JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

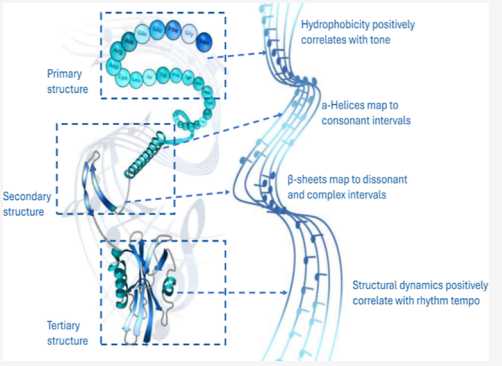
From Signal to Symphony: Exploring 2D Sequence Representations for Protein Function Prediction

Yiquan Wang, Minnuo Cai, Yuhua Dong, Yahui Ma, and Kai Wei*

Cite This: J. Chem. Inf. Model. 2025, 65, 12723−12736 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Predicting protein function from its primary sequence is a fundamental challenge in computational biology. While deep learning has excelled, the optimal representation of sequence data remains an open question. This study explores protein sonification—the conversion of amino acid sequences into 2D spectrograms—as a representation of this task. To facilitate this investigation, we developed a benchmark data set of 18,000 sequences spanning 12 functionally diverse protein classes. Our systematic evaluation suggests that the structural transformation from a 1D sequence to a 2D spectrogram may be a key contributor to the model's predictive performance. This observation is supported by ablation studies where models using either purely visual or acoustic features from the spectrogram demonstrated effective stand-alone performance, suggesting that the representation itself is a key source of this capability. For instance, a model using a sonification map without explicit biophysical meaning achieved 81.08% accuracy, while our biophysically informed model reached 84.00%, indicating that such domain knowledge may offer a modest performance benefit. When trained from scratch on our data set, our fusion model achieved performance comparable to or slightly exceeding that of standard transformer architectures like ESM-2 and ProtBERT, suggesting its potential for data efficiency in this specific context. The model's potential for generalizability was further supported by its performance on the external CARE enzyme classification benchmark, where it achieved 90.44% accuracy. Finally, as a proof-of-concept, we explore the utility of our encoding to guide a diffusion model in generating novel green fluorescent protein variants, which were assessed for structural viability using computational methods. Our work provides evidence suggesting that the utility of sonification in this context may stem largely from its representational structure, offering a perspective on feature engineering for biological sequences.



INTRODUCTION

Understanding the complex relationship between protein sequence, structure, and function is a paramount challenge in the biological sciences.[1−4] While dominant computational methods, such as protein language models (PLMs) and structure-based predictors, have achieved remarkable success, they each have inherent limitations.[5−7] PLMs, which process sequences linearly, can struggle to capture the long-range, quasi−periodic correlations that define global protein architecture.[8−10] This creates an opportunity for novel data representations that can efficiently extract holistic, functional information from primary sequences.

This study investigates protein sonification, the translation of sequence data into a rich, two-dimensional spectrogram.[11−13] Our approach leverages a powerful and broadly validated computational strategy: transforming 1D biological sequences into 2D representations to capture long-range interactions (LRIs) that are otherwise missed. This paradigm has become a pillar of modern bioinformatics, driving breakthroughs in fields as diverse as protein structure prediction with AlphaFold[14,15] 3D genome folding,[16,17] and RNA secondary structure analysis.[18] Motivated by these successes, we employ sonification to "fold" a 1D sequence into a 2D spectrogram. In this format, sequence-distant residues are brought into spatial proximity, enabling standard convolutional networks to model global functional signatures in a computationally efficient manner.

However, a critical question arises: if such a method is effective, then what is the source of its power? Is it the specific, biophysically inspired rules used for the translation, which could be viewed as speculative, or is it the fundamental act of transforming a 1D sequence into a 2D representation? To address this question directly, we moved beyond heuristic mappings to establish a quantitative sonification framework based on first-principles of amino acid physicochemical properties. We began by rigorously validating the core premise of this framework—the link between static sequence properties and true protein dynamics—with extensive molecular dynam-

Received: July 28, 2025
Revised: October 11, 2025
Accepted: October 24, 2025
Published: November 17, 2025

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736

ics (MD) simulations across diverse protein-fold families. Then, we conducted a series of controlled experiments on a broad and functionally diverse data set, comparing our theory-driven model against alternatives with inverted or randomized semantics. This investigation points toward the importance of the 1D-to-2D data representation itself, a central theme of this work. Our results suggest that functional patterns might be learned as emergent properties from this structured representation.

RELATED WORKS

Computational Approaches to Protein Function Prediction. The analysis of protein structure and function has long been propelled by experimental techniques such as X-ray crystallography and Cryo-EM, which provide high-resolution but often static snapshots of molecular machinery.[19,20] To overcome the throughput and dynamic limitations of these methods, computational approaches have become indispensable.[21] In recent years, deep learning has revolutionized the field, most notably with DeepMind's AlphaFold, which has achieved unprecedented accuracy in predicting protein structures from sequence.[7,22,23] These advances have driven new biological insights by providing reliable structural models for millions of proteins.[24]

However, a static structure does not equate to a complete function. Even state-of-the-art models face challenges in capturing the dynamics of flexible regions, modeling complex multiprotein interactions, and fully elucidating the functional context from structure alone.[25,26] This highlights a critical gap: there is an urgent need for novel data representations that can transcend 1D sequences and 3D structures to more comprehensively capture functional information. While models operating on 1D sequences (like PLMs[27]) or 3D structures[7,15,28] have been highly successful, there remains a need for novel data representations that can capture the hierarchical and quasi-structural information implicitly encoded in a protein's primary sequence in a more holistic manner.[29]

Sonification and Audio-Based Methods in Bioinformatics. One emerging avenue for developing such representations is to draw an analogy between the hierarchical complexity of proteins and music, whose mathematical and structural properties offer a rich framework for encoding biological data.[30,31] This concept was established in early pioneering works like "Protein Music",[32] which translated sequences into melody and bass lines for data analysis. The approach has since been adapted for diverse goals, including enhancing accessibility for the visually impaired by creating classical music from sequences[33,34] and assisting in complex tasks like protein alignment through combined auditory and visual representations.[35]

Building on this foundation, recent work has employed more sophisticated deep learning methods. A significant advancement is Buehler's "AttentionCrossTranslation" model, which introduced a powerful framework for unsupervised, bidirectional translation between musical and protein domains using interacting transformer networks.[12] A key strength of their model is its ability to perform bidirectional and cycle-consistent translations, ensuring high fidelity. This work established a powerful, generalizable methodology for finding hidden relationships between data types, moving the field toward fully automated pattern discovery. Concurrently, the "conversion of music to protein" (CoMtP) concept further demonstrated that musical scores can serve as templates for designing new functional peptides.[36]

While the unsupervised discovery of translation rules, as pioneered by Buehler et al., is a powerful approach for fundamental pattern analysis, our research addresses a different, complementary scientific goal. Our work is not aimed at discovering a universal mapping itself but rather at testing a specific, human-driven hypothesis: can a biophysically grounded, quantitative mapping serve as an effective feature engineering strategy for a downstream machine learning task, namely, protein function prediction? Therefore, our philosophy is distinct. We intentionally leverage existing scientific knowledge (i.e., physicochemical properties) to construct transparent and interpretable quantitative mapping. The primary goal is not translation fidelity for its own sake but the creation of a feature-rich 2D representation (the spectrogram) that is optimized for analysis by computer vision models. The effectiveness of this knowledge-driven feature engineering approach is evaluated by the accuracy achieved on the predictive task.

The Resonant Recognition Model: A Precedent for Signal-Based Functional Analysis. A significant precedent for treating biological sequences as signals to decode functions is the Resonant Recognition Model (RRM). This physico-mathematical framework posits that biomolecular interactions are governed by resonant energy transfer.[37,38] The core methodology converts an amino acid sequence into a numerical signal—often using the electron−ion interaction potential—and then applies Fourier analysis. This approach builds directly on foundational work that first demonstrated how periodicities in physicochemical properties, such as the "hydrophobic moment", strongly correlate with secondary structures like α-helices and β-sheets.[39] The RRM posits that a single, dominant peak in the resulting frequency spectrum—the "characteristic frequency"—is strongly correlated with the protein's specific biological function or interaction.[40] The practical utility of this model is well-established; it has been successfully used to design novel bioactive peptides[41] and, more recently, to explain complex phenomena like long-distance DNA−protein interactions.[42]

Furthermore, the RRM framework has proven to be both generalizable and extensible. Its principles have been successfully adapted for the classification of other macromolecules, serving as an effective preprocessing technique for distinguishing between different classes of RNA sequences.[43] Recognizing that a single power spectrum might not capture all available information, researchers developed the Complex RRM (CRRM). This extension utilizes both the real and imaginary components of the Fourier spectrum, providing a richer, two-dimensional view of the signal's frequency characteristics. The CRRM demonstrated superior performance in distinguishing subtle biological differences between closely related viral protein families, such as subtypes of Influenza A neuraminidase, where the traditional RRM was insufficient.[44]

Collectively, the extensive literature on RRM and its derivatives provides a powerful theoretical and empirical foundation for our work. It establishes that transforming the linear information on a biological sequence into the frequency domain is a biophysically grounded and highly effective strategy for functional analysis. The evolution from RRM to CRRM also highlights a clear trajectory: increasing the dimensionality of the signal representation can unlock deeper

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736

biological insights. Our sonification approach is conceptually aligned with this trajectory but takes a substantial leap forward by synthesizing multiple physicochemical properties into a complex, information-dense, two-dimensional spectrogram.

