an article which goes beyond the present findings by investigating the response of the double helix to forced changes in groove width (Boutonnet et al, 1992).

## Hydrogen Exchange: The Opening of Base Pairs

One specific, but very important type of local DNA deformation, involves the opening of base pairs. This process is a necessary part of both DNA transcription and replication. It has been extensively studied by hydrogen exchange (Englander & Kallenbach, 1983) which follows reactions involving the exchange of labile protons (imino and amino groups) belonging to the nucleic acid bases with the surrounding solvent. Such exchange, within double stranded nucleic acids, requires opening one or more base pairs in order to expose protons which are otherwise sterically hindered by their participation in hydrogen bonds. Despite the accumulation of a considerable amount of data concerning the exchange kinetics, little is known about the structure of the

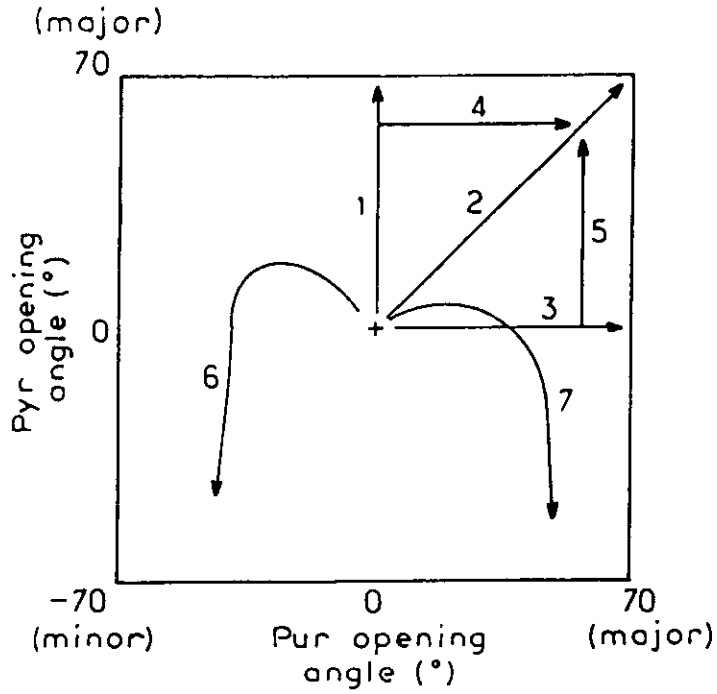
transitory open state. Moreover, since this exchange is a rather slow process typically occurring in the millisecond range (Guéron et al, 1987; Hartmann et al, 1986), it is well beyond the scope of normal molecular dynamics simulations.

In order to study the energetics of base pair opening we began by forcing the opening of the central pair within a B-DNA oligomer $(dA)_5.(dT)_5$, using distance constraints between pairs of atoms belonging to the Watson-Crick hydrogen bonds (see also Keepers et al, 1984). These studies showed that the bases opened mainly by a rotational movement and that a mean axis of rotation was perpendicular to the base plane and passed through a point close to the centre of the sugar ring attached to the opening base. This information was put to use in a preliminary study of opening with a simplified model of B-DNA consisting of five stacked A-T base pairs in the standard B-conformation (Arnott et al, 1980), without phospho-diester backbones. Many different opening pathways (see figure 11) could be tested rapidly using this model (Ramstein & Lavery, 1988, 1990). Interactions between the rotating bases showed that opening occurred most easily towards the major groove. Adenine and thymine could rotate into this groove either independently or in a concerted fashion. Rotating either base towards the minor groove led to severe steric hindrance, unless its partner was first rotated some way into the major groove. Both breaking the base pair hydrogen bonds and decreasing stacking interactions with the neighbouring base pairs were seen to contribute to the cost of the opening process. It was also found that a variety of open states could be formed with roughly similar stabilities. The various pathways studied and the deformation energies necessary to rotate a base by 50° are listed in table 7.

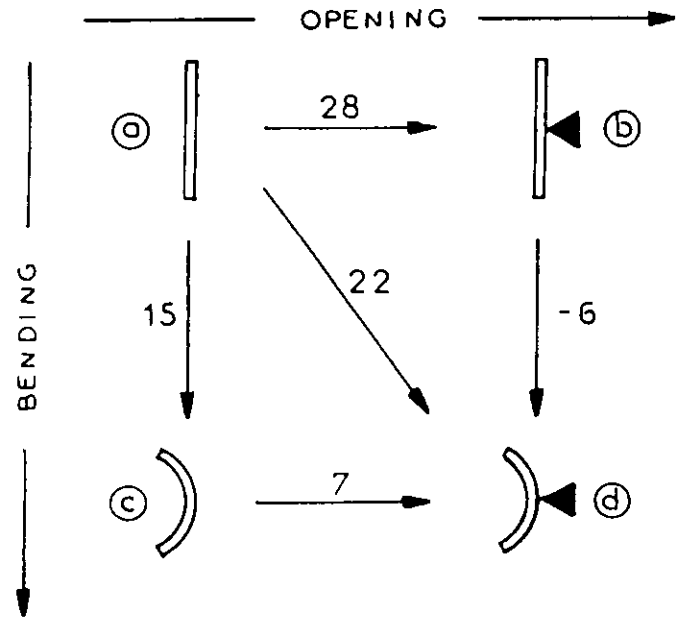
Table 7. Base pair opening energies as a function of the radius of curvature (R) of DNA. The open state corresponds to a 50° base rotation (or a 50° path length in the case of the concerted opening routes 6 and 7). The second column gives the energy deduced from the simplified base only model.

| Path | Base only | R=∇ | R=30Å | R=15Å |
|---|---|---|---|---|
| 1 | 15.5 | 28.4 | 24.7 | 21.8 |
| 2 | 23.5 | 37.4 | 33.2 | 33.8 |
| 3 | 16.4 | 30.0 | 26.5 | 30.2 |
| 4 | 23.5 | 39.2 | 34.4 | 35.7 |
| 5 | 23.5 | 37.5 | 33.2 | 34.1 |
| 6 | 10.2 | 29.3 | 24.8 | 32.7 |
| 7 | 13.5 | 25.0 | 20.2 | 26.5 |

The desire to extend these studies to a full DNA oligomer was one of main drives that led to the development of the Jumna program, since in order




(11) Schematic diagram of the base opening pathways studied.




(12) Thermodynamic cycle for DNA bending and thymine opening: bending favors opening and vice versa. (Deformation energies correspond to bending to a radius of curvature of 15Å and opening thymine towards the major groove by 50° rotation).

to control the opening rotation it was necessary to define nucleotide positions in terms of helicoidal coordinates during energy minimisation. This possibility had existed in the SIR (Successive Infinitesimal rotations) technique developed earlier with Heinz Sklenar (Sklenar et al, 1986; Lavery et al, 1986a, 1986b), but had been lost when this numerically inefficient approach was replaced by the Cinflex program (Lavery et al, 1986c) which defined molecular conformations only in terms of dihedral and valence angles.

