-CH₂- lengthening of the internucleotide linkage in the ApA dimer can improve its conformational compatibility with its natural polynucleotide counterpart

J. Hanuš¹,², *, I. Barvík¹, K. Ruszová-Chmelová¹,², J. Štěpánek¹, P.-Y. Turpin², J. Bok¹, I. Rosenberg³ and M. Petrová-Endová³

¹Charles University, Institute of Physics, Ke Karlovu 5, CZ-121 16 Praha 2, Czech Republic, ²Université Pierre et Marie Curie, LPBC (CNRS URA 2056), 4 Place Jussieu, Case 138, F-75252 Paris Cedex 05, France and ³Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, CZ-166 10 Praha 6, Czech Republic

Received June 28, 2001; Revised and Accepted October 22, 2001

ABSTRACT

The complete family of ApA phosphonate analogues with the internucleotide linkage elongated by insertion of a -CH₂- group was prepared and the hybridisation and structural properties of its members in interaction with polyuridylic acid were investigated using an original 2D Raman approach. Except for the conformationally restricted A⁴¹⁵pA(2′3′endo-5′) modification, all of the isopolar, non-isosteric analogues form triplex-like complexes with poly(rU) at room temperature, in which two polymer strands are bound by Watson–Crick and Hoogsteen bonds to a central pseudostrand consisting of a ‘chain’ of A-dimers. For all of these dimers, the overall conformation of the triplexes was found to be similar according to their extracted Raman spectra. A simple semi-empirical model was introduced to explain the observed dependency of the efficiency of triplex formation on the adenine concentration. Apparently, for most of the modifications studied, the creation of a stable complex at room temperature requires the formation of a central pseudostrand, consisting of several adenine dimers. Molecular dynamics calculations were finally performed to interpret the differences in ‘cooperative’ behaviour between the different dimers studied. The results indicate that the exceptional properties of the ApCH₂A(3′-5′) dimer could be caused by the 3D conformational compatibility of this modified linkage with the second (Hoogsteen) poly(rU) strand.

INTRODUCTION

Interactions between long nucleic acid chains (NA) and (oligo)nucleotides have been extensively studied since the 1960s to better understand the processes involved in NA strand hybridisation. The polyuridylic acid [poly(rU)] interaction with adenosine mono- and oligonucleotides has often served as a model system, since these compounds satisfy important criteria: the aqueous solubilities of the polymer and adenosine mono- and oligomers are high; poly(rU) has a minimal degree of self-interaction; the poly(rA)–poly(rU) duplexes are well characterised (1).

These interactions have been monitored under various physicochemical conditions by a variety of methods, among them UV absorption, IR, CD and optical rotation, calorimetry and equilibrium dialysis (2,3). It has been found that poly(rU) and adenosine (oligo)nucleotides have a strong tendency to form triplex structures with a 2:1 U:A stoichiometry whose stability depends on the concentration of monovalent and divalent ions in solution.

The simplest model for these studies is represented by the poly(rU) + adenosine system. Under the given salt conditions, the stability of poly(rU)–adenosine complexes essentially depends on the total concentration of adenosine in solution. The formation of a complex starts above a threshold concentration of adenosine and, moreover, the relative efficacy of adenosine binding to poly(rU) increases with the concentration of free adenosine in solution, i.e. the process of adenosine binding to poly(rU) has a cooperative character (4). This can be explained in the framework of a nearest neighbour theory: at moderate concentrations the first adenosine molecules serve as initiators for subsequent binding of other adenosine molecules to poly(rU) in some type of ‘cooperative’ manner (5–7). Interesting results concerning this initiation effect have also been obtained for interactions of poly(rU) with adenosine oligonucleotides (8,9).

A new impulse for re-opening the studies of polynucleotide interactions with some of their shorter counterparts appeared at the beginning of the 1980s with the introduction of the antisense drug strategy (10–15). In this perspective, since new ‘antisense-suitable’ internucleotidic backbone designs were sought, it seemed advantageous to test the properties of
synthetic compounds at the level of dinucleotides, which are readily obtainable with a wide variety of linkage modifications: thus, the relation between the primary structure of the linkage and the association properties of nucleoside dimers with their target counterparts can be easily studied for various families of modifications (16).

An interesting group of modified linkages can be achieved by inserting a methylene group in the phosphodiester linkage. This provides a family of non-isosteric but isopolar modifications, consisting of modified 2′-5′, 3′-5′ and even conformationally restricted 2′3′-5′ linkages. The striking nucleolytic stability of the phosphonate -O-P-CH2-O- type bond has been demonstrated in several syntheses of potent mononucleotide antivirals (17,18), as well as in studies of modified oligonucleotides (19). Despite the lengthening of the internucleotide backbone, members of this modification family remain able to hybridise with their phosphodiester counterparts (20).

The molecular structure of the simplest model of the phosphonate linkage, i.e. methyl-methoxymethylphosphonate (CH3-O-PO2-CH2-O-CH3), has been studied by ab initio quantum chemical calculations (21). It has been found that the secondary structure of an oligonucleotide is not drastically affected by phosphonate modification of the internucleotidic linkage (22). Structural analogy with the natural dinucleoside monophosphates has been shown in a 1H NMR study of some structural parameters of various modified nucleoside dimers (23). On the other hand, the only crystal structure yet known of a dinucleoside monophosphate analogue, i.e. guanosine-2′-O-(phosphono)methyl-O-5′-cytidine (24), indicates a high conformational flexibility of the linkage by distinguishing two different conformations of this dimer in the crystal.

Raman spectroscopy studies of A-A and A-U dimers bearing a phosphonate linkage indicated that neither the tendency of the dinucleotide bases to stack (25) nor their ability to bind to the natural ribonucleic acid chain (26) is reduced by the modification. The features of some of these phosphonate linkages have also been compared in a Raman study of triplex-like complexes formed between A-A modified dimers and natural poly(rU) (27). Significant differences were found in the relation between the efficiency of complex formation and the modified A-dimer concentration, depending on the type of modification.

