

# Conformational properties of DNA dodecamers containing four tandem repeats of the CNG triplets

Michaela Vorlícková\*, Miriam Zimulová, Jirí Kovanda, Petr Fojtík and Jaroslav Kypr

Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, CZ-612 65 Brno, Czech Republic

Received February 9, 1998; Revised and Accepted April 16, 1998

## ABSTRACT

We studied DNA dodecamers (CAG)<sub>4</sub>, (CCG)<sub>4</sub>, (CGG)<sub>4</sub> and (CTG)<sub>4</sub> by CD spectroscopy and polyacrylamide gel electrophoresis. Each dodecamer adopted several ordered conformers which denatured in a cooperative way. Stability of the conformers depended on the dodecamer concentration, ionic strength, temperature and pH. The dodecamers, having a pyrimidine base in the triplet center, generated foldbacks at low ionic strength whose stem conformations were governed by the GC pairs. At high salt, (CCG)<sub>4</sub> isomerized into a peculiar association of two strands. The association was also promoted by high oligonucleotide concentrations. No similar behavior was exhibited by (CTG)<sub>4</sub>. At low salt, (CGG)<sub>4</sub> coexisted in two bimolecular conformers whose populations were strongly dependent on the ionic strength. In addition, (CGG)<sub>4</sub> associated into a tetraplex at acidic pH. A tetraplex was even observed at neutral pH if the (CGG)<sub>4</sub> concentration was sufficiently high. (CAG)<sub>4</sub> was very stable in a monomolecular conformer similar to the known extremely stable foldback of the (GCGAAGC) heptamer. Nevertheless, even this very stable conformer disappeared if (CTG)<sub>4</sub> was added to the solution of (CAG)<sub>4</sub>. Association of the complementary strands was also strongly preferred to the particular strand conformations by the other couple, (CCG)<sub>4</sub> and (CGG)<sub>4</sub>.

## INTRODUCTION

Nuclear DNA molecules of higher organisms contain huge amounts of simple sequence repeats that expand in the human genome (1–3). Some of the expansions correlate with a number of serious, mainly neurological diseases (reviewed in 4). These pathological expansions have so far been mainly connected with the (CNG)<sub>n</sub> repeats, although many other repeats expand as well (3) and their expansions also bring about diseases (5). Many studies suggest that the simple sequence repeats can adopt unusual DNA conformations, which are likely to play a significant role in the expansions (reviewed in 6–8).

The (CNG)<sub>n</sub> repeats were shown to form intramolecular foldbacks (reviewed in 7) and/or bimolecular duplexes. Homoduplexes

and heteroduplexes of the (CNG)<sub>n</sub> repeats have been mostly characterized as slipped (9), compact (10) or flexible (11). The central triplet bases were extrahelical in a duplex formed by (CCG)<sub>2</sub> (12), or they introduced a conformational peculiarity in another way that is not yet understood. Furthermore, (CGG)<sub>n</sub> generated guanine tetraplexes promoted by cytosine methylation (13) or potassium cations (14). Still larger (CGG)<sub>n</sub> aggregates were promoted by acid pH (15).

DNA molecules often adopt more than a single conformation depending on a number of factors, including DNA concentration, pH, ionic strength and temperature. Here, we compare the conformational equilibria of (CAG)<sub>4</sub>, (CCG)<sub>4</sub>, (CGG)<sub>4</sub>, (CTG)<sub>4</sub>, and of their duplexes, to demonstrate the following points. (i) Although hairpins of the (CNG)<sub>4</sub> repeats are more stable than hairpins of other sequences of the same length, they are quantitatively destabilized in the presence of the complementary strand with which they unambiguously prefer to form a heteroduplex; (ii) this is even true of the very stable single-stranded conformer of (CAG)<sub>4</sub> which, remarkably, shares some properties with the well-known extremely stable hairpin of (GCGAAGC) (16); (iii) we found conditions to control tetraplex formation by (CGG)<sub>4</sub> without the oligonucleotide aggregation. At the oligonucleotide concentrations lower than required for the tetraplex stability, this dodecamer coexisted in two bimolecular structures whose populations were controlled by the ionic strength; and (iv) in contrast with (CAG)<sub>4</sub> and (CTG)<sub>4</sub>, the conformation of (CCG)<sub>4</sub> was oligonucleotide concentration dependent, like the conformation of (CGG)<sub>4</sub>. At low concentrations, (CCG)<sub>4</sub> folded back to generate a GC base-paired stem that was not significantly influenced by the extra cytosines. A very similar foldback was formed by (CTG)<sub>4</sub>. At high oligonucleotide concentrations, however, (CCG)<sub>4</sub> associated into a peculiar bimolecular structure providing a strong non-conservative CD spectrum that originated from a cytosine–cytosine interaction.

## MATERIAL AND METHODS

The (CNG)<sub>4</sub> dodecamers were synthesized and purified by Integrated DNA Technologies, Inc., and bought from East Port, Prague. They were dissolved in 0.5 mM NaCl, giving about 10 mM DNA nucleoside residue concentrations. Aliquots were diluted by 1 mM sodium phosphate and 0.3 mM EDTA, pH 7, for the particular measurements, and the dodecamer concentrations were determined

\*To whom correspondence should be addressed. Tel: +42 5 41517188; Fax +42 5 41240497; Email: mifi@ibp.cz

