

核酸および核酸タンパク質複合体の  
ダイナミクス解析に関する研究

中 村 周 吾



# もくじ

1	はじめに	3
I	生体分子の構造モデリング・ダイナミクス解析システムの構築	11
2	第 I 部の概要	13
3	配座エネルギー関数と座標系	14
3.1	AMBER の配座エネルギー関数	14
3.2	XYZ 系と二面角系	20
4	構造モデリング・ダイナミクス解析システムの構築	23
5	二面角系配座エネルギー 2 次微分行列の並列計算法の開発	26
5.1	二面角系における Hessian 計算法	28
5.2	Unify 方式による Hessian 並列計算法の開発	31
5.2.1	Unify 方式のアルゴリズム	31
5.2.2	Unify 方式による Hessian 計算の並列化効率	33
5.2.3	Unify 方式によるニュートン・ラフソン法の並列化効率	38
5.3	サブタスクグループ方式による Hessian 並列計算法の高速化	43
5.3.1	サブタスクと依存関係	43
5.3.2	サブタスク間の依存関係の解析	44
5.3.3	サブタスクグループ方式のアルゴリズム	44
5.3.4	遺伝的アルゴリズムを用いたサブタスク最適割り当ての探索	47
5.3.5	サブタスクグループ方式による Hessian 計算の高速化の結果	52
6	第 I 部のまとめ	59

<b>II</b>	<b>DNA ミニヘアピン分子の構造モデリングと熱安定性解析</b>	<b>61</b>
7	第 II 部の目的	63
8	Two-Stage 法の開発とミニヘアピン分子の構造モデリング	68
8.1	エネルギー最小化による構造モデリング法	68
8.2	Two-Stage 法の概要	70
8.3	初期構造の構築と計算パラメータ	72
8.4	構造モデリング結果	77
8.4.1	dGC(GNA)GC のモデリング	77
8.4.2	dGC(GAC)GC のモデリング	79
8.4.3	dGC(GAG)GC のモデリング	79
8.5	ループ部のエネルギー解析の結果	80
8.6	構造モデリング結果のまとめ	81
9	ミニヘアピン分子の熱安定性の解析	82
9.1	Poisson-Boltzmann 分子動力学法	82
9.2	反応座標の設定	85
9.3	構造サンプリングの方法	85
9.4	WHAM を用いた自由エネルギー計算の方法	87
9.5	反応座標に拘束をかけない分子動力学計算の結果	89
9.6	自由エネルギー解析の結果	89
9.7	熱安定性解析結果のまとめ	96
10	第 II 部のまとめ	97
<b>III</b>	<b>tRNA・アミノアシル tRNA 合成酵素のダイナミクス解析</b>	<b>99</b>
11	第 III 部の目的	101
12	基準振動モード解析法	112
12.1	基準振動モード解析法の概要	112
12.2	シミュレーションの手順	114
13	フリー状態の tRNA のダイナミクス解析	118
13.1	モード周波数の解析結果	118

13.2	各原子の熱ゆらぎの大きさ . . . . .	120
13.3	低周波数モードにおける熱ゆらぎ . . . . .	123
<b>14</b>	<b>tRNA・アミノアシル tRNA 合成酵素複合体のダイナミクス解析</b>	<b>130</b>
14.1	フリー状態の構造の構築 . . . . .	130
14.2	モード周波数の解析結果 . . . . .	131
14.3	各原子の熱ゆらぎの大きさ . . . . .	131
14.4	低周波数モードにおける熱ゆらぎ . . . . .	137
14.5	複合体の基準振動モード運動のフリー状態との比較 . . . . .	143
<b>15</b>	<b>第 III 部のまとめ</b>	<b>152</b>
<b>16</b>	<b>おわりに</b>	<b>157</b>
	謝辞	159
	Appendix	161
A	Two-Stage 法の理論 . . . . .	161
A.1	ポテンシャルのガウス関数での近似 . . . . .	161
A.2	ガウス関数に対する拡散方程式の解 . . . . .	162
A.3	Time-reversing procedure . . . . .	163
B	Poisson-Boltzmann 分子動力学法の理論 . . . . .	164
B.1	Poisson-Boltzmann 方程式 . . . . .	164
B.2	有限差分 Poisson-Boltzmann 法 . . . . .	165
B.3	Poisson-Boltzmann 分子動力学法 . . . . .	169
C	WHAM の理論 . . . . .	171
C.1	1 回のシミュレーションからの系の状態密度の見積もり . . . . .	172
C.2	複数のシミュレーションからの系の状態密度の見積もり . . . . .	173
D	基準振動モード解析の理論 . . . . .	175
	参考文献	179



はじめに





## 第 1 章

### はじめに

ノーベル医学・生理学賞の受賞学者である R. Dulbecco がヒトゲノム計画を提案<sup>1)</sup>してから 15 年が経ち、われわれはゲノム時代のまさにまっただなかにある。当時はヒトゲノム計画の現実性に懐疑的な意見もあったが、今やヒトのみならずさまざまな生物種でゲノム計画が進展している。タンパク質をコードしているヒトの遺伝子の総数がショウジョウバエの約 2 倍にあたる 3 万個から 4 万個程度であるという発表は記憶に新しい<sup>2,3)</sup>。

ゲノム計画は、対象となる生物種のゲノムの塩基配列を決定する作業であり、膨大なデータが日々蓄積されている。一方で、網羅的にタンパク質の立体構造を明らかにしようとする構造ゲノミクスと呼ばれる研究も盛んに行われている。タンパク質や核酸の立体構造データベースである Protein Data Bank (PDB) のエントリ数は、X 線結晶解析や NMR など、生体分子の立体構造を決定するための技術の発展によって、ここ 10 年で指数関数的に増大している。

このような中で、コンピュータを用いて、生体分子の果たす機能に物性やダイナミクス、熱安定性などの物理化学的な側面からアプローチするコンピュータ・シミュレーションの重要性が急速に高まりつつある。その主な理由として、以下の 3 つが考えられる。

1 つ目は、上で述べたような、ゲノム計画などのプロジェクトの進展や生化学実験の手法の発達などによる、遺伝子の塩基配列、タンパク質のアミノ酸配列、立体構造、ダイナミクス、機能などに関する膨大なデータの蓄積である。これらの大量のデータから意味のある情報を引き出すためには、コンピュータを用いることが必須である。構造ゲノミクスなどにより次々決定される構造を手がかりにタンパク質・核酸の機能や他の分子との相互作用を理解する際に、コンピュータ・シミュレーションによる分子のダイナミクス解析などから得られる知見は大きい。また構造データベースの拡大は、アミノ酸や塩基の配列と立体構造の関係に多くの知見をもたらし、

コンピュータによる構造モデリングの精度向上の大きな要因ともなっている。

2つ目は、シミュレーション技術の発達である。たとえば生体分子に対する溶媒効果やアミノ酸を置換したときの自由エネルギー変化などの計算アルゴリズム、エネルギーパラメータの改良などにより、コンピュータ・シミュレーションで生化学実験の結果を高い精度で再現できるようになってきている。また量子化学計算と分子力学計算を組み合わせた QM/MM と呼ばれる方法の開発などにより、化学反応を伴った分子全体にわたるシミュレーションも現実的なものとなりつつある。

3つ目は、コンピュータの高速化と並列計算技術の進展である。ハードウェアの技術革新は目ざましく、10 数年前のスーパーコンピュータと同程度かそれ以上の性能を安価なパーソナルコンピュータがもつ。またこれに加えて、1つの仕事を複数のプロセッサで分担して行う並列計算が、アルゴリズムの開発やハードウェアの発達によって、急速に普及しつつある。世界のスーパーコンピュータの性能上位 500 台の諸元をまとめた J. Dongarra らによるレポート<sup>4)</sup> では、2000 年 11 月現在そのすべてがプロセッサ数 16 以上の並列コンピュータであり、性能 1 位はプロセッサ数 8192 の IBM ASCI White となっている。このような流れを受けて、分子動力学法をはじめとする分子シミュレーション手法の並列計算アルゴリズム<sup>5-7)</sup> や並列アプリケーション実行基盤<sup>8-10)</sup> が開発されている。これらにより、計算可能な系のサイズが拡大し、精度を上げた大規模な計算を高速に行うことが可能となり、より現実に近い系でのシミュレーションができるようになってきている。

シミュレーションでは、モデル化された系において少数の分子のダイナミクスやその他の物性を解析し、可視化することができる。また手法や分子の種類を次々と変えた網羅的な解析が可能である。一方、生化学実験は、現実の系において多数の分子のふるまいを分子からのさまざまなシグナルを解析することで直接明らかにすることができる。これらのことに象徴されるように、シミュレーションと生化学実験は相補的な得意分野をもち、互いに得られた知見をフィードバックしあうことで非常に有益な情報を得ることができると考えられる。

さて、タンパク質分子については、数多くのシミュレーションがこれまで行われ、手法や結果の蓄積も多い。しかし核酸分子については、タンパク質に比べると研究例が少ないのが実情である。1977 年に J. A. McCammon らが BPTI (bovine pancreatic trypsin inhibitor) に分子動力学法を初めて適用してから 6 年後には、M. Levitt と B. Tidor らが DNA の分子動力学計算を相次いで行っている<sup>11-13)</sup>。しかし現在においても、核酸については配列ベースの解析が主で、シミュレーションの研究例は少ない。一方、核酸分子の立体構造やダイナミクスが機能を果たす上で大きな役割を担っていることはこれまでに数多く報告されており、PDB 中の核酸、核酸タンパ

ク質複合体のエントリ数もとくに最近急速に増えてきている。

本論文は以上のような時代背景をふまえ、核酸および核酸タンパク質複合体を対象として、コンピュータ・シミュレーションによって、構造モデリング、自由エネルギー解析、ダイナミクス解析を行った研究について述べる。

シミュレーションは大きく分けると、要素技術の開発と、生体分子への適用およびその結果の解析の 2 本の柱から成る。本研究では、まずシミュレーションを行うための要素技術の開発として、生体分子の構造モデリング・ダイナミクス解析システムの構築を行った (第 I 部)。このシステムは、筆者らが開発したエネルギー最小化手法を用いた生体分子の構造モデリング、Poisson-Boltzmann 分子動力学法による構造サンプリングと自由エネルギー解析、二面角基準振動モード解析法によるダイナミクス解析などを、並列計算技術を用いて統合的に、かつ大規模分子に対しても高速に行うことができる。シミュレーションの際に大きな計算コストのかかるエネルギー 2 次微分行列 (Hessian) の計算などについても、筆者らが開発した並列計算アルゴリズムを用いて実装されている。

次に、シミュレーション技術の生体分子への適用として、構築したシステムを用いて、核酸ミニヘアピン分子と核酸タンパク質複合体の構造モデリング、自由エネルギー解析、ダイナミクス解析を行った。核酸ミニヘアピン分子については、Two-Stage 法によって構造モデリングを行い、さらに Poisson-Boltzmann 分子動力学法と Weighted Histogram Analysis Method (WHAM) によってヘアピン形成自由エネルギープロファイルを求めた (第 II 部)。これにより、生化学実験で明らかになっている配列による熱安定性の違いを、自由エネルギープロファイルを直接比較することで解析することができた。また核酸タンパク質複合体については、ダイナミクス解析の手法の一つである二面角系基準振動モード解析を tRNA-ARS 複合体に適用し、フリー状態と複合体形成状態のダイナミクスとその違いを解析した (第 III 部)。これにより、フリー状態に内在する複合体状態でのモード運動などが明らかになり、tRNA-ARS の精密な分子認識へダイナミクスがどのように関わっているかを解析するための糸口を与えた。

本論文は以上のような 3 部で構成されている。以下にそれぞれの概要を述べる。

第 I 部では、生体分子の構造モデリング・ダイナミクス解析システムの構築について述べる。大規模な分子に対して精度の高い解析を行うためには膨大な計算コストがかかるため、並列計算技術を用いることが必須である。本研究では、構造モデリング、自由エネルギー解析、ダイナミクス解析の際に必要なさまざまな手法を統合的に利用できるようにするとともに、すべてに並列計算技術を適用した、構造モデリング・ダイナミクス解析システムを構築した<sup>14)</sup>。各手法は、それ自体が計算量を

軽減するように工夫され、また並列処理が可能な部分には並列計算アルゴリズムが用いられており、アルゴリズムの工夫と並列計算の両面から計算時間の短縮が図られている。

構造モデリングには、筆者らが開発したエネルギー最小化をベースにした Two-Stage 法および MCM 法、MCMS 法などの構造探索手法が用いられている。自由エネルギー解析は、筆者らのグループが開発した Poisson-Boltzmann 分子動力学法<sup>15,16)</sup> を用いて構造サンプリングを行い、複数のシミュレーション結果を信頼度が最大になるように統合することができる WHAM によって自由エネルギープロファイルを求める。ダイナミクス解析には、二面角系基準振動モード解析法を用いる。また、シミュレーションの際にしばしば計算が必要となる行列計算、エネルギー計算、エネルギーの 1 次微分 (Gradient)・2 次微分 (Hessian) 計算にはいずれも並列アルゴリズムが用いられ、大規模な分子についても高速に計算が行えるようになっている。

Hessian は生体分子のエネルギー関数の 2 次微分行列のことであり、エネルギー関数の形状情報を与える重要な量である。構造サンプリング、エネルギー極小化、ダイナミクス解析など、さまざまな解析アルゴリズムにおいて繰り返し Hessian の計算が必要となることが多い。しかし Hessian の計算には大きな計算コストがかかるため、とくに大規模な分子の計算では高速解法が必要である。そこで筆者らは、Unify 方式と呼ばれる Hessian の並列計算アルゴリズムを新たに開発した<sup>6)</sup>。このアルゴリズムを用いることで、Hessian 計算およびアルゴリズム内に Hessian 計算を含むニュートン・ラフソン法によるエネルギー極小化計算において 6 ~ 12 倍程度の並列化効率を達成することができた。

また Unify 方式では、とくに小さい系において、プロセッサ数の増加に伴って通信コストが原因で性能低下がみられるケースがあることがわかった。そこで、これを改良したサブタスクグループ方式を開発した<sup>17)</sup>。これは Hessian 計算を複数の部分問題 (サブタスク) に分割し、サブタスク間の依存関係を解析することで通信量を必要最小限に抑えるものである。さらに遺伝的アルゴリズムを用いて、プロセッサの負荷分散と通信量の低減を同時に達成できるように、サブタスクのプロセッサ割り当てを最適化する。このサブタスクグループ方式を Hessian 計算に適用したところ、プロセッサ数 60 で 32.6 倍のスピードアップを達成し、Unify 方式からさらに大きく性能を改善することができた。

システムは標準的なプログラミング言語と並列通信ライブラリを用いてプログラミングされており、分散メモリ型並列コンピュータ、共有メモリ型並列コンピュータ、ワークステーションクラスタなどのプラットフォームを問わず高い性能を発揮することができる。これを用いることで、大規模分子の構造モデリングやダイナミク

ス解析を高速に行うことが可能となった。

次にこのシステムを用いて、2種類の分子を対象に解析を行った。1つ目は核酸ミニヘアピン分子の熱安定性解析、2つ目は核酸タンパク質複合体の基準振動モード解析である。

第 II 部では、核酸ミニヘアピン分子の熱安定性解析を行った結果について述べる。DNA ミニヘアピン分子 dGC(GAA)GC は異常に高い熱安定性を示すことが知られており<sup>18)</sup>、ループ部の塩基をシステマティックに置換した生化学実験から、3番目と5番目の塩基がそれぞれ G, A であることが高い熱安定性の必要十分条件であることが明らかになっている<sup>19)</sup>。そこで、高い熱安定性を示す配列とそうでない配列を与えたときに、モデリングした構造の相違およびループ部塩基の相互作用の解析によって両者を区別できるかどうかを検証した。

この DNA ミニヘアピン分子の立体構造は、ループ部が GAA のものが NMR で解かれている<sup>20)</sup>。そこで、まずループ部配列が GAA, GGA, GCA, GTA, GAG, GAC の6つについて、筆者らが開発した Two-Stage 法<sup>21,22)</sup> と呼ばれる方法で構造モデリングを行った。この方法はエネルギー最小化を用いることで、塩基対を作る塩基などの事前知識なしに構造モデリングを行うことが可能であるという特長をもつ。Two-Stage 法によって得られた GAA のモデリング構造は、NMR で明らかになっている構造に極めて近く、Two-Stage 法の有効性を確認できた。また高い熱安定性を示す GGA, GCA, GTA についても、NMR 解析結果と矛盾しないモデリング構造を構築することができた。一方、高い熱安定性を示さない GAG, GAC についても、モデリング構造はヘアピン構造となった。またループ部の塩基間相互作用を解析したところ、熱安定なヘアピンを形成する配列とそうでない配列との間に大きな差はみられなかった<sup>23)</sup>。このことは、これらの分子の熱安定性の理解には、エネルギー（エンタルピー）だけでなく、エントロピーの効果も考慮することが必要であることを示唆する。

そこで、ループ部配列が GAA と GAG のものについて、筆者と同じ研究グループの Tazaki らが開発した Poisson-Boltzmann 分子動力学法<sup>15,16)</sup> を用いて、多数の構造サンプリングを行い、得られたデータを Weighted Histogram Analysis Method (WHAM)<sup>24,25)</sup> で統合して、ヘアピン形成自由エネルギープロファイルを求めた。その結果、安定なヘアピンを形成する GAA では、ヘアピン構造付近が自由エネルギーの谷となり、ヘアピンが open 構造に近づくにしたがって自由エネルギーは増加した。一方 GAG では、ヘアピン構造付近に自由エネルギーの谷はあるものの、ヘアピンが崩れた構造付近に自由エネルギーのより深い谷が存在していた。従来の自由エネルギー計算では安定な構造（今の場合はヘアピン構造）同士の自由エネルギー差をみることが

ほとんどであった。本研究では、対象とした DNA ミニヘアピン分子においては、分子動力学計算を用いた自由エネルギー解析によって求めたヘアピン形成自由エネルギープロファイルを直接比較することで熱安定性の配列依存性を議論できることを明らかにした<sup>26)</sup>。

第 III 部では、核酸タンパク質複合体であるトランスファー RNA (tRNA) とアミノアシル tRNA 合成酵素 (ARS) の複合体に基準振動モード解析を適用した結果について述べる。ARS は似たような立体構造をもつ多数の tRNA の中から、自分が担当する種類のものだけを非常に正確に認識し、アミノ酸を付加することができる。tRNA 中のさまざまな部位を塩基置換した実験で、ARS との接触面から離れた部位を置換した場合もアミノアシル化効率や ARS との結合定数に影響をおよぼすことが明らかになっており、この分子認識に tRNA のダイナミクスが関与している可能性が指摘されている。そこでこの分子認識機構の中でダイナミクスの果たしている役割を理解するための手がかりとして、ダイナミクス解析手法の一つである基準振動モード解析をフリー状態の tRNA、ARS および tRNA-ARS 複合体に適用して、フリー状態と複合体形成状態のダイナミクスの違いを解析した。

基準振動モード解析によって求めた各原子の熱ゆらぎの大きさは、X 線結晶解析から得られる温度因子データと高い相関を示し、本研究による解析の妥当性を示した。フリー状態の tRNA のダイナミクス解析では、tRNA 分子が球状タンパク質の  $1/4 \sim 1/10$  の低いモード周波数でゆっくりとした大きな運動をしていることが明らかになった<sup>27-30)</sup>。低周波数領域では、tRNA 分子はいくつかの領域に分かれて運動をしていた。一体となって運動する領域は計算対象とした 2 種類の tRNA (tRNA<sup>Phe</sup> と tRNA<sup>Asp</sup>) で異なっており、よく似た L 字型立体構造をとる両分子のダイナミクスには微妙な違いがあることが明らかになった。

また、Gln 系の tRNA (tRNA<sup>Gln</sup>) - ARS (GlnRS) 複合体について、フリー状態のモード運動と複合体状態のモード運動の詳細な解析を試みた<sup>31-35)</sup>。複合体形成によって、tRNA<sup>Gln</sup> と GlnRS が接触している広い範囲で原子の熱ゆらぎの大きな減少がみられた。これに対して、tRNA<sup>Gln</sup> の T アーム、D ループ、エキストラループ、GlnRS のアンチコドン認識ドメインなどでは、ゆらぎの減少の割合は比較的小さかった。低周波数領域では、tRNA<sup>Gln</sup> のアクセプタステム、アンチコドンアーム、中央部分の各領域と、GlnRS の触媒ドメイン、アンチコドン認識ドメインなどの各ドメインが、カップリングして大きく協調的に運動していた。これは球状タンパク質同士の低周波数領域のモード運動とは大きく異なる特徴的なものであった。また低周波数領域の主要な運動は、GlnRS はドメイン同士の相対運動であり、tRNA<sup>Gln</sup> は二重らせん部分のゆっくりとした変形であるという、タンパク質と核酸のダイナミク

スの特徴の違いもみられた。

複合体状態とフリー状態でのモード運動を相互に比較すると、GlnRS は複合体形成によっても低周波数領域でのモード運動の保存性が高く、複合体形成によってそのダイナミクスは大きな影響を受けていないことがわかった。tRNA<sup>Gln</sup> では、フリー状態のダイナミクスに、いくつかの複合体状態のモード運動が内在していることが明らかになった。また、複合体状態とフリー状態に対応するモードの周波数の大小関係は保存しておらず、tRNA<sup>Gln</sup> のダイナミクスは複合体形成によって大きな影響を受けていた。

tRNA-ARS の分子認識に tRNA のダイナミクスが関わっているとすれば、これらの複合体形成の前後で保存しているモード運動は認識に重要な役割を果たす可能性がある。本研究をさらにすすめて、多数のアミノ酸系について同様の解析を行い、その結果を比較することで、これまで両分子の接触部位周辺の静的な解析が中心だった tRNA-ARS の分子認識メカニズムの解析に、ダイナミクスの観点からの新たな知見をもたらす可能性があると思われる。

なお本論文では、章番号は全体を通して付けられている。

## 本論文で用いる略称

本論文では、以下の略称を用いている。

<b>r.m.s.d.</b>	最小 2 乗平均誤差 (root mean square deviation)
<b>PDB</b>	Protein Data Bank <sup>36,37</sup> )
<b>NMR</b>	核磁気共鳴 (nuclear magnetic resonance)
<b>tRNA</b>	トランスファー RNA (transfer RNA)
<b>ARS</b>	アミノアシル tRNA 合成酵素 (aminoacyl-tRNA synthetase)
<b>MD</b>	分子動力学法 (molecular dynamics simulation)
<b>DEM</b>	拡散方程式法 (Diffusion Equation Method)
<b>PBMD</b>	Poisson-Boltzmann 分子動力学法 (Poisson-Boltzmann Molecular Dynamics Method)
<b>FDPB</b>	有限差分 Poisson-Boltzmann 法 (Finite Difference Poisson-Boltzmann Method)
<b>WHAM</b>	Weighted Histogram Analysis Method
<b>RISM</b>	Reference Interaction Site Model
<b>MCM</b>	Monte Carlo Minimization
<b>MCMS</b>	Monte Carlo Minimization with Saturated Non-bonded Energy

また、分子グラフィックスの表示には、RasMol<sup>38</sup>) および LIVE<sup>39</sup>) を用いた。





## 第 I 部

# 生体分子の構造モデリング・ダイナミクス 解析システムの構築



## 第 2 章

### 第 I 部の概要

第 I 部では、筆者らが開発した生体分子の構造モデリング・ダイナミクス解析システムの概要について述べる。

まず分子シミュレーションの基礎となる配座エネルギーの概念を、本研究で用いている AMBER の関数系を例に述べる。次に分子の配座 (コンフォメーション) を記述する座標系として、XYZ 系と二面角系について説明する。

次に、開発した生体分子の二面角系構造モデリング・ダイナミクス解析システムの構成について述べる。

また、モデリングやダイナミクスの解析を行う際には、エネルギーの微分の計算が必要となることが多い。とくにエネルギーの 2 次微分行列 (Hessian) は、エネルギーそのものや 1 次微分に比べて計算コストが大きく、大規模分子の計算では全体の計算時間の大きな部分を占める。これを解決するために開発した、Hessian を高速に計算可能な並列計算アルゴリズムについて述べる。

## 第 3 章

### 配座エネルギー関数と座標系

#### 3.1 AMBER の配座エネルギー関数

生体分子のシミュレーションを行う際にもっとも重要な概念となるのは、分子の配座エネルギー（コンフォメーションエネルギー）である。分子や周辺環境（溶媒など）を構成する個々の原子の力学的な運動は、系のエネルギーによって規定される。エネルギーは、運動エネルギー  $K$  と配座エネルギー  $E$  の和で表されるが、前者の表現はきわめて簡単であるので、実際の分子シミュレーションに際しては、分子の配座エネルギーの決定が最初の課題となる。

分子を構成する各原子の座標を  $x$  とすると、配座エネルギーは  $E(x)$  と表すことができる。分子の中の各原子の配置、すなわちその分子の立体構造が決まると、エネルギー値 1 つが決まる関数である (Fig. 3.1)。分子が刻一刻と形を変えていくと、それに伴ってエネルギー値が変化してゆく。またエネルギー関数の微分は、エネルギー曲面の形状に関する情報を表し、分子のダイナミクスを解析するときの大きな手がかりとなる。

原理的には、配座エネルギー関数  $E(x)$  は量子力学の Schrödinger 方程式を解くことによって求められる。しかし生体分子のような多数の原子から成る複雑な系では、Schrödinger 方程式の解を解析的に求めることはできない。このため、近似的かつ経験的な  $E(x)$  の関数系が開発されている。分子シミュレーションにおいては、エネルギー関数は、大規模な分子でも計算が容易である程度に簡単な式で表現されていること、エネルギー極小化によって正しい構造が得られるように天然 (Native) 構造のエネルギーが低くなっていることなどが要請される。これらを満たす配座エネルギー関数系として、ECEPP<sup>40)</sup>、CHARMM<sup>41)</sup>、AMBER<sup>42,43)</sup> などが広く用いられている。これらの配座エネルギー関数と各種パラメータのセットは、力場 (フォースフィールド) と呼ばれ、生化学実験から求められるさまざまな物理量を精度よく再現できるように設計されている。

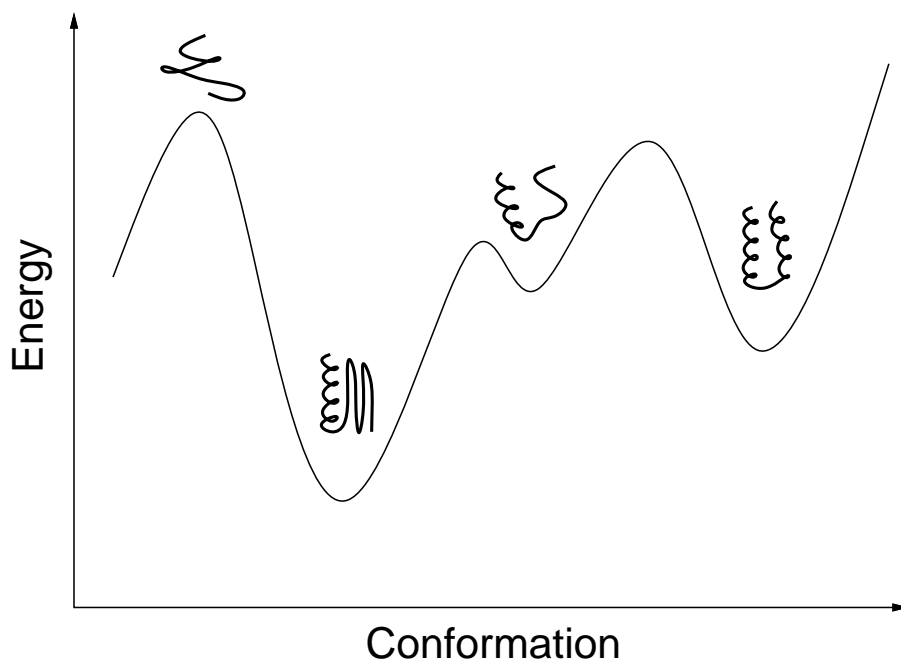


Figure 3.1 配座エネルギー関数の概念図．分子の1つの配座（コンフォメーション）に1つのエネルギー値が対応する．

ここでは、本研究で用いている AMBER のエネルギー関数についてその概略を説明する．AMBER のエネルギー関数は、バージョンによってパラメータの値や水素結合の取り扱いなどが若干異なるが、基本的な枠組みは 1984 年の開発当初と変わっていない．また CHARMM や ECEPP も、座標系やパラメータなどが多少異なっているものの、配座エネルギーを 2 原子間相互作用の和で表すという基本概念は同じである．

AMBER では、各原子を「原子種」に分類し、それに応じてパラメータを設定する．原子種は、たとえば同じ炭素原子でも、「カルボニル基の  $sp^2$  炭素」「6 員環の芳香族の  $sp^2$  炭素」「4 つの置換基をもつ  $sp^3$  炭素」などと細かく分類され、それぞれパラメータが決められている．

AMBER では、配座エネルギー  $E_{total}$  は、以下の複数の 2 原子間相互作用の和で記述される．

$$E_{total} = E_{len} + E_{ang} + E_{tor} + E_{vdw} + E_{el} + E_{hb} \quad (3.1)$$

ただし、

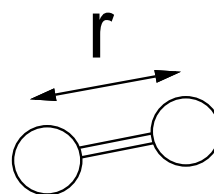
$E_{len}$	=	共有結合ポテンシャル
$E_{ang}$	=	結合角ポテンシャル
$E_{tor}$	=	二面角ポテンシャル
$E_{vdw}$	=	van der Waals ポテンシャル
$E_{el}$	=	静電ポテンシャル
$E_{hb}$	=	水素結合ポテンシャル

である．以下に，それぞれの表式を示す．

### 共有結合ポテンシャル

$$E_{len} = \sum_{i=1}^{N_r} k_r (r_i - r_0)^2 \quad (3.2)$$

$N_r$	=	2 原子間の共有結合の数
$r_i$	=	2 原子間の結合長 ( $\text{\AA}$ )
$r_0$	=	標準結合長 ( $\text{\AA}$ )
$k_r$	=	結合伸縮力の定数 ( $\text{kcal/mol}\cdot\text{\AA}^2$ )

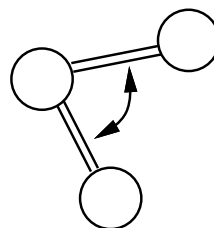


共有結合ポテンシャルは，共有結合している 2 原子間の結合長の，標準結合長からのずれの 2 乗に比例するバネ型の関数形をとる．パラメータ  $k_r$  は，2 原子の原子種の組み合わせごとに決められている．

### 結合角ポテンシャル

$$E_{ang} = \sum_{i=1}^{N_\theta} k_\theta (\theta_i - \theta_0)^2 \quad (3.3)$$

$N_\theta$	=	3 原子で形成される結合角の数
$\theta_i$	=	3 原子で形成される結合角 (deg)
$\theta_0$	=	標準結合角 (deg)
$k_\theta$	=	変角力の定数 ( $\text{kcal/mol}\cdot\text{deg}^2$ )

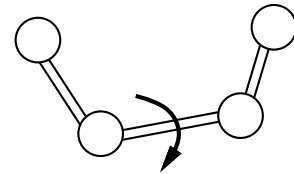


結合角ポテンシャルは，共有結合している 3 原子で形成される結合角の，標準結合角からのずれの 2 乗に比例するバネ型の関数形をとる．パラメータ  $k_\theta$  は，3 原子の原子種の組み合わせごとに決められている．

### 二面角ポテンシャル

$$E_{tor} = \sum_{i=1}^{N_\phi} \frac{k_n}{2} [1 + \cos(n\phi_i - \gamma)] \quad (3.4)$$

- $N_\phi$  = 4 原子で形成される二面角の数
- $\phi_i$  = 4 原子で形成される二面角 (deg)
- $k_n$  = 二面角のねじれ力定数 (kcal/mol·deg<sup>2</sup>)
- $n$  = 周期
- $\gamma$  = 位相 (deg)



二面角ポテンシャルは，共有結合している 4 原子で形成される二面角の拘束を表す． $n$  は二面角まわりの原子の対称性を反映し，たとえば  $sp^3$  型の炭素原子 C が作る結合 X-C-C-X (X は任意の原子種) の場合，3 回対称であるので  $n = 3$  となる．パラメータ  $k_n$ ， $n$ ， $\gamma$  は，4 原子の原子種の組み合わせごとに決められている．

### van der Waals ポテンシャル

$$E_{vdw} = \sum_{j=1}^{N_{atom}} \sum_{i>j}^{N_{atom}} \epsilon_{ij}^* \left[ \left( \frac{R_{ij}^*}{R_{ij}} \right)^{12} - \left( \frac{R_{ij}^*}{R_{ij}} \right)^6 \right] \quad (3.5)$$

- $N_{atom}$  = 系に含まれる原子数
- $\epsilon_{ij}^*$  = van der Waals 力の定数 (kcal/mol)
- $R_{ij}$  =  $i$  番目と  $j$  番目の原子間の距離 (Å)
- $R_{ij}^*$  =  $i$  番目と  $j$  番目の原子の van der Waals 半径の合計 (Å)

ただし， $\epsilon_{ij}^* = \sqrt{\epsilon_i^* \epsilon_j^*}$ ， $R_{ij}^* = R_i^* + R_j^*$  である． $\epsilon_i^*$  および  $R_i^*$  は，原子  $i$  の原子種ごとに決められている．

van der Waals ポテンシャルは，6-12 型の Lennard-Jones ポテンシャルと呼ばれる関数型が用いられている．これは，パウリの排他律に由来する原子間の衝突を避ける斥力と，原子内の双極子に起因する配向力・誘起力・分散力と呼ばれる引力を合わせたものである．前者はきわめて近距離でのみ有効であり，原子間距離  $r^{-12}$  に

比例する形となっている．後者はかなり遠距離まではたつき， $r^{-6}$  に比例する形となっている．和は，1-4 相互作用以上離れた原子ペアに関してとられる．ただし 1-4 相互作用とは，ある原子と，共有結合に沿った 3 つ先の原子（二面角を形成する端の原子同士）の関係のことである (Fig. 3.2)．また AMBER では，1-4 相互作用している原子ペアについては，斥力項の補正のために上記の式で計算された値を 2.0 で割ったものとする．

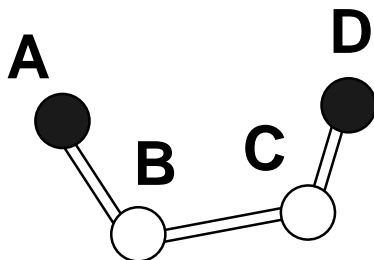


Figure 3.2 1-4 相互作用の概念図．原子 A と原子 D が 1-4 相互作用の関係にある．

### 静電ポテンシャル

$$E_{el} = \sum_{j=1}^{N_{atom}} \sum_{i>j}^{N_{atom}} \frac{q_i q_j}{\epsilon R_{ij}} \quad (3.6)$$

$N_{atom}$  = 系に含まれる原子数

$q_i$  =  $i$  番目の原子の電荷

$R_{ij}$  =  $i$  番目と  $j$  番目の原子の距離 (Å)

$\epsilon$  = 誘電率

静電ポテンシャルは，電荷  $q$  をもつ各原子間のクーロンポテンシャルの和で表される．AMBER では，各原子の中心に点電荷が割り当てられる．和は，1-4 相互作用以上離れた原子ペアに関してとられる．また AMBER では，1-4 相互作用している原子ペアについては，補正のために上記の式で計算された値を 2.0 (parm91<sup>42,43</sup>) 以前のバージョン)，または 1.2 (parm94<sup>44</sup>) 以降のバージョン) で割ったものとする．

誘電率  $\epsilon$  は巨視的な量であり，真空中や一様な物質の中以外では位置の関数となる．しかしこの効果をエネルギー関数に取り込むのは困難であるので，いくつかの近似式が提案されている．水分子などの溶媒を陽に取り入れない計算の場合，溶媒による静電遮蔽効果を原子間の距離の 1 次式

$$\epsilon = aR_{ij} \quad (3.7)$$



あるいは、シグモイダル関数

$$\varepsilon = D - \frac{D-1}{2} \exp(-sR_{ij})[(sR_{ij})^2 + 2sR_{ij} + 2] \quad (3.8)$$

として取り込むことが多い (Fig. 3.3)<sup>45,46</sup>。ただし  $a, D, s$  はユーザが定義するパラメータで、 $a = 4.0, D = 79.0, s = 0.3$  などがよく用いられる。溶媒分子を陽に取り入れる計算の場合は、 $\varepsilon = 1.0$  とする必要がある。

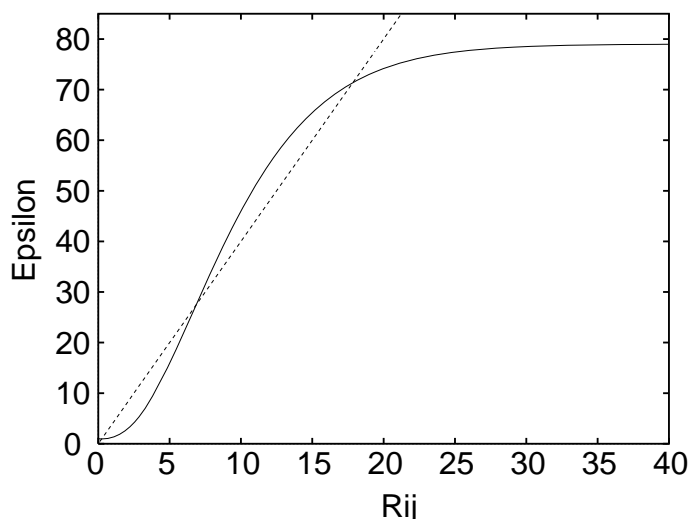


Figure 3.3 距離依存の誘電率  $\varepsilon$  の関数系の例．実線はシグモイダル関数 ( $D = 79.0, s = 0.3$ ) の例．破線は距離の 1 次式 ( $a = 4.0$ ) の例．

### 水素結合ポテンシャル

$$E_{hb} = \sum_{j=1}^{N_{hb}} \sum_{i>j}^{N_{hb}} \left( \frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right) \quad (3.9)$$

$N_{hb}$  = 水素結合している原子数

$R_{ij}$  =  $i$  番目と  $j$  番目の原子の距離 (Å)

$C_{ij}, D_{ij}$  = 水素結合パラメータ

水素結合ポテンシャルは、水素結合している 2 原子間の、van der Waals ポテンシャルと静電ポテンシャルの和を補正する目的で導入されている。水素結合している 2 原子間距離は、それぞれの van der Waals 半径の和よりもさらに近接する。このため、水素結合している場合は、van der Waals ポテンシャルの代わりに水素結合ポテンシャルを用いる。なお、AMBER の parm94 以降のフォースフィールドでは、この項は廃止されている。

### 3.2 XYZ系と二面角系

前節で述べた配座エネルギーは、分子の立体構造を変数とする関数となっている。分子を構成している各原子の座標を記述するにあたっては、以下の XYZ 系および二面角系と呼ばれる 2 種類の記述方法 (座標系) がよく用いられる。両者の概要を述べる。

#### XYZ系

XYZ 系は、分子を構成する各原子の直交座標系での位置  $(x, y, z)$  の組で、分子のコンフォメーションを記述するものである ( Fig. 3.4 )。分子に含まれる原子数を  $N$  とすると、自由度の数は  $3N$  個となる。XYZ 系では、原子それぞれの位置のみを記述するため、どの原子とどの原子が共有結合しているか、という概念はコンフォメーションの記述の中には現われない。このため分子内のすべての自由度を表現しうるが、逆に 1 つのコンフォメーションを記述するために必要な自由度の数が多いという欠点も併せもつ。

この座標系では、運動エネルギー  $K$  が、

$$K = \frac{1}{2} \sum_i m_i (\dot{x}_i^2 + \dot{y}_i^2 + \dot{z}_i^2) \quad (3.10)$$

と原子  $i$  の質量  $m_i$  と座標  $(x_i, y_i, z_i)$  の時間微分を用いて完全に表現することができるため、運動方程式を容易に表現することができる。このため分子動力学計算では、ほとんどの場合 XYZ 系が用いられる。

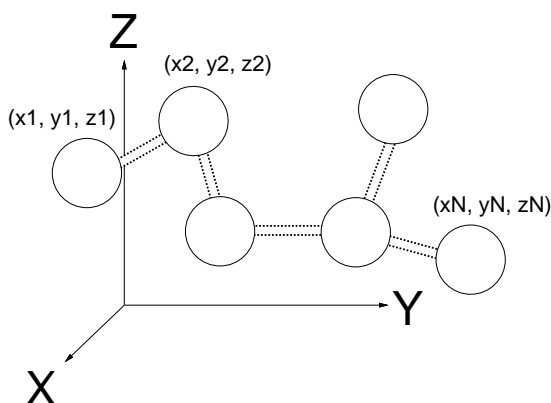


Figure 3.4 XYZ 系の概念図。

## 二面角系

二面角系は、共有結合している原子の結合長および結合角がコンフォメーションが変わっても一定であるとみなし、指定した共有結合（二面角）まわりの回転のみを自由度とするものである。この根拠は、結合長の伸縮運動と結合角の変角振動の時間オーダー（ $\sim 10^{-13}$  秒周期）が、共有結合まわりの回転の時間オーダー（ $\sim 10^{-10}$  秒周期）に比べてはるかに短いという点にある<sup>47)</sup>。

分子のコンフォメーションは、回転の自由度を与えられた二面角の値の組で記述される (Fig. 3.5)。分子は一体となって運動する「ユニット」と回転する二面角である「ボンド」から構成されると考える。これにより、1分子あたりの自由度の数は、XYZ系の $3 \times (\text{原子数})$ から、 $5 \sim 6 \times (\text{残基数})$ 程度にまで減少させることができる。このため、大きな生体分子をより少ない計算量で取り扱うことが可能となる。一方で、タンパク質-水系などの複数の分子の取り扱いや、運動方程式が複雑になるために分子動力学計算は困難になるという欠点がある。二面角系は基準振動モード解析法などでよく用いられる。

核酸分子は、塩基の部分が環構造になっているので1つのユニットとして扱うことが可能である。このためXYZ系に比べて自由度を大幅に減らせるという点で、二面角系が有効となる。本研究では、計算の大部分を二面角系で行っている。

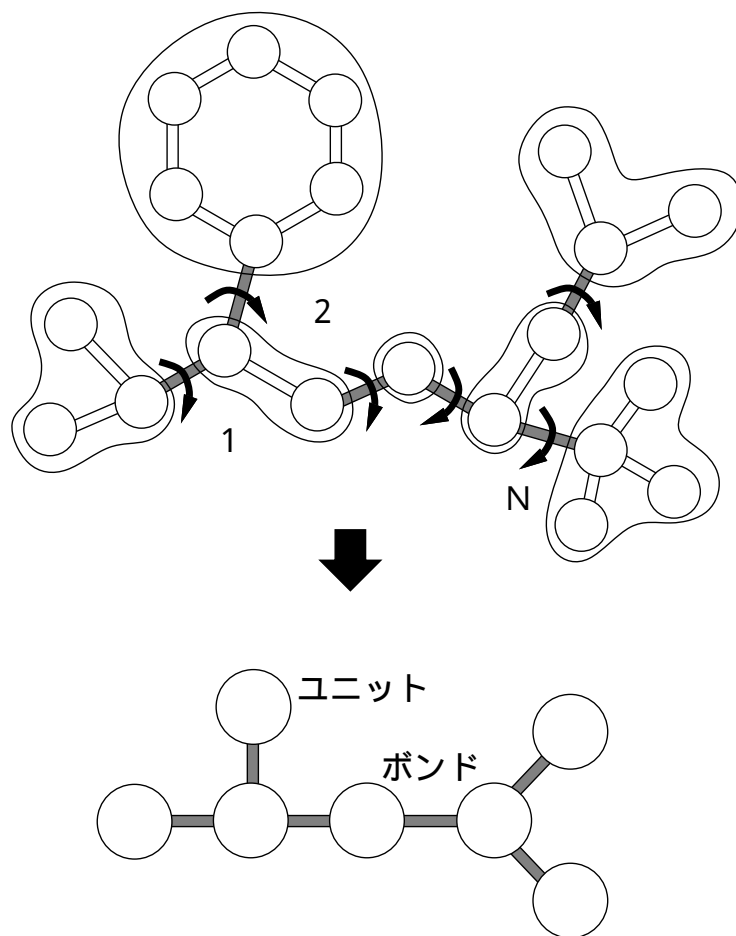


Figure 3.5 二面角系におけるコンフォメーション記述方法と，ボンドとユニットの概念図．一体となって運動する原子団は1つの「ユニット」として表される．ユニットとユニットをつなぐ回転する二面角を「ボンド」と呼ぶ．

## 第 4 章

### 構造モデリング・ダイナミクス解析システムの構築

核酸タンパク質複合体やタンパク質同士の複合体などの巨大な系を対象としてシミュレーションを行う際には、膨大な計算コストが必要となる。これを解決するためには、アルゴリズム自体を工夫することで計算量を軽減する方法と、複数のコンピュータを同時に用いて並列に計算を実行することで、1 台のコンピュータのみを用いた場合よりも大幅に計算時間を短縮する方法の主に 2 つが挙げられる。