The present work presents a systematic study of the process of triplex formation for a complete series of phosphonate modifications of the internucleotidic linkage (Fig. 1). Newly developed methods for treatment of a series of Raman spectra have been used, enabling the total vibrational information over the complete Raman spectra to be used for determining, much more accurately than before, the fractional amounts of the triplexes formed, their proper vibrational features and the various stability constants of formation. Moreover, molecular dynamics (MD) simulations of fully solvated species have also been performed to interpret the experimental results.

**MATERIALS AND METHODS**

Following a previously described method (27), a set of six diribonucleoside monophosphate analogues was synthesised, including 3′-5′ dimers modified by inserting a -CH2-group on the 3′- or 5′-side of the phosphate, 2′-5′ dimers also differing in the position of the inserted methylene group and both possible stereoisomers of the semi-rigid 2′3′-5′ modification (see Fig. 1). Natural ApA(3′-5′) and ApA(2′-5′) dimers were purchased from Sigma. All samples were purified and adjusted in the same manner, as described (27). Poly(rU) (potassium salt) was purchased from Pharmacia. NaCl and NaOH and MgCl2 (1 M) and HCl (32%) solutions of 99.5% purity were purchased from Fluka Biochemika and KH2PO4 (99.0% purity) from Sigma.

**Sample preparation**

A set (15–26 samples) of A-dimer/poly(rU) mixtures was prepared in the same manner as reported (27) for each dimer.
under study. The total nucleoside concentration (20 mM) and content of monovalent (Na\(^+\) + K\(^+\)) (120 mM) and divalent Mg\(^{2+}\) ions (2 mM) were kept constant, while the adenosine:uridine ratio gradually varied from pure poly(rU) to pure A-dimer in solution.

**Raman measurements**

Raman spectra were recorded in a Jobin Yvon T64000 CCD Raman spectrometer using the 488 nm Ar\(^+\) laser line for excitation. The laser power at the sample was 180 mW. Spectra of the samples, placed in a 10 µl microcell, were measured in a 90° scattering geometry. The total integration time per pixel was 500 s. Spectra of the samples within a particular series were measured in a random order to eliminate possible systematic errors. Spectra were not smoothed before analysis.

**Treatment of experimental data**

Methods based on factor analysis (FA) were used for the quantitative analysis of the spectral series. This method enables essential information to be extracted from a large set of data (28). It allows the original spectral set to be approximated by optimal linear combinations of suitably constructed ‘base’ (orthogonal) spectra. These so-called \( S_j \) subspectra are obtained along with a set of weights, \( w_j \), that express the statistical importance of each subspectrum in the original spectral series and a set of scores, \( V_j \), that represent normalised coefficients representing the relative presence of the \( S_j \) subspectrum in each experimental spectrum \( Y_i \). Experimental spectra \( Y_i \) are thus expressed as:

\[
Y_i = \sum_{j=1}^{n} w_j V_j S_j
\]

The approximation of the complete original set of spectra by the most significant \( m \) subspectra of FA is the best possible approximation obtainable by any reduced set of \( m \) orthogonal spectra. From the residual error of this approximation, the factor dimension of the problem can be derived, i.e. the number of independent spectra that are necessary to adequately reconstruct the original data set. For the current series of Raman spectra, this factor dimension ought to correspond to the number of spectrally distinguishable species present in the samples.

However, statistical methods of Raman data treatment, such as FA, are highly sensitive not only to the intrinsic information embedded in a series of spectra, but also to numerous possible artefacts, such as drift of the wavenumber scale, changes in the luminescence background and variations in the signal intensity or deformations caused by manual solvent subtraction. In the present work an original approach has been used that provides considerably better results than other methods commonly used in Raman spectroscopy. To correct the drifts of the \( x \)-scale, we employed an original automatic recalibration procedure, using a neon glow lamp spectrum measured after every single sample spectrum. The luminescence background was corrected and a solvent spectrum (50 mM phosphate buffer) was semi-automatically subtracted using a new POMLAZKA program developed in our laboratory. POMLAZKA also enables the sample spectra to be corrected for weak contributions of the capillary glass and to be normalised using the solvent spectrum as an intensity standard.

**Molecular dynamics simulations**

MD simulations of some fully solvated triplex model structures were carried out to support our interpretation of the experimental results. Initial structures were constructed using the Biopolymer module from the BIOSYM software package. CH\(_3\) groups were manually inserted into the backbone. Explicit Na\(^+\) counterions were placed at the phosphates/phosphonates of the nucleotides by the Edit module of AMBER 5.0. Nucleic acids with Na\(^+\) counterions were surrounded by a periodic box of \( \sim 4000 \) TIP3P (29) water molecules, which extended to a distance of \( \sim 10 \) Å (in each direction) from the nucleic acid atoms. This leads to a periodic box size of \( \sim 60 \times 40 \times 40 \) Å\(^3\).

The AMBER force field (29) does not contain the force constants needed to describe the modified parts of the internucleotidic linkages of the \( 3^{\text{CH}_2}pS' \) and \( 3'p^{\text{CH}_2}S' \) analogues. The force field parameters were therefore completed with those obtained from ab initio calculations on methylmethoxyphosphonate published elsewhere (30,31).

MD simulations were started with a wide set of initial structures covering the conformational space of modified segments. Fully solvated trajectories were computed using the SANDER module of the AMBER 5.0 software package, using the implemented Particle Mesh Ewald summation method for electrostatic interactions (32). The usual computational procedures and equilibration protocols enabling bad contacts to be avoided and poor bond, angle and dihedral deviations to be reduced in the modelled structure, were used as described (http://www.amber.uchicago.edu/amber/ tutorial/polyA-polyT). Finally, the system was heated from 100 to 300 K over 5 ps and then the production runs were initiated. MD trajectories were examined using the Analysis/BIOSYM module (torsion angles, hydrogen bond connections) and Dickerson’s NEWHELIX program (helical parameters).

