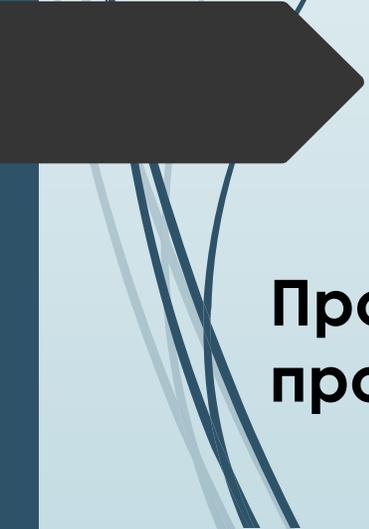


# «МЕТОДЫ, ИСПОЛЬЗУЮЩИЕСЯ БИОИНФОРМАТИКОЙ ДЛЯ АНАЛИЗА МАКРОМОЛЕКУЛ И СОЗДАНИЯ ЛЕКАРСТВ»



Проверил:  
проф. Сафонов М.А.

*Выполнили студентки:*

*Андреева Н.А.*

*Горбова Д.А.*

*группа 503(2)-Б ОЗО*

**Биоинформатика** — совокупность методов и подходов<sup>[1]</sup>, включающих в себя:

- математические методы компьютерного анализа в [сравнительной геномике](#) (геномная биоинформатика).
- разработку алгоритмов и программ для предсказания пространственной структуры биополимеров ([структурная биоинформатика](#)).
- исследование стратегий, соответствующих вычислительных методологий, а также общее управление информационной сложности биологических систем<sup>[2]</sup>.

В биоинформатике используются методы [прикладной математики](#), [статистики](#) и [информатики](#). Биоинформатика используется в [биохимии](#), [биофизике](#), [экологии](#) и в других областях.



Соотношение  
этапов  
развития  
биоинформатики

	Технология	Биоинформатика
1962		Молекулярные часы
1965	Секвенирование tRNA	База данных PIR
1970	Обратная транскрипция	Алгоритм выравнивания NW
1972	Клонирование	
1977	Секвенирование	База данных PDB
1980		Спец. выпуск NAR, Базы данных нукл. Послед.
1981		Алгоритм выравнивания SW
1982	Секвенирование ДНК фага лямбда	
1983	PCR	Алгоритм поиска по базе данных WL
1985	Секвенирование ДНК вирусов	FASTA - поиск по базе данных
1987		GeneBank , Профили
1989	Программа "Геном человека"	Swiss-Prot , NCBI
1991	EST	
1992	Первая хромосома дрожжей	BLOSSUM
1993	Автоматическое секвенирование	
1995	Первый геном бактерии	База данных SCOP
1996	Первый геном архейный	
1997		PSI-BLAST, Кластеры ортологичных генов
1998	Геном червя	
2001	Геном человека	



# Анализ геномов – что можно извлечь из генетических текстов

К настоящему времени полностью расшифрованы геномы около 30 биологических видов. В ближайшие годы ожидается завершение работ по анализу геномов еще несколько десятков видов, среди них – геномы ряда патогенных микроорганизмов; микроорганизмов, находящих применение в биотехнологии; геномов млекопитающих, в том числе – человека.

Примеры организаций, которые занимаются расшифровкой геномов:

- The Sanger Centre, Wellcome Trust, Великобритания;
- The Institute of Genomic Research, США

Хранением и систематизацией медико-биологической и биотехнологической информации:

- National Center for Biotechnology Information, США

Компьютерным анализом, а также активно использующих такую информацию в прикладных и фундаментальных исследованиях:

- Institut Pasteur, Франция.



**BLAST** находит области сходства между биологическими последовательностями. Программа сравнивает нуклеотидные или белковые последовательности в базы данных последовательностей и вычисляет статистическую значимость.



В биоинформатике **FASTA-формат** представляет собой текстовый формат для нуклеотидных или полипептидных последовательностей, в котором нуклеотиды или аминокислоты обозначаются при помощи однобуквенных кодов. Данный формат может содержать названия последовательностей и сопутствующие комментарии.

Простота FASTA-формата позволяет легко производить различные действия с последовательностями при помощи инструментов редактирования текста и скриптовых языков программирования, таких как Python, Ruby, Perl.

#### Format of a FASTA definition line

*space*  
>Seq1 [organism=Carpodacus mexicanus] [*space*clone=6b] [*space*actin (act) mRNA, partial CDS  
CCTTTATCTAATCTTTGGAGCAYGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATC

*hard return*

В 1999 году была опубликована первая работа, описывающая попытку выбора мишеней для действия лекарственных средств на основании сравнительного анализа генетической информации.



