Infrared Spectroscopy Reveals the Preferred Motif Size and Local Disorder in Parallel Stranded DNA G-quadruplexes

David A. Price,[a] Zachary J. Kartje,[a][c] Joanna A. Hughes,[a] Tayler D. Hill,[a] Taylor M. Loth,[a] Jonathan K. Watts,[c] Keith T. Gagnon,[a][b] and Sean D. Moran[a]*

[a] D.A. Price, Dr. Z.J. Kartje, Dr. J.A. Hughes, T.D. Hill, T.M. Loth, Prof. Dr. K.T. Gagnon, Prof. Dr. S.D. Moran*
Department of Chemistry and Biochemistry
Southern Illinois University Carbondale
1245 Lincoln Drive, Carbondale, IL, USA, 62901
E-mail: smoran@chem.siu.edu
[b] Prof. Dr. K.T. Gagnon
Department of Biochemistry and Molecular Biology
Southern Illinois University School of Medicine
1245 Lincoln Drive, Carbondale, IL, USA, 62901
[c] Dr. Z.J. Kartje, Prof. Dr. J.K. Watts
RNA Therapeutics Institute and Department of Biochemistry and Molecular Pharmacology
University of Massachusetts Medical School
Worcester, MA, USA, 01655

Supporting information for this article is given via a link at the end of the document.

Abstract: Infrared spectroscopy detects the formation of G-quadruplexes in guanine-rich nucleic acid sequences through shifts in the guanine C=O stretch mode. Here, we use ultrafast two-dimensional infrared (2D IR) spectroscopy and isotope substitution to show that these shifts arise from vibrational delocalization among stacked G-quartets. This provides a direct measure of the sizes of locally ordered motifs in heterogeneous samples with substantial disordered regions. We find that parallel-stranded, potassium-bound DNA G-quadruplexes are limited to five consecutive G-quartets and 3 – 4 consecutive layers are preferred for longer polyguanine tracts. The resulting potassium-dependent G-quadruplex assembly landscape reflects the polyguanine tract lengths found in genomes, the ionic conditions prevalent in healthy mammalian cells, and the onset of structural disorder in disease states. Our study describes spectral markers that can be used to probe other G-quadruplex structures and provides insight into the fundamental limits of their formation in biological and artificial systems.

Introduction

Guanine-rich nucleic acid sequences have a propensity to form non-canonical four-stranded motifs known as G-quadruplexes.[1] These structures have been hypothesized to regulate many cellular functions by acting as targets for protein binding,[1b, 2] controlling telomere maintenance and function,[2-3] selecting DNA replication origins,[4] modulating the processivities of polymerases in replication and transcription,[5] and influencing ribosome progression during translation.[1d, 5c] Their formation in guanine-rich tandem repeats and repeat expansions has also been implicated in multiple human diseases. For example, G-quadruplex stability in the *c-Myc* promoter controls tumor progression[5b] and G-quadruplexes in the *HOX11* breakpoint region contribute to T-cell leukemia by promoting chromosomal translocations.[6] Additionally, $(CGG)_n$ trinucleotide repeats in *FMR1* form fragile sites in chromosomes[5a] and the *C9orf72* $(GGGGCC)_n$ hexanucleotide repeat expansion produces aggregating mRNA transcripts in amyotrophic lateral sclerosis and frontotemporal dementia.[7] The formation of G-quadruplexes *in vivo* continues to be substantiated through genomic sequence analysis[1a, 4-5, 8] and their occurrence in cells has been demonstrated using fluorescently-labelled antibodies and small molecule probes.[1a, 5c]

The minimal structural unit of a G-quadruplex is a planar array of four guanine bases, known as a G-quartet.[1b-d, 9] To form a G-quadruplex, multiple G-quartets stack upon each other and the overall structure is stabilized by hydrogen-bonding between the Watson-Crick and Hoogsteen faces of guanine bases within G-quartets, by π-π stacking of bases in adjacent G-quartets, and by the coordination of alkali metal cations to guanine C6 carbonyl oxygens.[1d, 9-10] G-quadruplexes can adopt diverse structures, and high-resolution NMR and X-ray crystallography studies have revealed examples with parallel or antiparallel strand orientations and either *syn*- or *anti*- base conformations. These structural variations, in turn, depend on multiple factors including poly-guanine tract lengths,[11] backbone sugar substitution,[1d, 9-10] and the identities of bound metal cations.[10a] This information has proven useful for understanding molecular recognition of G-quadruplexes in biology, the development of molecular probes and therapeutic agents that target them,[1b, 5c, 8, 12] and emerging applications of G-quadruplexes and G-quartets in sensing and nanotechnology.[13] However, it remains unknown whether the conclusions from model studies can be applied generally to longer sequences that form heterogeneous structural ensembles.

Vibrational methods such as Fourier transform infrared (FTIR) spectroscopy, Raman spectroscopy, and multidimensional spectroscopies are sensitive probes of molecular structures because the frequencies and lineshapes of peaks depend on local electrostatic environments and coupling between vibrational modes. These factors produce characteristic spectral shifts associated with the formation of specific secondary structures in biomolecules, including both proteins and nucleic acids.[14] Although vibrational spectra provide lower

resolution information than NMR or X-ray crystallography, they are largely insensitive to sample morphology and are thus useful probes of local structures in complex samples.[15] In recent studies of G-quadruplexes, FTIR,[16] vibrational circular dichroism (VCD),[17] and surface-enhanced Raman spectroscopy (SERS)[18] have used shifts in the guanine C6=O6 carbonyl stretch, near 1700 cm^{-1}, to monitor their formation. These shifts have been alternately attributed to C=O bond lengthening upon metal coordination and Hoogsteen hydrogen-bonding, along with generalized base "stacking" effects.[14b, 16-18] These interpretations invoke local interactions of individual bases but ignore coupling effects that delocalize vibrations among multiple bases within ordered arrays. Vibrational coupling between stacked bases is a major factor in frequency and lineshape shifts in complementary duplex DNAs,[19] but to date its contribution to G-quadruplex spectra has remained unexplored. This gap in knowledge has resulted in incomplete and conflicting assignments of G-quadruplex spectra, which has hindered the application of vibrational spectroscopy to structural studies of G-quadruplex forming sequences.

Ultrafast two-dimensional infrared (2D IR) spectroscopy is a powerful technique for disentangling the physical phenomena that contribute to the vibrational spectrum of a molecule or complex.[20] Compared to linear techniques, 2D IR has numerous advantages including the ability to observe coupling directly through cross-peaks, to measure local dynamics through spectral diffusion, and to better resolve peaks in congested spectra.[15, 20-21] Recent studies of known duplex DNA structures have characterized sequence-dependent coupling between stacked bases,[19] coupling between bases and backbone phosphate groups,[22] and hydrogen-bonding dynamics within base pairs.[23] These studies provide a toolkit that can, in principle, be used to determine unknown motif structures from their local and collective vibrations. Precedent for this approach is provided by 2D IR studies of amyloid aggregates of peptides and proteins.[24] Like G-quadruplexes, amyloid β-sheets contain extended arrays of hydrogen-bonded residues. The analysis of coupling within these arrays,[15] aided by isotope editing[24a, 24b, 24e] and computational simulations,[24b, 24c] has produced models of amyloid structures with resolutions ranging from single-residues[24c] to entire domains of proteins,[24a] and 2D IR spectroscopy has even been applied to detect these structures in diseased human tissues.[25]

In this article, we examine the assembly of parallel-stranded DNA G-quadruplexes using the guanine C6=O6 carbonyl stretch as a probe of local structure. We measure trends in vibrational frequencies and lineshapes that depend on poly-guanine tract lengths, metal ion identities and concentrations, and oligonucleotide oligomerization states. Using 2D IR spectroscopy, we show that spectral shifts arise primarily from coupling of guanine base modes within ordered G-quadruplex motifs, allowing assignment of overlapping features to bases that are included or excluded from ordered arrays. Through isotope substitution we show that vibrations are delocalized among stacked G-quartets, so shifts in infrared spectra actually reflect the sizes of ordered G-quadruplexes. Using this information, we analyze the FTIR spectra of potassium-bound structures as potassium concentrations and guanine tract lengths are varied. From this data, we construct an empirical landscape for parallel-stranded G-quadruplex assembly that closely reflects the known and predicted behaviors of natural guanine-rich sequences. Our study suggests that fundamental limits of DNA G-quadruplex motif size exist and provides methods and insights for testing this hypothesis in increasingly complex natural and artificial sequences.

Results

FTIR features depend on metals and poly-guanine tract lengths

Thymine-capped DNA oligonucleotides containing tracts of four consecutive guanines assemble into parallel-stranded tetramers in the presence of Na^+ and K^+.[10b, 26] In representative NMR and X-ray crystal structures (Figure 1a), the resulting G-quadruplexes contain four stacked G-quartets separated by ~ 3 Å, with metal ions bound between layers. The all-*anti* guanine bases in these structures adopt nearly identical mean conformations, with an RMSD of 0.61 Å among common atoms, and similar helical twists resulting in ~30° angular offsets between adjacent layers (Figure S1). To test whether these similarities are maintained as the length of the poly-guanine tract is varied, we examined a series of alkali metal complexes formed by oligonucleotides (TG_nT) containing three to ten consecutive guanine bases (Table S1). Assuming idealized structures, each additional guanine in the sequence would result in one additional G-quartet layer in the stack (Figure 1b). However, because polyguanine tracts can fold into G-quadruplexes with diverse conformations,[1c, 11, 18] the formation of ideal tetramers is not guaranteed.

Between 1630 – 1710 cm^{-1}, the mid-IR spectra of these complexes contain multiple features associated with guanine and thymine base vibrations. Tokmakoff and co-workers termed these the G1 and T1-3 modes,[27] but here we refer to them as guanine I and thymine I-III to avoid confusion with common G-quadruplex nomenclature and the numbering of our oligonucleotide samples. The guanine I mode, which is dominated by the C6=O6 carbonyl stretching motion, is especially sensitive to G-quadruplex formation due to C=O bond lengthening upon metal coordination,[16b, 17] Hoogsteen hydrogen bonding,[16a, 18] and the putative vibrational coupling effects that are the subject of this investigation.[19] To quantify the combined effects of these interactions, we collected FTIR spectra of TG_nT structures formed in the presence varying concentrations of Li^+, Na^+, and K^+ (Figure 1c). Briefly, samples were prepared by heating tris-buffered D_2O solutions of TG_nT and metal chloride salts to 90 ℃, followed by slow cooling to room temperature to assemble the complexes. For samples prepared with Li^+, which is not expected to stabilize G-quadruplexes,[9-10] a single Gaussian-shaped peak appeared between 1665 – 1668 cm^{-1} for all guanine tract lengths (n). This frequency is similar to that of free guanosine monophosphate (GMP) (Figures S3, S4), and the invariant linewidth of this peak indicates that the guanine I mode is the primary absorber in this region while contributions from thymine-based modes are relatively minor. In contrast, the FTIR spectra of samples prepared with Na^+ and K^+ showed blue shifts in the guanine I peak maximum (ω_{max}) ranging from 3 – 15 cm^{-1}, depending on sequence and metal ion conditions. Broadly, these frequency shifts are consistent with G-quadruplex formation,[16a, 17, 19a] but the large range suggests substantial differences in structures or stabilities across the series.

