NMR Solution Structure of the Duplex Formed by Self-Pairing of α-L-Arabinopyranosyl-(4′⟶2′)-(CGAATTCG)

by **Marc-Olivier Ebert**[a][1], **Anatol Luther**[b][2], **Hoan K. Huynh**[b][3], **Ramanarayanan Krishnamurthy**[b], **Albert Eschenmoser**[a][b], and **Bernhard Jaun***[a]

[a]) Laboratorium für Organische Chemie, ETH-Hönggerberg, CH-8093 Zürich (fax: ++41-1-632-1475; e-mail: jaun@org.chem.ethz.ch)

[b]) The Skaggs Institute for Chemical Biology at The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

Dedicated to Professor *Dieter Seebach* on the occasion of his 65th birthday

The solution structure of the duplex formed by α-L-arabinopyranosyl-(4′→2′)-(CGAATTCG) was studied by NMR. The resonances of all H-, P- and most C-atoms could be assigned. Dihedral angles and distance estimates derived from coupling constants and NOESY spectra were used as restraints in a simulated annealing calculation, which generated a well-defined bundle of structures for the six innermost nucleotide pairs. The essential features of the resulting structures are an antiparallel, *Watson–Crick*-paired duplex with a strong backbone inclination of *ca.* −50° and, therefore, predominant interstrand base stacking. The very similar inclination and rise parameters of arabinopyranosyl-(4′→2′)-oligonucleotides and p-RNA explain why these two pentapyranosyl isomers are able to cross-pair.

Introduction. – We have reported earlier that 4′⟶2′ phosphodiester linked β-D-ribo-, β-D-xylo-, α-L-lyxo-, and α-L-arabinopyranosyl oligonucleotides (see *Fig. 1*) constitute base-pairing systems that form antiparallel duplexes according to the *Watson–Crick*-pairing mode [1]. All four isomers pair more strongly than RNA. While all members of this group can efficiently cross-pair with each other, none of them is able to communicate by base-pairing with either RNA or DNA [1b].

For β-D-ribopyranosyl-nucleic acid (p-RNA), the first isomer of the pentopyranosyl family studied experimentally, this pairing propensity had been predicted by qualitative conformational analysis on the basis of idealized pyranose chairs exhibiting cyclohexane geometry and backbone bonds of equal length [1c][2]. A comprehensive NMR study of a p-RNA octamer duplex confirmed the essential structure predictions, such as the nearly linear conformation and the very pronounced backbone inclination with predominant interstrand stacking (as opposed to intrastrand stacking found in DNA and RNA) [3]. The NMR-derived structure showed significant deviations of the backbone angles from the idealized model, which were interpreted in terms of adjustments needed to reduce the stacking distance (rise) and reach optimal purine-pyrimidine and purine-purine interstrand stacking.

---

[1]) Part of the planned Ph.D. thesis of *M.-O. E.*

[2]) Current address: *Polyphor*, CH-Basel.

[3]) Current address: *Astra Zeneca R&D Boston, Inc.*, Waltham, MA, USA.

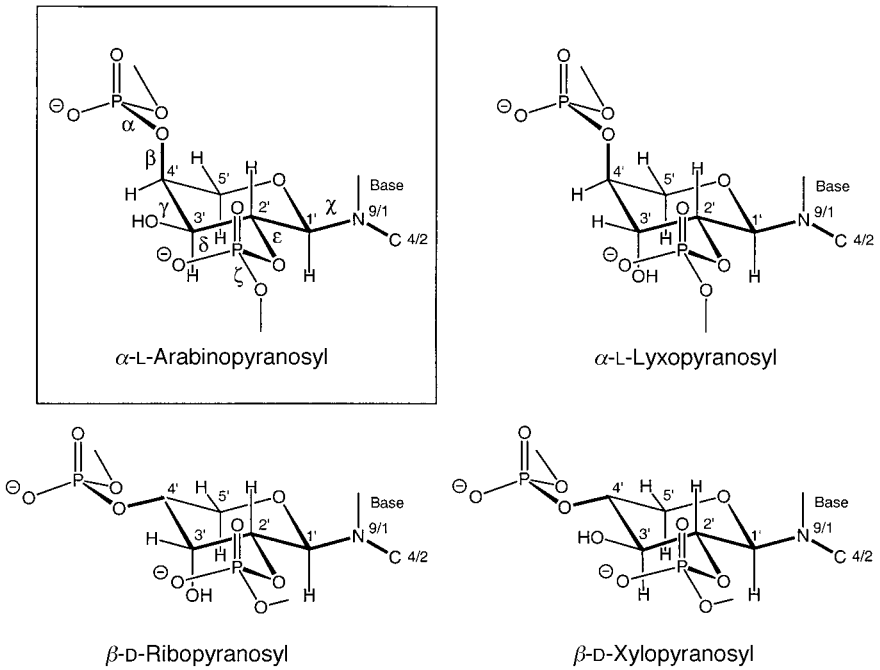

Fig. 1. *The four (4′⟶2′)-pentopyranose systems studied.* Atom numbering and definition of torsional angles were chosen in analogy to natural systems [20] and are shown in the formula of *α*-L-*arabinopyranosyl-(4′⟶2′)-oligonucleotides.* Backbone: $-O2'_{i-1}-P_i-O4'_i-C(4')_i-C(3')_i-C(2')_i-O2'_i-P_{i+1}$. Purine base: C(4)–N(9)–C(1′)–O5′–; pyrimidine base: –C(2)–N(1)–C(1′)–O5′–. Angles *α*, *β*, *γ*, *δ*, *ε*, *ζ*, and *χ* as indicated.

For *α*-arabinopyranosyl-(4′⟶2′)-oligonucleotides (apNA), qualitative conformational analysis predicts severe steric hindrance between the 3′-hydroxy- and the 2′-phosphodiester groups in the pairing conformation. We, therefore, expected that, for this stereoisomer, preorganization of the single strand towards the quasilinear conformation necessary for pairing would be severely hampered. Hence, the experimental result that *α*-arabinopyranosyl-NA is, in fact, one of the strongest oligonucleotide pairing systems known [1a] came as a surprise and prompted us to undertake a full NMR analysis of the duplex derived from *α*-L-arabinopyranosyl-(4′⟶2′)-(CGAATTCG) (**1**) having the same self-complementary base sequence studied earlier for p-RNA.

Here, we report the results of this study and discuss the structural factors of the experimentally determined conformation that may contribute to the extraordinary pairing strength of *α*-arabinopyranosyl-(4′⟶2′)-oligonucleotides and to their ability to cross-pair with the other three pentapyranosyl systems.

Results. – *Assignment of Nonexchangeable Nuclei.* The residue-specific assignment of the non-exchangeable protons was achieved in a standard way through a combination of COSY, TOCSY, and NOESY spectra measured in D_2O at 26.7°. Most

of the sugar protons could be assigned through COSY and TOCSY. The H–C(5′) protons were correlated to the rest of the sugar spin system by means of the H–C(1′)/H_{ax}–C(5′) NOEs. The C-atoms, H–C(6), and H–C(8) protons of the bases were assigned from HMBC and HSQC. Sequential connectivities were established by ^{1}H,^{31}P-COSY and were consistent with the NOESY data. Lists of all ^{1}H and ^{31}P resonances as well as the assignment of the ^{13}C resonances based on HSQC/HMBC are given in *Tables 1* and *2*, respectively.

Table 1. *^{1}H and ^{31}P-NMR Chemical Shifts* [ppm] *and Assignments for* **1** *at 26.7°*

	H–C(1′)	H–C(2′)	H–C(3′)	H–C(4′)	H_{eq}–C(5′)	H_{ax}–C(5′)
apC1	5.75	4.72	4.01	4.13	4.03	3.88
apG2	5.26	4.55	3.95	4.55	4.24	3.75
apA3	5.34	4.46	3.89	4.43	3.88	3.67
apA4	5.19	4.21	3.83	4.33	3.64	3.50
apT5	5.28	3.97	3.78	4.25	3.67	3.47
apT6	5.38	4.29	3.86	4.43	4.26	3.68
apC7	5.41	4.28	3.83	4.44	4.10	3.66
apG8	5.02	3.78	3.71	4.38	4.09	3.57
	H–C(6)[a])	H–C(5)[b])	H–C(2)	H–N(1)[c])	NH_2[d])	^{31}P[e])
apC1	7.87	6.26			7.52, 8.16*	
apG2	7.89			12.31	6.76	0.19
apA3	8.11		7.83		7.04	0.19
apA4	8.02		8.11		6.94, 7.68*	0.13
apT5	7.13	2.01		14.53		0.46
apT6	7.38	1.97		13.59		0.91
apC7	7.29	5.93			7.04, 8.17*	0.42
apG8	7.73			12.84 (br.)	6.62	0.13

[a]) Pyrimidine; corresponding purine proton = H–C(8). [b]) Cytidine: corresponding thymidine protons = Me–C(5). [c]) Purine; corresponding pyrimidine proton = H–N(3). [d]) H-Bonded NH_2 protons labeled with *. [e]) The ^{31}P resonances are listed in the 4′ → 2′ direction; chemical shifts in ppm *vs.* external 85% H_3PO_4.

