Green Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Green Chem.*, 2025, **27**, 6831

Computer-aided design of stability enhanced nicotinamide cofactor biomimetics for cell-free biocatalysis†

Alexandra P. Platt, [a] Heidi Klem, [b] Sam J. B. Mallinson, [c] Yannick J. Bomble*[c] and Robert S. Paton *[a]

Received 21st January 2025,
Accepted 2nd April 2025

DOI: 10.1039/d5gc00351b

rsc.li/greenchem

Cell-free biocatalysis (CFB) is an efficient and environmentally friendly method to synthesize molecules such as pharmaceuticals, biochemicals, and biofuels through the *in vitro* use of enzyme cascades. These enzymes often require redox cofactors to drive chemical reactions. Natural redox cofactors (NAD(P)H) are expensive to isolate, motivating synthetic nicotinamide cofactor biomimetics (NCBs) as a cost-effective solution. A select handful of NCBs have been identified as potential NAD(P)H alternatives with comparable or improved redox capabilities, however, they display a tendency to degrade in common buffers. In this study, a library of 132 NCB candidates is systematically generated, over 85% of which have not been characterized in the literature, to expand the diversity of currently explored NCBs. The decomposition mechanism of NCBs in phosphate is evaluated using density functional theory (DFT), revealing protonation at the nicotinamide C5 position as a reporter of cofactor stability. Based on this result, we trained a linear regression model on DFT calculated descriptors to predict NCB stability in phosphate buffer, achieving mean absolute error (MAE) and root mean squared error (RMSE) values within computational accuracy. Analysis of key atomic descriptors and qualitative trends in our dataset informed the design of novel NCB candidates we propose with optimized stability. This work enables researchers to predict the relative stability of NCBs before synthesis, thereby streamlining the process to make CFB more affordable and viable at industry scales.

Green foundation

1. This work systematically expands the chemical diversity of nicotinamide cofactor biomimetics (NCBs) and develops a computational model to predict chemical stability in phosphate buffer from calculated descriptors. The design of NCBs is crucial for large-scale implementation of cell-free biocatalysis, a chemical synthesis strategy popular for its low environmental impact and high yields.
2. Our NCB stability model provides an efficient and accurate (MAE 0.61 kcal mol^{-1}) process to filter the candidate pool, informing molecule selection for experimental synthesis and characterization. This work revealed important stability design features: an aryl *N*-substituent on the pyridine ring, an electron withdrawing C3-substituent, and a methyl C5-substituent.
3. Future work will explore additional properties of nicotinamide cofactor biomimetics, such as redox potentials and enzymatic binding affinity, to further promote cell-free biocatalysis at an industrially viable scale.

Introduction

With rising demand for stereo-, chemo-, and regioselective chiral molecule synthesis, interest in biocatalysis has also grown.[1–3] In particular, cell-free biocatalysis (CFB), the *in vitro* use of enzymes in chemical synthesis, has gained popularity in recent years due to its higher yields and specificity and lower environmental impact than current synthetic techniques. Compared with other catalysts currently used in the synthesis of biofuels, enzymatic catalysts and CFB have many desired traits for green chemistry.[3,4] While the reaction rates may be slow, enzymes show high yields under mild reaction conditions, as well as easy separation and purification of products. Enzymes in CFB are also able to be reused, working as true catalysts rather than requiring stoichiometric reagents to produce biofuels or other desired chemicals.[5] Furthermore, the use of enzymes to catalyze chemical reactions instead of traditional

[a]*Department of Chemistry, Colorado State University, Fort Collins, CO, 80523, USA. E-mail: robert.paton@colostate.edu*
[b]*Thermodynamics Research Center, Applied Chemicals and Materials Division, National Institute of Standards and Technology, Boulder, CO, 80305, USA*
[c]*Biosciences Center, National Renewable Energy Laboratory, Golden, CO, 80401, USA. E-mail: yannick.bomble@nrel.gov*
† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d5gc00351b

catalysts ensures the use of safer solvents and auxiliaries, one of the 12 principles of green chemistry.[6] CFB also offers improved yields and resistance to toxicity relative to *in vivo* biocatalysis by removing cellular metabolic requirements and burdens.[7–9] This aligns with another principle of green chemistry, the reduction of waste required to synthesize these platform chemicals. Biosynthetic techniques, especially CFB, have therefore grown in popularity in recent years. However, the cost of CFB is still a significant drawback,[5] with enzyme cofactors representing a significant fraction of the reaction setup price tag. Approximately 50% of known enzyme reactions require cofactors, the most popular being nicotinamide cofactors, which can be prohibitively expensive and unstable outside of cells.[10–12] Pathways can be designed to recycle these redox cofactors, and sacrificial substrates can be included where this is not possible, allowing for their inclusion in catalytic rather than stoichiometric quantities, but this is not always enough to offset costs and make enzymatic syntheses economical.[13–16] As a result, the development and implementation of synthetic nicotinamide cofactor biomimetics (NCBs) is an active area of research that aims to facilitate CFB in industrial settings and at scale.[13,17]

NCBs are small molecules designed to mimic the redox role of natural nicotinamide cofactors, while providing advantages such as ease and low cost of manufacturing, and tunable enzyme specificity to enable bioorthogonal redox cascades.[13,18–20] Simple NCBs contain a central pyridine ring to maintain the hydride transfer ability of their natural counterparts (Fig. 1A), and various ring substituents, especially at the nitrogen position, are introduced to tune chemical properties. Fig. 1B displays sixteen well-characterized cofactor mimics found in the literature,[16,19,21] three of which contain variations in the R_3 substituent such that they are not nicotinamide structures, however, they are still classified as NCBs for their ability to replace NAD(P)H. Not only are NCBs able to replicate the catalytic reductive properties of NAD(P)H, but they can, in principle, be used in key immobilization processes like those using nanostructures or hydrogels.[22] Many ene-reductases and some oxidoreductase enzymes have been shown to readily accept NCBs as redox cofactors.[23] Other types of enzymes do not always naturally accept NCBs, either due to poor binding or incompatible reduction potentials, but activity can be engineered for these synthetic cofactors.[15,24–27]

The use of NCBs makes CFB more industrially feasible due to their reduced cost compared to the natural nicotinamide cofactors. However, many NCBs tested to date are also at risk of decomposing at the nicotinamide moiety in buffers used in CFB.[28,29] This decomposition risk also applies to immobilized NCBs in reactors. Due to the difficulty of replacing decomposed cofactors in reactors, closer examination of NCB decomposition is crucial to the implementation of NCBs for industrial uses. The proposed mechanism of nicotinamide decomposition by Alivisatos *et al.* is through addition across the 5,6-double bond in the 1,4-dihydropyridine ring, depicted in Scheme S1.†[29] This decomposition occurs most quickly in a potassium phosphate buffer,[14,16] which often yields high

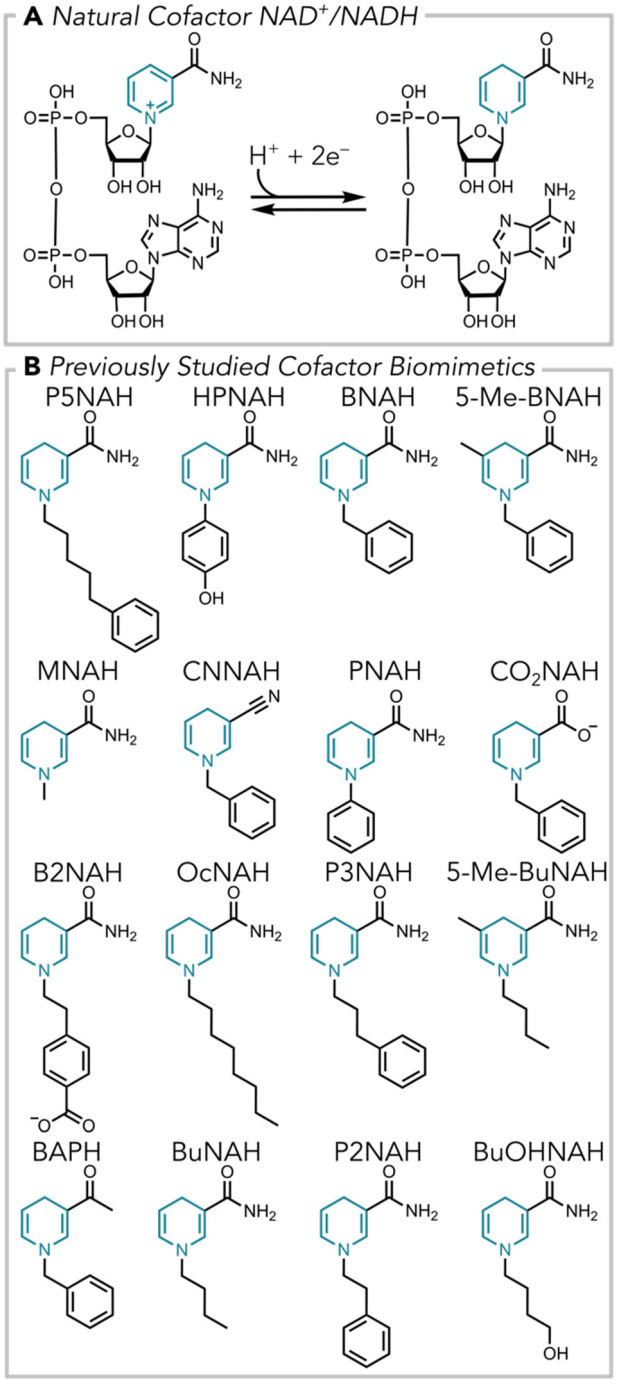

Fig. 1 (A) Reduction of NAD^+. (B) Selected NCBs previously seen in literature.[16,19,21] The 1,4-dihydropyridine in both NAD(P)H and NCBs, shown in teal, is featured as the one constant design component throughout this work.

enzyme activity.[23] To further promote the applicability of NCBs in CFB, we propose their stability should be included as a feature to optimize during engineering campaigns.

