Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Vibrations of Guanine-Cytosine pair in chloroform: an anharmonic computational study. Supporting Information

James Green, Roberto Improta

July 2019

As stated in the main text, all calculated spectra contained here in the SI have been plotted by convoluting each peak with a Lorentzian with half width at half maximum of 4 cm^{-1} unless otherwise stated. For a given level of theory in a particular spectral range, the spectra are shown normalised to the most intense peak unless otherwise stated.

Local vibrational contributions to normal mode assignments that appear in the tables are labelled following the same convention as the main text: ω_m^n(X-Y), where ω is the type of vibration, with ν = stretch, δ = scissor and γ = umbrella; m is the nucleobase the vibration is localised on; n contains additional information about the vibration, with s = symmetric, as = antisymmetric, and ot = overtone; and X-Y are the atoms involved in the vibration, with atomic numbering according to figure 1 in the main text. When multiple local vibrations contribute to a normal mode, they will be indicated with a '+' and listed in order of importance. Combination bands are shown with the notation ([mode 1],[mode 2]). Any other notation will be indicated in the footnotes of the tables.

1 Harmonic Calculations

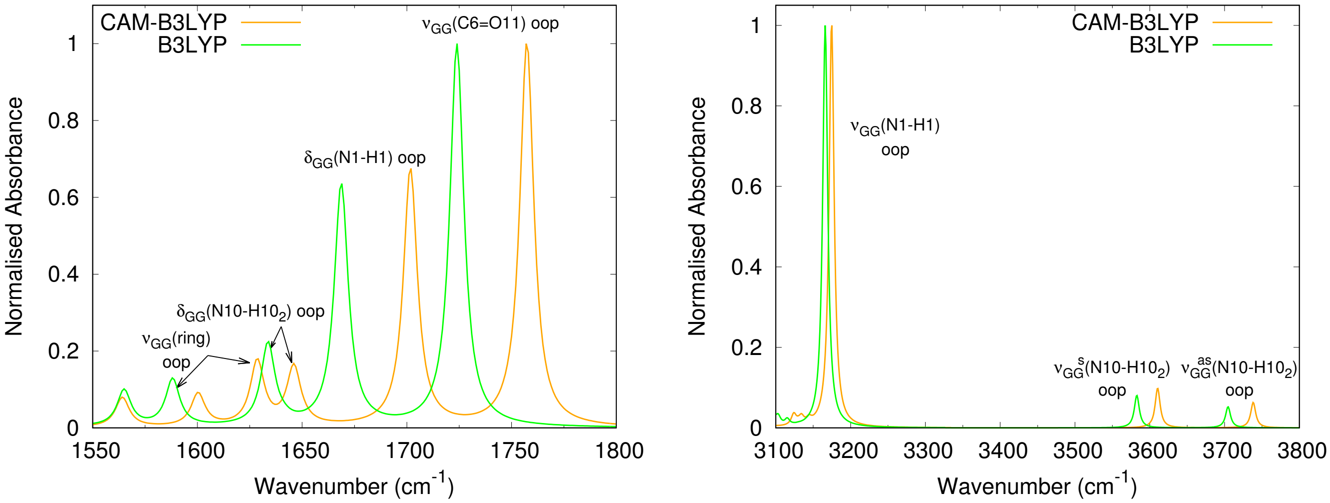

Figure S1: Spectra of cG=cG in PCM($CHCl_3$) at the CAM-B3LYP/6-311G(d,p) and B3LYP/6-311G(d,p) level of theory in the fingerprint (left) and NH stretch (right) regions.

Table S1: Harmonic frequencies (in cm^{-1}) at the CAM-B3LYP and B3LYP level of theory with 6-311G(d,p) basis in PCM(chloroform) for the cC, cG, cG≡cC, and cG=cG models described in the text. Compared to experimental results on TBDMS-C, -G, and -GC in chloroform.

Description*	CAM-B3LYP	B3LYP	Exp
C			
$\nu_{\mathrm{C}}^{\mathrm{as}}(\mathrm{N8'{-}H_2})$	3760	3731	$3534^{\mathrm{a,b}}$
$\nu_{\mathrm{C}}^{\mathrm{s}}(\mathrm{N8'{-}H_2})$	3630	3602	$3418^{\mathrm{a,b}}$
$\nu_{\mathrm{C}}(\mathrm{C2'{=}O7'})$	1748	1711	1659^{c}
$\nu_{\mathrm{C}}(\mathrm{C5'{=}C6'})$	1705	1668	1598^{c}
$\delta_{\mathrm{C}}(\mathrm{N8'{-}H_2})$	1637	1627	1640^{c}
G			
$\nu_{\mathrm{G}}^{\mathrm{as}}(\mathrm{N10{-}H_2})$	3726	3690	$3521^{\mathrm{a,b,d}}$
$\nu_{\mathrm{G}}^{\mathrm{s}}(\mathrm{N10{-}H_2}) + \nu_{\mathrm{G}}(\mathrm{N1{-}H1})$ ip	3618	3586	$3411^{\mathrm{a,b,d}}$
$\nu_{\mathrm{G}}^{\mathrm{s}}(\mathrm{N10{-}H_2}) + \nu_{\mathrm{G}}(\mathrm{N1{-}H1})$ oop	3611	3580	$3411^{\mathrm{a,b,d}}$
$\nu_{\mathrm{G}}(\mathrm{C6{=}O11})$	1803	1764	1686^{c}
$\delta_{\mathrm{G}}(\mathrm{N10{-}H_2}) + \delta_{\mathrm{G}}(\mathrm{N1{-}H1})$	1674	1655	1631^{c}
$\nu_{\mathrm{G}}(\mathrm{ring})$	1627		1596^{c}
GC			
$\nu_{\mathrm{G}}(\mathrm{N10{-}H10_f})$	3708	3683	$3491^{\mathrm{a,b}}$
$\nu_{\mathrm{C}}(\mathrm{N8'{-}H8'_f})$	3699	3677	$3491^{\mathrm{a,b}}$
$\nu_{\mathrm{G}}(\mathrm{N10{-}H10})$	3356	3340	$3303^{\mathrm{a,b}}$
$\nu_{\mathrm{C}}(\mathrm{N8'{-}H8'})$	3302	3287	$3145^{\mathrm{a,b}}$
$\nu_{\mathrm{G}}(\mathrm{N1{-}H1})$	3160	3147	$3145^{\mathrm{a,b}}$
$\nu_{\mathrm{G}}(\mathrm{C6{=}O11}) + \delta_{\mathrm{G}}(\mathrm{N1{-}H1}) + \delta_{\mathrm{G}}(\mathrm{N10{-}H_2}) + \delta_{\mathrm{C}}(\mathrm{N8'{-}H_2})$	1765	1732	$1689^{\mathrm{c,e}}$
$\nu_{\mathrm{C}}(\mathrm{C2'{=}O7'}) + \delta_{\mathrm{C}}(\mathrm{N8'{-}H_2}) + \delta_{\mathrm{G}}(\mathrm{N10{-}H_2}) + \nu_{\mathrm{G}}(\mathrm{C6{=}O11})$	1720	1693	$1650^{\mathrm{c,e}}$
$\delta_{\mathrm{C}}(\mathrm{N8'{-}H_2}) + \nu_{\mathrm{C}}(\mathrm{C5'{=}C6'}) + \delta_{\mathrm{G}}(\mathrm{N10{-}H_2})$	1711	1682	$1650^{\mathrm{c,e}}$
$\delta_{\mathrm{G}}(\mathrm{N10{-}H_2}) + \delta_{\mathrm{G}}(\mathrm{N1{-}H1}) + \delta_{\mathrm{C}}(\mathrm{N8'{-}H_2})$	1664	1638	$1610^{\mathrm{c,e}}$
GG			
$\nu_{\mathrm{GG}}^{\mathrm{as}}(\mathrm{N10{-}H_2})$ oop	3738	3705	3490^{d}
$\nu_{\mathrm{GG}}^{\mathrm{s}}(\mathrm{N10{-}H_2})$ oop	3611	3583	$\sim 3370^{\mathrm{d}}$
$\nu_{\mathrm{GG}}(\mathrm{N1{-}H1})$ oop	3175	3166	$\sim 3180^{\mathrm{d}}$
$\nu_{\mathrm{GG}}(\mathrm{C6{=}O11})$ oop + $\delta_{\mathrm{GG}}(\mathrm{N1{-}H1})$ oop + $\delta_{\mathrm{GG}}(\mathrm{N10{-}H_2})$ oop	1757	1724	-
$\delta_{\mathrm{GG}}(\mathrm{N1{-}H1})$ oop + $\delta_{\mathrm{GG}}(\mathrm{N10{-}H_2})$ oop	1702	1669	-
$\delta_{\mathrm{GG}}(\mathrm{N10{-}H_2})$ oop + $\delta_{\mathrm{GG}}(\mathrm{N1{-}H1})$ oop	1646	1634	-
$\nu_{\mathrm{GG}}(\mathrm{ring})$ oop + $\delta_{\mathrm{GG}}(\mathrm{N10{-}H_2})$ oop	1629	1588	-

[a] Ref. 1, [b] Ref. 2, [c] Ref. 3, [d] Ref. 4, [e] Ref. 5.

* ip = in phase, oop = out of phase, where for monomeric G ip and oop refer to combinations of local modes, whilst for the GG dimer they refer to the same local mode, but on each G.

