Exploring SARS-CoV-2 Main Protease Binding Site with PharmQSAR

Ariadna Llop-Peiró[1], Javier Luque[3], Tiziana Ginex[2], Santi Garcia-Vallvé[1], Gerard Pujadas[1]

ariadna.llop@urv.cat, fjluque@ub.edu, tiziana.ginex@pharmacelera.com, santi.garcia-vallve@urv.cat, gerard.pujadas@gmail.com

1) Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, 2) Pharmacelera SL, 3) Institut de Biomedicina (IBUB) and Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona

The SARS-CoV-2 Main Protease (Mpro) structure has been widely studied for computer-aided drug discovery. In this study, we conducted a 3D-QSAR analysis on 104 protein-ligand complexes from the Protein Data Bank (PDB), accompanied by their respective activity data (pIC_{50}). Despite variability in the activity sources, the model **demonstrated robustness** and **avoided overfitting**, capturing key structural features critical for ligand interaction. Importantly, this QSAR model was developed using crystallized bioactive poses by **aligning proteins rather than ligands**, enabling the exploration of a broader chemical space.

This study aims to analyze the contributions of these **104 ligands** to bioactivity for M-pro, define a pharmacophore model, and identify the key features necessary for an active lead compound.

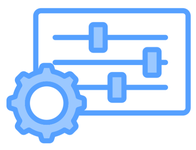
Additionally, we aim to develop a **QSAR model** capable of accurately predicting the activity values of potential M-pro inhibitors with high accuracy, covering a diverse chemical space.

1

217 co-crystallized ligands with known activity

Perform protein-ligand complex alignment

2



Refine set

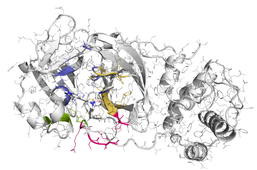
Ensure a proper pIC_{50} distribution

Discard ligands with protonation centers

Ensure similar positioning of flexible amino acids and loops

Ensure proper alignment of the M-pro binding site

3

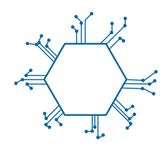


104 co-crystallized ligands

Input data:

- 3D structures (bioactive conformation) + atomic 3D hydrophobic descriptors
- Experimental activities (pIC_{50})

4



Train the PharmQSAR model

5

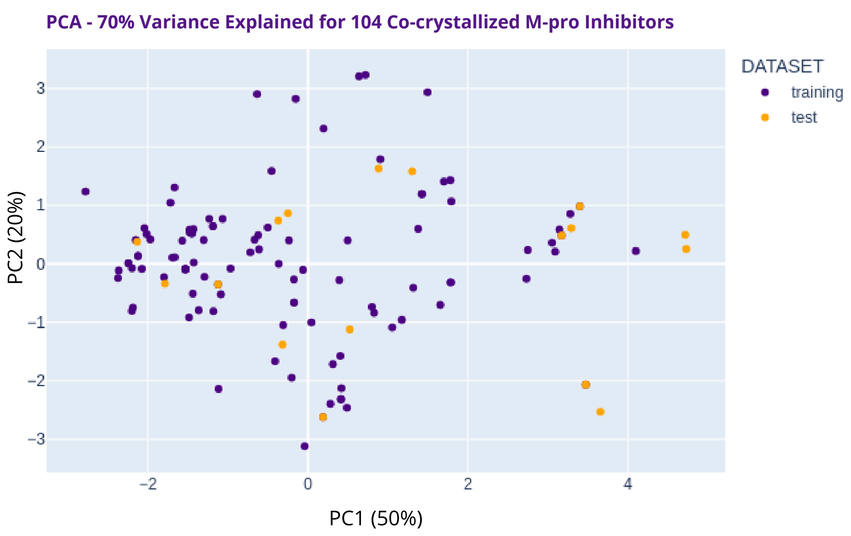
Validate

External set: 14 known inhibitors with unknown bioactive pose

Building the model

Known binding mode + Known activity

Chemical Space



The model was constructed using a **diverse dataset** of 104 co-crystallized M-pro inhibitors. The test set encompasses a **partially distinct chemical space** compared to the training set. The inhibitors exhibit an activity range from 4.5 to 8.5 pIC_{50}, covering all four key subpockets: S1, S2, S4, and S1'.

