



Динамика и потенциал производительности проекта добровольных вычислений SiDock@home

Никитина Наталия Николаевна
Институт прикладных математических исследований
Карельского научного центра РАН, г. Петрозаводск

1. Разработка лекарств

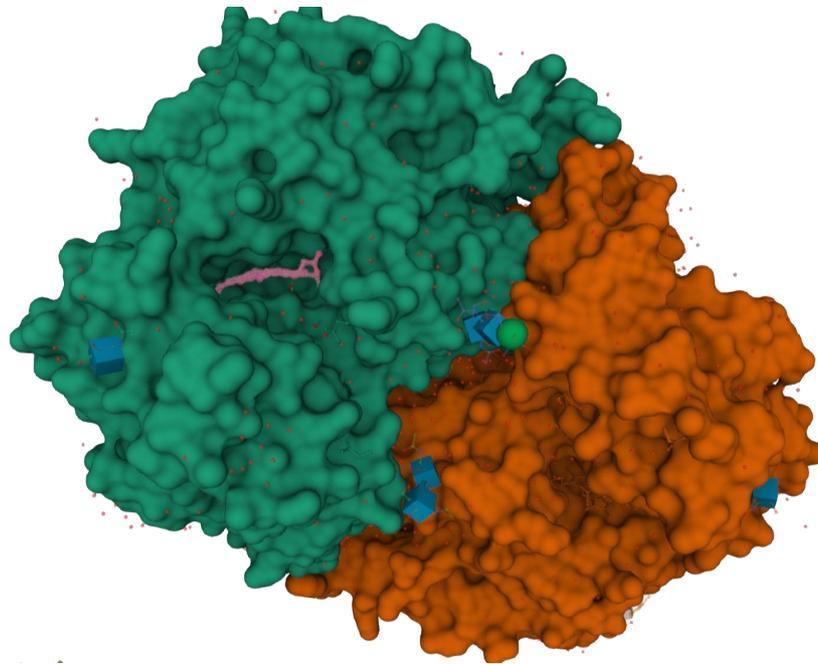


- Разработка лекарства – времяемкий и ресурсоемкий процесс
- Полный цикл разработки лекарства требует 10-15 лет и ~1 млрд \$
- На ранних этапах разработки лекарства важную роль играют высокопроизводительные вычисления (HPC, HTC)

1. Разработка лекарств

Мишень – макромолекула в организме человека, которая играет существенную роль в процессе заболевания и воздействие на которую молекулой лекарства даст желаемый терапевтический эффект.

Лиганд - химическое соединение, которое связывается с мишенью при помощи сил межмолекулярного взаимодействия и производит те или иные биохимические, физиологические или фармакологические эффекты.



Пример: циклооксигеназа-2, фермент, участвующий в воспалительном процессе в организме человека. Лиганд аспирина необратимо связывается с мишенью и препятствует развитию воспалительного процесса.

1. Разработка лекарств

ОПРЕДЕЛЕНИЕ
ЗАБОЛЕВАНИЯ

ИДЕНТИФИКАЦИЯ
ХИТОВ

ИДЕНТИФИКАЦИЯ
КАНДИДАТОВ

КЛИНИЧЕСКИЕ
ИССЛЕДОВАНИЯ

ПОСТМАРКЕТИНГОВЫЕ
ИССЛЕДОВАНИЯ,
ФАРМАКОНАДЗОР

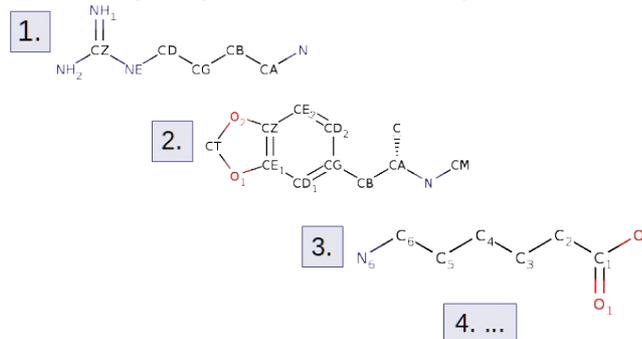


1. Разработка лекарств

Виртуальный скрининг - вычислительноемкая процедура, включающая в себя компьютерное моделирование связывания лиганда с мишенью (молекулярный докинг) на множестве независимых моделей молекул.

Результат - список хитов, упорядоченных по предсказанной способности связывания с мишенью.

