

## Bioinformática: Perspectivas na Medicina

### Bioinformatics: Perspectives in Medicine

Glaucius Oliva

Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, São Paulo, SP, Brasil

A medicina moderna requer a integração e análise de grandes quantidades de dados genômicos, moleculares, celulares, associados a informações clínicas, desta forma apresentando desafios fenomenais para a Bioinformática. Esta área interdisciplinar da ciência, que começou com a análise de seqüências gênicas e cresceu na direção do desenvolvimento de técnicas automatizadas de anotação de genomas e transcriptomas, está agora caminhando na direção das áreas emergentes da genômica integrativa e comparada, e buscando, em última instância uma medicina personalizada. Os impactos previstos vão da identificação de milhares de novos alvos terapêuticos para doenças genéticas e moleculares ao florescimento de uma medicina preventiva personalizada, com base no risco genético individual. A Farmacogenômica permitirá a otimização da eficácia e minimização dos efeitos colaterais dos medicamentos. A elucidação da estrutura tridimensional das proteínas humanas e de patógenos permitirá o desenvolvimento de novos medicamentos planejados quimicamente para interagir especificamente com seus alvos. A Terapia Gênica, em conjunto com a tecnologia de manipulação de células-tronco vislumbra a possibilidade de reposição corretiva de tecidos e genes. Esta revisão visita as principais tendências em bioinformática que detêm papel maior na busca de futuras descobertas biológicas e aplicações clínicas.

**Palavras-chave:** bioinformática, proteínas, genômica, biologia estrutural, farmacogenômica, SNP.

*Modern medicine requires the integration and analysis of large amounts of data, coming from genomic, molecular, cellular and clinical information, thus demanding a remarkable set of challenges to Bioinformatics. The evolution of this interdisciplinary area, which started with sequence analysis and has led to high-throughput whole genome or transcriptome annotation today, is now going to be directed towards recently emerging areas of integrative and translational genomics, and ultimately personalized medicine. The foreseen impacts cover from the identification of thousands of new therapeutic targets for genetic and molecular diseases to the flourishing of a new personalized medicine, based on the individual genetic risk. Pharmacogenomics will allow the optimization of the efficacy and minimization of side effects of drugs. The elucidation of the tridimensional structure of human and pathogen proteins should lead to the development of new drugs, designed to specifically interact with their targets. Gene therapy, in conjunction with stem cell Technologies opens the pathway to corrective reposition of genes. This review outlines the most promising trends in bioinformatics, which may play a major role in the pursuit of future biological discoveries and medical applications.*

**Key words:** Bioinformatics, proteins, genomics, structural biology, pharmacogenomics, SNP.

O conceito de Bioinformática é muito amplo e sua definição mais geral engloba qualquer análise de dados biológicos utilizando-se métodos computacionais. No entanto o entendimento usual desta moderna área do conhecimento está na utilização das ciências da informação e da implementação computacional de técnicas físicas e estatísticas aplicadas ao estudo dos sistemas biológicos, desde o nível molecular (p.ex. DNA, RNA, proteínas) até o dos organismos e sistemas complexos. Neste artigo serão predominantemente abordados os impactos do vertiginoso crescimento das informações genômicas e moleculares, na medicina do presente e do futuro. Neste contexto a

bioinformática está focada em três tipos de bancos de dados: seqüências genômicas, estruturas macromoleculares e experimentos de genômica funcional (p.ex. análises de transcrição gênica, proteômica, dados de interações entre genes expressos, etc.). No entanto a Bioinformática pode ser também aplicada na análise de outros tipos de dados como árvores filogenéticas, relações entre vias metabólicas, correlação com dados estatísticos sobre pacientes, populações e suas patologias entre outros. Um amplo espectro de técnicas é utilizado, como alinhamentos de seqüência primária, superposições entre estruturas tridimensionais de proteínas, construção de árvores filogenéticas, predição, modelagem e classificação de estruturas protéicas e de RNA, predição de função protéica e agrupamento de dados de expressão gênica. Esta plethora de dados e análises tem modificado de forma explosiva a forma como a medicina moderna é exercida, com amplo impacto no bem estar e qualidade de vida da sociedade humana.

Recebido em 20/12/2007

Aceito em 25/01/2008

Endereço para correspondência: Dr. Glaucius Oliva, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, C.P.369. 13560-970. São Carlos, SP, Brasil. Endereço eletrônico: oliva@ifsc.usp.br. Apoio: FAPESP, CNPq.

Gazeta Médica da Bahia

2008;78 (Suplemento 1):52-58.

© 2008 Gazeta Médica da Bahia. Todos os direitos reservados.

