Near-infrared emissive cyanine probe for selective visualization of the physiological and pathophysiological modulation of albumin levels.

Bidisha Biswas,[a] Surbhi Dogra,[a] Gourab Dey, [a] N. Arul Murugan,[b,c] Prosenjit Mondal,*, [a] Subrata Ghosh.*, [a]

[a]School of Basic Sciences, Indian Institute of Technology Mandi, Mandi-175001, Himachal Pradesh India.

[b]Department of Computer Science, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, S-100 44 Stockholm, Sweden

[c]Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India - 110020

E-mail: prosenjit@iitmandi.ac.in, subrata@iitmandi.ac.in

Table of Contents

Protein sensing studies:

The amount of DMSO used in the case of protein sensing studies reported in this work did not exceed 0.2 %.[1]

Calculation of quantum yield (Φ):

For determination of quantum yield, 1,1',3,3,3',3' –Hexamethylindotricarbocyanineiodideinin ethanol was used as the standard.[2, 3] The quantum yield values were further calculated following the equation given below[4]-

$$\Phi = \frac{A_s F_u n_u^2}{A_u F_s n_s^2} \times \Phi_s$$

A_s= the absorbance value of standard.

A_u= the absorbance value of unknown sample.

F_s = the fluorescence peak of standard.

F_u = the fluorescence peak of unknown sample.

n_s = the refractive index of the solvent used for standard (corrected to ethanol).

n_u = the refractive index of the solvent used for unknown sample (corrected to be PBS).

Φ_s = the quantum yield of standard (Φ_s= 0.28).

Φ_u = the quantum yield of unknown sample.

Determination of the experimental limit of detection:

The experimental limit of detection was calculated following the reported procedure.[5]The concentration of albumin at which there is more than 10 % of emission enhancement response for **CyG-NHS**, compared to the initial emission intensity of **CyG-NHS** was considered as the detection limit in this study.

Cell line and culture conditions:

Human hepatocellular carcinoma cell lines (HepG2) and human embryonic kidney cell line (HEK293A) were used for all the *invitro* experiments. HepG2 cells were maintained in Dulbecco's minimal essential media (DMEM) supplemented with 10% FBS (fetal bovine serum) and 1% penicillin streptomycin (PenStrep) and 293A cells were cultured in DMEM with 10% FBS, 1% non-essential amino acids (NEAA) and 1% PenStrep. The cells were maintained at 37 °C in 5% CO_2.

CyG-NHS treatment:

The dye was dissolved in DMSO to make a stock of 50 mM concentration. The working concentration used for the subsequent experiments is 50 µM. For the *in vitro* experiments HepG2 cells were incubated with the dye for one hour in serum free media at 37 ºC. DMSO was used as vehicle control.

Invitro **experiments:**

In vitro **cytotoxicity assay-** The cytotoxicity of CyG-NHS dye in HepG2 cells was determined using MTT assay. To proceed with the assay, 8000 cells were seeded in 96 well plate. After 24 hours, cells were treated with CyG-NHS for 24 hours at different concentrations (12.5, 25, 50, 75, 100 µM). Cells were then treated with MTT dye (0.5 mg/ml) for 4 hours. Media from the cells was removed and the formazan crystals thus formed were then dissolved in 200 µl of DMSO. The absorbance was then recorded at 570 nM using iMark plate reader (BIORAD). The cell viability was expressed as percent cell viability relative to control.

Immunocytochemistry-Approximately, 25000 cells were plated on coverslips. On the following day cells were then treated with 50 µM dye for one hour.[6] Post treatment cells were fixed in 4% paraformaldehyde for 10mins and then blocked in 2% FBS and 0.05% Tween-20 in PBS for 1 hour at room temperature. Cells were then incubated with Anti-tubulin primary antibody (Invitrogen- 62204) (1:500 dilution in blocking buffer) overnight at 4 degrees. Further the cells were washed thrice with PBS containing 0.05% Tween-20 and finally incubated with FITC conjugated secondary antibody (Jackson Immuno Research 715-095-150) for 2 h at room temperature. The coverslips containing cells were mounted using DAPI mounting media.

Hyperinsulinemia experimental protocol- Briefly, HepG2 cells were seeded on coverslip in complete media. The cells were treated with 100 nM insulin and co-treated with insulin and phenformin hydrochloride (1 mM) for 24 hrs in serum free media. Post treatment the cells were incubated with 50 µM dye for one hour. The cells were washed thrice in PBS and the fixed with 4% paraformaldehyde. The coverslips were mounted onto slides using DAPI

mounting media. The images were visualized using Zeiss fluorescence microscope. Furthermore, for the quantitative representation of the obtained data ImageJ software was used.

Cholesterol experimental protocol-As mentioned above, the cells were seeded on coverslips and then treated with cholesterol (5 µg/ml) and co-treated with cholesterol and phenformin hydrochloride (1 mM) for 24 hrs. The cells were then incubated with the 50 µM dye for one hour, followed by washing with PBS thrice. The cells were fixed with 4% paraformaldehyde and mounted onto slides using DAPI mounting media. Images were visualized using Zeiss fluorescence microscope. Furthermore, for the quantitative representation of the obtained data ImageJ software was used.
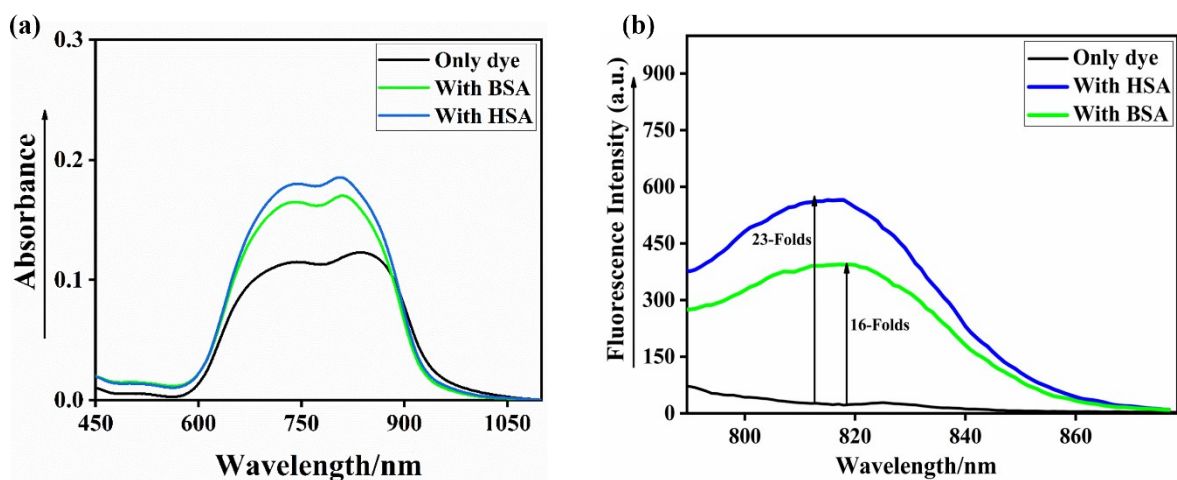