Assignment of Exchangeable Protons and Proof of Pairing Mode. The ^{1}H-spectrum in H_2O/D_2O 9:1 at 26.7° showed three sharp signals in the imino region (*Fig. 2*). When the temperature was lowered to 11.4°, an additional, broader signal appeared. Two of the four signals could be assigned to the imino protons of apT5 and apT6 by means of the H–N(3)/Me NOEs. The third sharp signal is expected to belong to H–N(1) of apG2, which was confirmed by the presence of H–N(3)(G)/H–N(4)(C) interstrand NOEs typically observed for *Watson–Crick* GC base pairs [4] (see *Fig. 3*). The amino resonances of the pairing partner apC7 as well as those of apC1 could be assigned from the strong H–N(4)/H–C(5) intraresidual NOEs. The assignment of the amino groups of apA3 and apA4 was also established *via* their pairing partners as described above. The amino signal of apG2 was assigned on the basis of the intraresidual NOE to the already known imino proton and the interstrand H–N(2)(apG2)/H–N(4)(apC7) correlation. The assignment of the remaining signal in the amino region to H–N(2)(apG8) could be confirmed by inter- and intrastrand NOEs to nonexchangeable protons. *Together, these findings demonstrate the formation of eight*

Table 2. *^{13}C-NMR Chemical Shifts* [ppm] *and Assignments for* **1** *at 26.7°*

	C(1′)	C(2′)	C(3′)	C(4′)	C(5′)	C(2)
apC1	85.3	75.9	75.7	71.3	72.1	160.7
apG2	83.4	78.6	74.9	76.4	70.2	157.2/156.4[b])
apA3	83.5	78.0	74.9	–[a])	70.1	154.7
apA4	83.4	78.3	–	75.6	69.8	155.8
apT5	83.8	76.9	74.7	75.4	69.9	153.4
apT6	83.9	77.0	74.7	–	69.7	152.9
apC7	84.6	76.8	–	–	70.1	159.1
apG8	84.6	73.0	75.1	76.3	69.8	157.2/156.4[b])
	C(4)	C(5)	C(6)	C(8)	Me–C(5)	
apC1	168.7	100.4	144.3			
apG2	152.9	117.2	162.1/161.1[b])	139.0		
apA3	150.9	120.4	157.4	142.0		
apA4	150.7	119.4	157.8	142.6		
apT5	170.3	114.2	140.3		15.4	
apT6	170.0	114.2	140.7		14.8	
apC7	167.9	100.4	143.7			
apG8	153.9	116.9	162.1/161.1[b])	139.4		

[a]) No assignment possible. [b]) Assigned as a pair; C(2) resonates at lower field [19].

base pairs in an antiparallel duplex pairing in the Watson–Crick *mode and exhibiting* C_2*-symmetry on the NMR time scale.*[4])

Conformation of the Backbone. The six-membered rings of the arabinopyranoses adopt a chair conformation. The large values for ^3J(1′,2′)' and ^3J(2′,3′) of *ca.* 9 Hz (*Table 3*) prove that these three protons occupy axial positions in the ring. The values for ^3J(3′,4′) of about 4 Hz (equatorial position of H–C(4′)) as well as the small values of ^3J(4′,5′) determined are also in agreement with a chair conformation. Further evidence comes from the strong NOEs observed between the three axial protons H–C(1′), H–C(3′), and H_{ax}–C(5′).

In the ^{1}H,^{31}P-COSY spectrum, four-bond cross-peaks between H–C(3′) and PO4′ were observed for residues apG2–apG8 (see *Fig. 4*). For residues apA4, apT6, apC7, and apG8 the ^4J(PO4′, H–C(3′)) coupling constants could be determined from the E.COSY pattern of the H–C(3′)/H–C(4′) DQF.COSY cross-peaks and have values of *ca.* 3 Hz. Since all PO4′/H–C(3′) cross-peaks in the ^{1}H,^{31}P-COSY spectrum show practically the same intensity, the three ^4J(PO4′,H–C(3′)) coupling constants that could not be determined directly must also have values of *ca.* 3 Hz. Observation of a W-type coupling of this magnitude is only possible with a nearly synplanar conformation of H–C(3′)–C(4′)–O4′–P [5].

The same procedure – direct determination of a subset from DQF.COSY and extrapolation based on cross-peak intensities in the H,P-COSY spectrum – was used to show that all vicinal ^3J(PO4′,H–C(4′)) coupling constants have values of 9–10 Hz.

---

[4]) Due to this twofold symmetry, residues are labeled apC1 through apG8 (each label corresponding to two symmetry-equivalent residues) for the discussion of the NMR data, whereas, for the structure calculations, the 16 residues are treated individually.

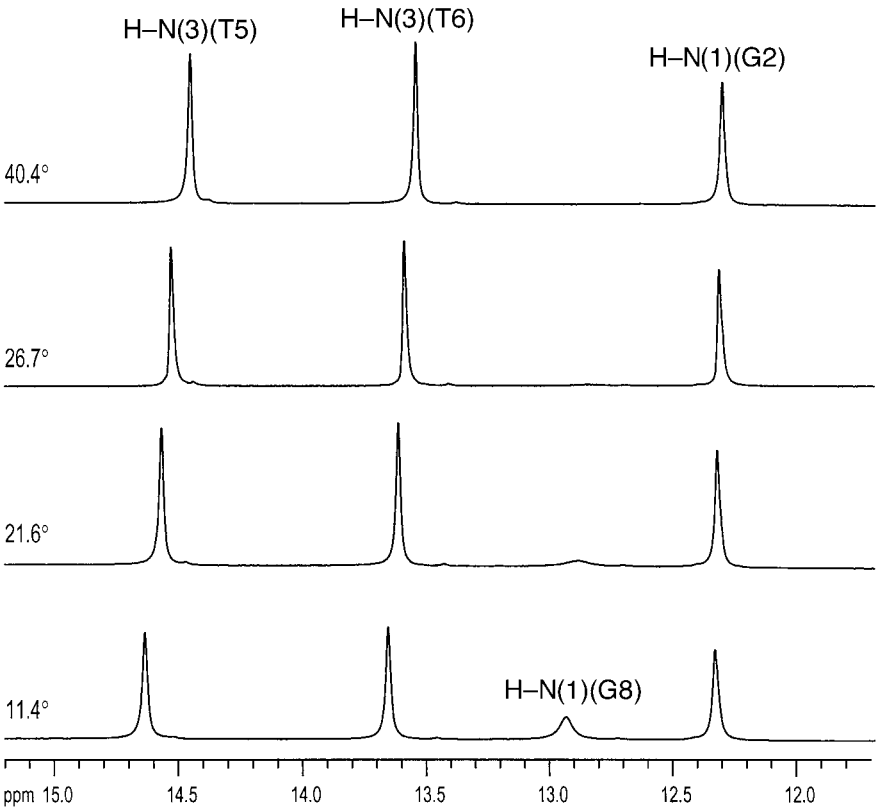

Fig. 2. *Imino-region of the ^{1}H-NMR spectrum of* **1** *in H_2O/D_2O 9:1, 50 mM sodium arsenate, at different temperatures* (for conditions: see *Exper. Part*)

Table 3. *Coupling Constants* [Hz] *of* **1** *at 26.7°* [a])

	$^3J(1',2')$	$^3J(2',3')$[b])	$^3J(3',4')$[b])	$^3J(4',5_{eq}')$[b])	$^3J(4',5_{ax}')$[b])	$^3J(2',PO2')$[c])	$^3J(4',PO4')$[d])	$^4J(3',PO4')$[d])
apC1	9.2	9.2	4.4	3.6	–[e])	8.7		
apG2	9.4	9.3	4.2	–	2.1	8	–	–
apA3	9.0	9.2	4.0	–	–	8.1	–	–
apA4	8.9	9.2	4.2	–	3.0	7.8	9.6	3.5
apT5	9.3	9.3	3.8	–	–	7.3	–	–
apT6	9.2	9.3	3.2	–	–	8.4	9.5	3.3
apC7	9.1	9.1	3.3	4.6	1.8	8.5	9.8	3.3
apG8	8.8	9.7	3.6	2.3	2.6		9.1	3.3

[a]) Estimated error $\leq$ 1.5 Hz. [b]) Determined by fitting the appropriate number of *Gauss* lines to traces of the COSY spectrum. [c]) Determined by fitting the appropriate number of *Gauss* lines to traces of the COSY spectrum with/without {^{31}P} decoupling. [d]) Determined by analysis of the E.COSY pattern present in the cross-peaks correlating H–C(3′) and H–C(4′). [e]) Not determined.

