# Electronic Supplementary Information

**PTML Models of Surface Transition Metal Nanoparticle Catalysts for Cross-Coupling Reactions**

Andrea Ruiz-Escudero,1,2 Sonia Arrasate,3 and Humberto González-Díaz,3,4\*

1 Department of Computer Science and Information Technologies, Faculty of Computer Science, CITIC-Research Centre of Information and Communication Technologies, University of A Coruña, Campus Elviña s/n, 15071, A Coruña, Spain.

2 IKERDATA S.L., ZITEK, University of Basque Country UPVEHU, Rectorate Building, 48940 Leioa, Spain.

3 Department of Organic and Inorganic Chemistry, Faculty of Science and Technology, University of The Basque Country (UPV/EHU), P.O. Box 644, 48080, Bilbao, Spain.

4 IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain.

**Table S1** Description of the symbols of the Molecular descriptors (Dk), functional variables (Vk) and condition/labelling variables (cj) used in the database.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Symbol** | **Description** | **Symbol** | **Name** | **Symbol** | **Name** |
| D1 | Size of the catalyst (nm) | D11 | Squared Ghose-Crippen octanol-water partition coefficient (logP2) of the electrophile | D21 | Ghose-Crippen octanol-water partition coefficient (logP) |
| D2 | Van der Waals radius of the catalyst’s metal (pm) | D12 | Number of substituted benzene C(sp2) of the electrophile | D22 | Squared Ghose-Crippen octanol-water partition coefficient (logP2) of the nucleophile |
| D3 | Electronegativity of the catalyst’s metal | D13 | Number of halogen atoms of the electrophile | D23 | Number of substituted benzene C(sp2) of the nucleophile |
| D4 | Sum of atomic van der Waals volumes (scaled on Carbon atom) of the electrophile | D14 | Number of 6-membered rings of the electrophile | D24 | Number of Boron atoms of the nucleophile |
| D5 | Sum of atomic Sanderson electronegativities (scaled on Carbon atom) of the electrophile | D15 | Sum of atomic van der Waals volumes (scaled on Carbon atom) of the nucleophile | D25 | Number of 6-membered rings of the nucleophile |
| D6 | Sum of atomic polarizabilities (scaled on Carbon atom) of the electrophile | D16 | Sum of atomic Sanderson electronegativities (scaled on Carbon atom) of the nucleophile | D26 | Sum of atomic van der Waals volumes (scaled on Carbon atom) of the base |
| D7 | Sum of Kier-Hall electrotopological states of the electrophile | D17 | Sum of atomic polarizabilities (scaled on Carbon atom) of the nucleophile | D27 | Sum of atomic polarizabilities (scaled on Carbon atom) of the base |
| D8 | Moriguchi octanol-water partition coefficient (logP) of the electrophile | D18 | Sum of Kier-Hall electrotopological states of the nucleophile | D28 | Number of substituted benzene C(sp2) of the product |
| D9 | Squared Moriguchi octanol-water partition coefficient (logP2) of the electrophile | D19 | Moriguchi octanol-water partition coefficient (logP) of the nucleophile | D29 | Number of 6-membered rings of the product |
