SUPPLEMENTARY INFORMATION

Deciphering the Protonation and Tautomeric Equilibria of Firefly Oxyluciferin by Molecular Engineering and Multivariate Curve Resolution

Mateusz Rebarz,[a,†] Boris-Marko Kukovec,[b,†] Oleg V. Maltsev,[c,†] Cyril Ruckebusch,[a] Lukas Hintermann,*[c] Panče Naumov,*[b] Michel Sliwa,*[a]

[a]Laboratoire de Spectrochimie Infrarouge et Raman (LASIR), CNRS UMR 8516/Université Lille Nord de France, Université Lille1 – Sciences et Technologies/Chemistry Department, bât C5/59655 Villeneuve d'Ascq Cedex, France. Email: michel.sliwa@univ-lille1.fr (M.S.).

[b]New York University Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates. Email: pance.naumov@nyu.edu (P.N.).

[c]Department Chemie, Technische Universität München, Lichtenbergstr. 4, 85748 Garching bei München, Germany. Email: lukas.hintermann@tum.de (L.H.).

Experimental

Synthesis:

General: DME (1,2-dimethoxyethane) was pre-dried by filtering through dried Al_2O_3 and stirring overnight with KOH + $FeSO_4 \cdot 7H_2O$. After filtration, it was distilled once, filtered through a column of dried Al_2O_3 and kept under argon over 4 Å molecular sieves. The water content (coulometric Karl Fischer titration) amounted to 42 ppm (*v*/*v*).

2-(6′-Hydroxybenzo[*d*]thiazol-2′-yl)-4-methoxythiazole (4-MeOxyLH) and **2-(6′-methoxybenzo[*d*]thiazol-2′-yl)-4-methoxythiazole (4,6′-DMeOxyL)**

A solution of diazomethane (0.2 mol/L; 9.5 mL, 1.9 mmol) in *t*BuOMe was added with a syringe pump (0.5 mL/h) to an argon bubbled suspension of **OxyLH$_2$** [1] (250 mg, 1.00 mmol) in DME (50 mL) and stirred overnight at r.t. The progress of the reaction was monitored by TLC (toluene/EtOAc). The reaction was quenched by addition of a few drops of acetic acid and SiO_2. After evaporate to dryness, the residue was placed on top of a chromatography column and eluted (SiO_2; toluene/EtOAc 10:1→5:1) to give 4-**MeOxyLH** (170 mg, 64%) and 4,6′-**DMeOxyL** (24 mg, 9%).

*Data for 4-**MeOxyLH***: ^{1}H NMR (500 MHz, DMSO-d_6, 27 °C): δ = 3.90 (s, 3 H, Me), 6.86 (s, 1 H, 5-H), 7.03 (dd, $^3J_{H,H}$ = 8.8 Hz, $^4J_{H,H}$ = 2.4 Hz; 1 H, 5'-H), 7.44 (d, $^4J_{H,H}$ = 2.4 Hz; 1 H, 7'-H), 7.88 (d, $^3J_{H,H}$ = 8.8 Hz; 1 H, 4'-H), 10.08 (s, 1 H, phenolic OH) ppm. ^{13}C NMR (91 MHz, DMSO-d_6, 25 °C): δ = 57.4 (4-O*C*H$_3$), 93.6 (C), 106.9 (C), 116.9 (C), 123.9 (C), 136.6 (C), 146.4 (C), 156.7 (C), 156.8 (C), 157.3 (C), 164.9 (C) ppm. HR-MS (EI): Calc. for $C_{11}H_8N_2O_2S_2$ 264.0022; Found 264.0014.

*Data for 4,6'-**DMeOxyL***: ^{1}H NMR (500 MHz, DMSO-d_6, 27 °C): δ = 3.86 (s, 3 H, CH$_3$), 3.91 (s, 3 H, CH$_3$), 6.89 (s, 1 H, 5-H), 7.17 (dd, $^3J_{H,H}$ = 9.0 Hz; $^4J_{H,H}$ = 2.6 Hz; 1 H, 5'-H), 7.74 (d, $^4J_{H,H}$ = 2.6 Hz; 1 H, 7'-H), 7.96 (d, $^3J_{H,H}$ = 9.0 Hz; 1 H, 4'-H) ppm. ^{13}C NMR (91 MHz, DMSO-d_6, 25 °C): δ = 55.8 (6'-O*C*H$_3$), 57.4 (4-O*C*H$_3$), 94.0 (C), 104.9 (C), 116.7 (C), 123.8 (C), 136.6 (C), 147.3 (C), 157.0 (C), 157.9 (C), 158.2 (C), 164.9 (C) ppm. HR-MS (EI): Calc. for $C_{12}H_{10}N_2O_2S_2$: 278.0178; found 278.0182.

2-(6′-Methoxybenzo[*d*]thiazol-2′-yl)-4-methoxythiazole (4,6′-DMeOxyL) by alkylation of 6′-**MeOxyLH:**

To an argon-bubbled suspension of 6′-**MeOxyLH** [1] (794 mg, 3.00 mmol) in DME (150 mL), a solution of diazomethane in *t*BuOMe (0.2 mol/L; 40.5 mL, 8.1 mmol) was added with a syringe pump (5.75 mL/h) and the reaction mixture was stirred overnight at r.t.. The progress of the reaction was monitored by TLC (toluene/EtOAc). The reaction was quenched by dropwise addition of acetic acid (0.5 mL) followed by SiO_2. After evaporation, the dry residue was placed on top of a chromatography column and eluted (SiO_2; toluene) to give 4,6′-**DMeOxyL** (580 mg, 69%).

Spectroscopic and titration experiments:

The UV/Vis spectra of compounds were recorded on a Shimadzu UV/Vis/NIR spectrophotometer UV-3600 using a series of phosphate buffers in the pH range 6–9 with ΔpH ≈ 0.1 for 6′-**MeOxyLH** and **HOxyLH** and in the pH range 6–11 for **OxyLH$_2$**, 4-**MeOxyLH** and 5,5-**DMeOxyLH**. Before each measurement, the weighed (7–9 mg) compound was dissolved in DMSO (5 mL) and an aliquot (2 μL) of this stock was pipetted into a quartz cuvette containing the buffer (2 mL) and shaked. To prevent any decomposition of the samples, the time interval between the dilution of the stock solution and the start of spectral acquisition did not exceed 30 s. The final concentration of the analyte was 4.80–7.00 μM (Table S1). Each spectrum was recorded in the 250–600 nm region against the same buffer. We could not detect any decomposition under these experimental conditions.

Data Analysis:

Self-modeling curve resolution

Among the mathematical and statistical methods to analyze spectroscopic mixture data, multivariate curve resolution describes a set of chemometrics tools for estimating pure component spectra and concentration profiles from data matrices of mixture spectra recorded from an evolving chemical system (any chemical system that change in systematic way as a function of e.g. time, temperature, pH). Self-modeling curve resolution does not require any assumptions except a bilinear model for the data.

