Supporting Information

S1 Parameters and energies

Table S1: Overview of the 23 production GA runs. μ refers to the target mean number of atoms during evolution of molecules. From run number 10 it was decided to evolve substituents of smaller size.

Scoring	Generations	Population	Mating pool	Mutation rate	nconfs	μ	STD(μ)
(1)	50	50	30	0.5	4	100	30
(1)	50	50	30	0.5	4	100	30
(1)	50	50	30	0.5	4	150	70
(1)	50	50	30	0.5	4	200	50
(1)	50	50	30	0.5	4	180	100
(1)	50	50	30	0.5	4	200	150
(1)	50	50	30	0.5	4	200	150
(3)	50	50	30	0.5	4	200	150
(3)	50	50	30	0.5	4	200	150
(2)	50	50	30	0.5	4	200	150
(1)	50	50	30	0.5	4	15	5
(1)	50	50	30	0.5	4	15	5
(1)	50	50	30	0.5	4	12	2
(1)	50	50	30	0.5	4	12	3
(2)	50	50	30	0.5	4	12	3
(1)	50	50	30	0.5	3	12	3
(2)	50	50	30	0.5	3	12	3
(2)	50	50	30	0.5	4	12	3
(2)	50	50	30	0.5	4	12	3
(2)	50	50	30	0.5	3	12	3
(2)	50	50	30	0.5	4	12	3
(2)	50	50	30	0.5	4	12	3
(2)	50	50	30	0.5	4	12	3

Table S2: Overview of computational details for DFT calculations.

	DFT Functional	
	PBE	B3LYP
Basis set	ZORA-def2-TZVP	ZORA-def2-TZVP
Mo specific basis set	SARC-ZORA-TZVP	SARC-ZORA-TZVP
Aux basis set	SARC/J	SARC/J
Aux acceleration	Split-RI-J	RIJCOSX
SCF Acceleration method	KDIIS / SOSCF	KDIIS / SOSCF
Calculation type	Optimization	Singlepoint
Dispersion	D3BJ	D3BJ
Solvent model	CPCM($\epsilon = 1.844$)	CPCM($\epsilon = 1.844$)

Table S3: RDKIT embedding default model parameters. Other possible embedding parameters was set to RDKIT defaults. *ignoreSmoothingFailures* was changed to True if an embedding failed. This setting removed the majority of embedding errors.

Keyword	Default value
maxAttempts	10
pruneRmsThresh	0.25
numCons	4
numThreads	1
useRandomCoords	True
ignoreSmoothingFailures	False

Table S4: Calculated reaction energies for all reaction steps in the Schrock cycle with corresponding thermal corrections. The *Label* column refers to the label used for the different reactions.

Reaction	Label	ΔE		Thermal correction(xTB)
		PBE-TZVP	B3LYP-TZVP	
$[Mo{-}N_2]$ + LH^+ ⟶ $[Mo{-}N_2H]^+$ + L	Mo_N2H+	6.48	14.62	-0.14
$[Mo{-}N_2H]^+$ + C ⟶ $[Mo{-}N_2H]$ + C^+	Mo_N2H	-24.58	-26.38	-0.15
$[Mo{-}N_2H]$ + LH^+ ⟶ $[Mo{-}N_2H_2]^+$ + L	Mo_N2H2+	-13.57	-22.14	-0.35
$[Mo{-}N_2H_2]^+$ + C ⟶ $[Mo{-}N_2H_2]$ + C^+	Mo_N2H2	10.96	0.82	-0.58
$[Mo{-}N_2H_2]$ + LH^+ ⟶ $[Mo{-}N_2H_3]^+$ + L	Mo_N2H3+	0.86	-0.66	-3.64
$[Mo{-}N_2H_3]^+$ + C ⟶ $[Mo{-}N]$ + NH_3 + C^+	Mo_N	-42.73	-51.48	-13.16
$[Mo{-}N]$ + LH^+ ⟶ $[Mo{-}NH]^+$ + L	Mo_NH+	-14.83	-12.69	2.27
$[Mo{-}NH]^+$ + C ⟶ $[Mo{-}NH]$ + C^+	Mo_NH	6.83	-9.41	-5.30
$[Mo{-}NH]$ + LH^+ ⟶ $[Mo{-}NH_2]^+$ + L	Mo_NH2+	-22.58	-19.39	1.96
$[Mo{-}NH_2]^+$ + C ⟶ $[Mo{-}NH_2]$ + C^+	Mo_NH2	-3.86	-21.05	-3.84
$[Mo{-}NH_2]$ + LH^+ ⟶ $[Mo{-}NH_3]^+$ + L	Mo_NH3+	-27.57	-30.97	3.24
$[Mo{-}NH_3]^+$ + C ⟶ $[Mo{-}NH_3]$ + C^+	Mo_NH3	10.47	1.52	-1.39
$[Mo{-}NH_3]$ + N_2 ⟶ $[Mo{-}NH_3N_2]$	Mo_N2_NH3	-13.81	-2.07	13.95
$[Mo{-}NH_3]$ ⟶ $[Mo]$ + NH_3	Mo	33.16	32.15	-11.51
$[Mo]$ + N_2 ⟶ $[Mo{-}N_2]$	Mo_N2_bare	-44.34	-23.67	10.89
$[Mo{-}NH_3N_2]$ ⟶ $[Mo{-}N_2]$ + NH_3	Mo_N2_exchange	2.63	10.55	-14.57
Cp^*_2Cr ⟶ $Cp^*_2Cr^+$		90.70	83.41	-0.09
$LutH^+$ ⟶ Lut		261.94	263.44	-8.11

Table S5: Calculated energies and spin states for all Schrock catalyst intermediates and additional reaction species. All values are in kcal/mol.

Intermediate	S	PBE-TZVP	B3LYP-TZVP	xTB thermal correction
[Mo]	$\frac{1}{2}$	-5488029.40	-5490352.92	1415.26
[Mo−N_2]	$\frac{1}{2}$	-5556819.30	-5559165.44	1418.07
[Mo−N_2H]$^+$	$\frac{1}{2}$	-5557074.75	-5559414.27	1426.05
[Mo−N_2H]	0	-5557190.02	-5559524.06	1425.98
[Mo−N_2H_2]$^+$	0	-5557465.54	-5559809.64	1433.75
[Mo−N_2H_2]	$\frac{1}{2}$	-5557545.28	-5559892.23	1433.25
[Mo−N_2H_3]$^+$	$\frac{1}{2}$	-5557806.35	-5560156.33	1437.73
[Mo−N]	0	-5522446.45	-5524771.87	1414.95
[Mo−NH]$^+$	$\frac{1}{2}$	-5522723.22	-5525048.01	1425.33
[Mo−NH]	0	-5522807.08	-5525140.82	1420.12
[Mo−NH_2]$^+$	$\frac{1}{2}$	-5523091.60	-5525423.65	1430.19
[Mo−NH_2]	1	-5523186.15	-5525528.12	1426.44
[Mo−NH_3]$^+$	1	-5523475.66	-5525822.53	1437.79
[Mo−NH_3]	$\frac{1}{2}$	-5523555.89	-5525904.42	1436.48
[Mo−NH_3−N_2]	$\frac{1}{2}$	-5592315.26	-5594695.35	1442.35
N2	0	-68745.56	-68788.85	-8.08
LuH+	0	-205371.32	-205519.50	75.82
Lu	0	-205109.38	-205256.05	67.70
CrCp2+	$\frac{3}{2}$	-1150972.06	-1151414.30	236.22
NH3	0	-35493.33	-35519.35	9.70
CrCp2	1	-1151062.76	-1151497.71	236.30