Цель - сократить время и стоимость начального этапа разработки лекарства.



- ⊖ Дорзоламид (торговое название Трусопт) - лекарство от глаукомы
- ⊖ Тирофибан (торговое название Аггратат) - антитромбоцитарный препарат
- ⊖ Саквинавир, ритонавир, индинавир, делавирдин - лекарства от СПИДа
- ⊖ Кладрибин - противоопухолевый химиотерапевтический препарат
- ⊖ Эпалрестат - препарат для лечения диабетической невропатии
- ⊖ Гефитиниб - лекарство, используемое при некоторых видах рака молочной железы, легких и других
- ⊖ Эрлотиниб - препарат, используемый для лечения рака легких и рака поджелудочной железы
- ⊖ Сароглитазар - препарат для лечения сахарного диабета 2 типа и дислипидемии
- ⊖ Дувелисиб — противоопухолевый препарат для лечения лимфом и лейкоза
- ⊖ Федратиниб - препарат для лечения заболеваний костного мозга

2. Проект SiDock@home для поиска лекарств



SiDock@home:

- Виртуальный скрининг лекарств
- Авторское ПО для молекулярного докинга CmDock (Curie Marie Dock) с открытым исходным кодом, <https://gitlab.com/Jukic/cmdock>
- Авторская библиотека ~1 000 000 000 моделей молекул
- Предшествующий проект: COVID.SI и другие наработки по поиску лекарств
 - COVID.SI – проект добровольных вычислений на собственной платформе (г. Любляна, Словения)
- URL проекта: <https://sidock.si/sidock/>



2. Проект SiDock@home для поиска лекарств

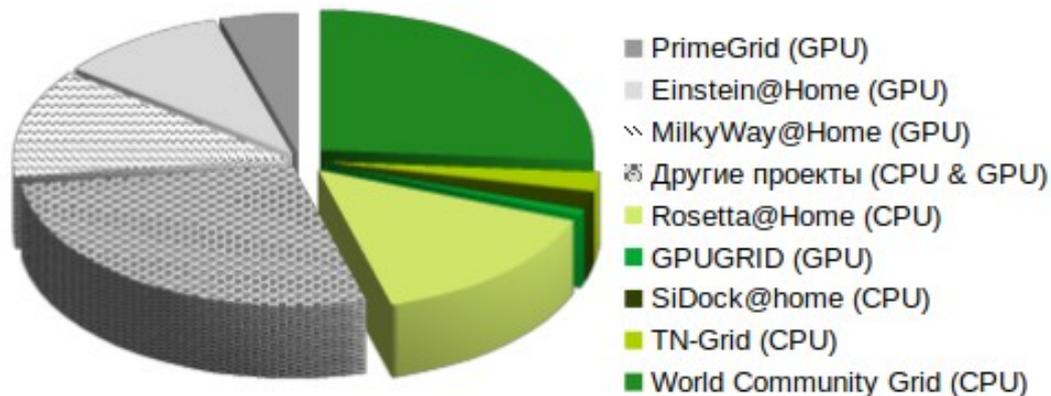
- Первая миссия: виртуальный скрининг для белков SARS-CoV-2
- Вычислительный эксперимент: мишень + библиотека молекул + протокол VC
 - 2 000 000 заданий
 - Вычислительные эксперименты № 1-4: 100%, найдены новые микромолярные ингибиторы шиповидного белка SARS-CoV-2
 - Вычислительные эксперименты № 5-20: 100%, результаты в обработке
 - Вычислительный эксперимент № 21: 1.95%
- Обработка результатов и отбор хитов виртуального скрининга
- Тестирование лучших результатов в лаборатории

2. Проект SiDock@home для поиска лекарств

- Серверная часть: BOINC-сервер, обработка результатов, архивация

Server	Functions	Characteristics
Cloud	BOINC server (6 months); development; storage and processing of I/O files (rxdock-boinc, cmdock-boinc)	Cloud virtual machine at 4 virtual cores of Xeon 6140, 8 Gb RAM, 32 Gb SSD, 512 Gb HDD
Humpback	BOINC server; storage and processing of I/O files (cmdock-boinc-zip)	HP DL 380 Gen8; 2x Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz (12 cores, 24 threads), 32 Gb RAM, 2x HDD 4Tb SAS, RAID 1.
Gray	Auxiliary server; storage of the library; archivation; computations	2x Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz (14 cores, 28 threads), 64 Gb RAM, 2x 10 Tb HDD, RAID 1.