In a bilinear model, a spectroscopic mixture represented by a matrix **D** containing *m* pH-dependent spectra (rows) registered at *n* wavelengths (columns) can be described according to Eq. (1):

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathbf{T}} + \mathbf{E} \quad (1)$$

where the matrices **C**(*m* x *N*) and $\mathbf{S}^{\mathbf{T}}$(*N* x *n*) contain the pH-dependent concentration profiles and the characteristic spectra of the *N* absorbing species in the mixture, respectively. The matrix **E**(*m* x *n*) contains the residual signal, which is mostly due to experimental noise.

Multivariate curve resolution – alternating least squares (MCR-ALS)

MCR-ALS [2,3] is one of the most widely used soft-modeling methods to decompose two-way data according to Eq. (1). Only a brief description of the main steps of the method for the application to spectrophotometric data is given in the following. More comprehensive information can be found in the literature where the potential of the method in equilibrium studies has been reported. [4] The number *N* of components required to describe the maximum of the variance in **D** is first estimated, *e.g.* from singular value decomposition [5].

The iterative alternating least squares (ALS) optimization then starts from initial estimates of **C** or $\mathbf{S}^{\mathbf{T}}$ and, in each cycle, the matrices **C** and $\mathbf{S}^{\mathbf{T}}$ are calculated under constraints.

Constraints are applied to enforce some generic knowledge about the concentration and spectra profiles during the ALS optimization, such as non-negativity of UV absorptivities, non-negativity and unimodality (one unimodal peak) of pH-dependent concentration profiles and closure (the sum of the concentrations of the detected component at any different pH is equal to the total concentration).

The quality of the decomposition is assessed from the lack of fit (lof, in %) between the experimental data matrix **D** and the data reproduced data matrix from the product $\mathbf{CS}^{\mathbf{T}}$, defined in Eq. (2):

$$lof(\%) = 100 \times \sqrt{\frac{\sum \left(d_{ij}^{*} - d_{ij}\right)^2}{\sum d_{ij}^2}} \qquad (2)$$

where d_{ij} is one element of the experimental matrix **D** and d_{ij}^{*} is the analogous element of the reproduced data matrix by the MCR-ALS model.

The bilinear model be formulated as in Eq. (1) can be applied to simultaneous analysis of multiple data sets in which information from one system monitored in different conditions is merged. In a multi-set approach, the single data matrices can be appended column-wise, one under the other. In this configuration, where the spectral matrix $\mathbf{S}^{\mathbf{T}}$ contains all the pure spectra of the different components present in the individual data matrices and the concentration matrix **C** is an augmented data matrix describing the pH-dependent concentration profiles in each of the q individual dataset, denoted [$\mathbf{C_1}$; $\mathbf{C_2}$; …; $\mathbf{C_q}$]. Note that a multi-set configuration does not imply either that all compounds are present in all experiments or that all experiments should share the same kinetic behavior.

The advantages of multi-set analysis are manifold; a better solution can be obtained regarding the extraction of profiles in **C** and $\mathbf{S}^{\mathbf{T}}$ because of the complementary information in each of the matrices that helps to model experiments with too overlapped or indistinguishable (rank-deficient) information. Constraints can also be tailored to set the correspondence among the different components in the different matrices forming the multi-set structure. This is a key point to ensure that the conditions are met for resolution of complex systems. By using the information of presence/absence of compounds in each matrix, a multi-set MCR strategy provides more reliable, meaningful and more robust solutions (less prone to ambiguities) compared to single dataset analysis. An increasing number of constraints and the simultaneous analysis of multiple data sets will reduce the range of feasible solutions.

One of the theoretical issues that have to be considered when using soft-modeling methods to resolve bilinear spectroscopic mixtures is the presence of rotational ambiguity in the results. This translates into the fact that the calculated profiles are not unique, i.e. a band of feasible profiles is fitting the data actually well for a defined set of mathematical and chemical constraints. The problem of calculating this range of feasible solutions is a very

challenging one. This problem was addressed and solved for a two component system in the seminal paper of Lawton and Sylvestre [6]. However, the method cannot be generalized to more than two components and since then the determination and visualization of rotational ambiguity has been a matter of extensive research. Several methods were published for finding feasible bands, including numerical [7,8], statistical [9] and geometric [10,11] ones. A generally applicable method for the exhaustive description of three component systems has been published very recently [12].

References:

[1] N. Suzuki, T. Goto. Arg. Biol. Chem. 36 (1972) 2213.
[2] R. Tauler, Chemometrics and Intelligent Laboratory Systems 30 (1995) 133.
[3] A. de Juan, R. Tauler, Critical Reviews in Analytical Chemistry 36 (2006) 163.
[4] H. Abdollahi, V. Mahdavi,Langmuir 23 (2007) 2362.
[5] G.H. Golub, C.F.V. Loan, Matrix Computation, The John Hopkins University Press, Baltimore, USA, 1996.
[6] W.H. Lawton, E.A. Sylvestre, Technometrics, 13 (1971) 617.
[7] P.J. Gemperline, Analytical Chemistry, 71 (1999) 5398.
[8] R. Tauler, Journal of Chemometrics, 15 (2001) 627.
[9] M.N. Leger, P.D. Wentzell, Chemometrics and Intelligent Laboratory Systems, 62 (2002) 171.
[10] R. Rajko, K. Istvan, Journal of Chemometrics, 19 (2005) 448.
[11] S. Miron, M. Dossot, C. Carteret, S. Margueron, D. Brie, Chemometrics and Intelligent Laboratory Systems, 105 (2011) 171.
[12] A. Golshan, H. Abdollahi, M. Maeder, Analytical Chemistry, 83 (2011) 836.

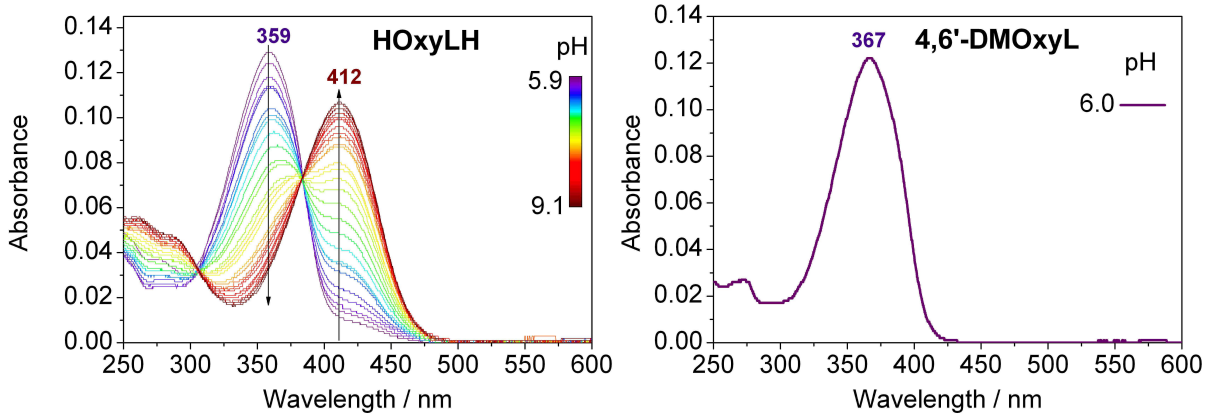


Fig. S1 The UV/Vis spectra of **HOxyLH** and 4,6′-**DMeOxyL** in phosphate buffer. The concentrations are given in Table S1.

