



**Universidade Federal do Rio Grande do Norte**  
**Centro de Biociências**  
**Departamento de Biofísica e Farmacologia**

**Disciplina de Bioinformática**

**Apostila de Introdução a Bioinformática**

Prof. Alexandre Queiroz  
Fevereiro de 2002

## 1. Biologia e Bioinformática

A bioinformática é um campo das ciências biológicas que está em rápido crescimento e está sendo desenvolvida para atender à necessidade de manipular-se com grandes quantidades de dados genéticos e bioquímicos. Estes dados, originados a partir do esforço individual de vários pesquisadores, estão relacionados entre si através de uma origem comum: as células dos organismos vivos. Para compreender a relação entre estas informações fragmentadas oriundas das diversas áreas da Biologia (tais como Biologia molecular, bioquímica estrutural, enzimologia, Biologia molecular, fisiologia e patologia), a bioinformática usa o poder computacional para catalogar, organizar e estruturar estas informações em uma entidade compreensiva e extremamente importante para a Biologia. Estas entidades são reflexões da organização celular da vida e seu denominador comum que é a evolução dos seres vivos a partir de uma forma ancestral comum.

Devido o fato da bioinformática ser ainda uma ciência recente ainda não se fez outra definição diferente ou mais precisa. Geralmente ela é referida como tendo a tarefa de organizar e analisar dados incrivelmente complexos resultantes de modernas técnicas de Biologia molecular e bioquímica. Para muitos a bioinformática é uma importante ferramenta para a compreensão de como as informações contidas nos genes são refletidas em características fisiológicas, como inteligência, crescimento dos cabelos ou susceptibilidade ao câncer. De uma forma geral, a bioinformática é apresentada como sendo a ciência da criação e manutenção de base de dados (ou Bancos de Dados) e eventualmente a simulação de fenômenos dos organismos vivos. A Bioinformática se baseia na premissa de que existe um relacionamento hierárquico entre as estruturas dos genes, seu arranjo em relação ao genoma, a

função das proteínas e as interações entre as proteínas em um organismo resultando em energia, metabolismo, reprodução e forma.

Os genes são as unidades hereditárias dos seres vivos. Sendo assim, a fidelidade das cópias das informações moleculares contidas nos genes é crucial para a viabilidade dos organismos. Contudo, níveis residuais de mutações são igualmente necessários para sua evolução e enriquecimento dos conjuntos de genes em populações. A partir da manipulação genética é possível expressar as informações contidas nas seqüências de bases de um ou mais genes e daí se comparar às propriedades esperadas com aquelas apresentadas pelas proteínas em relação à forma e função. A tecnologia do DNA recombinante pode então ser utilizada para sintetizar grandes quantidades de proteínas para posterior análise bioquímica e assim se confirmar às formas e funções previstas a partir das seqüências nos genes. A clonagem, o sequenciamento e a localização cromossômica são 3 aspectos bastante inter-relacionados necessários para a compreensão da Biologia molecular de um gene e seu produto, a proteína. A Biologia molecular proporciona a base para a investigação do genótipo através da Bioinformática.

Durante os primeiros anos do século 20, os bioquímicos utilizaram a química orgânica para descobrir e montar vias metabólicas, estudar a cinética das enzimas e determinar as relações entre estas vias e as doenças hereditárias. Vale ressaltar que tudo isto foi realizado sem nenhuma técnica de Biologia molecular, sem nenhuma informação sobre a estrutura das proteínas e em muitos casos sequer se sabia que eram os ácidos nucléicos as unidades hereditárias da vida e não as proteínas. Da mesma forma que a química foi útil para os bioquímicos no passado, e a Biologia molecular esta sendo hoje em dia, a bioinformática, através dos conhecimentos gerados a partir das bases de dados e das simulações também será bastante útil para os biocientistas no futuro.

## 2. A Internet e a Publicação científica

Não só apenas as mudanças nas técnicas bioquímicas influenciaram na produtividade e no sucesso da bioquímica mas, sobretudo, a interação entre os cientistas que através de esforços coletivos ou cooperativos buscaram realmente superar o esmagador e ao mesmo tempo tentador trabalho que havia pela frente. Tradicionalmente a descoberta de vias metabólicas ou o sequenciamento de genes era conduzido por cientistas em trabalhos individuais usando uma abordagem clone após clone.

