Белорусский государственный медицинский университет, г. Минск

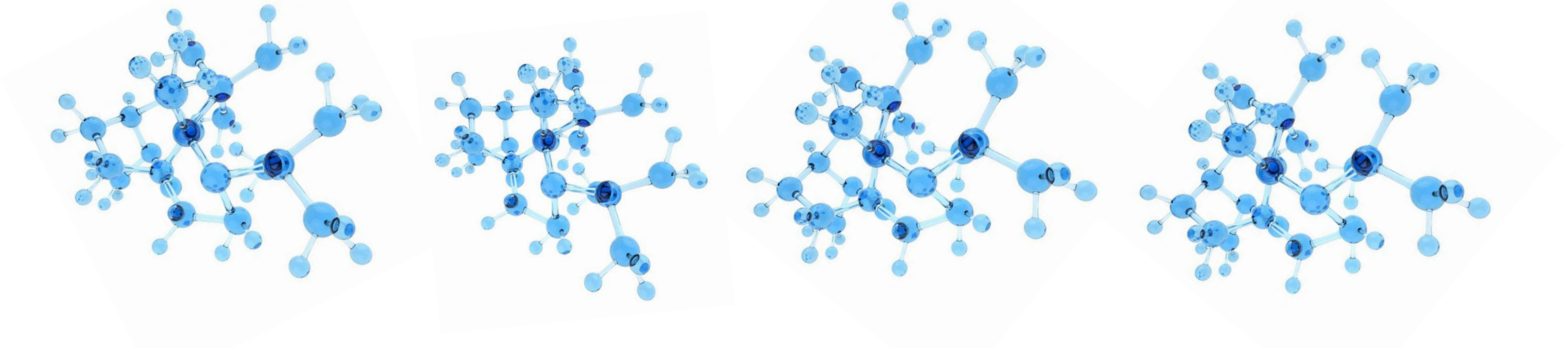
# IN SILICO ПОИСК ПОТЕНЦИАЛЬНЫХ ИНГИБИТОРОВ ГЕПАРАНАЗЫ (Q9Y251)

Кафедра фармакологии
Табакова Д. В., Терлецкая В. А.
3 курс, фармацевтический факультет
Научный руководитель ассист. Кашкур Ю.В.

**Цель:** поиск новых ингибиторов гепараназы на основе построенной модели фармакофора

**Задачи:**

1) Провести скрининг соединений, обладающих сродством к гепараназе;
2) Базируясь на центроидных молекулах, построить модель фармакофора;
3) На основе построенного фармакофора найти потенциальные ингибиторы гепараназы, обладающие удовлетворительными фармакокинетическими параметрами.



# Материалы и методы.

Данные для построения модели фармакофора были получены из баз данных Protein Data Bank и ChEMBL DB;

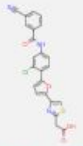
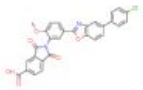
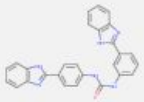
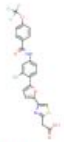
Скрипты, для обработки и кластеризации данных были написаны на языке программирования Python;

Построение фармакофора осуществлялось с помощью PharmaGist webserver;

Поиск новых соединений осуществлялся в базе данных ZINC 12 с помощью веб ресурса ZINCPharmer.

# Результаты и их обсуждение.

Для построения модели фармакофора использовалось ***504*** ингибитора гепараназы, найденные в базе данных **ChEMBL**

| Molecule ChEMBL ID | Compound Key | Standard Type | Standard Relation | Standard Value | Standard Units | pChEMBL Value | Comment | Assay ChEMBL ID | Assay Description |
|---|---|---|---|---|---|---|---|---|---|
|  CHEMBL197295 | 46 | IC50 | = | 3981.07 | nM | 5.40 | No Data | CHEMBL990587 | Inhibition of heparanase |
|  CHEMBL112739 | 20 | IC50 | = | 4466.84 | nM | 5.35 | No Data | CHEMBL990587 | Inhibition of heparanase |
|  CHEMBL200554 | 12 | Inhibition | < | 10.0 | % | No Data | No Data | CHEMBL869913 | Inhibitory acti against heparanase fr human platele at 0.83 uM |
|  CHEMBL192463 | 44 | IC50 | = | 1513.56 | nM | 5.82 | No Data | CHEMBL990587 | Inhibition of heparanase |

