МЕТОДЫ АНАЛИЗА СТРУКТУРЫ БЕЛКА

Пространственная (третичная) структура белков

Знание третичной структуры белков существенно углубляет наше понимание о том, как этот белок (фермент) работает (функция) и обеспечивает твёрдую почву для последующих попыток модификации этой функции - увеличение/уменьшение стабильности (температурной, pH, др.) и/или субстратной специфичности).

Глобулярные белки представляют наибольший интерес. Структура белков (взаимное положение тяжёлых атомов в пространстве) определяется экспериментально – методами рентгеноструктурного анализа, методом многомерного ЯМР.

В последние годы появились работы, посвящённые определению формы макромолекулярных комплексов методом криоэлектронной микроскопии.

Структуры макромолекул собраны в базы данных, основной из которых является **Protein Data Bank (PDB)**, поддерживаемая Research Collaboratory for Structural Bioinformatics (RCSB).

Банк белковых структур. Protein Data Bank (PDB)

http://www.rcsb.org/pdb

Research Collaboratory for Structural Bioinformatics

Каждая структура имеет свой идентификатор (4 символа) и ей соответствует файл, в котором приведены координаты тяжёлых атомов.

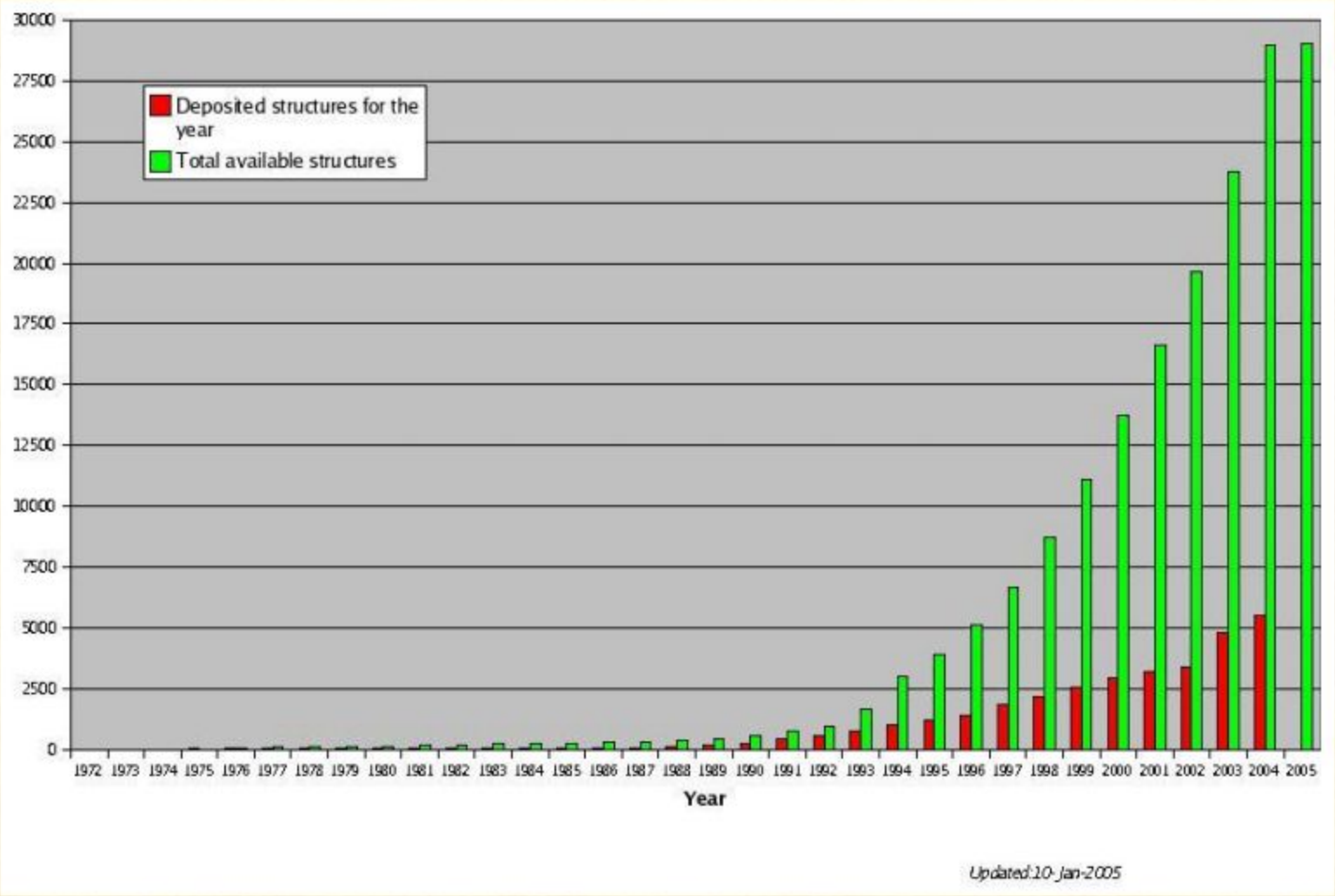
Структура PDB файла. (Brookhaven Protein Data Bank)

```
HEADER    OXIDOREDUCTASE(ALDEHYDE(D),NAD+(A))     06-JUL-94   2NAD      2NAD   2
COMPND    NAD-DEPENDENT FORMATE DEHYDROGENASE (E.C.1.2.1.2)             2NAD   3
COMPND   2 (HOLO FORM) COMPLEXED WITH NAD AND AZIDE                     2NAD   4
SOURCE    (METHYLOTROPHIC BACTERIUM PSEUDOMONAS SP. 101)                2NAD   5
AUTHOR    V.S.LAMZIN,Z.DAUTER,V.O.POPOV,E.H.HARUTYUNYAN,K.S.WILSON      2NAD   6
……
HELIX    1 H1A GLY A   55  LEU A   59  5 LEFT-HANDED                    2NAD 194
HELIX    2 H1  LEU A   59  GLY A   67  1 3/10 FOR RES 63 - 67           2NAD 195
……
ATOM    281  CB  LEU A  36      65.524  35.308   0.941  1.00 14.15     2NAD 584
ATOM    282  CG  LEU A  36      64.771  34.257   1.735  1.00 17.41     2NAD 585
ATOM    283  CD1 LEU A  36      64.749  34.636   3.209  1.00 28.05     2NAD 586
ATOM    284  CD2 LEU A  36      63.336  34.097   1.458  1.00 22.50     2NAD 587
……
END
```

Поступления в базу данных пространственных структур биополимеров

На 9.03.2010 в базе данных PDB (http://www.rcsb.org) было депонировано **63956** структур, из них более **59183** белковых структур.