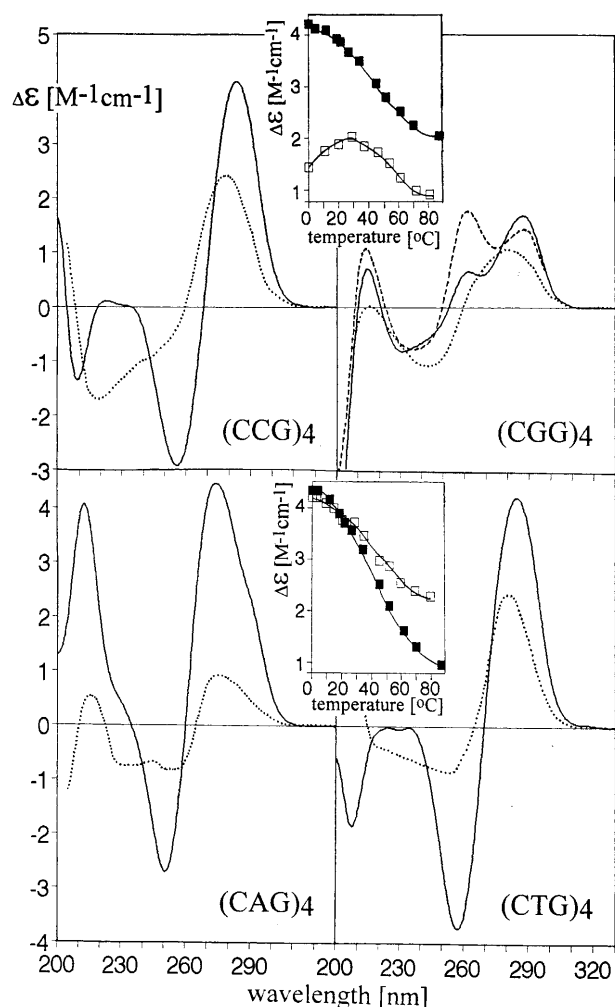
from the UV absorbance measured at room temperature on the UNICAM 5625 spectrometer, using the following molar extinction coefficients:  $\epsilon_{260} = 7650$  M $\cdot$ cm for (CCG) $_4$ ,  $\epsilon_{260} = 8300$  M $\cdot$ cm for (CGG) $_4$  (the sample had to be denatured for 15 min at 90°C prior to the UV absorption measurement),  $\epsilon_{255} = 9050$  M $\cdot$ cm for (CAG) $_4$  and  $\epsilon_{260} = 8300$  M $\cdot$ cm for (CTG) $_4$ . The molar extinction coefficients were determined using phosphodiesterase hydrolysed samples and the known monomer values. Further oligonucleotides, (C $_6$ G $_6$ ), (GA) $_5$ , (GA) $_5$ (GC) $_2$ (GA) $_5$  and (GCGAAGC), were synthesized and purified by Dr L. Arnold, Prague. Their molar extinction coefficients were assessed from their single-strand extinction coefficients calculated according to Gray *et al.* (17) and their hypochromic effects. The resulting values were  $\epsilon_{260} = 8090$  M $\cdot$ cm for C $_6$ G $_6$ ,  $\epsilon_{255} = 11\,450$  M $\cdot$ cm for (GA) $_5$ ,  $\epsilon_{255} = 10\,780$  M $\cdot$ cm for (GA) $_5$ (GC) $_2$ (GA) $_5$  and  $\epsilon_{255} = 10\,300$  M $\cdot$ cm for (GCGAAGC). After the oligonucleotide concentration was determined in 1 mM sodium phosphate + 0.3 mM EDTA, pH 7, the sample solutions were adjusted to the conditions specified in figure legends. Dependencies on the salt concentration were measured in 10 mM sodium phosphate while the Robinson–Britton buffer was used at acid pH and in the pH dependencies. In the salt dependencies, drops of concentrated NaCl or KCl were added up to 100 mM concentrations and then weighed amounts of solid salts were used instead. In both cases, the salt and DNA concentrations were corrected for the sample volume increase.

CD spectra were measured using the Jobin-Yvon Mark IV and Mark VI dichrographs in 0.01–5 cm pathlength Hellma cells placed in thermostated holders. The latter dichrograph was also used to measure the UV absorption spectra. Concentrations of DNA moved within 0.02–10 mM (in nucleoside residues) and the cells were always chosen to get the absorption values around 0.8, which gave optimum signal-to-noise CD ratios. Dependencies on pH were measured in 1 cm pathlength cells where the samples were titrated by the acid component of the buffer. The pH values were measured directly in the cells by a Radelkis pH-meter and a Hofer electrode.

Polyacrylamide gel electrophoresis was performed in a thermostated apparatus of submarine type (SE 600, Hofer Scientific, San Francisco). Polyacrylamide gels (20%, 29:1 monomer/bis ratio) had dimensions of 14  $\times$  16  $\times$  0.1 cm. The electrophoresis was run at 0°C for 20 h at 70 V (~5 V/cm) in buffers specified in figure legends. The gels were stained with ethidium and/or silver. Densitometry was performed using the Personal Densitometer SI, 375 A (Molecular Dynamics, Sunnyvale, CA).

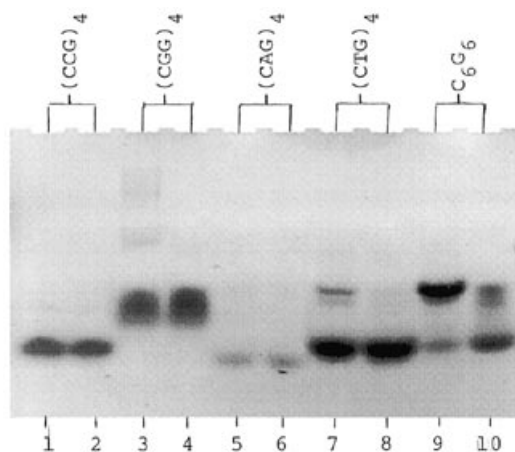
## RESULTS

The large amplitudes of bands in the CD spectra (Fig. 1) and their extensive reduction with increasing temperature indicated that the present (CNG) $_4$  strands of DNA assumed ordered conformers even at very low ionic strength, e.g. in 1 mM sodium phosphate, 0.3 mM EDTA, pH 7. The dodecamers exhibited signs of cooperative thermal melting (Fig. 1, inserts) like oligonucleotides of similar length and base composition, containing self-complementary nucleotide sequences. The CD spectra of the dodecamers having pyrimidine bases in the repeated motif center, i.e. (CCG) $_4$  and (CTG) $_4$  (Fig. 1), were similar. They had the positive maximum at 285 nm and the negative maximum at 255 nm, like the B-forms of (G+C)-rich DNAs, but their ellipticity amplitudes were much larger.



