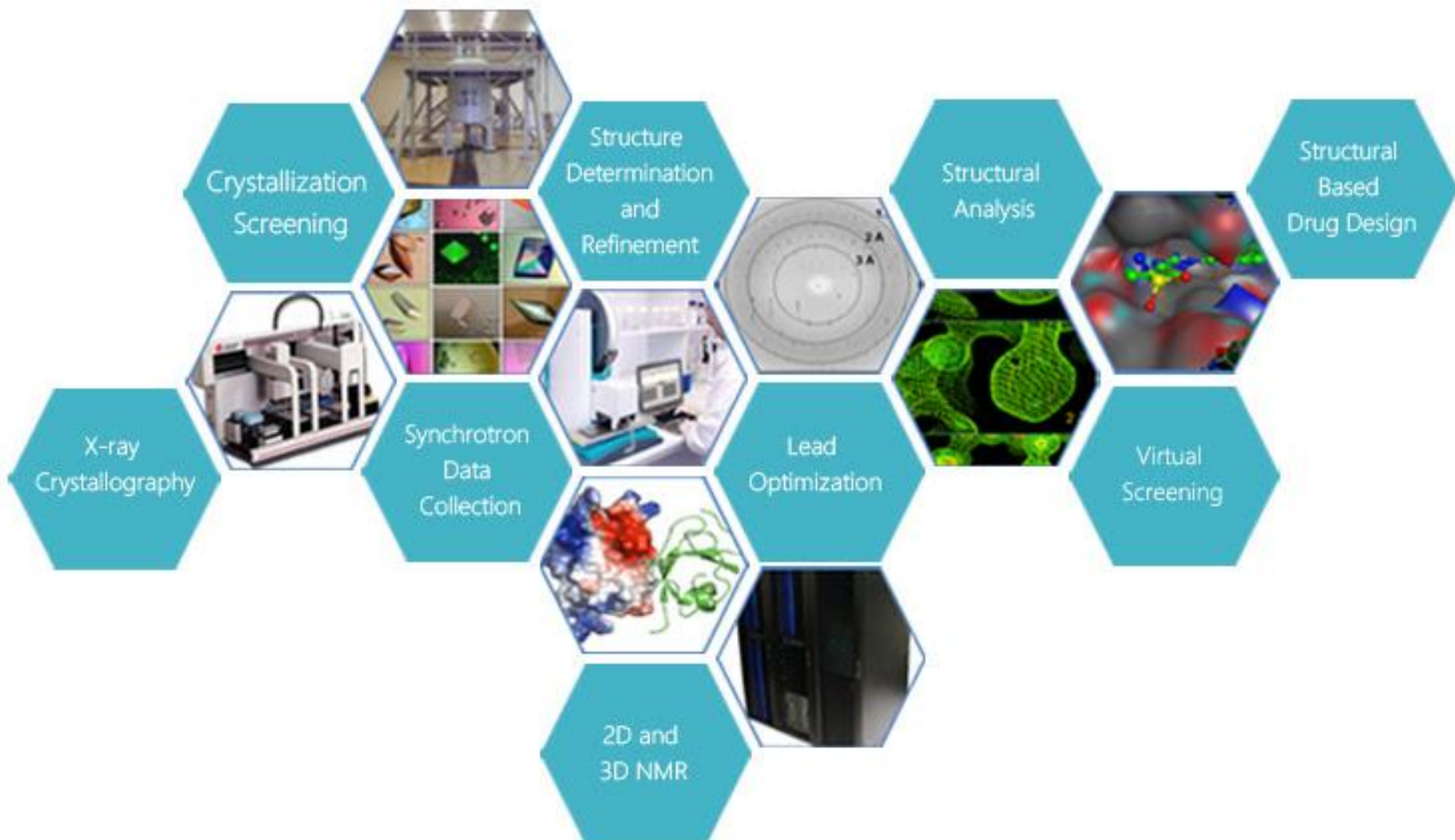


Лекция 7. Биоинформатика и разработка лекарств

Анна Станиславовна Карягина

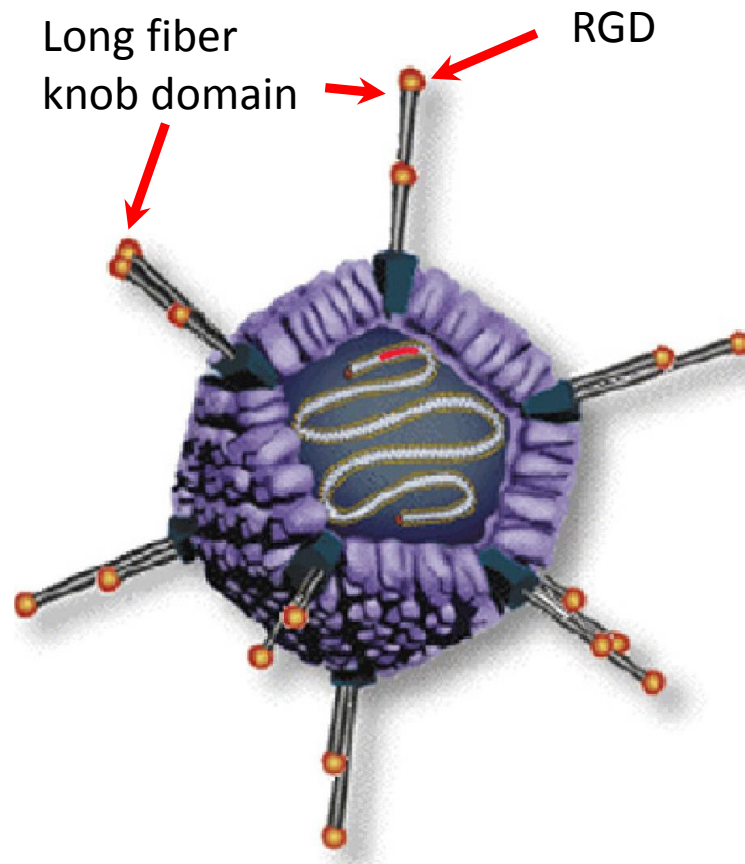
ФГБУ «ФНИЦ эпидемиологии и микробиологии им. Н.Ф. Гамалеи» Минздрава РФ



Как можно использовать биоинформатику для разработки лекарств?

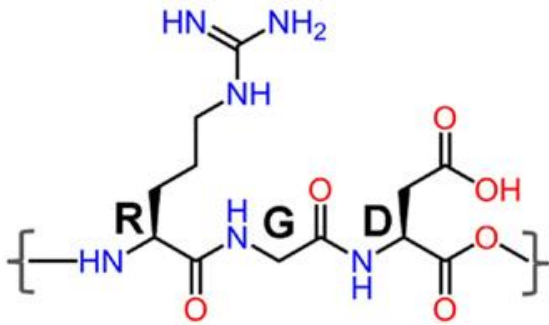
Пример использования «обычных» методов биоинформатики:

**Изменение тропизма вектора для генной терапии:
*компьютерный дизайн фибера рекомбинантного
птичьего аденовируса CELO***

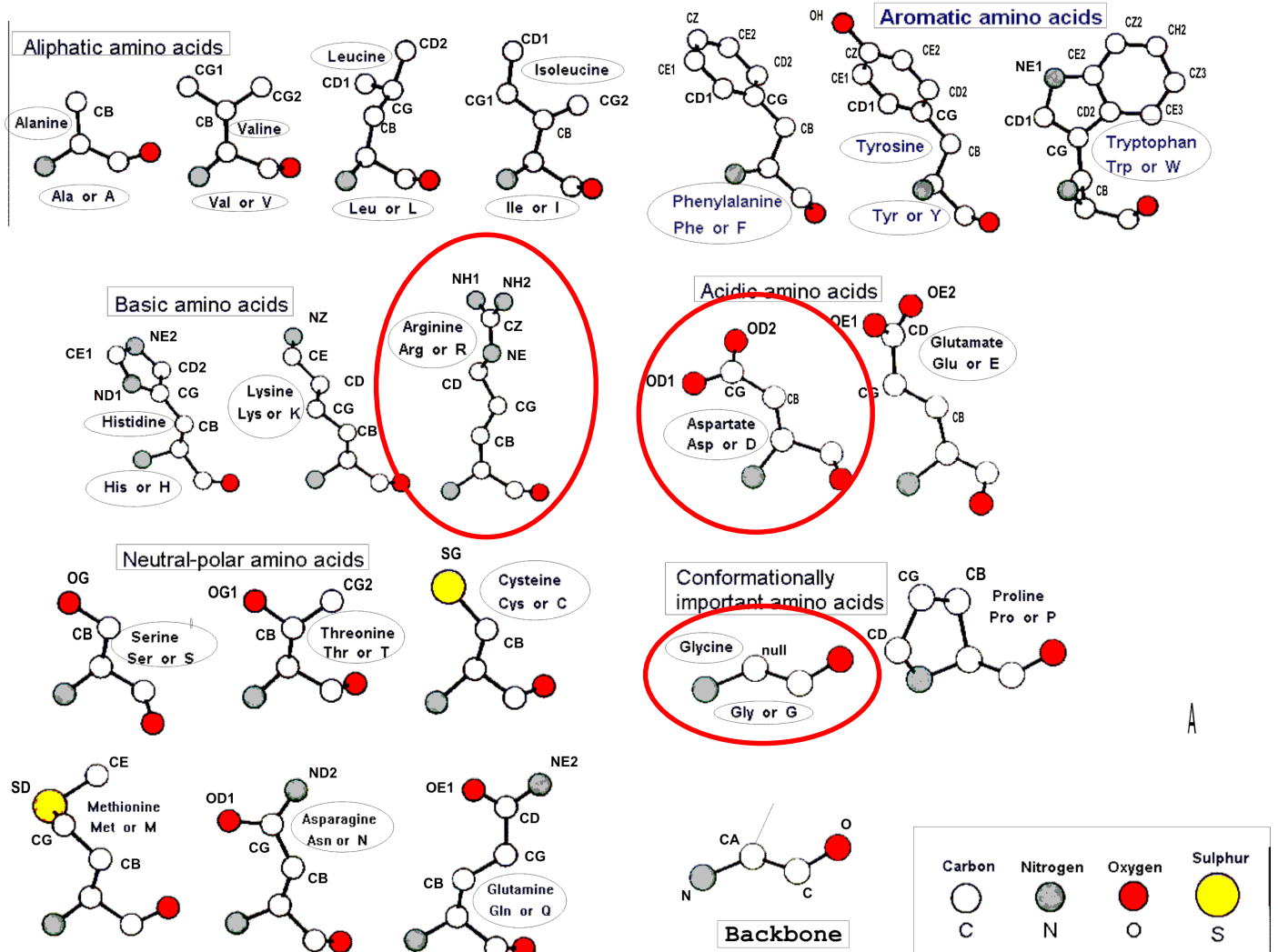


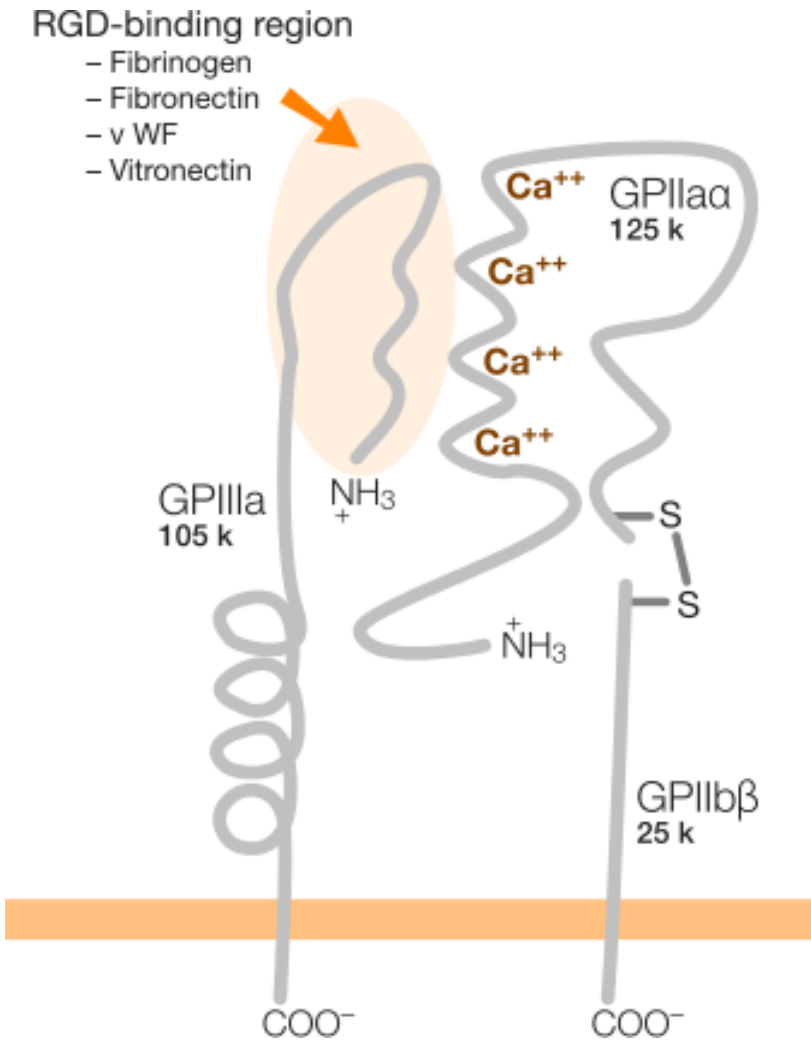
- Птичьи аденовирусы можно использовать в качестве векторов для генной терапии.
Преимущества по сравнению с аденовирусами человека:
- Вирус CELO (Chicken Embryo Lethal Orphan) можно наращивать в куриных эмбрионах (до 10^{12} вирусных частиц на эмбрион) – это гораздо проще и дешевле, чем выращивание вирусов человека в культурах эукариотических клеток.
- На вирус CELO отсутствует первичная иммунная реакция, поскольку люди не инфицируются этим вирусом.
- CELO не реплицируется в клетках человека и млекопитающих.
Недостаток:
- Вирус не способен проникать в некоторые типы клеток млекопитающих (клетки молочной железы).

