

# 情報資源から 生物資源をつくり出す

生命科学研究においてデータベースの重要性は高まり、その活用が研究の成果を左右する時代になっている。

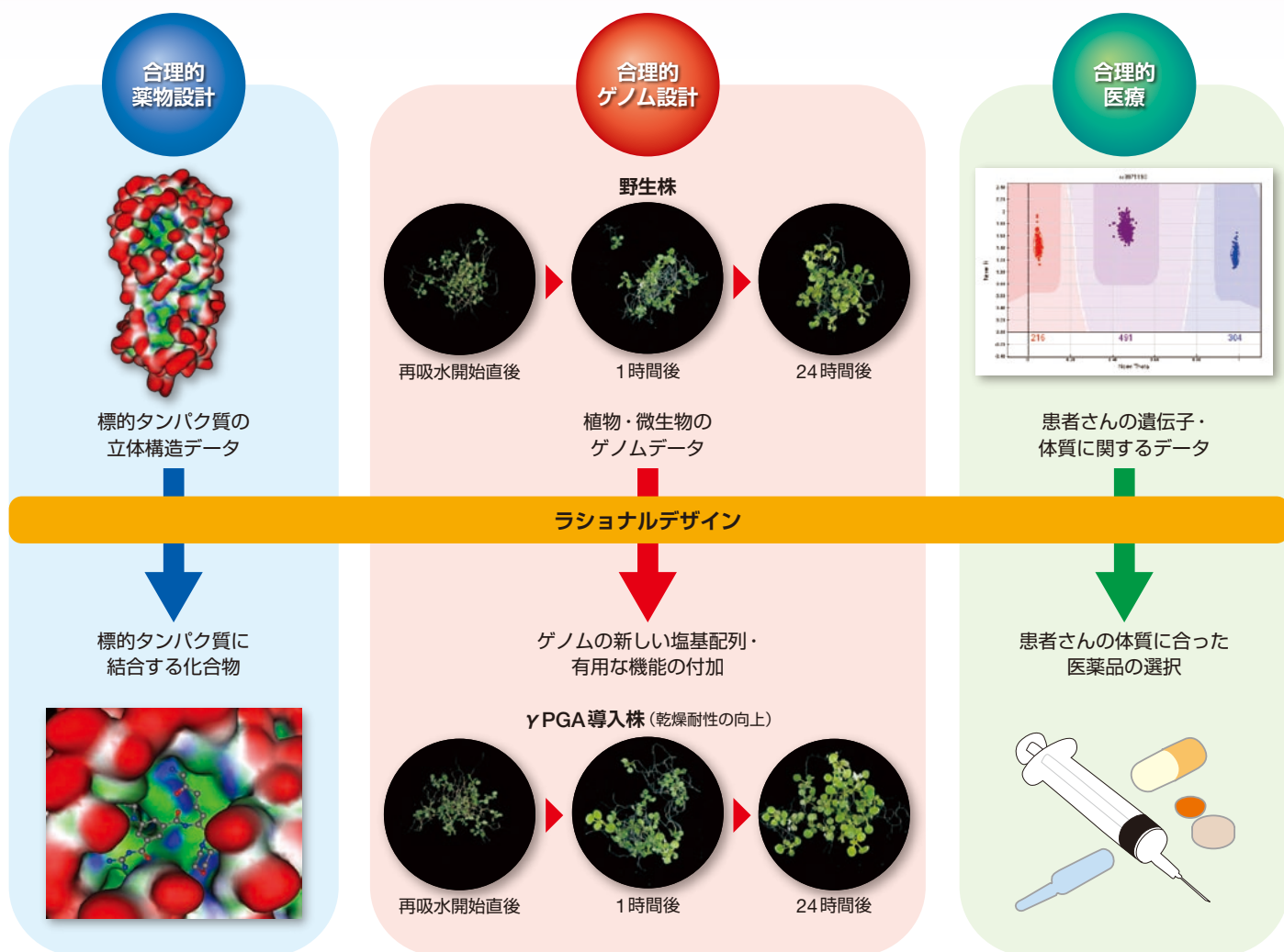
生命情報基盤研究部門では、日々の研究活動で生み出される膨大なデータを、コンピュータで自動解析可能なデータベースとして公開するための情報基盤“サイネス”を開発し、世界から注目を集めている。

