Descriptor - property relationships in heterogeneous catalysis: exploiting synergies between statistics and fundamental kinetic modelling

Supporting Information

Laura Pirro[†], Pedro S. F. Mendes[†], Stijn Paret[†], Bart D. Vandegehuchte[‡], Guy B. Marin[†], Joris W. Thybaut[†,]*

† Laboratory for Chemical Technology, Ghent University, Technologiepark 125, B-9052 Ghent, Belgium

‡ Total Research & Technology Feluy, Zone Industrielle Feluy C B-7181, Seneffe, Belgium

*joris.thybaut@ugent.be

S1. OCM fundamental kinetic model and catalyst descriptor summary

Table S1: Gas-phase reaction network; kinetic parameters are reported by Chen et al.[1]

Primary initiation	**Dehydrogenation of C_2-C_3**
$CH_4 + O_2 \rightleftharpoons CH_3\bullet + HO_2\bullet$	$C_2H_6 + H\bullet \rightleftharpoons C_2H_5\bullet + H_2$
$CH_3\bullet$ generation	$C_2H_6 + OH\bullet \rightleftharpoons C_2H_5\bullet + H_2O$
$CH_4 + H\bullet \rightleftharpoons CH_3\bullet + H_2$	$C_2H_6 + CH_3\bullet \rightleftharpoons C_2H_5\bullet + CH_4$
$CH_4 + O\bullet \rightleftharpoons CH_3\bullet + OH\bullet$	$C_2H_5\bullet + M \rightleftharpoons C_2H_4 + H\bullet + M$
$CH_4 + OH\bullet \rightleftharpoons CH_3\bullet + H_2O$	$C_2H_5\bullet + O_2 \rightleftharpoons C_2H_4 + HO_2\bullet$
$CH_4 + HO_2\bullet \rightleftharpoons CH_3\bullet + H_2O_2$	$C_2H_4 + O_2 \rightleftharpoons C_2H_3\bullet + HO_2\bullet$
$CH_3\bullet$ oxidation	$C_2H_4 + H\bullet \rightleftharpoons C_2H_3\bullet + H_2$
$CH_3\bullet + O_2 \rightleftharpoons CH_3O\bullet + O\bullet$	$C_2H_4 + OH\bullet \rightleftharpoons C_2H_3\bullet + H_2O$
$CH_3\bullet + O_2 \rightleftharpoons CH_2O + OH\bullet$	$C_2H_4 + CH_3\bullet \rightleftharpoons C_2H_3\bullet + CH_4$
$CH_3\bullet + HO_2\bullet \rightleftharpoons CH_3O\bullet + OH\bullet$	$C_2H_3\bullet + M \rightleftharpoons C_2H_2 + H\bullet + M$
Coupling Reactions	$C_2H_3\bullet + O_2 \rightleftharpoons C_2H_2 + HO_2\bullet$
$CH_3\bullet + CH_3\bullet + M \rightleftharpoons C_2H_6 + M$	$C_3H_8 + H\bullet \rightleftharpoons C_3H_7\bullet + H_2$
$C_2H_5\bullet + CH_3\bullet + M \rightleftharpoons C_3H_8 + M$	$C_3H_7\bullet + M \rightleftharpoons C_3H_6 + H\bullet + M$
$C_2H_4 + CH_3\bullet + M \rightleftharpoons C_3H_7\bullet + M$	$C_2H_6 \rightleftharpoons C_2H_5\bullet + H\bullet$
Oxidation of $CH_3O\bullet$ and CH_2O	**C_2 Oxidation**
$CH_3O\bullet + M \rightleftharpoons CH_2O + H\bullet + M$	$C_2H_5\bullet + HO_2\bullet \rightleftharpoons CH_3\bullet + CH_2O + OH\bullet$
$CH_2O + OH\bullet \rightleftharpoons CHO\bullet + H_2O$	$C_2H_4 + OH\bullet \rightleftharpoons CH_3\bullet + CH_2O$
$CH_2O + HO_2\bullet \rightleftharpoons CHO\bullet + H_2O_2$	$C_2H_3\bullet + O_2 \rightleftharpoons CH_2O + CHO\bullet$
$CH_2O + CH_3\bullet \rightleftharpoons CHO\bullet + CH_4$	**Hydrogen–oxygen reactions**
$CHO\bullet + M \rightleftharpoons CO + H\bullet + M$	$O_2 + H\bullet \rightleftharpoons OH\bullet + O\bullet$
$CHO\bullet + O_2 \rightleftharpoons CO + HO_2\bullet$	$O_2 + H\bullet + M \rightleftharpoons HO_2\bullet + M$
$CO + HO_2\bullet \rightleftharpoons CO_2 + OH\bullet$	$HO_2\bullet + HO_2\bullet \rightleftharpoons O_2 + H_2O_2$
	$H_2O_2 + M \rightleftharpoons OH\bullet + OH\bullet + M$

Table S2: Catalytic reaction network; Surface reactions[2] grouped by their function in terms of the desired product: positive catalytic function (leading to methane activation and C_{2+} products formation) in red, negative catalytic function (leading to CO_x products formation) in blue.

