ELSEVIER

Biophysical Chemistry 50 (1994) 323–343

Biophysical Chemistry

A database of 32 DNA triplets to study triple helices by molecular mechanics and dynamics

J.M. Piriou, Ch. Ketterlé, J. Gabarro-Arpa, J.A.H. Cognet, M. Le Bret *

Laboratoire de Physicochimie et Pharmacologie des macromolécules biologiques (INSERM U140, CNRS URA 147), 94805 Villejuif Cédex, France

(Received 9 August 1993; accepted 23 November 1993)

Abstract

We present here a database of 32 deoxyribonucleotide triplets, that can be used as building blocks of triple helix forming deoxyribonucleotides on a computer. This database is made of all the pairing schemes of the triplets ATT, GCC^+, ATA and GCG where the third base forms two hydrogen bonds with the purine of the first two Watson–Crick strands. The essential features of the known triple helices were preserved in the resulting structures. A triple helix can be easily built from any combination of these basic triplets. Four homogeneous and alternate triple helices thus obtained were studied by molecular mechanics and dynamics in vacuo. The results are in agreement with known experimental observations for ATT and suggest a possible structure for the GCG triple helix. In order to characterize the geometry of the structures obtained, the definitions of nucleic acid structure parameters (R.E. Dickerson et al., EMBO J. 8 (1989) 1–4) have been extended to triple helical polynucleotides.

Key words: Triple helix; Oligonucleotides; Triplet database; Molecular modeling; Molecular mechanics; Molecular dynamics

1. Introduction

Since the first report on triple helices [1], a considerable body of work has been accumulated and reviewed [2–5]. These last years, triple helices have regained considerable interest [6] mostly as the consequence of the massive development of molecular biology. Intramolecular triplexes may have a role in gene expression and recombination [4]. Intermolecular DNA triplexes find promising potential applications in chromosome mapping [7–10], human DNA cleavage [11] and gene targeting [12]. Specially designed oligonucleotides can inhibit the expression of genes that probably play a key role in cancer [13] or in viral diseases [14]. In the first triplexes to have been discovered, a pyrimidine strand is bound in the major groove and to the purine strand of the target double helix through two hydrogen bonds as described by Hoogsteen [15,16]. The structures of several triple helices have been determined by X-ray fiber diffraction [17,18] and segments of triple helices were found

* Corresponding author.

Abbreviations: rms, root mean square; MD, molecular dynamics.

SSDI 0301-4622(93)E0103-C

in tRNA X-ray crystal structures [19,20]. Beside this information, many qualitative structural facts have been established for triple helices: protonation of cytosine [21] in the third strand, existence of poly G.poly C.poly G triplexes [22], polarity of the third strand [5,23] regarding different anomers [24], or different sequences [25,26], determination of the puckers and of the forms of the three strands [27–29]. Molecular modeling [30] and molecular dynamics have been used to give a consistent picture of the structure of homogeneous or alternate triple helices [31,32]. The formation of DNA, RNA or DNA/RNA triplexes that seemed to be limited to homopurine tracts [7,33] has recently been extended to other cases: (i) homogeneous triple helices can accommodate some local replacements [21,34–39], (ii) a double helical DNA made of a purine tract followed by a pyrimidine tract can be recognized by a third strand [40–43], (iii) RecA protein-mediated pairing between single stranded DNA and duplex targets [44] seems to considerably stretch and untwist the DNA [45] and requires homology between the single stranded DNA and its target but not a specific sequence of the target.

The possibilities to construct these third strands to recognize a given double stranded target DNA are numerous and a structural or theoretical database is needed to help to discard the lame ones. In this paper, we have systematically constructed all triple helices where the first two strands are in a canonical Watson–Crick conformation. The third strand may be in eight different conformations: its sugar can be the α or β anomer, and lies anti or syn to the base. Finally, the third strand can be parallel or antiparallel to the first strand. The set of basic conformations could be extended by the inclusion of various sugar puckers. The C2' endo conformations have been considered in most calculations as based on recent experimental data [29,46].

The first objective of this work is to search for the proper combinations of anomers and of conformations that yield to geometrically uniform homogeneous triple helices. This search is meaningful if a complete set of combinations is explored, since a given helix may exist physically if basic geometrical criteria to form a helix are satisfied and if no other helix fulfill them better. This is why we have explored such a large number of conformations. The conformations that we found are in good agreement with known experimental features and were found to be stable when MD simulations in vacuo were performed.

2. Methods

2.1. Graphics

The molecular structures have been displayed on a Silicon Graphics IRIS 4D/310GT using the programs, MORCAD [47] and OCL [48].

2.2. Nomenclature

The double stranded DNA bases are assumed to be in the canonical Watson–Crick pairing. The purine of a given Watson–Crick base pair can be

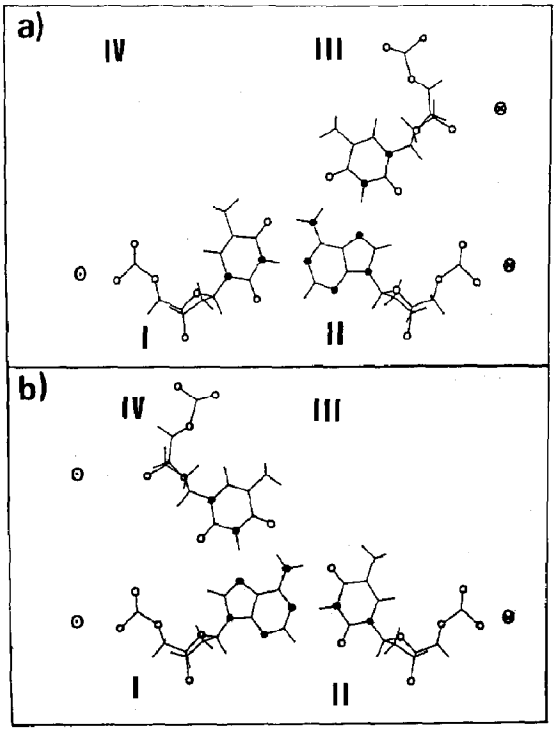


Fig. 1. Representation of the two homologous triplets (a) AToTβpt and (b) TAToβpt. The O5' to O3' of the strand in the first position points towards the reader.

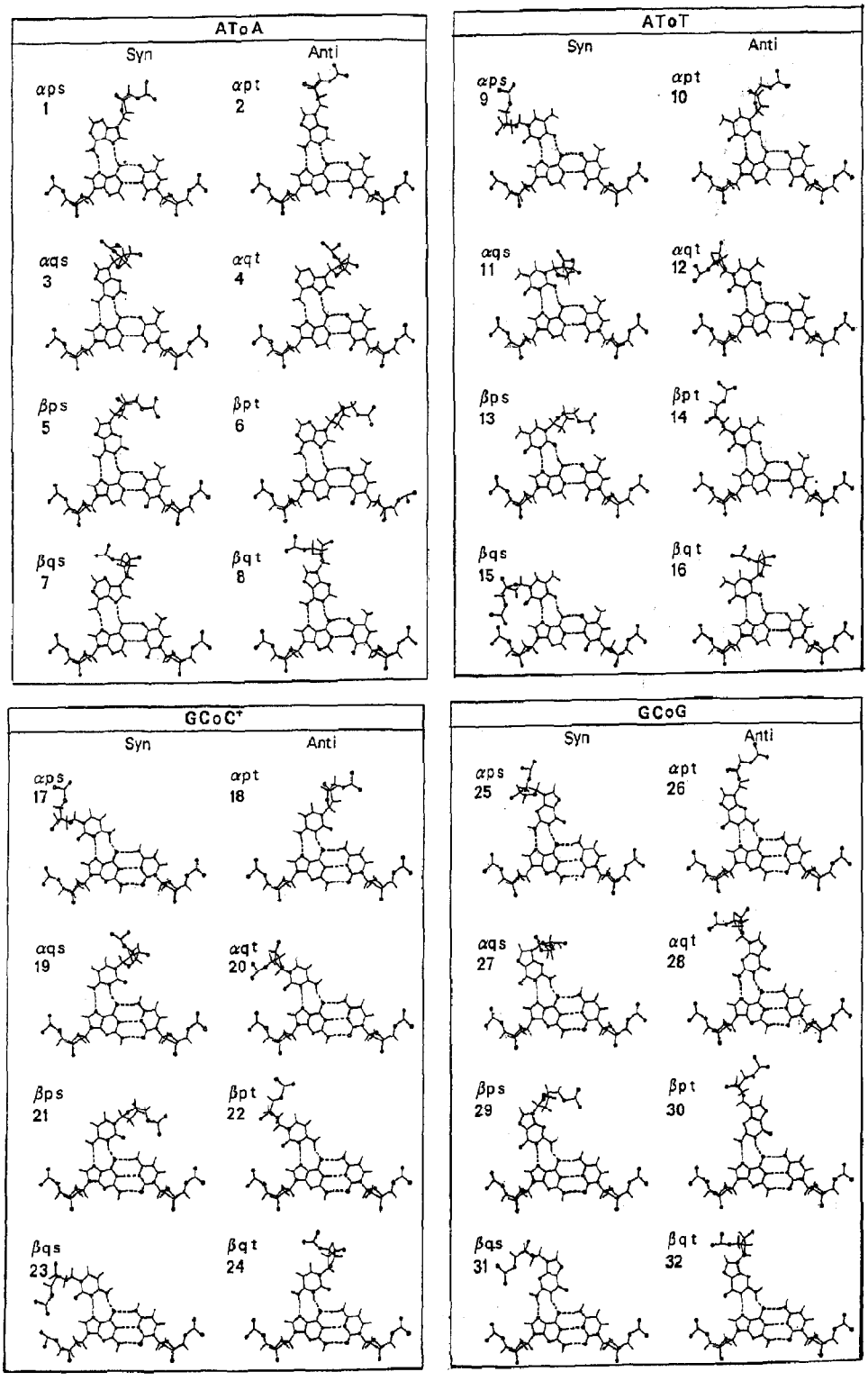


Fig. 2. Definition of the name conventions and representation of the 32 main triplets of the database. Nitrogen and oxygen atoms are represented respectively as black and white circles. The O5′ to O3′ direction of the strand in the first position (I) points towards the reader.

either in the first (Fig. 1b) or in the second strand (Fig. 1a) of the sequence. Correspondingly, we adopt the convention that the third strand occupies the positions referred as IV and III respectively, as indicated in the scheme of Fig. 1. Thus, the triplet in Fig. 1a was designated TATo, and the triplet in Fig. 1b AToT, where the symbol "o" designates the unoccupied position.

The various conformations of the third strand that were studied here are denoted with a three character word identifier such as αps. The anomers of the deoxyribose of the third strand may be alpha (α) or beta (β). The third strand may be parallel (p) or antiparallel (q) to the first strand. The glycosidic torsion of the third base can have either the *anti* (t) or the *syn* (s) orientation [49]. Fig. 2 summarizes the conventions schematically.