Задача: С целью изменения тропизма вектора (способности проникать в клетки другого типа, несущие на поверхности рецепторы - **интегрины**) было необходимо вставить аминокислотные остатки **RGD** в поверхностно экспонированный участок длинного фибера аденовируса CELO.



Что такое RGD?

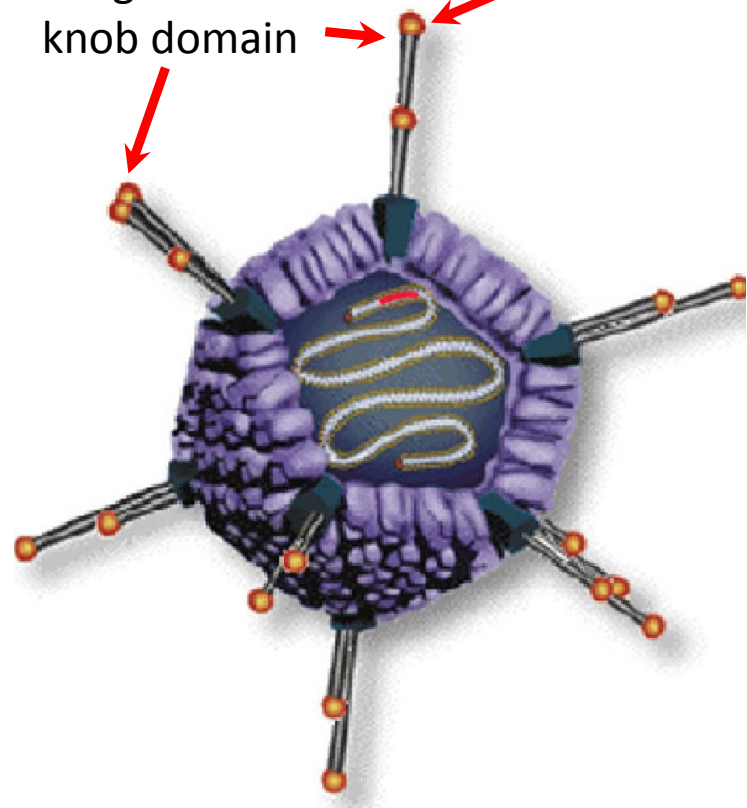




Интегрины — это трансмембранные гетеродимерные клеточные рецепторы, взаимодействующие с внеклеточным матриксом и передающие различные межклеточные сигналы. От них зависит форма клетки, её подвижность, они участвуют в регулировке клеточного цикла. Структурно интегриновые рецепторы представляют собой гетеродимеры — каждый состоит из одной альфа- и одной бета-субъединицы. Альфа-субъединицы определяют специфичность интегрин к лиганду, а бета-субъединицы связаны с компонентами цитоскелета и обеспечивают передачу сигнала в клетке. У человека описано 18 альфа- и 8 бета-субъединиц, при этом каждая альфа-субъединица образует комплекс только с определённым набором бета-единиц, что в итоге порождает 24 варианта димеров.

Long fiber
knob domain

RGD



Задача: Вставить аминокислотные остатки **RGD** в поверхностно экспонированный участок длинного фибера аденовируса CELO.

Проблема: Пространственная структура длинного фибера CELO была не расшифрована. При этом наблюдалось очень ограниченное сходство аминокислотных последовательностей между фибром CELO и аналогичными белками человеческих аденовирусов с известной трехмерной структурой.

Путь решения:

Применение биоинформатических подходов: выравнивание аминокислотных последовательностей с совмещением и введением нумерации предсказанных элементов вторичной структуры белка.

CELO large fiber knob domain

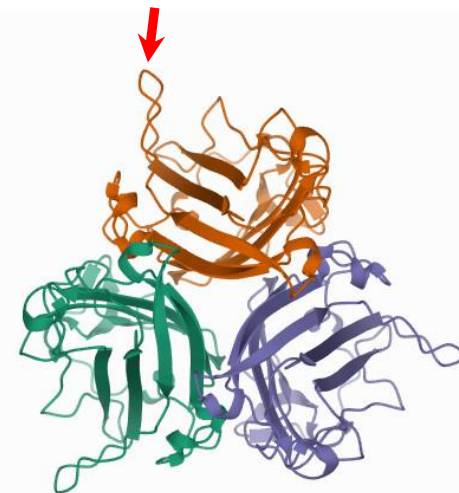
```
630          640          650          660 RGD          670          680          690          700          710
      G              H              I              J
      EEEEE          EEEEEEE          EEEEEEEEE          EE          EEEEE
SNLVTYTKIHFKPTGSQQQLQSLASGELVVAAAKSPVQTTKYELIYLGFTLKQNSSGTFNFFDPNASSDLSFLTPPIPFTYLGYYQ
```

Ad5 fiber knob domain

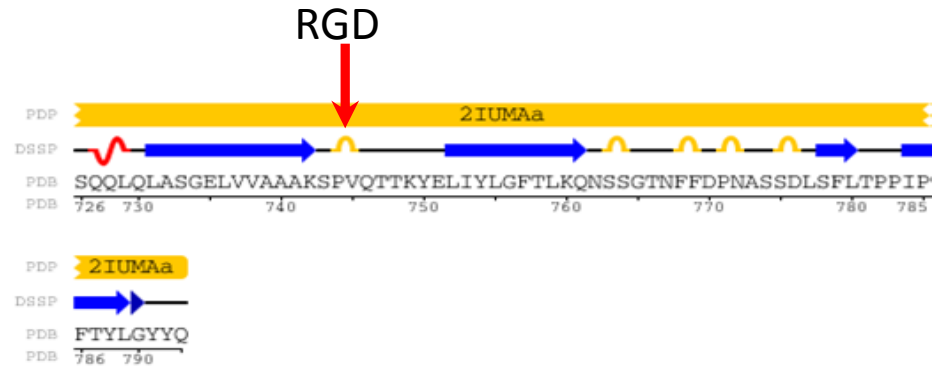
```
510          520          530          540          550          560          570          580
      G              H              I              J
      EEEEEEE          EEEEEEE          EEEEEEE          EE          EEEEE
KSHGKTAKSNIVSQVYLNQDKTKPVTLLTITLNGTQETGDTTPSAYSMSFSWDWSGHNYINEIFATSSYTFSYIAQE
```

EEEE – бета тяжи

Для вставки RGD в аминокислотной последовательности фибера CELO было выбрано место между тяжами H и I, соответствующее промежутку между H и I в структуре белка Ad5, локализирующемуся на поверхности.



Результат: Полученный модифицированный RGD вирус проникал в кроличьи клетки молочной железы *in vivo* с эффективностью, сравнимой с эффективностью человеческого Ad вируса типа 5. В сентябре 2007 года в *J. Virol* была опубликована наша статья.



JOURNAL OF VIROLOGY, Sept. 2007, Vol. 81, No. 18, pp. 9641–9652
 0022-538X/07/\$08.00+0 doi:10.1128/JVI.00534-07
 Copyright © 2007, American Society for Microbiology. All Rights Reserved.

Vol. 81, No. 18

Identification of HI-Like Loop in CELO Adenovirus Fiber for Incorporation of Receptor Binding Motifs[∇]

Denis Y. Logunov,¹ Olga V. Zubkova,¹ Anna S. Karyagina-Zhulina,¹ Eugenia A. Shuvalova,¹
 Andrei P. Karpov,¹ Maxim M. Shmarov,¹ Irina L. Tutykhina,¹ Yulia S. Alyapkina,¹
 Natalia M. Grezina,² Natalia A. Zinovieva,² Lev K. Ernst,²
 Alexandr L. Gintsburg,¹ and Boris S. Naroditsky^{1*}

*Gamaleya Research Institute for Epidemiology and Microbiology (GIEM), 123098, Gamaleya Street 18, Moscow, Russia,¹ and
 Center of Biotechnology and Molecular Diagnostic, All-Russian State Research Institute of Animal Breeding of
 the Russian Academy of Agricultural Science Dubrovitsy Settlement, Podolsk District, Moscow Region, Russia²*

Received 14 March 2007/Accepted 20 June 2007

Vectors based on the chicken embryo lethal orphan (CELO) avian adenovirus (Ad) have two attractive properties for gene transfer applications: resistance to preformed immune responses to human Ads and the ability to grow in chicken embryos, allowing low-cost production of recombinant viruses. However, a major limitation of this technology is that CELO vectors demonstrate decreased efficiency of gene transfer into cells

Structure of the C-terminal head domain of the fowl adenovirus type 1 long fiber

Pablo Guardado-Calvo,¹ Antonio L. Llamas-Saiz,² Gavin C. Fox,³
Patrick Langlois⁴ and Mark J. van Raaij^{1,2}

Correspondence
Mark J. van Raaij
vanraaij@usc.es

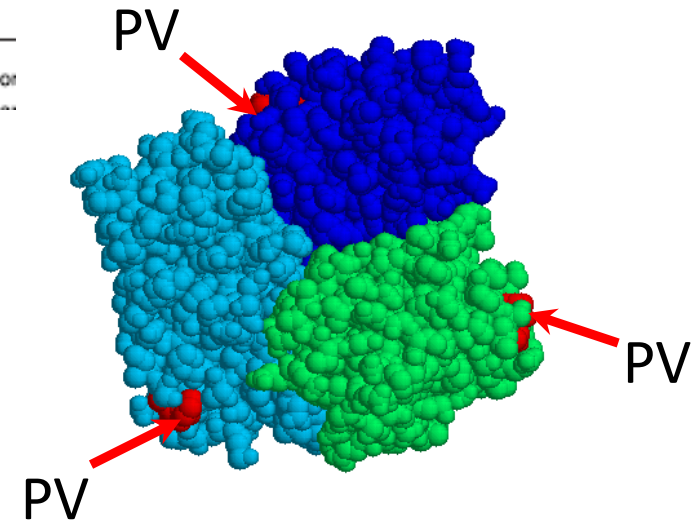
¹Departamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain

²Unidad de Rayos X, Laboratorio Integral de Dinámica y Estructura de Biomoléculas José R. Carracido, Edificio CACTUS, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain

³Spanish CRG Beam Line BM16, European Synchrotron Radiation Facility, 6 rue Jules Horowitz, BP 220, F-38043 Grenoble, France

⁴Agence Française de Sécurité Sanitaire des Aliments, Unité Génétique Virale et Biosecurité Site Les Croix, BP 53, F-22440 Ploufragan, France

Avian adenovirus CELO (chicken embryo lethal orphan virus, fowl adenovirus type 1) incorporates different hexameric fiber proteins extending from the same vertex base a long fiber



Месяцем позже в J. of General Virology была опубликована статья со структурой длинного фибера аденовируса CELO. Предсказанное место для введения RGD, действительно, оказалось на поверхности белковой глобулы, и представляло собой очень удобный для введения модификаций участок.

Safety and immunogenicity of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine in two formulations: two open, non-randomised phase 1/2 studies from Russia



Denis Y Logunov*, Inna V Dolzhikova*, Olga V Zubkova, Amir I Tukhvatulin, Dmitry V Shcheblyakov, Alina S Dzharullaeva, Daria M Grousova, Alina S Erokhova, Anna V Kovyrshina, Andrei G Botikov, Fatima M Izhaeva, Olga Popova, Tatiana A Ozharovskaya, Ilias B Esmagambetov, Irina A Favorskaya, Denis I Zrelkin, Daria V Voronina, Dmitry N Shcherbinin, Alexander S Semikhin, Yana V Simakova, Elizaveta A Tokarskaya, Nadezhda L Lubenets, Daria A Egorova, Maksim M Shmarov, Natalia A Nikitenko, Lola F Morozova, Elena A Smolyarchuk, Evgeny V Kryukov, Vladimir F Babira, Sergei V Borisevich, Boris S Naroditsky, Alexander L Gintsburg

Summary

Background We developed a heterologous COVID-19 vaccine consisting of two components, a recombinant adenovirus type 26 (rAd26) vector and a recombinant adenovirus type 5 (rAd5) vector, both carrying the gene for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike glycoprotein (rAd26-S and rAd5-S). We aimed to assess the safety and immunogenicity of two formulations (frozen and lyophilised) of this vaccine.

Methods We did two open, non-randomised phase 1/2 studies at two hospitals in Russia. We enrolled healthy adult volunteers (men and women) aged 18–60 years to both studies. In phase 1 of each study, we administered intramuscularly on day 0 either one dose of rAd26-S or one dose of rAd5-S and assessed the safety of the two components for 28 days. In phase 2 of the study, which began no earlier than 5 days after phase 1 vaccination, we administered intramuscularly a prime-boost vaccination, with rAd26-S given on day 0 and rAd5-S on day 21. Primary outcome measures were antigen-specific humoral immunity (SARS-CoV-2-specific antibodies measured by ELISA on days 0, 14, 21, 28, and 42) and safety (number of participants with adverse events monitored throughout the study). Secondary outcome measures were antigen-specific cellular immunity (T-cell responses and interferon- γ concentration) and change in neutralising antibodies (detected with a SARS-CoV-2 neutralisation assay). These trials are registered with ClinicalTrials.gov, NCT04436471 and NCT04437875.

