Please cite this article in press as: Anzenbacher et al., Toward data-science-guided prediction of enantiomeric excess in amines—A workflow method, Chem (2024), https://doi.org/10.1016/j.chempr.2024.06.032

Chem

CellPress

Preview

# Toward data-science-guided prediction of enantiomeric excess in amines—A workflow method

Pavel Anzenbacher, Jr.,[1,*] Sandra M. George,[1]
and Austin R. Sartori[1]

**Potential bottlenecks in high-throughput determination of enantiomeric excess in chiral primary amines is being addressed, by Howard et al. in this issue of *Chem*, by developing a data-driven methodology that replaces extensive calibration measurements and the corresponding calibration curves through a combination of machine learning methods and computations that produces a theoretical model and predicted calibration curves.**

Chiral compounds in the form of single optical isomers, typically enantiomers or diastereomers, are important starting materials for the synthesis of compounds of biological importance, including optically active pharmaceutical ingredients, herbicides, pesticides, drugs, catalysts, materials, and devices endowed with asymmetrical photophysical properties.[1,2]

Whether by asymmetric synthesis or the resolution of enriched mixtures and racemates, reliable enantiomeric excess (ee) determination methods are required to ensure and quantify the desired optical purity of ingredients and products. Typical methods used to this end include chiral liquid chromatography, capillary electromigration methods, and a plethora of chiral spectroscopy methods, including NMR using chiral solvating agents or lanthanide shift reagents.[3] Lately, chiroptical methods, namely circular dichroism spectroscopy (CD), came to the fore.[4]

However, except for CD, these methods are serial in nature and difficult to convert into parallel arrangements. Nevertheless, high-throughput ultrafast liquid chromatography (UHPLC)[5] or high-throughput screening (HTS) by NMR[6] using benchtop NMR instruments, among others, are vigorously pursued. CD (UV-Vis), on the other hand, has been adapted for medium-throughput screening in 96-well plate format.[7]

Anslyn[8] and Wolf[9] groups are at the forefront of developing high-throughput assays for determining ee that utilizes chemosensors in conjunction with CD (UV-Vis) using a CD multi-well plate reader.[7] This approach, however, relies on extensive calibration to achieve reliable determination of concentration and ee. Now, Anslyn and coworkers[10] utilize a previously described method based on condensation of a chiral amine with 3-hydroxypyridine-2-carboxaldehyde, which in the presence of Fe(II), forms a complex (Figure 1A) that exhibits a pronounced CD couplet at approximately 520 nm. The CD signal, emerging from the metal-to-ligand charge-transfer band (500–620 nm), stems from a mixture of two configurational isomers (*fac*/*mer*), each of which yields two helical isomeric forms ($\Delta$ and $\Lambda$). If the chiral amine is not enantiopure, 24 stereoisomers may form. Therefore, calibration curves must be prepared individually for each amine before performing the assay to assess the ee in the chiral amine samples. Anslyn and coworkers assert that the HPLC method, a currently prevalent method for ee determination, must be developed and calibrated using standards of known ee for each chiral analyte, which may produce a bottleneck for HTS workflows.

It is proposed that computational prediction of calibration curves from readily available structural parameters could greatly simplify the CD-based method. This advancement would circumvent the need for traditional chiral separation methods used in LC-based ee determination. Unlike chromatography, chiral spectroscopic techniques like CD and NMR with chiral additives do not necessitate enantiomer separation for analysis.

To validate this concept, Anslyn and coworkers devised a workflow aimed at creating a model capable of converting CD spectra derived from primary amines—whether pure or enantiomeric mixtures—into precise ee values using theoretical predictions. This innovation bypasses the resource-intensive development of chromatographic processes.

While promising, this approach may not universally apply across different methods involving varied chemosensor complexes, analyte classes, or impurities. Machine-learning-derived weights and variables are inherently specific to the system they were trained on. Nevertheless, the outlined strategy—a kind of procedural "recipe"—could be adapted with minor adjustments to construct new models suited to diverse scenarios.