Oxygen activation	$CHO^* + O^* \rightleftharpoons CO^* + OH^*$
$O_2 + 2^* \rightleftharpoons 2O^*$	$CO^* + O^* \rightleftharpoons CO_2^* + ^*$
Radical generation	$CO + ^* \rightleftharpoons CO^*$
$CH_4 + O^* \rightleftharpoons CH_3\bullet + OH^*$	$C_2H_4 + O^* \rightleftharpoons C_2H_4O^*$
$C_2H_6 + O^* \rightleftharpoons C_2H_5\bullet + OH^*$	$C_2H_4O^* + O^* \rightleftharpoons C_2H_3O^* + OH^*$
Regeneration of active sites	$C_2H_3O^* + O^* \rightleftharpoons CH_2O^* + HCO^*$
$2OH^* \rightleftharpoons H_2O^* + O^*$	$CH_3O\bullet + O^* \rightleftharpoons CH_2O + OH^*$
$H_2O^* \rightleftharpoons H_2O + ^*$	$CH_2O + O^* \rightleftharpoons CHO\bullet + OH^*$
Dehydrogenation to ethylene	$CHO\bullet + O^* \rightleftharpoons CO + OH^*$
$C_2H_5\bullet + O^* \rightleftharpoons C_2H_4 + OH^*$	**Coverage of active site**
Radical quenching	$CO_2 + ^* \rightleftharpoons CO_2^*$
$HO_2\bullet + O^* \rightleftharpoons O_2 + OH^*$	**Generation of HO_2 radical**
$HO_2\bullet + ^* \rightleftharpoons OH\bullet + O^*$	$H_2O_2 + O^* \rightleftharpoons HO_2\bullet + OH^*$
Non-selective oxidation	**Consumption of active O***
$C_2H_4 + O^* \rightleftharpoons C_2H_3\bullet + OH^*$	$H_2 + O^* \rightleftharpoons H\bullet + OH^*$
$CH_3\bullet + O^* \rightleftharpoons CH_3O^*$	$OH\bullet + O^* \rightleftharpoons O\bullet + OH^*$
$CH_3O^* + O^* \rightleftharpoons CH_2O^* + OH^*$	$H_2O + O^* \rightleftharpoons OH\bullet + OH^*$
$CH_2O^* + O^* \rightleftharpoons CHO^* + OH^*$	

Table S3: Catalyst descriptors in the OCM fundamental kinetic model, together with the relevant feasibility ranges and the corresponding references for those ranges.

OCM catalyst descriptor	Lower	Upper	Units	References

		bound	bound		
D_1	Reaction enthalpy of H-atom abstraction from CH_4	10	140	kJ/mol	[3, 4]
D_2	Chemisorption enthalpy of O_2	30	300	kJ/mol	[5]
D_3	Chemisorption enthalpy of CH_2O	50	150	kJ/mol	[3, 4]
D_4	Chemisorption enthalpy of $CHO^{\cdot}$	100	300	kJ/mol	[3, 4]
D_5	Chemisorption enthalpy of CO	50	200	kJ/mol	[6, 7]
D_6	Chemisorption enthalpy of CO_2	50	200	kJ/mol	[8-11]
D_7	Chemisorption enthalpy of H_2O	20	150	kJ/mol	[12]
D_8	Chemisorption enthalpy of C_2H_4O	20	100	kJ/mol	[3, 4]
D_9	Chemisorption enthalpy of $C_2H_3O^{\cdot}$	20	200	kJ/mol	[3, 4]
D_{10}	Sticking coefficient of O_2	10^{-3}	1	-	[3, 4]
D_{11}	Sticking coefficient of $CH_3^{\cdot}$	10^{-8}	10^{-2}	-	[13]
D_{12}	Sticking coefficient of CO	10^{-7}	10^{-2}	-	[3, 4]
D_{13}	Sticking coefficient of CO_2	10^{-5}	10^{-1}	-	[3, 4]
D_{14}	Sticking coefficient of H_2O	10^{-5}	10^{-1}	-	[3, 4]
D_{15}	Sticking coefficient of C_2H_4	10^{-8}	10^{-3}	-	[3, 4]
D_{16}	Density of active sites	10^{-11}	10^{-8}	mol/cm²	[14-16]

Table S4: Catalyst descriptors in the OCM microkinetic model.

