Суперкомпьютерный поиск ингибиторов фактора свертывания крови XIIa в китайской национальной библиотеке соединений

Д.К. Кутов, А.В. Сулимов, А.С. Тащилова,
И.С. Ильин, В.Б. Сулимов

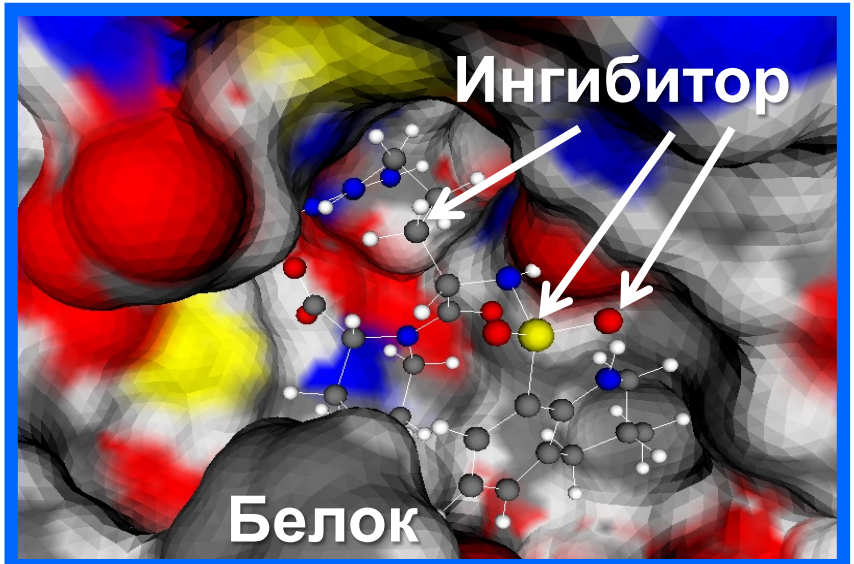
DIMONTA

ZKJT 张江集团 ZHANGJIANG GROUP

Работа поддержана грантом РНФ, Соглашение №21-71-20031

Парадигма действия лекарства

- Для многих заболеваний известны белки, определяющие развитие патологии
- Такие белки – мишени для действия лекарств
- Молекулы-ингибиторы лекарства связываются с активным центром белка и блокируют его работу



Блокирование белка приводит к излечению заболевания

Рациональная разработка лекарств

- Моделирование на атомно-молекулярном уровне увеличивает скорость разработки **ингибиторов**
- **Мощные суперкомпьютеры** дают возможность применять методы молекулярного моделирования
- Новые ингибиторы – основа для открытия **новых лекарств**

Суперкомпьютеры помогают найти молекулу лекарства

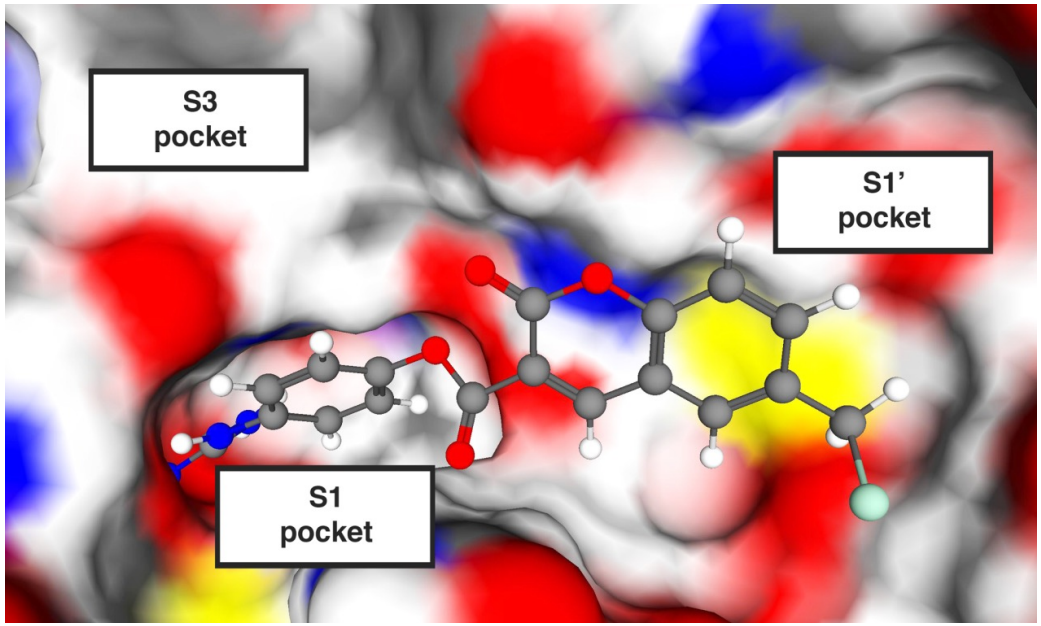
Антикоагулянты на основе ингибиторов фактора XIIa

