Electronic Supplementary Material (ESI) for New Journal of Chemistry.

SUPPORTING INFORMATION

Navigating Epoxidation Complexity: Building a Data Science Toolbox to Design Vanadium Catalysts

José Ferraz-Caetano[1], Filipe Teixeira[2], M. Natália D. S. Cordeiro[1]

1) LAQV-REQUIMTE – Department of Chemistry and Biochemistry – Faculty of Sciences, University of Porto - Rua do Campo Alegre, S/N, 4169-007 Porto, Portugal

2) CQUM – Centre of Chemistry, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

Corresponding Authors: **José Ferraz-Caetano** (jose.caetano@fc.up.pt)

M. Natália D.S. Cordeiro (ncordeir@fc.up.pt)

Model Database, Results and Code

The complete model database, with the list of compounds, model optimization features, full descriptor list and the repository for the sample code and data to produce the statistical results are available at the following online repository: https://github.com/jfcaetano/ESA-Database. Reactions were collected based on a literature review[1] found from 16 bibliographical sources[2-17], featuring homogeneous catalyst reactions published across different research articles

RDKit Full Descriptor List

Descriptors used in Machine Learning model development retrieved from RDKit software. The respective bibliographical foundation of each descripted is described in detail in the RDKit WebBook Documentation [18]. The descriptor calculations were made by converting each SMILES string (representing each molecular entry), running on top of Python version 3.9 [19] and using RDKit package, version 2022.09.4 [20].

Table S1

List of Descriptors Used in Machine Learning model development

Descriptor	Type	Meaning
BalabanJ	Structural	Distance sum of the two end-vertex for each edge. BalabanJ index has been proven to be relevant to network branching.
BertzCT	Structural	Complexity index, taking into account both the variety of kinds of bond connectivities and atom types; information contents related to bond connectivity and atom type diversity.
0χ , 1χ	Structural	This descriptor signifies a retention index (zero order) derived directly from gradient retention times.
0χn – 4χn	Structural	This descriptor signifies a retention index (zero order) derived directly from gradient retention times.
0χv – 4χv	Structural	This descriptor signifies atomic valence connectivity index (order 0).
VSA_EState1 – VSA_EState11	Surface Area	MOE-type (QSAR model) descriptors using EState indices and surface area contributions.
Hall Kier α	Structural	Descriptor fingerprint that displays the difference between active and inactive molecules.
HeavyAtomCount	Structural	The number of heavy atoms in the molecule.

HeavyAtomMolWt	Structural	The average molecular weight of the molecule ignoring hydrogens.
κ_1 , κ_2 , κ_3	Structural	This descriptor signifies # κ shape index: $(n-1) \times 2 / m^2$
Max and Min AbsEStateIndex	Electronic	Maximum and minimum Absolute E-State
Max and Min AbsPartialCharge	Electronic	Maximum and minimum Absolute Partial Charge
Max and Min EStateIndex	Electronic	Maximum and minimum E-State
Max and Min PartialCharge	Electronic	Maximum and minimum Partial Charge
Mol logP	Structural	Wildman-Crippen logP value.
MolMR	Electronic	Wildman-Crippen molar refractivity.
MolWt	Structural	The average molecular weight of the molecule.
NHOH Count	Structural	The number of NHs or OHs.
NO Count	Structural	The number of Nitrogens and Oxygens.
$n_{alicarb}$	Structural	The number of aliphatic carbocycles.
n_{alihet}	Structural	The number of aliphatic heterocycles.
n_{alirig}	Structural	The number of aliphatic rings.
$n_{arocarb}$	Structural	The number of aromatic carbocycles.
n_{arohet}	Structural	The number of aromatic heterocycles.
n_{arorig}	Structural	The number of aromatic rings.
n_{Ha}	Structural	The number of Hydrogen Bond Acceptors.
n_{Hd}	Structural	The number of Hydrogen Bond Donors.
n_{het}	Structural	The number of Heteroatoms.
n_{radele}	Structural	The number of radical electrons.

n_{rot}	Structural	The number of Rotatable Bonds.
n_{satrig}	Structural	The number of saturated rings.
n_{ele}	Structural	The number of valence electrons.
PEOE_VSA1 – PEOE_VSA14	Surface Area	MOE-type (QSAR model) descriptors using partial charges and surface area contributions.
SMR_VSA1 – SMR_VSA10	Surface Area	MOE-type (QSAR model) descriptors using MR contributions and surface area contributions.
SlogP_VSA1 – SlogP_VSA12	Surface Area	MOE-type (QSAR model) descriptors using logP contributions and surface area contributions.
TPSA	Surface Area	The total polar surface area of a molecule based upon fragment calculations.
VSA_EState1 – VSA_EState10	Surface Area	MOE-type (QSAR model) descriptors using EState indices and surface area contributions.

