

## 計算科学を取り入れた生体高分子の新機能解析法

田村 隆

(岡山大学大学院自然科学研究科)

化学という学問は、経験の積重ねの学問であるとされてきた。しかし、コンピュータの高速化、小型化、そして今後の潮流である並列化によって未知の分子の性質を予測してその機能をシミュレーションすることが可能になり、さらには任意の生理活性を設計することも可能になりつつある。有用分子の候補となる分子構造が2, 3個, あるいは10個以内であれば実際に調製して調べることもできるが、候補分子が数十, 数百も想定できる場合は何らかの方法で候補を絞らなくてはならない。その結果、実験化学者は後になって「自分は金の鉱脈をまるごと捨ててしまったのではないか。」という疑念に悩まされることになる。

量子化学は、分子の構造、性質、反応性、分子間の相互作用までも解き明かす理論であり、分子の大きさによって近似のやり方に違いがあるが、おなじ原理で既知あるいは未知の分子を記述できる。その普遍性から量子論のことを「第一原理」とも呼ぶこともある。量子化学で計算される波動関数 $\psi$ は分子の性質や反応性など「分子の内面」を表す情報であり、 $|\psi|^2$ を積分して重ねると「分子の形」を三次元構造として得ることができる。化学の先人達から見れば21世紀の化学者はとても幸福である。コンピュータをよき友として、分子の立体構造、極性や磁気的性質、分光スペクトル、反応性を前もって予見してそこから実験計画と考察を重ねて研究を進めることができるからである。

分子量の低い化合物を扱う有機化学者たちは、すでに量子論とコンピュータを研究の仲間として活用しはじめている。しかし、タンパク質や糖鎖などの生体高分子を扱う分野では、そのような「幸せそうな生化学者」はなかなか見当たらない。タンパク質のような分子量が大きいものには量子化学計算は当面は無理な話で、分子を「剛球とばね」に見立てて、必要ならば剛球に点電荷を張り付けた力学モデルであるMM (molecular mechanics) またはMD (molecular dynamics) を代用してきた。これらのモデルには、結合の切断と形成が関わる化学反応は予測できないのである

が、近縁同士のタンパク質の構造モデリングやタンパク質とリガンドの結合シミュレーションには役立つと期待され、創薬研究方面で使われてきた。しかし、ビオチン-ストレプトアビジン系のように開きっぱなしのポケットにリガンドがほぼそのままの姿勢で収まり、その結合が電荷相互作用のみで決まるケースであれば、有用なシミュレーション結果を提供してくれるが、実際には、そのような系はむしろ例外的といえる。やはりタンパク質の内部で働く弱い力を記述するためには「電子はどこでどうしているのか」を知る必要がある。最近では、現実的な折衷案としてQM/MM (quantum mechanics/molecular mechanics) 法により、ある範囲までを量子化学計算して、その境界の外側はMMの剛体モデルとうまく切り分けることが行われている。ただし、タンパク質分子内では数オングストロームの長距離間でも電子伝達が起こり得るので、「どこでQM/MMの境界線を引くか。」と悩ましい課題を振り切ることができない。

量子化学計算には、原子の成り立ちから始まってすべて計算する *ab initio* (=最初から) と、原子の性質などの計算を「実験値」を代入して端折ってしまい分子の反応性に直接関わりそうな項を近似的に計算する「半経験的」分子軌道計算がある。後者の方が計算の負担が少ないので、より大きな分子量のものを手軽に計算ができる。精度のことを割り切ってしまうえば市販のノートPC程度のもので数百原子の分子も扱える。ただし、この計算の過程において電子間反発を無視するなど、近似計算をどこまで容認するか様々な提案がありユーザーはそのパッケージである「ハミルトニアン」を選ばなくてはならない。ユーザーとなる実験家の多くは、ハミルトニアンの中身まで理解する人はまづいないので、AM1, PM3などのハミルトニアンを幾通りか変えて計算して、それぞれの計算結果に共通する傾向を比較することが必要である。あとは実験家なのだから実験して確かめることになる。実験家には、仮説をたてたり考察を練ったりする上で、まるで見てきたかのような反応シミュレーションがある方が楽しいに違いない。電子同士の反発など厳密な計算項目を省略した半経験的分子軌道計算は、絶対値として正しい値を算出しないことは保証付