Using Jumna it was possible to regenerate the pathways studied with the simple model, but now starting from an energy minimized oligomer conformation and including the hindrance and inter-nucleotide coupling due to the sugar-phosphate backbones (Ramstein & Lavery, 1988). The results shown in the 3rd column of table 11 indicate that the findings of the simple model were in fact quite realistic. Opening to 50° within the full oligomer is seen to be roughly 13 kcal/mol more difficult than within the base only model, but the relative energies associated with different pathways are hardly changed. It should also be remarked that opening one or both bases belonging to the central pair of the oligomer had very little effect on the rest of the molecule. The base pairs on either side of the opening pair showed only minor deformations and most of the backbone deformation was absorbed in the phosphodiester linkages on either side of the opening nucleotides.

An analysis of the deformation energy associated with opening showed that, for small opening angles, the major energy component came from base-base interactions. At roughly 30° rotation, the base pair hydrogen bonds were more or less destroyed, but base stacking interactions are still present and continue to exert a restoring force on the opening base. As the opening angle increases further, the limit of the flexibility of the phosphate-sugar backbones is reached and backbone deformation becomes the principal component of the deformation energy.

These studies with Jumna however brought to light a new and unexpected feature of the opening process. It was noted during the first studies, which involved rotating thymine into the major groove (path 1), that the stability of the open state could be improved by bending the oligomer towards the minor groove. Similar results were subsequently observed for most of the alternative opening pathways (see table 11). This effect was found to be due to strain within the backbones which builds up as DNA is bent. Opening a base pair releases this strain by locally increasing the flexibility of the double helix (Manning, 1983). Opening and bending were thus found to be closely linked. Bending DNA makes opening easier, opening a base pair makes bending DNA easier. The coupling of these deformations is summarized for the case of path 1 by the thermodynamic cycle shown in figure 12. A further implication of this coupling is that bending the double helix could serve as a mechanism to concentrate strain at a given point and triggering base pair opening.

In a subsequent study (Briki et al, 1991) used the deformation energy curve generated for path 1 as the basis for a Brownian dynamics study of the opening process. Such dynamics involve numerically integrating a Langevin equation which takes into account the equilibrium restoring force on the base (coming from our studies), a frictional drag and a random torque due to base-solvent interactions. This approach enables much longer simulations than conventional dynamics. Studies covering 0.08ms were made and showed a proportionality between the inverse log of the time interval separating 2 fluctuations and the opening angle achieved. Extrapolation to opening angles corresponding to breaking the base pair hydrogen bonds led to base pair lifetimes of 15ms at room temperature and an activation energy of roughly 20 Kcal/mol. Both these values are in good qualitative agreement with recent experimental measurements (Leroy et al, 1985). The physical interpretation of this study is that opening can occur by a stochastic random-walk process and its slow kinetics are associated with the large scale base rotations which have to be achieved.

## JUNCTIONS AND TRANSITIONS BETWEEN ALLOMORPHIC FORMS

We will end this section by discussing one of the first applications in which we were able to profit from the direct use of helicoidal parameters by Jumna. This involved the construction of junctions between the different allomorphic forms of DNA (Lavery 1988a, 1988b) and the search for transition pathways between these forms. We carried out such studies for both B-A and B-Z junctions for a single base sequence, poly(dG-dC).poly(dG-dC). Junctions were built using trinucleotide double-stranded fragments extracted from energy optimized hexameric oligomers in the A, B and Z conformations. After preliminary tests, we chose a higher electrostatic damping (S=0.356) for these minimisations than we normally use for B-DNA. This value which leads to a dielectric function close to that proposed by (Hingerty et al, 1985) and which led to better conformations for both the A and Z allomorphs, notably in terms of Xdisp and twist.

### *B-Z junctions*

Since the difference between the B and Z conformations are the most striking, it is perhaps interesting to consider this case first. As the base sequence chosen has a dinucleotide repeat, it is possible to build two distinct junctions, the interface between the two allomorphic forms occurring either at a 5'-CpG-3' step (termed type I) or at a 5'-GpC-3' step (termed form II). Each of these junctions was roughly formed on a graphics system by simply juxtaposing fragments of optimized Z-DNA and B-DNA. It is worth noting that this procedure avoided any temptation to maintain a common helical axis

across the junction, the juxtaposition primarily being aimed at rejoining the phosphodiester backbones.

The resulting structures were then analysed to obtain the "kink" parameters describing the axis disruption at the interface. With this data it was possible to reconstruct a hexanucleotide complex within Jumna and carry out a new energy minimisations. The results obtained showed that both B-Z junctions were indeed principally characterized by an important dislocation of the helical axis (roughly 10Å) and also by an important axis bend of 30°-40°. Apart from these features, very little change occurred in the conformations of the B and Z fragments on either side of the junctions.

The results shown in table 8 indicate that in the case of the type I junction the axis bends into the major groove of the B-fragment, while, for type II, junction bending is towards the minor groove. This can be explained by the fact that the B to Z transition occurs by right-handed rotation of the base pairs around the G -> C long axis (Harvey, 1983). During the transition, guanosine changes its glycosidic angle from anti to syn, thus hardly distorting the backbone linkages. In contrast, cytidine hardly changes its glycosidic angle. This would result in breaking the phosphodiester linkage on the 3'-side of the nucleotide unless some compensation is made. One way to achieve this is simply to incline the Z-fragment in the directions noted above - towards the major groove when the first Z base pair is GC and the minor groove when the first pair is CG. This observation also has an effect on the relative stability of the two junctions since, for type II junctions, the Z-fragment fits neatly into the B-DNA minor groove allowing considerable stacking interactions to be conserved at the interface. For type I junctions, steric hindrance between the B and Z bases at the interface forces a much greater rise and consequently inhibits stacking. The result is that the type I junction costs almost twice as much to form as the type II junction (see table 8).

Table 8. Energy optimised B-Z junction parameters (The kink direction is measured by right-hand rotation with respect to the major groove dyad at the junction. Translation Å, rotation degrees, energies in Kcal/mol)

| Parameter | Type I | Type II |
|---|---|---|
| Ax | -13.3 | -8.7 |
| Ay | -1.2 | -1.0 |
| Ainc | 13.8 | 1.8 |
| Atip | 42.1 | -27.6 |
| Rise | 7.1 | 1.3 |
| Twist | 4.3 | -17.2 |
| Kink angle | 44 | 28 |
| Direction | 49 | -123 |
| ΔE | 19.3 | 10.3 |

(13) Steps along the B-Z transition pathway. The top left hand image shows a type I B-Z junction with a B-DNA segment below and a Z-DNA segment above. Starting top left and rotating anti-clockwise to top right corresponds to the passage of one nucleotide pair from the B to the Z conformation and the creation of a type II junction.