# Активность молекул определялась на основе значений **IC50**.

```
data = pd.read_csv('Data.csv')
```

```
data.pivot_table(index = ['standard_type'], aggfunc='size')
```

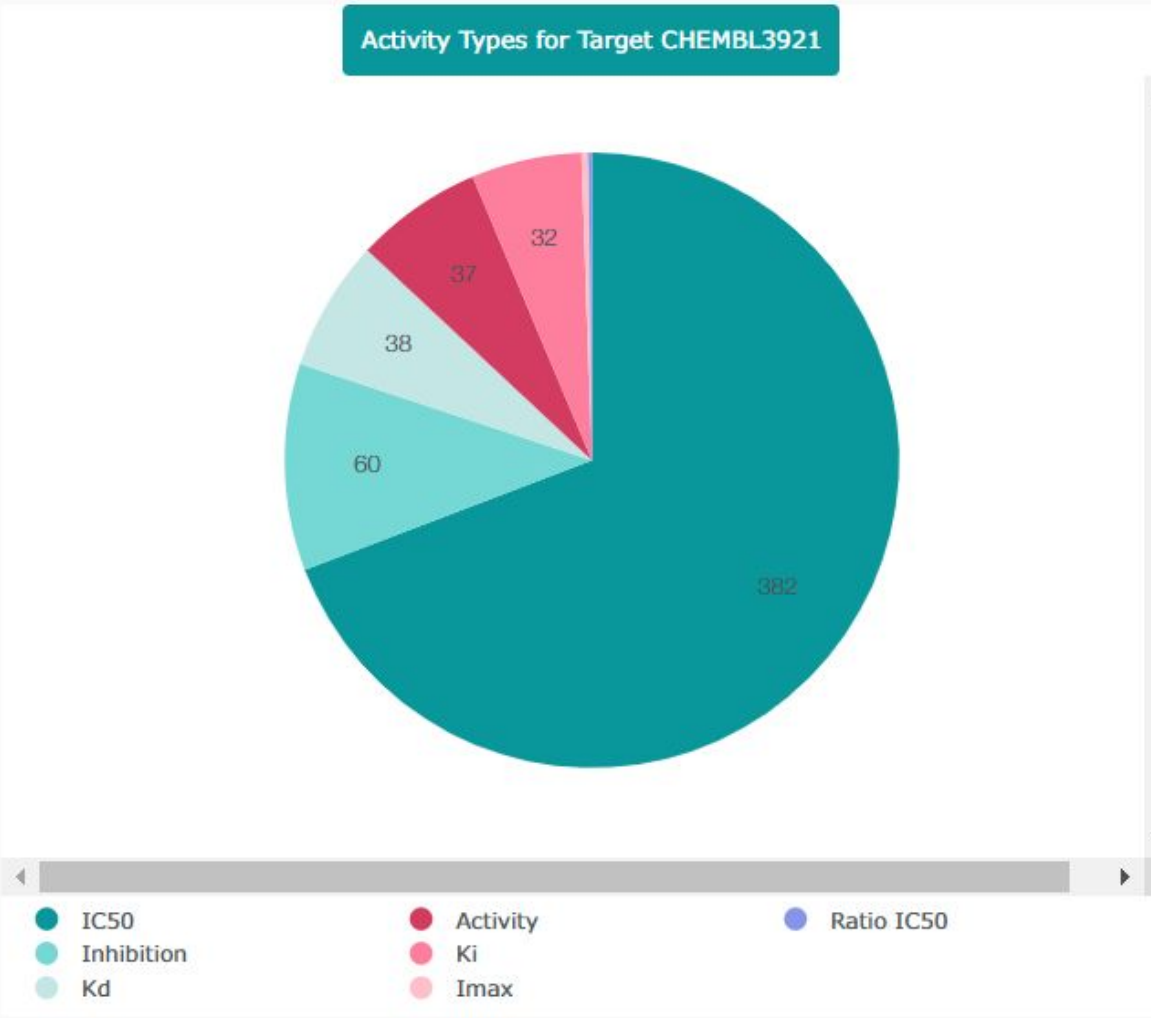
```
standard_type
IC50    172
dtype: int64
```

```
data = data.query("standard_type == 'IC50'")
```

```
data.head()
```

