Metabolic Engineering 87 (2025) 86–94

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/meteng

Deep learning for NAD/NADP cofactor prediction and engineering using transformer attention analysis in enzymes

Jaehyung Kim [a,1], Jihoon Woo [a,1], Joon Young Park [a], Kyung-Jin Kim [b], Donghyuk Kim [a,*]

[a] School of Energy and Chemical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, 44919, Republic of Korea
[b] School of Life Sciences, BK21 FOUR KNU Creative BioResearch Group, KNU Institute of Microbiology, Kyungpook National University, Daegu, 41566, Republic of Korea

ARTICLE INFO

Keywords:
NAD(P) specificity
Cofactor switching
Deep learning
Explainable AI
Protein engineering
Synthetic biology

ABSTRACT

Understanding and manipulating the cofactor preferences of NAD(P)-dependent oxidoreductases, the most widely distributed enzyme group in nature, is increasingly crucial in bioengineering. However, large-scale identification of the cofactor preferences and the design of mutants to switch cofactor specificity remain as complex tasks. Here, we introduce DISCODE (Deep learning-based Iterative pipeline to analyze Specificity of COfactors and to Design Enzyme), a novel transformer-based deep learning model to predict NAD(P) cofactor preferences. For model training, a total of 7,132 NAD(P)-dependent enzyme sequences were collected. Leveraging whole-length sequence information, DISCODE classifies the cofactor preferences of NAD(P)-dependent oxidoreductase protein sequences without structural or taxonomic limitation. The model showed 97.4% and 97.3% of accuracy and F1 score, respectively. A notable feature of DISCODE is the interpretability of its transformer layers. Analysis of attention layers in the model enables identification of several residues that showed significantly higher attention weights. They were well aligned with structurally important residues that closely interact with NAD(P), facilitating the identification of key residues for determining cofactor specificities. These key residues showed high consistency with verified cofactor switching mutants. Integrated into an enzyme design pipeline, DISCODE coupled with attention analysis, enables a fully automated approach to redesign cofactor specificity.

1. Introduction

NAD(H) and NADP(H), hereafter referred to as simply NAD and NADP, are essential cofactors ubiquitous in all domains of life forms, playing a pivotal role in the transferring reducing equivalents in most oxidoreductase reactions. Despite their near-identical structures, NADP is distinguished by an extra phosphomonoester moiety at the 2′ position of its adenine ribose. This slight structural variance leads to distinct enzymatic affinities for the two cofactors, facilitating functional segregation based on cellular demands (Agledal et al., 2010; Goldford et al., 2022; Russell and Cook, 1995). Managing the balance of these cofactors in line with cellular demands is complex, yet essential for processes like metabolic engineering or synthetic biology, where optimizing these balances is crucial for efficient biochemical production (Bennett et al., 2009; Wang et al., 2013). To overcome these challenges, 'cofactor switching' – altering an enzyme's native cofactor specificity to its alternative form – has emerged as a strategic approach (Chánique and Parra, 2018; Vidal et al., 2018; Wang et al., 2017). This can either replenish the cofactor supplies (Cheng et al., 2023; Jia et al., 2022; Ma et al., 2023; Son et al., 2023) or tailor the enzymatic cofactor preference to align with the host organism's metabolism (Jia et al., 2022; Meng et al., 2016; Pick et al., 2014). Furthermore, King & Feist conducted a comprehensive study using constraint-based modeling to analyze cofactor switching impacts. Their research showed that cofactor switching can enhance the production yields of various substances in *Escherichia coli* and *Saccharomyces cerevisiae* (King and Feist, 2014).

Extensive studies on NAD(P)-dependent enzymes have shed light on protein engineering techniques for cofactor switching. A predominant NAD(P) binding motif in these enzymes is the Rossmann fold (Medvedev et al., 2019, 2021; Rossmann et al., 1974). However, studies have also identified other NAD(P)-dependent oxidoreductases with different structural motifs, such as TIM barrel, 3-Dehydroquinate synthase-like

* Corresponding author.
E-mail address: dkim@unist.ac.kr (D. Kim).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ymben.2024.11.007
Received 28 March 2024; Received in revised form 25 September 2024; Accepted 17 November 2024
Available online 20 November 2024

fold, and FAD/NAD-binding fold (Brakoulias and Jackson, 2004; Campbell et al., 2013; Carpenter et al., 1998; Nagano et al., 2002). Since the initial breakthrough in protein engineering for cofactor switching (Scrutton et al., 1990), numerous instances of successful cofactor switching have been documented (Chánique and Parra, 2018). These studies highlight that cofactor preferences often hinge on specific residues near the adenine moiety of bound NAD(P) (Cahn et al., 2017; Carugo and Argos, 1997; Laurino et al., 2016). Additionally, the presence of glycine-rich motifs, GXXXXG/A, in the Rossmann fold, has been noted to influence enzyme's cofactor preferences (Dambe et al., 2006; Kleiger and Eisenberg, 2002). Nonetheless, these findings suggest that cofactor preferences are influenced by the overall structure of the binding pocket, and rational engineering requires thorough investigation, experiment, and expertise. Furthermore, approaches like random mutagenesis and screening face limitations due to the vast number of potential combination of mutants (Cahn et al., 2017; Naylor et al., 2001).

To overcome the complexities in designing cofactor switching, various computational strategies have been employed to either differentiate or redesign cofactor specificities. These methods range from physics-based simulations, sequence or structure-based studies, and machine learning-based predictions (Cahn et al., 2017; Cui et al., 2015; Geertz-Hansen et al., 2014; Kallberg and Persson, 2006; Kaminski et al., 2022; Khoury et al., 2009; Sugiki et al., 2022). Among these, machine learning-based Cofactory and Rossmann-toolbox stand out for their ability to perform high-throughput, sequence-based predictions of cofactor specificities (Geertz-Hansen et al., 2014; Kaminski et al., 2022). However, their effectiveness is largely contingent on identification of Rossmann fold motifs, limiting their applicability to variants of these motifs and other types of NAD(P)-dependent enzymes. Additionally, their use in mutant design is constrained by the computational costs involved in examining the vast array of possible sequence combinations.

Deep learning models, while achieving remarkable successes in analyzing diverse biological data, are often criticized as 'black boxes' due to their opaque reasoning processes. This lack of transparency hinders interpretability and trust in their decision-making capabilities. Explainable AI (XAI) emerges as a solution to address these limitations (Karim et al., 2023). Existing XAI methods, such as Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016), Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al., 2017), and Integrated Gradients (Sundararajan et al., 2017), have provided insights into the behavior of models in protein property prediction (Chen et al., 2021; Kim et al., 2021; Yang et al., 2023). However, the transformer architecture (Vaswani et al., 2017), currently the state of the art in natural language processing, offers a unique advantage for interpretation in protein sequence analysis. By analyzing its multi-head self-attention layers, we can directly assess the importance of each input token in the model's decision-making process, intuitively unveiling its focus from multiple perspectives. Moreover, the transformer excels at processing sequential data and capturing long-range dependencies, which are crucial aspects of protein sequences. This makes it more suitable than other deep learning models, like convolutional neural networks (CNN) that primarily extract local features and struggle with sequential data processing (Chandra et al., 2023). Recently, while transformer-based protein sequence models have emerged (Chandra et al., 2023; Lin et al., 2023), a limited number of studies explored their attention layers (Kim et al., 2023; Zhou et al., 2023), especially for practical bioengineering applications. This gap underscores the potential for further research in leveraging transformer interpretability for more insightful protein sequence analysis.

