Supporting Information for

Specific protein-labeling and ligand-binding position analysis with amidopyrene probes as LDI MS tags

Rei Watanabe,[a] Yaping Hu,[a] Keita Iio,[a] Kozo Yoneda,[a] Atsunori Hattori,[b] Atsushi Arai,[b] Hideo Kigoshi,*[,b] and Masaki Kita *[,a–c]

[a] *Graduate School of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, 305-8571, Japan.*
E-mail: kigoshi@chem.tsukuba.ac.jp
[b] *Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa, Nagoya 464-8601, Japan.*
E-mail: mkita@agr.nagoya-u.ac.jp
[c] *PRESTO, Japan Science and Technology Agency*

(19 pages)

Table S1. Tryptic peptides of the avidin labeled with probes **1**–**3** detected by MALDI MS.

	Observed (*m/z*) [a]								Calculated (*m/z*) [a]			
	Avidin + **1**		Avidin + **2**		Avidin + **3**		Avidin					
No.	Z18 (Fig. S2a)	G3000S (Fig. 3a)	Z18 (Fig. S2b)	G3000S (Fig. 3b)	Z18 (Fig. S2c)	G3000S (Fig. 3c)	Z18 (Fig. S2d)	G3000S (Fig. 3d)	apy-peptide	peptide	Start/end	Sequence [b]
1			774.5				774.5			774.4	147 / 152	LRTQ**K**E
					1101.7	**1101.5**			**1101.6**			
2										793.4	27 / 33	**K**C(Carbamidomethyl)SLTGK
					1120.6	**1120.5**			**1120.5**			
3	819.5	819.4	819.5	819.4	819.5	819.4	819.5	819.4		819.5	119 / 124	TMWLLR
4	919.5	919.5	919.6	919.5	919.6	919.5	919.6	919.5		919.5	139 / 146	VGINIFTR
5										1020.6	25 / 33	AR**K**C(Carbamidomethyl)SLTGK
					1347.8				**1347.7**			
6			1188.9			1188.7				1188.7	139 / 148	VGINIFTRLR
7			1235.7		1235.7		1235.7			1235.6	125 / 135	SSVNDIGDDWK
8	1425.7	1425.7	1425.8	1425.6	1425.8	1425.6	1425.9	1425.7		1425.7	84 / 95	TQPTFGFTVNWK
9			1563.9		1563.8	1563.7	1563.9			1563.7	125 / 138	SSVNDIGDDW**K**ATR
			1891.0	**1890.7**	**1891.0**	**1890.7**			**1890.9**			
10	1581.8		1581.9		1581.9		1582.0	1581.7		1581.8	83 / 95	RTQPTFGFTVNWK
11	1594.8		1595.0				1595.0			1594.8	70 / 83	ESPLHGTQNTIN**K**R
			1922.1	**1921.8**	**1922.1**	**1921.8**			**1921.9**			
12	1894.9	1894.8	1895.0	1894.7	1895.0	1894.7	1895.0	1894.8		1894.9	96 / 111	FSESTTVFTGQCFIDR
13		2002.9	2003.1		2003.1	2002.9	2003.2			2003.0	51 / 69	GEFTGTYITAVTATSNEIK
14	2036.7					2035.8		2035.9		2036.0	119 / 135	TMWLLRSSVNDIGDDWK
15	2195.0	2195.0	2195.2	2194.8	2195.1	2194.9	2195.2	2195.0		2194.0	96 / 114	FSESTTVFTGQCFIDRNGK
16	**2561.3**	**2561.2**							**2561.3**		125 / 138	SSVNDIGDDW**K**ATR + probe **1**
17	2664.4	2664.3	2664.6	2664.2	2664.5	2664.2	2664.6	2664.3		2663.3	96 / 118	FSESTTVFTGQCFIDRNGKEVLK
18			3052.7		3052.6	3052.3	3052.7			1835.9 [c]	34 / 50	WTNDLGS<u>N</u>MTIGAVNSR
			3296.9		3296.8	3296.5	3296.9	3296.6				
			3376.9		3376.8	3376.5	3376.9	3376.6				
19			3459.0		3458.9	3458.6	3459.0	3458.7		3457.7	83 / 111	RTQPTFGFTVNWKFSESTTVFTGQCFIDR
20	3579.4	3579.1	3579.4	3579.0	3579.3	3579.3	3579.4	3579.1		3578.8	51 / 83	GEFTGTYITAVTATSNEI**K**ESPLHGTQNTIN**K**R
						3906.2			**3905.9**			
21			3601.4							3600.7	84 / 114	TQPTFGFTVNWKFSESTTVFTGQCFIDRNGK

[a] The data represent the monoisotopic ion peaks $(M+H)^+$ values. Bold value means the data for apy-labeled peptides.

[b] "**K**" means apy-labeled lysine residue. Cysteine residues are carbamidomethylated.

[c] Calculated value without *N*-glycans on the N41 residue (underlined). The common *N*-terminal sequences (W34–S40) were determined by MS/MS analysis.

Table S2. Calculated distances between the Lys residues on avidin and the succinyl ester carbons in probes **1** and **2** in the most stable complex conformers obtained by docking simulations (see Figs. 4a and 4b).

Lys residue	Distances (Å)	
	1	**2**
27	39.33	42.01
33	32.33	34.35
69	20.59	21.99
82	45.64	48.35
95	17.57	19.64
114	40.31	42.39
118	32.12	34.29
135	8.48	8.58
151	44.99	47.35

Table S3. Boltzmann weights of the five lowest-energy conformers of **1** and **2** obtained by docking simulations (see Figs. 4a and 4b). Calculated distances between the ε-amino nitrogen atom of the K135 on avidin and the succinyl ester carbon in each conformer, and the RMSD values for the biotin ligand (15 atoms, red + blue) between the calculated probes and the original complexes on avidin are shown.

Figure 4a

1	⊿G (kcal/mol)	ratio (%)	Distance (Å)	RMSD (Å)
conformer 1	0.000	66.03	8.48	0.463
conformer 2	0.793	17.31	13.24	0.998
conformer 3	1.249	8.01	11.52	3.013
conformer 4	1.503	5.21	4.95	4.420
conformer 5	1.750	3.43	9.49	1.896

Figure 4b

2	⊿G (kcal/mol)	ratio (%)	Distance (Å)	RMSD (Å)
conformer 1	0.000	69.64	8.58	0.378
conformer 2	0.970	13.53	10.11	0.676
conformer 3	1.250	8.43	4.43	2.707
conformer 4	1.610	4.59	5.67	4.515
conformer 5	1.720	3.81	14.05	0.783

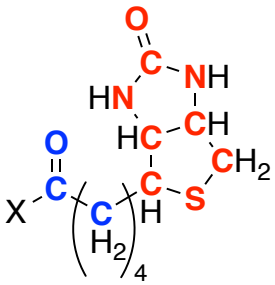


Table S4. Boltzmann weights of the five lowest-energy conformers of **1** and **2** obtained by docking simulations (see Figs. 4c–4e). RMSD values for the biotin ligand [15 atoms (red + blue) and bicyclic 9 atoms (red)] between the calculated probes and the original complexes on avidin are shown.

Figure 4c

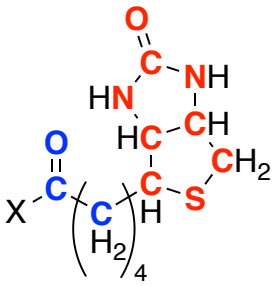
1	⊿G (kcal/mol)	ratio (%)	RMSD (Å)	
			15 atoms	9 atoms
conformer 1	0.000	72.91	0.786	0.198
conformer 2	0.899	15.96	7.300	8.177
conformer 3	1.421	6.61	6.215	7.354
conformer 4	2.003	2.47	19.683	23.184
conformer 5	2.113	2.05	7.278	7.767

Figure 4d

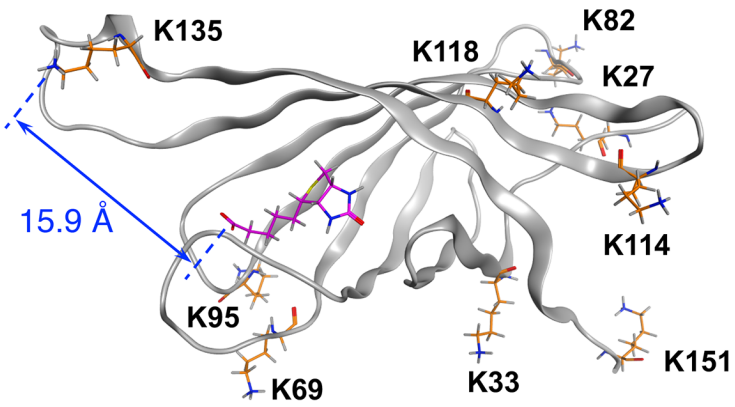
2	⊿G (kcal/mol)	ratio (%)	RMSD (Å)	
			15 atoms	9 atoms
conformer 1	0.000	83.00	0.332	0.305
conformer 2	1.329	8.80	1.267	1.157
conformer 3	1.553	6.03	3.725	0.854
conformer 4	2.512	1.19	5.951	4.483
conformer 5	2.624	0.99	6.260	7.422

Figure 4e

1	⊿G (kcal/mol)	ratio (%)	RMSD (Å)	
			15 atoms	9 atoms
conformer 1	0.000	52.74	4.459	0.952
conformer 2	0.478	23.53	4.959	2.486
conformer 3	1.008	9.61	1.436	1.411
conformer 4	1.106	8.14	11.078	11.394
conformer 5	1.287	5.99	11.413	11.668



(a)



(b)

```
 25   ARKCSLTGKWTNDLGSNMTIGAVNSRGEFTGTYITAVTATSNEIKE    70
 71   SPLHGTQNTINKRTQPTFGFTVNWKFSESTTVFTGQCFIDRNGKEV   116
117   LKTMWLLRSSVNDIGDDWKATRVGINIFTRLRTQKE             152
```

Figure S1. Structures of avidin and the complex with biotin. (a) X-ray structure of the biotin–avidin complex (PDB code: 2AVI, monomeric complex is shown). The ligand biotin and eight Lys residues are highlighted in magenta and orange, respectively. (b) Amino acid sequences of avidin. Lys residues are highlighted in bold red. *N*-glycosylated Asn residue is highlighted in blue and underlined.

