

バイオインフォマティクスの現状と動向

東京大学大学院農学生命科学研究科 教授
清水謙多郎

DNA^(*)が運び手となる遺伝情報は、複製によって世代から世代へと伝えられ、生体内では、DNAが複製によって2本鎖となり、RNAに転写され、タンパク質^(**)に翻訳される。タンパク質は固有の構造をもち、それらが複雑に相互作用することで、個体を形成し維持している。このような生命現象の解明に情報学(インフォマティクス)の手法や技術が盛んに用いられるようになってきている。

1980年代にヒトゲノムプロジェクトが発足して以来、大量のデータを処理するために、新しい情報処理技術の開発と情報インフラストラクチャの整備が進められてきた。現在では、ヒトをはじめと、さまざまな生物のゲノム(遺伝子のセット)が解読されているが、そうした中で、実用的なデータベースや解析ツールが開発されている。そもそも、生命科学では、生物という複雑な対象を研究するのに、データを蓄積し、それを解析することで、体系や規則(埋もれている情報)を抽出して、新しい知識を得るというアプローチがとられる。データが大量かつ多様に発生している現代の生命科学において、インフォマティクスはなくてはならないものとなっている。

バイオインフォマティクスは、こうした必然的ともいえる状況の中で登場した学問であり、まさに、20世紀に飛躍的な発展を遂げた情報学と生物学との融合によってできた学問である。バイオインフォマティクスには、大きく分けて、「ゲノム情報解析」、「遺伝子ネットワークの解析」、「タンパク質の構造の解析」の3つの領域がある。以下、それぞれの領域について概観し、さらに、我々が研究を行っている「タンパク質の構造の解析」について、多少詳しく説明する。

ゲノム情報解析とバイオインフォマティクス

バイオインフォマティクスは、もともとゲノム解析によって発生した大量かつ多様なデータを処

理するために必然的に生まれたもので、当初は実験プロジェクトをサポートすることを主目的としていた。ヒトゲノムは約30億ある塩基対から構成されているが、その塩基配列を決定するための大量のデータの解析には、インフォマティクス技術の導入が必要不可欠であった。ゲノムの塩基配列の決定では、シーケンサ(DNA配列決定自動化装置)から得られる断片的なDNA塩基配列を、コンピュータを用いてつながりのDNA配列に構成する作業が行われる。このとき、複数の断片で重複して決定される部分を検出し、これらを高等生物によく見られるDNA配列の繰り返し構造を考慮しつつ、つなぎ合わせる操作が必要となる。読みとり時のエラーなども考慮しなければならない。

また、決定されたDNA配列からコンピュータを用いて自動的に遺伝子領域を予測する技術が求められた。ヒトのような高等生物では、遺伝子の複数部分がつなぎ合わさって1つのタンパク質に翻訳される。そこで、DNA配列のどの部分がタンパク質に翻訳されるかを推定する遺伝子領域予測が必要となる。遺伝子領域予測では、特定の塩基の並びや塩基の出現頻度、特徴的なパターンを検出する操作が行われ、コンピュータの学習によってこれらを検出できるようにしているものもある。

こうした、DNA塩基配列の決定や遺伝子領域の予測には、膨大な計算量と記憶領域を必要とする。ヒトゲノムの配列決定を推進したセラ・ゲノミクス社では、2001年当時、合計1.7 Tera flopsの処理能力(900台以上のノードから構成されるクラスター型コンピュータ)と合計400 Tera Byteのディスク記憶を装備していた。

2003年3月にヒトゲノムの解読が完了したのをはじめ、マウスやイネ、シロイヌナズナ、さらに150種近くの微生物のゲノムの解読が完了し、さまざまな生物のゲノムの配列が、遺伝情報データベースに登録されている。しかし、DNAの塩基配列から遺伝子に対応するタンパク質のアミノ酸

