Mini Review

Mini Review on Multivariate Linear Regression for Homogeneous Catalysis

Shashank Krishna R[1] and Pradeep Ilayaperumal[1,2]*

[1]Department of Chemistry, Bishop Heber College, India

[2]Department of Chemistry, Indian Institute of Technology Delhi, India

Abstract

The computational approaches are advantageous to the rational design of catalysts elucidating mechanistic insights, identifying active sites, and predicting catalytic activity. Here, the post-2018 short review on homogeneous catalysis, activation, and mechanistic approach using Multivariate Linear Regression (MLR) in DFT calculations.

Keywords: Multivariate Linear Regression (MLR); Selectivity prediction; Turn-over-frequency; Conformational analyses; Mechanistic analysis; DFT calculations

Introduction

Since the advent of computational methods, Density Functional Theory (DFT) in particular, a huge number of reports concerning the mechanistic rationale have been published and reviewed. Computational methods have been highly useful to predict or characterize intermediates, to explain the reaction mechanism, as an ideal alternative, where the experimental procedures have proven difficult or impossible. Selectivity prediction, turn-over-frequency, conformational analyses, spectroscopic calculations, mechanistic analysis are notable applications of computational chemistry, to name a few. However, the trade-off between accuracy and computational cost has limited the usage of higher-level calculations like CCSD, CASSCF, etc. Accuracy and applicability of such methods and questions relating to errors has been raised. While mechanistic studies are capable of providing insights on reaction processes or predictions, the need for simpler and ample methods has tremendously grown. Big-data and machine learning has gained a burgeoning rise in physical sciences establishing themselves as 'sine qua non' tools for the future. Moreover, Statistical tools have been ubiquitous in research and are suitable and proven to afford requisite performance at a lower computational cost.

Multivariate Linear Regression (MLR) based reaction optimization and catalyst design is one of the evolving approaches as a predictive platform in this field. The correlation between the experimental outcomes (Selectivity, yield, etc.) and molecular descriptors (Steric, electronic parameters) are used in the aforementioned method. The MLR model substantially depends on the descriptors, which should be sufficient enough to rely upon, such that the results obtained are admissible. The details of the MLR model design are described earlier [1,2] and hence not expounded here.

Earlier in 2018, Matthew Sigman and coworkers gave a review on MLR models for reaction development [3]. Herein we attempt to

ICBC.MS.ID.000522.

provide a short review, post-2018, covering key papers that utilize MLR models for design/optimization of reactions. The review is limited by the boundary of homogenous catalysis, concerning the vastness of literature.

Catalysis Using MLR

Selectivity in catalysis

Site-selective catalysis using MLR: Matthew Sigman's group has done extensive work usage of MLR models in catalysis. Site-selective acylation by BINOL-CPA (Chiral Phosphoric Acid) catalysts was interrogated by MLR models, revealing that role of substitution at R3 was found to improve the selectivity and the presence of CH−π interaction was also found to favour the desired product by DFT calculations [4].

Enantioselective catalysis using MLR: Matthew S. Sigman, F. Dean Toste and Scott J. Miller jointly reported a study on C2-symmetric chiral phosphoric acid (CPA) and Phosphothreonine-embedded, peptidic phosphoric acid (pThr-type CPA) scaffolds that catalyse atroposelective cyclodehydration. Infra Red (IR) vibrations, natural bond orbitals (NBO) charges, and Sterimol values were used as descriptors (variables). The process was found to proceed through the transition state that minimizes the steric repulsions, which indeed controls the enantioselectivity of C2-type CPA. Whereas pThr-type catalyst eluded the steric repulsions by conformational adaptations [5].

Chemical shift as a descriptors: Zhao, S., et al. reported, utilizing MLR for ligand design for stereo-specific palladium (Pd) Catalysed Suzuki type cross-coupling reaction was reported by Sigman. Tolman Cone angle, Sterimol dimensions and electronic properties such as chemical shift was used as descriptors. With the design and optimization guidance of MLR, ligand-controlled stereoselectivity was achieved. The study is one of the model work that shows the level of insights one can obtain from this statistical tool and shown in the Figure 1 [6]. (A) Workflow of parameter generation and statistical modeling. ΔΔG‡, relative free energy of activation; f(xi), function of parameter xi. (B) Application of phosphine parameterization to ligand optimization of the reaction with Z = CO2Et. b:l = branched-to-linear ratio. A positive % ee value indicates net retention; a negative % ee value indicates net inversion. es, enantiospecificity; R2, coefficient of determination. Reproduced from Zhao, S., et al. [6] by permission of science.

