Supplementary Material for PCCP

Zwitterionic i-motif structures are preserved in DNA negatively charged ions produced by electrospray mass spectrometry

Frederic Rosu[1*], Valérie Gabelica[1], Laure Joly[1], Gilles Grégoire[2], Edwin De Pauw[1]

1 Mass Spectrometry Laboratory, Chemistry Institute (B6c), University of Liège, B-4000 Liège, Belgium

2 Laboratoire de Physique des Lasers, CNRS, UMR 7538, Institut Galilée, Université Paris 13, F-93430 Villetaneuse, France

* Corresponding author. E-mail: f.rosu@ulg.ac.be

Supporting information table of contents

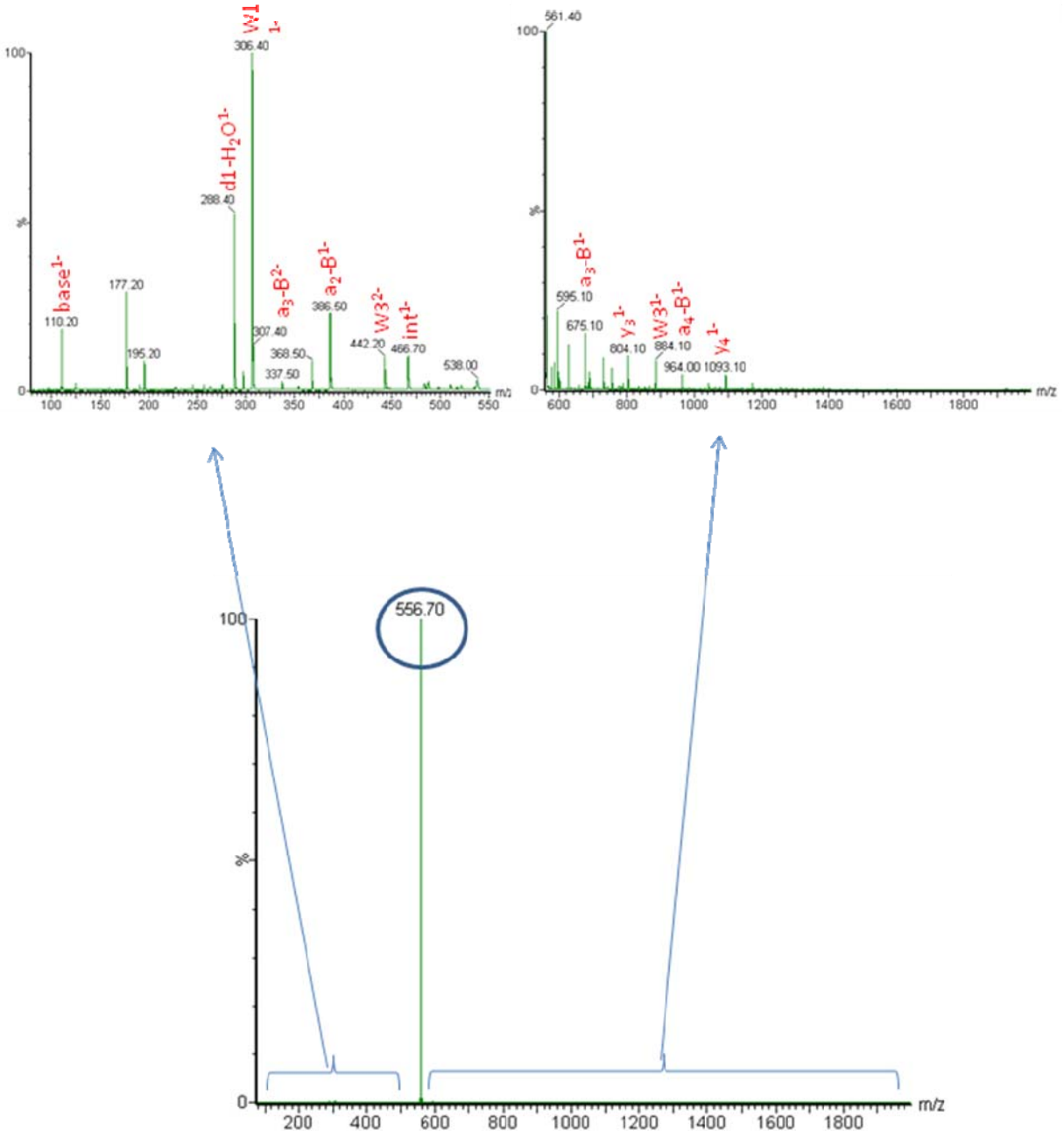