In this study, we present DISCODE (Deep learning-based Iterative pipeline to analyze Specificity of COfactors and to Design Enzyme), a broadly applicable model for classifying NAD/NADP cofactor specificity. This model incorporates the transformer architecture due to its explainability and ability to capture long-range dependencies common in protein sequences, further enhanced with ESM-2 embeddings (Lin et al., 2023; Vaswani et al., 2017). Leveraging the self-attention mechanism inherent to the transformer architecture (Chandra et al., 2023; Hao et al., 2021), DISCODE effectively predicts the cofactor preferences of NAD(P)-dependent oxidoreductases across a wide range of sequences. This makes it universally suitable for any NAD(P)-dependent oxidoreductases. Moreover, the attention-based interpretative capability of DISCODE allows for a meaningful representation of protein sequences in relation to cofactor specificities. This, in turn, offers valuable insights and guidelines for the design of site-directed mutants aimed at cofactor switching.

2. Methods

2.1. NAD(P)-dependent oxidoreductases dataset

NAD and NADP binding protein sequences were retrieved from the Swiss-Prot database (released May 2023) (Bateman et al., 2023). The enzymes showed dual selectivity were excluded. To prevent over-representation of redundant proteins, the collected sequences were clustered based on 90.0% similarity using the UCLUST algorithm built in USEARCH (Edgar, 2010). The remaining protein sequences were further manually curated based on several criteria: The reactions that are not related to transfer reducing equivalents, such as NAD, NADP synthesis, degradation, transport reactions, and allosteric bindings were excluded from the dataset. For multi-subunit enzymes, the only subunit directly binding with NAD or NADP (e.g. a reductase subunit) was retained in the dataset. Fragmented sequences and those with a length over 512 were excluded (this does not indicate that DISCODE has a sequence length limitation). The dataset was further trimmed in cases when other cofactors such as FAD is preferred over NAD or NADP based on literature search. To characterize the dataset, an EC number, taxonomy, and CATH code were assigned to each protein sequence. The EC number and taxonomy were retrieved from the metadata in UniProt (Bateman et al., 2023). For CATH classification, the dataset was annotated using InterProScan 5, and a CATH code was assigned based on G3DSA accession (Jones et al., 2014; Orengo et al., 1997). A total of 7,132 sequences are listed in Supplementary dataset 1.

2.2. Cofactor switching mutation dataset

To collect cofactor switching cases, sequences and kinetic parameters were retrieved from extensive literature search. Catalytic efficiency was defined (equation (1)) using K_{cat} and cofactor K_m values, and it was used to calculate the cofactor specificity ratio (equation (2)). Given that NAD(P)-dependent enzymes generally follow a Bi-Bi reaction mechanism, we used cofactor K_a value as the apparent K_m when the concentration of the other substrate remain constant. A case is considered reversed only if the cofactor specificity ratio exceeds 1.

$$\textit{Catalytic efficiency} = \frac{k_{cat}}{K_m} \quad \text{(equation 1)}$$

$$\textit{Cofactor specificity ratio} = \frac{CE_{NAD\ or\ NADP}}{CE_{NADP\ or\ NAD}} \quad \text{(equation 2)}$$

when calculating the catalytic efficiency, the kinetic parameters of both wild type and mutants were considered based on their respective ligand preferences: the wild type for its original preference and the mutants for the opposite preference. A total of 43 cases and their kinetic parameters are summarized in Supplementary dataset 2. The assay conditions (pH and temperature) for each case were within typical biological ranges.

2.3. Model architecture and training

DISCODE consists of eight encoder layers of the transformer (Vaswani et al., 2017) and two fully connected layers (Fig. 1A). Each

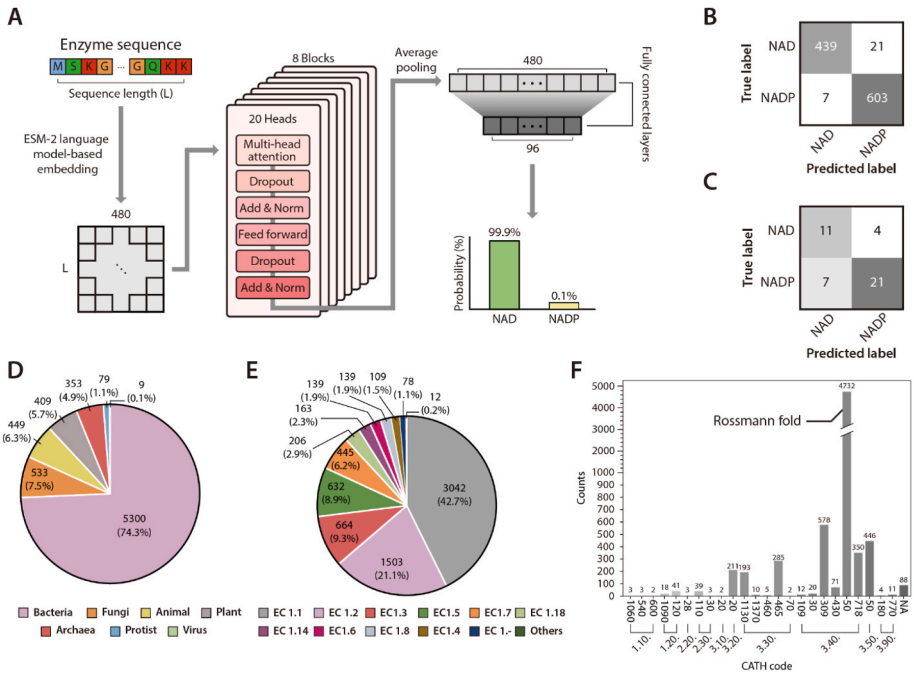

Fig. 1. Architecture overview of transformer-based enzyme sequence model, its performance assessment, and the characteristics of sequence dataset used for model training. A) Description of the DISCODE model architecture. This model was built upon a transformer framework consisting of eight consecutive encoder layers, each containing 20 self-attention heads. The initial step involves preprocessing the input enzyme sequence using the ESM-2 language model, which handles tokenization and embedding. Following this, the sequence undergoes processing through the encoder layers, global average pooling, and fully connected layers. In the final stage, sigmoid function is employed to transform the logits into probability scores, indicating cofactor preferences for either NAD or NADP. **B)** Confusion matrix of the test dataset. The transformer model underwent training, validation, and testing using the dataset split training, validation, and test sets in a 70:15:15 ratio. The generated confusion matrix for the test set demonstrates an accuracy performance of 97.4%. **C)** Confusion matrix of the mutation dataset. The transformer model, once trained, was applied to predict the preference labels of curated laboratory mutants, specifically those altered for NAD/NADP cofactor switching through site-directed mutagenesis. Notably, the model accurately predicted 32 out of 43 mutants that exhibited reversed cofactor specificity. **D)** Taxonomic classification of the enzyme sequences. Among the total 7,132 enzyme sequences analyzed, a substantial 74.3% originated from bacteria. The rest were distributed among five other groups: fungi, plants, animals, protists, and archaea. Additionally, there were 9 sequences that, unlike the rest, were derived from viruses rather than traditional organisms. **E)** Classification of enzyme sequences by EC number. In the dataset, the vast majority of enzyme sequences were identified as oxidoreductases (EC 1, 99.8%). The remaining 12 sequences fell into others categories. Within the EC 1 group, these enzyme sequences were further divided into 11 distinct subcategories, each characterized by specific functional attributes. **F)** CATH-based classification (Orengo et al., 1997). Most sequences belong to the alpha beta fold (3.-) class. This includes the Rossmann fold (3.40.50), a well-known structural motif for NAD(P) binding. A small fraction of sequences belonged to 1.- (mainly alpha) and 2.- (mainly beta) classes.