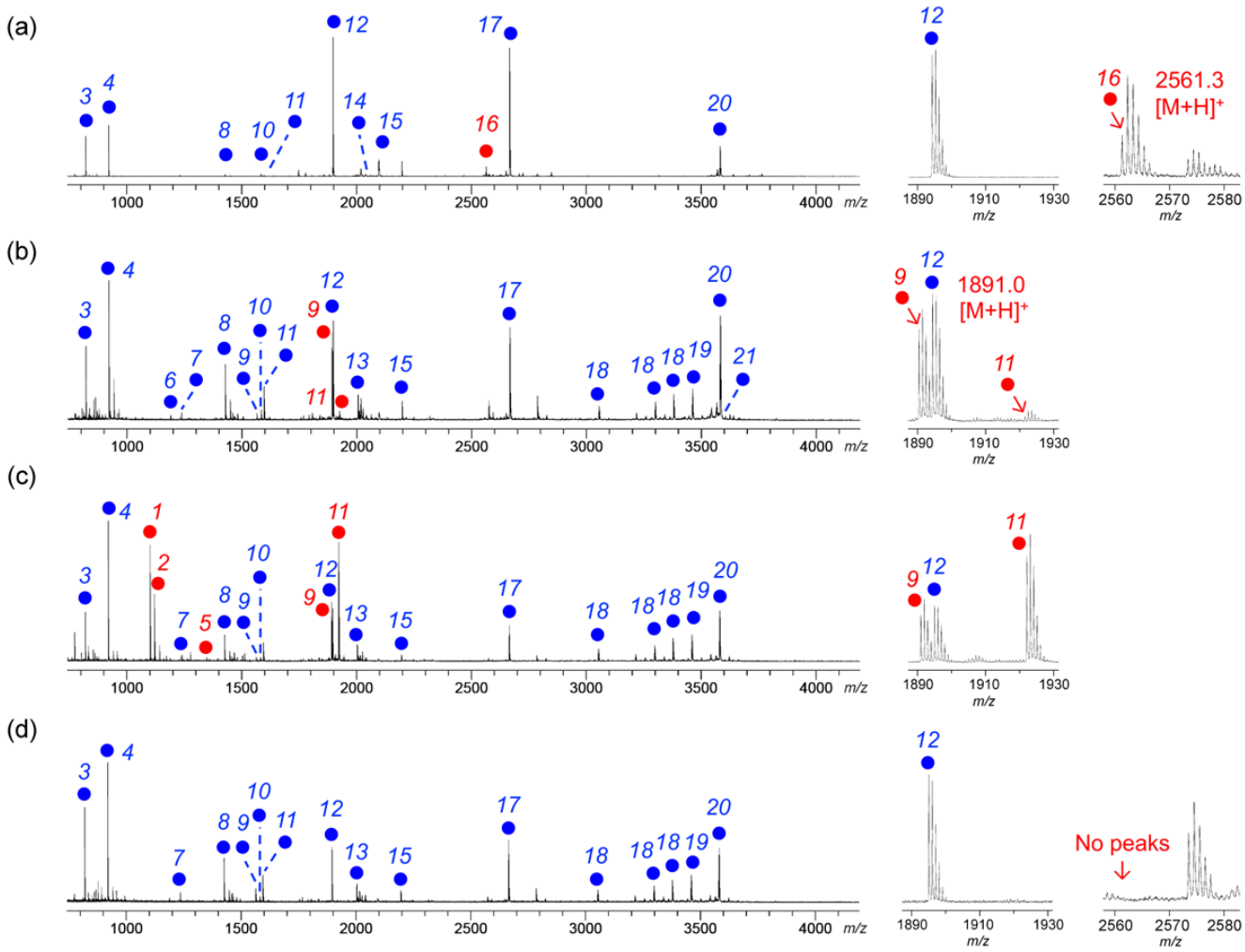


Figure S2. MALDI-TOF mass spectra of the tryptic peptides of avidin (2 pmol) labeled with probes **1** (a), **2** (b), or **3** (c), and unlabeled avidin (d). Desalting was performed using a reversed-phase ODS tip column (ZipTip® C_{18}). Red and blue circles indicate amidopyrene-labeled and unlabeled avidin peptides, respectively. Numerals indicate the positions of tryptic peptides in avidin (for details, see Table S1).

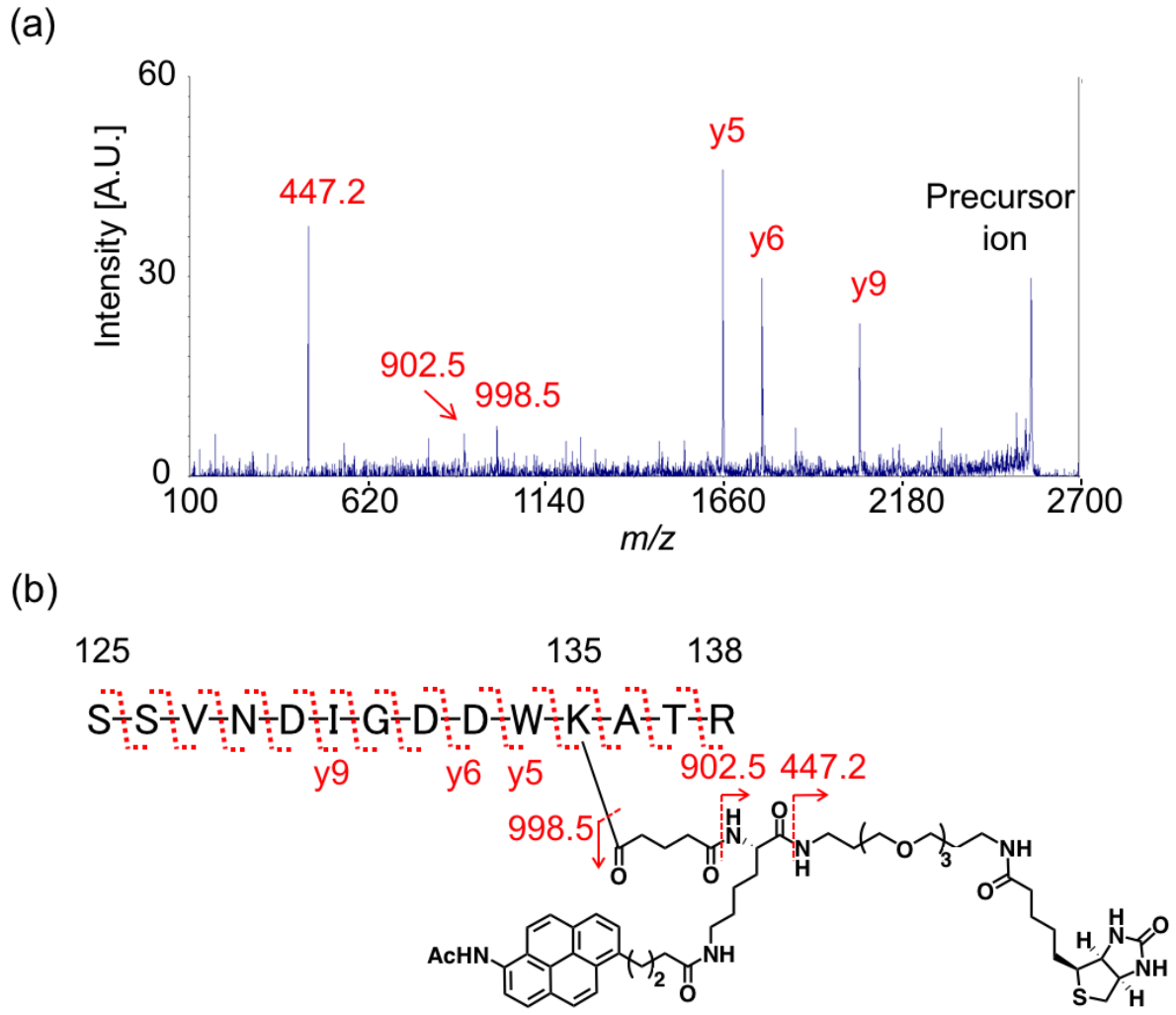
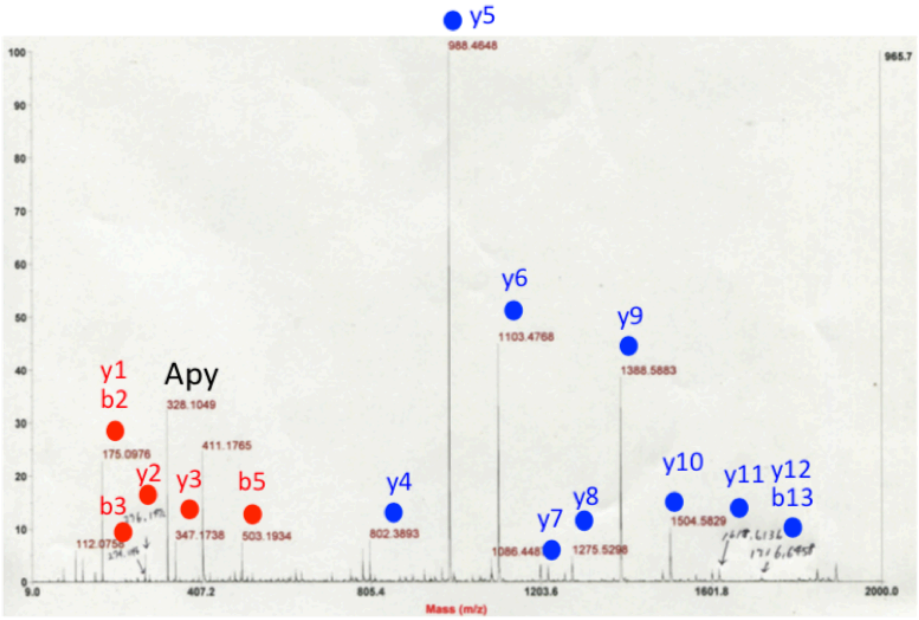


Figure S3. MALDI MS/MS analysis of the peptide labeled with probe **1** (No. 16, precursor ion: *m/z* 2561.4 for $[M+H]^+$, see Figs. 3a and S2a). (a) MALDI MS/MS. (b) Structure of the K^{135}-labeled peptide.

(a)



(b)

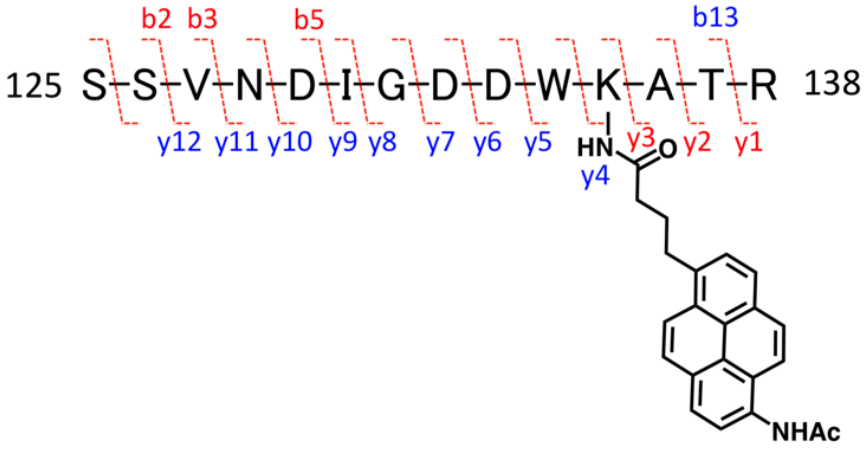


Figure S4. MALDI MS/MS analysis of the peptide labeled with probe **2** (No. 9, precursor ion: *m/z* 1890.6 for $[M+H]^+$, see Figs. 3b and S2b). (a) MALDI MS/MS. (b) Structure of the K^{135}-labeled peptide.

