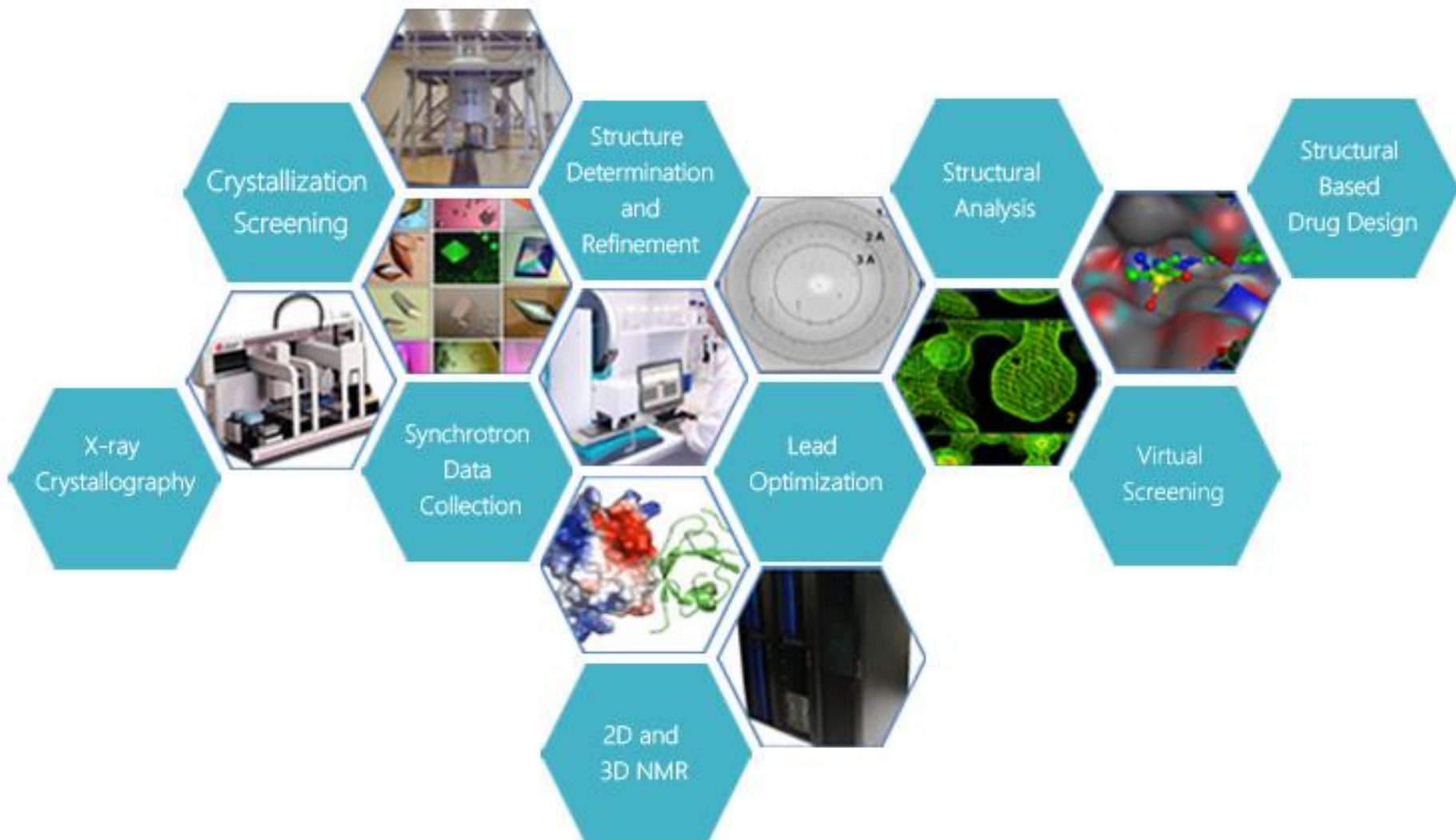


# Биоинформатика и разработка лекарств

Анна Станиславовна Карягина

НИИ ФХБ им.А.Н. Белозерского МГУ им. М.В Ломоносова

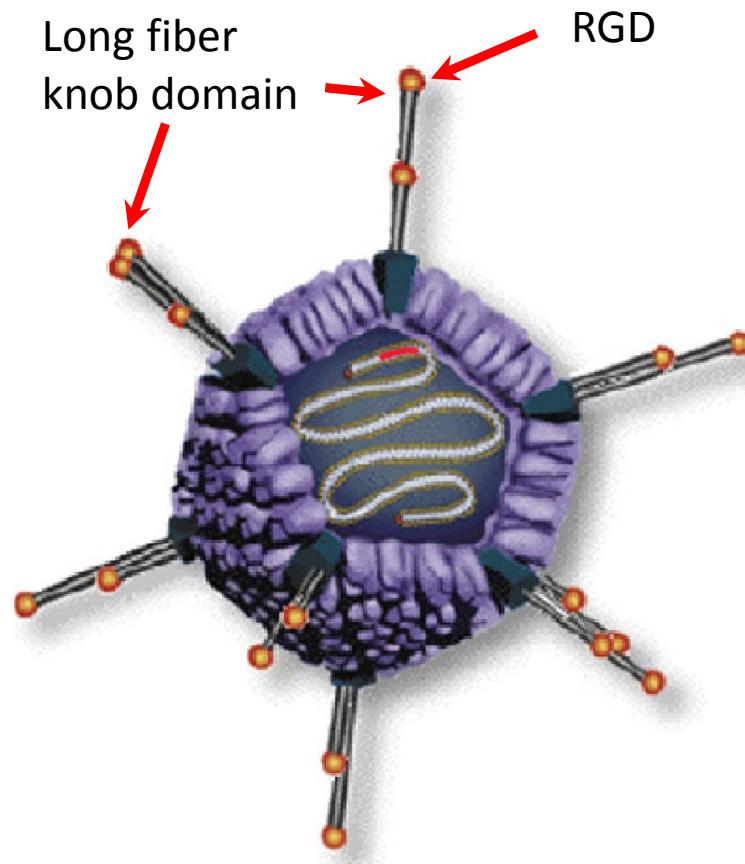
ФГБУ «ФНИЦ эпидемиологии и микробиологии им. Н.Ф. Гамалеи» Минздрава РФ



Как можно использовать биоинформатику для разработки лекарств?

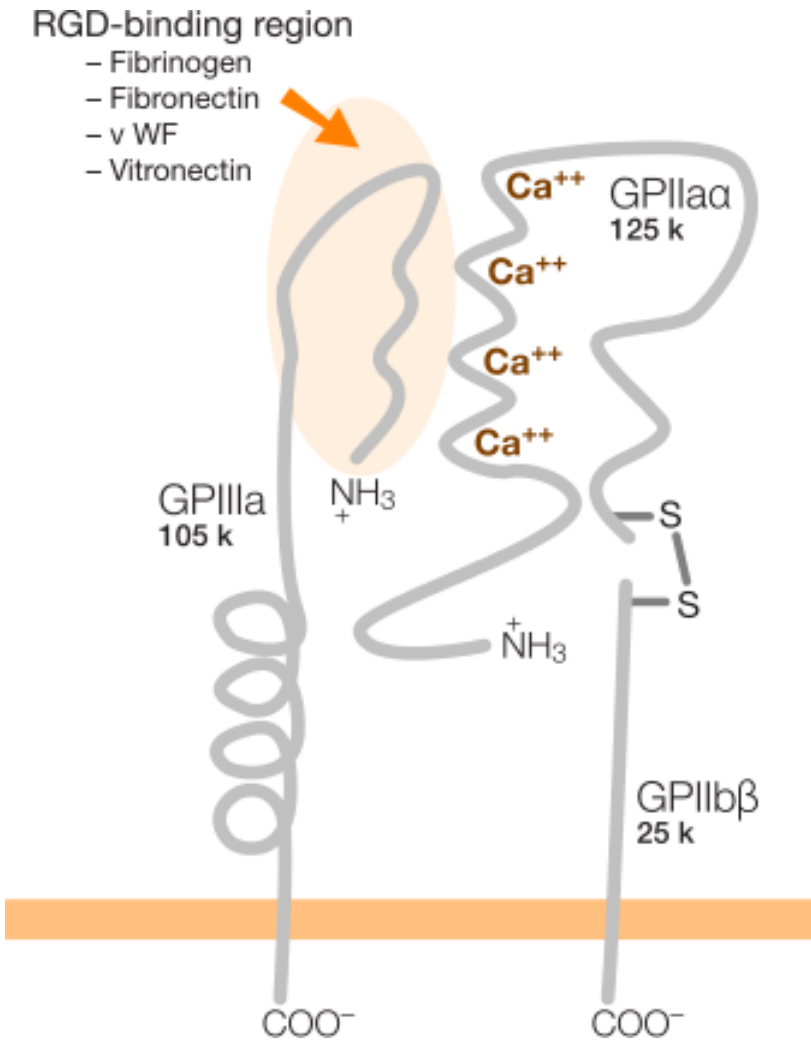
*Пример использования «обычных» методов биоинформатики:*

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

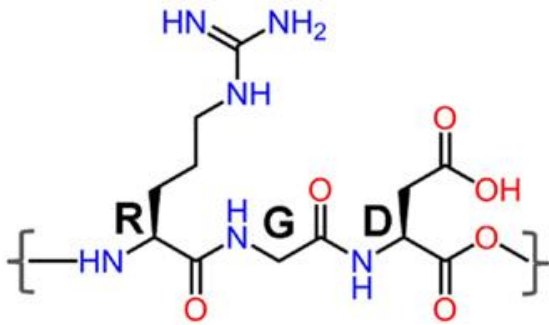


- Птичьи аденовирусы можно использовать в качестве векторов для генной терапии.  
*Преимущества по сравнению с аденовирусами человека:*
- Вирус CELO (Chicken Embryo Lethal Orphan) можно наращивать в куриных эмбрионах (до  $10^{12}$  вирусных частиц на эмбрион) – это гораздо проще и дешевле, чем выращивание 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).

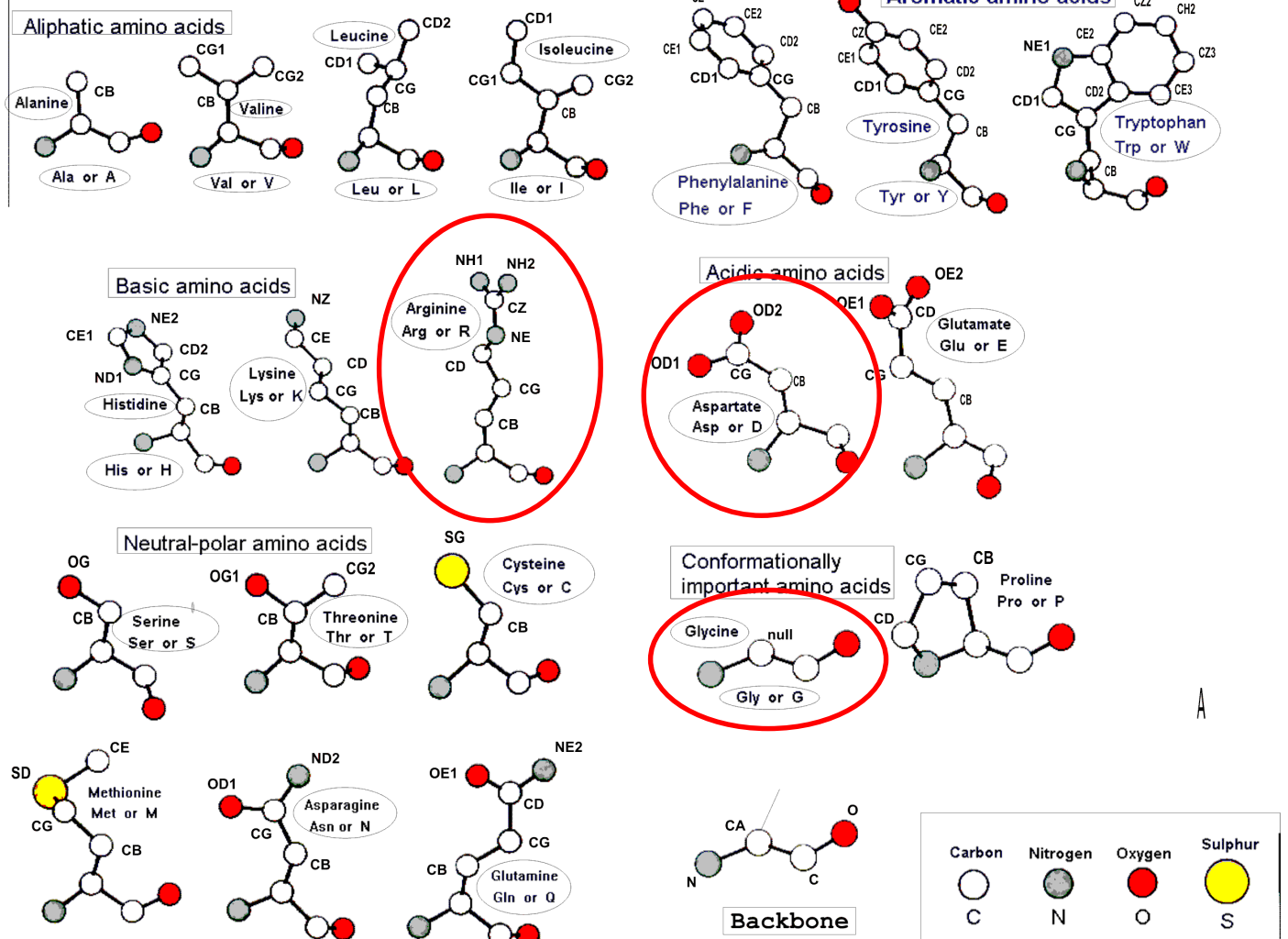
**Задача:** С целью изменения тропизма вектора (способности проникать в клетки другого типа, несущие на поверхности рецепторы - **интегрины**) было необходимо вставить аминокислотные остатки **RGD** в поверхностно экспонированный участок длинного фибера аденовируса CELO.

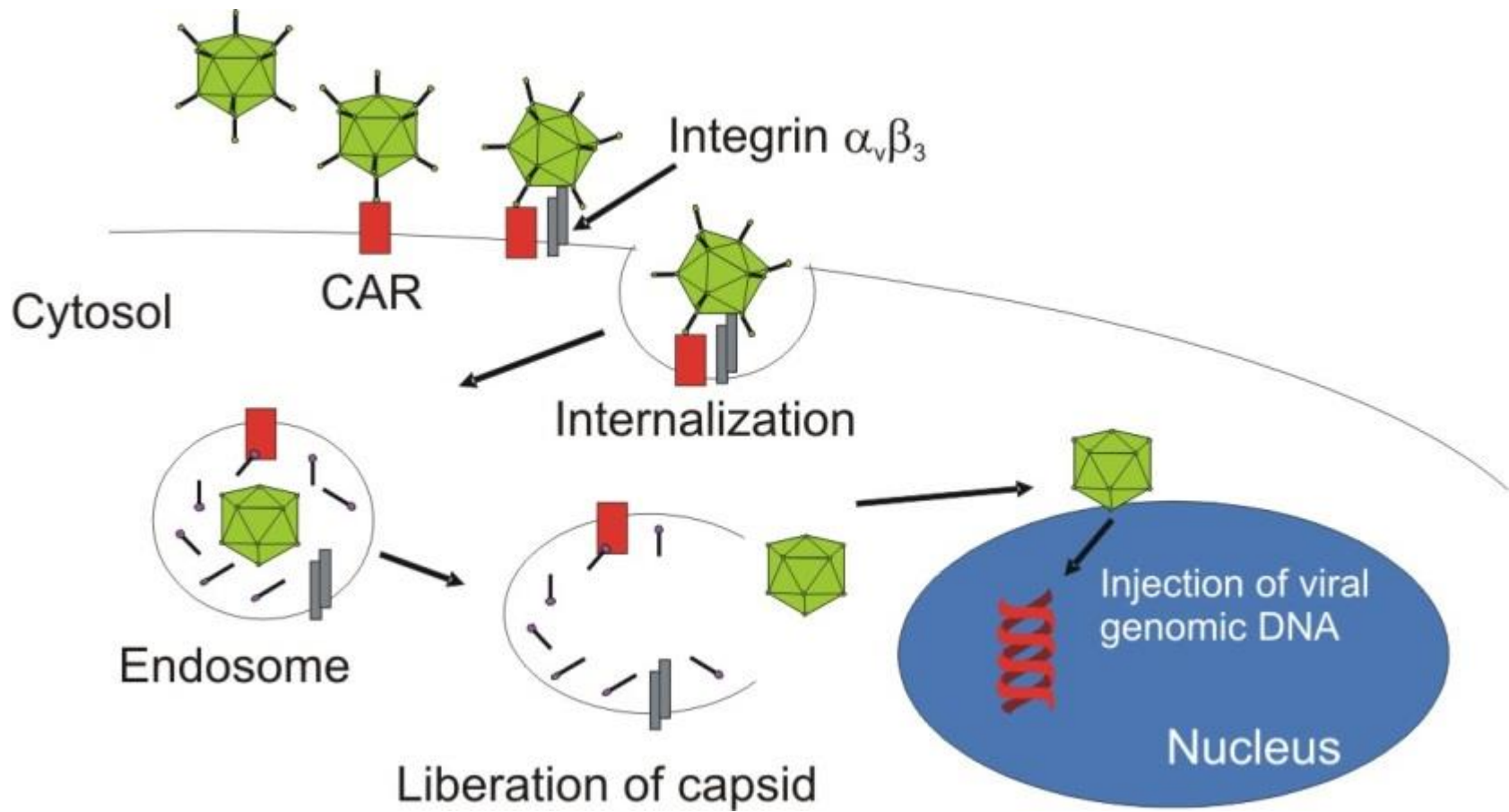


**Интегрины** — это трансмембранные гетеродимерные клеточные рецепторы, взаимодействующие с внеклеточным матриксом и передающие различные межклеточные сигналы. От них зависит форма клетки, её подвижность, они участвуют в регулировке клеточного цикла. Структурно интегриновые рецепторы представляют собой облигатные гетеродимеры — каждый состоит из одной альфа- и одной бета-субъединицы. Молекулярная масса субъединиц варьирует от 90 до 160 кДа. 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.