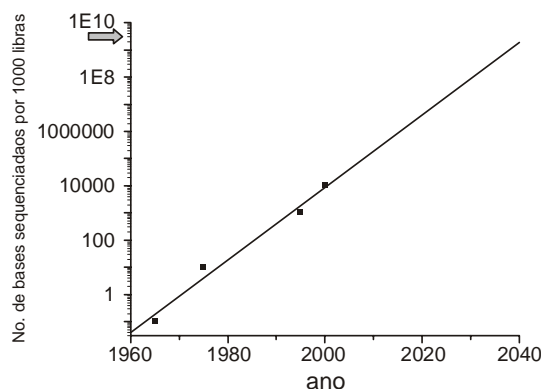
### A Explosão do Conhecimento Genômico e Molecular

Desde a seminal elucidação da estrutura em dupla-hélice do DNA por James D. Watson e Francis H.C. Crick em 1953, que abriu o caminho para a compreensão dos mecanismos de armazenamento, transcrição, tradução e replicação da informação genética, o crescimento do conhecimento genômico e molecular dos organismos vivos tem sido explosivo. Em dezembro de 2007 o banco de dados de nucleotídeos mantido pelo European Molecular Biology Laboratory (EMBL) registra mais de 190 bilhões de pares de base em 108 milhões de seqüências depositadas (ver <http://www3.ebi.ac.uk/Services/DBStats/> para atualizações diárias). Este esforço internacional tem resultado no seqüenciamento de genomas completos de 51 organismos do domínio Archea, 569 de Bacteria e 73 de Eukariota (<http://www.ebi.ac.uk/genomes/index.html>) que inclui, entre outros, o genoma humano, de chimpanzé, canino, bovino, murino e de diversos patógenos humanos. O custo por base seqüenciada tem caído de forma vertiginosa. O geneticista e escritor britânico Richard Dawkins, em seu livro *A Devil's Chaplain* (DAWKINS, 2004), faz uma análise retrospectiva do número de bases de DNA seqüenciadas por £ 1.000, em dados históricos dos últimos 40 anos, concluindo que em média, a cada 27 meses o custo é reduzido à metade. Como ilustrado na Figura 1, é possível estimar que no ano de 2050 deva ser possível obter um genoma pessoal ao custo de aproximadamente £ 100 ou cerca de R\$ 400,00 ! Esta perspectiva, somada à compreensão de como os polimorfismos na seqüência de nucleotídeos pode afetar a função biológica correspondente, fortemente sugere impactos antes inimagináveis na Medicina do futuro. Os impactos previstos vão da identificação de milhares de novos alvos terapêuticos para doenças genéticas e moleculares ao florescimento de uma medicina preventiva personalizada, com base no risco genético individual. A Farmacogenômica permitirá a otimização da eficácia e minimização dos efeitos colaterais dos medicamentos. A elucidação da estrutura tridimensional das proteínas humanas e de patógenos permitirá o desenvolvimento de novos medicamentos planejados quimicamente para interagir especificamente com seus alvos. A Terapia Gênica, em conjunto com a tecnologia de manipulação de células-tronco vislumbra a possibilidade de reposição corretiva de tecidos e genes.

Naturalmente, o seqüenciamento de um Genoma é apenas o primeiro passo nas ciências genômicas, necessariamente acompanhado de intensa atividade comparativa e experimental que permita a identificação dos genes ali codificados e suas funções (Genômica Funcional). Como cada gene de um organismo somente é transcrito em localizações, momentos e quantidades específicas, os estudos de Transcriptômica buscam determinar os padrões de transcrição gênica no espaço e no tempo, pela quantificação dos RNA mensageiros em diferentes tecidos e condições fisiológicas.

A tradução da informação contida nas moléculas de RNA mensageiro se dá pela síntese protéica ao nível ribossomal,

**Figura 1.** Custo de seqüenciamento do DNA. Número de bases de DNA seqüenciadas por £1.000, com 4 dados históricos dos últimos 40 anos, e extrapolação linear futura. (Adaptado por R.C.Garratt com base na figura original em Dawkins, R. "A devil's chaplain") (DAWKINS, 2004).



Tempo para reduzir o custo pela metade = 27 meses  
Custo de um genoma pessoal em 2050 ~ R\$ 400,00

modulada por multivariados fatores físico-químicos celulares, levando a um padrão definido de proteínas presentes nas células em cada instante de tempo, foco de atenção da Proteômica.

O que tem ficado evidente nos diversos projetos genoma concluídos, é que a complexidade de um organismo não está determinada pelo número de genes codificados em seu genoma, mas na expressão concertada de genes, sistêmica ou regulada por estímulos ambientais. As pesquisas na área de Interactômica visam determinar a existência de interações entre proteínas correlacionadas funcionalmente.

Em última instância é essencial entender como as proteínas agem, pela elucidação de suas estruturas moleculares tridimensionais e a relação entre a estrutura e a função biológica, que é o objetivo dos programas de Genômica Estrutural. Com estas informações pode-se então abordar, de forma completa, o complexo problema de identificar alvos macromoleculares de importância médica, que possam ser utilizados na descoberta de novos fármacos.