Introduction of MOPAC simulation for experimental biochemists

Takashi Tamura (Faculty of Agriculture, Okayama University, 1-1-1 Tsushima-naka, Okayama, 700-8530, Japan)

## テクニカルノート

きであるが、この欠点は実験家にはむしろ歓迎されるかもしれない。「実験よりも正しい」という *ab initio* 計算を提示されたならば、それはかえって重荷になりかねない。本稿では半経験的分子軌道計算 MOPAC の生化学への応用について紹介する。

### 1. MOPAC 計算のはじめ方

計算の対象となる分子の立体構造はユーザーが用意しなくてはならない。分子軌道計算を実行するソフトウェアにはパソコン上で、任意の原子を個々に並べてそれらを自由に結合する構造描画機能が付いているが、タンパク質などの巨大分子を原子レベルで立体的に正しく描くことはとても困難である。分子の三次元座標は、PDB ([http://www.pdbj.org/index\\_j.html](http://www.pdbj.org/index_j.html)) などのエクス線結晶学データベースを利用して得るのがもっとも一般的である。しかしそのようなデータがない新規構造の場合は ChemDraw (CambridgeSoft) などで描いた化学構造式を Chem3D (CambridgeSoft) で開くことで立体化ができる。Chem3D には Molecular Mechanics による構造最適化 (MM2) が実行可能なので、これで原子同士のぶつかり合いだけはまず解消できる。ここで不斉中心の立体配置が誤っていないことは確認したほうがよい。

Chem3D にも MOPAC 計算機能が搭載されているので、構造最適化だけを取り敢えず行いたい場合は Chem3D 上で MOPAC を実行しても構わない。しかし、立体配座の一部を変化させたときの生成熱の変化など、より詳細に配座変化とエネルギーの検討をしたい場合は、やはり MOPAC 計算パッケージが必要である。そこで ChemDraw で作成した構造を拡張子 .mop をつけて別名保存する。じつは MOPAC 自体は無料でダウンロード可能であり、それを使うための優れたインターフェースを提供してくれるサイトも存在する。例えば <http://winmostar.com/> からダウンロードできる WinMostar は無料のソフトウェアでありながら市販の WinMOPAC (富士通) などに匹敵する使い

勝手を提供してくれる。MOPAC 本体を無料で提供するサイトもここからリンクされている。一方、著者が使用している WinMOPAC (現 CygressMO) や CAChe (現 Cygress) などは市販のソフトである。無料で入手できるものをわざわざ市販品として購入する理由は、インターフェースの利便性、つまり二つの構造パラメータを同時に変化させた時のエネルギー平面を三次元の等高線表示ができるなど条件設定の利便性、および計算結果表示の機能性にあると著者は理解している。

### 2. 二面角を理解することが最初で最後のハードル

MM2 で構造最適化した分子は原子が互いに衝突を避けるだけの構造であり、まだ電子が入っていない「中身は空っぽの分子」である。分子軌道計算は、分子力場計算 MM2 で構成した殻に電子を流し込む作業とイメージすればよい。その作業を始めるにあたって MOPAC では分子の形を z-matrix という相対的座標によって記述する。

z-matrix では、分子中のある一つの原子 1 を原点として、それに結合する二つ目の原子 2 は [原子 1 からの距離] で表す。3 番目の原子 3 は、[原子 2 からの距離] と [原子 1, 原子 2, 原子 3 が成す角度] で表す。そして原子 4 は、[隣接原子からの距離] と [自身, 隣, その隣の原子が成す角度], そして [さらに三つ隣の原子] が成す二面角で表す。そして原子 5 以降は原子 4 と同様に、自分より若い番号の原子を選んで相対的座標を記述してゆく。図 1 のアセトアルデヒドで例示するように二面角とは 1 本のシグマ共有結合を挟んで作られる二つの平面が作る角度で表される。生化学では、二面角を「ねじれ」と定性的に表現することが多く、これを定量的に議論することは少ない。そのためか「距離や角度はわかるが、二面角がイメージしにくい。」という人も多い。二面角は結晶構造解析と同様に計算科学でも重要なパラメータなので、「最初で最後のハードル」と思って理解に努めることが必要である。二面角は、 $\alpha$  ヘリックスや  $\beta$  ストランド形成を定義する  $\phi$ ,  $\psi$

