Capturing Protein Dynamics and Their Determinants Using Explainable Artificial Intelligence

Faraneh Haddadi, Stanislav Mazurenko
faraneh.haddadi@recetox.muni.cz

MUNI | RECETOX

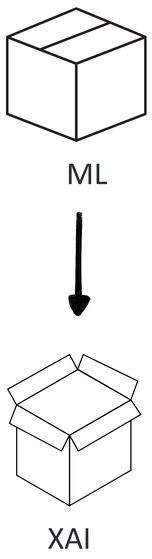
LOSCHMIDT LABORATORIES

Background

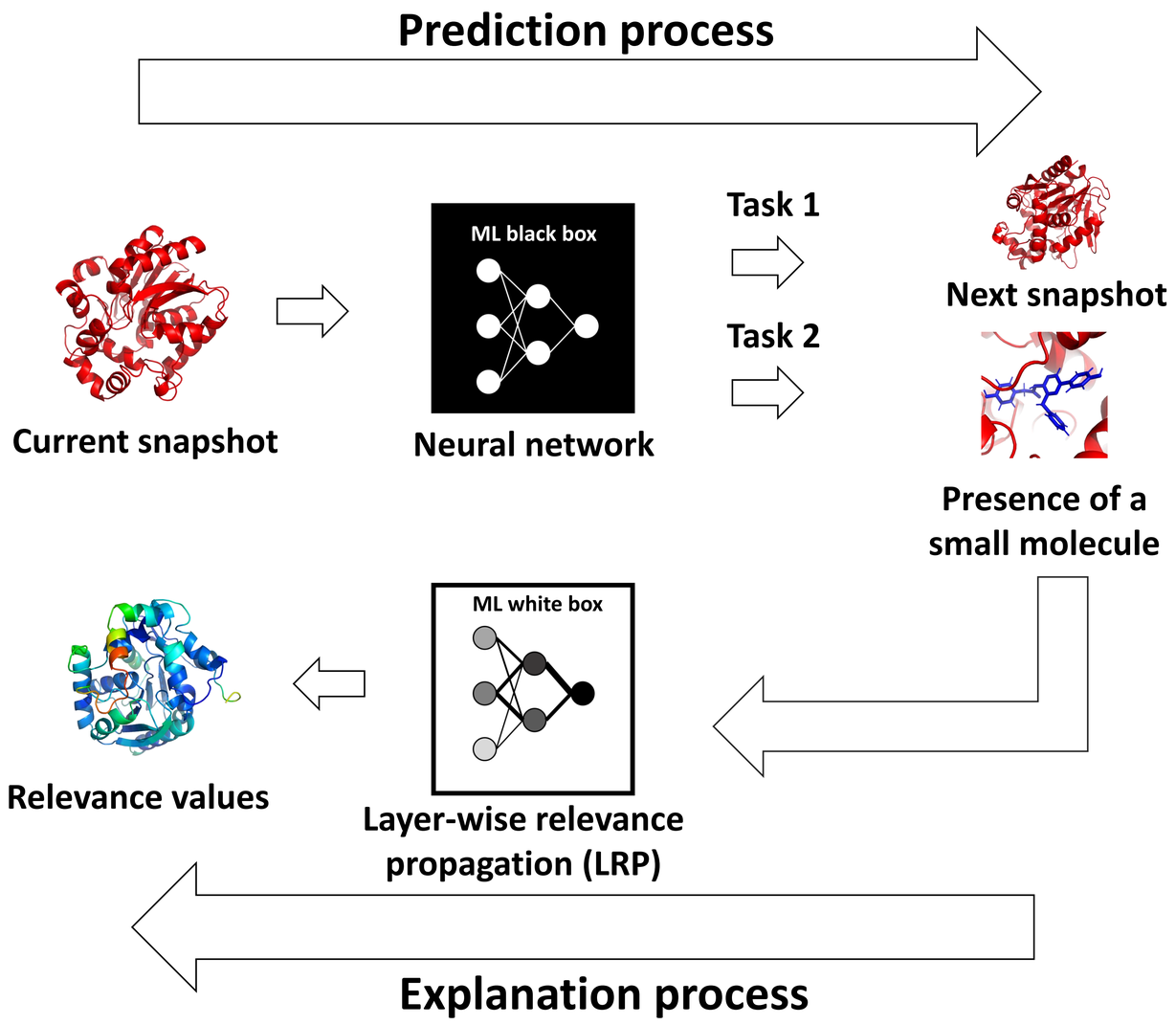
The complexity of biological systems makes it challenging to analyze their behavior using traditional methods alone. Machine learning (ML) and explainable artificial Intelligence (XAI) can help overcome this challenge by providing powerful tools for analyzing large and complex data sets, e.g., molecular dynamics (MD) data, identifying patterns and relationships, and making predictions.