### A Compreensão Estrutural dos Determinantes Genéticos

As proteínas, fenótipos principais dos genes e responsáveis pela maioria das atividades fundamentais dos organismos, são macromoléculas de peso molecular na ampla faixa de  $10^4$  a  $10^6$  unidades de massa atômica, constituindo-se em polímeros formados pela ligação covalente entre os 20 aminoácidos diferentes utilizados universalmente pelos organismos vivos em nosso planeta. Como característica fundamental para sua funcionalidade, a atividade biológica das proteínas depende criticamente de sua estrutura tridimensional. Técnicas físicas experimentais como a cristalografia de proteínas e a ressonância magnética nuclear, têm permitido conhecer a estrutura 3D de mais de 48.000 proteínas, registradas no Protein Data Bank (<http://www.rcsb.org>) em dezembro de 2007. Este

espetacular avanço da Bioinformática Estrutural nos últimos anos tem nos ajudado a compreender de forma mais abrangente a hierarquia modular da organização das macromoléculas biológicas. Como a pressão da seleção natural atua primeiramente na função biológica, é natural e de fato observado que as estruturas tridimensionais protéicas se conservam mais do que as seqüências de proteínas e genes codantes. Assim, as proteínas podem ser organizadas na forma de um número limitado de famílias estruturais, recentemente descritas de forma sistematizada na tabela periódica das proteínas (*The Protein Chart*) (GARRATT, R.C.; ORENGO, C.A., 2007), que permite uma visualização ao mesmo tempo abrangente e compacta dos princípios que dirigem a diversidade de estruturas protéicas.

Uma importante fronteira da moderna pesquisa em Bioinformática Estrutural está no desenvolvimento de metodologias teóricas que permitam a modelagem molecular das proteínas, a partir de suas seqüências de aminoácidos. Nos casos em que há significativa homologia com outras proteínas com estrutura 3D conhecida, pode-se fazer a modelagem molecular teórica para propor modelos para a proteína de interesse, para o que há inúmeras ferramentas computacionais disponíveis (EDWARDS, Y.J.K.; COTTAGE, A., 2003). Vários grupos tem avançado no desenvolvimento de ferramentas automáticas para a modelagem da estrutura tridimensional de proteínas, como o SWISS MODEL (SCHWEDE, 2003) e o GENO3D (COMBET, 2002), ainda que com confiabilidade limitada. No entanto, com o explosivo aumento no número de estruturas conhecidas experimentalmente, pode-se prever que em futuro não muito distante, será possível modelar com precisão razoável a estrutura da grande maioria das proteínas, a partir exclusivamente de suas seqüências de aminoácidos.

### **Fármacos Planejados a Partir da Estrutura 3D de Alvos Terapêuticos**

Nos casos em que proteínas são associadas a doenças moleculares ou a patógenos, a elucidação de sua estrutura permite o desenvolvimento de *fármacos inteligentes*, planejados de forma a interagir especificamente com seu receptor, maximizando a afinidade e minimizando os efeitos colaterais. Exemplos recentes de sucesso desta abordagem são os inibidores da protease do vírus HIV (KIM, 1995; KALDOR, 1997; WLODAWER, 2002), os fármacos antiinflamatórios inibidores específicos da enzima ciclo-oxigenase 2 (COX2) (KURUMBAIL, 1996; FLOWER, 2003), o antigripal Relenza™, inibidor específico da enzima viral neuraminidase (VON ITZSTEIN, 1993) e o anticancerígeno Glivec™, inibidor específico da proteína tirosina quinase (SCHINDLER, 2000; CAPDEVILLE, 2002).

No entanto a batalha pela saúde prossegue, com muitas áreas em que há poucas ou nenhuma alternativa terapêutica como câncer, doenças do envelhecimento e reações auto-imunes. Além disso, o aparecimento de novos e mais poderosos agentes infecciosos e o desenvolvimento de

resistência por microorganismos e vírus que se supunha controlados, claramente demonstram a permanente necessidade de busca por novos medicamentos. O mecanismo de ação de medicamentos envolve a interação, preferencialmente específica, entre os princípios ativos (geralmente moléculas pequenas) e receptores biológicos macromoleculares. O paradigma central de um bom fármaco é que este deve interagir especificamente com seus receptores, pela complementaridade estrutural entre o arranjo espacial dos grupos funcionais da micromolécula com o sítio receptor da macromolécula (VERLINDE, C.L.M.J.; HOL, W.G.J., 1994; BARREIRO, E.J.; FRAGA, C.A.M., 2001).

Desta forma, o atual estado da arte na descoberta de novos fármacos requer primeiramente a identificação de alvos macromoleculares relevantes para uma determinada condição patológica, seguida da elucidação de suas estruturas tridimensionais, de forma a permitir a aplicação da tecnologia do planejamento racional. Como a esmagadora maioria dos alvos são proteínas, fica evidente que encontrar os genes que codificam por proteínas-alvo é uma ferramenta poderosíssima e central na farmacologia moderna.

