A Data-Driven Approach to the Development and Understanding of Chiroptical Sensors for Alcohols with Remote γ-Stereocenters

Jordan J. Dotson[†], Eric V. Anslyn[‡]*, and Matthew S. Sigman[†]*

[†]Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States
[‡]Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States

ABSTRACT: Dynamic covalent chemistry-based sensors have recently emerged as powerful tools to rapidly determine the enantiomeric excess of organic small molecules. While a bevy of sensors have been developed, those for flexible molecules with stereocenters remote to the functional group that binds the chiroptical sensor remain scarce. In this study, we develop an iterative, data-driven workflow to design and analyze a chiroptical sensor capable of assessing challenging acyclic γ-stereogenic alcohols. Following sensor optimization, the mechanism of sensing was probed with a combination of computational parameterization of the sensor molecules, statistical modeling, and high-level density functional theory (DFT) calculations. These were used to elucidate the mechanism of stereochemical recognition and revealed that competing attractive non-covalent interactions (NCIs) determine the overall performance of the sensor. It is anticipated that the data-driven workflows developed herein will be generally applicable to the development and understanding of dynamic covalent and supramolecular sensors.

Introduction

The development of dynamic covalent chemistry (DCC) sensors can fundamentally impact the disparate fields of biology, medicine, and materials science.[1–4] At the heart of successful DCC-based sensors is a robust molecular recognition strategy that involves the binding of and the chemical or conformational response to the analyte. Recognition of molecule type is often comparatively simple to engineer since it relies on predicable functional group-specific reactions (e.g., nucleophilic attack of an alcohol on an iminium).[5] In contrast, a difficulty of sensor design often stems from having to discriminate between different molecules bearing the same or similar functional groups or having to respond to specific analyte stereochemistry.[3] Solving these types of challenges often involves creating sensors that recruit the subtle additive effects of modest (<2 kcal/mol) noncovalent interactions (NCIs). However, the development and understanding of these types of sensors is hampered by the difficulty of deconvoluting individual effects of the multiple, competing NCIs involved.

As examples, the challenge of designing sensors based on dynamic covalent chemistry coupled with weak NCIs is evident in the field of chiroptical sensing (Figure 1). Many chiroptical sensors **1** feature a configurationally labile set of CD-active chromophores that interact via exciton coupling (e.g., + and – configurations, depicted in red and blue, respectively).[2] In the absence of a chiral analyte, the (+)-**1** and (–)-**1** configurations are enantiomeric and exist in a 1:1 equilibrium ratio. As a result, the Cotton effects from each enantiomer of the sensors cancel with no net CD signal observed. However, upon binding an enantioenriched analyte, the NCIs between the analyte **2** stereocenter and the optically active chromophores (depicted with a grey arrow) can lead to a net CD signal, which in turn is used to determine the enantioenrichment of **2**. When the stereocenter is adjacent to the binding functional group (**2**, n = 0), repulsive steric interactions influence the relative twist of the chromophores, often leading to a substantial CD-signal (e.g., > 10 mdeg). When the stereocenter is remote (**2**, n ≥1) the steric interactions necessarily diminish, and multiple weak NCIs (including attractive interactions) could presumably induce a measurable CD-signal. However, there are a dearth of reports for chiroptical sensing of remote stereocenters.[6]
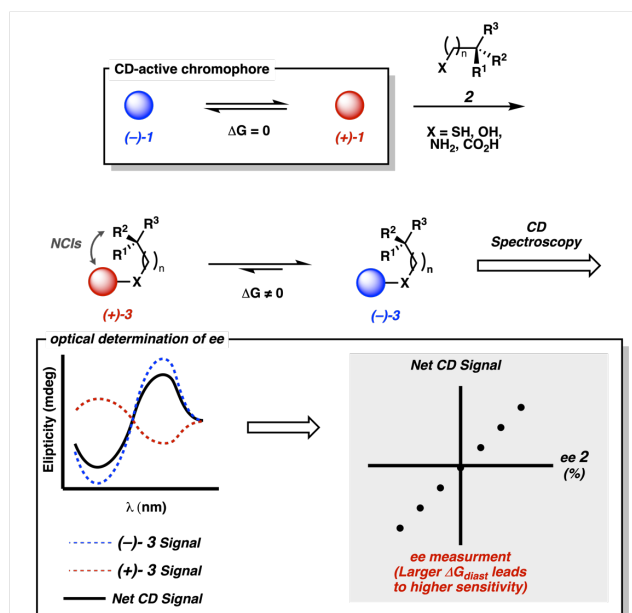

Figure 1. Covalent binding of an enantioenriched analyte **2** induces a conformational bias in the chiroptical sensor. The resultant CD-signal can be used to determine the enantioenrichment of **2**.

One of our groups has been particularly interested in developing chiroptical sensors of enantioenriched alcohols (Figure 2). We have shown that 2-formylpyridine (**4**) can efficiently react with dipicolylamine (**5**) and Zn(OTf)₂ in the presence of an acid catalyst (chloroethylmorpholine hydrochloride, CEM•HCl) to form hemiaminal **6** (Figure 2A).[2,7] The pyridine rings in **6** adopt enantiomeric CD-active helical configurations (P and M, depicted in red and blue, respectively). Upon binding of a chiral alcohol **7**, a series of diastereomers form by virtue of the pyridine helical twist, the chiral alcohol, and the newly formed point stereocenter of the hemiaminal ether functional group. If the alcohol is

1

enantioenriched, one configuration at the hemiaminal ether group dominates, which induces a bias toward a *P* or *M* twist of the pyridines, leading to exciton coupling and thus an observable CD signal. The magnitude of the CD signal can be correlated through a linear free energy relationship to differences in steric size of the R-groups on the stereocenter of the alcohol.[8] This sensor assembly has proven to be robust in detecting the enantioenrichment of α-stereogenic alcohols in many settings.[9–15]

More recently, we developed complex **8** that is capable of sensing the more challenging β-stereogenic alcohols (Figure 2B).[16] The design strategy hinged on the incorporation of an "appendage" off the pyridine ring closest to the hemiaminal group that could interact with the more remote stereocenter. While the assembly formation is identical to **9**, the chromophore responsible for sensing the chirality of the alcohols is an atropisomeric biaryl motif (depicted in red and blue). Importantly for the present work, however, designing sensors for the more challenging γ-stereogenic alcohols **11** remains a tremendous challenge (Figure 2C).
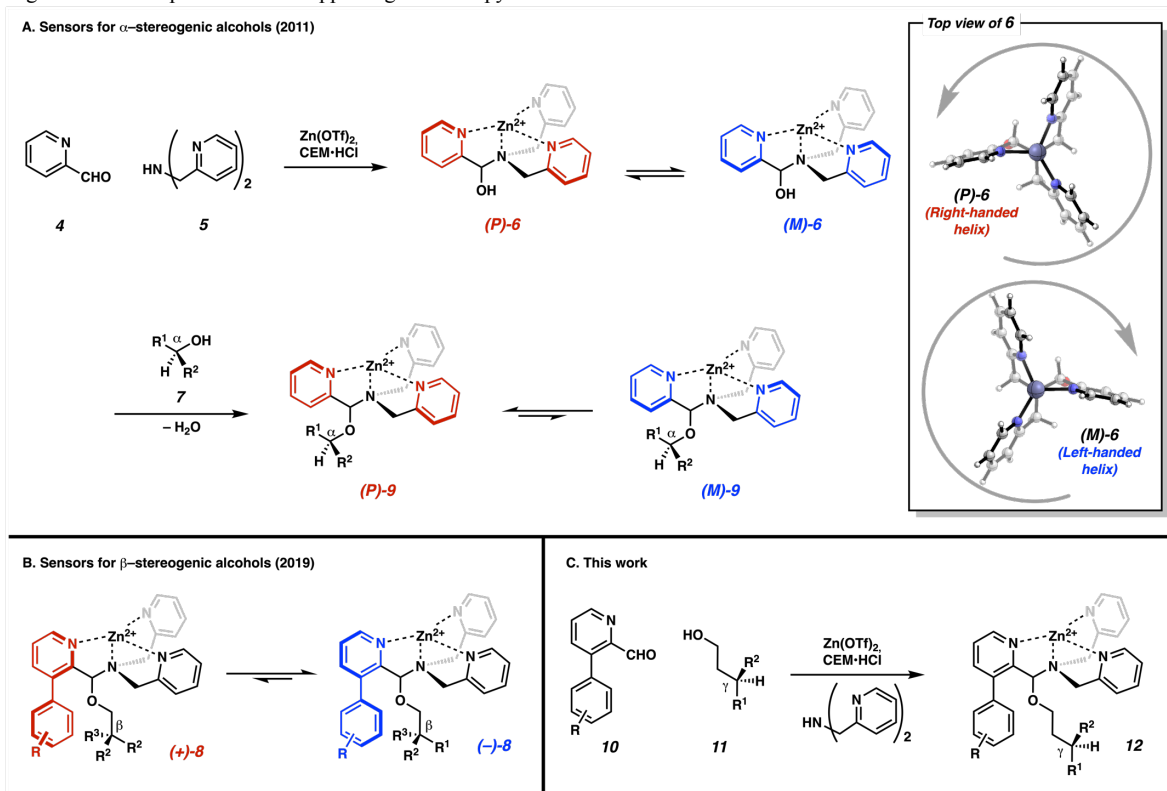

Figure 2. (A) The reversible covalent binding of enantioenriched **7** by assembly **6** favors one helically chiral configuration of the terpyridine motif. (B) Assembly **8** can be used to sense the enantioenrichment of β-stereogenic alcohols through induced atropisomerism of a biaryl axis. (C) This study focuses on the development of a chiroptical sensors for γ-stereogenic alcohols.