S2 Correlation plots

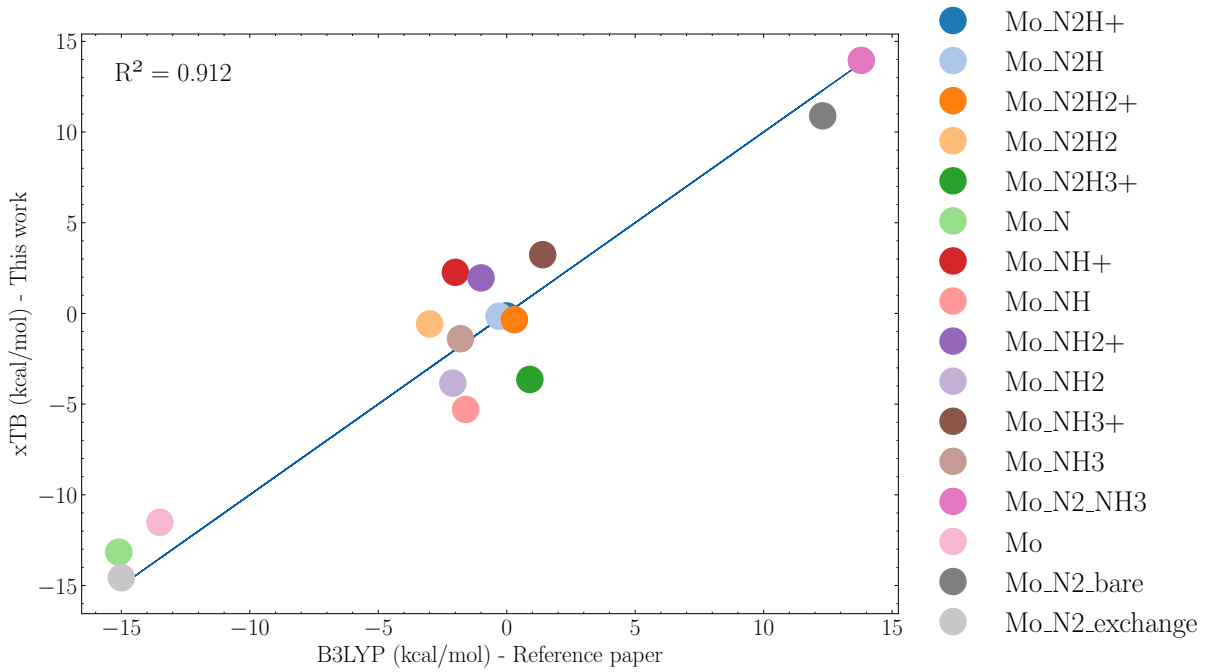


Figure S1: Correlation plot comparing the xTB thermal corrections obtained in this study (y-axis) with corresponding values obtained by Thimm et al. (x-axis). The corresponding reaction for each legend label is found in table S4.

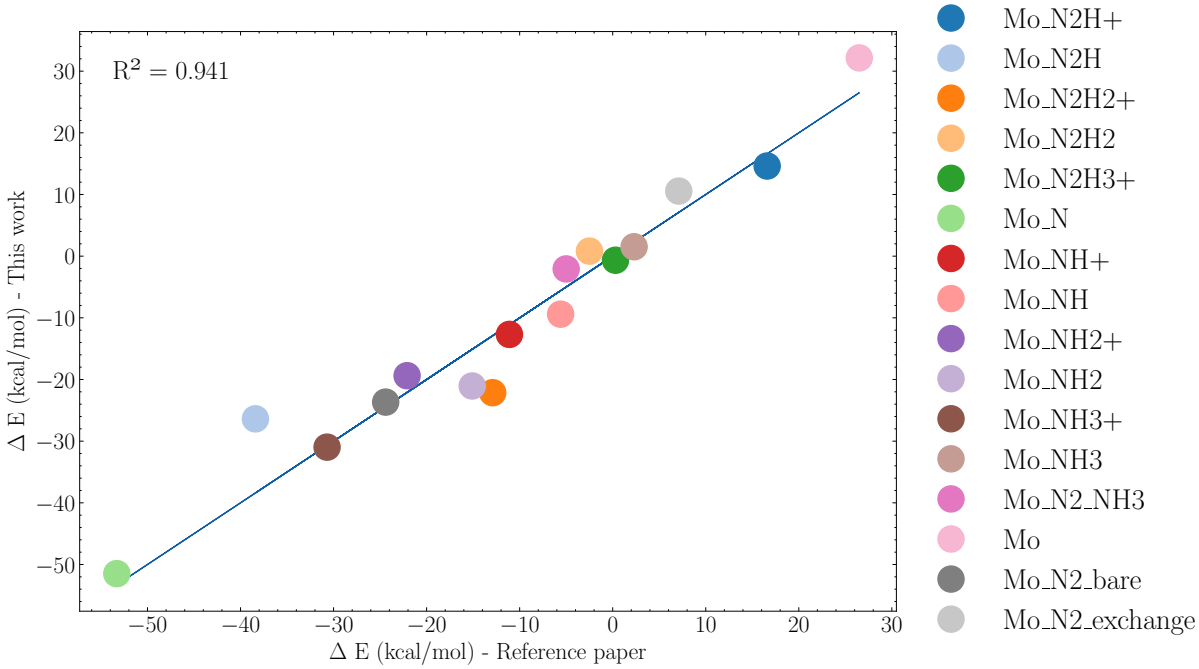


Figure S2: Correlation plot comparing the reaction values obtained in this study (y-axis) for PBE optimized intermediates followed by B3LYP singlepoints, with the corresponding values obtained by Thimm et al. (x-axis). The reaction values are based on electronic energies and the corresponding reaction for each legend label is found in table S4

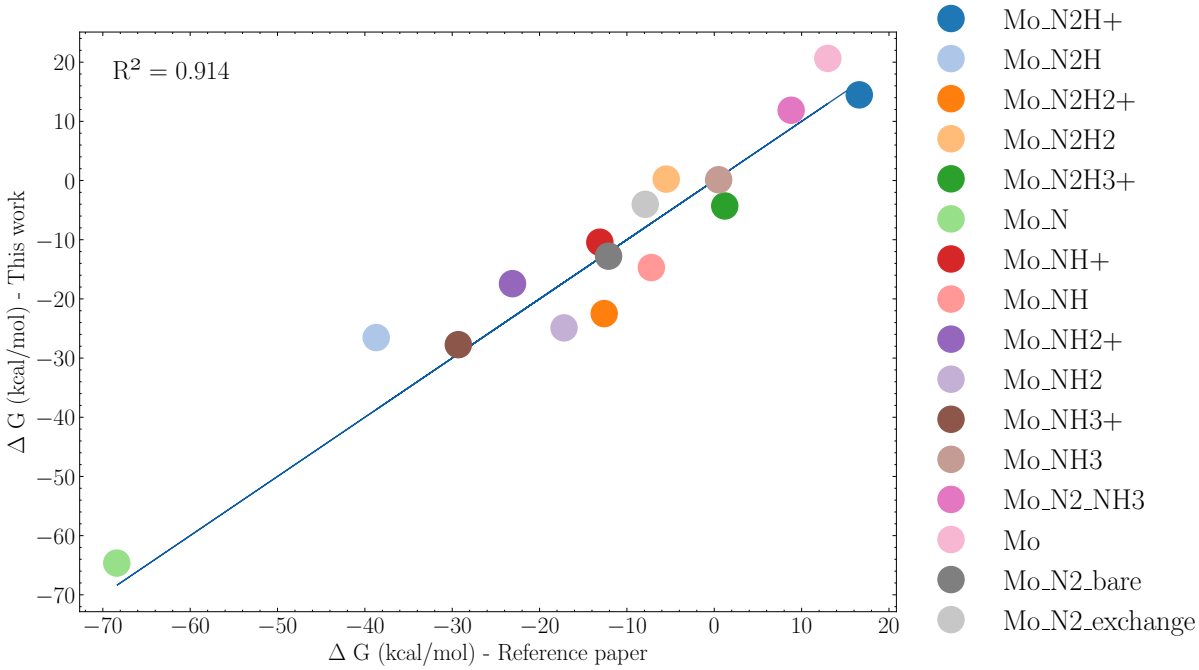


Figure S3: Correlation plot comparing the free energies obtained in this study (y-axis) for PBE optimized intermediates followed by B3LYP singlepoints and xTB vibrational corrections, with the corresponding free energies obtained by Thimm et al. (x-axis). The reaction values are based on free energies from the paper and electronic energies plus xTB thermal corrections for the values from this study. The corresponding reaction for each legend label is found in table S4.