Assim, todo programa moderno de descoberta de novos medicamentos inicia-se pela seleção de proteínas medicamento relevantes e sua validação como alvos terapêuticos. A Validação de Alvos é hoje uma atividade de pesquisa central na área do desenvolvimento farmacêutico, pois leva à redução do número de potenciais alvos (p.ex., todas as proteínas humanas ou de um dado parasita) para um número mais realístico relacionado ao estudo de uma doença em particular, permitindo assim iniciar um programa de planejamento racional de novos fármacos baseado em estruturas. A descoberta de novos fármacos é um empreendimento de alto custo e longa duração. É muito mais racional certificar-se amplamente que o alvo a ser escolhido de fato tem relação com a doença em foco e que interferir com ele deve efetivamente produzir os resultados biológicos ou terapêuticos esperados. Desta forma há intensa atividade de desenvolvimentos tecnológicos que possam acelerar os estágios iniciais de identificação e validação de alvos, minimizando o uso de recursos materiais e humanos desnecessariamente. Procura-se demonstrar a correlação entre uma dada classe de proteínas, genes ou vias metabólicas com uma doença humana específica, o mesmo valendo para doenças de animais ou plantas. Um estudo recente estimou que a validação de alvos representa hoje aproximadamente um quinto do custo e tempo envolvido em descoberta e desenvolvimento de novos fármacos (TOLLMAN, 2002), indicando que esta etapa representa um gargalo na indústria farmacêutica. A escolha de alvos moleculares terapêuticos adequados no processo de descoberta é uma vantagem competitiva extremamente importante em um setor caracterizado pelo alto risco e dependente de inovações constantes.

Para identificar um gene como tendo importância farmacológica, é preciso determinar sua função, de forma que as técnicas experimentais para a validação de alvos são

frequentemente aquelas utilizadas em Genômica Funcional. Entre as abordagens importantes ao problema destaca-se a Bioinformática, reconhecida como uma ferramenta essencial na análise de seqüências genômicas completas para a identificação e extração das seqüências codantes e correspondentes proteínas e que permite a comparação múltipla de seqüências entre diversas espécies. Atenção especial tem sido dada à genômica comparativa, na qual genomas completos de espécies próximas e/ou distantes podem indicar genes conservados de relevância funcional. Por exemplo, a análise do genoma de linhagens virulentas e não-virulentas de uma dada espécie bacteriana pode indicar os potenciais genes associados à sua patogenicidade (ROSAMOND, J.; ALLSOP, A., 2000). O tamanho e a complexidade dos bancos de dados de seqüências de ácidos nucléicos e proteínas requer o constante desenvolvimento de novas e poderosas ferramentas de software e hardware para o garimpo de informações.

Outra abordagem importante para a identificação de potenciais alvos terapêuticos são os estudos comparativos de análise de expressão gênica, com o uso de micro-arranjos de cDNA (PEROU, 2000), análise serial de expressão gênica (SAGE) (HERMEKING, H., 2003) e particularmente as tecnologias proteômicas de eletroforese multidimensional e espectrometria de massa (WANG, J.H.; HEWICK, R.M., 1999), que permitem identificar quais genes e proteínas estão sendo diferencialmente expressos. Como exemplo, pode-se buscar os genes mais ou menos expressos em tecidos tumorais ou que apresentem uma determinada característica patológica. Outra possibilidade é analisar os genes ativados ou desativados durante a transformação de parasitas do estágio não-infectivo para infectivo.

Há vários estudos funcionais *in vitro* e *in vivo* que tem sido utilizados em larga escala na identificação e/ou validação de alvos, frequentemente em contextos celulares e bioquímicos definidos. É o caso do sistema de duplo-híbrido em levedura, que pode indicar proteínas que interagem entre si (FIELDS, S.; SONG, O., 1989) e que, portanto, requerem ação conjugada para sua função. Outro exemplo é a clonagem por complementação, na qual linhagens celulares deficientes de genes metabolicamente importantes são complementadas por genes exógenos (TANAKA, 1990).

Uma técnica que tem tido grande importância na validação de alvos potencialmente interessantes é a adição ou remoção seletiva de genes (*knockins* e *knockouts*, respectivamente) seguida de observações sobre viabilidade do organismo transgênico e do fenótipo resultante naqueles que sobrevivem (GORDON, J.W.; RUDDLE, F.H., 1981; HARRIS, 2001). Quando o gene está presente em um genoma em cópia única, este processo é facilitado. Abordagens de deleção sistemática de genes em larga escala, pelo uso de técnicas como transposons e mutagênese insercional aleatória tem permitido a identificação de vários genes de interesse (ZAMBROWICZ, 1998).

Mais recentemente, grande interesse tem sido despertado com as técnicas de RNA-antisense e RNA de interferência,

que envolvem a utilização de oligonucleotídeos que são complementares a uma dada porção de uma molécula de mRNA e portanto capazes de inibir sua translação, splicing e transcrição (DEAN, 2001; HEASMAN, 2002). Na técnica de RNA de interferência (*RNAi*), um método de validação de alvos atual e muito popular, utiliza-se RNA de dupla fita (*dsRNA*), que quando introduzido em células é digerido a fragmentos curtos de pequenos RNA's de interferência (*siRNA*) que são replicados por uma RNA polimerase. Esta técnica tem sido aplicada com sucesso no silenciamento de genes de nematóides e moscas e tem amplo potencial de aplicação na validação de alvos terapêuticos (FJOSE, 2001; BRUMMELKAMP, T.R.; BERNARDS, R.; AGAMI, R., 2002)