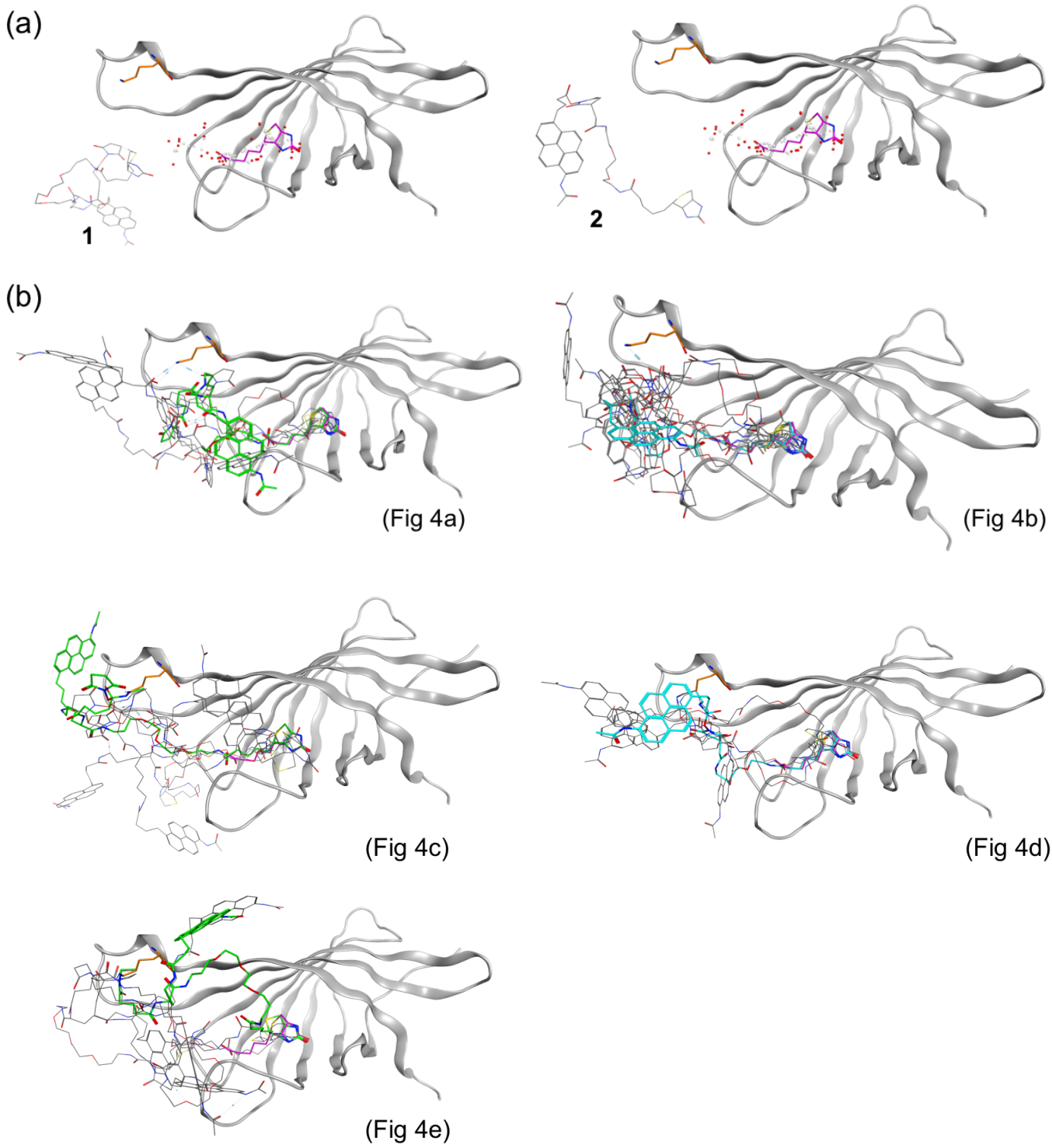
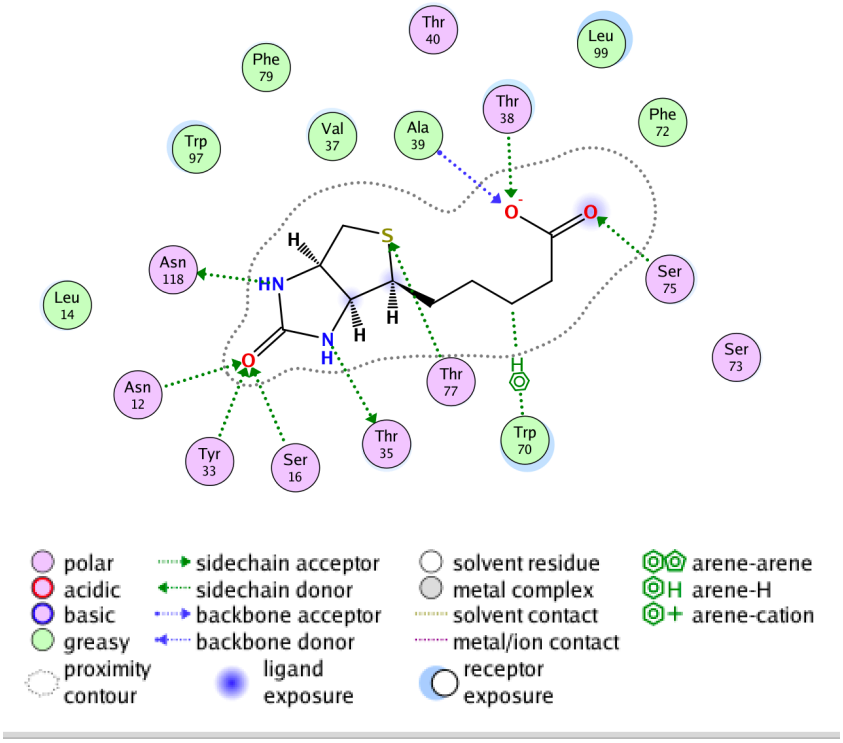


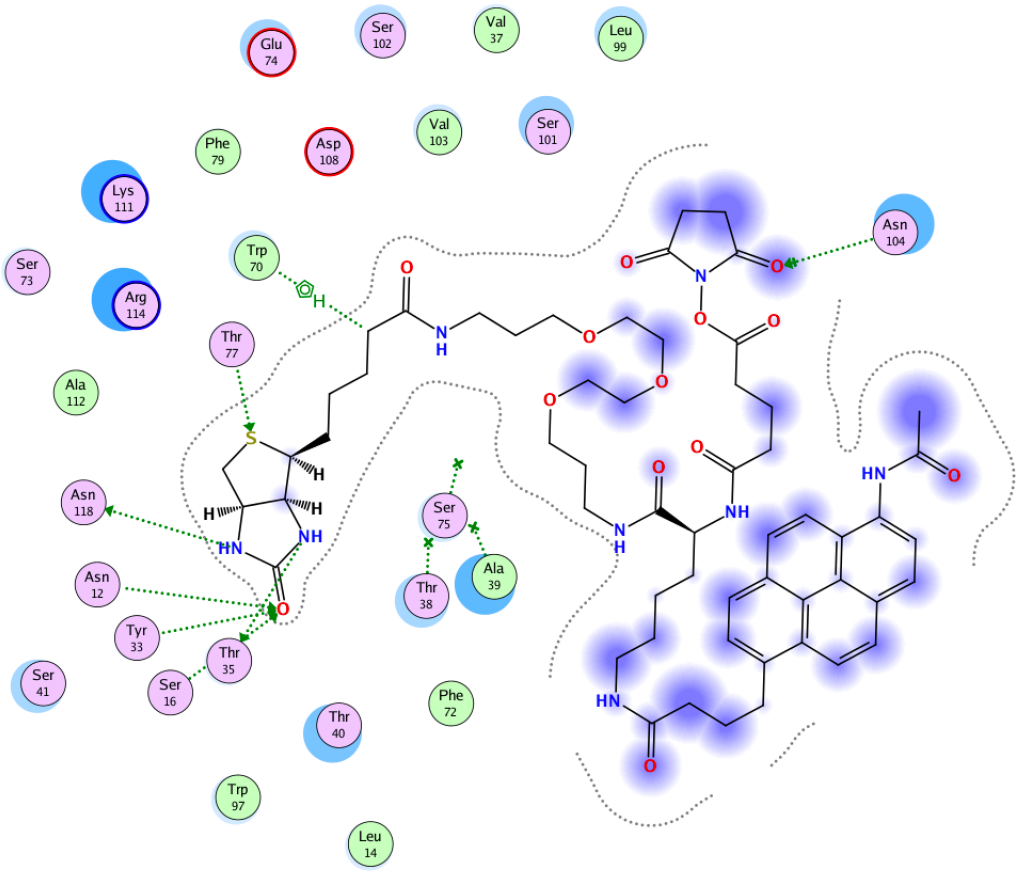
Figure S5. Molecular modeling of the complexes of avidin and amidopyrene biotin probes. (a) Initial model structures for docking simulations (Figs. 4a and 4b). The ligand biotin on the complex is shown in magenta. Probes **1** and **2** (grey) were initially located apart from the target protein avidin. The dummy ligands used for the docking simulation are shown as multiple white (hydrophobic) and red (hydrophilic) spheres, respectively, which represent 23 residues (N36, L38, S40, Y57, T59, V61–T64, W94, F96–S99, T101, F103, W121, L123–S126, R138, and N142) (b) Superimposed structures of the five lowest-energy conformers of **1** and **2** on avidin. The most stable conformers of **1** and **2** are highlighted in green and cyan, respectively.

(a)



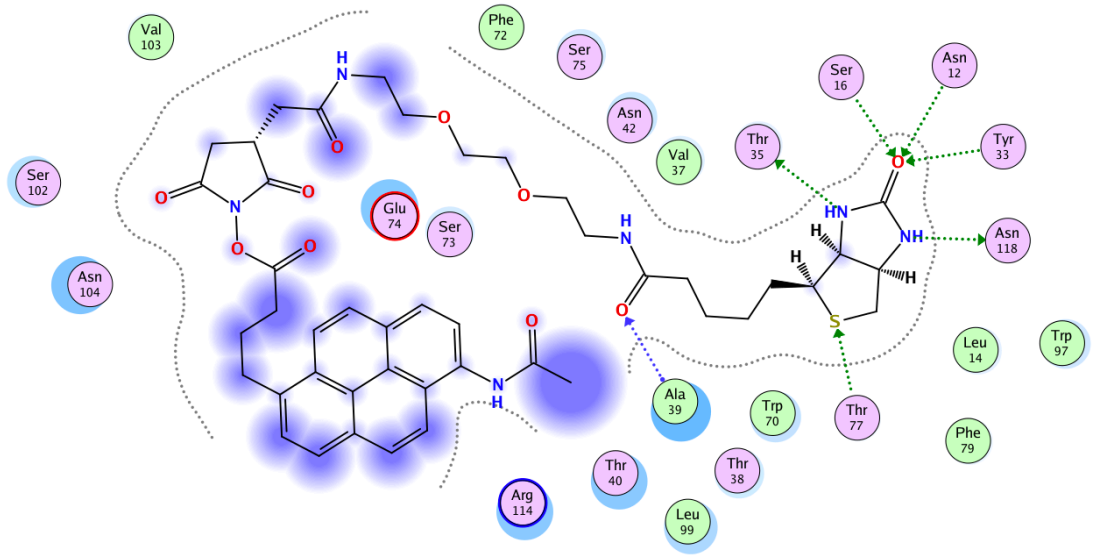
(Fig. S1a)

(b)

