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

Supporting Information: In silico design of fluorescent biomarkers

Andrea Echeverri*[1,2], Candice Botuha[3], Tatiana Gómez[4], Eleonora Luppi[5], Julia Contreras-Garcia[5], and Carlos Cárdenas[1,6]

[1]Departamento de Física, Facultad de Ciencias, Universidad de Chile, Las Palmeras 3425, Casilla 635, Santiagio, Chile

[2]Laboratoire de Chimie Théorique, Sorbonne Université, 4 Pl Jussieu, 75005, Paris, Francia.

[3]IPCM, Sorbonne Université and CNRS, 4 Pl Jussieu, 75005, Paris, Francia.

[4]Theoretical and Computational Chemistry Center, Institute of Applied Sciences, Faculty of Engineering, Universidad Autonoma de Chile, El Llano Subercaceaux 2801, Santiago, Chile

[5]Laboratoire de Chimie Théorique, Sorbonne Université and CNRS, 4 Pl Jussieu, 75005, Paris, Francia.

[6]Centro para el Desarrollo de la Nanociencia y la Nanotecnología (CEDENNA), Santiago, Chile

*Email: andrea.echeverri_cardenas@sorbonne-universite.fr

A Supporting Information: In silico design of ESIPT - based fluorescent probes for the biomarkers detection

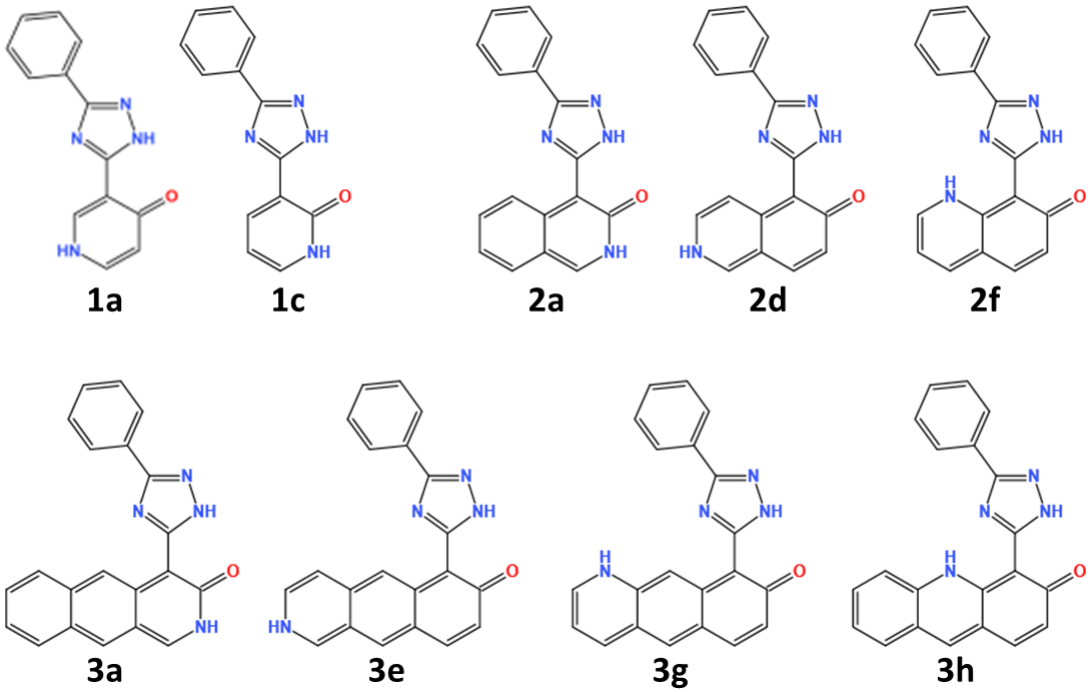

Figure SI 1: All possible keto$_2$ forms in the sets studied in this article. In this case, the hydrogen is on the nitrogen atom in pyridine (**1a** and **1c**), (iso)quinoline (**2a**, **2d** and **2f**) or benzoquinoline (**3a**, **3e**, **3g** and **3h**).