The alternating directions of bending for the two types of B-Z junction led us to suggest a possible transition pathway between these allomorphs. Since the base pairs must rotate 180° in passing from the B to Z conformation it is necessary to facilitate this process by opening up the equivalent of an intercalation site on one side of the transiting pair. The location of this site should be above GC pairs and below CG pairs due to the right-handed sense of rotation around the G -> C axis. As a given base pair undergoes its rotation, the junction changes its type. For rotating CG pairs, the junction changes from type I to type II. This means that the Z-fragment changes its bending direction from the major groove to the minor groove, or, in other words, rotates anti-clockwise around the G -> C axis of the rotating pair. The consequence of these contrary rotations is that stacking energy at the junction is recovered before the rotating pair has completed its full 180° turn. For CG pairs the same is true, the base pair and the junction bend again rotate in opposite directions. We concluded that the B-Z interface can progress along the double helix by a flip-flop movement coupling base pair rotation and junction bending. Some preliminary calculations of this pathway (illustrated in figure 13) were carried out by constraining the tip angle of the bases at the B-Z interface. The results suggested that the energy barrier should not exceed roughly 25 Kcal/mol. The most difficult point along the path was found roughly at the half-way stage when the rotating base pair was vertical and entirely contained within the double helix. At this point, the phosphodiester strands are forced apart and are under strain. In one simulation, this strain resulted in breaking the transiting pair, whose component bases then comfortably stacked on one another within the helix cavity. Further rotation pulled the bases apart and reformed the usual Watson-Crick base pair. Further studies however remain to be done to refine this pathway.

It will be noted that the pathway we propose concentrates on base pair rotation rather than the passage from a right-handed helix to a left-handed helix, which is often thought of as the major difference between B- and Z-DNA. In fact, this change can occur very easily and has little influence on the B-Z interface which, according to our model, has relatively weak base stacking, almost no twist and is moreover disrupted by the transitory unstacking necessary to rotate intact base pairs. This study should also illustrate the advantage of being able to manipulate, constrain and minimize DNA in terms of helicoidal coordinates. This facility led to our being able to form junctions which were essentially limited to a single dinucleotide step. Recent experimental results from (Dai et al, 1992) appear to confirm this finding.

### *B-A junctions*

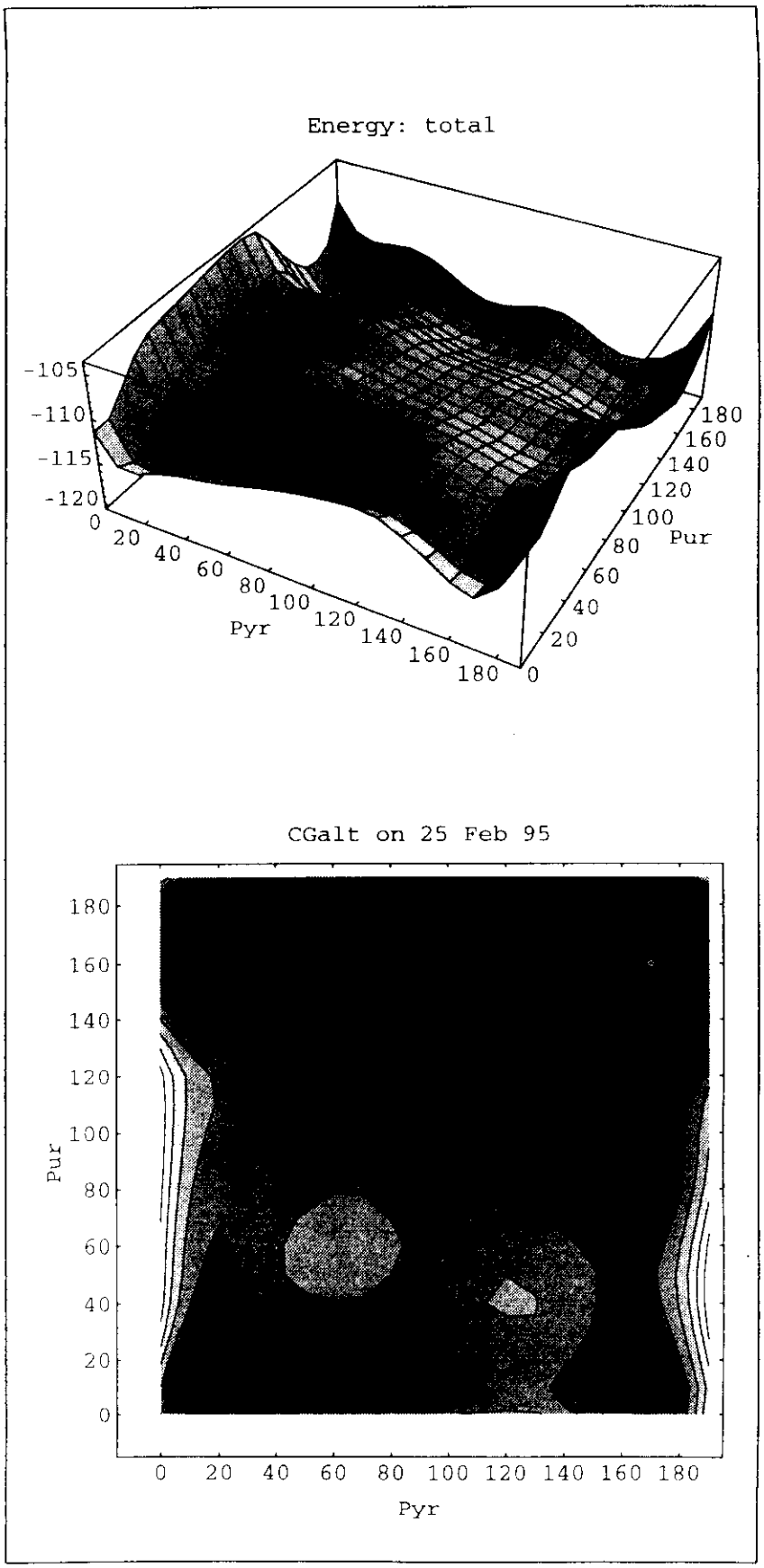
We now turn to B-A junctions, which were formed in the same way as the B-Z junctions discussed above, again starting from optimized oligomer conformations of each allomorph. Not surprisingly, it was found that B-A

junctions could be formed at almost no energy cost since neither unstacking or backbone strain are induced at the B-A interface. As table 9 shows significant axis breaking occurs at the junctions due to the larger base pair Xdisp and shift found in the A allomorph. Changes in inclination also lead the A oligomer to bend by roughly 15° in a direction intermediate between the 5'-3' strand and the minor groove of the B oligomer. There are small differences between type I and type II junctions due to the influence of the alternating base sequence, but these are insignificant compared to the differences found at the B-Z interface. One interesting feature observed during the formation of these junctions was that although relatively little helical deformation occurred in either the B or A oligomer fragment, there was a change in backbone geometry, the first A oligomer sugar in the 5'-3' chain flipping from its normal C3'-endo pucker to either O1'-endo or C2'-endo.