Figure S1: Example of product ion spectrum resulting from IRMPD: MS/MS spectrum of the $(dC_6)^{3-}$ ion irradiated at 1650 cm^{-1}. Parent ion is circled in blue. "w_n" and "a_n-B" fragment ions are labeled according to the nomenclature by McLuckey [see: Wu, J.; McLuckey S. A. *Int. J. Mass Spectrom.* (**2004**), 237, 197-241]. "int" indicates an internal fragment. "-Base" indicates a neutral base loss, and the nature of the base is indicated in parentheses.

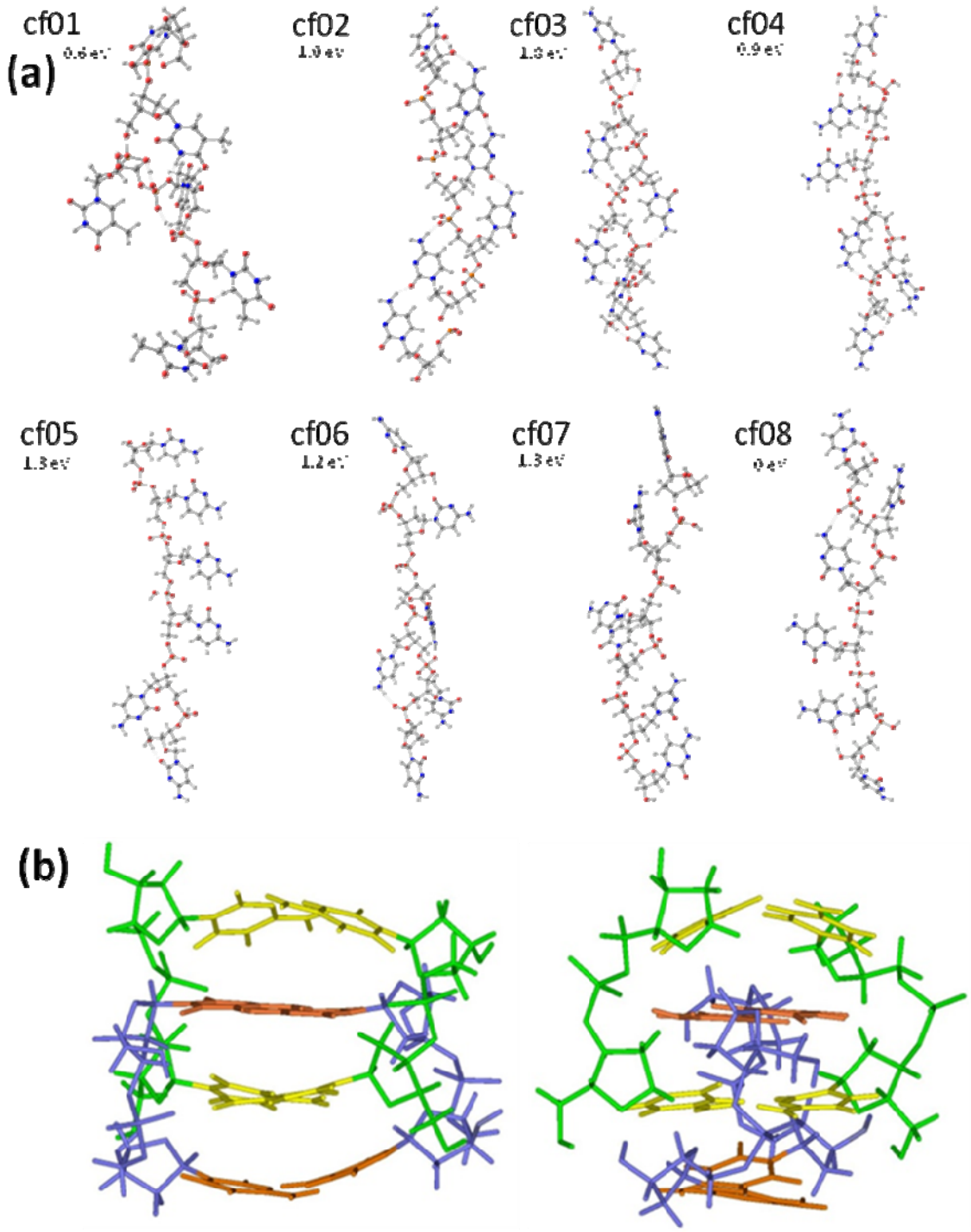

Figure S2: (a) Structure of 6-mer oligonucleotide d(CCCCCC)$^{3-}$ obtained after 50 ns MD using amber parm99 followed by geometry optimization using DFT B3LYP at the 3-21G* level. (b) Optimized structure of the model i-motif used for the calculation. This complex structure has four stacked C-H^{+}-C base pairs

