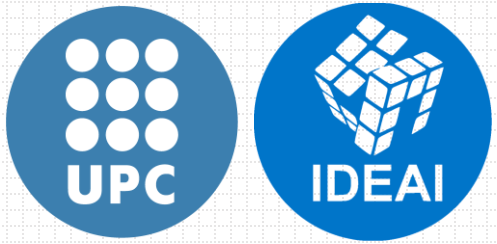
IWBBIO-2022
9th International Work-Conference on Bioinformatics and Biomedical Engineering

A Deep Learning-based method for uncovering GPCR ligand-induced conformational states using interpretability techniques



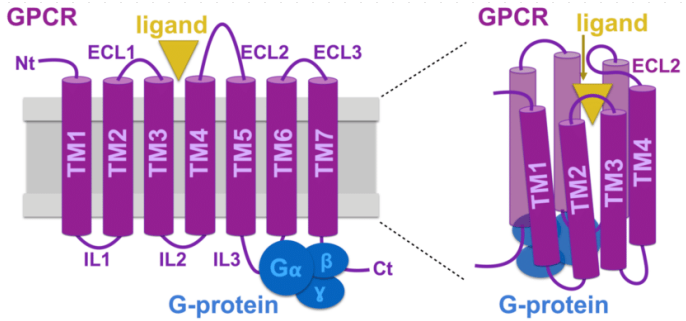
Mario A. Gutiérrez-Mondragón,
Caroline König,
Alfredo Vellido.

Gran Canarias, Spain, June 2022

Contents

Study target:
G protein-coupled receptors (GPCRs).

- Primary transmembrane receptors for signal transduction.
- These proteins act as receptors of plenty of signaling molecules (ligands) - from hormones, neurotransmitters, to photons of light.
- Understanding the functional properties of these receptors is critical to deciphering the signaling process.



Schneider, J., Korshunova, K., Musiani, F., Alfonso-Prieto, M., Giorgetti, A., & Carloni, P. (2018). Predicting ligand binding poses for low-resolution membrane protein models: Perspectives from multiscale simulations. *Biochemical and biophysical research communications*, *498*(2), 366-374.

Data:
Molecular Dynamic Simulations (MD).

- MD has become an established technique to explore the conformational space of proteins at an atomic level.
- MD simulations provided missing information on the dynamics of the receptors.
- The investigation of the structural information of the receptor has relevant implications in pharmacoproteomic.

Tool to perform analysis:
Deep Learning (DL)– based Models

- DL are suitable tools for knowledge discovery.
- DL models have proven to be relevant tools to alleviate and complement the challenges faced by Bioinformatics.
- However, the interpretability of the results has been poorly investigated.

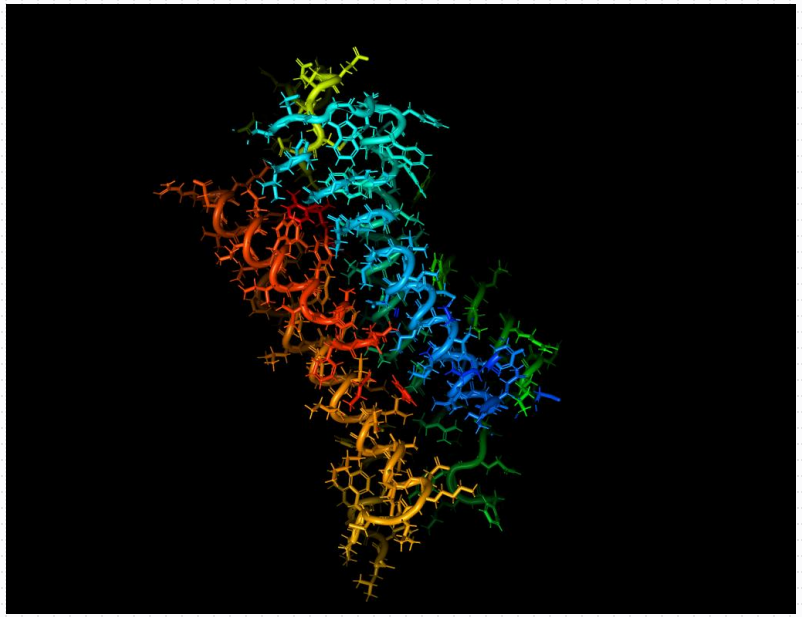
Methodology:

- Data Gathering - GPCR **β2-adrenergic(β2AR)** receptor with full agonist, inverse agonist and ligand-free structure.
- Data Transformation to Residue Interaction Networks.
- Model proposed: Convolution Neural Network on a Supervised classification problem on agonist-specific responses.
- Motif (residues or groups of residues) identification that induce ligand-dependent conformations trough interpretability algorithms. In particular, Layer-Wise Propagation Relevance algorithm.