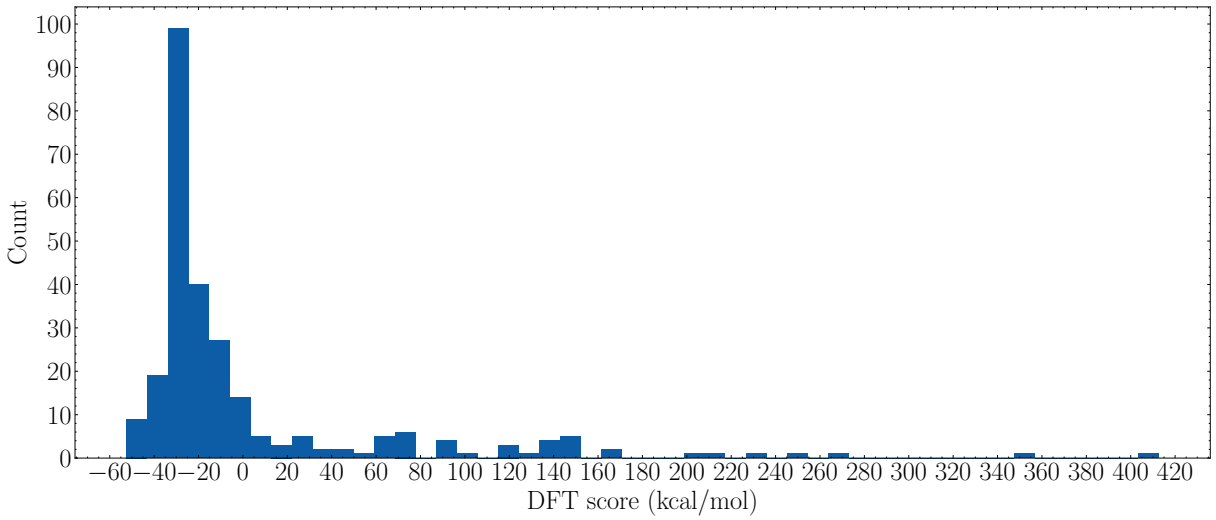


Figure S4: Distribution of singlepoint energies on 299 candidate molecules resulting from 23 GA runs.

S3 Improving GA

The GA went through several iterations and the results from the first iterations helped apply constraints that improved GA performance. Early iterations of the GA would heavily favor evolution of molecules with multiple amines. An examination of the results revealed that this was due to the fact that the amines on the substituents would form bonds to molybdenum for one of the intermediates in the scoring function but not the other. Which would lead to misleadingly low reaction energies and therefore high probabilities for mating pool selection. Molecules such as this was observed to have extremely poor reaction energies at the DFT level. An example of this is shown in figure S5. Figure S6 also shows an extreme case where the primary amines boosted the score. Therefore the GA was exploiting an artifact in the xTB level of theory to evolve molecules it would think were suitable, when in reality they were not. It was decided to restrict the GA ability to create such amine rich molecules. If a molecule after mating had more than one primary amine, these were replaced with hydrogens. This proved to significantly increase diversity in the final populations as the GA could no longer simply latch on to amine saturated molecules that would seemingly outperform any other molecules at the xTB level.

Another early issue was the possibility of a detachment of the N_xHy moieties during the xTB optimizations in the scoring functions. An example of this is shown in figure S7. This could also lead to artificially low scores. To prevent this, constrains were added to the Mo and N_xHy and a full relaxation within the GA was not performed. This would prevent the detachment of the moieties at the cost of only optimizing to a partially relaxed structure. This was deemed sufficient for the coarse screening of conformers within the GA. When the conformer search was performed in the DFT verification, the structure was allowed to fully relax in the scoring function.

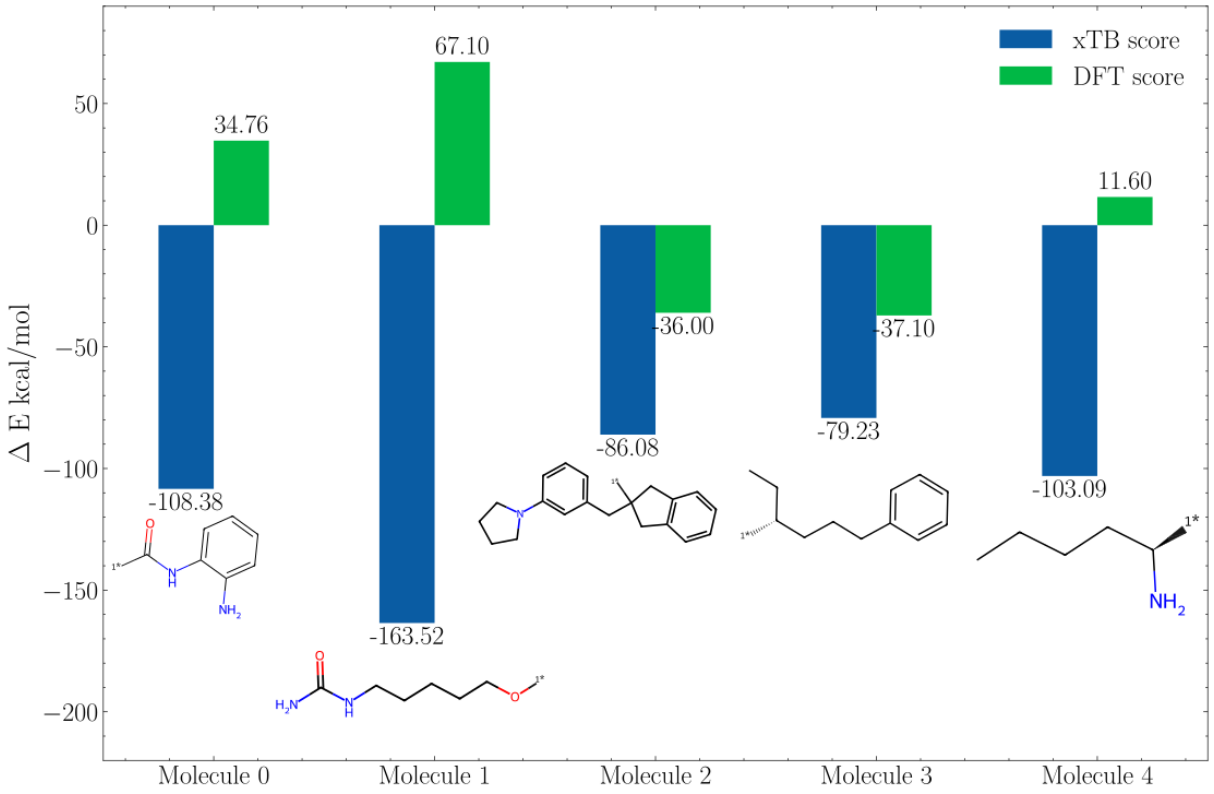


Figure S5: Comparison of DFT and xTB scores for a selection of five molecules from early iterations of the genetic algorithm for the $NH_3 \longrightarrow N_2$ scoring function. The molecules are placed next to their corresponding bars and the 1* marker denotes the core attachment point.