Findings Between June 18 and Aug 3, 2020, we enrolled 76 participants to the two studies (38 in each study). In each study, nine volunteers received rAd26-S in phase 1, nine received rAd5-S in phase 1, and 20 received rAd26-S and

Lancet 2020; 396: 887–97

Published Online
September 4, 2020
[https://doi.org/10.1016/S0140-6736\(20\)31866-3](https://doi.org/10.1016/S0140-6736(20)31866-3)

This online publication has been corrected. The corrected version first appeared at thelancet.com on January 7, 2020

See [Comment](#) page 868

*Contributed equally

Federal State Budget Institution "National Research Centre for Epidemiology and Microbiology named after Honorary Academician N F Gamaleya" of the Ministry of Health of the Russian Federation, Moscow, Russia (D Y Logunov DSc, I V Dolzhikova PhD,

Что такое лекарство?

Лекарственное средство, лекарственный препарат, медикамент, лекарство — вещество или смесь веществ синтетического или природного происхождения в виде лекарственной формы (таблетки, капсулы, раствора, мази и т. п.), применяемое для профилактики и лечения заболеваний.

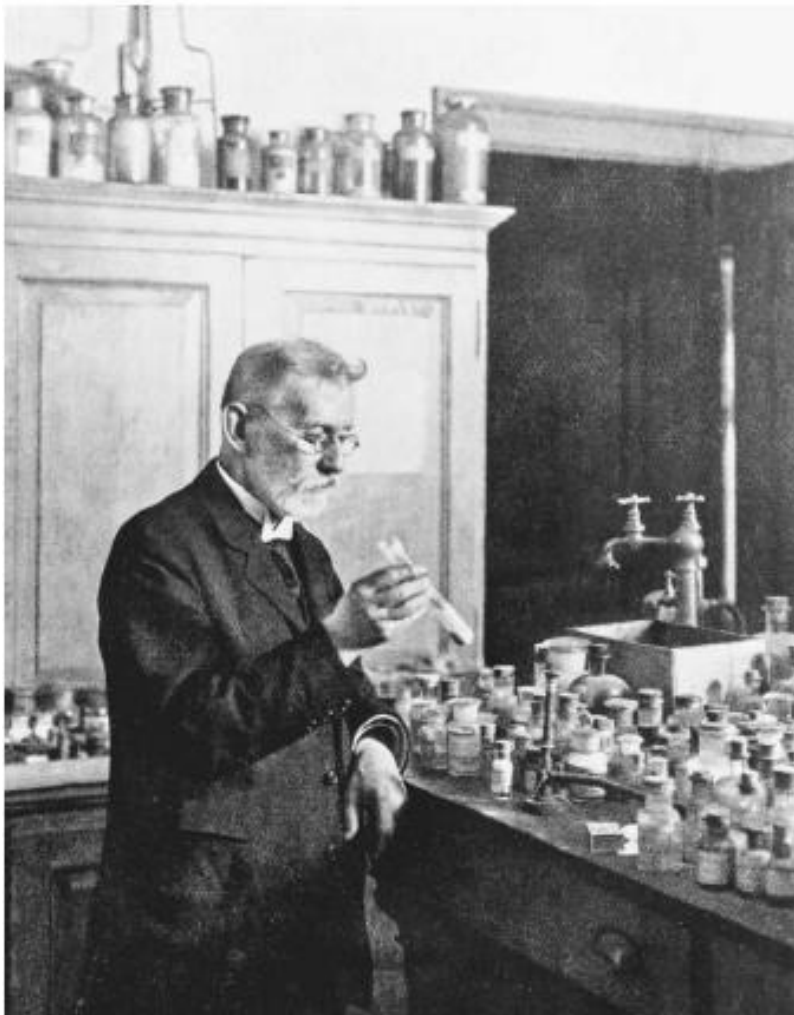


Сколько существует лекарств?



A – 3000 B – 7000 C – 15000 D - 30000

- К настоящему времени в мире разработано немного более 7 тысяч лекарственных препаратов.
- Большинство из них появилось в последние 25 лет.



Немного истории...

Основателем современной химиотерапии считается Пауль Эрлих (1854–1915), нобелевский лауреат 1908 г. (совместно с Мечниковым) по медицине и физиологии.

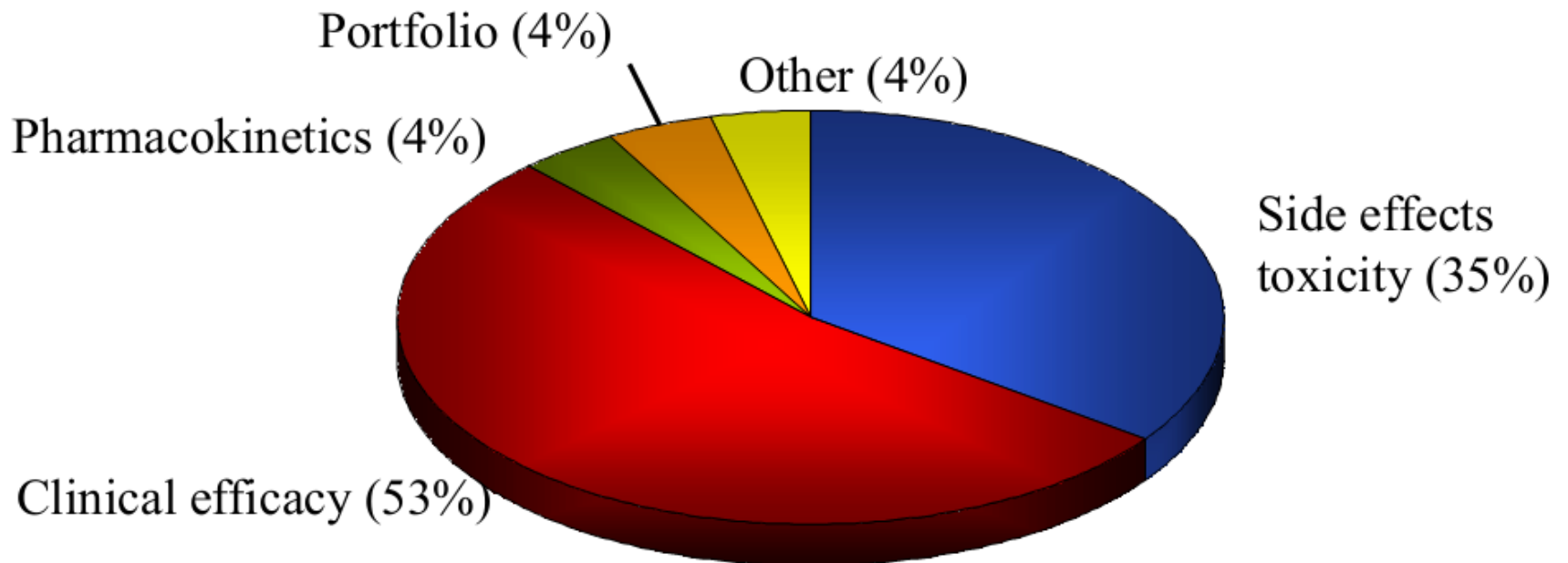
Он выдвинул гипотезу о существовании *хеморецепторов* — специальных тканевых структур, специфически взаимодействующих с химическими веществами, и постулировал возможность использования этого феномена в терапии различных заболеваний.

На протяжении многих лет он разрабатывал лекарства на основе мышьяка и в конце концов достиг невероятного успеха — получил абсолютно безвредное и очень эффективное лекарство от сифилиса.

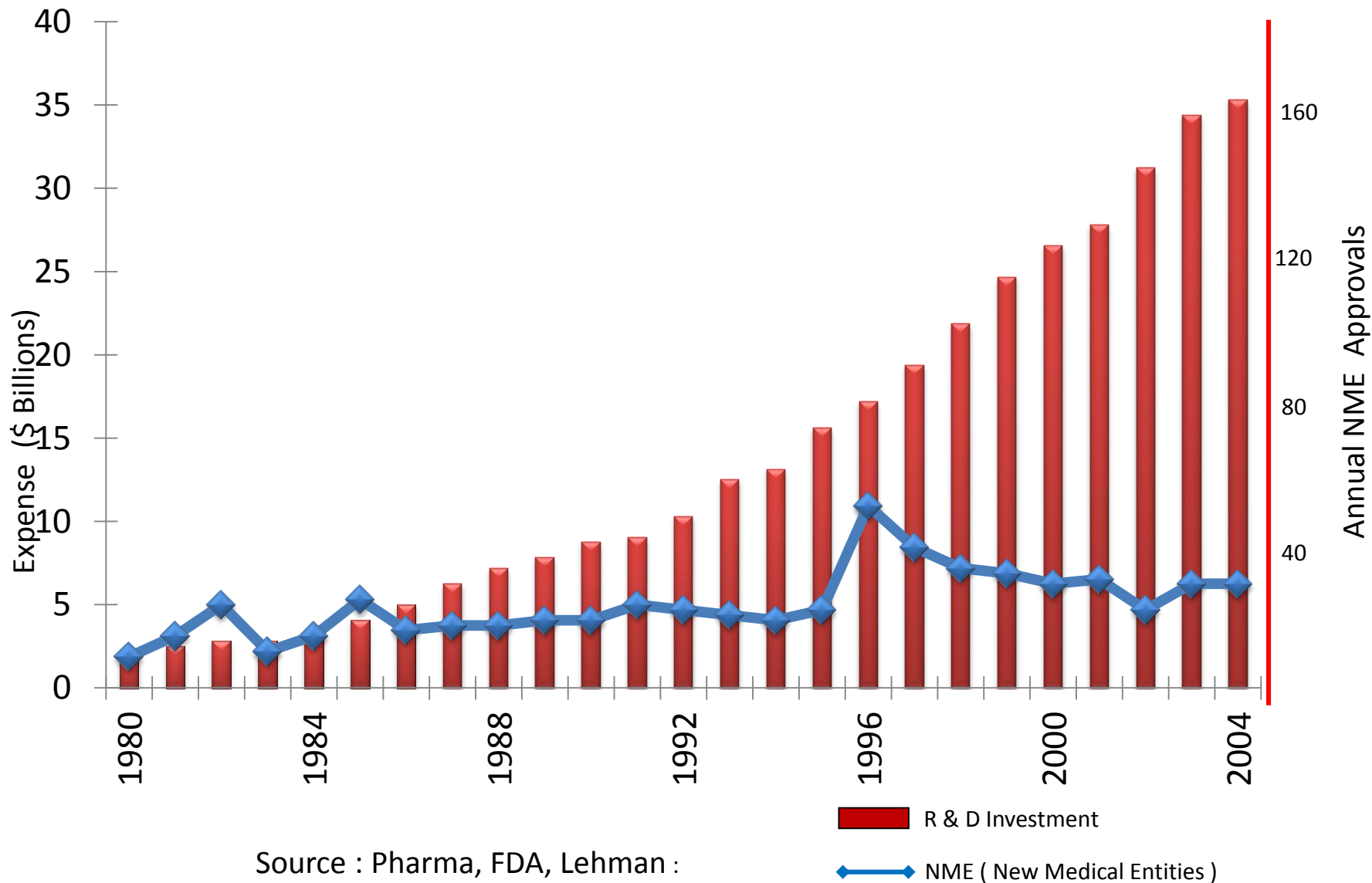
Особенности разработки лекарств на современном этапе

- Очень сегментированный рынок – самая большая компания (Pfizer) имеет только 11% рынка
- Высокий уровень риска: разработка лекарства длится 10-20 лет, при этом большинство лекарств не выходит на рынок
- Регистрация препаратов – отдельный сложный и длительный процесс

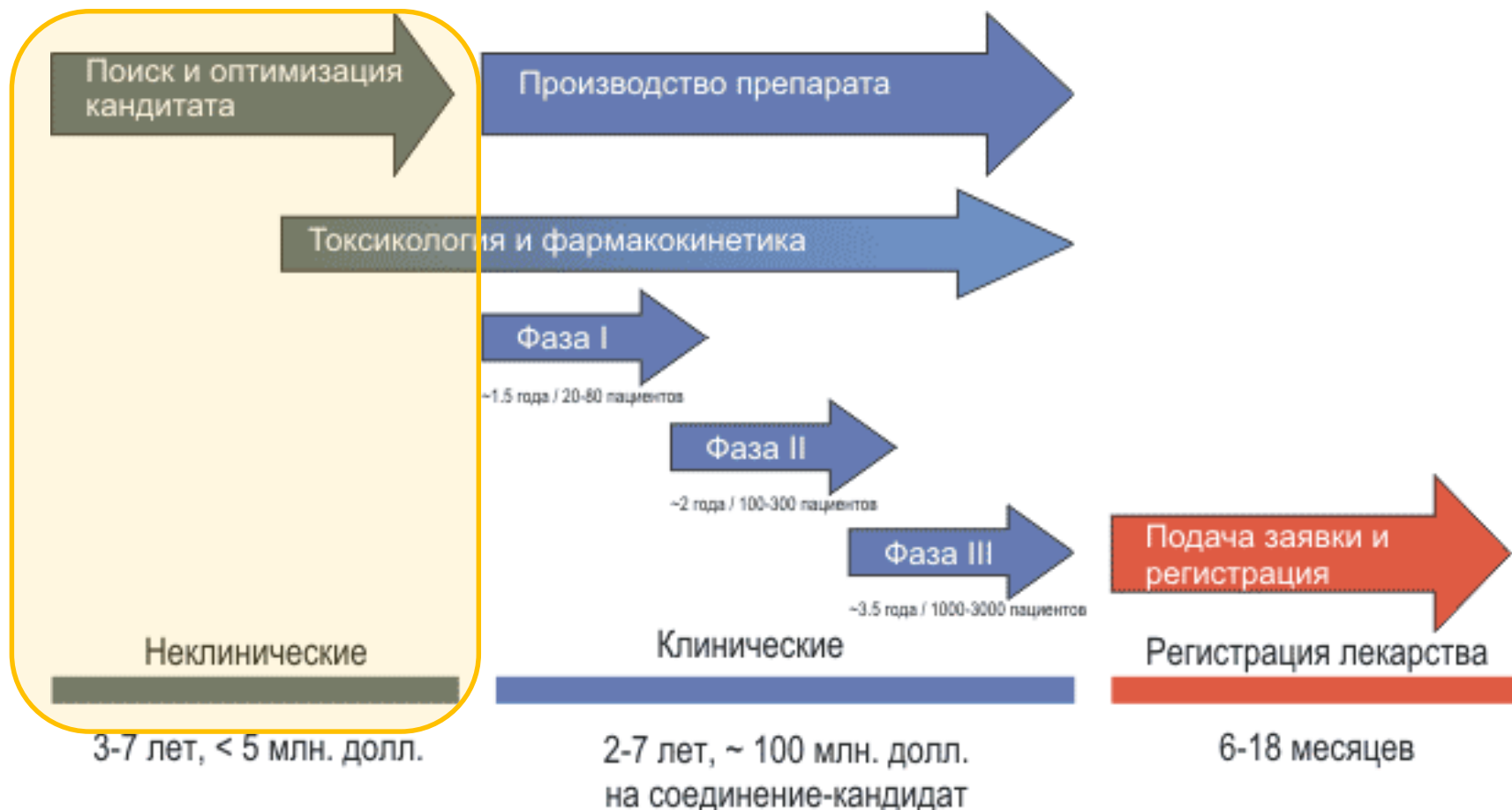
Phase III failures 1992 – 2002



Ежегодные вложения в разработку лекарств и количество одобренных препаратов



Этапы создания нового лекарства



Затраты на полный цикл разработки одного нового лекарства обычно составляют от 900 миллионов до 2 миллиардов долларов. Типичное время разработки лекарства – 10-15 лет.