Molecule	S_0			S_1
	E (keto-enol)	E ($keto_2$-enol)	%enol	E (keto-enol)
1	0.43	N/A	100	-0.25
1a	0.36	-0.04	18	-0.26
1b	0.42	N/A	100	-0.30
1c	0.32	0.08	94	-0.18
1d	0.42	N/A	100	-0.26
2	0.32	N/A	100	-0.03
2a	0.26	-0.17	0.2	-0.13
2b	0.30	N/A	100	-0.11
2c	0.31	N/A	100	-0.06
2d	0.27	-0.04	18	-0.02
2e	0.30	N/A	100	-0.07
2f	0.25	0.07	91	-0.01
3	0.20	N/A	100	0.03
3a	0.26	-0.10	2	0.13
3b	0.24	N/A	100	0.26
3c	0.25	N/A	100	-0.07
3d	0.25	N/A	100	0.02
3e	0.24	0.12	98	0.06
3f	0.24	N/A	100	0.01
3g	0.25	0.16	99.8	0.06
3h	0.45	-0.09	94	-0.02

Table SI 1: Relative energies in eV to enol form corrected by zero point in S_0 for all molecules studied in this article. Keto is referred to the isomer that results after TPI take place; and $keto_2$ the hydrogen is bond to the nitrogen that belongs to the pyridine ring or conjugated rings. We use N/A in molecules where $keto_2$ geometry doesn't exist.

Molecule	Coefficient		
	HONTO, LUNTO	HONTO-1,LUNTO+1	others
1	0.9201	0.0684	<0.0097
$\mathbf{1b}_{e}nol$	09440	0.0477	< 0.0082
$\mathbf{1b}_{k}eto$	0.9866	0.0100	<0.0030
$\mathbf{1c}_{e}nol$	0.8975	0.0632	<0.0295
$\mathbf{1c}_{k}eto$	0.9764	0.0145	<0.0053
$\mathbf{3c}_{e}nol$	0.9705	0.0247	<0.0073

Table SI 2: Natural transition orbital[43] coefficients for the following molecules have been determined: **1** and **3c** in their enol forms, and **1b** and **1c** for both isomers, enol and keto. These calculations were done using the ω-b97xd functional and a 6-311++g(d,p) basis set, with the PCM approximation applied to simulate the solvent effects

In all instances, it is noteworthy that the contributions of "HONTO" and "LUNTO" consistently exceed 0.9. This observation confirms that it suffices to focus our analysis primarily on HONTO and LUNTO because they have the largest contribution.

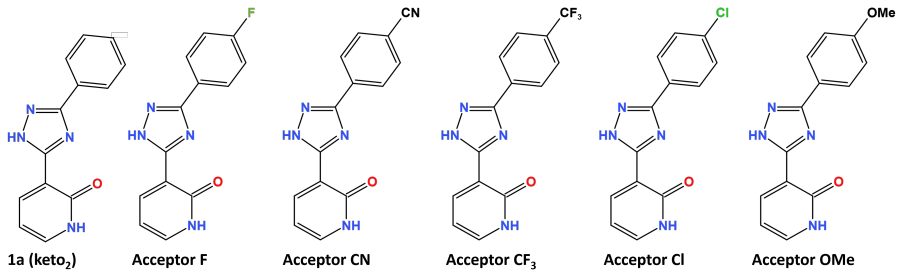

Figure SI 2: Molecules used for validating the computational methodology. These types of molecules have a design of charge transfer processes. The pyridine group serves as the donor fragment, the triazole group functions as the bridge fragment, and there are different acceptor groups.

Molecule	$\lambda_{emission}$ (EtOH)		$\lambda_{emission}$ (DCM)		$\lambda_{emission}$ (MeCN)	
	Exp.	Calc.	Exp.	Calc.	Exp.	Calc.
1a keto$_2$	392	346	392	343	392	342
Acceptor F	393	343	393	346	393	342
Acceptor CN	390	342	390	345	390	342
Acceptor CF$_3$	388	342	388	345	388	341
Acceptor Cl	390	343	390	346	390	342
Acceptor OMe	394	345	395	349	395	344

Table SI 3: Experimental emission wavelengths (measured by the Institut Parisien de Chimie Moléculaire, Sorbonne University, Paris) and the corresponding calculated values (in nm) for the molecules depicted in Figure SI 2 are provided. The calculations were done using the ω-b97xd functional along with a 6-311++g(d,p) basis set and applying the PCM approximation to simulate the effects of ethanol (ETOH), dichloromethane (DCM), and acetonitrile (MeCN) solvents.

