

論文 / 著書情報
Article / Book Information

題目(和文)	電子波動関数の解析によるタンパク質の構造構築原理と機能部位予測
Title(English)	
著者(和文)	福島健太郎
Author(English)	Kentaro Fukushima
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第7246号, 授与年月日:2008年3月26日, 学位の種別:課程博士, 審査員:櫻井 実
Citation(English)	Degree:Doctor of Engineering, Conferring organization: Tokyo Institute of Technology, Report number:甲第7246号, Conferred date:2008/3/26, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

電子波動関数の解析によるタンパク質の構造構築原理 と機能部位予測



300065481

東京工業大学附属図書館

06-5734-2097

生体分子機能工学専攻

福 島 健太郎

目次

第1章 緒言	2
第2章 電子波動関数の解析に基づくタンパク質の構造-機能相関と機能部位予測法の開発	28
第3章 超分子計算による水和状態タンパク質の電子波動関数の解析	59
第4章 タンパク質のフォールディング過程における電子波動関数-構造相関の解析	85
第5章 総括	111

謝辞

第1章 緒言

1. 1 ポスト構造ゲノム

1. 1. 1 全塩基配列の決定

1991年にスタートしたヒトゲノム計画は、人間の全生命情報の核となる核酸塩基の配列決定と発現機能の解析が目標だった。当初ヒトゲノムの全塩基配列決定は2010年以降になると予想されていたが、DNAシーケンサとそれに連携する情報技術との融合が強力なツールとなり、2001年には国際共同チームと民間のCelera社（米）によって、人類の全塩基配列決定の論文が有名誌に同時発表されるという急展開を迎える^{1,2}。ヒトの“基本設計図”である全配列決定に伴い、予想されていた10万というヒトの全遺伝子の数も予想を大きく下回る約32000個と推定され、2007年現在ではその数を下回るという報告もある。従来の「ヒトゲノムの情報は無限に等しい」という過去の考えは完全に抹消され、生命科学の研究の主な関心は配列ゲノム科学から機能ゲノム科学へと移ることになる。

生命の遺伝情報の本体はDNAであるが、その表現型はタンパク質をはじめとする生体分子である。生命は生体分子間の相互作用によって高度に複雑化されたシステムであり、核酸、脂質、糖などの生体高分子や、リガンドなどとの分子間相互作用のネットワーク、すなわち生体内パスウェイを明らかにしていく必要がある。多様化した相互作用は分子個々の構造と密接に関係しており、ゲノムを分子レベルかつ体系的に読み解くには、これらの機能を解明することが必要である。

1. 1. 2 生体分子の機能と構造

生体分子の機能を解明するために重要なステップは、その立体構造を明らかにすることである。生体分子の構造と機能の因果関係に初めて明白な根拠を与えたのは1953年のWatsonとCrickが発表したDNAのX線結晶構造である³。構造解析されたDNAの二重らせん構造によって、転写機構が分子レベルで解明されたことは歴史的に有名である。その後、構造解析の対象は糖や脂質、タンパク質などの様々な生体分子に及んだ。1958年にJohn Kendrewはミオグロビンの立体構造をX線結晶構造によって明らかにした⁴。DNAの構造が明らかにされた当初は、タンパク質も規則的で単純な構造であると考えられていたが、実際には規則構造や対称性も無い複雑な高次構造であることが判明した。

タンパク質は自然環境下で一定の立体構造を持ち、その形に即した生化学的機能を発

現する。1960 初頭に Anfinsen はリボヌクレアーゼの変性実験からタンパク質の高次構造は一次配列（遺伝子）によって完全に決定されるという説を提唱し（Anfinsen Dogma）⁵、タンパク質の構造と機能（遺伝子）の対応関係が明らかになった。他にも数多くの変性したタンパク質が *in vivo* でフォールディングして機能を発現することが明らかにされた。リボソーム内でターゲット配列が転写された後、タンパク質の高次構造が形成される一連の過程では、細胞内に広く存在する分子シャペロンという熱ショックタンパク質群が関与している。しかし、このタンパク質はミスフォールディングしたタンパク質の正しい構造形成と会合体形成を阻止する役目を担っているため、タンパク質の高次構造形成は基本的に自発的過程であると考えられている^{6,7}。このように、タンパク質の機能はその立体構造と密接に関係している。

1. 1. 3 構造ゲノムプロジェクト

1970 年代の遺伝子工学の発展と 1990 年初頭に開発された放射光による X 線放射技術の進歩や核磁気共鳴（NMR）や電子線構造解析を利用した構造解析法の発展によって、タンパク質の立体構造の決定に関する実験的なプロセスがハイスループット化（効率化・自動化）された⁸。2000 年から国際的な協力体制のもと、様々な構造ゲノムプロジェクトが開始され、ゲノム情報を元に網羅的かつ体系的な構造決定を目的とした「構造プロテオミクス」の時代が到来し、米国では各ファミリーを代表した一万程度の構造決定を目標とした。我が国でも「タンパク 3000」プロジェクトが立ち上がり、2002 年から 2007 年までの 5 年間にタンパク質の全基本構造の 1/3 と推定されている 3000 個の構造解析を目標に掲げた。その結果、巨大膜タンパク質の立体構造を含む約 2500 の構造決定に成功した⁹。構造プロテオミクスの成果は、生命科学の発展のみならず、難病治療やゲノム創薬の特許に直結することから、今後の進捗状況が一国の生命科学の明暗を分けるとさえ言われている。

迅速なタンパク質の構造決定技術が確立されたことによって既知の構造の数が飛躍的に増加する一方、構造と機能の因果関係や機能を解析する自動化された技術は未だ確立されていない。現在構造決定されているタンパク質の構造数は 3 万を超え、その数は増加の一途を辿る傾向にある。蓄積された膨大な量のゲノム情報が生命活動においてどのような意味を持っているかの解析、すなわち遺伝子型と個体における表現型とを結ぶ作業が、これからのゲノム科学の一つの大きな課題である。現在ヒトの遺伝子から転写されるタンパク質の約 4 割の機能は不明であり、構造が既知であり機能が未知であるタンパク質が急速に増加している。今後の予想される「ポスト構造ゲノム」の時代は、立体構造が決定されたタンパク質の構造から未知機能をゲノムスケールで予測する方法論の確立が目標となるだろう¹⁰。

1. 1. 3 バイオインフォマティックス

生物を分子レベルで理解するためには、遺伝子産物であるタンパク質の機能を同定しなければならない。しかし、膨大な数に及ぶ全てのタンパク質に関して、その機能を実験的手法によって同定することは現実的ではない。さらに、特定の機能に対する活性測定を行う現在の実験手法では、単一のタンパク質が持つ複数の機能を同時に解析することが困難である。

1970年代から1980年代の配列決定技術の進歩とそれに伴うデータベースの急速な進歩から膨大な生命情報が蓄積され、それらの情報科学的な解析が必要とされた。その解析手法の中で最も注目されるようになったのは、機能既知のタンパク質との配列類似性の検索である。核酸配列やアミノ酸配列の並びを「情報」として捉えることで、既知機能と未知機能の配列間の相同性を評価し、共有されている機能を類推する手法である。コンピュータの演算は配列型の情報が扱いやすいため、情報の処理が非常に効率的に行える。また、生命科学の高度な専門知識を必要としないため、数学・情報工学の研究者にも急速に関心が高まった。

遺伝子配列を情報科学的に解析する技術は注目を集め、1990年代後半以降は「バイオインフォマティックス」と呼ばれる統合的な学問分野として広く認知されるようになった。地上には多種多様な生物種が存在しているが、それぞれの生物はゲノム上で様々なタンパク質をコードしている。これらの機能は生物種によって独立に獲得された訳ではなく、現在の生物に至る進化の過程で遺伝子重複や種分化に伴う共通の祖先遺伝子からの分岐や融合によって生成した遺伝子によってコードされている。従って、相同性が高い遺伝子は互いに比較的最近分化した遺伝子であり、転写されたタンパク質の機能も類似している可能性が高い。タンパク質の情報の全ては配列に「押し込められている」はずであり、配列の相同性によって機能の類似性を推測するアプローチは科学的な正当性を持っている。

バイオインフォマティックスの登場は、従来の探索型の研究では困難であったゲノムワイドな生命科学の研究スタイルが確立される起点となり、遺伝子の分化・融合やタンパク質の種分化・機能といった生物進化の論題が急速に具体化されるきっかけとなった。バイオインフォマティックスは、従来の生物学の研究である *in vitro* (=試験管内で) と対比され、*in silico* (=半導体で) と呼ばれるようになった。

1. 2 経験則に基づくタンパク質の *in silico* 機能予測

1. 2. 1 配列情報に基づく機能予測法

遺伝子の相同性からタンパク質の機能を予測する具体的な方法は、**Open Reading Frame (ORF)** 遺伝子に由来するアミノ酸配列の相同性を比較することによって、タンパク質間に共有されている機能を経験的に予測する方法である。高い配列相同性を示す遺伝子間では機能も共有される場合が多いことが経験的に知られているため、配列の比較から機能の類縁関係を推測可能である。

既に決定されている 200 種以上の生物の全核酸配列情報を元に、タンパク質の個別の機能やタンパク質間相互作用によるシグナル伝達機構の予測がゲノムワイドに行われている (2007 年現在)。配列比較とは、核酸配列もしくは転写された複数のアミノ酸配列同士を比較し、相同性を置換・ギャップ・欠損を考慮しながら互いの類似性を最大化するような組み合わせを数学的に検索する手法を指し、一般に配列アラインメントと呼ばれている。配列の類似性を検索する代表的なアルゴリズムに **BLAST¹¹** や **FASTA¹²** が知られている。

2 つの配列を比較対象とした配列アラインメントにはダイナミック・プログラミング (**DP**) という手法が用いられる。DP では 2 つの配列から塩基配列を要素に持つ行列を作成し、配列の置換・ギャップ・欠損の頻度を統計的に数量化したスコア表 (**PAM¹³**、**BLOSUM¹⁴**) を用いることで配列間の相同性を最大化する。このように、スコア表を用いることによって配列比較の経路を構築する手法は多くの配列アラインメントに共通する最も基本的な手法である。2 つの配列間だけでなく、複数の配列を同時にアラインメントする手法をマルチプルアラインメントと呼ぶ。この手法は、DP の多次元化や逐次的な DP の実行 (**ClustalW¹⁵**)、配列対を乱数によってグループ化 (**PRRN**) することによって実行される^{16,17}。実行されたアラインメントの結果を評価する際、優位性、すなわち結果が偶然によるものなのか進化的に保存された結果なのかを定量化するパラメータが必要であり、代表的なものとして **Stephan** の考案した偶然頻度 (**E-value¹¹**) が知られている。**E-value** は配列の一致が偶然起こりうる期待値を表し、値が低ければ配列の相同性は偶然には起こり得ず、統計的に保存度が高いことを意味する。

タンパク質の機能を相同性の比較から予測する場合、全配列の相同性だけではなく局所的な配列部位の保存度を評価する必要がある。例えば、酵素の触媒残基やリガンドの結合部位などの機能部位に相当した配列が挙げられる。局所的な配列部位の保存性を評価する方法の例として、**Pei** と **Grishin** は従来の保存スコアに出現する配列の頻度を取り入れたスコアによって、その部分的な配列保存性をユークリッド距離によって評価している¹⁸。**Shannon** は問い合わせ配列の確率的優位性を情報エントロピー型によって評価し、

両者に共通することは部分的な配列の保存度と変異度を定性的に評価している¹⁹。配列情報だけではなくアミノ酸の化学的性質も考慮したスコアが Mirney と Shakhnovich らによって考案され、配列を構成しているアミノ酸配を極性や電荷の種類による化学的性質の保存スコアを配列比較に導入している²⁰。その後、保存配列の解析は大規模かつ系統的に行われるようになり、実践的な数多くの方法論が考案された。細胞レベルの機能予測法として、保存配列の分析によって代謝サイクルにおける拮抗薬の予測法²¹や、相互作用するタンパク質ネットワークの予測法^{22,23}が提案されている。分子レベルの機能予測法として、機能に重要な配列モチーフの予測法²⁴や標的分子の認識部位を予測法^{25,26}などが提案されている。配列比較によって得られた情報は、実験によって同定された生化学の情報と統合され管理されるようになった。データベース KEGG²⁷では、代謝の文献と配列との相関関係をグラフ理論によってモデル化することでデータを集中的に管理している。

相同性配列を見つけ出して類似機能を予測する方法は、タンパク質の *in silico* 機能予測として最も一般的な方法論として知られるようになった。特にタンパク質の機能を細胞レベルで予測するための大規模な解析手法は「システムバイオロジー」と呼ばれ、今後の成果が期待されている分野である。

1. 2. 2 分子の形状に基づく機能予測法

タンパク質の機能はその高次構造に由来し、配列の相同性が高くても構造の違いによって全く異なる機能を発現するタンパク質も多く存在する。配列・機能相関を用いた機能予測の限界は、「配列→立体構造→機能」という図式において、立体構造をスキップすることによって引き起こされている。この限界を超えるためには、立体構造・機能の相関に基づいた予測法を確立しなければならず、タンパク質の機能を予測するためには遺伝子の比較だけではなく、高次構造の特性も比較する必要がある。

タンパク質の構造は繰り返しの構造単位が存在するため、系統的に分類することが可能である。SCOP²⁸や CATH²⁹ではタンパク質の二次構造の構成要素や構造モチーフのパターンによって構造を経験的な階層性によって分類している。両者は Helix、Sheet の組み合わせによる幾何学的な構造 (Class)、構造モチーフや二次構造の空間配置を共有する部分構造 (Fold)、Fold が共通し機能を共有するタンパク質の一群 (Family) によって系統的な分類を行っている。データベース DALI³⁰と CE³¹では、ユークリッド距離 (幾何学距離) を用いることによって、2つのタンパク質構造の類似性を定量的に分類している。E-value と二次構造の比較によって、タンパク質の構造を時系列で分類することが可能になった (3D-Blast³²)。

タンパク質は配列の相同性が比較的低い場合においても、構造の類似性によって類似機能を予測できる場合が多く、特にドメインを共有するタンパク質は機能を共有することが多いことが経験的に知られている。部分的に特徴のある二次構造の繰り返し単位をモ

モチーフ（ジंकフィンガー、ロイシンジッパー等）と呼び、同種の機能を持つタンパク質構造の間で共有されることが多く、データベース化もされている（Prosite³³）。構造特定の機能を発現する構造の最小単位（狭義の意味では構造モチーフの集合体）は構造モジュールと呼ばれ、タンパク質の機能発現^{34,35}やタンパク間相互作用ネットワーク³⁶を構造単位で理解する上で重要な概念となっている。同一のファミリー間では配列よりもむしろ機能部位の構造が保存される傾向があるという報告もあり³⁷、配列情報からは探れない“構造モジュール”を同定することは構造ゲノム科学の興味深いテーマの一つとなっており、代表的な機能予測法の一つとして位置づけられている。

X線結晶解析やNMRによって決定されたタンパク質の構造情報は、配列情報と比べて情報量が圧倒的に多い。配列は一次元情報であるのに対し、立体構造の情報は3次元であり、原子種や結合情報などを含めると情報の自由度が急激に増加するからである。タンパク質の構造情報を定性的・定量的に分析するため、その形状に関する様々な評価関数が必要となる。現在タンパク質の構造分析法は、表面形状や原子間の接触距離をパラメータ化した統計的な手法が主流である。分子表面にある窪みの容積や形状を判別・分析することによって、クレフトを検出する方法^{38,39}や、仮想プローブ（:CH₃など）を用いてクレフトを探索する方法⁴⁰、多面体分割によるクレフト探索法（CAST⁴¹）が知られている。また、クレフトの構造類似性は空間モチーフと呼ばれ、これを比較することによって類似機能のタンパク質を検索する様々なアルゴリズムが提案されている。

タンパク質の静的な構造だけでなく、構造の自由度に注目した解析法も登場した。代表的な例は、アミノ酸側鎖のコンフォメーション自由度^{42,43}やDisorder領域⁴⁴に関する解析がある。決定された構造は、位置に関する情報に限らず電子密度や温度因子などの構造安定性や機能に重要な実験情報も含んでおり、データベース化されている（DisEMBL⁴⁵）。

天然型のタンパク質に関して、その配列が変異を起こすと機能が大きく変化する場合は多いことが良く知られているが、それらの中には天然構造と比較して大きく構造変化を起こしている場合も多く、構造と機能の相関を考察する上で重要な情報を含んでいる。現在PDBに登録されているデータの中で、変異の入った構造は数多く登録されている。天然型の構造と変異の入った構造の組み合わせデータを統合的に管理している（PMD⁴⁶）。

1. 2. 3 数理モデルに基づく機能予測法

近年、タンパク質の構造に関する統計解析法と配列情報を融合させることによって、機能を予測する新規の方法論が数多く考案されている⁴⁷。分子系統樹の構築によって進化的に類縁関係にある残基をタンパク質の構造にマッピングすることによって機能に重要な構造クラスターを評価するEvolutionary Trace法（ET法^{48,49}）や、遺伝子の大域的な領域と局所的な領域に保存されている部位の両者を評価することで、機能を共有するタン

パク質の機能部位を網羅的に予測する 3 次元クラスター解析法などが代表的な研究例である^{50,51}。

構造の多様性を進化的な観点から機能を評価するこれらの方法は、配列特異的な構造特性や進化的に保存される傾向がある構造モジュールの発見など、ゲノムワイドな構造解析法として現在最も有用な方法論の一つである。その代表的な手法として、曖昧さを検索するファジー関数を応用した例⁵²や、クレフトの構造を学習アルゴリズムであるニューラルネット⁵³やサポートベクターマシーン (SVM)⁵⁴を応用した研究例がある。セリンプロテアーゼに代表されるように、機能に重要なリレー型の 3 つ組触媒基が同定されている構造は、データベース化 (SUMO⁵⁵) されており、類似の空間モチーフから触媒機能が予測可能である。表面形状の分析法だけではなく、平均力ポテンシャル⁵⁶⁻⁵⁸や接触ポテンシャル⁵⁹からタンパク質の機能を分類する方法や、SVM や空間アラインメント法によって触媒残基の配置を分類・予測する方法論も報告されている⁶⁰⁻⁶³。

タンパク質の構造に関する統計解析法の発展によって、タンパク質の高次構造に関する構造の特性を半定量的に評価することが可能となったばかりでなく、構造と機能の因果関係を体系的に結びつける様々なモデルが提案された。基質結合様式や水素結合ネットワークのパターン等から、今まで注目されていなかった構造様式の多様性と機能との相関関係が示唆された⁶⁴。

1. 3 第一原理に基づくタンパク質の *in silico* 機能予測

1. 3. 1 タンパク質の機能と物理化学

このような背景から、タンパク質の構造情報と実験情報を統合も進んでいる。変異によるタンパク質の熱力学的な影響を統合的に管理しているデータベース (Protherm⁶⁵) も登場している。このデータベースでは、変異を入れた立体構造と分光実験によって得られた自由エネルギー変化や熱容量等を統合的に管理している。

現在バイオインフォマティクスに代表される *in silico* 機能予測法では、経験則を用いた手法が主である。この方法は、配列→類縁関係→機能という順番に相関関係を探るため、できる限り遠い類縁関係を検出する努力がなされる。しかし、遠い類縁関係から機能を推定することには原理的に限界があることを考えると、進化的類縁関係のみに依存しないアプローチが必要になってくる。構造生物学のドグマとして考えられている配列・構造・機能の相関は、常にはっきりとしている訳ではなく、立体構造が分かれば直ちに機能が解明される訳ではない⁶⁶。すなわち、機能未知であるタンパク質の立体構造から、これらの生化学的機能を特定することは自明ではなく、重要だが極めて困難な課題なのである。この問題点をクリアするためには、類縁関係に頼ることのない第一原理計算に基づく機能予測法を確立しなければならない。その手段の一つは、計算化学を用いることでタンパク質構造に関する物理化学的な諸性質を考慮することである。

1. 3. 2 *in silico* 機能予測の方法論

計算化学の魅力は、経験論に頼ることなく、分子の化学的性質を原子レベルや電子レベルで推測可能な点である。そのため、*in vivo* の実験結果からは観測困難な現象を議論するのに役立つ。計算技術の進歩から、タンパク質をはじめとする巨大分子も計算化学の研究対象となり、生体分子の機能を原子レベルで予測することが可能になった。近年この分野はケモインフォマティクス (ケモは化学の意味) と呼ばれることが多く、バイオインフォマティクスと並ぶ *in silico* 機能予測法の一つとなった。

計算化学には様々な理論体系があるが、代表的なものとして古典力学や統計力学、量子力学が挙げられる。古典力学を適用した例は、経験的な調和型ポテンシャルによって分子に働く力を計算する方法である分子力学法 (MM)、アンサンブルを構築することで時間発展する分子の挙動をシミュレーションする分子動力学法 (MD) が代表的である。古典力学を基礎にした方法は、モンテカルロ法 (MC) 等の統計力学を併用する場合も多い。量子力学 (QM) を適用した例は、その理論を分子に適用した量子化学が知られてい

る。具体的には、第一原理に従って Schrödinger 波動方程式の解を求め、その振幅（電子の密度）や固有値（電子のエネルギー）を得る。現在、タンパク質のような巨大な分子構造に適用されているのは古典力学を中心とした理論である。

タンパク質を対象とした計算例で最も一般的かつ具体的な例は、分子力学と静電気学の Poisson-Boltzmann 方程式を融合した手法である。この方法によると、タンパク質の局所的な静電相互作用や極性相互作用が計算可能である。この計算手法が広まった背景には、近年確立された Finite Difference Poisson-Boltzmann 法 (FDPB) と、その巨大分子への近似理論の発展がある^{67,68}。FDPB では、水和したタンパク質を、分子内部が低誘電率（比誘電率 2-20）、水を高誘電率（比誘電率 78.4）の二層でモデル化する。その後、溶質—溶媒の界面（Surface Area (SA)）を考え、境界が満たす条件式（Poisson-Boltzmann 方程式）を数値的に解く。塩強度を効率的に計算する線形近似法の登場も重要であった^{69,70}。Bashfold と Karplus は FDPB と MC を融合させて、タンパク質表面の残基に関する pKa 値を算出する方法を考案した^{71,72}。この方法により、表面残基の解離状態を予測するなど、加水分解酵素の触媒機能を第一原理的に予測することが可能となった。FDPB の計算は簡易かつ良好な計算結果が得られることが確かめられている。振動シュタルクスpekトルの実験から、FDPB の結果を直接比較している⁷³。

第一原理に基づく機能予測法を開発するためには、構造と機能に関して系統的な解析が必要となる。コンピュータの演算能力が向上したことによって、数多くの酵素タンパク質が計算対象とされる中、網羅的な数や種類のタンパク質をターゲットにした研究事例も登場した。その際、予測法のベンチマークとなったのは、実験的に同定されている機能部位の予測であった。Elcock はタンパク質の機能部位を実験的に変異させることによって構造が安定化されることが多い事実から、機能部位は構造を不安定化するという仮説を立て、216 個の酵素について表面に存在する解離性残基の静電ポテンシャルを計算することによって網羅的に機能部位を予測する方法論を提案した⁷⁴。Ondretchen は解離性残基の電荷とそれらの pH 曲線を理論的に計算することによって、構造と機能が共に異なる酵素の触媒基を予測することが可能であることを報告し⁷⁵、予測法の統計的な優位性⁷⁶や他の手法との精度の比較⁷⁷を行っている。

FDPB の計算結果とバイオインフォマティックスの手法を融合した予測法も数多く報告されている。Jones⁷⁸や Tsuchiya⁷⁹はタンパク質の露出表面積から構築した統計関数とタンパク質表面の静電ポテンシャルから、DNA 結合部位を予測する方法論を提案している。Kinoshita は、タンパク質表面の静電ポテンシャルに関する包絡面（コノリー面）を系統的に分類することにより、タンパク質の機能と種類を予測する方法論を提案している^{80,81}。Warwicker はタンパク質の機能を表面形状の分析と静電ポテンシャルの計算によって詳細に議論し、分子表面のポテンシャル勾配から酵素と非酵素を判別する方法論を提案している^{82,83}。Ohta は 67 個の酵素タンパク質について、配列の保存性、表面クレフトの探索、静電ポテンシャルの解析結果を統合することによって、より高い確率で機能部位を予測

する方法論を提案している⁸⁴。

FDPB とその関連手法における問題点は、その計算結果が実験構造の状態に依存することである。PDB の登録構造が増加する一方、同種のタンパク質でも複数の実験構造が登録されることが多くなった。従来の機能予測法は実験構造に頼っているため、結晶構造の選び方が予測結果を左右する可能性もある。Nielsen らは複数のリゾチームの結晶構造 (41 構造) を対象に、構造の依存性と計算した pKa の値を詳細に検討した⁸⁵。その結果、多くの結晶構造に関して、機能部位の解離状態を予測することが可能であると報告している。しかし、リゾチームは大量精製が可能であるため、比較的結晶状態の良いものだけが登録されていると言われ、賛否両論である。現在では、計算結果ができる限り初期構造に依存しないように、水溶液中のアンサンブル平均 (MD のトラジェクトリーや NMR アンサンブル) を用いた計算例が多い。

機能予測法をテーマにした研究に関して、近年報告例が急増しているのは定量的構造活性相関 (Quantitative Structure-Activity Relationship (QSAR)) に関連させた研究である。このような背景 (流行) もあり、機能予測法の研究分野ではタンパク質 (レセプター) の構造単体を扱った内容から、タンパク質-基質やタンパク質-タンパク質に関する「相互作用様式の予測」に移行してきている (2007 年現在)。

QSAR を背景とした機能予測法の研究は、従来の FDPB と Docking 法を併せた方法が代表的である⁸⁶⁻⁸⁹。より高度な方法論として、Docking 法と MD を用いた方法もある。この方法では、CD 4 とその阻害剤との振動モードを計算することで機能部位を予測する方法²⁴や、T4-Lysozyme を対象に Linear Binding Energy (LIE) 法を用いた基質結合部位の予測法が提案されている⁹⁰。ホモロジーモデリングと Docking を用いて、結合部位を予測する方法が提案されている^{91,92}。さらに、タンパク質・基質間だけでなく、タンパク質間の相互作用形態を予測するアルゴリズムも登場している⁹³。ヒドラーゼを対象に、相互作用部位を予測した例が報告されている⁹⁴。近年では、タンパク質溶液のシミュレーションによる機能予測法の例もある⁹⁵。機能不明の酵素の機能予測や⁹⁶、KEGG からサンプリングした中間体の予測⁹⁷など、大がかりな機能の予測例も報告されている。

機能予測法の方法論が高度になる一方、計算結果をデータベース化する動きも活発化している。Tsuchiya は、DNA 結合部位の予測サーバを構築している (preDs)⁹⁸。Kinoshita らは、タンパク質構造の表面形状と FDPB で求めた静電ポテンシャルをデータベース化し、類似性から類似機能を予測するサーバを構築した (eF-Site⁹⁹ と eF-seek¹⁰⁰)。Onta はタンパク質表面のポテンシャルと分子形状、保存配列のデータを統合し、活性部位予測サーバを公開している (P-cats¹⁰¹)。Atanasov らはタンパク質の pKa 値をデータベース化している (PHEPS¹⁰²)。Jones は FDPB を適用して計算したタンパク質分子間の相互作用部位を、データベース化している (SHARP2¹⁰³)。Mandel はタンパク質分子表面の静電ポテンシャルを可視化するサーバを構築している (Pfplus¹⁰⁴)。Baker は PDB ファイルから、pKa の値を基にプロトン化状態を自動判別するサーバを構築している (PDB2PQR¹⁰⁵)。Zhou

は GB モデルからタンパク質の溶媒和を自動的に算出するサーバを立ち上げている (GBr6NL¹⁰⁶)。Endo らは、タンパク質の基準振動解析の結果をデータベース化している (ProMode¹⁰⁷)。

タンパク質の機能は、活性部位近傍のアミノ酸の突然変異することによる基質特異性や活性の変化によって獲得したものである。このような背景を踏まえると、第一原理に基づいた機能部位の予測法の研究とは、「タンパク質がその物性と機能をどのようにして進化させてきたのか」という、生命科学の問いに対して答えを提示し得る重要な研究分野であるのかもしれない。DNA 結合型のタンパク質を対象に、お互いの Homologue である複数の構造に関して DNA 間との相互作用を系統的に計算した結果、静電的に反発に寄与する残基とそのコンフォメーション変化が、残基の変異によって変化しやすい相互作用の原因になっていることが報告されている¹⁰⁸。複数のキナーゼを対象として pKa 計算を実行した結果、機能に重要なモチーフ内で保存されている Cys の役割が説明されている¹⁰⁹。プロテアーゼスーパーファミリーに属する複数のタンパク質を対象に MD を実行した結果、構造や配列の相同性が低いにもかかわらず、機能に固有のダイナミクスが保存されていることが報告されている¹¹⁰。

このようにタンパク質を対象とした第一原理計算のデータが蓄積されることで、タンパク質の機能に対して「物性と進化」という視点から統一的な解釈が得られるかもしれない。

1. 4 電子状態に基づく *in silico* 機能予測の可能性

1. 4. 1 タンパク質の機能と電子状態

タンパク質・核酸等の生体高分子を対象とした計算は、古典論に基づく方法が主流である。しかし、タンパク質の機能を突き詰めて考えると、結合の生成・消滅、プロトンの移動、スピン平衡といった量子力学的な過程に辿り着く。タンパク質の機能は、現実には観測可能な物理量だけでなく、分子界面の電子分極や光との相互作用など、古典論では説明できない様々な性質によって支配されている。そのため、タンパク質の機能を厳密にモデル化するためには、量子力学の適用を避けることはできない。代表的な例は、金属タンパク質の機能である。全てのタンパクの半数近くは遷移金属などの金属原子を必要としており、その大部分は生体に必須な金属である。これらの金属イオンはタンパク質構造の安定性に寄与するだけでなく、鉄ポルフィリンやクロロフィル、鉄硫黄クラスターとして組み込まれ、活性中心の構造を形成している。これらは非常に活性が高く様々な反応性を示す。例えばヘムタンパクに限っても、酸化還元や酸素運搬や電子移動など、その機能は多岐にわたっており、ヘムの様々なスピン状態を介して機能を制御している。例えば定量的な軸を入れて考えてみる。金属タンパク質は生体内で働くため体温で反応が進行する必要があるが、これは僅か 0.1eV 程度のエネルギー変化で反応していることを意味する。以上の例に代表されるように、巨大分子系という困難な状況下で、電子状態まで考慮した高精度の計算手法が必要である。

1. 4. 2 巨大分子に関する電子状態の計算法

無機化学・有機化学の分野で分子の電子状態を記述することに成功を収めた理論は量子化学であり、その代表的な方法に分子軌道法 (Molecular Orbital (MO)) や原子価結合理論 (Valence Bond (VB))、密度汎関数法 (Density Functional Theory (DFT)) などがある。分子軌道法は数学的には基底関数の相互作用行列に対しその固有値 (エネルギー) と固有ベクトル (波動関数) を求める理論である。密度汎関数法は系の基底状態の全エネルギーが電子密度の汎関数として表すことができる数学的定理に基づいた理論である。

量子化学は分子のサイズに対して膨大な計算量を必要とする。初期の量子化学計算は、数十から数百原子程度の低分子の解を得る程度の極めて非実用的なものであった。一般に分子軌道 (Hartree-Fock 方程式) の解を求めると、系の粒子数 N に対して N^4 のオーダーで計算量が増加する。そのため、タンパク質のような巨大分子を計算するには計算手法を工夫する必要があった。

第一の工夫は、タンパク質の構造を分割して計算することである。タンパク質の機能部位は分子構造の限られた一部分であるため、予め機能部位の位置が分かっている場合、機能部位は演算コストの大きな計算を、その他の部分をよりコストの少ない計算を実行すれば非常に効率的である。その代表的な計算手法として、Hybrid Quantum Mechanics Molecular Mechanics (QM/MM) 法¹¹¹や ONIOM 法¹¹²が知られている。QM/MM は、分子構造を QM 領域（機能部位）と MM 領域（その他の分子構造）に分割して計算する方法である。各領域では別々に解（エネルギー）が求められるため、領域間の相互作用や Linker Bond（異なる領域を跨いだ結合）の取り扱いなどの面倒な処理が必要となる。そのため、系が複雑になると計算が理論的に困難である場合が多かった。しかし近年、この問題を克服する様々な近似手法が提案され、周期境界（Ewald¹¹³）や溶媒（Generalized Born^{114,116}）などの重要なモデルを扱うことが可能になった。ONIOM 法は、分子構造に対して多層構造（領域間で構造の重複がある）の計算精度を適用する手法である。ONIOM 法の利点は、ポテンシャル勾配と二次微分が求められるため、振動数のような性質が厳密に計算可能な点である。QM/MM 法や ONIOM 法はタンパク質の反応機構に関する有用な情報を与えてくれる方法である。

第2の工夫は、巨大分子を対象とした近似理論の適用である。分子軌道法は、その近似レベルによって様々な方法に分かれ、(1) π 電子系に適用して成功を収めたヒュッケル法などに代表される経験的方法、(2) 経験的パラメータ等を計算に取り入れる半経験的方法(semi empirical)、(3) 経験的な値を一切使わずに第一原理からすべて計算によって求める非経験的方法(ab initio)の3つに大別される。巨大分子に対する近似理論は、主に半経験的方法を背景に発展してきた。代表的な近似法として、Yang による Divide-and-conquer (D&C)¹¹⁶ や Stewart による MOZYME¹¹⁷、Anikin らによる Local SCF¹¹⁸ がある。これらの近似法は、系のサイズ N に対して N^2 以下の計算コストに抑えることができるため、Linear-Scaling 法と呼ばれている。近似法が発表された当初は、コンピュータの演算能力が限られていたため、タンパク質のような巨大分子の計算には半経験的分子軌道法を適用するのが主流であった。しかし近年では、タンパク質の ab initio 計算を現実的な計算コストの範囲内で可能にする近似法が登場している。代表的な例として、Kitaura の Fragment Molecular Orbital (FMO) 法^{119,120}、Kashiwagi の Quasi-Cannonical Localized Orbital (QCLO) 法¹²¹、Zhang の Molecular Fractionation with Conjugate Caps 法 (MFCC)¹²²、Imamura の Elongation 法^{123,124} がある。

1. 4. 3 *in silico* 機能予測法としての可能性

近年、近似理論の発展やコンピュータの処理能力が急速に向上したことに伴い、量子化学計算によって計算可能な分子のサイズは大きく拡大した。

QM/MM 法の原案が発表されて以降、酵素反応のシミュレーションを中心に数多くの研

究例がある（代表的な総説を引用）¹²⁵⁻¹²⁷。最新の研究では、MD や MC を取り入れることで、自由エネルギー計算に基づく反応解析を実現した研究例も増えている。しかし、QM/MM は機能部位の位置が既知である必要があるため、機能部位の予測法に代表される機能予測法としての研究例は少数である。そのため、本章では具体的な研究例の引用は割愛する。

タンパク質分子の電子状態の研究例を取り上げると、半経験的分子軌道法を用いた研究例がある。近年 York によって、RNA の電子状態に関するデータベースが構築された¹²⁸。

Sato と Kashiwagi は、QCLO によってインスリンの DFT 計算を成功させている。

タンパク質を取り扱った量子化学計算では、FMO 法を適用した研究例が際立って多い。その理由としては、高度な計算精度が保証されるだけでなく、フラグメント（残基）のペアごとに相互作用の大きさを計算できるため、分子設計や機能部位予測を行う際に利便性が高いことが挙げられる。Fukuzawa はエストロゲン受容体^{129,130}や転写因子¹³¹を対象に全電子計算を行い、リガンドとレセプターとの相互作用部位を詳細に議論している。Ishida は QM/MM 計算によってモデリングしたムターゼの反応中間体を対象に、全電子計算によって遷移構造を安定化する機能部位を議論している¹³²。Komeiji は、水和したユビキチンに関して全電子計算を行い、分子界面で起こる電子移動量や双極子の大きさを定量化している¹³³。Ito は核内受容体 Retinoid X について全電子計算を行い、補酵素を介して受容体が活性化する詳細なメカニズムを提唱している¹³⁴。FMO の有用性が評価されているもう一つの理由として、応用理論の発展がある。Mochizuki は基底関数の二次摂動まで考慮することによって、軌道間相互作用（電子移動）を解析する手法を提案している（CAFI）¹³⁵。Nakano は、FMO の相互作用形式をクラスター分析によって判別し、スクリーニングを行うアルゴリズムを提案している（VISCANA）¹³⁶。

分子の電子状態に関する研究は計算コストの面から主に静的構造を対象としてきた。しかし、時間発展する電子の挙動を追跡することができなければ、化学反応を再現ができない。そのため、電子状態と原子核のダイナミクスを同時に扱える計算手法の導入が望まれる。Car-Parrinello は、電子の波動関数に自由度を与え、原子の運動と共に波動関数の時間変化を記述する効率的な運動方程式を提案した（CP 法¹³⁷）。CP 法では、電子密度と外場が一对一対応する性質を利用することで、（Kohn-Sham 方程式の固有値問題の解を直接求めることなく）時間発展する電子のハミルトニアン記述することが可能である。初期の頃、CP 法は表面化学の分野で用いられることが多く、液体の相転移¹³⁸や触媒の重合反応¹³⁹を再現することに成功している。最新では、CP 法を生体分子へ応用した研究例も報告されている。代表的な例に、リボザイムのリン酸基に関する加水分解反応¹⁴⁰やプロテアーゼの加水分解反応¹⁴¹や阻害剤との相互作用¹⁴²、PYP の発色団に関するプロトン移動¹⁴³、Azurin の電子励起¹⁴⁴、ミオグロビンの酸素のラジカル化機構¹⁴⁵、シトクロム c のヘム鉄に関する酸化機構¹⁴⁶がある。より複雑な作用機構を持った系の適用が可能になったことで、今後増え報告されるようになった。