A ampla divulgação das novas descobertas e desenvolvimento de novas técnicas envolvem o livre acesso a informações que estão sob forte pressão do poder econômico. Para que o fluxo destas informações seja rápido, a Internet aparece como uma ferramenta vital para os cientistas, porque ela permite a proliferação e o compartilhamento de grande quantidade de informações contidas em bancos de dados centralizados. O fluxo de informações científicas através da Internet proporciona 3 grandes vantagens para os cientistas: primeira, proporciona o acesso à informação para qualquer um que possua um computador conectado a Internet e um browser (navegador), sendo esta uma excelente forma de democratização da informação. Em segundo lugar, as informações armazenadas em bancos de dados centralizados apresentam redundância em vários graus. Isto ocorre devido ao fato de que vários pesquisadores podem contribuir com a mesma informação para o banco de dados. Por conseguinte o acesso a estas informações pela Internet proporciona um amplo e eficiente controle de qualidade para as seqüências armazenadas. Em terceiro lugar, as informações armazenadas nos bancos de dados a respeito de um determinado organismo fornecem valiosas informações que poderão ajudar os cientistas a descobrirem e compreenderem várias vias metabólicas de outros organismos através da comparação de genes

homólogos. Por exemplo, na tentativa de curar doenças humanas, muitos genes envolvidos com várias vias metabólicas importantes para estas doenças foram estudados em leveduras. Esta abordagem se fundamenta na relação existente entre todos os seres vivos baseado na evolução. Assim surge uma nova abordagem da Biologia, a genômica comparativa entre espécies lineares.

O lançamento do Projeto Genoma Humano no final dos anos 80 (e muitos projetos genomas de outros organismos) foi um evento decisivo para o desenvolvimento da bioinformática. Os biólogos moleculares que participam do projeto genoma não conseguem ir muito longe na pesquisa em seqüências lineares inteiras devido a limitações técnicas. Sendo assim, eles simplesmente produzem bits de milhares de pedaços pequenos das seqüências de DNA que serão posteriormente montados como um gigantesco quebra-cabeças, dando origem a seqüência genômica completa. A bioinformática é a ferramenta analítica que eles precisam para ajudá-los a montar a seqüência genômica e finalmente revelar a informação biológica relacionada a esta seqüência. A informação genômica obtida é bastante relevante, porque a função de um gene não está apenas na seqüência de codificação para uma proteína, mas também na organização destes genes nos genomas. Os organismos não usam os genes de forma isolada, mas sim em grupos. A compreensão de como o padrão de atividade destes genes é controlado durante o ciclo de vida de um organismo ainda está longe de ser totalmente compreendida pelos biólogos. Cada uma das células de um organismo pluricelular contém um pacote completo de genes, mas apenas uma pequena parte é utilizada pela célula (a maioria dos genes que nunca serão utilizados é inativada). A combinação de genes ativados numa célula define seu destino biológico. Isto é conhecido como diferenciação celular e, normalmente é um processo irreversível durante a vida de uma célula. A genômica funcional é uma consequência direta da informação

acumulada nas seqüências genômicas e da organização destas dentro do genoma. Através da análise dos padrões de atividade dos genes nas células, tecidos e órgãos, se sabe agora a localização precisa no genoma de muitas seqüências relacionadas a importantes questões médicas. Devido ao enorme potencial econômico destes achados científicos, como o desenvolvimento de novas drogas, os interesses comerciais estão fortemente relacionados a estes protejo genoma e as brigas pelos direitos de patentes das seqüências de DNA ainda desconhecidas tem sido comparada à corrida do ouro no século 19.

### 3. Genômica e Proteômica

O esforço dos biólogos moleculares para sequenciar DNA em suas pesquisas nas universidades com interesse em diversos aspectos da Biologia celular e bioquímica resultou em uma coleção aleatória de seqüências de genes distribuídos em vários bancos de dados públicos. A vantagem da abordagem gene a gene é que ela já traz consigo a informação sobre a função associada com o gene. Por outro lado, a abordagem de sequenciar maciçamente genomas completos produz um acúmulo sistemático de seqüências de DNA das quais não se tem nenhum conhecimento sobre sua fisiologia e função, pondo a Biologia molecular de ponta-cabeça. Tradicionalmente os cientistas precisavam fazer uma varredura de linhagens inteiras de células ou tecidos animais para identificar um novo gene (através da técnica de Northern Blot). Agora eles podem fazer essa varredura eletronicamente nos banco de dados públicos para a descoberta de novos genes ou fragmentos de genes (Northern eletrônico). Isto gerou um excelente atalho em relação ao método tradicional de isolamento de mRNA e sequenciamento de proteínas que eram necessários para se obter às seqüências de DNA. A sistemática de detecção e anotação de proteínas