Table 14. Energy optimised B-A junction parameters (The kink direction is measured by right-hand rotation with respect to the major groove dyad at the junction. Translational parameters in Å, rotational parameters in degrees, energies in Kcal/mol)

| Parameter | Type I | Type II |
|---|---|---|
| Ax | 1.6 | 2.2 |
| Ay | 1.5 | 0.3 |
| Ainc | -9.6 | -10.8 |
| Atip | -7.5 | -13.3 |
| Rise | 3.1 | 4.0 |
| Twist | 30.2 | 27.9 |
| Kink angle | 12 | 17 |
| Direction | -155 | -142 |
| ΔE | 0.5 | 0.4 |

Lastly, a few words on the B-A transition. It can be remarked that while the B to Z transition we have discussed must progress step by step along the helix, the B to A transition can reasonably occur smoothly for a whole polynucleotide. An insight into this conformational change, which is principally characterized by the C2'-endo (S) to C3'-endo (N) sugar repuckering, can be gained by extending the sugar plane energy maps discussed earlier in this chapter. Figure 14 illustrates the results obtained for poly(dG-dC).poly(dG-dC) with sugar puckers ranging between 0° and 180°. Beyond the classical B minimum and the SE:SE sub-state already described (see figure 7), we can now see an A-form minimum in the lower left-hand corner of the figure and two hybrid conformations which combine equal




(14) Energy surface and contour map for the CG alternating sequence as a function of its sugar phase angles: the full north-south region (Energy in Kcal/mol, phase in degrees).

numbers C2'-endo and C3'-endo puckers. The lower energy regions of the map occur around its periphery indicating that it is preferable for one sugar to make the S-N repuckering transition at a time. Note that the hybrid forms, which consequently appear to be intermediates in the B-A transition, combine S and N sugar puckers in each strand of the helix and are thus unlike the heteronomous DNA structures where only the two strands differ from one another (Arnott et al, 1983; see, however, Rao & Kollman, 1985 for modeling of intra-strand mixed sugars). It is also found that the hybrid form having N sugars associated with its pyrimidines is more stable, in line with the fact that intermediate O1'-endo puckers are only found for pyrimidine nucleotides. This adiabatic map, whose analysis is still underway, can also certainly give us helpful insights into the mechanism of the transition.

## CONCLUSIONS

I have tried to give an overview of our approach to modeling nucleic acids, both in terms of the theoretical tools which have been developed and in terms of the application of these tools to biologically relevant problems. Several areas of theory clearly need further development. Amongst these I would place, firstly, new methods for rapidly and realistically modeling the influence of solvent and of counterions, secondly, techniques for studying dynamic processes which are slower than the time limits imposed by classical MD simulations and, thirdly, extensions of large scale models for studying systems with hundreds of nucleotides. There is also certainly room to improve the basic force fields applied to nucleic acid modeling which, in general, have received much less attention than that directed towards protein studies.

On the positive side, I think the use of a combination of helicoidal and internal coordinates for modeling nucleic acids has played a major role in any progress we have achieved. This direction of research which grew out of collaborations with Heinz Sklenar has led to set of successively more sophisticated programs, the latest of which, Jumna, really opened the door to the studies of structure, flexibility and conformational transitions which have been described. I think we can also say today that the problem of analyzing the conformation of nucleic acids has been solved. Even if a consensus has not yet been reached on the best way to calculate the helicoidal parameters of irregular nucleic acids, everyone in the field is now aware of the problems that obtaining such a description poses and, more importantly still, that such detailed descriptions are necessary for understanding base sequence effects.

Our studies of base sequence effects are far from their ultimate goal of predicting fine structure and flexibility from base sequence alone. However I think we now start to see a pathway which leads in the right direction. Recent work on both regular and irregular sequences has emphasized the central role of sugar puckering in determining other structural features of the nucleic acid helices. This realisation has led us to adiabatic mapping in the sugar plane

which, I hope, we enable us to make an important step forwards. Understanding the link from base-base interactions to the choice of sugar geometry is clearly at the heart of the story.

## REFERENCES


Arnott, S. (1970). The geometry of nucleic acids. Progress in Biophys. and Mol. Biology 21, 265-319.

Arnott, S. & Hukins, D.W.L. (1973). Refinement of the structure of B-DNA and implication for the analysis of X-ray diffraction data from fibres of biopolymers. J. Mol. Biol. 81, 93-105.

Arnott, S. & Selsing, E. (1974). Structures for the polynucleotide complexes Poly(dA).Poly(dT) and Poly(dT).Poly(dA).Poly(dT). J. Mol. Biol. 88, 509-521.

Arnott, S., Chandrasekaran, R. & Leslie, A.G.W. (1976). Structure of the single-stranded polyribonucleotide polycytidylic acid. J. Mol. Biol. 106, 735-738.

Arnott, S., Chandrasekaran, R., Birdsall, D.L., Leslie, A.G. W. & Ratliff, R.L. (1980). Left handed DNA helices. Nature 283, 743-745 and coordinates communicated to our laboratory by S. Arnott.

Arnott, S., Chandrasekaran, R., Hall, I.H. & Puigjaner, L.C. (1983). Heteronomous DNA. Nucleic Acids Res. 11, 4141-4155.

Bashford D. (1991). Electrostatic effects in biological molecules. Curr. Opinion in Struct. Biol. 1, 175-184.

Battacharya, D. & Bansal, M. (1988). A general procedure for generation of curved DNA molecules. J. Biomol. Struct. Dyn. 6, 93-104.

Beveridge, D.L., Swaminathan, S., Ravishankar, G., Withka, J.M., Srinivasan, J., Prévost, C., Louise-May, S., Langley, D.R., DiCapua, F.M. & Bolton, P.H. (1992). Molecular dynamics simulations on the hydration, structure and motions of DNA oligomers. In Topics in molecular and structural biology 17. Water and biological macromolecules (Westhof, E. Ed.) in press.

Boutonnet, N., Hui, X. & Zakrzewska, K. (1992). Looking into the grooves of DNA. Biopolymers in press.

Brahms, S., Fritsch, V., Brahms, J.G. & Westhof, E. (1992). Investigations on the dynamic structures of adenine- and thymine-containing DNA. J. Mol. Biol. 223, 455-476.

Briki, F., Ramstein, J., Lavery, R. & Genest, D. (1991). Evidence for the stochastic nature of base pair opening in DNA : a brownian dynamics simulation. J. Am. Chem. Soc. 113, 2490-2493.

Broitman, S.L., Im, D.D. & Fresco, J.R. (1987). Formation of the triple-stranded polynucleotide helix, poly(A.A.U). Proc. Natl. Acad. Sci. (USA) 84, 5120-5124.

Calladine, C.R. (1982). Mechanics of sequence-dependent stacking of bases in B-DNA. J. Mol. Biol. 161, 343-352.

Chou, S.H., Cheng, J.W., Fedoroff, O.Y., Chuprina, V.P. & Reid, B.R. (1992). Adjacent G:A mismatch base pairs contain $B_{II}$ phosphodiesters in solution. J. Am. Chem. Soc. 114, 3114-3115.