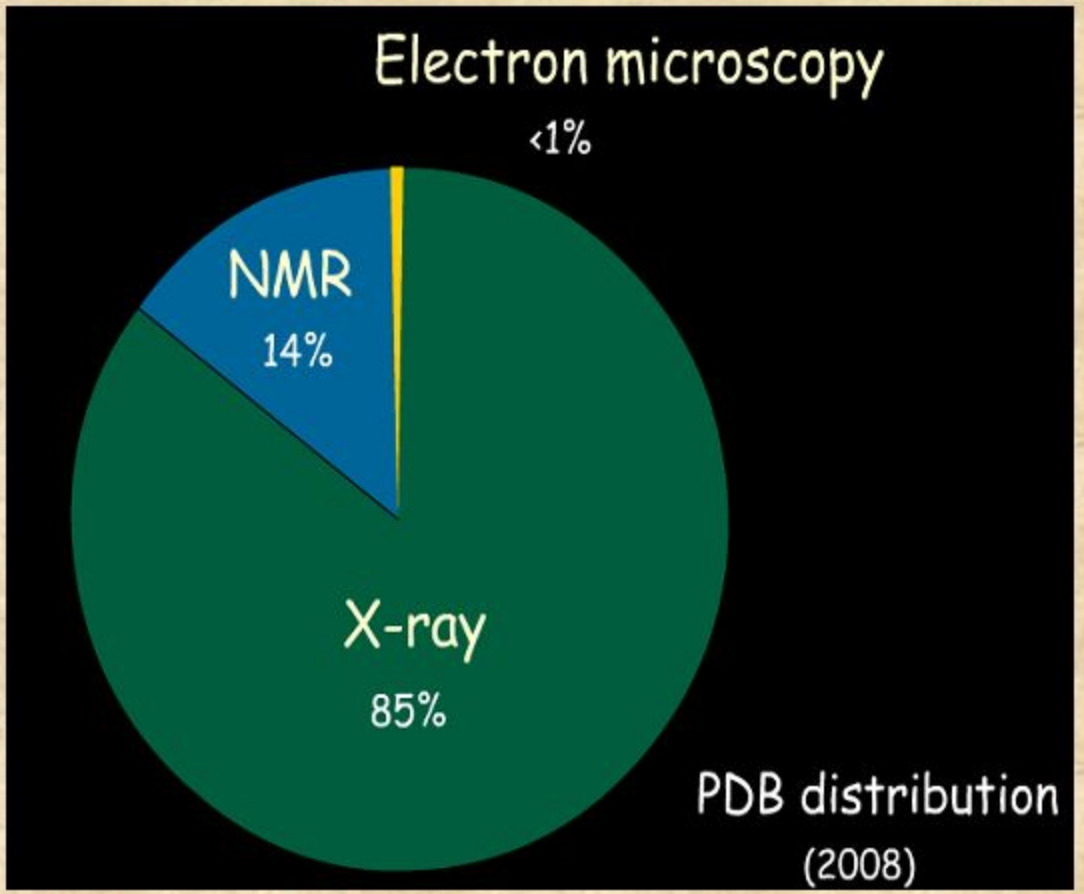
Динамика роста количества разрешённых структур



PDB сегодня

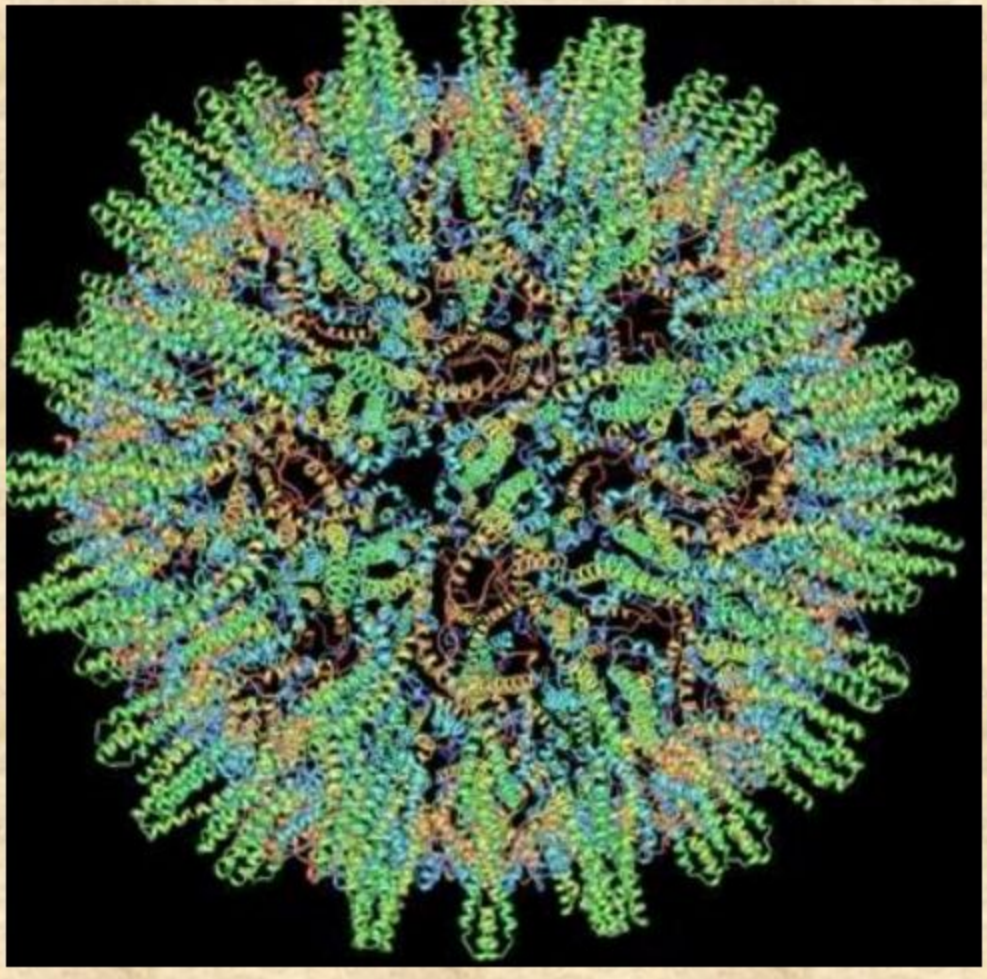
Экспериментальный метод	Белки	Нуклеиновые кислоты	Белок-НК комплексы	Другое	Всего
Рентгеноструктурный анализ	65950	1346	3261	2	70559
ЯМР	8185	979	186	7	9357
Электронная микроскопия	284	22	116	0	422
Гибридные методы	44	3	2	1	50
Другое	140	4	5	13	162
Всего:	74603	2354	3570	23	**80550**

Распределение разрешённых структур по экспериментальным методам



Molecular Conceptor v. 2.11, Synergix ltd., USA

Преимущества метода рентгеноструктурного анализа.



Molecular Conceptor v. 2.11, Synergix ltd., USA

- принципиально достижимо высокое разрешение. Разрешение выше 1Å позволяет определять степень протонирования а/к остатков в белках
- возможность разрешать структуры объектов большого размера (вирусная капсида, рибосома, фотосинтетический реакционный центр, т.д.), состоящих из нескольких десятков тысяч атомов.

Ограничения метода рентгеноструктурного анализа биомолекул

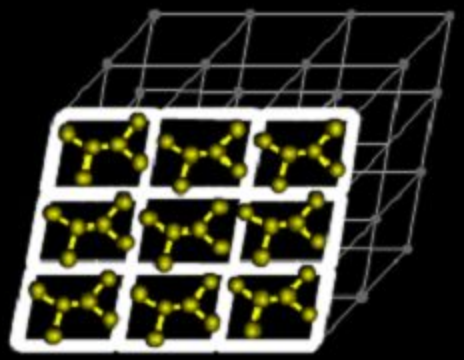
static view

slow procedure

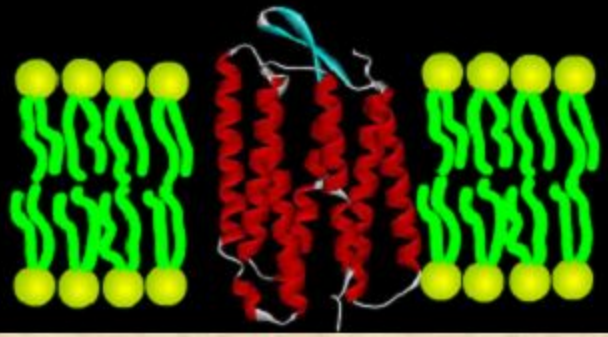
S	M	T	W	T	F	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