analisadas por gel de eletroforese bidimensional (proteômica) fornece informações relevantes sobre a fisiologia permitindo a detecção de novas e importantes proteínas associadas com o desenvolvimento, com o envelhecimento e com as doenças. Esta informação é fundamentalmente importante para se compreender como a informação genética é lida e implementada no desenvolvimento e no funcionamento de um organismo viável.

Muitos cientistas acreditam que os estudos das seqüências completas obtidas a partir dos modernos projetos de sequenciamento genético, irão contribuir muito para uma melhor compreensão da Biologia dos organismos. Outra contribuição importante está relacionada a grande quantidade de DNA que não é codificado em muitos organismos. Este DNA, até bem pouco tempo era desconhecido e não se sabia sua função. O conhecimento destas seqüências irá indicar, no futuro, o caminho a ser seguido pelos cientistas no sentido desta descoberta. A promessa dos projetos genoma é a compreensão da vida!

Será que o DNA é a única forma de armazenamento de informação que é herdada? Não é tão simples afirmar que toda a informação herdada está unicamente na seqüência de DNA, pois devido o fato do mesmo ser incapaz de se replicar sozinho fora do ambiente celular, se faz necessário que, de alguma forma, o DNA seja lido. Na célula esta leitura é feita por proteínas. A organização cromossômica, a interação e o arranjo dos complexos DNA/proteínas também são parte da informação herdada, não apenas a seqüência do DNA. Os banco de dados estão se expandindo muito rapidamente, alguns são atualizados diariamente para acomodar dados novos e disponibiliza-los para a comunidade científica. De abril de 1988 até o final de 1999 os projetos genoma cobriram 83 espécies com 21 projetos completados (predominantemente microorganismos) e no ano 2000 existiam 62 projetos em andamento. Estes projetos genoma

são realizados através da clonagem automática por PCR (polimerase chain reaction) ou reação em cadeia da polimerase, que amplifica o DNA e pelos sistemas de sequenciamento automático. Para a reconstrução dos espaços vazios (gap-free) e união das seqüências contíguas (contigs) oriundas da fragmentação aleatória dos cromossomos (shot-gun) são utilizados softwares especiais que, realizam esta função em toda a seqüência do genoma base por base.

Inicialmente, a bioinformática havia sido uma colaboração entre diferentes grupos de pesquisas em diferentes países. Hoje a bioinformática está sendo transformada em uma ciência independente graças ao advento dos bancos de dados centralizados, a comunicação via Internet, e aos projetos genoma que impulsionam o incrível aperfeiçoamento das técnicas de clonagem e sequenciamento. Atualmente, a ciência tem sido, considerada um excelente negócio e, conseqüentemente, muitas organizações públicas e privadas têm sido fundadas com o objetivo de sequenciar genomas completos, mapear todos os genes e criar bancos de dados que relacionem às seqüências com a estrutura e função celular.

#### **4. Computação em Biologia e Medicina**

A bioinformática utiliza como ferramentas a matemática aplicada e a computação. A Biologia molecular de hoje seria impossível sem os recursos de bioinformática, tais como o armazenamento, distribuição e atualização das informações, as análises estatísticas, a modelagem de dados e a simulação de fenômenos biológicos em computador. Pesquisas e tratamentos médicos, neurobiologia e o uso de sofisticados equipamentos de laboratório seriam impossíveis sem os computadores. A medicina moderna utiliza muitos equipamentos analíticos e a realidade virtual para ajudar os médicos em seus diagnósticos, como na inserção de sondas miniaturizadas no interior de vasos e na realização de delicadas técnicas de microcirurgias. A neurobiologia está começando a mapear a anatomia cerebral e a composição celular, assim como os projetos genoma estão mapeando os cromossomos. A neuroinformática também é um novo e emergente ramo da bioinformática. Ela surgiu através do esforço colaborativo entre neurologistas, (cientistas que estudam cognição) e psicólogos. O cérebro e os neurônios estão sendo encarados como um sistema complexo que serve de modelo para o desenvolvimento de redes neurais de computadores, devido à forma com que os neurônios trabalham. Algoritmos genéticos e raciocínio não linear atualmente têm sido utilizados para o desenvolvimento de inteligência artificial e evolução computacional. (Veja: The Genetic Algorithms Archive – um arquivo mantido por Alan C. Shultz no Centro para pesquisa aplicada em Inteligência artificial – <http://www.aic.nrl.navy.mil/galist/>).