Cruse, W.B.T., Salisbury, S.A., Brown, T., Costick, R., Eckstein, F. & Kennard, O. (1986). Chiral phosphothionate analogues of B-DNA. The crystal structure of Rp(GpSCpGpSCpGpSC). J. Mol. Biol. 192, 891-905.

Dai, Z., Dauchez, M., Thomas, G. & Peticolas, W.L. (1992). Base sequence criteria and cartesian coordinates for stable B/Z and B/Z/B junctions in relaxed DNA. J. Biomol. Struct. Dynam. 9, 1155-1183.

Davis M.E. & McCammon J.A. (1990). Electrostatics in biomolecular structure and dynamics. Chem. Rev. 90, 509-521.

Dickerson, R.E. (1988). Usual and unusual DNA structures: A summing up. In "Unusual DNA structures" (Wells, R.D. and Harvey, S.C., Eds.) Springer-Verlag New York pp.287-306.

Dickerson, R.E. & Drew, H.R. (1981). Structure of a B-DNA Dodecamer: the influence of Base Sequence on helix Structure. J. Mol. Biol. 149, 761-786.

Dickerson, R.E., Goodsell, D.S., Kopta, M.L. & Pjura, P.E. (1987). The effect of crystal packing on oligonucleotide double helix structure. J. Biomol. Struct. Dyn. 5,557-579.

Dickerson, R.E., Bansal, M., Calladine, C.R., Diekmann, S., Hunter, W.N., Kennard, O., Lavery, R., Nelson, H.C.M., Olson, W.K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D.M., Tung, C.-S., von Kitzing, E., Wang, A.H.-J. & Zhurkin, V.B. (1989). Definitions and nomenclature of nucleic acid structure parameters. J. Mol. Biol. 205, 787-791.

Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, I. & Dickerson, R.E. (1981). Structure of a B-DNA dodecamer: Conformation and dynamics. Proc. Natl. Acad. Sci. (USA) 78, 2179-2183.

Englander, S.W. & Kallenbach, N.R. (1983). Hydrogen exchange and structural dynamics of proteins and nucleic acids. Quart. Rev. Biophys. 16, 521-655.

Fratini, A.V., Kopka, M.L., Drew, H.R. & Dickerson R.E. (1982). Reversible bending and helix geometry in a B-DNA dodecamer : CGCGAATT$^{Br}$CGCG. J. Biol. Chem. 257, 14686-14707.

Friedman R.A. & Honig B. (1992). The electrostatic contribution to DNA base-stacking interactions. Biopolymers 32, 145-159.

Fritsch, V. & Westhof, E. (1990). Minimisation and molecular dynamics of Z-DNA modified by acetylaminofluorene. In "Modeling of molecular structures and properties" (Rivail, J.L., Ed.) Elsevier pp. 627-634.

Fukui, T. & Ikehara, M. (1979). Polynucleotides XLVII. Synthesis and properties of poly(2-methylthio- and 2-ethylthioadenylic acid). Formation of non-Watson-Crick type complexes. Biochem. Biophys. Acta. 562, 527-533.

Gann, E.K. (1961). "Fate is the Hunter" Simon & Schuster New York pp.257.

Gilson, M.K. & Honig, B. (1988). Calculation of the total energy of a macromolecular system: Solvation energies, binding energies and conformational analysis. Proteins: Structure, function and Genetics 4, 7-18.

Gilson, M.K. & Honig, B. (1991). The inclusion of electrostatic hydration energies in molecular mechanics calculations. J. Computer-Aided Mol. Design 5, 5-20.

Gochin, M. & James, T. L. (1990). Solution structure studies of $d(AC)_4.d(GT)_4$ via restrained molecular dynamics simulations with NMR constraints derived from two-dimensional NOE and double quantum filtered COSY experiments. A.C.S. Biochemistry 29, 11172-11180.

Goodfellow, J.M. & Williams, M.A. (1992). Molecular dynamics. Curr. Opinion in Struct. Biol. 2, 211-216.

Gronenborn, A. M., Clore G. M. & Kimber, B. J. (1984). An investigation into the solution structure of two self-complementary DNA oligomers, 5'-d(CGTACG) and 5'-d(ACGCGCGT), by means of NOE mesurements. Biochem. J. 221, 723-736.

Grzeskowiak, K., Yanagi, K., Privé, G.G. & Dickerson R.E. (1991). The structure of B-helical CGATCGATCG and comparison with CCAACGTTGG. J. Biol. Chem. 266, 8861-8883.

Guéron, M., Kochoyan, M. & Leroy, J.L. (1987). A single mode of DNA base-pair opening drives imino proton exchange. Nature 328, 89-92.

Guschlbauer, W., Chantot, J.-F. & Thiele, D. (1990). Four-stranded nucleic acid structures 25 years later: From guanosine gels to teleomer DNA. J. Biomol. Struct. Dynam. 8, 491-511.

Hao, M.-H. & Olson, W.K. (1989). Molecular modeling and energy refinement of supercoiled DNA. J. Biomol. Struct. Dyn. 7, 661-692.

Hard, T., Fan, P., Magde, D. & Kearns, D.R. (1989). On the flexibility of DNA : time-resolved fluorescence polarization of intercalated quinacrine and 9-amino-6-chloro-2-methoxyacridine. J. Phys. Chem. 93, 4338-4345.

Hartmann, B., Leng, M. & Ramstein, J. (1986). Poly(dA-dT).poly(dA-dT) two-pathway proton exchange mechanism. Effect of general and specific base catalysis on deuteration rates. A.C.S. Biochemistry 25, 3073-3077.

Hartmann, B., Piazzola, D. & Lavery, R. (1992). $B_I$-$B_{II}$ transitions in B-DNA. Nucleic Acids Res. submitted.

Harvey, S.C. (1983). DNA structural dynamics: Longitudinal breathing as a possible mechanism for the B

Harvey, S.C. (1989). Treatment of electrostatic effects in macromolecular modeling. Proteins: Structure, function and Genetics 5, 78-92.

Harvey, S.C., Prabhakaran, M. & McCammon, J.C. (1985). Molecular dynamics simulation of phenylalanine transfer RNA. I. Methods and general results. Biopolymers 24, 1169-1188.

Hayes, D.M., Kollman, P.A. & Rothenberg, S. (1977). A conformational analysis of $H_3PO_4$, $H_2PO_4^-$, $HPO_4^{2-}$ and related model compounds. J. Am. Chem. Soc. 99, 2150-2154.

Heinemann, U. & Alings, C. (1989). Crystallographic study of one turn of G/C-rich B-DNA. J. Mol. Biol. 210, 369-381.

Heinemann, U. & Hahn, M. (1992). CCAGGC$m^5$CTGG Helical fine structure, hydration and comparison with CCAGGCCTGG. J. Biol. Chem. 267, 7332-7341.

Hélène, C. & Toulmé, J.J. (1990). Specific regulation of gene expression by antisense, sense and antigene nucleic acids. Biochem. Biophys. Acta 1049, 99-125.