Experimental Descriptor List

Experimental descriptors used in Machine Learning model development (Table S2) retrieved from each bibliographical source. Specific solvent parameters were in particular retrieved from the Minnesota Solvent Descriptor Database [21].

Table S2

List of experimental descriptors wsed in machine learning model development.

Descriptor	Type	Meaning
Temp_K	Experimental	Reaction temperature in Kelvin.
Yield	Experimental	Reaction yield.
EE	Experimental	Reaction enantiomeric excess.
Configuration	Experimental	Product geometrical configuration.
Substrate_quant_mmol	Experimental	Quantity of reaction substrate in mmol.
Catalyst_quant_mmol	Experimental	Quantity of reaction catalyst in mmol.
Ligand_quant_mmol	Experimental	Quantity of reaction catalyst ligand in mmol.
Oxidant_quant_mmol	Experimental	Quantity of reaction oxidant in mmol.
Additive_quant_mL	Experimental	Volume of reaction additive in mL.

Solution_vol_mL	Experimental	Volume of reaction solution in mL.
Time_h	Experimental	Reaction duration in hours.
Solvent	Experimental	Reaction solvent.
Solv_opt_freq20	Experimental	Solvent index of refraction at optical frequencies at 293 K.
Solv_opt_freq25	Experimental	Solvent index of refraction at optical frequencies at 298 K.
Solv_Hbond_ac	Experimental	Solvent Abraham's hydrogen bond acidity.
Solv_Hbond_bs	Experimental	Solvent Abraham's hydrogen bond basicity.
Solv_surf_tens	Experimental	Solvent surface tension.
Solv_diele_cst	Experimental	Solvent dielectric constant at 298 K.
Solv_aromcity	Experimental	Solvent aromaticity: fraction of non-hydrogenic solvent atoms that are aromatic carbon atoms.
Solv_elect_halo	Experimental	Solvent electronegative halogenicity: fraction of non-hydrogenic solvent atoms that are F, Cl, or Br

References

(1) Ferraz-Caetano, J.; Teixeira, F.; Cordeiro, M. N. Systematic Development of Vanadium Catalysts for Sustainable Epoxidation of Small Alkenes and Allylic Alcohols. *International Journal of Molecular Sciences* **2023**, *24* (15). DOI: 10.3390/ijms241512299.

(2) Alvarez, H. M.; de Andrade, J. L.; Pereira, N.; Muri, E. M. F.; Horn, A.; Barbosa, D. P.; Antunes, O. A. C. Catalytic oxidation of isosafrol by vanadium complexes. *Catalysis Communications* **2007**, *8* (9), 1336-1340. DOI: https://doi.org/10.1016/j.catcom.2006.11.021.

(3) Barlan, A. U.; Zhang, W.; Yamamoto, H. Development and application of versatile bis-hydroxamic acids for catalytic asymmetric oxidation. *Tetrahedron* **2007**, *63* (27), 6075-6087. DOI: https://doi.org/10.1016/j.tet.2007.03.071.

(4) Bolm, C.; Kühn, T. Asymmetric Epoxidation of Allylic Alcohols Using Vanadium Complexes of (N)-Hydroxy-[2.2]paracyclophane-4-carboxylic Amides. *Synlett* **2000**, *2000* (06), 0899-0901. DOI: 10.1055/s-2000-6702.

(5) Bourhani, Z.; Malkov, A. V. Ligand-accelerated vanadium-catalysed epoxidation in water. *Chemical Communications* **2005**, (36), 4592-4594, 10.1039/B509436D. DOI: 10.1039/B509436D.

(6) Bryliakov, K. P.; Talsi, E. P.; Stas'ko, S. N.; Kholdeeva, O. A.; Popov, S. A.; Tkachev, A. V. Stereoselective oxidation of linalool with tert-butyl hydroperoxide, catalyzed by a vanadium(V) complex with a chiral terpenoid ligand. *Journal of Molecular Catalysis A: Chemical* **2003**, *194* (1), 79-88. DOI: https://doi.org/10.1016/S1381-1169(02)00527-7.

(7) Ghaffari, A.; Behzad, M.; Dutkiewicz, G.; Kubicki, M.; Salehi, M. Crystal structure, electrochemistry, and catalytic studies of a series of new oxidovanadium(IV) Schiff-base complexes derived from 1,2-diphenyl-1,2-ethylenediamine. *Journal of Coordination Chemistry 65* (5), 840-855. DOI: 10.1080/00958972.2012.662275.

