

# Modelador de estruturas

Especialista em explicar o envelhecimento de proteínas e do DNA, físico brasileiro ajudou a restaurar o genoma do mamute lanoso

MARCOS PIVETTA E RICARDO ZORZETTO

José Nelson Onuchic gosta de ser livre para se mover entre diferentes temas que pesquisa. “A beleza da vida acadêmica é poder estudar o que se bem entende”, afirmou em entrevista a *Pesquisa FAPESP* em 9 de outubro.

U Pesquisador da Universidade Rice, nos Estados Unidos, Onuchic estava de passagem pelo Brasil. Veio participar de um simpósio sobre temas atuais em biofísica molecular e foi ao Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), em Campinas, onde funciona a fonte de luz síncrotron Sirius, propor um acordo de colaboração. Ele está interessado em trazer para o país uma técnica de sequenciamento de genomas desenvolvido por colaboradores nos Estados Unidos e em usar o Sirius para analisar a estrutura tridimensional de genomas.

Filho de professores universitários, os matemáticos Nelson Onuchic [1926-1999] e Lourdes de La Rosa Onuchic, que, aos 93 anos, ainda está na ativa, José Nelson graduou-se em engenharia elétrica em 1980 e em física em 1981 pela Universidade de São Paulo (USP) em São Carlos.

Após o doutorado no Instituto de Tecnologia da Califórnia (Caltech), orientado pelo físico John Hopfield, que ganhou o Nobel de Física de 2024, retornou brevemente a São Carlos antes de se tornar professor da Universidade da Califórnia em San Diego (UCSD). Lá, passou a estudar o envelhecimento de proteínas e, com colaboradores, propôs dois conceitos que o tornaram referência na área. Em 2011, mudou-se para a Rice, no Texas, onde iniciou uma linha de pesquisa em câncer. Aos 66 anos, ele publicou mais de 400 artigos científicos, citados 45 mil vezes por outros trabalhos, e é codiretor do Centro de Física Biológica Teórica, que realiza estudos em áreas da fronteira do conhecimento. Leia, a seguir, os principais trechos da entrevista.



Onuchic, durante passagem por São Paulo no início de outubro

**Em outubro, o senhor participou de um simpósio sobre biofísica molecular em São Paulo. Do que falou?**

Da estrutura do genoma, de como ele está organizado no núcleo das células. Na maior parte do tempo, a molécula de DNA [que abriga os genes e os trechos não codificadores de proteínas] se apresenta na forma de cromatina, que é o DNA levemente envelado em torno de proteínas. A estrutura tridimensional da cromatina é importante para controlar quais genes serão lidos e quando. Ela pode ocultar alguns genes enquanto expõe outros à maquinaria de leitura das células. Os livros convencionais de genética molecular não falam da estrutura tridimensional dessa molécula, um conhecimento que muda rapidamente.

**Em 2016, o senhor publicou um artigo na *PNAS* falando justamente sobre como a estrutura do DNA muda e permite que o gene seja lido, não?**

Exatamente. Começamos a trabalhar nessa área com Erez Aiden, do Baylor



College of Medicine, que usa uma técnica chamada Hi-C contact map. Ela permite identificar, por exemplo, trechos da cromatina que se encontravam distantes na molécula de DNA, mas ficavam espacialmente próximos na fita enovelada. Usamos esse mapa de contatos em duas dimensões para gerar a estrutura 3D da cromatina. O modelo evoluiu e, hoje, não precisamos mais do Hi-C. Agora, pegamos segmentos de 50 mil pares de bases do genoma, 10 vezes o tamanho de um gene, e separamos cada segmento em categorias. Na versão mais simples do modelo, a gente divide a cromatina em duas categorias. Chamamos de A a eucromatina, menos condensada e mais fluida, onde há mais genes expressos, e de B a heterocromatina, mais enovelada e com menos informação genética. O modelo também usa informações da técnica ChIP-seq, que investiga como as proteínas estão associadas ao DNA, para obter informações epigenéticas, ou seja, do padrão de ativação dos genes, e classificar os tipos de cromatina.