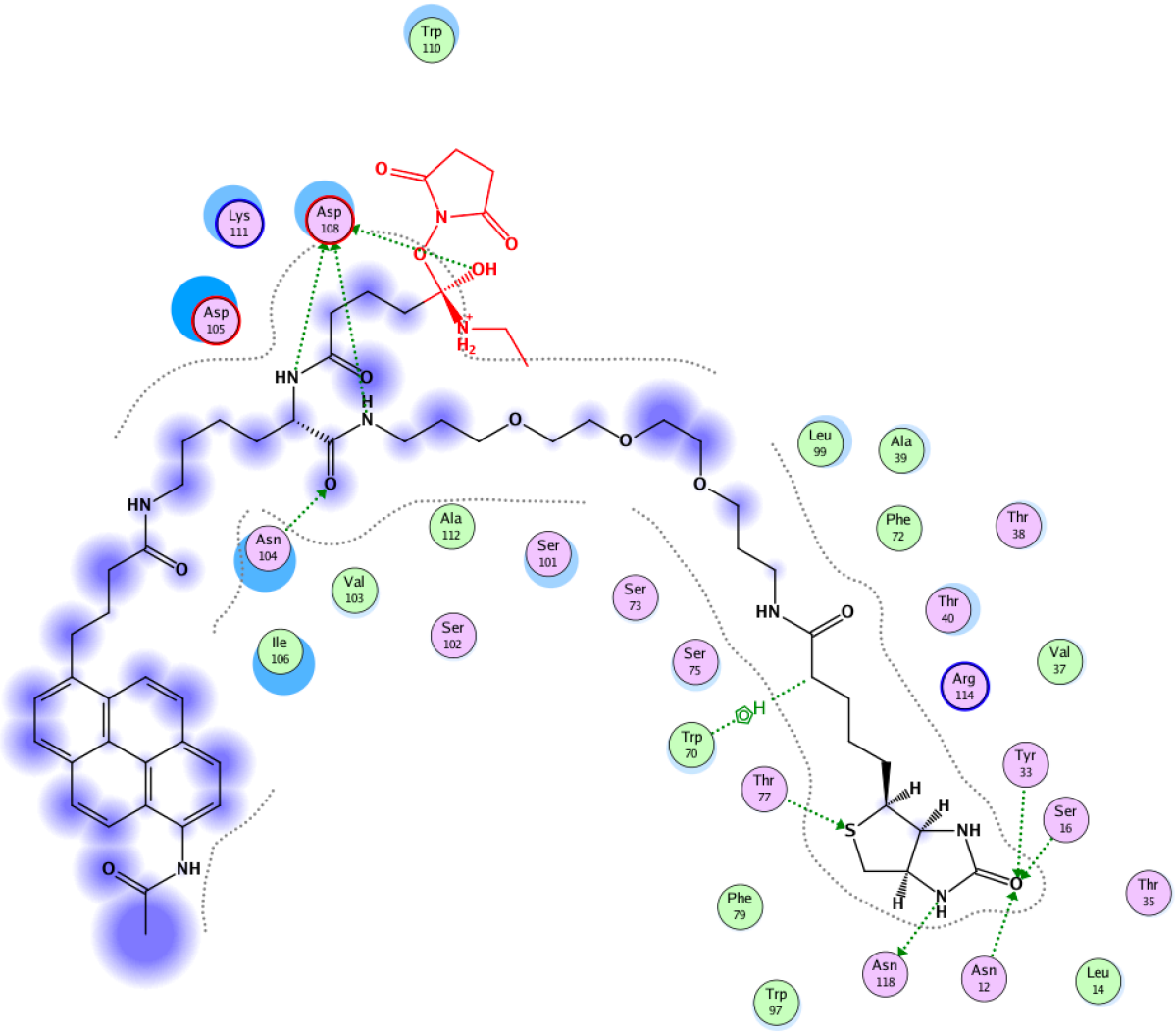


(Fig. 4a)

Figure S6. LIG-PLOT analysis. (a) Interaction of the ligand biotin with avidin on the crystal structure (PDB ID: 2avi). (b) Interactions of the amidopyrene biotin probes with avidin on the calculated complex structures in Fig. 4. The amino acid residue numbers in circles are those of mature proteins (24 aa smaller than the original, see Fig. S1).

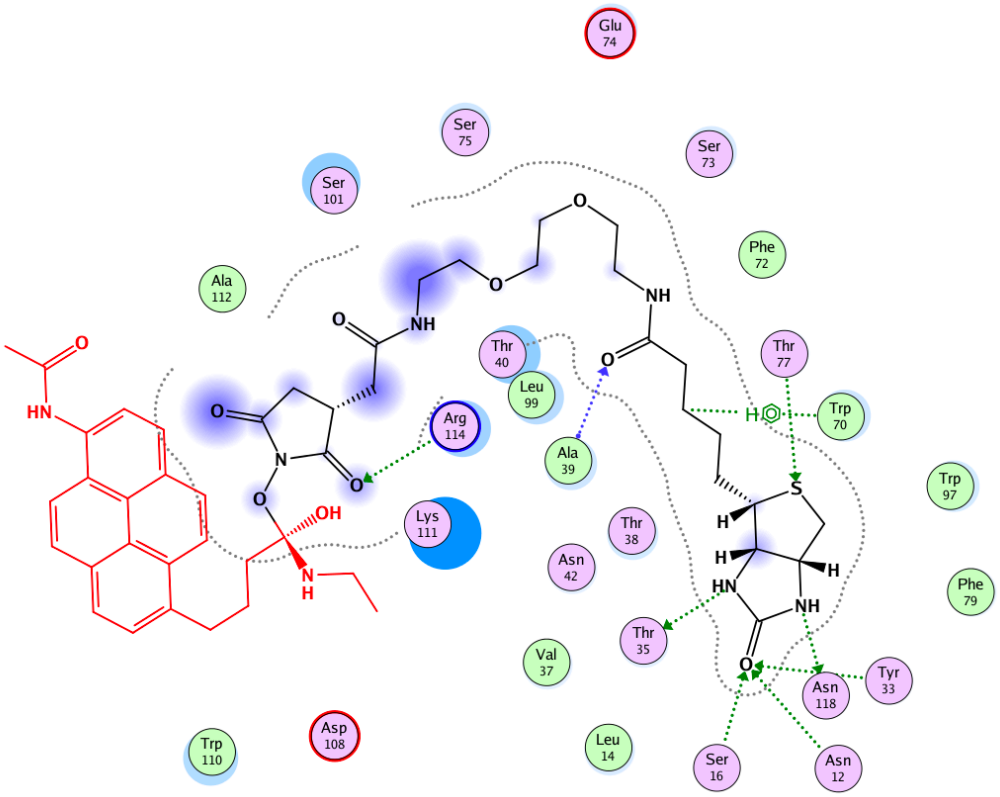


(Fig. 4b)

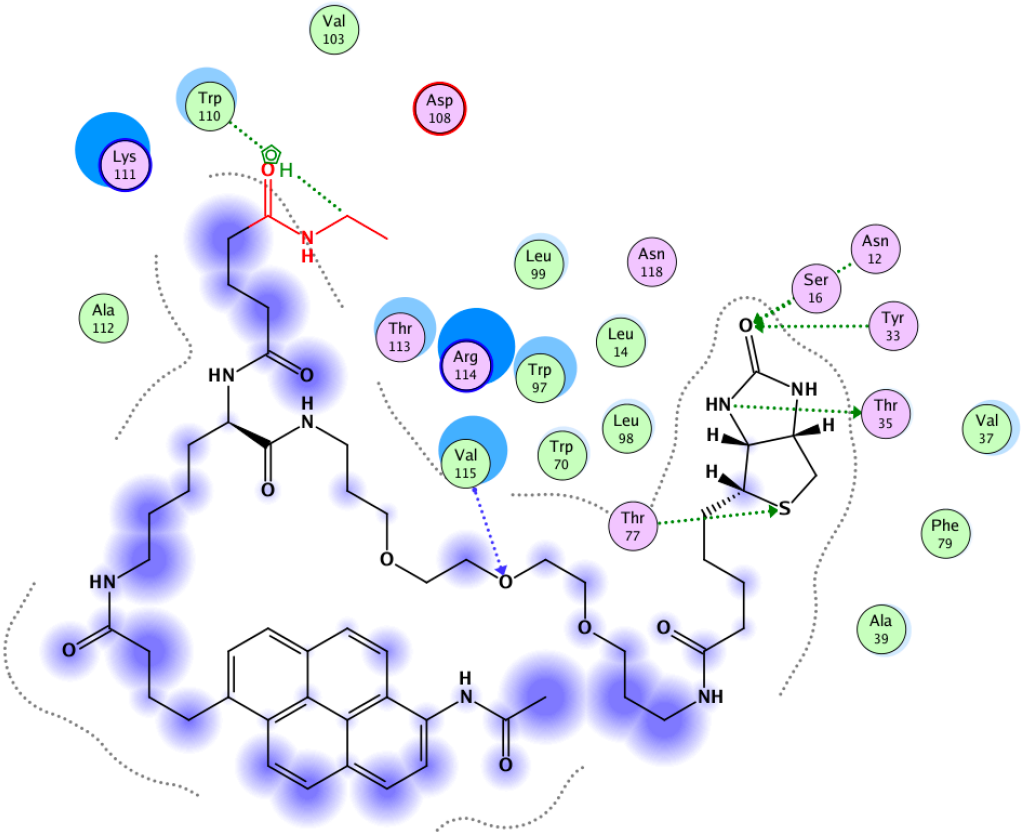


(Fig. 4c)

Figure S6. (continued)



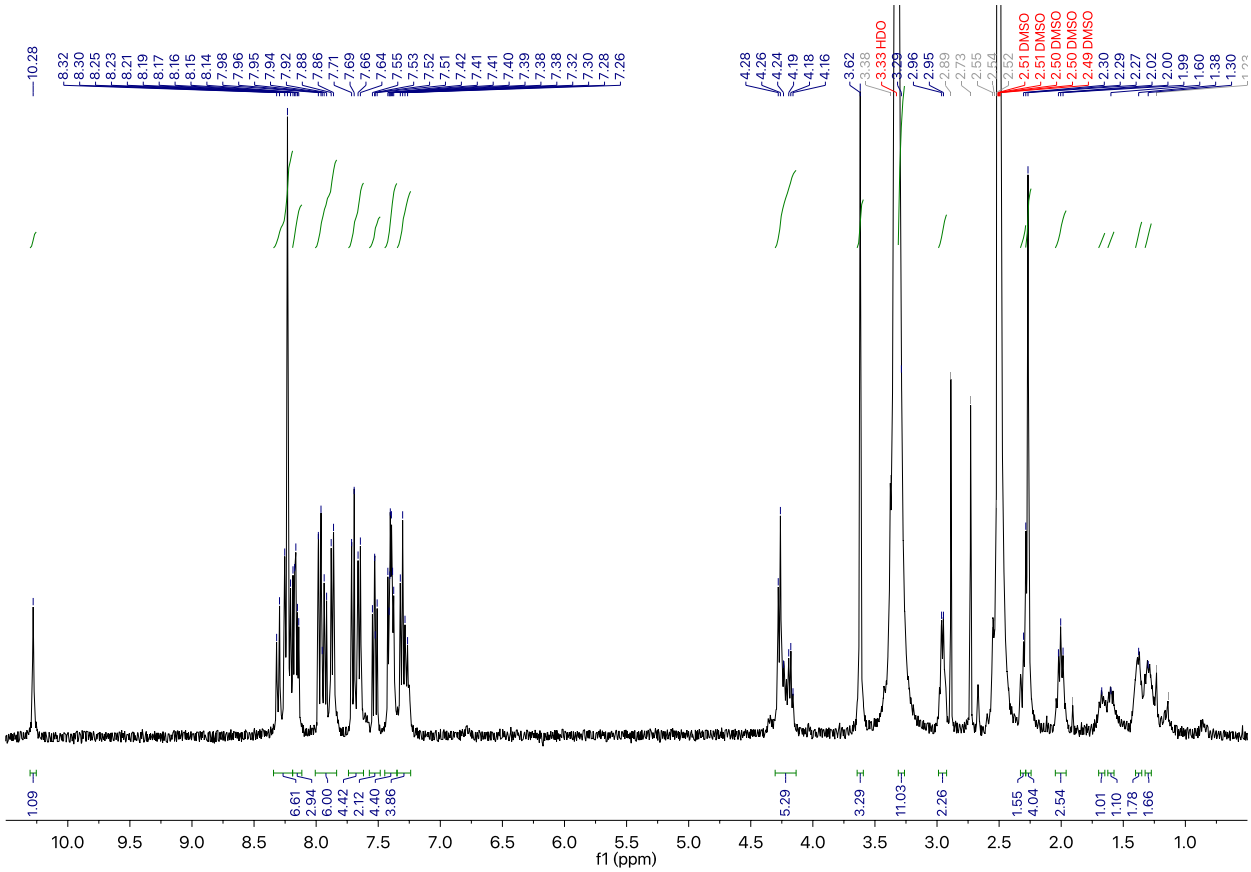
(Fig. 4d)