| | molecule_chembl_id | Smiles | standard_type | s |
|---|---|---|---|---|
| 0 | CHEMBL200215 | COc1ccc(C(=O)Nc2ccc(CNc3ccc(-c4nc5ccccc5[nH]4)... | IC50 | |
| 1 | CHEMBL199576 | CCCOc1ccc(C(=O)Nc2ccc(-c3nc4ccccc4[nH]3)cc2)cc1Br | IC50 | |
| 2 | CHEMBL194140 | COc1ccc(NC(=O)/C=C/c2ccc(-c3nc4cc(CC(=O)O)ccc4... | IC50 | |
| 3 | CHEMBL194231 | O=C(O)Cc1ccc2oc(-c3ccc(/C=C/C(=O)Nc4ccc(Br)cc4... | IC50 | |
| 4 | CHEMBL196338 | COc1ccc(NC(=O)/C=C/c2ccc(-c3nc4cc(CC(=O)O)ccc4... | IC50 | |

## Associated Bioactivities



# Для удобства использования значения IC50 были конвертированы в **pIC50**

```
def calcPic50(st_values):
    for i in range(0, len(st_values)):
        st_values[i] = st_values[i]*10**-9
        st_values[i] = round(-np.log10(st_values[i]), 3)
    return st_values
```

```
st_val = list(data['standard_value'])
```

```
pic50_vals = calcPic50(st_val)
```

```
data['pIC50'] = pic50_vals
```

```
data.head()
```

| | molecule_chembl_id | Smiles | standard_type | standard_value | standard_units | assay_type | assay_organism | target_organism | pIC50 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | CHEMBL200215 | COc1ccc(C(=O)Nc2ccc(CNc3ccc(-c4nc5ccccc5[nH]4)... | IC50 | 230.0 | nM | B | Homo sapiens | Homo sapiens | 6.638 |
| 1 | CHEMBL199576 | CCCOc1ccc(C(=O)Nc2ccc(-c3nc4ccccc4[nH]3)cc2)cc1Br | IC50 | 2200.0 | nM | B | Homo sapiens | Homo sapiens | 5.658 |
| 2 | CHEMBL194140 | COc1ccc(NC(=O)/C=C/c2ccc(-c3nc4cc(CC(=O)O)ccc4... | IC50 | 3000.0 | nM | B | Homo sapiens | Homo sapiens | 5.523 |
| 3 | CHEMBL194231 | O=C(O)Cc1ccc2oc(-c3ccc(/C=C/C(=O)Nc4ccc(Br)cc4... | IC50 | 400.0 | nM | B | Homo sapiens | Homo sapiens | 6.398 |
| 4 | CHEMBL196338 | COc1ccc(NC(=O)/C=C/c2ccc(-c3nc4cc(CC(=O)O)ccc4... | IC50 | 450.0 | nM | B | Homo sapiens | Homo sapiens | 6.347 |

Активными считались соединения со значением **pIC50 ≥ 6.0**.