The base in the third strand can be paired to the purine in two different ways: αps, αqt, βpt and βqs triplets have the Hoogsteen bond pairing [15,16], while αpt, αqs, βps and βqt triplets have a reverse Hoogsteen pattern [50] (see Fig. 2). The hydrogen bond pairing geometry in the third strand can be deduced from the following two rules: (i) any triplet whose identifier differs by an even number of changes from βpt has the Hoogsteen pattern, (ii) otherwise, it has the reverse Hoogsteen pattern.

To clarify the definitions, three examples will be taken. (i) the Arnott [17,18] classical triple helix is denoted by AToTβpt: the deoxyribose of

Table 1
Values of the helical parameters of various oligonucleotides either compiled from the literature, or used in this database or obtained after energy refinement (E. refin. columns). Lengths and angles are respectively in angstrom and deg. Helical repeats are in number of base pairs/turn or in number of triplets/turn. The parameters are global or local according to the g or l character shown in the second column. By definition global shift, slide, tilt and roll are nil for a uniform structure. The 3rd column "Sign", indicates whether either choice of EMBO reference frames [52] leaves the parameter unchanged (+) or reverses its sign (−). The values of the helical parameters are shown here with their mean values and their standard deviations for six triplexes [18], and for two duplexes [55]. The standard deviations are computed from the six or two available values and have no thermodynamical meaning in contrast to the fluctuations shown in Table 8

		Sign	Six triplexes of ref. [18]	AToTβpt		Six triplexes of ref. [18]	AToTβpt		A-DNA ref. [55]	B-DNA ref. [55]
		1–2, 3	1–2	database 1–2	E. refin. 1–2	3	database 3	E. refin. 3	1–2	1–2
X-displacement (Å)	g	+	−3.0 ± 0.4	−2.8	−2.9	−3.2 ± 0.1	−2.6	−2.6	−4.5	0.2
Y-displacement (Å)	g	−	−0.1 ± 0.1	0.3	0.1	−1.6 ± 0.1	−1.8	−1.6	0.1	0.1
inclination (deg)	g	+	7.9 ± 1.7	2.9	1.7	14.3 ± 2.4	0.0	1.1	20.5	−4.7
Tip (deg)	g	−	−3.6 ± 1.4	−1.3	1.3	−0.3 ± 1.0	0.0	−2.1	0.4	−0.1
shift (Å)	g	−	0	0	0	0	0		0	0
slide (Å)	g	+	0	0	0	0	0		0	0
rise (Å)	g	+	3.1 ± 0.1	3.3	3.1				2.6	3.4
tilt (deg)	g	−	0	0	0	0	0		0	0
roll (deg)	g	+	0	0	0	0	0		0	0
helical repeat	g	+	11.5 ± 0.5	12.0	11.8				11.0	10.0
helical twist (deg)	g	+	31.4 ± 1.3	30.0	30.5				32.7	36.0
tilt (deg)	l	−	2.0 ± 0.5	0.7	−0.5	0.1 ± 0.5	0.0	2.3	−0.2	0.1
roll (deg)	l	+	4.5 ± 1.0	1.6	1.8	8.0 ± 1.6	0.0	0.3	11.8	−3.0
shear (Å)	l	−	0.5 ± 0.2	0.4	−0.2	−0.1 ± 0.1	−0.1	0.1	0.0	0.0
stretch (Å)	l	+	0.4 + 1.1	0.0	0.1	2.8 ± 0.0	2.8	2.9	0.0	0.0
stagger (Å)	l	+	−0.2 ± 0.1	−0.3	−0.2	−0.1 ± 0.0	−0.5	−0.2	0.1	0.0
buckle (deg)	l	−	1.6 ± 0.8	−0.1	−0.5	1.4 ± 0.2	−8.3	−3.3	0.0	0.0
propeller twist (deg)	l	+	−13.8 ± 3.3	−12.1	−9.3	−9.4 ± 1.5	−3.8	−3.4	11.2	−1.2
opening (deg)	l	+	4.8 ± 0.8	1.0	−0.2	−3.9 ± 3.8	−3.6	0.0	0.3	0.8

the third strand has the β anomery; its third strand is parallel to the first purine strand; the third base is in the anti conformation. (ii) For the TAoTβpt, TAoTβqt, GCoGβpt and GCoGβqt triplets our convention is identical to the definition given by Durland et al. [51] for β triplets. (iii) Triplets AToAαqt and TAAoαpt have different names but have identical structures.

2.3. Helical parameters

The helical parameters were defined and computed with the object command language, OCL [48] and MORCAD [47] programs. Through these programs, the user can at will define sets of atoms, attach to them user defined reference frames O*xyz* and displace them around through user defined transformations such as rotations and translations.

In this work, the reference axes for the target double stranded DNA have been chosen according to the definitions adopted at the EMBO Workshop [52–54]. Table 1 shows the helical parameters of strands 1 and 2 that can be found from the X-ray fiber diffraction data in triplexes [18] and duplexes [55]. Three groups of helical parameters apply to strands 1 and 2 and respectively describe the position of the base pair with respect to the global helical axis O*z* (*X* and *Y* displacement, inclination and tip parameters in Table 1), the position of two consecutive base pairs relative to each other (shift through helical twist in Table 1), and the position of the two bases within a base pair (shear through opening

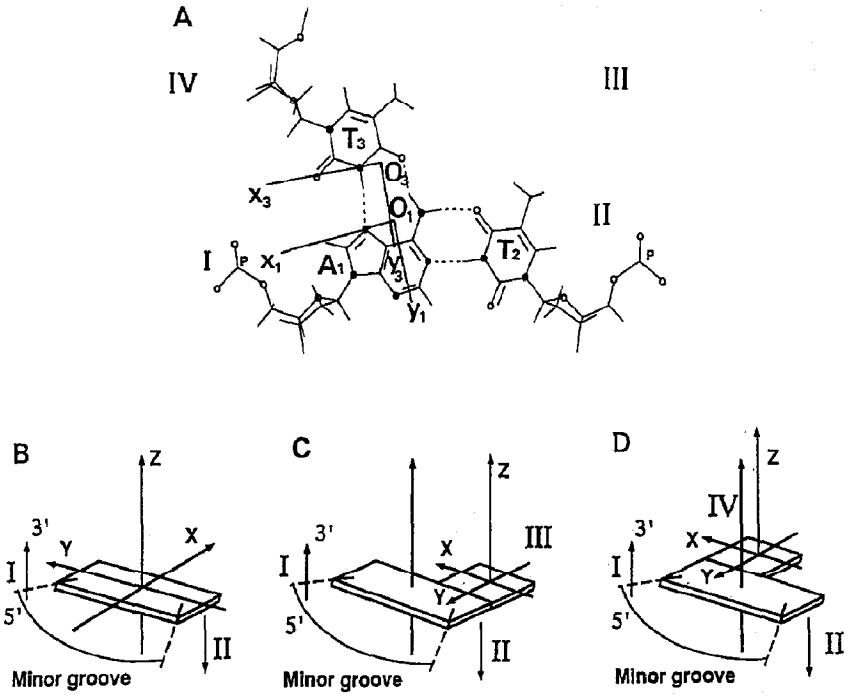


Fig. 3. (A) Construction of the orthonormal reference frames attached to bases 1 and 3 of the AToT triplet to compute helical parameters. The O*x* axis of these reference frames, is defined as the axis that passes through the two non hydrogen atoms of one base involved in hydrogen bonding with the other base. In the example shown here, the origin O3 of the reference is the middle of the line joining the atom O4 to the atom N3 of the Thymine T3. The origin O1 is the middle of the line joining the atom N6 to the atom N7 of adenine A1. Nitrogen and oxygen atoms are represented as black and white circles respectively and the phosphorous atoms can be identified through the character "P". (D) Orientation of a reference frame for the double helix according to EMBO conventions [52]. (C) and (D): Orientation of a reference frame for the triple helix according to our proposal to extend EMBO conventions.

in Table 1). For the first two groups of parameters, the Oy axis or long axis of a base pair is defined as the line passing through the C6 and C8 atoms. For the last group, Oy is defined as described in the literature [52–54]. All parameters were computed as explained in ref. [54].

It would be artificial to define a specialized long axis and a reference frame for every type of pairings between strands 1 and 3. It is sufficient to characterize the geometry and position of the third base with respect to the first base to provide a complete geometrical description. Thus we did not compute helical parameters for the base pair 1–3 as for the Watson–Crick strands, but solely for base 3. It required the construction through OCL of two new local reference frames attached to the triplet first and third bases that are bound through two hydrogen bonds as shown in Fig. 3.

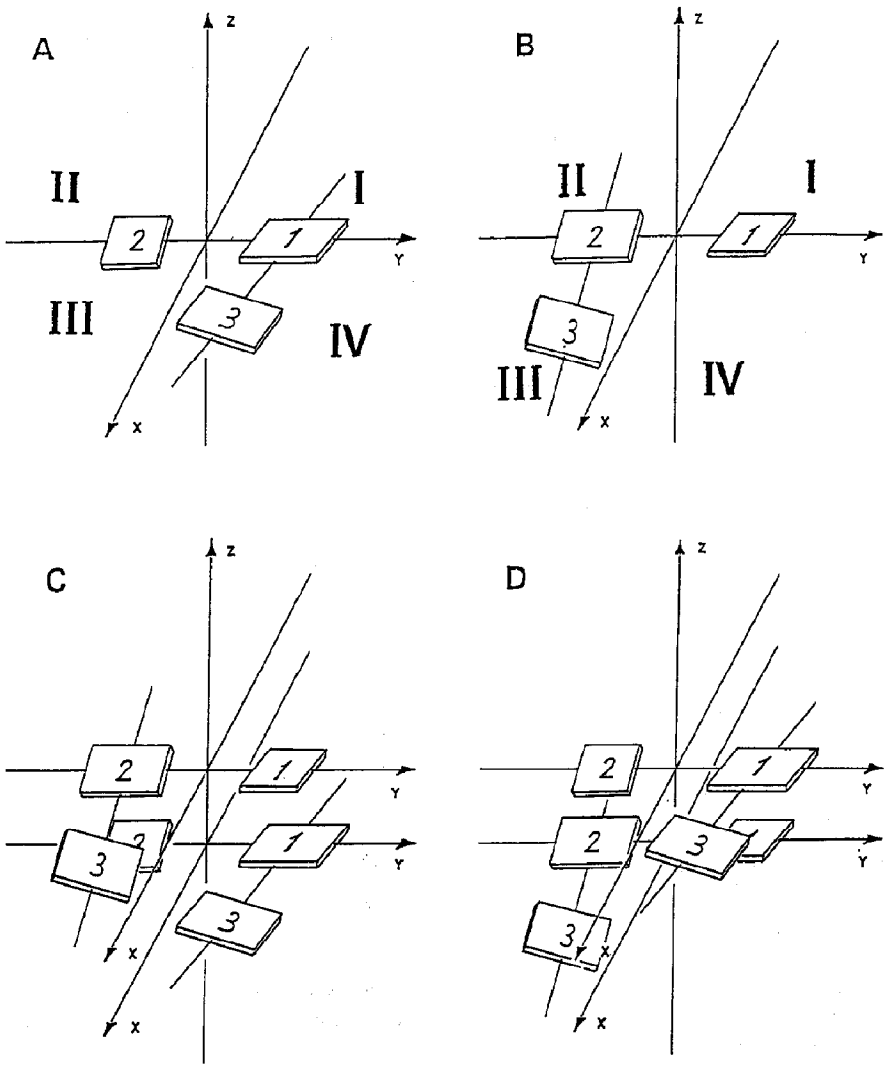


Fig. 4. Schematic representation of a triplet AToT (A) drawn after X-ray fiber diffraction data [17] and of the same triplet rotated 180° about the Ox axis (B). Propeller twists 1–2, and 1–3, and inclination of base pair 1–3 are exaggerated for sake of illustration of the helical symmetry of these parameters. All triplets are represented in the same reference frame [52]. Strand numbering is arbitrary and is defined in (A).

