Published in final edited form as:

Anal Biochem. 2007 August 1; 367(1): 56–67.

Diffusion ordered NMR spectroscopy for analysis of DNA secondary structural elements:

“To: DNA Recombination Techniques and Nucleic Acids”

Attila Ambrus^{*,1} and Danzhou Yang^{*,1}

Department of Pharmacology and Toxicology, University of Arizona, Tucson, Arizona, 85721, USA

Abstract

Structure determination of secondary DNA structural elements, such as G-quadruplexes, gains an increasing importance as fundamental physiological roles are being associated with the formation of such structures *in vivo*. A truncated native DNA sequence generally requires further optimization to obtain a candidate with desired NMR properties for structural analysis in solution. The optimum sequence is expected to form one dominant, stable molecular entity in solution with well-resolved NMR peaks. However, DNA sequences are prone to form structures composed of one, two, three or four strands depending on sequence and solution conditions. The thorough characterization of the molecularity (stoichiometry and molecular weight) and appropriate solution conditions for sequences with different modifications traditionally applies analytical techniques that generally do not represent the solution conditions for NMR structure determination. Here we present the application of diffusion ordered NMR spectroscopy as a useful analytical tool for the optimization and analysis of DNA secondary structural elements, specifically, the DNA G-quadruplex structures, including those formed in the human telomeric sequence and in the promoter regions of bcl-2 and c-myc genes.

Keywords

DNA; NMR; DOSY; quadruplex; bcl-2; human telomere; c-myc

Structural analysis of unusual DNA motifs with particular emphasis on G-quadruplexes and i-motifs has a central role in the identification and characterization of molecular targets related to these structures [1-7]. Such structures are implicated in several physiologically important regulatory processes, with special emphasis on carcinogenesis [4-12]. Most of the structural studies on these elements have been performed by NMR spectroscopy. In most cases the related nucleotide sequence bearing the actual physiological function *in vivo* needs to be truncated/elongated and/or mutated in order to obtain a successful candidate for NMR structure determination [3-7,13]. This NMR candidate should be a well-defined molecular entity in physiologically relevant solution conditions (ionic strength, pH, temperature) representing the same conformation as the original sequence. Appropriate modification of the sequence by mutation/truncation/elongation will force the molecule to form a single stable structure without altering crucial connectivities (conformation) of the structure.

*To whom correspondence should be addressed at the Department of Pharmacology and Toxicology, College of Pharmacy, University of Arizona, 1703 E. Mabel St., P.O. Box 210207, room 101/406, Tucson, Arizona 85721, USA; Tel: +1 520 626 6749/626 5969, Fax: +1 520 626 2466, E-mail: ambrus@pharmacy.arizona.edu, yangd@pharmacy.arizona.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Sequences comprising multiple guanine runs show a propensity to form G-quadruplex structures [14,15]. The physiologically relevant form of G-quadruplexes is unimolecular as these structures generally fold up at the ends of telomeres from single-stranded DNA [5,7] or loop out in the promoter regions from double-stranded DNA [3,4,6]. Depending on sequence and solution conditions, some sequences form very stable unimolecular quadruplexes in physiological buffer conditions even at millimolar concentrations while other sequences tend to form higher order structures composed of multiple strands. The actual molar ratio of the desired molecular entity (oligomerization state) in the sample is also a function of strand concentration.

Diffusion ordered NMR spectroscopy is a method for the identification of different molecular weight (macro)molecules present in solution at millimolar concentrations, while the successful acquisition of the signal requires a sufficient concentration of each component [16-33]. The method uses a special pulse-sequence to obtain a series of NMR spectra using field-gradients and is capable of determining diffusion coefficients of molecules without any concentration-gradient [19,20]. The intensity of the detected proton signal belonging to a particular molecular entity at a certain gradient level is dependent mostly on the rate of diffusion of that molecule as the gradient eliminates more signal intensity of molecules in faster motion. The analysis of attenuation in signal intensity as a function of diffusion time, gradient strength or gradient time (depending on the experimental set-up) provides the related diffusion coefficients. The number of identifiable components in solution is only limited by resolution (if species exchange slowly on NMR time scale) [25,26]. Comparing the diffusion coefficients of target molecules to that of reference(s), the molecular weight can be calculated [19]. The MWs of the applied references should be extended to both sides of the MW range of the molecule being analyzed, taking the most probable oligomerization orders expected into consideration. It is preferable, but not a requirement, that the chemical nature of the references and the molecule in question is similar; what is more important is the shape, which determines how molecules behave during diffusion [19].

The first pulse sequence for diffusion NMR was established in the 1960s [16]. Since then several 1-3D versions of diffusion NMR experiments have been developed even in combination with regular magnetization transfer experiments (COSY, NOESY, TOCSY, HMQC, etc.) to provide MW information for samples being analyzed [32,34-39]. Different forms of diffusion NMR experiments have been applied for diverse varieties of systems including inorganic materials [37,40-42], organic molecules [17,23,24,28,43-49] and polymers of different kinds (inorganic and organic polymers, polysaccharides, proteins) [21,37,50-62]. Application for nucleic acids has been rather limited because the well-defined double helical structures have been investigated for most cases. However, the diffusion NMR experiment can be a useful analytical tool for the unusual DNA secondary structures, such as G-quadruplexes that can comprise one, two and four strands. In this contribution we show that 2D DOSY NMR is capable of determining the molecular weights of DNA G-quadruplexes in NMR samples and thus is a useful technique during the optimization of sequence and solution conditions for structure determination of DNA secondary structural motifs. We present the DOSY analysis of several molecular systems including the physiologically relevant promoter quadruplexes of the bcl-2 and the c-myc genes and the quadruplexes formed in the human telomeric sequence.

