

Supporting information

Uncovering Novel Liquid Organic Hydrogen Carriers: A Systematic
Exploration of Chemical Compound Space using Cheminformatics and
Quantum Chemical Methods

Hassan Harb¹, Sarah N. Elliott², Logan Ward³, Ian T. Foster³, Stephen J. Klippenstein², Larry A.
Curtiss¹, and Rajeev Surendran Assary^{1, *}

¹Materials Science Division, Argonne National Laboratory, Lemont, IL 60439, U.S.A

²Chemical Sciences and Engineering Division, Argonne National Laboratory, Lemont, IL 60439,
U.S.A

³Data Sciences and Learning Division, Argonne National Laboratory, Lemont, IL 60439, U.S.A

* Corresponding Author (RSA): assary@anl.gov, Phone: +1 630-252-3536

Table of Contents

Figure S1: Comparison of Tanimoto similarity index (left) and our provisional scoring criteria (right) with DFT-calculated dehydrogenation enthalpies of 700 molecules down selected from GDB-17 database. DFT calculations performed with ω b97X-D/6-31G(2df,p) model chemistry.

Text S1: Details of Provisional Scoring Criteria

Text S2: Limitations of using Tanimoto Similarity index.

Table S1: List of datasets compiled in this study.

Figure S2: Plot of residual error of ω b97X-D in Exp-31 dataset.

Figure S3: Mean-absolute errors of B3LYP, ω b97X-D, M06, and B3LYP + GD3 in Exp-31 dataset.

Figure S4: Schematic overview of screening methodology.

Table S2: Experimental dataset.

Table S3: Benchmark dataset.

Table S4: Dehydrogenation scoring criteria.

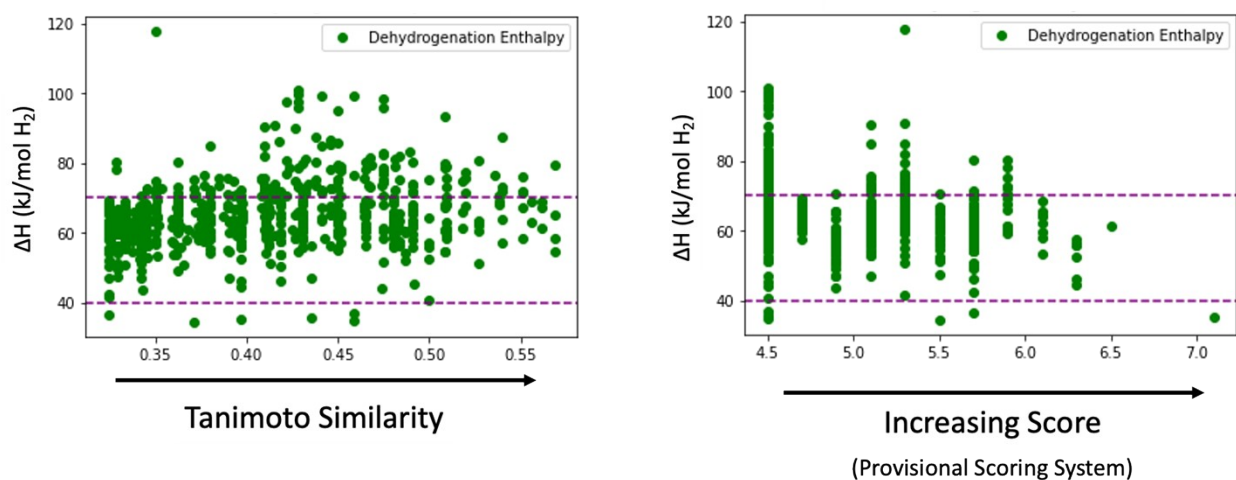


Figure S1: Comparison of Tanimoto similarity index (left) and our provisional scoring criteria (right) with DFT-calculated dehydrogenation enthalpies of 700 molecules down selected from GDB-17 database. DFT calculations performed with ω b97X-D/6-31G(2df,p) model chemistry. The

plots show that our provisional scoring criteria presents a better correlation with desired range of dehydrogenation enthalpies for LOHCs (shown as dashed purple lines).

