DOI: 10.1002/chem.200901248

Rapid Stoichiometric Analysis of G-Quadruplexes in Solution

Patrick Groves*[b] and Mateus Webba da Silva*[a]

Guanine-rich sequences of nucleic acids may fold into secondary structural folds called quadruplex architectures. These architectures are a rapidly growing theme of interest with promising repercussions in our understanding of biology and practical applications in medical fields, materials science, and biotechnology. Currently, there is thus interest in determining both topology and their atomic detail 3D structures. The establishment of solution conditions suitable for structural studies of G-quadruplex architectures requires the determination of the level of oligomerization (stoichiometry) of DNA strands. Various analytical techniques are currently applied for the routine assessment of the stoichiometry that generally include conditions not representative of the environment in which the structural studies are performed. The approaches that include the appropriate conditions are not suitable for routine use (reviewed in reference [1]). The method most frequently employed is based on serial dilutions from the folded quadruplex concentration to low concentrations up to a point in which the concentration of the unfolded state becomes significant.[2] There is at least one limiting disadvantage for this method. It may take weeks for the equilibrium to be reached at low concentrations; or worse, for some multimeric sequences, the unfolded state may not be achieved for concentrations at which NMR measurements are possible. This limitation may be overcome by unfolding the quadruplex at high equilibrium concentrations by raising the temperature to achieve a state in which both structured and unfolded forms co-exist.[3,4] However, this method has at least two limitations: 1) the melting temperature of the quadruplex may be above the boiling point of the solvent, and 2) that NMR signals due to the unfolded state have to be clearly identified and should not be confused with any other folding state. In principle, information regarding the oligomeric state, as well as purity, of biomolecules can also be routinely obtained in a few minutes from the diffusion coefficient of 1 mg mL^{-1} concentration samples by utilizing diffusion-ordered NMR spectroscopy (DOSY).[5,6] In a 2D DOSY experiment, a series of spin-echo spectra is measured with different pulsed field gradient strengths, and the signal decays are analyzed to extract a set of diffusion coefficients that make up the diffusion domain of a spectrum. The method has been previously used to determine the molecular weight of unimolecular quadruplexes.[7] Herein we describe a protocol for the rapid characterization of the stoichiometry of nucleic acid quadruplex folds.

The diffusion coefficient of a molecule is shape dependent, as shown by Wilkins et al. for peptides and proteins under native and unfolded conditions.[8] We constructed calibration curves using globular proteins[9] and polyethylene glycol (PEG) molecular-weight standards.[10] The difference in shape between proteins and PEGs results in two different correlation curves (Figure 1A). In principle, since quadruplexes can neither be more disordered than PEGs [below the PEG curve, $\log D = -8.119 - 0.490 \log \mathrm{MW}$ ($r^2 = 0.996$, 7 points—one point is not shown)], nor more compact than proteins [above the protein curve, $\log D = -8.305 - 0.381 \log \mathrm{MW}$ ($r^2 = 0.996$, 5 points)], their folded state should fall in the region between the two curves; the Goldilocks zone. Moreover, the diffusion coefficient of the folded globular protein is a better indicator of the hydrodynamic radius of the folded quadruplex, and thus, the molecular weight associated with a particular diffusion coefficient falling closest to the protein line should reflect the stoichiometry of the quadruplex fold.

We have thus measured the diffusion coefficient from the aromatic proton region of DNA guanine-rich sequences with the potential to fold into quadruplexes, both in the presence and virtual absence of cations. The presence of cations is necessary for DNA quadruplex formation; whilst their absence is not achievable in solubilized nonmodified nucleic acids due to the anionic nature of the nucleic acid

[a] Dr. M. Webba da Silva
School of Biomedical Sciences
University of Ulster, Cromore Road
Coleraine, BT51 1SA (UK)
Fax: (+44) 28-7032 4375
E-mail: mm.webba-da-silva@ulster.ac.uk

[b] Dr. P. Groves
Department of Biological Chemistry
Instituto de Tecnologia Química e Biológica
Address 2 Av. da República (EAN)
2781-901 Oeiras (Portugal)