20 человеколет на GroEL

contact effect

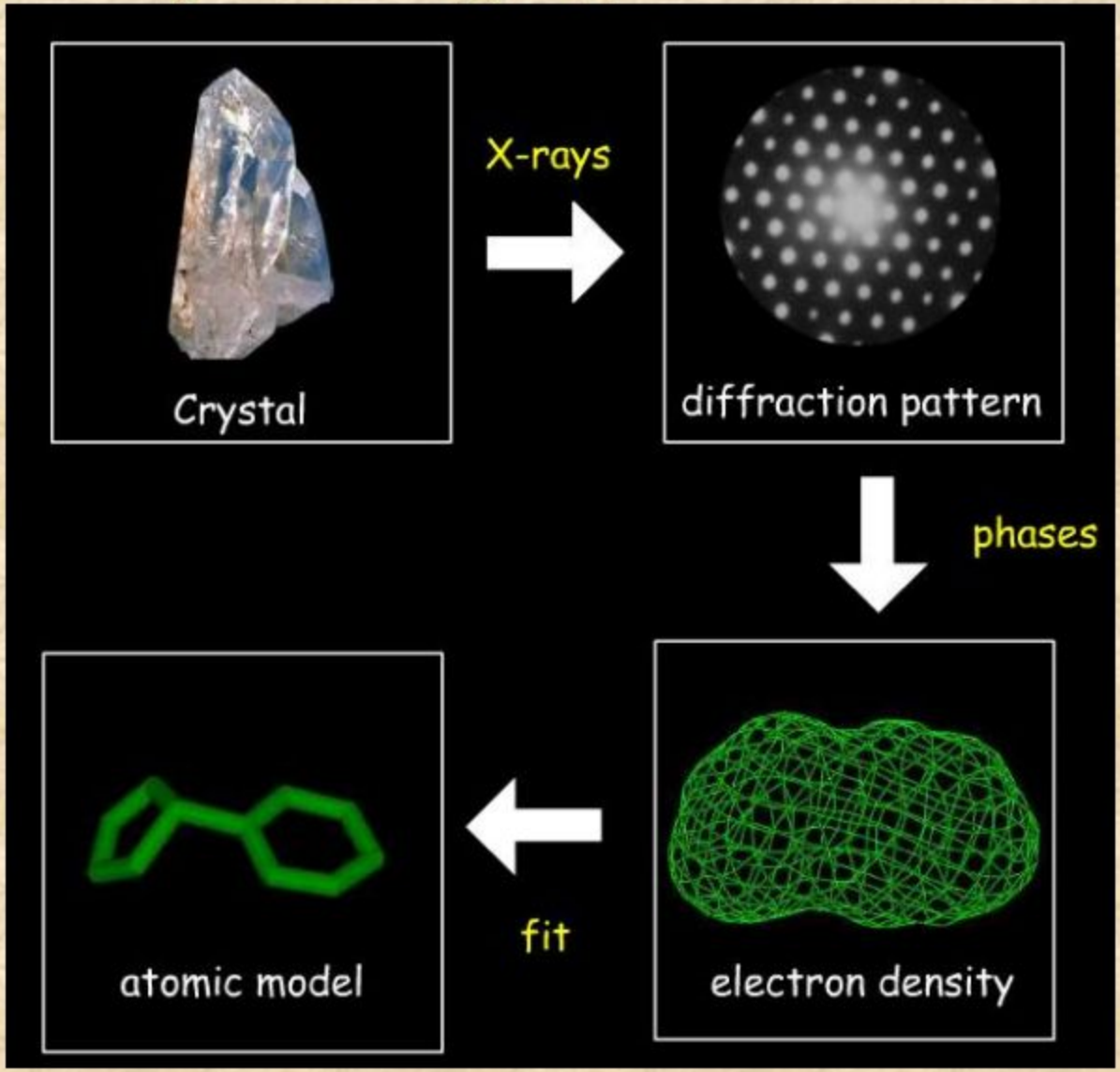


can not be solved



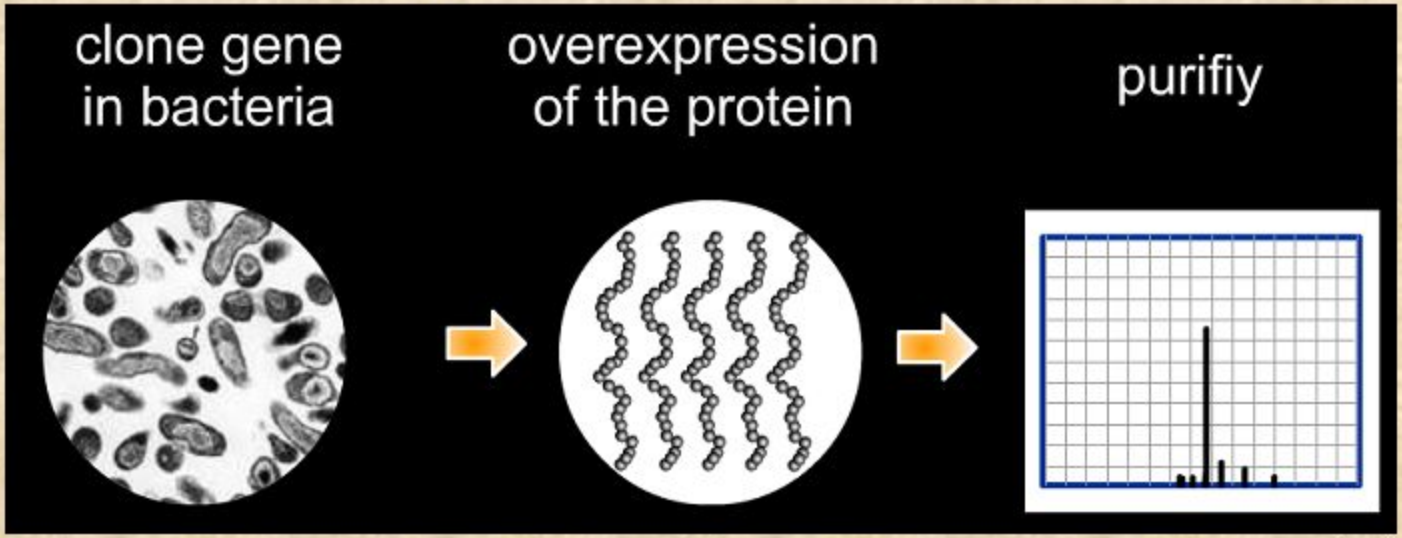
Molecular Conceptor v. 2.11, Synergix ltd., USA

