Title: Screening for Generality in Asymmetric Catalysis

One-Sentence Summary: A new method for high-throughput enantioselectivity determination on diverse substrates enables an intentional approach for the identification of broadly effective asymmetric catalysts.

Authors: Corin C. Wagen\textsuperscript{1}\textsuperscript{†}, Spencer E. McMinn\textsuperscript{2}\textsuperscript{†}, Eugene E. Kwan\textsuperscript{3}\textsuperscript{*}, Eric N. Jacobsen\textsuperscript{1}\textsuperscript{*}

Affiliations:
\textsuperscript{1}Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA
\textsuperscript{2}Discovery Chemistry, Merck & Co. Inc, Boston, MA 02115, USA
\textsuperscript{3}Process Research and Development, Merck & Co. Inc, Boston, MA 02115, USA
*Corresponding authors. Email: eugene.kwan@merck.com; jacobsen@chemistry.harvard.edu
\textsuperscript{†}These authors contributed equally.

Abstract
Research in the field of asymmetric catalysis over the past half century has resulted in landmark advances, enabling the efficient synthesis of chiral building blocks, pharmaceuticals, and natural products. A small number of asymmetric catalytic reactions have been identified that display high selectivity across a broad scope of substrates; not coincidentally, these are the reactions that have the greatest impact on how enantioenriched compounds are synthesized. We postulate that substrate generality in asymmetric catalysis is rare not simply because it is intrinsically difficult to achieve, but also because of the way chiral catalysts are identified and optimized. Typical discovery campaigns rely on a single model substrate, and thus select for high performance in a narrow region of chemical space. Here, we put forth a practical approach for using multiple model substrates to select simultaneously for both enantioselectivity and generality in asymmetric catalysis from the outset. Multi-substrate screening is achieved by conducting high-throughput chiral analyses via supercritical fluid chromatography-mass spectrometry (SFC-MS) with pooled samples. When applied to Pictet–Spengler reactions, the multi-substrate screening approach revealed a promising and unexpected lead for the general enantioselective catalysis of this important transformation.

Introduction
Since the discovery that chiral phosphine-rhodium(I) complexes catalyze the highly enantioselective hydrogenation of certain dehydroamino acids (1,2), asymmetric synthesis with small-molecule catalysts has been demonstrated in a dazzling variety of contexts (3). It is now widely appreciated that high enantioselectivity, typically defined as >90\% enantiomeric excess (ee), is often attainable for specific model substrates in a reaction of interest. While these model substrates may be the targets of a specific synthetic campaign, they are more frequently chosen on the basis of accessibility, ease of chiral analysis, lack of peculiar structural features, or similarity to substrates studied successfully in related reactions.

To this day, high enantioselectivity remains the \textit{sine qua non} of asymmetric catalysis development efforts; it is only after that condition is met with a model substrate that the substrate scope and limitations of the reaction are evaluated (Fig. 1A). Given the truism that “you get what you screen for” (4), an unintended consequence of this paradigm is that optimization in this manner fundamentally selects for success with substrates that are similar to the model. In rare cases, asymmetric transformations with broad substrate scope do emerge, and those are typically the ones that are most impactful as they can be applied predictively in new contexts (5–9). But such
generality is effectively accidental. In the vast majority of cases, the scope is limited, researchers strain to identify enough high ee examples to fill the “Substrate Scope” table requisite for publication, and the methods remain underutilized because synthetic practitioners shy away from trying unproved or unpredictable chemistry.

Optimization against multiple, diverse substrates simultaneously rather than against a single model substrate would shift the focus in asymmetric catalysis discovery efforts from identifying circumscribed examples of high enantioselectivity to revealing more general solutions (Fig. 1B). This approach was proposed in 1999 by Kagan et al. (10,11) and articulated compellingly in recent studies by MacMillan et al. (12) and List et al. (13), but its adoption has been largely precluded by the challenges associated with conducting large numbers of chiral analyses on a variety of products. Even though methods for ee determination have advanced substantially over the past several decades, their development and application remain laborious and time-consuming. Accordingly, researchers will only commit to the laborious task of developing and applying ee-determination methods for multiple products after success with a model substrate has been achieved.