A.1 Set 1

All solution spectroscopic measurements were performed in spectroscopic grade solvents (Carlo Erba Reagents) at R.T in a 1 cm path length quartz cuvette. A double-beam UV-Vis CARY 50 (Agilent Technologies) spectrophotometer was used to collect the absorption spectra and measurements were compared to a blank solution (solvent only). Steady-state emission spectra in solution were recorded on a Varian Cary Eclipse fluorescence spectrometer (Agilent) The signal was collected at 90° with respect to the excitation beam (right angle mode) for solution studies.

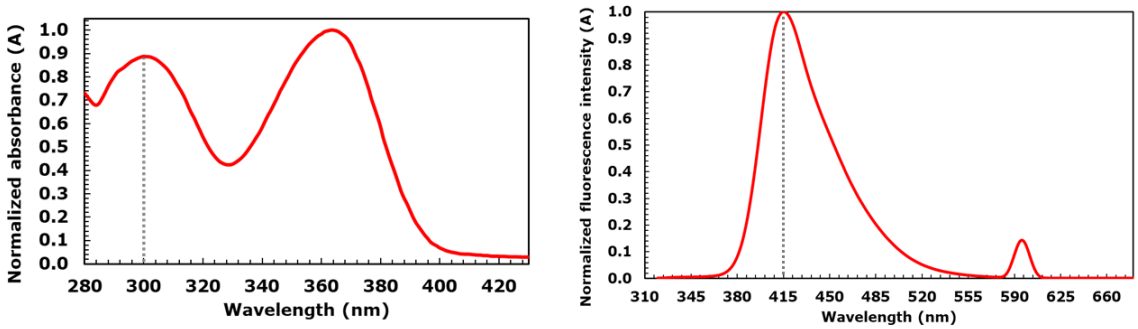

Figure SI 3: Experimental UV-Vis absorption (left) and fluorescence emission (right) spectra in an aqueous solvent with 1% dimethylsulfoxide (DMSO) for molecule **1b** at a concentration of $5 * 10^{-6} M$. The vertical dashed lines indicate the position of the absorption peak (left) for the enol isomer and the fluorescence emission peak (right) from the keto isomer, with a 300 nm excitation wavelength.

Molecule	$\lambda_{absorbtion}$	$\lambda^{enol}_{emission}$	$\lambda^{keto}_{emission}$	Stokes^{enol}	Stokes^{keto}	$f^{keto}_{emission}$	$f_{absorption}$
1	267	305	363	39	97	0.28	0.25
1a	264	306	351	43	87	0.38	0.40
1b	273	315	381	42	108	0.26	0.21
1c	248	303	328	55	80	0.31	0.33
1d	273	316	374	43	101	0.37	0.40

Table SI 4: Calculated wavelengths, λ, for absorption and emission (keto and enol form) in nm, Stokes shifts from enol and keto isomer emission (nm), and oscillator strength for the process absorption (enol) and emission (keto) for the molecules of set 1.

