Electronic Supplementary Material (ESI) for RSC Advances.

Supplementary Materials

Structure-based design and synthesis of a novel long-chain 4"-alkyl ether derivative of EGCG as potent EGFR inhibitor: *In-vitro and in-silico* studies

Satyam Singh[1], Revathy Sahadevan[2], Rajarshi Roy[1], Mainak Biswas[3], Priya Ghosh[1], Parimal Kar[1], Avinash Sonawane[1], Sushabhan Sadhukhan[2,4,5*]

[1]Department of Biosciences and Biomedical Engineering, Indian Institute of Technology Indore, Madhya Pradesh 453 552, India.

[2]Department of Chemistry, Indian Institute of Technology Palakkad, Kerala 678 623, India.

[3]School of Biotechnology, KIIT Deemed to be University, Bhubaneswar, Orissa 751 024, India.

[4]Physical & Chemical Biology Laboratory, Indian Institute of Technology Palakkad, Kerala 678 623, India.

[5]Department of Biological Sciences & Engineering, Indian Institute of Technology Palakkad, Kerala 678 623, India.

*Corresponding author. *E-mail address*: sushabhan@iitpkd.ac.in

Table of Contents

1. Characterization of 4"-alkyl EGCG derivatives

1. 4"-hexyl EGCG (4"-C_6 EGCG)

The compound was obtained with 60 % yield. MS (ESI): m/z calculated for $[C_{28}H_{30}O_{11} + H]^+$ = 543.18, observed = 543.25. FTIR (cm^{-1}): 3354, 2963, 2924, 1690, 1605, 1520, 1458, 1342, 1258, 1234, 1188, 1142, 1096, 1020, 857, 802, 733, 624, 532. ^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm)= 0.84 (3H, t, *J* = 7.05 Hz), 1.24-1.35 (6H, overlapping), 1.61 (2H, m), 2.65 (1H, dd, *J* = 17.00 Hz, 4.20 Hz), 2.93 (1H, dd, *J* = 17.32 Hz, 4.00 Hz), 3.89 (2H, t, *J* = 6.80 Hz), 4.96 (1H, m), 5.38 (1H, bs), 5.83 (1H, bs), 5.92 (1H, bs), 6.40 (2H, s), 6.81 (2H, s). ^{13}C NMR (125 MHz, DMSO-d_6): δ (ppm)= 165.44, 156.97, 156.92, 156.01, 151.11, 146.08, 139.42, 132.82, 129.05, 124.59, 109.15, 105.92, 97.78, 96.04, 94.82, 76.84, 72.31, 68.91, 31.51, 31.10, 29.80, 26.11, 25.35, 22.49, 14.34.

2. 4"-octyl EGCG (4"-C_8 EGCG)

The compound was obtained with 66 % yield. MS (ESI): m/z calculated for $[C_{30}H_{34}O_{11} + H]^+$ = 571.21, observed = 571.30. FTIR (cm^{-1}): 3348, 2924, 2854, 1690, 1605, 1520, 1458, 1342, 1234, 1188, 1143, 1095, 1025, 863, 818, 764, 726, 463. ^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm)= 0.84 (3H, t, *J* = 7.05 Hz), 1.22-1.27 (8H, overlapping), 1.33 (2H, m), 1.61 (2H, m), 2.64 (1H, dd, *J* = 16.65 Hz, 3.50 Hz), 2.93 (1H, dd, *J* = 16.65 Hz, 3.50 Hz), 3.88 (2H, t, *J* = 6.80 Hz), 4.97 (1H, m), 5.37 (1H, bs), 5.83 (1H, bs), 5.92 (1H, bs), 6.40 (2H, s), 6.81 (2H, s). ^{13}C NMR (125 MHz, DMSO-d_6): δ (ppm)= 165.43, 156.92, 156.88, 156.01, 151.11, 146.08, 139.41, 132.82, 129.04, 124.58, 109.15, 105.92, 97.78, 96.04, 94.82, 76.84, 72.28, 68.90, 31.66, 30.00, 29.84, 29.09, 25.70, 22.51, 14.38.

3. 4"-decyl EGCG (4"-C_{10} EGCG)

The compound was obtained with 72 % yield. MS (ESI): m/z calculated for $[C_{32}H_{38}O_{11} + H]^+$ = 599.24, observed = 599.35. FTIR (cm^{-1}): 3350, 2924, 2855, 1690, 1605, 1520, 1458, 1342, 1234, 1188, 1140, 1095, 1024, 865, 810, 764, 717, 520. ^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm)= 0.84 (3H, t, *J* = 7.05 Hz), 1.22-1.32 (12H, overlapping), 1.35 (2H, m), 1.61 (2H, m), 2.65 (1H, dd, *J* = 17.50 Hz, 3.98 Hz), 2.93 (1H, dd, *J* = 16.80 Hz, 4.00 Hz), 3.88 (2H, t, *J* = 6.80), 4.95 (1H, m), 5.38 (1H, bs), 5.83 (1H, bs), 5.93 (1H, bs), 6.40 (2H, s), 6.81 (2H, s). ^{13}C NMR (125 MHz, DMSO-d_6): δ (ppm)= 165.43, 156.97, 156.92, 156.01, 151.10, 146.08,

139.41, 132.82, 129.05, 124.59, 109.15, 105.92, 97.79, 96.04, 94.82, 76.84, 72.30, 68.90, 31.72, 29.84, 29.38, 29.12, 26.11, 25.69, 22.52, 14.38.

4. 4" dodecyl-EGCG (4"-C_{12} EGCG)

The compound was obtained with 75 % yield. MS (ESI): m/z calculated for $[C_{34}H_{42}O_{11} + H]^+$ = 627.27, observed = 627.35. FTIR (cm^{-1}): 3356, 2924, 2855, 1690, 1605, 1520, 1458, 1342, 1234, 1188, 1141, 1095, 1018, 865, 810, 764, 724, 455. ^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm)= 0.85 (3H, t, *J* = 7.05 Hz), 1.18-1.23 (16H, overlapping), 1.33 (2H, m), 1.61 (2H, m), 2.65 (1H, dd, *J* = 21.80 Hz, 4.70 Hz), 2.93 (1H, dd, *J* = 17.00 Hz, 4.20 Hz), 3.88 (2H, t, *J* = 6.80 Hz), 4.96 (1H, m), 5.38 (1H, bs), 5.83 (1H, bs), 5.93 (1H, bs), 6.40 (2H, s), 6.81 (2H, s). ^{13}C NMR (125 MHz, DMSO-d_6): δ (ppm)= 165.42, 156.99, 156.93, 156.03, 151.12, 146.10, 139.42, 132.83, 129.02, 124.59, 109.15, 105.92, 97.77, 96.03, 94.82, 76.84, 72.29, 68.87, 31.74, 30.03, 29.87, 29.45, 29.14, 25.72, 22.53, 14.40.

5. 4" tetradecyl-EGCG (4"-C_{14} EGCG)

The compound was obtained with 70 % yield. MS (ESI): m/z calculated for $[C_{36}H_{46}O_{11} + H]^+$ = 655.30, observed = 655.40. FTIR (cm^{-1}): 3356, 2924, 2854, 1689, 1604, 1520, 1458, 1342, 1234, 1188, 1141, 1095, 1030, 865, 802, 763, 624, 527. ^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm)= 0.85 (3H, t, *J* = 7.05 Hz), 1.18-1.30 (20H, overlapping), 1.35 (2H, m), 1.61 (2H, m), 2.65 (1H, dd, *J* = 19.20 Hz, 3.20 Hz), 2.93 (1H, dd, *J* = 17.40 Hz, 3.90 Hz), 3.88 (2H, t, *J* = 6.80 Hz), 4.97 (1H, m), 5.38 (1H, bs), 5.83 (1H, bs), 5.93 (1H, bs), 6.40 (2H, s), 6.81 (2H, s). ^{13}C NMR (125 MHz, DMSO-d_6): δ (ppm)= 165.42, 156.99, 156.93, 156.02, 151.11, 146.09, 139.42, 132.82, 129.04, 124.59, 109.15, 105.92, 97.78, 96.04, 94.82, 76.84, 72.29, 68.89, 31.73, 31.11, 29.87, 29.48, 29.44, 25.72, 22.53, 14.38. A combination of ^{1}H and ^{13}C NMR experiment was done by means of Heteronuclear Multiple-Bond Correlation method (HMBC) as shown in Fig. 2. The cross peak connecting 4" carbon and the protons from the -OCH_2- of the alkyl chain (as highlighted in the inset of Fig. 2) confirms that they are directly connected through the ethereal bond.

6. 4"-hexadecyl EGCG (4"-C_{16} EGCG)

The compound was obtained with 50 % yield. MS (ESI): m/z calculated for $[C_{38}H_{50}O_{11} + H]^+$ = 683.34, observed = 683.35. FTIR (cm^{-1}): 3364, 2924, 2855, 1690, 1605, 1520, 1458,

1342, 1234, 1155, 1142, 1095, 1034, 868, 818, 764, 718, 462. 1H NMR (500 MHz, DMSO-d_6): δ (ppm)= 0.85 (3H, t, *J* = 7.05 Hz), 1.23 (24H, overlapping), 1.34 (2H, m), 1.61 (2H, m), 2.65 (1H, dd, *J* = 15.40 Hz, 3.35 Hz), 2.93 (1H, dd, *J* = 16.70 Hz, 3.85 Hz), 3.88 (2H, t, *J* = 7.00 Hz), 4.96 (1H, m), 5.38 (1H, bs), 5.82 (1H, bs), 5.93 (1H, bs), 6.40 (2H, s), 6.81 (2H, s). ^{13}C NMR (125 MHz, DMSO-d_6): δ (ppm)= 165.44, 156.95, 156.91, 155.99, 151.08, 146.05, 139.41, 132.80, 129.05, 124.58, 109.14, 105.92, 97.79, 96.04, 94.82, 76.84, 72.31, 68.91, 31.71, 29.96, 29.84, 29.44, 29.10, 25.69, 22.51, 14.36.

7. 4"-octadecyl EGCG (4"-C_{18} EGCG)

The compound was obtained with 52 % yield. MS (ESI): m/z calculated for [$C_{40}H_{54}O_{11}$ + H]$^+$ = 711.37, observed = 711.35. FTIR (cm^{-1}): 3364, 2924, 2855, 1690, 1605, 1520, 1458, 1342, 1234, 1188, 1142, 1095, 1034, 866, 818, 764, 717, 463. ^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm)= 0.85 (3H, t, *J* = 6.65 Hz), 1.19-1.29 (28H, overlapping), 1.33 (2H, m), 1.61 (2H, m), 2.66 (1H, dd, *J* = 14.50 Hz, 4.10 Hz), 2.93 (1H, dd, *J* = 14.90 Hz, 4.15 Hz), 3.88 (2H, t, *J* = 7.00 Hz), 4.98 (1H, m), 5.38 (1H, bs), 5.82 (1H, bs), 5.93 (1H, bs), 6.40 (2H, s), 6.81 (2H, s). ^{13}C NMR (125 MHz, DMSO-d_6): δ (ppm)= 165.44, 156.93, 155.76, 151.09, 146.06, 139.41, 132.80, 129.04, 124.88, 109.14, 105.92, 96.62, 94.91, 72.30, 31.70, 29.97, 29.83, 29.42, 29.40, 29.27, 25.69, 22.51, 14.37.

2. HPLC analysis of 4"-alkyl EGCG derivatives

Purity analysis was carried out using Shimadzu HPLC-20AP instrument with a Sprite TARGA C18 column (40 × 2.1 mm, 5 µm) monitoring at 210 and 254 nm. Solvents for the HPLC were water with 0.1% acetic acid (solvent A) and acetonitrile with 0.1% acetic acid (solvent B). Compounds were eluted at a flow rate of 0.7 mL/min with a gradient of 5% solvent B for 1 min, followed by a linear gradient from 5% to 60% solvent B over 4 min, continued for 1 min, followed by changing the solvent B from 60% to 95% in 3 min, continued for another 3 min, and finally, it was brought down to 5% solvent B in 1.5 min and then continued for another 1.5 min before the method stopped. Column was always washed with 50% Solvent B followed by 95% Solvent B before sample injection.

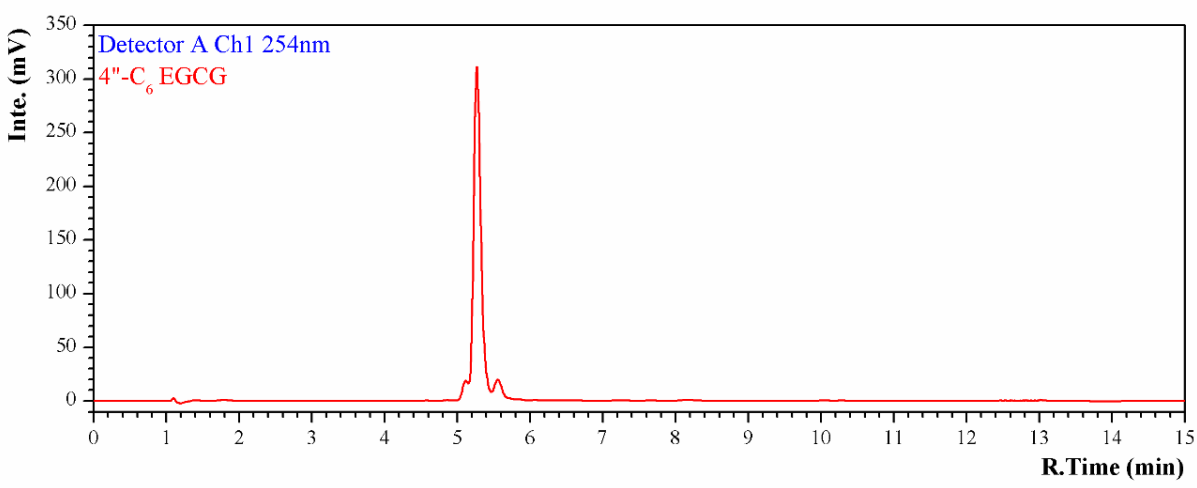


Fig. S1. HPLC chromatogram of 4"-C_6 EGCG.