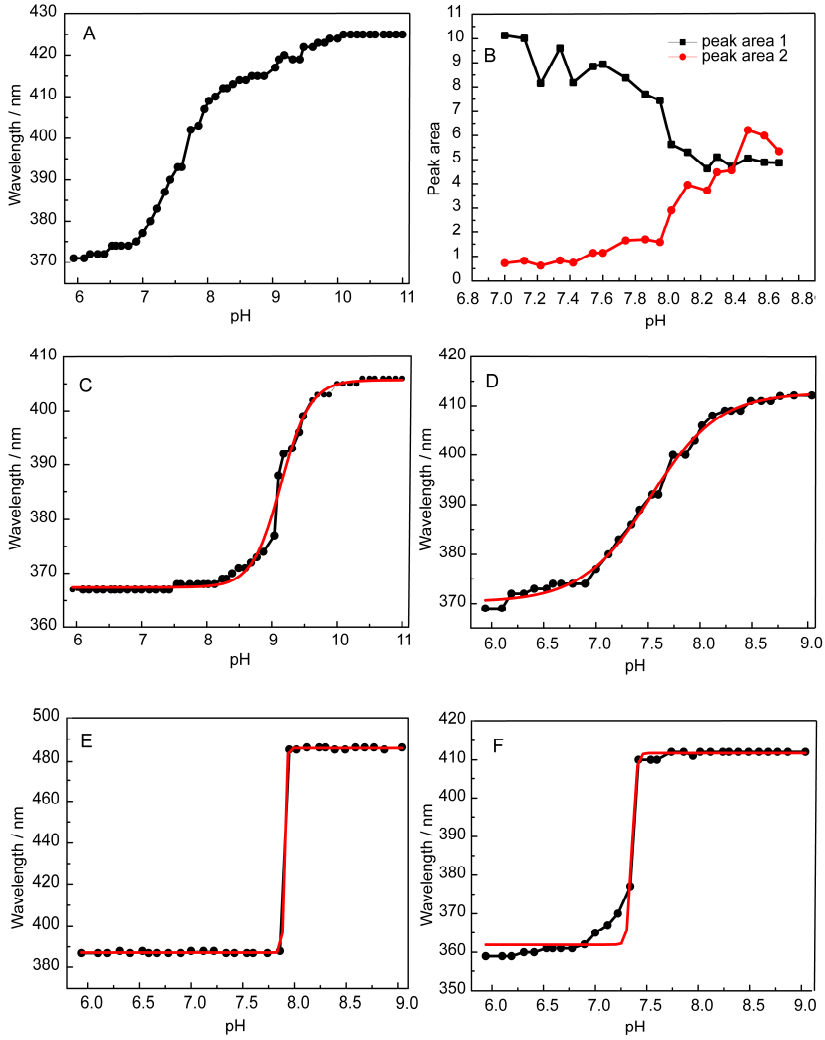


Fig. S2 The dependence of λ^{abs}_{max} on pH for **OxyLH$_2$** (A), 4-**MeOxyLH** (C), 6′-**MeOxyLH** (D), 5,5-**DMeOxyLH** (E) and **HOxyLH** (F), shown in the pH range 6–11 for **OxyLH$_2$** and 4-**MeOxyLH** and in the pH range 6–9 for others. The peak area fitted plots for **OxyLH$_2$** are also shown (B).

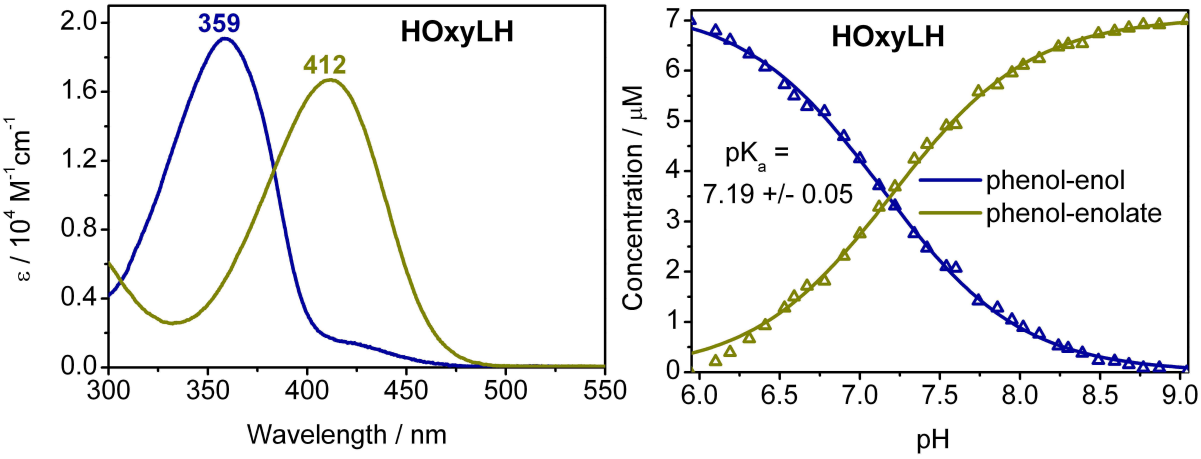


Fig. S3 Pure absorption spectra (left) and corresponding concentration profiles (right) of species obtained by MCR-ALS for **HOxyLH**.

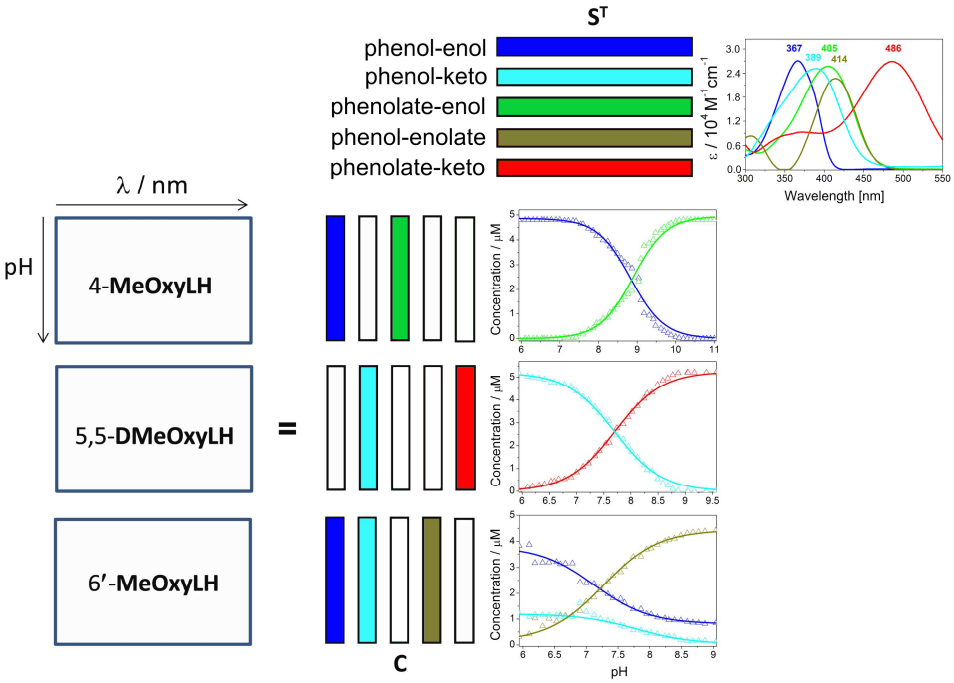


Fig S4 Multi-set MCR-ALS analysis of 6′-**MeOxyLH$_2$**, 4-**MeOxyLH** and 5,5-**DMeOxyLH**

