Supporting information for the article

Combinatorial Explosion in Homogeneous Catalysis: Screening 60,000 Cross-coupling reactions

Enrico Burello[a], David Farrusseng[b],* and Gadi Rothenberg[a],*

[a] van 't Hoff Institute of Molecular Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands. [b] Institut de Recherches sur la Catalyse, CNRS, 2 avenue Albert Einstein, F-69626, Villeurbanne Cedex, France. E-mail: farrusseng@catalyse.cnrs.fr, gadi@science.uva.nl

Descriptors used in this study

Table S1 shows the number and type of descriptors selected after the Relief and PCA variable selection process. Each reaction is characterized by a set of steric and electronic descriptors pertaining the chemicals involved in the Heck reaction (R_1, R_2, ligand and solvent) together with the experimental conditions (catalyst precursor, time, temperature and Pd loading). The Relief algorithm selects descriptors and experimental conditions which display a statistically significant correlation with the responses TON and TOF. The PCA analysis discards redundant descriptors by selecting a set of uncorrelated parameters. This variable selection process reduces the chances of over-learning and gives a simpler model which is easier to interpret.

Starting with 74 descriptors the Relief algorithm selects 26 and 33 descriptors for TON and TOF respectively. The principal component analysis finally reduces the dimensionality to 17 descriptors for TON and 20 for TOF.

Table S1. List of descriptors used by the Relief and the PCA algorithms

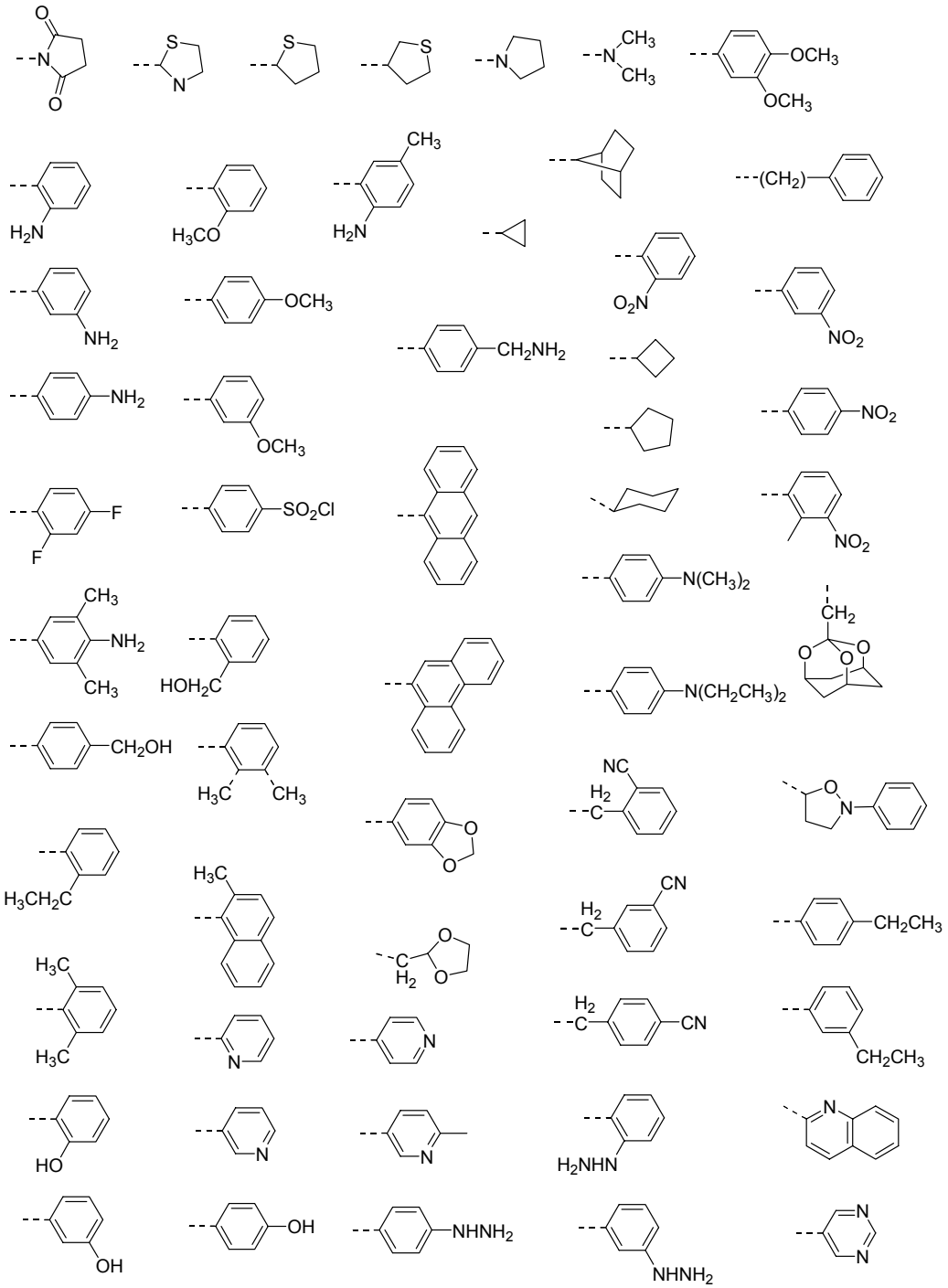
		TON		TOF	
Descriptors		**Relief**	**PCA**	**Relief**	**PCA**
R_1					
1	$Hammett_m$	HOMO	HOMO	$Hammett_m$	$Hammett_{p(+)}$
2	$Hammett_p$	LUMO	LUMO	$Hammett_p$	$Hammett_{p(-)}$
3	$Hammett_{p(+)}$	GAP	GAP	$Hammett_{p(+)}$	V
4	$Hammett_{p(-)}$	MW	S_{occ}	$Hammett_{p(-)}$	
5	HOMO	S		MW	
6	LUMO	V		S	
7	GAP	A		V	
8	dipole	S_{occ}		A	
9	MW				