**RESULTS AND DISCUSSION**

**Determination of the complexes**

For each type of A-dimer a set of Raman spectra corresponding to poly(rU)/dimer mixtures with various adenosine:uridine ratios was obtained. FA applied to each set led to a factor dimension equal to three, indicating the formation of only one stoichiometric type of complex: the three independent spectral components correspond to unbound A-dimer, unbound poly(rU) and the complex. By comparing the spectral shapes of the FA components (see below), closely similar local arrangements of the complexes can easily be inferred for all types of internucleosidic linkage modification and, considering our previous results (27), we can assume that these complexes are actually triplexes.

A single exception was observed: for the \( A^{3'pA}(2'3'\text{-endo}-5') \) (no. 8) dimer a factor dimension of only two was obtained. The spectra of its various mixtures with poly(rU) were equal to the concentration weighed sums of the spectra from unbound species. This constitutes a spectroscopic confirmation of non-interacting components, i.e. the \( A^{3'pA}(2'3'\text{-endo}-5') \) dimer has been proved not to bind to poly(rU).

Our aim was to determine the triplex concentrations in every mixed solution with the highest possible precision. It is worth pointing out that the features of a Raman spectrum are mostly sensitive to the local environment of the molecule under study,
therefore the experimental data will provide information in terms of nucleoside units (adenosines and uridines) bound in the triplex structure. In contrast to our previous work (27), where the concentrations of bound and unbound species were determined by monitoring only two Raman marker bands, here the whole spectral information contained in each complete Raman spectrum was employed. Let us present the procedure of data processing which was developed and applied.

It is expected that for a particular A-dimer the set of Raman spectra \( Y(v) \) corresponding to its various mixtures with poly(rU) can be represented by various superpositions of the spectra of unbound A-dimer and poly(rU), on the one hand, [they can be directly measured and denoted, normalised for unit concentration, as \( A(v) \) and \( U(v) \)] and of the spectrum of the triplex [a priori unknown and denoted as \( T(v) \)], on the other:

\[
Y(v) = c_{A\text{free}}^i A(v) + c_{U\text{free}}^i U(v) + c_i^T T(v) \quad i = 1, 2, \ldots, N
\]

where \( N \) is the number of measured mixed solutions. \( c_{A\text{free}}^i \), \( c_{U\text{free}}^i \) and \( c_i^T \) represent the concentrations of unbound adenosine, unbound uridine and triplex, respectively. They are related to the (known) total nucleoside concentrations of each species in the sample \( (c_A^i \) and \( c_U^i \)) via simple relations:

\[
c_{A}^i = c_{A\text{free}}^i + c_i^T \quad c_{U}^i = c_{U\text{free}}^i + 2c_i^T \quad i = 1, 2, \ldots, N
\]

The new approach is based on the spectra calculated as differences between the actual experimental spectra and the sum of the unbound species spectra weighted by the total concentration of each nucleoside:

\[
\text{Diff}(v) = Y(v) - c_{A}^i A(v) - c_{U}^i U(v) \quad i = 1, 2, \ldots, N
\]

\( \text{Diff}(v) \) spectra actually represent only those spectral changes caused by formation of the complex. In the case of no interaction between the two species, \( \text{Diff}(v) \) is expected to be null. The significance of these difference spectra is illustrated in Figure 2. Spectra provided by the set of mixed poly(rU):ApA(3'–5') samples are drawn in a rainbow colour scale, from violet for pure adenosine dimer to red for pure poly(rU). While the \( Y(v) \) Raman spectra (Fig. 2A) reflect primary changes in the nucleoside composition and the dominant bands are either of violet-blue or red-yellow colour (corresponding to samples where one of the nucleosides prevails), the maxima of the \( \text{Diff}(v) \) spectra (Fig. 2B) are of greenish colour (corresponding to samples with the highest content of complexes). Figure 2 also shows that various changes in the spectral features of the straight Raman spectra (changes in the main band intensities and less visible changes in some marker band positions) are reduced to the single spectral shape of the \( \text{Diff}(v) \) spectrum, whose overall intensity only varies with the A-dimer:poly(rU) ratio.

An important relation follows from equations 2–4:

\[
\text{Diff}(v) = c_i^T [T(v) - A(v) - 2U(v)] = c_i^T \Delta T(v) \quad i = 1, 2, \ldots, N
\]

This indicates that in the case of only one type of complex, \( \text{Diff}(v) \) spectra should have the same spectral shape \( \Delta T(v) \) and their amplitudes should be proportional to the complex concentration \( c_i^T \) in the ith sample: this is actually observed in Figure 2B.

After calculating the \( \text{Diff}(v) \) difference sets for all of the investigated A-dimers, these sets have been in turn treated by FA: except for the A\(^{\text{Cl}}\)pA(2'3'-endo-5') (no. 8) dimer, which forms complexes very poorly, a factor dimension equal to one has been obtained. This again proves the presence of only one type of complex. Optimal normalised spectra \( \Delta T(v) \) can then be derived from the first FA spectral component (subspectrum)

\[
\Delta T(v) = (1/\gamma) S_{yi}(v)
\]

where \( \gamma \) is a proportionality coefficient between the triplex concentration and the weighted FA score for each sample:

\[
c_i^T = \gamma w_i V_{yi} \quad i = 1, 2, \ldots, N
\]

The dependency of the weighted scores \( w_i V_{yi} \) on the fraction of adenosine in the mixture yields a maximum located between 30 and 40% adenosine for all the dimers under study [except for A\(^{\text{Cl}}\)pA(2'3'-endo-5')], which is expected for a 1:2 A:U stoichiometry of the complexes.