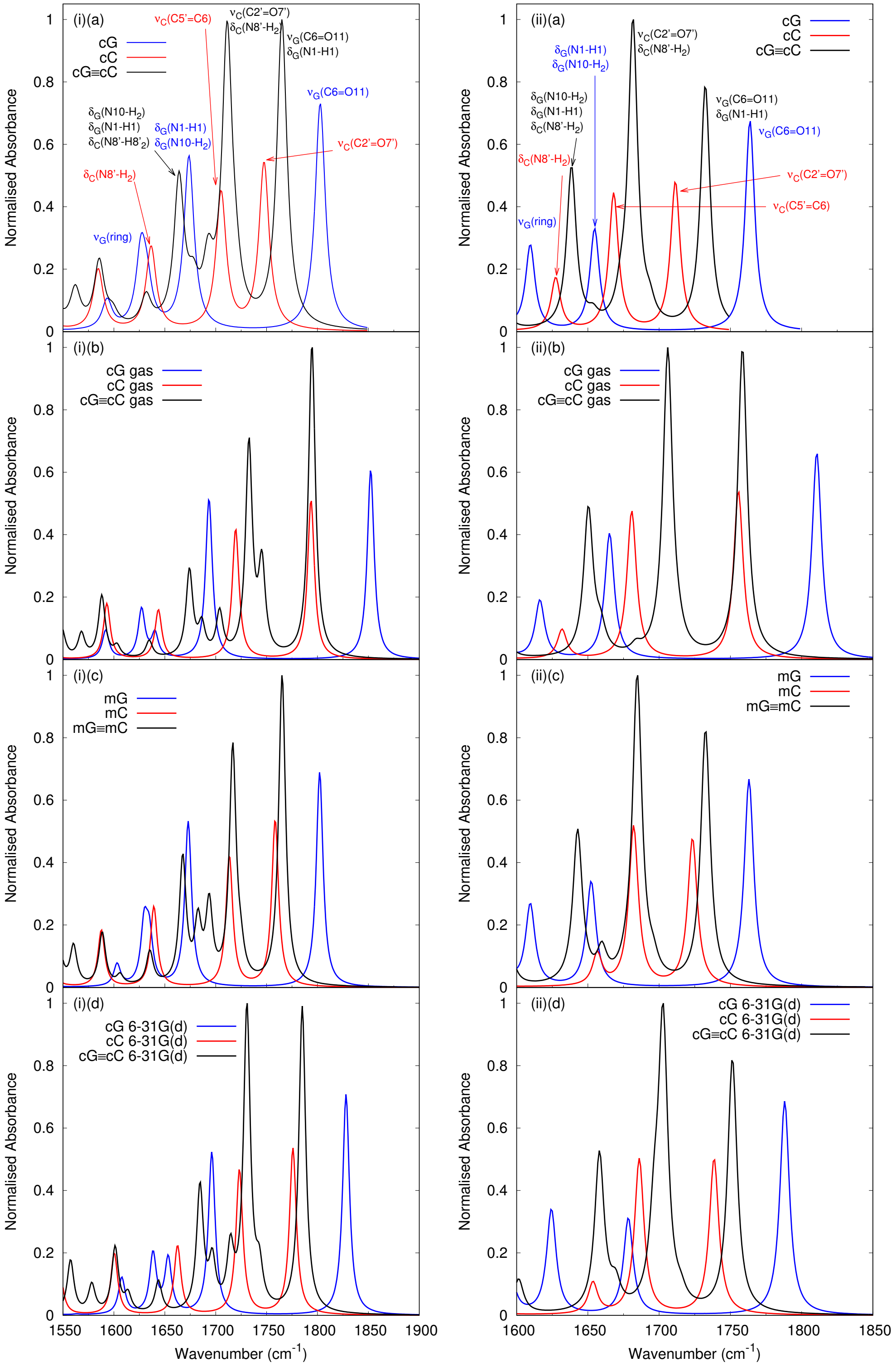

Figure S2: Harmonic spectra for GC systems in the fingerprint region computed at the (i) CAM-B3LYP and (ii) B3LYP level. Panel (a) cG≡cC, cG and cC in PCM(chloroform) with 6-311G(d,p) basis (as in figure 3 in the main text); (b) cG≡cC, cG and cC in the gas phase with 6-311G(d,p) basis; (c) mG≡mC, mG and mC in PCM(chloroform) with 6-311G(d,p) basis; (d) cG≡cC, cG and cC in PCM(chloroform) with 6-31G(d) basis.

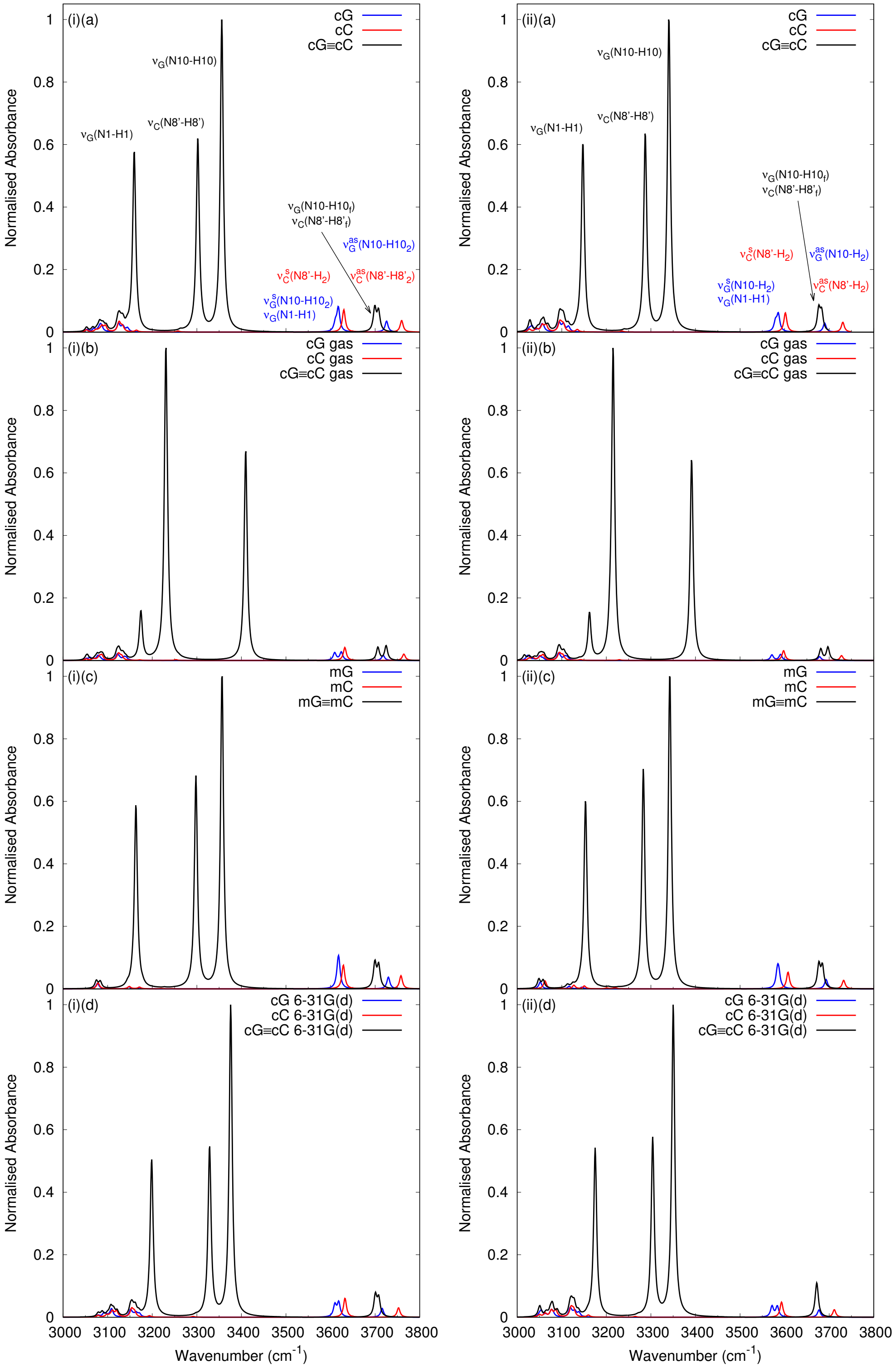

Figure S3: Harmonic spectra for GC systems in the NH stretching region computed at the (i) CAM-B3LYP and (ii) B3LYP level. Panel (a) cG≡cC, cG and cC in PCM(chloroform) with 6-311G(d,p) basis (as in figure 5 in the main text); (b) cG≡cC, cG and cC in the gas phase with 6-311G(d,p) basis; (c) mG≡mC, mG and mC in PCM(chloroform) with 6-311G(d,p) basis; (d) cG≡cC, cG and cC in PCM(chloroform) with 6-31G(d) basis.

2 Anharmonic calculations

2.1 1-methylcytosine

2.1.1 Gas Phase

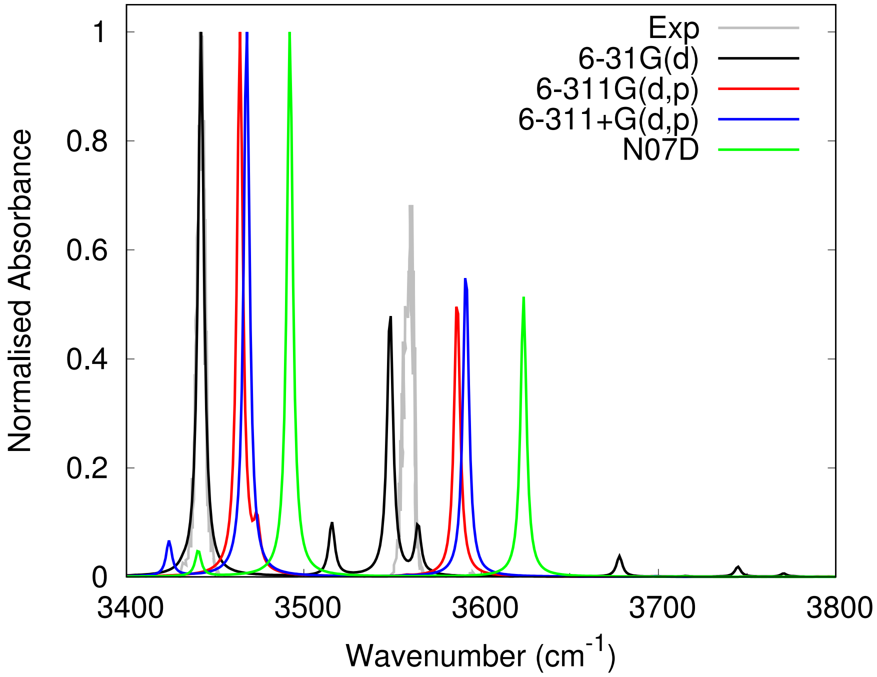

Figure S4: Linear anharmonic infrared spectra computed in the gas phase for 1-methylcytosine in the NH stretching region with B3LYP, illustrating the use of the N07D basis. Black curve: 6-31G(d); red-curve: 6-311G(d,p); blue curve: 6-311+G(d,p); green curve: N07D. The experimental spectrum reproduced from Ref. 6 is shown in grey.