METHODS

Data Set Curation and Preprocessing. To construct a benchmark for evaluating protein function prediction that represents a broad spectrum of biological roles, we curated a data set of 18,000 sequences from 12 distinct protein classes: Enzyme, Structural, Transport, Storage, Signaling, Receptor, Regulatory, Immune, Chaperone, Cell Adhesion, Motor, and Antimicrobial. Protein sequences were sourced from several authoritative public databases to ensure diversity and quality. These included the NCBI RefSeq collection,[45] the UniProt Knowledgebase (UniProtKB),[46] and specialized databases such as the Antimicrobial Peptide Database,[47] the Transporter Classification Database (TCDB)[48] for transport proteins, and the CARE database[49] for enzymes. To ensure data quality and reduce redundancy, we employed the CD-HIT clustering algorithm,[50] removing all sequences with an identity of over 90% to any other sequence in the data set. The final data set was balanced, containing 1500 sequences for each of the 12 classes. For model training and evaluation, we partitioned the data into a training set (70%, 12,600 sequences), a validation set (10%, 1800 sequences), and an independent test set (20%, 3600 sequences). This stratified split ensures that each data partition maintains the original class distribution.[51]

Quantitative, Biophysically Grounded Sonification Framework. To move beyond heuristic rules and establish a rigorous foundation, we developed a quantitative framework that maps fundamental amino acid physicochemical properties to musical elements. The core of our framework is the translation of the primary sequence into a musical score, which is then rendered as a 2D spectrogram for analysis. Our central scientific question was whether the model's performance stems from the specific biophysical semantics encoded in the mapping rules or from the structural transformation of the 1D sequence into a 2D representation. To answer this, we implemented and compared three distinct mapping schemes.

Theory-Driven Quantitative Model. This model serves as our principled, theory-driven starting point. Our focus on hydrophobicity is motivated by the foundational discovery that periodicities in this property along the primary sequence are strong determinants of secondary structure,[39] with further computational studies confirming that this global patterning of hydrophobic and polar residues is a dominant force in determining the overall protein fold, often overriding the intrinsic propensities of individual amino acids[52] Building on this, our model is based on the biophysical hypothesis that stable, hydrophobic residues in the protein core correspond to low-frequency collective motions. To implement this, we map higher hydrophobicity to lower pitches, a design inspired by psychoacoustic analogies linking low frequencies to stability.[53] The validity of this choice is empirically demonstrated in our results (Table 2), where this mapping provides a clear performance advantage over the control models. The pitch of each note is determined by the hydrophobicity of the corresponding amino acid, following the formula:

$$\text{note} = \text{round}(\text{base\_note} - (H \times S))$$

where note is the resulting MIDI pitch number, base_note is a reference pitch set to 60 (Middle C), H is the Kyte-Doolittle hydrophobicity index of the amino acid, and S is a scaling factor set to 3. This formula directly implements our hypothesis that a higher hydrophobicity corresponds to lower, more stable-sounding pitches. Timbre, or the tonal quality, is determined by amino acid polarity based on the Zimmerman scale, with polar residues mapped to "bright" timbres (e.g., piano) and nonpolar residues to "dark" timbres (e.g., cello).

Control Models. To deconstruct the source of performance, we designed two control models:

- Inverted Semantics Model: As a direct control for the directionality of the biophysical analogy, we inverted the core mapping rule such that higher hydrophobicity was mapped to higher musical pitches (note = round-(base_note + ($H \times S$))), while all other rules remained identical.
- Semantic Ablation Model: To isolate the effect of the 2D representation structure itself, we created a model with a fixed, but random, mapping between the 20 amino acids and a set of 20 distinct pitches. This removes any biophysical meaning from the pitch assignment while preserving a consistent 1D-to-2D transformation.

Biophysical Validation of the Sequence-to-Dynamics Mapping Principle. To establish a rigorous biophysical foundation for our entire sonification framework, we performed a critical computational experiment. The objective was to validate our core premise: that it is valid and meaningful to map static sequence information onto a dynamic temporal medium like music. We hypothesized that if our encoding philosophy is sound, then the static features we use as proxies for dynamics (like B-factors for flexibility) must correlate with "ground truth" physical dynamics observed in simulation.

MD Simulations. To assess the generalizability of our findings, we performed MD simulations on four proteins representing highly diverse structural fold families: T4 Lysozyme (PDB ID: 2LZM, an α/β protein), a de novo designed TIM-barrel (PDB ID: 5BVL, an α/β protein), a Villin Headpiece (PDB ID: 2F4K, an all-α protein), and green fluorescent protein (GFP) (PDB ID: 1EMA, an all-β barrel

Table 1. Correlation between Experimental B-Factors and MD-Derived RMSF across Diverse Protein Folds

PDB ID	Protein-fold family	Pearson's r	P-value
2F4K	all-α (Villin headpiece)	0.9030	$6.58 \times 10{-}13$
5BVL	α/β (TIM-Barrel)	0.6768	$1.85 \times 10{-}25$
2LZM	α/β (T4 Lysozyme)	0.5478	$3.15 \times 10{-}14$
1EMA	all-β (GFP)	0.3653	$4.01 \times 10{-}8$

Table 2. Performance Comparison of Different Sonification Mapping Schemes on the 12-Class Dataset[a]

mapping scheme	core principle	accuracy (%)	precision	recall	F1-score
theory-driven	biophysical semantics	84.00	0.84	0.84	0.84
inverted semantics	inverted biophysics	82.69	0.83	0.83	0.83
semantic ablation (random)	structure-only	81.08	0.81	0.81	0.81

[a]All models use the identical end-to-end fusion architecture and were evaluated on the same independent test set to deconstruct the source of performance.

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736

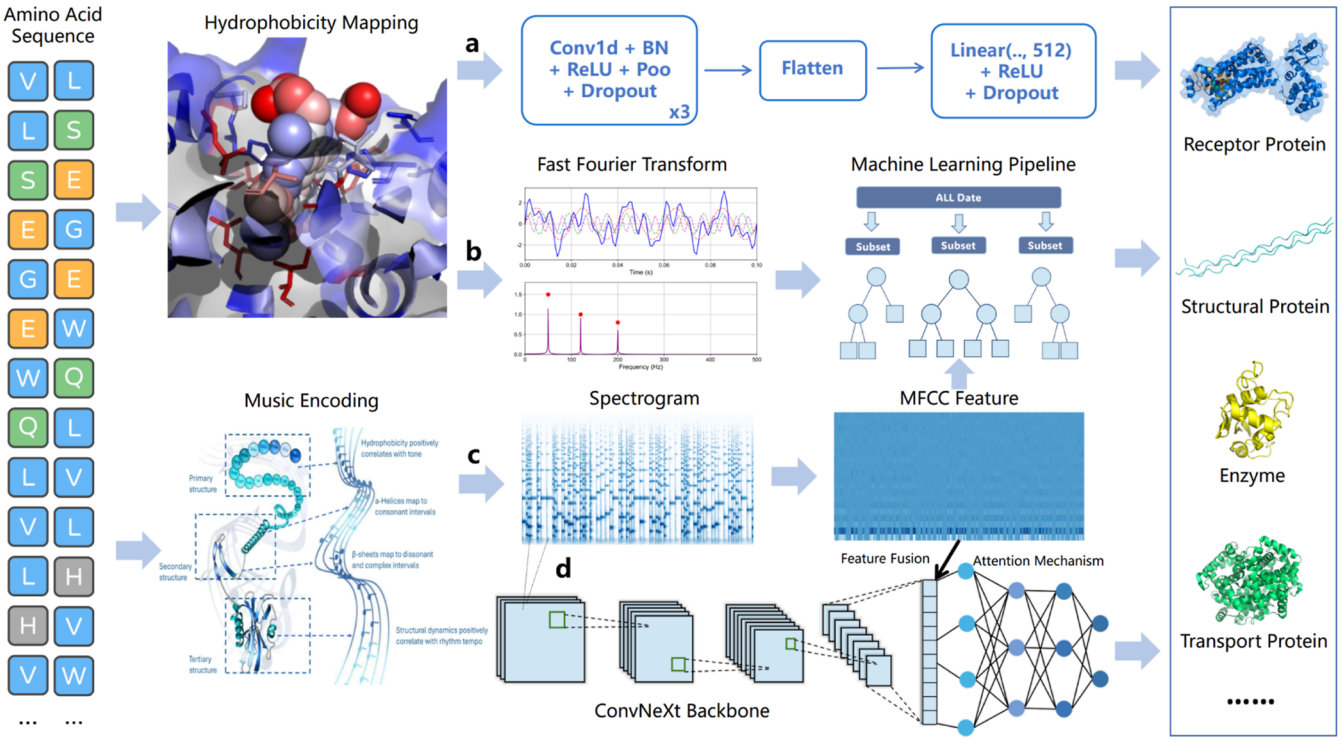


Figure 1. Our analytical pipeline is designed as a systematic comparative study to demonstrate the progressive benefits of evolving both data representation and model complexity. The workflow encompasses four distinct models: (a) 1D CNN on 1D signal: a deep learning model that learns features directly from the raw 1D physicochemical signal. (b) Classical ML on 1D signal: a baseline model that relies on spectral features extracted via FFT from the 1D signal. (c) Classical ML on 2D spectrogram: a model that uses engineered audio features Mel-frequency cepstral coefficients (MFCC) from spectrograms. (d) End-to-End fusion model on 2D spectrogram: our final model, which employs a CNN and an attention mechanism to automatically learn and fuse features from the 2D spectrograms.

protein). All simulations were conducted using the GROMACS package with the AMBER99SB-ILDN force field and the TIP3P water model. Following standard system preparation and equilibration, a 100 ns production MD simulation was carried out for each system in the *NPT* ensemble. The stability of each trajectory was confirmed by the convergence of the backbone root-mean-square deviation (RMSD). A detailed simulation protocol is provided in the Supporting Information.

Data Analysis. For each trajectory, the root-mean-square fluctuation (RMSF) of each residue's Cα atom was calculated from the full 100 ns production run using the `gmx rmsf` tool, serving as the ground-truth measure of dynamic flexibility. The experimental B-factors were extracted from the corresponding PDB files. Given that these two metrics are theoretically related,[54] we quantified their linear relationship by calculating the Pearson correlation coefficient (r) and its corresponding p-value. This direct comparison validates the link between the static crystallographic data and the dynamic behavior.

Comparative Models for Protein Function Classification. Our analytical pipeline was designed to rigorously investigate the interplay between the nature of the input signal (1D vs 2D) and the model's learning paradigm. All sonification-based models were trained using data generated by the three mapping schemes described previously, with the performance in comparative tables corresponding to the best-performing theory-driven model unless stated otherwise.

