

# A MATEMÁTICA DA VIDA



## Júnior Barrera

Após séculos de especialização do conhecimento, chegamos às últimas ‘letras’ do genoma humano. Agora, para entender como os genes interagem e controlam o metabolismo, a bioinformática busca modelos matemáticos que descrevam esse mecanismo biológico. Para Júnior Barrera, livre-docente do Departamento de Ciência da Computação do Instituto de Matemática e Estatística (IME) da Universidade de São Paulo (USP), vivemos um momento histórico, de aproximação entre as ciências exatas e a biologia, com uma perspectiva de integração dos conhecimentos. Barrera é presidente do Núcleo de Bioinformática da USP (Bioinfo), criado há cerca de três anos no IME, que atua em parceria com outras unidades da USP e diversos institutos, como o Ludwig e o National Human Genoma Research Institute (NHGRI). No ano passado, a Capes reconheceu a importância estratégica da bioinformática para o Brasil, aprovando a criação de dois programas de doutorado na área: um na USP e outro na Universidade Federal de Minas Gerais (UFMG). O primeiro começou no segundo semestre de 2002, integrando oito institutos da USP. Para marcar o início desse curso, foi realizada em maio, em Ribeirão Preto, a Conferência Internacional de Bioinformática e Biologia Computacional (ICoBiCoBi), que atraiu cerca de 400 participantes. Nessa oportunidade, Barrera conversou com *Ciência Hoje* sobre os desafios que os programas vão enfrentar, como o estabelecimento de uma linguagem comum entre as áreas, e sobre as enormes possibilidades de pesquisa que a bioinformática oferece.

**Marcelo Alexandre Barbosa**  
Especial para *Ciência Hoje*/SP

### O que é exatamente a bioinformática?

Para expor a minha idéia do que é bioinformática, vou abordar antes um conceito mais abrangente: a modelagem de fenômenos biológicos, que consiste na construção de modelos quantitativos, expressos através de equações matemáticas. Tais modelos simulam esses fenômenos no computador, fornecendo explicações precisas sobre os mecanismos observados. A modelagem de fenômenos de biologia molecular é o objetivo central do que se chama hoje de bioinformática, termo que ficou mais conhecido com os projetos de seqüenciamento de genomas.

Uma das áreas da biologia mais avançadas em modelagem matemática é a neurociência. Essa área estuda sinais bioelétricos, que podem ser medidos e modelados por técnicas de engenharia elétrica, uma linha de pesquisa que apresentou resultados muito interessantes nos últimos 30 anos. Por exemplo, pode-se simular a forma da propagação do sinal elétrico em um neurônio por um sistema de equações diferenciais que descreve o fluxo iônico pela sua membrana. Também pode-se simular a interação de uma rede de neurônios que controlam o movimento de músculos de algum subsistema. Esses avanços aconteceram mais precocemente em neurociência, principalmente pela capacidade de medir fenômenos bioelétricos. Medir os sinais reais é muito importante, porque permite avaliar a acurácia do modelo. A simulação de um modelo correto vai produzir sinais similares aos medidos, um modelo incorreto não teria essa propriedade. Mais recentemente, algo análogo aconteceu em biologia molecular. Existem tecnologias modernas que permitem medir os sinais produzidos pelos genes. Então, em princípio é possível construir e testar modelos para explicar o controle das células pelas redes de genes. Assim, as pesquisas em biologia molecular vão fazer um caminho análogo às pesquisas em neurociências.