Table S2: Anharmonic frequencies ω (with diagonal anharmonicities Δ_{ii} in parentheses) in cm^{-1}; and dipole strength D in cgs units of 10^{-40} esu^{2}cm^2 for the meC model. Calculated in the gas phase with CAM-B3LYP and B3LYP with various basis sets as detailed in the column headers, and compared to experimental results.[6,7]

Description	6-31G(d)		6-311G(d,p)		6-311+G(d,p)		Exp
	ω (Δ_{ii})	D	ω (Δ_{ii})	D	ω (Δ_{ii})	D	
			CAM-B3LYP				
$\nu_{\mathrm{C}}^{\mathrm{as}}(\mathrm{N8'{-}H_2})$	3607 (90.3)	47	3680 (95.0)	67	3685 (89.1)	53	3557
$\nu_{\mathrm{C}}^{\mathrm{s}}(\mathrm{N8'{-}H_2})$	3496 (77.8)	102	3545 (78.3)	205	3541 (74.7)	243	3438
$\nu_{\mathrm{C}}(\mathrm{C2'{=}O7'})$	1808 (16.1)	231	1785 (15.9)	585	1763 (26.8)	110	1716
$\nu_{\mathrm{C}}(\mathrm{C5'{=}C6'})$	1713 (5.4)	633	1700 (11.7)	1380	1679 (-2.8)	861	1688
$\delta_{\mathrm{C}}(\mathrm{N8'{-}H_2})$	1629 (16.7)	448	1617 (11.6)	523	1589 (13.6)	770	1603
			B3LYP				
$\nu_{\mathrm{C}}^{\mathrm{as}}(\mathrm{N8'{-}H_2})$	3549 (94.5)	32	3586 (95.1)	45	3591 (93.4)	57	3557
$\nu_{\mathrm{C}}^{\mathrm{s}}(\mathrm{N8'{-}H_2})$	3442 (81.9)	67	3464 (79.8)	90	3468 (78.6)	104	3438
$\nu_{\mathrm{C}}(\mathrm{C2'{=}O7'})$	1766 (16.6)	685	1745 (16.6)	878	1724 (24.3)	1063	1716
$\nu_{\mathrm{C}}(\mathrm{C5'{=}C6'})$	1670 (9.9)	214	1653 (4.7)	491	1642 (-2.4)	667	1688
$\delta_{\mathrm{C}}(\mathrm{N8'{-}H_2})$	1615 (16.2)	152	1593 (15.9)	392	1588 (15.9)	461	1603

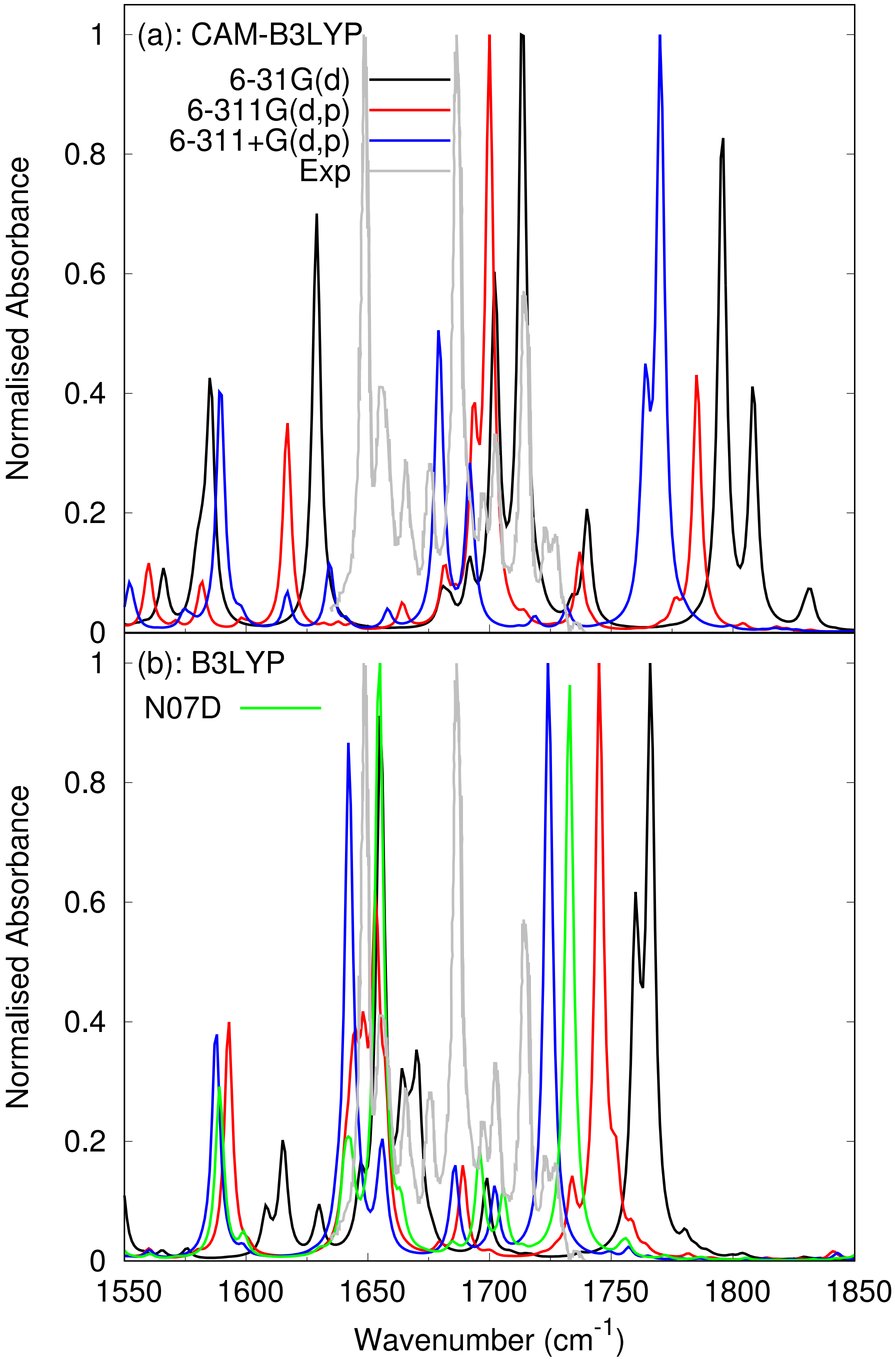

Figure S5: Linear anharmonic infrared spectra computed in the gas phase for 1-methylcytosine in the fingerprint region with (a) CAM-B3LYP and (b) B3LYP. Black curve: 6-31G(d); red-curve: 6-311G(d,p); blue curve: 6-311+G(d,p); green curve for B3LYP: N07D. The experimental spectrum reproduced from Ref. 6 is shown in grey. Assignments of fundamentals and dipole strengths in Table S2

2.1.2 Chloroform solution

Table S3: Anharmonic frequencies ω (with diagonal anharmonicities Δ_{ii} in parentheses) in cm^{-1}; and dipole strength D in cgs units of 10^{-40} $\mathrm{esu}^2\mathrm{cm}^2$ for the meC model. Calculated with CAM-B3LYP and B3LYP in PCM(chloroform) with various basis sets as detailed in the column headers, and compared to experimental results.

Description	6-31G(d)		6-311G(d,p)		6-311+G(d,p)		Exp
	ω (Δ_{ii})	D	ω (Δ_{ii})	D	ω (Δ_{ii})	D	
			CAM-B3LYP				
$\nu_{\mathrm{C}}^{\mathrm{as}}(\mathrm{N8'{-}H_2})$	3630 (84.7)	99	3581 (72.4)	71	3571 (72.9)	78	3534[a,b]
$\nu_{\mathrm{C}}^{\mathrm{s}}(\mathrm{N8'{-}H_2})$	3519 (73.2)	208	3471 (66.2)	80	3462 (66.6)	103	3418[a,b]
$\nu_{\mathrm{C}}(\mathrm{C2'{=}O7'})$	1744 (-3.8)	717	1721 (1.1)	1599	1690 (15.8)	1081	1659[c]
$\nu_{\mathrm{C}}(\mathrm{C5'{=}C6'})$	1707 (4.1)	1627	1692 (8.3)	1149	1686 (4.3)	877	1598[c]
$\delta_{\mathrm{C}}(\mathrm{N8'{-}H_2})$	1682 (9.0)	363	1595 (27.2)	338	1598 (17.2)	939	1640[c]
			B3LYP				
$\nu_{\mathrm{C}}^{\mathrm{as}}(\mathrm{N8'{-}H_2})$	3534 (90.4)	69	3537 (86.7)	146	3538 (87.4)	321	3534[a,b]
$\nu_{\mathrm{C}}^{\mathrm{s}}(\mathrm{N8'{-}H_2})$	3427 (79.0)	144	3438 (73.9)	129	3441 (74.2)	230	3418[a,b]
$\nu_{\mathrm{C}}(\mathrm{C2'{=}O7'})$	1728 (27.5)	939	1693 (14.8)	1369	1660 (12.1)	2473	1659[c]
$\nu_{\mathrm{C}}(\mathrm{C5'{=}C6'})$	1662 (10.0)	360	1655 (20.7)	109	1644 (2.6)	177	1598[c]
$\delta_{\mathrm{C}}(\mathrm{N8'{-}H_2})$	1536 (33.8)	589	1636 (-6.6)	1103	1673 (13.2)	1006	1640[c]