Descriptor	Unit	Definition	Impact in the kinetic model
D1	kJ/mol_{CH4}	Reaction enthalpy of H-atom abstraction of CH_4 that represents the methane activation occurring through the breaking of a C–H bond assisted by an adsorbed oxygen species, O*, and leading to a methyl radical [17]	Key parameter in the calculation of the activation energies of all the H-atom abstraction reactions via Eley-Rideal mechanism
D2 - D9	kJ/mol_i	Chemisorption enthalpies of O_2, CH_2O, $CHO\cdot$, CO, CO_2, H_2O, C_2H_4O, $C_2H_3O\cdot$, which are adsorbed from the gas phase onto the catalyst surface	Used, together with the standard enthalpy of formation of the corresponding gas species, to calculate the standard enthalpy of formation of surface species via thermodynamic consistency within a catalytic cycle
D10 - D15	-	Sticking probabilities of O_2, $CH_3\cdot$, CO, CO_2, H_2O, C_2H_4, representing the ratio of the number of the respective molecules or radicals actually adsorbing on a clean surface to the total number of them colliding with it [18]	Used to correct the maximum value of the pre-exponential factors, calculated via collision theory, of the adsorption steps, which are assumed non-activated
D16	mol/m^2	Density of active sites	Crucial parameter in calculating the reaction rates of the catalytic steps and, hence, affecting the methane conversion and C_2 selectivity in OCM in a significant way

Kinetic details on the descriptors which were found to be significant in the OCM application of the present methodology:

- D11 is the sticking coefficient of methyl radicals on the active site, i.e. the multiplier of the pre-exponential factor of the following reaction:

$CH_3\bullet + O^* \rightleftharpoons CH_3O^*$ Eq. (1)

D11 has a crucial kinetic role in steering the selectivity of the OCM reaction. In fact, a high sticking coefficient of methyl radicals (D11) favours $CH_3\cdot$ oxidation on the catalyst surface[19] (Eq. (1)) instead of $CH_3\cdot$ coupling in the gas phase:

$CH_3\bullet + CH_3\bullet + M \rightleftharpoons C_2H_6 + M$ Eq. (2)

(where M represents a third body).

- D1 is the enthalpy of the reaction of atomic hydrogen abstraction from CH_4 (kJ):

$CH_4 + O^* \rightleftharpoons CH_3\bullet + OH^*$ Eq. (3)

Obviously, low values of D1 favour methane activation[19].

- D16 is the density of active sites * (mol/m^2), which generate activated oxygen species on the catalyst surface:

$O_2 + 2^* \rightleftharpoons 2O^*$ Eq. (4)

This species is responsible for the desired methane activation (Eq. (3)) but also for undesired deep oxidation surface reactions, indicated by e.g. Eq. (1).

- D2 is the chemisorption enthalpy of O_2 (enthalpy of the reaction in Eq. (4)). In combination with D1, it determines the stability of the hydroxyl species on the catalyst surface. The latter affects significantly the methane activation and the C_2 selectivity, mainly via the formation of adsorbed atomic oxygen[20]:

$2OH^* \rightleftharpoons H_2O^* + O^*$ Eq. (5)

- D15 is sticking coefficient of ethylene on the catalyst surface, which is the multiplier of the pre-exponential factor of the following reaction:

$C_2H_4 + O^* \rightleftharpoons C_2H_4O^*$ Eq. (6)

A low value of D15 is, hence, desirable to avoid consecutive oxidation reactions and maximize the yield of the desired, intermediate product: C_2H_4.

S2. Discovery library of *virtual* OCM catalysts

In *Step 1* of the methodology, the Fast Flexible Filling (FFF)[21] DoE technique was applied.

This methodology is based on hierarchical clustering[22]: a large number of random points (N >> n) is generated and grouped into clusters using the Fast Ward algorithm[23], with n representing the desired number of design points (in the present work, the number of *virtual* catalysts in each library to be designed). Ward's minimum variance criterion tends to produce clusters with approximately the same number of observations, which is ideal for the design of numerical experiments. The centroid of each cluster is taken as a design point, and each design point is treated as being the representative of a region in the design space. Therefore, FFF designs satisfy the *minimax* criterion[24], which aims at minimizing the maximum distance between non-design points in the design space and their nearest design point. This is opposed to the *maximin* criterion[24], implemented in other space-filling design techniques such as sphere packing[25], which seeks to maximize the minimum distance between any pair of design points.

The design of experiments was implemented via the statistical software JMP® 13.2.1[26].

FFF was applied to sample *virtual* catalysts in the [0,1]^16 descriptors' space. Subsequently, a linear transformation was used to relate the [0,1] range to the selected value ranges for each descriptor reported in Table S3.

For descriptors 10 to 16, which have variability ranges spanning more than three order of magnitudes, the logarithm was considered, in order to reduce the skewness of the distributions.

A 2D visualization of the obtained 16-dimensional library is shown in Figure S1 below.