- При дефиците фактора XIIa лабораторные животные значительно меньше страдают от развития индуцированного тромбоза
- Пациенты с дефицитом FXIIa демонстрируют нормальную гемостатическую способность
- Дефицит фактора XIIa защищает от тромбоза, не вызывая спонтанного кровотечения

Блокируют внутренний путь активации свертывания крови без нарушения внешнего пути и реакций плазменного гемостаза

Атомистическая модель фактора XIIa

Разработка новых ингибиторов фактора свертывания XIIa методами молекулярного моделирования



Активный сайт белка представлен карманами связывания: S1, S1' and S3

PDB ID 6B74

- **Структура белка закристаллизована с ингибитором**
- **Разрешение = 2.3 Å**
- **Успешный докинг нативного лиганда, RMSD = 1.93 Å**
- **Населенность первого кластера 49**
- **Скоринг-функция SOL -5.19 kcal/mol**

Китайская национальная библиотека соединений (CNCL)

国家化合物样品库
CHINESE NATIONAL COMPOUND LIBRARY

- Библиотека CNCL содержит около 1 миллиона низкомолекулярных органических соединений
- Библиотека соединений разбита на 7 частей в двухмерном формате SDF
- Все соединения из библиотеки CNCL доступны для заказа и экспериментального тестирования

Фильтрация и подготовка лекарственных соединений CNCL

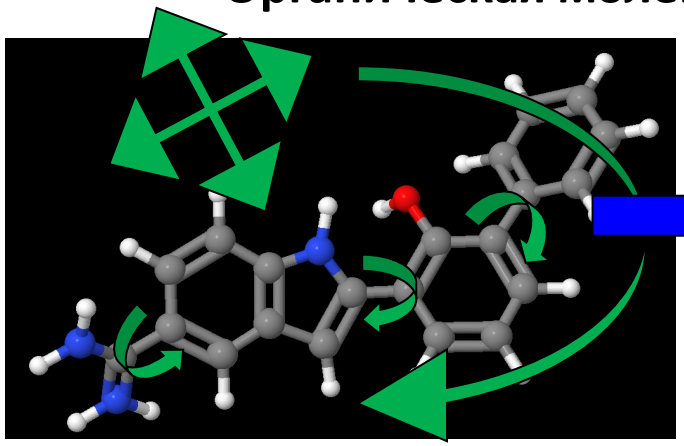
- Фильтрация по молекулярному весу (соединения легче 200 атомных единиц массы отбрасывались)
- Фильтрация по числу подвижных степеней свободы (использовалось число степеней свободы менее 10)
- Фильтрация функциональных групп, которые нежелательны в составе лекарств (оксимы и т.д.)
- Фильтрация атомов бора (бор не типизируется в используемом силовом поле MMFF94)
- Протонирование при pH = 7.4 и преобразование в трехмерный формат (LigPrep, Schrodinger Suite 2015)

Программа докинга SOL

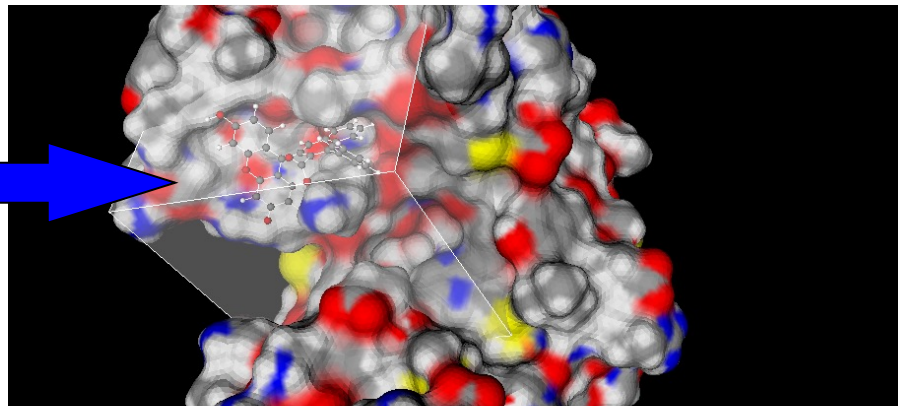
- Силовое поле MMFF94 фирмы Merck
- Предварительно рассчитанная сетка потенциалов для пробных атомов в поле белка (жесткий белок)
- Генетический алгоритм глобальной оптимизации
- Время расчетов на 1 CPU от 2 до 4 часов