Figure S1. (a) Absorption spectra and (b) emission spectra of **CyG-NHS** in the presence and absence of BSA and HSA (15 µM) in phosphate buffer (PBS 10X, pH= 7.4).
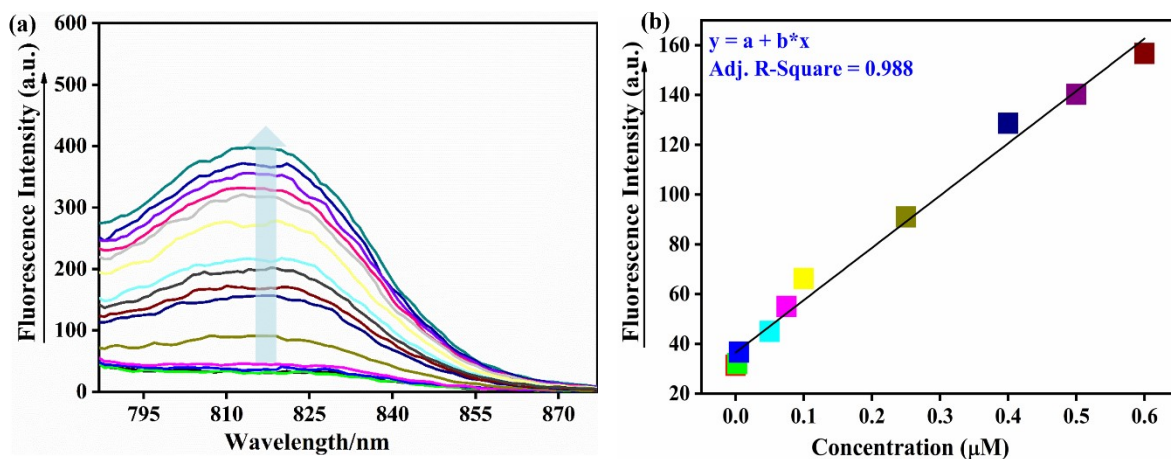
Figure S2. (a) Emission profile of **CyG-NHS** (10 µM) with gradually increasing amounts of bovine serum albumin (BSA) from (0-15 µM) in phosphate buffer (PBS 10X, pH= 7.4); (b) Linear correlation of the emission response of **CyG-NHS** with incremental addition of BSA in phosphate buffer (PBS 10X, pH= 7.4).
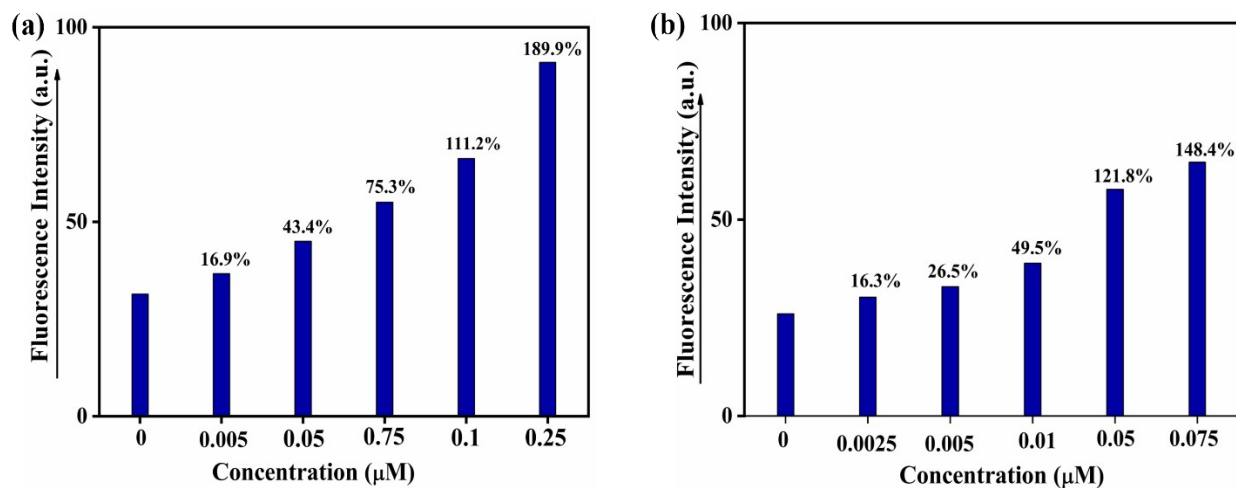