Merging the need to develop sensors for γ-stereogenic alcohols with the challenges of identifying what structural features are required to achieve this goal, we sought to develop an integrated synthetic and data science-based workflow (Figure 3). With such a remote stereocenter, significant steric interactions with the three pyridines or the atropisomeric chromophore were anticipated to be minimal, while the possibility of attractive NCIs could arise that would place a twist into either chromophore.[17] Ideally, this workflow would allow one to simultaneously optimize, as well as understand, the performance of chiroptical sensors that require multiple weak NCIs to function effectively. Four key stages were envisioned: 1) a simple machine-learning (ML) enabled strategy would be used to carefully identify highly diverse appendages (blue in structure **10**) that are synthetically accessible, 2) the selected compounds would then be synthesized and tested, 3) the resultant data would be used to train statistical models that correlate structural features to CD signal intensity, and 4) interrogation of the models would provide insight to inform next generation designs. This workflow is iterative in nature, wherein the mathematical model is applied to quantitatively predict the performance of new assemblies and ultimately deconvolute the noncovalent interactions critical to the success or failure of the chiroptical sensor. Herein, we present the successful application of this workflow to the identification and interrogation of chiroptical sensors of γ-stereogenic alcohols. Further, we were able to bolster the insight acquired from statistical modelling with DFT-level calculations to probe the underlying NCIs responsible for the observed sensor performance.
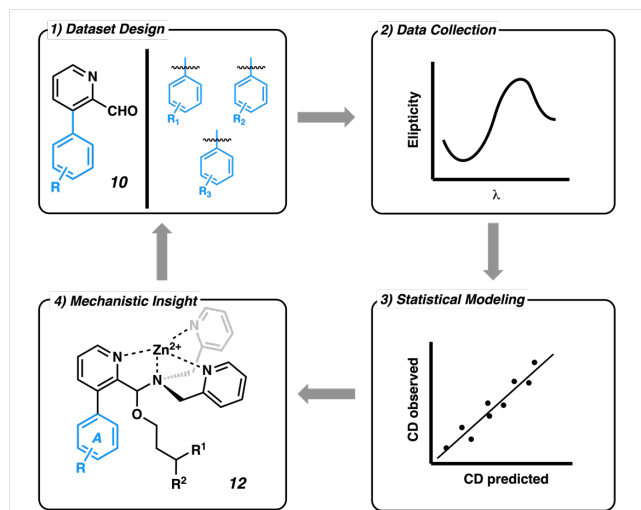
Figure 3. Data-science workflow for the development and understanding of chiroptical sensors.

Results and Discussion

Initial Design and Evaluation

To effectively use this workflow to design chiroptical sensors for γ-stereogenic alcohols, we had to select which modular structural element of the assembly to systematically vary (Figure 4A). We identified the A-ring (highlighted in blue) as ideal. Diverse moieties could be easily introduced using a metal-catalyzed cross-coupling reaction and would also presumably be well positioned to interact with the γ-stereocenter. To ensure that the biaryl aldehydes **10** selected for evaluation are chemically diverse, we employed a data-driven substrate selection strategy (Figure 4B). This was comprised of an initial generation of a data-rich virtual library of arylboronic acids **13** followed by a machine learning-based selection strategy.[18] The virtual library was first curated by computationally assessing all commercially available and in-house aryl boronic acids (a total of 5,136 boronic acids) (Step 1). To accomplish this, the geometry of each substrate was minimized and then 1,357 computationally inexpensive 2D and 3D Quantitative Structure Activity Relationship (QSAR) parameters - simple molecular descriptors routinely used in medicinal chemistry - were calculated for each boronic acid (Step 2, see SI for details).[19] With this computational data set, the boronic acids were analyzed based on their similarity in descriptor space by performing principal component analysis (PCA) to reduce the dimensionality of the computed descriptors. A small, representative set was then selected for biaryl aldehyde synthesis by employing K-means clustering.[20,21] This machine learning algorithm is used to group or "cluster" structures based on structural similarity (Step 3). Through this process, nine boronic acids (described below) were identified from distinct clusters, procured, and subjected to a Suzuki cross-coupling with aryl bromide **14** to produce the initial library of biaryl aldehydes **10a** – **10k** (Step 4).[22]

We note that the selection strategy depicted in Figure 4B resulted in a small, highly diverse library of candidates. Intuitively, this is inferred from a simple inspection of the structures and substitution patterns of **10a** – **10k**. The library features both electron rich and electron poor arenes (e.g., **10a** and **10d**, respectively), substrates with extended π surface area (**10f**, **10g**, and **10h**), and a multitude of substitution patterns (e.g., **10c**, **10e**, and **10k**). Moreover, the chemical diversity resulting from our workflow can be visualized (each member plotted as a blue point) using a Uniform Manifold Approximation and Projection (UMAP) visualization plot (see plot depicted in step 3).[23] With UMAP, similar boronic acids are positioned close to one another in their descriptor space as demonstrated by their grouping in "islands". The compounds that were selected from our workflow (plotted with yellow circles) are well spread across the representation, reflecting the statistical diversity of our library. It should be noted that the UMAP axes enable

visualization of the variance in the two plotted dimensions, similar to a principal component representation.
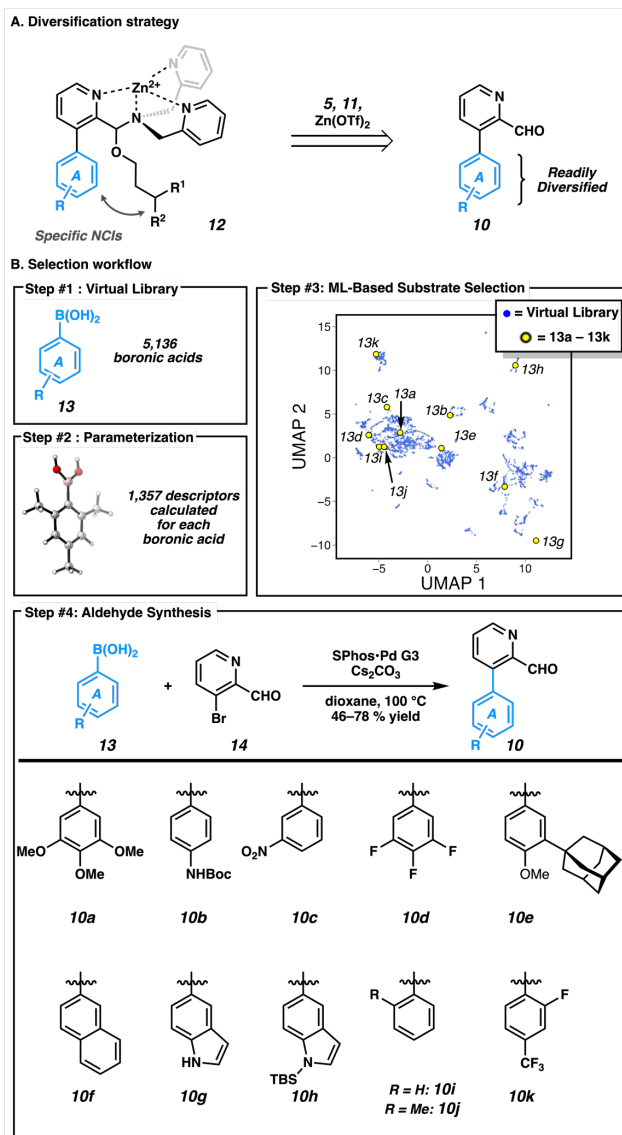

Figure 4. (A) The strategy to diversify assemblies **12** is enabled by the modular introduction of diverse A-rings in aldehyde **10**. (B) The data-science-based workflow to select diverse A-ring motifs begins with curation and parameterization of an arylboronic acid virtual library (1-2), followed by (3) ML-based substrate selection, and (4) synthesis of biaryl aldehydes **10**.