encoder layer comprises 20 heads and employs the GELU (gaussian error linear unit) activation function internally. At the final stage of the model, the sigmoid function is utilized as the activation function, while binary cross entropy is applied as the loss function, respectively. The input protein sequence is pre-processed by ESM-2 transformer (Lin et al., 2023) for tokenization and embedding. The dimension of the embedded sequences remains at (L, 480; L is length of input sequence) until it leaves the encoder. Using global average pooling (GAP), a feature vector is extracted from the final layer of the encoder with 480 dimensions. The feature vector is subsequently processed by two fully connected layers with a decreasing number of nodes (96 and 2, respectively), followed by label prediction using the sigmoid function. Total 7,132 NAD (P)-dependent oxidoreductase sequences were divided into training, validation, and test datasets, with the proportions approximately 70:15:15, respectively. During training, early stopping was employed by monitoring the validation loss (Supplementary Fig. 1). DISCODE was implemented using PyTorch v.2.0.1 (Paszke et al., 2019). The well-trained model and its associated codes are available in the GitHub repository (https://github.com/SBML-Kimlab/DISCODE).

2.4. Benchmarking of models

To evaluate predictive performance of DISCODE, the model was compared with Cofactory (Geertz-Hansen et al., 2014) and Rossmann-toolbox (Kaminski et al., 2022) using the test dataset described in above section (1,070 sequences). For benchmarking using the test dataset, since both Cofactory and Rossmann-toolbox require Rossmann motif detection, only the sequences that succeeded in the detection were evaluated and cases which fail to detect the motifs were treated as ‘not predicted’. For the three tools, predictions were designated as specific cofactor only in cases where the label value with the highest probability was over 0.5. If all label probabilities were below 0.5, the sequence was considered as ’others’. In addition, FAD and SAM labels in results of Cofactory and Rossmann-toolbox were designated as ‘others’. As metrics for comparison of the results, accuracy and F1 score were calculated for NAD and NADP labels, where TP, TN, FP, and FN denoted true positive, true negative, false positive, and false negative, respectively (equations (3) and (4)). Macro-scale accuracy and F1 score were calculated as by averaging corresponding metrics.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad \text{(equation 3)}$$

$$F1\ score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad \text{(equation 4)}$$

2.5. Model interpretation

Attention weights of an input sequence can be obtained from the 20 heads within each of 8 attention layers of DISCODE. The overall shape of attention layers is [8, 20, L, L] where L is a length of an input sequence.

To visualize a maximum attention map, the maximum value among [L, L] shaped attention matrix in each head was taken and the shape of the maximum attention map is [8, 20] to indicate the maximum value of each head. In addition, to calculate the attention sum for each residue, attention values from all attention layers were first summed by position. The resulting matrix was then summed along the vertical axis to produce a 1-dimensional array with a shape of [L]. If the overall attention score matrices are denoted as A and if its axes are *i*, *j*, *k*, and *l*, equation (5) describes how the attention sum array is derived, which contains the attention sum of each residue.

$$\textit{Attention sum array} = \sum_{j}\sum_{k}\sum_{l} A_{i,j,k,l} \quad \text{(equation 5)}$$

Then, outlier residues were chosen based on a criterion given by equation (6), using the mean (*M*) and the standard deviation (*S*) calculated from the attention sum array. Only positions that satisfy the threshold are selected as outlier residues. Our test results for other threshold metrics are shown in Supplementary Fig. 6.

$$\textit{Attention score}_i > M + 2{*}S \quad \text{(equation 6)}$$

3. Results

3.1. Development of a transformer-based model for predicting NAD(P) preference

DISCODE, a transformer-based deep neural network model, was developed for the purpose of determining the preference of NAD(P)-dependent oxidoreductases between NAD(H) or NADP(H), based on their protein sequences. This model was designed to analyze entire protein sequences for preference prediction, thereby avoiding limitations associated with specific structural motifs like the Rossmann fold. A secondary aim of this model was to provide explainable AI functionality. Therefore, we adopted a transformer architecture, which is effective not only for sequential data processing but also for interpreting the results. The model incorporates 8 encoder layers from the transformer architecture (Vaswani et al., 2017), each featuring 20 heads (Fig. 1A). Protein sequences undergo initial preprocessing using the ESM-2 transformer model for embedding (Lin et al., 2023), after which the resulting matrix is input into the multi-head attention layer. Following this, the feature vector is processed via average pooling and fully connected layers to make a prediction.

For model training, protein sequences of NAD(P)-dependent oxidoreductases were sourced from the Swiss-Prot database (released May 2023) (Bateman et al., 2023). These sequences were further carefully curated, excluding enzymes with dual cofactor selectivity, resulting in a collection of 7,132 non-redundant sequences (detailed in Methods). The dataset was then randomly partitioned into training, validation, and test datasets following a 70:15:15 ratio. To avoid overfitting during the training process, strategies such as early stopping, batch normalization and dropout were implemented (Fig. 1A, Supplementary Fig. 1). Consequently, the model demonstrated a high accuracy of 0.9738 on the test dataset (Fig. 1B).

In the creation of NAD(P) preference prediction model with universal applicability, an extensive array of protein sequences was gathered. Each sequence was analyzed for its taxonomic origin, EC number, and CATH identifier (Supplementary dataset 1) (Orengo et al., 1997). The majority of these sequences were bacterial (5,300, 74.3%) and archaeal (353, 4.9%). A significant fraction also originated from the eukaryotic domain (1,470, 20.6%), including fungi (533, 7.5%), animals (449, 6.3%), plants (409, 5.7%), and protists (79, 1.1%). Additionally, 9 virus-derived sequences (0.1%) were identified (Fig. 1D). The dataset predominantly consisted of oxidoreductases (EC 1), totaling 7,120 sequences (99.8%), spread across more than 10 sub-categories (Fig. 1E). According to CATH identifiers, the most common structural group was the Rossmann fold (3.40.50.-, 4,732, 66.3%) and its strict portion (3.40.50.720, 4,248, 59.6%). Nonetheless, a significant proportion showed structural variations while retaining the alpha-beta fold characteristic of the Rossmann domain—97.2% for the 3.- (alpha-beta) group and 80.8% for the 3.40.- (alpha-beta-alpha sandwich) group. Notably, 2.8% (200) of sequences had different CATH identifiers starting with 1 (mainly alpha) or 2 (mainly beta) (Fig. 1F). These included non-Rossmann folds like TIM barrel (3.20.20), FAD/NAD(P)-binding domain superfamily (3.50.50), and 3-dehydroquinate synthase-like fold (1.20.1090.10 or NA). This diversity underscores the importance of not relying solely on the Rossmann fold domain for a universally applicable model, considering the convergent or divergent evolution of NAD(P)-dependent oxidoreductases.