# Что такое RGD?





## CELO large fiber knob domain

630            640            650            660            670            680            690            700            710  
          G                            H                            I                            J  
          EEEE                            EEEEEEE                            EEEEEEEEEEE                            EE                            EEEEE  
SNLVTYTKIHFKPTGSQQQLQLASGELVVA~~AAKSPV~~QTTKYELIYLGFTLKQNS SGTNFFDPNASSDLS **FLTPPIPF**TYLGYQ

## Ad5 fiber knob domain

510            520            530            540            550            560            570            580  
          G                            H                            I                            J  
          EEEEEEEE                            EEEEEEE                            EEEEEEE                            EE                            EEEEE  
KSHGKTAKSNIVSQVYLN~~GDKTKPV~~TLTITLNGTQETGDTTPSAYSMSFSWDWSGHNYINEI **FATSSYTF**SYIAQE

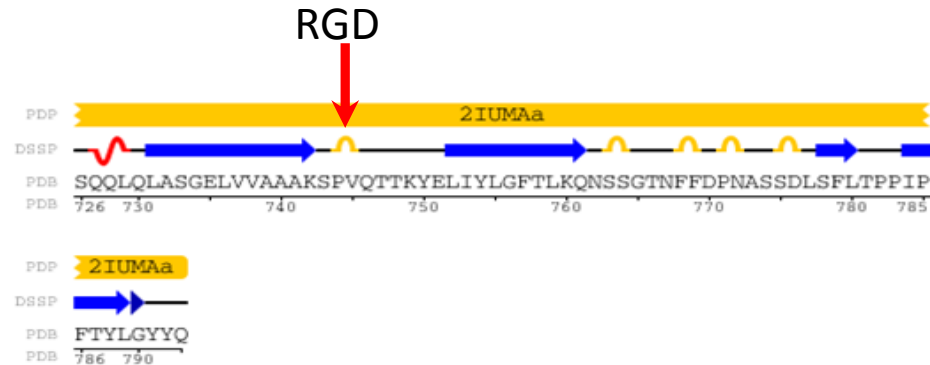
EEEE – бета тяжи

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

**Путь решения:**

Применение биоинформатических подходов: множественное выравнивание аминокислотных последовательностей с совмещением предсказанных элементов вторичной структуры белка.

**Результат:** В аминокислотной последовательности фибера CELO было выбрано место для вставки RGD, по нашим предсказаниям, локализирующееся на поверхности. 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<sup>∇</sup>

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

*Gamaleya Research Institute for Epidemiology and Microbiology (GIEM), 123098, Gamaleya Street 18, Moscow, Russia,<sup>1</sup> 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, Russia<sup>2</sup>*

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,<sup>1</sup> Antonio L. Llamas-Saiz,<sup>2</sup> Gavin C. Fox,<sup>3</sup>  
Patrick Langlois<sup>4</sup> and Mark J. van Raaij<sup>1,2</sup>

Correspondence  
Mark J. van Raaij  
vanraaij@usc.es

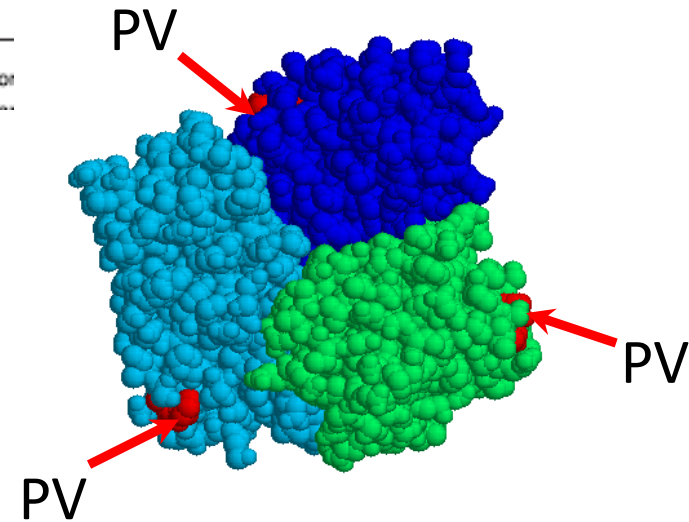
<sup>1</sup>Departamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain

<sup>2</sup>Unidad 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

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

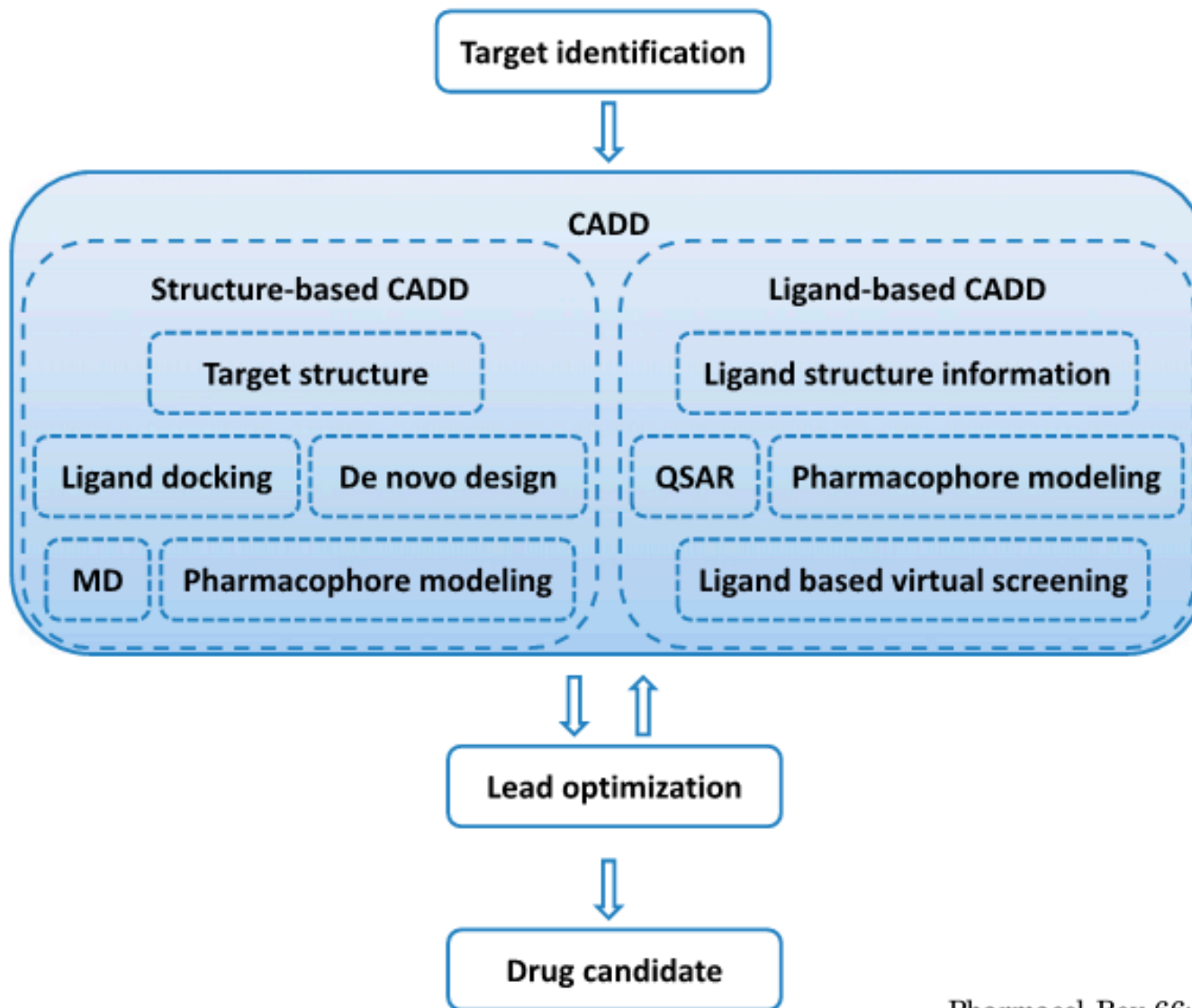
<sup>4</sup>Agence Française de Sécurité Sanitaire des Aliments, Unité Génétique Virale et Biosecurité Site Les Croix, BP 53, F-22440 Ploufragan, France

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.

# Методы, используемые при компьютерной разработке лекарств Computer-Aided Drug Design (CADD)



# Что такое лекарство?

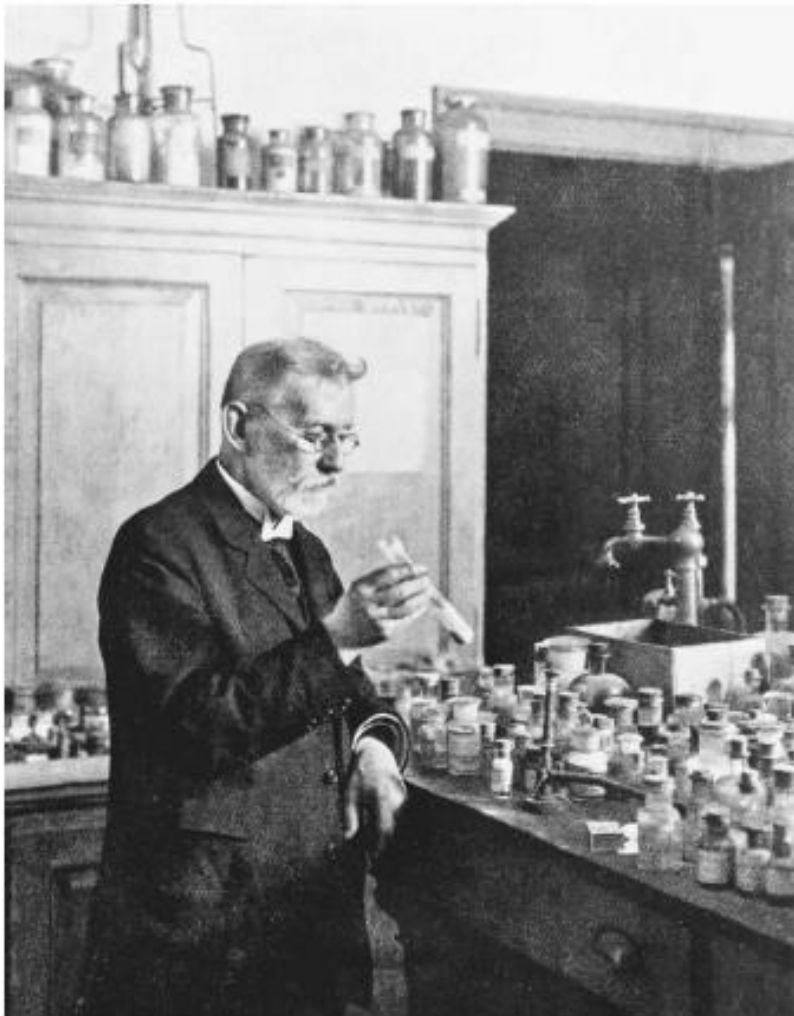
**Лекарственное средство, лекарственный препарат, медикамент, лекарство** — вещество или смесь веществ синтетического или природного происхождения в виде лекарственной формы (таблетки, капсулы, раствора, мази и т. п.), применяемое для профилактики и лечения заболеваний.



# Сколько существует лекарств?



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



# Немного истории...

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

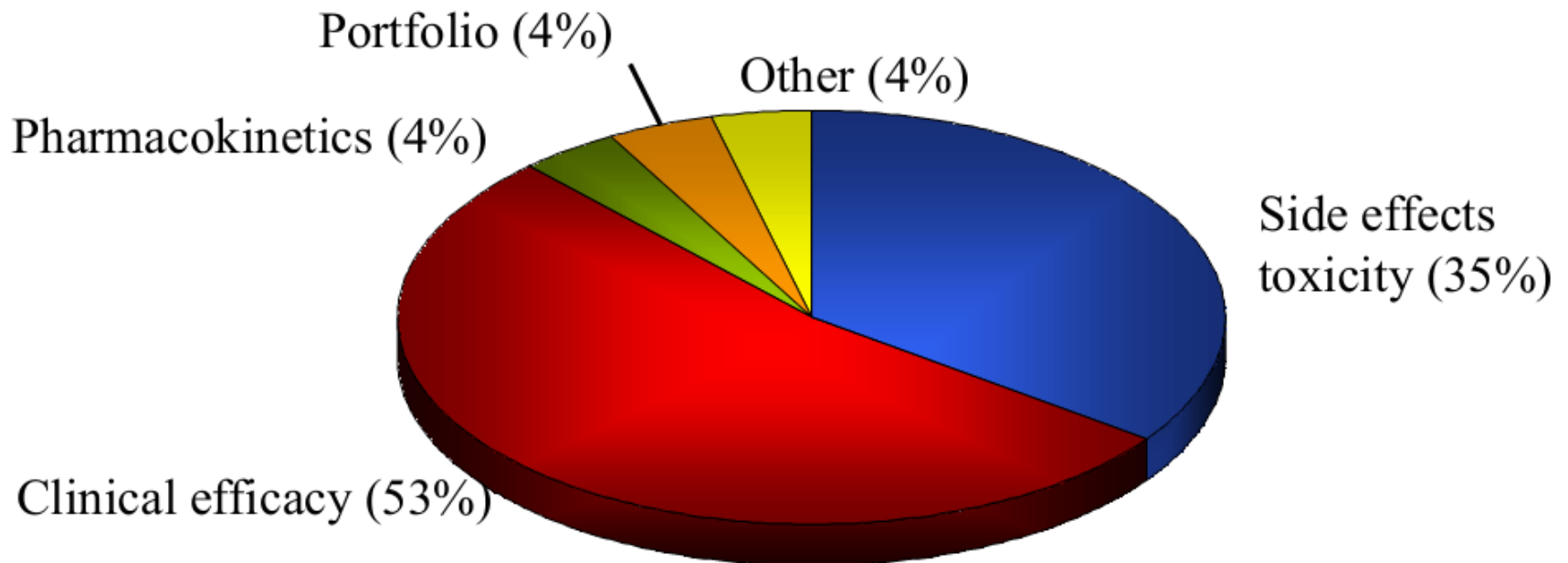
Он выдвинул гипотезу о существовании *хеморецепторов* — специальных тканевых структур, специфически взаимодействующих с химическими веществами, и постулировал возможность использования этого феномена в терапии различных заболеваний.

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.

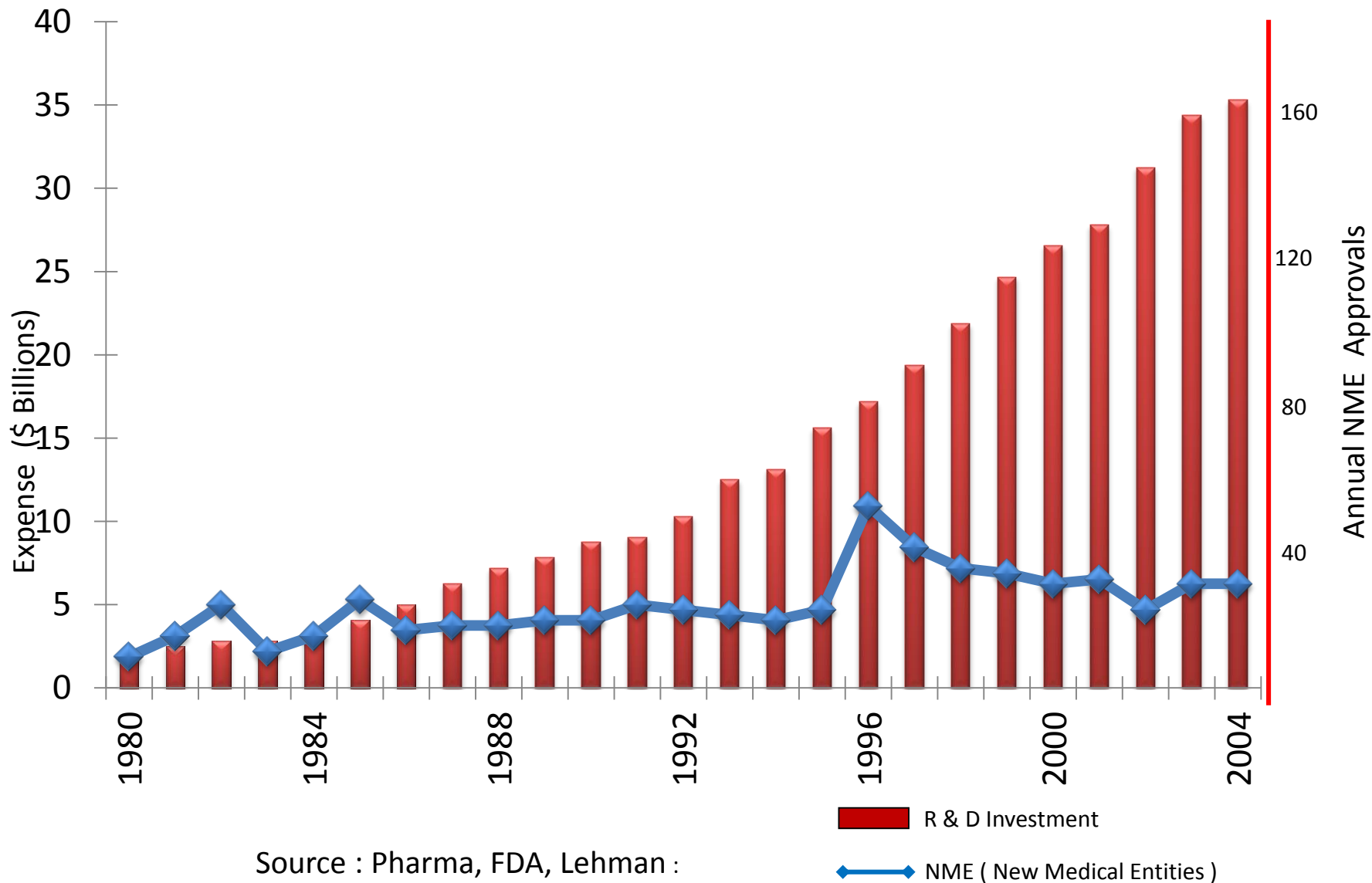
# Особенности разработки лекарств на современном этапе