The most commonly applied analytical methods for ee determination involve high-performance liquid chromatography (HPLC) or supercritical fluid chromatography (SFC) with chiral stationary phases (CSPs) in conjunction with isocratic elution and UV-Vis detection (Fig. 1C). Such methods are inherently limited in throughput because of the requirement for baseline separation and the removal of interfering signals before analysis. While there has been long-standing interest in developing high-throughput chiral analysis methods, the techniques identified to date have been tailored for repeated analyses of specific analyte classes of interest. For instance, the Anslyn and Wolf groups have developed circular-dichroism-based sensors that enable ee determination for compounds containing chelating functional groups (14,15). Other approaches have employed chiral 19F-NMR shift reagents (16), fluorescent DNA biosensors (17), mass-tagging with pseudo-enantiomers (18), and selective enzymatic oxidation (19,20). Although these methods can provide high accuracy and sample throughput, more general analytical strategies are needed to conduct effective multi-substrate screening across a range of chemical space.

We envisioned that combining conventional chromatographic techniques and sample pooling might allow facile and simultaneous ee determination for a diverse panel of products. In particular, we investigated the coupling of SFC, which offers very short analytical runtimes and excellent resolution for most substrate classes, with mass spectrometry (MS) as the detection method (21). By choosing substrate combinations that generate products of differing mass, combining aliquots drawn from multiple independent reactions into a single analysis vial, and subjecting the sample to SFC-MS and extracted ion chromatogram (EIC) analysis, we anticipated that it would be possible to obtain many ee measurements simultaneously (Fig. 1D).

This strategy of using SFC-MS to analyze pooled samples has many attractive features. First, SFC already enjoys widespread use and has been applied successfully to the rapid analysis and separation of a wide variety of compounds (21). By applying MS as the detection method, it is possible to use an elution gradient; a single, generic gradient can be applied to analytes possessing a wide range of polarities, greatly simplifying analytical method development (22). Recent advances in immobilization strategies for polysaccharide chiral selectors have given rise to commercially available CSPs with excellent separation properties and improved robustness, increasing the possibility of analyzing crude reaction mixtures of a wide variety of products using only a small number of columns (23). Finally, the EIC analysis would enable ee determinations when the product enantiomers are separated but co-elute with other products or residual reaction components. Overall, we anticipated that the combination of SFC-MS and sample pooling would raise the throughput of ee determination to the point where it may become practical to optimize directly for both enantioselectivity and generality by allowing high-throughput analysis of multiple crude reaction mixtures simultaneously.
**Fig. 1. Approaches to the discovery and analysis of enantioselective reactions.** A) The standard approach to discovery of new asymmetric catalytic reactions involves optimization around a single model substrate. The scope of the method is then examined in a separate exercise, often resulting in methods that are only highly effective for substrates similar to the model. B) Optimization of asymmetric transformations via multi-substrate screening improves the chances of identifying more general catalysts and conditions, particularly if screening is performed across a broad cross-section of substrate space. C) Conventional ee determination of single isolated products via chromatography on CSPs and detection by UV-Vis. D) Proposed approach to simultaneous, high-throughput ee determination of multiple products. Crude reaction mixtures are pooled and analyzed by SFC-MS. Signals due to resolved products are detected without interference by other products or other reaction components of different mass.

**Methods**
The performance of SFC-MS for rapid ee determination of single and pooled samples was first assessed with a set of commercially available compounds using standard chromatographic columns. When pure samples of known enantiomeric composition were analyzed at high concentrations (10 mM), an unacceptably large root-mean-square error (RMSE) of 22% resulted, with the greatest deviations for scalemic mixtures (Fig. S2). A major source of error originates from the nonlinear relationship between detector response and concentration, which is exacerbated at high signal intensities and therefore results in underestimations of the ee (Fig. S2). However, the signal dependence on concentration approaches linearity at lower sample concentrations (<0.1 mM, Fig. S3), and the RMSE in the analyses could thus be reduced to 3%, which is comparable to standard HPLC-UV conditions (24). Consequently, all subsequent samples were analyzed at the lowest concentrations that still afforded good signal-to-noise (typically 0.1 mM, assuming 100% theoretical yield), a practice that was convenient for high-throughput assays of reactions carried out at micromole-scale.

