Volume 16 Number 11 1988 Nucleic Acids Research

Bulge-out structures in the single-stranded trimer AUA and in the duplex (CUGGUGCGG).(CCGCCCAG). A model-building and NMR study

Yvonne Th.van den Hoogen, Aart A.van Beuzekom, Erik de Vroom, Gijs A.van der Marel, Jacques H.van Boom and Cornelis Altona

Gorlaeus Laboratories, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands

Received April 14, 1988; Accepted May 6, 1988

ABSTRACT

Model-building studies were carried out on the trimer AUA. Bulge-out structures which allow incorporation into a continuous RNA helix were generated and energy-minimized. All geometrical features obtained by previous NMR studies on purine-pyrimidine-purine sequences are accounted for in these models. One of the models was used to fit into a double helical fragment. Only minor changes were necessary to construct a central bulge-out in an otherwise intact duplex.

NMR and model-building studies were performed on the duplex (CUGGUGCGG)•(CCGCCCAG) which contains an unpaired uridine residue. NOE data, chemical-shift profiles and imino-proton resonances provided evidence that the extra U is bulged out of the duplex. The relatively small dispersion in ^{31}P chemical shifts (≈0.7 ppm) indicate the absence of t/g or g/t combinations for the phosphodiester angles ζ/α. An energy-minimized model of the duplex, which fits the present collection of data, is presented.

INTRODUCTION [1,2]

An extra unpaired base on one strand is frequently found to occur in both double-helical RNA and DNA. In DNA these extra bases are considered to play an important role in frameshift mutagenesis [3]. In RNA several functions for these extra bases, denoted bulges, are proposed. In some cases the bulges are found to serve as protein recognition features [4,5]; the formation of certain lariat structures requires a single bulge base [6]; and, more generally, the bulges are considered to give rise to a locally flexible structure [7,8]. In the forementioned cases, the bulges are part of large systems, such as ribosomal RNA, which prevents detailed analysis of the local structure. Therefore it appears of interest to study model compounds in order to obtain more information about fragments containing an extra base. Such studies are designed to provide an answer to the question whether the extra base is stacked into the duplex or remains bulged-out. Thus far, little is known about the behaviour of bulges in double-stranded RNA. However, several studies on single-stranded RNA and on double-helical DNA fragments have been performed.

From studies on single-stranded RNA fragments [9,10,11] it is known that trimers with a purine-pyrimidine-purine sequence prefer to adopt a structure in which the pyrimidine is bulged-out and the purines stack upon each other in a 1-3 fashion. Recently, several models for bulged-out trimers have been generated [12,13] but none of these allows extension of stacking interactions in both the 3' and 5' direction and fits all the NMR data at the same time. To find such a structure appears important because it is often postulated to exist in a RNA double helix [4-8].

From several studies on double-helical DNA fragments containing an extra base it was concluded that an extra purine stacks into the helix [14-17]. In contrast, the extra dC in $d(CA_3CA_3G)\cdot d(CT_6G)$ duplex was considered to be bulged out [18]. Finally, a recent investigation from this laboratory [19] showed that the extra dT in d(CTGGTGCGG)•d(CCGCCCAG) is stacked inside the duplex. The question arises whether or not similar observations can be made for double-stranded RNA fragments containing an extra base.

In the present report, NMR and model-building studies on bulges in RNA are dealt with. A preliminary report on some models has been presented elsewhere [20]. First, a model-building study of the single-strand bulge-out trimer AUA is presented. A model which fits all NMR data and allows extension at both the 3' and 5' side is fitted into a duplex structure and is again energy-minimized. In order to determine experimentally whether an extra uridine favours to reside in the interior of a RNA duplex or prefers to bulge-out we have carried out NMR studies on (CUGGUGCGG)•(CCGCCCAG), henceforth denoted (9+8)-mer. Chemical-shift and NOE data as well as information from imino-proton and ^{31}P resonances are used to monitor the behaviour of the unpaired base in the (9+8)-mer. A comparison with the data obtained for the deoxyribose counterpart d(CTGGTGCGG)•d(CCGCCCAG) is given.

MATERIALS AND METHODS

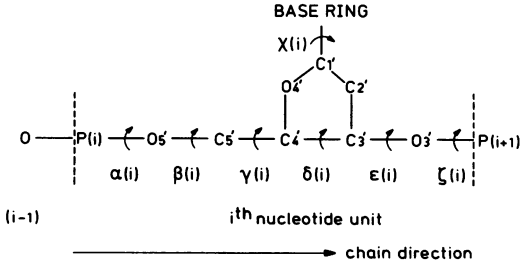
The octamer CCGCCCAG and nonamer CUGGUGCGG were synthesized via a modified phosphotriester approach [21]. After purification, the fragments were treated with a Dowex cation-exchange resin (Na^+ form) to yield the sodium salt. NMR samples were lyophilized three times from 99.8% D_2O and finally taken up in 99.95% D_2O. A trace of EDTA was added and the pH was adjusted to 7 (meter reading); Me_4NCl was used as internal reference. For most practical purposes the Me_4NCl scale can be converted to the DSS scale by setting $\Delta\delta(Me_4NCl\text{-}DSS) = 3.18$ ppm. In order to observe imino-proton resonances of the (9+8)-mer a 3.5 mM sample was prepared in a H_2O/D_2O (90/10) solvent mixture; in this case DSS was used as reference compound.

Spectra of the separate octamer and nonamer as well as of the (9+8)-mer 1:1 mixture were recorded at 3.5 mM nucleotide concentration. In order to obtain an 1:1 mixture of the nonamer and octamer the concentrations of the separate single-strands were determined spectrophotometrically and the solutions were mixed in the correct ratio. One-dimensional ^1H NMR spectra of the D_2O samples were measured on a Bruker WM-300 spectrometer equipped with an Aspect 2000 computer. Spectra were recorded at temperatures between 0 °C and 100 °C at regular intervals in order to obtain chemical-shift vs temperature profiles. The residual HDO peak in the D_2O samples was suppressed by means of irradiation at the HDO frequency. The imino-proton spectra were recorded on a Bruker MSL-400 spectrometer interfaced with an Aspect 3000 computer. The H_2O signal in these spectra was minimized by means of a time-shared long pulse in combination with a data-shift accumulation routine [22,23]. A phase-sensitive NOESY experiment of the (9+8)-mer was recorded at 295 K on a Varian VXR-500 spectrometer. A mixing time of 0.8 s was used; 512 spectra of 2 K were collected. Before Fourier transformation both the t_2 and t_1 domains were zero-filled once, the t_2 domain was apodized with a Gaussian window and the t_1 domain with a sine-bell-squared window. After Fourier transformation the low-field half of the ω_2 domain and the ω_1 domain were base-line corrected with a third degree polynomial function.