Text S1: Details of Provisional Scoring Criteria

In the course of developing our provisional criteria for identifying potent LOHC candidates, we identified 11 key structural descriptors from extensive literature review. Initially, these descriptors were each assigned an arbitrary score based on their perceived significance in LOHC performance. Subsequently, these scores were optimized against the established 'Bench-93' benchmark dataset using multivariate linear regression, a well-regarded statistical technique for modeling multiple variables. This optimization refined the arbitrary scores into more accurate descriptor weights, leading to the formulation of our provisional scoring method depicted in Equation 4 in the main text, where each descriptor's weight is multiplied by its appearance count in a molecule.

In the interest of demonstrating the validity of our scoring method, we performed a comparative analysis involving 700 molecules, assessing these using DFT calculations, Tanimoto similarity coefficients, and our provisional scoring criteria. This comparison, illustrated in Figure S1, reveals a more robust correlation between our provisional scoring criteria and the DFT-calculated dehydrogenation enthalpies of the molecules, which were selectively extracted from the GDB-17 database and evaluated using the ω B97X-D/6-31G(2df,p) model chemistry. Particularly noteworthy is the better alignment of our scoring criteria with the desired range of dehydrogenation enthalpies for LOHCs (represented by the dashed purple lines), underscoring its potential for guiding the selection of promising LOHC candidates.

Text S2: Limitations of using Tanimoto Similarity index.

As depicted in Figure S1, our search strategy begins with the application of the Tanimoto similarity index, narrowing down the candidates to a subset of all possible LOHC systems within the chemical compound space (CCS). This approach inherently restricts the selection to molecules that bear resemblance to known LOHCs, potentially omitting other viable candidates residing in different subspaces of the CCS. Other methodologies, such as the one reported in [reference PubChem paper], utilize distinct techniques and have the capability to target other promising LOHC systems. These alternative approaches may explore regions of the CCS that might not overlap with the subspace studied here, thereby underscoring the importance of diverse strategies in the ongoing exploration and discovery of new LOHCs.

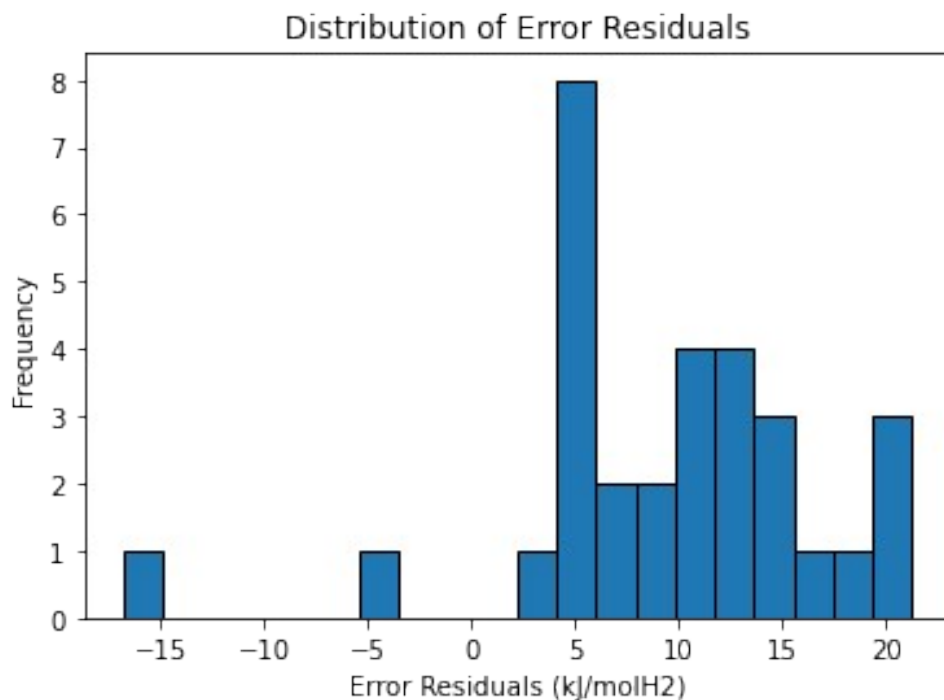


Figure S2: Distribution of residual errors of DFT results (ω b97X-D/6-31G(2df,p)) as compared to experimental results (Exp-31). The plot shows that majority of the datapoints (29) show positive residual errors, indicating that the model chemistry used overestimates the dehydrogenation enthalpies.