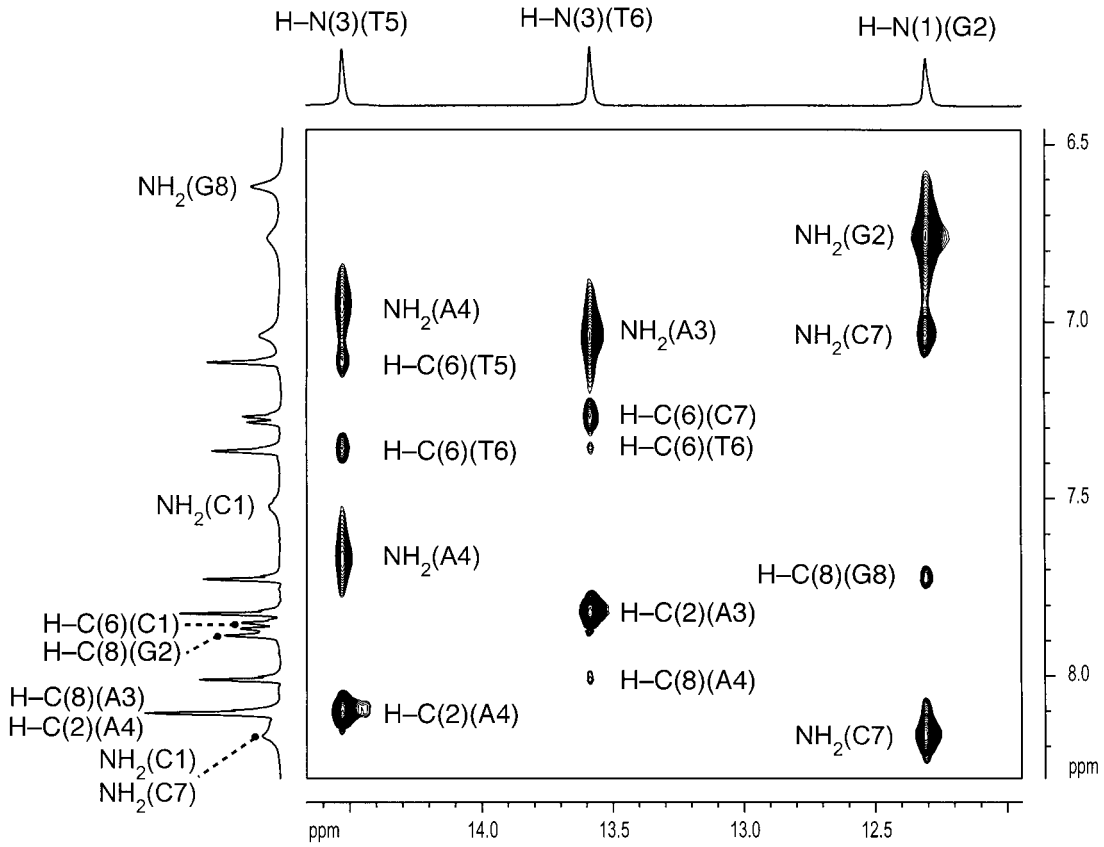

Fig. 3. *Expansion of the NOESY spectrum* ($t_m = 150$ ms) *of* **1** *in* H_2O/D_2O *9:1, containing cross-peaks between imino and amino/aromatic protons* (for conditions, see *Exper. Part*)

This together with the observed four-bond coupling allows us to conservatively restrain the exocyclic torsion angle β to $-150 \pm 30°$.

The second exocyclic torsion angle ε is not unambiguously determined by the value of *ca.* 8 Hz measured for ^3J(PO2′,H–C(2′)), which is consistent with $\varepsilon = 0° \pm 10°$, $120 \pm 10°$, $-150 \pm 15°$, or $-90 \pm 15°$. Additional knowledge of ^3J(PO2′,C(1′)) and/or ^3J(PO2′,C(3′)), however, would allow us to find the correct solution for ε. Because of severe overlap, none of the ^3J(PO2′,C) values could be determined directly from the proton decoupled ^{13}C-spectrum. Therefore, determination of the ^3J(PO2′,C(1′) coupling constants from ^{13}C-traces of HSQC spectra with and without ^{31}P-decoupling during t_1 was tried [6]. Unfortunately, the uncertainty of this method proved to be too large to allow extraction of reliable values for ^3J(PO2′,C(1′)) in those cases where no peak splitting but only a broadening of the line due to the coupling was observed. The exceptions were the C(1′)-signals of apG2 and apC7, which both consisted of doublets with splitting of *ca.* 8 Hz that collapsed to a singlet under ^{31}P decoupling. If we assume that all residues have values of ε in the same of the four sectors given above, the ^3J(C1′,P2′)

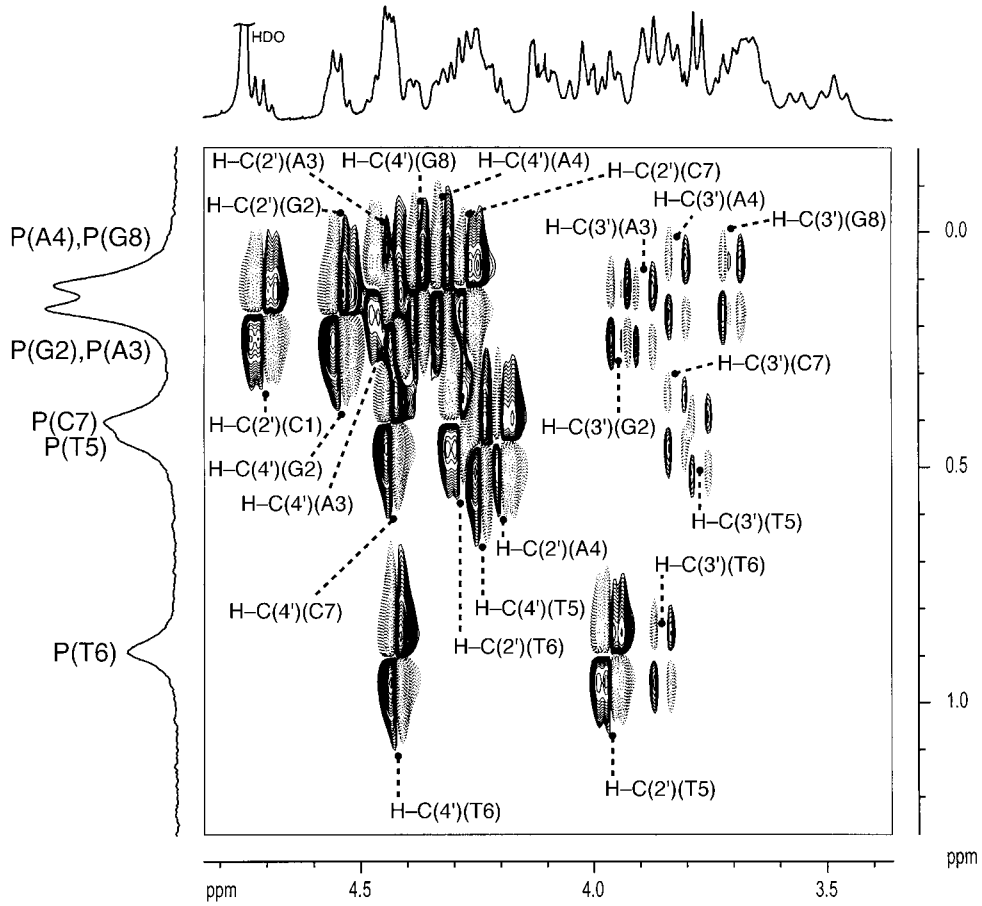

Fig. 4. *H,P-COSY Spectrum of* **1** *in* D_2O *showing the three-bond H–C(2′)/PO2′ and H–C(4′)/PO4′ cross-peaks* (high intensity) *and the four-bond H–C(3′)/PO4′ cross-peaks* (low intensity)

values determined for apG2 and apC7, together with ^3J(PO2′,H–C(2′)) (which is known for residues apC1–apC7) confine ε to a value of $-95 \pm 15°$. The *Karplus* functions used for derivation of dihedral angles from coupling constants [7] and the resulting angles are shown in *Fig. 5*.

The analysis of the glycosidic torsion angle χ was based on cross-peak volumes in the NOESY spectra. The values of 2.3–2.8 Å and 3.8–4.2 Å obtained for the distances between ribose protons at C(2′) and C(1′), respectively, and the pyrimidine/purine H–C(6)/H–C(8) correspond to a value of χ in the anti-domain of *ca.* $-120°$. This is in agreement with the *Watson–Crick*-pairing mode demonstrated through NOEs between exchangeable protons (see above).