## 5. Algoritmos Computacionais

Os computadores são essenciais para o processamento de grandes quantidades de dados em tempo hábil para seu estudo. Contudo, os computadores precisam de instruções, processo chamado de intervenção humana. Este processo de instrução analítica humana pode demandar muito tempo para resolver um problema usando o computador. Atualmente muitos dos processos automatizados têm como objetivo ensinar os computadores a tomarem decisões num determinado futuro (como o reconhecimento do contexto de um problema). Sistemas inteligentes são programas que realizam determinadas tarefas que requerem enorme poder computacional. As situações da vida real nunca são totalmente reproduzíveis para os computadores comuns e muitas decisões baseadas atualmente na intervenção humana estão sendo projetadas para serem manipuladas por redes neurais (Neural Networks –NNs) que são sistemas que possuem a habilidade de aprender. As redes neurais embora sejam uma boa promessa para o futuro, ainda são difíceis de se aplicar de forma bem sucedida em problemas relativos a manipulação de símbolos e memória. Além disso não existe um método de treinamento de NNs que permita, de forma mágica, que elas criem alguma informação que não esteja contida nos dados previamente informados (Neural Network FAQ; [ftp://ftp.sas.com/pub/neural/FAQ.html](http://ftp.sas.com/pub/neural/FAQ.html)). Alguns algoritmos podem ser desenvolvidos para funcionarem de forma eficiente em computadores que respondem questões repetitivas onde os dados informados e obtidos variam constantemente e podem ser gerados e ajustados pelo próprio processo como num feedback.

O poder dos computadores é inegável. A facilidade de se escrever um texto em um processador de texto, por exemplo, se tornou uma ferramenta tão popular que hoje em dia pode ser considerada indispensável. Embora a formatação de um texto possa ser feita em questão de

segundos, a facilidade de se mudar o layout do texto e das figuras tem aumentando em grande quantidade o desperdício de papel, simplesmente porque nós ainda queremos ver como vai ficar o produto final! Por outro lado às várias cópias produzidas das várias versões de um texto podem ser economizadas pelas versões eletrônicas vistas no próprio monitor. Corretores ortográficos e gramaticais são bons exemplos de um simples algoritmo que pode fazer uma medida analítica da linguagem. Uma das armadilhas dos corretores ortográficos está na leitura de uma determinada palavra dentro de um contexto, o computador não pode encontrar um erro ortográfico se a palavra estiver correta para um outro contexto (por exemplo: conserto e concerto). A habilidade da mente humana em analisar e reconhecer corretamente, falas, estilos e gramática não é reproduzida satisfatoriamente pelos computadores porque mentes e computadores trabalham diferentemente. O funcionamento da correção ortográfica é muito parecido com a análise de dados científicos e sua subsequente interpretação quando feito pelos computadores, mas apenas sob o restrito controle da mente humana.

Os computadores são excelentes ferramentas para soluções numéricas (análises e simulações), controle de máquinas, edição e busca de caracteres em textos, busca de relação entre dados e gerenciamento de base de dados. As ultimas três aplicações são cruciais para a bioinformática.

## 6. Diferentes tipos de computadores para diferentes tarefas

Os computadores pessoais (PCs) são multifacetados e são usados para várias tarefas como processamento de texto, planilhas de cálculo, apresentações e acesso a Internet. Com softwares especializados também podem controlar equipamentos de laboratório como o pClamp da Axon Instruments (<http://www.axon.com/>) um software aplicativo largamente usado em eletrofisiologia para o controle e medida da atividade elétrica em neurônios, na medida da concentração de íons, para analisar os padrões de hibridização de fragmentos de DNA, para guiar microeletrodos em neurocirurgia funcional e no diagnóstico e monitoramento de distúrbios do movimento (ex.: Mal de Parkinson). A versatilidade, velocidade e o aumento do poder computacional dos PCs em redes locais, têm guiado os cientistas no trabalho com modelagem molecular e alinhamento de múltiplas seqüências de DNA (evolução molecular) independentemente da existência de supercomputadores. Os laboratórios podem ser equipados com várias opções de interfaces com PCs que ajudam nas pesquisas em aplicações que atendam as necessidades específicas dos experimentos.