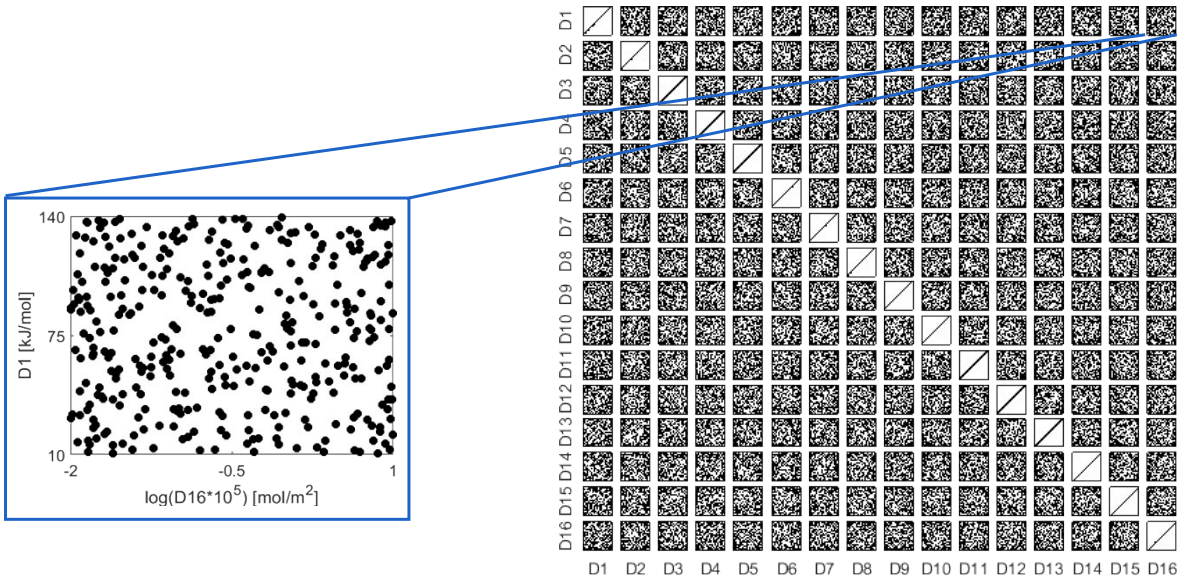

Figure S1. Scatterplot matrix representation of the 16-dimensional discovery library of virtual catalysts obtained in step 1 of the proposed methodology for the OCM application. Each of the 320 points in each square, such as the one in the zoom framed in blue, represents a virtual catalyst.

S3. Additional information about dataset 1

Table S5: Composition of the catalysts in the dataset of Kondratenko et al.[27] and experimentally determined performance.

Group	Catalysts	Composition [wt%]								Performance [-]		
		La	Mg	Sr	Ba	Li	Na	Cs	Mn	X_{CH_4}	S_{C_2}	Y_{C_2}
Pure La_2O_3	La	100								0.39	0.35	0.14
La-Sr	LaSrMn	90.8		9.1					0.1	0.38	0.37	0.14
	LaSrBa	90.8		9.1	0.1					0.39	0.42	0.16
	LaSrLi	90.8		9.1		0.1				0.41	0.42	0.17
	LaSrNa	90.8		9.1			0.1			0.39	0.41	0.16
	LaSrCs	90.8		9.1				0.1		0.39	0.38	0.15
La-Ba	LaBaMn	98.9			1.0				0.1	0.39	0.30	0.12
	LaBaLi	99.8			0.1	0.1				0.39	0.40	0.15
	LaBaNa	90.8			0.1		9.1			0.35	0.40	0.14
	LaBaCs	99.8			0.1			0.1		0.38	0.39	0.15
La-Mg	LaMgMn	90.8	9.1						0.1	0.35	0.29	0.10
	LaMgSr	83.8	8.3	8.3						0.39	0.38	0.15
	LaMgBa	90.1	9.0		0.9					0.40	0.40	0.16
	LaMgLi	90.8	9.1			0.1				0.39	0.36	0.14
	LaMgNa	90.8	9.1				0.1			0.39	0.36	0.14
	LaMgCs	90.1	9.0					0.9		0.38	0.34	0.13
La-Alkali	LaLiMn	83.3				8.3			8.3	0.35	0.31	0.11
	LaNaMn	90.1					9.0		0.9	0.38	0.37	0.14
	LaCsMn	83.3						8.3	8.3	0.34	0.26	0.09
	LaNaLi	83.3				8.3	8.3			0.38	0.34	0.13
	LaNaCs	90.8					9.1	0.1		0.39	0.37	0.15
	LaCsLi	90.8				9.1		0.1		0.40	0.37	0.15
Pure MgO	Mg		100							0.36	0.28	0.10
Mg-Sr	MgSrMn		83.3	8.3					8.3	0.25	0.04	0.01
	MgSrBa		83.3	8.3	8.3					0.40	0.42	0.17
	MgSrLi		83.3	8.3		8.3				0.28	0.37	0.10
	MgSrNa		83.3	8.3			8.3			0.39	0.40	0.16
	MgSrCs		83.3	8.3				8.3		0.41	0.41	0.17
Mg-Ba	MgBaMn		90.8		9.1				0.1	0.35	0.34	0.12
	MgBaLi		90.8		0.1	9.1				0.09	0.71	0.06
	MgBaNa		90.1		0.9		9.0			0.03	0.53	0.02
	MgBaCs		83.3		8.3			8.3		0.39	0.39	0.15
Mg-La	MgLaMn	9.1	90.8						0.1	0.34	0.26	0.09
	MgLaSr	8.3	83.3	8.3						0.38	0.37	0.14
	MgLaBa	9.0	90.1		0.9					0.37	0.34	0.13
	MgLaLi	8.3	83.3			8.3				0.37	0.28	0.10
	MgLaNa	8.3	83.3				8.3			0.41	0.39	0.16
	MgLaCs	8.3	83.3					8.3		0.37	0.32	0.12
Mg-Alkali	MgLiMn		90.8			9.1			0.1	0.17	0.57	0.10
	MgNaMn		90.8				9.1		0.1	0.33	0.03	0.01
	MgCsMn		90.8					9.1	0.1	0.33	0.28	0.09
	MgNaLi		90.8			9.1	0.1			0.09	0.70	0.06
	MgNaCs		90.8				0.1	9.1		0.37	0.30	0.11
	MgCsLi		90.8			9.1		0.1		0.11	0.69	0.07