Hingerty, B., Richie, R.H., Ferrel, T.L. & Turner, J.E. (1985). Dielectric effects in biopolymers : the theory of ionic saturation revisited. Biopolymers 24, 427-439.

Hogan, M., LeGrange, J. & Austin, R.H. (1983). Dependence of DNA helix flexibility on base composition. Nature 304, 752-754.

Jain, S. & Sundaralingam, M. (1989). Effect of crystal packing environment of conformation of the DNA duplex. J. Biol. Chem. 264, 12780-12784.

Jayaram, B., Swaminathan, S., Beveridge, D.L., Sharp, K. & Honig, B. (1990). Monte Carlo simulation studies on the structure of the counterion atmosphere of B-DNA. Variations on the primitive dielectric model. Macromolecules 23, 3156-3165.

Juffer, A.H., Botta, E.F.F., van Keulen, B.A.M., van der Ploeg, A. & Berendsen, H.J.C. (1991). The electric potential of a macromolecule in a solvent: A fundamental approach. J. Comp. Phys. 97, 144-170.

Kaluarachchi, K., Meadows, R. P. & Gorenstein, D. G. (1991). How accurately can oligonucleotide structures be determined from the hybrid relaxation rate matrix: NOESY distance restrained molecular dynamics approach. A.C.S. Biochemistry 30, 8785-8797.

Keepers, J., Kollman, P.A. & James, T.L. (1984). Molecular mechanical studies of base-pair opening in d(CGCGC).d(GCGCG), $dG_5.dC_5$, d(TATAT).d(ATATA) and $dA_5.dT_5$ in the B and Z forms of DNA. Biopolymers 23, 2499-2511.

Klysik, J., Rippe, K. & Jovin, T.M. (1991). Parallel-stranded DNA under topological stress : rearrangement of $(dA)_{15}.(dT)_{15}$ to a $d(A.A.T)_n$ triplex. Nucleic Acids Res. 19, 7145-7154.

Kollman, P.A., Keepers, J.W. & Weiner P.K. (1982). Molecular-mechanics studies on $d(CGCGAATTCGCG)_2$ and $dA_{12}.dT_{12}$ : an illustration of the coupling between sugar repuckering and DNA twisting. Biopolymers 21, 2345-2376.

Lavery, R. (1988a). DNA flexibility under control : the Jumna algorithm and its application to BZ junctions. In "Unusual DNA structures" (Wells, R.D. and Harvey, S.C., Eds.), Springer-Verlag pp.189-206.

Lavery, R. (1988b). Junctions and bends in nucleic acids : a new theoretical modeling approach. In "Structure and expression. Vol.3, DNA bending and curvature" (Olson, W.K., Sarma, R.H., Sarma, M.H. and Sundaralingam, M., Eds). Adenine Press New York pp.191-211.

Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. J. Biomol. Struct. Dyn. 6, 63-91.

Lavery, R. & Sklenar, H. (1989). Defining the structure of irregular nucleic acids : conventions and principles. J. Biomol. Struct. Dyn. 6, 655-667.

Lavery, R. & Sklenar, H. (1990). A quantitative description of the conformation of biological macromolecules. In "Structure and methods. Vol.2, DNA Protein complexes and proteins" (Sarma, R.H. and Sarma, M.H., Eds.) Adenine Press pp.215-235.

Lavery, R., Zakrzewska, K. & Pullman, A. (1984). Optimized monopole expansions for the representation of the electrostatic properties of the nucleic acids. J. Comp. Chem. 5, 363-373.

Lavery, R., Sklenar, H., Zakrzewska, K. & Pullman, B. (1986a). The flexibility of the nucleic acids: (II) the calculation of internal energy and applications to mononucleotide repeat DNA. J. Biomol. Struct. Dyn. 3, 989-1014.

Lavery, R., Sklenar, H. & Pullman, B. (1986b). The flexibility of the nucleic acids : (III) the interaction of an aliphatic diamine, putrescine, with flexible B-DNA. J. Biomol. Struct. Dyn. 3, 1015-1031.

Lavery, R., Parker, I. & Kendrick, J. (1986c). A general approach to the optimization of the conformation of ring molecules with an application to valinomycin. J. Biomol. Struct. Dyn. 4, 443-461.

Lavery, R., Zakrzewska, K., Sun, J.S. & Harvey, S.C. (1992). A comprehensive classification of nucleic acid structural families based on strand direction and base pairing. Nucleic Acids Res. 20, 5011-5016.

Leroy, J.L., Broseta, D. & Guéron, M. (1985). Proton exchange and base-pair kinetics of poly(rA).poly(rU) and poly(rI).poly(rC). J. Mol. Biol. 184, 165-178.

Leroy, J.L., Charretier, E. , Kochoyan, M. & Guéron, M. (1988). Evidence from base-pair kinetics for two types of adenine tract structures in solution : their relation to DNA curvature. A.C.S. Biochemistry 27, 8894-8898.

Manning, G.S. (1983). Breathing and bending fluctuations in DNA modelled by an open-base-pair kink coupled to axial compression. Biopolymers 22, 689-729.

Mazur, J. & Jernigan, R.L. (1991). Distance-dependent dielectric constants and their application to double helical DNA. Biopolymers 31, 1615-1629.

Metzler, W. J., Wang, C., Kitchen, D. B., Levy, R. M. & Pardi, A. (1990). Determining local conformational variations in DNA nuclear magnetic resonance structures of DNA duplexes d5CGCCTAATCG) and d(CGTCACGCGC) generated using back-calculation of the NOE spectra, a distance geometry algorithm and constrained molecular dynamics. J. Mol. Biol. 214, 711-736.

McLachlan, A.D. (1979). Gene duplications in the structural evolution of chymotrypsin. J. Mol. Biol. 128, 49-79.

Michel, F. & Westhof, E. (1990). Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J. Mol. Biol. 216, 585-610.

Millar, D.P., Robbins, R.J. & Zewail, A.H.(1982) . Torsion and bending of nucleic acids studied by sub-nanosecond time-resolved fluorescence depolarization of intercalated dyes. J. Chem. Phys. 76, 2080-2094.

Morvan, F., Rayner, B., Imbach, J.L., Chang, D.K. & Lown, J.W. (1987). α-DNA-III. Characterization by high field $^{1}$H-NMR, anti-parallel self-recognition and conformation of the unnatural hexadeoxyribonucleotides α-[d(CpApTpGpCpG)] and α-[d(CpGpCpApTpG)]. α-Oligodeoxynucleotides as potential cellular probes for gene control. Nucleic Acids Res. 15, 4241-4255.

Moult, J. (1992). Electrostatics. Curr. Opinion in Struct. Biol. 2, 223-229.

Nikonowicz, E.P., & Gorenstein, D.G. (1990). Two-dimensional $^{1}$H and $^{31}$P NMR spectra and restrained molecular dynamics structure of a mismatched GA decamer oligodeoxyribonucleotide duplex. A.C.S. Biochemistry 29, 8845-8858.

Olson, W.K. & Sussman, J.L. (1982). How flexible is the furanose ring? 1. A comparison of experimental and theoretical studies. J. Am. Chem. Soc. 104, 270-278.