Another opportunity for improvement in the NCB field is expansion of the chemical space. Within the realm of simple NCBs, described by Black *et al.* as structures which do not

contain nucleotide components, there are vast possibilities for the R_1 and R_2 substituents (*N*-substituents and C3-substituents, respectively).[25] However, only a small number of NCBs have been synthesized and characterized *in vitro*. Tan *et al.* investigated chemical and physical properties of NCBs with a methyl R_3 substituent (C5-substituent), pointing out the lack of research into modifications at the R_3 position.[16] However, they proposed adding an R_3 methyl substituent to only two NCB structures of the eleven they report. Additionally, R_2 variants have only been examined for NCBs with a benzyl R_1 substituent. In our literature review, we found sixteen simple synthetic NCBs out of which eleven R_1, four R_2, and two R_3 substituents are explored (Fig. 1B).[16,19,21] By computationally expanding the possible NCB structures, we can examine a larger area of chemical space than previously available without the need for experimental characterization of potentially unstable structures.[3,30,31]

In this work, we expand the chemical space of NCB candidates through a combinatorial generation of simple NCBs from the previously evaluated substituents, resulting in a library of 132 NCB candidates (including the sixteen previously proposed) (Fig. 2, Step I). Quantum chemical calculations are used to provide a mechanistic understanding of NCB decomposition in phosphate and determine an approach to conduct high throughput estimates of NCB stability (Fig. 2, Steps II & III). Finally, we curate a dataset of quantum chemical descriptors for each NCB candidate (Fig. 2, Step IV). This dataset is subjected to qualitative and quantitative analysis and evaluated by numerous predictive model schemes (Fig. 2, Step V). We use the results of these analyses to design novel, more stable NCBs. Our work allows quick, high-throughput stability screening of an expanded space of NCB candidates to select top choices for downstream, more costly experimental validation.

Experimental

Conformational sampling was performed using CREST version 3.0[32] using the GFN2-xTB semi-empirical level of theory[33] with xTB version 6.7.0 (Fig. 2, Step II).[34] SMILES (Simplified Molecular Input Line Entry System) text representations of each NCB can be found at **https://github.com/aplatt22/ncb_stability** and in the ESI† and were converted to xyz-coordinate files using OpenBabel version 3.1.1.[35] For structures in our library, CENSO version 1.2.0[36] was used to further refine the ensemble before clustering into 10 or fewer representative structures. Transition state conformers were generated using

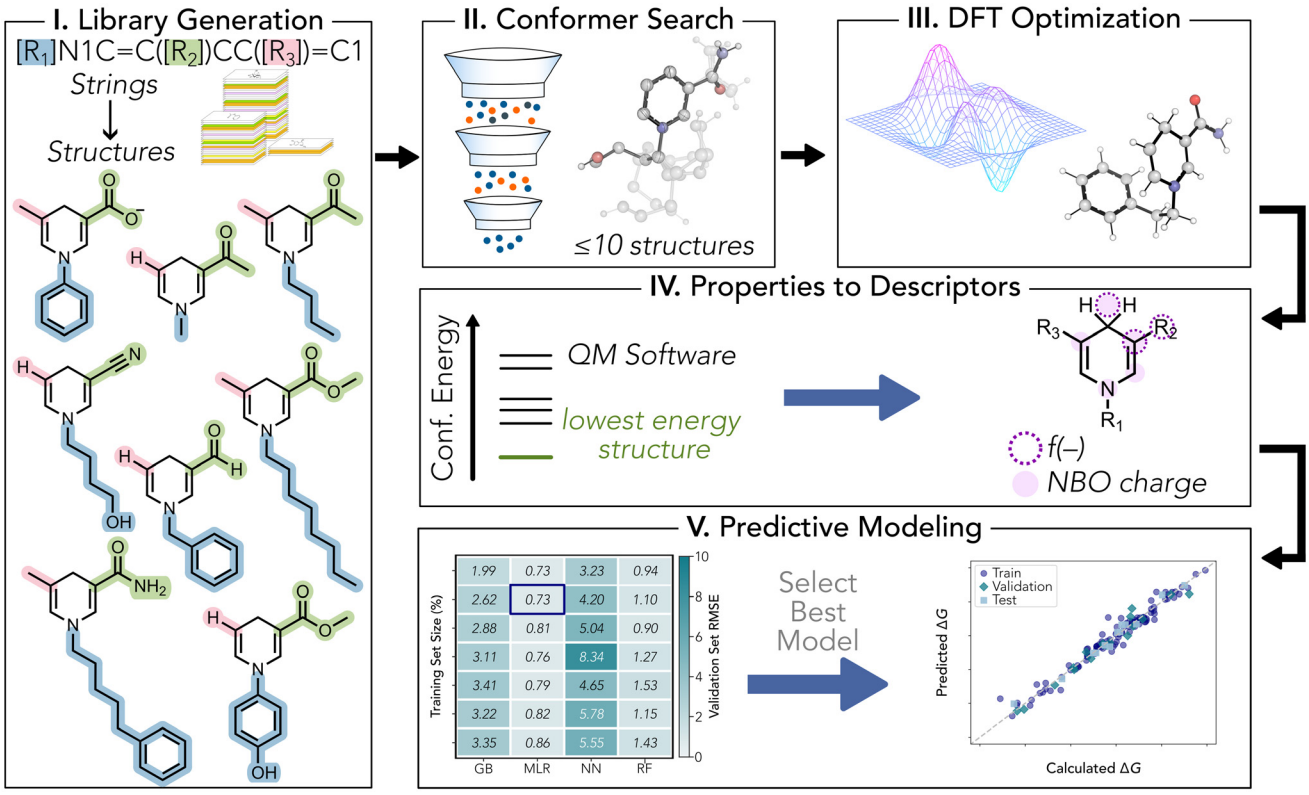

Fig. 2 Computational workflow presented in this work for the design of stability-enhanced NCBs. (I) Combinatorial addition of substituents at the R_1, R_2, and R_3 positions to generate a library of 132 NCBs. (II) Conformational sampling was performed on our NCB library using CREST and CENSO, followed by PCA and k-Means clustering to result in 10 or fewer conformers for each NCB. (III) DFT geometry optimization using PBE0-D3(BJ)/6-31+G(d)/SMD(water) on NCB conformers to obtain final structures. (IV) Calculation of atomic descriptors of the core 1,4-dihydropyridine ring and substituents, including atomic charges and condensed Fukui indices. (V) Predictive modeling for NCB stability, testing multiple models.

constrained CREST conformational sampling without CENSO to refine the ensembles. More details can be found in section 4 of the ESI.†

Density functional theory (DFT) calculations were performed using Gaussian 16, Revision C.01 (Fig. 2, Step III).[37] All geometry optimizations used PBE0-D3(BJ)/6-31+G(d)/SMD (water) as the level of theory.[38–44] Vibrational frequency analyses at the geometry optimization level of theory confirmed the nature of transition structures and ground state structures based on the presence of only one imaginary normal mode or none at all, respectively. Intrinsic Reaction Coordinate (IRC) calculations also used this level of theory to ensure the nature of transition structures during the mechanistic study. Additional single-point energy corrections with larger basis sets were used to obtain more accurate energies. For the calculation of all barrier heights, we used ωB97M-V/def2-TZVP/SMD (water) to obtain these energies.[38,45,46] Calculations which used the ωB97M-V functional were done using Orca version 5.0.3 [47] with support from libXC version 5.1.0.[48] Energy corrections for structures in our library and the newly-designed NCBs were performed with PBE0-D3(BJ)/def2-TZVP/SMD (water). Free energies were calculated and potential energy surfaces constructed using GoodVibes version 3.2.[49]

Natural bonding orbital (NBO) program version 7.0.5 [50] was used to obtain atomic partial charges. NBO partial charges at ten key atoms were gathered from the ground state structure of each reduced cofactor. Additionally, nucleophilic and electrophilic condensed Fukui indices were calculated using NBO partial charges for each after the subtraction and addition of an electron, respectively (Fig. 2, Step IV).[51]

Predictive modeling was done using ROBERT[52] software version 1.0.6 (Fig. 2, Step V). Our final model was trained using DFT-level descriptors after testing numerous model architectures (Fig. S7†) before settling on the highest-performing one, a multivariate linear regression model with a 101 : 18 : 13 train : validation : test split. Default settings were used except permutation feature importance analysis (PFI). PFI scores were calculated in ROBERT and tabulated, then used to isolate the 8 most important descriptors to train the model. More details about the use of ROBERT software for predictive modeling, including model architectures and data splits tested, permutation feature importance analysis, and verification of model performance and generalizability, can be found in section 7 of the ESI.†

Results and discussion

Expanding the NCB candidate space

To expand the limited space of NCB candidates, we assembled a library of 132 NCBs, over 85% of which have not (to our knowledge) been previously evaluated. Each molecule contains