(in yellow and brown). The connecting backbone is made of eight sugars and four phosphate groups (green and blue). The charge state is zero.

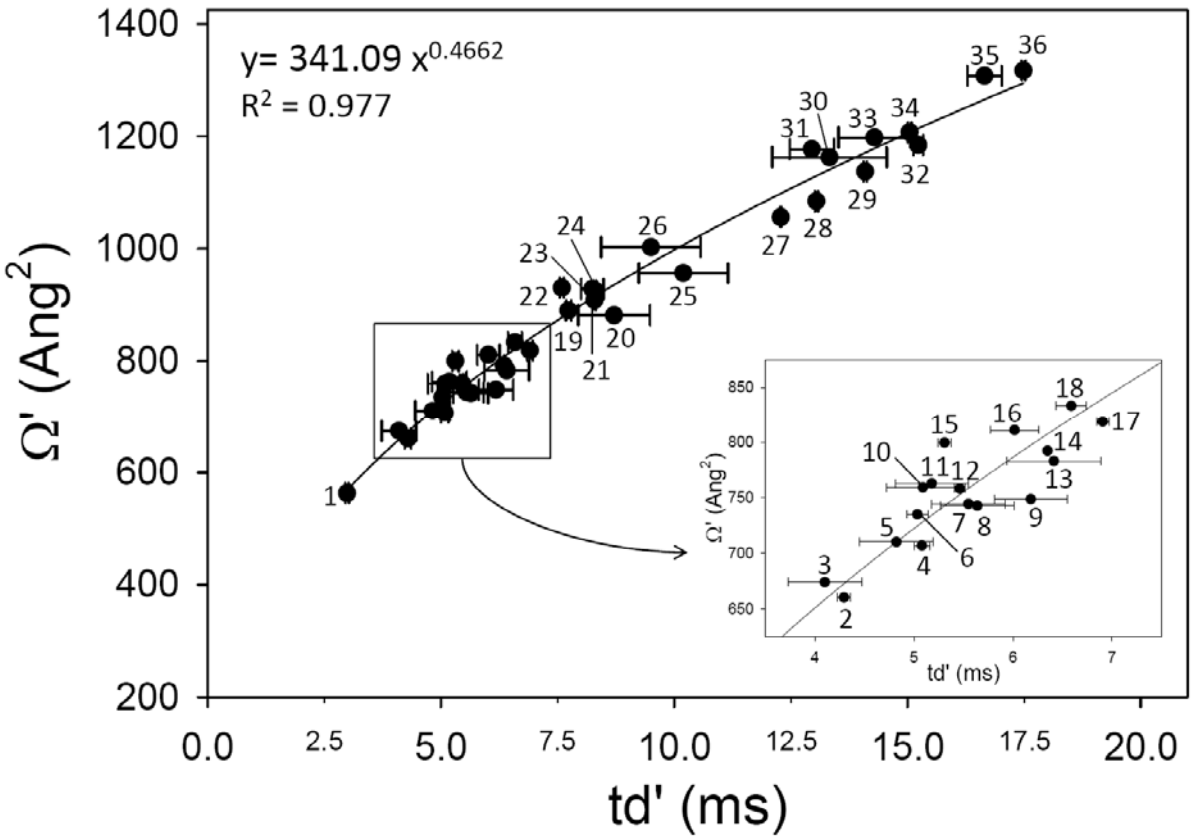

Figure S3: Calibration curve to convert arrival time into collision cross section. The effective drift time t'_D (in ms) is calculated by $t'_D = t_D - \frac{C\sqrt{\frac{m}{z}}}{1000}$, where m/z is the mass-to-charge ratio of the observed ion and C is equal to 1.41 in our instrument. Published cross sections (Ω in Å^2) are converted to Ω', which is proportional to the mobility of the ion, by equation $\Omega' = \frac{\Omega}{z}\sqrt{\mu}$, where z is the charge, μ the reduced mass of the ion (m_{gas} = 28 for nitrogen). The effective drift time is plotted against corrected collision cross section and fitted with a power relationship according to $\Omega' = A \cdot (t'_D)^B$. The list of used calibrants is given in Supporting Information Table S1, next page.