[a] Ref. 1, [b] Ref. 2, [c] Ref. 3

2.2 9-methylguanine monomer and dimer

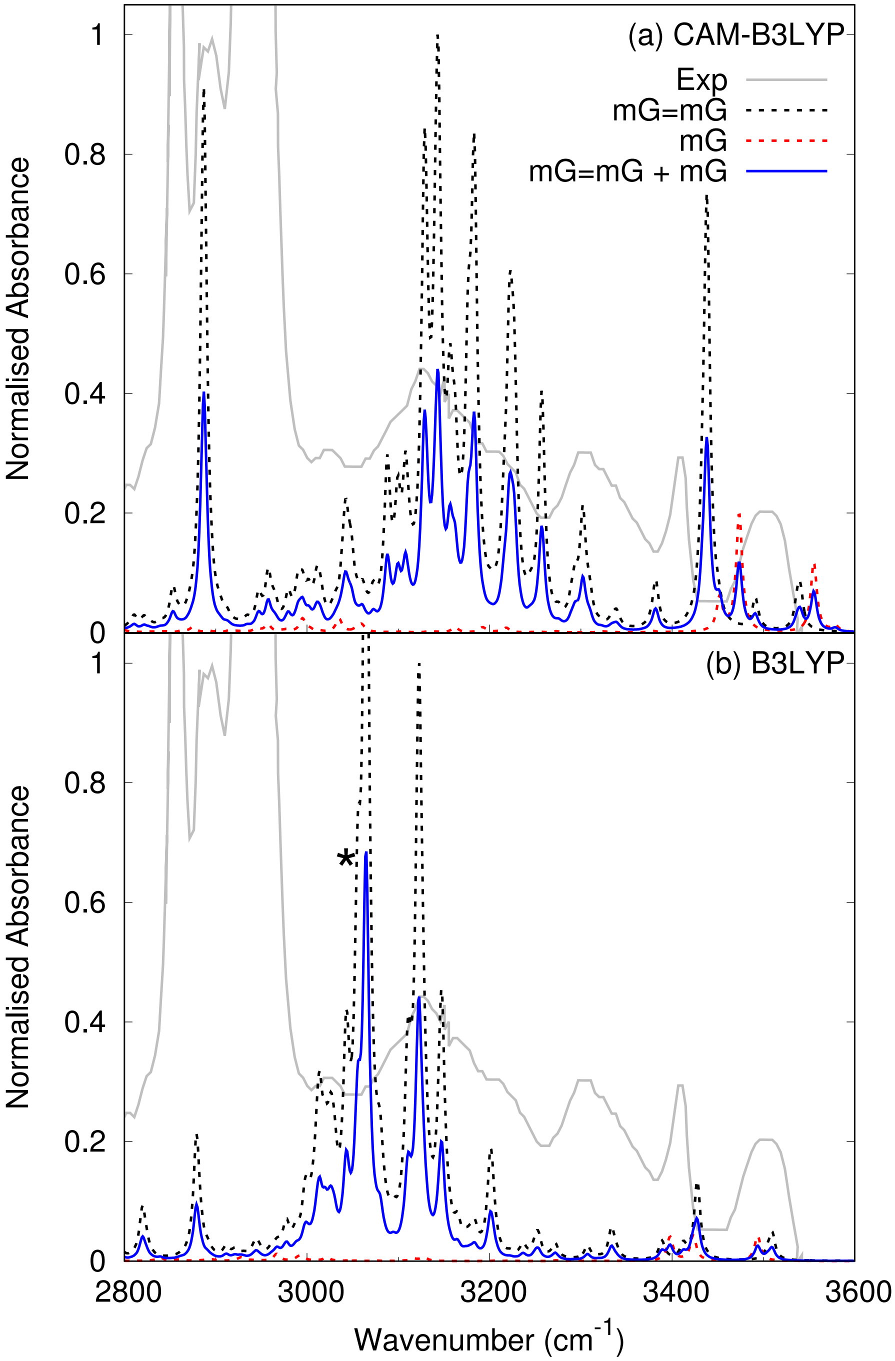

Figure S6: Anharmonic spectrum in PCM(chloroform) at (a) CAM-B3LYP and (b) B3LYP level of theory with 6-31G(d) basis for mG=mG (black dashed), mG (red dashed), and the combination of the two in the ratio of monomer to dimer in a nominal 2 mM solution of G ([G] = 0.778 mM, [GG] = 0.611 mM), according to the equations in Ref. 4. Experimental spectra also shown in grey. A potential anomalous peak is marked with a * and not used for normalisation. The experimental spectrum is normalised to the most intense non-anomalous peak on the G + GG spectrum, to aid comparison.

Table S4: Anharmonic frequencies ω of fundamentals (with diagonal anharmonicities Δ_{ii} in parentheses), in cm^{-1}; and dipole strength D in cgs units of 10^{-40} esu^{2}cm^2 for the mG and mG=mG model in the NH stretch region. Calculated with CAM-B3LYP and B3LYP in PCM(chloroform) with the 6-31G(d) basis set, and compared to experimental results.

Description*	CAM-B3LYP		B3LYP		Exp
	ω (Δ_{ii})	D	ω (Δ_{ii})	D	
G					
$\nu_{\mathrm{G}}^{\mathrm{as}}(\mathrm{N10{-}H_2})$	3555 (89.4)	93	3494 (95.3)	78	3521[a,b,c]
$\nu_{\mathrm{G}}(\mathrm{N1{-}H1})$	3474 (158)	160	3423 (158)	84	3411[a,b,d]
$\nu_{\mathrm{G}}^{\mathrm{s}}(\mathrm{N10{-}H_2})$	3452 (56.3)	45	3398 (69.1)	79	3411[a,b,d]
GG					
$\nu_{\mathrm{G1}}^{\mathrm{as}}(\mathrm{N10{-}H_2})$	3540 (91.3)	50	3510 (92.6)	87	3490[c]
$\nu_{\mathrm{G2}}^{\mathrm{as}}(\mathrm{N10{-}H_2})$	3537 (91.3)	26	3505 (92.6)	21	3490[c]
$\nu_{\mathrm{G1}}^{\mathrm{s}}(\mathrm{N10{-}H_2})$	3437 (79.7)	14	3412 (79.6)	47	$\sim$ 3370[c]
$\nu_{\mathrm{G2}}^{\mathrm{s}}(\mathrm{N10{-}H_2})$	3438 (81.0)	550	3428 (79.6)	263	$\sim$ 3370[c]
$\nu_{\mathrm{GG}}(\mathrm{N1{-}H1})$ oop	2887 (40.3)	868	2879 (53.8)	472	$\sim$ 3180[c]
$\nu_{\mathrm{GG}}(\mathrm{N1{-}H1})$ ip	2854 (250)	24	2820 (232)	194	$\sim$ 3180[c]

[a] Ref. 1, [b] Ref. 2, [c] Ref. 4

[*] ip = in phase and oop = out of phase combination of the same vibration on different G in the GG dimer.

Table S5: Anharmonic frequencies of combination bands ω_{cb} (with anharmonic fundamental frequencies of the two modes involved ω_1 and ω_2 in parentheses), in cm^{-1}; and dipole strength D in cgs units of 10^{-40} esu^{2}cm^2 for the mG=mG model in the NH stretch region. Calculated with CAM-B3LYP and B3LYP in PCM(chloroform) with the 6-31G(d) basis set.

Description*	ω_{cb} ([ω_1, ω_2])	D
CAM-B3LYP		
([ν_{GG}(ring) oop + δ_{GG}(N1−H1) oop + ν_{GG}(C6=O11) oop], [ν_{GG}(ring) ip + δ_{GG}(N1−H1) ip + ν_{GG}(C6=O11) ip])	3088 ([1540,1555])	212
([ν_{GG}(ring) ip + δ_{GG}(N1−H1) ip + ν_{GG}(C6=O11) ip], [ν_{GG}(ring) oop + δ_{GG}(N10−H$_2$) oop])	3129 ([1555,1575])	545
([ν_{GG}(ring) oop + δ_{GG}(N1−H1) oop + ν_{GG}(C6=O11) oop], [ν_{GG}(C4=N3) oop + δ_{GG}(N10−H$_2$) oop])	3143 ([1540,1601])	658
†([γ_{GG}(C9−H9$_3$) ip],[ν_{GG}(C6=O11) oop + δ_{GG}(N1−H1) oop])	3157 ([1351,1735])	238
([ν_{GG}(ring) ip + δ_{GG}(N1−H1) ip + ν_{GG}(C6=O11) ip], [δ_{GG}(N10−H$_2$) oop])	3177 ([1555,1628])	307
([ν_{GG}(ring) oop + δ_{GG}(N1−H1) oop + ν_{GG}(C6=O11) oop], [δ_{GG}(N10−H$_2$) ip])	3183 ([1540,1642])	605
†([γ_{GG}(C9−H9$_3$) oop],[ν_{GG}(C6=O11) ip + δ_{GG}(N1−H1) ip])	3222 ([1468,1762])	273
([ν_{GG}(ring) ip + δ_{GG}(N1−H1) ip + ν_{GG}(C6=O11) ip], [δ_{GG}(N1−H1) oop + ν_{GG}(C2−N10) oop])	3227 ([1555,1669])	266
([ν_{GG}(ring) oop + δ_{GG}(N1−H1) oop + ν_{GG}(C6=O11) oop], [δ_{GG}(N1−H1) ip + ν_{GG}(C2−N10) ip])	3257 ([1540,1717])	326
B3LYP		
([ν_{GG}(C2−N10) oop + δ_{GG}(N1−H1) oop + ν_{GG}(N9−C9) oop], [δ_{GG}(N1−H1) ip + δ_{GG}(N10−H$_2$) ip + ν_{GG}(C2−N10) ip])	3014 ([1378,1646])	314
([ν_{GG}(C2−N10) ip + δ_{GG}(N1−H1) ip + ν_{GG}(N9−C9) ip], [δ_{GG}(N1−H1) oop + δ_{GG}(N10−H$_2$) oop])	3020 ([1381,1643])	581
†([γ_{GG}(C9−H9$_3$) ip],[δ_{GG}(N10−H$_2$) oop + δ_{GG}(N1−H1) oop])	3056 ([1399,1604])	1009
†([γ_{GG}(C9−H9$_3$) ip],[δ_{GG}(N10−H$_2$) ip + δ_{GG}(N1−H1) ip])	3065 ([1399,1614])	3167
†([γ_{GG}(C9−H9$_3$) ip],[δ_{GG}(N1−H1) oop + δ_{GG}(N10−H$_2$) oop])	3080 ([1399,1643])	212
([ν_{GG}(ring) oop + ν_{GG}(C6=O11) oop + δ_{GG}(N1−H1) oop], [ν_{GG}(C4=N3) ip])	3110 ([1546,1570])	543
([ν_{GG}(ring) oop + δ_{GG}(N1−H1) oop], [δ_{GG}(N10−H$_2$) oop + δ_{GG}(N1−H1) oop]	3123 ([1532,1604])	1806
([ν_{GG}(ring) ip + ν_{GG}(C6=O11) ip + δ_{GG}(N1−H1) ip], [δ_{GG}(N10−H$_2$) ip + δ_{GG}(N1−H1) ip])	3127 ([1530,1614])	270
([ν_{GG}(ring) oop + δ_{GG}(N1−H1) oop], [δ_{GG}(N1−H1) ip + δ_{GG}(N10−H$_2$) ip + ν_{GG}(C2−N10) ip])	3148 ([1532,1646])	810
([ν_{GG}(ring) oop + δ_{GG}(N1−H1) oop], [δ_{GG}(N1−H1) ip + ν_{GG}(C6=O11) ip])	3201 ([1532,1698])	314

* ip = in phase and oop = out of phase combination of the same vibration on different G in the GG dimer.

† Likely an anomaly, due to the choice of molecular model.