Materials and Methods

Materials

Chemicals for solid phase DNA chemistry were all purchased from Glen Research; all other chemicals are from Sigma. The two low-MW references used are: Trypan Blue ($C_{34}H_{24}N_6O_{14}S_4Na_4$, MW=960.8 g/mol; MW=868.8 g/mol without the four Na^+ in solution) and Xylene Cyanole FF ($C_{25}H_{27}N_2O_6S_2Na$, MW=538.6 g/mol; 515.6 g/mol without the

Na⁺ in solution). The two high-MW references are Lysozyme (from chicken egg white, MW=14,300 Da) and Ribonuclease A (from bovine pancreas, MW=13,700 Da).

DNA synthesis and purification

All the sequences were synthesized as single strands by standard solid phase β -cyanoethyl-phosphoramidite chemistry on a 1 μ mol scale on an Expedite 8909 Nucleic Acid Synthesis System (Perceptive Biosystems) in DMT-on mode. Sequences were cleaved from the solid phase synthesis columns and deprotected by NH₃ at 55 °C for 14 h and were subsequently purified by HPLC. Elimination of the DMT group was carried out by 80 % AcOH treatment for 1 h. DNA was transferred to a pure water phase by ethyl ether extraction and was further purified by HPLC and successive dialysis against 150 mM NaCl and H₂O. Samples were lyophilized from pure water and dissolved generally in 50 mM K-PO₄, 140 mM KCl, pH 7.0, 5 % D₂O (for field lock), if not stated otherwise. The sample concentration was determined by absorption at 260 nm using calculated extinction coefficients. The following sequences are used as case studies: a mutated bcl-2 promoter DNA (G₃CGCG₃AGGAATTG₃CG₃, MW=7291.76 Da) [4,6], a 22-nt human telomeric DNA (Tel22: AG₃(T₂AG₃)₃ [1], MW=6966.58 Da), a mutated 22-nt human telomeric DNA (MW=6951.58 Da; to be published elsewhere) and a mutated human c-myc promoter DNA (TG(AG₃TG₃T)₂A₂, MW=6991.59 Da) [3].

Mass Spectrometry

Mass spectrometry analyses were performed on a Thermoquest LCQ Classic and a Bruker Reflex III MALDI-TOF instrument. Samples measured by MALDI-TOF were mixed with a saturated solution of sinapinic acid (3,5-Dimethoxy-4-hydroxycinnamic acid) in 70:30 H₂O:Acetonitrile containing 0.1 % TFA, in such a way that the total amount of material deposited was 1-2 pmols. In general, a laser attenuation of 60-80 % was used. ESI experiments were performed by the HPLC-MS technique. Protein solutions were run with the moving phase of 1:1 MeOH:H₂O containing 2 % AcOH while the small molecule references were run in 1:1 MeOH:ACN.

NMR experiments

All measurements were performed on a Bruker DRX-600 with a TXI gradient probe-head in 5 mm NMR tubes with limited sample volumes (~0.4-0.5 ml) to use the linear range of the gradient coils. For integrity and purity analysis, 1D proton spectra were recorded before all the DOSY experiments, generally by the standard Bruker microprogram *zgpr* that applies presaturation for water suppression. The standard Bruker microprogram *stebugp1s19* was applied to obtain diffusion ordered spectroscopy (DOSY) spectra at 298 K. The pulse-program applies stimulated echoes using bipolar gradient pulses for diffusion and 3-9-19 pulses with gradients for water suppression. For each FID, 512 transients were collected with 3 s relaxation delay and a 20 μ s delay for binomial water suppression. 4096 data points in the F2 dimension (20 ppm) and 16-32 data points (gradient strengths) in the F1 dimension were collected for all experiments. Final data sizes were 4096 $\times$ 1024. Exponential multiplication was applied in F2 with 1 Hz line broadening. The diffusion time (Δ) and the gradient length (δ) were set to 100 ms and 1ms, respectively, while the recovery delay after gradient pulses was 200 μ s. Two types of data analyses were applied to the raw experimental data. For automatic 2D-processing, the standard 2D DOSY processing protocol was applied in XWINNMR software with logarithmic scaling in the F1 (diffusion coefficient) dimension. For manual curve-fitting, the intensities of selected peaks in the 1D proton spectra measured at different gradient strengths were fitted using the equation $I = I_0 \exp(-D\gamma^2 g^2 \delta^2 (\Delta - \delta/3))$ [$\rightarrow \sqrt{-\ln(I/I_0)} = \sqrt{D^*} g$] [19] to obtain the apparent diffusion coefficient D^* . In this theoretical equation the following physical quantities are symbolized: I, the actual (measured) peak intensity; I_0 , peak intensity at zero gradient

strength; D , diffusion coefficient; γ , gyromagnetic ratio (of proton); g , gradient strength; δ , length of gradient; and Δ , diffusion time. Theoretically the length of gradient and the diffusion time can also be incremented in diffusion experiments, however, most pulse-schemes modify the gradient strength (g). Since D , γ , Δ and δ are constant, in $D\gamma^2g^2\delta^2(\Delta-\delta/3)$ they can be converted to be under a new constant, D^* ($D^*=cD$, where $c=\gamma^2\delta^2(\Delta-\delta/3)$ and is a constant). By mathematical rearrangement of the original equation and substitution of the new constant (D^*), a linear equation is deduced [$\sqrt{-\ln(I/I_0)}=\sqrt{D^*}g$] (see above), that is applicable in determining the diffusion coefficient. On these plots, gradient strengths are represented as the linearly changing increments of the total gradient strength between 5% and 95% (16 or 32 increments were applied). As shown in the equation, the slope of the fitted line is equal to the square root of D^* , so D^* can be calculated from the value of the slope. The actual molecular weights relative to the references can thus be determined by the following equation, $\log(D_1/D_2)=1/3*\log(MW_2/MW_1)$, where $D_1/D_2=D^*/D^*$ [19]. This equation assumes that the molecules being compared have the same overall shapes and relaxation properties.

