NOVA SCHOOL OF SCIENCE & TECHNOLOGY

LAQV requimte | Laboratório Associado

UCIBIO @ REQUIMTE

# COMPUTATIONAL TARGETING OF IMMUNE AND GLYCO-IMMUNE CHECKPOINTS IN CAR T-CELL THERAPY

Patrícia Isabel Saúde Sobral,
PhD Candidate in Sustainable Chemistry (FCT NOVA & Stemmatters)

Supervisors:
Dr. Florbela Pereira (LAQV, Cheminformatics Lab); Dr. Paula Videira (UCIBIO, Glycoimmunology Group) and Dr. Rachel Paterson (Stemmatters)

**BCADD Symposium - 19 October 2026**

# Introduction

- **Cancer** is a leading global cause of death, highlighting the need for better therapies, with **CAR T-cell therapy** emerging as a promising new cancer treatment approach.
- **CAR T-cell therapy** reprograms T-cells to target and kill cancer cells.
- Its efficacy is still limited by the immunosuppressive tumor microenvironment (TME).
    - Tumors evade immune attack through inhibitory pathways, including glycan-mediated immune suppression and immune checkpoint signaling.
        - **Siglecs** are inhibitory receptors on immune cells that bind tumor-associated **sialic acids**, reducing T-cell activation and cytotoxicity.

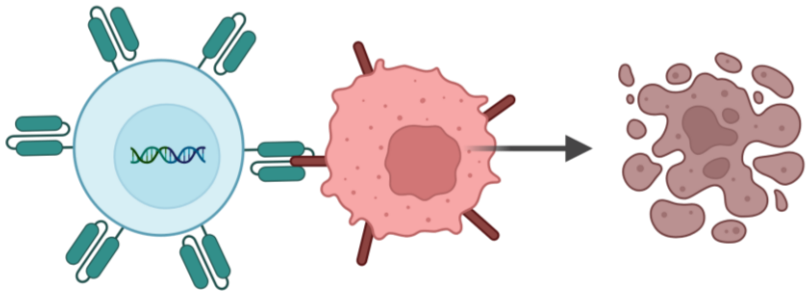


Fig.1. CAR T-cell therapy.

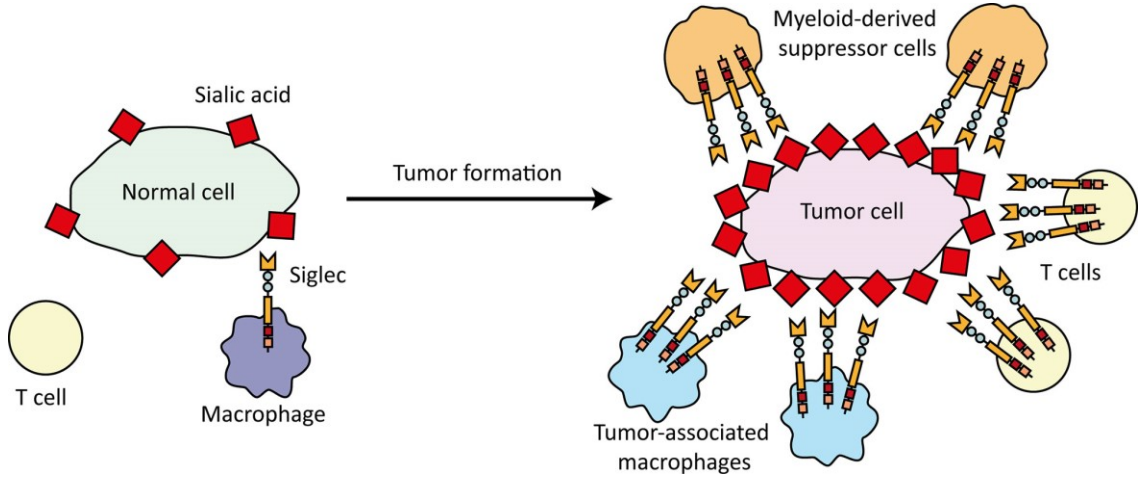



Fig. 2. Schematic overview of changes in the Siglec-sialoglycan axis upon tumor formation. *Adapted from van Houtum & Adema, G.J. (2021). Siglec Signaling in the Tumor Microenvironment. Frontiers in Immunology, 12, 790317.*

# Introduction

• Another key suppressive axis is **PD-1/PD-L1**, where tumor PD-L1 binding to PD-1 on T-cells leads to T-cell exhaustion and impaired antitumor activity.

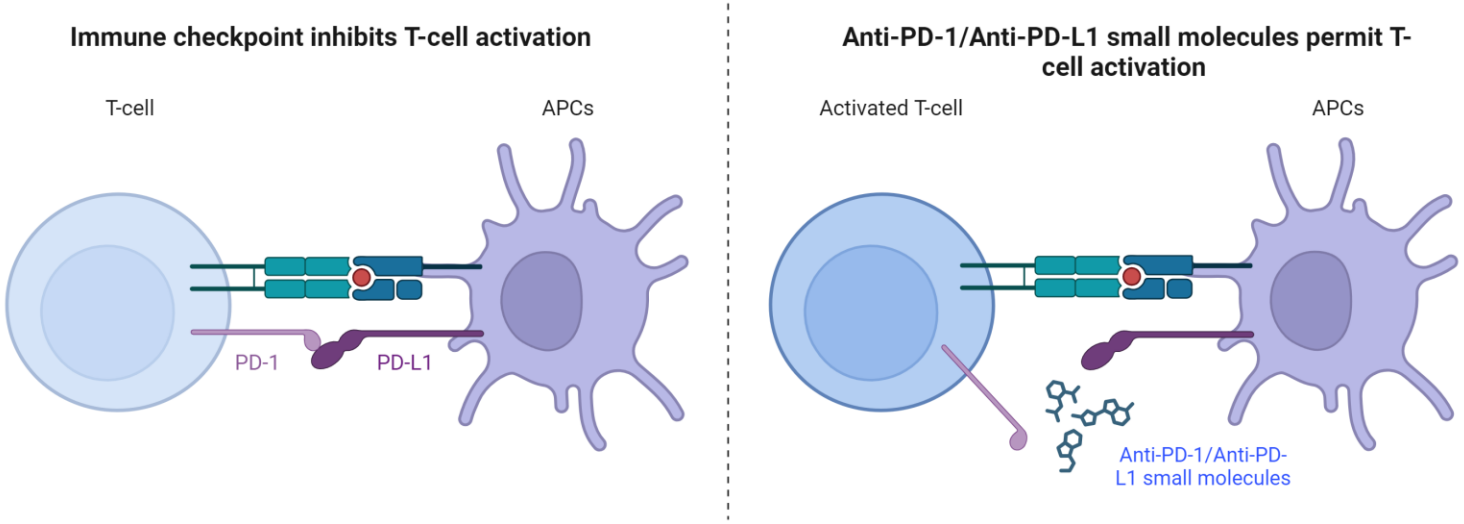



Fig.3. Targeting PD-1/PD-L1 with small molecules inhibitors.