The two non-hydrogen atoms of the first base that are hydrogen bound to strand 3 define by their middle the origin, O_1, of the first reference frame and by the line that join them the O_1x_1 axis. The two non hydrogen atoms of the triplet third base that are hydrogen bound to strand 1 define by their middle the origin, O_3, of the reference frame of the third base and by the line that join them the O_3x_3 axis.

The orientation of the axes is based on the following analysis. The EMBO conventions [52–54] specify that to orient a reference frame attached to a base pair or to a single base of double strand DNA, a strand numbering (1) and (2) must be chosen and the Oz axis is, then, oriented in the 5′ to 3′ direction of strand 1. The Ox axis is in the base plane and is oriented from the minor to the major groove. The Oy axis is oriented so that the reference frame is right-handed. This is summarized in Fig. 3B, and defines a particular reference frame. If the strand numbering is reversed, a second frame is obtained where Oy and Oz are oriented in the opposite direction. Helical parameters computed with these frames are identical in absolute values, but their signs change as indicated in the third column of Table 1 labelled "Sign". Thus frame orientation is important. As the orientation of the Ox axis is not specified by the EMBO conventions for the third strand of a triple helix, we propose the following extensions. First choose a helical position numbering I through IV in a circular order. The Oz axis is perpendicular to the mean plane defined by the six-membered rings of the bases, and is oriented as in the Watson–Crick base pair. Figs. 1a and 1b show the axis 5′ to 3′ applied on the first helical position pointing towards the reader. We can see how to reverse the orientation of a whole triplet by changing each orientation of the bases and swapping the positions I and IV with the positions II and III. So, the triplet TAToβqt is the same triplet as AToTβpt. By our conventions, we orient the Ox axis, from helical position III to IV as shown in Figs. 3C and 3D. The Oy axis is defined so that the final orthonormal reference frame is right-handed. As before, if the Watson–Crick strands 1,2 are numbered 2,1, helical parameters computed with these frames are identical in absolute values and their signs change exactly as indicated previously, in the second column of Table 1.

2.4. Construction of the 32 triplets

This construction is based on the following three simplifying hypotheses:

(i) *Helical symmetry or pseudodyadic symmetry is imposed on helical parameters of the third strand.* We have seen that it was sufficient to construct the triplets with a purine in first position, AToT, $GCoC^+$, AToA, GCoG, to deduce the triplets with the pyrimidine in first position TATo, CGC^+o, TAAo, and CGGo. However two given triplets AToT, for instance, and TATo obtained by rotating 180° AToT around the axis Ox, cannot stack simply along the Oz helical axis as illustrated in Fig. 4. In Fig. 4A we can see a schematic triplet showing the characteristic helical parameters that are found by X-ray fiber diffraction [17,18]: a large propeller twist for base pairs 1–2 and 1–3 and a large inclination for base 3, as can be read from Table 1. In Fig. 4B, the same triplet is shown rotated by 180° around Ox. As shown on Figs. 4C and 4D, some helical parameters such as the propeller twist between bases 1 and 2 do not create any important discontinuity at the junction formed by a triplet and the same triplet turned by 180° about Ox. However, other helical parameters such as the inclination of base 3 give rise either to steric clashes, Fig. 4C, or to discontinuities as shown schematically in Fig. 4D.

At this stage of model building, our aim was to create a simple database of triplets, that can be easily and properly stacked by a simple helical transformation. Thus, we decided to construct triplets for which all characteristic helical parameters are sufficiently preserved by pseudodyadic symmetry (rotation of 180° degrees about Ox, Fig. 4). The Watson–Crick base pair in a triplet has sufficient pseudodyadic symmetry, i.e. helical parameters marked "−" are small (Table 1) to be kept unchanged in their original conformation as defined by fiber diffraction [17]. This is not the case for the third nucleotide. In order to prevent these collisions, the third nucleotide was moved

as a block so that its base lies in the plane xOy. This amounts to cancel the inclination and tip of the third base (Table 1). As shown later, these helical parameters are partially restored by energy minimizations.

(ii) *The most external hydrogen bond is taken as a template to define base pairing.* The hydrogen bonds between the bases of strands 1 and 3 are not parallel unlike the Watson–Crick hydrogen bonds between those of strands 1 and 2. They make an angle of 36° in AToT, $GCoC^+$, GCoG and AToA (reverse Hoogsteen pattern) triplets and an angle of about twice this value in triplet AToA in the Hoogsteen pattern. Hydrogen bond involving the N7 of the first strand purine is more distant from the helix axis (3.8 Å in AToT) than the other hydrogen bond (1.5 Å in AToT), and is found to be more instable, or liable to be disrupted under mechanical stress. This tendency is further enhanced if we observe that the geometry of hydrogen bonding has an optimal arrangement that is slightly distorted when the third base is forced to lie in the plane xOy, as explained previously. Thus to improve the stability of our structures, we set the position of the third base by using the most external hydrogen bond as a template as described below. This procedure yields two further benefits: hydrogen bonding can be defined simply for all different kind of triplets, and the final pairing remains close to experimental structural data on triple helices and tRNA.

The construction of the starting structures of the database was achieved by a program, solely written with these two hypotheses. This program, written in OCL [48] and MORCAD [47], is based on four modules ROT, POS, XOY and ALPHA that operate on a copy of either the first, "1", or the second nucleotide, "2", as indicated in Table 2, to produce the third nucleotide. Table 2 shows how these modules are combined to produce the starting triplet structures

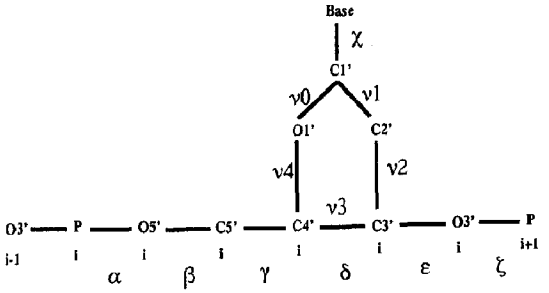
- ROT rotates the base of this third nucleotide by 180° about C1′–N1 or C1′–N9.
- POS casts the position of the third nucleotide into proper hydrogen bonding by two translations and a rotation. Depending on the triplet, these translations and rotation are those required to reproduce the most external hydrogen bond as in AToTβpt [17], adenine.inosine.inosinc [18], or as in tRNA [19,20].
- XOY translates third nucleotide to set the base in the xOy plane.

Table 2
Construction of all the starting triplet structures for $GCoC^+$. Each row summarizes the construction of the conformation indicated in column 1. The starting conformation of $GCoC^+$ is built from the canonical Arnott [15,17,18] AToTβpt. The functions ROT, POS, XOY and ALPHA are applied to the base indicated in the second column if the box is filled with a cross

Triplet	Transformed nucleotide	ROT	POS	XOY	ALPHA
αps	C^+				×
αpt	C^+	×	×		×
αqs	C		×	×	×
αqt	C	×	×	×	×
βps	C^+	×	×		
βpt	C^+				
βqs	C	×	×	×	
βqt	C		×	×	

Table 3
Data in deg taken from the literature concerning the torsion angles shown in Scheme 1. Rows 2 and 3: values for double stranded A- and B-DNA from X-ray fiber diffraction [55]. Rows 4 and 5: values and standard deviation for the six triple helices from X-ray fiber diffraction [18]

	Torsion angles						
	α	β	γ	δ	ϵ	ζ	χ
A-DNA	−90	−149	47	83	175	−45	−153
B-DNA	−39	−151	31	157	159	−99	−95
3rd strand	−58	173	51	86	−148	−81	−153
3rd strand stand. dev.	9	4	7	7	6	10	10



Scheme 1. Definition of the torsion angles.

– ALPHA rotates the deoxyribose of the third nucleotide by 120° and the H1′ corresponding atom by −120° about an axis parallel to the helical axis passing through C1′.

(iii) *The dihedral angles are forced to classical values in two stages.* In their classical work, Arnott et al. [18] assumed that the conformations of the sugar puckers in the triple helices were in C3′ endo. Several observations in solution have shown that the C2′ endo conformation can be dominant [56] or important [39,46,57]. In all 32 triplets, the sugar puckers of all nucleotides are forced to adopt a C2′ endo conformation.

Up to this point, we have set the third nucleotide with respect to the double helix and forced the sugar conformation. The triplet would be completely determined with values of the torsion angles of the sugar phosphate chain. (See Scheme 1 for their definitions). Table 3 shows that torsion angles α, β, through ζ stay in the classical conformation, respectively, g−, t, g+, g+/t, t, g− for A−, B− DNA [55] or known DNA triple helices [18]. This observation leads us to search triplexes that could yield triple helices with torsion angles values around those of Table 3. For this reason, an AToTβpt triple helix with three repeats was minimized under no constraint. It yielded a geometrically uniform structure characterized by the torsion angles given in Table 4. This set of torsion angles served as the basic template about which all other triple helices were explored. 32 triple helices with five repeats were energy refined, first with pucker constraints and, then under torsion angle constraints of Table 4. The 32 central triplets were extracted from these structures to form the final database (Fig. 4).