Results

Selection of reference materials

As discussed above, the molecular weights of the applied references have to be selected on the basis of the MW of the molecule in question. Since the purpose of the DOSY analysis is to determine the oligomeric state(s) present, the referencing MW calibration curve needs to cover a reasonable MW region according to the possible oligomeric states of the molecule under analysis. Beyond an appropriate and well-defined molecular weight stability, the purity, overall shape, diffusion and basic NMR properties (line-width, peak-separation, relaxation) are of great importance (rather than chemical similarity) in the selection of the adequate references. Optimal referencing in this study was achieved by applying two low- and two high-MW reference materials (see the Materials and Methods section) that cover the MW range of 0.5-14 kDa (the MW of a monomeric quadruplex is ~ 7 kDa) and possess all the desirable properties described above. As part of the characterization of the applied references, HPLC-MS and MALDI-TOF Mass Spectrometry analyses were carried out on the reference molecules to confirm their integrity, molecular weight and purity (Figure 1).

Bcl-2 promoter G-quadruplex

The first example is the result of our sequence optimization for the major G-quadruplex formed in the bcl-2 promoter region [4,6]. This sequence is relatively (compared to, for example, the c-MYC quadruplex) unstable in time and can form aggregates, especially at higher concentrations. The maximum strand concentration that could be successfully applied for long NMR experimentation is ~ 1.5 mM. A fresh sample at ~ 1 mM, however, could always be safely used. Figure 2A shows the bcl-2 DNA promoter sequence used for structural studies, with one major quadruplex species present (indicated by 12 well-resolved imino peaks in H_2O). As seen in Figure 2A, the same major species is formed at 0.1 mM and 1.5 mM, which is used for NMR structure determination. The NMR titration experiment (Figure S1B) shows that this major species is from a unimolecular structure. Figure 2B shows the 2D DOSY analysis of the final NMR sample used for structure determination (in D_2O). The diffusion coefficient scale shows only one major molecular weight present in the system. The F1 projection of the 2D DOSY experiment averages the contributions of different peaks and provides a way to estimate the D value.

The apparent diffusion coefficient in the bcl-2 G-quadruplex system was also obtained by manual curve fitting of the DOSY data. Specifically, we selected a well-resolved peak of the DNA (designated by “x” on the 1D 1H (upper) projection, Figure 2B), measured its intensity as a function of gradient strength and fitted the data to the linearized theoretical equation (see

Materials and Methods). Figure 3 shows the fitting analysis of the selected peaks of bcl-2 and the external reference standards (see Materials and Methods). The slope of the fitted line is equal to the square root of D^* , so D^* can be calculated from the value of the slope. Furthermore, as shown in Materials and Methods, $D_1/D_2 = D_1^*/D_2^*$, and $\log(D_1/D_2) = 1/3 \log(M_2/M_1)$, thus the MW of the target molecule can be calculated from those of the reference molecules. This value is specific to the selected peak, however, if we consider the same or similar relaxation properties for all protons in the molecule, it can be accepted as the diffusion coefficient of an average proton in the molecule. To confirm this, we have also performed the same curve fitting for other protons of the bcl-2 sequence, which gave rise to the same result in the experimental error range. By this calculation we obtained the MW for the bcl-2 promoter G-quadruplex, which equals the calculated MW within the experimental error range of $\sim 3\%$. The observed error can come from both the experimental errors and the different relaxation properties (and shapes) of the reference and DNA molecules, as the two theoretical equations do not contain relaxation terms. As shown below, the apparent diffusion coefficient (D^*) values (and thus the resultant actual D values) are considerably dependent on the nature of solution conditions. Since solution conditions remarkably alter diffusion behaviors, the same buffer conditions need to be used for the target DNA molecule and the reference molecules.

Human telomeric mutant G-quadruplex

The 22-nt human telomeric sequence (Tel22) forms a major basket-type quadruplex structure in Na^+ solution [1], however, the sequence forms multiple conformations in the physiologically relevant K^+ solution (Figure 6A; the number of imino protons is much more than 12) [5,7]. A mutated 22-nt Tel22 sequence was identified and has been shown to form one major G-quadruplex conformation (Figure 4A, upper projection). The NMR titration experiment was performed for Tel22-Mut. The slope of ~ 1 indicates that this major species is from a unimolecular structure (Figure S2). Tel22-Mut shows a remarkably different NMR signature from that of Tel22, with the number of imino protons being less than twelve. The exchangeable imino proton region is the most sensitive to changes in a G-quadruplex structure. This indicates that the conformation of a G-quadruplex forming nucleotide sequence is sensitive to mutations, especially to ones in structurally deterministic regions. The well-resolved NMR spectra indicate that this sequence is appropriate for NMR structure determination, which will be reported elsewhere. In Figure 4A, the 2D DOSY projection analysis of Tel22-Mut at 1 mM strand concentration in K^+ solution is shown. As seen there is one dominant projection peak. Since the appropriate buffer solution conditions for Tel22-Mut are different from that for the bcl-2 promoter G-quadruplex system, separate experiments were recorded for the MW references in the Tel22-Mut solution conditions. The molecular weight of Tel22-Mut was determined using low-MW references (Figure 4B) and high-MW references (Lysozyme and Ribonuclease A, Figure 6D&E). The experimental error for the Tel22-Mut MW was $\sim 11\%$ with the low-MW standards and was $\sim 8\%$ with the high-MW standards.