# Objectives: Overview

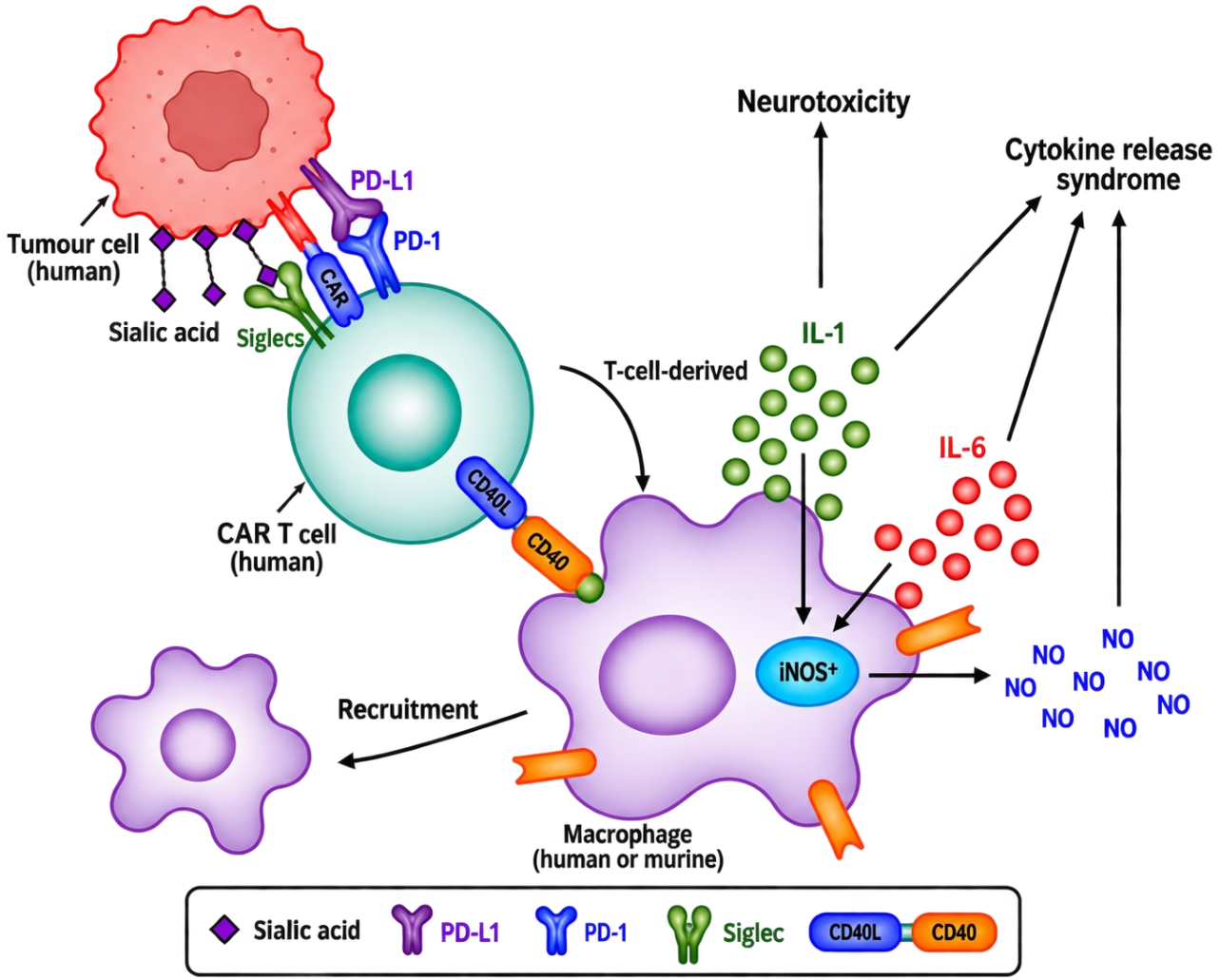
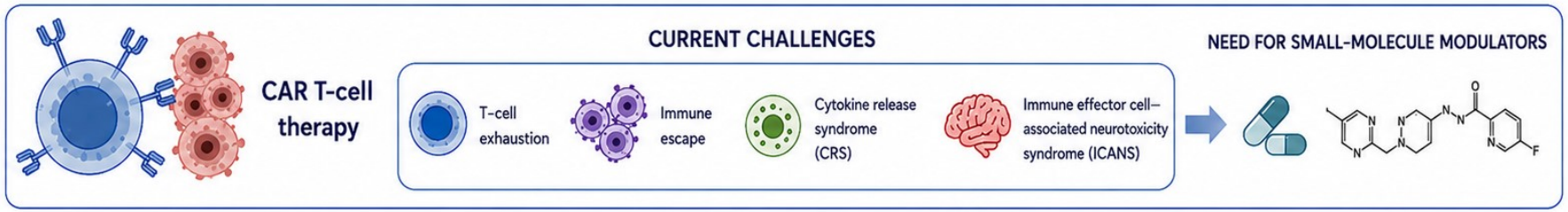



Fig. 4. Immune checkpoint signalling and inflammatory pathways in CAR T-cell therapy.

Our research focuses on three axes: **PD-1/PD-L1** and **Siglec-7/9** to reduce inhibitory signalling and CAR T-cell exhaustion, and **IL-1R1** to limit inflammation-related toxicity.

CAR T MATTERS

# Objectives: Overview




AI-DRIVEN QSAR PLATFORM – ONE STRATEGY, MULTIPLE TARGETS

**1. PD-1/PD-L1 AXIS**

*Novel compound discovery*

- ChemoDOTS analogue library (derived from parent compounds)
- Previously developed QSAR model
- Virtual screening
- **Novel hit candidates**

**Aim**

Discover novel small-molecule inhibitors to restore T-cell activation and enhance CAR T-cell efficacy.

**2. SIGLEC-7/9 AXIS**

*Drug repurposing*

- Bioactivity datasets (ChEMBL • PubChem • Literature)
- QSAR model
- Virtual screening
- **Repurposed hit candidates**

**Aim**

Identify repurposed small-molecule modulators to restore anti-tumour immunity through Siglec blockade.

**3. INFLAMMATORY AXIS (IL-1R1 / IL-6)**

*Drug repurposing*

- Bioactivity datasets (ChEMBL • PubChem • Literature)
- QSAR model
- Virtual screening
- **Repurposed hit candidates**

**Aim**

Identify repurposed small-molecule modulators targeting IL-1R1 and IL-6 signalling to mitigate CAR T-cell–associated toxicities.

**OVERALL GOAL → Enhance CAR T-cell activity and reduce related toxicities**

# PD-1/PD-L1 axis

- Programmed Death-1 (PD-1) is an immune checkpoint receptor expressed on T cells.
- PD-L1 (Programmed Death-Ligand 1) is its primary ligand, often expressed on tumor cells and immune cells.
- The PD-1/PD-L1 interaction inhibits T cell activation, leading to immune suppression.

→ **Targeting PD-1/PD-L1 with immune checkpoint inhibitors has revolutionized cancer immunotherapy.**

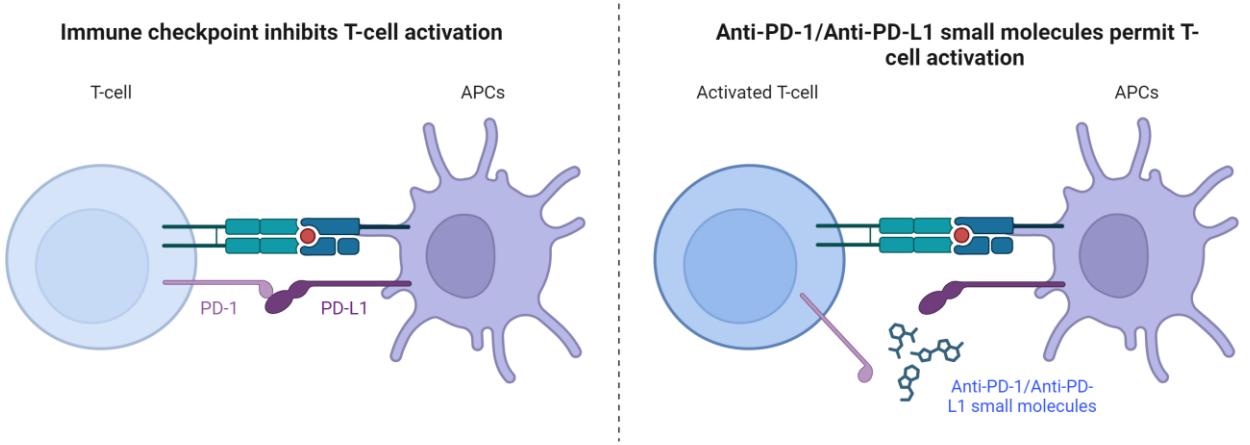



Fig.3. Targeting PD-1/PD-L1 with small molecules inhibitors.

# PD-1/PD-L1 axis

The **DC Matters** project aimed to improve cancer immunotherapy by combining **dendritic cell (DC) vaccines** with **small-molecule immune checkpoint inhibitors of PD-1/PD-L1.**



/2025 4:30:00 AM.
n 3.0 Unported Licence.