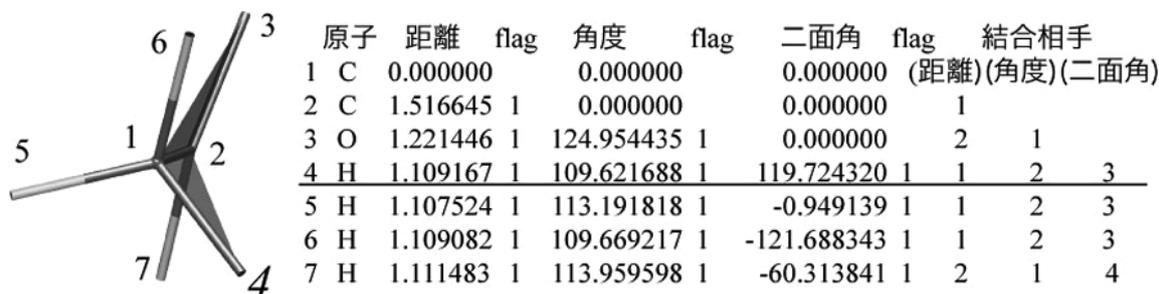


図 1 アセトアルデヒドの z-matrix

アセトアルデヒドの H4 は、C1 との距離 1.109167Å, H4-C1-C2 の角度 109.621688°, H4-C1-C2-O3 の二面角が O3 を基準にして右 (+) に 119.724320° と記述される。

と同じと言えばやや親近感を持てるかもしれない。水素 H4 を z-matrix で記述する時、C1-C2 間の結合軸に沿って C1 原子から C2 原子を眺めると、O3 を基準にして右側、つまり時計回りに水素 H4 による二面角が形成される。その角度を時計回りに 180 度までを正の数で表し、180 度を超えていけば O3 を基準に反時計回りに測って -180 度までの負の値で表す。

このような z-matrix が MOPAC 計算に分子構造を入力するときに必要とされる。自分でテキストファイルに書き出して作成することもできるが ChemDraw で作成した MOP ファイルを MOPAC に読み込んだときに z-matrix は自動的に作られる。そしてハミルトニアンを選択して分子軌道計算が始まると、より安定な構造を求めて個々の結合距離や角度、二面角の各パラメータも微妙にあるいは大きく変化してゆく。このとき構造最適化を許す場合は、z-matrix 上の距離、角度、二面角のそれぞれに付く「flag」の値を 1 とする。フラッグを 0 とすれば、そのパラメータは固定したままで他のパラメータだけが変ってゆく。さらに人為的にある部分の距離、角度、二面角を変えながらエネルギー（生成熱）の変化を見たいという場合は、ある二面角の「フラッグ」を -1 にして設定したい二面角の数値を列挙しておく、それぞれの二面角についての分子の最安定構造と生成熱を列挙してくれる。

### 3. インフルエンザノイラミニダーゼ糖鎖の MOPAC 計算

ある程度、巨大な生体分子でもこのような計算を実行することができる。糖鎖はエックス線結晶構造解析が困難でありタンパク質のように数多くの立体構造が報告されていない。その構造は「揺らぎ」の中にあり、低いエネルギー障壁で隔てられた複数の安定構造間を移ろっていると考えられている。MOPAC 計算ならばノートパソコン程度のありふれた計算機で、糖鎖の立体構造変化とそれに伴うエネルギー変化を予想できる。