Программа **CATS** предназначена для анализа геномов с целью поиска белков, которые могли бы рассматриваться как наиболее предпочтительные мишени для действия лекарственных веществ.





## Три группы методов:

1. распознавание фолда (укладки, упаковки) с использованием библиотеки известных фолдов;
2. предсказания *abinitio* на основе знаний об атомных взаимодействиях и архитектуре белковой глобулы;
3. моделирование по гомологии.



Распознавание фолда – это стадия для построения модели трехмерной структуры белка. Оно применяется, если отсутствует информация о близких гомологах исследуемого белка, пространственная структура которых расшифрована ранее. Хотя при этом удастся предсказать корректно укладку для ~75% белков, "разрешение" построенной таким образом модели не достаточно, чтобы использовать ее в дальнейших исследованиях как базовую для выявления механизма функционирования макромолекул.



При предсказании *abinitio* целью является построение модели 3D структуры без использования знаний по структуре гомологов. Эти методы близки к методам предсказания фолда как по точности распознавания, так и по "разрешению".

Предсказание трёхмерной структуры белка по известной аминокислотной последовательности осуществляется наиболее успешно, когда известна пространственная структура одного или нескольких его гомологов.

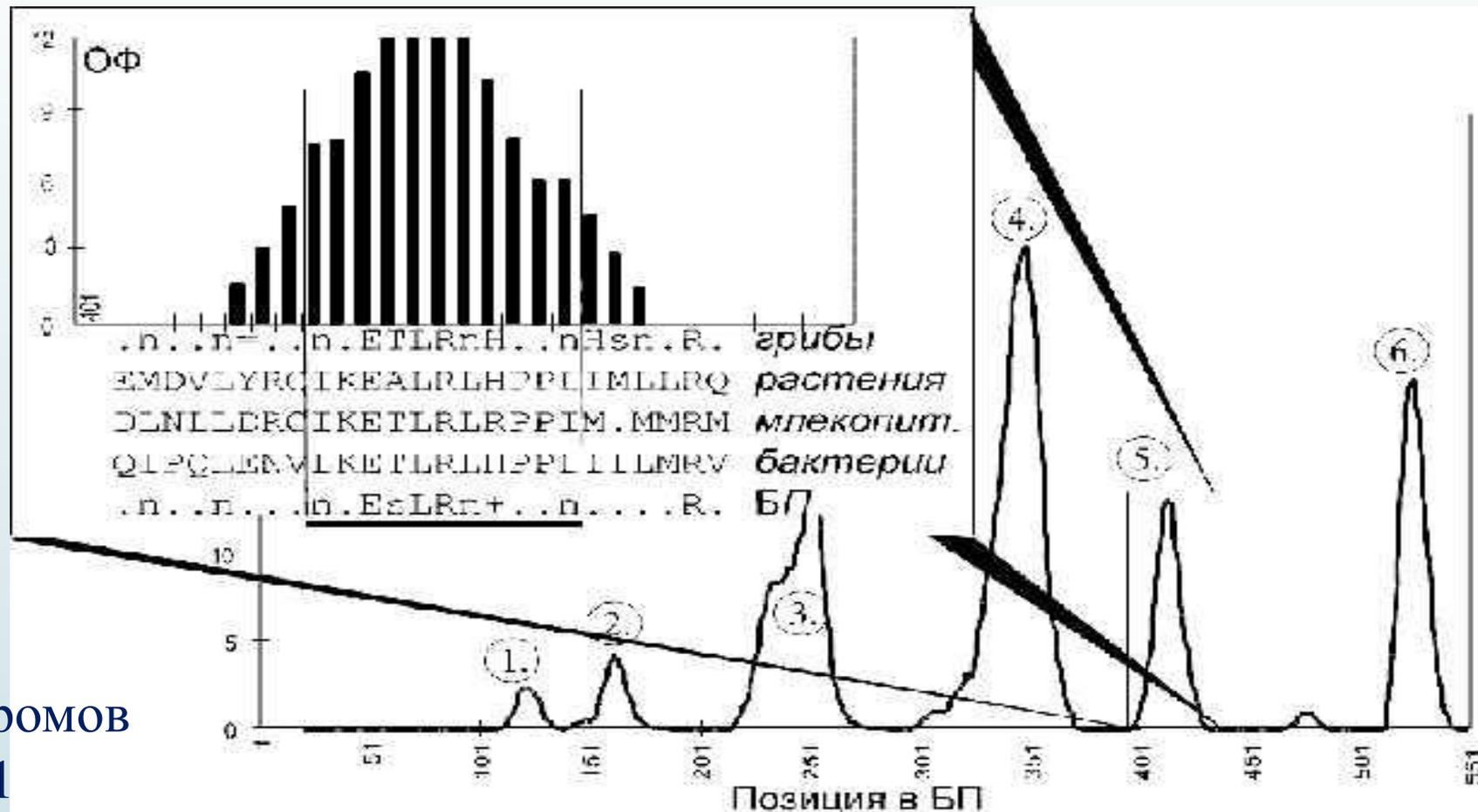
A dark grey arrow points to the right from the left edge of the slide. Below it, several thin, curved lines in shades of blue and grey sweep across the left side of the slide.