We next turned our attention to the evaluation of assemblies **12**, which were readily formed from the reaction of biaryl aldehydes **10** with Zn(OTf)$_2$, dipicolylamine, and an enantioenriched γ-stereogenic alcohol (either **11α** or **11β**, Figure 5A).[24] The resultant assemblies **12α** and **12β** were assessed via CD spectroscopy (subscript α and β denote the use of alcohols **11α** and **11β**, respectively). In nearly all cases, we observed a common Cotton effect at 267 nm (see representative spectra, Figure 5B).[25] On the basis of previous reports, it was concluded that this feature arose from exciton coupled circular dichroism (ECCD) of the helical pyridine chromophore (depicted in blue).[15,15] The intensity of the CD signal varied substantially depending on the structure of **12**, with signals of up to 22.7 mdeg observed for **12dα** and near-zero signals for assemblies **12bβ**, **12gβ**, and **12hβ** (Figure 5C).[26] The diversity of responses is likely a result of the intentional incorporation of disparate

structural features as a function of the data-driven design strategy. Additionally, no trends relating the structure of **12** to the CD spectrum were intuitively obvious.
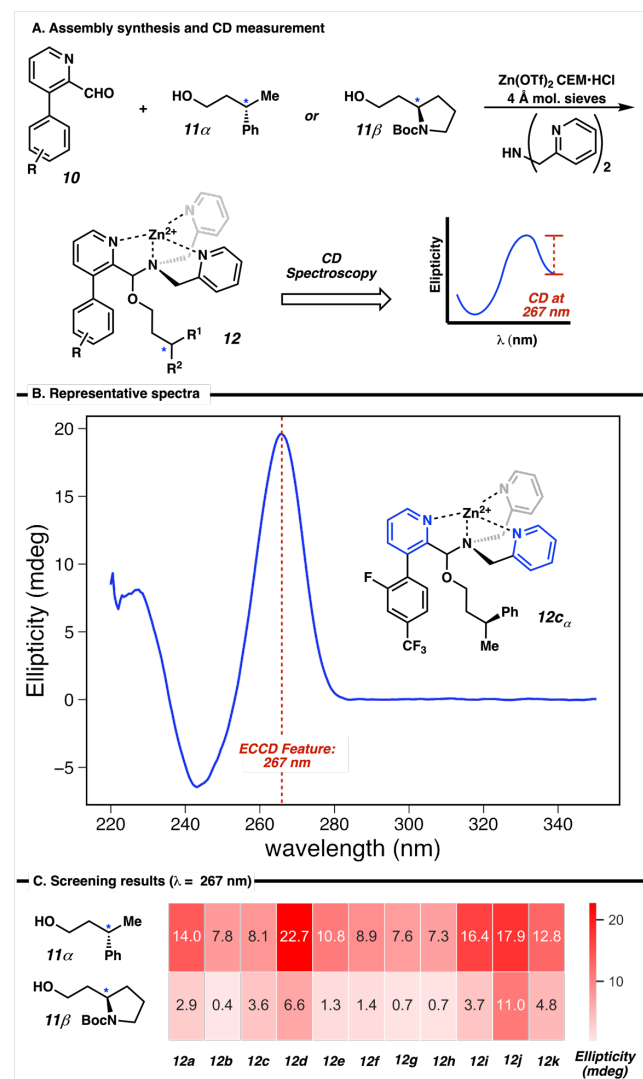

Figure 5. (A) Synthesis and CD measurements of assemblies **12**. (B) The representative spectrum of assembly **12c$_\alpha$** shows clear Cotton effect at 267nm. (C) The heat map depicts the CD signal (mdeg, normalized for absorbance) at the λ_{max} (267 ± 2 nm) for assemblies **12a$_\alpha$** – **12k$_\alpha$** and **12a$_\beta$** – **12k$_\beta$**.

In order to correlate CD intensity to the structural features of **12**, we next employed a two-step statistical modelling protocol (Figure 6A &

B).[27–30] This was accomplished by computing a range of physical organic descriptors to reflect the physical properties of **12a** – **12k**. These were computed using a simplified chemical structure **15**, which captured the essential structural elements anticipated to influence the CD response to the chiral alcohols.[31] The descriptors included sterimol values (steric measurements),[32,33] global electronic terms such as HOMO/LUMO energies and polarizability, as well as local electronic terms reflected by NBO charges on various atoms (see SI for full parameter list). Using a forward stepwise linear regression algorithm to correlate experimental CD intensities to computed parameters, a four-term statistical model was found (Figure 6A).[34] This model included one parameter to account for the alcohol **11α**/**11β** and three derived from the biaryl pyridine motif. The alcohol parameter (classifier) was a simple unitless value (1 for **11α**, –1 for **11β**), used to assign which alcohol was used in the assembly. The remaining three parameters, NBO$_{ipso}$, ΔE_{FMO}, and ortho B5 (a sterimol value), were derived from **15** (Figure 6B).[35] While the parameters NBO$_{ipso}$ and ΔE_{FMO} had a relatively even distribution of values across **15a** – **15k** (see SI), the B5 (Sterimol parameter reflecting size) for the ortho substituent had only three discrete values. Assemblies stemming from biaryl aldehydes with H at both A-ring ortho positions (**12a** – **12i**) all had an ortho B5 of 1.1Å and only **12j** and **12k** had larger measurements. Therefore, the parameter ortho B5 can be considered a classification term that reflects whether **12** has an ortho substituent.

The first iteration of the workflow accomplished one of our two initial goals, i.e., it identified which biaryl aldehyde (**10j**) would lead to assemblies with large enough CD signals to accurately determine the *ee* of both **11α** and **11β**. To our knowledge, this is the first chiroptical sensor that can assess a stereocenter so remote to the binding functional group. However, the second primary goal, to deconvolute the NCIs that underpin a successful sensor, was not yet realized. The model depicted in Figure 6A is statistically sound, but the descriptor terms do not give clear insight into the NCIs that determine the magnitude of the CD signal. For instance, it is challenging to understand the physical significance of the NBO charge on the ipso carbon (NBO$_{ipso}$, highlighted with a green sphere). Furthermore, the need to classify two aldehydes that incorporate an ortho-substituted A-ring (only **10j** and **10k**) reduces the utility of the statistical modelling approach.

As a result, we revisited the workflow discussed above for a second iteration with the goal of exploring regions of chemical space that should be further sampled. Specifically, the under representation of ortho substituents in the initial library led us to the hypothesis that incorporation of additional examples would enhance our application of regression analysis. In turn, the improved regression model would likely provide a more detailed understanding of the origin of the NCIs at play in determining large CD signals.
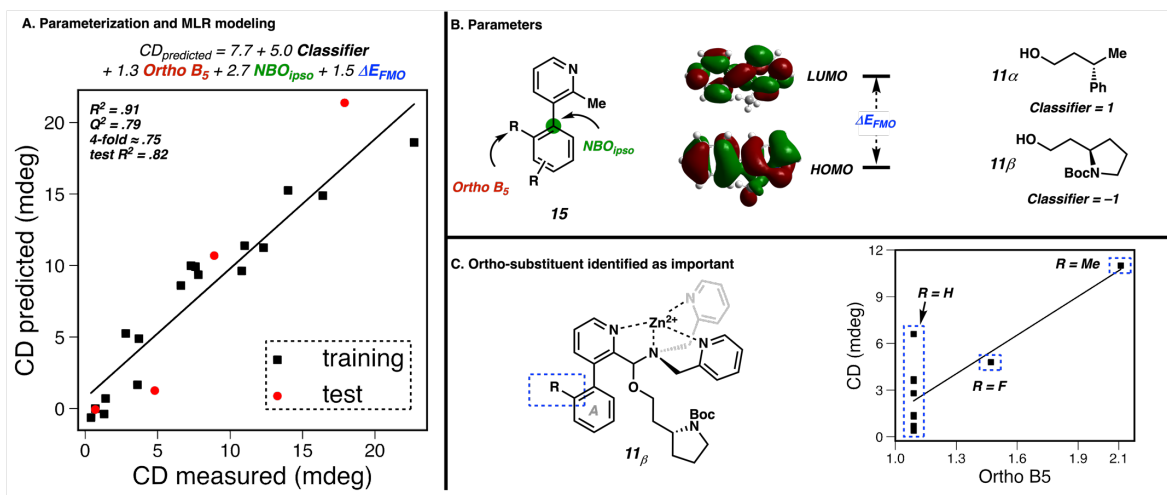
Figure 6. (A) First-generation MLR modeling of CD signal intensity for sensor **12**. (B) Alcohol classifier and DFT-parameters that were used to develop an MLR model to predict CD-signal intensity. (C) Role of A-ring ortho sterics on CD performance.

Second Generation

To better understand of the role of ortho substituents in the chiroptical sensing of γ-stereogenic alcohols, we next focused on the selection, synthesis, and evaluation of aldehydes **10l** – **10p** (Figure 7). Boronic acid precursors **13l** – **13p** were selected using the ML-based selection workflow described in Figure 4 (**13l** – **13p** depicted with orange diamonds, Figure 7A). It should be noted that this selection protocol focused on a narrower region of chemical space (ortho-functionalized compounds only) and that this is reflected by the close grouping of **13l** and **13n** – **13p** on the UMAP plot. The resultant assemblies **12l** – **12p** were then evaluated via CD spectroscopy. As was anticipated from the MLR model shown above, ortho-functionalized assemblies **12j** – **12p** tended to give comparatively higher CD intensities than **12a** – **12i** (Figure 7C). Nevertheless, the varied substituents in the aldehyde structures provided meaningful trends difficult to ascertain by simple inspection.

Once more, we turned to statistical modeling to assess assemblies **12a** – **12p** (Figure 8).[36] The best model obtained was able to adequately account for both assemblies **12α** and **12β** suggesting common NCIs intrinsic to the biaryl motif are similarly important whether alcohol **11α** or **11β** are used. The model featured a good correlation (R² = 0.91) and robust internal validation measures (Q² = 0.85 and 4-fold = 0.84). Furthermore, an external validation was prepared by initially partitioning the data into training and test set (80/20 split, see SI for details) and then evaluating the ability of the model to predict the test set data points. It was found that the model reliably predicted the CD signals of the test set (test R² = 0.91).