GPCR-β2-Adrenergic Receptor cloud-based MD simulations on Google Exacycle.

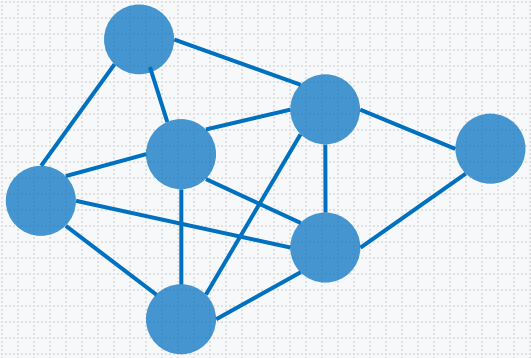
Structure	Description	Type	Simulations	Length
β2ar2RH1-a	β2ar2-inactive	apo (no ligand)	10,000	~12.5 ns
β2ar2RH1-b		full agonist (BI-167107)		
β2ar2RH1-c		partial inverse agonist (carazolol)		
β2ar2RH1-icl3		Icl3 modeled – apo (no ligand)		
β2ar3pg0-a	B2ar2-active	apo (no ligand)		
β2ar3pg0-b		full agonist (BI-167107)		
β2ar3pg0-c		partial inverse agonist (carazolol)		
a2a3eml	B2ar2-active	apo (no ligand)		
B1ar2y02		apo (no ligand)		

Study Case



Kohlhoff, K. J., Shukla, D., Lawrenz, M., Bowman, G. R., Konerding, D. E., Belov, D., Altman, R.B., Pande, V. S.: **Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways**. Nature chemistry, 6(1), 15-21 (2014).

3.1 Residue interaction network representation

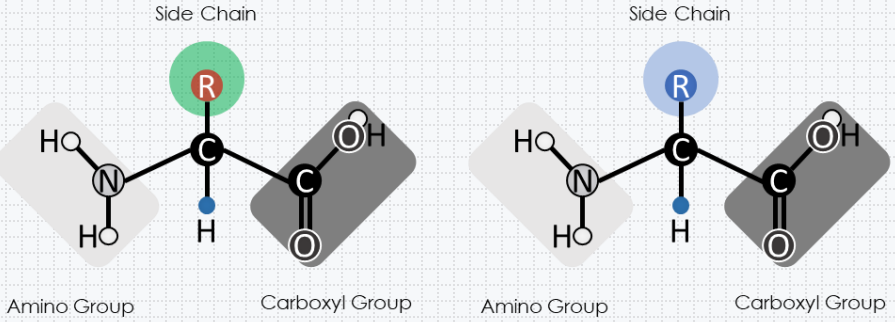


Brinda, K., Vishveshwara, S.: A network representation of protein structures: implications for protein stability. Biophysical journal 89(6), 4159–4170 (2005)

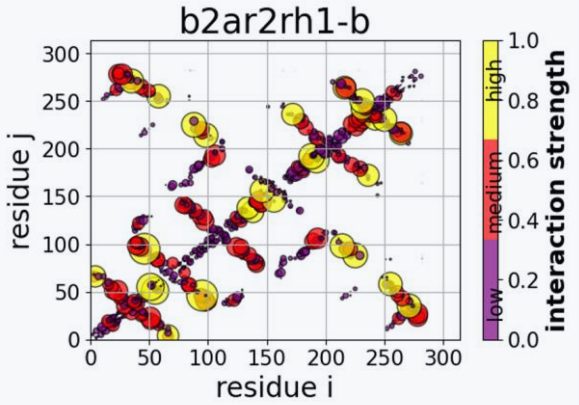
- Network representation of the protein structure to facilitate the study and analysis.
- It represent the three dimensional structure into a two-dimensional space.
- Each amino acid (residue) refers to a node, and the strength of the noncovalent interactions between two amino acids is evaluated for edge determination.