3.2. Evaluation of the predictive accuracy of DISCODE and analysis of cofactor switching mutation impacts

The predictive performance of DISCODE was benchmarked against other cofactor prediction models, namely Cofactory (Geertz-Hansen et al., 2014) and Rossmann-toolbox (Kaminski et al., 2022). Cofactory utilizes a feed-forward fully connected network for classifying NAD, NADP, and FAD labels, whereas Rossmann-toolbox employs a convolutional neural network to classify NAD, NADP, FAD, and SAM labels. These tools, reliant on the Rossmann fold sequence detection, had their accuracy assessed only in scenarios where such detection was feasible. The evaluation used a test dataset (detailed in Methods) for comparison and benchmarking. DISCODE surpassed both Cofactory and Rossmann-toolbox in terms of accuracy and F1 scores (Table 1). Impressively, DISCODE managed to process all 1,070 sequences in the dataset, in contrast to Cofactory which processed 568, and Rossmann-toolbox which processed 698 due to their preprocessing steps (Table 1). Furthermore, when focusing on the Rossmann fold subset of the test dataset, comprising 635 sequences with the 3.40.50.720 CATH identifier, DISCODE continued to exhibit higher accuracy and F1 scores compared to the alternative tools (Table 1). This was noteworthy, even as Cofactory and Rossmann-toolbox displayed improved accuracies, aligning with their previously reported performance (Kaminski et al., 2022).

The effectiveness of DISCODE was further assessed using a dataset of experimentally verified mutations. Following specific selection using the criteria (detailed in Methods), a collection of 43 instances of site-directed mutagenesis was compiled, each demonstrating a complete shift in cofactor specificity. For all these instances, the model accurately predicted all 43 labels of the wild-type sequences. When applied to the mutant sequences, DISCODE correctly predicted the cofactor specificity reversal in 32 out of the 43 cases (74.4%) (Fig. 1C). This indicates that while most evolutionary traits from the full-length sequence still retain features favoring the original cofactor preference, the model could sensitively discern the effects of the cofactor switching mutations.

In summary, DISCODE exhibits higher accuracy and a more extensive range of predictive capabilities compared to other available tools. Furthermore, it is able to reasonably predict the outcomes of mutations, though with a marginally lower level of accuracy compared to its predictions for wild-type cases.

3.3. Analyzing protein sequences for cofactor specificities using attention-based interpretation

In an effort to comprehend how DISCODE discerns cofactor specificity, an analysis of the attention layers of the model was conducted, as attention weights can be interpreted as indicating the relative importance between tokens in this binary classification. This examination was exemplified by analyzing the attention weights of *Neurospora crassa* L-arabinitol 4-dehydrogenase (PDB: 3M6I), an NAD-dependent Rossmann fold enzyme whose cofactor switching has been well studied. The attention weights across its 160 heads (20 heads per each of the 8 layers) showed a vertical pattern (Supplementary Fig. 2). Within these heads,

Table 1
Evaluation of DISCODE predictive accuracy relative to other models.

Dataset	Classification Methods	True label	Prediction				Accuracy	F1 score
			NAD	NADP	Others	No		
Test dataset	Cofactory	NAD	230	9	21	200	0.7465	0.7455
		NADP	62	194	52	302		
	Rossmann-toolbox	NAD	239	18	26	177	0.8209	0.8175
		NADP	19	334	62	195		
	DISCODE	NAD	439	21	0	0	0.9738	0.9732
		NADP	7	603	0	0		
Rossmann fold subset	Cofactory	NAD	209	9	5	42	0.8361	0.8359
		NADP	60	104	5	111		
	Rossmann-toolbox	NAD	239	18	5	3	0.9124	0.9104
		NADP	19	334	13	4		
	DISCODE	NAD	253	12	0	0	0.9764	0.9756
		NADP	3	367	0	0		

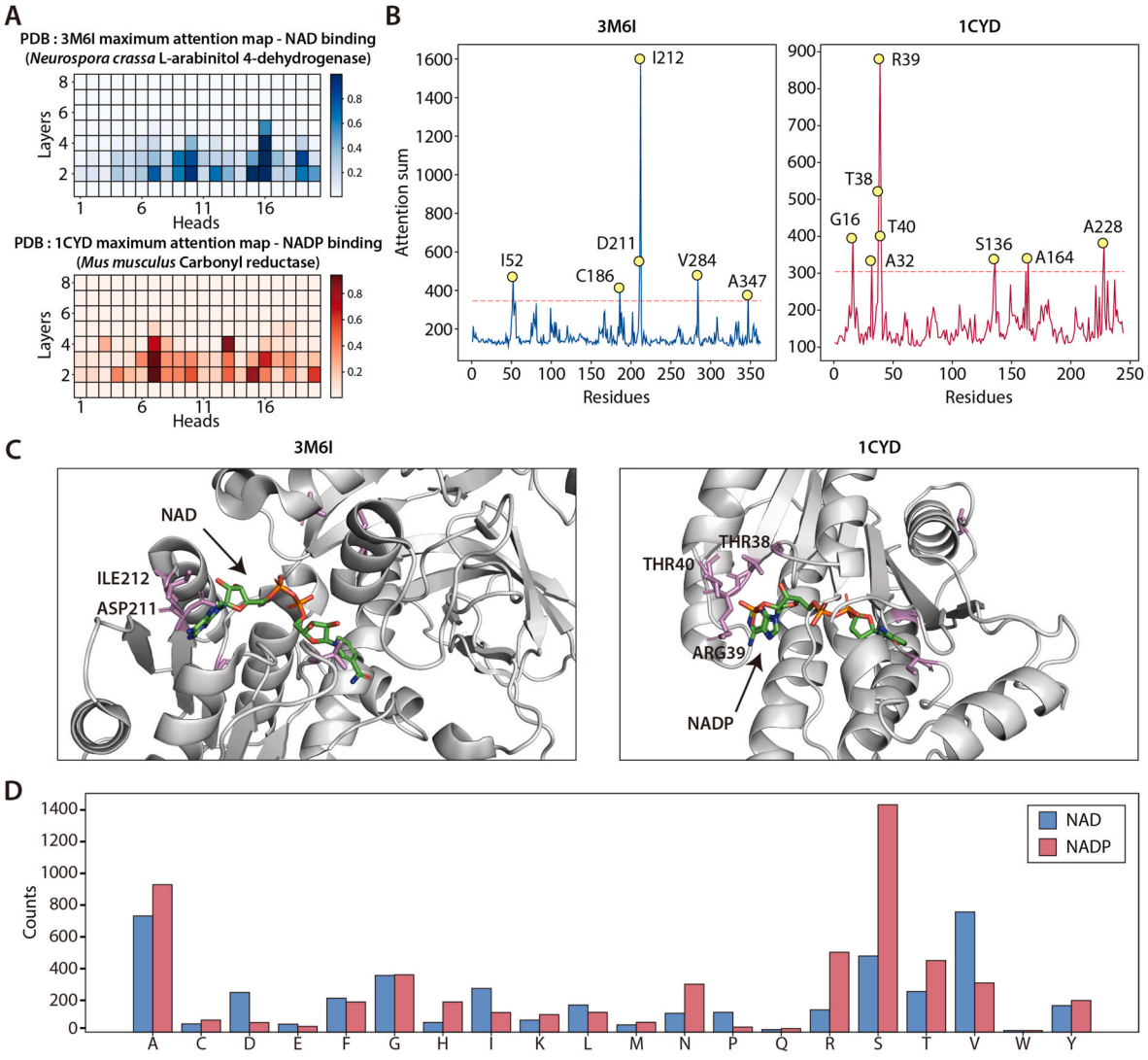

Fig. 2. Interpretation of the trained transformer model focusing on attention mechanisms to identify key residues determining cofactor specificity between NAD and NADP. A) Representation of maximum attention maps of 3M6I and 1CYD, corresponding to NAD- and NADP-dependent enzymes, respectively. In these maps, each value signifies the maximum attention weight for each respective head. Notably, the attention maps for these two enzymes exhibited distinct patterns. **B)** Interpretation of the model based on the attention sum. The red dotted lines indicate the threshold for identifying outlier residues, and yellow circles highlight the outlier residues. **C)** Crystal structures of 3M6I and 1CYD. The outlier residues are highlighted in purple. In 3M6I, residues D211 and I212 are located in proximity to adenine and ribose moieties in NAD. In contrast, in 1CYD, residues T38, R39, and T40 are positioned close to the adenine, ribose, and phosphate moieties of NADP. This highlights the correlation between the attention sum and structurally significant positions that influence cofactor specificities. **D)** Frequency distribution of amino acids in outlier residues from the test dataset. The amino acid counts exhibited varying distributions between NAD- and NADP-dependent enzymes.