As in the previous iteration of MLR modeling, we evaluated a multitude of steric and electronic parameters in this second model (see SI for more details). Those parameters that were included in the final model are depicted in Figure 8B & 8C and include the same alcohol classifier and ΔE$_{FMO}$ terms used in the previous model, as well as two new steric terms, B$_{max}$ 2Å and B$_{max}$ 8.5Å. The latter two parameters are advanced sterimol descriptors recently introduced by Paton and coworkers.[37] The parameters were measured by first defining an axis along the pyridine C$_4$–C$_{1'}$ bond (depicted with a grey line, Figure 8C). The maximum and minimum steric measurements perpendicular to this axis (B$_{max}$ and B$_{min}$, respectively) were collected at 0.5 Å intervals along the axis. The B$_{max}$ measurements collected at 2.0 Å and 8.5 Å led to the best MLR model.
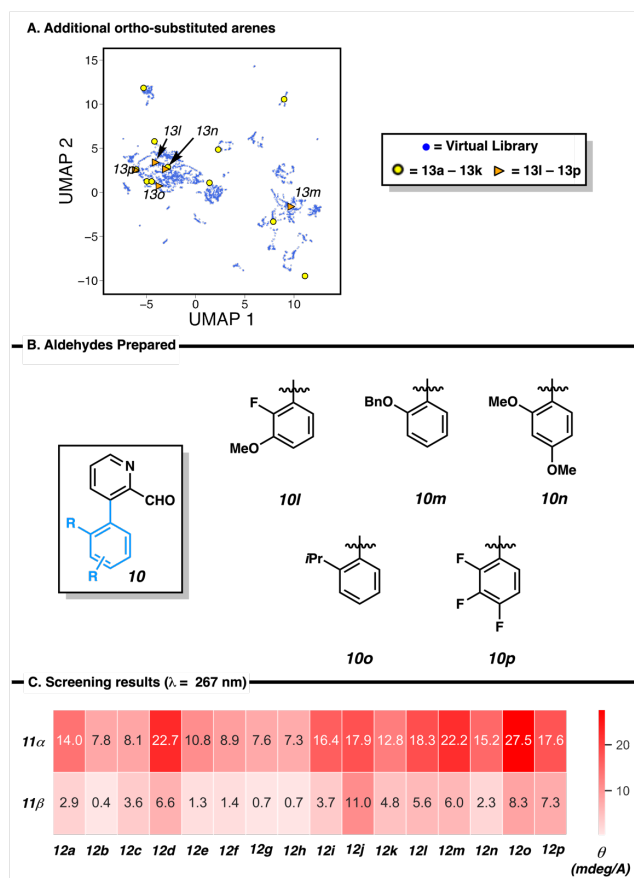

Figure 7. (A) UMAP plot depicting additional, ortho-substituted arylboronic acids **13l** – **13p**. (B) Chemical structures of ortho-functionalized aldehydes **10l** – **10p**. (C) The heat map depicts the CD signal (mdeg, normalized for absorbance) at the λ$_{max}$ (267 ± 2 nm) for assemblies **12a$_α$** – **12p$_α$** and **12a$_β$** – **12p$_β$**..
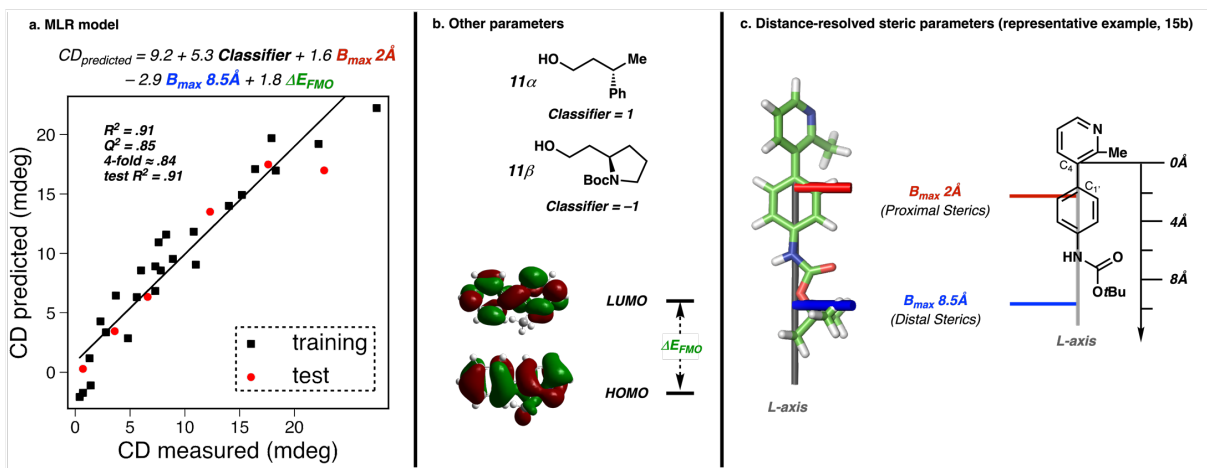
5

Figure 8. (A) Second-generation MLR modeling of CD signal intensity for sensor **12**. (B) Alcohol classifier and ΔE_{FMO} parameters. (C) Distance-resolved steric parameters that describe the steric profile of the A-ring.

In order to further validate the robustness of the statistical model, we synthesized and tested an additional validation set of 4-biaryl aldehydes **10q** – **10t** that were not used for statistical modeling (Figure 9). These were synthesized from boronic acids that were already on hand in our laboratory via a Suzuki cross-coupling reaction. The aldehydes were then tested with alcohols **11α** and **11β** and the CD spectra of the resultant assemblies were measured. The original model was then retrained using all assemblies **12a** – **12p**. We were pleased to find that assemblies **12q** – **12t** were well predicted by the model (validation R^2 = 0.81). It should be noted that the compounds in the validation set featured both electron-poor (**12t**), electron-rich (**12s**), and π-extended (**12q**) arenes and a common 3,5-disubstitution pattern that was absent from the aldehydes used for model generation. This highlights the ability of the model to effectively predict the performance of unique, out-of-sample substrates.

While the second model was only modestly better than the first in terms of regression statistics (R^2, test R^2, etc.), it was significantly more interpretable. Both the "hardness" term (ΔE_{FMO}) and alcohol classifier term were conserved across both models. However, the relatively opaque NBO_{ipso} value and the ortho-substituent classifier were replaced by two intriguing steric parameters presented above. These were particularly interesting because of their contrasting correlation to CD signal. While steric bulk was positively correlated to CD signal at 2.0 Å, it was negatively correlated at 8.5 Å.[38] As discussed below, these pronounced distance-resolved effects provided insight into subtle intramolecular NCIs within assemblies **12** that critically effect CD signal intensity.
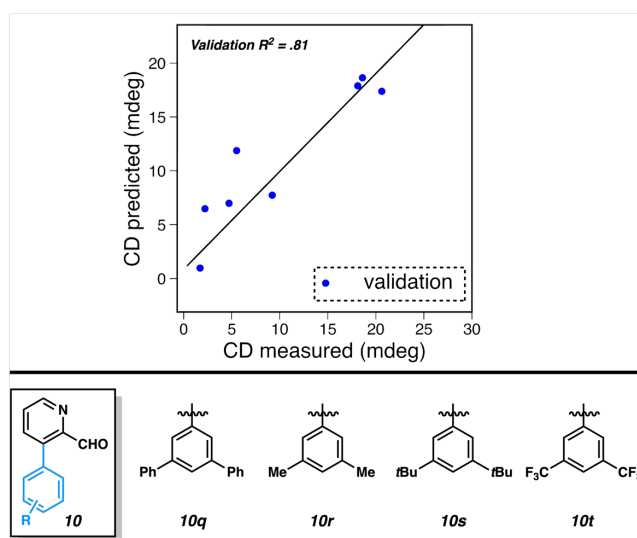

Figure 9. The second-generation MLR model accurately predicted 8 assemblies deriving from aldehydes **10q** – **10t** and alcohols **11α** and **11β** (R^2 = 0.82).

Computational Investigation

Armed with a statistical model composed of mechanistically interpretable parameters, we sought to answer the following question: what is the mechanism for the transfer of stereochemistry from the CD-silent point stereocenter of the alcohol to the CD-active terpyridine helix? As stereoenrichment of the terpyridine is the basis for the CD signal, an underlying understanding of this phenomena would provide a blueprint for what structural features make an effective sensor.