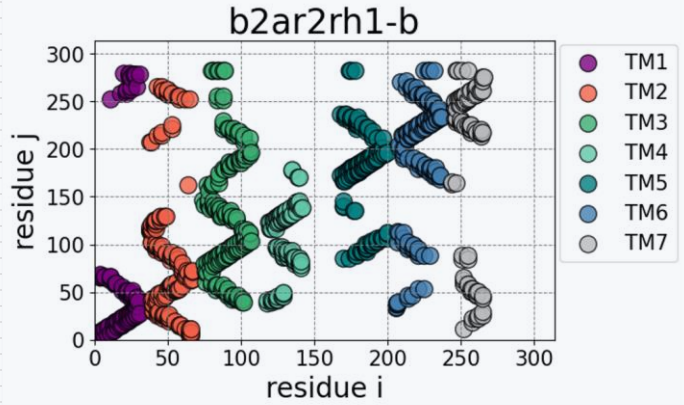
- The interaction strength is evaluated as a percentage:

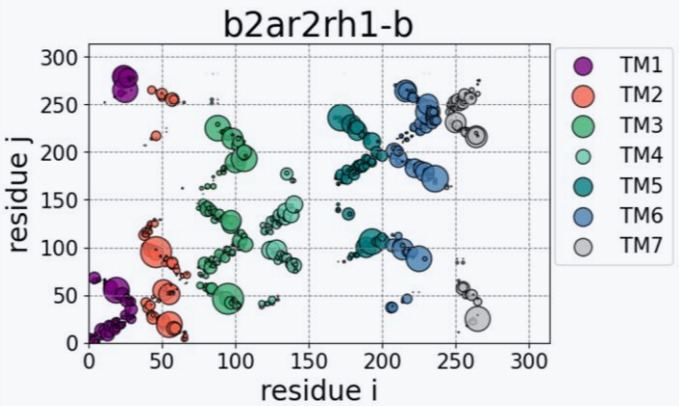
$$I_{ij} = \frac{n_{ij}}{\sqrt{N_{ij}}} \times 100 \rightarrow edge\ if\ i \neq j\ and\ I_{ij} > I_{min}$$



- n_{ij} refers to the number of distinct atom pairs between the side chains of the residues *i* and *j*, which come within a cutoff distance (4.5 Å as a default).
- Ni and Nj represent normalization factors.