RSC Advances

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *RSC Adv.*, 2025, **15**, 2298

**Advancements in drug discovery: integrating CADD tools and drug repurposing for PD-1/PD-L1 axis inhibition†**

Patrícia S. Sobral,[ab] Tiago Carvalho,[bcd] Shiva Izadi,[e] Alexandra Castilho,[e] Zélia Silva, *[bcd] Paula A. Videira*[bcd] and Florbela Pereira *[a]



International Journal of *Molecular Sciences*

MDPI

*Review*

**Computational Approaches Drive Developments in Immune-Oncology Therapies for PD-1/PD-L1 Immune Checkpoint Inhibitors**


**Patrícia S. Sobral [1,2,3], Vanessa C. C. Luz [2,3], João M. G. C. F. Almeida [2], Paula A. Videira [2,3] and Florbela Pereira [1,*]**

1 LAQV and REQUIMTE, Department of Chemistry, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
2 UCIBIO, Applied Molecular Biosciences Unit, Department of Life Sciences, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
3 Associate Laboratory i4HB—Institute for Health and Bioeconomy, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
* Correspondence: florbela.pereira@fct.unl.pt

# PD-1/PD-L1 axis

**Previous work - DC Matters**

**Dataset:** 29,197 molecules retrieved from ChEMBL, PubChem and literature

- Actives: 431 | Inactives: 28,766

**Split into:**

- Training: 28,319 (403 actives)
- Test 1: 878 (14 actives)
- Test 2: 1,000 (14 actives)

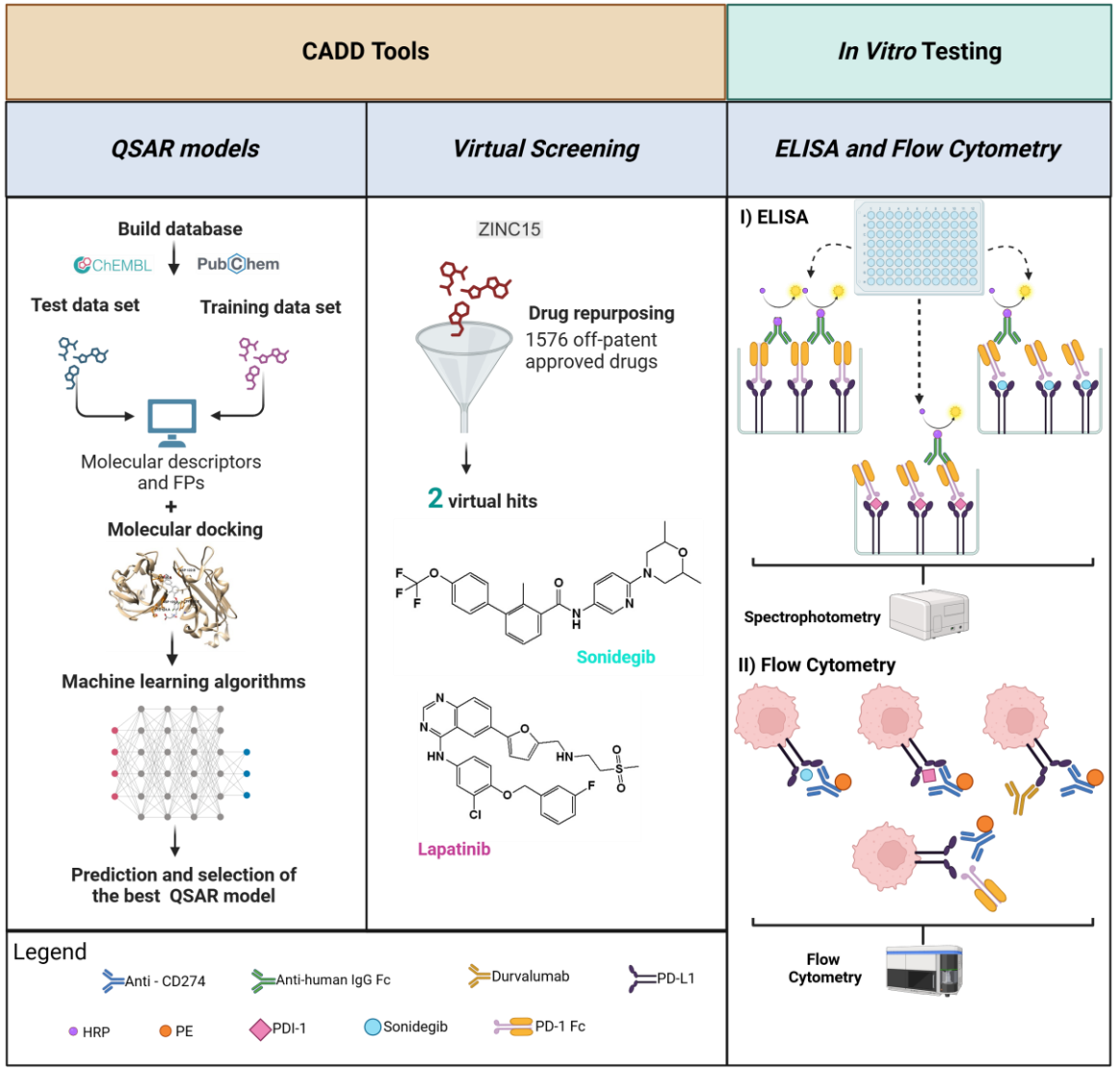
**Descriptors & FPs:**

- RDKit: 242 1D/2D + 3D descriptors
- Fingerprints: MACCS (166), Morgan (1024), RDKit (2048)
- 3D Grids: Electrostatic and voxel-based **(GUIDEMOL)**

**Docking**

- Autodock Vina

Fig.6. DC Matters project.



# PD-1/PD-L1 axis

**Previous work - DC Matters**

- **Choosing the best QSAR Classification model**
  - **Random Forest**

Table 1. Exploration of two models containing 3D grid descriptors for building PD-L1 classification models using the RF algorithm for the training set in an OOB estimation.The best model is highlighted in bold.

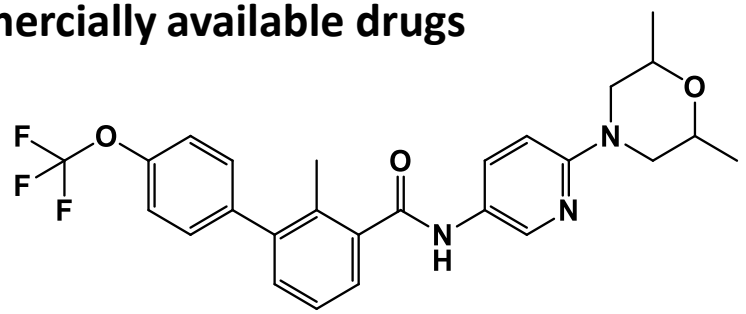
| Descriptors | # | SE[a] | SP[b] | $Q^c$ | MCC[d] |
|---|---|---|---|---|---|
| Morgan FP + RDKit 3D + grid of voxel – Log*P* | 3008 | 0.935 | 0.999 | 0.999 | 0.948 |
| Morgan FP + RDKit 3D + grid of voxel – MR | | **0.938** | **1.000** | **0.999** | **0.954** |

[a] Sensitivity, the ratio of true positive to the sum of true positive and false positive. [b] Specificity, the ratio of true negative to the sum of true negative and false negative. [c] Overall predictive accuracy, the ratio of the sum of true positive and true negative to the sum of true positive, true negative, false positive and false negative. [d] Matthews correlation coefficient.