Além de apresentar correlação causal com a ocorrência de doenças, um bom alvo terapêutico para o desenvolvimento de fármacos deve ter propriedades características que permitam a ligação de moléculas pequenas adequadas à farmacologia, ou seja, de baixo peso molecular e balanço equilibrado de polaridade (para viabilizar a solubilidade e ligação ao receptor) e lipofilicidade (para proporcionar a absorção através de membranas lipídicas). Desta forma, o conjunto de alvos para fármacos é a intersecção entre o genoma drogável (= todos os genes que codificam por proteínas potencialmente capazes de ligar moléculas pequenas) e do conjunto dos genes correlacionados a doenças. Estimativas recentes sugerem que no genoma humano espera-se encontrar de 600 a 1.500 bons alvos para fármacos (HOPKINS, A.L.; GROOM, C.R., 2002).

### O Desafio das Doenças Infecciosas Tropicais

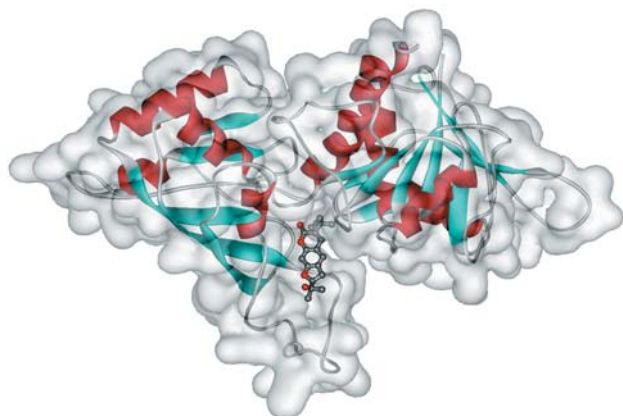
As doenças infecciosas tropicais continuam a afetar seriamente o desenvolvimento social e econômico dos países mais pobres, afetando de forma desproporcional as populações marginalizadas. Estudos recentes indicam que nos países de baixa renda as doenças infecciosas representam até 45% das causas de morte registradas (WHO, 1999). A doença de Chagas, causada pelo protozoário *Trypanosoma cruzi*, afeta particularmente os países das Américas Central e do Sul, onde ainda exerce forte impacto negativo sobre a saúde e qualidade de vida das populações mais pobres. Estima-se que haja 16 a 18 milhões de infectados com o *T. cruzi* e cerca de 120 milhões de pessoas sob risco de contrair a doença. Além da mortalidade, os infectados pelo parasita sofrem de severos problemas cardíacos, gastrointestinais e mesmo dano neurológico, resultando em impacto estimado de 580 mil DALYs (*Disability Adjusted Life Years*, uma medida quantitativa do impacto da doença na perda de anos de vida saudável). Os tratamentos quimioterápicos disponíveis, nifurtimox (Lampit) e benznidazol (Rochagan), apresentam baixa eficácia e alta toxicidade, razão pela qual o nifurtimox não é comercializado no Brasil, Argentina, Chile e Uruguai há alguns anos (FAIRLAMB, 1999). Além disso, tanto o benznidazol como o nifurtimox apresentam sérios efeitos colaterais como: hiporexia, perda de peso, náuseas, vômitos,

alergia cutânea e neuropatia periférica e tem eficácia duvidosa, uma vez que algumas cepas estudadas apresentam susceptibilidade diferente para esse tipo de droga (CASTRO, J.A.; DE MECCA, M.M.; BARTEL, L.C., 2006).

Uma diferença fundamental entre tripanossomatídeos e o hospedeiro humano é a compartimentalização metabólica da glicólise, com nove das enzimas glicolíticas seqüestradas em uma organela denominada glicosomo (MICHELS, 2006). Esta diferença resulta em diferenças estruturais significativas com as enzimas homólogas humanas, tornando-se, portanto, alvos atrativos para o desenvolvimento de inibidores específicos. Estudos experimentais e de simulação computacional do controle do fluxo metabólico na via glicolítica tem demonstrado que os alvos preferenciais para inibição são o transportador de glicose e as enzimas aldolase e gliceraldeído-3-fosfato desidrogenase (GAPDH) (ALBERT, 2005). Desde a elucidação da estrutura da TcGAPDH em nosso laboratório (SOUZA, 1998), temos desenvolvido diversos estudos de busca por inibidores em produtos naturais (TOMAZELA, 2000; MORAES, 2003), estruturas de complexos proteína-ligante (PAVÃO, 2002; CASTILHO, 2003), estudos computacionais quantitativos da relação estrutura-atividade (QSAR) (LEITÃO, 2004), planejamento e desenvolvimento de análogos (DE MARCHI, 2004; ALVIM, 2005).