Схема рентгеноструктурного исследования

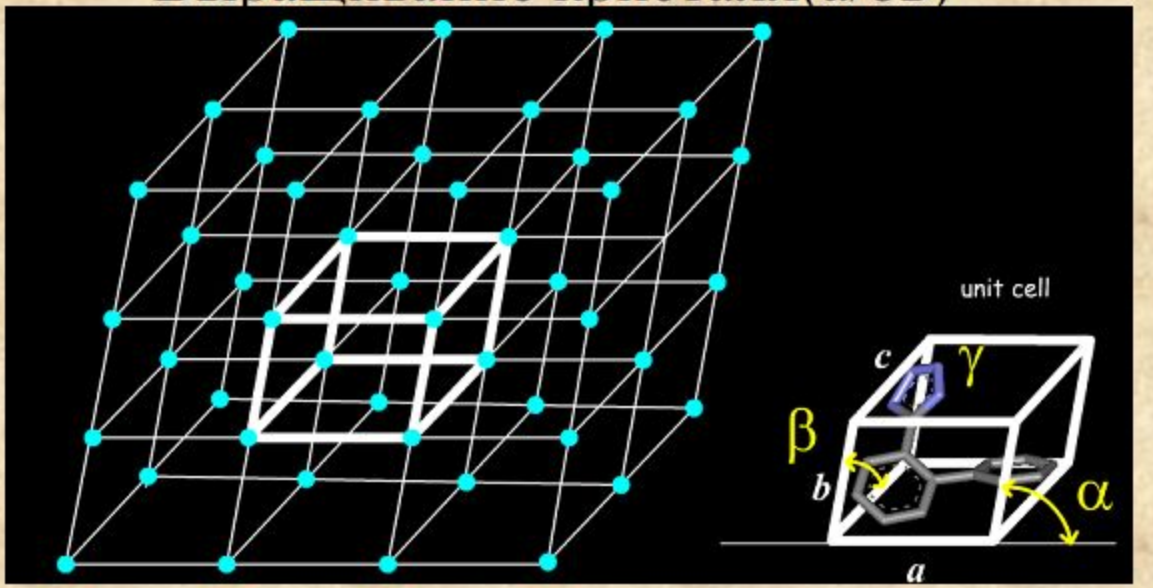


Molecular Conceptor v. 2.11, Synergix ltd., USA

Наработка и очистка белка



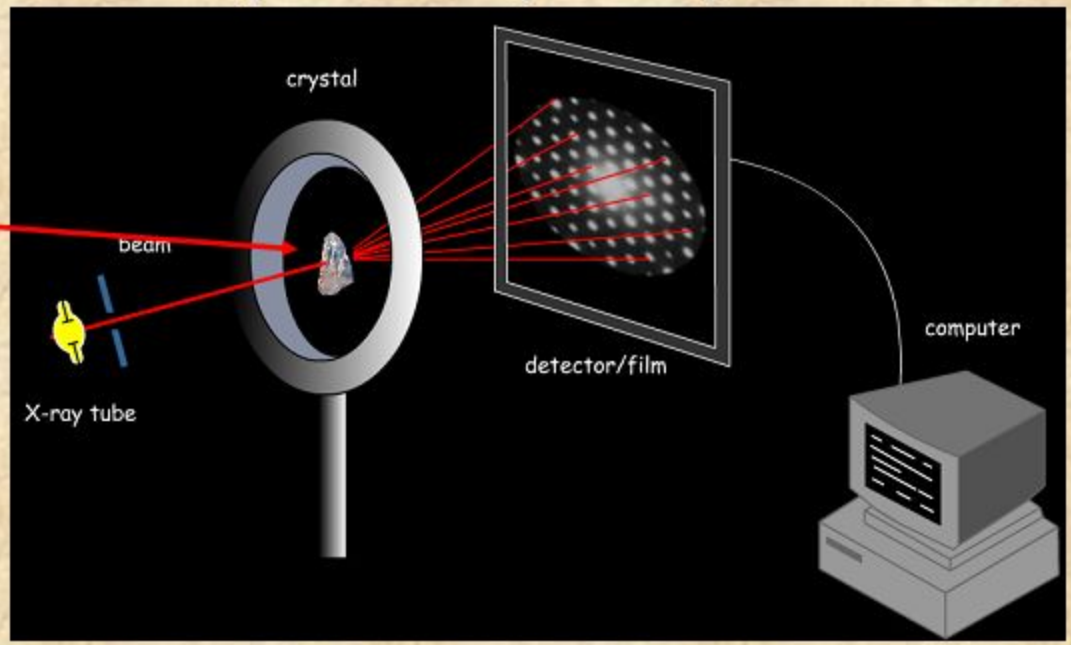
Выращивание кристалл(а/ов)



Molecular Conceptor v. 2.11, Synergix ltd., USA

Снятие рентгенограмм кристаллов

Регулярный
Кристалл
Размером
От 0,3 мм



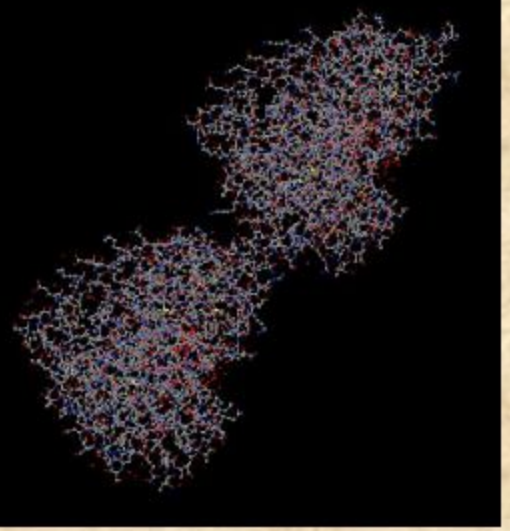
Molecular Conceptor v. 2.11, Synergix ltd., USA

Определение координат тяжёлых атомов биомолекулы

Resolution = 1.9 A

R-factor = 11%

					X	Y	Z
ATOM	17	C4	C A	1	15.917	28.226	25.120
ATOM	18	N4	C A	1	14.828	27.477	25.444
ATOM	19	C5	C A	1	15.719	29.442	24.471
ATOM	20	C6	C A	1	16.843	30.171	24.101
ATOM	21	P	G A	2	22.409	31.286	21.483
ATOM	22	O1P	G A	2	21.822	31.459	20.139
ATOM	23	O2P	G A	2	23.536	32.157	21.851
ATOM	24	O5*	G A	2	22.840	29.751	21.498
ATOM	25	C5*	G A	2	23.543	29.175	22.594
ATOM	26	C4*	G A	2	23.494	27.709	22.279
ATOM	27	O4*	G A	2	22.193	27.252	22.674



Molecular Conceptor v. 2.11, Synergix ltd., USA

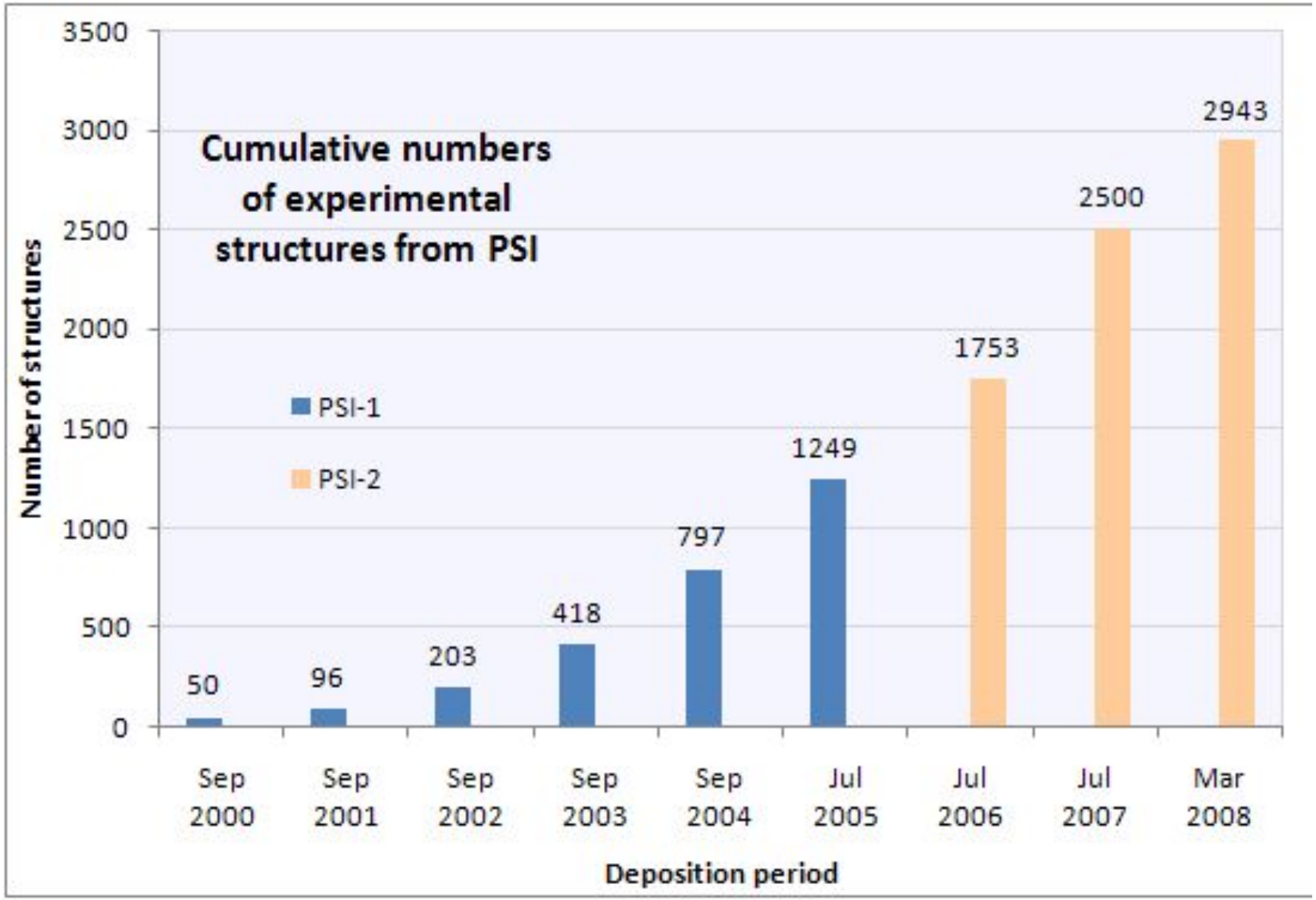
Protein Structure Initiative (NIGMS, NIH, USA, 2001-2010, 2011-2015 ?)

