

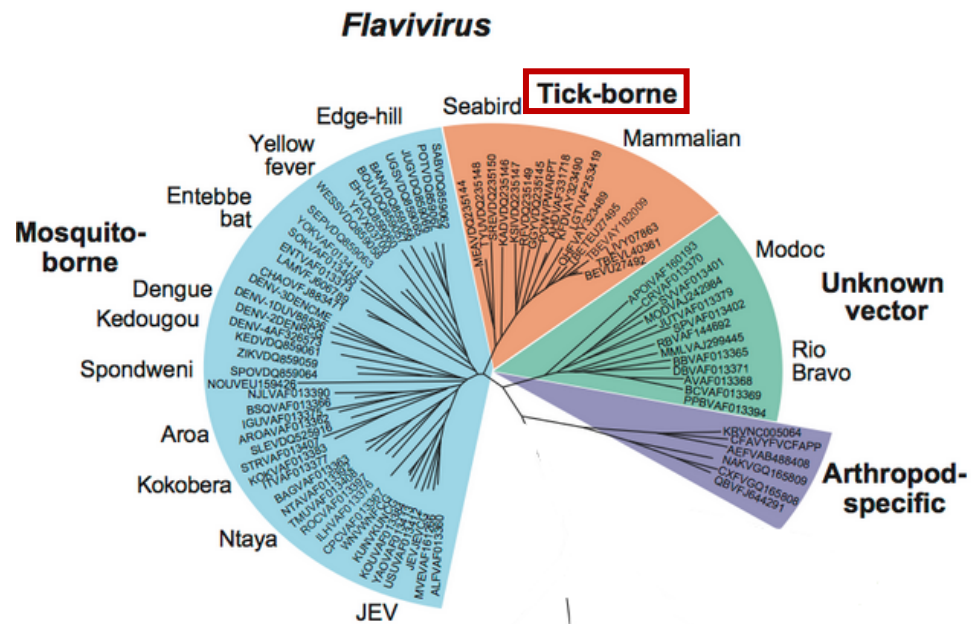
ФГБНУ «Федеральный научный центр исследований и разработки иммунобиологических препаратов им. М.П. Чумакова РАН»

Виртуальный скрининг ингибиторов метилтрансферазы вируса клещевого энцефалита на основе методов машинного обучения

В.С. Фроленко, А.А. Никитина, Д.И. Осолодкин

Вирус клещевого энцефалита

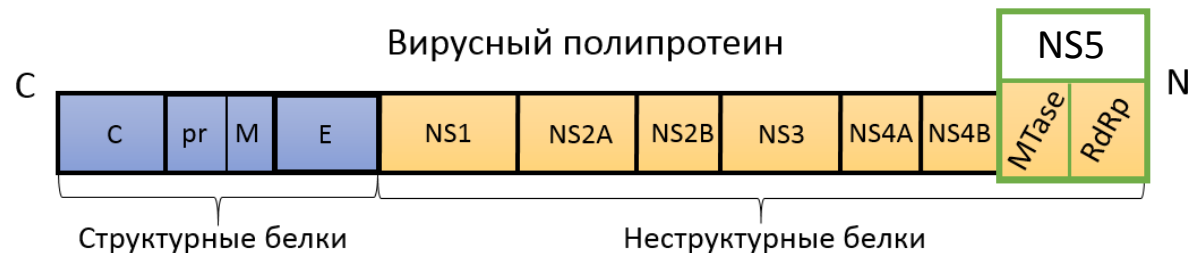
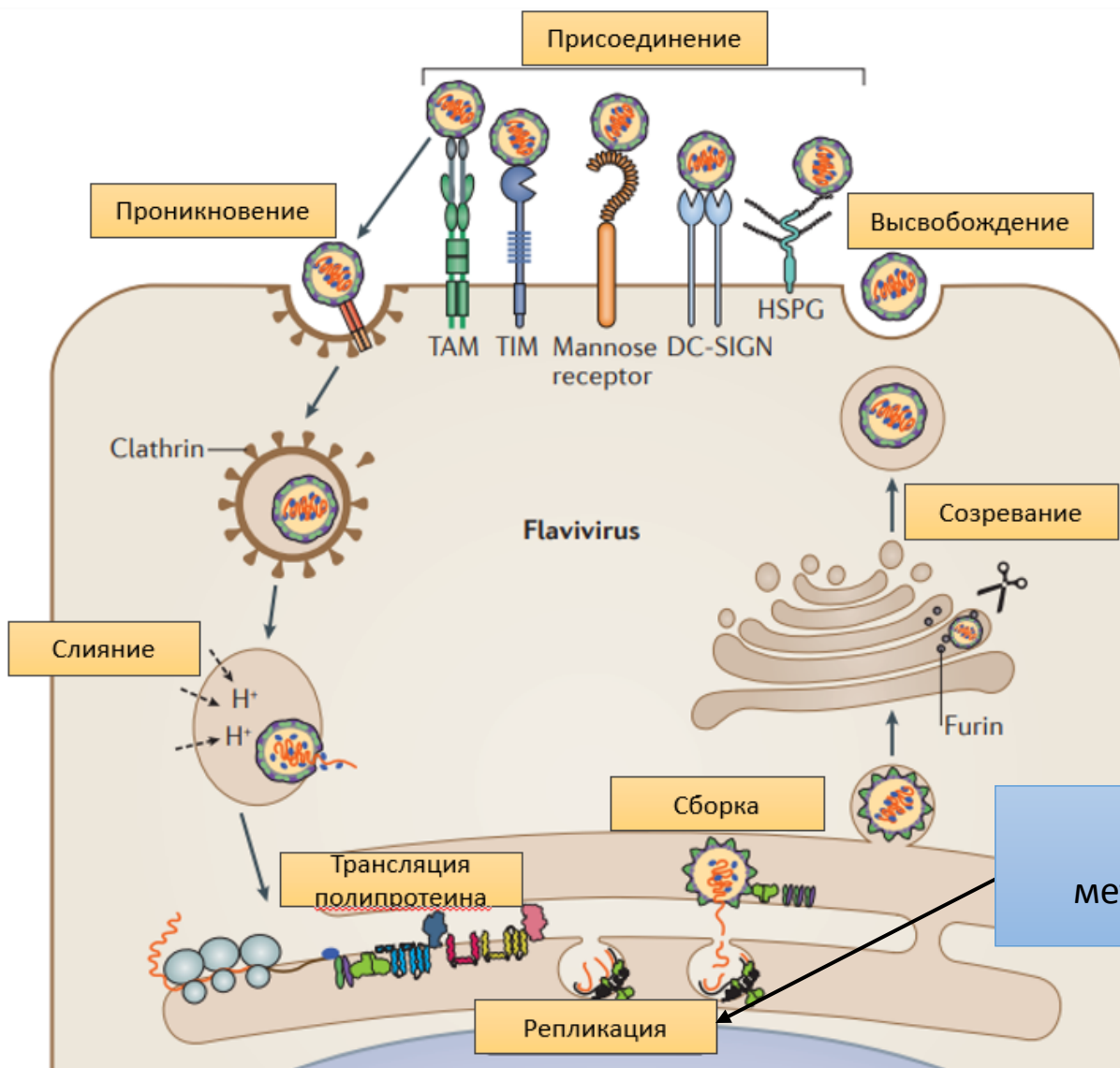
Клещевой энцефалит (КЭ) – вирусное инфекционное заболевание, поражающее центральную нервную систему человека. Вирус передается через укусы инфицированных клещей