Table 4
Values in deg to which the torsion angles are constrained when all the sugar puckers of all strands of AToTβpt are forced to either the C2′ or the C3′ endo conformations

	Torsion angles				
	α	β	γ	ϵ	ζ
C2′ endo	−65	177	60	177	−100
C3′ endo	−74	173	62	−169	−64

2.5. *Molecular mechanics and dynamics calculations*

Energy minimizations were run on IBM RS6000/530 and SGI 4D310 work-stations using the MORMIN program [47], a rapid and efficient quasi-Newtonian minimizer of AMBER [58–60] force field. The net charges on the protonated cytosine were calculated ab initio using the QUEST module of AMBER so that the set of parameters is as homogeneous as possible (Table 5). To simulate the screening effect of the solvent, a gas phase potential was employed where the dielectric constant [61], D_{ij}, is proportional to the distance d_{ij} separating a pair of atoms: $D_{ij} = Cd_{ij}$. C was taken as 4 Å^{-1}. All atom pairs were included in the calculations of nonbonding interactions. Minimizations were stopped when the root mean square of the energy gradient was less than 0.1 kcal/Å. In order to change the sugar ring puckers it was sufficient to force the torsion angle δ (Scheme 1) to δ_0 ($\delta_0 = 144°$ for C2′ endo conformation [62]) by addition for each constrained sugar of the penalty $k(\delta - \delta_0)^2$ to the energy function of AMBER. We have constrained the torsional angles ν_0, ν_1, ν_2, ν_3, and ν_4 [63] to the values of −22, 33, −33, 22, and 0°

Table 5
Net charges on the atoms of the protonated cytosine as computed from the ab initio QUEST module of AMBER

O5′	−0.509	C5′	0.118	H5′	0.021	H5″	0.021
C4′	0.036	H4′	0.056	O1′	−0.368	C1′	0.376
H1′	0.009	N1	−0.067	C2	0.719	O2	−0.388
N3	−0.569	H3	0.372	C4	0.845	N4	−0.734
HN4A	0.354	HN4B	0.376	C5	−0.546	H5	0.190
C6	0.259	H6	0.134	C3′	0.233	H3′	0.025
C2′	−0.307	H2′	0.081	H2″	0.081	O3′	−0.509

respectively to force the sugar pucker in C2′ endo conformation (Scheme 1). For these torsion angles, the force constant was set to 50 kcal/mol deg^{-2}.

Starting structures were minimized with the MORMIN program [47]. The Verlet algorithm [64] was used with a time interval of 0.002 ps per step. Hydrogen atom bond lengths were constrained with the SHAKE algorithm [65]. A temperature coupling parameter [66] of 0.4 ps was used. MD runs were started by assigning random velocities, that followed a Gaussian distribution at 50 K. The molecule was heated from 50 to 300 K in stages of 50 K over a period of 11 ps followed by a 5 ps period during which random thermal velocities corresponding to a Gaussian distribution were repeatedly reassigned at intervals of 0.2 ps. Further equilibration took place over 4 ps. During the entire 20 ps preparation procedure, each atom of the molecule was bound to the starting structure by very weak harmonic constraints. Constraints were determined in such a way that the molecule could fluctuate as freely as possible without diverging in a major way from the starting structure [67]. Kinetic, potential and total energy of the molecule remained at equilibrium values during the entire 100 ps run without drifting. During the runs the temperature fluctuated around 300 K by ±15 K. The computation time for 120 ps in vacuo MD runs on heptamers took 13 h 20 min on an IBM RS6000/530 workstation.

The hydrogen bonds of the four extreme base triplets were strengthened with a weak harmonic potential to avoid end-fraying and prevent the trimolecular complex from disrupting from its ends. The force constant was set to 5 kcal/mol/$Å^2$ and the equilibrium hydrogen bond length to 2 Å.

3. Results

3.1. Geometrical analysis of the triple helices built from the triplet database

All distinct homogeneous triple helices with five repeats were generated by successive helical transformations starting with the corresponding database triplet, each repeat is obtained from the preceding one with a rotation around the axis Oz of 30.0° or 32.7°, and a translation along Oz of 3.26 Å or 3.04 Å. We searched the triple helices that admitted the set of torsional angles given in Table 4 and that did not get distorted by energy minimization to the point of losing their helical symmetry. These losses of helical symmetry occur as stacking or pairing of the third strand progressively deteriorates along these short triple helices. Table 6 summarizes the results. The triplets, marked with an X, cannot form a homogeneous triple helix, mostly because atoms 05′ and 03′ of neighbor nucleotides on the third strand are closer than 1.5 Å. These conformations are called unfavorable, however this does not exclude the possibility that they can exist as short linker stretches. A second class of triple helices, marked with the star symbol * can be eliminated: they do form an homogeneous triple helix but strands 1 and 3 are closer to each other than strands 1 and 3 of ATo$T\beta$pt (Table 6). Infinite regular triple helices have not been studied since no boundary conditions have been used in the minimizations. However, it may be possible to build geometrically uniform homogeneous triple helices with the triplets that are left blank in the Table 6.

Table 6
Possibility of a geometrically uniform homogeneous triple helix. Triplets marked "X" are unfavorable: they cannot form geometrically uniform homogeneous triple helix within our three simplifying hypotheses. Under these hypotheses, the triplets marked "*" are unlikely to form geometrically uniform homogeneous triple helices because of the close proximity of strands 1 and 3

	αps	αpt	αqs	αqt	βps	βpt	βqs	βqt
AToA			×	×	×	×		
AToT	*		×		×		*	
$GCoC^+$	*		×		×		*	
GCoG	*		×		×		*	

As summarized in Table 6, ATo$T\alpha$pt, ATo$T\alpha$qt, ATo$T\beta$pt, ATo$T\beta$qt as well as $GCoC^+\alpha$pt, $GCoC^+\alpha$qt, $GCoC^+\beta$pt and $GCoC^+\beta$qt can form geometrically uniform homogeneous helices whereas ATo$T\alpha$qs, ATo$T\beta$ps, $GCoC^+\alpha$qs and $GCoC^+\beta$ps form very poor triple

Table 7
Torsion angles for the third strand of AToTβpt, GCoC$^+$βpt and GCoGβqt when all the sugars are in C2′ endo or in C3′ endo conformation (column 2) obtained by the minimizer when it is launched from the conformation taken from the database

	Torsion angles						
		α	β	γ	δ	ε	ζ
AToTβpt	2	−68	173	59	111	−177	−88
GCoC$^+$βpt	2	−68	171	60	110	−176	−88
GCoC$^+$βpt	3	−74	175	63	78	−168	−159
GCoGβqt	2	−82	137	49	141	−88	168

helices. AToTαpt and AToTαqt have close energy values and cannot be distinguished on this ground. AToTαpt and AToTβpt are experimentally observed in homogeneous triple helices whereas AToTαqt is observed in mixed triple helices containing both AToTα and GCoC$^+$α [24]. AToTβqt and GCoC$^+$βqt have not yet been observed experimentally presumably because AToTβpt and GCoC$^+$βpt are more favorable and perfectly isomorphic.

Many different workers have shown that the two purine tracts are antiparallel to each other in purine-pyrimidine-purine type triple helix. It has been proposed for homogeneous GCoG in plasmid DNA [68]. It has been demonstrated for homogeneous GCoG and for triple oligonucleotides containing both AToA and GCoG triplets [23,69]. The classical work of Cooney et al. [14] showed that the binding of a 27-base long oligonucleotide to DNA represses transcription of the human c-myc gene in vitro. They showed that the third strand binds antiparallel to the G-rich polypurine sequence [25]. Table 6 shows that two structures, GCoGβpt and GCoGβqt can

Table 8
Values of the helical parameters of AToTβpt, GCoC$^+$βpt, GCoGβqt and GCoGβqt/AToAβqt obtained from the 100 ps production phase of the dynamics. Lengths and angles are respectively in angstrom and deg. Helical repeat are in number of triplets/turn. The fluctuations are given for AToTβpt and should be multiplicated by up to 1.5 for the other triple helices. The fluctuations are here thermodynamical quantities. In principle they are inversely proportional to the second derivative of the energy

	AToTβpt		GCoC$^+$βpt		GCoGβqt		GCoGβqt/AToAβqt	
	1–2	3	1–2	3	1–2	3	1–2	3
X-displacement (Å)	0.5 ± 0.2	0.4 ± 0.2	0.4	0.3	0.5	0.5	0.5	0.3
Y-displacement (Å)	−1.4 ± 0.2	1.5 ± 0.2	−1.6	1.4	−1.5	1.3	−1.4	1.4
inclination (deg)	7.1 ± 5	−6.8 ± 8	6.1	−20.7	0.6	27.0	8.7	−17.4
tip (deg)	2.9 ± 5	10.8 ± 7	−2.5	5.7	1.8	3.7	4.2	11.0
shift (Å)	3.5 ± 0.6	2.0 ± 1	−0.2	−0.1	0.6	1.7	0.4	1.7
slide (Å)	2.4 ± 0.7	2.6 ± 1	−2.9	2.8	−2.0	1.9	−2.2	2.4
rise (Å)	3.1 ± 0.3	3.1 ± 0.3	3.0	3.0	3.2	3.3	3.0	3.1
tilt (deg)	−1.6 ± 3	−7.4 ± 9	0.1	0.9	−1.0	0.8	−1.2	−3.4
roll (deg)	3.3 ± 5	−2.0 ± 8	5.0	−11.5	−2.7	−11.7	3.2	8.0
helical repeat	11.0 ± 1	11.1 ± 2	11.0	10.8	9.7	10.0	10.4	10.2
helical twist (deg)	32.8 ± 3	32.4 ± 7	32.9	33.4	37.1	35.9	34.6	35.4
shear (Å)	−0.2 ± 5	0.1 ± 0.5	−0.5	0.8	−1.0	0.1	−0.4	0.2
stretch (Å)	0.1 ± 0.2	−2.8 ± 0.2	0.0	−2.6	−0.1	−2.7	0.0	−2.9
stagger (Å)	0.2 ± 0.5	0.4 ± 0.5	−0.1	−0.5	−0.1	0.7	0.2	0.3
buckle (deg)	4.1 ± 10	2.5 ± 10	−3.7	−11.4	6.8	−31.1	4.2	−13.4
propeller twist (deg)	−19.6 ± 10	16.0 ± 10	−13.3	1.5	0.7	7.3	14.2	15.9
opening (deg)	−1.2 ± 5	−3.3 ± 5	−1.3	−11.0	0.4	−3.3	0.1	0.8