| D10 | Ghose-Crippen octanol-water partition coefficient (logP) | D20 | Squarded Moriguchi octanol-water partition coefficient (logP2) of the nucleophile |  |  |
| V1 | Time of the catalyst treatment (min) | V18 | Power of the microwave (W) (Step 1) | V35 | Quantity of the nucleophile (mmol) (Step 2) |
| V2 | Minimum quantity of the metal absorbed in the catalyst (µg) | V19 | Dipole moment of the solvent 1 (D) (Step 1) | V36 | Quantity of the base (mmol) (Step 2) |
| V3 | Maximum quantity of the metal absorbed in the catalyst (µg) | V20 | Molecular weight of the solvent 1 (g/mol) (Step 1) | V37 | Base equivalent (Step 2) |
| V4 | Minimum quantity of the metal released from the catalyst (µg) | V21 | Density of the solvent 1 (g/mL) (Step 1) | V38 | Oxidant equivalent (Step 2) |
| V5 | Maximum quantity of the metal released from the catalyst (µg) | V22 | Volume of the solvent 1 (mL) (Step 1) | V39 | Quantity of the oxidant (mol) (Step 2) |
| V6 | 1st dimension of the support (mm) | V23 | Solvent 1 equivalent (Step 1) | V40 | Additive equivalent (Step 2) |
| V7 | 2nd dimension of the support (mm) | V24 | Molecular weight of the solvent 2 (g/mol) (Step 1) | V41 | Quantity of the additive (mmol) (Step 2) |
| V8 | Quantity of the electrophile (mmol) | V25 | Density of the solvent 1 (g/mL) (Step 1) | V42 | Power of the microwave (W) (Step 2) |
| V9 | Nucleophile equivalent (Step 1) (Step 1) | V26 | Volume of the solvent 2 (mL) (Step 1) | V43 | Volume of the solvent 1 (Step 2) (mL) |
| V10 | Quantity of the nucleophile (mmol) (Step 1) | V27 | Solvent 2 equivalent (Step 1) | V44 | Dipole moment of the solvent 1 (D) (Step 2) |
| V11 | pKa of the base | V28 | Dipole moment of the solvent 2 (D) (Step 1) | V45 | Volume of the solvent 2 (mL) (Step 2) |
| V12 | Base equivalent (Step 1) | V29 | Total dipole moment (Step 1) | V46 | Dipole moment of the solvent 2 (D) (Step 2) |
| V13 | Quantity of the base (mmol) (Step 1) | V30 | Temperature (ºC) (Step 1) | V47 | Total dipole moment (Step 1) |
| V14 | Oxidant equivalent (Step 1) | V31 | Time (h) (Step 1) | V48 | Temperature (ºC) (Step 2) |
| V15 | Quantity of the oxidant (mol) (Step 1) | V32 | Was the reaction mixed? Y/N (Step 1) | V49 | Time (h) (Step 2) |
| V16 | Additive equivalent (Step 1) | V33 | Minimum times the catalyst was reused | V50 | Was the reaction mixed? Y/N (Step 1) |
| V17 | Quantity of the additive (mmol) (Step 1) | V34 | Nucleophile equivalent (Step 2) |  |  |
| c1 | Characteristics of the support of the catalyst | c4 | Generation of the synthesis of the SAM/SGM | c8 | Yield detection |
| c2 | Form of the metal in the support | c5 | Catalyst metal | c9 | Step where the nucleophile was added |
| c3 | Treatment/Procedure of the catalyst support | c6 | Time of reuse of the catalyst |  |  |