Service R.B., *Science* **319**, 1610 (2008)

PSI

- **PSI-I** – 2000-2005. Цель разработка высокопроизводительных методов для анализа структуры белка. Бюджет 270 млн. $.1100 структур, 700 уникальных. Объекты - *Arabidopsis thaliana*, *Caenorhabditis elegans* and *Mycobacterium tuberculosis*.
- **PSI-II** – 2005-2010. Цель – масштабные исследования по определению структуры белков, преимущественно уникальных. Бюджет 325 млн. $. 4800 структур, 4100 уникальных.
- **PSI-III** – использование разработанных подходов для решения важных биологических проблем, связанных с медициной.

Результаты PSI-I и PSI-II



Многомерная ЯМР спектроскопия

Преимущества:

- молекулы в растворе (тяжёлая вода), не нужно выращивать кристалл
- положения атомов водорода м.б. определены
- информация о динамике атомов м.б. определена

В белках 10^3 – 10^4 протонов
Метильные и метиленовые группы 0,8-3,5 ppm, ароматические, индольные и иммидазольные кольца 6,5-8 ppm
В ДНК/РНК H-2/H-8 пуринов 8,4-9 ppm, H-5 пиримидинов 6,3-6,6 ppm, H-6 8,0-8,5 ppm, метильная группа тимидина 2,3-2,4 ppm.

Molecular Conceptor v. 2.11, Synergix ltd., USA

Многомерная ЯМР спектроскопия

Преимущества:

- молекулы в растворе (тяжёлая вода), не нужно выращивать кристалл
- положения атомов водорода м.б. определены
- информация о динамике атомов м.б. определена

В белках 10^3 – 10^4 протонов
Метильные и метиленовые группы 0,8-3,5 ppm, ароматические, индольные и иммидазольные кольца 6,5-8 ppm
В ДНК/РНК H-2/H-8 пуринов 8,4-9 ppm, H-5 пиримидинов 6,3-6,6 ppm, H-6 8,0-8,5 ppm, метильная группа тимидина 2,3-2,4 ppm.

Molecular Conceptor v. 2.11, Synergix ltd., USA

Bruker Biospin AVANCE 1000
The World's First 1 Gigahertz NMR Spectrometer

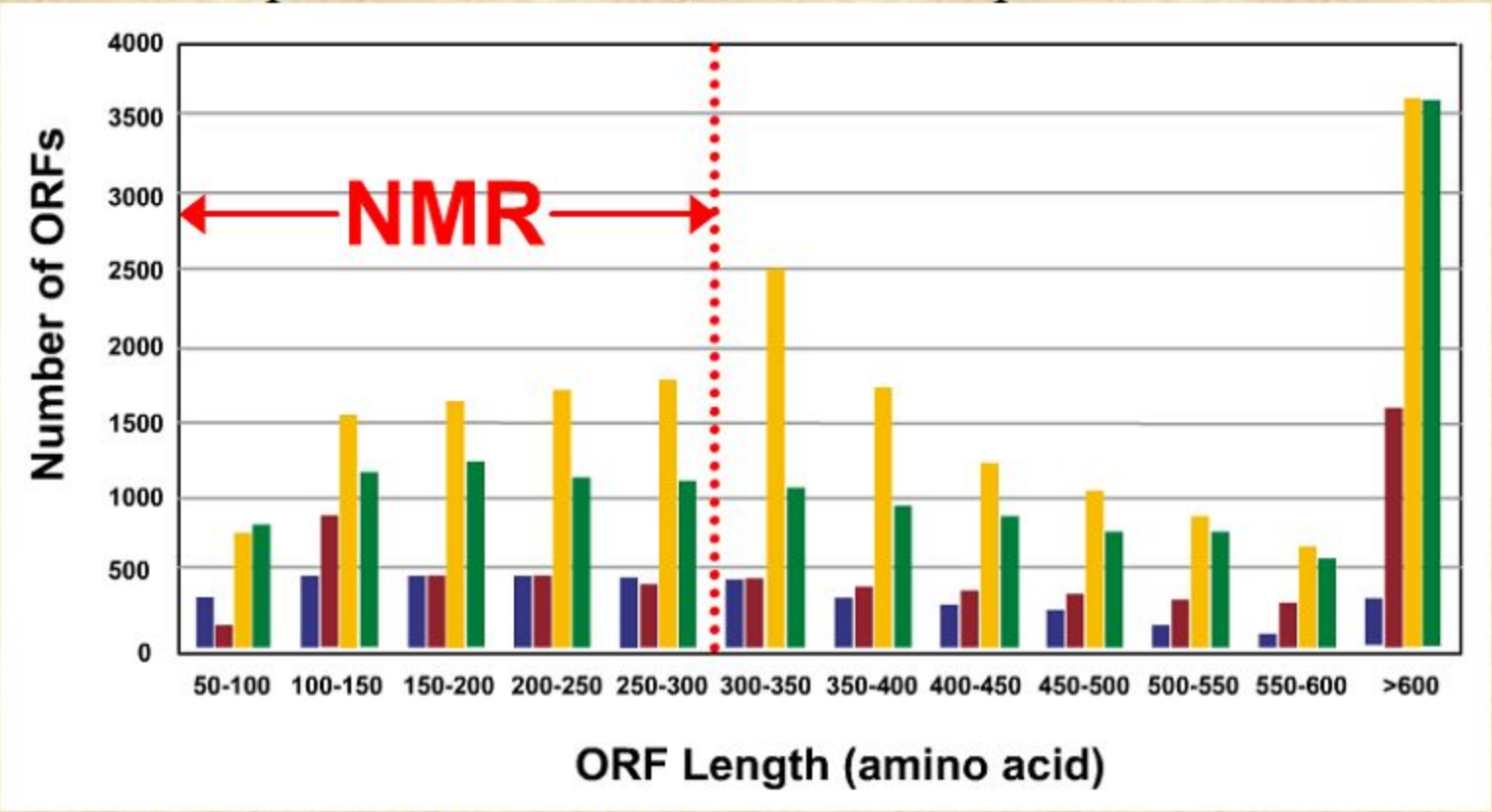
World's First 1 Gigahertz NMR Spectrometer based on unique 23.5 Tesla Standard-Bore, Persistent Superconducting Magnet (12-tonne, 4.5-metre-tall machine)

Dr. Lyndon Emsley, European Centre for High Field NMR (CRMN) in Lyon, France.
Nature **463**,605(2010).

