Supporting Information for

Machine Learning Dihydrogen Activation in the Chemical Space Surrounding Vaska's Complex

Pascal Friederich,[1,2,3] Gabriel dos Passos Gomes[1,3], Riccardo De Bin,[4] Alán Aspuru-Guzik,[1,3,5,6,] and David Balcells[7,]**

[1]Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada.

[2]Institute of Nanotechnology, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany.

[3]Department of Computer Science, University of Toronto, 214 College St., Toronto, Ontario M5T 3A1, Canada.

[4]Department of Mathematics, University of Oslo, P.O. Box 1033, Blindern, N-0315, Oslo, Norway

[5]Vector Institute for Artificial Intelligence, 661 University Ave. Suite 710, Toronto, Ontario M5G 1M1, Canada.

[6]Lebovic Fellow, Canadian Institute for Advanced Research (CIFAR), 661 University Ave, Toronto, ON M5G 1M1, Canada.

[7]Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Oslo, P.O. Box 1033, Blindern, N-0315, Oslo, Norway

Table of Contents

DFT data exploration and cleaning. Figure S1 shows all activation barriers $\Delta E^{\ddagger}_{HH}$ as a function of the respective H-H distance in the transition state $d^{\ddagger}_{HH}$. The resulting scatter plot was used to identify outliers. The data points within the range 2.00 Å < $d^{\ddagger}_{HH}$ < 2.25 Å are associated with calculations that, at the **DFT-2** stage, converged into spurious transition states nearby the equilibrium geometry of the dihydride product. These data points, which have the longest $d^{\ddagger}_{HH}$ values and $\Delta E^{\ddagger}_{HH} < 0$, were thus excluded. The data points with $d^{\ddagger}_{HH}$ < 0.80 Å were also excluded since they were identified as non-activated dihydrogen structures – their mean $d^{\ddagger}_{HH}$ value (0.78 Å) is very close to the interatomic distance of the isolated H_2 molecule optimized at the same level of theory (0.77 Å). After excluding the outliers, the refined {$d^{\ddagger}_{HH}$, $\Delta E^{\ddagger}_{HH}$} space included the 1,947 data points used to train and test the machine learning models (*vide infra*). The final success rate over the attempted 2,574 H_2-activation barriers was 75.6%.