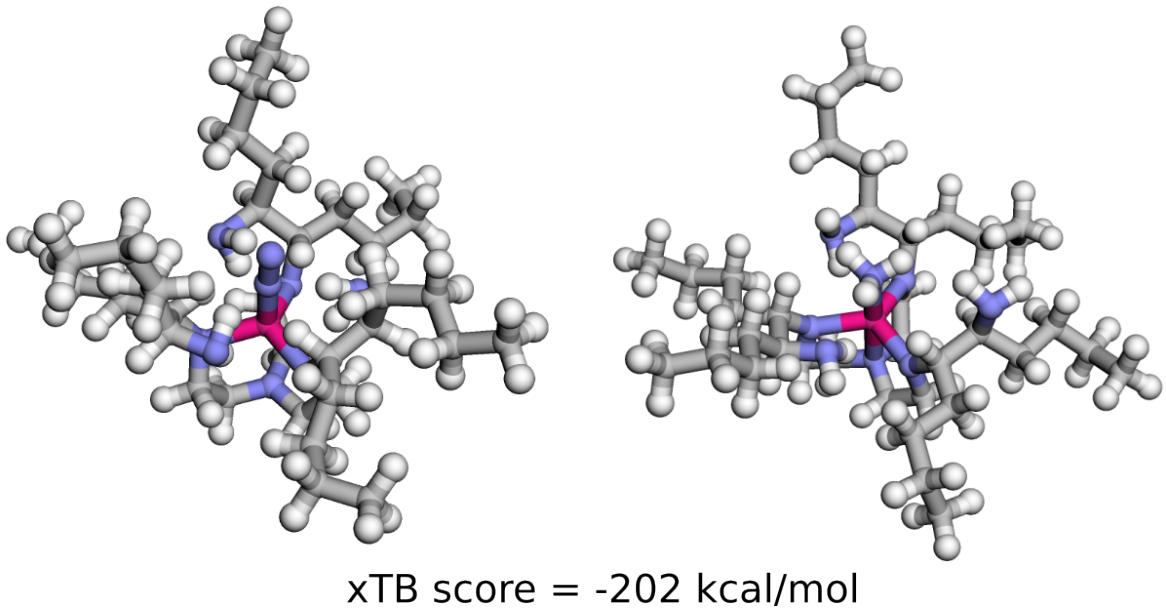


Figure S6: Example of a well scoring molecule form early iterations of the GA where content of primary amines was too high. The scoring was done on the [Mo-NH3] $\longrightarrow$ [Mo-N2] reaction step. The reaction energy appeared excessively low due to the approximate nature of xTB. This was not reflected at the PBE level.

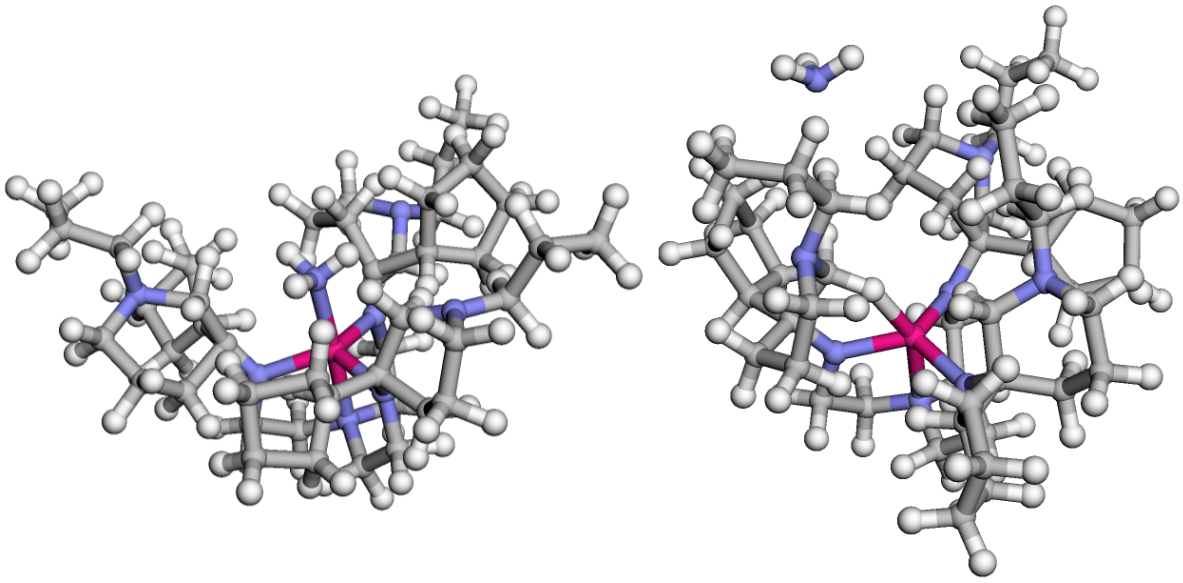


Figure S7: Example of NH_3 detaching from the [Mo] complex during the conformer search. The structure on the left is the starting conformer from the embedding and the structure on the right is the lowest energy conformer found for the [Mo−NH_3] intermediate out of 100 conformers. The four lowest energy conformers all had this detachment of the NH_3. The fifth lowest conformer had 4 kcal/mol higher energy with the ammonia still attached.

S3.1 xTB and DFT reaction energy comparison

Figure S8 shows the xTB reaction energies as compared to DFT reaction energies of the final 20 catalysts.

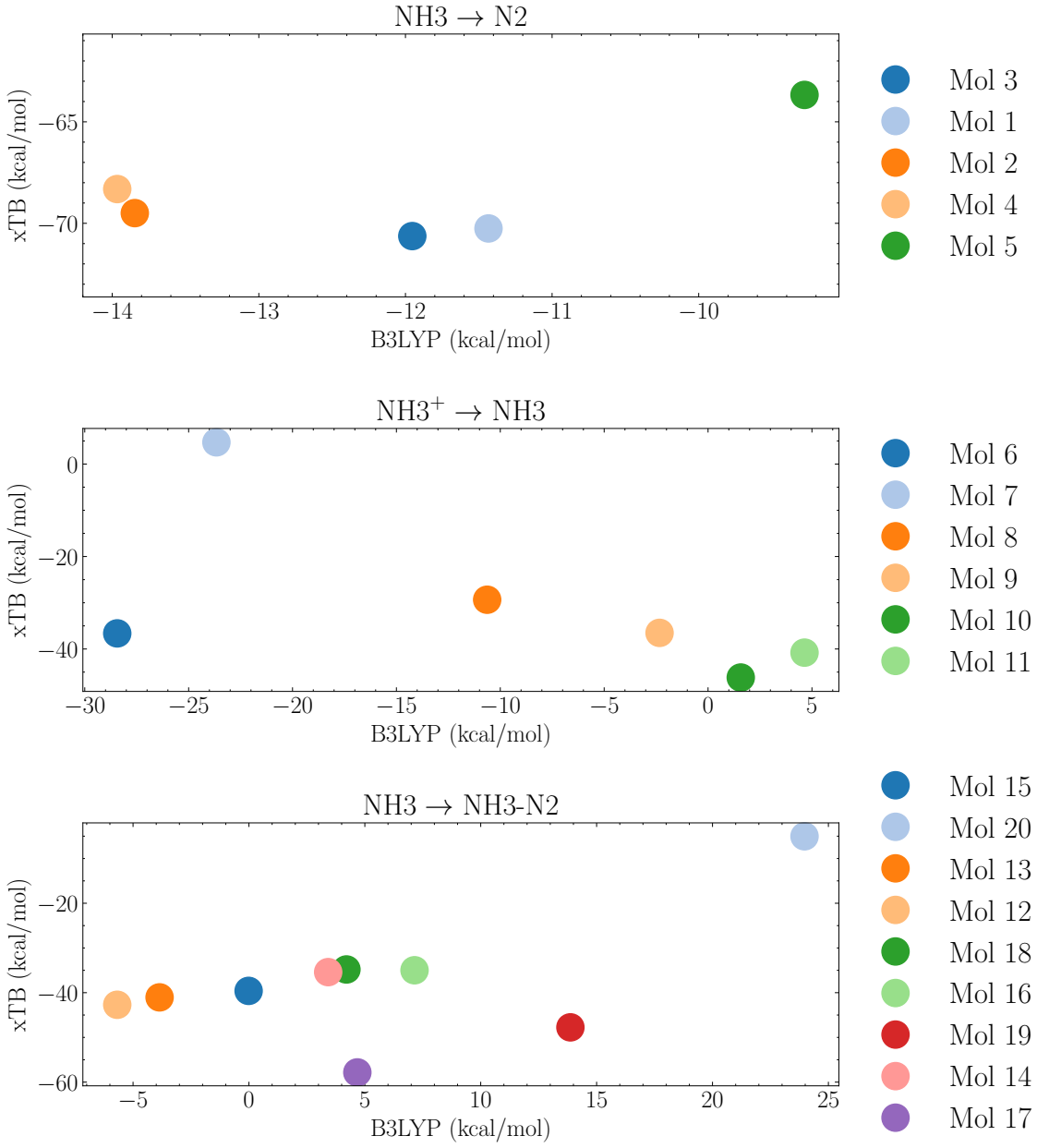


Figure S8: Comparison of scoring step electronic reaction energies (ΔE) for the final 20 molecules in table S6. The xTB reaction energies were obtained from the xTB vibrational calculations used for DFT reaction profiles. The DFT energies were obtained from the calculations used to obtain said reaction profiles. The title of each subplot indicate the reaction step the energy points correspond to.