Использование новых технологий, в частности, основанных на методах био- и химиоинформатики, может существенно ускорить и удешевить стадию поиска и оптимизации лидерного соединения.

Как сравнить эффективность использования компьютерных технологий и традиционных методов при разработке лекарств?

Пример:

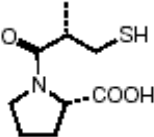
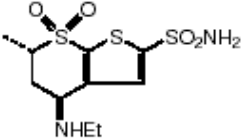
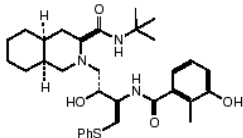
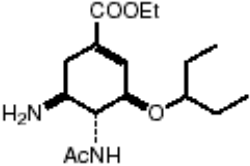

Специалисты фирмы Pharmacia (сейчас – часть фирмы Pfizer) использовали компьютерные методы для поиска ингибиторов *тирозин-фосфатазы-1В* с целью разработки лекарства от диабета.

В результате было отобрано 365 соединений, 127 из которых демонстрировали ингибирующий эффект (**hit rate 35%**).

Одновременно проведенный традиционный высокопроизводительный скрининг (HTS) среди 400000 соединений позволил отобрать 81 ингибитор (**hit rate 0.021%**).

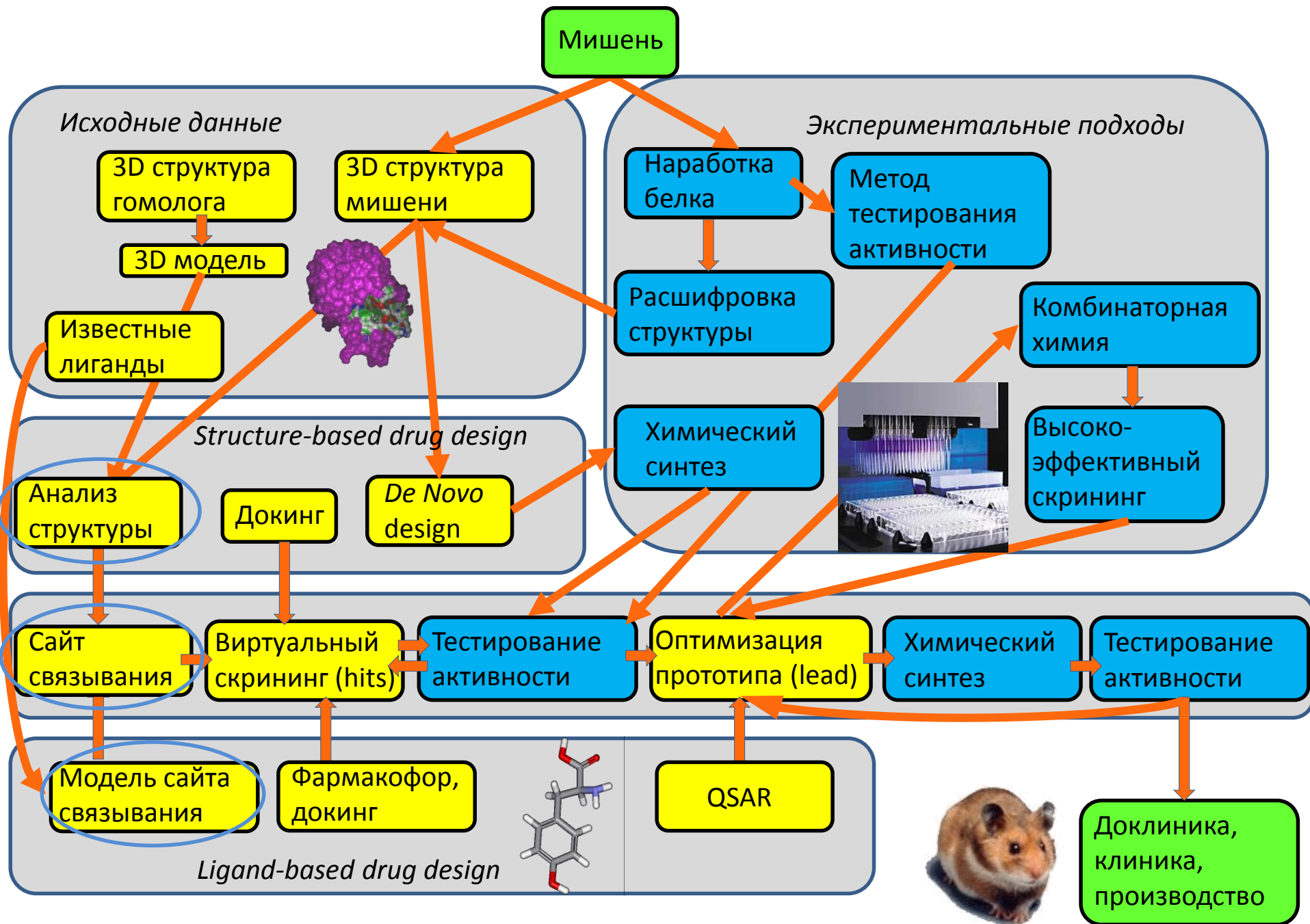
Doman et al. J Med Chem 45: 2213–2221.

Примеры лекарств, разработанных с применением компьютерных подходов

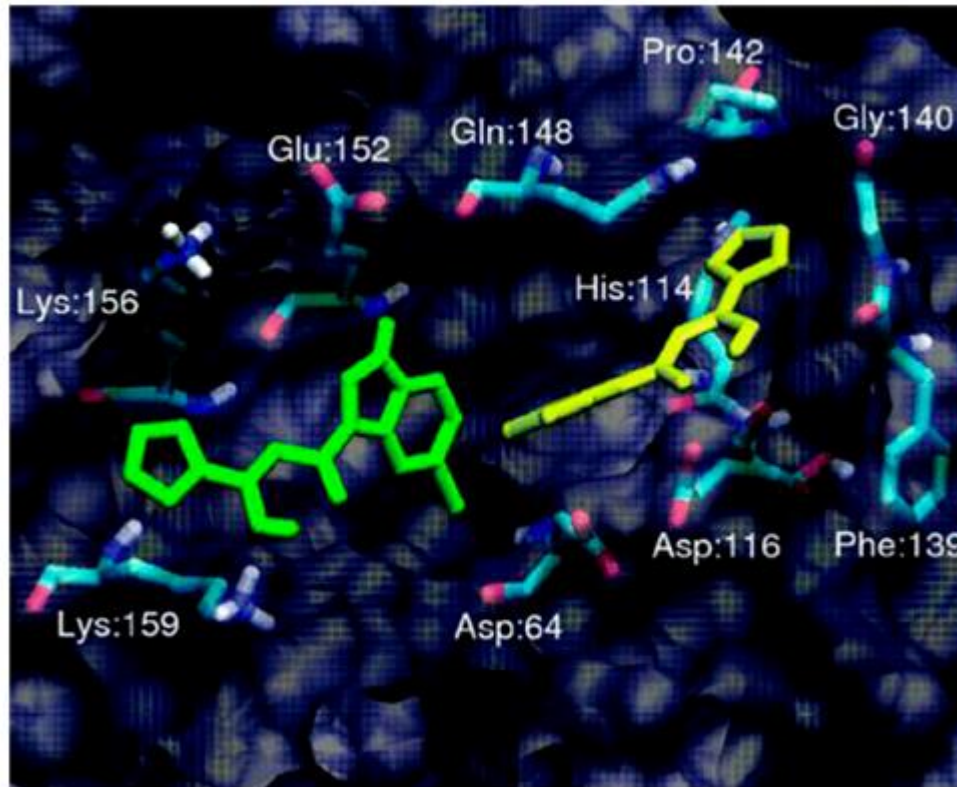
Capoten	Captopril		ACE (angiotensin- converting- enzyme)	Hypertension	1981	Bristol- Myers Squibb
Trusopt	Dorzolamide		Carbonic anhydrase	Glaucoma	1995	Merck
Viracept	Nelfinavir		HIV protease	HIV/ AIDS	1999	Agouron (Pfizer) and Lilly
Tamiflu	Oseltamivir		Neuraminidase	Influenza	1999	Gilead and Roche
Gleevec	Imatinib		BCR- Abl	Chronic myelogenous leukaemia	2001	Novartis



Основные стратегии, используемые при компьютерной разработке лекарств



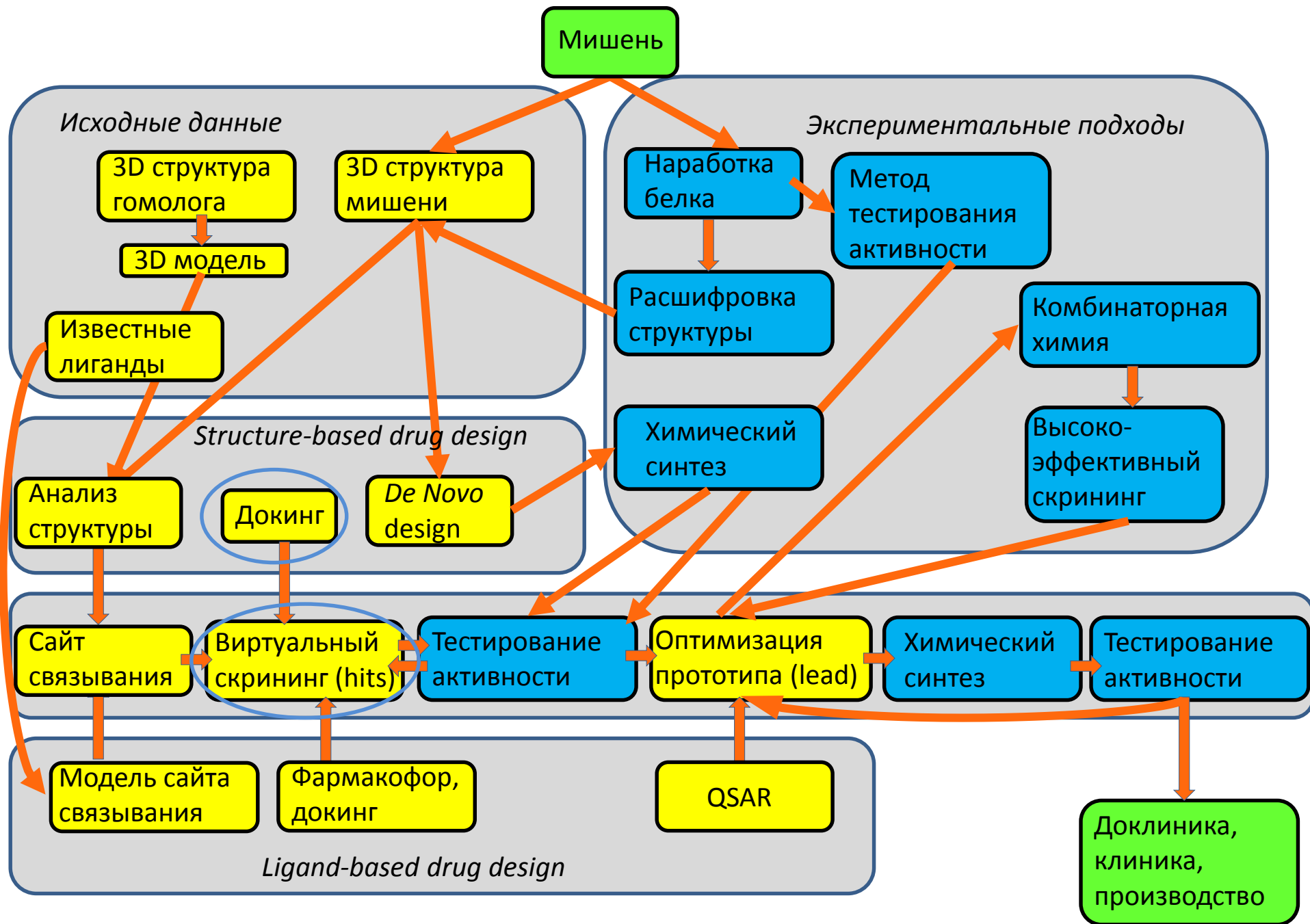
Анализ структуры белка-мишени и его комплексов



Результат анализа структуры не всегда очевиден:

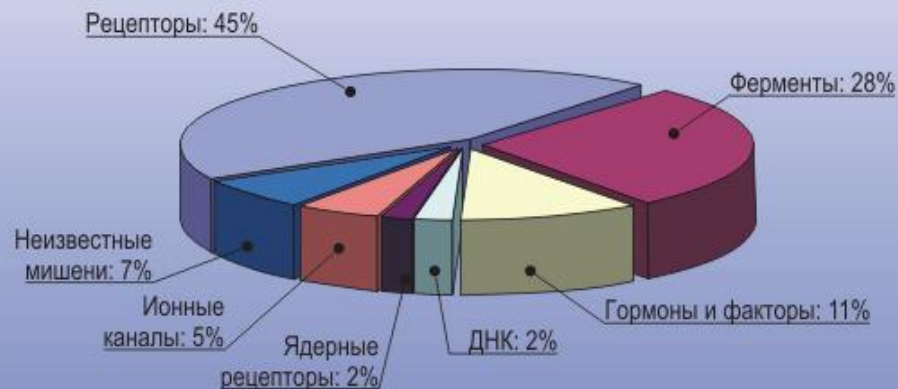
Методом симуляции молекулярной динамики исследовали структуру комплекса интергазы ВИЧ с известным ингибитором. При этом нашли новое место связывания, отличное от того, которое обнаружили при расшифровке структуры с помощью РСА. Это привело к разработке в дальнейшем фирмой Merck лекарства ралтегравира, используемого для лечения СПИДа (Summa et al., 2008).