Non-covalent interactions as a descriptor: Dean Toste and M.S.Sigman jointly reported a study on highly enantioselective allenoate-Claisen rearrangement using doubly axially chiral phosphoric Acid (DAP) as the catalyst [7]. Non-Covalent interactions (NCI) are well known for their ability to impact the course and selectivity of reactions. Herein different interactions between catalyst and allylamine were taken as descriptors. Dispersion potential related terms were also used. While other such MLR studies focused on the design of catalysts, this work was focused upon the mechanistic elucidation of the reaction. The

effect of structural change on enantioselectivity was investigated. It was concluded (conformational change) the presence of stronger interactions increased enantioselectivity and the presence of ortho or meta substituents tends to favour one TS over another by structural change. One other study by Sigman and Toste on N-heterocyclic (NHC) gold(III) complexes in Diels–Alder reactions use MLR model to analyse both substrate and catalyst trends. A broad range of substrates was profiled and analysed. Non-Covalent Interaction (NCI) (π-π interaction between the proximal double bond of the substrate and the catalyst aromatic group) was the dictating factor in enantioselectivity [8].

Natural Bond Orbitals and Steric parameters as the descriptor: J. P. Reid., et al. reported, A study on Minisci reaction of diazines was reported wherein Multi-variate Linear Regression was used to study the mechanistic features, to guide the prediction of enantioselectivities and guiding the development of reaction to substrates such as pyrimidines and pyrazines.9 For the parameter selection, product structures were used. Steric parameters and NBO were used as descriptors. The results of models were suggestive of a pyrimidine to be amendable to the same reaction conditions.

Chemoselective catalysis using MLR: S.Chang and Co-workers reported an extensive computational study utilizing a Multi-variate Linear regression model for the design of Iridium catalyst for C-H amidation, where η5-cyclopentadienyl ancillary (CpX) and LX co-ligand were modified.10 NBO Charges, orbital energies, Sterimol parameters, Cone angle were taken as descriptors. Chemoselectivity of higher orders was achieved for acylnitrene transfer to either of sp2 spirocyclization or sp3 C-H insertion. The study serves as another model exhibiting the extent to which MLR methods can successfully predict or guide the development of catalysts [10].

Electrocatalysis using MLR: Mechanistic features of electro-catalytic reactions have also been interrogated by MLR models. Stability of cobalt (I) intermediate by bidentate cum-electronically asymmetric ligands was studied and correlation between thermo-dynamic and ligand properties with disproportionation rates were examined using MLRs [11]. Another study over Co(I) complex bearing two N,N,N-tridentate ligands a redox catalyst, was performed to study the reaction mechanisms. Electro-Analytical and kinetic studies were carried out in coherence with MLR models to delineate the mechanistic features [12].

Conclusion

The aforementioned studies are lucid examples demonstrating the potential of Multi-Variate Linear Regression (MLR) models as an effective and efficient platform to study, analyse and design catalysts yielding the desired selectivity and reactivity and can be carried out in synergy with experiments. One other factor to note is that most of the data used in the studies have been validated by methods such as LOO (leave-one-out) or other methods to ensure the robustness of the method used, as every value used is of immense importance in the determination of results. The regression models have existed

in statistics for years. Yet the putative use of MLR in chemistry is emerging as a powerful tool recently. We believe that the MLR methods can be of use in deciphering the intricacies of reaction mechanisms and towards the development of new reactions.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

References

1. MS Sigman, KC Harper, EN Bess, A Milo (2016) Accounts of Chemical Research 49: 1292-1301.

2. FD Toste, MS Sigman, SJ Miller (2017) Accounts of Chemical Research 50: 609-615.

3. CB Santiago, JY Guo, MS Sigman (2018) Chemical Science 9: 2398-2412.

4. J Li, S Grosslight, SJ Miller, MS Sigman, FD Toste (2019) ACS Catalysis 9: 9794-9799.

5. Y Kwon, J Li, JP Reid, JM Crawford, R Jacob, et al. (2019) Journal of the American Chemical Society 141: 6698-6705.

6. S Zhao, T Gensch, B Murray, ZL Niemeyer, MS Sigman, et al. (2018) Science 362: 670.

7. J Miró, T Gensch, M Ellwart, SJ Han, HH Lin, et al. (2020) Journal of the American Chemical Society 142: 6390-6399.

8. JP Reid, M Hu, S Ito, B Huang, CM Hong, et al. (2020) Chemical Science 11: 6450-6456.

9. JP Reid, RSJ Proctor, MS Sigman, RJ Phipps (2019) Journal of the American Chemical Society 141: 19178-19185.

10. Y Hwang, H Jung, E Lee, D Kim, S Chang (2020) Journal of the American Chemical Society 142: 8880-8889.

11. DP Hickey, C Sandford, Z Rhodes, T Gensch, LR Fries, et al. (2019) Journal of the American Chemical Society 141: 1382-1392.

12. T Tang, C Sandford, SD Minteer, MS Sigman (2021) Chemical Science 12: 4771-4778.