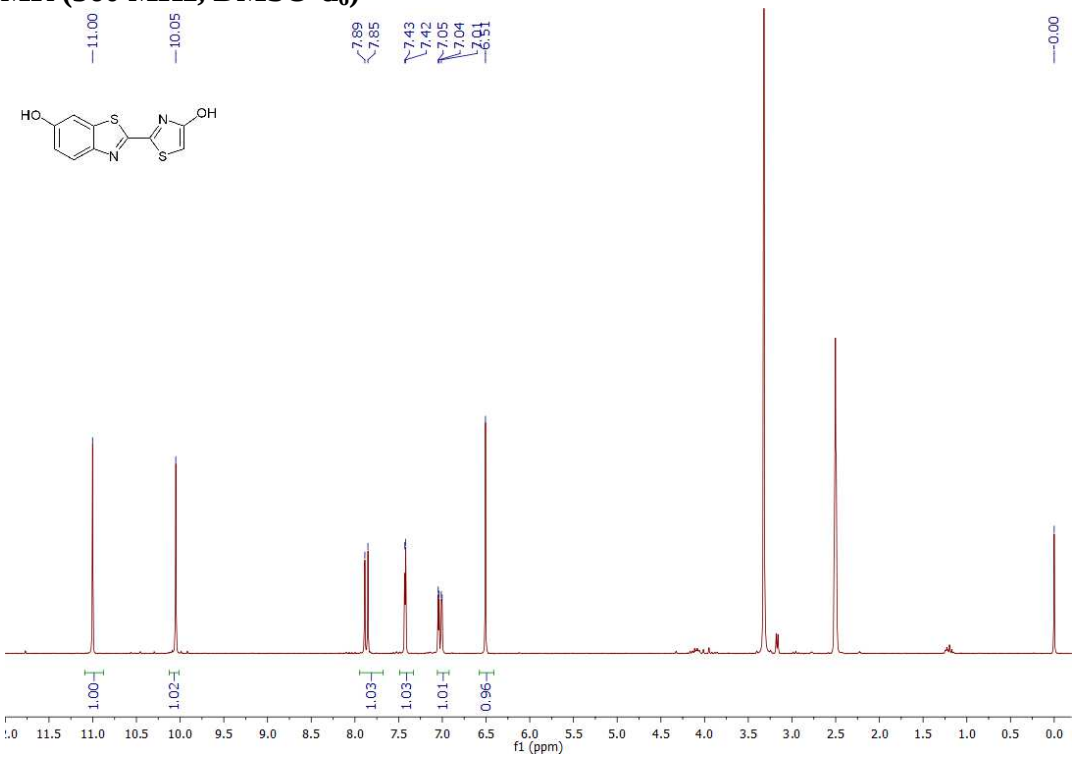
Table S1. Final concentration of the compounds in the buffer solution used for measurement of the UV/Vis spectra.

Compound	*c* / µM
OxyLH$_2$	6.11
HOxyLH	7.00
4-**MeOxyLH**	4.80
6'-**MeOxyLH**	5.37
5,5-**DMeOxyLH**	5.22
4,6-**DMeOxyL**	6.61

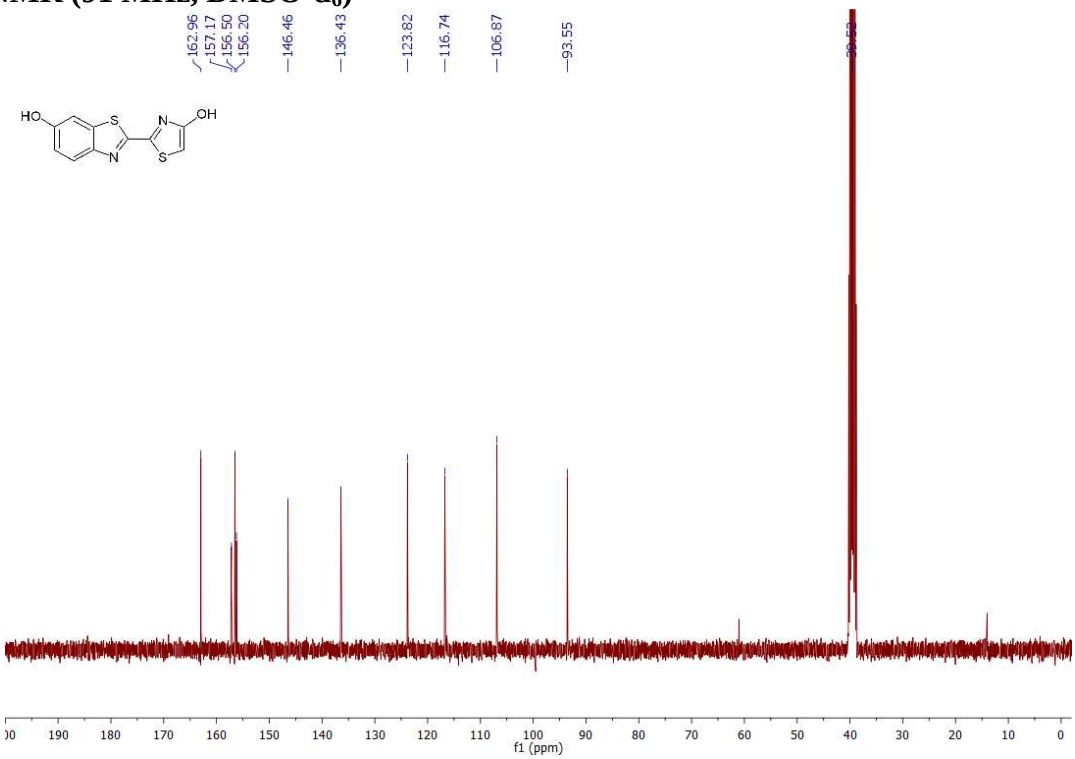
NMR Spectra:

2-(6′-Hydroxybenzo[*d*]thiazol-2′-yl)-4-hydroxythiazole ($OxyLH_2$)

^{1}H NMR (360 MHz, DMSO-d_6)