Baseline Models on 1D Physicochemical Signals. We first established performance baselines by using the primary 1D physicochemical sequence. Each protein sequence was converted to a numerical signal by mapping amino acids to their Kyte-Doolittle hydrophobicity values.

Classical ML with Spectral Features. We applied a Fast Fourier Transform (FFT) to each 1D signal to extract spectral features, which were then used to train an ensemble machine learning model.

1D Convolutional Neural Network. We also applied a 1D convolutional neural network (CNN) directly to the 1D signals to establish a modern deep learning baseline.

Sonification-Based Models on 2D Spectrograms. Next, we advanced the data representation by translating protein sequences into 2D spectrograms via sonification.

Classical ML with Acoustic Features. To isolate the impact of the 2D representation, we extracted Mel-Frequency-Cepstral Coefficients (MFCCs) from each spectrogram and fed them into the same classical ML pipeline.

Model Architecture and Ablation Study. Finally, we combined the advanced 2D representation with a sophisticated, end-to-end deep learning model. The final model features a dual-branch architecture, fusing features from a Visual Branch (a ConvNeXt-Tiny backbone processing the spectrogram image[55]) and an Acoustic Branch (a GRU with attention processing the MFCC sequence). The concatenated features are passed to a final classification head.

To deconstruct the contributions of each modality and investigate the source of the model's performance, we designed two additional model configurations for an ablation study. The Visual-Only Model consists solely of the ConvNeXt-Tiny backbone, processing the complete spectrogram as an image. Conversely, the Acoustic-Only Model utilizes only the GRU with an attention mechanism, taking the MFCC sequence derived from the spectrogram as its input. These configurations allow for a direct comparison of the predictive power inherent

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736

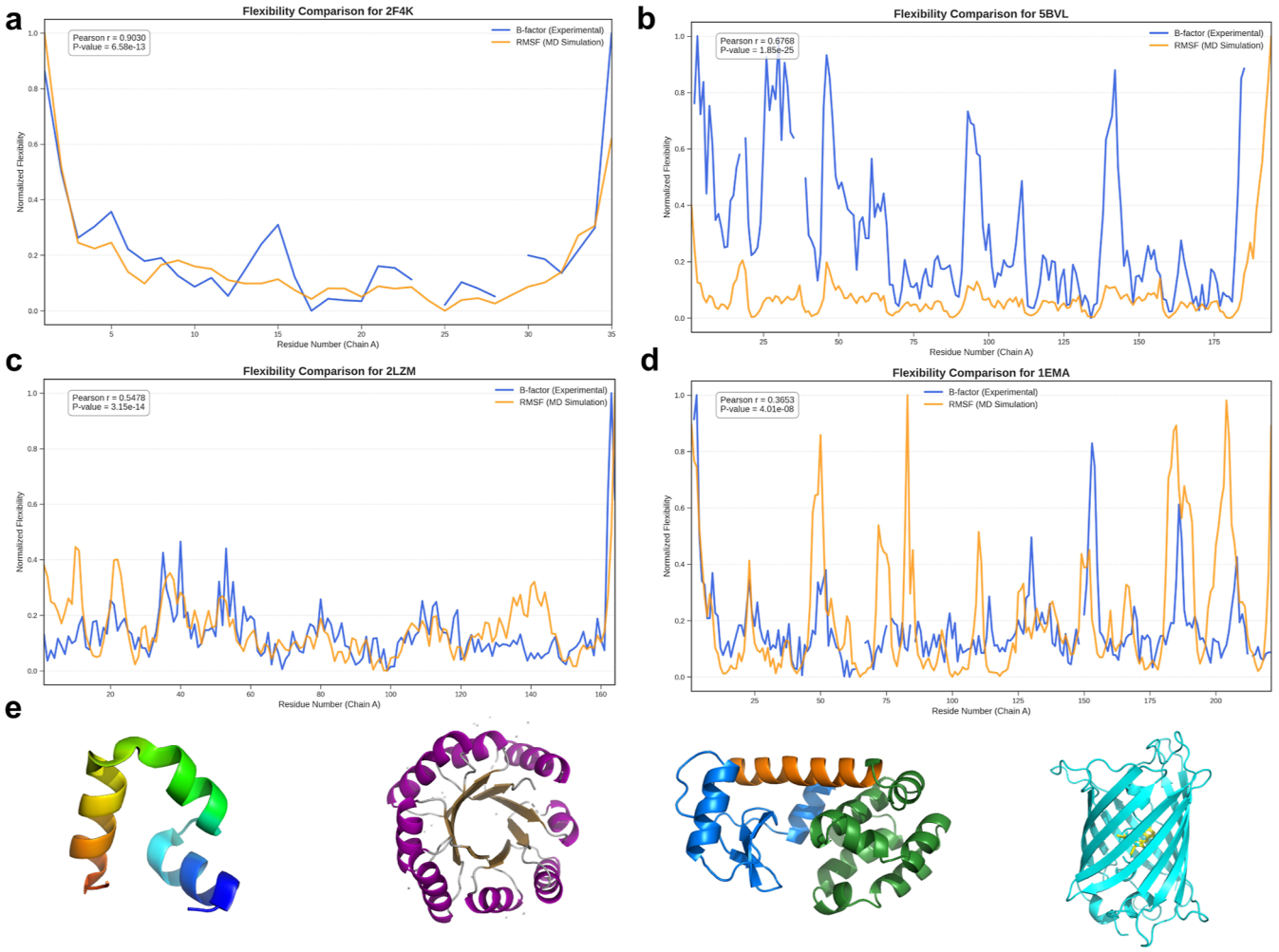


Figure 2. Assessing B-factors as a proxy for dynamic flexibility across diverse protein-fold families. Comparison of experimental flexibility (crystallographic B-factors, blue) and simulated dynamic flexibility (RMSF from 100 ns MD simulations, orange) for four proteins representing distinct structural classes. The plots show statistically significant positive correlations for (a) the all-α Villin Headpiece (2F4K, $r = 0.903$), (b) the α/β TIM-Barrel (5BVL, $r = 0.677$), (c) the α/β T4 Lysozyme (2LZM, $r = 0.548$), and (d) the all-β GFP (1EMA, $r = 0.365$). All correlations are highly significant ($p < 1 \times 10^{-7}$). This analysis provides biophysical support for the principle of mapping static structural properties to dynamic features in our sonification framework. (e) Cartoon representations of the protein structures analyzed in (A–D), shown in corresponding order from left to right.

in the visual versus the acoustic interpretation of the sonified data.

Benchmark Models for Comparison. To situate our model's performance, we benchmarked it against prominent PLMs and classical homology-based methods.

PLMs (ESM-2 and ProtBERT). We evaluated two prominent PLMs: ESM-2 (8M[56] and 35M[57] parameter versions)[5] and ProtBERT.[6] Both were tested under two conditions: a fine-tuning setting using official pretrained weights, and a from-scratch training setting using only our data set, which provides a direct comparison of data efficiency and the inherent capabilities of the architectures.

Homology-Based Search (BLAST + kNN). As a classical baseline, we implemented two BLAST-based k-Nearest Neighbors (kNN) classifiers.[58,59] The primary benchmark searches against the manually curated Swiss-Prot database to simulate a real-world annotation scenario. A secondary baseline searches against our own training set, serving as a measure of performance based on sequence similarity within our specific data distribution. We report the more challenging Swiss-Prot results in the main text.

Generative Design of GFP Variants. To demonstrate the practical utility of our encoding, we integrated it into a conditional diffusion model to guide the de novo design of GFP variants. The model was trained on experimental fluorescence data[60] and employed a multiobjective selection strategy that balanced predicted fitness with a harmonic score derived from our sonification framework, aiming to generate novel sequences that were both high-functioning and structurally viable.

RESULTS

Biophysical Foundation of the Sonification Framework. Before evaluating classification models, we first sought to validate the core premise of our quantitative sonification framework: mapping static sequence properties to a dynamic medium is a biophysically meaningful approach. To address the potential limitations of a single-protein validation, we extended our analysis to four proteins representing diverse fold families (all-α, all-β, and two distinct α/β topologies), each simulated for a comprehensive 100 ns (Figure 1).