Figure S3. (a) Emission profile of **CyG-NHS** (10 µM) showing the percentage increment of the emission response in the presence of BSA and HSA in phosphate buffer (PBS 10X, pH= 7.4).
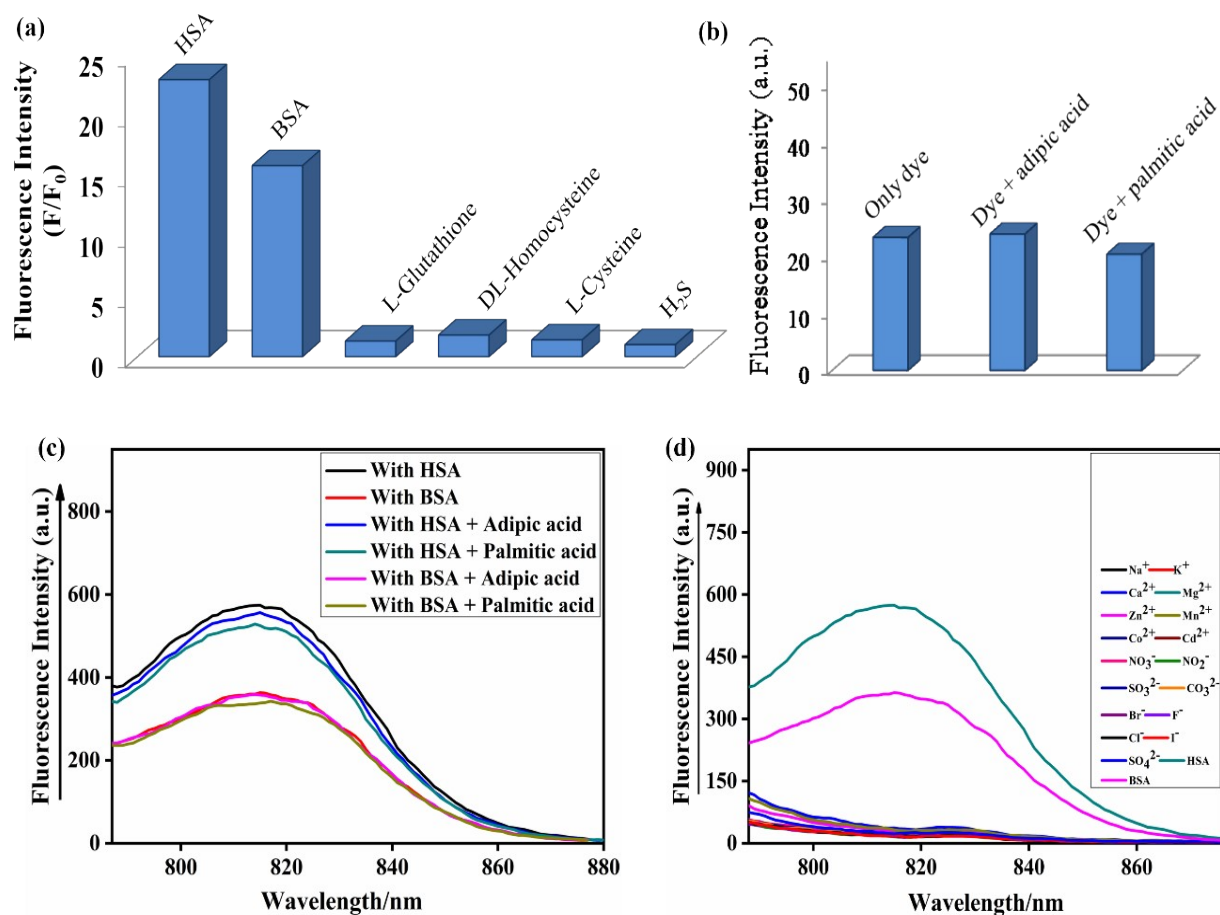
Figure S4. (a) Fluorescence emission spectra of **CyG-NHS** (10 µM) in the presence of biothiols (1 mM); (b) Fluorescence emission spectra of **CyG-NHS** (10 µM) and (c) the labeling efficiency of the probe towards BSA (15 µM) and HSA (15 µM) in the presence of 1 mM lipid (Palmitic acid and Adipic acid); and (d) the fluorescence emission spectra of **CyG-NHS** (10 µM) in the presence of various cations and anions (1 mM) in phosphate buffer (PBS 10X, pH= 7.4).
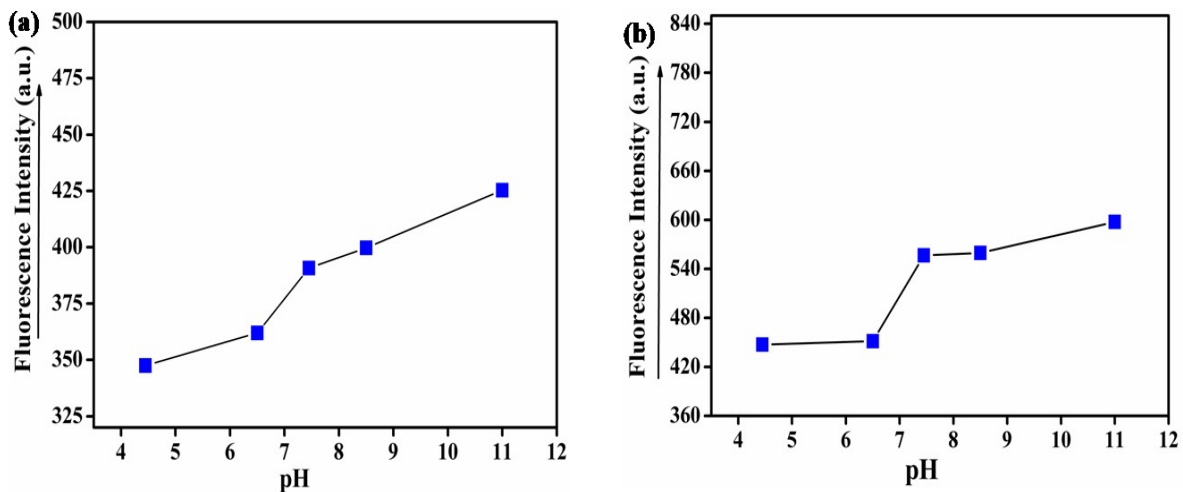
Figure S5. (a) Fluorescence profile of **CyG-NHS** (10 µM) with (a) BSA and (b) HSA under various pH (4.45-11) in phosphate buffer (PBS 10X, pH= 7.4).
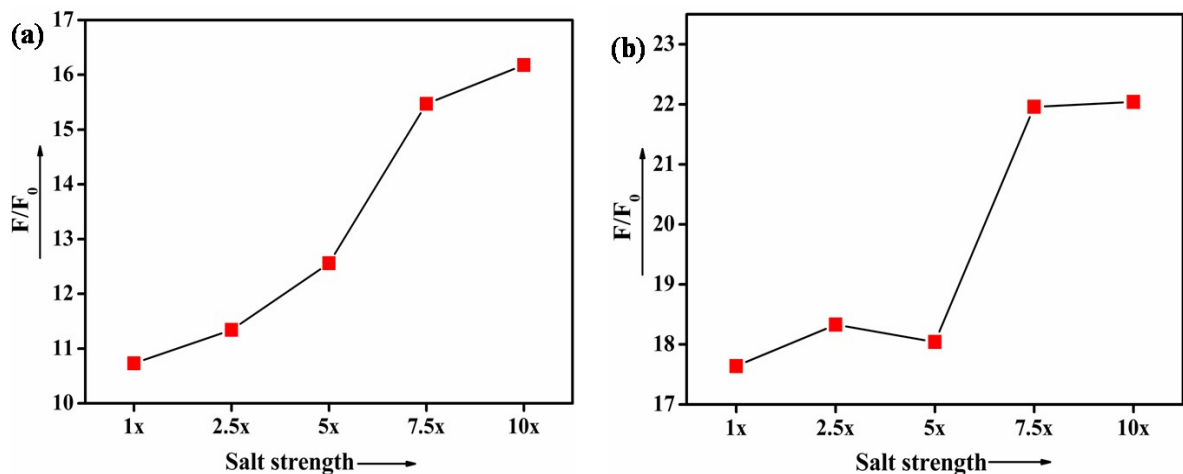

Figure S6. Fluorescence profile of **CyG-NHS** (10 µM) with (a) BSA and (b) HSA in phosphate buffer (pH= 7.4) under various salt concentrations (PBS 1X-10X).
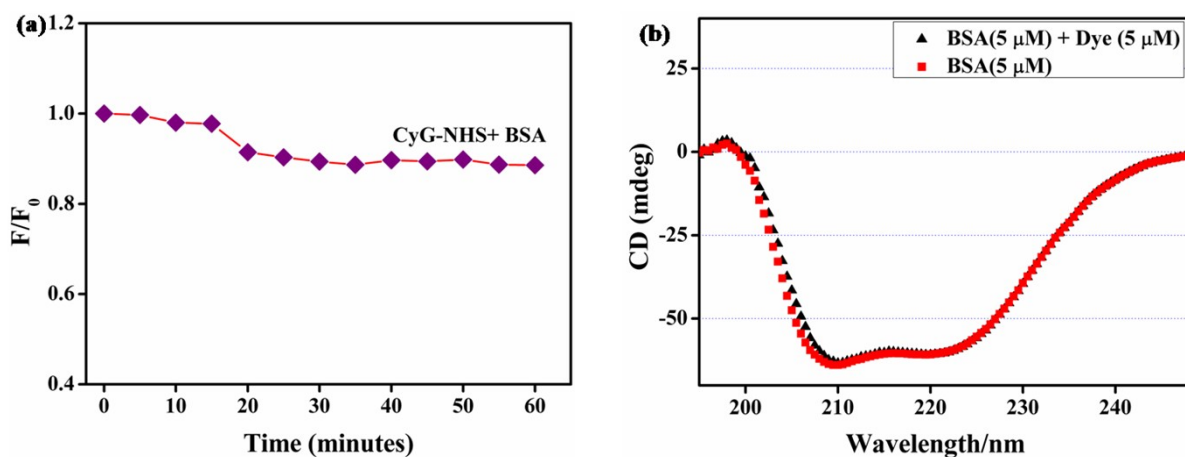
Figure S7. (a)Time dependent emission behavior of **CyG-NHS** (10 µM) in presence of BSA (15 µM) in phosphate buffer (PBS 10X, pH= 7.4) under continuous light exposure using 160 W mercury lamp; (b) Circular Dichroism (CD) spectra of HSA (5 µM) and BSA (5 µM) in the presence and absence of **CyG-NHS** (5 µM) in phosphate buffer (PBS 10X, pH= 7.4).
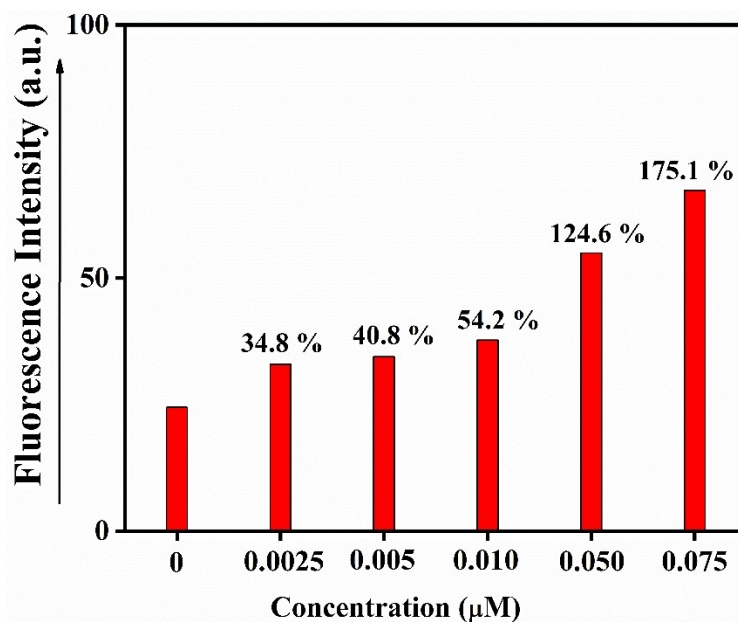

Figure S8. Emission profile of **CyG-NHS** (10 µM) showing percentage increment of the emission response in blood serum samples.