Table S1: list of reference ions used for the negative ion mode calibration of the collision cross sections in our Synapt HDMS.

Peak annotation if Fig. S3	Molecule	Reference	Charge state	Mobility peak[a]
1	dT_{10}	[1]	6-	1
2	dT_{10}	[1]	5-	1
3	$[d(CG)_{11}]_2$	[2]	11-	1/2
4	dT_{10}	[1]	4-	1
5	dGAGGGTGGGGAGGGTGGGGAAG	[3]	6-	1/2
6	dGAGGGTGGGGAGGGTGGGGAAG	[3]	5-	1/2
7	$(d(CG)_9)_2$	[4]	9-	1
8	$[(d(CG)_9)_2 + NH_4^+]$	[4]	9-	1
9	$(d(CG)_{11})_2$	[2]	11-	2/2
10	$(d(CG)_7)_2$	[4]	7-	1
11	$[(d(CG)_7)_2 + NH_4^+]$	[4]	7-	1
12	$dGGG(TTAGGG)_3$	[3]	5-	1
13	dT_{10}	[1]	3-	1
14	dGAGGGTGGGGAGGGTGGGGAAG	[3]	6-	2/2
15	dGAGGGTGGGGAGGGTGGGGAAG	[3]	5-	2/2
16	$d(TTAGGG)_6$	[3,5]	7-	1/2
17	$[(dTG_4T)_4 + 3NH_4^+]$	[5]	5-	1
18	$d(TTAGGG)_4$	[5]	5-	1
19	$d(TTAGGG)_6$	[3,5]	6-	1
20	cytochrome C (horse)	[6]	8-	1/2
21	$dGGG(TTAGGG)_3$	[3]	4-	1
22	$d(TTAGGG)_6$	[3,5]	7-	2/2
23	$d(TTAGGG)_6$	[3,5]	8-	1
24	dGAGGGTGGGGAGGGTGGGGAAG	[3]	4-	1
25	cytochrome C (horse)	[6]	7-	1/2
26	cytochrome C (horse)	[6]	9-	1/2
27	apomyoglobin (sperm whale)	[6]	17-	1
28	apomyoglobin (sperm whale)	[6]	16-	1
29	apomyoglobin (sperm whale)	[6]	15-	1
30	cytochrome C (horse)	[6]	8-	2/2
31	dT_{10}	[1]	2-	1
32	apomyoglobin (sperm whale)	[6]	14-	1
33	cytochrome C (horse)	[6]	7-	2/2
34	apomyoglobin (sperm whale)	[6]	13-	1
35	apomyoglobin (sperm whale)	[6]	12-	1
36	cytochrome C (horse)	[6]	5-	1

[a] Indicates whether there was a single peak in the arrival time distribution (1), or two peaks. If there were two peaks, indicates if this is the peak with the lowest collision cross section (1/2) or the highest collision cross section (2/2).

Reference List for Table S1.

1. Hoaglund, C. S.; Liu, Y.; Ellington, A. D.; Pagel, M.; Clemmer, D. E. Gas-phase DNA: oligothymidine ion conformers. *J. Am. Chem. Soc.* **1997,** *119*, 9051-9052.

2. Baker, E. S.; Bowers, M. T. B-DNA helix stability in a solvent-free environment. *J. Am. Soc. Mass Spectrom.* **2007,** *18*, 1188-1195.

3. Gabelica, V.; Baker, E. S.; Teulade-Fichou, M.-P.; De Pauw, E.; Bowers, M. T. Stabilization and structure of telomeric and c-myc region intramolecular G-quadruplexes: The role of central cations and small planar ligands. *J. Am. Chem. Soc.* **2007,** *129*, 895-904.

4. Gidden, J.; Ferzoco, A.; Baker, E. S.; Bowers, M. T. Duplex formation and the onset of helicity in poly d(CG)$_n$ oligonucleotides in a solvent-free environment. *J. Am. Chem. Soc.* **2004,** *126*, 15132-15140.