Spotlights

- ML learns patterns from data to make predictions. XAI provides transparent explanations of ML predictions ensuring accountability and reliability.
- The project aims to combine protein engineering with data from molecular dynamics simulations using advanced XAI methods.
- We are using Layer-wise Relevance Propagation (LRP) in XAI, a key XAI method.
- LRP visualizes input feature and neuron contributions to predictions by propagating output through network layers, identifying influential neurons and assigning relevance values.

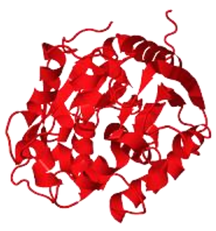


Overview of the project

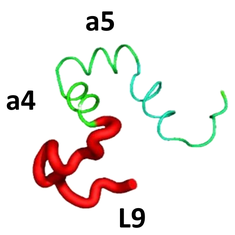


Use cases

Luciferases



Anc$^{HLD-RLuc}$

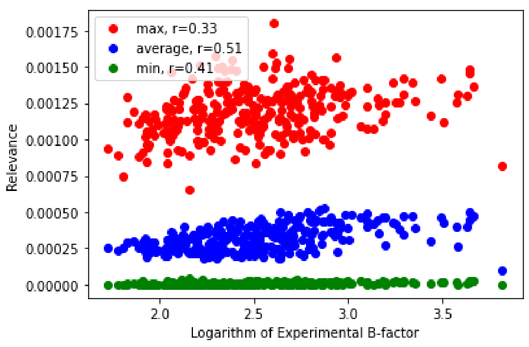


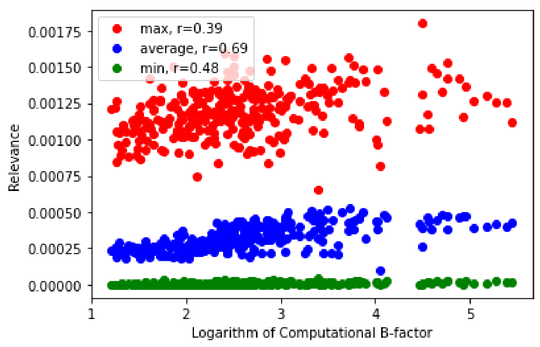
Important elements

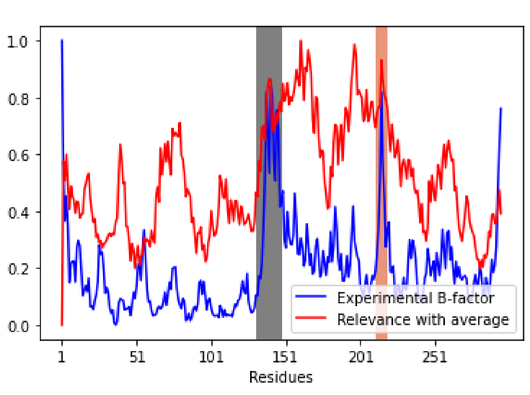
Luciferases enable bioluminescence, with Anc$^{HLD-RLuc}$ being a hybrid of haloalkane dehalogenases and *Renilla* luciferase. Its dynamic secondary structure, like alpha helices and loops, affects its function in bioluminescence assays.

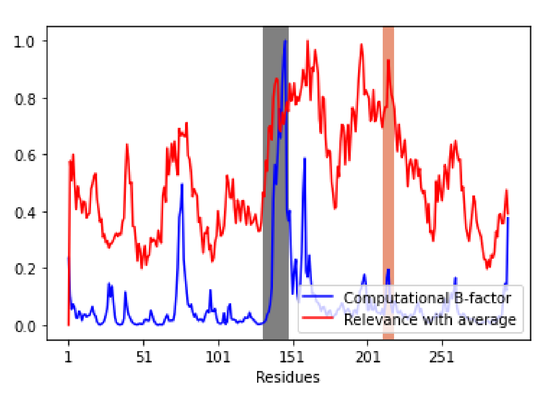
Using MD trajectories of Anc$^{HLD-RLuc}$, this study created a supervised learning task, treating snapshots as input and predicting subsequent snapshots 5 ns later. Relevance values are aggregated per residue using average and maximum methods over simulation time.

