CBR Teaching Questions and Answers

Introduction to Bioinformatics

Target: MCL1
PDB ID: 4HW3
Description of the MCL1 target

Discovery of Potent Myeloid Cell Leukemia 1 (Mcl-1) Inhibitors Using Fragment-Based Methods and Structure-Based Design

Myeloid cell leukemia 1 (MCL1) is a protein that is overexpressed in many types of cancer. MCL1 promotes the survival of tumor cells and results in these cells avoiding apoptosis. Targeting MCL1 with inhibitors could therefore lead to potential treatments for a variety of cancer types.

In recent years many MCL1 inhibitors have been developed. Some MCL1 inhibitors are in clinical trials, whereas others are already being used at the current time. Examples are: Maritoclax, VU661013, AZD5991, S63845, Motixafortide, (R)-(-)-Gossypol acetic acid, Sabutoclax, TW-37 and Gambogic acid.

The MCL1 protein is ecoded by the MCL1 gene. The protein is part of the Bcl-2 family. The Bcl-2 family is important in the regulation of apoptosis at the mitochondrion. MCL1 has two isoforms that are created through alternative splicing. Isoform 1 enhances cell survival by inhibiting apoptosis and isoform 2 promotes cell death by promoting apoptosis. MCL1 is a very important protein in the Bcl-2 family, since the loss of MCL1 gene has shown to result in embryonic death. Research has indicated that MCL1 interacts with: BAK1, BCL2L11, BID, BAD, DAD1, PMAIP1, PCNA, TCTP, and TNKS.

Isoform 1:
UNIPROT ID: Q07820
Sequence length: 350
Sequence:MFGLKRNAVIGLNLYCGGAGLGAGSGGATRPGGRLLATEKEASARREIGGGEAGAVIGGSAGASPPST
LTPDSRRVARPPPIGAEVPDVTATPARLLFFAPTRRAAPLEEMEAPAADAIMSPEEELDGYEPEPLGKRPAVLPLLEL
VGESGNNTSTDGSLPSTPPPAEEEEDELYRQSLEIISRYLREQATGAKDTKPMGRSGATSRKALETLRRVGDGVQRN
HETAFQGMLRKLDIKNEDDVKSLSRVMIHVFSDGVTNWGRIVTLISFGAFVAKHLKTINQESCIEPLAESITDVLVRTK
RDWLVKQRGWDGFVEFFHVEDLEGGIRNVLLAFAGVAGVGAGLAYLIR
Mass: 37,337 Da

Using BLAST to find identical proteins in other species:
Length: 350 AA
Identity: 22.1 - 100
Species: Human, Zebrafish, Rat, Bovine, Mouse

Main Accession: Q07820. Secondary accessions: B2R6B2, D3DV03, D3DV04, Q9HD91, Q9NRQ3

Most similar in other species are 3 proteins matching at 99.7%:
1: MCL1: A0A2R9BYH6 in Pan paniscus (Pygmy chimpanzee)(Bonobo) with 350 AA
2: CK820_G0046961: A0A6D2X745 in Pan troglodytes (chimpanzee) with 350 AA
3: MCL1: K7D0E2 in Pan troglodytes (chimpanzee) with 350 AA

These 3 are the same entry but 3 potential isoforms. For each isoform:
Status:
unreviewed

Mass (Da):
37278 Da

Most similar in Homo Sapiens:
cDNA FLJ54274

Length:
279 AA

Accession number:
B4DU51

Mass (Da):
30205 Da

Questions (1):
- Which target is more similar compared to the original target?

MCL1 and CK820 that are found in other species are more similar to human MCL1 compared to the most similar protein in humans
- Did you expect this?