certain patterns exhibited uniformly low attention weights, while others had more pronounced attention weights concentrated at specific positions (Supplementary Fig. 2). These latter instances were highlighted by visualizing the maximum attention of each head, which is likely to provide residue-specific insights into cofactor specificities (Fig. 2A). A residue-specific attention sum was calculated by summing all attention weights for each position, based on the vertical pattern and focused attention on certain residues (equation (5)). Consequently, the attention sum for 3M6I was visualized (Fig. 2B), revealing 6 outlier positions according to a threshold (equation (6)). Among these, D211 and I212 stood out with higher scores. In comparison with the experimental structure (PDB: 3M6I), both D211 and I212 were part of a structural motif in the Rossmann fold, near the adenine moiety of NAD. The original research (Bae et al., 2010) indicated the critical role of D211 and I212 in NAD specificity, with a D211S/I212R double mutant exhibiting a reversed cofactor specificity toward NADP. Furthermore, D211 might serve as an evolutional signature of NAD-specific dehydrogenase (Laurino et al., 2016).

In a second example, using *Mus musculus* carbonyl reductase (PDB: 1CYD), an NADP-dependent Rossmann fold enzyme whose cofactor switching also has been well studied, analysis of its maximum attention map revealed distinct patterns from those of 3M6I (Fig. 2A). While high maximum attentions were similarly noticed in layers 2, 3, 4, and 5, the specific heads within these layers exhibiting high scores varied considerably. This variation suggests that different heads in the model may contribute differently to the prediction of NAD and NADP preferences. To investigate this, the average maximum attention maps were calculated for both NAD (460 sequences) and NADP (610 sequences) cases from the 1,070 sequences in the test dataset (Supplementary Fig. 3). These averaged maps mirrored the patterns observed in 3M6I and 1CYD (Fig. 2A), supporting the hypothesis that these trends are not exclusive to just these two instances. The attention sum analysis of 1CYD identified eight outlier residues, with T38, R39, and T40 showing higher attention sums. These residues are located within the structural motif of the Rossmann fold, adjacent to the adenine moiety of NADP (PDB: 1CYD; Fig. 2C). According to its cofactor switching study, T38 is crucial for interaction with the 2′-phosphate moiety of NADP and serves as the counterpart of Asp in NAD(H)-dependent oxidoreductases (Nakanishi et al., 1997). Notably, the T38D mutant was reported to showing a sufficient shift in cofactor specificity towards NAD (Nakanishi et al., 1997). These findings imply that in both NAD and NADP cases, the attention sum of each residue is a reliable indicator of its significance in determining cofactor specificity.

To know which amino acids are frequently discerned as important by the model for each cofactor, we performed a cumulative count of the amino acid types from the outlier residues identified in the attention analysis (Fig. 2D). This analysis encompassed the 1,070 sequences from the test dataset, highlighting the differences in the amino acid frequency of outlier residues between NAD and NADP-dependent enzymes. In the case of NAD-dependent enzymes, amino acids such as Asp, Ile, Pro, and Val were found to be more prevalent compared to their occurrence in NADP-dependent enzymes (Fig. 2D). On the other hand, for NADP-dependent enzymes, amino acids like Arg, His, Ser, and Thr were observed more frequently than in NAD-dependent enzymes (Fig. 2D). These variations in amino acid distribution identified by the model consistent with previous studies, which empirically analyzed amino acid frequencies in the context of mutations known to influence cofactor switching (Cahn et al., 2017; Chánique and Parra, 2018).

Overall, the attention sum generated by the model serves as a reliable indicator of the significance of particular residues in establishing cofactor specificities. This interpretation is bolstered by the analysis of two enzymes with distinct cofactor specificities, the aggregated data from maximum attention maps, and the patterns observed in amino acid distributions. Consequently, the attention analysis enables a deeper understanding of the 'black box' of the model, shedding light on its predictive processes. Notably, the optimized attention layers refined during training appear to effectively capture changes most relevant to NAD/NADP specificity, highlighting their role in discerning subtle differences between cofactors. This suggests that the model not only performs well in prediction but also suggests a strategic method for ranking the importance of residues when designing mutations.

3.4. Utilizing attention analysis for mutation cases and a pipeline for a design of cofactor switching

In evaluating the relationship between residues identified through attention analysis and those used in experimental cofactor design, we compared the outlier residues from the wild-type sequences in the mutation dataset (detailed in Methods) with the positions of experimentally verified side-directed mutagenesis for cofactor switching. The findings revealed that in 81.4% of the instances (35 out of 43 cases), there was at least one point of correspondence between the identified outliers and the mutation sites (Fig. 3A). This underscores the model's capability in identifying viable candidates for mutation, particularly for cofactor switching purposes. It is important to note, however, that since these mutations are not derived from high-throughput experiments, it is not feasible to verify every outlier residue. Moreover, the lack of overlap in the remaining seven cases does not necessarily imply incorrect predictions, given the variety of potential strategies in mutation design. For instance, the Q362K mutant of the human malic enzyme (PDB: 1PJ3) (Hsieh et al., 2006), which is known for its reversed cofactor specificity, does not reside within its Rossmann fold motif (Supplementary Fig. 4). However, the substitution of 345D or 346K, as suggested by the attention analysis, could also serve as promising targets for cofactor switching (Supplementary Fig. 4).

For the development of complex mutants that go beyond single-site mutagenesis, the attention sum of a mutant sequence was analyzed. Taking the NAD-dependent *Escherichia coli* pyruvate dehydrogenase enzyme (PDB: 4JQ9) (Bocanegra et al., 1993), DISCODE accurately predicted its preference for NAD and pinpointed nine outlier residues in the wild-type sequence. Among these outliers, E205 and M206 were also identified in experimental designs for cofactor switching. When E205V or M206R were individually mutated, the model still predicted a preference for NAD. However, introducing a E205V/M206R double mutation shifted the model prediction towards NADP preference, with a probability of 0.9860 (Fig. 3B). Intriguingly, the attention analysis of the double mutant highlighted F207K as a new outlier residue, aligning with experimental designs (Fig. 3C). For the assessment of the E205V/M206R/F207K triple mutant indicated an NADP preference with a slightly increased probability of 0.9957 (Fig. 3B). This outcome implies that reselecting outliers from mutant sequences can potentially uncover more effective mutation targets for the modified sequence, compared to those originally identified from the wild-type sequence.

