Supporting Information

for

Hunting for organic molecules with artificial intelligence:

Molecules optimized for desired excitation energies

Masato Sumita*,†,‡, Xiufeng Yang†,‡, Shinsuke Ishihara*, Ryo Tamura*,‡,¶, and
Koji Tsuda*†,‡,¶

* Center for Advanced Intelligence Project, RIKEN, 1-4-1 Nihombashi, Chuo-ku, Tokyo, 103-0027, Japan.
† International Center for Materials Nanoarchitectonics (WPI-MANA), National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki, 305-0044, Japan.
‡ Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa, Chiba, 277-8561, Japan.
¶ Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki, 305-0047, Japan.

Correspondence and requests for materials should be addressed to M. S. (email: masato.sumita@riken.jp) or to K. T. (email: tsuda@k.u-tokyo.ac.jp)
1. Materials

Compound I (2-methyl-oxazole) is commercially available and was purchased from J&W Pharmlab LLC (catalog No. 56R0594). Compound II (1,2-dimethy-1H-imidazol-5-ol), compound III (4-methyl-6-quinolinol), compound IV (5-methylnaphtalene-2-ol) and compound VI (1-(dimethylamino)-2,3-butanedione) were obtained from Tokyo Chemical Industry Co., Ltd. (TCI) upon custom synthesis. Compound V (N-(2-hydroxybenzyl)-N-methylnitrous amide) was obtained from HeBei Sundia Meditech Company, Ltd. upon custom synthesis. All chemical compounds obtained by custom synthesis satisfy reagent-grade purity (> 96 %), and were used as received.

2. Characterization

**Compound II**: $^1$H-NMR (in CDCl$_3$) in ppm: 4.04 (q, 2H, CH$_2$), 3.05 (s, 3H, CH$_3$), 2.20 (t, 3H, CH$_3$). $^{13}$C-NMR (in CDCl$_3$) in ppm: 181.2, 163.2, 58.3, 26.4, 15.9. LC-MS (m/z): calculated for [C$_5$H$_8$N$_2$O] = 112.06 m/z, found 113.3 m/z (M+H$^+$). Purity (GC): 98.9%. Note that $^1$H- and $^{13}$C-NMR spectra indicate that compound II mainly exist as keto-form in tautomerism.

**Figure S1.** $^1$H-NMR spectrum of compound II (as prepared). This chart is measured by TCI that performed custom synthesis and shown with permission from TCI.
Figure S2. $^{13}$C-NMR spectrum of compound II (as prepared). This chart is measured by TCI that performed custom synthesis and shown with permission from TCI.

Figure S3. $^1$H-NMR spectrum of compound II measured in two weeks after the reagent bottle was opened in air for taking out sample. After the bottle was opened, the bottle was securely sealed and stored at −20 °C for two weeks. Compared with $^1$H-NMR spectra of as-prepared compound, signals from some impurities are observed beside main signals around 2.0–3.5 ppm. This chart is measured by authors using AL300 BX NMR spectrometer (JEOL, Tokyo, Japan).
Figure S4. Photograph of compound II.
**Compound III:** $^1$H-NMR (in DMSO-$d_6$) in ppm: 10.04 (s, 1H, OH), 8.56 (s, 1H, ArH), 7.89 (d, 1H, ArH), 7.34-7.25 (m, 3H, ArH), 2.60 (s, 3H, CH$_3$). $^{13}$C-NMR (in DMSO-$d_6$) in ppm: 156.4, 147.7, 143.6, 142.7, 132.0, 130.1, 122.8, 122.3, 106.0, 19.2. LC-MS (m/z): calculated for [C$_{10}$H$_9$NO] = 159.07 m/z, found 160.2 m/z (M+H$^+$). Purity (GC): 96.6%.

**Figure S5.** $^1$H-NMR spectrum of compound III. This chart is measured by TCI that performed custom synthesis and shown with permission from TCI.
**Figure S6.** $^{13}$C-NMR spectrum of compound III. This chart is measured by TCI that performed custom synthesis and shown with permission from TCI.

**Figure S7.** Photograph of compound III.
**Compound IV**: $^1$H-NMR (in CDCl$_3$) in ppm: 7.90 (d, 1H, ArH), 7.53 (d, 1H, ArH), 7.31 (t, 1H, ArH), 7.16-7.10 (m, 3H, ArH), 4.87 (br, 1H, OH), 2.64 (s, 3H, CH$_3$). $^{13}$C-NMR (in CDCl$_3$) in ppm: 153.0, 134.8, 134.3, 128.1, 126.4, 126.2, 124.9, 124.5, 117.3, 110.2, 19.4. LC-MS (m/z): calculated for [C$_{11}$H$_{10}$O] = 158.07 m/z, found 159.0 m/z (M+H$^+$). Purity (LC): 99.5%.