Computational methods:

We have studied the binding mode, conformational dynamics and binding free energy of CyG-NHS within HSA using molecular docking, molecular dynamics and molecular mechanics-Generalized Born surface area (MM-GBSA) approach. The molecular docking has been carried out using 3D structure reported for HSA in protein databank (the reference ID is 1AO6)[7]. The structure has been reported with a resolution of 2.5 Å. Firstly, the molecular structure for CyG-NHS has been built using Molden software.[8] The molecular geometry of the ligand has been optimized by using Gaussian 09[9] software by employing B3LYP/6-31G* level of theory. The final Gaussian output file has been used to generate .mol2 using Ambertools[10] which is then subsequently converted to .pdbqt file using Mgltools.[11]The pdbqt file of the ligand and PDB file for the target HSA protein were used as the inputs for carrying out molecular docking studies by employing Autodock4.0 software.[11] Since the binding sites for this specific ligand is not known, we carried out a blind docking by specifying grid box size which can enclose the whole protein. However, the grid box spacing was chosen as the default 0.375 Å. The numbers of grid points along x, y and z directions for the HSA were 270 x 195 x 230. The molecular docking procedure has been carried out by choosing Lamarckian genetic algorithm for finding the global minimum energy structure for the protein-ligand complex. The algorithm generates tens of thousands of conformations and orientations for the ligand within the defined grid box and computes the docking energies for them and the minimum energy structures are identified by comparing the scores for all the conformations. Use of genetic algorithm allows the program to find the global energy minimum in the potential energy surface of the protein-ligand complex. As many as 500 low energy binding modes/poses were identified for the ligand within the HSA target which was analyzed.

The least energy binding mode has been used for generating the input structure for the subsequent molecular dynamics simulation. The hydrogens were added to the ligand in its most stable binding mode (using Chimera[12] Add H option) as obtained from docking calculation. Followed by this, the charges for all atoms in the ligand were computed by employing B3LYP/6-31G* level of theory by keeping the orientation and conformation unaltered. The molecular dynamics simulations were carried out for the high affinity binding

mode of the ligand within HSA target by using Amber18 software.[13] There were multiple binding sites found for the ligand. However, the ligand in the high affinity binding site is a reasonable representation for the protein-ligand complex, the calculations were carried out for this alone. With the increasing concentrations of the solute, the other binding sites will become relevant. We have used GAFF,[14] FF99SB[15] force-fields for describing the ligand and protein respectively while TIP3P forcefield was used for water solvent. The force-field is made of inbuilt set of parameters describing the potential energy functions describing the intra and intermolecular interactions of the small molecular and biomolecular systems. The albumin ligand complex was solvated in the water solvent and sufficient number of counter ions was added to neutralize the systems. Firstly, minimization run was carried out to remove any strains and hot spots within the complex. Followed by this constant volume simulation and simulations in isothermal-isobaric ensemble were carried out. The time step for the integration of equation of motion was 2 fs. Followed by the equilibration run for a time scale 5 ns, the production runs were carried out for a time scale of 100 ns. As a reference system, the ligand in water also has been simulated in isothermal-isobaric ensemble for a total time scale of 100 ns. The conformational dynamics of ligand in water and in HSA protein were analyzed. The dihedral angles defining the relative orientations of different fragments in the molecules were computed as a function of time and the distributions of the angles were computed using CPPTRAJ module of Amber tools. Further, the binding free energy was estimated for the ligand in HSA protein using MM-GBSA approach. The binding free energy calculations were carried out for 2500 configurations corresponding to the last 5 ns time scale from the production runs. The binding free energies in this approach are defined as the sum of van der Waals, electrostatic, polar, and non-polar solvation energies. The free energies are computed for the protein, ligand and complex, and the binding free energy is computed as the difference between the complex and individual subsystems. The coordinates for the protein, ligand and complex for the free energy calculations were extracted from the single trajectory which reduces the computational cost associated with running three independent simulations for these systems. The MMPBSA.py[16] module which is part of the Amber tools has been used for this set of binding free energy calculations.

Table S1: The binding free energy of the ligand with HSA protein and different contributions.

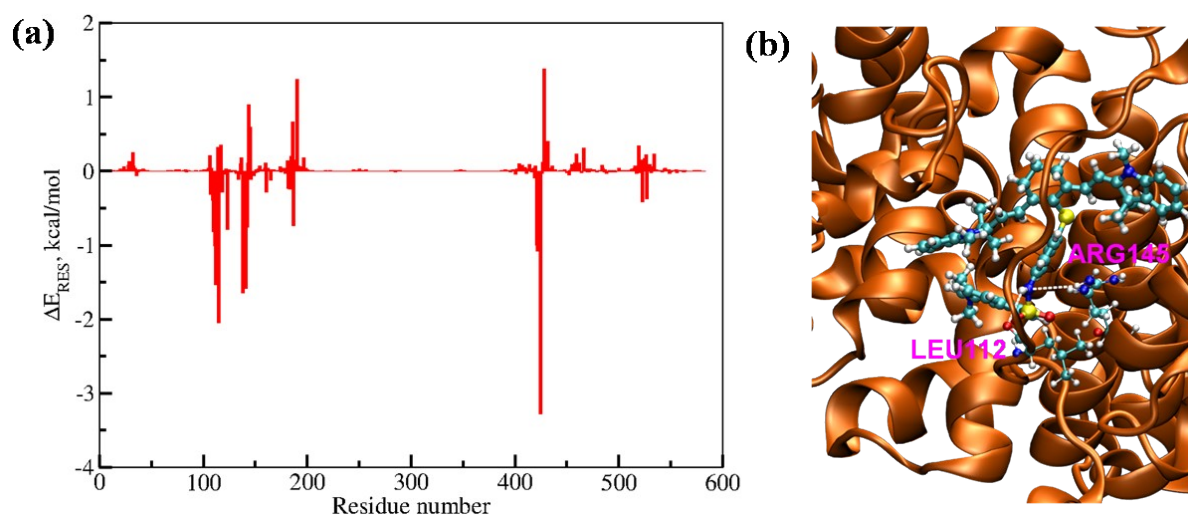| Target	E_{vdw}	E_{elec}	$E_{pol\ solvation}$	$E_{np\ solvation}$	ΔG, kcal/mol
HSA	-66.9	-84.8	127.9	-8.0	**-31.8**