---

[1]Center for Photochemical Sciences and Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403, USA

*Correspondence: pavel@bgsu.edu

https://doi.org/10.1016/j.chempr.2024.06.032

Please cite this article in press as: Anzenbacher et al., Toward data-science-guided prediction of enantiomeric excess in amines—A workflow method, Chem (2024), https://doi.org/10.1016/j.chempr.2024.06.032

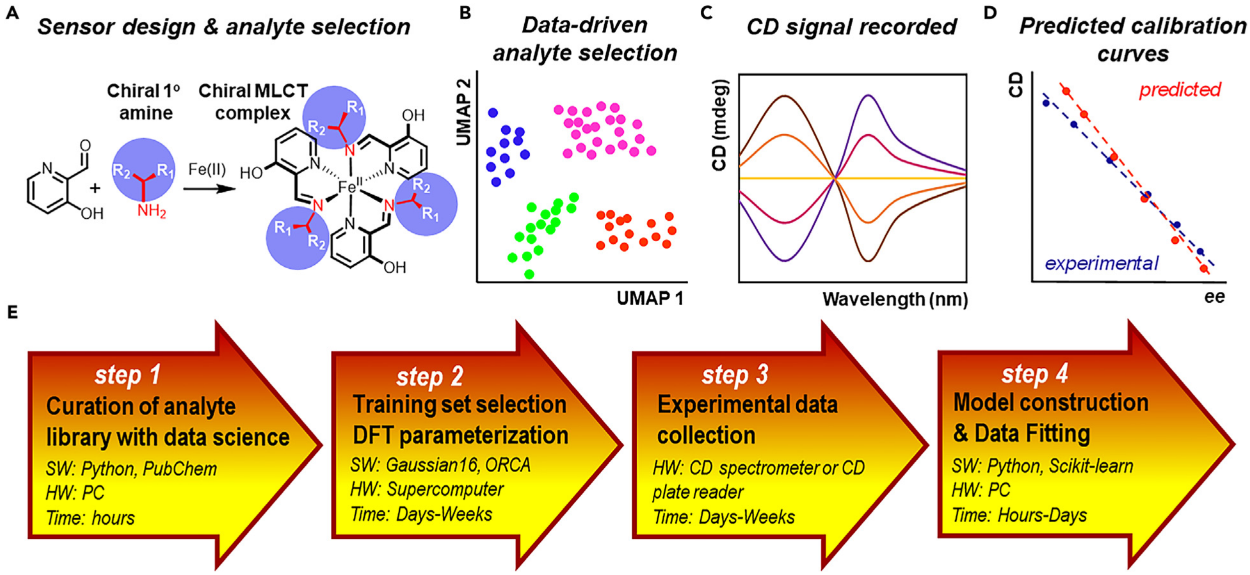

**Figure 1. The workflow of the data-driven methodology to measure and predict the enantiomeric excess (ee) in primary amines**

(A) Formation of an octahedral Fe(II) complex utilized in CD spectroscopy.

(B) The pattern of structurally distinct clusters of primary amines for method training was identified through two-dimensional UMAP analysis.

(C) The Fe(II)-amine complex recorded CD spectra for the calibration experiment.

(D) Multiple linear regression (MLR) calibration allows us to predict the maximum CD signal observed for enantiopure amines.

(E) A general workflow for a data-driven high-throughput determination of enantiomeric excess in chiral primary amines.

The methodology relies on quantum chemical descriptors such as Sterimol parameters and other structural parameters (e.g., partial charges, natural bond orbital populations) to comprehensively characterize each analyte. Regression models are then employed to predict the maximum CD signal for optically pure amines. Such models, computationally generated using cost-effective DFT techniques, aim to establish a correlation between analyte structure and CD signal, which is crucial for accurate ee determination. This computational approach aligns with current trends in asymmetric catalyst development, underscoring its robustness and applicability in advancing chirality sensing methods.