- Клиентская часть: BOINC-клиенты сообщества добровольных вычислений



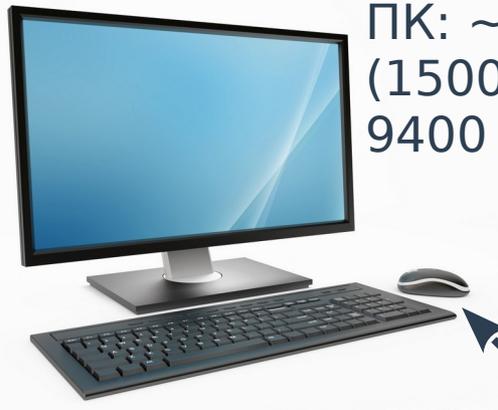
Число активных компьютеров в BOINC-проектах по данным сайта BOINCstats (ноябрь 2022 г.)

2. Проект SiDock@home для поиска лекарств

- 2 450 активных участников, 6 000 активных компьютеров, 40 терафлопс
- 9 900 зарегистрированных пользователей, 43 000 компьютеров за 2 года
- **Science United:** с декабря 2020 г. проект входит в Science United (новую модель добровольных вычислений на основе BOINC, нацеленную на расширение сообщества добровольных вычислений; >6000 компьютеров)
- **Gridcoin:** с июня 2021 г. проект входит в белый список Gridcoin (Gridcoin, GRC – криптовалюта с открытым исходным кодом, начисляемая за участие в научных проектах добровольных вычислений)
- Соревнования в сообществе добровольных вычислений:
 - февраль 2021 г., SiDock@home first challenge
 - март 2021 г., Formula Boinc Sprint
 - май 2021 г., BOINC Pentathlon, 10 000 активных компьютеров, 268 терафлопс
 - сентябрь 2021 г., SiDock@home September Sailing
 - ноябрь 2021 г., SiDock@home challenge against coronavirus
 - январь 2022 г., Charity Event 2022

2. Проект SiDock@home для поиска лекарств

ПК: ~25000
(15000 Windows,
9400 Linux, 600 MacOS)



Серверы: ~6140 (Xeon, Opteron, EPYC)



Наши компьютеры
(32050 с total_credit > 0)

ARM: ~680



Сетевые устройства: 7

3. Производительность проекта SiDock@home

Показатели производительности Desktop Grid:

- Вычислительная мощность (число операций в единицу времени)
- Пропускная способность (число заданий в единицу времени)
- Длительность вычислительного эксперимента
- Надежность (вероятность получения корректного результата)

- Планирование вычислительных экспериментов (время, бюджет)
- Планирование итеративного обмена данными с лабораторией
- ...в условиях гетерогенности и ненадежности Desktop Grid

3. Производительность проекта SiDock@home

- Пропускная способность как число заданий в единицу времени (LHC@Home): наглядна при сравнении физических и вычислительных экспериментов или HPC-систем
- Flops (число операций с плавающей точкой в секунду): осложняется выбором бенчмарка и гетерогенностью; универсальный показатель, общепринятый в научном сообществе
- Бенчмарк Whetstone считается показательным для приложений, интенсивно использующих CPU для арифметических операций
- Система начислений BOINC-кредитов основана на оценке flops

3. Производительность проекта SiDock@home

- Система BOINC-кредитов для унификации и ранжирования вкладов в BOINC-проект, сравнения BOINC-проектов или оценки производительности BOINC в целом