Глобальная оптимизация на многомерной энергетической поверхности

Органическая молекула



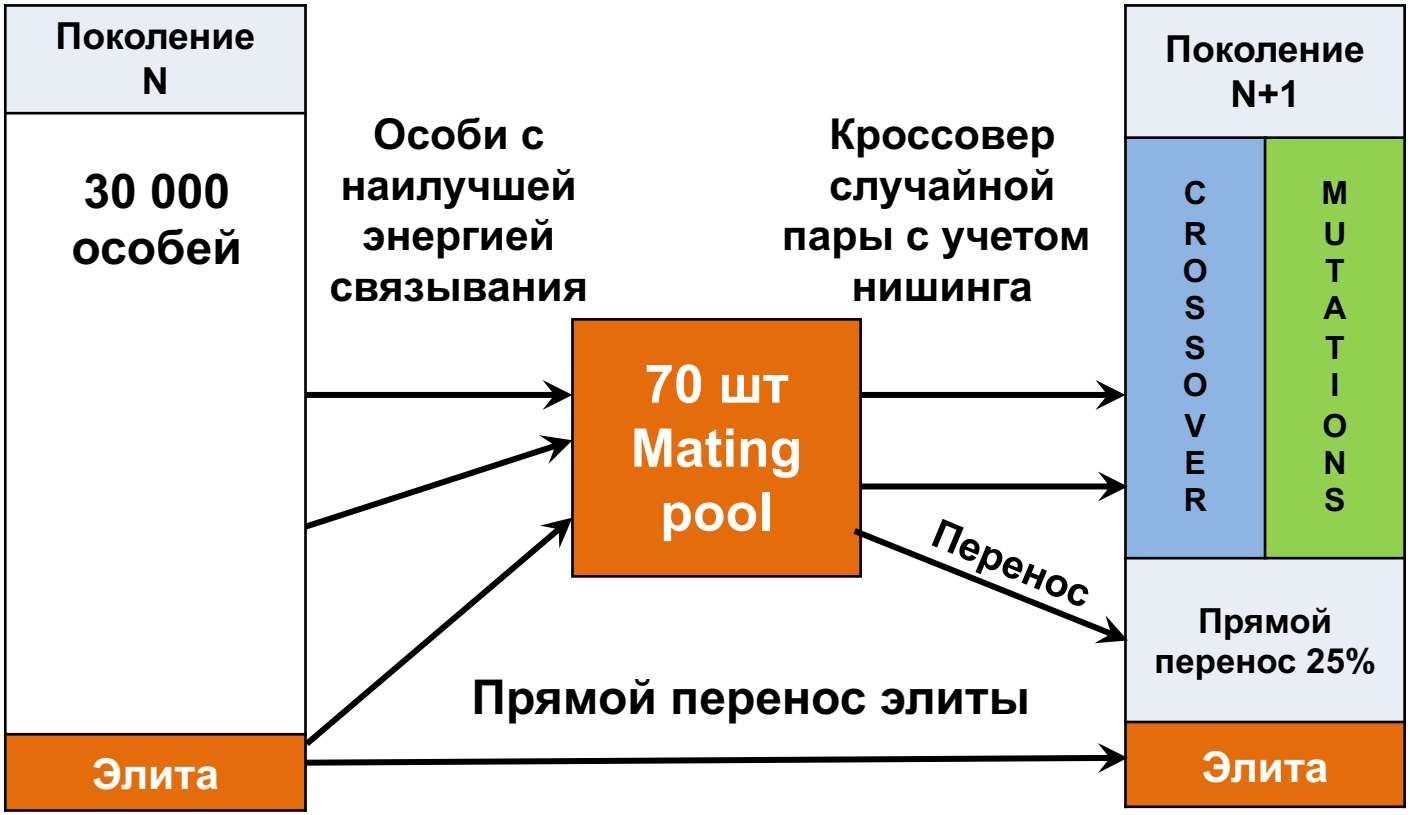
Активный центр белка-мишени



Для библиотеки CNCL необходимо 230 000 узлочасов

Параметры генетического алгоритма (ГА)

- **Количество запусков**
  - **50**
- **Размер популяции**
  - **30 000**
- **Число поколений**
  - **1 000**



Разработка быстрых параметров ГА

- Подготовка 32 новых параметров генетического алгоритма программы SOL (1 независимый запуск)
- Докинг 12 различных соединений базы ВГУ в Mpro SARS-CoV-2 (до 17 торсионных степеней свободы)
- Выбор 8 новых параметров по скорости