Estima-se que apenas 1% de todos os microprocessadores no mundo estejam nos PCs. Os outros 99% estão envolvidos em muitos outros produtos espalhados pelo mundo, como aeronaves, sistemas de refrigeração e aquecimento, instrumentos de laboratório, sistemas de segurança e outros dispositivos. Estes processadores são comumente referidos como “firmware”, que são chips que tem como função especial realizar tarefas sem a necessidade de programação.

A ciência requer o uso de muitos equipamentos que usam processadores, como cromatógrafos a gás, balanças computadorizadas e espectrofotômetros. Os espectrofotômetros são usados para ler o espectro de absorção luminosa

de diferentes comprimentos de onda, incluindo medidas em tempo real para o monitoramento de mudanças na composição química de soluções. Os cromatógrafos fazem a separação de misturas moleculares como componentes individuais de acordo com seu tamanho e solubilidade. Estes equipamentos são controlados por microprocessadores construídos especificamente para este fim, porém com interfaces para computadores remotos. Em geral os microprocessadores são acessíveis através de pequenas telas que mostram uma ou mais linhas de código ou comando de texto que pode ser digitado ou selecionado a partir de um menu. Eles funcionam essencialmente como caixas eletrônicas, onde um pequeno teclado pode ser usado para acessar sua conta bancária, você interage com o computador que realiza a transferência de seu dinheiro ou efetua o pagamento de uma conta, mas você não pode editar um texto ou alterar o programa.

## 7. Supercomputadores

Os supercomputadores são usados para tarefas que demandem grande poder computacional e que necessitem grande capacidade de armazenamento e memória. Eles são os servidores principais dos provedores de Internet e a maior parte utiliza o sistema operacional UNIX. Este sistema tem sido implementado em muitas comunidades científicas devido a sua alta capacidade e baixo preço.

Em 20 de junho de 1997 pesquisadores do centro de supercomputadores de Pittsburgh Nos Estados Unidos e a Universidade de Stuttgart na Alemanha conectaram supercomputadores dos dois lados do atlântico via rede de pesquisas de alta velocidade. Esta foi a primeira vez em uma rede de telecomunicação de alta velocidade (Very High Speed Backbone Network Service – vBNS) para transmissão de dados de computador através do atlântico. Considerado como protótipo para a rede internacional de alta performance, um projeto colaborativo entre o processador CRAY T3E com outro processador de alta performance de Stuttgart. O trabalho de computação realizado por dois supercomputadores em diferentes locais de trabalho e conectados entre si é conhecido como “metacomputação”. O link Pittsburgh-Stuttgart criou um sistema virtual de 1024 processadores com a performance teórica de 675 bilhões de cálculos por segundo! O projeto permitiu o desenvolvimento de várias redes de pesquisas através da conexão de alta velocidade entre os dois centros através do atlântico. Cada nova rede estabelecida nos últimos anos permitiu a transmissão de informações a uma velocidade 100 vezes mais rápida que a Internet convencional. Por exemplo, o vBNS que conecta os centro de supercomputadores dos Estados Unidos, atualmente pode transmitir em velocidades em torno de 622 milhões de bits por segundo, velocidade suficiente para transferir toda a Enciclopédia Britânica em menos de 10 segundos.

Finalmente, a Internet é uma rede de supercomputadores e estações de trabalho (PCs) conectados através de switches, roteadores e cabos de fibra óptica. A grande força da Internet é o modo interativo remoto. A maior parte dos aplicativos disponíveis na World Wide Web (www) roda através de supercomputadores remotos dispensando a necessidade de downloads de softwares para análise local. A bioinformática é um exemplo de que a Internet tem sido parte integral da pesquisa científica e que sua necessidade e uso através de programas remotos se tornaram cada vez mais importante.