Table S6: Process data concerning the dataset of Kondratenko et al.[27].

Variable	Experimental	Simulation
Reactor Type	Isothermal Fixed-Bed	
Operating Temperature (K)	1073	
Operating Pressure (bar)	1	
Feed Gas	CH_4 + Air	
CH_4/O_2 (mol/mol)	2	
Space Time (kg_{cat} s/$mol_{CH4,0}$)	71	
Catalyst/Diluent (SiC)	1:3	
Catalyst Particles Diameter (μm)	250-450	350

Table S7: Results of the Kruskal-Wallis test applied to the descriptors distributions in the four clusters obtained from the discovery library for the dataset of Kondratenko et al.[27]. A p value < 0.05 has been considered as threshold for significance: the null hypothesis of equal mean ranks among the distributions of the descriptor in different clusters can be rejected. The strength of the effect is defined according to the growing order of the H-statistics obtained during the test[28]. In red, the descriptors which have been considered significant for the targeting procedure.

Descriptor	p-value	Strength of the effect
D11	**<0.0001**	**115.9**
D1	**<0.0001**	**54.7**
D16	**<0.0001**	**26.2**
D15	**<0.0001**	**22.9**
D7	0.10	6.3
D8	0.34	3.3
D14	0.36	3.2
D6	0.44	2.7
D9	0.44	2.7
D10	0.52	2.3
D2	0.62	1.8
D3	0.68	1.5
D13	0.82	0.9
D5	0.88	0.7
D4	0.91	0.6
D12	0.92	0.5

Table S8: Results of the Mood's median test applied to the descriptors distributions in the four clusters obtained from the discovery library for the dataset of Kondratenko et al.[27]. A p value < 0.05 has been considered as threshold for significance: the null hypothesis of equal medians among the distributions of the descriptor in different clusters can be rejected. The strength of the effect is defined according to the growing order of the H-statistics obtained during the test[29]. The results are in good agreement with the ones reported in Table S7. In red, the descriptors which have been considered significant for the targeting procedure.

Descriptor	p-value	Strength of the effect
D11	**<0.0001**	**106.06**
D1	**<0.0001**	**46.0**
D15	**<0.0001**	**21.5**
D16	**0.0005**	**17.9**
D13	0.31	3.6
D6	0.34	3.3
D9	0.41	2.9
D12	0.43	2.8
D10	0.52	2.3
D2	0.60	1.9
D7	0.64	1.7
D14	0.64	1.7
D5	0.65	1.6
D3	0.67	1.5
D8	0.77	1.1
D4	0.95	0.4

S4. Additional information about dataset 2

Table S9. Characterization of the dataset of Kuś et al.[30] and experimentally determined performance for four different catalysts. The calcination atmosphere used in the catalyst synthesis has been altered for some catalysts. Surface basicity has been measured via temperature-programmed desorption of CO_2 and is reported as total µmol CO_2 desorbed per m² of catalyst surface area.

Oxide	Calcination	Basicity [µmol m^{-2}]	X_{CH_4} [-]	S_{C_2} [-]	Y_{C_2} [-]
La_2O_3	O_2 (a)	39.3	0.40	0.33	0.13
	He (b)	37.8	0.36	0.33	0.12
Nd_2O_3	O_2 (a)	17.2	0.38	0.31	0.12
	He (b)	19.0	0.36	0.30	0.11
ZrO_2	Air (a)	0.5	0.16	0.27	0.04
	He (b)	0.4	0.12	0.23	0.03
Nb_2O_5	Air	0.2	0.06	0.19	0.01

Table S10: Process data concerning the dataset of Kuś et al[30].