	Popul (K)	10%	STD	300	300	300	300	100	100	100	100	75	75	75	75	30	30	30	30	20	20	20	10	10	10	1	0.5	0.5	0.15	0.1	0.1	0.1	0.05	0.05	0.05
Best scor	N. of gen (K)	-10%	STD	1.5	1	0.75	0.5	1.5	1	0.75	0.5	1.5	1	0.75	0.5	1.5	1	0.75	0.5	1	0.75	0.5	1	0.75	0.5	2	20	2	30	100	50	20	100	20	2
VGY_0005893	17	-1.70	STD-1	114%	121%	86%	32%	75%	86%	88%	82%	80%	0%	93%	95%	68%	0%	0%	99%	21%	0%	-11%	0%	0%	21%	-31%	-28%	-26%	-74%	-60%	-40%	-82%	-176%	-50%	-196%	-296%
VGY_0003845	15	-2.20	STD-1	67%	31%	61%	53%	33%	13%	64%	29%	46%	24%	29%	20%	29%	13%	0%	-16%	14%	24%	-28%	0%	0%	-21%	20%	-23%	-28%	-41%	-51%	-57%	-45%	-128%	-87%	-160%	-138%
VGY_0006019	14	-2.30	STD-1	44%	38%	46%	36%	37%	40%	35%	38%	33%	43%	38%	12%	33%	0%	0%	43%	35%	0%	19%	0%	0%	33%	13%	0%	0%	-44%	-10%	38%	-35%	0%	-17%	-20%	-150%
VGY_0002846	12	-3.66	STD-1	49%	43%	49%	44%	44%	42%	0%	46%	45%	38%	40%	37%	45%	21%	0%	35%	43%	40%	13%	14%	-20%	0%	-12%	-37%	-47%	-63%	-33%	-27%	-78%	-55%	-32%	-104%	-143%
VGY_0003161	11	-4.72	STD-1	0%	0%	0%	0%	11%	-20%	-15%	-18%	-18%	-14%	0%	-11%	-19%	0%	0%	-16%	-11%	0%	0%	-23%	0%	-23%	-34%	-63%	-30%	-56%	-44%	-22%	-43%	-64%	-52%	-71%	-74%
VGY_0014738	9	-4.11	STD-1	47%	0%	33%	22%	13%	0%	0%	0%	48%	48%	46%	0%	0%	15%	0%	0%	-12%	0%	0%	0%	34%	0%	32%	-40%	-34%	-55%	-35%	0%	-28%	0%	-29%	-71%	-76%
VGY_0005008	8	-5.72	STD-1	0%	0%	0%	0%	0%	0%	-14%	0%	-16%	-16%	0%	0%	0%	-26%	0%	-16%	-17%	0%	0%	0%	0%	0%	0%	-18%	-31%	-32%	-32%	-26%	-18%	-34%	-25%	-34%	-74%	-98%
VGY_0004990	7	-5.52	STD-1	22%	0%	0%	17%	0%	0%	0%	17%	0%	17%	0%	0%	0%	0%	0%	16%	-15%	17%	17%	16%	0%	0%	0%	-24%	-21%	-30%	16%	0%	-13%	-30%	-37%	-46%	-43%
VGY_0015030	6	-6.17	STD-1	0%	0%	-15%	0%	-20%	0%	0%	0%	0%	0%	-12%	0%	-16%	0%	0%	0%	0%	0%	-13%	-20%	0%	0%	0%	-31%	-27%	-13%	-18%	-27%	-28%	-32%	-15%	-45%	-87%
VGY_0015029	5	-6.53	STD-1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-19%	-34%	-20%	-46%	-19%	-38%	0%	-21%	-27%	-48%	-52%
VGY_0006593	4	-5.98	STD-1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-17%	-13%	-13%	-21%	-13%	-21%	-18%	-15%	-26%	-43%
VGY_0016121	3	-6.40	STD-1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-15%	-21%	-22%	-17%	-24%	-14%	0%	-28%	-47%