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736

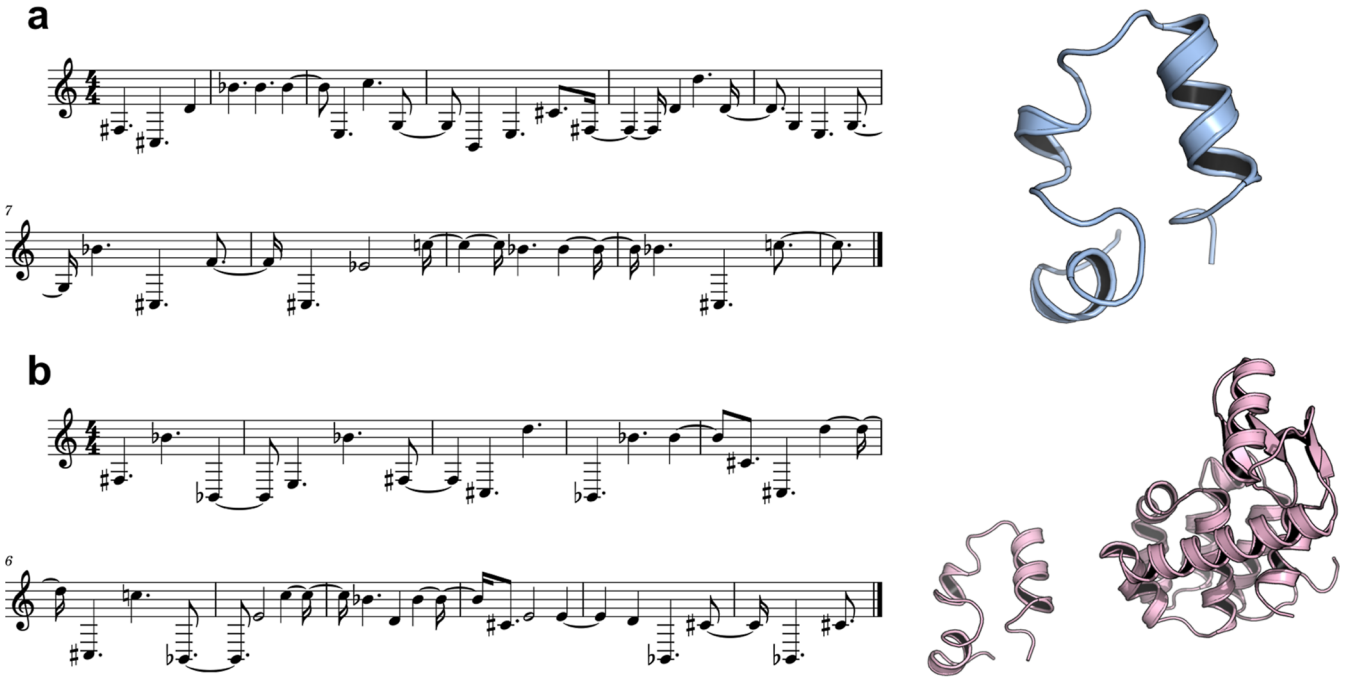


Figure 3. Visualization of protein-to-music translation as sheet music. Musical scores were generated from the first 30 amino acids of two representative proteins. (a) Villin Headpiece (1VII): this structurally stable protein translates into a score with a constrained melodic range and regular rhythm. (b) T4 Lysozyme (2LZM): this complex protein produces a score with greater melodic leaps and more intricate rhythms. These visual differences are quantitatively supported: the T4 Lysozyme fragment exhibits a 35.1% higher melodic complexity score and a 7.4% higher rhythmic complexity score compared to the Villin fragment. This is consistent with the greater diversity in its physicochemical properties.

Table 3. Comparative Performance of Classification Models on the Independent Test Set, Demonstrating the Impact of Evolving Data Representation (1D vs 2D) and Model Paradigm

data	model	accuracy (%)	precision	recall	F1-score
1D signal	FFT + classical ML	45.86	0.45	0.46	0.43
	1D CNN (deep learning)	61.31	0.61	0.61	0.60
2D spectrogram	MFCC + classical ML	69.47	0.69	0.69	0.69
	fusion model (deep learning)	84.00	0.84	0.84	0.84

Our analysis revealed statistically significant positive correlations between the experimental B-factors and the simulation-derived RMSF values for all four proteins, as detailed in Table 1. Notably, the all-α Villin Headpiece (2F4K) showed a strong correlation (r = 0.9030), while even more complex proteins such as the TIM-barrel (5BVL) and T4 Lysozyme (2LZM) demonstrated significant correlations. These results provide quantitative evidence supporting the use of static crystallographic data as a proxy for dynamic flexibility across different protein architectures. This supports the premise that the fundamental principle of our encoding is grounded in the physical reality of protein dynamics, strengthening the biophysical rationale for our methodology. The correspondence is visualized in Figure 2.

Visual and Quantitative Interpretation of the Protein Score. To provide an intuitive illustration of our encoding framework, we translated protein sequences not only into spectrograms for machine analysis but also into conventional musical scores for human interpretation. This allows for a direct visual inspection of how a protein's physicochemical properties are mapped onto musical language. To illustrate this, we selected two proteins with contrasting architectures: the small, highly stable, all-α-helical Villin Headpiece (1VII) and the larger, structurally complex T4 Lysozyme (2LZM), which contains a mix of helices, sheets, and flexible loops.

As shown in Figure 3, the resulting scores are visually and quantitatively distinct, reflecting their underlying biochemical differences. The score for the stable Villin Headpiece (Figure 3a) exhibits a higher degree of regularity. Its melody, determined by amino acid hydrophobicity, shows a less dramatic fluctuation, and its rhythm, governed by molecular weight, is more uniform. In contrast, the score for the complex T4 Lysozyme (Figure 3b) is noticeably more varied, featuring greater melodic leaps and more intricate rhythmic patterns.

To formalize this visual observation, we calculated complexity metrics for the first 30 amino acid fragment. The 'melodic complexity' was quantified as the average absolute difference in hydrophobicity between adjacent residues, while 'rhythmic complexity' was defined as the standard deviation of molecular weights. The analysis indicates that the T4 Lysozyme fragment is more complex, possessing a 35.1% higher melodic complexity score (4.58 vs 3.39) and a 7.4% higher rhythmic complexity score (29.70 vs 27.65) than the Villin fragment. This visual and quantitative comparison suggests that our framework can translate core structural and chemical features into distinct, interpretable musical patterns.

Beyond Semantics: The Role of Representational Structure. To ascertain whether the predictive power of sonification stems from the specific semantic encoding rules or the underlying structural transformation, our investigation was

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736

Table 4. Ablation Study of the Fusion Model on the Independent Test Set[a]

model configuration	input features	architecture	accuracy (%)	F1-score (macro)
fusion model (full)	spectrogram + MFCC	ConvNeXt + GRU	84.00	0.84
visual-only	full spectrogram	ConvNeXt	76.56	0.77
acoustic-only	MFCC	GRU + attention	77.17	0.77

[a]Performance of the full model is compared against its individual visual and acoustic branches.

Table 5. Comparison of Performance and Computational Complexity against State-of-the-Art Models[a]

model	accuracy (%)	parameters (M)	GFLOPs	inference latency (ms/seq)	training paradigm
ours (fusion model)	84.00	28.57	4.45	3.57	from scratch (12.6k sequences)
pretrained language models					
ProtBERT	93.14	378.95	773.61	56.73	pretrained (BFD)
ESM-2 (35M)	91.78	33.49	68.05	12.16	pretrained (UR50)
ESM-2 (8M)	90.53	7.50	15.13	5.58	pretrained (UR50)
language models trained from scratch					
ESM-2 (35M)	79.64	33.49	68.05	12.16	from scratch (12.6k sequences)
ESM-2 (8M)	78.53	7.50	15.13	5.58	from scratch (12.6k sequences)
ProtBERT	76.61	378.95	773.61	56.73	from scratch (12.6k sequences)
homology-based methods					
BLAST + kNN (k = 5, Swiss-Prot)	87.56 (on classified)	N/A	N/A	~338	N/A (search on 0.57 M sequences)
BLAST + kNN (k = 5, internal)	83.47	N/A	N/A	~80	N/A (search on 14.4k sequences)

[a]Latency for our model and PLMs was measured as the average of 100 runs on a single NVIDIA 4090 24GB GPU. The BLAST + kNN latencies were measured on a system with a 20 vCPU Intel Xeon Platinum 8470Q processor and are reported per query sequence. GFLOPs for PLMs are calculated for a sequence length of 1024. "From Scratch" models were trained on our data set (12,600 sequences) without any pretraining. Pretrained models utilize weights trained on large external data sets like BFD or UR50. BLAST accuracy for the Swiss-Prot search is reported only on the 61.86% of test sequences for which homologues could be found.

conducted on the comprehensive 12-class, 18,000-sequence data set. The results, summarized in Table 2, provide quantitative evidence suggesting that the structural transformation from a 1D sequence to a 2D spectrogram is a major contributor to the observed performance improvement.

Our theory-driven quantitative model, which incorporates biophysical information, achieved an accuracy of 84.00%. However, the most striking finding is that the control models with inverted and random semantics still achieved remarkably high accuracies of 82.69% and 81.08%, respectively. All three 2D sonification-based models substantially outperformed our strongest 1D baseline (61.31%, see Table 3). This result indicates that the vast majority of the predictive power is unlocked by the 2D representation itself, which enables the deep learning model to learn long-range dependencies and complex textural patterns. While the structural transformation is primary, biophysical semantics provide a consistent and measurable performance advantage (a nearly 3-point gain from the random to the theory-driven model), suggesting they offer a valuable inductive bias rather than being an absolute requirement.

The Synergy of Representation and Architecture. Building on the finding that 2D sonified representations are effective, we conducted a broader comparative analysis to understand the synergy between data representation and model architecture. The results, summarized in Table 3, reveal a clear trend for our 12-class benchmark.