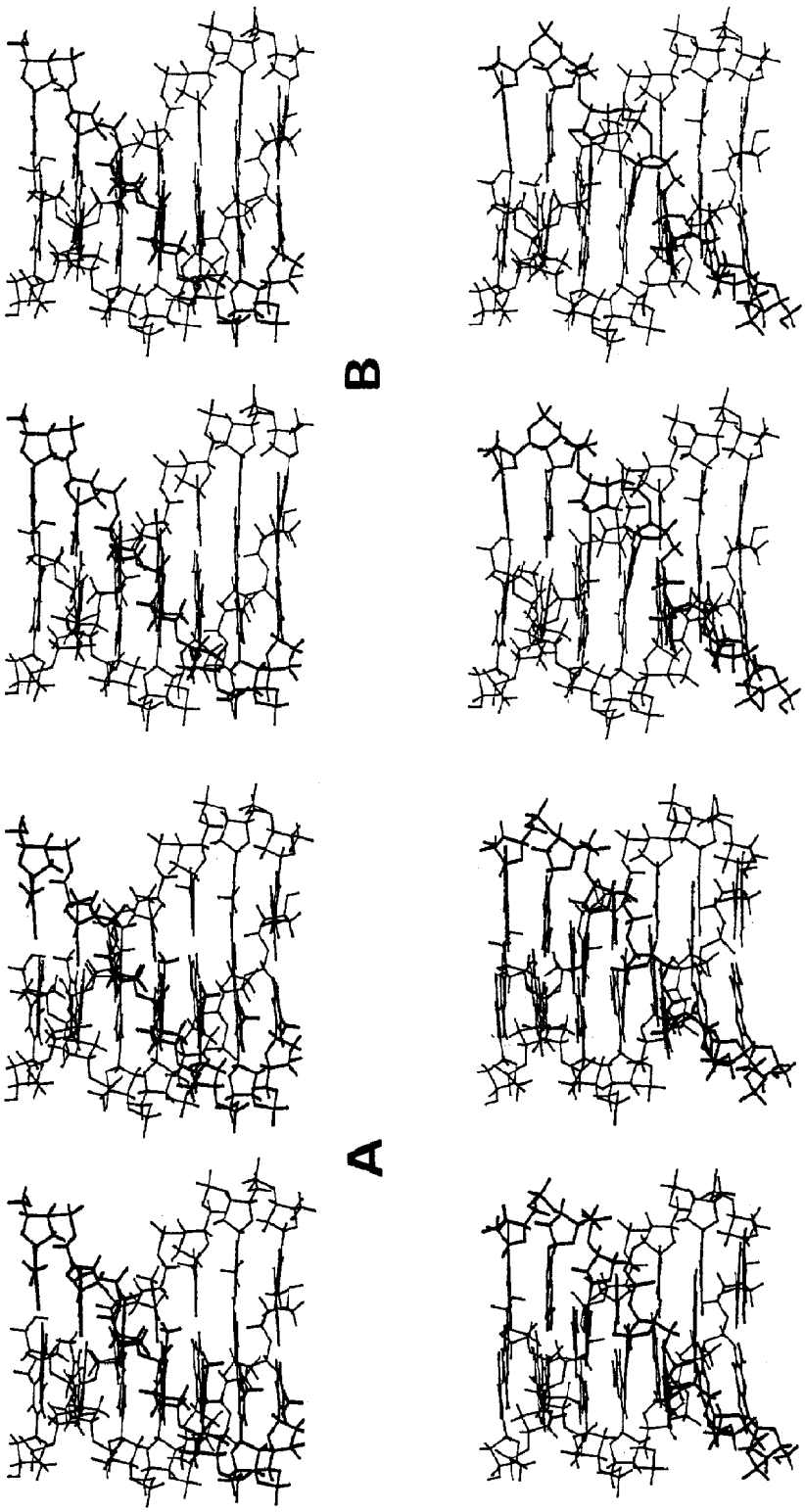


Fig. 5. Stereographic views of the minimized structures of AToTβpt (A), GCoC$^+$βpt (B), GCoGβqt (C) and (GCoG/AToA)βqt (D).

form homogeneous helices. It has been shown that both can be observed according to target DNA sequence [70].

Several workers [23,69] demonstrated that the third strand is antiparallel to the first for triple helices containing both AToA and GCoG triplets and the two types of stacking AToA-GCoG and GCoG-AToA. Table 6 rules out AToAβpt and AToAβps but does not discriminate between AToAβqs and AToAβqt. Full relaxation by energy minimization of the triple helix heptamers AToTβpt, GCoCβpt and GCoGβqt yields to the molecules characterized by the torsion angles values given in Table 7. Helical parameters for the energy refined structure obtained from the database AToTβpt triplet are given in Table 1. Torsion angle values and helical parameters for AToTβpt agree quantitatively well with X-ray fiber diffraction data [18] if one considers that these data are currently the best data available and are somewhat approximate. Table 7 indicates that GCoGβqt should be mostly found in B II conformation where the torsion angles ϵ and ζ are respectively (g −, t) instead of (t, g −) as in the more classical B I conformation [62].

3.2. Helical parameters

The EMBO workshop [52] set up a new convention for the analysis of double stranded nucleic acids. In this work we have made use of those definitions for the Watson–Crick double helical part of our structures. As for the third strand, we have extended the convention in the way that follows: we define the roll, tilt, inclination, tip, twist, *X*-displacement, *Y*-displacement and rise parameters as done for a classical Watson–Crick base pair, yet the reference frame is constructed with a single base. Definitions of the buckle, propeller twist, opening, shear, stretch and stagger are based on the construction of two reference frames: one for the purine (first or second strand) and the other for the third strand base. Due to the impossibility to reproduce the definitions relative to a classical Watson–Crick base pair in the case of a Hoogsteen or "pseudo-Hoogsteen" base pair, the reference frames that we have devised are based on hydrogen bonds. Each base is provided with its own reference frame. Both frames are linked in an OCL procedure [48], and the desired parameters are obtained through a "list wedge" command in the MORCAD program [47].

To get into more details, there are two types of reference frames: in the local frames, parameters (roll, tilt, buckle, propeller twist, opening, shear, stretch, and stagger) are measured with respect to a local axis described below, whereas in the case of "global" frames, parameters (inclination, tip, *X*-displacement, *Y*-displacement, rise) are measured with respect to the global helix axis. The *Z* axis of the local frame is perpendicular to the mean plane of the base and points in the 5′ to 3′ direction of the first strand. The *Y* axis joins the two atoms involved in the hydrogen bonds. If one or both of the atoms are a proton, the nitrogen to which it is bound is taken into account. The *X* axis points towards the major groove of the helix. The *Y* axis is oriented so that (*X*, *Y*, *Z*) is a right-handed reference frame.

The major advantage of this procedure is that despite the heterogeneity of the bases – and subsequently of the hydrogen bonding patterns – all triplets are treated with a high degree of consistency. For the third strand, the Stretch parameter value of 2.9 Å mostly reflects the separation between the origins of the reference frames in the original construction (see Fig. 3A).

3.3. Molecular dynamics on homogeneous and alternate structures

As already mentioned, the models obtained by molecular mechanics calculations on the structures directly built from the triplet database present characteristics of regularity and homogeneity, at least for those which are supposed to give rise to geometrically uniform helices. We should now validate the database by showing that these regular structures remain stable during a dynamics run. Our MD runs were performed on heptamers as a compromise between structuraly meaningful results and reasonable computer time bounds. We have stacked 7 identical triplets upon each other making the homogeneous AToTβpt and GCoC$^+$βpt, GCoGβqt triple helices. We

also have built d(GAGAGAG)2.d(CTCTCTC) by the alternate stacking of GCoGβqt and AToAβqt triplets. The minimized structures of these heptamers are shown in Fig. 5 and are used as the starting structures of the dynamics (time $t = -20$ ps).

The rms deviations from their starting structures of the four heptamers are plotted versus time t in Fig. 6. In the negative times, the atom motions are restrained to their starting positions by a force constant so that the rms remains small. At time 0, these restraints are released, the production phase begins and the rms grows up as expected. For AToTβpt and the alternate GCoGβqt/AToAβqt (Figs. 6A and 6D), the rms deviations reach soon their plateau value at 1.5 Å and 2.1 Å. The $GCoC^{+}$βqt heptamer has the worse rms curve. The rms shows a first plateau (1.2Å), then increases around 35 ps and stabilizes at 3.2 Å (Fig. 6B). For GCoGβqt, the rms stays around 1.9 Å, and for a short time culminates at 2.8 Å (48 ps) before it goes back to the earlier value showing the stability of the starting structure (Fig. 6C).

The torsion angles of the first two Watson–Crick strands have been plotted but no special behavior could be detected if the extremities are ignored (results not shown). The average values during the production phase are close to the starting values. The fluctuations are between 8° and 15° in the order $(\epsilon, \beta) < (\gamma, \alpha) < \zeta < (\delta, \chi)$ for AToTβpt, GCoGβqt and for the alternate GCoGβqt/AToAβqt. The torsion angles of the Watson–Crick strands of $GCoC^{+}$βqt have a similar behavior except for the central C. For this one neutral C, the fluctuations of α and γ are 45 and 31° respectively. The values of α cram about 290°, but have a well populated tail ranging from 100° to 290°. The variations of γ are anticorrelated with α (the correlation factor is equal to -0.93) and follow the relationship $\gamma = 240° - 0.63\alpha$, where the definition interval of α ranges from 0° to 360°.

In contrast to the first two Watson–Crick

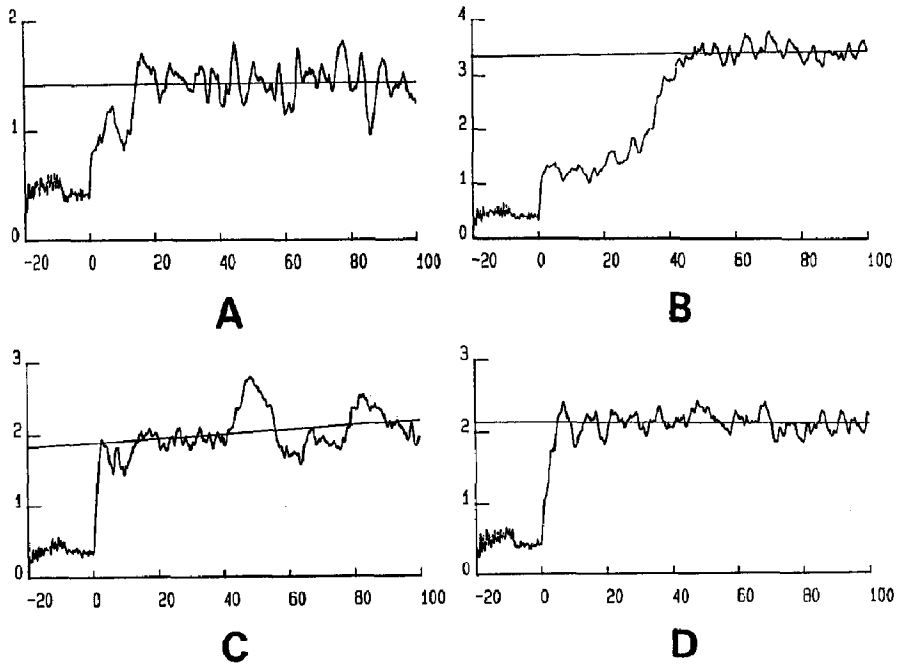


Fig. 6. The rms deviation from the starting structures ($t = -20$ ps) shown in Fig. 5 versus time: AToTβpt (A), $GCoC^{+}$βpt (B), GCoGβqt (C) and (GCoG/AToA)βqt (D). The phases of heating (11 ps), randomization (5 ps) and equilibration have negative times. The phase of production (100 ps) starts from 0 ps.

strands, the third strand is much more flexible and its torsion angles fluctuate accordingly. During the production phase, some angles behave in an analogous way for the four heptamers we have studied. For instance, the glycosidic angle, χ, is in the range (g −, t) for all triple helices (Fig. 7). However most torsion angles averages and fluctuations largely depend on the sequence. The pseudorotational angle (not shown) and its related torsion angle δ fluctuate much so that the sugar puckers populate both the C2′ endo and C3′ endo conformations for AToTβpt (Fig. 7A). The triple helice GCoC$^+$βpt is here again the least stable. Torsion angles α, ϵ and ζ of the third strand have large fluctuations (Fig. 7B). For instance the α torsion angle of C6 fluctuates very much from g − to t. The puckers of GCoGβqt are exceptionally stable. For some residues of GCoGβqt, ϵ and ζ torsion angles of the third strand are in the g − and t configuration respectively, at least two times during the 100 ps of production. Therefore, there is a tendency for the structure to evolve to a B II form [71] (Fig. 7C). So we have performed MD calculations, starting from a B II form constrained conformation. In that case, the central bases of the third strand remained in B II throughout the run and the extreme ones flipped to B I and remained in that conformation.

