



意志のある一歩が未来を拓く 研究応援プロジェクト

第30回 リバネス研究費

日本マイクロソフト賞 決定! 採択者インタビュー



◎ 日本マイクロソフト賞

学術的価値ある研究や社会課題の解決につながる研究を応援するため、自社が提供する、クラウドコンピューティングシステム「Microsoft Azure」を使った研究テーマを募集。

助成内容 研究費 50 万円 + 1200 万円相当の Microsoft Azure クラウド利用権

採択テーマ 『ヒトタンパク質の相互作用ネットワークを網羅的に予測し活用するクラウドシステム開発』

情報科学と生命科学をつなぐ開発者を目指して

～「Microsoft Azure」で誰もが使えるタンパク質間相互作用予測システムを作る～



東京工業大学 情報理工学院 情報工学系 助教

大上 雅史さん

東京工業大学の大上雅史さんが開発したタンパク質間相互作用の計算ソフト「MEGADOCK」は、「現在わかっている全てのタンパク質の関係性を網羅的に計算し、予測する」システムだ。これを、多くの研究者に活用されるシステムにするべく、「Microsoft Azure」と連携し、新たな研究を開始している。大上さんがこの研究を通じて実現したい世界や、研究への想いを伺った。

画面の中に、生命現象が浮かび上がってくることに感動して

中学校卒業後、高等専門学校（高専）に進学し、情報工学を学んでいた大上さんが、生物学に出会ったのは高専5年生の卒業研究のとき。当時、急性白血病に関連したDNAマイクロアレイのデータから、病気の発症リスクやタイプを分類を機械学習で解析するプログラムを開発していた。研究を進めるうち、生物学に対する印象が大きく変わったという。「プログラミングや電子工学にどっぷりついていた当時の自分にとって、人の健康や病気といった『生々しいもの』がパソコンの画面の中に出てくる、ということがとても衝撃でした。そ

れまで暗記科目という印象しかなかった生物学の背後に物理や化学、情報科学が重要な要素として存在していることを実感し、興味を持つようになったという。それまではシステムエンジニアやシステムインテグレータとして企業で活躍する未来を描いていたが、「情報科学によって、生命現象の解明や、ひいては病気の治療や予防にも貢献できる研究をしてみたい」と考えるようになり、高専卒業後、現在所属する東京工業大学への編入を決めた。

タンパク質集団の関係性を計算で予測する

大学編入後から一貫して大上さんが取り組むのは、タンパク質間相互作用ネット

ワークの網羅的予測システムの開発だ。タンパク質のかたち（構造）は多種多様であるが、構造解析技術が進み、現在ではデータベースに11万件ものタンパク質構造情報が集積されるようになった。分子生物学の実験研究分野では、タンパク質1つ1つの構造や機能の解析ももちろん重要だが、細胞内に集団として存在するタンパク質間関係性も近年注目されるようになっていく。しかし現状では、1組のタンパク質ペアについて精細に調べたり、複数のタンパク質間の全体像を粗く調べるといった実験的な解析がほとんどである。「タンパク質の構造データが蓄積されてきた今、情報科学の知見によって、タンパク質間相互作用を網羅的に計算で予測するというのは、と

てもチャレンジングで面白い。実験データが少ないタンパク質に対して、関係の深いタンパク質群を妥当に予測・提示できるシステムは価値があるのでは。そう考えた大上さんは、「MEGADOCK」の開発を開始。「ファンデルワールス力や静電的な相互作用、疎水・親水性といった3種類の関数を使って、ある程度正確に相互作用が計算できる」という知見を取り入れ、タンパク質を簡易的なモデルに変換して高速で計算する方法の検討を続けた。そして、スーパーコンピュータを使ってわずか1日で計算を可能にする、「MEGADOCK」のシステムを生み出したのだ。

誰もが使える、生命現象を解き明かすツールを目指す

「MEGADOCK」により、タンパク質間相互作用を「短時間で、大規模に」計算・予測できるようになった。「しかし、単にできるようにしただけでは意味がありません。例えばDNA配列の相同性を解析する『BLAST』があそこまで生命科学研究に浸透したのは、手法が優れていることももちろんですが、ウェブブラウザ上で気軽に解析ができるようになったことも大きな要因です。『MEGADOCK』も、例えば予測の結果を共有して誰もが活用できるかたちにすることや、利便性の良いインターフェースが必要だと思いますし、その実現が開発者の使命だと思っています」。そう話す大上さんが挑戦するのが、今回のAzureを使った研究だ。計算を高速にしたとはいえ、現在の「MEGADOCK」は普通のパソコンで動かす場合、10万種のタンパク