1. 5 本研究の意義と目的

タンパク質の 3 次元立体構造から機能を予測する方法は、コンピュータ性能の向上と相まって、目覚しく進歩してきている。しかし、これらの方法の多くは、ある程度の予測力を持っているものの、本質的に古典的力学や静電気学に基づいているため、タンパク質の機能、特に酵素機能に対して重要な視点が欠けていると思われる。触媒反応の過程で、酵素自身は基質と反応し、化学的な修飾（化学結合の開裂や電子移動など）を受け、最終段階で元の状態に戻る。すなわち、酵素の触媒残基は基質に対して反応物である。酵素の触媒反応以外にも、タンパク質の構造と機能の関係を理解する上で、量子力学的な効果を考慮しなければならない例は少なくない。多数のタンパク質を研究対象にした場合、そのような量子力学的な効果を考慮するためには、タンパク質の電子波動関数に注目するのが良いだろう。なぜなら、電子波動関数は分子構造によって一義的に決定されるため、タンパク質の分子構造に固有の機能は、電子波動関数の性質に直接反映されるからである。

本研究では、タンパク質の電子波動関数、立体構造及び機能の 3 者間の関係（波動関数－立体構造相関）を明らかにし、電子波動関数の性質から機能を予測する方法論の確立を目指した。それによって、与えられた立体構造から機能を予測する方法及びその構築原理そのものに解釈を与えることを目的とした。

本論文の構成は、以下の通りである。

第 2 章 電子波動関数の解析に基づくタンパク質の構造-機能相関と機能部位予測法の開発

構造、機能のいずれも既知で、冗長性やホモロジーに重複の無い 112 の酵素を選出し、それらの全てに対して **linear-scaling** 分子軌道計算によって正準軌道 (MO) を求めた。その結果、3 つの重要な結論に到達した。第一に、タンパク質の MO は階層的な様式で構築されており、電子の非局在化の度合いによって 3 つのグループに分類できることを見出した。これをエネルギーの降順にグループ A、B、C と呼ぶことにすると、1) グループ A は非常に局在化した MO から成る、2) グループ B は電子分布がほぼタンパク質構造全体に広がった MO を含む、3) グループ C はタンパク質の共有結合に沿って形成された中程度の広がり of MO から成っている、ことが判明した。また、フロンティア軌道はいずれのタンパク質でもグループ A に属し、非常に局在化した性質を持っていることが明らかとなった。第二に、水和はタンパク質の電子構造に大きな影響を与え、特にフロンティア軌道を活性部位に局在化させる役割を果たしていることを見出した。これより、タンパク質の水和は、天然構造に対する熱力学的安定化作用だけではなく触媒機能発現にも重要な役割を果たし

ていることが示唆された。3 番目として、グループ A の分析からタンパク質の機能部位を予測するプロトコールを提案し、この手法が有用であることを実証した。

第 3 章 超分子計算による水和状態タンパク質の電子波動関数の解析

本章では、タンパク質と水和圏からなる超分子の電子波動関数を解析することによって、タンパク質と水分子間に起こる量子力学的な過程を詳細に分析した。その結果、タンパク質と第一水和圏の水分子との間に顕著な電子移動が起こること、及び水和に伴うエネルギーの安定化は主に第一水和圏と第二水和圏から生じることを見出した。また、この超分子計算の結果と連続体近似計算の結果を比較し、タンパク質の電子波動関数、分子内部エネルギー及びタンパク質－水界面での相互作用に関して、両者の予測結果には顕著な違いがあることを見出した。特に量子力学的効果をあらわに考慮できる前者の計算から、タンパク質と水から成る超分子では系全体に広がる波動関数が形成され、これにより系全体が安定化されること、及び触媒部位では水分子との間で高エネルギー状態にある電子を介した特異的な電子移動相互作用が生じることが判明した。

第 4 章 タンパク質のフォールディング過程における電子波動関数－構造相関の解析

本章では 3 つの小タンパク質に対して、フォールディング過程と電子波動関数の関係を詳細に解析した。具体的には、各タンパク質の分子動力学計算を実行することによってフォールディング過程をシミュレーションした後、サンプリングした構造全てに対して分子軌道計算を実行し波動関数の時間発展に関する解析を行った。その結果、フォールディング過程では主鎖間の水素結合や疎水コアを形成する残基間で電子移動が起こることが判明し、フォールディングの駆動力として残基間電子移動が重要であることが示唆された。また、フォールディング過程を 2 章で得た波動関数－立体構造相関の観点から解析することにより、電子分布の広がりとは構造形成の間に直接的因果関係が存在することを示唆する結果も得た。

1. 6 引用文献

1. Lander ES et al. The Genome International Sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature* 2001;15:860-921.
2. Venter JC, Adams MD, Myers EW. The sequence of human genome. *Science* 2001; 291:1304-1351.
3. Watson J, Crick F, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953;171:737-738.
4. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff HW, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* 1958;181:662-666.
5. Anfinsen CB. Principles that govern the folding of protein chains. *Science* 1973;181: 223-230.
6. Ellis RJ, Hartl FJ. Principles of folding in the cellular environment. *Curr. Opin. Struct. Biol.* 1999;9:102-110.
7. Pelham HRB. Intracellular membrane traffic: Getting proteins sorted. *Phil. Trans. R. Soc.* 1999;B354:1472-1478.
8. Berman M, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res.* 2000;28:235-242.
9. Oshima T, Nakamura H. Overview: Progress on and trends in the world Protein 3000 Project. *Tanpakushitsu Kakusan Koso.* 2005;50:831-835.
10. Brenner SE. A tour of structural genomics. *Nature Rev. Genet.* 2001;2:801-809.
11. Pei J, Grishin NV. AL2CO: Calculation of positional conservation in a protein sequence alignment. *Bioinformatics* 2001;17:700-712.
12. Pearson WR. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* 1990;183:63-98.
13. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc.Natl.Acad.Sci. USA* 1992;89:10915-10919.
14. Altschul SF. Amino acid substitution matrices from an information theoretic perspective. *J. Mol. Biol.* 1991;219:555-565.
15. Higgins DG, Thompson JD, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22: 4673-4680.
16. Gotoh O. Multiple sequence alignment algorithms and applications. *Adv. Biophys.* 1999;36:159-206.
17. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped

- BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389-3402.
18. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* 1991;9:56-68.
 19. Mirney LA, Shakhnovich EI. Universal conserved positions in protein folds: Reading evolutionary signals about stability, folding kinetics and functions. *J. Mol. Biol.* 1999;291:177-196.
 20. Oliveira L, Paiva PB, Paiva ACM, Vriend G. Identification of functionally important residues with the use of entropy-variability plots. *Proteins* 2003;52:544-552.
 21. Fruih T, Saika H, Svensson L, Pitterna T, Sakurai J, Okada T, Urade Y, Oda K, Fujitani Y, Takimoto M, Yamaura T, Inui T, Matatani M, Takai M, Umemura I, Teno N, Murata T. RL2500: A potent ETb selective endothelin antagonist. *Bioorg. Med. Chem. Lett.* 1991;6:2323-2326.
 22. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles., 1999 *Proc. Natl. Acad. Sci. USA*;96:4285-4288.
 23. Enright AJ, Iliopoulos I, Kyripides NC, Ouzounis CA. Protein interaction maps for complete genomes based on gene fusion events. *Nature* 1999;402:86-90.
 24. Edward MM, Matteo P, Ho-leung N, Danny WR, Todd O, Eisenberg YD. Detecting protein function and protein-protein interactions from genome sequence. *Science* 1999;285:751-753.
 25. Chakrabarti R, Klibanov AM, Friesner RA. Sequence optimization and designability of enzyme active sites. *Proc Natl Acad Sci U S A.* 2005;102:12035-12040.
 26. Chakrabarti R, Klibanov AM, Friesner RA. Computational prediction of native protein ligand-binding and enzyme active site sequences. *Proc Natl Acad Sci U S A.* 2005;102:10153-10158.
 27. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 2006;34:354-357.
 28. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 1995;247:536-540.
 29. Pearl FMG, Lee D, Bray JE, Sillitoe I, Todd AE, Harrison AP, Thornton JM, Orengo CA. Assigning genomic sequences to CATH. *Nucleic Acids Res.* 2000;28:277-282.
 30. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 1993;233:123-138.
 31. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* 1998;11:739-748.
 32. Yang JM, Tung CH. Protein structure database search and evolutionary classification. *Nucleic*

- Acids Res. 2006;34:3646-3659.
33. Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJ. The PROSITE database. *Nucleic Acids Res.* 2006;34:227-230.
 34. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell Biology. *Nature* 1999;402:47-52.
 35. Pereira-Leal JB, Enright AJ, Ouzounis CA. Detection of functional modules from protein structure. *Proteins* 2004;54:49-57.
 36. Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci USA.* 2003;100:12123-12128..
 37. Laskowski RA, Watson JD, Thornton JM. Protein function prediction using local 3D templates. *J Mol Biol.* 2005;351:614-626.
 38. Brandy Jr GP, Stouten PFW. Fast visualization of protein binding pocket with PASS. *J. Comp. Mol. Des.* 2000;14:383-401.
 39. Ben-Shimon A, Eisenstein M. Looking at enzymes from the inside out: the proximity of catalytic residues to the molecular centroid can be used for detection of active sites and enzyme-ligand interfaces. *J Mol Biol.* 2005;351:309-326.
 40. Laurie AT, Jackson RM. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics.* 2005;21:1908-1916.
 41. Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci.* 1998;7:1884-1897.
 42. Petock JM, Torshin IY, Weber IT, Harrison RW. Analysis of protein structures reveals regions of rare backbone conformations at functional sites. *Proteins* 2003;53: 872-879.
 43. Shapovalov MV, Dunbrack RL Jr. Statistical and conformational analysis of the electron density of protein side chains. *Proteins.* 2007;66:279-303.
 44. Yang AY, Källblad P, Mancera RL. Molecular modelling prediction of ligand binding site flexibility. *J Comput Aided Mol Des.* 2004;18:235-250.
 45. Linding R, Jensen LJ, Diella, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. *Structure* 2003;11:1453-14539.
 46. Kawabata T, Ota M, Nishikawa K. "The protein mutant database" *Nucleic Acids Research* 1999;27:355-357.
 47. Whisstock JC, Lesk AM. Prediction of protein function from sequence and structure. *Quart. Rev. Biol.* 2003;36:307-340.
 48. Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* 1996;257:342-358.
 49. Yao H, Kristensen DM, Mihalek I, Sowa ME, Shaw C, Kimmel M, Kavraki L, Lichtarge O. An accurate, sensitive and scalable method to identify functional sites in protein structures. *J. Mol.*

- Biol. 2004;326:255-261.
50. Landgraf R, Fisfer D, Eisenberg D. Analysis of hereglin symmetry by wighted evolutionary tracing. *Protein Eng.* 1999;12:943-951.
 51. Landgraf R, Xenarios L, Eisenberg D. Three dimensional cluster analysis identifies interfaces and functional residue cluster in proteins. *J. Mol. Biol.* 2001;307:1487-1502.
 52. Cammer SA, Hoffman BT, Speir JA, Canady MA, Nelson MR, Knutson S, Gallina M, Baxter SM, Fetrow JS. Structure-based active site profiles for genome analysis and functional family subclassification. *J Mol Biol* 2003;28:334:387-401.
 53. Gutteridge A, Bartlett GJ, Thorton JM. Using a neural network and spatial clustering to predict the location of active sites in enzymes. *J Mol Biol.* 2003;330:719-734.
 54. Cai Y, Liu X, Xu X, Chou K. Support vector machines for predicting HIV protease cleavage sites in protein. *J. Comput. Chem.* 2002;23:267-274.
 55. Jambon M, Andrieu O, Combet C, Deléage G, Delfaud F, Geourjon C. The SuMo server: 3D search for protein functional sites. *Bioinformatics.* 2005;21:3929-3930.
 56. Sippl M. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J Mol Biol.* 1990;213:859-883.
 57. Sippl MJ. Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.* 1995;5:229-235.
 58. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring functions to predict protein-ligand interaction. *J Mol Biol.* 2000;295:337-356.
 59. Miyazawa S, Jernigan RL. Estimation of effective inter-residue contact energies from protein crystal structures: Quasi-chemical approximation, *Macromolecules. J Mol Biol.* 1985;18:534-552.
 60. Panchenko AR, Kondrashov F, Bryant S. Prediction of functional sites by analysis of sequence and structure conservation. *Protein Sci.* 2004;13:884-892.
 61. Cai Y, Zhou GP, Jen ZH, Lin SL, Chou KC. Identify catalytic triads of serine hydrolases by support vector machine. *J Theo Biol.* 2004;228:551-557.
 62. Chou K, Cai Y. A novel approach to predict active sites of enzyme molecules. *Proteins.* 2004;55:77-82.
 63. Meng EC, Polacco BJ, Babbitt PC. Superfamily active site templates. *Proteins.* 2004;55:962-976.
 64. Shulman-Pleg A, Nussinov R, Wolfson HJ. Recognition of functional sites in protein structures. *J Mol Biol.* 2004;339:607-663.
 65. Kumar MD, Bava KA, Gromiha MM, Prabakaran P, Kitajima K, Uedaira H, Sarai A. A: ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Res.* 2006;34:204-206.

66. Russell RB, Sasieni PD, Sternberg MJ. Supersites within superfolds. Binding site similarity in the absence of homology. *J Mol Biol* 1998;282:903-918.
67. Warwicker J, Watson NC. Calculation of the electric field potential in the active site cleft due to alpha-helix dipoles. *J Mol Biol.* 1982;157:671-679.
68. Honig B, Sharp K, Yang A. Macroscopic models of aqueous solution: Biological and chemical application. *J Phys Chem.* 1993;97:1101-1109.
69. Antosiewicz J, Briggs JM, Elcock AH, Gilson MK, Mccammon JA. Computing ionization states of proteins with a detailed charge model. *J Computat Chem.* 1996;17:1633-1644.
70. Rocchia W, Alexov E, Honig B. Extending the applicability of the nonlinear poisson-boltzmann equation: Multiple dielectric constants and multivalent ions. *J Phys Chem. B* 2001;105:6507-6514.
71. Bashford D, Karplus M. Multiple-site titration curves of proteins: An analysis of exact and approximate methods for their calculation. *J Phys Chem.* 1991;95: 9556-9561.
72. Bashford D, Gerwert K. Electrostatic calculations of the pKa values of ionizable bacteriorhodopsin. *J Mol Biol.* 1992;224:473-486.
73. Suydam IT, Snow CD, Pande VS, Boxer SG. Electric fields at the active site of an enzyme: direct comparison of experiment with theory. *Science.* 2006;14;313:200-204.
74. Elcock AH. Prediction of functionally important residues based solely on the computed energetics of protein structure. *J Mol Biol.* 2001;312:885-896.
75. Ondrechen MJ, Clifton JG, Ringe D. THEMATICS: A simple computational predictor of enzyme function from structure. *Proc Natl Acad Sci. USA* 2001;98: 12473-1247.
76. Ko J, Murga LF, André P, Yang H, Ondrechen MJ, Williams RJ, Agunwamba A, Budil DE. Statistical criteria for the identification of protein active sites using theoretical microscopic titration curves. *Proteins* 2005;59:183-195.
77. Wei Y, Ko J, Murga LF, Ondrechen MJ. Selective prediction of interaction sites in protein structures with THEMATICS. *BMC Bioinformatics* 2007;8:119.
78. Jones S, Shanahan HP, Berman HM, Thornton JM. Using electrostatic potential to predict DNA-binding proteins. *Nucleic Acids Res.* 2003;31:7189-7198.
79. Tsuchiya Y, Kinoshita K, Nakamura H. Structure-based prediction of DNA-binding sites on proteins using the empirical preference of electrostatic potential and shape of molecular surfaces. *Proteins* 2004;55:885-894.
80. Kinoshita K, Nakamura H. Identification of protein biochemical functions by similarity search using the molecular surface database eF-site. *Protein Sci.* 2003;12:1589-1595.
81. Kinoshita K, Nakamura H. Identification of the ligand binding sites on the molecular surface of proteins. *Protein Sci.* 2005;14:711-718.
82. Bete P, Warwicker J. Enzyme/Non-enzyme discrimination and prediction of enzyme active site

- location using charge-based methods. *J Mol Biol.* 2004;340:263-276.
83. Greaves R, Warwicker J. Active site identification through geometry-based and sequence profile-based calculations: Burial of catalytic clefts. *J Mol Biol.* 2005;349:547-557.
 84. Ohta M, Kinoshita, Nishikawa K. Prediction of catalytic residues in enzymes based on known tertiary structure, stability profile, and sequence conservation. *J Mol Biol.* 2003;327:1053-1064.
 85. Nielsen JE, Mccammon JA. On the evaluation and optimization of protein X-ray structures for pKa calculation. *Protein Sci.* 2003;12:313-326.
 86. Mincer JS, Schwartz SD. A computational methods to identify residues important in creating protein promoting vibration in enzymes. *J Phys Chem. B* 2003;107:366-371.
 87. Reddy SY, Bruice TC. Determination of enzyme mechanics by molecular dynamics: Studies on quono proteins, methanol dehydrogenase, and soluble glucose dehydrogenase. *Protein Sci.* 2004;13:1965-1978.
 88. Kortvelyesi T, Dennis S, Silberstein M, BrownIII L, Vajda S. Algorithms for computational mapping of proteins. *Proteins.* 2003;51:340-351.
 89. Silberstein M, Dennis S, BrownIII L, Kortvelyesi T. Identification of substrate binding sites in enzymes by computational solvent mapping. *J Mol Biol.* 2004;332: 1095-1113.
 90. Kong R, Tan JJ, Ma XH, Chen WZ, Wang CX. Prediction of the binding mode between BMS-378806 and HIV-1 gp120 by docking and molecular dynamics simulation. *Biochim Biophys Acta.* 2006;1764:766-772.
 91. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA. Predicting absolute ligand binding free energies to a simple model site. *J Mol Biol.* 2007;371:1118-1134.
 92. Costanzi S, Tikhonova IG, Ohno M, Roh EJ, Joshi BV, Colson AO, Houston D, Maddileti S, Harden TK, Jacobson KA. P2Y1 antagonists: combining receptor-based modeling and QSAR for a quantitative prediction of the biological activity based on consensus scoring. *J Med Chem.* 2007;50:3229-3241.
 93. Bjelic S, Aqvist J. Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site. *Biochemistry.* 2004;43:14521-14528.
 94. McGuffee SR, Elcock AH. Atomically detailed simulations of concentrated protein solutions: the effects of salt, pH, point mutations, and protein concentration in simulations of 1000-molecule systems. *J Am Chem Soc.* 2006;128:12098-12110.
 95. Zhang C, Liu S, Zhou Y. Docking prediction using biological information, ZDOCK sampling technique, and clustering guided by the DFIRE statistical energy function. *Proteins.* 2005;60:314-318.
 96. Sen TZ, Kloczkowski A, Jernigan RL, Yan C, Honavar V, Ho KM, Wang CZ, Ihm Y, Cao H, Gu

- X, Dobbs D. Predicting binding sites of hydrolase-inhibitor complexes by combining several methods. *BMC Bioinformatics*. 2004;17:205-210.
97. Hermann JC, Marti-Arbona R, Fedorov AA, Fedorov E, Almo SC, Shoichet BK, Raushel FM. Structure-based activity prediction for an enzyme of unknown function. *Nature*. 2007;448:775-779.
 98. Hermann JC, Ghanem E, Li Y, Raushel FM, Irwin JJ, Shoichet BK. Predicting substrates by docking high-energy intermediates to enzyme structures. *J Am Chem Soc*. 2006;128:15882-15891.
 99. Yuko Tsuchiya , Kengo Kinoshita , Haruki Nakamura PreDs: a server for predicting dsDNA-binding site on protein molecular surfaces. *Bioinformatics*. 2005;21:1721-1723.
 100. Kinoshita K, Murakami Y, Nakamura H. eF-seek: prediction of the functional sites of proteins by searching for similar electrostatic potential and molecular surface shape. *Nucl. Acids Res*. 2007;35:398-402.
 101. Kinoshita K, Ota M. P-cats: prediction of catalytic residues in proteins from their tertiary structures. *Bioinformatics*. 2005;21:3570-3571.
 102. Kantardjiev AA, Atanasov BP. PHEPS: web-based pH-dependent Protein Electrostatics Server. *Nucleic Acids Res*. 2006;34:43-47.
 103. Murakami Y, Jones S. SHARP2: protein-protein interaction predictions using patch analysis. *Bioinformatics*. 2006;22:1794-1795.
 104. Shazman S, Celniker G, Haber O, Glaser F, Mandel-Gutfreund Y. Patch Finder Plus (PFplus): a web server for extracting and displaying positive electrostatic patches on protein surfaces. *Nucleic Acids Res*. 2007;35:526-530.
 105. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res*. 2007;35:522-525.
 106. Tjong H, Zhou HX. GBr6NL: a generalized Born method for accurately reproducing solvation energy of the nonlinear Poisson-Boltzmann equation. *J Chem Phys*. 2007;126:195102-195107.
 107. Wako H, Kato M, Endo S. ProMode: a database of normal mode analyses on protein molecules with a full-atom model. *Bioinformatics*;2004;20:2035-2043.
 108. Chen YC, Wu CY, Lim C. Predicting DNA-binding amino acid residues from electrostatic stabilization upon mutation to Asp/Glu and evolutionary conservation. *Proteins*. 2007;67:671-680.
 109. Carnevale V, Raugei S, Micheletti C, Carloni P. Convergent dynamics in the protease enzymatic superfamily. *J Am Chem Soc*. 2006;128:9766-9772.
 110. Wang PF, Flynn AJ, Naor MM, Jensen JH, Cui G, Merz KM Jr, Kenyon GL, McLeish MJ. Exploring the role of the active site cysteine in human muscle creatine kinase. *Biochemistry*.

2006;45:11464-11472.

111. Warshel A, Levitt MJ. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol.* 1976;103:227-249.
112. Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and $\text{Pt}(\text{P}(\text{t-Bu})_3)_2 + \text{H}_2$ Oxidative Addition. *J Phys Chem.* 1996;100:19357-19363.
113. Nam K, Gao J, York DM. "An efficient linear-scaling Ewald method for long-range electrostatics in combined QM/MM calculations", *J Chem Theory Comput* 2005;1:2-13.
114. Schaefer P, Ricciardi D, Cui Q. Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules. *J Chem Phys.* 2005;123:14905-14919.
115. König PH, Ghosh N, Hoffmann M, Elstner M, Tajkhorshid E, Frauenheim T, Cui Q. Toward theoretical analysis of long-range proton transfer kinetics in biomolecular pumps. *J Phys Chem. A* 2006;110:548-563.
116. Yang W. Linear Scaling Quantum Mechanical Studies of Biological Macromolecules in Solution. *Phys Rev Lett.* 1991;66:1438-1441.
117. Stewart JJP. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *Int J Quantum Chem.* 1996;58:133-146.
118. Anikin NA, Anisimov VM, Bugaenko VL, Bobrikov VV, Andreyev AM. LocalSCF method for semiempirical quantum-chemical calculation of ultralarge biomolecules. *J Chem Phys.* 2004;121:1266-1270.
119. Kitaura K, Sawai S, Asada T, Nakano T, Uebayasi M. Pair interaction molecular orbital method: An approximate computational method for molecular interactions. *Chem Phys Lett.* 1999;312:319-324.
120. Kitaura K, Sawai S, Asada T, Nakano T, Uebayasi M. Fragment molecular orbital method: An approximate computational method for large molecules. *Chem. Phys. Lett.* 1999;313:701-706.
121. Kashiwagi H, Iwai H, Tokieda K, Era M, Sumita T, Yoshihiro T, Sato F. "Convergence process with quasi-canonical localized orbital in all-electron SCF calculation on proteins", *Mol Phys.* 2003;101:81-86.
122. He X, Zhang JZ. A new method for direct calculation of total energy of protein. *J Chem Phys.* 2005;122:31103-31104.
123. Imamura A, Aoki Y, Maekawa K. A Theoretical Synthesis of Polymers by Using Uniform Localization of Molecular Orbitals----- Proposal of an Elongation Method. *J Chem Phys.* 1991;95:5419-5431.
124. Aoki A, Imamura A. Local Density of States of Aperiodic Polymers Using the Localized

- Orbitals from an ab initio Elongation Method. *J Chem Phys.* 1992;97:8432-8440.
125. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MH. Electrostatic basis for enzyme catalysis. *Chem Rev.* 2006;106:3210-3235.
126. Bruice TC. Computational approaches: reaction trajectories, structures, and atomic motions. Enzyme reactions and proficiency. *Chem Rev.* 2006;106:3119-3139.
127. Senn HM, Thiel W. QM/MM studies of enzymes. *Curr Opin Chem Biol.* 2007;11:182-187.
128. Giese TJ, Gregersen BA, Liu Y, Nam K, Mayaan E, Moser A, Range K, Faza ON, Lopez CS, de Lera AR, Schaftenaar G, Lopez X, Lee TS, Karypis G, York DM. QCRNA 1.0: a database of quantum calculations for RNA catalysis. *J Mol Graph Model.* 2006;25:423-433.
129. Fukuzawa K, Mochizuki Y, Tanaka S, Kitaura K, Nakano T. Molecular interactions between estrogen receptor and its ligand studied by the ab initio fragment molecular orbital method. *J Phys Chem B.* 2006;110:16102-16110.
130. Fukuzawa K, Kitaura K, Uebayasi M, Nakata K, Kaminuma T, Nakano T. Ab initio quantum mechanical study of the binding energies of human estrogen receptor alpha with its ligands: an application of fragment molecular orbital method. *J Comput Chem.* 2005;26:1-10.
131. Fukuzawa K, Komeiji Y, Mochizuki Y, Kato A, Nakano T, Tanaka S. Intra- and intermolecular interactions between cyclic-AMP receptor protein and DNA: Ab initio fragment molecular orbital study. *J Comput Chem.* 2007;28:2237-2239.
132. Ishida T, Fedorov DG, Kitaura K. All electron quantum chemical calculation of the entire enzyme system confirms a collective catalytic device in the chorismate mutase reaction. *J Phys Chem B.* 2006;110:1457-1463.
133. Komeiji Y, Ishida T, Fedorov DG, Kitaura K. Change in a protein's electronic structure induced by an explicit solvent: an ab initio fragment molecular orbital study of ubiquitin. *J Comput Chem.* 2007;28:1750-1762.
134. Ito M, Fukuzawa K, Mochizuki Y, Nakano T, Tanaka S. Ab initio fragment molecular orbital study of molecular interactions between liganded retinoid X receptor and its coactivator: roles of helix 12 in the coactivator binding mechanism. *J Phys Chem. B* 2007;111:3525-3533.
135. Mochizuki Y, Fukuzawa K, Kato A, Tanaka S, Kitaura K, Nakano T. A Configuration Analysis for Fragment Interaction. *Chem Phys Lett.* 2005;410:247-253.
136. Amari S, Aizawa M, Zhang J, Fukuzawa K, Mochizuki Y, Iwasawa Y, Nakata K, Chuman H, Nakano T. VISCANA: visualized cluster analysis of protein-ligand interaction based on the ab initio fragment molecular orbital method for virtual ligand screening. *J Chem Inf Model.* 2006;46:221-230.
137. Car R, Parrinello M. Unified approach for molecular-dynamics and density-functional theory. *Phys Lev Lett.* 1985;55:2471-2474.
138. Kuo IF, Mundy CJ. An ab initio molecular dynamics study of the aqueous liquid-vapor interface.

- Science. 2004;303:658-660.
139. Boero M, Parrinello M.; Terakura K. Ziegler-Natta heterogeneous catalysis by first principles computer experiments. *Surface Science*. 1999;438:1-8.
 140. Boero M, Terakura K, Tateno M. Catalytic role of metal ion in the selection of competing reaction paths: a first principles molecular dynamics study of the enzymatic reaction in ribozyme. *J Am Chem Soc*. 2002;124:8949-8957.
 141. Raugei S, Cascella M, Carloni P. A proficient enzyme: insights on the mechanism of orotidine monophosphate decarboxylase from computer simulations. *J Am Chem Soc*. 2004;126:15730-15737.
 142. Vidossich P, Carloni P. Binding of phosphinate and phosphonate inhibitors to aspartic proteases: a first-principles study. *J Phys Chem. B* 2006;110:1437-1442.
 143. Leenders EJ, Guidoni L, Röthlisberger U, Vree de J, Bolhuis PG, Meijer EJ. Protonation of the chromophore in the photoactive yellow protein. *J Phys Chem. B* 2007;111:3765-3773.
 144. Cascella M, Magistrato A, Tavernelli I, Carloni P, Rothlisberger U. Role of protein frame and solvent for the redox properties of azurin from *Pseudomonas aeruginosa*. *Proc Natl Acad Sci. U S A* 2006;103:19641-19646.
 145. De Angelis F, Jarzecki AA, Car R, Spiro TG. Quantum chemical evaluation of protein control over heme ligation: CO/O₂ discrimination in myoglobin. *J Phys Chem. B* 2005;109:3065-3070.
 146. Furlan S, Penna GL, Banci L, Mealli C. Ab initio molecular dynamics of heme in cytochrome c. *J Phys Chem B*. 2007;111:1157-1164.

第2章 電子波動関数の解析に基づくタンパク質の構造・機能相関と機能部位予測法の開発

2. 1 緒言

周知の通り、構造ゲノミクスとはタンパク質の生産と構造決定のためのハイスループットな方法により、膨大な数のタンパク質の構造を解く国際的な試みである¹⁻⁴。このプロジェクトはタンパク質ファミリーに関する我々の構造的な知見を大きく広げることに成功してきた^{5,6}。特に、機能が既知で構造類似性のあるタンパク質が存在しないような新規な折り畳み構造を発見することに貢献した^{7,8}。その結果、現在、*in silico* 計算によってタンパク質の機能（その機能部位）に関する予測法の必要性が増大している⁹⁻¹¹。一般的に、新規の折り畳み構造に対する機能部位の予測は、それらの生物学的機能の理解への出発点というだけでなく、タンパク・基質間の相互作用部位に関する情報を必要としたドラッグデザインへの応用にとっても重要である¹²。

いくつかの場合、特に酵素において機能部位として可能性がある部位は、分子の全体的な構造類似性もしくは局所的な構造パターンの類似性から推測が可能である¹³⁻¹⁵。もちろん、このような知識ベースの方法は、標的のタンパク質が完全に新規の折り畳み構造を持つ場合には適用できない。そのため、タンパク質の *ab initio* 機能予測は今後の構造ゲノミクス時代におけるコンピュータ生物学にとって非常に重要な課題となる。そのための良い出発点は、構造と機能・物理との間に存在する因果関係を明らかにすることである。コンピュータを用いてタンパク質の3次元構造のみから機能を予測する方法はかなり進歩してきた。これらの方法では、与えられた3次元構造の活性部位を見出すために様々な物理化学的な情報を利用する。例えば Ondrechen らによると、触媒反応に関係した解離性残基の理論的な pH 滴定曲線は、しばしば異常な挙動を示し、予想外の平坦な（曲線）領域と pKa 値のシフトを示す¹⁶。このように理論的な滴定曲線（THEMATICS）は酵素の触媒部位を見出すのに有用である^{16,17}。実際、近年のベンチマーク¹⁸では、THEMATICS は他の3次元構造に基づく方法と比較して^{19,20}、良好な精度かつ低い偽陽性の割合という点で触媒残基の予測に関してより良い結果を与えることを示している。もう一つのアプローチは、機能部位がタンパク質の構造を熱力学的に不安定化することが多いという実験的な事実に基づいている。Elcock は、実験で不安定化に寄与することが知られている機能に重要な残基が、Finite Difference Poisson-Boltzmann (FDPB) の計算で最も不安定化を引き起こす残基であり、計算した荷電性残基の静電エネルギーと進化的な保存度合との間に一般的な関係が存在することを立証した²¹。このように検出された不安定部位は、特に糖鎖や小リガンドが結合す

る機能部位を特定するために有用な情報となる²²。近年、経験的なエネルギー関数から得られた触媒残基の安定性プロファイルと立体構造や配列保存を組み合わせた統合的手法が開発された²³⁻²⁷。エネルギー的な性質に加え、タンパク質のダイナミクスを解析することも活性部位残基を特定するのに有力である。タンパク質のダイナミクスに関する研究も機能部位の位置に有用な情報を提供する。Yang と Bahar は 98 の非冗長かつホモロジーの無い酵素群に関して、Gaussian Network Model (GNM) から予測した分子全体の Hinge (運動) の中心と、実験的に同定された触媒部位が共局在化していることを示した²⁸。

以上の 3 次元構造に基づく方法は、ある程度の予測力を持っているものの、本質的に古典的な力学や静電気学に基づいている。そのため、タンパク質の機能、特に酵素機能の重要な観点の一つ欠けていると思われる。触媒反応の過程で、酵素自身は基質と反応し、化学的な修飾 (化学結合の開裂や電子移動など) を受け、最終段階で元の状態に戻る。別の言い方をすれば、酵素の触媒残基は基質が反応する相手となる。それゆえ、酵素の触媒部位は分子構造全体の中で電子論的に活性のある部位に位置しているという仮説が立てられる。よく知られているように、フロンティア軌道理論は分子の反応部位を特定するための良い指針を与えてくれる。この理論によると、求電子反応は最高被占軌道 (HOMO) の電子密度が最大の部位で起こり、求核反応は最低空軌道 (LUMO) の電子密度が最大の部位で起こる²⁹。我々が以前に行った研究では、水和した RnaseT₁ のフロンティア軌道は触媒部位にフロンティア軌道が局在化していることを示した³⁰。酵素では反応には直接関与しないが、基質の結合に必須である別のタイプの機能部位が存在する。48G7 抗体の親和性成熟に関する我々の最近の研究では、フロンティア軌道はハプテン (アナログ基質) の結合部位に局在化していた³¹。これらの研究は、フロンティア軌道の分析によりタンパク質の機能を予測ができる可能性が高いことを明白に示している。従って、このアイディアに沿った系統的な研究は、現在最も重要である。

今回、構造と機能の知られている、冗長性やホモロジーに重複の無い 112 の酵素を選出し、それらの全てについて Linear-Scaling 分子軌道計算³²による正準軌道 (MOs) を求めた。以上の結果から、タンパク質の MOs は電子の非局在化の度合いによって 3 つのグループに分類できることを明らかにした。最初のグループは非常に局在化した MOs に関係し、2 番目のグループは電子がほぼタンパク質構造全体に広がった MOs であり、3 番目のグループはタンパクの共有結合に沿って形成された中程度の大きさの MOs である。あるタンパク質のフロンティア軌道が含まれている、最初の MOs に関しては系統的な解析を実行した。そして、フロンティア軌道の位置と実験的に決定された各酵素の機能部位を比較した。その結果、フロンティア軌道と酵素の機能部位は、水和した際に共局在化する明瞭な傾向が見つかった。最後に、フロンティア軌道の探索がどの程度酵素の機能部位を予測する力があるかを議論する。

2. 2 方法

酵素の選出

基本的に、活性部位の既知である酵素構造は以下のデータベース^{33,34}から取得した。

- ・ CATRES(<http://www.ebi.ac.uk/thornton-srv/databases/CATRES/>)
- ・ Catalytic Site Atlas(<http://www.ebi.ac.uk/thornton-srv/databases/CSA/>)

その後、以下の条件を満たす構造を選出した。

1. 分解能が 3Å より良い X 線結晶構造
2. C 末端を除く重原子に欠陥のない構造
3. 天然構造 (1LXA, 1BWP, 1CHK, 1YTW, 2AID, 12AS だけは変異体)
4. 一次配列に関して冗長性の無い構造

この時点で、178 の構造を選出した。次に、異なる EC 分類の間での内訳が、PDB に登録されている全酵素に対応した値とほぼ一致するように選出した。最終的に 112 の酵素を選出した (Table S1)。

構造最適化と分子軌道計算

ある酵素が対象なサブユニットから構成される場合、2 個の等価な座標データから一方を削除した。遷移金属や阻害剤、補酵素が存在する場合には、それらを除外した。これらの前処理を行った後、分子軌道計算に用いた初期構造を以下の手順で作成した。

1. 欠落している水素原子を結晶構造の重原子に付加
2. Arg、Lys、Asp、Glu のイオン化状態を pH=7 の条件下で設定
3. タンパク内部にある His のプロトン化状態は、その周囲にある極性残基との間に形成される水素結合ネットワークが最適化されるようなパターンに決定
4. Asn と Gln の側鎖に関するコンフォメーションは、その周囲にある極性残基との間に形成される水素結合ネットワークが最適化されるような配向に決定

これら 1-4 のステップはプログラム Reduce³⁵を用いて実行した。初期構造の水素原子の位置は我々が開発したタンパク質構造の高速最適化アルゴリズム³⁶によって実行した。その結果得られた構造は、MOZYME³⁴として知られている代表的な Linear-Scaling 分子軌道法、AM1 ハミルトニアン³⁷を用いて一点構造の計算を実行した。溶媒効果の計算は、COSMO³⁸として知られている連続体近似を比誘電率 78.4 によって実行し、酵素の各原子には Klamt らによって提案された原子半径³⁹を設定した。MOZYME の計算によって得られた一電子波動関数は、ユニタリー変換によって全て正準軌道に変換した。全ての MO 計算は MOPAC20002 パッケージ⁴⁰を使って実行した。Lysozyme に関する *ab initio* Hartree-Fock

の計算には Gaussian03 program⁴¹ を用いた。

分子軌道の解析

得られた正準軌道は以下の手順で解析した。

$$\phi_i = \sum_m^n c_{i,m} \chi_m \quad (1)$$

$$\sum_m^n c_{i,m}^2 = 1 \quad (2)$$

$$D_i(a) \equiv \sum_{m \in a} c_{i,m}^2 \quad (3)$$

$$D_i = \frac{1}{N} \frac{1}{\sum_a D_i^2(a)} \quad (4)$$

(1)式の準位 i の分子軌道 ϕ_i は n 個の原子軌道 χ_m ($m=1,2,\dots,n$) の線形結合によって記述され、(2)は ZDO 近似⁴² による ϕ_i の規格化条件である。(3)で定義したパラメータ $D_i(a)$ は準位 i の分子軌道に関する原子 a 上の電子密度である。ある残基の全原子に関して $D_i(a)$ の和をとることによって、その残基上の準位 i の分子軌道に関する電子密度が評価できる。本研究では、 ϕ_i の電子密度が特定の残基で 0.2 以上の値であれば、その軌道は残基に局在化していると判定した。(4)式の D_i は軌道 i の非局在化の程度を定量化し³⁰、 N はある酵素の原子数である。電子が複数の全原子に非局在化している場合を考えるなら D_i は 1 になる。逆に単一の原子に局在化しているなら D_i は $1/N$ となる。

