

**Докинг и постпроцессинг 1 миллиона  
молекул базы данных CNCL для разработки  
ингибиторов Mpro SARS-CoV-2**

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



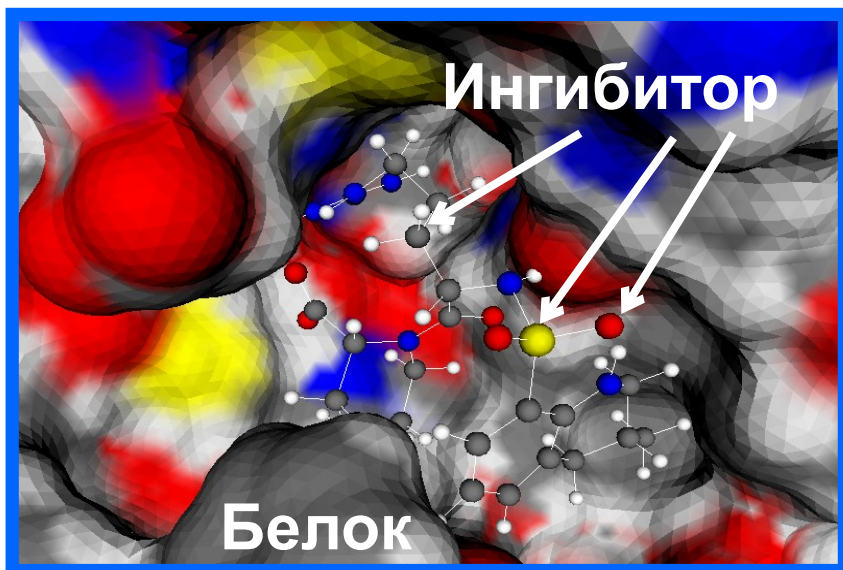
**Работа поддержана грантом РФФ, Соглашение №21-71-20031**

# Обзор публикаций по разработке препаратов прямого действия на вирус SARS-CoV-2

- ▶ К концу 2024 года опубликовано **около 1 тысячи** статей на тему разработки противовирусных препаратов прямого действия вируса SARS-CoV-2
- ▶ **Молекулярный докинг** является основным инструментом обнаружения хитов
- ▶ Для поиска ингибиторов все чаще используются **ресурсы суперкомпьютеров**
- ▶ Широко используются **базы данных имеющихся в наличии** низкомолекулярные соединения
- ▶ Избегание синтеза новых соединений на начальном этапе разработки лекарств сокращает время

# 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**

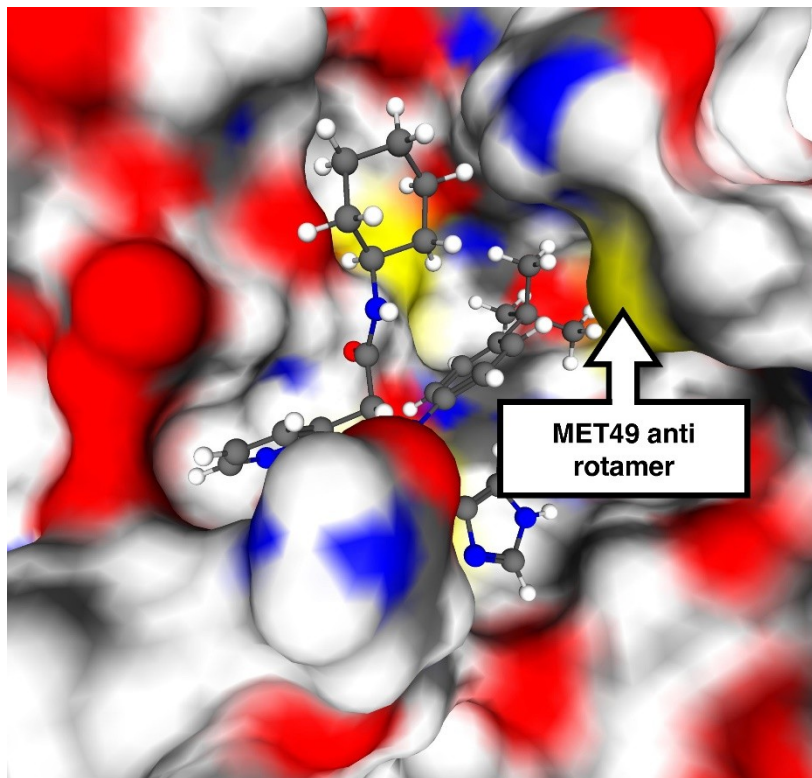


Суперкомпьютеры  
помогают найти  
молекулу лекарства

# Терапевтические мишени коронавируса SARS-CoV-2

- ▶ **Главная протеаза (3CLpro или Mpro)**
- ▶ **Папаино-подобная протеаза (PLpro)**
- ▶ **Вирусный неструктурный протеин nsp14**
- ▶ **Неструктурная эндорибонуклеаза nsp15**
- ▶ **Вирусная 2'-О-метилтрансфераза (nsp16)**
- ▶ **Вирусная геликаза (helicase)**
- ▶ **РНК-зависимая РНК-полимераза (RdRP)**
- ▶ **S-протеин (Spike-протеин)**