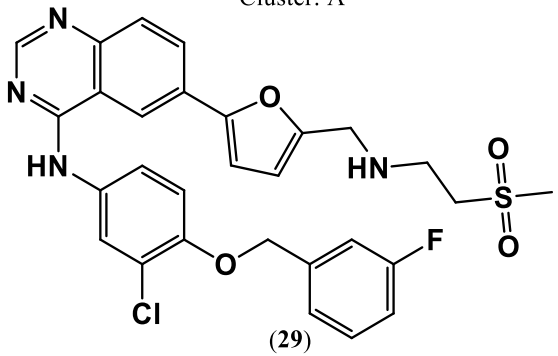
# PD-1/PD-L1 axis

**Previous work - DC Matters**

- **Virtual screening of off patent and commercially available drugs**



(**28**)
Predicted as active
Probability of being active: 0.72
Cluster: A



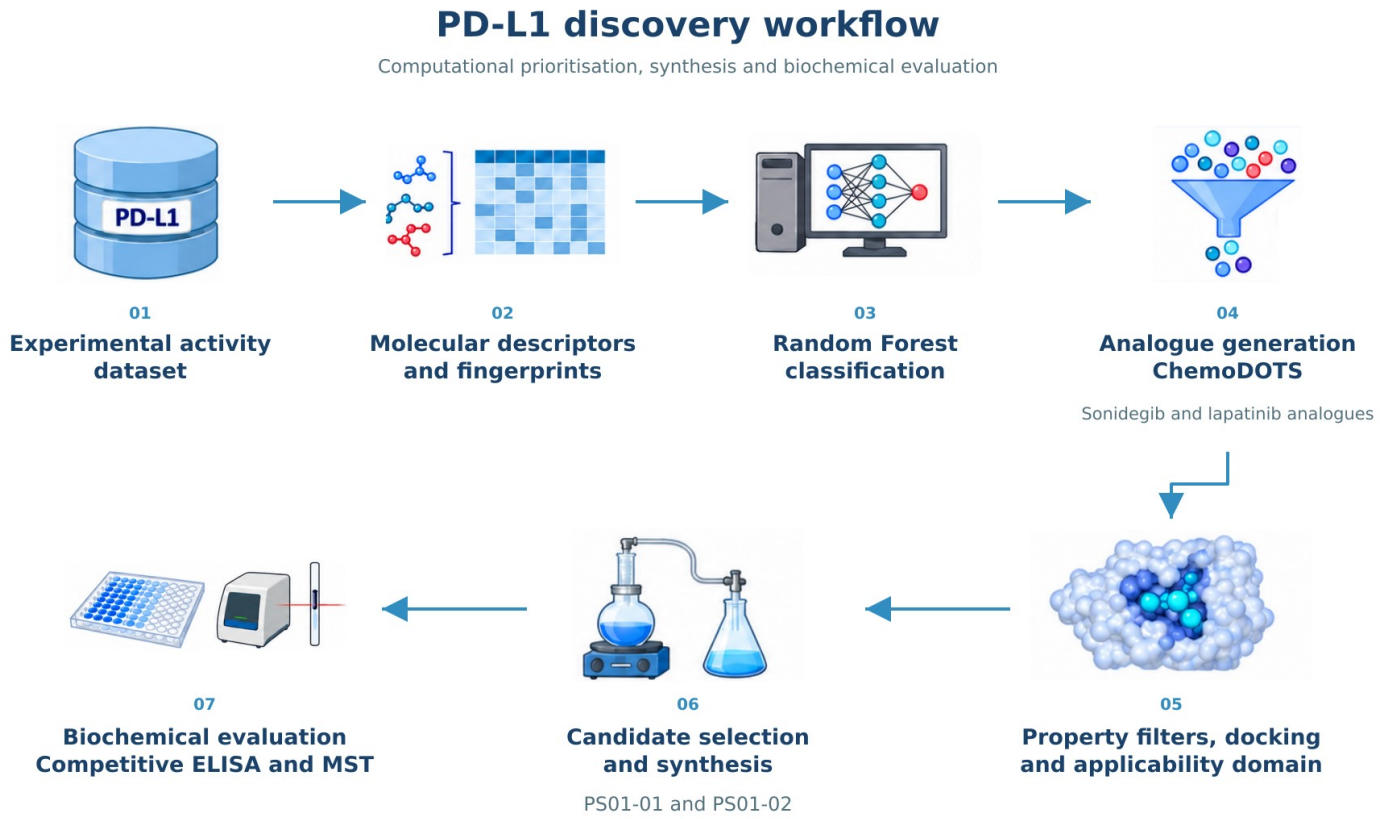
(**29**)
Predicted as active
Probability of being active: 0.64
Cluster: A

Fig. 7. Chemical structures of lead-like PD-L1 inhibitors: sonidegib (28) and lapatinib (29).

# PD-1/PD-L1 axis

**Previous work - DC Matters**

- Unfortunately, **Lapatinib suffers from poor solubility, while Sonidegib exhibits limited potency.**




➢ **CAR T-Matters goal**

**Identification of Sonidegib and Lapatinib analogues as potential PD-L1 inhibitors, by applying the QSAR Classification model developed during the DC-Matters project to a dataset of new compounds.**

# PD-1/PD-L1 axis

**Tasks:**

**1. Virtual Library Construction**

- Use ChemoDOTS to create a PD-L1 inhibitor library.




**Lapatinib** → **TOTAL OF MOLECULES: 4.027.039**

**Sonidegib** → **TOTAL MOLECULES: 2.059.809**

Adding filters (Lipinski's rule of five*, Veber, Goose, Muegge, LogS ≥ -4, nº of rings and QED (Quantitative Estimate of Drug-likeness))

**Lapatinib** → **TOTAL OF MOLECULES: 1000**

**Sonidegib** → **TOTAL MOLECULES: 1064**

*with the exception of MW ≤ 550 instead of 500.

# PD-1/PD-L1 axis

**Tasks:**

**2. Virtual Screening and Selection**

- Apply QSAR model (DC Matters) and Random Forest for screening.

**Results**

- **120 analogues of Sonidegib predicted active (>72%)**
- **0 analogues of Lapatinib predicted active**
  - Probably due to the drug-like filters applied to ensure higher solubility than Lapatinib (the most similar analogue has only 0.193 tanimoto similarity with Lapatinib)

# PD-1/PD-L1 axis

## Tasks:

### 2. Virtual Screening and Selection

Table 2. Top 10 prioritized PD-1/PD-L1 inhibitor candidates based on QSAR activity probability, docking score, and protein–ligand interaction profile.

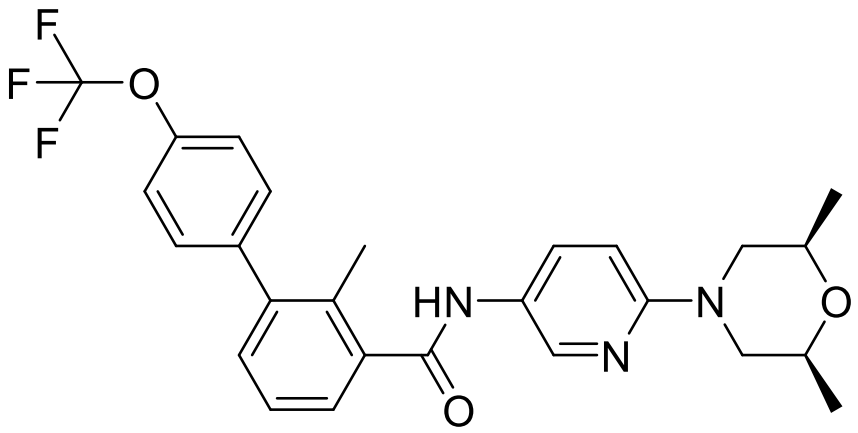
| # | Name | $\Delta G_B$[1] | Interaction | | | Activity probability |
|---|---|---|---|---|---|---|
| | | | Hydrophobic residues | H-bond residues | p- Stacking residues | |
| 1 | **1957** | **-11.1** | Ile54[3], Ile54[2], Tyr56[3], Tyr56[2], Met115[3], Met115[2], Ala121[3], Ala121[2], Tyr123[2] (equal to BMS) | Gln66[2] | | **0.86** |
| 2 | **2013** | **-11.5** | Ile54[3], Tyr56[3], Tyr56[2], Met115[3], Met115[2], Ala121[3], Ala121[2], Tyr123[2] (equal to BMS), Lys124[2] | - | - | **0.86** |
| 3 | **1956** | **-11.1** | Ile54[3], Tyr56[3], Tyr56[2], Met115[3], Met115[2], Ala121[3], Ala121[2], Tyr123[2] (equal to BMS) | Gln66[2] | - | 0.84 |
| 4 | **1825** | **-12.5** | Ile54[3], Tyr56[3], Tyr56[2], Met115[3], Met115[2], Ala121[3], Ala121[2], Tyr123[2] (equal to BMS) | Gln66[2]; Asn63[3] | | 0.83 |
| 5 | **1792** | **-11** | Ile54[3], Tyr56[3], Tyr56[2], Met115[3], Met115[2], Ala121[3], Ala121[2] (HB), Tyr123[2] | Phe19[2]; Ala121[2]; Gln66[2] | | 0.80 |
| 6 | **1838** | **-10.3** | Ile54[3], Ile54[2], Tyr56[3], Tyr56[2], Met115[3], Met115[2], Ala121[3], Ala121[2], Tyr123[2] (equal to BMS) | Asn63[3] | | 0.80 |
| 7 | **1910** | **-9.9** | Ile54[3], Ile54[2], Tyr56[3], Tyr56[2], Met115[3], Met115[2], Ala121[3], Ala121[2], Tyr123[2] (equal to BMS) | Gln66[2] | | 0.80 |
| 8 | **1959** | **-11.5** | Ile54[3], Ile54[2], Tyr56[3], Tyr56[2], Met115[3], Met115[2], Ala121[3], Ala121[2], Tyr123[2] (equal to BMS) | Gln66[2] | | 0.80 |
| 9 | **1701** | **-12.9** | Ile54[3], Ile54[2], Tyr56[3], Tyr56[2], Met115[3], Met115[2], Ala121[3], Ala121[2], Tyr123[2] (equal to BMS) | Gln66[2]; Gln66[3] | | 0.79 |
| 10 | **1714** | **-10.2** | Ile54[3], Ile54[2], Tyr56[3], Tyr56[2], Met115[3], Met115[2], Ala121[3], Ala121[2], Tyr123[2], Tyr123[3] (equal BMS) | - | - | 0.79 |
| - | **Sonidegib** | **-11** | Ile54[2], Tyr56[2], Met115[2], Met115[3], Ala121[2], Ala121[3], Tyr123[3] | Gln66[2], Ala121[3] | --- | **0,72** |
| - | **BMS-200** | **-12** | **Ile54[3], Tyr56[3], Met115[3], Ala121[3], Ala121[2], Tyr123[2]** | Lys124[2] | Tyr56[3] | - |