И

```
data = data.query("pIC50 >= 6.0")
```

```
data.head()
```

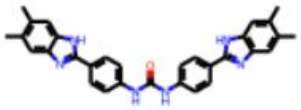
| | molecule_chembl_id | Smiles | standard_type | standard_value | standard_units | assay_type | assay_organism | target_organism | pIC50 |
|---|---|---|---|---|---|---|---|---|---|
| **0** | CHEMBL200215 | COc1ccc(C(=O)Nc2ccc(CNc3ccc(-c4nc5ccccc5[nH]4)... | IC50 | 230.0 | nM | B | Homo sapiens | Homo sapiens | 6.638 |
| **1** | CHEMBL194231 | O=C(O)Cc1ccc2oc(-c3ccc(/C=C/C(=O)Nc4ccc(Br)cc4... | IC50 | 400.0 | nM | B | Homo sapiens | Homo sapiens | 6.398 |
| **2** | CHEMBL196338 | COc1ccc(NC(=O)/C=C/c2ccc(-c3nc4cc(CC(=O)O)ccc4... | IC50 | 450.0 | nM | B | Homo sapiens | Homo sapiens | 6.347 |
| **3** | CHEMBL196336 | O=C(O)Cc1nc(-c2ccc(-c3ccc(NC(=O)/C=C/c4ccc(OC(... | IC50 | 600.0 | nM | B | Homo sapiens | Homo sapiens | 6.222 |
| **4** | CHEMBL196826 | O=C(O)Cc1ccc2oc(-c3ccc(NC(=O)/C=C/c4cccc(Br)c4... | IC50 | 750.0 | nM | B | Homo sapiens | Homo sapiens | 6.125 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **60** | CHEMBL200597 | Cc1cc2nc(-c3ccc(NC(=O)Nc4ccc(-c5nc6cc(C)c(C)cc... | IC50 | 560.0 | nM | B | Homo sapiens | Homo sapiens | 6.252 |
| **61** | CHEMBL4291586 | O=C(O)Cc1ccc2[nH]c(-c3ccc(NC(=O)Nc4ccc(-c5nc6c... | IC50 | 980.0 | nM | B | Homo sapiens | Homo sapiens | 6.009 |
| **62** | CHEMBL4290499 | CC(NC(=O)Cc1ccc2oc(-c3ccc(NC(=O)Nc4ccc(-c5nc6c... | IC50 | 370.0 | nM | B | Homo sapiens | Homo sapiens | 6.432 |
| **63** | CHEMBL4283311 | O=C(O)CNC(=O)Cc1ccc2oc(-c3ccc(NC(=O)Nc4ccc(-c5... | IC50 | 450.0 | nM | B | Homo sapiens | Homo sapiens | 6.347 |
| **64** | CHEMBL4280766 | CC(=O)N[C@H]1[C@@H](O[C@@H]2C(C(=O)O)O[C@@H](O... | IC50 | 5.0 | nM | B | Homo sapiens | Homo sapiens | 8.301 |

65 rows × 9 columns

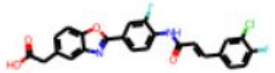
Далее с помощью алгоритма кластеризации **Butina** наиболее структурно схожие соединения были объединены в **6** кластеров, из которых были отобраны **6** центроидных молекул.

```
print("Centroid molecules from first 6 clusters:")
# Draw molecules
Draw.MolsToGridImage(
    [compounds[clusters[i][0]][0] for i in range(6)],
    legends=[compounds[clusters[i][0]][1] for i in range(6)],
    molsPerRow=5,
)
```

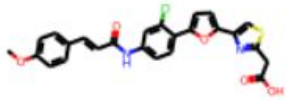
```
Centroid molecules from first 6 clusters:
```