	Popul (K)	2	STD	300	300	300	300	100	100	100	100	75	75	75	75	30	30	30	30	20	20	20	10	10	10	1	0.5	0.5	0.15	0.1	0.1	0.1	0.05	0.05	0.05
Time, m	N. of gen (K)	0.5	STD	1.5	1	0.75	0.5	1.5	1	0.75	0.5	1.5	1	0.75	0.5	1.5	1	0.75	0.5	1	0.75	0.5	1	0.75	0.5	2	20	2	30	100	50	20	100	20	2
VGY_0005893	17	17	STD-1	0.02	0.07	0.11	0.16	0.19	0.22	0.31	0.4	0	0.31	0.46	0	0	0	0	0	0	2.83	0	2.43	4.25	3.4	4.25	8.5	29.1	5.67	51	9.36	3.58	6.99	18.55	5.15	26.15	340
VGY_0003845	15	11	STD-1	0.02	0.06	0.07	0.1	0.15	0.15	0.31	0.31	0.42	0.2	0.32	0	0	0	0	0	0	0	0	0	2.75	2.2	3.67	5.5	19.4	4.13	38.8	6.47	3.33	4.82	12.94	3.53	17.37	165
VGY_0006019	14	17	STD-1	0.02	0.07	0.1	0.13	0.19	0.21	0.31	0.43	0	0.38	0.49	0	0	0	0	0	0	2.13	0	2.13	3.4	3.4	4.25	8.5	23.7	5.26	48.6	9.62	3.54	7.13	19.25	6.11	29.14	340
VGY_0002846	12	14	STD-1	0.02	0.07	0.09	0.13	0.2	0.22	0.36	0.38	0	0.25	0.48	0	0	0	0	0	0	0	0	0	3.5	3.5	4.67	7	24	5.79	49.4	7.71	3.27	6.94	15.85	5.12	25.45	280
VGY_0003161	11	14	STD-1	0.02	0.07	0.15	0.18	0.22	0.22	0.31	0	0	0.35	0.42	0	0	0	0	0	0	2.8	2.33	3.5	4.67	3.5	4.67	7	25.5	5.15	49.4	8.75	3.28	7.57	18.26	5.68	28.97	280
VGY_0014738	9	16	STD-1	0.02	0.06	0.11	0.13	0.22	0.21	0.31	0	0	0.25	0.42	0.47	0	0	0	0	0	0	0	2.29	5.33	4	4	8	27.4	5.25	45.7	10.4	3.71	6.76	16.27	5.52	28.24	240
VGY_0005008	8	13	STD-1	0.02	0.06	0.09	0.16	0.19	0.25	0.28	0.46	0	0.29	0.42	0	0	0	0	0	0	2.17	0	2.17	3.25	4.33	4.33	6.5	27.9	4.67	45.9	8.13	2.97	5.95	17.33	5.31	26	390
VGY_0004990	7	14	STD-1	0.02	0.09	0.09	0.11	0.25	0.18	0.28	0.38	0	0.26	0.38	0	0	0	0	0	0	2.33	0	0	2.8	4.67	4.67	7	33.6	5.6	52.5	9.44	3.77	6.56	18.26	7.37	26.25	420
VGY_0015030	6	6	STD-1	0.02	0.04	0.07	0.11	0.13	0.14	0.21	0.23	0	0.26	0.25	0.46	0	0	0	0	0	0	0	0	3	3	6	6	15	2.67	18	4.39	0	3.43	8.78	3.36	16.36	120
VGY_0015029	5	9	STD-1	0.02	0.06	0.1	0.12	0.28	0.22	0.3	0.38	0	0.24	0.41	0	0	0	0	0	2.25	2.25	0	3	4.5	3	4.5	9	17.4	2.84	31.8	5.14	2.8	5.93	11.49	6.51	38.57	270
							0.18	0.17	0.29	0.36	0	0.26	0.4	0	0	0	0	0	0	2.5	0	2.5	3.33	3.33	5	10	17.1	2.51	24	5.31	2.38	3.87	11.11	3		
							0.26	0.19	0.29	0.42	0	0.32	0.42	0	0	0	0	0	0	2.2	2.2	2.75	3.67	3.67	5.5	11	14	2.53	24.4	4.96	2.39	4.1	11.19	5		

Оценка качества новых параметров ГА

- 8 параметров генетического алгоритма программы SOL (5 независимых запусков)
- Докинг 14 соединений базы ВГУ, отобранных на экспериментальную проверку в Mpro SARS-CoV-2
- Выбор 2 новых параметров