- 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

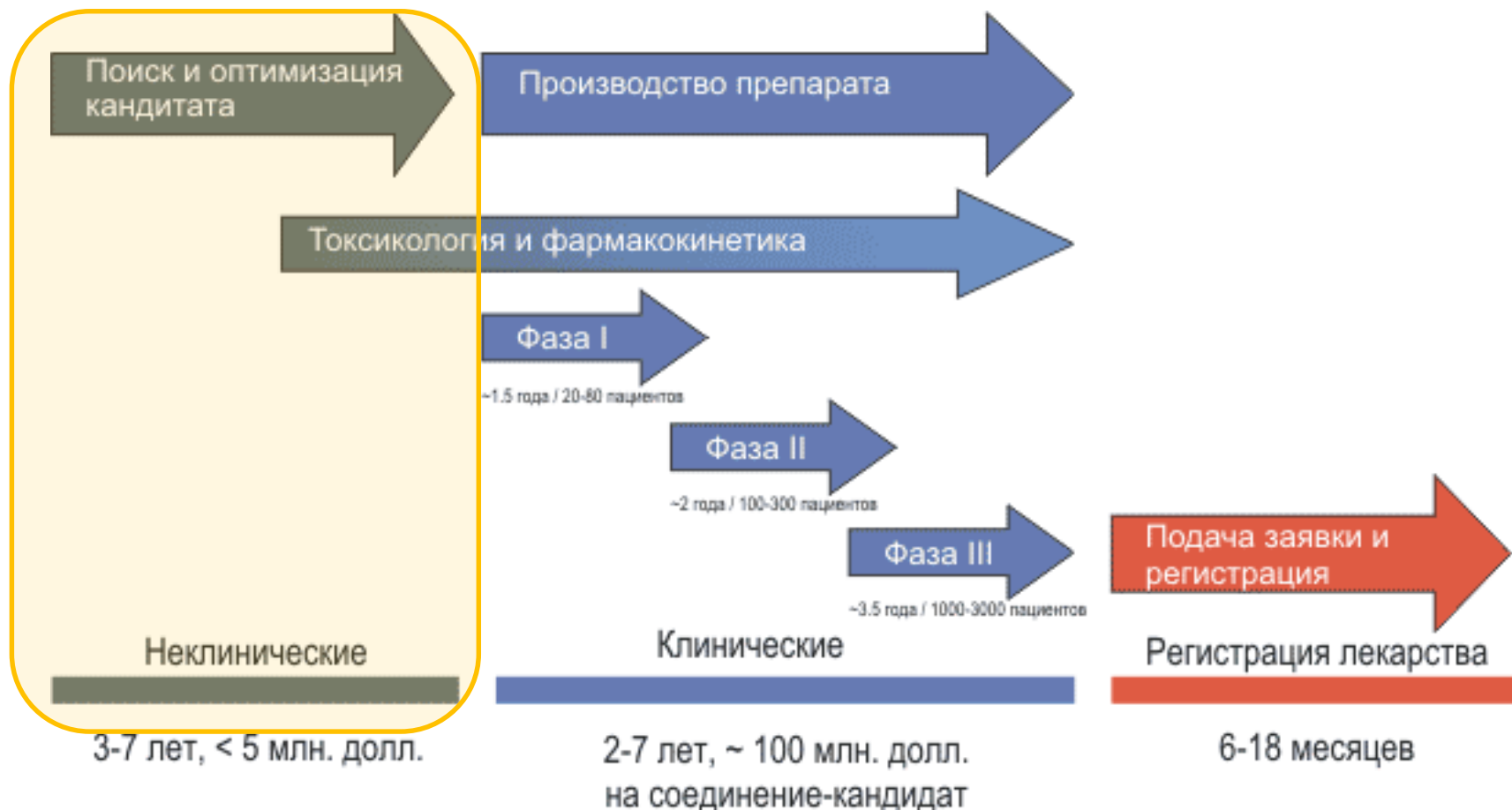


# Этапы создания нового лекарства





# 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 (hits, **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.

# Использование новых технологий при разработке лекарств



*Болезнь*

**ГЕНОМИКА, ПРОТЕОМИКА, ФАРМАКОЛОГИЯ**

*Идентификация мишеней*

**ВЫСОКОПРОИЗВОДИТЕЛЬНЫЙ СКРИНИНГ**

*Скрининг до 100000 соединений в день на активность в отношении мишени*

**ВИРТУАЛЬНЫЙ СКРИНИНГ**

*Использование компьютера для поиска потенциальных лигандов*

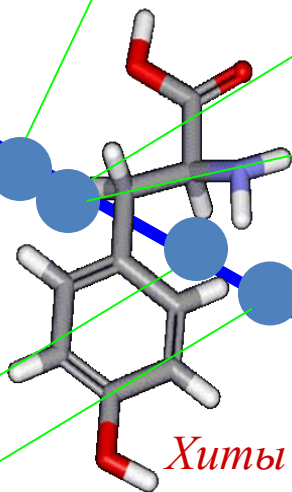
**ФАРМАКОФОРНЫЙ СКРИНИНГ**

*Поиск потенциальных лигандов, если структура мишени неизвестна*

**МОЛЕКУЛЯРНОЕ МОДЕЛИРОВАНИЕ**

*Моделирование структуры мишени и ее комплекса с лигандом для повышения активности отобранных соединений*

*Мишень*



*Хиты*

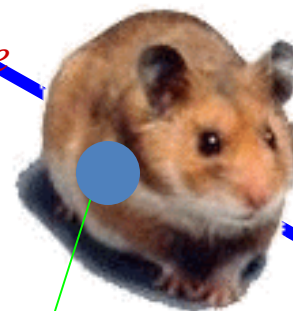
*Лидерное соединение*

**КОМБИНАТОРНАЯ ХИМИЯ**

*Быстрое получение множества соединений*

**IN VITRO & IN SILICO ADME (absorption, distribution, metabolism and excretion) МОДЕЛИРОВАНИЕ**

*Тестирование на культурах тканей и компьютерное моделирование снижают количество экспериментов на животных*

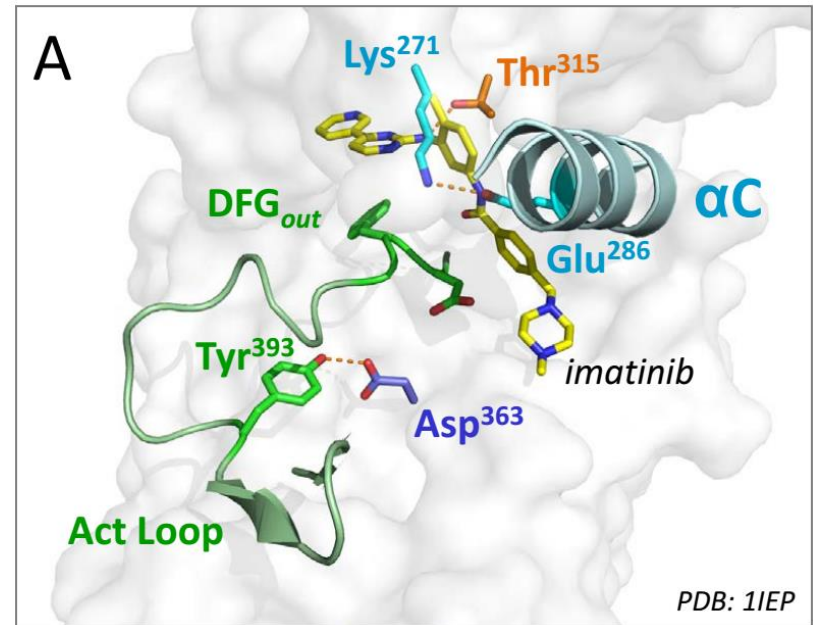
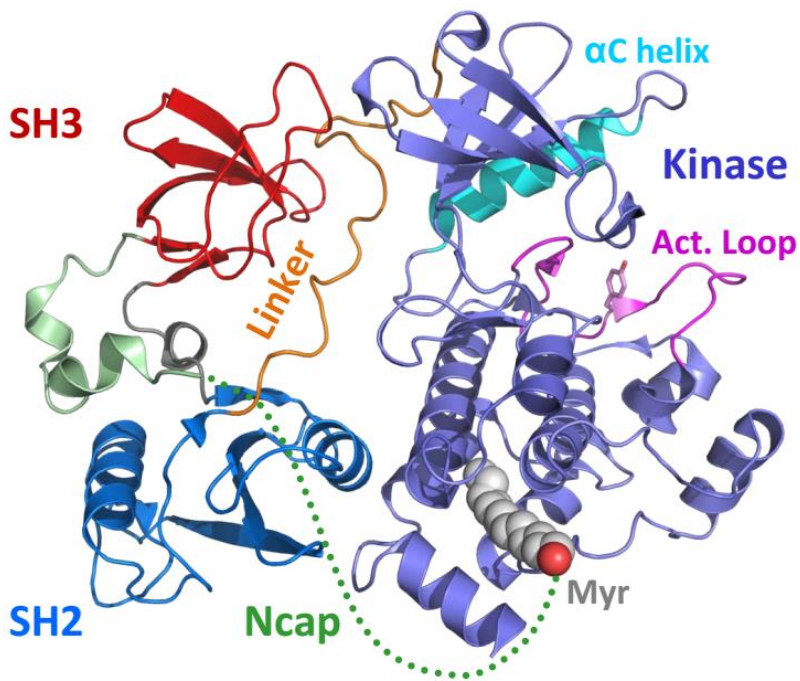


*Доклинические исследования*

# Примеры лекарств, разработанных с применением компьютерных подходов

Capoten	Captopril	<chem>CC1(C)CC(C(=O)CSC1)C(=O)O</chem>	ACE (angiotensin- converting- enzyme)	Hypertension	1981	Bristol- Myers Squibb
Trusopt	Dorzolamide	<chem>CC1(C)C(S(=O)(=O)N)C(S1)C(=O)NCC</chem>	Carbonic anhydrase	Glaucoma	1995	Merck
Viracept	Nelfinavir	<chem>CC(C)(C)NC(=O)N(C)C1CCN(C1)C(C)C(O)C(=O)NCC2=CC=C(O)C=C2</chem>	HIV protease	HIV/ AIDS	1999	Agouron (Pfizer) and Lilly
Tamiflu	Oseltamivir	<chem>CC(C)C(O)C(=O)NCC1=CC=C(N)C=C1C(=O)OCC</chem>	Neuraminidase	Influenza	1999	Gilead and Roche
Gleevec	Imatinib	<chem>CC1=CC=C(C=C1)C(=O)NCC2=CC=CC=C2N3C=NC=CC3</chem>	BCR- Abl	Chronic myelogenous leukaemia	2001	Novartis

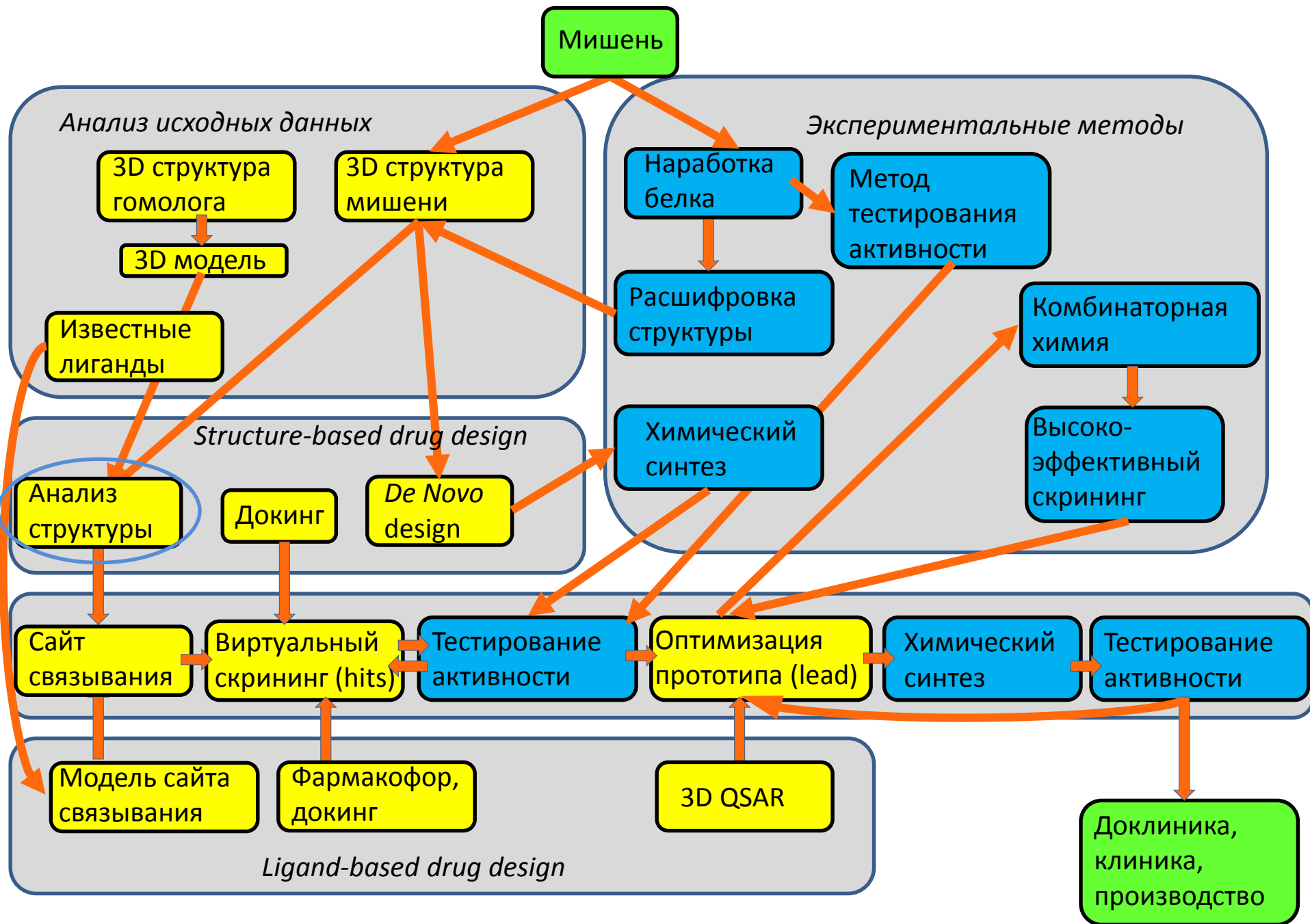




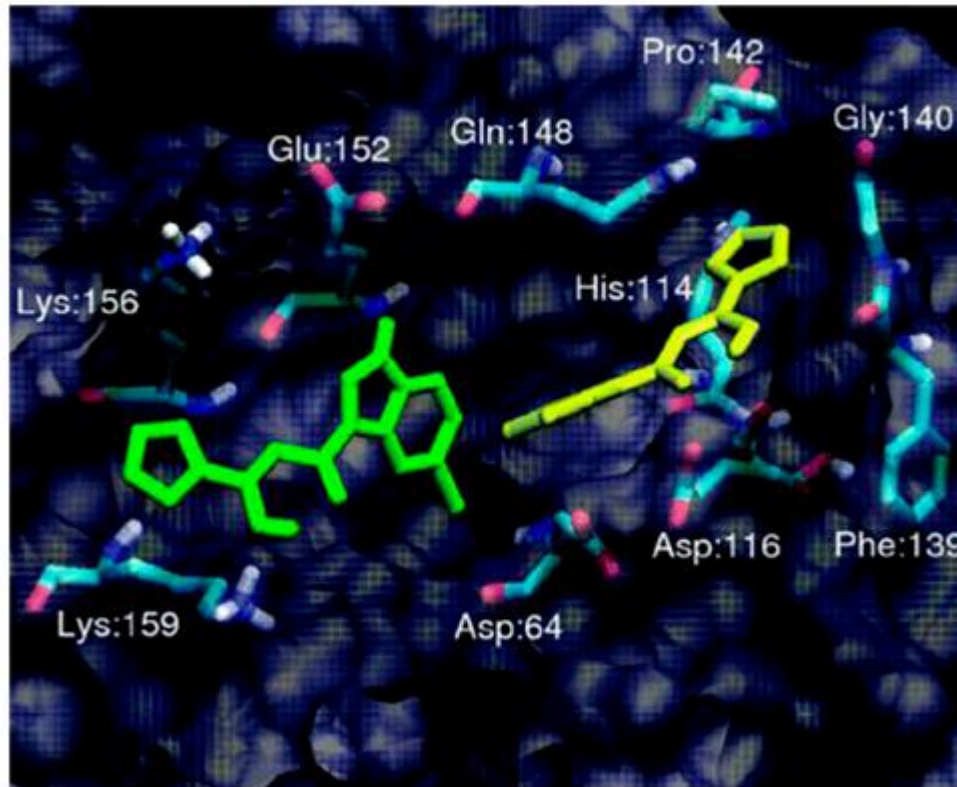
Иматиниб - ингибитор тирозиновой протеинкиназы Bcr-Abl, ген которой находится на «филадельфийской хромосоме» (Ph), образующейся вследствие реципрокной транслокации между 9-й и 22-й хромосомами.

Иматиниб селективно подавляет пролиферацию и вызывает апоптоз клеточных линий, позитивных по Bcr-Abl, а также лейкозных клеток при Ph-положительном хроническом миелолейкозе и остром лимфобластном лейкозе.

# Основные стратегии, используемые при компьютерной разработке лекарств



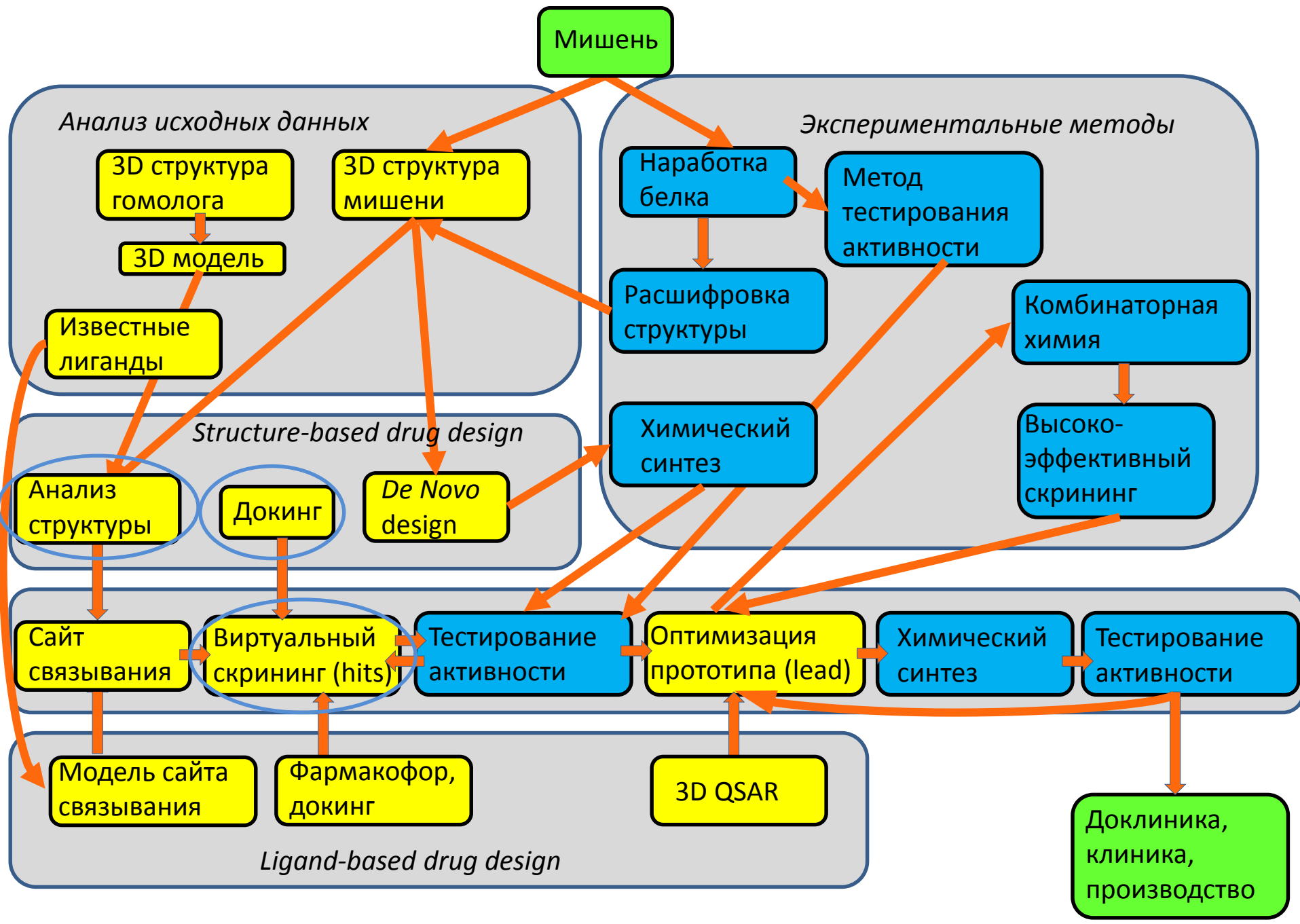
# Анализ структуры белка-мишени и его комплексов



*Пример:*

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 ралтегавира, используемого для лечения СПИДа (Summa et al., 2008).