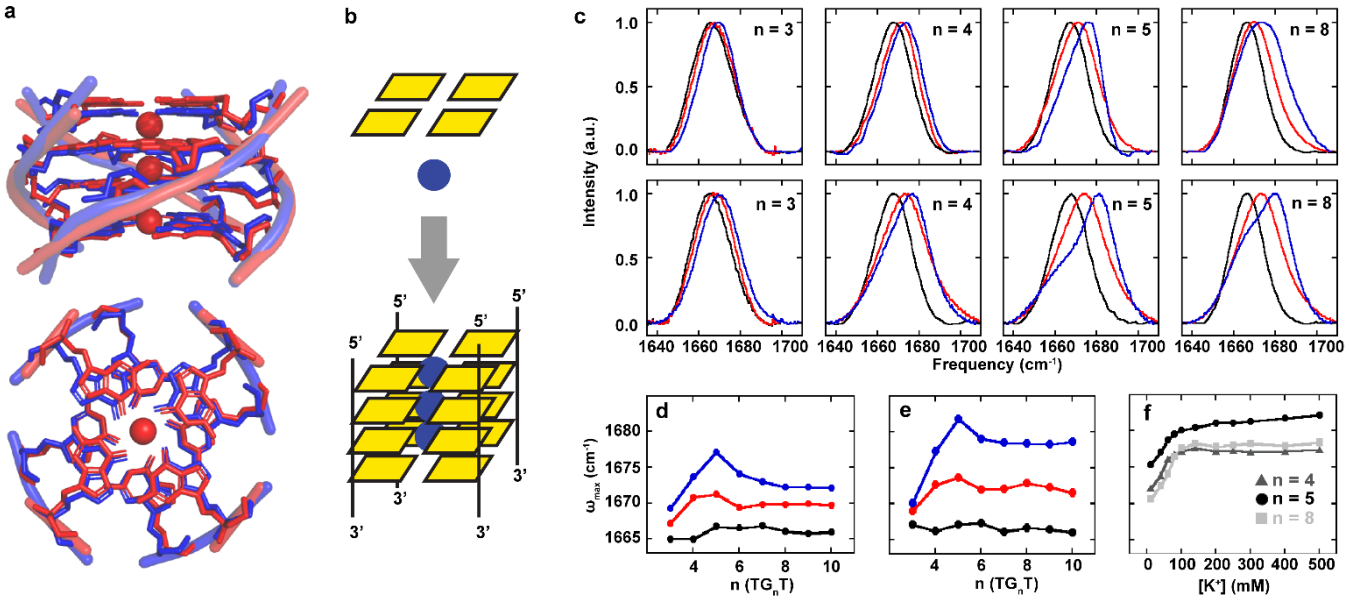

Figure 1. G-quadruplex structure and FTIR analysis of TG_nT. (a) Structural overlay of G-quadruplexes (n = 4) bound to K^+ (blue) (PDB ID: 139D, Ref 26) and Na^+ (red) (PDB ID: 244D, Ref 10b). (b) Schematic model of metal-bound TG_4T showing idealized axial extension with an increase in n. (c) Guanine I region FTIR spectra for selected TG_nT formed with Li^+ (black), Na^+ (red) and K^+ (blue) at 40 mM (top) and 500 mM (bottom) metal salts. All spectra are normalized to the maximum intensity to clarify the shifts and representative non-normalized spectra are presented in Figure S4. (d,e) Guanine I ω_{max} for Li^+-TG_nT, Na^+-TG_nT and K^+-TG_nT at (d) 40 mM and (e) 500 mM salts. (f) ω_{max} of guanine I FTIR spectra of K^+-TG_4T (triangles), K^+-TG_5T (circles), and K^+-TG_8T (squares) at varying concentrations of K^+. ω_{max} values are averages of three independent trials, and error bars are omitted for clarity. All FTIR spectra and tables of averages, with standard deviations, are available in Figure S2 and Table S2.

At low (40 mM) metal concentrations, the FTIR spectra of short (n = 3, 4) Na^+- and K^+- TG_nT complexes are nearly identical (Figure 1c), reflecting the structural similarity expected from NMR and crystallographic data (Figure 1a).[10b, 26] As the guanine tract length is increased, the ω_{max} values of Na^+-TG_nT and K^+-TG_nT spectra diverge, with K^+-TG_nT appearing at higher frequencies in all samples and noticeable broadening of the spectra with large n. These differences are increased at high (500 mM) metal ion concentrations, where a clear low-frequency shoulder also appears in the spectra of the potassium complexes. In Figure 1d,e, the ω_{max} divergence between Na^+-TG_nT and K^+-TG_nT is apparent for all tract lengths, but the individual frequency trends are not monotonic. Here, ω_{max} increases from n = 3 to n = 5 and then reverts to intermediate values at large n for complexes formed with both metal ions. This reversion is indicative of a structural transition that occurs beyond a length threshold of n = 5, since ideal G-quartet stacks would most likely approach a guanine I frequency maximum asymptotically as the number of layers is increased.[19, 28] Importantly, the same length threshold is observed irrespective of metal ion identity and concentration. At n = 5, the shift in K^+-TG_5T spectra (up to +15 cm^{-1} vs. GMP) approaches that of five consecutive, vibrationally coupled guanine bases in duplex DNA.[19] However, at the same Na^+ concentration, ω_{max} only shifts by ~7 cm^{-1}, indicating that metal ion identity has a strong influence on the frequencies of the observed vibrational modes.

Using native polyacrylamide gel electrophoresis (PAGE) and size-exclusion chromatography (SEC) (Figure S5), we found that the ω_{max} reversion in Na^+- and K^+-TG_nT coincides with the onset of oligomeric polymorphism when n > 5. Mixtures of tetramers and higher-order oligomers occur for these sequences, consistent with the broadening of the FTIR spectra at large n. Mixtures of oligomers were also observed for samples (n > 5) prepared with Li^+, indicating that long poly-guanine tracts have an intrinsic propensity to aggregate independent of metal binding. When TG_nT oligos were disaggregated with LiOH prior to assembly,[29] the observed ω_{max} trends in FTIR spectra were indistinguishable (within error) from those obtained by heating alone (Figure S6). Thus, the frequency trends reported in Figure 1c reflect optimal local structures for each complex and not trapped states resulting from incomplete disaggregation. Using UV circular dichroism (UV-CD) spectroscopy, we characterized the strand topologies of representative metal-TG_nT complexes (Figure S7). In all samples, an absence of substantial ellipticity near 295 nm shows that parallel-stand orientations are maintained across the series.[1b, 11] Mean molar ellipticities per guanine residue (260 nm) reflect the known metal-dependent stability series (K^+ > Na^+ > Li^+)[9-10] and a decrease in G-quadruplex content above the length threshold. Notably, the UV-CD spectra of Li^+-TG_5T samples still display some parallel-stranded G-quadruplex character, even though the corresponding FTIR spectrum shows no shift in ω_{max}. Thus, the frequency of the guanine I peak in infrared spectra is a more sensitive probe of metal-dependent G-quadruplex assembly than commonly used circular dichroism techniques.

Finally, we examined the $[K^+]$ dependence of guanine I shifts in TG_4T, TG_5T and TG_8T. The resulting complexes represent the structurally-defined tetramer, tetramers at the FTIR-defined guanine tract length threshold, and a mixture of oligomers formed beyond the threshold, respectively. For all three sequences, ω_{max} increased sharply up to ~100 mM K^+ and then remained approximately constant up to 500 mM K^+ (Figure 1f). These trends show that K^+ binding reaches saturation near the typical potassium concentration (130-140 mM) found within healthy mammalian cells,[5b, 30] and that the same saturation limit occurs even when mixtures of different structures are formed. In

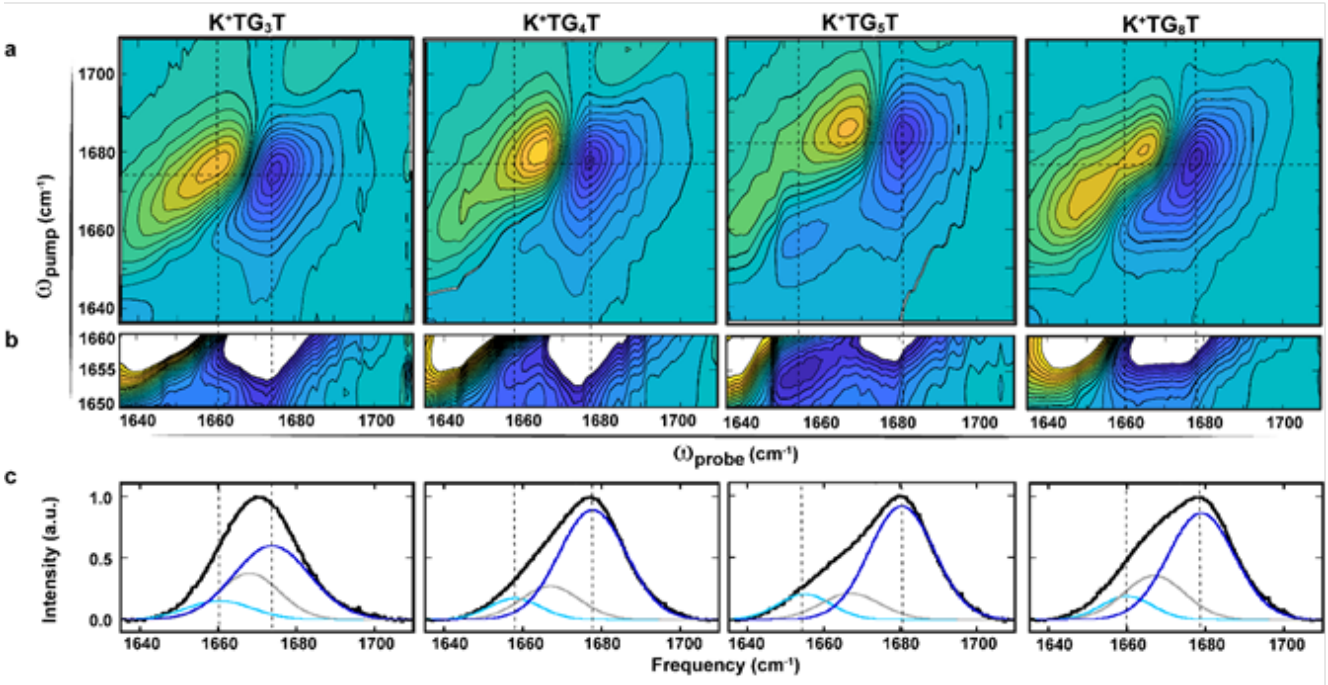

Figure 2. 2D IR spectra and FTIR analysis for selected K^+-TG_nT complexes formed at 140 mM K^+. (a) 2D IR spectra of the guanine I region, normalized to the largest diagonal v(0-1) intensity in each spectrum. Horizontal dashed lines show shifts in ω_{pump} values of high frequency guanine I diagonal peaks as n is varied. (b) Detail of the low-frequency region, scaled to 20% of (a), showing the appearance of the low frequency guanine I peaks. (c) Normalized FTIR spectra of K^+-TG_nT complexes fit to three Gaussian components: high-frequency guanine I (dark blue), low-frequency guanine I (light blue) and disordered bases (grey). Vertical dashed lines (all panels) show the correspondence of Gaussian maxima in (c) to 2D IR features in (a) and (b).