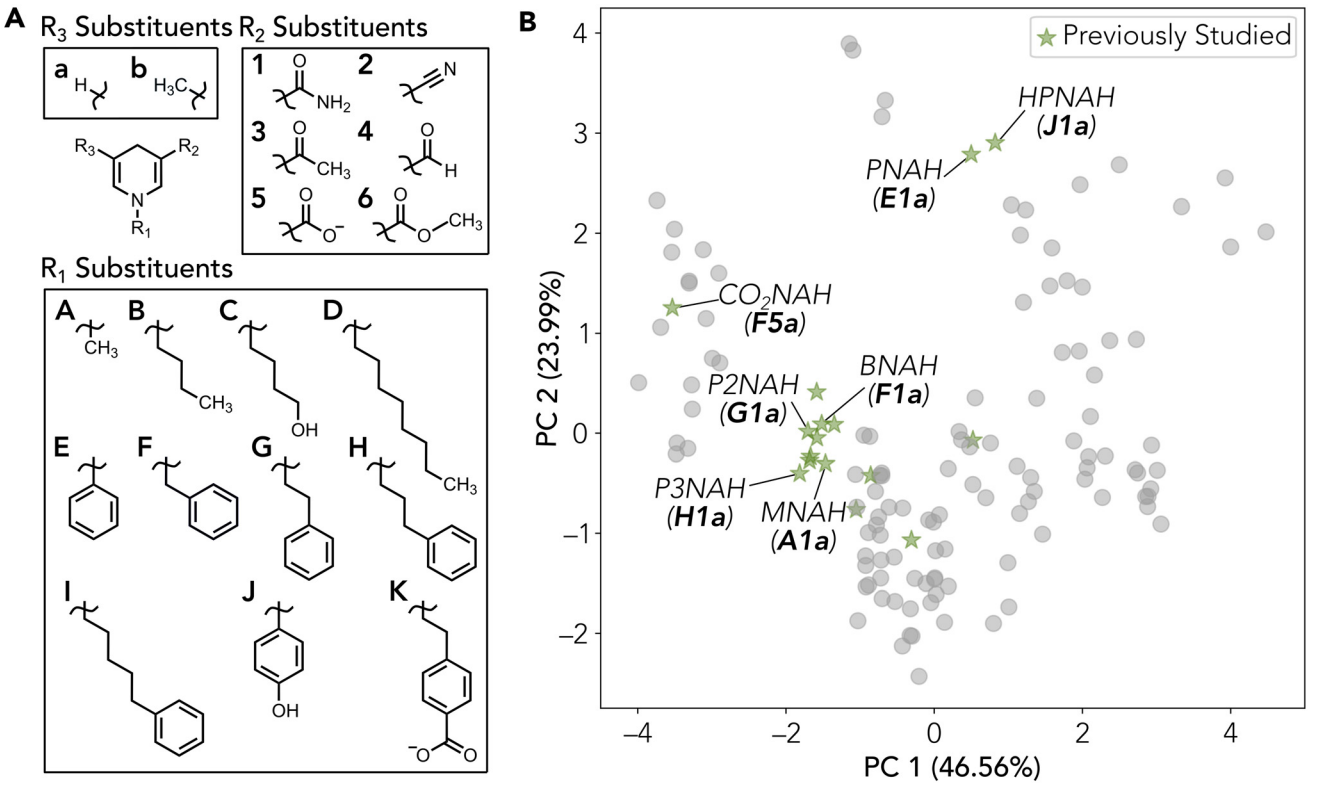

Fig. 3 (A) Visual representation of our substrate scope. Substituent possibilities were grouped in all possible combinations to create 132 structures identified by substituent identities (ex. P2NAH = G1a). (B) Principal component analysis (PCA) representation of the scope of our computational library, highlighting previously studied structures (green stars). Select NCBs have been labeled.

the redox active 1,4-dihydropyridine ring and is uniquely identified by the combination of R_1, R_2, and R_3 substituents (Fig. 3A). Each substituent possibility was combinatorically added to the core 1,4-dihydropyridine moiety. Substituents were selected based on previously evaluated NCBs to improve the likelihood that the resultant molecules will maintain some semblance of those that have been experimentally validated, increasing the likelihood that these untested NCBs maintain the activity of their natural counterparts. However, we also included two additional R_2 substituents which have not appeared in the literature to help examine substituent effects within the library, **4** and **6** in Fig. 3A. These substituents were added due to their electron-withdrawing capabilities, which we hypothesize will improve cofactor stability and allow further examination of the electronic effect of R_2 substituents on stability while maintaining chemical similarity to precedented mimics. To refer to specific molecules we will adopt a three-character naming convention detailed in Fig. 3A, where the first, second, and third characters correspond to the R_1 (**A–K**), R_2 (**1–6**), and R_3 (**a**, **b**) substituent identities, respectively. For example, all NCBs with an octyl substituent at the R_1 position are named **D****, where the asterisks represent wildcard characters that could be any R_2 and R_3 substituent, respectively. Similarly, all NCBs which have a methyl group as the R_3 substituent are named ****b**. In addition to the reduced form of each NCB, our automated workflow also generates the oxidized and decomposed species to study substituent effects on stability or redox activity. Generation of this expansive library was automated using an in-house developed Python script that is publicly available at **https://github.com/aplatt22/ncb_stability** and can be openly used to further expand this library with different substituents. More details about the generation of our NCB library can be found in section 1 the ESI.†

After SMILES (text-based) representations of each NCB in our library were generated, we followed an in-house automated protocol to perform conformational sampling and obtain optimized geometries at the DFT level of theory. This in-house script can be found at **https://github.com/aplatt22/ncb_stability**. Our conformational sampling procedure, using CREST to generate conformers and CENSO then CREST clustering to refine the ensemble, was used to isolate ten or fewer conformers of each NCB which were used for DFT optimizations and the curation of electronic-structure derived descriptors. A description of all calculated descriptors can be found in section 6 of the ESI.† Principal component analysis (PCA) was applied to visualize the variations within our dataset of NCBs featurized by calculated atomic descriptors (Fig. 3B). The first two principal components account for 71.55% of the total variance in the data, providing a sufficient representation of the diversity in our NCB library. Most of the experimental structures occupy a localized region of the chemical space, alluding to an improvement in molecular diversity. This robust library of structures allows more research into various substituent effects, opening the door for tuning specific properties of NCBs, such as reduction potential and stability in buffer.

Modeling NCB decomposition in phosphate buffer

Having expanded NCB candidate space, we aimed to understand the stability in terms of the NCB decomposition mechanism. NCBs which have been studied experimentally are noted to have high decomposition rates in phosphate buffer, which is problematic for CFB applications where enzymes often have highest activity in phosphate buffers.[14,16,23] Thus, we modeled the decomposition mechanism of NCBs using a negatively charged phosphate molecule to approximate the buffer solution. Referencing a previously-proposed decomposition scheme,[29] we modeled the decomposition of four NCBs which have experimental decomposition rates in phosphate buffer: MNAH (**A1a**), BNAH (**F1a**), P2NAH (**G1a**), and P3NAH (**H1a**). The activation barriers obtained are consistent with a process that is kinetically feasible at ambient temperatures. DFT modeling of the decomposition pathway suggests that the process occurs in a stepwise manner (Fig. 4). The first step involved C5 protonation *via* dihydrogen phosphate to form Int I, followed by attack of hydrogen phosphate at C6 to form the functionally inactive product. A concerted mechanism could also be envisioned; however, it is not supported at this level of theory (see section 3 of the ESI† for discussion). In each evaluated NCB, TS I corresponds to the highest energy barrier. As a result, these NCBs follow the stability trend of **F1a** = **G1a** > **A1a** > **H1a**, which is comparable to the experimental study that found the stability in phosphate buffer of **F1a** and **G1a** to be equivalent and higher than that of **A1a** and **H1a**.[14] The species product labeled on the reaction scheme in Fig. 4 does not necessarily represent the final decomposed species, and has been known to further decompose into another final product.[29] However, because experimental rates of decomposition are measured by the disappearance of the 1,4-dihydropyridine moiety and our calculated values are consistent with experiment, we do not consider further decomposition.

After identifying TS I as a predictor of NCB stability, we sought a simpler stability representation that would be more suitable for high-throughput calculations. Optimization of transition state structures often requires more computationally intensive protocols and manual intervention. Therefore, we investigated whether the free energy difference of this first elementary step (ΔG) was an appropriate alternative. Calculating ΔG requires only the optimization of minima structures and is therefore computationally cheaper. Such an approach is well grounded in studies of catalytic reactivity, where the thermodynamics of an elementary step is used to quantitatively assess the kinetic feasibility, formalized by the Evans–Polanyi principle[53] and linear (free) energy relationships (LFERs).[54]

To ensure that this assumption holds, we optimized transition state structures for a subset of 34 NCBs from our library. These 34 representative structures were selected using a binning technique based on important features of the molecules (details in section 3 of the ESI†). There is a highly correlated LFER between thermodynamic (ΔG) and kinetic ($\Delta G^{\ddagger}$) quantities, demonstrated by a Pearson's correlation coefficient