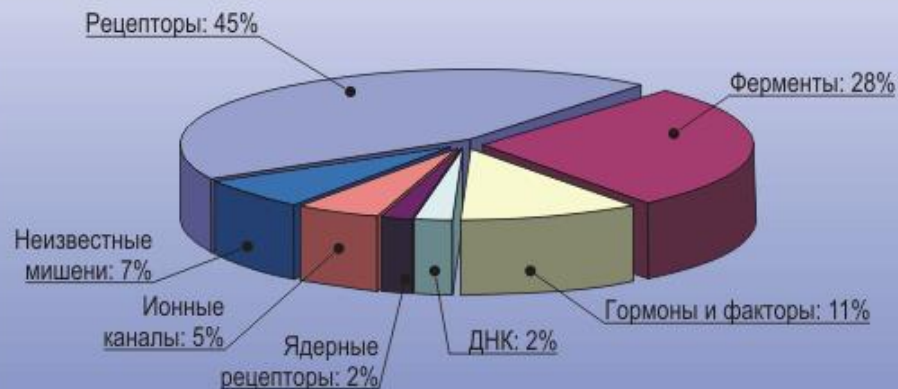
# Основные стратегии, используемые при компьютерной разработке лекарств





# Structure-based drug design: молекулярный докинг

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

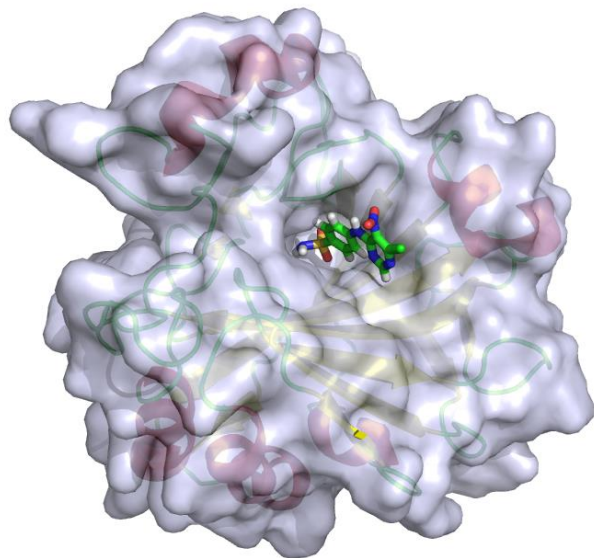


Подавляющее большинство лекарств – низкомолекулярные ингибиторы белков.

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

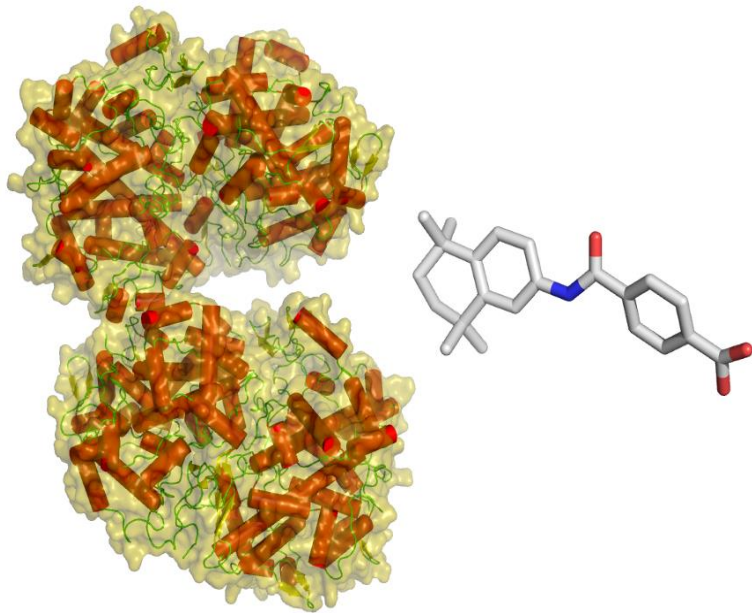
Даже среди гораздо меньшего количества соединений можно найти такое, которое будет связываться с интересующим нас белком.

**Как это сделать?**

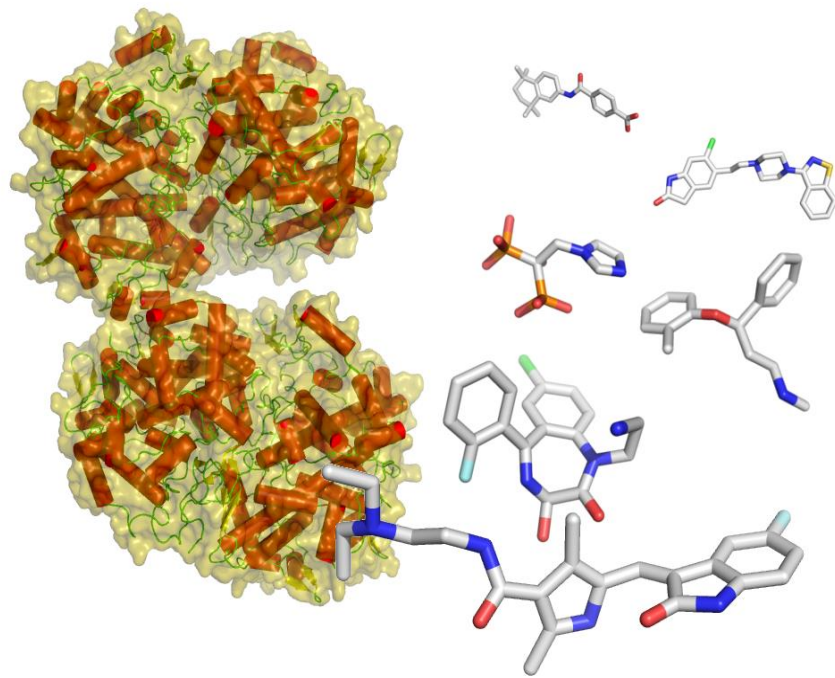


Нужно найти соединение, подходящее к определенному месту белка, например, активному центру, как ключ к замку.





**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.



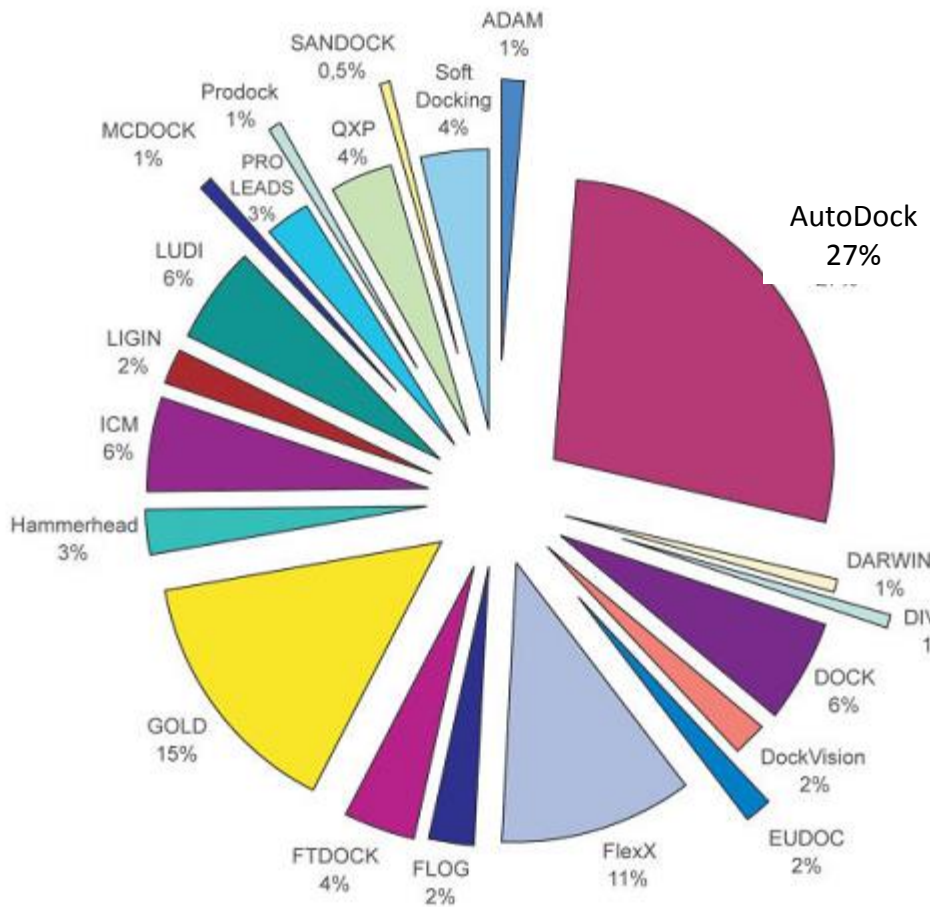
**Виртуальный скрининг** (virtual screening) – последовательный докинг множества различных низкомолекулярных веществ против одного белка с целью отобрать для экспериментальной проверки наиболее перспективные. Позволяет повысить вероятность обнаружения ингибитора интересующего нас белка по сравнению со 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](http://www.tripos.com))

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

Gold ([http://www.ccdc.cam.ac.uk/products/life\\_sciences/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](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>)

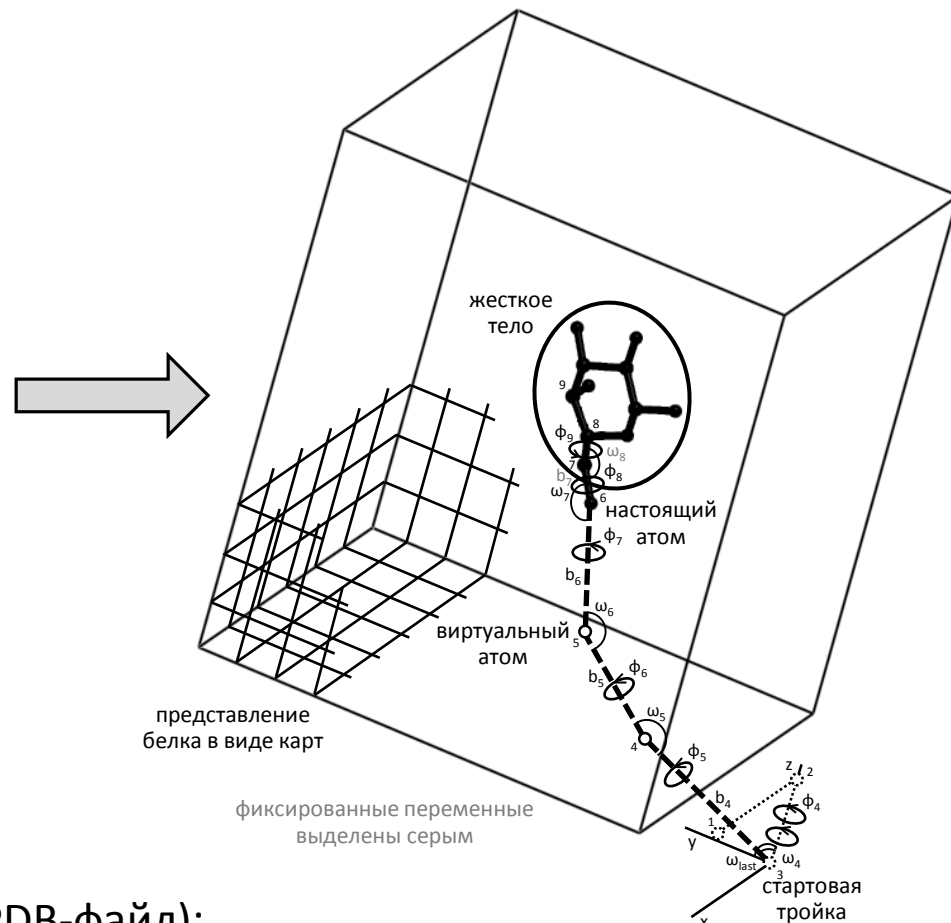
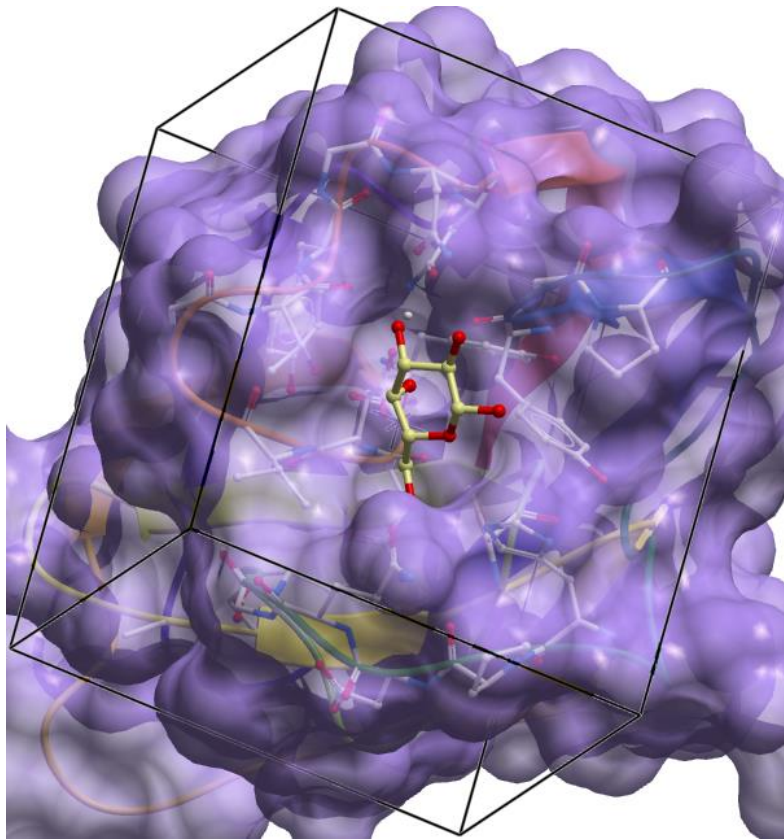
Glide (<http://www.schrodinger.com/ProductDescription.php?mID=6&slD=6&clD=0>)

Программы для виртуального скрининга:

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

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

# ICM: докинг лиганда в белок, подготовительный этап



Подготовка белка (исходные данные – PDB-файл):

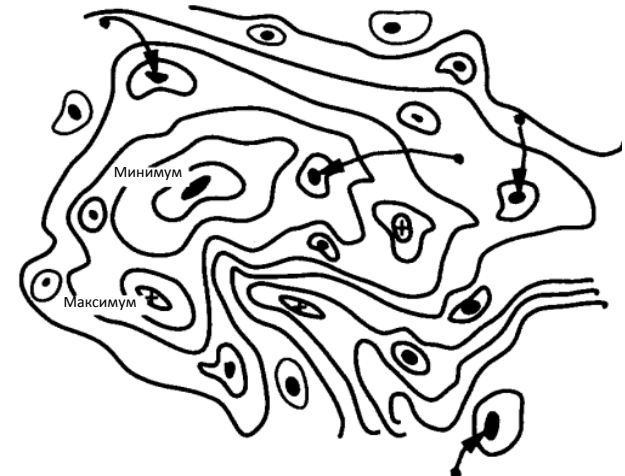
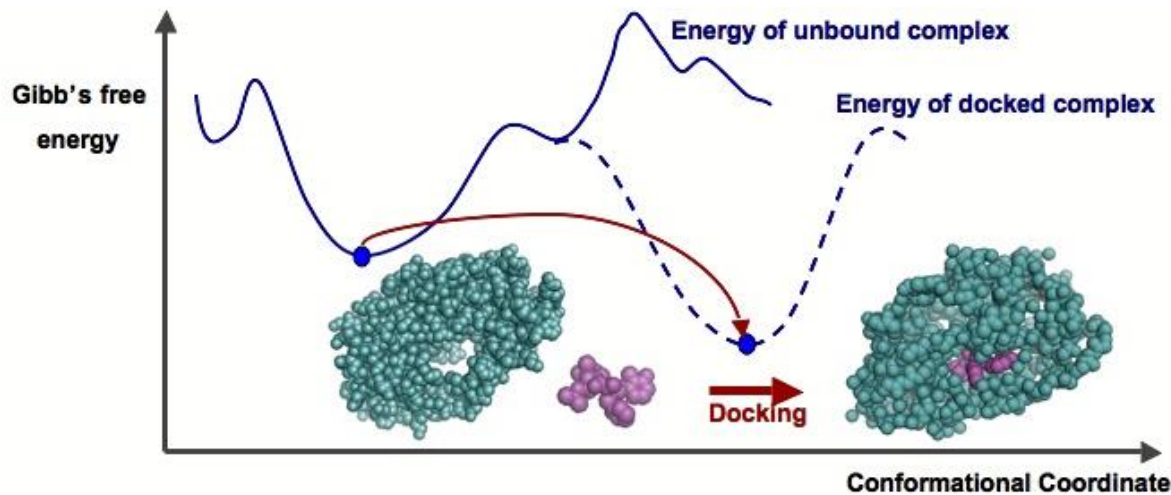
Удаление воды и альтернативных конформаций боковых цепей, добавление атомов водорода, назначение парциальных зарядов, **расчет карт потенциалов нескольких типов** (шаг по умолчанию – 0.5 Å).