A novel iterative pipeline for the automatic design of cofactor switching mutants has been established, leveraging an attention-based approach to select residues from both wild-type and mutant sequences. The pipeline operates through four iterative stages: (i) DISCODE first predicts the cofactor preference of the input sequence; (ii) it then identifies potential substitution sites via attention analysis; (iii) mutant sequences are subsequently generated by altering amino acids at these identified positions; (iv) finally, the model evaluates the cofactor preference of each mutant, selecting those with a reversed cofactor preference (Fig. 3D). For more complex designs involving multiple mutations ($N \geq 2$, where N represents the number of mutations), combinations are generated using outlier residues from both the wild-type sequence and mutant sequences derived from the previous iteration round (Fig. 3D).

This pipeline is capable of generating designs for cofactor switching in a computationally efficient manner, suitable for use on a standard desktop computer, even for intricate designs. It also offers options to prioritize either maximizing the probability of preference changes or optimizing computational efficiency. In the case of the mutation dataset,

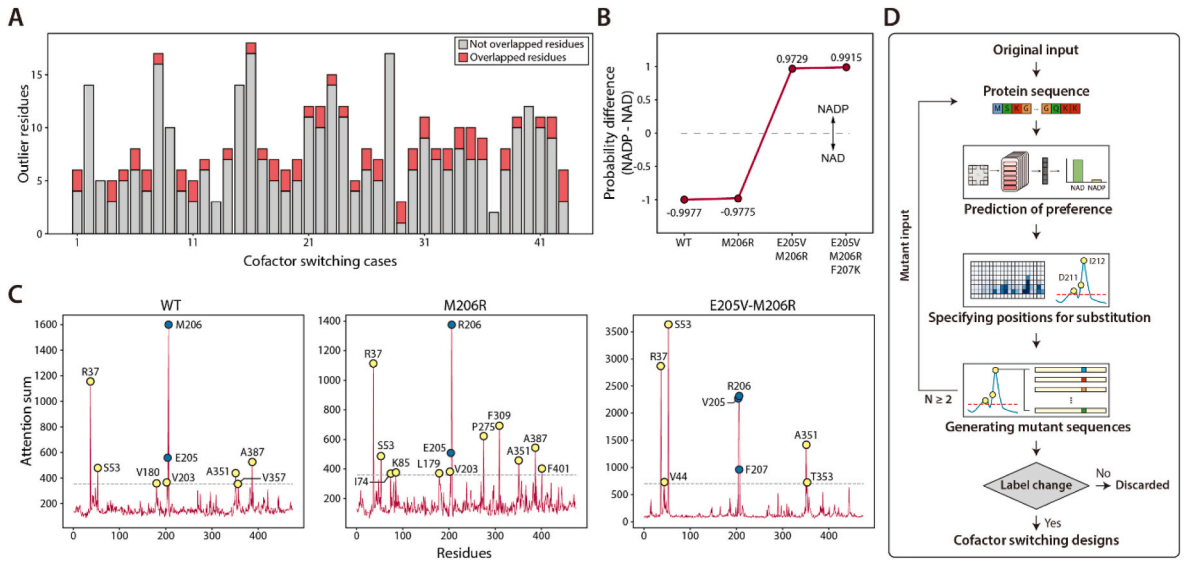

Fig. 3. The mutation design pipeline facilitated by DISCODE for cofactor switching. A) Overlaps between outlier residues and experimental designs for cofactor switching. Each value of horizontal axis specifies corresponding indices of cofactor switching cases in Supplementary dataset 2. In 35 out of the 43 cases analyzed (81.4%), there was at least one instance of overlap. **B)** Sequential mutation impact on label prediction for 4JQ9. The *Escherichia coli* pyruvate dehydrogenase (4JQ9) and its M206R mutants were identified as having a preference for NAD. In contrast, E205V/M206R and E205V/M206R/F207K mutants were predicted to prefer NADP. **C)** Attention analysis of 4JQ9 during sequential mutations. Outlier residues are depicted as circles, with those experimentally verified as mutations highlighted in blue. **D)** Schematic diagram representing the designing pipeline of DISCODE. This pipeline is structured into four essential stages: 1) Predicting the cofactor preference label for the input enzyme sequence; 2) Identifying suitable positions for mutation within the sequence; 3) Creating mutant sequences for evaluation and screening; 4) The model then predicts the cofactor preference labels of each mutant, selectively retaining mutants that demonstrate a reversal in their original label.

the pipeline successfully designed cofactor switching mutants in the majority of instances (41 out of 43 cases), with the time required ranging from seconds to a several hours (Supplementary dataset 3). In each case, the designed mutant with the highest probability of cofactor reversal was selected and compared with the verified mutations. As a result, 82.9% (34 out of 41 cases) showed overlaps in the mutation positions (Supplementary Fig. 5). When comparing the entire set of designed mutants, the overlap increased slightly to 87.8% (36 out of 41 cases).

4. Conclusion

In this study, we present DISCODE, a novel transformer-based model, purpose-built for NAD(P) cofactor preferences in enzymes. Our findings demonstrate that DISCODE is capable of accurately classifying protein sequences, free from the confines of structural or taxonomic limitations. Furthermore, DISCODE proves to be an effective tool for predicting cofactor switching mutation sequences and identifying key residues through attention analysis. A key feature of DISCODE is its designing pipeline, which facilitates the fully automated generation of cofactor switching designs, operating independently of additional data requirements.

The exploration of deep learning models, particularly the attention layers of the transformer architectures, has sparked considerable interest in recent years (Kovaleva et al., 2019; Sundararajan et al., 2017; Vaswani et al., 2017; Wiegreffe and Pinter, 2019). However, their application in biological research remains relatively limited. This study showcases a practical application of interpreting the attention layers in a protein sequence-based model. The fully automated design pipeline with DISCODE effectively overcomes the computational hurdles typically encountered in *in silico* screening due to combinatorial complexity. This makes DISCODE accessible for researcher, whether designing a small number of rational mutations or generating a vast array of protein designs, which could potentially be useful for biofoundry-like platforms (Lee et al., 2023).

Nevertheless, it is crucial to acknowledge that DISCODE can produce incorrect predictions in scenarios where the NAD(P) utilization in proteins is ambiguous or poorly defined. In these instances, it is advisable to use DISCODE in conjunction with experimental evidence or annotation tools such as Pfam (Mistry et al., 2021) or COFACTOR (Zhang et al., 2017). As the field of high-throughput protein engineering progresses (Madhavan et al., 2021), further enhancements and learning about cofactor switching mutations are anticipated to enhance the capability of our model. Looking ahead, our research aims to apply mutations designed by DISCODE in practical experimental settings, potentially offering new insights and advancements in the field.

CRediT authorship contribution statement

Jaehyung Kim: Writing - review & editing, Data curation, Investigation, Formal analysis, Visualization, Methodology, Software. **Jihoon Woo:** Writing - original draft, Writing - review & editing, Conceptualization, Data curation, Investigation, Formal analysis, Visualization. **Joon Young Park:** Visualization, Writing - review & editing. **Kyung-Jin Kim:** Writing - review & editing. **Donghyuk Kim:** Writing - review & editing, Conceptualization, Supervision, Project administration, Funding acquisition, Resources. J.K and J.W contributed equally to this work.

Declaration of competing interest

The authors declare that there is no conflicts of interest.

Acknowledgements

This research was supported by National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) [2021M3A9I4024840, 2022M3A9I5018934, RS-2023-00208026, RS-2024-00398252]. This research was funded by The Circle Foundation (Republic of Korea) as UWCC selected as the '2020 TCF Innovative Science Project'.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymben.2024.11.007.