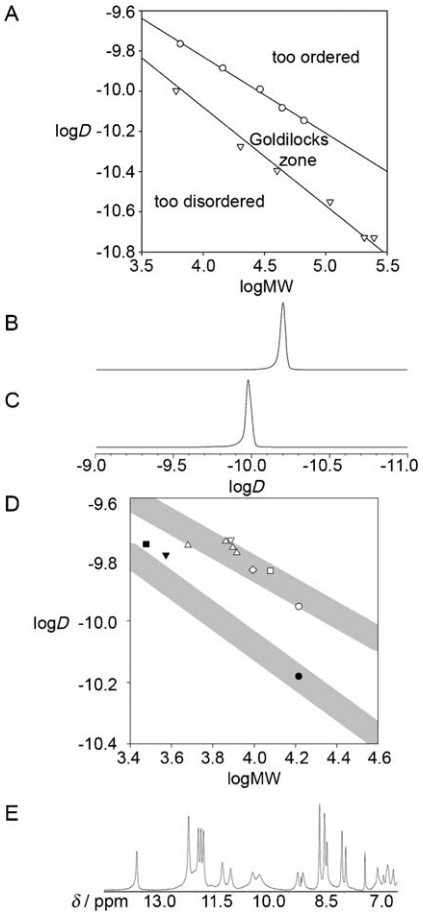

Figure 1. Diffusion analysis of quadruplexes. A) Plot of $\log D$ versus logMW for a series of proteins (circles) and polyethylene glycols (triangles). The Goldilocks zone of the graph for data interpretation is indicated. B) and C) Averaged diffusion data from the aromatic region (5.7–8.7 ppm) of the 2D DOSY spectrum are shown for the telomeric sequence (B) in ^2H_2O and (C) in the presence of 120 mM potassium. D) Diffusion data for various DNA sequences known to form monomeric, dimeric, or tetrameric quadruplexes under known salt conditions (open symbols) and in the absence of salt (filled symbols). Data were obtained for the telomeric sequence d[$AG_3(TTAG_3)_7TTA$] (circles) that forms a monomeric quadruplex; d($GCGGTG_3T$) (squares) that forms a tetrameric quadruplex; d($G_4T_4G_4$) (inverted triangles) that forms a dimeric quadruplex; TBA–d($G_2T_2G_2TGTG_2T_2G_2$) and the three telomeric sequences d[$TAG_3(TTAG_3)_3$], d[$TAG_3(TTAG_3)_3TT$], and d[$A_3G_3(TTAG_3)_3AA$] that form monomeric quadruplexes (triangles) and TG_5T that forms a tetrameric quadruplex (diamond). The shaded lines represent $\pm 0.04 \log D$ units from the protein calibration line. E) Exchangeable proton region showing spread of chemical shifts denoting distinct chemical environments experienced by imino (10.5 to 13.5 ppm) and amino (8.5 to 10.5 ppm) protons in d[$(GCGGTG_3T)_4$].

backbone. We measured the diffusion coefficient of seven known quadruplex folds of different stoichiometry: four monomers, a dimer, and two tetramers; and established the stoichiometry for the folding of the human telomeric sequence d[$AG_3(TTAG_3)_7TTA$].

The human telomeric sequence d[$AG_3(TTAG_3)_7TTA$] comprises eight guanine segments. The single peak in the average diffusion profiles for the telomeric sequence, for both the disordered state, and in 120 mM potassium, suggests a single species (Figure 1B and C), respectively. In virtual absence of cations, the telomeric sequence is expected to be in a disordered state and the diffusion coefficient falls close to the PEG calibration line when plotted against the log MW of single strand (Figure 1D). In 120 mM potassium the diffusion coefficient of the telomeric sequence changes (Figure 1B and C). Since formation of quadruplexes depends on the availability of cations this change in diffusion coefficient is in line with the process of self-assembly. We observe that the folded monomeric value falls close to the globular protein calibration line and within the Goldilocks zone; within $\pm 0.04 \log D$ units from the protein calibration line. This is consistent with the formation of a unimolecular structure, potentially consisting of tandem quadruplex domains (currently under investigation). The tandem structure would be expected to result in a small deviation from the protein calibration curve towards the PEG line, since it is expected to be rather more prolate spheroid than globular proteins, which are expected to be spherical.