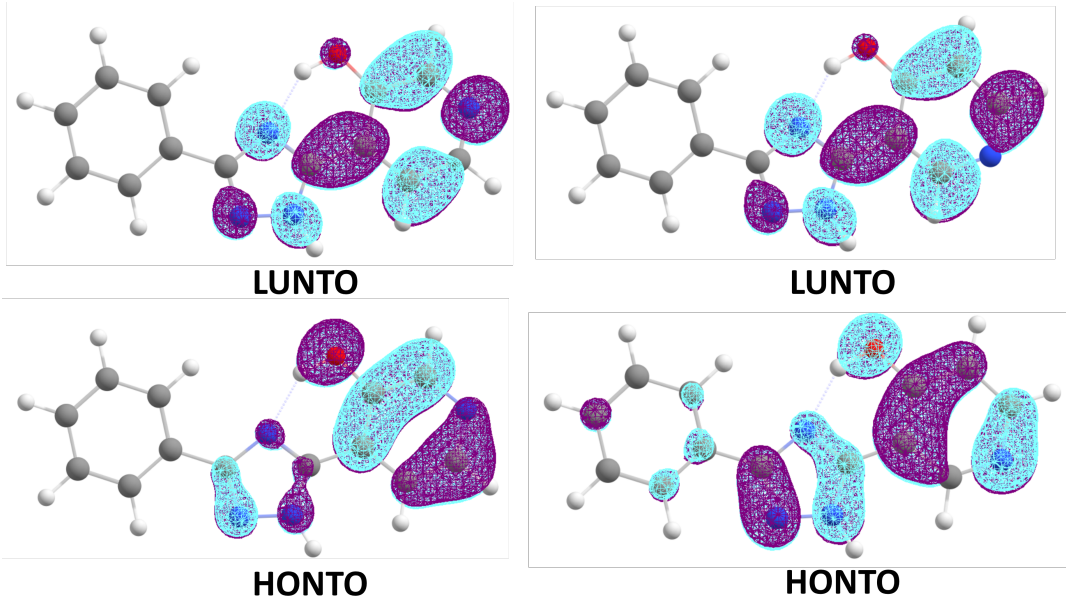

Figure SI 4: NTO for **1b** (left) and **1c** (right) in the enol isomer. To model the aqueous solvent, we employed the PCM approximation. The system's computations were performed using ω-b97xd/6-311++g(d,p) method.

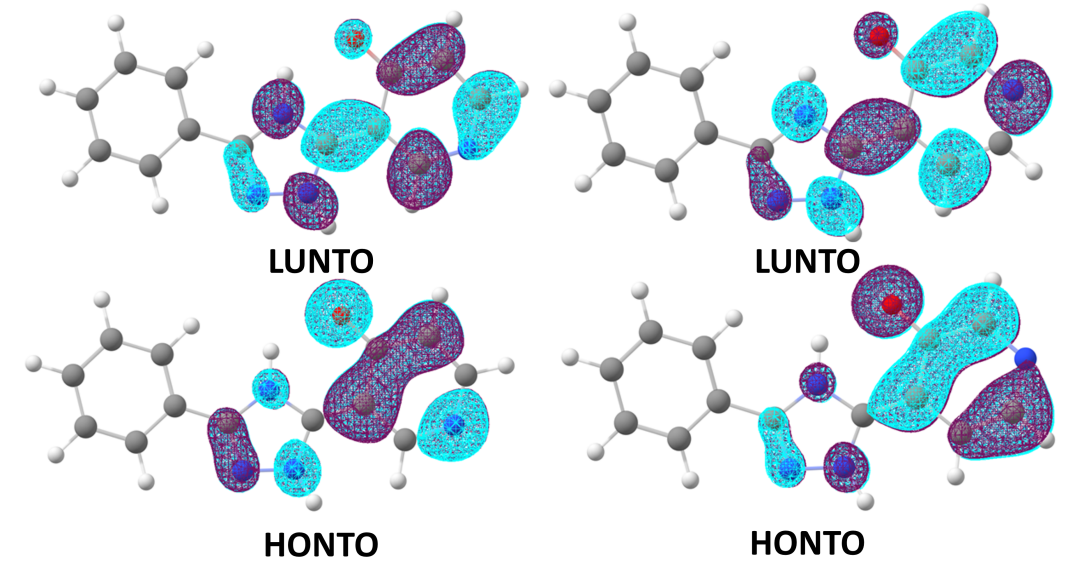

Figure SI 5: NTO for **1b** (left) and **1c** (right) in the keto isomer. To model the aqueous solvent, we employed the PCM approximation. The system's computations were performed using ω-b97xd/6-311++g(d,p) method.

A.2 Set 2

Molecule	$\lambda_{absorbtion}$	$\lambda^{enol}_{emission}$	$\lambda^{keto}_{emission}$	Stokes^{enol}	Stokes^{keto}	$f^{keto}_{emission}$	$f_{absorption}$
2	299	362	393	63	94	0.24	0.22
2a	314	382	429	68	115	0.23	0.25
2b	308	371	415	63	107	0.22	0.19
2c	298	375	407	77	109	0.20	0.19
2d	286	301	372	75	86	0.26	0.24
2e	310	376	417	66	107	0.24	0.24
2f	292	363	375	71	82	0.21	0.21

Table SI 5: Calculated wavelengths, λ, for absorption and emission (keto and enol form) in nm, Stokes shifts from enol and keto isomer emission (nm), and oscillator strength for emission from keto tautomer and process absorption for the molecules of set 2.