インフルエンザのノイラミニダーゼは糖鎖付きのタンパク質構造が報告されている数少ない例であり、PDB に登録されている 1F8B 1I7F, 7NN9, 1XOE, 2QWA には糖鎖構造が含まれる<sup>1)</sup>。それらをアスパラギン残基に結合している N-アセチルグルコサミンを基準にして重ね合わせると N-アセチルグルコサミン a と b の間での微細なねじれの違いが d, e, f のマンノース部分の位置にまで波及していることが分かるが、これらの構造が熱的にどれがどれだけ安定なのかは、構造を眺めていても見当がつかない。

そこで a と b の間での二面角の変化によって糖鎖全体の揺らぎが得られると仮定して配座変化のエネルギーを MOPAC で計算した。そこで PDB データの一つである

7NN の糖鎖の N-アセチルグルコサミン a と b の間で二つの二面角 ( $\phi$  と  $\psi$ ) を変化させる MOPAC 計算を行った (図 2B, C)。この糖鎖構造が水溶液中にあると想定するために分子を水の誘電率 78.4 の雰囲気下におく設定もした。その入力コマンドは富士通の Cygress MO では COSMO in Water である。ただし COSMO in Water は扱う分子がほぼ球状であるときに最も有効とされており、いかなる分子でも水溶液中の溶質として計算できるわけではないので注意が必要である。ここでの糖鎖モデルは水分子を巻き込んで全体的に楕円球になるようにしている。

計算が終了したときに示される構造は、あくまで最初の構造から出発して得られる最安定構造なので、本当は全く異なるもっと安定な構造が別にある可能性も捨てきれな

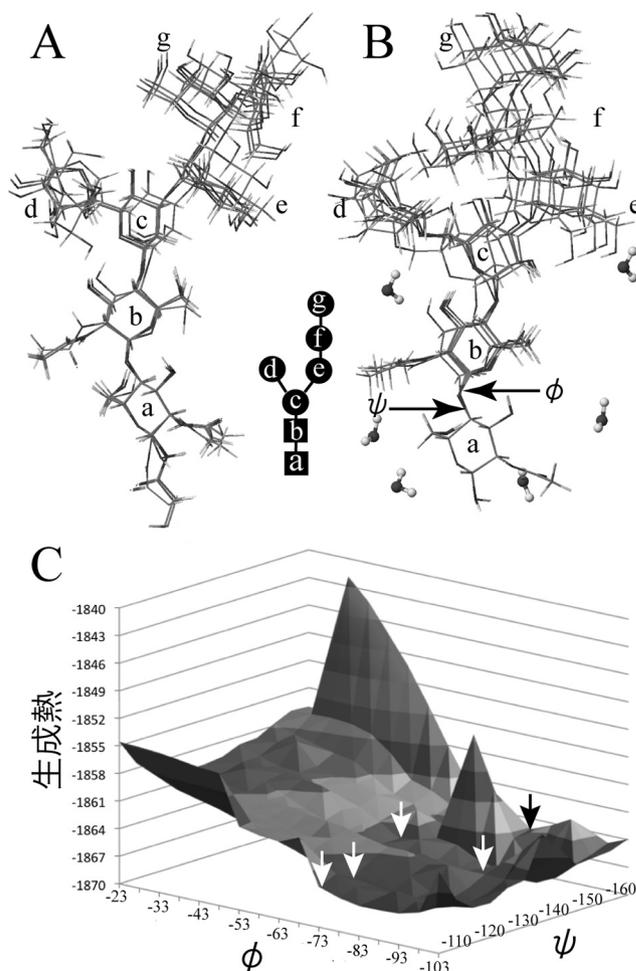


図2 インフルエンザウイルス表層のノイラミニダーゼの糖鎖構造

A. エックス線結晶構造の糖鎖構造の重ね合わせ。B. N-アセチルグルコサミン a, b 間の二つの二面角 ( $\phi$ ,  $\psi$ ) を変化させた時に得られる局所安定構造の重ね合わせ。C. 生成熱のエネルギー平面図。  $\phi$  は、-23.26 ~ -103.25 の範囲、  $\psi$  は -105.42 から -160.43 の範囲でローカルミニマム構造を算出した。