# PD-1/PD-L1 axis

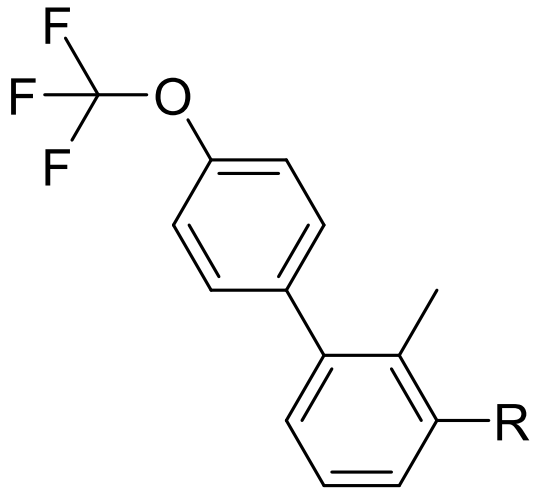
**Tasks:**

**2. Virtual Screening and Selection**

Sonidegib

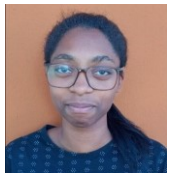



**NEW**
**SYNTHESIZED MOLECULES**
**PS01 and PS02**



# PD-1/PD-L1 axis



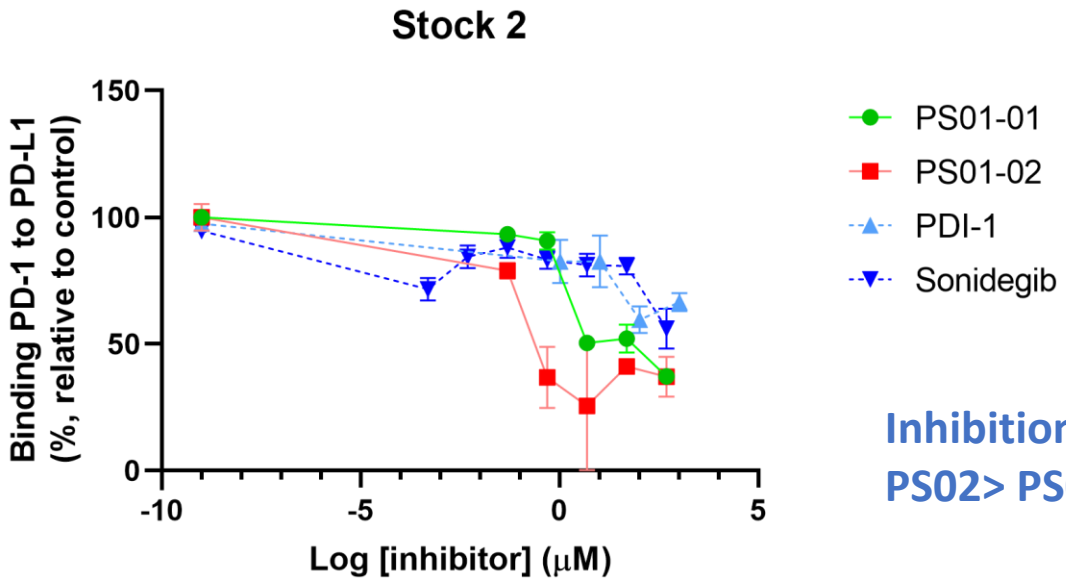


**Tasks:**

**3. Biological Activity Evaluation**

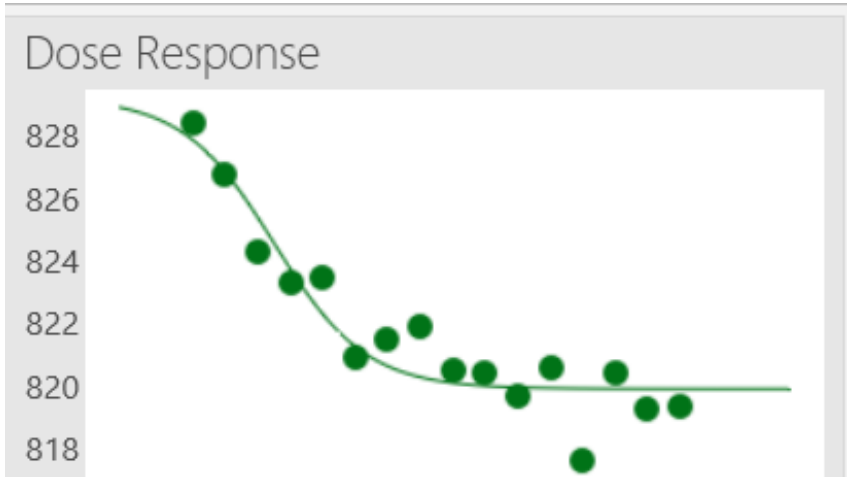
- Preliminary biochemical results

## Competitive ELISA




**Inhibition**
**PS02> PS01> Sonidegib**

## Microscale thermophoresis




**Reported Kd ≈ 200 nM**

Preliminary binding result

# Siglec-7/9 axis

- **Siglecs** are inhibitory receptors expressed on immune cells.
- They recognize **sialoglycans, or sialic acids**, which are often overexpressed on tumor cells, leading to immune suppression.

**Siglec-7**

- Found manly on NK cells, monocytes and **T cells**.
- Plays a role in tumor immune evasion.

**Siglec-9**

- Expressed mainly on neutrophils, monocytes and **T cells**.
- Implicated in cancer immune suppression and chronic inflammation.

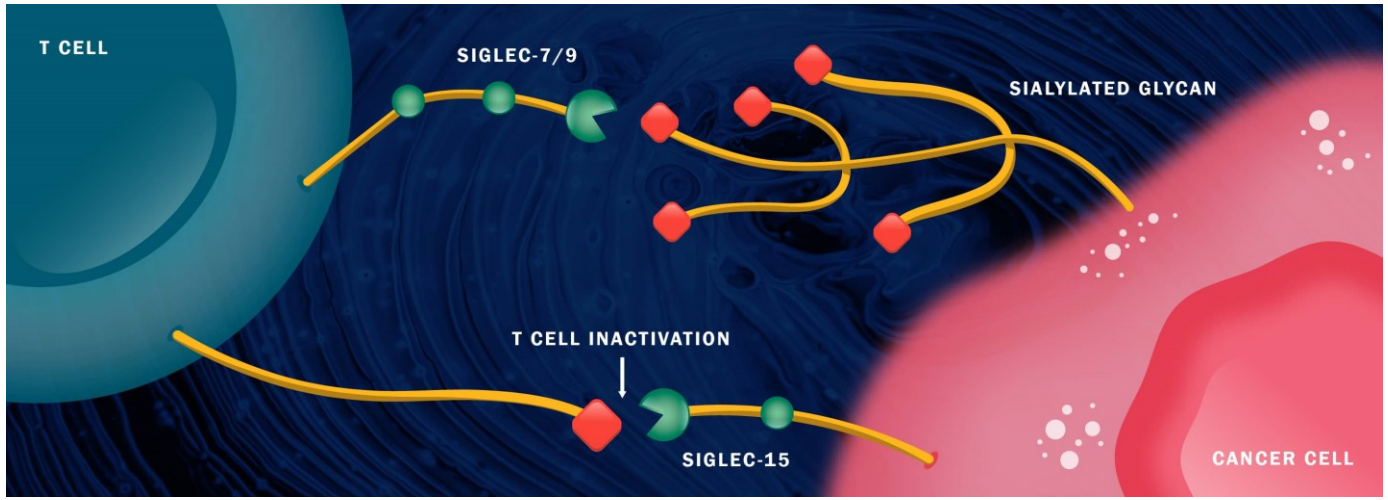



Figure 8. Inhibitory receptors Siglec-7 and Siglec-9 are expressed on T cells. Image from R&D Systems (https://www.rndsystems.com/products/siglecs-targets-immune-regulation-and-cancer-therapy).

# Siglec-7

## Methodology:

### 1st Step. Siglec-7 Inhibitors Database Construction with IC50 values

- Extract  small molecules from ChEMBL, PubChem, Reaxys, and literature.