	Случаи инфицирования в год
Вирус денге	300 млн
Вирус Западного Нила	100 тыс
Вирус желтой лихорадки	200 тыс
Вирус клещевого энцефалита	15 тыс

Относится к роду *Flavivirus* (семейство *Flaviviridae*) [1]

Жизненный цикл флавивирусов



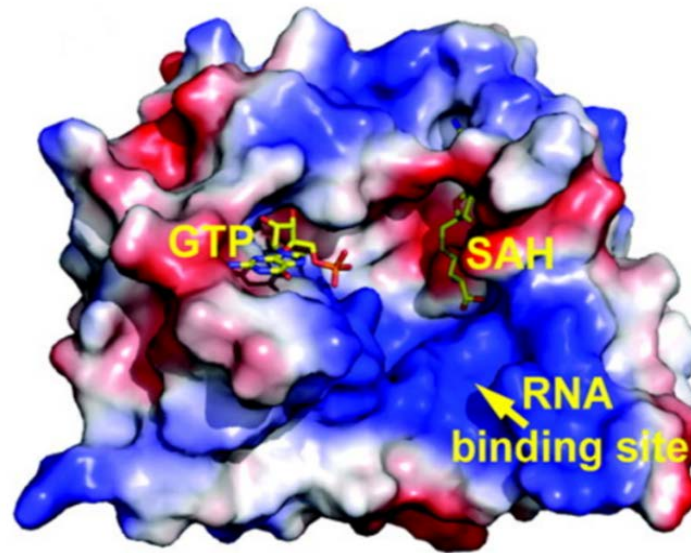
- МТ необходима для формирования структуры кэпа первого типа (m7GpppAm) генома вирусной РНК
- Действие фермента происходит на стадии репликации вирусного генома

Ингибиторы метилтрансферазы

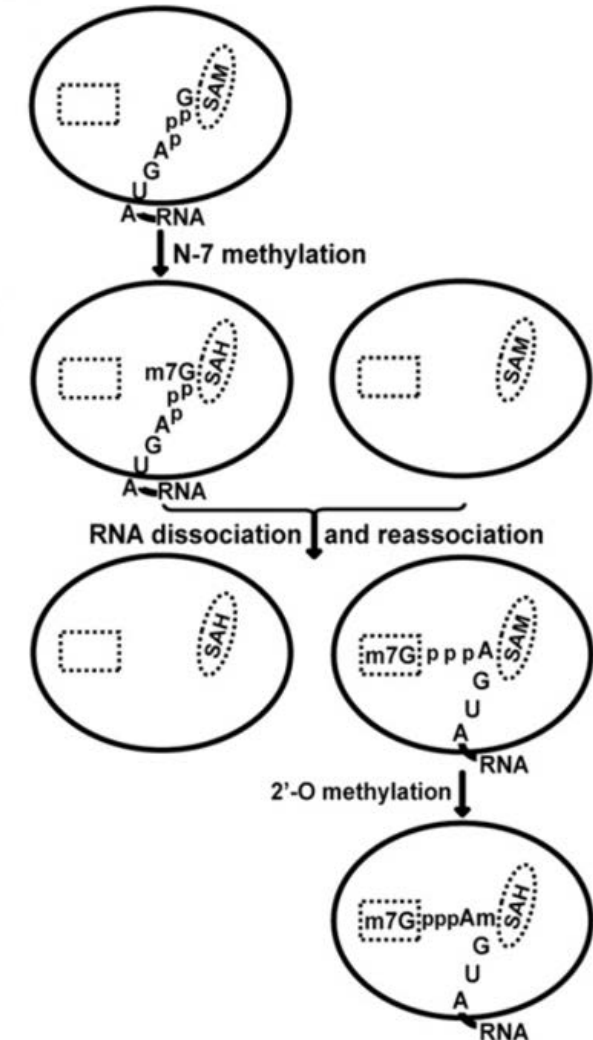
Механизм действия метилтрансферазы (МТ)

Метилирование вирусной РНК происходит по следующему механизму:

- Связывание молекулы SAM с МТ
- Метилирование гуанозина по положению N7, образование структуры кэп 0 (7meGpppA-РНК)
- Диссоциация комплекса
- Связывание кэп 0 (7meGpppA-РНК) с МТ в сайте связывания с GTP
- Метилирование аденозина в положении 2'-О, образование структуры кэп 1 (7meGpppA2'-O-me-RNA)



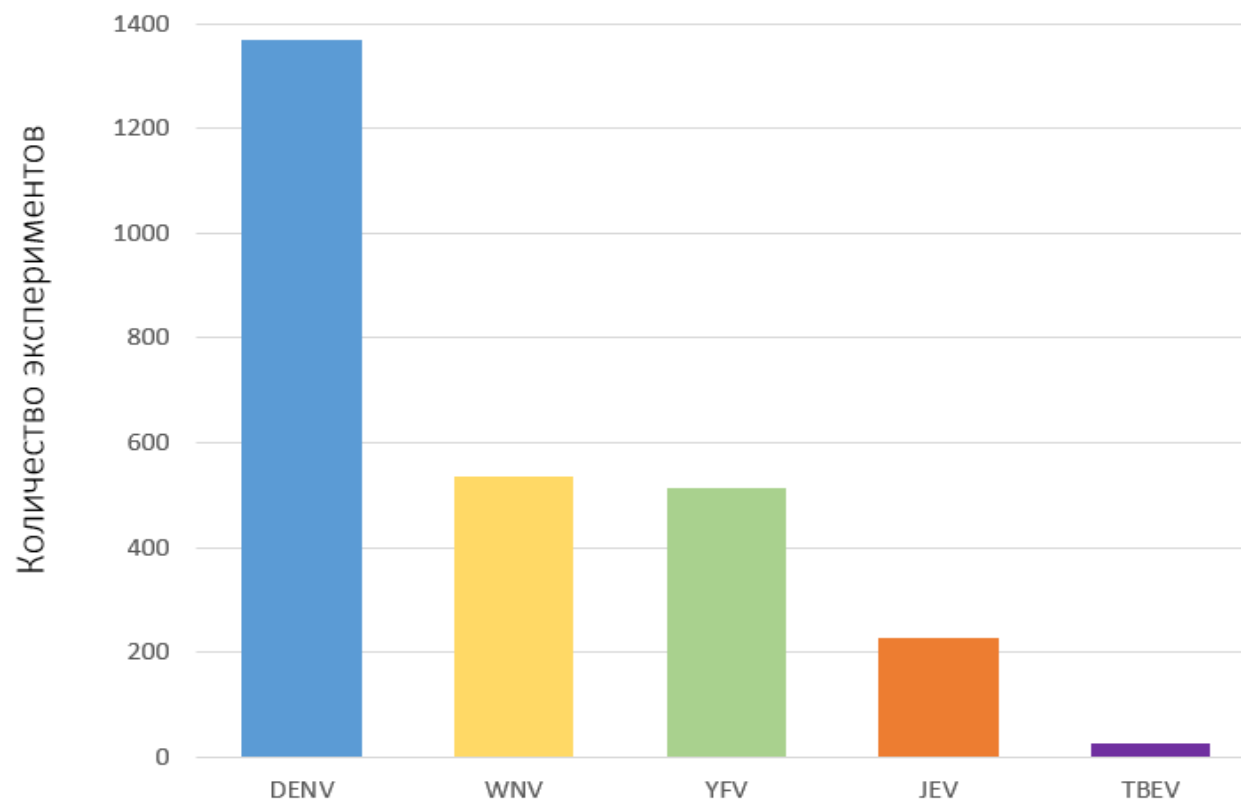
Сайты связывания МТ с субстратом [3]