本研究では、構造モデリング、自由エネルギー解析、ダイナミクス解析の際に必要なさまざまな手法を並列計算技術を用いて統合した解析システム NORMA (Normal mode and Molecular mechanics Analysis system) を構築した。各手法は、それぞれが計算量を軽減するように工夫され、また並列処理が可能な部分には並列計算アルゴリズムが用いられており、上記の両面から計算時間を短縮する工夫がなされている。

構造モデリングについては、エネルギー最小化による構造モデリング手法「Two-Stage 法」を開発し<sup>21)</sup>、実装した。この方法は、エネルギー関数を平滑化することでエネルギーのローカルミニマムにとらわれることなくグローバルミニマムを探索することができる。Two-Stage 法については、第 II 部で詳しく述べる。また Monte Carlo 法とエネルギー極小化を組み合わせた MCM (Monte Carlo Minimization) 法、さらに MCM 法において van der Waals ポテンシャルと静電ポテンシャルを原子間距離の近いところで平滑化することで構造探索能力を高めた MCMS (Monte Carlo Minimization with Saturated Non-bonded Energy) 法が実装されている。

自由エネルギー解析については、構造サンプリングに Poisson-Boltzmann 分子動力学法<sup>15,16)</sup>を用いている。これは筆者らのグループが開発したもので、溶媒を連続体近似し、静電遮蔽効果を Poisson-Boltzmann 方程式を用いて求めることにより、効率のよい構造サンプリングを可能にしたものである。また複数の分子動力学計算から信頼度が最大になるように系の状態密度を求めることのできる WHAM 法を用

いて、自由エネルギープロファイルの解析を行う。これらの手法については、第 II 部で詳しく述べる。

ダイナミクス解析については、二面角系基準振動モード解析法の並列アルゴリズムが実装されている。基準振動モード解析法は、生体分子のダイナミクスを少ない計算コストで解析することができる方法である。また系を二面角系で取り扱うことで、とくに環を多く含む核酸分子の場合に必要な計算量を大きく軽減することができる。二面角系基準振動モード解析法については、第 III 部で詳しく述べる。

これらの計算の際には、連立 1 次方程式、固有値問題、一般化固有値問題、行列のかけ算などの行列演算を計算することが度々必要となる。これらの大規模な行列の並列計算は、ScaLAPACK (Scalable LAPACK) ライブラリ<sup>48-51</sup>) を用いて実装されている。

また分子のコンフォメーションに対するエネルギー値とその 1 次微分、2 次微分の計算が必要となる。NORMA システムでは、各プロセッサで担当する原子ペアがなるべく均等になるように並列化されたエネルギー計算法を用いている。とくにエネルギー極小化や基準振動モード解析の際に計算コストの大きな部分を占めるエネルギー 2 次微分行列 (Hessian) については、「Unify 方式」およびそれをさらに改良した「サブタスクグループ方式」と呼ばれる並列計算アルゴリズムを開発し、実装した。これについては、次章で詳しく述べる。なお、1 次微分 (Gradient) の計算については、Hessian の Unify 方式と同様の方法で並列化されている。

以下に、NORMA システムの概要を示す。システムは、1 つのマスターモジュールと、配座エネルギー計算などの各種演算を行う複数のモジュールから構成されている。これは、各モジュールでの計算を並列に行うことにより高速化を図るためと、計算に必要な作業領域を分散配置することによって、限られたメモリなどの計算機資源の範囲内でできるだけ大きな系を扱うことを可能にするためである。マスターモジュールには 1 つのプロセッサ (マスタープロセッサ) が割り当てられ、スレーブモジュールには 1 つまたは複数のプロセッサを割り当てることができる。各スレーブモジュールに割り当てられるプロセッサの台数は、ユーザが自由に指定することができる。また計算機資源が十分にあれば、1 つのプロセッサに複数のモジュールを重ねて割り当てられることも可能である。このシステムのプロットを Fig. 4.1 に示す。

プログラムは広く用いられている C 言語によって記述され、通信には汎用的な MPI (Message Passing Interface) ライブラリを用いている。このため、分散メモリ型並列コンピュータ、共有メモリ型並列コンピュータなど、さまざまなプラットフォームで動作可能であり、これまで、ワークステーションクラスタ (HP C180, Sun ULTRA 5, Compaq DS20 など)、分散メモリ型並列コンピュータ日立 SR2201, IBM SP/2,

共有メモリ型並列コンピュータ Sun Ultra Enterprise などでの動作を確認している。

このシステムを用いることにより，大規模分子の二面角系での構造モデリング・ダイナミクス解析が可能となった．第 II 部の核酸ミニヘアピン分子のモデリング，第 III 部の核酸タンパク質複合体の基準振動モード解析は，このシステムを用いて解析を行っている．

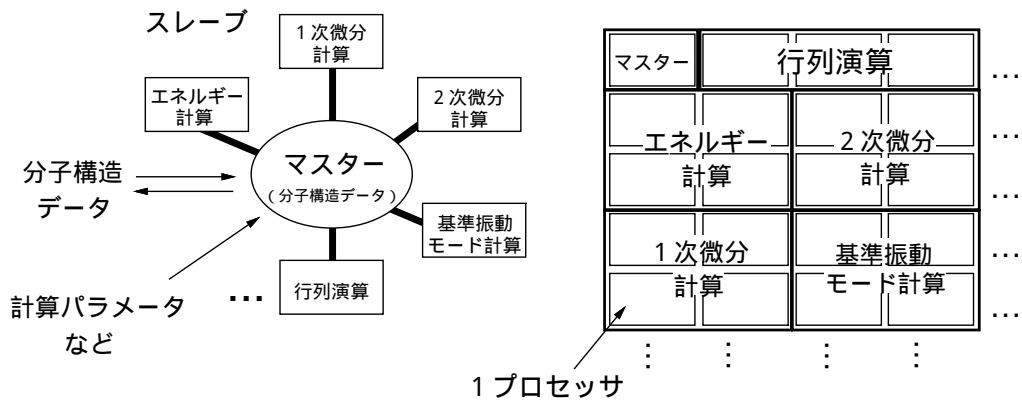


Figure 4.1 システムにおけるマスターモジュールとスレーブモジュールの概念図．

## 第 5 章

### 二面角系配座エネルギー 2 次微分行列の並列計算法の開発

配座エネルギーの 1 次微分ベクトル (Gradient) と 2 次微分行列 (Hessian) は、生体分子のシミュレーションを行う際に、しばしば計算が必要となる。Gradient はエネルギー曲面の勾配を表し、Hessian は上に凸か下に凸かというような Gradient の変化の様子を表す。

Noguti らは、タンパク質分子のエネルギー曲面の形状を Hessian から見積り、分子が柔らかい方向を優先的に探索することで Monte Carlo 法のサンプリング効率を 50~500 倍に改善する方法を開発した<sup>52)</sup>。分子の配座エネルギー極小化でよく用いられるニュートン・ラフソン法 (Newton-Raphson Method) では、Hessian と Gradient を用いてエネルギーの 1 次微分がゼロとなる方向を算出する。筆者らが開発したエネルギー最小化法である Two-Stage 法 (第 II 部で述べる) でも、局所探索の過程で Hessian を用いている。また、ダイナミクスの解析手法である基準振動モード解析 (第 III 部で述べる) では、まず Hessian の固有値がすべて正となるまでエネルギー極小化を行い、続いて Hessian を含む一般化固有値問題を解くことで基準振動モードを求める。

これらのシミュレーションでは、アルゴリズム中に Hessian を何度も計算する必要があるが、エネルギーの微分、とくに Hessian の計算には、エネルギー値そのものや Gradient に比べて大きな計算コストがかかる。このため、Hessian を高速に計算することはシミュレーション全体の高速化につながり、重要な意義をもつ。

本研究では、主に二面角系で分子のコンフォメーションを記述するので、以下は二面角系について述べる。二面角系では、たとえばベンゼン環などのように、タンパク質の主鎖などに比べるとほとんど形を変えずに一体となって運動する部分を、内部変形を伴わない剛体 (ユニット) として取り扱うことができるため、系全体の自由度を XYZ 系に比べて大幅に減らし、計算量を減少させることができる。XYZ 系では各原子の座標が独立しているために、エネルギー関数をそれぞれの原子の X, Y,



Z で微分することで容易に 1 次微分, 2 次微分が求まるのに対して, 二面角系では一体となって運動している原子同士の相関関係を考慮しなければならないので, その微分の計算は自明ではない. 通常の計算では Hessian の計算には自由度  $n$  の 4 乗の計算コストがかかるため, とくに大規模な分子になるとその計算に膨大な時間を要することになる. Gō らは, この二面角系 Hessian の高速計算法に取り組み, 計算コストを  $n^4$  から  $n^2$  に減少させるアルゴリズムを開発した<sup>53-57</sup>). しかしこのアルゴリズムを用いても, 第 III 部で扱う核酸タンパク質複合体のような大規模な分子の場合は, やはり膨大な計算時間がかかってしまう.

一方で, 近年のコンピュータ・サイエンスの進展によって, 並列計算技術が急速に発達している. 並列計算技術とは, 1 つの作業を複数の独立な作業単位に分割できる場合に, それらを複数のコンピュータで分担して行うことで, 1 台あたりの計算コストを軽減し, 作業全体の計算時間を短縮させるという概念である. 生体分子のシミュレーションにおいては, たとえば前章で述べた AMBER のエネルギー関数は, 原子ペア間のポテンシャルの和で記述されており, 1 つ 1 つの原子ペアに対する計算は互いに独立である. このため, 配座エネルギーに関する計算は並列計算に適した問題であるといえる. 並列計算アルゴリズムをうまく適用することで, 大規模な分子に対する計算を高速に行うことが可能となる.

二面角系あるいは XYZ 系での生体分子の配座エネルギーとその 1 次微分の並列計算方法は, 1990 年代の始め頃からさまざまなものが開発されており<sup>58</sup>), 現在ではアルゴリズムによる並列計算のほか, ハードウェアを用いた並列化の試みもなされている<sup>59</sup>). 二面角系では, Totrov らが共有メモリ型並列コンピュータを対象に, 配座エネルギー, 1 次微分, 溶媒接触表面積 (solvent accessible surface) などを並列計算するアルゴリズムを開発している<sup>60</sup>). しかし, 二面角系については, より計算コストの高い Hessian の並列計算アルゴリズムは開発されておらず, また分散メモリ型並列コンピュータにおいても高い性能を発揮することができるエネルギーとその微分の計算アルゴリズムは存在しなかった.

以上をふまえて, 第 5 章では, 二面角系 Hessian の並列計算アルゴリズムの開発について述べる. はじめに「Unify 方式」と呼ばれる方法による並列化をグルタミン tRNA (tRNA<sup>Gln</sup>) およびグルタミンイル tRNA 合成酵素 (GlnRS) の 2 つの分子に適用した結果<sup>6</sup>) を示したあと, Unify 方式をさらに改良して性能向上を図った「サブタスクグループ方式」の結果を示す.

## 5.1 二面角系における Hessian 計算法

AMBER 型の配座エネルギー  $E$  は、原子ペア間の距離 ( $r = \{r_{ij}\}$ ) の関数のもの  $E(r)$  と、二面角 ( $\theta = \{\theta_a\}$ ) の関数のもの  $E(\theta)$  に分けられる。

$$E = E(r) + E(\theta) \quad (5.1)$$

ある二面角  $\theta_a$  での  $E$  の 1 次微分は、次のようになる。

$$\frac{\partial E}{\partial \theta_a} = \sum_{i,j} \frac{\partial E(r)}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial \theta_a} + \frac{\partial E(\theta)}{\partial \theta_a} \quad (5.2)$$

右辺の第 2 項は、 $\theta$  の関数の  $\theta_a$  での微分であるので、その計算は容易である。また右辺第 1 項の  $\partial E(r)/\partial r_{ij}$  の部分も、 $r$  の関数の  $r_{ij}$  での微分であるので、容易に計算が可能である。右辺第 1 項の  $\partial r_{ij}/\partial \theta_a$  の部分は、ある二面角  $\theta_a$  がわずかに回転したときに、原子  $i$  と原子  $j$  の間の距離  $r_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$  (ただし  $\mathbf{x}_i, \mathbf{x}_j$  は、それぞれ原子  $i, j$  の XYZ 系での座標を表す位置ベクトル) がどれだけ変化するかということを表す (Fig. 5.1)。

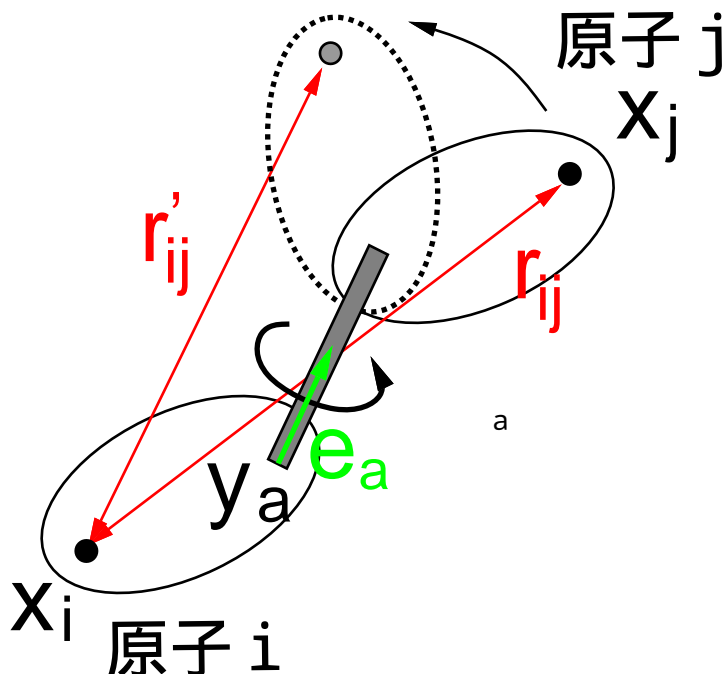


Figure 5.1 二面角  $\theta_a$  が  $\Delta\theta_a$  だけ回転したときの 2 原子間の距離変化 ( $\Delta r_{ij} = r'_{ij} - r_{ij}$ ) を示す概念図。  $\mathbf{x}_i, \mathbf{x}_j$  はそれぞれ原子  $i, j$  の位置ベクトル、  $\mathbf{e}_a$  は二面角  $\theta_a$  の方向を表す単位ベクトル、  $\mathbf{y}_a$  は二面角  $\theta_a$  のボンドの一端の位置ベクトル。

これを式で書くと，以下のようになる．

$$\frac{\partial r_{ij}}{\partial \theta_a} = \begin{cases} \mathbf{q} \cdot \{\mathbf{e}_a \times (\mathbf{x}_i - \mathbf{y}_a)\} & (\text{原子 } i \text{ が原子 } j \text{ の先のユニットにある場合}) \\ -\mathbf{q} \cdot \{\mathbf{e}_a \times (\mathbf{x}_j - \mathbf{y}_a)\} & (\text{原子 } j \text{ が原子 } i \text{ の先のユニットにある場合}) \\ 0 & (\text{原子 } i, j \text{ が同じユニット内にあるとき}) \end{cases} \quad (5.3)$$

ただし， $\mathbf{e}_a$  は二面角  $\theta_a$  の方向を表す単位ベクトル（方向余弦）， $\mathbf{y}_a$  は二面角  $\theta_a$  のボンドの一端の位置ベクトルであり， $\mathbf{q} \equiv (\mathbf{x}_i - \mathbf{x}_j)/r_{ij}$  である．これを用いると，エネルギー  $E(r)$  の二面角  $\theta_a, \theta_b$  での 2 次微分は以下のようにかける．

$$\frac{\partial^2 E(r)}{\partial \theta_a \partial \theta_b} = -S_{ab}(\mathbf{e}_a, \mathbf{e}_a \times \mathbf{y}_a) \mathbf{R}_{ab} \begin{pmatrix} \mathbf{e}_b \\ \mathbf{e}_b \times \mathbf{y}_b \end{pmatrix} \quad (5.4)$$

ただし  $S_{ab}$  は，二面角  $\theta_a$  が二面角  $\theta_b$  の先にある場合は  $-1$ ，それ以外の場合は  $1$  となる． $\mathbf{R}_{ab}$  は  $6 \times 6$  の正方行列で，その要素はボンドペア  $(a, b)$  の外側にあるすべての原子ペア  $(i, j)$  についての，位置ベクトル  $\mathbf{x}_i, \mathbf{x}_j$  およびその間の距離  $r_{ij}$  の関数となっている． $\mathbf{R}_{ab}$  の具体的な表式は複雑であり，詳細については Gō らの原論文<sup>53-56</sup>を参照されたい．

Gō らは，この行列  $\mathbf{R}_{ab}$  の計算を工夫することで，Hessian を自由度の  $n^2$  のコストで解析的に計算する方法を開発した．

このアルゴリズムの説明のために以下の表記法を導入する (Fig. 5.2) ．

- $(a, b)$  : ボンド  $a$  と  $b$  の組 (ボンドペア) ．
- $(a^*, b^*)$  : ボンド  $a$  と  $b$  の外側のユニットの組 ．
- $(a', b')$  : ユニット  $a^*$  と  $b^*$  の外側のボンドの組 ．
- $k(a, b)$  : ボンド  $a$  と  $b$  の間の距離 ．  $a$  と  $b$  の間に挟まれたユニットの個数と定義される ． Fig. 5.2 上の例では，ボンド  $a, b$  の間にはユニットが 2 つ存在するので， $k(a, b) = 2$  である ． なお，同じ  $k(a, b)$  の値をもつ  $(a, b)$  同士を，「同じステージ  $k$  に属する」と表現する (Fig. 5.2 下) ．
- $k_{max}$  : 対象分子の中で現われる  $k(a, b)$  の最大値 ．

さて， $\mathbf{R}_{ab}$  の計算は，以下のように展開することができる ．

$$\mathbf{R}_{ab} = l_{a^*b^*} + \sum_{a'} \mathbf{R}_{a'tb} + \sum_{b'} \mathbf{R}_{ab'b'} - \sum_{a', b'} \mathbf{R}_{a'tb'} \quad (5.5)$$

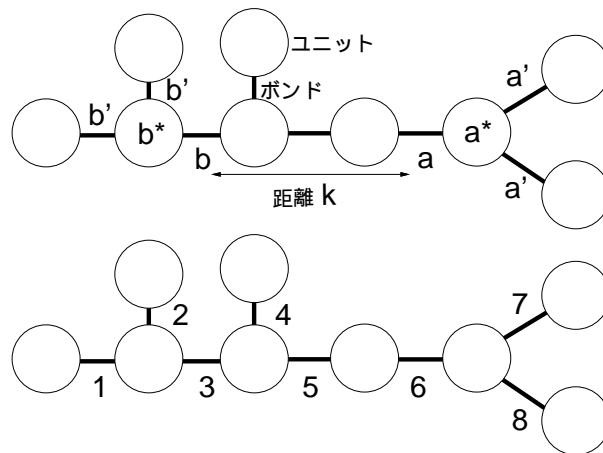


Figure 5.2 (上)  $a^*$ ,  $a'$  とボンド間距離の概念図 . (下) 「ステージ」の概念図 .  $(a, b) = \{(1, 6), (2, 6), (3, 7), (3, 8), (4, 7), (4, 8)\}$  は, いずれも  $k(a, b) = 3$  であるので, 同じ「ステージ3」に属する .

```

for (k = max_k; k >= 0; k--)
  for (ステージkに属するボンドペア(a,b))
     $I_{a^*b^*}$  の計算
     $R_{ab}$  の計算
     $R_{ab} = I_{a^*b^*} +$ 
       $R_{a'b} + R_{ab'} - R_{a'b'}$ 
    Hessab の計算

```

Figure 5.3 Gō らによる Hessian 計算アルゴリズム .

ただし,  $l_{a^*b^*}$  は  $6 \times 6$  の非対称行列で, その各要素はユニット  $a^*$  と  $b^*$  に含まれるすべての原子ペアの相互作用を表す.

ここで,  $k(a, b) = k$  であるときに, 以下の式が常に成り立つ.

$$k(a', b) = k(a, b') = k + 1 \quad (5.6)$$

$$k(a', b') = k + 2 \quad (5.7)$$

これらと式 (5.5) から,  $k(a, b) = k$  であるボンドペア  $(a, b)$  についての  $R_{ab}$  の計算を行う際には,  $k + 1$  と  $k + 2$  に対応する  $R$  だけを保存しておけばよいことがわかる.

そこで, すべての原子ペアについての相互作用を計算するときに,  $k(a, b)$  が大きい値のボンドペアから順に計算を行い, 前の計算結果を逐次利用することで計算量を減らすことができる.

まず  $k(a, b) = k_{max}$  である, すなわちステージ  $k_{max}$  に属するボンドペア  $(a, b)$  について  $R_{ab}$  の計算を行う (Fig. 5.3). 次にステージ  $k_{max} - 1$  に属する  $(a, b)$  についての計算を, ステージ  $k_{max}$  で計算した結果を用いながら行う. その次には, ステージ  $k_{max} - 2$  に属する  $(a, b)$  についての計算を, ステージ  $k_{max}$  とステージ  $k_{max} - 1$  で計算した結果を用いながら行う. これを順次繰り返し,  $k(a, b) = 0$  までのボンドペアについて計算する. これが Gō らによる Hessian 計算アルゴリズムの概要である.

## 5.2 Unify 方式による Hessian 並列計算法の開発

### 5.2.1 Unify 方式のアルゴリズム

Hessian の並列計算法として, 前章で述べた Gō らによるアルゴリズムをベースにして, 「Unify 方式」と呼ばれるアルゴリズムを開発した<sup>6)</sup>. 計算を行うためのシステムは, マスター/スレーブ方式で実現されている. マスター/スレーブ方式とは, 1つの「マスター」プロセッサが, 他の「スレーブ」プロセッサに仕事を依頼し, スレーブはマスターから依頼された仕事を終わると終了通知メッセージをマスターに送信して, 再びマスターからの仕事待ち状態になる, という並列計算実行方式である (Fig. 5.4). 今の場合, マスターはスレーブに Hessian の各要素の計算を割り当て, スレーブは自分に割り当てられた要素の計算をすべて終わると, その結果をマスターに返す.

Unify 方式は, Hessian の各要素  $Hess_{ab}$  を計算する際に, 同じステージに属する  $(a, b)$  については  $R_{ab}$  を互いに独立に計算できることから, これらの計算を複数プロセッサで分割して行うというものである. 概略を Fig. 5.5 に示す. 具体的な手順は, 以下のようになる.

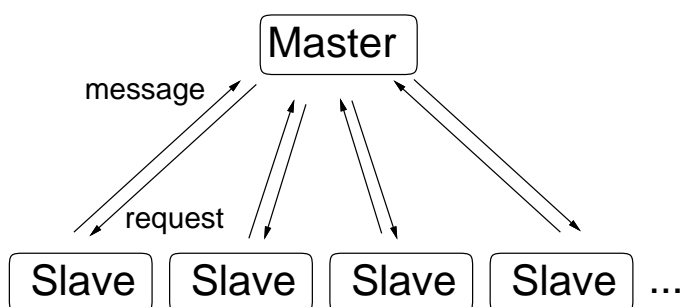


Figure 5.4 マスター/スレーブ方式の概念図 .

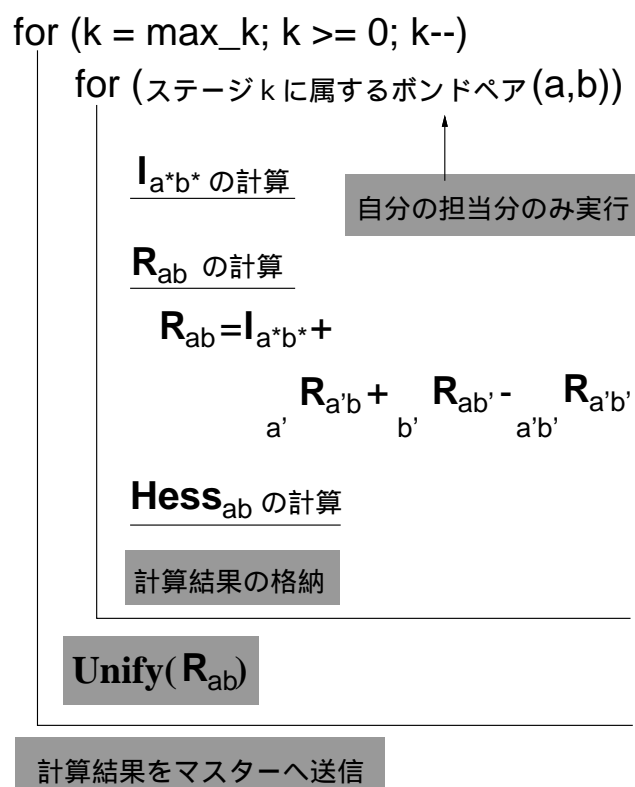


Figure 5.5 Unify 方式による Hessian の並列計算アルゴリズム .

1. 計算の最初は,  $k = k_{max}$  とする .
2. ステージ  $k$  に属するボンドペアをプロセッサに均等に割り当てる .
3. 割り当てられた  $(a, b)$  について  $l_{a*b^*}$  と  $R_{ab}$  を計算する . (ステップ A)
4. 計算した  $R_{ab}$  に対応する Hessian の要素  $Hess_{ab}$  を計算する . (ステップ B)
5. プロセッサに割り当てられたステージ  $k$  に属する  $(a, b)$  すべてについて, ステップ A およびステップ B を実行する .
6. 自分が計算した  $R_{ab}$  を, 他のすべてのプロセッサへ配布する . これを Unify ルーティンと呼ぶ . (ステップ C)
7.  $k$  を 1 つ減らして, ステップ 2 へ戻る .
8.  $k = 0$  の計算の終了後に, 各プロセッサはマスターに自分が計算した  $Hess_{ab}$  の値を送信する . マスターはスレーブからの結果を受信し, Hessian の全要素に統合する . (ステップ D)

Unify ルーティン (ステップ C) は, 自分が計算した  $R_{ab}$  を他のプロセッサに配布するとき用いる . Unify ルーティンの概要を Fig. 5.6 に示す . 各プロセッサでの送信と受信をそれぞれ同時に行うために, 効率がよい . プロセッサ  $i$  は, まずプロセッサ  $(i + 1)$  に自分の計算部分を送信し, プロセッサ  $(i - 1)$  からのデータを受信する . ついで, 自分がプロセッサ  $(i - 1)$  から直前に受信した部分をプロセッサ  $(i + 1)$  へ送信し, プロセッサ  $(i - 1)$  からのデータを受信する . これをプロセッサ数  $p$  のときに  $(p - 1)$  回繰り返すと, すべてのプロセッサは  $R_{ab}$  全体をもつことができる .

### 5.2.2 Unify 方式による Hessian 計算の並列化効率

Unify 方式による Hessian 並列計算法と, この Hessian 計算法を組み込んだニュートン・ラフソン法によるエネルギー極小化計算を, 実際の生体分子に適用して, その並列化の効果を検証した . ニュートン・ラフソン法の際に必要な配座エネルギーとその 1 次微分 (Gradient) の計算も並列化している (第 4 章参照) .

対象分子は, 74 ヌクレオチドから成るグルタミン tRNA ( $tRNA^{Gln}$ ) と, 540 残基から成るグルタミンイル tRNA 合成酵素 (GlnRS) である . 自由度の数は,  $tRNA^{Gln}$  が 441, GlnRS が 2621 であった . 計算は東京大学情報基盤センターの分散メモリ型並列コンピュータ 日立 SR2201 で行った .

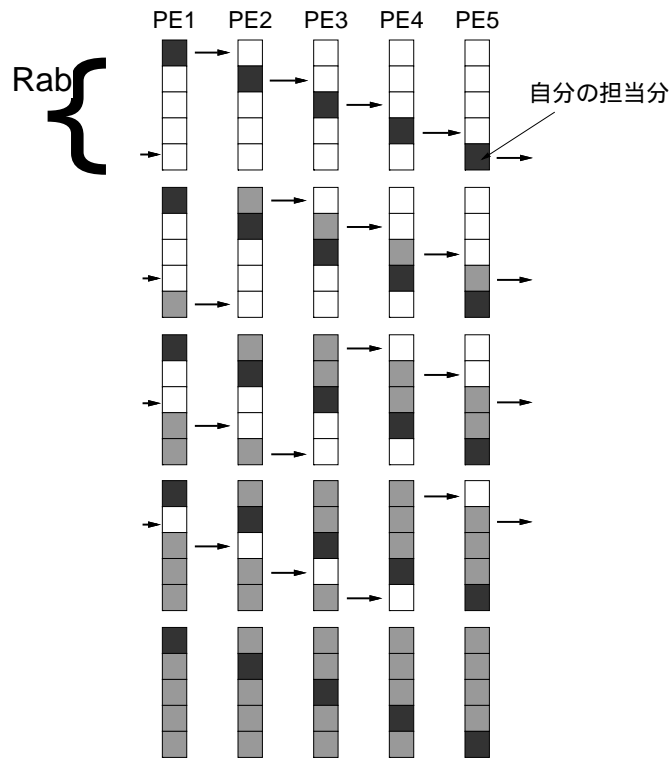


Figure 5.6 Unify ルーティンのアルゴリズム．プロセッサ (PE) 数が 5 台の場合．

以下に結果を示す． $\text{tRNA}^{\text{Gln}}$  および  $\text{GlnRS}$  について，プロセッサ数と計算時間の関係をプロットしたものが Fig. 5.7 である．グラフの横軸は，Hessian 計算に用いたプロセッサ数である．また縦軸は，「スピードアップ」(speedup) である．1 プロセッサを用いたときの計算時間を 1 とし，それよりも  $N$  倍速い計算の場合，スピードアップは  $N$  となる．実線は理想的なスピードアップ，すなわち 2 プロセッサを用いたときにスピードアップが 2，4 プロセッサを用いたときにスピードアップが 4 となる直線である．破線は  $\text{tRNA}^{\text{Gln}}$  におけるスピードアップの実測値，点線は  $\text{GlnRS}$  におけるスピードアップの実測値である．

Unify 方式による Hessian 計算の並列化効率は， $\text{tRNA}^{\text{Gln}}$  の場合は 24 プロセッサを用いたときがもっとも高く 1 プロセッサの 6.8 倍， $\text{GlnRS}$  の場合は 54 プロセッサを用いたときがもっとも高く 1 プロセッサの 11.2 倍であった． $\text{tRNA}^{\text{Gln}}$ ， $\text{GlnRS}$  いずれの場合も，プロセッサ数が 24 を越えるとスピードアップが理想直線から大きくはずれ，とくに  $\text{tRNA}^{\text{Gln}}$  の場合は 54 プロセッサを用いた場合に 24 プロセッサでの計算時間よりも多くの計算時間がかかってしまっていた．

このスピードアップの理論上限を求めるために，Amdahl の法則<sup>61)</sup> を適用する．Amdahl の法則によれば，プロセッサ数  $N_{\text{proc}}$  とスピードアップ  $S$  の間には以下の関



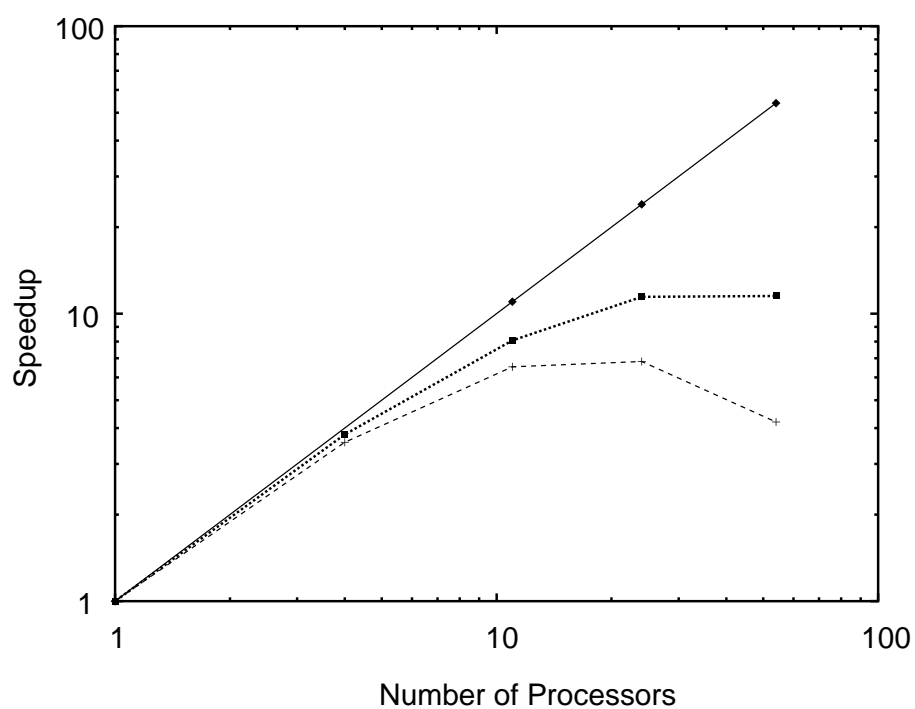


Figure 5.7 Unify 方式による Hessian 計算の並列化効率．実線は理想的なスピードアップ．破線は tRNA<sup>Gln</sup> におけるスピードアップの実測値．点線は GlnRS におけるスピードアップの実測値．

係式が成り立つ．

$$S = \frac{N_{\text{proc}}}{1 + \alpha (N_{\text{proc}} - 1)} \quad (5.8)$$

ただし， $\alpha$ はこのシステムのボトルネックを表す量である．ボトルネックとは，全体の計算時間  $T$  を，並列化によって短縮可能な部分  $T_p$  と並列化によっても変わらない部分  $T_s$  に分けたときの， $T (= T_s + T_p)$  の中に  $T_s$  が占める割合 ( $T_s/T$ ) のことである．並列化によって短縮できる  $T_p$  をゼロにした極限では，全体の計算時間は  $T_s$  になるので， $T/T_s = 1/\alpha$  が，システムを理想的に最適化したときのスピードアップの上限となる．また，式 (5.8) を変形すると，以下の式が得られる．

$$\frac{N_{\text{proc}}}{S} = 1 + \alpha (N_{\text{proc}} - 1) \quad (5.9)$$

このことから， $N_{\text{proc}}/S$  を  $N_{\text{proc}} - 1$  に対してプロットすると， $\alpha$  は得られたグラフの傾きとして求められることがわかる．GlnRS の結果について， $N_{\text{proc}}/S$  を  $N_{\text{proc}} - 1$  に対してプロットしたものが Fig. 5.8 である．グラフを  $y$  切片が 1 の直線でフィッティングすると  $y = 0.065x + 1$  となったことから，このシステムのボトルネック  $\alpha$  は 6.5% であり，スピードアップ  $S$  の理論上限は  $1/\alpha = 15.4$  倍であることがわかった．

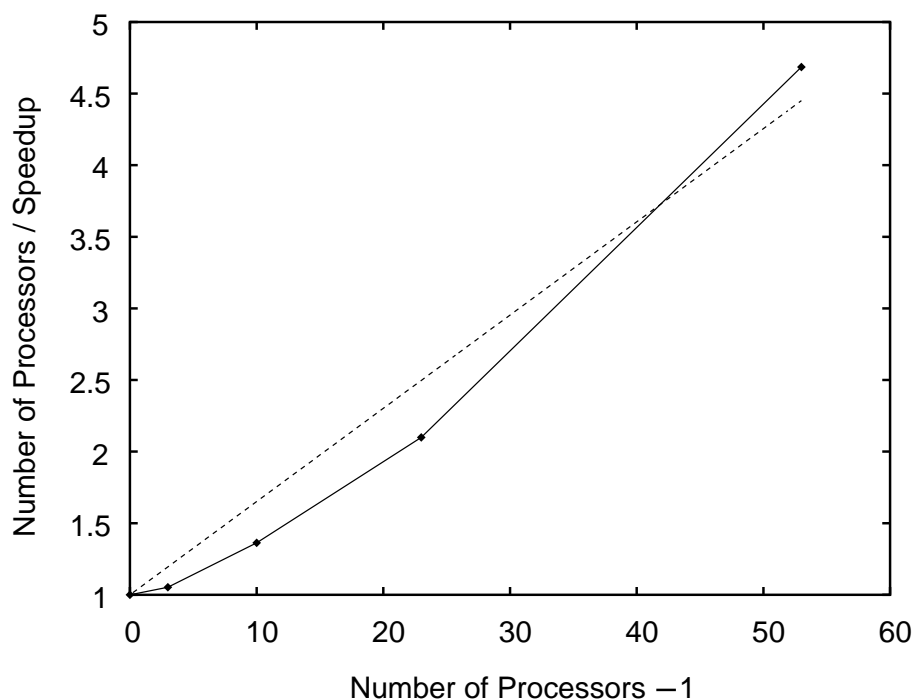


Figure 5.8  $N_{\text{proc}}/S$  を  $N_{\text{proc}} - 1$  に対してプロットしたもの．実線は GlnRS の場合の実測値で，破線はこれにフィッティングした  $y$  切片が 1 の直線を表す．

プロセッサ数が 24 を越えたときの性能低下の原因を解析するため，第 5.2.1 節で述べた Unify 方式での各ステップにおける計算時間を詳細に解析した．これをまとめたものが Table 5.1 および Table 5.2 である．

Table 5.1 tRNA<sup>Gln</sup> の Hessian 計算過程のステップ A-D における計算時間 (秒) .

プロセッサ数	1	4	11	24	54
(A) $l_{a^*b^*}$ と $R_{ab}$ の計算	13.95	3.49	1.35	0.64	0.32
(B) Hess <sub>ab</sub> の計算	0.55	0.14	0.06	0.02	0.01
(C) Unify ルーティン	0.00	0.54	0.96	1.59	3.25
(D) Hess <sub>ab</sub> のマスターへの送信	0.01	0.04	0.04	0.04	0.05

Table 5.2 GlnRS の Hessian 計算過程のステップ A-D における計算時間 (秒) .

プロセッサ数	1	4	11	24	54
(A) $l_{a^*b^*}$ と $R_{ab}$ の計算	258.94	62.55	22.33	10.37	4.71
(B) Hess <sub>ab</sub> の計算	20.67	5.25	1.87	0.86	0.37
(C) Unify ルーティン	0.00	8.00	11.16	14.29	20.33
(D) Hess <sub>ab</sub> のマスターへの送信	0.75	1.16	1.25	1.28	1.18

ステップ A ( $l_{a^*b^*}$ ,  $R_{ab}$  の計算) およびステップ B (Hess<sub>ab</sub> の計算) においては，並列化効率はほぼ 100% であった．しかし Unify ルーティンにおいては，プロセッサ数が増えるにつれて計算時間が大きくなっている．

Unify ルーティンの計算時間  $t_u$  は，送信すべき要素データ ( $R_{ab}$ ) の量  $d$  と，通信の立ち上げに要する時間  $\beta$  に依存する．データ量  $d$  は各プロセッサに分割されるので，1 回の送信で送られるデータ量は  $d/N_{\text{proc}}$  となり，プロセッサ数が増えると減少する．一方必要なデータ送信回数は  $N_{\text{proc}} - 1$  であるので，プロセッサ数が増えると増加する．以上から  $t_u$  を式で表すと，次のようになる．

$$t_u = \frac{d}{N_{\text{proc}}}(N_{\text{proc}} - 1) + \beta(N_{\text{proc}} - 1) \quad (5.10)$$

上式において，プロセッサ数  $N_{\text{proc}}$  が増えたときの極限を考えると， $t_u \sim \beta N_{\text{proc}}$  となるので，Unify ルーティンはプロセッサ数に比例して時間がかかることになる．

また，tRNA<sup>Gln</sup> および GlnRS について，ボンドペア ( $a, b$ ) の間の距離  $k(a, b)$  に対するボンドペアの数の分布をプロットしたものが，Fig. 5.9 である．tRNA<sup>Gln</sup>，

GlnRS いずれの場合も，ボンド間距離  $k(a, b)$  が 0 から  $k_{max}$  に増えるのにしたがって，ボンドペア数はほぼ直線的に減少している．ボンドペア数が少ないと，1 プロセッサあたりの計算量が少なくなるので，相対的に Unify に要する時間の割合が増え，プロセッサ数が増えたときの性能低下が大きくなると考えられる．また tRNA<sup>Gln</sup> は全体の自由度の数が GlnRS よりも少なく，一層この傾向が強いと考えられる．

このことを示すために， $k(a, b)$  に対して要した計算時間をプロットしたものを Fig. 5.10 に示す．11 プロセッサと 54 プロセッサのグラフを比較すると，横軸  $k(a, b)$  が大きくなるほど 11 プロセッサに対する 54 プロセッサの計算時間が長くなり，プロセッサ数が増えたときに性能低下を起こしていることがわかる．tRNA<sup>Gln</sup> においては，ほとんどの  $k(a, b)$  において 54 プロセッサの方が 11 プロセッサよりも計算時間が長くなってしまっている．

以上のような理由で，24 プロセッサ以上の場合には，計算時間の大半を Unify ルーティンが占めるようになり，並列化の効果が大きいステップ A, B の割合が相対的に減少して，性能低下を招いたと考えられる．

Unify ルーティンでは，各プロセッサが自分に割り当てられたステージ  $k$  における Hessian 要素の計算が終了したあと，全プロセッサに自分が計算した  $R_{ab}$  を送信している．しかし  $R_{ab}$  の計算には，ステージ  $k+2$  およびステージ  $k+1$  に属する  $R$  行列の要素すべてが必要なわけではなく， $(a', b)$ ， $(a, b')$ ， $(a', b')$  に対応する  $R$  行列のデータのみが必要である．そこで， $R_{ab}$  と  $R_{a'b}$ ， $R_{ab'}$ ， $R_{a'b'}$  の間の依存関係を解析して，通信コストが最小となるようにプロセッサに  $R_{ab}$  の計算を割り当てるのが Unify 方式の改良として考えられる．これは，第 5.3 節で述べるサブタスクグループ方式で実現されている．

### 5.2.3 Unify 方式によるニュートン・ラフソン法の並列化効率

Hessian，Gradient，エネルギーの並列計算によって，ニュートン・ラフソン法によるエネルギー極小化の 1 ステップの計算時間がどのくらい短縮されるかを解析した．ニュートン・ラフソン法は，関数の極小化に広く使われる一般的な方法である．

関数  $E(\theta)$  の Gradient ベクトルの第  $a$  成分を，

$$g_a(\theta) \equiv \frac{\partial E(\theta)}{\partial \theta_a} \quad (5.11)$$

とし， $E(\theta)$  の Hessian の第  $(a, b)$  成分を

$$F_{ab}(\theta) \equiv \frac{\partial^2 E(\theta)}{\partial \theta_a \partial \theta_b} \quad (5.12)$$

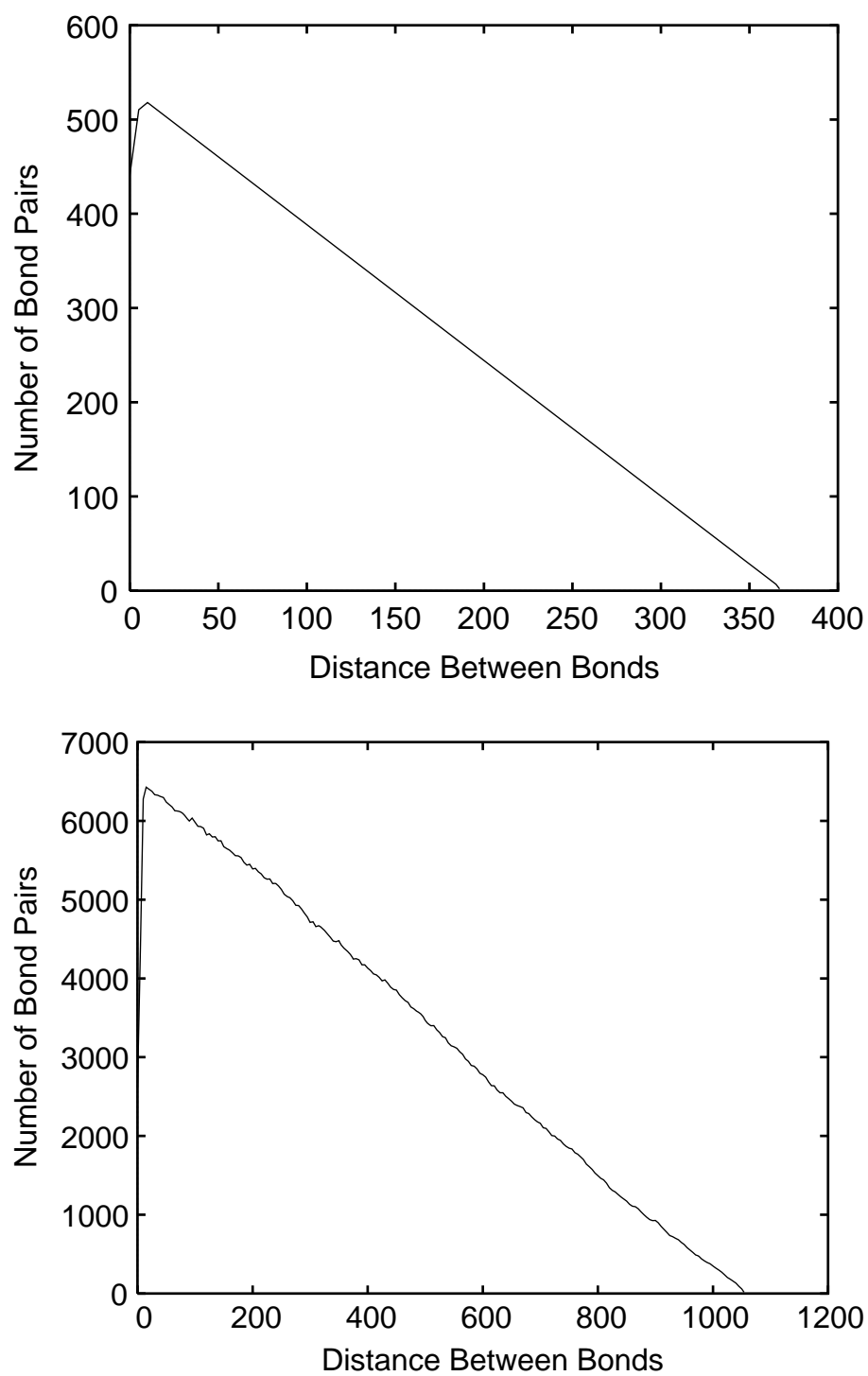


Figure 5.9 ボンドペア間の距離に対するボンドペア数の分布 . (上) tRNA<sup>Gln</sup> の場合 . (下) GlnRS の場合 .