## 8. Limitações na análise computacional – A promessa paradoxal da Internet.

Quando um experimento é terminado, os dados acumulados precisam ser analisados ou processados. Isto inclui a tabulação de bases de dados, realização de testes estatísticos ou correlacionais e, o mais importante de todos, a seleção dos dados que podem ser analisados e utilizados para interpretação dos fenômenos. Este último é um processo que é independente do computador e evidentemente dependa da confiança do cientista em seus experimentos. O julgamento da qualidade dos dados pode enganar aos olhos do observador, uma vez que o que se espera ver está baseado na hipótese na qual os experimentos foram fundamentados.

A intuição e a interferência dos cientistas são os fatores mais importantes na tomada de decisões corretas. É claro que os computadores podem ajudar, mas apenas com softwares que tenham sido desenvolvidos especificamente para aquela análise, ou seja através da intervenção humana. Enquanto que o processo analítico pode ser feito exclusivamente por computadores, a interpretação dos dados não pode ser feita por um computador por si só. Contudo, computadores podem ajudar a análise de dados avaliando cada



variável de um experimento e lendo os dados de forma mais consistente, o que não poderia ser observado superficialmente pelo cientista.

Nenhum manual técnico, livro ou a Internet podem sobrepor a experiência. O estudo e o ganho de experiência são os fatores mais importantes neste estágio do processo científico. É aqui que a Internet oferece uma importante ajuda para os cientistas. Contudo, as informações encontradas em livros texto, páginas com instruções e protocolos de laboratório não são substitutos para a experiência prática, embora a natureza interativa da Internet possa causar impressão contrária. Atualmente a Internet não é mais interativa que ler um livro em uma biblioteca ou realizar um curso tradicional de longa distância. A razão de insistir neste ponto é que sabemos que a precisão dos dados gerados por computador depende do quanto correto são as informações fornecidas e manuseadas pelo operador humano. Não é nenhum segredo que as bases de dados de DNA e proteínas contêm erros conseqüentes da origem e da forma com que as seqüências foram obtidas. O sucesso da bioinformática está diretamente vinculado a uma anotação completa e confiável das seqüências nas bases de dados e com a precisão dos dados biológicos. Este processo de checagem de precisão das anotações não é automatizado e não será por muito tempo. Para tornar as bases de dados fontes confiáveis de informação, são necessários muitos especialistas para analisar palavra por palavra.

A comparação de seqüências de genes usando programas como o BLAST é razoavelmente fácil, mas a compreensão verdadeira do que os resultados desta comparação significam pode ser extremamente difícil. Os resultados dependem dos tipos de seqüências comparadas e da origem destas seqüências. Ou seja é necessário se compreender a Biologia por trás da seqüência que está armazenada na base de dados e por comparação chegar a alguma conclusão a respeito das novas seqüências. Os projetos genomas

produziram muitas seqüências sem nenhuma função biológica associada. Obviamente esse era seu objetivo. Espera-se que se possam extrair informações ordenadas de modo a orientar novos experimentos que colaborem para desvendar a Biologia por trás destas seqüências.

## **9. A necessidade de Ferramentas computacionais**

Atualmente a Biologia e a medicina são campos multifacetados concentrando um ou mais aspectos da vida. As neurociências concentram-se na Biologia dos neurônios. A bioquímica se detém na química dos organismos vivos. A Biologia molecular estuda as interações biológicas em nível molecular e suas interferências para as células e para o organismo como um todo. A virologia e a bacteriologia concentram-se nos ciclos de vidas de vírus e de bactérias, respectivamente. Muitos outros aspectos da Biologia se concentram em tópicos específicos da Biologia que são importantes para cada campo particular. É indiscutível a importância que cada um destes aspectos da Biologia pode ter para o outro. A grande sobreposição de dados que existe nas informações fornecidas por cada um destes campos agora está se tornando mais óbvia com o avanço nas ferramentas de manipulação de dados e, conseqüentemente, necessita do uso de ferramentas computacionais mais elaboradas.