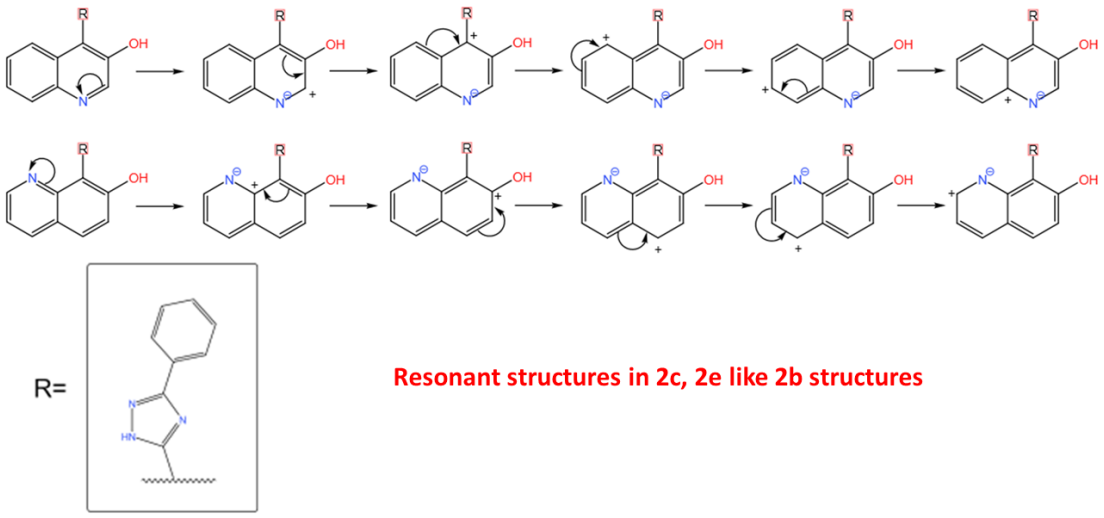

Figure SI 6: Resonant structures in the (iso)quinoline ring, in enol isomer, for molecules **2b** and **2f** in Set 2. The resonant structures for molecules **2c** and **2e** have the same positive partial charges that **2b** molecule. The inset shows the R group of the molecule not involved in the resonant structures.

A.3 Set 3

Molecule	$\lambda_{absorbtion}$	$\lambda^{enol}_{emission}$	$\lambda^{keto}_{emission}$	Stokes^{enol}	Stokes^{keto}	$f_{absorption}$	$f^{keto}_{emission}$
3	354	434	459	80	105	0.21	0.28
3a	377	478	523	101	146	0.20	0.24
3b	369	503	493	134	125	0.19	0.29
3c	353	440	505	87	151	0.18	0.26
3d	353	445	492	92	139	0.19	0.24
3e	349	435	469	86	120	0.20	0.23
3f	363	445	493	83	130	0.20	0.25
3g	349	437	472	88	123	0.18	0.25
3h	339	416	441	77	102	0.18	0.30

Table SI 6: Calculated wavelengths, λ, for absorption and emission (enol and keto forms) in nm, Stokes shift from enol and keto isomer emission (nm), and oscillator strength for the process absorption (enol) and emission (keto) for the molecules of set 3.