Derivation of Distance and Dihedral Angle Restraints from NMR Data. We derived 71 interproton-distance restraints for nonexchangeable protons and 26 distance restraints involving imino protons from the NOESY spectra in D_2O and H_2O/D_2O 9 : 1, respectively. The distances most important for the definition of the global structure

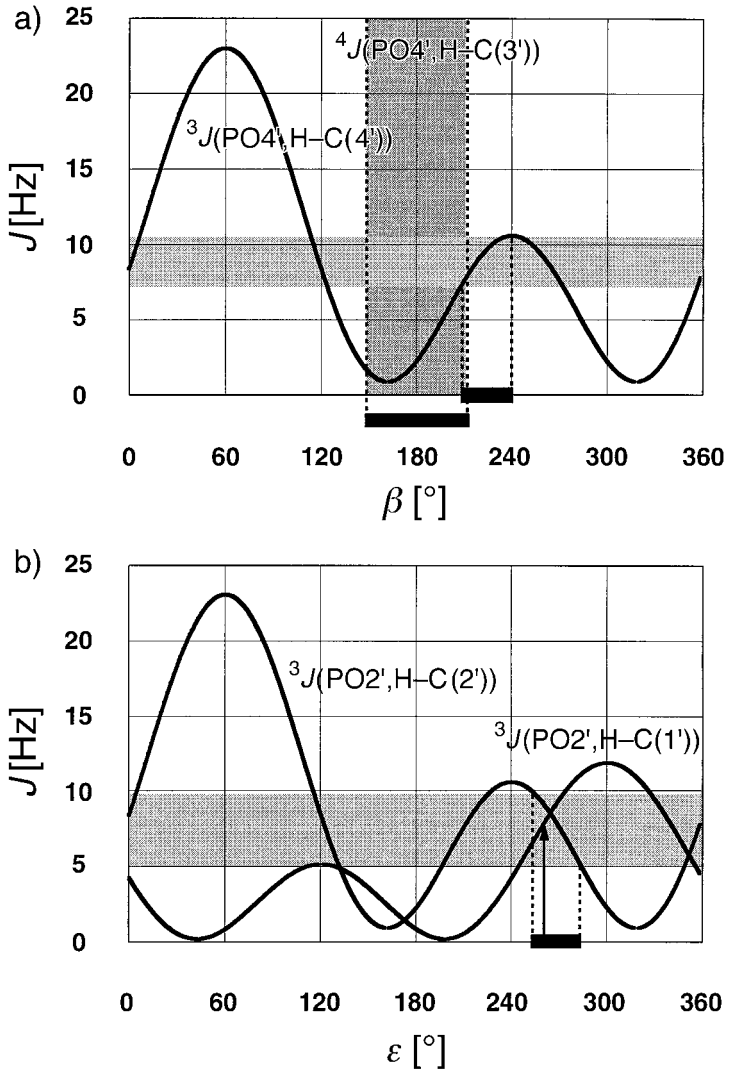

Fig. 5. a) *Derivation of the dihedral angle β from* ^{3,4}J*(H,P) coupling constants.* The horizontal band encompasses the experimental ^{3}J(PO4′,H−C(4)) coupling constants including an estimated error of 1.5 Hz. b) *Derivation of the dihedral angle ε from* ^{3}J*(H,P) and* ^{3}J*(C,P) coupling constants.* Horizontal band as in *a*), bar = region of ε consistent with the ^{3}J(PO2′,H−C(2′)) values and ^{3}J(PO2′,C(2′)) of apG2 and apC7 (arrow). *Karplus* relation for vicinal couplings from [7].

and the backbone inclination of the duplex[5]) are shown schematically in *Fig. 6*. They consist mainly of the sequential and interstrand correlations of the imino protons H−N(1) and H−N(3) and distances involving H−C(2) of the central adenines. With

[5]) The definition of the backbone inclination angle used here is identical to that proposed by *M. Egli* (personal communication), see *Fig. 6*,*a* (see also *Fig. 5* in [3b]).

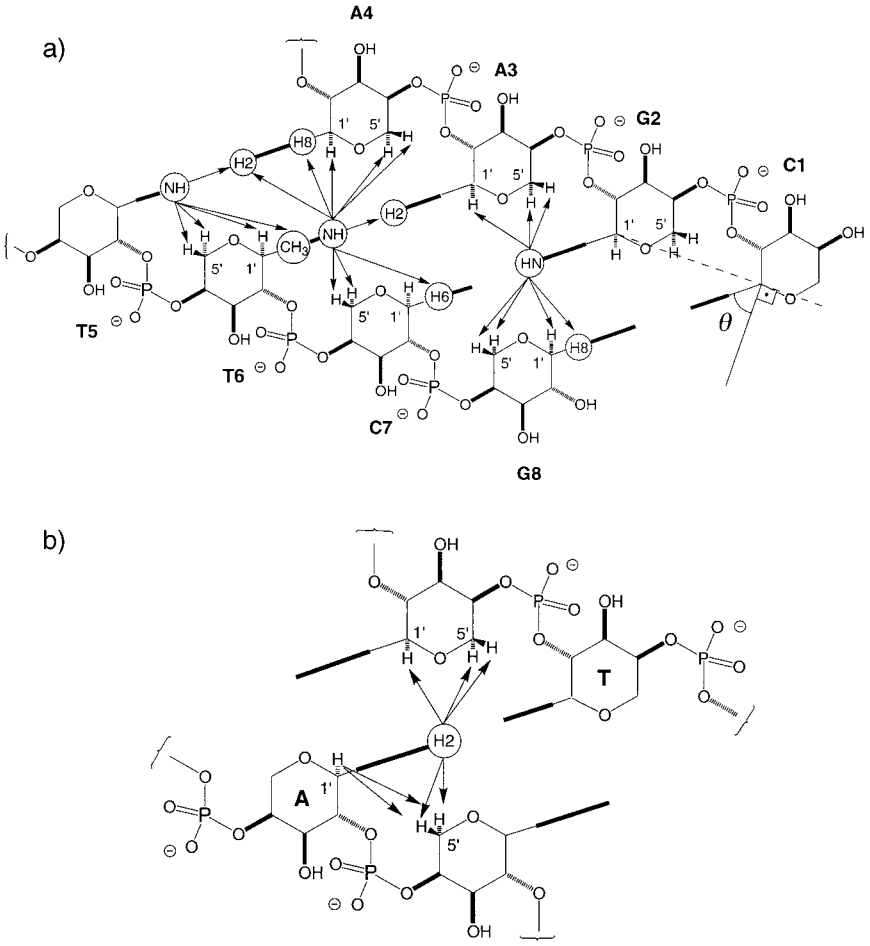

Fig. 6. *Schematic representation of the NOE-derived distances defining the main geometric features of the duplex.* a) *Sequential and interstrand NOEs of the imino protons.* The angle-θ-defining backbone inclination is shown for apC1/apG2. b) *NOEs of nonexchangeable protons.*

the exception of the distances between the pyrimidine/purine protons H–C(6)/H–C(8) and ribose protons H–C(1′) and H–C(2′), intraresidual-distance restraints were not incorporated into the calculation. Instead, the chair conformation of the ring was enforced by dihedral-angle restraints between the axial atoms ($180 \pm 20^\circ$). Since the observation of four imino resonances and their NOEs independently demonstrated the presence of eight *Watson–Crick* base pairs, additional H-bonding restraints corresponding to this pairing mode were introduced and treated like NOE restraints. Judging from the only scarcely intersecting regions found for β based on J(PO4′, H–C(4′)) on the one hand and the observed four-bond coupling between H–C(3′) and PO4′ on the other hand, the values measured for these coupling constants are expected to be biased to some extent by dynamical averaging. This lead us to constrain β and ε rather loosely to $-150 \pm 30^\circ$ and $-90 \pm 30^\circ$, respectively. α, ζ, and χ were left unrestrained.

Structure Calculation with NMR-Derived Restraints. A set of 96 structures was generated by simulated annealing with restrained torsion-angle molecular-dynamics calculations [8]. Each run started from two single strands with newly randomized backbone conformation. To reduce the influence of the particular implementation of electrostatic- and *Lennard–Jones* interactions on the result, the final minimization was carried out with the simplified force field used during annealing [4][9]. From the resulting structures, 46 were selected that had no NOE-distance violation > 0.2 Å, no dihedral-angle violation $> 5°$ and good local geometry in terms of bond lengths, bond angles, and improper torsions (see *Exper. Part* and *Table 4*).

Table 4. *Structure-Determination Statistics for the 46 Converged Structures of* **1** (see *Exper. Part*)

Number of NOE distance restraints[a])	
Intranucleotide	16
Internucleotide	57
Interstrand	24
Base pair	14
Dihedral angle restraints[a])[b])	44
Rmsd of restraints	
Distance restraints [Å]	0.0137 ± 0.0017
Dihedral restraints [°]	0.0326 ± 0.0191
Rmsd from equilibrium geometry of the force field	
Bonds [Å]	0.0040 ± 0.0001
Angles [°]	0.7026 ± 0.0064
Impropers [°]	0.0961 ± 0.0245
Atomic rmsd [Å]	
Residues 2–7[c])	1.12
Residues 3–6	0.64

[a]) According to the C_2 symmetry of the duplex, each entry corresponds to the number of pairs of restraints. [b]) Angles β and ε were restrained for residues 2–7 only. [c]) The superposition was made for the backbone atoms O2′, C(2′), C(3′), C(4′), O4′, and P. The PO4′ of the first selected residue and PO2′ of the last one were excluded from superposition.