### Como é possível representar fenômenos biológicos com algo tão abstrato como equações matemáticas?

As idéias só podem ser comunicadas através de uma linguagem, um conjunto de regras de descrição associado a significados. Existem linguagens de diversas naturezas, que são mais adequadas para expressar um tipo de idéia do que outra. Por exemplo, para descrever uma música, usam-se notas em um pentagrama. Para relatar um sentimento, compõe-se uma poesia ou uma música. Para falar de política, escrevem-se textos. Para descrever fenômenos mecânicos, usam-se equações matemáticas. Essas equações concentram toda a informação necessária a res-

peito do fenômeno e permitem fazer predições. Por exemplo, sabendo-se a lei de variação de velocidade de um carro em uma estrada e a distância a ser percorrida, pode-se prever o tempo a ser gasto no trajeto. De forma análoga, conhecendo-se as leis de fluxo de íons na membrana dos neurônios, pode-se deduzir a forma do potencial de ação. Conhecendo-se as regras de interação entre os genes, pode-se prever a via metabólica a ser ativada. Quando os fenômenos biológicos passaram a ser medidos quantitativamente, a linguagem natural para descrevê-los passou a ser a matemática, a exemplo do que aconteceu na física. Antigamente, quando os fenômenos biológicos eram descritos de forma apenas qualitativa, a linguagem especializada era a mais adequada. Em muitos casos, a linguagem especializada ainda é a única forma de descrever fenômenos biológicos, dada a sua complexidade.

## Quando os fenômenos biológicos passaram a ser medidos quantitativamente, a linguagem natural para descrevê-los passou a ser a matemática, a exemplo do que aconteceu na física

### O que isso representa para a ciência?

Olhando em perspectiva a história da evolução do conhecimento, os primeiros esforços para o entendimento da natureza foram muito unificados. O mesmo pesquisador estudava filosofia, matemática, biologia, astronomia etc. Com o tempo, percebeu-se a complexidade dos fenômenos e a necessidade de abordagens mais especializadas. Esse caminho mostrou-se muito produtivo e o conhecimento avançou durante séculos dessa forma. Agora, depois de tanto tempo de especialização, curiosamente começa a surgir um processo inverso, que retorna àquela idéia inicial de unificação. Então, segmentos do conhecimento, como biologia e ciências exatas, que tiveram cursos evolutivos muito distintos, começam a se aproximar. A razão disso é que foram descobertos fenômenos muito básicos em biologia e técnicas para medi-los, ao mesmo tempo que a matemática e a computação adquiriram maturidade conceitual e tecnológica. Isso criou o cenário que permite essa perspectiva de conhecimento integrado que, a meu ver, é um marco histórico. Acredito que a época contemporânea, quando for estudada daqui a séculos, vai estar fortemente marcada por essa unificação de conhecimentos de origens diferentes. Vivemos, com certeza, um momento histórico, de aproximação das áreas de biologia e ciências exatas. ▶

## **Nesse contexto, o que representa o seqüenciamento do genoma humano, recentemente finalizado?**

Ainda temos um caminho muito longo pela frente. Estamos de fato interessados no comportamento dinâmico dos genes, porque eles têm um papel de agentes de controle, regulando grande parte dos fenômenos que ocorrem nos organismos, que são cadeias de reações químicas chamadas de vias metabólicas. Quem define a ordem dessas reações são as enzimas, diretamente construídas por proteínas produzidas pelos genes. Estes interagem e a dinâmica de sua interação define a ordem em que as proteínas vão ser geradas e, conseqüentemente, a ordem em que as reações vão ocorrer. É exatamente o mecanismo dessa interação que ainda temos muita dificuldade de entender. Então, os genes são os elementos que definem o controle do desenvolvimento dos organismos. Conhecer o código genético é um passo importante para esse estudo, mas há várias etapas subseqüentes necessárias para se entender esse comportamento dinâmico, que é o objetivo que se vislumbra como mais importante.