Variable	Experimental	Simulation
Reactor Type	Isothermal Fixed-Bed	
Operating Temperature (K)	1033	
Operating Pressure (bar)	1	
Feed Gas	CH_4 + Air	
CH_4/O_2 (mol/mol)	2	
Space Time (kg_{cat} s/$mol_{CH4,0}$)	1031	
Catalyst Particles Diameter (µm)	300-600	450

Table S11: Description of the clusters obtained in the performance-based comparison via k-means clustering applied to the dataset of Kuś et al.[30].

Cluster	Colour	Number *virtual* catalysts	Number *real* catalysts	X_{CH4} [%] (mean and standard deviation)	S_{C2} [%] (mean and standard deviation)	C_2 Yield [%] (mean and standard deviation)
C1	Red	37	0	8.9 ± 7.1	50.2 ± 13.3	4.6 ± 4.4
C2	Blue	67	0	55.5± 7.5	67.6± 7.5	37.9 ± 9.0
C3	Green	80	4	37.9± 6.3	37.1± 9.7	14.4 ± 5.2
C4	Yellow	136	3	22.9 ± 7.6	6.2 ± 8.2	1.4 ± 2.0

Table S12: Results of the Mann-Whitney test applied to the descriptors distributions in the four clusters obtained from the discovery library for the dataset of Kuś et al.[30]. A p value < 0.05 has been considered as threshold for significance: the null hypothesis of equal mean ranks among the distributions of the descriptor in different clusters can be rejected. The strength of the effect is defined according to the growing order of the χ^2 statistics obtained during the test[28]. In red, the descriptors which have been considered significant for the targeting procedure.

Descriptor	p-value	Strength of the effect
D1	**0.0002**	**14.3**
D2	**0.0031**	**8.7**
D15	0.08	3.1
D13	0.12	2.4
D4	0.23	1.5
D3	0.23	1.5
D9	0.31	1.0
D12	0.53	0.40
D11	0.59	0.29
D16	0.59	0.29
D14	0.63	0.23
D7	0.78	0.08
D6	0.85	0.04
D8	0.85	0.04
D10	0.90	0.02
D5	0.90	0.02

Table S13: Results of the Mood's median test applied to the descriptors distributions in the four clusters obtained from the discovery library for the dataset of Kuś et al.[30]. A p value < 0.05 has been considered as threshold for significance: the null hypothesis of equal medians among the distributions of the descriptor in different clusters can be rejected. The strength of the effect is defined according to the growing order of the χ^2 statistics obtained during the test[29]. The results are in good agreement with the ones reported in Table S11. In red, the descriptors which have been considered significant for the targeting procedure.

Descriptor	p-value	Strength of the effect
D1	**0.0021**	**9.5**
D2	**0.0279**	**4.8**
D14	0.19	1.7
D13	0.19	1.7
D4	0.19	1.7
D9	0.19	1.7
D15	0.66	0.19
D12	0.66	0.19
D11	0.66	0.19
D16	0.66	0.19
D3	0.66	0.19
D7	0.66	0.19
D6	0.66	0.19
D8	0.66	0.19
D10	0.66	0.19
D5	0.66	0.19

S5. Additional Information about dataset 3

Table S14. Data of Malekzadeh et al.[31], experimentally determined performance for $MO_x/Na_2WO_4/SiO_2$ catalysts and their electrical properties (electrical conductivity, semiconductor type and metal oxide band gap). The catalysts are ordered according to increasing electrical conductivity.

Catalyst	$\sigma \times 10^6$ [Ω^{-1}]	Semiconductor type	Band gap [eV]	X_{CH_4} [-]	S_{C_2} [-]	Y_{C_2} [-]
$V/Na_2WO_4/SiO_2$	0.1	n	2.1	0.10	0.12	0.01
$Zn/Na_2WO_4/SiO_2$	20	n	3.3	0.09	0.63	0.06
$Cr/Na_2WO_4/SiO_2$	80	n, p	1.9	0.10	0.24	0.02
$Fe/Na_2WO_4/SiO_2$	100	n	1.0	0.15	0.60	0.09
$Co/Na_2WO_4/SiO_2$	303	p	0.9	0.16	0.68	0.11
Na_2WO_4/SiO_2	333	-	-	0.11	0.63	0.07
$Mn/Na_2WO_4/SiO_2$	20000	p	0.3	0.20	0.80	0.16

Table S15: Process data concerning the dataset of Malekzadeh et al.[31].

Variable	Experimental	Simulation
Reactor Type	Isothermal Fixed-Bed	
Operating Temperature (K)	1048	
Operating Pressure (bar)	1	
Feed Gas	$CH_4 + O_2$	
CH_4/O_2 (mol/mol)	7.5	
Space Time (kg_{cat} s/$mol_{CH4,0}$)	5.2	
Catalyst Particles Diameter (µm)	-	300 (assumed)

Table S16: Description of the clusters obtained in the performance-based comparison via k-means clustering applied to the dataset of Malekzadeh et al.[31].