(8) Han, L.; Liu, C.; Zhang, W.; Shi, X.-X.; You, S.-L. Dearomatization of tryptophols via a vanadium-catalyzed asymmetric epoxidation and ring-opening cascade. *Chemical Communications* **2014**, *50* (10), 1231-1233, 10.1039/C3CC47921H. DOI: 10.1039/C3CC47921H.

(9) Makita, N.; Hoshino, Y.; Yamamoto, H. Asymmetric Epoxidation of Homoallylic Alcohols and Application in a Concise Total Synthesis of (−)-α-Bisabolol and (−)-8-epi-α-Bisabolol. *Angewandte Chemie International Edition* **2003**, *42* (8), 941-943. DOI: https://doi.org/10.1002/anie.200390250 (acccessed 2023/11/07).

(10) Malkov, A. V.; Czemerys, L.; Malyshev, D. A. Vanadium-Catalyzed Asymmetric Epoxidation of Allylic Alcohols in Water. *The Journal of Organic Chemistry* **2009**, *74* (9), 3350-3355. DOI: 10.1021/jo900294h.

(11) Michaelson, R. C.; Palermo, R. E.; Sharpless, K. B. Chiral hydroxamic acids as ligands in the vanadium catalyzed asymmetric epoxidation of allylic alcohols by tert-butyl hydroperoxide.

Journal of the American Chemical Society **1977**, *99* (6), 1990-1992. DOI: 10.1021/ja00448a059.

(12) Monfared, H. H.; Bikas, R.; Mayer, P. Homogeneous green catalysts for olefin oxidation by mono oxovanadium(V) complexes of hydrazone Schiff base ligands. *Inorganica Chimica Acta* **2010**, *363* (11), 2574-2583. DOI: https://doi.org/10.1016/j.ica.2010.04.046.

(13) Murase, N.; Hoshino, Y.; Oishi, M.; Yamamoto, H. Chiral Vanadium-Based Catalysts for Asymmetric Epoxidation of Allylic Alcohols. *The Journal of Organic Chemistry* **1999**, *64* (2), 338-339. DOI: 10.1021/jo9821933.

(14) Pereira, C.; Leite, A.; Nunes, A.; Rebelo, S. L. H.; Rangel, M.; Freire, C. Oxidovanadium(IV) Complexes of 3-Hydroxy-4-pyrone and 3-Hydroxy-4-pyridinone Ligands: A New Generation of Homogeneous Catalysts for the Epoxidation of Geraniol. *Catalysis Letters* **2010**, *135* (1), 98-104. DOI: 10.1007/s10562-010-0281-8.

(15) Sheng, M. N.; Zajacek, J. G. Hydroperoxide oxidations catalyzed by metals. III. Epoxidation of dienes and olefins with functional groups. *The Journal of Organic Chemistry* **1970**, *35* (6), 1839-1843. DOI: 10.1021/jo00831a027.

(16) Zhang, W.; Yamamoto, H. Vanadium-Catalyzed Asymmetric Epoxidation of Homoallylic Alcohols. *Journal of the American Chemical Society* **2007**, *129* (2), 286-287. DOI: 10.1021/ja067495y.

(17) Zhang, W.; Basak, A.; Kosugi, Y.; Hoshino, Y.; Yamamoto, H. Enantioselective Epoxidation of Allylic Alcohols by a Chiral Complex of Vanadium: An Effective Controller System and a Rational Mechanistic Model. *Angewandte Chemie International Edition* **2005**, *44* (28), 4389-4391. DOI: https://doi.org/10.1002/anie.200500938 (acccessed 2023/11/07).

(18) *RDKit: Open-source cheminformatics - Descritptor Webbook*. https://www.rdkit.org/docs/GettingStartedInPython.html#list-of-available-descriptors (accessed March 1, 2022).

(19) *Python Software Foundation - Python Language Reference, version 3.9.8*. http://www.python.org (accessed.

(20) Landrum, G. *RDKit: Open-source cheminformatics 2022_09_4 (Q3 2022) Release - January 16, 2023*. 2023. http://www.rdkit.org/ (accessed January 18, 2023).

(21) Marenich, A. V. K., Casey P; Thompson, Jason D; Hawkins, Gregory D; Chambers, Candee C; Giesen, David J; Winget, Paul; Cramer, Christopher J; Truhlar, Donald G. *Minnesota Solvation Database (MNSOL) version 2012. Retrieved from the Data Repository for the University of Minnesota*. 2020. https://doi.org/10.13020/3eks-j059. (accessed.