^{31}P spectra (broad-band ^1H decoupled) were recorded on a Bruker WM-300 spectrometer at an operating frequency of 121.6 MHz.

RNA models were generated with the aid of the programs BUILDER [24] and MacroModel [25] on an Evans and Sutherland PS-350 computer-graphics system, interfaced to a VAX 11/750. The structures were energy-minimized with the all-atom version of the molecular-mechanics program AMBER [26].

The recommended IUPAC-IUB nomenclature is used throughout this work [27,28]. A schematic representation is displayed in Scheme I.

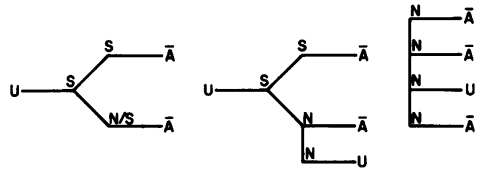


Scheme I. Conformational notation of oligonucleotides [27,28].

RESULTS AND DISCUSSION

Single-stranded bulge-out structures

Geometrical characteristics. From an NMR study on thirteen trimers Lee and Tinoco [9] demonstrated that those with a purine-pyrimidine-purine sequence, ACG, AUG and GUG, display unexpected chemical-shift behaviour: the interior pyrimidine protons are less shielded than expected from the constituent dimers whereas the terminal purine residues show an excess of shielding. These data led the authors [9] to propose a structure in which the interior pyrimidine is bulged out and a next-nearest-neighbour stacking interaction is present between the terminal purine residues. Parallel studies carried out in this laboratory [11,29-31] on sequences containing $\bar{A}$ ($\bar{A}$ = N^6-dimethyladenosine) and U bases supported this proposal and provided more geometrical information about bulge-out structures. A schematic representation of the preferred conformations of these sequences is shown in Scheme II. The favoured ribose ring conformations are indicated.



Scheme II. Preferred conformations of ĀUĀ, ĀUĀU and ĀĀUĀ. The major conformation of the ribose rings is indicated.

The trimer ĀUĀ spontaneously forms a bulge-out structure in which the Ā(1) and U(2) ribose rings prefer the S-type conformer and the Ā(3) sugar exists in a 50/50 N-type/S-type equilibrium [11] (Scheme II). Because the rotamer population around ϵ is correlated with the N/S equilibrium [30] one should expect the presence of a considerable amount of ϵ^- rotamer for both Ā(1) and U(2). Indeed, the respective values of $J_{C4'P}$ and $J_{C2'P}$ [30] as well as large (W-path) long-range $J_{H2'P}$ couplings [11] indicate the presence of ϵ^- rotamer in the first two residues of ĀUĀ. The backbone torsion angles β and γ in ĀUĀ generally prefer to adopt the β^t and γ^+ rotamer ranges, respectively [11]. However, the preferences for β^t (77 ± 5%) and γ^+ (63-76% in the Ā(3) residue) are not overly prominent, which fact supports our earlier proposal that the bulge-out does not represent a single well-defined conformation but rather an assembly of forms. Finally, the Ā(1) base in ĀUĀ

Table 1. Characteristics of the four models generated for AUA.

Model	A(1)	U(2)	A(3)
1	S anti ϵ^-	β^t γ^+ S anti ϵ^-	β^t γ^+ S anti
2	S anti ϵ^-	β^t γ^+ S anti ϵ^-	β^t γ^+ N anti
3	S syn ϵ^-	β^t γ^+ S anti ϵ^-	β^t γ^+ S anti
4	S syn ϵ^-	β^t γ^+ S anti ϵ^-	β^t γ^+ N anti

prefers to adopt the syn orientation whereas the U(2) and $\bar{A}$(3) bases favour the anti position [31].

Extension of the trimer at the 3'-side with an U residue results in a side-by-side occurrence of a bulge-out structure in the $\bar{A}U\bar{A}$- part together with a strong stacking proclivity in the $\bar{A}U$- part [29] (Scheme II). The main difference between the $\bar{A}U\bar{A}$- part of $\bar{A}U\bar{A}U$ and $\bar{A}U\bar{A}$ lies in the outspoken preference for N-type conformer in $\bar{A}$(3) of $\bar{A}U\bar{A}U$ whereas in $\bar{A}U\bar{A}$ a 50/50 N/S mixture is found. From these observations we conclude that $\bar{A}$(3) in the bulge-out sequence $\bar{A}U\bar{A}$ remains free to select either the N-type or the S-type sugar conformation.

Extension of the trimer $\bar{A}U\bar{A}$ with an $\bar{A}$ base at the 5'-side results in a disappearance of the bulge-out structure in the -$\bar{A}U\bar{A}$ part [31] (Scheme II). The ribose rings of the $\bar{A}$(2) and U(3) residues in $\bar{A}\bar{A}U\bar{A}$ prefer the N-type conformer in contrast to the corresponding sugars in $\bar{A}U\bar{A}$. Another difference between $\bar{A}\bar{A}U\bar{A}$ and $\bar{A}U\bar{A}$ appears the anti orientation of the $\bar{A}$(2) base in $\bar{A}\bar{A}U\bar{A}$, whereas in $\bar{A}U\bar{A}$ a preference for the syn position of the 5' base was observed. These data were interpreted in terms of a 5' → 3' conformational transmission effect in which the bulge-out disappears when the $\bar{A}$(2) residue in $\bar{A}\bar{A}U\bar{A}$ is forced into the N-type sugar due to the $\bar{A}$(1) - $\bar{A}$(2) stack. Consequently, ϵ(2) flips into the trans conformer and the $\bar{A}$(2) base into the anti orientation which factors result in an increase of stack between $\bar{A}$(2) and U(3) at the expense of the next-nearest-neighbour bulge-out interaction. Studies on $U\bar{A}U\bar{A}$ [31] and $\bar{A}U\bar{A}U\bar{A}$ [32] support this conclusion. Thus the S-type sugar together with the ϵ^- rotamer in the $\bar{A}$(1) residue seems to be a prerequisite for the formation of bulge-out structures in $\bar{A}U\bar{A}$.