#### Como ele funciona?

O modelo tem quatro partes. Três são mais importantes. Primeiro, consideramos o DNA um polímero *soft*, que pode ser cortado, como naturalmente ocorre por causa de certas proteínas. Depois, o modelo separa os tipos A e B por semelhança. Os A tendem a se agrupar com os A e os B com os B. A terceira parte é o cromossomo ideal, em que partes espacialmente próximas da cadeia tendem a se atrair. Isso cria contatos locais, que geram compactação e formação de hélices. Aumentando os contatos locais, evita-se a formação de nós. Isso é importante nos cromossomos. Quando seus trechos são expressos, eles precisam ser desenrolados e não se quer que surjam nós. O modo como a cromatina se enovela evita isso. Uma parte da compactação local dizemos que é motorizada, promovida por proteínas. Em 95% do tempo, a célula se encontra em interfase, quando a cromatina está desenrolada e é duplicada. Nos 5% restantes, ela entra em mitose, que é a divisão celular. Na mitose, aumenta

a compactação e o cromossomo é mais enovelado. Nosso modelo mostra que, aumentando a motorização, aumentam-se a compactação e a formação de hélices.

#### O que governa tudo isso são as características químicas e elétricas das moléculas?

Isso. Mas é preciso lembrar que o genoma tem funções muito específicas, que precisam ser preservadas por sua conformação. O genoma tem de apresentar a capacidade de ser transcrito [lido pela maquinaria da célula e codificado na forma de RNA], de duplicar e de se separar. Apesar de ele ser tratado pelos modelos como um polímero, assim como as proteínas, a funcionalidade do genoma e das proteínas é distinta. As proteínas são mais rígidas, enquanto o genoma é mais maleável. Trabalhamos hoje para saber como as características estruturais do genoma são importantes para sua função. Isso já foi feito para proteínas. Por muito tempo se trabalhou em enovelamento de proteínas para definir a estrutura delas. Mas, no fundo, o que se quer saber é a função da proteína.

#### Mas, no caso das proteínas, a estrutura define a função, não?

Muitos livros de biologia falam da relação entre estrutura e função. Mas entender a função é mais complicado do que isso. Muitas proteínas têm várias estruturas. Uma parte das proteínas é ordenada. Outra é desordenada e se torna ordenada quando se liga a outra molécula. Quando se começou a determinar as primeiras estruturas de proteína, iniciou-se pelas mais fáceis, as enzimas, que têm estrutura bem definida. Nelas, a função está estritamente relacionada à forma. Só que as proteínas são muito mais diversas do que isso. A capacidade de esses polímeros, feitos pela combinação de 20 tipos de aminoácidos, assumirem estruturas diferentes é espantosa. Elas podem ser enzimas, sinalizadores, fibras.

#### Como os físicos ajudam os biólogos a entender essa complexidade?

A questão do enovelamento de proteínas pode ser analisada de duas maneiras. O biólogo pergunta: se te dou essa sequência de aminoácidos, consegue determinar a estrutura e a função da proteína? Como físico, pensei diferente: a partir de uma sequência de aminoácidos, é

possível saber se ela tem ou não estrutura definida? Estou mais interessado em saber o que difere uma proteína de um polímero aleatório de aminoácidos. Partindo dessa perspectiva, propus em 1992 [em um artigo na revista *PNAS*, assinado com Peter Leopold e Mauricio Montal, da Universidade da Califórnia em San Diego] o conceito de funil de enovelamento. Em 1995, Paul Wolynes, Joe Bryngelson, Nicholas Socci, Zaida Luthey-Schulten e eu finalizamos o conceito de paisagem de energia.

#### O que dizem essas propostas?

Desde os anos 1960, havia um paradoxo sobre o enovelamento de proteínas. Com base no estudo de pequenas proteínas globulares, havia se assumido que elas se enovelam e alcançam a configuração final quando atingem um estado de energia mínima. Mas as proteínas podem assumir um número tão grande de conformações que levaria um tempo muito longo para encontrar esse estado. Um polímero aleatório tem vários estados mínimos de energia que são estruturalmente muito diferentes. Na proteína, propusemos a ideia que não basta ter um estado mais atrativo do que outros. Ele precisa ser mais atrativo para permitir encontrar a estrutura desejada e menos para aquelas que não se quer. O número de estados favoráveis diminui à medida que se aproxima da estrutura mais estável.

#### Daí a ideia de funil?

Sim. Tem ainda um conceito que chamamos de entropia configuracional. Se uma proteína estiver acima da temperatura de enovelamento, haverá vários estados possíveis. Se estiver abaixo, haverá um estado que começa a prevalecer, mais estável. Esse estado tem energia atrativa suficiente para compensar a entropia configuracional contra a qual compete. Ou seja, há uma temperatura em que a energia ganha da entropia. Só que se quer que essa temperatura que favorece o enovelamento seja atingida antes que a molécula da proteína alcance o chamado estado vítreo e perca mobilidade, ficando presa ali, como em uma armadilha. A ideia da paisagem de energia diz que há um estado atrativo que é muito mais profundo do que o das armadilhas. É possível entender muito da proteína a partir desta hipótese: se a proteína enovela, ela