Реакция метилирования [3]

Поиск исследований

Статистика по экспериментам противовирусной активности базы данных PubChem BioAssay



*Обозначения: DENV – вирус денге, WNV – вирус лихорадки Западного Нила, YFV – вирус Желтой лихорадки, JEV – вирус Японского энцефалита, TBEV – вирус клещевого энцефалита

Крупномасштабное исследование, проведенное Southern Research Institute [4]:

PubChem About Blog Submit Contact Search PubChem

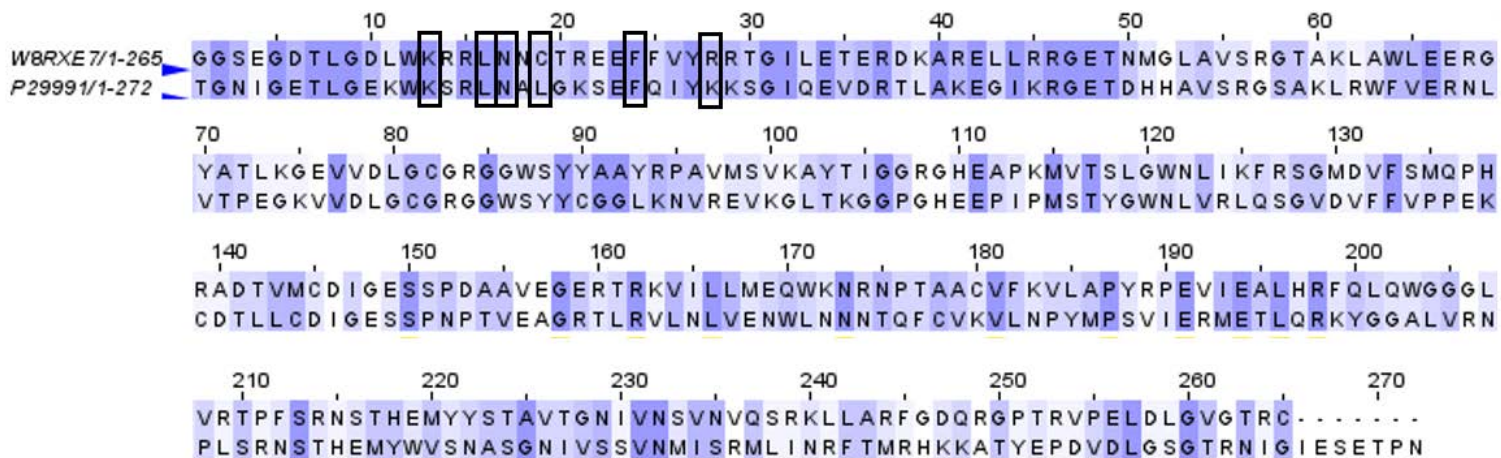
BIOASSAY RECORD

Primary and Confirmatory Screening for Flavivirus Genomic Capping Enzyme Inhibition

Cite Download

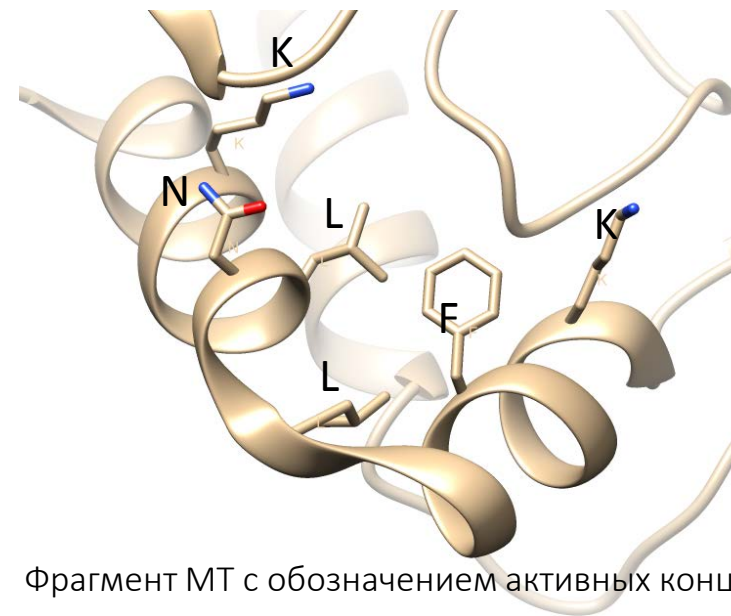
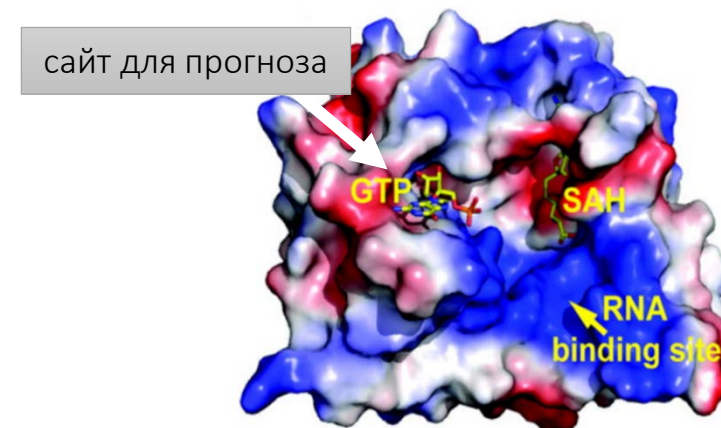
PubChem AID	588689
Protein Target	Chain A, RNA-directed RNA polymerase NS5 Taxonomy Dengue virus 2 16681-PDK53
Source	Southern Research Specialized Biocontainment Screening Center
External ID	CEGtase_01
BioAssay Type	Confirmatory
Tested Substances	All (338,853) Active (1,013) Inactive (336,307) Data Table
Tested Compounds	All (338,623) Active (1,013) Inactive (336,086)
Version	1.1 Revision History
Status	Live
Dates	Deposit: 2011-10-26 Modify: 2011-10-26

Сравнение аминокислотных последовательностей флавивирусов



ВКЭ Штамм Абсеттаров: код PDB W8RXE7,
 Вирус Денге 2: код PDB 29991

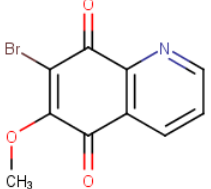
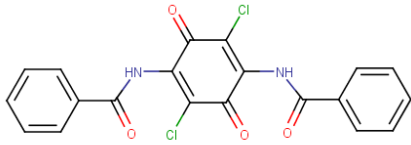
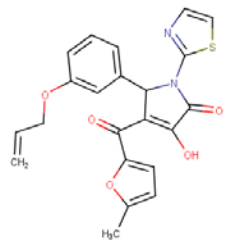
Процент идентичности аминокислотных остатков исследуемых флавивирусов равен 52%, но ответственные за связывание с ГТФ аминокислотные остатки – Lys13, Leu16, Asn17, Cys19, Phe24, Arg28 – достаточно консервативны у ДЕНВ и ВКЭ.