Fig. 7,a, shows a bundle of the 15 structures lowest in energy. The superposition, which was made for the backbone atoms of residues 2–7 of each strand, shows a well-defined region of four base pairs in the middle of the duplex while fraying out towards the ends of the strands. For short oligonucleotides, it is generally observed that the terminal base pairs are kinetically more labile than the inner ones [10] and, in our case, this is demonstrated by the exchange broadening of the imino resonance of the outermost base pairs. It should be emphasized, however, that the fraying out seen in *Fig. 7,a*, is not the consequence of higher dynamics but arises because the termini of the duplex are less well-defined by the NMR data.

The picture is mirrored in the variation of the dihedral angles and their standard deviations along the backbone (*Table 5* and *Fig. 8*). The angles of the outermost residues show markedly different mean values and higher standard deviation than in the rest of the molecule. Interestingly, some backbone angles of the inner residues show significant sequence specific variations with rather low standard deviations. In particular, β angles of residues apT6 and apT14 are *ca.* $-135°$ compared to *ca.*

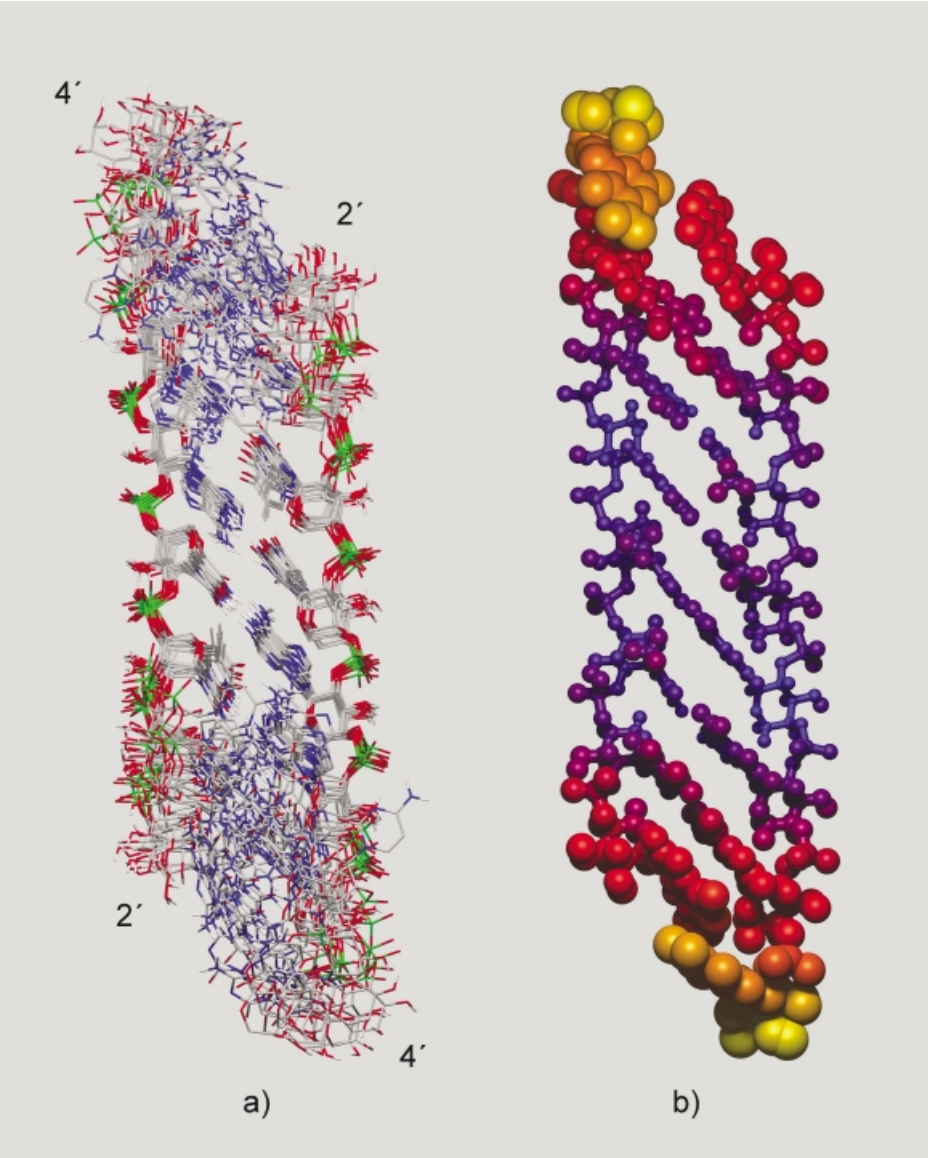

Fig. 7. a) *Superposition of the 15 structures lowest in energy out of a set of 46 accepted structures* (see *Exper. Part*). The superposition is based on the backbone atoms of residues 2–7 of each strand. b) *Representative single structure from the set shown in* a). The rms-deviation of the corresponding atom position in the full set is illustrated by the size of the atoms and by color.

$-155°$ elsewhere, and angles ε of apA4 and apA12 are quite drastically changed from typical values of *ca.* $-110°$ to *ca.* $-73°$. Close inspection reveals that these local variations involve the two phosphodiester groups that flank the thymidines of the apT5 · apA12 and apT13 · apA4 base pairs and are associated with the change from purine · pyrimi-

Table 5. *Statistics of Backbone-Angle Values* (see *Fig. 8*). A: Ensemble discussed in the text (see *Table 4* and *Exper. Part*). B: Ensemble consisting of 47 structures generated and selected by the same protocol and acceptance criteria but with additional restraint on the coplanarity of base pairs ($k_{plan} = 0.5$ kcal/mol). C: Ensemble consisting of 20 structures generated as for B but with $k_{plan} = 1.0$ kcal/mol.

	α			β			ε		
	A	B	C	A	B	C	A	B	C
apC1	–[a])	–	–	–	–	–	−53±51	−55±51	−62±51
apG2	−56±37[b])	−64±40	−61±42	−154±15	−155±15	−156±18	−105±15	−104±14	−105±13
apA3	−34±22	−45±11	−45±11	−153±11	−150±5	−149±6	−118±3	−116±4	−116±4
apA4	−30±7	−26±8	−25±9	−154±3	−157±4	−158±4	−73±11	−69±11	−65±6
apT5	−70±8	−66±8	−68±6	−157±8	−161±6	−163±4	−104±11	−84±14	−82±15
apT6	−65±14	−76±12	−79±11	−135±10	−144±7	−143±6	−103±14	−103±14	−105±13
apC7	−60±11	−62±9	−59±12	−145±8	−144±10	−145±8	−105±16	−102±19	−94±21
apG8	−17±27	−15±29	−23±29	−168±24	−172±22	−171±24	–	–	–
apC9	–	–	–	–	–	–	−44±51	−78±49	−62±52
apG10	−58±41	−51±39	−58±42	−152±18	−157±16	−151±17	−108±13	−107±13	−101±13
apA11	−33±18	−40±18	−50±8	−153±10	−152±11	−148±7	−118±3	−115±4	−115±5
apA12	−28± 8	−28±8	−27±7	−155±3	−157±3	−157±3	−74±13	−68±10	−67±9
apT13	−67±10	−68±7	−66±7	−158±6	−160±7	−162±5	−102±12	−85±14	−81±14
apT14	−67±16	−77±11	−81±12	−136±11	−143±7	−144±6	−100±15	−107±13	−109±13
apC15	−62±10	−59±12	−61±9	−145±9	−143±9	−140±10	−101±18	−101±18	−102±17
apG16	−21±29	−18±26	−19±25	−170±24	−169±23	−168±15	–	–	–

	ζ			χ		
	A	B	C	A	B	C
apC1	174±32	168±36	175±33	−79±27	−82±29	−80±26
apG2	−161±17	−159±8	−159±6	−118±11	−122±8	−123±7
apA3	−148±7	−154±6	−154±7	−106±5	−113±2	−115±2
apA4	−157±10	−163±7	−165±6	−132±3	−132±3	−134±3
apT5	−144±14	−150±12	−151±10	−129±9	−139±4	−140±4
apT6	−147±9	−146±10	−146±10	−130±7	−129±7	−128±7
apC7	−167±21	−171±16	−172±18	−115±8	−120±7	−123±8
apG8	–	–	–	−139±13	−138±12	−144±8
apC9	165±32	176±33	180±42	−74±23	−86±29	−91±27
apG10	−162±14	−161±13	−157±7	−119±10	−123±8	−123±6
apA11	−150±7	−152±6	−153±6	−107±4	−113±2	−114±3
apA12	−160±9	−162±8	−164±6	−131±4	−133±3	−134±3
apT13	−143±16	−150±11	−150±11	−129±8	−139±5	−140±4
apT14	−146±9	−145±9	−145±10	−131±6	−127±7	−126±7
apC15	−169±19	−171±19	−168±13	−117±6	−120±7	−119±6
apG16	–	–	–	−138±13	−142±9	−143±10

[a]) Not defined. [b]) Average and standard deviation.