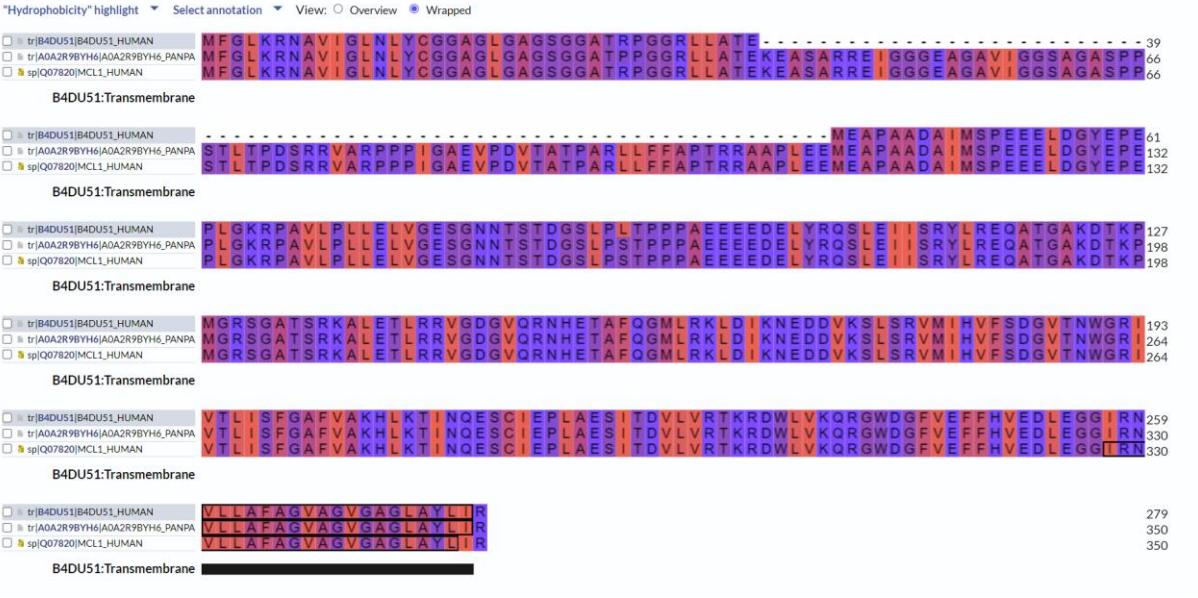
We did expect to find a more similar protein in a similar organism because the proteins between organisms are often very closely related to eachother as was also indicated by the lectures of this course

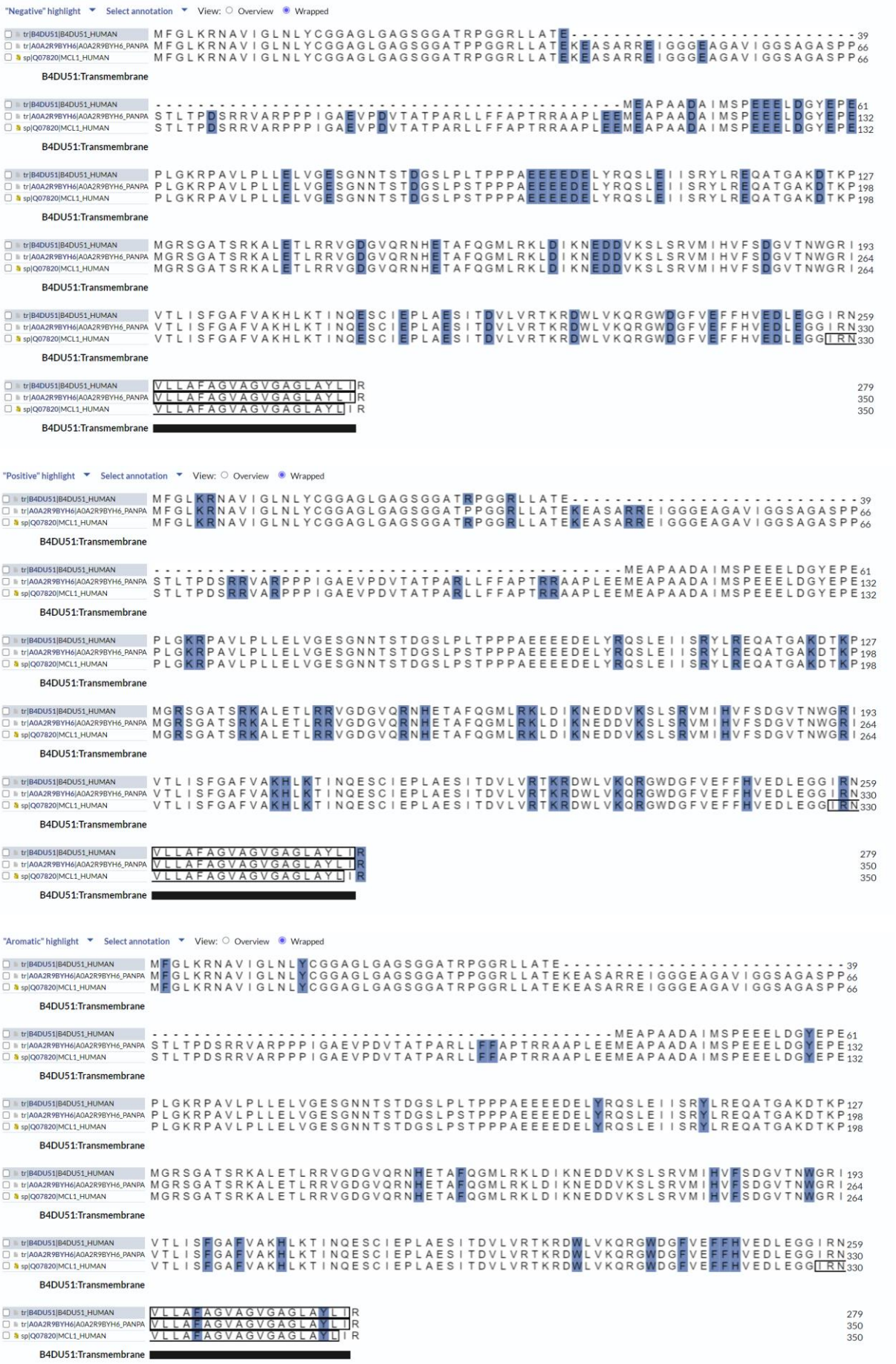
![image.png](attachment:image.png)
![image-2.png](attachment:image-2.png)
![image-3.png](attachment:image-3.png)
![image-4.png](attachment:image-4.png)
![image-5.png](attachment:image-5.png)

