Códigos de comunicação digital reproduzem e identificam mutações em sequência de DNA

Fenômenos podem passar a ser analisados por métodos quantitativos

CARMO GALLO NETTO
carmo@reitoria.unicamp.br

Foto: Antoninho Perri

O professor Reginaldo Palazzo Júnior, coordenador da linha de pesquisa: nova fase exigirá maior diálogo interdisciplinar

Andréa Santos Leite da Rocha: "Os resultados se mostraram mais promissores quando utilizamos a estrutura de anel"

Luzinete Cristina Bonani de Faria: "Fomos ampliando a aplicação do processo"

Tanto os que creem em uma força superior criadora como os que atribuem ao acaso a formação do universo poderão, no futuro, admitir que até as sequências de DNA obedecem ou podem ser explicadas através de uma modelagem matemática. Na física e na química, o uso de equações matemáticas para explicar, quantificar e prever a possibilidade de ocorrência de transformações naturais ou provocadas se tornou rotineiro. Na biologia esse recurso é bem mais recente e restrito. Vários pesquisadores das áreas de teoria de informação e codificação, principalmente nos EUA e Europa, têm procurado reproduzir sequências de DNA através de estruturas matemáticas com o objetivo de melhor compreender o funcionamento do sistema biológico. A primazia coube a um grupo de pesquisadores da Unicamp em colaboração com a USP que estabeleceu uma relação matemática entre um código numérico e a sequência do DNA, sigla que identifica em inglês o ácido desoxirribonucléico – portador dos genes dentro das células.

Os pesquisadores verificaram que existe uma relação entre sequências de DNA e códigos corretores de erros (ECC em inglês). Estes códigos fazem parte do cotidiano dos que usam a internet, celulares, TVs, CDs, pen-drives. De forma geral, eles estão presentes na comunicação via satélite, nas comunicações internas de um computador e no armazenamento de dados. A utilização destes códigos tem como objetivo a correção de erros que ocorrem durante a transmissão ou armazenamento da informação.

A associação de códigos corretores de erros com sequências de DNA constitui objeto de pesquisa desde os anos 80. Vários pesquisadores têm se dedicado a essa procura e na proposição de hipóteses sem conseguir estabelecer relações matemáticas com o DNA. O grupo brasileiro estabeleceu essa relação com diferentes sequências de DNA que constituem o genoma (exons, íntrons, DNA repetitivo, sequência de direcionamento, proteínas, hormônios, gene, etc) até chegarem na reprodução do genoma completo de um plasmídeo.

No Brasil, essa possibilidade de utilização de um código matemático que transcrevesse a sequência de DNA foi proposta pelo professor Reginaldo Palazzo Júnior, do Departamento de Telemática da Faculdade de Engenharia Elétrica e de Computação (FEEC) da Unicamp, a duas de suas alunas de doutorado que com ele também haviam realizado o mestrado: Andréa Santos Leite da Rocha e Luzinete Cristina Bonani de Faria, ambas graduadas em matemática pela PUC-Campinas. Elas se propuseram inicialmente a estudar o transporte de proteínas mitocondriais.

Explicam que as mitocôndrias, organelas responsáveis pela respiração celular, apesar de conter o seu próprio DNA e toda maquinaria necessária para fabricar proteínas, sintetizam somente um pequeno número delas. A grande maioria das proteínas mitocondriais é codificada por genes que ficam no núcleo das células, sintetizadasno citosol e posteriormente enviadas para as mitocôndrias. Esclarecem que, neste caso, a proteína é considerada como a informação que será enviada para a organela, existindo um código padrão para transmiti-la. Nos trabalhos desenvolvidos, elas mostraram que o modelo que empregaram se ajusta a diferentes sequências de DNA.

Na verdade, a modelagem matemática só pode ser desenvolvida com base nos comportamentos biológicos já conhecidos envolvendo os sistemas celulares. Essa modelagem se revela válida na medida em que referenda os fatos descritos pela biologia e se mostra ainda altamente positiva se permitir entender anomalias observadas nos sistemas celulares e até possibilitar previsões de novas descobertas.