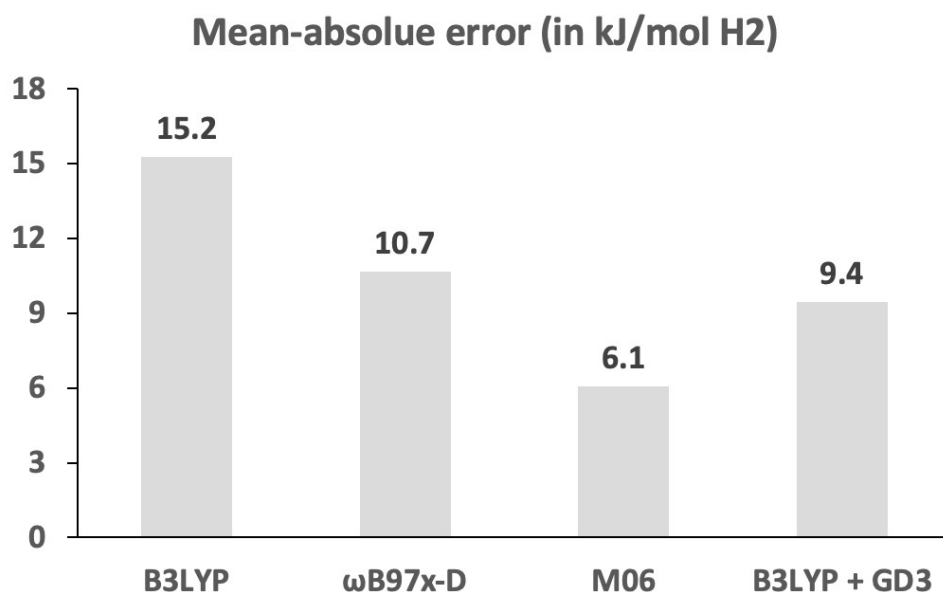


Figure S3: Mean absolute errors of four model chemistries: B3LYP, ω b97X-D, M06, and B3LYP + GD3 (Grimme's D3 empirical dispersion). All calculations were performed using 6-31G(2df,p) basis set. In the selection of the functional for the computational study, the ω b97x-d functional

was chosen over others like B3LYP, M06, and B3LYP+GD3, despite their respective Mean Absolute Errors (MAEs). The choice of wb97x-d was driven by its ability to handle both short-range and long-range interactions, making it ideal for representing aromatic species. A comparative study identified it as one of the best for non-covalent interactions, and it provides a robust and physically meaningful description. The 6-31g(2df,p) basis set was selected for its performance across various functionals, and the aug-cc-pvtz basis set was used for heavier atoms to accurately account for higher angular momentum orbitals.