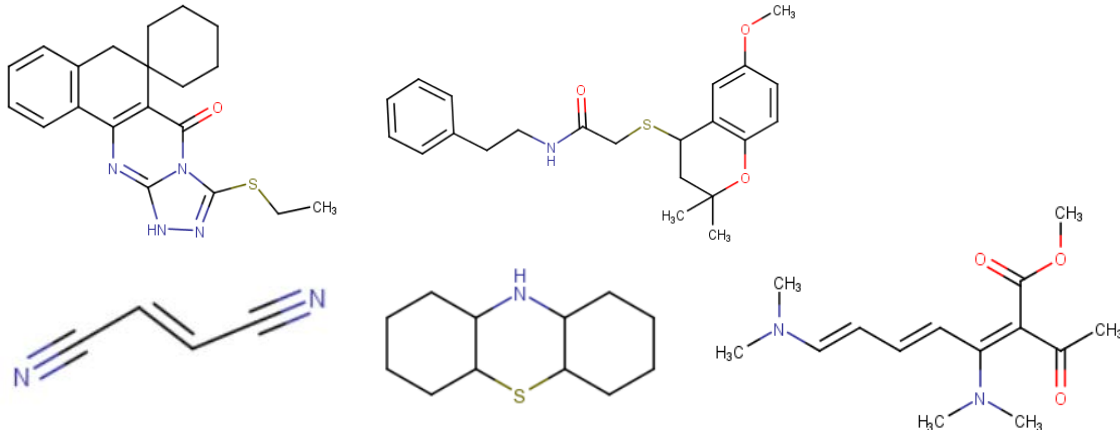
Фрагмент МТ с обозначением активных концов связывания

Выборка для обучения

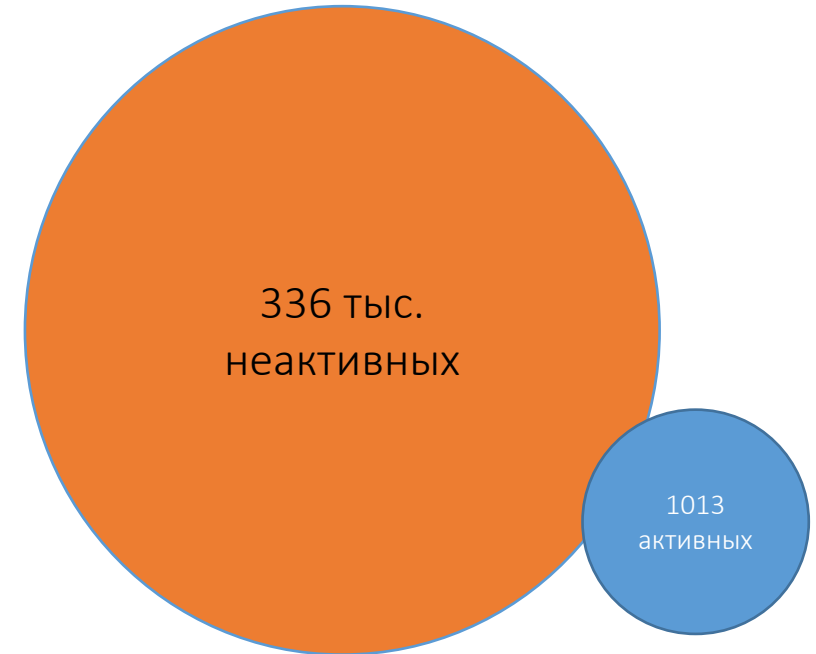
Наиболее активные соединения

CID: 265935	CID: 1897383	CID: 654648
 IC50 = 3.41 μ M	 IC50 = 3.53 μ M	 IC50 = 5.64 μ M

Неактивные соединения (PI<28.5)