In this study, Anslyn and coworkers selected a reaction that had previously been studied (Figure 1). They used a large number (1,826) of low-level descriptors such as heteroatoms, topological descriptors, calculated partition coefficients, and other parameters used in quantitative structure-activity relationships to create a broad pool of information comprising enough information to map the chemical space of the computational library of chiral primary amines. From this large pool of descriptors, only the descriptors comprising a higher amount of variance were retained, and the remaining descriptors espousing high variance were then projected into two dimensions using Uniform Manifold Approximation Projection (UMAP). UMAP analysis produced a pattern of five structurally distinct clusters of primary amines, and the amines for this study were selected from each cluster to represent the main structural features to generate a training set that comprises a maximum diversity-variance. Such a diverse training set is believed to capture the salient structural features within the 54-amine dataset of 28 enantiomer pairs (14 different structures with both enantiomers). Also, 26 amines were used without their opposite enantiomer.

Each amine in the training set underwent parameterization by modeling its corresponding imine and optimizing its structures. Subsequently, the structures were subjected to a Conformer-Rotamer Ensemble Search Tool (CREST) followed by further optimization of the conformers. Furthermore, single-point energy calculations were then performed. Various weighted geometric, electronic, and steric descriptors, including natural bond orbital charges and Sterimol parameters, were derived from these energies. These descriptors were utilized in multiple linear regression (MLR) models used for asymmetric catalysis and ee determination assays to provide insights into the chemical properties influencing ee determination, crucial for practitioners in chemistry. MLR was employed to select the most relevant chemical descriptors correlating with the experimental CD data. However, the CD signal of certain amines was observed to be affected by incomplete complex formation. In this study, computational

Please cite this article in press as: Anzenbacher et al., Toward data-science-guided prediction of enantiomeric excess in amines—A workflow method, Chem (2024), https://doi.org/10.1016/j.chempr.2024.06.032

corrections were applied to address incomplete complex formation without modeling the entire Fe(II) complex, which is not always feasible and would be computationally expensive. A computational workaround was employed to exclude amines unlikely to form the Fe(II) complex, resulting in a refined set of 25 amines that significantly improved model predictions related to CD spectra for complexes of chiral primary amines with Fe(II). The data suggest that differences in steric demand at the stereogenic center dominate variations in CD spectra, with models capturing these steric contributions adequately predicting CD spectra of the selected amines. The method was validated in a mid-throughput manner using 96-well plates, demonstrating its practical potential.

To showcase the practicality of the model in predicting the ee of actual samples, a single-blind ee determination was conducted using the EKKO CD plate reader to measure the ee in solutions for four different enantiomer pairs at various ee levels. The analyses of the experimental calibration curves resulted in a low mean absolute error (MAE) of 4.5%, whereas the predicted calibration curves yielded an MAE of 8.8%. The experimental and the predicted calibration curves intersect near 0 mdeg, corresponding to racemic mixtures, and the difference between the predicted and experimental calibration curves increases with increasing ee values due to the slope divergence in the two lines. Thus, the magnitude of the errors increases at the higher ee values, thereby impacting the most interesting regions from the perspective of practical ee determination.

The present methodology has specific requirements (Figure 1E). First and foremost, modeling chiroptical sensors requires knowledge in many fields, including analytical, computational, and physical organic chemistry. For starters, access to libraries and services (e.g., Reaxys) and knowledge of physical chemistry, QSAR, MORDRED, and RDKit descriptors, etc., are required. As always, chemical intuition is a large part of the expertise and may impact the selection of library samples, molecular diversity, etc., without which any calculations may not provide a usable model. Then, there is the issue of calculations, supercomputers for the DFT calculations, as well as the days, or weeks, of runtime.

Experimental work, however, is more straightforward. With access to a CD spectrometer or plate reader, a skilled technician or MS/PhD student can perform these tasks adequately. Yet, constructing a sufficiently diverse training library poses challenges, mainly due to the need for multiple sample replicas and the lower sensitivity of the method, which together require substantial standard/analyte quantities (1–10 mg). Moreover, accurate model evaluation demands proficiency in statistical regression techniques.