Основные стратегии, используемые при компьютерной разработке лекарств



Structure-based drug design: молекулярный докинг

Биохимическая классификация мишеней, используемых в современной фармацевтической промышленности (всего: 483)

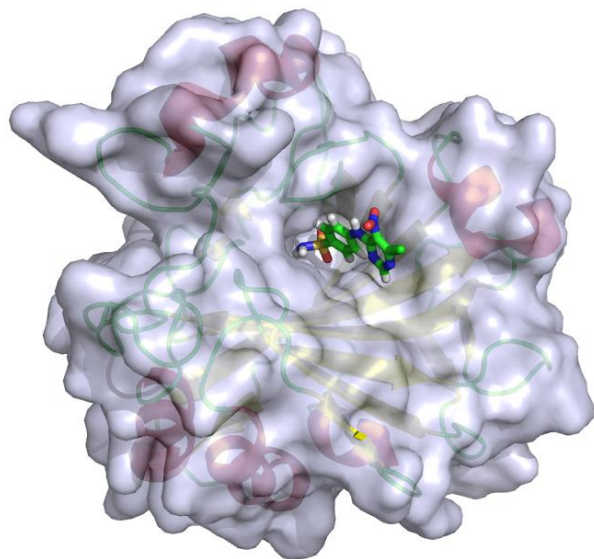


Подавляющее большинство лекарств – низкомолекулярные ингибиторы белков.

PubChem – наиболее полная база данных известных низкомолекулярных соединений, содержит > 37 миллионов в-в.

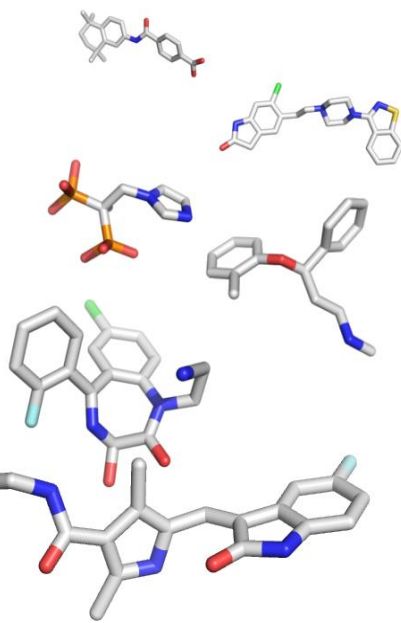
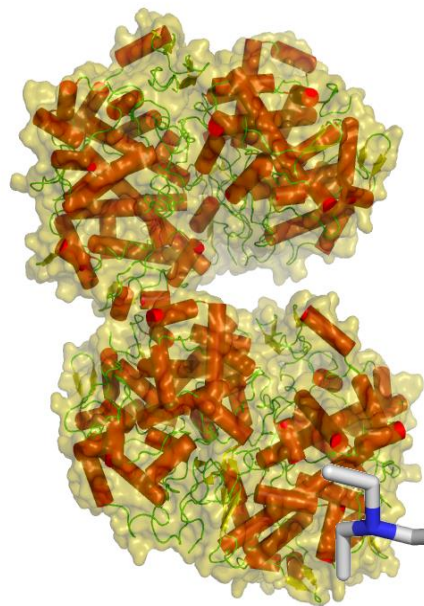
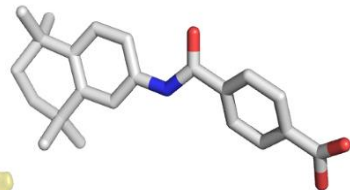
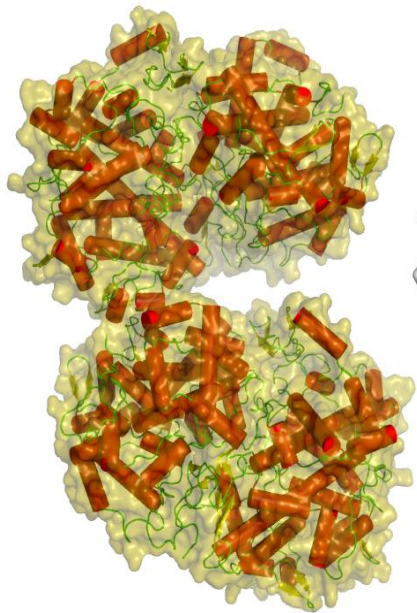
Даже среди гораздо меньшего количества соединений можно найти такое, которое будет связываться с интересующим нас белком.

Как это сделать?



Нужно найти соединение, подходящее к определенному месту белка, например, активному центру, как ключ к замку.





Докинг (docking) – моделирование взаимодействия двух молекул, чаще всего белка и низкомолекулярного вещества – предполагаемого лиганда. Позволяет найти оптимальную конформацию и расположение лиганда в месте связывания и оценить силу взаимодействия.

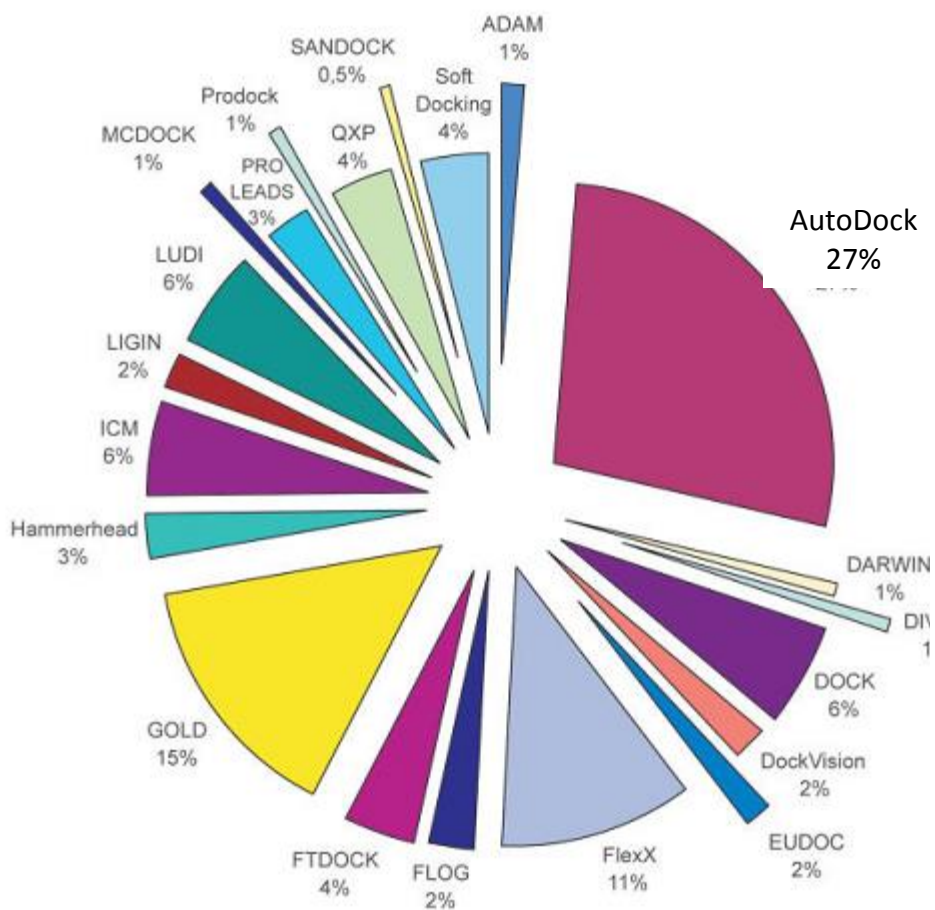
Виртуальный скрининг (virtual screening) – последовательный докинг множества различных низкомолекулярных веществ против одного белка с целью отобрать для экспериментальной проверки наиболее перспективные. Позволяет повысить вероятность обнаружения ингибитора интересующего нас белка по сравнению со случайным поиском по банку веществ, т.е., сократить количество в-в, требующих экспериментальной проверки.

Список баз данных химических соединений

Database	Type	Size
PubChem (Wheeler et al., 2006)	Biologic activities of small molecules	~40,000,000
Accelrys Available Chemicals Directory (ACD) (Accelrys, 2012)	Consolidated catalog from major chemical suppliers	~7,000,000
PDBeChem (Dimitropoulos, 2006)	Ligands and small molecules referred in PDB	14,572
Zinc (Irwin and Shoichet, 2005)	Annotated commercially available compounds	~21,000,000
LIGAND (Goto et al., 2002)	Chemical compounds with target and reactions data	16,838
DrugBank (Wishart et al., 2006)	Detailed drug data with comprehensive drug target information	6711
ChemDB (Chen et al., 2005, 2007)	Annotated commercially available molecules	~5,000,000
WOMBAT Data base (World of Molecular BioAcTivity) (Ekins et al., 2007; Hristozov et al., 2007)	Bioactivity data for compounds reported in medicinal chemistry journals	331,872
MDDR (MDL Drug Data Report) (Hristozov et al., 2007)	Drugs under development or released; descriptions of therapeutic	180,000
3D MIND (Mandal et al., 2009).	Molecules with target interaction and tumor cell line screen data	100,000

Pharmacol Rev 66:334–395

Программы для докинга и виртуального скрининга



FlexX (<http://www.biosolveit.de/FlexX/>)

Dock (<http://dock.compbio.ucsf.edu>)

AutoDock (<http://autodock.scripps.edu>)

AutoDock Vina (<http://vina.scripps.edu>)

Surflex (<http://www.biopharmics.com>, www.tripos.com)

Fred (<http://www.eyesopen.com/products/applications/fred.html>)

Gold (http://www.ccdc.cam.ac.uk/products/life_sciences/gold/)

PLANTS (<http://www.tcd.uni-konstanz.de/research/plants.php>)

3DPL (<http://www.chemnavigator.com/cnc/products/3dpl.asp>)

Molegro Virtual Docker (<http://www.molegro.com>)

ICM Pro (http://www.molsoft.com/icm_pro.html)

Q-Pharm (<http://www.q-pharm.com>)

Ligand fit, Libdock and CDocker

(<http://accelrys.com/services/training/life-science/StructureBasedDesignDescription.html>)

DockSearch (<http://www.ibmcm.sk.ru>)

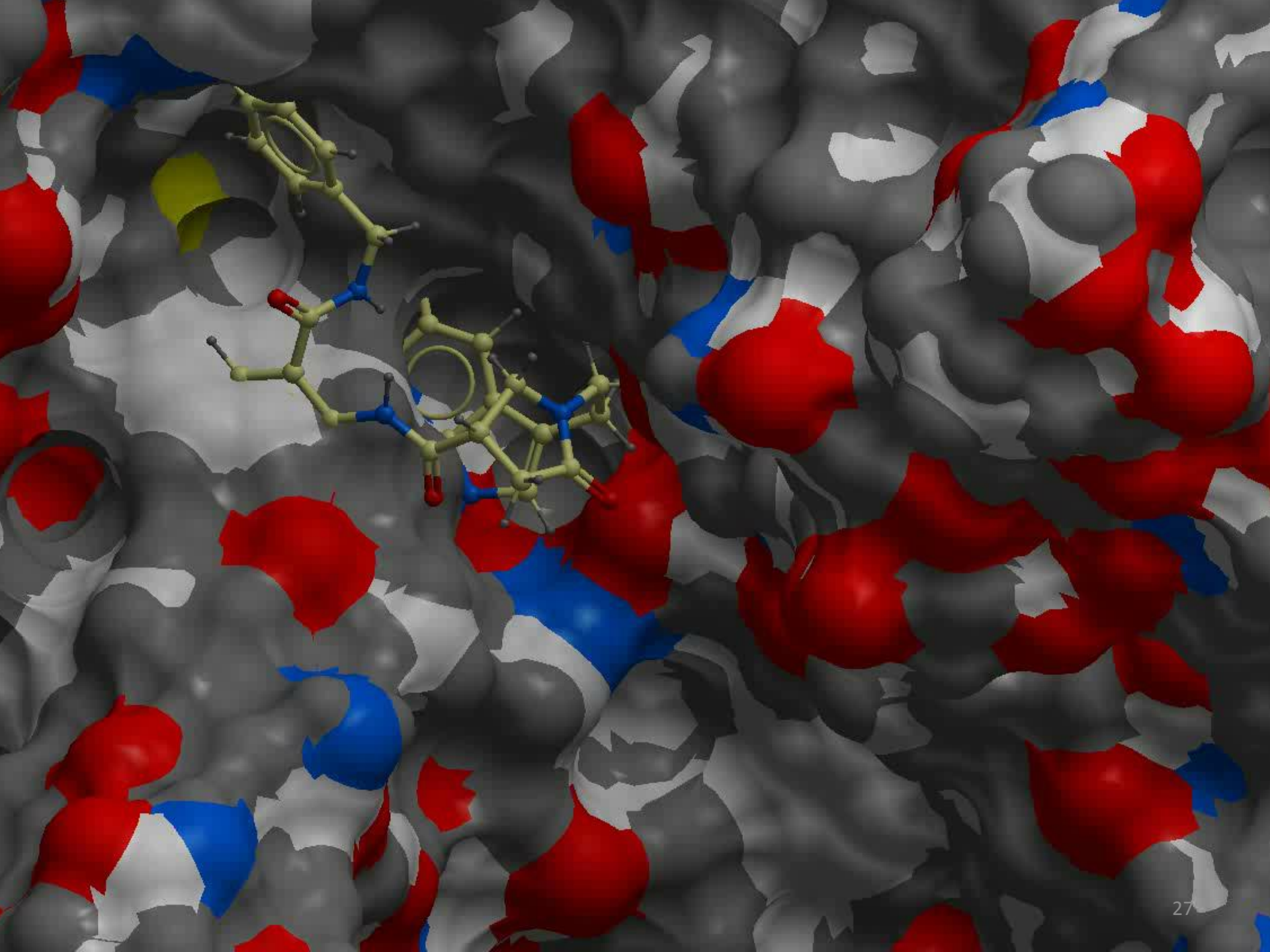
eHiTS (<http://www.simbiosys.ca/ehits/index.html>)