conseguiu otimizar a estrutura estável sobre as outras. Com base nessa ideia, parto da estrutura da proteína e crio um modelo em que todos os estados nativos [estáveis] são atrativos e os não nativos são repulsivos. Com esse modelo, é possível entender toda a transição e todos os estados intermediários do enovelamento. Em um artigo publicado neste ano na revista *Science*, minha equipe e a de Paul Whitford, da Universidade Northwestern, colaborou com a de Walther Mothes, da Universidade Yale. Mothes havia usado tomografia eletrônica criogênica para definir as estruturas que a proteína *spike* do novo coronavírus assume durante a invasão celular. Mas eles obtiveram estruturas de baixa resolução. Nós já tínhamos feito um modelo da invasão celular e viu-se que as configurações da *spike* era muito próximas das obtidas por eles. Então, trabalhamos juntos para fazer um modelo combinando dados experimentais com nossas simulações.

#### E o que concluíram?

A *spike* está na superfície do coronavírus. Quando adere ao receptor na superfície da célula humana, a *spike* sofre transformações estruturais que aproximam a membrana do vírus e a da célula.

#### Esse mecanismo já tinha sido visto em outros vírus?

A primeira proteína em que vimos foi na hemaglutinina, do vírus da influenza.

#### É difícil ver porque está tudo sempre em movimento.

Por isso, tiveram de usar anticorpos que se ligavam aos estados intermediários da *spike* para ficarem estáveis. Estamos interessados nisso. Se tenho um anticorpo que se conecta a um estado intermediário, posso modificar esse anticorpo para que interrompa a transição para o estado final. Se bloquear essa transição, talvez se impeça o vírus de entrar na célula.

#### Recentemente o senhor participou de um estudo sobre o genoma do mamute lanoso. O que viram?

O pessoal que iniciou o projeto estava trabalhando nele havia anos e, no final, tinha de fazer a modelagem com os dados e nos chamaram. Eles tiveram de desenvolver estratégias para coletar amostras do material genético do animal, que estava congelado havia 52 mil anos. Quando conseguiram, fizeram um mapa de Hi-C e observaram que havia alguma estrutura preservada. Esse DNA foi congelado e ressecado. Apesar de o cromossomo estar quebrado em diversos segmentos, os pedaços não haviam se movido muito e a estrutura 3D do DNA foi parcialmente preservada. Com o Hi-C e modelagem, conseguimos recuperar essa estrutura.

#### Como seu grupo ajudou?

Dissemos: “Apesar de esse DNA ter 52 mil anos, acreditamos que as regras de como se organizam e de como os trechos são conectados permanecem as mesmas”. Decidimos, então, olhar o genoma dos primos deles, os elefantes. Usamos as informações do genoma dos elefantes para reconstruir o do mamute lanoso. Depois, com nossos modelos do genoma, conseguimos reduzir muito o nível de ruído da amostra e gerar a estrutura do genoma em três dimensões. Em seguida, fomos ver quais partes do genoma do mamute eram mais ativas e quais eram menos. Também olhamos quais regiões genéticas eram ativas no mamute e não são no elefante. Uma é a que gera os pelos, muito mais abundantes nos mamutes. Isso indica que o que fizemos era consistente.

#### Como avalia o cenário da pesquisa hoje no Brasil?

“  
**Usamos as informações do genoma dos elefantes para reconstruir o do mamute lanoso**”

Estamos dando pouca atenção à pesquisa básica. Não focar nela é queimar o futuro. A ciência básica é o depósito da capacidade intelectual da sociedade, enquanto a ciência aplicada, que é importante, gera riqueza. No mundo atual, a segurança nacional é mais associada ao controle do conhecimento do que ao poderio militar. No Brasil, todo mundo quer escrever patente. O número de patentes desnecessárias geradas no país é impressionante. Elas são importantes, mas o controle do conhecimento deve ser mais valorizado.

#### **Não é assim também no mundo?**

Não. Nos Estados Unidos, a gente escreve patente esporadicamente. Todo mundo sabe que manter boas patentes não é simples. É caro, precisa manter um corpo jurídico para protegê-la. É preciso ser seletivo. Mas esse quadro é parte do aprendizado. Tínhamos uma lei de proteção intelectual fraca no Brasil. Estamos aprendendo como fazer.