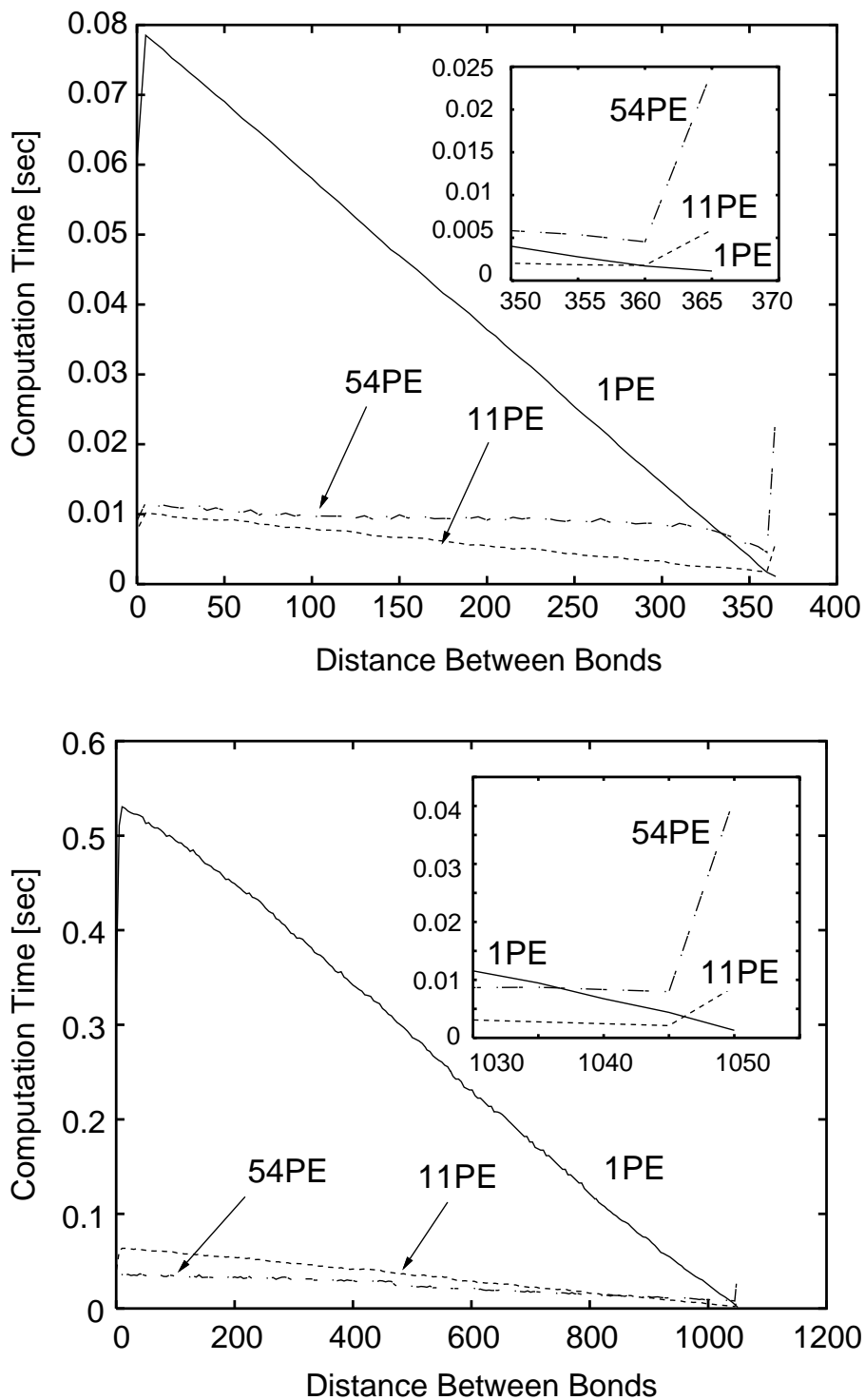


Figure 5.10 ボンドペア間の距離に対する計算時間．プロセッサ (PE) 数が, 1, 11, 54 について示した．また各グラフの中に  $k_{max}$  付近の拡大図を示す．(上)  $tRNA^{Gln}$  の場合．(下) GlnRS の場合．

とする．また Hessian 全体を，

$$\mathbf{F}(\theta) \equiv \{F_{ab}(\theta)\} \quad (5.13)$$

と表記する．

関数  $E(\theta)$  の極小条件は，「すべての自由度について傾き (Gradient) がゼロ」であるので，すべての  $a$  について， $g_a(\tilde{\theta}) = 0$  となる  $\tilde{\theta}$  の組を求めることになる． $\theta^i$  を  $\tilde{\theta}$  の第  $i$  次近似値とするととき，極小点における Gradient は，

$$g_a(\tilde{\theta}) = g_a(\theta^i) + \sum_b F_{ab}(\theta^i) (\tilde{\theta}_b - \theta_b^i) + (\text{高次の項}) = 0 \quad (5.14)$$

と展開される．高次の項を無視すると， $\theta^i$  よりもよい極小点の近似値として，

$$\theta_a^{i+1} = \theta_a^i - \sum_b [F^{-1}(\theta^i)]_{ab} g_b(\theta^i) \quad (5.15)$$

を得る．ただし  $a, b = 1, 2, \dots, n$  ( $n$  は自由度の数) である．この操作を繰り返すことで，極小点への到達を図るのがニュートン・ラフソン法である．1 変数関数での簡単な概念図を Fig. 5.11 に示す．

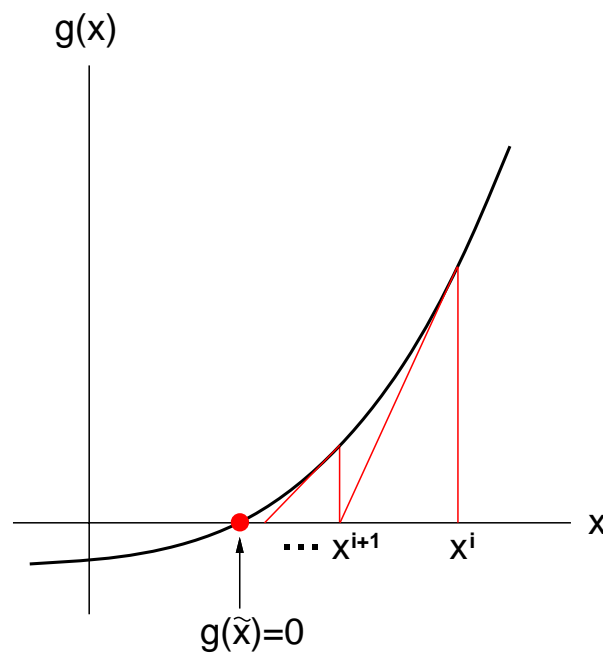


Figure 5.11 1 変数関数でのニュートン・ラフソン法概念図．操作の反復にしたがって， $x^i$ ， $x^{i+1}$  と解が真の解  $\tilde{x}$  に近づいていく．

実際には，式 (5.15) の右辺第 2 項をそのまま用いるのではなく，右辺第 2 項で表される方向にエネルギー曲面を探索し，現在のエネルギー値よりも下がる点を見つ

けてそこへ進むということがよく行われる．このため，1ステップのニュートン・ラフソン法の計算には，1回の Hessian 計算，1回の Gradient 計算，複数回のエネルギー計算が含まれることになる．

並列化を行った部分の計算時間は，1ステップの計算のうちの 98 ~ 99% を占めているため，並列化を行うことは非常に効果的であると考えられる．

Gradient 計算にそれぞれ 1, 2, 3, 6, 8 プロセッサを用い，Hessian 計算にそれぞれ 1, 4, 11, 24, 54 プロセッサを用いた結果について示す．エネルギー計算は，Hessian/Gradient の計算と同時にには行わないため，Gradient と Hessian を行うプロセッサ全台を用いて計算している．

tRNA<sup>Gln</sup> と GlnRS について，用いたプロセッサ数の合計に対するスピードアップを示したものが，Fig. 5.12 である．ニュートン・ラフソン法の 1 ステップの計算には，tRNA<sup>Gln</sup> の場合は 5 回のエネルギー計算，GlnRS の場合は 4 回のエネルギー計算が含まれていた．

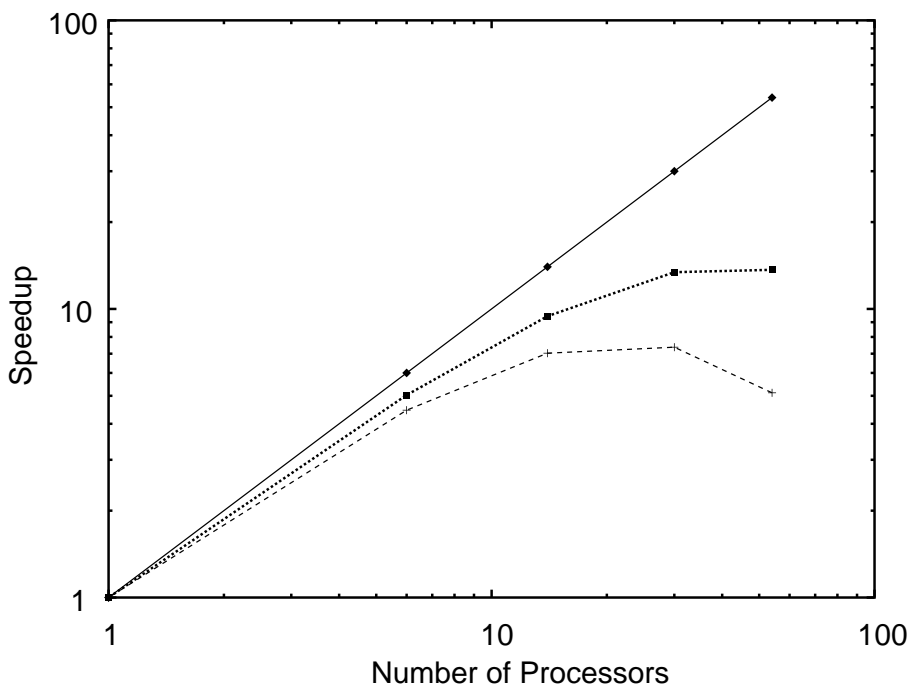


Figure 5.12 Unify 方式によるニュートン・ラフソン法の 1 ステップの並列化効率．実線は理想的なスピードアップ．破線は tRNA<sup>Gln</sup> におけるスピードアップの実測値．点線は GlnRS におけるスピードアップの実測値．

Gradient とエネルギーの計算においては，ほぼ 100% の並列化効率を得られた．得られた最大のスピードアップは，tRNA<sup>Gln</sup> では 30 プロセッサを用いた場合の 6.3 倍，GlnRS では 62 プロセッサを用いた場合の 12.0 倍であった．計算時間全体に対



する Hessian 計算の占める割合は 40–56% 程度であり, tRNA<sup>Gln</sup> の場合のプロセッサ数が大きいところでの性能の低下は Hessian 計算の性能低下によるものと考えられる. 次節で述べるサブタスクグループ方式など, Hessian 計算のさらなる高速化によって, ニュートン・ラフソン法の性能向上が見込まれる.

### 5.3 サブタスクグループ方式による Hessian 並列計算法の高速化

前章で述べた Unify 方式による Hessian 計算の並列化アルゴリズムは, GlnRS 分子に適用した場合, 62 プロセッサを用いた場合で約 11 倍の並列化効率を実現する方法であるが, 計算に用いるプロセッサ数が増えると通信コストの増大によって性能低下がおきるという欠点があることが明らかになった.

そこで, Unify 方式を改良した方法として, サブタスクグループ方式による Hessian 並列計算アルゴリズムの開発について述べる. この方法は Hessian の要素 1 つ 1 つの計算を 1 つの「サブタスク」と定義し, サブタスク間の依存関係を解析して必要なデータのみを通信することによって, 通信コストを減らし, 高いスケラビリティ (並列化効率) を実現するものである. また, 遺伝的アルゴリズムを用いて, 実行環境にあわせてサブタスクの各プロセッサへの割り当て方法を最適化し, さらなる性能向上を図る.

#### 5.3.1 サブタスクと依存関係

サブタスクとは, 作業全体 (タスク) の中のまとまった部分単位のことである. あるサブタスク (A) は, 別のサブタスク (B) の結果を入力として実行されることがある. このとき, サブタスク A の実行を開始するためには, サブタスク B が終了している必要がある. これを, 「サブタスク A はサブタスク B に依存している」と表現する. また, サブタスク B をサブタスク A の「先行サブタスク (predecessor)」, 逆にサブタスク A をサブタスク B の「後続サブタスク (successor)」と呼ぶ.

今の Hessian 並列計算の場合, サブタスクは Hessian の要素 1 つ ( $Hess_{ab}$ ) の計算に対応する. このサブタスクを  $S_{ab}$  と表すことにする. また, サブタスク 1 つ 1 つには固有のサブタスク番号 (subtask ID) をつけるものとする. 第 5.1 節で述べたように,  $Hess_{ab}$  は  $6 \times 6$  の正方行列  $R_{ab}$  から求められ,  $R_{ab}$  の計算には  $R_{a'b}$ ,  $R_{ab'}$ ,  $R_{a'b'}$  が必要となる. つまり  $Hess_{ab}$  を計算するためには,  $Hess_{a'b}$ ,  $Hess_{ab'}$ ,  $Hess_{a'b'}$  の結果が必要である. このことから, サブタスク  $S_{ab}$  はサブタスク  $S_{a'b}$ ,  $S_{ab'}$ ,  $S_{a'b'}$  に依存しており,  $S_{ab}$  の先行サブタスクが  $S_{a'b}$ ,  $S_{ab'}$ ,  $S_{a'b'}$  ということになる. なお, 1 組のボンドペア  $(a, b)$  に対して,  $(a', b)$ ,  $(a, b')$ ,  $(a', b')$  はそれぞれ複数組が存在することも

あるので (Fig. 5.2 参照),  $S_{a'b}$ ,  $S_{ab'}$ ,  $S_{a'b'}$  もそれぞれ 1 つとは限らず, 複数存在する可能性がある.

### 5.3.2 サブタスク間の依存関係の解析

サブタスク間の依存関係の解析には, 筆者らのグループで開発した “Parsley” と呼ばれる並列アプリケーション実行基盤のためのサブタスク間依存関係解析ルーティンを使用する. Parsley は, サブタスク間の依存関係に基づいて, マスタープロセッサがスレーブプロセッサに最適なサブタスクの割り当てを行い, 並列アプリケーションの効率のよい実行を可能にする基盤ソフトウェアである<sup>5,8,62,9,10</sup>. 各プロセッサは, サブタスク間の依存関係に基づいて, 自分が計算した  $R_{ab}$  のデータを必要とするサブタスク (すなわち  $S_{ab}$  の後続サブタスク) を担当するプロセッサにのみ,  $R_{ab}$  のデータを送信する. これによって, Unify 方式よりも通信データ量を減らすことができる.

また, プロセッサ間の通信は立ち上げに時間を要するため, 同じデータ量の送受信であれば, 1 回の通信で送受信を行った方が, データを複数回に分けて送受信するよりも高速である. このため同じステージに属するサブタスクは独立に計算できることが保証されていることを利用して, 1 つのプロセッサに割り当てられている同じステージに属するサブタスク群は, ひとまとめにグループ化して, グループ単位で必要データの送受信を行うことにする (Fig. 5.13).

### 5.3.3 サブタスクグループ方式のアルゴリズム

サブタスクグループ方式によるアルゴリズムを Fig. 5.14 に示す.

具体的な手順は以下ようになる.

1. 計算の最初は,  $k = k_{max}$  とする.
2. ステージ  $k$  に属するサブタスクをプロセッサに割り当てる.
3. 各プロセッサは, サブタスク間の依存関係にしたがって, 先行サブタスクを担当したプロセッサからの  $R_{ab}$  データを受信する. (ステップ A)
4. 割り当てられた  $(a, b)$  について  $l_{a*b}$  と  $R_{ab}$  を計算する. (ステップ B)
5. 計算した  $R_{ab}$  に対応する Hessian の要素  $Hess_{ab}$  を計算する. (ステップ C)
6. プロセッサに割り当てられたステージ  $k$  に属するすべてのサブタスクについて, ステップ B およびステップ C を実行する. (ステップ D)

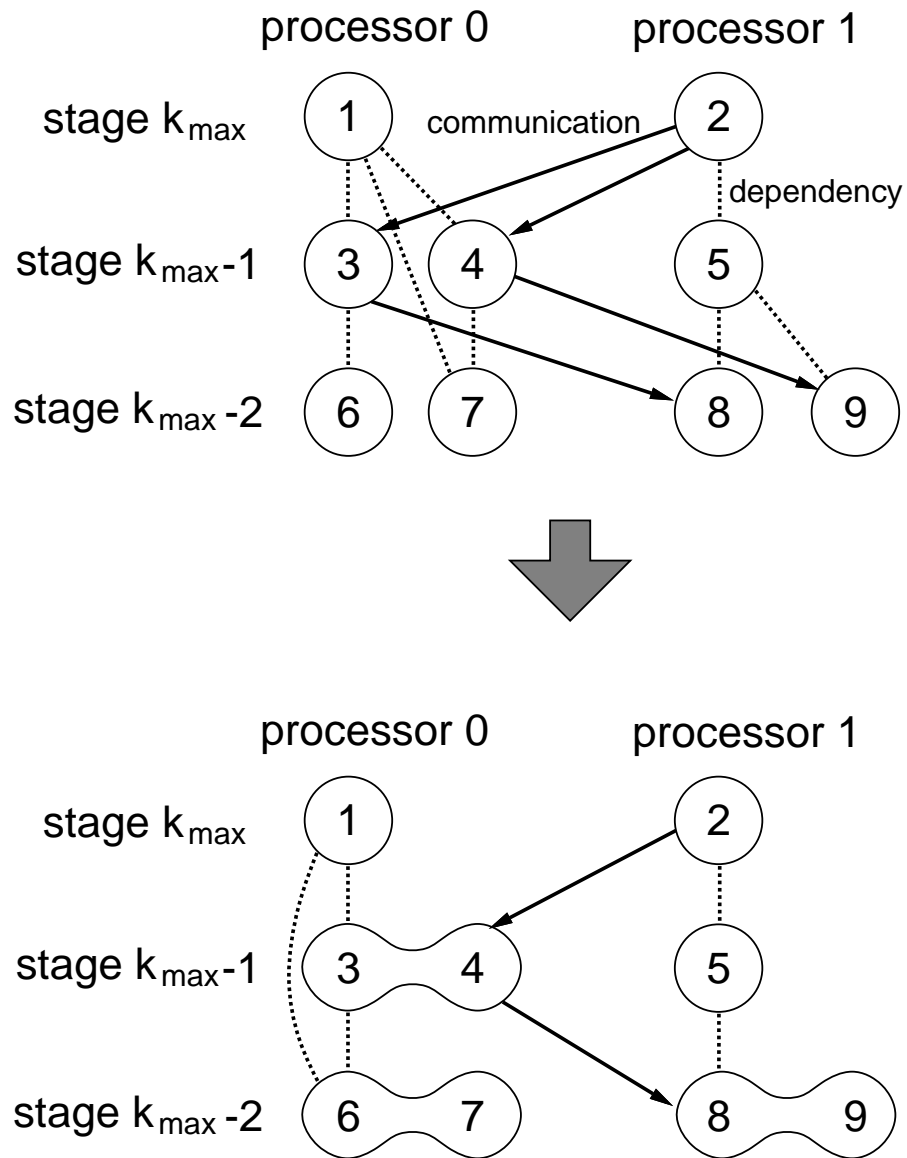


Figure 5.13 サブタスクのグループ化の概念図．円はサブタスク，矢印はデータ通信（依存関係を伴う），点線はデータ通信を伴わない依存関係を表す．上図において，同じステージに属するサブタスクをグループ化したものが下図である．グループ単位で通信を行うことで，通信回数が減少していることがわかる．

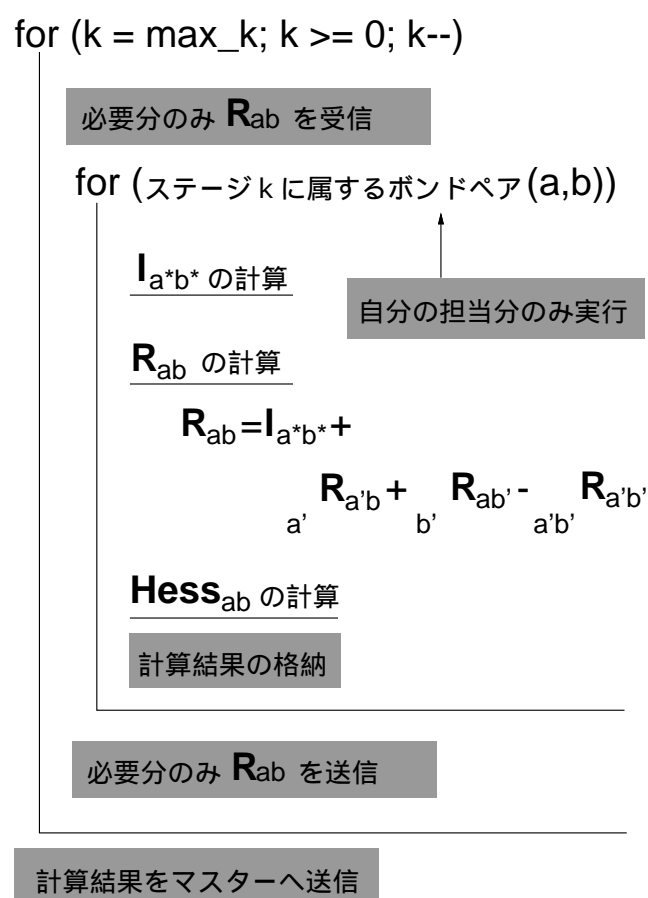


Figure 5.14 サブタスクグループ方式による Hessian の並列計算アルゴリズム .

7. サブタスク間の依存関係にしたがって、後続サブタスクを担当するプロセッサへ、自分が計算した  $R_{ab}$  データを送信する。(ステップ E)
8.  $k$  を 1 つ減らして、ステップ 2 へ戻る。
9.  $k = 0$  の計算の終了後に、各プロセッサはマスターに自分が計算した  $\text{Hess}_{ab}$  の値を送信する。マスターはスレーブからの結果を受信し、Hessian の全要素に統合する。

このように、1 つのサブタスクの計算は、先行サブタスクからのデータ受信 (ステップ A)、 $\text{Hess}_{ab}$  の計算 (ステップ B, C, D)、後続サブタスクへのデータ送信 (ステップ E) という 3 つのフェーズに分けられる。

#### 5.3.4 遺伝的アルゴリズムを用いたサブタスク最適割り当ての探索

サブタスクグループ方式では、依存関係に基づいて、必要な  $R_{ab}$  データのみを通信する。あるプロセッサが 1 つのサブタスクを実行したときに、そのサブタスクの後続サブタスクがすべて自分自身に割り当てられているとすると、 $R_{ab}$  を他のプロセッサに送信する必要がなくなる。これは、プロセッサへのサブタスクの割り当て方を最適化することで、 $R_{ab}$  のやりとりに必要なデータ量を最小限に抑えることができる可能性があることを意味する。

極端な例は、すべてのサブタスクを 1 台のプロセッサで実行する場合である。このときはすべてのサブタスクについてすべての後続サブタスクを自分自身で実行するので、 $R_{ab}$  を通信する必要がまったくない。しかし実際には、多数のサブタスクを 1 台のプロセッサで実行するので、計算負荷がそのプロセッサに集中し、全体の実行時間は大きくなってしまふ。このように、サブタスクの割り当て方の最適解は、計算負荷の分散と通信量の低減という、2 つの要素のトレードオフで決まることがわかる。

この最適なサブタスクの割り当て方の決定は、組み合わせ最適化問題であり、しらみつぶ的に解を探索するとその計算時間は指数関数的に増大してしまふ。そこで、このような問題に対してしばしば用いられる最適化手法である遺伝的アルゴリズムを用いて、サブタスクの割り当て方の最適化を試みた。遺伝的アルゴリズムは、最適解の候補を多数の個体 (染色体) の集団として表現し、染色体間の交叉や突然変異によって新たな候補を生成して集団にバリエーションを与え、定められた適応度の高い個体だけを次世代に選択的に継承することで、最終的に適応度のもっとも高

い最適解を得ようとするものである．遺伝的アルゴリズムは，以下のような特徴をもつ．

- 複数の個体（解の候補）から複数の個体へ探索を進めるので，初期値に比較的依存しにくい．
- 適応度の値を利用するだけで，適応度の勾配（形状）などの情報を用いないので，適応度を表す関数の性質がよくわからないような場合にも適用可能である．
- 確率的な遷移ルールにしたがって挙動するので，局所最小点にとどまらず，大域的な最小点に到達できる可能性が高い

遺伝的アルゴリズムによる最適化の際には，まず各プロセッサの動作タイミングをモデル化し，あるサブタスクの割り当て方を行ったときの実行時間全体を見積もる「シミュレータ」を構築することが必要となる．このため，1つのサブタスクの実行時間全体 ( $t_i$ ) を，先行サブタスクからの  $R_{ab}$  データの受信に要する時間 ( $t_i^{recv}$ )， $R_{ab}$  と  $Hess_{ab}$  の計算そのものに要する時間 ( $t_i^{calc}$ )，後続サブタスクへの  $R_{ab}$  データの送信に要する時間 ( $t_i^{send}$ ) に分け，それぞれを以下の式にしたがってモデル化した．

$$t_i = t_i^{recv} + t_i^{calc} + t_i^{send} \quad (5.16)$$

$$t_i^{recv} = w^{recv} + c^{recv} \times u_i^{recv} \quad (5.17)$$

$$t_i^{calc} = c^{calc} \times u_i^{calc} \quad (5.18)$$

$$t_i^{send} = w^{send} + c^{send} \times u_i^{send} \quad (5.19)$$

ただし， $u_i^{recv}$ ， $u_i^{calc}$ ， $u_i^{send}$  は，サブタスク  $i$  での受信，計算，送信に必要なデータ量を表す．また  $c^{recv}$ ， $c^{calc}$ ， $c^{send}$  は，それぞれ受信，計算，送信時間のデータ量に対する比例係数， $w^{recv}$ ， $w^{send}$  は，それぞれ受信，送信の立ち上げに必要な時間である．自分自身へのデータの送受信に要する時間はゼロとする．なお，以後  $t_i$  をサブタスク  $i$  の「実行時間」， $t_i^{recv}$  を「受信時間」， $t_i^{calc}$  を「計算時間」， $t_i^{send}$  を「送信時間」と呼ぶことにする．

$c$  と  $w$  は計算を行うコンピュータ（プラットフォーム）ごとに計測を行って求める．これらのパラメータは，プラットフォームが変わらない限り，計算対象分子によらず一定である．また  $u_i$  は，サブタスクごとにユーザが与える． $l_{a^*b^*}$  の計算の大部分はユニット  $a^*$  とユニット  $b^*$  に属する原子同士の相互作用の計算であるので，サブタスク  $S_{ab}$  に必要な計算量は，ユニット  $a^*$  とユニット  $b^*$  に含まれる原子の数の積にほ

ば比例する．そこでこの値を  $u_i^{calc}$  として用いる．受信，送信に必要なデータ量  $u_i^{recv}$  ,  $u_i^{send}$  は，プロセッサへのサブタスクの割り当て方によって決まる．

各プロセッサの動作タイミングの模式図を Fig. 5.15 に示す．このモデルを用いたシミュレータによって，あるサブタスク割り当てに対する実行時間全体を見積もり，これを遺伝的アルゴリズムを用いて最適化することで，最適なサブタスク割り当てを探索する．

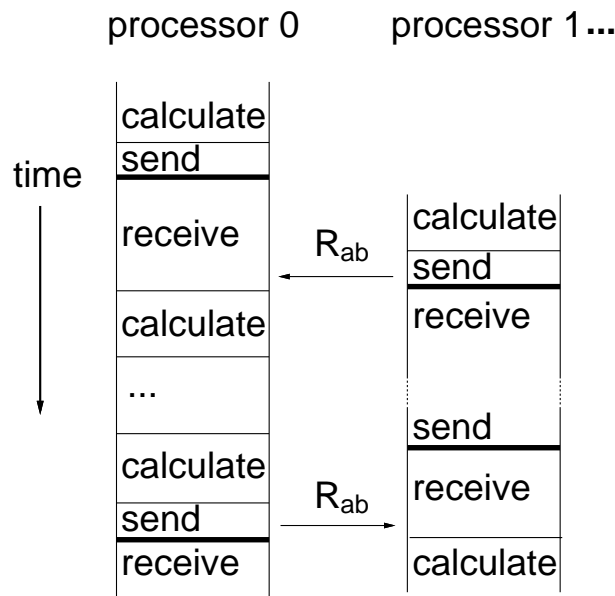


Figure 5.15 各プロセッサの動作タイミングの模式図．プロセッサ 1 は後続サブタスクを実行するプロセッサ 0 に，必要な  $R_{ab}$  データを送信している．プロセッサ 0 は， $R_{ab}$  データを受信し，それをを用いて割り当てられたサブタスクグループの計算を終えたあと，後続サブタスクを実行するプロセッサ 1 に必要な  $R_{ab}$  データを送信している．

サブタスクの割り当て方と染色体（遺伝子）との対応づけの方法は，以下のように行う．サブタスクの総数を  $N_S$  個とすると，1つの染色体  $x$  は  $N_S$  個の数値の列で表され，この数値1つが1つの遺伝子  $x_i$  ( $i = 1, \dots, N_S$ ) に対応する． $i$  番目の遺伝子  $x_i$  は， $i$  番目のサブタスクを割り当てるプロセッサの番号を表す．これにより，1つの染色体は，1組のサブタスクのプロセッサ割り当てを表すことになる．

たとえばプロセッサ 0 ~ 2 の 3 台に以下のように 8 つのサブタスクを割り当てるとする．

プロセッサ番号	担当サブタスク
0	1, 3, 4
1	5, 8
2	2, 6, 7

この割り当て方を表す染色体は、

	1	2	3	4	5	6	7	8
染色体	0	2	0	0	1	2	2	1

となる。

遺伝的アルゴリズムの手順には、標準的なものを用いる。具体的には、以下のように行う。

1. まず、最適化したい変数（今の場合はサブタスクの割り当て方）を「染色体」として表現する。
2.  $N$  個の染色体（最適解の候補）を生成して、「親」グループを作る。
3. 2つの親個体（親染色体）をその適応度に比例した確率で選びだし（ペアリング）、両者の間の交叉によって2つの子個体を生成する。
4. 生成した子個体に突然変異を加える。
5. 手順3, 4を、子個体の総数が  $N$  になるまで繰り返す。
6. 親と子を合わせた  $2N$  個の個体の「適応度」を計算し、上位  $N$  個体を次世代の親グループとする。
7. 手順3-6を  $N_{gen}$  世代にわたって繰り返す。

こうすることで、世代を経るにつれて適応度の高い個体（染色体）がグループ内に残ることになる。今の場合、「適応度」とはそのサブタスク割り当てでの Hessian 計算時間の見積値から求める。個体  $x^j$  の割り当て方による Hessian 計算時間の見積値を  $C^j$  とすると、その適応度  $f^j$  は、グループ内のもっとも大きな見積値 ( $C_{\max}$ ) を用いて、

$$f^j = (C_{\max} - C^j) + 0.001 \times C_{\max} \quad (5.20)$$



と表す．右辺第 1 項まででは，もっとも大きな見積値 ( $C^j = C_{\max}$ ) をもつ個体の適応度がゼロとなり，ペアリングの際に選ばれなくなってしまうので，右辺第 2 項の補正項を加えている．

手順 2 では，1 個体はサブタスク数が各プロセッサでなるべく均等になるようにし，残りの  $(N - 1)$  個の個体はその突然変異体とした．

手順 3 では，交叉部位を 1 箇所ランダムに選び，交叉部位以降の部分記号列を入れ換える (Fig. 5.16 上)．手順 4 では，各遺伝子座において，あらかじめ定められた突然変異率  $p_m$  で，その遺伝子座の遺伝子を変異させる (Fig. 5.16 下)．

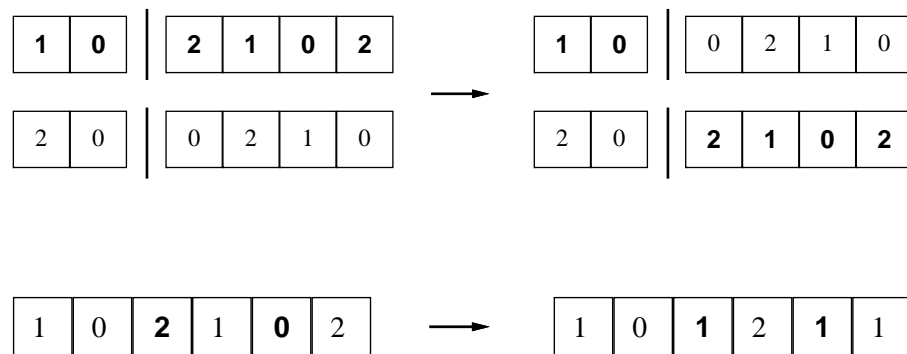


Figure 5.16 遺伝的アルゴリズムにおける遺伝子の操作の例．(上) 交叉 (cross over)．(下) 突然変異 (point mutation)．

探索範囲は非常に広いので，遺伝的アルゴリズムによる探索を，以下の「計算負荷の分散の最適化」と「実行時間全体の最適化」の 2 段階で行う．

### 計算負荷の分散の最適化

この段階の目的は，各ステージにおける各プロセッサの負荷をできるだけ均等にすることである．これにより， $R_{ab}$  の送受信のタイミングが各プロセッサでそろうために受信の際の待ち時間を少なくできる．ステージ  $k$  における適応度を  $T_k$  とすると，

$$T_k = \max_{(p,q)} |T_k^p - T_k^q| \tag{5.21}$$

と定義する． $p$  および  $q$  はプロセッサ番号である． $T_k^p$  および  $T_k^q$  は，それぞれプロセッサ  $p, q$  にステージ  $k$  において割り当てられたサブタスクグループの実行時間全体である．

### 実行時間全体の最適化

この段階の目的は、Hessian 計算全体に要する実行時間を短縮することである。あるプロセッサでの実行時間はそのプロセッサの各ステージでの実行時間の和で表され、全体の実行時間はもっとも遅いプロセッサの実行時間となるので、Hessian 計算全体の実行時間  $T$  は、

$$T = \max_p \sum_k T_k^p \quad (5.22)$$

と計算できる。  $p$  はプロセッサ番号である。この  $T$  を適応度とした最適化を行う。

### 遺伝的アルゴリズムによる最適化の手順

遺伝的アルゴリズムによる最適化では、計算負荷の分散の最適化を行ったあとで実行時間全体の最適化を行う。遺伝的アルゴリズムの個体数や突然変異率などのパラメータは、さまざまなものを試して最適と思われる値を探索した結果、Table 5.3 の値を用いた。なお、遺伝的アルゴリズムの計算には、Compaq DS20 クラスタ (16 プロセッサ) を用いた。計算時間は、対象とするプラットフォームのプロセッサ数にもよるが、およそ 12 時間程度であった。

Table 5.3 遺伝的アルゴリズムで用いたパラメータ

個体数	$N$	1600	(負荷分散の最適化の場合)
	$N$	160	(実行時間全体の最適化の場合)
突然変異率	$p_m$	0.00003	

計算負荷の分散の最適化では、各ステージについて、1000 世代にわたって最適解に変化がみられなければ探索終了とした。また実行時間全体の最適化では、1000 世代にわたって最適解の変化が 0.5% 以下の改善しかみられなければ探索終了とした。

式 (5.16) の受信、計算、送信時間のモデルで用いるパラメータは、SR2201 上で測定プログラムを実行しておおまかに値を見積もり、さらに値を微調整して実測値との整合性をとることによって、Table 5.4 のように求めた。

### 5.3.5 サブタスクグループ方式による Hessian 計算の高速化の結果

Unify 方式でとくに大きな性能低下がみられた tRNA<sup>Gln</sup> 分子について、サブタスクグループ方式による二面角系 Hessian の計算を行った。自由度の数は 441 であるので、サブタスクの数 (Hessian の要素数) は、 $441 \times (441 + 1) / 2 = 97461$  個であ

Table 5.4 受信, 計算, 送信時間のモデルで用いたパラメータ.

$c^{recv}$	$3.50 \times 10^{-8}$ sec/byte
$c^{calc}$	$4.93 \times 10^{-6}$ sec/byte
$c^{send}$	$1.00 \times 10^{-8}$ sec/byte
$w^{recv}$	$1.00 \times 10^{-8}$ sec
$w^{send}$	$7.10 \times 10^{-5}$ sec

る. 計算は, Unify 方式と同じく東京大学情報基盤センターの分散メモリ型並列コンピュータ 日立 SR2201 で行った.

#### シミュレータの性能評価

第 5.3.4 節で述べたように, サブタスクのプロセッサへの 1 つの割り当て方を与えたときに, その実行時間を見積もるシミュレータを構築した. シミュレータで見積もった実行時間が現実を正しく反映しているかどうかを確認するために, プロセッサ数を 1~60 の範囲で変えたときのシミュレータによる実行時間の見積もりと実際の実行時間を比較した (Fig. 5.17). 割り当ては, 各ステージにおいて, 各プロセッサができるだけ均等にサブタスク数を担当するようにしている. 点線がシミュレータによる見積値, 実線が実測値である.

シミュレータによる実行時間の見積値は, 実測値とよい一致を示している. このことから, このシミュレータによって計算された実行時間の見積値を適応度として, 遺伝的アルゴリズムによってサブタスクの最適割り当てを探索することは妥当であると思われる.

また, プロセッサ数が 16 のときに, 各プロセッサにおける受信・計算・送信・実行時間全体の, 見積値と実測値を示したものが Fig. 5.18 である. 横軸は 0 から 15 のプロセッサ番号で, 縦軸は時間である. 実測値を実線で, 見積値を点線で示している. また赤は実行時間全体, 緑は受信時間, 青は計算時間, マゼンタは送信時間である. とくに計算時間, 送信時間はよく一致している. また受信時間には若干の誤差が見られるが, 各プロセッサでかかる時間の傾向はよく再現できている. 実行時間全体については, 計算時間の誤差と受信時間の誤差が相補的であるために, 見積値と予測値にほぼ一定の誤差がみられる. これは, 計算時間の見積り誤差から発生する「待ち」時間が受信時間に含まれるためであると思われる.

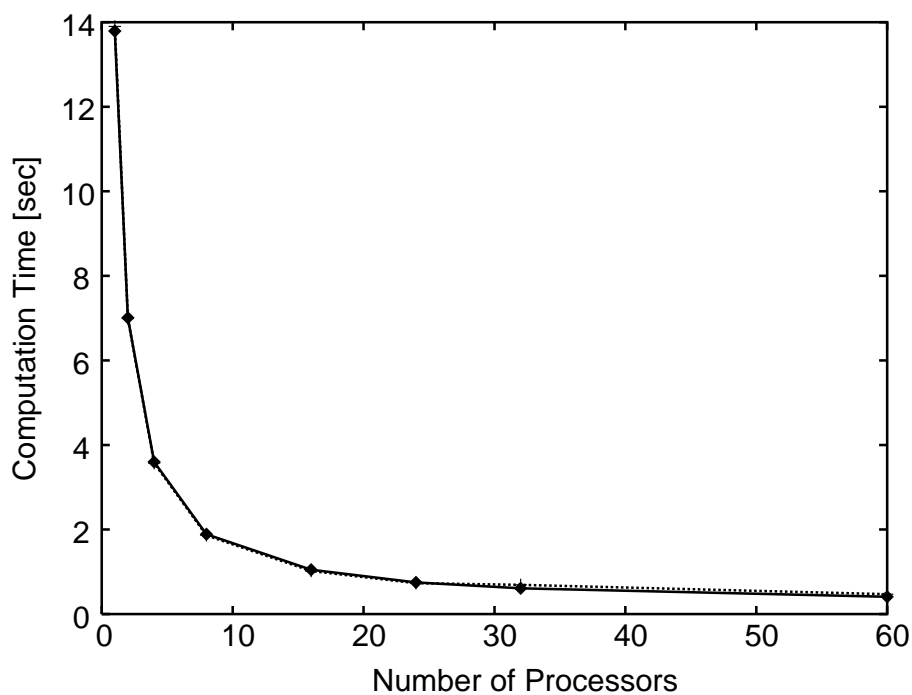


Figure 5.17 実行時間のシミュレータによる見積値 (点線) と実測値 (実線) の比較 .