Na^+-TG_nT, saturation also occurs between 40 mM and 140 mM Na^+ (Figure S8), but fully sodium-bound complexes are unlikely under cellular conditions due to the predominance of K^+ ions in the cytoplasm.

Guanine I peak splitting in K^+-TG_nT

For a detailed characterization of the guanine I shifts in metal-TG_nT complexes, we turned to 2D IR spectroscopy. In a 2D IR experiment, a series of femtosecond mid-IR pulses interrogates correlations between vibrational modes; a pair of pump pulses excites vibrations and a probe pulse measures the system response following a variable waiting time, T.[20] In a 2D IR spectrum, peak pairs along the diagonal ($\omega_{pump} = \omega_{probe}$) reflect the v(0-1) and v(1-2) transitions of normal modes. These peaks have opposite signs and their separations along ω_{probe}, or diagonal anharmonic shifts ($\Delta\omega_a$), report the anharmonicities of the normal modes. Diagonal slices through 2D IR spectra reflect the same v(0-1) transitions observed by FTIR spectroscopy, but because 2D IR signals depend on the fourth powers of transition dipole moments ($|\mu|^4$), peaks in congested spectra are better resolved and strong transitions are enhanced. Additionally, cross-peaks appearing off-diagonal reveal coupling between different normal modes. Finally, as T is scanned, the evolution of the intensities and lineshapes of the diagonal peak pairs reports vibrational relaxation and spectral diffusion.[21] Together, these advantages can provide clear insight into molecular structure as well as information that can be used to interpret the corresponding FTIR spectra.[15]

In this study, we focused on K^+-TG_nT complexes, which show the largest spectral changes upon assembly (Figure 1) and are most likely to form under physiological conditions.[30] We collected 2D IR spectra of representative K^+-saturated complexes (n = 3, 4, 5 and 8; [K^+] = 140 mM) at pump-probe waiting times of T = 0.0 ps, where spectral diffusion is minimized (Figure 2a). In all cases, we observe an intense diagonal peak pair between ω_{pump} = 1670 cm^{-1} and 1681 cm^{-1} that follows the same trend as the FTIR ω_{max}. Close inspection of the 2D IR spectra shows substantial guanine-tract dependent variations in diagonal features below ~1670 cm^{-1}. Specifically, a second peak pair grows in intensity and shifts to lower frequency (ω_{pump} = 1655 cm^{-1}) as n approaches 5 and then reverts when n = 8. This feature is also observed at 40 mM and 500 mM K^+, but is somewhat attenuated below the K^+ saturation limit (Figure S9, S10), so it also arises from guanine I vibrations that are modulated by G-quadruplex assembly. For the FTIR spectra of the same complexes (Figure 2c), Gaussian deconvolution of the two resolved features observed in the 2D IR spectra requires a third component at ~1667 cm^{-1} to produce consistent fits (Figure S11). This feature is not clearly resolved in the congested 2D IR spectra, but its appearance at the same frequency as free GMP and Li^+/TG_nT suggests that weak intensity in that region arises from guanine or thymine bases that are excluded from ordered metal-bound arrays.

To confirm this, we monitored the T-dependent decay of the 2D IR signal in K^+-TG_5T (Figure 3). In Figure 3a, spectra collected at T = 0.0 ps, 0.6 ps, and 1.5 ps show clear spectral diffusion in the observed peaks and rapid loss of signal at intermediate frequencies (~1667 cm^{-1}); only the intense 1681 cm^{-1} peak pair is resolved at T = 1.5 ps. The diagonal v(0-1) intensity was measured as a function of T at the center frequencies extracted from the Gaussian fit of the corresponding FTIR spectrum (Figure 3b). All relaxation curves, spanning T = 0.0 to 2.5 ps, were fit to single exponential decays with static offsets. The resolved features at 1655 cm^{-1} and 1681 cm^{-1} show-

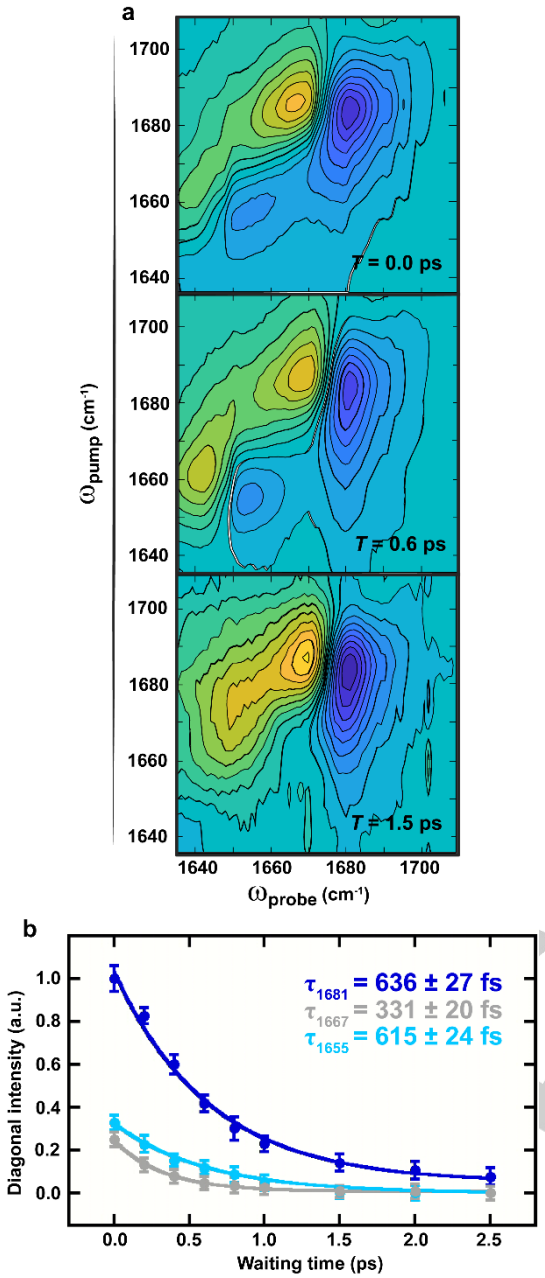

Figure 3. Waiting time dependence of K^+-TG_5T spectra. (a) 2D IR spectra at (Top-Bottom) T = 0.0 ps, 0.6 ps, and 1.5 ps. (b) Decay of ν(0-1) intensities at 1681 cm^{-1} (dark blue), 1667 cm^{-1} (grey), and 1655 cm^{-1} (light blue) to T = 2.5 ps. Intensities were normalized to the negative 1681 cm^{-1} ν(0-1) signal at 0.0 ps. Error bars indicate standard deviations of measurements. Lifetimes (τ) were extracted from single exponential functions with static offsets.

ed similar decay lifetimes of τ_{1655} = 615 ± 24 fs and τ_{1681} = 636 ± 27 fs, respectively. In contrast, the intensity at 1667 cm^{-1} decayed much more rapidly, with a lifetime of τ_{1667} = 331 ± 20 fs. Similar behaviors were observed across the guanine tract length series (Figure S12, S13, S14) and the lifetime of the high-frequency ν(0-1) signal in K^+-TG_5T is shortened below the saturation limit. These differences further support the assignment of the two resolved peaks to ordered, metal-bound guanine arrays and the ~1667 cm^{-1} signal to disordered or solvent-exposed regions where environmental fluctuations result in more rapid spectral diffusion.[31] Hereafter, we refer to the high-frequency and low-frequency K^+-TG_5T diagonal features at 1655 cm^{-1} and 1680 cm^{-1}, which vary with guanine tract length, as guanine I(L) and guanine I(H), respectively.

Vibrational coupling and delocalization shifts guanine I features

The appearance of multiple diagonal features in K^+-TG_nT 2D IR spectra, which vary with guanine-tract length and metal ion concentration, can provide useful information for the construction of structural models of the resulting complexes. However, even though these features can be distinguished by their relaxation behaviors and can be described by Gaussian fits to FTIR spectra, the congestion of the guanine I region allows various physical interpretations of the data. First, the overall broadening of the guanine I profile could be due to increased structural heterogeneity in the sample as n is increased, thus producing an increase in inhomogeneous contributions to the signals. This clearly occurs above the n = 5 length threshold with structural polymorphism observed by PAGE and SEC (Figure S5), but could also occur from n = 3 – 5 with the formation of minor sub-populations containing unobserved oligomerization states or an increase in environmental distributions within individual tetramers. To a first approximation, such an interpretation is supported by weak to absent cross-peaks between guanine I(L) and guanine I(H) signals, which suggests that they could arise from separate or highly-flexible complexes where guanine I vibrations are not coupled. However, because the overall frequency profiles actually narrow below the K^+ saturation limit of 100 mM, this interpretation remains questionable. Second, the opposing shifts of guanine I(L) and guanine I(H) could arise from coupling of guanine I vibrations among bases in ordered arrays, which in turn delocalizes the normal modes and leads to splitting of the signal. A similar phenomenon has been shown in complementary guanine-cytosine duplexes (e.g., dG_5C_5)[19], where inter-layer coupling between guanine carbonyl modes forms vibrational excitons along strands and contributes to a shift of the guanine I feature to ~1680 cm^{-1}, similar to our guanine I(H). However, we only observe a small (± 1.8 cm^{-1}) variation in guanine I(H) $\Delta\omega_a$ values within the K^+-TG_nT series (Figure S15), indicating that any such delocalization has a small effect on guanine I anharmonicity compared to exciton formation in other biomolecular assemblies such as amyloids.[24e] Important structural differences between duplex DNA and G-quadruplexes prohibit direct comparisons between the two systems. For example, intense cytosine carbonyl features in dG_5C_5 appear at the same frequency as guanine I(L) in K^+-TG_5T, and likely obscure this feature if it is present.[19] Furthermore, in-plane coupling within guanine and cytosine base pairs, which is the primary contributor to the appearance of cross-peaks,[19b,32] also contributes to the blue shift in the guanine diagonal feature, and it is not clear if the same phenomenon occurs in G-quartets. Finally, metal binding itself can shift guanine carbonyl frequencies[16b] and must be accounted for in any description of the K^+-TG_nT spectra.