# Первая модель главной протеазы SARS-CoV-2

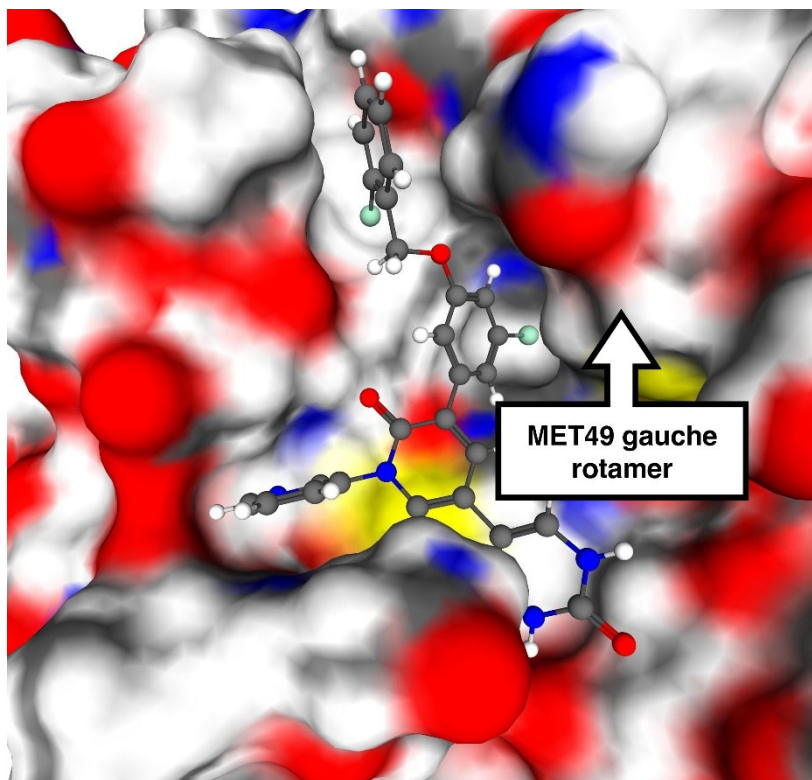


## PDB ID **6W63**

- ▶ Разрешение = 2.1 Å
- ▶ Закристаллизован с нековалентным ингибитором **X77** (IC<sub>50</sub> = 4.1 μM)
- ▶ Докинг нативного лиганда, RMSD = 1.31 Å
- ▶ Населенность 1-го кластера 11
- ▶ Скоринг-функция SOL **-6.27 ккал/моль**

Активный центр 6W13 достаточно открытый, что позволяет исследовать связывание самых разнообразных лигандов

# Новая модель главной протеазы SARS-CoV-2



## PDB ID 7L13

- ▶ Разрешение = 2.17 Å
- ▶ Закристаллизован с более сильным нековалентным ингибитором XF7 (IC<sub>50</sub> = 0.018 μM)
- ▶ Докинг нативного лиганда, RMSD = 1.5 Å
- ▶ Населенность 1-го кластера 12
- ▶ Скоринг-функция SOL -8.09 ккал/моль

Активный центр 7L13 более узкий за счет закрытого кармана, который образуется посредством подвижного остатка MET49

# 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

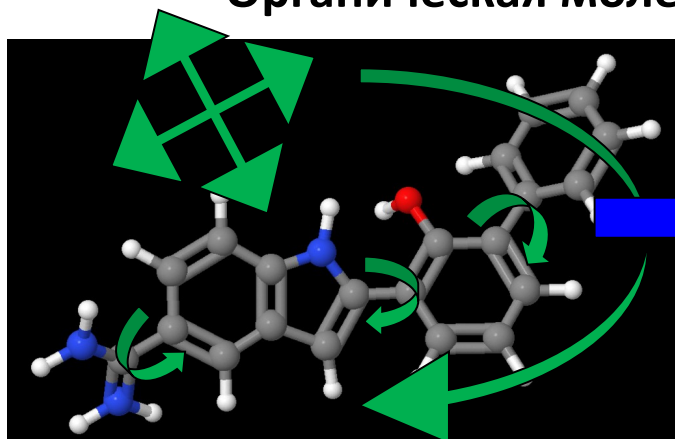
- ▶ Фильтрация **по молекулярному весу** (соединения легче 200 атомных единиц массы отбрасывались)
- ▶ Фильтрация **по числу подвижных степеней свободы** (число степеней свободы не более 10)
- ▶ Фильтрация **функциональных групп**, которые нежелательны в составе лекарств (оксимы и т.д.)
- ▶ Фильтрация **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