For the third strand of alternate GCoGβqt/

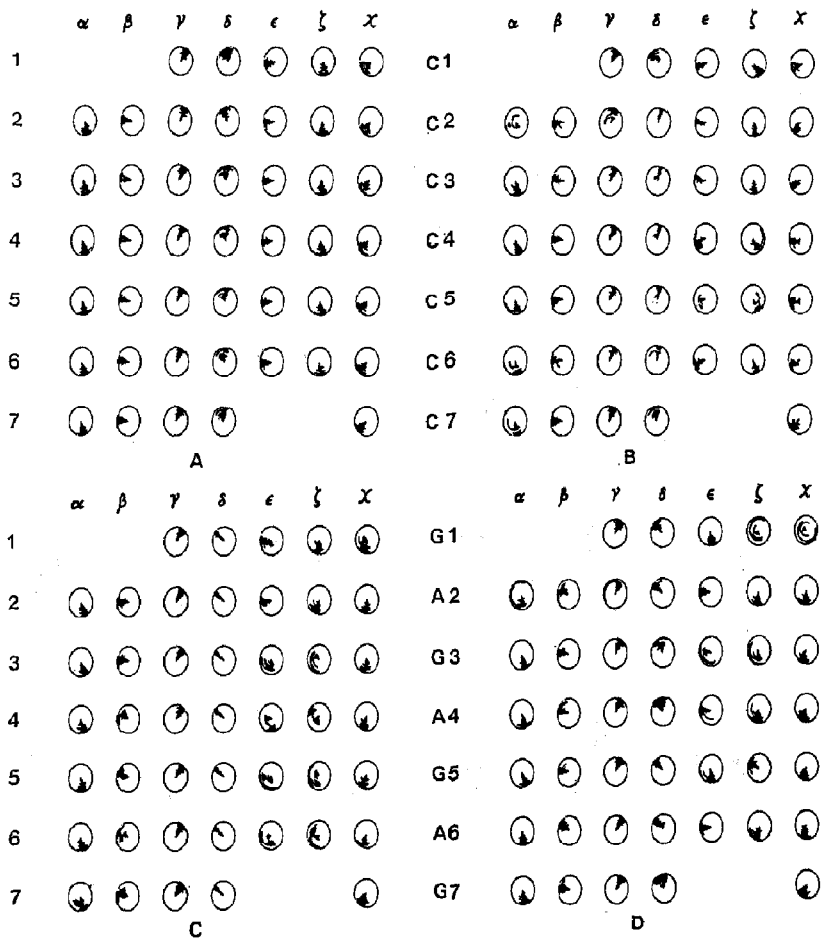


Fig. 7. Dial plots of the third strand torsion angles over the MD production phase for AToTβpt (A), GCoC$^+$βpt (B), GCoGβqt (C) and (GCoG/AToA)βqt (D). Each dial is a trigonometric circle with 0° on the right at 3 o'clock and positive angles are anti-clockwise.

AToAβqt the triplets of different types have not the same behavior. The tendency of the GCoGβqt triplet to the B II form is observed here too (Fig. 7D). The values of the torsion angles ϵ and ζ switch between (g −, t) and (t, g −) as the triplets GCoGβqt alternate with the triplets AToAβqt. Two different triplets, stacked on each other, can form a smooth triple helix and still keep their intrinsic properties.

The helical parameters for AToTβpt and alternate GCoGβqt/AToAβqt have fluctuations that keep low amplitudes. As can be seen from the evolution of the helical parameters with time, the inclination of all the base pairs have wide fluctuations relative to the other parameters most of the time. For GCoC$^+$βpt, the inclinations of the base pairs are in the range from −8° to 23° (Fig. 8). The Opening parameter between the bases of the first strand and the bases of the third strand decreases between 31 ps and 88 ps and is anticorrelated to shear. In contrast, the shear and stretch parameters are correlated and reach the values of 4.6 Å and −0.5 Å respectively (Fig. 8).

Most of the helical parameters of the GCoGβqt heptamer fluctuate around their starting value. The temporary wagging in the rms deviation at 56 ps (Fig. 6) is accompanied by a change in the propeller twist, opening and shear parameters between the bases of the first and third strands (Fig. 8).

4. Discussion

4.1. Name conventions and construction rules

In this work we have constructed a complete database of triplets with three construction rules concerning helical symmetry, pairing and torsion angles. The first rule is not unreasonable since it

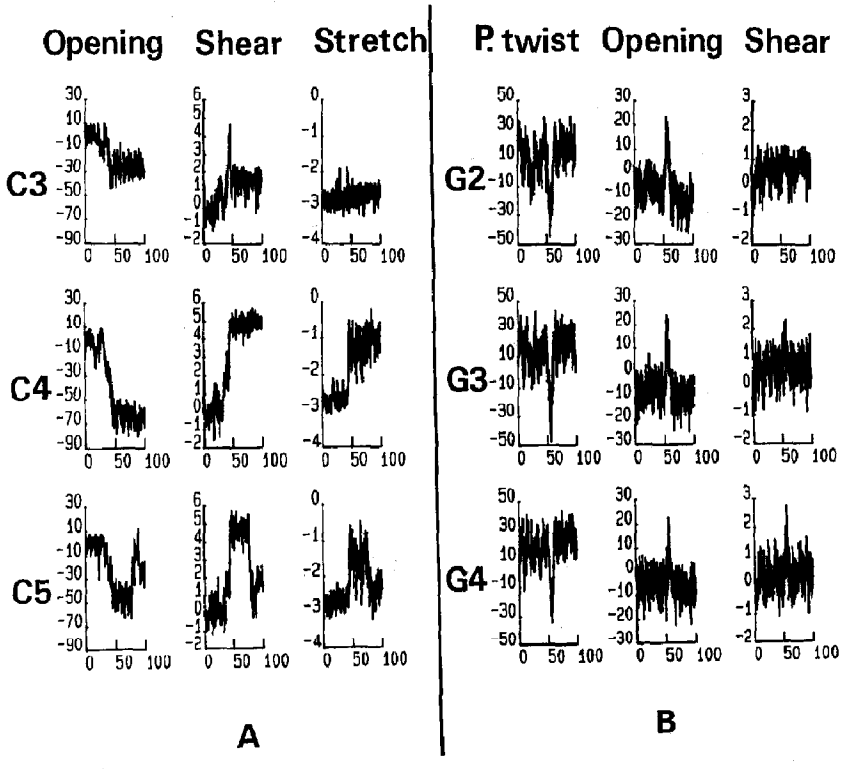


Fig. 8. Variations of helical parameters over the MD production phase for the third strand (relatively to the first strand) of GCoC$^+$βpt (A) and GCoGβqt (B). Opening and propeller twist are in deg. Shear and stretch are in angstrom.

amounts in database triplets to cancel only two helical parameters tip and inclination. Furthermore these helical parameters are partially recovered by energy minimization and bear little incidence on the rest of the helical parameters as shown in Table 1. The second rule was set to establish similar pairings for all triplets in accordance with known molecular structures. The third rule serves to construct a sugar phosphate chain to connect in a general way a triplet to any other database triplet.

Once the database was completed we searched for geometrically uniform homogeneous triple helices, whose torsion angles of the sugar phosphate chains qualitatively agree with data of Table 3 and 4. The conformations of the third strand could not depart too much from those of a classical DNA. It is remarkable that this method is sufficiently restrictive to eliminate more than half of possible triplets and sufficiently general to keep all structures that have been directly or indirectly observed. One can make use of qualitative experimental data in combination with our results to infer pairing and molecular structure as we have shown for GCoGβqt and AToAβqt. As a result, the third rule turns out to be a very reasonable assumption to construct this triplet database.

4.2. Isomorphism and importance of the of C1', P and O3' radii

The crystal structure of tRNA-Phe [20] revealed the presence of three different base triplets in the D stem region. One of them involves bases C13, G22 and m^7G46 and has the conformation CGGoβpt. In tRNA-Arg the corresponding bases are U13, A22, and A46. If the interactions between the phosphate of A9 with the amino groups of C13 and m^7G46 are ignored, it can be conjectured that the triplet C13 G22 m^7G46 can be replaced by U13 A22 A46 without much changing the sugar-phosphate backbone. It was argued that the triplet U13 A22 A46 should have the form TAAoβpt because the geometry of TAAoβpt is more isomorphic than TAAoβps to that of CGGoβpt, as shown in Figs. 9A and 9B [72]. The triplet TAToαqt and CGC^+oαqt are truly isomorphic (i.e. the sugar phosphate chain between different triplets is identical to that between identical triplets) as are TAToβqt and CGC^+oβqt. Figs. 9A and 9B show that there is qualitatively the same relationship of isomorphism between TAToαqt and CGC^+oαqt as between TAAoβpt and CGGoβpt. But isomorphism is not a sufficient explanation since triple helices containing alternate TAAoβpt and CGGoβpt have been

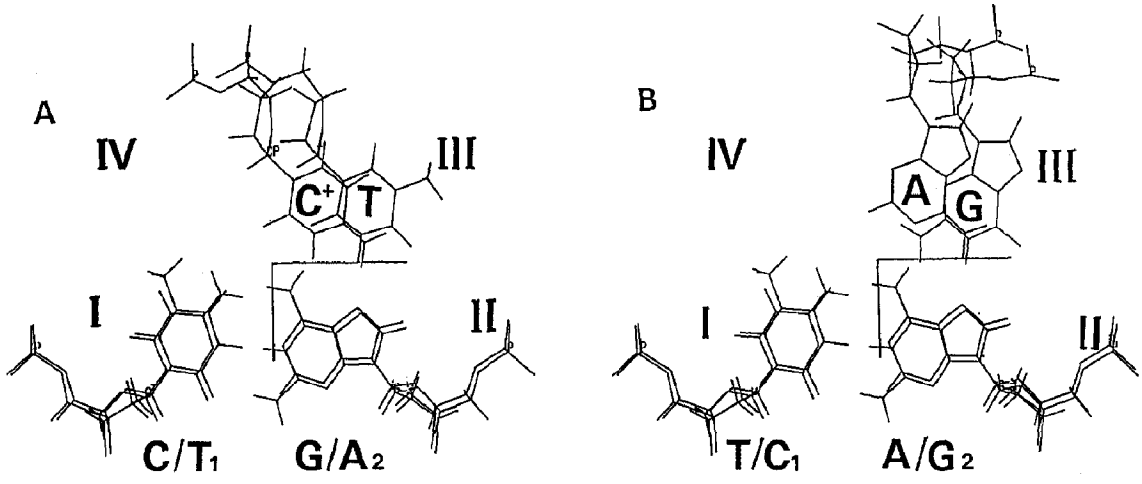


Fig. 9. View of the superimposed triplets TAToαpt and CGC^+oαpt (A) and of the superimposed triplets TAAoβqt and CGGoβqt (B). The reference frame shown here is the same as in Fig. 3 and has been used to represent all triplets. Its Oz axis, which is perpendicular to the figure plane, is the global helical axis.

observed while it is not the case for TAToαqt and CGC$^+\alpha$qt [5].