A key consideration relevant to this question was whether the configuration at the aminal stereocenter (R vs S) played a role in biasing the terpyridine configuration (M vs P). One of our groups had previously discovered a correlation (R^2 = 0.97) between the *dr* of **8** and the measured CD signal intensity (Figure 10A).[8,15] On the basis of this striking correlation, it was concluded that a point-to-point-to-helix mechanism of stereochemical transfer was operative (Figure 10B). This can be conceptually deconstructed into two steps: (1) the alcohol stereocenter (highlighted with a blue star) causes one of the hemiaminal ether epimers (stereocenter highlighted with a grey sphere) to be enriched. In turn, the configuration of the aminal influences the energetics to favor one of the two terpyridine configurational twists. It was determined that

the (*S*)-aminal led to (*P*)-helicity and the (*R*)-aminal led to (*M*)-helicity.[15] Therefore, the ratio of aminal epimers directly controls the magnitude of the CD signal.
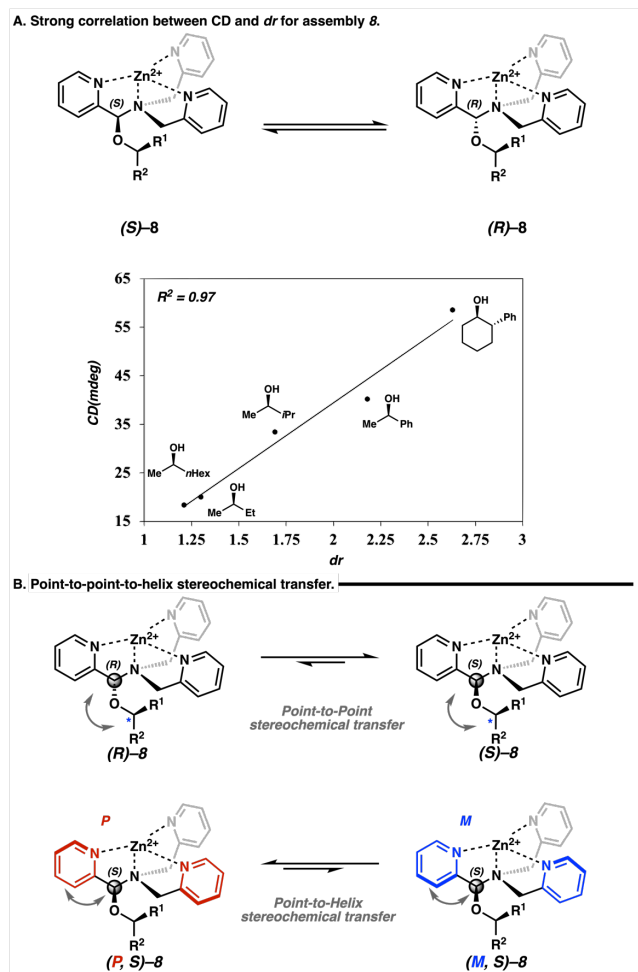

Figure 10. (A) Strong correlation is present between the aminal stereocenter and CD signal intensity for assemblies **8**. (B) The point-to-point-to-helix mechanism of stereochemical transfer from the configurationally inert point stereocenter to the helically chiral terpyridine stereocenter.

Given the good correlation previously observed between CD signal intensity and *dr* for assembly **8**, we questioned whether a similar relationship existed for the γ-stereogenic alcohol-derived assemblies **12** (Figure 11). In order to test this, we measured the *dr* of several assemblies **12_α** via ¹H-NMR spectroscopy as previous reported.[39,40] Although a similar range of values was observed (*dr* =1.1 – 2.2), there was no correlation with CD signal intensity. This suggested that either: (1) another variable was at play that obscured the correlation, or (2) the hemiaminal ether stereochemistry did not strongly influence the pyridine configuration and, instead, another mechanism for stereochemical transfer was involved.

In order to interrogate the mode of stereochemical transfer from point to helical chirality, we performed a computational study of assemblies **12b_α** and **12o_α**, as well as the previous assembly **8** (Figure 12).[41] Given the presumed importance of both proximal and distal steric effects revealed by MLR modelling, **12o_α** was selected because it had an ortho substituent on the A-ring, while **12b_α** was para-substituted. Additionally, **12o_α** and **12b_α** were the best and worst-performing assemblies tested during model construction respectively and would, therefore, reflect the limiting cases. Assembly **8** was selected as it would provide a benchmark into the origin of the impact of hemiaminal-ether stereo-

chemistry on the CD response. For each assembly analyzed, both aminal epimers were subjected to a gas-phase molecular mechanics-based conformational search using the OPLS3e forcefield.[42] From this, five distinct conformers were identified for **8** while the conformationally flexible **12b_α** and **12o_α** resulted in numerous conformers (92 – 145 per epimer). To limit computational resources, the latter two were clustered (17 – 32 clusters per epimer) based on atomic RMSD and one representative conformer was selected from each cluster. All conformers from **8** and the selected conformers from **12b_α** and **12o_α** were then refined via DFT calculations. Geometry optimization was conducted using the B3LYP functional with the 6-31G(d,p) basis set in Gaussian16.[43] Single-point corrections were then carried out with the M06-2X[44] functional and def2-TZVP[45] basis set. Solvation effects of acetonitrile were considered using the PCM solvation model.[46]
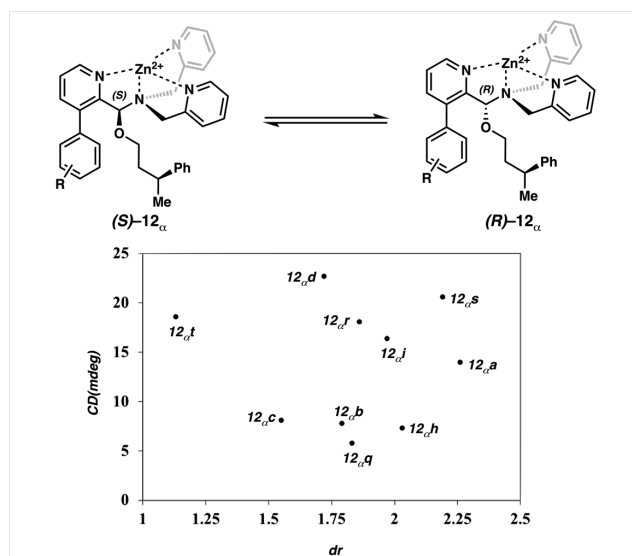

Figure 11. No clear relationship is present between hemiaminal ether stereocenter and CD signal for assemblies **12**.

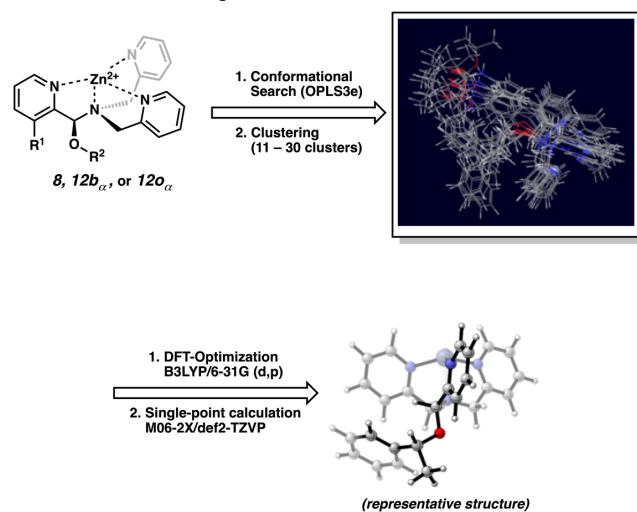*(representative structure)*

Figure 12. Computational workflow employed a molecular mechanics-based conformational search and then used a clustering algorithm to select 11 – 30 representative conformers. These were further refined with DFT optimization and single-point energy calculations. Note: the acetonitrile molecule that coordinates to Zn^{2+} was removed from the CYLview renderings for clarity.

The lowest-energy structures of both (*P, S*)–**8** and (*M, S*)–**8** (Figure 13) were analyzed revealing a large energy difference between the (*M*) and (*P*) configurations ($\Delta G_{M/P}$ = 2.2 kcal/mol).[47] This likely stems from the difference in orientation of the alkoxy group between the (*P*) and (*M*) conformers. (*P, S*)–**8** positions the C–O bond of the alkoxy group

(highlighted in blue) downward with respect to the N–Zn bond (highlighted in red). This positioning of the alkoxy group is termed axial and likely benefits from minimized steric interactions with the pyridine motifs.[48] In contrast, the alkoxy group is projected roughly perpendicular to the Zn–N bond in the **(M, S)–8** conformer. This placement of the alkoxy group, termed equatorial, is likely disfavored due to energetically unfavorable repulsive interaction with the pyridine rings. The substantial energetic preference for one terpyridine configuration is consistent with the notion that the hemiaminal ether stereocenter controls the helicity.
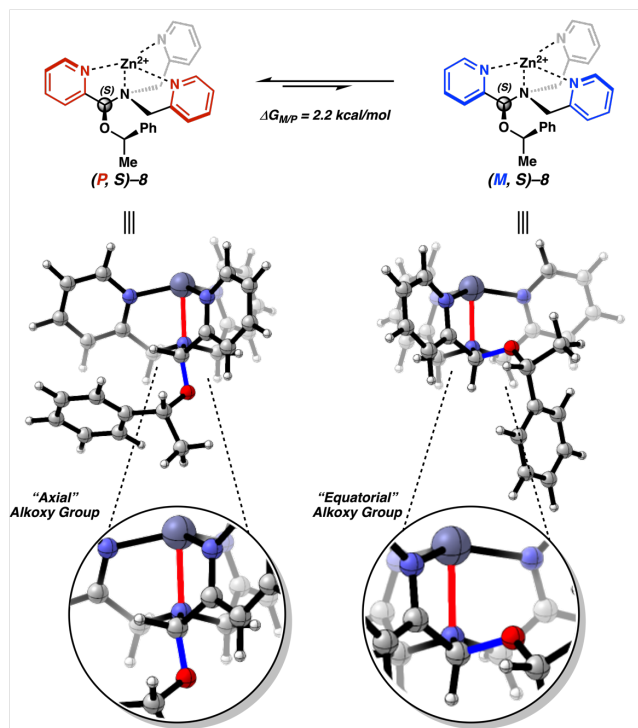

Figure 13. Discrete conformational ensembles for both (*M*) and (*P*) terpyridine configuration (the acetonitrile molecule coordinate to Zn^{2+} was removed from the CYLview renderings for clarity).