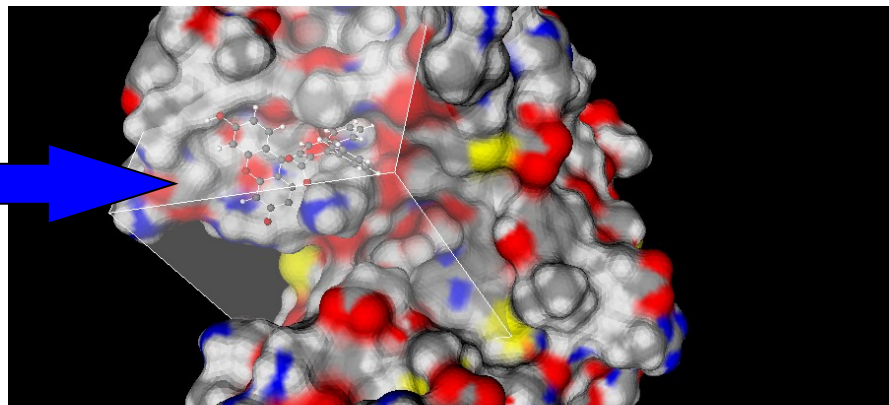
- ▶ Силовое поле MMFF94 фирмы Merck
- ▶ Предварительно рассчитанная сетка потенциалов для пробных атомов в поле белка (жесткий белок)
- ▶ Генетический алгоритм глобальной оптимизации
- ▶ Время расчетов на 1 CPU от 2 до 4 часов