В настоящее время разработано достаточно большое число различных подходов к сравнительному моделированию. Одним из наиболее широко используемых является метод, первоначально разработанный Бланделом и реализованный в программе COMPOSER комплекса молекулярного моделирования SYBYL (TRIPOS, Inc.).

Сравнительная оценка различных подходов к предсказанию пространственной структуры белка по аминокислотной последовательности традиционно проводится в Асиломаре (Калифорния, США).

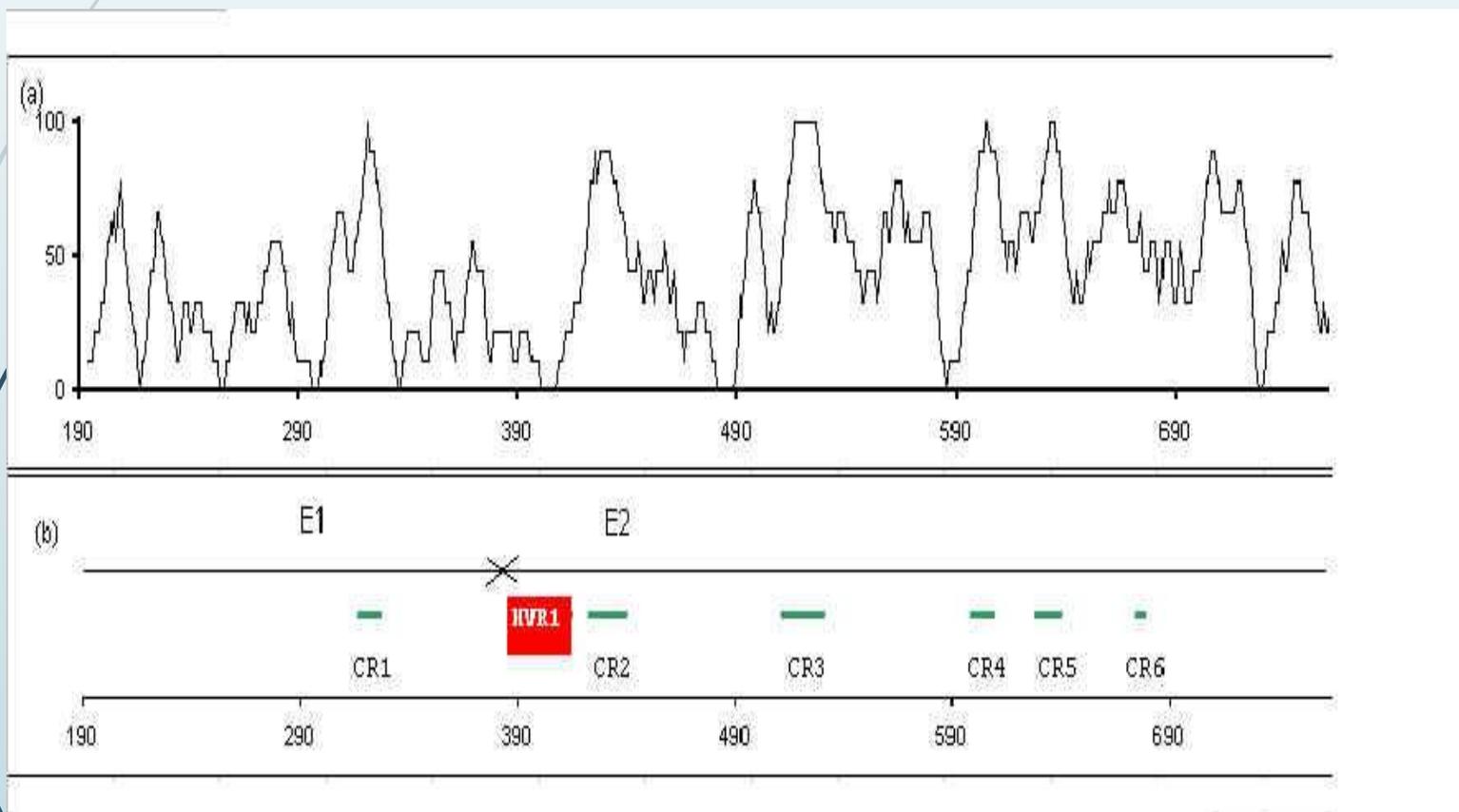


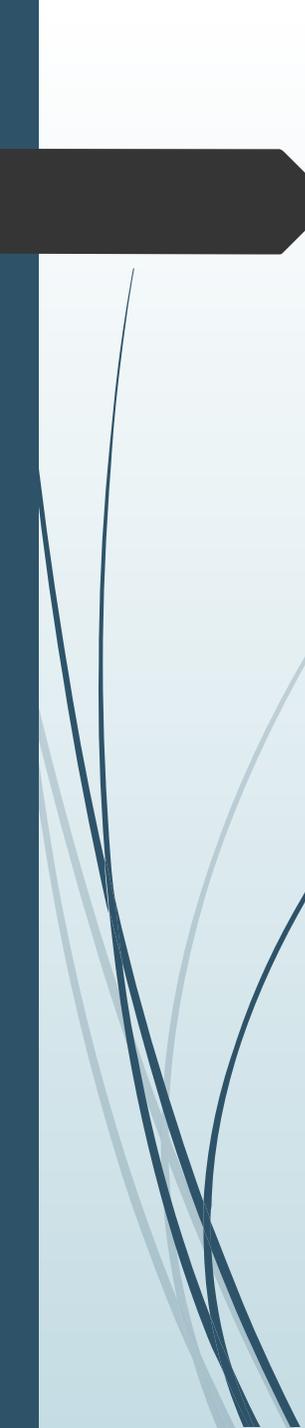
## Анализ консенсусной последовательности цитохромов Р450 семейства CYP51



Оценочная функция (ОФ), базирующаяся на статистическом критерии Шермана, имеет 6 выраженных пиков, которые соответствуют структурно-функциональным мотивам. Пятый мотив увеличен и на врезке приведен соответствующий участок выравнивания консенсусных последовательностей. Консенсусные символы 'n'-(ILMV), '='-(DEQN), '+'-(RKH), 's'-(ASTG). Точкой обозначен любой остаток. Мотив подчеркнут в последовательности белка-прародителя (БП).

Характеристика выровненных 827 последовательностей оболочечных белков E1 и E2 ВГС человека. Процентное содержание преобладающих остатков в позициях выравнивания; б) высококонсервативные участки (CR1 - CR6) и гипервариабельный участок HVR1.



- 
- Методы биоинформатики эффективно используются для выяснения механизма взаимодействия макромолекул (узнавания).
  - Методы "стыковки" (докинга) или нахождения в белках мест взаимодействия с низкомолекулярными лигандами, или друг с другом начинают доминировать не только в конструировании новых лекарств, но и в исследованиях механизма взаимодействия (узнавания) белковых молекул.
  - Методы молекулярного моделирования с последующим докингом и молекулярной динамикой являются в важным методическим инструментом для исследования механизма функционирования макромолекул.



Биоинформатика является базовой дисциплиной, прежде всего, при поиске мишеней действия новых лекарственных препаратов.

В оценке перспективности конкретной мишени учитываются также возможности нахождения соответствующих лигандов (ингибиторов или активаторов). В процессе поиска базовых структур новых лигандов и на этапе оптимизации свойств веществ-кандидатов широко используются компьютерные методы.