**Figure 1.** Low and high temperature CD spectra of (CNG) $_4$  dodecamers of DNA measured in 1 mM sodium phosphate and 0.3 mM EDTA, pH 7. The DNA concentration was about  $7.5 \times 10^{-4}$  M. The solid traces and the dashed trace of (CGG) $_4$  were measured at 0°C. Dotted traces: 86, 81, 86 and 80°C with (CCG) $_4$ , (CGG) $_4$ , (CAG) $_4$  and (CTG) $_4$ , respectively. The dashed trace of (CGG) $_4$  was measured immediately after diluting the sample from the stock solution. The solid traces were measured after the sample exposure to 90°C for 15 min, and cooling. Inserts: temperature-induced ellipticity [at 285 nm, except (CAG) $_4$  at 275 nm]. Temperature-induced ellipticity changes of (CCG) $_4$  (top, filled squares), (CGG) $_4$  (top, open squares), (CAG) $_4$  (bottom, filled squares) and (CTG) $_4$  (bottom, open squares).

The purine-rich (CNG) $_4$  dodecamers were different. The CD spectrum of (CAG) $_4$  (Fig. 1), including the strong positive maximum at 272 nm and the positive band at 210 nm, was more similar to the CD spectrum of the (GA) $_n$  homoduplex (see below) than to the CD spectrum of the (G+C)-rich duplexes of DNA. Another peculiar CD spectrum was displayed by (CGG) $_4$  (Fig. 1). Immediately after diluting the aliquot taken from the concentrated stock solution, the sample displayed the dominant positive band at 260 nm, the amplitude of which increased with the oligonucleotide concentration (see below) and decreased with time after the dilution. In contrast with the remaining three present dodecamers, the CD spectrum of (CGG) $_4$  irreversibly changed upon thermal denaturation and renaturation. Its positive



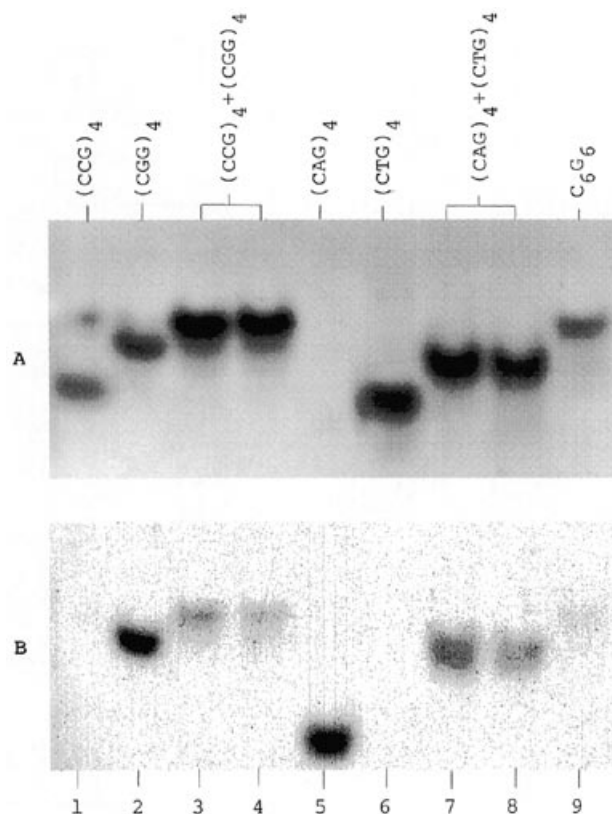
**Figure 2.** A polyacrylamide gel run in 1 mM sodium phosphate and 0.3 mM EDTA pH 7, at 0°C. The gel was stained with ethidium. From left to right: lanes 1 and 2, (CCG)<sub>4</sub>; lanes 3 and 4, (CGG)<sub>4</sub>; lanes 5 and 6, (CAG)<sub>4</sub>; lanes 7 and 8, (CTG)<sub>4</sub>; lanes 9 and 10, (C<sub>6</sub>G<sub>6</sub>) marker. The samples run in the even lanes were denatured for 15 min at 90°C before loading.

band at 260 nm disappeared and was substituted by a maximum at 290 nm (Fig. 1).

Figure 2 shows a polyacrylamide gel of the present four dodecamers electrophoresed in 1 mM sodium phosphate + 0.3 mM EDTA at 0°C. The marker bands were provided by another DNA dodecamer (C<sub>6</sub>G<sub>6</sub>), which coexisted as a bimolecular duplex and a foldback under the conditions of electrophoresis. Thermal denaturation/renaturation switched a portion of the bimolecular duplexes into the foldbacks. All of the five dodecamers were run in two samples, one without and the other after the thermal denaturation/renaturation treatment. All four samples of the pyrimidine-rich (CNG)<sub>4</sub> dodecamers migrated like the foldback of (C<sub>6</sub>G<sub>6</sub>), except for a small portion of (CTG)<sub>4</sub> that was a bimolecular duplex before the denaturation/renaturation treatment. Remarkably, the foldback of (CAG)<sub>4</sub> migrated faster than the foldbacks of the other present (CNG)<sub>4</sub> dodecamers. It stained very poorly with ethidium but well with silver (Figs 2 and 3), which is characteristic of the DNA homoduplex of (GA)<sub>n</sub> (J.Kovanda and J.Kypr, unpublished data).

In contrast with the above three dodecamers, the strands of (CGG)<sub>4</sub> failed to form species migrating in the polyacrylamide gels as an ordered monomolecular conformer. The oligonucleotide migrated in a broad smear of bands, corresponding to two associated oligonucleotide strands, up to species migrating between the duplex and foldback. In addition, there were traces of four and eight associated (CGG)<sub>4</sub> strands that disappeared upon denaturation.