S4 Final population of 43 molecules

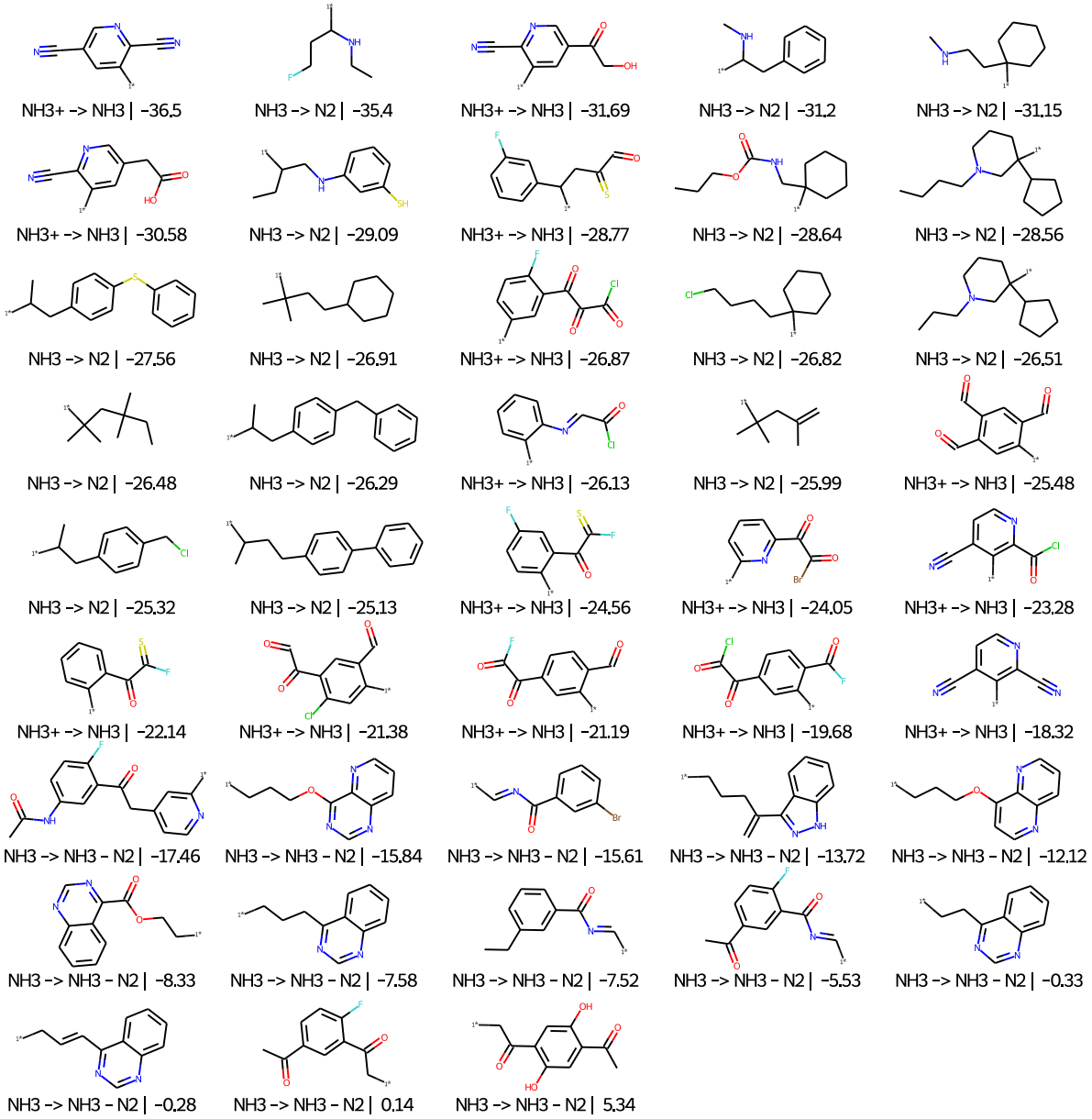


Figure S9: The 43 molecules left in the end of the DFT verification workflow. The label to the left of the | separator indicate the scoring function used when the molecules was scored in a GA run. The value to the right of the separator indicate the PBE evaluated energy (kcal/mol) on the optimized catalyst intermediates with the molecule acting as replacement for the HIPT substituent. The 1* on each molecule indicates the attachment point.

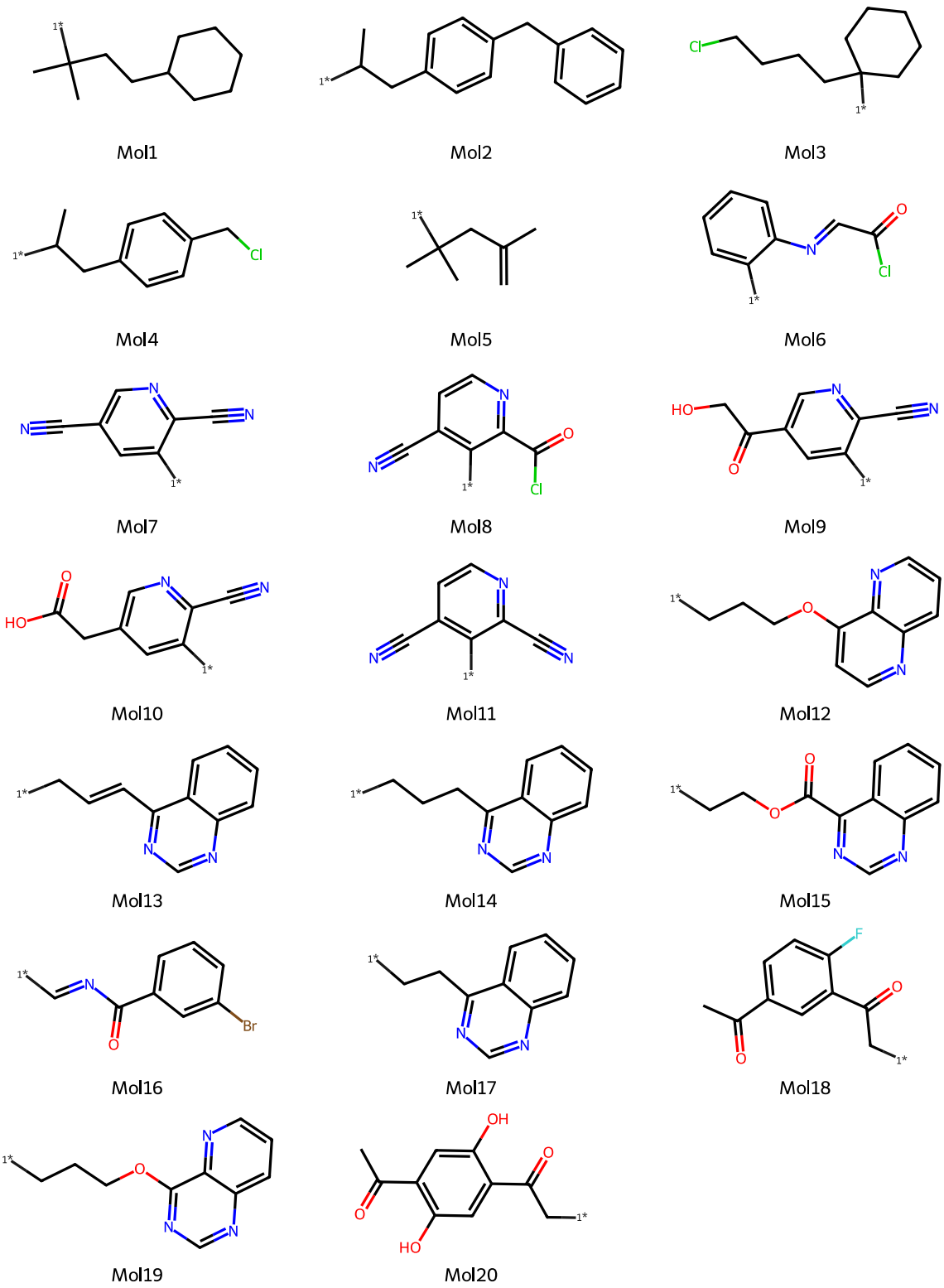


Figure S10: The 20 molecules for which full reaction profiles were obtained. Labels correspond to the ones found in tables 1 and S6.