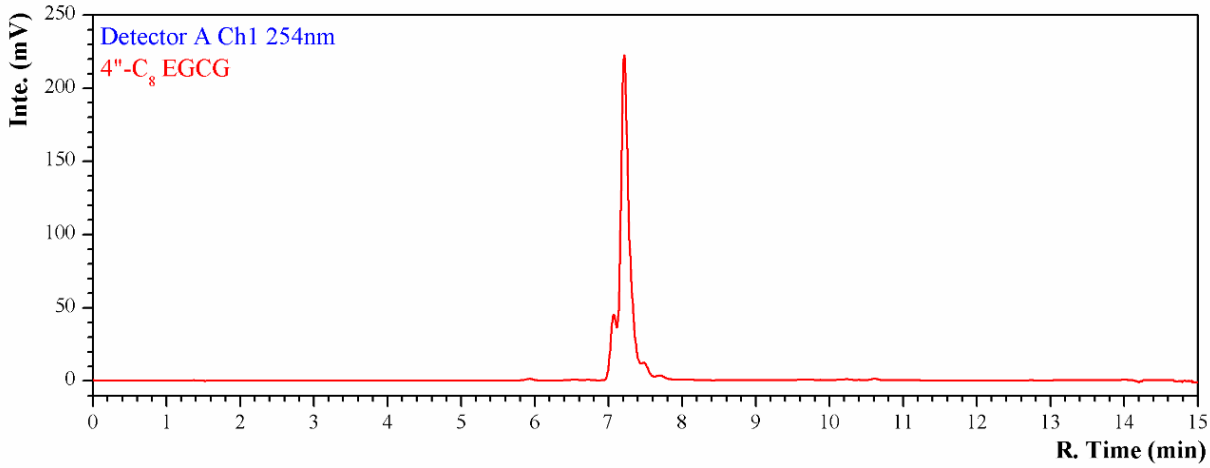


Fig. S2. HPLC chromatogram of 4"-C_8 EGCG.

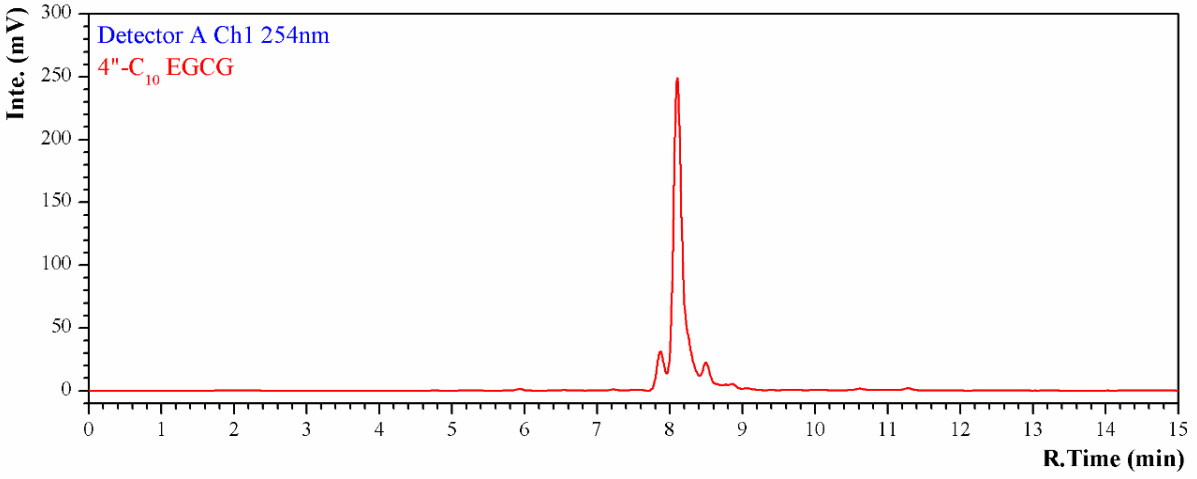


Fig. S3. HPLC chromatogram of 4"-C_{10} EGCG.

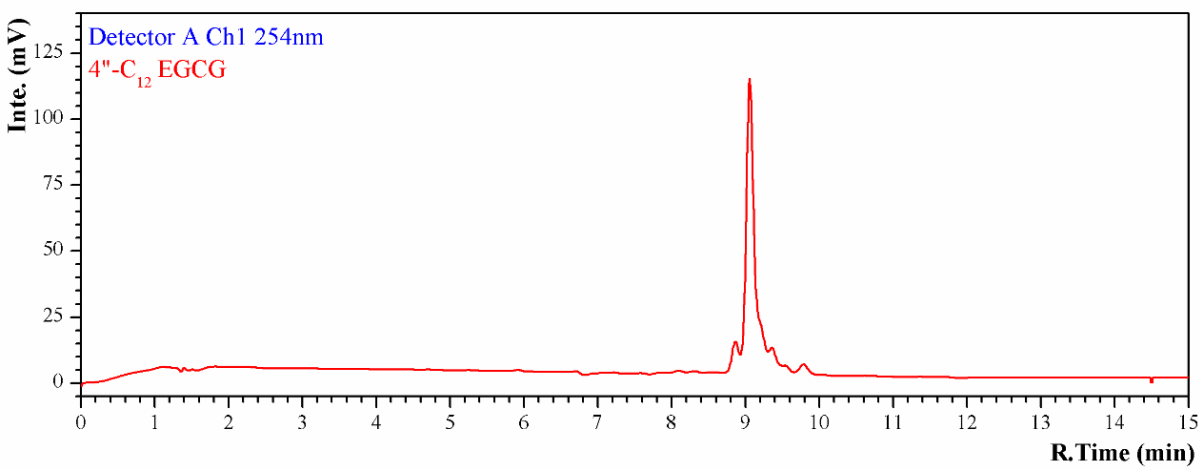


Fig. S4. HPLC chromatogram of 4"-C_{12} EGCG.

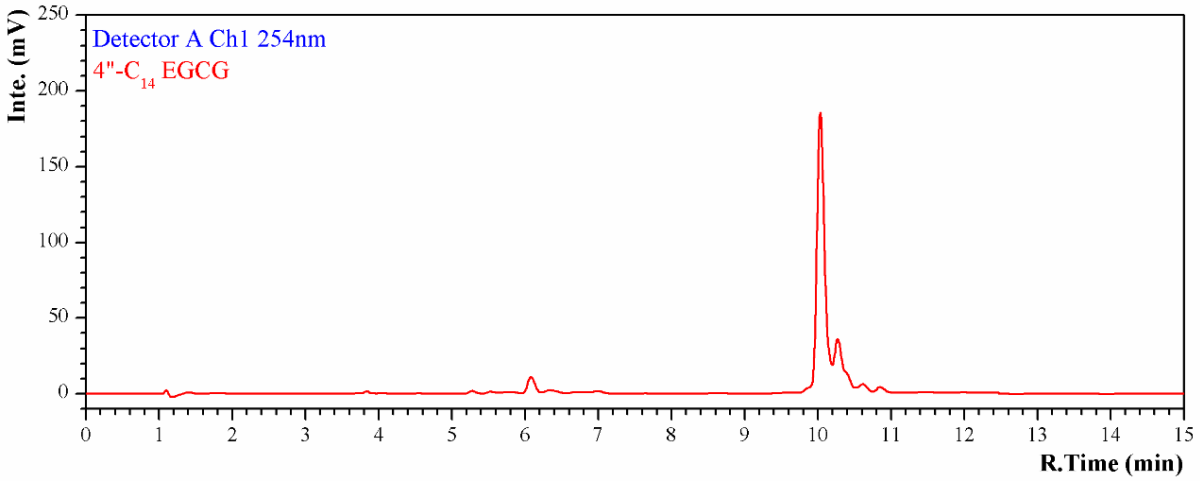


Fig. S5. HPLC chromatogram of 4"-C_{14} EGCG.

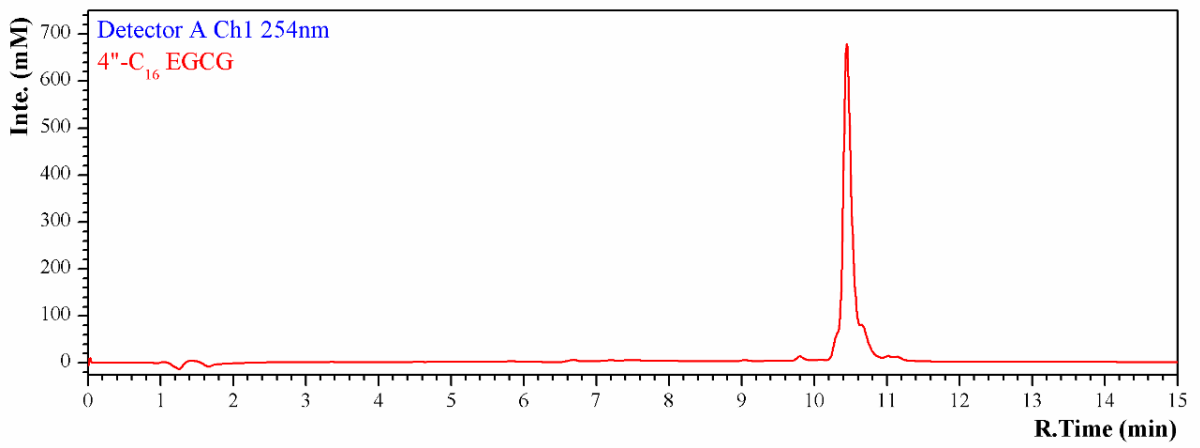


Fig. S6. HPLC chromatogram of 4"-C_{16} EGCG.

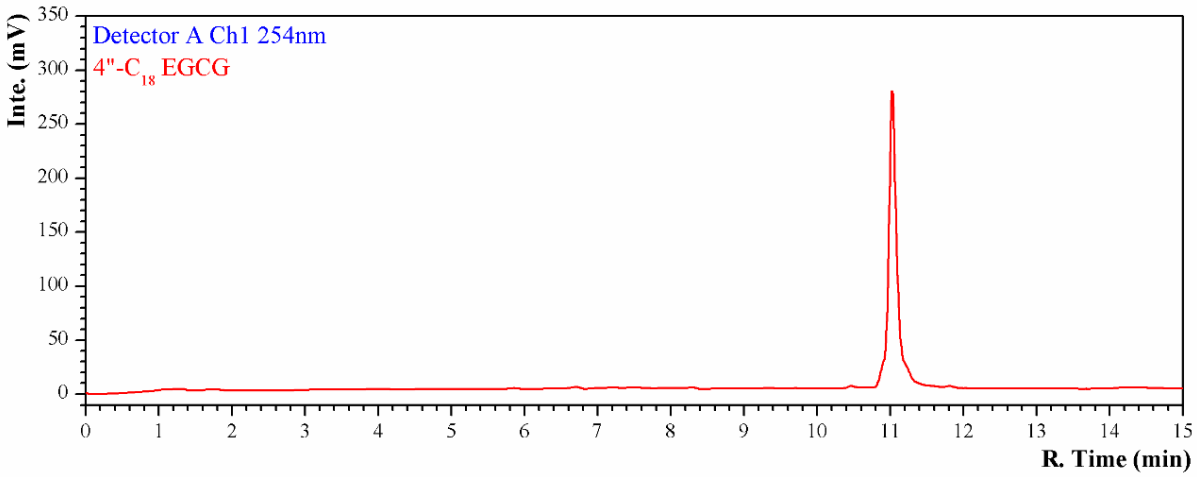


Fig. S7. HPLC chromatogram of 4"-C_{18} EGCG.

3. *Mass spectra of 4''-alkyl EGCG derivatives*

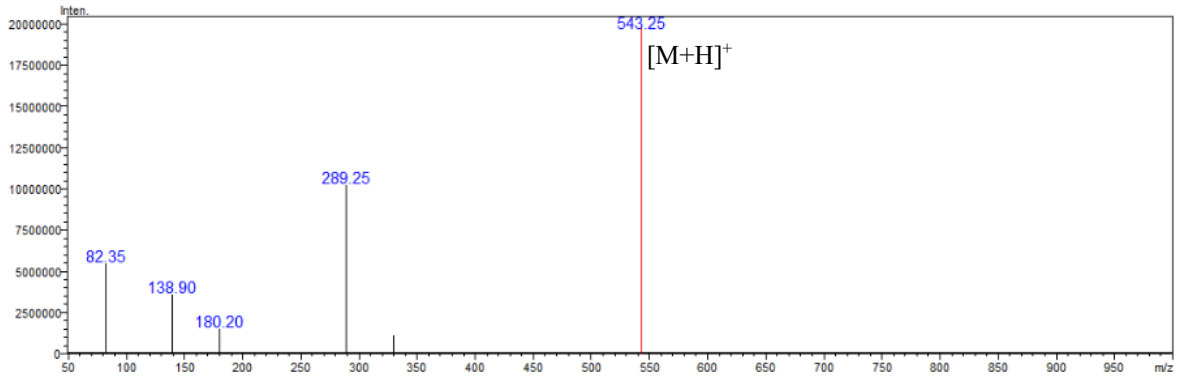


Fig. S8. Mass spectrum of 4"-C_6 EGCG (Exact mass 542.18 Da).