Глобальная оптимизация на многомерной энергетической поверхности

Органическая молекула



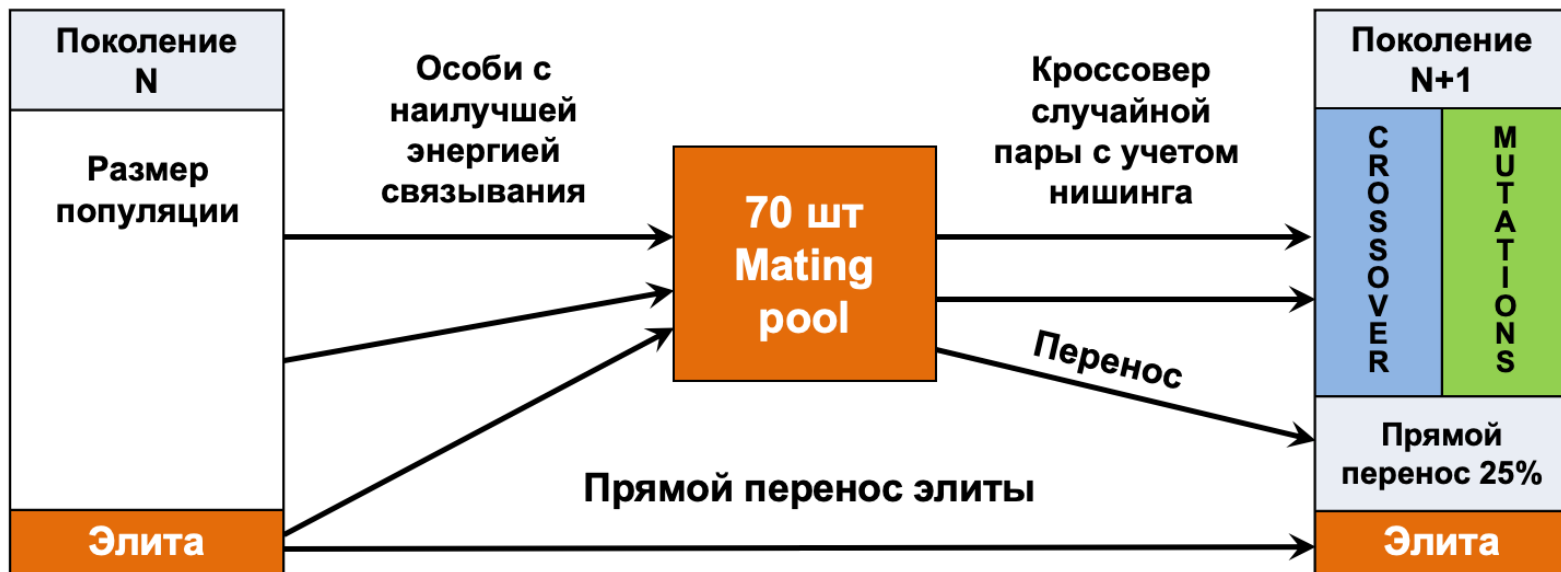
Активный центр белка-мишени



Для библиотеки CNCL необходимо 230 000 узлочасов

# Параметры генетического алгоритма

	СТАНДАРТНЫЕ	БЫСТРЫЕ
▶ Количество запусков	▶ 50	▶ 5
▶ Размер популяции	▶ 30 000	▶ 10 000
▶ Число поколений	▶ 1 000	▶ 1 000
▶ Время расчетов	▶ 2-4 часа	▶ 4-8 минут



# Виртуальный скрининг базы CNCL

## 1. Докинг на быстрых параметрах ГА

- ▶ 988 340 соединений
  - ▶ Подготовка 988 340 трехмерных конформеров
  - ▶ Докинг программой SOL (быстрые параметры)
  - ▶ Критерии отбора соединений
    - ❑ SOL score < -5.5 ккал/моль
  - ▶ Отобрано 12 600 соединений
- ▶ Время расчетов 8 436 узлочасов

# Виртуальный скрининг базы CNCL

## 2. Докинг на стандартных параметрах ГА

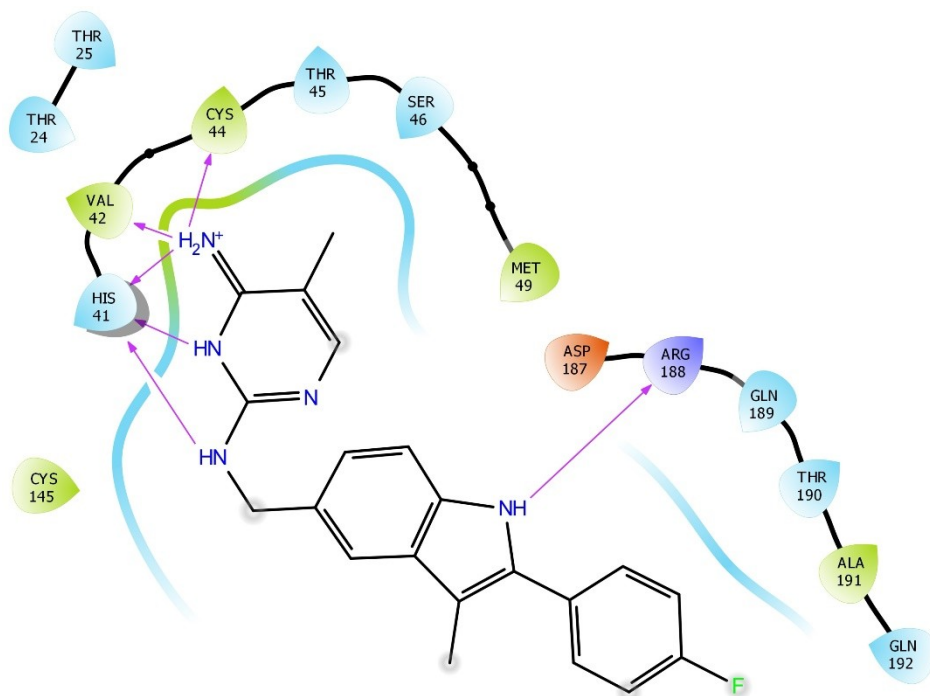
- ▶ 12 600 соединений
- ▶ Подготовка 25 165 трехмерных конформеров
- ▶ Докинг программой SOL (стандартные параметры)
- ▶ Критерии отбора соединений
  - ❑ SOL score < -6.75 ккал/моль
  - ❑ SOL score < -5.75 ккал/моль & 1-ый кластер  $\geq 30$  (из 50)
- ▶ Отобрано 3 131 соединения
- ▶ Время расчетов 7 475 узлочасов