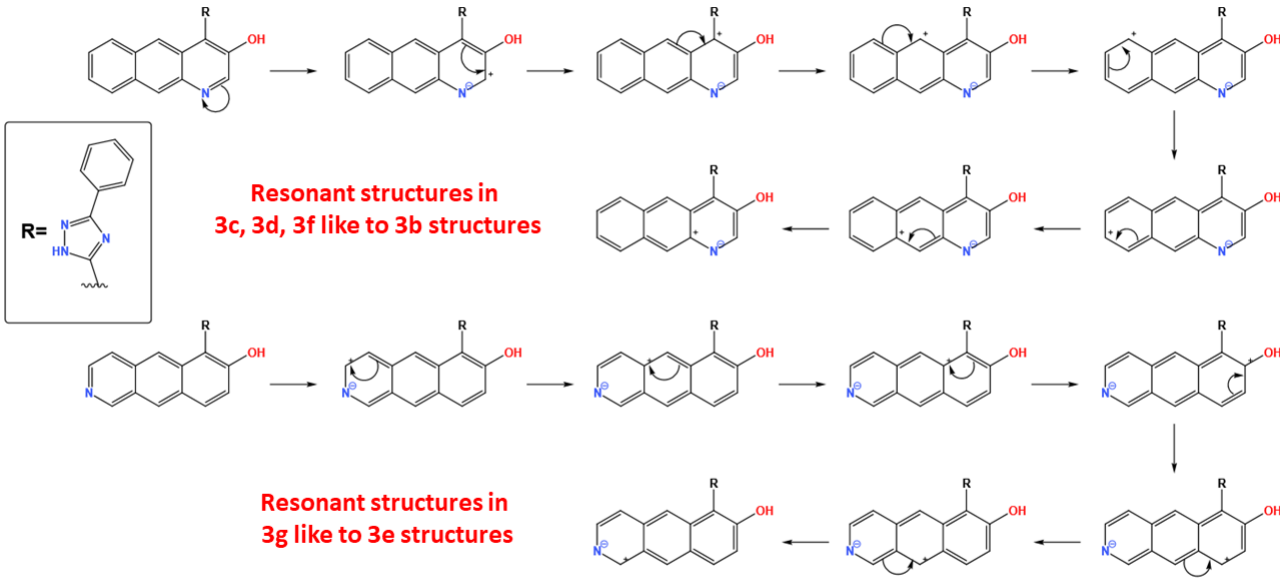

Figure SI 7: Resonant structures in the benzoquinoline ring, in enol isomer, for molecules **3b** and **3e** in Set 3. The resonant structures for molecules **3c**, **3d**, and **3f** have the same positive partial charges that **3b** molecule and the resonant structures for molecule **3g** are similar to **3e**. The inset shows the R group of the molecule not involved in the resonant structures.

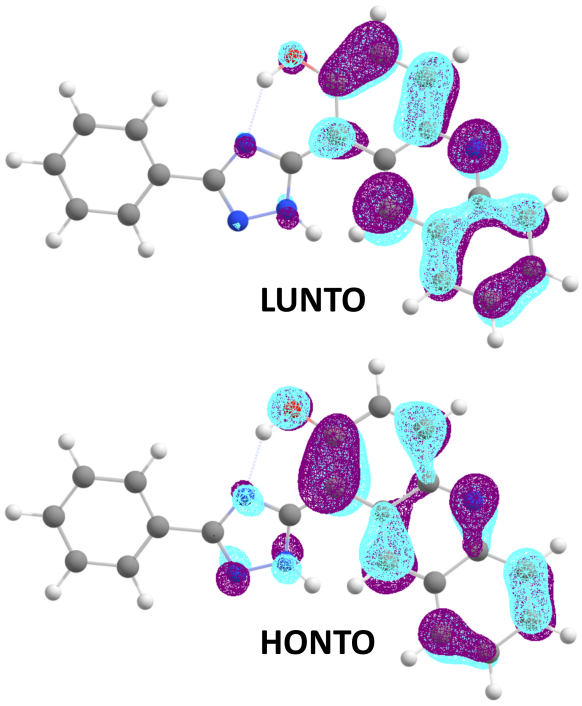

Figure SI 8: NTO for **3c** in the enol isomer. To model the aqueous solvent, we employed the PCM approximation. The system's computations were performed using ω-b97xd/6-311++g(d,p) method.

A.4 Set 4

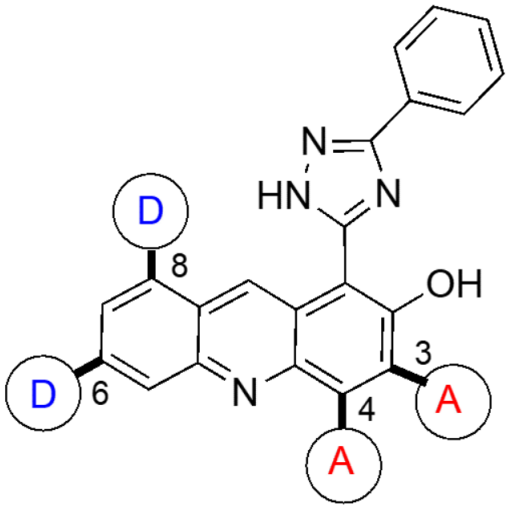

Figure SI 9: Design of optimized 2-hydroxy acridine substituted 1,2,4-triazoles to shift emission near infra-red.