Best Performing Pharmacophore Models:

Q(ESP)+R³

QE + R³ stands for a pharmacophore obtained by combining QM-derived ESP (QE) atomic partial charges with the third power of the atomic radii (R³)

Field Description		Contribution	
Field (E)	**Field (NE)**	**ESP charges**	**R³**
ESP charges	R³	11.7%	88.3%

E = Electrostatic;
NE = Non electrostatic

HyPhar (*logP*+HB)

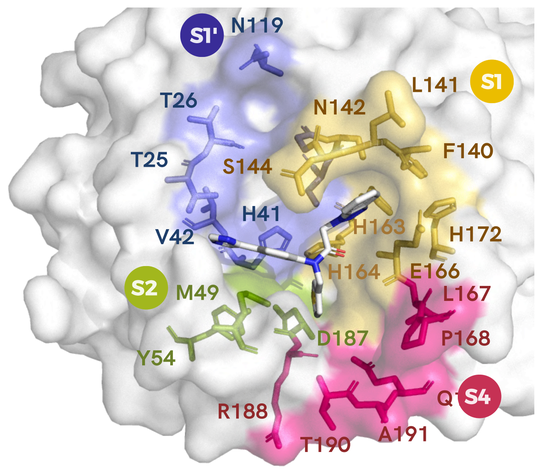
HyPhar parameters: 3D hydrophobicity pattern of a molecule (*logP*) with HB donor (HBD) and acceptor (HBA) descriptors

Field Description		Contribution		
Field (NE)	**Field 2**	**Hydrophobic**	**HBA**	**HBD**
logPtot	HBD/A	61.3%	28.7%	10.0%

NE = Non electrostatic;
HB = Hydrogen Bond

The pharmacophore models were built using the standard PLS algorithm and subjected to a leave-one-out (LOO) cross-validation (Q^2) to identify the optimal number of latent variables.

M-pro: ligand-protein interactions



PDB ID: 7LMF

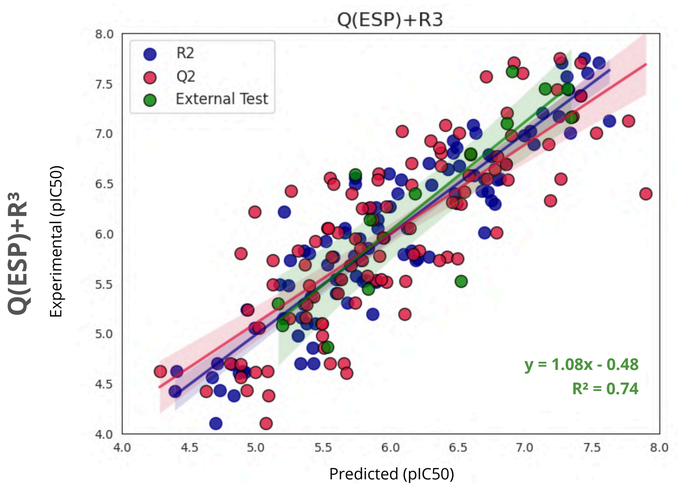
HyPhar (Hydrophobic Pharmacophore)

$$\log P_X = \sum_{i=1}^{N} \log P_{X,i} = \sum_{i=1}^{N} -\frac{\Delta G^{o/w}_{X,i}}{2.303RT} \quad (X:\ ele,\ cav)$$