「データベースは単なるデータの入れ物ではありません。新しいタイプの生命進化を生み出す場です。

データベースから有用な遺伝子を選択し、ゲノムを設計して生物に戻すことで、情報資源から有用な生物資源をつくり出すことができるようになってきました。では、どのようなデータベースがあれば

合理的な設計“ラショナルデザイン”を実現できるか。それが私たちの課題です」と、豊田哲郎 部門長は語る。

サイネスを利用してゲノムのデザイン技能を競う、第1回“合理的ゲノム設計コンテスト (GenoCon)”も開催中だ。



**ラショナルデザイン**

論理的なプログラムをつくり、データベースの情報に基づいて薬物や医療、ゲノムを合理的に設計する“ラショナルデザイン”が行われている。豊田部門長らは、納豆のネバネバ成分 ( $\gamma$ PGA) をつくり出す三つの酵素群を合理的ゲノム設計に基づいて実験植物であるシロイヌナズナのゲノムに導入した。 $\gamma$ PGA導入株を乾燥状態に置くと、野生株と同様に乾燥するが、水を与えると吸収率が高く、葉の成長が良くて生存率も高い。

合成生物学には強力なデータベース基盤が不可欠。  
 情報資源から環境と社会を支える新しい  
 生物資源を創造することを目指しています。

## 豊田哲郎

生命情報基盤研究部門 部門長



とよだ・てつろう。1968年、東京都生まれ。薬学博士。東京大学大学院薬学系研究科博士課程修了。医薬分子設計研究所、理研ゲノム科学総合研究センターゲノム知識ベース研究開発チーム チームリーダーを経て、2008年4月より現職。専門はタンパク質立体構造に基づく合理的薬物設計、バイオインフォマティクス、情報合成生物学。

### ■ ラショナルデザインの時代へ

「将来、油田開発からゲノム開発へ、資源の源泉がシフトしていきます」。豊田哲郎部門長の第一声だ。ゲノムにはさまざまな物質を生命活動で生み出すための情報が書かれており、「第二の油田」とも呼ばれている。

豊田部門長は1990年代、民間の研究所でマラリア治療薬の開発をしていた。「薬の開発では、疾患に関連するタンパク質にちょうど結合してその働きを調整できるように、化合物の形を設計することが重要です。しかし、タンパク質の構造は複雑で、やみくもに設計してもうまくいきません。そこで、論理的なプログラムをつくり、タンパク質の立体構造の情報に基づいて合理的な設計“ラショナルデザイン”が行われています」。しかし、完璧な設計ができて既存の有機化学の技術では、その化合物をつくれなことがある。豊田部門長は何度も壁にぶつかっていた。

一方、2000年代になると、ヒトをはじめさまざまな生物のゲノムの解読が本格化した。「ゲノムには4種類の塩基の並び方によって遺伝情報が書かれており、遺伝子領域の塩基配列をもとにタンパク質が作られます。ゲノムの塩基配列を自在に設計できたら、化合物を使うより確実に生物に新しい機能を持たせることができる。そのための情報資源が急増していました」

同じころ、医療の分野でもラショナルデザインが注目され始めていた。患者さんの遺伝子を調べ、その情報をもとに一人ひとりの体質に合った医薬品と治療を設計しようというもので、“オーダーメイド医療”とも呼ばれる。データベースを使った合理的アプローチはあらゆる分野に波及し、それまでの勘と経験のみに頼る古いやり方を変えつつあった。