2.3 1-methylcytosine 9-methylguanine WC pair

Table S6: Coordinates of the mG≡mC model in Å, optimised at the CAM-B3LYP/6-31G(d) and B3LYP/6-31G(d) level in PCM(chloroform)

Atom	CAM-B3LYP			B3LYP		
	x	y	z	x	y	z
N	-4.629519	-0.354252	0.008967	-4.651958	-0.359749	0.039331
C	-4.954701	0.985238	0.027386	-4.977760	0.984351	0.101387
N	-3.917883	1.773582	0.026933	-3.936446	1.779339	0.095397
C	-2.838910	0.910528	0.008111	-2.854593	0.916340	0.025639
C	-3.265050	-0.411589	-0.002998	-3.280987	-0.415343	-0.010215
N	-2.525848	-1.538659	-0.022860	-2.539700	-1.541760	-0.079743
C	-1.228751	-1.292208	-0.032674	-1.236522	-1.291275	-0.116297
N	-0.699963	-0.024837	-0.020788	-0.708465	-0.019786	-0.078178
C	-1.435332	1.163690	-0.000532	-1.448843	1.175031	-0.007236
O	-0.855138	2.255829	0.008089	-0.866193	2.272119	0.020418
N	-0.348061	-2.305596	-0.062744	-0.354855	-2.309249	-0.226060
H	0.666612	-2.163880	-0.029798	0.653678	-2.169831	-0.099445
H	-0.723964	-3.240174	-0.037628	-0.730816	-3.240256	-0.122865
H	-5.989093	1.303340	0.039267	-6.012049	1.300918	0.148576
C	-5.533058	-1.486667	0.011479	-5.560521	-1.495218	0.029283
H	-5.332373	-2.130694	-0.846940	-5.418162	-2.087945	-0.878483
H	-5.413939	-2.070561	0.926949	-5.383904	-2.133767	0.899428
H	-6.557398	-1.117905	-0.048022	-6.586256	-1.123482	0.058689
N	2.214700	0.321556	-0.004840	2.227620	0.322161	-0.019012
C	2.779503	1.529942	-0.008025	2.797270	1.535953	-0.037602
C	4.203448	1.686723	-0.003371	4.223271	1.691923	-0.018919
C	4.943467	0.556340	0.006068	4.967931	0.556231	0.023518
N	4.365899	-0.673000	0.010862	4.389191	-0.676499	0.047461
C	2.971270	-0.801231	0.005311	2.983191	-0.805324	0.026412
O	2.493252	-1.942154	0.010159	2.503135	-1.950684	0.051296
N	1.976558	2.594296	-0.016232	1.989359	2.604023	-0.077719
H	0.954517	2.478133	-0.011017	0.966841	2.488638	-0.054909
H	2.364921	3.524672	-0.014965	2.376570	3.535826	-0.072032
H	4.667473	2.663658	-0.006764	4.690862	2.668152	-0.036392
H	6.027102	0.571507	0.010564	6.052107	0.573663	0.041333
C	5.156938	-1.898061	0.021533	5.185229	-1.904514	0.094107
H	4.930185	-2.497353	-0.861471	4.970543	-2.523209	-0.780213
H	4.922141	-2.487112	0.909299	4.939390	-2.476087	0.991924
H	6.213907	-1.632983	0.024814	6.242715	-1.637022	0.105601
H	0.329808	0.082987	-0.021639	0.321846	0.089442	-0.081998

Table S7: Anharmonic frequencies ω of fundamentals and overtones (with diagonal anharmonicities Δ_{ii} in parentheses), in cm^{-1}; and dipole strength D in cgs units of 10^{-40} $\mathrm{esu}^2\mathrm{cm}^2$ for the mG≡mC model in the NH stretch region. Calculated with CAM-B3LYP and B3LYP in PCM(chloroform) with the 6-31G(d) basis set, and compared to experimental results.

Description	CAM-B3LYP		B3LYP		Exp
	ω (Δ_{ii})	D	ω (Δ_{ii})	D	
$\nu_\mathrm{G}(\mathrm{N10{-}H10_f})$	3545 (137)	95.6	3480 (144)	12	3491[a,b]
$\nu_\mathrm{C}(\mathrm{N8'{-}H8'_f})$	3531 (134)	120	3489 (144)	143	3491[a,b]
$\nu_\mathrm{G}(\mathrm{N10{-}H10})$	3155 (165)	975	3112 (169)	541	3303[a,b]
$\nu_\mathrm{C}(\mathrm{N8'{-}H8'})$	3067 (204)	2527	3028 (226)	1209	3145[a,b]
$\nu_\mathrm{G}(\mathrm{N1{-}H1})$	2932 (220)	303	2867 (210)	718	3145[a,b]
$\delta^\mathrm{ot}_\mathrm{C}(\mathrm{N8'{-}H_2})$	3332 (-5.8)	303	3270 (-9.3)	205	-

[a] Ref. 1, [b] Ref. 2

Table S8: Anharmonic frequencies of combination bands ω_cb (with anharmonic fundamental frequencies of the two modes involved ω_1 and ω_2 in parentheses), in cm^{-1}; and dipole strength D in cgs units of 10^{-40} $\mathrm{esu}^2\mathrm{cm}^2$ for the mG≡mC model in the NH stretch region. Calculated with CAM-B3LYP and B3LYP in PCM(chloroform) with the 6-31G(d) basis set.

Description	ω_cb ($[\omega_1, \omega_2]$)	D
CAM-B3LYP		
†([$\gamma_\mathrm{G}(\mathrm{C9{-}H9_3})$],[$\nu_\mathrm{G}(\mathrm{C6{=}O11})$ + $\delta_\mathrm{G}(\mathrm{N1{-}H1})$ + $\delta_\mathrm{C}(\mathrm{N8'{-}H_2})$])	3189 ([1493,1752])	1841
([$\nu_\mathrm{C}(\mathrm{C4'{=}C5'})$],[$\delta_\mathrm{G}(\mathrm{N10{-}H_2})$ + $\delta_\mathrm{G}(\mathrm{N1{-}H1})$])	3239 ([]1559,1682])	371
([ν_G(ring) + $\delta_\mathrm{G}(\mathrm{N10{-}H_2})$ + $\nu_\mathrm{G}(\mathrm{C6{=}O11})$],[$\delta_\mathrm{C}(\mathrm{N8'{-}H_2})$])	3246 ([1581,1663])	2441
([ν_G(ring) + $\delta_\mathrm{G}(\mathrm{N1{-}H1})$ + $\nu_\mathrm{G}(\mathrm{C6{=}O11})$], [$\nu_\mathrm{C}(\mathrm{C2'{=}O7'})$ + $\delta_\mathrm{C}(\mathrm{N8'{-}H_2})$ + $\nu_\mathrm{G}(\mathrm{C6{=}O11})$ + $\delta_\mathrm{G}(\mathrm{N10{-}H_2})$])	3251 ([1537,1717])	455
([$\delta_\mathrm{G}(\mathrm{N10{-}H_2})$ + $\delta_\mathrm{G}(\mathrm{N1{-}H1})$],[$\delta_\mathrm{C}(\mathrm{N8'{-}H_2})$])	3315 ([1637,1663])	698
B3LYP		
†([$\gamma_\mathrm{C}(\mathrm{C1'{-}H1'_3})$],[$\delta_\mathrm{G}(\mathrm{N1{-}H1})$ + $\delta_\mathrm{G}(\mathrm{N10{-}H_2})$])	3101 ([1520,1623])	276
([$\nu_\mathrm{C}(\mathrm{C4'{=}C5'})$ + ν_C(ring)], [$\delta_\mathrm{G}(\mathrm{N10{-}H_2})$ + $\delta_\mathrm{C}(\mathrm{N8'{-}H_2})$ + $\nu_\mathrm{C}(\mathrm{C5'{=}C6'})$ + $\nu_\mathrm{C}(\mathrm{C2'{=}O7'})$])	3195 ([1534,1662])	219
†([ν_G(ring) + $\delta_\mathrm{G}(\mathrm{N1{-}H1})$ + $\gamma_\mathrm{G}(\mathrm{C9{-}H9_3})$ + $\gamma_\mathrm{C}(\mathrm{C1'{-}H1'_3})$], [$\nu_\mathrm{G}(\mathrm{C6{=}O11})$ + $\delta_\mathrm{G}(\mathrm{N1{-}H1})$ + $\delta_\mathrm{C}(\mathrm{N8'{-}H_2})$])	3208 ([1460,1718])	204
([$\delta_\mathrm{G}(\mathrm{N1{-}H1})$ + $\delta_\mathrm{G}(\mathrm{N10{-}H_2})$ + ν_G(ring)], [$\delta_\mathrm{C}(\mathrm{N8'{-}H_2})$ + $\nu_\mathrm{C}(\mathrm{C2'{=}O7'})$ + $\delta_\mathrm{G}(\mathrm{N10{-}H_2})$ + $\nu_\mathrm{G}(\mathrm{C6{=}O11})$])	3313 ([1623,1696])	221
([$\delta_\mathrm{C}(\mathrm{N8'{-}H_2})$ + $\nu_\mathrm{C}(\mathrm{C5'{=}C6'})$ + $\delta_\mathrm{G}(\mathrm{N1{-}H1})$], [$\delta_\mathrm{C}(\mathrm{N8'{-}H_2})$ + $\nu_\mathrm{C}(\mathrm{C2'{=}O7'})$ + $\delta_\mathrm{G}(\mathrm{N10{-}H_2})$ + $\nu_\mathrm{G}(\mathrm{C6{=}O11})$])	3324 ([1630,1696])	223
([$\delta_\mathrm{G}(\mathrm{N10{-}H_2})$ + $\delta_\mathrm{C}(\mathrm{N8'{-}H_2})$ + $\nu_\mathrm{C}(\mathrm{C5'{=}C6'})$ + $\nu_\mathrm{C}(\mathrm{C2'{=}O7'})$], [$\delta_\mathrm{C}(\mathrm{N8'{-}H_2})$ + $\nu_\mathrm{C}(\mathrm{C2'{=}O7'})$+ $\delta_\mathrm{G}(\mathrm{N10{-}H_2})$ + $\nu_\mathrm{G}(\mathrm{C6{=}O11})$])	3353 ([1662,1696])	222

† Likely an anomaly, due to the choice of molecular model.