We next considered the computed structures for **12b$_\alpha$**, the worst-performing assembly tested during model generation (Figure 14). While we investigated both hemiaminal ether epimers, we will discuss **(R)–12b$_\alpha$** as it was calculated to be more stable than **(S)–12b$_\alpha$** (not shown) by ca. 1.8 kcal/mol. It was noted that the 30 conformers assessed at the DFT level of theory could all be categorized into two distinct conformational ensembles: one with (*M*) and one with (*P*) terpyridine configurations. We then compared the energies and structures of the lowest-energy conformer from each ensemble. As was the case for **8**, the configuration of the terpyridine dramatically impacted the orientation of the alkoxy group for **(R)–12b$_\alpha$**. In this case, the (*P*) and (*M*) configurations coincide with the alkoxy group being placed equatorial and axial, respectively. Unlike **8**, however, the energy difference between **(P, R)–12$_\alpha$** and **(M, R)–12$_\alpha$** were comparatively small ($\Delta G_{M/P} = 0.1$ kcal/mol). The low energy difference between helical conformers is consistent with the modest CD-signal observed experimentally.
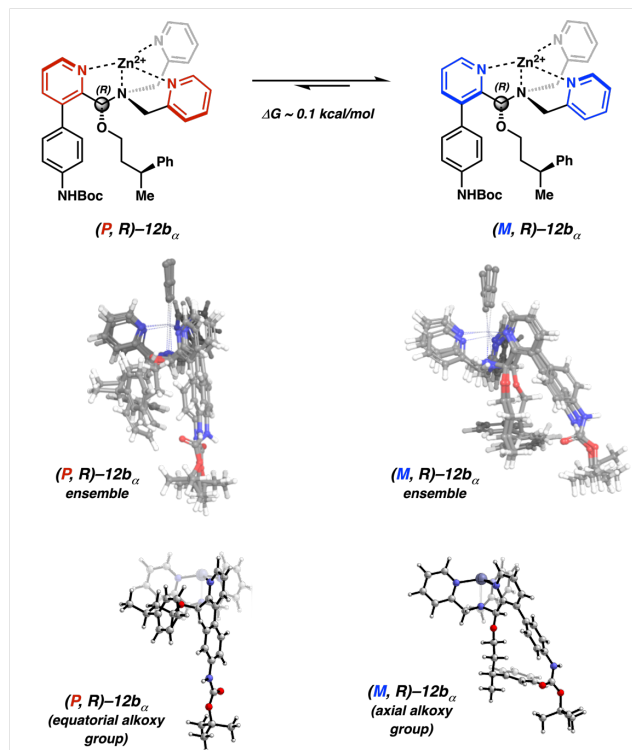

Figure 14. Assembly **12b$_\alpha$** features a modest energy difference between (*M*) and (*P*) configurations (CYLview renderings of the lowest-energy conformers from the (*P*) and (*M*) conformational ensembles are depicted). Note that the conformational ensembles are only depicted with three conformers for clarity.

Finally, we turned our attention to **12o$_\alpha$**, the best-performing assembly assessed in this study (Figure 15). Once more, we will only discuss the lowest-energy hemiaminal ether epimer, which was **(S)–12o$_\alpha$**. All conformers could again be categorized into two distinct conformational ensembles, each with an opposite pyridine helicity. The (*M*)-terpyridine configuration resulted from the equatorial alkoxy orientation while the (*P*)-configuration favored the axial alkoxy positioning. Unlike both **8** and **12b$_\alpha$**, in which the axial alkoxy orientation was favored, the conformer with the equatorial orientation of the alkoxy group **(M, S)–12o$_\alpha$** was preferred, now by 2.7 kcal/mol.[49] This comparatively large energy difference likely results in a substantial enrichment of the (*M*) terpyridine configuration and is consistent with the large CD-signal observed for **12o$_\alpha$**.
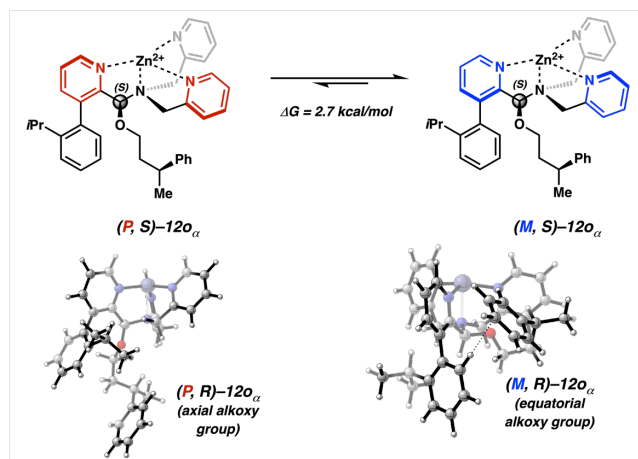
Figure 15. Assembly **12o$_\alpha$** strongly favors the (*M*) terpyridine configuration (CYLview renderings of the lowest-energy conformers from the (*P*) and (*M*) conformational ensembles are depicted). The acetonitrile molecule coordinated to Zn^{2+} was removed from the CYLview renderings for clarity.

Our combined computational and statistical modeling efforts suggest that attractive π–π interactions as well as London dispersion forces, play a key role in controlling CD-signal intensity by modulating $\Delta G_{M/P}$ (Figure 16). For assembly **(R)–12b$_\alpha$**, we observed attractive NCIs that stabilize both the (*P*)-equatorial and (*M*)-axial conformations. London-dispersion contacts can be seen in the former between the γ-phenyl ring and both the pyridine rings and the A-ring ortho-protons (see NCI plots, Figure 16B).[50,51] However, long-range dispersive interactions between the alkoxy chain and the A-ring para-substituent also stabilize the (*M*)-axial conformation. The presence of distal dispersive contacts (i.e., the 4-bocamino group), therefore, stabilizes **(M,R)–12b$_\alpha$** relative to **(P,R)–12b$_\alpha$** and attenuates $\Delta G_{M/P}$. In contrast, more significant attractive NCIs were observed for the (*M*)-equatorial conformer of **(S)–12o$_\alpha$** than in the (*P*)-axial conformer (Figure 16C). In addition to extensive dispersion interactions, the former benefits from a T-stacked interaction[52] between the A-ring ring ortho-proton and the γ-phenyl moiety (see dark blue attractive T-stacking interaction in the inset, Figure16C). The limited rotation enforced by the ortho isopropyl group results in an increased dihedral angle of the biaryl which, in turn, enhances this T-stacked interaction.

The computational analysis discussed herein, details the complexities of dynamic covalent sensors that exist as a mixture of multiple equilibrating diastereomeric and conformational isomers (i.e. **16**, Scheme 16A). While we have analyzed this in detail in the preceding paragraphs, the overarching conclusion can be summarized as follows. Substitution at the A-ring ortho and para positions has opposite, competing effects on the CD-signal intensity. Ortho steric bulk energetically favors one of the helical diastereomers by reinforcing a key T-stacking interaction (i.e., Figure 16C, inset). This leads to a large, net CD-signal from exciton coupling of the energetically preferred helical chromophore. Para steric bulk, in contrast, stabilizes the minor isomer leading to a smaller energetic preference for one terpyridine helix. This means that the exciton coupling of the terpyridine chromophores with opposite helicity will largely offset one another and only a weak net CD signal will be observed. The competing roles of para and ortho substituents are also reflected in the final MLR model (Figure 8). Steric bulk at the para position is reflected by the term B$_{max}$ 8.5Å and is negatively correlated to CD signal intensity, while steric bulk at the ortho position, reflected by B$_{max}$ 2Å, is positively correlated. Taken together this illuminates the following design principle. Installation of an ortho-functionalized A-ring and omission of para-substituents will maximize efficient stereochemical transfer from the hemiaminal ether stereocenter to the terpyridine helix and will thereby maximize the CD signal for chiral alcohols of this type.