In addition, we also carried out the manual calculation of diffusion coefficients of Tel22-Mut and the two low-MW references (Figure 5). One isolated peak was selected for the major species of Tel22-Mut (Figure 5C, diamond) for the curve-fitting calculation. The fitting of peak intensities against the gradient strength was plotted in Figure 5A (in insets the direct intensity-gradient ratio plots are seen). Similarly, one peak is selected for each of the low-MW references (Figure 4B), and the intensities of these peaks were plotted against the gradient strength. The comparison of the slope values of Tel22-Mut and the references gave rise to a MW of Tel22-Mut, which is within $\sim 13\%$ error from the calculated value. A minor species was also detected for Tel22-Mut (Figure 5C, square). We also selected one isolated peak for this minor species and performed manual curve-fitting calculation to determine the state of this minor species (Figure 5A). Interestingly, the result gives rise to the same MW as the major species of Tel22-Mut, indicating that the minor species is also unimolecular.

Human telomeric wild-type G-quadruplexes

As shown in Figure 6A, the wide-type Tel22 sequence forms multiple conformations in potassium solution. We have also carried out 2D DOSY experiment for Tel22 in potassium solution using the same solution conditions as those of Tel22-Mut. The F1 projection of the 2D DOSY is shown in Figure 6B. It appears that Tel22 has a close shoulder peak in its DOSY projection, the intensity of which is comparable to the dominant peak. The diffusion coefficients of both peaks are very similar to that of Tel22-Mut. Tel22 in K^+ solution forms a mixture of two unimolecular conformations (Figure S3B) [5,7,64], whose folding and consequently the overall shapes are different, which may correspond to the two slightly different projection peaks. The MS analysis of the Tel22 sample showed no higher or lower molecular weights present in this sample (data not shown).

c-myc promoter G-quadruplex

We have also applied DOSY for a c-myc promoter DNA sequence that has been shown to form the physiologically relevant major G-quadruplex structure in the silencer element of the c-myc gene [3,63]. The wild-type sequence (AG₃TG₄AG₃TG₄) with a minimal length (18-nt) for the formation of G-quadruplex adopts a mixture of four unimolecular quadruplex conformations in K^+ solution [63]; its 1D 1H NMR spectrum consists of broad envelopes (data not shown). The DOSY analysis of this 18-mer sequence also shows broadening towards higher molecular weights even at low millimolar concentrations (data not shown), suggesting possible multimer formation as well. The optimization of this sequence for NMR structure determination has been performed by elongation and mutation, resulting in a modified sequence of (TGAG₃TG₃TAG₃TG₃TAA) which selected out the predominant G-quadruplex formed in the c-MYC promoter sequence [63]. The structure of this successful NMR candidate has already been solved in our laboratory [3]. The DOSY F1 projection of this sequence can be seen in Figure 6C, representing a sharp dominant peak at the monomeric molecular weight at a millimolar NMR concentration (referencing standards are not shown).

Discussion

Molecular weight determination with DOSY-NMR is sensitive to the shape and relaxation properties of the molecules, as well as other solution conditions. In addition, exchange processes, such as observed in the case of Tel22, can also broaden the peaks and decrease precision. Therefore the accuracy of molecular weight determination by DOSY-NMR cannot be compared to high-resolution methods like mass spectrometry. However, DOSY-NMR can provide very useful information on the stoichiometry of macromolecules, including multi-stranded DNA conformations in the actual NMR sample used for structure determination. This information about the oligomeric states (of, e.g., G-quadruplexes) cannot be obtained by any other analytical technique regularly applied in the DNA field (see below). Since the NMR effect is rather weak, the optimal DNA strand concentration for NMR experiments is in the millimolar range. During the optimization of the DNA sequence for NMR studies, a selection of different analytical tools (including NMR spectroscopy) can be used for the characterization of a sequence in a given set of solution conditions. However, most methods are used at a concentration that is several orders of magnitude lower than that used for NMR spectroscopy. For instance, using electromobility shift assay (EMSA), the G-quadruplex-forming DNA sequence can be compared to other DNA sequences with various lengths forming extended single stranded structures, unimolecular, dimeric or tetrameric G-quadruplex structures. The results give a good indication of structural preference for the DNA sequence, but the concentration range of EMSA experiments is micromolar compared to a millimolar range for the NMR sample. The three orders of magnitude difference in working concentration and the unpredictable effects of gel analysis may not provide unambiguous information for the actual settings in solution NMR conditions. The same argument is valid for size-exclusion

chromatographic analyses of molecular weights of nucleic acids, which involve extensive dilution on the column as well as matrix effects. Mass spectrometry is capable of preserving and analyzing non-covalent adducts, but its working concentration is generally even lower than that of the methods mentioned above and the ionization process might be destructive. Both CD spectroscopy and analytical ultracentrifugation have the advantage of being matrix-free analyses. CD spectroscopy as a qualitative method can be useful especially for defining the probable folding pattern of the G-quadruplexes [5]; however, it generally cannot provide information on the extent of oligomerization, and its working concentration is about two orders of magnitude lower than that of NMR. Analytical ultracentrifugation has the same working concentration as CD spectroscopy. In contrast, NMR titration methods can be used for the determination of the extent of oligomerization of G-quadruplexes in the actual NMR sample environment [5,6,13,15]. During the analysis, the melting point of the DNA structural element is determined first (50% transformation to the unfolded conformation of a single strand, represented by very sharp resonances in the 1D ^1H NMR spectrum). Subsequently, the sample is titrated at the melting temperature with a concentrated solution of the same DNA strand and the ratio of the folded and the unfolded forms is determined as a function of the overall strand concentration. Then the slope of the concentration dependence curve is determined from the intensities (integrals) of the two peaks, one belonging to the folded form and one to the unfolded (melted) form. If the ratio, i.e., the slope, of the two forms is around 1, then the folded structure represents a unimolecular species (1:1 stoichiometry); if the ratio is close to an integer other than one, then this number provides the oligomeric state. The method works well when one major single G-quadruplex species is present, but it is generally difficult to separate amongst signals belonging to different oligomeric states when they co-exist in solution (especially because the oligomer peaks can be much broader) and the correct analysis can be obscured.