Pattabiraman, N. (1986). Can the double helix be parallel? Biopolymers 25, 1603-1606.

Pilch, D.S., Levensen, C. & Shafer, R.H. (1991). Structure, stability, and thermodynamics of a short intermolecular purine-purine-pyrimidine triple helix. A.C.S. Biochemistry 30, 6081-6087.

Pleij, C.W.A. (1990). Pseudoknots: A new motif in the RNA game. Trends in Biochem. Sciences 15, 143-147.

Poncin, M., Hartmann, B. & Lavery, R. (1992). Conformational sub-states in B-DNA. J. Mol. Biol. 226, 775-794.

Poncin, M., Piazzola, D. & Lavery, R. (1992) DNA flexibility as a function of allomorphic conformation and of base sequence. Biopolymers 32, 1077-1103.

Prévost, C., Louise-May, S., Ravishankar, G., Lavery, R. & Beveridge, D.L. (1992). Biopolymers in press.

Privé, G.G., Heinemann, U., Chandrasekharan, S. Kan, L.S., Kopta, M.L. & Dickerson, R.E. (1987). Helix geometry, hydration, and GA mismatch in a B-DNA decamer. Science 238, 498-504.

Privé G.G., Yanagi, K. & Dickerson, R.E. (1991). Structure of the B-DNA decamer CCAACGTTGG and comparison with isomorphous decamers CCAAGATTGG and CCAGGCCTGG. J. Mol. Biol. 217, 177-199.

Pullman, B. (1990). Molecular mechanisms of specificity in DNA-antitumor drug interactions. In "Theoretical biochemistry and molecular biophysics" (Beveridge, D. L. and Lavery, R., Eds.) pp.193-205.

Ramsing, N.B. & Jovin, T.M. (1988). Parallel stranded duplex DNA. Nucleic Acids Res. 16, 6659-6676.

Ramstein, J. & Lavery, R. (1988). Energetic coupling between DNA bending and base pair opening. Proc. Natl. Acad. Sci. (USA) 85, 7231-7235.

Ramstein, J. & Lavery, R. (1990). Base pair opening pathways in B-DNA. J. Biomol. Struct. Dyn. 7, 915-933.

Rao, S.N. & Kollman, P. (1985). On the role of uniform and mixed sugar puckers in DNA double helical structures. J. Am. Chem. Soc. 107, 1611-1617.

Ravishankar, G., Swaminathan, S., Beveridge, D.L., Lavery, R. & Sklenar, H. (1989). Conformational and helicoidal analysis of 30ps of molecular dynamics on the d(CGCGAATTCGCG) double helix : "curves", dials and windows. J. Biomol. Struct. Dyn. 6, 669-699.

Rich, A., Davies, D.R., Crick, F.H.C. & Watson, J.D. (1961). The molecular structure of polyadenylic acid. J. Mol. Biol. 3, 71-86.

Roongta, V.A., Jones, C.R. & Gorenstein, D.G. (1990). Effects of distortions in the deoxyribose phosphate backbone conformation of duplex oligodeoxyribonucleotide decamers containing GT, GA, GG, AC and GU base pair mismatch on $^{31}P$ NMR spectra. A.C.S. Biochemistry 29, 5245-5258.

Rose, I.A., Hanson, K.R., Wilkinson, K.D. & Wimmer, M.J. (1980). A suggestion for naming faces of ring compounds. Proc. Natl. Acad. Sci. (USA) 77, 2439-2441.

Rosenberg, J.M., Seeman, N.C., Day, R.O. & Rich, A. (1976). RNA double helices generated from crystal structures of double helical dinucleoside phosphates. Biochem. Biophys. Res. Com. 69, 979-987.

Saenger, W. (1984). "Principles of nucleic acid structure". Springer-Verlag, New-York.

Sarai, A., Mazur, J., Nussinov, R. & Jernigan, R.L. (1989). Sequence dependence of DNA conformational flexibility. A.C.S. Biochemistry 28, 7842-7849.

Schlick, T., Peskin, C., Broyde, S. & Overton, M. (1987). An analysis of the structural and energetic properties of deoxyribose by potential energy methods. J. Comp. Chem. 8, 1199-1224.

Schlick, T. & Olson, W.K. (1992). Supercoiled DNA energetics and dynamics by computer simulation. J. Mol. Biol. 223, 1089-1119.

Schmitz, U., Zon, G. & James, T.L. (1990). Deoxyribose conformation in d(GTATATAC) : evaluation of sugar pucker by simulation of double quantum filtered COSY cross-peaks. A.C.S. Biochemistry 29, 2357-2368.

Schmitz, U., Pearlman, D.A. & James T.L. (1991). Solution structure of d(GTATATAC) via restrained molecular dynamics simulation with nuclear magnetic resonance constraints derived from relaxation matrix analysis of two-dimensional NOE experiments. J. Mol. Biol. 221, 271-292.

Searle, M.S. & Wakelin, L.P.G. (1990). Sequence specific conformation of a DNA decamer containing an adenine tract studied in solution by $^{1}$H-NMR spectroscopy. Biophys. Biochem. Acta 1049, 69-77.

Shakked, Z. & Rabinovich, D. (1986). The effect of the base sequence on the fine structure of the DNA double helix. Prog. Biophys. Molec. Biol. 47, 159-195.

Sharp, K. (1991). Incorporating solvent and ion screening into molecular dynamics using the finite-difference Poisson-Boltzmann method. J. Comp. Chem. 12, 454-468.

Sklenar, H., Lavery, R. & Pullman, B. (1986). The flexibility of the nucleic acids: (I) "SIR", a novel approach to the variation of polymer geometry in constrained systems. J. Biomol. Struct. Dyn. 3, 967-987.

Sklenar, H., Eisenhaber, F., Poncin, M. & Lavery, R. (1990). Including solvent and counterion effects into the force fields of macromolecular mechanics : the field integrated electrostatic approach (Fiesta). In "Theoretical biochemistry and molecular biophysics. A comprehensive survey" (Beveridge, D.L. and Lavery, R., Eds.) Adenine Press, New York, pp.317-335.

Sklenar, V., & Bax, A. (1987). Measurements of $^{1}$H-$^{31}$P NMR coupling constants in double-stranded DNA fragments. J. Am. Chem. Soc. 109, 7525-7526.

Soumpasis, D.M. & Tung, C.-S. (1988). A rigourous basepair oriented description of DNA structures. J. Biomol. Struct. Dyn. 6, 397-420.

Sprous, D. & Harvey, S.C. (1992). A three dimensional model for nodule DNA. J. Biol. Chem. 267, 5502.

Srinivasan, A.R. & Olson, W.K. (1987). Nucleic acid model building : the multiple backbone solutions associated with a given base morphology. J. Biomol. Struct. Dyn. 4, 895-938.

Srinivasan, J., Withka, J.M. & Beveridge, D.L. (1990). Molecular dynamics of an *in vacuo* model of duplex d(CGCGAATTCGCG) in the B-form based on the amber 3.0 force field. Biophys J. 58, 523-547.