配列が得られたとしても、ある生物について理解するには、それだけでは不十分であり、そのためには、翻訳されるタンパク質の機能に関する情報が必要である。

タンパク質の機能を調べる最も簡単な方法は、配列の類似性を用いる方法である。すべての生物は共通の祖先から分岐して進化してきたと考えられるが、数十億年という長い年月の間で、各生物種の塩基配列は大きく変化してきた。同じ機能をもつ遺伝子については、突然変異による配列の変化が蓄積されている。つまり、同じ機能をもつ遺伝子は、古い年代に分化した生物種間ほど変異の差が大きくなる傾向にある。逆に言えば、配列が似ているほど、比較的新しく分化した遺伝子であり、機能も似ているという推測ができる。このように、共通の祖先をもつことにより配列が似ていることをホモロジーがあるといい、そうした遺伝子やタンパク質を検索することをホモロジー検索という。すでに機能のわかっている遺伝子は、データベースに特徴が登録されているので、新たに見つかった遺伝子の特徴をもとに検索して比較、機能を予測することができるのである。図1は、いくつかの生物種の、タンパク質シトクロームCのアミノ酸配列を比較したものである。古くに分岐した生物ほど、ヒトとの変異が大きいがわかる。アミノ酸一致率は、マウス91%、ミツバチ73%、イースト菌64%、テトラヒメナ46%、RP25336%で、チンパンジーはヒトと100%同じである。

配列の類似性は、配列を並べて比較することにより評価する。塩基もアミノ酸も文字で表現されるので、配列の比較は文字列の比較の問題に帰着される。アミノ酸の場合は、進化の過程で変異が起きやすいものはそれだけ似ているとして、単に一致しているかどうかではなく、「変異のしやすさ」を示すスコアで評価する。また、進化の過程で、塩基あるいはアミノ酸の挿入や欠失が生じることがあり、そうした部分はギャップ (-) で対応づける (図1)。文字列の類似性を評価するのに、最適解を求めるには動的計画法が用いられるが、データベースの検索には時間がかかるため、通常は近似手法が用いられている。DNAやタンパク質の配列では、一般に機能的に重要な部分は、進化の過程で変化しにくく、特定の配列パターン

をもっていることが多い。このような部分をモチーフといい、それを登録したデータベースも用意され、機能を調べる際に重要な手がかりとなっている。

```

ヒト      @OVERGKVFLLMCCQCHTVRGGHRTGPHLHGLFGKTKQAFQSTYAAHNNKGIIGHEITLMEYLNPKETIPOTRFVGHKKEERASGLATLEKATNE
マウス   @OVERGKELFYQACQCHTVRGGHRTGPHLHGLFGKTKQAFQSTYAAHNNKGIIGHEITLMEYLNPKETIPOTRFVGHKKEERASGLATLEKATNE
ミツバチ @PEERGGELFYQACQCHTVRGGHRTGPHLHGLFGKTKQAFQSTYAAHNNKGIIGHEITLMEYLNPKETIPOTRFVGHKKEERASGLATLEKATNE
イースト菌 @AERGAATLFTKTRLCQCHTVRGGHRTGPHLHGLFGKTKQAFQSTYAAHNNKGIIGHEITLMEYLNPKETIPOTRFVGHKKEERASGLATLEKATNE
テトラヒメナ @AAGKROIFDQSCCHALK--@QSTAFPLVGGVIGKAKQEK--PATEHRRGGGQITWHEITLMEYLNPKETIPOTRFVGHKKEERASGLATLEK
RP253      REIAKRCGLCHSLKLDGPNLGRRLNINIVGRKASITDVEYSFAISKGLGVDNENLFAFLHEPSTAPGTGHSFAGISEKPDLDLVLLFLNNYV

```

図1. シトクロームCのアミノ酸配列の比較

このように、DNAやタンパク質の配列は、過去の進化の情報を含んでいる。そのため、それらの配列の類似性から進化の過程を推定することができ、例えば、クラスタ解析で、類似した配列から順にまとめていくことにより、進化系統樹を形成することができる。また、生物種の違いが遺伝子のレベルでどこが違っていることによるものを明らかにすることもできる。過去の生物の化石によらず、現存する生物の分子 (配列) を比較することで進化に関する情報をさかのぼって得ることができるのである。

遺伝子ネットワークの解析

個体や細胞の機能を理解するには、遺伝子や遺伝子の産物であるタンパク質同士がどのように関係しているかという相互作用に関する情報が必要である。そこで、それらの相互作用の関係をつないでネットワークを構築することにより、生命のはたらきをシステムとして理解しようとする研究が行われている。こうしたネットワークは、反応の経路 (パスウェイ) を示すもので、遺伝子産物のタンパク質や代謝産物の低分子化合物、多糖などが生成されていく関係やそこに関与する酵素などが記述される。