Подготовка лиганда (исходные данные – файл в 2D sdf-формате):

Добавление атомов водорода, назначение парциальных зарядов, построение трехмерной структуры, **перевод декартовых координат во внутренние.**

# ISM: докинг лиганда в белок, алгоритм

Необходимо найти стабильную конформацию комплекса и оценить ее свободную энергию ( $E$ ), т.е. найти **глобальный минимум свободной энергии**. Основные вклады в  $E$  известны, но аналитически найти ее глобальный минимум **невозможно**, так как функция имеет огромное количество переменных и локальных минимумов.

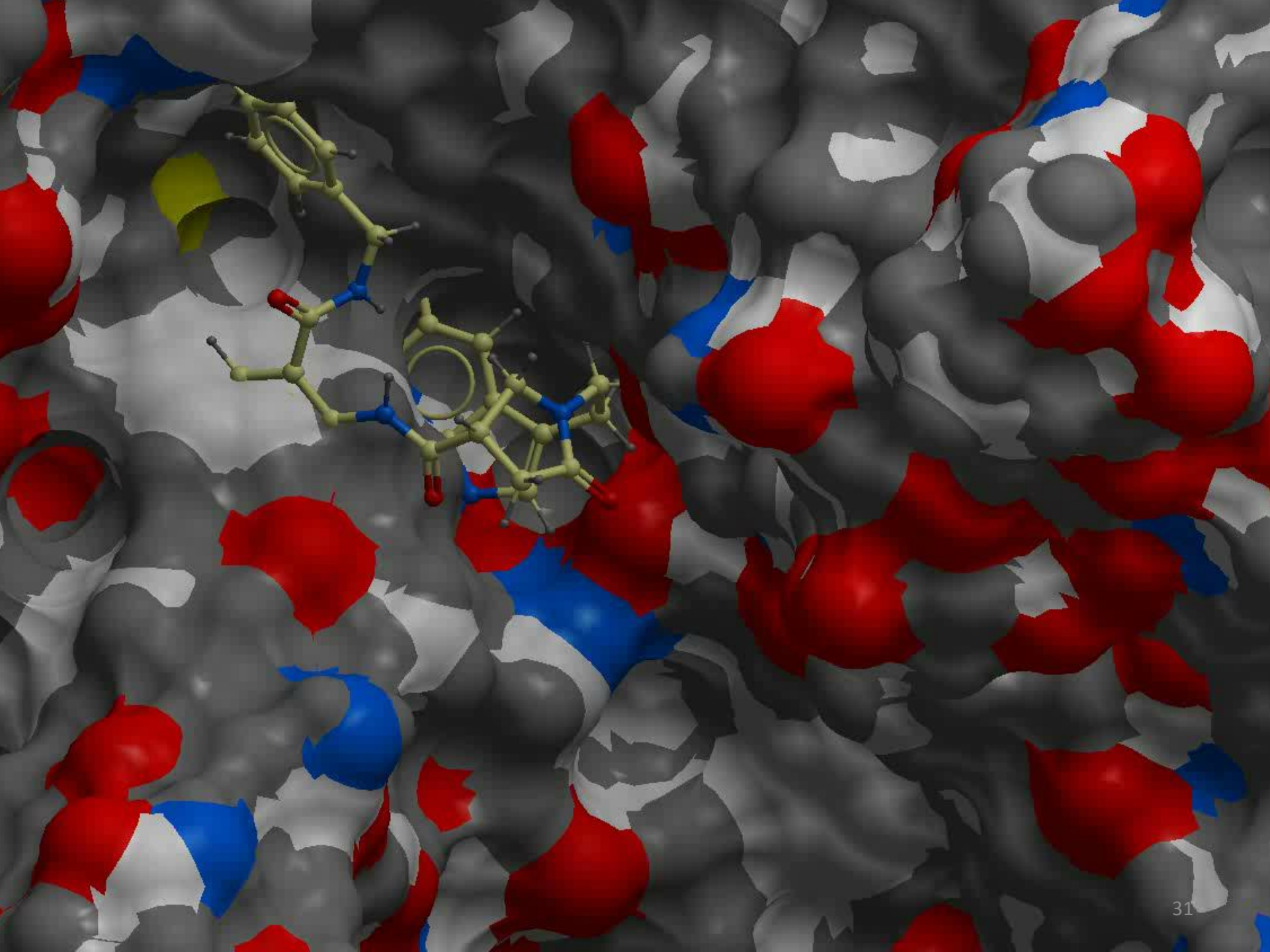


Пример двумерной проекции графика свободной энергии

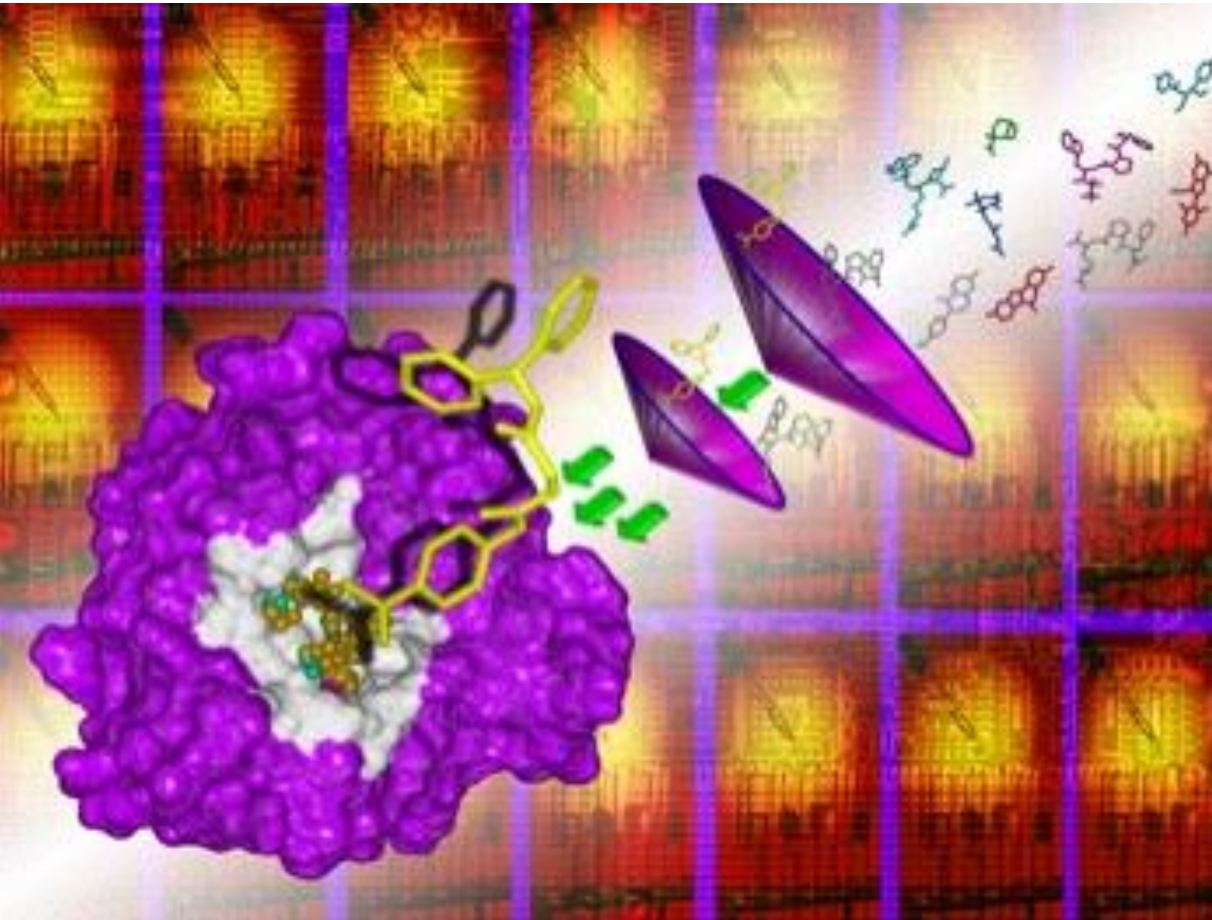
Решение проблемы – стохастический докинг методом Монте-Карло.

Основная идея — случайное изменение положений лиганда и анализ изменений энергии связывания лиганда. На основании изменения энергии осуществляется, например, выбор предпочтительного направления случайных изменений или принятие решения об отказе/принятии изменения.

Альтернативные подходы: генетический алгоритм, использование преобразований Фурье, молекулярная динамика (самый реалистичный и самый медленный – используется для уточнения).



# Подбор низкомолекулярных ингибиторов белков с помощью виртуального скрининга: общая схема



Библиотека химических соединений



Фильтры  
(правило пяти  
Липински)



Виртуальный скрининг

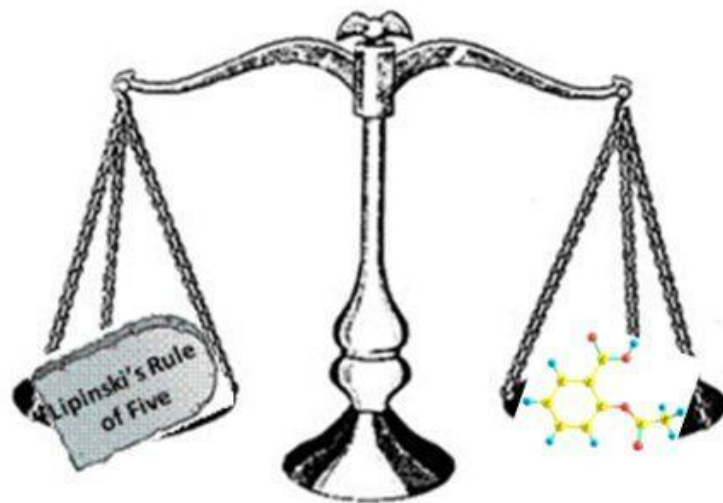


Отбор лучших по  
оценочной функции  
соединений-ингибиторов  
(scoring, hits)



# 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...

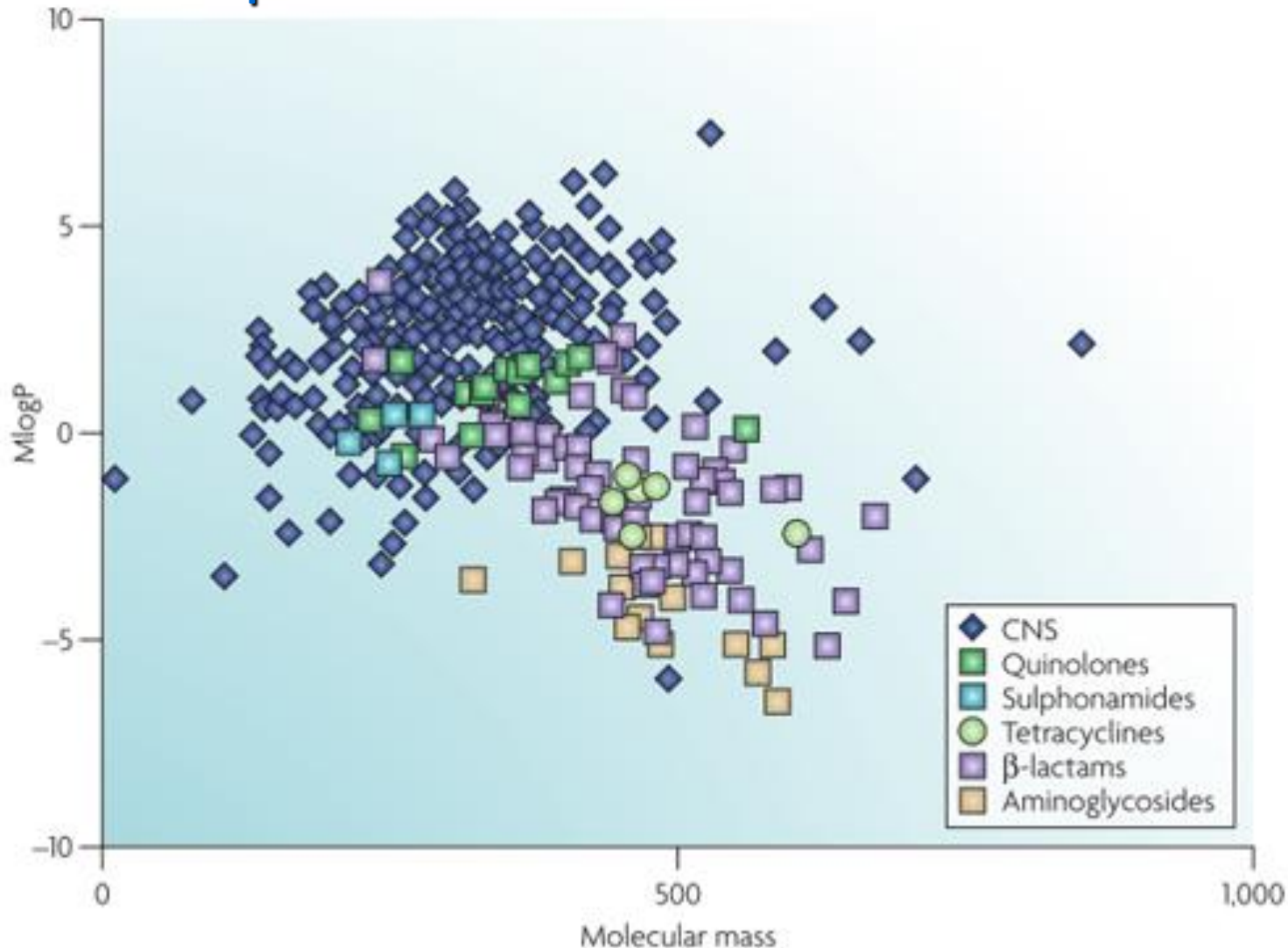
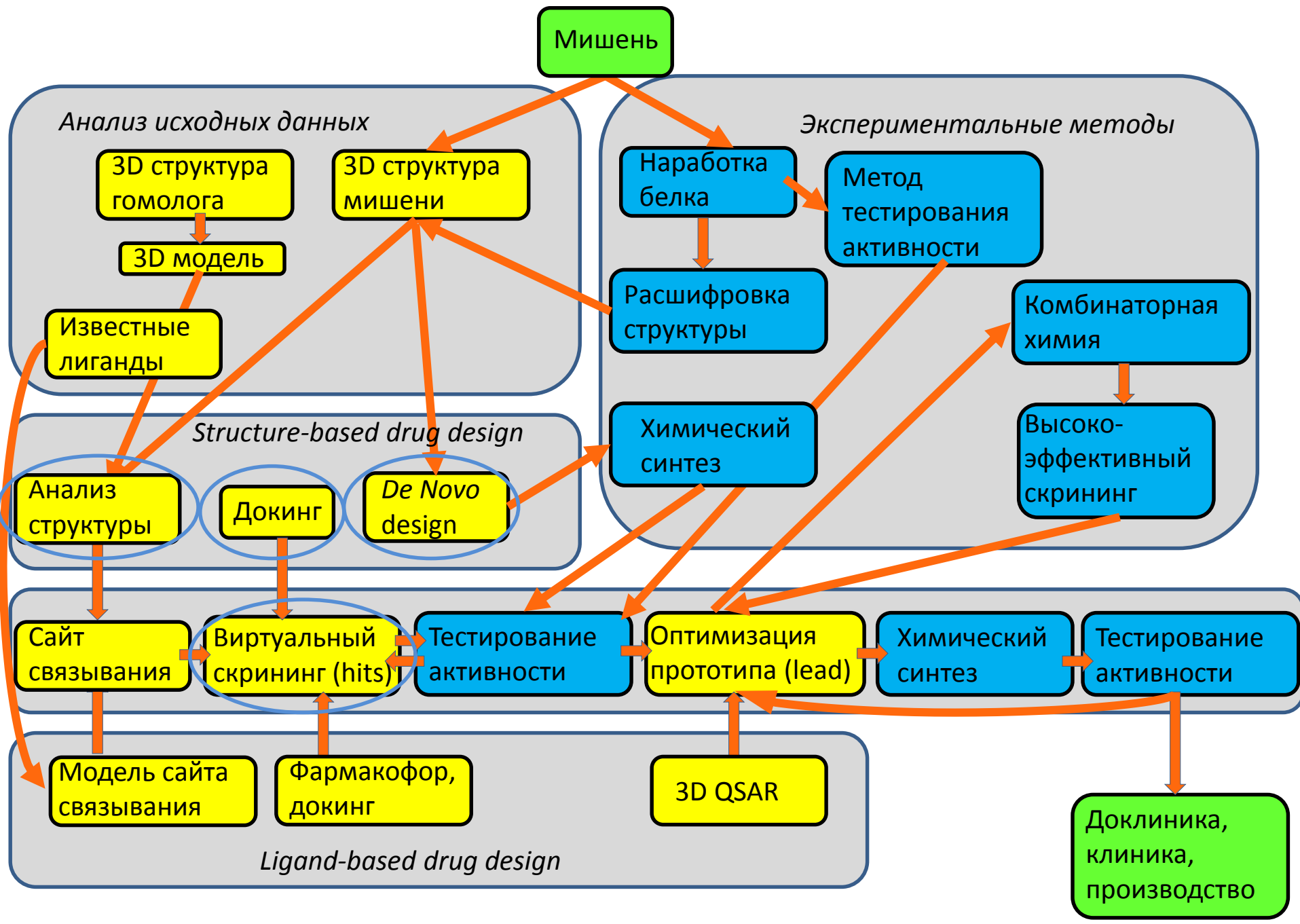
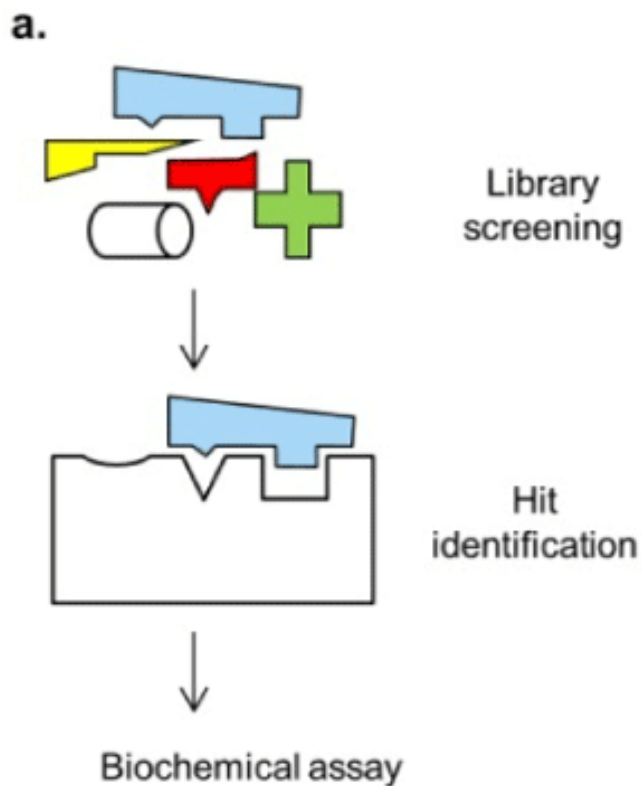