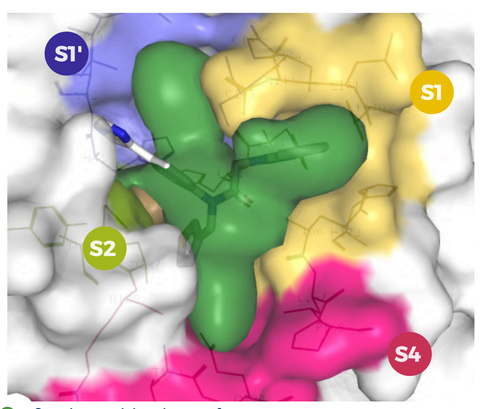
Testing the model

Unknown binding mode + Known activity

QSAR

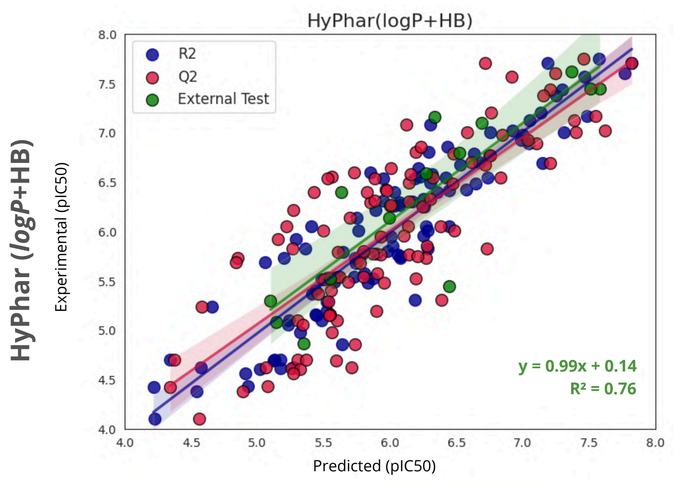


Isocontour Maps

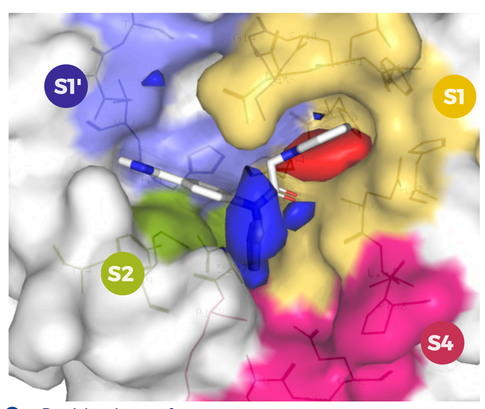


Steric positive isosurf.
Steric negative isosurf.

QSAR



Isocontour Maps

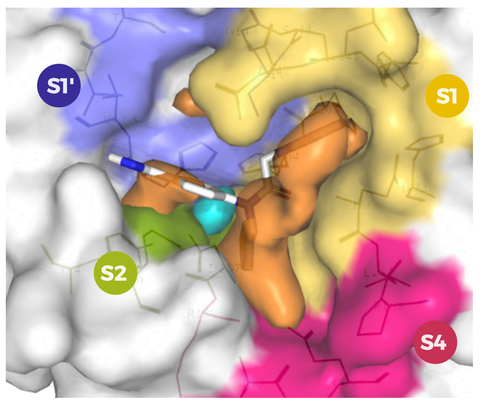


HBA

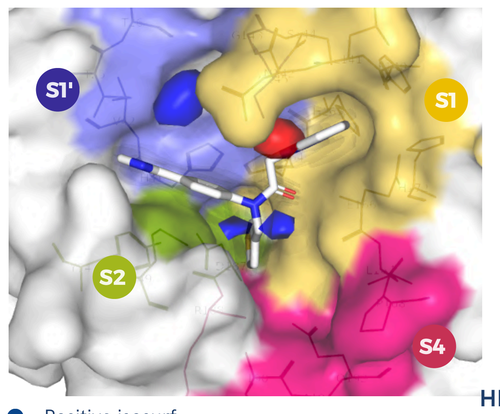
Positive isosurf.
Negative isosurf.

Both QSAR models, Q(ESP)+R³ and Hyphar (*logP*+HB), accurately predict the pIC_{50} activity of known SARS-CoV-2 M-pro inhibitors. The Q(ESP)+R³ model shows an 88% steric contribution, with the isosurface (green) highlighting the importance of occupying the S1 and S1' subpockets, while filling the S2 subpocket is less critical (brown).

Similarly, the Hyphar model demonstrates comparable contributions across its fields. The *logP* map (orange) analysis indicates that hydrophobic groups, such as aromatic rings, favor the S1 and S4 subpockets. In contrast, H-bond acceptor (HBA) groups, such as carbonyls or pyridines, contribute favorably to the S2 subpocket (blue). Additionally, in the S1' subpocket, H-bond donors (HBD), such as amino groups, are associated with higher activity.



logP positive isosurf.
logP negative isosurf.



HBD

Positive isosurf.
Negative isosurf.

Conclusions

- The Q(ESP)+R³ and HyPhar (*logP*+HB) models allow for the prediction of pIC50 values of M-pro inhibitors and provide a better understanding of the components responsible for their bioactivity.
- The use of crystallographic poses in 3D-QSAR modeling enhances the accuracy and robustness of the predictions by preserving the bioactive conformations of the ligands.
- Ensuring the correct ligand conformation of the test set is crucial for reliable activity prediction.

This work was supported by the project PID2022-138327OB-I00 financed by MCIN/AEI/10.13039/501100011033/FEDER, UE