例えば、生物は、生きていくためにいろいろな物質を代謝している。代謝とは、外界から摂取した物質を素材として生命活動に必要な物質を合成したり、細胞内の物質を分解してエネルギーを獲得したりすることをいう。代謝では、酵素が基質に作用し、別な化合物に変換していく過程が繰り返される。このような酵素反応のパスウェイをコンピュータ内で構築し、代謝のメカニズムを理解しようとする研究が行われている。

また、細胞は、増殖や分化など、必要な機能の変化はほとんど外界からの情報によって引き起こされている。細胞は、このようなシグナルを受けると、これを電気信号（細胞内電位の変化）に変換し、これによってタンパク質のリン酸化や脱リン酸化、さらにそれに伴うタンパク質の活性化、不活性化が引き起こされる。これをシグナル伝達といい、環境の変化に対する細胞の応答、栄養源の獲得、また細胞間のコミュニケーションの基本的な手段となっている。こうしたシグナル伝達も、分子間相互作用のネットワークとして記述し、システムとして理解しようという研究が行われている。

こうした研究において、基本となるのが、パスウェイをデータベースに登録する作業である。代謝やシグナル伝達の経路は、これまでも出版物として参照することができた。しかし、それでは、特定の化合物や酵素の名前からパスウェイを検索したり、さらにそれをもとに高度な解析をしたりすることは困難である。データベース化することにより、生物種間のデータを比較することで、ネットワークを補完したり、それぞれの生物の特徴を明らかにしたりすることも容易に行える。さまざまな生物種のゲノム解析が行われ、情報が蓄積してきている現在、データベースの有用性はますます増大している。京都大学バイオインフォマティクスセンターが開発し公開しているパスウェイデータベースKEGG⁽³⁾などがパスウェイのデータベースの例である。

パスウェイは、コンピュータ上での解析を容易にするため、グラフで表現されることが多い。グラフ理論を用いることで、大規模なシステムでも、効率的に経路を探索したり、グラフの類似性を調べたりすることができる。また、単に解析するだけでなく、構築したパスウェイをもとに、生化学的なパラメータを与えて、パスウェイの動的な挙動をシミュレートする研究も行われている。こうした研究を推し進めれば、細胞全体のシミュレーションにつながる。

パスウェイのデータベースの構築にあたっては、相互作用に関する情報を、学術論文やデータベースのアノテーション（機能に関する注釈）をもとに抽出する。専門家が一つ一つ吟味できれば信頼性も上がるが、膨大なデータに対応するのは困難であるため、自然言語処理技術を用いて自動

的に抽出するシステムが利用されることもある。

また、遺伝子のネットワークを解明するために、DNAマイクロアレイを用いた実験が行われている。DNAマイクロアレイは、専用の基板の上にDNAサンプルを二次元アレイ状に高密度に配置し、同時にそれらの発現パターンを解析できるようにした装置をいう。DNAマイクロアレイは、例えば、ゲノム中の全遺伝子を用意しておき、異なるcDNA（mRNAからDNAを逆転写したもので、翻訳に必要な塩基配列のみを含む）と二本鎖を形成させることで、さまざまな条件下における遺伝子発現プロファイルの変化を調べることができる。ある遺伝子の発現量の変化が、他の遺伝子の発現にどのような影響を及ぼしているのかも調べることができ、その結果、どの遺伝子がどの遺伝子を活性化したり、抑制したりするかが分かる。また、特定の遺伝子の変異パターンのすべてを用意しておき、個人のDNAのパターンと照合して、病気の診断や治療に利用したりすることができる。

DNAマイクロアレイで、DNAを並べる数ミリ四方のシートをチップと呼ぶ。チップという用語が用いられるようになったのは、半導体チップの作製に使われているフォトリソグラフ技術を用いてシリコンウェハース上の細かい升目の中にオリゴペプチド、さらにオリゴヌクレオチドを合成することに成功したのが始まりである。DNAを高密度に精確に配置するのに、プリンタで使われているインクジェット方式が威力を発揮している。

DNAマイクロアレイのデータ解析では、スポットの画像データから発現強度データへの変換、ノイズを排除し意味のあるデータを抽出するための統計的手法、それに関連した遺伝子の発現データの可視化、関連する遺伝子を発見するためのクラスタリングなどの処理でインフォマティクスが利用されている。

タンパク質の構造と機能

タンパク質は、そのアミノ酸配列に基づき、一定の環境下で固有の構造をとる。それらの構造に基づいて、非常に高い触媒機能をもったり、ある基質を特異的に選択して反応したりする。タンパク質の機能は、その構造によって発現するので、タンパク質の機能を把握するのに、タンパク質の