Figure 4 | **The chemical diversity of antibacterials is different to other drugs.** A plot of calculated logP and molecular mass of marketed drugs for central nervous system disorders compared with marketed antibacterial classes. CNS drugs (similar observations for other mammalian target classes; data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

# Основные стратегии, используемые при компьютерной разработке лекарств



# Виртуальный скрининг с помощью молекулярного докинга



# Fragment based drug design (*de novo*)

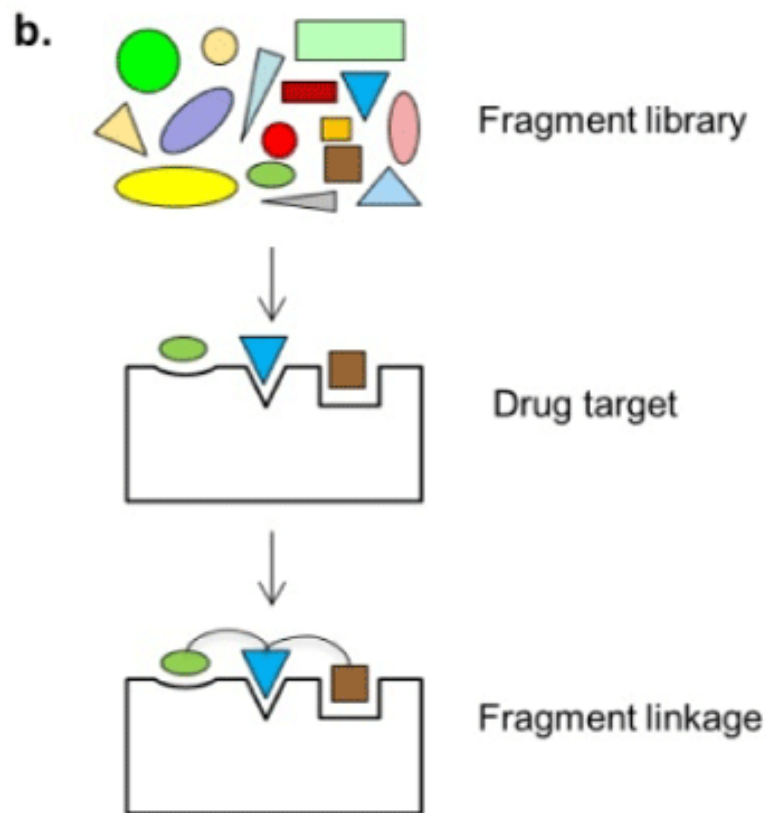
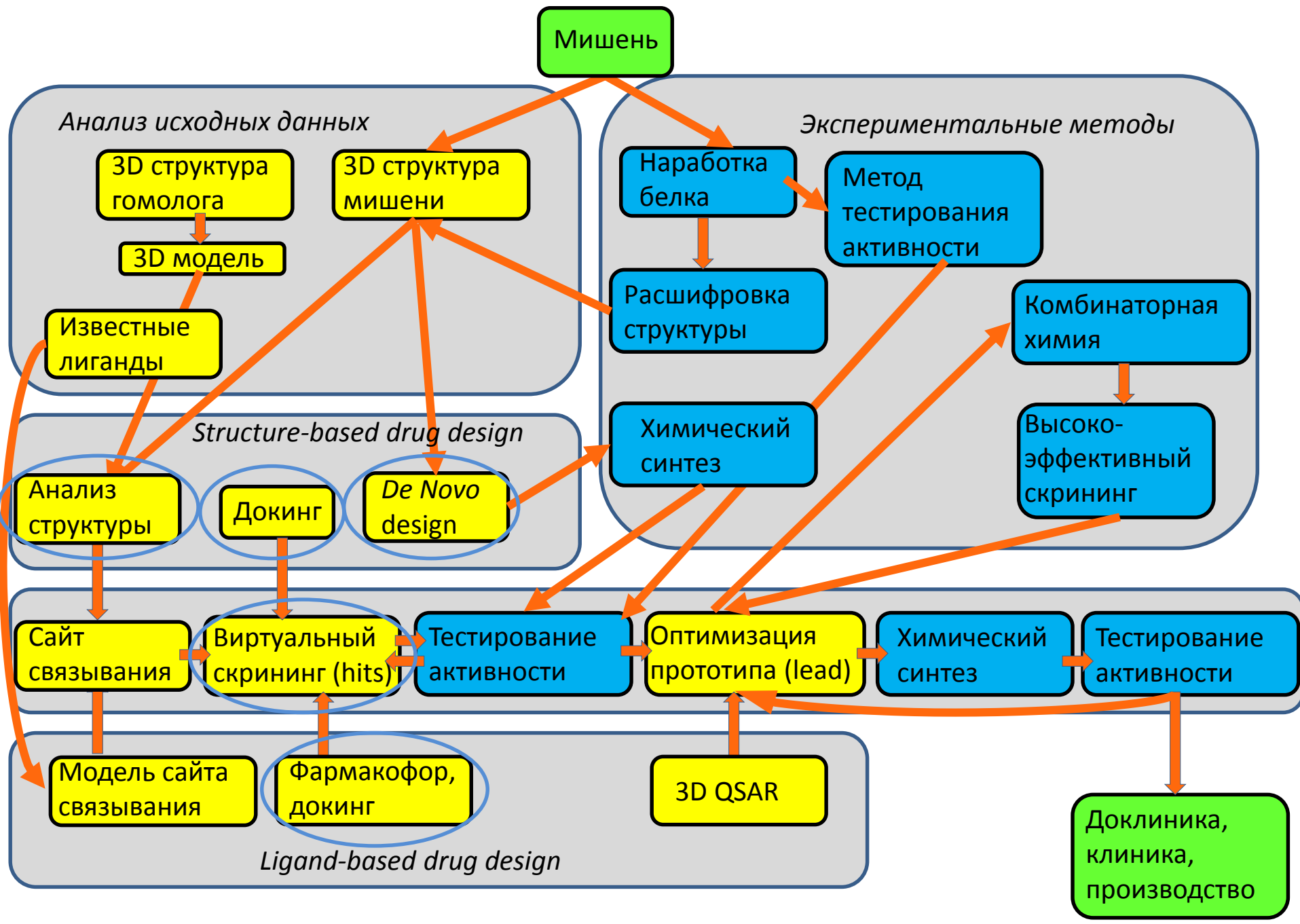
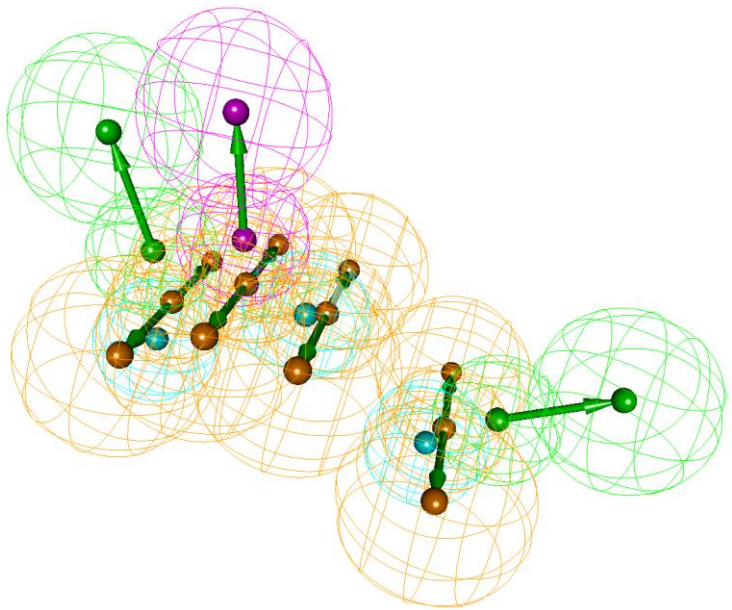


Fig from: Park S, Mann J, Li N (2013) Targeted Inhibitor Design: Lessons from Small Molecule Drug Design, Directed Evolution, and Vaccine Research. Chem Eng Process Tech 1: 1004.

# 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).

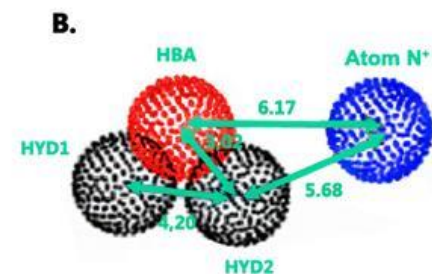
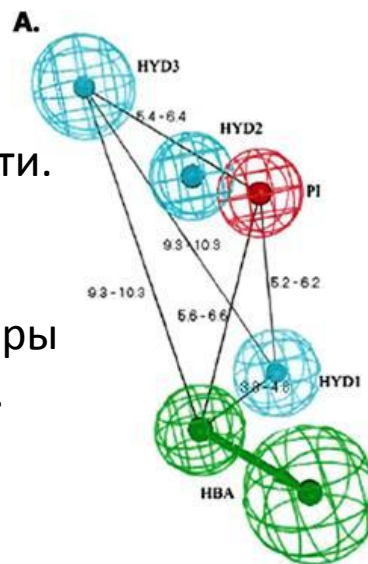
В 1977 г. это определение было модифицировано Питером Гундом (Peter Gund): Фармакофор — это набор структурных признаков в молекуле, которые распознаются биологическими рецепторами и являются ответственными за биологическую активность молекулы (Gund. *Prog. Mol. Subcell. Biol.* 1977, 5: pp 117—143).

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](#)

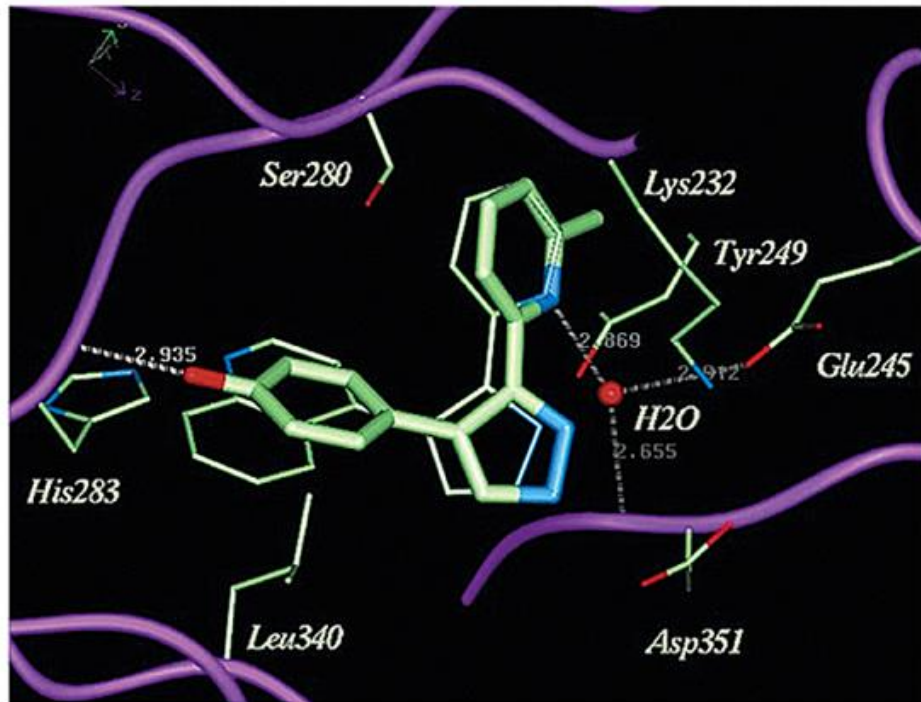
ICM Pro (<http://www.molsoft.com/gui/ph4.html>)

DISCOtech

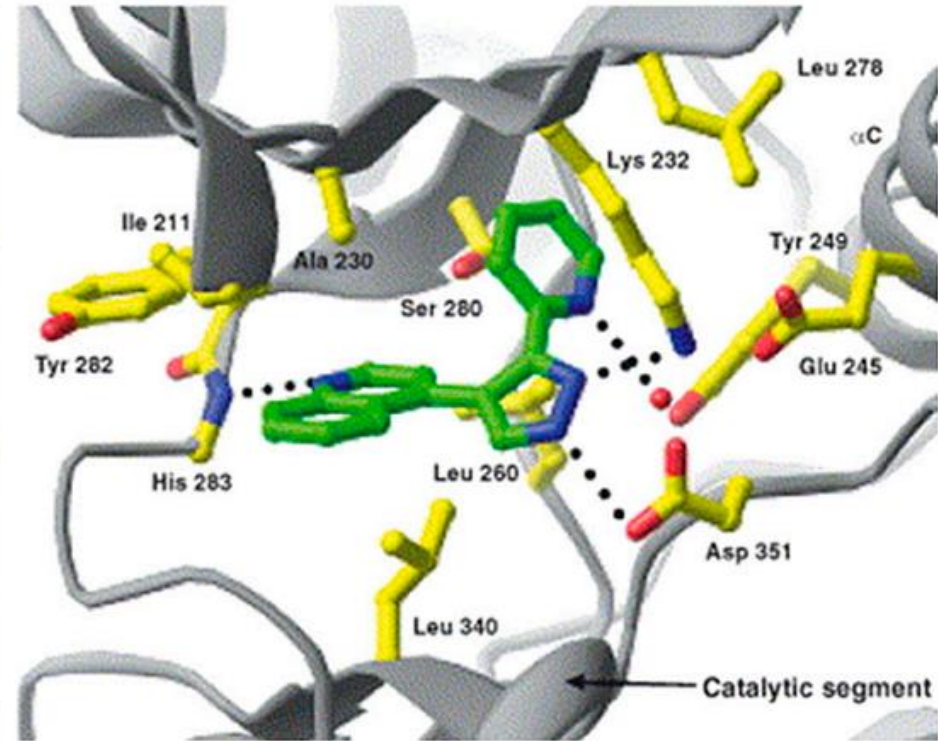
([http://www.tripos.com/index.php?family=modules,SimplePage,,,&page=sybyl\\_discotech](http://www.tripos.com/index.php?family=modules,SimplePage,,,&page=sybyl_discotech))

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

I)



II)



Pharmacol Rev 66:334–395

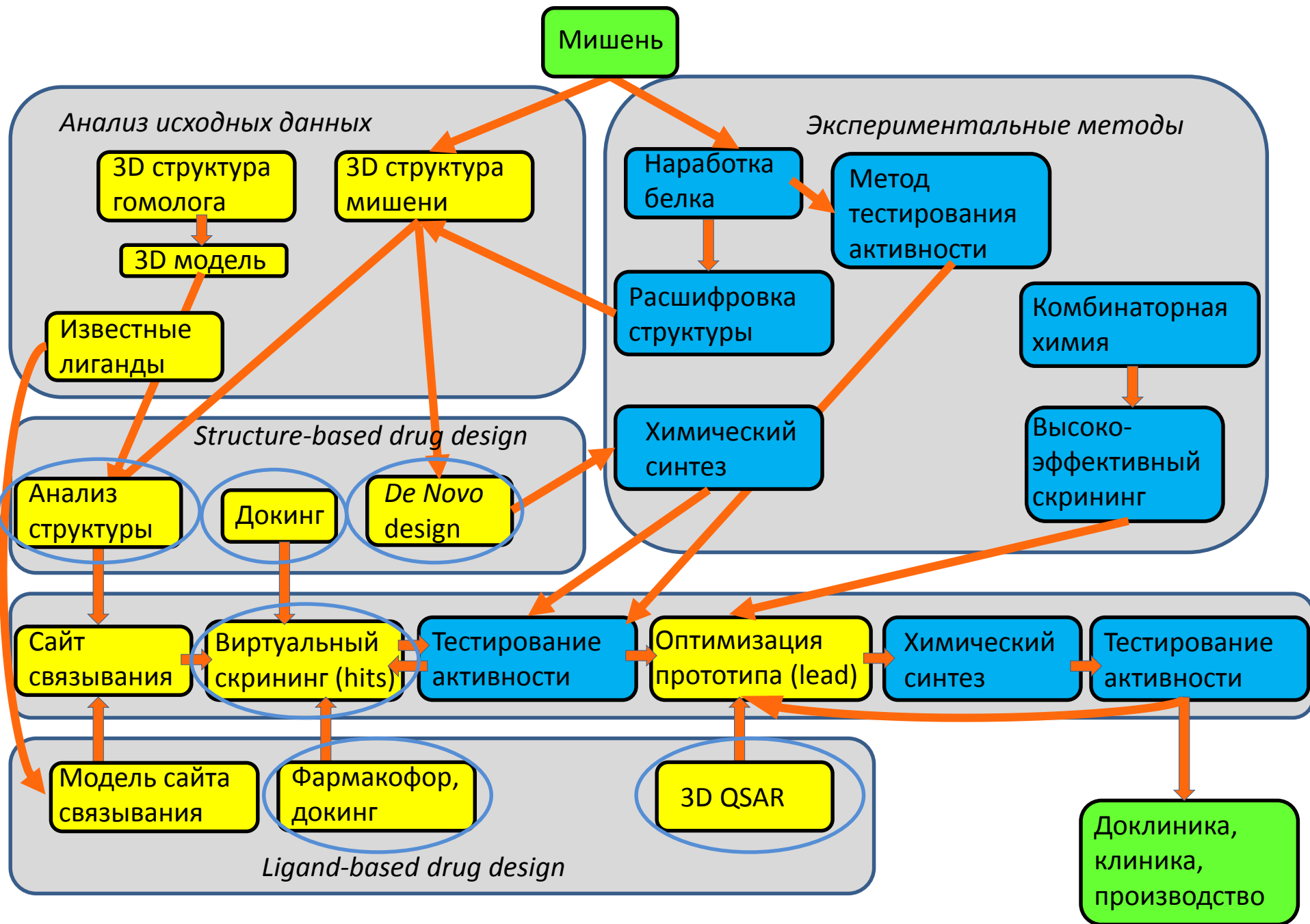
*Пример:* Поиск ингибиторов киназного домена рецептора TGF $\beta$ 1 (Transforming growth factor beta 1).