S5 Supplementing reaction profiles

S5.1 PBE reaction profile for reference

The reaction profile for the PBE optimized structures are shown in figure S11. As mentioned in the main section, there were many problems with SCF convergence. Both for the GA proposed molecules, but also in particular for the reference Schrock catalyst structures. Many things were tried in order to obtain the relaxed structures which the energy profiles in S11 are based on. In order to converge the structures, the solution was to use a smaller steps size for internal coordinates in the geometry optimization. The orca keyword *MaxStep* had a default of 0.3 and this was set to 0.03 for the Schrock catalyst optimizations for all 15 intermediates. This severely increased computation time, but the SCF failure was avoided.

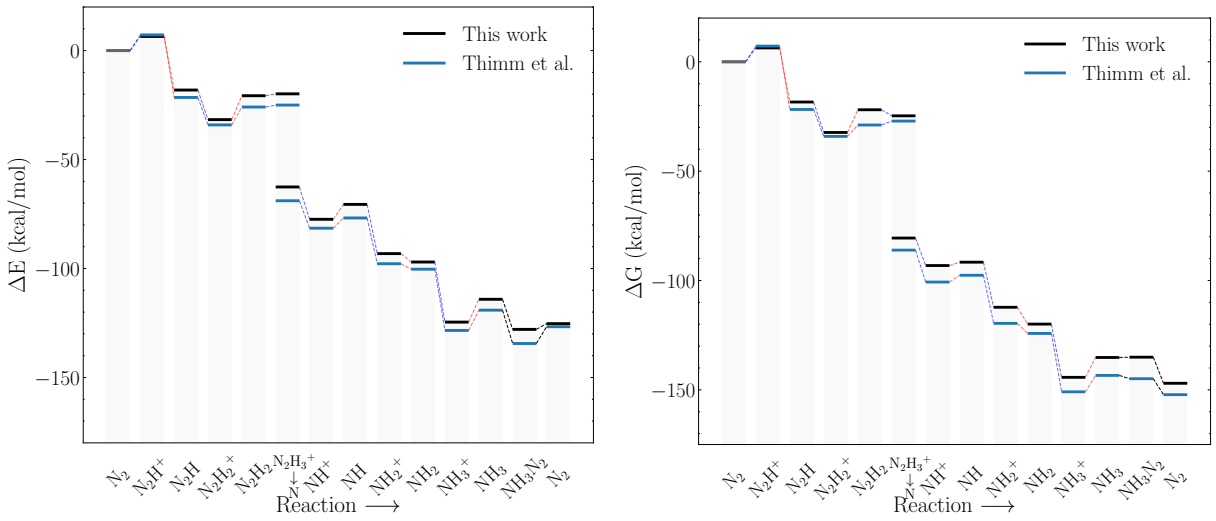
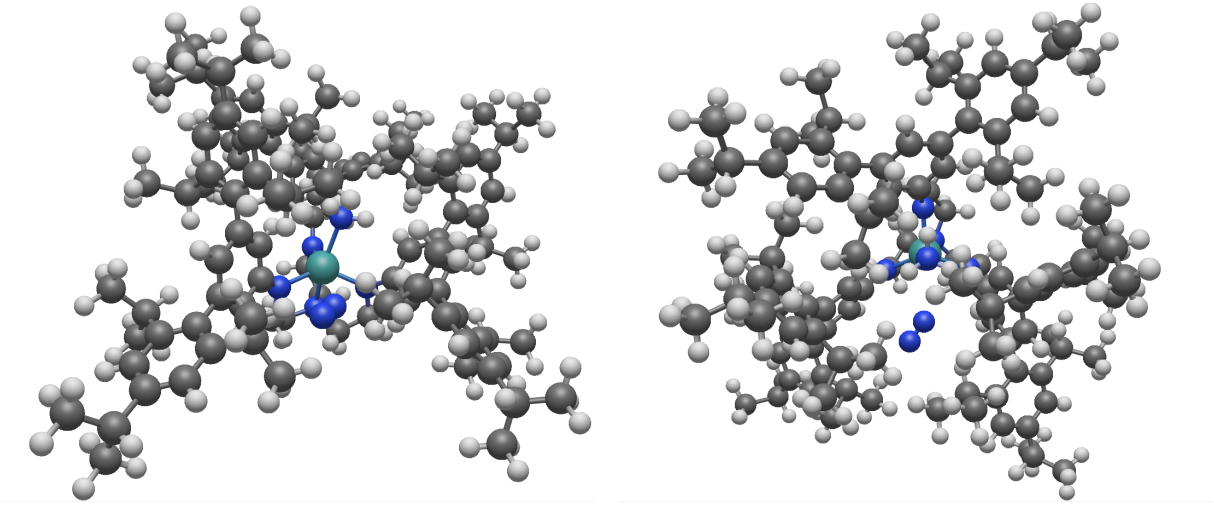


Figure S11: Reaction profiles for the Schrock catalyst calculated with PBE optimizations as compared to Thimm et al. S11a) Electronic energies, S11b) Free energies where the energies obtained in this work have been augmented with xTB vibrational corrections instead of DFT. Dotted blue lines indicate proton transfer and red lines indicate electron transfer.

S5.2 B3LYP reaction profiles

Figures S15, S16, S14, S13 contain the reaction profiles for the remaining molecules from table S6 that were not discussed in the main paper.



(a) **(b)**

Figure S12: PBE optimized Mo−N_2−NH_3 structure. Left) Side view, Right) Top view. N_2 was observed to have difficulty binding to the molybdenum and residing in nearby space. The distance from the nearest of the two nitrogen atoms to the molybdenum was measured to be 3.076 Å.

Table S6: Molecules with fully converged catalytic cycles at the end of the DFT verification. The energies in ΔE indicate the electronic energy difference of the scoring step used. They were obtained from PBE optimized structures. [1*] denotes the attachment point and all values are in kcal/mol. Molecules marked in bold were selected for further analysis in section 3.3.

Schrock catalyst:	$NH_3 \longrightarrow N_2$ (1) $NH_3^+ \longrightarrow NH_3$ (2) $NH_3 \longrightarrow NH_3{-}N_2$ (3)	7.86 0.13 11.87
SMILES	Scoring	ΔE
	$NH_3 \longrightarrow N_2$	
[1∗]**C(C)(C)CCC1CCCCC1**	**Mol1**	-11.43
[1∗]C(C)Cc1ccc(Cc2ccccc2)cc1	Mol2	-13.85
[1∗]C1(CCCCCl)CCCCC1	Mol3	-11.95
[1∗]C(C)Cc1ccc(CCl)cc1	Mol4	-13.97
[1∗]C(C)(C)CC(=C)C	Mol5	-9.28
	$NH_3^+ \longrightarrow NH_3$	
[1∗]c1ccccc1N=CC(=O)Cl	Mol6	-28.44
[1∗]c1cc(C#N)cnc1C#N	Mol7	-23.66
[1∗]**c1c(C#N)ccnc1C(=O)Cl**	**Mol8**	-10.63
[1∗]c1cc(C(=O)CO)cnc1C#N	Mol9	-2.33
[1∗]c1cc(CC(=O)O)cnc1C#N	Mol10	1.57
[1∗]c1c(C#N)ccnc1C#N	Mol11	4.64
	$NH_3 \longrightarrow NH_3 - N_2$	
[1∗]**CCCOc1ccnc2cccnc12**	**Mol12**	-5.68
[1∗]CC=Cc1ncnc2ccccc12	Mol13	-3.86
[1∗]CCCc1ncnc2ccccc12	Mol14	3.42
[1∗]CCOC(=O)c1ncnc2ccccc12	Mol15	-0.01
[1∗]C=NC(=O)c1cccc(Br)c1	Mol16	7.14
[1∗]CCc1ncnc2ccccc12	Mol17	4.68
[1∗]CC(=O)c1cc(C(C)=O)ccc1F	Mol18	4.21
[1∗]CCCOc1ncnc2cccnc12	Mol19	13.87
[1∗]CC(=O)c1cc(O)c(C(C)=O)cc1O	Mol20	23.97