Given the intention to carry out analyses of multiple compounds rapidly and simultaneously with the fewest possible injections, we recognized that baseline separation of all enantiomeric pairs would not always be achieved. We therefore sought to develop a general peak-fitting method to extract accurate integrations from partially separated enantiomers. Fitting the experimental data of baseline-resolved peaks produced from scalemic mixtures of known composition to linear combinations of peak functions revealed that the SFC-MS peak shapes were not well reproduced by any single known peak functions. However, a “Frankenstein” model involving a piecewise combination of Gaussian and Voigt functions was found to provide excellent fits (Fig. S1). We combined this peak model with parameter-fitting methods in a convenient web-based application for analyzing chromatographic data. Accurate ee determinations were obtained when the protocol was applied to the analysis of the same scalemic standards on columns that produced only partial separation of the enantiomers, even in the challenging cases of high-ee samples where the first major peak tailed into the second minor peak (Figs. S4-S5).

Fig. 2: Development of the SFC-MS method. (A) Analysis of 20 pooled racemic compounds by SFC-MS. The RMSE is 7% and is dominated by a few outliers (n=58). (B) Heatmap visualization of the total ion chromatogram. Highly ionizable compounds produce ion suppression (dark vertical streaks). (C) Pairwise experiments with racemates confirm that co-elution of different compounds can result in ion suppression and inaccuracies in the ee determinations. This problem was circumvented through the application of “smart pooling” of the product (see text).
With a rapid and accurate SFC-MS ee-determination method in hand, we evaluated its accuracy for analysis of pooled samples. A mixture of 20 commercially available racemic compounds (mostly pharmaceuticals; Table S3) at low concentration was analyzed with five chromatographic columns, registering ees of $-1\pm 7\%$, with several significant outliers (Fig. 2A). The origin of these outliers was identified upon inspection of the total ion chromatograms displayed as two-dimensional heatmaps (Fig. 2B). The bright horizontal lines represent unavoidable contaminants (e.g., polyethylene glycol) and did not interfere with the analyses. However, the dark vertical bands are the result of ion suppression by strongly ionizing species, which results in reduced intensity for any co-eluting ions (25,26). If one enantiomer falls under such a band and its partner does not, the apparent ee is distorted in favor of the unaffected peak. This effect is illustrated in pairwise experiments with racemates (Fig. 2C).

Recognizing the need to minimize ion suppression and increase the capacity of our workflow to handle large panels of products, we introduced an alternative pooling scheme. Rather than combining all compounds into a single vial and injecting the entire pool onto each column (“simple pooling”), we pooled subsets of products into multiple vials and injected each vial only once (“smart pooling”, see SI for details). This latter procedure reduces ion suppression, avoids the unnecessary injection of compounds onto columns that do not separate them, and was applied in the screens discussed below.

**Results and Discussion**

We selected asymmetric catalysis of the Pictet–Spengler reaction as a timely and synthetically relevant platform to illustrate the application of the new high-throughput enantioselectivity determination methodology and the “screening for generality” concept. The condensation of tryptamines with aldehydes or ketones to generate tetrahydro-β-carbolines (Fig. 3A) is a venerable reaction with crucially important applications in laboratory and biological synthesis (27,28). The reaction has inspired intensive research efforts in search of asymmetric catalytic variants, and so far over a dozen distinct catalytic systems have been described (29–43). Each study relied on optimization of catalyst and conditions around very limited numbers of model substrates, and resulted in the identification of highly enantioselective reactions. By traditional standards, this output of new catalysts, high ee examples, and publications in high-impact journals can certainly be viewed as a major success. Yet despite this apparent progress, none of the published methods has found widespread application, and the chemist interested in carrying out an enantioselective catalytic Pictet–Spengler reaction on a never-before-tested substrate combination would be hard-pressed to know which system to try. We sought to establish whether screening across broad stretches of substrate space might prove informative in that regard and possibly enable the identification of general systems.