**Table S2** Statistical results of the MLR models for each IFPTML partition.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PTML partition table** | **Procedure** | **Variables** | | | **Multiple R** | **Multiple R²** | ***p-*value** | **Std.Err. of Estimate** |
| **1a** | FWS | * ∆V1(cI) * ∆V14(cIII) * ∆V17(cIII) | * ∆V30(cIII) * ∆V33(cIII) * ∆V44(cIII) | * ∆D5(cI) * ∆D26(cI) * ∆D29(cIII) | 0.8397 | 0.7051 | < 0.05 | 14.7105 |
| FWS  +  EGS | * ∆[D16∙(V10+ V35)](cI) * ∆[V11∙V13](cIII) * ∆[V29∙(V22+V26)](cIII) | * ∆[D4∙V8](cI) * ∆[D2∙V2](cI) * ∆V30(cII) | * ∆V31(cI) * ∆V33(cI) | 0.8210 | 0.6740 | < 0.05 | 15.4089 |
| **2a** | FWS | * ∆V1(cI) * ∆V14(cIII) | * ∆V18(cIII) * ∆V30(cIII) | | 0.8265 | 0.6831 | < 0.05 | 15.2147 |
| FWS  +  EGS | * ∆[D2∙V2](cI) * ∆[D5∙V8](cI) * ∆[D16∙∙(V10+ V35)](cII) * ∆[V29∙(V22+V26)](cII) | * ∆[V11∙V13](cII) * ∆V1(cI) * ∆V14(cIII) * ∆V18(cII) | * ∆V30(cIII) * ∆V31(cI) * ∆V33(cI) | 0.8333 | 0.6944 | < 0.05 | 14.9894 |
| **3a** | FWS | * ∆V1(cI) * ∆V14(cII) | * ∆V18(cI) * ∆V30(cII) | * ∆V33(cI) * ∆D24(cI) | 0.8371 | 0.7007 | < 0.05 | 14.7995 |
| FWS  +  EGS | * ∆[D15∙(V10+ V35)](cII) * ∆[V11∙V13](cII) * ∆[V29∙(V22+V26)](cII) | * ∆[D6∙V8](cII) * ∆[D2∙V2](cI) * ∆V30(cII) | * ∆V31(cI) * ∆V33(cI) | 0.8177 | 0.6686 | < 0.05 | 15.5863 |
| **4a** | FWS | * ∆[D16∙(V10+ V35)](cI) * ∆V12(cI) * ∆V13(cI) | * ∆V14(cI) * ∆V20(cI) * ∆V27(cI) | * ∆V31(cI) * ∆V33(cI) * ∆D1(cI) | 0.8373 | 0.7012 | < 0.05 | 14.8087 |
| FWS  +  EGS | * ∆[D17∙(V10+ V35)](cI) * ∆[V11∙V13](cI) * ∆[V29∙(V22+V26)](cI) | * ∆[D6∙V8](cI) * ∆[D2∙V2](cI) * ∆V30(cI) | * ∆V31(cI) * ∆V33(cI) | 0.8157 | 0.6653 | < 0.05 | 15.6647 |
| **1b** | FWS | * ∆V30(cIII) | * ∆V36(cI) | | 0.8205 | 0.6732 | < 0.05 | 15.3634 |
| FWS  +  EGS | * ∆[D16∙V10](cI) * ∆[V11∙V13](cIII) * ∆V29∙(∆V22+∆V26)(cIII) | * ∆[D2∙V2](cI) * ∆[D4∙V8](cI) * ∆V30(cIII) | * ∆V31(cI) * ∆V33(cI) | 0.8181 | 0.6694 | < 0.05 | 15.5775 |
| **2b** | FWS | * ∆[D5∙V8](cI) * ∆V8(cII) * ∆V10(cI) | * ∆V15(cII) * ∆V18(cII) * ∆V23(cII) | * ∆V21(cIII) * ∆V30(cIII) * ∆D26(cIII) | 0.8437 | 0.7118 | < 0.05 | 14.5530 |
| FWS  +  EGS | * ∆[D2∙V2](cI) * ∆[D5∙V8](cI) * ∆[D16∙(V10+ V35)](cI) | * ∆V15(cII) * ∆V18(cII) * ∆V23(cII) | * ∆V31(cI) * ∆V31(cIII) * ∆V33(cI) | 0.8204 | 0.6730 | < 0.05 | 15.5010 |
| **3b** | FWS | * ∆[V19∙V22](cI) * ∆[D4∙V8](cI) * ∆V8(cI) | * ∆V8(cII) * ∆V15(cI) * ∆V18(cI) | * ∆V22(cII) * ∆V30(cII) * ∆D26(cII) | 0.8440 | 0.7124 | < 0.05 | 14.5367 |
| FWS  +  EGS | * ∆[V11∙V13](cII) * ∆[V29∙(V22+V26)](cII) * ∆[D15∙(V10+ V35)](cII) | * ∆[D2∙V2](cI) * ∆[D6∙V8](cII) * ∆V30(cII) | * ∆V31(cI) * ∆V33(cI) | 0.8181 | 0.6693 | < 0.05 | 15.5787 |
| **4b** | FWS | * ∆V12(cI) | * ∆V25(cI) | | 0.8248 | 0.6803 | < 0.05 | 15.2639 |
| FWS  +  EGS | * ∆[D5∙V8](cI) * ∆[V29∙(V22+V26)](cI) * ∆[D17∙(V10+ V35)](cI) | * ∆[D2∙V2](cI) * ∆V12(cI) * ∆V30(cI) | * ∆V31(cI) * ∆V33(cI) | 0.8289 | 0.6870 | < 0.05 | 15.1549 |