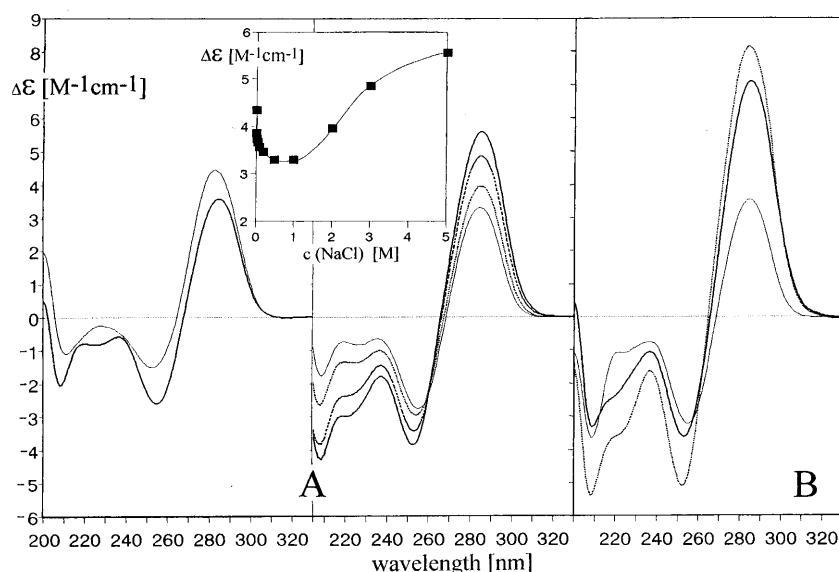
All of the events the described above concerned a very low ionic strength, i.e. a millimolar sodium cation concentration. The ionic strength increase to 10 mM sodium phosphate + 100 mM NaCl at pH 7, completely switched (C<sub>6</sub>G<sub>6</sub>) to a bimolecular duplex (Fig. 3), whereas (CAG)<sub>4</sub>, (CTG)<sub>4</sub> and a major part of (CCG)<sub>4</sub> remained at the position of foldbacks. (CAG)<sub>4</sub> displayed the fastest migration again (Fig. 3B). The sample of (CGG)<sub>4</sub> provided only a single sharp band under these conditions, corresponding to a bimolecular structure. In accordance with the electrophoretic results, the addition of 100 mM NaCl induced



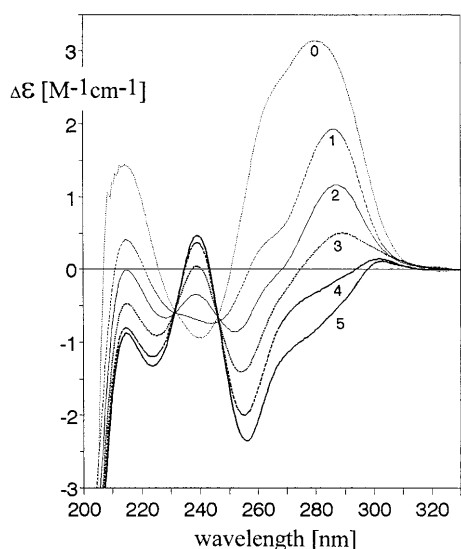
**Figure 3.** A polyacrylamide gel run in 10 mM sodium phosphate, 0.1 mM EDTA, pH 7, 0.1 M NaCl at 0°C, and stained with (A) ethidium and (B) silver. The oligonucleotides were denatured for 15 min at 90°C before loading. Lane 1, (CCG)<sub>4</sub>; lane 2, (CGG)<sub>4</sub>; lanes 3 and 4, (CCG)<sub>4</sub>·(CGG)<sub>4</sub>; lane 5, (CAG)<sub>4</sub>; lane 6, (CTG)<sub>4</sub>; lanes 7 and 8 (CAG)<sub>4</sub>·(CTG)<sub>4</sub>; lane 9, (C<sub>6</sub>G<sub>6</sub>) marker. Equimolar quantities of the complementary oligonucleotides were mixed and immediately loaded onto the gel (lanes 3 and 7) or heated for 15 min at 90°C, slowly cooled to room temperature and loaded (lanes 4 and 8).

almost no changes in the CD spectra of (CAG)<sub>4</sub> and (CTG)<sub>4</sub> (not shown). On the other hand, the CD spectra of (CCG)<sub>4</sub> (Fig. 4) and especially (CGG)<sub>4</sub> (Fig. 5) were sensitive even to small ionic strength changes.

The change from 0 to 100 mM NaCl suppressed the whole CD spectrum of (CCG)<sub>4</sub> (Fig. 4A, left). The spectrum, and mainly its positive part, was further suppressed at still higher salt concentrations. The suppression was non-cooperative and presumably reflected conformational modifications within the oligonucleotide foldback. However, molar NaCl concentrations caused a different effect, i.e. a positive band amplification combined with the negative CD band deepening (Fig. 4A, right). These changes resulted in the CD spectra intersection in the isoelliptic point at 260 nm, indicating that the oligonucleotide switched into another conformer. This conformer was also promoted at high oligonucleotide concentrations (Fig. 4B). If the ionic strength was increased in this sample, the CD amplitudes became still larger. The conformer, however, disappeared upon the sample dilution (Fig. 4B). It also disappeared upon salt dilution if it was stabilized by the high NaCl concentrations. These properties did not allow us to detect the conformer properly under the conditions of electrophoresis. Yet it was clear that the conformer of (CCG)<sub>4</sub>, providing the huge positive CD band at 284 nm and the negative bands at 252.5 and



**Figure 4.** Changes in the CD spectra of  $(CGG)_4$  induced by increasing the NaCl (A) or  $(CGG)_4$  (B) concentration. (A) DNA concentration (0.8–0.6 mM) in 0.1 cm cell, temperature 25°C. Left panel, (thin trace) 1 mM sodium phosphate + 0.3 mM EDTA, pH 7; (bold trace) 10 mM sodium phosphate, 0.3 mM EDTA, pH 7, and 0.1 M NaCl. Right panel, 10 mM sodium phosphate, 0.3 mM EDTA, pH 7, and (from the thinnest to the boldest trace) 0.5, 2.0, 3.0 and 5 M NaCl. Insert, salt-induced changes in the CD spectra of  $(CGG)_4$  monitored at 285 nm. (B) Bold trace, 9.5 mM  $(CGG)_4$  (0.01 cm cell) in 1 mM sodium phosphate and 0.3 mM EDTA, pH 7, at 0°C. Dotted trace, 6.2 mM  $(CGG)_4$  in 1 mM sodium phosphate, 0.3 mM EDTA, 0.5 M KCl, pH 6.5, at 23.5°C. Thin trace, the latter sample diluted by the same buffer to a 0.5 mM DNA concentration and measured at 23.5°C (in 0.1 cm cell).