構造に関する情報は非常に有用である。

現在、タンパク質の構造を網羅的に決定し、これらの機能を解析する国家的プロジェクト（タンパク質3000）が進められている。タンパク質の構造は、X線結晶構造解析やNMRによって決定されるが、これらの作業には多くのコストと時間を必要とする。そこで、代表的なタンパク質が構造決定の対象として選ばれている。代表的とは、タンパク質のアミノ酸の配列を互いに比較し、類似性の高いものをグループにまとめて（クラスタ解析による）、それぞれのグループの中から代表を選ぶ、というもので、さらに、ゲノム解読が完了したヒトなどの生物のタンパク質で、大量にサンプルが得られ、機能的に重要と考えられるものを優先させる。

クラスタ解析における配列類似性の基準としては、たとえば、アミノ酸一致度（配列を構成するアミノ酸のうち、一致しているものの割合）が35%以上というような基準が用いられている。その程度の一致度があれば、それらのタンパク質が進化的に同一の祖先タンパク質を起源とする可能性が高い。また、タンパク質は、進化の過程で、フォールド（基本的な骨格構造）が保存されることが経験的に知られている。上のようにグループをまとめたとき、それらのタンパク質の構造は互いに似ていると予想される。

こうしたことから、グループの中の1つのタンパク質の構造が決定されれば、それをもとに、残りのタンパク質の構造を、コンピュータによる構造モデリングによって求めることができると期待される（図2）。コンピュータによるモデリングは、X線結晶構造解析やNMRに比べてはるかに少ないコストと時間で行うことができるので、精度の良い予測が可能であれば、その有用性は非常に大きい。

また、タンパク質の機能は、分子間の相互作用によって実現されるので、例えば、酵素と基質、タンパク質とDNA、タンパク質とタンパク質などのドッキングシミュレーションは、今後ますます重要になってくると考えられる。ドッキングシミュレーションでは、分子表面の形状（凹凸）や電荷（正負）によって結合するかどうかの評価される。ドッキングシミュレーションは、医薬品開

発においても、非常に重要な技術である。

アミノ酸配列

```

MADKWTGKVTIKVQNWTDALFSLTVHAFVLEPTAGQFTKLGLEIDGERVORAYSTVNSPEINED
LEFYLVTVPEGRILSPRIALAKPGDQVQVVSARGHFFVLDVPHICTIUMKLTQPTIGPYLSI
LRLGKLDLDRFNKLVLVHAARYAADLSYLEIMQBLEKRYEKGKLRIQTVVSRETAAGSLTGRIP
ALIESGELESTIGLFMKNKESHVMVLCGNEQVVRDTQQLLKEKTRQMTGHLRRRFGHMTAEHYW
  
```

立体構造



図2. タンパク質の構造モデリング（フラボドキシンを SWISS-MODEL でモデリング）

タンパク質は、生体内では固定した構造をとっているのではなく、周囲の溶媒分子の影響を受け、常に揺らいでいる。また、分子間の相互作用によって構造を変化させることもある。このようなタンパク質の動的な構造（動き）を解析するには、分子シミュレーションを用いる。分子シミュレーションには、個々の原子を確率的に動かすモンテカルロ法と、ニュートンの運動方程式を積分して、原子の座標を時系列で求めていく分子動力学法があり、タンパク質のシミュレーションでは分子動力学法がよく用いられている。

以下では、我々の研究室の研究テーマである、タンパク質の構造モデリングと、シミュレーションについて、もう少し詳しく説明する。

タンパク質の構造モデリング

タンパク質の構造モデリング手法としてよく用いられているのは、ホモロジーモデリングである。ホモロジーモデリングとは、構造未知のタンパク質（ターゲット）の構造を、構造既知のタンパク質の中から手本（テンプレート）となる構造を選んで、それをもとに予測する手法をいう。

ホモロジーモデリングは、良いテンプレートが得られる場合には、比較的精度の高い予測ができる。では、こうしたテンプレートが得られない場合はどうすればよいだろうか？先に述べたような、タンパク質の構造決定のプロジェクトが進むにつれ、そうしたケースは実際には少なくなってくると予想される。しかしながら、現在でも、新しいフォールドは発見され続けており、また、ターゲットのタンパク質のすべてにわたって良いテ