Por essa razão, Palazzo que é especialista na chamada teoria matemática de comunicação, área de estudo da transmissão de todo o tipo de informação e de seus códigos, sugeriu inicialmente às suas orientandas que procurassem especialistas da área médica da Universidade para inteirarem-se dos conteúdos biológicos. Foi lá que, depois dos primeiros contatos, receberam a orientação para procurarem o professor Márcio de Castro Silva Filho na Escola Superior de Agronomia Luiz de Queiroz (Esalq) da USP. Estabeleceu-se então um diálogo envolvendo um geneticista especializado em transporte de proteínas, o engenheiro elétrico Reginaldo Palazzo e as matemáticas Andréa e Luzinete. Iniciou-se então a abordagem para descrever uma estrutura matemática nas sequências de DNA.

O grupo passou a contar também com a participação do engenheiro de computação João Henrique Kleinschmidt, ex-aluno de doutorado da FEEC e atualmente professor da Universidade Federal do ABC, em Santo André. A modelagem matemática que se mostrou adequada foi aplicada tanto em sequências de DNA de fungos, vírus, plantas, bactérias e humanos, quanto no gene "TRAV7" do *Homo sapiens* e no genoma do plasmídeo (DNA extra-cromossomais) da bactéria *Lactococcus lactis*, verificando que a maioria destas sequências foram geradas por um código matemático.

O feito dos pesquisadores apresenta uma solução importante e inovadora para a biologia em que os fenômenos podem passar a ser analisados por métodos quantitativos. Palazzo considera que essa nova fase exigirá maior diálogo entre biólogos, matemáticos e engenheiros eletrônicos e afirma estar plenamente convencido de que a teoria da informação constitui uma ferramenta adequada para intercâmbio com a biologia molecular. Ele reconhece que essa interdisciplinaridade ainda está distante, embora a seu ver o grupo tenha dado um grande passo inicial.

A linha de pesquisa orientada por Palazzo desenvolve programas computacionais para a geração, classificação e análise mutacional e de polimorfismos em sequências de DNA. Ela abre a possibilidade, a partir de um código e de uma solução matemática, de corrigir uma mutação ou um erro celular, de produzir proteínas desejadas e também de encontrar proteínas ainda desconhecidas existentes nas células. Em decorrência, o método pode ser aplicado em projetos e pesquisas com a finalidade de criar novas funções para determinada sequência de DNA através de mutações segundo as necessidades comerciais e científicas, afirmam Andréa e Luzinete. As pesquisadoras, que são colegas desde a graduação e trabalharam juntas também no mestrado na mesma linha de pesquisa envolvendo genética, iniciaram no doutorado o trabalho com análises mutacionais em sequências de DNA. Elas explicam que o tema, desenvolvido conjuntamente por elas, por questões formais deu origem a duas teses que se complementam. A tese de Andréa foi apresentada no final do ano passado e a de Luzinete deverá estar concluída até maio.

Explicações

Andréa conta que no início estudaram trabalhos centrados na reprodução de sequências de DNA através de códigos de corretores de erros. Constataram que as abordagens neles utilizadas não eram capazes de reproduzir as sequências. Para ela, ao iniciar os trabalhos, o grupo teve a felicidade na escolha das estruturas matemáticas, corpo e anel, utilizadas nas construções de códigos capazes de reproduzir sequências de DNA. "Iniciamos com as sequências de direcionamento organelares por causa do comprimento pequeno, o que diminui a complexidade dos cálculos numa primeira aproximação. A partir daí consideramos sequências envolvendo outras funções biológicas e comprimentos variados. Os resultados se mostraram mais promissores quando utilizamos a estrutura de anel, apesar de pouco difundida na área de códigos", explica Andréa.

Luzinete acrescenta que, devido à característica multidisciplinar da pesquisa, a fundamentação e a caracterização de sistemas biológicos associados a transportes de proteínas através da teoria da informação e codificação foram possíveis de ser realizadas a contento. Como os códigos corretores de erros são os mesmos usados no sistema de comunicação para armazenar e transmitir informações de uma fonte para um receptor, foram estabelecidas uma analogia entre o sistema de comunicação e o sistema biológico por considerarem que nele existe uma informação armazenada que é transmitida. "A partir de sequências específicas, fomos ampliando a aplicação do processo e conseguimos resultados interessantes, válidos para células eucarióticas e procarióticas que envolvem mundos biológicos bastante distintos".