In summary, from these studies we have deduced that in the bulge-out structure of $\bar{A}U\bar{A}$ the following characteristics are preferred: $\bar{A}$(1) favours S-type sugar, ϵ^- conformer and syn for the base; U(2) prefers β^t, γ^+, S-type sugar, ϵ^- and anti; $\bar{A}$(3) β^t, γ^+ and anti whereas the sugar can adopt both N and S. These geometrical data should be accounted for in model building studies.

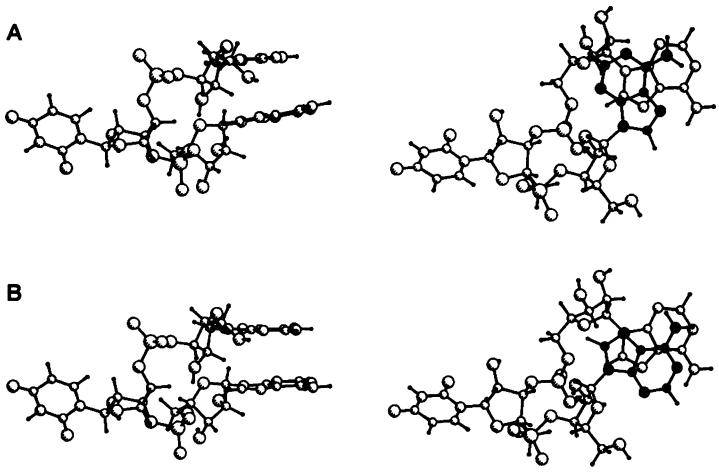


Figure 1. Energy-minimized structure of (A) model 2 and (B) model 4 (Table 2). The structures on the left-hand side are drawn parallel to the first base while those on the right-hand side are drawn perpendicular to the first base. The H-bridge between 2'-OH(1) and an oxygen of P(3) is indicated with an arrow.

Model building. Recently, Lee [13] reported several possible conformations for AAA, one of which (structure B1 in [13]) represented a bulge-out structure that in principle can be incorporated into a RNA helix. Perhaps unfortunately, the geometrical features of this structure do not correspond with the above-mentioned experimentally deduced characteristics of $\bar{A}U\bar{A}$. Moreover, several bulge-out structures for $\bar{A}U\bar{A}$ were generated earlier [12] but none of these allows incorporation into a RNA helix. In order to obtain a fitting structure further model-building studies were carried out.

All geometrical information obtained by the NMR studies, *vide supra*, was used to generate structures with the aid of the program BUILDER [24]. The torsion angles α and ζ were varied in order to obtain the desired structural aspects, such as the 1,3 stacking interaction between the purines, while maintaining the possibility to incorporate the trimer into a RNA helix.

Four different models for AUA (non-methylated) were generated differing in $\chi(1)$ (syn/anti) and $\delta(3)$ (N/S) (Table 1). These structures were energy-minimized with the all-atom version of the molecular-mechanics program AMBER [26]. The structures thus obtained for the models 2 and 4 (Table 1) are shown in Fig. 1. The approximately parallel purine base planes in the various models are at a distance of circa 3.5 Å and the base-base overlap appears satisfactory (Fig. 1). Calculated energies and the backbone torsion

Table 2. Backbone torsion angles and energies for the energy-minimized models of AUA (Table 1).

Model	Energy[a]	Residue	α	β	γ	δ	ε	ζ	χ	P	ϕ_m
	kcal/mol		degrees								
1	-94	A(1)			62	134	273	261	-152	143	40
		U(2)	72	197	55	127	276	257	-148	133	41
		A(3)	37	164	47	149			-147	167	35
2	-91	A(1)			61	139	274	256	-148	148	41
		U(2)	69	202	58	135	278	262	-143	142	40
		A(3)	312	196	69	85			-170	10	39
3	-93	A(1)			59	136	268	260	42	143	42
		U(2)	69	198	57	132	280	258	-147	139	41
		A(3)	36	159	51	146			-156	162	36
4	-91	A(1)			58	140	272	254	41	148	43
		U(2)	69	204	59	135	278	273	-144	142	40
		A(3)	318	196	71	83			-172	12	39

a. $\epsilon_{diel} = 2.5$ was chosen instead of the usual $\epsilon = R$.

angles for the four models are listed in Table 2. It is seen, Table 2, that the structures turn out to have almost identical energy and backbone angles, except for α(3) which appears correlated to δ(3). The main backbone features of the bulge-out structures in AUA appear to be a ζ^-/α^+ phosphodiester torsion angle combination around P(2) and a ζ^-/α^- or ζ^-/α^+ combination around P(3) depending upon the sugar conformation of A(3). A remarkable feature which is shared by the four calculated models appears to be a H-bridge between the 2'-OH of the A(1) residue and a phosphate oxygen of P(3) (Fig. 1). However, an NMR study of the hybrid dAr(UA) [31] revealed that a bulge-out structure is formed in dAr(UA), despite the absence of the 2'-OH in the dA(1) residue. This is taken to mean that the H-bridge seen in the calculated RNA structures cannot represent a major stabilizing factor for bulge-out structures.

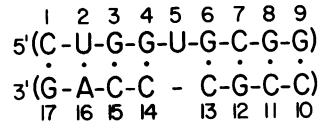
We conclude that several single-stranded bulge-out structures can be generated that appear to be in agreement with the data from NMR studies. Structure 2 (Table 2, Fig. 1) was used to fit into a double-stranded fragment, *vide infra*.