変性 Lysozyme のモデリング

変性した Lysozyme の構造は、600K の高温条件下で分子動力学 (MD) シミュレーションを 1ns の時間実行することによって取得した。初期構造は、PDB コード 1HEL の構造⁴³ を、最急降下法によるエネルギー最小化計算を実行することによって取得し、ジスルフィド結合-S-S-は 2 つの-S-H 基に還元した。MD シミュレーションは仮想溶媒モデル (一般化ボルンモデル⁴⁴) を用いて実行した。温度制御は Week-Coupling 法⁴⁵ によって実行した。全ての MD 計算は、Amber8 パッケージを用いて実行した⁴⁶。

2. 3 結果

タンパク質の分子軌道の解析

我々は構造と機能が共に既知である 112 の酵素を選出した (Table 1S)。Table I は 6 つの EC 間の酵素数の内訳を示しており、異なる EC クラス間での相対的な割合は、それに対応する PDB に登録されている全酵素に対応する割合とほぼ同じになるように取得した。112 酵素では酵素当たりのアミノ酸残基が 100 残基から 500 残基の範囲に分布しており、平均して 271 である。2. 2 方法に述べた手順に従い、これらの酵素の分子軌道計算を実行した。

式(1)に示すように、ある MO はタンパク質の各原子上にある原子軌道の線形結合で表される。今回我々が使用した半経験的分子軌道法では、水素原子は 1 個の 1s 原子軌道で記述され、第一周期の元素 (C・N・O) は 4 つの原子価軌道 ($2s \cdot 2p_x \cdot 2p_y \cdot 2p_z$) によって記述される。あるタンパク質の全 MOs の数は(1)式で用いられている原子軌道の数と等しい。研究対象の 112 の酵素では、タンパク質当たりの MOs の数が 3095 軌道から 19058 軌道までの範囲にあり、平均して 10638 軌道であった。1 個のタンパク質分子でさえ、全 MOs を一つずつ調べるのは困難である。代わりに、電子の非局在性(MO の大きさ)や軌道のエネルギーなどの物理的な性質によって、それらを分類することを試みた。

Table I. 6 つの EC クラスに関する酵素の割合

EC	This study		PDB	
	No. of structures ^a	Ratio (%)	No. of structures ^b	Ratio (%)
EC1	19	16.9	3609	17.9
EC2	21	18.8	5346	26.6
EC3	49	43.6	8233	40.9
EC4	12	10.7	1409	7.0
EC5	8	8.0	891	4.4
EC6	3	2.7	625	3.1

^a 本研究に用いた 112 の酵素 ^b PDB に登録されている全酵素

電子の非局在化の度合いは(4)式に示した D 値によって定量化できる。D 値が大きいほど電子は非局在している。そこで、タンパク質にある全 MOs の D 値を軌道エネルギーに対し

てプロットした。Figure 1(a)は気相中の Lysozyme (HEWL、PDB:1HEL) に対してプロットした例である。HEWL は合計 4964 の軌道を持ち、それらは 2796 の被占軌道と 2195 の空軌道から成る。予想した通り、被占軌道と空軌道の両方はバンド様の構造を構成し、HOMO と LUMO がそれぞれ -11.960 と -10.484 eV に位置していた。Figure 1(a) では以下の特徴が指摘できる。1) D 値はバンドギャップ近傍でかなり小さいこと、2) 被占バンドでは、HOMO から約 -10eV 離れたところで、すなわち -25eV 付近に最大ピークが存在する、3) この最大ピークの左側にいくつかのピークがあり、4) 空バンドには唯一つのピークが存在する。軌道エネルギーが全体的に高いエネルギー領域にシフトしてバンドギャップは増大したが、これらの特徴は水和した状態の酵素(Figure 1(b))にも見られた。そして、研究した他の全ての酵素の MO も、同様の D 値の分布パターンを示した(データ未提出)。

上述した通り、Figure 1(a)(b)の両データは、結果がパラメータや方法に依存する (AM1 ハミルトニアン)、半経験的分子軌道法によって得られたものである。ここで、D 値の分布パターンの普遍性を調べるため、気相中の HEWL に関して *ab initio* Hartree-Fock 計算を実行した。Figure 1(c)は STO-3G レベルの計算で得られた D 値の分布を示しているが、このレベルの計算では C・N・O のような非水素の原子は最小基底 (原子軌道: 1s, 2s, 2px, 2py, 2pz) で近似される。Figure 1(c)を描く際、Figure 1(a)との直接比較をするため、1s 内殻軌道からの寄与は省いた。Figure 1(c)は Figure 1(a)と全体として類似しており、このことは D 値の分布パターンは使用した計算手法 (近似のレベル) にそれほど敏感には依存しないことを示している。

まとめると、Figure 1(a)-(c)に見られる独特の D 値の分布パターンは、タンパク質の固有の性質に由来するものと推測され、おそらく 3 次元構造によるものである。これを確認するために、我々は変性した HEWL に関する分子軌道計算を実行した。これらの構造は、600K の高温条件下 (2. 2 方法) で行った分子動力学シミュレーションの軌跡から得られたものである。Figure 1(d)は真空中での変性した構造に対する D 値の分布を示している。Figure 1(a)と比較すると -25eV あたりの最も大きいピークの高さが変性によって大きく下がっているが、被占バンドの他の領域の D 値には (変性してもその高さは) ほぼ影響がない。それゆえ、この最も高いピークに属する MOs は 3 次構造の構築と密接に関わっている。空バンド (unoccupied band) では、-4 eV 付近の最も高いピークは変性によって減少しており、このピークに属する分子軌道は -25 eV にある最大ピークに対応した反結合性軌道 (anti-bonding counterparts) であることを示唆している。

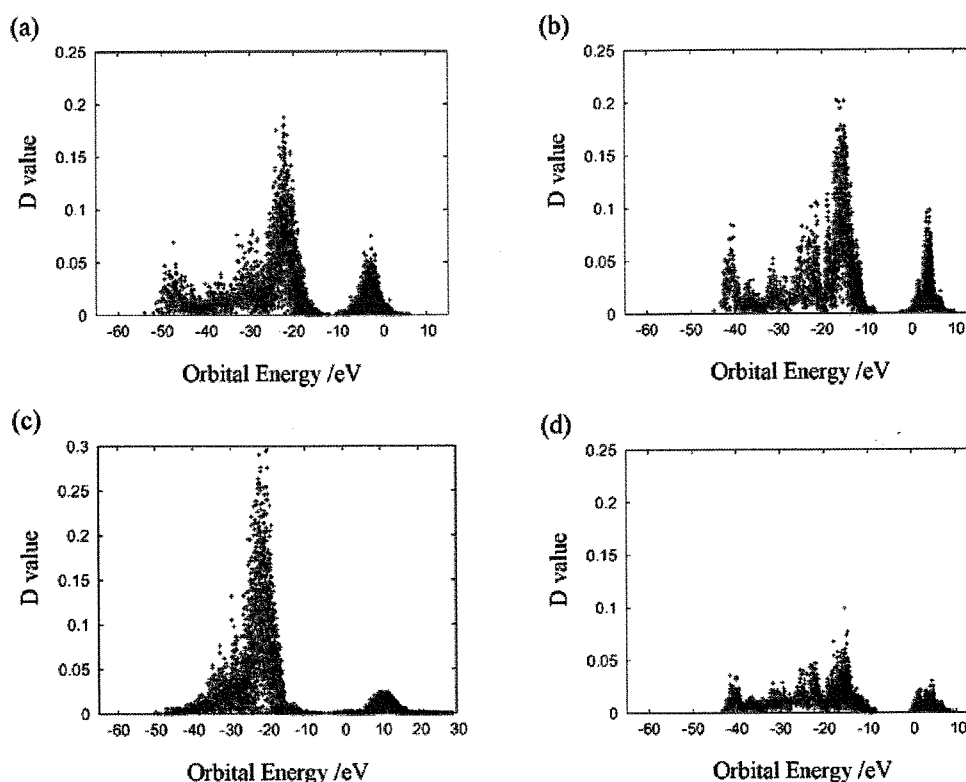


Figure 1. HEWL の天然構造と変性構造の軌道エネルギーに対する D 値のプロット。(a) 真空中の天然構造 (b) 水和した天然構造 (c) 孤立状態の天然構造 (d) 水和状態での変性構造 (a)(b)(d)は MOZYME 計算、(c)は ab-initio Hartree-Fock 計算 (STO-3G) によって得た。

D 値の分布から、HEWL の分子軌道は少なくとも 3 つのグループに分類してよいだろう。A) バンドギャップに近接し、極めて局在化した軌道 B) 非常に非局在化し、真空中では -25 eV、水中では -18 eV にピークを作る軌道 c) 最大ピークの傍らでエネルギーが最も低い側にある中程度に非局在化した軌道。ここで再度強調すべきことは、これらの分類はタンパク質が真空中か水中であるかに関わらず維持されていることである。これらの A-C のグループを詳細に特徴付けるため、それぞれのグループの代表的な軌道を可視化した。水和した酵素に関する結果を Figure 2 に示す。Figure 2(a) はグループ A の代表として HOMO と HOMO-1 を表している。明らかに、これらの軌道は非常に局在化し、それぞれ Asp52 と Glu35 の上に最大振幅を持っており、これらの部位はリゾチームの触媒基である。Figure 2(b) は、グループ B の最も D 値が大きい軌道を示している。この軌道は分子全体に非局在化した軌道であり、よく調べると近接した残基間で through-space 相互作用(電子雲の重なり)が形成されていることが分かる。なぜこのような軌道の非局在化が形成されるかはあとで議論する (2.4 考察)。Figure 2(c) は Figure 1(b) の -43 eV のピークに対応する軌道を示している。この軌道は共有結合に沿って最大振幅を持つ。このような特徴は、約 -45 から

-35eV の間の（エネルギーが）低い領域にある軌道の中に共通に見られた。

他の酵素の分子軌道群も A)-C) の 3 グループに分類できるだろう。というのは、それらも Figure 1(a)(b) に類似した D 値の分布パターンを示すからである（データ未提出）。以上の HEWL では、グループ A の軌道、HOMO と HOMO-1 は酵素の触媒部位に共局在化している（Figure 2(a)）。グループ A の探索は酵素の機能部位の予測には重要であり、同様の結果が他の酵素にも期待される。これは次の節で詳細に議論するつもりである。グループ B の軌道に占有された電子は、その非局在性から多くの数の原子に共有されており、それゆえこのタイプの軌道は酵素の構造全体（3 次構造）、つまり、フォールドした状態の安定化に寄与しているのだろう。対照的に、グループ C の軌道はタンパク質の共有結合の構築に関与しているのだろう。要するに、タンパク質の MOs は、その機能的な役割に応じて大まかに 3 つのグループに分類できると解釈することが妥当である。

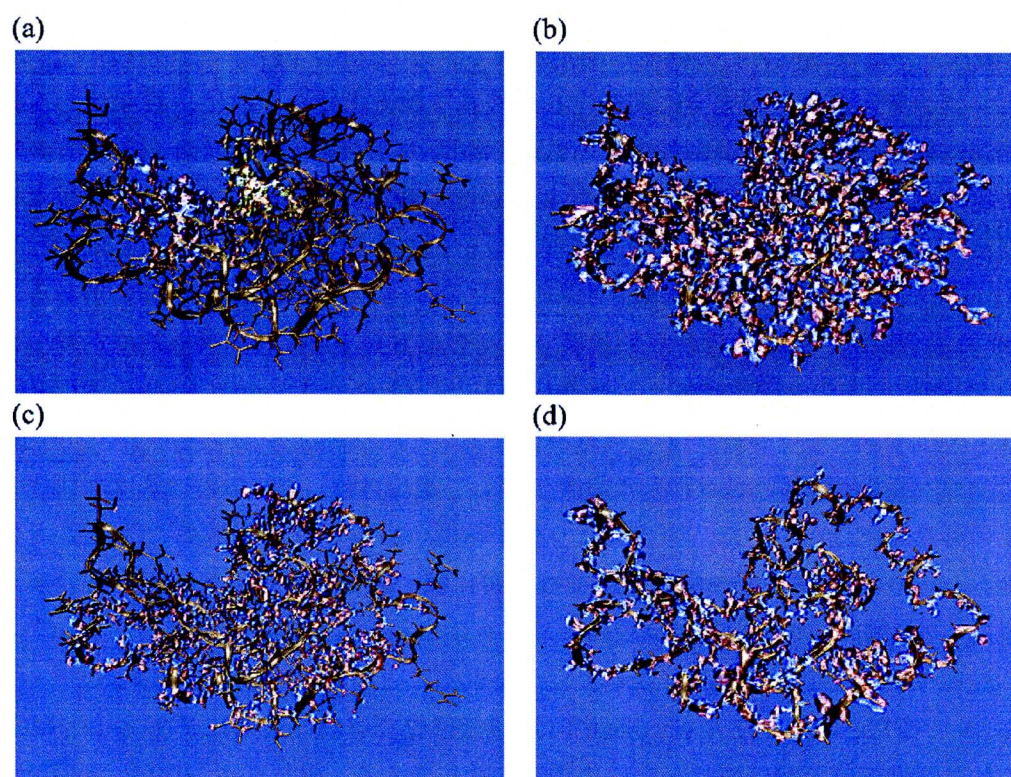


Figure 2. HEWL と poly-Gly の分子軌道に関する 3 次元グラフィックス。(a) HEWL の HOMO (赤と青) と HOMO-1 (緑と黄色) (b) HEWL の代表的な group B 軌道 (c) HEWL の代表的な group C 軌道 (d) poly-Gly の代表的な group B 軌道 各 MO イメージは振幅 ± 0.0001 (e/Bohr^3) の表面を表す。poly-Gly の主鎖は HEWL と同一。

フロンティア軌道と酵素の機能部位との関係

フロンティア電子軌道理論によれば、HOMO と LUMO は高い化学反応性を持ち、それらは酵素の活性部位に位置することが予想される。厳密な意味では、“フロンティア軌道”という用語は与えられた分子の HOMO と LUMO のみを指す。しかしながら、広範囲の酵素に適切な理論を見いだすためにはフロンティア軌道の定義を修正したほうが良いだろう、というのはタンパク質の MOs はバンドに類似した構造を形成するためである。この目的のため、HOMO-LUMO バンドギャップ付近に位置する MO を 2 つのグループに分割した。1 つは HOMO- $\Delta\epsilon$ から HOMO の被占 MOs を含み、他方は LUMO から LUMO+ $\Delta\epsilon$ の空の MOs を含む。 $\Delta\epsilon$ は MOs の数を表す整数もしくは eV で与えられたエネルギー値を表す。以下、前者と後者はそれぞれ HOMO バンドと LUMO バンドと命名し、フロンティア軌道はこれから 2 つのバンドに含まれた MOs とする。

選出した 112 酵素に関して、データベース CATRES/CSA^{35,36} では、基質結合部位を含む 419 アミノ酸残基が活性部位として機能が同定されている (Table 1S)。これらの中で、上で定義したフロンティア軌道が完全に局在化した残基を数え上げた。次に、我々はある軌道がどの残基に局在化しているかを判定するため、以下の基準を採用した。ある軌道の電子密度が特定の残基上で 0.2 以上の値ならば、軌道がその残基に局在化していると判断する (2. 2 方法)。各酵素について、このような解析を孤立状態と水和状態で別々に実行した。Table II に見られるように、孤立状態と水和状態に関する結果には、大きな違いがある。例えば $\Delta\epsilon=0$ の場合には、孤立状態と水和状態でそれぞれ 8 個と 28 個の活性部位が選出され、後者の 28 部位の中の 2 部位は同一の酵素に属した (Carbon-Oxygen Lyase, PDBID: 5ENL)。 $\Delta\epsilon$ の増加に伴い、フロンティア軌道が局在化した活性部位の数は増加する。 $\Delta\epsilon=50$ の場合には、293 個の活性部位が水和した酵素に関して選出され、それらの残基は本研究で扱った全ての酵素を網羅していた。酵素当たり分子軌道の平均の数は 10638 であることを考慮すると、 $\Delta\epsilon=50$ はエネルギー領域全体では非常に狭い範囲に対応している。言い換えれば、酵素はこの特別な MOs を機能発現のために有効利用している。そして、フロンティア軌道と活性部位の共局在化が水和によって促進されることを強調するべきである。

もちろん、 $\Delta\epsilon$ で指定されたエネルギー範囲では、機能部位ではない残基に局在化した MOs もあり、これは活性部位予測の観点からは偽陽性の結果である。それゆえ、Table II のみからでは活性部位の MOs は非活性部位の MOs よりもバンドギャップ近傍に位置する傾向が高いか否かを判定できない。このことを調べるため、我々は $P_{active}(\Delta\epsilon)$ と $P_{all}(\Delta\epsilon)$ の 2 種類の確率を以下に定義した。

$$P_{active}(\Delta\epsilon) = \frac{\text{指定したエネルギー領域}\Delta\epsilon\text{に軌道が落ち込んだ活性部位数}}{\text{112酵素に関する活性部位の総数}}$$

$$P_{all}(\Delta\epsilon) = \frac{\text{指定したエネルギー領域}\Delta\epsilon\text{に軌道が落ち込んだ活性部位数}}{\text{112酵素に関する残基の総数}}$$

Table II. $\Delta\epsilon$ で定義したフロンティア軌道に属する MOs と機能部位が共局在化した酵素・機能部位の数

$\Delta\epsilon^a$	Isolated state		Hydrated state	
	No. of	No. of	No. of	No. of
	enzymes ^b	functional sites ^c	enzymes ^b	functional sites ^c
0	8 (1.9)	8 (7.1)	28 (6.7)	27 (24.1)
5	31 (7.4)	25 (22.3)	89 (21.2)	62 (53.4)
10	53 (12.7)	38 (33.9)	136 (32.5)	89 (79.5)
30	107 (25.5)	63(56.3)	235 (59.1)	108 (96.4)
50	155 (37.0)	77(68.6)	293 (69.0)	112 (100.0)

a フロンティア軌道の閾値（整数）。フロンティア軌道は被占軌道では HOMO から HOMO- $\Delta\epsilon$ 、空軌道では LUMO から LUMO+ $\Delta\epsilon$ を参照する。

b 機能部位が定義したフロンティア軌道に属する MOs と共局在化した酵素数と相当した機能部位数。括弧内の指標は全酵素（112）の対する割合（%）。

c 定義したフロンティア軌道に属する MOs と共局在化した機能部位の数。括弧内の指標は全機能部位（419）の対する割合（%）。

分母の $P_{active}(\Delta\epsilon)$ と $P_{all}(\Delta\epsilon)$ はそれぞれ 419 と 20780 であり、 $P_{active}(\Delta\epsilon)$ は Table II の 5 列目に示した確率に等しい。Figure 3(a)(b) では両方の確率が $\Delta\epsilon$ に対してプロットされている。ただし、これらは水和した酵素の結果を示している。 $\Delta\epsilon$ が 0 から 50 の時、 $P_{active}(\Delta\epsilon)$ は $P_{all}(\Delta\epsilon)$ の 3 倍以上大きい。 $\Delta\epsilon = 50$ はエネルギー単位で 1.0 eV の間隔とほぼ等しい。それゆえ、Figure 3(b) では $\Delta\epsilon < 1.0$ eV の時に $P_{active}(\Delta\epsilon)$ が $P_{all}(\Delta\epsilon)$ よりもかなり大きい。これらの結果は、活性部位に局在化する傾向の高い分子軌道がバンドギャップ付近に集中していることを意味する。

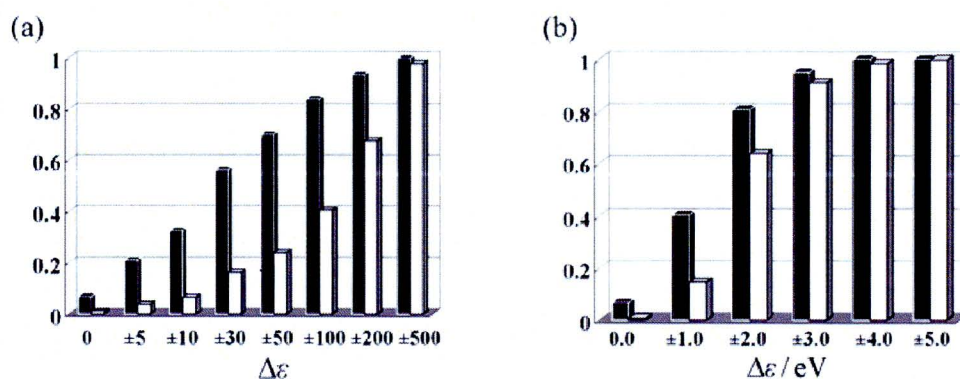


Figure 3. $P_{active}(\Delta\epsilon)$ (black) と $P_{all}(\Delta\epsilon)$ (white)の比較。(a) $\Delta\epsilon$ は整数値 (b) $\Delta\epsilon$ は eV 単位

以上の結果は Figure 4(a)でより明白に見られる。Figure 4(a)ではバンドギャップの近くに位置する被占・空軌道のエネルギーを、各酵素の HOMO と LUMO エネルギーを横軸としてそれぞれプロットしている。例えば、リゾチーム (HEWL) は HOMO と LUMO のエネルギーが、それぞれ -8.091 と -2.325 eV なので、その被占・空軌道のエネルギーを横軸値の -8.091 と -2.325 eV にそれぞれプロットした。次に、活性部位に局在化した MOs のデータと非機能部位に局在化された MOs のデータをそれぞれ赤と緑の点によって表した。HOMO と LUMO の両方が活性部位に局在化する理想的な場合には、2 つの赤い点が対角線上に並ぶはずである。研究対象とした水和した酵素では、赤の点是对角線近傍に集中し (Figure 4(a))、孤立 (系) の酵素では赤の点は大きくばらついている (Figure 4(b))。これらの結果は、明らかに、水和に伴って活性部位の MOs がバンドギャップ付近に位置するように、エネルギーレベルの再序列化が起こることを示している (Table II)。したがって、酵素の活性部位のより良い予測に電子構造計算への水和効果の導入は不可欠である。

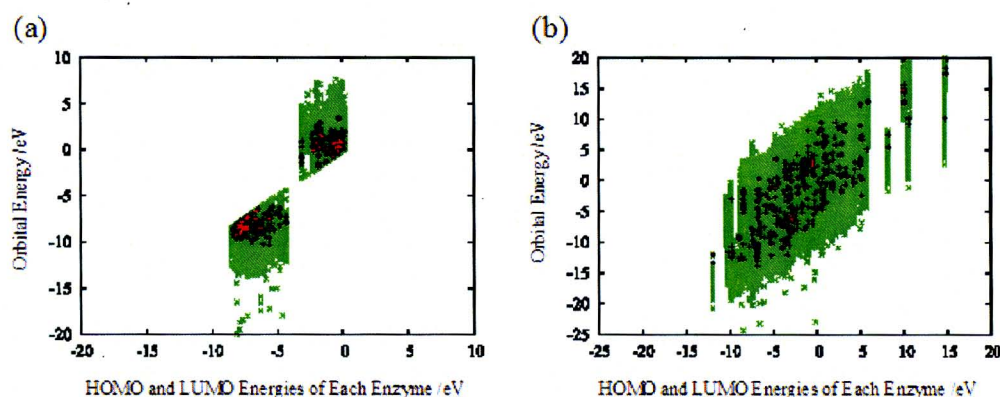


Figure 4. HOMO・LUMO バンドにある MOs エネルギーに関する二次元プロット。(a) 水和した酵素 (b) 孤立系の酵素 横軸の値は酵素の HOMO・LUMO のエネルギーを示し、対応する縦軸の値は酵素の HOMO-LUMO バンド付近に位置する MO のエネルギーである。横軸

には 112 酵素に関する HOMO・LUMO の全数に対応する 112×2 のデータがある。共局在化した MOs のエネルギーは、それぞれ赤と緑で機能部位と非機能部位を表している。

ある $\Delta\epsilon$ で定義された HOMO/LUMO バンドにどのような種類の残基がより集中しているかを調べることは興味深い。この目的のため、研究対象の 419 の活性部位を以下の 7 つのグループに分類した。これらは、酸性タイプ(Asp・Glu)、塩基性タイプ(Lys・Arg)、His、極性タイプ(Gln・Asn・Ser・Thr)、Cys、芳香環状タイプ(Trp・Phe・Tyr)、その他である。Figure 5(a)では研究対象の活性部位の総数 (419) に対する各グループの母比率を示している。Figure 5(a)は、軌道のエネルギーが $\Delta\epsilon = 50$ で定義した HOMO/LUMO バンドに落ち込んだ活性部位を分類した結果を表し、データは水和したタンパク質に関するものをまとめたものである。Figure 5(b)に示すように、酸性タイプの残基は 34%の確率で見いだされ、それらの母集団中での割合である 28%よりも高い (Figure 5(a))。このように、酸性タイプの活性部位残基に局在化した MOs はバンドギャップ付近に位置する有意な傾向がある。対照的に、極性タイプの残基(Gln・Asn・Ser・Thr)は、(特に Ser は)触媒残基として重要な役割を果たす場合が多いにもかかわらず、それら母比率(16%)と比較して相対的に HOMO と LUMO バンドで見出す頻度が低くなっている。Figure 5(c)は、同じ HOMO/LUMO バンドに軌道のエネルギーが落ち込んだ非活性部位に対して類似の解析を行った結果を示している。芳香族タイプの残基が最も高い確率である 32%で見出せることに注意すべきである。機能部位予測の観点から、それらは偽陽性の主な原因である。このような都合の悪い結果の原因は後で議論する。

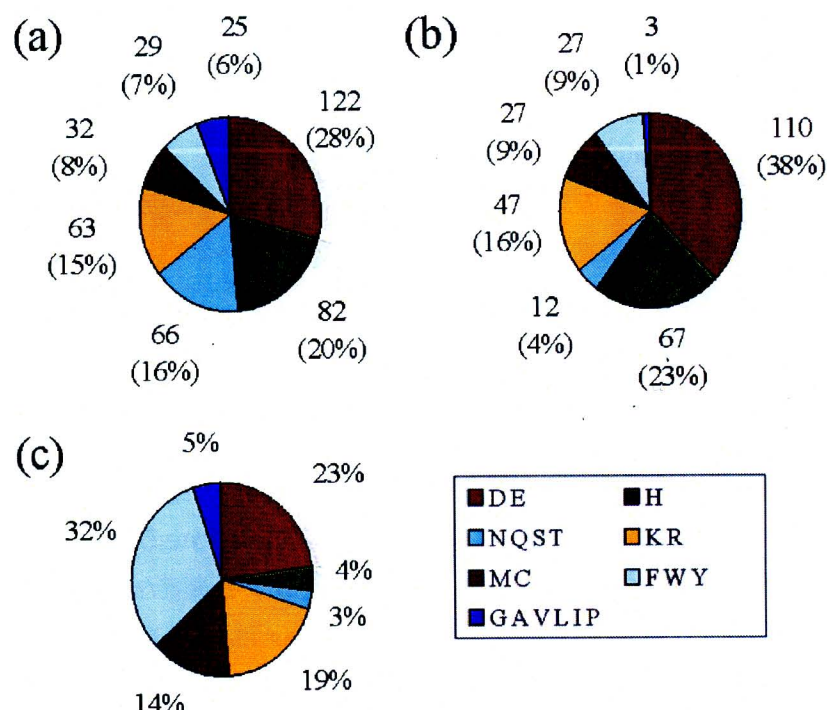


Figure 5. (a) 化学的な性質による機能部位（全 419 残基）の分類。 (b) 軌道が $\Delta\epsilon=50$ で定義した HOMO・LUMO バンドにある機能部位の分類、データは水和した条件。 (c) 軌道が HOMO・LUMO バンドにある非機能部位の分類（偽陽性）

HOMO/LUMO バンドをより詳細に分析するため、前に定義した $P_{active}(\Delta\epsilon)$ と $P_{all\ e}(\Delta\epsilon)$ に類似した確率を各化学種について計算した。例えば、酸性残基の $P_{active}(\Delta\epsilon)$ と $P_{all\ e}(\Delta\epsilon)$ は以下のように計算される。

$$P_{active}(\Delta\epsilon) = \frac{\text{指定したエネルギー領域}\Delta\epsilon\text{に軌道が落ち込んだ酸性活性部位数}}{\text{112酵素に関する酸性活性部位の総数}}$$

$$P_{all}(\Delta\epsilon) = \frac{\text{指定したエネルギー領域}\Delta\epsilon\text{に軌道が落ち込んだ酸性活性部位数}}{\text{112酵素に関する酸性残基の総数}}$$

Figure 6 は、eV で与えられた $\Delta\epsilon$ に対して以上の確率プロットした結果である。酸性タイプの残基の場合は、 $\Delta\epsilon < 0$ の領域で $P_{active}(\Delta\epsilon)$ は $P_{all}(\Delta\epsilon)$ よりも顕著に大きい、 $\Delta\epsilon > 0$ の領域では順序が逆転している。His と Cys の場合には、調べたエネルギー領域全体にわたって $P_{active}(\Delta\epsilon)$ は $P_{all}(\Delta\epsilon)$ よりも少し大きい値を持つ。他の化学種では、エネルギー領域全体にわたって $P_{active}(\Delta\epsilon)$ の値は $P_{all}(\Delta\epsilon)$ の値とそれほど大きな違いはない。したがって、Figure 3・4(a)から推論されるように、フロンティア軌道群と活性部位の共局在化は主に酸

性のタイプの残基と His、Cys の寄与であると見なせる。

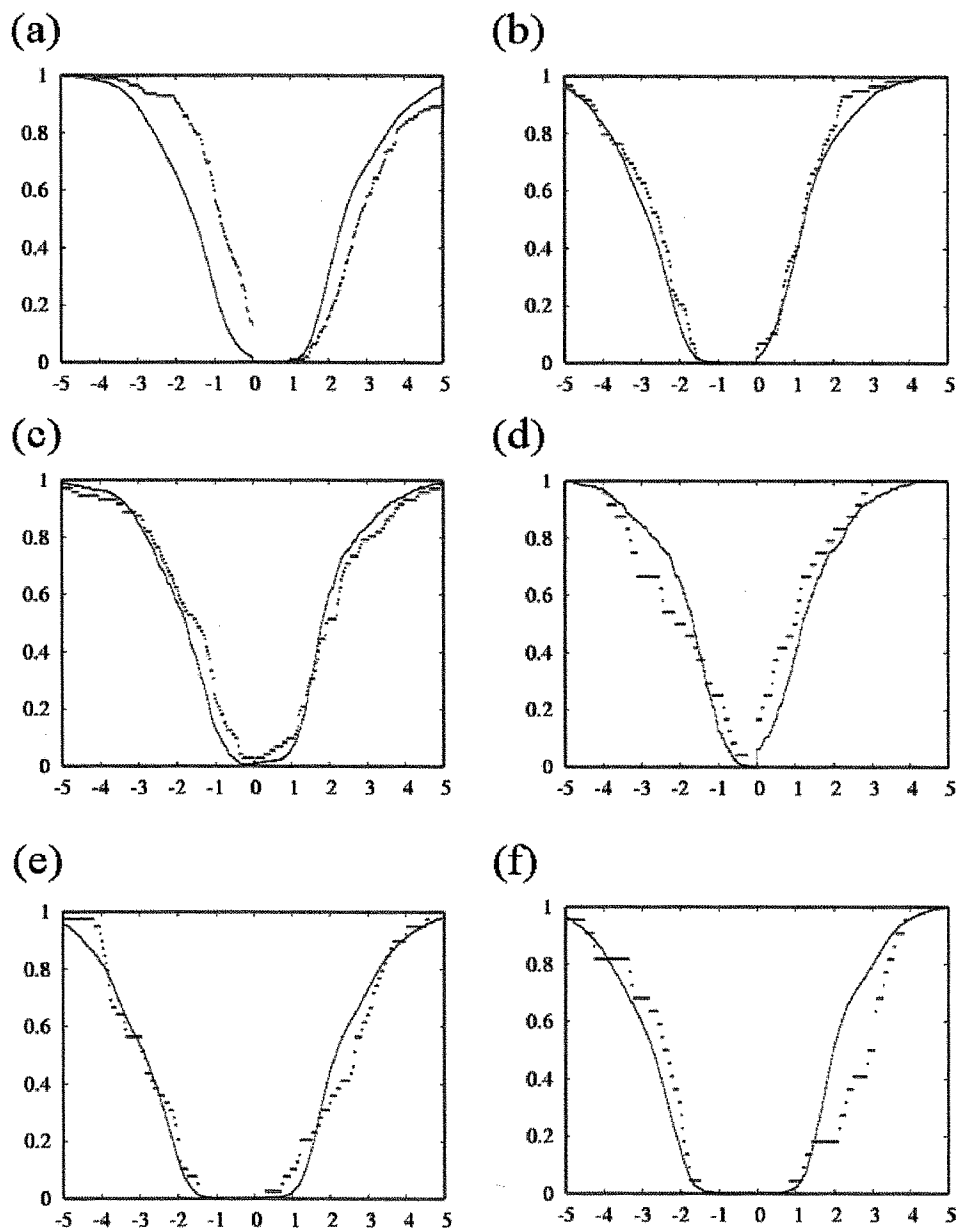


Figure 6. $\Delta\epsilon$ (eV)に対する $P_{active}(\Delta\epsilon)$ (点線) と $P_{all}(\Delta\epsilon)$ (実線) のプロット。(a) Asp と Glu (b) Arg と Lys (c) His (d) Cys (e) Ser と Thr (f) Asn と Gln

2. 4 考察

タンパク質の分子軌道の特徴

言うまでもないが、分子の電子波動関数とエネルギーは 3 次元構造(原子座標)から Schrödinger 方程式を解くことによって一義的に決まる基本量である。原理的に、電子波動関数は分子の全ての物理・化学的性質に関する情報を暗に含んでいる。それゆえ、タンパク質の電子波動関数の詳細な分析は、構造と機能の因果関係について良い洞察を提供すると期待できる。しかし、タンパク質科学の量子化学計算はタンパクの切り出されたモデルもしくは量子力学部分と古典的(部分)の混合系(混合 QM/MM 系⁴⁷⁾)に適用されてきた。これらのアプローチはもちろん、活性部位での化学反応の精密なエネルギー論を展開するは強力であるが、膨大なタンパク質について構造と機能の統一的な知見を得るには適切ではないだろう。それに代わって、本研究は、タンパク質、特に酵素の電子波動関数について調査した初めての系統的な試みである。

タンパク質は極性やイオン、疎水性側鎖を含む様々な単量体ユニットから成る、共重合体(ヘテロポリマー)である。結果、タンパク質の MOs は非常に複雑な構造を持っている。しかし、D 値の分布を解析することによって、それらをグループ A-C の 3 種類に分類することが可能となった。以下、各グループの性質をより詳細に議論する。

以下の 2 つの状態を満たす軌道をグループ A と呼ぶことにする。すなわち、D 値が非常に小さく (<0.01)、それらの軌道エネルギーが与えられたタンパク質の HOMO か LUMO に近接する軌道である。ここでは、どのタイプの MOs がこれらの条件を満たすかを考える。(4)式から、ある MO が 0.01 程度の D 値を持っているならば、その空間分布は多くとも 1 つのアミノ酸残基しか収まらない小空間に限られていると期待できる。一方、分子の高エネルギー被占軌道は、一般に極性(荷電性)残基の孤立電子対と芳香環側鎖の π 軌道に生成するが、これらは単一側鎖の外に広がらない。それゆえ、これらの種類の軌道は以上の 2 つの条件を同時に満たし、活性部位に位置するかどうかに関わらずグループ A の主要なメンバーとなる。

重要なことは、なぜタンパク質にグループ B のような非常に非局在化した軌道が形成されるのか理解することである。このことを調べるため、HEWL の場合について軌道エネルギーに対して状態密度をプロットした (Figure 7(a))。明らかに、状態密度と D 値分布のプロファイルは良く類似している (Figure 7(b))。特に状態密度が大きいほど D 値が大きくなることに注目すべきである。したがって、エネルギー的に近接した軌道間相互作用が分子軌道間の混合を引き起こし、グループ B のような非常に非局在化した MOs の構築につながったと解釈するのが妥当である。恐らく 2 種類の軌道間相互作用は、このような分子軌道の大きな非局在化に関係している。一つは Figure 2(b)で例に挙げたアミノ酸残基の側鎖を

介した **through-space** 相互作用であり、波動関数は残基間の空間で相対的に大きな振幅を持つ。他者は主鎖のアミド基間の軌道相互作用である。タンパク質内のアミド結合はほとんど等しいエネルギーを持つため、大きく非局在化した軌道を容易に形成する。このようなタイプの軌道を拾い出すため、Figure 1(b) (HEWL) に示したのと同じ主鎖構造を持つ仮想タンパク質 (poly-Gly) について MO 計算を実行した。Poly-Gly の中で最も大きな D 値を持つ軌道を Figure 2(d) に示した。この軌道は主鎖に沿って非局在化している。タンパク質の 3 次元構造には、両方のグループ B 軌道が不可欠であろう。

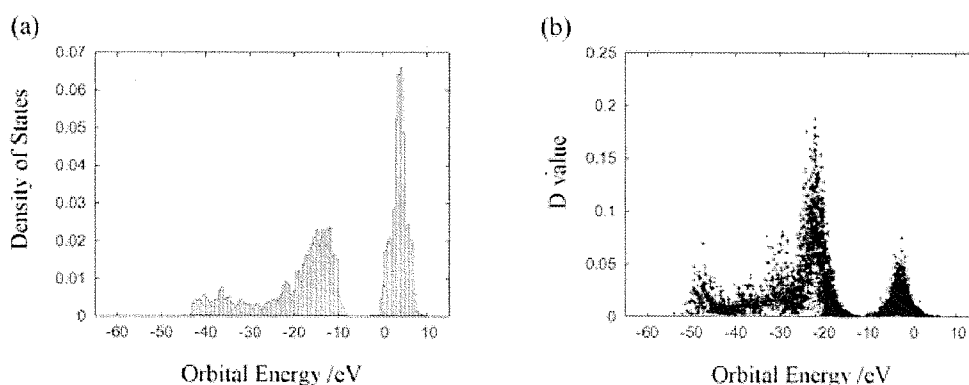


Figure 7. 水和した HEWL の D 値と状態密度の比較。(a) 軌道エネルギー E に対する状態密度 (b) D 値のデータは Figure 2(a) を参照した。状態密度はエネルギーが E から $E+\Delta\epsilon$ にある軌道数であり、 $\Delta\epsilon$ は 5 eV にとった。