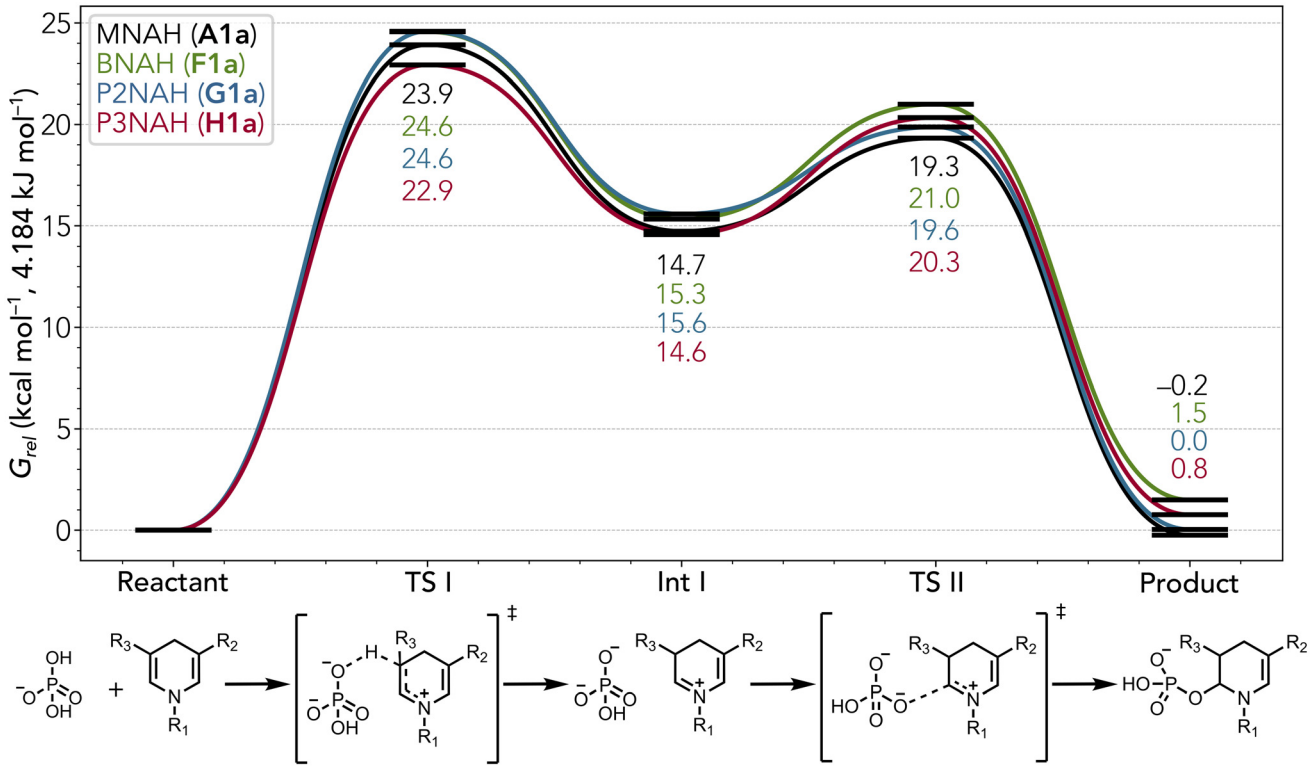

Fig. 4 Potential energy surface for the decomposition reaction of NCBs for MNAH (A1a), BNAH (F1a), P2NAH (G1a), and P3NAH (H1a). Calculations were performed with the level of theory ωB97M-V/def2-TZVP/SMD(water)//PBE0-D3(BJ)/6-31+G(d)/SMD(water).[38–46]

(R^2) of 0.92 (Fig. 5 and Table S1†). Furthermore, we compared thermodynamic stabilities of five NCBs with rates of decomposition from literature. Though quantitative decomposition rates were not available, we achieved qualitative accuracy in ranking the stabilities of **B1a**, **B1b**, **D1a**, **F1a**, and **F1b** from Tan *et al.* with 100 mM NCB concentrations.[16] Due to high correlation between ΔG and $\Delta G^{\ddagger}$, as well as qualitative agreement with experiment, we decided to model NCB stability through only a thermodynamic analysis of the first step, a procedure that is more amenable to high-throughput computation and analysis.

Having determined an appropriate metric for NCB stability, we analyzed the relative stability of our structures (Fig. S4†). The stabilities of structures in our NCB library range from −141.2 to −121.2 kcal mol^{-1} (−591.0 to −507.2 kJ mol^{-1}). The majority of our library shows increased stability when compared with a common mononucleotide NCB, nicotinamide mononucleotide ($NMNH^-$), which has a stability of −134.1 kcal mol^{-1} (−560.9 kJ mol^{-1}). These data suggest that we will be able to design NCBs with higher stability than those currently used in CFB.

Predictive modeling of NCB stability

To train a machine learning model to predict NCB stability in buffer, we had to identify descriptors to effectively describe the NCBs in ways relevant to decomposition. Because decomposition in phosphate buffer occurs within the conserved, redox active 1,4-dihydropyridine ring, this motif was our focus while describing NCBs. DFT-level descriptors were calculated at each atomic position within the 1,4-dihydropyridine ring (N1, C2, C3, C4, C5, and C6), as well at the first atom of each substituent (R_1, R_2, & R_3) and the descriptor values of both hydrogens at the C4 position (C4H) were averaged for a total of ten descriptors from eleven atomic loci. Calculated descriptors include partial atomic charges derived from natural population analysis (NPA),[50] as well as the electrophilic and nucleophilic condensed Fukui indices ($f(+)$ and $f(-)$, respectively).[51] Further explanation of descriptors can be found in the ESI.† Upon examination, these descriptors are able to effectively capture different substituent identities at each position, so no other features were considered in this study (Fig. S6†).

To limit computational cost as we moved forward, descriptors were calculated only for the reduced species of each NCB. Using a single species for descriptor calculation also reduces the computational cost of making future predictions with out-of-sample NCBs. We selected the reduced cofactor species for descriptor generation due to its role as the "universal intermediate" for both decomposition and catalytic reduction, allowing this set of descriptors to be used in a future multi-objective optimization for both stability and reduction potential. Permutation feature importance (PFI) analysis was implemented to isolate and use only the most important descriptors for NCB stability to help prevent model overfitting. PFI begins with a model trained on all descriptors, then systematically permutates one descriptor at a time to determine

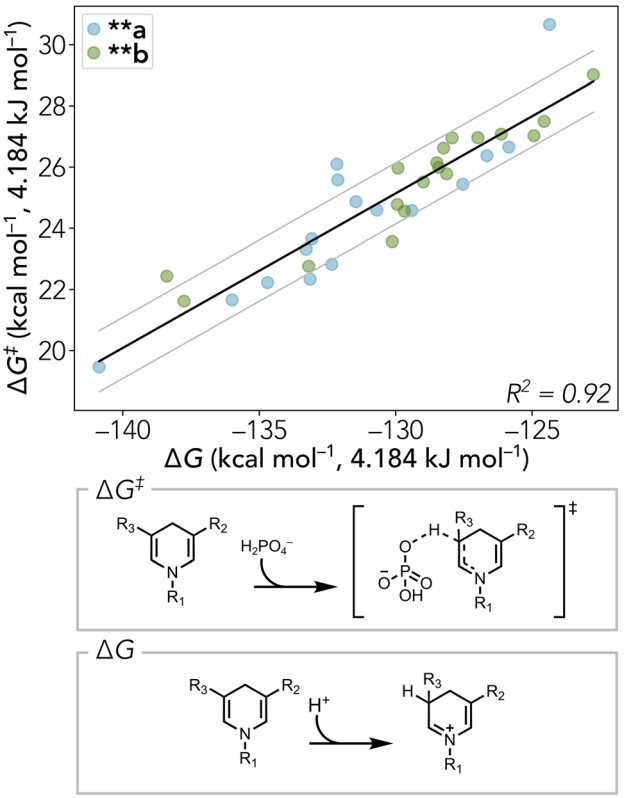

Fig. 5 Correlation between ΔG and $\Delta G^{\ddagger}$ for the first step of NCB decomposition with grey lines denoting 1 kcal mol^{-1} (4.184 kJ mol^{-1}) greater or less than the linear fit (black) for these data. Values for ΔG were obtained with PBE0-D3(BJ)/def2-TZVP/SMD(water)//PBE0-D3(BJ)/6-31+G(d)/SMD(water).[38–44] Values for $\Delta G^{\ddagger}$ were computed with ωB97M-V/def2-TZVP/SMD(water)//PBE0-D3(BJ)/6-31+G(d)/SMD (water).[38–46] Shown below the graph are schemes used to calculate each variable.

the impact of that descriptor on the overall model. The worse a model performs when a descriptor is permutated, the more important that descriptor is to making predictions. This analysis is automated in ROBERT, resulting in a PFI score for each descriptor (Table S5†). The eight descriptors with the highest PFI scores were used in our model, including atomic charges at N1, C2, C3, C5, and C4H and electrophilic condensed Fukui indices at C3, R_2, and C4H. We also performed a comparison of these properties calculated from the lowest energy conformer, which was used in model training, with the Boltzmann weighted properties for our representative subset of 34 NCBs. The maximum differences between values for all atomic charges and electrophilic condensed Fukui indices of these NCBs were 0.001 and 0.008, respectively. The differences are low relative to the magnitude of the values, suggesting we do not need to include the full conformational ensemble for our descriptors.

An array of four model types and seven train : validation : test splits were used to train possible DFT-level models for predicting stability using ROBERT (Fig. S7†). We selected the highest-performing model architecture after default PFI analysis based on the root mean squared error (RMSE) values of the validation set, a multivariate linear regression (MLR) employing a 101 : 18 : 13 train : validation : test split (Fig. 6). PFI scores were then further analyzed to limit model training to only the eight most important physical properties of the NCBs. This number of descriptors was selected because it was the lower limit at which we still had high model performance (Table S6†). The final descriptors included in our model are the NBO partial charges at N1, C2, C3, C5, and C4H, as well as $f(-)$ at C3, R_2, and C4H. While our list of key descriptors does not include direct contributions from the R_1 or R_3 substituents, properties from N1 and C5 are used to predict stability, which we believe incorporate effects from their respective substituents. In support of this, descriptors of R_3 were highly correlated with the C5 NBO partial charge and not considered for model training. More details regarding model training can be found in section 7 of the ESI.†

This analysis led to a high-performing model, shown by a high correlation (test set R^2 = 0.98), as well as mean absolute error (MAE) and RMSE within DFT accuracy levels (≤3 kcal mol^{-1}, 12.552 kJ mol^{-1}).[55] Furthermore, ROBERT automates the additional evaluation of model performance, see **https://github.com/aplatt22/ncb_stability/tree/main/Data/ROBERT_Results** for results. This extra evaluation includes testing for worse performance when training with mean stability values, shuffled stability values or one-hot encoded descriptors, and performing 5-fold cross-validation to ensure our model is robust. Our model passed these tests, demonstrating a generalizable model with an appropriate train : validation : test split. The data was also checked for outliers and the ratio of data points to descriptors as part of this additional model evaluation. More details regarding this evaluation can be found in section 7 of the ESI.†