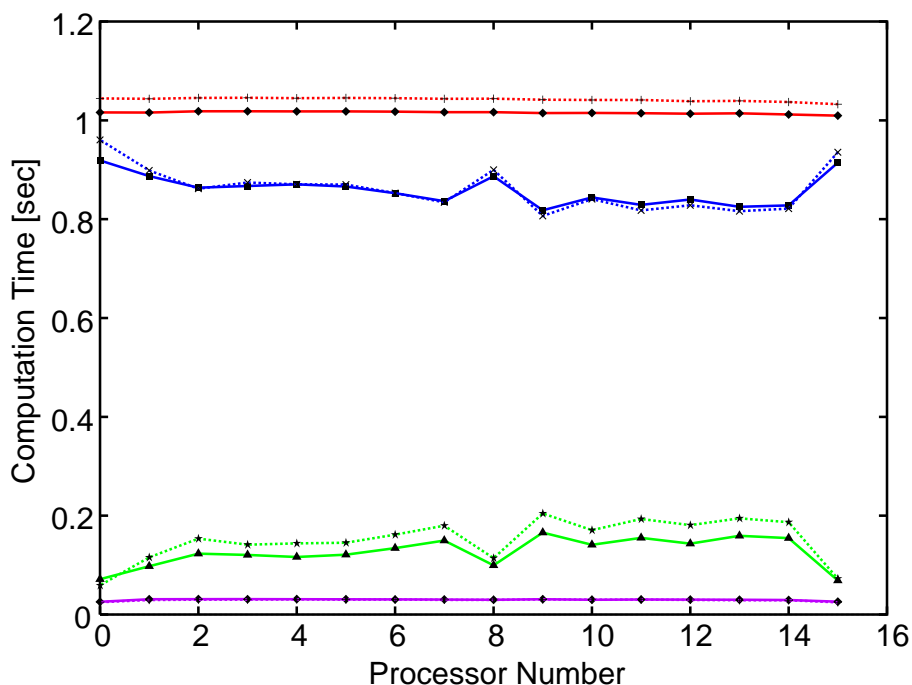


Figure 5.18 16 プロセッサのときの各プロセッサにおける実行時間の見積値と実測値の比較．実線は実測値，点線は見積値を表す．赤は実行時間全体，緑は受信時間，青は計算時間，マゼンタは送信時間である．

## サブタスクグループ方式による Hessian 計算の並列化効率

サブタスクグループ方式による Hessian 並列計算の，プロセッサ数に対するスピードアップを解析したものが，Fig. 5.19 である．横軸にプロセッサ数，縦軸にスピードアップをとっている．点線がサブタスクグループ方式によるスピードアップ，実線が理想的なスピードアップである．比較のために Unify 方式を破線で示してある．

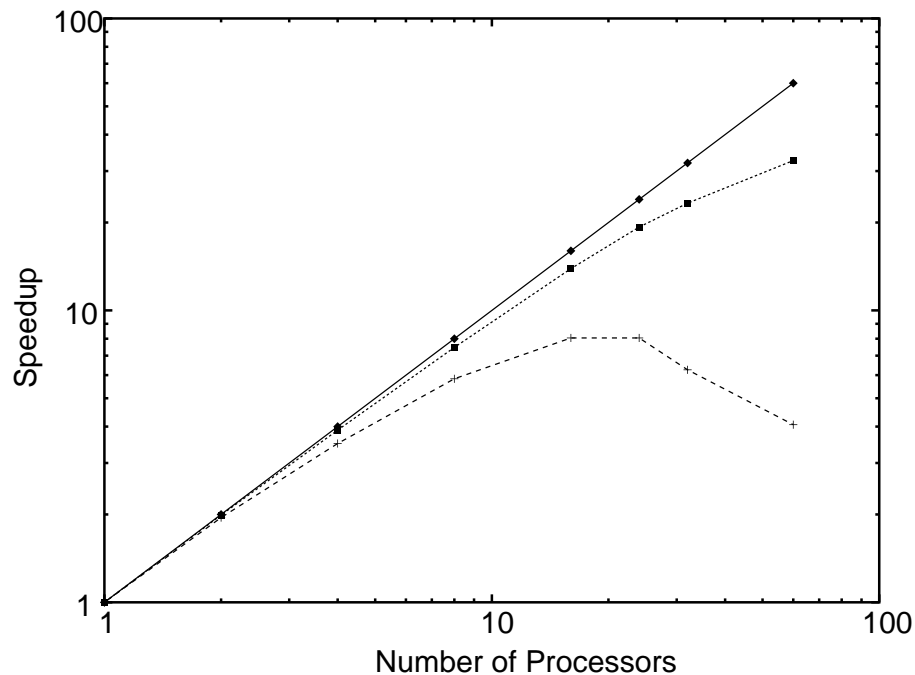


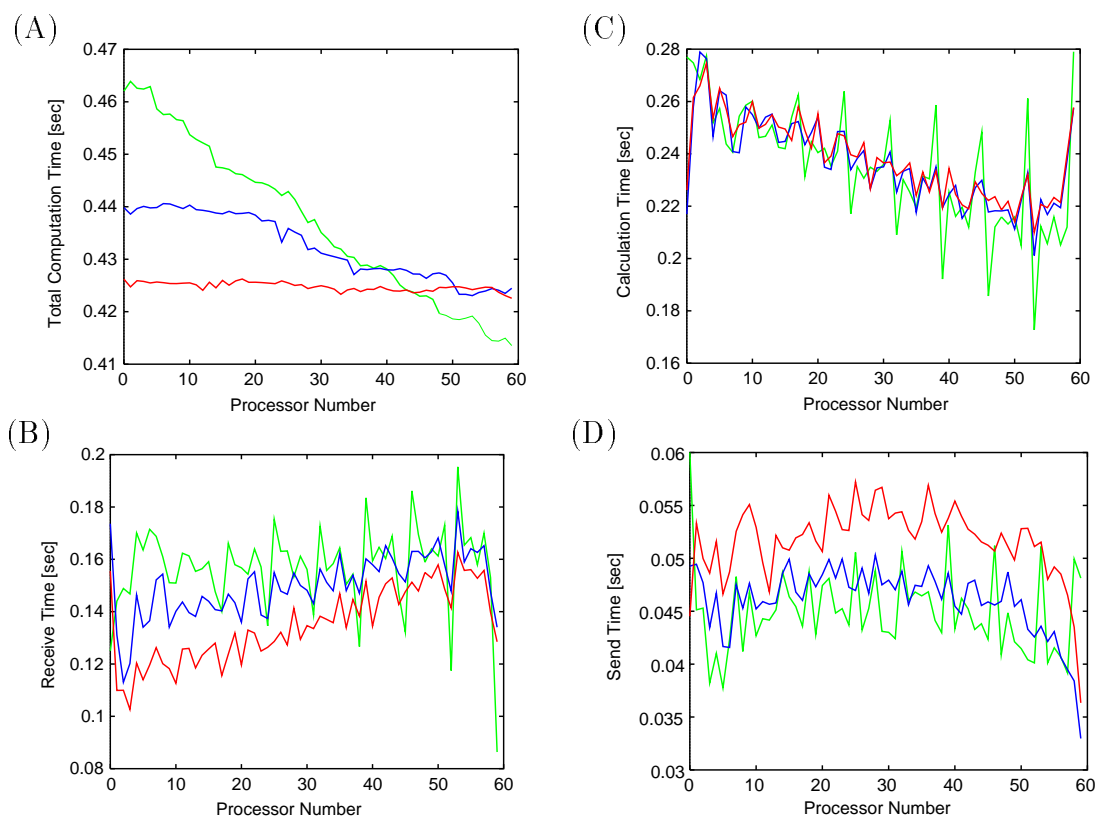
Figure 5.19 サブタスクグループ方式による  $\text{tRNA}^{\text{Gln}}$  の Hessian 計算時間の並列化効率．比較のため，Unify 方式による結果と併せて示した．(実線) 理想的なスピードアップ．(破線) Unify 方式によるスピードアップ．(点線) サブタスクグループ方式によるスピードアップ．

Unify 方式では，プロセッサ数 16 付近から性能の大幅な低下がみられたが，サブタスクグループ方式では，プロセッサ数が増えたときの性能低下は Unify 方式に比べて抑えられていた．60 プロセッサを用いたときに 32.6 倍のスピードアップが得られた．

第 5.2.2 節と同様，Amdahl の法則にしたがってシステムのボトルネック  $\alpha$  を計算したところ 1.3% となり，スピードアップの理論上限は  $1/\alpha = 75.5$  倍であることがわかった．

遺伝的アルゴリズムによるサブタスク割り当て最適化の効果を調べるために，遺伝的アルゴリズムを適用する前，計算付加の分散の最適化を行った後，実行時間全体の最適化を行った後の，各プロセッサの受信，計算，送信時間をそれぞれ比較し

た．60 プロセッサを用いた場合について Fig. 5.20 に示す．縦軸が時間，横軸がプロセッサ番号 (0 から 59 まで) であり，(A) に実行時間全体，(B) に受信時間，(C) に実行時間，(D) に送信時間を示す．それぞれのグラフでは，緑が最適化前，青が負荷分散の最適化後，赤が実行時間全体の最適化後である．



**Figure 5.20** 60 プロセッサにおける各プロセッサの実行時間全体と，受信，計算，送信時間．緑は遺伝的アルゴリズムによる最適化前で，各ステージにおいて担当するサブタスク数が各プロセッサでなるべく均等になるようにしたもの．青は負荷分散の最適化後．赤は実行時間全体の最適化後．(A) 実行時間全体．(B) 受信時間．(C) 計算時間．(D) 送信時間．

2 段階の遺伝的アルゴリズムによる最適化の効果は， Fig. 5.20 A の実行時間全体のグラフに明確にみることができ．最適化前には各プロセッサでの実行時間のばらつきが大きいのにに対して，最適化が進むにつれてばらつきが減少していく様子がわかる．またもっとも遅いプロセッサの実行時間（すなわち全体の実行時間）も，最適化が進むにつれて短縮されている．

計算時間 (Fig. 5.20 C) については，最適化前は各ステージごとにサブタスク数なるべく均等になるように割り当てられているが，各ステージに属するサブタスク数がプロセッサを下回る場合の影響や，サブタスクによる計算量のばらつきによって各プロセッサでの差が大きい．これが遺伝的アルゴリズムによる最適化によって

差が小さくなっている。

また受信時間 (Fig. 5.20 B) は, ほとんどのプロセッサで, 最適化が進むにつれて時間の短縮がみられる。送信時間 (Fig. 5.20 D) については, 最適化が進むにつれて逆に時間が増大しているが, 受信時間, 計算時間に比べて送信時間の絶対値が小さいため, 実行時間全体に対する影響は小さい。

以上のように, 遺伝的アルゴリズムによる最適化によって, 実行時間の短縮が実現できていることが確認された。サブタスクグループ方式では, Unify 方式ではプロセッサ数が増えたときに性能低下を招いた tRNA<sup>Gln</sup> のような小さな分子についても, 高い並列化効率を実現できた。さらに最適化の性能向上を図るための方法として, シミュレータで用いている各プロセッサの動作タイミングのモデルとパラメータの改良が挙げられる。60 プロセッサの場合に, 実行時間全体の最適化終了後のサブタスクの割り当て方に対して, シミュレータの見積値と実測値を示したものが, Fig. 5.21 である。

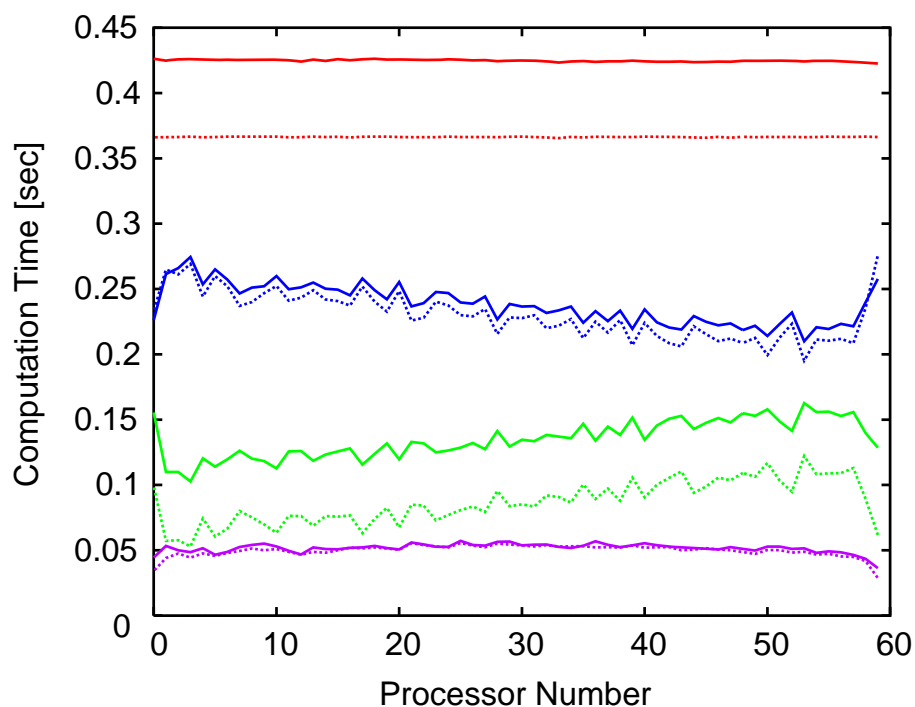


Figure 5.21 60 プロセッサ数のときの各プロセッサにおける実行時間の見積値と実測値の比較。実線は実測値, 点線は見積値を表す。赤は実行時間全体, 青は計算時間, 緑は受信時間, マゼンタは送信時間。

現在は式 (5.16)~(5.19) のような単純なモデルを用いているが, 実際には送受信時間は通信するバイト数の 1 次式で完全に記述することはできず, また計算時間も式

(5.18) の表現だけでは不十分である．このために，とくに受信時間について，シミュレータによる見積値と実測値の違いがみられる．プロセッサ数が大きいところでは，1 プロセッサあたりに割り当てられるサブタスクの数が減るために，計算時間が実行時間全体に占める割合が小さくなる．また送信時間は，絶対値が小さいために，実行時間全体に占める割合はもともと小さい．このことから，プロセッサ数が大きいところでは，受信時間が実行時間全体に対して支配的になり，受信時間のずれが実行時間全体のずれを生むことになる．

16 プロセッサを最適化前の割り当て方に用いた Fig. 5.18 と比較すると，計算時間の実行時間全体に占める割合が小さくなり，実行時間の誤差の大部分が受信時間の誤差に由来していることがわかる．より詳細な動作タイミングのモデル式を用いることで，受信・計算・送信時間の見積値と実測値の差を減少させ，さらに最適なサブタスクの割り当て方を遺伝的アルゴリズムによって発見できる可能性があると思われる．

プラットフォームと用いるプロセッサ数が変わらなければ，求められた最適な割り当て方を繰り返し利用することが可能である．Hessian 計算は1つの計算の中で繰り返し行われる場合が多いので，遺伝的アルゴリズムを用いた最適化のように時間をかけて最適割り当てを求めてもその利用価値は十分にある．最適な割り当ての探索によってよりよい解が得られれば，プログラムを一切変更することなく，高速化が図れることも利点である．また問題を「サブタスク間の依存関係に基づくサブタスク割り当ての最適化」という枠組みに抽象化して考えると，Hessian 計算だけでなく，分子動力学計算や他の情報工学の問題などにも幅広く適用できると思われる．



## 第 6 章

### 第 I 部のまとめ

第 I 部では、生体分子のコンピュータ・シミュレーションの基礎概念となる配座エネルギー関数と座標系、そして Hessian 並列計算法の開発を中心に、二面角系における生体分子の構造モデリング・ダイナミクス解析システムの構築について述べた。

このシステムは、生体分子の構造モデリング、Poisson-Boltzmann 分子動力学法による構造サンプリングと自由エネルギー解析、二面角基準振動モード解析法によるダイナミクス解析などを、並列計算技術を用いて統合的に、かつ大規模分子に対しても高速に行うことができる。

構造モデリングについては、エネルギー最小化をベースにした Two-Stage 法を開発し実装した。また自由エネルギー解析には、構造サンプリングのために Poisson-Boltzmann 分子動力学法、サンプリングした構造群から自由エネルギーを計算するために WHAM 法を用いた。ダイナミクス解析には、二面角系基準振動モード解析法を用いた。

さらに、配座エネルギー計算、1 次・2 次微分計算、行列計算、基準振動モード計算などの並列計算アルゴリズムを実装した。これにより、大規模な分子の解析を高速に行うことが可能となった。

Hessian 並列計算については、計算量を通常の  $n^4$  から  $n^2$  に減らした Gō らによるアルゴリズムをベースに、Unify 方式と呼ばれる並列計算アルゴリズムを開発した。Unify 方式を tRNA<sup>Gln</sup> および GlnRS の 2 種類の分子の Hessian 計算に適用した結果、tRNA<sup>Gln</sup> では 24 プロセッサを用いたときに 6.8 倍、GlnRS では 54 プロセッサを用いたときに 11.2 倍のスピードアップが得られた。また Hessian 計算をアルゴリズム内に含むニュートン・ラフソン法によるエネルギー極小化計算では、tRNA<sup>Gln</sup> で 6.3 倍、GlnRS で 12.0 倍のスピードアップであった。

Unify 方式はプロセッサ数が大きいところでは通信コストの増大による性能低下がみられたために、サブタスクグループ方式と呼ばれるアルゴリズムで、さらなる高

速化を試みた．サブタスク間の依存関係の解析によって必要なデータのみ通信を行い，またサブタスクのプロセッサへの割り当て方を 2 段階の遺伝的アルゴリズムで最適化した結果，大幅な性能向上がみられた． $tRNA^{Gln}$  にこの方法を適用したところ，60 プロセッサを用いたときに 32.6 倍のスピードアップを達成した．サブタスク実行時間の見積りを行うためのシミュレータの性能を向上させることで，さらなる高速化を図ることができると考えられる．

## 第 II 部

# DNA ミニヘアピン分子の構造モデリングと 熱安定性解析



## 第 7 章

### 第 II 部の目的

第 II 部では、非常に高い熱安定性をもつことが知られている DNA ミニヘアピン分子、およびそのループ部の塩基置換体に対して、筆者らが開発した生体分子構造モデリング手法 (Two-Stage 法) を適用して構造モデリングを行い、ついでモデリングした構造をもとに分子動力学計算と WHAM (Weighted Histogram Analysis Method) を用いて、DNA ミニヘアピン分子のヘアピン形成自由エネルギーの配列依存性解析を行った結果について述べる。

核酸分子は、多くの場合二重らせん構造をとっているが、環境などによってそれ以外の部分構造をとることも多い。ヘアピン構造はそのような核酸分子の部分構造の 1 つで、核酸の機能と深い関わりがあることが多くの研究によって示唆されている。DNA ヘアピン構造が、topoisomerase II<sup>63)</sup>、HIV-1 integrase<sup>64)</sup>、SMC (structure maintenance of chromosomes)<sup>65)</sup> などの多くのタンパク質のインタラクション部位になっているという報告がある。またメッセンジャー RNA の C(UUCG)G 配列を逆転写酵素が読めないという指摘があり<sup>66)</sup>、この配列は安定なヘアピン構造をとると考えられている。1 本鎖の環状 DNA をゲノムにもつ G4 フェージ DNA では、複製開始領域に安定なヘアピン構造をとることが知られている GCGAAAGC 配列をもつ<sup>67)</sup>。

核酸ミニヘアピン分子は、3~4 塩基程度の小さなヘアピンループと 2~4 塩基対程度の短いステムから成るコンパクトな構造をもつ 1 本鎖のオリゴヌクレオチドである。核酸中のヘアピン構造のみをいわば取り出したような分子であるため、この核酸ミニヘアピン分子の塩基配列と熱安定性の関係解析をすすめることによって、核酸ヘアピン構造の熱安定性や核酸分子の立体構造予測などに重要な知見が得られることが期待される。これまで、生化学的手法を用いて求められた熱力学的パラメータから、DNA および RNA について、とくに高い熱安定性を示す特定の配列が数種類報告されている。

Antao らは、DNA および RNA の 4 塩基ループについて、ループ部の塩基配列とループを閉じる塩基対 (loop-closing base pair) をいろいろに変えたときの熱力学的パラメータを求め、rGGAC(UNCG)GUCC、dGATC(GNRA)GATC などが安定であることを明らかにした<sup>68,69)</sup> (塩基配列表記のかっこはループ部分を表す。また先頭の「r」は RNA、「d」は DNA を表す)。Hirao らは、dGC(GAAA)GC、dGC(GAA)GC およびそれらの塩基置換体の融解温度や電気泳動の際の泳動度を網羅的に調べ、安定な DNA 配列に対応する配列をもつ RNA の安定性が低いことを報告している<sup>70)</sup>。また Yoshizawa らは dGC(NNN)GC の 3 塩基ループの電気泳動の際の泳動度を調べ、GNA がもっとも安定であり、また高いヌクレアーゼ耐性をもつことを明らかにしている<sup>19)</sup>。これらの研究により安定であるとされた配列は、生体内でもリボゾーム RNA、ファージ DNA など多くの DNA、RNA から発見されている。またミニヘアピン分子の応用として、mRNA の 3' 末端に安定なヘアピン構造を含む「キャップ DNA」を付加することによって mRNA を安定化し、細胞外タンパク質合成系における翻訳効率の上昇を図るという応用も試みられている<sup>71)</sup>。最近さまざまな生物種において全ゲノムの塩基配列決定が盛んに行われているが、塩基配列決定の過程で特定の配列がたまたま 3' 末端にあらわれたときに band compression による読み間違いが起きることが DNA、RNA の双方について報告されているが、このような配列の多くはこれらの安定なミニヘアピン分子と同じ塩基配列をもつという興味深い事実もある<sup>72,73)</sup>。

いくつかのミニヘアピン分子は、NMR によって立体構造が決定されており、ステムに近いループ部の塩基の非ワトソンクリック塩基対やループ部の塩基のステム部の塩基上へのスタッキングなど、安定なヘアピン構造のもつ特徴が明らかになった<sup>74-76,20)</sup>。最近では、Ennifar らが RNA の UUCG タイプのループについて X 線結晶解析を行い、NMR による構造との違いを議論している<sup>77)</sup>。また Amri らは、1 塩基から成るループの安定性について議論している<sup>78)</sup>。

コンピュータを用いた核酸ミニヘアピン分子の安定性についての研究は、これまでいくつかのタイプの分子について行われてきた。構造モデリングの例では、Erie らが、Monte Carlo 法とエネルギー極小化を用いて DNA の  $CT_nG$  ループと  $CA_nG$  ループ ( $n = 3, 4, 5$ ) の配座エネルギー値を解析し、同じ塩基配列でも複数のコンフォメーションをとる可能性があることを示唆している<sup>79,80)</sup>。また Kratky は、JUMNA プログラムと Bremermann optimization method と呼ばれる方法を用いて RNA の 4 塩基ループ (GNNA) および RNA の 3 塩基ループ (UUU, RUU, UUR) をモデリングし、ループ配列を変化させたときの全体構造の変化が非常に小さいことを見出ししている<sup>81)</sup>。ミニヘアピン分子の安定性を議論した研究としては、Singh らの分

子動力学計算によるアプローチがある<sup>82)</sup>。これは RNA の UUCG タイプのループの C を U に塩基置換したときの自由エネルギー変化を、分子動力学計算を用いた Free Energy Perturbation 法により見積もったものである。

Erie らや Kratky による研究は、ミニヘアピン分子のループ部の塩基配列を変えたときの配座エネルギーの違いをシステムティックに解析した意義深いものである。しかし用いている構造モデリング方法はその配列がヘアピン構造をとると仮定したものであり、構造探索範囲は限られたものとなっている。また Singh らによる自由エネルギー変化の計算は、安定なヘアピン同士のわずかな安定性の違いについて議論したものである。ミニヘアピン分子の熱安定性解析をさらにすすめるためには、安定なヘアピンを形成する配列とそうでない配列の違いについても議論する必要があると考えられる。

これらをふまえて、第 II 部では、DNA ミニヘアピン分子 dGC(GAA)GC の熱安定性に対するループ塩基配列の依存性を、構造モデリングと分子動力学法を用いた自由エネルギー計算を用いて解析する。

DNA ミニヘアピン分子 dGC(GAA)GC は、ステム部が 2 塩基対、ループ部が 3 ヌクレオチドであり、もっとも小さなミニヘアピン分子の 1 つである。先に述べた Yoshizawa らによる研究は、3 番目から 5 番目までのループ部の 3 つの塩基を  $4^3 = 64$  通りのすべてに変えたときの電気泳動の泳動度を調べたもので、これにより 3 番目と 5 番目の塩基がそれぞれ G, A であることが、高い熱安定性を獲得するための必要十分条件であることが明らかになった<sup>19)</sup>。また、この分子の重水素置換速度が B-DNA や Z-DNA に比べて非常に遅いことからその安定性を議論した研究<sup>83)</sup> や、DNA 1 本鎖の 3' 末端にこの分子をつけたときの 3' exonuclease 耐性と hybridization の解析から、このヘアピン分子の open-close の安定性を議論した研究もある<sup>84)</sup>。

DNA ミニヘアピン分子 dGC(GAA)GC の立体構造は、Hirao らによって、NMR で決定されている<sup>20)</sup>。この構造を Fig. 7.1 に示す。また構造の模式図を Fig. 7.2 に示す。この分子には、以下のような構造上の特徴がある。

- G3-A5 が非ワトソクリック塩基対を形成し、loop closing base pair となる (Fig. 7.2 右)。
- パッカリングはすべて C2'-endo (B-form) である。
- A4 と A5 の間の糖 - リン酸骨格に「折れ」がある。
- G3 と A4 はスタッキングしている。これが非常に高い熱安定性の要因であると考えられている。

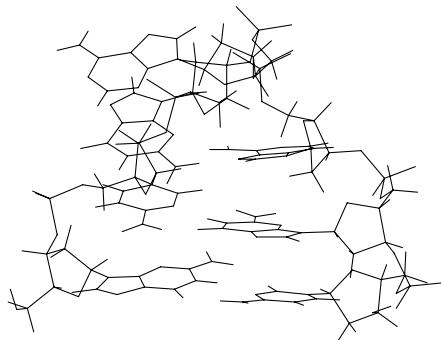


Figure 7.1 NMR によって決定された dGC(GAA)GC の構造 .

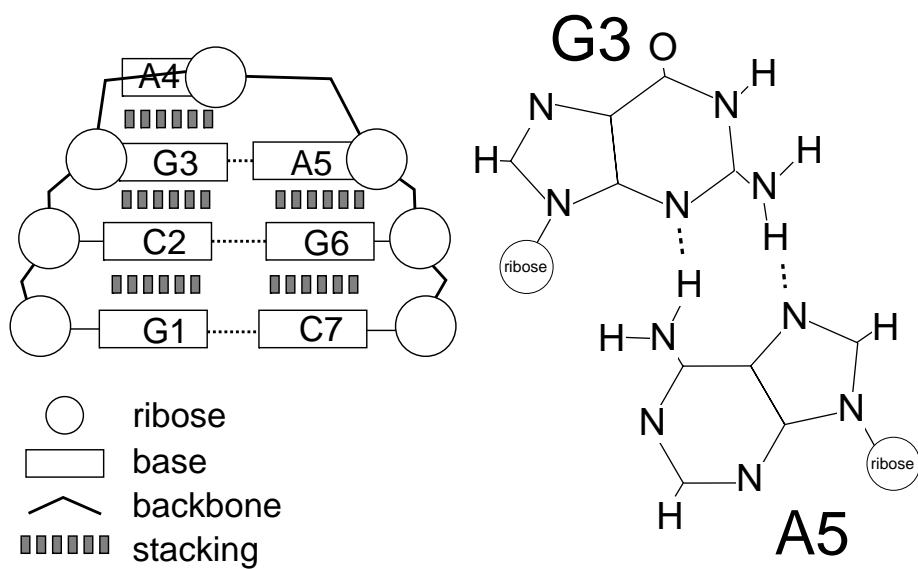


Figure 7.2 (左) dGC(GAA)GC の構造の模式図 . (右) G3-A5 (非ワトソンクリック塩基対) の模式図 .



また安定なヘアピンを形成する他のループ配列 GGA, GTA, GCA についても, NMR による解析から GAA と似た構造をとることが示唆されている.

本研究では, 安定なヘアピン構造を形成する dGC(GNA)GC (N=A, G, T, C) と, ヘアピンを形成しない dGC(GAC)GC, dGC(GAG)GC を対象とする. NMR で原子座標が決定されているのは GAA ループだけであるので, そのほかの構造を, 筆者らが開発した Two-Stage 法と呼ばれる方法を用いてモデリングする. これは, どの塩基とどの塩基が塩基対を形成するかの仮定なしに, 配座エネルギー最小化によって構造モデリングを行うところに大きな特徴がある. ついで Poisson-Boltzmann 方程式法によって静電遮蔽効果を考慮した分子動力学計算をもとに, モデリングしたミニヘアピン分子のヘアピン形成自由エネルギープロファイルを WHAM (Weighted Histogram Analysis Method) を用いて作成し, ループ配列によるプロファイルの比較を行う. Poisson-Boltzmann 方程式法は計算対象となる分子周辺の溶媒を連続体近似して扱う方法で, 水やイオンからの静電遮蔽効果を高速に計算することができる. WHAM は複数の分子動力学計算を用いてサンプリングされた構造群から, 自由エネルギープロファイルを求める方法である. これらを用いて,

1. 安定なヘアピン構造を形成する dGC(GNA)GC と, ヘアピン構造を形成しない配列 dGC(GAC)GC, dGC(GAG)GC のモデリング構造および配座エネルギー値に違いがみられるか.
2. 両者のヘアピン形成自由エネルギープロファイルに違いがあるか.

について考察する.

## 第 8 章

# Two-Stage 法の開発とミニヘアピン分子の構造モデリング

生体分子のうち、タンパク質の構造モデリング手法は、モデリング対象のタンパク質とホモロジーが高い構造既知のタンパク質を鋳型として構造を組み上げるもの (comparative modeling), 全体のフォールドが似ている既知構造を探索するもの (fold recognition), 配座エネルギー関数の最小化により第一原理的に構造をビルディングするもの (global optimization) など、さまざまなものが開発されている。これに対して核酸では、2次構造予測は Zuckerkandl アルゴリズム<sup>85)</sup> などが広く利用されているものの、立体構造については、有効なモデリング手法がいまだ存在しないのが現状である。

最近では 16S rRNA など原子レベルでの核酸構造情報も次第に増えてきているが、タンパク質ほど既知構造のデータベースは充実しておらず、データベースをもとにしたアプローチは有効ではない。そこで、配座エネルギー関数を最小化する手法の適用を考える。

### 8.1 エネルギー最小化による構造モデリング法

配座エネルギー最小化による構造モデリングの方法は、native 構造が (もっとも低いかどうかは議論の余地があるが) 非常に低い配座エネルギー値をとることを前提に、配座エネルギー関数の最小点 (グローバルミニマム) を探索し、第一原理的に構造モデリングを行うものである。

エネルギー最小化による生体分子の構造モデリングに伴う困難は、以下の 2 点である。

1. 探索すべき配座空間が膨大であること

2. 多数のローカルミニマムが存在し、グローバルミニマムへの到達が困難であること (multiple minima problem)

1. は、仮に各ヌクレオチドが 3 通りの離散的なコンフォメーションしかとらないとしても、ヌクレオチド数 100 の核酸の構造は  $3^{100} \sim 10^{48}$  通りと非常に多いことから明らかである。2. は、初期構造をわずかに変えてエネルギー極小化を行うと、異なる初期構造によって別々の構造がエネルギー極小構造として得られることからわかる (Fig. 8.1)。これをエネルギー極小化の初期値依存性と呼ぶ。

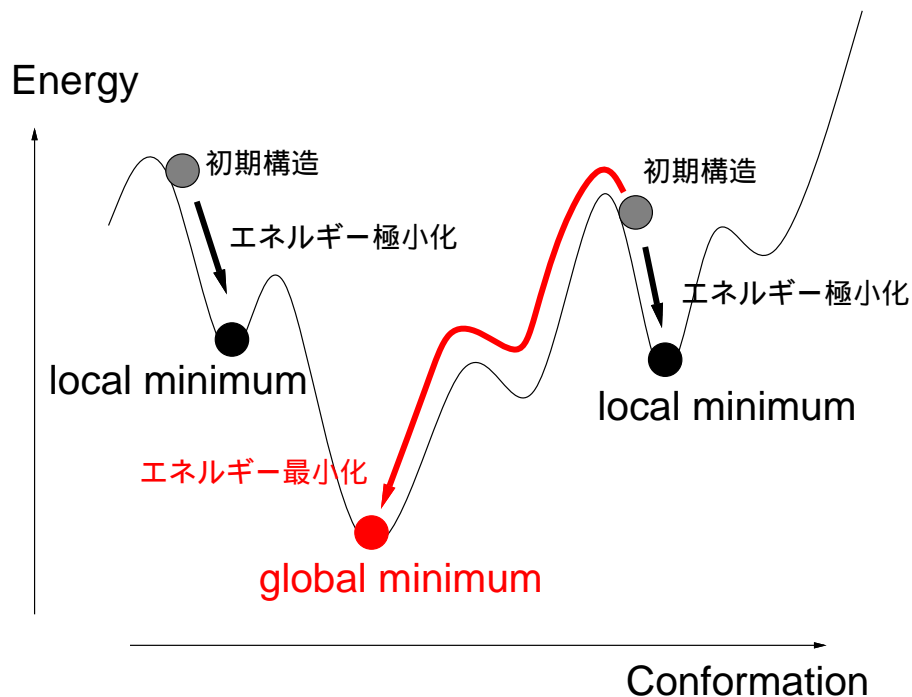


Figure 8.1 Multiple minima problemとエネルギー極小化・最小化の概念図。エネルギー極小化では、初期値によって異なるローカルミニマムに到達する。エネルギー最小化は、ローカルミニマムにとらわれることなく、グローバルミニマムに到達できる。

このため、エネルギー最小化によって生体分子の構造をモデリングするためには、ローカルミニマムにとらわれずにグローバルミニマムに到達する工夫が必要である。これを実現する方法として、シミュレーテッド・アニーリング法に代表される Monte Carlo 法を基礎とした方法が広く用いられている。これに対して、エネルギー曲面を変形し、直接的にグローバルミニマムへの到達を容易にする方法も開発されている。その中の一つである拡散方程式法 (Diffusion Equation Method, DEM) は、拡散方程式を利用してエネルギー関数を平滑化し、エネルギー関数がただ 1 つのローカルミニマム (= グローバルミニマム) をもつ状態を作り出してから、平滑化した関数を

次第にもとの関数に戻すというアルゴリズムで，もとの関数のグローバルミニマムを探索する手法である．この方法は，アルゴリズムに Monte Carlo 法のようなランダムな要素がなく，高速に探索が行えるという利点がある．筆者らは，この拡散方程式法に周辺探索を加えた新たな手法 Two-Stage 法を開発し，拡散方程式法の限界を大幅に広げた<sup>21)</sup>．本研究では，この Two-Stage 法を用いて，DNA ミニヘアピン分子の構造モデリングを行う．

## 8.2 Two-Stage 法の概要

拡散方程式法 (DEM)<sup>86-88)</sup> および Two-Stage 法<sup>21)</sup> の概略について述べる．Two-Stage 法のアルゴリズムの詳細については，Appendix A に記載する．

前節で述べたように，生体分子の配座エネルギー関数の形状は非常に複雑であり，多数のローカルミニマムが存在する．これを平滑化して，ミニマムが 1 つまたは非常に少数である状態を作れば，そのミニマムを探索することは従来のエネルギー極小化の方法を用いて容易にできる．

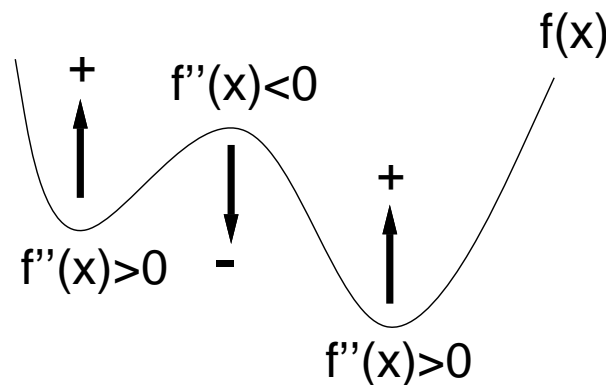


Figure 8.2 関数とその関数の 2 次微分を用いて平滑化する概念図．

ある関数  $f(x)$  を平滑化することを考える．Fig. 8.2 のように， $f(x)$  の「上に凸」の部分に負の値，「下に凸」の部分に正の値を加えると， $f(x)$  は平滑化される．ここで  $f(x)$  の 2 次微分  $f''(x)$  に着目すると， $f''(x)$  は  $f(x)$  が上に凸のところでは負，下に凸のところでは正の値をもつ．したがって， $f(x)$  に  $f''(x)$  を加えると， $f(x)$  は平滑化されることになる．

さらに，仮想的な時間  $t$  という量を導入する． $f(x)$  を  $x$  と  $t$  の関数  $F(x, t)$  に拡張し，次式のように時刻  $t$  から  $\Delta t$  だけ時間が経つと， $\Delta t$  倍の 2 次微分  $\nabla^2 F(x, t)$  を

もとの関数に加えることにする．つまり，時間が経過すればするほど，平滑化を進めるということである．

$$F(\mathbf{x}, t + \Delta t) = F(\mathbf{x}, t) + \nabla^2 F(\mathbf{x}, t) \times \Delta t \quad (8.1)$$

これを变形すると，

$$\nabla^2 F(\mathbf{x}, t) = \frac{F(\mathbf{x}, t + \Delta t) - F(\mathbf{x}, t)}{\Delta t} \quad (8.2)$$

となる．これは次式の拡散方程式と呼ばれる偏微分方程式に対応する．拡散方程式は，水面に立つ波が水面を広がっていく様子や金属板の端を熱したときの熱の伝わりなどを表すことができる．時刻  $t = 0$  のときの初期条件を式 (8.4) のように  $f(\mathbf{x})$  として与えると，時刻  $t$  が大きくなるほど，もとの関数  $f(\mathbf{x})$  をより平滑化した関数  $F(\mathbf{x}, t)$  が方程式の解となる．

$$\nabla^2 F(\mathbf{x}, t) = \frac{\partial F(\mathbf{x}, t)}{\partial t} \quad (8.3)$$

$$F(\mathbf{x}, 0) = f(\mathbf{x}) \quad (8.4)$$

今の場合， $\mathbf{x}$  は分子のコンフォメーション（原子座標）であり， $f(\mathbf{x})$  はもとのエネルギー関数である．

拡散方程式では，平滑化された関数（ある時刻  $t$  での方程式の解）のエネルギー極小点は，もとの関数とは位置が異なる．このため，もとの関数のグローバルミニマムを見つけるために，以下のような手順（time-reversing procedure）を行う．

1. 時刻  $t$  を十分に大きな値 ( $t = t_{max}$ ) とする．このときエネルギー関数は，十分に平滑化され，唯一の極小点をもっているとする．この極小点は従来のエネルギー極小化の方法で容易に到達可能である．
2. この極小点を初期値として，時刻  $t$  をある小さな値  $\delta t$  だけ戻した（小さくした）ときの  $F(\mathbf{x}, t_{max} - \delta t)$  での極小点を求める．
3. さらにその極小点を初期値として  $F_i(\mathbf{x}, t_{max} - 2\delta t)$  での極小点を求める．
4. 手順 3 の操作を， $t$  を  $\delta t$  ずつ減らしながら順次行い，最終的に求まる  $F(\mathbf{x}, 0) = f(\mathbf{x})$  での極小点が，エネルギー最小点（グローバルミニマム）になると考える．

Time-reversing procedure の模式図を Fig. 8.3 に示す． $t = t_{max}$  のときの唯一の極小点が  $t = 0$  でのグローバルミニマムに成長すれば，この time-reversing procedure によってもとの関数のグローバルミニマムを見つけることができる．拡散方程

式法はアルゴリズムに確率的な要素を含んでおらず、決定論的に計算を進めることができるため、高速な計算が可能である。

しかし実際に計算を行ってみると、この time-reversing procedure の過程で、現在いる極小点の周辺にさらに深い極小点が成長し、結果としてグローバルミニマムに到達できないことが多い。このため、time-reversing procedure の各ステップで、現在いる極小点の周辺を探索し、さらに深いミニマムがあればそちらへ移るというプロセスを加える。これを Two-Stage 法と呼ぶ (Fig. 8.4)。周辺探索は、 $t > 0$  においては極小点でのエネルギー 2 次微分行列 (Hessian) の固有ベクトル方向に最急降下法とニュートン・ラフソン法による探索を行う (line search)。また  $t = 0$  においては、Monte Carlo Minimization 法<sup>89)</sup> を改良した MCMS 法 (Monte Carlo Minimization with Saturated non-bonded energy) を用いる。Monte Carlo Minimization 法は、Monte Carlo 法の各ステップにおいてエネルギー極小化を行うもので、エネルギー極小点のみをたどる Monte Carlo 法に相当する。MCMS 法は、van der Waals ポテンシャル、水素結合ポテンシャル、静電ポテンシャルの  $r$  が小さいところを  $r$  の 4 次関数で平滑化し、さらにコンフォメーション探索を Hessian の固有ベクトルの方向を考慮しながら行うものである。以上により、拡散方程式法では到達できなかった深いミニマムを探索することができる。なお、この平滑化された曲面での Hessian を用いた line search 法が有効であることは、のちに Pappu らによっても確認されている<sup>90,91)</sup>。

同じミニヘアピン分子について拡散方程式法と Two-Stage 法で実際にエネルギー最小化を行ったときのエネルギー遷移の例を Fig. 8.5 に示す。

### 8.3 初期構造の構築と計算パラメータ

構造モデリングの初期構造は、以下のようにして構築した。例として、dGC(GAA)GC の初期構造を Fig. 8.6 に示す。

1. NMR で決定された dGC(GAA)GC の構造に、AMBER4.0 の共役勾配法によるエネルギー極小化を 5000 ステップ適用する。
2. ループ部の二面角をランダムな値に設定し、引き延ばした構造を作成する。これを dGC(GAA)GC の初期構造とする。
3. dGC(GAA)GC の初期構造から、A4 を G, T, C にそれぞれ置き換え、1000 ステップの共役勾配法によるエネルギー極小化を適用する。これらを、dGC(GGA)GC, dGC(GTA)GC, dGC(GCA)GC の初期構造とする。

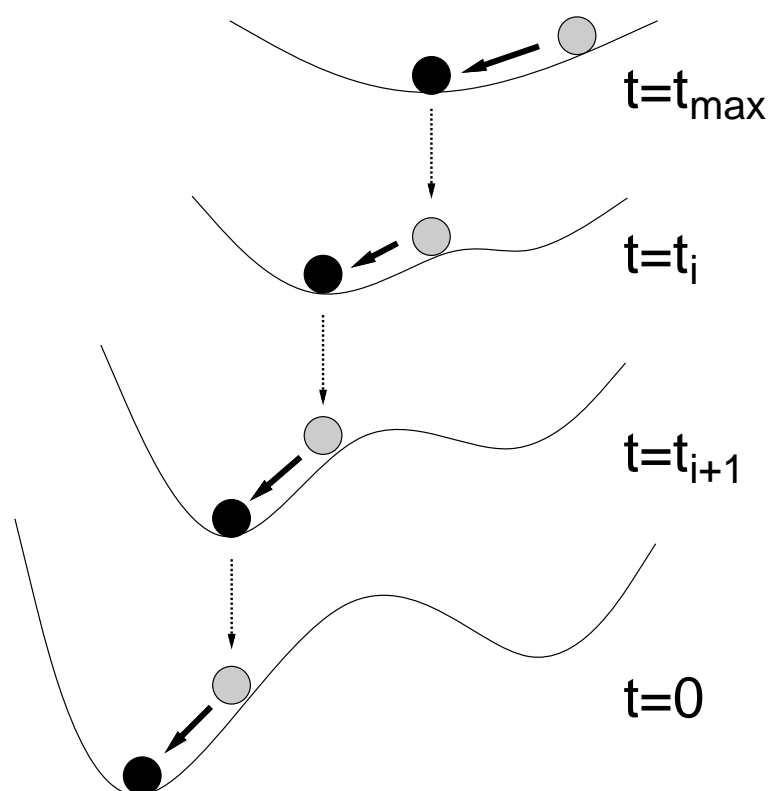


Figure 8.3 Time-reversing procedure の概念図 .

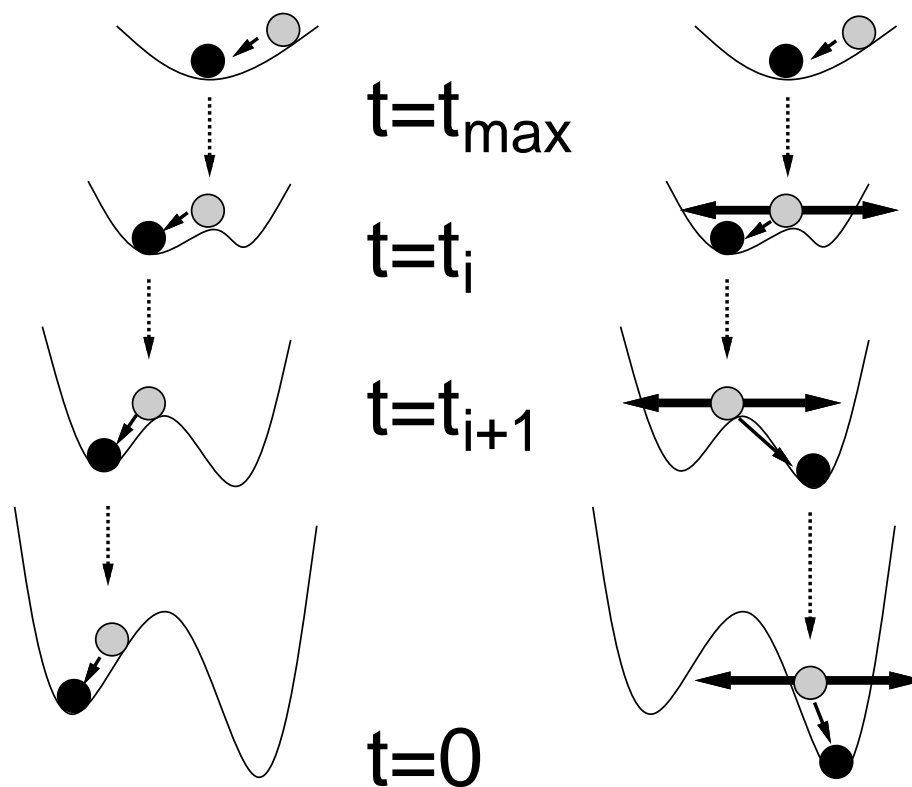


Figure 8.4 拡散方程式法と Two-Stage 法の time-reversing procedure における概念図 . (左) 拡散方程式法 . (右) Two-Stage 法 . Two-Stage 法では , 周辺探索によって , より深いミニマムが見つかったとそちらを採用する .



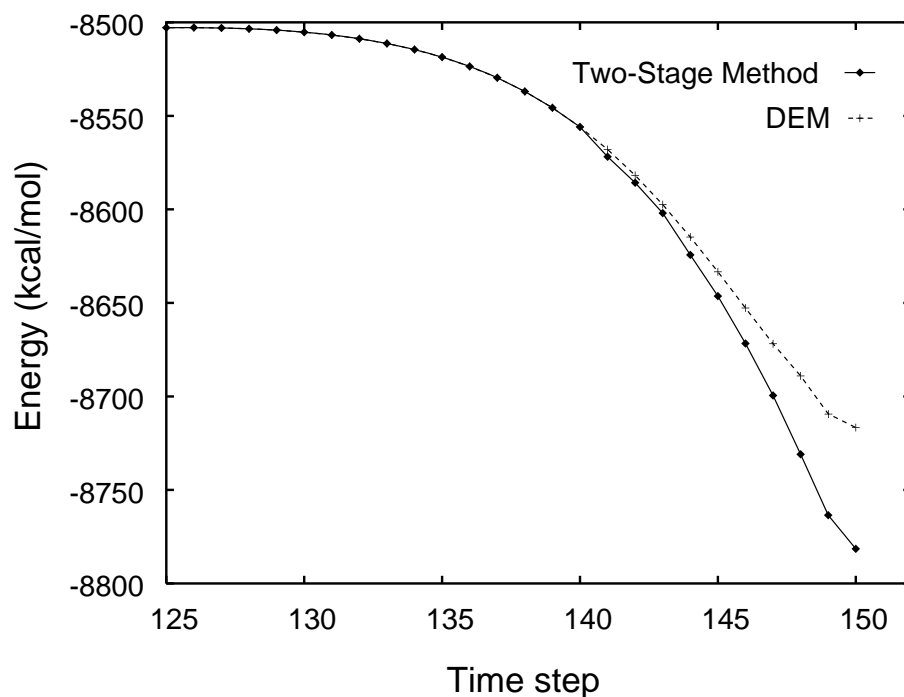


Figure 8.5 拡散方程式法 (DEM) および Two-Stage 法によるエネルギー最小化でのエネルギー遷移の例。破線が拡散方程式法，実線が Two-Stage 法である。Two-Stage 法では 141 ステップ目で周辺探索によって，より低いエネルギー値をもつミニマムが見つかり，そこで分岐が起きて，最終的に拡散方程式法よりも低いエネルギー値をもつミニマムに到達できていることがわかる。

4. dGC(GAA)GC の初期構造から , A5 を C, G にそれぞれ置き換え , 1000 ステップの共役勾配法によるエネルギー極小化を適用する . これらを , dGC(GAC)GC, dGC(GAG)GC の初期構造とする .