既述の通り、グループ C 軌道は共有結合に沿って比較的大きな振幅を持っているため、タンパク質全体もしくはその断片に対する共有結合の骨格形成に寄与している。しかし、グループ B とグループ C の間には決定的な境界は無い。もし、Figure 1(a) を例にとれば、エネルギーの範囲が -35 から -30 eV までに位置する大半の MOs に関して、その起源が **through-space** (空間を通り抜けた相互作用) か **through-bond** (共有結合を介した相互作用) かどうかを厳密に区別できない。この領域では、軌道エネルギーの増加に伴って、分子軌道の性質がグループ C からグループ B に僅かに変化しているように見て取れる。

酵素の電子状態に関する水和の効果

Figure 1(a)(b) に挙げたように、水和は酵素の MOs について主に 2 つの変更を引き起こす。1) エネルギーが高い側へ全体的にシフトすること 2) バンドギャップの増加 である。本計算では、溶媒効果は水と等しい 78.4 の誘電率を持っている分極可能な連続体を介した酵素の静電相互作用しか考慮していない。一般に分子の水和は反作用場を生成するために溶質を取り囲む溶媒 (誘電場) の分極を引き起こし、これが溶質の背景になる。その結果、

溶質は分極され、溶質自体の内部エネルギーが増加する。これは、水和によって軌道エネルギーが高い方へシフトすることを説明する。

上の発見 2)はバンドギャップ付近のエネルギーレベルは溶媒によって特に強い摂動を受けることを示している。これは Figure 4(a)(b)が互いに大きく異なるプロファイルであることを反映している。すなわち、Figure 4(a)では研究対象の酵素の HOMO_s が分布しているエネルギー領域は、LUMO_s の分布領域と明確に分けられているが、このような分離は孤立した酵素 (Figure 4(b)) では不明瞭である。Figure 8 は各酵素の被占・空軌道のエネルギー領域を示している。孤立した状態では、酵素種によって HOMO と LUMO のエネルギーは異なり、HOMO-LUMO バンドギャップが非常に小さい (Figure 8(a)) 一方で、水和状態では異なった酵素で HOMO と LUMO のレベルが殆ど一線上にあり、HOMO-LUMO ギャップがより大きくなっている (Figure 8(b))。その結果、構成 MO_s の分布したエネルギー空間 (という視点) から見た各酵素の性質は、水中では曖昧になる。このような水和によって誘起される軌道エネルギーのシフトは、フロンティア軌道と機能部位の共局在化を促進する (Table II, Figure 3・4)。

タンパク質科学では、水がタンパク質の構造を維持する際に重要な役割を果たすということが広く受け入れられている。例えば疎水性相互作用は、変性状態と比較して天然型タンパク質の界面構造を安定化することが知られている。本研究は、そのような熱学的効果に加え、水和がタンパク質の電子状態への大きな影響を及ぼし、その化学反応性を変調させることを明らかにした。

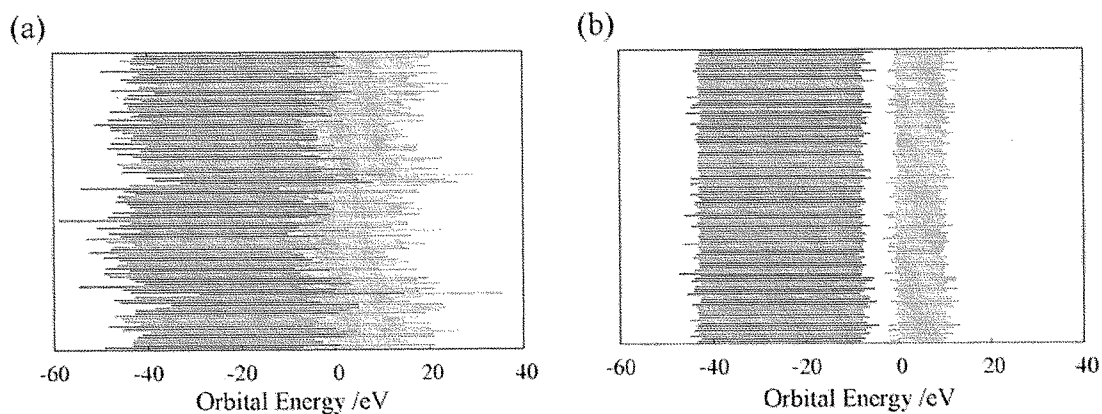


Figure 8. 研究に用いた各 112 酵素に関する被占・空軌道バンドのエネルギー領域。(a) 孤立系での結果 (b) 水和状態での結果。112 酵素のデータは 2 つの平行線である。すなわち、暗灰色と明灰色はそれぞれ酵素の被占・空軌道バンドにまたがっている。

機能部位の予測可能性

最後に、酵素の機能的に重要な残基を特定することに関してフロンティア軌道の探索がどの程度有用であるかを議論する。方法の性能を測るため 4 つの指標を導入した。これらは、回収率 (Recall)、Precision (精度)、透過率 (Filtration Ratio (FR))、偽陰性率 (False Positive Rate (FPR)) である。文献 21 から、これらの指標は以下のように定義される。

$$\text{Recall} = \frac{\# \text{活性部位として予測した残基}}{\# \text{機能同定されている活性部位残基}}$$

$$\text{Precision} = \frac{\# \text{活性部位として予測した残基}}{\# \text{予測法が検出した残基}}$$

$$\text{FR} = \frac{\# \text{予測法が検出した残基}}{\text{全ての残基}}$$

$$\text{FPR} = \frac{\# \text{偽陰性の値}}{\text{機能が同定されていない全ての残基}}$$

Table III. 機能部位予測法の結果比較

Method	Recall (%)	Precision (%)	Filtration Rate (%)	False Positive (%)
This Study $\Delta\epsilon=0$	6.8	12.5	0.9	0.76
This Study $\Delta\epsilon=5$	26.4	8.2	4.2	3.97
This Study $\Delta\epsilon=10$	39.3	7.0	7.4	7.05
This Study $\Delta\epsilon=30$	62.2	4.8	18.3	17.67
This Study $\Delta\epsilon=50$	77.1	4.1	26.9	26.22
THEMATICS ^a Z=0.99	53.4	21.9	3.2	2.60
Q-SiteFinder ^a (top 3 sites)	65.6	5.4	14.9	14.3
SARIG ^a	57.4	8.1	8.7	8.15

^a これらのデータは文献 18 を引用している。THEMATICS に関しては Z スコアの閾値として 0.99 を引用した。方法の詳細は対応する原論文にて参照のこと。THEMATICS が 16 と 17、Q-SiteFinder が 19、SARIG が 20。

低い FR と FPR 値で Recall と Precision が最大化される時に性能が良いと判別する。Table III では、この研究の方法と他の 3 次元構造に基づいた方法、すなわち THEMATICS・Q-Site Finder・SARIG（残基間相互作用グラフ（理論）に基づく構造分析）とを比較している。現在の結果は Table 1S で挙げた 112 酵素の平均値であるが、後者 3 つの方法は他の 75 酵素に関する結果である。本研究と文献ではサンプルが異なっているが、Recall・FR はデータベース CATRES/CSA に対して測定されており、結果は同定された既知の活性部位である。このように、Table III から大雑把な性能の比較が可能である。現在の方法では、 $\Delta\epsilon$ の減少に従って、Precision・FR・FP が良くなるが Recall は悪くなる。 $\Delta\epsilon$ が 30 以上である場合、Recall が他の方法と同等になる。これらの傾向から、最適な $\Delta\epsilon$ は 30 付近だろう。その際の成功率は、THEMATICS や SARIG には敵わないが Q-Site Findern には匹敵する。

Table III によって示したように、現在の方法は比較的大きな FP 値を与える傾向がある。既に Figure 5(c)で指摘したように、HOMO バンドと LUMO バンドに芳香族の残基(Tyr、Phe、Trp)の π 電子が高い確率で集まる傾向があることによる。研究対象の全 419 の機能部位について、機能部位として同定されている芳香性の残基数は僅か 29 であった(Figure 5(a))。そして通常、Tyr を除いて、それらを触媒部位に見つけるのは稀である。これらの事実を考慮すると、芳香性残基に関する雑音（予測誤差）から信号を分離（情報を抽出）するために、より一層の努力することは無意味であるように思われる。それゆえ、芳香性残基は機能部位予測の標的から除いても良いだろう。芳香性残基に局在化した MOs を除いても、 $\Delta\epsilon$ で定義した区間には、まだ他種の偽陽性があるだろう。偽陽性の中から、本当の機能部位に局在化した MOs を特定する見込みは、その残基の化学的な性質に依存する。Figure 6 に示した $P_{active}(\Delta\epsilon)$ と $P_{all}(\Delta\epsilon)$ の相対値は、そのような可能性を測るものである。もし-1.0 から 1.0 までのエネルギー区間($\Delta\epsilon=1.0$ eV)に注目すれば、Glu・Asp の被占バンド側の $P(\Delta\epsilon)_{active}$ は、 $P(\Delta\epsilon)_{all}$ よりもかなり大きな値を持ち、His と Cys の $P(\Delta\epsilon)_{active}$ は $P(\Delta\epsilon)_{all}$ と比べて僅かに大きい値を持つ。しかし、他の種類の残基では明白な信号が観察されなかった。したがって、本研究のアプローチは、潜在的に機能部位が Glu・Asp・His・Cys に位置している場合に限る、予測する力があるのだろう。

以上の議論により、酵素の機能部位を予測するプロトコールを以下のように提案した。

- 1)分子軌道計算を実行
- 2)HOMO から HOMO- $\Delta\epsilon$ ($\Delta\epsilon=50$)までのエネルギー範囲にある被占 MOs を取得
- 3)Glu と Asp に局在化した MOs を選出し、その中から潜在的な機能部位に局在化した最もエネルギーの高い分子軌道を候補として選出
- 4)他の潜在的な活性部位 MOs を、その位置とエネルギーが 3)で選出した軌道と近いという仮定のもと、 $\Delta\epsilon$ によって定義された HOMO と LUMO バンドから探し出す

以上のプロトコールにより、 $\Delta\epsilon$ を全て 50 にした場合について、活性部位の特定に成功した例を Figure 9 に示す。ジエノニルコア・イソメラーゼの場合 (Figure 9(a))、Asp176 に局在化した HOMO が最初の活性部位 MOs として選出された (ステップ 3))。次に、HOMO-1

から HOMO- $\Delta\epsilon$ までの各分子軌道を可視化し、3)で記述した空間・エネルギーの条件を満たす MOs を探した。結果として、HOMO-1 から HOMO-11 までが潜在的な活性部位 MOs であると判別した。これらは Glu194 と Asp204 にそれぞれ局在化しており、Asp176 と共にクラスターを形成していた。実際、これら 3 つの残基は実験的に同定された酵素の触媒残基であった^{48,49}。

カルボキシペプチターゼとリン酸ジアミナーゼに関しては、それぞれ HOMO-5 と HOMO がステップ 3)で選出された：これらは Glu80 と Asp72 にそれぞれ局在化していた。次のステップ 4)では Figure 9(b)(c)に示したものとは異なった潜在的な機能部位残基が選出された。ステップ 4)の際、Glu・Asp とは異なった機能部位残基^{50,51}も（時には）選出が可能である。

近年のコンピュータのハードとソフトの発展により、酵素の正準軌道は Reasonable な CPU 時間で得ることが可能になった：例えば Pentium4 (3.2 GHz clock)と MOPAC2002 program³⁷を用いれば、酵素当たりの一点計算は僅か 1 時間程度で終了する。このように、ステップ 1)は律速ではないが、ステップ 4)は最も時間を要する。なぜなら、MOs の可視化やイメージの分析のような（自動化ができない）手作業を要するからである。この点において、我々の方法は、他の 3 次元構造に基づいた他の手法^{16,17,19,20}と比較して有利であるとは言えない。一方、我々の量子化学計算に基づいた手法は酵素の触媒反応に関係した電子的な効果を直接考慮している特徴がある。理想的には、高い予測力やハイスループットの要求を満たすために、（他の手法を）補助する方法で使われるべきである。そのような組み合わせた方法については現在探索中である。

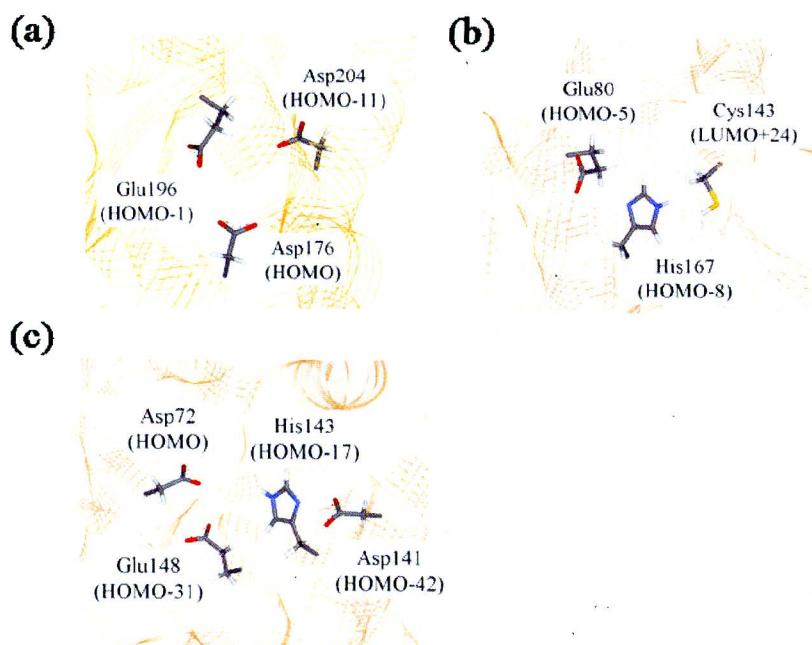


Figure 9. 分子軌道計算に基づいた機能部位予測の例。(a) Dienoyl-coa isomerase (PDB:1DCI). (b) Carboxyl peptidase (PDB:1A2Z). (c) Phosphate deaminase (PDB:1CD5).

2. 5 結論

今回の計算から、我々は3つの重要な結論に到達した。1番目は、タンパク質のMOsは階層的な様式で構築されており、エネルギーの降順にグループA、B、Cと定義できる。フロンティア軌道はグループA軌道に属し、非常に局在化した性質を持っている。これは酵素反応に関する特異性の発現と関連しているのだろう。もし、化学的に活性化された軌道がタンパク質全体にわたって空間的に非局在化していたら、一連の反応（触媒反応）は特定の基質を選択すること無く、酵素表面でランダムに起こってしまうだろう。2番目は、タンパク質の電子構造は水和によって大きな影響を受け、水和は分子軌道の相対準位の変化を引き起こすこと。この結果、フロンティア軌道は活性部位に局在化する。そのため、タンパク質の水和は天然構造の安定化に重要な役割を果たすだけでなく、触媒機能にも重要な役割を果たす。3番目は、現実的な適用にはいくつか問題点は残っているが、フロンティア軌道の分析がタンパク質の機能部位の識別に関して有望なことである。

Table S1. 研究対象にした酵素の一覧

	PDB ^a	a.a. ^b	Protein	RES ^c	EC ^d	Active Sites
1	1ytw	282	Tyrosine phosphatase	2.40	3.1.3.48	E290,H350,H353 W354,P355,D356 Q357,T358,H402 C403, R409,T410
2	1xffA	237	Glucosamine fructose-6 phosphate	1.80	2.6.1.16	C1
3	1apxA	248	Peroxidase	2.20	1.11.1.11	R38,H42,N71
4	2jcw	152	Superoxide dismutase	1.70	1.15.1.1	H46,H48,H63 H71,H80,D83 H120,R143
5	1cseE	274	Serine proteinase	1.20	3.4.21.62	D32,H64,S221
6	1axwA	263	Thymidylate synthase	1.70	2.1.1.45	C146
7	1qumA	278	Endonuclease iv	1.55	3.1.21.2	E261
8	8tlnE	315	Thermolysin	1.60	3.4.24.27	E143,H231
9	1bs4A	167	Peptide deformylase	1.90	3.5.1.31	G45,Q50,L91 E133
10	1pud	371	Transglycosylase	1.85	2.4.2.29	D102
11	5enl	435	Glycerate hydrolase	2.20	4.2.1.11	E168,E211,H373 K345,K396
12	1ula	288	Purine nucleoside phosphorylase	2.75	2.4.2.1	H86,E89,N243
13	1eugA	224	Glycosylase	1.60	3.2.2.3	D64,H187
14	1d4aA	272	Quinone reductase	1.70	1.6.99.2	G149,Y155,H161
15	1chkB	237	Chitosanase	2.40	3.2.1.132	E22,D40
16	1chmA	400	Creatine amidinohydrolase	1.90	3.5.3.3	H232,E262,E358
17	1aru	335	Peroxidase	1.60	1.11.1.7	R52,H56,D57 G75,D77,S79 N93,H184,D246
18	1a8uA	276	Chloroperoxidase t	1.60	1.11.1	S98,D228,H257
19	1a2zA	219	Carboxyl peptidase	1.73	3.4.19.3	E80,C143,H167
20	1lba	149	T7 lysozyme	2.20	3.5.1.28	Y46,K128
21	1mbb	339	Acetylenolpyruvylglucosamine	2.30	1.1.1.158	R159,S229,E325
22	1idjA	358	Pectin lyase a	2.40	4.2.2.10	D154,R176,R236
23	1daaA	276	Aminotransferase	1.94	2.6.1.21	Y31,K145,E177

						L201
24	1azsA	189	Adenylyl cyclase	2.30	4.6.1.1	D396,I397,D440
25	4kbpA	423	Purple acid phosphatase	2.70	3.1.3.2	H202,H295,H296
26	1hfs	159	Stromelysin-1	1.70	3.4.24.17	E202,M219
27	1fmcA	254	7alpha-hydroxysteroid dehydrogenase	1.80	1.1.1.159	S146,Y159
28	1mpyA	306	Catechol 2,3-dioxygenase	2.80	1.13.11.2	H199,H246,Y255
29	2abk	210	Endonuclease iii	1.85	4.2.99.18	E112,K120,D138
						K191
30	1aq0A	305	Glucanase	2.00	3.2.1.73	E288,E232
31	1ako	267	Exonuclease iii	1.70	3.1.11.2	H259
32	1og1A	222	Adp ribosyltransferase 2	2.00	2.4.2.31	L6,T9,K66,N87 R91,R126,G127 T130,F132,Q144 S147, R184, Q187 E189
33	1apyB	140	Aspartylglucosaminidase	2.00	3.5.1.26	T183,T201,T234 G235
34	1jdw	359	L-arginine/glycine amidinotransferase	1.90	2.1.4.1	D254,H303,C407
35	1bjpA	61	Tautomerase	2.40	5.3.2.0	P1,F50,R39
36	7rsa	123	Ribonuclease a	1.26	3.1.27.5	H12,K41,V43, N44, T45,H119 F120,D121,S123
37	1xnb	184	Endo-1,4-beta-xylanase	1.50	3.2.1.8	Y5,Y69,E78 Y80,Y166,E172
38	1b73A	251	Glutamate racemase	2.30	5.1.1.3	D7,S8,C70 C178
39	1fua	205	L-fucose-1-phosphate aldolase	1.92	4.1.2.17	E73,H92,H94 H155
40	1eptB	84	Pyrrolidone carboxyl peptidase	1.80	3.4.21.4	D102
41	1adeA	430	Adenylosuccinate synthetase	2.00	6.3.4.4	D13,H41,K140 R147,Q224
42	2acy	97	Acylphosphatase	1.80	3.6.1.7	R23,N41
43	2plc	273	Phospholipase c	2.00	3.1.4.10	H45,D46,R84

						H93,D278
44	1tyfA	182	Clp peptidase	2.20	3.4.21.92	G68,S97,M98 H122,D171
45	1ppn	211	Papain	1.60	3.4.22.2	C25,H159,N175
46	1lcb	315	Thymidylate synthase	2.50	2.1.1.45	E60,R178,C198 S219,D221,D257 H259
47	1csn	292	Casein kinase-1	2.00	2.7.1.-	D131
48	1cqqA	179	Rhinovirus 3c protease	1.85	3.4.22.28	H40,E71,G145 C147
49	3ecaA	325	Asparaginase type ii	2.40	3.5.1.1	T12,Y25,T89 D90,K162
50	1ril	78	Ribonuclease h	2.80	3.1.26.4	D10,E48,D70
51	1cex	196	Cutinase	1.00	3.1.1	S120,D175,H188
52	1a4iA	294	Methylenethf dehydrogenase	1.50	1.5.1.5	K56
53	1dizA	281	3-methyladenine glycosylase ii	dna 2.50	3.2.2.21	Y222,D238,W272
54	1c2tA	208	Transformylase	2.10	2.1.2.2	D144
55	1gl3A	366	Dehydrogenase	2.50	1.2.1.11	C135,Q162,H274 R267
56	1a05A	356	Isopropylmalate dehydrogenase	2.00	1.1.1.85	Y140
57	1pgs	310	Endoglycosidase	1.80	3.5.1.52	D60,E206
58	1zio	216	Adenylate kinase	1.96	2.7.4.3	K13,R127,R160 D162,D163,R171
59	1chd	197	Cheb methylesterase	1.75	3.1.1.61	S164,T165,H190 M283, D286
60	2hgsA	471	Glutathione synthetase	2.10	6.3.2.3	R125,S159,G369 R450
61	1bzyA	213	Phosphoribosyltransferase	2.00	2.4.2.8	E133,D134,D137 K165,R169
62	1hel	128	Hen egg-white lysozyme	1.70	3.2.1.17	E35,D52
63	2phkA	276	Phosphorylase kinase	2.60	2.7.1.38	D149,K151
64	2cpo	297	Chloroperoxidase	2.16	1.11.1.10	N12,C29,N93 E104,H105,S108 E183,F186,N216

						T238,S239,S241 S242,S248,T250 S251,T252
65	2gb0A	386	Sarcosine oxidase	1.85	1.5.3.1	H45,R49,H269 C315
66	1e2aA	101	Enzyme iia	2.30	2.7.1.69	H78,Q80,D81 H82
67	1bxoA	322	Penicillopepsin	0.95	3.4.23.20	D33,D213
68	1ivhA	386	Isovaleryl-coa dehydrogenase	2.60	1.3.99.10	E254
69	1ra2	158	Dihydrofolate reductase	1.60	1.5.1.3	I5,M20,D27 L28, F31,L54 I94
70	1nzyA	268	Coenzyme a dehalogenase	1.80	3.8.1.6	F64,H90,G114 W137,D145
71	1al6	434	Citrate synthase	1.85	4.1.3.7	H274,H320,D375
72	12asA	326	Asparagine synthetase	2.20	6.3.1.1	D46,R100,Q116
73	1bg0	355	Arginine kinase	1.86	2.7.3.3	R126,E225,R229 R280,R309
74	1a2pA	107	Barnase	1.50	3.1.27	K27,R59,R87 H102
75	1pjbA	360	L-alanine dehydrogenase	2.10	1.4.1.1	K74,H95,E117 D269
76	1aj8A	370	Citrate synthase	1.90	4.1.3.7	H262,D312
77	1ctt	293	Cytidine deaminase	2.20	3.5.4.5	H102,E104,C129 C132
78	2aidA	98	Hiv protease	2.20	3.4.23.16	D25
79	1tph1	244	Triosephosphate isomerase	1.80	5.3.1.1	K13,H95,S96 E165
80	1afwB	392	3-ketoacetyl-coa thiolase	1.80	2.3.1.16	C125,H375,C403 G405
81	1b93A	147	Methylglyoxal synthase	1.90	4.2.99.11	H19,G66,D71 D91 H98,D101
82	1bwp	211	Acetylhydrolase	2.10	3.1.1.47	S47,G74,N104 D192,H195
83	1lxa	261	Acyltransferase	2.60	2.3.1.129	H125

84	1bn8A	398	Pectate lyase	1.80	4.2.2.2	D184,D223,D227 R279
85	1jejA	350	Dna beta-glucosyltransferase	2.50	2.4.1.27	E22,D100
86	2bmiB	229	Beta-lactamase	2.00	3.5.2.6	H82,H84,D86 H145,C164,H206
87	1mrj	246	Ribosome-inactivating protein	1.60	3.2.2.22	E160
88	1rpt	341	Prostatic acid phosphatase	3.00	3.1.3.2	R11,H12,R15 R79,H257,D258
89	1froA	175	Lactoylglutathione lyase	2.20	4.4.1.5	H126,E172
90	1eyiA	330	Fructose-1,6-bisphosphatase	2.32	3.1.3.11	D68,D74,E98
91	2ctc	306	Carboxypeptidase a	1.40	3.4.17.1	Y248,E270
92	2acs	313	Aldose reductase	1.76	1.1.1.21	Y48,K77,C298 H110
93	1dcia	274	Dienoyl-coa isomerase	1.50	5.3.1	D176,E196,D204
94	1nid	332	Nitrite reductase	2.20	1.7.99.3	H95,D98,H100 H135,C136,H145 M150,H255
95	3csma	255	Chorismate mutase	3.00	5.4.99.5	R16,R157,K168 E246
96	1kas	410	Beta-ketoacyl acp synthase ii	2.40	2.3.1.41	C163,H303,H340 F400
97	1bolA	221	Ribonuclease rh	2.00	3.1.27.1	H46,E105,H109
98	1dzcA	182	Dtdp-4-dehydrorhamnose 3,5-epimerase	2.17	5.1.3.13	D170,H63
99	1hfc	156	Fibroblast collagenase	1.56	3.4.24.7	E219
100	1l4dA	248	Plasminogen	2.90	3.4.21.7	H603,S608,D646
101	1cd5A	265	Phosphate deaminase	2.30	5.3.1.10	D72,D141,H143 E148
102	1dnpA	468	Dna photolyase	2.30	4.1.99.3	W306,W359,W382
103	1ahjA	197	Nitrile hydratase	2.65	4.2.1.84	C113,C115,S114
104	1alkA	448	Alkaline phosphatase	2.00	3.1.3.1	S102,R166
105	2lzm	163	T4 lysozyme	1.70	3.2.1.17	E11,D20,T26
106	1uae	417	Enolpyruvyl transferase	1.80	2.5.1.7	N23,C115,D305 R397

107	1pmi	439	Phosphomannose isomerase	1.70	5.3.1.8	Y287,R304
108	1a4sA	502	Betaine aldehyde dehydrogenase	2.10	1.2.1.8	N166,C297,E263
109	4pgaA	329	Glutaminase-asparaginase	1.70	3.5.1.38	T20,Y34,T100 D101 K173
110	1pfkA	319	Phosphofructokinase	2.40	2.7.1.11	G11,R72,T125 D127 R171
111	1b3rA	427	S-adenosylhomocysteine hydrolase	2.80	3.3.1.1	N109,D130,K185 D189,C194
112	1poa	117	Phospholipase a2	1.50	3.1.1.4	D23,C28,R30 G32,H47,N49 D93, N111

^a PDB と分子鎖の ID

^b 各酵素を構成する残基数

^c 分解能 [Å].

^d IUPAC に基づく酵素番号

2. 6 引用文献

1. Burley SK, Almo SC, Bonanno JB, Capel M, Chance MR, Gaasterland T, Lin D, Šali A, Studier WF, Swaminathan S. Structural genomics: beyond the human genome project. *Nature Genet* 1999;23:151-157.
2. Burley SK, Bonanno JB. Structural genomics of proteins from conserved biochemical pathways and processes. *Curr Opin Struct Biol* 2002;12:383-391.
3. Zhang C, Kim S-H. Overview of structural genomics: from structure to function. *Curr Opin Chem Biol* 2003;7:28-32.
4. Chandonia J-M, Brenner ES. The impact of structural genomics: Expectations and outcomes. *Science* 2006;311:347-351.
5. Dietmann S, Holm L. Identification of homology in protein structure classification. *Nature Struct Biol* 2001;8:953-957.
6. Todd AE, Orengo CA, Thornton JM. Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol* 2001;307:1113-1143.
7. Conte LL, Ailey B, Hubbard TJP, Brenner SE, Murzin AG, Chotia C. SCOP: a structural classification of protein database. *Nucl Acids Res* 2000;28:257-259.
8. Orengo CA, Miche AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH: a hierarchic classification of protein domain structures. *Structure* 1997;5:1093-1108.
9. Lankowski RA, Luscombe NM, Swindells MB, Thornton JM. Protein Clefts in molecular recognition and function. *Protein Sci* 1996;5:2438-2452.
10. Brady Jr. GP, Stouten PFW. Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des* 2000;14:383-401.
11. Yao H, Kristensen DM, Mihalek I, Sowa ME, Shaw C, Kimmel M, Kavraki L, Lichtarge O. An accurate, sensitive, and scalable method to identify functional sites in protein structures. *J Mol Biol* 2003;326:255-261.
12. Chakrabarti R, Klibanov AM, Friesner RA. Computational prediction of native protein ligand-binding and enzyme active site sequences. *Proc Natl Acad Sci USA* 2005;102: 10153-10158.
13. Wallace AC, Laskowski RA, Thornton JM. Derivation of 3D coordinate templates for searching structural databases: Application to the Ser-His-Asp catalytic triads of the serine proteinases and lipases. *Protein Sci* 1996;5:1001-1013.
14. Wallace AC, Borkakoti N, Thornton JM. TESS: A geometric hashing algorithm for deriving 3D coordinate templates for searching structural databases: Application to enzyme active sites. *Protein Sci* 1997;6:2308-2323.

15. Pazos F, Sternberg MJ. Automated prediction of protein function and detection of functional sites from structure. *Proc Natl Acad Sci USA* 2004;101:14754-14759.
16. Ondrechen MJ, Clifton JG, Ringe D. THEMATICS: A simple computational predictor of enzyme function from structure. *Proc Natl Acad Sci USA* 1999;98:12473-12478.
17. Ko J, Murga LF, André P, Yang H, Ondrechen MJ, Williams RJ, Agunwamba A, Budil DE. Statistical criteria for the identification of protein active sites using theoretical microscopic titration curves. *Proteins* 2005;59: 183-195.
18. Wei Y, Ko J, Murga LF, Ondrechen MJ. Selective prediction of interaction sites in protein structures with THEMATICS. *BMC Bioinformatics* 2007;8:119.
19. Laurie, ATR, Jackson RM. Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. *Bioinformatics* 2005;21:1908-1916.
20. Amitai G, Shemesh A, Sitbon E, Shklar M, Netanel D, Venger I, Pietrokovski S. Network analysis of protein structures identifies functional residues. *J Mol Biol.* 2004;344:1135–1146.
21. Elcock AH. Prediction of functionally important residues based solely on the computed energetics of protein structure. *J Mol Biol* 2001;312:885-896.
22. Dessailly BH, Lensink MF, Wodak SJ. Relating destabilizing regions to known functional sites in proteins. *BMC Bioinformatics* 2007;8:141.
23. Ota M, Kinoshita K, Nishikawa K. Prediction of catalytic residues in enzymes based on known tertiary structure, stability profile, and sequence conservation. *J Mol Biol* 2003;327: 1053-1064.
24. Kinoshita K, Nakamura H. Identification of protein biochemical functions by similarity search using the molecular surface database eF-site. *Protein Sci* 2003;12:1589-1595.
25. Bate P, Warwicker J. Enzyme/non-enzyme discrimination and prediction of enzyme active site location using charge-based methods. *J Mol Biol* 2004;340:263-276.
26. Greaves R, Warwicker J. Active site identification through geometry-based and sequence profile-based calculations: Burial of catalytic clefts. *J Mol Biol* 2005;349:547-557.
27. Ben-Shimon A, Eisenstein M. Looking at enzymes from the inside out: The proximity of catalytic residues to the molecular centroid can be used for detection of active sites and enzyme–ligand interfaces. *J Mol Biol* 2005;351:309-326.
28. Yang L-W, Bahar I. Coupling between catalytic site and collective dynamics: A requirement for mechanochemical activity of enzymes. *Structure* 2005;13:893-904.
29. Fukui, K., Yonezawa, T., Nagata, C. and Shing, H.. Molecular-orbital theory of

- orientation in aromatic, heteroaromatic, and other conjugated molecules. *J Chem Phys* 1954;20:722-725.
30. Ohno K, Kamioya N, Naoki A, Inoue Y, Sakurai M. Effects of hydration on the electronic structure of an enzyme: Implications for the catalytic function. *J Am Chem Soc* 2001;123:8161-8162.
 31. Ohno K, Wada M, Saito S, Inoue Y, Sakurai, M.. Quantum chemical study on the affinity maturation of 48G7 antibody. *J Mol Struct THEOCHEM* 2005;722:203-211.
 32. Stewart JJP. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *Int J Quantum Chem* 1996;58:133-146.
 33. Bartlett GJ, Porter CT, Borkakoti N, Thornton JM. Analysis of catalytic residues in enzyme active sites. *J Mol Biol* 2002;324:105-121.
 34. Porter CT, Bartlett GJ, Thornton JM. The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucl Acids Res* 2004;32:129-133.
 35. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: Using hydrogen atom constants in the choice of side-chain amide orientation. *J Mol Biol* 1999;285:1735-1747.
 36. Wada M, Sakurai M. A quantum chemical method for rapid optimization protein structures. *J Comput Chem* 2004;26:160-168.
 37. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. AM1: A new general purpose quantum mechanical molecular model. *J Am Chem Soc* 1985;107:3902-3909.
 38. Klamt A, Schüürmann G. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J Chem Soc Perkin Trans* 1993;2:799-805.
 39. Klamt A, Jonas V, Burger T, Lohrentz JCW. Refinement and parametrization of COSMO-RS. *J Phys Chem A* 1998;102:5074-5085.
 40. Stewart JJP. MOPAC2002. Fujitsu, Ltd., Tokyo, Japan(2002)
 41. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery Jr. JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC,

- Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA. Gaussian 03 Revision B 05. Wallingford CT: Gaussian Inc; 2004.
42. Pople JA, Beveridge DL. Approximate Molecular Orbital Theory. New York: McGraw-Hill; 1970. Chapters 3 and 4.
 43. Wilson KP, Malcolm BA, Matthews BW. Structural and thermodynamic analysis of compensating mutations within the core of chicken egg white lysozyme. *J Biol Chem* 1992;267:10842-10849.
 44. Tsui V, Case DA. Theory and application of the generalized solvation model in macromolecular simulations. *Nucl Acid Sci* 2001;56:275-291.
 45. Harvey SC, Tan RK, Cheatham TE. The flying ice cube: Velocity scaling in molecular dynamics leads to violation of energy equipartition. *J Comput Chem* 1998;19:726-740.
 46. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. The amber biomolecular simulation program. *J Comput Chem* 2005;26:1668-1688.
 47. Warshel A, Levitt MJ. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol* 1976;103:227-249.
 48. Oliva G, Fontes MR, Garratt RC, Altamirano MM, Calcagno ML, Horjales E. Structure and catalytic mechanism of glucosamine 6-phosphate deaminase from *Escherichia coli* at 2.1 resolution. *Structure* 1995;3:1323-1332.
 49. Horjales E, Altamirano MM, Calcagno ML, Garratt RC, Oliva G. The allosteric transition of glucosamine-6-phosphate deaminase: the structure of the state at 2.3 resolution. *Structure Fold Des* 1999;7:527-37.
 50. Modis Y, Filppula SA, Novikov DK, Norledge B, Hiltunen JK, Wierenga RK. The crystal structure of dienoyl-CoA isomerase at 1.5 Å resolution reveals the importance of aspartate and glutamate sidechains for catalysis. *Structure* 1998;6:957-70.
 51. Singleton MR, Isupov MN, Littlechild JA. X-ray structure of pyrrolidone carboxyl peptidase from the hyperthermophilic archaeon *Thermococcus litoralis*. *Structure* 1999;7: 237-244.