# Моделирование взаимодействия лиганд-мишень

Трехмерная структура  
биологической макромолекулы-мишени

Анализ структуры  
места связывания  
лиганда

Компьютерная модель  
лиганд-связывающего участка

Формализованное  
описание лиганд-  
связывающего участка

Компьютерные  
базы данных по  
структуре  
малых молекул

Компьютерный  
"слепок" с места  
связывания лиганда

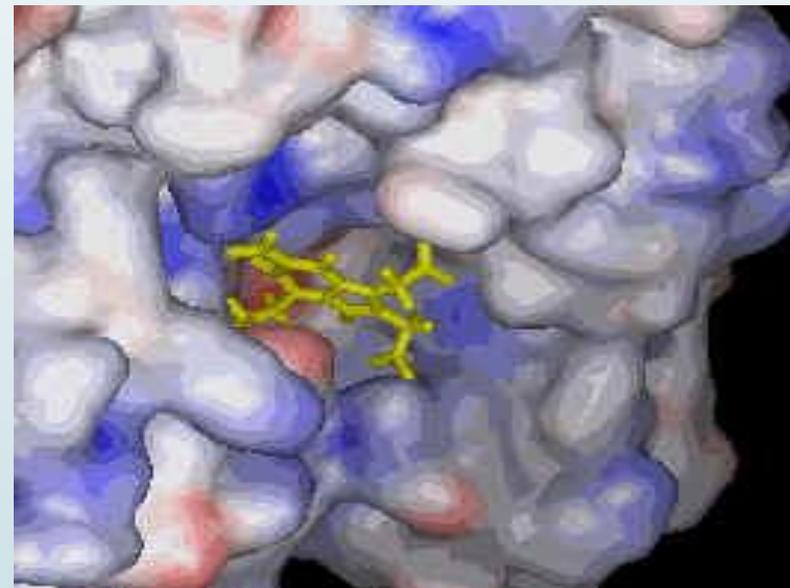
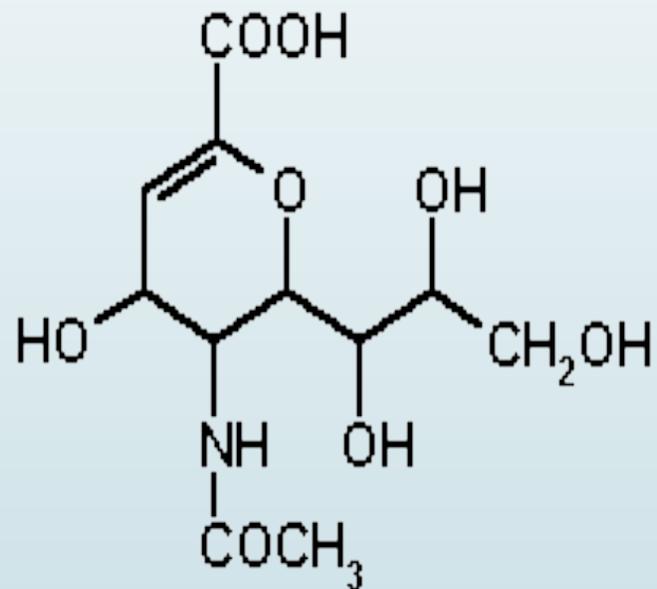
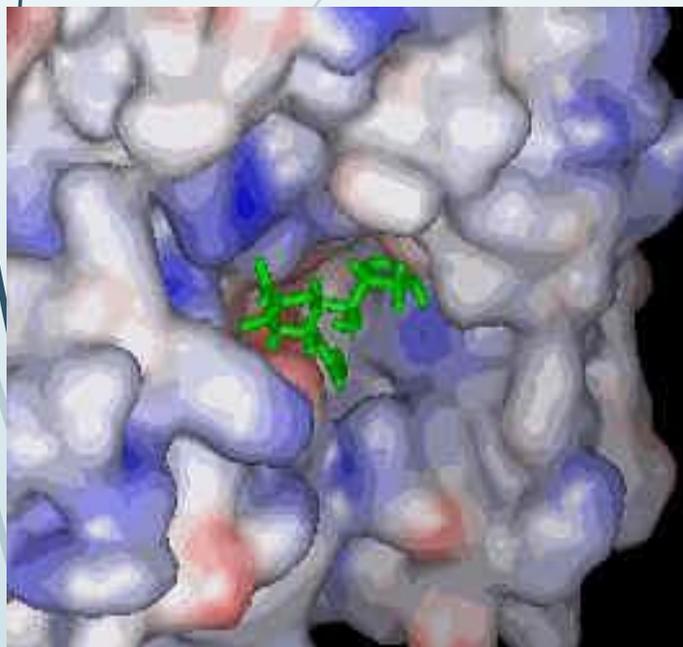
Скрининг  
малых  
молекул

Кандидаты на роль базовых  
структур для новых лекарств

Испытание биологической  
активности

Фармакологические  
испытания

Комплекс нейраминидазы вируса гриппа с известным ингибитором N-ацетил-2,3-дегидро-2-деоксинейраминовой кислотой (А) и найденным путем скрининга баз данных коммерчески доступных низкомолекулярных соединений новым лигандом «L8» (В).



# Заключение

**Благодаря техническому прогрессу биоинформатика смогла далеко продвинуться вперед и во многом облегчить труд ученых со всего мира. Без компьютерных биоинформационных технологий развитие геномных исследований было бы невозможным. Компьютерный поиск генов особенно важен для исследования генома человека, так как методы классической генетики имеют ограниченное применение в этом случае - ведь человек, в отличие мух-дрозофил, не может быть объектом искусственного мутагенеза или иных генетических экспериментов.**