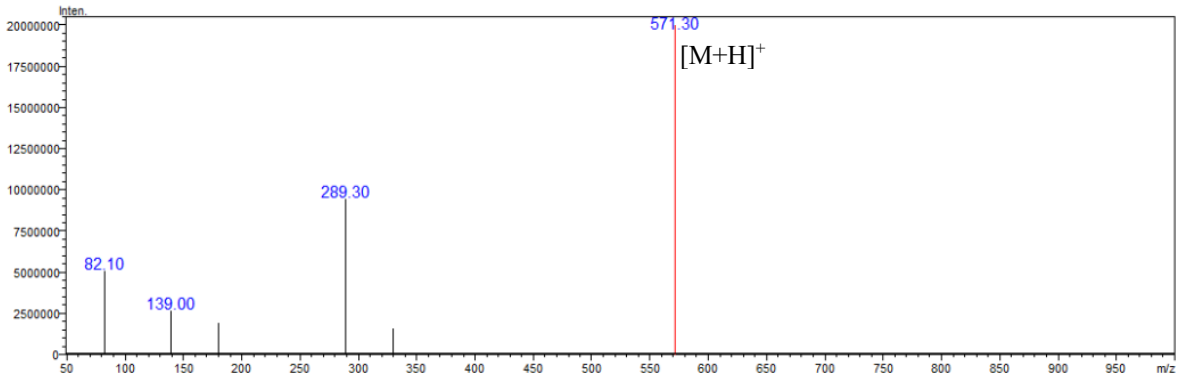


Fig. S9. Mass spectrum of 4"-C_8 EGCG (Exact mass 570.21 Da).

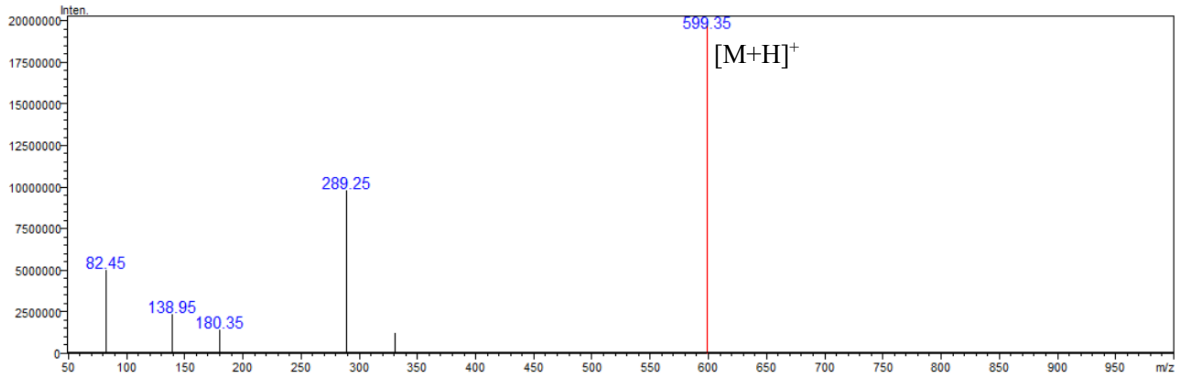


Fig. S10. Mass spectrum of 4"-C_{10} EGCG (Exact mass 598.24 Da).

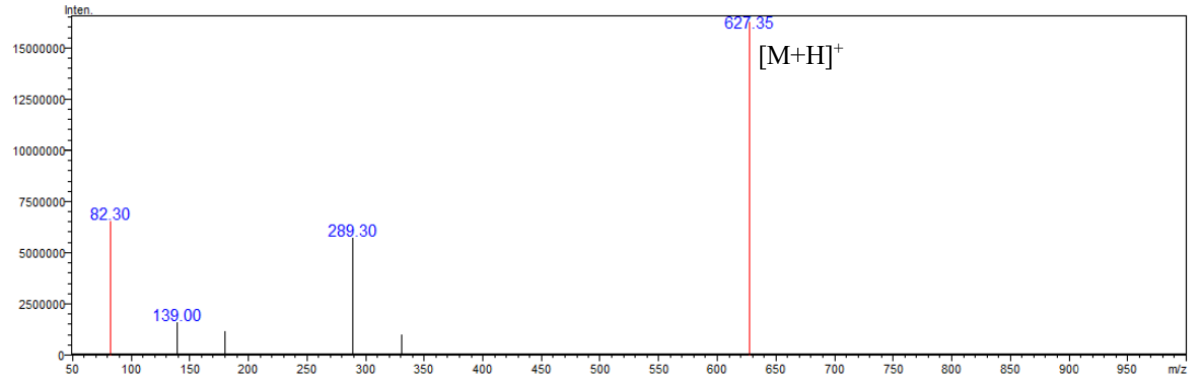


Fig. S11. Mass spectrum of 4"-C_{12} EGCG (Exact mass 626.27 Da).

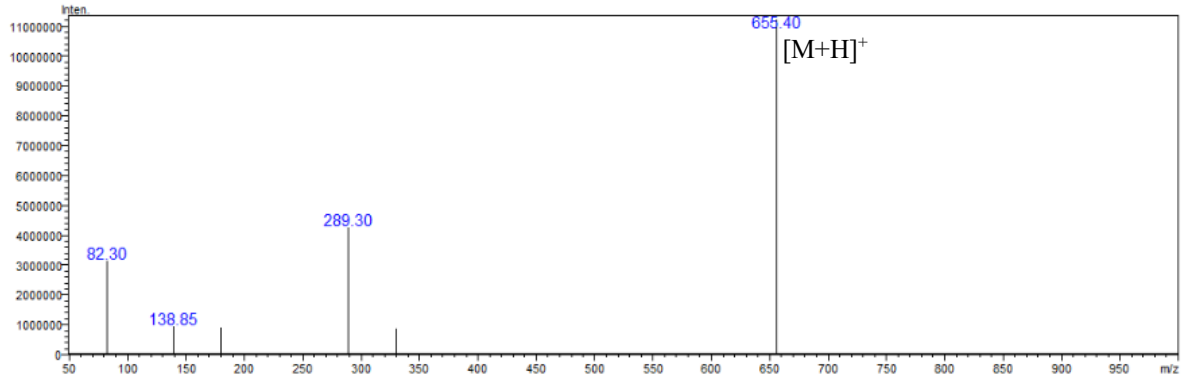


Fig. S12. Mass spectrum of 4"-C_{14} EGCG (Exact mass 654.30 Da).

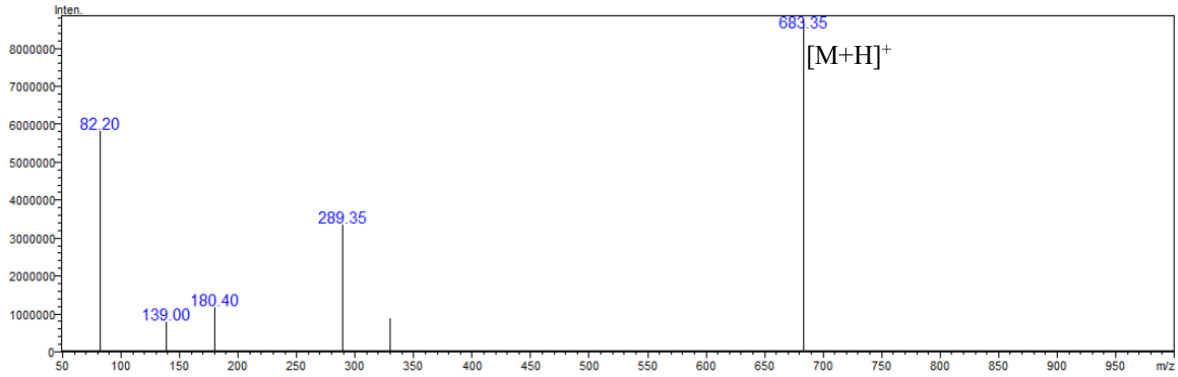


Fig. S13. Mass spectrum of 4"-C_{16} EGCG (Exact mass 682.34 Da).

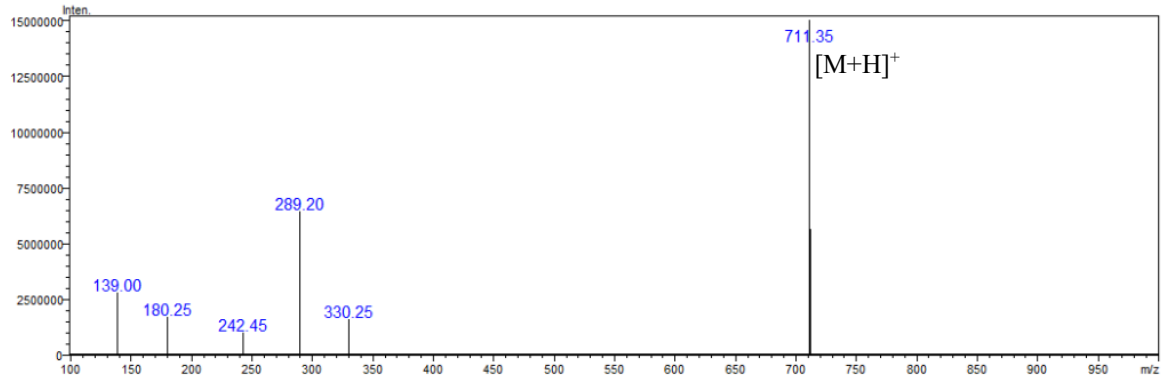


Fig. S14. Mass spectrum of 4"-C_{18} EGCG (Exact mass 710.37 Da)

4. ^{1}H NMR spectra of 4''-alkyl EGCG derivatives

^{1}H, ^{13}C, and HMBC NMR were recorded in Bruker AVANCE III 500 FT NMR spectrometer. Peaks at 3.5, 2.5 and 2.09 ppm are from residual H_2O, DMSO, acetone respectively.

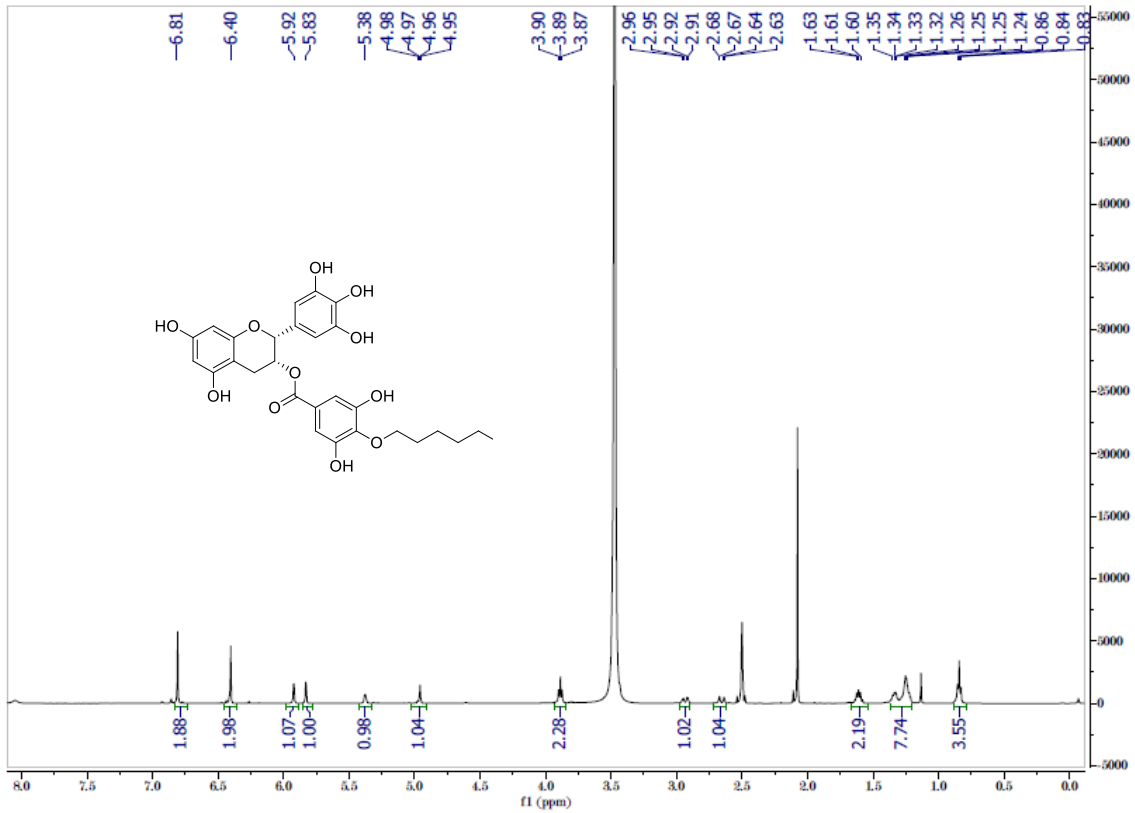


Fig. S15. ^{1}H NMR spectrum of 4"-C_6 EGCG.