Glide (<http://www.schrodinger.com/ProductDescription.php?mID=6&slD=6&clD=0>)

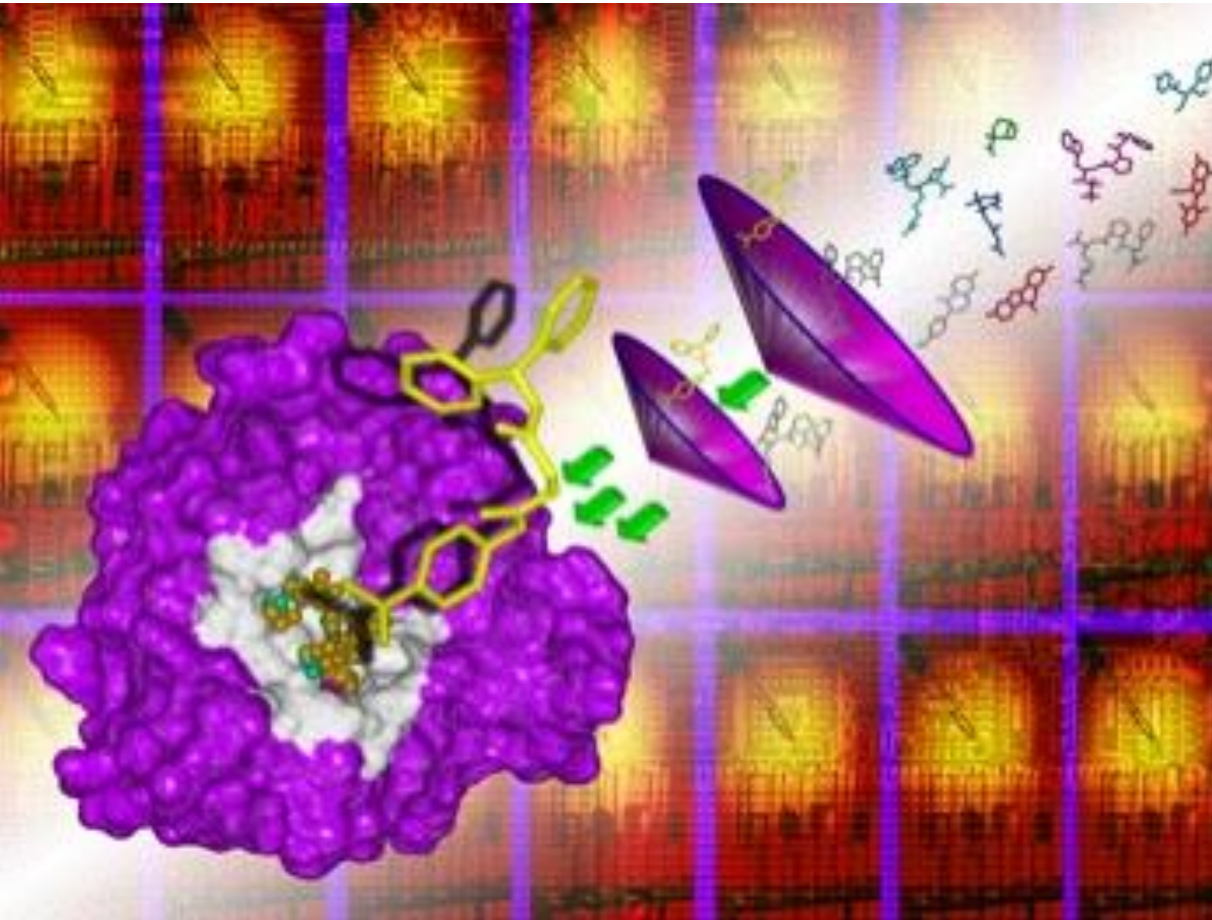
Программы для виртуального скрининга:

VSDocker (<http://bio.nnov.ru/projects/vsdocker2>)

DOVIS (<http://www.bhsai.org/>)



Подбор низкомолекулярных ингибиторов белков с помощью виртуального скрининга: общая схема



Библиотека химических соединений



Фильтры
(правило пяти
Липински)



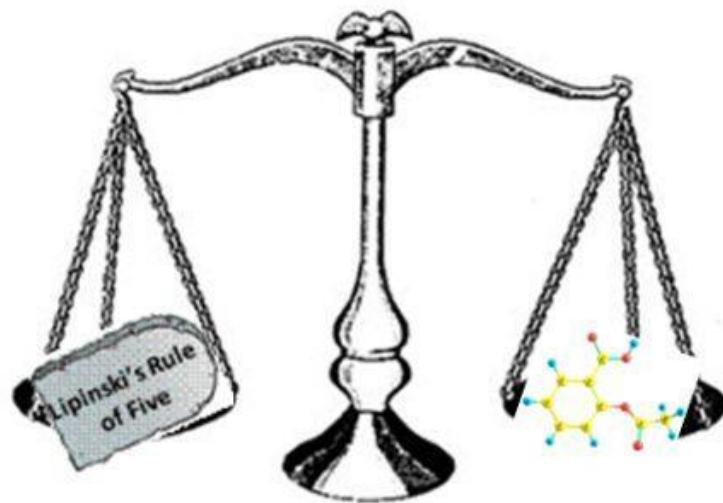
Виртуальный скрининг



Отбор лучших по
оценочной функции
соединений-ингибиторов
(scoring, hits)

Фильтры.

«Правило пяти» Липински.



Christopher Lipinski, PhD

Лекарство должно иметь:

- молекулярную массу не более 500 Да;
- липофильность $\log P < 5$
(P – коэффициент распределения в системе *n*-октанол/вода);
- не более пяти доноров водородной связи;
- не более 10 атомов азота и кислорода
(грубая оценка числа акцепторов водородных связей).

Не всё так просто...

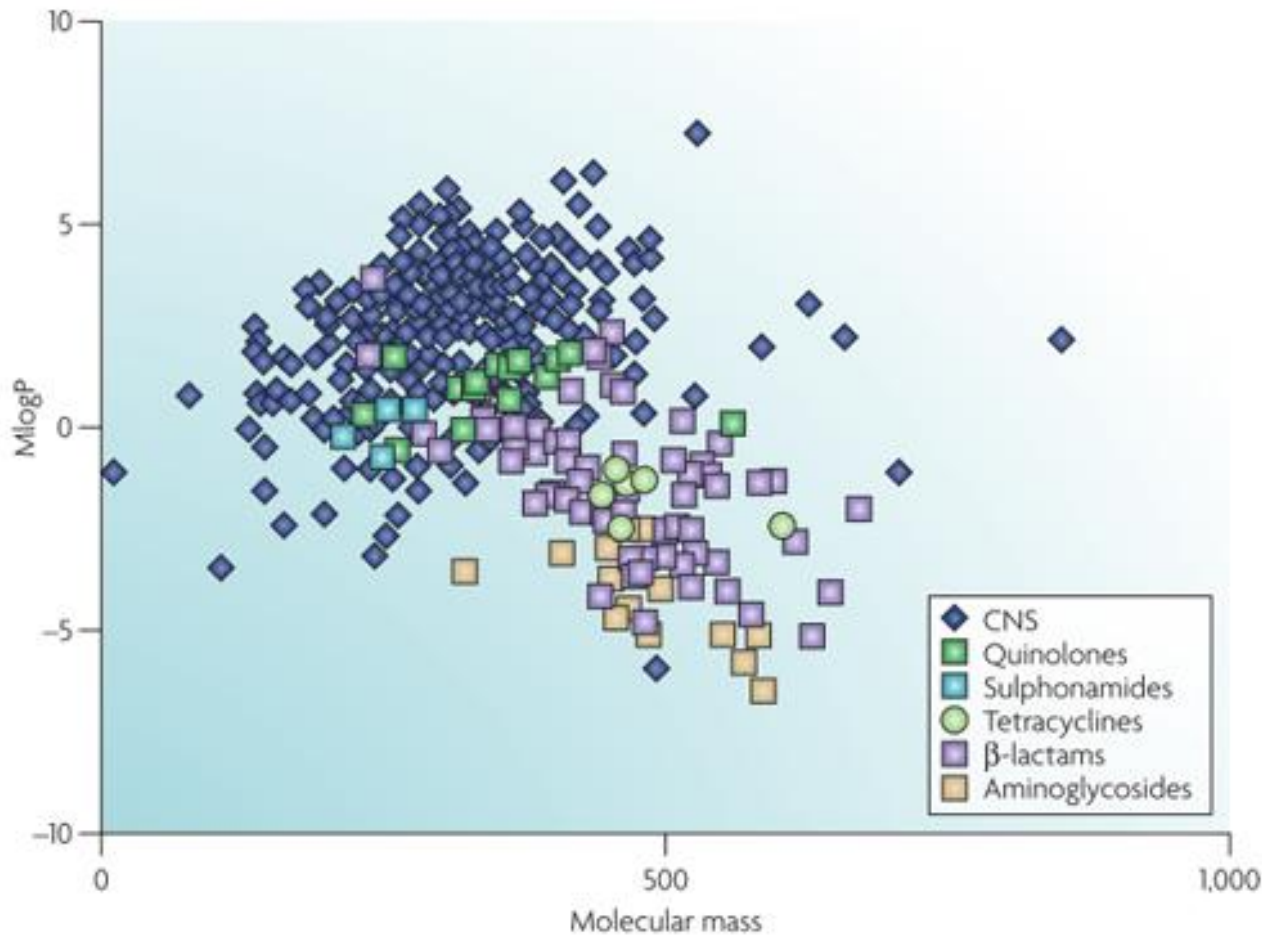
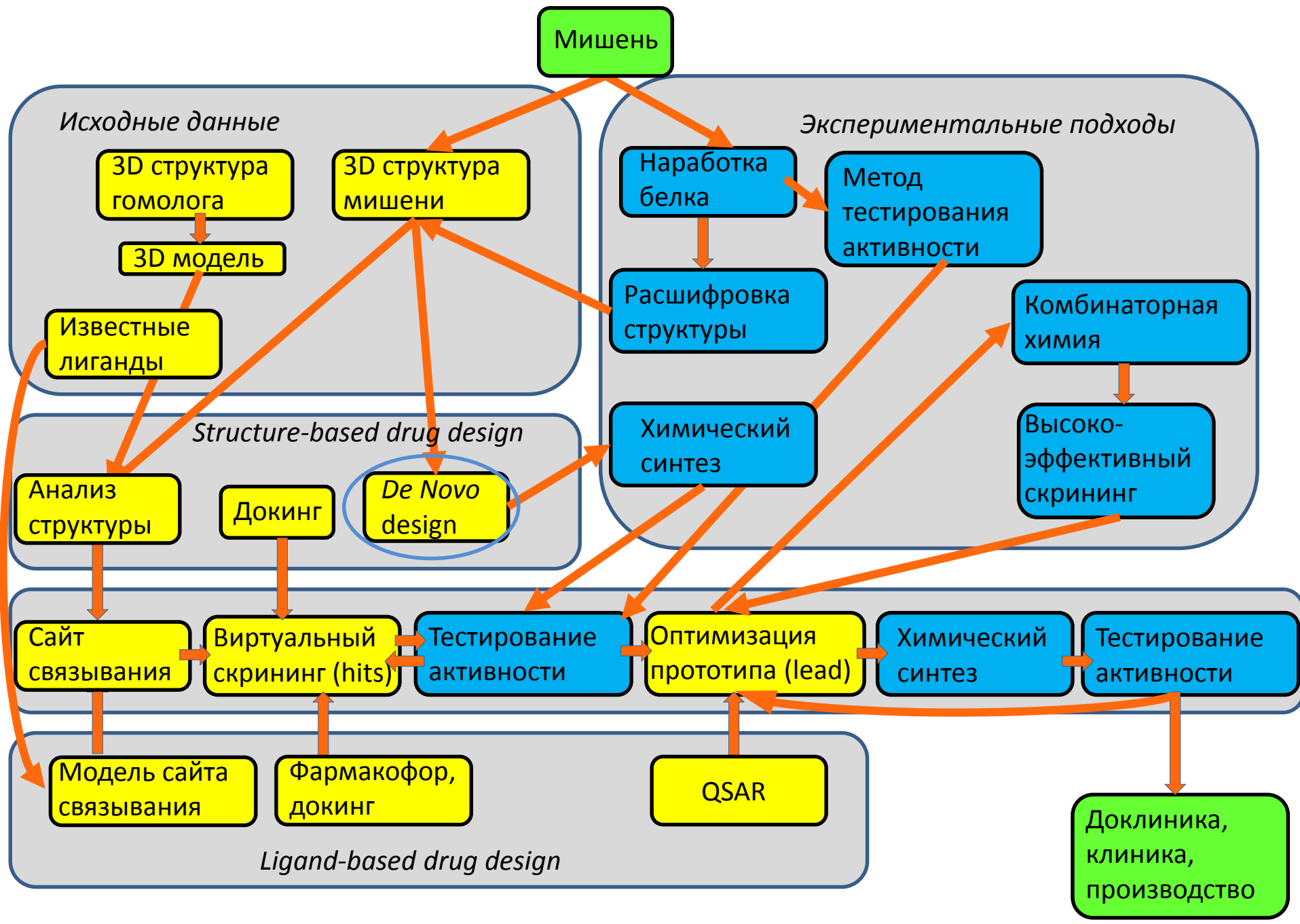