As further assessment of the performance of our model, we calculated Spearman rank-order correlations comparing stability from our model predictions with those from experiment. With the limited experimental stability data available, a problem we address with our computational NCB library, we are unable to compare our model predictions to many experimental stabilities. Using the limited data, we are able to examine rankings of the 5 NCB structures from Tan *et al.*[16] Our model ranks these cofactors **B1a** < **D1a** < **F1a** < **B1b** < **F1b**, which is the same ordering found experimentally, earning a Spearman correlation of 1.000. When comparing our model predictions to experimental data of 4 NCBs from Nowak *et al.*,[14] our model predictions showed rankings, in order of increasing stability, of **H1a** < **A1a** < **G1a** < **F1a** and the experimental rankings are **A1a** < **H1a** < **F1a** < **G1a**, earning a Spearman correlation of 0.600. The rankings for the Nowak *et al.* experimental results were not predicted as well as for the Tan *et al.* results, which can be explained by the nature of the experimental stability data. For the Nowak *et al.* set of NCBs, the range of calculated stabilities is only 1.9 kcal mol^{-1} (8.0 kJ mol^{-1}), while the Tan *et al.* set has a range of 5.2 kcal mol^{-1} (22.0 kJ mol^{-1}). Compared to our full computational library, which has a range in stabilities of 20.0 kcal mol^{-1} (83.8 kJ

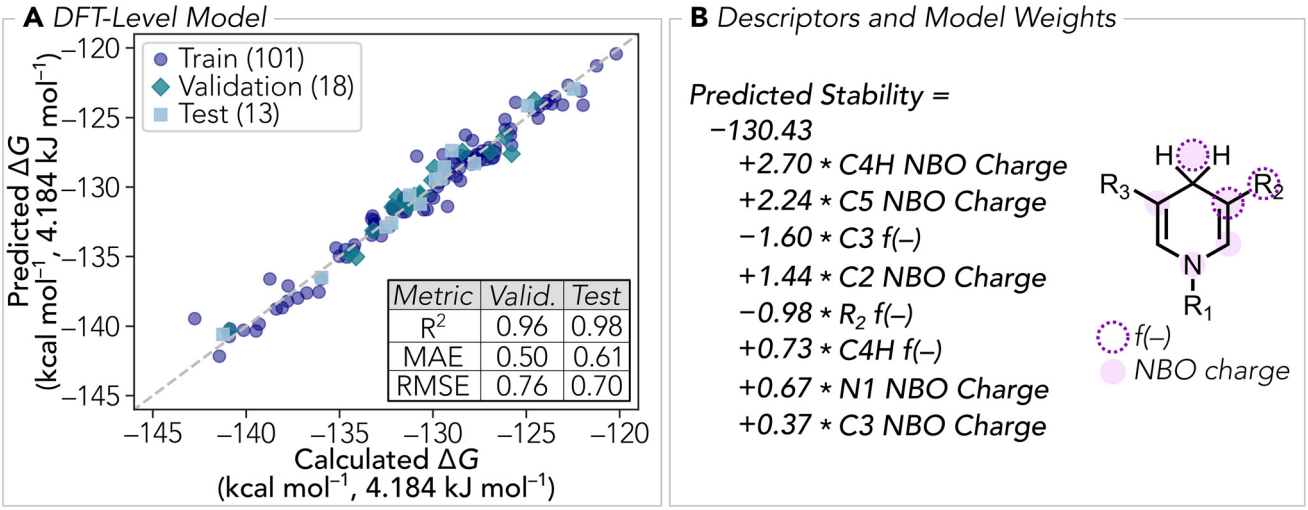

Fig. 6 (A) Parity plot and results for the best-performing predictive model trained with DFT-level descriptors, a MLR with a 101 : 18 : 13 train : validation : test split using the eight most important descriptors. Model performance metrics are shown at the bottom right of the graph for the validation and test sets. (B) Visual representation of the eight descriptors used for model training and MLR equation from standardized descriptor data. Raw weights can be found in the ESI.†

mol^{-1}), this subset of experimental stability data is not very diverse, so this available experimental data gives limited insight for our chemically diverse dataset. Additionally, all of the structures for which we have experimental stability data are similar, each with **1** for the R_2 substituent, only examining different R_3 substituents for 2 R_1 substituents, and only sampling 5 R_1 substituents of the 11 possibilities, further demonstrating the limited insight from comparing model results with current experimental data.

Finally, we tested our model performance with a Spearman correlation over our full NCB library, which a gave rank-order correlation of 0.974 when comparing model predictions to calculated stabilities, demonstrating the ability of our model to correctly order NCBs with respect to their relative stabilities (Table S10†). This feature is ideal for an optimization protocol, where only the most stable NCBs are considered for experimental validation, so getting the correct stability ranking is as effective as predicting the actual value. We also explored a model trained with semi-empirical descriptors in an attempt to reduce the computational cost of making predictions, but the difference in performance relative to the cost of each method was not enough to favor the semi-empirical model (see section 7 of the ESI†).

Examining model and descriptor trends

After training a regression model to predict NCB stability, we used that information to design NCBs with substituent groups outside the scope of those previously identified. To do this, we further examined the descriptors which were used in training the model by examining the relationship between each descriptor and the NCB stability. In Fig. 7, we show a subset of these relationships and label the data according to substituent positions to highlight how certain descriptors can discriminate by substituent identity. Examining the relationship of $f(-)$ at the C4H atoms with stability (Fig. 7A), there are two substituents which yield higher stabilities, corresponding to R_1 substituents **E** and **J**. Each of these substituents have an aryl substituent bound directly to the N atom in the 1,4-dihydropyridine ring. Furthermore, the two most stable NCBs in our library are **E4b** and **J4b**, suggesting that aryl R_1 substituents are a favorable design component to enhance NCB stability (Table S11†). This suggests that alterations to an aryl R_1 substituent could be a viable path for further NCB chemical space exploration. Other substituent identities also impact stability, but not to the extent of that seen from R_1 aryl substituents.

We performed a similar analysis of the R_2 substituent (Fig. 7B), showing that the aldehyde moiety (**4**) tends to have higher stability than the other functional groups represented in our library at this position. The five most stable NCB structures in our library, and fifteen of the twenty most stable, have **4** as the R_2 substituent. We hypothesize that this trend is due to the highly electron-withdrawing nature of aldehydes. Though **4** is not a substituent previously evaluated in literature, we added it to our library as an additional substituent to determine if the electron-withdrawing characteristics made a large impact on NCB stability. Based on these data, this is a key design feature in optimizing stability in NCBs. The next most stable substituent, **2**, is also highly electron-withdrawing as a nitrile, which has been evaluated previously.

Finally, it is clear when looking at the relationship between C4H partial atomic charge and stability colored by R_3 substituent (Fig. 7C) that NCBs with an R_3 methyl substituent have higher stability than those with only hydrogen at that position. While examining NCBs with similar C4H partial charges, those with **b** at R_3 demonstrate greater stability. Furthermore, our eleven most stable NCBs have a methyl substituent at the

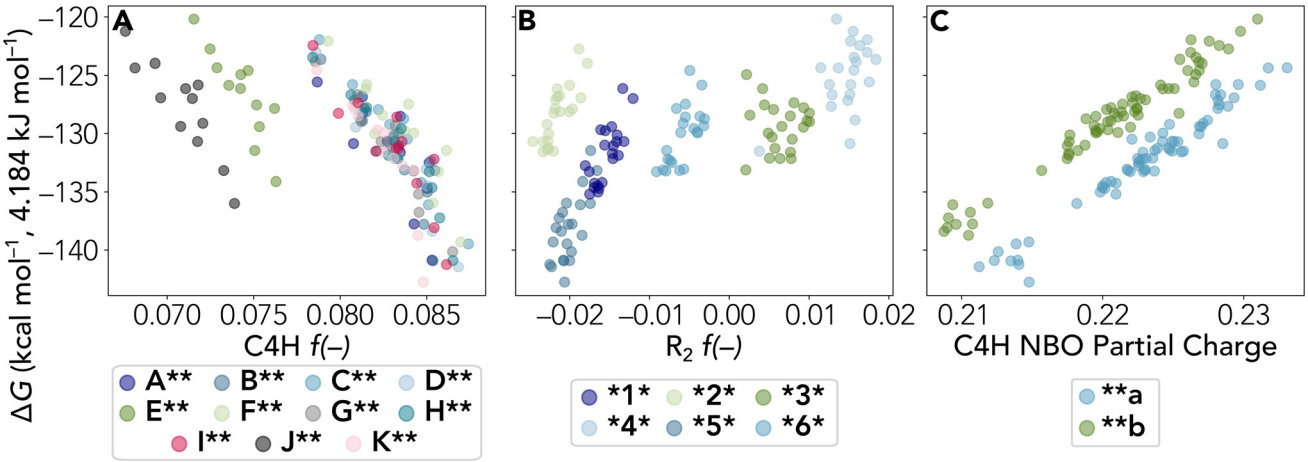

Fig. 7 Relationships between select important DFT-level descriptors and NCB stability. (A) Relationship between C4H *f*(−) and stability colored by R_1 substituent. (B) Relationship between R_2 *f*(−) and stability colored by R_2 substituent. (C) Relationship between C4H NBO partial charge and stability colored by R_3 substituent.

R_3 position. Tan *et al.* hypothesized that this improved stability is due to kinetic effects,[16] and we are still able to replicate this trend even though our analysis is strictly thermodynamic. This observation could be caused by additional steric effects, such as lower flexibility in the 1,4-dihydropyridine ring with greater congestion at the C5 carbon with a methyl substituent rather than a hydrogen.