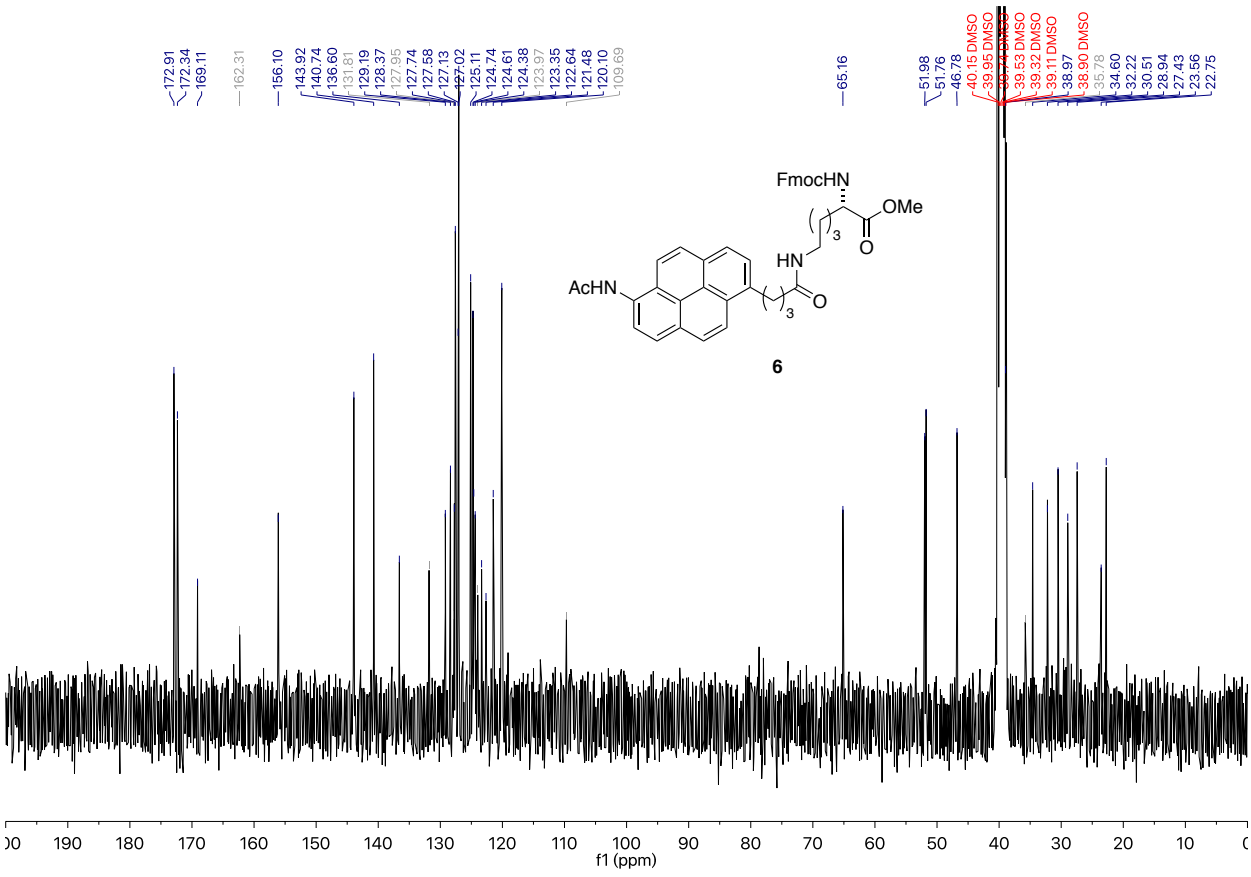
(Fig. 4e)

Figure S6. (continued)

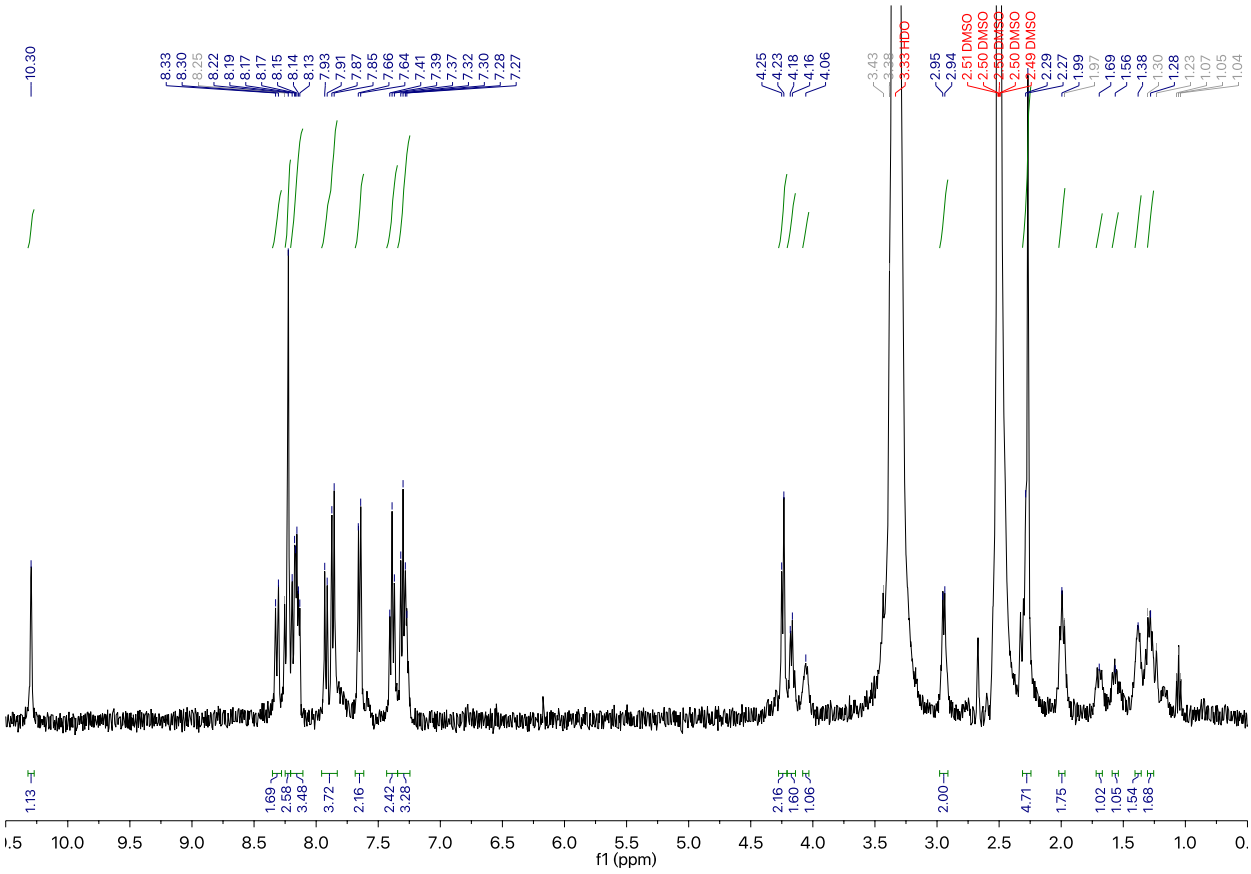
NMR spectra of new compounds.



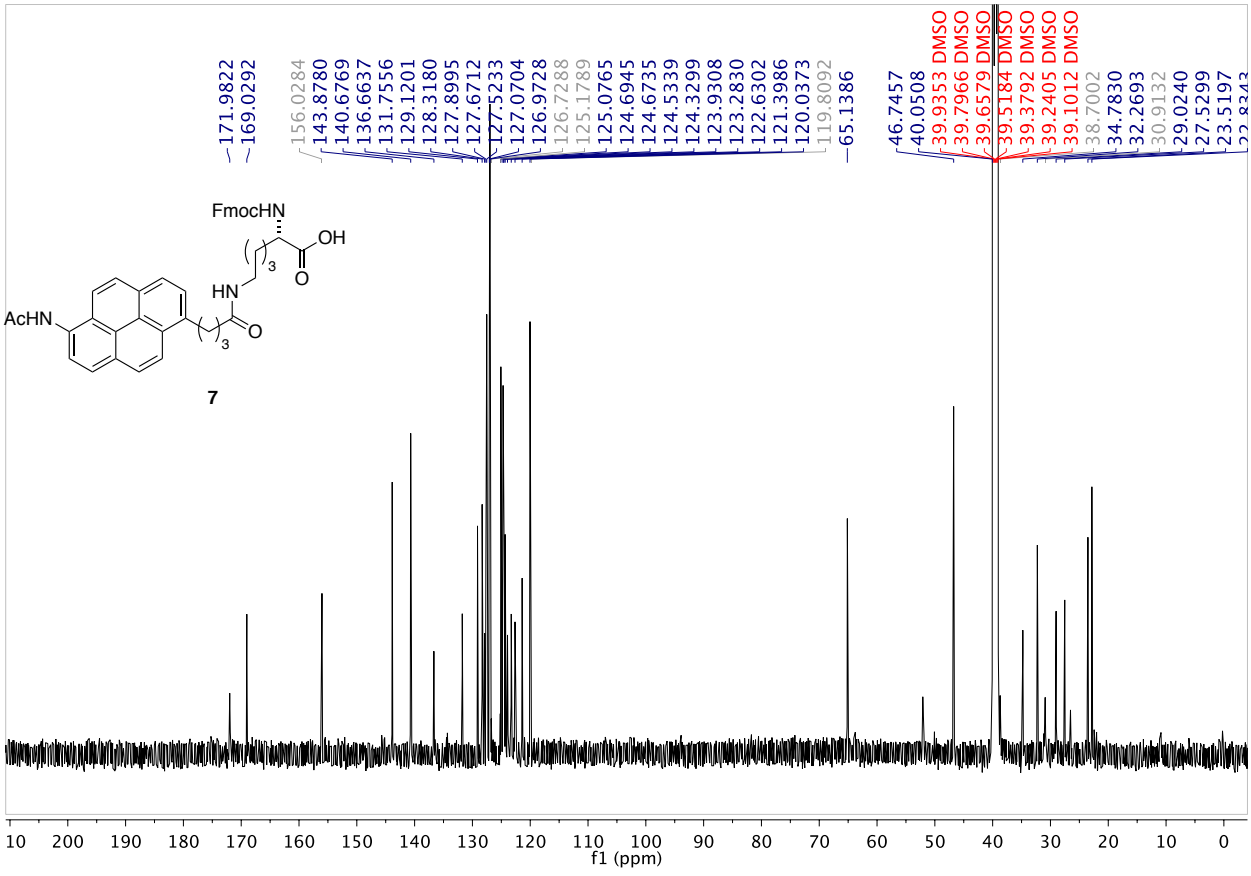
^{1}H NMR spectrum of **6** (400 MHz, DMSO-d_6).