dine to pyrimidine·purine base pairing in the middle of the sequence (see *Fig. 10,a*). They prevent formation of a gap between the two innermost base-pairs by introduction of a slight kink between the fifth and sixth nucleotide (from the 4′-end) in each strand. This adjustment is particularly remarkable because the two innermost base pairs are the only ones for which the energetic contribution of interstrand stacking is expected to be weak since it involves two pyrimidines. In fact, the stacking distance between the two central base pairs is *ca.* 3.9 Å, whereas, for the outer pairs where purine-pyrimidine interstrand stacking occurs, it is even slightly higher. Generally, these values are too

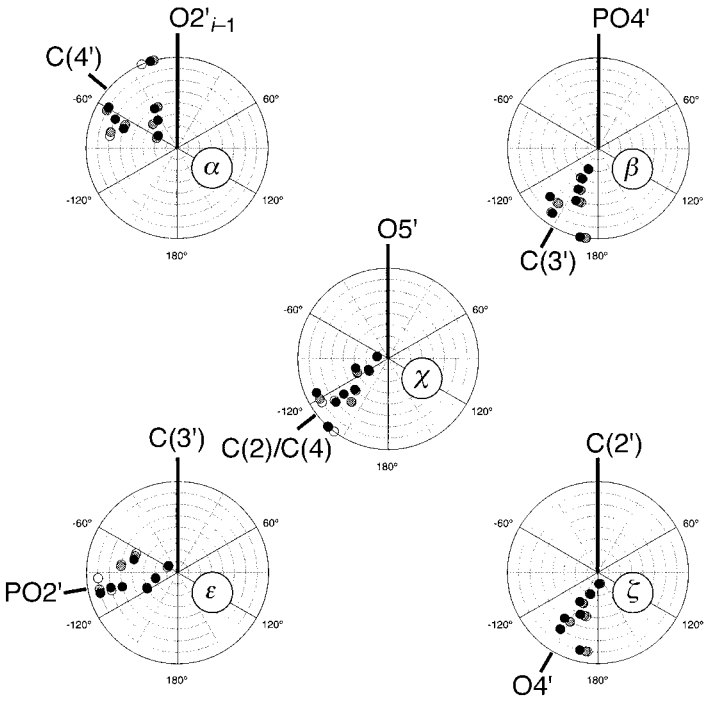

Fig. 8. *Average dihedral backbone angles of all accepted structures.* The values for the individual residues for one strand are shown on concentric circles (innermost = apC1; outermost = apG8). The symmetry equivalent residues in the other strand show values that are within the standard deviations of the displayed values. Different symbols represent calculations with increasing restraints on the coplanarity of the base pairs (black $k_{plan} = 0.0$ kcal/mol; gray: $k_{plan} = 0.5$ kcal/mol; open: $k_{plan} = 1$ kcal/mol).

high if one assumes that the intrastrand-stacking distances of 3.4–3.6 Å typically found in X-ray structures of natural oligonucleotides reflect optimal π–π stacking [11].

Comparison of individual structures from the superposition reveals that the apparent disorder in the terminal regions seen in *Fig. 7,a* is, to a large degree, the result of superimposing structures with widely different overall bending in the central, best-defined region of the backbone. While the major geometric features such as the strong inclination of *ca.* $-50°$[5])(see *Fig. 6*) and dominant interstrand stacking are conserved in all structures (see *Fig. 10,a*), the degree of overall bending of the duplex towards the base-bearing side (convex side, see Footnote 6 in [3a]) show large variation. This can be seen more clearly in *Fig. 9* where the backbone has been replaced by a spline through the C(1′) atoms. The short range nature of NOEs and the accumulation of tolerance limits of local dihedral-angle restraints along the strand make it impossible to determine the actual degree of this bending by NMR. For helical structures like RNA or DNA this effect is much less severe because the helical backbone intrinsically serves as a scaffold, limiting the structural variability over a medium range of several residues. We could not identify a correlation between the extent of the bending and the variation in total restraining energy or any other term of the minimal force field used for the simulated-annealing calculations. The change in global shape is locally reflected in an

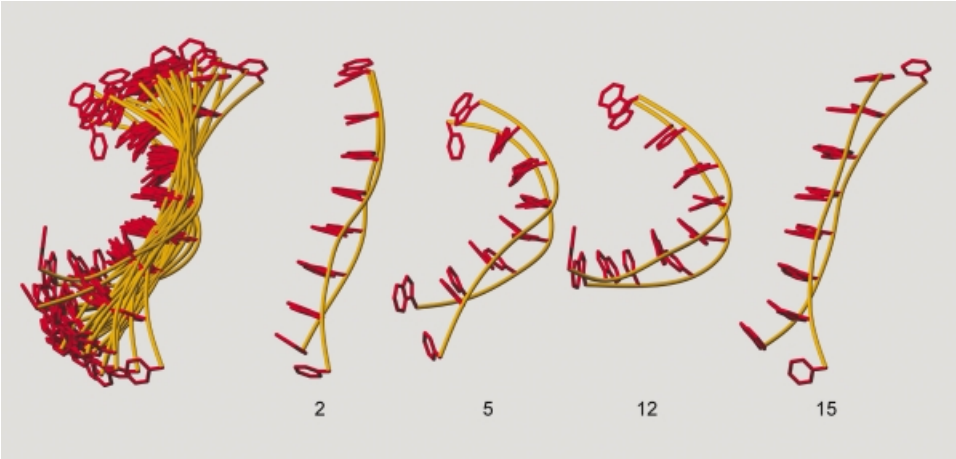

Fig. 9. *Overall bending of the duplex.* Left: *Schematic representation of the bundle shown in* Fig. 7,a *with all backbone atoms replaced by a spline through the C(1′) atoms* (side-on view). Right: *Individual members of the bundle illustrating the variation of overall bending* (structures are ordered according to total energy).

increased propeller twist and more prominent buckling of the base pairs together with a slight positive roll of adjacent bases on the same strand.

The result of enforcing a more-regular base pairing was studied by introducing a weak restraint to force the two bases into a common plane, an approach often used for RNA structures [9a]. With a force constant of 0.5 kcal/mol on all atoms of the pairing bases the bending was clearly attenuated, together with a slight decrease in the mean stacking distance (see *Fig. 10,b*). The stacking distance was only minimally decreased further when a force constant of 1 kcal/mol for the co-planarity restraint was used (*Fig. 10,c*) but now, only 20 structures fulfilled the selection criteria. As expected, the extent of the bending was now correlated with the total energy of the structure, but the backbone had still enough degrees of freedom to allow overall bending and a slight positive roll.

Discussion. – The most characteristic structural features of the experimentally derived solution structure of **1** are the strong backbone inclination of *ca.* $-50°$[5]) (see *Fig. 6,a*) and the band-like overall shape of the duplex. None of the members of the structural bundle derived shows significant helicity, although, due to their short-range nature, the experimental restraints do not define this parameter directly.

If we accept the postulate that achievement of optimal interstrand base stacking is one of the major driving forces determining the conformation of pentopyranosyl-NA duplexes, the deviation of the backbone geometry of arabinopyranosyl-(4′⟶2′)-oligonucleotides from the idealized conformation can be rationalized. A significant part of the deviation consists of adjustments needed to assume a nearly linear single strand conformation with real bond lengths and bond angles as opposed to the idealized structure on which the original conformational analysis was based.

Reduction of the backbone angle β from 180° (idealized) to *ca.* 155° (average NMR-derived value) leads to a reduction of the distance between stacked base pairs