第3章 超分子計算による水和状態タンパク質の電子波動関数の解析

3. 1 緒言

多くの生体分子が機能を発現するためには水の存在は不可欠である。多くのタンパク質分子は水溶液中に存在し、その機能を発現するためには周囲の水分子の存在が欠かせない。タンパク質の機能に固有の立体構造を安定化するためにも水分子の存在は不可欠であり¹、酵素の触媒機能に代表されるタンパク質の化学的な機能は水分子によって媒介されている。タンパク質分子と水分子との相互作用は水分子が持つ特徴的な水素結合の様式に由来している²。そのため、タンパク質の機能に関する水の重要性を明らかにするためには、その関係性を原子レベルで議論する必要がある。

近年のコンピュータを用いた研究は、タンパク質の機能に関与する水の様々な挙動を追跡することを可能にし、水分子に関する多くの知見をもたらした。その代表的な研究例は分子力場を適用した分子動力学計算によるシミュレーションである。分子力場ではクーロンポテンシャルと Lennard-Jones 型のポテンシャルに水素結合などの相互作用を直接考慮することが可能である。分子動力学を適用した多くの研究例によって、タンパク質の構造安定性やダイナミクスにとって、水分子の存在が重要な役割を持つことが知られるようになり³⁻⁵、水分子とタンパク質分子に関する様々な知見を得ることが可能になった。

しかし、タンパク質と水分子間の相互作用は、これらの古典力学では考慮できない様々な相互作用を含んでいる。タンパク質分子の内部には強い電場勾配が存在した不均一系であるため、電子の分極や移動などの量子的な過程を考慮することが必要になる⁶。定量的な視点から有機分子に関する例を挙げる。例えば、分子内相互作用の 10%程度は電子分極に関係する項だと考えられているが⁷、水分子が関与する場合にはその割合が 20%近くにまで増大することが知られている⁸。古典力場の静電相互作用項に電子分極の項を経験的に導入することによって、密度や拡散係数⁹、アミノ酸残基の水和自由エネルギー¹⁰が改良されることも報告されている。タンパク質と水分子との間に物理化学的に信頼性の高い知見を見出すためには、タンパク質・水分子の界面に起こる電子状態の考慮は決して欠かすことができない。

この問題点を克服するためには、通常の古典力場に基づいた方法に加えて分極力場¹¹を適用する工夫が報告されている。しかし、経験的に分極を考慮する方法を用いても、電子移動が考慮されていないため必ずしも十分とは言えない。タンパク質と水分子間の量子的な効果を取り入れる他の手法として、QM/MM 法を適用した研究がある¹²。この手法は局所

的に電子状態を考慮することが可能であり、活性部位など特定部位で起こる化学反応を再現するためには非常に有用な方法である。しかし、この手法では、水溶性タンパク質の水和状態に対し、分子全体の量子状態が溶媒からどのような影響を受けているかを評価することができない。それを可能とするアプローチの一つは、タンパク質と水殻から形成される超分子に対して量子化学計算を実行することである。

与えられたタンパク質の分子構造に対して分子軌道を得るためには非常に多くの計算コストを要する。タンパク質分子が水和した状態にある場合、分子構造の表面に配位した水の分子数が多数に及び、計算コストは膨大になる。そのため、溶媒の寄与を計算する際には、水分子を直接計算する代わりに PCM¹³ や COSMO¹⁴ に代表される連続体近似によって近似する仮想溶媒を適用するのが一般的である。連続体近似では水分子を誘電体と双極子を誘起する反応場として捉える近似法であり、バルクに存在する水分子による界面の電子分極を効率的に取り込むことが可能である。そのため、物性値の予測や計算コストの面では非常に有用なモデルであり、分子の双極子モーメントや振動数といった物性値を結果的に妥当な精度で計算することができる。その反面、水分子を明示的に扱わないために生じる問題点も存在する。タンパク質の分子表面と水分子に起こる電子移動や水分子の水素結合に関係した構造特異的な相互作用がその例である。電子移動相互作用は電子分極に基づく相互作用に匹敵する強さを持つことも報告されており、タンパク質分子の化学的な性質を議論する上で重要な物理量であると考えられる。

水和したタンパク質分子では溶質分子と水分子間に多体間の量子効果が存在しており、その効果は長距離相互作用を生み出していることが、理論的な立場^{15,16}と実験結果^{17,18}の両面から示唆されている。そのため、タンパク質の電子状態を正確に知るためには、タンパク質と水分子からなる超分子の電子状態を考慮する必要がある。タンパク質の分子構造が持つ電子状態に関して、配位している水分子の影響を分析することは、タンパク質の機能と水分子の関係をより明白にすることになるだろう。

第2章では水分子を仮想溶媒（連続体近似）によって考慮した。本章の目的は、タンパク質の電子波動関数に対して前章では考慮していなかった水分子の量子力学的寄与を明確にすることである。そのために、タンパク質と水和圏（タンパク質の分子表面を覆う水分子の殻）から構成された超分子の電子状態を解析する。その際、電子波動関数の性質と水和圏の大きさとの関係を明らかにし、第2章で得た電子波動関数に対する水和の効果がどの範囲の水分子による寄与なのかを明確にする。また、仮想溶媒計算と超分子計算の結果を比較することによって、前者の粗い近似の適用限界等を明確にする。以上より、水和がタンパク質の波動関数をどのように変調するのかを明らかにし、前章で提案したタンパク質の機能予測法の妥当性を検討する。

3. 2 方法

タンパク質の選出

触媒反応と機能部位が良く研究されている代表的な加水分解酵素を研究対象とし、構造は Protein Data Bank (PDB) ¹⁹ から取得した (Figure 1)。その際、構造の選出基準を以下のように設定した。

- ・ いずれも異なった構造 Class (データベース SCOP で用いられているタンパク質の分類) に属すること
- ・ サブユニットを持たず単一の構造ドメイン
- ・ 水素原子以外に原子の欠損が無い
- ・ 分解能が 2.0Å 以上の X 線結晶構造
- ・ 触媒部位の残基の位置が同定されている
- ・ 総残基数が 100 から 200 程度の構造

以上の条件を満たすタンパク質の構造から、分子の総電荷が中性及び正・負である計 4 構造を選択した。選出したタンパク質は、Phospholipase a2 (PDBID:1A3D²⁰)、Serine proteinase (PDBID:2SFA²¹)、Thioesterase (PDBID:1IVN²²)、Hen Egg White Lysozyme (PDBID:3LZT²³) である。取得したタンパク質に関する構造情報の詳細を Tabel I に示す。

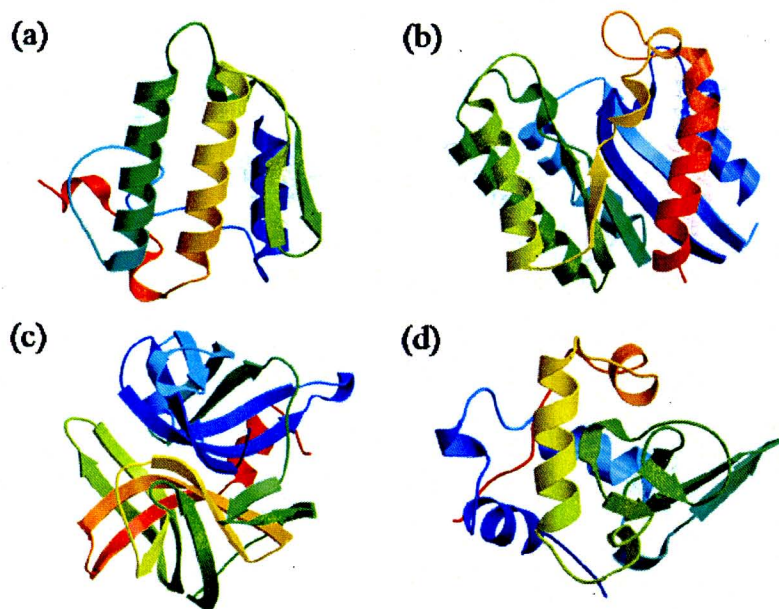


Figure 1. 選出したタンパク質の立体構造 (a) Phospholipase a2, (b) Serine proteinase, (c) Thioesterase, (d) Lysozyme

Table I. 研究対象に選出したタンパク質

Enzyme	Structure ^a	Num. Seq ^b	PDB	Resolution	Active Sites	Charge ^c
Phospholipase a2	all a	119	1A3D	1.80	G29 D47 D93	-3
Serine proteinase	all b	191	2SFA	1.60	H35 D65 G145 S147 S162	0
Thioesterase	a+b	178	1IVN	1.90	S10 G44 N73 D154 H157	-1
Lysozyme	a / b	129	3LZT	0.92	E35 D52	+8

a SCOP に基づく構造の Class

b 溶質タンパク質の構成残基の総数

c 溶質タンパク質の総電荷数

モデル構造の構築

取得した結晶構造に水素原子を付加した後、立方体のセル中に 1 分子のタンパク質を入れ、空隙に水分子を充填した。その際、系を中性化するために Na^+ と Cl^- を添加した。セルのサイズは溶質とセルとの最小距離が $8\text{\AA}$ になるように設定した (Figure 2(a))。収束判定を $0.1\text{ kcal/mol} \cdot \text{\AA}$ の条件で最急降下法によるエネルギー最小化計算を行った後、50000 ステップの分子動力学計算を実行した。0 ステップから 12500 ステップの間で昇温操作を行い、37500 ステップから 50000 ステップの間は $300\text{K} \cdot 1\text{ bar}$ の条件で平衡化を行った。分子力場にはタンパク質に Amber99 力場²⁴、水分子に TIP3P²⁵ を割り当てた。静電相互作用と van der Waals 相互作用については、原子間距離 $8\text{\AA}$ 以内の場合は直接計算し、長距離相互作用は Particle Mesh Ewald 法²⁶を適用して計算した。水素の伸縮運動は Shake 法²⁷によって拘束した。温度と圧力は Berendsen による Weak-Coupling 法²⁸を用いて制御した。分子動力学計算によって得られた最終構造に対してエネルギー最小化計算を行った後、タンパク質の重心を単位セルの中心に移動した。得られた構造から、以下の 2 種類のモデル (計 5 つ) のモデルを作成した (Figure 2(b))。

モデル A : タンパク質のみ

モデル B : タンパク質と第 N 層の水和殻 ($N=1,2,3,4$)

その際、第 1 水和殻はタンパク質と水素結合している水分子群、第 N 水和殻 (N が 2 以上) は内側の水和殻と水素結合している水分子群とした。水素結合の有無を判定する際に用いた基準は次の通りである。すなわち、アクセプター原子とドナー原子との距離が $3.5\text{\AA}$ 以下、

水素を挟んだ角度が 120° 以上の構造とした (Figure 2(c))。

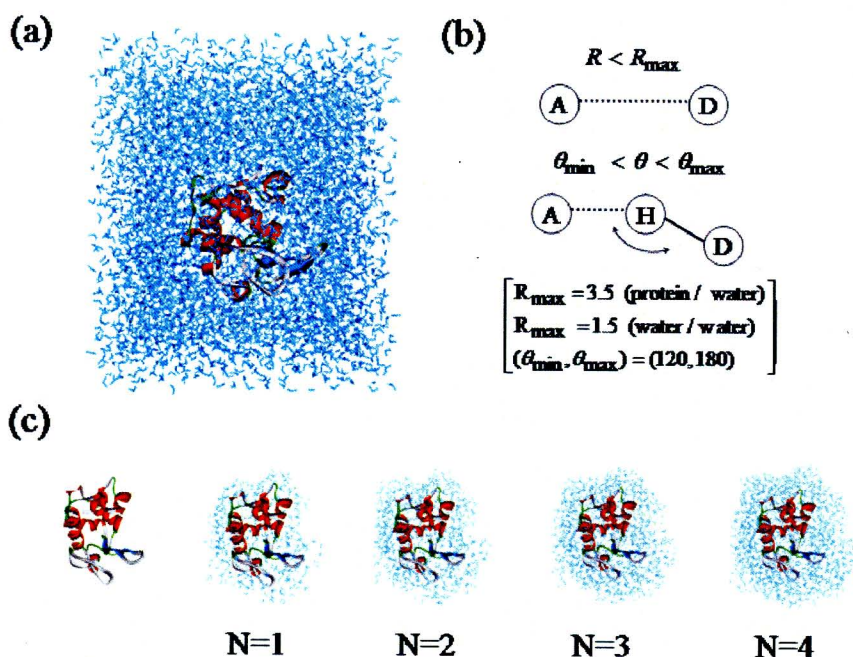


Figure 2. モデル作成法. (a)分子動力学計算によって得られた最終構造、(b)モデル作成時に適用した水素結合の判定方法、(c)水和したタンパク質の構造。N は水和圏の層数を表している。

分子軌道(MO)計算

得られたモデル構造から Na^+ と Cl^- イオンを除去した後、MOZYME 法²⁹ (AM1 ハミルトニアン³⁰) を用いて一点構造の SCF 計算を実行した。電子移動による相互作用 (CT) については、半経験的分子軌道法のハミルトニアンの中では AM1 が良好な結果を与えることが知られている³¹。溶媒効果の計算は、COSMO 法¹⁴ を用いて行い、その際比誘電率を 78.4、タンパク質の各原子には Klamt らによって提案された原子半径³³ を設定した。各 MOZYME の計算によって得られた一電子波動関数はユニタリー変換によって正準軌道に変換した。全ての MO 計算は MOPAC20002 パッケージ³⁴ を使って実行した。

溶媒和エネルギーの解析

Figure 2 で示した超分子系の水和エネルギー ΔE_{solv} は以下の式に基づいて算出した。

$$\Delta E_{solv} = E_{protein-wat} - (E_{protein} + E_{wat}) \quad (1)$$

$$\Delta E_{solv} = E_{protein-wat} - (E_{protein} + E_{wat}) \quad (2)$$

$E_{protein-wat}$ はタンパク質と水分子の複合体 (以下超分子と呼ぶ) の全エネルギーを表し、 $E_{protein}$ と E_{wat} はそれぞれタンパク質と水分子の全エネルギーを表す。水分子は水和圏全体を指す。

Figure 2 の超分子の周囲に仮想溶媒を置いた場合の溶媒和エネルギー $\Delta E_{solv}(imp)$ は Figure 3 の E_2 と E_5 の和に対応する。しかし、 E_5 を直接求めることはできないため、以下の方法で算出した。仮想溶媒の存在下では以下の式が成り立つ (Figure 3)。

$$E_2 + E_3 + E_5 = E_1 + E_4 \quad (3)$$

ゆえに、

$$\Delta E_{solv}(imp) = E_2 + E_5 = E_1 + E_4 - E_3 \quad (4)$$

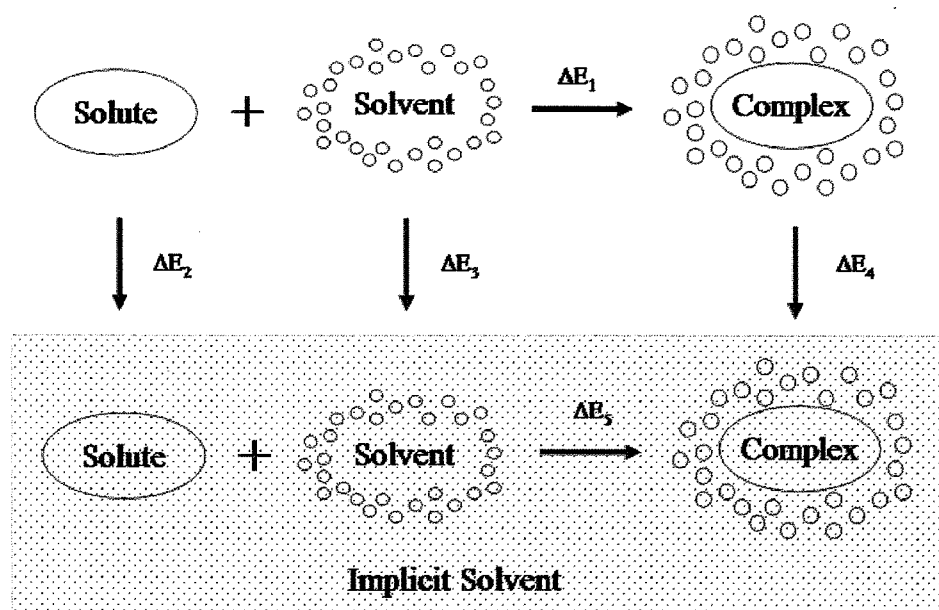


Figure 3. タンパク質の水和に関する熱力学サイクル。“Implicit Solvent”の領域 (灰色) は仮想溶媒が存在している状態を示す。 E_2 と E_3 と E_4 は、それぞれタンパク質と水分子、タンパク質と水の複合体の仮想溶媒中への移行に伴う全エネルギーの変化量を示す。

波動関数の解析

以下では、MOZYME 計算から得られた正準軌道を波動関数もしくは単に MO と呼ぶこと

にする。得られた MO は以下の式に基づいて解析した。

$$\phi_i = \sum_m^n c_{i,m} \chi_m \quad (5)$$

$$\sum_m^n c_{i,m}^2 = 1 \quad (6)$$

$$D_i(a) \equiv \sum_{m \in a} c_{i,m}^2 \quad (7)$$

$$D_i = \frac{1}{N} \frac{1}{\sum_a^N D_i^2(a)} \quad (8)$$

$$P_i = \left(\sum_{l \in S, W} C_{l,i} \right)^2 - \left\{ \left(\sum_{m \in S} C_{m,i} \right)^2 + \left(\sum_{n \in W} C_{n,i} \right)^2 \right\} \quad (9)$$

$$CP_a = \sum_{i=1}^{N_{occ}} e_i \times \left[\left(\sum_{l \in a, W} C_l \right)^2 - \left\{ \left(\sum_{m \in a} C_m \right)^2 + \left(\sum_{n \in W} C_n \right)^2 \right\} \right] \quad (10)$$

(5)式は MO ϕ_i が n 個の原子軌道の線形結合から成っていることを示し、(6)式は ZDO 近似³⁵による ϕ_i の規格化条件である。(7)で定義したパラメータ $D_i(a)$ は準位 i の MO に関する原子 a 上の電子密度である。ある残基の全原子に関して $D_i(a)$ の和をとることによって、その残基上の準位 i の MO に関する電子密度が評価できる。本研究では、 ϕ_i の電子密度が特定の残基で 0.2 以上の値であれば、その軌道は残基に局在化していると判定した。(8)式の D_i は軌道 i の非局在化の程度を定量化し³⁶、 N はあるタンパク質の原子数である。電子が複数の全原子に非局在化している場合を考えるなら D_i は 1 になる。逆に単一の原子に局在化しているなら D_i は $1/N$ となる。

(9)式の P_i は軌道 i に関するタンパク質分子と任意の水分子間の電子の交換を定量化したものであり、 S と W はそれぞれタンパク質分子と水分子の基底空間である。タンパク質分子-水分子間の電子の交換が無視できる場合、 P_i は 0 になる。 P_i の符号は、 i 番目の MO の S 及び W 上での波動（振幅）の正負に依存して変わる。 P_i が正の値ならば、 S と W 上でいずれも振幅が正か負であることを示し、両者は結合性であることに対応する。負の値ならば S と W の反結合性を示す。(10)式の CP_a は、アミノ酸残基 a と水分子間に対して算出された P_i の値を軌道エネルギーの重み付けをして加算したエネルギー指標値である。(10)式右辺の $[\]$ 内は、 P 値の定義式を参照すると、残基 a と水の間のみに電子の交換の程度を評価することに相当する。したがって、残基 a と水が、高い軌道エネルギー（負値）をもつ MO 上

で、両者が結合性の相互作用をするときは CP_a 値に絶対値の大きい負の寄与をする。それに対して、これらが反結合性の相互作用をするときは CP_a 値に正の寄与をすることになる。それゆえ、被占軌道全体についての和である CP_a が負であれば、全体として残基 a と水は結合性の相互作用することを意味する。

3. 3 結果

超分子の溶媒和エネルギーと電子移動

Table II は方法の節で述べた手順に従って得た水和したタンパク質（超分子）に関する系の規模を示している。超分子に関して、タンパク質 1 分子当りに配位した水分子数の数は第 1 水和圏（水層の厚みは 3.5Å）で 450 から 650 分子程度、第 4 水和圏（水層の厚みは 8Å）で 1500-2000 分子の範囲にあった。配位した水分子の数はタンパク質の表面積にほぼ比例しており、タンパク質分子に水素結合した水分子がタンパク質の表面から層状に構造化していることを確認した。

Table II. 水和タンパク質のモデルに含まれる水分子

Enzyme	SAS ^a	Volume ^b	The number of water molecules ^c			
			N = 1	N = 2	N = 3	N = 4
Phospholipase a2	6900	15243	454	775	1159	1548
Serine proteinase	7949	22418	508	883	1301	1733
Thioesterase	9259	25545	647	1078	1540	2064
Lysozyme	7111	17008	502	832	1214	1587

a 溶質分子の表面積 (Å²)

b 溶質分子の総容積 (Å³)

c 水和殻に含まれる水の分子数 (N=1,2,3,4)

Figure 4(a)に水和に伴う ΔE_{solv} の値(式(1))を棒グラフで示した。以下、 ΔE_{solv} の値を安定化エネルギーと呼ぶことにする。棒グラフの長さは安定化エネルギーの大きさを表し、黒色が濃い順に第 1 水和圏から第 4 水和圏の内訳を示している。選出した全てのタンパク質について、安定化エネルギーの大きさを水和圏の大きさ（層数）ごとに分析すると、タンパク質分子から距離が近い水和圏であるほど安定化エネルギーへの寄与が大きい。第 1 水和圏の寄与が最も大きく、第 4 水和圏まで考慮した場合を基準にすると、Phospho Lipase a2 では 65%、Serine Protease では 56%、Thioesterase では 62%、Lysozyme では 67%の寄与がある。この値はタンパク質分子の種類によってばらつきがあるが、第 2 水和圏までを考慮

した場合には、Phospho Lipase a2 では 85%、Serine Protease では 85%、Thioesterase では 84%、Lysozyme では 85%の寄与を示し、タンパク質分子の構造によらず均一の値となる。Figure 4(c)は水和に伴ってタンパク質分子から水分子へ移動した電荷量 (e) を表している。タンパク質分子から水和圏に移動する電荷 (電子) の量は水和圏が大きなほど増大するが、その大半は溶質から第 1 水和圏への移動である。Phospho Lipase a2 では 3.87e、Serine Protease では 3.73e、Thioesterase では 5.78 e、Lysozyme では 2.93e の電子が第 1 水和圏に移動する。第 4 水和圏まで考慮した場合を基準にすると、タンパク質から水和圏への電子移動の 95% 以上が第 1 水和圏の寄与である。

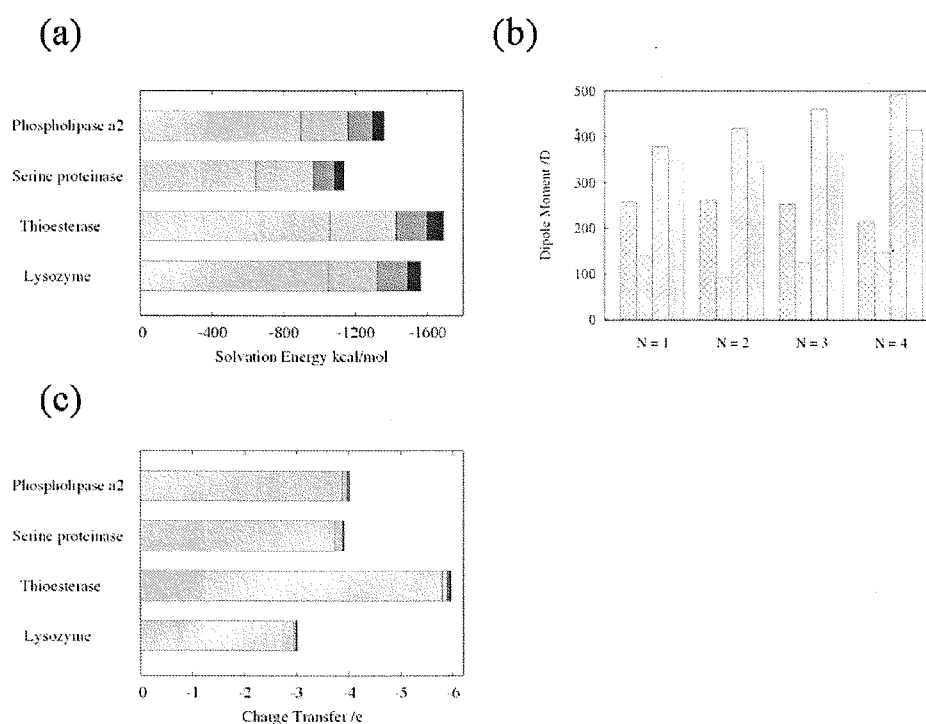


Figure 4. 様々な物理量の水和圏サイズ依存性。(a) 安定化エネルギー、(b)タンパク質から水和圏に移動した電荷量。色の濃くなる順に第 1, 2, 3 及び 4 水和圏のデータを表す。

Figure 5 はアミノ酸残基の相対露出表面積 (SAS (%)) とその残基から第 1 水和圏に移動した電荷量 (e) の関係を示している。ここで示した電荷移動量は、超分子中の着目する残基の総電荷から孤立タンパク質中でのその残基の総電荷を差し引いた値として評価している。厳密には、この差の中には、その残基から周囲のアミノ酸残基への電子移動も含まれているが、すべての残基について加算した値が Figure 4(c)とほぼ一致したので、ここでは近似的に水への電荷移動量とみなした。SAS は、着目する残基の露出表面積をその残基の総表面積で除した値である。Figure 5 より、SAS が大きな残基ほど電荷移動に大きく寄与するわけではない。特に大きく寄与する残基は SAS の広範囲の領域に散在している。これらの

残基を化学的な性質に従って分析すると、アニオン性の残基であった。他の多くの残基で電荷移動量は $0.1e$ 以下であるのに対し、これらアニオン性の残基では $0.2e$ 以上の電子移動に寄与する場合も多かった。アニオン性の残基の存在部位と電荷移動量の関係については興味ある問題であるが、今のところ解析が終了していない。

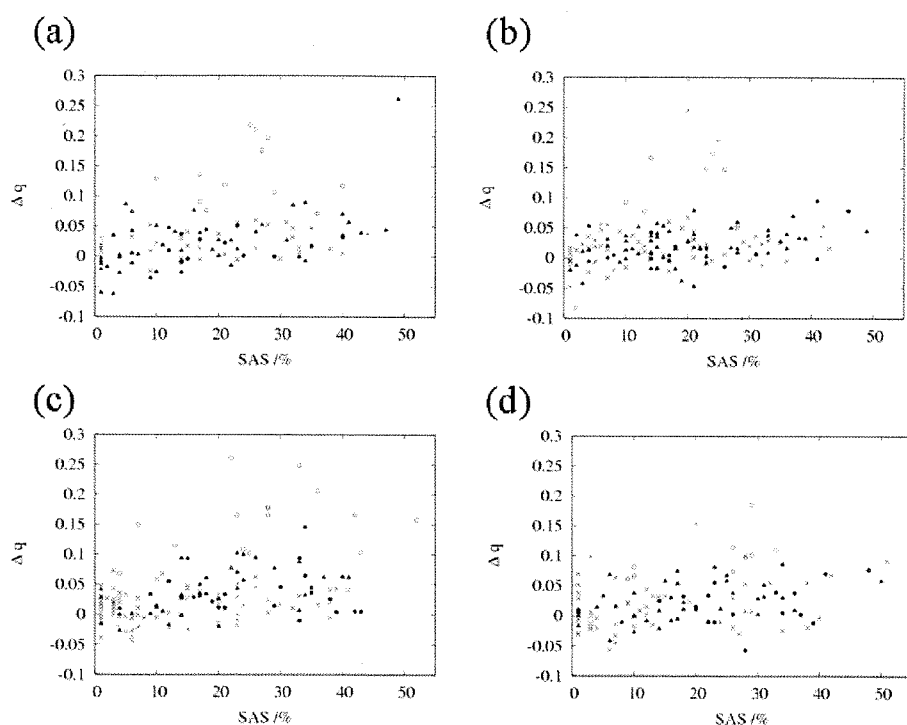


Figure 5. アミノ酸残基の相対露出表面積 (SAS %) とその残基から第1水和圏に移動した電荷量 Δq との関係。×は非極性残基、▲は非解離性の極性残基、○はアニオン性の残基、●はカチオン性の残基を表している。(a) Phospholipase a2, (b) Serine proteinase, (c) Thioesterase, (d) Lysozyme

超分子の電子状態

Figure 6 は MO のエネルギー分布を分析した結果である。孤立状態のタンパク質の軌道エネルギーを基準に考えると、超分子の被占軌道は低エネルギー側にシフトし、空軌道は高エネルギー側にシフトしている。その結果、超分子形成によって HOMO-LUMO ギャップは増大する。この性質は全てのタンパク質に関して確認でき、HOMO-LUMO ギャップは $4eV$ から $8eV$ まで増大した。

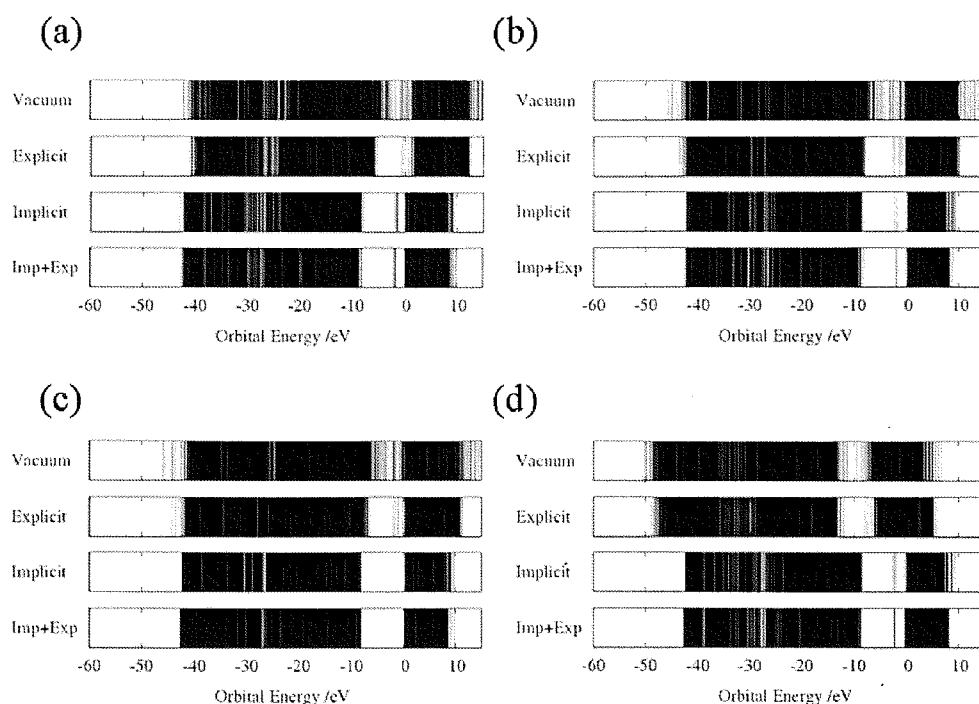


Figure 6. タンパク質の軌道エネルギー分布。“Vacuum”と“Implicit”は、それぞれ気相中と仮想溶媒中で計算した結果を表している。“Explicit”と“Imp+Exp”はタンパク質分子と第2水和圏からなる超分子に対して、それぞれ孤立系の結果と仮想溶媒を用いた結果を示している。(a) Phospholipase a2, (b) Serine proteinase, (c) Thioesterase, (d) Lysozyme

Figure 7 は Figure 6 に対応する分子軌道の状態密度を表している。孤立状態のタンパク質の状態密度 ((a)~(d)において Vacuum に対応する図) には HOMO-LUMO ギャップを境に2つの大きなピークが存在する。超分子の場合の状態密度は ((a)~(d)において Explicit に対応する図)、これらのピークに加え低エネルギー領域に別のピークを1つ有する。このピークの意味は後の節で議論する。

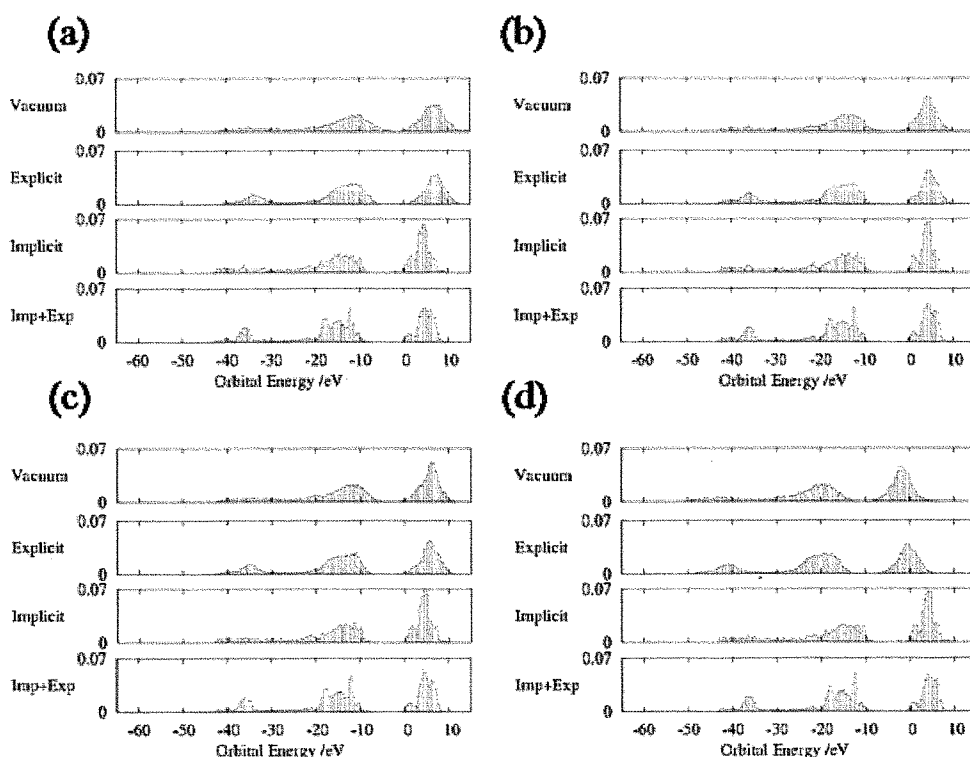


Figure 7. 各タンパク質の状態密度。状態密度は、エネルギー区間 $\Delta E=0.5$ eV ごとに状態数をカウントした値である。“Vacuum”と“Implicit”は、それぞれ気相中と仮想溶媒中で計算した結果を表している。“Explicit”と“Imp+Exp”はタンパク質分子と第2水和圏からなる超分子に対して、それぞれ孤立系の結果と仮想溶媒を用いた結果を示している。(a) Phospholipase a2, (b) Serine proteinase, (c) Thioesterase, (d) Lysozyme

Figure 8 はD値を軌道エネルギーに対してプロットした結果を示している。ここで(a)~(d)各図において、一番上の段（孤立状態のタンパク質）、2段目（超分子の中の水クラスターのみを取り出してMO計算した結果）及び3段目（超分子に対するデータ）に着目すると次の特徴が見出される。まず、孤立状態のタンパク質のD値は前章の結果と同様のピーク分布を示している。すなわち、前章の記法に従うと、グループB（タンパク質全体に非局在化したMO）、グループC（共有結合に形成するMO）、及び空軌道領域での大きなピークの存在という特徴づけができる。超分子中の水分子クラスターのD値もやはり3つの領域に分布している。これは水クラスターを一つの巨大分子と考えれば説明がつく。一方、超分子のD値の分布と形状は、孤立状態タンパク質の場合の分布と水クラスターのそれとの単純な重ねあわせではないように見える。このことは、(d)の lysozyme のデータから明らかである。すなわち、孤立状態タンパク質と水クラスターのグループB領域のピーク位置は若干ずれているが、超分子を形成するとこれらは融合して一つの大きなピークとなる。孤立状態の蛋白質と比べてピーク強度は4倍近くなる。しかし、グループC領域のピークは両者の分布を単純に重ね合わせたものに近い。本研究では、タンパク質の基底状態に興味が

あるので、空軌道領域については議論を省く。

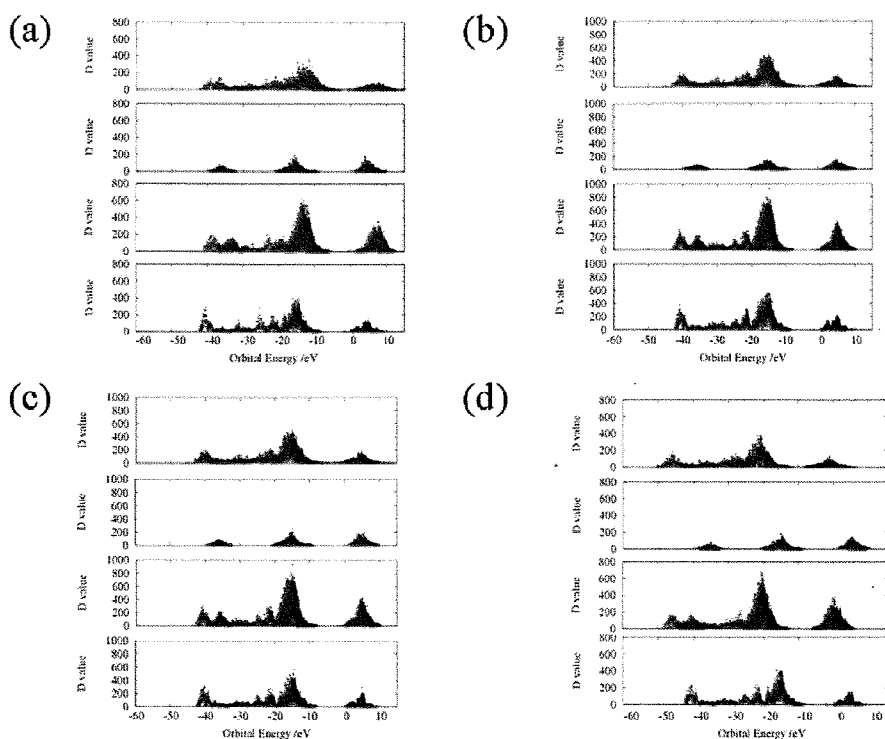


Figure 8. 孤立状態及び水和状態モデルに対するD値の解析結果。上から順番に、孤立状態のタンパク質、第2水和圏まで含む超分子中の水クラスターのみを取り出して計算した結果、その超分子全体についての結果、及びタンパク質のみを仮想溶媒に埋め込んだモデルに対する結果を表す。横軸は軌道エネルギー (eV)、縦軸は D 値を示している。(a) Phospholipase a2, (b) Serine proteinase, (c) Thioesterase, (d) Lysozyme

Figure 9 は、超分子が形成する各 MO 中の電子分布をタンパク質部位とそれを取り囲む水クラスターに分割して表示したものである。ただし、1 本 1 本のヒストグラムは、エネルギー範囲 0.5 eV 内にある複数の MO についての結果をまとめたものである。このエネルギー区間の設定は状態密度の算出のときと同一であるので、各図の一番外側の外形は、Figure 7 に示した状態密度の外形に一致している。Figure 9 の最も興味ある特徴は、グループ B 領域にある。すなわち、タンパク質中の電子数のピーク位置と水クラスター中の電子数のピーク位置が一致し、この領域内のいずれの軌道エネルギーの区間でも、タンパク質中の電子と水クラスター中の電子が共存していることである。これは、形成された MO がタンパク質あるいは水クラスターのどちらかに偏在しておらず、両領域に広がっていることを意味する。一方、グループ C の領域を見てみると、タンパク質中の電子数のピーク位置と水クラスター中の電子数のピーク位置が必ずしも一致していない。つまり、どちらか一方に電子が局在する軌道が多く残っているように見え、全体的に、グループ B 領域ほどにはタ

ンパク質部位と水クラスター部位の波動関数の融合が進んでいないと判断される。ただし、以下(Figure 10)で述べるように、グループ C においても若干の融合は起こっている。