5. Baker, E. S.; Bernstein, S. L.; Gabelica, V.; De Pauw, E.; Bowers, M. T. G-quadruplexes in telomeric repeats are conserved in a solvent-free environment. *Int. J. Mass Spectrom.* **2006,** *253*, 225-237.

6. Liu, Y.; Valentine, S. J.; Clemmer, D. E. (unpublished results), http://www.indiana.edu/~clemmer/Research/cross%20section%20database/Proteins/protein_cs.htm .

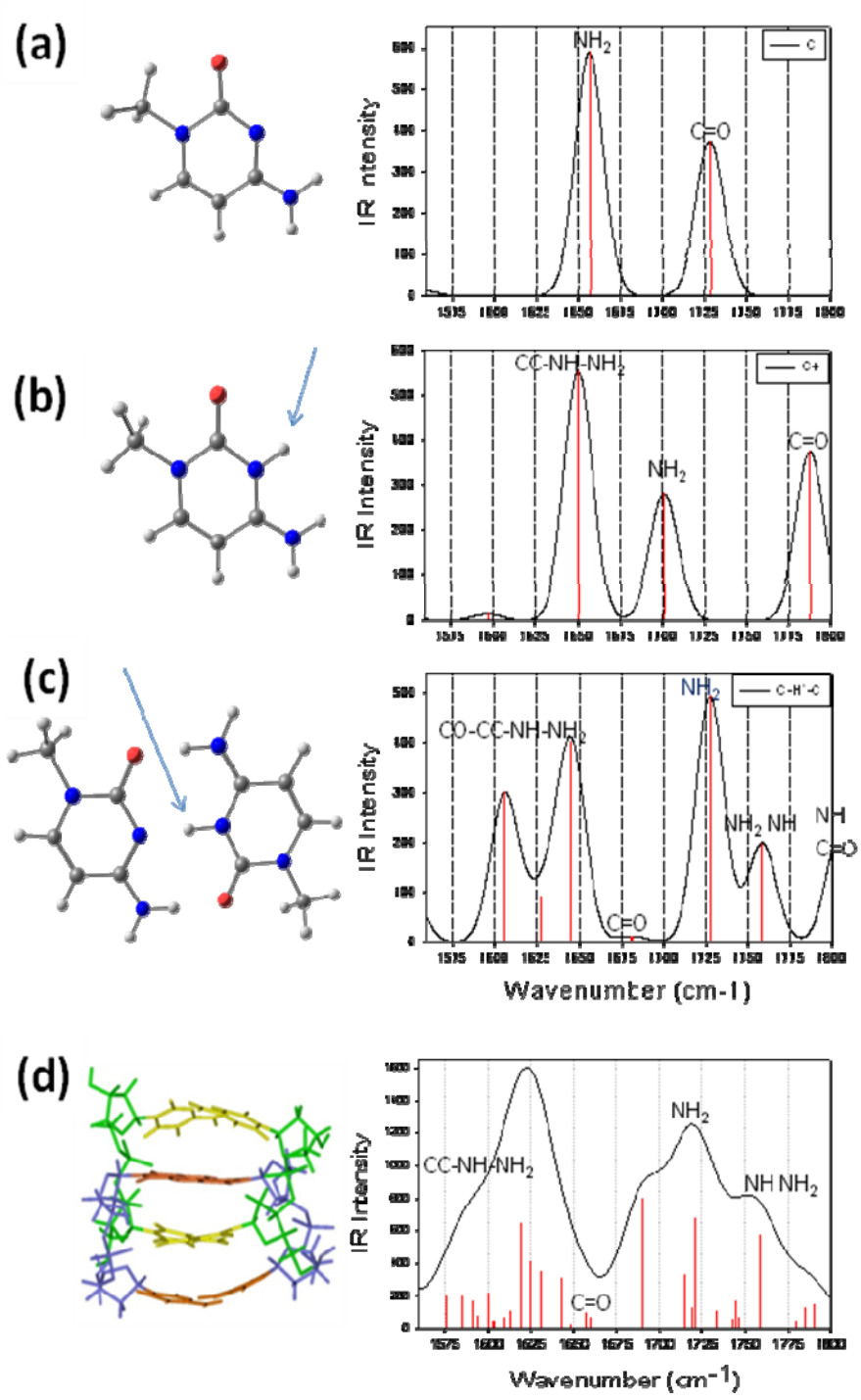

Figure S4: Structure and calculated IR spectrum of (a) methylcytosine (methylation in the N1 position) (b) the protonated (N3 position) methylcytosine C-H^+ (c) the protonated base pair C-H^+-C forming the i-motif (d) the i-motif (d(C2)4)) used to model larger system. The methylation in N1 position is to mimic

the sugar position. Calculation have been performed using DFT B3LYP at the 3-21G* level. The scaling factor is 0.985. The arrows show the protonation sites.