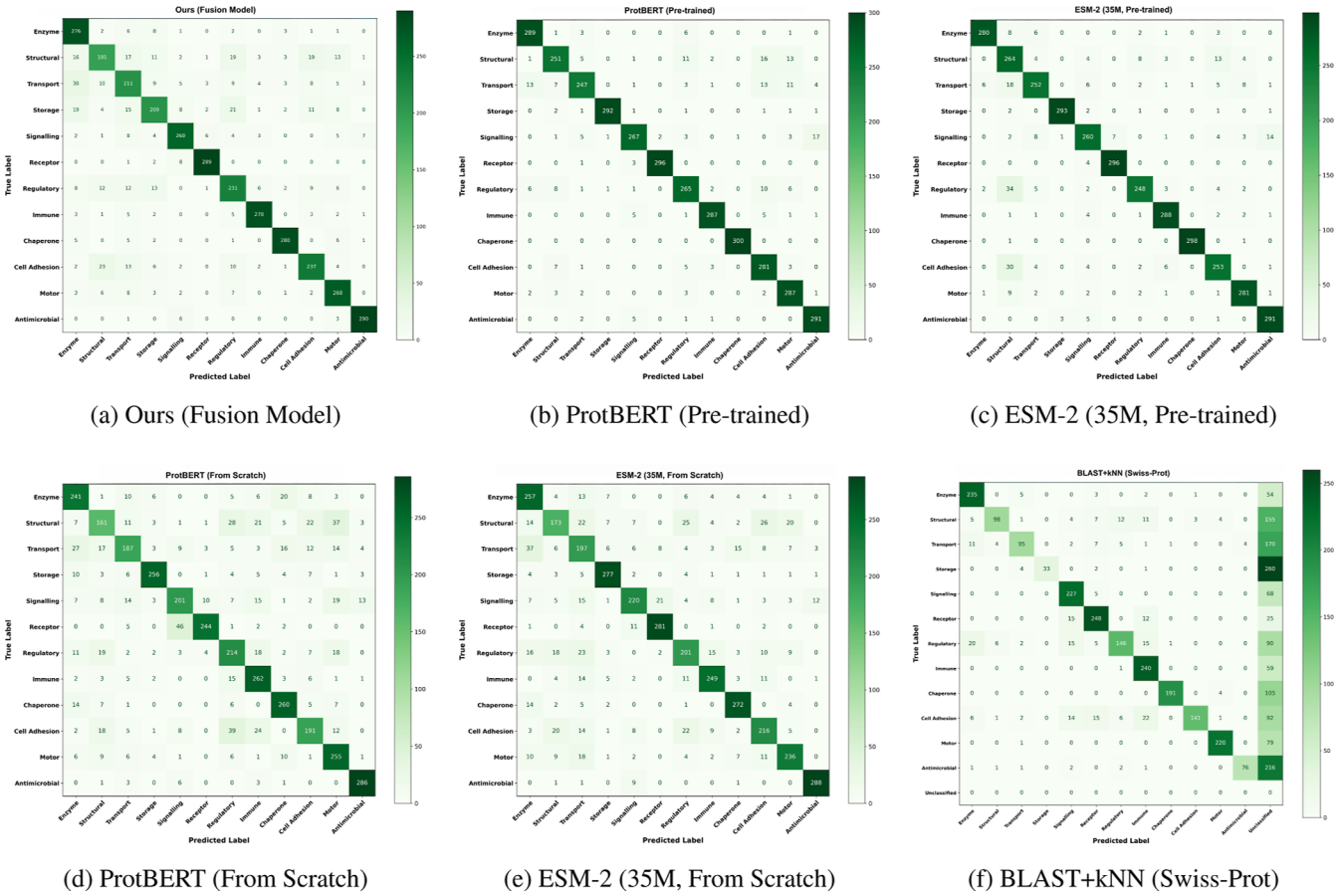
Our investigation began with the 1D physicochemical signal. The classical approach (FFT plus ML) achieved an accuracy of 45.86%. Applying a modern 1D CNN improved the performance to 61.31%. We then advanced the data representation to 2D spectrograms using theory-driven mapping. A classical ML pipeline with engineered MFCC features achieved 69.47% accuracy, a significant improvement over both 1D methods, confirming the intrinsic value of the 2D representation. The highest performance was achieved by pairing the advanced representation with an advanced model: our end-to-end fusion model reached a final accuracy of 84.00%. This progression suggests that high performance is unlocked when an enriched 2D representation is paired with a deep learning architecture capable of exploiting its complexity.

Ablation Study: Deconstructing the Source of Predictive Power. To further investigate the individual contributions of the visual and acoustic feature streams, we conducted an ablation study. As shown in Table 4, our full fusion model achieved the highest accuracy at 84.00%. Notably, the single-branch models also demonstrated substantial predictive capability. The Visual-Only model, relying on the ConvNeXt backbone to process the raw spectrogram, achieved 76.56% accuracy, while the Acoustic-Only model, using the GRU to process extracted MFCC features, reached 77.17% accuracy.

The strong performance of both unimodal configurations suggests that the 2D spectrogram generated through sonification is an information-rich representation. It indicates that functionally relevant patterns are encoded in a manner that is accessible to both direct visual analysis (as an image) and acoustic feature extraction (as a temporal signal). This finding implies that the effectiveness of our approach is not solely dependent on the final fusion architecture but is fundamentally rooted in the descriptive power of the 2D representation itself.

Comparative Benchmarking against Established Methods. To situate our model's performance in a broader context, we benchmarked our final fusion model against established PLMs (ProtBERT and ESM-2) and homology-based search methods. The results are detailed in Table 5, and visualized in Figure 4.

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736



(a) Ours (Fusion Model) (b) ProtBERT (Pre-trained) (c) ESM-2 (35M, Pre-trained)

(d) ProtBERT (From Scratch) (e) ESM-2 (35M, From Scratch) (f) BLAST+kNN (Swiss-Prot)

Figure 4. Comparative confusion matrices on the independent 12-class test set. The top row compares our model (a) against high-performing pretrained models: ProtBERT (b) and ESM-2 (c). The bottom row directly compares models trained from scratch—ProtBERT (d) and ESM-2 (e)—and the homology-based baseline (f), highlighting our model's data efficiency.

As expected, large pretrained models achieved the highest overall accuracies, with ProtBERT reaching 93.14% and ESM-2 reaching up to 91.78%. This outcome is attributable to their extensive pretraining on massive external data sets (e.g., BFD[61] and UR50[62]), which provides them with extensive prior knowledge derived from evolutionary and structural patterns.

However, a more direct comparison of model architecture and representational efficiency emerges when this pretraining advantage is removed. When trained from scratch on our 12,600 sequence training set, our fusion model (84.00%) achieved a higher accuracy than the transformer-based architectures tested. It surpassed the ESM-2 models (78.53–79.64%) and, notably, also the much larger ProtBERT model (76.61%). This finding suggests that, for the task and data set presented here, our sonification process may offer a data-efficient representation, allowing a vision-based model to learn relevant patterns effectively under data-limited conditions. This indicates that under these data-limited conditions, a vision-based model using our spectrogram representation was able to learn relevant functional patterns more effectively than the tested transformer architectures could from raw 1D sequences.

The homology-based BLAST + kNN approach against the Swiss-Prot database achieved a high accuracy of 87.56%; however, this was only on 61.86% of test sequences for which it could find reliable homologues, leaving the remaining 38% unclassified. As an internal control, applying the same method against our own training set yielded an accuracy of 83.47% across all test sequences. This comparison indicates that while our model holds a slight performance edge over the internal BLAST, a significant portion of the classification task within this data set can be solved by sequence similarity alone. Thus, our model offers a viable alternative that does not rely on finding close homologues in a reference database, suggesting its potential utility for sequences that lack clear evolutionary relatives.

Validation on an External Enzyme Classification Benchmark. To evaluate the generalizability of our framework, we tested our approach on a larger data set constructed from the external CARE enzyme classification benchmark.[49] For this validation, we retrained our fusion model from scratch on a new data set of 35,000 sequences. The model, still employing our theory-driven mapping, achieved a competitive overall accuracy of 90.44% on the 7000 sequence test set (Table 6). This strong performance on a large-scale, independent benchmark demonstrates that our sonification-based approach is not overfit to our initial data set and generalizes well to new classification tasks.

Table 6. Performance of the Fusion Model on the CARE Benchmark Test Set

metric	precision	recall	F1-score	accuracy (%)
macro average	0.91	0.90	0.90	90.44
weighted average	0.91	0.90	0.90	

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736

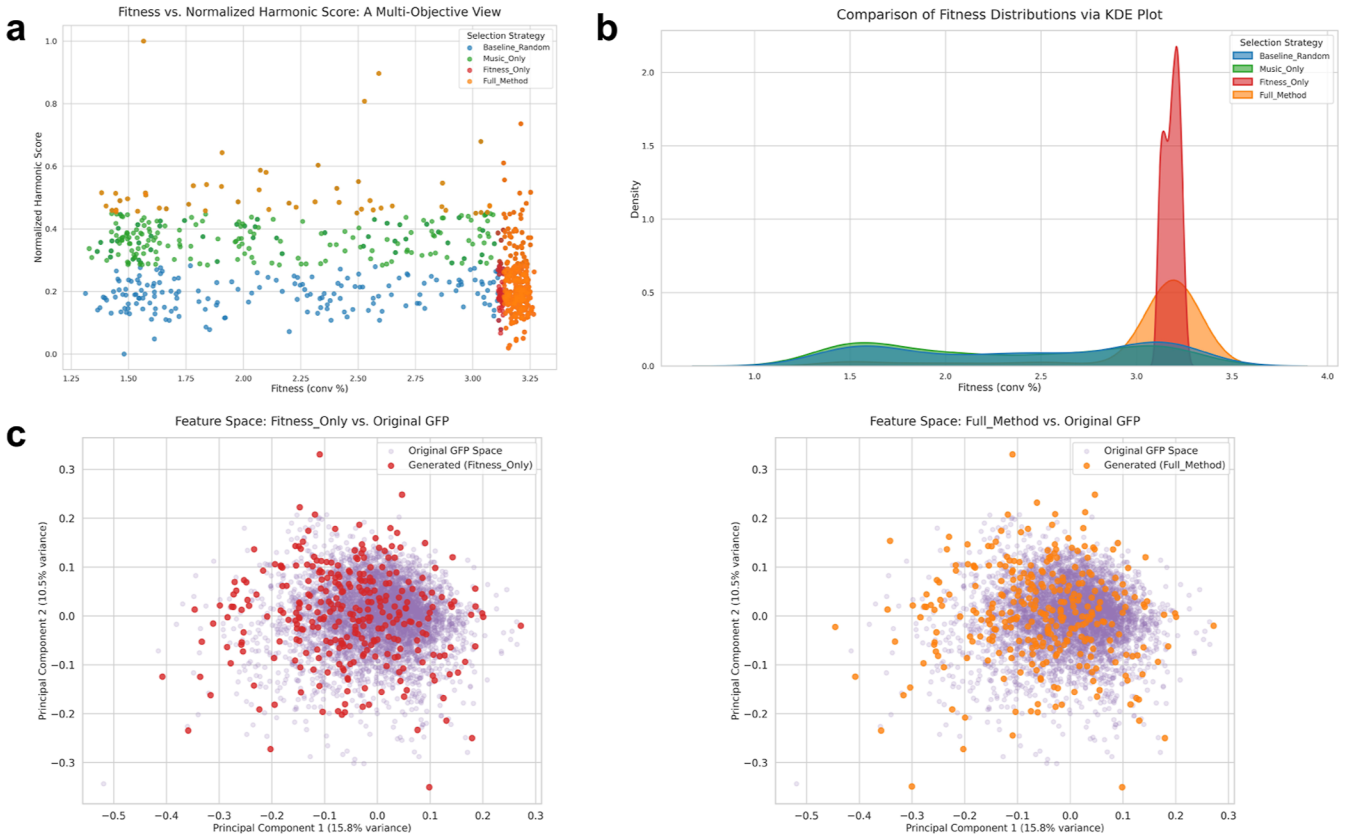


Figure 5. Comparison of selection strategies for multi-objective protein optimization (A) multi-objective landscape: this scatter plot shows the relationship between predicted fitness and normalized harmonic score. The `Full_Method` (orange) explores regions of higher harmonic scores, achieving a mean harmonic score of 0.8396 ± 0.0029. In contrast, the `Fitness_Only` population (red) is concentrated in a zone of high fitness (mean = 3.1831 ± 0.0404) but is constrained to a narrow range of lower harmonic values (mean = 0.8386 ± 0.0018). (B) Fitness distributions: the Kernel density estimate plot compares the fitness distributions. The `Full_Method` exhibits a significantly broader distribution (fitness = 3.0271 ± 0.4414) compared to the sharply peaked `Fitness_Only` population (fitness = 3.1831 ± 0.0404). The large standard deviation of the `Full_Method` quantitatively reflects the mixed nature of its population, comprising both high-performance variants and other sequences that contribute to its exploratory capability. (C) Sequence space visualization: principal component analysis visualizes the distribution of selected sequences. The `Fitness_Only` population forms a compact cluster, corresponding to a diversity score of 7.0261%. The `Full_Method` population occupies a visibly larger area of the sequence space, achieving a higher diversity score of 7.1159% and suggesting its ability to generate a more varied set of high-performing sequences.