Table S9: Anharmonic frequencies ω of fundamentals (with diagonal anharmonicities Δ_{ii} in parentheses), in cm^{-1}; and dipole strength D in cgs units of 10^{-40} $\mathrm{esu}^2\mathrm{cm}^2$ for the fully deuterated and partially deuterated mG≡mC model in the ND stretch region. Calculated with CAM-B3LYP and B3LYP in PCM(chloroform) with the 6-31G(d) basis set, and compared to experimental results.[3]

Description	CAM-B3LYP		B3LYP		Exp
	ω (Δ_{ii})	D	ω (Δ_{ii})	D	
	Fully Deuterated				
$\nu_{\mathrm{C}}(\mathrm{N8'{-}D8'_f})$	2628 (66.8)	59	2602 (73.3)	67	2603
$\nu_{\mathrm{G}}(\mathrm{N10{-}D10_f})$	2622 (68.2)	112	2594 (71.1)	119	2603
$\nu_{\mathrm{G}}(\mathrm{N10{-}D10})$	2349 (80.6)	5.7	2328 (82.4)	143	2338
$\nu_{\mathrm{C}}(\mathrm{N8'{-}D8'})$	2289 (86.2)	2060	2268 (87.1)	872	2338
$\nu_{\mathrm{G}}(\mathrm{N1{-}D1})$	2203 (122)	834	2190 (131)	705	2197
	Partially Deuterated				
$\nu_{\mathrm{C}}(\mathrm{N8'{-}H8'_f})$	3525 (139)	86	3485 (145)	75	-
$\nu_{\mathrm{G}}(\mathrm{N10{-}H10_f})$	3408 (143)	89	3498 (152)	82	-
$\nu_{\mathrm{G}}(\mathrm{N10{-}D10})$	2385 (107)	280	2354 (98)	869	2338
$\nu_{\mathrm{C}}(\mathrm{N8'{-}D8'})$	2317 (118)	326	2306 (108)	682	2338
$\nu_{\mathrm{G}}(\mathrm{N1{-}D1})$	2216 (130)	298	2218 (139)	267	2197

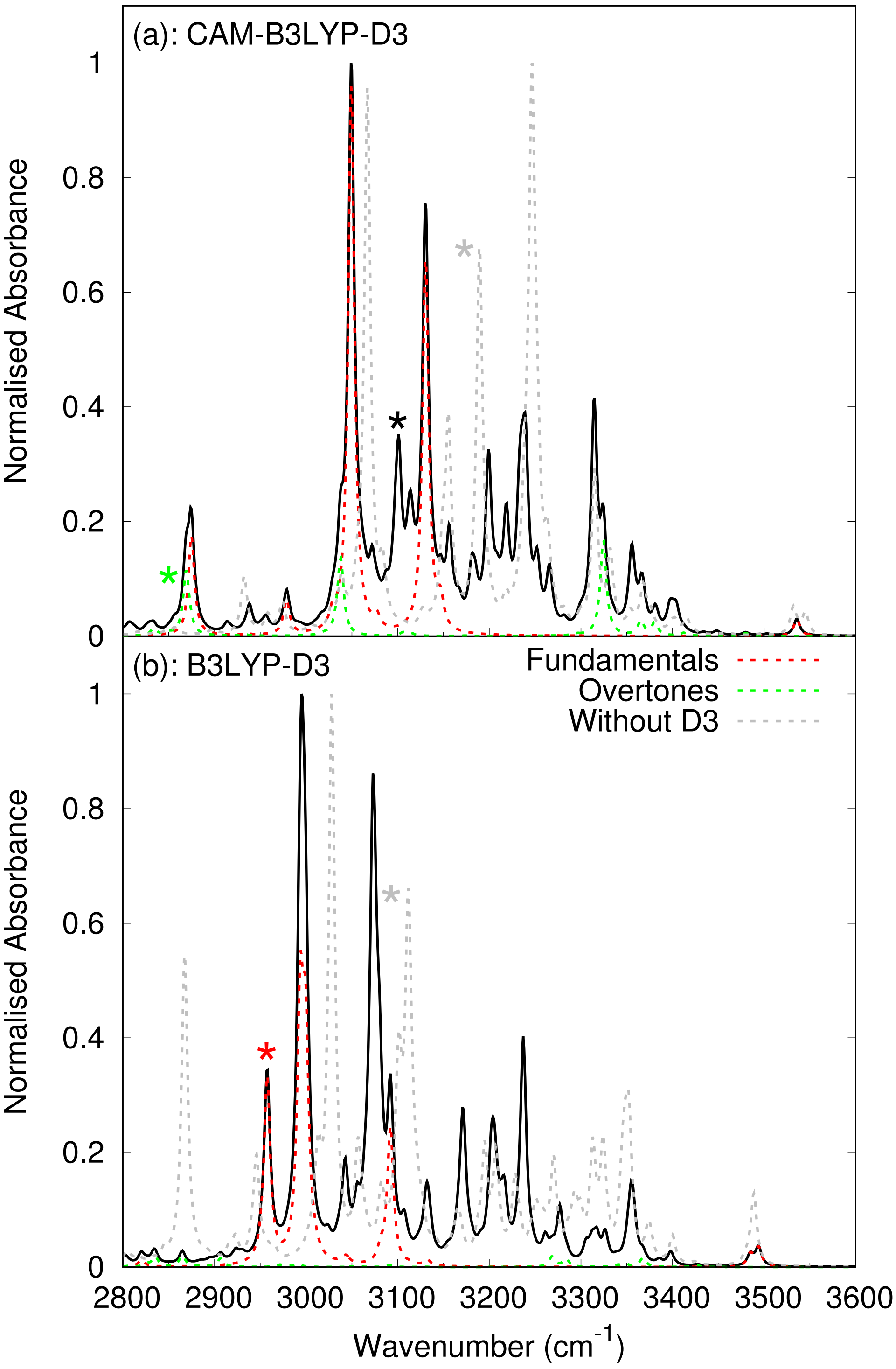

Figure S7: Anharmonic infrared spectra computed in chloroform for mG≡mC at the (a) CAM-B3LYP-D3 and (b) B3LYP-D3 level of theory with 6-31G(d) basis, where D3 indicates Grimme's D3 dispersion correction with Becke-Johnson damping.[8] The complete anharmonic spectrum is in black, the fundamentals in dashed red, overtones in dashed red, and spectra without the D3 dispersion correction (i.e. Figure 9 in the main text) in dashed grey. Possible anomalous peaks in the D3 corrected spectra are marked with a coloured *, the colour depending on whether it is a fundamental (red), overtone (green) or combination band (black), whilst the grey * indicates possible anomalous peaks in the spectra without D3 dispersion corrections.

Figure S7 illustrates the effect on the B3LYP and CAM-B3LYP spectra of including Grimme's D3 dispersion correction with Becke-Johnson damping.[8] The hydrogen bonded N-H stretch fundamental bands are each red shifted by 20–50 cm^{-1} due to an increase in HB strength for the dispersion corrected calculations, with all three HB distances shortening, as shown in Table S10. The dispersion correction also leads to a reduction in intensity of the combination bands, indicating that this may go some way to remedying the issue of overestimation of intensity. Furthermore, there are fewer anomalous combination band peaks due to the umbrella motion of the methyl groups on guanine and cytosine. However, there is an anomalous overtone on the CAM-B3LYP-D3 spectrum from a wagging motion of the methyl group on guanine, whilst for B3LYP-D3 there is an intense peak due to the symmetric and antisymmetric C-H stretches on the methyl group on guanine. Furthermore, for B3LYP-D3 the intensity due to the G(N1-H1) stretch is greatly underestimated, close to zero on the spectrum. Clearly, further investigation beyond the scope of this study is necessary regarding the introduction of dispersion effects on anharmonic spectra.

Table S10: Hydrogen bond distances (Å) for the mG≡mC model calculated by CAM-B3LYP and B3LYP in PCM only, with 1 and 2 explicit $CDCL_3$ molecules, and with D3 dispersion correction with Becke-Johnson damping[8] in PCM only.

Hydrogen Bond	PCM only	1 $CDCl_3$	2 $CDCL_3$	D3 dispersion
CAM-B3LYP				
G(N10-H10)-OC	1.840	1.870	1.859	1.824
C(N8'-H8')-OC	1.823	1.817	1.829	1.807
G(N1-H1)-N	1.900	1.912	1.905	1.881
B3LYP				
G(N10-H10)-OC	1.868	1.902	1.888	1.835
C(N8'-H8')-OC	1.847	1.838	1.855	1.810
G(N1-H1)-N	1.921	1.931	1.926	1.878

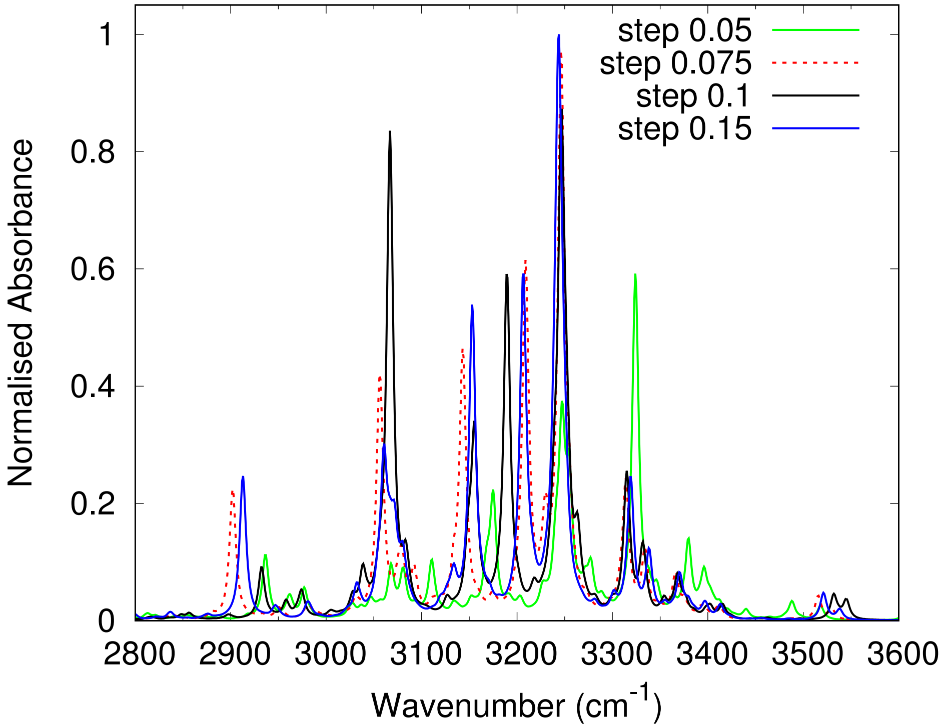

Figure S8: Anharmonic infrared spectra computed in chloroform for mG≡mC at the CAM-B3LYP/6-31G(d) level, by using different step size (in Å) in the numeric differentiation in the anharmonic calculations.