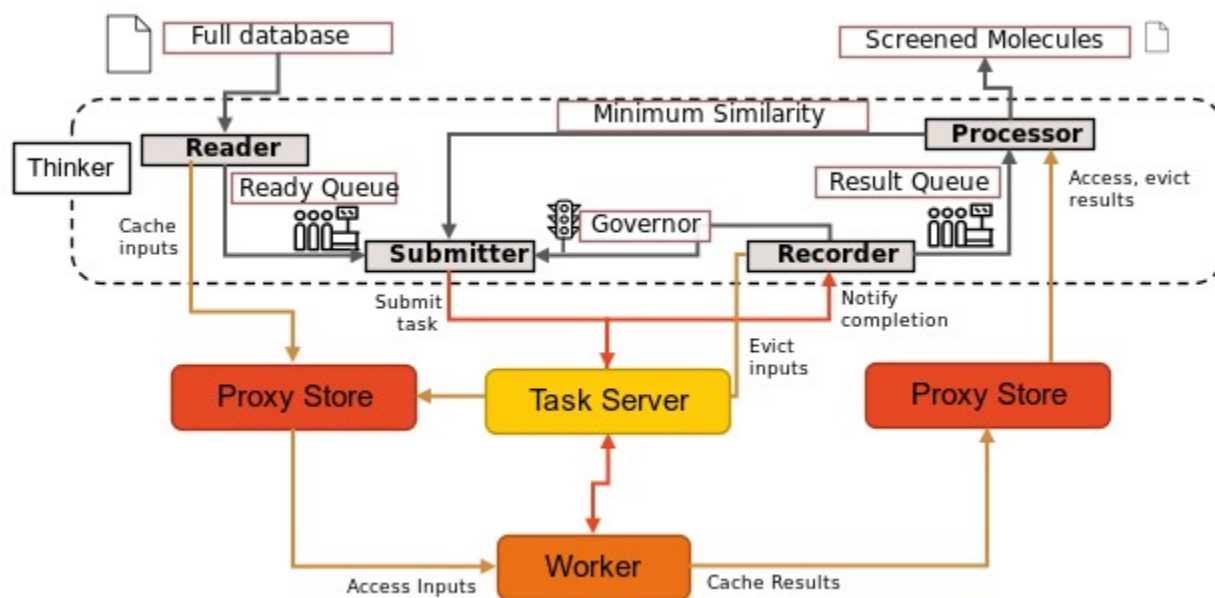


Figure S4: Schematic overview of screening methodology. Full description can be found in the main text.

Label	Number of Reactions	Comments
Exp-31	31 reactions	Experimentally available data (Table S2)
Bench-93	93 reactions	DFT-calculated data, permutations based on Exp-31 (Table S3)
ZINC-88	88 reactions	DFT-calculated, obtained from ZINC-15 screening
GDB-3000	3000 reactions	DFT-calculated, obtained from GDB-17 screening

Table S1: List of the datasets produced in this study and a short description of each. SMILES strings and all DFT-calculated dehydrogenation enthalpies of GDB-3000 and ZINC-88 are publicly available on GitHub: <https://github.com/HydrogenStorage/LOHC>. Details on the different

datasets generated in this study and descriptors involved in the dehydrogenation scoring criteria, and experimental and benchmark datasets are provided in the supplementary information. Python scripts for finding similar molecules within GDB-17 or other large datasets can be found on GitHub: <https://github.com/HydrogenStorage/screening-large-databases>.

Table S2: SMILES Structures, experimental values, and DFT-calculated values for the dehydrogenation enthalpies of the experimental set. MAE = 11.0 and RMSD 12.7. All energies are given in units of kJ/mol of H₂.

Label	Smiles (dehydrogenated)	ωB97X-D	Expt.	Ref
1	<chem>N1N=CC=N1</chem>	5.5	22.3	1
2	<chem>CN1C=CN=C1</chem>	44.1	39.1	1
3	<chem>N1C=CN=C1</chem>	44.5	39.8	1
4	<chem>C1=C3C(=CC2=C1[N](C(=C2C)C)C(=C([N]3C)C)C</chem>	54.8	40.7	1
5	<chem>C1=CC=C2C(=C1)C3=CC=CC=C3N2(CC)</chem>	64.2	50.6	2
6	<chem>C1=CC=C2C(=C1)C3=CC=CC=C3N2</chem>	63.9	51.1	2
7	<chem>CN1C=CC2=CC=CC=C21</chem>	62.9	51.9	3
8	<chem>CC1=CC2=CC=CC=C2N1</chem>	60.0	55.2	4
9	<chem>C1=CNC=C1</chem>	69.2	56.1	1
10	<chem>C1=CC=C2C(=C1)C=CN2</chem>	68.0	56.6	4
11	<chem>C1=CN=CC=C1N</chem>	52.6	56.7	5
12	<chem>C1=CC=CC2=C1OC3=C2C=CC=C3</chem>	66.4	56.7	1
13	<chem>C1=CN=CC=N1</chem>	66.7	56.9	6
14	<chem>C1=CN=CN=C1</chem>	64.8	60.1	1
15	<chem>C1=CC=C(C=C1)O</chem>	76.8	61	1
16	<chem>C1=CC=C2C(=C1)N=C3C=CC=CC3=N2</chem>	66.1	61.3	7
17	<chem>C1=CC=C2C(=C1)C=CC=N2</chem>	74.9	61.9	8
18	<chem>C1=CC=NC=C1</chem>	73.8	62.3	1
19	<chem>CC1=CC=CC=C1CC2=CC=CC=C2</chem>	76.7	63.5	8
20	<chem>C1=CC=C(C=C1)N</chem>	75.6	64.0	9,10
21	<chem>CC1=CC2=CC=CC=C2C=C1</chem>	73.1	65.3	8
22	<chem>CC1=C(C=CC=C1CC2=CC=CC=C2)CC3=CC=CC=C3</chem>	68.0	65.4	8
23	<chem>C1=CC=C(C=C1)C2=CC=CC=C2</chem>	70.8	66.6	1
24	<chem>C1=CC=C2C=CC=CC2=C1</chem>	80.7	66.7	11
25	<chem>CC1=CC=CC=C1</chem>	73.3	68.3	11
26	<chem>C1=CC=CC=C1</chem>	83.6	68.6	11
27	<chem>CC1=CC(=CC=C1)CC2=CC=CC=C2</chem>	77.1	71	8
28	<chem>C1C=CC=C1</chem>	125.3	107.5	1
29	<chem>C1CC=CC=C1</chem>	135.0	113.8	10
30	<chem>C1C=CCC=C1</chem>	134.6	113.8	10
31	<chem>C1CCC=CC1</chem>	138.3	118.83	10