Double-stranded bulge-out structures

NMR studies on (CUGGUGCGG)•(CCGCCCAG)

Assignment of non-exchangeable protons. The assignment of most of the H8, H2 and H6 resonances of the separate single-strand nonamer and octamer was made

by comparison with other compounds at 70 °C. The numbering of the residues is indicated in Scheme III. The parameters for the calculation of proton NMR



Scheme III. Numbering of the residues in the (9+8)-mer duplex.

chemical-shifts of oligoribonucleotides derived by Bell et al. [33] appeared to be very useful. In the case of the nonamer CUGGUGCGG the two H6 resonances belonging to the C residues could be distinguished from those of the U residues due to the difference in J_{56} between H6 of C residues (7.6 Hz) and J_{56} of H6 of U residues (8.1 Hz). Subsequently, the H6 resonances of the C residues could be assigned unambiguously to C(1) and C(7) and those of the U residues to U(2) and U(5) with the aid of the parameters mentioned above [33]. In an analogous fashion most of the H8 resonances of the nonamer and the resonances of the G residues were assigned. Due to almost identical chemical-shifts a definitive distinction between H8(4) and H8(9) and between H6(10) and H6(14) is precluded at this moment.

The assignment of the base and H1' proton resonances of the (9+8)-mer is based upon a careful analysis of a NOESY experiment; the assignment followed the procedures described by several authors [34-36]. A contour plot of the region in which the H8/H6/H2 - H1'/H5 connectivities are found is shown in Fig. 2. The H8/H6 base protons and H1' sugar protons were assigned via the H1'(i-1) - H8/H6(i) - H1'(i) NOE pathway. Subsequently, the H5 resonances were identified through the strong intranucleotide H6/H5 NOEs.

In the case of the octamer strand of the duplex sequential assignment from H6(10) up to H8(17) was easily made (Fig. 2). Note that this assignment includes NOEs between C(13) and C(14). Also in other parts of the spectrum interresidue NOEs were observed between C(13) and C(14). This feature is of particular interest because these residues flank the unpaired U in the opposite strand. The presence of interresidue NOEs between C(13) and C(14) strongly indicates that the extra U does not lead to an unusually large distance between C(13) and C(14). The assignment of the base and H1' resonances of the octamer strand is completed with the assignment of H2(16). As expected, the H2(16) displays both an intrastrand NOE to the H1' of its

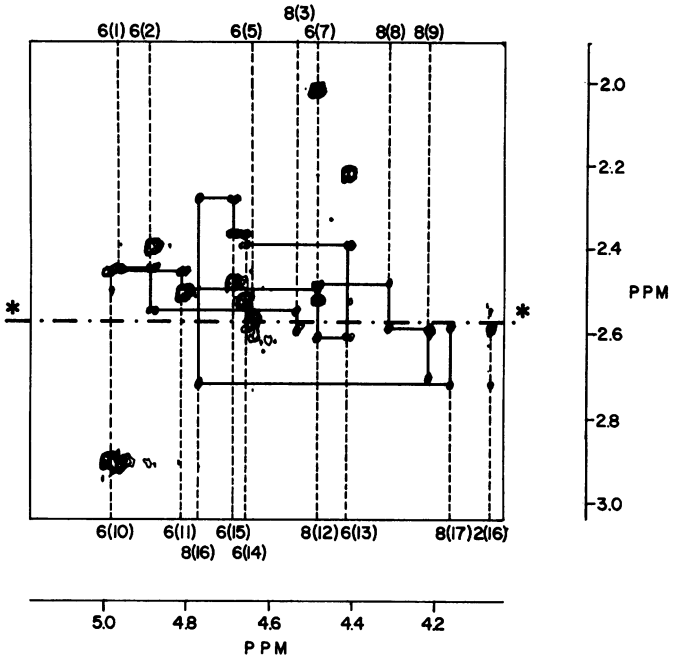


Figure 2. Contour plot of part of the NOESY spectrum of the (9+8)-mer duplex; 3.5 mM, 295 K, pH 7. The region in which H2/H8/H6 - H1'/H5 connectivities are found is shown.

3' neighbour, H1'(17), and an interstrand NOE to the 5' neighbouring residue on the opposite strand, H1'(3).

Unfortunately, the assignment of the resonances of the nonamer strand in the duplex appeared less straightforward. Although H6(2)-H1'(1) and H8(3)-H1'(2) connectivities were easily found, the sequential assignment could not be continued further than G(3). From C(7) to G(9) again a sequential assignment could be taken up (Fig. 2). However, the missing links in the G(4)-U(5)-G(6) part remain of particular interest because these in principle can provide an answer to the question whether or not the extra U is stacked in the duplex. Although the H8(4) and H8(6) could not be assigned due to overlap, no anomalous intense NOEs between H8 and H1' resonances are observed in this region. This is taken to mean that all G residues adopt an anti orientation in contrast to the observation of a syn position for the first purine in single-stranded purine-pyrimidine-purine sequences. The H6(5) was easily identified because of its characteristic J_{56} coupling

Table 3. Chemical shifts of base and H1' protons of the (9+8)-mer duplex (3.5 mM, 295 K); shift reference Me_4NCl.

Residue	H8/H6	H2/H5	H1'
	ppm		
C(1)	4.978	2.893	2.420
U(2)	4.885	2.363	2.520
G(3)	4.530		2.570
G(4)	n.a.		n.a.
U(5)	4.636	2.550	n.a.
G(6)	n.a.		n.a.
C(7)	4.474	1.980	2.493
G(8)	4.305		2.563
G(9)	4.203		2.683
C(10)	5.004	2.887	2.426
C(11)	4.809	<2.473>[a]	
G(12)	4.474		n.a.
C(13)	4.395	2.187	2.577
C(14)	4.657	2.493	2.360
C(15)	4.683	2.450	2.333
A(16)	4.773	4.048	2.247
G(17)	4.149		2.697

n.a = not assigned
a. near isochronous resonances

constant (8.1 Hz) displayed by the uridine H6 resonance. Subsequently, the H5(5) is assigned via the intranucleotide H6-H5 NOE. No interresidue NOEs between U(5) and any of the protons of its neighbours, G(4) and G(6), can be detected. For example, it is clearly seen (line indicated with asterisk, Fig. 2) that no internucleotide NOEs between H5(5) and protons of its neighbours are present in this part of the spectrum. At this stage we tentatively suggest that the present NOE data point to the existence of a bulge-out of the extra U residue. Chemical-shift and imino-proton data presented below provide more supporting evidence for this proposal. The chemical shifts of the assigned protons are given in Table 3.