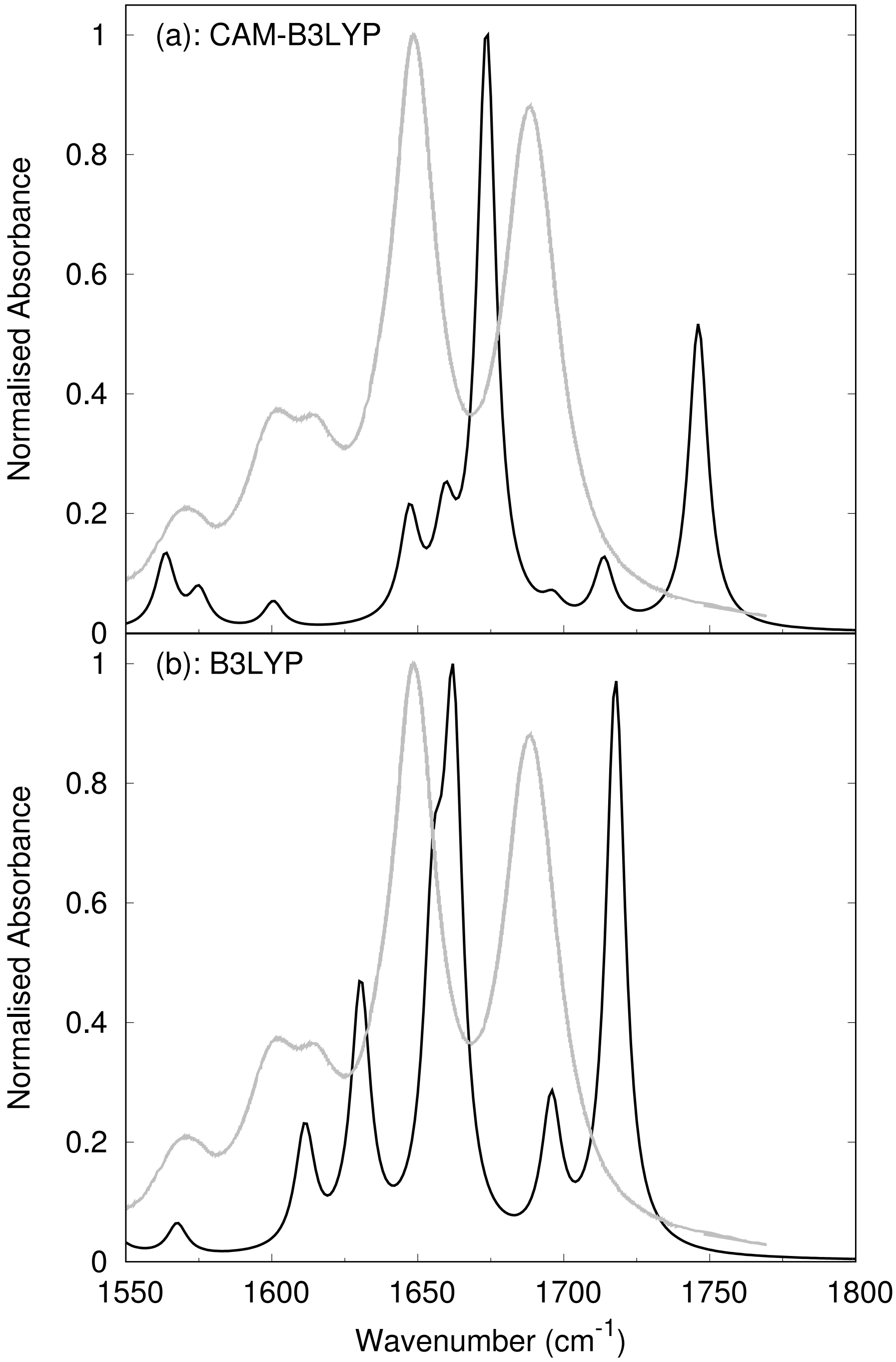

Figure S9: Anharmonic infrared spectra computed in PCM(chloroform) in the fingerprint region for mG≡mC at the (a) CAM-B3LYP and (b) B3LYP level of theory with a 6-31G(d) basis. Experimental spectrum in grey.[3]

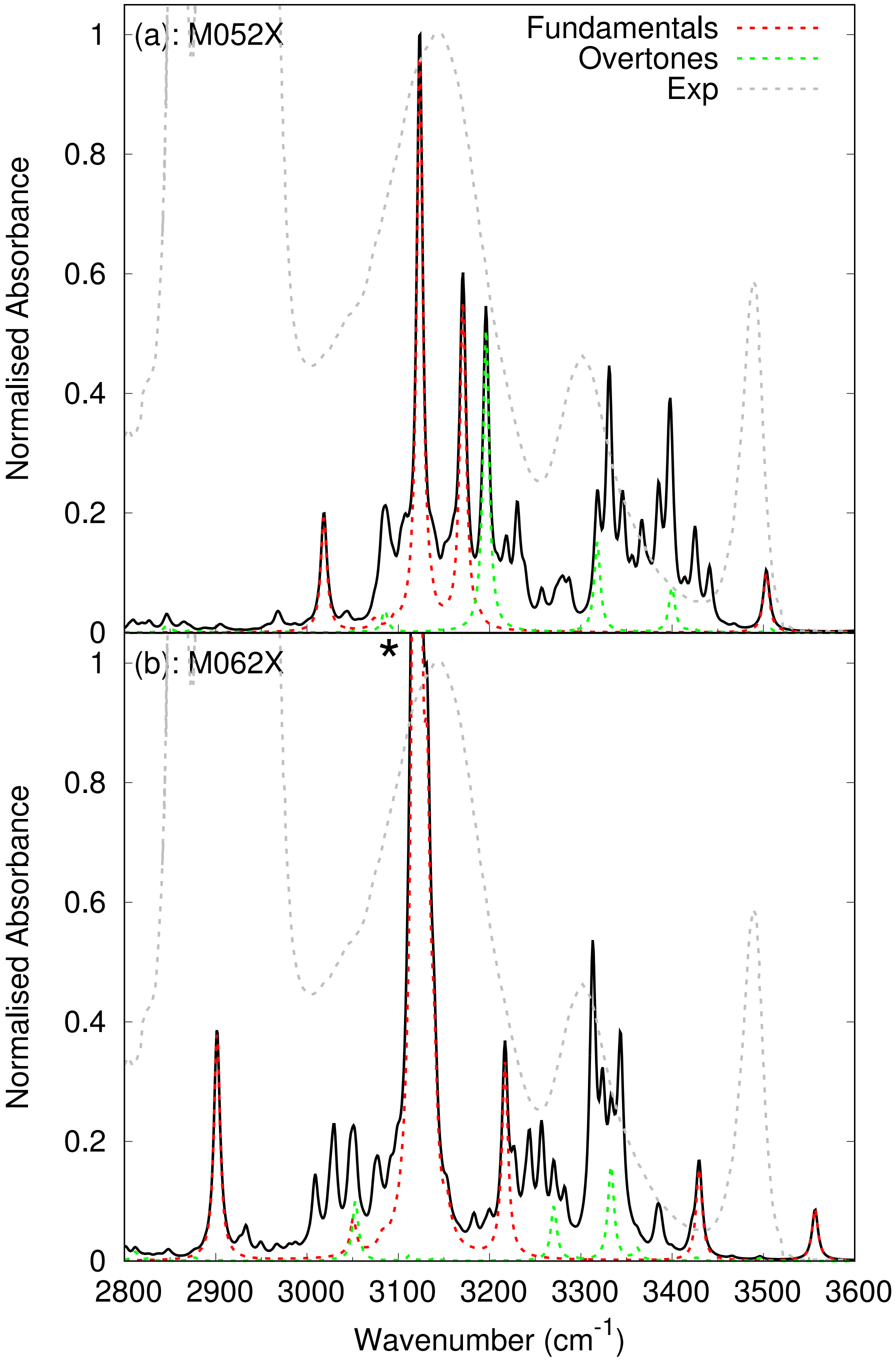

Figure S10: Anharmonic infrared spectra computed in PCM(chloroform) for mG≡mC at the (a) M052X and (b) M062X level of theory with a 6-31G(d) basis. Fundamentals are shown with a red dashed line, overtones with a green dashed line, and experimental spectrum with a grey dashed line.[2] The portion of the M062X spectrum marked with a * that goes off the scale is an anomalous peak due to the choice of computational model. The spectrum is normalised to the portion of the same peak due to a non-anomalous vibration (as it is the most intense vibration).

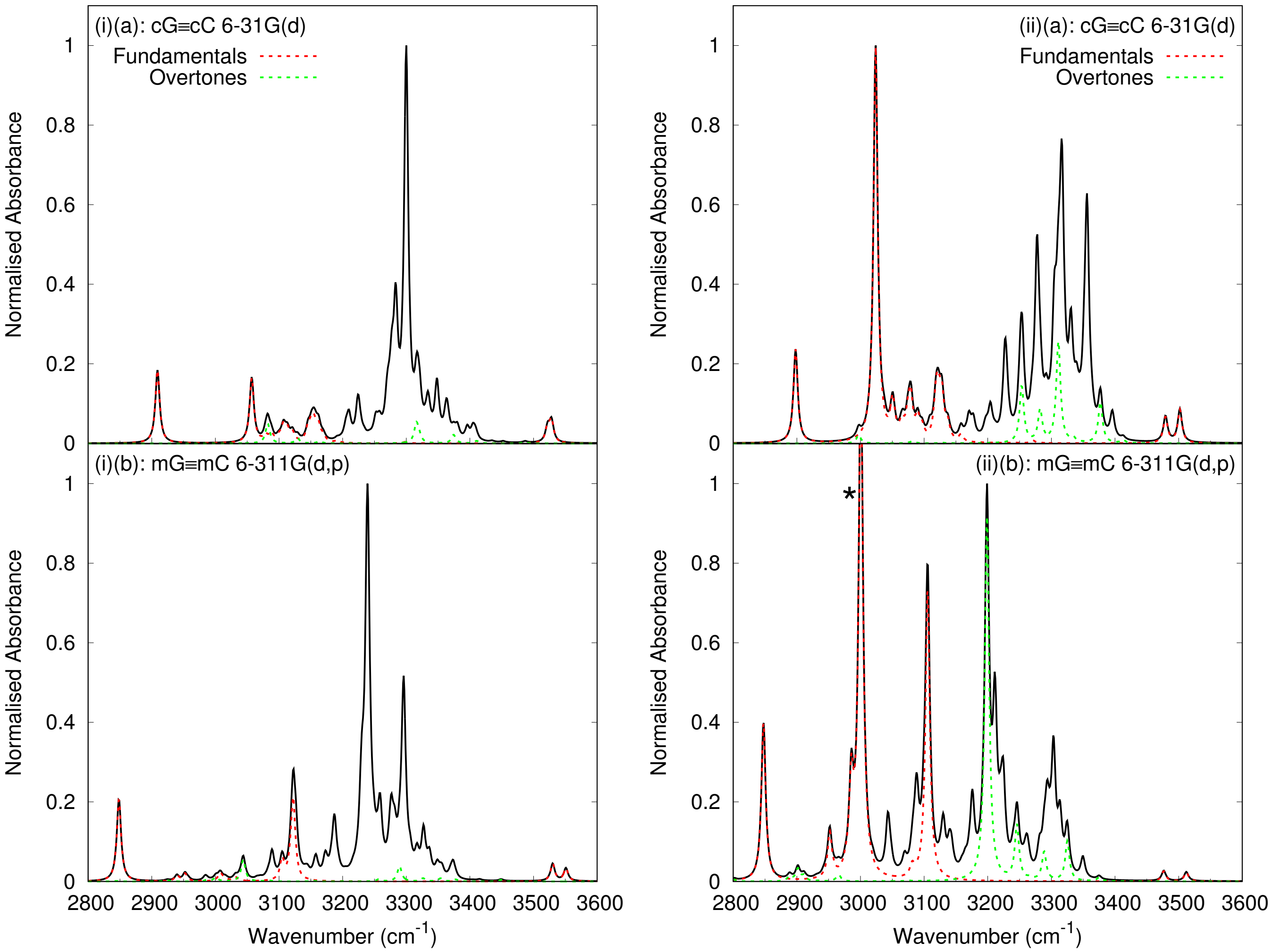

Figure S11: Anharmonic infrared spectra computed in PCM(chloroform) in the NH stretch region for (a) cG≡cC model with a 6-31G(d) basis and (b) the mG≡mC model with a 6-311G(d,p) basis at the (i) CAM-B3LYP and (ii) B3LYP level of theory. The * indicates a possible anomalous peak due to the choice of computational model, and the spectra are normalised to non-anomalous peaks.