Table S3: SMILES Structures and DFT-calculated dehydrogenation enthalpies for the benchmark set. All energies are given in units of kJ/mol of H₂.

Label	SMILES (Dehydrogenated)	ωB97X-D
1	C1=CC=CC=C1	83.6
2	CC1=CC=CC=C1	73.3
3	C1=CC=C2C=CC=CC2=C1	71.3
4	C1=CC=C2C=CC=CC2=C1	80.7
5	CC1=CC2=CC=CC=C2C=C1	66.9
6	CC1=CC2=CC=CC=C2C=C1	73.1
7	C1=CC=C(C=C1)C2=CC=CC=C2	70.6
8	C1=CC=C(C=C1)C2=CC=CC=C2	70.8
9	CC1=CC=CC=C1CC2=CC=CC=C2	76.0
10	CC1=CC=CC=C1CC2=CC=CC=C2	76.7
11	CC1=CC(=CC=C1)CC2=CC=CC=C2	77.0
12	CC1=CC(=CC=C1)CC2=CC=CC=C2	77.1
13	C1=CC=C2C(=C1)C3=CC=CC=C3N2	57.1
14	C1=CC=C2C(=C1)C3=CC=CC=C3N2	63.9
15	C1=CC=C2C(=C1)C3=CC=CC=C3N2(C)	55.6
16	C1=CC=C2C(=C1)C3=CC=CC=C3N2(C)	64.0
17	C1=CC=C2C(=C1)C3=CC=CC=C3N2(CC)	54.9
18	C1=CC=C2C(=C1)C3=CC=CC=C3N2(CC)	64.2
19	C1=CC=C2C(=C1)C3=CC=CC=C3N2(CCC)	54.4
20	C1=CC=C2C(=C1)C3=CC=CC=C3N2(CCC)	63.1
21	C1=CC=C2C(=C1)C=CN2	68.0
22	CN1C=CC2=CC=CC=C21	62.9
23	CC1=CC2=CC=CC=C2N1	60.0
24	CC1=CC2=CC=CC=C2N1C	61.5
25	CC1=C(C=CC=C1CC2=CC=CC=C2)CC3=CC=CC=C3	68.0
26	C1=CC=C2C(=C1)N=C3C=CC=CC3=N2	66.1
27	C1=CC(N)=C2C(=C1)N=C3C=CC=C(N)C3=N2	55.5
28	C1=CC(N)=C2C(=C1)N=C3C=CC=C(N)C3=N2	54.9
29	C1=CC(N(C)(C))=C2C(=C1)N=C3C=CC=C(N(C)(C))C3=N2	61.5
30	C1=CC(N(C)(C))=C2C(=C1)N=C3C=CC=C(N(C)(C))C3=N2	63.2
31	CN1N=CC=N1	18.8
32	CN1C=CN=C1	44.1
33	CN1C=CC=N1	35.2
34	C1CC=CC=C1	135.0
35	C1=CC=C(C(=C1)N)N	57.7