This preview article highlights the challenging entry barrier for creating a predictive chirality sensing model using CD spectroscopy. Conversely, UHPLC offers a robust, cost-effective, single-chemist-operable commercial solution at an acceptable price (<$500,000) for high-throughput setups. Nevertheless, all LC methods, including UHPLC, remain inherently serial methods.

This brings the final considerations: *quo vadis* (where are you going)? And also, *cui bono* (who benefits)? The initial steps toward significant change always appear daunting and require substantial investment, some of which may not be recovered. However, with practice comes perfection, and as these methods are practiced, they will be improved and made perfect in the future. The primary beneficiaries will be large research operations with qualified physical chemists, machine learning experts, computational chemists, and analytical chemists. Furthermore, we predict that large enterprises in chiral catalyst discovery, chiral transformations, and large-scale drug discovery will adopt these methods. This journey will be thrilling, and we urge readers to study this publication by Anslyn and coworkers. For those interested in this field, it promises to open new horizons.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF CHE 2102581 to P.A.) and Bowling Green State University (Building Strength Award, project no. 3300368 to P.A.). A.R.S. is supported by Taller Family Graduate Fellowship.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

1. Brooks, W.H., Guida, W.C., and Daniel, K.G. (2011). The Significance of Chirality in Drug Design and Development. Curr. Top. Med. Chem. *11*, 760–770. https://doi.org/10.2174/156802611795165098.
2. Ham, S.-H., Han, M.J., and Kim, M. (2024). Chiral Materials for Optics and Electronics: Ready to Rise? Micromachines *15*, 528.
3. Polavarapu, P.L. (2018). Chiral analysis : advances in spectroscopy. In Chromatography and Emerging Methods, Second edition ed.Chromatography and Emerging Methods (Elsevier).
4. Woody, R.W. (1995). Circular dichroism. In Methods Enzymol., *246*Methods Enzymol. (Academic Press), pp. 34–71.
5. Kaplitz, A.S., Kresge, G.A., Selover, B., Horvat, L., Franklin, E.G., Godinho, J.M., Grinias, K.M., Foster, S.W., Davis, J.J., and Grinias, J.P. (2020). High-Throughput and Ultrafast Liquid Chromatography. Anal. Chem. *92*, 67–84. https://doi.org/10.1021/acs.analchem.9b04713.
6. Wu, B., Barile, E., De, S.K., Wei, J., Purves, A., and Pellecchia, M. (2015). High-throughput screening by nuclear magnetic resonance (HTS by NMR) for the identification of PPIs antagonists. Curr. Top. Med. Chem. *15*, 2032–2042. https://doi.org/10.2174/1568026615666150519102459.
7. Herrera, B.T., Pilicer, S.L., Anslyn, E.V., Joyce, L.A., and Wolf, C. (2018). Optical Analysis of Reaction Yield and Enantiomeric Excess: A New Paradigm Ready for Prime Time. J. Am.

Chem. Soc. *140*, 10385–10401. https://doi.org/10.1021/jacs.8b06607.

8. Dragna, J.M., Pescitelli, G., Tran, L., Lynch, V.M., Anslyn, E.V., and Di Bari, L. (2012). In Situ Assembly of Octahedral Fe(II) Complexes for the Enantiomeric Excess Determination of Chiral Amines Using Circular Dichroism Spectroscopy. J. Am. Chem. Soc. *134*, 4398–4407. https://doi.org/10.1021/ja211768v.

9. Sripada, A., Thanzeel, F.Y., and Wolf, C. (2022). Unified sensing of the concentration and enantiomeric composition of chiral compounds with an achiral probe. Chem *8*, 1734–1749. https://doi.org/10.1016/j.chempr.2022.03.008.

10. Howard, J.R., Shuluk, J.R., Bhakare, A., and Anslyn, E.V. (2024). Data-science-guided calibration curve prediction of an MLCT-based ee determination assay for chiral amines. Chem *10*, ■■■.