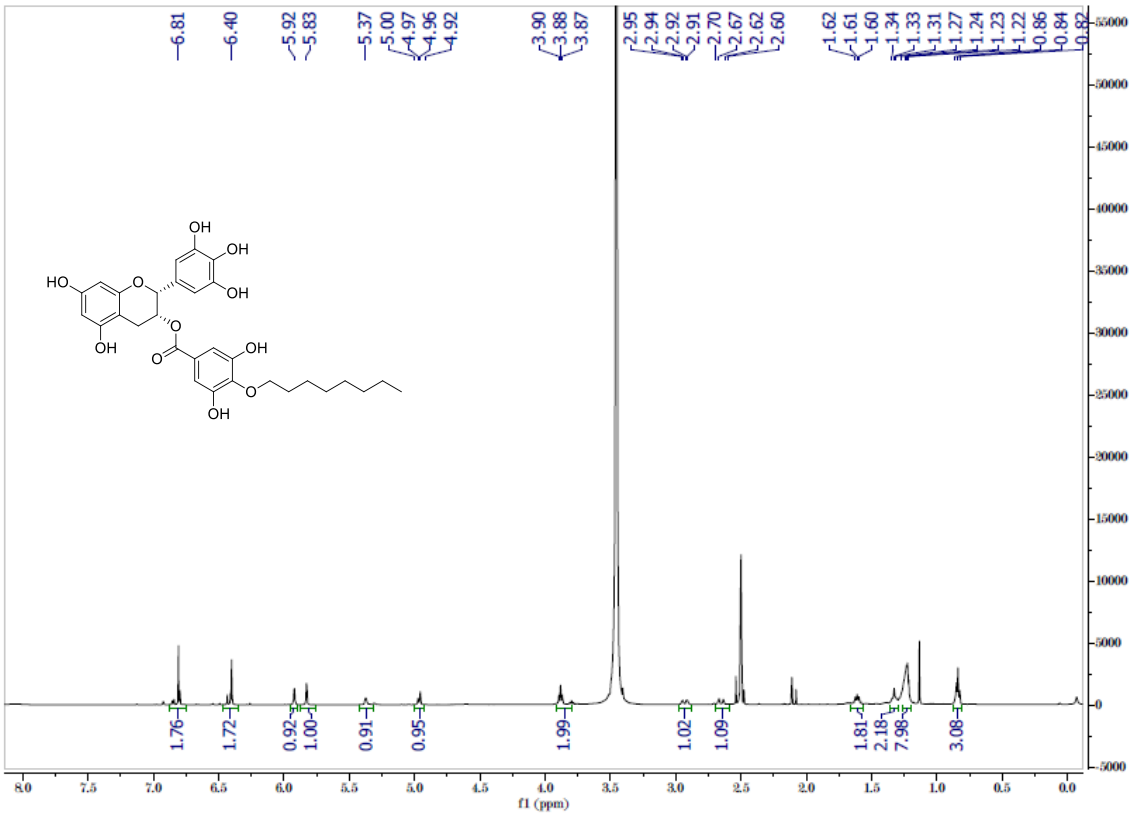


Fig. S16. ^{1}H NMR spectrum of 4"-C_8 EGCG.

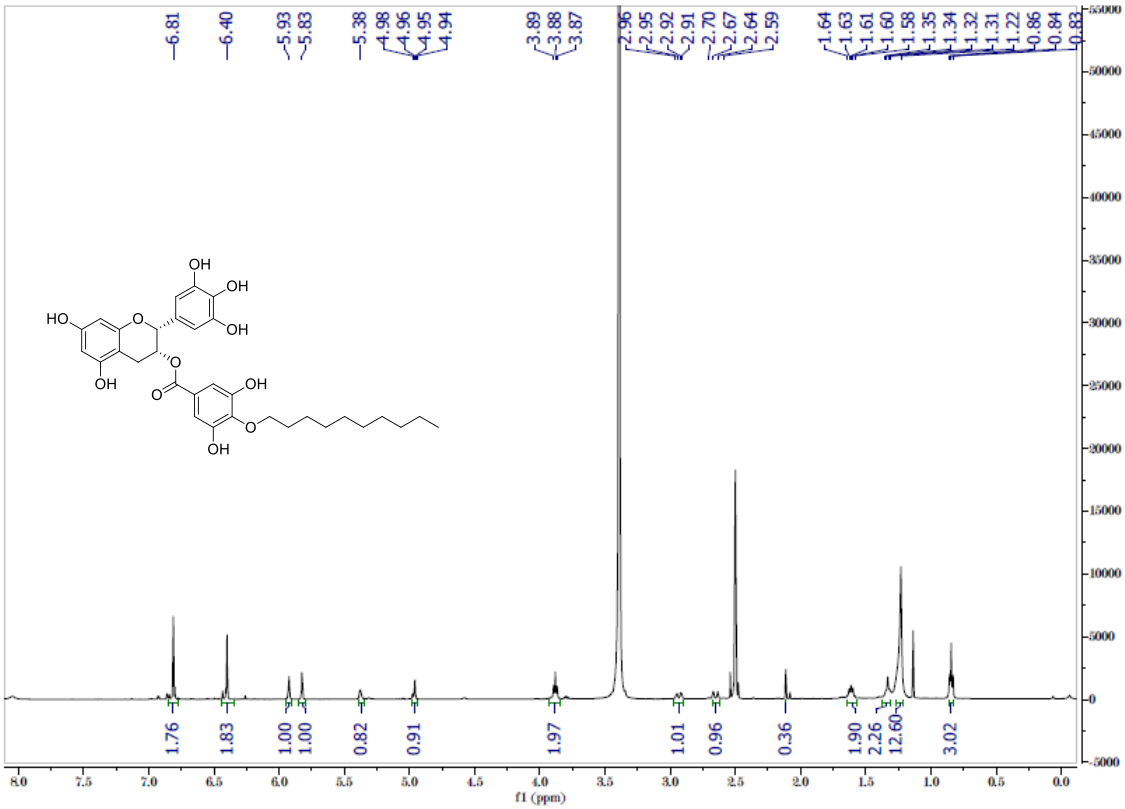


Fig. S17. ^{1}H NMR spectrum of 4"-C_{10} EGCG.

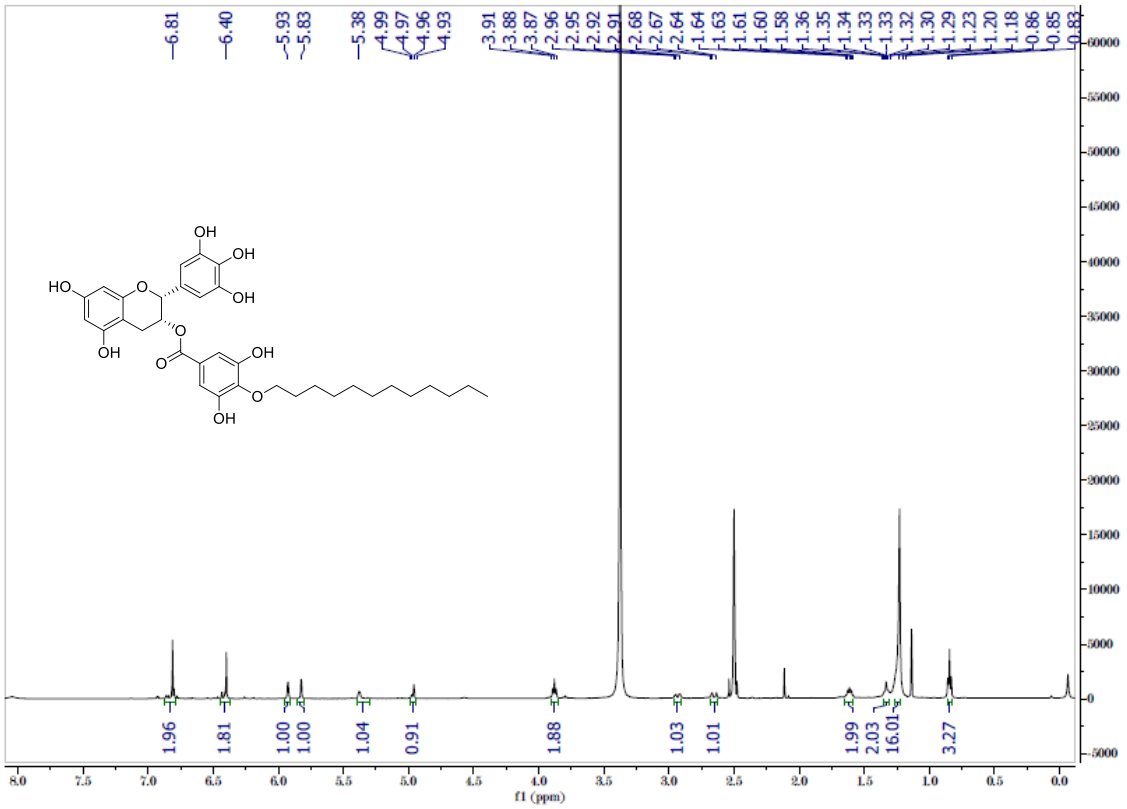


Fig. S18. ^{1}H NMR spectrum of 4"-C_{12} EGCG.

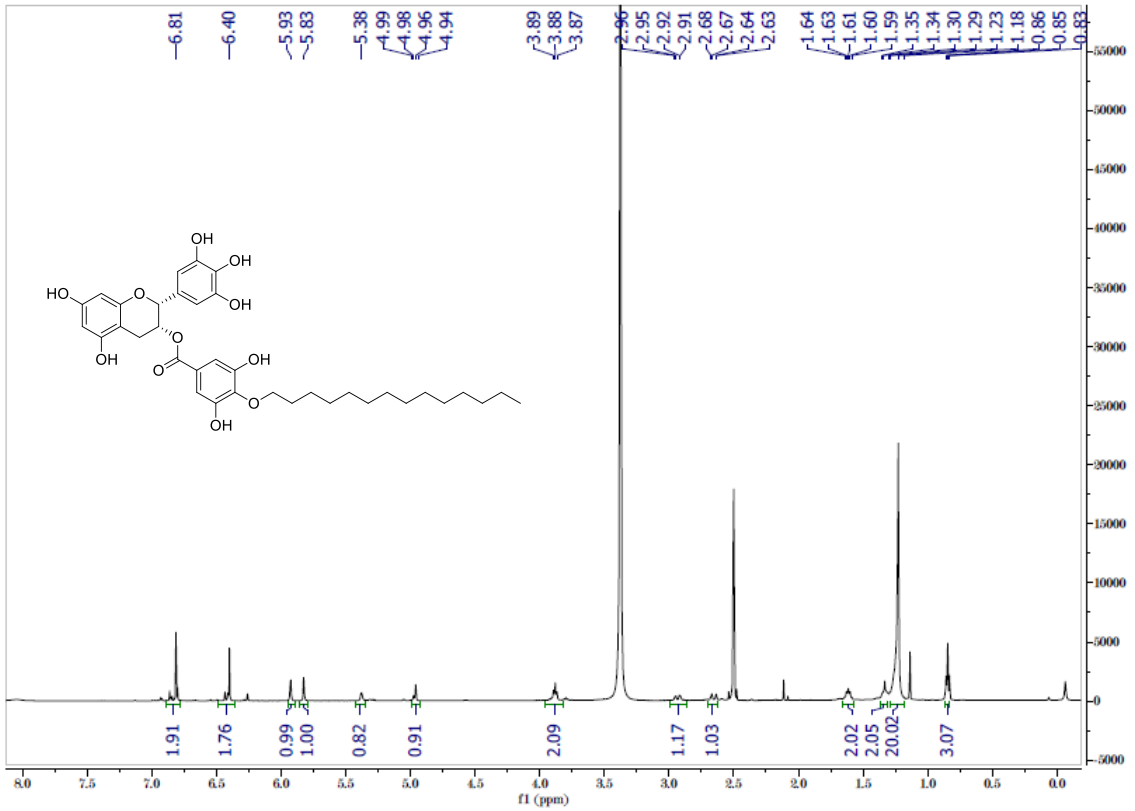


Fig. S19. ^{1}H NMR spectrum of 4"-C_{14} EGCG.

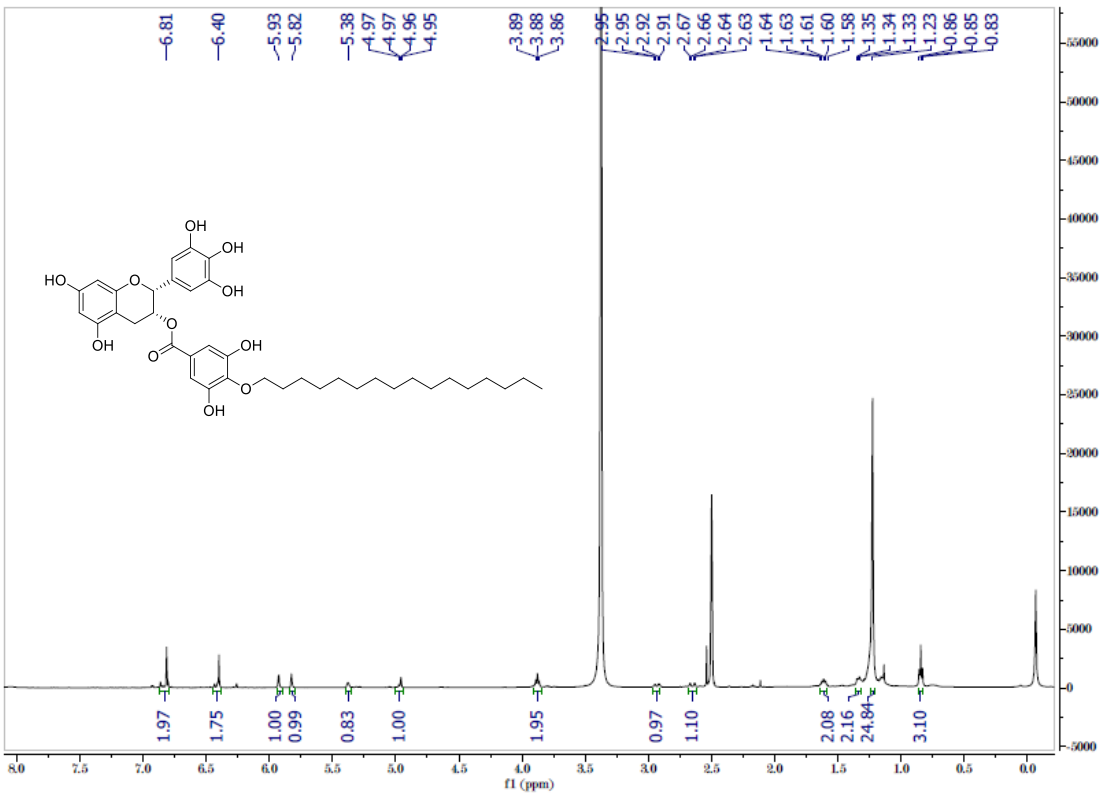


Fig. S20. ^{1}H NMR spectrum of 4"-C_{16} EGCG.