To complete the data analysis it is necessary to determine the complex concentrations, i.e. the \( \gamma \) factor value. In a previous paper it has been shown that the most intense Raman features of the triplexes formed by the various A-dimers are indeed similar (27), thus one can expect that the \( \gamma \) value be practically the same for all triplexes. Upper and lower estimations of \( \gamma \) may be performed. The upper one is limited by the maximum possible concentration of the triplex, which must not exceed the total concentration of adenosine and half that of uridine. The samples yielding the highest FA scores, i.e. Ap\(^{\text{Cl}}\)pA(3'–5') (no. 3) and Ap\(^{\text{Cl}}\)pA(2'–5') (no. 4) dimers, led to \( \gamma \leq 22 \times 10^{-6} \) A reasonable minimum value of \( \gamma \) can be found from a construction of the pure triplex spectrum. From relations 5 and 6 it follows that the triplex spectrum \( T(v) \) can be constructed using the first subspectrum \( S_{yi}(v) \) provided by FA of the \( \text{Diff}(v) \) set and the normalised spectra of unbound A-dimer and poly(rU):

\[
T(v) = A(v) + 2U(v) + [(1/\gamma) S_{yi}(v)]
\]

The minimum \( \gamma \) value which does not give rise to negative peaks in such a constructed \( T(v) \) spectrum is \( \gamma_{\text{min}} = 20 \times 10^{-6} \). Consequently, a value of \( \gamma = (21 \pm 1) \times 10^{-6} \) has been used to determine the concentrations of the triplexes and their normalised Raman spectra from the results of FA applied to the sets of \( \text{Diff}(v) \) spectra for all the dimers except A\(^{\text{Cl}}\)pA(2'3'-endo-5').

Analysis of the triplex spectra

Although the various modified A-dimers may differ from each other in their binding affinity for the natural poly(rU) counterpart, the spectra of the triplexes (constructed using relation 8) are practically identical (Fig. 3B), except in the 830–930 cm\(^{-1} \) region corresponding to the sugar–phosphate backbone vibrations (similar variations can also be found in the spectra of the pure A-dimers) due to the differences in the primary structure of the linkages.

One of the main spectral features distinguishing all the triplex spectra (Fig. 3B) from the mere sum of the spectra of the free components is the band at 813 cm\(^{-1} \). This band is the symmetrical stretching mode of the sugar–phosph(on)ate backbone and is known as a conformational marker of the C3'–endo puckering of the furanose ring (33,34) found in the A-genus NA conformation. The presence of this band in the current triplex spectra (and the absence of a 842 cm\(^{-1} \) band, marker of C2'-endo puckering; 33,34) is the signature of the dominant C3'-endo furanose puckering in poly(rU) and ApA dimers for all of the analogues under study. In the low frequency region it
is worth noting the presence of one additional, medium intensity band at 630 cm$^{-1}$ due to triplex formation, instead of broad, weak bands around 640 cm$^{-1}$ in both the A-dimer and poly(rU) spectra.

The formation of hydrogen bonds between adenine and uridine bases is nicely reflected in a significant downshift of the 1626 cm$^{-1}$ uridine band, which is representative of coupled stretching vibrations of the C5=C6 with the C4=O carbonyl double bond and thus is directly influenced by hydrogen bond formation between the bases. This band moves to $\sim 1617$ cm$^{-1}$ in all of the triplex spectra. This is due to the replacement of the hydrogen bond to water by a hydrogen bond to the adenine NH$_2$ group combined with dipole coupling between vibrations of the neighbouring U residues (35).

In this region another strong and broad uridine band can be found, lying in pure poly(rU) spectra at $\sim 1685$ cm$^{-1}$ and arising from several overlapping bands related to the vibrations of both carbonyl groups of U and their coupling. In the triplex spectra a strong decrease in its intensity can be observed (Fig. 3B) along with splitting into several peaks at 1693 cm$^{-1}$, with a shoulder at 1672 cm$^{-1}$ and a band at 1727 cm$^{-1}$. Again, this spectral pattern, including the relative intensities of the bands, is very similar for all of the complexes formed.

Another important feature of the triplex spectra (Fig. 3B) is the hypochromism of the bands related to the main ring modes of base vibrations, namely the ring breathing vibrations at 729 (A) and 782 cm$^{-1}$ (U), uracil bands at 1230 and 1396 cm$^{-1}$ and

---

**Figure 2.** (A) Raman spectra of poly(rU):ApA(3'–5') mixtures (after background and solvent subtraction and normalisation). (B) Difference spectra calculated from the Raman spectra shown in (A), according to relation 4.
The ring breathing vibrations at 782 (U) and 729 cm\(^{-1}\) (A), already mentioned for their hypochromism, exhibit a different behaviour. While the U mode remains at 782 cm\(^{-1}\), the adenine breathing mode changes its frequency in the triplex. The dependence of this band on the type of linkage in the dimer is interesting: for the \(\text{3'-5'}\) linkage it shifts down to 726 cm\(^{-1}\) in the triplex while for the \(\text{2'-5'}\) linkage it shifts up to 733 cm\(^{-1}\) and is accompanied by a shoulder at \(\sim 720\) cm\(^{-1}\). For \(\text{A}^\text{CH2A}(2'3'\text{exo-5'})\) this mode shifts slightly up to 730 cm\(^{-1}\), with a well-pronounced shoulder at 720 cm\(^{-1}\).

**Efficiency of triplex formation**

Figure 4 shows the dependencies of the complex concentrations on the adenosine fraction in the mixtures for the various modified A-dimers. The top left graph shows theoretical curves obtained from the equilibrium equation for the direct formation of triple helical complexes from independent units (see equation 9). In the case of extremely high affinity (stability constant \(K = 10^{12}\) L mol\(^{-1}\)), the shape is almost triangular, with a sharp maximum at 33% adenosine. This corresponds to almost ideal triplex formation, i.e. when all the possible binding sites are occupied. As the stability constant is decreased, the shape becomes more rounded and the maximal concentration of triplexes also decreases. However, the position of the maximum remains at 33% adenosine.

The other graphs in Figure 4 show the measured concentrations of the complexes (symbols) compared to the theoretical curve for ideal complex formation (line). Full and empty symbols correspond to experiments carried out at room temperature (20°C) and 4°C, respectively. Let us first discuss the experiments carried out at room temperature.