	Popul (K)	3%	STD	50	30	20	10	8	5	5	3
	N. of gen (K)	-3%	STD	0.5	1	1	1	1	2	1	1
VGY_0015946	7	-7.31	STD	0%	0%	-4%	0%	0%	0%	0%	-6%	-42%
VGY_0016826	6	-6.63	STD	0%	-10%	-11%	-11%	-3%	-4%	0%	-5%	-6%
VGY_0005098	5	-7.25	STD	0%	-9%	-9%	-12%	0%	-14%	-10%	-10%	-9%
VGY_0015689	5	-6.94	STD	0%	0%	0%	0%	0%	-4%	0%	0%	0%
VGY_0018302	5	-6.91	STD	0%	0%	0%	0%	0%	0%	0%	0%	0%
VGY_0005103	4	-7.54	STD	0%	-9%	-9%	-8%	-11%	-9%	-8%	-9%	-9%
VGY_0006276	4	-6.60	STD	0%	0%	0%	0%	0%	0%	0%	0%	-11%
VGY_0006563	4	-7.16	STD	0%	0%	0%	0%	0%	0%	0%	0%	0%
VGY_0013885	4	-6.64	STD	0%	0%	0%	0%	0%	0%	0%	0%	-10%
VGY_0002900	3	-6.72	STD	0%	0%	-3%	0%	0%	-9%	0%	-9%	0%
VGY_0002405	2	-6.95	STD	0%	0%	0%	0%	-3%	0%	0%	0%	0%
VGY_0002425	2	-7.13	STD	0%	0%	0%	0%	0%	0%	0%	0%	0%
VGY_0003063	2	-7.00	STD	0%	0%	0%	0%	0%	0%	0%	0%	0%
VGY_0003291	2	-7.37	STD	0%	0%	0%	0%	0%	0%	0%	0%	0%

STD	50	30	20	10	8	5	5	3
STD	0.5	1	1	1	1	2	1	1
0	14	11	16	34	39	33	67	138
0	14	11	16	33	45	32	65	117
0	13	11	16	32	40	32	63	116
0	14	11	16	32	40	32	64	115
0	15	11	16	32	39	32	63	98
0	13	11	16	33	41	33	66	115
0	13	10	16	32	40	33	62	135
0	13	11	16	32	41	33	64	115
0	13	11	16	32	39	32	63	106
0	13	11	16	32	39	32	64	114
0	13	11	17	33	41	34	68	117
0	13	11	16	33	41	34	66	110
0	12	10	15	31	39	31	62	107
0	13	11	16	33	41	33	64	104

Определение быстрых параметров ГА

- 2 параметра генетического алгоритма программы SOL (5 независимых запусков)
- Докинг полной базы ВГУ (14 000 соединений / 42 000 конформеров) в Mpro SARS-CoV-2
- Отбор 20 соединений, посланных на эксперимент при использовании стандартных параметров SOL

Ligname	# 30K-1K (STD)	Best scoring STD	N. of run	# 10K-1K	Best scoring 10K-	N. of run	# 5K-2K	Best scoring 5K-2K	N. of run
VGY_0224133	1	-7.83	50	1	-7.81	5	2	-7.77	5
VGY_0041452	2	-7.78	50	2	-7.78	5	1	-7.78	5
VGY_0005103	6	-7.54	50	130	-6.86	5	95	-6.92	5
VGY_0015946	21	-7.31	50	145	-6.81	5	172	-6.76	5
VGY_0225101	35	-7.22	50	76	-6.98	5	68	-6.99	5
VGY_0006563	40	-7.16	50	31	-7.16	5	31	-7.15	5
VGY_0223350	42	-7.14	50	35	-7.13	5	32	-7.14	5
VGY_0002425	43	-7.13	50	36	-7.13	5	35	-7.13	5
VGY_0019038	91	-6.98	50	90	-6.95	5	87	-6.96	5
VGY_0015689	114	-6.94	50	57	-7.03	5	137	-6.83	5

Ligname	# 30K-1K (STD)	Best scoring STD	N. of run	# 10K-1K	Best scoring 10K-	N. of run	# 5K-2K	Best scoring 5K-2K	N. of run
VGY_0018302	128	-6.91	50	97	-6.93	5	107	-6.90	5
VGY_0214225	200	-6.77	50	168	-6.77	5	166	-6.76	5
VGY_0002900	237	-6.72	50	632	-6.40	5	327	-6.56	5
VGY_0013355	242	-6.72	50	241	-6.67	5	214	-6.70	5
VGY_0031572	295	-6.66	50	244	-6.67	5	223	-6.69	5
VGY_0016826	334	-6.63	50	314	-6.60	5		-5.92	5
VGY_0045297	350	-6.61	50		-5.69	5		-6.17	5
VGY_0006276	365	-6.60	50	274	-6.64	5	395	-6.51	5
VGY_0006802	477	-6.51	50	421	-6.51	5		-6.23	5
VGY_0015693	781	-6.37	50		-6.25	5		-6.13	5