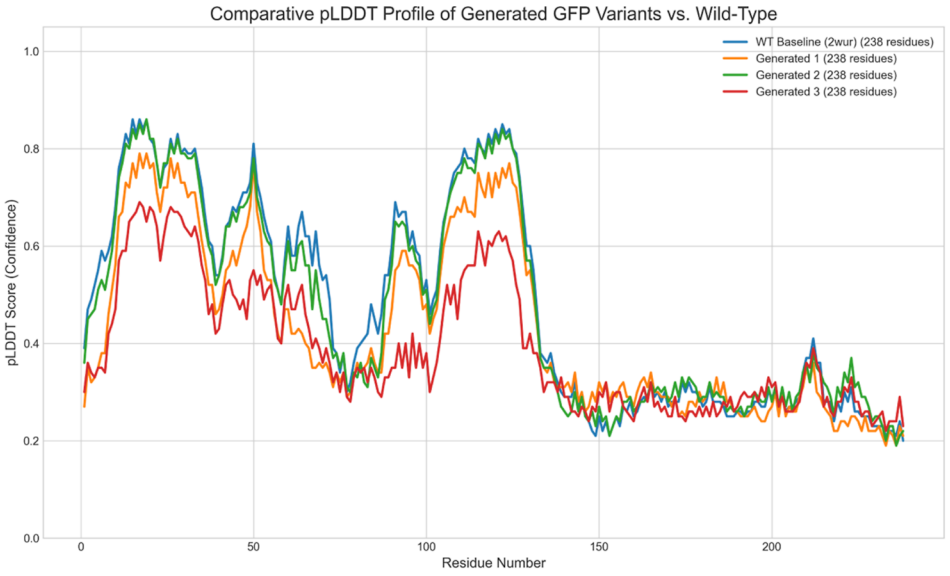
Generative Validation: De Novo Design of Viable GFP Variants. To demonstrate the practical utility of our encoding in a generative context, we integrated our framework into a conditional diffusion model to guide the de novo design of GFP variants. Our goal was to balance the dual objectives of high functional fitness and sequence diversity, as selecting solely for the highest fitness can lead to a narrow set of solutions. To achieve this, we implemented a two-stage selection strategy. First, we identified an elite pool of candidates with the highest predicted fitness, ensuring a baseline of high performance. Then, within this elite pool, we selected the final variants using a composite score that weighed both predicted fitness and the harmonic score derived from our sonification framework.

As shown in Figure 5, this strategy yielded promising results. Compared with a fitness-only selection method, our approach generated a population of variants that not only maintained high fitness levels but also exhibited greater sequence diversity and explored a broader region of the harmonic score landscape. To validate the structural integrity of these designs, computational analysis using ESMFold was performed. The results (Figure 6) indicated that the generated variants largely retained the canonical GFP β-barrel fold with high confidence and low RMSD values (0.062−0.265 Å) relative to the wild-type, providing evidence that our sonification-derived harmonic score may serve as a useful proxy for structural viability in this context.

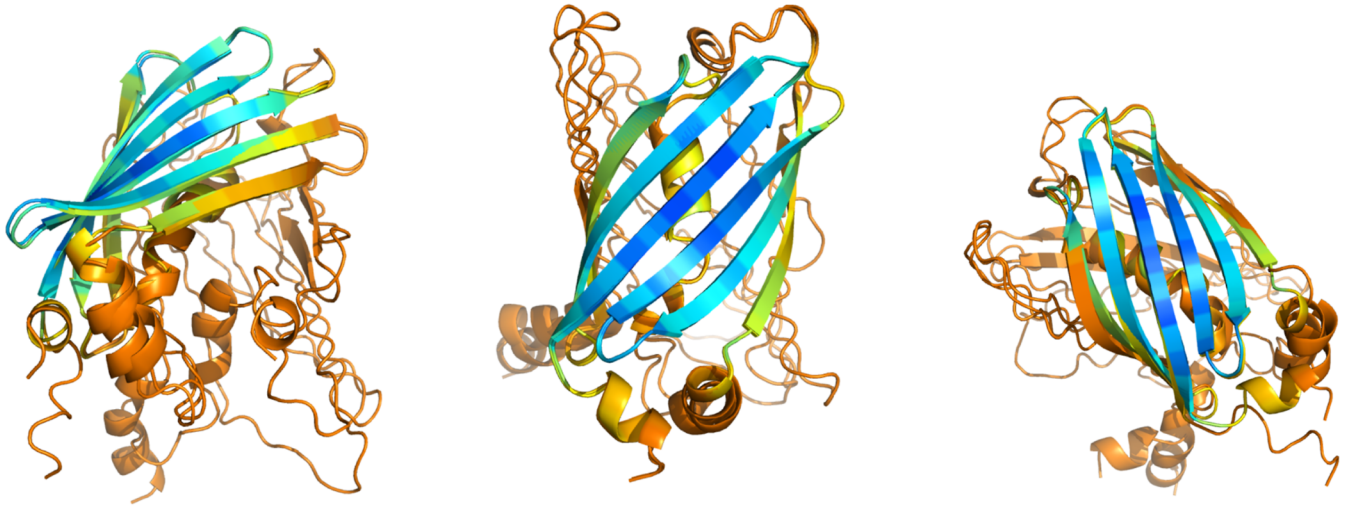
DISCUSSION

A central question motivating this study was whether our encoding rules were arbitrary or essential for performance. Our results provide a nuanced answer: the primary benefit of sonification appears to stem from the structural transformation of a 1D sequence into a 2D, information-dense representation. The strong performance of the semantic ablation model (81.08% accuracy) suggests that a deep learning model can learn functional patterns from the spectrogram's structure alone, treating them as emergent properties. This ability to extract meaningful regularities from a complex, semantically arbitrary representation resonates with neurophysiological findings on pattern recognition in uncertain musical contexts.[63] However, this does not render the encoding rules meaningless. The incremental performance gains from the inverted (82.69%) and our theory-driven (84.00%) models indicate that biophysically grounded semantics provide a valuable inductive bias, an idea supported by our MD simulations that

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736



(a) Comparative per-residue confidence plot (pLDDT)



(b) Generated Variant 1 (c) Generated Variant 2 (d) Generated Variant 3

Figure 6. Computational validation of generated GFP variants. (a) Per-residue confidence scores (pLDDT) from ESMFold for the wild-type baseline and three generated variants. (b–d) Predicted 3D structures of the three variants, colored by pLDDT score (blue: >0.9 and orange: <0.5). The generated variants appear to retain the essential β-barrel fold.

validate the premise that mapping static sequence properties to a dynamic medium is a physically meaningful approach. This reframes our method within a broader intellectual tradition of cross-pollination between biology and music, where such analogies have been used both for analysis and the creation of novel bioinspired materials.[64−66]

This 1D-to-2D transformation is a powerful and established principle and is not a new hypothesis. Its success is a cornerstone of landmark models like AlphaFold, which converts sequence information into 2D distance maps to capture the spatial relationships between all residue pairs.[14,15] The same paradigm is fundamental to state-of-the-art methods in genomics[16,17] and RNA secondary structure prediction,[18] where predicting 2D interaction maps from 1D sequences is crucial for modeling LRIs. This principle is not confined to biology; it is a core strategy in machine learning for capturing LRIs, for instance, in graph neural networks where multiscale representations have yielded order-of-magnitude performance gains on LRI-centric benchmarks.[67,68] By "folding" a sequence into a spectrogram, we bring distant residues into proximity, allowing a standard CNN to model these global dependencies efficiently, without the quadratic cost of attention mechanisms.[69] This echoes other computational paradigms where function is derived from global network properties.[70]

The richness of this representation is further evidenced by our ablation study. The strong standalone performance of both the Visual-Only and Acoustic-Only models demonstrates that functional information is encoded in a way that is accessible through different analytical lenses. This aligns with a growing body of work showing sonification can be a valuable complement to purely visual methods for data discovery in the life sciences,[71,72] and is conceptually analogous to how the mammalian auditory cortex utilizes joint spectro-temporal features for robust sound recognition.[73]

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736

Positioning our work relative to state-of-the-art PLMs highlights a trade-off. While large, pretrained models like ProtBERT and ESM-2 achieve superior accuracy, their performance advantage is removed when training is restricted to our data set. Our fusion model's competitive performance of our fusion model when trained from scratch suggests that the 2D representation is highly data-efficient for this task. This efficiency, however, comes with the overhead of a two-step process involving preconversion of sequences to spectrograms. This points toward a promising future direction: a truly end-to-end architecture such as a Protein Spectrogram Transformer (PST). Inspired by the audio spectrogram transformer,[74] a PST could learn to generate an optimal 2D representation directly from the 1D sequence, combining the representational power of our approach with the end-to-end learning paradigm of PLMs.