#### **Quantas patentes seu grupo gera por ano?**

Uma patente a cada quatro ou cinco anos e uns 15 artigos por ano. O que vale para a patente vale para o artigo científico. Há uma tendência de olhar para os números da produção científica ou de patentes quando ocorre a promoção de um pesquisador. Mas, pergunto: “Alguém já leu os artigos? Sabe do que tratam? Qual o impacto do trabalho”. É fácil pegar esses números. Faz parte do processo, mas acho que as coisas estão evoluindo na direção certa.

#### **No Brasil também?**

No Brasil em particular. A qualidade da ciência brasileira melhorou muito. Existem vários grupos que deixaram de ser seguidores e hoje competem de igual para igual em nível internacional. Isso se aplica, por exemplo, a boa parte dos Cepid [os Centros de Pesquisa, Inovação e Difusão financiados pela FAPESP].

#### **Nos próximos dias, irá ao Sirius, em Campinas. O que fará?**

Vou falar com o Antônio José Roque da Silva, diretor-geral do CNPEM [Centro Nacional de Pesquisa em Energia e Materiais] e pesquisadores do Laboratório Nacional de Biociências. Queremos estabelecer uma cooperação entre o nosso grupo e o de Erez Aiden com o CNPEM.



## **Estamos dando pouca atenção à pesquisa básica. Não focar em ciência básica é queimar o futuro**

Queremos aproveitar a janela de oportunidade do Sirius, que tem uma capacidade existente em poucos lugares do mundo. O Sirius consegue fazer tomografia com uma resolução que outros equipamentos não conseguem. Queremos usá-la para fazer tomografia de genomas.

#### **O que querem olhar no genoma?**

O Erez Aiden e a Olga Dudchenko, colaboradores do nosso grupo, estão desenvolvendo um projeto chamado DNA Zoo, Zoológico de DNA. Eles querem sequenciar o DNA de várias espécies de plantas e animais. O desafio de sequenciar o genoma dessas espécies é que não há referência para eles. Ao sequenciar esses genomas, haverá muitos erros. A Olga desenvolveu uma estratégia que usa a técnica de Hi-C [que permite saber quais partes estão próximas espacialmente] para corrigir os erros durante o alinhamento. É um sequenciamento muito mais barato. Ela e Erez estão usando a estratégia para sequenciar o genoma de várias espécies. Queremos transferir essa tecnologia para cá, para fazer o genoma de espécies brasileiras.

#### **O que a luz síncrotron permitirá ver?**

Como o genoma aparece em três dimensões no núcleo das células.

#### **Por que isso é importante?**

O genoma é espetacular. Na divisão celular, ele passa por uma mudança estrutural enorme. Como físico, quero saber se esse sistema consegue se descoordenar e se formar de novo ou se ele tem memória que permite sair do estado em que se encontra na interfase, passar para o da mitose e voltar. Se tem memória, como ela funciona? Qual a consequência disso? Quais os mecanismos de controle?

#### **O que o levou a mudar da Universidade da Califórnia para a Rice?**

Vários motivos. Na ciência, como em tudo na vida, há uma crise de meia-idade. San Diego foi muito bom para mim. Cheguei em 1990, me deram o cargo de professor, o *tenure*, em 1992, e o de titular em 1995. Em 2006, fui eleito para a Academia Nacional de Ciências. Eu estava com 54 anos, havia feito várias coisas na área de proteínas e pensei: “Se continuar aqui, farei isso pelo resto da vida. Que tal mudar?”. Na época, a Rice me ofereceu a possibilidade de iniciar estudos em câncer e o Instituto do Texas de Prevenção e Pesquisa do Câncer, o CPRIT, deu ao nosso centro um orçamento de US\$ 10 milhões para começar um projeto de teoria em câncer.

#### **O que estuda sobre o câncer?**

Fizemos modelos de metabolismo trabalhando com redes genéticas. Quando cheguei lá, comecei a trabalhar com Eshel Ben-Jacob [da Universidade de Tel Aviv, em Israel] e propusemos um modelo para explicar a ativação de genes que leva as células a se tornarem invasivas no câncer quando sofrem a transição epitelial-mesenquimal. Nessa transição, as células epiteliais [estáticas, como as da pele ou as que revestem órgãos internos] sofrem alterações bioquímicas e adquirem características de células mesenquimais [capazes de migrar e invadir tecidos]. Mostramos que ela é governada pela interação de um gene com um microRNA. Essa interação permite gerar estados híbridos, em que as células têm tanto as características epiteliais quanto as mesenquimais simultaneamente. Mostramos que isso pode ocorrer quando as células são submetidas a estresse. As células híbridas conseguem se mover e aderir, criando *clusters*, mais difíceis de serem destruídos. ●