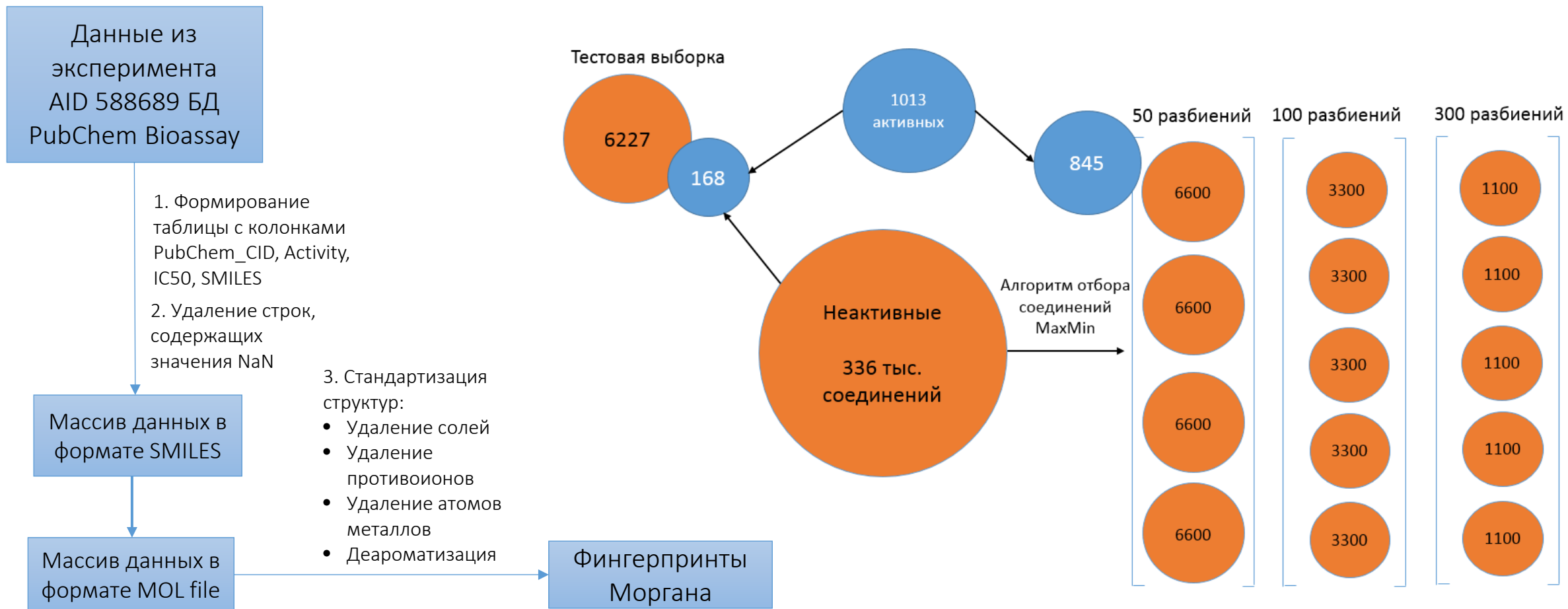


Проблема несбалансированности выборки



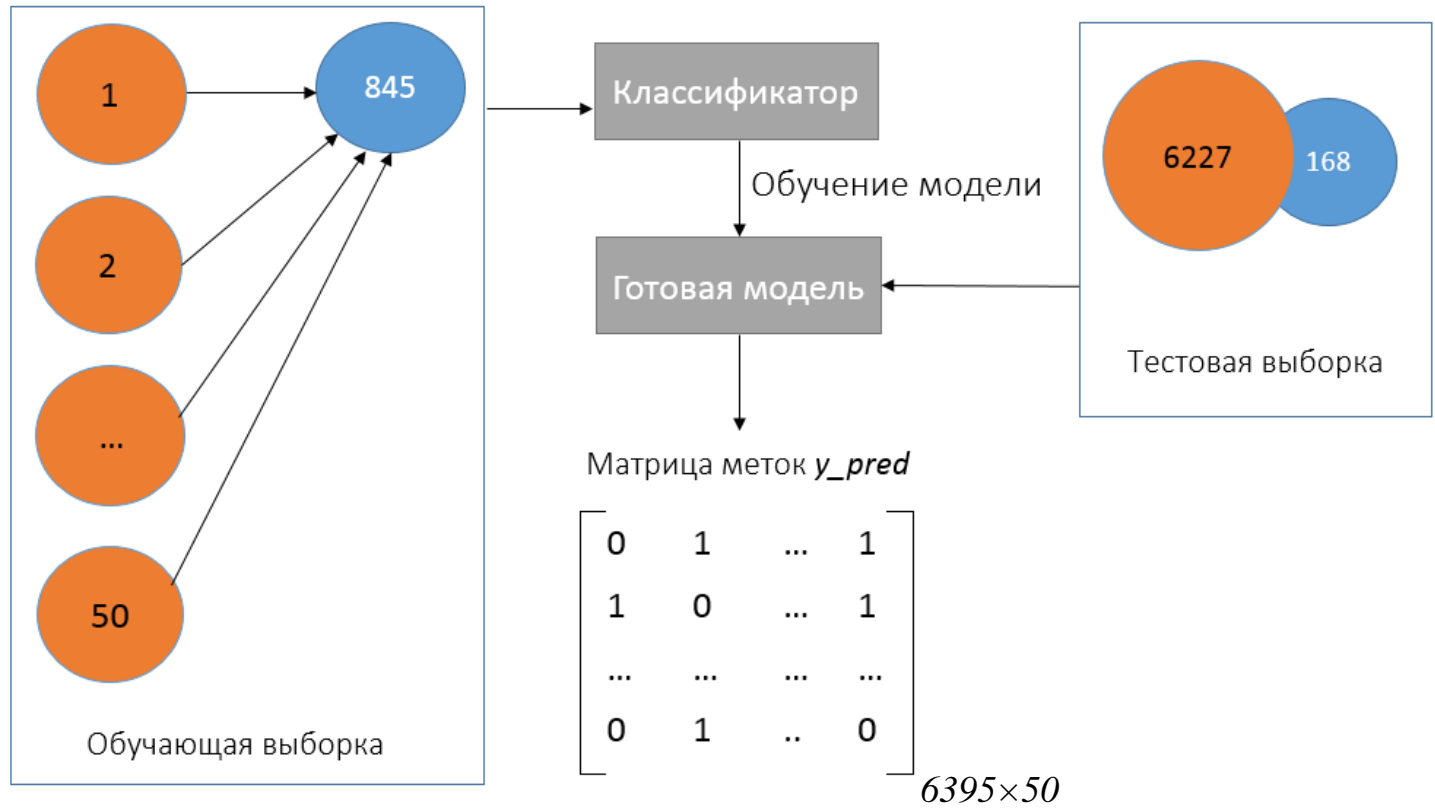
- Сокращение количества неактивных соединений до оптимального соотношения активных к неактивным (1:1)
- Подбор балансирующих коэффициентов на этапе обучения модели
- Разбиение набора неактивных соединений на подвыборки

Подготовка данных



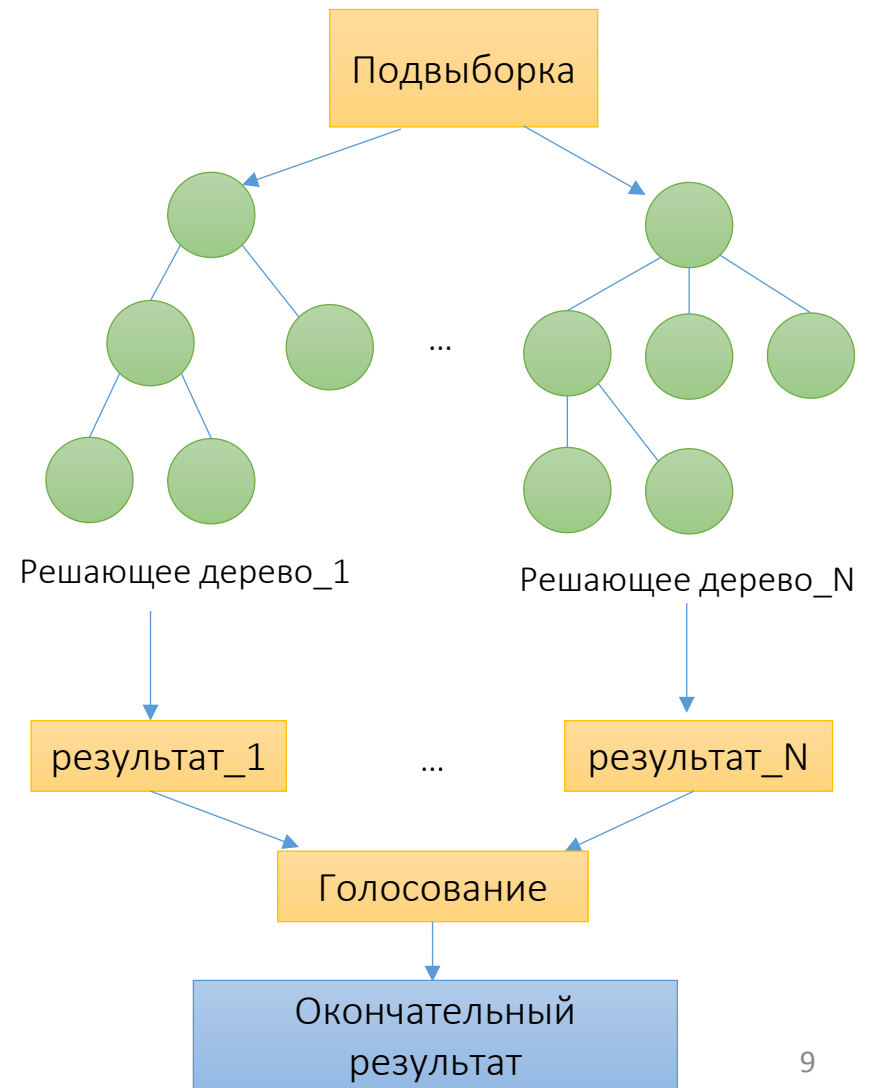
- Разбиения набора неактивных соединений на подвыборки