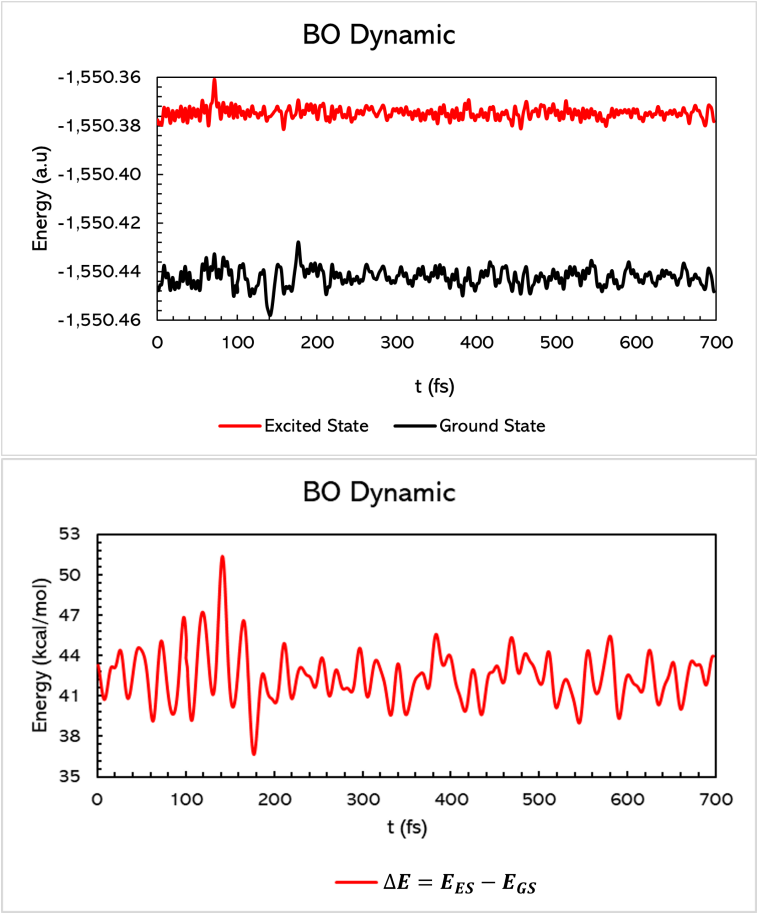

Figure SI 10: Born-Oppenheimer ab initio molecular dynamics were conducted with a total simulation time of 700 fs and a time step of 0.5 fs. Upper is a plot of energy (a.u) for both Ground and Excited states as a function of time (fs) during the dynamic simulation. Down illustrates the difference in energy (kcal/mol) between the excited and ground states over time (fs). BOMD was performed using the PCM model to simulate an aqueous solvent.

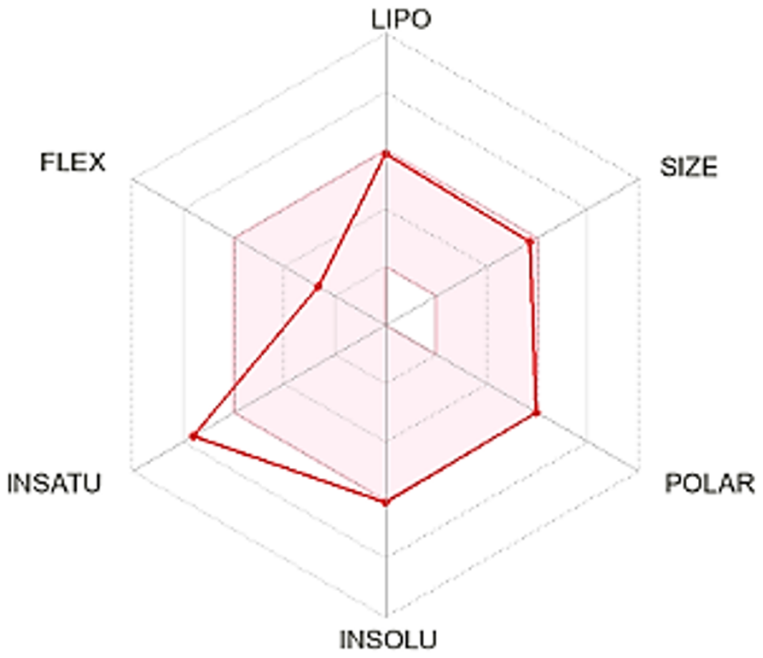

Figure SI 11: Findings related to the physicochemical properties of the **3c** molecule within the context of the biocompatibility analysis (conducted using the freely available online resource for ADME[36] properties http://www.swissadme.ch/.)