## テクニカルノート

い。この点は研究者の注意深い観察と考察が要求される。計算機はただひたすら与えられた条件での最安定構造を求めるのみである。ところで、グリコシド結合には酸素原子を挟んで自由回転できる単結合が二つ存在する。それらを変化させてほかの構造パラメータをすべて最適化したときの構造とエネルギーを立体的に描くことができる(図2C)。そこで得られた最も安定な構造を五つピックアップして重ね合わせたものが図2Bである。これらの構造の生成熱の差は1.52 kcal/mol以内なので生理的条件下ではほとんど等価ではないだろうか。この重ね合わせをアニメーションにすると糖鎖の揺らぎを動的にシミュレーションできて圧巻である。

### 4. パッカリングの微細な違いが生み出す補酵素特異性

NADやNADPは酸化還元反応に関わる補酵素であり、生物界に普遍的に分布する。その選択性は $k_{cat}/K_m$ の値として数千倍程度もの選択性になるのでNAD、NADPのどちらを使うかは、酵素特有の性質といえる。酵素を産業利用する場合、NADはNADPよりも安くても安定なのでNAD要求型ならば商品になるがNADP要求性であればよほど高い採算が見込まれないと開発の対象になりにくいという現実がある。そこで1980年代の後半から部位特異的変異導入法によりNADP型酵素をNAD型に変換するタンパク質工学研究がさかんに行われた。

誘導適合(induced fit)説で著名なD. Koshlandが結晶構造を解析した大腸菌のイソクエン酸脱水素酵素(NADP型ICDH)と大阪大学大学院の今田らが構造解析したイソプロピルリンゴ酸脱水素酵素(NAD型)は立体構造的にも極めて類似しており補酵素認識機構を解明するモデル酵素、あるいは「双子の酵素」といわれてきた。両タンパク質の立体構造を詳細に比較したHurleyらは、アデニン環を取り囲む七つのアミノ酸残基を適切に変異置換すれば大腸菌ICDHをNADP要求性からNAD要求性に変換できることを示した<sup>2)</sup>。後に今田らがNAD要求型のICDHを結晶構造解析した結果、Hurleyの目論んだアデニン環結合様式と一致していた<sup>3)</sup>。

NADとNADPの違いはアデノシンの2'水酸基についてのリン酸基である。それぞれの酵素ではリン酸基を認識するArg残基、リボース環のジオールを認識するAsp残基が同定されており、その違いは歴然としている。その違いがアデニン環の結合位置まで変えてしまうのはなぜだろうか。その理由もMOPAC計算を使って示すことができる。NAD要求型ICDHに結合したNADとNADP要求型ICDHに結合したNADPをそれぞれ取り出してタンパク質を取り出した状態でMOPACによる構造最適化をかけて重ね

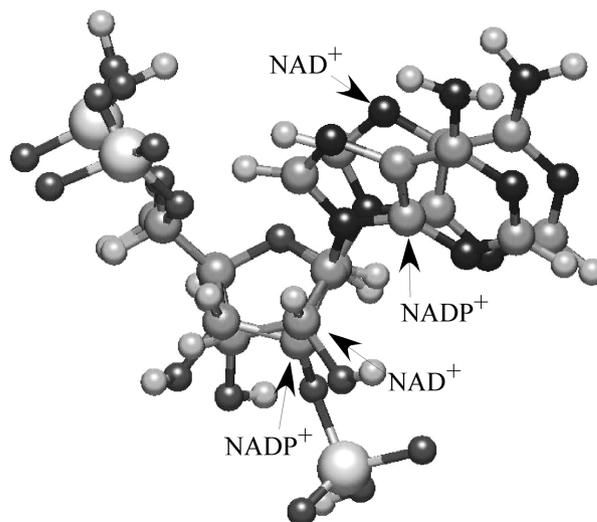


図3 補酵素特異性を決定するリボース環の安定配座解析  
イソプロピルリンゴ酸脱水素酵素(1IH6.pdb)のNADPの2'-ホスホアデノシン部分とイソクエン酸脱水素酵素(2D4V.pdb)のNADの相当部位をMOPAC計算より構造最適化を行った。