Selection of the panel of model substrates represents a key step in any screening-for-generality exercise, and we sought to maximize the structural and functional diversity of aldehyde and tryptamine combinations. We constructed an in silico library of 340 potential tetrahydro-β-caroline products, generated molecular fingerprints for each member, and performed UMAP dimensionality reduction to generate a two-dimensional representation of the diversity of potential products (Fig. 3B) (44). By combining insights gained from this representation with practical considerations such as reactivity, ease of separation, and commercial availability, we winnowed the candidates down to a panel of 14 representative products (Fig. 3A). Notably, these products encompass regions of chemical space that are largely unexplored by reported methodologies.

Pictet–Spengler reactions can proceed via N-acyl-, N-protio-, and N-alkyl-iminium ion intermediates, and enantioselective catalytic variants have been identified for each of these manifolds (29–43). We selected reactions between N-benzyl tryptamines and aldehydes as a particularly convenient platform to survey the substrate/catalyst landscape. In particular, reactions with N-benzyl tryptamines were found to remain homogenous, enabling the reactions to be run in a format that is highly amenable to parallel screening with standard equipment (96 well plates, no stirring, 0.01 mmol scale). Acceptable agreement between the
enantioselectivity values obtained by single sample SFC-UV and pooled sample SFC-MS was achieved for the resulting N-benzyl-tetrahydro-β-carboline products (8% RMSE, Fig. S32).

Fig. 3. High throughput ee-determination of enantioselective catalytic Pictet–Spengler reactions (A) The 14-member panel of products (left) used to study the Pictet–Spengler reaction and a map (right) of potential
products (grey) with previously reported products from the literature (blue) vs. our panel (red). (B) Enantioselectivity screen using 14 previously reported organocatalysts against the 14-member panel. Reactions with weakly acidic H-bond-donor catalysts i–ix and xii–xiv were run with benzoic acid as a co-catalyst. Empty squares represent low-yielding reactions.

We evaluated a set of previously reported chiral Brønsted-acid and H-bond-donor catalysts across the selected panel. The resulting ee data (Fig. 3B) reveal that different catalyst classes respond very differently to variations in substrate. In some cases, such as thioureas i–vii, moderate (e.g., 20–40%) enantioselectivity was observed fairly consistently across most substrates. Other catalysts afford higher (e.g., >60%) enantioselectivity, but only for specific subsets of the substrate space; for example, Miller squaramide ix is particularly effective in catalyzing the reaction between neutral indoles (X=Y=H) and aryl aldehydes, whereas the SPINOL-phosphoric acid xii reported by Lin and Wang et al. stood out in reactions of the electron-deficient indole (X=Cl, Y=H) (34, 43).

Tabulation of the data as a heatmap (Fig. 3B) provides a visually straightforward tool for identifying correlations in behavior between related catalysts. For instance, chiral phosphoric acids xi and xii perform poorly with the electron-rich indole (X=H, Y=OMe), and generally much better with the electron-poor or neutral tryptamine analogs. Similarly, comparison of ee values between rows reveals correlations between substrates. For example, compound 23, a prototypical minimally functionalized model product derived from N-benzyl tryptamine and benzaldehyde, responds similarly to catalyst effects as do products 24 and 28 also possessing the neutral indole. However, product 23 is a very poor model for products possessing electron-rich indoles such as 11, 12, 16, or 19. These findings highlight the risks associated with optimizing around a single model substrate combination and the value of employing multi-substrate screening for the accurate assessment and optimization of a given methodology.

To assist in quantifying the level of generality displayed by each catalyst, we constructed a generality metric $g$ that summarizes a collection of enantioselectivity values into a number between 0 and 1, where 1 represents a completely general catalyst that induces 100% ee in every reaction surveyed (see SI Section 3.9 for further discussion). By this analysis, catalyst xii stands out among those surveyed as the most promising from a generality standpoint.