Um exemplo é o complexo da enzima com o produto natural *chalepina*, uma cumarina extraída da árvore de mata atlântica *Pilocarpus spicatus*, elucidado por cristalografia de raios-X (CASTILHO, 2003) e representado na Figura 2. Com base nesta estrutura, diversos análogos foram planejados e sintetizados, buscando aperfeiçoar a afinidade e especificidade dos ligantes (ALVIM, 2005; COLLINS, F.S.; BROOKS, L.D.; CHAKRAVARTI, A., 1998), maximizando assim seu potencial farmacológico. Estas metodologias tem, portanto, criado perspectivas promissoras de descoberta de novas alternativas quimioterápicas para o tratamento da doença de Chagas.

**Figura 2.** Estrutura cristalográfica do complexo entre a enzima gliceraldeído-3-fosfato desidrogenase de *Trypanosoma cruzi* e o inibidor cumarínico *chalepina*, extraído da planta *Pilocarpus spicatus*.



## SNPs e Diferenças Genéticas-Estruturais

O tipo mais comum de variações genéticas entre os humanos é o Polimorfismo de Nucleotídeo Único (SNP, do inglês Single nucleotide polymorphism), que representam cerca de 90% das diferenças de seqüência do código genético entre indivíduos (COLLINS, F.S.; BROOKS, L.D.; CHAKRAVARTI, A., 1998). Os estudos realizados até o presente momento indicam que há cerca de 15 milhões de posições no genoma humano nas quais pode haver diferenças entre duas pessoas ou populações. Em meados de 2007 mais de três milhões destas posições tem sido mapeadas e catalogadas no *HapMap* (<http://www.hapmap.org>), uma base de dados internacional publica dedicada a auxiliar os cientistas a encontrar genes associados com doenças humanas e respostas diferenciadas a fármacos. Em sua já tradicional edição de final de ano, a conceituada revista científica Americana *Science* escolheu como maior avanço científico de 2007 os estudos da Variação Genética Humana (PENNISI, 2007). Nestes estudos tem sido utilizados chips de DNA que podem analisar cerca de 500 mil SNPs por vez, em amostras genômicas de centenas ou milhares de pessoas com e sem uma doença em particular. Pela correlação entre SNPs e sintomas, pode-se determinar o grau de susceptibilidade de desenvolvimento de uma determinada patologia associado a cada uma destas diferenças genéticas.

Um grande desafio nesta área é entender *como e quando* estas variações causam as doenças a elas relacionadas, um passo essencial para permitir avançar na sua prevenção ou cura. SNPs em regiões codantes (cSNPs) ou regulatórias são as mais prováveis de afetar a função gênica (SYVANEN, 1999). Regiões gênicas amplamente estudadas em diferentes indivíduos tem mostrado que cerca de metade dos cSNPs são silenciosas e o restante causam mutações do tipo *missense*, ou seja, que alteram um códon fazendo com que este codifique um aminoácido diferente (CARGILL, 1999). Alguns destes *missense* cSNPs são neutros e não causam alterações detectáveis no fenótipo gênico, enquanto outras podem resultar em diferenças em características individuais, diferentes respostas a fármacos ou mesmo susceptibilidades distintas a doenças.

Uma maneira de abordar esta questão é mapear as mudanças genéticas nas correspondentes estruturas protéicas como codificadas pelos genes, e buscar entender o efeito produzido, experimental ou teoricamente (por técnicas de modelagem molecular). O grande volume de dados experimentais sobre o efeito de mutagênese sítio-dirigida em estrutura e função de proteínas torna possível prever o possível efeito de uma mutação do tipo *missense* cSNP resultar em alterações na função molecular. No modelo proposto por Wang e Moulton (2001) e implementado de forma automatizada no servidor <http://www.SNPS3D.org> (YUE, P.; MELAMUD, E.; MOULT, J., 2006), cada mutação é associada ao seu impacto sobre um ou mais papéis que o dado resíduo de aminoácido exerce na estrutura protéica, como estabilidade no

enovelamento, conexão com ligantes, catálise, regulação catalítica direta ou alostérica, ou modificação pós-traducional, entre outras. Algumas mutações podem ser consideradas conservativas e neutras, enquanto outras produzem mudanças no caráter hidrofóbico, interações de van der Waals, ligações eletrostáticas, interações de hidrogênio ou pontes di-sulfeto.

O resultado destas análises tem indicado que 90% das mutações do tipo *missense* conhecidamente causativas de doenças e que foram analisadas por este modelo computacional, afetam predominantemente a *estabilidade* das proteínas, através de variados fatores energéticos (YUE, P.; LI, Z.; MOULT, J., 2005). Surpreendentemente, mais de 70% das variações genéticas humanas são neutras, o remanescente estando potencialmente envolvidas em doenças poligênicas.

### Perspectivas Futuras

Diversos estudos tem se proposto a avaliar o potencial impacto futuro da Bioinformática (BAYAT, 2002). O volume sem precedentes de dados biológicos e clínicos que tem sido produzidos desde a disseminação explosiva das ciências genômicas tem demandado interpretação e análise somente possíveis com as técnicas em evolução da Bioinformática. Uma das perspectivas futuras de curto prazo é que as análises da bioinformática vão identificar mais genes associados a doenças e simultaneamente permitir o desenvolvimento de novos fármacos planejados para o seu tratamento. A Bioinformática deverá também auxiliar na identificação de genes de susceptibilidade e elucidar as vias patogênicas envolvidas nas doenças, desta forma oferecendo novas oportunidades para o desenvolvimento de tratamentos mais específicos e eficientes. Por exemplo, alvos potenciais envolvidos em diferentes tipos de câncer têm sido recentemente identificados através da análise de perfis de expressão gênica (GRAEBER, T.G.; EISENBERG, D., 2001).