The bioavailability assessment, as illustrated in Figure SI 11, and the properties listed in Table SI 7 collectively suggest that the proposed molecule falls within the range of biocompatibility. Nevertheless, a notable concern arises regarding its solubility. It's worth emphasizing that the molecule exhibits characteristics that indicate it can serve as a substrate for permeable glycoprotein (P-gp), which facilitates its transport across biological membranes. Additionally, the molecule demonstrates inhibitory effects on substrates CYP2C19, CYP3A4, and CYP2C9. It is essential to approach these estimates with caution, considering that the original article reporting these interactions was based on data from imbalanced training and testing sets[36].

Shifting our focus to the Druglikeness models proposed by Lipinski, Ghose, Veber, Egan, and Muegge, these sets of rules and criteria are widely employed in pharmaceutical research and drug design to assess a chemical compound's potential for drug development. These rules rely on physicochemical and structural properties to identify compounds with a higher likelihood of success in terms of absorption, distribution, metabolism, excretion, and toxicity (ADMET). Remarkably, four out of the five models indicate that the compound is likely to exhibit biocompatibility.

Additionally, the acridine core is found in the most studied DNA-intercalating agents[37].

Physicochemical Properties		Water Solubility	
Formula	C27H22N8O	Log S (ESOL)	-6.06
Molecular weight	474.52 g/mol	Solubility	4.17e-04 mg/ml ; 8.79e-07 mol/l
Num. heavy atoms	36	Class	Poorly soluble
Num. arom. heavy atoms	25	Log S (Ali)	-7.24
Fraction Csp3	0.15	Solubility	2.76e-05 mg/ml ; 5.81e-08 mol/l
Num. rotatable bonds	4	Class	Poorly soluble
Num. H-bond acceptors	6	Log S (SILICOS-IT)	-8.54
Num. H-bond donors	2	Solubility	1.35e-06 mg/ml ; 2.85e-09 mol/l
Molar Refractivity	139.93	Class	Poorly soluble
TPSA	128.75 Å^2		
Lipophilicity		Druglikeness	
Log Po/w (iLOGP)	2.04	Lipinski	Yes
Log Po/w (XLOGP3)	4.80	Ghose	No; ; 1 violation: MR>130
Log Po/w (WLOGP)	4.42	Veber	Yes
Log Po/w (MLOGP)	1.08	Egan	Yes
Log Po/w (SILICOS-IT)	3.87	Muegge	Yes
Consensus Log Po/w	3.24	Bioavailability Score	0.55
Pharmacokinetics		Medicinal Chemistry	
GI absorption	Low	PAINS	0 alert
BBB permeant	No	Brenk	1 alert: polycyclic aromatic hydrocarbon 2
P-gp substrate	Yes	Leadlikeness	No; 2 violations: MW>350, XLOGP3>3.5
CYP1A2 inhibitor	No	Synthetic accessibility	3.55
CYP2C19 inhibitor	Yes		
CYP2C9 inhibitor	Yes		
CYP2D6 inhibitor	No		
CYP3A4 inhibitor	Yes		
Log Kp (skin permeation)	-5.79 cm/s		

Table SI 7: The ADME (Adsorption, Distribution, Metabolism, Excretion) values for molecule **3c** were determined through calculations conducted on the website http://www.swissadme.ch/. The table furnishes us with key ADME characteristics essential for assessing the potential biocompatibility of a molecule. In this context, GI (gastrointestinal), BBB (blood-brain barrier), P-gp (permeable glycoprotein), CYP (cytochromes P450), PAINS (pan assay interference compounds), and Brenk refer to specific factors and properties relevant to our evaluation.

Consequently, many acridine derivatives show mutagenic properties[38]. Nevertheless, acridines are widely used for their broad biological activities arising from their interactions with DNA. A number of acridines derivatives have been developed as antibacterial, antiprotozoal, and anticancer drugs[39, 40]. It's worth highlighting that this structural motif is present in currently approved drug molecules (acriflavine as antibacterial drug or amsacrine as antileukemia drug)[41].

In terms of synthetic feasibility, the program assigns it a score of 3.55 on a scale of 1 to 10, with 1 indicating ease and 10 indicating significant difficulty. This score implies that synthesizing the compound should not pose substantial challenges. Indeed, acridine is a versatile scaffold which can undergo many regioselective substitution. Substitution reactions at the acridine ring are also well documented[42]. Looking carefully to the retrosynthetic analysis of the target molecule, substitution at the position 1 and 2 of the acridine core to install respectively the triazole moiety and the hydroxyl group could be done using the synthetic strategy previously set up by the organic chemist co-author[25].