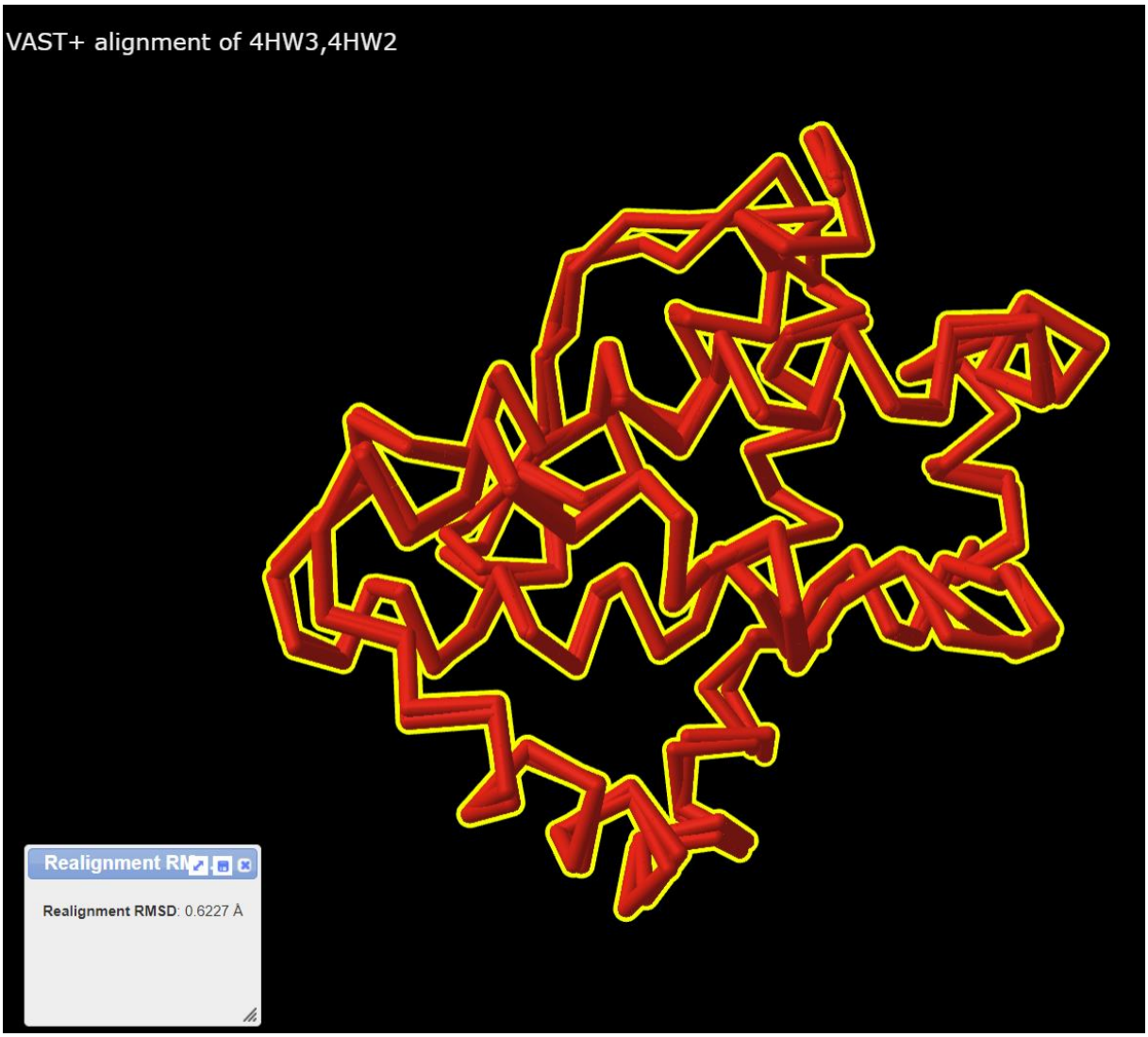
Q07820 is more similar to A0A2R9BYH6

In case of structure similartity: 4HW3 is most similar to 4HW2
![image-6.png](attachment:image-6.png)

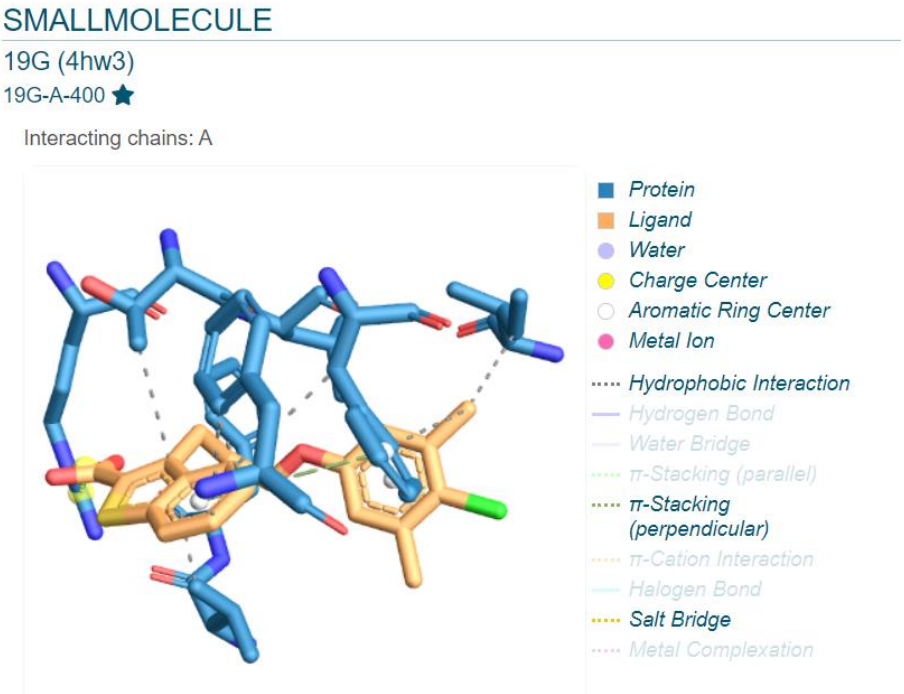

"Similarity" highlight
Select annotation
View: Overview Wrapped
tr|B4DU51|B4DU51_HUMAN
tr|A0A2R9BYH6|A0A2R9BYH6_PANPA
sp|Q07820|MCL1_HUMAN
B4DU51:Transmembrane


"Hydrophobicity" highlight
Select annotation
View: Overview Wrapped
tr|B4DU51|B4DU51_HUMAN
tr|A0A2R9BYH6|A0A2R9BYH6_PANPA
sp|Q07820|MCL1_HUMAN
B4DU51:Transmembrane


"Negative" highlight
Select annotation
View: Overview Wrapped
tr|B4DU51|B4DU51_HUMAN
tr|A0A2R9BYH6|A0A2R9BYH6_PANPA
sp|Q07820|MCL1_HUMAN
B4DU51:Transmembrane
"Positive" highlight
Select annotation
View: Overview Wrapped
tr|B4DU51|B4DU51_HUMAN
tr|A0A2R9BYH6|A0A2R9BYH6_PANPA
sp|Q07820|MCL1_HUMAN
B4DU51:Transmembrane
"Aromatic" highlight
Select annotation
View: Overview Wrapped
tr|B4DU51|B4DU51_HUMAN
tr|A0A2R9BYH6|A0A2R9BYH6_PANPA
sp|Q07820|MCL1_HUMAN
B4DU51:Transmembrane


VAST+ alignment of 4HW3,4HW2
Realignment RM
Realignment RMSD: 0.6227 Å

Machine Learning



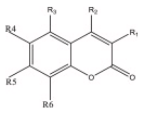
The following paper is a relevant paper for 4HW3 and MCL1. This paper was used to answer some of the questions.

