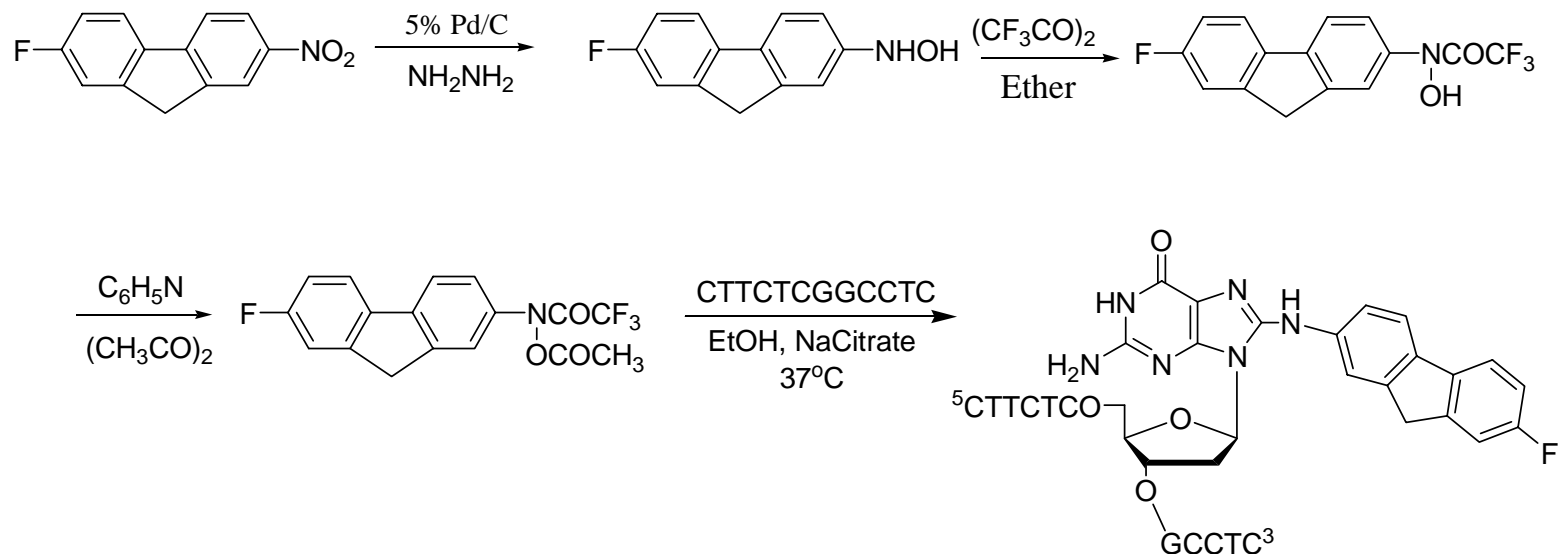# Supporting Information (Figures)

Spectrocopic and Theoretical Insights into the Sequence Effect on the Aminofluorene-induced Conformational Heterogeneity and *E. coli* Nucleotide Excision Repair

S. Meneni, S. M. Shell, Lan Gao, P. Jurecka, J. Sponer, W. Lee, Y. Zou, M. P. Chiarelli, and B. P. Cho*

## Fig. S1. Synthesis of FAF-modified 12-mer oligonucleotides

N-Acetoxy-N-2-(trifluoroacetylamino)-7-fluorofluorene was prepared starting from 2-fluoro-7-nitrofluorene according to the literature procedures. An unmodified 12-mer oligonucleotide (shown here -CG*G- sequence) was treated with the activated FAF derivative in ethanol and pH 6.5 sodium citrate buffer and incubated at 37 ºC for 36 hours. See Experimental section for details.

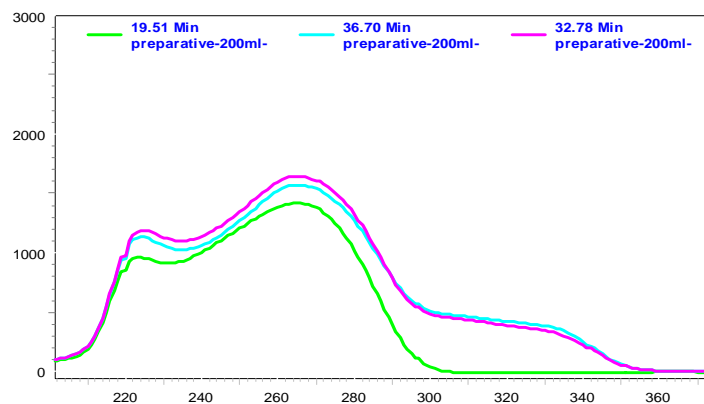# Fig. S2. HPLC profiles of FAF-modified 12-mer -CG*G- Seq

After ether workup, the aqueous layer was concentrated and purified by reversed phase preparative HPLC [Waters Xterra MS, 10x50mm, 2.5m, a 60-min gradient of 3-15% acetonitrile in 0.1M ammonium acetate, 2 mL/min]. The modified sequence were characterized by UV, and a standard enzymatic digestion analysis.
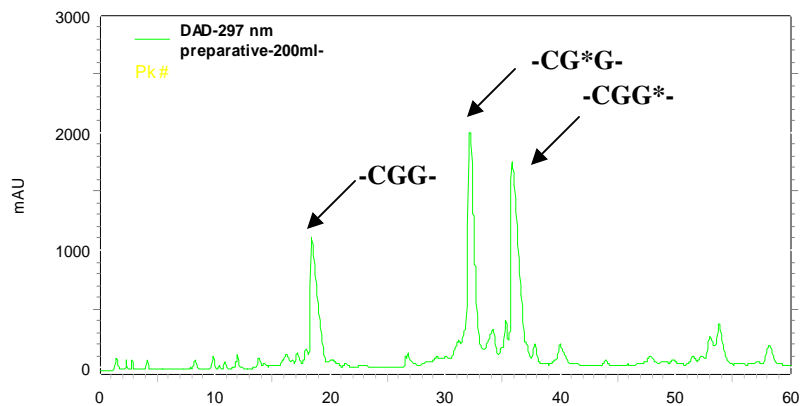
A. HPLC Chromatogram of the reaction mixture from seq. **-CG*G-** and the activated FAF-ester.
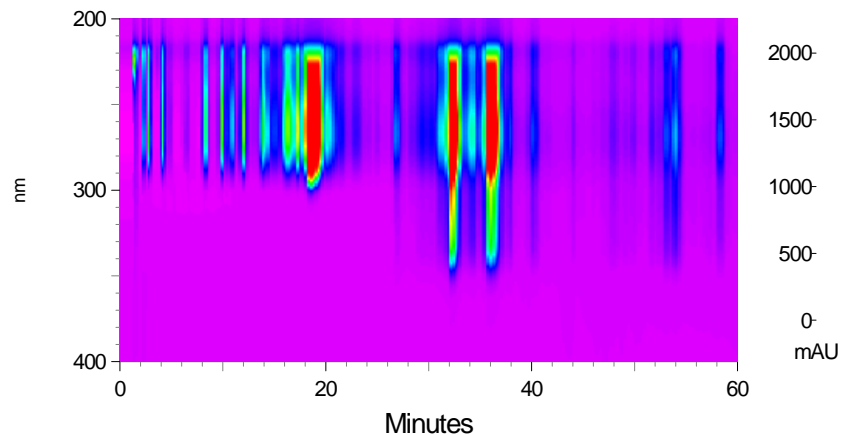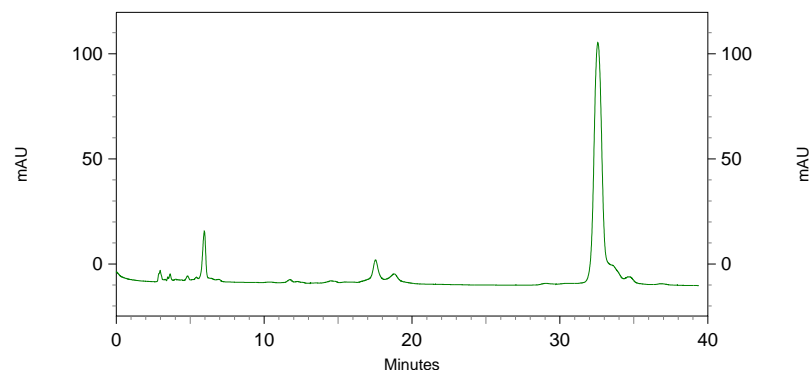