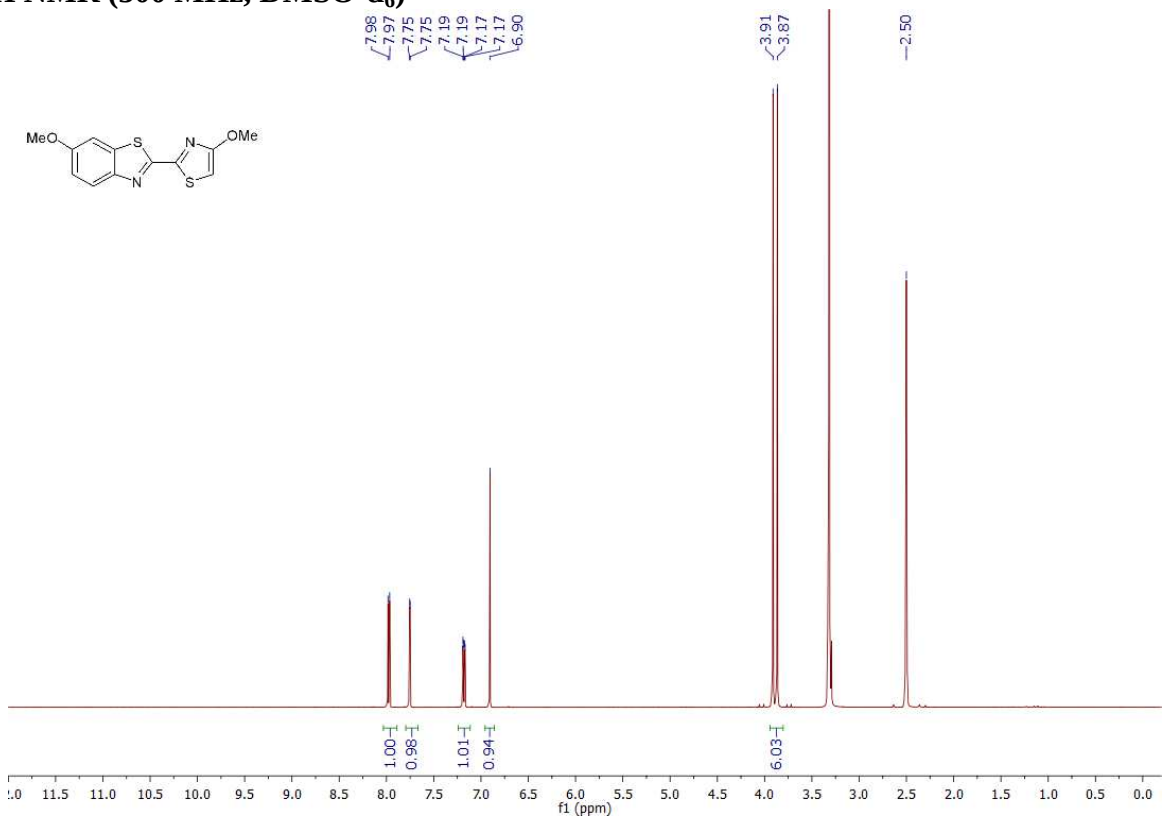


^{13}C NMR (91 MHz, DMSO-d_6)

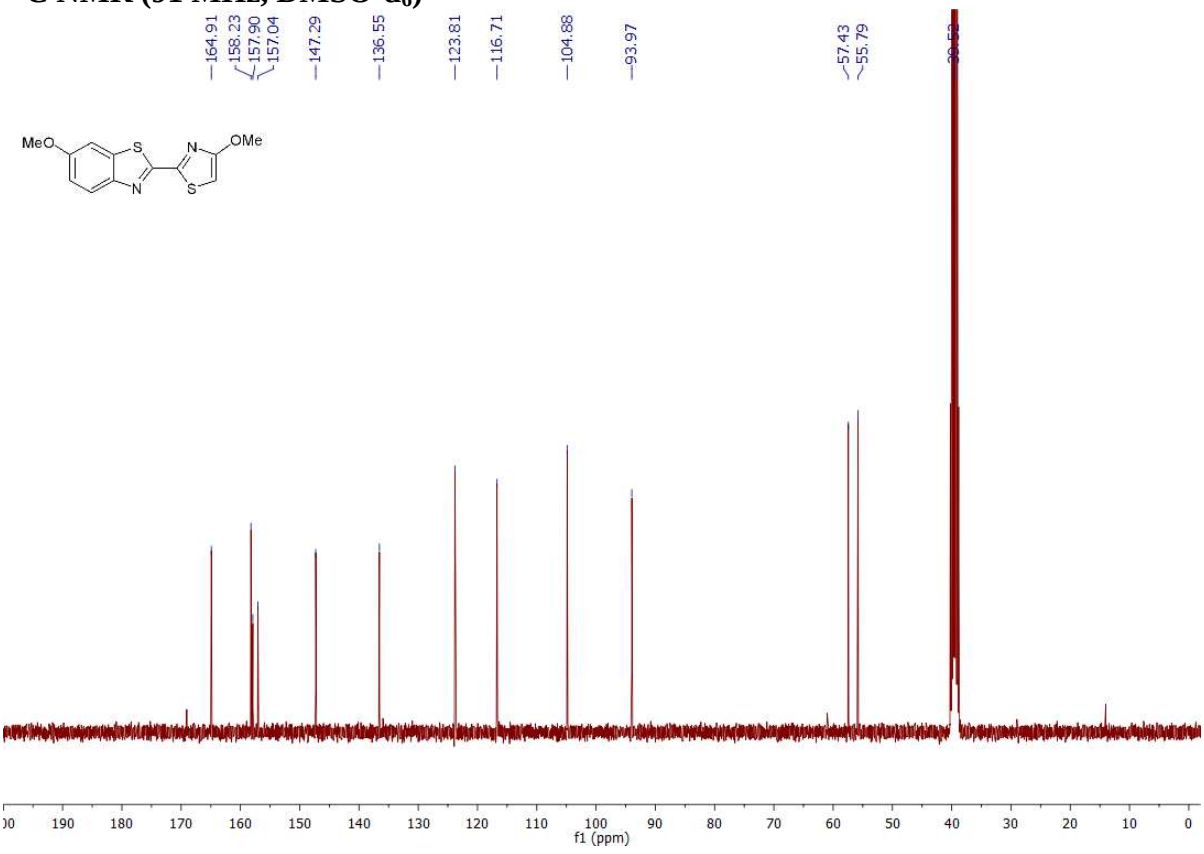


2-(6′-Methoxybenzo[*d*]thiazol-2′-yl)-4-methoxythiazole (4,6′-DMeOxyL)

^{1}H NMR (500 MHz, DMSO-d$_6$)