Out-of-sample NCB design

To test our hypotheses, we created out-of-sample NCBs by varying the identity of our substituents and testing the stability. We first examined new aryl R_1 substituents with the most stable R_2 (**4**) and R_3 (**b**) substituents (Fig. 8, **R_1sub-I–R_1sub-V**). We also designed a NCB with an R_2 amide (**1**) and R_3 hydrogen (**a**) with a highly electron-withdrawing R_1 substituent (**R_1sub-VI**).[56] We wanted to design this NCB to look more similar to current structures to ensure the calculated properties are reasonable because the majority of structures in literature include **1** as the R_2 substituent and **a** and the R_3 substituent.

In addition to altering the R_1 substituent, we also looked at how changing the R_2 and R_3 substituents impacted stability. To do this, we designed four structures with new R_2 substituents. These moieties included a thioaldehyde (**R_2sub-I**), thioketone (**R_2sub-II**), trifluoromethyl (**R_2sub-III**), and nitro (**R_2sub-IV**). Thioamides have literature precedent as dinucleotide biomimetic cofactors, which share most of their structure with NAD(P)H, and are proposed to behave similarly to their amide counterparts but have higher reduction potentials than the natural cofactors.[57] We selected a thioketone rather than a thioamide because ketones demonstrate higher stability than amides in our library. Additionally, the increased stabilities shown by R_2 aldehydes led us to also choose a thioaldehyde due to the increased stability shown by aldehydes in our study. Electron-withdrawing groups seem to improve stability at the R_2 position, with aldehydes performing well, so we selected a highly electron-withdrawing substituent, trifluoromethyl, to further investigate this trend. This reasoning also led us to design an NCB with a nitro R_2 substituent. Additionally, we designed two structures with a trifluoromethyl R_3 substituent (**R_3sub-I** & **R_3sub-II**). We selected this moiety as a possible R_3 substituent to test if electron-withdrawing effects at this position would have any influence on stability in addition to the bulkiness of the substituent. Trifluoromethyl substituents have only a slightly larger size than the previously tested methyl substituent but are more electron-withdrawing.

The new proposed NCBs were among the most stable tested, with three structures surpassing stability of those currently in our library: **R_3sub-II** (−117.1 kcal mol^{-1}, −490.1 kJ mol^{-1}), **R_2sub-IV** (−120.0 kcal mol^{-1}, −502.3 kJ mol^{-1}), and **R_1sub-IV** (−117.3 kcal mol^{-1}, −490.9 kJ mol^{-1}) (Fig. S11†). The increased stability of these NCBs relative to the remainder of the library is promising since it shows that we can use trends in existing data to propose and successfully design new NCBs with enhanced stability. Furthermore, the higher stability seen with the new substituents at each position supports our hypotheses that NCBs with aryl R_1 substituents, electron-withdrawing R_2 substituents, and bulky and/or electron-withdrawing R_3 substituents are likely to be stable structures. With this new information about how substituent identity and properties at each position impact stability, we can make informed decisions to eventually suggest novel NCBs which will be more resistant to decomposition *via* C5–C6 saturation.

In addition to high thermodynamic stability, these out-of-sample NCBs also show high kinetic stability and correlation between ΔG and $\Delta G^{\ddagger}$, as was determined with NCBs in our library (R^2 = 0.92) (Fig. S12 and Table S12†). We also used these novel structures to test our model's performance with out-of-sample structures. All except three of these out-of-sample NCBs showed accurate predictions (<3 kcal mol^{-1}, 12.552 kJ mol^{-1}, prediction error), with each poor prediction occurring for structures with out-of-sample R_2 groups. The accurate predictions on unseen substructures gives confidence

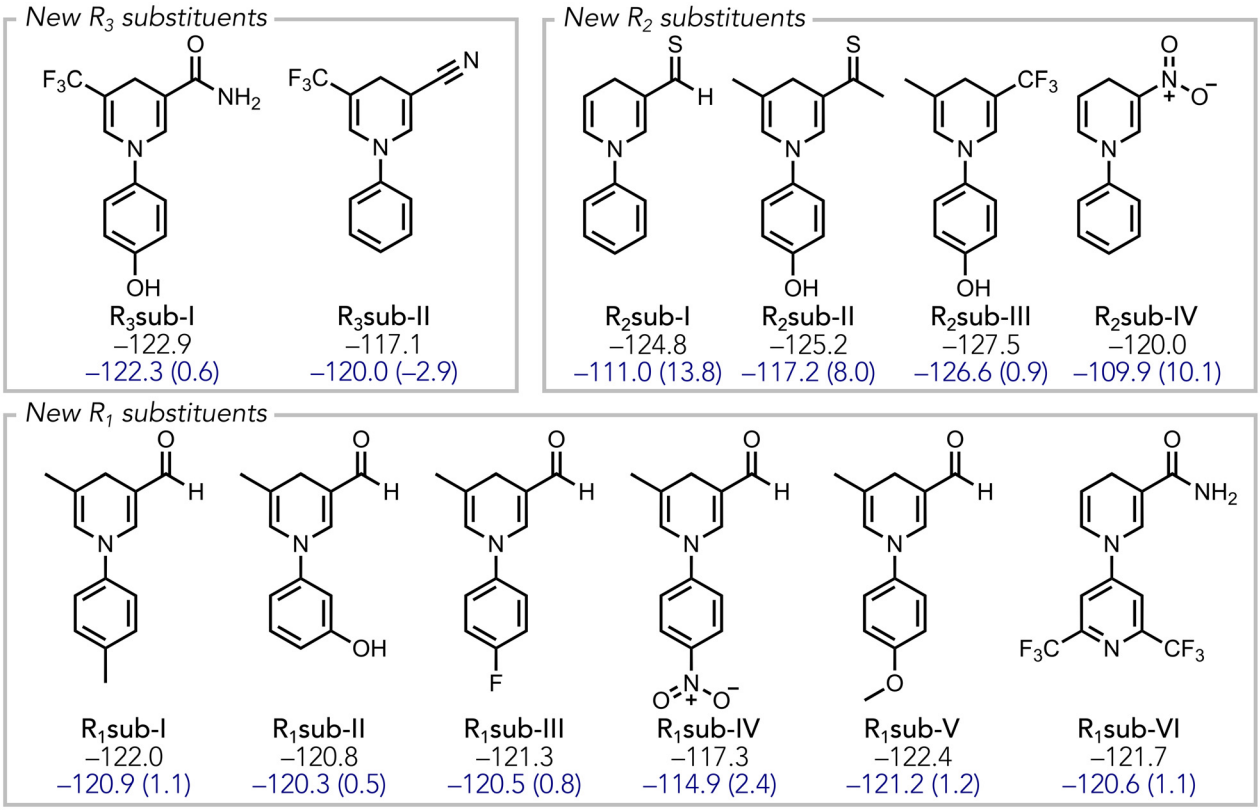

Fig. 8 Novel NCBs designed to improve stability based on trends in stability data. Shown below the structure names in black are the DFT-calculated stabilities. Also shown are the predicted stabilities from our predictive model in blue with prediction errors shown in parentheses. All values have units of kcal mol^{-1} (4.184 kJ mol^{-1}).

in the predictive ability of our model. Additionally, the model tends to predict stabilities higher than the benchmark DFT calculations, especially for the three out-of-sample structures errors in stability prediction for these out-of-sample structures assume higher stability than calculated using DFT, a trend present for all except one structure. Errors tending this way reduce the chance that synthetic chemists disregard highly stable structures due to our model when choosing stability-ranked NCB candidates to characterize experimentally.

However, there were also some structures which showed low model accuracy (error >3 kcal mol^{-1}, 12.552 kJ mol^{-1}), likely due to the type of descriptors utilized to make predictions and absence of sufficient structure representations in the training data. When analysing these structures with poor model accuracy, **R_2sub-I** (error of 13.8 kcal mol^{-1}, 57.7 kJ mol^{-1}), **R_2sub-II** (error of 8.0 kcal mol^{-1}, 33.5 kJ mol^{-1}), and **R_2sub-IV** (error of 10.2 kcal mol^{-1}, 42.6 kJ mol^{-1}), we found that each NCB had at least three descriptors used in the model which were outside the range seen by the initial library. For example, **R_2sub-I** and **R_2sub-II** each had $f(-)$ values at the R_2 substituent (−0.09 and −0.03, respectively) that were outside the observed range (−0.02 to 0.02), likely causing poor stability predictions in these structures. The other NCB which had a poor prediction was **R_2sub-IV**. Our model likely failed in making an accurate prediction of stability for **R_2sub-IV** because all other structures within our library have a neutral carbon atom at R_2, so the electronic descriptors at this atom and C3 (directly bound to R_2) are beyond the scope of our trained model.

While the expansion of our NCB space with these out-of-sample structures showed that our model carries predictive power even with substrates it was not trained with, the model can be adapted over time through expansion of the training data, which would be recommended if a substituent chemically distinct from our current library is found to be relevant in NCB design. Nevertheless, our model tends to predict stabilities higher than the benchmark DFT calculations, especially for the three high-error out-of-sample structures. This error trend reduces the chance that a highly stable structure would fail to advance to an experimental stage in a high-throughput NCB discovery pipeline, and even though some "false positives" might appear in the top ranks, our results show that is unlikely. For example, if an experimental group chose to synthesize and characterize the top 10% of stability-ranked structures, including the out-of-sample designs, this would result in two of the fifteen candidates (rounded up), with DFT calculated ΔG values that would have excluded them from the top 10%, but that still place them in the top 20%. The design of our computational workflow and ability of our model to predict the correct rankings of stability, even of newly designed NCB structures, reduces the chance that experimental che-

mists create additional reaction waste by synthesizing NCBs with low stability.