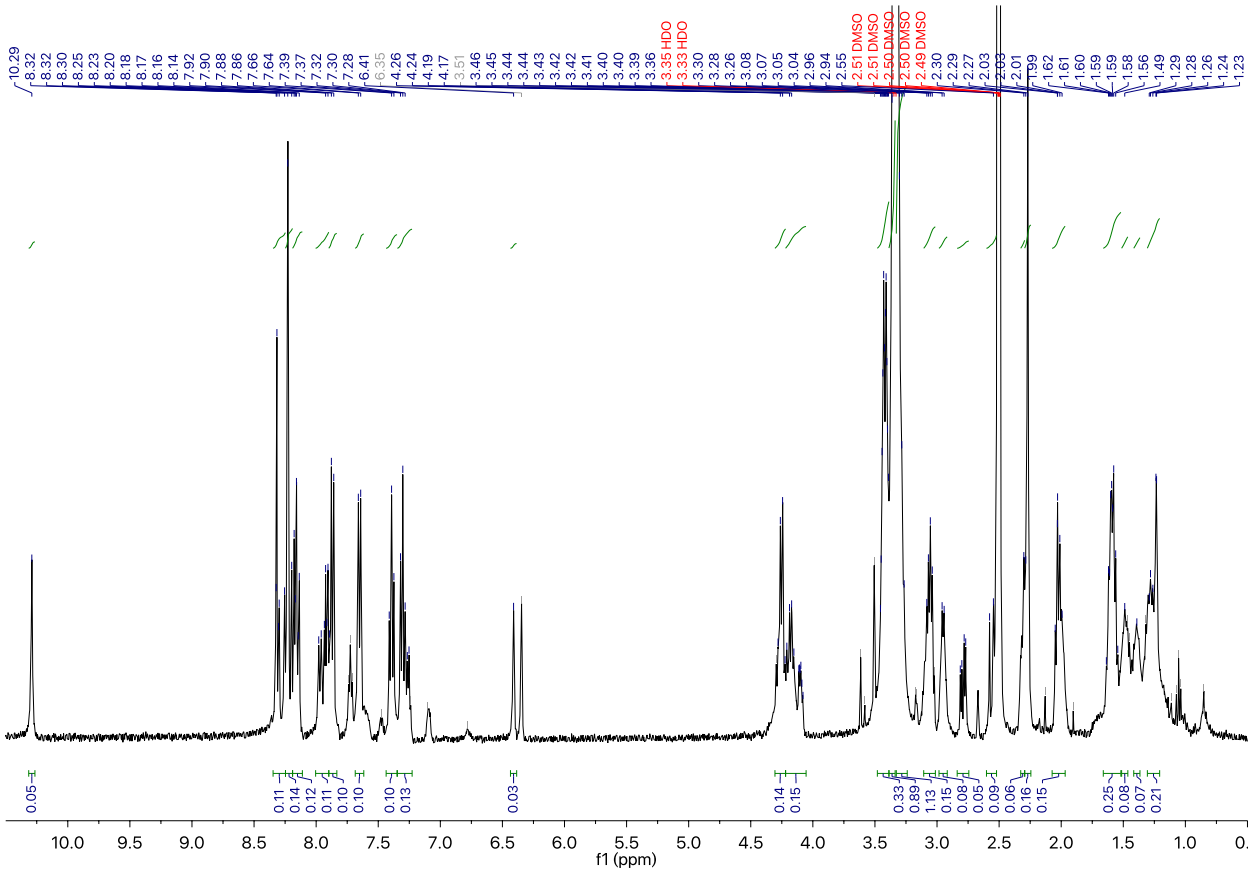
^{13}C NMR spectrum of **6** (100 MHz, DMSO-d_6).



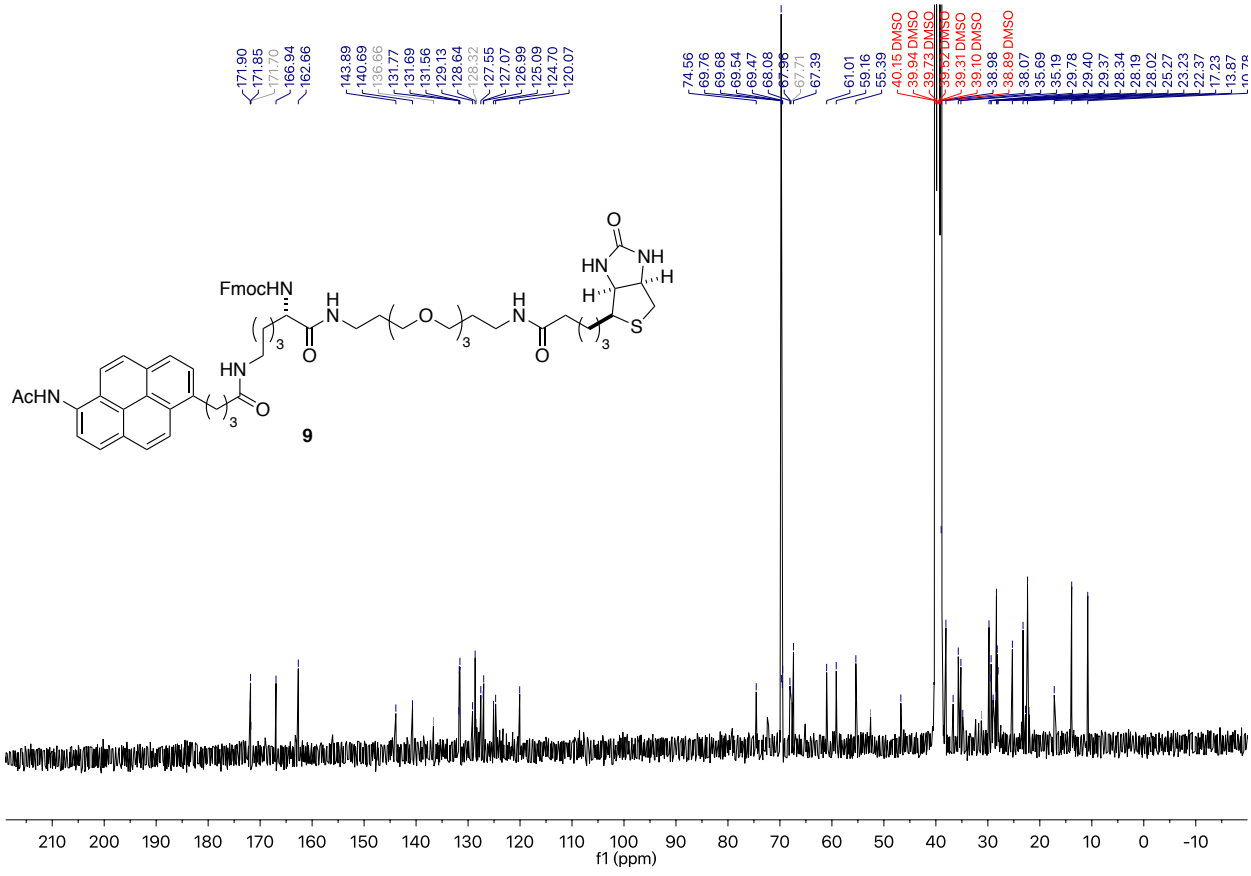
^{1}H NMR spectrum of **7** (400 MHz, DMSO-d_6).



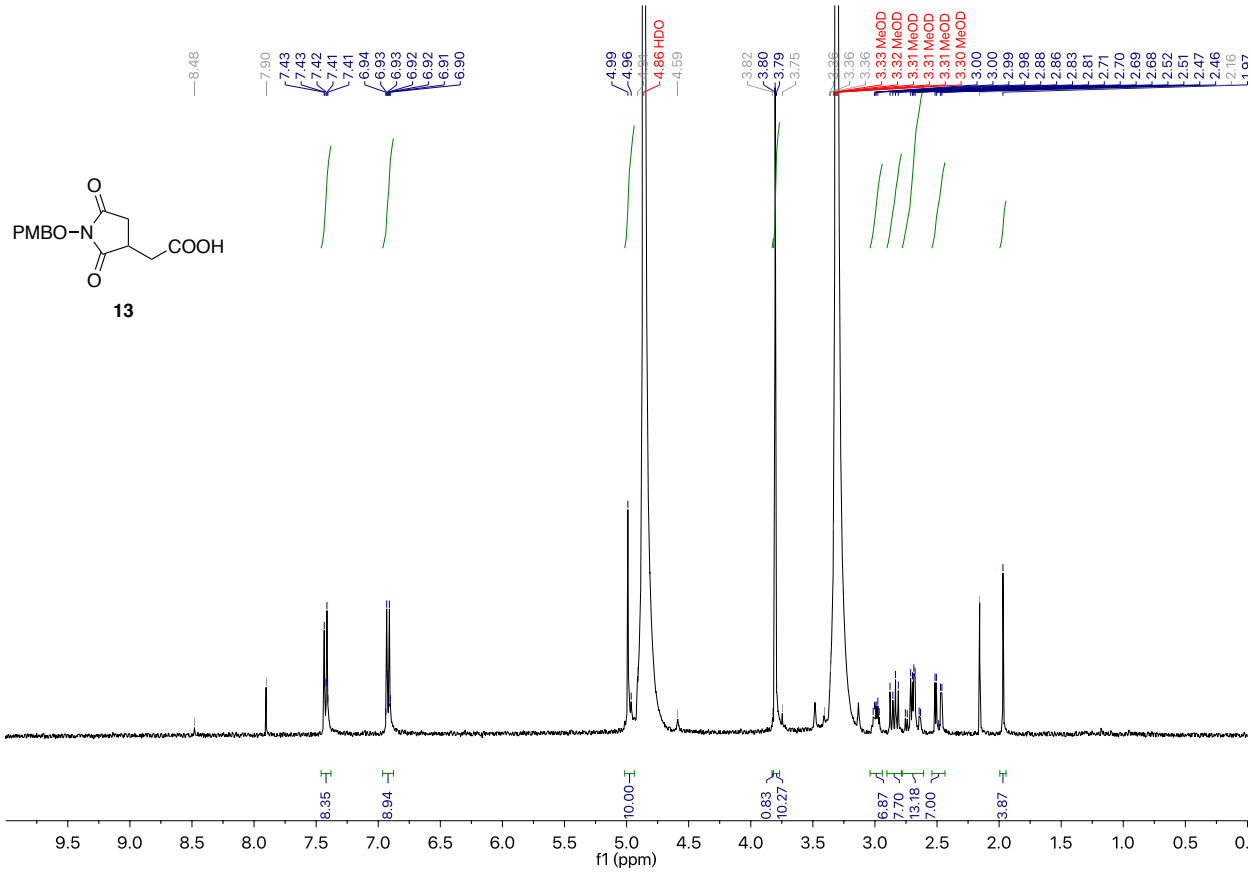
^{13}C NMR spectrum of **7** (150 MHz, DMSO-d_6).



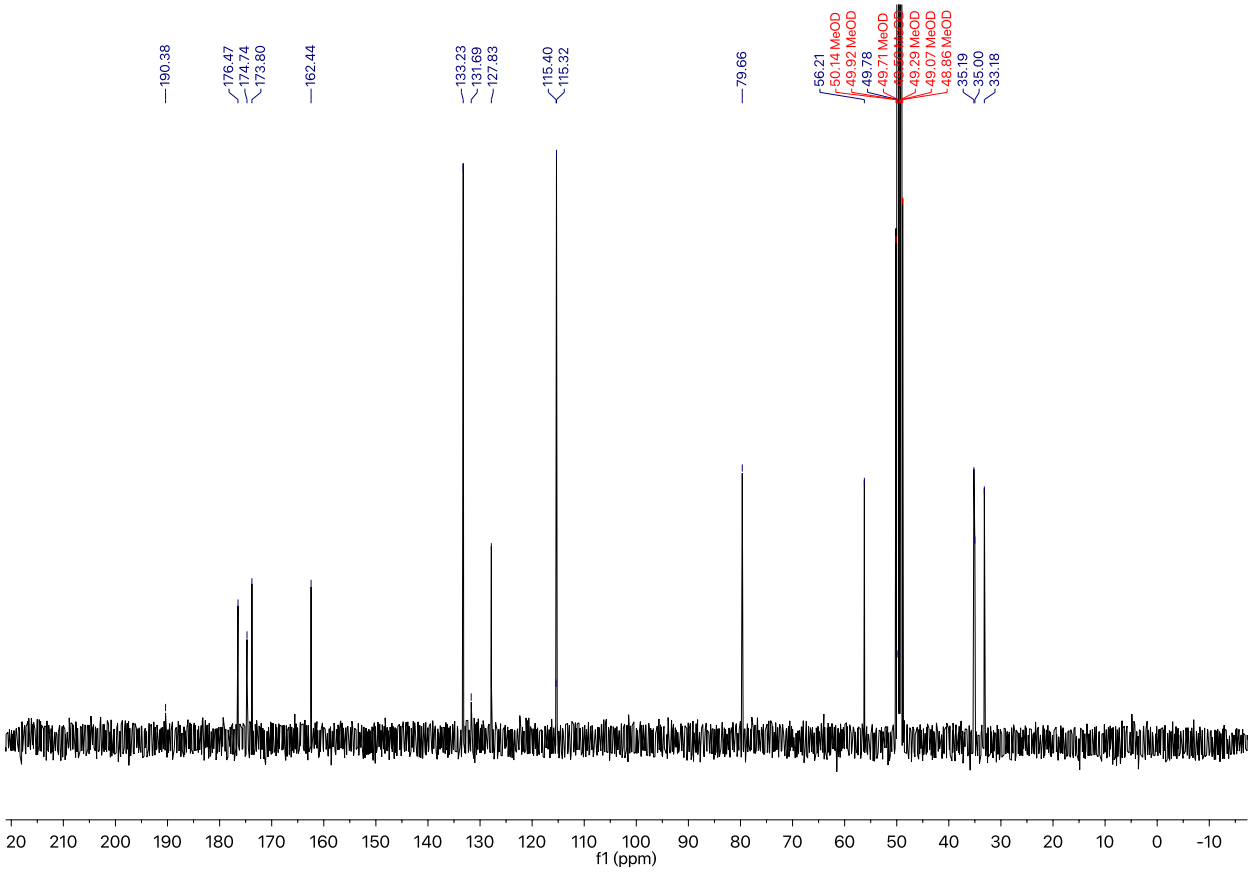
^{1}H NMR spectrum of **9** (400 MHz, DMSO-d_6).