**Acredito que a época contemporânea, quando for estudada daqui a séculos, vai estar fortemente marcada por essa unificação de conhecimentos de origens diferentes. Vivemos, com certeza, um momento histórico, de aproximação das áreas de biologia e ciências exatas**

**O senhor já disse que existem 'redes de comunicação' no sistema gênico, onde a 'unidade de informação' é a proteína. Há alguma previsão de quando vamos começar a entender as mensagens?**

Ninguém é capaz de prever quando conheceremos bem as redes de regulação gênica. São inúmeras redes controlando organismos diferentes e, embora elas tenham propriedades comuns, são diferentes umas das outras e regulam fenômenos diferentes. Os pesquisadores ficarão muito contentes quando conseguirem entender algum dos subsistemas de um organismo simples: como funciona seu controle metabólico, descrito por modelos matemáticos. Isso, a rigor, ainda não foi feito e é um tema de pesquisa sobre o qual se debruçam as grandes universidades e os centros de pesquisa.

## **A bioinformática é um dos caminhos para se chegar a isso?**

Sim, é um dos elementos, mas o avanço do conhecimento é complexo e envolve pesquisadores de bioquímica, de diversas especialidades de biologia, química, medicina, computação, matemática fundamental, estatística, física e engenharia. Todos eles têm a sua contribuição a dar. As pesquisas exigem equipes multidisciplinares grandes. Juntar esse conhecimento é um esforço complexo, que envolve ações científicas integradoras e ações administrativas para viabilizá-las, como a criação de grupos de pesquisa com perspectiva multidisciplinar, programas de doutorado, eventos para fomentar a integração de especialistas de áreas diferentes. Enfim, uma série de ações em conjunto, com apoio das instituições que reconheçam a importância desse projeto. Em todos os países desenvolvidos observamos esforços dessa natureza e, no Brasil, não está sendo diferente.

## **Como ocorre essa integração de conhecimentos na bioinformática?**

Hoje há muitos projetos em andamento que utilizam o seqüenciamento de organismos ou outras formas de medir propriedades nesse nível básico, como a expressão gênica, através de imagens de cDNA *Microarray*, ou da espectrometria de massas, que identifica e separa proteínas. Tudo isso é armazenado em grandes bancos de dados e o objetivo é, a partir desses dados e do conhecimento anterior acumulado na literatura, construir modelos explicativos de fenômenos biológicos. É aí que aparece o nome *data mining* (mineração de dados), que são apli-

cações de procedimentos de matemática, computação e estatística, utilizadas para cruzar informações ou generalizar conceitos. A bioinformática especializa técnicas de *data mining* para aplicações em biologia, onde são usadas para tentar extrair informação das bases de dados. Simplificando, temos os bancos de dados, o conhecimento acumulado ao longo da história e o biólogo, que quer conhecimento novo. Com a informação e as ferramentas disponíveis, ele levanta hipóteses (possíveis modelos), que em geral são muitas. Quando for para o laboratório, em vez de fazer experimentos para confirmar todas as hipóteses, ele vai testar somente aquelas selecionadas pelo *data mining*, que serve para juntar todas as informações e reduzir as hipóteses candidatas. Mas essa redução é drástica para tornar factível olhar os casos mais interessantes e testá-los.

O que ele confirma vai para o banco de dados e para a literatura. Isso eu chamo de espiral de avanço do conhecimento.

**Além da genética e neurociência, existem outros campos em que se espera uma interação com a bioinformática?**

Toda a biologia deve se beneficiar dos desenvolvimentos em bioinformática. Contudo, a bioinformática está longe de ser uma obra acabada: inúmeros avanços em matemática e computação devem permitir aperfeiçoar as suas técnicas. De fato, creio que no futuro a competitividade de um grupo multidisciplinar de pesquisas vai depender muito da sua capacidade de desenvolvimento de novas técnicas de mineração e interpretação dos dados. A tendência é que progressivamente os grupos tenham acesso aos mesmos dados e, então, a capacidade de extrair informação dos dados vai fazer toda a diferença.