From a comparison of the theoretical curve with the experimental data, we can distinguish two different groups of samples. The first family consists of two modified dimers where the methylene group has been inserted closer to the 5'-position of the phosphate group, i.e. \(\text{ApCH2A}(2'5')\) (no. 4) and \(\text{ApCH2A}(3'5')\) (no. 3) (bottom graphs of Fig. 4). These samples behave in agreement with the theoretical equilibrium equations and the efficiency of complex formation remains high whatever the A-dimer concentration. In the second family of samples, including natural \(\text{ApA}(3'5')\), only weak complex formation can be observed at low A-dimer concentration. However, the efficiency of triplex formation increases when the A-dimer concentration rises. This "retardation" in formation of triplex as a function of A-dimer concentration resembles some kind of cooperative effect. Because of this, the maximal possible concentration of triplexes was reached at a higher adenosine fraction (>33%) and was lower than for the first family; this effect was extremely pronounced for natural \(\text{ApA}(2'5')\), whose overall complex stability at room temperature seems to be rather low.

For some samples of the second family, which showed cooperative behaviour at room temperature, we also carried out experiments at 4°C. The results (empty symbols in Fig. 4) showed not only the expected increase in complex stability (i.e. a higher maximal concentration), but also disappearance of the retardation ("cooperative") effect: the maximal concentration of complexes was reached at a 1:2 A:U ratio and complex formation was efficient even at low adenosine concentrations. The difference between the behaviours at room and low temperature was especially dramatic for \(\text{ApA}(2'5')\) (no. 2),
but may also be clearly observed for the other two samples [ApA(3′-5′) (no. 1) and ACHpA(2′-3′exo-5′) (no. 7)].

Analysis of the dependency of triplex formation efficiency on sample composition

A detailed interpretation of the concentration dependencies obtained from the sets of Raman spectra (Fig. 2) requires consideration of the possible ways A-dimers can bind to poly(rU). Particular complex forms can be characterised in terms of stability constants.

The first, simplest approximation corresponds to a theoretical binding of one molecule of A-dimer to two U-dimers forming the triplex U₂A₂U₂ segment, i.e. we consider that the dimer (U₂) segments in poly(rU) can bind to the A-dimer independently of each other. The concentration of bound A-dimer c_D would then fulfill the relation:

$$K = \frac{c_D}{[c_{\text{Afree}}/2][c_{\text{Ufree}}/2]^2}$$

where \(c_{\text{Afree}}\) and \(c_{\text{Ufree}}\) are the concentrations of free adenosine and uridine base units.

Figure 4. Dependencies of triplex concentrations on the proportion of A in the mixtures.
To compare this relation with our sets of experimental data, they have been plotted in a format highlighting the differences in the efficiency of complex formation for the various samples, i.e. as the ratios between the bound and the total adenosine concentrations (\(c_{\text{A bound}}/c_{\text{A}}\)) versus the total adenosine concentration (Fig. 5).

The theoretical curves in the left top graph of Figure 5 demonstrate that equation 9 results in \(c_{\text{A bound}}/c_{\text{A}}\) values that steadily increase with decreasing adenosine concentration. At room temperature only the experimental results obtained for Ap\(^{CH2}\)A(3'-5') (no. 3) and Ap\(^{CH2}\)A(2'-5') (no. 4) satisfy this equation well. However, this agreement obviously arises from almost complete complex formation over the whole concentration range. In that case the concentration of triplex units is principally controlled by available binding sites.

In contrast, all data corresponding to natural ApA (nos 1 and 2) and A\(^{CH2}\)pA modified dimers (nos 5 and 6) show a decrease in complex formation efficiency at low adenosine concentrations. This indicates that a more complex process of formation is underway, probably involving the need for several molecules of...
A-dimer to bind simultaneously to poly(rU) strands so as to achieve sufficient complex stability, i.e. a cooperative effect of triplex formation.

To quantitatively characterise the formation of longer strings of A-dimers bound side-by-side in the triplex chain, one can use a model analogous to that of the nearest neighbour approximation developed for oligonucleotide duplex formation (38). In contrast to this case, an A-dimer bound to two poly(rU) strands in a triplex structure via two pairs of Watson–Crick and two pairs of Hoogsteen hydrogen bonds is considered as the elementary binding unit. The initial step, i.e. the first binding process, and the prolongation step, when another A-dimer binds in the neighbourhood of the initiated triplex chain, have to be distinguished.

If we consider as the initiation step the binding of one A-dimer to two free poly(rU) chains, the concentration $c$ of triplex bound in this way will fulfil the equation:

$$K'_1 = c^{1D}(c_{\text{Ufree}}^2)(c_{\text{Afree}}^2)^{\lambda} = c^{1D}(c_{\text{Ufree}}^2)(c_{\text{Afree}}^2)^{\lambda}$$

where $\lambda$ represents the average length of the poly(rU) chain in nucleoside units. In fact, considering that the average length of poly(rU) in our case was 2000 base units, we would expect that not one but many more single A-dimers can bind independently to a poly(rU) chain. This situation might also be approximated by the right side of equation 10a, although here $\lambda$ means the effective length of independently interacting parts of the poly(rU) chain. The value of $\lambda$ (unknown, but the same for various A-dimers) can be included in the value of the stability constant and equation 10a converted to a form analogous to that of equation 9.

$$K'_1 = K'(\lambda^2/4) = c^{1D}(c_{\text{Ufree}}^2)(c_{\text{Afree}}^2)^{\lambda}$$

