

# Биоинформатика

---

Область науки, в которой решаются биологические задачи с помощью вычислительных методов математики и информационных технологий

# Биоинформатика

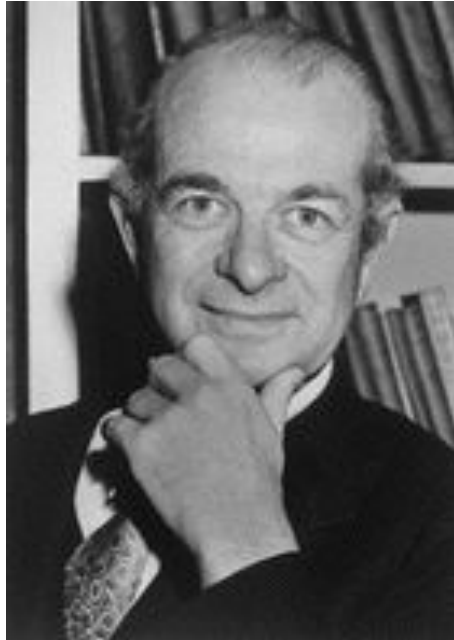
---

- Разработка алгоритмов для анализа биологических данных большого объема
  - Алгоритм поиска генов в геноме
- Анализ и интерпретация различных типов биологических данных таких, как нуклеотидные и аминокислотные последовательности, домены белков, структура белков и т.д.
  - Изучение структуры активного центра белка
- Разработка программного обеспечения для управления и быстрого доступа к биологическим данным
  - Создание банка данных аминокислотных последовательностей

# Пионеры биоинформатики

---

1962



## Лайнус Полинг

- Анализ аминокислотных последовательностей глобинов нескольких позвоночных
- Гипотеза **молекулярных часов**

Zuckerlandl, E., and L. Pauling. **1962**. Molecular disease, evolution, and genic heterogeneity. Horizons in Biochemistry, Academic Press, New York, 189-225.

Zuckerlandl, E., and L. Pauling. **1965**. Evolutionary divergence and convergence in proteins. Evolving Genes and Proteins, Academic Press, New York, 97-166.

# Пионеры биоинформатики

1965



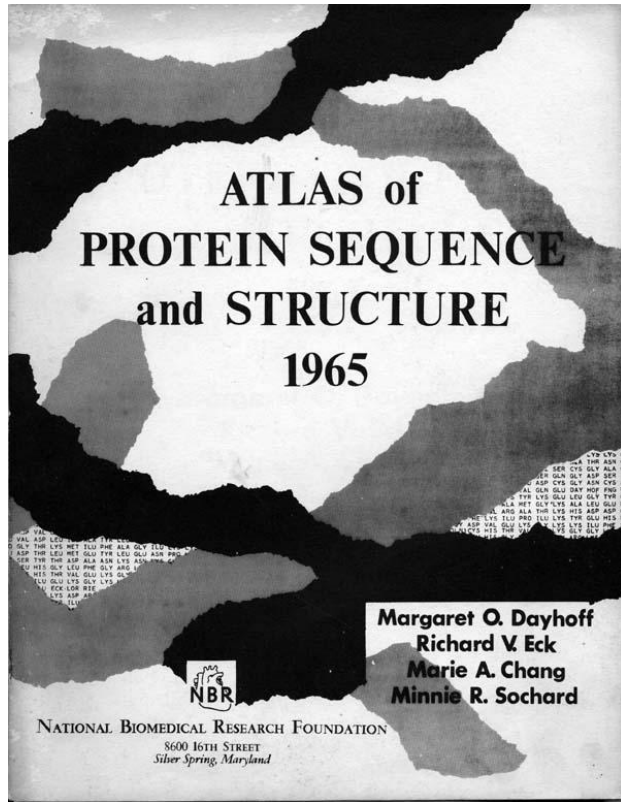
## Маргарет Дейхофф

- Однобуквенный код аминокислот  
A,C,D,E,F,G,H...
- Матрицы аминокислотных замен  
PAM (Point Accepted Mutation)

Атлас последовательностей белков и их структур (1965)

# Первый “банк данных”

1965 -1978



Атлас белковых  
последовательностей и  
их структур

Первая версия атласа содержала описание **65 !** последовательностей белков

# Банк данных PIR-PSD

---

1984



**PIR** Protein Information Resource

**PSD** (**P**rotein **S**equences **D**atabase) – банк данных классифицированных и аннотированных белковых последовательностей

- одно из подразделений **PIR**
- прямой наследник атласа Дейхофф

Текущий релиз **80** (31 декабря 2004) содержит **283 416** документов

---

<http://pir.georgetown.edu/pirwww/search/textpsd.shtml>

# Банк данных Swiss-Prot

---

1986



**Swiss-Prot** – база знаний о  
белковых последовательностях

- Курируемая база данных
- “**Золотой стандарт**” аннотации

# Банк данных Swiss-Prot

---



С 1987 поддерживается в сотрудничестве между

Swiss Institute of Bioinformatics (**SIB**)  
European Bioinformatics Institute (**EBI**)



## Амос Байрох

Руководитель группы Swiss-Prot в  
Швейцарском Институте Биоинформатики



# Банк данных Swiss-Prot

## Статистика роста количества документов



Текущий релиз **45.5** (4 января 2005) содержит **167 089** документов

# Документ банка данных Swiss-Prot



The image shows a screenshot of a Swiss-Prot database entry. The entry is displayed in a table-like format with various fields. The top part contains the protein name and accession number. Below that, there is a section for the protein description, followed by a section for the amino acid sequence. The sequence is presented in a multi-line format with residue numbers. The bottom part of the entry contains additional information such as references and cross-references to other databases.

Описание документа:  
идентификатор,  
имя, дата создания и модификации

Аннотация  
последовательности

Последовательность

# Банк данных TrEMBL

---



**TrEMBL (Tr**anslated **EMBL)**

Формальная трансляция всех кодирующих  
нуклеотидных последовательностей из банка EMBL

Автоматическая классификация и аннотация

Текущий релиз **29.0** (1 февраля 2005) содержит **1 589 670** документов

# Тенденция объединения

---

2002



**PIR** Protein Information Resource



# Банк данных UniProt

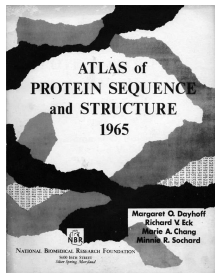
---



## UniProt (Universal Protein Resource)

- UniProt Knowledgebase – **UniProt**
- UniProt Archive – **UniParc**
- UniProt Reference – **UniRef**

кол. док. = 46 105 397



Автоматическая трансляция



кол. док. = 167 089



кол. док. = 1 589 670

Ручная аннотация