Stofer, E. & Lavery, R. (1992). Measuring the geometry of DNA grooves. in preparation.

Strobel, S.A. & Dervan, P.B. (1990). Site-specific cleavage of a yeast chromosome by oligonucleotide-directed triple-helix formation. Science 249, 73-75.

Subramanian, P.S., Ravishankar, G. & Beveridge, D.L. (1988). Theoretical considerations on the "spine of hydration" in the minor groove of d(CGCGAATTCGCG).d(CGCGAATTCGCG): Monte Carlo computer simulation. Proc. Natl. Acad. Sci. (USA) 85, 1836-1840.

Sun, J.S., François, J-C., Lavery, R., Saison-Behmoaras, T., Montenay-Garestier, T., Thuong, N.T. & Hélène, C. (1988). Sequence-targeted cleavage of nucleic acids by oligo-a-thymidylate-phenanthroline conjugates : parallel and antiparallel double helices are formed with DNA and RNA, respectively. A.C.S. Biochemistry 27, 6039-6045.

Sun, J.S., Mergny, J.-L., Lavery, R., Montenay-Garestier, T. & Hélène, C. (1991). Triple helix structures : sequence dependence, flexibility and mismatch effects. J. Biomol. Struct. Dyn. 9, 411-424.

Swaminathan, S., Ravishankar, G., Beveridge, D.L., Lavery, R., Etchebest, C. & Sklenar, H. (1990). Conformational and helicoidal analysis of the molecular dynamics of proteins : curves, dials and windows for a 50ps dynamic trajectory of BPTI. Proteins : Structure, Function and Genetics 8, 179-193.

Swaminathan, S., Ravishanker, G. & Beveridge, D.L. (1991). Molecular dynamics of B-DNA including water and counterions: A 140ps trajectory for d(CGCGAATTCGCG) based on the Gromos force field. J. Am. Chem. Soc. 113, 5027-5040.

Tan, R.K.Z. & Harvey, S.C. (1989). Molecular mechanics model of supercoiled DNA. J. Mol. Biol. 205, 573-591.

Tan, R.K.Z. & Harvey, S.C. (1990). Succinct macromolecular models: Application to supercoiled DNA. In "Theoretical biochemistry and molecular biophysics" (Beveridge, D. L. and Lavery, R., Eds.) pp.125-137.

Tilton, R F., Weiner, P.K. & Kollman P.A. (1983). An analysis of sequence dependence of the structure and energy of A- and B-DNA models using molecular mechanics. Biopolymers 22, 969-1002.

Trifonov E.N. (1991) DNA in profile. Trends in Biochem. Sciences 16, 467-470.

Tumanyan, V.G. & Esipova, N.G. (1975). Investigation of fibrous structures II. General method for the elucidation of the conformaitons of the complementary nucleic acids. Biopolymers 14, 2231-2246.

Uhlenbeck, O.C. (1991). Catalytic RNAs. Curr. Opinion in Struct. Biol. 1, 459-463.

von Kitzing, E. & Diekmann, S. (1987). Molecular mechanics calculations of $dA_{12}.dT_{12}$ and of a curved molecule $d(GCTCGAAAA)_4.d(TTTTTCGAGC)_4$. Eur. J. Biophys. 15, 13-26.

von Kitzing, E., Lilley, D.M. & Diekmann, S. (1990). The stereochemistry of a four-way DNA junction: A theoretical study. Nucleic Acids Res. 18, 2671-2683.

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

Watson, J.C. & Crick, F.H.C. (1953) A structure for deoxyribose nucleic acid. Nature 171, 737-738.

Wells, R.D. & Harvey, S.C. (Eds.) (1988). "Unusual DNA structures". Springer-Verlag, New York.

Westhof, E. (1992). Westhof's rule. Scientific correspondence. Nature 358, 459-460.

Westhof, E. & Beveridge, D.L. (1990). Hydration of nucleic acids. In Water science reviews 5 (Franks, F. Ed.) Cambridge University press pp.24-136.

Westhof, E., Romby, P., Ehresmann, C. & Ehresmann, B. (1990). Computer-aided structural biochemistry of ribonucleic acids In "Theoretical biochemistry and molecular biophysics" (Beveridge, D.L. and Lavery, R., Eds.) Adenine press New York, pp.399-409.

Withka, J.M., Swaminathan, S. & Bolton, P.H. (1990). NOE's in duplex DNA depend on orientations of internuclear vectors to the symmetry axis. J. Magn. Res. 89, 386-390.

Withka, J.M., Swaminathan, S., Beveridge, D.L. & Bolton, P.H. (1991). Time dependence of nuclear overhauser effects of duplex DNA from molecular dynamics trajectories. J. Am. Chem. Soc. 113, 5041-5049.

Withka, J.M., Swaminathan, S., Beveridge, D.L. & Bolton, P.H. (1992). Toward a dynamical structure of DNA : comparison of theoretical and experimental NOE intensities. Science 255, 597-599.

Yanagi, K., Privé, G.G., & Dickerson, R.E. (1991). Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. J. Mol. Biol. 217, 201-214.

Zakrzewska, K. (1992). Static and dynamic conformational properties of AT sequences in B-DNA. J. Biomol. Struct. Dyn. 9, 681-690.

Zakrzewska, K. & Lavery, R. (1989). Theoretical studies of groove binding drugs with DNA. In "Computer-aided molecular design" (Richards, W.G., Ed.) IBC Technical Services Ltd., U.K., pp.129-145.

Zauhar, R.J. (1991). The incorporation of hydration forces determined by continuum electrostatics into molecular mechanics simulations. J. Comp. Chem. 12, 575-583.

Zhurkin, V.B., Lysov, Yu.P. & Ivanov, V.I. (1978). Different families of double-stranded conformations of DNA as revealed by computer calculations. Biopolymers 17, 377-412.

Zhurkin, V.B., Poltev, V.I. & Florent'ev, V.L. (1980). Atom-atom potential functions for conformational calculations of nucleic acids. Mol. Biol. 14, 1116-1130.

Zhurkin, V.B., Lysov, Yu.P., Florentiev, V.L. & Ivanov, V.I. (1982). Torsional flexibility of B-DNA as revealed by conformational analysis. Nucleic. Acids Res. 10, 1811-1830.

Zhurkin, V.B., Gorin, A.A., Charakhchyan, A.A. & Ulyanov, N.B. (1990). Sequence-dependent variability of A- and B-DNA. Bending, torsional and stretching flexibility, interconversion of sugar rings and bulged-out structures. In "Theoretical biochemistry and molecular biophysics" (Beveridge, D. L. and Lavery, R., Eds.) pp.411-431.

Zhurkin, V.B., Ulyanov, N.B., Gorin, A.A. & Jernigan, R.L. (1991). Static and statistical bending of DNA evaluated by Monte Carlo simulations. Proc. Natl. Acad. Sci. (USA) 88, 7046-7050.