10	S				
11	V				
12	Θ				
13	Ω				
14	Ω_{max}				
15	S_{occ}				
16	A				
17	R_{max}				
R_2					
18	HOMO	LUMO	LUMO	HOMO	HOMO
19	LUMO	GAP	GAP	LUMO	LUMO
20	GAP	dipole	dipole	GAP	V
21	dipole	MW	A	dipole	S(ethylene)/S
22	MW	S		MW	Ω
23	S	V		S	dipole
24	V	A		V	
25	Θ			Θ	
26	Ω			S(ethylene)/S	
27	Ω_{max}			Ω_{max}	
28	S_{occ}			S_{occ}	
29	A			A	
30	R_{max}			R_{max}	
31	S/S(ethylene)				
Ligand					
32	q_1	q_1	q	q_1	q_2
33	q_2	q_2	HOMO	q_2	HOMO
34	HOMO	HOMO	LUMO	HOMO	LUMO
35	LUMO	LUMO	GAP	LUMO	S_{occ}
36	GAP	GAP	S_{occ}	S_{occ}	A
37	dipole	Ω_{max}		A	R_{max}
38	MW	S_{occ}		R_{max}	
39	S				
40	V				
41	Θ				
42	Ω				
43	Ω_{max}				
44	S_{occ}				
45	A				
46	R_{max}				

Solvent					
47	n	q	q	Ω_{max}	Ω_{max}
48	ε_r				
49	dipole				
50	MW				
51	d^{25}				
52	V				
53	ΔH				
54	ΔU				
55	Δ_H				
56	AN				
57	DN				
58	π^*				
59	HOMO				
60	LUMO				
61	Θ				
62	Ω				
63	Ω_{max}				
64	S_{occ}				
65	A				
66	R_{max}				
67	q				
68	$q_{2(-)}$				
69	$q_{1(+)}$				
70	$q_{2(+)}$				
Experimental conditions					
71	Cat. precursor	T	T	T	T
72	Time	Pd loading	Pd loading	Pd loading	Pd loading
73	T	Cat. precursor	Cat. precursor	Cat. precursor	Cat. precursor
74	Pd loading			Time	Time

Set of virtual monophosphine ligands

The following building blocks were used to construct the 61 symmetrical monophosphine ligands with furmula PR_3 *in silico*. the broken lines show the connectivity of the R groups.


O
N
O
S
N
S
S
N
CH3
N
CH3
OCH3
OCH3
H2N
H3CO
CH3
H2N
(CH2)
NH2
OCH3
CH2NH2
O2N
NO2
NH2
OCH3
NO2
F
F
SO2Cl
NO2
N(CH3)2
CH3
NH2
CH3
HOH2C
CH2
O
O
O
N(CH2CH3)2
CH2OH
H3C
CH3
O
O
NC
H2
C
O
N
H3CH2C
H3C
CN
H2
C
CH2CH3
H3C
H3C
O
C
H2
O
N
N
H2
C
CN
CH2CH3
HO
N
N
H2NHN
N
OH
NHNH2
NHNH2
N
N