Figs. 10, 11 and 12 respectively give the positions of the C1′, P, and O3′ atoms for all database triplets. All C1′, P, and O3′ atoms, of the three strands, for all triplets capable of forming geometrically uniform homogeneous triple helices, AToTαqt, AToTβpt, GCoC$^+\beta$pt, GCoC$^+\alpha$qt, AToAβqt, GCoGβqt, except AToTαpt are contained within closely spaced circles as shown in Fig. 10 for C1′: 6.8 Å and 8.0 Å; in Fig. 11 for P: 10.2 Å and 11.4 Å; and in Fig. 12 for O3′: 9.2 Å and 11.0 Å. This geometrical analysis singles out AToTαpt (and possibly GCoC$^+\alpha$pt) among triplets capable of forming geometrically uniform homogeneous triple helices. In a first approach, within the assumption of identical pucker confor-

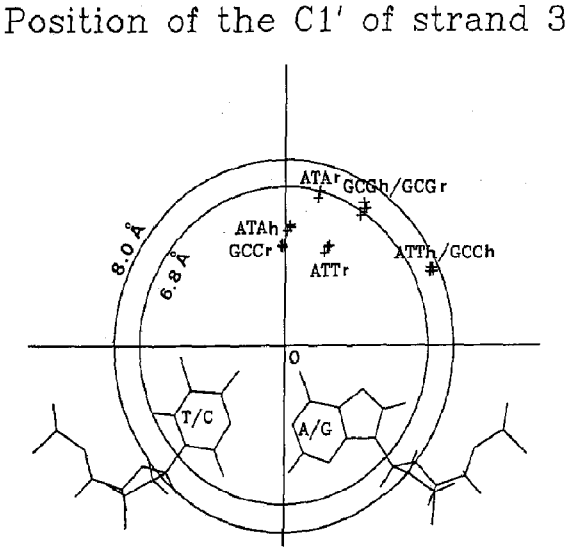


Fig. 10. Representation of the positions of the C1′ atoms with respect to the double strand for database triplets that can form geometrically uniform homogeneous triple helices, AToTαpt, AToTαqt, AToTβpt, GCoC$^+\beta$pt, GCoC$^+\alpha$qt, AToAβqt, GCoGβqt, and GCoC$^+\alpha$pt. The positions depend little on the α/β and p/q features of the conformation (here chosen in anti). The C1′ atoms in all three strands corresponding to all these latter triplets, except AToTαpt and GCoC$^+\alpha$pt, are contained within two circles of radii 6.8 Å and 8.0 Å.

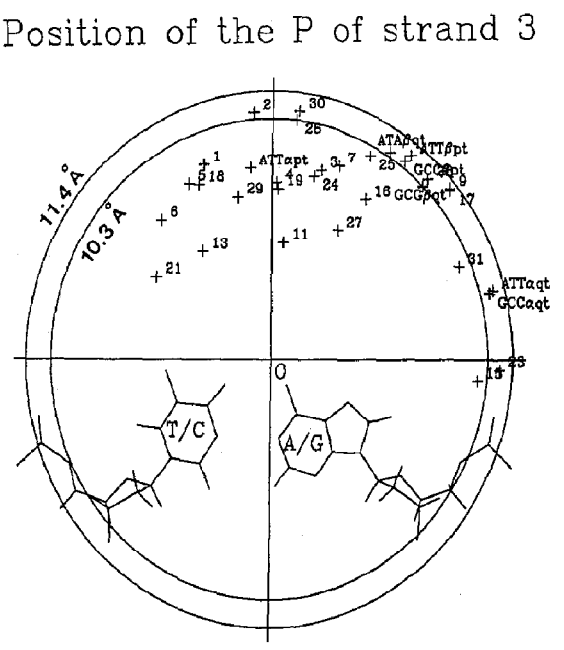


Fig. 11. Representation of the positions of the P atoms with respect to the double strand for all database triplets. Triplets are referred by the number given in Fig. 2. Triplets capable of forming geometrically uniform homogeneous triple helices, AToTαpt, AToTαqt, AToTβpt, GCoC$^+\beta$pt, GCoC$^+\alpha$qt, AToAβqt, GCoGβqt, and triplet GCoC$^+\alpha$pt are referred explicitly by their names. P atoms in all three strands corresponding to all these latter triplets, except AToTαpt and GCoCαpt, are contained within two circles of radii 10.3 Å and 11.4 Å.

mation, and within some limits that remain to be measured, the length of the sugar phosphate chain can be viewed as constant. In a helical geometry, the radii of key atoms such as C1′, P, and O3′ is normally proportional to the length of the sugar phosphate chain per helical turn, therefore these radii should be nearly identical as we observe it. These observations suggest that two factors at least are important for building geometrically uniform triple helices: isomorphism and similitude of helical radii for key atoms like C1′, P, and O3′ within limits given above.

4.3. Stability of homogeneous triple helices by molecular dynamics

A necessary condition for the structural validity of our triple helices, is that they remain stable

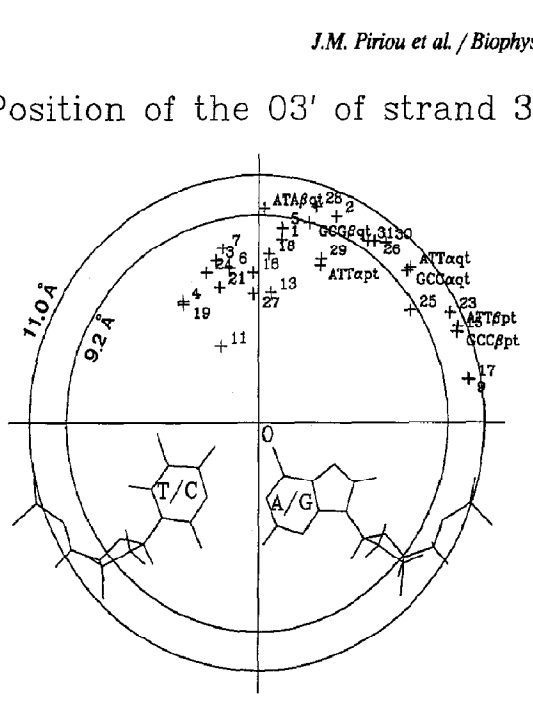


Fig. 12. Representation of the positions of the O3′ atoms with respect to the double strand for all database triplets. Triplets are referred by the number given in Fig. 2. Triplets capable of forming geometrically uniform homogeneous triple helices, AToTαpt, AToTαqt, AToTβpt, GCoC$^+$βpt, GCoC$^+$αqt, AToAβqt, GCoGβqt, and triplet GCoC$^+$αpt are referred explicitly by their names. O3′ atoms in all three strands corresponding to all these latter triplets, except AToTαpt and GCoCαpt, are contained within two circles of radii 9.2 Å and 11.0 Å.

when MD runs are performed on them. This is actually the case for the AToTβpt, GCoGβqt and alternate GCoGβqt/AToAβqt heptamers and to a lesser degree for the GCoC$^+$βpt triple helix, as far as base pairing, the torsion angles along the sugar-phosphate backbone, the rms deviation and the helical parameters are concerned. The stability depends on many subtle factors. It decreases when the size of the fragments grows or when complex disruption is not balanced by the introduction of ad hoc forces to reinforce the hydrogen bonds of the four extreme triplets. These additional forces do not seem to perturb the structures. The fluctuations that we observed are smaller than the fluctuations obtained with dynamics performed in the explicit presence of water molecules and counterions [32]. Explicit treatment of solvent should affect the fluctuations in two ways. On one hand, the motions should be slowed down, just because of viscosity. Helix bending must be simultaneous with the motions of solvent molecules. On the other hand, the average conformation of the triplex depends on the distribution of the explicit solvent charges. As this distribution varies, the conformation of the triplex changes and this make the fluctuations increase.

Most torsion angles of the sugar-phosphate backbone kept their starting values during the runs. The transition (g−, g+) to (g+/t, t) of the α and γ torsion angles which had been already described [32] has also here been observed in the purine Watson–Crick strand of purine-pyrimidine-pyrimidine triplexes. We could give a linear relationship between α and γ. As far as ϵ and ζ of the third stand are concerned, the B I to B II transition in the GCoGβqt triplet is new. The fact that this feature is preserved in alternate sequence is most interesting.

The values of all the helical parameters for AToTβpt, GCoGβqt and alternate GCoGβqt/AToAβqt are not significantly different from those of the starting structures. Moreover the values of the fluctuations suggest that the variations of the helical parameters are in an acceptable range. The values of the helical twist for the third strand may appear relatively high (up to 36°, for GCoGβqt); nevertheless, one must bear in mind that the database triplets were built by adjusting the mean plane of the third base in the XOY plane, thus encouraging those high values. From the MD runs of the triplexes, we could estimate the fluctuations of the Watson–Crick helical twist in the range 3° to 5°, depending on the sequence, which seems reasonable when compared to the experimental value, concerning duplexes: 4° [73]. The fluctuations in roll and tilt for the first two strands in the triplets compare well with a persistence length of 500 Å [74].

5. Conclusion

In this work, we have constructed a complete structural database of 32 triplets that encompasses all triplets AToT, GCoC$^+$, AToA, and GCoG where the third strand is bound to the

purine strand through two hydrogen bonds. The third strand can be geometrically characterized by the extension of the definition of helical parameters in agreement with the convention of the EMBO workshop. This database was constructed upon three simple building rules based on helical symmetry, base pairing and torsion angles and its validity has been checked using MD runs in vacuo. The search of geometrically uniform homogeneous triple helices has led us to recover structural results established for triplets AToT and $GCoC^+$, and to propose structures for GCoG and GCoG/AToA triple helices. We have shown that GCoG is most likely in βqt conformation and could have a B II conformation. Our results suggest that isomorphism and helical radii are important concepts to understand triple helices. A triplet by triplet study is all the more interesting since we could show in some cases that the triplets kept their individual properties during the MD runs.