あわせてみる(図3)。MOPAC計算によりリン酸基のついたリボース環は3'endoパッカリング(3'炭素がリボース環平面から外れている)として最安定であるが、NADのアデノシンリボース環では2'endoパッカリングが安定構造として収束する。この違いはリボース環ではごく僅かな差異であるがN-グリコシド結合を介してアデニン環にまで至ると、アデノシンの位置が大きく異なる。つまり補酵素特異性を変更するには単にリン酸を認識するArg残基から2',3'ジオールを捉えるAsp残基に変更するだけでは不十分であり、アデニン環の位置までNADの最安定構造にあわせる必要がある。酵素は補酵素の最も安定な構造に忠実に自らのポケットの内装を用意しているという点で、これもinduced fitの一つの事例と言えるかもしれない。

### 5. ハイドライド形成を促進するNADアミド基の二面角

分子軌道計算は、タンパク質の構造を原子レベルの解像度から電子のレベルで観察するという点なので、化学反応に関わる興味深い知見も提供してくれる。イソクエン酸脱水素酵素が触媒するNADへのハイドライド転移は生物界に普遍的な酸化還元反応である。しかし、ニコチンアミド環でハイドライド水素が発生する理由は実はいまだ明らかにされていない。ハイドライドを放出するだけの低い酸化還元電位を持つNADHがなぜ水中のプロトンに電子を奪われないのか。考えてみると不思議な話である。これまで報告されてきた類縁の脱水素酵素の結晶構造の多くは、触媒反応が進行するニコチンアミド環周辺の電子密度が不鮮明な解析結果が多く、「ハイドライドがなぜ発生するの

か」という疑問に答えてくれる結晶構造上の知見は乏しい。ニコチンアミド環の周辺に酸化還元に関与できるアミノ酸残基が存在しないことも、ハイドライド形成の謎を深める要因であった。

今田らが報告したイソクエン酸脱水素酵素では、緩衝液中のクエン酸が本来の基質であるイソクエン酸結合箇所きれいに収まっており補酵素 NAD 近傍の C-d, D-i の二つのループも閉じてタンパク質全体として最も閉じたコンフォメーションを取っていた。そして活性中心のニコチンアミド環が珍しく鮮明な電子密度を与えたのだが、奇妙なことにそのアミド基の窒素が C4 炭素に対して *cis* 側にあ

