

# **Disciplina do PPBiotec**

## **BIOTEC017: Bioinformática**

### **Prof. Teodiano Freire Bastos**

Graduação em Engenharia Elétrica - UFES, 1987

Doutorado em Ciências Físicas (Eletricidade e Eletrônica) - UCM/Espanha, 1995

Pós-Doutorado em Interfaces Cérebro-Computador - UAH/Espanha, 2005

Pós-Doutorado em Próteses Mioelétricas de Membro Superior - RMIT/Austrália, 2012

# Ementa

---

- Objetivos do Curso
- Conceitos de Bioinformática
- Níveis de Informação Biológica (Jéssica e Midori)
- Alinhamentos (Mariana e Kellyn)
- Pacote de Softwares Blast (Marcela)
- Projetos Genoma (Ligia e Jucimara)
- Modelos Tridimensionais (Danilo e João)
- Biologia de Sistemas (Diego e Fernanda) e Leonardo
- Filogenia (Natalia e Aricia) e Mateus
- Dinâmica Molecular (Erica e Juliana)
- Atracamento (Andrés e Estavão)
- Dicroísmo Circular (Iara e Simone)
- Infravermelho (Andreia e Tadeu)
- Ressonância Magnética Nuclear (RMN) (Raissa e Taciane)
- Cristalografia (Lauziene e Alexandre)

**Bioinformática**  
**(da Biologia à Flexibilidade Molecular)**

Autor: Hugo Verli (org)

1ª Edição, Sociedade Brasileira de Bioquímica e  
Biologia Molecular – SBBq, São Paulo

2014

Disponível em:

<http://www.ufrgs.br/bioinfo/ebook/>





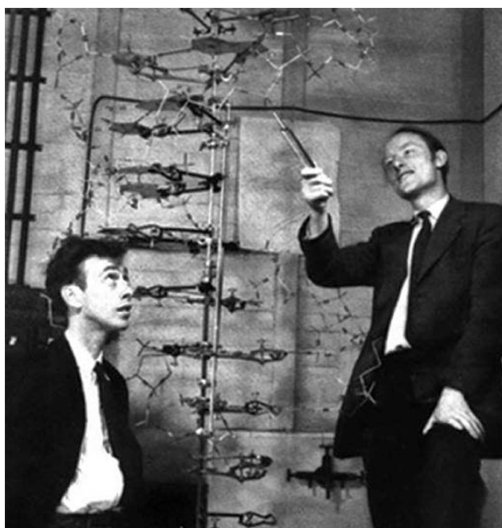
# Origens da Bioinformática

## Bioinformática Tradicional (Clássica)

Aborda principalmente problemas relacionados a sequências de nucleotídeos e aminoácidos

## Bioinformática Estrutural

Aborda questões biológicas de um ponto de vista tridimensional, abrangendo a maior parte das técnicas compreendidas pela química computacional ou modelagem molecular



### Pioneiros:

- James Watson e Francis Crick (em frente a uma estrutura em hélice da molécula de DNA, em 1950 – Revista Nature)
  - Permitiram o entendimento de como aquela "sequência de letras" (as bases do DNA) se organizam tridimensionalmente

### Outras contribuições para o advento da Bioinformática:

- Linus Pauling e Robert Corey (no início da década de 1950)
- Gopalasamudram N. Ramachandran (início da década de 1960)
  - Ofereceram as bases para a compreensão da estrutura tridimensional de proteínas

# Origens da Bioinformática

## Primeiro uso de computador em Bioinformática:

- A visualização de estruturas tridimensionais de moléculas por computador foi feita somente em 1966, por Cyrus Levinthal
  - Publicação na revista *Scientific American* do trabalho desenvolvido no *Massachusetts Institute of Technology* por John Ward e Robert Stotz