ンプレートが得られるとは限らず、良いテンプレートが得られない部分は、モデリングで補う必要がある。

そこで、タンパク質の構造モデリングにおいて、新規フォールド法と呼ばれる別なアプローチが用いられることがある。新規フォールド法は、文字通り、既知の構造にない、新しいフォールドも予測できることを目的としたもので、最もチャレンジングな手法は、ターゲットのアミノ酸配列だけからテンプレートを用いずに、タンパク質の立体構造を予測するというものである。そもそも、アンフィンゼンの仮説に従えば、天然状態のタンパク質は、自由エネルギーを最小とするような構造にフォールドされているはずである。(アンフィンゼンは、タンパク質の変性実験を通して、タンパク質の構造は壊れても元の構造に戻ることができること、そうした能力は、物質世界を支配する熱力学の自由エネルギー最小則に基づいていることを証明した。)しかし、コンピュータ上で自由エネルギーが最小の構造を見つけるには、多数のコンフォメーションを生成する必要がある、現在のコンピュータの処理能力では膨大な計算時間を必要とする。また、物理ポテンシャルの最小構造を探索すること自体、分子のコンフォメーション空間が広大でエネルギー曲面の形状が複雑である(多数の極小点が存在する)ことから、非常に困難になっている。

近年、新規フォールド法として最も成功を収めている手法は、フラグメントアセンブリ方式と呼ばれるものである。フラグメントアセンブリ方式は、特定の長さ(3残基, 9残基など)の短いフラグメント(配列断片)がとり得る部分的な構造を求め、それらを組み合わせてタンパク質全体の構造を構築し、それらを統計ポテンシャルで評価して、その最小構造を探索するというものである。統計ポテンシャルとは、タンパク質の構造決定には、いろいろな物理的要因が働いているが、それらを個々に計算しようとせず、構造データベースの内容から、とり得る構造の傾向を確率・統計的に算出し、ポテンシャルとして定義したものをいう。例えば、タンパク質分子の中での個々のアミノ酸の埋もれ度、アミノ酸間の距離、構造のコンパクトさなどの構造特徴に対して、それらのとり

やすさを確率の $-\log$ をとってポテンシャルの形で表わしたものをいう。構造モデリングでは、これを最小にする(確率が最大になる)構造を探索することになる。統計ポテンシャルは、英語では、**knowledge-based potential, database-derived potential**と呼ばれており、まさに経験に基づく、データベース(既知の情報)から導出されるポテンシャルである。

我々の研究グループでは、ABLEという、フラグメントアセンブリ方式に基づくタンパク質構造モデリングシステムを開発している。図3に、その基本的なモデリング手順を示す。

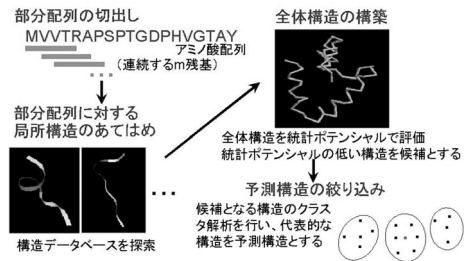


図3. フラグメントアセンブリ方式による構造モデリング

タンパク質分子のシミュレーション

タンパク質の動きをシミュレートするのに最もよく用いられているのは分子動力学法である。分子動力学法は、多原子系における原子の運動を、原子間に働く力を計算しながら、個々の原子に対するニュートンの運動方程式 $F_i = m_i \frac{d^2 r_i}{dt^2}$ を積分することにより求める手法である。このような積分は解析的に計算することが不可能であるため、有限差分法を用いて数値的に計算する。その基本手順は、粒子の位置や速度を短い時間刻みで、離散的に計算していくというものである。すなわち、ある時刻において原子間に働く力を計算し、それに基づいて原子の速度と座標を求め、その求めた位置で新たに原子間に働く力を計算する、という操作の繰返しである。シミュレーションの条件として、温度や圧力を一定にする場合、その調整のための操作も繰返しに含まれる。

原子に働く力は、

$$F_i = - \sum_{j \neq i} \text{grad}_{\theta_{ij}}(\eta_{ij}) = - \sum_{j \neq i} \frac{\partial \theta_{ij}(\eta_{ij})}{\partial r} \cdot \frac{r_j - r_i}{r_{ij}}$$