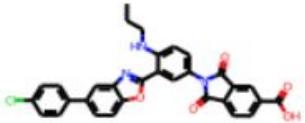
CHEMBL200597



CHEMBL195907

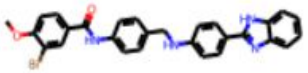


CHEMBL371702



CHEMBL193906

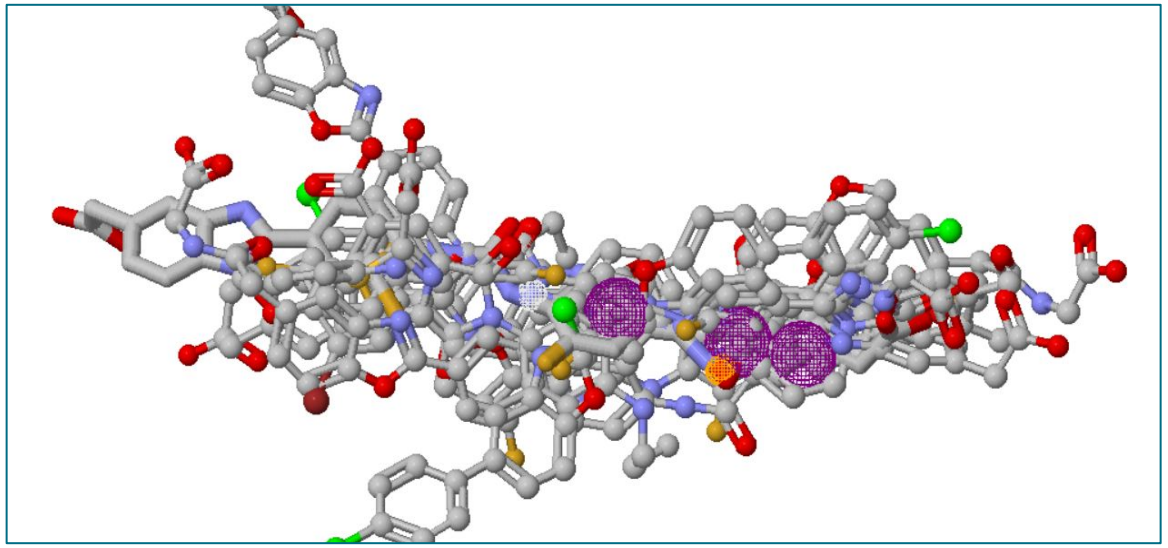
CHEMBL4283311



CHEMBL200215

С помощью **PharmaGist** webserver на основе 4-х молекул была построена модель фармакофора. Фармакофор представляет собой **6** фармакофорных центров:

- 1 донор водорода,
- 2 акцептора водорода,
- 3 ароматических цикла.



Pharmacophore | Filters | Viewer | Submit Query

| | Pharmacophore Class | x | y | z | Radius | Enabled | |
|---|---|---|---|---|---|---|---|
| > | HydrogenDonor | -2.76 | 0.38 | 0.00 | 0.50 | ☑ | |
| > | HydrogenAcceptor | 10.51 | 0.44 | 0.65 | 0.50 | ☑ | |
| > | HydrogenAcceptor | 3.32 | -1.92 | 0.60 | 0.50 | ☑ | |
| > | Aromatic | 5.96 | -1.53 | 0.17 | 1.10 | ☑ | |
| > | Aromatic | 3.90 | -1.07 | 0.13 | 1.10 | ☑ | |
| > | Aromatic | 0.01 | -0.18 | 0.03 | 1.10 | ☑ | |

Results

| Name | RMSD | Mass | RBnds |
|---|---|---|---|
| ZINC02384699 | 0.729 | 489 | 6 |
| ZINC09274463 | 0.721 | 488 | 8 |
| ZINC08592990 | 0.721 | 461 | 5 |
| ZINC40399157 | 0.713 | 392 | 5 |
| ZINC40267303 | 0.713 | 378 | 4 |
| ZINC40398886 | 0.713 | 410 | 5 |
| ZINC32093915 | 0.713 | 364 | 3 |
| ZINC40399061 | 0.713 | 363 | 3 |
| ZINC40399020 | 0.702 | 420 | 6 |
| ZINC40398982 | 0.702 | 382 | 8 |
| ZINC40399054 | 0.702 | 337 | 5 |
| ZINC40285420 | 0.702 | 352 | 6 |
| ZINC32105588 | 0.702 | 338 | 5 |
| ZINC40399148 | 0.702 | 366 | 7 |
| ZINC40398871 | 0.702 | 384 | 7 |
| ZINC40399032 | 0.685 | 453 | 4 |
| ZINC40398929 | 0.685 | 431 | 6 |
| ZINC40398882 | 0.685 | 417 | 5 |
| ZINC40399009 | 0.685 | 415 | 6 |
| ZINC40399059 | 0.683 | 392 | 4 |
| ZINC36359488 | 0.683 | 393 | 4 |
| ZINC40399024 | 0.683 | 476 | 5 |

<< < 1 2 3 4 5 6 7 8 > >>

377 hits
9.918s

Для поиска новых соединений в базе данных ZINC 12 использовался веб-сервер **ZINCPharmer**.