Chemical-shift analysis. Chemical shifts are known to be quite useful to monitor conformational phenomena like stacking and duplex formation. As these features appear temperature dependent one can follow the stack $\rightleftarrows$ unstack and duplex $\rightleftarrows$ monomer equilibria with the aid of chemical-shift vs

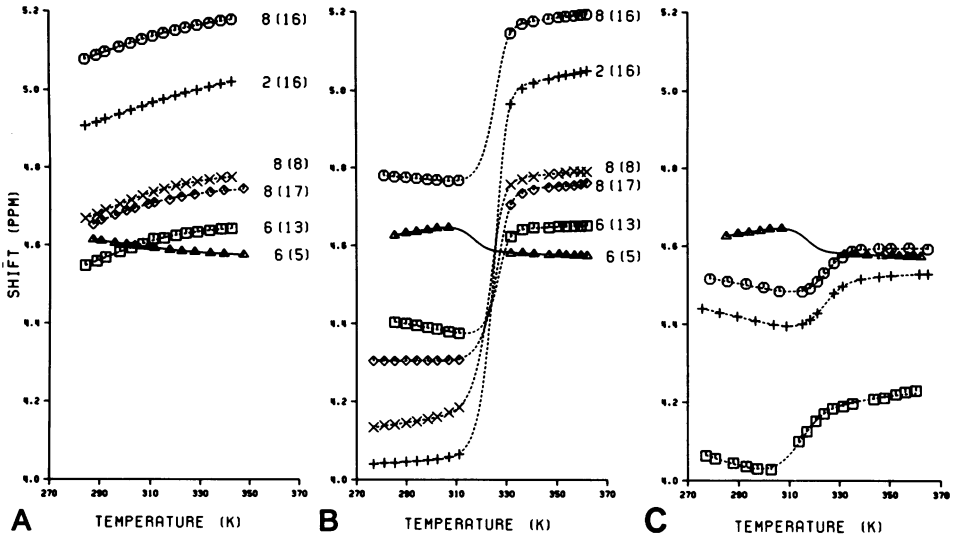


Figure 3. Chemical-shift vs temperature profiles of (A) several protons of the single-strands CCGCCCAG and CUGGUGCGG (B) several protons of the (9+8)-mer duplex (C) H6(5) of the RNA (9+8)-mer duplex (Δ); H6(7) of $(CACAUGUG)_2$ (O); H6(5) of $(CACAUGUG)_2$ (+); H6(5) of the DNA (9+8)-mer duplex (□).

temperature profiles. Because the (9+8)-mer duplex consists of two non-selfcomplementary strands, the shift profiles of the two single strands could be monitored separately. When stacking occurs, a noticeable shielding of the protons lying above or below a purine residue is expected. Some of the shift profiles obtained for the separate single-stranded octa- and nonamer are shown in Fig. 3a. It is seen that all protons, except H6(5), become more or less shielded at low temperature. This shielding remains in line with the usual gradual increase of stacking upon lowering the temperature. Since the U(5) residue in this sequence is located between the two G residues, one would expect a considerable shielding for the H6(5) proton in case this residue becomes stacked between its neighbours. The deviant behaviour of the H6(5) corresponds exactly to the behaviour of the H6(3) in UGUG [37] in which strand the U(3) was shown to prefer a bulged-out position. Thus, one may safely conclude that U(5) in single-stranded CUGGUGCGG takes up a bulge-out position.

In Fig. 3b the shift profiles of the corresponding protons in the (9+8)-mer duplex are shown. In general, these profiles display a sigmoidal shape with two plateau regions. The low-temperature plateau region corresponds to the duplex form, whereas the high-temperature plateau region represents the

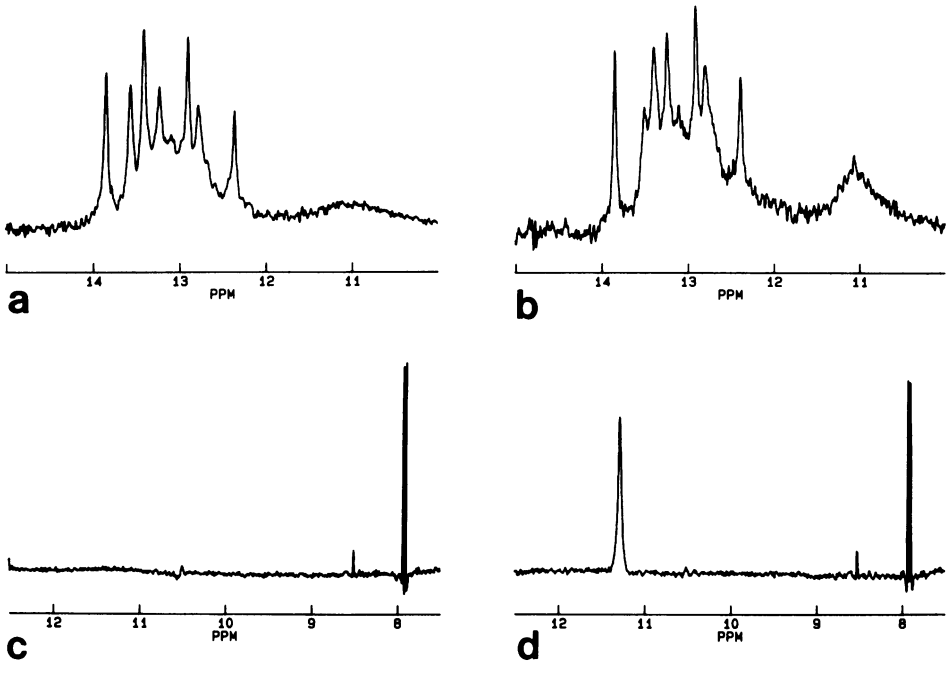


Figure 4. Imino-proton spectra of (A) the (9+8)-mer duplex, 3.5 mM, 280 K, pH 7.4 (B) the (9+8)-mer duplex, 3.5 mM, 280 K, pH 5.9 (C) the monomer mpUpm, 5mM, 280 K, pH 7 (D) the monomer mpUpm, 5mM, 280 K, pH 4.