antrain = 1134. bntrain = 851

**Table S3** Statistical results of the MLR models for each IFPTML partition without the double carbonylation set (ntrain = 790).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PTML partition table** | **Procedure** | **Variables** | | | **Multiple R** | **Multiple R²** | ***p-*value** | **Std.Err. of Estimate** |
| **1** | FWS | * ∆V30(cIII) * ∆V8(cIII) | * ∆V24(cI) | | 0.8732 | 0.7625 | < 0.05 | 11.8014 |
| FWS  +  EGS | * ∆[D4∙V8](cII) * ∆[D16∙(V10+ V35)](cIII) * ∆V29∙(∆V22+∆V26)(cIII) | * ∆[D2∙V2](cI) * ∆[V11∙V13](cIII) * ∆V30(cIII) | * ∆V31(cI) * ∆V33(cI) | 0.8601 | 0.7398 | < 0.05 | 12.4799 |
| **2** | FWS | * ∆[D15∙(V10+ V35)](cI) * ∆V2(cIII) * ∆V8(cI) | * ∆V8(cIII) * ∆V15(cI) * ∆V18(cII) | * ∆V29(cII * ∆V30(cIII * ∆D26(cIII) | 0.8985 | 0.8072 | < 0.05 | 10.7476 |
| FWS  +  EGS | * ∆[V29∙(V22+V26)](cI) * ∆[D15∙(V10+ V35)](cI) * ∆[D11∙V13](cIII) | * ∆[D6∙V8](cIII) * ∆[D2∙V2](cI) * ∆V30(cII) | * ∆V31(cI) * ∆V33(cI) | 0.8496 | 0.7218 | < 0.05 | 14.9894 |
| **3** | FWS | * ∆V1(cI) * ∆V8(cI) * ∆V12(cII) | * ∆V15(cI) * ∆V18(cI) * ∆V29(cI) | * ∆V29(cII * ∆V30(cI) * ∆D26(cII) | 0.8954 | 0.8017 | < 0.05 | 10.9022 |
| FWS  +  EGS | * ∆[D15∙(V10+ V35)](cII) * ∆[V11∙V13](cII) * ∆[V29∙(V22+V26)](cII) | * ∆[D6∙V8](cII) * ∆V30(cII) * ∆[D2∙V2](cI) | * ∆V31(cI) * ∆V33(cI) | 0.8604 | 0.7404 | < 0.05 | 12.4653 |
| **4** | FWS | * ∆V1(cI) * ∆V12(cI) * ∆V13(cI) | * ∆V15(cI) * ∆V19(cI) * ∆V25(cI) | * ∆D21(cI) * ∆D13(cI) * ∆D27(cI) | 0.8952 | 0.8013 | < 0.05 | 10.9113 |
| FWS  +  EGS | * ∆[D2∙V2](cI) * ∆[V29∙(V22+V26)](cI) * ∆[D17∙(V10+ V35)](cI) | * ∆[D6∙V8](cI) * ∆[V11∙V13](cI) * ∆V30(cI) | * ∆V31(cI) * ∆V33(cI) | 0.8723 | 0.7609 | < 0.05 | 11.9625 |