Site-specific isotope substitution has emerged as a powerful method to investigate both local and distributed vibrational phenomena in complex biomolecules and biomolecular assemblies. The incorporation of an isotope "label" into one or more residues does not the perturb molecular structure but shifts the vibrational frequency of the corresponding local mode. This simultaneously separates the si-

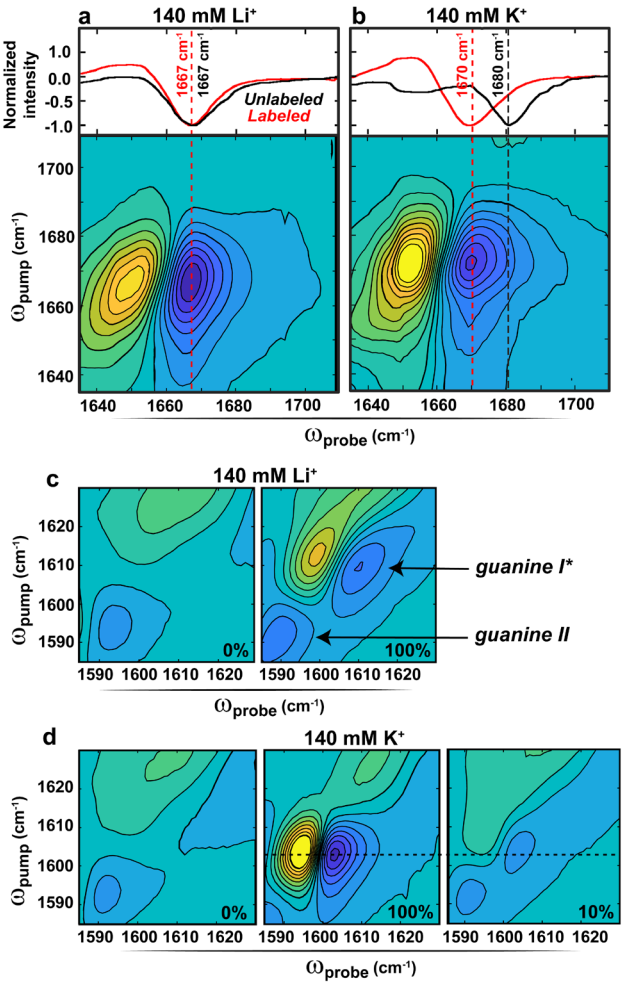

Figure 4. 2D IR spectra of $^{15}N,^{13}C$ labeled TG_5T, d(T-G-G-G*-G-G-T). (a,b) Bottom: 2D IR spectra of labeled TG_5T showing the unlabeled guanine I region in the presence of 140 mM (a) Li^+ and (b) K^+. Top: Diagonal slices through unlabeled (black) and labeled (red) in the presence of 140 mM (a) Li^+ and (b) K^+. (c) 2D IR spectra of TG_5T in the presence of Li^+ showing the guanine I*/guanine II region for (left) 0% labeled and (right) 100% labeled TG_5T. (c) 2D IR spectra of K^+-TG_5T showing the guanine I*/guanine II region for samples containing (left) 0% labeled, (middle) 100% labeled, and (right) 10% labelled TG_5T. Spectra covering the entire labeled and unlabeled frequency range (1550 – 1700 cm^{-1}) for fully labeled and unlabeled samples are presented in Figure S19.

gnal of a labeled residue from congested spectra and decreases the effects of coupling between labeled and unlabeled residues.[33] Thus, the behaviors of single labels can be analyzed in near isolation[34], coupling between pairs of labeled residues can be quantified[35], and the same concepts can be extended to uniformly labeled segments of larger structures.[24a, 24b] Importantly, perturbing the frequency of a labeled residue introduces a vibrational defect, and the resulting change in coupling can influence the IR-active modes of unlabeled regions.[20a] This effect is particularly strong if the label exists in a structural motif where vibrations are extensively delocalized, such as peptide or protein amyloid β-sheets.[24a, 33] In such systems, judicious choice of label position can therefore provide extensive information about vibrations in associated sets of residues and aid in separating the effects of inhomogeneity from the effects of coupling on observed spectra. Here, we apply this strategy to TG_5T in the presence of Li^+ and K^+ (140 mM), which displays the largest metal-dependent guanine I shifts. For 2D IR experiments (Figure 4), we synthesized a labeled variant of TG_5T by incorporating a uniformly $^{15}N,^{13}C$ labeled guanine residue at the central position, resulting in the sequence d(T-G-G-G*-G-G-T). This labeling scheme produces a quasi-symmetric isotope content along the strand, with the central labeled guanine (*) flanked by two pairs of unlabeled (natural abundance) guanines as well as two terminal thymines.

First, we focus on the unlabeled guanine I signal, which appears between 1630 – 1710 cm^{-1}; features associated with labeled bases appear at lower frequencies (*vide infra*). For samples prepared with 140 mM Li^+, the unlabeled bases produced a single peak pair at ω_{pump} = 1667 cm^{-1} (Figure 4a, bottom), which is identical to that of both free GMP (Figure S3) and unlabeled TG_5T in the presence of Li^+ (Figure S16). In diagonal slices (Figure 4a, top) we observe no measurable difference in the frequencies of labeled and unlabeled oligonucleotides in this region, so the presence of the label does not perturb the unlabeled residues within the strands. This result is consistent with minimal stabilization of G-quadruplexes by Li^+[9-10], where weak coupling between guanine carbonyls does not produce a frequency shift in the guanine I feature. In K^+-TG_5T, the formation of a stable G-quadruplex likely results in a labeled central G-quartet flanked by two unlabeled G-quartets (Figure S17). In the 2D IR spectrum of K^+-TG_5T (Figure 4b), labeling collapses the complex profile observed in Figure 2a into a single unlabeled guanine I peak pair at ω_{pump} = 1670 cm^{-1}. This result is consistent with the disruption of delocalization among unlabeled G-quartets by an isotope-induced frequency shift within the central layer of the complex. Notably, this 1670 cm^{-1} signal lies between the frequencies of free GMP (1667 cm^{-1}) and that of K^+-TG_3T formed under the same conditions (1674 cm^{-1}), so it could arise from coupling within the two pairs of unlabeled G-quartets. Although the loss of splitting in the unlabeled guanine I feature upon the introduction of label is compelling evidence for exciton formation, features associated with the labeled bases must be considered to further support this conclusion.

To do this, we examined the behavior of the labeled (guanine I*) peaks, which occur between 1590 – 1620 cm^{-1}. In samples prepared with 140 mM Li^+, the guanine I* signal appears at ω_{pump} = 1611 cm^{-1} near the unlabeled guanine II (ring mode) peak pair at ω_{pump} = 1595 cm^{-1} (Figure 4c). The guanine I* peak is elongated along the diagonal with nodal line and center line slopes of 2.1 and 1.9, respectively, indicating a broad structural distribution of weakly metal-associated oligonucleotides.[36] In these spectra, a weak guanine I/I* cross peak appears at $\omega_{pump},\omega_{probe}$ = 1611, 1667 cm^{-1} (Figure S18a) and a minor (~4 cm^{-1}) red shift in the guanine II feature is observed upon labeling, suggesting residual coupling between guanine ring modes in the Li^+ samples that may explain the parallel-stranded G-quadruplex character observed in the CD spectra (Figure S7). For K^+-TG_5T, the guanine I* peak pair appears at ω_{pump} = 1604 cm^{-1} (Figure 4d) and is narrowed compared to the same feature in Figure 4c, with nodal line and center line slopes of 5.8 and 4.6, respectively, reflecting a decrease in inhomogeneity compared to the Li^+ samples. Because only a single guanine I* feature is observed along the diagonal, the aforementioned model of a single labeled G-quartet (Figure S17) is supported, and no substantial population

of non-tetrameric K^+-TG_5T is present in the sample. A strong guanine I/I* cross-peak appears at $\omega_{pump},\omega_{probe}$ = 1604, 1670 cm^{-1} (Figure S18b), indicating enhanced inter-layer coupling between carbonyl modes upon K^+ binding. As in the corresponding Li^+ spectra, the unlabeled guanine II feature shifts to lower frequencies but this effect is larger for K^+ complexes, supporting the conclusion that ring modes couple between layers to produce the ~1595 cm^{-1} feature even in samples prepared with Li^+. Notably, no cross-peaks are apparent between guanine I* and guanine II features in the spectra of Li^+ or K^+ samples, so inter-layer coupling between labeled carbonyl and unlabeled ring modes is weak.

Comparing the guanine I* frequencies of labeled TG_5T in the presence of Li^+ and K^+-TG_5T (Figure 4c,d), we observe a -7 cm^{-1} potassium-dependent shift. This shift results from a combination of hydrogen-bonding between bases, metal ion coordination, inter-layer coupling, and coupling between the four guanines in the central G-quartet of K^+-stabilized complexes.[16-17] Because these phenomena arise from cooperative formation of the complexes, the relative contributions of each cannot be extracted from individual spectra without the aid of high-level *ab initio* calculations.[27, 32, 37] However, because each guanine in the central G-quartet is in a different TG_5T strand, a comparison of the spectra of isotopically dilute complexes allows the contribution of in-plane coupling to the observed frequency shift to be extracted experimentally. We collected 2D IR spectra of a series of K^+-TG_5T samples prepared with varying ratios of labeled and unlabeled oligonucleotides (Figure S20). In these spectra, a progressive shift of the guanine II feature to lower frequencies confirms efficient formation of the mixed complexes. Surprisingly, the frequency of the guanine I* did not shift as the proportion of labeled strands decreased from 100% to 10% (Figure 4d). This result indicates that in-plane coupling within a G-quartet has a minimal effect on the guanine I frequency, possibly due to the near-orthogonal orientations of the four transition dipoles (Figure S1), and that the -7 cm^{-1} shift arises from metal ion coordination and hydrogen-bonding between guanine bases. Based on prior work,[16b] it is likely that C=O bond lengthening upon metal ion coordination is primary contributor to this shift.