O incrível avanço na Biologia e na medicina durante as últimas décadas tem criado um feedback positivo, onde cada novo achado serve como guia para o crescimento e popularidade de um determinado campo da Biologia. Isto também gera um crescimento exponencial dos dados biológicos. Desta forma também se faz necessário um poderoso e eficiente sistema de manipulação de dados biológicos, o que parece óbvio, mas sem o poder dos computadores isto seria inimaginável. Os computadores são agora parte integrante do mundo biológico e sem eles, os atuais avanços na Biologia

e na medicina seriam, sem dúvida, impossíveis. A parceria entre estas criaturas não vivas (computadores) e a Biologia criou a necessidade de se fundir certos aspectos destas duas ciências. Novos campos, como a ciência da computação em medicina, a bioengenharia e a Biologia computacional estão em franca ascensão e estão ganhando grande respeito dentro das ciências da vida. O objetivo destes campos é a rápida análise de dados biológicos e a descoberta de informações biológicas desconhecidas. Estas informações podem ser ferramentas úteis em avanços terapêuticos no prolongamento e melhoramento da qualidade de vida.

A introdução da Internet na Biologia foi um fato bastante positivo, tem aumentado dramaticamente a comunicação entre os pesquisadores e reduzido a repetição de trabalhos em várias áreas da pesquisa científica em todo o mundo. A existência de sistemas gerenciadores de dados como o National Center for Biotechnology Information (NCBI - <http://www.ncbi.nlm.nih.gov>) e o European Bioinformatics Institute (EBI - <http://www.ebi.ac.uk/>) tem aumentado a eficiência de muitos pesquisadores em todo o mundo e unindo cientistas oriundos de diferentes disciplinas. O crescimento exponencial dos dados biológicos requer uma organização específica nas bases de dados através de sistemas especializados para cada tipo de informação. Por exemplo, os dados biológicos pertinentes às proteínas precisam ser separados dos polinucleotídeos (DNA e RNA). Os bancos de dados de proteínas (PDB – Protein Data Bank) são exemplos de sistema onde os dados das proteínas são armazenados especificamente como estruturas protéicas. Assim como muitos outros bancos de dados biológicos, o PDB permite a comparação entre diversas moléculas estocadas. No PDB, estas informações podem ser obtidas a partir da classificação estrutural das proteínas. Base de dados como o SCOP (Structural classification of Proteins) são ferramentas úteis na caracterização de macromoléculas em relação as

demais em um sistema biológico. A simples separação de moléculas em categorias específicas não é o suficiente. Os sistemas de gerenciamento de dados devem ser capazes de mostrar a relação existente nas informações contidas nas moléculas de interesse. As informações em um arquivo específico devem conter links para dados relacionados em outros sites relevantes. Por exemplo, no sumário do arquivo PDB da mioglobina de cavalo devem existir várias opções de links para várias informações relacionadas à molécula da mioglobina. Estas informações relativas a um determinado dado permite-nos avaliar outras informações potencialmente úteis sobre outras moléculas em outros sistemas e em outros servidores na Internet.

## 10. Novas descobertas em Pesquisa científica com computadores

Problemas biológicos ainda sem solução são agora os principais temas da Biologia computacional no mundo. A Bioquímica, a Biofísica, a Biologia molecular, a Biologia evolucionária, a Bioinformática, as neurociências e a Farmacologia são apenas alguns dos campos das ciências naturais que têm sido significativamente influenciados pelas ferramentas computacionais. Ao contrário dos fenômenos da física, os fenômenos biológicos, até recentemente, eram considerados imprevisíveis e muitos dos seus aspectos eram considerados indescritíveis. A introdução de ferramentas computacionais na Biologia tem reduzido brutalmente os problemas com o manuseio de dados e mais importante que isso, tem permitido a observação da relação existente entre as moléculas biológicas em seus respectivos campos da Biologia. As novas informações biológicas aliadas com o aumento da nossa habilidade em prever fenômenos biológicos têm reforçado muito o avanço da Biologia. A idéia da Biologia como uma ciência previsível é um estímulo para muitos cientistas, o que tem afastado a visão das ciências biológicas como ficção científica. A sociedade como um todo tem dedicado grande atenção à Biologia e a Medicina nas últimas décadas graças aos avanços ocorridos como a introdução de novas drogas e tratamentos que prolongam e aumentam a qualidade de vida. Estas novidades tem sido instrumento para colocar vários campos da Biologia, como a Bioquímica e a Biologia Molecular no pedestal da ciência

Adaptado de **Bioinformatic Basics – Applications in Biological Science and Medicine**. Hooman H. Rashidi e Buehle. CRC EUA 2000.