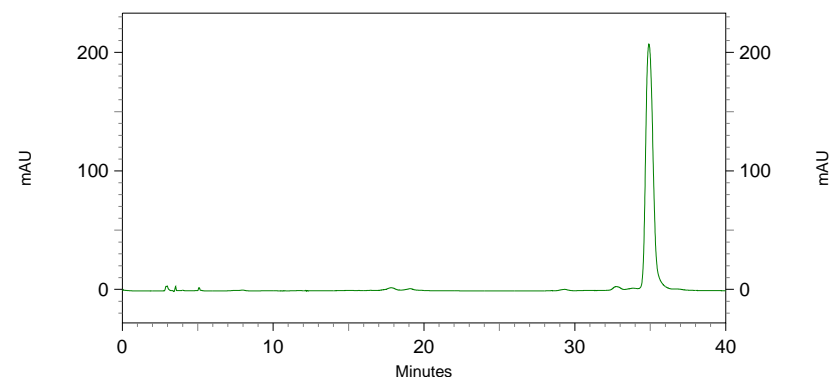**B.** Contour chromatogram of the reaction mixture of seq. **-CG*G** -

C. UV absorption spectra of the modified seq. **-CG*G** and the control

Fig. S3. HPLC profiles of purified FAF-modified 12-mer **-CG*G Seq**

A. HPLC chromatogram of purified seq. **-CG*G-** .

B. HPLC chromatogram of purified seq. **-CGG*-**

dC, dT, dA

dG-C8-FAF
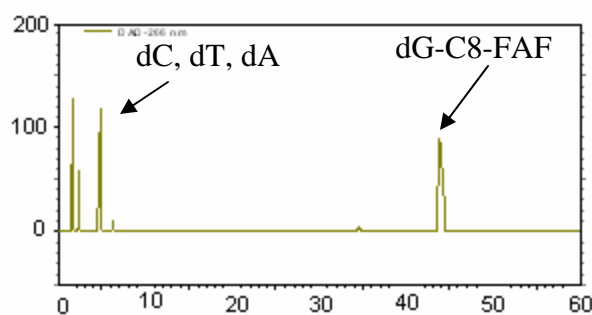
C. HPLC chromatogram of an enzyme digestion mixture from FAF-modified seq. **-CG*G-** . HPLC system with a gradient system involving 2 % (10 min), 5% (20 min), and 5 - 40% acetonitrile in pH 7.0 ammonium acetate buffer (10 mM) with a flow rate of 2.0 mL/min.

Sequence Analyses by Enzyme Digest/LC/MS

The modified ODNs were sequenced using a Waters ESI-mass spectrometer (LC-TOF-MS) in the negative ion mode based on the exonuclease (3'-5', or 5'-3') strategies described previously for identifying the position of bases modified by arylamines (1-3). Briefly, 5-10 mg of AF-modified ODNs were combined with ca. 0.01 units of a 3'- 5'- or a 5'- 3'-exonuclease in a 1 mM solution of $MgCl_2$ and incubated for different times up to 2 hours. Enzyme digests were separated online using a ten-minute water/acetonitrile gradient. The aqueous phase was 5 mM in both ammonium acetate and dimethylbutylamine (pH 5.6) and the acetonitrile was made 0.1% in formic acid. The spray voltage was 3.4 kV.

The ESI-TOF mass spectrometry analysis of each FAF-modified ODN was carried out to determine the position of the FAF-modification. Molecular weight analyses indicated that all modified ODNs were modified by a single FAF moiety (**Table S1**). **Fig. S4** shows a representative molecular weight spectrum of the early eluting FAF-modified –CG*G- sequence. Digest fragments were described using the system of nomenclature proposed by McLuckey and Habibigoudarzi (2). The spectrum did not show any singly charged molecular ions, but rather exhibited the expected multiple charged ions upon electrospray with the triply charged species as a base peak. Other sequences have been characterized similarly and the results are summarized in **Table S1**.

Each of the ODNs that contained more than one guanine (-AGG- and -CGG- sequences) was digested with both a 3'-5' and a 5'-3' exonuclease to determine the position of the modified guanine in the oligonucleotide chain, as described previously for modified ODNs containing only a single guanine (3). The LC/MS analyses of the digests of the two modified fractions in **Fig. S2** are described here to illustrate how the sites of modification were determined. As shown in **Fig. S1** and **S2**, reaction of the ODN 5'-CTTCTCG7G8CCTC-3' with activated FAF produced two hydrophobic peaks that eluted at approximately 33 and 37 min). Mass spectrometry analyses of these two peaks indicated that the structures of the two molecules generating these peaks both contained a single FAF modification and differed only in the position of the guanine base modified by the FAF.