## Sistematização do conhecimento acerca da estrutura tridimensional dos efetores da informação genética, as proteínas:

- Publicação, em 1965, do *Atlas of Protein Sequence and Structure*, organizado por diversos autores, dentre os quais destaca-se Margaret Dayhoff
  - Margaret Dayhoff é considerada a precursora do que se entende hoje por “Bioinformática”, tanto em sua faceta voltada para sequências quanto para estruturas
  - Foi uma das pioneiras no uso de computadores para o estudo de biomoléculas, incluindo tanto ácidos nucleicos quanto proteínas
  - É ela que inicia o uso da representação de uma única letra para descrever cada aminoácido (ver Tabela), ao invés das usuais três letras, em uma época em que os dados eram armazenados em cartões perfurados
  - Desenvolveu as primeiras matrizes de substituição e fez importantes contribuições no desenvolvimento dos estudos filogenéticos
  - Teve participação importante no desenvolvimento de métodos para o estudo de moléculas por cristalografia de raios-X

## Nomes dos 20 aminoácidos codificadores de proteínas junto a suas representações em 1 e 3 letras

Aminoácido	Representação de 3 letras	Representação de 1 letra
Alanina	Ala	A
Cisteína	Cys	C
Ác. aspártico	Asp	D
Ác. glutâmico	Glu	E
Fenilalanina	Phe	F
Glicina	Gly	G
Histidina	His	H
Isoleucina	Ile	I
Lisina	Lys	K
Leucina	Leu	L
Metionina	Met	M
Asparagina	Asn	N
Prolina	Pro	P
Glutamina	Gln	Q
Arginina	Arg	R
Serina	Ser	S
Treonina	Thr	T
Valina	Val	V
Triptofano	Trp	W
Tirosina	Tyr	Y



IBM 7090, computador que Margaret Dayhoff utilizou no início de seus trabalhos (NASA Ames Research Center, 1961)



# Origens da Bioinformática

## Primeiros estudos acerca da dinâmica e do enovelamento de proteínas por simulações de dinâmica molecular:

- Michael Levitt e Arieh Warshel utilizaram em 1970 computadores mais poderosos e o conhecimento sobre os determinantes da estrutura e da dinâmica proteica para ganhar em 2013 o Prêmio Nobel de Química

## Avanços continuados no entendimento de biomoléculas e emprego de técnicas computacionais:

- Aumento na obtenção de informações de alta qualidade sobre a estrutura 3D de biomoléculas
  - Suporte para o desenvolvimento de campos de força cada vez mais precisos
  - Novas abordagens que possibilitam o alinhamento de sequências cada vez mais distantes evolutivamente

## Avanços e barateamento no poder computacional aplicado ao Projeto Genoma Humano

- Computadores cada vez mais rápidos permitem processamento de um gigantesco e crescente volume de sequências de genes cujas relações evolutivas e funcionais precisam ser elucidadas, como ponto de partida para novos desenvolvimentos terapêuticos
- É possível identificar um novo candidato a receptor alvo de novos fármacos a partir de organismos muito distantes evolutivamente de nós, como leveduras, bactérias ou mesmo plantas

Agraciados pelo prêmio Nobel de química de 2013, os Professores Martin Karplus, Michael Levitt e Arieh Warshel



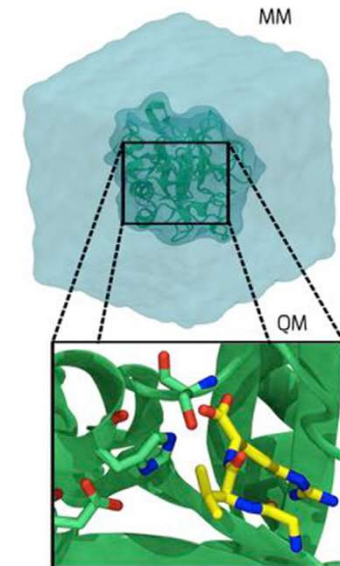
Martin Karplus



Michael Levitt



Arieh Warshel



# Origens da Bioinformática

---



## **Novos conhecimentos do funcionamento de sistemas biológicos:**

- Avanços em Bioinformática permitem estudos de transcriptoma, metaboloma ou glicoma, potencializando tanto aplicações terapêuticas quanto biotecnológicas