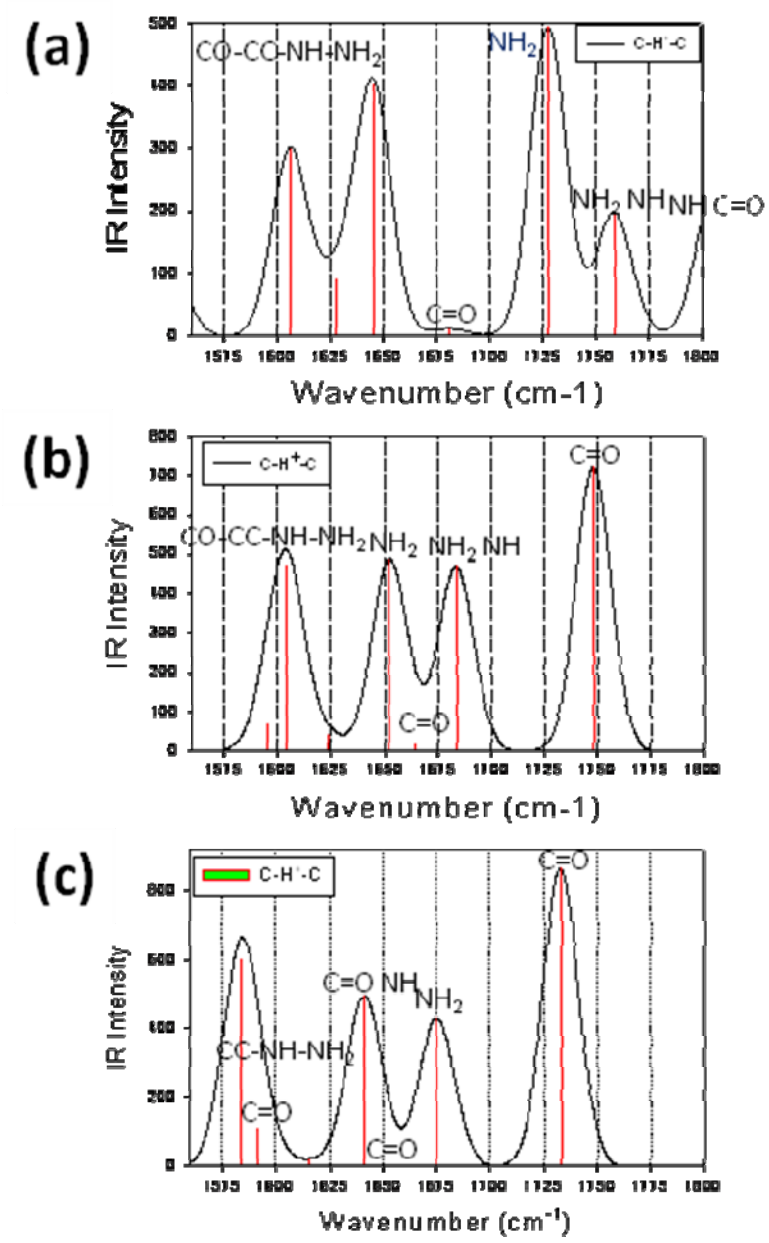

Figure S5: Influence of the basis set used for the calculation of the IR spectrum of the protonated base pair C-H$^+$-C forming the i-motif (a) 3-21G*, (b) 6-31G*, (c) 6-31+G**.