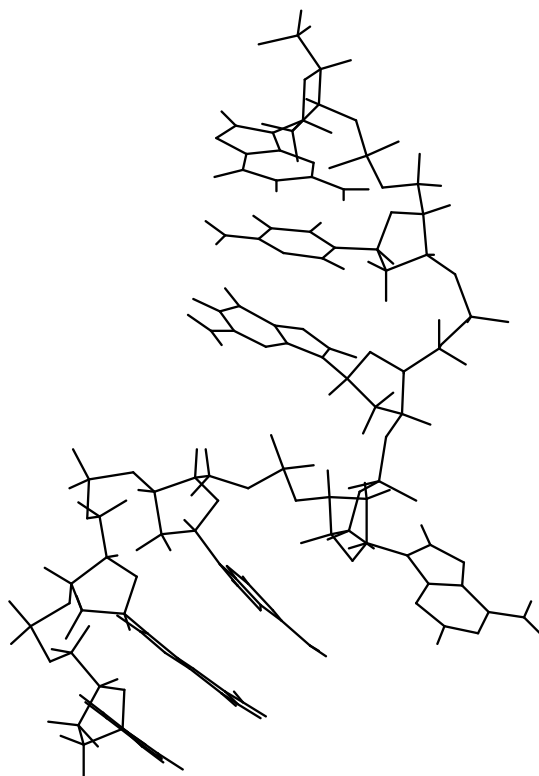


Figure 8.6 dGC(GAA)GC の初期構造 .

これらの初期構造を用いて , Two-Stage 法によるエネルギー最小化を行う . 計算に用いたパラメータは以下のとおりである .

- フォースフィールドは AMBER の parm91 を用いる . Two-Stage 法では , 拡散方程式の解を解析的に求めるために , エネルギー関数をガウス関数の和で近似する (Appendix A 参照) .
- $t_{max} = 5000.0$  とする .
- Time-reversing procedure は 600 ステップとする .
- 誘電率は , Two-Stage 法では  $\epsilon = 4r$  (ただし  $r$  は原子間距離) , それ以外では  $\epsilon = 80$  とする .

## 8.4 構造モデリング結果

### 8.4.1 dGC(GNA)GC のモデリング

dGC(GNA)GC のエネルギー最小化によるモデリング結果を示す。GAA については、NMR 構造を初期構造として AMBER4.0 の共役勾配法によるエネルギー極小化を行った構造と、本研究でのエネルギー最小化による最終構造との重原子について計算した最小 2 乗平均誤差 (root mean square deviation, r.m.s.d.) は  $0.42 \text{ \AA}$  であり、これはエネルギー最小化によって GAA のヘアピン構造が正しくモデリングできたことを表している。エネルギー最小化の過程を Fig. 8.7 に示す。

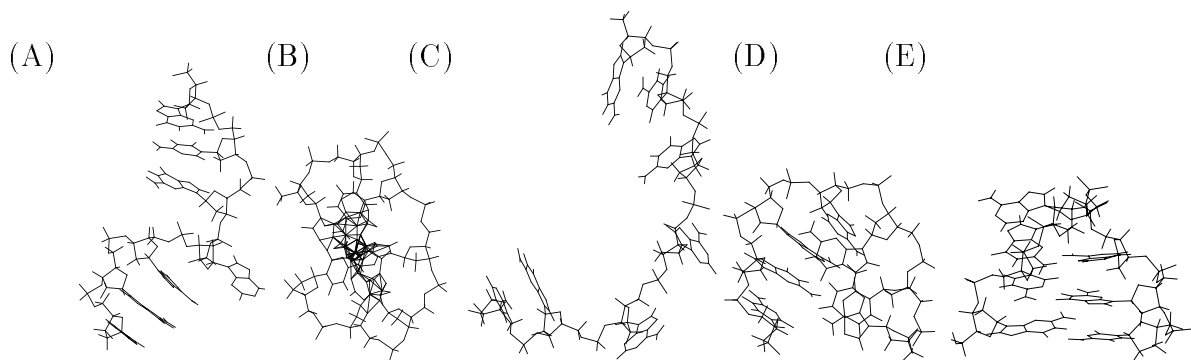


Figure 8.7 dGC(GAA)GC のエネルギー最小化の過程。(A) 初期構造。(B) Two-Stage 法が 87.5% 終了したときの構造。(C) Two-Stage 法が 93.8% 終了したときの構造。(D) Two-Stage 法で  $t = 0$  となったときの構造。(E) MCMS 法による周辺探索を行った最終構造。

また、dGC(GGA)GC, dGC(GTA)GC, dGC(GCA)GC のモデリング結果は、Fig. 8.8 のようになった。

いずれの構造も、loop closing base pair や糖 - リン酸骨格の折れ曲がりについての情報を与えていないにも関わらず、G3 と A5 の間の非ワトソンリック結合による loop closing, G1 から A4 および A5 から C7 へのスタッキング, A4 と A5 の間の糖 - リン酸骨格の急激な折れ曲がり、コンパクトな立体構造など、NMR 解析から得られている構造の特徴をもつものであった。

G3 と A5 の非ワトソンリック結合は、GAA の場合と比較すると GGA, GTA, GCA のいずれの場合も r.m.s.d. が  $0.57 \text{ \AA}$  であった。また、4 番目の塩基 (G, T, C) が G3 の上にスタックする位置は、いずれの場合もほとんど変わらず、この位置がエネルギー的に有利であると考えられる。

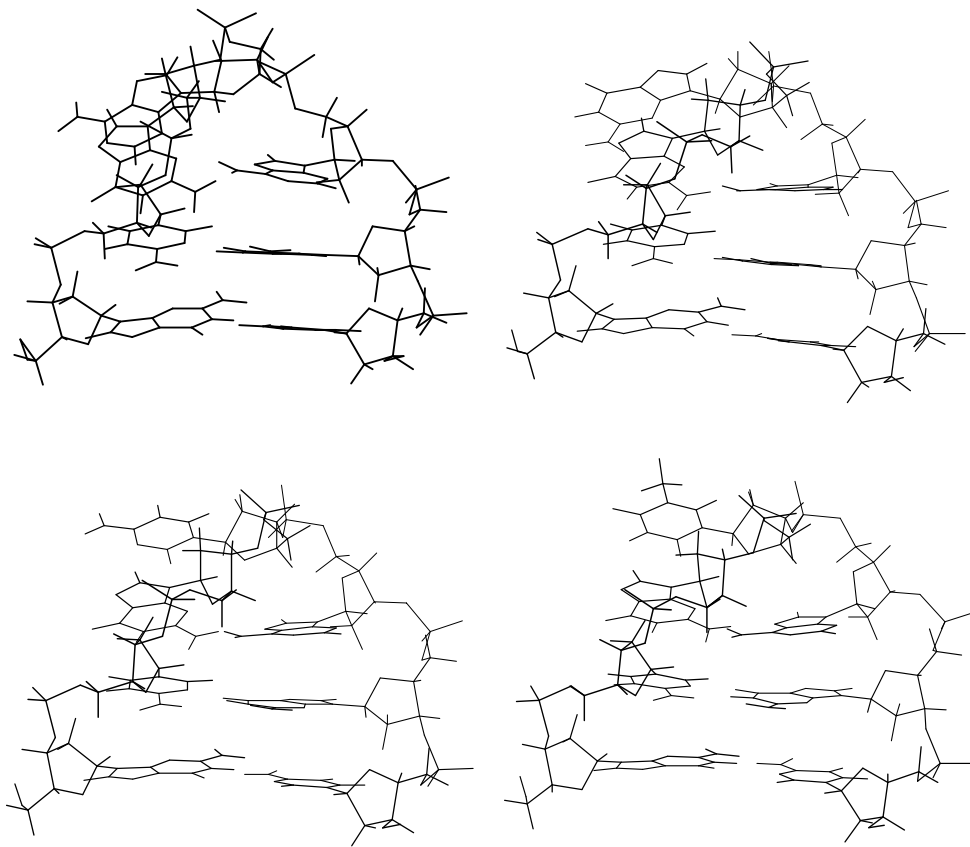


Figure 8.8 dGC(GAA)GC およびそのループ部置換体のエネルギー最小化によるモデリング構造 . (左上) ループ部配列が GAA . (右上) GGA . (左下) GCA . (右下) GTA .

## 8.4.2 dGC(GAC)GC のモデリング

ループ部の塩基配列が GAC の場合のモデリング結果を Fig. 8.9 に示す．生化学実験によるとこの配列は安定なヘアピン構造は形成しないとされているが，エネルギー最小化によるモデリング構造はヘアピン構造になった (Fig. 8.9 A)．Loop closing base pair は G3-C5 となったが，ワトソクリック結合ではなく，C1' 間の距離がワトソクリック結合よりも小さい結合様式となった (Fig. 8.9 B)．また A4 は G3 の上にスタックした．Bhaumik らは分子動力学計算によってワトソクリック結合した 3 塩基対から成るステムの上に 1 つの塩基がスタックした構造は安定に存在できないことを見出している<sup>92)</sup>．本研究によるモデリングで G3-C5 が非ワトソクリック結合になったことは，これを裏付けるものとなっている．

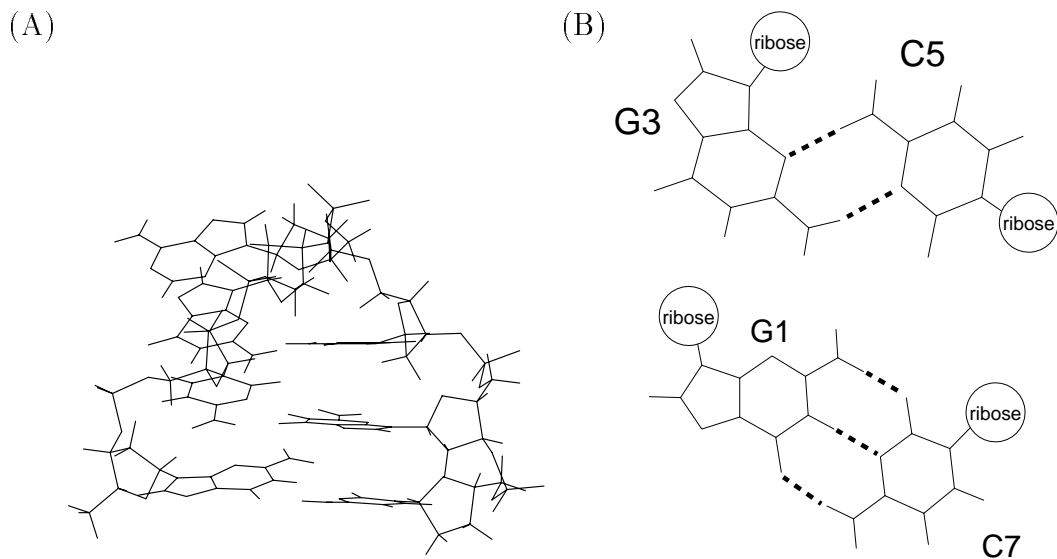
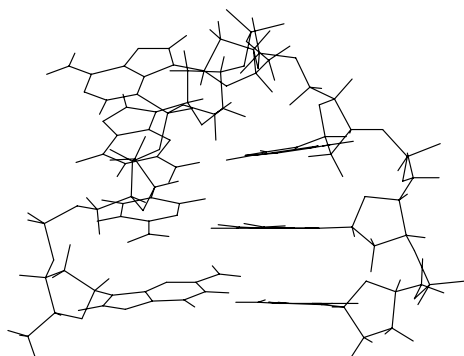


Figure 8.9 (A) dGC(GAC)GC のモデリング結果．(B 上) G3-C5 の結合様式．C1' 同士の距離は 8.1 Å．(B 下) 通常の G-C のワトソクリック結合様式．C1' 同士の距離は 10.6 Å．いずれも破線は水素結合を表す．

## 8.4.3 dGC(GAG)GC のモデリング

ループ部の塩基配列が GAG の場合のモデリング結果を Fig. 8.10 に示す．生化学実験によるとこの配列は GAC と同様，安定なヘアピン構造は形成しないとされているが，エネルギー最小化によるモデリング構造はヘアピン構造になった．Loop closing base pair は G3-G5 となり，両者の C1' 間の距離は，GAA の G3-A5 における C1' 距離とほぼ同じであった．また A4 は G3 の上にスタックした．

(A)



(B)

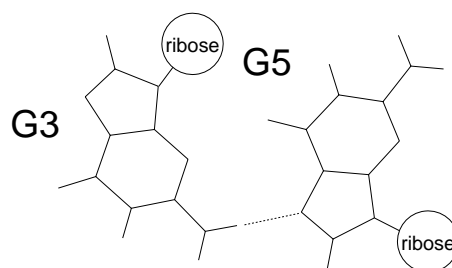


Figure 8.10 (A) dGC(GAG)GC のモデリング結果 . (B) G3-G5 の結合様式 . C1' 同士の距離は 8.0 Å . 破線は水素結合を表す .

### 8.5 ループ部のエネルギー解析の結果

モデリングを行った配列は、いずれもループ部の塩基が dGC(GAA)GC と 1 塩基異なっているだけであるので、ヘアピン形成におけるエンタルピー効果の違いの大部分は、ループ部の塩基間の相互作用によるものと考えられる。そこで、各モデリング構造についてループ部の塩基間のエネルギー値を求めたものが、Table 8.1 である。

Table 8.1 各ループ配列におけるヘアピンモデリング構造のエネルギー値 (kcal/mol) .

loop seq.	GAA	GGA	GTA	GCA	GAC	GAG
energy	-73.4	-73.2	-71.8	-71.9	-71.2	-70.0

安定なヘアピン構造を形成する配列同士を比較すると、GGA のエネルギー値は GAA とほぼ同程度であるのに対して、GTA, GCA はそれよりも 1.5 kcal/mol 程度高いエネルギー値となった。また、安定なヘアピンを形成しない GAC のエネルギー値は、ヘアピンを形成する GTA, GCA と同程度であった。ヘアピンの安定性を議論する場合は、ヘアピンを形成していない状態とヘアピンを形成した状態のエネルギー差が重要である。ヘアピンを形成していないランダム状態では、塩基間相互作用は隣りあう塩基同士によるものがほとんどであると考え、dGC(GCA)GC と dGC(GAC)GC はともに 4 つの G-C 間相互作用、1 つの G-A 間相互作用、1 つの A-C 間相互作用

をもつため、ランダム状態でのエネルギー値は同程度であると考えられる。ヘアピン形成時にもエネルギー値に差がないことを考えると、ヘアピン形成におけるエネルギー（エンタルピー）の効果は GCA, GAC でほぼ同程度である。また、GAG のエネルギー値は GGA よりも 3.2 kcal/mol, GTA よりも 1.8 kcal/mol 高かった。GCA と GAC のときと同様に、dGC(GGA)GC と dGC(GAG)GC はともに 3 つの G-C 間相互作用, 2 つの G-A 間相互作用, 1 つの G-G 間相互作用をもつため、ランダム状態でのエネルギー値は同程度であると考えられる。ループ部の塩基間相互作用の大きさ全体に対する割合を考慮すると、GAG と GGA のエネルギー差は有意であるとはいえない。これらの結果は、ヘアピンの安定性の差はエネルギー（エンタルピー）の解析だけでは不十分であることを示している。

## 8.6 構造モデリング結果のまとめ

Two-Stage 法を用いたエネルギー最小化による構造モデリングの結果、dGC(GAA)GC の NMR 構造に非常に近い構造がモデリング結果として得られ、また安定なヘアピン構造をとることが知られている dGC(GGA)GC, dGC(GTA)GC, dGC(GCA)GC について、NMR 解析から示唆されている GAA と似た構造が得られた。これに対し、安定なヘアピン構造を形成しない dGC(GAC)GC, dGC(GAG)GC についても、エネルギー最小構造としてヘアピン構造が得られた。またループ部の塩基間の相互作用の大きさをエネルギー値から見積もったところ、安定なヘアピンを形成しない GAC で GCA, GTA と同程度、また GAG の場合も安定なヘアピンを形成する配列と大きく変わらないという結果となった。このことはヘアピンの安定性における、エンタルピー以外のエントロピーの影響を示唆していると考えられる。次章では、分子動力学計算を用いてダイナミクスも考慮することで、これらのミニヘアピン分子の安定性の違いを議論できるかどうかを検証する。

## 第 9 章

### ミニヘアピン分子の熱安定性の解析

前章のミニヘアピン分子の構造モデリング結果をもとに、安定なヘアピンを形成する dGC(GAA)GC と安定なヘアピンを形成しない dGC(GAG)GC について、分子動力学法 (molecular dynamics simulation, MD) を用いて多数の構造サンプリングを行い、ヘアピン形成の自由エネルギープロファイルを求めて、それらの熱安定性について考察する。本研究で用いた分子動力学法は、筆者と同じ研究グループの Tazaki らが開発した Poisson-Boltzmann 方程式により静電遮蔽効果を取り込むもので<sup>15,16)</sup>、溶媒を連続体近似することで効率のよいサンプリングを可能にしたものである。また複数の分子動力学計算から得られた多数のサンプリングデータをもとに自由エネルギープロファイルを求める方法として、WHAM (Weighted Histogram Analysis Method)<sup>93,94,24)</sup> を用いる。以下にそれぞれの方法の概要を説明する。

#### 9.1 Poisson-Boltzmann 分子動力学法

ミニヘアピン分子のヘアピン形成自由エネルギーを求めるためには、ヘアピン構造、あるいはヘアピンを形成していない状態での多数のコンフォメーションをサンプリングする必要がある。サンプリングを行う際には、シミュレートしている系が定常状態にあることが前提となる。しかし生体分子の周辺に存在する水やイオンなどの溶媒分子を陽 (explicit) に取り込んだ計算をする場合、多数の溶媒分子を含んだ系全体が平衡に達するまでには多大な計算時間が必要となる。

Poisson-Boltzmann 分子動力学法 (PBMD) は、生体分子周辺の水やイオンなどの溶媒分子を、Poisson-Boltzmann 方程式を用いて連続体近似し、それらからの寄与を静電遮蔽効果として計算に取り入れる方法である。この方法は、1つ1つの水分子やイオンを陽に取り扱わないため、系の収束が非常に速く、効率のよい構造サンプリングが可能となる。

サンプリングを行う空間を、溶媒領域、溶質領域およびその間の境界領域に分け、



それぞれの領域で誘電率  $\epsilon$  , Debye 遮蔽定数  $\kappa$  を設定する (Fig. 9.1) . 境界領域は通常 , 溶質分子の周辺 1.4 Å (水分子の半径) 程度に設定する .

PBMD における静電ポテンシャルと力の計算のアルゴリズムの概要を以下に示す . 詳細については , Appendix B を参照されたい .

1. シミュレーションを行う空間を格子に区切り , 有限差分 Poisson-Boltzmann 法 (Finite Difference Poisson-Boltzmann Method, FDPB) を用いて , 格子点上の電場を計算する .
2. 1. で求めた電場をできるだけ再現するように , 溶質分子の各原子内部に「仮想電荷」を割り当てる .
3. 静電ポテンシャルを原子ペアの 2 体間力で表すために , 各原子の電荷  $q_j$  と割り当てられた仮想電荷をもとに実効電荷  $q_j^e$  を計算する .
4. 各原子の電荷  $q_j$  と実効電荷  $q_j^e$  から , 式 (9.2) と式 (9.3) によって静電ポテンシャルおよび各原子にかかる力 (ポテンシャルの 1 次微分) の計算を行う . 分子内相互作用や van der Waals 相互作用などの他のポテンシャルについての計算は , 通常分子動力学計算と同様に行う .
5. 実効電荷の大きさは溶質分子のコンフォメーションに依存するので , 一定ステップ数 (通常は数千ステップ程度) ごとに , 手順 1~3 によって実効電荷を計算しなおす .

原子  $j$  によって作り出される静電ポテンシャルは次式で近似できる .

$$\phi_j(d_j) = \begin{cases} q_j^e \frac{1}{4\pi\epsilon_s} \frac{\exp(-\kappa_s d_j)}{d_j} & b_j < d_j \\ Q_j \frac{1}{4\pi\epsilon_s} \frac{1}{d_j} + \alpha_j & r_j < d_j \leq b_j \\ q_j \frac{1}{4\pi\epsilon_p} \frac{1}{d_j} + \beta_j & d_j \leq r_j \end{cases} \quad (9.1)$$

ただし ,  $r_j$  は原子  $j$  の原子半径 ,  $b_j$  は  $r_j + r_{probe}$  ,  $Q_j$  は原子  $j$  の内部と表面に存在する電荷の総量 ,  $q_j$  は原子  $j$  の電荷である . また  $\alpha_j$  と  $\beta_j$  は定数で ,  $\phi_j(d_j)$  が連続関数となるように定義される .

これを用いると , 系全体の静電ポテンシャル  $E^{el}$  は次式で与えられる .

$$E^{el} = \frac{1}{2} \sum_k q_k \sum_j \phi_j(r_{jk}) \quad (9.2)$$

ただし ,  $r_{jk}$  は原子  $j$  と原子  $k$  との間の距離である .

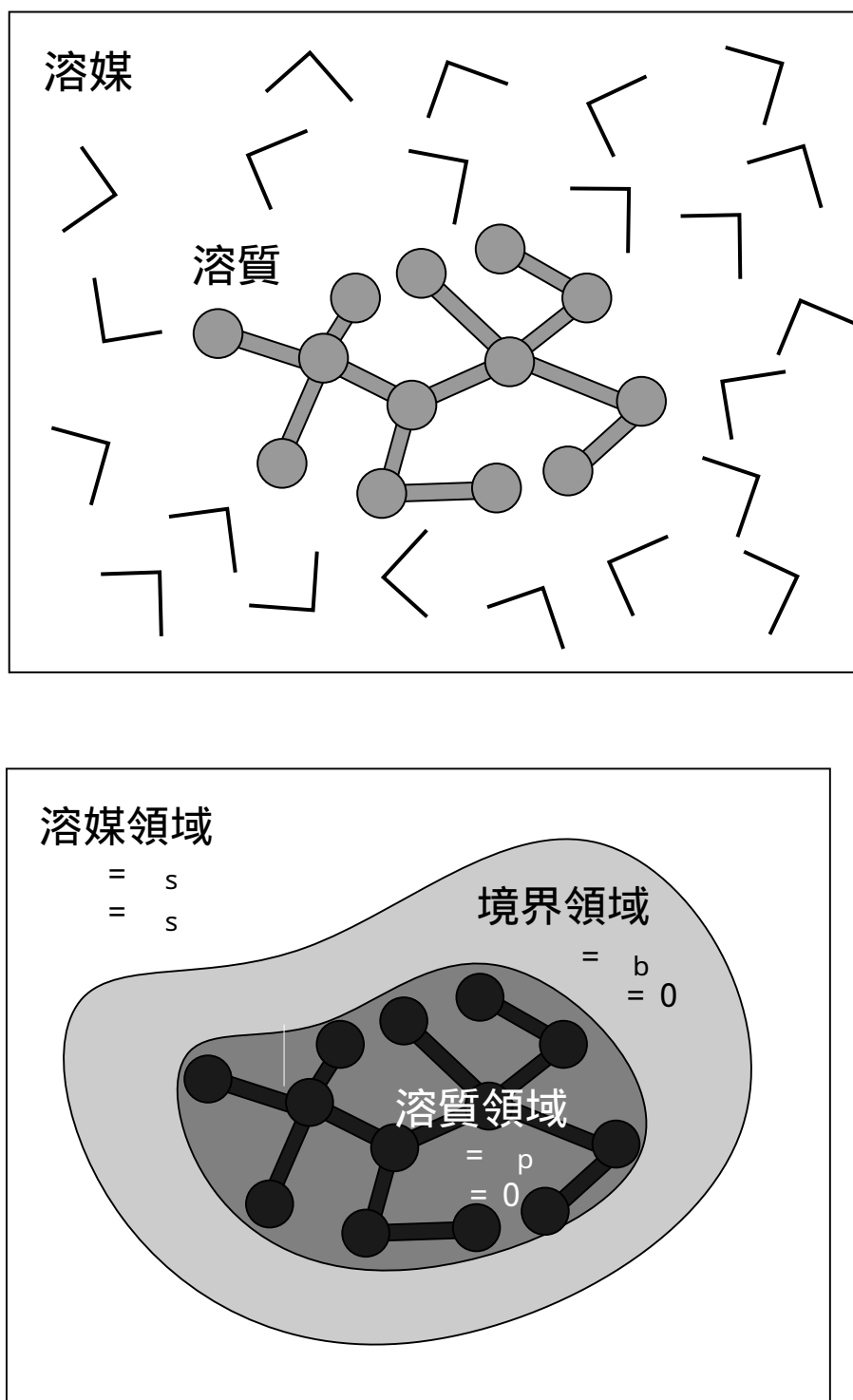


Figure 9.1 溶媒の取り扱いの概念図．(上) 溶媒分子を陽に取り込んだ場合．(下) 溶媒を連続体近似した場合．

各原子に加わる溶媒からの力と真空中での静電気力の和は，実効電荷  $q^e$  の構造依存性は小さいとして実効電荷の微分項をゼロとおくと，次式で表される．

$$\mathbf{F}^{el} = \frac{1}{2} \sum_{i \neq j} (q_j^e q_i + q_i^e q_j) \frac{1 + \kappa_s r_{ij}}{4\pi\epsilon_s} \frac{\exp(-\kappa_s r_{ij})}{r_{ij}^2} (\mathbf{r}_i - \mathbf{r}_j) \quad (9.3)$$

ただし  $f(r) = \exp(-\kappa_s r)/4\pi\epsilon_s r$  である．

## 9.2 反応座標の設定

本研究では，ヘアピン形成の自由エネルギープロファイルを求める際の反応座標 (reaction coordinate,  $\xi$ ) を，ヘアピン形成時に水素結合を形成しているステム端の塩基対 G1 の H1 と C7 の N3 の間の距離に設定する (Fig. 9.2)．この反応座標  $\xi$  に沿って自由エネルギーの値をプロットしたものを自由エネルギープロファイルとして求めることになる．

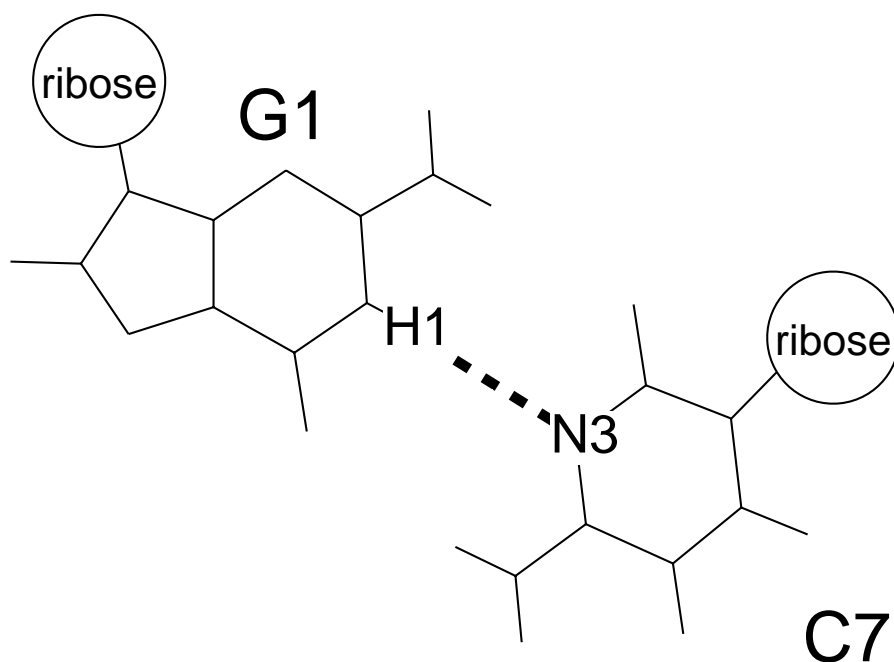


Figure 9.2 反応座標  $\xi$  は G1 の H1 と C7 の N3 の間の距離 (図中の破線) に設定する．ヘアピン構造形成時は，この 2 つの原子は水素結合している．

## 9.3 構造サンプリングの方法

構造サンプリングを行う際に，1 回の分子動力学計算では，配座空間のごく一部しか探索することができない．また通常の分子動力学計算では，エネルギーの山付

近に対応する構造がサンプリングされることは確率的に少ない．自由エネルギープロファイルを求める際には，できるだけ広い範囲の探索が必要である．このため，分子のコンフォメーションに拘束を加え，拘束のかけ方を変えた複数の分子動力学計算を行って，全体としてサンプリングされる空間が配座空間を広く覆うようにする (Fig. 9.3)．これにより，確率的にとりにくいコンフォメーションも強制的にサンプリングすることができる．また自由エネルギーは，それぞれの分子動力学計算から得られた構造群をまとめて計算する．これらにより，反応座標に沿ったサンプリングを十分に行うことができるようになる．

具体的には，式 (9.4) で表されるような，反応座標  $\xi$  に沿ったハーモニックな拘束ポテンシャル  $E_c$  を加える．この拘束ポテンシャルのサンプリングに及ぼす影響は，あとで述べる WHAM によって取り除くことができる．

$$E_c = c (\xi - \xi_0)^2 \quad (9.4)$$

ただし， $c$  は拘束の強さを表す定数， $\xi_0$  は拘束中心の反応座標の値である．

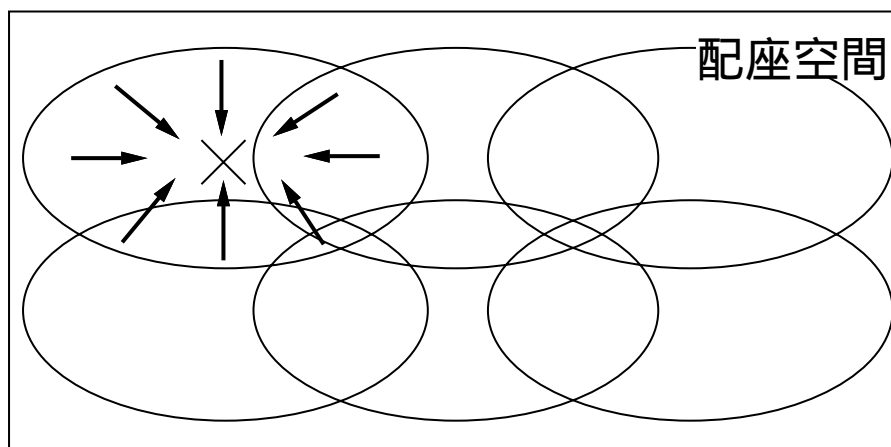


Figure 9.3 拘束をかけた構造サンプリングの概念図．1回のMDは，楕円で示した領域をサンプリングする．各MDでは，サンプリング中心 (図中の×印) からコンフォメーションが大きくはずれないような拘束がかけられる．拘束のかけ方を各MDによって変えることで，複数のMD全体として配座空間を広く覆うことができる．

計算に用いたパラメータは，以下の通りである．

- フォースフィールドは AMBER の parm91 を用いる．
- FDPB の格子サイズは  $0.7\sim 0.8 \text{ \AA}$  とする．

- イオン強度は dGC(GAA)GC の NMR による構造決定のときの条件と合わせて 0.24045 とする．誘電率  $\epsilon$  は，溶媒領域  $\epsilon_s$  が 78.5，溶質領域  $\epsilon_p$  が 2.0，境界領域  $\epsilon_b$  が 4.0 とし，境界領域の幅  $r_{probe}$  は 1.4 Å とする．
- 初期構造の反応座標  $\xi_0$  を中心として，拘束ポテンシャル  $E_c = c(\xi - \xi_0)^2$  を加える．拘束の強さ  $c$  は，1.0 ~ 12.0 に設定する．
- 初期構造に対して，最初に AMBER4.0 の最急降下法と共役勾配法によるエネルギー極小化を 1000 ステップ行い，原子同士の接触などを緩和する．その後分子動力学計算 (MD) を行う．
- MD のタイムステップは 1.0 fs とする．
- MD は 0 K から始めて，30000 ステップ目までで 300 K に系を heat up する．70000 ステップ目以降は温度制御を行わない free run とする．80000 ステップ目から 130000 ステップ目までの 50 ps の間に，50 ステップごとにサンプリングを行う (各 MD で 1000 個ずつサンプリングされることになる)．
- Free run の中で反応座標  $\xi$  が最大のものを次の MD の初期構造とする．
- 反応座標  $\xi$  が 6 Å を越えるまでの合計 12 本の MD の結果を統合して，WHAM により自由エネルギープロファイルを計算する．

#### 9.4 WHAM を用いた自由エネルギー計算の方法

Weighted Histogram Analysis Method (WHAM) は，複数の MC，MD シミュレーションから得られたサンプリングデータをまとめて自由エネルギープロファイルを求めるために開発された方法であり<sup>93,94,24)</sup>，これまでに terminally blocked tripeptide (反応座標は O1-H5 間距離)<sup>95)</sup>，three-helix bundle protein (反応座標は慣性半径)<sup>96)</sup>，アデノシン分子 (反応座標は二面角  $\gamma$  と  $\chi$ )<sup>25)</sup>，Protein G の segment B1 の熱安定性解析<sup>97)</sup>，疎水性相互作用の解析<sup>98)</sup> などに応用されている．またこの方法を  $\lambda$  ダイナミクス法<sup>99)</sup> と呼ばれる自由エネルギー解析手法と組み合わせた方法も開発されている<sup>100)</sup>．

WHAM のアルゴリズムの概要は，以下のとおりである．詳細については，Appendix C を参照されたい．

系の「状態」をエネルギー  $E$  と反応座標  $\xi$  の関数であると考える．また，エネルギーが  $E \sim E + \Delta E$  かつ反応座標が  $\xi \sim \xi + \Delta\xi$  の範囲に含まれる系の状態の数

を  $\Omega(E, \xi)$  個とする．Metropolis の Monte Carlo 法<sup>101)</sup> や分子動力学計算などのカノニカルアンサンブルを生成するシミュレーションを用いると，エネルギーが  $E \sim E + \Delta E$  かつ反応座標が  $\xi \sim \xi + \Delta\xi$  の状態が出現したステップ数  $N(E, \xi)$  を数えることで，この  $\Omega(E, \xi)$  を見積もることができる．

複数のシミュレーション結果から，真の状態密度  $\Omega(E, \xi)$  をできるだけ高い精度で推測するためには，各シミュレーションから得られた状態密度  $\Omega_j(E, \xi)$  に，それぞれの「信頼度」にしたがった重み  $w_j$  をつけて，

$$\Omega(E, \xi) = \sum_{j=1}^R w_j(E, \xi) \Omega_j(E, \xi) \quad (9.5)$$

$$\sum_{j=1}^R w_j(E, \xi) = 1 \quad (9.6)$$

とし，誤差を最小にするように  $w_j$  を決めればよい．ただし  $R$  はシミュレーションの本数である．このような  $w_j$  を求め，状態密度を求める式に代入すると，次式が得られる．

$$\Omega(E, \xi) = \frac{\sum_i \frac{N_i(E, \xi)}{g_i}}{\sum_m \frac{n_m}{g_m Z_m} \exp[-\beta_m(E + E_m)]} \quad (9.7)$$

$$Z_i = \sum_{E, \xi} \Omega(E, \xi) \exp[-\beta_i(E + E_i)] \quad (9.8)$$

ただし， $E_i$  は  $i$  番目のシミュレーションで系に加えた拘束エネルギーである．式 (9.7) と式 (9.8) を連立させて  $\Omega(E, \xi)$  を求めればよいが，これは非線形連立方程式なのでそのままでは解くことができない．このため，まず式 (9.7) で  $\Omega(E, \xi)$  を求め，それを式 (9.8) に代入して  $Z_i$  を求め，それを再び式 (9.7) に代入して  $\Omega(E, \xi)$  を求める，という反復計算を  $Z_i$  が収束するまで行うことによって解く． $\Omega(E, \xi)$  が求めれば，

$$P(E, \xi) = \frac{1}{Z} \Omega(E, \xi) \exp[-\beta E] \quad (9.9)$$

によって状態  $(E, \xi)$  が観測される確率が求まる．反応座標  $\xi$  についての自由エネルギープロファイル  $F$  は，

$$F(\xi) = -\log P(\xi) \quad (9.10)$$

$$P(\xi) = \int P(E, \xi) dE \quad (9.11)$$

によって求まる．

## 9.5 反応座標に拘束をかけない分子動力学計算の結果

第 8 章でモデリングした構造のうち，dGC(GAA)GC および dGC(GAG)GC に対して，300K において， $\xi$  に拘束をかけない PBMD によるサンプリングを行ったときの，各原子のゆらぎの大きさの比較をしたものを，Fig. 9.4 に示す．

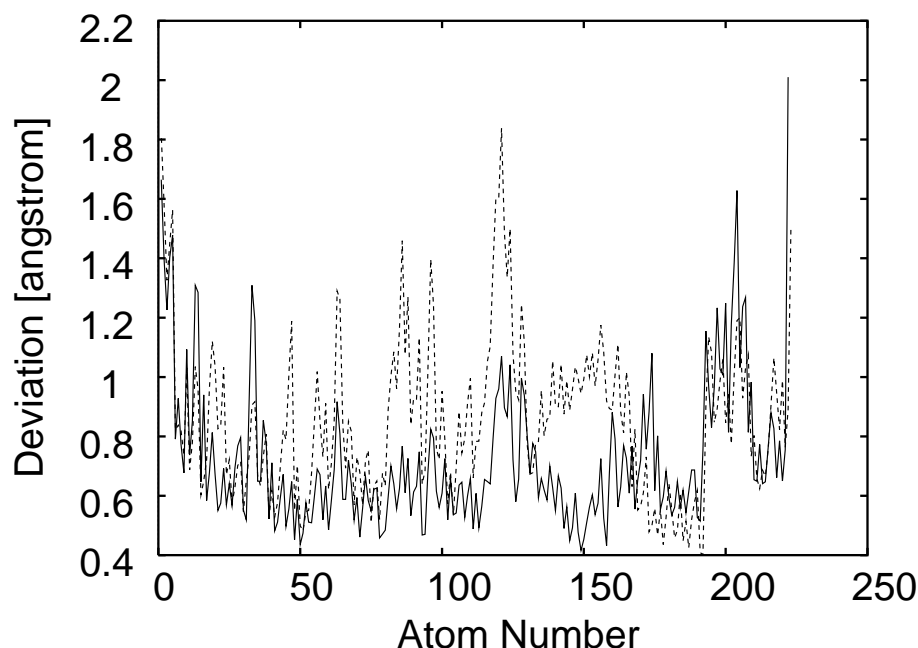


Figure 9.4 拘束をかけない分子動力学計算から求めた各原子の熱ゆらぎの大きさ．(破線) dGC(GAG)GC．(実線) dGC(GAA)GC．

第 8.5 節では，両者のエネルギー最小構造におけるループ塩基間のエネルギー値に大きな違いはなかったが，分子動力学計算では，dGC(GAG)GC は，dGC(GAA)GC に比べて，とくにループ部のゆらぎが大きくなっている．熱ゆらぎの大きさを構造に色づけして表示したものを，Fig. 9.5 に示す．

GAA はループ部よりもむしろステム端の G1，C7 のゆらぎが大きく，GAG では逆にループ部，とくに A4 と G5 の間の，糖 - リン酸骨格が折れ曲がっている部分でゆらぎが大きくなっている．

## 9.6 自由エネルギー解析の結果

dGC(GAA)GC および dGC(GAG)GC について，先に述べた PBMD と WHAM によってヘアピン形成の自由エネルギープロファイルを求めた．

反応座標  $\xi$  (G1 の H1 と C7 の N3 の間の距離) の 1.5 ~ 6 Å の範囲を，12 本の

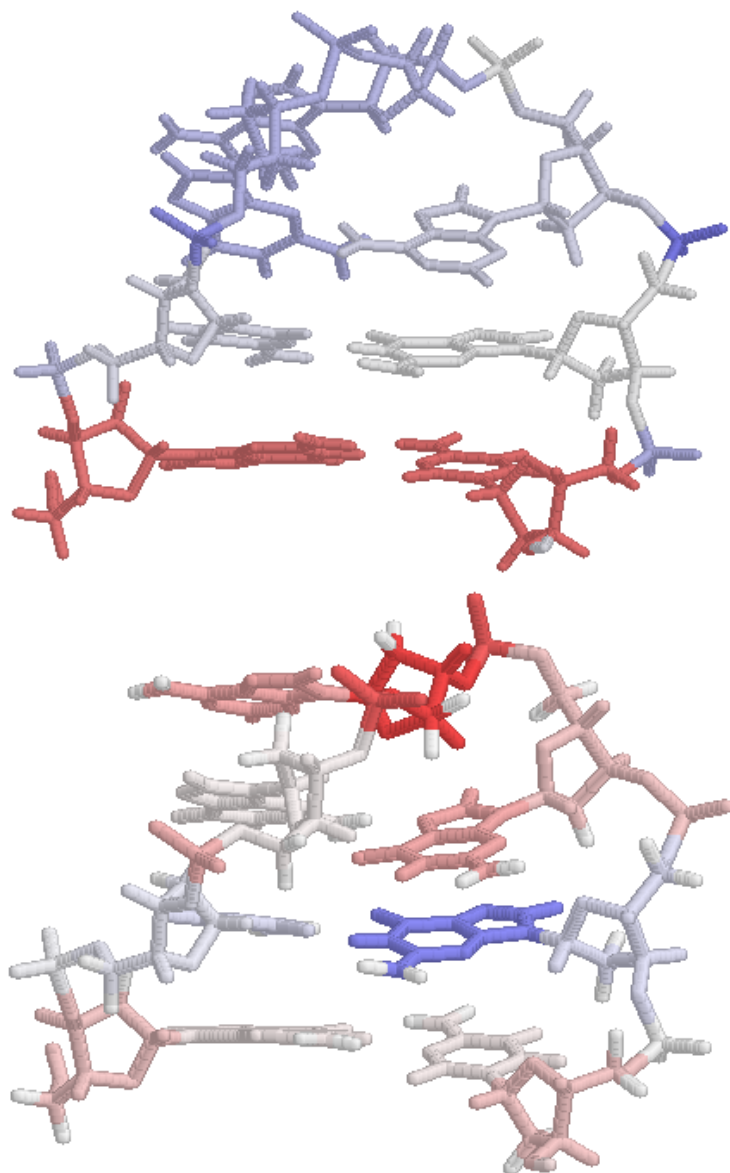


Figure 9.5 分子動力学計算から求めた各原子の熱ゆらぎの大きさを色で表したものの。青から赤になるにしたがって、ゆらぎが大きいことを示す。(上) dGC(GAA)GC。(下) dGC(GAG)GC。



PBMD で探索した . 各 PBMD の run での式 (9.4) における拘束パラメータは , Table 9.1 のようにした .

Table 9.1 PBMD の各 run での反応座標  $\xi$  に関する拘束 .

run 番号	1	2	3	4	5	6
$c$	1.5	1.5	6.0	12.0	12.0	12.0
$\xi_0$	1.959	2.427	2.427	2.648	2.927	3.206
run 番号	7	8	9	10	11	12
$c$	12.0	12.0	12.0	6.0	6.0	6.0
$\xi_0$	3.558	3.971	4.369	4.759	5.322	6.054

dGC(GAA)GC の各 run における  $\xi$  に対する頻度分布は , Fig. 9.6 のようになった .  $\xi$  が 1.5 ~ 6 Å の範囲を十分に覆い尽くすサンプリングができていることがわかる .