Even if there are a few instances when the co-existence of multiple oligomer states of G-quadruplexes does not hinder the possibility of recording and even interpreting the NMR data, this condition should be avoided for better spectrum quality and easier interpretation. In most cases the exchange effects (among oligomer states) broaden the signals and decrease the peak intensity and resolution, resulting in more overlapping spectra which are challenging to interpret. In these cases the solution conditions (concentration, pH, temperature, ionic strength) must be optimized to generate one major form of the G-quadruplex in solution, which should show promising NMR spectral qualities. If this does not work for the quadruplex sequence, other sequence modifications must be considered. Therefore, the nature and number of molecular entities present in the NMR sample are vital pieces of information for the efficient structural analysis of G-quadruplexes, which necessitate a methodology by which this information can be obtained.

The appearance of envelopes and broader signals in the 1D ^1H spectrum does not necessarily mean the existence of multimers or different conformations in exchange. It is our experience that similar effects in the 1D proton NMR spectra can occur when DNA purification for a G-quadruplex forming sequence was not performed carefully enough and some synthesis intermediates irreversibly coordinate to the target DNA sequence. There are several critical points during the purification process when this coordination can happen, e.g., when sequences are not cut off in a timely fashion from the solid synthesis matrix or the HPLC columns are not thoroughly washed between purification steps, to name a few. A promising NMR sequence candidate can be misinterpreted if this unexpected coordination is confused with multimerization of G-quadruplexes. This confusion can be clarified with the same sample, after recording the first 1D ^1H NMR spectrum for sequence optimization, by performing a 2D DOSY experiment. If there is one distinct peak in DOSY, careful purification can provide better NMR signatures. However, it needs to be noted that, for some unimolecular G-quadruplex candidates, good spectral qualities can only be achieved by additional modifications in structurally non-critical regions of the G-quadruplex, e.g., in loop or capping regions, or/and by careful tuning

of the solution conditions (especially temperature and buffer composition). If multiple peaks with distinct molecular weights appear in the diffusion NMR spectra, then further optimization of solution conditions (including DNA strand concentration) may be needed, or, if this strategy fails, new G-quadruplex sequence modifications must be considered.

The sequence and solution condition optimization of several G-quadruplex sequences that are under structural analysis in our laboratory have been performed by systematic modifications and testing [3-7]. The identification of the best NMR candidates is performed by the combination of analytical methods mentioned above together with the 2D DOSY-NMR technique. Since the optimized sequences in the applied solution conditions do not show considerable additional oligomer formation, the NMR structure determination was/is reliable. For the Tel22-Mut sequence that shows only one unimolecular conformation, structure determination is in progress in our laboratory.

In summary, we have demonstrated that the diffusion NMR technique is particularly useful in the sequence and solution condition optimization for NMR structure determination of DNA molecules, especially the DNA G-quadruplexes, which exhibit high propensity for oligomer formation. This technique is convenient since the experiment does not need a different sample preparation strategy and can be performed on the same (high concentration) sample used for regular NMR analyses, such as signal resolution, line width, and structure determination. The ultimate advantage of the method is that it is able to determine the molecular weight and its distribution with good accuracy *in situ* in the actual NMR sample tube being investigated for structural information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank to Dr. Árpád Somogyi (Department of Chemistry, University of Arizona) for the mass spectra and the valuable consultations. We also thank Dr. Megan Carver and Ms. Tiffanie Bialis for proofreading the manuscript. This research was supported by the National Institutes of Health (1K01CA83886 and 1S10 RR16659 to D. Yang).

Abbreviations used

NMR, Nuclear Magnetic Resonance; DOSY, Diffusion Ordered Spectroscopy; MS, Mass Spectrometry; DNA, Deoxyribonucleic Acid; AcOH, Acetic Acid; HPLC, High Performance Liquid Chromatography; MW, Molecular Weight; MALDI, Matrix Assisted Laser Desorption/Ionization; ESI, Electrospray Ionization; TOF, time-of-flight; Tel22, human telomeric 22-mer DNA; Tel22-Mut, mutant of the Tel22 sequence.

References

1. Wang Y, Patel DJ. Solution Structure of the Human Telomeric Repeat D[Ag(3)(T(2)Ag(3))3] G-Tetraplex. *Structure* 1993;1:263–282. [PubMed: 8081740]
2. Parkinson GN, Lee MPH, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* 2002;417:876–880. [PubMed: 12050675]
3. Ambrus A, Chen D, Dai JX, Jones RA, Yang DZ. Solution structure of the biologically relevant g-quadruplex element in the human c-MYC promoter. implications for g-quadruplex stabilization. *Biochemistry* 2005;44:2048–2058. [PubMed: 15697230]
4. Dai JX, Dexheimer TS, Chen D, Carver M, Ambrus A, Jones RA, Yang DZ. An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc* 2006;128:1096–1098. [PubMed: 16433524]