Each prolongation step is expected to lead to the same change in the Gibbs potential $\Delta G$ and can therefore be characterised by the same stability constant $K'_2$:

$$K'_2 = c^{2D}(c_{\text{Ufree}}^2)(c_{\text{Afree}}^2)^{\lambda} = c^{2D}(c_{\text{Ufree}}^2)(c_{\text{Afree}}^2)^{\lambda}$$

where $c^{2D}$, $c^{3D}$, $c^{4D}$, ... are the concentrations of triplex segments including 2, 3, 4, ... A-dimers. The total concentration of triplex expressed in terms of U:A*U triplet concentration will then be:

$$c^2 = 2c^{2D} + 4c^{3D} + 6c^{4D} + 8c^{5D} + \ldots$$

$$= \sum_{i=1}^{\infty} (2i-1)(c_{\text{Ufree}}^2)^iK_{i}^{1D}$$

where $L$ represents the maximum effective length of the triplex chain (in numbers of adenosine dimers). This model can be extended to cases where the prolongation stability constant $K'_n$ depends on the length of the clusters (this might be especially important for very short clusters). Relation 11 is in this case generalised to an equation containing several independent stability constants:

$$c^2 = \sum_{i=1}^{\infty} rK_{i}(c_{\text{Ufree}}^2)^i + \sum_{i=1}^{\infty} \frac{1}{i}(c_{\text{Ufree}}^2)^i$$

where $n$ is the number of adenosine dimers in the largest cluster characterised by an individual stability constant.

To test the agreement between equation 12 and the experimental results obtained for various A-dimers, least square fits were carried out. In using relations 3, equation 12 enables the $c_{\text{Afree}}^n/c^A$ ratio value for each $c^A$ to be determined unambiguously if $K_1$ and $K_2$ are known. The fitting procedure is based on a SIMPLEX algorithm, where at each step the $c_{\text{Afree}}^n/c^A$ value is numerically calculated as the solution of a high degree polynomial equation. For each set of experimental data a series of fits has been applied for $n$ and $L$ values varying from 1 to 6 and 1 to 50, respectively.

For ApA, $A^{\text{CHD}}$pA and $A^{\text{CHD}}$A dimers (nos 1, 2 and 5–7) the fitted curves were systematically above the experimental data in the region of high $c^A$ values (region of A-dimer excess). This disagreement was not reduced on increasing the $n$ value, i.e. the number of independent stability constants. This can be interpreted as follows: while equation 12 predicts an almost 100% efficiency of poly(rU) binding in the triplex structure for an adenosine excess, the experimental data obviously indicate that there remains some percentage of uridine units of poly(rU) which cannot participate in triplex structures. Such an effect is understandable since the dispersion of the poly(rU) strand lengths must lead to free (‘dangling’) ends of the longer strand of the two poly(rU) molecules, creating a particular triplex structure.

By employing relations 13 instead of 3, an excellent agreement of the fits with the experimental data has been easily achieved even for $n = 1$. In fact, the fitting results revealed that other stability constants apart from $K_1$ and $K_2$ were redundant, thus indicating that equation 11 was adequate (together with relations 13) in describing the experimental data.

As for the dependence of the fit on $L$ (the maximum effective length of the triplex chain), the sum of the squared deviations showed an almost flat minimum between values of $L = 4$–7 and $L = 7$–20 (depending on the various A-dimers); this value corresponds to approximately one turn of the helix. The limitation of the highest possible $L$ values obtained for some A-dimers (see Table 1) can plausibly be interpreted as follows: prolongation of the triple-helical structure can lead to increased tension in the poly(rU) strands and thus the stability of the triplex chain decreases when a certain length ($L$ value) is exceeded.

To analyse the fitted parameters we compared the values obtained for the minimum of the sum of the squared deviations for all A-dimers (see Table 1). The theoretical $c_{\text{Afree}}^n/c^A$ curves obtained as the best fits are shown in Figure 5.

Table 1 displays the results obtained at room temperature (20°C) for all A-dimers of the set (see Fig. 1) except $A^{\text{CHD}}$pA (2’-endo-5’), which is unable to form complexes with poly(rU). In fact, we found that the parameters $K_1$, $K_2$ and $\zeta$ cannot be unambiguously determined either in the case of very effective binding [see $A^{\text{CHD}}$A (3’-5’)] (no. 3) or, in contrast, for a low tendency to form triplex complexes [see ApA (2’-5’)] (no. 2). In the former case [ApA$^{\text{CHD}}$A (3’-5’)] (no. 3) the complex concentration is mainly controlled by the total concentration of available nucleoside units, given by $c^A$. The fit is thus less sensitive to $K_1$ and practically insensitive to $K_2$: for this...
seems to support the hypothesis that the enthalpy term may be for the studied complexes also do not differ significantly. This forming dimers and the MD data suggest that the water shells of the triplex chains are very similar for all triplex-

ments of the triplex chains are very similar for all triplex-

improves the ability to form longer chains to a level comparable with those of 3'-5' linkages.

**Table 1.** Results of the fit of experimental data shown in Figure 5 by relations 11 and 13

<table>
<thead>
<tr>
<th>Type of dimer</th>
<th>Temperature</th>
<th>$T_{max} - T_{min}$</th>
<th>$\log K_1$</th>
<th>$\log K_2$</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ap(A3'-5')</td>
<td>20°C</td>
<td>6-12</td>
<td>-0.53</td>
<td>2.98</td>
<td>0.86</td>
</tr>
<tr>
<td>Ap(A2'-5')</td>
<td>20°C</td>
<td>4-7</td>
<td>-0.57</td>
<td>2.75</td>
<td>1.00</td>
</tr>
<tr>
<td>Ap(CH2ApA3'-5')</td>
<td>20°C</td>
<td>-</td>
<td>&gt;3.75</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ap(CH2ApA2'-5')</td>
<td>20°C</td>
<td>7-15</td>
<td>1.67</td>
<td>3.48</td>
<td>0.92</td>
</tr>
<tr>
<td>Ap(CH2ApA3'-5')</td>
<td>20°C</td>
<td>5-10</td>
<td>-0.62</td>
<td>3.06</td>
<td>0.86</td>
</tr>
<tr>
<td>Ap(CH2ApA2'-5')</td>
<td>20°C</td>
<td>6-10</td>
<td>-0.18</td>
<td>3.15</td>
<td>0.88</td>
</tr>
<tr>
<td>Ap(CH2Ap(2'3'exo-5'))</td>
<td>20°C</td>
<td>7-20</td>
<td>-0.74</td>
<td>2.99</td>
<td>0.87</td>
</tr>
<tr>
<td>Ap(A3'-5')</td>
<td>4°C</td>
<td>7-15</td>
<td>2.23</td>
<td>3.71</td>
<td>0.90</td>
</tr>
<tr>
<td>Ap(A2'-5')</td>
<td>4°C</td>
<td>5-8</td>
<td>1.62</td>
<td>3.32</td>
<td>1.00</td>
</tr>
<tr>
<td>Ap(CH2Ap(2'3'exo-5'))</td>
<td>4°C</td>
<td>&gt;4.05</td>
<td>-</td>
<td>-</td>
<td>0.94</td>
</tr>
</tbody>
</table>

*The value was kept fixed during the fit.