質の相互作用の可能性を全て計算しようとすると数百年かかってしまう。そこで、これまでスーパーコンピュータで動いていた「MEGADOCK」をクラウド上に持ち込むことで、さながら「ミニ」スパコンのようなクラウド計算環境上でシステムを動かせるようにし、さらに計算結果を簡単に閲覧・操作できるインターフェースを作ろうと考えているのだ。これができれば、実験を専門とする研究者でも、自分の興味のあるタンパク質がどのようなタンパク質と相互作用しやすいのかはもちろん、タンパク質間ネットワークの地図の中でおおよその部分に位置するタンパク質なのかといったことも、パソコン上で気軽に調べることが可能になる。そんな世界を大上さんは目指している。

「MEGADOCK」システムを「Microsoft Azure」上に構築することに、大上さんは別の期待も寄せている。それは生命学者と情報学者の共同研究の「加速」だ。「例えば、物理化学的に相互作用しやすいという計算が出ていても、実際の細胞内ではそもそもそれらのタンパク質が会うのは不可能、というケースもあります。システムを作る側である情報学者と、使う側である生命学者が連携し、知識を融合させることで互いの研究にフィードバックがかかり、研究がより広がっていくものと思っています」。

研究費で自分の新しい挑戦を加速する

大上さんがこの研究を始めたときは、自分のこれまでの分野と全く違う言葉が飛び

交う世界を理解するのにとても苦労したという。しかし、「様々な研究者とディスカッションを重ね、その間にある課題を見つけ、ツールを整備したことで、生命現象の解明に貢献できるようになってきた。情報科学と生命科学の『つなぎ役』であることが自分の研究の魅力だと思っています」。そう語る大上さんが研究する上で大切にしているのは、自分の描いた研究だけに一直線になりすぎず、ちょっと横道にそれたことをしてみることに。開発者として企業か大学かといったフィールドへのこだわりはあまりないが、大学は「学生の教育」という研究活動とは異なる役割も担えることが魅力だと大上さんは考えている。「例えば、研究費の申請書類なんか書いてたら、研究が進まないじゃないかと思う学生は多いかもしれませんが、しかし、『申請書を書く』ことは普段の研究とはまた違う頭を使うもので、学生も、自分の研究を見つめ直す良い機会になると思います」。大上さんが初めて研究費を申請したのは大学院修士課程1年のときに申請したリバネス研究費だった。その経験が今の研究への姿勢につながっている。今回挑戦するAzureを使った研究も、リバネス研究費の告知を見て、「1千万円分のMicrosoft Azureの利用権が手に入ったら何ができるだろう？」と考えてみたことが始まりだった。「趣味や、本業とは違う活動も大切に。若いうちに挑戦できる今回のような研究費には、ぜひトライしてみたら良いのではと思いますよ」。

(文 楠晴奈)

Microsoft Azure が、あなたの研究を加速します

Microsoft Azure は、分析、コンピューティング、データベース、モバイル、ネットワーク、ストレージ、Webなどを統合した、現在急成長を遂げているクラウドサービスです。

マイクロソフトの研究部門は、Artificial intelligenceをはじめとした様々な研究活動を行っています。Microsoft Azure はそういった研究活動の知見を生かし、開発された、研究者のデータ収集・分析処理をサポートします。コンピュータサイエンスが専門ではない方でも使いやすいツールとなっており、Microsoft Azure を使ってこれまでできなかったデータ分析に挑戦できます。

●第33回リバネス研究費 日本マイクロソフト賞募集開始！

Microsoft Azure を用いて学術研究を加速させる、または社会課題を解決する研究アイデアを募集します。採択者には研究費および総額約1億円分のクラウド利用権が付与されます。クラウドコンピューティングを用いて研究を加速させるチャンスです。皆様のご応募をお待ちしております。(初めての方には、使い方のサポートも致します)

詳細はこちらから：<https://r.lne.st/2016/09/01/microsoft33/>