monomer. Most protons display a large difference in chemical shift between monomer and duplex. A Tm (the midpoint of the duplex $\rightleftarrows$ monomer transition) of 320-325 K is observed for these protons at this nucleic-acid concentration (3.5 mM). These features consistently point to the formation of a stable duplex. However, an interesting exception occurs in the case of the chemical-shift profile of H6(5). One notes only a small change in chemical shift of H6(5) during the duplex $\rightleftarrows$ monomer transition. Moreover, this change occurs in an opposite direction compared to the behaviour of the other protons monitored. Let us now combine this piece of pertinent information with our knowledge of U's (or T's) that are known to enter into stacking interactions with neighbouring purines. In Fig. 3c the shift profile of H6(5) is compared with those displayed by H6(5) and H6(7) in $(CACAUGUG)_2$ [37] and by H6 of dT(5) in the (9+8)-mer DNA counterpart [19]. In the latter three profiles the respective H6 resonances become shielded upon duplex formation due to the ring-current anisotropy of the neighbouring A and G bases. Finally, the shift of the H6(5) in the present RNA (9+8)-mer

corresponds with its shift in the single strand at low temperature (Fig. 3a). These facts are taken as a strong evidence that the U(5) residue in the (9+8)-mer does not reside in the interior of the duplex.

In summary, we conclude that a stable duplex is formed between the octamer and nonamer, in which the unpaired uridine residue prefers to adopt a bulge-out position. Note the difference in behaviour between an unpaired dT residue which stacks into a DNA duplex [19] and an unpaired U residue which bulges out from a RNA duplex. Similar differences in behaviour were also observed in case of single-stranded RNA and DNA fragments [37]. In UĀUĀ and UGUG the U(3) residue was found to bulge out [37], whereas in d(TATA) [38] and d(TGTG) [39] the dT(3) is stacked in a normal fashion. The poor stacking proclivity of U compared to dT was proposed to account for these observations. In a similar vein we now suggest that the behaviour of an unpaired residue in a duplex is determined mainly by the stacking interactions between this particular residue and its neighbours.

Exchangeable protons. Fig. 4a shows the low-field part of the ^1H NMR spectrum of the (9+8)-mer duplex measured in H_2O/D_2O (90/10) at pH = 7.4, 280 K. The signal at 13.9 ppm corresponds well with the position usually found for A•U base pairs and is therefore ascribed to the only A•U pair present. Between 13.6 and 12.6 ppm several resonances are observed. The positions of these resonances remain in line with those reported for G•C base pairs and are collectively ascribed to the seven G•C pairs in the (9+8)-mer duplex. Finally, a very broad signal is observed at 11.1 ppm and this resonance is tentatively assigned to the N_3H of U(5).

Again, one notes the difference in behaviour between the present RNA (9+8)-mer duplex and its DNA counterpart studied earlier [19]. In the latter DNA (9+8)-mer a sharp imino-proton resonance was observed for the N_3H of dT(5) at 10.6 ppm at pH 7, in contrast to expectations for "free" (non-hydrogen-bonded) dT residues, which normally display a broad signal at 11.2 ppm at neutral pH [40]. Both the sharpness of the signal and the upfield shift of the N_3H of dT(5) compared to the "free" imino-proton supported our conclusion that the extra dT stacks into the duplex. However, in the RNA (9+8)-mer, the N_3H resonance of U(5) only becomes sharper when the pH is lowered (5.9 in Fig. 4b). This behaviour corresponds closely to that displayed by a "free" imino-proton of a uridine monomer. For example, in mpUpm at pH = 7 no imino-proton is observed (Fig. 4c), but a sharp signal is seen at pH = 4 (Fig. 4d). Moreover, the position of the U(5) in the (9+8)-mer accords well with that seen in the monomer. These data strongly suggest:

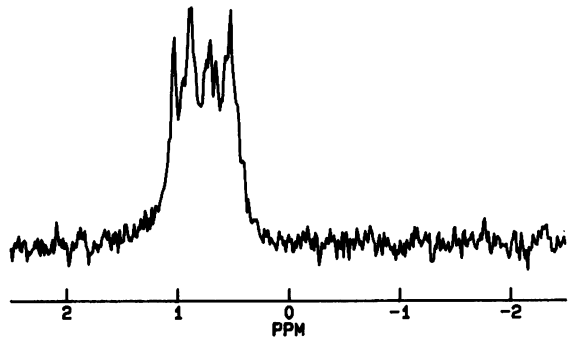


Figure 5. ^{31}P spectrum of the (9+8)-mer; 3.5 mM, 300 K, pH 7. Shifts are relative to 3'5'-cAMP.

(a) that the imino-proton of U(5) is not protected against rapid exchange with the solvent and (b) that this particular proton resides outside the shielding zone of the neighbouring G(4) and G(6) bases. These observations remain perfectly in line with our conclusion formulated above, *viz* that the extra U residue prefers to take up a position extraneous to the continuous helix.

^{31}P chemical shifts. ^{31}P chemical shifts can be used to obtain some pieces of information about the preferred conformations around the phosphodiester torsion angles ζ/α. In particular, the torsion angle combinations t/g and g/t for ζ/α can be distinguished from the g/g combinations [41]. This is of special interest in model-building studies because t/g or g/t combinations lead to an extended phosphodiester linkage. Note that in the single-stranded bulge-out model no such extended linkages appear present, *vide supra*.

In the (9+8)-mer under investigation all ^{31}P resonances are found clustered in a relatively narrow spectral region of 0.7 ppm (Fig. 5). This is taken to mean that no extended phosphodiester linkage, i.e. no t/g or g/t combinations for ζ/α, are present in the duplex. Therefore, it appears perfectly reasonable to use the calculated single-strand bulge-out structures as a starting point in model-building studies of the (9+8)-mer duplex.

Model building

Thus far, models for a duplex containing an unpaired base have been reported only for DNA fragments. Moreover, in these particular models the extra residue was stacked inside the duplex. Therefore, it seems desirable to construct a model for the (9+8)-mer duplex with a bulged-out U residue.