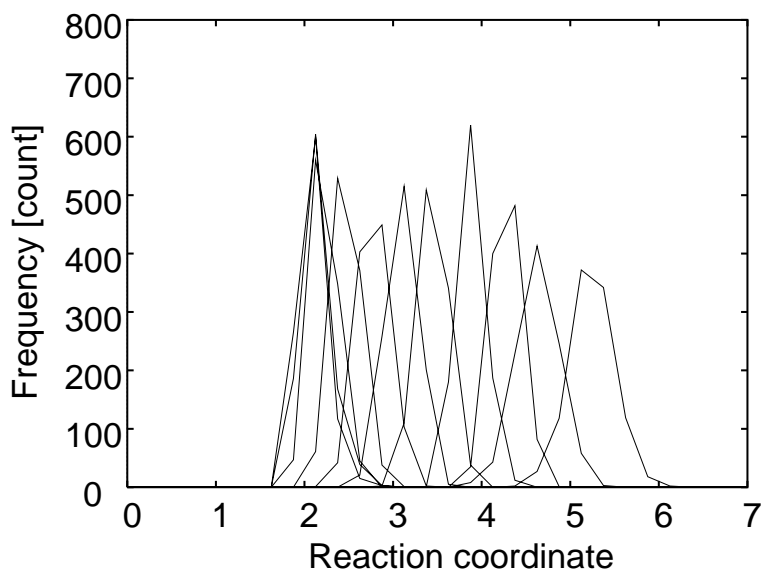


Figure 9.6 各 run に対する頻度分布のヒストグラム .

dGC(GAA)GC および dGC(GAG)GC のサンプリング構造群に対して WHAM を適用することで得られた自由エネルギープロファイルを , Fig. 9.7 に示す .

dGC(GAA)GC では , ヘアピン構造に対応する反応座標  $\xi = 2$  Å 付近に自由エネルギーの谷があり ,  $\xi$  が大きくなるにしたがって自由エネルギーは高くなった . 一方 , dGC(GAG)GC の自由エネルギープロファイルでは ,  $\xi = 2$  Å 付近のヘアピン構造

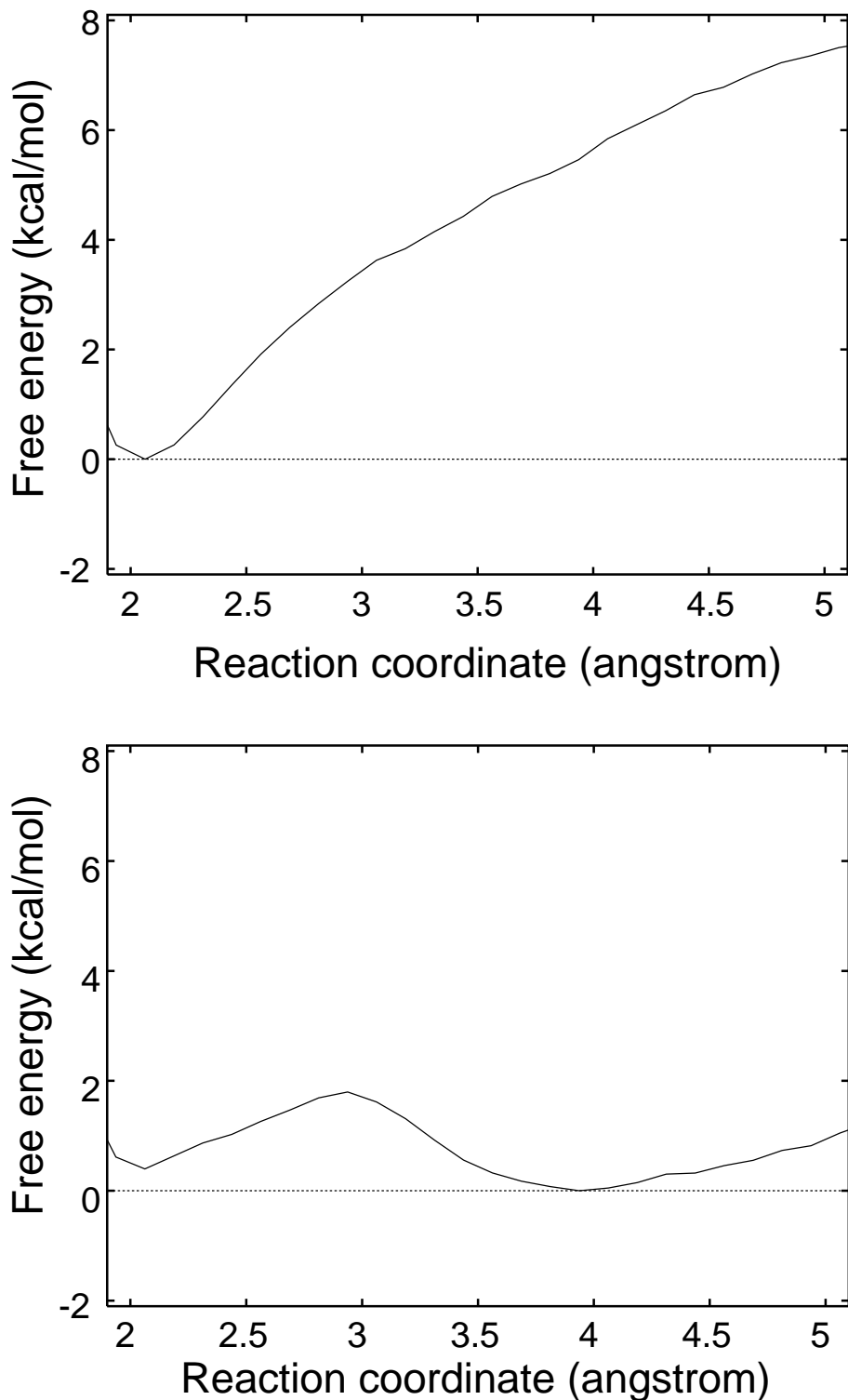


Figure 9.7 WHAMにより求められた自由エネルギープロファイル。(上)dGC(GAA)GC。(下)dGC(GAG)GC。

よりもさらに低い谷が  $\xi = 4 \text{ \AA}$  付近にあり，ヘアピン構造が最安定ではないことが明らかになった．

$\xi$  が  $1.5 \sim 6 \text{ \AA}$  の範囲において， $\xi$  に対する存在確率分布を自由エネルギープロファイルから求めたものを，Fig. 9.8 に示す．GAA の場合は， $\xi$  が  $1.7 \sim 3 \text{ \AA}$  にしか分布がみられないが，GAG の場合は， $1.7 \sim 3 \text{ \AA}$  と  $3 \sim 6 \text{ \AA}$  の2つのピークが見られた．GAG の  $3 \sim 6 \text{ \AA}$  のピークの高さは，もう一方のピークの2倍程度であった．

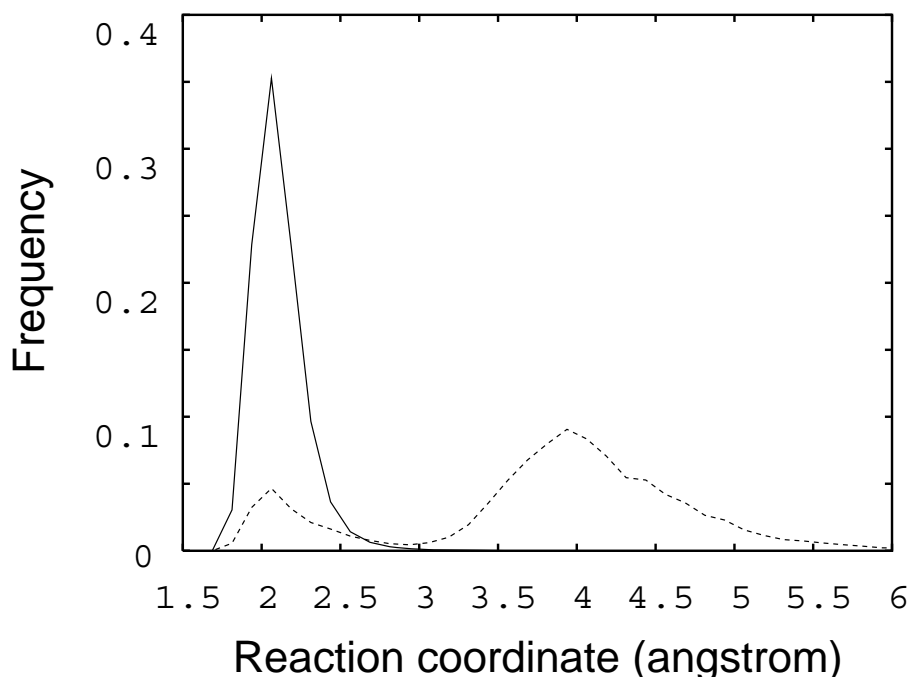


Figure 9.8 WHAMにより求められた確率分布．(実線)  $dGC(GAA)GC$ ．(破線)  $dGC(GAG)GA$ ．

また，自由エネルギープロファイルを，エンタルピー項とエントロピー項にわけて解析したものを Fig. 9.9 に示す．

エンタルピーとエントロピーについては，用いた構造サンプリング数がまだ十分でなく誤差が大きいと思われるので，定量的な議論は難しいと考えられるが，定性的には，以下のようなことが Fig. 9.9 から読み取れる．まず  $dGC(GAA)GC$  のヘアピン構造が安定なのは，主にエンタルピーの寄与による．また  $dGC(GAG)GC$  については， $\xi = 4 \text{ \AA}$  付近の構造は， $\xi = 2 \text{ \AA}$  付近のヘアピン構造に比べてエンタルピー的に安定になっていると考えられる．しかし  $\xi = 4 \text{ \AA}$  付近の構造の周辺をエネルギー極小化を用いて探索した結果，前章でモデリングしたヘアピン構造 ( $\xi = 2 \text{ \AA}$  付近に対応) よりも低いエネルギー値をもつ構造は得られなかった．このことから，

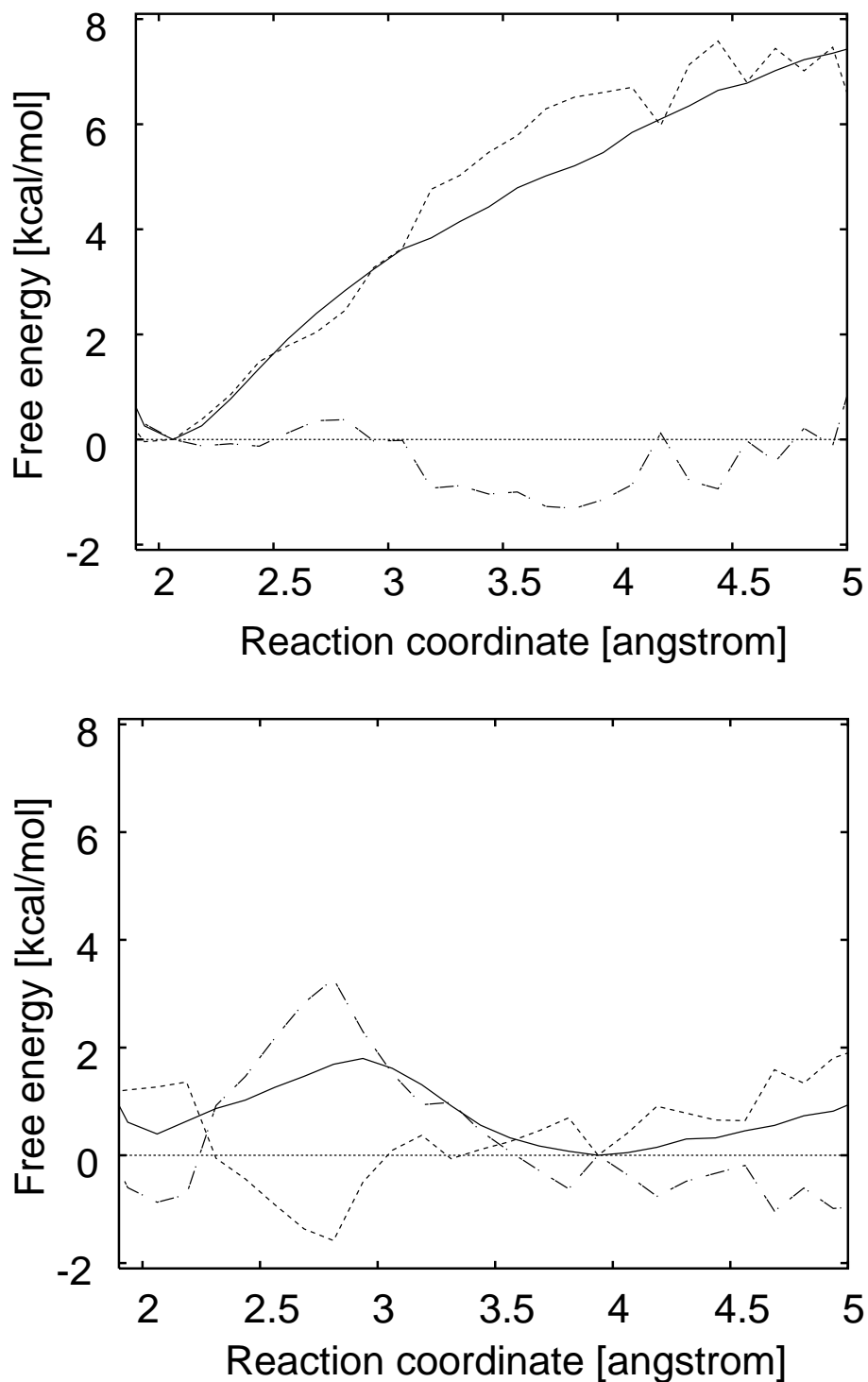


Figure 9.9 WHAM により求められた自由エネルギー (実線), エントロピー (1点鎖線), エンタルピー (破線). (上) dGC(GAA)GC. (下) dGC(GAG)GC.

モデリングで得られたエネルギー最小構造付近のエネルギー曲面の形状は狭い谷となっており，分子動力学計算によるサンプリングでは，このような狭い谷は探索されず， $\xi = 4 \text{ \AA}$  付近のヘアピン構造でない構造がエンタルピー的に安定になったと思われる．

GAA と GAG について，ヘアピン構造に対応する  $\xi = 2 \text{ \AA}$  付近，および  $\xi = 4 \text{ \AA}$  付近に対応する構造の例を，それぞれ Fig. 9.10 と Fig. 9.11 に示す．

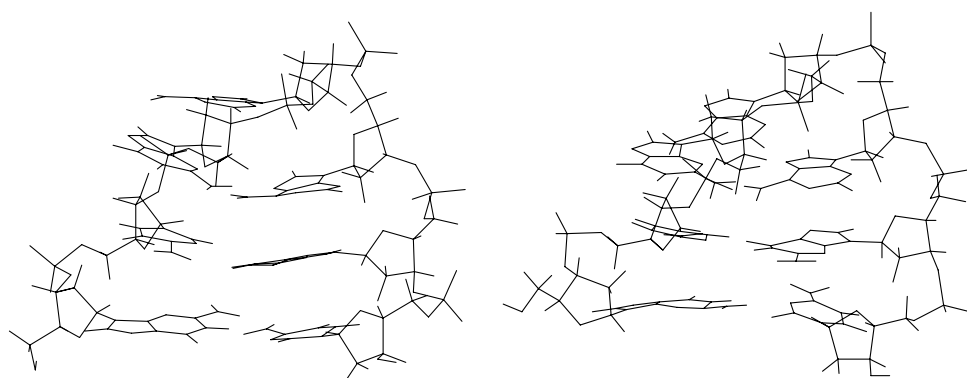


Figure 9.10 サンプリングされた dGC(GAA)GC の構造の例．(左)  $\xi = 2 \text{ \AA}$  付近．(右)  $\xi = 4 \text{ \AA}$  付近．

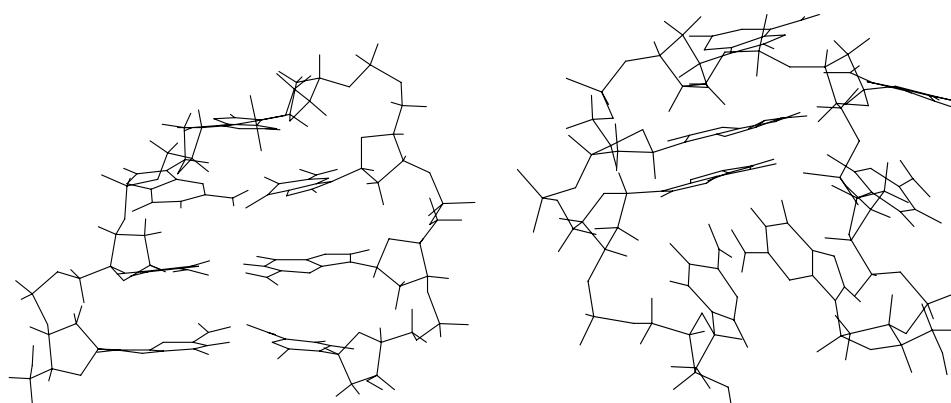


Figure 9.11 サンプリングされた dGC(GAG)GC の構造の例．(左)  $\xi = 2 \text{ \AA}$  付近．(右)  $\xi = 4 \text{ \AA}$  付近．

GAA については， $\xi = 2 \text{ \AA}$ ， $\xi = 4 \text{ \AA}$  のいずれもヘアピン構造であった．GAG については， $\xi = 2 \text{ \AA}$  はヘアピン構造であるが， $\xi = 4 \text{ \AA}$  では，ヘアピン構造から base pairing やスタッキング様式が崩れたコンパクトな構造をとっていた．

### 9.7 熱安定性解析結果のまとめ

前章で、Two-Stage 法による構造モデリングを行った dGC(GAA)GC および dGC(GAG)GC について、分子動力学計算によって多数のコンフォメーションをサンプリングして自由エネルギープロファイルを求めた。前章のエネルギー解析では、ループ塩基間の相互作用の大きさは GAA, GAG で大きな違いはなかったが、自由エネルギープロファイルからは、dGC(GAA)GC はヘアピン構造 ( $\xi = 2 \text{ \AA}$  付近) が安定であること、dGC(GAG)GC は、ヘアピン構造が崩れた構造 ( $\xi = 4 \text{ \AA}$  付近) が安定であることが明らかになった。このことは、ヘアピン構造の熱安定性を議論する際にはエネルギー最小構造のエネルギー値の解析だけでは不足であること、分子動力学計算によって得られた構造群を WHAM でまとめて解析することで、安定なヘアピン構造をとる分子とそうでない分子との熱安定性の違いを、明らかにすることができることを示している。

サンプリングされた構造の数がまだ十分ではないために、ヘアピン形成自由エネルギーのエンタルピー項・エントロピー項についての定量的な考察はできないが、今後さらに多数の構造サンプリングを行うことで、定量的な解析を行うことができるようになると考えられる。

## 第 10 章

### 第 II 部のまとめ

第 II 部では、DNA ミニヘアピン分子の熱安定性の配列依存性について、Two-Stage 法を用いた構造モデリングと、PBMD および WHAM を用いたヘアピン形成自由エネルギー計算による解析を行った。構造モデリングは、安定なヘアピン構造を NMR 解析結果と矛盾なく生成することができた。またエネルギー最小構造では、安定なヘアピンを形成する分子と形成しない分子とのエネルギー値に大きな違いは見られなかった。一方、分子動力学計算で多数の構造をサンプリングし、エントロピーの効果も考慮することで、両者の安定性を区別できることが明らかになった。

本研究による手法の改善点については、まず構造モデリングについては、本研究では探索により求めたエネルギー最小構造をモデリング構造として用いたが、天然構造がエネルギー最小ではない場合を考慮することが考えられる。このためには、モデリングの過程でエネルギーが低い構造を数多くサンプリングし、もっともらしい構造についてのエネルギー値以外の知見に基づいてモデリング構造を生成することが必要である。

タンパク質の立体構造予測においては、多数の候補構造群から天然構造をピックアップするための方法として、類似構造同士をクラスタリングし、最大クラスタの中心 (平均) 構造を予測構造とする手法が広まりつつある。このような考え方を、本手法に応用することも可能であると考えられる。

また二面角系では、溶媒効果を陽に取り込んだ計算が難しいが、二面角系においても RISM (Reference Interaction Site Model) 理論を用いて溶媒効果を見積もる方法<sup>102)</sup> などが開発されており、本研究でもこれらの応用が考えられる。

構造サンプリングについては、人工的な非カノニカルアンサンブルを生成することでエネルギー曲面の山や谷を越えた広範な探索を行うことができる手法が、マルチカノニカル法、レプリカ交換法などをはじめとしていくつか提案されている<sup>103-107)</sup>。これらの方法を用いてさらに十分な探索を行うことで、エンタルピー・エントロピー

の定量的な解析まで行えることが期待される．これらの手法を採り入れて，dGC(NNN)GC のループ部のすべての塩基配列に対して，システマティックな構造モデリングと自由エネルギー解析を行うことが，今後の課題である．その結果は，DNA ミニヘアピン分子のみならず，RNA ミニヘアピン分子やその他の核酸分子の安定性とその塩基配列依存性などの研究に大きな示唆を与えると期待される．



## 第 III 部

# tRNA・アミノアシル tRNA 合成酵素の ダイナミクス解析



## 第 11 章

### 第 III 部の目的

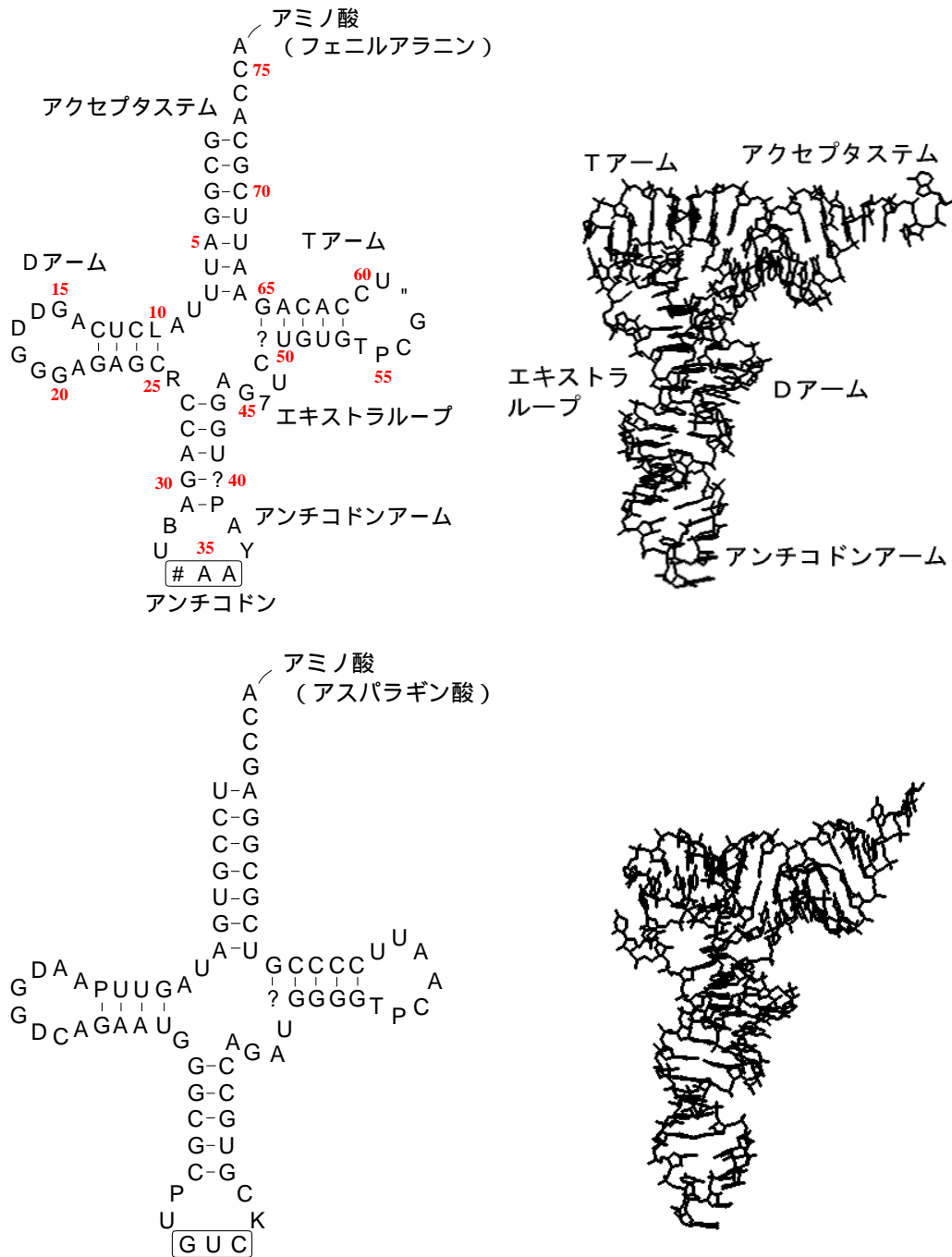
第 III 部では、トランスファー RNA (tRNA) およびアミノアシル tRNA 合成酵素 (ARS) に二面角系基準振動モード解析法を適用して、tRNA および ARS のフリー状態と複合体形成状態でのダイナミクスを解析した結果について述べる。

DNA 上にコードされた遺伝子がタンパク質として発現する過程には、さまざまな分子が関わっているが、トランスファー RNA (tRNA) は、その中でも中心的な役割を果たす。tRNA は、リボソームなどの働きを借りながら、メッセンジャー RNA (mRNA) に転写された遺伝情報 (コドン) に対応するアミノ酸を運んで次々に並べていく。ヌクレオチドとアミノ酸の橋渡しをしている、遺伝情報の翻訳の根本を担う分子である<sup>108)</sup>。

アミノアシル tRNA 合成酵素 (ARS) は、天然に存在する 20 種類のアミノ酸の 1 つ 1 つに対応して存在する。ARS は、自分が担当するアミノ酸とそれに対応する 1 組の tRNA を認識して複合体を形成し、tRNA の 3' 末端に対応するアミノ酸を付加する機能をもつ。ARS が担当する tRNA を、その ARS の「同族 (cognate) tRNA」と呼ぶ。tRNA に対応しないアミノ酸が付加されると、遺伝情報が誤って翻訳されることになるので、ARS の同族 tRNA の認識は非常に高精度である。翻訳の誤り確率は 1 つのアミノ酸あたり 1/10000 以下であるといわれている。

ARS は生物種・アミノ酸系によって非常に多様な構造であるのに対して、tRNA は、ミトコンドリア由来の一部のものを除くといずれも共通のクローバー型の 2 次構造をしており、立体構造は特徴的な L 字型構造である<sup>109)</sup>。このため ARS は同族 tRNA を、よく似た他の tRNA の中からわずかな違いを認識して区別していると考えられる。

tRNA の立体構造は、1973 年に Kim らによって初めて明らかにされた<sup>110)</sup>。例として、*S. cerevisiae* 由来の tRNA<sup>Phe</sup>、tRNA<sup>Asp</sup> の 2 次構造および立体構造を Fig. 11.1 に示す。tRNA のクローバー型 2 次構造は、アクセプタステム、D アーム、アンチ



**Figure 11.1** tRNA の 2 次構造 (左) および立体構造 (右) . 上に *S. cerevisiae* tRNA<sup>Phe</sup> (PDB エントリ名: 4TRA) , 下に tRNA<sup>Asp</sup> (PDB エントリ名: 2TRA) の例を示した . 4TRA の 2 次構造の中の赤数字はヌクレオチド番号を示す . また 2 次構造の中の A, G, U, C 以外の記号は修飾塩基を示す . L: N2-methylguanosine (m2G), D: dihydrouridine (D), R: N2,N2-dimethylguanosine (m22G), B: 2'-O-methylcytidine (Cm), #: 2'-O-methylguanosine (Gm), Y: wybutosine (yW), P: pseudouridine (psi), ?: 5-methylcytidine (m5C), 7: 7-methylguanosine (m7G), ”: 1-methyladenosine (m1A), K: 1-methylguanosine (m1G) .

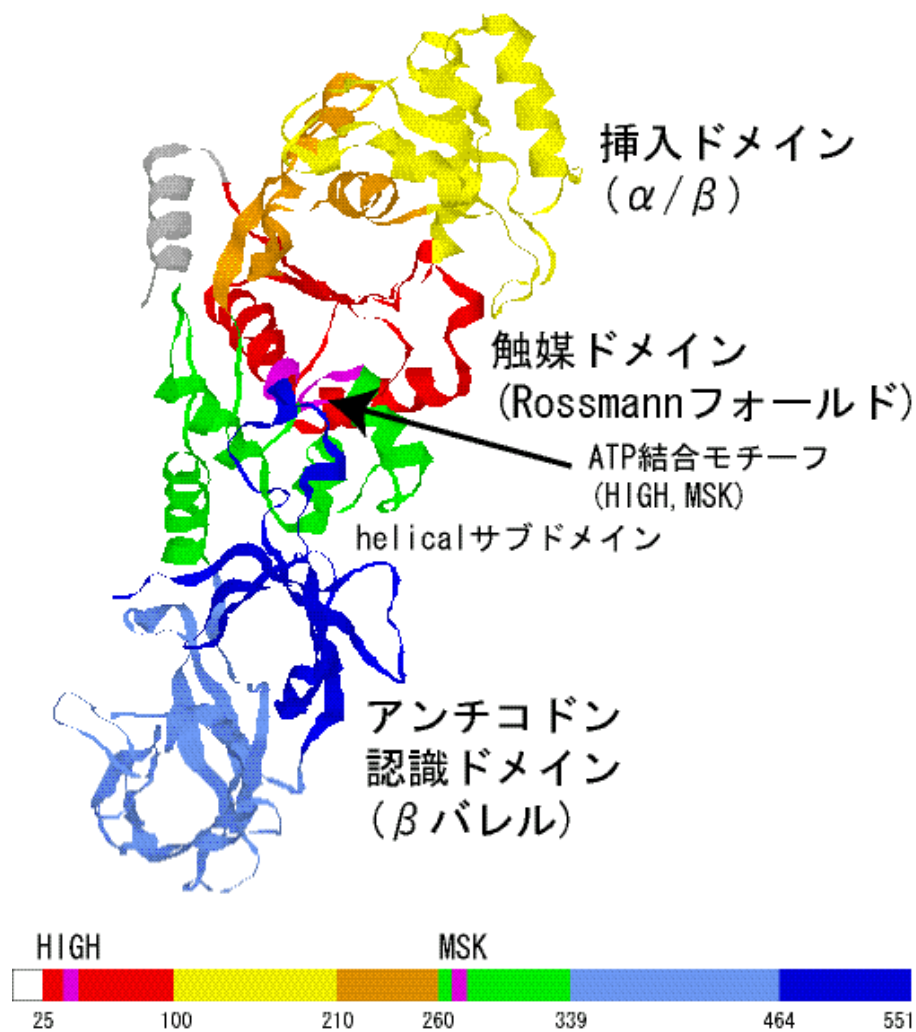
コドンアーム，T アームの 4 本の主要な二重らせん構造が形作っており，アンチコドンアームと T アームの間にエクストラループがある．エクストラループの長さは tRNA の種類によって異なるが，主要なアームの長さはほとんど同じである．tRNA の 3' 末端（アクセプタステムの先端）に，ARS によって対応するアミノ酸が付加され（アミノアシル化），アミノアシル tRNA となる．また 34, 35, 36 番目の 3 塩基がアンチコドンであり，リボソームの中でメッセンジャー RNA (mRNA) 上のコドンと対合する．なお，アンチコドンアーム，D アーム，T アームはステム-ループから成るヘアピン構造をとっており，たとえばアンチコドンアームのステム部は「アンチコドンステム」，ループ部は「アンチコドンループ」などと呼ぶ．

一方 ARS は，分子量やフォールドがさまざまにアミノ酸配列の相同性も互いに低い，1990 年に ATP 結合モチーフの違いによって Class I, Class II の 10 種類ずつの 2 つのクラスに分けられることが報告された<sup>111)</sup>．ARS のクラス分けを Table 11.1 に示す<sup>112-114)</sup>．ARS は ATP 結合モチーフをもつ触媒ドメイン，アンチコドンを認識するアンチコドン認識ドメインなど，いくつかのドメインに分けることができる．Class I では，ARS の触媒ドメインは平行  $\beta$  シートを基本とする Rossmann フォールドであり，His-Ile-Gly-His (HIGH) と Lys-Met-Ser-Lys (KMSK) という共通の ATP 結合モチーフをもつ．これに対して Class II の ARS の触媒ドメインは，7 本の逆平行  $\beta$  シートで構成され，モチーフ 1, 2, 3 という共通の ATP 結合モチーフをもつ．機能的には Class I の ARS は対応する tRNA の 3' 末端（アデニン）のリボース環 2'-OH にアミノ酸を付加するが，Class II はリボース環の 3'-OH にアミノ酸を付加する，という違いがある．Class I は側鎖が大きく疎水性の高いアミノ酸に対応し，Class II は側鎖が小さく親水性の高いアミノ酸に対応する．それぞれのクラスは，対応するアミノ酸の性質に基づいて，さらにそれぞれ 3 つのサブクラスに分けられる．同じサブクラスの ARS 間では，アミノ酸配列の相同性が高くなる．

ARS の構造の例として，*E. coli* 由来の GlnRS (Class I) の立体構造<sup>115)</sup> を Fig. 11.2 に示す．それぞれのドメインの位置は，対応した色でアミノ酸配列の中に示してある．また，HIGH および MSK (*E. coli* GlnRS では KMSK モチーフの最初の Lys は Val になっている) の ATP 結合モチーフの位置をマゼンタで表示している．

GluRS とのアミノ酸配列・構造を比較した研究<sup>116,113,114)</sup> により，GlnRS と GluRS は，触媒ドメインとそこに挿入された  $\alpha/\beta$  構造のドメインの構造が非常によく似ているが，C 末端側のアンチコドン認識ドメインは大きく構造が異なっていることが明らかになった．この C 末端のドメインは，進化の過程であとから付加されたものと考えられている．

tRNA-ARS 複合体は結晶化が困難であり，その立体構造は長い間ベールに包まれ



**Figure 11.2** *E. coli* GlnRS の立体構造 (PDB エントリ名: 1GTR) . アクセプタステムを認識し, HIGH/MSK の ATP 結合モチーフをもつ触媒ドメイン (赤・橙), all  $\beta$  のバレル構造をもつアンチコドン認識ドメイン (淡青・青), 進化の過程で挿入されたと考えられる  $\alpha/\beta$  構造の挿入ドメイン (黄) などいくつかのドメインに分けられる. 下に各ドメインのアミノ酸配列上の位置を示す. また HIGH/MSK のモチーフをマゼンタで示した.

**Table 11.1** ARS のクラス分け . ATP 結合モチーフによって Class I, II に分かれ , さらにアミノ酸の性質に基づいてサブクラスに分けられる .  $\alpha$  などはサブユニット構成を示す . PheRS は Class II であるがアミノアシル化部位は A76 の 2'-OH である . また LysRS については , 真核生物とほとんどの真正細菌については Class II であるが , 真正細菌と古細菌の中には Class I に属するものもある . AlaRS と GlyRS は , 生物種によって異なる 2 つの構造をとる . 同じサブクラスの ARS 間では , アミノ酸配列の相同性が高くなる .

	Class I		Class II	
モチーフ	HIGH, KMSK		モチーフ 1 (P), 2 (FRXE), 3 (GXGXGXER)	
触媒ドメイン	Rossmann フォールド		逆平行 $\beta$ シート	
アミノアシル化部位	2'-OH		3'-OH	
サブクラス	<b>Class Ia</b>	<b>Class Ib</b>	<b>Class IIa</b>	<b>Class IIb</b>
	ArgRS $\alpha$	GlnRS $\alpha$	GlyRS $\alpha_2$	AsnRS $\alpha_2$
	CysRS $\alpha$	GluRS $\alpha$	HisRS $\alpha_2$	AspRS $\alpha_2$
	IleRS $\alpha$	<b>Class Ic</b>	ProRS $\alpha_2$	LysRS $\alpha_2$
	LeuRS $\alpha$	TrpRS $\alpha_2$	SerRS $\alpha_2$	<b>Class IIc</b>
	MetRS $\alpha_2$	TyrRS $\alpha_2$	ThrRS $\alpha_2$	AlaRS $\alpha$ or $\alpha_4$
	ValRS $\alpha$			GlyRS $\alpha_2\beta_2$
				PheRS $\alpha_2\beta_2$

ていた．このため，tRNA-ARS の精緻な分子認識メカニズムの解明のために，tRNA の塩基配列レベルでの多数の研究が 1970 年代から 1990 年代初め頃にかけて行われた．その結果，tRNA 中の 1 つまたは少数の特定塩基を置換すると，tRNA のアミノアシル化効率が大きく減少することが明らかになり，これらの塩基はその tRNA を特徴づける「アイデンティティー」と呼ばれている<sup>117)</sup>．これまでさまざまな生物種・アミノ酸系で数多くのアイデンティティーが報告されているが，その大部分はアクセプタステム，アンチコドンアームに集中している<sup>118)</sup>．

1989 年に Rould らによって，*E. coli* tRNA<sup>Gln</sup>-GlnRS 複合体の立体構造が，X 線結晶解析によって初めて明らかにされた<sup>115)</sup>．さらにその後の X 線結晶解析技術の進歩により，多くの生物種・アミノ酸系において，フリー状態の ARS あるいは tRNA-ARS 複合体の構造が原子レベルで決定されるようになり<sup>119,112)</sup>，報告されていたアイデンティティーと ARS とのインタラクションに関する議論が活発となった．

tRNA，ARS およびそれらの複合体の立体構造が明らかにされたことで，これらの分子のコンピュータを用いた解析も行われるようになった．

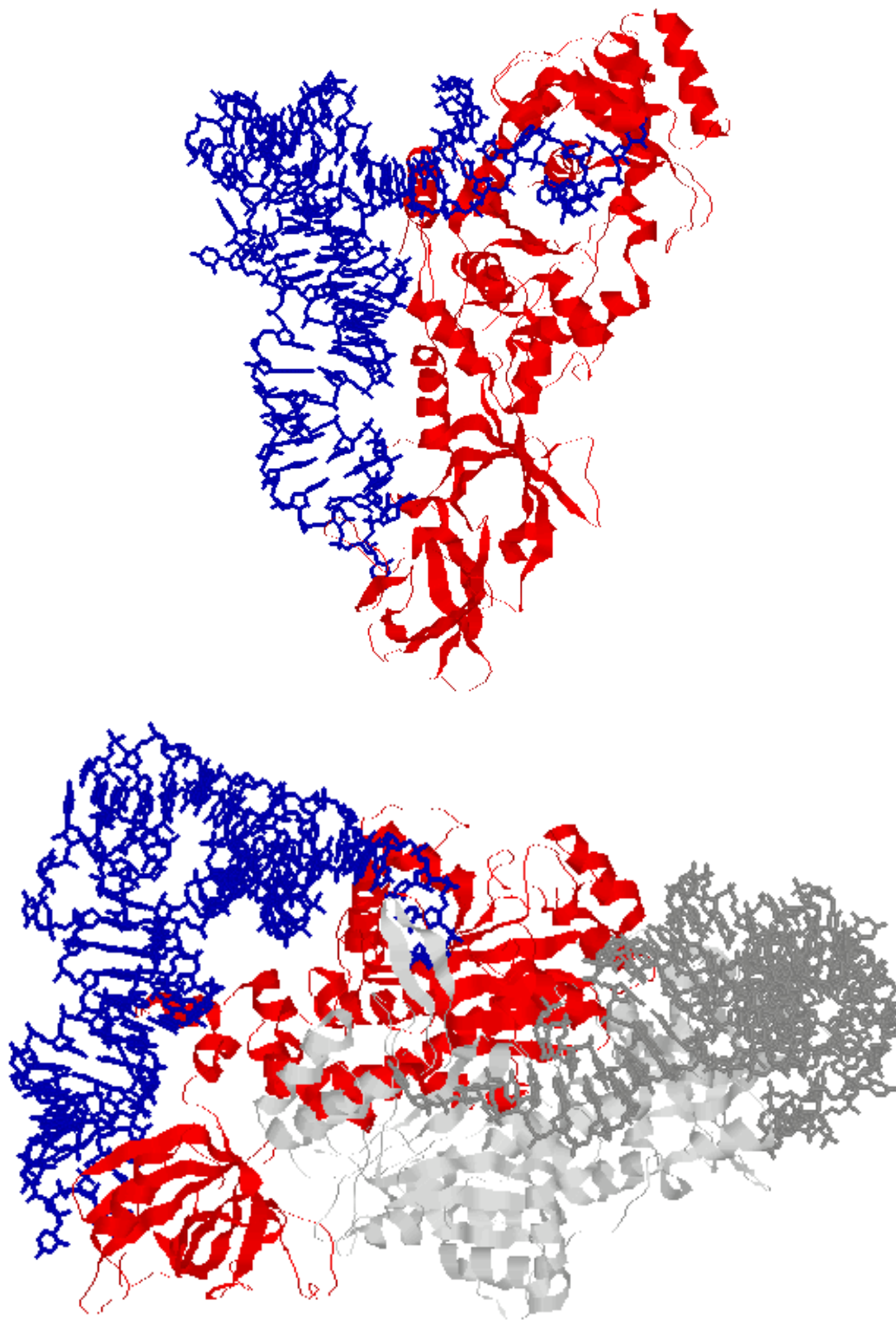
Harvey らは，24 ピコ秒の分子動力学計算を *S. cerevisiae* tRNA<sup>Phe</sup> に適用し，その熱ゆらぎの大きさが球状タンパク質よりも大きいことを見出した<sup>120,121)</sup>．筆者らは，二面角系基準振動モード解析法を *S. cerevisiae* tRNA<sup>Phe</sup> および tRNA<sup>Asp</sup> に適用して，両分子のダイナミクスを 0.2 ピコ秒周期の高周波数領域から 50 数ピコ秒周期の低周波数領域まで周波数領域にわけて解析し，tRNA の低周波数領域での hinge bending motion の様子などを明らかにした<sup>27,30)</sup>．その後 Matsumoto らは，筆者らが固定して扱ったヌクレオチドのリボース環のパッカリングの自由度を取り込んだ基準振動モード解析法を *S. cerevisiae* tRNA<sup>Phe</sup> に適用し，計算されたダイナミクスへのパッカリングの影響について考察している<sup>122)</sup>．Auffinger らは，分子動力学法を *S. cerevisiae* tRNA<sup>Asp</sup> のアンチコドンヘアピン部分や全体構造に適用し，tRNA を分子動力学法を用いてシミュレートするときの計算方法やパラメータ設定などについて考察している<sup>123-125)</sup>．Simonson らは，AspRS の活性部位でのアミノ酸結合自由エネルギーを，溶媒の連続体近似 (Poisson-Boltzmann 法) と分子動力学法を用いて計算している<sup>126-128)</sup>．

また tRNA-ARS 複合体の解析例には以下のようなものがある．Tateno らは，分子動力学法とコンピュータモデリングを用いて，tRNA<sup>Glu</sup> と GluRS の複合体モデルを構築した<sup>129,130)</sup>．Bahar らは，*E. coli* tRNA<sup>Gln</sup>-GlnRS 複合体のダイナミクスを，簡略化したポテンシャルを用いた Gaussian Network Model (GNM) という基準振動モード解析法に類似した方法で解析し，GlnRS の  $\beta$  パレルの熱ゆらぎが大きく，逆に Rossmann フォールド周辺の熱ゆらぎが小さいことなどを見出している<sup>131)</sup>．



比較的小規模な *E. coli* tRNA<sup>Gln</sup>-GlnRS 複合体でも総原子数が 1 万を超える巨大分子であり、他のアミノ酸系では複合体中に tRNA あるいは ARS が複数分子含まれる場合も多い (Fig. 11.3)。このため tRNA-ARS 複合体へのコンピュータの適用例は、非常に限られたものになっている。

さて、tRNA-ARS の分子認識メカニズムの研究は、これまで両分子の接触部位 (インタフェース) でのアミノ酸とヌクレオチドの相互作用の解析が中心であり、複合体の X 線結晶解析構造をベースにした「静的な」観点からのものであった。しかし、複合体状態とフリー状態での tRNA および ARS 分子の構造を比較することで、tRNA-ARS 複合体の形成には両分子の構造変化を伴うことが従来から明らかにされていた。このことから、ARS が正確に tRNA をアミノアシル化する過程において、分子の構造変化・ダイナミクスが深く関わっていることが示唆される。最近になって、このことが「tRNA コア」に着目した研究によって、さらに裏付けられることになった。tRNA コアとは、tRNA の特徴的な L 字型立体構造の「角 (かど)」にあたる部分のことであり、D アーム、T アーム、エキストラループが会合し、塩基や糖-リン酸骨格の間で高次水素結合が形成されている。*E. coli* の Cys,<sup>132,133</sup> Ile,<sup>134</sup> Phe,<sup>135,136</sup> Ser,<sup>137</sup> Arg<sup>137</sup> および *S. cerevisiae* の Phe<sup>138</sup> などの系で、tRNA コア部位の塩基を置換した場合に tRNA のアミノアシル化効率が大きく低下することが知られていた。これに加えて、最近のコア内の高次水素結合やスタッキング様式に関する詳細な解析から、これらの塩基が tRNA のアミノアシル化に重要な役割を果たしている可能性が指摘されつつある。Hou らは、*E. coli* tRNA<sup>Cys</sup> のもつ G15:G48 および A13:A22 という特徴的なコア内の塩基対が *H. influenzae* tRNA<sup>Cys</sup> でも保存していることをゲノム配列解析から見出し、さらにそれらの塩基を置換したときの  $k_{cat}/K_m$  の低下の様子などから、これら 2 つの生物種の tRNA<sup>Cys</sup> のコア内塩基対が機能的にも同等の特徴をもつことを明らかにした<sup>139</sup>。また、筆者と同じ研究グループの Sagara らは、さまざまな生物種やアミノ酸系の tRNA の塩基配列を相互に比較し、tRNA コア付近にその tRNA を特徴づける塩基が多く現われることが明らかにしている<sup>140</sup>。Christian らは、*E. coli* tRNA<sup>Cys</sup> の tRNA コア部位の塩基を置換して  $k_{cat}$  と  $K_m$  を調べることで、高次水素結合だけでなくコア部位の塩基のスタッキングがアミノアシル化に重要であることを示した<sup>141</sup>。Bullock らは、*in vitro* で wild type よりも強く *E. coli* GlnRS と結合する *E. coli* tRNA<sup>Gln</sup> の置換体をスクリーニングして、tRNA<sup>Gln</sup> のエキストラループの CAUUC が AGGU に置換された配列 (aptamer) が、wild type の約 30 倍の結合能をもつことを明らかにし、さらにこの aptamer-GlnRS 複合体の X 線結晶解析を行った<sup>142</sup>。その結果、aptamer と GlnRS の結合様式や GlnRS の構造に wild type との大きな違いはなく、(1) 複合体形成に必要な構造変化を ap-