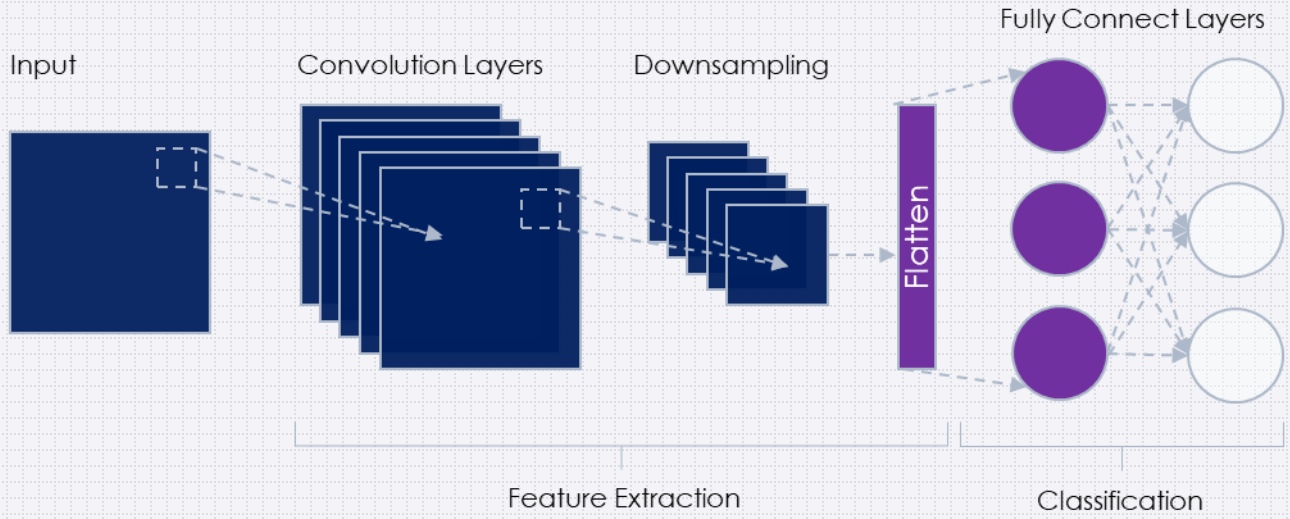


- Computations were made using PSN module from Wordom Software.

- Data Splits.

class	description	# Training samples	# Test samples
2RH1-b	β2AR- full agonist	8,089	5,393
2RH1-c	β2AR- inverse agonist	7,711	5,141
2RH1-icl3	apo (no ligand)	7,659	5,106
	Total:	23,459	15,640

- Model Architecture Proposed.



- Overview on Convolution Neural Networks.

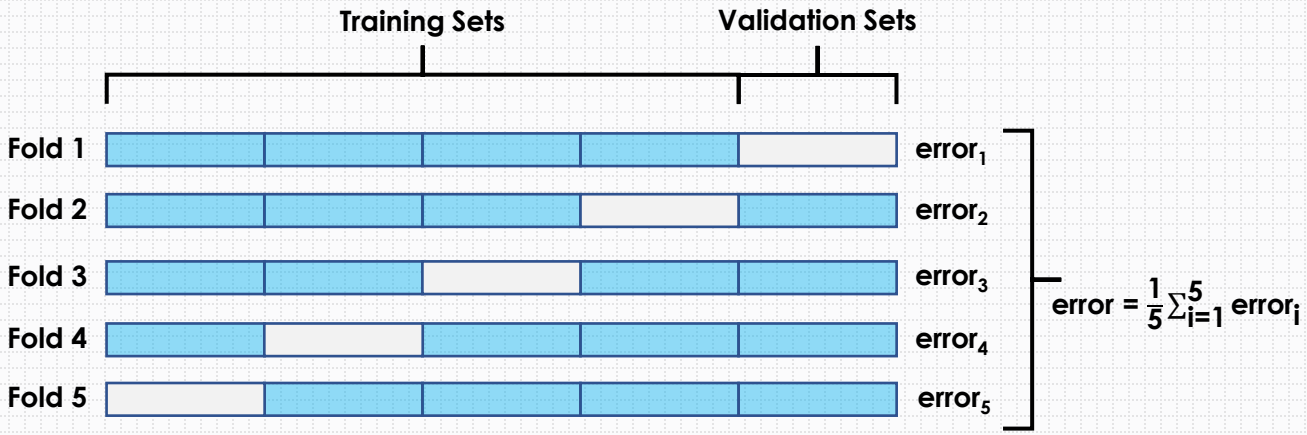
Layer(type)	Output Shape	# Parameters
Conv2d-1	[-1, 32, 310, 310]	832
ReLu-2	[-1, 32, 310, 310]	0
MaxPool2d-3	[-1, 32, 155, 155]	0
Conv2d-4	[-1, 32, 151, 151]	25,632
ReLu-5	[-1, 32, 150, 150]	0
MaxPool2d-6	[-1, 32, 75, 75]	0
Flatten-7	[-1, 18,000]	0
Linear-8	[-1, 32]	5, 760, 032
ReLu-9	[-1, 32]	0
Dropout-10	[-1, 32]	0
Linear-11	[-1, 3]	99

Total params: 5, 768, 595
Trainable params: 5, 786, 59

*Algorithms and computations were developed using PyTorch and Python.