5. Ambrus A, Chen D, Dai JX, Bialis T, Jones RA, Yang DZ. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res* 2006;34:2723–2735. [PubMed: 16714449]
6. Dai JX, Chen D, Jones RA, Hurley LH, Yang DZ. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res* 2006;34:5133–44. [PubMed: 16998187]
7. Dai JX, Punchihewa C, Ambrus A, Chen D, Jones RA, Yang DZ. Structure of the intramolecular human telomeric G-quadruplex in potassium solution: A novel adenine triple formation. *Nucleic Acids Res.* March 29;2007 published online
8. Hurley LH. DNA and its associated processes as targets for cancer therapy. *Nat. Rev. Cancer* 2002;2:188–200. [PubMed: 11990855]
9. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U. S. A* 2002;99:11593–11598. [PubMed: 12195017]
10. Rezler EM, Bearss DJ, Hurley LH. Telomeres and telomerases as drug targets. *Curr. Opin. Pharmacol* 2002;2:415–423. [PubMed: 12127874]
11. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. The role of secondary DNA structures in silencing transcription. *Eur. J. Cancer* 2002;38:S115–S115.
12. Grand CL, Bearss DJ, Von Hoff DD, Hurley LH. Quadruplex formation in the c-MYC promoter inhibits protein binding and correlates with in vivo promoter activity. *Eur. J. Cancer* 2002;38:S106–S107.
13. Phan AT, Modi YS, Patel DJ. Propeller-type parallel-stranded g-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc* 2004;126:8710–8716. [PubMed: 15250723]
14. Simonsson T. G-quadruplex DNA structures - Variations on a theme. *Biol. Chem* 2001;382:621–628. [PubMed: 11405224]
15. Phan, AT.; Gueron, M.; Leroy, JL. *Nuclear Magnetic Resonance of Biological Macromolecules*. 338. Academic Press Inc; San Diego: 2001. p. 341-371.
16. Stejskal EO, Tanner JE. Spin Diffusion Measurements - Spin Echoes in Presence of a Time-Dependent Field Gradient. *J. Chem. Phys* 1965;42:288.&
17. Morris KF, Stilbs P, Johnson CS. Analysis of Mixtures Based on Molecular-Size and Hydrophobicity by Means of Diffusion-Ordered 2d Nmr. *Anal. Chem* 1994;66:211–215.
18. Barjat H, Morris GA, Smart S, Swanson AG, Williams SCR. High-Resolution Diffusion-Ordered 2d Spectroscopy (Hr-Dosy) - a New Tool for the Analysis of Complex-Mixtures. *J. Magn. Reson. Ser. B* 1995;108:170–172.
19. Waldeck AR, Kuchel PW, Lennon AJ, Chapman BE. NMR diffusion measurements to characterise membrane transport and solute binding. *Prog. Nucl. Magn. Reson. Spectrosc* 1997;30:39–68.
20. Johnson CS. Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Prog. Nucl. Magn. Reson. Spectrosc* 1999;34:203–256.
21. Gounarides JS, Chen AD, Shapiro MJ. Nuclear magnetic resonance chromatography: applications of pulse field gradient diffusion NMR to mixture analysis and ligand-receptor interactions. *J. Chromatogr. B* 1999;725:79–90.
22. Loening NM, Keeler J, Morris GA. One-dimensional DOSY. *J. Magn. Reson* 2001;153:103–112. [PubMed: 11700086]
23. Hodge P, Monvisade P, Morris GA, Preece I. A novel NMR method for screening soluble compound libraries. *Chem. Commun* 2001:239–240.
24. Simpson AJ. Determining the molecular weight, aggregation, structures and interactions of natural organic matter using diffusion ordered spectroscopy. *Magn. Reson. Chem* 2002;40:S72–S82.
25. Cabrita EJ, Berger S. HR-DOSY as a new tool for the study of chemical exchange phenomena. *Magn. Reson. Chem* 2002;40:S122–S127.
26. Cabrita EJ, Berger S, Brauer P, Karger J. High-resolution DOSY NMR with spins in different chemical surroundings: Influence of particle exchange. *J. Magn. Reson* 2002;157:124–131. [PubMed: 12202141]