**Unfortunately, only 91 compounds have been tested so far, and only 49 of them has IC50 values!**

- **Lack of experimental data is one of the biggest issues for machine learning.**
- **How can we try to overcome this lack of data?**

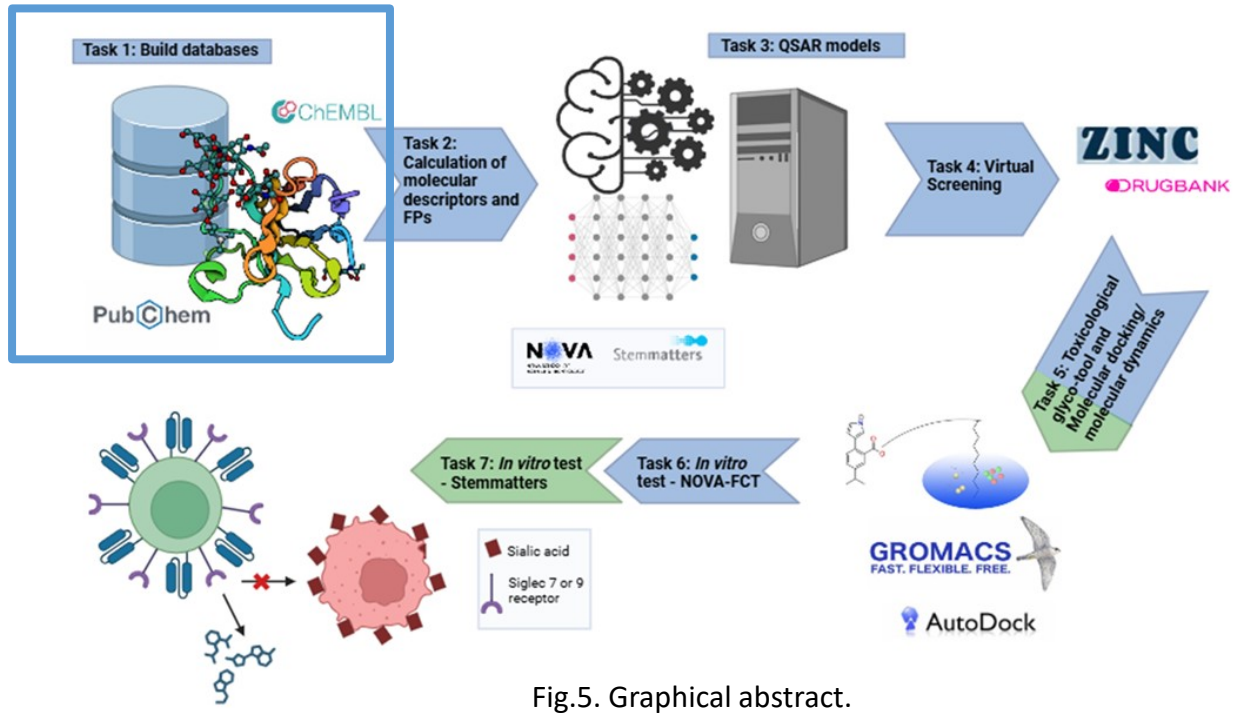



Fig.5. Graphical abstract.

# Siglec-7

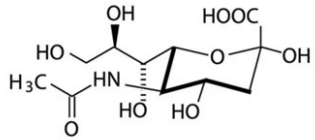
## Methodology:

### 1st Step. Siglec-7 Inhibitors Database Construction with IC50 value or/and percent inhibition

- Extract small molecules from ChEMBL. ✅

### 1st Step. Sialic acid analogues Database Construction

- Extract small molecules from PubChem using a Similarity Tanimoto tool.
  - **N-acetylneuraminic acid (Neu5Ac) -** The predominant sialic acid on human cell surfaces.




N-acetyl-α-neuraminic acid (Neu5Ac)

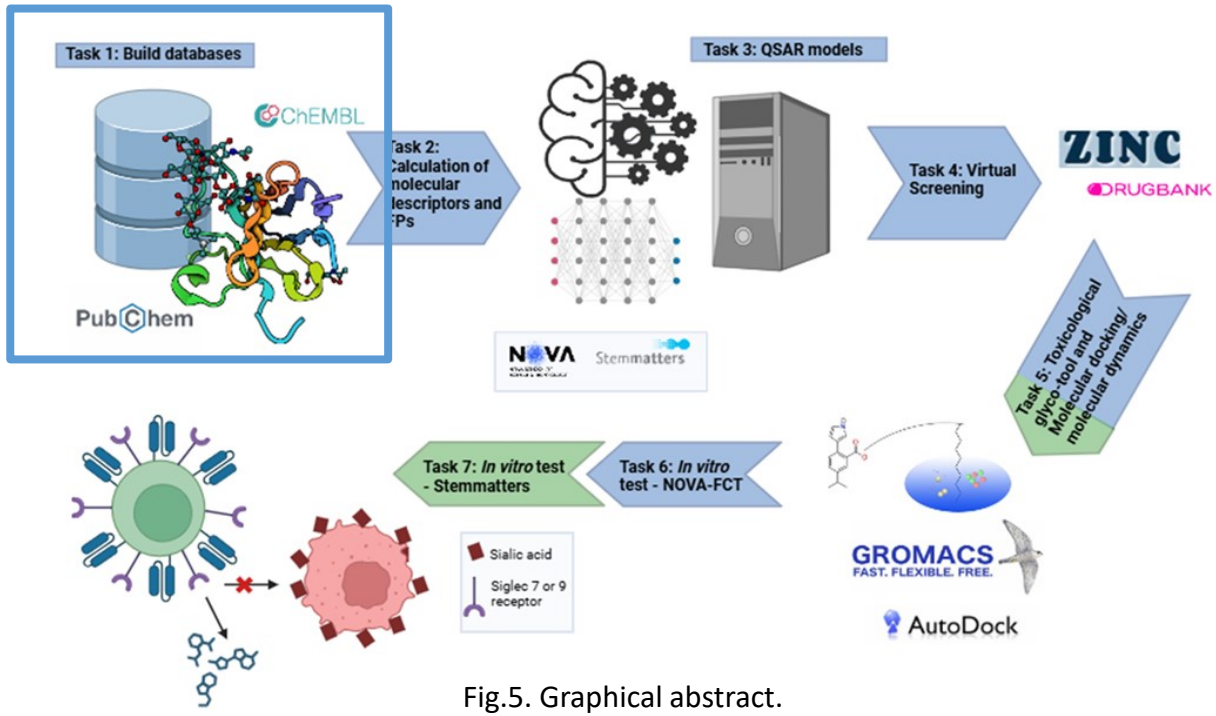


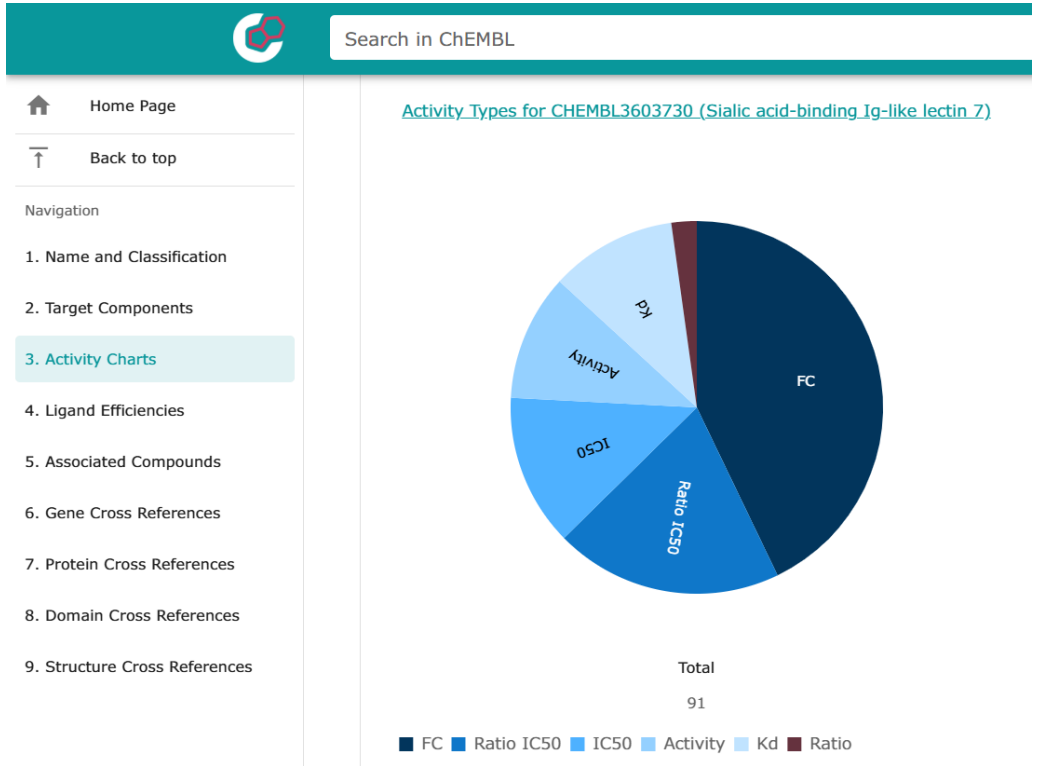

Fig.5. Graphical abstract.