Виртуальный скрининг базы CNCL

1. Докинг на быстрых параметрах ГА

- 977 240 соединений
- Подготовка 977 240 трехмерных конформеров
- Докинг программой SOL (быстрые параметры)
- Критерии отбора соединений
  - SOL score < −5.85 ккал/моль & 1-ый кластер ≥ 1 (из 5)
- Отобрано 9 113 соединений
- Время расчетов 7 236 узлочасов

Виртуальный скрининг базы CNCL

2. Докинг на стандартных параметрах ГА

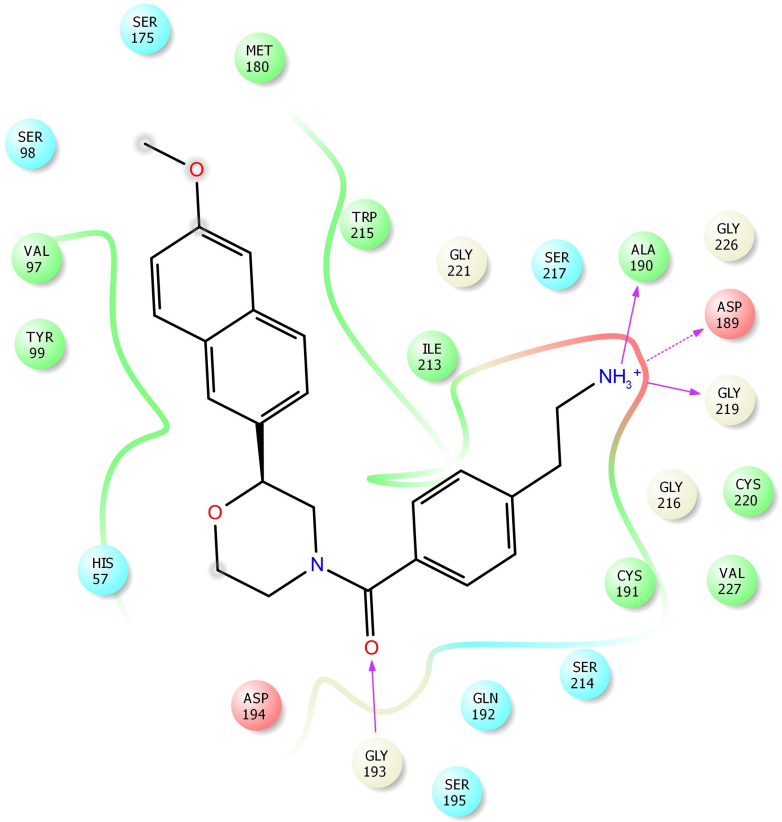
- 9 113 соединений
- Подготовка 22 477 трехмерных конформеров
- Докинг программой SOL (стандартные параметры)
- Критерии отбора соединений
  - SOL score < −8.5 ккал/моль & 1-ый кластер ≥ 1 (из 50)
  - SOL score < −7.3 ккал/моль & 1-ый кластер ≥ 30 (из 50)
- Отобрано 162 соединения
- Время расчетов 950 узлочасов

Виртуальный скрининг базы CNCL

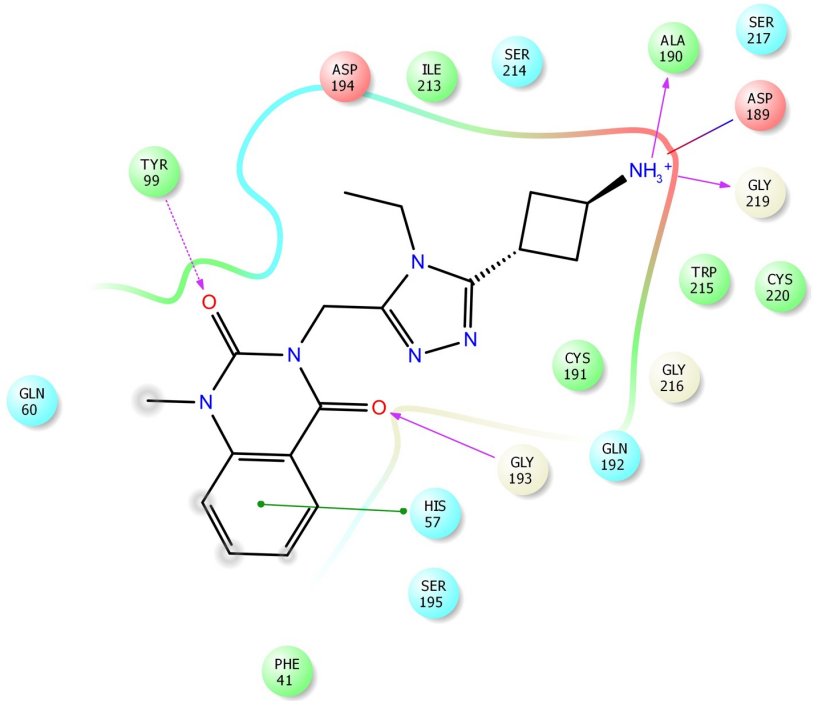
3. Расчет энтальпии связывания (MOPAC)