# Виртуальный скрининг базы CNCL

## 3. Расчет энтальпии связывания (MOPAC)

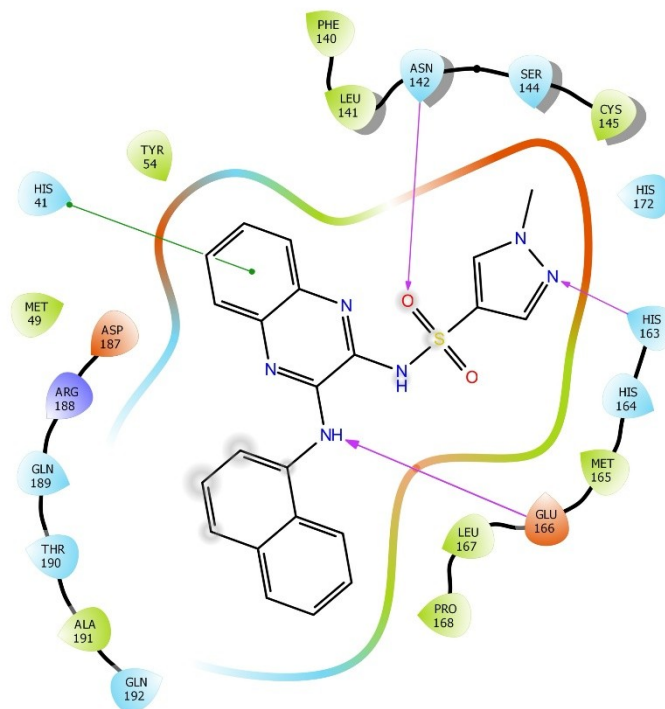
- ▶ Подготовка 3 131 комплексов и 7 022 конформеров
  - ▶ Расчеты квантово-химическим полуэмпирическим методом PM7, с учетом воды в модели COSMO
    - ❑  $\Delta H_{\text{bind}} = E_{\text{Complex}} - E_{\text{Ligand}} - E_{\text{Protein}}$
    - ❑ Для комплекса используется лучшая поза после докинга
    - ❑ Для лиганда используются все начальные конформации
  - ▶ Отобрано 17 соединений
    - ❑ Отбор по значениям сора SOL и энтальпии связывания
    - ❑ Анализ взаимодействия и создания водородных связей
- ▶ Время расчетов 2 000 узлочасов

# Взаимодействие найденных кандидатов в ингибиторы с аминокислотными остатками



**CL01907735**

(6 водородных связей)



**CL01766894**

(3 водородные связи, 1 пи-стекинг)

# Выводы

- ▶ Разработана новая более перспективная модель **7L13** главной протеазы SARS-CoV-2
- ▶ Для виртуального скрининга **сотен тысяч соединений** применена двухэтапная процедура докинга с использованием собственной программы SOL
- ▶ Использование суперкомпьютера «Ломоносов-2» позволило провести исследования (**18 000 узлочасов**)
- ▶ Использование **миллионной библиотеки CNCL** позволило выявить 17 кандидатов в ингибиторы, относящихся к **различным химическим классам**
- ▶ Выявление специфических взаимодействий с белком-мишенью: найденные кандидаты в ингибиторы имеют до **6 водородных связей** с белком-мишенью



# Спасибо за внимание



**Алексей  
Сулимов  
Данил**

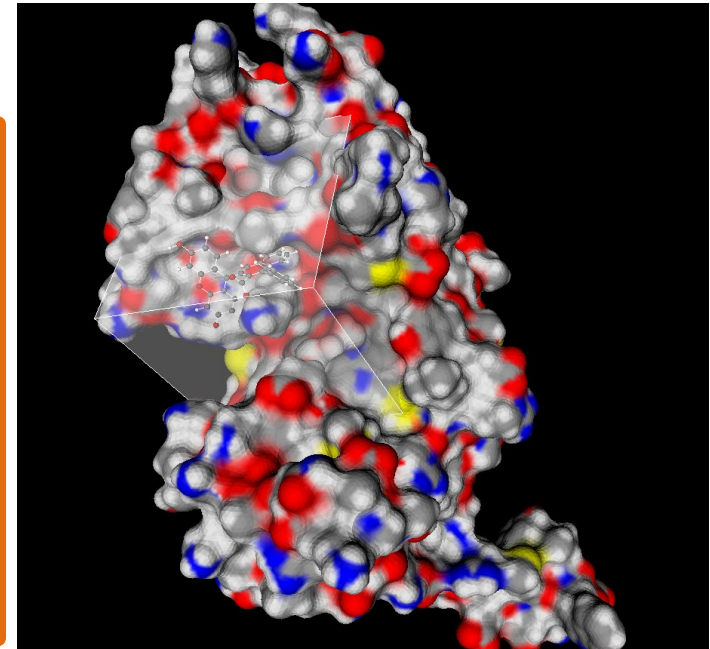
**Кутов  
Иван  
Ильин**

**Владимир  
Сулимов**

**Работа поддержана грантом РФ,  
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 (MORAC)

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