The LC-MS analyses of all digests showed a mixture of ODNs that yielded primarily $(M-2H)^{2-}$ ions upon electrospray. **Fig. S6** shows ESI-MS spectra of some of the ODN fragments generated by 3'-5' exonuclease digestion of the early eluting fraction. **Fig. S6A** shows several ions that are formed by the sequential elimination of unmodified bases near the 3'-terminus. The ion at m/z 1120.7 (labeled B7) is indicative of the position of the unmodified guanine. The m/z value of this ion is consistent with the heptaODN, 5'-CTTCTCG(FAF)-3', formed by the loss of a guanine residue from the octaODN, 5'-CTTCTCG(FAF)G-3', that forms doubly- and triply-charged ions at m/z 1,285.7 and m/z 856.8 (labeled B8), respectively. The ion at m/z 857.7 in **Fig. S6B** (labeled B6) is formed by the loss of the FAF-modifed guanine from the B7 fragment in **Fig. S6A**.

The LC-mass spectra of several of the ODN fragments generated in the 5'-3' hydrolysis of the late eluting peak are shown in **Fig. S7**. The observation of Y5 and Y6 ions at m/z 981.2 and m/z 816.7 is consistent with (M-2H)2- ions formed from 5'-GG(FAF)CCTC-3' (Y6) and 5'-G(FAF)CCTC-3' (Y5), respectively. The observation of these two ions indicate that G7 (the guanine farthest from the 3' terminus) is unmodified.