CONCLUSION

The development of DCC-based sensors remains a broadly applicable endeavor within the field of supramolecular chemistry. However, the identification and deconvolution of subtle NCIs that are important to sensor performance remains an unsolved problem. In this study, we used a combination of computational parameterization, statistical modeling, and high-level DFT calculations to develop and gain a detailed understanding of the first reported chiroptical sensor for γ-stereogenic alcohols. By performing two iterations of a data-driven optimization workflow, we were able to identify a highly effective sensor and produce a robust, interpretable statistical model. This provided the basis to deconvolute the roles of distal and proximal steric bulk on sensor performance by mathematically relating the distance resolved sterimol parameters B$_{max}$ 2 Å and B$_{max}$ 8.5 Å to CD-signal intensity. We then performed high-level DFT calculations to understand the physical significance of the steric parameters. These calculations both revealed a likely sensing mechanism and suggested that both distal and proximal substituents influence sensor performance through attractive NCIs. This work demonstrates the effectiveness of an iterative data-driven approach to sensor design and showcases the utility of computational parameterization and statistical modeling to deconvolute competing weak NCIs. We anticipate that the workflows described herein can be readily adopted for the design of other DCC-based assemblies and in the design and understanding of new supramolecular systems at large.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Detailed experimental procedures and compound characterization data (PDF)

AUTHOR INFORMATION

Corresponding Author

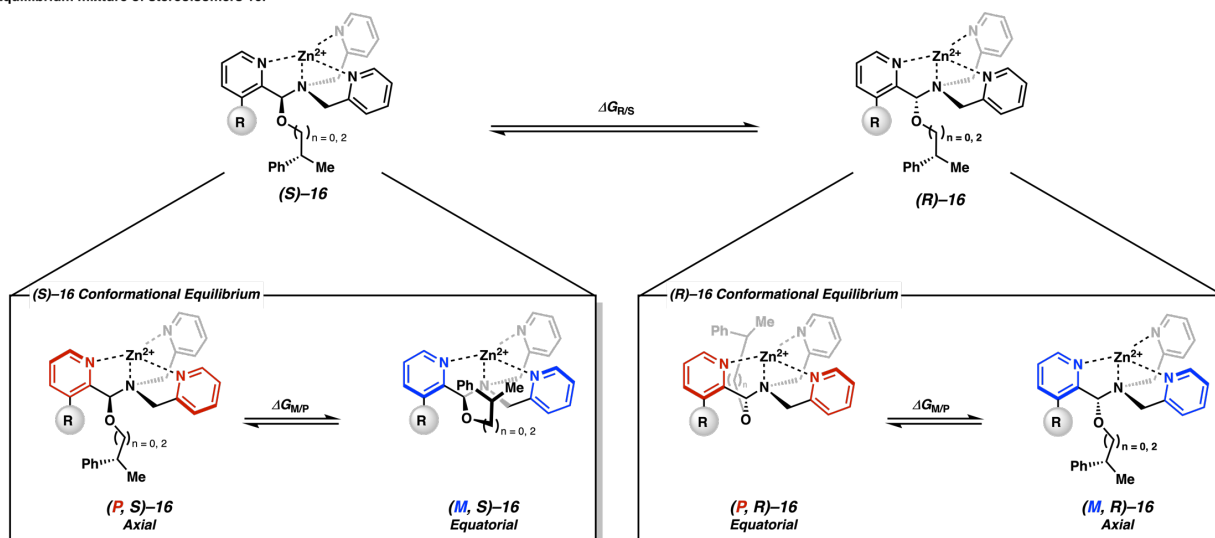*anslyn@austin.texas.edu
*matt.sigman@utah.edu

Notes

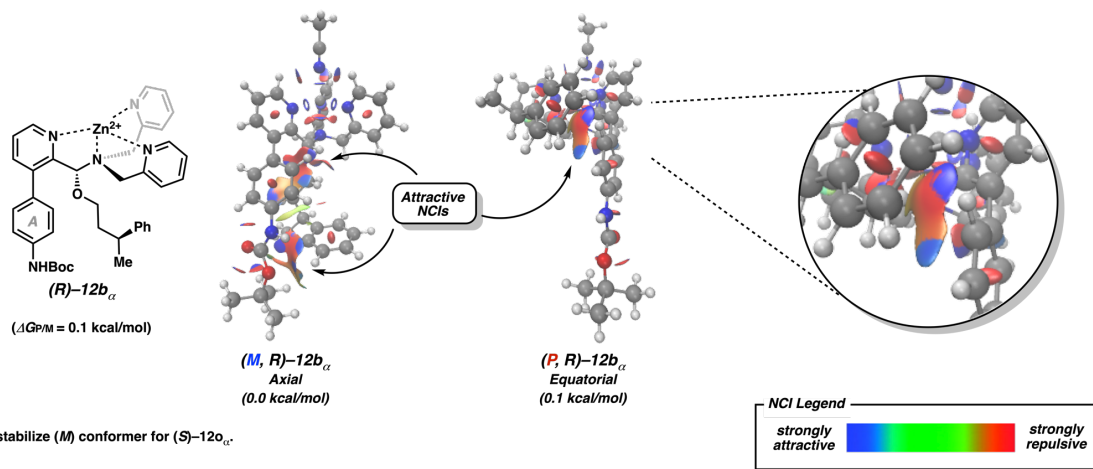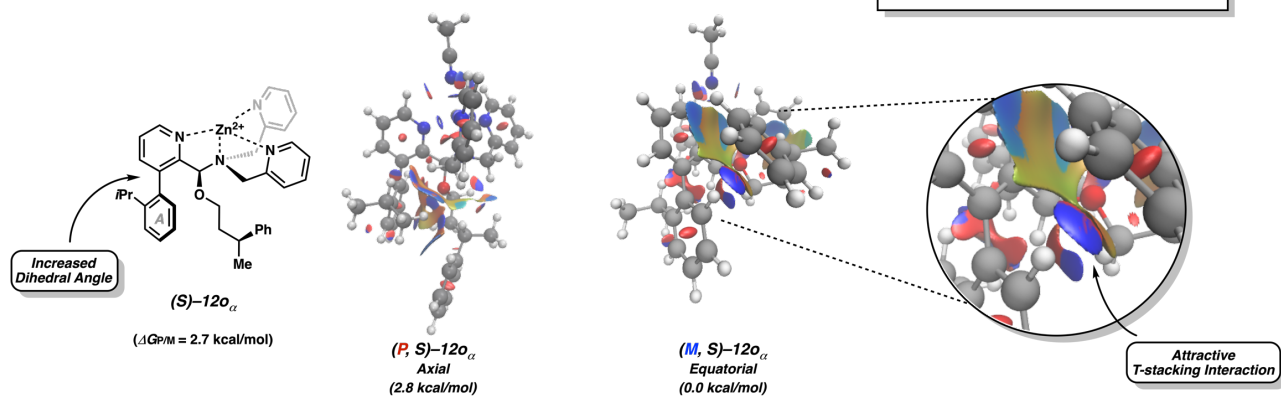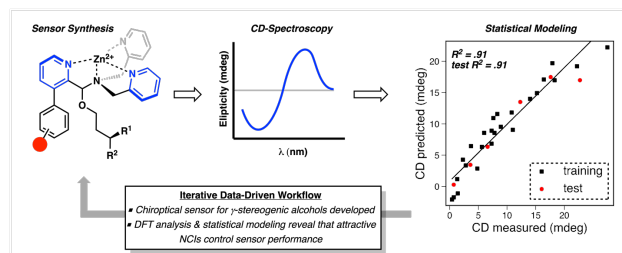The authors declare no competing financial interests.

ACKNOWLEDGMENT

Figure 16. (A) Equilibrium between (*R*) and (*S*) configurations of the hemiaminal ether stereocenter, (*M*) and (*P*) helicities of the terpyridine, and axial/equatorial orientation of the alkoxy moiety. (B) NCI plots of (*M, R*)–12b$_\alpha$ and (*P, R*)–12b$_\alpha$. (B) NCI plots of (*M, S*)–12o$_\alpha$ and (*P, S*)–12o$_\alpha$.

SYNOPSIS TOC

[1] Jin, Y.; Chao, Y.; Denman, R. J.; Zhang, W. Recent Advances in Dynamic Covalent Chemistry. *Chem. Soc. Rev.* **2013,** *42,* 6634–6654.

[2] Herrera, B. T.; Pilicer, S. L.; Anslyn, E. V.; Joyce, L. A.; Wolf, C. Optical Analysis of Reaction Yield and Enantiomeric Excess: A New Paradigm Ready for Prime Time *J. Am. Chem. Soc.* **2018,** *140,* 10385–10401.

[3] You, L.; Zha, D.; Anslyn, E. V. Recent Advances in Supramolecular Analytical Chemistry Using Optical Sensing. *Chem. Rev.* **2015,** *115,* 7840–7892.

[4] Zhang, X.; Yin, J.; Yoon, J. Recent Advances in Development of Chiral Fluorescent and Colorometric Sensors. *Chem. Rev.* **2014,** *114,* 4918–4959.

[5] Rowan, S. J.; Cantrill, S. J.; Cousins, G. R. L.; Sanders, J. K. M.; Stoddart, J. F. Dynamic Covalent Chemistry. *Angew. Chem., Int. Ed.* **2002,** *41,* 898–952.

[6] An example of an innovative chiroptical sensor for challenging substrates can be seen in the following reference: Wang, L.-L.; Quan, M.; Yang, T.-L.; Chen, Z.; Jiang, W. A Green and Wide-Scope Approach for Chiroptical Sensing of Organic Small Molecules through Biomimetic Recognition in Water. *Angew. Chem., Int. Ed.* **2020,** *59,* 23817–23824.

[7] You, L.; Berman, J. S.; Anslyn, E. V. Dynamic Multi-Component Covalent Assembly for the Reversible Binding of Secondary Alcohols and Chirality Sensing *Nat. Chem.* **2011,** *3,* 943–948.

[8] You, L.; Berman, J. S.; Lucksanawichien, A.; Anslyn, E. V. Correlating Sterics Parameters and Diastereomeric Ratio Values for a Mult-Component Assembly To Predict Exciton-Coupled Circular Dichroism Intensity and Thereby Enantiomeric Excess of Chiral Secondary Alcohols. *J. Am. Chem. Soc.* **2012,** *134,* 7126–7134.