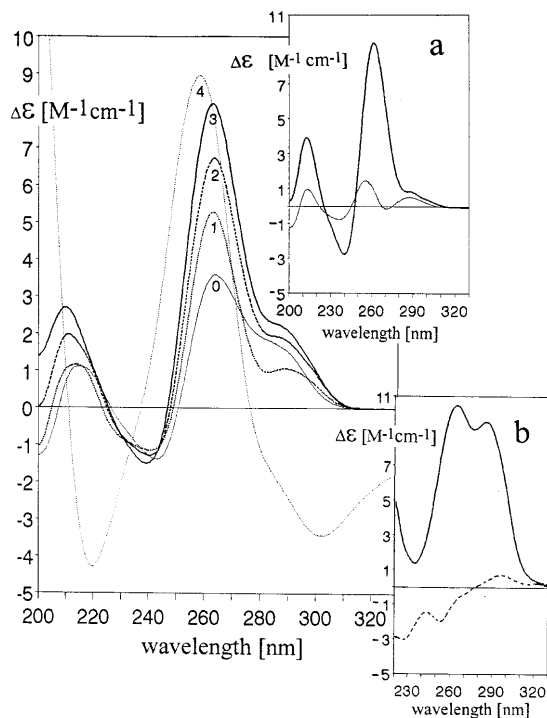
**Figure 5.** Oligonucleotide concentration-dependent and salt-dependent CD spectra of denatured/renatured  $(CGG)_4$  measured at 25°C. Trace 0, 0.02 mM  $(CGG)_4$  in 0.02 mM sodium phosphate, pH 7. Traces 1–3, 0.7 mM  $(CGG)_4$  in 1, 5 and 10 mM sodium phosphate, pH 7, respectively. Traces 4 and 5, 0.7 mM  $(CGG)_4$  in 10 mM sodium phosphate plus 0.1 and 3 M NaCl, respectively.

208.5 nm, included a strand association. Traces in the electrophoretic patterns (Fig. 3) suggested that the association was a bimolecular complex.  $(CTG)_4$  did not form this bimolecular complex at either high salt or at its own high concentration. The reason is probably that clustered cytosines are required for the complex stability because it was easily adopted, e.g. by  $(CCCG)_5$

(M. Vorlíčková *et al.*, unpublished data). This bimolecular complex is being analyzed further in our laboratory.

In line with the above electrophoretic results (Figs 2 and 3), the CD spectrum of  $(CGG)_4$  was also very sensitive to ionic strength (Fig. 5). The spectrum almost reverted owing to the ionic strength increase causing the smear (Fig. 2) coalescence into the single sharp band (Fig. 3). The CD spectrum could also be manipulated by diluting the oligonucleotide sample in 1 mM sodium phosphate. Then, the long wavelength part of the CD spectrum increased substantially (Fig. 5). All of the CD spectra intersected in the isoelectric points at 231 and 246 nm, indicating that  $(CGG)_4$  coexisted in two structures whose equilibrium depended on the ionic strength. A bimolecular duplex was stable at higher ionic strengths while another conformer was preferred at very low salt and oligonucleotide concentrations. This conformer was also promoted by slightly increased temperatures. Denaturation caused opposite changes in the CD spectra. All three other present  $(CNG)_4$  dodecamers were stable foldbacks under these conditions and we see no reason why the foldback of  $(CGG)_4$  should be less stable (18). Hence the apparent absence of foldbacks of  $(CGG)_4$  in the gels was likely to originate from an association of two foldbacks. This two-foldback association was probably the conformer stable at the very low ionic strength.

The third conformer of  $(CGG)_4$ , namely the guanine tetraplex, was stabilized by high oligonucleotide concentrations present in the stock solution. Its diagnostic positive CD band (19) disappeared upon thermal denaturation but it was restored and amplified by 0.5 M KCl added to the sample of 10 mM, but not 1 mM,  $(CGG)_4$ . The addition of KCl triggered a process with slow kinetics (Fig. 6). Within the first 2 h, the CD spectra intersected in the isoelectric points at 231 and 244 nm. These changes reflected the guanine tetraplex formation, while large aggregates formed later as suggested by non-zero ellipticity above 320 nm

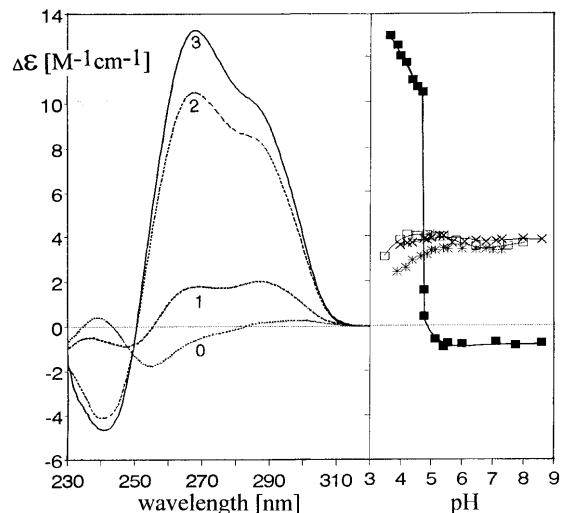


**Figure 6.** KCl-induced time-dependent changes in the CD spectra of  $(CGG)_4$ . Temperature 23.5°C. Trace 0, 9.4 mM  $(CGG)_4$  in 1 mM sodium phosphate and 0.3 mM EDTA, pH 7, no denaturation/renaturation.  $(CGG)_4$  (7 mM) measured immediately (trace 1) and 1 (trace 2), 2 (trace 3) and 39 h (trace 4) after the concentration of KCl was raised from 0 to 0.5 M at pH 6.5. Insert a, CD spectra of  $(TGG)_4$  at 25°C. Thin trace, 0.7 mM oligonucleotide in the Robinson–Britton buffer, pH 7; bold trace, the same sample measured immediately after raising the concentration of KCl from 0 to 0.1 M. The spectrum did not change with time. Insert b, CD spectra of 1 mM  $(CGG)_4$  in the Robinson–Britton buffer, pH 5.3, measured immediately (dashed trace) or 35 days after the addition of 0.5 M KCl (solid trace).