<table>
<thead>
<tr>
<th>Type of dimer</th>
<th>Temperature</th>
<th>$T_{max} - T_{min}$</th>
<th>$\log K_1$</th>
<th>$\log K_2$</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ApCH2ApA(3'-5')</td>
<td>11°C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ApCH2ApA(2'-5')</td>
<td>13°C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Because the two terms are of opposite sign and probably of comparable value, a relatively small change in one of them is able to cause a relatively large change in the stability constant. The entropy term involves fixation of two poly(rU) strands and their ordering near the binding site, the geometrical arrangement of the A-dimer and changes in the hydration shell around the complex. As follows from the Raman spectra, the arrangements of the triplex chains are very similar for all triplex-forming dimers and the MD data suggest that the water shells for the studied complexes also do not differ significantly. This seems to support the hypothesis that the enthalpy term may be responsible for the observed differences in $\log K_1$. This expectation is in accord with the low temperature data obtained for the A-dimers exhibiting typical ‘cooperative’ curves at room temperature. When the temperature was lowered to 4°C, the $K_1$ value was so increased that the cooperative effect almost disappeared (Fig. 5 and Table 1). More definite conclusions are, however, expected only from a systematic study of the 3D dependencies of Raman spectra on both the concentration ratio and the temperature, planned for the near future. The data obtained for Ap(A2'-5') at room temperature indicate that its low binding ability does not result from a low efficiency of the initiation step but from a low ability to form long triplex chains. The flexibility increase caused by incorporation of a methylene group into the 2'-5' phosphate linkage [i.e. in Ap(CH2ApA2'-5') (no. 4) and Ap(CH2ApA2'-5') (no. 6)] obviously improves the ability to form longer chains to a level comparable with those of 3'-5' linkages.

**Figure 6.** Example of a simulated system, rU12*(ApA 3'-5')2*rU12.

Although the Raman features of the triplex exhibited a great similarity for the various modified A-dimers [except for the position of the adenine breathing mode, sensitive to the 3' or 2' binding site of the phosph(on)ate linkage], thus reflecting common structural properties, the stability of the complexes and the necessity to simultaneously bind several adenosine dimers differed significantly from one compound to another. To explore the basic reasons for these differences, a MD simulation of the natural A-dimer (ApA 3'-5') and two of its modified congeners [Ap(CH2ApA3'-5') (no. 3) and Ap(CH2ApA3'-5') (no. 5)] was carried out.

Four model structures representing dodecamer triplex chains have been studied. Each model system consisted of two mutually antiparallel uridine dodecamers and a pseudouracil consisting of six A-dimers bound by Watson–Crick and Hoogsteen hydrogen bridges to the first and second rU12 strands, respectively (Fig. 6). The three dimer–oligomer systems, namely rU12*[Ap(A3'-5')6*rU12, rU12*[Ap(CH2ApA3'-5')6*rU12 and rU12*[Ap(CH2ApA3'-5')6*rU12, were completed by a system with a natural adenosine dodecamer strand, rU12*rA12.*rU12.

The first MD runs, however, led to a break-up of all of the model triplex structures, despite the fact that those formed with their DNA equivalents, e.g. dT12*[dApdA3'-5')6*dT12, were stable. A careful investigation of the critical structural changes before the break revealed that the problem resides in the force constants of the glycosidic torsion angle of uridine in the newest version of the AMBER force field. An energetic profile of this torsion had been optimised for some intervals of torsion.
angles but remains underestimated for other intervals of torsion angles. When the force constants from the previous version of AMBER were used, the instability of the RNA triplexes disappeared.

For all four systems, 100 ps MD runs yielded stable conformations. A-dimers with natural phosphodiester linkages adopted a $\beta$-g-g geometry. For the rU$_{12}$·tA$_{12}$·rU$_{12}$ system some temporary transitions to the $\beta$-gt conformation were found for the adenylic strand. For both modified linkages the $\beta$-g-g-g conformation has been found to be most stable. The preferred ribose puckering was C3′-endo, but for the modified A-dimers the riboses bound at the C5′ site showed a C4′-exo puckering.

Average helical parameters of the duplex fragments involved in U and A strands (or pseudostrands) bound by Watson–Crick hydrogen bonds were close to those characteristic for the A-form of RNA. The main overall deviation was an increased inclination (39) for all systems. It was caused by a high inclination of the adenine moieties at the ends of the complexes.

The main significant differences between the investigated complexes were found in the stability of the Hoogsteen hydrogen bridges. Except for the rU$_{12}$·[ApCH$_2$A(3′-5′)]$_6$·rU$_{12}$ complex, several Hoogsteen hydrogen bonds for the inner adenosines in average MD structures exceeded the limit of 2.1 Å (Fig. 7).

For natural complexes the central Hoogsteen bonds (Ade N6…O4 Ura; Fig. 8A, bonds 2 and 3) are disturbed. This can be explained by an analysis of the behaviour of the sugar–phosphate backbones of the A and the Hoogsteen uridine strand during the MD runs. The phosphate groups of the A strand can get close to the OH group at the C2′ site of the uridine ribose. In this situation an additional hydrogen bond can be established between both groups (Fig. 8A, bond 1), which correlates with the lengthening of the central Hoogsteen bond.