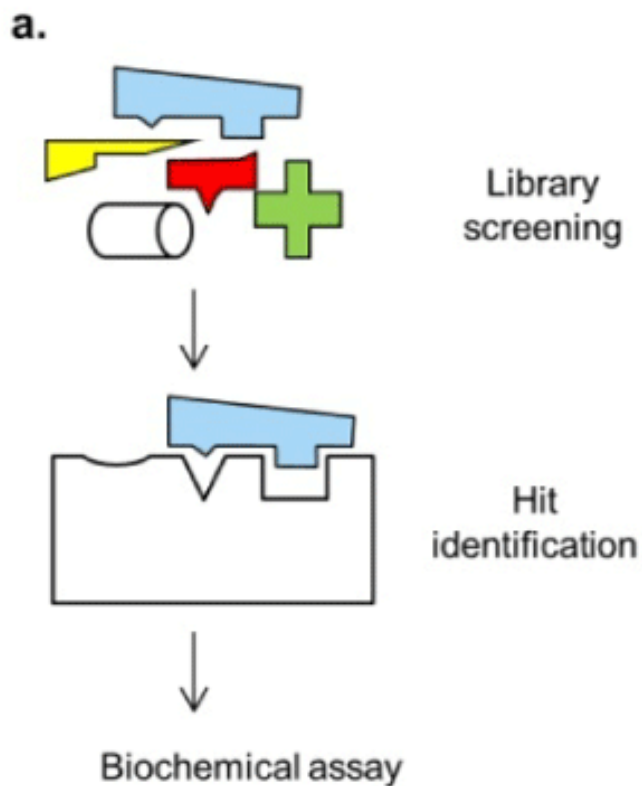
Figure 4 | **The chemical diversity of antibacterials is different to other drugs.** A plot of calculated logP and molecular mass of marketed drugs for central nervous system disorders compared with marketed antibacterial classes. CNS drugs (similar observations for other mammalian target classes; data not shown) closely follow Lipinski's rule of five. Antibacterial molecules are on average more hydrophilic and slightly larger.

Лекарства для ЦНС в большинстве случаев удовлетворяют правилам Липински.
Антибактериальные препараты, как правило, более гидрофильны и большей мол. массы.

Основные стратегии, используемые при компьютерной разработке лекарств



Виртуальный скрининг с помощью молекулярного докинга



Fragment based drug design (*de novo*)

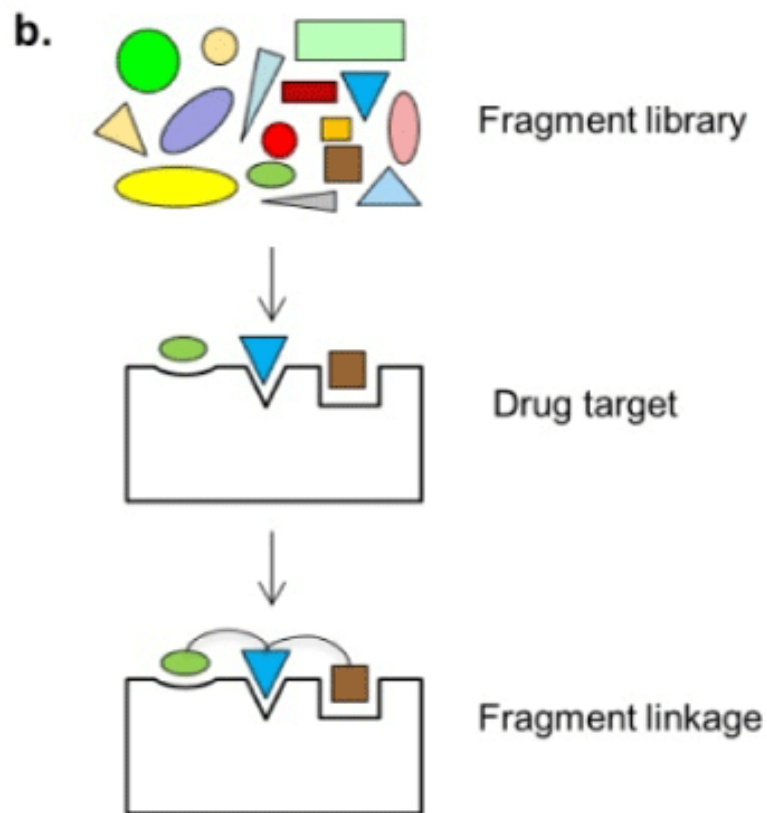
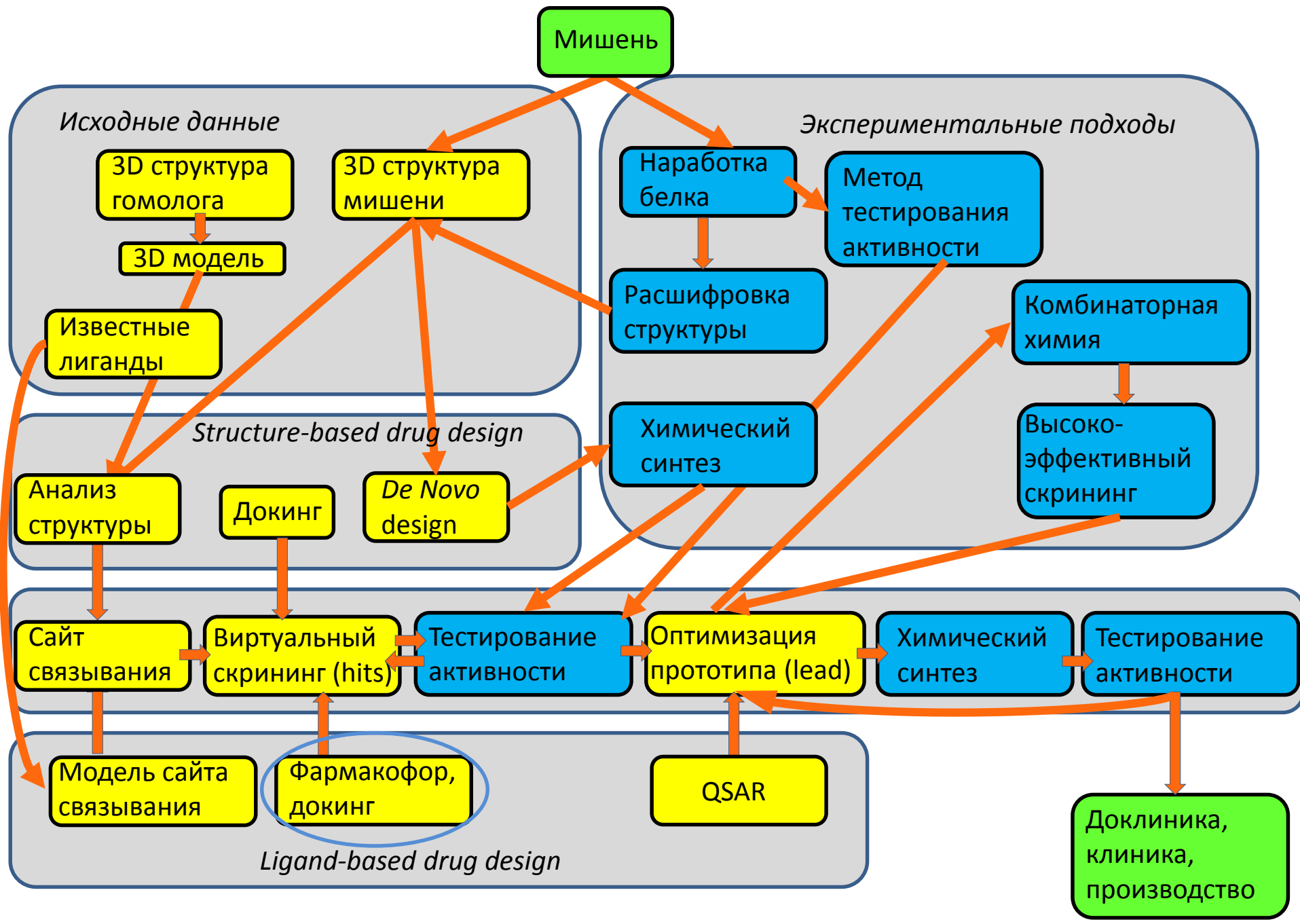
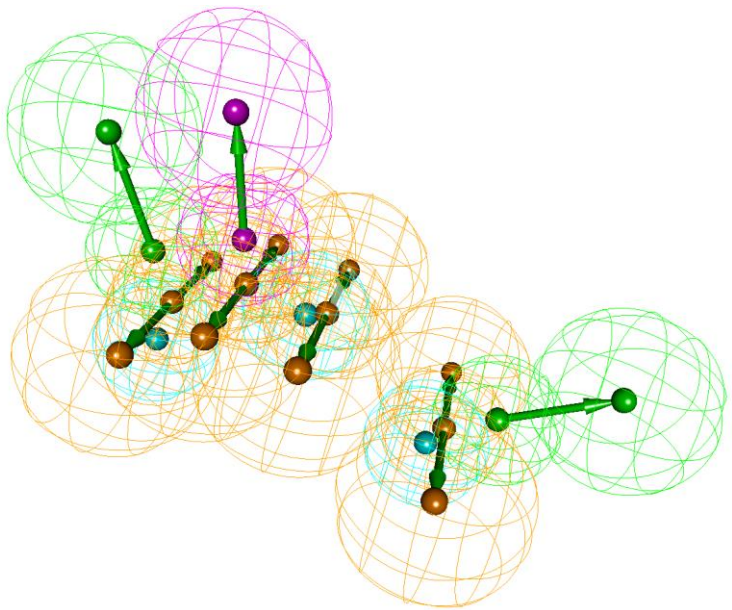


Fig from: Park S, Mann J, Li N (2013) Targeted Inhibitor Design: Lessons from Small Molecule Drug Design, Directed Evolution, and Vaccine Research. Chem Eng Process Tech 1: 1004.

Основные стратегии, используемые при компьютерной разработке лекарств



Ligand-based drug design: построение модели фармакофора



Термин **фармакофор** был введен Паулем Эрлихом в 1909 г.: **Фармакофор** — это молекулярный остов, который несет (*фор*) существенные признаки, ответственные за биологическую активность лекарства (*фармако*) (Ehrlich *Dtsch. Chem. Ges.* 1909, 42: p.17).

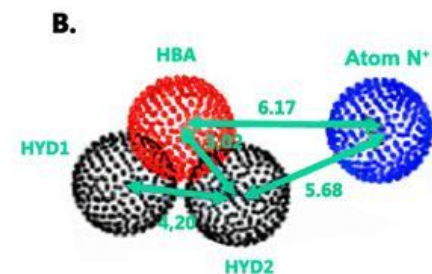
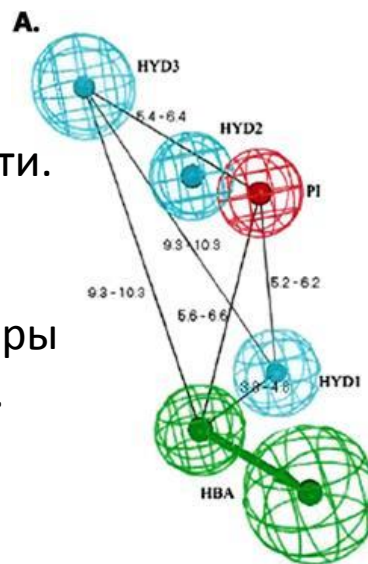
В 1977 г. это определение было модифицировано Питером Гундом (Peter Gund): **Фармакофор** — это набор структурных признаков в молекуле, которые распознаются биологическими рецепторами и являются ответственными за биологическую активность молекулы (Gund. *Prog. Mol. Subcell. Biol.* 1977, 5: pp 117—143).

Определение ИЮПАК (International Union of Pure and Applied Chemistry): **Фармакофор** — это набор пространственных и электронных признаков, необходимых для обеспечения оптимальных супрамолекулярных взаимодействий со специфической биологической мишенью, которые могут вызывать (или блокировать) ее биологический ответ.

Под фармакофорными признаками обычно понимаются фармакофорные центры и интервалы расстояний между ними, необходимые для проявления данного типа биологической активности.

Типичные фармакофорные центры: гидрофобные области, ароматические кольца, доноры и акцепторы водородной связи, анионные и катионные центры.

При фармакофорном поиске проводится поиск соответствия между описанием фармакофора и характеристиками молекул из базы данных, находящихся в допустимых конформациях.