XAI revealed differences between relevance values and B-factors, identifying crucial protein residues for motion overlooked by B-factors.









Element	Residues
L9	133-149
a4	150-157
a5'	159-168
L14	213-220

Protein	Order of computational B-factor [1]	Last layer relevancies order	
		Average	Maximum
Anc$^{HLD-RLuc}$	L9 > α4 > α5'	α5' > α4 > L9	L9 > α5'> α4

ApoE4

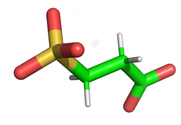
ApoE4 is a genetic variant of protein ApoE. Its dynamics influence its function in lipid metabolism and clearance of amyloid beta, a key molecule in Alzheimer's disease.

We calculated how inter-resiedue distances differ between consecutive snapshots of ApoE4 dynamics.

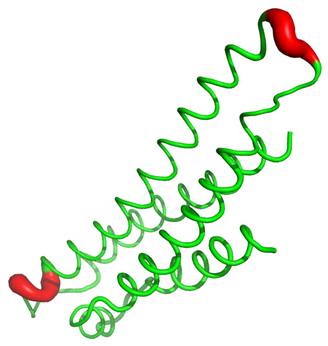
We employed a fully connected neural network with two hidden layers to distinguish between snapshots with ApoE4 alone (label = 0) or in the presence of a small molecule 3-SPA (label = 1). Our focus was on four key alpha helices (residues: 24-161).

Our findings highlighted residues Val85, Ala86, Glu87, Val122, Gln123, and Ala124, as having the highest relevance values.

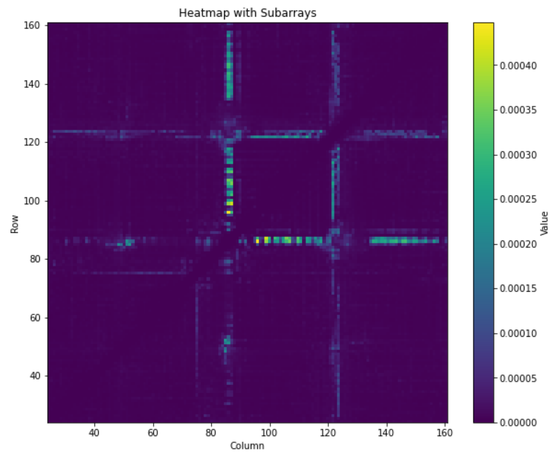
ApoE4



3-SPA structure



Elements identified as relevant in ApoE4



	Train		Cross validation		Test	
	Accuracy	F1 measure	Accuracy	F1 measure	Accuracy	F1 measure
Baseline (Logistic Regression)	0.52	0.68	0.49	0.64	0.49	0.64
Fully connected network (with 2 hidden layers)	0.85	0.86	0.79	0.79	0.82	0.83

Conclusion

This project demonstrates that XAI effectively captures the dynamic features of Anc$^{HLD-RLuc}$ and ApoE4. In Anc$^{HLD-RLuc}$, there is a stronger correlation between B-factors and relevance values in loops compared to helices, suggesting that the network captures loop dynamics more accurately. Regarding ApoE4, LRP reveals highest relevance values where the helices H2-H3 and H3-H4 are interconnected, highlighting their crucial role in distinguishing between binding and unbinding 3-SPA.

[1] Schenkmayerova, A., Pinto, G.P., Toul, M. et al. Engineering the protein dynamics of an ancestral luciferase. *Nat Commun* 12, 3616 (2021). https://doi.org/10.1038/s41467-021-23450-z

Published paper:
Kuba, P., Kohout, P., Haddadi, F. et al. Machine Learning-Guided Protein Engineering. *ACS Catalysis* 13, 21 (2023). https://doi.org/10.1021/acscatal.3c02743