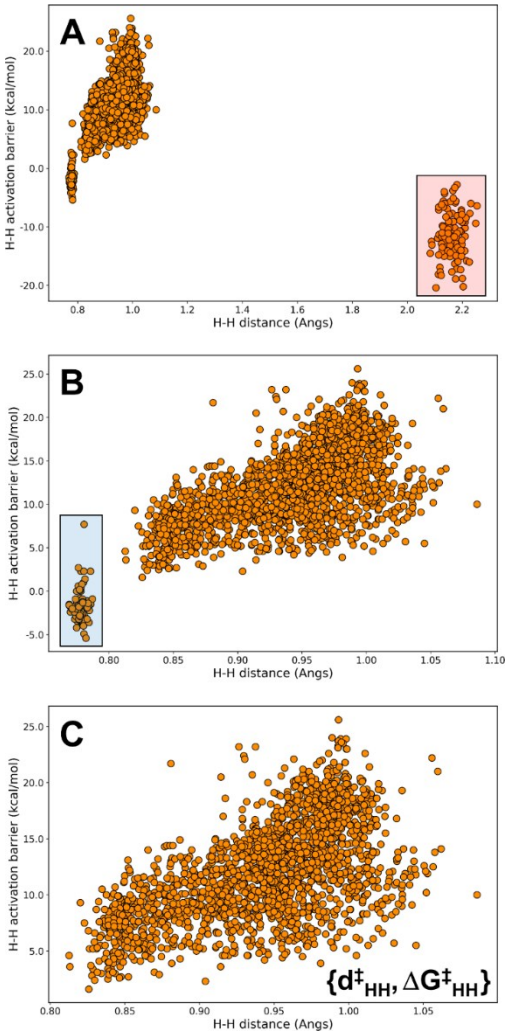

Figure S1. The scatter plot **A** contains all the 2,197 data points converged in the DFT calculations. The transparent red box highlights the dihydride outliers. After excluding the dihydrides, plot **B** shows the remaining data (2,087 points), with the transparent blue box highlighting the dihydrogen outliers. Plot **C** shows the final {$d^{\ddagger}_{HH}$, $\Delta E^{\ddagger}_{HH}$} space of DFT data (1,947 points) after excluding all outliers.

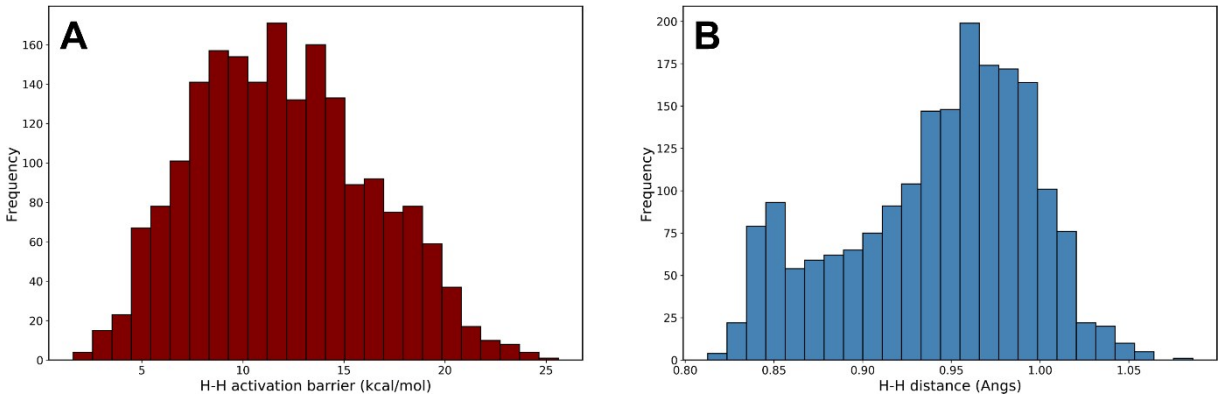

Figure S2. Histograms showing the distribution of the H-H activation barriers (**A**) and distances (**B**).

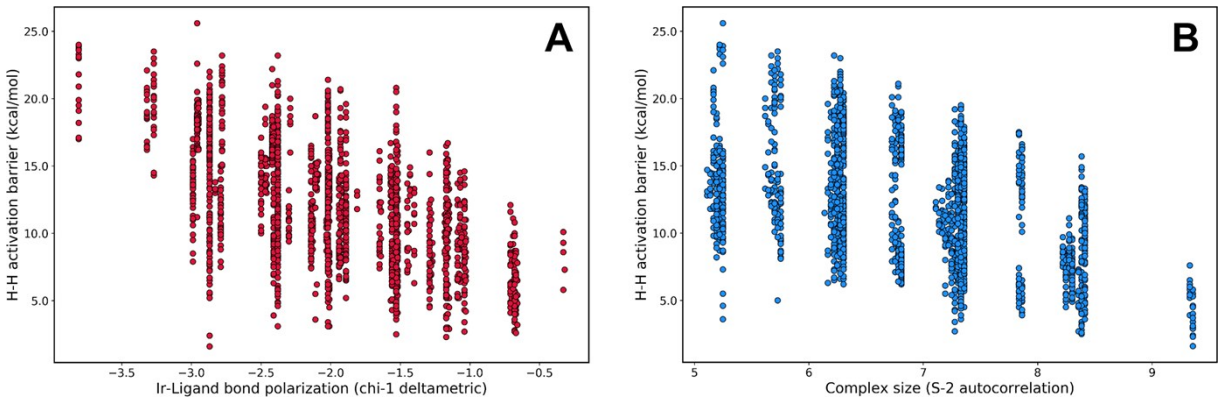

Figure S3. Scatter plots of the H-H activation barriers vs. **A)** χ_1 (*i.e.*, polarization of the Ir-Ligand bonds) and **B)** S_2 (*i.e.*, size of the metal complexes at depth = 2).