27. Lucas LH, Otto WH, Larive CK. The 2D-J-DOSY experiment: Resolving diffusion coefficients in mixtures. *J. Magn. Reson* 2002;156:138–145. [PubMed: 12081451]
28. Viel S, Mannina L, Segre A. Detection of a pi-pi complex by diffusion-ordered spectroscopy (DOSY). *Tetrahedron Lett* 2002;43:2515–2519.
29. Huo R, Wehrens R, van Duynhoven J, Buydens LMC. Assessment of techniques for DOSY NMR data processing. *Anal. Chim. Acta* 2003;490:231–251.
30. Thrippleton MJ, Loening NM, Keeler J. A fast method for the measurement of diffusion coefficients: one-dimensional DOSY. *Magn. Reson. Chem* 2003;41:441–447.
31. Cobas JC, Martin-Pastor M. A homodecoupled diffusion experiment for the analysis of complex mixtures by NMR. *J. Magn. Reson* 2004;171:20–24. [PubMed: 15504677]
32. Nilsson M, Gil AM, Delgadillo I, Morris GA. Improving pulse sequences for 3D diffusion-ordered NMR spectroscopy: 2DJ-IDOSY. *Anal. Chem* 2004;76:5418–5422. [PubMed: 15362901]
33. Huo R, Wehrens R, Buydens LMC. Improved DOSY NMR data processing by data enhancement and combination of multivariate curve resolution with non-linear least square fitting. *J. Magn. Reson* 2004;169:257–269. [PubMed: 15261621]
34. Wu DH, Chen AD, Johnson CS. Three-dimensional diffusion-ordered NMR spectroscopy: The homonuclear COSY-DOSY experiment. *J. Magn. Reson. Ser. A* 1996;121:88–91.
35. Gozansky EK, Gorenstein DG. DOSY-NOESY: Diffusion-ordered NOESY. *J. Magn. Reson. Ser. B* 1996;111:94–96. [PubMed: 8620290]
36. Barjat H, Morris GA, Swanson AG. A three-dimensional DOSY-HMQC experiment for the high-resolution analysis of complex mixtures. *J. Magn. Reson* 1998;131:131–138. [PubMed: 9533915]
37. Stchedroff MJ, Kenwright AM, Morris GA, Nilsson M, Harris RK. 2D and 3D DOSY methods for studying mixtures of oligomeric dimethylsiloxanes. *Phys. Chem. Chem. Phys* 2004;6:3221–3227.
38. Nilsson M, Gil AM, Delgadillo I, Morris GA. Improving pulse sequences for 3D DOSY: COSY-IDOSY. *Chem. Commun* 2005:1737–1739.
39. Bradley SA, Krishnamurthy K, Hu H. Simplifying DOSY spectra with selective TOCSY edited preparation. *J. Magn. Reson* 2005;172:110–117. [PubMed: 15589414]
40. Harris RK, Kinnear KA, Morris GA, Stchedroff MJ, Samadi-Maybodi A, Azizi N. Silicon-29 diffusion-ordered NMR spectroscopy (DOSY) as a tool for studying aqueous silicates. *Chem. Commun* 2001:2422–2423.
41. Ribot F, Escax V, Martins JC, Biesemans M, Ghys L, Verbruggen I, Willem R. Probing ionic association on metal oxide clusters by pulsed field gradient NMR spectroscopy: The example of Sn-12-oxo clusters. *Chem.-Eur. J* 2004;10:1747–1751.
42. Eilertsen JL, Hall RW, Simeral LS, Butler LG. Tools and strategies for processing diffusion-ordered 2D NMR spectroscopy (DOSY) of a broad, featureless resonance: an application to methylaluminoxane (MAO). *Anal. Bioanal. Chem* 2004;378:1574–1578. [PubMed: 15214419]
43. Hinton DP, Johnson CS. Diffusion Ordered 2d-Nmr Spectroscopy of Phospholipid-Vesicles - Determination of Vesicle Size Distributions. *J. Phys. Chem* 1993;97:9064–9072.
44. Diaz MD, Berger S. Studies of the complexation of sugars by diffusion-ordered NMR spectroscopy. *Carbohydr. Res* 2000;329:1–5. [PubMed: 11086680]
45. Kapur GS, Findeisen M, Berger S. Analysis of hydrocarbon mixtures by diffusion-ordered NMR spectroscopy. *Fuel* 2000;79:1347–1351.
46. Simpson AJ, Kingery WL, Spraul M, Humpfer E, Dvortsak P, Kerssebaum R. Separation of structural components in soil organic matter by diffusion ordered spectroscopy. *Environ. Sci. Technol* 2001;35:4421–4425. [PubMed: 11757596]
47. Schraml J, Blechta V, Soukupova L, Petrakova E. DOSY of silylated saccharides. *J. Carbohydr. Chem* 2001;20:87–91.
48. Zhang F, Yu XY, Chen Z, Lin S, Liu SX. Characterizations of some N-substituted-salicylhydrazide in mixtures by NMR diffusion ordered spectroscopy. *Chin. J. Struct. Chem* 2003;22:287–292.
49. Kaucher MS, Lam YF, Pieraccini S, Gottarelli G, Davis JT. Using diffusion NMR to characterize guanosine self-association: Insights into structure and mechanism. *Chem.-Eur. J* 2004;11:164–173.
50. Chen A, Wu DH, Johnson CS. Determination of Molecular-Weight Distributions for Polymers by Diffusion-Ordered Nmr. *J. Am. Chem. Soc* 1995;117:7965–7970.

51. Jayawickrama DA, Larive CK, McCord EF, Roe DC. Polymer additives mixture analysis using pulsed-field gradient NMR spectroscopy. *Magn. Reson. Chem* 1998;36:755–760.
52. Jerschow A, Muller N. Diffusion-separated nuclear magnetic resonance spectroscopy of polymer mixtures. *Macromolecules* 1998;31:6573–6578.
53. Liu ML, Toms HC, Hawkes GE, Nicholson JK, Lindon JC. Determination of the relative NH proton lifetimes of the peptide analogue viomycin in aqueous solution by NMR-based diffusion measurement. *J. Biomol. NMR* 1999;13:25–30.
54. Palczewska M, Groves P, Ambrus A, Kaleta A, Kover KE, Batta G, Kuznicki J. Structural and biochemical characterization of neuronal calretinin domain I-II (residues 1-100) - Comparison to homologous calbindin D-28k domain I-II (residues 1-93). *Eur. J. Biochem* 2001;268:6229–6237. [PubMed: 11733019]
55. Johnston MR, Latter MJ. Characterization of porphyrin supramolecular complexes using NMR diffusion spectroscopy. *J. Porphyr. Phthalocyanines* 2002;6:757–762.
56. Viel S, Capitani D, Mannina L, Segre A. Diffusion-ordered NMR spectroscopy: A versatile tool for the molecular weight determination of uncharged polysaccharides. *Biomacromolecules* 2003;4:1843–1847. [PubMed: 14606917]
57. Zhao TJ, Beckham HW, Gibson HW. Quantitative determination of threading in rotaxanated polymers by diffusion-ordered NMR spectroscopy. *Macromolecules* 2003;36:4833–4837.
58. Groves P, Palczewska M, Molero MD, Batta G, Canada FJ. Protein molecular weight standards can compensate systematic errors in diffusion-ordered spectroscopy. *Anal. Biochem* 2004;331:395–397. [PubMed: 15265747]
59. Rajagopalan S, Chow C, Raghunathan V, Fry CG, Cavagnero S. NMR spectroscopic filtration of polypeptides and proteins in complex mixtures. *J. Biomol. NMR* 2004;29:505–516. [PubMed: 15243181]
60. Auguin D, Gostan T, Delsuc MA, Roumestand C. NMR determination of the Pleckstrin Homology domain oligomerisation state of the protein AKT2. *C. R. Chim* 2004;7:265–271.
61. Groves P, Rasmussen MO, Molero MD, Samain E, Canada FJ, Driguez H, Jimenez-Barbero J. Diffusion ordered spectroscopy as a complement to size exclusion chromatography in oligosaccharide analysis. *Glycobiology* 2004;14:451–456. [PubMed: 14693914]
62. Dyrby M, Petersen M, Whittaker AK, Lambert L, Norgaard L, Bro R, Engelsen SB. Analysis of lipoproteins using 2D diffusion-edited NMR spectroscopy and multi-way chemometrics. *Anal. Chim. Acta* 2005;531:209–216.
63. Seenisamy J, Rezler EM, Powell TJ, Tye D, Gokhale V, Joshi CS, Siddiqui-Jain A, Hurley LH. The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc* 2004;126:8702–8709. [PubMed: 15250722]
64. Xu Y, Noguchi Y, Sugiyama H. The new models of the human telomere d[AGGG(TTAGGG)3] in K⁺ solution. *Bioorg. Med. Chem* 2006;14:5584–91. [PubMed: 16682210]

Fig1.