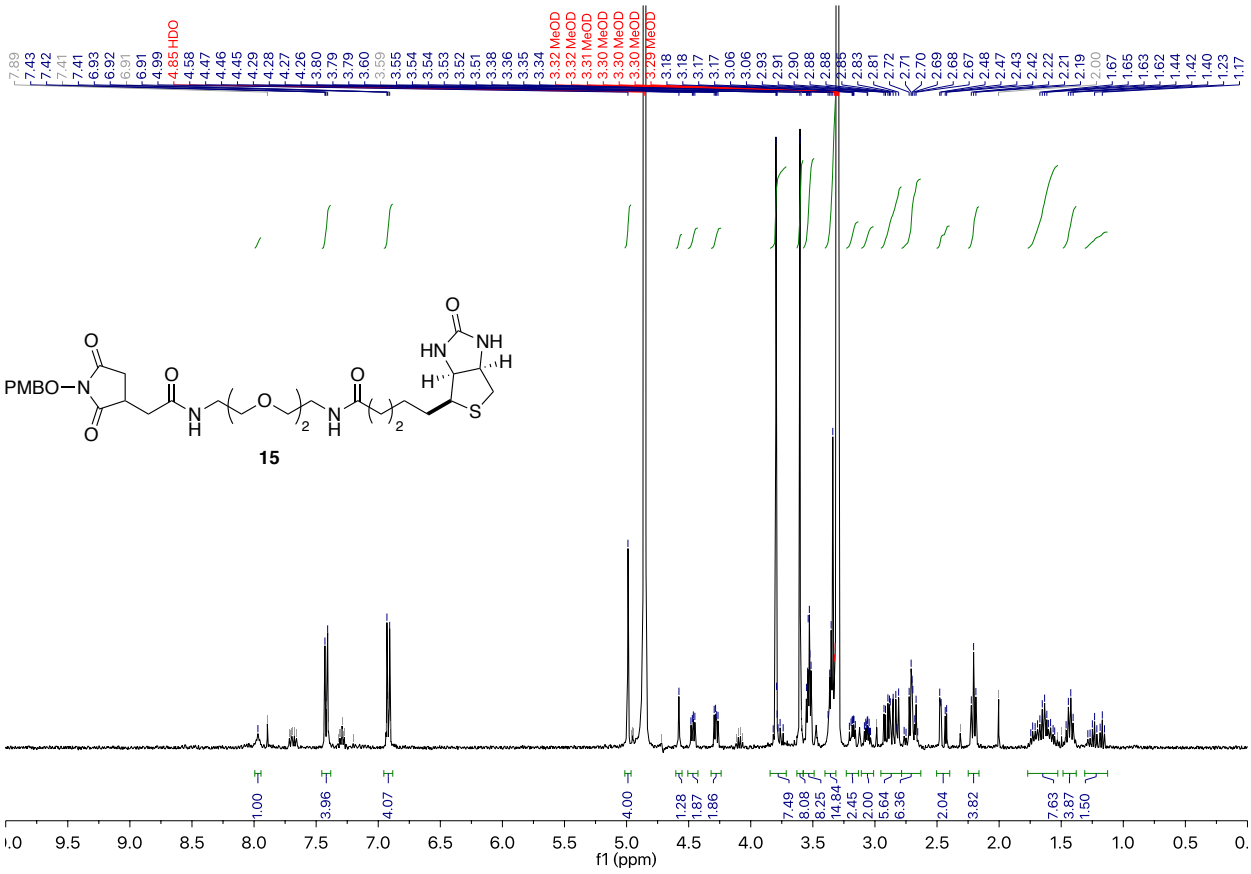
^{13}C NMR spectrum of **9** (100 MHz, DMSO-d_6).



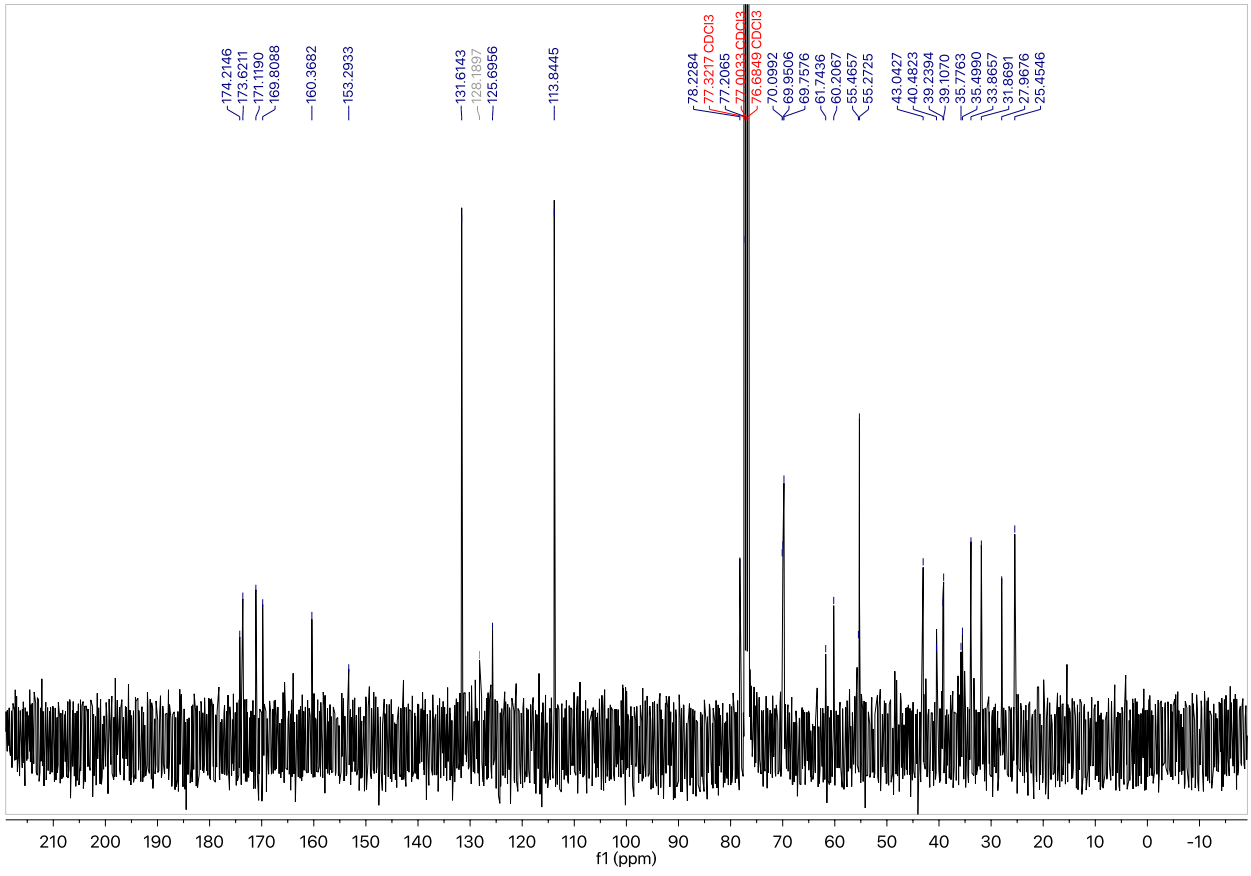
^{1}H NMR spectrum of **13** (400 MHz, CD_3OD).



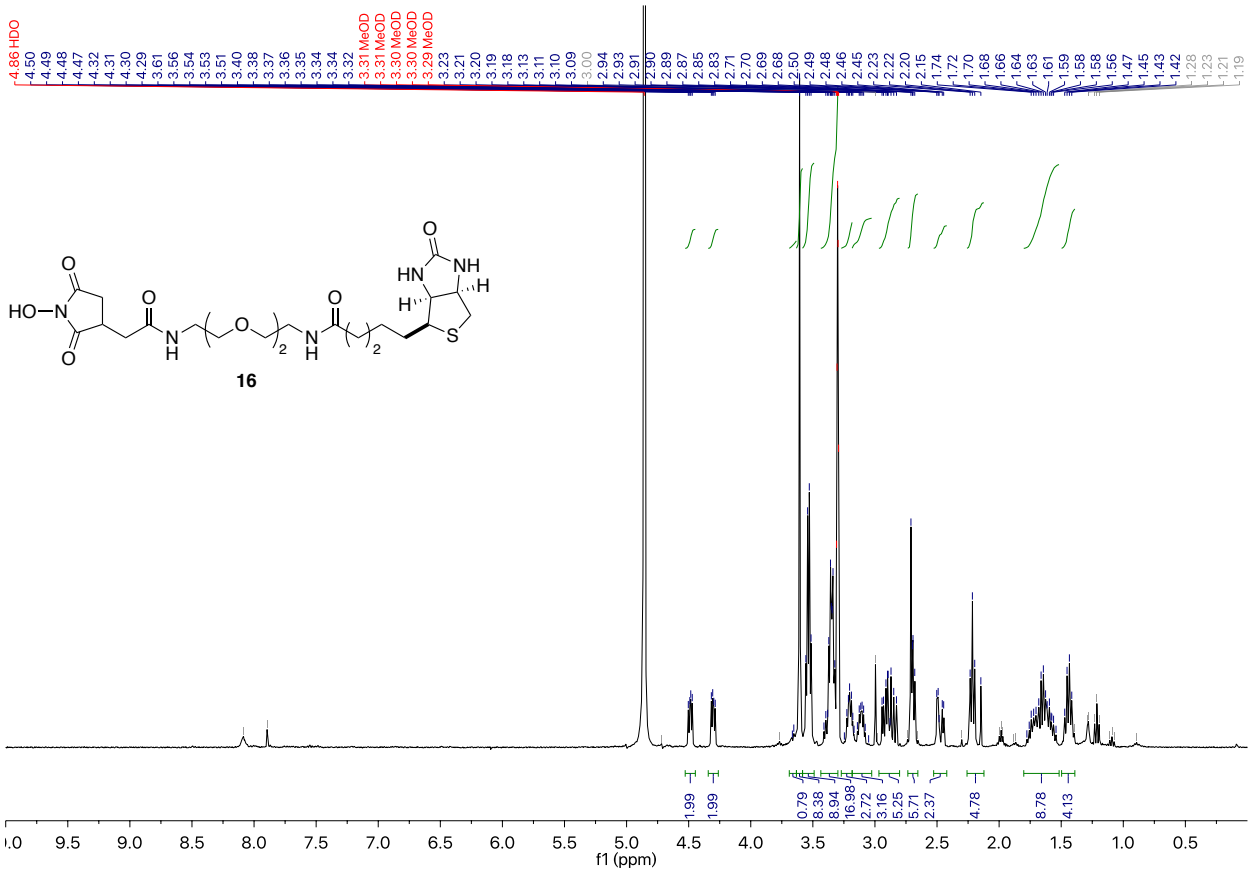
^{13}C NMR spectrum of **13** (100 MHz, CD_3OD).