Обучение модели



50 – количество подвыборок
6395 – количество соединений
в тестовой выборке

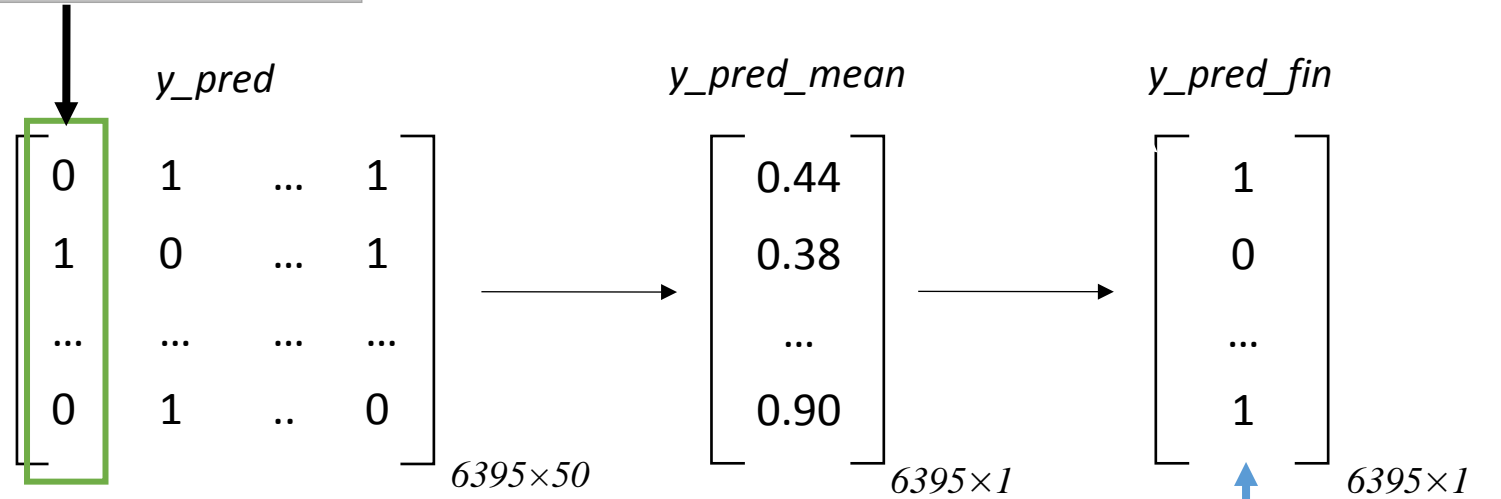
Метод случайного леса



Метод сдвига порога (Threshold Tuning)

Принцип:

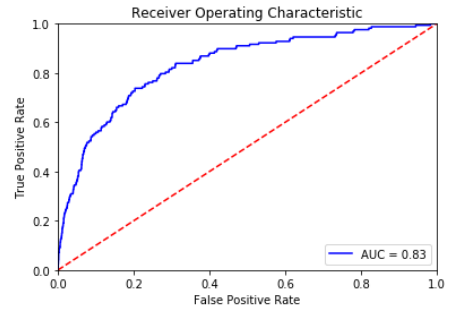
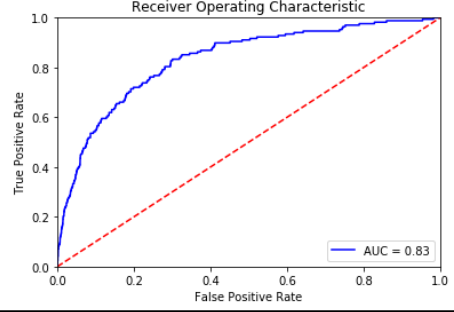
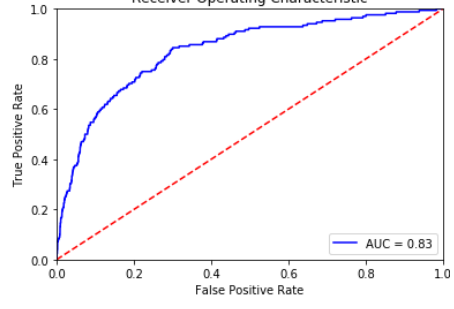
«Вердикт 1-го классификатора»



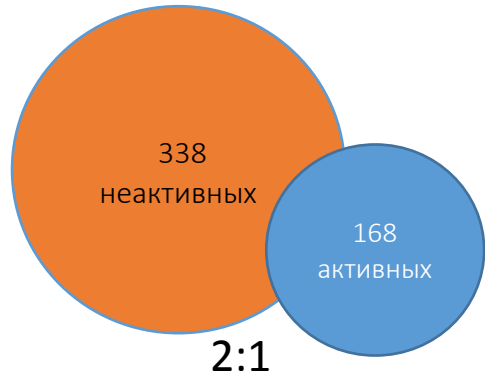
$$T = \frac{\sum_{i=1}^n y_i}{n}$$

Если $T > 0.4$, $T' := 1$

Анализ качества моделей

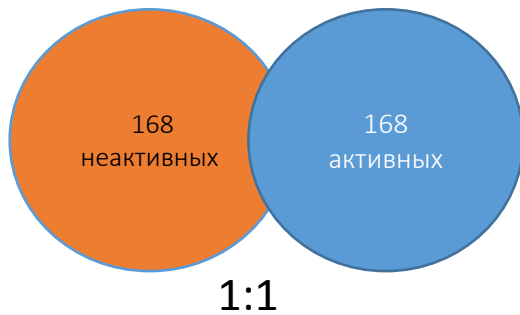
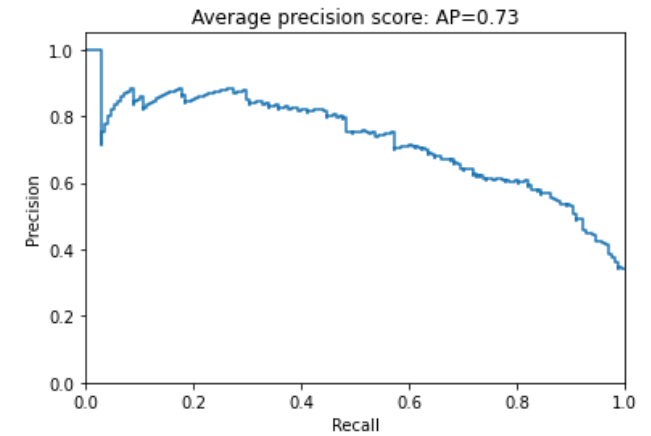
	Метка класса	Precision	Recall	F1-score	ROC	BEDROC
N = 50:	0	0.98	0.99	0.98		0.4879
	1	0.34	0.21	0.26		
N = 100:	0	0.99	0.99	0.98		0.4854
	1	0.25	0.27	0.26		
N = 300:	0	0.99	0.98	0.99		0.4909
	1	0.21	0.31	0.25		