Cluster	Colour	Number *virtual* catalysts	Number *real* catalysts	X_{CH4} [%] (mean and standard deviation)	S_{C2} [%] (mean and standard deviation)	C_2 Yield [%] (mean and standard deviation)
C1	Red	127	0	2.7 ± 3.0	93.0 ± 4.8	2.5 ± 2.8
C2	Blue	48	1	22.1 ± 5.4	89.7 ± 6.6	20.0 ± 5.6
C3	Green	57	4	6.3 ± 6.0	57.9 ± 12.0	3.8 ± 3.8
C4	Yellow	88	2	3.1 ± 3.0	5.3 ± 8.4	0.2 ± 0.5

Table S17: Results of the Kruskal-Wallis test applied to the descriptors distributions in the four clusters obtained from the discovery library for the dataset of Malekzadeh et al.[31]. A p value < 0.05 has been considered as threshold for significance: the null hypothesis of equal mean ranks among the distributions of the descriptor in different clusters can be rejected. The strength of the effect is defined according to the growing order of the χ^2 statistics obtained during the test[28]. In red, the descriptors which have been considered significant for the targeting procedure.

Descriptor	p-value	Strength of the effect
D11	**<0.0001**	**71.9**
D1	**<0.0001**	**53.0**
D16	**0.0020**	**12.47**
D15	**0.0072**	**9.9**
D10	0.11	4.5
D4	0.23	2.9
D6	0.26	2.7
D3	0.33	2.2
D13	0.35	2.1
D5	0.47	1.5
D7	0.53	1.3
D8	0.53	1.3
D2	0.60	1.0
D9	0.73	0.6
D14	0.76	0.5
D12	0.85	0.3

Table S18: Results of the Mood's median test applied to the descriptors distributions in the four clusters obtained from the discovery library for the dataset of Malekzadeh et al.[31]. A p value < 0.05 has been considered as threshold for significance: the null hypothesis of equal medians among the distributions of the descriptor in different clusters can be rejected. The strength of the effect is defined according to the growing order of the χ^2 statistics obtained during the test[29]. The results are in good agreement with the ones reported in Table S15. In red, the descriptors which have been considered significant for the targeting procedure.

Descriptor	p-value	Strength of the effect
D11	**<0.0001**	**55.2**
D1	**<0.0001**	**38.9**
D15	**0.0018**	**12.6**
D16	**0.0031**	**11.5**
D4	0.20	3.2
D3	0.25	2.8
D10	0.31	2.3
D13	0.41	1.8
D6	0.42	1.7
D14	0.53	1.3
D12	0.58	1.1
D2	0.59	1.1
D7	0.64	0.9
D8	0.72	0.7
D9	0.80	0.4
D5	0.92	0.2

According to the results above, descriptors 15 and 16 were included in the targeting procedure.

However their impact in discriminating between the three targeted libraries resulted to be

minimal, as show in the figure below, where the box plots for each descriptor show a high overlap and no real trend can be identified.

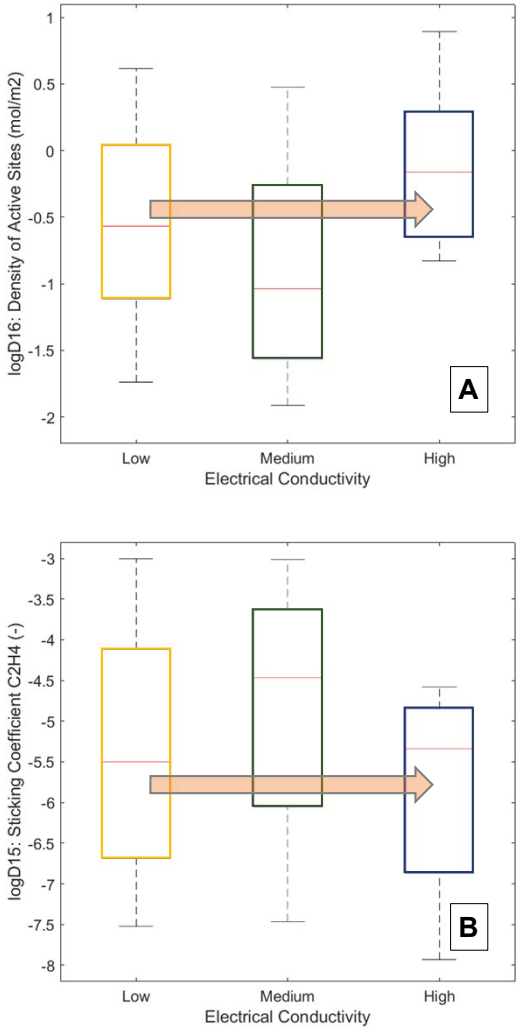

Figure S2. Comparison of the distributions of descriptors logD16: density of active sites (A) and logD15: sticking coefficient of C_2H_4 (B) in the virtual libraries which have been generated and tested in order to target the performances of real catalysts from the dataset of Malekzadeh et al.[31] with increasing (from C4 to C2) electrical conductivity.