Figure S9. (a) Decomposition analysis and the residue wise contributions to the total binding free energies in kcal/mol; (b) Hydrogen bonds between the LEU112 (carbonyl carbon) and ARG145 (Guanidinium NH_2) with -NH group of the probe are shown.
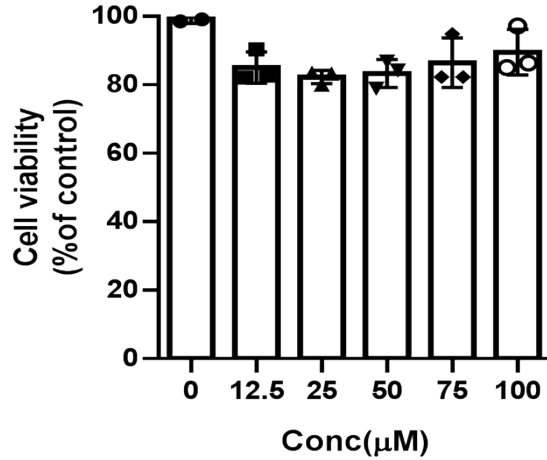
Figure S10. Cell viabilities of HepG2 cells treated with different concentrations (0–100 µM) of **CyG-NHS** for 24 h.
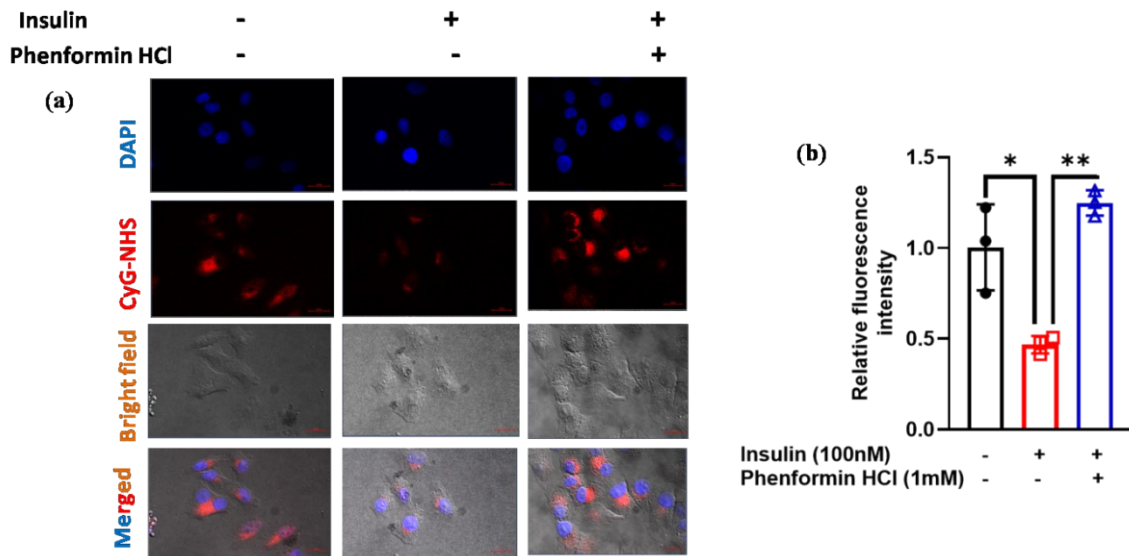

Figure S11. (a) Fluorescence images of HepG2 cells treated with **CyG-NHS**. The cells were co-treated with insulin (100 nM) and Phenformin HCl (1 mM) for 24 h. Post treatment the cells were incubated with **CyG-NHS** (50µM) for 60 min. Scale bar-20 µm; (b) Quantified graphical plot for the fluorescence intensity obtained using ImageJ software. *P < 0.05, **P < 0.01compared with the untreated control; values expressed as mean ± SEM (ANOVA followed by Bonferroni's Multiple Comparison)
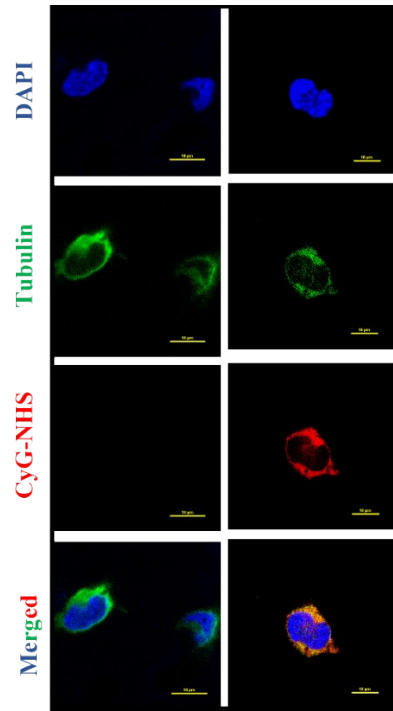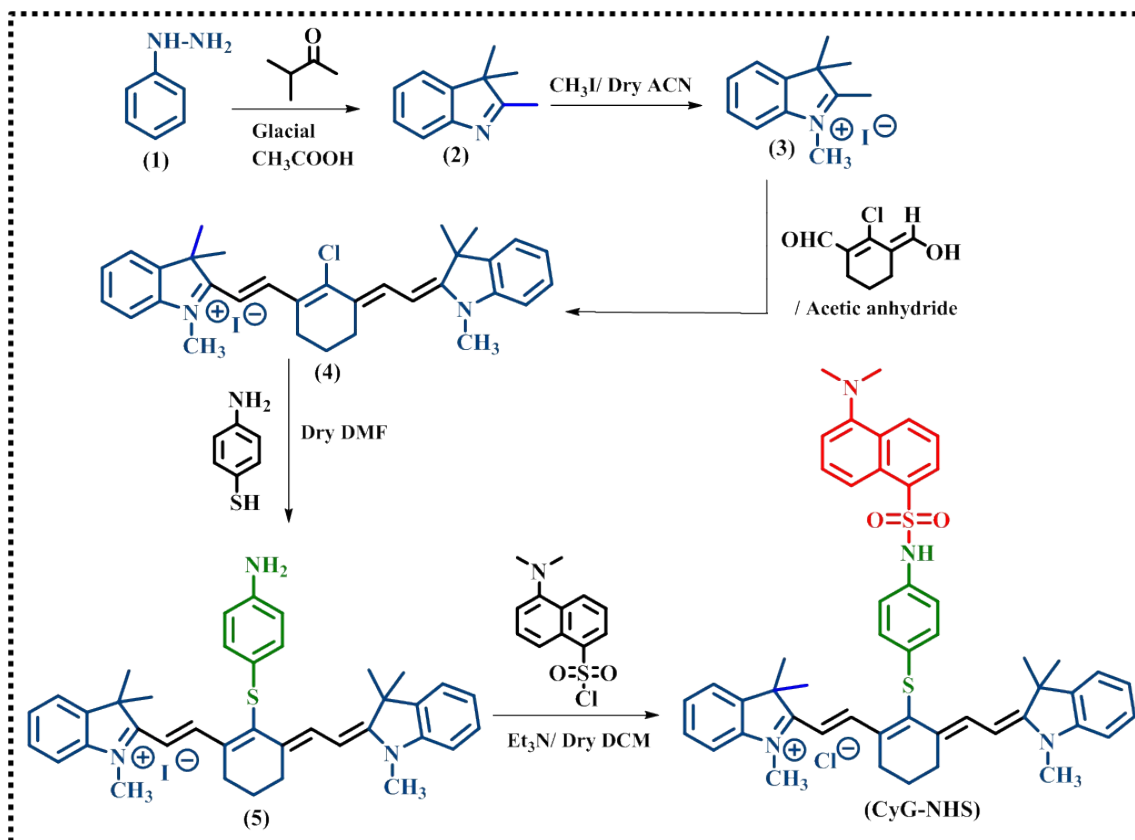
Figure S12. The confocal images depict the colocalization of **CyG-NHS** with the tubulin in the cytosol in HepG2 cells. Cells were treated with **CyG-NHS** for one hour and then immunostained for tubulin. The scale bar is 10 μM.
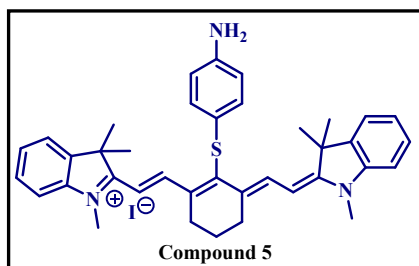
Scheme S1. Synthetic route for development of probe **CyG-NHS**.