We acknowledge the limitations of our study. Our data set, while functionally diverse, is modest by the standards of large-scale pretraining—a well-recognized challenge often necessitating customized approaches or domain adaptation techniques.[75] Similarly, our generative experiments on GFP, while a compelling proof-of-concept, require validation across a broader range of protein families to establish the generalizability of our harmonic score as a proxy for structural viability. Nevertheless, this initial validation is critical, as it suggests that features derived from our framework could serve as a practical regularizer in the de novo protein design cycle, where ensuring a viable three-dimensional fold is a cornerstone of success.[76,77]

In summary, this work establishes protein sonification as a viable strategy for creating computationally efficient and information-dense representations for function prediction. The primary benefit stems from the structural shift to a 2D format, which allows deep learning models to learn emergent functional patterns, with biophysical semantics providing a beneficial but secondary guiding principle.

CONCLUSION

In this study, we developed and evaluated a quantitative sonification framework as a novel method for protein analysis. Our work demonstrates that translating protein sequences into 2D spectrograms creates a feature-rich representation that is effective for function prediction, showing signs of data efficiency and generalizability across different protein classification tasks. Through a series of systematic experiments, we provide evidence that the primary source of the model's predictive power stems from the structural transformation of 1D sequence data into a 2D format. This allows for the capture of complex patterns as emergent properties, a process that is further enhanced but not solely dependent on biophysically informed encoding rules. The utility of this representation extends beyond predictive tasks, as shown by our proof-of-concept where the encoding guided a generative model in designing novel, structurally viable GFP variants. This study contributes to a systematic evaluation of protein sonification as a principled feature engineering strategy, highlighting its potential for both predictive and generative protein modeling. Ultimately, our work reinforces the paradigm of translating biological sequences into other domains, not merely for artistic inspiration,[78] but as a structured approach for creating powerful, bioinspired analytical methods and materials.[65,66]

ASSOCIATED CONTENT

Data Availability Statement

The code is available at https://github.com/wyqmath/Symphony_of_Fate.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.5c01768.

Detailed methodology of the protein-to-music sonification framework, including quantitative mapping rules, technical implementation for audio synthesis, and a discussion of alternative schemes; protocols for MD simulations and RMSF analysis; complete implementation details for all classification systems, including the end-to-end fusion model and all benchmark models (1D-CNN, PLMs, homology-based); implementation of the generative framework for GFP variants, detailing the fitness predictor, Harmonic score, and directed evolution strategy; and case study on spectral feature analysis of protein structures (PDF)

AUTHOR INFORMATION

Corresponding Author

Kai Wei − *Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Urumqi 830049 Xinjiang, China;* orcid.org/0000-0001-5431-8290; Email: kaiwei@xju.edu.cn

Authors

Yiquan Wang − *Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Urumqi 830049 Xinjiang, China; College of Mathematics and System Science, Xinjiang University, Urumqi 830046 Xinjiang, China;* orcid.org/0000-0003-1417-5752

Minnuo Cai − *Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Urumqi 830049 Xinjiang, China;* orcid.org/0009-0007-7151-8068

Yuhua Dong − *Advanced Research Institute of Multidisciplinary Sciences, Beijing Institute of Technology, Beijing 100081, China;* orcid.org/0009-0001-4314-6806

Yahui Ma − *Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Urumqi 830049 Xinjiang, China;* orcid.org/0009-0004-7443-5645

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.5c01768

Author Contributions

Yiquan Wang: conceptualization, methodology, software, investigation, writing − original draft, and writing − review and editing. **Minnuo Cai:** software, investigation, visualization, writing − original draft, and writing − review and editing. **Yuhua Dong:** conceptualization, methodology, software, investigation, writing − original draft, and writing − review and editing. **Yahui Ma:** formal analysis and writing − original draft. **Kai Wei:** project administration and writing − review and editing.

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736

Funding

This work was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (Grant No. 2024D01C216) and the "Tianchi Talents" introduction plan.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors extend their sincere gratitude to colleagues, mentors, and Shenzhen X-Institute for their invaluable discussions and support, and to the anonymous reviewers for their constructive feedback, which significantly improved this manuscript. Sincere thanks are also extended to Chenyu Shen for her invaluable help with the musical scores. We also acknowledge the public databases that made this work possible, including the National Center for Biotechnology Information (NCBI), UniProt, BRENDA, TCDB, and APD.

■ REFERENCES

(1) Hegyi, H.; Gerstein, M. The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J. Mol. Biol.* **1999**, *288* (1), 147−164.
(2) Sadowski, M. I.; Jones, D. T. The sequence−structure relationship and protein function prediction. *Curr. Opin. Struct. Biol.* **2009**, *19* (3), 357−362.
(3) Whitford, D. *Proteins: Structure and Function*; John Wiley & Sons, 2013.
(4) Koehler Leman, J.; Szczerbiak, P.; Renfrew, P. D.; Gligorijevic, V.; Berenberg, D.; Vatanen, T.; Taylor, B. C.; Chandler, C.; Janssen, S.; Pataki, A.; et al. Sequence-structure-function relationships in the microbial protein universe. *Nat. Commun.* **2023**, *14* (1), 2351.
(5) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **2023**, *379* (6637), 1123−1130.
(6) Elnaggar, A.; Heinzinger, M.; Dallago, C.; Rehawi, G.; Wang, Y.; Jones, L.; Gibbs, T.; Feher, T.; Angerer, C.; Steinegger, M.; Bhowmik, D.; Rost, B. ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44* (10), 7112−7127.
(7) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with alphafold. *Nature* **2021**, *596* (7873), 583−589.
(8) Wang, L.; Li, X.; Zhang, H.; Wang, J.; Jiang, D.; Xue, Z., Wang, Y.. A comprehensive review of protein language models. **2025**, arXiv preprint arXiv:2502.06881.
(9) Bepler, T.; Berger, B. Learning the protein language: Evolution, structure, and function. *Cell Syst.* **2021**, *12* (6), 654−669.
(10) Chen, J. Y.; Wang, J. F.; Hu, Y.; Li, X. H.; Qian, Y. R.; Song, C. L. Evaluating the advancements in protein language models for encoding strategies in protein function prediction: a comprehensive review. *Front. Bioeng. Biotechnol.* **2025**, *13*, 1506508.
(11) Yu, C. H.; Qin, Z.; Martin-Martinez, F. J.; Buehler, M. J. A self-consistent sonification method to translate amino acid sequences into musical compositions and application in protein design using artificial intelligence. *ACS Nano* **2019**, *13* (7), 7471−7482.
(12) Buehler, M. J. Unsupervised cross-domain translation via deep learning and adversarial attention neural networks and application to music-inspired protein designs. *Patterns* **2023**, *4* (3), 100692.
(13) Milazzo, M.; Anderson, G. I.; Buehler, M. J. Bioinspired translation of classical music into de novo protein structures using deep learning and molecular modeling. *Bioinspiration Biomimetics* **2022**, *17* (1), 015001.
(14) Wang, S.; Sun, S.; Li, Z.; Zhang, R.; Xu, J. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput. Biol.* **2017**, *13* (1), No. e1005324.
(15) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577* (7792), 706−710.
(16) Fudenberg, G.; Kelley, D. R.; Pollard, K. S. Predicting 3D genome folding from DNA sequence with Akita. *Nat. Methods* **2020**, *17* (11), 1111−1117.
(17) Schwessinger, R.; Gosden, M.; Downes, D.; Brown, R. C.; Oudelaar, A. M.; Telenius, J.; Teh, Y. W.; Lunter, G.; Hughes, J. R. DeepC: predicting 3D genome folding using megabase-scale transfer learning. *Nat. Methods* **2020**, *17* (11), 1118−1124.
(18) Bugnon, L. A.; Di Persia, L.; Gerard, M.; Raad, J.; Prochetto, S.; Fenoy, E.; Chorostecki, U.; Ariel, F.; Stegmayer, G.; Milone, D. H. sincFold: end-to-end learning of short-and long-range interactions in RNA secondary structure. *Briefings Bioinf.* **2024**, *25* (4), bbae271.
(19) Nogales, E.; Scheres, S. H. Cryo-em: A unique tool for the visualization of macromolecular complexity. *Mol. Cell* **2015**, *58* (4), 677−689.
(20) Yip, K. M.; Fischer, N.; Paknia, E.; Chari, A.; Stark, H. Atomic-resolution protein structure determination by cryo-em. *Nature* **2020**, *587* (7832), 157−161.
(21) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **2009**, *19* (2), 120−127.
(22) AlQuraishi, M. Machine learning in protein structure prediction. *Curr. Opin. Chem. Biol.* **2021**, *65*, 1−8.
(23) Meng, Y.; Zhang, Z.; Zhou, C.; Tang, X.; Hu, X.; Tian, G.; Yang, J.; Yao, Y. Protein structure prediction via deep learning: an in-depth review. *Front. Pharmacol* **2025**, *16*, 1498662.
(24) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. Alphafold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50* (D1), D439−D444.
(25) Hertig, S.; Latorraca, N. R.; Dror, R. O. Revealing atomic-level mechanisms of protein allostery with molecular dynamics simulations. *PLoS Comput. Biol.* **2016**, *12* (6), No. e1004746.
(26) Vander Meersche, Y.; Cretin, G.; Gheeraert, A.; Gelly, J. C.; Galochkina, T. Atlas: protein flexibility description from atomistic molecular dynamics simulations. *Nucleic Acids Res.* **2024**, *52* (D1), D384−D392.
(27) Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; Fergus, R. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118* (15), No. e2016239118.
(28) Krishna, R.; Wang, J.; Ahern, W.; Sturmfels, P.; Venkatesh, P.; Kalvet, I.; Lee, G. R.; Morey-Burrows, F. S.; Anishchenko, I.; Humphreys, I. R.; et al. Generalized biomolecular modeling and design with rosettafold all-atom. *Science* **2024**, *384* (6693), No. eadl2528.
(29) Kyro, G. W.; Qiu, T., Batista, V. S. A model-centric review of deep learning for protein design. **2025**, arXiv:2502.19173.
(30) Benson, D. *Music: A Mathematical Offering*; Cambridge University Press, 2006.
(31) Musimathics, G. L. *The Mathematical Foundations of Music Volume 1*; MIT Press, 2011.
(32) King, R. D.; Angus, C. G. Pm—protein music. *Bioinformatics* **1996**, *12* (3), 251−252.
(33) Takahashi, R.; Miller, J. H. Conversion of amino-acid sequence in proteins to classical music: search for auditory patterns. *Genome Biol.* **2007**, *8* (5), 405.
(34) Mössinger, J. The music of life. *Nature* **2005**, *435* (7040), 280.
(35) Ghavami, S.; Toozandehjani, H.; Ghavami, G.; Sardari, S. Innovative protein translation into music and color image applicable for assessing protein alignment based on bio-mimicking human perception system. *Int. J. Biol. Macromol.* **2018**, *119*, 896−901.