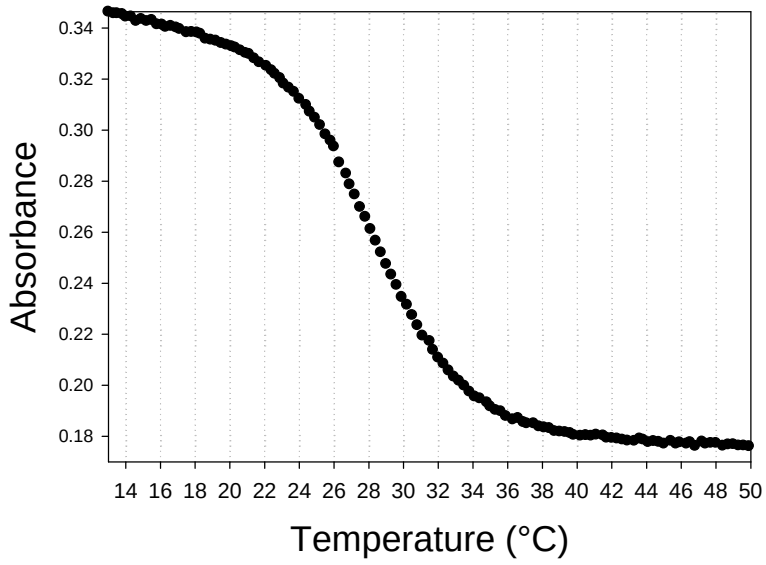

Figure S6: Thermal denaturation followed at 295 nm of the human intramolecular i-motif (d(CCCAAT)$_3$CCC) in 150 mM NH_4Ac. The temperature gradient was 0.1°C/min

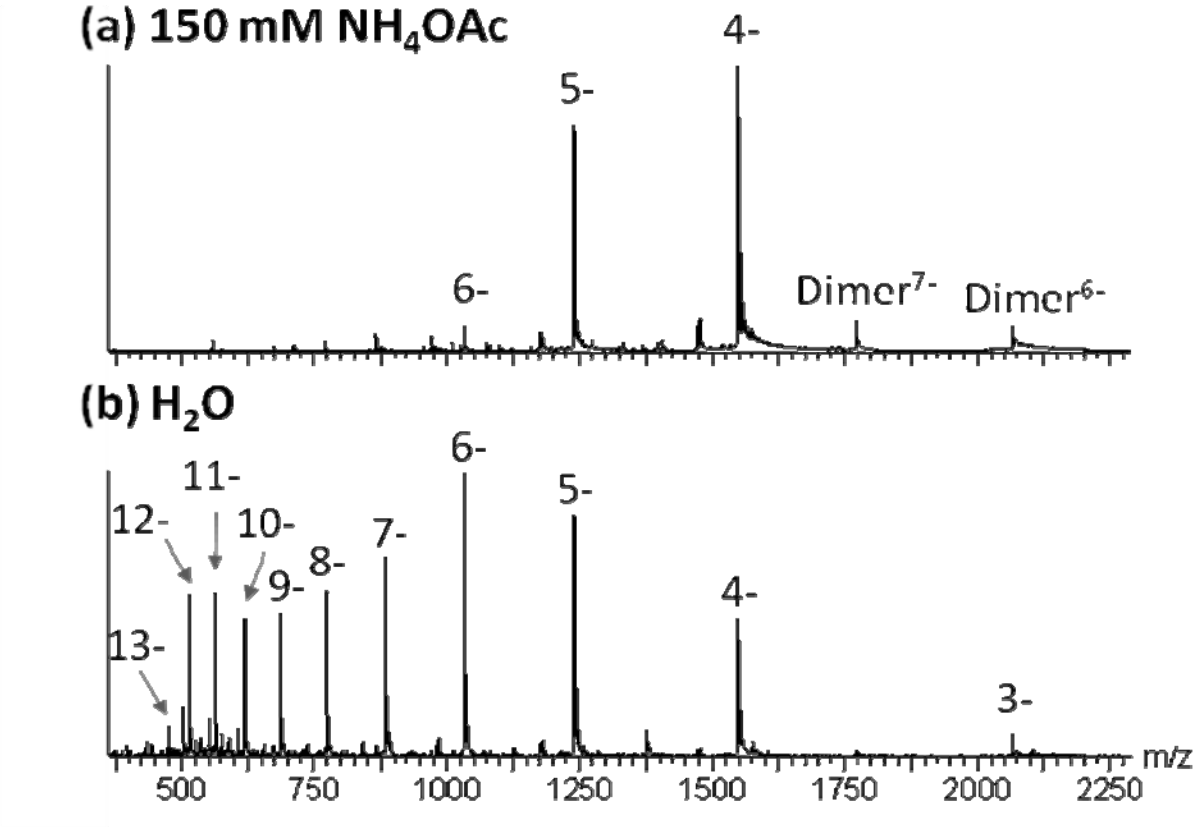

Figure S7: Electrospray spectra of the human telomeric C-rich sequence $(CCCAAT)_3CCC$. (a) 5 µM in DNA strand in 150 mM NH_4OAc solution. (b) 5 µM DNA solution in pure water: a bimodal charge state distribution is observed.