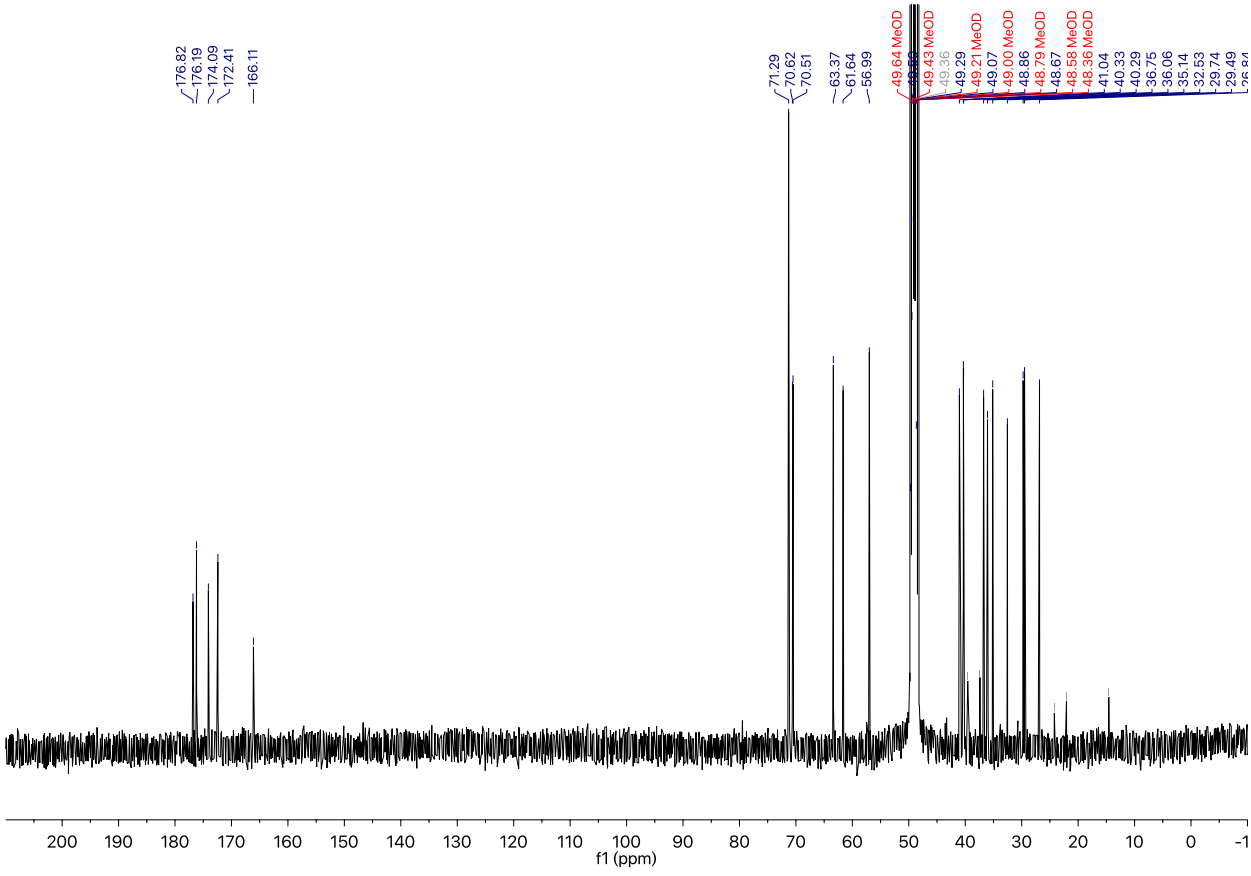
^{1}H NMR spectrum of **15** (400 MHz, CD_3OD).



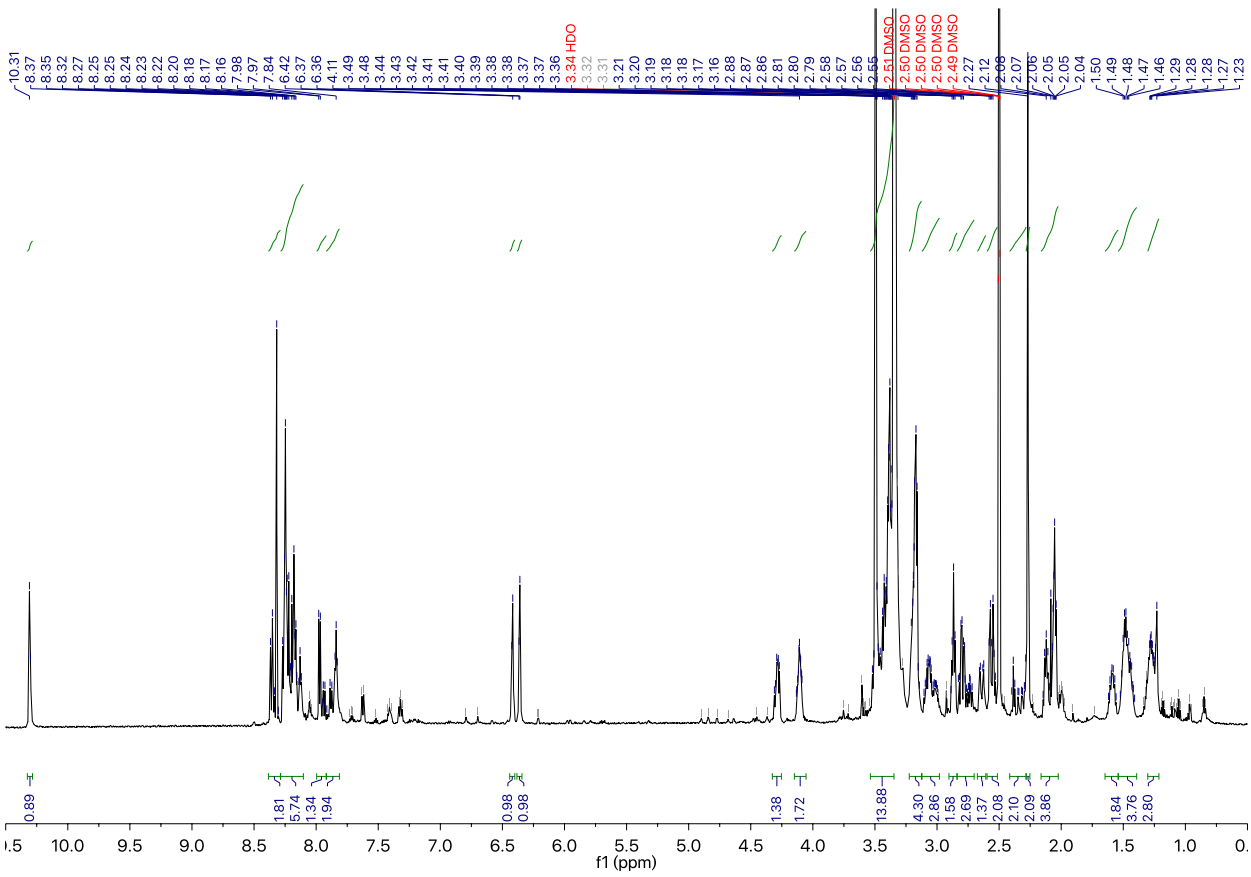
^{13}C NMR spectrum of **15** (100 MHz, $CDCl_3$).



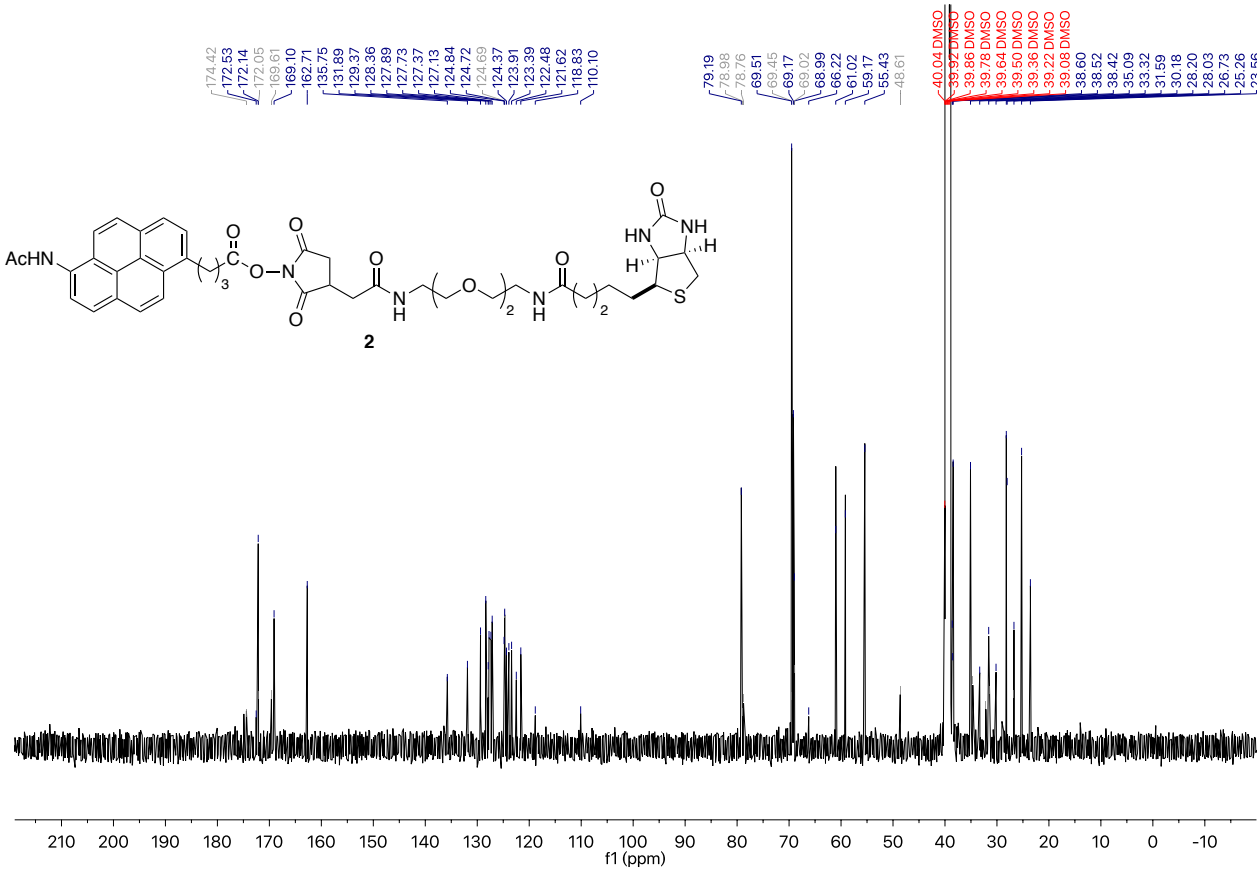
^{1}H NMR spectrum of **16** (400 MHz, CD_3OD).



^{13}C NMR spectrum of **16** (100 MHz, CD_3OD).



^{1}H NMR spectrum of **2** (600 MHz, DMSO-d_6).



^{13}C NMR spectrum of **2** (150 MHz, DMSO-d_6).