[9] Lin C.-Y.; Lim, S.; Anslyn, E. V. Model Building Using Linear Free Energy Relationship Parameters−Eliminating Calibration Curves for Optical Analysis of Enantiomeric Excess. *J. Am. Chem. Soc.* **2016,** *138*, 8045–8047.

[10] Herrera, B. T.; Moor, S. R.; McVeigh, M.; Roesner, E. K.; Marini, F.; Anslyn, E. V. Rapid Optical Determination of Enantiomeric Excess, Diastereomeric Excess, and Total Concentration Using Dynamic-Covalent Assemblies: A Demonstration Using 2-Aminocyclohexanol and Chemometrics. *J. Am. Chem. Soc.* **2019,** *141,* 11151–11160.

[11] Jo, H. H.; Edupuganti, R.; You, L.; Dalby, K. N.; Anslyn, E. V. Mechanistic studies on covalent assemblies of metal-mediated hemiaminal ethers. *Chem. Sci.* **2015,** *6,* 158–164.

[12] Giuliano, M. W.; Lin, C. Y.; Romney, D. K.; Miller, S. J.; Anslyn, E. V. A Synergistic Combinatorial and Chiroptical Study of Peptide Catalysts for Asymmetric Baeyer–Villiger Oxidation. *Adv. Synth. Catal.* **2015,** *357,* 2301−2309.

[13] Jo, H. H.; Gao, X.; You, L.; Anslyn, E. V.; Krische, M. J. Application of a High-Throughput Enantiomeric Excess Optical Assay Involving a Dynamic Covalent Assembly: Parallel Asymmetric Allylation and ee Sensing of Homoallylic Alcohols. *Chem. Sci.* **2015,** *6,* 6747–6753.

[14] Lin, C.-Y.; Giuliano, M. W.; Ellis, B. D.; Miller, S. J.; Anslyn, E. V. From Substituent Effects to Applications: Enhancing the Optical Response of a Four-Component Assembly for Reporting ee Values. *Chem. Sci.* **2016,** *7,* 4085–4090.

[15] You. L.; Pescitelli, G.; Anslyn, E. V.; Di Bari, L. An Exciton-Coupled Circular Dichroism Protocol for the Determination of Identity, Chirality, and Enantiomeric Excess of Chiral Secondary Alcohols. *J. Am. Chem. Soc.* **2012,** *134,* 7117–7125.

[16] Minus, M. B.; Featherston, A. L.; Choi, S.; King, S. C.; Miller, S. J.; Anslyn, E. V. Reengineering a Reversible Covalent-Bonding Assembly to Optically Detect ee in β-Chiral Primary Alcohols. *Chem* **2019,** *5,* 3196–3206.

[17] Notably, the importance of attractive NCIs in stabilizing molecular conformations is well-documented in the area of asymmetric catalysis. For further discussion of this, see the following reference: Knowles, R. R.; Jacobsen, E. N. "Attractive Noncovalent Interactions in Asymmetric Catalysis: Links Between Enzymes and Small Molecule Catalysts," *Proc. Natl. Acad. Sci. USA* **2010,** *107,* 20678–20685.

[18] The strategy attaining diverse substrates through virtual library curation followed by a clustering-based selection was elegantly demonstrated in the following reference: Kariofillis, S. K.; Jiang, S.; Zuranski, A. M.; Gandhi, S. S.; Martinez Alvarado, J. I.; Doyle, A. G. Using Data Science to Guide Aryl Bromide Substrate Scope Analysis in a Ni/Photoredox-Catalyzed Cross-Coupling with Acetals as Alcohol-Derived Radical Sources. ChemRxiv. Preprint. https://chemrxiv.org/engage/chemrxiv/article-details/60ce1ef1e21133ea62dfc625.

[19] QSAR parameters were computed using the Mordred Python library. For further discussion, see the following reference: Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. *J. Chemoinf.* **2018,** *10,* 4.

[20] The clustering of molecules based on their descriptors to achieve high chemical diversity has been previously demonstrated in the field of medicinal chemistry. For further discussion, see: Pötter, T.; Matter,

H. Random or Rational Design? Evaluation of Diverse Compound Subsets from Chemical Structure Databases. *J. Med. Chem.* **1998,** *41,* 478–488.

[21] MacKay, D. J. C. An Example Inference Task: Clustering. In *Information Theory, Inference and Learning Algorithms*; Cambridge University Press, 2003; pp 284–292.

[22] It should be noted that **13i** and **13j** were not chosen using the workflow outlined in Figure 4. For further discussion see the SI.

[23] McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018,** *3,* 861.

[24] It should be noted that both enantiomers of alcohol **11β** were tested in several assemblies. As expected, equal and opposite Cotton effects were observed (see SI for more details).

[25] This Cotton effect λ_{max} deviated by ± 2 nm at the most, however, it typically only deviated by ± 0.5 nm.

[26] All reported CD-signals were normalized by dividing the measured CD-signal (mdeg) by the absorbance.

[27] Our group has extensively employed DFT-parameterization followed by multivariate linear regression for reaction development. For further discussion see references 28–30.

[28] Sigman, M. S.; Harper, K. C.; Bess, E. N.; Milo, A. The Development of Multidimensional Tools for Asymmetric Catalysis and Beyond. *Acc. Chem. Res.* **2016,** *49,* 1292–1301.

[29] Santiago, C. B.; Guo, J.-Y.; Sigman, M. S. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chem. Sci.* **2018,** *9,* 2398–2412.

[30] Crawford, J. M.; Kingston, C.; Toste, F. D.; Sigman, M. S. Data Science Meets Physical Organic Chemistry. *Acc. Chem. Res.* **2021,** *54,* 3136–3148.

[31] Parameters were acquired from truncated biaryl **15** instead of the full assembly **12** to minimize the computational resources used.

[32] Verloop, A. *Drug Design,* Vol. III; Ariens, E. J., Ed.; Academic Press: New York, 1976.

[33] Brethomé, A. V.; Fletcher, S. P.; Paton, R. S. Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters. *ACS Catal.* **2019,** *9,* 2313–2323.

[34] It should be noted that assembly **12cα** was an outlier and was not included in model development. It is likely that the nitroaromatic moiety engages in distinct π-π NCIs with the γ-phenyl ring of **11α** that are not accounted for by our models.

[35] For further information about parameter collection, see discussion in the SI.

[36] As in the modeling depicted in Figure 6, assembly **12cα** was an outlier and was excluded from model development.

[37] Luchini, G.; Paton, R. S. DBSTEP: DFT Based Steric Parameters. **2021,** DOI: 10.5281/zenodo.4702097.

[38] It should be noted that Paton's advanced distance-resolved sterimol descriptors were critically important as the simple sterimol measurements of the arene as a whole were poorly correlated to CD signal. The opposite correlations for proximal and distal encumbrance likely counteract one another when distance-resolved parameters are not used.

[39] The *dr* of assemblies **12** was conveniently measured by integrating the aminal methine proton, who's signal was well-resolved for almost all assemblies observed. The assemblies that did not feature two clear, well-resolved methine signals for the two diastereomers were not included in the plot. It should be noted that *dr* measurements were conducted using racemic **11α** due to cost considerations.

[40] We did not conduct a similar analysis of the **11β** derived assemblies, because the NMR-timescale rotation of the Boc motif broadened the ^{1}H NMR signals and obscured the direct measurement of *dr*.

[41] While we could have chosen assemblies deriving from either alcohol **11α** or **11β**, we chose the former because the added conformational dynamics of the latter's carbamate moiety would complicate analysis.

[42] Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019,** *15,* 1863–1874.

[43] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F.Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J.Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N.Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M.Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A.Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C.Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B.Foresman, D. J. Fox. *Gaussian 16, Revision A.03*; Gaussian, Inc.: Wallingford, CT, 2016.

[44] Zhao, Y.; Druhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008,** *120,* 215–241.

[45] Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005,** *7,* 3297–3305.

[46] Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005,** *105,* 2999–3094.

[47] The (*S*)-aminal diastereomer was chosen because it had been previously determined to be the most energetically favored. For further discussion, see reference 15.

[48] It is likely that the axial conformation is also weakly stabilized by an anomeric-like effect in which the nitrogen lone pair hyperconjugatively donates into the σ* orbital of the neighboring C–O bond. This stereoelectronic effect is strongly conformation dependent and will be absent in the equatorial conformer.

[49] The energy difference reported was obtained by measuring the energy difference between lowest-energy conformers from both the (*M*)- and (*P*)-terpyridine conformational ensembles.

[50] Conteras-Garcia, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W. NCIPLOT: A Program for Plotting Noncovalent Interaction Regions. *J. Chem. Theory Comput.* **2011,** *7,* 625–632.

[51] Johnson, E. R.; Keinan, S.; Mori-Sanchez, P.; Contreras-Garcia, J.; Cohen, A. J.; Yang, W. Revealing Noncovalent Interactions. *J. Am. Chem. Soc.* **2010,** *132,* 6498–6506.

[52] For a detailed discussion of T-stacked π–π interactions, see the following reference: Sinnokrot, M. O.; Sherrill, C. D. Substituent Effects in π–π Interactions: Sandwich and T-Shaped Configurations. *J. Am. Chem. Soc.* **2004,** *126,* 7690–7697.