Esraa Albasher Osman, Mohammed Abdalrahman Abdalla, Mohja Omer Abdelraheem, Mubarak Fadlalla Ali, Shima Albasher Osman, Yasmin Mohamed Tanir, Mohammed Abdelrahman, Walaa Ibraheem, Abdulrahim A. Alzain, Design of novel coumarins as potent Mcl-1 inhibitors for cancer treatment guided by 3D-QSAR, molecular docking and molecular dynamics, Informatics in Medicine Unlocked, Volume 26, 2021, 100765, ISSN 2352-9148, https://doi.org/10.1016/j.imu.2021.100765.

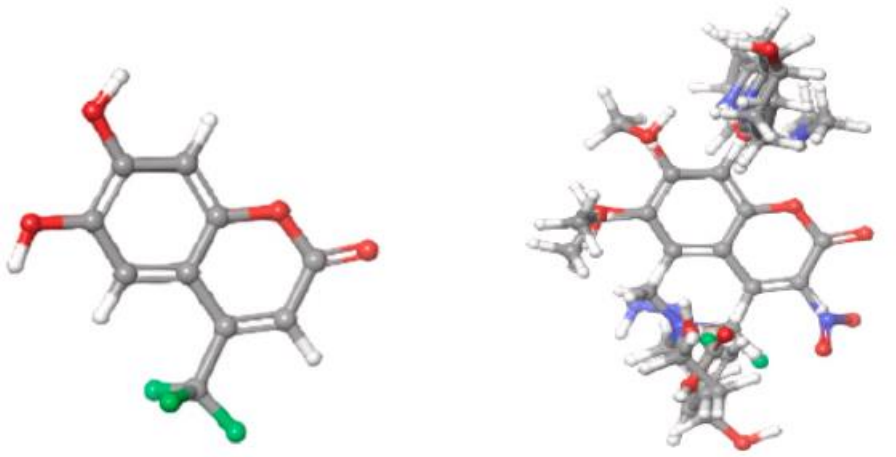
(https://www.sciencedirect.com/science/article/pii/S2352914821002392)

Abstract: Myeloid cell leukemia-1 (Mcl-1) is a protein that belongs to a large group of proteins called B cell lymphoma-2 (Bcl-2) which are involved in controlling of apoptosis via interacting with other Bcl-2 family. Various studies showed that Mcl-1 is overexpressed in cancer cells. Thus, it is a promising target for cancer treatment. In the present study, in silico drug design approaches were applied on a library of 33 coumarin derivatives with Mcl-1 inhibitory activity. Firstly, 3D-QSAR study was performed using Gaussian field-based QSAR resulting in a good predictive model with r2 value of 0.80 and q2 value of 0.81 and Pearson-r value of 0.95. Depending on the established model, 10 novel designed compounds were predicted to have more than 10-fold inhibitory activity against Mcl-1 compared to the reference compound. Moreover, these designed molecules have good ADMET properties within the recommended range. Secondly, molecular docking and molecular dynamic simulations were performed to locate the newly designed inhibitors into protein active site to discover how they fit together in the most reliable conformations and address the molecular interactions stability. The in silico methods offer guidance to rational drug design of new compounds with improved potency.