Most solution NMR structural studies are currently performed on much shorter DNA sequences ranging from about 9 to 35 nucleotides. We have thus investigated if the level of multimerization for known quadruplexes could be predicted with this method. In Figure 1D, the diffusion coefficients for the known quadruplexes formed by the monomeric thrombin binding aptamer d[$G_2T_2G_2TGTG_2T_2G_2$], the dimeric d[$(G_4T_4G_4)_2$], the tetrameric d[$(TG_5T)_4$] and the monomeric telomeric sequences d[$TAG_3(TTAG_3)_3$], d-[$TAG_3(TTAG_3)_3TT$], and d[$A_3G_3(TTAG_3)_3AA$], fall on, the protein line in accordance with their known level of multimerization.[11] In the virtual absence of cations (in water) the diffusion coefficients of the short DNA sequence that forms the dimeric d[$(G_4T_4G_4)_2$] deviates from the PEG line. This is either due to non-specific interactions, or because the hydrodynamic radius of a small nucleic acid oligonucleotide should be more spherical.

The short sequence d($GCGGTG_3T$) as well as its analogues d(GCG_2TG_2AT), d(GCG_2AG_2AT), and d(GCG_2TTG_2AT), all fold into homomultimeric bimolecular dimers in sodium buffer solution.[12] The molecular weight of the folded 36 nucleotides quadruplex is thus comparable to that of the tetrameric d[$(TG_5T)_4$]. Indeed, the diffusion coefficient for both tetrameric folds is also similar and fall in the protein line- Figure 1D. In the virtual absence of cations (in water) the diffusion coefficient of the small (9 nt) sequence deviates from the PEG line; Figure 1C. This is either due to non-specific aggregation or, that the hydrodynamic radius of a small nucleic acid sequence should be more spherical. This effect has been previously observed with small oligosaccharides.[10] In contrast to the telomeric sequence, the quadruplex formed by the d($GCGGTG_3T$) sequence in 120 mM sodium diffuses more slowly than its "cation-free" form. In principle this is interpreted as the result of a substantial increase in the hydrodynamic radius as compared to the increase resultant from a monomeric fold. Indeed, by plotting the diffusion coefficient calculated against the expected monomeric, dimeric, and tetrameric

masses the tetrameric molecular weight is closest to the protein line. Inspection of the exchangeable proton resonances of the sample under the same experimental conditions demonstrates that the sample is folded; Figure 1E. Thus, the d-($GCGGTG_3T$) sequence folds into a tetrameric architecture in sodium buffer.

The short sequences that form the dimeric d[$(G_4T_4G_4)_2$], and the tetrameric d[$(TG_5T)_4$] demonstrate that some caution is warranted when interpreting the topology of the unfolded state of small oligonucleotide sequences. Moreover, proton exchange of basic imino and amino groups with solvent, the presence of low molecular-weight buffers, or the presence of small molecule stabilizers can lead to an averaged diffusion coefficient that is weighted towards the fast diffusion coefficient of the smaller molecules, which may fall below the PEG correlation curve. For this reason, $\log D$ data are best obtained from non-exchanged aromatic protons in regions that do not contain amide signals. Markedly different environmental conditions that lead to high solution viscosities should also be avoided.

We have thus shown that by comparing the diffusion coefficient of folded quadruplexes to the calibration curves of diffusion coefficients versus molecular weight of globular proteins, representing folded quadruplexes, and polyethylene glycol (PEG), which represents single-strand oligonucleotides, we are able to rapidly establish the stoichiometry of DNA sequences in solution. We have also shown that the hydrodynamic radius of folded quadruplexes is comparable to that of folded globular proteins. With correct data interpretation, this method can be a useful tool for the rapid stoichiometric characterization of quadruplexes in solution conditions when optimizing samples for structural studies.