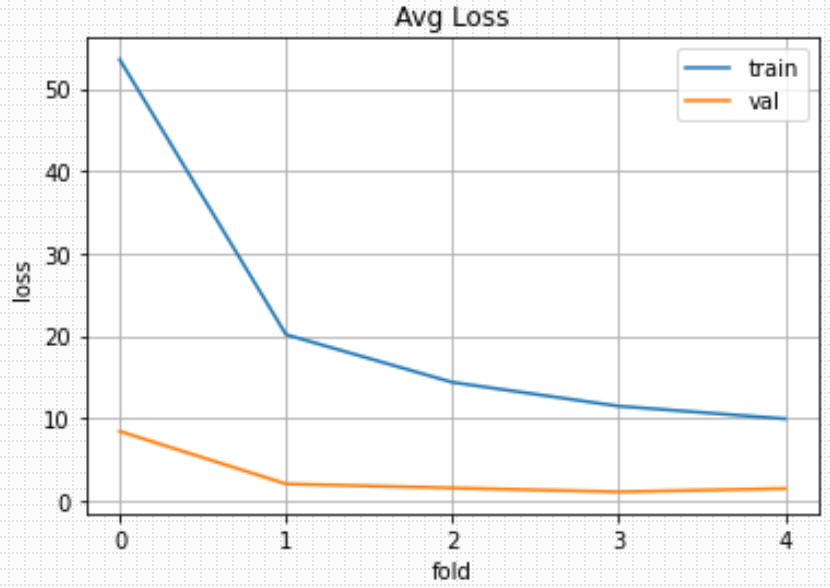
5. Results and Discussion

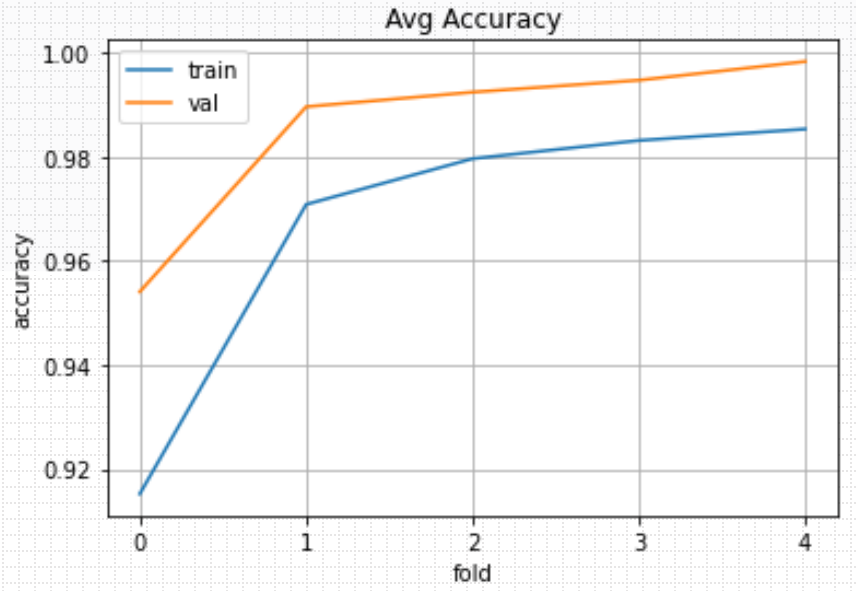
- Cross K-Fold Validation Training



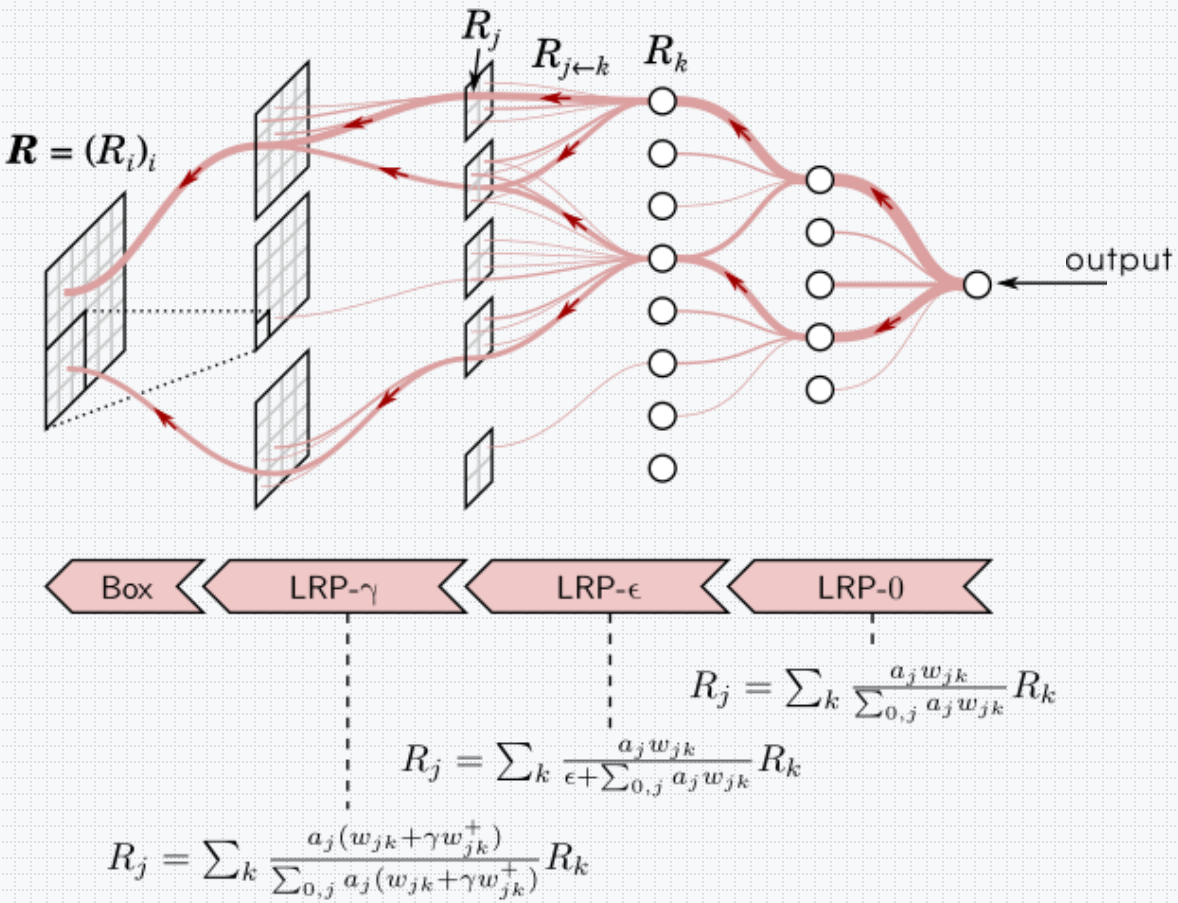
# fold	AVG Training Loss	AVG Validation Loss	AVG Training Accuracy	AVG Validation Accuracy
1	53.5658	8.3702	0.9152	0.9541
2	20.1192	1.9798	0.9709	0.9897
3	14.3261	1.4840	0.9797	0.9925
4	11.4291	1.0138	0.9832	0.9948
5	9.8927	1.4139	0.9854	0.9984
Fold AVG:	**21.8665**	**2.8523**	**0.9668**	**0.9859**