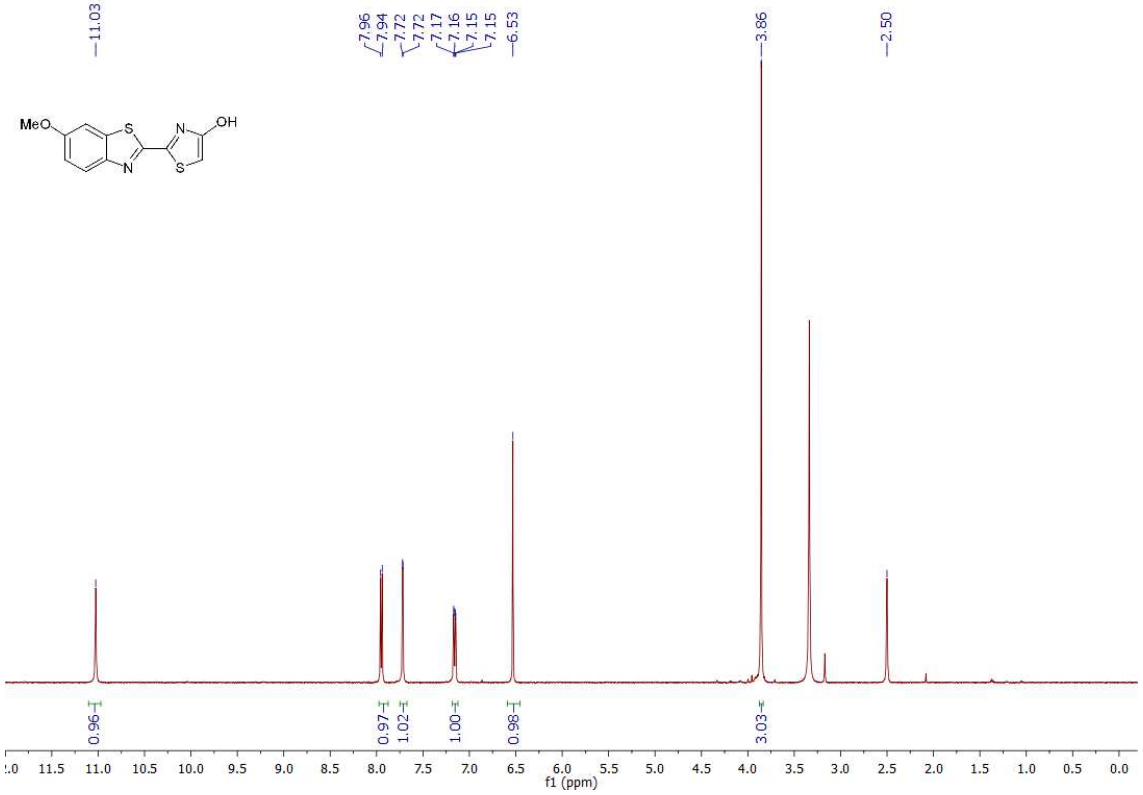


^{13}C NMR (91 MHz, DMSO-d$_6$)

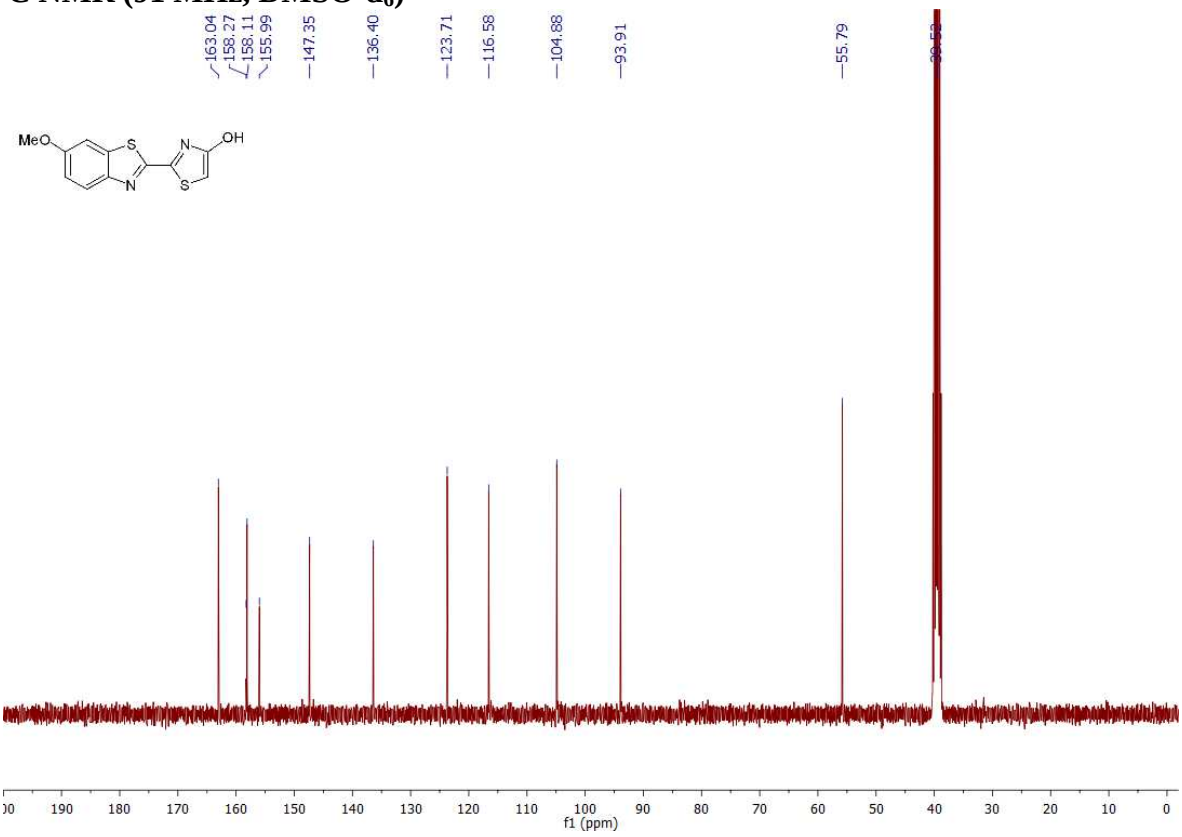


2-(6′-Methoxybenzo[*d*]thiazol-2′-yl)-4-hydroxythiazole (6′-MeOxyLH)

^{1}H NMR (500 MHz, DMSO-d$_6$)