Acknowledgment

We are most grateful to M. Leminor (INSERM) for access to the VAX 8530 computer. MLB is a recipient of grants from Ligue Nationale contre le Cancer, Association pour la Recherche contre le Cancer and Université Pierre et Marie Curie. JMP was supported by a grant from Association pour le Développement de la Formation par la Recherche Biomédicale (ADFRB). The authors thank Dr. W. Guschlbauer and Dr. G.V. Fazakerley for helpful discussions and Dr. L. Deleu for his help in the starting phase of the work.

References

[1] G. Felsenfeld, D.R. Davies and A. Rich, J. Am. Chem. Soc. 79 (1957) 2023–2024.
[2] G. Felsenfeld and H.T. Miles, Ann. Rev. Biochem. 36 (1967) 407–448.
[3] A.M. Michelson, J. Massoulié and W. Guschlbauer, Prog. Nucl. Acid Res. 6 (1967) 83–141.
[4] R.D. Wells, D.A. Collier, J.C. Hanvey, M. Shimizu and F. Wohlrab, FASEB J. 2 (1988) 2939–2949.
[5] C. Hélène and J.J. Toulmé, Biochim. Biophys. Acta 1049 (1990) 99–125.
[6] A.S. Moffat, Science 252 (1991) 1375–1376.
[7] H.E. Moser and P.B. Dervan, Science 238 (1987) 645–650.
[8] S.A. Strobel, H.E. Moser and P.B. Dervan, J. Am. Chem. Soc. 110 (1988) 7927–7929.
[9] S.A. Strobel, L.A. Doucette-Stamm, L. Riba, D.E. Housman and P.D. Dervan, Science 254 (1991) 1639–1642.
[10] B.M.J. Revet, E.P. Sena and D.A. Zarling, J. Mol. Biol. 232 (1993) 779–791.
[11] L.J. Ferrin and R.D. Camerini-Otero, Science 254 (1991) 1494–1497.
[12] D. Praseuth, L. Perrouault, T. Le Doan, M. Chassignol, N. Thuong and C. Hélène, Proc. Natl. Acad. Sci. USA (1990) 85, 1349–1353.
[13] M. Cooney, G. Czernuszewicz, E.H. Postel, S. Flint and M.E. Hogan, Science (1988) 241 456–459.
[14] W.M. McShan, R.D. Rossen, A.H. Laughter, J. Trial, D.J. Kessler, J.G. Zendegui, M.E. Hogan and F. Orson J. Biol. Chem. (1992) 267 5712–5721.
[15] K. Hoogsteen, Acta Cryst. 12 (1959) 822–823.
[16] K. Hoogsteen, Acta Cryst. 16 (1963) 907–916.
[17] S. Arnott and E. Selsing, J. Mol. Biol. 88 (1974) 509–521.
[18] S. Arnott, P.J. Bond, E. Selsing and P.J. Smith, Nucl. Acids Res. 10 (1976) 2759–2770.
[19] S.H. Kim, F.L. Suddath, G.J. Quigley, A. McPherson, J.L. Sussman, A.J.H. Wang, N.C. Seeman and A. Rich, Science 185 (1974) 435–440.
[20] J.D. Robertus, J.E. Ladner, J.T. Finch, D. Rhodes, R.S. Brown, B.F.C. Clark and A. Klug, Nature 250 (1974) 546–551.
[21] D. Thiele and W. Guschlbauer, Biopolymers 8 (1969) 361–378.
[22] C. Mark and D. Thiele, Nucl. Acids Res. 5 (1978) 1017–1028.
[23] P.A. Beal and P.B. Dervan, Science 251 (1991) 1360–1363.
[24] J.S. Sun, C. Giovannangeli, J.C. Francois, R. Kurfurst, T. Montenay-Garestier, U. Asseline, T. Saison-Behmoaras, N.T. Thuong and C. Hélène, Proc. Natl. Acad. Sci. USA 88 (1991) 6023–6027.
[25] R.H. Durland, D.J. Kessler, M. Duvic and M.E. Hogan in: Molecular basis of specificity in nucleic acid–drug interactions, eds. B. Pullman and J. Jortner (Kluwer, Dordrecht, 1990) pp. 565–578.
[26] J. Bernués, R. Beltrán, J.M. Casasnovas and F. Azorín, Nucl. Acids Res. 18 (1990) 4067–4073.
[27] R.F. Macaya, P. Schultze and J. Feigon J. Am. Chem. Soc. 114 (1992) 781–783.
[28] M. Ouali, R. Letellier, F. Adnet, J. Liquier, J.S. Sun, R. Lavery and E. Taillandier, Biochemistry 32 (1993) 2098–2103.
[29] G. Raghunathan, H.T. Miles and V. Sasisekharan, Biochemistry 32 (1993) 455–462.
[30] C.A. Laughton and S. Neidle, Nucl. Acids Res. 20 (1992) 6535–6541.
[31] H.W.Th. van Vlijmen, G. Ramé, and B.M. Pettitt, Biopolymers 30 (1990) 517–532.

[32] C.A. Laughton and S. Neidle, J. Mol. Biol. 223 (1992) 519–529.
[33] J.C. Hanvey, M. Shimizu and R.D. Wells, J. Biol. Chem. 264 (1989) 5950–5956.
[34] L.C. Griffin and P.B. Dervan, Science 245 (1989) 967–971.
[35] L.E. Xodo, G. Manzini and F. Quadrifoglio, Nucl. Acids Res. 18 (1990) 3557–3564.
[36] M.D. Distefano, J.A. Shin and P.B. Dervan, J. Am. Chem. Soc 113 (1990) 5901–5902.
[37] J.L. Mergny, J.S. Sun, M. Rougée, T. Montenay-Garestier, F. Barcelo, J. Chomillier and C. Hélène, Biochemistry 30 (1991) 9791–9798.
[38] R.F. Macaya, D.E. Gilbert, S. Malek, J.S. Sinsheimer and J. Feigon, Science 254 (1991) 270–274.
[39] H.U. Stilz and P.B. Dervan, Biochemistry 32 (1993) 2177–2185.
[40] D.A. Horne and P.B. Dervan, J. Am. Chem. Soc. 112 (1990) 2435–2437.
[41] M. Riordan and J.C. Martin, Nature 350 (1990) 442–443.
[42] S. McCurdy, C. Moulds and B. Froehler, Nucleosides Nucleotides 10 (1990) 287–290.
[43] S.D. Jayasena and B.H. Johnston, Biochemistry 32 (1993) 2800–2807.
[44] C.M. Radding, Ann. Rev. Biochem. 47 (1978) 847–880.
[45] E.H. Egelman, Cur. Opin. Struct. Biol. 3 (1993) 189–197.
[46] F.B. Howard, H.T. Miles, K. Liu, J. Frazier, G. Raghunathan and V. Sasisekharan, Biochemistry 31 (1992) 10671–10677.
[47] M. Le Bret, J. Gabarro-Arpa, J. Ch. Gilbert and Cl. Lemaréchal, J. Chim. Phys. Phys.-Chim. Biol. 88 (1991) 2489–2496.
[48] J. Gabarro-Arpa, J.A.H. Cognet and M. Le Bret, J. Mol. Graph. 10 (1992) 166–173.
[49] Y.-K. Cheng and B.M. Pettitt, Prog. Biophys. Mol. Biol. 58 (1992) 225–257.
[50] A.E.V. Haschemeyer and H.M. Sobell, Acta Cryst. 18 (1965) 525–532.
[51] R.H. Durland, D.J. Kessler, S. Gunnell, M. Duvic, B.M. Pettitt and M.E. Hogan, Biochemistry 30 (1991) 9246–9255.
[52] R.E. Dickerson et al., EMBO J. 8 (1989) 1–4.
[53] R. Lavery and H. Sklenar, J. Biomol. Struct. Dyn. 6 (1989) 65–667.
[54] E. von Kitzing and S. Diekmann, Eur. Biophys. J. 15 (1987) 13–26.
[55] S. Arnott, P. Campbell-Smith and R. Chandrasekaran, in Handbook of Biochemistry and Molecular Biology, ed. G.D. Fasman, Vol. 2, 3rd Ed., (CRC Press, Boca Raton, 1976) pp. 411–422.
[56] J. Liquier, P. Coffinier, M. Firon and E. Taillandier, J. Biomol. Struct. Dyn. 9 (1991) 437–445.
[57] S.S. Blommers, M.M.W. Mooren, D.E. Pulleyblank, M.M.J. Wijmenga and C.W. Hilbers, Nucl. Acids Res. 18 (1990) 6523–6529.
[58] P.K. Weiner and P.A. Kollman, J. Comput. Chem. 2 (1981) 287–303.
[59] S.J. Weiner, P.A. Kollman, D.T. Nguyen and D.A. Case, J. Comput. Chem. 7 (1986) 230–252.
[60] U.C. Singh, P.K. Weiner, J.W. Caldwell and P.A. Kollman, AMBER 3.0. University of California, San Francisco (1986).
[61] B. Gelin and M. Karplus, Biochemistry 18 (1979) 1256–1268.
[62] R.E. Dickerson, M. Kopka and P. Pjura, in: Biological macromolecules and assemblies, Vol. 2. Nucleic acids and interactive proteins, eds. F.A. Jurnak and A. McPherson (Wiley, New York, 1985).
[63] C. Altona and M. Sundaralingam, J. Am. Chem. Soc. 94 (1972) 8205–8212.
[64] L. Verlet, Phys. Rev. 159 (1967) 98–103.
[65] J.P. Ryckaert, G. Cicotti and H.J.C. Berendsen, J. Comput. Phys. 23 (1977) 327–341.
[66] U.C. Singh, S.J. Weiner and P.A. Kollman, Proc. Natl. Acad. Sci. USA 82 (1985) 755–759.
[67] J.A.H. Cognet, J. Gabarro-Arpa, Ph. Cuniasse, G.V. Fazakerley and M. Le Bret, J. Biomol. Struct. Dyn. 7 (1990) 1095–1115.
[68] F.M. Chen, Biochemistry 30 (1991) 4472–4479.
[69] D.S. Pilch, C. Levenson and R.H. Shafer, Proc. Natl. Acad. Sci. USA 87 (1990) 1942–1946.
[70] J.S. Sun, T. de Bizemont, G. Duval-Valentin, T. Montenay-Garestier and C. Hélène, C.R. Acad. Sci. Paris, Série III 313 (1991) 585–590.
[71] G. Gupta, M. Bansal and V. Sasisekharan, Proc. Natl. Acad. Sci. USA 77 (1980) 6468–6470.
[72] C.R. Cantor and P.R. Schimmel, in: Biophysical chemistry (Freeman, San Francisco, 1980).
[73] D. Shore and R.L. Baldwin, J. Mol. Biol. 170 (1983) 983–1007.
[74] H. Eisenberg, Accounts Chem. Res. 20 (1987) 276–282.