References

1. Brown, K., Harvey, C.A., Turteltaub, K.W., Shields, S.J. (2003) Structural characterization of carcinogen-modified oligodeoxynucleotide adducts using matrix-assisted laser desorption/ionization mass spectrometry *J. Mass Spectrom*, . **38**, 68–79.
2. McLuckey, S.A. and Habibigoudarzi, S. (1993) Decompositions of multiply charged oligonucleotide anions *J. Am. Chem. Soc*, . **115**, 12085–12095 .
3. Wu, H. and Aboleneen, H. (2000) Sequencing oligonucleotides with blocked termini using exonuclease digestion and electrospray mass spectrometry *Anal. Biochem*, **287**, 126–135.
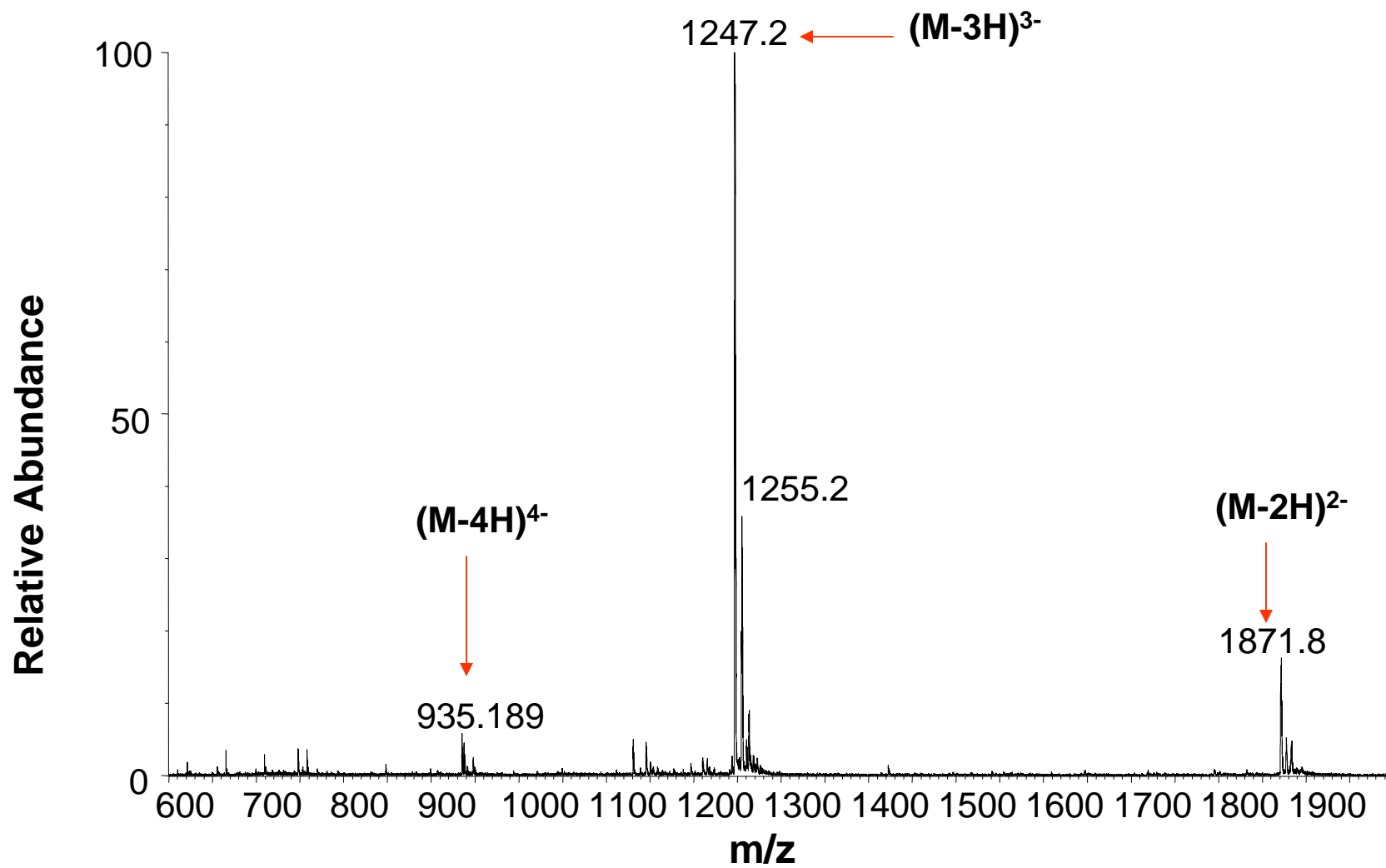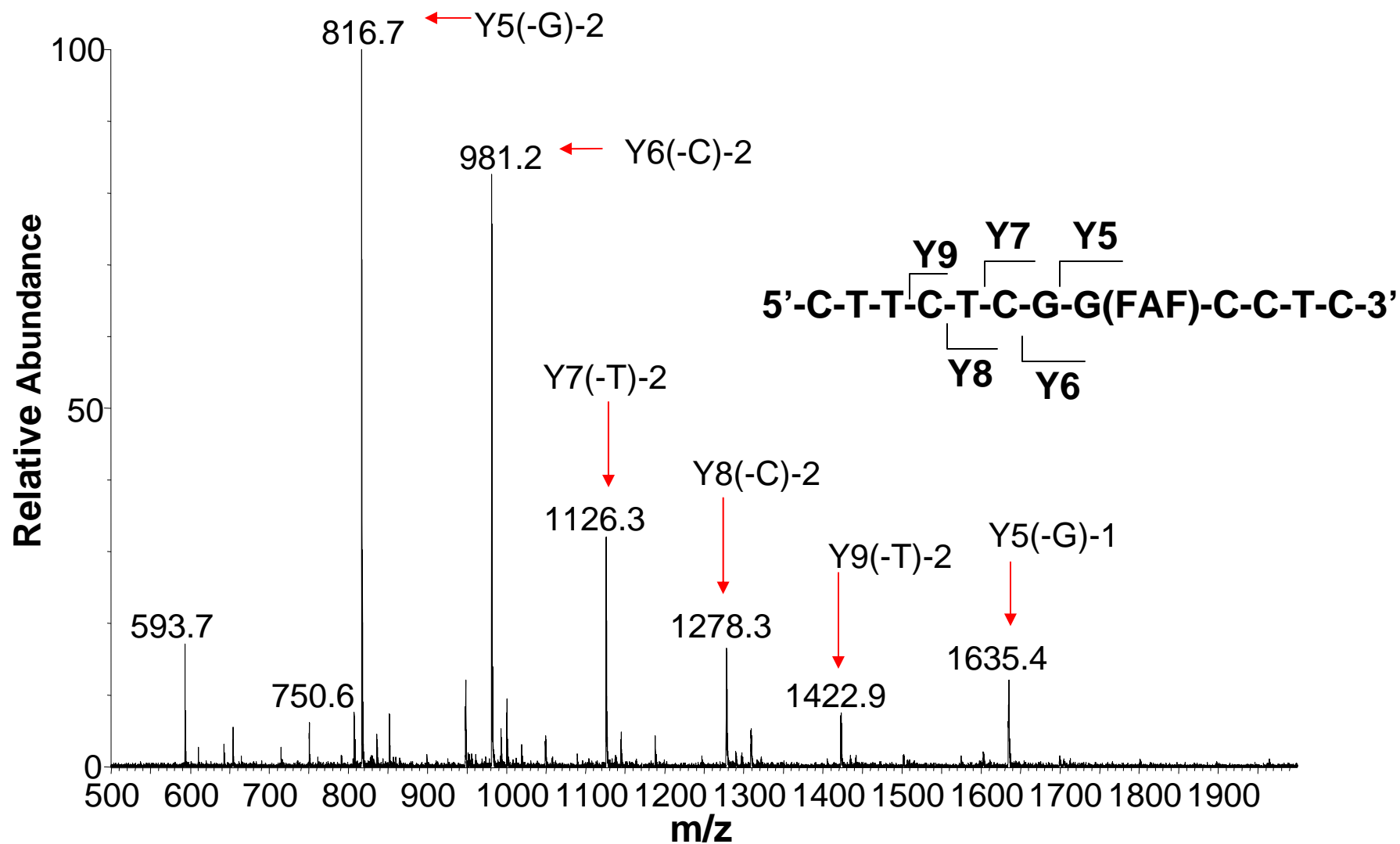
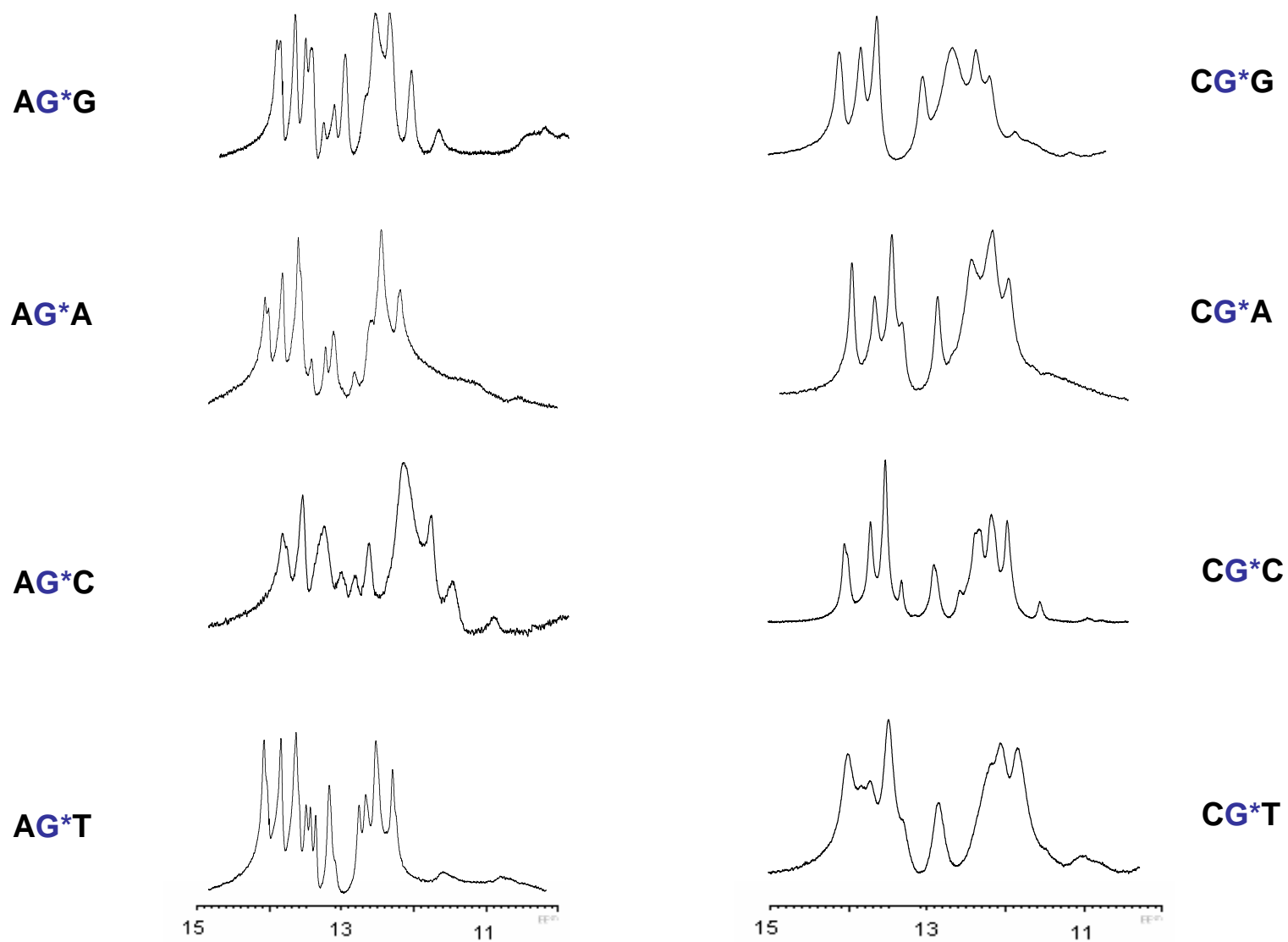**Fig. S4**: TOF-LC/MS molecular weight spectrum of the FAF-modified-CG*G- sequence.
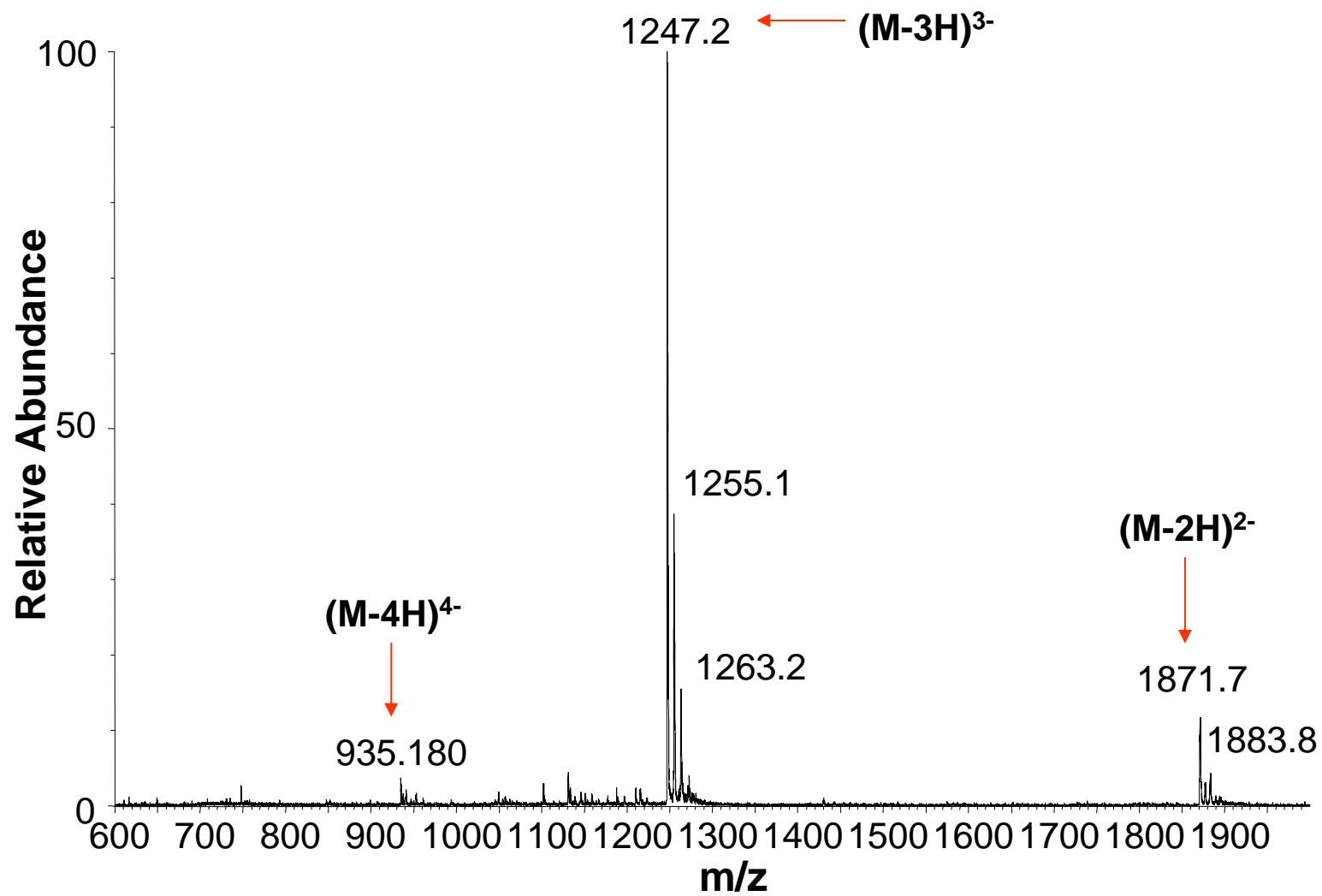
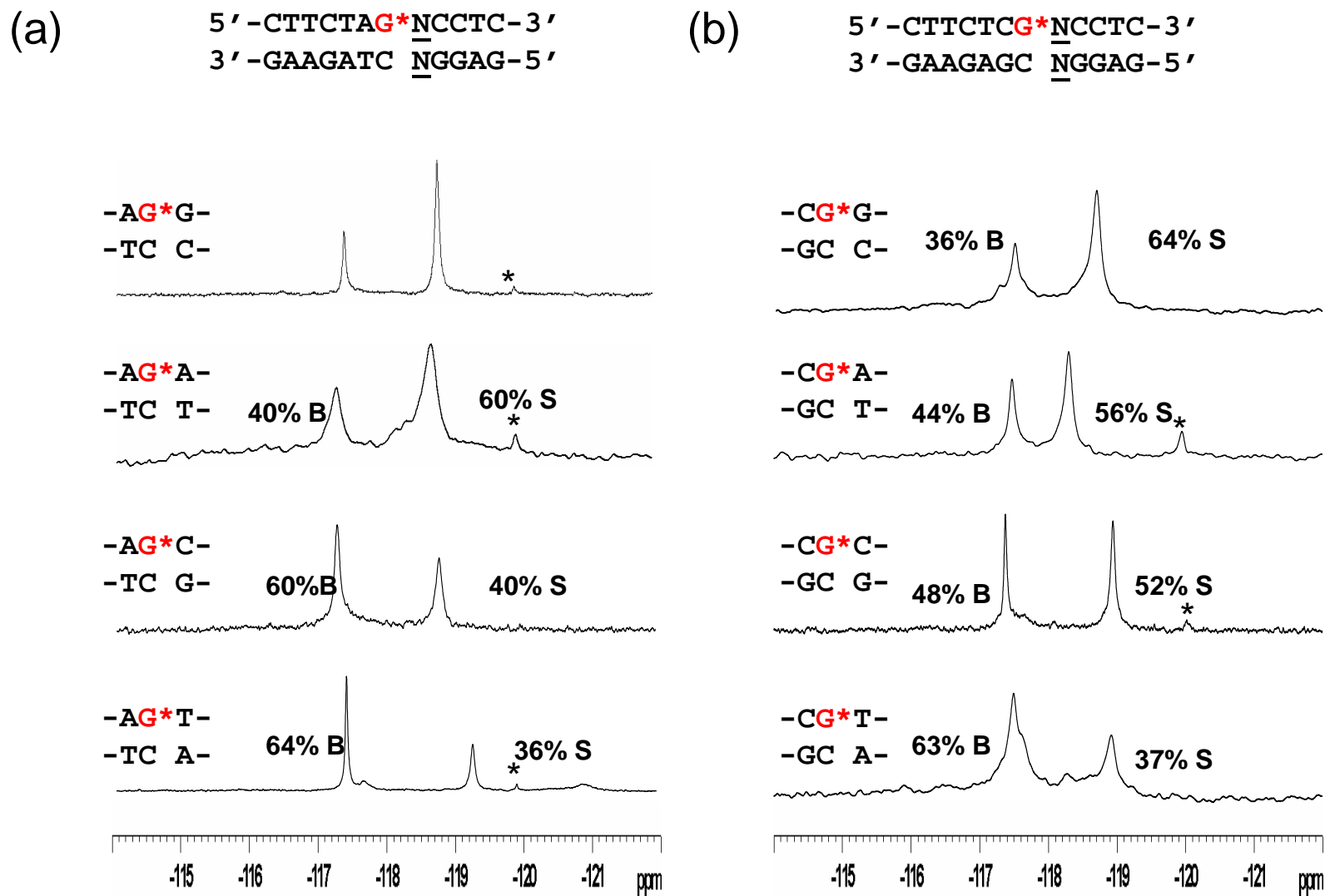**Fig. S5**: TOF-LC/MS molecular weight spectrum of the FAF-modified -CGG*- sequence.
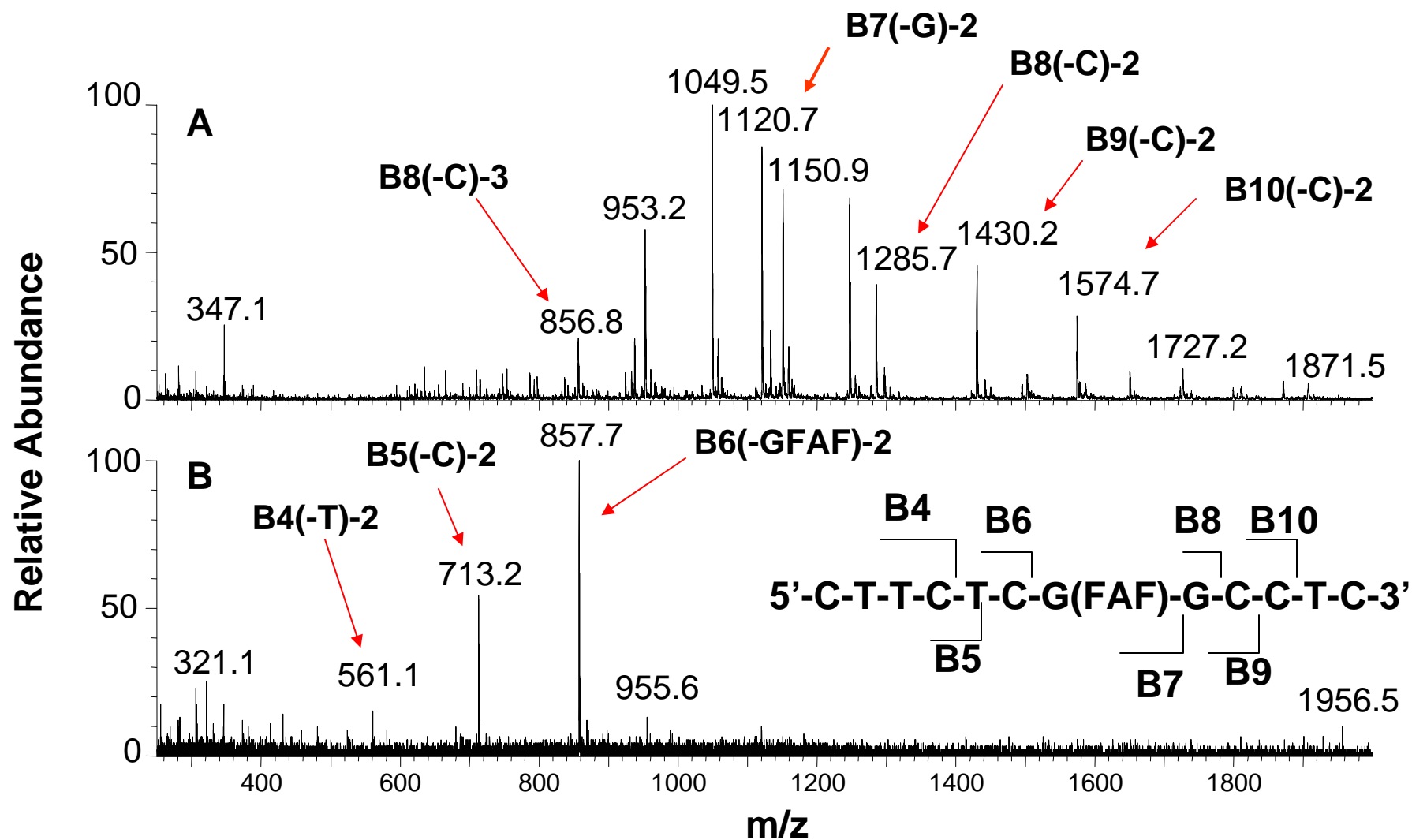
Fig. S6: TOF-LC/MS spectrum of the FAF-modified -CG*G- oligonucleotide digest carried out with 3'-5'-exonuclease.