36	C1=CC=NN=C1	74.9
37	C1=CN=NN=C1	72.7
38	C1=CN=NN1	24.4
39	C1=CN=NC=N1	69.4
40	N1C=NC=N1	14.6
41	C1=CC(=CC(=C1)N)N	56.6
42	C1=CN=CN=C1	64.8
43	N1C=CN=C1	44.5
44	N1=CNC=N1	27.6
45	C1=C(C=C(C=C1N)N)N	57.7
46	C1=NC=NC=N1	44.3
47	C1C=CCC=C1	134.6
48	C1=CC(=CC=C1N)N	70.5
49	C1=CN=CC=N1	66.7
50	CN1N=CC=N1	0.0
51	CN1C=NC=N1	4.0
52	N1N=CC=N1	5.5
53	CN1C=NN=C1	23.2
54	C1=CC=C(C=C1)N	75.6
55	C1CCC=CC1	138.3
56	C1C=CC=C1	125.3
57	N1C=CC=N1	39.1
58	C1=CC=NC=C1	73.8
59	C1=CNC=C1	69.2
60	C1=CN(CCCC)C=C1	71.6
61	C1=CN(CC)C=C1	70.5
62	C1=CN(C)C=C1	72.7
63	C1=CN(CCC)C=C1	68.8
64	C1=CN(C=C1)N2C=CC=C2	75.2
65	C1=CN(CC)C(CC)=C1	65.0
66	C1=CN(C)C(C)=C1	70.6
67	C1=CN(CCC)C(CCC)=C1	61.5
68	C1=CC2=C(C=CC=N2)N=C1	58.3
69	C1=CCN2C1=CC=C2	91.3
70	CCC1=NC=CN=C1	64.7
71	C1=CNC(CC)=C1	68.9
72	CCC1=NC2=CC=CC=C2C=C1	67.8
73	CC1=NC=CN=C1	64.2
74	C1=CNC(C)=C1	65.8

75	CC1=NC2=CC=CC=C2C=C1	66.3
76	CCCC1=NC=CN=C1	61.7
77	C1=CNC(CCC)=C1	67.6
78	CCCC1=NC2=CC=CC=C2C=C1	67.8
79	CCC1=NC(=CC=C1)CC	61.8
80	C1=CC2=C(C=CC(C)=N2)N=C1C	54.6
81	CC1=NC(=CC=C1)C	68.9
82	CCCC1=NC(=CC=C1)CCC	58.2
83	C1=CN=CC=C1C2=CC=NC=C2	61.4
84	C1=CC2=CC=CN2C=C1	74.3
85	C1=CC=C2C=NC=CC2=C1	65.3
86	C1=C2C(=CC(=N2))N=C1	119.4
87	C1=CC=C2C(=C1)C=CC=N2	74.9
88	C1=CC=C2C(=C1)N=CC=N2	64.4
89	C1=CN=CC=C1N	52.6
90	C1=C3C(=CC2=C1[N](C(=C2C)C)C)C(=C([N]3C)C)C	54.8
91	C1=CC=C(C=C1)O	76.8
92	C1=COC=C1	85.5
93	C1=CC=CC2=C1OC3=C2C=CC=C3	66.4

Table S4: Dehydrogenation scoring criteria. Columns include the descriptor's tag, its weight, and a brief description. The predicted dehydrogenation enthalpy depends on the counts of the descriptors multiplied by their respective weights. All weights are in units of 0.1 kJ/mol of H₂.

Entry	Descriptor	Weight	Remarks
1	% wt. h2	[0, 1]	cutoff parameter, score under 5.5 is 0
2	sp3_carb	0.1153	sp3 carbons
3	sp2_carb	0.2872	sp2 carbons
4	sp_carb	1.1298	sp carbons
5	M-no5mem	-1.2127	monocyclic rings that aren't 5-membered rings
6	M-5mem	-1.6513	5-membered monocyclic rings
7	M-nitcont	-0.0955	monocyclic rings with a nitrogen atom in the ring
8	M-SOcont	0.9387	monocyclic rings with a S or O atoms in the ring
9	B-no5mem	5.2805	bicyclic rings that don't have any 5-membered rings
10	B-5mem	5.0358	5-membered ring containing bicyclic rings
11	B-nitcont	-0.3393	bicyclic rings with a nitrogen atom in any ring
12	B-SOcont	-0.0191	bicyclic rings with a sulfur or oxygen atoms in any ring
13	P-no5mem	1	polycyclic (3 or more) rings that don't have any 5-membered rings
14	P-5mem	3.491	5-membered ring containing polycyclic rings
15	P-nitcont	0.5179	polycyclic rings with a nitrogen atom in any ring
16	P-SOcont	0.0603	polycyclic rings with a sulfur or oxygen atoms in any ring
17	nonCfuse	0.52	non-carbon atoms that are the fusion point of two rings
18	SObadpos	0.9095	S or O is gamma to a fusion point
19	SO_adjrng	2	two fused rings both have a S or O
20	3mem_bi+	3.9448	3-membered ring is fused to another ring
21	one-posi	-0.0735	substituents on a ring (including another ring)
22	onethree	-0.2623	1,3 substituent or 1,3 nitrogen interactions on a ring