$$C_{h\tau} = T_{h\tau} \cdot a_h \cdot CS$$

$h \in H$ - клиент Desktop Grid

a_h - пиковая производительность CPU клиента по оценке BOINC

τ - задание, которое выполняется на клиенте h

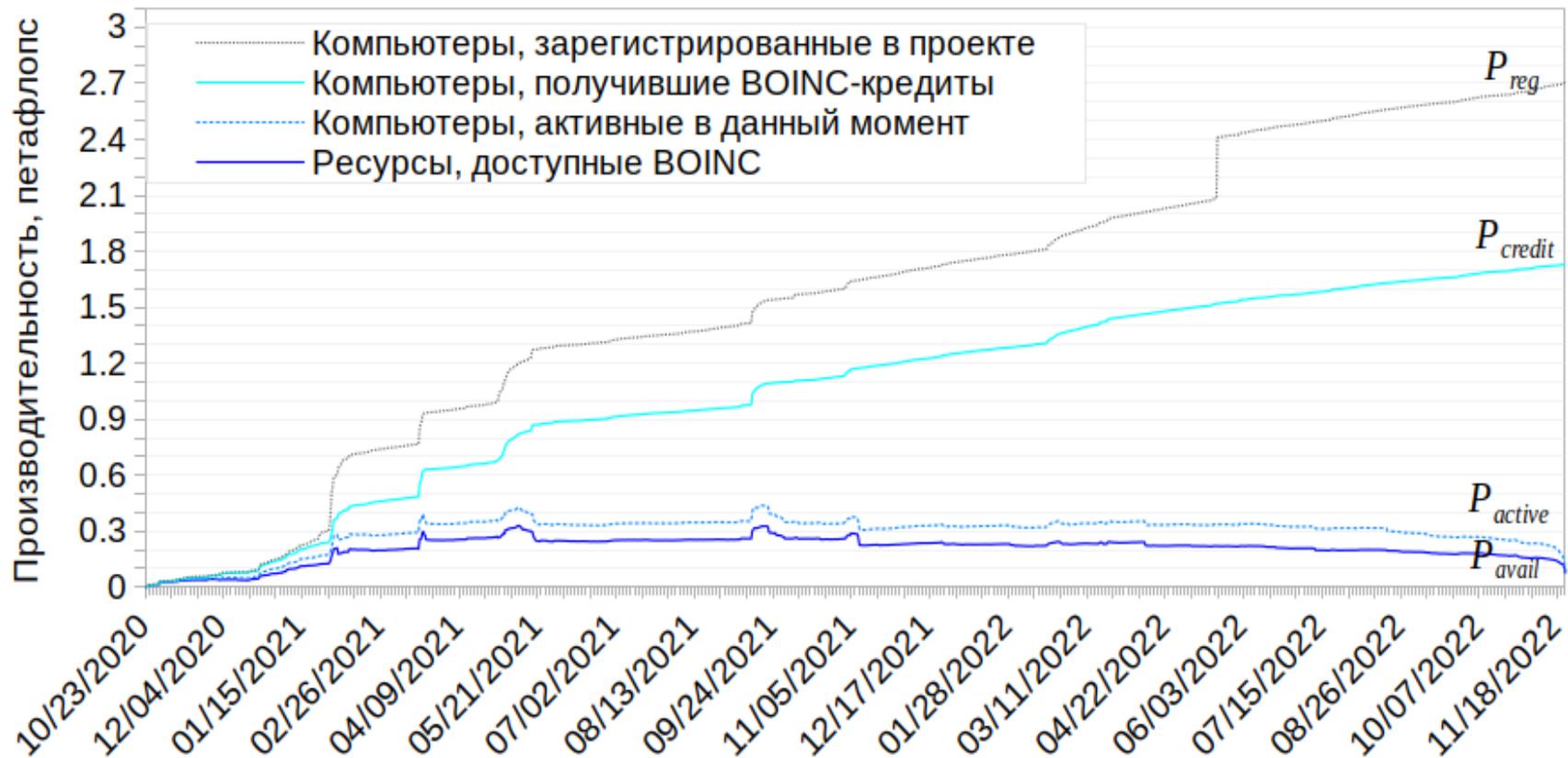
$T_{h\tau}$ - время расчетов

$CS = \frac{200}{86400} \times 10^9$ - константа *Cobblestone* для унификации объема полезной работы гетерогенных компьютеров по сравнению с эталонным компьютером производительностью 1 гигафлопс согласно бенчмарку Whetstone и получающим 200 BOINC-кредитов в день

$C_{h\tau} = T_{h\tau} \cdot a_h \cdot CS$ - число BOINC-кредитов, которые клиент получит за задание