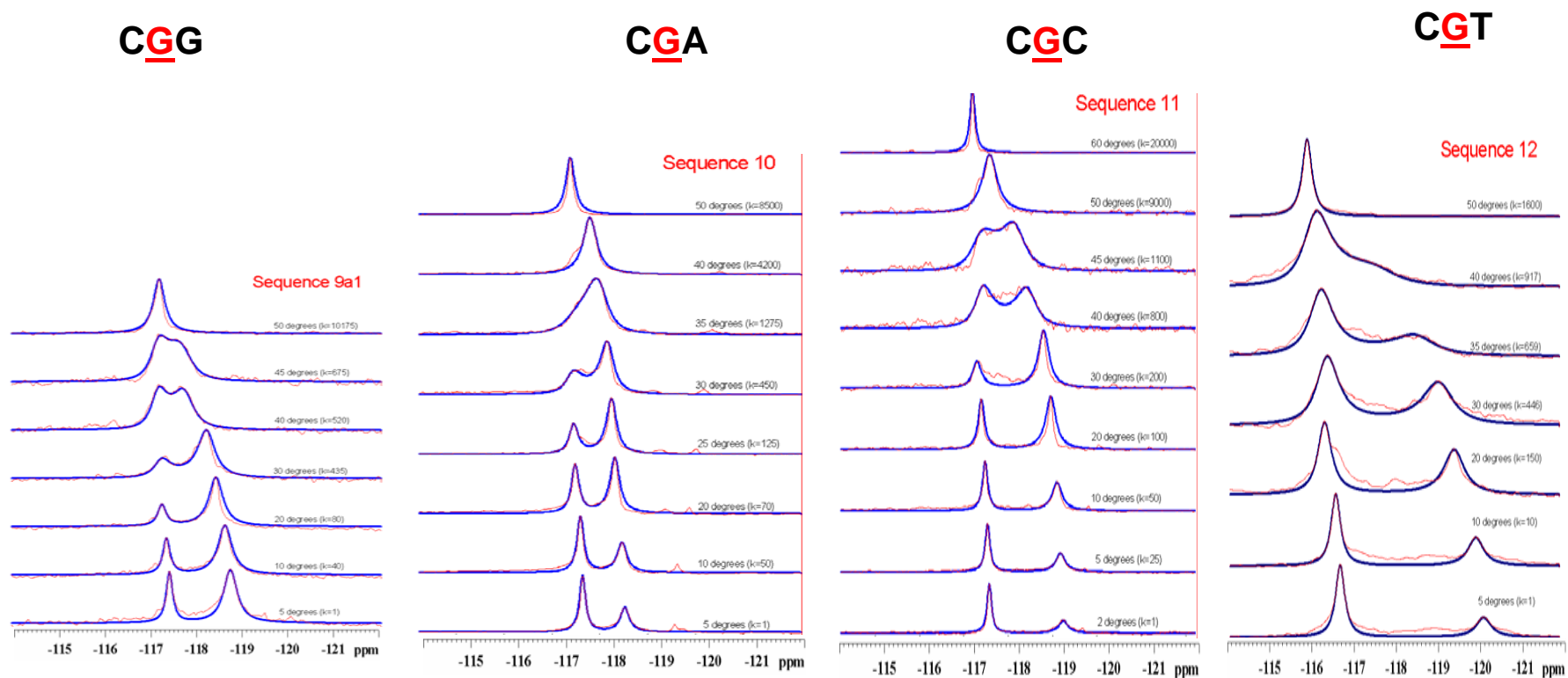**Fig. S7**: TOF-LC/MS spectrum of the FAF-modified -CGG*- oligonucleotide digest carried out with 5'-3'-exonuclease.