From the spectra of unlabeled and labeled TG_5T in the presence of Li^+ (Figure 4a,c), we find an isotope shift of approximately -56 cm^{-1}. Combining this with the -7 cm^{-1} metal-dependent shift and the 1604 cm^{-1} guanine I* signal in K^+-TG_5T (Figure 4d), we estimate the frequency of the unlabeled central layer to be near 1660 cm^{-1} in the potassium-saturated complex. This estimate, which neglects inter-layer coupling effects, lies between the frequencies of the guanine I(L) and GMP-like disordered signals in the unlabeled G-quadruplex (Figure 2). However, because coupling between the central G-quartet and the flanking layers is clearly apparent via the $\omega_{pump},\omega_{probe}$ = 1604, 1670 cm^{-1} guanine I/I* cross-peak in labeled K^+-TG_5T (Figure S18,19) but no appreciable cross-peak appears between guanine I(L) and guanine I(H) in the spectrum of unlabeled K^+-TG_5T (Figure 2a), the latter features most likely arise from delocalization of guanine I vibrations among coupled layers and not local structural heterogeneity.

In summary, heavy isotope labeling of the central G-quartet within ordered K^+-TG_5T tetramers not only shifts the labeled modes to lower frequencies but also perturbs the observed modes of the adjacent unlabeled G-quartets. In the labeled K^+-TG_5T 2D IR spectra, the labeled guanine I* and unlabeled guanine I features both appear as narrow diagonal peak pairs (Figure 4b,d), and cross-peaks between guanine I* and guanine I modes (Figure S18,19) indicate vibrational coupling between G-quartet layers. Although isotope labeling within coupled arrays is well-known to redistribute intensities between modes and influence their inhomogeneity[33], the diagonal features in the labeled K^+-TG_5T spectrum cannot account for the observed broadening or the resolved guanine I(L) and guanine I(H) peaks in spectra of unlabeled samples. This is particularly clear in the high-frequency guanine I(H) feature, which appears to be entirely dependent on ordered assembly of near degenerate guanine carbonyl modes and scales with the guanine tract length in K^+-TG_nT tetramers (Figure 2a). Thus, the observed broadening is unlikely to arise from redistribution of intensity among localized vibrations and is better understood as a result of vibrational delocalization among bases in ordered arrays. Interestingly, guanine I frequencies depend on axial degeneracy between G-quartets (Figure 4b) but guanine I* frequencies are independent of equatorial degeneracy within them (Figure 4d). From this, we conclude that the guanine I(L) and guanine I(H) features in K^+-TG_nT reflect the formation of axially-oriented excitons due to coupling of guanine carbonyls in stacked G-quartets.

The appearance of the guanine I(H) near 1680 cm^{-1} is similar to that of intra-strand excitons in dG_5C_5 described by Krummel *et al*.[19], which coincidentally matches the limit of five consecutive guanine bases per strand before the onset of polymorphism in K^+-TG_nT. However, although the high-frequency guanine carbonyl feature appears at ~1680 cm^{-1} in both systems, this shift (+13 cm^{-1} vs to GMP) arises from different phenomena. In dG_5C_5, coupling between complementary GC base pairs and between stacked guanines within a strand both contribute to the observed shift, with intra-strand coupling constants on the order of +10 cm^{-1}.[19] In contrast, in K^+-TG_5T the in-plane (G-quartet) contribution is near zero. The overall +13 cm^{-1} guanine I(H) blue shift in K^+-TG_5T also includes a -7 cm^{-1} metal-dependent red shift of individual K^+-bound guanines, so we estimate that the shift due to coupling between G-quartets is approximately +20 cm^{-1}. The source of this increase vs. duplex DNA is not clear from the data, but it could arise from larger positive coupling constants due to the helical twist of parallel-stranded G-quadruplexes or enhanced mechanical coupling via coordinated metal ions.[20a] Finally, although the guanine I(L) features observed in our spectra appear to follow the excitonic model, they were not clearly observed in either experimental or simulated spectra of guanines stacked within a single strand in dG_5C_5 and other similar systems.[19b, 32, 37d] Based on the structure of a G-quadruplex, both intra-strand and inter-strand coupling could contribute to the formation of excitons among stacked G-quartets, yielding potentially richer manifolds of allowed transitions. However, the molecular-level details of these phenomena remain to be determined by calculations and further isotope labeling experiments. Approaches to this question are highlighted in the Discussion.

Quantitative structural analysis from FTIR spectra

To determine the structural landscape of K^+-TG_nT complexes across the guanine-tract length series and with varying [K^+], it is necessary to find the sizes of ordered G-quadr-

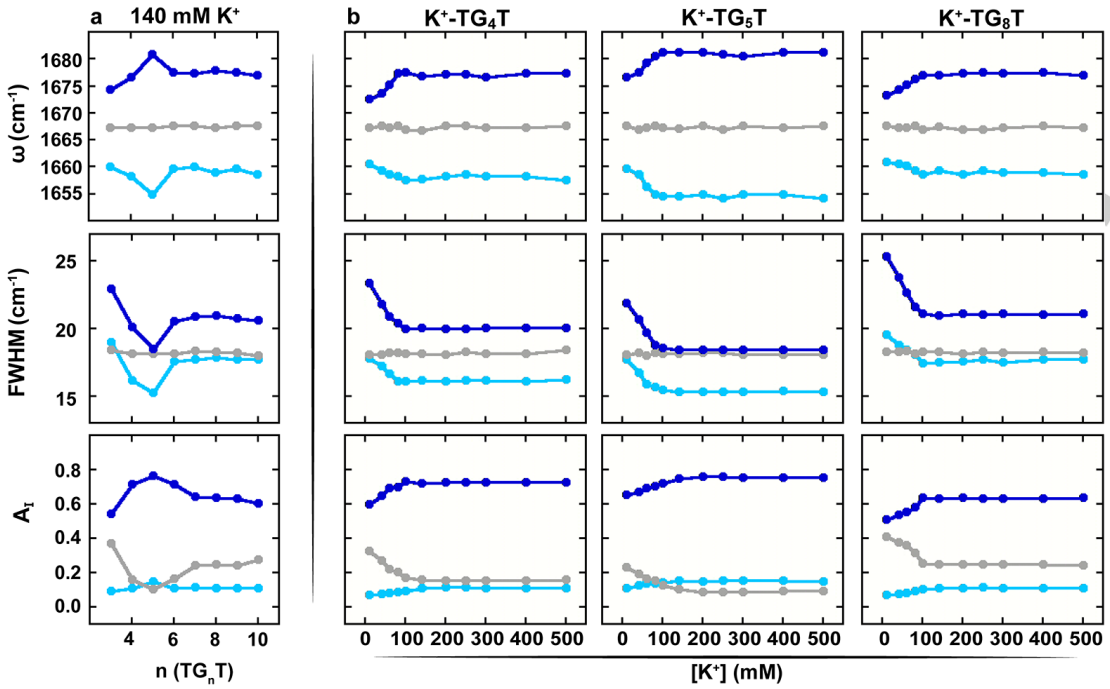

Figure 5. Gaussian deconvolution of K^+-TG_nT FTIR spectra. In all panels, values for the high frequency (H), disordered, and low frequency (L) components are shown in dark blue, grey, and light blue, respectively. (a) Fit parameters for K^+-TG_nT formed with 140 mM K^+: ω_{ctr} (top), FWHM (middle), A_I (bottom). (b) Fit parameters from for K^+-TG_4T (left), K^+-TG_5T (middle) and K^+-TG_8T (right) at varying [K^+]. All points are averages of three independent trials and error bars are omitted for clarity. All values with standard deviations are reported in Tables S3, S4, S5, S6.

uplex motifs as well as the relative contributions of ordered and disordered regions to the ensembles. 2D IR spectroscopy provided the assignments necessary to do so, but because 2D IR signal scales as $|\mu|^4$ and coupling modulates μ, the intensities of diagonal peaks cannot be used to calculate structural distributions without detailed knowledge of the relevant transition dipole magnitudes.[20a] Moreover, even if such information was available, interference from v(1-2) transitions obscures some diagonal intensity and introduces uncertainty into the analysis. FTIR spectroscopy, where peaks scale as $|\mu|^2$, is actually advantageous in this situation because no v(1-2) interference occurs and total intensity conserved as coupling strengths change. In our data, Gaussian deconvolution of the FTIR signal closely reproduced the 2D IR features shown in Figure 2. We applied the same analysis to the FTIR spectra (Figure 1, S2b) of saturated (140 mM K^+) K^+-TG_nT over the entire tract length series as well as selected K^+-TG_nT (n = 4, 5, 8) prepared with varying [K^+]. The center frequencies (ω_{ctr}), full width at half-maximum linewidths (FWHM) and fractional integrated areas (A_I) extracted from resulting the fits are compiled in Figure 5.

For K^+-TG_nT formed with 140 mM K^+ (Figure 5a), the guanine I(H) and guanine I(L) center frequencies diverge symmetrically about the mean with a maximum separation ($\Delta\omega_{ctr}$) of 26 cm^{-1} at n = 5; above n = 5, $\Delta\omega_{ctr}$ is stable at ~19 cm^{-1}, which is similar to the divergence at n = 4. As the peaks diverge, they also narrow from FWHM maxima at n = 3 to minima at n = 5, followed by a sharp return to a value near n = 3– 4 for longer tracts. Throughout the series, the intermediate (disordered) component remains stable within error at 1667 cm^{-1} and its FWHM is invariant with n. Finally, we observe an inverse relationship between the fractional areas of the guanine I(L) and guanine I(H) versus the disordered components that reverses at the n = 5 threshold. Considering the large axial coupling effects observed via isotope labeling of K^+-TG_5T (Figure 4), this is consistent with a transition between structures containing coupled and partially decoupled G-quadruplexes. We note that the quality of the three-component fits is somewhat reduced above the n = 5 threshold, likely due to a distribution of structures in polymorphic K^+-TG_nT. Additionally, the consistent ~5 cm^{-1} difference between the guanine I(H) and guanine I(L) FWHM values may be due to intrinsic differences in the linewidths of two different axially-oriented excitons, error associated with baseline subtraction, or the presence of weak and unresolved high-frequency signals such as thymine I[27] that are not considered in the fitting model but contribute to the apparent guanine I(H) FWHM. Although not all of the TG_nT modes are accounted for in this analysis, the results agree well with an increase in ordered G-quartet stack height up to a limit of n = 5, after which structural defects persist and ordered sets of only 3 – 4 adjacent G-quartets occur.