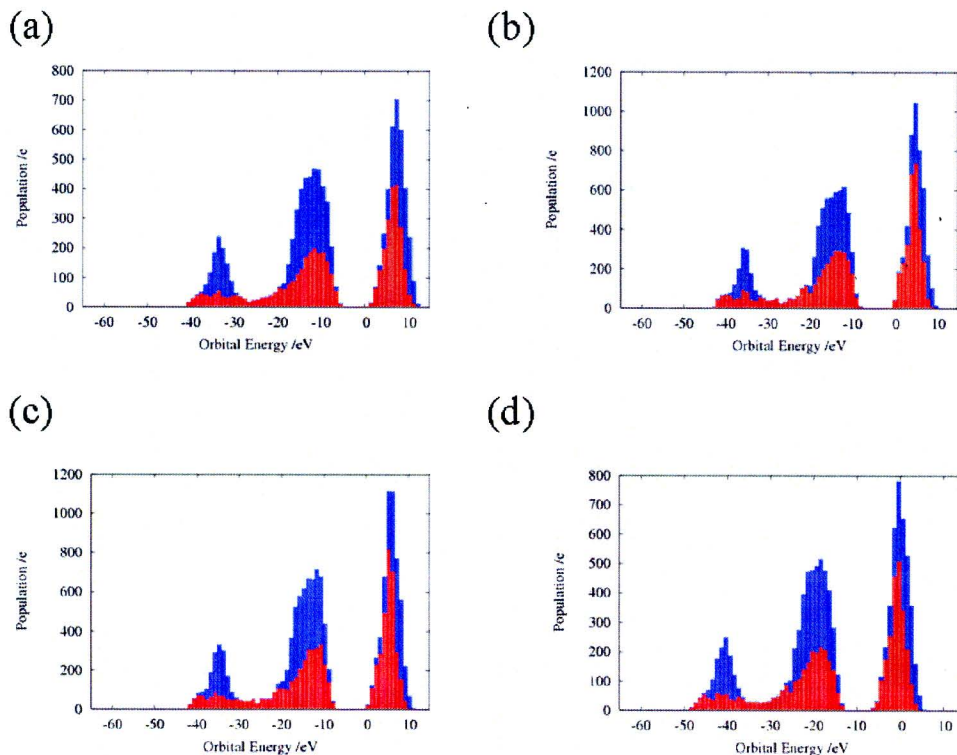


Figure 9. 超分子の各 MO 中の電子分布。タンパク質分子（赤）と水分子領域（青）それぞれに存在する電子数をバーの長さで表した。水分子は第 2 水和圏まで考慮している。(a) Phospholipase a2, (b) Serine proteinase, (c) Thioesterase, (d) Lysozyme

Figure 10 では超分子の場合の P 値を軌道エネルギーに対してプロットした。興味あることに、先に述べたグループ B 及び C いずれの領域においても、正・負の方向にそれぞれ 2 つのピークが存在する。これは正に、タンパク質部位の波動関数と水クラスターの波動関数が同位相あるいは逆位相で重なりを起こし、2 つの融合した MO が生じたことを表しており、Figure 8 と 9 で述べたことを支持するものである。(a)~(d)いずれにおいても、グループ B 領域の P 値とグループ C 領域のそれには大きさの違いがあり、後者の方が大きい。このことは、タンパク質部位と水クラスターの相互作用が、前者より後者の方が強いことを示している。これは Figure 8 と 9 における両領域ピーク強度の関係と逆であるが、矛盾するものではない。前章で示したように、グループ B 内の MO はアミノ酸側鎖間の弱い through-space 相互作用によって非局在化した軌道であるが、これが水クラスター領域まで広がるときにも同じくアミノ酸-水間の through-space 相互作用が効いてくると考えられる。一方、グループ C 領域では、共有結合もしくはそれに準ずる相互作用で軌道が広がってい

る。おそらく、超分子のグループ C 領域では、アミノ酸-水間の水素結合を通して、タンパク質と水が結合していると考えられる。水素結合は1種の **through-bond** 相互作用なので、グループ B 領域で起こっている **through-space** 相互作用よりは強く、先に述べた P 値の大小関係が生じる、と解釈される。

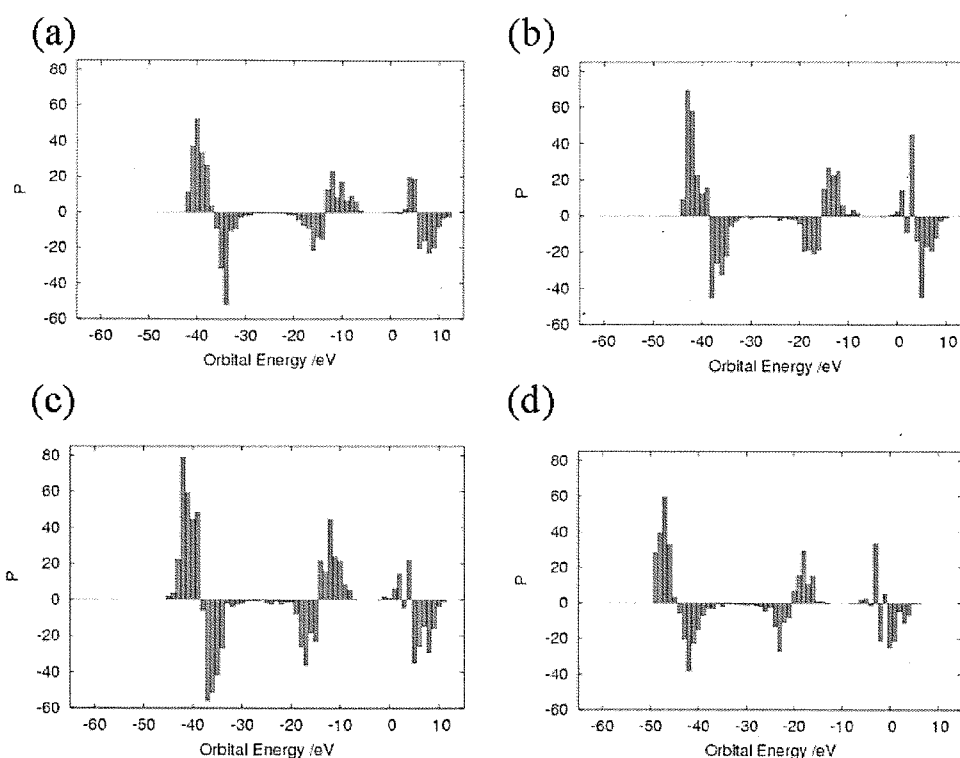


Figure 10. 第2水和圏まで含んだ超分子の軌道エネルギーに対する P 値の分布。(a) Phospholipase a2, (b) Serine proteinase, (c) Thioesterase, (d) Lysozyme

Figure 11 は超分子のMOを可視化したものである。上段にある左右の図では、それぞれグループ C とグループ B に対応するMOの中で代表的なものを選出して可視化した。グループ C のMOは超分子の比較的狭い空間領域に局在化している。上で予想されたように、その局在位置はタンパク質分子や水分子の共有結合とタンパク質-水クラスター界面に形成された水素結合に対応している。対照的に、グループ B の MO は超分子全体を覆う空間に分布している。一方、超分子の HOMO 及び LUMO (下段左右) は分子構造の限られた空間に局在化している。

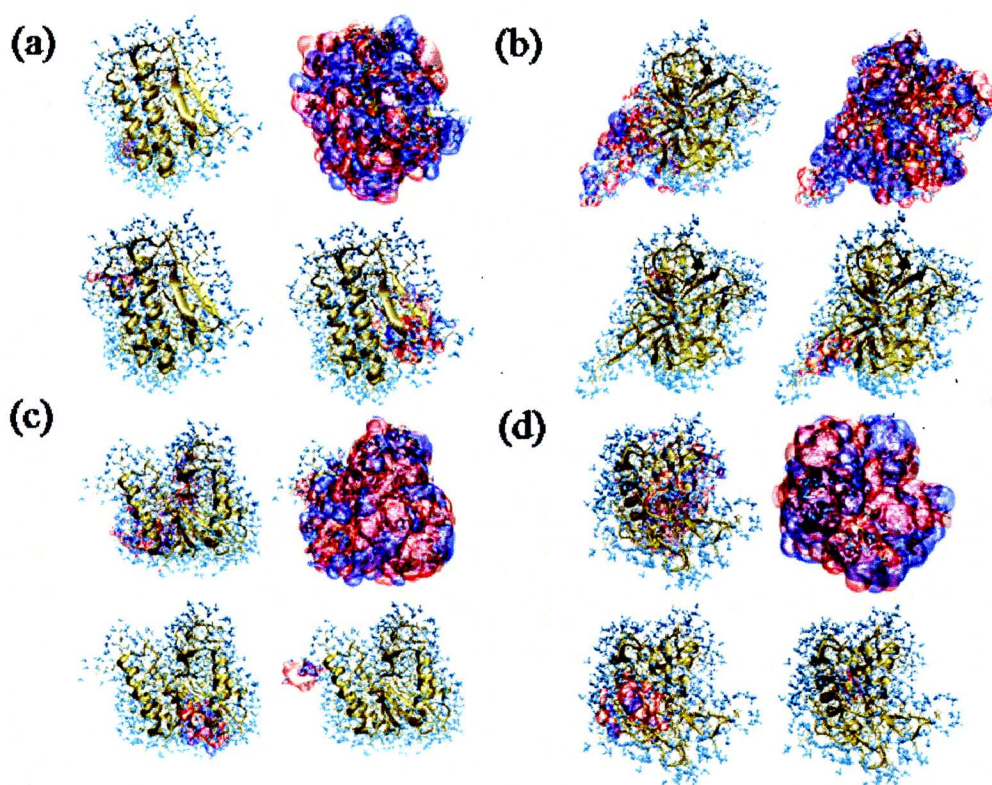


Figure 11. 超分子の軌道エネルギーに対する P 値の分布。水分子は第 2 水和圏まで考慮している。(a) Phospholipase a2, (b) Serine proteinase, (c) Thioesterase, (d) Lysozyme

Figure 12 は超分子中のタンパク質分子の残基が示す CP パラメータを示している。この値は残基によって特徴があり、CP パラメータの大きい（負値の小さい）残基は一次配列上の特定の領域に密集している。例えば、Figure 12(b)を見ると、Serine proteinase の 125 番目から 135 番目までの連続した残基の CP パラメータは-14 以上の値を示しており、この値は他の残基と比較して高い値である。CP パラメータが高い残基は、その部位でタンパク質と水分子が高エネルギーの電子を介して結合していることを示している。Figure 13 は超分子の CP パラメータをタンパク質の立体構造上にマッピングしたものである。Lysozyme は-19 以上、それ以外の酵素は-14 以上の CP パラメータを示す残基をマッピングした。CP パラメータの値が高い値を示す部位はタンパク質分子の一部に局所化しており、それはクレフト構造にある。この部位は、ここで調べた 4 つのタンパク質については実験的に同定されている触媒部位と一致している。

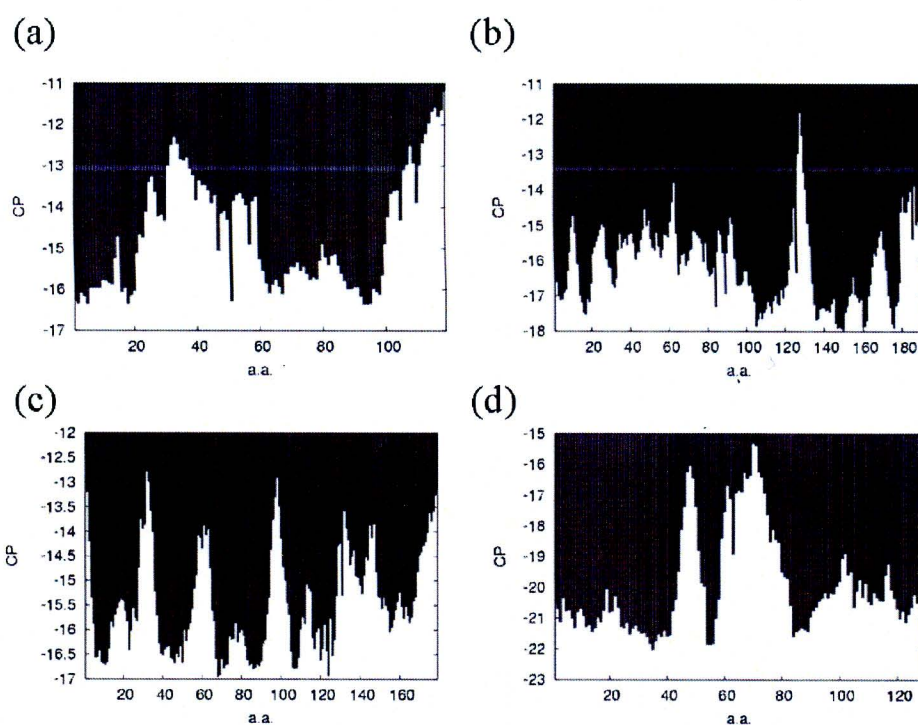


Figure 12. 超分子の CP パラメータを各残基に対してプロットした結果。(a) Phospholipase a2, (b) Serine proteinase, (c) Thioesterase, (d) Lysozyme

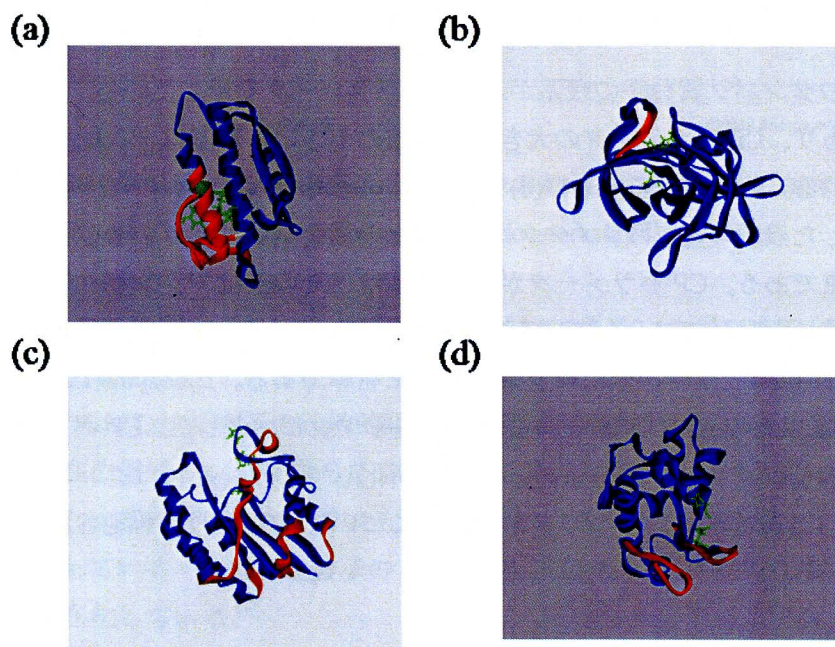


Figure 13. CP パラメータをタンパク質の立体構造上にマッピングした結果。赤色は CP パラメータが高い残基を表し、実験によって同定された触媒残基を緑で表している。(a) Phospholipase a2, (b) Serine proteinase, (c) Thioesterase, (d) Lysozyme

超分子計算と仮想溶媒計算の比較

Figure 14 は仮想溶媒を用いた場合の安定化エネルギー（式(4)の $\Delta E_{\text{solv}}(\text{imp})$ ）の分析結果である。タンパク質分子のみに仮想溶媒を適用すると（下から2番目の棒の長さ）、第2水和圏まで含む超分子（各図の一番下の棒）と同程度の安定化エネルギーが得られた。超分子に仮想溶媒を適用すると、安定化エネルギーはタンパク質分子のみに適用した場合に比べ大きく減少した。前者のエネルギー値を基準にした後者の値は、Phospho Lipase a2 の場合 67%-75%、Serine proteinase では 46-62%、Thioesterase では 61%-72%、Lysozyme では 87-93% 程度である。また、仮想溶媒を適用しない場合と比較して、水和圏のサイズが安定化エネルギーの値に与える影響は小さくなった。

Figure 14(b)は水-タンパク質間の電荷移動量を仮想溶媒を適用した場合について分析した結果を示している。縦軸が各タンパク質を表し横軸は電荷移動量を表す。各タンパク質の分析結果の最下部の棒グラフは最下部の棒グラフは孤立系の結果を示している。孤立系の結果は Figure 4 と同じデータを示している。その他の棒グラフは下から順に、仮想溶媒と第1水和圏から第四水和圏が共存した場合の結果を表す。これより、電荷移動の量は仮想溶媒の有無に関してあまり影響されない、ことがわかる。

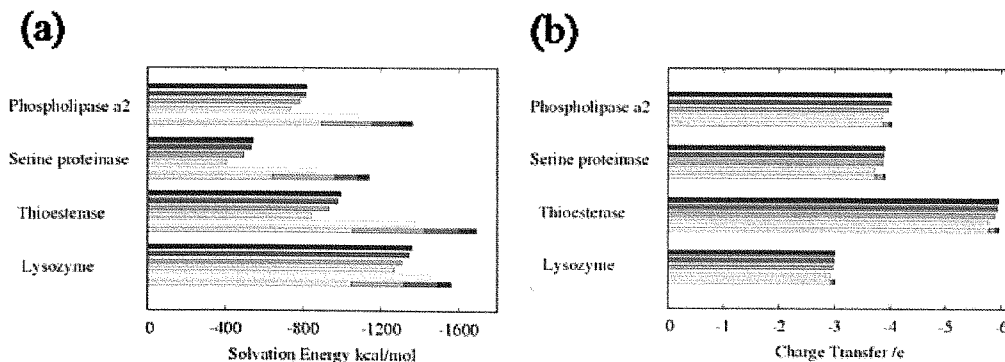


Figure 14. 仮想溶媒を用いた場合の水和に伴う各タンパク質の安定化エネルギーと電荷移動量。各タンパク質の分析結果において、最下部の棒グラフは孤立系の結果を示している。孤立系の結果は Figure 4 と同じデータを示しており、色の濃くなる順に第1, 2, 3及び4水和圏のデータを表す。第1水和圏から第4水和圏まで仮想溶媒を用いなかった場合。その他の棒グラフは下から順に仮想溶媒と第1水和圏から第4水和圏が共存した場合の結果。
(a) 安定化エネルギーの変化量 (kcal/mol) (b) タンパク質分子から水和圏に移動した電荷量 [e]

Figure 6 の MO のエネルギー分布を分析した結果をみると、仮想溶媒を用いた場合、軌道

エネルギーのシフトする方向はタンパク質分子の種類によって大きく異なった。Serine Protease と Thioesterase では、仮想溶媒中に移行するに伴い、被占軌道は低エネルギー側にシフトし、空軌道は高エネルギー側にシフトした。このエネルギーシフトの大きさは超分子系の場合と同程度であった。しかし、Phospho Lipase a2 では被占軌道と空軌道のエネルギーは両方とも低エネルギー側にシフトし、Lysozyme では被占軌道と空軌道のエネルギーは両方とも高エネルギー側にシフトした。

Figure 7 の状態密度をみると、仮想溶媒を用いた場合、気相中の超分子と比較して気相中の超分子と比較して空軌道が示す状態密度のピークがシャープになり、被占軌道が示す状態密度分布のピークの位置が不明瞭になる。これは、LUMO に接する領域の軌道が高エネルギー側にシフトし、HOMO にエネルギーが近い領域の軌道が低エネルギー側にシフトするからである。その結果、仮想溶媒を用いるとバンドギャップは増大する。仮想溶媒と超分子の両方が存在する場合、被占軌道領域で3つの極値が存在する。また、仮想溶媒のみが存在している場合と比較して、空軌道のピークがブロード化する。

3. 4 考察

水和タンパク質の電子状態

一般にタンパク質と水分子の間には電子移動が起こる。実際、Merz らによると、CspA では 1e 程度の電子がタンパク質から水和圏に移動すると報告している³⁷。本章 Figure 4 の結果では、彼らの結果を大きく上回る 3e から 5e の電子移動が確認できた。また、Table I の右端に示したタンパク質の総電荷と電荷の移動量は相関がないため、タンパク質－水間の電子移動は単純にタンパク質をマクロに見るだけでは説明できず、部位特異的に起こる現象であると推測される。Figure 5 の解析から、電子移動は水分子に露出したアニオン性の残基から起こっている。Figure 4 の結果から、タンパク質分子と水分子間に起こる電子移動は、主に第 1・第 2 水和圏に存在する水分子群の寄与による。

これらの水分子群は、タンパク質分子と互いに電子を共有し合っているという点でより外側の水分子とは明白に区別できる。Figure 10 の結果から P 値が正の値を持つ領域が 2 つ存在している。そのため、超分子中のタンパク質部位と水和圏の両方にまたがる MO は、軌道エネルギーによって 2 種類のグループに分類可能である。それらのうちで相対的に低いエネルギー領域に存在する MO に注目すると、Figure 8 の結果から、D 値最大ピークと比較すると中程度の値を示す。これらの MO は、第 2 章で示したグループ C 軌道（共有結合に沿って、比較的中程度の空間分布を持つ軌道グループ）の性質と類似している。また、Figure 11 から、水との共有結合領域もしくはタンパク質の界面付近に局所的に分布している。これらの MO はタンパク質－水分子間で比較的近距离で働く相互作用と関係しており、水素結合に代表される through-bond 相互作用に対応していると考えられる。一方、P 値が正を示す 2 つの領域のうち高エネルギー側に存在する MO は、D 値の最大ピークに対応している。そのため、これらの MO は超分子内の比較的広い空間に分布している。これらはアミノ酸残基あるいは水分子間の電子雲の重なりによって系全体に分布を持った through-space 相互作用に対応すると考えられる。これらの MO は超分子が占有する空間全体に広がっており、前章で述べたグループ B の MO に対応している。したがって、超分子全体、言い換えると水和状態の安定化に寄与していると考えられる。

これら 2 種の MO はいずれもタンパク質と溶媒の間が結合性か非結合性かで異なる 2 種の MO を伴っている。これらのうち結合性の MO は、水和状態の安定化に寄与していると考えられる。一方、非結合性の MO は、それぞれの分子あるいは分子クラスターを安定化することに寄与していると考えられる。

残基の示す CP の値に分布があることから、水分子と相互作用する電子のエネルギーは構造依存的な値を示す。特に CP 値が高い部位は HOMO とエネルギー準位が近接した MO 中の電子密度が高い部位に対応する。したがって、化学的活性の高い部位であることが予想

される。タンパク質の表面付近に存在する水分子は、重要なものとそうでないものが存在することが知られている。特に、表面に強く結合した水和水はタンパク質の機能に重要な役割を持つことが知られている。それに対して、結合水の外部にある水分子はバルク層として区別され、運動性が高いため構造依存的にタンパク質の機能に寄与することはない。これらの水分子は、主に溶質の構造安定性に寄与すると考えられている。今回我々が CP の解析によって明らかにした水分子は、タンパク質の表面に存在する特別な水分子である。その存在の重要性を電子レベルかつ定量的に特定することが可能になったことを強調しておきたい。

仮想溶媒と実溶媒の比較

孤立したタンパク質に対して仮想溶媒を用いた場合、第 1 水和圏までを含む超分子の場合と比較して、安定化エネルギーは大きな値を示した。しかし、第 1 水和圏を含んだ超分子モデルに仮想溶媒を適用すると、その安定化エネルギーは大きく低下した。孤立したタンパク質に対する仮想溶媒の効果は主に表面に存在する極性残基に対する静電遮蔽効果である。したがって、これらの結果は、仮想溶媒を直接タンパク質の周囲に置いた場合、この遮蔽効果が過大評価され、結果としてタンパク質の水和によるエネルギーの安定化が過大評価されてしまうことを示している。

一方、超分子に仮想溶媒を用いた場合、仮想溶媒はタンパク質の表面ではなく主に水和圏に影響する。したがって、仮想溶媒は溶媒間の相互作用に大きな影響を与えることが考えられる。水和圏が大きくなるほど安定化エネルギーは増大する傾向にあるが、仮想溶媒を適用した場合、その安定化エネルギーの値は飽和する。第 1 水和圏のモデルと第 4 水和圏のモデルを比較し、安定化エネルギーの差を分析すると、-467 eV から -637 eV までの差が生じている。それに対して、仮想溶媒を適用した場合には、平均値として -81 eV から -148 eV 程度の差しか生じていない。このように安定化エネルギー差が相対的に小さな値に抑えられていることから、仮想溶媒は溶媒間の（静電）相互作用を緩和する作用をしていること判断できる。仮想溶媒を適用すると、真空のときとタンパク質分子 HOMO-LUMO ギャップは増大し、被占軌道バンドと空軌道バンドはより大きく分離された状態になる。

3. 5 結論

本章では、タンパク質と水和圏からなる超分子の電子波動関数を解析することによって、タンパク質と水分子間に起こる量子力学的な過程を詳細に分析した。その結果、タンパク質と第一水和圏の水分子との間に顕著な電子移動が起こること、及び水和に伴うエネルギーの安定化は主に第一水和圏と第二水和圏から生じることを見出した。また、この超分子計算の結果と連続体近似計算の結果を比較し、タンパク質の電子波動関数、分子内部エネルギー及びタンパク質－水界面での相互作用に関して、両者の予測結果には顕著な違いがあることを見出した。特に量子力学的効果をあらわに考慮できる前者の計算から、タンパク質と水から成る超分子では系全体に広がる波動関数が形成され、これにより系全体が安定化されること、及び触媒部位では水分子との間で高エネルギー状態にある電子を介した特異的な電子移動相互作用が生じることが判明した。第2章では、タンパク質の構造安定性に関与するグループ B 軌道の存在を示したが、水和したタンパク質ではグループ B 軌道が水和圏を包括することによって超分子構造全体を安定化している。

3. 6 引用文献

52. Creighton TE, Freeman WH. *Proteins: Structure and Molecular Properties*. Freeman, New York 1993;35-38.
53. Jeffrey GA, Saenger W. *Hydrogen Bonding in Biological Structures*. Springer-Verlag Berlin and Heidelberg. 1991.
54. Levitt M, Sharon R. Accurate simulation of protein dynamics in solution. *Proc Natl Acad Sci. USA* 1988;85:7557-7561.
55. Kitao A, Hirata F, Go N. The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chem Phys*. 1991;158:447-472.
56. Kim B, Young T, Harder E, Friesner RA, Berne BJ. Structure and dynamics of the solvation of bovine pancreatic trypsin inhibitor in explicit water: a comparative study of the effects of solvent and protein polarizability. *J Phys Chem. B* 2005;109:16529-16538.
57. Brooks CL, Pettit BM, Karplus M. Structural and Energetic Effects of Truncating Long Ranged Interactions in Ionic and Polar Fluids. *J Chem Phys*. 1985;83:5897-5908.
58. Gao J, Xia X. A Priori Evaluation of Aqueous Polarization Effects Through Monte Carlo QM - MM Simulations. *Science*. 1992;258:631-635.
59. Kitaura K, Morokuma K. A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *Int J Quantum Chem*. 1976;10:325-340.
60. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. *J Phys Chem*. 1987;91:6289-6271.
61. Tannor DJ, Marten B, Murphy R, Friesner RA, Sitkoff D, Nicholls A, Ringnalda M, Goddard IWA, Honig B. *J Am Chem Soc*. 1994;116:11875-11882.
62. Patel S, Mackerell AD Jr, Brooks CL 3rd. CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J Comput Chem*. 2004;25:1504-1514.
63. J Gao, X Xia. A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science*. 1992;258;631-635.
64. Miertus S, Scrocco E, Tomasi J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilization of ab initio Molecular Potentials for the Prevision of Solvent Effects. *Chem Phys*. 1981;55:117-129.
65. Klamt A, Schüürmann G. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J Chem Soc Perkin Trans*. 1993;2:799-805.

66. Alexei AS. Long-distance electron tunneling in proteins. *Theor chem acc.* 2003;110:291-306.
67. Yoshioka Y, Kitagawa Y, Yamaguchi K, Nakamura T, Saito I. Experimental and Thoretical Studies on the Selectivity of GGG Triplets toward One-Electron Oxidation in B-Form DNA. *J Am Chem Soc.* 1999;121:8712-8719.
68. Hall DB, Holmlin RE, Barton JK. Oxidative DNA damage through long-range electron transfer. *Nature.* 1996;382:731-735.
69. Cowan JA, Upmacis RK, Beratan DN, Onuchic JN, Gray HB. Long-range electron transfer in myoglobin. *Ann N Y Acad Sci.* 1988;550:68-84.
19. Berman M, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourno PE. The protein data bank. *Nucleic Acids Res.* 2000;28:235-242.
20. Segelke BW, Nguyen D, Chee R, Xuong NH, Dennis EA. Structures of two novel crystal forms of *Naja naja naja* phospholipase A2 lacking Ca²⁺ reveal trimeric packing. *J Mol Biol.* 1998;279:223-232.
21. Kitadokoro K, Tsuzuki H, Nakamura E, Sato T, Teraoka H. Purification, characterization, primary structure, crystallization and preliminary crystallographic study of a serine proteinase from *Streptomyces fradiae* ATCC 14544. *Eur J Biochem.* 1994;220:55-61.
22. Lo YC, Lin SC, Shaw JF, Liaw YC. Crystal structure of *Escherichia coli* thioesterase I/protease I/lysophospholipase L1: consensus sequence blocks constitute the catalytic center of SGNH-hydrolases through a conserved hydrogen bond network. *J Mol Biol.* 2003;330:539-551.
23. Walsh MA, Schneider TR, Sieker LC, Dauter Z, Lamzin VS, Wilson KS. Refinement of triclinic hen egg-white lysozyme at atomic resolution. *Acta Crystallogr D Biol Crystallogr.* 1998;54:522-546.
24. Wang J, Cieplak P, Kollman PA, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem.* 2000;21:1049-1074.
25. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. "Comparison of simple potential functions for simulating liquid water". *J Chem Phys.* 1983;79:926-935.
26. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen L. A smooth particle mesh Ewald method. *J Chem Phys.* 1995;103:8577-8593.
27. Ryckaert JP, Ciccotti G, Berendsen HJC. "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes". *J Comput Phys.* 1977;23:327-341.
28. Harvey SC, Tan RK, Cheatham TE. The flying ice cube: Velocity scaling in molecular dynamics leads to violation of energy equipartition. *J Comput Chem.* 1998;19:726-740.
29. Stewart JJP. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *Int J Quantum Chem.* 1996;58:133-146.

30. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. AM1: A new general purpose quantum mechanical molecular model. *J Am Chem Soc.* 1985;107:3902-3909.
31. Vaart DVA, Badry DB, Charles L. Brooks, III, and Kenneth M. Merz, Jr. Are Many-Body Effects Important in Protein Folding? *J Phys Chem. B* 2000;104:9554-9563.
32. Vaart DVA, Merz MK. Charge transfer in biologically important molecules: comparison of high-level ab initio and semiempirical methods *Int J Quant Chem.* 2000;77:27-43.
33. Klamt A, Jonas V, Burger T, Lohrentz JCW. Refinement and parametrization of COSMO-RS. *J Phys Chem. A* 1998;102:5074-5085.
34. Stewart JJP. MOPAC2002. Fujitsu, Ltd., Tokyo, Japan(2002) MOPAC2002
35. Pople JA, Beveridge DL. Approximate Molecular Orbital Theory. New York: McGraw-Hill; 1970. Chapters 3 and 4.
36. Ohno K, Wada M, Saito S, Inoue Y, Sakurai, M. Quantum chemical study on the affinity maturation of 48G7 antibody. *J Mol Struct. THEOCHEM* 2005;722:203-211.
37. Nadig G, Van Zant LC, Merz KM, *J Am Chem Soc.* 1998;120:5593-5594.
38. Giancoli DC. Physics for Scientists and Engineers, with Modern Physics. Prentice-Hall Englewood Cliffs New Jousey 1995

第4章 タンパク質のフォールディング過程における電子波動関数－構造相関の解析

4. 1 緒言

タンパク質の機能は生物系の事実上全ての生物学的過程に関与している。多くのタンパク質はリボソーム上で数十から数百のポリアミドとして合成される。この鎖が機能するためには、個々のタンパク質に固有の天然三次元構造に折り畳まれることが必要である。一般的に、タンパク質の機能は3次元立体構造によって一義的に決まると考えられている。このフォールディング過程は素早く起こり、それは非常に特異的な過程である。新しく合成されたポリペプチド配列は、構造の自由度が無数に存在するにもかかわらず、正しいフォールディングへ到達する道筋を素早く見つけることができる (Anfinsen ドグマ)¹。このような構造空間の探索が如何にして起こるのかを発見することは、生命科学の最も大きな課題の一つである。フォールディングにおいて確立されている原理は、天然状態が一定の条件下で最小自由エネルギーに対応していることである²。重要な問題は、タンパク質がどのようにして妥当な時間内にその最小エネルギーを見つけることができるかということである (Levinthal Paradox)。そのため、タンパク質の折り畳み機構を厳密に解明するためには、分子のコンフォーメーションが天然構造に折り畳まれる際に要求される物理的な過程を理論的な枠組みから詳細に理解する必要がある。フォールディングの過程を物理的に説明するモデルとして、Pathway モデル³や Funnel 型モデル^{4,5}が提案されてきた。しかし、これらのモデルによって予想される過程と矛盾した例も多く、未だ確立されたモデルは存在していない。

多くの立体構造が解かれるに従って、タンパク質の構造形成に関する研究分野ではコンピュータを用いた手法が重要視されるようになった。代表的な例はタンパク質の分子構造をビーズ状の剛体と Go-Type ポテンシャルによって粗視化する手法 (Coarse-Grained Model)、古典力場を用いた手法 (分子動力学法 ; Molecular Dynamics) が知られている。これらの研究結果が蓄積されるにつれ、生体分子の構造形成やダイナミクスが急速に明らかにされるようになった。コンピュータの演算能力が飛躍的に増加するのに伴い、その時間スケールや粒子数 (原子数) は現実的なオーダーに近づいている^{6,7}。分子動力学法による構造形成のシミュレーションは、配列情報や構造比較等の経験的な情報を入れないことから、第一原理フォールディングシミュレーション (Ab Initio Molecular Dynamics) とも呼ばれる⁸。構造形成のシミュレーションに成功した例として、Turn^{9,10}や Helix¹¹⁻¹³、B Hairpin¹⁴等のペプチドに関する研究例が報告されている。近年では、Three Helix Bundle Protein¹⁵や

Trp-Cage¹⁶⁻¹⁹、BBA5²⁰ など、小規模のタンパク質分子のフォールディング研究も報告されるようになった。

しかしながら、分子動力学計算によるシミュレーションには理論的な問題点が存在し、明確な適用限界が存在する。それは、タンパク質のように多数の原子で構成されかつその間で複雑な相互作用ネットワークを形成している系では、そのポテンシャルエネルギー曲面は非常に複雑で、自由エネルギー最小の構造を求めることが困難な場合が多いことである。この問題を克服するためには、シミュレーションによる物理量計算の精度の向上が必要である。解決策の一つはシミュレーションの統計精度を改良することである。その方法論としてマルチカノニカル法²¹やレプリカ交換法^{22,23}、シミュレーテッド・テンパリング法^{24,25}に代表される拡張アンサンブル法などの洗練された手法が知られている。これらの手法は、非ボルツマン因子に基づく配置を発生させることでポテンシャルエネルギー空間上の一次元酔歩を実現し、シミュレーションがエネルギー極小状態に留まるのを避ける手法である。もう一つの課題は、ポテンシャルエネルギーの精度を改良することである。これは一つに、従来の古典力学をベースにしたポテンシャルエネルギー関数に電子的な効果を導入することに対応する²⁶。DNA 核酸対が起こすスタッキングや多くの有機分子の結晶充填は π - π 相互作用によって安定化されていることが知られている²⁷。タンパク質の内部には極性基の孤立電子対や共役系の π 電子に起因する電子的な相互作用が存在し、構造形成の過程にも重要な役割を担うと考えられる²⁸。共役系の π 電子だけでなく、近年ではCH- π ²⁹やOH- π として知られている弱い相互作用にも注目されている^{30,31}。タンパク質が構造形成の過程で探索する経路を説明するには、従来の古典的なモデルには含まれない相互作用を考慮することによって、より複雑な多体問題を解き明かす必要がある。

本章ではタンパク質分子の構造形成に伴う電子波動関数の性質を追跡する。具体的には、小タンパク質の構造形成を分子動力学法によってシミュレーションし、得られた構造変化の軌跡を対象に分子軌道の解析を行う。

4. 2 方法

研究対象のタンパク質

実験的に安定な構造を形成することが知られている2種類のペプチド (Amphiphilic Helix と Trp-Zipper) 及び小タンパク質 (Trp-Zipper) を研究対象とした (Figure 1)。これらの分子は水中で構造化することが知られており、側鎖によって分子がコンパクトに充填されているため、コンフォメーションの自由度が比較的制限されている。このような理由からモデルとして理想的な分子である。

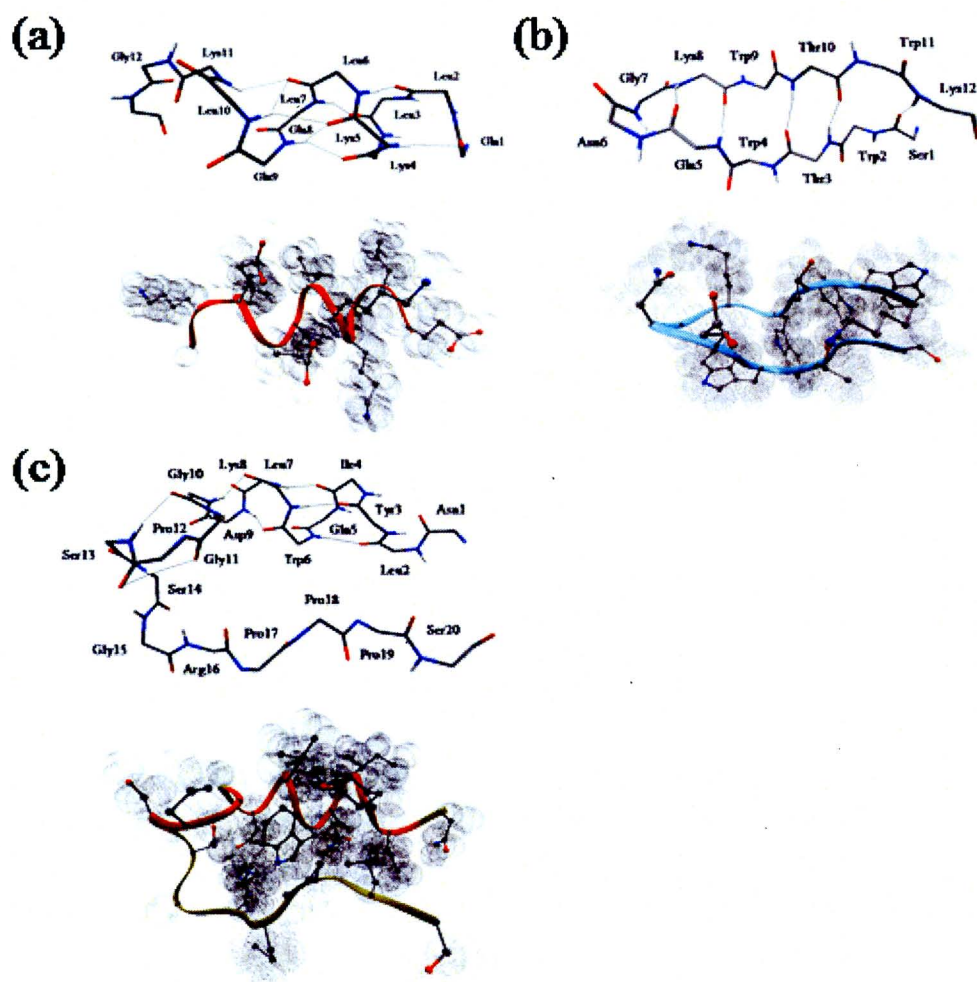


Figure 1. 研究対象の分子構造。上段は主鎖構造を Ball-Stick モデルで描画したグラフィックを表す。構成残基の名称を太字で表記し、水素結合は点線で表している。下段は主鎖構造を Ribbon 表示し、側鎖の排除空間を Dot モデルで表したグラフィックを表す。(a) Amphiphilic