- Подготовка 162 комплексов и 474 конформеров
- Расчеты квантово-химическим полуэмпирическим методом PM7, с учетом воды в модели COSMO
  - $\Delta H_{bind} = E_{Complex} – E_{Ligand} – E_{Protein}$
  - Для комплекса используется лучшая поза после докинга
  - Для лиганда используются все начальные конформации
- Отобрано 18 соединений
  - Отбор по значениям скора SOL и энтальпии связывания
  - Анализ взаимодействия и создания водородных связей
- Время расчетов 32 узлочаса

Взаимодействие найденных кандидатов в ингибиторы с аминокислотными остатками



CL01982900
(4 водородные связи)



CL01813516
(4 водородные связи, 1 пи-стекинг)

Выводы

- Для программы докинга SOL разработаны быстрые параметры генетического алгоритма, которые ускоряют расчеты в 30 раз
- Разработана новая методика виртуального скрининга сотен тысяч соединений с использованием программы докинга SOL
- Использование миллионной библиотеки CNCL позволило выявить разнообразные химические кандидаты в ингибиторы: 18 выбранных лигандов принадлежат к 5 различным химическим классам
- Выявление специфических взаимодействий с белком-мишенью: найденные кандидаты в ингибиторы имеют от 2 до 7 водородных связей с белком

Спасибо за внимание

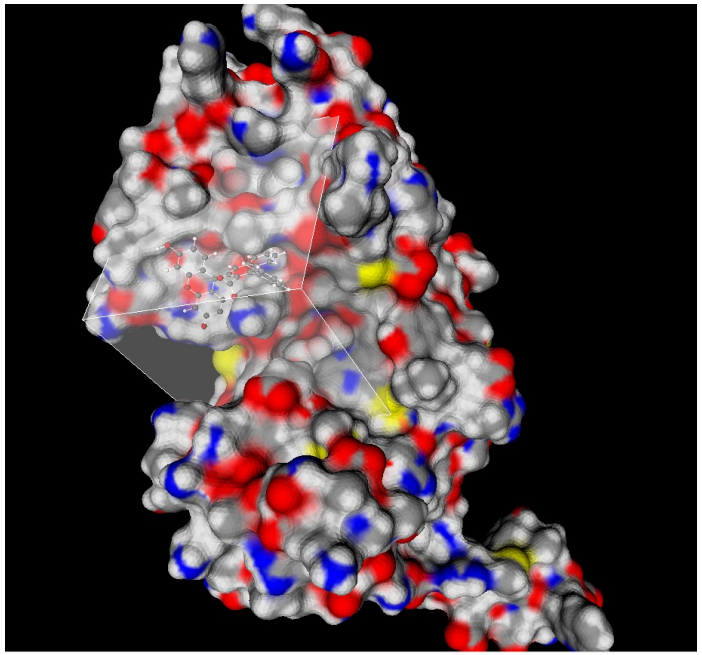
... Каждый медицинский метод есть инновация; а кто не хочет применять новые средства, должен ждать новых бед ...

Francis Bacon (1561-1626)
OF INNOVATIONS

Работа поддержана грантом РНФ, Соглашение №21-71-20031

Докинг (SOL)

- Силовое поле MMFF94 фирмы Merck
- Жесткий белок
- Предварительно рассчитанная сетка потенциалов для пробных атомов в поле белка
  - Кулоновское взаимодействие
  - Ван-дер-ваальсово взаимодействие
  - Энергия десольватации – обобщенная модель Борна



- Генетический алгоритм глобальной оптимизации
- Целевая функция: E (лиганда в поле белка) + E (внутренних напряжений лиганда)

Постпроцессинг (MOPAC)

- Расчет энтальпии связывания ΔH_{bind} новым квантово-химическим полуэмпирическим методом PM7, с учетом воды в модели COSMO
  - Улучшенные дисперсионные взаимодействия
  - Улучшенное описание водородных и галогенных связей
- ΔH_{bind} = E(Complex) – E(Ligand) – E(Protein)
  - Оптимизация в вакууме методом PM7
  - Для комплекса используется лучшее положение после докинга
  - Для лиганда используются все начальные конформации
- Отбор соединений на эксперимент