Fig. S8: Imino proton spectra of the -AG*N- and -CG*N- duplex series

Fig. S9: ^{19}F NMR spectra for (a) the -AG*N- and (b) -CG*N- duplexes measured in the -115 ~ -122 ppm range at 20 °C (G* = FAF-adduct. N = G, A, C, T). Percent populations of the B-type and S-conformers are indicated. *=impurity

Fig. S10: Examples of simulated dynamic NMR spectra for the -CG*N- duplex series
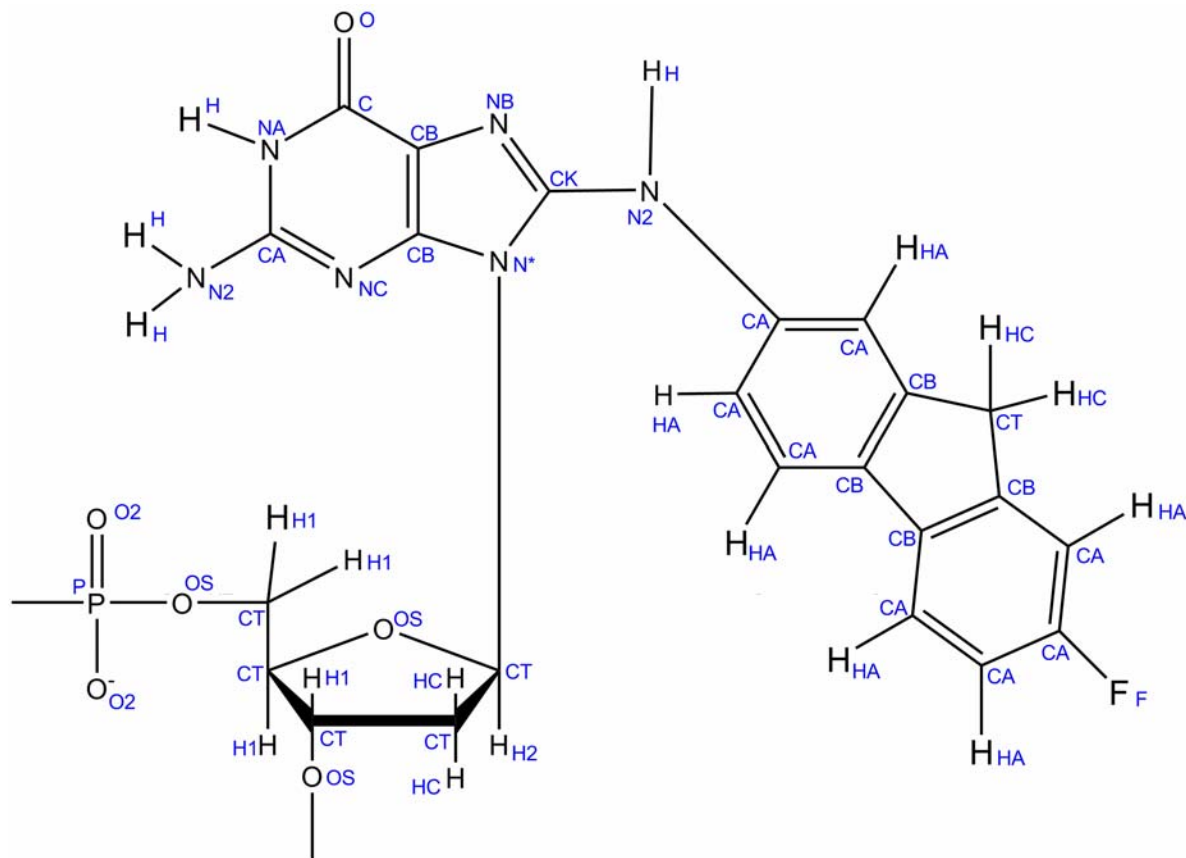
Fig. S11. The atom type assignment for FAF-dG.