(Fig. 6). Just before the aggregation, the CD spectrum of  $(CGG)_4$  was very similar to the CD spectrum of the tetraplex of  $(TGG)_4$  (20; Fig. 6, insert a) induced by 0.1 M KCl at neutral pH. However, the concentration of  $(TGG)_4$  was an order of magnitude lower.

Below millimolar oligonucleotide concentrations, the tetraplex of  $(CGG)_4$  could only be induced at acid pH (15) where its CD spectrum was, however, rather different. Conditions stabilizing the acid tetraplex led to the formation of large aggregates (15). We reproduced this observation (Fig. 6, insert b) and reduced the reported scattering due to having a lower KCl concentration. No KCl, and even smaller oligonucleotide concentrations, were sufficient to stabilize the acid tetraplex of  $(CGG)_4$  around pH 4.5 (Fig. 7). This acid tetraplex of  $(CGG)_4$  arose with a slow kinetics (hours) but much faster than at pH 5.4 with KCl (days, ref. 15). Our acid tetraplex of  $(CGG)_4$  (Fig. 7) gave no CD signals above 320 nm. Under comparable conditions, no similar changes were observed in the CD spectra of the remaining three present  $(CNG)_4$  dodecamers (Fig. 7).

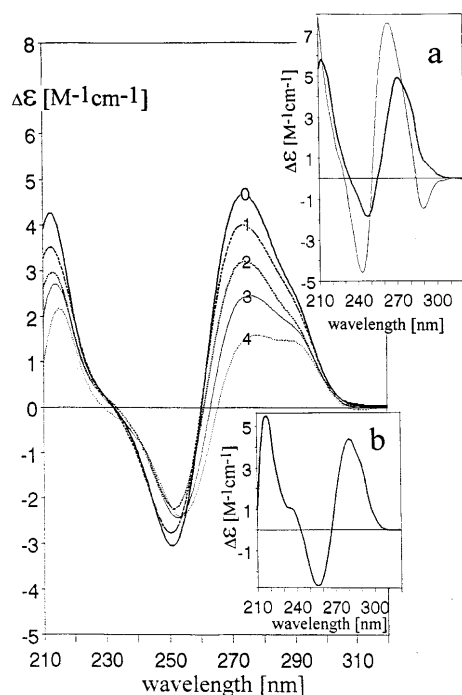
Unlike  $(CCG)_4$  and  $(CGG)_4$ ,  $(CAG)_4$  exhibited no oligonucleotide concentration dependent properties. It was a very stable foldback under most of the examined conditions. The similarity of the CD spectrum of  $(CAG)_4$  (Fig. 8) to the CD spectrum of  $(GA)_n$  (21; Fig. 8) suggested that  $(CAG)_4$  was



**Figure 7.** pH-induced changes in the CD spectra of  $(CGG)_4$ . Temperature 25°C. Left panel, 0.07 mM  $(CGG)_4$  in the Robinson–Britton buffer at pH 8.6 (trace 0), spectra measured 2 and 19 h after adjusting the pH to 4.7 (traces 1 and 2, respectively) and pH 3.7 (trace 3). Right panel, pH-induced changes in the CD spectra of  $(CGG)_4$  (filled squares),  $(CCG)_4$  (open squares),  $(CAG)_4$  (asterisks) and  $(CTG)_4$  (crosses) monitored at 268, 285, 275 and 285 nm, respectively.

different from the foldbacks of  $(CCG)_4$  or  $(CTG)_4$ . Fast migration in the polyacrylamide gels, failure to be stained with ethidium and a strong staining by silver (Fig. 3) support this notion. The CD spectrum of  $(CAG)_4$  was redshifted in comparison with the CD spectrum of  $(GA)_5$  (Fig. 8,a), but the redshift and the shoulder of an enhanced positive ellipticity around 290 nm originated from the chromophores of cytosine, as indicated by a comparison of the CD spectra of  $(GA)_5$  and  $(GA)_5(GC)_2(GA)_5$  where similar differences were observed (Fig. 8,a). The characteristic positive CD band originated from the unusual  $(CAG)_4$  conformation, and not simply from its primary structure, because the positive band was much reduced at high salt concentrations (Fig. 8), making the CD spectrum of  $(CAG)_4$  similar to the CD spectra of natural DNAs. Remarkably, the low-salt conformer of  $(CAG)_4$  provided an almost identical CD spectrum like the known (16 and references therein) extremely stable foldback of  $(GCGAAGC)$  (Fig. 8,b).

It follows from the above data that all four present  $(CNG)_4$  dodecamers folded into ordered conformers in the absence of the complementary strand. Addition of the complementary strand, however, immediately destabilized the ordered conformers, giving rise to the complementary strand heteroduplexes (Figs 3 and 9). No heating, annealing etc., necessary to prepare random sequence DNA duplexes, were needed for the complementary  $(CNG)_4$  strand heteroduplex formation both at the low (1 mM sodium phosphate) and moderate (0.3 M NaCl, not shown) ionic strengths. The duplex of  $(CCG)_4 \cdot (CGG)_4$  migrated like the duplex of  $(C_6G_6)$ . The duplex of  $(CTG)_4 \cdot (CAG)_4$  migrated slightly faster but much slower than the homoduplexes of the individual strands (Fig. 3). Non-complementary strands did not associate into heteroduplexes in any case (not shown). The CD spectral similarities of  $(CAG)_4$  and  $(GA)_n$  caused us to mix  $(CAG)_4$  with  $(GA)_5$ , but these oligonucleotides also did not mutually interact. The heteroduplexes were neither generated upon mixing  $(CG)_4$  with  $(CTG)_4$  nor  $(CCG)_4$ . This choice was