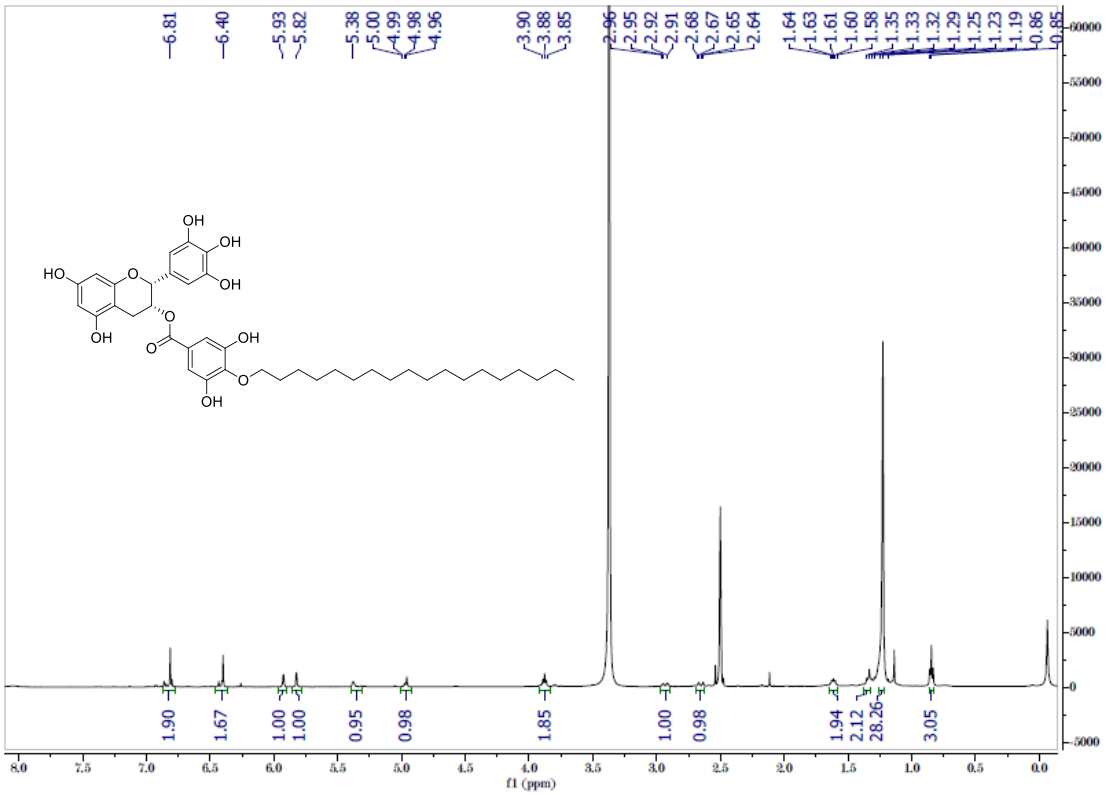


Fig. S21. ^{1}H NMR spectrum of 4"-C_{18} EGCG.

5. *^{13}C NMR spectra of 4"-alkyl EGCG derivatives*

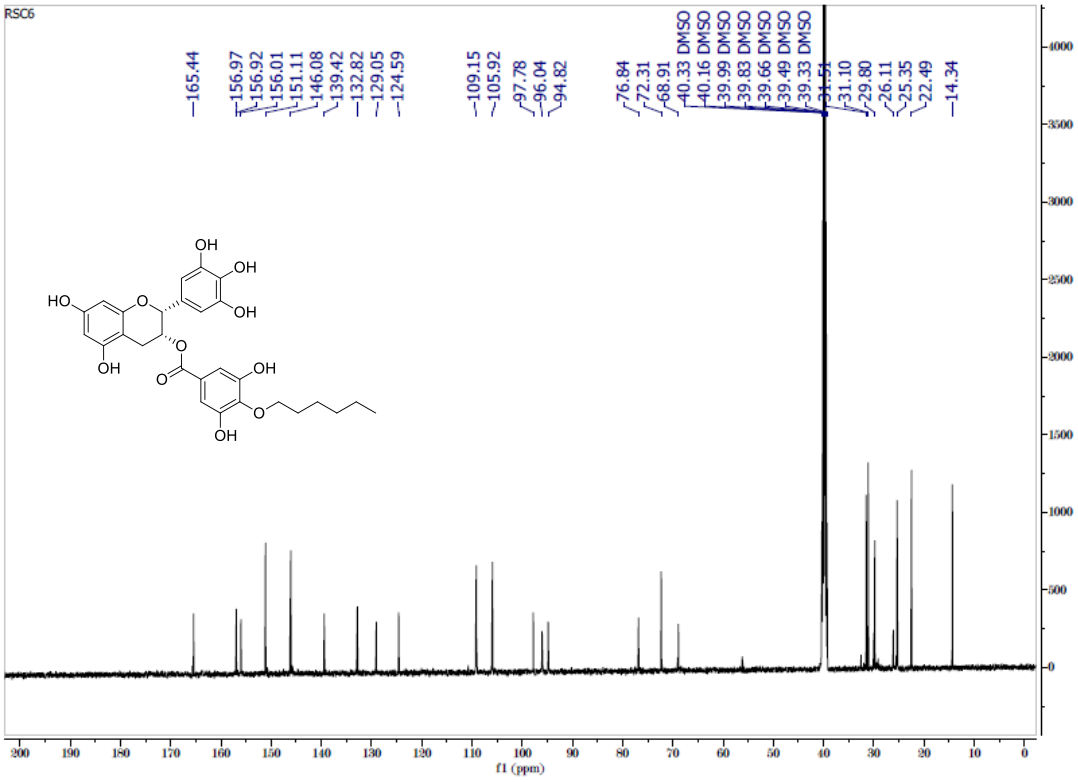


Fig. S22. ^{13}C NMR spectrum of 4"-C_6 EGCG.

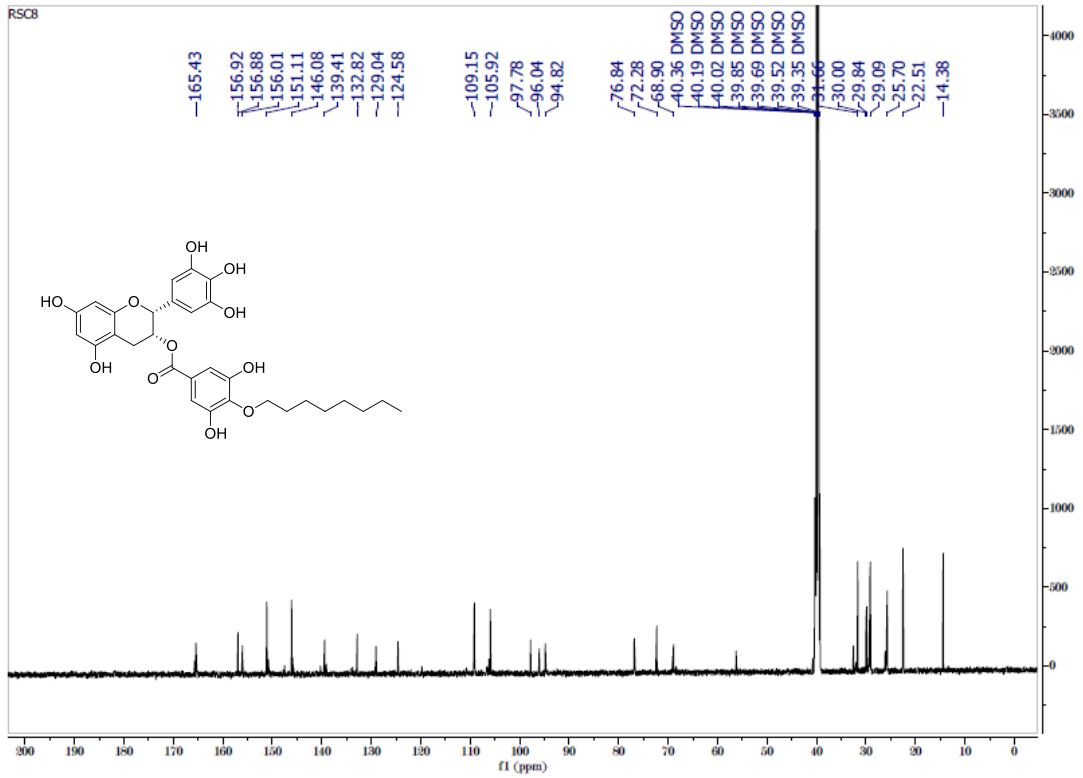


Fig. S23. ^{13}C NMR spectrum of 4"-C_8 EGCG.

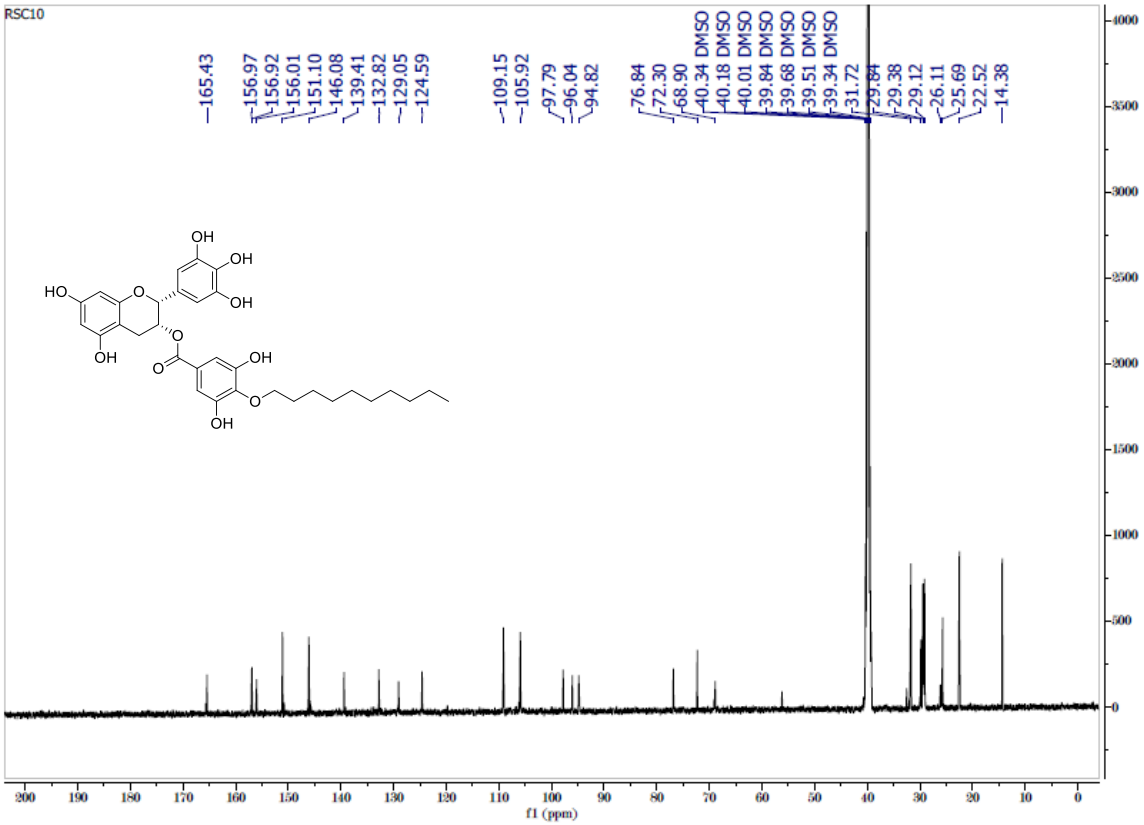


Fig. S24. ^{13}C NMR spectrum of 4"-C_{10} EGCG.