В результате было найдено **377** потенциальных ингибиторов гепараназы в диапазоне значений RMSD ***0,729 – 0,135***.

Results

| Name | RMSD | Mass | RBnds |
|---|---|---|---|
| ZINC00826908 | 0.135 | 533 | 9 |
| ZINC00826908 | 0.135 | 533 | 9 |
| ZINC00826908 | 0.135 | 533 | 9 |

<< < 11 12 13 14 15 16 17 18 > >

377 hits
9.918s

**RMSD –** root-mean-square deviation (среднеквадратичное отклонение)

# Соединения были отобраны в соответствии с правилом **Липинского.**

## Lipinski Filter

```
smiles = list(data['smiles'])
a = LipinskiCalc(smiles_2)
```

```
data = a.lipinski_table
data
```

| | Smile | Molecular_weight | Acceptors | Donors | LogP | Fulfill |
|---|---|---|---|---|---|---|
| **0** | Oc1ccc(-c2cn3nc(N4CCOCC4)sc3n2)cc1 | 302.083747 | 7 | 1 | 2.0000 | True |
| **1** | NNC(=O)c1ccc(-c2nc3ccc(C(=O)NN)cc3[nH]2)cc1 | 310.117824 | 5 | 5 | 0.4369 | True |
| **2** | Nc1ccc(-c2cn3c(C(=O)N4CCOCC4)csc3n2)cc1 | 328.099397 | 6 | 1 | 2.1174 | True |
| **3** | COCCn1ccc2nc3ccn(-c4ccn[nH]4)c(=O)c3cc2c1=O | 337.117489 | 7 | 1 | 1.0701 | True |
| **4** | CC(=O)Nc1ccc(-c2nc3ccc(OC(C)=O)cc3c(=O)o2)cc1 | 338.090272 | 6 | 1 | 2.7387 | True |
| **...** | ... | ... | ... | ... | ... | ... |
| **108** | O=C(Nc1ccc(-c2nc(-c3ccccc3)c3cc(NC(=O)C4CCCCC4... | 532.283826 | 4 | 2 | 8.0014 | False |
| **109** | O=C(Cc1ccccc1)Nc1ccc(-c2nc(-c3ccccc3)c3cc(NC(=... | 548.221226 | 4 | 2 | 7.3262 | False |
| **110** | CC(C)(C)c1cc(NC(=O)Nc2ccc(-c3cn4c(n3)sc3cc(OCC... | 560.220575 | 9 | 2 | 5.8565 | False |
| **111** | COc1c(OC2OC(COC3OCC(O)(CO)C3O)C(O)C(O)C2O)cc2o... | 594.158470 | 15 | 8 | -1.4766 | False |
| **112** | COC(=O)C1OC(Oc2cc(OC(C)=O)c3c(=O)c(-c4ccc(O)cc... | 628.142820 | 15 | 1 | 2.1627 | False |