For rU$_{12}$·[A$_{3}$CH$_2$pA(3′-5′)]$_6$·rU$_{12}$, the perturbation concerns the second lateral Hoogsteen hydrogen bond (Ade N7…O4 Ura; Fig. 8B, bond 2). This is due to a steric conflict between the oxygen atom of the adenosine phosphate group and the O$_2$ of uridine participating in the hydrogen bond (Fig. 8B, interaction 1).

For rU$_{12}$·[ApCH$_2$A(3′-5′)]$_6$·rU$_{12}$ the position of the phosphate group in the adenosine strand is moved away from uridine in the Hoogsteen strand and, instead, the methylene group, which does not induce any weakening of the Hoogsteen bonds, is close to the uridine (Fig. 8C).

Figure 7. Lengths of hydrogen bonds (in Å) in average MD structures for individual U-A-U triads (numbering starts at the 5′-end of the Watson–Crick (rU)$_{12}$). Filled circles correspond to Ade N1…N3 Ura Watson–Crick pairs, empty circles to Ade N1…O2 Ura Watson–Crick pairs, filled triangles to Ade N1…N3 Ura Hoogsteen pairs and empty triangles to Ade N1…O2 Ura Hoogsteen pairs.

Figure 8. Interaction of A-dimers (red) with uracil moieties of poly(rU) (green). The normal hydrogen bonds (black) as well as the disturbed hydrogen bonds (purple) are indicated in the figures, atoms involved in the disturbed hydrogen bonds or steric conflicts are coloured yellow. (A) ApA(3′-5′). The additional hydrogen bond (1) between the P-O group of the dimer and the ribose ring of poly(rU) can disturb the central Hoogsteen hydrogen bonds (2,3). (B) A$^{3\text{CH}_2}$pA(3′-5′). The oxygen atom of the adenosine phosphate group and the O$_2$ of uridine are involved in a steric conflict (1) that can disturb the lateral Hoogsteen hydrogen bond (2). (C) A$^{3\text{CH}_2}$pA(3′-5′). All the hydrogen bonds are well formed.
Proper 3D arrangement of the $\text{Ap}^{\text{CH2A}}(3'-5')$ linkage (with no steric conflict nor additional hydrogen bonds) may explain the outstanding properties of this A-dimer with respect to its propensity for building stable triplex structures with poly(rU). It is worth mentioning that it is also the only phosphonate internucleotide linkage that can lead to an oligonucleotide with alternation of modified and natural linkages, which, besides conferring a high resistance to nucleases, is able to activate RNase H (19), a necessary precondition for antisense activity.

CONCLUSIONS

The newly developed procedure for processing 2D Raman data (representing Raman spectra dependencies on the concentration ratio of interacting species) provides direct information about the formation, stoichiometry and concentration of complexes formed in mixed samples of modified A-dimers and poly(rU). At the same time, this method allows pure Raman spectra of the complexes to be obtained, which provide information about their structure. The next step will be extension of this procedure to process 3D data, including temperature as another independent variable.

The ability of a family of modified A-dimers to bind their natural nucleic acid counterpart poly(rU) has been monitored as a function of internucleotide linkage modification, consisting of lengthening of the linkage by inserting a methylene group in various places. Except for the $\text{A}^{\text{CH2A}}(2'3'\text{endo}-5')$ modification, all of the isopolar, non-isotERIC analogues formed triplex-like complexes with poly(rU) at room temperature, where two polymer strands were bound by Watson–Crick and Hoogsteen bonds to a central pseudostrand consisting of a ‘chain’ of A-dimers. Raman spectra showed that the overall conformations of the triplexes obtained for the various modified adenosine dimers are very similar and the first poly(rU) strand and the pseudo-strand of dimers joined by Watson–Crick bonds adopt a conformation typical of the A-form RNA duplex. The only systematic differences found between the 2'-5' and 3'-5' families of dimers involved an adenosine breathing vibration and reflect different torsion angles of the glycosidic linkage.

A simple semi-empirical model enabled the observed dependence of triplex formation efficiency on the adenosine concentration to be entirely described. For most of the analogues tested (including the natural 3'-5' phosphate linkage) it has been shown that creation of a stable complex requires formation of a central pseudostrand chain consisting of several adenosine dimers. These pseudostrands may have different average lengths, depending on the type of dimer and the A:U ratio, and in the resulting clusters the $U_2:A_2^*U_2$ units have similar spectral properties. The ‘cooperative’ aspect of complex formation involves a decrease in A-dimer binding efficiency when adenosine concentration decreases, although the number of effective binding sites increases on the poly(rU) strand. For natural 2'-5' ApA complex formation is poor, even at high A-dimer concentrations, probably for steric reasons.

For $\text{Ap}^{\text{CH2A}}$ dimers (nos 3 and 4), i.e. when the methylene group is inserted close to the 5'-position, the initial binding conditions are much more favourable compared to other modified or natural linkages. MD simulations indicate that these differences are consistent with very good compatibility of the 3D conformation of the phosphonate, including the methylene group, toward the second (Hoogsteen) poly(rU) strand. There is no additional hydrogen bond between the A-dimer linkage and the poly(rU) ribose [as for ApA(3'-5')] and no steric conflict between the modified linkage and uracil [as for $\text{A}^{\text{CH2A}}(3'-5')$]. Although these properties do not directly give an advantage to $\text{Ap}^{\text{CH2A}}$ linkages for duplex formation with natural nucleic acids, they can serve as an indication of suitable structures enabling good interactions with other molecules over the major groove.

ACKNOWLEDGEMENTS

This work was supported by the Grant Agency of the Czech Republic (project nos 203/01/1166 and 202/01/P102), the Grant Agency of the Czech Academy of Sciences (project no. A4055902), the Czech Ministry of Education, Youth and Sports (project no. VS97113) and the Czech–French Collaboration Project (Barrande no. 98065). J.H. gratefully acknowledges the French government’s support of his stay at LPBC.

REFERENCES