Em uma perspectiva temporal de maior prazo, as análises bioinformáticas integrando dados genômicos, moleculares, patológicos e clínicos deverão revelar potenciais reações adversas de fármacos em indivíduos específicos, através de testes genéticos simples. Em última instância a farmacogenômica permitirá utilizar a informação genética pessoal no tratamento das doenças, trazendo à tona uma nova era de medicina personalizada, na qual pacientes poderão levar, em um "cartão genético" todas as informações sobre seu próprio perfil genético para a eficácia e efeitos colaterais de certos fármacos.

Como resultado destes avanços, a Bioinformática deverá aproximar os biólogos moleculares e pesquisadores clínicos, capitalizando na ampla disponibilidade de dados e nas ferramentas da biologia computacional. Os médicos do futuro estão sendo formados agora, e certamente aqueles que terão mais sucesso são os que transitarem com igual frequência e facilidade entre a prática clínica e o laboratório de pesquisas, apoiados pelo uso extensivo de ferramentas computacionais.

### Agradecimento

O autor agradece fortemente ao Dr. R.C.Garratt pelas prazerosas e produtivas discussões sobre o tema deste artigo, e pela gentil cessão da imagem que compõe a Figura 1 deste texto.

### Referências

1. ALBERT, M.A. *et al.* Experimental and *in silico* analyses of glycolytic flux control in bloodstream form *Trypanosoma brucei*. *J Biol Chem*, 280: 28306-28315, 2005.
2. ALVIM, J. *et al.* Preparation and evaluation of a coumarin library towards the inhibitory activity of the enzyme gGAPDH from *Trypanosoma cruzi*. *J Braz Chem Soc*, 16: 763-773, 2005.
3. BARREIRO, E.J.; FRAGA, C.A.M. Química Medicinal: as bases moleculares da ação dos fármacos. 1ª Edição, Porto Alegre: Artmed Editora, 2001.
4. BAYAT, A. Science, medicine, and the future: Bioinformatics. *BMJ*, 324:1018-1022, 2002.
5. BRUMMELKAMP, T.R.; BERNARDS, R.; AGAMI, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science*, 296: 550-553, 2002.
6. CAPDEVILLE, R. *et al.* Gleevec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat Rev Drug Discov*, 1: 493-502, 2002.
7. CARGILL, M. *et al.* Characterization of single nucleotide polymorphisms in coding regions of human genes. *Nat Genet*, 22:231-238, 1999.
8. CASTILHO, M.S. *et al.*, Evidence for the two phosphate binding sites of an analogue of the thioacyl intermediate for the *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase-catalyzed reaction, from its crystal structure. *Biochemistry*, 42: 7143-7151, 2003.
9. CASTRO, J.A.; DE MECCA, M.M.; BARTEL, L.C. Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). *Hum Exp Toxicol*, 25: 471-479, 2006.
10. COLLINS, F.S.; BROOKS, L.D.; CHAKRAVARTI, A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res*, 8:1229-1231, 1998.
11. COMBET, C. *et al.* Geno3D: Automatic comparative molecular modelling of protein. *Bioinformatics*, 18: 213-214, 2002.
12. DAWKINS, R. A devil's chaplain: reflections on hope, lies, science and love", A Mariner Book, Houghton Mifflin Company, Boston - New York, 2004.
13. DE MARCHI, A.A. *et al.* New 3-piperonylcoumarins as inhibitors of glycosomal glyceraldehyde-3-phosphate dehydrogenase (gGAPDH) from *Trypanosoma cruzi*. *Bioorg Med Chem Lett*, 12: 4823-4833, 2004.
14. DEAN, N.M. Functional genomics and target validation approaches using antisense oligonucleotide technology. *Curr Opin Biotechnol*, 12: 622-625, 2001.
15. EDWARDS, Y.J.K.; COTTAGE, A. Bioinformatics methods to predict protein structure and function - A practical approach. *Mol Biotechnol*, 23: 139-166, 2003.
16. FAIRLAMB, A.H. Future prospects for the chemotherapy of Chagas' disease. *Medicina (B Aires)*, 59: 179-187, 1999.
17. FIELDS, S.; SONG, O. A novel genetic system to detect protein-protein interactions. *Nature*, 340: 245-246, 1989.
18. FJOSE, A. *et al.* RNA interference: mechanisms and applications. *Biotechnol Annu Rev*, 7: 31-57, 2001.
19. FLOWER, R.J. The development of COX2 inhibitors *Nat Rev Drug Discov* 2, 179-191 (2003).
20. GARRATT, R.C.; ORENGO, C.A. The Protein Chart, Wiley-VCH Verlag, Weinheim, 2007.
21. GORDON, J.W.; RUDDLE, F.H. Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science*, 21, 1244-1246, 1981.