Synthesis of intermediates 2-4:

Intermediates 2-4 have been synthesized following reported literature methods.[17-19]

Synthesis of intermediate compound 5[20]

Compound 5

Compound 4 (1 equiv) was taken in a round bottomed flask under nitrogen atmosphere followed by addition of dry DMF. To this solution 4-aminothiophenol (5.5 equiv) previously dissolved in dry DMF was added using a syringe. The resultant mixture was allowed to stir at room temperature for 1 h and the progress of the reaction was monitored by TLC. Once the reaction was complete, the reaction mixture was poured over ice water. As soon as the ice melted, the aqueous layer was washed with EtOAc and was extracted using a separating funnel. The organic layer was then dried over sodium sulfate before concentrating in rota vapor keeping temperature below 50 °C. Finally, the compound was purified through column chromatography using a mixture of DCM-methanol as eluent. The deep blue solid obtained as product was further subjected to ^{1}H- NMR, ^{13}C-NMR, and mass characterization.

**Spectroscopic details of compound 5:**

^{1}H NMR (500 MHz, CDCl$_3$): δ 8.74-8.71 (d, J=14.45 Hz, 2H), 7.35-7.32 (t, J=7.55 Hz, 2H), 7.28-7.27 (d, J= 7.55 Hz, 2H), 7.19-7.16 (m, 2H), 7.12-7.10 (d, J=7.55 Hz, 2H), 7.00-6.98 (d, J=8.25 Hz, 2H), 6.60- 6.58 (d, J=8.9 Hz, 2H), 6.16- 6.13 (d, J=13.75 Hz, 2H), 3.65 (s, 6H), 2.72- 2.70 (m, 4H), 1.98-1.94 (m, 2H), 1.53 (s, 12H).

^{13}C NMR (125 MHz, CDCl$_3$): δ172.57, 154.57, 146.42, 145.24, 142.73, 140.90, 134.32, 128.61, 128.17, 125.00, 124.67, 122.02, 116.21, 110.47, 101.54, 49.02, 32.27, 27.90, 26.66, 20.78

HRMS (ESI): Calc. for C$_{38}$H$_{42}$N$_3$S [M-Cl]$^+$: 572.3094; Found: 572.2994

Physical Properties: Deep blue colored solid. (50 mg, 42.7 % yield).

Synthesis of compound CyG-NHS:

CyG-NHS
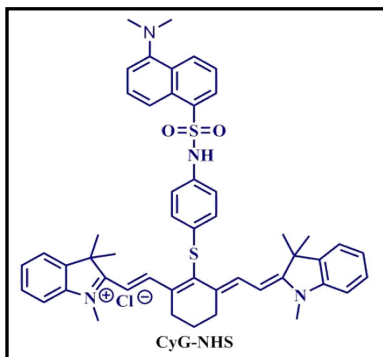
Compound 5(1equiv) dissolved in dry DCM was taken in a dried sealed tube with N_2 atmosphere. To this solution, triethylamine (20 µL) was added and stirred for 10 minutes. Dansyl chloride (3 equiv) was then added to the resultant mixture and allowed to stir overnight at 35-37 ºC. Once the reaction was complete, the reaction mixture was extracted with water and dichloromethane followed by drying of the organic layer over sodium sulfate. The organic layer was then concentrated using rotavapor and was subjected to purification using column chromatography using a mixture of DCM-methanol as eluent. The deep green solid obtained as product was further subjected to ^{1}H- NMR, ^{13}C-NMR, and mass characterization.

Spectroscopic details of CyG-NHS:

^{1}H NMR (500 MHz, CDCl$_3$): δ 8.53-8.49 (m, 3H), 8.23-8.22 (d, J= 8.9 Hz, 1H), 8.14- 8.12 (d, J= 7.55 Hz, 1H), 7.53-7.50 (m, 1H), 7.34-7.31 (m, 2H), 7.25- 7.16 (m, 7H), 7.10-7.09 (d, J= 7.55 Hz, 2H), 7.01-7.00 (d, J= 7.55 Hz, 1H), 6.92- 6.90 (d, J=8.25 Hz, 2H), 6.09- 6.07 (d, J=13.75 Hz, 2H), 3.59 (s, 6H), 2.66 (s, 6H), 2.63- 2.61 (t, J= 6.2 Hz, 4H), 1.88-1.86 (m, 2H), 1.29 (s, 12H).

^{13}C NMR (125 MHz, CDCl$_3$): δ172.68, 151.96, 151.25, 145.96, 142.67, 140.90, 135.80, 135.14, 133.95, 131.67, 130.06, 129.52, 129.44, 129.00, 128.59, 128.19, 126.73, 125.06, 122.76, 122.03, 121.56, 119.52, 115.17, 110.54, 101.52, 48.97, 45.48, 45.22, 31.61, 29.66, 27.63, 26.30, 20.67

HRMS (ESI): Calc. for $C_{50}H_{53}N_4O_2S_2$ [M-Cl]$^+$: 805.3604; Found: 805.3604

Physical Properties: Deep green colored solid. (20 mg, 29 % yield).