As [K^+] is increased (Figure 5b), the TG_4T, TG_5T, and TG_8T spectra also show divergence of ω_{ctr}, FWHM narrowing, and A_I increases in the guanine I(H) and guanine I(L) signals up to ~100 mM K^+, after which they remain constant within experimental error. Again, ω_{ctr} and FWHM are invariant for the disordered component. From the A_I trends, we find that disordered or uncoupled bases (including thymines) only contribute ~10% of the total FTIR signal in saturated K^+-TG_5T but about 25% of the signal in saturated K^+-TG_8T. Therefore, at the n = 5 threshold nearly all guanine bases are included in the stable, coupled array, and above the threshold multiple bases are excluded from it. This is consistent with the prediction of ordered stacks of 3 – 4 G-quartets in G-quadruplexes formed by

longer oligos, in which localized motifs are interspersed with disordered loops or unoccupied metal sites.

The K^+-TG_nT assembly landscape

These results allow the construction of an empirical landscape for the assembly of ordered K^+-TG_nT G-quadruplexes (Figure 6). From FTIR and 2D IR spectroscopy, a guanine tract length threshold at n = 5 and a saturation threshold at 100 mM K^+ delineate four quadrants containing parallel-stranded oligomers that form G-quadruplexes as well as different types and degrees of local disorder. In Figure 6a, short, K^+-saturated complexes are highly-ordered both within and between G-quartet layers and excitons are extensively delocalized along the G-quadruplex axis. In Figure 6b, vacant ion binding sites at low [K^+] attenuate exciton lengths and guanine I splitting. In Figure 6c, G-quadruplexes containing 3 – 4 adjacent G-quartets form with some bases excluded from the ordered stacks. Here, both tetramers and higher-order oligomers can form, as indicated by both PAGE and SEC, and the absence of ~1680 cm^{-1} guanine I(H) signal and a resolved guanine I(L) peak in the 2D IR spectra of longer (n ≥ 6) tracts shows that even the tetrameric sub-population contains local structural defects. For tracts of 6 – 10 guanines, multimers of monomeric parallel-stranded G-quadruplexes[38] are unlikely because few nucleotides are available to form the required loops; instead, excluded bases may be available for association with additional strands, resulting in the formation of higher-order oligomers resembling G-wires.[39] Finally, in Figure 6d the combination of vacant sites and excluded bases produces mixtures of oligomers with only marginal G-quadruplex character.

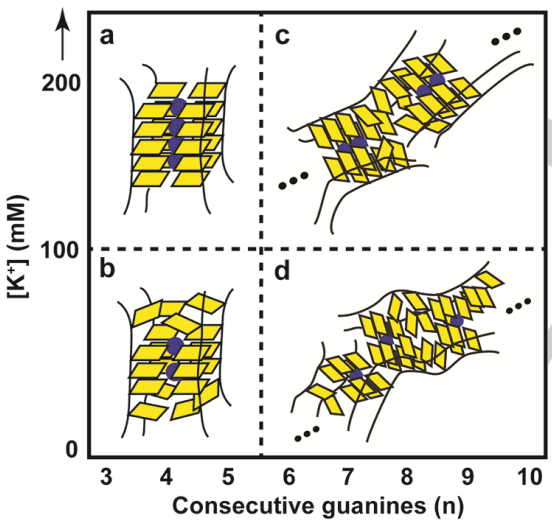

Figure 6. Structural models of K^+-TG_nT complexes defined by guanine tract length (*n*) and potassium concentration thresholds. Guanine bases are shown in yellow and K^+ ions in blue. Dashed lines indicate the experimentally-defined thresholds in n (vertical) and [K^+] (horizontal). (a) K^+-saturated tetramers. (b) Tetramers with vacant sites and local disorder. (c) Representative section of a higher-order oligomer with locally-ordered, K^+-saturated motifs (n = 3 – 4). (d) Representative section of an unsaturated higher-order oligomer.

Discussion

In this work, we showed that vibrational coupling is the primary contributor to the characteristic guanine C6=O6 peak shifts in the vibrational spectra of G-quadruplexes. Beginning with the known structures of parallel-stranded complexes containing four stacked G-quartets, we examined trends on guanine I vibrations as poly-guanine tract lengths and metal ion conditions were varied. We found that both K^+- and Na^+-TG_nT complexes transition from tetramers to mixtures of oligomers above five consecutive guanine bases, and that K^+-TG_nT complexes are fully formed above 100 mM K^+, independent of tract length. Using 2D IR spectroscopy, we resolved overlapping guanine I peaks between 1630 cm^{-1} and 1710 cm^{-1} and assigned them to signals from excitonically-coupled and disordered base vibrations. Through heavy isotope substitution, we found that excitons delocalized among stacked G-quartets are the primary contributors to the observed frequency and lineshape shifts. Using this information, we quantified the structural contributions to G-quadruplex FTIR spectra across a range of guanine tract lengths and potassium ion concentrations. Combining these results, we constructed a landscape for the assembly of K^+-TG_nT complexes and found that short (n = 3 – 5) locally ordered G-quadruplexes are favored in parallel stranded structures even when additional G-quartet layers could potentially form.

The large contribution of inter-layer coupling to guanine I shifts is a key piece of information that must be considered in any application of vibrational spectroscopy to determining G-quadruplex structures. Vibrational coupling between bases, not perturbations arising from metal cation binding or hydrogen-bonding,[16-17] is the primary source of commonly observed blue shifts in guanine C=O peaks in the infrared spectra of G-quadruplexes. Thus, this spectral feature does not necessarily correlate to the stability of the structure, but instead reports the sizes of motifs formed above a minimum stability threshold. This is an important distinction that motivates the re-examination of previous results and informs the design of future experiments.

The sensitivity of coupling effects to variations in G-quadruplex structures is clear from our data. Extensive vibrational delocalization in K^+-TG_nT motifs allowed different structural contributions to their spectra be resolved and analyzed. Because this phenomenon depends on both the degeneracy of local modes and the distances and orientations of transition dipoles in repeated patterns, even small discontinuities in stacking or variations in metal coordination are sensed. For example, guanine I frequency shifts in Na^+-TG_nT are attenuated compared to K^+-TG_nT, even though the same assembly thresholds are observed. Previous structures of Na^+-bound G-quadruplexes suggest that the inter-layer distance is larger than that of K^+-bound structures, which may contribute to this effect.[40] Alternatively, the smaller cationic radius of Na^+ could alter local coordination geometries or metal site occupancies in both tetramers and higher-order species. A similar size effect has been observed in the selectivity filters of K^+-selective ion channels such as KcsA, where Na^+ binding restricts ion transport and exchange through the channel.[41] It remains to be determined if these small variations have any significance in the biological function of G-quadruplexes. However, the ability of infrared spectroscopy to detect structural variations and defects in G-quadruplexes through their effects on vibrational delocalization makes it a valuable complement to NMR and X-

ray crystallographic approaches, where local changes in G-quartet stacking can be obscured in ensemble averages.

Because G-quadruplexes form a diverse family of structures,[1, 13d] extension of our methods to related systems will require additional information about the strengths and orientations of guanine I transition dipoles in metal-bound states as well as the associated coupling constants. These values can be determined experimentally through further isotope labelling studies[24c, 35] and comparisons of linear and third-order vibrational spectra.[28] Values for model systems can also be determined through *ab initio* or semiempirical calculations, and further applied to calculate FTIR and 2D IR spectra using the results of molecular dynamics simulations.[19, 24b, 27, 32, 37] Theoretical spectra can then be compared to experimental results to assign vibrational features and test structural models. The combination of these methods can be used to further refine our K^+-TG_nT models by resolving the remaining ambiguity in disordered spectral features, which have undetermined locations in the structures and may result from both conformational and vibrational effects.[20a] Other G-quadruplex topologies, such as antiparallel structures with mixed *syn*/*anti* base conformations,[1, 13d] have different relative transition dipole orientations so these approaches must be applied separately in each case. However, when assignments can be made, they can be straightforwardly applied to FTIR spectra to gain a detailed understanding of G-quadruplex assembly landscapes and the structures contained therein. A similar strategy can be used to analyze other infrared features, such as the guanine II ring mode, which appears to be sensitive to base stacking but relatively insensitive to metal ion coordination (Figure 4). Extension of this approach will be especially useful for distinguishing structures in cases where competing models have been proposed. [24b]

Although our K^+-TG_nT study is limited to parallel-stranded DNA G-quadruplexes, the assembly landscape is surprisingly consistent with the known and predicted behaviors of model systems as well as guanine-rich sequences in biology and disease. The highly-ordered tetrameric region (Figure 6a) corresponds to the typical guanine tract lengths in genomes[1c, 4-5] and the prevailing ionic conditions within healthy mammalian cells.[5b, 30] Most high-resolution structures have guanine tracts that fall within this range, and molecular dynamics simulations of K^+-bound dG_8 have shown a preference for five ordered G-quartets bound to four K^+ ions in long trajectories.[17] Recently, a similar length preference was discovered for stable i-motif formation by complementary poly-cytosine tracts,[42] suggesting that evolution may have co-selected sequences that are pre-disposed to ordered assembly. Outside this region, structural data is sparse, but many biological systems appear to obey the same threshold behavior determined in our study. For example, decreases in intracellular K^+ concentrations destabilize G-quadruplexes formed by short guanine tracts in the *c-Myc* promoter region, allowing transcription of oncogenes and promoting tumor progression (*cf.* Figure 6b).[5b] Additionally, short guanine tracts (n = 3 – 4) are commonly separated by 1 – 2 non-guanine bases in repeat expansions[5a, 7] and 2 – 4 in telomeres.[1c, 2-3, 9, 38] These tracts may form localized G-quadruplex domains separated by loops, either with antiparallel strands or parallel strands approximating the structural distributions above the n = 5 length threshold (*cf.* Figure 6c,d), which may further aggregate in long repeat sequences.[7b, 39] Although these similarities may be circumstantial, they suggest a fundamental limit to G-quartet stacking in G-quadruplexes, which can be tested using the methods and insights presented here. It is likely that analogous landscapes also exist for alternative G-quadruplex topologies, and based on the available information similar stacking limits likely apply. Further examination of these limits would provide valuable information for sequence-based structural predictions,[1a, 4-5, 5c] the development of molecular probes and therapeutics that target G-quadruplexes,[1b, 5c, 8, 12] and design principles that can be used in nucleic acid-based sensors and nanotechnology.[13]

Conclusion

We have shown that infrared spectroscopy detects the sizes of locally-ordered G-quadruplex motifs through the extent of vibrational delocalization among their constituent bases. We used assignments of spectral features derived from 2D IR spectra to characterize the structures accessible to model G-quadruplex complexes. By combining the sensitivity of 2D IR spectroscopy with the quantitative value of FTIR spectroscopy, we found a structural landscape that bears striking similarities to the behaviors of guanine-rich sequences *in vitro* and *in vivo*. Using our insights and approaches, it will be possible to determine additional structural details in diverse G-quadruplexes through future FTIR and 2D IR experiments combined with established theoretical methods. The results reported here are an important advance in nucleic acid structural biology, and with frequency-resolved IR microscopy[43] and the recent advent of 2D IR imaging techniques,[44] G-quadruplex structures may even be identified and resolved in structurally heterogeneous assemblies, including engineered materials and biological samples that are relevant to disease.