Essa modelagem matemática, afirma o professor Palazzo, que permite caracterizar um sistema biológico, pode possibilitar, através de um programa computacional, a realização de análises que levem a resultados concretos. Porque, lembra ele, a pesquisa na área da saúde é, em geral, demorada e dispendiosa. E, mesmo assim, não raro, depois de um longo tempo não se conseguem resultados positivos. Com a utilização de um software, a caracterização e a simulação do que se pretende pode ser realizada previamente no computador. As previsões teóricas podem ser testadas em laboratório de uma forma sistemática e eficaz, tornando os resultados muito mais confiáveis. A modelagem matemática pode possibilitar a cura de uma doença de forma mais rápida".

Para o professor, a teoria se mostrou válida para certos fatos concretos, mas admite que ainda há muito a ser feito. Ele considera que a linha adotada está sendo delineada e lembra que alguns pesquisadores chegam a afirmar que não existe uma estruturamatemática que possa ser associada a sequências de DNA. "Nós estamos mostrando que existe", afirma o docente. Ele insiste que efetivamente o processo adotado inova na modelagem em relação aos trabalhos propostos anteriormente e revela-se bastante coerente e consistente com o modelo biológico vigente.

Perspectivas

Os pesquisadores consideram que, por tratar-se de uma proposta nova, que se insere numa área estratégica de pesquisa e de inovação tecnológica e, mais, que a detenção do conhecimento necessário se resume a poucos pesquisadores, é imprescindível que esforços sejam despendidos na formação de recursos humanos para o estabelecimento de uma massa crítica de especialistas na área. Para isso, consideram fundamental o respaldo institucional quanto ao reconhecimento da relevância em prover uma infraestrutura de pesquisa bem como no estabelecimento de um programa de pós-graduação.

Andréa e Luzinete alertam que para manterem a fronteira nessa linha de pesquisa será necessário conseguir mais investimentos e recursos humanos que garantam continuidade na obtenção de novos resultados, sem o que temem o comprometimento dos esforços até agora despendidos. Para o professor Palazzo, parcerias com empresas privadas são bem vindas, pois os objetivos e as dinâmicas do mundo acadêmico e empresarial são diferentes: "A dinâmica atrelada às instituições de fomento à pesquisa não atendem às necessidades impostas pelo célere desenvolvimento que o trabalho impõe". Para ele, a pesquisa assume particular importância para laboratórios que desenvolvem fármacos, indústrias têxteis, empresas de agronegócios e de biotecnologia de um modo geral.

Autora ganhou 3 prêmios

Em decorrência do doutorado que tratou do "Modelo de Sistemas de Comunicação Digital para o Mecanismo de Importação de Proteínas Mitocondriais através de Códigos de Corretores de Erros", Andréa ganhou três importantes prêmios na área de engenharia. Recebeu o Prêmio Marechal-do-Ar Casimiro Montenegro Filho, concedido pela Secretaria de Assuntos Estratégicos da Presidência da República, atribuído à melhor tese de doutorado na área de engenharia elétrica. Ela foi tambéma grande vencedora do Prêmio Brasil de Engenharia 2010, cuja tese, além de obter o primeiro lugar na categoria pós-graduação da área temática "Tecnologias Digitais e de Comunicação", recebeu também o Grande Prêmio Inovação, concedido ao melhor entre todos os trabalhos inscritos na competição. Além dos troféus e prêmio em dinheiro, Andréa ganhou uma viagem a Paris com acompanhante. O Prêmio Brasil de Engenharia 2010 foi uma realização do Sindicato dos Engenheiros no Distrito Federal e do Instituto Atenas de Pesquisa e Desenvolvimento, com apoio oficial do Ministério da Ciência e Tecnologia e o patrocínio máster da Petrobrás.