Figure S12. ^{1}H and ^{13}C NMR spectra of compound 5

S17
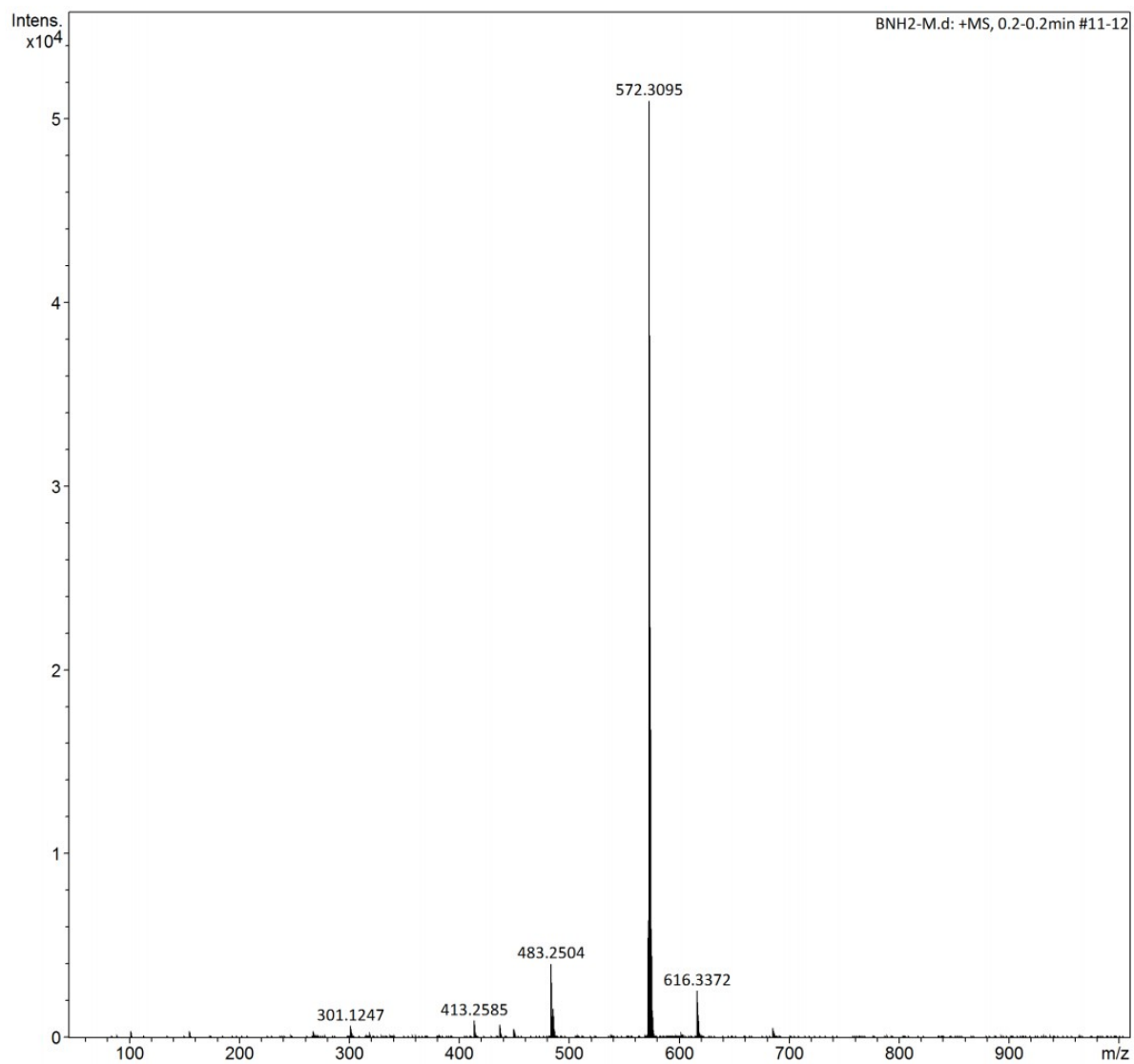
Figure S13. HRMS spectra of compound **5.**
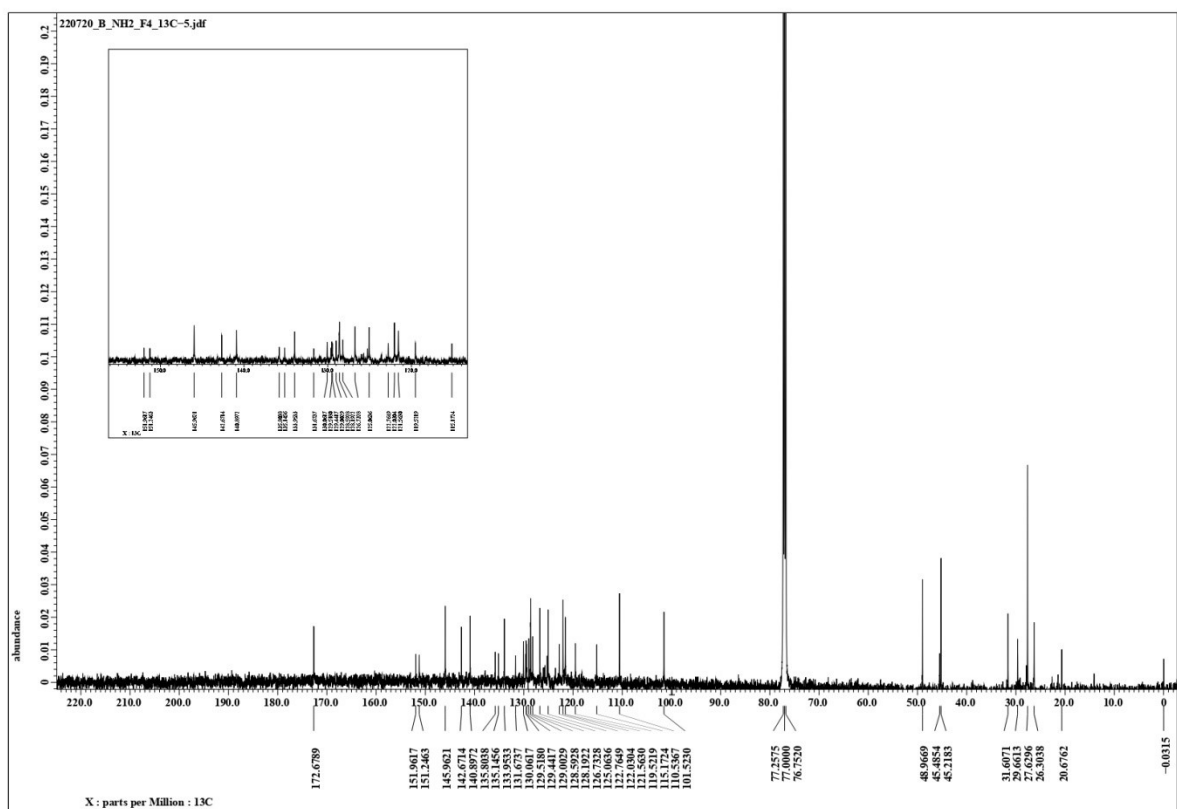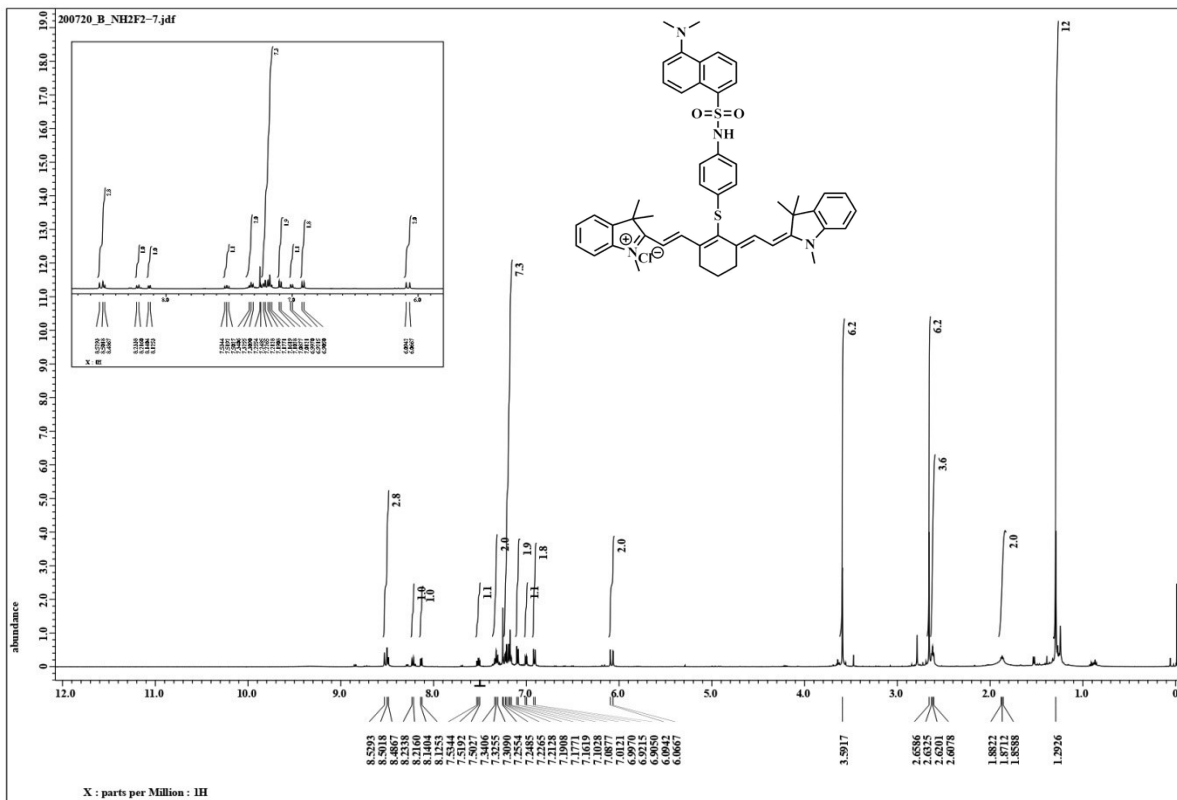
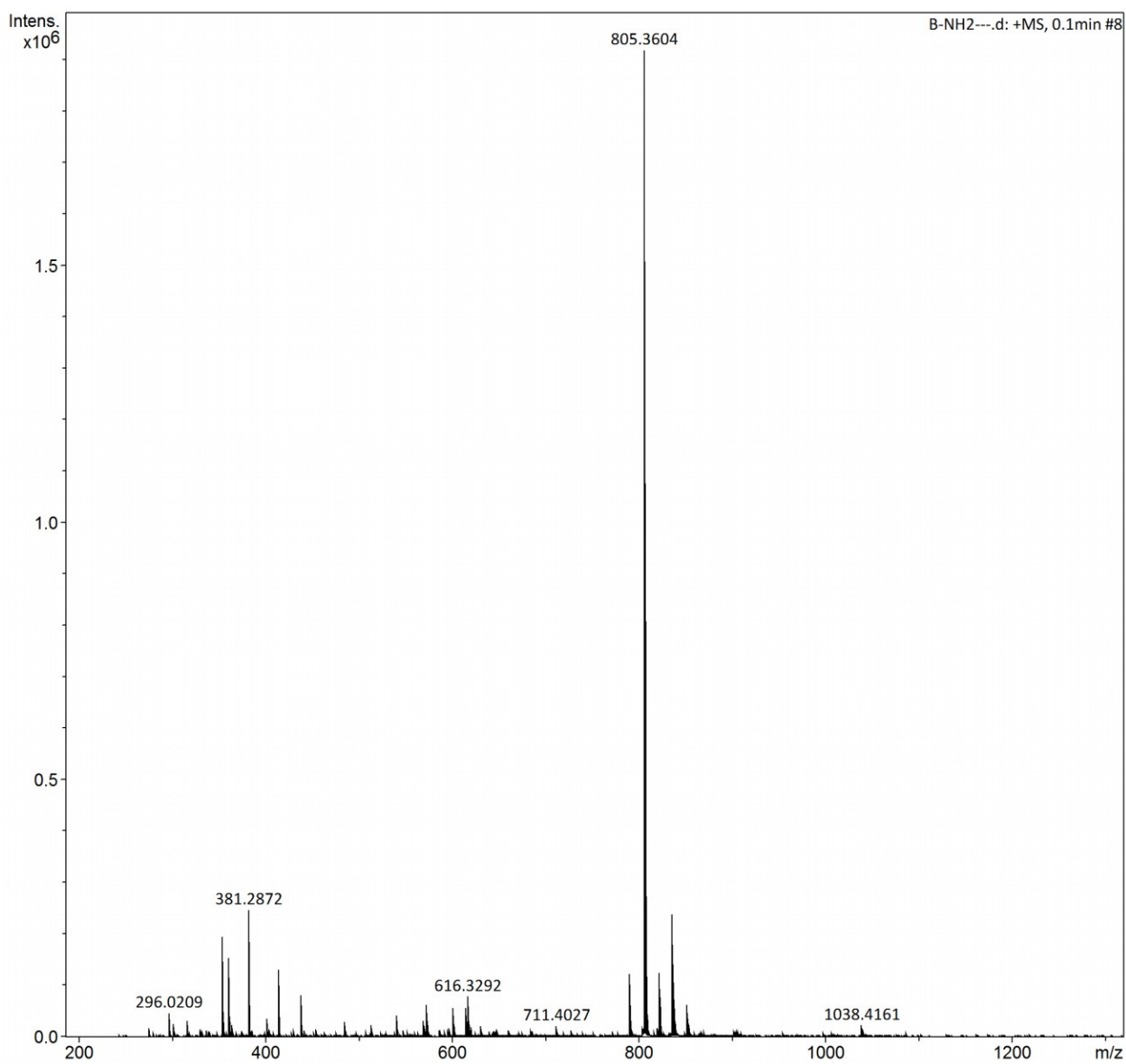**Figure S14.** ^{1}H and ^{13}C NMR spectra of **CyG-NHS**.