Компания Eli Lilly использовала традиционный HTS 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 $\beta$ 1R (Singh et al., 2003a). Было отобрано 87 соединений, лучшее из которых (II) совпало по структуре с соединением, отобранным на основании традиционного HTS 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). При прогнозировании свойств на качественном уровне (например, будет ли данное химическое соединение обладать данным видом биологической активности) говорят о решении *классификационной* задачи, тогда как при прогнозировании числовых значений свойств говорят о решении *регрессионной* задачи.

При векторном описании химической структуре ставится в соответствие вектор молекулярных дескрипторов:

- Фрагментные дескрипторы
- Физико-химические дескрипторы
- Квантово-химические дескрипторы
- Дескрипторы молекулярных полей
- Фармакофорные дескрипторы
- Дескрипторы молекулярного подобия

## 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

## Примеры прогнозирования свойств ADMET (Absorption, Distribution, Metabolism, and Excretion)

### Фармакокинетические свойства

- Проникновение через гематоэнцефалический барьер
- Скорость проникновения через кожу

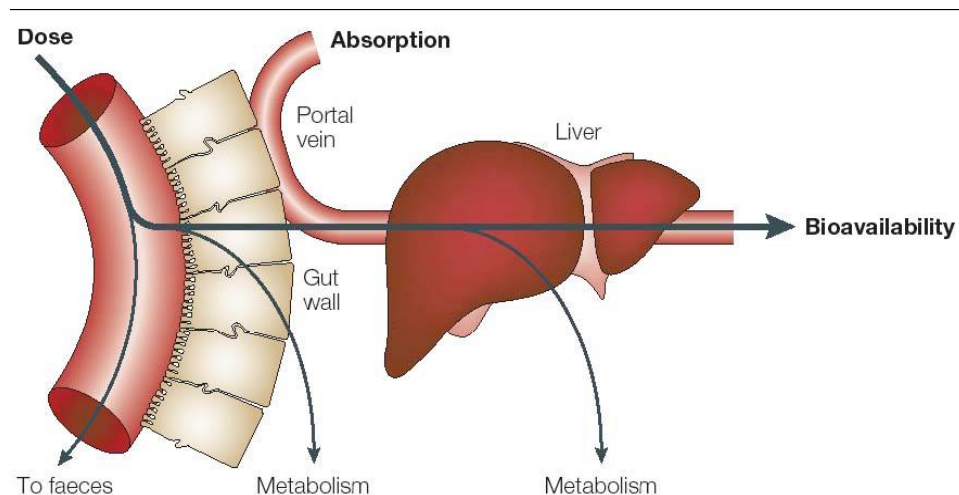
### Метаболизм

- Сайты ароматического гидроксирования при метаболической активации цитохромом P450

## Примеры прогнозирования биологической активности

Спектр биологической активности

Принадлежность к фармакологическим группам

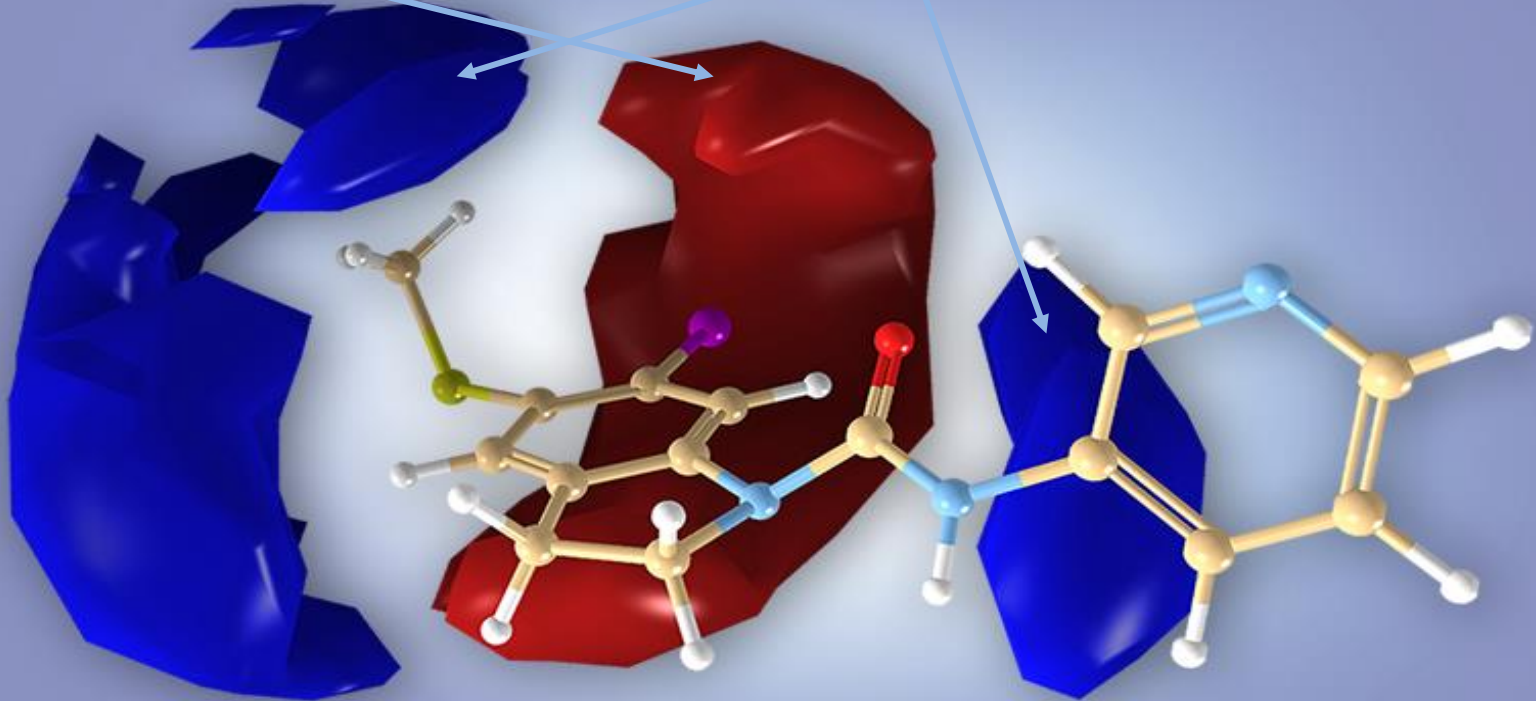


# 3D-QSAR (Quantitative Structure-Activity Relationship) анализ

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

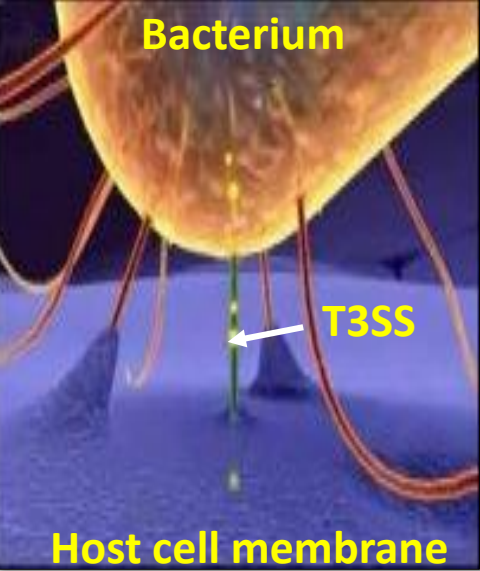
Взаимодействие с синими областями положительно влияет на активность лиганда





*Пример:*

**Подбор ингибиторов АТФазы  
III транспортной системы грам-  
отрицательных бактерий –  
возбудителей инфекционных  
заболеваний**

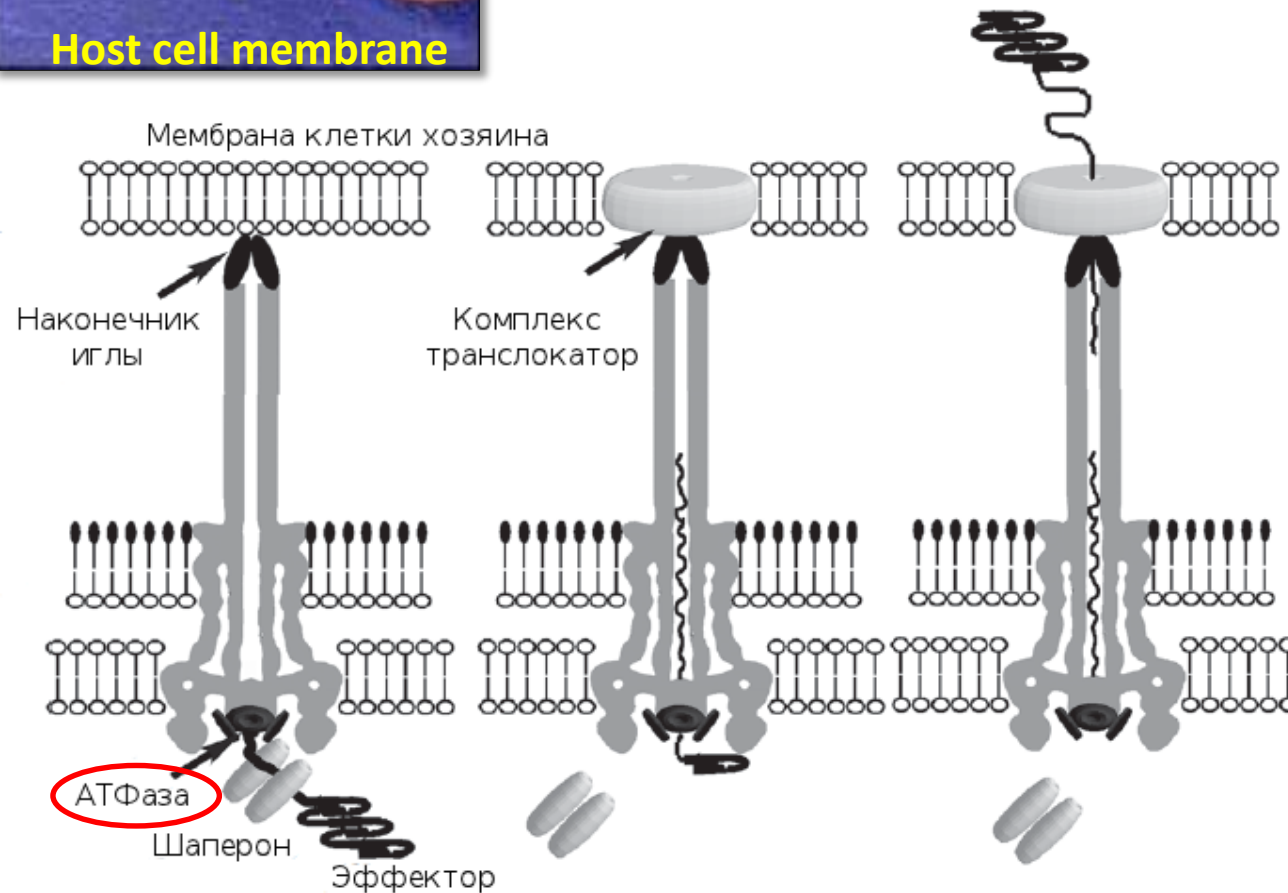


Роль III транспортной системы - перенос факторов вирулентности (белков-эффекторов) из бактериальной клетки непосредственно в цитоплазму клетки-хозяина.

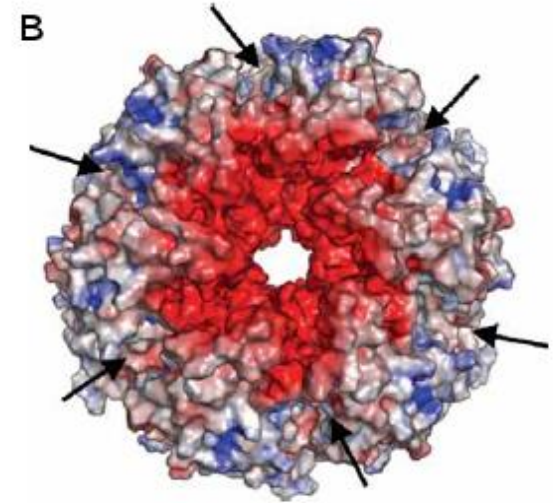
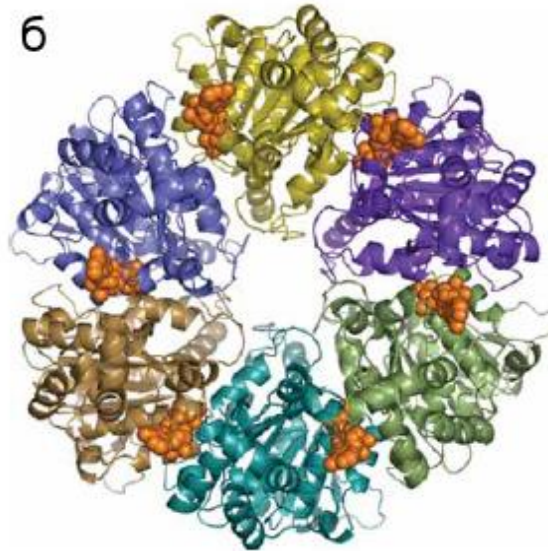
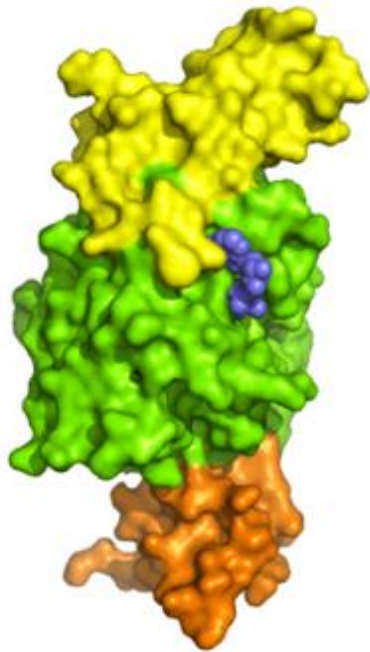
- **АТФаза** снабжает энергией весь комплекс и является ключевым компонентом системы

- Мутанты по гену АТФазы ТТС утрачивали вирулентность

- АТФаза ТТС является привлекательной мишенью для подбора ингибиторов – кандидатов лекарственных средств для борьбы с хламидиозом



# Структура АТФазы ТТС



- Имеет вид гомогексамерного кольца
- Каталитический сайт находится между двумя субъединицами
- Следовательно, для проведения эффективного скрининга необходимо иметь трехмерную модель димера АТФазы ТТС

# Построение димера АТФазы SctN хламидий в программе MODELLER

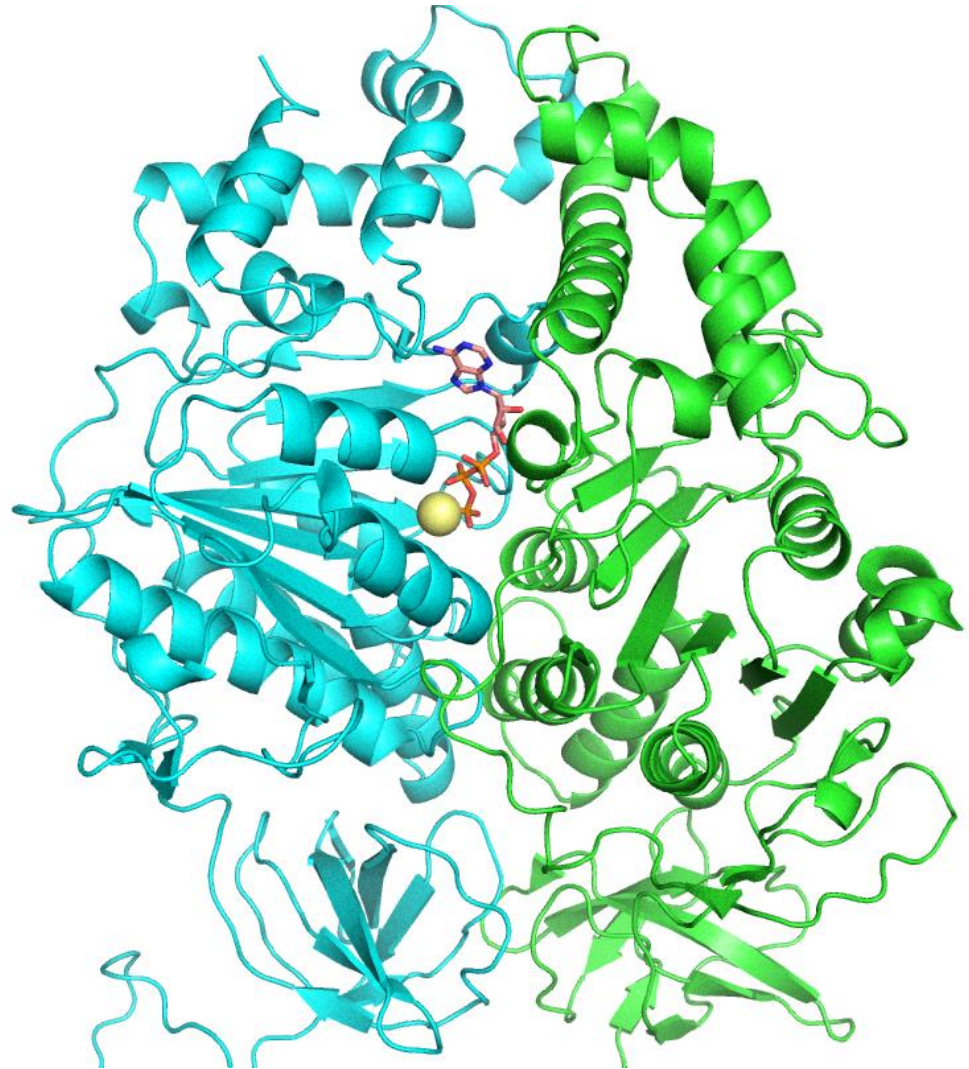
Трёхмерные структуры  
гомологов

+

Выравнивание аа  
последовательностей

=

Трёхмерная модель  
димера АТФазы  
хламидий:





# Виртуальный скрининг

- Программа для проведения виртуального скрининга MolSoft ICM
- Библиотека низкомолекулярных химических соединений (220 000)
- ~7000 отобранных кандидатов со скором лучше -20.