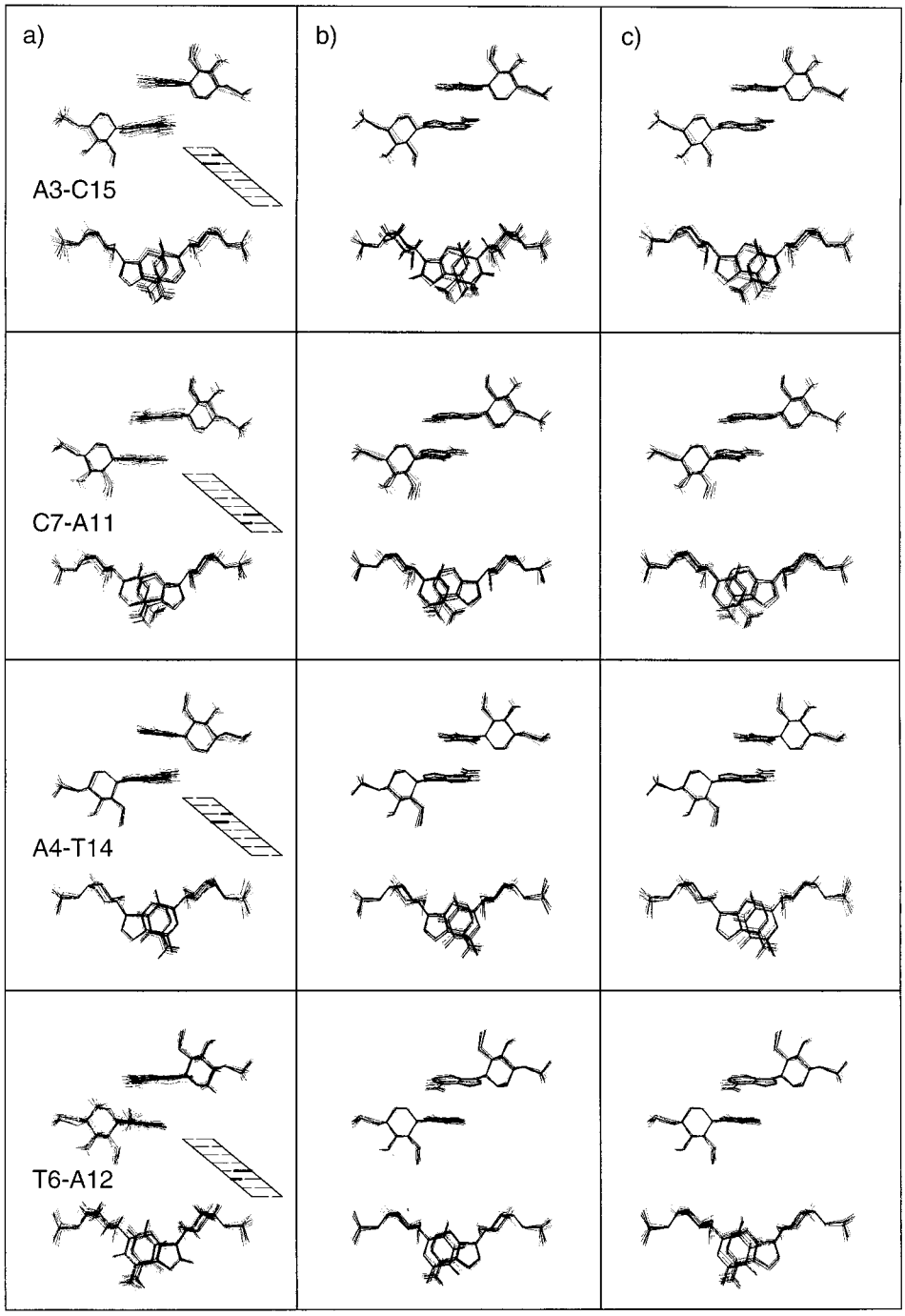

Fig. 10. *Graphical representation of interstrand stacking involving residues 2–7 of the duplex*. For each pair, all structures of the bundle shown in *Fig. 7,a* were superimposed based on all atoms of the two involved nucleotides. From left to right: calculations with increasing coplanarity restraints: *a*) $k_{plan} = 0.0$ kcal/mol, *b*) $k_{plan} = 0.5$ kcal/mol, *c*) $k_{plan} = 1.0$ kcal/mol).

(rise) and, at the same time, increases the inclination, concomitant to an increase in the slide between subsequent base pairs and improved interstrand stacking. As a consequence of the axial attachment of the phosphodiester group at C(4′), the backbone of arabinopyranosyl duplexes becomes more deformed than that of p-RNA to reach optimal interstrand base stacking. On the other hand, the steric clash between the equatorial $H_{eq}-C(5')$ proton and the aromatic ring of the next base upstream, which may well limit the possible inclination in p-RNA, is reduced or even absent in the arabinopyranosyl system because the axial 4′-substituent introduces a step between neighboring residues moving the $H_{eq}-C(5')$ below the edge of the aromatic ring. This is clearly reflected in the chemical shifts of these protons, which, in contrast to p-RNA, are practically not shielded by the ring current.

The large deviation of ε from idealized $-60°$ to values between -100 and $-120°$ greatly reduces the steric interaction between one of the phosphate O-atoms of the 2′-phosphodiester group and the equatorial 3′-OH by raising the phosphodiester group above the mean plane of the pyranose chair. This change in geometry also reduces the step introduced by the axial orientation of the 4′-phosphodiester linkage in the idealized model (shift between stacked base pairs). Model considerations indicate that this reduction of $\pi-\pi$ stacking could in principle be avoided by changing the glycosidic torsion angle χ, by introducing helicity, or by means of a 'crankshaft' movement around ε (and α). The experimental structure bears out that the latter adjustment dominates.

The question as to why arabinopyranosyl duplexes show such an extraordinary pairing strength compared to p-RNA is difficult to answer with only the static geometric picture of the experimentally determined conformation in hand. The analysis given above indicates that, to achieve efficient interstrand stacking and base-pairing, the arabinopyranosyl-(4′⟶2′)-oligonucleotides undergo deformation of the backbone further away from the idealized model than does p-RNA. Since the deviation from optimal bond staggering is expected to introduce strain energy, the free energy of formation of an arabinopyranosyl-(4′⟶2′)-oligonucleotide in the pairing conformation should be higher than that of a corresponding p-RNA strand. However, this strain energy would manifest itself in the free energy of pairing only, if the unpaired single strand were able to adopt a more relaxed conformation of lower energy. If, on the other hand, the steric interactions between the 3′-OH and the 2′-phosphodiester groups enforce the same backbone-angle deformation in the unpaired single strand and in the duplex, such preorganization by steric constraint would reduce the loss of entropy that opposes duplex formation. For the homobasic octameric duplexes $(rpA_8)(rpT_8)$ and $(apA_8)(apT_8)$, the enthalpies of pairing are equivalent within 2 kcal/mol whereas the arabinopyranosyl system shows an entropy of pairing that is 6.8 eu less-negative than that of p-RNA [1a]. However, the thermodynamic data for pairing of pentopyranosyl-(4′⟶2′)-oligonucleotides reported in [1a,b] show a strong dependence of the enthalpic *vs.* entropic components on the sequence and, as a whole, do not give a coherent picture with regard to the preorganization hypothesis. Therefore, the role of preorganization in the pairing process of pentopyranosyl-(4′⟶2′)-oligonucleotides remains to be demonstrated experimentally, *e.g.*, by studying the conformational dynamics of non-complementary single strands.

This work was supported by *Novartis AG*, CH-Basel, *Firmenich SA*, CH-Geneva, *The Skaggs Foundation*, LaJolla, CA, USA and the *Schulleitung* of the ETH-Zürich. *M.-O. E.* and *B. J.* thank Dr. *D. Moskau* of *Bruker Biospin AG*, CH-Fällanden for access to a 600 MHz NMR spectrometer.

Experimental Part

Synthesis. The preparation of ap(CGAATTCG) (**1**) on a 10-mg scale was carried out with 2′-*O*-(4′,4″-dimethoxytrityl)-3′-*O*-benzoyl-protected 4′-phosphoramidite monomer units. The monomer synthesis as well as synthesis of the octamer strand **1** on solid support will be reported separately in this journal [12]. T_m-(UV/VIS): 78.1° ($c = 1$ μM **1** in 0.15M NaCl, 0.01M phosphate buffer pH 7.0, 0.1 mM Na_2EDTA).[6])

Sample Preparation. The oligomer (Na salt, 446 OD_{260}, 4.9 μmol) was dissolved in 0.6 ml sodium arsenate buffer (50 mM), prepared by dissolving the appropriate amount of Na_2HAsO_4 in H_2O, adjusting the pH to 7.0 with HCl/NaOH, lyophilizing and redissolving in D_2O). The resulting 8.2 mM soln. of oligomer was centrifuged and decanted into the NMR tube. For measurements involving exchangeable protons, the sample was lyophilized and redissolved in 0.6 ml H_2O/D_2O 9:1.

NMR Spectra: Bruker AMX-500, *DRX-500*, and *DRX-600* spectrometers at 26.7°, except were indicated otherwise. Quadrature detection was achieved with TPPI. The solvent signal in H_2O/D_2O was suppressed by means of WATERGATE [13]. The spectra were processed on a SGI O_2 workstation with XWINNMR 2.6 and analyzed with FELIX [14].

1D-Spectra (*DRX-/AMX-500*): ^{1}H-NMR Spectra, with 128 scans and 88k data points were measured and processed without further zero-filling. The spectra in D_2O were recorded with and without ^{31}P-decoupling during acquisition, with a spectral width of 4529 Hz. For the spectra in H_2O/D_2O (T = 11.4, 21.6, 26.7, 29.8, 40.4, and 48.1°) the spectral width was 12019 Hz and multiplication with an exponential function was used for apodization. A broad-band decoupled ^{31}P-NMR spectrum (64 scans and 32k data points) was recorded in D_2O with a spectral width of 1532 Hz. The data were apodized with an exponential function and zero-filled to 128k. A ^{1}H-decoupled ^{13}C-NMR spectrum was recorded in D_2O with 166284 scans and 90906 data points over a spectral width of 22727 Hz. The data were apodized with a gaussian function and zero-filled to 256k.

2D-Experiments: Parameters for individual experiments are given in *Table 6*.