References

1. Q. Chen, P. M. Couwenberg and G. B. Marin, *Catal. Today*, 1994, **21**, 309-319.
2. P. N. Kechagiopoulos, J. W. Thybaut and G. B. Marin, *Ind. Eng. Chem. Res.*, 2014, **53**, 1825-1840.
3. V. I. Alexiadis, M. Chaar, A. van Veen, M. Muhler, J. W. Thybaut and G. B. Marin, *Appl. Catal. B-Environ.,* 2016, **199**, 252-259.
4. J. W. Thybaut, J. Sun, L. Olivier, A. C. Van Veen, C. Mirodatos and G. B. Marin, *Catal. Today*, 2011, **159**, 29-36.
5. Y. P. Arnaud, *Appl. Surf. Sci.*, 1992, **62**, 37-45.
6. A. Zecchina, D. Scarano, S. Bordiga, G. Spoto and C. Lamberti, in *Advances in Catalysis*, Academic Press, 2001, vol. 46, pp. 265-397.
7. V. E. Ostrovskii, *Thermochim. Acta*, 2009, **489**, 5-21.
8. M. Casarin, D. Falcomer, A. Glisenti and A. Vittadini, *Inorg. Chem.*, 2003, **42**, 436-445.
9. C. Chu, Y. Zhao, S. Li and Y. Sun, *J. Phys. Chem. C*, 2016, **120**, 2737-2746.
10. A. Auroux and A. Gervasini, *J. Phys. Chem. US*, 1990, **94**, 6371-6379.
11. M. Cutrufello, I. Ferino, E. Rombi, V. Solinas, G. Colón and J. Navío, *J. Therm. Anal. Calorim.*, 2003, **72**, 223-229.
12. J. Majzlan, L. Mazeina and A. Navrotsky, *Geochim. Cosmochim. Ac.*, 2007, **71**, 615-623.
13. Y. Tong and J. H. Lunsford, *J. Am. Chem. Soc.*, 1991, **113**, 4741-4746.
14. S. Lacombe, C. Geantet and C. Mirodatos, *J. Catal.*, 1995, **151**, 439-452.
15. D. Wolf, M. Slinko, E. Kurkina and M. Baerns, *Appl. Catal. A-General*, 1998, **166**, 47-54.
16. H. Zanthoff, Z. Zhang, T. Grzybek, L. Lehmann and M. Baerns, *Catal. Today*, 1992, **13**, 469-480.
17. J. H. Lunsford, *Langmuir*, 1989, **5**, 12-16.
18. J. A. Dumesic, D. F. Rudd, L. M. Aparicio, J. E. Rekoske and A. A. Treviño, *ACS Professional Reference Book,* Washington, DC*,* 1993, 315.
19. J. Sun, J. W. Thybaut and G. B. Marin, *Catal. Today*, 2008, 137, 90-102.
20. V. I. Alexiadis, J. W. Thybaut, P. N. Kechagiopoulos, M. Chaar, A. C. van Veen, M. Muhler and G. B. Marin, *Appl. Catal. B-Environ.*, 2014, 150, 496-505.
21. R. Lekivetz and B. Jones, *Qual. Reliab. Eng. Int.*, 2015, **31**, 829-837.
22. S.C. Johnson, Psychometrika, *1967*, **32**, 241-254.
23. J.H. Ward, *J. Am. Stat. Assoc.*, 1963, **58**, 236-244.
24. M.E. Johnson, L.M. Moore and D. Ylvisaker, *J. Stat. Plan. Infer*., 1990, **26**, 131-148.
25. L. Pronzato and W.G. Müller, *Stat. Comput*., 2012, **22**, 681-701.
26. SAS, JMP 13.2.1, https://www.jmp.com.
27. E. V. Kondratenko, M. Schluter, M. Baerns, D. Linke and M. Holena, *Catal. Sci. Technol.*, 2015, **5**, 1668-1677.
28. W. H. Kruskal and W. A. Wallis, *J. Am. Stat. Assoc.*, 1952, **47**, 583-621.
29. A. M. Mood, *Ann. Math. Stat.*, 1954, **25**, 514-522.
30. S. Kuś, M. Otremba and M. Taniewski, *Fuel*, 2003, **82**, 1331-1338.
31. A. Malekzadeh, A. Khodadadi, M. Abedini, M. Amini, A. Bahramian and A. K. Dalai, *Catal. Commun.*, 2001, **2**, 241-247.