Mass spectrometry analyses of the reference materials in the DOSY study. A. HPLC-MS analysis of Lysozyme ($MW_{\text{theor.}}=14,307$ Da). The main peak from the HPLC run ($RT=14.16\text{--}15.41$ min) is analyzed. In the inset the different ionization states are seen. B. The MALDI-TOF analysis of Ribonuclease A ($MW_{\text{theor.}}=13,700$ Da) in linear (positive) acquisition mode. The singly and doubly charged states are seen. C. HPLC-MS analysis of Xylene Cyanole FF ($MW_{\text{theor.}}=538.6$ g/mol as salt; 515.6 g/mol after dissociation in buffer). The main peak from the HPLC run ($RT=3.20\text{--}4.25$ min) is analyzed in negative mode. The peak at 553.0 is a K^+ adduct. D. The MALDI-TOF analysis of Trypan Blue ($MW_{\text{theor.}}=960.8$ g/mol as salt; 868.8 g/mol after dissociation in buffer) in reflectron (negative) acquisition mode. All MS spectra (in A-D subfigures) confirm the accuracy of the MW references used for our DOSY analyses.

Fig2.

A., 1D ^1H imino NMR spectral region of the bcl-2 promoter DNA G-quadruplex. As seen there are (the characteristic) 12 well-resolved imino protons in both spectra recorded at two concentrations. This shows that the result of the NMR titration experiment (in which a single monomeric species is present, Fig. S1) in the 0.1 mM strand concentration range, is still valid at higher concentrations as well (in samples used for structure determination). B., DOSY analysis of the optimized bcl2 promoter sequence. The spectrum shows one relatively sharp peak in the F1 (diffusion coefficient) dimension representing one dominant molecular species. The peak selected for fitting analysis (Fig.3) is designated by “x” in the F2 (upper) projection.

Conditions: A., 25 °C, 20 mM K-PO₄, 40 mM KCl, pH 7.0 (H₂O), B., 25 °C, 20 mM K-PO₄, 40 mM KCl, pH 7.0 (D₂O).

Fig3.

A., Peak intensity is plotted against the gradient strength for the selected peak in Fig.2 for the bcl2 promoter G-quadruplex. The inset shows refitting of the more reliable data points (the two fittings result in a ~20% difference in the final result). B., The same type of fitting for the two low-MW standards (peaks applied are designated in the D subfigure). C., peak intensity plotted against gradient strength for the low-MW standards for a demonstration of the actual raw data. D., 1D ^1H NMR spectrum and peaks used for the fitting analysis (B and C subfigures) of the two low-MW standards (in mixture). Calculation of MW of DNA according to referencing shows the right MW with ~3% overall error. Conditions: 25 °C, 20 mM K-PO₄, 40 mM KCl, pH 7.0 (D₂O).

Fig4.

A., 2D DOSY analysis of the optimized Tel22-Mut sequence. The 1D ^1H projection shows well-resolved aromatic and imino peaks (candidate for structure determination) and the F1 projection shows one dominant peak representing one major molecular species at 1 mM DNA strand concentration. B., 2D DOSY analysis of the low-MW standards (mixed to a 1 mM concentration, each) in the same solution conditions as for the DNA in subfigure A. The F1 projection shows two major peaks (and some degradation products). The calculation of the MW of the DNA is accurate (see text). Peaks selected for fitting analysis (Fig.5B) are designated by the same labels as in Fig.5B in the F2 (upper) projection. Conditions: 25 °C, 50 mM K-PO₄, 140 mM KCl, pH 7.0 (H₂O).

Fig5.

A., Converted peak intensity is plotted against the gradient strength for the selected peaks in subfigure C for the Tel22-Mut sequence. The inset shows the raw peak intensity-gradient strength data. This plot shows that the two peaks selected (one belonging to the high intensity methyl peaks, one to the low intensity ones in subfigure C) represent the same MW species but two different conformations. B., The same type of fitting and inset as for subfigure A, used for the low-MW standards (peaks applied are designated in Fig.4B). Calculation of MW shows correct results for the DNA (see text) C., 1D ^1H NMR methyl spectrum region of Tel22-Mut and peaks used for the fitting analysis in subfigure A. Conditions: 25 °C, 50 mM K- PO_4 , 140 mM KCl, pH 7.0 (H_2O).

Fig 6.

The 1D ¹H NMR spectrum (A) and the F1 DOSY projection (B) of the Tel22 sequence at 5 mM DNA concentration. The F1 projection shows one dominant peak and a shoulder peak with comparable intensity. The D value for the dominant peak is different from that of the Tel22-Mut, probably because of the different folding (see text) and consequently the molecular shape in solution. C., DOSY analysis of the optimized c-MYC promoter sequence TG (AG₃TG₃T)₂A₂. The spectrum shows one dominant peak with slight broadening towards higher MWs at 3 mM DNA concentration. The shift in the logD values (compared to Tel22/Tel22-Mut) is because of the difference in MW, folding, temperature and buffer (optimized condition; referencing in this condition is not shown). D. and E., F1 projections of the DOSY analysis of the two high MW reference materials for the MW calculation of Tel22-Mut:

Ribonuclease A (MW=13,700 Da, D) and Lysozyme (MW=14,300 Da, E). Conditions for A,B,D,E: 25 °C, 50 mM K-PO₄, 140 mM KCl, pH 7.0, and for C: 20 °C, 25 mM K-PO₄, 70 mM KCl, pH 7.0.