2.3.1 Explicit chloroform

Protons of chloroform solvent molecules are slightly acid and structures where solute/solvent HB are present are possible. Actually, our preliminary MD analysis in explicit solvent molecules indicate that 2/3 solvent molecules are more strongly bound to the solute, the two most important interactions involving the G and the C carbonyl. As a first step towards the inclusion of solute/solvent interactions in our vibrational treatment, we have thus studied two computational models (see Figure S12), where one or two $CDCl_3$ molecules are explicitly included in our calculations (N.B. we included $CDCl_3$ rather than $CHCl_3$ as this was the solvent used experimentally,[2] without allowing exchange of the nucleoside protons).

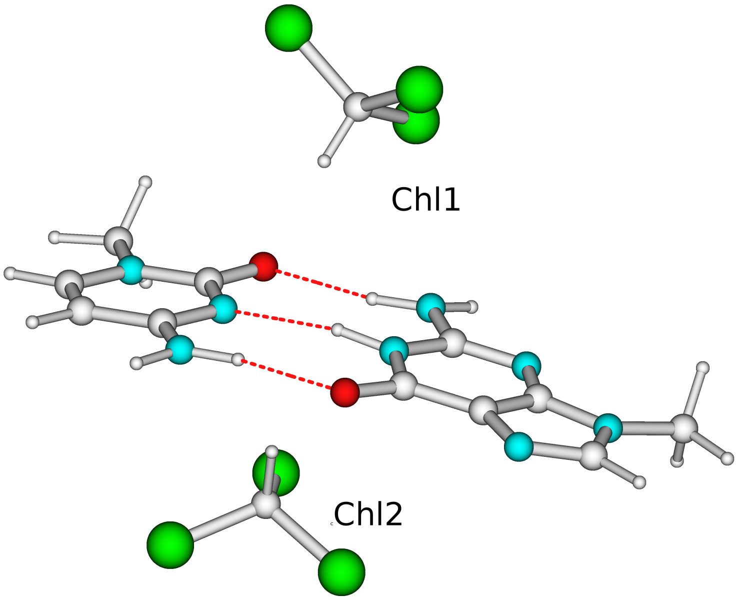

Figure S12: Computational models including one (only Chl1 is present) and two (both Chl1 and Chl2 are considered) explicit $CDCl_3$ molecules

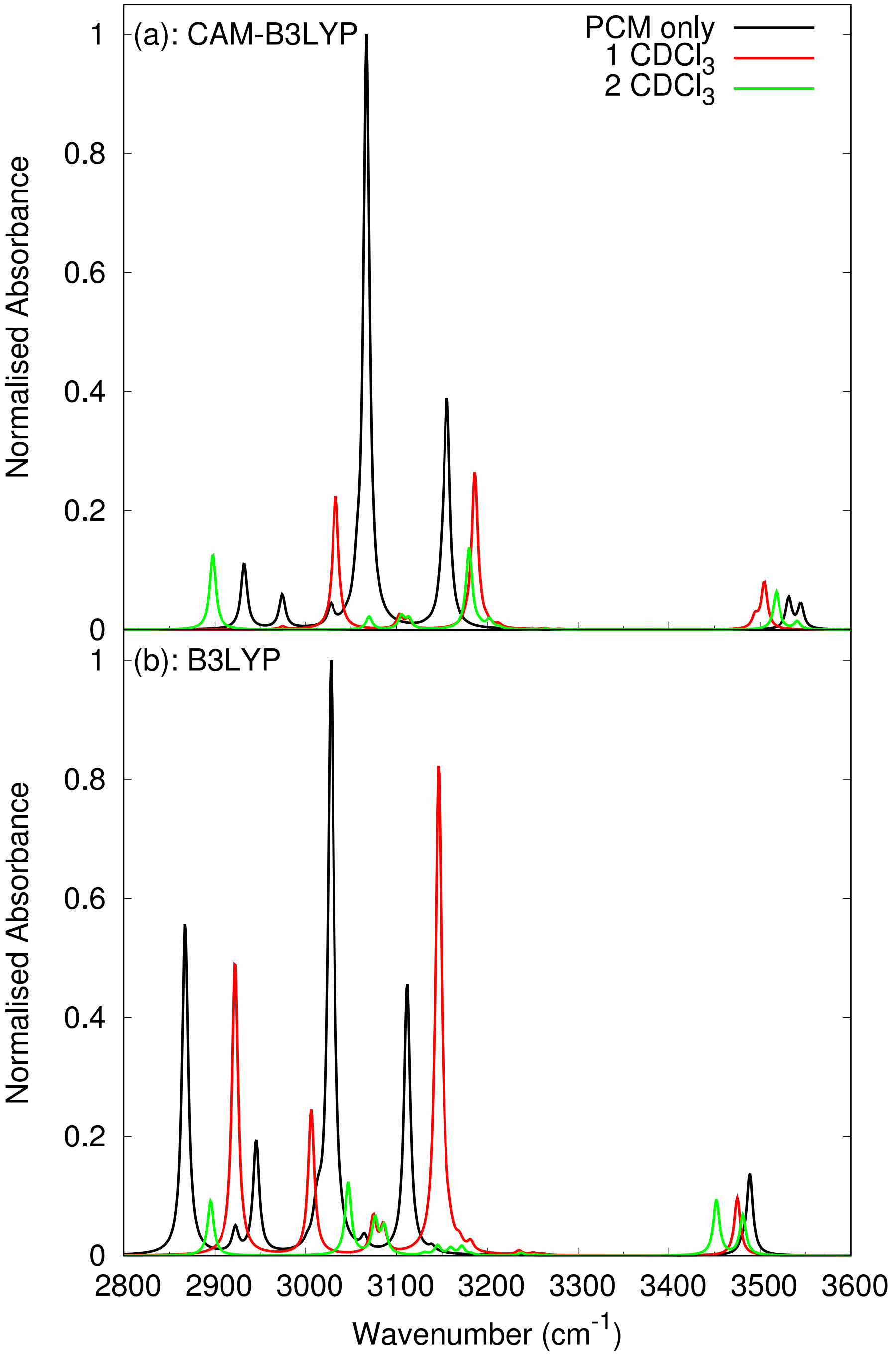

Figure S13: Anharmonic IR spectra showing only the fundamentals in the N-H stretching region, computed for mG≡mC at the (a) CAM-B3LYP and (b) B3LYP level of theory with a 6-31G(d) basis, with chloroform solvent modelled by PCM only (as in main text), and by PCM with 1 or 2 explicit $CDCl_3$ molecules, with positions shown in figure S12. Calculated spectra normalised to most intense peak on the PCM only calculations.

2.3.2 Deuterated systems

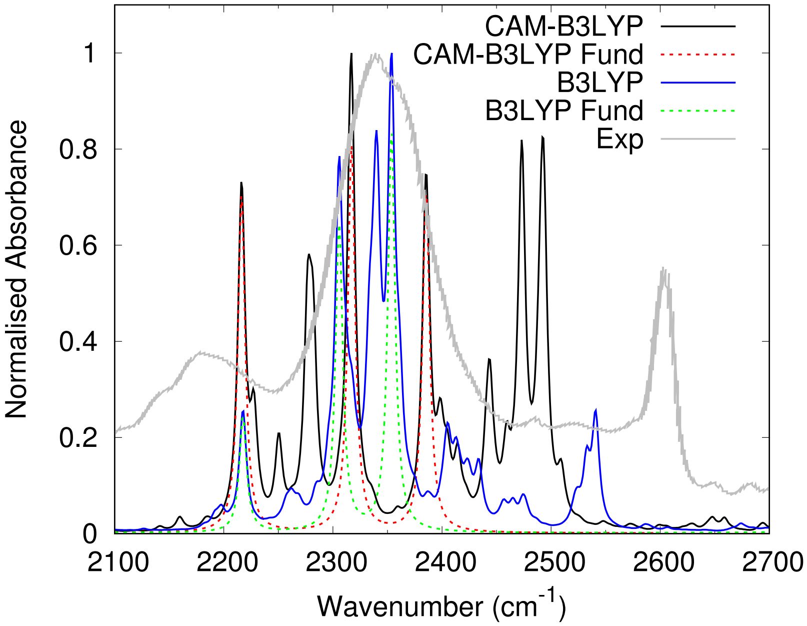

Figure S14: Anharmonic infrared spectra computed in PCM(chloroform) for partially deuterated mG≡mC at the CAM-B3LYP (black, with fundamentals in dashed red) and B3LYP (blue, with fundamentals in dashed green) level of theory with a 6-31G(d) basis. Experimental spectrum in grey.[3] Assignments of fundamentals in table S9

References

[1] N. K. Schwalb, T. Michalak and F. Temps, *J. Phys. Chem. B*, 2009, **113**, 16365.

[2] M. Yang, L. Szyc, K. Röttger, H. Fidder, E. T. J. Nibbering, T. Elsaesser and F. Temps, *J. Phys. Chem. B*, 2011, **115**, 5484.

[3] H. Fidder, M. Yang, E. T. J. Nibbering, T. Elsaesser, K. Röttger and F. Temps, *J. Phys. Chem. B*, 2013, **117**, 845.

[4] R. A. Ingle, G. M. Roberts, K. Röttger, H. J. Marroux, F. Dönnichsen, M. Yang, L. Szyc, Y. Harabuchi, S. Maeda, F. Temps and A. J. Orr-Ewing, *Chem. Phys.*, 2018, **515**, 480 – 492.

[5] K. Röttger, H. J. B. Marroux, M. P. Grubb, P. M. Coulter, H. Bohnke, A. S. Henderson, M. C. Galan, F. Temps, A. J. Orr-Ewing and G. M. Roberts, *Angew. Chem., Int. Ed.*, 2015, **54**, 14719.

[6] M. Szczesniak, M. Nowak and K. Szczepaniak, *J. Mol. Struct.*, 1984, **115**, 221–224.

[7] K. Kuczera, M. Szczesniak and K. Szczepaniak, *J. Mol. Struct.*, 1988, **172**, 73 – 88.

[8] S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.