Despite the encouraging results obtained with catalyst xii, it is apparent from the data in Fig. 3B that reactions of the electron-rich indole are particularly challenging for that, and indeed all catalysts in the screen. We zeroed in on that substrate class by evaluating several solvent–catalyst combinations for product 16 (Fig. S40). Improved results were obtained with polar aprotic compounds such as 2-methyl THF (2MT) and ethyl acetate (EA), prompting us to evaluate these solvents across the entire substrate panel (Fig. 4A). Notably, the solvent effects were highly catalyst-dependent, with enhanced enantioselectivities observed in reactions performed in 2MT for chiral phosphoric acid xii, but no systematic improvement observed with H-bond donor catalyst i. The substantial benefits of 2MT were possibly missed in the original work that led to the development of catalyst xii (34) because the authors used only a single model product, similar to 23 and related N-protected analogs. While changing from PhMe to 2MT leads to a small decrease in the enantioselectivity of 23, most other substrates benefit substantially, with the increase from 1% to 62% ee for product 11 serving as a particularly notable example. These results illustrate how performing screens on multiple substrates simultaneously can provide valuable and otherwise elusive insights. When xii was used to generate product 6 in a standard, rather than high-throughput, experimental format, a slightly improvement in enantioselectivity from 86 to 91% ee was observed (Fig. 4B), thereby demonstrating the successful translation of the HTE-based screens and analyses to laboratory-scale reactions.
Figure 4: Solvent studies and scale-up validation. (A) Solvent screening reveals the beneficial effects of 2-methyl THF (2MT) compared to toluene (PhMe) and ethyl acetate (EA) for most substrate combinations. (B) Scaleup of substrate 6 with conventional glassware and analysis demonstrates that hits from the multi-substrate screens can translate to preparative-scale reactions.
Relative to traditional ee-determination methods, the analytical workflow outlined in this study trades a modest decrease in accuracy for the ability to analyze multiple crude reaction mixtures simultaneously with substantial reductions in method development and analysis time. As a result, optimization of catalyst structure and reaction conditions is more readily performed across a variety of substrate combinations, thereby increasing the chances of identifying general protocols. If no broadly general protocols are found for a given reaction of interest, the broad survey of the catalyst/substrate landscape can be used to ensure that highly effective catalyst-substrate combinations are not missed, and to identify “islands” of high enantioselectivity involving specific catalyst structures and subsets of the substrate space.

The survey of the substrate/catalyst landscape in the asymmetric Pictet–Spengler reaction reveals how the standard approach of basing optimization campaigns on a single model system has indeed resulted in substrate-specific islands of high enantioselectivity for this reaction. However, an unexpected finding from this study is that there is already a family of catalysts (exemplified by xii) that displays very promising levels of generality across a wide range of substrate combinations. The uniquely interesting features of xii was likely masked by the fact that it was identified through an evaluation of a fairly narrow region of substrate space (34). Similarly, the beneficial effect of 2-methyl THF as a solvent in reactions with catalyst xii was masked by the anomalous behavior of the prototypical model substrates for that reaction. Thus, the use of multi-substrate screening revealed a promising lead for the development of a highly general and enantioselective Pictet–Spengler reaction, and will hopefully prove useful in the identification and discovery of general chiral catalyst systems for other transformations of interest.

References and Notes


**Acknowledgments:** We thank J. Gair, R. Klausen, R. Liu, A. Makarov, L. Nogle, and E. Regalado for helpful discussions.

**Funding:** NIH grant GM043214 (E.N.J.), NSF predoctoral fellowship DGE1745303 (C.C.W.)

**Author Contributions:** C.C.W, E.E.K, and E.N.J conceived the work, C.C.W., S.E.M., and E.E.K. designed and validated the analytical method, C.C.W. and S.E.M. conducted high-throughput screens, E.E.K and E.N.J. supervised, and all authors wrote and edited the manuscript.

**Competing interests:** Authors declare that they have no competing interests.

**Data and materials availability:** All data are available in the main text or the supplementary materials. The Python library used for SFC-MS peak deconvolution and analysis is available on Github under the GPL 3.0 license (https://github.com/corinwagen/chromatics).

**Supplementary Materials**

Materials and Methods

Supplementary Text

Figs. S1 to S47

Tables S1 to S10

References (45–48)
Supplementary Data