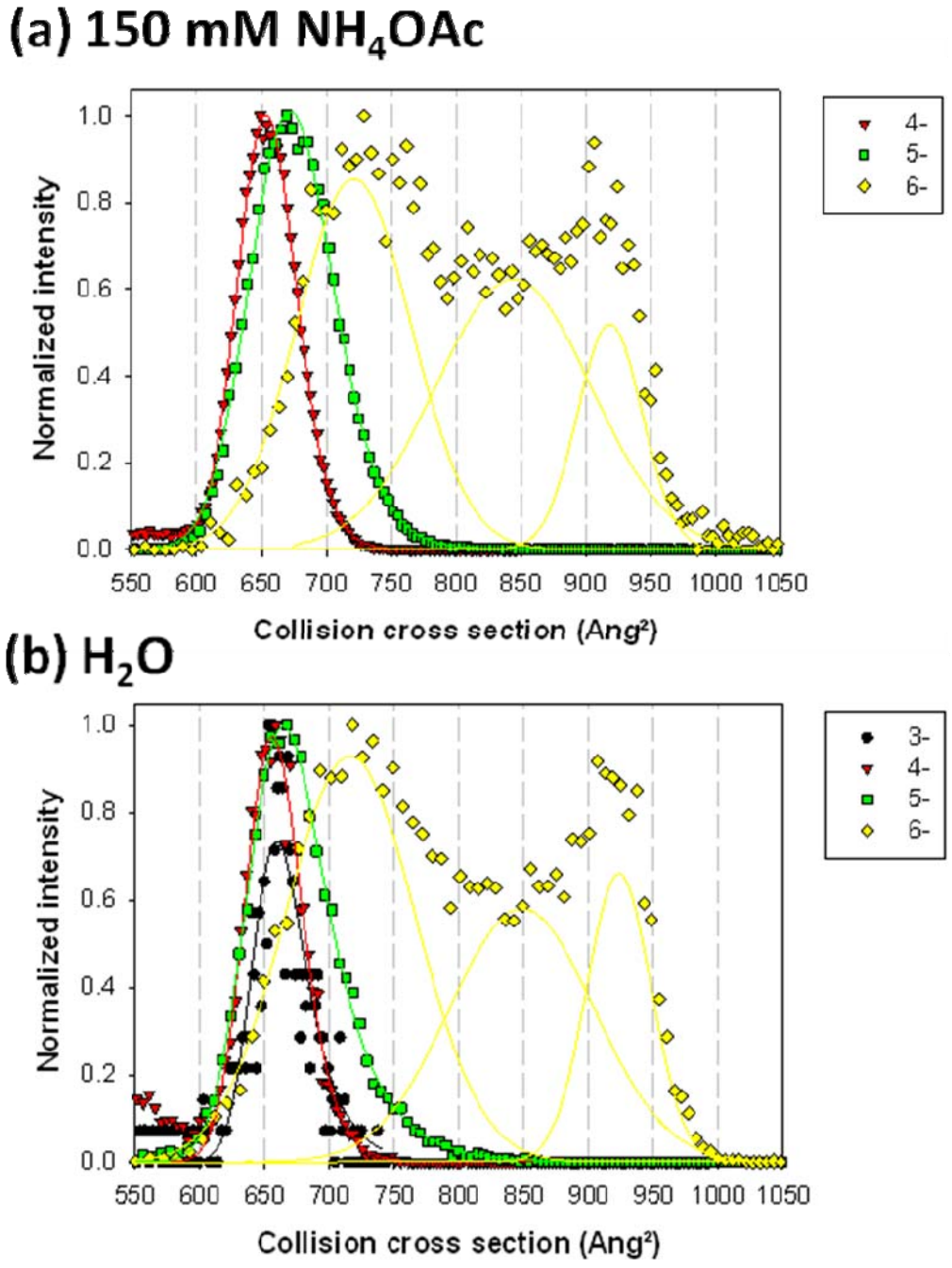

Figure S8: Experimental cross section of the human telomeric C-rich sequence $(CCCAAT)_3CCC$. (a) In 150 mM NH4OAc solution. (b) in pure water. The collision cross section for a given charge state are the same in ammonium acetate compared to water.