Keywords: Mcl-1; Apoptosis; Gaussian based 3D-QSAR; Docking; Molecular dynamics



Cpd	R_1	R_2	R_3	R_4	R_5	R_6	IC_{50} (μM)
1	H	H	H	OH	OH	H	8.77 ± 1.16
2	H	CH_3	H	OH	OH	H	6.28 ± 1.08
3	H	C_6H_5	H	OH	OH	H	3.27 ± 1.77
4	H	CF_3	H	OH	OH	H	1.21 ± 0.56
5	H	CH_2Cl	H	OH	OH	H	5.82 ± 1.18
6	H	CH2OH	H	OH	OH	H	4.01 ± 0.95
7	H	CH_2N_3	H	OH	OH	H	34.62 ± 1.06
8	H	CH_2COOH	H	OH	OH	H	33.73 ± 0.86
9	H	H	H	OH	OH	$CH_2N(CH_3)_2$	53.6 ± 1.68
10	H	H	$CH_2N(CH_3)_2$	OH	OH	H	19.92 ± 0.94
11	H	H	H	OH	OH	$CH_2N(CH_2)_4CHOH$	21.79 ± 1.06
12	H	H	$CH_2N(CH_2)_4CHOH$	OH	OH	H	14.69 ± 3.16
13	H	H	H	OH	OH	$CH_2N(CH_2)_4$	27.97 ± 1.02
14	H	CH_3	H	OH	OH	$CH_2N(CH_2)_4CHOH$	59.22 ± 0.62
15	H	CH_3	H	OH	OH	$CH_2N(CH_2)_4$	11.34 ± 1.51
16	H	H	H	OCH_3	OH	H	22.59 ± 1.49
17	H	H	H	OCH_3	OCH_3	H	25.14 ± 1.53
18	NO_2	H	H	OCH_3	OCH_3	NO_2	12.21 ± 1.29
19	H	H	H	H	OH	OH	11.77 ± 5.12
20	H	CH_3	H	H	OH	OH	11.21 ± 1.57
21	H	CH_2Cl	H	H	OH	OH	18.61 ± 4.33
22	H	CH_2OH	H	H	OH	OH	32.23 ± 7.45
23	H	H	H	H	OCH_3	OH	16.93 ± 4.70
24	H	CH_3	H	H	OCH_3	OH	16.93 ± 4.70
25	H	CH_2Cl	H	H	OCH_3	OH	30.57 ± 6.83
26	H	CH_2OH	H	H	OCH_3	OH	32.89 ± 7.45
27	H	CH_3	H	H	OH	H	103.70 ± 0.11
28	H	CH_3	H	H	OH	NO_2	26.29 ± 4.14
29	H	CH_3	H	Cl	OH	H	28.57 ± 5.47
30	H	CH_3	H	CH_2CH_3	OH	H	30.88 ± 6.12
31	H	H	H	OH	H	H	83.69 ± 5.64
32	H	H	H	H	OH	H	21.04 ± 4.41
33	H	H	H	OCH_3	OH	OH	32.1 ± 6.90



A B

Download : Download high-res image (292KB) Download : Download full-size image

Fig. 1. (A) Structure 4 (B) Alignment of all 33 structures.

4HW3 InChI key: GFWMYYSJLSUPMD-UHFFFAOYSA-N

SMILES: CC1=CC(=CC(=C1Cl)C)OCCCC2=C(SC3=CC=CC=C32)C(=O)O

There are 36 similar compounds found. Changing the Tanimoto threshold to 80% results in >1000 similar compounds found.

ChEMBL ID: CHEMBL2314173

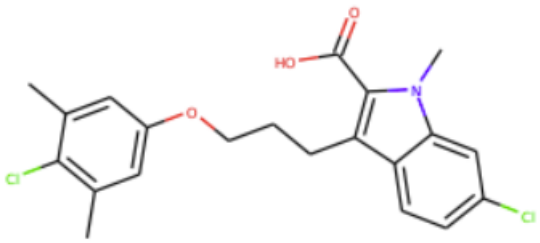
At 95% similarity only 1 compound was found. At 90% similarity still only 1 compound was found.

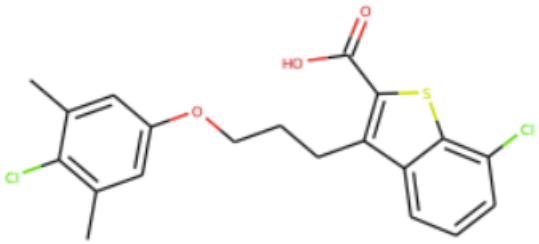
Although in CHEMBL only 1 similar compound was found and in PubChem there were 36 found (both at 90% similarity), this can be explained.

CHEMBL finds similar compounds based on the exact structure, whereas PubChem finds compounds that are similar in any type of aspect. This can be also similar in the amount of each atom but with completely different positions compared to 4HW3.

Using the ZINC site to find similar compounds with a Tanimoto 70 resulted in 9 hits.

These compounds are all very similar and not just similar based on the amount of atoms in the molecule.





Differences are small with for example an Oxygen one position different or a difference like either a Nitrogen or a Sulfur.

Quiz at the end of Machine Learning:

Question 1:
ML can be used by creating an algorithm that learns from the data. The ML Algorithm will get better at predicting outcomes after the algorithm gets better and more data. This will result in a better and better model that can eventually be used for virtual screening since the model can predict what will work.
Question 2:
Linear regression, logistic regression, decision trees, random forest
Question 3:
A large dataset is first needed. Next to this the molecules in the dataset need to have their own unique code, and a label. Then a ML algorithm can be used in order to train a model.