# Siglec-7

## Methodology:

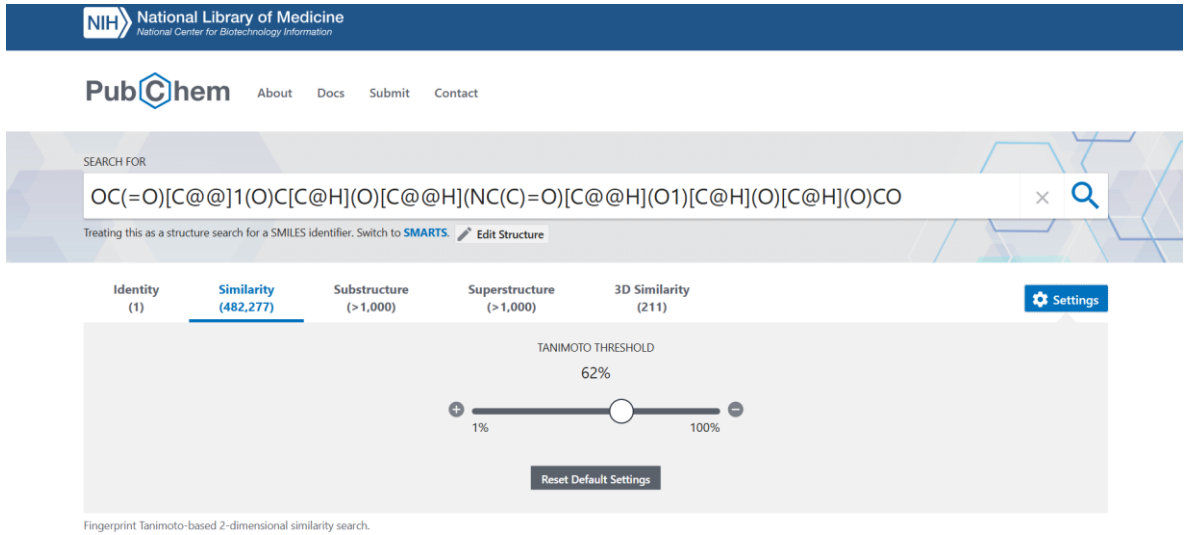
Siglec-7 ➡ **91 compounds**

- **This dataset – external validation test**




Sialic acid analogues ➡ **97,965 compounds** identified with ≥62% similarity to sialic acid and filtered using **Lipinski's Rule of Five**.

- **This dataset – our training (90%) and test set (10%)**



# Siglec-7

## Methodology:

**2nd Step. Standardize and generate optimized 3D structures**

- **Difficulties** ⚠: The generation of 3D structures was a bit difficult due to the specificities of carbohydrates.

**RDKit Advantages:**

•**Better conformer generation** using **ETKDG** (Experimental-Torsion Distance Geometry):

- Captures realistic ring puckering & torsions in **complex sugars**.

•**Stereochemistry handling** is more reliable:

- Accurate chiral center preservation, crucial for carbohydrates.

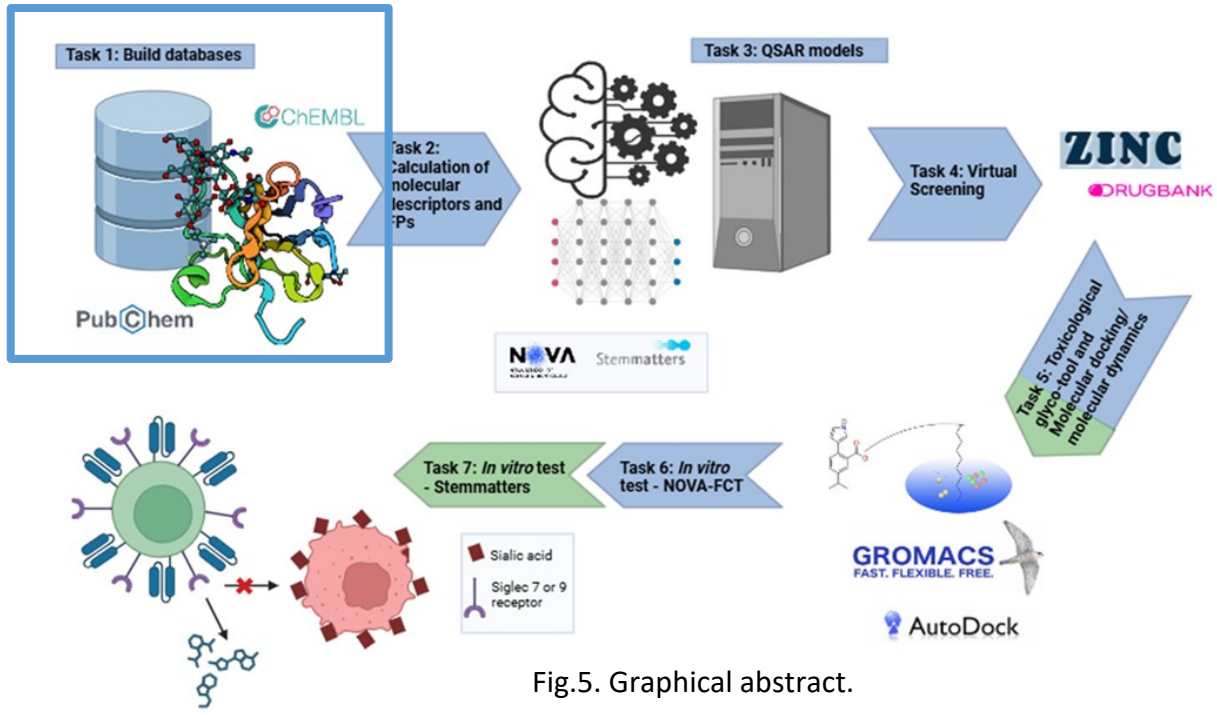



Fig.5. Graphical abstract.

# Siglec-7

## Methodology:

### 3rd Step: Performing docking against the treated structure of Siglec-7 using GlycoTorch (carbohydrates ⚠).

Docking of all molecules using **GlycoTorch** against Siglec-7 structure (2G5R).

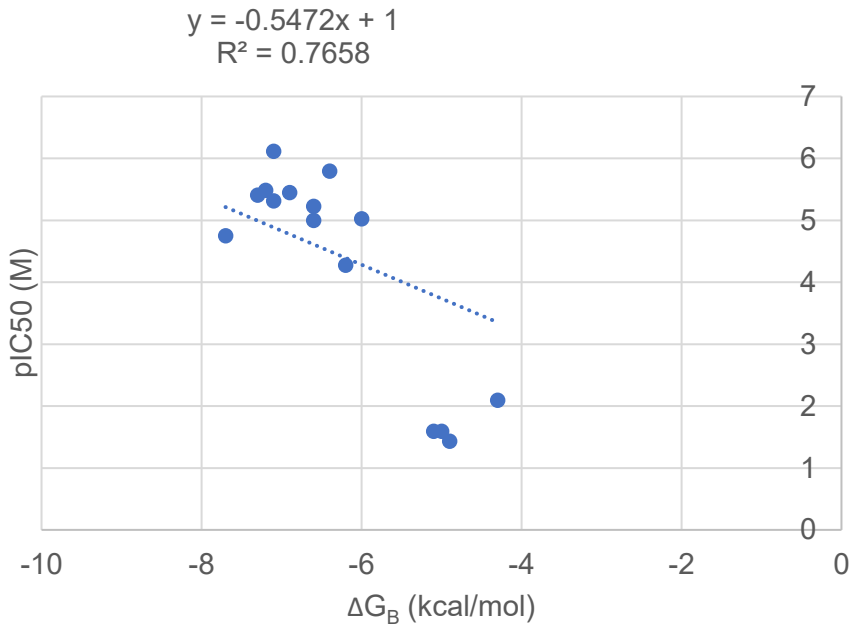
Table 2. Comparision of GlycoTorch and Autodock Vina.

| | GlycoTorch Vina | AutoDock Vina |
|---|---|---|
| Carbohydrate-specific logic | ✅ Yes | ❌ No |
| Glycosidic torsion handling | ✅ Optimized | ❌ Poor |
| Scoring for polar ligands | ✅ Improved for sugars | ❌ Biased for drug-like ligands |
| Ring conformer stability | ✅ Maintained | ❌ Often distorted |
| Use case | Sugars & glycomimetics | General small molecules |

# Siglec-7

## Methodology:

### 4th Step: Check if there is correlation between activity ($pIC_{50}$) and docking scores




- The correlation between the pIC50 values and docking scores of the dataset retrieved from ChEMBL was calculated:
  - **Pearson Correlation: –0.875104463**
- This strong negative correlation is consistent with the expected trend:
  - Higher pIC50 (stronger activity, lower IC50)
  - More negative docking score (better predicted binding affinity)

Therefore, as biological potency increases, the docking score becomes more negative, indicating good agreement between experimental activity and docking predictions.