Helix, (b) Trp-Zipper, (c) Trp-Cage

Amphiphilic Helix は 12 残基から構成された α -Helix モチーフをもつ (Protein Data Bank ID:1AL1)³²。この分子の特徴は、分子表面に露出した Leu から形成された疎水面と反対側に露出した Glu 及び Lys から形成された親水面により両新媒性を示すことである³³。この構造は、主鎖間に形成された水素結合の他、Leu によって充填された疎水核と Lys と Glu の側鎖間に形成された塩橋によって安定化されている³⁴。

Trp-Zipper は 12 残基から構成された β -hairpin モチーフをもつが (Protein Data Bank ID:1LE1)³⁵、これは向き合った 2 つの Trp ペアが疎水核を形成することで安定化され、298K の条件下では、Pro と Trp のコンフォメーションが異なる 3 種類の安定構造 (TZ1, TZ2, TZ3) が存在することも確認されている。この構造は構造形成に関する好適なモデルとして良く用いられてきた。その結果、Trp が会合した後に主鎖に沿って水素結合が形成される機構 (Zipping 仮説³⁶) が予想されてきたが、近年では疎水核の崩壊後に主鎖間の水素結合が再構成される機構 (Reptation 仮説^{37,38}) が有力な説である。しかし、両者の真偽は未だ定かではない。

Trp-Cage は 20 残基から構成された最小のタンパク質構造の一つと考えられている。ホルモン・ペプチド (Exendin-4) の C 末端領域が水中で構造化していることが NMR によって確認された (Protein Data Bank ID:1L2Y)^{39,40}。その結果によると、N 末端を基準として 3-6 の領域は α -Helix を、11-14 の領域は 3_{10} -Helix を、末端領域は Coil を形成する。二次構造は Trp9 と Gly11 の主鎖間に形成される水素結合と、Trp9 の側鎖と Arg16 の主鎖との間に形成される水素結合によって安定化される。Tyr3、Trp9、Pro12 及び Pro18 の側鎖によって形成される疎水核がフォールディングの駆動力であると考えられている³⁹。

分子動力学 (MD) シミュレーション

各小タンパク質に関して、それらと同一の一次配列を持つ直鎖状のポリアミドをモデリングし、1000 ステップの最急降下法によりエネルギー最小化計算を実行した。その後、Amphiphilic Helix と Trp-Zipper は 100 ns、Trp-Cage は 1000 ns の分子動力学計算を実行した。溶媒は仮想溶媒 (Generalized Born (GB) モデル⁴¹) を用いた。分子力場は Amber99 力場⁴² を割り当てた。静電相互作用と van der Waals 相互作用にはカットオフを用いず直接計算を行った。タイムステップは 2 fs とし、水素の伸縮運動は Shake 法⁴³ によって拘束した。温度と圧力は Berendsen による Week-Coupling 法⁴⁴ を用いて制御した。実験構造と比較するため、系の温度は実験条件と一致させた。Amphiphilic Helix は 298K、Trp-Zipper は 288K、Trp-Cage は 296 K の条件で温度制御を行った^{33,35,39}。折り畳み構造の再現には脱溶媒和のペナルティー項を導入することが効率的である⁴⁵。そのため、GB 法に Surface Area (SA) 法⁴⁶ を組み込んだ GBSA 法を適用した。その際、SA 法の表面張力は $0.005\text{kcal/mol}\text{\AA}^2$ に設定し

た⁴⁶。構造は 20 ps ごとにサンプリングした。

トラジェクトリーの解析

上のMDシミュレーションで得られた3次元構造を実験値と比較するため、NMR構造を基準にしたRMSD (Root Mean Square Deviation) を算出した。参照部位は全重原子とし、構造の重ね合わせはLSQ Fit (Least-Squares Fit) アルゴリズムを適用した。回転半径の算出には全原子を用いて行った。水素結合の有無を判定する際に用いた基準は、アクセプター原子とドナー原子との距離が3.5Å以下であり水素を挟んだ角度が120°以上の構造とした。

コンフォメーションの分類と構造のサンプリングを行うため、主成分解析 (PCA) を実行した。運動している原子*i*と原子*j*の共分散行列*C_{ij}*は以下の式になる：

$$C = (q_i - \langle q_i \rangle) \times (q_j - \langle q_j \rangle)$$

q は空間座標を表し、時間平均を $\langle \rangle$ と表記した。得られた行列はエルミート行列であるため、以下のようにスペクトル分解 (対角化) が可能である。

$$C_{ij} = \sum_{n=1}^N \lambda_n v_n$$

v_n は固有ベクトルを表し、 λ_n は v_n の固有値である。データ解析には、上位の λ_n に対応した2ベクトル (第一主成分：PCA1と第二主成分：PCA2) を用いた。主成分空間上に投影された全構造に対してRMSDによるクラスタリングを実行した^{47,48}。その際、構造類似性の判定基準となるRMSD (クラスター半径) は6Åとした。

自由エネルギー地形の解析は以下の方法によって実行した⁴⁹。

主成分空間上の微小区間(*i, j*)に構造を見出す確率を*P(i, j)*とすると、以下の式が成立する。

$$1 = \sum_i^N \sum_j^N P(i, j)$$

N は主成分軸の分割個数を表している。得られた確率分布から分配関数を一定 (=1) と見なすことで、*G_{ij}*を以下のように算出した。

$$P(i, j) = \frac{e^{-K_B T \cdot \Delta G_{ij}}}{Z} = \frac{e^{-K_B T \cdot \Delta G_{ij}}}{\sum_i^N \sum_j^N e^{-K_B T \cdot \Delta G_{ij}}}$$

$$G_{ij} \equiv -K_B T \cdot \text{Log}(P(i, j))$$

得られた *G_{ij}* を主成分空間上にマッピングすることによって、自由エネルギー地形の概形

を得た。

残基の配向の時間発展の解析には時間相関の算出を行った。分子動力学シミュレーションは時々刻々データ値を発生するため、着目する性質のある瞬間での値ともっと後の時刻 t での値との相関は容易に計算できる。物理量 P の時刻 t_0 及び時刻 t の間の自己相関関数は(規格化型) 以下のように書ける。

$$C(t) = \frac{1}{N} \sum_{i=1}^N \frac{\langle P_i(t) \times P_i(t_0) \rangle}{\langle P_i(t_0) \times P_i(t_0) \rangle}$$

本章の議論では、 N はアミノ酸残基 a の原子数を表し、 P はアミド結合の N-H に関する方向ベクトルである。 $\langle \rangle$ は時間平均を表す。シミュレーションはある有限時間だけしか実行できないため、“減衰の遅れ”による誤差が生じる。そのため、一般には時間相関関数をそれとフーリエ変換の関係にあるスペクトル密度へ変換する (Wiener-Khinchin の定理)。

$$C(\tau) = \int_{-\infty}^{+\infty} C(t) \cdot e^{-i2\pi\tau t} dt$$

τ は周波数領域における緩和時定数で、 i は虚数単位である。この変換には高速フーリエ法 (FFT) を用いた。このようにして算出した $C(\tau)$ はモデルに依存しない挙動を示し初期構造の依存性を考慮する必要が無い。

トラジェクトリーにはフォールディング軌道と (時間反転した) アンフォールディング軌道が混在している。特定の天然構造へのフォールディング軌道を抽出するため、接触パラメータを分析した。フォールディングの速度はタンパク質構造内部にある残基のトポロジー (充填率) と強く相関していることが知られている。Baker はフォールディングの度合いをコンタクトオーダー (RCO) というパラメータを用いて定量化したが、これは反応座標の決定に有力なパラメータである⁵⁰

$$RCO = \frac{1}{LN} \sum_N \Delta S_{ij}$$

N はコンタクトの総数、 L は残基数の総数である。 ΔS_{ij} はコンタクトしている残基 (i, j) 間の一次配列間の距離を与える関数であり、コンタクトの判定には一般的にコンタクトマップ (Average Distance Map, ADM) の成分が割り当てられる。ADM の構築方法は残基の全構成原子間の距離に関する平均値を採用し、閾値を 6Å とした⁵¹。

分子軌道計算

得られたサンプリング構造に関して半経験的分子軌道計算を実行した。計算には

MOZYME⁵²として知られている代表的な Linear-Scaling 分子軌道法、AM1 ハミルトニアン⁵³を用いた。MOZYME の計算によって得られた一電子波動関数は、ユニタリー変換によって正準軌道に変換した。計算は MOPAC20002 パッケージ⁵⁴を使って実行した。

波動関数の解析

得られた正準軌道（以下 MO）は以下の手順で解析した。

$$\phi_i = \sum_m^n c_{i,m} \chi_m \quad (1)$$

$$\sum_m^n c_{i,m}^2 = 1 \quad (2)$$

$$D_i(a) \equiv \sum_{m \in a} c_{i,m}^2 \quad (3)$$

$$D_i = \frac{1}{N} \frac{1}{\sum_a^N D_i^2(a)} \quad (4)$$

$$D = \sum_{i=1}^{occ} D_i \quad (5)$$

(1)式の準位 i の分子軌道 ϕ_i は n 個の原子軌道 χ_m ($m=1,2,\dots,n$) の線形結合によって記述され、(2)は ZDO 近似⁵⁵による ϕ_i の規格化条件である。(3)で定義したパラメータ $D_i(a)$ は準位 i の分子軌道に関する原子 a 上の電子密度である。ある残基の全原子に関して $D_i(a)$ の和をとることによって、その残基上の準位 i の分子軌道に関する電子密度が評価できる。本研究では、 ϕ_i の電子密度が特定の残基で 0.2 以上の値であれば、その軌道は残基に局在化していると判定した。(4)式の D_i は軌道 i の非局在化の程度を定量化し⁵⁶、 N はある酵素の原子数である。電子が複数の全原子に非局在化している場合を考えるなら D_i は 1 になる。逆に単一の原子に局在化しているなら D_i は $1/N$ となる。(5)で定義した D は全被占軌道について D_i の和を算出した値である。本章では D 値を(5)によって定義する。

4. 3 結果

フォールディングシミュレーション

Figure 2 は、3つのタンパクのダイナミクスを分析した結果を表わす。Amphiphilic Helix では 0 ns から 1.44 ns の領域で水素結合数が時間経過に従って増加し、RMSD が減少する。その後、10.38 ns から 10.44 ns までの領域で RMSD が時間経過と共に減少する。回転半径は 0 ns から約 10 ns にかけて減少し、その後一定の値を示す。Trp-Zipper では 0 ns から約 5.0 ns の領域で水素結合数が時間経過に従って増加するが、RMSD は減少する。その後、12.18 ns から約 20 ns までの領域で RMSD は減少する。後者の時間領域では前者の時間領域と比較して RMSD の減少は比較的緩やかである。回転半径は 0 ns から約 2.0 ns にかけて減少する。Trp-Cage では 0 ns から 75.16 ns までの時間領域で RMSD は一定に保たれるが、水素結合数は緩やかに増加する。その後 75.64 ns までの時間領域で RMSD は減少する。その後は 637.76 ns から 659.32 ns までの時間領域を除き、RMSD は比較的安定する。一方、回転半径は 0 ns から約 10 ns までの時間領域で減少し、その後はほぼ一定の値を保つ。また、約 40 ns から約 170 ns までの時間領域と約 40 ns から約 170 ns までの時間領域で周期的に値の高低値を繰り返す。

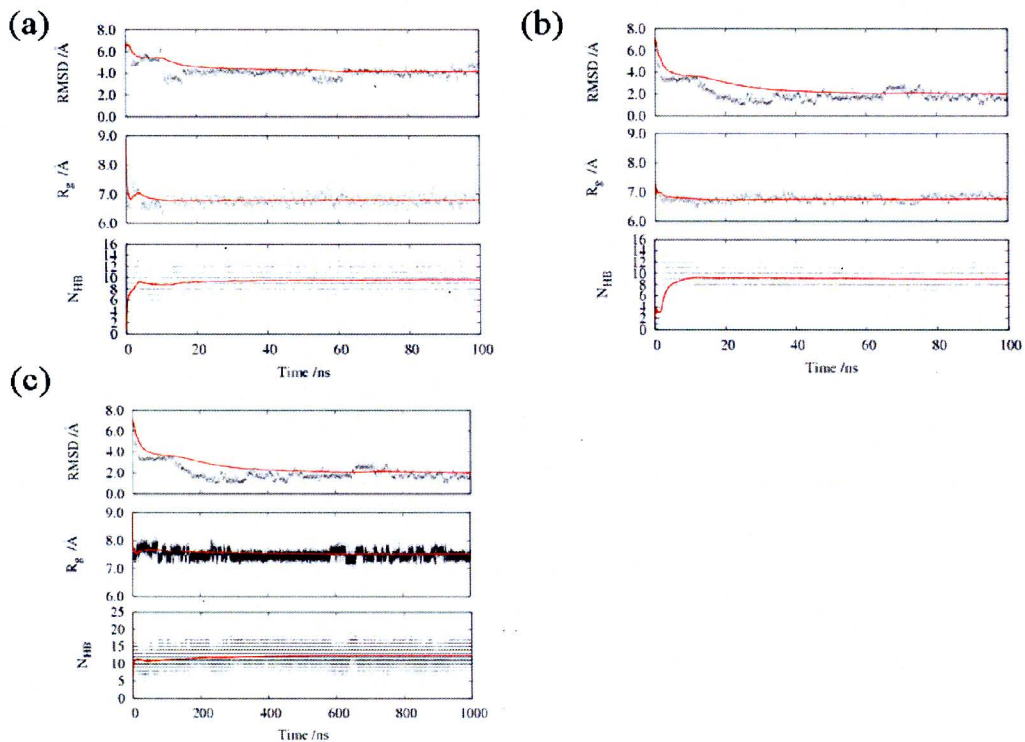


Figure 2. 構造パラメータ (RMSD、回転半径、水素結合数) の時間発展 (あるいは依存性)。黒点はサンプリング構造から直接計算された各パラメータ値を表し、赤線は対応する時間までの平均値を表す。(a) Amphiphilic Helix, (b) Trp-Zipper, (c) Trp-Cage

Figure 3 は、フォールディング過程で得られた代表的なコンフォメーションを可視化したものである。Amphiphilic Helix は約 1.6 ns で Helix の一部が構造化し、その後約 10 ns 程度の時間で Helix の巻き戻りが進行する。その後、安定に Helix 構造を保ちながら約 11.48 ns で実験構造に最も近い構造を形成する (RMSD=1.964 Å)。Trp-Zipper は 1.18 ns で主鎖構造がターン状に折り畳まれ、約 17 ns で構造全体が安定化する。その後、31.86 ns で実験構造に最も近い構造を形成する (RMSD=0.809 Å)。Trp-Cage は約 46 ns で一部の bundle を形成し Loop を安定化させ、徐々に残りの α -Helix を形成する。その後、構造全体を安定に保ちながら、659.32 ns で実験構造に最も近い構造を形成する (RMSD=3.032 Å)。

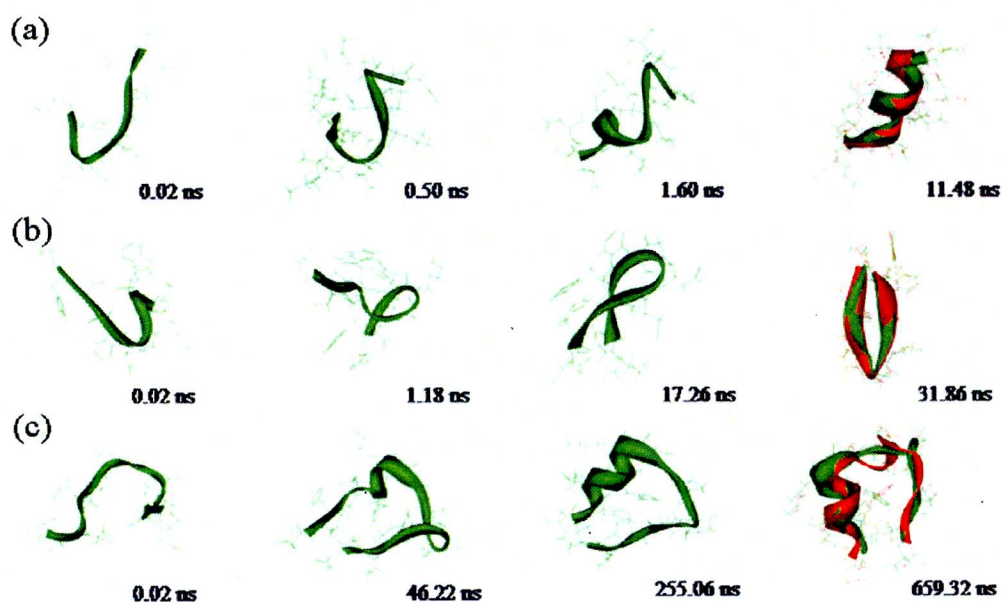


Figure 3. フォールディング過程の代表構造。左端の構造はシミュレーション初期の構造を示し、右端は最小の RMSD を与えたシミュレーションの構造 (緑) と実験構造 (赤) を重ね合わせて得られた構造。左から 2 番目と 3 番目の構造は、その間の時間領域で得られた代表的な構造を示している。これらの構造は水素結合数が一定値を示す時間領域で最初の構造を選出した。右下の数字は構造に対応する時間を示す。(a) Amphiphilic Helix, (b)

Trp-Zipper, (c) Trp-Cage

構造クラスターの分析

Figure 4はシミュレーションで得られたトラジェクトリーを主成分解析によってクラスター分析した結果である。Amphiphilic Helixの場合は、第1主成分(PCA1)と第2主成分(PCA2)の寄与がそれぞれ33.27%と19.24%の合計52.51%であった。これらを軸とする2次元平面上で見るとサンプリングされたコンフォメーションは11個のクラスターを形成している。Trp-Zipperの場合は、PCA1とPCA2の寄与がそれぞれ34.93%と19.02%の合計53.95%で、4つのクラスターが見出された。Trp-Cageの場合は、PCA1とPCA2の寄与がそれぞれ36.35%と23.41%の合計59.76%で、15のクラスターが見出された。

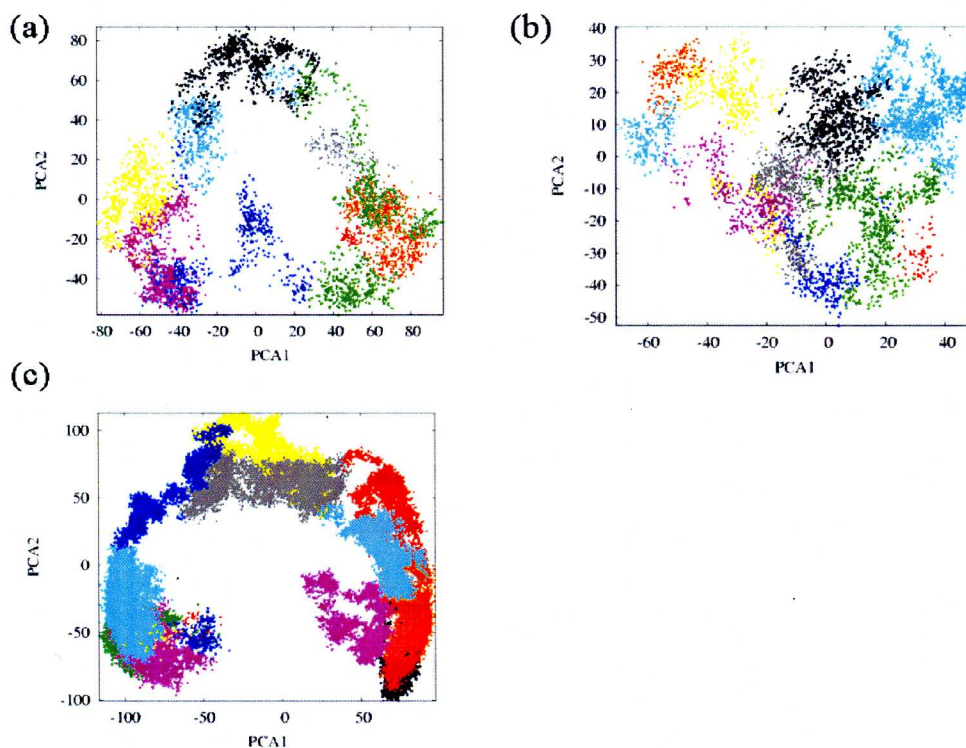


Figure 4. 主成分解析の結果をクラスター分析した結果。クラスター半径は6Åで、異なるクラスターを色分けして区別している。(a) Amphiphilic Helix, (b) Trp-Zipper, (c) Trp-Cage

Figure 5は発生した構造の主成分空間上での数密度を表わしている。いずれのタンパク質の場合も局所的に数密度が集中している箇所がある。Amphiphilic HelixとTrp-CageはPCA1軸の両端にピークが存在し、Trp-ZipperではPCA1とPCA2の最大領域から中央(平均構造)に伸びた連続的な領域にピークが存在する。

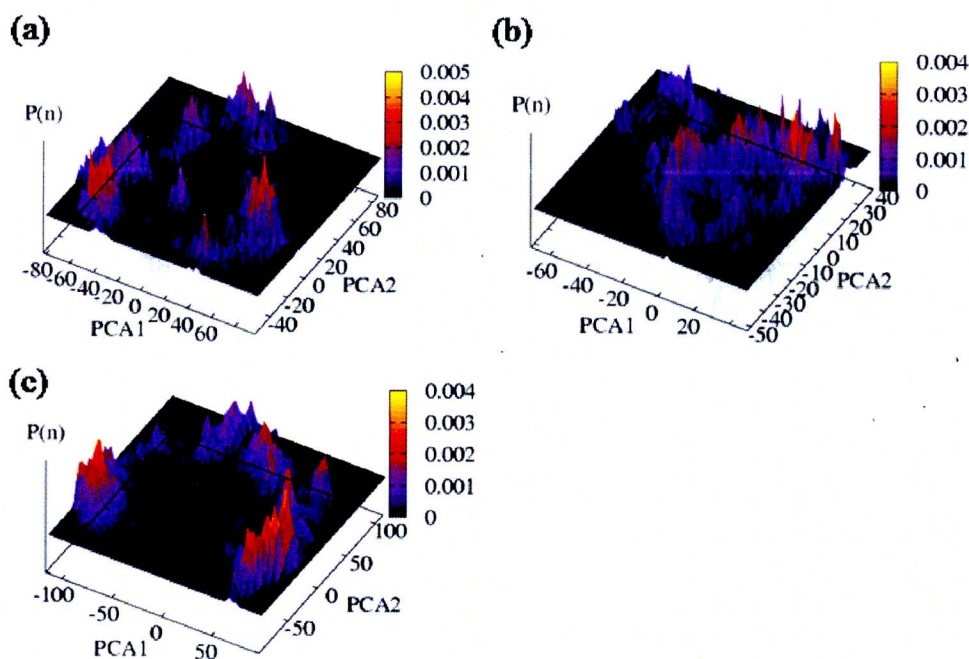


Figure 5. 主成分空間上に発生した構造の数密度 $P(n)$ を分析した結果。数密度の算出は 100×100 の格子点によって区切られた微小区間内で行った。(a) Amphiphilic Helix, (b) Trp-Zipper, (c) Trp-Cage

Figure 6 は主成分空間上にクラスタリングした構造に対して自由エネルギーを算出した結果である (算出法は方法の節を参照)。自由エネルギーは、Amphiphilic Helix では 0-6 kcal/mol、Trp-Zipper では 0-6 kcal/mol、Trp-Cage では 0-7 kcal/mol の範囲に分布している。Figure 4 に示した各構造クラスターの境界領域では、自由エネルギーの相対的に低い領域が存在している。一方、各構造クラスターの中央部分では、Amphiphilic Helix と Trp-Zipper では自由エネルギーが相対的に高く、Trp-Cage では相対的に低い値を示す。

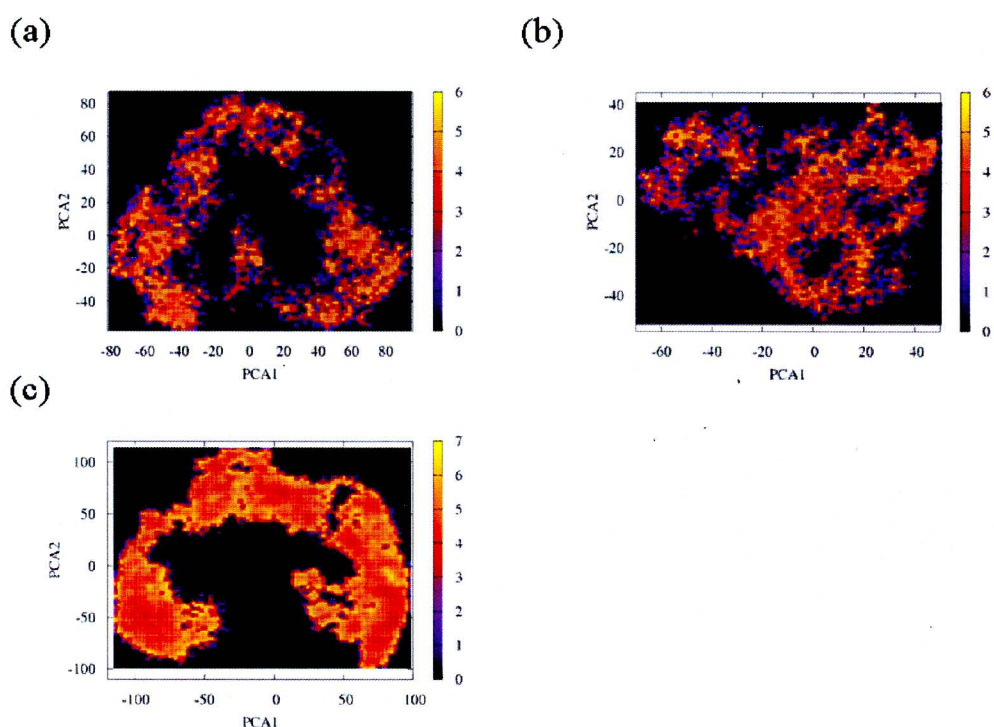


Figure 6. 主成分空間上における自由エネルギーマップ。エネルギー値は Figure5 で算出した数密度を用いて算出した。(a) Amphiphilic Helix, (b) Trp-Zipper, (c) Trp-Cage

Figure 7 は相対コンタクトオーダー (RCO) の値と実験構造に対する RMSD の値から反応座標を分析した結果である。シミュレーションによって発生した構造は様々なコンフォメーションを酔歩している。初期構造から天然類似の構造にフォールドする経路を考えると、その構造空間は必ずしも最短の経路を選択している訳ではないように思われる。マップ中にはいくつか数密度が粗い箇所（密度の極小点）が存在する。この領域では構造が不安定化しているため、フォールディング過程の進行が妨げられると考えられる。Amphiphilic Helix では RMSD が約 4.0 Å の領域に数密度が粗い 3 つの極小点が形成され、各 RCO の値は 13 と 22 と 38 である。最も RCO の小さい極小点の領域付近では勾配が最も大きい。後者の 2 点に関しては勾配が同程度である。Trp-Zipper では RMSD と RCO の値がそれぞれ約 1.5 Å と約 3.2 の領域にそれぞれ極小点が存在する。前者は勾配が急な領域に位置し、後者は比較的緩やかな勾配中に存在する。2 つの領域では数密度が異なり、勾配が急な領域の極値と勾配が緩やかな領域の極値の両者を比較すると、後者の数密度は前者の数密度の 4 倍の値を示す。Trp-Cage にも極小点は存在し、RMSD と RCO の値がそれぞれ約 4.3 Å と約 31 の領域に位置する。

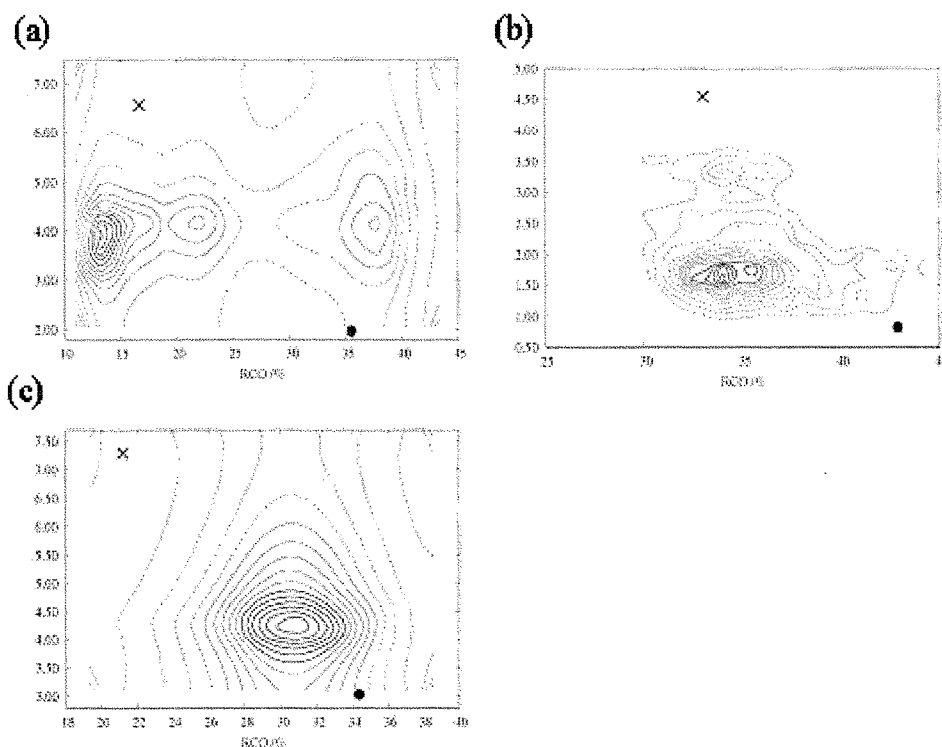


Figure 7. 相対コンタクトオーダー (RCO) に対して RMSD を解析した結果。構造の数密度を等高線で表している。×印と黒丸はそれぞれシミュレーションの開始時と最小の RMSD を与える構造に対応する点を示している。(a) Amphiphilic Helix, (b) Trp-Zipper, (c) Trp-Cage

Figure 8 はアミノ酸残基の配向に関する時間相関関数を解析した結果である。Amphiphilic Helix では 20 ns 以降に Leu2 と Lys5 の配向が揺らぐ。Trp-Zipper では全 Trp の配向がシミュレーションの全時間領域にわたって揺らぐ。特に 60 ns 以降では Trp4 の配向が大きく揺らいでいる。Trp-Cage では、Tyr3 と Trp9 の配向に特徴的な挙動が見られた。500 ns 以降 Trp9 の配向が揺らいだ後、600 ns 以降で Pro18 の配向が大きく揺らいだ。

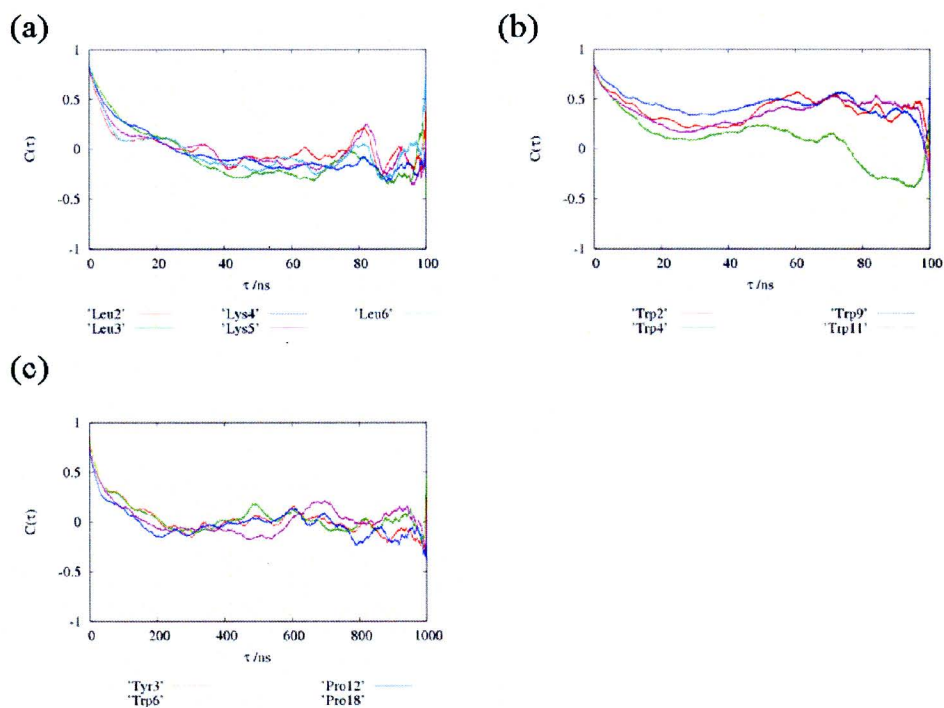


Figure 8. アミノ酸残基の配向に関する時間相関関数。配向は主鎖のカルボニル基の方向ベクトルで表した。(a) Amphiphilic Helix, (b) Trp-Zipper, (c) Trp-Cage

構造形成と波動関数

Figure 9 はシミュレーションの過程で変動する残基の電荷を解析したものである。シミュレーションの過程で残基の持つ電荷は約 0.03 から 0.05 程度の揺らぎを示した。全体的には、分子鎖末端は電荷揺らぎが比較的小さく、内部の特定部位では電荷の揺らぎが大きい、と言える。より詳細に見ると、Amphiphilic Helix では電荷を持った Lys4・Glu8・Glu9・Lys11 の他、Leu2 の電荷が大きく揺らいだ。特に C 末端近くの連続した領域 (Glu8・Glu9・Leu10・Lys11) は電荷の揺らぎが大きい。Trp-Zipper では Gly7 を除く Glu5 から Trp11 までの連続した領域で電荷の揺らぎが大きい。Trp-Cage では電荷を持った Glu5 の他、疎水性残基である Leu2・Ile4・Gln5・Gly11 に関する電荷の揺らぎが大きく Pro19 の電荷は特に揺らぎが大きい。

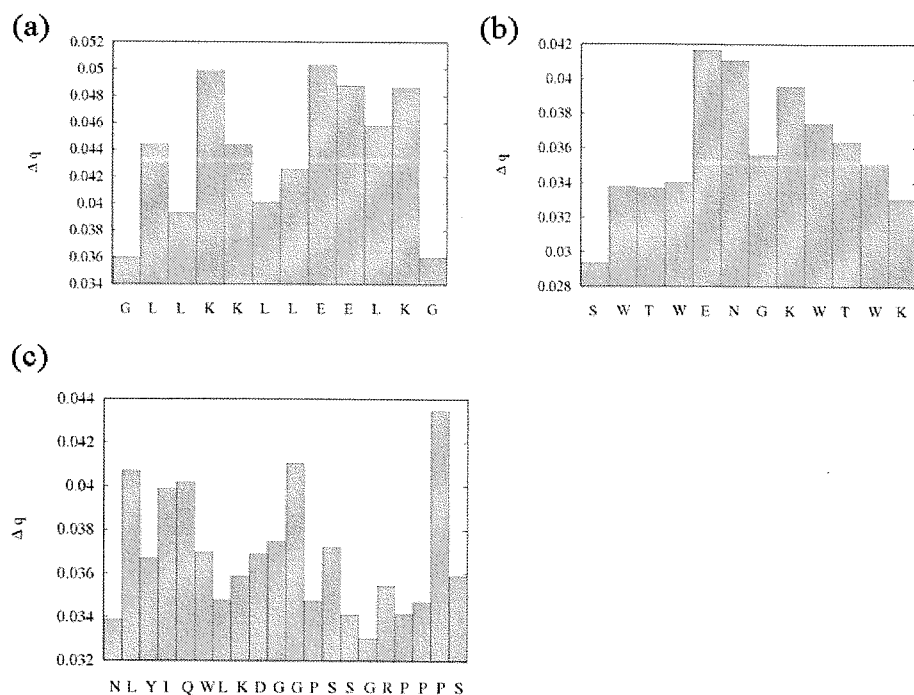


Figure 9. アミノ酸残基の正味電荷に関する揺らぎの大きさ（根二乗平均値）。 (a) Amphiphilic Helix, (b) Trp-Zipper, (c) Trp-Cage

Figure 10. はシミュレーションの過程で変動する残基の電荷に関する相関行列を解析したものである。配列が連続した残基間では電荷の相関が大きく位相は負である。特に分子鎖末端と、その手前に位置する残基との間には比較的強い負の相関がある。その他の残基間では正負の位相が異なる相関が散在している。Amphiphilic Helix では天然構造で水素結合の組になっている Gly1・Lys4・Lys5・Glu9 との間で負の相関が強い。また、Leu7・Leu10 と Lys9・Gly12 との間で正の相関が強い。Trp-Zipper では Glu5・Lys12 及び Asn6・Lys8・Trp9 間に負の相関がある。Trp-Cage では Ile4・Trp6 間で負、Ile4・Leu7 間で正の相関があり、N 末端と C 末端領域の 2 残基 (Pro19・Ser20) 間で負の相関がある。

Figure 10. アミノ酸残基の正味電荷に関する揺らぎの相関（根二乗平均値）。自己相関に関するデータは省いている。 (a) Amphiphilic Helix, (b) Trp-Zipper, (c) Trp-Cage