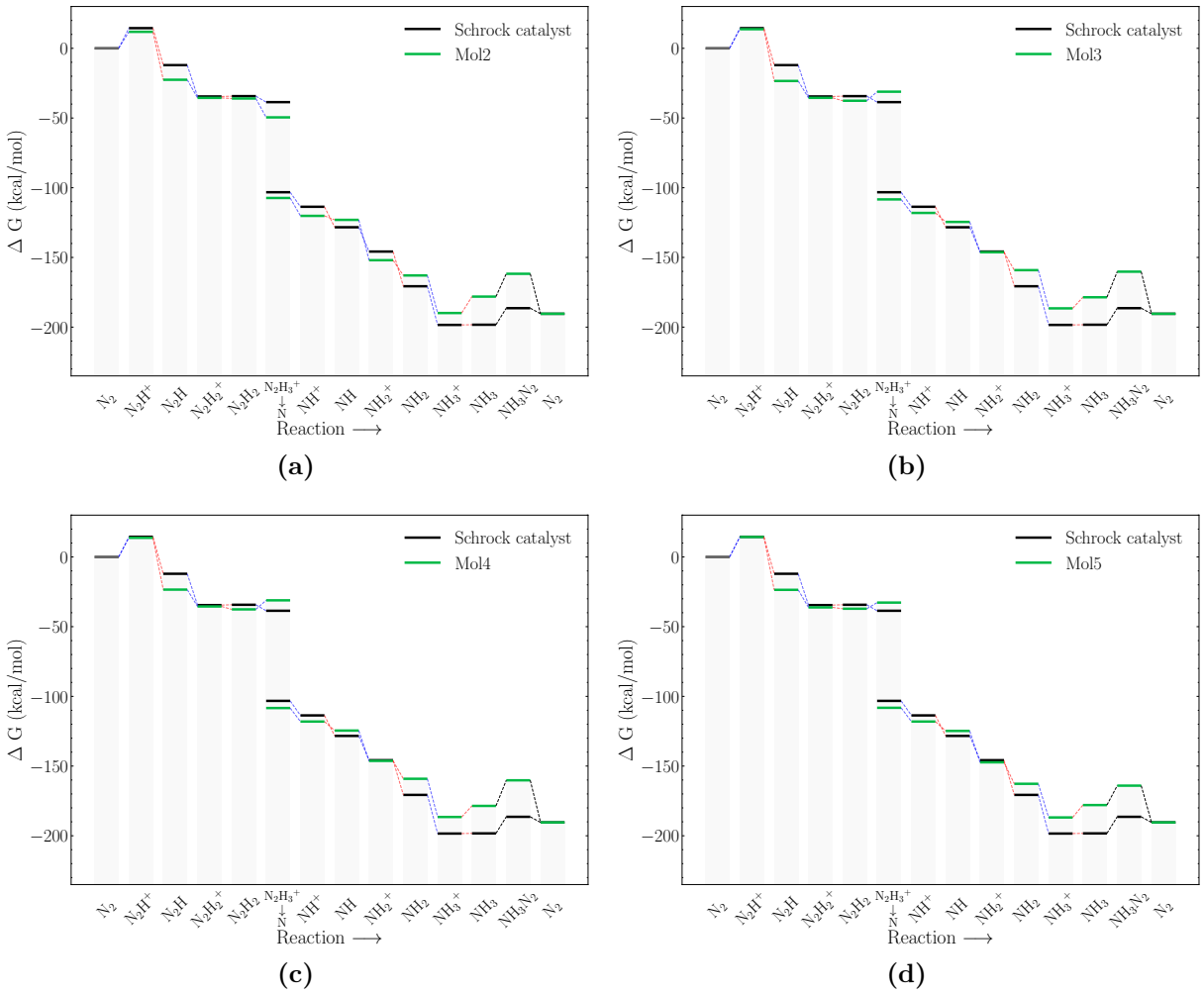


Figure S13: Reaction profiles for Mol(2-5).

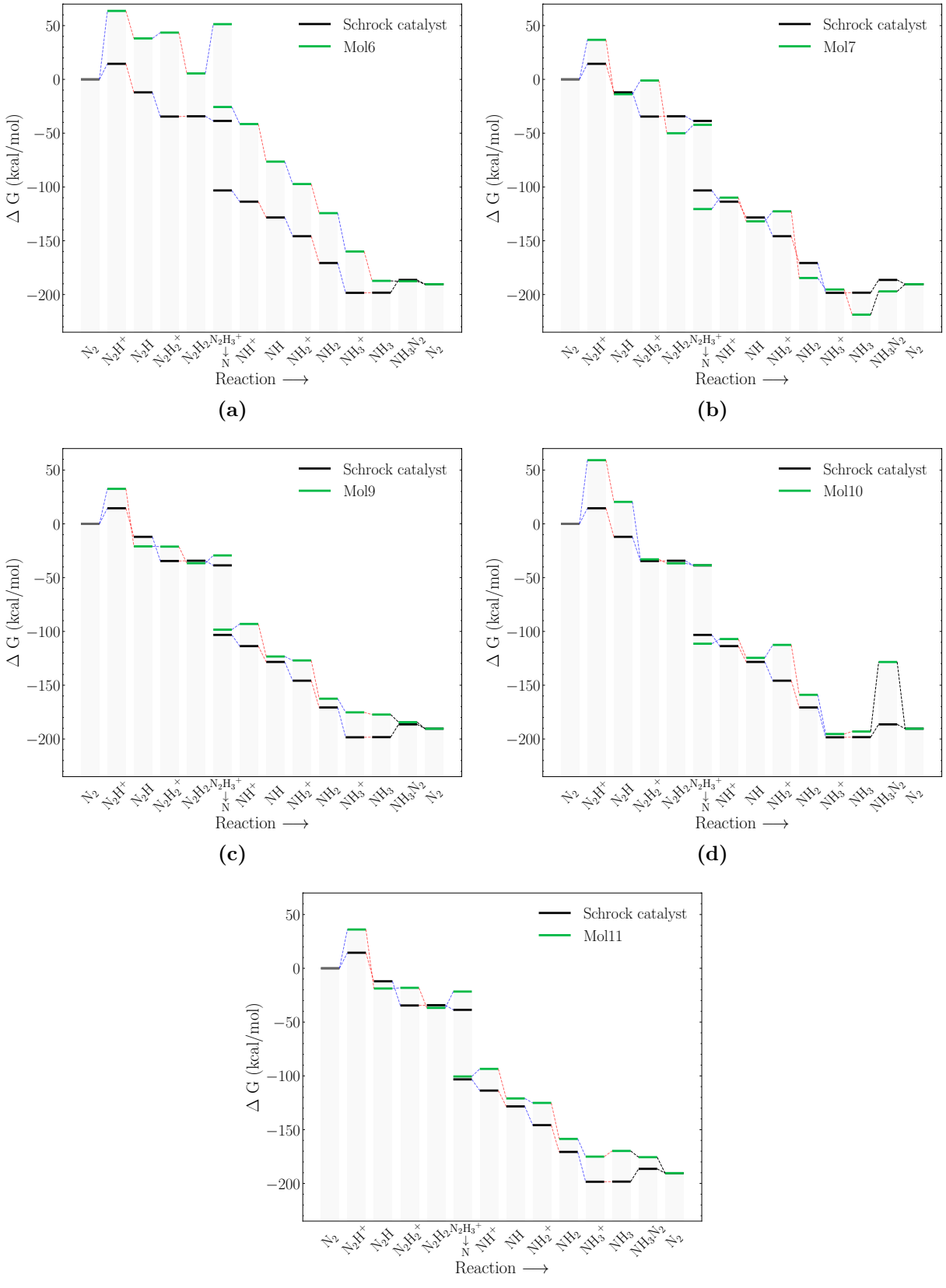


Figure S14: Reaction profiles for Mol(6-11), excluding Mol8 discussed in the main section.