€11.7-million (US$16.3-million)

http://www.bruker-biospin.com/av1000-dir.html

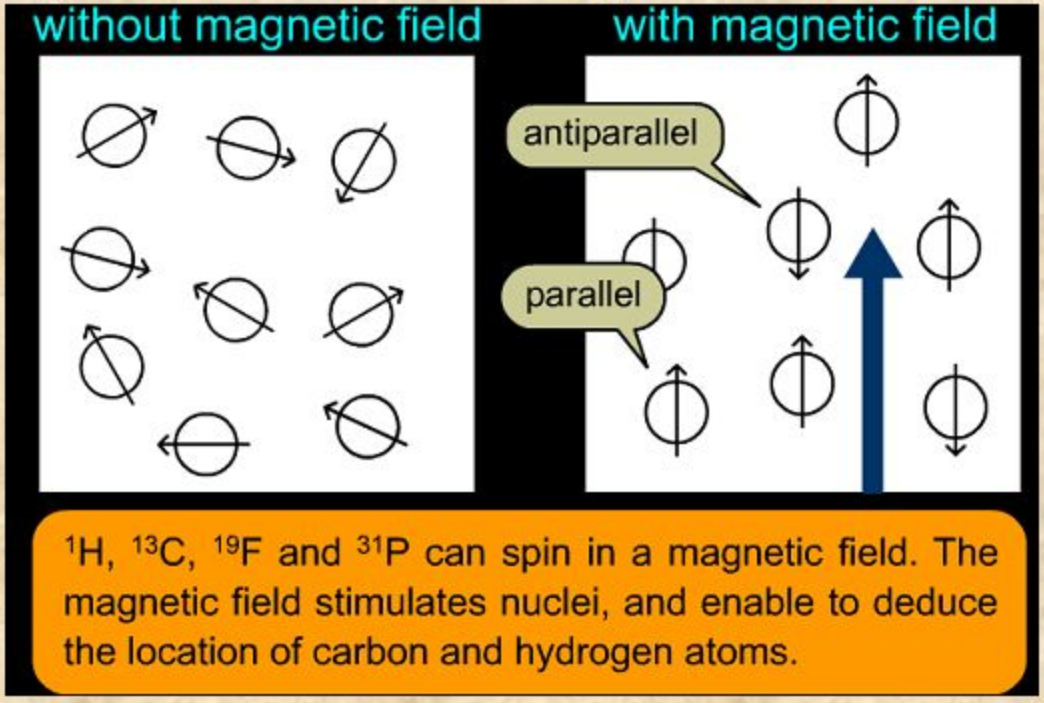
Ограничение метода многомерного ЯМР

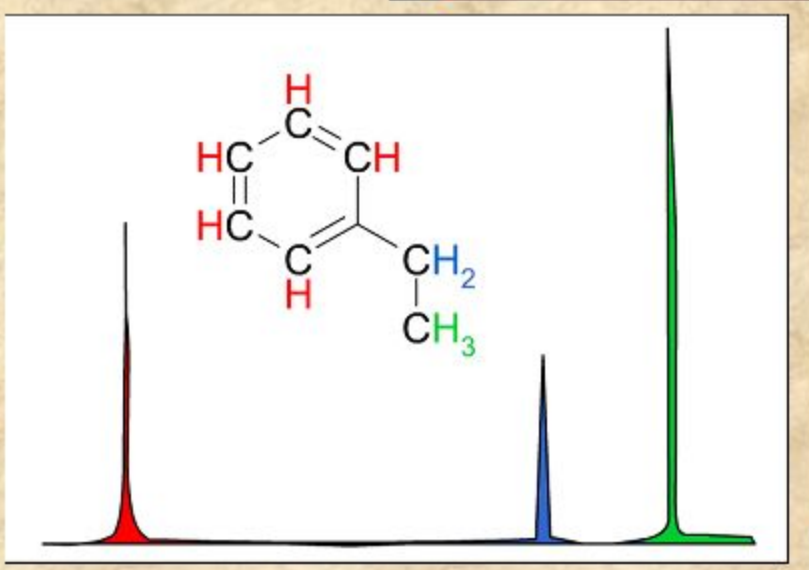


Molecular Conceptor v. 2.11, Synergix ltd., USA

- Структура белков < 300 а/к остатков м.б. Определена этим методом, что составляет менее половины известных белковых последовательностей. Известны примеры разрешения структур белков из 700 а/к.
- Дороговизна получения образцов с изменённым изотопным составом (^{13}C, ^{19}F, ^{31}P)
- Невысокая точность разрешения структуры