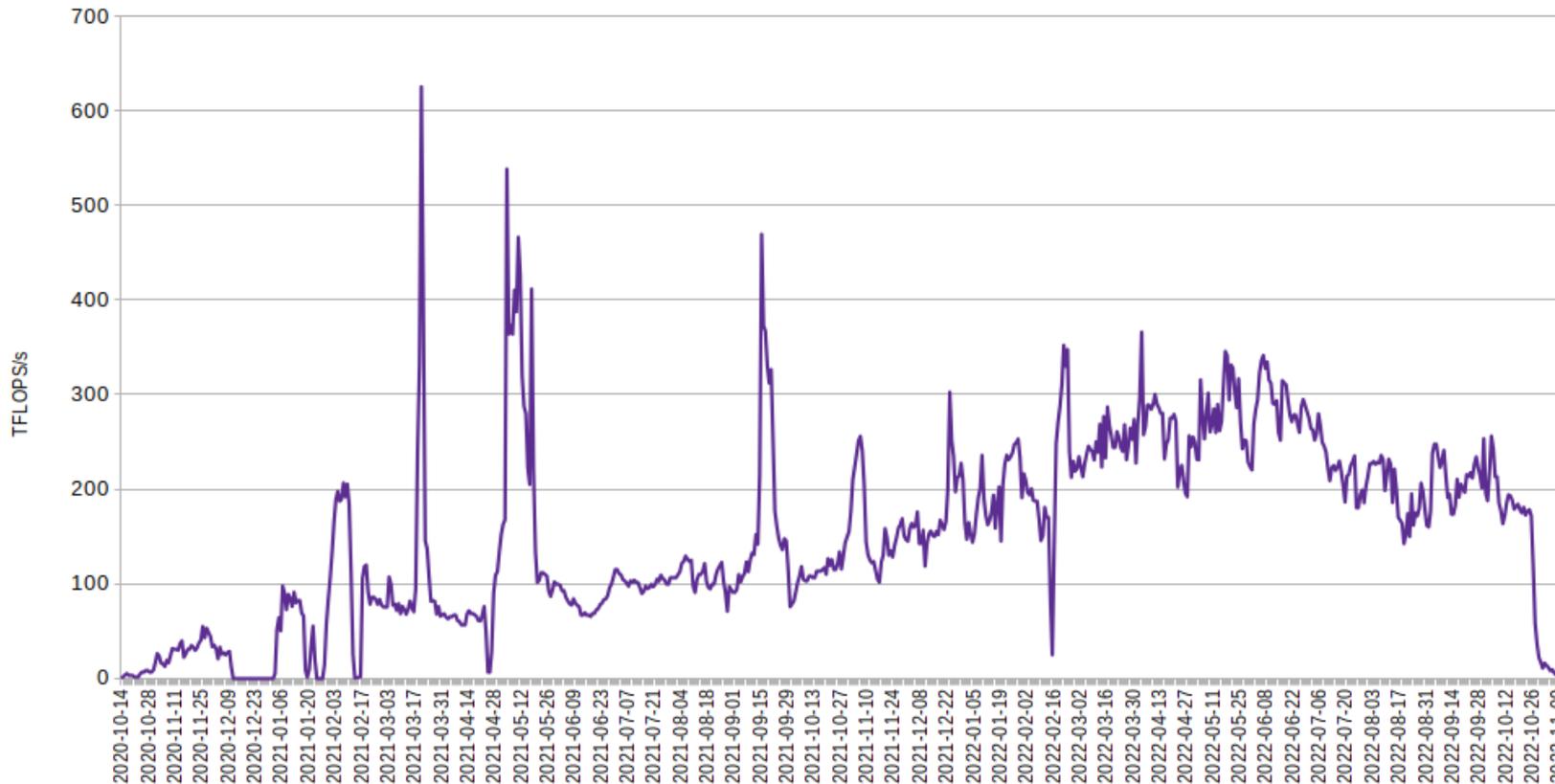
3. Производительность проекта SiDock@home

а) Статистика по вычислительным узлам на основе данных BOINC.