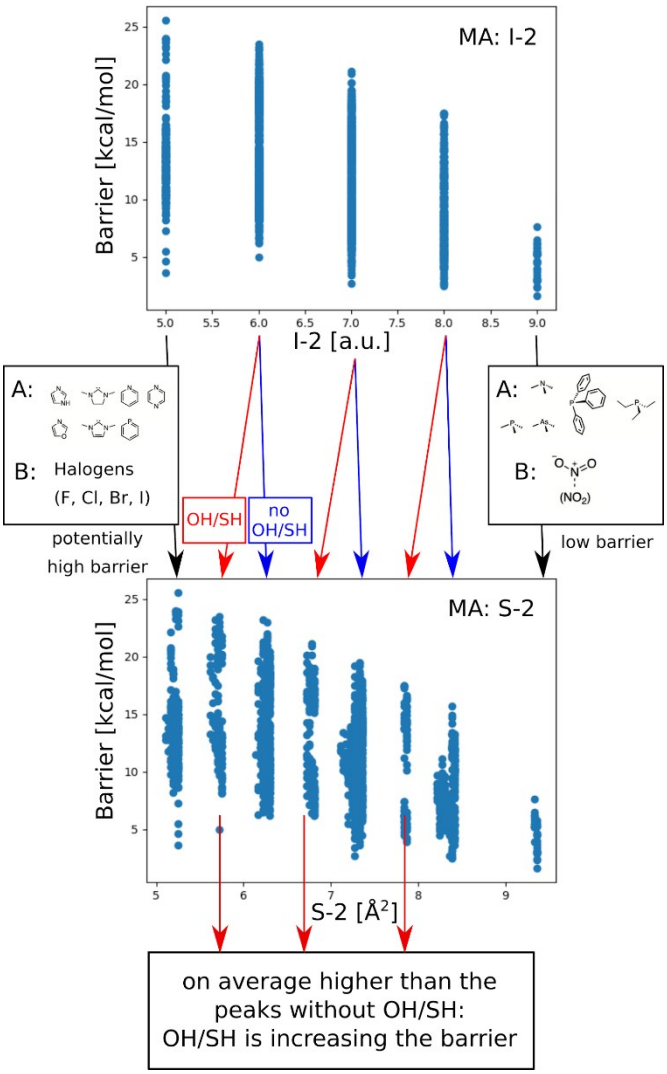

Figure S4. Combined interpretation of the MA features I-2 and S-2. The I-2 feature counts the number of atoms at a distance of two chemical bonds from Ir, whereas the S-2 feature accounts for the size of the atoms placed at the same distance from the metal center.

The I-2 vs. $\Delta E^{\ddagger}_{HH}$ scatter plot shown in **Figure S4** shows five distinct values, corresponding to the number of second nearest atoms to Ir. The peak at I-2 = 9 corresponds to all complexes in which the A ligands (Figure 2) are a combination of PR_3, AsR_3, or NR_3, the B ligand is nitro, and the C ligand is arbitrary - no other combination yields I2 = 9. Interestingly, this combination of ligands guarantees an H_2-activation barrier lower than 10 kcal/mol. In the S-2 distribution, we can see eight distinct peaks instead of the five yielded by I-2. Since all atoms at d = 2 are heavy atoms (*i.e.,* not H), the I-2 peak at 9 corresponds to the S-2 peak at ~9.5. The same is true for the correlating I-2 = 5 and S-2 ≈ 5.2 peaks, which originate from a combination of two heterocyclic A ligands, a halide B ligand, and an arbitrary C ligand. In the three intermediate I-2 peaks at 6, 7, and 8, the B ligand can contain the OH or SH ligands, which are the only placing an H atom at d = 2. This singularity splits these I-2 peaks into the six intermediate peaks observed in the S-2 scatter plot. Combined with the fact that the left-hand peaks in each pair yields $\Delta E^{\ddagger}_{HH}$ values that, on average, are higher than those in the right-hand side peaks, these results show that the OH (and SH) ligands are increasing the $\Delta E^{\ddagger}_{HH}$ barrier, in line with the fingerprint analysis shown in **Figure S6** (panel D).