References

- (1) He, T.; Pei, Q.; Chen, P. Liquid Organic Hydrogen Carriers. *Journal of Energy Chemistry* **2015**, *24* (5), 587–594. <https://doi.org/10.1016/j.jechem.2015.08.007>.
- (2) Biniwale, R.; Rayalu, S.; Devotta, S.; Ichikawa, M. Chemical Hydrides: A Solution to High Capacity Hydrogen Storage and Supply. *International Journal of Hydrogen Energy* **2008**, *33* (1), 360–365. <https://doi.org/10.1016/j.ijhydene.2007.07.028>.
- (3) Rao, P. C.; Yoon, M. Potential Liquid-Organic Hydrogen Carrier (LOHC) Systems: A Review on Recent Progress. *Energies* **2020**, *13* (22), 6040. <https://doi.org/10.3390/en13226040>.
- (4) Konnova, M. E.; Li, S.; Bösmann, A.; Müller, K.; Wasserscheid, P.; Andreeva, I. V.; Turovtzev, V. V.; Zaitsau, D. H.; Pimerzin, A. A.; Verevkin, S. P. Thermochemical Properties and Dehydrogenation Thermodynamics of Indole Derivates. *Ind. Eng. Chem. Res.* **2020**, *59* (46), 20539–20550. <https://doi.org/10.1021/acs.iecr.0c04069>.
- (5) Cui, Y.; Kwok, S.; Bucholtz, A.; Davis, B.; Whitney, R. A.; Jessop, P. G. The Effect of Substitution on the Utility of Piperidines and Octahydroindoles for Reversible Hydrogen Storage. *New J. Chem.* **2008**, *32* (6), 1027. <https://doi.org/10.1039/b718209k>.
- (6) Clot, E.; Eisenstein, O.; Crabtree, R. H. Computational Structure–Activity Relationships in H₂ Storage: How Placement of N Atoms Affects Release Temperatures in Organic Liquid Storage Materials. *Chem. Commun.* **2007**, No. 22, 2231–2233. <https://doi.org/10.1039/B705037B>.
- (7) Niermann, M.; Beckendorff, A.; Kaltschmitt, M.; Bonhoff, K. Liquid Organic Hydrogen Carrier (LOHC) – Assessment Based on Chemical and Economic Properties. *International Journal of Hydrogen Energy* **2019**, *44* (13), 6631–6654. <https://doi.org/10.1016/j.ijhydene.2019.01.199>.
- (8) Aakko-Saksa, P. T.; Cook, C.; Kiviaho, J.; Repo, T. Liquid Organic Hydrogen Carriers for Transportation and Storing of Renewable Energy – Review and Discussion. *Journal of Power Sources* **2018**, *396*, 803–823. <https://doi.org/10.1016/j.jpowsour.2018.04.011>.
- (9) NIST: Reaction Enthalpy Was Calculated Form the Enthalpies of Formation of Reactants and Products, Which Were Obtained from the NIST Chemistry WebBook.
- (10) Linstrom, P. NIST Chemistry WebBook, NIST Standard Reference Database 69, 1997. <https://doi.org/10.18434/T4D303>.
- (11) Kariya, N.; Fukuoka, A.; Ichikawa, M. Efficient Evolution of Hydrogen from Liquid Cycloalkanes over Pt-Containing Catalysts Supported on Active Carbons under “Wet–Dry Multiphase Conditions.” *Applied Catalysis A: General* **2002**, *233* (1–2), 91–102. [https://doi.org/10.1016/S0926-860X\(02\)00139-4](https://doi.org/10.1016/S0926-860X(02)00139-4).