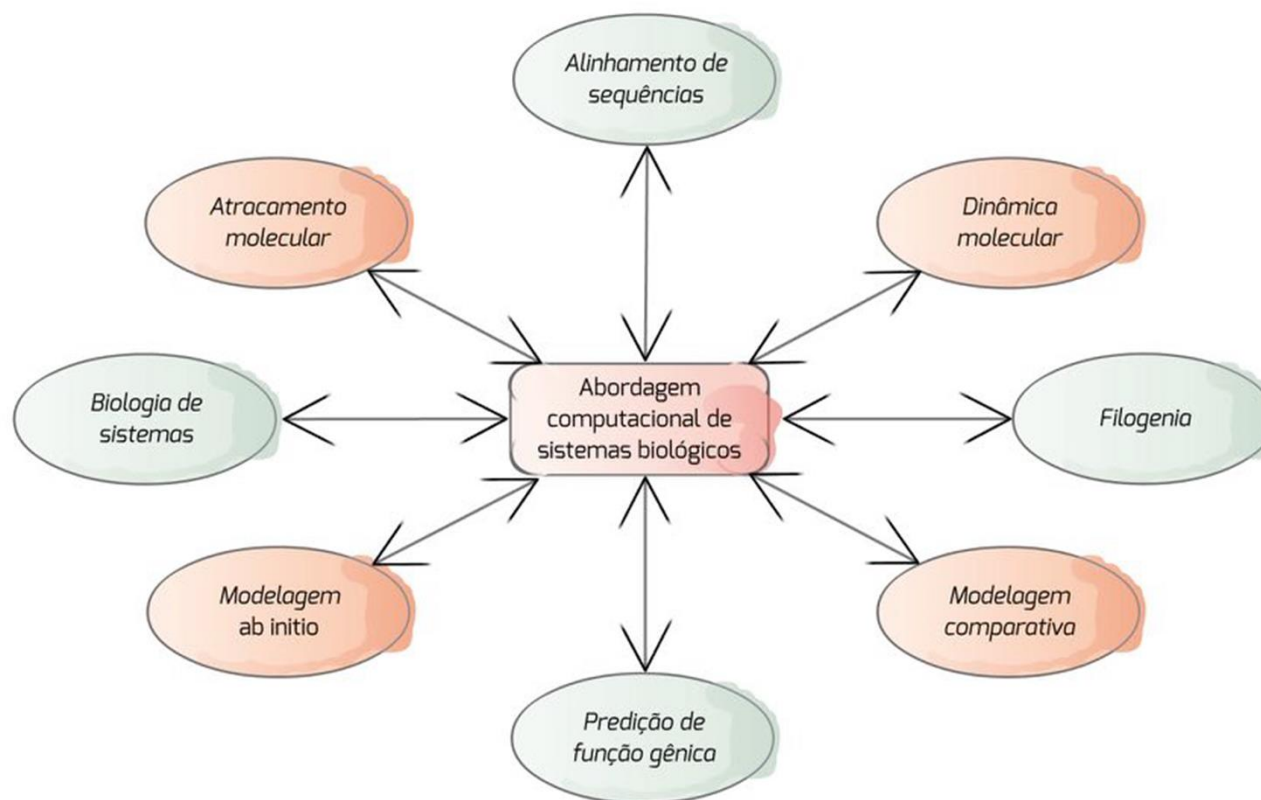
## **Avanços necessários em Bioinformática (hardware e software):**

- É possível realizar em um computador pessoal simples alinhamentos com algumas centenas de sequências sem maiores dificuldades, e ter uma resposta quase imediata
- Entretanto, para realizar uma simulação por dinâmica molecular de uma única proteína, é necessário, neste mesmo computador, de alguns meses
- **Bioinformática: uma das áreas do conhecimento mais acessíveis para novos pesquisadores**



# Problemas Alvo

- Sequências de Biomoléculas
- Estrutura de Biomoléculas



## Representação de algumas das principais áreas da Bioinformática

- Em Cinza: metodologias envolvidas principalmente com sequências de biomoléculas
- Em Laranja: metodologias que lidam majoritariamente com estruturas 3D de biomoléculas

# Sequências de Biomoléculas

---



- **Comparações entre sequências (alinhamento)**
- **Identificação de padrões em sequências (assinaturas)**
- **Caracterização de relações evolutivas (filogenia)**
- **Construção e anotação de genomas**
- **Construção de redes (biologia de sistemas)**