Data availability

The DISCODE model and associated codes are freely available at the public GitHub repository (https://github.com/SBML-Kimlab/DISCODE).

References

Agledal, L., Niere, M., Ziegler, M., 2010. The phosphate makes a difference: cellular functions of NADP. Redox Rep. 15, 2–10.

Bae, B., Sullivan, R.P., Zhao, H., Nair, S.K., 2010. Structure and engineering of L-arabinitol 4-dehydrogenase from Neurospora crassa. J. Mol. Biol. 402, 230–240.

Bateman, A., Martin, M.J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., Gonzales, L.J.D., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Joshi, V., Jyothi, D., Kandasaamy, S., Lock, A., Luciani, A., Lugaric, M., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Mishra, A., Moulang, K., Nightingale, A., Pundir, S., Qi, G.Y., Raj, S., Raposo, P., Rice, D.L., Saidi, R., Santos, R., Speretta, E., Stephenson, J., Totoo, P., Turner, E., Tyagi, N., Vasudev, P., Warner, K., Watkins, X., Zellner, H., Bridge, A.J., Aimo, L., Argoud-Puy, G.L., Auchincloss, A.H., Axelsen, K.B., Bansal, P., Baratin, D., Neto, T. M.B., Blatter, M.C., Bolleman, J.T., Boutet, E., Breuza, L., Gil, B.C., Casals-Casas, C., Echioukh, K.C., Coudert, E., Cuche, B., de Castro, E., Estreicher, A., Famiglietti, M.L., Feuermann, M., Gasteiger, E., Gaudet, P., Gehant, S., Gerritsen, V., Gos, A., Gruaz, N., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Kerhornou, A., Le Mercier, P., Lieberherr, D., Masson, P., Morgat, A., Muthukrishnan, V., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Poux, S., Pozzato, M., Pruess, M., Redaschi, N., Rivoire, C., Sigrist, C.J.A., Sonesson, K., Arighi, C.N., Armin-ski, L., Chen, C.M., Chen, Y.X., Huang, H.Z., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q.H., Wang, Y.Q., Zhang, J., Bye-A-Jee, H., Zaru, R., Sundaram, S., Wu, C.H., Consortium, U., 2023. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. 51, D523–D531.

Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., Rabinowitz, J.D., 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat. Chem. Biol. 5, 593–599.

Bocanegra, J.A., Scrutton, N.S., Perham, R.N., 1993. Creation of an NADP-dependent pyruvate dehydrogenase multienzyme complex by protein engineering. Biochemistry-Us 32, 2737–2740.

Brakoulias, A., Jackson, R.M., 2004. Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: an automated all-against-all structural comparison using geometric matching. Proteins 56, 250–260.

Cahn, J.K.B., Werlang, C.A., Baumschlager, A., Brinkrnann-Chen, S., Mayo, S.L., Arnold, F.H., 2017. A general tool for engineering the NAD/NADP cofactor preference of oxidoreductases. ACS Synth. Biol. 6, 326–333.

Campbell, E., Chuang, S., Banta, S., 2013. Modular exchange of substrate-binding loops alters both substrate and cofactor specificity in a member of the aldo-keto reductase superfamily. Protein Eng. Des. Sel. 26, 181–186.

Carpenter, E.P., Hawkins, A.R., Frost, J.W., Brown, K.A., 1998. Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis. Nature 394, 299–302.

Carugo, O., Argos, P., 1997. NADP-dependent enzymes .1. Conserved stereochemistry of cofactor binding. Protein Struct. Funct. Genet. 28, 10–28.

Chandra, A., Tünnermann, L., Löfstedt, T., Gratz, R., 2023. Transformer-based deep learning for predicting protein properties in the life sciences. Elife 12.

Chánique, A.M., Parra, L.P., 2018. Protein engineering for nicotinamide coenzyme specificity in oxidoreductases: attempts and challenge. Front. Microbiol. 9.

Chen, J.R., Cheong, H.H., Siu, S.W.I., 2021. xDeep-AcPEP: deep learning method for anticancer peptide activity prediction based on convolutional neural network and multitask learning. J. Chem. Inf. Model. 61, 3789–3803.

Cheng, F., Wei, L., Wang, C.J., Liang, X.H., Xu, Y.Q., Xue, Y.P., Zheng, Y.G., 2023. Switching the cofactor preference of formate dehydrogenase to develop an NADPH-dependent biocatalytic system for synthesizing chiral amino acids. J. Agric. Food Chem. 71, 9009–9019.

Cui, D.B., Zhang, L.J., Jiang, S.Q., Yao, Z.Q., Gao, B., Lin, J.P., Yuan, Y.A., Wei, D.Z., 2015. A computational strategy for altering an enzyme in its cofactor preference to NAD(H) and/or NADP(H). FEBS J. 282, 2339–2351.

Dambe, T.R., Kühn, A.M., Brossette, T., Giffhorn, F., Scheidig, A.J., 2006. Crystal structure of NADP(H)-dependent 1,5-anhydro-D-fructose reductase from at 2.2 Å resolution: construction of a NADH-accepting mutant and its application in rare sugar synthesis. Biochemistry-Us 45, 10030–10042.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.

Geertz-Hansen, H.M., Blom, N., Feist, A.M., Brunak, S., Petersen, T.N., 2014. Cofactory: sequence-based prediction of cofactor specificity of Rossmann folds. Proteins 82, 1819–1828.

Goldford, J.E., George, A.B., Flamholz, A.I., Segre, D., 2022. Protein cost minimization promotes the emergence of coenzyme redundancy. Proc. Natl. Acad. Sci. U.S.A. 119, e2110787119.

Hao, Y.R., Dong, L., Wei, F.R., Xu, K., 2021. Self-attention attribution: interpreting information interactions inside transformer. Aaai Conf Artif Inte 35, 12963–12971.

Hsieh, J.Y., Liu, G.Y., Chang, G.G., Hung, H.C., 2006. Determinants of the dual cofactor specificity and substrate cooperativity of the human mitochondrial NAD(P)+-dependent malic enzyme: functional roles of glutamine 362. J. Biol. Chem. 281, 23237–23245.

Jia, Q., Zheng, Y.C., Li, H.P., Qian, X.L., Zhang, Z.J., Xu, J.H., 2022. Engineering isopropanol dehydrogenase for efficient regeneration of nicotinamide cofactors. Appl. Environ. Microbiol. 88.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W.Z., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240.

Kallberg, Y., Persson, B., 2006. Prediction of coenzyme specificity in dehydrogenases/reductases - a hidden Markov model-based method and its application on complete genomes. FEBS J. 273, 1177–1184.

Kaminski, K., Ludwiczak, J., Jasinski, M., Bukala, A., Madaj, R., Szczepaniak, K., Dunin-Horkawicz, S., 2022. Rossmann-toolbox: a deep learning-based protocol for the prediction and design of cofactor specificity in Rossmann fold proteins. Briefings Bioinf. 23.

Karim, M.R., Islam, T., Shajalal, M., Beyan, O., Lange, C., Cochez, M., Rebholz-Schuhmann, D., Decker, S., 2023. Explainable AI for bioinformatics: methods, tools and applications. Briefings Bioinf. 24.