# Структуры 16 отобранных для дальнейшей работы соединений:

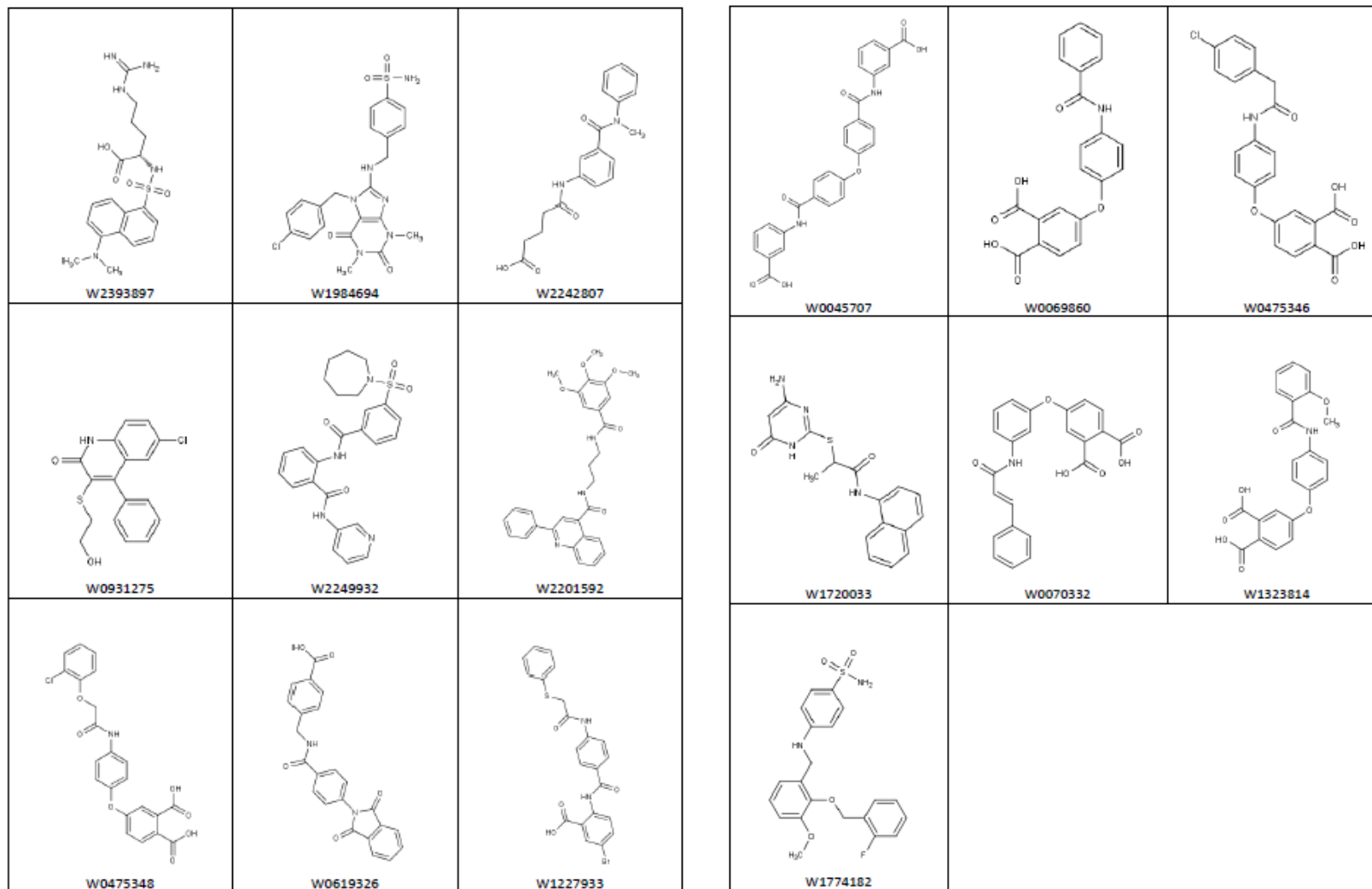
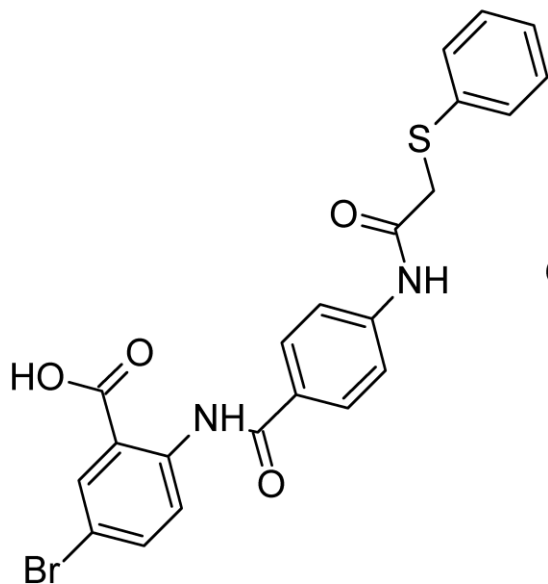
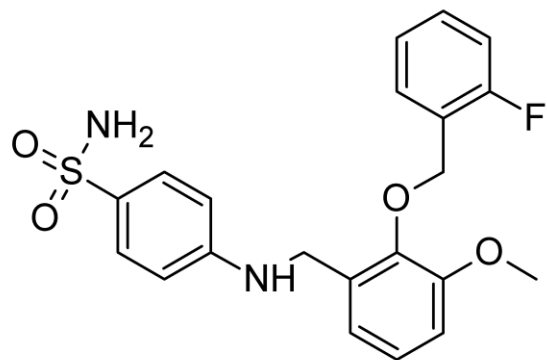


Table S2. Anti-chlamydial activity of experimentally tested compounds

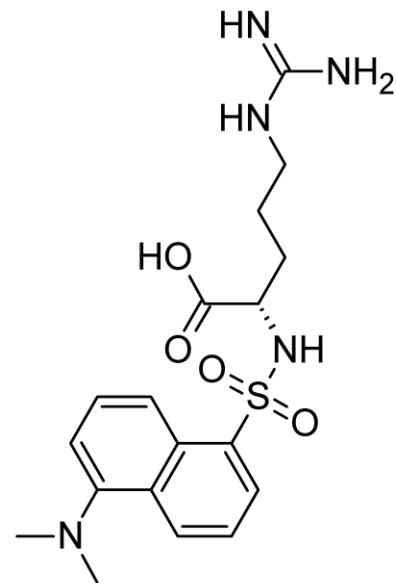
compound	reduction in the number of chlamydial inclusions, % to control		
	50 $\mu$ M	75 $\mu$ M	100 $\mu$ M
W2393897	20	40	50
W1984694	0	0	10
W2242807	0	0	0
W1720033	0	0	0
W0931275	0	0	10
W2249932	0	0	0
W2201592	insoluble		
W0070332	0	5	30
W0475348	0	15	25
W0619326	0	5	5
W1227933	70	97	100
W1774182	90	sporadic inclusions	100
W0045707	0	0	0
W0069860	0	0	0
W0475346	0	0	15
W1323814	0	0	0



W1227933

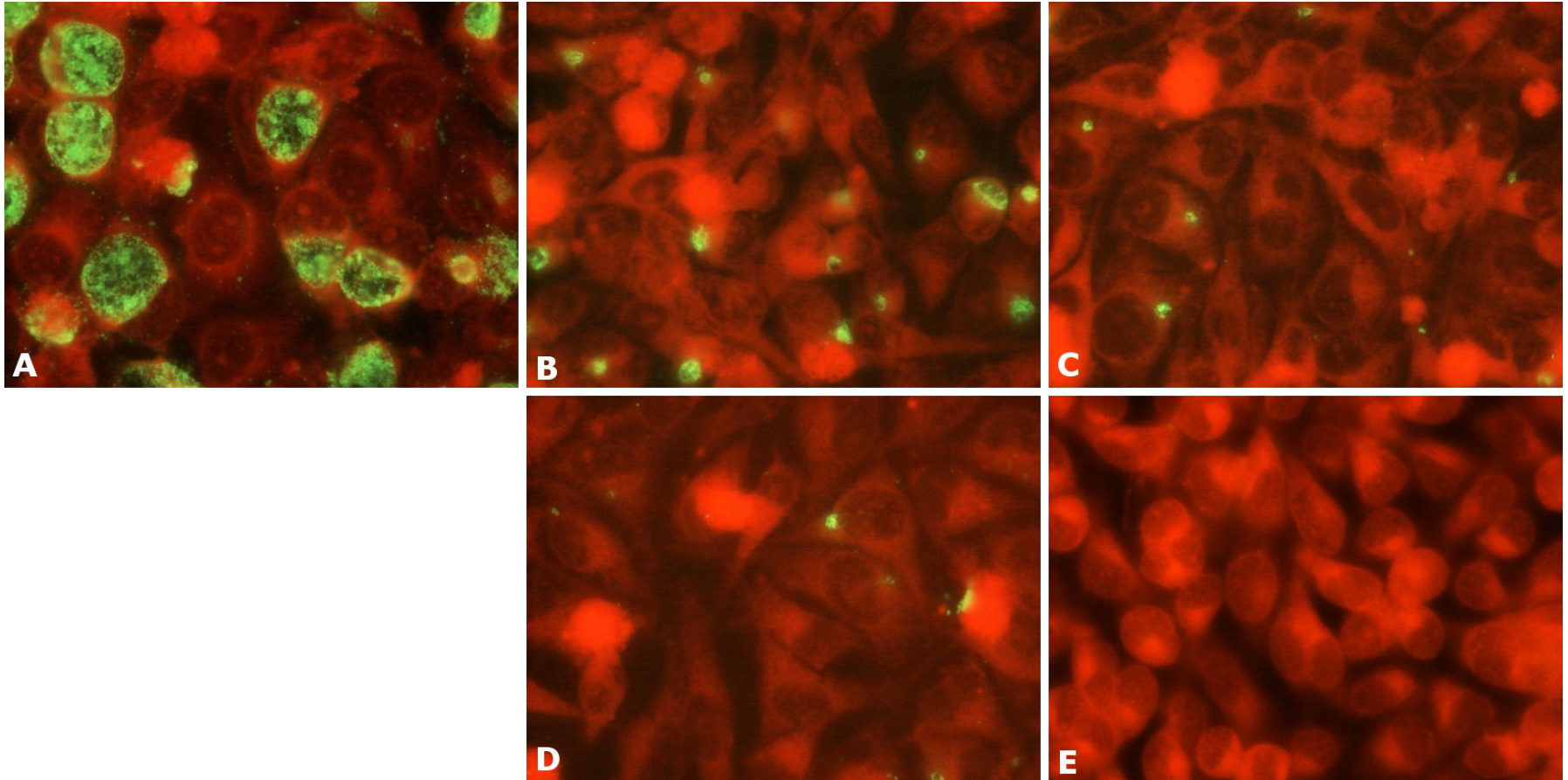


W1774182



W2393897

**Chemical structures of three compounds that fully (W1227933 and W1774182) or partially (W2393897) suppress chlamydial development within eukaryotic cells.**

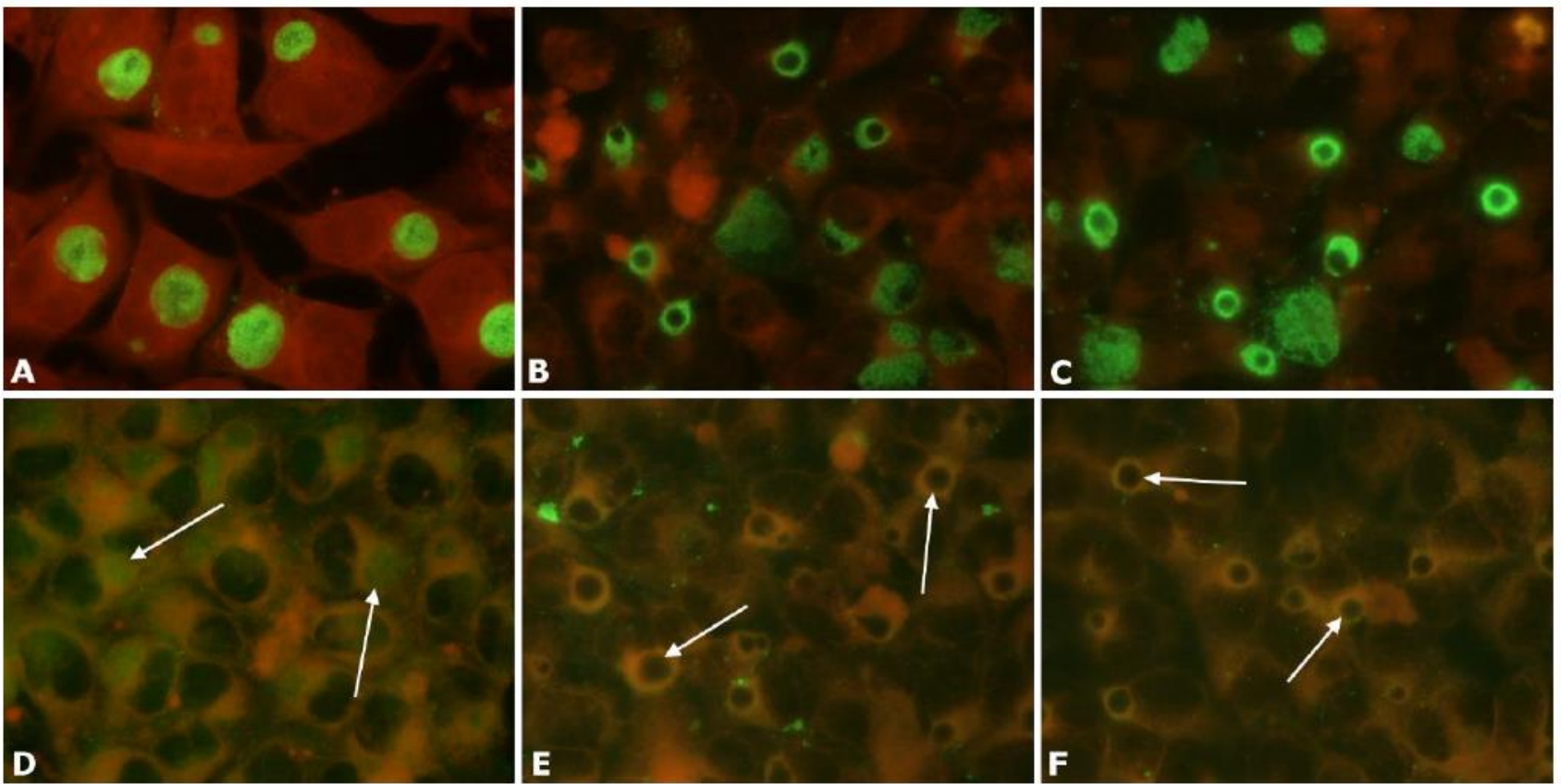


**Effect of compounds W1227933 and W1774182 on chlamydial survival and development in eukaryotic cells. McCoy cells are stained red with bromophenol blue, chlamydial inclusions are stained green using FITC-conjugated anti-MOMP antibodies.**

**A – control (no inhibitor);**

**B, C – compound W1227933 added at 50 (B) and 75 (C)  $\mu$ M;**

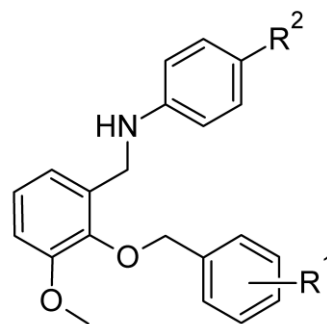
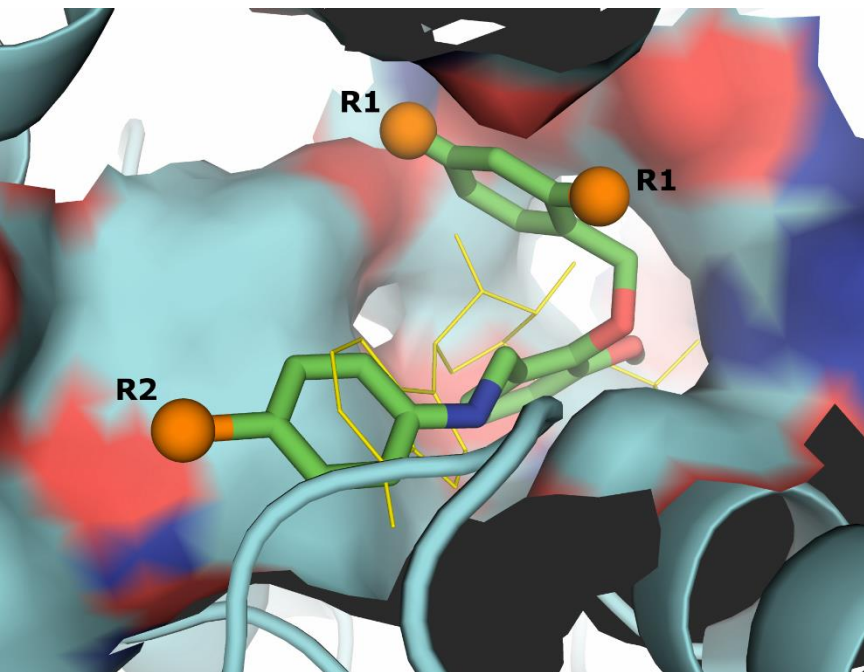
**D, E – compound W1774182 added at 50 (D) and 75 (E)  $\mu$ M.**



**Effect of compounds W1227933 and W1774182 on IncA translocation by T3SS.**

**A – control (no inhibitor) chlamydia-infected McCoy cells, chlamydial inclusions are stained green with FITC-conjugated anti-MOMP antibodies; B, C – same with compounds W1227933 (B) or W1774182 (C) at 25  $\mu$ M; D - control (no inhibitor) chlamydia-infected McCoy cells, chlamydial inclusions (two example inclusions are marked with arrows) are stained with primary polyclonal anti-IncA antibodies and FITC-labeled secondary antibodies, E, F – same with compounds W1227933 (B) or W1774182 (C) at 25  $\mu$ M. In F and E chlamydial inclusions can be seen as dark circles, but no IncA-specific staining is visible (example inclusions are marked with arrows).**

Дальнейшая разработка антихламидийных препаратов связана с синтезом производных активных соединений и характеристикой их свойств



25-33

25:  $R^1 = H$ ,  $R^2 = SO_2NH_2$ ;

26:  $R^1 = 4-F$ ,  $R^2 = SO_2NH_2$ ;

29:  $R^1 = 2-F$ ,  $R^2 = COOCH_3$ ;

30:  $R^1 = 2-F$ ,  $R^2 = 4-CN$ ;

31:  $R^1 = 2-F$ ,  $R^2 =$

32:  $R^1 = 2-F$ ,  $R^2 =$

33:  $R^1 = 4-Cl$ ,  $R^2 = SO_2NH_2$

Docked pose of W1774182 in SctN ATP-binding site. Base structure of W1774182 is shown as sticks with green carbon atoms; places where R1 and R2 substituents were introduced are shown as orange spheres. AMP-PNP in native conformation is shown as yellow thin sticks.