Анализ качества моделей

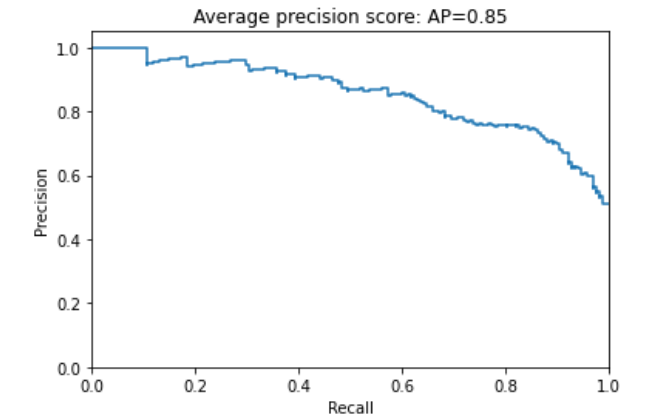


Метка класса	Precision	Recall	F1-score
0	0.85	0.75	0.80
1	0.60	0.74	0.66

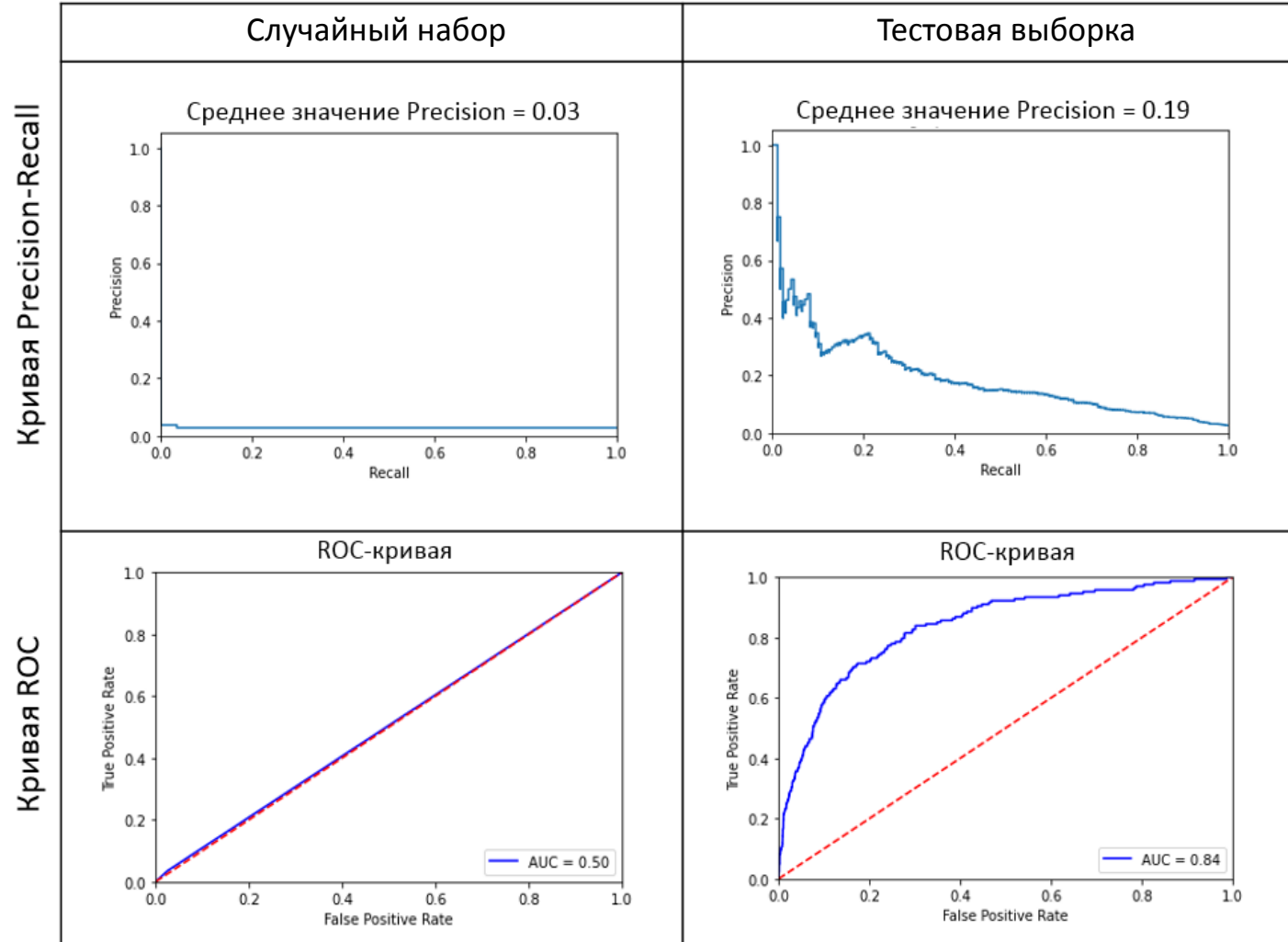
Кривые Precision-Recall



Метка класса	Precision	Recall	F1-score
0	0.75	0.77	0.76
1	0.77	0.74	0.75



Проверка случайности прогноза модели



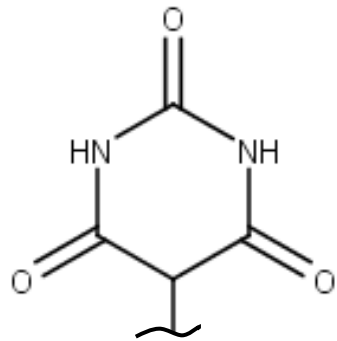
Применение модели на внешней выборке

- В качестве внешней выборки использовался набор соединений ФНЦИРИП им. М.П. им. М.П. Чумакова РАН, включающий в себя 1125 соединений.
- По результатам прогноза активность была предсказана для 19 соединений. Согласно ранее проведенным экспериментам 9 из них проявили ингибиторную активность по отношению к репродукции ВКЭ *in vitro* [5].