(36) Su, J.; Zhou, P. Musical protein: Mapping the time sequence of music onto the spatial architecture of proteins. *Comput. Methods Programs Biomed.* **2024**, *252*, 108233.
(37) Cosic, I. Macromolecular bioactivity: is it resonant interaction between macromolecules?-theory and applications. *IEEE Trans. Biomed. Eng.* **1994**, *41* (12), 1101−1114.
(38) Cosic, I. *The Resonant Recognition Model of Macromolecular Bioactivity: Theory and Applications*; Birkhäuser, 2012.
(39) Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81* (1), 140−144.
(40) Cosic, I. The resonant recognition model of bio-molecular interactions: possibility of electromagnetic resonance. *Pol J Med Phys Eng.* **2001**, *7* (1), 73−87.
(41) Cosic, I.; Pirogova, E. Bioactive peptide design using the resonant recognition model. *Nonlinear Biomed. Phys.* **2007**, *1* (1), 7.
(42) Cosic, I.; Cosic, D. Dna-protein interactions at distance explained by the resonant recognition model. *Int. J. Sci.* **2024**, *13* (11), 1−5.
(43) Bueno de Souza, F.; Pimenta-Zanon, M. H.; Henriques, D.; Pinto, M. A.; Balsa, C.; Rufino, J.; Lopes, F. c. M. Resonant recognition model as a preprocessing technique for rna classification. In *International Conference on Advanced Research in Technologies, Information, Innovation and Sustainability*; Springer, 2024; pp 3−17.
(44) Chrysostomou, C.; Seker, H.; Aydin, N.; Haris, P. I. Complex resonant recognition model in analysing influenza a virus subtype protein sequences. In *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine*; IEEE, 2010; pp 1−4.
(45) O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretdin, A.; Bao, Y.; Blinkova, O.; Brover, V.; Chetvernin, V.; Choi, J.; Cox, E.; Ermolaeva, O.; Farrell, C. M.; Goldfarb, T.; Gupta, T.; Haft, D.; Hatcher, E.; Hlavina, W.; Joardar, V. S.; Kodali, V. K.; Li, W.; Maglott, D.; Masterson, P.; McGarvey, K. M.; Murphy, M. R.; O'Neill, K.; Pujar, S.; Rangwala, S. H.; Rausch, D.; Riddick, L. D.; Schoch, C.; Shkeda, A.; Storz, S. S.; Sun, H.; Thibaud-Nissen, F.; Tolstoy, I.; Tully, R. E.; Vatsan, A. R.; Wallin, C.; Webb, D.; Wu, W.; Landrum, M. J.; Kimchi, A.; Tatusova, T.; DiCuccio, M.; Kitts, P.; Murphy, T. D.; Pruitt, K. D. Reference sequence (refseq) database at ncbi: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733−D745.
(46) Bateman, A.; Martin, M. J.; Orchard, S.; Magrane, M.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E. H.; Britto, R.; Bye-A-Jee, H.; Cukura, A.; et al. Uniprot: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51*, D523−D531.
(47) Wang, G.; Schmidt, C.; Li, X.; Wang, Z. APD6: the antimicrobial peptide database is expanded to promote research and development by deploying an unprecedented information pipeline. *Nucleic Acids Res.* **2025**, gkaf860.
(48) Saier, M. H., Jr; Reddy, V. S.; Moreno-Hagelsieb, G.; Hendargo, K. J.; Zhang, Y.; Iddamsetty, V.; Lam, K.; Tian, N.; Russum, S.; Wang, J.; et al. The transporter classification database (TCDB): 2021 update. *Nucleic Acids Res.* **2021**, *49* (D1), D461−D467.
(49) Anandkumar, A.; Arnold, F.; Liu, S.; Mora, A.; Wittmann, B.; Yang, J.; Yue, Y. Care: a benchmark suite for the classification and retrieval of enzymes. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 3094−3121.
(50) Li, W.; Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22* (13), 1658−1659.
(51) Japkowicz, N.; Stephen, S. The class imbalance problem: A systematic study. *Intell. Data Anal.* **2002**, *6*, 429−449.
(52) Bellesia, G.; Jewett, A. I.; Shea, J. Sequence periodicity and secondary structure propensity in model proteins. *Protein Sci.* **2010**, *19* (1), 141−154.
(53) Spence, C. Crossmodal correspondences: A tutorial review. *Atten. Percept. Psychophys.* **2011**, *73* (4), 971−995.
(54) Kuzmanic, A.; Zagrovic, B. Determination of ensemble-average pairwise root mean-square deviation from experimental b-factors. *Biophys. J.* **2010**, *98* (5), 861−871.
(55) Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE, 2022; pp 11976−11986.
(56) Facebook AI. Model repository for esm2_t6_8m_ur50d. 2022. https://huggingface.co/facebook/esm2_t6_8M_UR50D (accessed Aug 26, 2025).
(57) Facebook AI. Model repository for esm2_t12_35m_ur50d. 2022. https://huggingface.co/facebook/esm2_t12_35M_UR50D (accessed Aug 26, 2025).
(58) Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; McGinnis, S.; Madden, T. L. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **2008**, *36* (Web Server), W5−W9.
(59) Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13* (1), 21−27.
(60) Sarkisyan, K. S.; Bolotin, D. A.; Meer, M. V.; Usmanova, D. R.; Mishin, A. S.; Sharonov, G. V.; Ivankov, D. N.; Bozhanova, N. G.; Baranov, M. S.; Soylemez, O.; et al. Local fitness landscape of the green fluorescent protein. *Nature* **2016**, *533* (7603), 397−401.
(61) BFD Datasets. Bfd. 2018. https://bfd.mmseqs.com/ (accessed Oct 01, 2025).
(62) nferruz. Ur50_2021_04. 2022. https://huggingface.co/datasets/nferruz/UR50_2021_04 (accessed Aug 26, 2025).
(63) Brattico, E.; Delussi, M. Making sense of music: Insights from neurophysiology and connectivity analyses in naturalistic listening conditions. *Hear. Res.* **2024**, *441*, 108923.
(64) Bountouridis, D.; Brown, D. G.; Wiering, F.; Veltkamp, R. C. Melodic similarity and applications using biologically-inspired techniques. *Appl. Sci.* **2017**, *7* (12), 1242.
(65) Giesa, T.; Spivak, D. I.; Buehler, M. J. Reoccurring patterns in hierarchical protein materials and music: the power of analogies. *BioNanoScience* **2011**, *1*, 153−161.
(66) Wong, J. Y.; McDonald, J.; Taylor-Pinney, M.; Spivak, D. I.; Kaplan, D. L.; Buehler, M. J. Materials by design: Merging proteins and music. *Nano Today* **2012**, *7* (6), 488−495.
(67) Dong, H.; Xu, J.; Yang, Y.; Zhao, R.; Wu, S.; Yuan, C.; Li, X.; Maddison, C. J.; Han, L.; MeGraph: Capturing long-range interactions by alternating local and hierarchical aggregation on multi-scaled graph hierarchy. In *37th Conference on Neural Information Processing Systems*; NIPS, 2023; pp 63609−63641.
(68) Wang, Y.; Cheng, C.; Li, S.; Ren, Y.; Shao, B.; Liu, G.; Zheng, N. Neural P^3M: A Long-Range Interaction Modeling Enhancer for Geometric GNNs. In *38th Conference on Neural Information Processing Systems*; NIPS, 2024; pp 120336−120365.
(69) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *31st Conference on Neural Information Processing Systems*; NIPS, 2017.
(70) Amitai, G.; Shemesh, A.; Sitbon, E.; Shklar, M.; Netanely, D.; Venger, I.; Pietrokovski, S. Network analysis of protein structures identifies functional residues. *J. Mol. Biol.* **2004**, *344* (4), 1135−1146.
(71) Braun, R.; Tfirn, M.; Ford, R. M. Listening to life: Sonification for enhancing discovery in biological research. *Biotechnol. Bioeng.* **2024**, *121* (10), 3009−3019.
(72) Martin, E. J.; Meagher, T. R.; Barker, D. Using sound to understand protein sequence data: new sonification algorithms for protein sequences and multiple sequence alignments. *BMC Bioinf.* **2021**, *22*, 456.
(73) Patil, K.; Pressnitzer, D.; Shamma, S.; Elhilali, M. Music in our ears: the biological bases of musical timbre perception. *PLoS Comput. Biol.* **2012**, *8* (11), No. e1002759.
(74) Yuan, G., Chung, Y.-A., Glass, J. AST: Audio spectrogram transformer. **2021**, arXiv preprint arXiv:2104.01778.

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736

(75) Orouji, S.; Liu, M. C.; Korem, T.; Peters, M. A. Domain adaptation in small-scale and heterogeneous biological datasets. *Sci. Adv.* **2024**, *10* (51), No. eadp6040.
(76) Hellinga, H. W. Rational protein design: combining theory and experiment. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (19), 10015−10017.
(77) Schmidt, T.; Bergner, A.; Schwede, T. Modelling three-dimensional protein structures for applications in drug design. *Drug Discovery Today* **2014**, *19* (7), 890−897.
(78) Tay, N. W.; Liu, F.; Wang, C.; Zhang, H.; Zhang, P.; Chen, Y. Z. Protein music of enhanced musicality by music style guided exploration of diverse amino acid properties. *Heliyon* **2021**, *7* (9), No. e07933.

https://doi.org/10.1021/acs.jcim.5c01768
J. Chem. Inf. Model. 2025, 65, 12723−12736