Table 6. *Parameters Used for the 2D-Experiments*

Experiment	MHz	SW1/SW2 [Hz][a])	TD1/TD2[b])	S1/S2[c])	NS/DS[d])	d [s][e])	Apod.[f])
DQF.COSY	500.03	1929	8k/512	8k/1k	16/16	1	g/g
TOCSY	500.03	4771	4k/512	4k/1k	16/16	2.5	g/g
H,P-COSY	500.03	4529/3094	2k/512	2k/1k	32/16	0.075	g/g
NOESY	500.13	5208	4k/480	4k/1k	64/16	2.0	q/q
NOESY	600.13	6720	4k/512	4k/1k	32/16	2.5	q/q
HSQC	500.03	4529/22727	2k/512	2k/1k	16/16	2.5	q/g
HSQC	500.13	4032/3144	2k/1k	2k/2k	32/96	1.5	g/g
HMBC	500.03	2893/22727	1k/512	1k/1k	32/16	2.5	q/q
NOESY (H_2O)	500.03	12019	4k/1350	4k/1350	32/16	2.5	q/g

[a]) Spectral width. [b]) Number of recorded data points. [c]) Size of the processed spectrum (real part). [d]) Number of scans/dummy scans. [e]) Recycle delay. [f]) Function used for apodization (g: *Gauss* function, q: $\sin^2$ shifted by $-\pi/2$).

DQF.COSY: For the determination of ^{1}H,^{31}P-coupling constants, a set of two ^{31}P-coupled/decoupled spectra were measured under identical conditions. Decoupling was achieved by a 180°-refocussing pulse in the middle of t_1 and WALTZ16-cpd decoupling during acquisition.

TOCSY: A version with 120 ms DIPSI2 mixing was used with a low-power purge pulse of 30 μs and *z*-filter delays of 2 and 3 ms.

---

6) Calculation of T_m from the thermodynamic data reported in [3a] gives a value of 51° for the corresponding p-RNA sequence.

NOESY: Spectra in D_2O were measured with mixing times of 50, 100, and 150 ms (*AMX-500*) and 75 and 150 ms (*DRX-600*). Spectra in H_2O/D_2O were measured with mixing times of 80, 100, and 150 ms.

HSQC: Besides a HSQC spectrum with full spectral width in the ^1H and ^{13}C dimension, a set of four HSQC-spectra with smaller spectral widths (confined to the sugar regions) and better resolution in t_1 was recorded with and without ^{31}P-decoupling in t_1 and/or t_2 for the determination of $^{13}C,^{31}P$-coupling constants.

Generation of Distance Restraints. After inspection of the build-up curves calculated from the NOESY spectra in D_2O at mixing times of 50, 100, and 150 ms, correlations with possible contributions of spin diffusion were excluded from the generation of distance restraints at this stage. Also excluded were cross-peaks with signal-to-noise ratios ≤ 1 at mixing time 50 ms and peaks overlapping more than 30% at their basis. The overlapping peaks were assigned an upper bound of 6.0 Å. For the remaining build-up curves, the slope at $t = 0$ was determined by linear regression over the first two mixing times. Distances of 2.5 Å between pyrimidine protons H–C(5) and H–(6) (apC1 and apC7) and of 1.8 Å between H_{eq}–C(5′) and H_{ax}–C(5′)(apC7 and apC8) were used as reference values with which to generate distance restraints with upper and lower bounds of $\pm 20\%$ of the calculated value. For distances involving one Me group, a multiplicity correction of 20% was applied [15]. All cross-peaks discarded in this first step were now assigned a distance upper bound of 7.0 Å. 71 distances involving nonexchangeable protons were generated this way (see *Table 4*). The same procedure was applied to generate 26 distances between imino protons and nonexchangeable protons from the NOESY spectra in H_2O/D_2O. H-bonding restraints between base pairs 2–7 were introduced as NOE restraints between the heavy atoms involved and assigned a value of 3.1 ± 0.4 Å [4]. A second restraint between the H-atom and the acceptor (2.2 ± 0.2 Å) was used to enforce the linearity of the H-bond. An additional planarity restraint ($k_{plan} = 500$ kcal/mol) was introduced to force the donor H-atom into the base plane containing the acceptor atom and prevent excessive staggering of the outer base pairs.

Structure Calculation. The calculations were performed with X-PLOR 98.1 [16]. The topallhdg.dna and parallhdg.dna files were modified to cope with the new arabinose moiety (equilibrium values for bonds and angles are based on the X-ray structure of a monomer). All parameters concerning the base atoms as well as the force constants ($k_{bond} = 1000$ kcal/mol, $k_{angle} = 500$ kcal/mol, $k_{improper} = 500$ kcal/mol) were left unchanged. The torsion-angle dynamics protocol applied was according to the scheme proposed by *Stein et al.* [8] for a DNA duplex. The starting structures were generated by torsion-angle randomization of two energy minimized single strands at the beginning of each cycle. Final structures were accepted if they showed no NOE violation > 0.2 Å, no violation of an angle restraint $> 5°$, and no deviation from equilibrium bond lengths, angles, and impropers of > 0.05 Å, $> 5.0°$, and $> 1°$, respectively. Figures were made with the MOLMOL [17] and POV-Ray 3.1 g [18] software.

REFERENCES

[1] a) M. Beier, F. Reck, T. Wagner, R. Krishnamurthy, A. Eschenmoser, *Science* **1999**, *283*, 699; b) O. Jungmann, H. Wippo, M. Stanek, H. K. Huynh, R. Krishnamurthy, A. Eschenmoser, *Org. Lett.* **1999**, *1*, 1527; c) S. Pitsch, S. Wendeborn, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **1993**, *76*, 2161; d) S. Pitsch, R. Krishnamurthy, M. Bolli, S. Wendeborn, A. Holzner, M. Minton, C. Lesueur, I. Schlönvogt, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **1995**, *78*, 1621; e) F. Reck, H. Wippo, R. Kudick, R. Krishnamurthy, A. Eschenmoser, *Helv. Chim. Acta* **2001**, *84*, 1778; f) T. Wagner, H. K. Huynh, R. Krishnamurthy, A. Eschenmoser, *Helv. Chim. Acta* **2002**, *85*, 399.

[2] A. Eschenmoser, M. Dobler, *Helv. Chim. Acta* **1992**, *75*, 218.

[3] a) I. Schlönvogt, S. Pitsch, C. Lesueur, A. Eschenmoser, B. Jaun, R. Wolf, *Helv. Chim. Acta* **1996**, *79,* 2316; b) S. Illin, I. Schlönvogt, M.-O. Ebert, B. Jaun, H. Schwalbe, *Chem. Biochem.* **2002**, *3*, 93.

[4] G. Varani, F. Aboul-ela, F. H.-T. Allain, *Prog. Nucl. Magn. Reson. Spectrosc.* **1996**, *29*, 51.

[5] M. J. J. Blommers, D. Nanz, O. Zerbe, *J. Biomol. NMR* **1994**, *4*, 595.

[6] H. Schwalbe, W. Samstag, J. W. Engels, W. Bermel, C. Griesinger, *J. Biomol. NMR* **1993**, *3*, 479.

[7] S. S. Wijmenga, M. M. W. Mooren, C. W. Hilbers, in 'NMR of Macromolecules. A Practical Approach', Ed. G. C. K. Roberts, IRL Press, Oxford, 1993, p. 217.

[8] E. G. Stein, L. M. Rice, A. T. Brünger, *J. Magn. Reson.* **1997**, *14*, 51.

[9] a) J. Cromsigt, B. van Buuren, J. Schleucher, S. Wijmenga, *Methods Enzymol.* **2001**, *338*, 371; b) N. B. Ulyanov, T. L. James, *Methods Enzymol.* **1995**, *261*, 90.

[10] M. Guéron, J.-L. Leroy, *Methods Enzymol.* **1995**, *261*, 383.

[11] R. E. Dickerson, in 'Oxford Handbook of Nucleic Acid Structure', Ed. S. Neidle, Oxford University Press, Oxford, 1999, p 145.

[12] M. Beier, O. Jungmann, A. Luther, H. K. Huynh, R. Krishnamurthy, A. Eschenmoser, in preparation.
[13] M. Piotto, V. Saudek, V. Sklenar, *J. Magn. Reson., Ser. A* **1993**, *102*, 241.
[14] http://www.accelrys.com
[15] D. Neuhaus, M. P. Williamson, 'The Nuclear Overhauser Effect in Structural and Conformational Analysis', 2nd edn., John Wiley & Sons, Inc., New York, 2000, p. 503.
[16] a) A. T. Brünger, 'X-PLOR. A System for X-Ray Crystallography and NMR', Yale University Press, New Haven, 1992; b) J. Badger, R. A. Kumar, P. Yip, S. Szalma, *Proteins: Struct. Funct. Genet.* **1999**, *35*, 25.
[17] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **1996**, *14*, 51.
[18] http://www.povray.org
[19] S. S. Wijmenga, B. N. M van Buuren, *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287.
[20] J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, K. Wüthrich, *Pure Appl. Chem.* **1998**, *70*, 117.

Received July 1, 2002