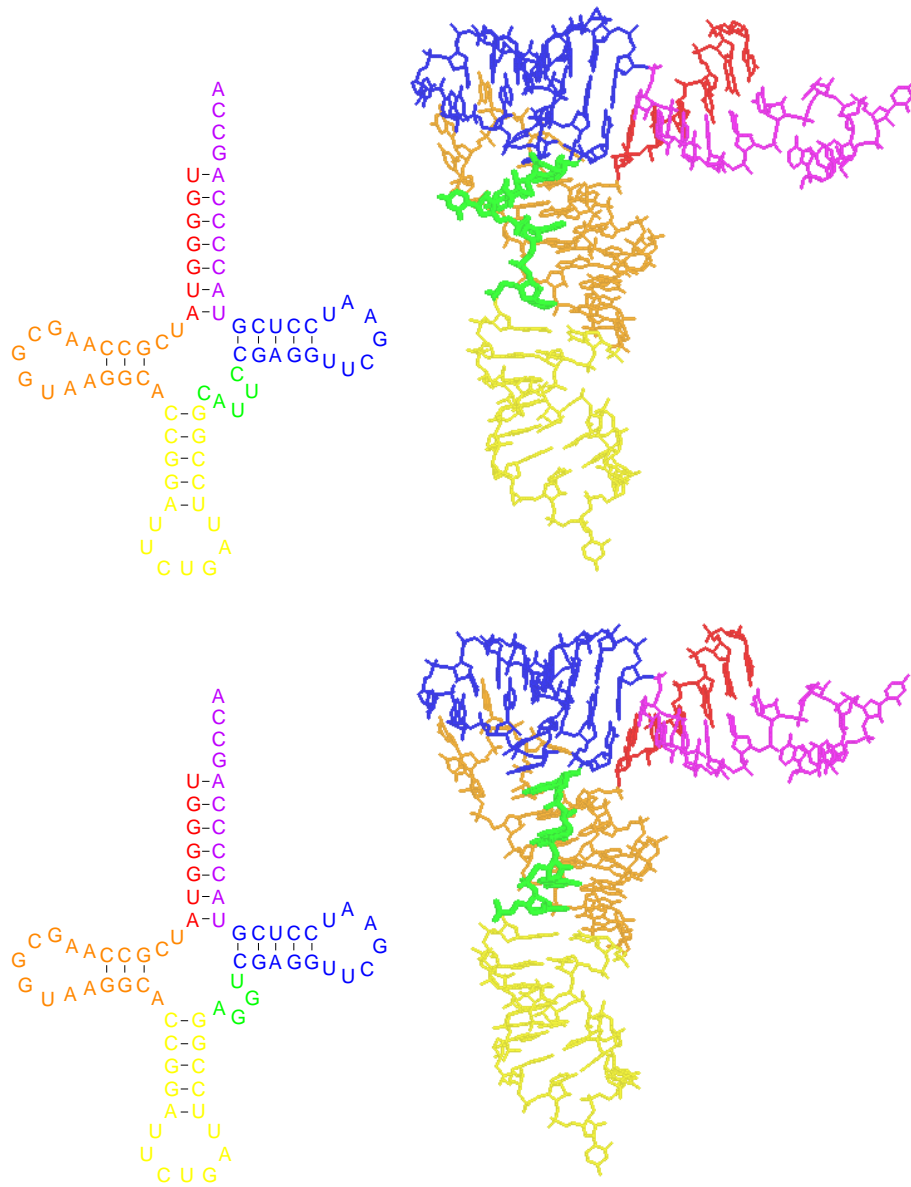
**Figure 11.3** Gln系およびAsp系のtRNA-ARS複合体の結晶構造．青のワイヤーフレーム表示がtRNA，赤のリボン表示がARS．(上) *E. coli* のGln系 (PDB エントリ名: 1GTR)．(下) *S. cerevisiae*のAsp系 (PDB エントリ名: 1ASY)．tRNA<sup>Asp</sup> 1分子+AspRS 1分子が，さらに2量体を形成している (もう一方のtRNA<sup>Asp</sup>-AspRSペアをグレーで表示)．

tamer の方がより低い自由エネルギーで起こせる，(2) GlnRS と結合したあとのダイナミクスに差があり，aptamer の方がエントロピー的に有利になっている，などの可能性が指摘されている．

非常に興味深いのは，これらの tRNA コア領域の塩基は，ARS の接触部位とは離れた位置にあることである．Bullock らの見出した aptamer は wild type の tRNA<sup>Gln</sup> と数塩基しか異なっていない (Fig. 11.4)．両分子の GlnRS との解離定数の違いがダイナミクスの違いによるとすれば，これは数塩基の違いからもたらされていることになる．このことを考えると，アミノ酸系による tRNA の特徴として，「静的な」塩基配列や構造だけでなく，似たような L 字型立体構造をもつ tRNA のダイナミクスが相互に異なり，ARS がこの違いを認識している可能性が十分に考えられる．

コンピュータ・シミュレーションを用いることで，フリー状態あるいは複合体を形成したときの分子のもつドメイン間運動や柔らかさなど，分子全体にわたる「動的な」特徴を直接的に解析することができる．しかし，先に述べたように，tRNA-ARS 複合体のシミュレーションには膨大な計算コストが必要となる．たとえば先に述べた Bahar らによる GNM を用いた解析では，計算量を減少させるために，(1) アミノ酸・ヌクレオチドの間の相互作用は，1 つの作用点 (アミノ酸の場合は C<sup>α</sup> 原子，ヌクレオチドの場合はリン原子) の間でのみ働くとする，(2) アミノ酸・ヌクレオチドの種類によらず，相互作用を表すパラメータは共通であるとする，(3) アミノ酸・ヌクレオチドのゆらぎは空間的に等方的であるとする，などの大幅なモデルの簡略化が行われている．Bahar らは，GNM を用いて X 線結晶解析から得られた tRNA<sup>Gln</sup>-GlnRS の温度因子を比較的よい精度で再現しており<sup>131)</sup>，また同じグループの Keskin らは，Rossmann フォールドをもつタンパク質同士でも，構造のバリエーションによってグループ分けをすると，それぞれのグループ内で固有の似たダイナミクスをもつ，という興味深い結果を導いている<sup>143)</sup>．さらに GNM を拡張した ANM (Anisotropic Network Model)<sup>144)</sup> で上に述べた (3) の欠点を解決しようとする試みもある．しかし依然として詳細なダイナミクスの解析を行うには精度が不十分である．とくに tRNA-ARS の分子認識では，先に述べたように少数の塩基の置換によってアミノアシル化効率が大きく変化するために，そのダイナミクスの解析には AMBER や CHARMM など，原子レベルでの精密なポテンシャルを用いることが必要であると思われる．このためには膨大な計算コストが必要であり，大規模な分子に対応するために並列計算技術を用いることが必須となる．

以上をふまえて，第 III 部では，tRNA-ARS の分子認識へのダイナミクスの関与を検討する第一歩として，フリー状態の tRNA<sup>Phe</sup> と tRNA<sup>Asp</sup>，および tRNA<sup>Gln</sup>-GlnRS 複合体に対して二面角系基準振動モード解析法を適用し，そのダイナミクス



**Figure 11.4** *E. coli* tRNA<sup>Gln</sup>(上) と aptamer(下) の 2 次構造および立体構造．左に 2 次構造の模式図，右に立体構造を示す．立体構造の図には，2 次構造の図と対応して，塩基配列に沿った色づけがしてある．tRNA<sup>Gln</sup> と aptamer は，エキストラループ(緑)部分の塩基のみが異なっている．

を解析することを目的とする．基準振動モード解析法は，分子の熱ゆらぎを各周波数成分に分け，分子がどの周波数でどのような運動（モード運動）をしているかを解析する方法である．分子動力学法に比べて少ない計算量で，分子全体にわたる熱ゆらぎの振動運動を解析することができる．さらに第 I 部で構築した並列計算システムを用いることで，tRNA-ARS 複合体分子のような巨大分子に対して，AMBER などの高い精度のポテンシャルを用いた解析が適用可能となった．これらを利用して，アミノ酸系によるモード運動の相違，複合体形成によるモード運動の保存・変化を解析し，複合体の形成が tRNA および ARS 分子のダイナミクスに与える影響，フリー状態と複合体状態で共通する運動成分を明らかにする．

## 第 12 章

### 基準振動モード解析法

#### 12.1 基準振動モード解析法の概要

この節では、基準振動モード解析法の概略について、同じく生体分子のダイナミクスの代表的な解析手法である分子動力学法と対比しつつ述べる。基準振動モード解析法のアルゴリズムについては、Appendix Dにその詳細を述べる。

基準振動モード解析法は、生体分子のポテンシャルエネルギーが、注目する自由度 (パラメータ) に関する多次元放物面 (2次形式) で表せるという仮定のもとに、その分子のダイナミクスに関する特性を計算するものである。ポテンシャルエネルギーの極小点近傍での小さな分子内振動、すなわち分子全体の並進や回転の運動を除いた、小さな立体構造変化のダイナミクスを対象としており、タンパク質や核酸のダイナミクス解析などに広く使われている。

系に与える自由度 (今の場合は二面角系なので、自由度は1つ1つの二面角となる) を  $\mathbf{q} = \{q_j\}$  とし、極小点近傍でポテンシャルエネルギー関数  $E(\mathbf{q})$  が多次元放物面で近似できると考えると、 $E(\mathbf{q})$  は極小点での2次微分行列 (Hessian)  $F_{ij}$  と極小点での値  $\mathbf{q}^0 = \{q_j^0\}$  を用いて、

$$E(\mathbf{q}) = \frac{1}{2} \sum_{i,j} F_{ij} (q_i - q_i^0)(q_j - q_j^0) \quad (12.1)$$

$$F_{ij} = \frac{\partial^2 E}{\partial q_i \partial q_j} \quad (12.2)$$

と表すことができる。ただし  $i, j$  は自由度の番号である。また運動エネルギー  $K$  は、

$$K = \frac{1}{2} \sum_{i,j} H_{ij} \dot{q}_i \dot{q}_j \quad (12.3)$$

$$H_{ij} = \sum_l m_l \frac{\partial \mathbf{r}_l}{\partial q_i} \cdot \frac{\partial \mathbf{r}_l}{\partial q_j} \quad (12.4)$$

と書ける．ただし  $m_l$  は各原子の質量， $r_l$  は各原子の座標を表す位置ベクトルである．これらを用いると，Lagrange の運動方程式は，

$$H A \Lambda = F A \quad (12.5)$$

$$A_{ij} = \omega_i^2 \delta_{ij} \quad (12.6)$$

という一般化固有値問題になり，これを規格化条件

$$A^T H A = \delta_{ij} \quad (12.7)$$

のもとで解く．その解  $q_j$  は，重ね合わせの係数  $A_{j\lambda}$  と固有振動数  $\omega_\lambda$  を用いて，

$$q_j = q_j^0 + \sum_{\lambda} A_{j\lambda} Q_{\lambda} \quad (12.8)$$

$$Q_{\lambda} = \alpha_{\lambda} \cos(\omega_{\lambda} t + \beta_{\lambda}) \quad (12.9)$$

と表すことができる． $Q_{\lambda}$  は基準振動モードである．各基準振動モードの振幅  $\alpha_{\lambda}$  は，熱平衡状態ではエネルギーが各モードに等分配されることから，以下のように求められる．

$$\alpha_{\lambda} = \sqrt{\frac{2kT}{\omega_{\lambda}^2}} \quad (12.10)$$

ただし  $T$  は系の絶対温度， $k$  はボルツマン定数である．なお熱ゆらぎの大きさは，第 III 部のすべての計算において系が  $T = 300$  K であるときに励起される値を示してある．

式 (12.8) は，自由度  $q_j$  の極小点近傍での振動が，自由度数個の基準振動モード  $Q_{\lambda}$  の線形和で表せることを示している．重ね合わせ係数 ( $A_{j\lambda}$ ) と固有振動数 ( $\omega_{\lambda}$ ) は，Hessian を含む一般化固有値問題である式 (12.5) を解くことで得られる．また， $A_{j\lambda}$  と  $Q_{\lambda}$  が求まれば，二面角や原子座標の熱ゆらぎなど，種々の物理量の時間平均が計算できる．

基準振動モード解析法の特徴を，分子動力学法と対比すると以下ようになる．

分子動力学法：

- Newton の運動方程式を，小さい時間刻み幅で次々と数値積分することで，分子の形の刻一刻のスナップショットを順次求めていく．
- 非平衡過程や大きなコンフォメーション変化も解析可能である．

- 長周期の振動モードを得るための計算コストが大きくなる．
- XYZ 系で計算されることが多い．

基準振動モード解析法：

- エネルギー曲面が多次元放物面であると仮定して，Lagrange の運動方程式に対応する一般化固有値問題を解くことで，解析的に振動モードを求める．
- 1つのエネルギー極小点付近の小さな振動運動のみ解析可能である．
- 長周期の振動モードを簡単に求められる．
- 二面角系で計算されることが多い．これによって，幅広い周波数範囲における原子の熱運動の様子を分子全体にわたって解析できる．

分子動力学法は，実時間での分子の運動をそのままシミュレートするので，たとえば 50 ピコ秒周期の振動を観測するためには，少なくともその数倍のシミュレーション時間が必要となる．これに対して基準振動モード解析法は，1点での計算で済むことから，分子全体にわたる長周期の運動を少ない計算時間で求めることができる．両者の概念図を Fig. 12.1 に示す．

生体分子では，少数の低周波数領域の基準振動モードが全体の運動におよぼす寄与が非常に大きいことが知られている<sup>27,30,145-149</sup>．このため少ないモード数の和で分子全体にわたるダイナミクスの特徴を表すことができ，複合体状態とフリー状態でのダイナミクスの比較も容易になる．

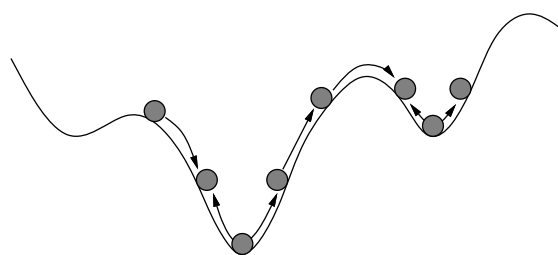
本研究では，環を多く含む核酸分子でとくに自由度を減らせること，大きな系でも長周期の振動モードを少ない計算量で求められることなどの利点を考慮して，ダイナミクス解析に二面角系基準振動モード解析法を用いる．複数分子を二面角系で扱う方法は，Higo らによるアルゴリズム<sup>150,151</sup>を用いている．基準振動モード解析法は大きな構造変化を扱うことができないが，タンパク質分子同士のドッキングの際の構造変化が，少数の基準振動モード運動の和で記述できたという報告もあり<sup>152</sup>，基準振動モードで tRNA，ARS および複合体のダイナミクスの特徴をとらえることは有効であると考えられる．

## 12.2 シミュレーションの手順

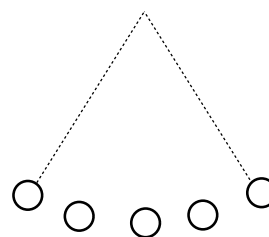
一般化固有値問題の式 (12.5) を解くにあたっては，Hessian ( $F$ ) が正定値 (すべての固有値が正) でなければならない．これは基準振動モード解析の「解析を行う点で



### 分子動力学法

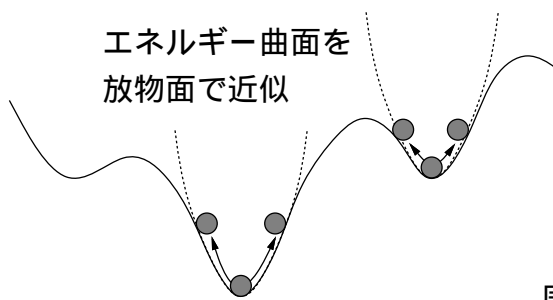


時間軸に沿った分子の形の  
スナップショットを順次求める



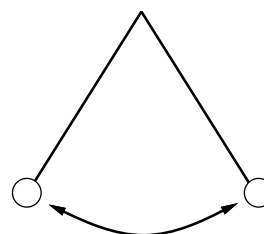
スナップショットを  
解析することで  
ダイナミクスを記述

### 基準振動モード解析法



エネルギー曲面を  
放物面で近似

極小点の 2 次微分から  
振動の周期・振幅を求める



周期・振幅から解析的に  
ダイナミクスを記述

Figure 12.1 分子動力学法と基準振動モード解析法 の概念図 .

は、エネルギー曲面はすべての自由度について下に凸の放物面となっている」という必要条件を満たすためである。固有値が負のモードが存在するということは、エネルギー曲面はその固有値に対応する固有ベクトル方向に「上に凸」になっており、現在いる「頂点」から固有ベクトル方向にたどればエネルギーがさらに下がるということの意味する。このため、本研究では、最急降下法 (Steepest Descent Method) とニュートン・ラフソン法 (Newton-Raphson Method) の 2 種類のエネルギー極小化のアルゴリズムを組み合わせて使い、すべての固有値が正となるような十分に低いエネルギー値をもつ極小点を探索する。

本研究での基準振動モード解析計算の流れを以下に示す。

1. PDB (Protein Data Bank) の座標データを初期構造とする。
2. X 線結晶解析データの場合は水素原子に関する座標情報がないので、AMBER4.0 を用いて生成する。
3. 水素原子のみについて、AMBER4.0 を用いて、共役勾配法 (Conjugate Gradient Method) によるエネルギー極小化を最大 5000 ステップ行う (1 ステップあたりのエネルギー変化が  $1.0 \times 10^{-4}$  kcal/mol 以下になった時点で終了する)。
4. 全原子について、AMBER4.0 を用いて、共役勾配法によるエネルギー極小化を最大 5000 ステップ行う (1 ステップあたりのエネルギー変化が  $1.0 \times 10^{-4}$  kcal/mol 以下になった時点で終了する)。
5. 二面角系において、最急降下法とニュートン・ラフソン法を組み合わせ、Hessian の固有値がすべて正となり、かつ 1 ステップあたりのエネルギー変化が  $1.0 \times 10^{-5}$  kcal/mol 以下となるまでエネルギー極小化を行う。
6. 得られたエネルギー極小構造に対して、基準振動モード解析計算を実行する。

結晶構造と基準振動モード解析に用いたエネルギー極小構造との r.m.s.d. (tRNA はリン原子, ARS は C $\alpha$  原子について計算) は、tRNA<sup>Phe</sup> が 2.6 Å, tRNA<sup>Asp</sup> が 2.4 Å, tRNA<sup>Gln</sup>-GlnRS 複合体が 1.1 Å である。

フォースフィールドには、AMBER の parm91<sup>42,43</sup>) を用いた。また AMBER によるエネルギー極小化では誘電率を  $\epsilon = 4r$  ( $r$  は原子間距離) とし、二面角系の計算では誘電率を式 (3.8) の  $\epsilon =$  シグモイダル関数 ( $D = 79.0, s = 0.3$ ) を用いて水和効果を近似する。

二面角系でリボース環のパッカリングを扱う方法は、Tomimoto らによって開発されており<sup>153,154)</sup>、本研究のシステムでも導入されているが、現在では複合体計算の際の計算コストの問題から、tRNA のリボース環のパッカリングの自由度を固定して扱っている。Harvey らによる tRNA<sup>Phe</sup> の分子動力学計算では、24 ピコ秒のシミュレーション中にパッカリングが変化したヌクレオチドは A46 と A76 の 2 つだけであったことが報告されている<sup>155)</sup>。また、Matsumoto らによってフリー状態の tRNA についてこの自由度を考慮したときの効果が検証されているが<sup>122)</sup>、パッカリングの固定は分子全体の基準振動モード結果に大きな影響を与えるものではない。これらのことから、パッカリングを固定して扱うことは妥当であると考えられるが、エネルギー極小化・基準振動モード計算のさらなる高速化を図ることで、パッカリングの自由度も計算に含めることは今後の課題である。

なお計算には、東京大学情報基盤センターの日立 SR2201 を用いた。計算時間は、Gln 系複合体の場合 (64 CPU を使用) で、エネルギー極小化に約 144 時間、基準振動モード解析計算に約 1.3 時間、tRNA<sup>Gln</sup> の場合 (16 CPU を使用) で、エネルギー極小化に約 38 分、基準振動モード解析計算に約 1 分であった。

## 第 13 章

### フリー状態の tRNA のダイナミクス解析

この章では、フリー状態の *S. cerevisiae* 由来の 2 つの tRNA 分子、tRNA<sup>Phe</sup> (PDB エントリ名: 4TRA) および tRNA<sup>Asp</sup> (PDB エントリ名: 2TRA) の基準振動モード解析結果について述べる。

#### 13.1 モード周波数の解析結果

基準振動モード解析では、分子の運動を、設定した自由度数のモード運動の和で表す。Fig. 13.1 は、そのモード周波数の分布を tRNA<sup>Phe</sup>、tRNA<sup>Asp</sup> について示したものである。周波数の単位には、その周波数に対応する光の波数を用いている。波数  $k$  ( $\text{cm}^{-1}$ ) と周波数  $\nu$  (Hz) の対応は、次式で表される。

$$k = \frac{\nu}{c} \quad (13.1)$$

ただし  $c$  は光の速度 ( $2.9979 \times 10^{10} \text{cm/sec}$ ) である。

これらの周波数分布の最低値・最高値・平均値および  $20 \text{cm}^{-1}$  以下のモード数の全体に対する割合を示したものが Table 13.1 である。

Table 13.1 フリー状態の tRNA<sup>Phe</sup> および tRNA<sup>Asp</sup> のモード周波数分布。単位は  $\text{cm}^{-1}$ 。

分子	最低	最高	平均	$20 \text{cm}^{-1}$ 以下のモード数 (%)
tRNA <sup>Phe</sup>	0.79	197.4	67.3	15.0
tRNA <sup>Asp</sup>	0.60	199.7	70.6	15.9

両分子のモード周波数分布に大きな違いはなかった。最低周波数は、球状タンパク質で報告されている値<sup>156,157)</sup> の  $1/4 \sim 1/10$  程度となっている。すなわち、tRNA は

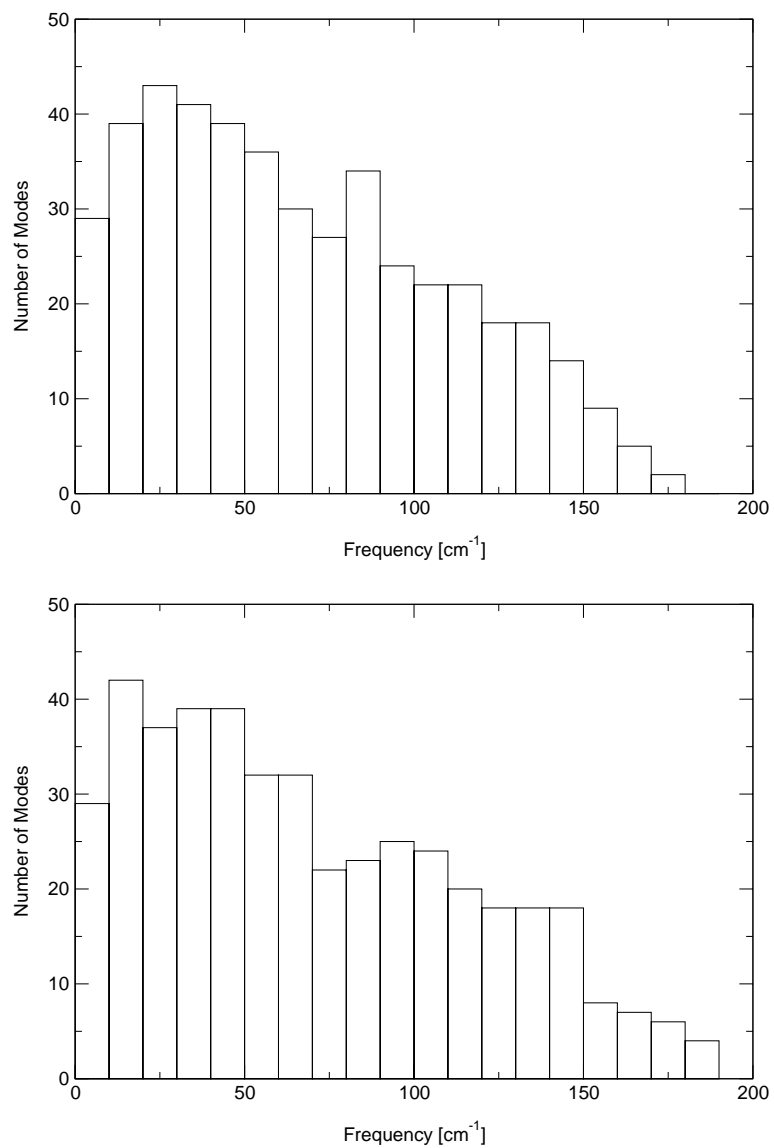


Figure 13.1 フリー状態の tRNA 分子のモード周波数分布 . (上) tRNA<sup>Phe</sup> . (下) tRNA<sup>Asp</sup> .

球状タンパク質に比べて、4 倍から 10 倍程度遅い振動モードをもっている。これは tRNA 分子の特異的な L 字型立体構造に起因するものと思われる。

### 13.2 各原子の熱ゆらぎの大きさ

フリー状態の tRNA<sup>Phe</sup> と tRNA<sup>Asp</sup> における、基準振動モード解析から得られた熱ゆらぎの大きさと X 線結晶解析の温度因子データから得られた熱ゆらぎの大きさの比較を Fig. 13.2 に示す。値はヌクレオチドごとに平均したものである。また、温度因子 ( $B$ ) と熱ゆらぎ ( $d$ ) の間の関係は、

$$B = 8\pi^2 d^2 / 3 \quad (13.2)$$

で表される。

基準振動モード解析法で求められた熱ゆらぎの大きさは、tRNA<sup>Phe</sup>、tRNA<sup>Asp</sup> いずれの場合も、L 字型立体構造の両端にあたるアクセプタステムおよびアンチコドンループ付近で大きく、2~4 Å 程度であった。

tRNA<sup>Phe</sup> および tRNA<sup>Asp</sup> において、温度因子と基準振動モード解析結果は相関係数がそれぞれ 0.60、0.51 と比較的良好に一致した。これは本研究のシミュレーションの妥当性を示している。温度因子と基準振動モード解析結果の違いは、主にアクセプタステム、アンチコドンループなどの L 字型立体構造の先端部分で、基準振動モード解析の方がゆらぎが大きくなっていることに起因する。この原因として、結晶構造中では、分子が密に並んで結晶格子を形成しているため、隣の分子の影響で先端部分の動きが抑えられるという要素があると考えられる。Auffinger らは、同じ tRNA<sup>Asp</sup> 分子の 500 ピコ秒間の分子動力学計算結果から同様の各原子のゆらぎの大きさを計算しているが<sup>125)</sup>、本研究の基準振動モード解析による結果と同様、L 字型構造の先端では温度因子を大きく上回るゆらぎを検出している。

また 4TRA (tRNA<sup>Phe</sup>) の座標が決定されたのは 1987 年であるが、2000 年に 4TRA と同じ結晶を使って tRNA<sup>Phe</sup> の原子座標と温度因子を決め直したデータが PDB に登録された (エントリ名: 1EVV)<sup>158)</sup>。この 1EVV の温度因子から求めた熱ゆらぎを 4TRA の基準振動モード解析結果と比較したものが Fig. 13.3 である。

相互の相関係数は、基準振動モード解析結果と 4TRA の温度因子が前述のように 0.60 であるのに対して、基準振動モード解析結果と 1EVV の温度因子は 0.75 と、より高い一致を示している。1EVV の温度因子と 4TRA の温度因子は 0.89 であった。

1EVV での tRNA<sup>Phe</sup> の立体構造は、4TRA と比較するとリン原子に関する r.m.s.d. が 0.77 Å である。この程度違いであれば、基準振動モード解析による各原子の熱ゆ

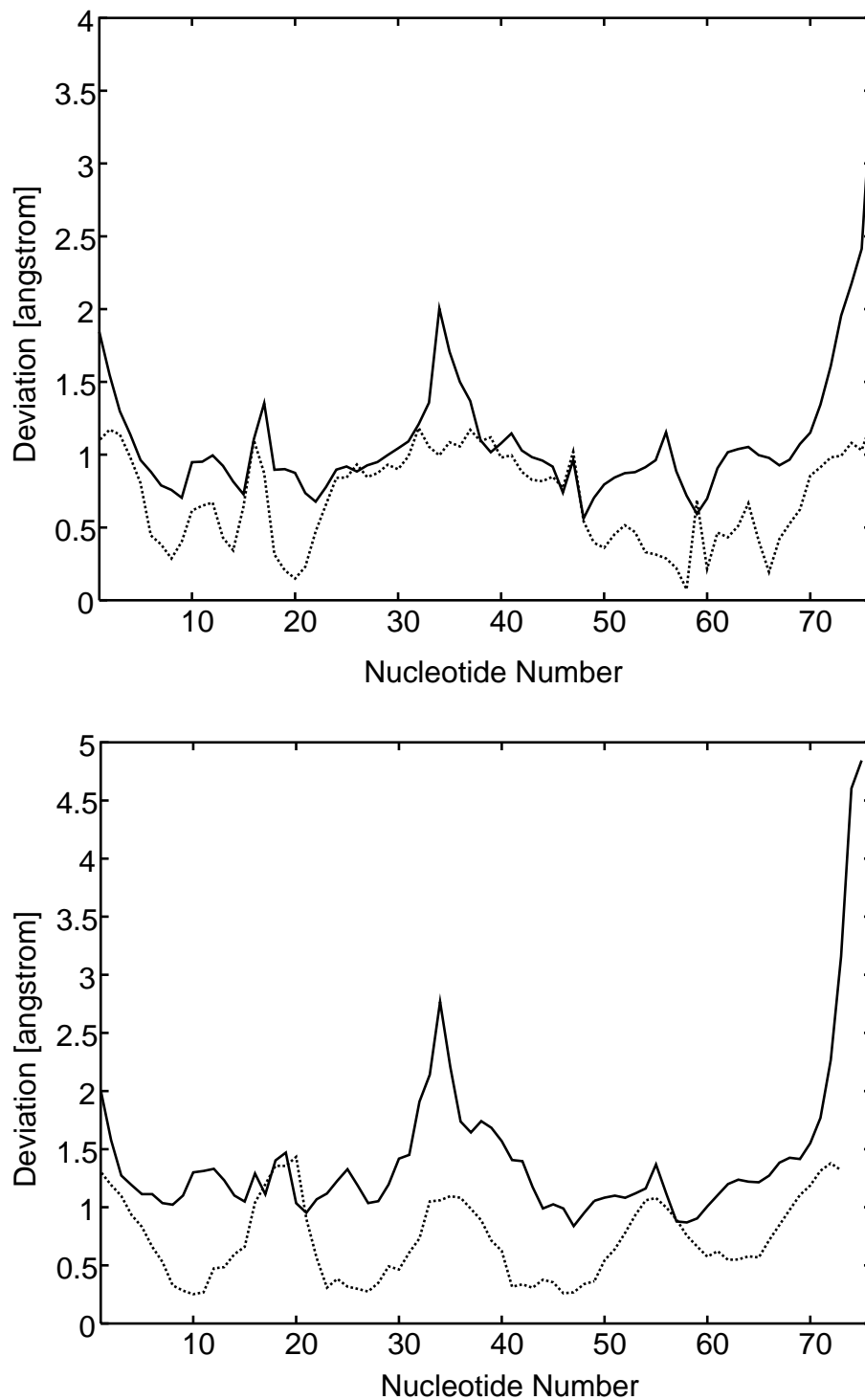
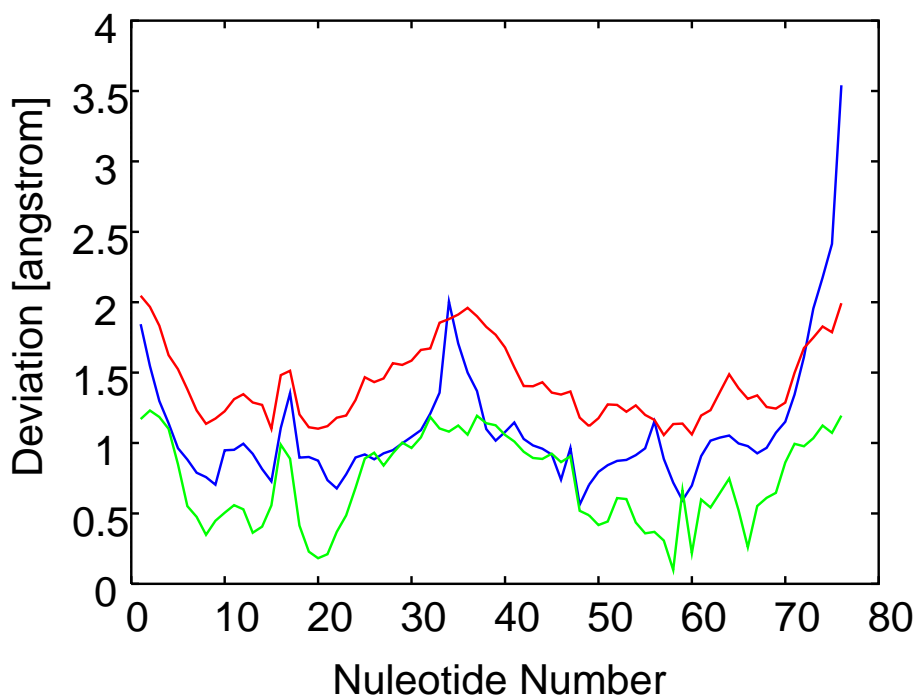


Figure 13.2 フリー状態の tRNA<sup>Phe</sup>(上), tRNA<sup>Asp</sup>(下) における, 基準振動モード解析から得られた熱ゆらぎ(実線)と X 線結晶解析から得られた温度因子から計算した熱ゆらぎ(破線)の大きさの比較. 実線と破線の相関係数は, それぞれ 0.60, 0.51.



**Figure 13.3** 4TRA, 1EVV の温度因子から求めた原子の熱ゆらぎの大きさと 4TRA の基準振動モード解析結果から求めた各原子の熱ゆらぎの大きさの比較。(青) 基準振動モード解析結果。(赤) 1EVV の温度因子。(緑) 4TRA の温度因子。相互の相関係数は、青 - 緑が 0.60, 青 - 赤が 0.75, 赤 - 緑が 0.89。



らぎの大きさは 4TRA と 1EVV でほぼ同じであると考えられる。また、温度因子は原子の熱ゆらぎに加えて結晶格子の乱れの要素を含むため、L 字型構造の端のような部分を除いて基準振動モード解析結果よりも値が大きくなることが多いが、4TRA ではそうっていないのに対して 1EVV ではそれがあてはまっている。これらのことから、tRNA<sup>Phe</sup> の熱ゆらぎを見積もるにあたっては、1EVV の温度因子の方が 4TRA のものよりも適当であり、基準振動モード解析結果から計算された熱ゆらぎとの相関係数は 0.7 を越えると考えられる。また、4TRA と同様に 1987 年に原子座標が決定された tRNA<sup>Asp</sup> についても、現在の X 線結晶解析技術で原子座標や温度因子を決め直せば、基準振動モード解析との相関係数は 2TRA の値 (0.51) よりも上昇する可能性があると考えられる。

### 13.3 低周波数モードにおける熱ゆらぎ

基準振動モード解析の結果得られた tRNA<sup>Phe</sup> および tRNA<sup>Asp</sup> のモード運動を個別に示す ( Fig. 13.4 および Fig. 13.5 )。tRNA のモード運動は、高周波成分は数ヌクレオチドの範囲に限られた局所的なものでその振幅は小さく、低周波数寄りになるにつれて分子全体にわたる協調的な動きがあらわれてくる<sup>27,30</sup>。運動全体への寄与は、低周波成分からのものが大部分であるので、ここでは、最低周波数から 3 つずつのモード運動を示すことにする。各原子の熱ゆらぎの方向と大きさを矢印で表示している。また見やすいように矢印の大きさは 5 倍に拡大してある。

tRNA<sup>Phe</sup>、tRNA<sup>Asp</sup> いずれの場合も、最低モード周波数に対応するモード運動は、L 字型で作られる平面内の bending motion であり、L 字型立体構造の端 (アクセプタステムとアンチコドンループ) が近づいたり離れたりする方向に大きく運動していることが明らかになった。また 2 番目、3 番目に低い周波数モードに対応する運動は、アンチコドンアーム、アクセプタステムそれぞれが、L 字型で作られる平面内、あるいはそれに垂直な方向に運動していた。モード運動の周期は、tRNA<sup>Phe</sup> が 42.2, 32.7, 26.9 ピコ秒 (波数は 0.79, 1.02, 1.24cm<sup>-1</sup>)、tRNA<sup>Asp</sup> が 55.5, 47.0, 31.8 ピコ秒 (波数は 0.60, 0.71, 1.05cm<sup>-1</sup>) であった。

これらのモード運動では、隣り合う原子のモード運動を表す矢印は似た方向を向いており、tRNA を構成している原子が協調的に集団運動 (collective motion) をしている様子がうかがえる。そこで、これらのモード運動をさらに詳細に解析するために、DynDom<sup>159</sup> というツールを適用した。DynDom は、同じタンパク質分子の 2 つの構造データを与えると、両者の間を結ぶような変形において、一体となって動く領域とそれらをつなぐ領域を同定し、さらに一体となって運動する領域同士の

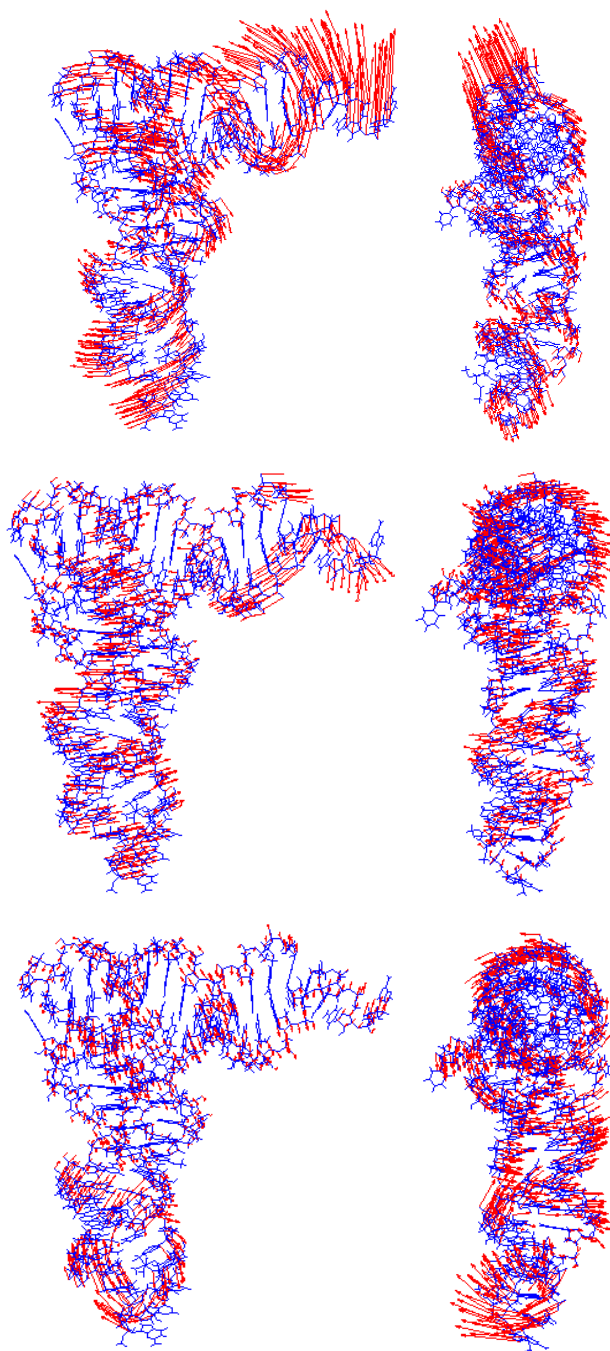


Figure 13.4 tRNA<sup>Phe</sup> の最低周波数領域のモード運動．最低周波数モードから3つのモードについて示す．右図は，左図において紙面に向かって左側から見たもの．青は分子のワイヤーフレーム表示，赤の矢印は，矢印の根元の原子のモード運動の方向と大きさ（片側の振幅）を示す．矢印の大きさは見やすいように5倍に拡大してある．モード周波数は，上から  $0.79\text{cm}^{-1}$  ,  $1.02\text{cm}^{-1}$  ,  $1.24\text{cm}^{-1}$  .

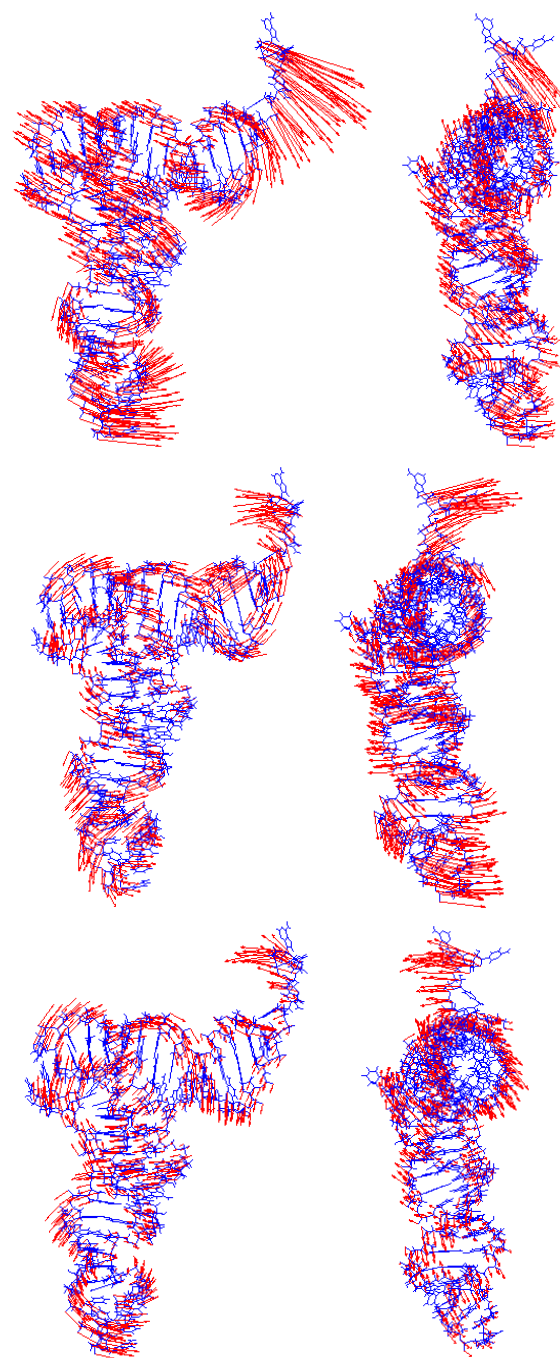


Figure 13.5 tRNA<sup>Asp</sup> の最低周波数領域のモード運動．最低周波数モードから 3 つのモードについて示す．右図は，左図において紙面に向かって左側から見たもの．青は分子のワイヤーフレーム表示，赤の矢印は，矢印の根元の原子のモード運動の方向と大きさ（片側の振幅）を示す．矢印の大きさは見やすいように 5 倍に拡大してある．モード周波数は，上から  $0.60\text{cm}^{-1}$  ,  $0.71\text{cm}^{-1}$  ,  $1.05\text{cm}^{-1}$  .