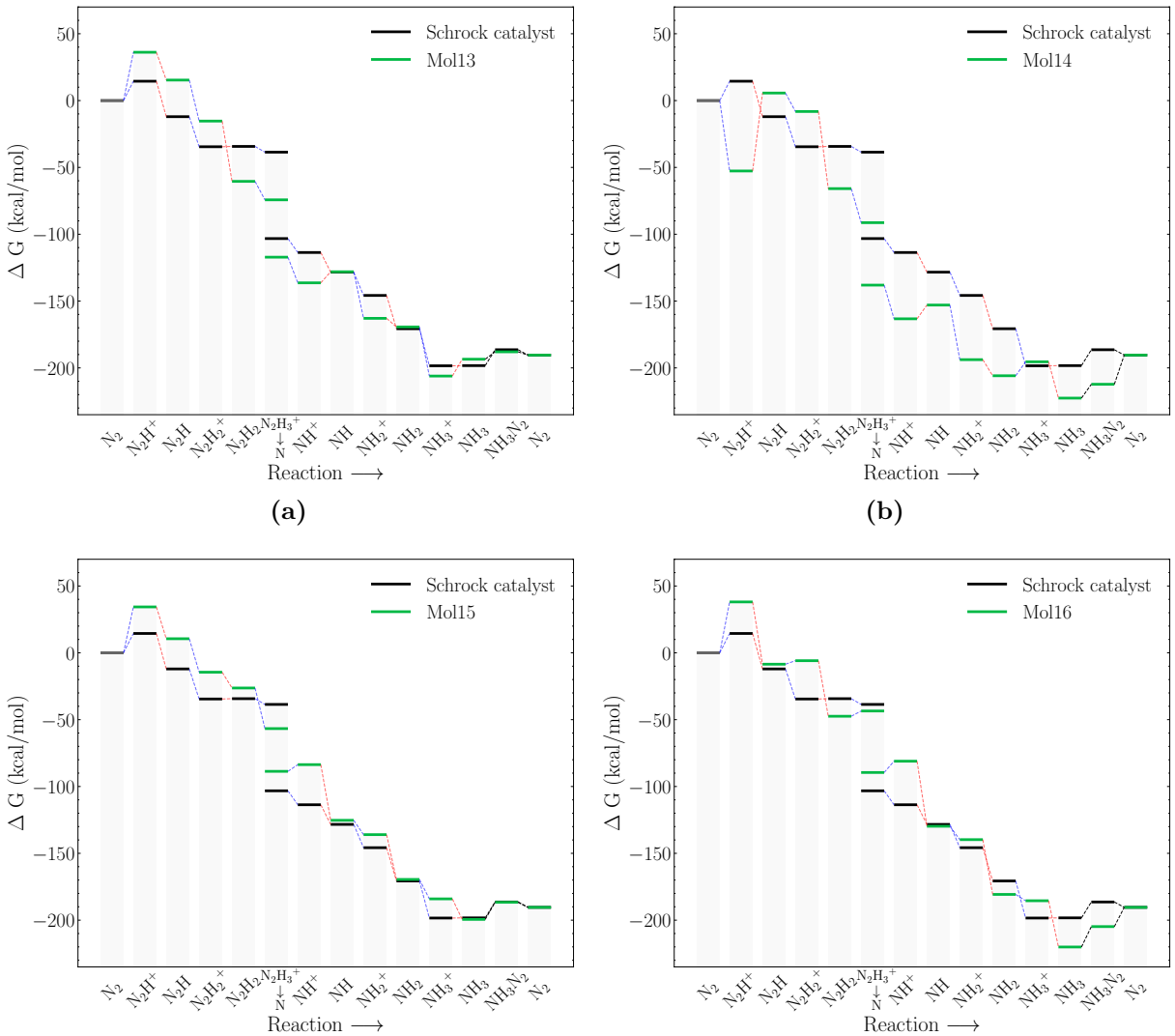


Figure S15: Reaction profiles for Mol(13-16).

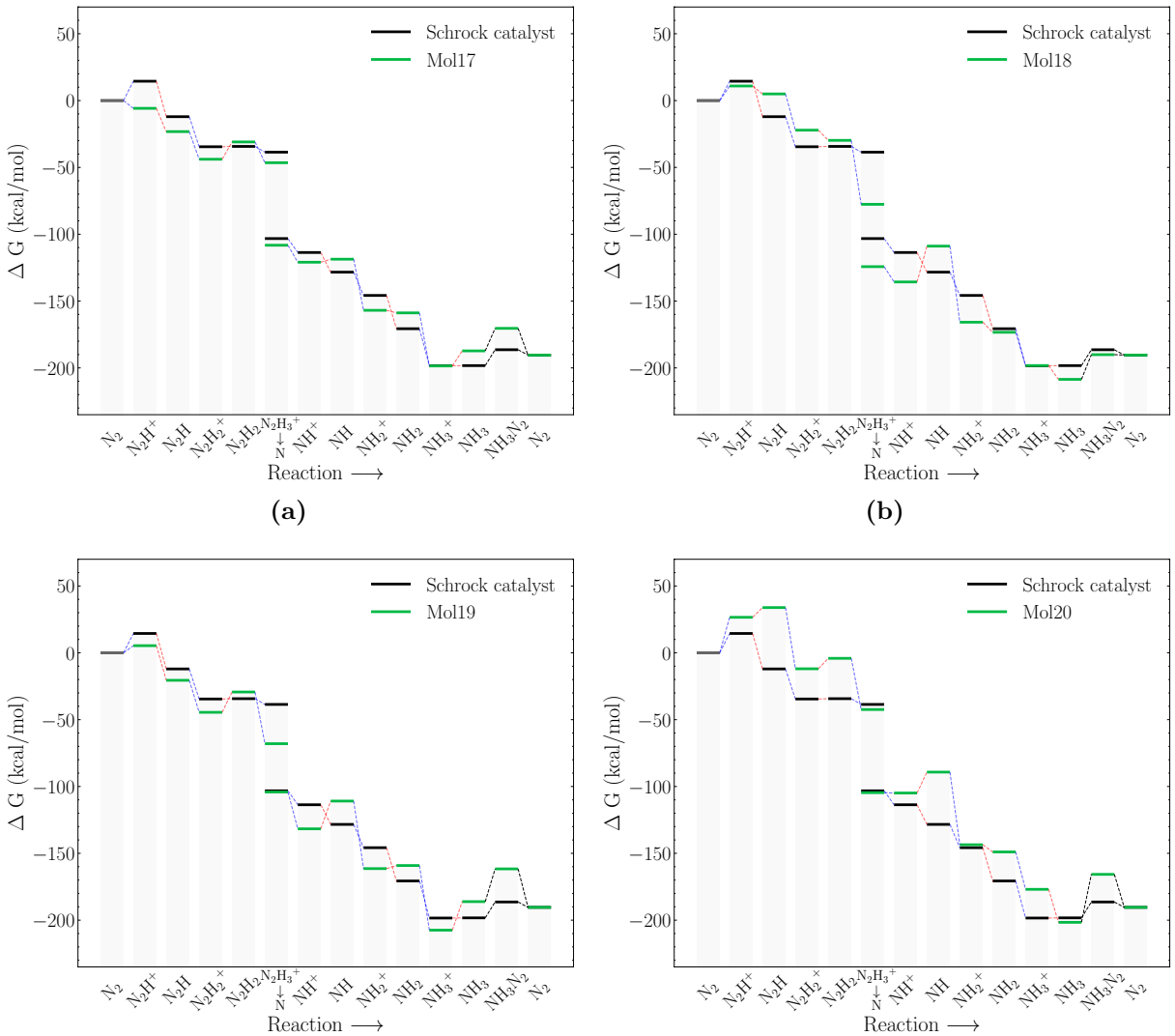


Figure S16: Reaction profiles for Mol(17-20).