**Qual o valor estratégico da bioinformática para o Brasil?**

A bioinformática tem um papel estratégico para o país. Imagine que nós nos capacitamos apenas para fazer experimentos biológicos e as correspondentes medidas. Isso seria como dispor de lençóis petrolíferos em águas profundas e não ter tecnologia para fazer a extração do óleo. Montar bases de informações sem ter instrumentos para encontrar as relações e extrair conhecimento dessas bases é fazer a coisa pela metade. Precisamos das duas partes. Nos capacitar para fazer medidas dos organismos que nos interessam e que terão um papel econômico e social importante, mas também nos preparar para extrair informação das bases de dados geradas.

**É nesse contexto que se insere o programa de doutorado?**

O doutorado tem o papel de capacitar pessoas para trabalharem na área de bioinformática e de fomentar pesquisas inéditas. Os alunos devem aprender técnicas de matemática, física, computação, estatística, biologia e bioquímica, de forma a estarem aptos a explorar grandes volumes de dados em busca de informação que permita construir novos modelos para fenômenos biológicos de interesse.

**Como está sendo o diálogo entre alunos de exatas e biológicas?**

Sem dúvida, um dos grandes desafios do nosso programa é desenvolver uma linguagem comum entre biólogos e matemáticos, e esse doutorado tem ca-

racterísticas particulares para permitir uma melhor integração. Uma regra é que cada aluno precisa ter dois orientadores, um especialista da área de exatas e outro da de biológicas. Outra característica é que aceitamos alunos de formação tanto em ciências exatas quanto em biológicas, o que não é usual. Definimos um elenco de disciplinas básicas, onde alunos oriundos de ciências exatas fazem cursos de ciências biológicas e vice-versa. Em outros programas, a banca se reúne apenas no final para analisar os resultados, mas no novo programa temos a intenção de fazer um acompanhamento anual dos alunos e realizar *workshops*, para que os professores do programa observem o trabalho de

**As pesquisas exigem equipes multidisciplinares grandes. Juntar esse conhecimento é um esforço complexo, que envolve ações científicas integradoras e ações administrativas para viabilizá-las**

seus colegas e dêem sugestões. Contudo, ainda é cedo para dar uma resposta mais categórica, porque o nosso programa está começando, tem apenas seis meses de vida e 15 alunos aceitos, sendo 40% de biológicas e 60% de exatas.

**O senhor fez sua carreira em processamento digital de imagens. Como passou a se interessar pela área de biologia?**

O processamento de imagens é uma área de aplicação na qual trabalhei bastante, mas do ponto de vista de fundamento matemático, a área em que trabalho mesmo é a de reconhecimento de padrões, ou *data mining*. Minhas contribuições mais significativas estão no desenvolvimento dessas técnicas. Foi esse vínculo conceitual que me ligou à área de biologia.

Há uns quatro anos, fui convidado a trabalhar com o pessoal do NHGRI (National Human Genome Research Institute), de Washington. O diretor do instituto na época, Jeff Trent, me convidou para participar em trabalhos na área de *cDNA Microarray*, em conjunto com o grupo liderado por Edward Dougherty, meu velho companheiro de pesquisas. Participávamos de muitas reuniões em Washington com pesquisadores de câncer do NHGRI. Acabei conhecendo também pessoas aqui na USP. Hugo Armelin, do Instituto de Química [IQ], motivou-se muito pela área e propôs um projeto temático pioneiro à Fapesp [Fundação de Amparo à Pesquisa do ▶

Estado de São Paulo], o Cage [Cooperation for Analysis of Gene Expression], que integra colegas do IQ e do IME. Hernan Chamovich e Carlos Humes Jr., que na época ocupavam cargos administrativos na USP, reconheceram a importância estratégica da área e foram grandes incentivadores da intensificação das atividades na universidade. Logo depois da criação do Cage, colegas do departamento começaram a trabalhar no projeto VGDN [Viral Genetic Diversity Network], de análise de diversidade viral no estado de São Paulo. Percebemos, então, muita similaridade nos problemas estudados no Cage e VGDN e vantagens em nos organizarmos em um núcleo (Bioinfo), que permitiria compartilhamento e aproveitamento mais produtivo de recursos.