Программное обеспечение для моделирования фармакофоров

[Discovery Studio](#) — от [Accelrys](#)

[LigandScout](#) ([external link](#)) — от [Inte:ligand](#)

[Phase](#) — от [Schroedinger](#)

[MOE](#) — от [Chemical Computing Group](#)

ICM Pro (<http://www.molsoft.com/gui/ph4.html>)

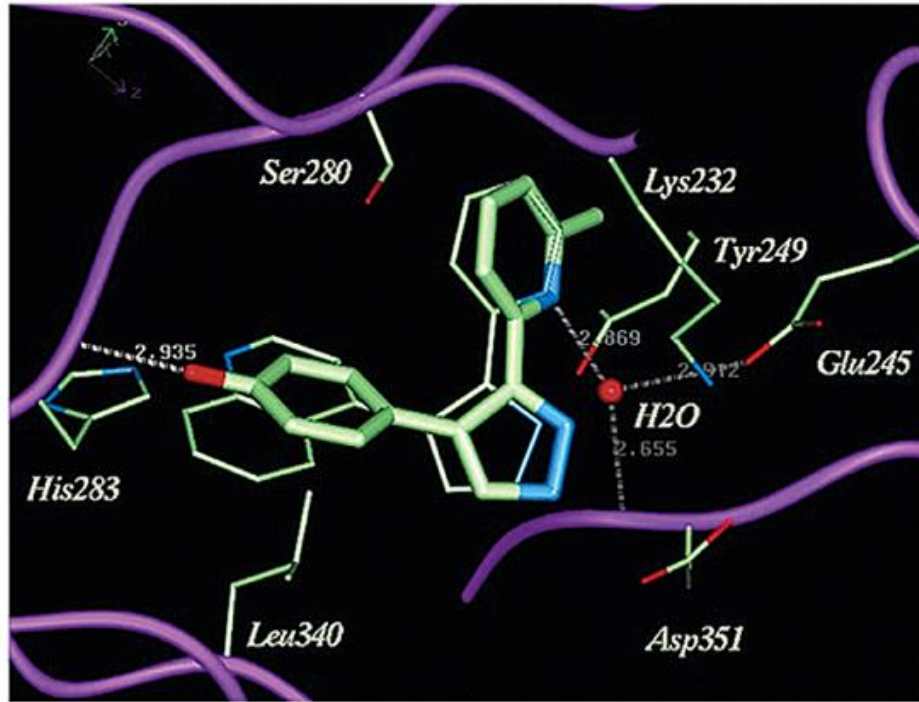
DISCOtech

(http://www.tripos.com/index.php?family=modules,SimplePage,,,&page=sybyl_discotech)

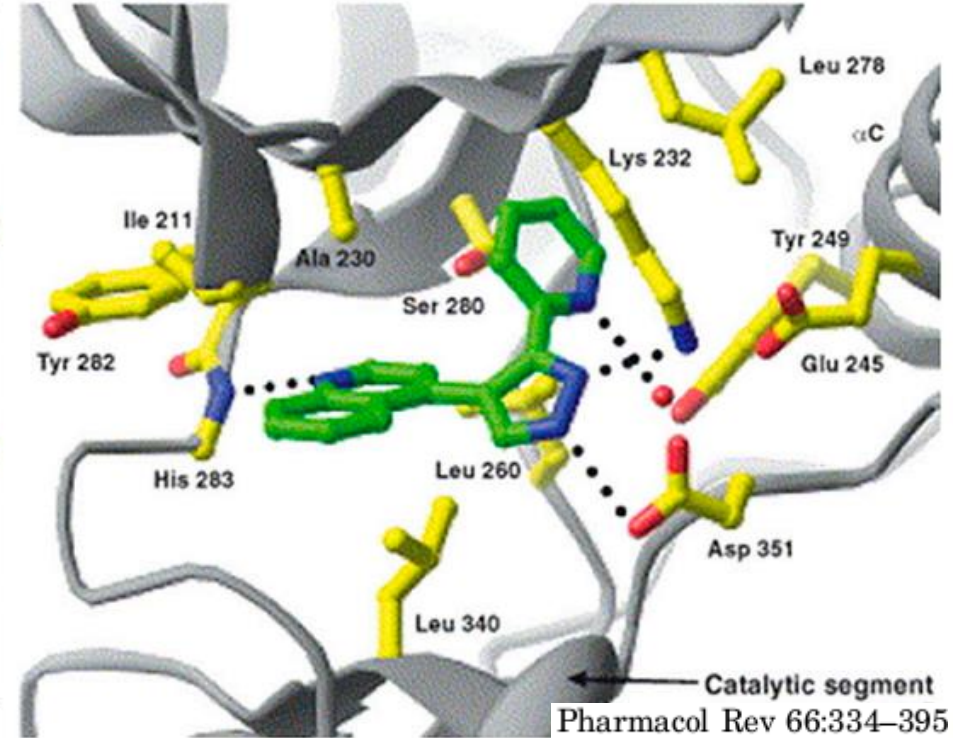
PharmaGist (<http://bioinfo3d.cs.tau.ac.il/PharmaGist/>)

Традиционный химический скрининг и виртуальный скрининг могут дать одинаковый результат

I)



II)

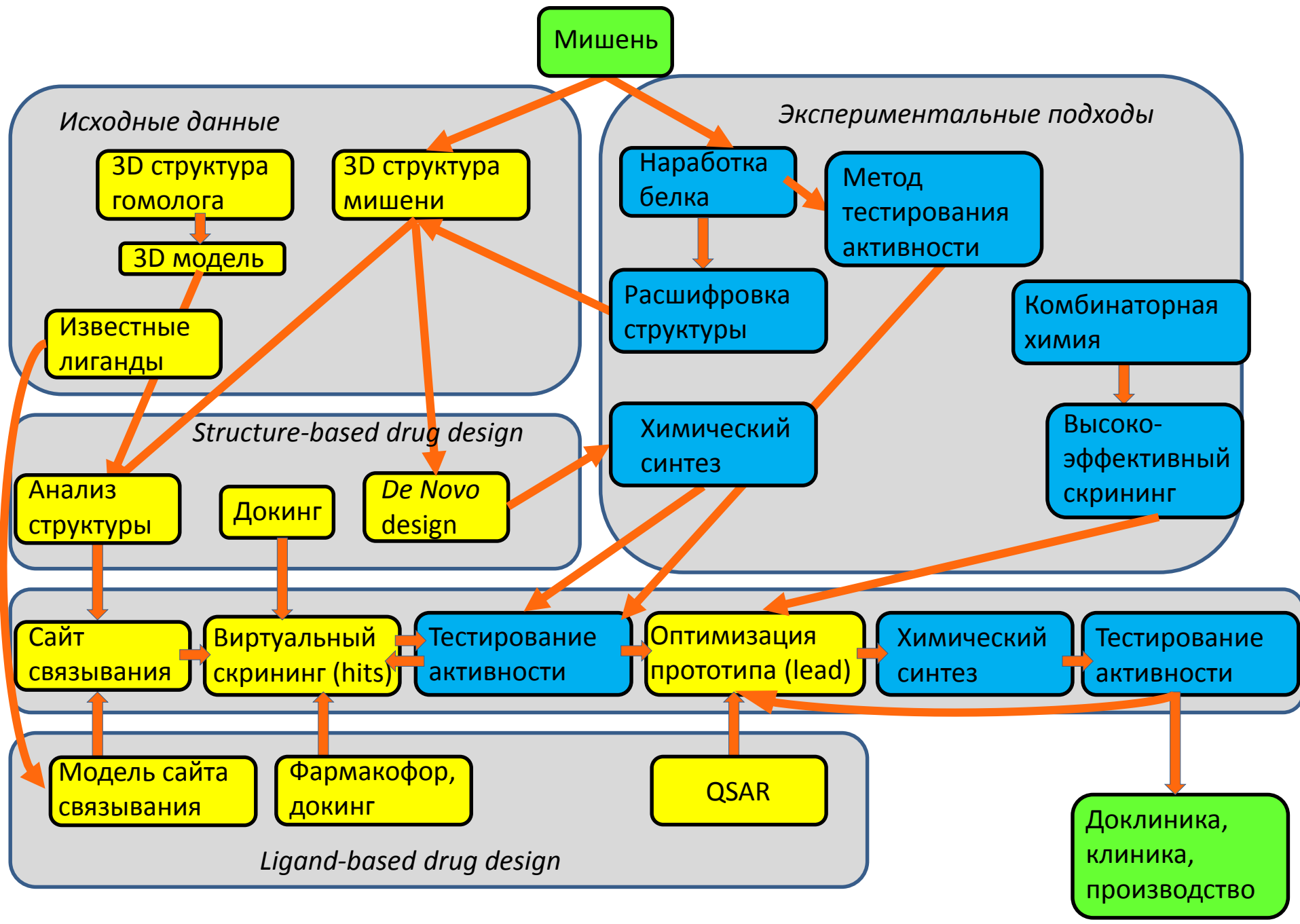


Пример: Поиск ингибиторов киназного домена рецептора TGF β 1 (Transforming growth factor beta 1).

Компания Eli Lilly использовала традиционный химический скрининг для поиска лидерного соединения, затем провела его оптимизацию (I) (Sawyer et al., 2003).

Biogen Idec использовала виртуальный скрининг на основе структуры более слабого ингибитора и TGF β 1R (Singh et al., 2003a). Было отобрано 87 соединений, лучшее из которых (II) совпало по структуре с соединением, отобранным на основании традиционного химического скрининга в Eli Lilly.

Основные стратегии, используемые при компьютерной разработке лекарств



QSAR (Quantitative Structure-Activity Relationship) анализ

Поиск количественных соотношений структура-свойство — процедура построения моделей, позволяющих по структурам химических соединений предсказывать их разнообразные свойства.

Поиск количественных соотношений структура-свойство основан на применении методов математической статистики и машинного обучения для построения моделей, позволяющих по описанию структур химических соединений предсказывать их свойства (физические, химические, биологическую активность).

Ресурсы, позволяющие строить новые модели структура-свойство

[Online Chemical Modeling \(OCHEM\)](#) — информационный и вычислительный ресурс, позволяющий работать через Web-интерфейс с базой данных по органическим соединениям и их свойствам, пополнять её, осуществлять в ней поиск и формировать выборки, рассчитывать широкий набор молекулярных дескрипторов, строить количественные модели структура-свойство и применять их для прогнозирования свойств новых соединений

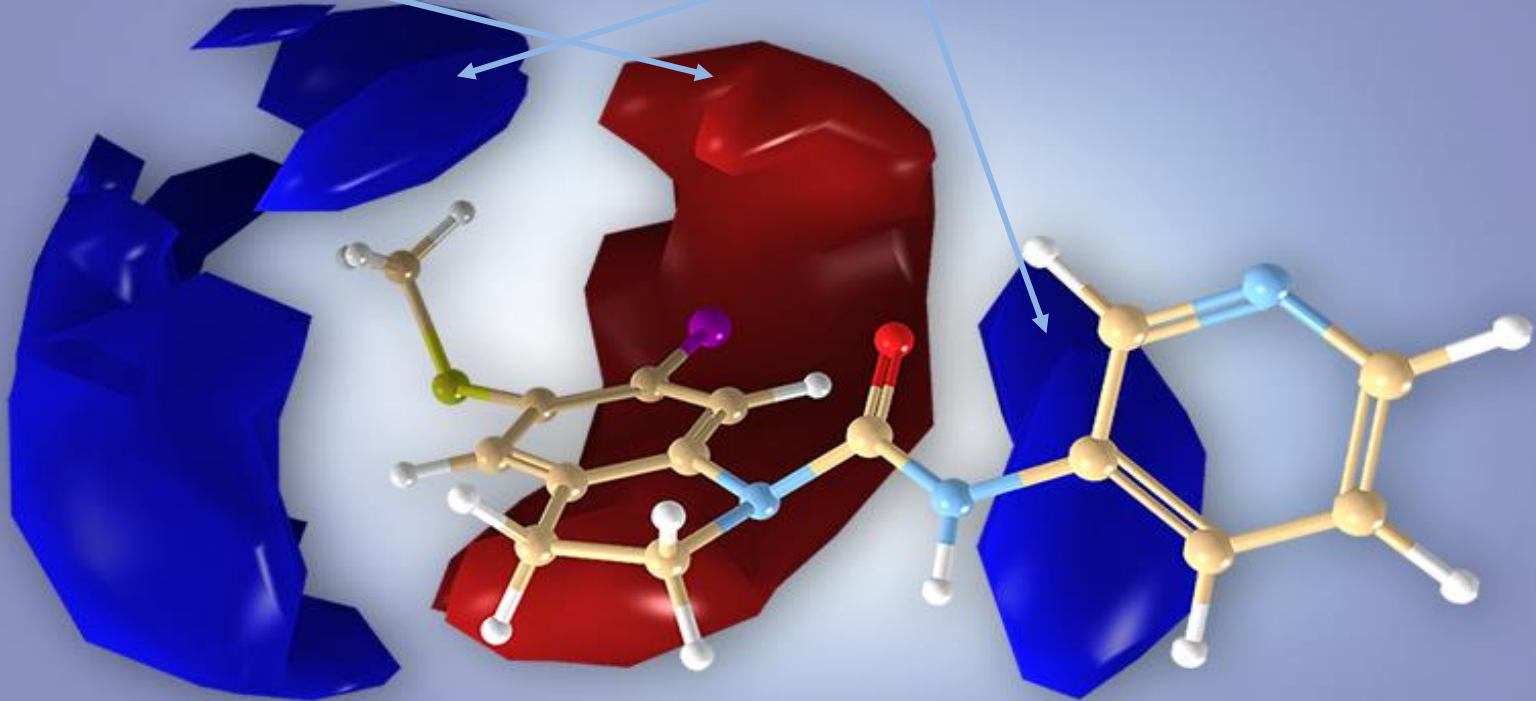
[Chembench](#) — ресурс, позволяющий строить модели структура-свойство и использовать их для прогнозирования

3D-QSAR (Quantitative Structure-Activity Relationship) анализ

3D-QSAR – это расширение классических QSAR подходов с учетом трехмерной структуры лигандов для предсказания их биологической активности с использованием современных хемометрических методик.

Взаимодействие с красными областями негативно отражается на активности лиганда

Взаимодействие с синими областями положительно влияет на активность лиганда



Основные стратегии, используемые при компьютерной разработке лекарств