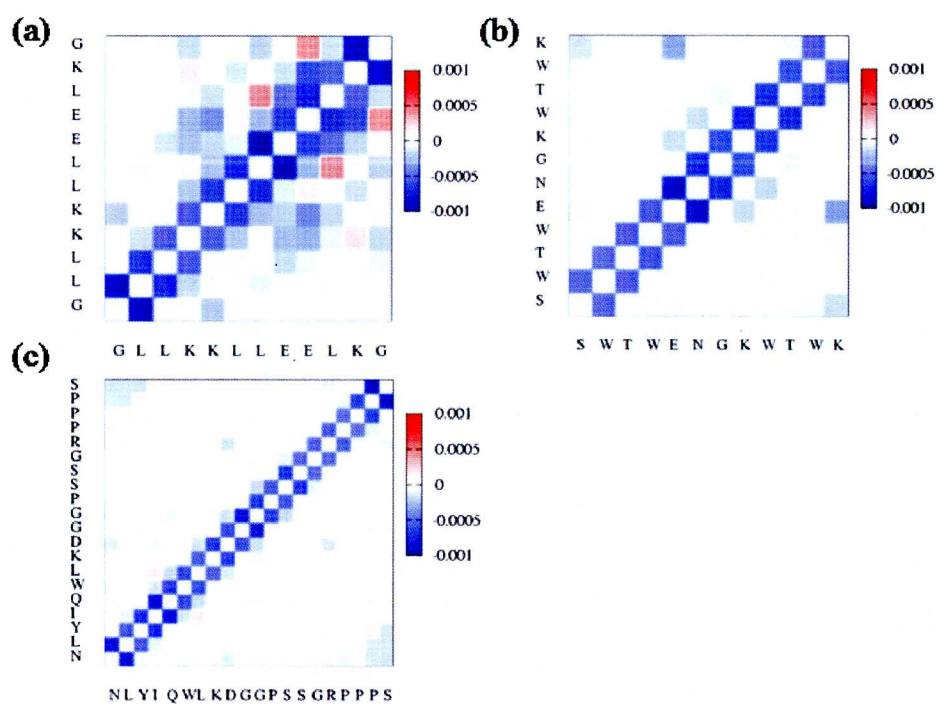


Figure 11 はシミュレーションの過程で変動する軌道エネルギーを解析したものである。エネルギーの揺らぎは最大で 0.7 eV 程度である。相対的にエネルギーの揺らぎが大きな軌道は被占軌道の低エネルギー領域に存在する。

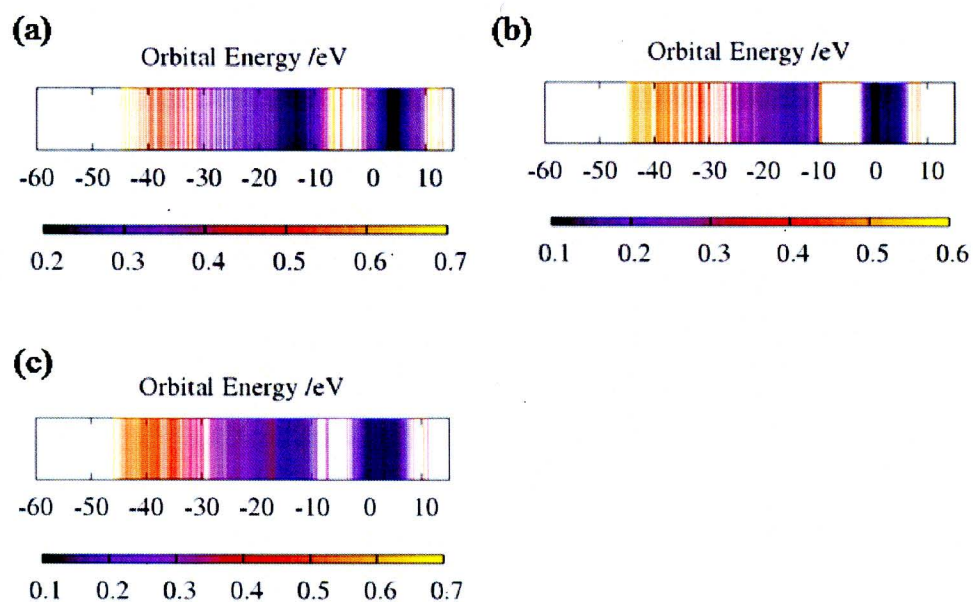


Figure 11. 分子軌道のエネルギーに関する揺らぎの大きさ。線分は軌道エネルギーの平均値を示し、軌道エネルギーの根二乗平均値によって色分けしている。 (a) Amphiphilic Helix, (b) Trp-Zipper, (c) Trp-Cage

Figure 12 は相対コンタクトオーダー (RCO) と D 値を時間に対してプロットした結果である。RCO の増加している領域では D 値は増加する。Amphiphilic Helix ではシミュレーション初期の約 10 ns までの時間領域で RCO と D 値が共に増加する。この時間から約 36 ns になるまで RCO の値は緩やかに増加するが、D 値は減少を始める。その後両者は緩やかに減少を続ける。Trp-Zipper ではシミュレーション初期から 10 ns まで RCO と D 値が共に大きく増加する。その後両者の値は緩やかに増加を続ける。Trp-Cage では RCO は約 60 ns まで上昇した後下降に転じるが、D 値はシミュレーション初期から減少を続ける。

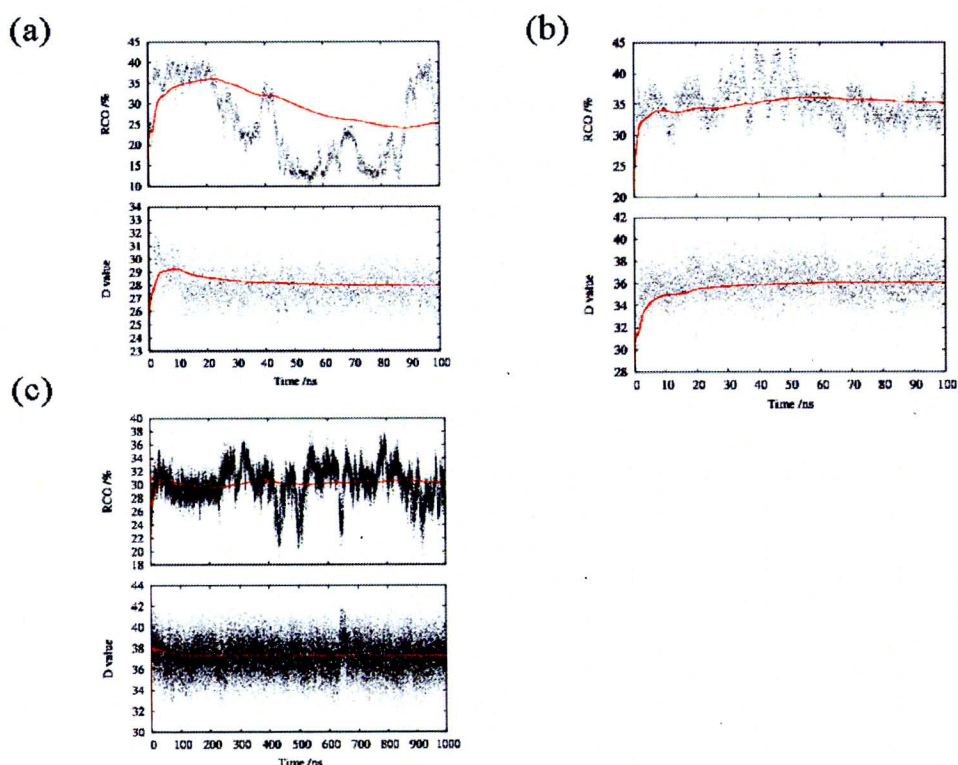


Figure 12. 相対コンタクトオーダー (RCO) と D 値の時間発展。黒点は各サンプリング構造から計算した値を示し、赤線は対応する時間までの平均値を表す。 (a) Amphiphilic Helix, (b) Trp-Zipper, (c) Trp-Cage

Figure 13 は相対コンタクトオーダーと D 値の確率分布をヒストグラムによって分析した結果である。RCO の値は分子によって異なる個数のピークが存在するが、D 値の確率分布

はただ 1 つのピークをもつ。RCO の確率分布のピークは Amphiphilic Helix では 3 カ所、Trp-Zipper では 2 カ所、Trp-Cage では 1 カ所に存在する。

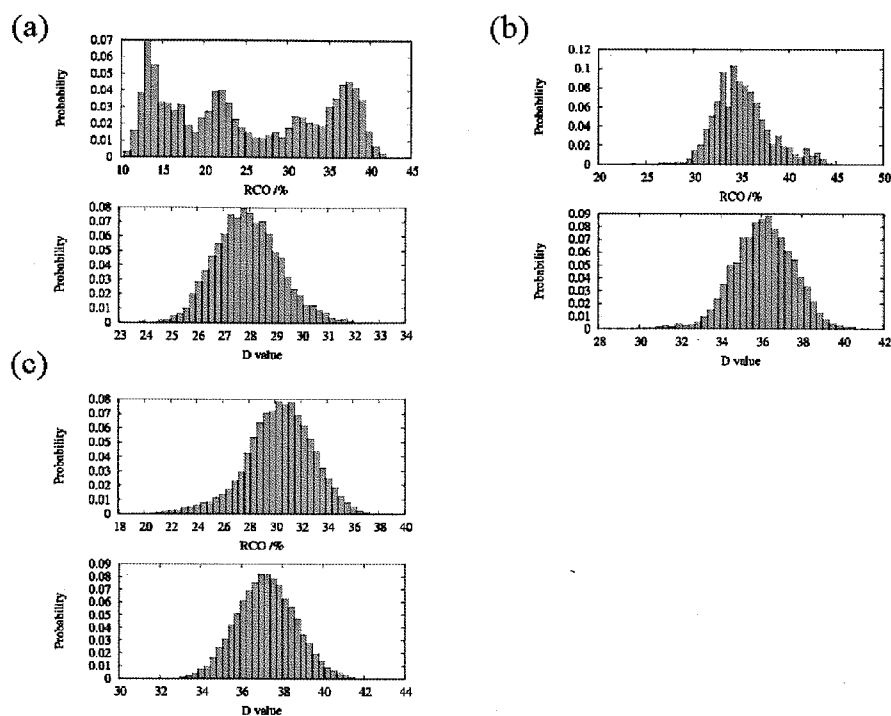


Figure 13. 相対コンタクトオーダーと D 値のヒストグラム。(a) Amphiphilic Helix, (b) Trp-Zipper, (c) Trp-Cage

Figure 14 は主成分解析によって得られたクラスター (Figure 4) の RMSD と D 値を Figure 8 に重ねてプロットした結果である。クラスターの重心に最も近い構造 (以下、重心構造と呼ぶ) とクラスター内で最大の D 値を与える構造 (以下、D 構造と呼ぶ) に関して、両者の描く構造空間の軌跡を比較してみる。すると、重心構造の描くフォールディングの経路は相対的に不連続であり、広い範囲に離散している。D 構造の描く軌跡は比較的実験構造を目指す経路に沿っている。

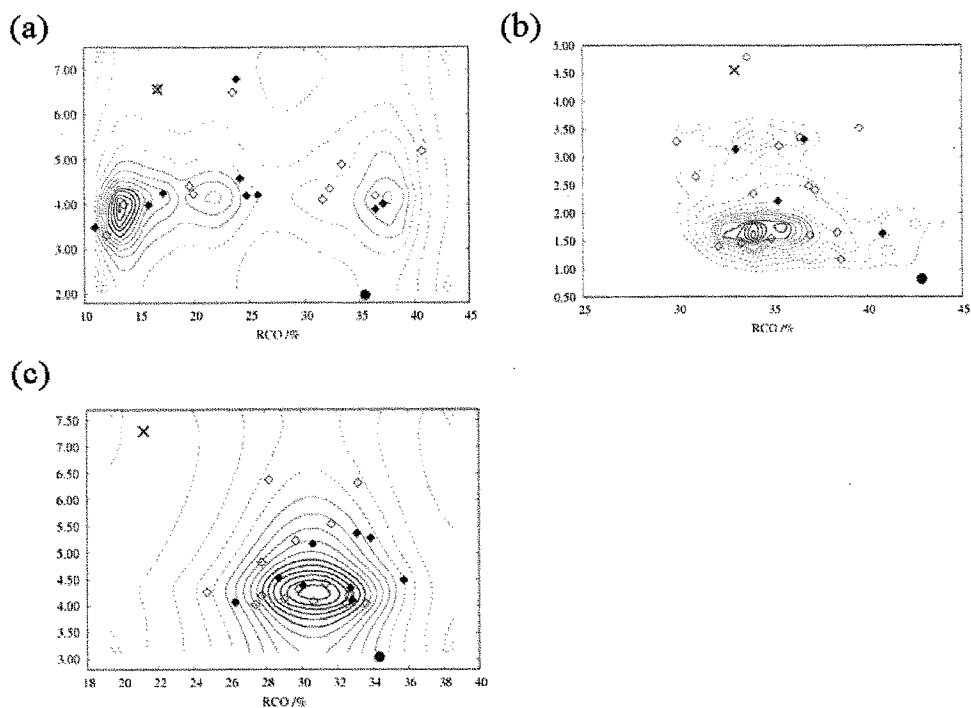


Figure 14. 主成分解析によって得られたクラスターの情報 (Figure 4) を Figure 8 に重ねてプロットした結果。 □はクラスターの重心に最も近い構造、×はクラスター内で最大の D 値を示す構造を表している。 (a) Amphiphilic Helix, (b) Trp-Zipper, (c) Trp-Cage

4. 4 考察

フォールディングシミュレーションの結果

選出したペプチド (Amphiphilic Helix、Trp-Zipper) と小タンパク質 (Trp-Cage) は実験構造に近い構造にフォールドした。Amphiphilic Helix と Trp-Zipper に関しては分子の主鎖と側鎖の両方に関して良好な精度でフォールドし、アミノ酸残基の配向まで一致している (Figure 3)。Trp-Cage の場合は、実験構造で存在の確認されている α -Helix 構造とループに関してはフォールドが完了している。シミュレーションによって得られたコンフォメーションには分布があるが、Figure 5 から、構造は特定構造空間上の特定部位に局在化している。これらの構造は比較的 RMSD が高い領域に位置している (Figure 7)。

Figure 12 の結果から、ある時間を境に RCO 値が大きな変動を示しているが、この時間はコンフォメーションが再構成される状態遷移の境界であると推測できる。Figure 8 の結果から、フォールディングの過程で特定の残基の配向 (相関関数) が時間依存的に変化している。時間発展と共に主鎖コンフォメーションの変化が段階的に起こるが、その状態間遷移は特定の残基の配向の変化によって引き起こされるものと推測される。Amphiphilic Helix では、シミュレーションの開始時から Leu2・Leu3・Leu6 に関する配向の素早い緩和が起こり、20 ns までランダムな配向を維持する。一方、Lys の配向は緩やかに緩和するが、同じく 20 ns でランダムな配向を示すようになる。両グループの配向は 20 ns 付近で全て緩和され、これらの残基が完全に自由度を獲得したことによって、構造全体を再構築する駆動力が生まれると考えられる。Trp-Zipper では 30 ns 付近で RCO の大きな揺らぎが起こるが、この領域は Trp9 の配向相関関数が最小値を示す時間領域である。Trp9 は分子構造の重心近くに位置し、分子構造の安定性に重要な役割を持っている⁵⁷。分子内部を充填する Trp9 の配向が揺らぐため、その変化が近接した残基に伝播し、より広範囲のコンフォメーション変化を引き起こすものと推測できる。Trp-Cage の RCO は 400 ns 付近の領域で大きく変化する。その領域では Trp9 と Pro12 の配向相関関数が突然変化しており、両者の配向は一時的に固定される。Trp9 と Pro12 との間には側鎖を介して (疎水性) 相互作用しており、Trp-Cage 内部の疎水コアを形成している。両残基が特定の配向に誘導されることによって、内部エネルギーの蓄積が起こり、大きな構造変化が引き起こされるものと推測できる。

シミュレーションの問題点を挙げるとすれば、実験構造に対する厳密な再現性という意味から、Trp-Cage のシミュレーションは若干不十分であると言える。具体的には、 3_{10} -Helix の構造と一部の残基側鎖の配向が実験構造と比較して僅かにずれている。この問題点を解消するには、いくつか改良の余地はある。初期構造の選択方法や GB モデルのポテンシャルの種類によって改善できる可能性もある⁵⁸。他にも、実溶媒の適用やシミュレーションの統計量を増やすことが重要な課題であろう。しかし、本章の目的とする議論をするためには、

フォールディングの精度として最低限の条件はクリアしているだろう。

フォールディング過程における電子移動と波動関数の特徴

タンパク質の立体構造形成は電子波動関数の空間的な広がりと関係している。Figure 9 から、フォールディング過程では電荷の揺らぎが生じる。この現象が静電環境の変化に起因するのであれば、揺らぎの大きな部位は荷電性や極性の残基に集中するはずである。しかし、電荷の揺らぎが大きな部位には、それらの残基だけではなく、Leu や Pro、Trp といった残基も含まれる。このことから、小タンパク質の構造形成過程に伴った分子内部の電子的環境変化は、残基の化学構造のみからは説明できない。したがって、構造変化に特異的な電子状態が存在しているのだろう。Figure 10 に示した解析から、残基の持つ電荷の揺らぎは残基間 (分子内) に起こる電子移動とよく相関している。Amphiphilic Helix では、Gly1、Lys5 及び Glu9 の 3 者間に相対的に大きな負の相関がある。これらの残基は α -Helix の構成に携わり主鎖間で水素結合を形成する残基である。これらの残基間では、フォールディングの過程で残基の主鎖間に形成された水素結合を介して電子移動が起こっている。

それでは、フォールディングに伴った電子移動は第 2, 3 章で示したグループ B 軌道に起因する性質なのだろうか。Figure 11 からフォールディングに伴うエネルギーの揺らぎが大きな領域は、グループ B とグループ C の境界に位置する領域である。

Figure 12 の結果から、フォールディングの初期に分子の大まかな構造が形成される時間領域において、RCO と D 値は共に増加する。言い換えると、構造が形成され分子内部が充填されることと D 値の値が増大することは対応している。Figure 13 に見られるように、RCO と D 値の分布にはピークが存在している。Figure 7 から、RCO のピークは構造形成の障壁に対応している。RCO が立体構造を示すパラメータであるのに対し、D 値はピークを持つ確からしい状態を有する統計量としての性質を持つ。反応座標の曲面に極小点に対応する幾何配置をえるためには

従って、D 値によって最適化される構造は一義的に決定される。

Figure 14 に示した通り、D 値を最大化するように構造をサンプリングすることによって、フォールディング経路の中でも比較的効率の良い経路が選択可能である。主成分クラスターの重心の描く軌跡は非常に広範囲の構造空間を移動し、分子の構造を“探索”しているように見て取れる。その一方で D 値の値からサンプリングした構造の軌跡は、比較的無駄なくフォールドする経路を示す。

このように、波動関数の特性を考慮することによって、従来の古典力学的手法では評価できなかった新しい知見を得ることができた。この結果を考慮すると、フォールディングの問題の真の解決のためには、ポテンシャル関数に電子状態を反映する相互作用を組み込んだ新規な方法論を導入が必要であろう。

4. 5 結論

代表的なペプチドと小タンパク質のフォールディングをシミュレーションすることに成功し、その過程で変化する波動関数の性質を詳細に分析した。フォールディングする過程で分子内部に電子移動が起こり、その方向は構造形成に重要な相互作用に対応していた。具体的には、主鎖間の水素結合や疎水コアを形成する残基間であった。この結果は、タンパク質が固有の立体構造を形成する過程において必須な分子内部の相互作用には電子移動が重要な役割を担っている可能性を示唆している。電子波動関数の解析によって、古典的なシミュレーション手法では追跡できない相互作用を評価することができた。また、フォールディングの経路を最適化することが新たな探索法として、方法論の有用性も主張することができた。第2章ではリゾチームの天然構造と比較して変性構造の示すD値のピーク値が顕著に減少したことから、電子波動関数の広がりタンパク質の構造形成（構造安定性）に対応するという仮説を立てた。本章では、実際に小タンパク質が構造形成する過程をシミュレーションすることによって、立体構造の構築と電子波動関数の広がりとの相関関係をより明確に支持する結果を得た。

4. 6 引用文献

1. Anfinsen CB. Principles that govern the folding of protein chains. *Science* 1973;181: 223-230.
2. Goldberger RF, Epstein CJ, Anfinsen CB. Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. *J Biol Chem.* 1963;238:628-635.
3. Levinthal C. "Are there pathways for protein folding?". *J Chim Phys* 1968;65:44-45.
4. Wolynes PG, Onuchic JN, Thirumalai D. Navigating the folding routes. *Science.* 1995;267:1619-1620.
5. Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. R Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. *Proc Natl Acad Sci. U S A* 1995;92:3626-3630.
6. Daggett V, Fersht A. The present view of the mechanism of protein folding. *Nat Rev Mol Cell Biol.* 2003;4:497-502.
7. Daggett V, Fersht AR. Is there a unifying mechanism for protein folding? *Trends Biochem Sci.* 2003;28:18-25.
8. Shirts M, Pande VS. COMPUTING: Screen Savers of the World Unite! *Science.* 2000; 290:1903-1904.
9. Mohanty D, Elber R, Thirumalai D, Beglov D, Roux B. Kinetics of peptide folding: computer simulations of SYPPFDV and peptide variants in water. *J Mol Biol.* 1997;272:423-442.
10. Xiongwu W, Shaomeng W. Folding Studies of a Linear Pentamer Peptide Adopting a Reverse Turn Conformation in Aqueous Solution through Molecular Dynamics Simulation. *J Phys Chem. B* 2000;104:8023-8034.
11. Hummer G, García AE, Garde S. Helix nucleation kinetics from molecular simulations in explicit solvent. *Proteins.* 2001;42:77-84.
12. Daura X, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Reversible peptide folding in solution by molecular dynamics simulation. *J Mol Biol.* 1998;280:925-932.
13. Wu X, Wang S. Helix folding of an alanine-based peptide in explicit water. *J Phys Chem. B* 2001;105:2227-2235.
14. Daura X, Gademann K, Schäfer H, Jaun B, Seebach D, van Gunsteren WF. The beta-peptide hairpin in solution: conformational study of a beta-hexapeptide in methanol by NMR spectroscopy and MD simulation. *J Am Chem Soc.* 2001;123:2393-2404.
15. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science.* 1998;282:740-744.
16. Pitera JW, Swope W. Understanding folding and design: replica-exchange simulations of "Trp-cage" miniproteins. *Proc Natl Acad Sci. USA* 2003;100:7587-7592.

17. Qiu L, Pabit SA, Roitberg AE, Hagen SJ. Smaller and faster: the 20-residue Trp-cage protein folds in 4 micros. *J Am Chem Soc.* 2002;124:12952-12953.
18. Snow CD, Zagrovic B, Pande VS. The Trp cage: folding kinetics and unfolded state topology via molecular dynamics simulations. *J Am Chem Soc.* 2002;124:14548-14589.
19. Chowdhury S, Lee MC, Xiong G, Duan Y. Ab initio folding simulation of the Trp-cage mini-protein approaches NMR resolution. *J Mol Biol.* 2003;327:711-717.
20. Jang S, Shin S, Pak Y. Molecular dynamics study of peptides in implicit water: ab initio folding of beta-hairpin, beta-sheet, and beta beta alpha-motif. *J Am Chem Soc.* 2002;124:4976-4977.
21. Berg BA, Neuhaus T. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys Rev Lett.* 1992;68:9-12.
22. Hukushima K, Nemoto K. Exchange Monte Carlo and Application to Spin Glass Simulations. *J Phys Soc Jpn.* 1996;65:1604-1608.
23. Hukushima K, Takayama H, Nemoto K. Application of an extended ensemble method to spin glasses. *Int J Mod Phys.* 1996;7:337-344.
24. Lyubartsev P, Martsinovskii AA, Shevkunov SV, Vorontsov-Velyaminov PN. "New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles". *J Chem.Phys.* 1992;96:1776-1783.
25. Marinari E, Parisi G. Simulated tempering: A new Monte Carlo scheme. *Europhys Lett.* 1992;19:451-458.
26. Patel S, Mackerell AD Jr, Brooks CL 3rd. CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J Comput Chem.* 2004;25:1504-1514.
27. Hunter CA. Sequence-dependent DNA structure. The role of base stacking interactions. *J Mol Biol.* 1993;230:1025-1054.
28. Burley SK, Petsko GA. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science.* 1985;229:23-28.
29. Brandl M, Weiss MS, Jabs A, Sühnel J, Hilgenfeld R C-H...pi-interactions in proteins. *J Mol Biol.* 2001;307:357-377.
30. Scheiner S, Kar T. Effect of solvent upon C Prajapati RS, Sirajuddin M, Durani V, Sreeramulu S, Varadarajan R. Contribution of cation-pi interactions to protein stability. *Biochemistry.* 2006;45:15000-15010.
31. Samanta U, Pal D, Chakrabarti P Packing of aromatic rings against tryptophan residues in proteins. *Acta Crystallogr D Biol Crystallogr.* 1999;55:1421-1427.
32. Eisenberg D, Wilcox W, Eshita SM, Pryciak PM, Ho SP, DeGrado WF. The design, synthesis, and crystallization of an alpha-helical peptide. *Proteins.* 1986;1:16-22.

33. Hill CP, Anderson DH, Wesson L, DeGrado WF, Eisenberg D. Crystal structure of alpha 1: implications for protein design. *Science*. 1990;249:543-546.
34. Privé GG, Anderson DH, Wesson L, Cascio D, Eisenberg D. Packed protein bilayers in the 0.90 Å resolution structure of a designed alpha helical bundle. *Protein Sci*. 1999;8:1400-1409.
35. Cochran AG, Skelton NJ, Starovasnik MA. Tryptophan zippers: stable, monomeric beta-hairpins. *Proc Natl Acad Sci. USA*. 2001;98:5578-5583.
36. Muñoz V, Thompson PA, Hofrichter J, Eaton WA. Folding dynamics and mechanism of beta-hairpin formation. *Nature*. 1997;390:196-199.
37. Dinner AR, Lazaridis T, Karplus M. Understanding beta-hairpin formation. *Proc Natl Acad Sci. USA* 1999;96:9068-9073.
38. G. Wei, P. Derreumaux, N. Mousseau. "Sampling the complex energy landscape of a simple beta-hairpin". *J Chem Phys*. 2003;119:6403-6406.
39. Neidigh JW, Fesinmeyer RM, Andersen NH. Designing a 20-residue protein. *Nat Struct Biol*. 2002;9:425-430.
40. Neidigh JW, Fesinmeyer RM, Prickett KS, Andersen NH. Exendin-4 and glucagon-like-peptide-1: NMR structural comparisons in the solution and micelle-associated states. *Biochemistry*. 2001;40:13188-13200.
41. Tsui V, Case DA. Theory and application of the generalized solvation model in macromolecular simulations. *Nucl Acid Sci*. 2001;56:275-291.
42. Wang J, Cieplak P, Kollman PA, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem*. 2000;21:1049-1074.
43. Ryckaert JP, Ciccotti G, Berendsen HJC. "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes". *J Comput Phys*. 1977;23:327-341.
44. Harvey SC, Tan RK, Cheatham TE. The flying ice cube: Velocity scaling in molecular dynamics leads to violation of energy equipartition. *J Comput Chem*. 1998;19:726-740.
45. Daidone I, Ulmschneider MB, Di Nola A, Amadei A, Smith JC. Dehydration-driven solvent exposure of hydrophobic surfaces as a driving force in peptide folding. *Proc Natl Acad Sci. USA* 2007;104:15230-15235.
46. Sitkoff, Doree, Kim A. Sharp and Barry Honig, "Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models", *J. Phys. Chem*. 1994;98:1978-1988.
47. Feig M, Karanicolas J, Brooks CL 3rd MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J Mol Graph Model*. 2004;22:377-395.
48. Xu S, Kamath MV, Capson DW. Selection of Partitions from a Hierarchy. *Pattern Recognition Letters*. 1993;14:7-15.

49. Jang S, Kim E, Shin S, Pak Y. Ab Initio Folding of Helix Bundle Proteins Using Molecular Dynamics Simulations. *J Am Chem Soc.* 2003;125:14841-14846.
50. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol.* 1998;277:985-994.
51. Sergiy O. Garbuzynskiy, Michail Yu. Lobanov and Oxana V. Galzitskaya To be folded or to be unfolded? *Prot Sci.* 2004;13:2871-2877.
52. Stewart JJP. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *Int J Quantum Chem.* 1996;58:133-146.
53. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. AM1: A new general purpose quantum mechanical molecular model. *J Am Chem Soc.* 1985;107:3902-3909.
54. Stewart JJP. MOPAC2002. Fujitsu, Ltd., Tokyo, Japan(2002) MOPAC2002
55. Pople JA, Beveridge DL. Approximate Molecular Orbital Theory. New York: McGraw-Hill; 1970. Chapters 3 and 4.
56. Ohno K, Wada M, Saito S, Inoue Y, Sakurai, M. Quantum chemical study on the affinity maturation of 48G7 antibody. *J Mol Struct. THEOCHEM* 2005;722:203-211.
57. Snow CD, Qiu L, Du D, Gai F, Hagen SJ, Pande VS. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. *Proc Natl Acad Sci. USA* 2004;101:4077-4082.
58. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins.* 2004;55:383-394.

第5章 総括

本研究によって得られた結論をまとめると以下の通りとなる。

第2章の前半では、タンパク質・酵素の立体構造、機能及び波動関数の間のユニバーサルな関係を明らかにすべく、EC 分類のすべてをカバーする構造既知・機能既知の 112 個の酵素に対して全原子分子軌道計算を実行し、得られた電子状態を比較・検討した。このようにタンパク質あるいは酵素分子全体の電子状態を直接的に評価し、しかも系統的な解析を行ったのは本研究がはじめてである。電子状態の解析では特に、立体構造の構築と直接結びつくと考えられる、波動関数(分子軌道あるいは MO と同義)そのものの特質を抽出することに重点を置いた。その際、1 分子あたり 10000~20000 個程度も存在する波動関数を系統的かつ迅速に比較検討するための指標として、MO の広がり (非局在化) を表す D 値パラメータを導入した。このパラメータの解析により、タンパク質・酵素の被占軌道は階層的な様式で構築されており、電子の非局在化の度合いによって 3 つのグループ A~C に分類可能であることが判明した。グループ A は、HOMO-LUMO バンドギャップ近傍に存在し空間的に局在化した MO から成る。グループ B は through-space 相互作用によりタンパク質構造全体に広がった MO の集合であり、グループ C は through-bond 相互作用により共有結合に沿って形成された MO から成る。この MO の分類は、タンパク質中の電子の波動パターンを明らかにしたことに相当する。一般に、分子軌道計算は近似のレベルによってエネルギー、電子密度等に対し数値的に異なった値を与えるが、波動のパターン (節面の位置と数) に関しては一定の結果を与える。したがって、上の分類 A~C は計算精度に依存しないタンパク質の MO の普遍的特性であると考えてよい。また、立体構造の崩壊に伴って、このような分類が不明瞭になる (グループ B の消失) ことも考慮すると、分類 A~C は天然状態におけるタンパク質に特異な性質の一つであると考えられる。重要なことは、このような分類 A~C は単に MO の形式的分類に留まらず、タンパク質の構造構築原理や機能発現機構を解く手がかりを与える、ことである。すなわち、グループ A は HOMO、LUMO を含むため化学反応活性の高い MO の集合体であり、これを解析することにより、酵素であれば機能部位の予測につながる。グループ B は立体構造の構築・安定性に関係しており、その解析は“タンパク質はどのようにしてフォールドするか (フォールディング問題)”というタンパク質科学の最大の問題の一つに量子力学から回答を与える切り口となる。グループ C は分子の共有結合骨格 (水素結合ネットワーク含む) の形成と安定性を解く鍵となる MO である。

第2章後半では、上で述べた 112 個の酵素についてグループ A に属する MO の局在場所と酵素の機能部位のそれとの関係を詳細に調べ、次の結論を得た。まず、真空状態に比べ水和状態の酵素では HOMO、LUMO あるいはエネルギー的にそれらに近い MO (これら

をまとめてフロンティア軌道と呼んだ)が高い確率で活性部位に局在している、ことを示した。それゆえ、水はタンパク質の立体構造安定性のみならず、酵素の活性発現そのものにも重要な寄与をしている、といえる。これらのことを総合すると、水溶性酵素の立体構造はフロンティア軌道が活性部位に局在化するように進化してきた、という仮説が成り立つかもしれない。次に、この 112 個の酵素に対するベンチマークデータを使って、逆に、フロンティア軌道の探索から機能部位がどの程度の精度で予測できるかを定量的に示し、その結果を既存の予測法と比較した。その結果、芳香族残基に対して比較的高い false positive 値を示すものの、他の方法に匹敵し得る予測精度があることを明らかにした。最終的に、フロンティア軌道の探索から機能部位（特に、触媒基となる Glu や Asp）を予測するプロトコルを適用事例とともに提案し、本博士論文研究の目的の一つである機能予測法の開発を達成した。

前述したとおり第 2 章において、タンパク質の波動関数に与える水の効果について新規な知見を得たが、そこでは水を連続体近似で扱っていたため、水そのものの量子力学的効果（水自身の波動関数の効果）は考慮されていなかった。そこで、第 3 章では、タンパク質と水クラスタからなる超分子系の計算を行い、この巨大系の量子力学的性質を明らかにするとともに連続体近似（仮想溶媒）の問題点を論じた。選択された 4 つのタンパク質に対する超分子計算から得られた重要な結果は次の通りである。すなわち、1) タンパク質から水クラスタに 3e から 5e 程度の電子移動が起こるが、その影響範囲はほぼ第 2 水和圏までである、言い換えると、水の量子力学的効果は第 2 水和圏まで考えれば十分である、2) タンパク質のグループ B 軌道と水クラスタ（第 2 水和圏までを含む）のグループ B 軌道の間に融合（互いの波動関数の重ね合わせ）が起こり、新たに超分子系全体に広がる結合性 MO と反結合性 MO が形成される、3) グループ C 軌道については、タンパク質と水クラスタの間に若干の融合は見られるが、基本的には各系に独立に局在化した波動状態が支配的である、4) 連続体近似は、分子の内部エネルギーについて過大評価する傾向がある。

1) で述べた電子移動は、タンパク質部位特異的に起こるものであり、特に解離した Glu および Asp が大きな寄与をしていることが明らかになった。さらに、高エネルギーの被占軌道（反応活性の高い軌道）からの電子移動がどのような残基から起こるかを明確にするために CP パラメータ（第 3 章式(10)）を導入し、4 つのタンパク質について解析した。その結果、興味あることに、触媒部位近傍が高い値を示した。これは、すなわち超分子計算と CP パラメータの解析を組み合わせれば、機能部位の予測が可能であることを示唆している。超分子計算では水の量子力学的効果（水素結合などの特異的相互作用など）が考慮されているので、第 2 章の連続体モデルによる予測法より精度的に優れているかもしれない。しかしながら、一つの超分子系の結果を得るには、MD シミュレーションによる水クラスタ構造の決定と超分子全体の全原子計算が必要であり、連続体水モデルによる計算とは違い、多くの系にルーティン的に適用するのは困難である。今後は、第 2 章で提案した連続体水モデルによる機能予測法の欠点を補うような形の 2 次的・部分的利用法を考えるべきで

あろう。一方、2) で述べたグループB領域に現れる超分子系全体に広がる結合性MOの存在は、従来熱力学的観点から言われてきた水のタンパク質構造安定化効果を量子力学の言葉で実証したものとみることもできる。タンパク質のグループB軌道は第2章で述べたように立体構造を安定化する役割を持つが、これが水クラスターのグループB軌道と融合することにより、電子の動ける空間的範囲がより広がり、よりいっそうの電子的安定化効果がもたらされるからである。

グループB軌道がタンパク質の立体構造を安定化する役割を持つならば、当然、フォールディング過程でも重要な役割を演じているはずである。第4章では、このことを検証するため、小タンパク質のフォールディング過程を解析した。その結果、古典的MDシミュレーションから得られたフォールディングの自由エネルギーランドスケープ上で、出発構造、いくつかのローカルミニマム及び最終構造（天然構造に近い構造）を比較的短距離に結ぶ経路は、各ローカルミニマム（通過点）での構造のD値をなるべく大きくするような経路であった。このことは、フォールディングは、D値が大きい構造（軌道の広がりが大きく電子的に安定な構造）を選択しながら進行していくと解釈することができる。また、D値が大きい軌道が発達する駆動力は、分子内部での局所的電子移動であることも示唆された。もう少し具体的に言うと、たとえば電子移動により形成される水素結合がD値を大きくし、これが発達すると α ヘリックスのような二次構造形成が起こるというようなイメージである。これは従来の古典力学的取扱いでは捉えることのできなかった、フォールディング過程に関する全く新しい描像である。

言うまでも無く、タンパク質は特定の立体構造を持つことにより機能を発現する。したがって、タンパク質科学にとって重要な課題は立体構造の構築原理及び構造-機能相関の解明である。一方、タンパク質の電子波動関数は立体構造によって一義的に決定され、そのタンパク質の力学的・化学的性質に関するすべての情報が電子波動関数に含まれる。本研究では、以上述べた通り、この量子力学の基本原則に従って、タンパク質科学の基本問題に新しい解釈を与えることに成功した。本研究では、電子波動関数が持つ情報の一部を引き出し、その利用法を示したに過ぎないが、電子波動関数からの生体分子・生体反応へのアプローチ“波動関数生物学”は、確実に幕を開けたといえよう。

謝辞

本論文の執筆に関して全面的に指導していただいた櫻井教授に心より深く感謝致します。指導面のみならず、研究の面白さや研究者としての誇りなど、様々なことを教えていただきました。私が研究生活によって成長できたのは、ひとえに先生のおかげです。公私にわたり御懇篤なる御指導とご厚情を賜り、心よりお礼申し上げます。

研究生活を支えてくれた櫻井研究室の皆様には深く感謝いたします。長い 6 年の研究生活を有意義に過ごすことができたのは研究室の皆様のおかげです。皆様のご厚情に感謝いたします。生活面で相談に乗っていただいた先輩の新井博士に感謝いたします。研究の面白さを共有し合った同期の宮川、後輩の野村君、雄花山君、岡部君、林智彦君に感謝します。研究室に明るい雰囲気を提供してくれた渡邊さん、河崎さんに感謝します。

最後に、このような素晴らしい環境で学ぶ機会を与えてくれた両親に感謝します。

2007 年 12 月

福島 健太郎