Experimental Section

The sequences d($GCGGTG_3T$) and d($AG_3(TTAG_3)_7TTA$) were purchased from Invitrogen HPLC purified. The eluent containing the pure oligonucleotide was dialyzed several times alternatively against water and salt buffer; 115 mM KCl, 5 mM potassium phosphate (pH 6.9) for d-($AG_3(TTAG_3)_7TTA$), d($G_2T_2G_2TGTG_2T_2G_2$), and d[$(TG_5T)_4$]; 115 mM NaCl, 5 mM potassium phosphate (pH 6.9) for d($GCGGTG_3T$); and 100 mM NaCl, 20 mM sodium phosphate (pH 6.8) for d[$(G_4T_4G_4)_2$]. These samples were lyophilized and re-suspended in 99.996 % ^2H_2O. We constructed calibration curves using globular proteins (purchased from Sigma)[9] and polyethylene glycol (PEG) molecular-weight standards (purchased from Fluka).[10] DOSY data was collected with the stebpgp1s19 pulse sequence on (2D Stimulated Echo experiment using bipolar gradients and WATERGATE) Bruker Avance 500 MHz instruments. All experiments were run at 20 °C was run in 32 linear steps over a 5–95 % range of the gradient (maximum = 50 G cm^{-1}).

Acknowledgements

The authors thank Prof. Jesús Jiménez-Barbero for access to NMR facilities at the Centro de Investigaciones de Biologicas in Madrid (Spain); the Portuguese NMR spectrometers are part of The National NMR Network (REDE/1517/RMN/2005), supported by "Programa Operacional Ciência e Inovação (POCTI) 2010" and Fundação para a Ciência e a Technologia. This work was also supported by European Community (EU-NMR, Contract # RII3–026145); and COST Action MP0802.

Keywords: analytical methods · diffusion coefficient · NMR spectroscopy · nucleic acids · structure elucidation

[1] M. Webba da Silva, *Methods* **2007**, *43*, 264–277.
[2] J. L. Leroy, K. Gehring, A. Kettani, M. Gueron, *Biochemistry* **1993**, *32*, 6019–6031.
[3] "Investigation of Unusual DNA Motifs": A. T. Phan, M. Gueron, J. L. Leroy in *Nuclear Magnetic Resonance of Biological Macromolecules Part A, Vol. 338*, **2001**, Academic Press, New York, pp. 341–371.
[4] A. T. Phan, D. J. Patel, *J. Am. Chem. Soc.* **2003**, *125*, 15021–15027.
[5] C. S. Johnson, Jr., *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 203–256.
[6] J. C. Cobas, P. Groves, M. Martin-Pastor, A. De Capua, *Curr. Anal. Chem.* **2005**, *1*, 289–305.
[7] A. Ambrus, D. Yang, *Anal. Biochem.* **2007**, *367*, 56–67.
[8] D. K. Wilkins, S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, L. J. Smith, *Biochemistry* **1999**, *38*, 16424–16431.
[9] P. Groves, M. Palczewska, M. D. Molero, G. Batta, F. J. Canada, *Anal. Biochem.* **2004**, *331*, 395–397.
[10] P. Groves, M. O. Rasmussen, M. D. Molero, E. Samain, F. J. Canada, H. Driguez, J. Jimenez-Barbero, *Glycobiology* **2004**, *14*, 451–456.
[11] a) P. Schultze, N. V. Hud, F. W. Smith, J. Feigon, *Nucleic Acids Res.* **1999**, *27*, 3018–3028; b) A. T. Phan, V. Kuryavyi, K. N. Luu, D. J. Patel, *Nucleic Acids Res.* **2007**, *35*, 6517–6525; c) A. Ambrus, D. Chen, J. X. Dai, T. Bialis, R. A. Jones, D. Z. Yang, *Nucleic Acids Res.* **2006**, *34*, 2723–2735; d) C. Bardin, J. L. Leroy, *Nucleic Acids Res.* **2007**, *36*, 477–488.
[12] a) M. Webba da Silva, *Biochemistry* **2003**, *42*, 14356–14365; b) M. Webba da Silva, *Biochemistry* **2005**, *44*, 3754–3764; c) J. L. Mergny, A. De Cian, S. Amrane, M. Webba da Silva, *Nucleic Acids Res.* **2006**, *34*, 2386–2397.

Received: May 11, 2009
Published online: May 6, 2010