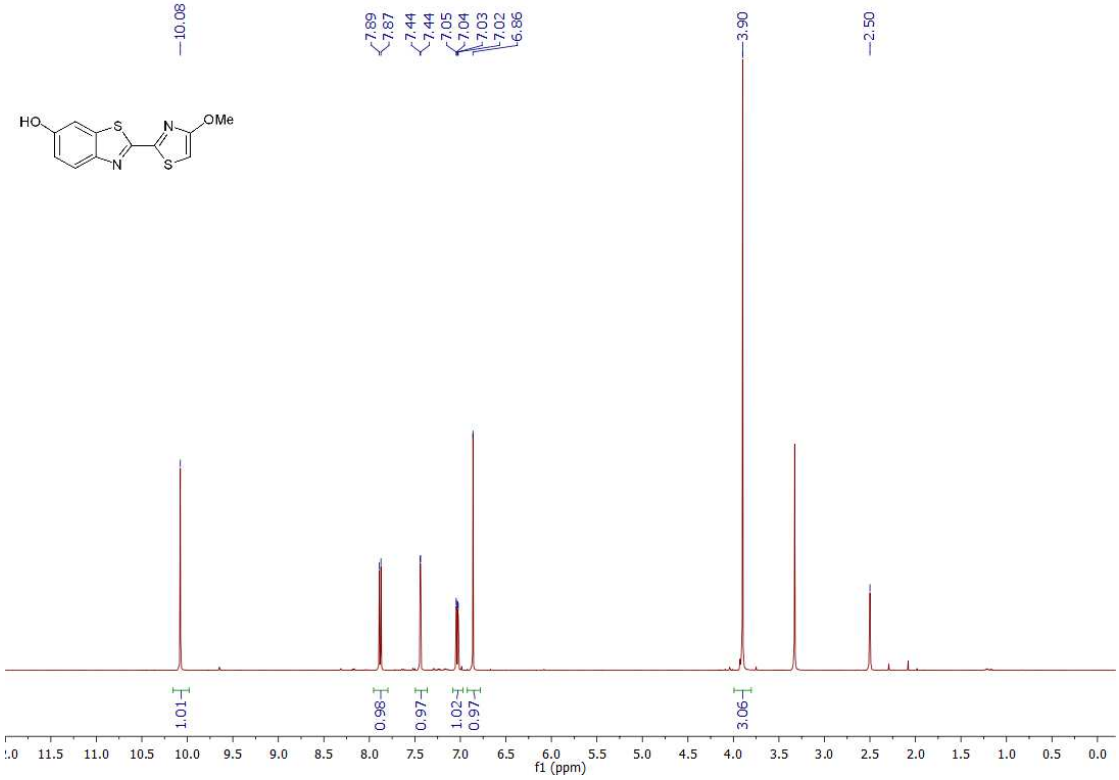


^{13}C NMR (91 MHz, DMSO-d$_6$)

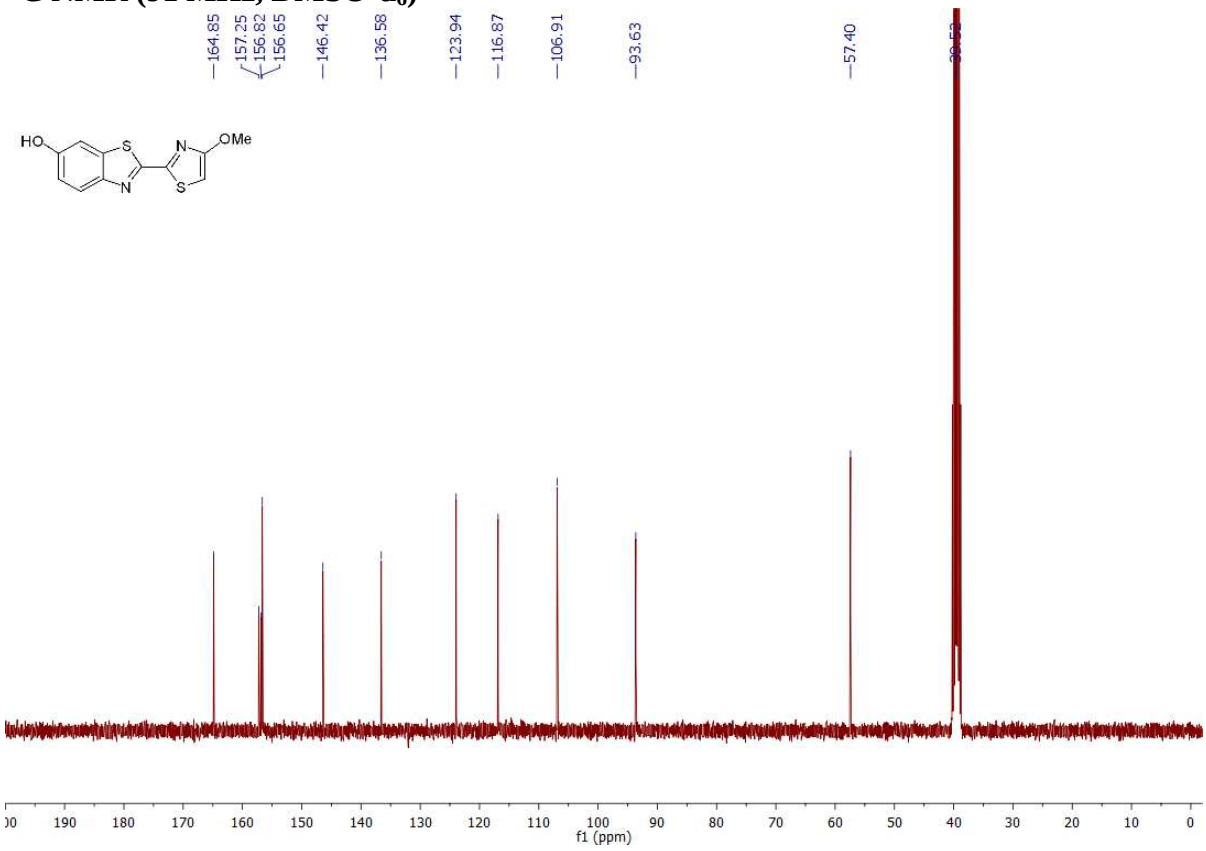


2-(6′-Hydroxybenzo[*d*]thiazol-2′-yl)-4-methoxythiazole (4-MeOxyLH)

^{1}H NMR (500 MHz, DMSO-d$_6$)

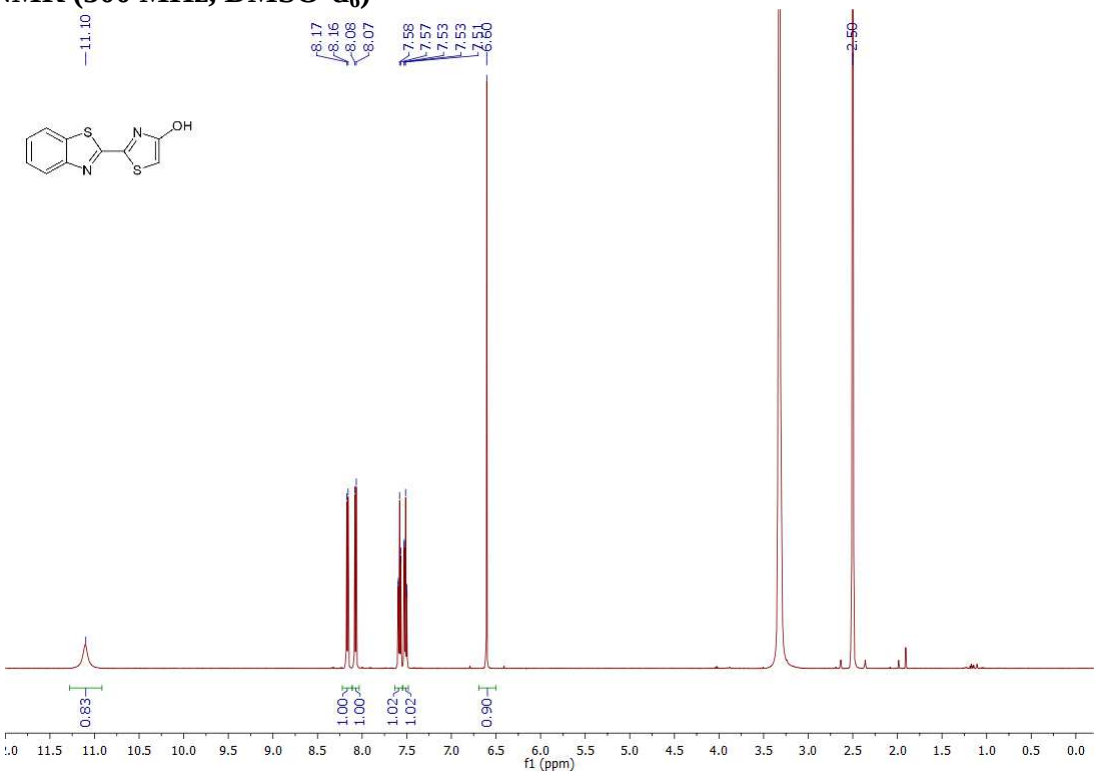


^{13}C NMR (91 MHz, DMSO-d$_6$)