**Figure S8.** $^1$H-NMR spectrum of compound IV. This chart is measured by TCI that performed custom synthesis and shown with permission from TCI.
**Figure S9.** $^{13}$C-NMR spectrum of compound IV. This chart is measured by TCI that performed custom synthesis and shown with permission from TCI.

**Figure S10.** Photograph of compound IV.
**Compound V**: $^1$H-NMR (in DMSO-d$_6$) in ppm: 9.79 (s, 1H, OH), 7.17 (t, 1H, ArH), 7.09 (d, 1H, ArH), 6.87 (d, 1H, ArH), 6.80 (t, 1H, ArH), 5.26 (s, 2H, benzyl-CH$_2$), 2.89 (s, 3H, CH$_3$). $^{13}$C-NMR (in DMSO-d$_6$) in ppm: 156.6, 129.9, 129.4, 121.1, 119.2, 115.4, 51.7, 31.1. LC-MS (m/z): calculated for [CsH$_{10}$N$_2$O$_2$] = 166.07 m/z, found 167.0 m/z (M+H$^+$). Purity (LC): 99.2%.

**Figure S11.** $^1$H-NMR spectrum of compound V. This chart is measured by authors using AL300 BX NMR spectrometer (JEOL, Tokyo, Japan).

**Figure S12.** $^{13}$C-NMR spectrum of compound V. This chart is measured by authors using AL300 BX NMR spectrometer (JEOL, Tokyo, Japan).
Figure S13. Photograph of compound V.
**Compound VI:** $^1$H-NMR (in CDCl$_3$) in ppm: 6.31 (s, 1H, C=CH), 6.2-6.0 (br, 0.5H, OH), 3.10 (s, 6H, N(CH$_3$)$_2$), 2.14 (s, 3H, CH$_3$). $^{13}$C-NMR (in CDCl$_3$) in ppm: 187.9, 132.6, 130.0, 42.3, 21.2. LC-MS (m/z): calculated for [C$_6$H$_{11}$NO$_2$] = 129.08 m/z, found 130.4 m/z (M+H$^+$). Purity (GC): 96.9%. Note that $^1$H- and $^{13}$C-NMR spectra indicate that compound VI mainly exist as enol-form in tautomerism.

![Diagram of tautomerism](image)

**Figure S14.** $^1$H-NMR spectrum of compound VI. This chart is measured by TCI that performed custom synthesis and shown with permission from TCI.
**Figure S15.** $^{13}$C-NMR spectrum of compound VI. This chart is measured by TCI that performed custom synthesis and shown with permission from TCI.

**Figure S16.** Photograph of compound VI.
Table S1. Energies of keto/enol-form of II

|                | Keto       | Enol       |
|----------------|------------|------------|
| Energy / $E_h$ | -377.99028 | -377.96296 |
| Relative energy / kJ mol$^{-1}$ | 0.0         | 71.72      |

Figure S17. Computational UV-vis spectra for keto/enol-forms of II.

Table S2. Energies of syn/anti-conformers in keto/enol-forms of VI

|                | syn       | anti      |
|----------------|-----------|-----------|
| keto           |           |           |
| Energy / $E_h$ | -437.99645 | -438.01127 |
| Relative energy / kJ mol$^{-1}$ | 62.62      | 23.71     |
| enol           |           |           |
| Energy / $E_h$ | -438.02030 | -438.01697 |
| Relative energy / kJ mol$^{-1}$ | 0.0        | 8.74      |
**Figure S18.** Computational UV-vis spectra keto/enol-forms of VI
3. Dependence of solvent and concentration

**Compound I**

![Experimental UV-vis absorption spectra of I](image1)

**Figure S19.** Experimental UV-vis absorption spectra of I with various concentration and solvent.

**Compound II**

![Experimental UV-vis absorption spectra of II](image2)

**Figure S20.** Experimental UV-vis absorption spectra of II with various concentration and solvent.
Compound III

Figure S21. Experimental UV-vis absorption spectra of III with various concentration and solvent.
Compound IV

![Experimental UV-vis absorption spectra of IV with various concentration and solvent.](image1)

Figure S22. Experimental UV-vis absorption spectra of IV with various concentration and solvent.

Compound V

![Experimental UV-vis absorption spectra of V with various concentration and solvent.](image2)

Figure S23. Experimental UV-vis absorption spectra of V with various concentration and solvent.
Compound VI

Figure S24. Experimental UV-vis absorption spectra of VI with various concentration and solvent.