113 rows × 6 columns

Не более 5 донорных водородных связей

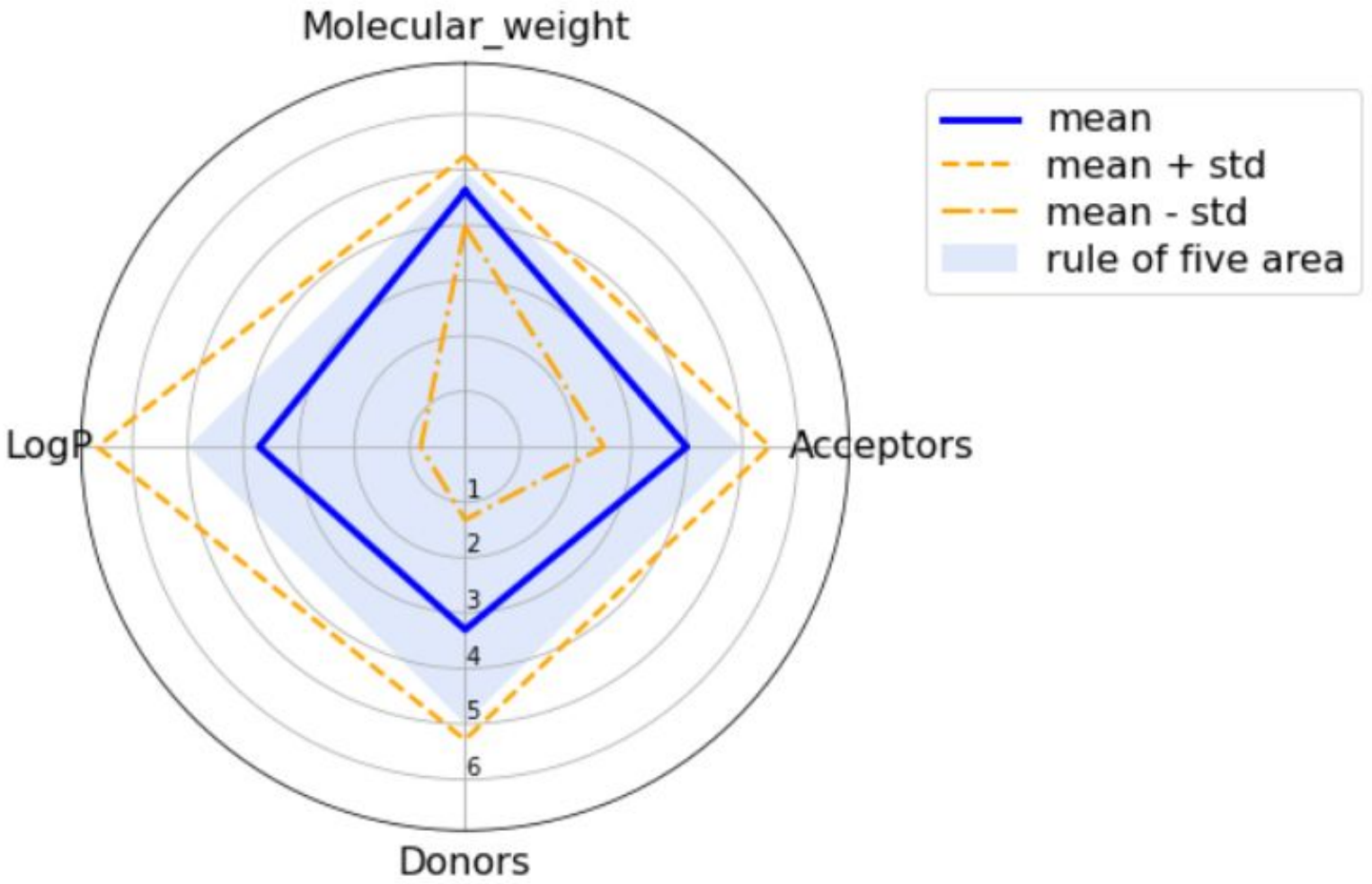
Не более 10 акцепторных водородных связей

Молекулярная масса соединения менее 500 а.е.м.

Коэффициент распределения октанол-вода (log *P* ), не должен превышать 5

## Draw Lipinski radar plot

```
a.visualize(data_fulfilled)
```



Из оставшихся ***113*** молекул были исключены молекулы из списка **PAINS**.

В результате осталось ***28*** молекул.

## Filter PAINS

```
p = PAINS(list(data_fulfilled['Smile']))
```

```
p.get_pains()
```

```
[{'smiles': 'C/C(=N/NC(=O)c1cc(-c2ccc(O)cc2)nc2ccccc12)c1cc(Cl)ccc1O',
  'pains': 'hzone_phenol_A(479)'},
 {'smiles': 'Cc1ccc(/N=N/c2c(S(=O)(=O)[O-])cc3c(ccc4nc(-c5ccc(N)cc5)sc43)c2O)c(C)c1',
  'pains': 'azo_A(324)'}]
```

```
no_pains = p.exclude()
len(no_pains)
```

```
28
```

После удаления соединений, содержащих нежелательные подструктуры, отобрали ***22*** молекулы, обладающие удовлетворительными фармакокинетическими параметрами.

```
u = UnwantedSubs(no_pains)
```

```
unwanted, clean = u.get_unwanted()
```