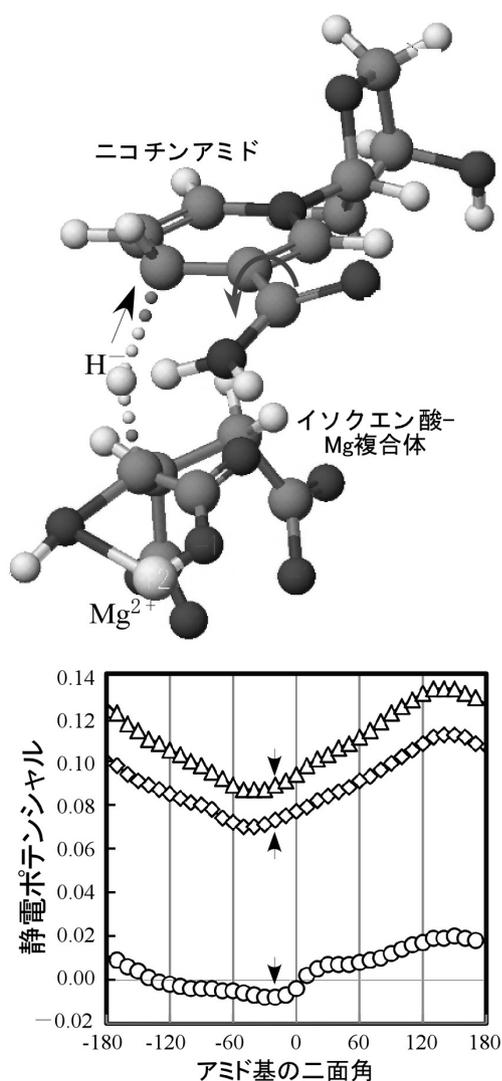


図4 ニコチンアミドの二面角とハイドライド形成  
この触媒反応は吸熱反応なので「遷移状態」はプロダクトライクな構造、つまり NADPH に近い。NADH のアミド基を回転させた時の pro-R 水素の静電ポテンシャルを AM1 ( $\Delta$ ), PM3 ( $\diamond$ ), MNDO ( $\circ$ ) で計算した。結晶構造に見られる  $-20^\circ$  付近で pro-R 水素がもっとも強く負の電荷を帯びる。

り  $-21^\circ$  の二面角で基質側に傾いていた<sup>3)</sup>。教科書で学ぶ NAD(P) のアミド基は窒素が C4 炭素に対して *trans* にある構造が一般的であり、我々が学生時代に暗記したニコチンアミドの構造でもある。そこでアミド基の二面角を変えたときの転移する水素の電荷を MOPAC で計算した結果、C4 の二つの水素原子はアミド基の  $\text{NH}_2$  が近づく *cis* 配座で最も大きな負電荷を持ち、二面角  $-20^\circ$  付近で最も大きな負電荷を帯びると予測された (図 4)。この計算によれば、教科書で学ぶ *trans* 配座の構造では C4 位の水素が負電荷を持つことはあり得ない。じつはアミド基  $\pi$  電子はピリジン環  $\pi$  電子とは共役しておらず、水溶液中ではアミド基は自由に回転できる。そして酵素に固定されて反応するときだけ二面角  $-20^\circ$  をとり、ハイドライドが入り出すという魅力的な仮説が提案された。つまり酵素の働きは、基質をニコチンアミド環の近くに配置するとともにアミド基の二面角を  $-20^\circ$  に固定してハイドライド形成を促すことにあると考えられる。

## 6. シミュレーションは第二の産業革命

コンピュータの発達によって、省力化、省コスト、省エネの恩恵を得ることができるのは化学の分野だけにとどまらない。コンピュータは、機械、電気、新素材などあらゆる科学分野の研究スピードと効率を飛躍的に高めており、シミュレーションは第二の産業革命ともいわれている。しかし計算科学の理想的な用法は、実験者自身が計算に携わることである。実験者は、自身の扱う分子について何が本質的でなにが不要であるかを知っている。本稿で紹介した簡便な計算シミュレーションに馴れば、さらに大型のシミュレーションへも楽々ステップアップ可能である。現在、QM/MM 法のような部分的な量子化学計算ではなく、タンパク質分子の全電子計算を *ab initio* で計算するプログラムも開発されつつある<sup>4)</sup>。そのプロジェクトでは実験生化学者をユーザーとして想定しており、PC 端末から Py-MOL や PDB Viewer 程度に操作性の高い GUI (graphic user interface) を介して、神戸に建設中の次世代型スパコンに計算指令を投げて、返ってきた結果を視覚化して美しく表示するシステムである。微分積分も忘れてしまった著者のような非理系の生化学者でも幸せな研究生活を送れる計算システムが構築されることを期待している。

- 1) Smith, B.J., Colman, P.M., Von Itzstein, M., Danylec, B., & Varghese, J.N. (2001) *Protein Sci.*, 10, 689-696.
- 2) Hurley, J.H., Chen, R.R., & Dean, A.M. (1996) *Biochemistry*, 35, 5670-5678.
- 3) Imada, K., Tamura, T., Takenaka, R., Kobayashi, I., Namba, K., & Inagaki, K. (2008) *Proteins*, 70, 63-71.
- 4) <http://www.ciiss.iis.u-tokyo.ac.jp/>