Figure S15. HRMS spectra of **CyG-NHS.**

References:

1.   B. Ojha and G. Das, *Photochemical & Photobiological Sciences*, 2011, **10**, 554-560.

2.   C. Würth, M. Grabolle, J. Pauli, M. Spieles and U. Resch-Genger, *Nature protocols*, 2013, **8**, 1535-1550.

3.   K. Rurack and M. Spieles, *Analytical chemistry*, 2011, **83**, 1232-1242.

4.   D. F. Eaton, *Pure and Applied Chemistry*, 1988, **60**, 1107-1114.

5.   B. Liu, Y. Pang, R. Bouhenni, E. Duah, S. Paruchuri and L. McDonald, *Chemical Communications*, 2015, **51**, 11060-11063.

6.   S. Dogra, A. K. Kar, K. Girdhar, P. V. Daniel, S. Chatterjee, A. Choubey, S. Ghosh, S. Patnaik, D. Ghosh and P. Mondal, *Nanomedicine: Nanotechnology, Biology and Medicine*, 2019, **17**, 210-222.

7.   S. Sugio, A. Kashima, S. Mochizuki, M. Noda and K. Kobayashi, *Protein engineering*, 1999, **12**, 439-446.

8.   G. Schaftenaar and J. H. Noordik, *Journal of computer-aided molecular design*, 2000, **14**, 123-134.

9.   R. A. Gaussian09, *Inc., Wallingford CT*, 2009, **121**, 150-166.

10.   D. Case, T. Darden, T. Cheatham III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang and K. Merz, AmberTools 16, University of California, San Francisco, 2016.

11.   G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *Journal of computational chemistry*, 2009, **30**, 2785-2791.

12.   E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *Journal of computational chemistry*, 2004, **25**, 1605-1612.

13.   D. Case, I. Ben-Shalom, S. Brozell, D. Cerutti, T. Cheatham III, V. Cruzeiro, T. Darden, R. Duke, D. Ghoreishi and M. Gilson, *University of California, San Francisco*, 2018.

14.   J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *Journal of computational chemistry*, 2004, **25**, 1157-1174.

15.   L. Wickstrom, A. Okur and C. Simmerling, *Biophysical journal*, 2009, **97**, 853-856.

16.   B. R. Miller III, T. D. McGee Jr, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *Journal of chemical theory and computation*, 2012, **8**, 3314-3321.

17. Y. Li, Y. Sun, J. Li, Q. Su, W. Yuan, Y. Dai, C. Han, Q. Wang, W. Feng and F. Li, *Journal of the American Chemical Society*, 2015, **137**, 6407-6416.

18. C. Hu, W. Sun, J. Cao, P. Gao, J. Wang, J. Fan, F. Song, S. Sun and X. Peng, *Organic letters*, 2013, **15**, 4022-4025.

19. Z. Guo, S. Nam, S. Park and J. Yoon, *Chemical Science*, 2012, **3**, 2760-2765.

20. D. L. Gallaher Jr and M. E. Johnson, *Analyst*, 1999, **124**, 1541-1546.