Физические принципы метода ядерного магнитного резонанса

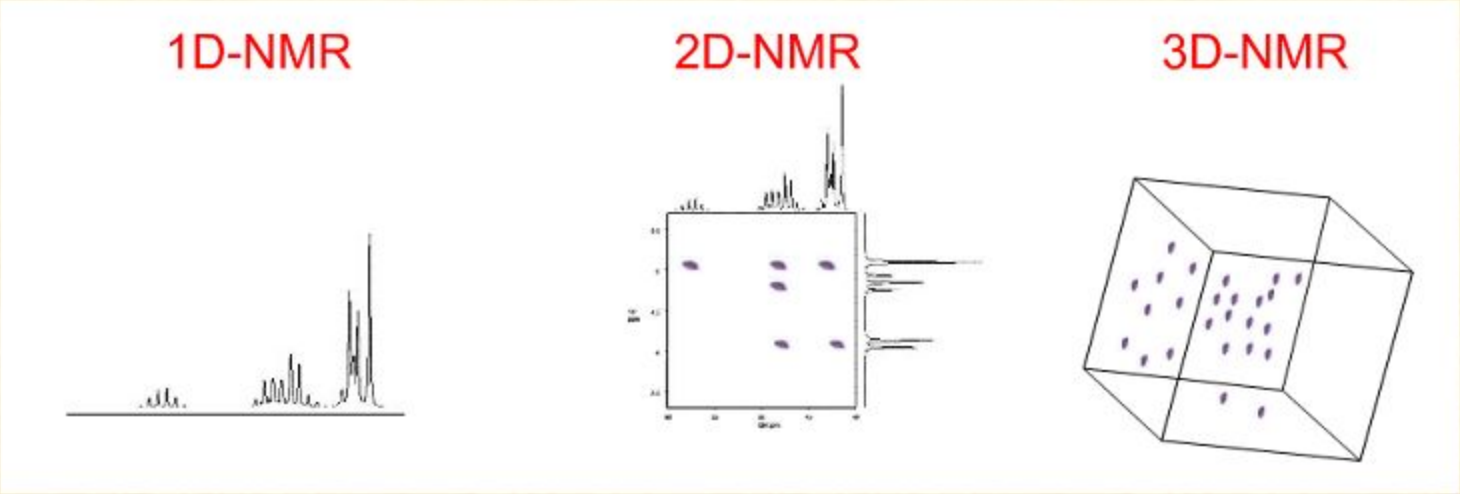


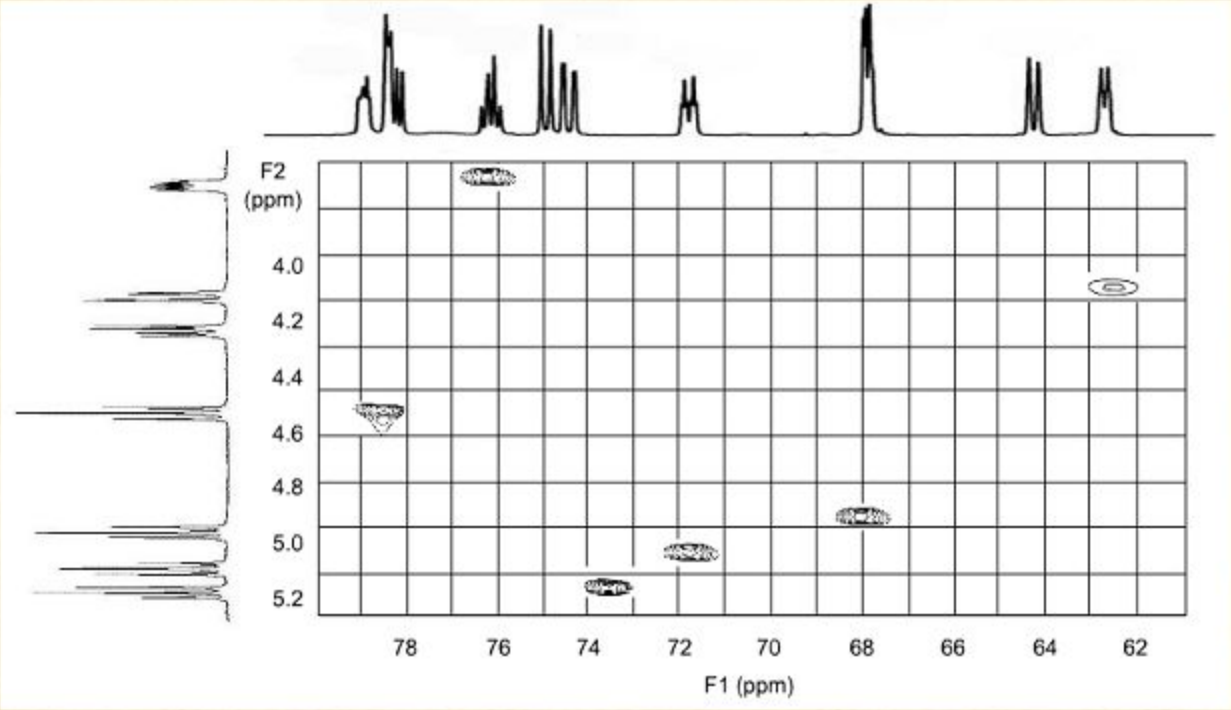


Характерные спектры химических групп и соединений

Molecular Conceptor v. 2.11, Synergix ltd., USA

Сбор данных

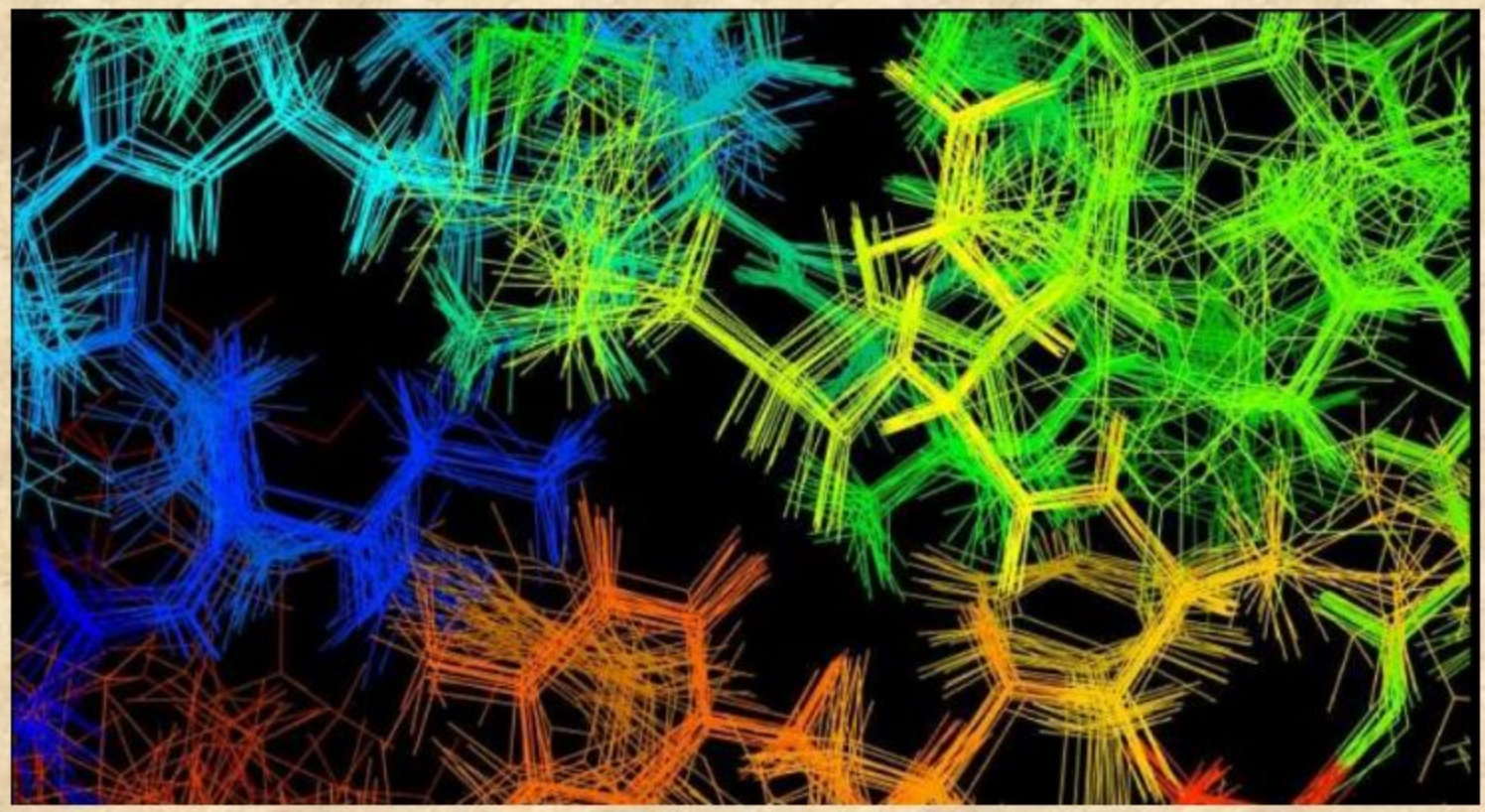




Анализ,
Соотнесение
(assignment)