- Learning Curves



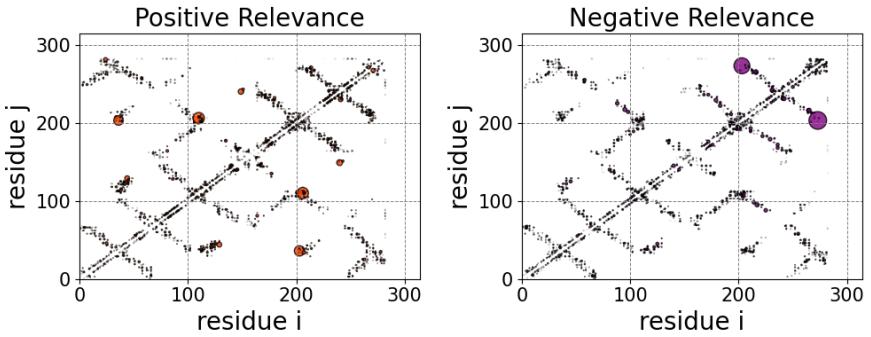


- Layer-Wise Relevance Propagation

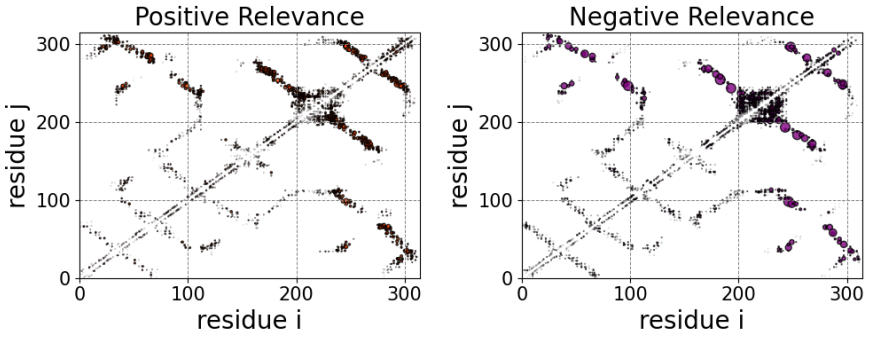


Montavon, G., Binder, A., Lapuschkin, S., Samek, W., & Müller, K. R. (2019). Layer-wise relevance propagation: an overview. *Explainable AI: interpreting, explaining and visualizing deep learning*, 193-209.

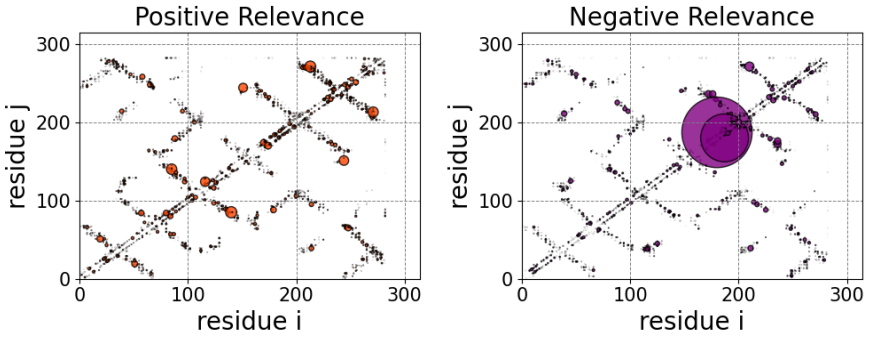
a) AVG Relevance β2ar2 with full agonist.



b) AVG Relevance β2ar2 with inverse agonist.



c) AVG Relevance β2ar2 free ligand.



- Motifs Identification

5.1 Interpretability of the Results

- Relevance Per Transmembrane.

a) Transmembrane AVG Relevance β2ar2 with full agonist.

region	# RI	AVG R
TM1	178	0.002775
TM2	219	0.003977
TM3	236	0.003365
TM4	126	0.005040
TM5	211	0.004188
TM6	**269**	**0.005906**
TM7	186	0.003853

b) Transmembrane AVG Relevance β2ar2 with inverse agonist.

region	# RI	AVG R
TM1	183	0.006238
TM2	239	0.007591
TM3	260	0.013146
TM4	**116**	**0.018101**
TM5	219	0.011499
TM6	291	0.011066
TM7	213	0.012098

c) Transmembrane AVG Relevance β2ar2 free ligand.

region	# RI	AVG R
TM1	214	0.005606
TM2	179	0.002913
TM3	251	0.002294
TM4	164	0.002551
TM5	445	0.003197
TM6	**257**	**0.012646**
TM7	66	0.002494

5. Conclussions

- The study of the conformational space of GPCRs has critical implications in Drug Discovery process.
- Molecular Dynamics and Machine Learning-based models enable a further understanding of molecular processes providing relevant insights into the dynamics of proteins.
- The proposed methodology addresses the transformation of the raw simulations into a Protein Interaction Network.
- We have presented a Convolution Neural Network with high accuracy on a classification problem to provide evidence of the ligand-specific GPCR activity.
- We also include an interpretability study of the prediction results. Therefore, we can identify relevant features (motifs) that underlie conformational rearrangements influenced by ligands.
- Importantly, we provide the trustworthiness of the proposed model and a method to further assess its predictions in this domain.

Thank You!

mario.alberto.gutierrez@upc.edu
caroline.leonore.konig@upc.edu
avellido@cs.upc.edu