のように、ポテンシャル関数の微分で与えられる。具体的なポテンシャルとしては、共有結合ポ

テンシャル（バネの弾性エネルギーで近似することが多い）、クーロンポテンシャル、分子間力を表すファンデルワールスポテンシャルなどがある。

図4は、タンパク質BPTI（892原子）と水（15843原子）の分子動力学シミュレーションを行ったときの様子を示したものである。中央のタンパク質の周辺の小さな点1つ1つが水分子である。

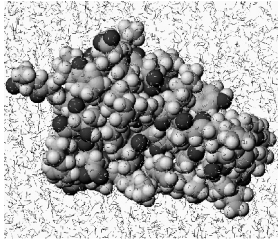


図4. タンパク質BPTIと水の動力学シミュレーション

分子動力学法は、原子の運動そのものを解析するのに用いられることもあるが、近年はむしろ長時間シミュレーションを実行して、多数のコンフォメーションをサンプリングし、それらをもとに解析を行う場合が多い。例えば、生命科学の分野において、以下のような目的に用いられている。

- (1) 構造の精密なモデリング
- (2) タンパク質のフォールディング過程のシミュレーション
- (3) 分子間相互作用の詳細な解析
- (4) 物理化学量の計算

タンパク質のシミュレーションは、大規模で複雑な系を扱い、しかも生体内に近い環境での精密な計算が必要であることから、計算量は膨大になり、多大な計算時間を必要とする。このため、並列計算による高速化は不可欠なものになっている。例えば、IBMでは、Peta flops（1秒間に1000兆回の浮動小数点演算を実行できる能力、flopsは、floating point operations per secondの略）の演算能力をもつ並列コンピュータを構築し、タンパク質のフォールディングなどに応用しようというBlue Geneプロジェクト構想を進めている。Blue Geneプロジェクトは、ハードウェア、基本ソフトウェア、アプリケーションの統合的な開発を目指すもので、主要なアプリケーションとして、分子シミュレーションを筆頭に挙げている。

まとめ

バイオインフォマティクスの分野は、非常に広く多岐にわたっている。もともと、生物学とインフォマティクスの境界領域の学問として登場したが、物理学や化学の知識はますます重要になってきている。いろいろなバックグラウンドをもつ人がいろいろな角度から取り組むべき分野である。バイオインフォマティクスは、オーダーメードの医薬品や機能性食品の開発、遺伝子治療、環境修復など、さまざまな応用が考えられ、それらは、我々の生活に結びついた非常に重要なものである。この分野にかけられる期待は非常に大きい。ただ、それだけに、慎重に解決しなければならない重要な問題が山積している。いまの高校生たちが社会の一線で活躍し始めるころ、現在、言われている応用のいくつかは現実のものとなり、バイオインフォマティクスは新たな展開を見せているかも知れない。しかし、生命現象の解明という根本的な問題も、医薬、食品、環境といった応用に関する問題も、永遠の課題と言って過言ではない奥の深い問題である。できるだけ多くの高校生がこの分野に興味をもっていただきたいと考えている。

参照

- *1) DNA（デオキシリボ核酸）は、塩基、リン酸、デオキシリボースと呼ばれる糖の3つのパーツから構成され、この3つのパーツがヌクレオチドという一つのユニットとして鎖状に繋がっている。糖とリン酸は共通で、塩基は、アデニン、グアニン、チミン、シトシンの4種類があり、それぞれ1文字で、A、G、T、Cと表記される。遺伝情報は、この4種類の塩基の並び方によって記述され、DNAは、これらの塩基の配列として表現される。DNAは通常、2本の鎖がより合わさったらせん構造をとっており、これは、それぞれの鎖の塩基と塩基が、互いにA-T、C-Gの対（塩基対）で結合することにより形成される。
- *2) タンパク質は、アミノ酸がペプチド結合と呼ばれる結合で鎖状につながった分子である。各アミノ酸は、中心となる炭素原子（Ca原子）に、水素（-H）、アミノ基（-NH₂）、カルボキシル基（-COOH）、側鎖が結合した構造になっている。タンパク質を構成するアミノ酸は基本的には20種類であり、それぞれ異なる側差をもつ。タンパク質は、これら20種類のアミノ酸の並び方によって、固有の構造と機能をもつ。その並び方を表したのがアミノ酸配列である。アミノ酸も、アラニン(A)、システイン(C)、アスパラギン酸(D)、グルタミン酸(E)のように1文字で表記され、これらの並びでアミノ酸配列を表す。
- *3) <http://www.genome.ad.jp/kegg/kegg2.html>