Molecular Conceptor v. 2.11, Synergix ltd., USA

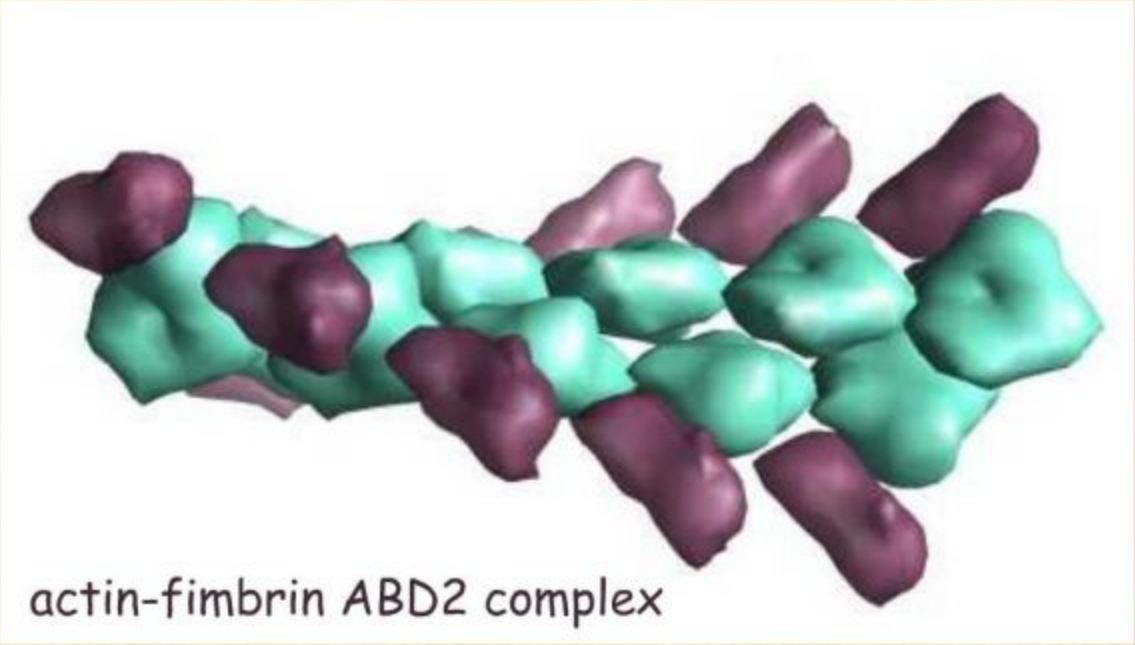
Определение координат атомов молекулы



Molecular Conceptor v. 2.11, Synergix ltd., USA

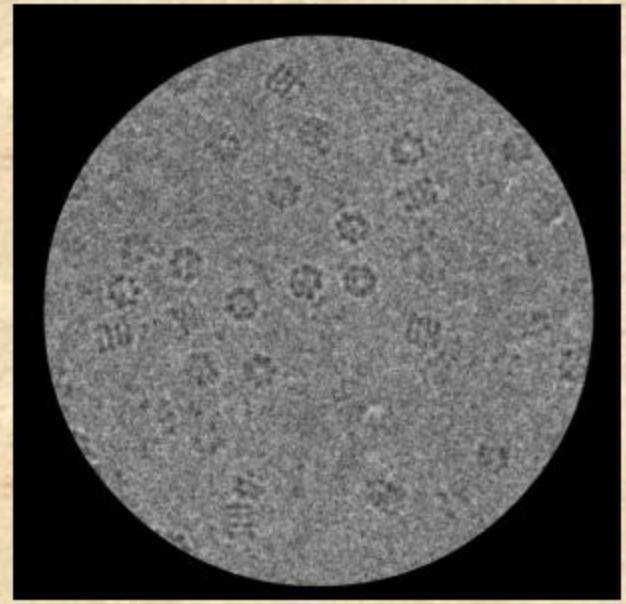
Для структур разрешённых методом многомерного ЯМР представлено 10-20 структур. Усреднённая структура имеет наибольшую достоверность.

Электронная микроскопия



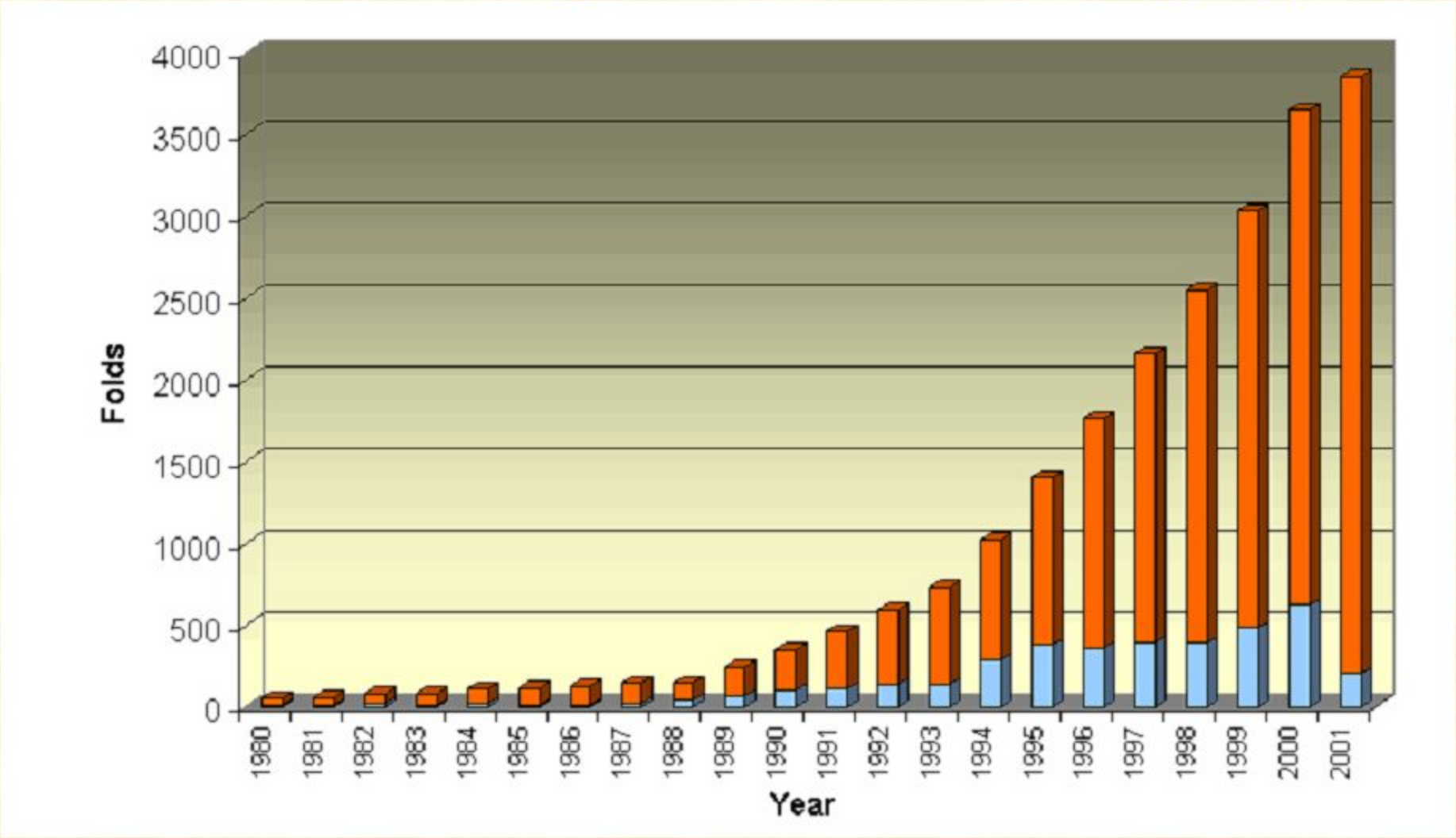
Определяется форма крупных межмолекулярных комплексов методом диффракции электронных пучков. Типичное разрешение этого метода 3-5 Å не позволяет определять координаты атомов.

Molecular Conceptor v. 2.11, Synergix ltd., USA



Образцы в замороженном состоянии, что предотвращает радиационные повреждения и удерживает их в нативном состоянии.

Количество новых фолдов в банке белковых структур



Голубым цветом количество «новых» фолдов, оранжевым количество «старых» фолдов. PSI ставит своей целью разрешение структур с новыми фолдами в ущерб биологической значимости белка, что вызывает критику оппонентов.

Актуальность разработки методов моделирования белковых структур

Увеличивающийся разрыв между определёнными первичными структурами белков (11,000,000 в базе SWIS-PROT/TrEMBL) и количеством разрешённых белковых структур (59,000 из которых только около 5,500 непохожи друг на друга) указывает на необходимость разработки методов предсказания третичных структур белков.

Поскольку в настоящее время *отсутствуют надёжные теоретические методы** прямого предсказания третичной структуры белков из первичной, методы предсказания должны основываться на структурах известных белков.

Это подразумевает, что *структуры неизвестного и опорного (reference) белков должны быть похожи или гомологичны. О степени гомологичности белков судят по степени идентичности их первичных структур*.

Анализ эволюционно связанных белков показал, что при небольших эволюционных расстояниях степень их гомологичности, определённая по первичной структуре, достаточно высока.

При возрастании эволюционного расстояния гомология по последовательности становится трудно уловимой, однако укладка полипептидной цепи (фолд/folding) остаются схожими.

Суть моделирования третичной структуры по гомологиям – выравнивание последовательностей моделируемого и опорного белков с последующим переносом элементов структуры опорного белка на моделируемый.

* В работе Qian *et al*., Nature 450, 259-264(2007) была аккуратно de novo предсказана третичная структура белка из 112 а/к остатков. ROSETTA@HOME сеть включает более 70,000 компьютеров.