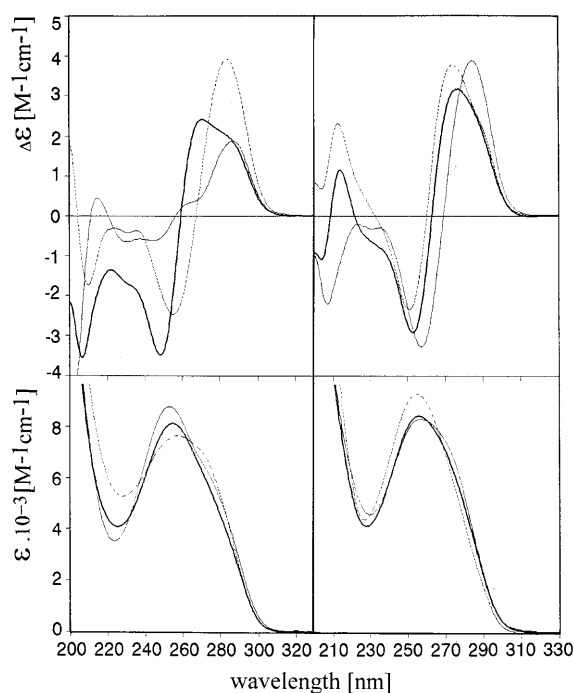
**Figure 8.** CD spectra of 0.4 mM (CAG)<sub>4</sub> in 1 mM sodium phosphate + 0.3 mM EDTA, pH 7, 0°C (trace 0); the same sample at 22°C (trace 1), then in 10 mM sodium phosphate and 0.3 mM EDTA, pH 7, plus 0 (trace 2), 2 (trace 3) and 5 M (trace 4) NaCl. Insert a, (thin trace) (GA)<sub>5</sub> in 10 mM sodium phosphate, 0.3 mM EDTA, pH 7, plus 0.5 M KCl, 0°C; (bold trace) (GA)<sub>5</sub>(GC)<sub>2</sub>(GA)<sub>5</sub> in 1 mM sodium phosphate + 0.3 mM EDTA, pH 7, 0°C. Insert b, (GCGAAGC) in 1 mM sodium phosphate and 0.3 mM EDTA, pH 7, 0°C.

motivated by an expectation that the central pyrimidine bases of the CTG or CCG triplets might bulge out to facilitate formation of eight consecutive GC pairs with the (CG)<sub>4</sub> octamer. However, this scenario did not work either. Heteroduplexes were only formed by the complementary dodecamers. Their CD and UV absorption spectra (Fig. 9) were different from the averages of the CD and UV spectra of their constituent strands. The CD spectrum of (CCG)<sub>4</sub>·(CGG)<sub>4</sub> resembled the CD spectra of (G+C)-rich DNAs, e.g. poly(GC). The positive band of (CAG)<sub>4</sub>·(CTG)<sub>4</sub> was slightly higher than with native DNAs containing 66% GC pairs.

## DISCUSSION

The primary structure of genomic DNAs undergoes permanent alterations that are the source of not only evolution but also diseases. Recently, dynamic mutations have been discovered (5) the essence of which is a change in the number of tandem copies in the simple sequence repeats, or microsatellites, which constitute a substantial part of the human genome (J.Kypr, D.Haring and M.Matula, unpublished data). Hence it was interesting to learn that expansions of the (CNG)<sub>n</sub> and other simple sequence repeats probably stood behind a number of serious, mostly neurological diseases (4).

It follows from a systematic work of our and other laboratories that the simple sequence repeats differ from non-repetitive sequences by conferring an extended conformational variability on DNA (22). Indeed, many recent studies have pointed to this conformational variability as the source of the simple sequence



**Figure 9.** CD (top panels) and UV absorption (bottom panels) spectra of: left panels, (CCG)<sub>4</sub> (dashed traces), (CGG)<sub>4</sub> denatured for 15 min at 90°C (thin traces) and the duplex of (CCG)<sub>4</sub>·(CGG)<sub>4</sub> (bold traces); right panels, (CAG)<sub>4</sub> (dashed traces), (CTG)<sub>4</sub> (thin traces) and the duplex of (CAG)<sub>4</sub>·(CTG)<sub>4</sub> (bold traces). Temperature 21.5°C, 0.7 mM oligonucleotide concentrations in 1 mM sodium phosphate + 0.3 mM EDTA, pH 7.

repeat expansions (reviewed in 6–8). The (CNG)<sub>n</sub> repeat conformations have so far mostly been studied by proton NMR and gel electrophoresis (7) which have many advantages but also weak points. The weak points include the fact that these methods permit the study of the oligonucleotides under only a limited number of the experimental conditions that are known to affect DNA conformation. Here, we offer a complementary study based on CD spectroscopy which permits to map the whole conformational spaces of the analyzed oligonucleotides and characterize how, and under what conditions, the oligonucleotides isomerize from one conformer into another.

The first of the major points of the present work is a demonstration that each of the four analyzed (CNG)<sub>4</sub> oligonucleotides can adopt more than a single conformer, depending on its concentration, presence of the complementary strand, ionic strength, pH and temperature. All of the four present oligonucleotide motifs were reported to adopt very stable foldbacks (reviewed in 7). In line with these reports we observed very stable foldbacks within 0–0.3 M NaCl with (CCG)<sub>4</sub>, (CTG)<sub>4</sub> and (CAG)<sub>4</sub>, but only bimolecular species with (CGG)<sub>4</sub>. All of these ordered (CNG)<sub>4</sub> conformers disappeared, however, in the presence of complementary strand (but not strands containing other sequences). This is the second major point of the work. The complementary (CNG)<sub>4</sub> strands are remarkably efficient reciprocal chaperones refolding the conformers containing non-canonical base–base interactions into the heteroduplexes of complementary sequences. The Watson–Crick base pairing is possible throughout the heteroduplexes but it is not certain that all of the complementary

bases participate in the Watson–Crick pairs in the  $(CAG)_n$ · $(CTG)_n$  and  $(CCG)_n$ · $(CGG)_n$  duplexes (8–10).

The third remarkable fact is that  $(CAG)_4$  adopts a structure very similar to the well-known extremely stable foldback of (GCGAAGC). This structure of  $(CAG)_4$  is simultaneously very similar to the homoduplex of  $(GA)_5$ . The available data suggest that the structure of  $(CAG)_4$  contains GC pairs in the stem, like the foldback of (GCGAAGC), yet the adenines influence the structure of  $(CAG)_4$  in such a way that its CD spectrum resembles the CD spectrum of the  $(GA)_5$  homoduplex. This suggests that the guanines of  $(CAG)_4$  not only pair with the cytosines, but that they significantly interact with the adenines as well. The adenines are probably highly tilted and stacked in the loop of (GCGAAGC). This point will certainly motivate further studies.