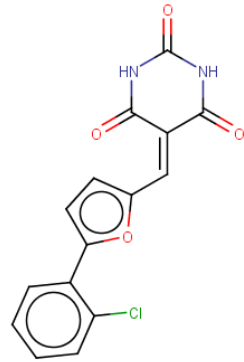
Противовирусная активность *in vitro* (ВКЭ, клетки СПЭВ)

ID	EC50, мкМ
STOCK1S-01217	>50
STOCK1S-17696	3,8252 ± 0,4989
STOCK1S-65797	>50
STOCK1S-80114	11,0212 ± 3,6898
STOCK1S-80492	>50
STOCK2S-24613	>50
STOCK2S-35871	1,0651 ± 0,4014
STOCK2S-37596	1,13 ± 0,174
STOCK2S-51582	>50
STOCK2S-61618	>50
STOCK2S-62110	>50
STOCK2S-63499	>50
STOCK3S-98701	21,691 ± 7,804
STOCK4S-04943	16,411 ± 3,369
STOCK4S-08490	1,869 ± 0,116
STOCK4S-13035	2,342 ± 0,060
STOCK4S-27505	>50
STOCK5S-01065	>50
STOCK5S-21328	10,569 ± 3,907

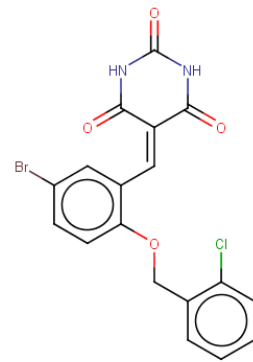
Классы соединений



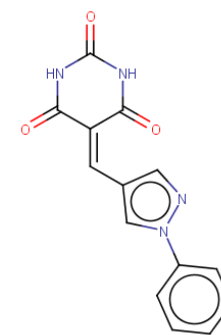
1,3-дiazинан-2,4,6-трион



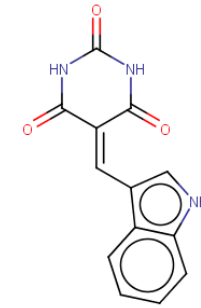
STOCK1S-17696
EC50 >50



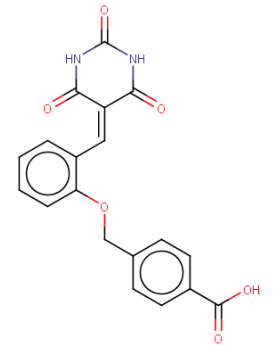
STOCK1S-80114
EC50 = 11 ± 4



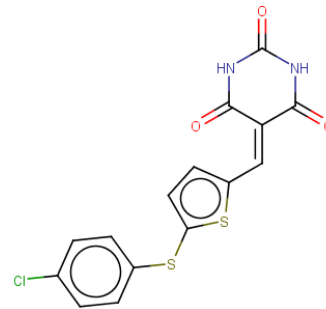
STOCK3S-98701
EC50 = 22 ± 8



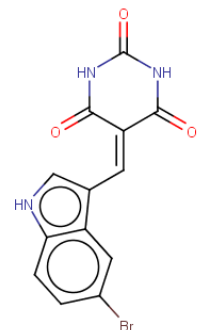
STOCK1S-01217
EC50 >50



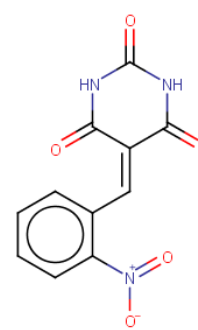
STOCK2S-24613
EC50 >50



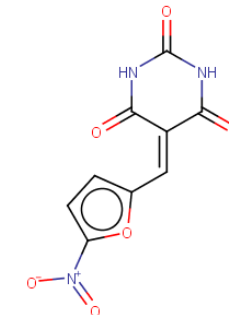
STOCK4S-08490
EC50 = 1,87 ± 0,12



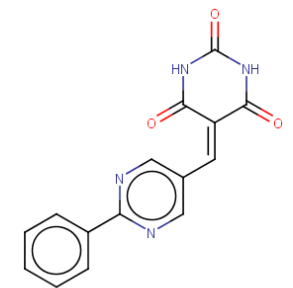
STOCK4S-13035
EC50 = 2,34 ± 0,06



STOCK1S-65797
EC50 >50

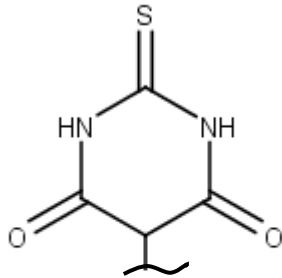


STOCK1S-80492
EC50 >50

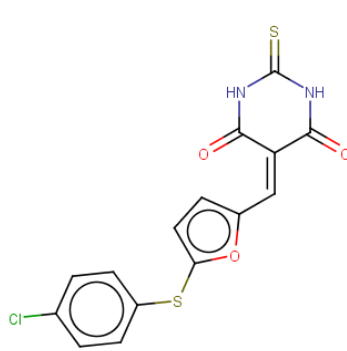


STOCK4S-27505
EC50 >50

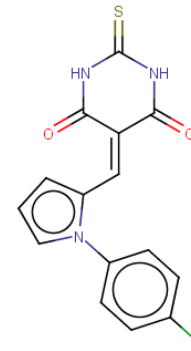
Классы соединений



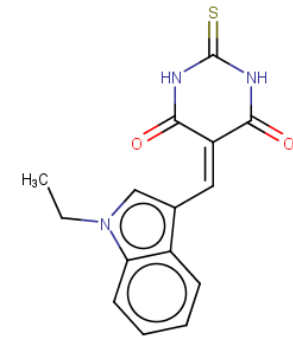
2-сульфанилиден-1,3-дiazинан-4,6-дион



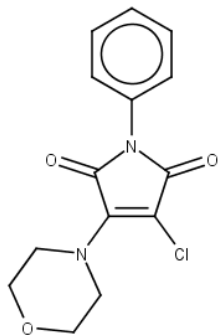
STOCK2S-35871
EC50 = 1,06 ± 0,40



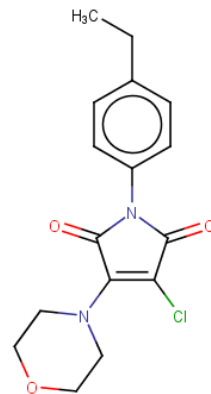
STOCK2S-37596
EC50 = 1,13 ± 0,17



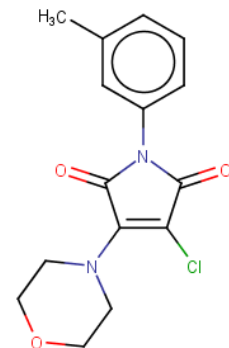
STOCK4S-04943
EC50 = 16,41 ± 3,37



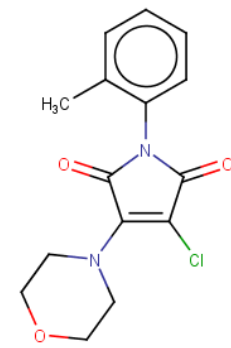
3-хлор-4-(морфолин-4)-1-фенил-2,5-
дигидро-1H-пиррол-2,5-дион



STOCK2S-61618
EC50 >50



STOCK2S-51582
EC50 >50



STOCK2S-63499
EC50 >50

Выводы

- Построена модель, позволяющая прогнозировать активность ингибиторов метилтрансферазы ВКЭ
- Проведен прогноз для внешней выборки соединений ФНЦИРИП им. М.П. им. М.П. Чумакова РАН
- Ведется разработка методики экспериментального определения активности ингибиторов метилтрансферазы ВКЭ

Благодарности