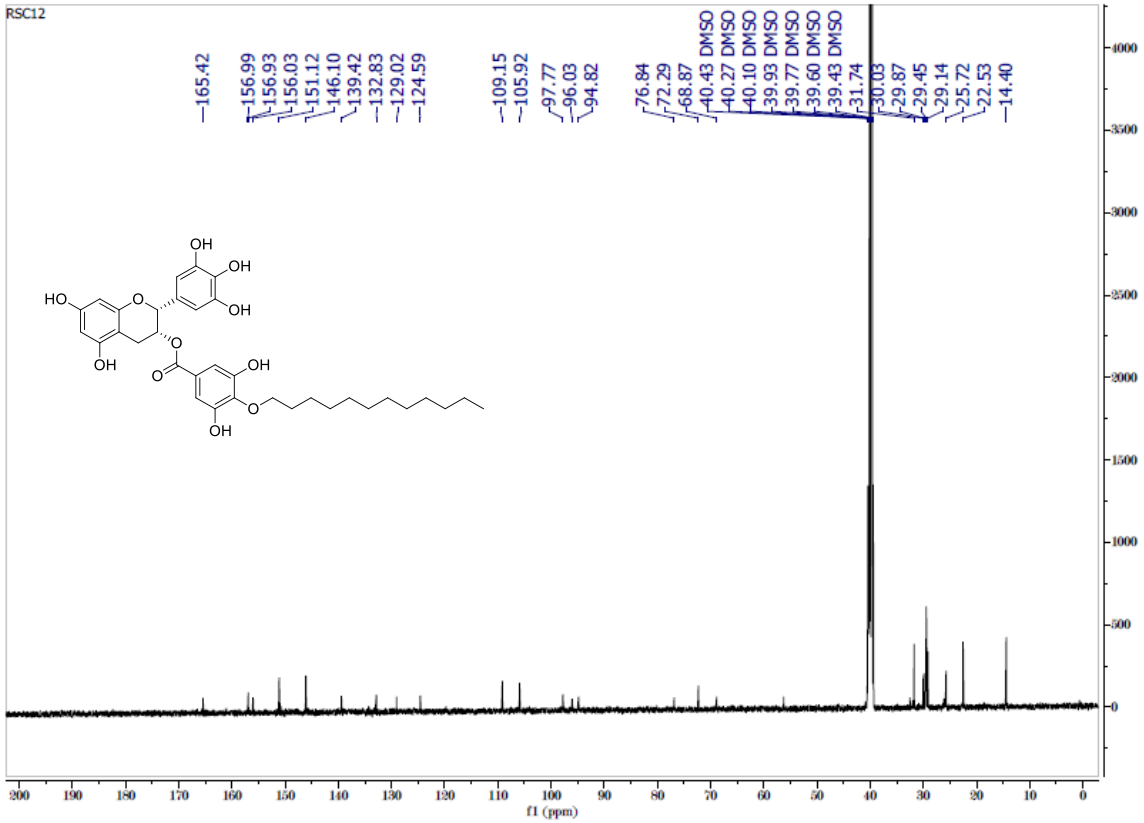


Fig. S25. ^{13}C NMR spectrum of 4"-C_{12} EGCG.

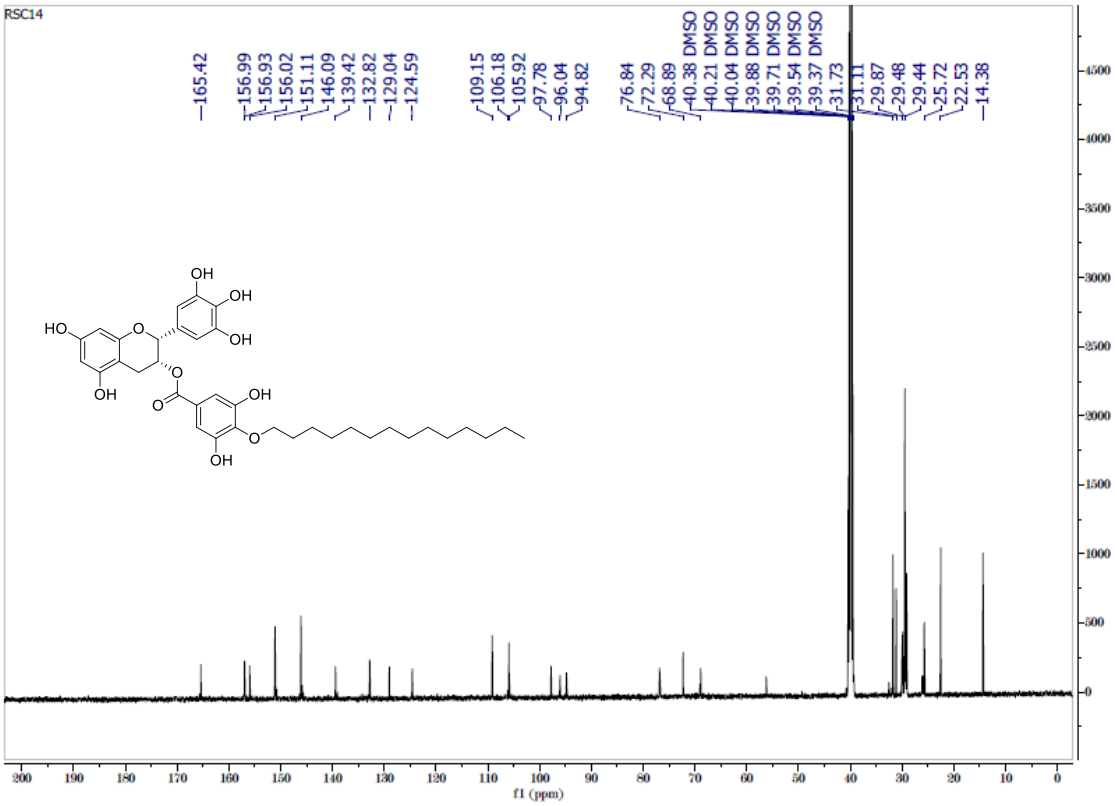


Fig. S26. ^{13}C NMR spectrum of 4"-C_{14} EGCG.

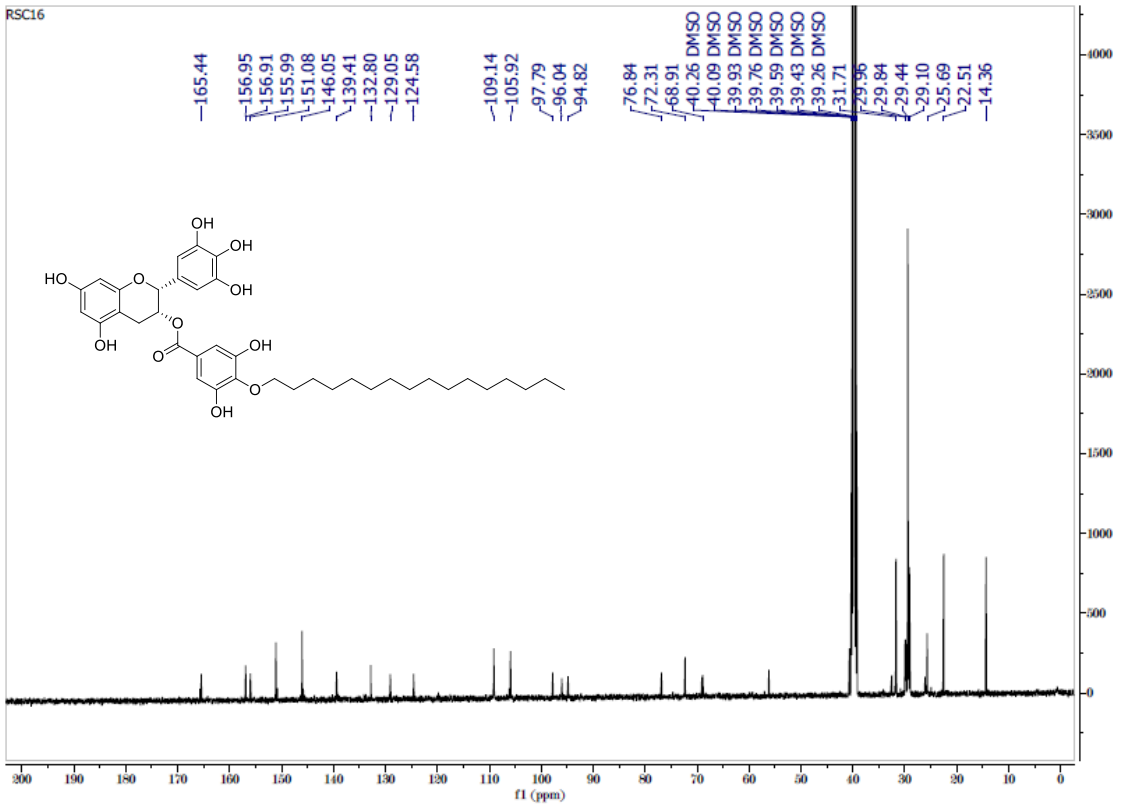


Fig. S27. ^{13}C NMR spectrum of 4"-C_{16} EGCG.

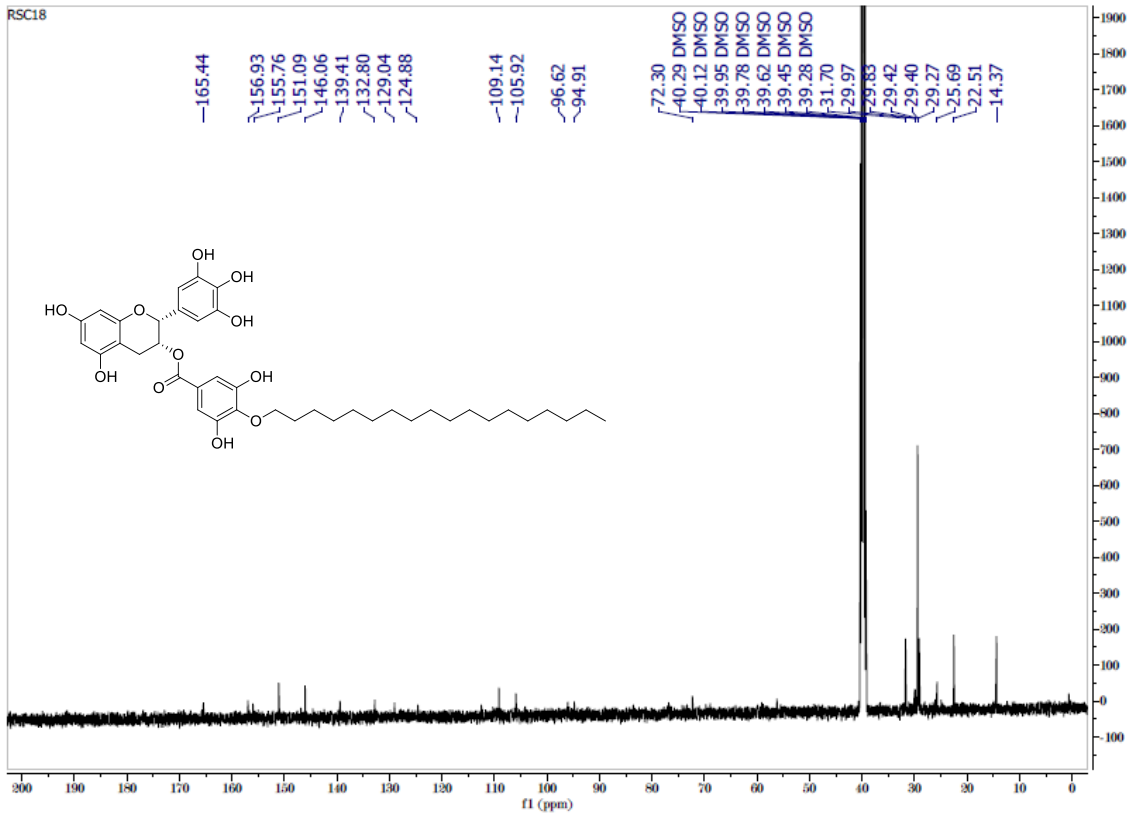


Fig. S28. ^{13}C NMR spectrum of 4"-C_{18} EGCG.

6. *FTIR spectra of 4''-alkyl EGCG derivatives*

FTIR spectra were recorded in Shimadzu Fourier transformed infra-red spectrophotometer (Model: IR Tracer 100).

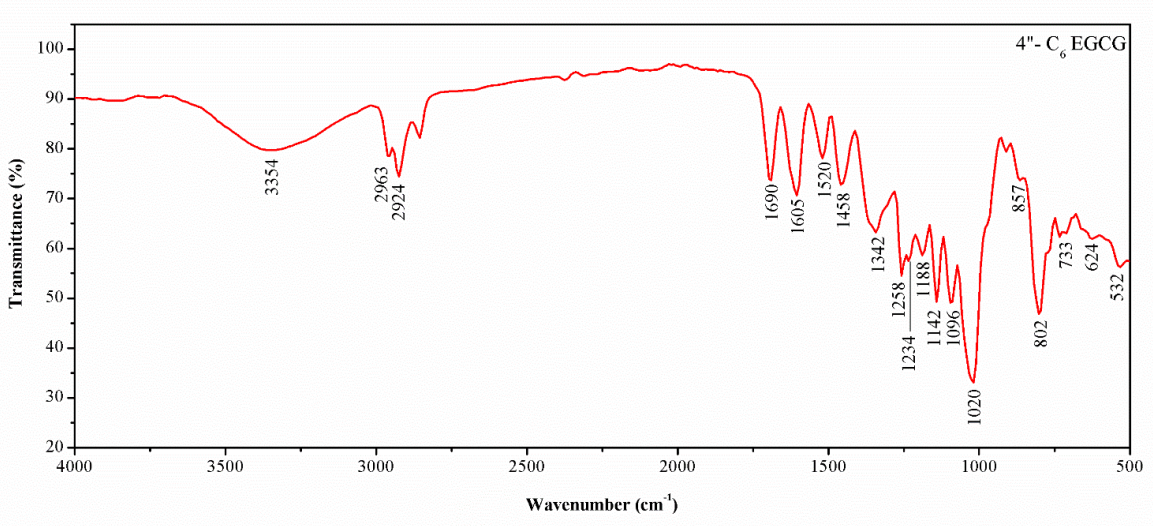


Fig. S29. FTIR spectrum of 4''-C_6 EGCG.