If guanine was the central triplet base in the  $(CNG)_4$  dodecamer, then we failed to observe any monomolecular species under the conditions when the other three  $(CNG)_4$  dodecamers were foldbacks. However  $(CGG)_{15}$  generates a very stable foldback (18) and we see no reason for a lower stability of the foldback of  $(CGG)_4$  as compared with the foldbacks of the other three present  $(CNG)_4$  dodecamers. Hence the apparent absence of foldbacks of  $(CGG)_4$  in the gels suggests that the foldbacks of  $(CGG)_4$  associate. The associations are only stable at low ionic strength. It is possible that they result from kissing of the foldback loops or from guanine tetrad formation by Hoogsteen paired GG pairs in the foldback stems. At moderate and high ionic strengths, the  $(CGG)_4$  dodecamer switched into a bimolecular homoduplex whose negative long wavelength part of the CD spectrum indicated a significant overwinding caused by the bulky GG pairs. At high oligonucleotide concentrations,  $(CGG)_4$  associated into a neutral guanine quadruplex like  $(TGG)_4$ . This observation has not yet been reported with any unmethylated sample of  $(CGG)_n$ . The neutral quadruplex of  $(CGG)_4$  was, however, not very stable owing to a competing CG pairing which did not destabilize the quadruplex of  $(TGG)_4$ . At acid pH, a guanine  $(CGG)_4$  tetraplex appeared at relatively very low oligonucleotide concentrations. Its aggregations were avoided in the absence of potassium cations.

If the central triplet base was thymine or cytosine, then the GC pairing entirely governed the foldback conformation. The foldbacks seemed to be very similar and, as usual, both were destabilized by the increasing ionic strength. High salt concentrations, however, induced a peculiar bimolecular complex of  $(CCG)_4$  which was also stabilized by high oligonucleotide concentrations. Its CD spectrum was reminiscent of the hemiprotonated cytosine duplex (23) or intercalated tetraplex (24). The two consecutive cytosines were

essential for stability of the  $(CCG)_4$  conformer because  $(CTG)_4$  exhibited no similar behavior. Stacking of two cytosines has been shown to give rise to cytosine protonation in a  $(CCG)_{15}$  hairpin even at neutral pH, but the protonation was not accompanied by  $CC^+$  pair formation (25). Further studies are in progress in our laboratory of the bimolecular complex of  $(CCG)_4$ .

## ACKNOWLEDGEMENT

This work was supported by grant no. 204/95/1270 kindly awarded to M.V. by the Grant Agency of the Czech Republic.

## REFERENCES

- 1 Tautz, D., Trick, M. and Dover, G.A. (1986) *Nature*, **322**, 652–656.
- 2 Han, J., Hsu, H., Zhu, Z., Longshore, J.W. and Finley, W.H. (1994) *Nucleic Acids Res.*, **22**, 1735–1740.
- 3 Mrázek, J. and Kypr, J. (1994) *Miami Bio/Technol. Short Reports*, **5**, 39.
- 4 Ashley, C.T., Jr and Warren, S.T. (1995) *Annu. Rev. Genet.*, **29**, 703–728.
- 5 Richards, R.I. and Sutherland, G.R. (1997) *Trends Biochem. Sci.*, **22**, 432–436.
- 6 Wells, R.D. (1996) *J. Biol. Chem.*, **271**, 2875–2878.
- 7 Mitas, M. (1997) *Nucleic Acids Res.*, **25**, 2245–2253.
- 8 Kypr, J. and Vorlíčková, M. (1997) *Chemická Listy*, **91**, 570–574.
- 9 Pearson, C.E. and Sinden, R.R. (1996) *Biochemistry*, **35**, 5041–5053.
- 10 Mitchell, J.E., Newbury, S.F. and McClellan, J.A. (1995) *Nucleic Acids Res.*, **23**, 1876–1881.
- 11 Bacolla, A., Gellibolian, R., Shimizu, M., Amirhaeri, S., Kang, S., Ohshima, K., Larson, J.E., Harvey, S.C., Stolar, B.D. and Wells, R.D. (1997) *J. Biol. Chem.*, **272**, 16783–16792.
- 12 Gao, X., Huang, X., Smith, G.K., Zheng, M. and Liu, H. (1995) *J. Am. Chem. Soc.*, **117**, 8883–8884.
- 13 Fry, M. and Loeb, L.A. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4950–4954.
- 14 Usdin, K. and Woodford, K.J. (1995) *Nucleic Acids Res.*, **23**, 4202–4209.
- 15 Chen, F.-M. (1995) *J. Biol. Chem.*, **270**, 23090–23096.
- 16 Yoshizawa, S., Kawai, G., Watanabe, K., Miura, K. and Hirao, I. (1997) *Biochemistry*, **36**, 4761–4767.
- 17 Gray, D.M., Hung, S.-H. and Johnson, K.H. (1995) *Methods Enzymol.*, **246**, 19–34.
- 18 Mitas, M., Yu, A., Dill, J. and Haworth, I.S. (1995) *Biochemistry*, **34**, 12803–12811.
- 19 Penázová, H. and Vorlíčková, M. (1997) *Biophys. J.*, **73**, 2054–2063.
- 20 Chen, F.-M. (1997) *Biophys. J.*, **73**, 348–356.
- 21 Rippe, K., Fritsch, V., Westhof, E. and Jovin, T.M. (1992) *EMBO J.*, **11**, 3777–3786.
- 22 Vorlíčková, M. and Kypr, J. (1985) *J. Biomol. Struct. Dyn.*, **3**, 67–83.
- 23 Antao, V.P. and Gray, D.M. (1993) *J. Biomol. Struct. Dyn.*, **10**, 819–839.
- 24 Gehring, K., Leroy, J.-L. and Guéron, M. (1993) *Nature*, **363**, 561–565.
- 25 Yu, A., Barron, M.D., Romero, R.M., Christy, M., Gold, B., Dai, J., Gray, D.M., Haworth, I.S. and Mitas, M. (1997) *Biochemistry*, **36**, 3687–3699.