Experimental Section

Sample Preparation. All chemicals were purchased from Fisher Scientific (Hampton, NH) and were used as received. Unlabeled TG_nT oligonucleotides were purchased from Integrated DNA Technologies, Inc. (Coralville, IA) and labeled TG_5T was synthesized as described below. The sequences of the oligonucleotides are shown in Table S1. Stock solutions (10 mM) of DNA were prepared in D_2O and allowed to exchange for 3 hours at ambient temperature. For analysis, 3 mM solutions of each oligonucleotide were prepared separately in buffered D_2O containing 40 mM Tris (pD 7.2) and varying concentrations of metal chloride salt (LiCl, NaCl, or KCl). Each sample was then heated to 90 °C for 3 minutes to disaggregate oligonucleotides and cooled slowly to room temperature to assemble their metal complexes.[11]

To ensure that the results for longer TG_nT were not biased by aggregates formed during prior purification, we also pre-treated selected oligonucleotides in base (LiOH) prior to metal complexation following the method of Bardin and Leroy.[29] Briefly, lyophilized TG_nT oligonucleotides were resuspended in 0.1M LiOH, heated to 95°C for 5 minutes, and cooled slowly to room temperature. After cooling, samples were dialyzed once against 0.1 M LiCl to remove base and then three times against H_2O to remove Li^+ ions. Finally, samples were lyophilized and prepared as described above. The FTIR spectra of LiOH-treated samples did not differ, outside of experimental error, from those prepared by heating alone (Figure S6).

Synthesis of ^{13}C, ^{15}N TG_5T oligonucleotide. The isotopically-labeled oligonucleotide was synthesized on a 2-μmole scale on an

Applied Biosystems 394 DNA synthesizer. Oligonucleotides were assembled on 1,000 Å CPG functionalized with Unylinker at 42 µmol/g (ChemGenes, Wilmington, MA). DNA Guanine/Thymine CED phosphoramidites (ChemGenes) and 2'-deoxyguanosine phosphoramidite ($^{13}C_{10}$, 98%; $^{15}N_5$, 98%) (Cambridge Isotope Laboratories, Andover, MA) were dissolved in acetonitrile to 0.15 M; the coupling time was 25 seconds for each unlabeled base and 10 minutes for the labeled base. Cleavage from the support as well as removal of the Unylinker moiety and nucleobase protecting groups were effected by treatment with concentrated aqueous ammonia for 16 hours at 55 °C. The deprotected product was resuspended in water and its mass verified using ESI-MS in negative ion mode (calculated mass 2,207.3 Da, found 2,206.4 Da). Small molecule impurities were removed with Amicon Ultra centrifugal filters (3 kDa molecular weight cutoff, Millipore Sigma). Labeled TG_5T was purified by reverse-phase HPLC using a linear (0 – 100%) gradient of acetonitrile in 0.1 M TEAA on a 4.6 x 150 mm C18 5 µm column. The purified oligonucleotide was lyophilized and resuspended in water twice and then dialyzed against water three times to ensure the removal of HPLC solvents. The dialyzed product was lyophilized and re-dissolved to 10 mM in D_2O, allowed to exchange for 3 hours, and treated as described in Sample Preparation (above).

FTIR Spectroscopy. Samples were placed in demountable transmission cells between two 2 mm thick CaF_2 windows separated by 50 µm PTFE spacers. Sample and background spectra were collected with 128 scans at room temperature using a Jasco 6800 FTIR spectrometer equipped with a single-channel mid-band MCT detector. Spectra were collected with a resolution of 0.5 cm^{-1}. The spectrometer and sample chamber were purged with dry air to eliminate interference from water vapor during all experiments. Buffer backgrounds were collected separately and subtracted from sample spectra, and residual baseline corrections were performed using low-order polynomial fits in MATLAB. Frequencies of the maxima (ω_{max}) were determined from the zeroes of first derivatives of each spectrum.

FTIR Fitting Analysis. For the analysis reported in the article text, nonlinear least-squares fits to three Gaussian functions were performed in MATLAB using background-corrected FTIR spectra. For each function, ω_{ctr} was constrained within 15 cm^{-1} windows centered on the corresponding 2D IR peak maxima, FWHM was constrained within a range of 5 – 40 cm^{-1}, and positive intensity was allowed to vary freely.

To ensure that our analysis was robust, we also considered alternative fitting models. In Figure S21, three Gaussian functions were used to fit spectra of K^+-TG_nT samples prepared with 140 mM K^+ but the ω_{ctr} values of guanine I(L) and guanine I(H) peaks were fixed those of K^+-TG_5T (1655 cm^{-1} and 1681 cm^{-1}). Based on residuals, the quality of fits for all other n was poor, so we concluded that ω_{ctr} must vary to reproduce the FTIR profile. In Figure S22, a reduced model of two Gaussian functions was applied with the same constraints used in the initial analysis. Here, fit quality improved but the low-frequency function did not reproduce the guanine I(L) peaks observed in the corresponding 2D IR spectra (Figure 2).

Polyacrylamide Gel Electrophoresis. 18% native polyacrylamide gels were prepared using 19:1 acrylamide:bis-acrylamide supplemented with Tris buffer (pH = 7.2). Gels were run at room temperature using 20 mM Tris, 0.5 mM EDTA and 40 mM of the appropriate alkali metal salt (KCl, NaCl, or LiCl) as the running buffer. Gels were fixed at 4 °C in 50% ethanol, 37% H_2O, 10% acetic acid, and 3% glutaraldehyde for 30 minutes, and stained with methylene blue.

Size-exclusion chromatography. Size-exclusion chromatography was performed using a Hitachi Elite LaChrom HPLC system equipped with a Zenix SEC-100 column (Sepax Technologies, Inc.). All samples were prepared as described above and diluted with buffer to 300 µM concentrations immediately prior to injection in 5 µL volumes. Samples were eluted by flowing with 40 mM Tris buffer (pH 7.2) supplemented with appropriate metal concentration (LiCl, NaCl, or KCl), at 0.35 mL/min. Elution profiles were measured via absorbance at 220 nm.

Circular Dichroism Spectroscopy. Metal-TG_nT samples were prepared identically to the FTIR and 2D IR samples (3 mM TG_nT), but buffers were prepared with H_2O instead of D_2O. For analysis, each sample was placed in a demountable 0.01 mm quartz cell (Starna Cells, Inc., Atascadero, CA). UV-CD spectra were collected at 25 °C in 1 nm steps over a wavelength range of 220-330 nm, using an Aviv Biomedical 410SF spectrometer. Buffer backgrounds collected in the same cells were subtracted from sample spectra.

2D IR Spectroscopy. For 2D IR spectroscopy, 1.75 mJ, ~100 fs pulses (790 nm) from an UpTek Phidia Ti:Sapphire amplifier (Uptek Solutions Corp., Bohemia, NY) were directed into a TOPAS-Prime OPA equipped with a $AgGaS_2$ difference frequency generator (Light Conversion, Vilnius, Lithuania) tuned to generate ~17 mJ mid-IR pulses centered at 6000 nm. Mid-IR pulses were split into pump and probe beams in a 2DQuick Array spectrometer (Phasetech Spectroscopy, Inc., Madison, WI) equipped with a 128x128 pixel liquid nitrogen-cooled MCT array detector. Samples were placed in demountable transmission cells between two 2 mm thick CaF_2 windows separated by 50 µm PTFE spacers. The pump beam was shaped with a Ge acousto-optic modulator to produce pump pulse pairs, and overlapped with the probe on the sample in partially-colinear pump-probe geometry. Signals were dispersed on the detector using a 30 line/mm grating in a Princeton Instruments Acton SpectraPro SP-2150 spectrometer and self-hetereodyned with residual probe light. All spectra were collected with parallel pump and probe polarizations and pump-probe waiting times (T) were set relative to the temporal overlap resulting in maximal signal intensity. Time-domain 2D IR spectra were collected with 107 24 fs pump delays with four-frame phase cycling to reduce interference from pump scatter. During data collection, the amplifier, OPA, and spectrometer were purged continuously with dry air to reduce the interference from water vapor. Probe frequencies were calibrated to the absorbances of 4-nitrobenzaldehyde (1605, 1676, and 1709 cm^{-1}), used as an external standard. Spectra were processed in MATLAB as described previously, using low-order polynomial background correction and Hamming apodization.[24a, 45] Diagonal anharmonic shifts ($\Delta\omega_a$) were measured by fitting horizontal slices of 2D IR spectra to two pseudo-Voigt functions of opposite signs using a nonlinear least-squares routine in Origin. Nodal line and center line slopes were determined using linear slopes calculated within the diagonal v(0-1) FWHM as described previously.[24a, 36]

Acknowledgements

This research was supported via a Team Development Grant from the SIU School of Medicine, to K.T.G. and S.D.M. T.D.H. is an NSF Graduate Fellow (DGE 1545870). T.M.L. is an NSF-REU participant under grant 1757954. The authors thank Dr. Truc Chi Pham and Prof. Abby Parrill of the University of Memphis Chemistry Department for use of their CD spectrometer.

Keywords: DNA • G-quadruplex • Infrared • 2D IR • Vibrational Coupling

[1] a) J. Spiegel, S. Adhikari, S. Balasubramanian, *Trends Chem.*; b) J. L. Huppert, *Chem. Soc. Rev.* **2008**, *37*, 1375-1384; c) S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, S. Neidle, *Nucleic Acids Res.* **2006**, *34*, 5402-5415; d) M. M. Fay, S. M. Lyons, P. Ivanov, *J. Mol. Biol.* **2017**, *429*, 2127-2147.