「ゲノムのラショナルデザインの時代が来る」。そう考えた豊田部門長は2001年、理研ゲノム科学総合研究センター（GSC）にゲノム知識ベース研究開発チームを立ち上げた。当時、GSCでは和田昭允<sup>あきよし</sup> 初代センター長のもと、生命を丸ごとデータ化することで生命を総合的に理解しようとする壮大な試みが世界に先駆けて行われており、大量のデータを生み出していた。そのとき、豊田部門長にはデータベースが新しい生命進化の場のように見えていた。

### ■ データベースは生命進化の場

生命は、遺伝情報の複製と自然淘汰による選択を繰り返して進化してきた。遺伝情報の記録はRNAとDNAに限られていたが、今はデータベースにも記録されている。「生命情報

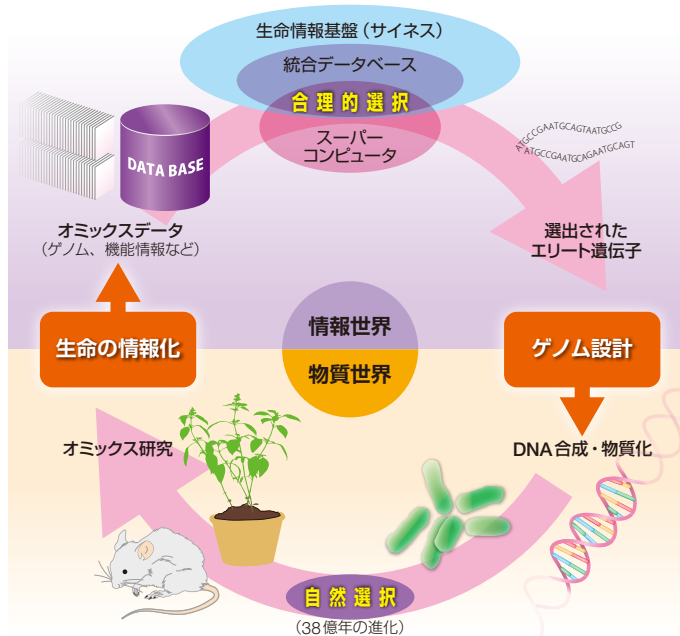


図1 オミックスで情報空間にまで広がった遺伝子進化のサイクル  
 記録媒体の種類を乗り越えて永続しつづける情報が生命の本流であり、「利己的な遺伝子」とも呼ばれる。約38億年の間、遺伝情報はRNAやDNAなど物質の世界だけで複製と選択を繰り返し、進化してきた。近年のオミックス技術革命の結果、遺伝情報はデータベース上でも複製と選択が可能になり、生命進化の場が情報世界にまで広がっている。情報世界で有用な遺伝子を選び、適切な安全管理基準のもとで生物に戻すことで、有用な生物資源が新たに創造される。

の記録媒体が、物質世界から情報世界に広がったのです。生命進化が大きく変わった瞬間に、私たちは遭遇しています」

データベースに記録された遺伝情報は、インターネットを通じて地球規模で複製され続けている。また、データベース上で有用な遺伝子の選択もできる。そして、その遺伝子をさらに良いものへと合理的に設計し、DNA合成技術を使って実際に生物のゲノムに導入すれば、新たな有用生物資源をつくり出すことができる。「データベースは遺伝情報の複製と選択を行う場、つまり新しい生命進化の場なのです」(図1)

それは夢ではなく、生命はすでにデータベースによって進化を始めている。「私たちは、データベースからγPGAという納豆のネバネバ成分をつくる酵素群を選び、合理的なゲノム設計を行って植物に導入することで、乾燥に強い植物をつくり出すことに成功しました」(8ページの図中央)

再生医療に応用できると期待されるiPS細胞(人工多能性幹細胞)も同様にして生まれたものだ。京都大学の山中伸弥教授らは、理研でつくった完全長cDNA(DNAの遺伝子領域が転写されたmRNAを鋳型にしてつくったDNA)のデータベースを使い、あらゆる細胞に分化できるES細胞(胚性幹細胞)で発現している遺伝子を選び出した。そしてヒトの皮膚細胞のゲノムに導入することで、iPS細胞という新しい生物資源をつくり出すことに成功したのだ。