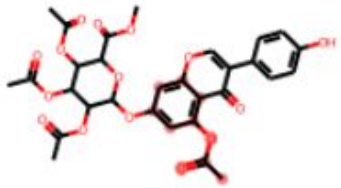
```
unwanted
```

```
[{'smiles': 'COc1ccc2c(c1)sc1nc(-c3ccc(O)cc3)c(NCCCOC(C)C)n12',
  'name': 'Aliphatic-long-chain',
  'smarts': '[R0D2][R0D2][R0D2][R0D2]'},
 {'smiles': 'Nc1ccc(-c2nc3cc(Oc4ccc5[nH]c(-c6ccc(N)cc6)nc5c4)ccc3[nH]2)cc1',
  'name': 'aniline',
  'smarts': 'c1cc([NH2])ccc1'},
 {'smiles': 'Nc1ccc(-c2nc3cc(Oc4ccc5oc(-c6ccc(N)cc6)nc5c4)ccc3o2)cc1',
  'name': 'aniline',
  'smarts': 'c1cc([NH2])ccc1'},
 {'smiles': 'O=C(CCl)Nc1ccc(-c2nc(-c3ccccc3)c3cc(NC(=O)CCl)ccc3n2)cc1',
  'name': 'alkyl-halide',
  'smarts': '[CX4][Cl,Br,I]'},
 {'smiles': 'C/C(=N/NC(=O)c1cc(-c2ccc(O)cc2)nc2ccc(Br)cc12)c1ccccc1OC(C)C',
  'name': 'Oxygen-nitrogen-single-bond',
  'smarts': '[OR0,NR0][OR0,NR0]'},
 {'smiles': 'COC(=O)C1OC(Oc2cc(OC(C)=O)c3c(=O)c(-c4ccc(O)cc4)coc3c2)C(OC(C)=O)C(OC(C)=O)C1OC(C)=O',
  'name': '>2EsterGroups',
  'smarts': 'C(=O)O[C,H1].C(=O)O[C,H1].C(=O)O[C,H1]'},
 {'smiles': 'COC(=O)C1OC(Oc2cc(OC(C)=O)c3c(=O)c(-c4ccc(O)cc4)coc3c2)C(OC(C)=O)C(OC(C)=O)C1OC(C)=O',
  'name': 'phenol-ester',
  'smarts': 'c1ccccc1OC(=O)[#6]'}]
```

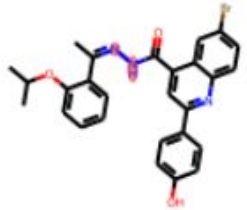
```
print(f'Unwanted found: {len(unwanted)} \nclean: {len(clean)}')
```

```
Unwanted found: 7
clean: 22
```

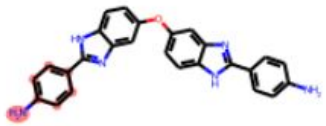
```
u.visualize(unwanted[:7])
```



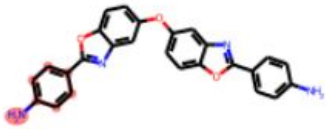
phenol-ester



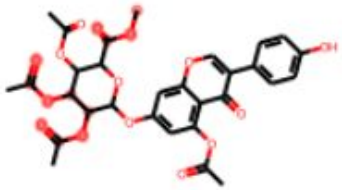
Oxygen-nitrogen-single-bond



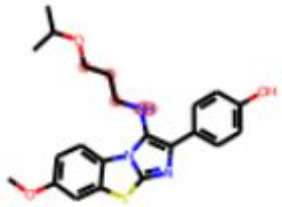
aniline



aniline



>2EsterGroups



Aliphatic-long-chain

alkyl-halide

# Выводы.

Из 504 соединений было отобрано 6 центроидных молекул, которые стали основой для построения фармакофора;

Найдено 377 потенциальных ингибиторов гепараназы, из которых 22 молекулы обладают удовлетворительными фармакокинетическими параметрами;

В дальнейшем планируется докинг наиболее удачных кандидатов.

*Благодарим за внимание!!!*