Architectures and features used in the neural networks

Table S1. Neural networks in **Experiment 1** based on different hyperparameter sets and trained on MAD3 features, 20% and 80% test data, rmsprop optimizer, relu activation and adaptive learning rate.

ID	Training fraction	HL1	HL2	HL3	HL4	L2 reg.	Dropout	patience
NN1	80	801	-	-	-	0.000178	1.527922	20
NN2	80	459	78	-	-	0.000123	0.000391	47
NN3	80	954	374	128	-	1e-06	1e-08	45
NN4	80	883	84	136	31	1e-06	5.281509	24
NN5	20	232	-	-	-	0.002843	2.196804	27
NN6	20	380	151	-	-	0.000770	1.503655	30
NN7	20	91	369	140	-	0.1	1e-08	23
NN8	20	584	94	41	20	0.000441	1.625533	29

Table S2. Results of the hyperparameter search from **Experiment 1**. *n* is the number of architectures trained by the Bayesian optimization algorithm used for hyperparameter optimization.

train:test:val		80:10:10			20:40:40	
# Layers	r^2	MAE	*n*	r^2	MAE	*n*
1	0.678780	1.819524	663	0.644204	1.973622	620
2	**0.785287**	1.516890	372	**0.710946**	1.765680	701
3	0.766591	**1.437768**	347	0.689518	1.856310	401
4	0.714288	1.555579	666	0.706799	**1.745859**	452

Table S3. Results of the hyperparameter search for **Experiment 2**. *n* is the number of architectures trained by the Bayesian optimization algorithm used for hyperparameter optimization. Lowest mean absolute errors (objective of the Bayesian optimization) are found for depth = 5. FA features perform significantly better than MA, MAD, and MD.

f. sets	FA			MA			MAD			MD		
depth	r^2	MAE	*n*	r^2	MAE	*n*	r^2	MAE	*n*	r^2	MAE	*n*
1	0.35	2.71	633	0.59	2.19	507	0.57	2.23	487	0.59	2.19	555
2	0.67	1.95	651	**0.71**	1.80	509	**0.72**	1.71	532	0.69	1.71	731
3	0.69	1.81	583	0.70	1.80	472	0.72	1.73	456	**0.70**	1.77	712
4	0.76	1.58	537	0.69	**1.78**	502	0.69	1.72	505	0.70	1.78	552
5	**0.78**	**1.51**	433	0.68	1.83	466	0.71	**1.67**	603	0.69	**1.76**	629

Table S4. Results of the hyperparameter search for **Experiment 3**. *n* is the number of neural network architectures trained by the Bayesian optimization algorithm used for hyperparameter optimization.

train:test:val	80:10:10			20:40:40		
# Layers	r^2	MAE	*n*	r^2	MAE	*n*
1	0.769	1.473	660	0.668	1.930	625
2	0.810	1.277	514	0.742	1.632	631
3	**0.853**	1.123	705	**0.764**	1.577	526
4	0.845	**1.121**	456	0.763	**1.575**	416