[2] A. L. Moye, K. C. Porter, S. B. Cohen, T. Phan, K. G. Zyner, N. Sasaki, G. O. Lovrecz, J. L. Beck, T. M. Bryan, *Nat. Commun.* **2015**, *6*, 7643.

[3] F. W. Smith, J. Feigon, *Nature* **1992**, *356*, 164-168.

[4] A. R. Langley, S. Graf, J. C. Smith, T. Krude, *Nucleic Acids Res.* **2016**, *44*, 10230-10247.

[5] a) R. G. Thys, C. E. Lehman, L. C. T. Pierce, Y.-H. Wang, *Curr. Genomics* **2015**, *16*, 60-70; b) H. Tateishi-Karimata, K. Kawauchi, N. Sugimoto, *J. Am. Chem. Soc.* **2018**, *140*, 642-651; c) C. K. Kwok, C. J. Merrick, *Trends Biotechnol.* **2017**, *35*, 997-1013.

[6] M. Nambiar, M. Srivastava, V. Gopalakrishnan, S. K. Sankaran, S. C. Raghavan, *Mol. Cell. Biol.* **2013**, *33*, 4266-4281.

[7] a) B. Zhou, C. Liu, Y. Geng, G. Zhu, *Sci. Rep.* **2015**, *5*, 16673; b) A. Jain, R. D. Vale, *Nature* **2017**, *546*, 243.

[8] E. Ruggiero, S. N. Richter, *Nucleic Acids Res.* **2018**, *46*, 3270-3283.

[9] J. R. Williamson, M. K. Raghuraman, T. R. Cech, *Cell* **1989**, *59*, 871-880.

[10] a) D. Bhattacharyya, G. Mirihana Arachchilage, S. Basu, *Front. Chem.* **2016**, *4*, 38; b) G. Laughlan, A. Murchie, D. Norman, M. Moore, P. Moody, D. Lilley, B. Luisi, *Science* **1994**, *265*, 520-524.

[11] P. A. Rachwal, T. Brown, K. R. Fox, *Biochemistry* **2007**, *46*, 3036-3044.

[12] a) D. Yang, K. Okamoto, *Future Med. Chem.* **2010**, *2*, 619-646; b) S. Asamitsu, S. Obata, Z. Yu, T. Bando, H. Sugiyama, *Molecules* **2019**, *24*, 429; c) G. W. Collie, G. N. Parkinson, *Chem. Soc. Rev.* **2011**, *40*, 5867-5892.

[13] a) L. Stefan, D. Monchaud, *Nat. Rev. Chem.* **2019**; b) S. Modi, D. Bhatia, F. C. Simmel, Y. Krishnan, *J. Phys. Chem. Lett.* **2010**, *1*, 1994-2005; c) J. Ida, S. K. Chan, J. Glökler, Y. Y. Lim, Y. S. Choong, T. S. Lim, *Molecules* **2019**, *24*, 1079; d) S. A. Dvorkin, A. I. Karsisiotis, M. Webba da Silva, *Sci. Adv.* **2018**, *4*, eaat3007.

[14] a) A. Barth, *Biochim. Biophys. Acta* **2007**, *1767*, 1073-1101; b) E. Taillandier, J. Liquier, in *Methods in Enzymology, Vol. 211*, Academic Press, **1992**, pp. 307-335.

[15] S. D. Moran, M. T. Zanni, *J. Phys. Chem. Lett.* **2014**, *5*, 1984-1993.

[16] a) M. R. Guzman, J. Liquier, S. K. Brahmachari, E. Taillandier, *Spectrochim. Acta. A, Mol. Biomol. Spectros.* **2006**, *64*, 495-503; b) V. Andrushchenko, P. Bouř, *J. Phys. Chem. B* **2009**, *113*, 283-291.

[17] V. Andrushchenko, D. Tsankov, M. Krasteva, H. Wieser, P. Bouř, *J. Am. Chem. Soc.* **2011**, *133*, 15055-15064.

[18] Y. Li, X. Han, S. Zhou, Y. Yan, X. Xiang, B. Zhao, X. Guo, *J. Phys. Chem. Lett.* **2018**, *9*, 3245-3252.

[19] a) A. T. Krummel, M. T. Zanni, *J. Phys. Chem. B* **2006**, *110*, 13991-14000; b) A. T. Krummel, P. Mukherjee, M. T. Zanni, *J. Phys. Chem. B* **2003**, *107*, 9165-9169.

[20] a) P. Hamm, M. Zanni, *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, Cambridge, **2011**; b) N. T. Hunt, *Chem. Soc. Rev.* **2009**, *38*, 1837-1848; c) Y. S. Kim, R. M. Hochstrasser, *J. Phys. Chem. B* **2009**, *113*, 8231-8251.

[21] M. C. Thielges, M. D. Fayer, *Acc. Chem. Res.* **2012**, *45*, 1866-1874.

[22] G. Hithell, P. M. Donaldson, G. M. Greetham, M. Towrie, A. W. Parker, G. A. Burley, N. T. Hunt, *Chem. Phys.* **2018**, *512*, 154-164.

[23] C. Greve, T. Elsaesser, *J. Phys. Chem. B* **2013**, *117*, 14009-14017.

[24] a) S. D. Moran, A. M. Woys, L. E. Buchanan, E. Bixby, S. M. Decatur, M. T. Zanni, *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 3329-3334; b) L. E. Buchanan, J. K. Carr, A. M. Fluitt, A. J. Hoganson, S. D. Moran, J. J. de Pablo, J. L. Skinner, M. T. Zanni, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 5796-5801; c) L. Wang, C. T. Middleton, S. Singh, A. S. Reddy, A. M. Woys, D. B. Strasfeld, P. Marek, D. P. Raleigh, J. J. de Pablo, M. T. Zanni, J. L. Skinner, *J. Am. Chem. Soc.* **2011**, *133*, 16062-16071; d) N. Demirdöven, C. M. Cheatum, H. S. Chung, M. Khalil, J. Knoester, A. Tokmakoff, *J. Am. Chem. Soc.* **2004**, *126*, 7981-7990; e) C. H. Londergan, J. Wang, P. H. Axelsen, R. M. Hochstrasser, *Biophys. J.* **2006**, *90*, 4672-4685.

[25] A. M. Alperstein, J. S. Ostrander, T. O. Zhang, M. T. Zanni, *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 6602-6607.

[26] Y. Wang, D. J. Patel, *J. Mol. Biol.* **1993**, *234*, 1171-1183.

[27] C. S. Peng, K. C. Jones, A. Tokmakoff, *J. Am. Chem. Soc.* **2011**, *133*, 15650-15660.

[28] J. P. Lomont, J. S. Ostrander, J.-J. Ho, M. K. Petti, M. T. Zanni, *J. Phys. Chem. B* **2017**, *121*, 8935-8945.

[29] C. Bardin, J. L. Leroy, *Nucleic Acids Res.* **2008**, *36*, 477-488.

[30] R. F. Burton, *Comp. Biochem. Physiol. A Comp. Physiol.* **1983**, *76*, 161-165.

[31] C. T. Middleton, L. E. Buchanan, E. B. Dunkelberger, M. T. Zanni, *J. Phys. Chem. Lett.* **2011**, *2*, 2357-2361.

[32] C. Lee, K.-H. Park, J.-A. Kim, S. Hahn, M. Cho, *J. Chem. Phys.* **2006**, *125*, 114510.

[33] D. B. Strasfeld, Y. L. Ling, R. Gupta, D. P. Raleigh, M. T. Zanni, *J. Phys. Chem. B* **2009**, *113*, 15679-15691.

[34] a) A. M. Woys, Y.-S. Lin, A. S. Reddy, W. Xiong, J. J. de Pablo, J. L. Skinner, M. T. Zanni, *J. Am. Chem. Soc.* **2010**, *132*, 2832-2838; b) H. T. Kratochvil, J. K. Carr, K. Matulef, A. W. Annen, H. Li, M. Maj, J. Ostmeyer, A. L. Serrano, H. Raghuraman, S. D. Moran, J. L. Skinner, E. Perozo, B. Roux, F. I. Valiyaveetil, M. T. Zanni, *Science* **2016**, *353*, 1040-1044.

[35] A. M. Woys, A. M. Almeida, L. Wang, C.-C. Chiu, M. McGovern, J. J. de Pablo, J. L. Skinner, S. H. Gellman, M. T. Zanni, *J. Am. Chem. Soc.* **2012**, *134*, 19118-19128.

[36] Q. Guo, P. Pagano, Y.-L. Li, A. Kohen, C. M. Cheatum, *J. Chem. Phys.* **2015**, *142*, 212427.

[37] a) S. Hahn, S. S. Kim, C. Lee, M. Cho, *J. Chem. Phys.* **2005**, *123*, 084905; b) A. Moran, S. Mukamel, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 506-510; c) A. V. Cunha, A. S. Bondarenko, T. L. Jansen, *J. Chem. Theory Comput.* **2016**, *12*, 3982-3992; d) C. Lee, M. Cho, *J. Chem. Phys.* **2007**, *126*, 145102.

[38] R. Hansel, F. Lohr, S. Foldynova-Trantirkova, E. Bamberg, L. Trantirek, V. Dotsch, *Nucleic Acids Res.* **2011**, *39*, 5768-5775.

[39] S. Kolesnikova, E. A. Curtis, *Molecules* **2019**, *24*, 3074.

[40] E. Largy, J. L. Mergny, V. Gabelica, *Metal ions in life sciences* **2016**, *16*, 203-258.

[41] P. Stevenson, C. Götz, C. R. Baiz, J. Akerboom, A. Tokmakoff, A. Vaziri, *J. Phys. Chem. B* **2015**, *119*, 5824-5831.

[42] A. M. Fleming, Y. Ding, R. A. Rogers, J. Zhu, J. Zhu, A. D. Burton, C. B. Carlisle, C. J. Burrows, *J. Am. Chem. Soc.* **2017**, *139*, 4682-4689.

[43] M. Paulite, Z. Fakhraai, I. T. S. Li, N. Gunari, A. E. Tanur, G. C. Walker, *J. Am. Chem. Soc.* **2011**, *133*, 7376-7383.

[44] a) C. R. Baiz, D. Schach, A. Tokmakoff, *Opt. Express* **2014**, *22*, 18724-18735; b) J. S. Ostrander, A. L. Serrano, A. Ghosh, M. T. Zanni, *ACS Photonics* **2016**, *3*, 1315-1323.

[45] C. T. Middleton, A. M. Woys, S. S. Mukherjee, M. T. Zanni, *Methods* **2010**, *52*, 12-22.