Conclusions

This work introduces an NCB library which covers more chemical space than has been studied before, generating a library of 132 unique NCBs to expand the candidate space and allow us to examine substituent stability effects in more detail than was previously possible. With the limited number of experimental structures with decomposition data available, we also isolated the mechanism of decomposition of nicotinamide NCBs in a phosphate buffer, suggesting decomposition occurs in a stepwise manner where the first decomposition event (TS I) is highest in energy. Furthermore, there is a strong correlation between the thermodynamic and kinetic data, so we simplified our model of stability and used ΔG for the proton transfer step of decomposition (Reactant to Int I).

Using our NCB library and simplified model of stability, we fit a regression model on DFT-calculated electronic descriptors to predict NCB stability in phosphate buffer and used the results of our model to design new NCBs proposed to demonstrate higher stability. Our model has high accuracy, shown by low MAE and RMSE values, as well as high ranked-order correlations, demonstrating successful rankings of NCB stabilities, even if we cannot predict the correct quantitative stability. Using trends in key descriptors, we drew qualitative conclusions about designing stable cofactors and tested these hypotheses through the design of novel NCBs, most of which showed high stability. We also tested these structures as out-of-sample predictions for our predictive model, through which we discovered strengths and weaknesses of the model. In the future, we also hope to expand the library of NCBs to improve generalizability of our predictive models.

As shown by this work, the stability of NCBs can be successfully tuned. We now have an initial pass at designing more effective NCBs, narrowing the search to only NCBs which demonstrate stability. Following this work, we can continue down the molecular design pathway, focusing on cofactor activity and optimization of other desirable properties. Additionally, this work integrates only publicly available packages into an automated workflow available at **https://github.com/aplatt22/ncb_stability**, highlighting the transparency and reproducibility of this work. Future work in this area can be efficiently adapted to build predictive models for other key NCB properties, such as reduction potential or solubility. The binding affinity of NCBs within an enzyme active site has been studied through enzyme engineering,[58] so provided that the NCBs demonstrate high stability and solubility, as well as an optimized reduction potential, we are confident that the NCBs can successfully be used for industrial CFB. This work provides a robust computational study of new NCB structures with the goal of motivating experimental validation, a direction currently pursued. Tuning NCBs and their properties on a case-by-case basis with workflows like ours will increase the usefulness of CFB, reducing waste from synthesis of NCBs known to have low stability and making the method more efficient and feasible for use in industrial settings.

Author contributions

A. Platt: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft, and writing – review & editing. H. Klem: conceptualization, methodology, supervision, visualization, and writing – review & editing. S. Mallinson: writing – review & editing. Y. Bomble: conceptualization, funding acquisition, supervision, writing – review & editing. R. Paton: conceptualization, funding acquisition, project administration, resources, supervision, writing – review & editing.

Data availability

The data supporting this article have been included either as part of the ESI† and on GitHub at **https://github.com/aplatt22/ncb_stability**.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work utilized the Alpine high performance computing resource at the University of Colorado Boulder, jointly funded by the University of Colorado Boulder, the University of Colorado Anschutz, and Colorado State University, and the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) through allocation TG-CHE180056.

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy under the Cell Free and Immobilization Technologies initiative. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

H. Klem was supported by a NIST National Research Council (NRC) Research Postdoctoral Associateship Award in the Thermodynamics Research Group of the Materials

Measurement Laboratory during the preparation of this manuscript.

We would like to thank Louis DeLescure (Sanofi Pharmaceuticals) and Andrei F. Kazakov (Thermodynamics Research Center, NIST) for their time and effort reviewing the manuscript during the internal NIST editorial review process.

References

1 T. Matsuda, R. Yamanaka and K. Nakamura, Recent Progress in Biocatalysis for Asymmetric Oxidation and Reduction, *Tetrahedron: Asymmetry*, 2009, **20**(5), 513–557, DOI: 10.1016/j.tetasy.2008.12.035.

2 R. C. Simon, F. G. Mutti and W. Kroutil, Biocatalytic Synthesis of Enantiopure Building Blocks for Pharmaceuticals, *Drug Discovery Today: Technol.*, 2013, **10**(1), e37–e44, DOI: 10.1016/j.ddtec.2012.08.002.

3 R. A. Sheldon, The E Factor 25 Years on: The Rise of Green Chemistry and Sustainability, *Green Chem.*, 2017, **19**(1), 18–43, DOI: 10.1039/C6GC02157C.

4 P. Lozano and E. García-Verdugo, From Green to Circular Chemistry Paved by Biocatalysis, *Green Chem.*, 2023, **25**(18), 7041–7057, DOI: 10.1039/D3GC01878D.

5 L. Wang, H. Wang, J. Fan and Z. Han, Synthesis, Catalysts and Enhancement Technologies of Biodiesel from Oil Feedstock – A Review, *Sci. Total Environ.*, 2023, **904**, 166982, DOI: 10.1016/j.scitotenv.2023.166982.

6 P. T. Anastas and J. C. Warner, *Green Chemistry: Theory and Practice*, Oxford University Press, Oxford, New York, 2000.

7 H. Liu and J. U. Bowie, Cell-Free Synthetic Biochemistry Upgrading of Ethanol to 1,3 Butanediol., *Sci. Rep.*, 2021, **11**(1), 9449, DOI: 10.1038/s41598-021-88899-w.

8 S. Sherkhanov, T. P. Korman, S. Chan, S. Faham, H. Liu, M. R. Sawaya, W.-T. Hsu, E. Vikram, T. Cheng and J. U. Bowie, Isobutanol Production Freed from Biological Limits Using Synthetic Biochemistry, *Nat. Commun.*, 2020, **11**(1), 4292, DOI: 10.1038/s41467-020-18124-1.

9 T. P. Korman, P. H. Opgenorth and J. U. Bowie, A Synthetic Biochemistry Platform for Cell Free Production of Monoterpenes from Glucose, *Nat. Commun.*, 2017, **8**(1), 15526, DOI: 10.1038/ncomms15526.

10 K. D. Wolfe, M. Alahuhta, M. E. Himmel, Y. J. Bomble, G. K. Jennings and D. E. Cliffel, Long-Term Stability of Nicotinamide Cofactors in Common Aqueous Buffers: Implications for Cell-Free Biocatalysis, *Molecules*, 2024, **29**(22), 5453, DOI: 10.3390/molecules29225453.

11 J. A. Rollin, T. K. Tam and Y.-H. P. Zhang, New Biotechnology Paradigm: Cell-Free Biosystems for Biomanufacturing, *Green Chem.*, 2013, **15**(7), 1708–1719, DOI: 10.1039/C3GC40625C.

12 C. E. Paul, I. W. C. E. Arends and F. Hollmann, Is Simpler Better? Synthetic Nicotinamide Cofactor Analogues for Redox Chemistry, *ACS Catal.*, 2014, **4**(3), 788–797, DOI: 10.1021/cs4011056.

13 T. Knaus, C. E. Paul, C. W. Levy, S. de Vries, F. G. Mutti, F. Hollmann and N. S. Scrutton, Better than Nature: Nicotinamide Biomimetics That Outperform Natural Coenzymes, *J. Am. Chem. Soc.*, 2016, **138**(3), 1033–1039, DOI: 10.1021/jacs.5b12252.

14 C. Nowak, A. Pick, L.-I. Csepei and V. Sieber, Characterization of Biomimetic Cofactors According to Stability, Redox Potentials, and Enzymatic Conversion by NADH Oxidase from Lactobacillus Pentosus, *ChemBioChem*, 2017, **18**(19), 1944–1949, DOI: 10.1002/cbic.201700258.

15 C. Nowak, A. Pick, P. Lommes and V. Sieber, Enzymatic Reduction of Nicotinamide Biomimetic Cofactors Using an Engineered Glucose Dehydrogenase: Providing a Regeneration System for Artificial Cofactors, *ACS Catal.*, 2017, **7**(8), 5202–5208, DOI: 10.1021/acscatal.7b00721.

16 Z. Tan, Y. Han, Y. Fu, X. Zhang, M. Xu, Q. Na, W. Zhuang, X. Qu, H. Ying and C. Zhu, Investigating the Structure-Reactivity Relationships Between Nicotinamide Coenzyme Biomimetics and Pentaerythritol Tetranitrate Reductase, *Adv. Synth. Catal.*, 2022, **364**(1), 103–113, DOI: 10.1002/adsc.202100726.

17 C. E. Paul, S. Gargiulo, D. J. Opperman, I. Lavandera, V. Gotor-Fernández, V. Gotor, A. Taglieber, I. W. C. E. Arends and F. Hollmann, Mimicking Nature: Synthetic Nicotinamide Cofactors for C=C Bioreduction Using Enoate Reductases, *Org. Lett.*, 2013, **15**(1), 180–183, DOI: 10.1021/ol303240a.

18 D. Ji, L. Wang, S. Hou, W. Liu, J. Wang, Q. Wang and Z. K. Zhao, Creation of Bioorthogonal Redox Systems Depending on Nicotinamide Flucytosine Dinucleotide, *J. Am. Chem. Soc.*, 2011, **133**(51), 20857–20862, DOI: 10.1021/ja2074032.

19 I. Zachos, C. Nowak and V. Sieber, Biomimetic Cofactors and Methods for Their Recycling, *Curr. Opin. Chem. Biol.*, 2019, **49**, 59–66, DOI: 10.1016/j.cbpa.2018.10.003.

20 C. E. Paul and F. Hollmann, A Survey of Synthetic Nicotinamide Cofactors in Enzymatic Processes, *Appl. Microbiol. Biotechnol.*, 2016, **100**(11), 4773–4778, DOI: 10.1007/s00253-016-7500-1.