# Estruturas de Biomoléculas

---



- **Obtenção de modelos 3D para proteínas e outras biomoléculas (por exemplo, modelagem comparativa)**
- **Identificação do modo de interação de moléculas (atracamento)**
- **Seleção de compostos com maior potencial de inibição (atracamento)**
- **Caracterização da flexibilidade molecular (dinâmica molecular)**
- **Avaliação do efeito de mudanças na estrutura e ambiente molecular na dinâmica e função de biomoléculas (dinâmica molecular)**

# Processamento em CPUs e GPUs

## CPUs (Central Processing Units – Unidades de Processamento Central):

- Partes integrantes dos computadores responsáveis pelas execuções dos programas (no caso, programas de Bioinformática)
- Atualmente os computadores possuem vários núcleos, chamados de CPUs de processadores múltiplos (*multi-core processing*), o que implica em grande velocidade de processamento

## GPUs (Graphical Processing Units – Unidades de Processamento Gráfico):

- São processadores utilizados para representações gráficas em computadores (localizados nas placas de vídeo dos computadores)
- Possuem centenas a milhares de núcleos de processamento, permitindo uma grande aceleração na manipulação de polígonos e formas geométricas (aplicações 3D)

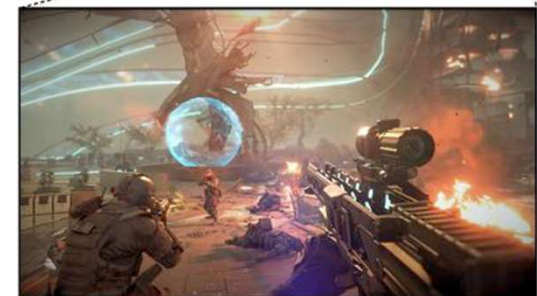
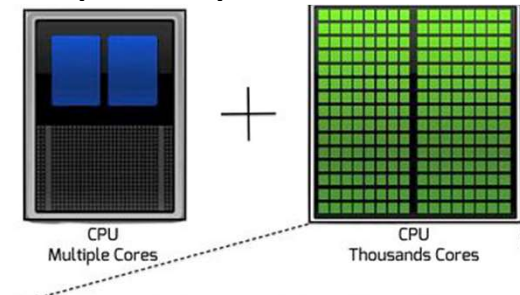
## Aplicativos de Bioinformática que utilizam CPUs e GPUs:

- Alinhamento de sequências à filogenia
- Atracamento molecular à dinâmica molecular

## Determinação da Estrutura de Biomoléculas através da Bioinformática:

- Compreensão de como a natureza evoluiu, como os organismos funcionam, como os processos patológicos se desenvolvem (e podem ser tratados) e como as enzimas exercem suas funções catalíticas

Representação dos núcleos de processamento em CPUs e GPUs. O grande número de núcleos em GPUs permite a realização de cálculos complexos rapidamente



# Importância da Bioinformática

---



## Entendimento de como as proteínas se enovelam

- Construção de novas proteínas, capazes de adotar formas que a natureza não previu até o momento, enzimas aptas a catalizar reações de importância econômica, com menor toxicidade
- Possibilidade de planejamento racional de enzimas e proteínas envolvidas na detoxificação de áreas
- Obtenção de modelos de qualidade próxima àquela de métodos experimentais para proteínas com no mínimo 30% de identidade com outras proteínas de estrutura 3D já determinada
- Refinamento de estruturas cristalográficas utilizando métodos computacionais, agregando explicitamente informações ausentes nos experimentos (como a flexibilidade molecular)
- Construção de alças flexíveis, de difícil observação experimental, mas que podem ser abordadas por diferentes métodos computacionais
- Construção computacional de estruturas 3D de moléculas de RNA
- Uso de estratégias computacionais para lidar com moléculas muito flexíveis (por exemplo, membranas biológicas)
- Simulações computacionais para descrever estruturas de macromoléculas biológicas, não observáveis nos experimentos, para determinar estruturas com resolução atômica
- Uso de métodos computacionais para estudo dos carboidratos (que não parecem sofrer enovelamento nem adotar tipos de estrutura secundária em solução)
- Estimação computacional da estrutura de glicanas com graus variados de complexidade com grande precisão (métodos experimentais possuem grandes dificuldades em abordar)