Computations were carried out with the AMBER 8.0 suite of programs (UCSF, San Francisco, CA), the Cornell et al. force field (*J. Am. Chem. Soc.* **1995** 117:5179) and the PARM 99 parameter set (Cheatham et al., *J.Biomol. Struct.& Dyn., 16*, 845-862 (1999). The force field was parametrized for the FAF-dG consistently with the rest of the force field. Partial charges for FAF-dG were obtained as described by Cieplak et al. (*J. Comp. Chem.* **1995** 16:1357). HF calculations with the 6-31G* basis set (Hehre et al., *J. Chem. Phys.* **1972** 56:2257) were used to calculate the electrostatic potential using Gaussian 03 (Gaussian, Inc., Wallingford, CT.), and the restrained electrostatic potential (RESP) fitting algorithm (Bayly et al., *J. Phys. Chem.* **1993** 97:10269; Zhang et al., *Biophys. J.* **2006** 90:1865) was employed to fit the charge to each atom center. Partial charges were separately computed for S and B conformers. Missing bond length and angle equilibrium values, force constant parameters for bond lengths, angles, and dihedral angles were assigned by analogy with chemically similar atom types present in the GAFF (General Amber Force Field) parameter set (Wang et al, *J. Comp. Chem.* **2004** 25:1157) or the AMBER PARM99 force field.
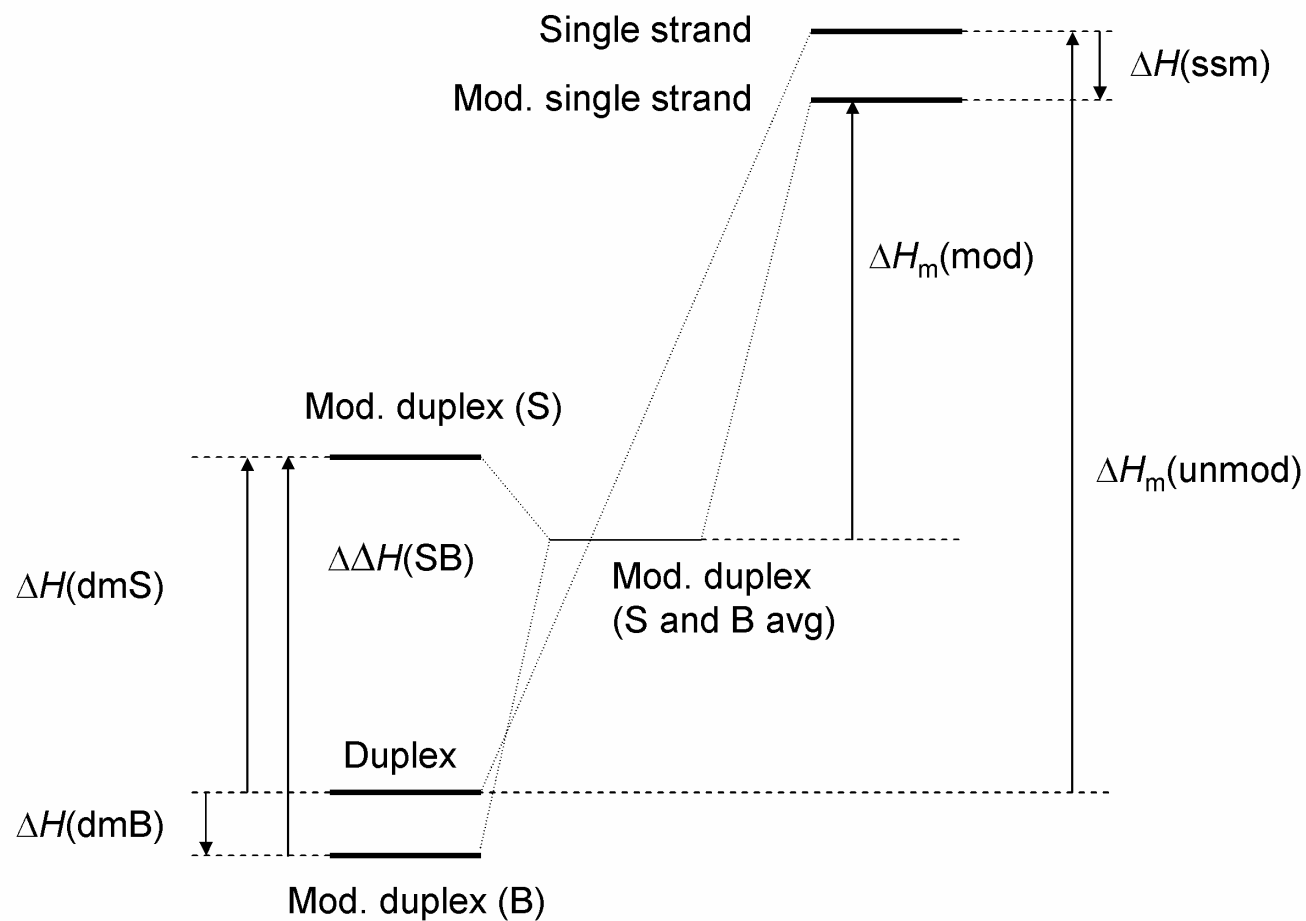
Fig. S12. Assumed ordering of the unmodified and FAF-modified duplexes and the corresponding single strands on the enthalpy scale: mod (modified); unmod (unmodified), dmS (modified S-conformer), dmB (modified B-conformer), ssm (modified single strand).

Fig. S12 continue…

Thermodynamics of the intercalated duplexes.
Fig. S12 shows the assumed ordering of the duplex, modified duplex, single strands and modified single strands on the enthalpy scale. In this scheme, the ΔHm(mod) and ΔHm(unmod) are enthalpies of melting and correspond to the values in Table 1 column 3 (unmodified values in parentheses) in the article. Since the FAF-modified duplex exists in two distinct forms, S and B, the enthalpy of its melting is a difference between enthalpy of single strands and enthalpy of the mixture of S and B: ΔHm(mod) = H(ss) – (XSH(S) + XBH(B)), where XS and XB are molar fractions of S and B forms at the conditions of melting. ΔH(ssm) and ΔH(dmB) are enthalpy changes connected with FAF modification of the single strand and duplex. These could be approximated by solvation enthalpy of attached FAF, which will roughly cancel out for single strand (melted state) and duplex in the case of the B conformer (solvation of FAF is likely to be the same both in B and in the melted single strand). Finally, ΔΔH(S-B) is the enthalpy difference between S and B conformers of the modified duplex and was approximated by the corresponding energy difference ΔΔEint(S-B) corrected for the solvation energy difference, ΔΔEsolv(S-B) (see Table 3 in the text).

Thermodynamics of melting.
UV-melting experiments showed that, in terms of Gibbs energy, the FAF modification destabilizes duplexes by 1.8 ~ 3.3 kcal/mol (Table 1 in the article). This thermodynamic destabilization is of an enthalpic nature as ΔΔH is always positive and greater (in absolute value) than ΔΔG. It is, however, largely cancelled by entropy effects. Unfortunately, the ΔΔH of melting can hardly be evaluated directly as a difference between the duplex and the melted single strands, since the structure of the latter is likely to be fluctuating and is not very well defined. Nevertheless, a closer look at Figure 6 reveals that the DDH of melting could be linked with the calculated energy difference between S and B conformers, ΔΔEtot(S-B). Under the experimental conditions both S and B conformers coexist in the system, therefore the resulting DDH of melting must be a weighted average of both S and B enthalpy contributions. If we assume that the hydration enthalpy of the FAF moiety is similar both in the melted single strand and in the B duplex (see also Figure 7), this average can be approximated as ΔΔH = XSΔΔH(S-B), where XS is molar fraction of S at the conditions of melting. In this formula we will use %S/100 (as measured by NMR at 20°C) as an estimate of XS for simplicity (XS at the melting temperature should be slightly larger). Now, if we assume that energy constitutes the main part of enthalpy, we can use ΔΔEtot(S-B) (Table 3 in the text) instead of ΔΔH(S-B) in the previous expression. The result shown in the last column of Table 3 (ΔΔE(S-B) weighted) represents our theoretical estimate of the ΔΔH destabilization of duplex upon FAF modification.