22. GRAEBER, T.G.; EISENBERG, D. Bioinformatic identification of potential autocrine signaling loops in cancers from gene expression profiles. *Nat Genet*, 29:295-300, 2001.
23. HARRIS, S. Transgenic knockouts as part of highthroughput, evidence-based target selection and validation strategies. *Drug Discov Today*, 6: 628-636, 2001.
24. HEASMAN, J. Morpholino oligos: making sense of antisense? *Dev Biol*, 243: 209-214, 2002.
25. HERMEKING, H. Serial analysis of gene expression and cancer. *Curr Opin Oncol*, 15: 44-49, 2003.
26. HOPKINS, A.L.; GROOM, C.R. The druggable genome. *Nat Rev Drug Discov*, 1: 727-730, 2002.
27. KALDOR, S.W. *et al.* Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J Med Chem*, 40:3979-3985, 1997.
28. KIM, E.E. *et al.* Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *J Am Chem Soc*, 117:1181-1182, 1995.
29. KURUMBAIL, R.G. *et al.* Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature*, 384: 644-648, 1996.
30. LEITÃO, A. *et al.* Structure-activity relationships of novel inhibitors of glyceraldehyde-3-phosphate dehydrogenase. *Bioorg Med Chem Lett*, 14: 2199-2204, 2004.
31. MICHELS, P.A. *et al.* Metabolic functions of glycosornes in trypanosomatids. *Biochim Biophys Acta*, 1763: 1463-1477, 2006.
32. MORAES, V.R.D. *et al.* Enzymatic inhibition studies of selected flavonoids and chemosystematic significance of polymethoxylated flavonoids and quinoline alkaloids in *Neoraputia* (Rutaceae). *J Braz Chem Soc*, 14: 380-387, 2003.
33. PAVÃO, F. *et al.* Structure of *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase complexed with chalepin, a natural product inhibitor, at 1.95 angstrom resolution. *FEBS Lett*, 520: 13-17, 2002.
34. PENNISI, E. Breakthrough of the year: human genetic variation. *Science*, 318: 1842-1843, 2007.
35. PEROU, C.M. *et al.* Molecular portraits of human breast tumours. *Nature*, 406: 747-752, 2000.
36. ROSAMOND, J.; ALLSOP, A. Harnessing the Power of the Genome in the Search for New Antibiotics. *Science*, 287: 1973-1976, 2000.
37. SCHINDLER, T. *et al.* Structural mechanism for STI571 inhibition of Abelson tyrosine kinase. *Science*, 289: 1938-1942, 2000.
38. SCHWEDE, T. *et al.* SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res*, 31: 3381-3385, 2003.
39. SOUZA, D.H.F. *et al.* *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase: structure, catalytic mechanism and targeted inhibitor design. *FEBS Lett*, 424: 131-135, 1998.
40. SYVANEN, A.C. *et al.* Enthusiasm mixed with skepticism about singlenucleotide polymorphism markers for dissecting complex disorders. *Eur J Hum Genet*, 7:98-101, 1999.
41. TANAKA, K. *et al.* Analysis of a human DNA excision repair gene involved in group A xeroderma pigmentosum and containing a zincfinger domain. *Nature*, 348: 73-76, 1990.
42. TOLLMAN, P. *et al.* A Revolution in R&D: how genomics and genetics are transforming the biopharmaceutical industry. *Boston Consulting Group Report*, 26 November, 1-64, 2002.
43. TOMAZELA, D.M. *et al.* Pyrano chalcones and a flavone from *Neuraputia magnifica* and their *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase-inhibitory activities. *Phytochemistry*, 55: 643-651, 2000.
44. VERLINDE, C.L.M.J.; HOL, W.G.J. Structure-based drug design: progress, results and challenges. *Structure*, 2: 577-587, 1994.
45. VON ITZSTEIN, M. *et al.* Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, 363: 418-423, 1993.
46. WANG, J.H.; HEWICK, R.M. Proteomics in drug discovery. *Drug Discov Today*, 4: 129-133, 1999.
47. WANG, Z.; MOULT, J. SNPs, protein structure, and disease. *Hum Mutat.*, 17: 263-270, 2001.
48. WHO. "Removing Obstacles to Healthy Development", WHO Report on Infectious Diseases, WHO, <http://www.who.int/infectious-disease-report/index-rpt99.html>, 1999.
49. WLODAWER, A. Rational approach to AIDS drug design through structural biology. *Annu Rev Med*, 53: 595-614, 2002.
50. YUE, P.; LI, Z.; MOULT, J. Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol*, 353:459-473, 2005.
51. YUE, P.; MELAMUD, E.; MOULT, J. SNPs3D: Candidate gene and SNP selection for association studies. *BMC Bioinformatics*, 7: 166, 2006.
52. ZAMBROWICZ, B.P. *et al.* Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature*, 392, 608-611, 1998.