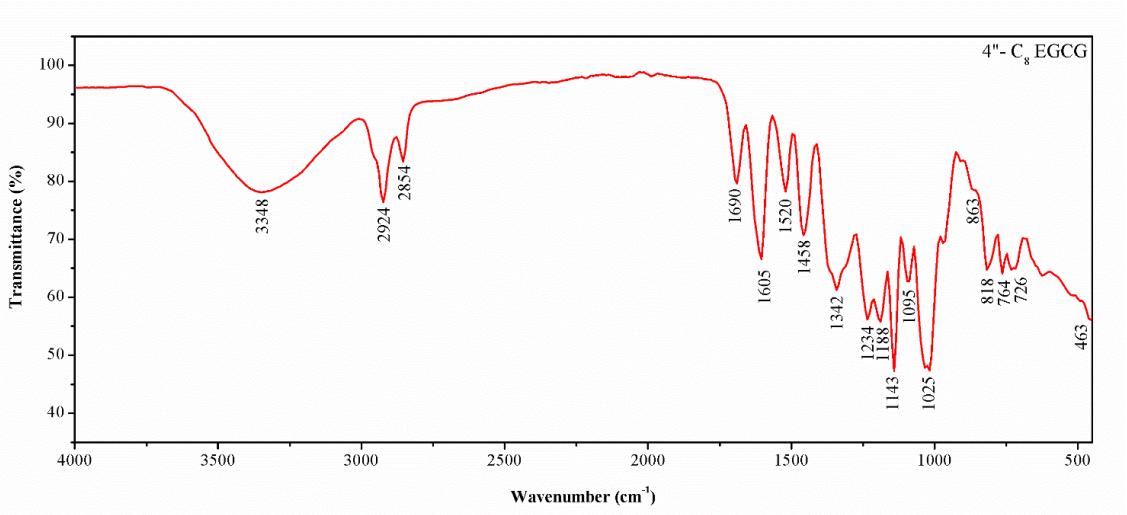


Fig. S30. FTIR spectrum of 4''-C_8 EGCG.

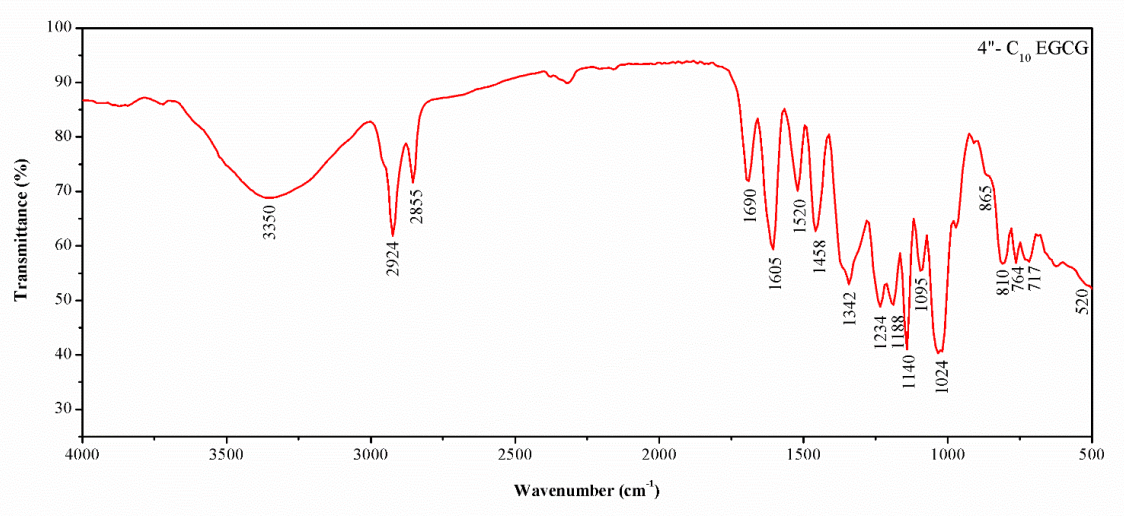


Fig. S31. FTIR spectrum of 4"-C_{10} EGCG.

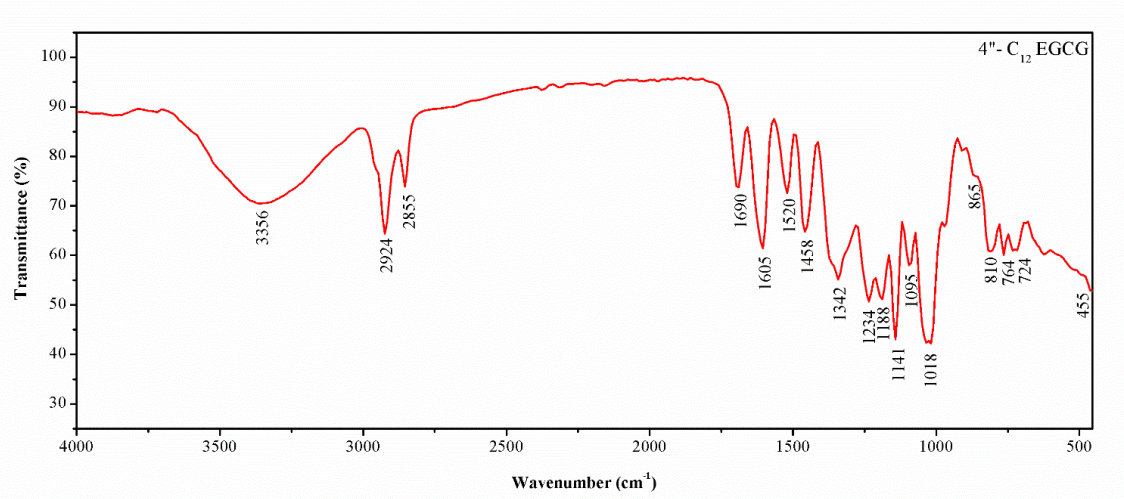


Fig. S32. FTIR spectrum of 4"-C_{12} EGCG.

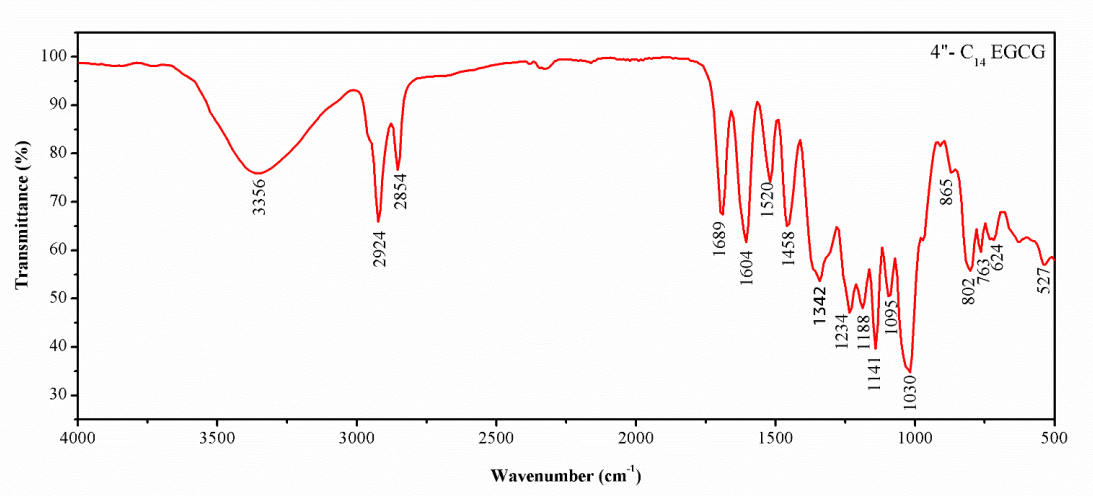


Fig. S33. FTIR spectrum of 4"-C_{14} EGCG.

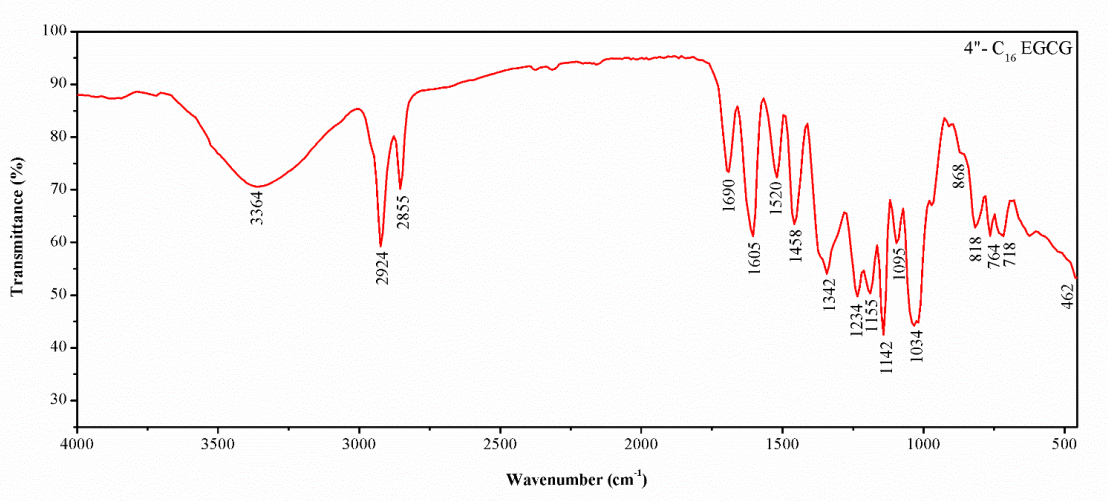


Fig. S34. FTIR spectrum of 4"-C_{16} EGCG.

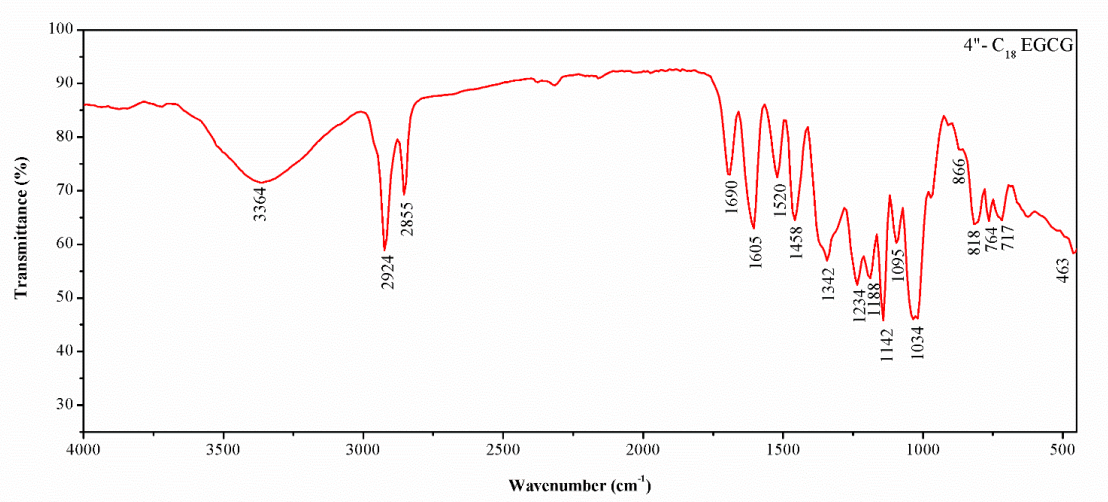


Fig. S35. FTIR spectrum of 4"-C_{18} EGCG.

7. *Stability assay*

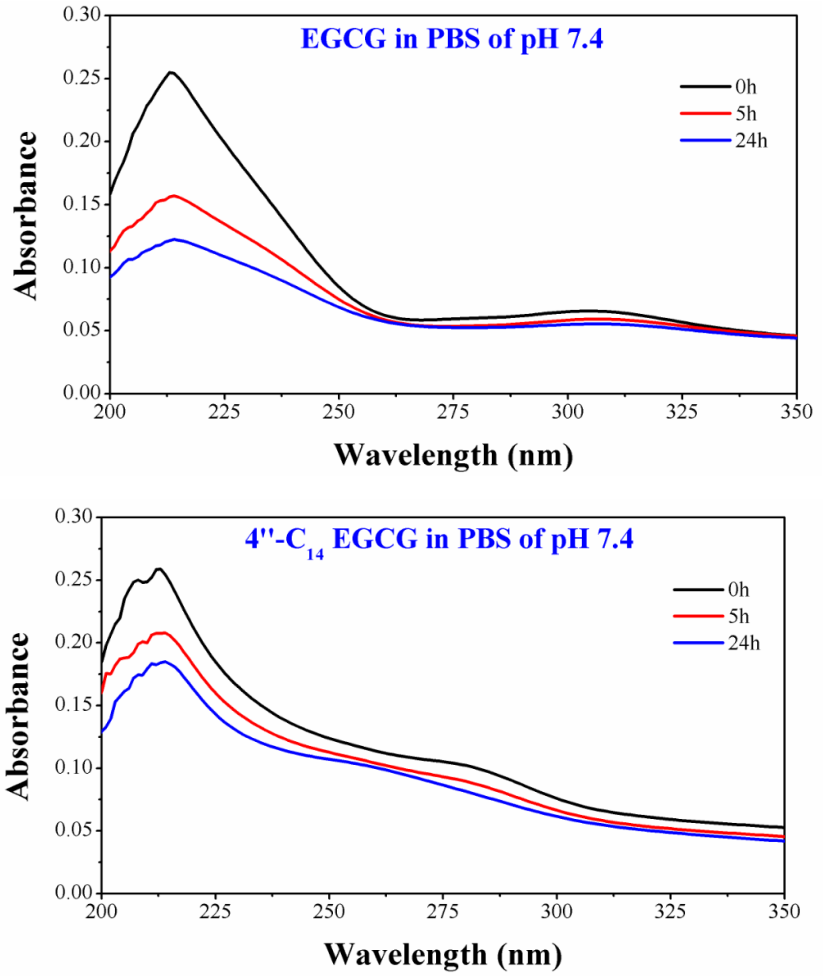
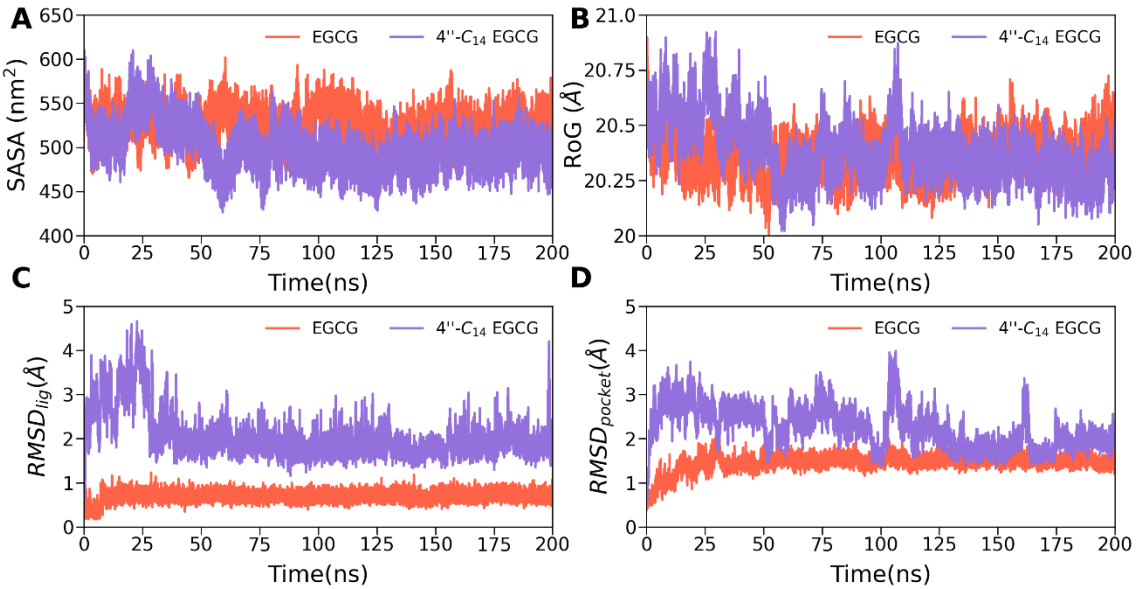