### **Quais os principais projetos em que o Bioinfo está envolvido?**

Cooperamos com colegas da área de biologia em mais de uma dezena de problemas de aplicação. Vou ci-

mento refinado do ciclo do parasita, teste de drogas, evolução epidemiológica etc.

### **Poderia citar outros projetos de pesquisa inovadores relacionados à bioinformática?**

Na área de medicina existem muitas pesquisas em andamento que envolvem os fundamentos que estamos discutindo. Existe um grande esforço para desenvolver técnicas para o diagnóstico de câncer que possam detectar um tecido maligno muito precocemente e com grande precisão. A precocidade do diagnóstico pode aumentar muito as chances de um paciente, enquanto a precisão permite definir a terapia mais adequada e o grau de agressividade do tratamento prescrito. Uma classe de técnica muito estudada é a caracterização de um tecido pela expressão de um conjunto pequeno de genes. A ideia é que a expressão desses genes funcione como uma assinatura que caracteriza o tipo de câncer. Dessa forma, mesmo cânceres difíceis de serem detectados histologicamente poderiam ser diagnosticados.

Outro estudo de que tomei conhecimento é a tentativa de pesquisadores de New Jersey de modificar geneticamente células nervosas, fazendo com que algumas se tornem regenerativas. Essa possibilidade seria uma esperança para pacientes com lesão na coluna causada por algum acidente. Outros estudos ambiciosos, esses realizados aqui no ICB-USP, tentam modificar geneticamente mosquitos transmissores de doen-

ças para que possam passar a proteção da vacina em vez de parasitas.

Em virologia o seqüenciamento de vírus em pacientes infectados, por exemplo, por HIV, já é uma prática corrente, que permite avaliar mutações do vírus e o efeito de drogas nos pacientes. A partir da análise evolutiva de populações de vírus, alguns até sonham com a possibilidade de prever com quais vírus teremos que conviver em anos futuros e nos prepararmos antecipadamente para isso.

Na agricultura e na pecuária os projetos não são menos ambiciosos. A indústria de papel sonha com árvores mais ricas em celulose, os produtores de laranja sonham com árvores mais resistentes a pragas, a indústria de álcool gostaria de retardar o período de floração da cana-de-açúcar para torná-la mais produtiva, os avicultores gostariam de aves com mais peito e coxa, animais geneticamente modificados poderiam produzir substâncias de interesse medicinal. O potencial é imenso, esperamos que venha a ser usado apenas para o bem-estar da humanidade. ■

## **Toda a biologia deve se beneficiar dos desenvolvimentos em bioinformática. Contudo, a bioinformática está longe de ser uma obra acabada, inúmeros avanços em matemática e computação devem permitir aperfeiçoar as suas técnicas**

tar apenas alguns nos quais estou pessoalmente envolvido. Estudamos com Hugo Armelin problemas de controle do ciclo celular. Uma pergunta-chave nesse estudo é como as redes de expressões gênicas regulam a duplicação das células. Para tanto são necessários experimentos sofisticados de medida de sinais biológicos e modelagem dos sistemas dinâmicos que descrevem as redes de regulação gênica. Estudamos com Luís Fernando Reis e Helena Brentani, do Instituto Ludwig, técnicas para identificar genes que permitam separar tecidos cancerígenos dos não-cancerígenos ou mesmo diagnosticar o tipo de câncer de um tecido. Conduzimos estudos similares em parceria com o grupo do NHGRI, liderado por Jeff Trent, e da Texas A & M, liderado por Edward Dougherty. Pesquisamos com Sandro de Souza, também do Instituto Ludwig, genes raros e outras estruturas em cromossomos humanos. Estudamos com Hernando del Portillo e outros colegas do Instituto de Ciências Biomédicas da USP diversos aspectos do desenvolvimento da malária: entendi-