「データベースをデザインすることは、生命進化の場をデザインすることと同じです。情報資源から地球と社会を支える新しい生物資源を創造するにはデータベースをどうデザインすべきか、進化の場という視点で見ると面白い」

### ■ データ共有の仕組みづくりは世界的な課題

豊田部門長が率いる生命情報基盤研究部門(BASE)は、GSCの再編成によって、2008年4月にスタートした。「BASE誕生のきっかけは、2006年に行われた第6回の理研アドバイザリー・カウンスル(RAC)でした」

RACとは理研が独自に開催している外部評価制度で、世界的な科学者・有識者が理研の活動全般を評価し、理事長に対して提言を行う。「2006年のRACでは、“理研には100以上のデータ公開ウェブサイトがあり、それぞれの質は高いが、提供方法に問題がある”と指摘されました。“きちんとデータ公開しているのになぜ”と、みんなが戸惑いました」

理研のほとんどのデータ公開ウェブサイトは人が直接閲覧するためのものであり、コンピュータが自動的にアクセスして解析することができない。また、多様なデータを標準化して共有化していく組織的な体制もなく、非効率的であり、研究成果の総合的な見せ方としても不十分である。これがRACが指摘した問題点であり、“その問題点を解決しなさい。そのためにデータベースの専門家を置くべきである”という提言を受けた。RACが指摘した問題点は、世界

中のデータベースに共通していた。「RACはどこも成功していない課題の解決を、理研に求めてきたのです。そして、私が理研におけるデータベースのディレクターとして、この問題に取り組むことになりました」

データ公開サイトの多くは変化の速いウェブ標準規格の動向に付いていけず、このまま無秩序に増え続けたら收拾がつかなくなる。データベースの維持コストも大きな負担になっている。だから統合が必要なのだが、「それが難しい」と豊田部門長。「統合化の失敗を山ほど見てきました。すでに何百とあるデータベースをすべて標準化し直すことは不可能です。そこで、国際標準規格を備えた万能な入れ物をつくり、その中にデータを移すだけで自動的に標準化して収録・公開でき、共有化を容易にする新しい発想の統合データベースの開発にチャレンジしました」

### ■ 世界が注目する“サイネス”

豊田部門長がまず手掛けたのが、ライフサイエンス分野を主体にしたデータベースの総合インキュベーション基盤システム“サイネス”(SciNeS: Scientists' Networking System)である(図2)。「サイネスの最大の特徴は、次世代ウェブの国際標準規格“セマンティックウェブ”と“クラウドコンピューティング”を採用したことです」

セマンティックウェブとは、現在広く使われているワールドワイドウェブ(WWW)の発展形だ。WWWは、ネットワーク上に置かれた文書などをハイパーリンクでつないでいるため、人が文書を読み、内容を解釈してリンクをたどりながら調べていくのには適している。しかし、ハイパーリンクは位置情報だけで、文書の間はどういう関連性があるかまでは表現していないので、コンピュータが自動でアクセスして解析することはできない。一方、セマンティックウェブでは、すべてのデータに意味を付け、リンクにその関連性を表現している。それならば多様なデータでもコンピュータが自動アクセスして解析が可能だ。

クラウドコンピューティングとは、ウェブブラウザを介してアプリケーションを使うコンピュータの新しい利用形態である。研究者は自分でサーバを維持する必要がなく、サイネスの中にバーチャルラボをつくり、データを入れていくだけでいい。データを自動的に処理して、国際標準規格のデータベースとして公開してくれる。「論文の発表には学術雑誌というメディアが発達していますが、データベースには専門のメディアがありませんでした。サイネスはデータベースを発表する、世界初の学術メディアでもあります」

2009年3月にサイネスの運用開始が発表されるや、大きな注目を集めた。「セマンティックウェブは以前から知られていましたが、大規模なデータベースをつくるのは難しいといわれていました。項目ごとにセキュリティーを管理できる新しい技術を加えることで、私たちが初めてそれを実現し