Fig S36: UV-Visible absorption spectra of EGCG and 4″-C_{14} EGCG in PBS buffer of pH 7.4.

8. *MD simulation data*



.

Fig. S37. MD simulation to analyse the interaction between EGFR and EGCG/4″-C14 EGCG (A) Time evolution of solvent accessible surface area (SASA) of EGFR kinase domain, (B) Time evolution of radius of gyration of EGFR kinase domain, (C) Time evolution of RMSD of respective ligands in EGFR complex, (D) Time evolution of RMSD of binding pocket residues (5 Å from the respective ligands) in EGFR complex.

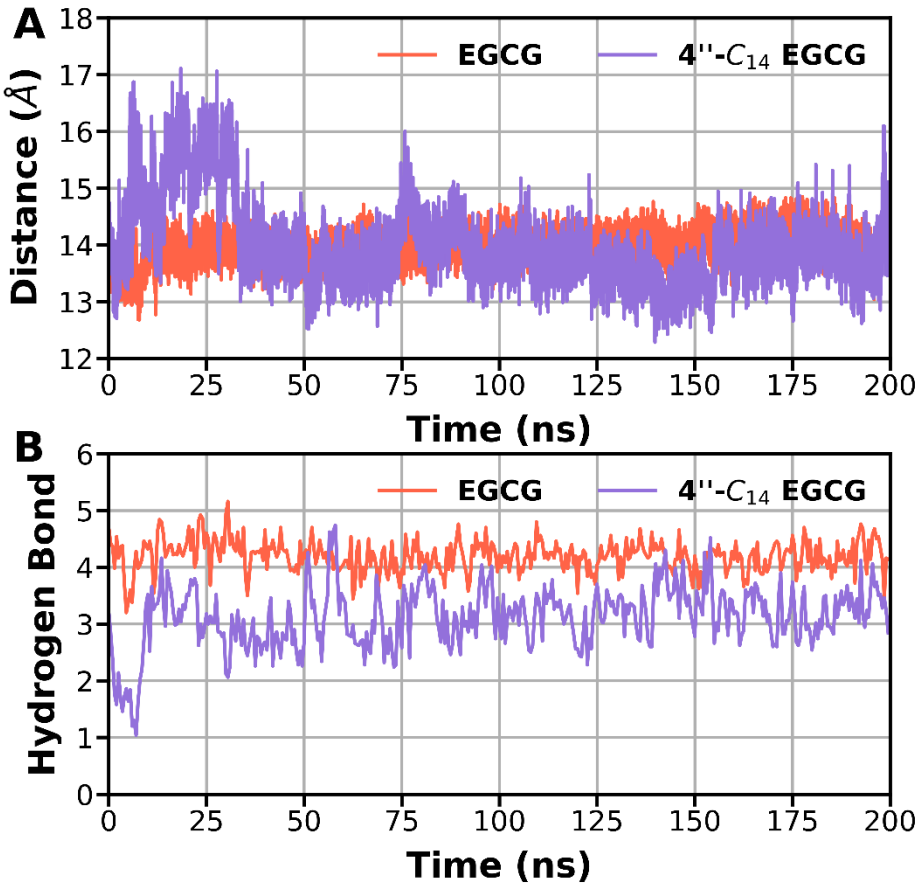


Fig. S38. The center of mass distance and the number of hydrogen bonds calculated between ligand and protein. (A) Centre of mass (COM) distance between EGFR and its respective ligands, (B) Time evolution of hydrogen bond between receptor and ligands.

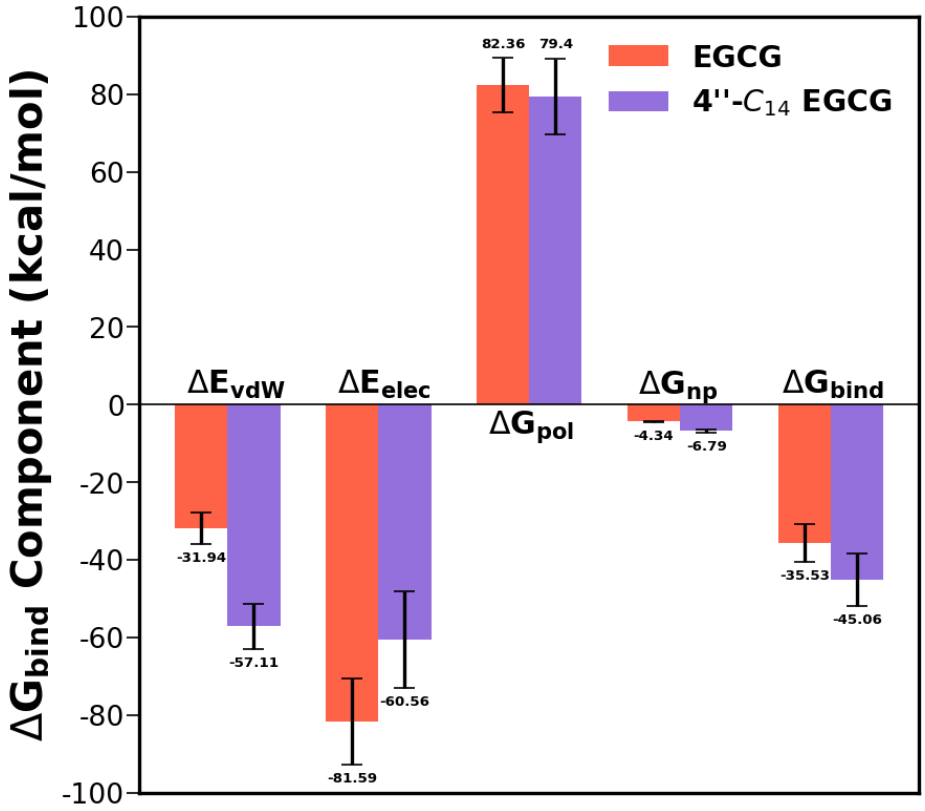


Fig. S39. Binding free energy (kcal/mol) for the receptor-ligand complexes.

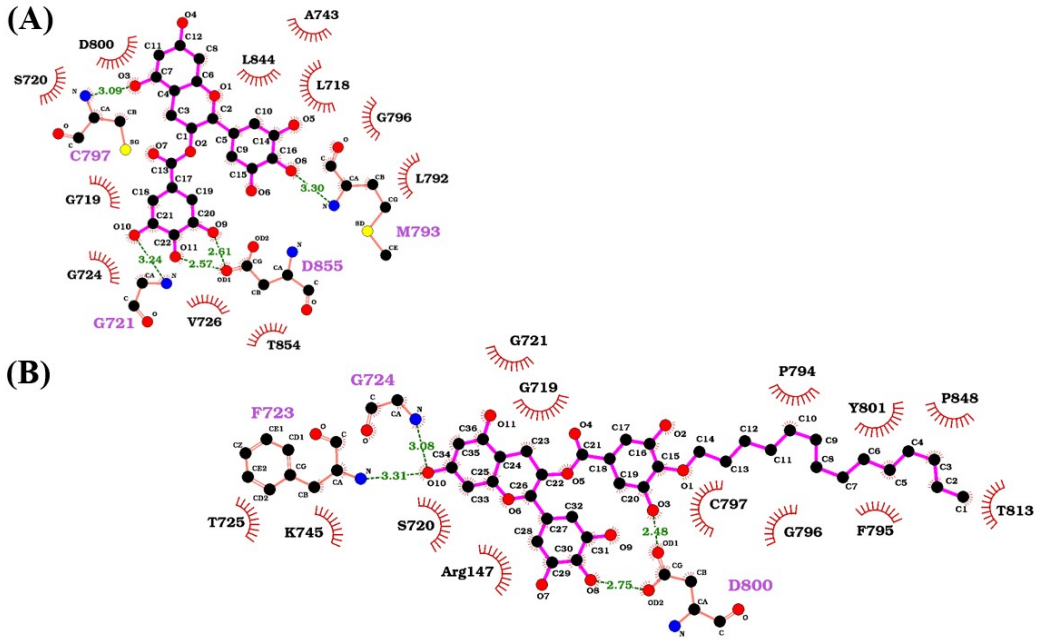


Fig. S40. Ligand-protein 2D interaction profiles for (A) EGCG, and (B) 4"-C14 EGCG. Here green dashed line is used to denote hydrogen bonds and red semi-circle is used to show the hydrophobic contacts.

Table S1. Decomposition of the binding free energy (kcal/mol) in residue level. Overall residue contribution (below -1 kcal/mol) is listed.

Residue	T_{vdw}	T_{ele}	T_{pol}	T_{np}	T_{side}	T_{back}	T_{total}
EGCG							
D855	2.30	-26.00	18.25	-0.17	-5.53	-0.09	-5.62
V726	-2.17	0.40	-0.31	-0.33	-2.09	-0.32	-2.41
L718	-2.41	0.21	0.56	-0.33	-2.05	0.07	-1.98
G796	-1.58	-1.28	1.41	-0.23	-0.59	-1.09	-1.68
C797	-1.05	-1.16	0.99	-0.17	-0.75	-0.65	-1.40
L844	-1.18	-0.11	0.21	-0.27	-1.29	-0.05	-1.34
M793	-0.00	-1.96	0.72	-0.06	-0.38	-0.91	-1.29
G719	-1.03	-0.90	0.94	-0.12	-0.31	-0.80	-1.11
4''-C_{14} EGCG							
D800	1.59	-24.30	19.42	-0.31	-3.40	-0.21	-3.61
L718	-2.57	-0.71	1.73	-0.45	-1.94	-0.05	-1.99
R841	-2.90	-2.50	3.91	-0.48	-1.89	-0.09	-1.98
C797	-2.01	-0.94	1.33	-0.21	-1.16	-0.66	-1.82
V726	-1.62	-0.13	0.22	-0.25	-1.54	-0.24	-1.78
F795	-2.29	-0.19	1.19	-0.36	-1.29	-0.35	-1.64
S720	-1.74	-0.59	1.16	-0.19	-0.36	-0.99	-1.35
G796	-1.08	-0.77	0.68	-0.17	-0.39	-0.94	-1.33

Table S2. Occupancy of receptor-ligand hydrogen bonds throughout simulation time-length.

Acceptor	Donor	Distance (Å)	Angle (°)	Occupancy (%)
EGCG				
O3	C797@N-H	2.90	156.08	36.35
M793@O	O5-H13	2.73	150.77	78.63
D855@OD2	O9-H16	2.64	165.06	64.83
D855@OD2	O11-H18	2.66	162.43	64.42
D855@OD1	O11-H18	2.67	159.43	43.40
D855@OD1	O9-H16	2.65	163.94	36.86
D800@OD1	O3-H9	2.66	162.34	30.29
D800@OD2	O3-H9	2.66	162.14	27.17
4''-C_{14} EGCG				
D800@OD1	O8-H40	2.68	163.37	49.88
D800@OD2	O8-H40	2.68	163.11	40.65
D800@OD1	O7-H39	2.65	165.24	31.19
D800@OD2	O7-H39	2.66	164.68	27.65
D800@OD1	O9-H41	2.66	164.48	18.94
D800@OD2	O9-H41	2.66	164.41	14.88
D855@OD2	O11-H46	2.69	162.41	12.35
M793@O	O2-H32	2.79	147.39	11.46
D800@OD1	O3-H35	2.64	163.79	10.44
D800@OD2	O3-H35	2.63	163.14	10.27