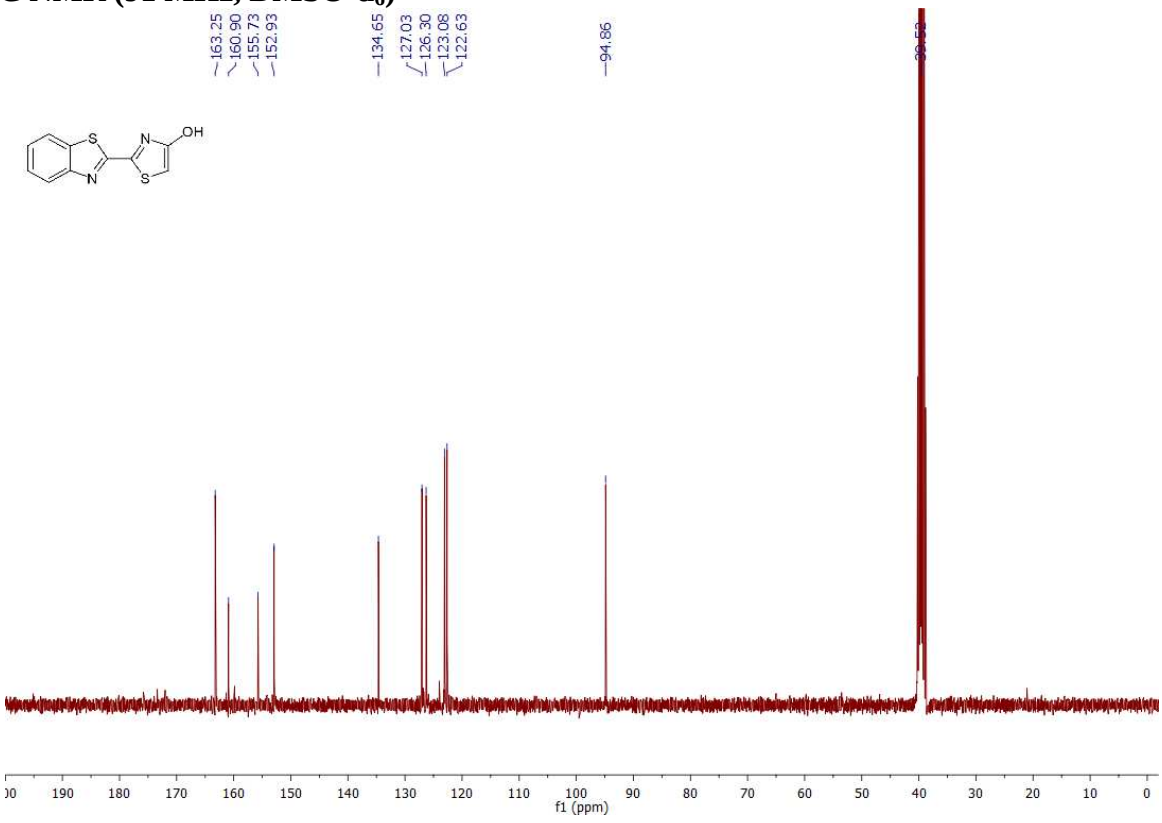


2-(Benzo[*d*]thiazol-2′-yl)-4-hydroxythiazole (HOxyLH)

^{1}H NMR (500 MHz, DMSO-d$_6$)

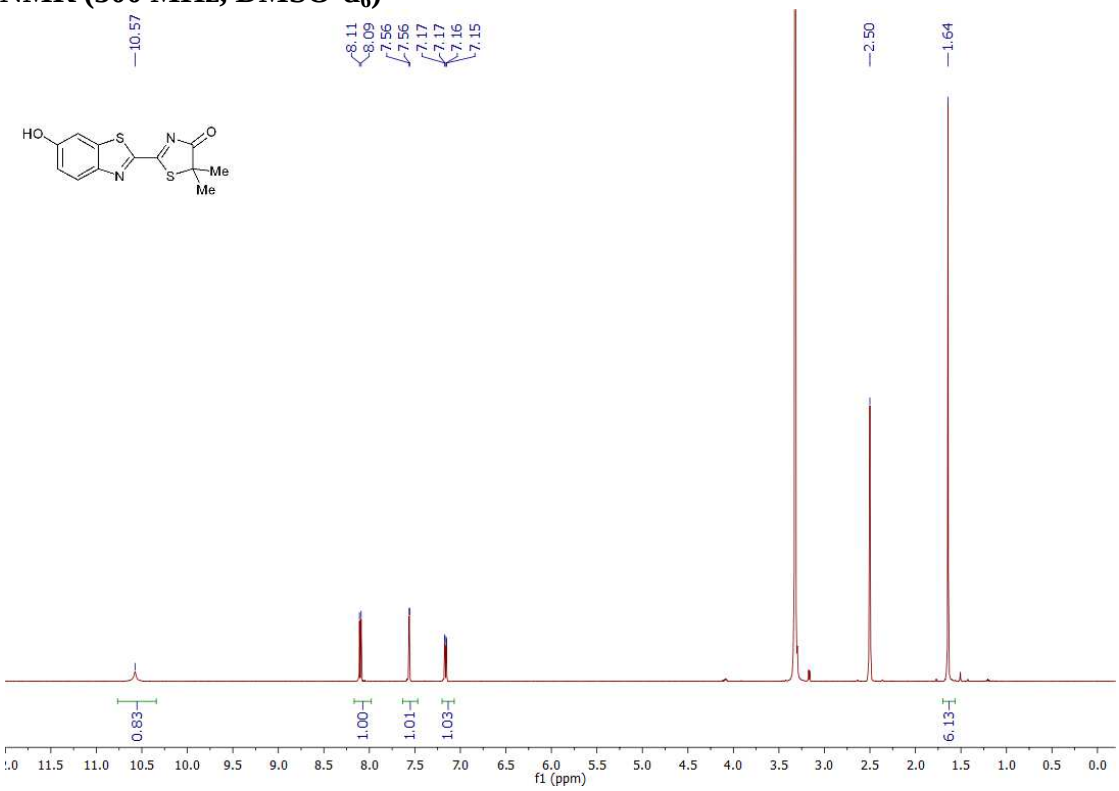


^{13}C NMR (91 MHz, DMSO-d$_6$)



2-(6′-Hydroxybenzo[*d*]thiazol-2′-yl)-5,5-dimethylthiazol-4(5*H*)-one (5,5-DMeOxyLH)

^{1}H NMR (500 MHz, DMSO-d$_6$)



^{13}C NMR (91 MHz, DMSO-d$_6$)