Khoury, G.A., Fazelinia, H., Chin, J.W., Pantazes, R.J., Cirino, P.C., Maranas, C.D., 2009. Computational design of xylose reductase for altered cofactor specificity. Protein Sci. 18, 2125–2138.

Kim, G.B., Gao, Y., Palsson, B.O., Lee, S.Y., 2021. DeepTFactor: a deep learning-based tool for the prediction of transcription factors. P Natl Acad Sci USA 118.

Kim, G.B., Kim, J.Y., Lee, J.A., Norsigian, C.J., Palsson, B.O., Lee, S.Y., 2023. Functional annotation of enzyme-encoding genes using deep learning with transformer layers. Nat. Commun. 14.

King, Z.A., Feist, A.M., 2014. Optimal cofactor swapping can increase the theoretical yield for chemical production in and. Metab. Eng. 24, 117–128.

Kleiger, G., Eisenberg, D., 2002. GXXXG and GXXXA motifs stabilize FAD and NAD(P)-binding Rossmann folds through C -H•••O hydrogen bonds and van der Waals interactions. J. Mol. Biol. 323, 69–76.

Kovaleva, O., Romanov, A., Rogers, A., Rumshisky, A., 2019. Revealing the dark secrets of BERT. In: 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (Emnlp-Ijcnlp 2019), pp. 4365–4374.

Laurino, P., Tóth-Petróczy, A., Meana-Pañeda, R., Lin, W., Truhlar, D.G., Tawfik, D.S., 2016. An ancient fingerprint indicates the common ancestry of rossmann-fold enzymes utilizing different ribose-based cofactors. PLoS Biol. 14.

Lee, D.-H., Kim, H., Sung, B.-H., Cho, B.K., Lee, S.-G., 2023. Biofoundries: bridging automation and biomanufacturing in synthetic biology. Biotechnol. Bioproc. Eng.

Lin, Z.M., Akin, H., Rao, R.S., Hie, B., Zhu, Z.K., Lu, W.T., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., Costa, A.D., Fazel-Zarandi, M., Sercu, T., Candido, S., Rives, A., 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science 379, 1123–1130.

Ma, W., Geng, Q., Chen, C., Zheng, Y.C., Yu, H.L., Xu, J.H., 2023. Engineering a formate dehydrogenase for NADPH regeneration. Chembiochem, e202300390.

Madhavan, A., Arun, K.B., Binod, P., Sirohi, R., Tarafdar, A., Reshmy, R., Kumar Awasthi, M., Sindhu, R., 2021. Design of novel enzyme biocatalysts for industrial bioprocess: harnessing the power of protein engineering, high throughput screening and synthetic biology. Bioresour. Technol. 325, 124617.

Medvedev, K.E., Kinch, L.N., Schaeffer, R.D., Grishin, N.V., 2019. Functional analysis of Rossmann-like domains reveals convergent evolution of topology and reaction pathways. PLoS Comput. Biol. 15.

Medvedev, K.E., Kinch, L.N., Schaeffer, R.D., Pei, J.M., Grishin, N.V., 2021. A fifth of the protein world: rossmann-like proteins as an evolutionarily successful structural unit. J. Mol. Biol. 433.

Meng, H.K., Liu, P., Sun, H.B., Cai, Z., Zhou, J., Lin, J.P., Li, Y., 2016. Engineering a D-lactate dehydrogenase that can super-efficiently utilize NADPH and NADH as cofactors. Sci Rep-Uk 6.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021. Pfam: the protein families database in 2021. Nucleic Acids Res. 49, D412–D419.

Nagano, N., Orengo, C.A., Thornton, J.M., 2002. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J. Mol. Biol. 321, 741–765.

Nakanishi, M., Matsuura, K., Kaibe, H., Tanaka, N., Nonaka, T., Mitsui, Y., Hara, A., 1997. Switch of coenzyme specificity of mouse lung carbonyl reductase by substitution of threonine 38 with aspartic acid. J. Biol. Chem. 272, 2218–2222.

Naylor, C.E., Gover, S., Basak, A.K., Cosgrove, M.S., Levy, H.R., Adams, M.J., 2001. NADP and NAD binding to the dual coenzyme specific enzyme glucose 6-phosphate dehydrogenase:: different interdomain hinge angles are seen in different binary and ternary complexes. Acta Crystallogr. D 57, 635–648.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., Thornton, J.M., 1997. Cath - a hierarchic classification of protein domain structures. Structure 5, 1093–1108.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z.M., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J.J., Chintala, S., 2019.

PyTorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32 (Nips 2019), 32.
Pick, A., Ott, W., Howe, T., Schmid, J., Sieber, V., 2014. Improving the NADH-cofactor specificity of the highly active AdhZ3 and AdhZ2 from Escherichia coli K-12. J. Biotechnol. 189, 157–165.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why should I trust you?" Explaining the predictions of any classifier. In: Kdd'16: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144.
Rossmann, M.G., Moras, D., Olsen, K.W., 1974. Chemical and biological evolution of a nucleotide-binding protein. Nature 250, 194–199.
Russell, J.B., Cook, G.M., 1995. Energetics of bacterial-growth - balance of anabolic and catabolic reactions. Microbiol. Rev. 59, 48–62.
Scrutton, N.S., Berry, A., Perham, R.N., 1990. Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. Nature 343, 38–43.
Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: visual Explanations from deep networks via gradient-based localization. Ieee I Conf Comp Vis. 618–626.
Son, H.F., Yu, H., Hong, J., Lee, D., Kim, I.K., Kim, K.J., 2023. Structure-guided protein engineering of glyceraldehyde-3-phosphate dehydrogenase from corynebacterium glutamicum for dual NAD/NADP cofactor specificity. J. Agric. Food Chem. 71, 17852–17859.
Sugiki, S., Niide, T., Toya, Y., Shimizu, H., 2022. Logistic regression-guided identification of cofactor specificity-contributing residues in enzyme with sequence datasets partitioned by catalytic properties. ACS Synth. Biol. 11, 3973–3985.
Sundararajan, M., Taly, A., Yan, Q.Q., 2017. Axiomatic attribution for deep networks. Pr Mach Learn Res 70.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. Adv Neur In. 30.
Vidal, L.S., Kelly, C.L., Mordaka, P.M., Heap, J.T., 2018. Review of NAD(P)H-dependent oxidoreductases: properties, engineering and application. Bba-Proteins Proteom 1866, 327–347.
Wang, M., Chen, B.Q., Fang, Y.M., Tan, T.W., 2017. Cofactor engineering for more efficient production of chemicals and biofuels. Biotechnol. Adv. 35, 1032–1039.
Wang, Y.P., San, K.Y., Bennett, G.N., 2013. Cofactor engineering for advancing chemical biotechnology. Curr. Opin. Biotechnol. 24, 994–999.
Wiegreffe, S., Pinter, Y., 2019. Attention is not not Explanation. In: 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (Emnlp-Ijcnlp 2019), pp. 11–20.
Yang, T.H., Yu, Y.H., Wu, S.H., Zhang, F.Y., 2023. CFA: an explainable deep learning model for annotating the transcriptional roles of -regulatory modules based on epigenetic codes. Comput. Biol. Med. 152.
Zhang, C., Freddolino, P.L., Zhang, Y., 2017. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. Nucleic Acids Res. 45, W291–W299.
Zhou, Z., Yeung, W., Gravel, N., Salcedo, M., Soleymani, S., Li, S., Kannan, N., 2023. Phosformer: an explainable transformer model for protein kinase-specific phosphorylation predictions. Bioinformatics 39.