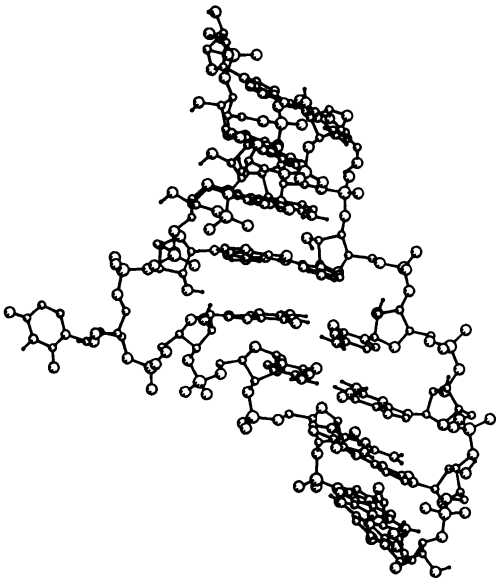


Figure 6. Energy-minimized structure of the (9+8)-mer duplex.

The features of this model should at least agree with the information obtained from the present NMR studies. The most important criteria which should be met are: (a) the extra U residue resides outside the duplex; (b) all base pairs remain intact; (c) no t/g or g/t combinations of torsion angles around ζ/α are present and (d) no unusually large distance is found between residues C(13) and C(14). The single-stranded bulge-out structure 2 (Table 2, Fig. 1) was used as a starting point for a model-building study. This model was fitted into a (9+8)-mer duplex with the aid of the program MacroModel [25]. The model thus obtained was energy-minimized with the all-atom version of the molecular-mechanics program AMBER [26]. The energy-minimized structure, which fits all our experimental criteria, is shown in Fig. 6. The backbone torsion angles, sugar-ring pseudorotation parameters and glycosyl torsion angles are listed in Table 4. From both Fig. 6 and Table 4 it is seen that a continuous helix is formed in which the bulge manifests itself as a local phenomenon. Comparison of the torsion angles of the G(4)-U(5)-G(6) part with those of the single-stranded model 2 (Table 2) reveals that only minor changes are required to incorporate the single-stranded fragment into a helix. The torsion angles of the remaining residues of the (9+8)-mer appear within the ranges normally observed for A-RNA i.e.: α^-, β^t, γ^+, N-type sugar, ϵ^-, ζ^- and χ anti. In addition all base pairs

Table 4. Torsion angles of the model of the (9+8)-mer duplex after energy-minimization (ε = R). For notation see scheme I and [27,28].

Strand	Residue	α	β	γ	δ	ε	ζ	χ	P	ϕ_m
		degrees								
I	C(1)			46	72	250	276	-170	5	44
	U(2)	283	153	67	70	200	301	-165	16	46
	G(3)	287	175	71	69	198	294	-170	16	39
	G(4)	293	185	69	139	277	260	-142	156	39
	U(5)	53	181	56	140	281	302	-152	146	44
	G(6)	312	186	88	67	201	288	-168	31	45
	C(7)	288	186	61	77	199	295	-169	3	42
	G(8)	284	180	65	76	197	299	-166	8	42
	G(9)	280	185	56	83			-157	17	38
II	C(10)			47	70	248	284	-171	4	46
	C(11)	288	172	69	68	197	297	-160	20	46
	G(12)	283	172	67	72	198	302	-173	11	46
	C(13)	288	182	71	71	193	297	-172	11	45
	C(14)	289	182	69	75	196	295	-168	14	42
	C(15)	287	186	66	74	197	292	-163	11	42
	A(16)	287	180	65	78	197	299	-163	15	34
	G(17)	280	184	56	87			-152	20	36

remain intact and the distance between C(13) and C(14) appears normal (3.6 Å). In summary, a bulged-out U residue seems to behave as a strictly local disturbance of a regular A-type RNA helix.

CONCLUSIONS

Several bulge-out structures for single-stranded trimers have been generated earlier [12,13], but none of these models allowed extension of stacking interactions in both the 3' and 5' direction and at the same time fitted all the NMR data. In the present study we report several models which not only agree with the geometrical characteristics deduced from NMR studies but also easily fit into a continuous RNA helix.

In addition, the duplex between (CUGGUGCGG) and (CCGCCCAG), which contains an extra U residue, was investigated experimentally. All data (NOE, chemical-shift and imino-proton information) indicate that the U remains bulged out of an otherwise intact duplex. Only minor changes in torsion angles of the single-strand bulge-out model were necessary to construct a

bulge-out in this (9+8)-mer duplex. A remarkable difference between RNA and DNA is noted: in the RNA (9+8)-mer the extra U is bulged out, whereas in the DNA counterpart the extra dT is stacked into the duplex [19]. These observations correspond with those previously reported for single-stranded RNA and DNA fragments [37]. It is suggested that the difference in stacking proclivity between U and dT mainly determines the difference in behaviour of the unpaired bases in the (9+8)-mer duplexes.

ACKNOWLEDGEMENTS

This research was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO). Spectra were recorded on the 500 MHz spectrometer in the Department of Chemistry in Groningen and on the 300-400 MHz NMR facility in the Department of Chemistry at Leiden. We wish to thank Drs. C. Erkelens, Dr. R.M. Scheek and Ing. K. Dijkstra for technical assistance. Dr. R.M. Liskamp provided the opportunity to perform model-building studies with the program MacroModel at Leiden; Dr. E. Westhof and Mr. P. Dumas (Laboratoire de Cristallographie Biologique, Strasbourg) and Dr. H.E. Bosshard (EMBL, Heidelberg) gave invaluable assistence during our first modelling attempts.