21 I. Zachos, S. Güner, A. Essert, P. Lommes and V. Sieber, Boosting Artificial Nicotinamide Cofactor Systems, *Chem. Commun.*, 2022, **58**(85), 11945–11948, DOI: 10.1039/D2CC03423A.

22 B. Reus, M. Damian and F. G. Mutti, Advances in Cofactor Immobilization for Enhanced Continuous-Flow Biocatalysis, *J. Flow Chem.*, 2024, **14**(1), 219–238, DOI: 10.1007/s41981-024-00315-2.

23 C. Nowak, B. C. Beer, A. Pick, T. Roth, P. Lommes and V. Sieber, A Water-Forming NADH Oxidase from Lactobacillus Pentosus Suitable for the Regeneration of Synthetic Biomimetic Cofactors, *Front. Microbiol.*, 2015, **6**, 957, DOI: 10.3389/fmicb.2015.00957.

24 E. King, S. Maxel and H. Li, Engineering Natural and Noncanonical Nicotinamide Cofactor-Dependent Enzymes: Design Principles and Technology Development, *Curr.*

Opin. Biotechnol., 2020, **66**, 217–226, DOI: 10.1016/j.copbio.2020.08.005.

25 W. B. Black, L. Zhang, W. S. Mak, S. Maxel, Y. Cui, E. King, B. Fong, A. Sanchez Martinez, J. B. Siegel and H. Li, Engineering a Nicotinamide Mononucleotide Redox Cofactor System for Biocatalysis, *Nat. Chem. Biol.*, 2020, **16**(1), 87–94, DOI: 10.1038/s41589-019-0402-7.

26 D. Meng, M. Liu, H. Su, H. Song, L. Chen, Q. Li, Y. Liu, Z. Zhu, W. Liu, X. Sheng, C. You and Y.-H. P. J. Zhang, Coenzyme Engineering of Glucose-6-Phosphate Dehydrogenase on a Nicotinamide-Based Biomimic and Its Application as a Glucose Biosensor, *ACS Catal.*, 2023, **13**(3), 1983–1998, DOI: 10.1021/acscatal.2c04707.

27 Y. Liu, Y. Cong, C. Zhang, B. Fang, Y. Pan, Q. Li, C. You, B. Gao, J. Z. H. Zhang, T. Zhu and L. Zhang, Engineering the Biomimetic Cofactors of NMNH for Cytochrome P450 BM3 Based on Binding Conformation Refinement, *RSC Adv.*, 2021, **11**(20), 12036–12042, DOI: 10.1039/D1RA00352F.

28 P. L. Hentall, N. Flowers and T. D. H. Bugg, Enhanced Acid Stability of a Reduced Nicotinamide Adenine Dinucleotide (NADH) Analogue, *Chem. Commun.*, 2001, (20), 2098–2099, DOI: 10.1039/B107634P.

29 S. G. A. Alivisatos, F. Ungar and G. Abraham, Non-Enzymatic Interactions of Reduced Coenzyme I with Inorganic Phosphate and Certain Other Anions, *Nature*, 1964, **203**(4948), 973–975, DOI: 10.1038/203973a0.

30 D. Dell'Angelo, 13 - Computational Chemistry and the Study and Design of Catalysts, in *Green Chemistry and Computational Chemistry*, ed. L. Mammino, Advances in Green and Sustainable Chemistry, Elsevier, 2022, pp. 299–332. DOI: 10.1016/B978-0-12-819879-7.00010-6.

31 *Green chemistry and computational chemistry: A wealth of promising synergies – ScienceDirect*. https://www.sciencedirect.com/science/article/abs/pii/S2352554123001857 (accessed 2025-03-10).

32 P. Pracht, S. Grimme, C. Bannwarth, F. Bohle, S. Ehlert, G. Feldmann, J. Gorges, M. Müller, T. Neudecker, C. Plett, S. Spicher, P. Steinbach, P. A. Wesołowski and F. Zeller, CREST—A Program for the Exploration of Low-Energy Molecular Chemical Space, *J. Chem. Phys.*, 2024, **160**(11), 114110, DOI: 10.1063/5.0197592.

33 C. Bannwarth, S. Ehlert and S. Grimme, GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions, *J. Chem. Theory Comput.*, 2019, **15**(3), 1652–1671, DOI: 10.1021/acs.jctc.8b01176.

34 C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, Extended Tight-Binding Quantum Chemistry Methods, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **11**(2), e1493, DOI: 10.1002/wcms.1493.

35 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, Open Babel: An Open Chemical Toolbox., *J. Cheminform.*, 2011, **3**(1), 33, DOI: 10.1186/1758-2946-3-33.

36 S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher and M. Stahn, Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules, *J. Phys. Chem. A*, 2021, **125**(19), 4039–4054, DOI: 10.1021/acs.jpca.1c00971.

37 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 and D. J. Fox, *Gaussian 16, Revision C.01*, 2016.

38 A. V. Marenich, C. J. Cramer and D. G. Truhlar, Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions, *J. Phys. Chem. B*, 2009, **113**(18), 6378–6396, DOI: 10.1021/jp810292n.

39 C. Adamo and V. Barone, Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model, *J. Chem. Phys.*, 1999, **110**(13), 6158–6170, DOI: 10.1063/1.478522.

40 T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. V. R. Schleyer, Efficient Diffuse Function-Augmented Basis Sets for Anion Calculations. III. The 3-1+G Basis Set for First-Row Elements, Li–F, *J. Comput. Chem.*, 1983, **4**(3), 294–301, DOI: 10.1002/jcc.540040303.

41 W. J. Hehre, R. Ditchfield and J. A. Pople, Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules, *J. Chem. Phys.*, 1972, **56**(5), 2257–2261, DOI: 10.1063/1.1677527.

42 P. C. Hariharan and J. A. Pople, The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies, *Theor. Chim. Acta*, 1973, **28**, 213–222, DOI: 10.1007/BF00533485.

43 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu, *J. Chem. Phys.*, 2010, **132**(15), 154104, DOI: 10.1063/1.3382344.

44 S. Grimme, S. Ehrlich and L. Goerigk, Effect of the Damping Function in Dispersion Corrected Density Functional Theory, *J. Comput. Chem.*, 2011, **32**(7), 1456–1465, DOI: 10.1002/jcc.21759.

45 F. Weigend and R. Ahlrichs, 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**(18), 3297–3305, DOI: **10.1039/B508541A**.

46 N. Mardirossian and M. Head-Gordon, ωB97M-V: A Combinatorially Optimized, Range-Separated Hybrid, Meta-GGA Density Functional with VV10 Nonlocal Correlation, *J. Chem. Phys.*, 2016, **144**(21), 214110, DOI: **10.1063/1.4952647**.

47 F. Neese, Software Update: The ORCA Program System—Version 5.0, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2022, **12**(5), e1606, DOI: **10.1002/wcms.1606**.

48 S. Lehtola, C. Steigemann, M. J. T. Oliveira and M. A. L. Marques, Recent Developments in Libxc—A Comprehensive Library of Functionals for Density Functional Theory, *SoftwareX*, 2018, **7**, 1–5, DOI: **10.1016/j.softx.2017.11.002**.

49 G. Luchini, J. V. Alegre-Requena, I. Funes-Ardoiz and R. S. Paton, GoodVibes: Automated Thermochemistry for Heterogeneous Computational Chemistry Data, *F1000Research*, 2020, **9**, 291, DOI: **10.12688/f1000research.22758.1**.

50 E. D. Glendening, C. R. Landis and F. Weinhold, NBO 7.0: New Vistas in Localized and Delocalized Chemical Bonding Theory, *J. Comput. Chem.*, 2019, **40**(25), 2234–2241, DOI: **10.1002/jcc.25873**.

51 W. Yang and R. G. Parr, Hardness, Softness, and the Fukui Function in the Electronic Theory of Metals and Catalysis, *Proc. Natl. Acad. Sci. U. S. A.*, 1985, **82**(20), 6723–6726, DOI: **10.1073/pnas.82.20.6723**.

52 D. Dalmau and J. V. Alegre-Requena, ROBERT: Bridging the Gap Between Machine Learning and Chemistry, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2024, **14**(5), e1733, DOI: **10.1002/wcms.1733**.

53 M. G. Evans and M. Polanyi, Inertia and Driving Force of Chemical Reactions, *Trans. Faraday Soc.*, 1938, **34**(0), 11–24, DOI: **10.1039/TF9383400011**.

54 P. R. Wells, Linear Free Energy Relationships, *Chem. Rev.*, 1963, **63**(2), 171–219, DOI: **10.1021/cr60222a005**.

55 M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller and K. Burke, Quantum Chemical Accuracy from Density Functional Approximations via Machine Learning, *Nat. Commun.*, 2020, **11**(1), 5223, DOI: **10.1038/s41467-020-19093-1**.

56 I. Kageyuki, J. Li and H. Yoshida, Platinum–P(BFPy)3-Catalyzed Regioselective Diboration of Terminal Alkynes with (Pin)B–B(Aam), *Org. Chem. Front.*, 2022, **9**(5), 1370–1374, DOI: **10.1039/D1QO01903A**.

57 W. B. Black, S. Perea and H. Li, Design, Construction, and Application of Noncanonical Redox Cofactor Infrastructures, *Curr. Opin. Biotechnol.*, 2023, **84**, 103019, DOI: **10.1016/j.copbio.2023.103019**.

58 S. J. B. Mallinson, D. Dessaux, S. Barbe and Y. J. Bomble, Computer-Aided Engineering of a Non-Phosphorylating Glyceraldehyde-3-Phosphate Dehydrogenase to Enable Cell-Free Biocatalysis, *ACS Catal.*, 2023, **13**(17), 11781–11797, DOI: **10.1021/acscatal.3c01452**.