✓ Calculate $pIC_{50}$ for the sialic acid analogues dataset

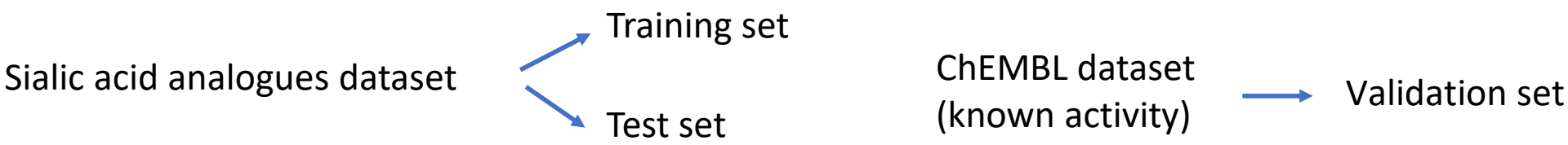
# Siglec-7

## Methodology:

### 5th Step: Partition of the training and test sets

- The sialic acid analogue dataset was divided into training and test sets using an automated Python script.
- Molecular descriptors relevant for sugars—especially sialic-acid derivatives—were calculated for all compounds.
- Based on these descriptors, the compounds were grouped into defined clusters.
- The training and test sets were then created by sampling proportionally from each cluster.

➢ This ensured that both sets were **balanced and representative** of the overall chemical space.



# Siglec-7

## Methodology:

### 6th Step: Calculate descriptors and building QSAR Regression models

- Several fingerprints and molecular descriptors were calculated.
- A first model using **lightGBM** was built with **Morgan fingerprints**.

| | |
|---|---|
| Mean absolute error (MAE): | 0.069500 |
| Mean squared error (MSE): | 0.009976 |
| Root mean squared error (RMSE): | 0.099882 |
| R square (R^2): | 0.921709 |

➢ **Training Set (10-fold Cross-Validation)**

- Low prediction error and high correlation during cross-validation.

| | |
|---|---|
| Mean absolute error (MAE): | 0.150359 |
| Mean squared error (MSE): | 0.037809 |
| Root mean squared error (RMSE): | 0.194445 |
| R square (R^2): | 0.710554 |

➢ **Test Set Performance**

- Shows **good generalization**, though with some performance drop compared to training.

**Overall**

- Cross-validation confirms **robust training performance**.
- Test set results indicate the model maintains **reasonable predictive power** on unseen compounds.
- Performance gap suggests **some overfitting**, but results remain strong for a chemical descriptor-based regression task.

# Siglec-7

## Methodology:

**6th Step: Calculate descriptors and building QSAR models**

- **PROBLEM** ⚠: The predictive performance on the external validation set (91 experimentally tested compounds) was very poor.
  - While docking score correlates with activity in the 91-compound set, this may not apply to all sialic acid derivatives, as **docking is less reliable for carbohydrates due to their flexibility and polar interactions.**

**NEW APPROACH – QSAR Classification model**

- Integrate **docking scores with residue-level interaction data** and **evaluate the correlation between essential binding residues and biological activity in the 91-compound dataset.**
  - **PYMOL and PLIP**

# Siglec-7

## Methodology:

### 6th Step: Calculate descriptors and building QSAR models

**NEW APPROACH – QSAR Classification model**

- **Random Forest (R and scikit-learn)** and **LightGBM** were evaluated.
- As **Random Forest** showed consistently robust performance, the scikit-learn implementation was selected for further studies because of its computational efficiency, allowing rapid evaluation of multiple molecular descriptor sets.

# Siglec-7

## Methodology:

### 6th Step: Calculate descriptors and building QSAR models

Table 3. Performance of Random Forest classification models implemented in Scikit-learn using balanced class weights across different molecular descriptor sets.

| Descriptor | Sensitivity | Specificity | Accuracy (Q) | MCC |
|---|---|---|---|---|
| **RDKit 2D** | **0.8261** | **0.9673** | **0.9660** | **0.3900** |
| **Mordred 2D** | 0.8152 | 0.9691 | 0.9676 | 0.3943 |
| **RDKit 3D** | 0.5543 | 0.9280 | 0.9245 | 0.1750 |
| **Mordred 3D** | 0.7283 | 0.9623 | 0.9601 | 0.3240 |
| **Morgan** | 0.6087 | 0.9817 | 0.9782 | 0.3729 |
| **Sialic Acid** | 0.8261 | 0.9602 | 0.9590 | 0.3578 |
| **Feature Morgan** | 0.5870 | 0.9795 | 0.9758 | 0.3445 |
| **MACCS** | 0.6739 | 0.9655 | 0.9627 | 0.3128 |
| **Avalon** | 0.6196 | 0.9801 | 0.9767 | 0.3668 |
| **Atom Pair** | 0.6739 | 0.9788 | 0.9759 | 0.3859 |
| **MAP4** | 0.5109 | 0.9907 | 0.9862 | 0.4120 |
| **RDKit2D + RDKit3D + Mordred3D** | **0.8152** | **0.9666** | **0.9652** | **0.3815** |
| **Morgan + RDKit3D + Mordred3D** | 0.7500 | 0.9685 | 0.9664 | 0.3612 |
| **MAP4 + RDKit3D + Mordred3D** | 0.6848 | 0.9706 | 0.9679 | 0.3415 |
| **Mordred2D + Mordred3D + RDKit3D** | 0.7826 | 0.9697 | 0.9679 | 0.3826 |

# Siglec-7

## Methodology:

The QSAR models did not provide sufficiently convincing external validation results. We therefore took a step back to reassess the docking protocol before further model development.

**01**

### Validate redocking

Rescoring and redocking studies using smina, vinardo and GNINA.

**02**

### Investigate binding-site waters

Compare docking with and without crystallographic or predicted water molecules.

**03**

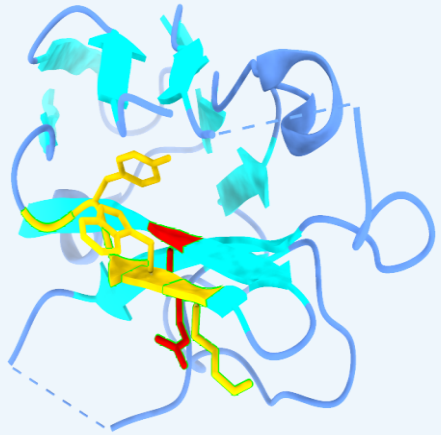
### Refine the protocol

Examine ligand protonation, receptor preparation and scoring functions.

**Goal: establish a reliable docking workflow before revisiting QSAR modelling and compound prioritisation.**

# Siglec-9

## Siglec-9 | Planned homology-modelling strategy



**Siglec-7 · 2G5R**

Experimental template

**Sequence alignment**

↓

**MODELLER**

Homology modelling



**Siglec-9 model**

To be generated and evaluated

Image: AlphaFold reference, not MODELLER output

**Structural validation · AlphaFold comparison · Binding-pocket assessment**

Future application: molecular docking and compound prioritisation

# Conclusions and next steps

## PD-L1 | Experimental progress

Two sonidegib analogues synthesised and evaluated by ELISA and MST.

**PS01-02 > PS01-01 > Sonidegib**

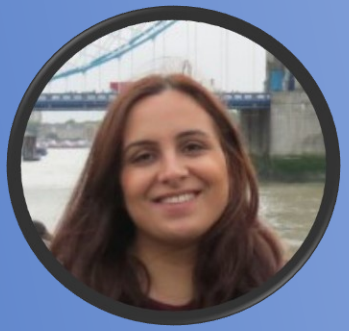
## Siglec-7 | Method refinement

Limited experimental data and poor external QSAR validation motivated a reassessment of the docking protocol.

**Protonation • Waters • Scoring**

### NEXT STEPS

**Refine docking validation** → **Prioritise Siglec-7/9 hits** → **Biochemical & cellular testing**



# Thank you!

This research was funded by the Fundação para a Ciência e a Tecnologia (FCT), Portugal, through the research units LAQV-REQUIMTE (UIDB/50006/2020), UCIBIO (UIDB/04378/2020 and UIDP/04378/2020) and the Associate Laboratory i4HB (LA/P/0140/2020); by the European Commission through the GLYCOTwinning project (Grant Agreement No. 101079417); by the DCMatters project (NORTE-01-0247-FEDER-047212; AVISO No. 17/SI/2019, Ref. 47212); by the CAR T-Matters project (COMPETE2030-FEDER-01477200); and by FCT through the PhD fellowship 2024.04128.BDANA.