REFERENCES

1. This is part 59 of the series Nucleic Acid Constituents of this laboratory. For part 58 see [19].
2. Abbreviations used: $\bar{A}$, N^6-dimethyladenosine; mpUpm, 3',5'-bis-(methyl phosphate) ester of uridine; Me_4NCl, tetramethylammoniumchloride; DSS, sodium 2,2-dimethyl-2-silapentane-5-sulphonate; NOESY, two-dimensional nuclear Overhauser spectroscopy.
3. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol. 31*, 77-84.
4. Peattie, D.A., Douthwaite, S., Garrett, R.A. & Noller, H.F. (1981) *Proc. Natl. Acad. Sci. USA 78*, 7331-7335.
5. Romaniuk, P.J., Lowary, P., Wu, H.-N, Stormo, G. & Uhlenbeck, O.C. (1987) *Biochemistry 26*, 1563-1568.
6. Schmelzer, C. & Schweyen, R.J. (1986) *Cell 46*, 557-565.
7. White, S.A. & Draper, D.E. (1987) *Nucleic Acids Res. 15*, 4049-4064.
8. De Wachter, R., Chen, M.-W. & Vandenberghe, A. (1984) *Eur. J. Biochem. 143*, 175-182.
9. Lee, C.-H. & Tinoco, I. Jr. (1980) *Biophys. Chem. 11*, 283-294.
10. Lee, C.-H. (1983) *Eur. J.Biochem. 137*, 347-356.
11. Lankhorst, P.P., Groeneveld, C.M., Wille, G., van Boom, J.H., Altona, C. & Haasnoot, C.A.G. (1982) *Recl.Trav.Chim.Pays-Bas 101*, 253-263.
12. Altona, C. in *Structure and Dynamics of RNA* (1986) (Van Knippenberg, P.H. & Hilbers, C.W. Eds.) NATO ASI Series A, Life Sciences, vol. 110, pp 15-30, Plenum Press, New York.

13. Lee, C.-H (1983) *Eur. J. Biochem. 137*, 357-363.
14. Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Itakura, K. & Breslauer, K.J. (1982) *Biochemistry 21*, 445-451.
15. Hare, D., Shapiro, L. & Patel, D.J. (1986) *Biochemistry 25*, 7456-7464.
16. Woodson, S.A. & Crothers, D.M. (1987) *Fifth Conversation in Biomolecular Stereodynamics* (Ed. Sarma, R.H.) State University of New York at Albany, Adenine Press, p 64.
17. Roy, S., Sklenar, V., Appella, E. & Cohen, J.S. (1987) *Biopolymers 26*, 2041-2052.
18. Morden, K.M., Chu, Y.G., Martin, F.H. & Tinoco, I. Jr. (1983) *Biochemistry 22*, 5557-5563.
19. Van den Hoogen, Y.Th., van Beuzekom, A.A., van den Elst, H., van der Marel, G.A., van Boom, J.H. & Altona, C. *Nucleic Acids Res. in press.*
20. Van den Hoogen, Y.Th., van Beuzekom, A.A. & Altona, C. (1987) *Fifth Conversation in Biomolecular Stereodynamics* (Ed. Sarma, R.H.) State University of New York at Albany, Adenine Press, pp 274-275.
21. De Vroom, E., Fidder, A., Marugg, J.E. van der Marel, G.A. & van Boom, J.H. (1986) *Nucleic Acids Res. 14*, 5885-5900.
22. Haasnoot, C.A.G. & Hilbers, C.W. (1983) *Biopolymers 22*, 1259-1266.
23. Roth, K., Kimber, B.J. & Feeney, J. (1980) *J. Magn. Res. 41*, 302-309.
24. Mellema, J.-R, van Kampen, P.N., Carlson, C.N., Bosshard, H.E. & Altona, C. (1983) *Nucleic Acids Res. 11*, 2893-2905.
25. Copyright Columbia University (1986) New York, NY.
26. Weiner, S.J., Kollman, P.A., Nguyen, D.T., & Case, D.A. (1986) *J. Comp. Chem. 7*, 230-235.
27. IUPAC-IUB Nomenclature Commission (1983) *Eur. J. Biochem. 131*, 9-15.
28. IUPAC-IUB Nomenclature Commission (1986) *J. Biol. Chem. 261*, 13-17.
29. Hartel, A.J., Wille, G., van Boom, J.H. & Altona, C. (1981) *Nucleic Acids Res. 9*, 1405-1423.
30. Lankhorst, P.P., Haasnoot, C.A.G., Erkelens, C. & Altona, C. (1984) *J. Biomol. Struct. Dyns. 1*, 1387-1405.
31. Van den Hoogen, Y.Th., Treurniet, S.J., Roelen, H.C.F.P., de Vroom, E., van der Marel, G.A., van Boom, J.H. & Altona, C. (1988) *Eur. J. Biochem. 171*, 155-162.
32. Lankhorst, P.P., Wille, G., van Boom, J.H., Haasnoot, C.A.G., & Altona, C. (1983) *Nucleic Acids Res. 11*, 2839 - 2856.
33. Bell, R.A. Everett, J.R., Hughes, D.W., Coddington, J.M., Alkema, D., Hader, P.A. & Neilson, T. (1985) *J. Biomol. Struct. Dyns. 2*, 693-707.
34. Haasnoot, C.A.G., Westerink, H.P., van der Marel, G.A. & van Boom, J.H. (1983) *J. Biomol. Struct. Dyns. 1*, 131-149.
35. Scheek, R.M., Boelens, R., Russo, N., van Boom, J.H. & Kaptein, R. (1984) *Biochemistry 23*, 1371-1376.
36. Clore, G.M. & Gronenborn, A.M. (1985) *FEBS Lett. 179*, 187-198.
37. Van den Hoogen, Y.Th., Erkelens, C., de Vroom, E., van der Marel, G.A., van Boom J.H. & Altona, C. (1988) *Eur. J. Biochem. in press.*
38. Mellema, J.-R., Pieters, J.M.L., van der Marel, G.A., van Boom, J.H., Haasnoot, C.A.G. & Altona, C. (1984) *Eur. J. Biochem. 143*, 285-301.
39. Rinkel, L.J., van der Marel, G.A., van Boom, J.H. & Altona, C. (1987) *Eur. J. Biochem. 166*, 87-101.
40. Haasnoot, C.A.G., den Hartog, J.H.J., de Rooij, J.F.M., van Boom, J.H. & Altona, C. (1979) *Nature 281*, 235-236.
41. Gorenstein, D.G., Findlay, J.B., Momii, R.K., Luxon, B.A. & Kar, D. (1976) *Biochemistry 15*, 3796-3803.