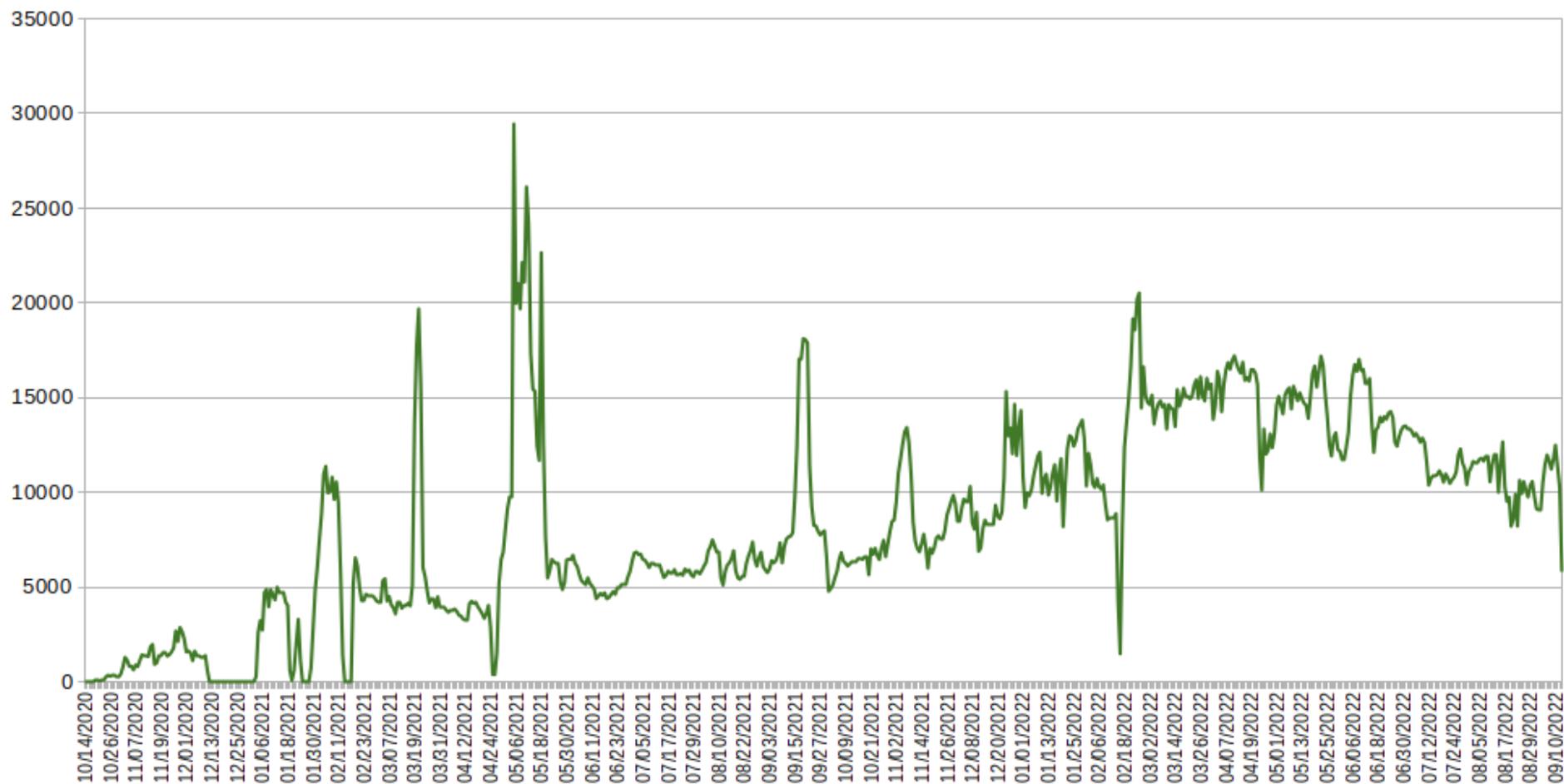
3. Производительность проекта SiDock@home

б) Статистика по BOINC-кредитам, назначенных за задания по сравнению с компьютером AMD Ryzen 9 3900X 12-Core (22500 BOINC-кредитов в сутки, 0.5 TFLOPS/s - производительность по Linpack).

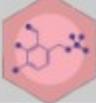
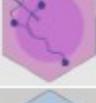
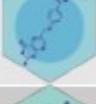


3. Производительность проекта SiDock@home

в) Число потоков (число CPU-дней в день).
Сегодня: ~80000 заданий в день, или 40 млн лигандов



3. Производительность проекта SiDock@home

	BOINC-кредиты	Число участников
	250 000	1939
	1 000 000	752
	2 500 000	329
	5 000 000	164
	10 000 000	73
	25 000 000	25
	50 000 000	11
	100 000 000	8
	250 000 000	1

4. Планы на будущее

- Расширение библиотеки лигандов (+ 1 000 000 000)
- Виртуальный скрининг для новых мишеней SARS-CoV-2, вируса Эбола и др.
- Новая схема генерации, валидации и ассимиляции заданий, новые алгоритмы управления заданиями для повышения эффективности виртуального скрининга: на основе собранной статистики и результатов итеративного обмена данными с лабораторией
- Поддержка новых платформ и GPU

Спасибо за внимание!

E-mail: nikitina@krc.karelia.ru

URL: <https://sidock.si/sidock/>