Feature correlations

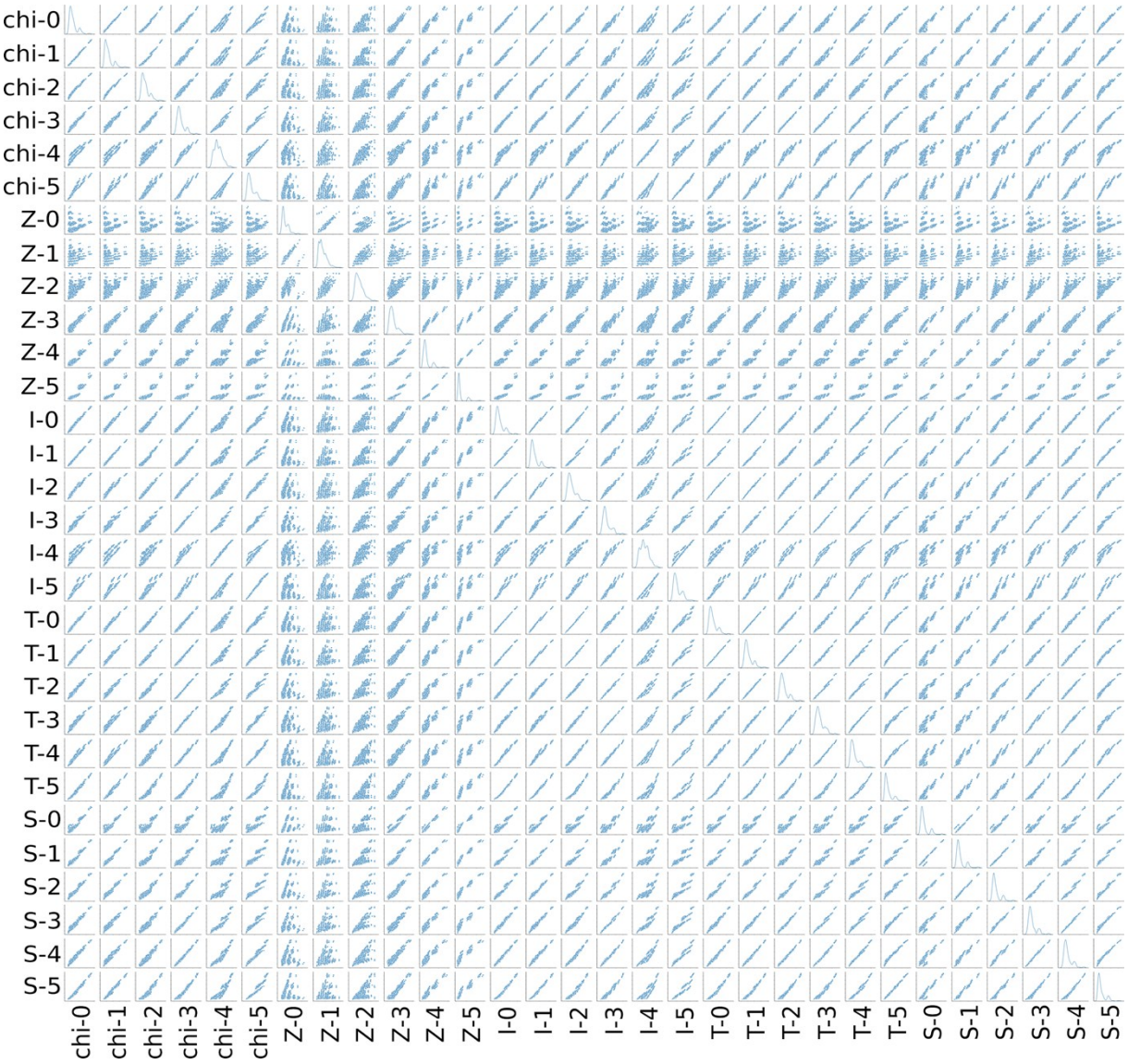

Figure S5. Correlation pairplot between all FA5 features (chi1-chi5, Z1-Z5, I1-I5, T1-T5, and S1-S5). All features apart from Z-features have high correlations with each other.