相対運動を解析できるツールである．本研究では，この DynDom を核酸分子も扱えるように拡張したものをを用いている．

DynDom を用いて，tRNA<sup>Phe</sup> および tRNA<sup>Asp</sup> の最低周波数付近のモード運動（1番目から3番目まで）を解析した結果を Fig. 13.6 および Fig. 13.7 に示す．

青，赤，黄の領域は，それぞれの領域内で原子が一体となって集団運動をしていることを表している．緑の領域は集団運動している領域同士をつなぐ部分であり，灰色は内部変形が大きい（一体となって運動しているとはいえない）領域を示す．また図中の矢印は，集団運動している領域同士の相対運動が hinge bending motion であるとしたときの hinge の軸方向を示したものである．たとえば矢印の頭が赤で，棒が青であった場合，赤の領域と青の領域の相対運動は，その矢印の方向に hinge 軸をもつ hinge bending motion である，ということを表す．

Fig. 13.6 および Fig. 13.7 において tRNA<sup>Phe</sup> と tRNA<sup>Asp</sup> のモード運動を比較すると，立体構造の似ている両者の低周波領域のモード運動には違いがあることがわかる．tRNA<sup>Phe</sup> では，最低周波数モードと2番目に低いモードに対応する運動は，アクセプタステムと残りの部分との相対運動が主である．また3番目に低いモードに対応する運動は，アンチコドンアームと残りの部分の相対運動が主である．いずれの場合も，tRNA<sup>Phe</sup> を2つの領域に分けた運動になっている．これに対して tRNA<sup>Asp</sup> では，最低周波数から3つのモードはいずれも，アクセプタステム，アンチコドンアーム，残りの部分という3つの領域が相対運動をしている．

また，tRNA<sup>Phe</sup>，tRNA<sup>Asp</sup> および第14章で扱う tRNA<sup>Gln</sup> の各原子の熱ゆらぎの大きさ（ヌクレオチド平均）を，ヌクレオチド番号に沿ってプロットし比較したものを，Fig. 13.8 に示す．それぞれの tRNA はヌクレオチド数が異なるが，tRNA のアームやループなどの2次構造を考慮しながらアラインメントしたデータベース tRNA compilation<sup>160,161</sup> にしたがってヌクレオチド番号を合わせてある．

おおよその様子は似ているが，tRNA<sup>Phe</sup> のゆらぎは他の2つの分子に比べて全体的に小さく，また3つの分子の間でゆらぎのピーク位置に違いがみられる．

現時点では，*S. cerevisiae* では，AspRS の立体構造は解かれているが，PheRS の立体構造は PDB に登録されていないので，tRNA<sup>Phe</sup> と PheRS との結合様式が明らかではなく，これらの tRNA のモード運動の違いが tRNA-ARS の分子認識に本質的に重要かどうかという議論は現段階では難しい．しかし将来的に *S. cerevisiae* の PheRS の構造が明らかになれば，Phe 系および Asp 系複合体に対して次章で述べるように基準振動モード解析を適用し，これらのフリー状態でのダイナミクスと複合体でのダイナミクスの違いなどを解析することで，分子認識へのダイナミクスの関与の可能性が議論できると思われる．

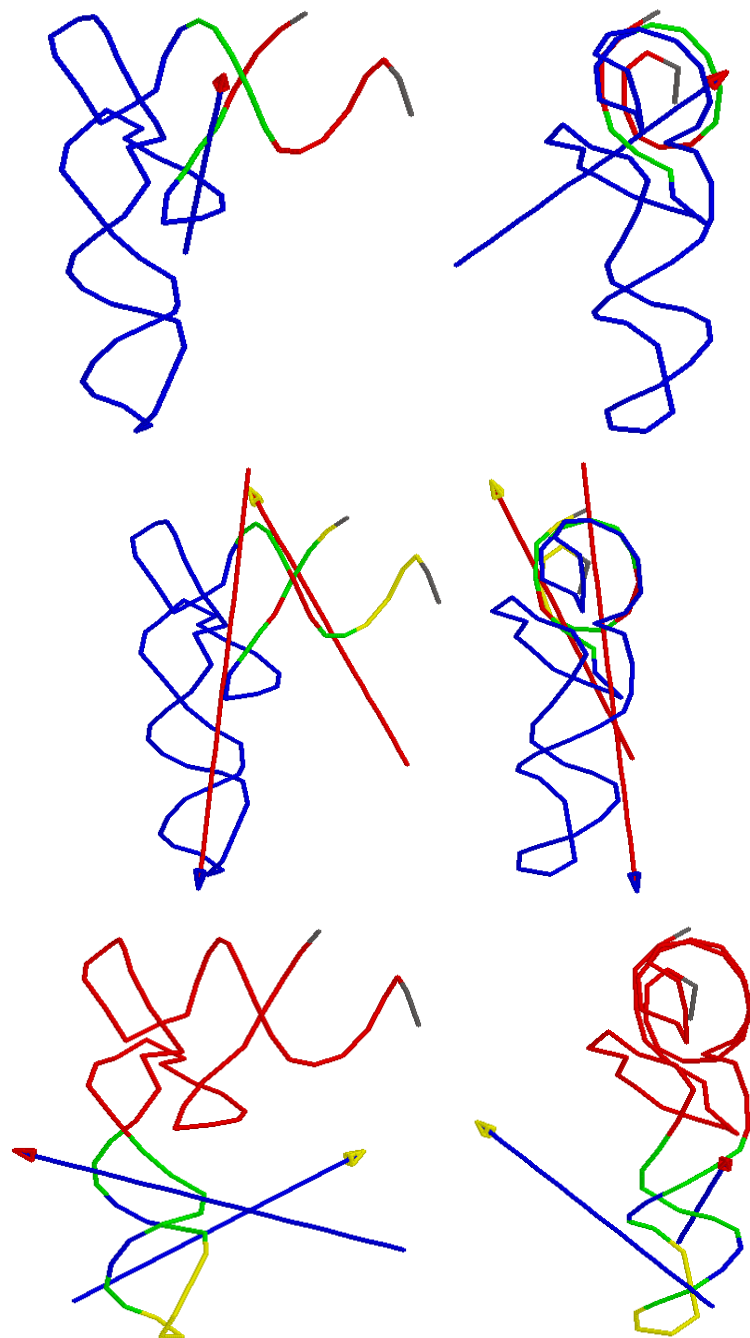


Figure 13.6 tRNA<sup>Phe</sup> の低周波数領域のモード運動を DynDom で解析したもの．最低周波数モードから 3 つのモードについて示す．一体となって運動している領域は同じ色となっている．またそのような領域同士の相対運動の軸が，双方の色をもつ矢印で表されている．なお，緑色はそのような領域の間をつなぐ部分と判定されたところであり，灰色は内部変形が大きかった部分である．

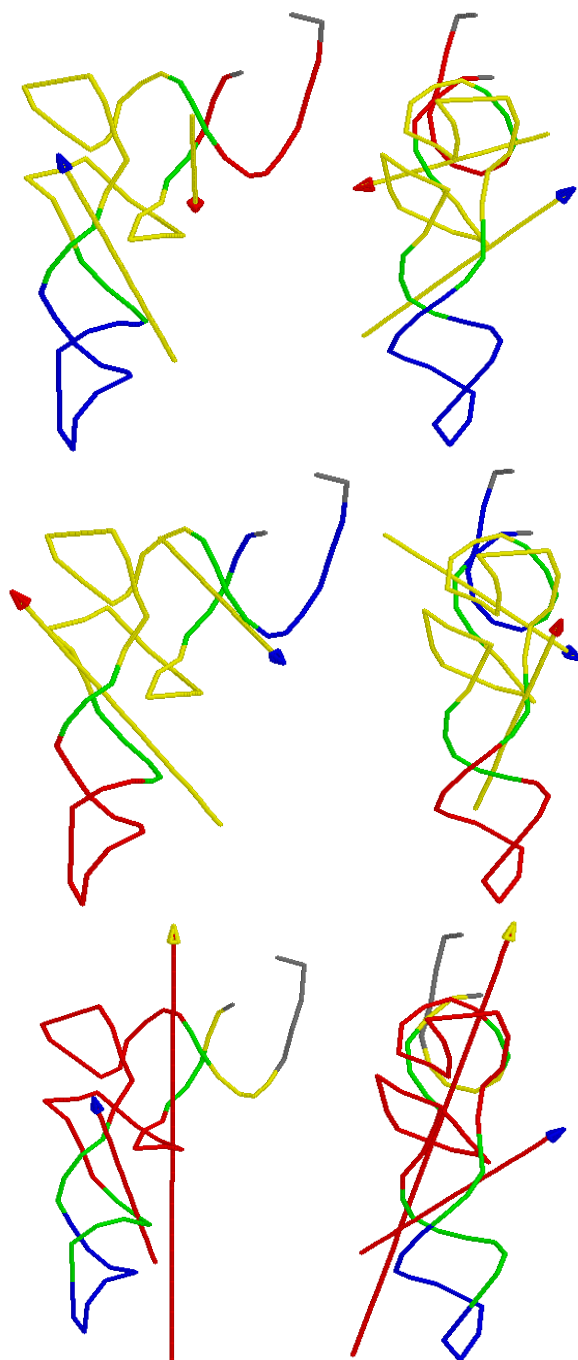


Figure 13.7 tRNA<sup>ASP</sup> の低周波数領域のモード運動を DynDom で解析したもの．最低周波数モードから 3 つのモードについて示す．一体となって運動している領域は同じ色となっている．またそのような領域同士の相対運動の軸が，双方の色をもつ矢印で表されている．なお，緑色はそのような領域の間をつなぐ部分と判定されたところであり，灰色は内部変形が大きかった部分である．

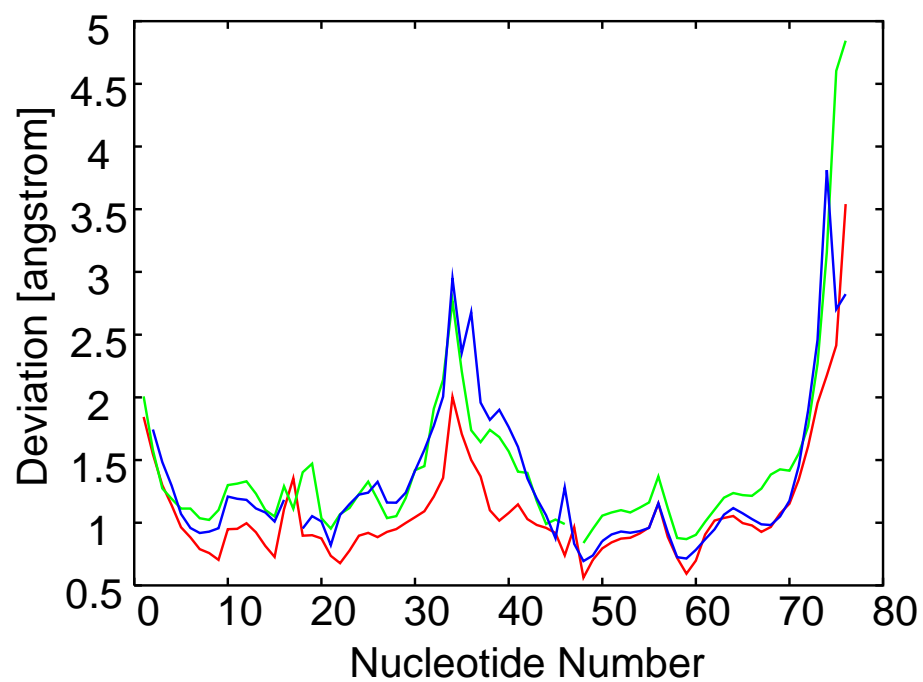


Figure 13.8 基準振動モード解析から得られたフリー状態の tRNA<sup>Phe</sup>(赤), tRNA<sup>Asp</sup>(青), tRNA<sup>Gln</sup>(緑)における熱ゆらぎの大きさの比較. ヌクレオチド番号は, tRNA compilation<sup>160,161</sup> にしたがって合わせてある.

## 第 14 章

# tRNA・アミノアシル tRNA 合成酵素複合体のダイナミクス解析

この章では、*E. coli* tRNA<sup>Gln</sup>-GlnRS 複合体について、フリー状態と複合体形成状態の tRNA および ARS の基準振動モードの違いを解析した結果について述べる。

### 14.1 フリー状態の構造の構築

現時点で、PDB にはフリー状態の tRNA<sup>Gln</sup> および GlnRS の構造は登録されていない。このため、以下の手順で計算に用いるフリー状態の tRNA<sup>Gln</sup> と GlnRS の構造を構築する。

1. *E. coli* tRNA<sup>Gln</sup>-GlnRS 複合体の結晶構造 (PDB エントリ名: 1GTR) をもとにして、第 12.2 節の手順にしたがって複合体の初期構造を構築する。
2. 手順 1 の構造から tRNA<sup>Gln</sup> のみ、および GlnRS のみを取り出し、二面角系において最急降下法とニュートン・ラフソン法によるエネルギー極小化をそれぞれについて行う。このとき、Hessian の固有値がすべて正となり、かつ 1 ステップあたりのエネルギー変化が  $1.0 \times 10^{-5}$  kcal/mol 以下となるまでエネルギー極小化を行う。
3. 手順 2 で得られた tRNA<sup>Gln</sup>、GlnRS のエネルギー極小構造に対して、基準振動モード解析計算をそれぞれ実行する。

1GTR 中の結晶構造との r.m.s.d. (tRNA<sup>Gln</sup> はリン原子、GlnRS は C<sup>α</sup> 原子について計算) は、tRNA<sup>Gln</sup> が 1.6 Å、GlnRS が 0.56 Å である。

このように、本研究では、複合体中の構造を初期構造としてエネルギー極小化を行った構造を、フリー状態の tRNA<sup>Gln</sup>、GlnRS の構造として用いた。これは、両分



子のフリー状態でのモード運動を複合体状態でのモード運動と比較するにあたって、複合体状態との差を「相手分子が存在しない」という要素以外にできるだけ少なくした方が、複合体形成に伴うダイナミクス変化を直接的に解析できると考えたためである。

これらのフリー状態の構造には、たとえば複合体中の構造を初期構造として、水分子を陽に考慮した分子動力学計算を長時間行って多数の構造をサンプリングし、それらの平均構造を用いるという考え方もある。これを試すことは今後の課題である。

## 14.2 モード周波数の解析結果

複合体およびフリー状態の tRNA<sup>Gln</sup>、GlnRS 分子について、求められた基準振動モードのモード周波数分布を Fig. 14.1 に示す。また、これらの周波数分布の最低値・最高値・平均値および 20cm<sup>-1</sup>以下のモード数の全体に対する割合を Table 14.1 に示す。

Table 14.1 tRNA<sup>Gln</sup>-GlnRS 複合体とフリー状態の tRNA<sup>Gln</sup> および GlnRS のモード周波数分布の比較。単位は cm<sup>-1</sup>。

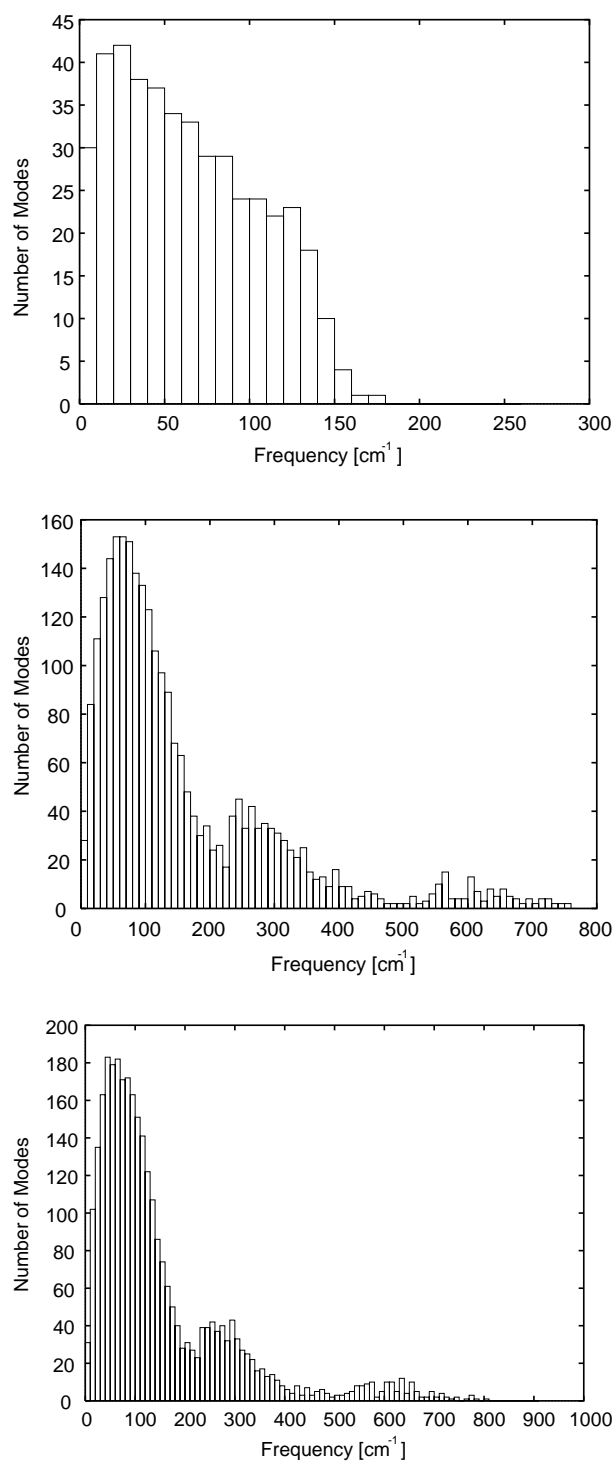
分子	最低	最高	平均	20cm <sup>-1</sup> 以下のモード数 (%)
tRNA <sup>Gln</sup>	0.70	261.7	65.2	16.1
GlnRS	1.54	775.9	159.1	4.3
複合体	1.42	910.7	150.4	4.3

フリー状態の tRNA<sup>Gln</sup> のモード周波数分布は、Fig. 13.1 での tRNA<sup>Phe</sup>、tRNA<sup>Asp</sup> と似た分布を示した。また最低周波数は、Table 13.1 の tRNA<sup>Phe</sup>、tRNA<sup>Asp</sup> とほぼ同じであった。

複合体形成の結果、tRNA<sup>Gln</sup> の最低周波数は高周波数寄りに、また GlnRS の最低周波数は、わずかに低周波数寄りにシフトした。複合体のモード周波数の分布は、GlnRS の分布と大きな差はみられなかった。

## 14.3 各原子の熱ゆらぎの大きさ

複合体状態の tRNA<sup>Gln</sup> および GlnRS の基準振動モード解析から得られた熱ゆらぎの大きさと X 線結晶解析の温度因子データから得られた熱ゆらぎの大きさの比較を Fig. 14.2 に示す。値は、残基・ヌクレオチドごとに平均したものである。



**Figure 14.1** モード周波数の分布．(上) フリー状態の tRNA<sup>Gln</sup>．(中) フリー状態の GlnRS．(下) tRNA<sup>Gln</sup>-GlnRS 複合体．

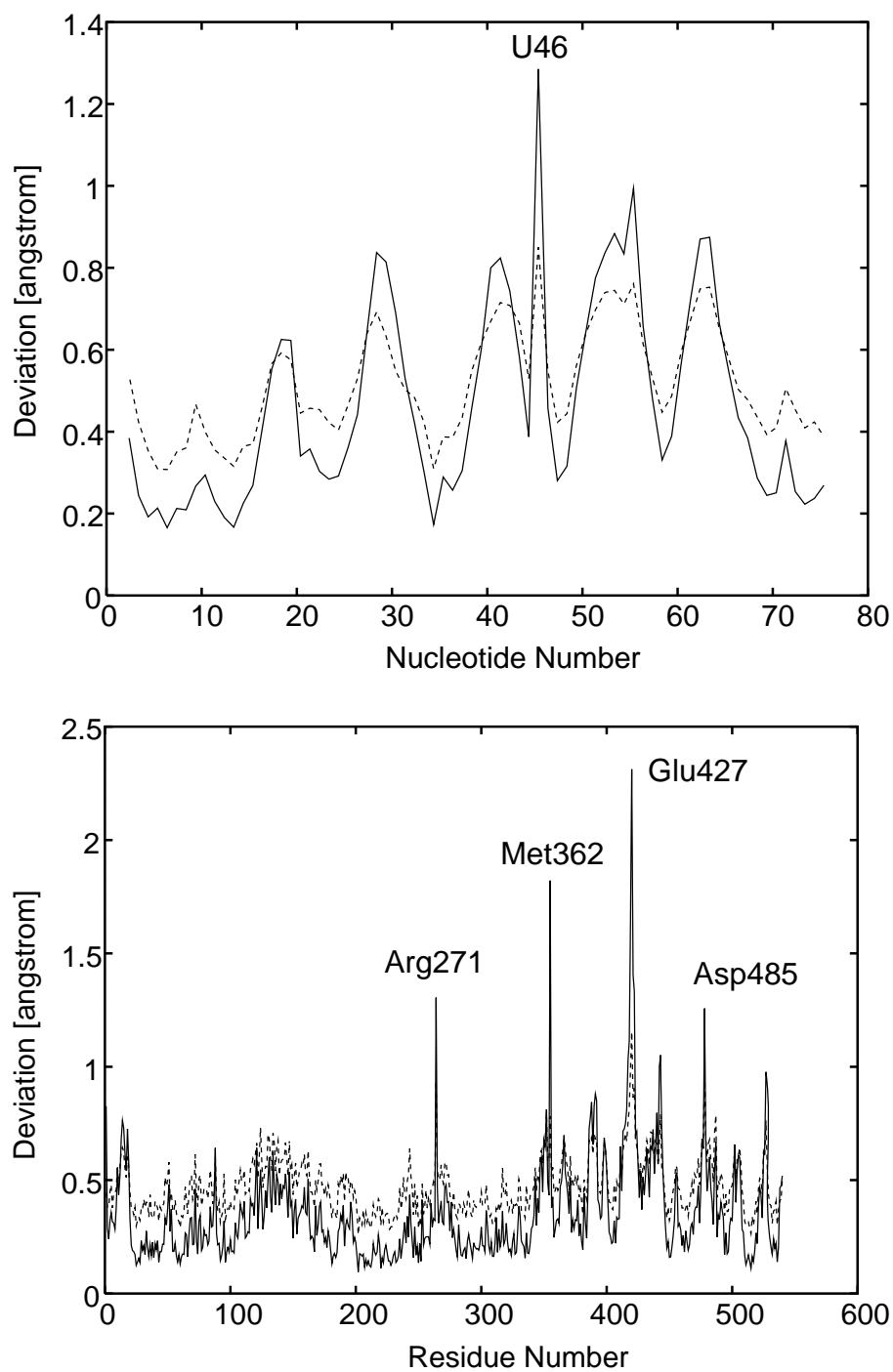


Figure 14.2 複合体中における、基準振動モード解析から得られた熱ゆらぎ (実線) と X 線結晶解析から得られた温度因子から計算した熱ゆらぎ (破線) の大きさの比較。(上) tRNA<sup>Gln</sup> の部分。相関係数は 0.97。(下) GlnRS の部分。相関係数は 0.90。

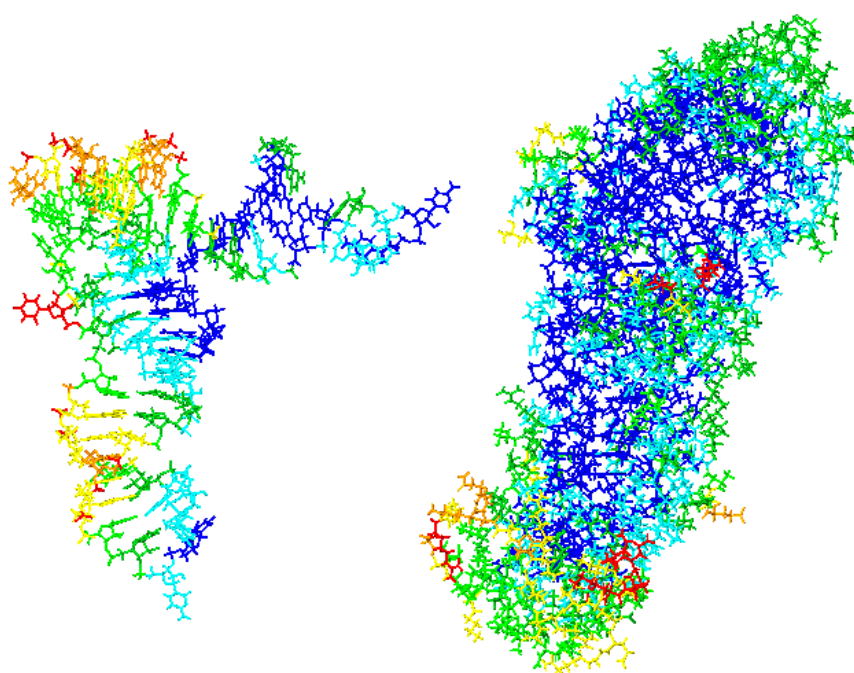


Figure 14.3 複合体中の tRNA<sup>Gln</sup>(左) および GlnRS(右) の熱ゆらぎの大きさ．赤い色ほどその残基が大きな熱ゆらぎをしており，青い色ほどその残基のゆらぎが小さいことを示す．両者の中間付近が緑色である．見やすいように，tRNA<sup>Gln</sup> と GlnRS を離して表示している．

tRNA<sup>Gln</sup>, GlnRS のいずれの場合も, 温度因子と基準振動モード解析結果は非常によく一致し, tRNA<sup>Phe</sup>・tRNA<sup>Asp</sup> のときと同様, 本研究のシミュレーションの妥当性を示している. 相関係数は, それぞれ 0.97, 0.90 であった. 各残基・ヌクレオチドの熱ゆらぎの大きさを立体構造に色づけして示したものが Fig. 14.3 である. 左が tRNA<sup>Gln</sup>, 右が GlnRS で, 赤い色ほどその残基・ヌクレオチドが大きな熱ゆらぎをしており, 青い色ほど逆にその残基・ヌクレオチドの熱ゆらぎが小さいことを表す. 両者の中間付近が緑色である. なお, 図では見やすいように tRNA<sup>Gln</sup> と GlnRS を離して表示している.

ゆらぎの大きさは, tRNA<sup>Gln</sup>, GlnRS のいずれについても, 一部の残基やヌクレオチドを除いて 1 Å 以下であった. 複合体中の tRNA<sup>Gln</sup> は, アクセプトステムやアンチコドンアームの 3' 末端側の鎖などの GlnRS との接触面でゆらぎが抑えられている一方で, 接触面から遠い側の D アームや T アーム, 分子外に大きく突出している U46, アンチコドンアームなどで, ゆらぎが大きい部分がみられる. tRNA<sup>Gln</sup> のコアは, U8-C16, A21-C27, G43-C48, A59 付近であるが, この領域は前述の U46 を除くと比較的熱ゆらぎが小さい. これはコア内の高次水素結合やスタッキングによってこの領域が安定化されていることを示すと思われる. また複合体中の GlnRS は, 分子外へ露出している Arg271, Met362, Glu427, Asp485 でゆらぎのピークがみられ, アンチコドン認識ドメインで比較的大きなゆらぎがみられた. アミノアシル化の活性部位付近ではゆらぎは小さかった.

複合体状態とフリー状態の熱ゆらぎの大きさの差を示したものが Fig. 14.4 である. 正の値は複合体形成によって熱ゆらぎが大きくなったことを, 負の値は複合体形成によって熱ゆらぎが抑えられたことを表す. 図の上にヌクレオチド・残基番号に対するプロットを, 下に立体構造に色づけしたものを示す. 色づけでは, ゼロに近い値ほど赤い色を, 負の絶対値が大きいほど青い色をつけてある. 両者の中間付近が緑色である.

複合体形成によって, ほとんどすべてのヌクレオチド・アミノ酸においてゆらぎの大きさは減少した. とくに tRNA<sup>Gln</sup> と GlnRS が接触している広い範囲で, ゆらぎの大きな減少がみられた. これに対して, tRNA<sup>Gln</sup> の T アーム, D ループ, エキストラループ, GlnRS のアンチコドン認識ドメインなどでは, ゆらぎの減少の割合は比較的小さかった.

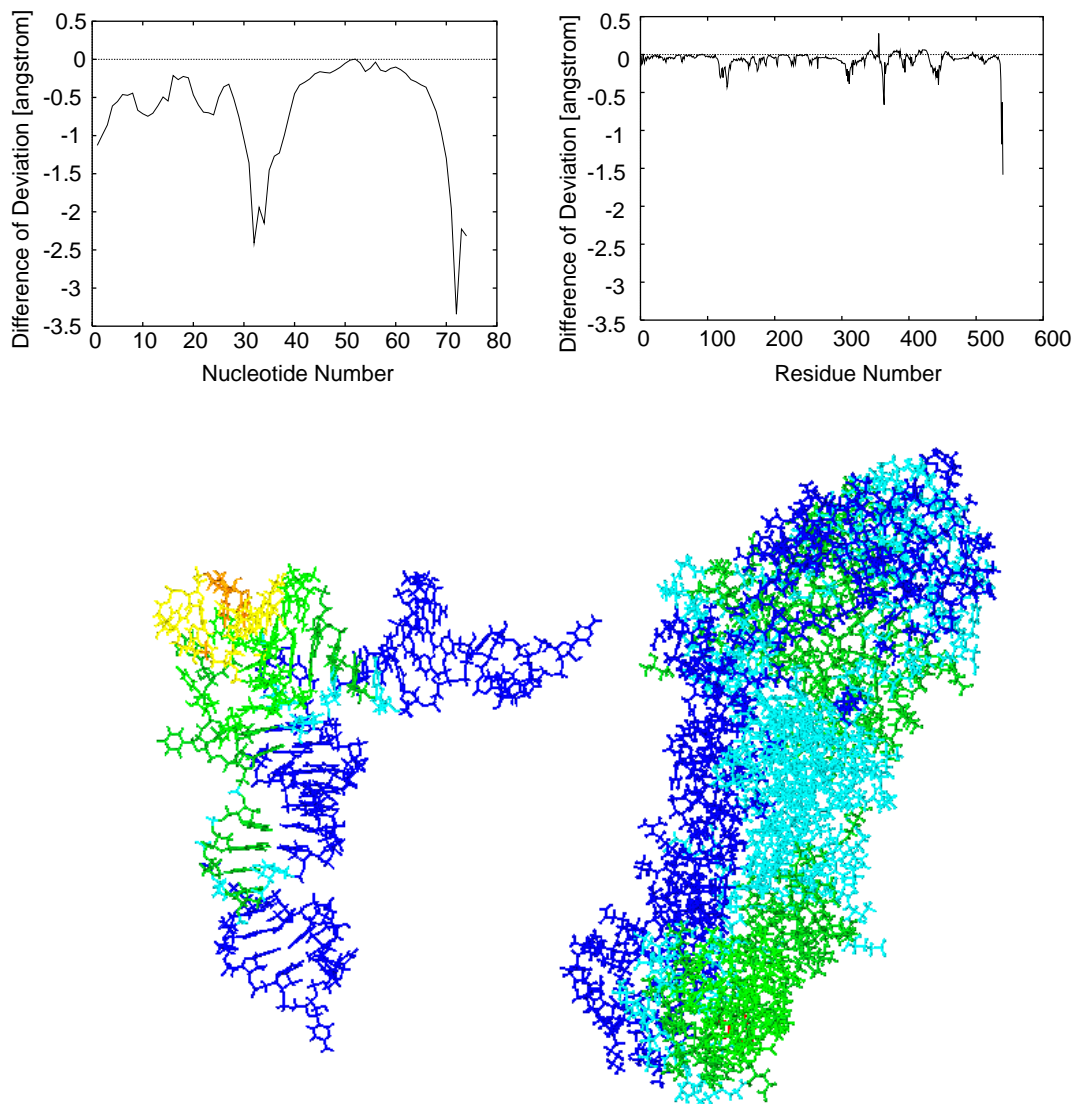


Figure 14.4 複合体を形成したときの  $\text{tRNA}^{\text{Gln}}$  および GlnRS の熱ゆらぎの変化の大きさ。上のグラフはフリー状態に対する複合体形成時での値の差を、ヌクレオチド・残基の番号に沿ってプロットしたもの。(左上)  $\text{tRNA}^{\text{Gln}}$ 。(右上) GlnRS。下は、上のグラフを立体構造の中で色づけして表示したもの。青い色ほど、フリー状態と比較したときに、複合体形成によってその残基の熱ゆらぎがより抑えられたことを示し、赤い色ほど、複合体形成によるその残基のゆらぎの抑制が小さいことを示す。両者の中間付近が緑色である。見やすいように、 $\text{tRNA}^{\text{Gln}}$  と GlnRS を離して表示している。

#### 14.4 低周波数モードにおける熱ゆらぎ

基準振動モード解析の結果得られた tRNA<sup>Gln</sup>-GlnRS 複合体の低周波領域のモード運動を個別に示す．Fig. 14.5 は，最低周波数から 3 つのモード運動における各原子の熱ゆらぎの大きさと方向を矢印で示したものである．青が tRNA<sup>Gln</sup>，赤が GlnRS の原子の運動を示す．矢印の大きさは，見やすいように 15 倍に拡大して表示している．

フリー状態の tRNA<sup>Phe</sup> あるいは tRNA<sup>Asp</sup> のときには L 字型で作られる平面内の bending motion がみられたが，複合体の低周波数領域ではこのような運動はみられなかった．また特徴的なのは，tRNA<sup>Gln</sup> と GlnRS の運動がお互いに大きくカップリングしていることである．Ishida らによる subtilisin と eglin c のタンパク質-タンパク質複合体に基準振動モード解析を適用した結果では，最低周波数から 3 つのモードはいずれも 2 つのタンパク質の間の相対運動であり，それぞれのタンパク質の内部変形はそれよりも高周波寄りにあらわれている<sup>162)</sup>．これに対して tRNA<sup>Gln</sup>-GlnRS 複合体の最低周波数領域では，tRNA<sup>Gln</sup> と GlnRS が一体となって運動しており，様子が大きく異なっている．モード運動の周期は，23.5, 17.7, 13.5 ピコ秒 (波数は 1.42, 1.88, 2.47 cm<sup>-1</sup>) であった．また，各周波数モードでの原子の熱ゆらぎの平均値をモード番号に沿ってプロットしたものを Fig. 14.6 に示す．最低周波数領域の少数のモード運動のサイズが，高周波寄りのモード運動に比べて非常に大きいことがわかる．

基準振動モード解析では，tRNA<sup>Gln</sup> および GlnRS 内のある原子  $\alpha$  の平均位置  $r_\alpha$  からのゆらぎ  $\Delta r_\alpha$  を，次式のように分子内の変形に起因する成分 (内部運動)  $\Delta r_\alpha^i$  と分子間の相対運動に起因する成分 (外部運動)  $\Delta r_\alpha^e$  に分けることができる．

$$\Delta r_\alpha = \Delta r_\alpha^i + \Delta r_\alpha^e \quad (14.1)$$

$$\langle \Delta r_\alpha^2 \rangle = \langle (\Delta r_\alpha^i)^2 \rangle + \langle (\Delta r_\alpha^e)^2 \rangle + 2\langle \Delta r_\alpha^e \cdot \Delta r_\alpha^i \rangle \quad (14.2)$$

ただし  $\langle \rangle$  は時間平均を表す．式 (14.2) の右辺第 1 項が内部運動のみに起因するゆらぎ，第 2 項が外部運動のみに起因するゆらぎで，第 3 項は外部運動と内部運動の相関の大きさを表す．

最低周波数に対応するモード運動において，tRNA<sup>Gln</sup>-GlnRS 複合体の外部運動のみに起因するゆらぎを取り出したものが，Fig. 14.7 である．青が tRNA<sup>Gln</sup>，赤が GlnRS の各原子の運動で，矢印の大きさは見やすいように 15 倍に拡大してある．

最低周波数モードでの外部運動は，tRNA<sup>Gln</sup> が GlnRS との接触を保ちながら，GlnRS の上を滑るような運動であった．外部運動の大きさは，tRNA<sup>Gln</sup> の場合で内部運

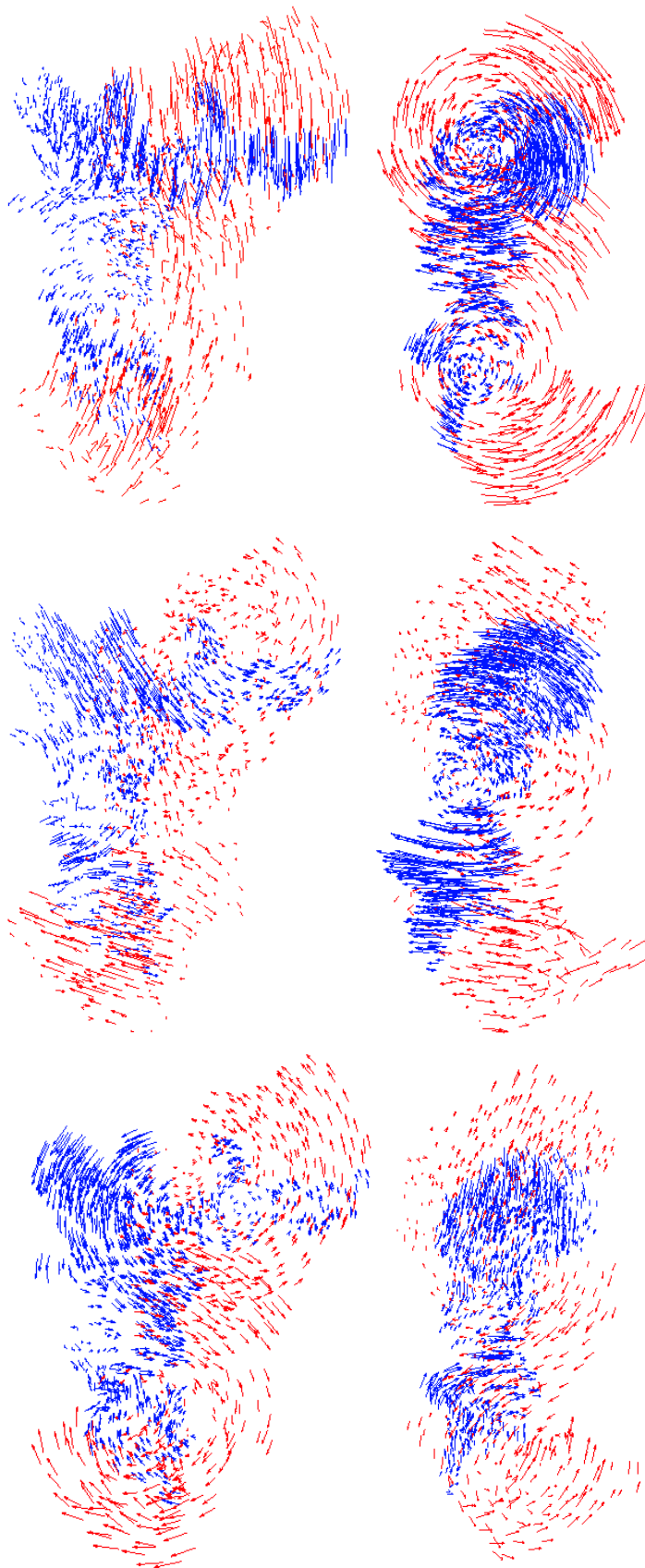


Figure 14.5 tRNA<sup>Gln</sup>-GlnRS 複合体の低周波数領域のモード運動の様子．最低周波数モードから3つのモードについて示す．青が tRNA<sup>Gln</sup>，赤が GlnRS．矢印の大きさは見やすいように15倍に拡大してある．モード周波数は，上から  $1.42\text{cm}^{-1}$ ， $1.88\text{cm}^{-1}$ ， $2.47\text{cm}^{-1}$ ．



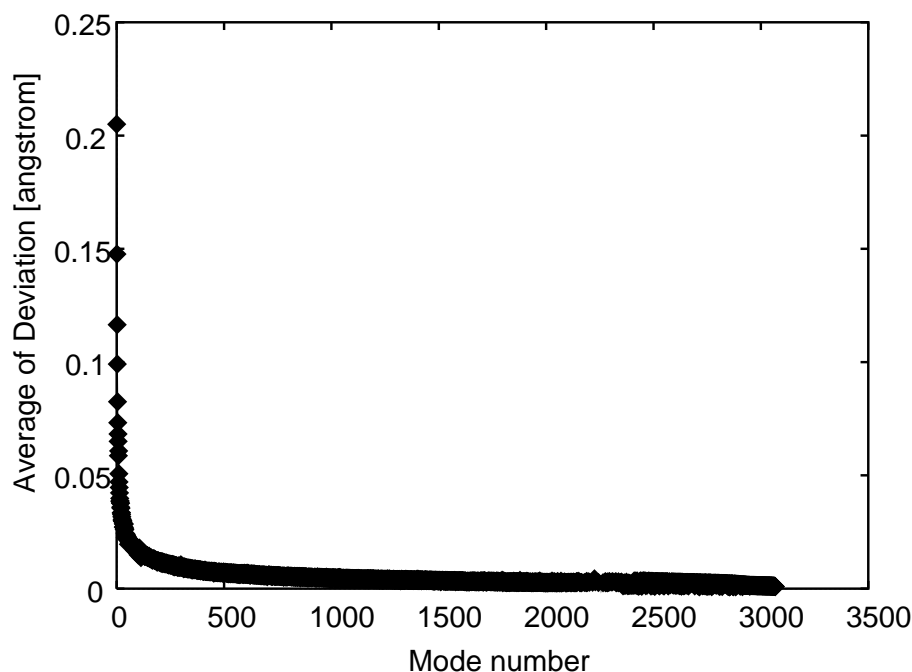
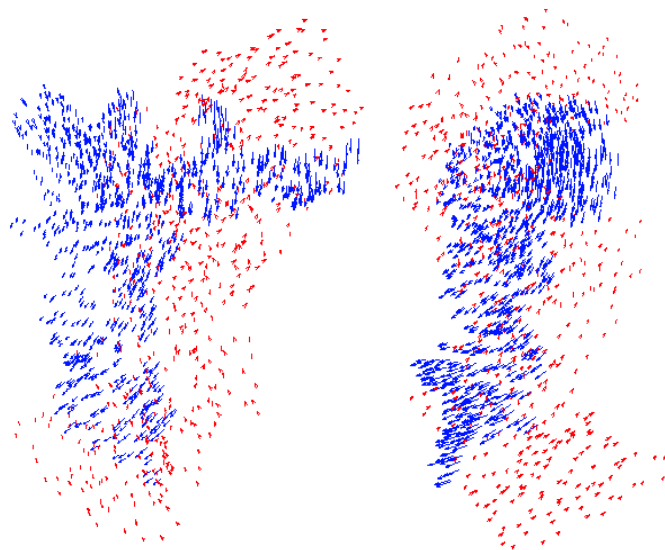


Figure 14.6 モード番号に対する，各原子の熱ゆらぎの大きさの平均値．

動の平均 26%，GlnRS の場合で平均 4.5% と小さく，先に述べたように低周波数領域では内部運動からの寄与が大部分であることを示している．

式 (14.2) の右辺第 3 項で表される内部運動と外部運動の相関を示したものが，Fig. 14.8 である．上のグラフは相関の大きさを残基・ヌクレオチド番号に沿ってプロットしたものである．また下は上のグラフを立体構造の中に色づけして示したものであり，絶対値が大きな正の相関があるほど赤い色，絶対値が大きな負の相関があるほど青い色，ゼロ付近は緑色で表している．内部運動と外部運動の相関値は，式 (14.2) で示したように内部運動ベクトル  $\Delta r_{\alpha}^i$  と外部運動ベクトル  $\Delta r_{\alpha}^e$  の内積であるから，内部運動ベクトルと外部運動ベクトルが同じような方向を向いていると正の値，逆の方向を向いていると負の値，直交していればゼロとなる．

内部運動と外部運動の相関値は，tRNA<sup>Gln</sup> と比べると GlnRS では全体として小さい．tRNA<sup>Gln</sup> では，とくにアクセプタステムおよびアンチコドンループで絶対値の大きな負の相関がみられ，逆に T アーム付近では正の相関となっている．これは，Fig. 14.3 で，複合体中の tRNA<sup>Gln</sup> と GlnRS の接触部位のゆらぎが抑えられているのは，内部運動 (tRNA<sup>Gln</sup> および GlnRS 双方の内部変形) と外部運動 (両分子が形を変えない相対運動) が逆の方向であることで，全体としてゆらぎが小さくなっているということを示している．この描像は，subtilisin-eglin c 複合体の場合と似てい



**Figure 14.7** tRNA<sup>Gln</sup>-GlnRS 複合体の最低周波数モードにおける外部運動成分．青が tRNA<sup>Gln</sup>，赤が GlnRS．矢印の大きさは見やすいように 15 倍に拡大してある．

る<sup>162)</sup>．

先に述べたように，tRNA<sup>Gln</sup>-GlnRS 複合体は，低周波数領域では tRNA<sup>Gln</sup> と GlnRS が一体となった協調的な運動がみられた．これをさらに詳細に解析するために，第 13.3 節と同様に DynDom を使って，一体となって運動している領域の同定を試みた．最低周波数から 3 つのモード運動の内部運動に DynDom を適用した結果を Fig. 14.9 に示す．それぞれの左の図では，tRNA<sup>Gln</sup> をワイヤーフレーム表示，GlnRS をリボン表示し，一体となって運動している部分は，黄，青，赤，マゼンタで色づけしてある．また緑は一体となって運動している領域をつなぐ部分であり，内部変形が大きかった部分は灰色となっている．右の図では，一体となって運動している領域同士の相対運動の hinge 軸方向が，双方の色をもつ矢印で示されている．

DynDom を用いた解析から，低周波数領域では，GlnRS の主要な運動はドメイン間の相対運動であることが明らかになった．Fig. 11.2 に示したように，GlnRS は構造的にいくつかのドメインに分かれるが，Fig. 14.9 に示した最低周波数モードから 3 つのモードにおいては，GlnRS の触媒ドメインと tRNA<sup>Gln</sup> のアクセプタステム，あるいはアクセプタステムと T アームの両方が同じ黄色になっており，両者が一体となって運動していることを示している．また同様に，GlnRS のアンチコドン認識ドメインと tRNA<sup>Gln</sup> のアンチコドンアーム (青色) が，一体となって運動している．GlnRS の helical サブドメイン (第 2, 3 番目のモードでの赤い部分) は，それ自身がほぼひとかたまりで運動している．このように，GlnRS の低周波数領域では，ド

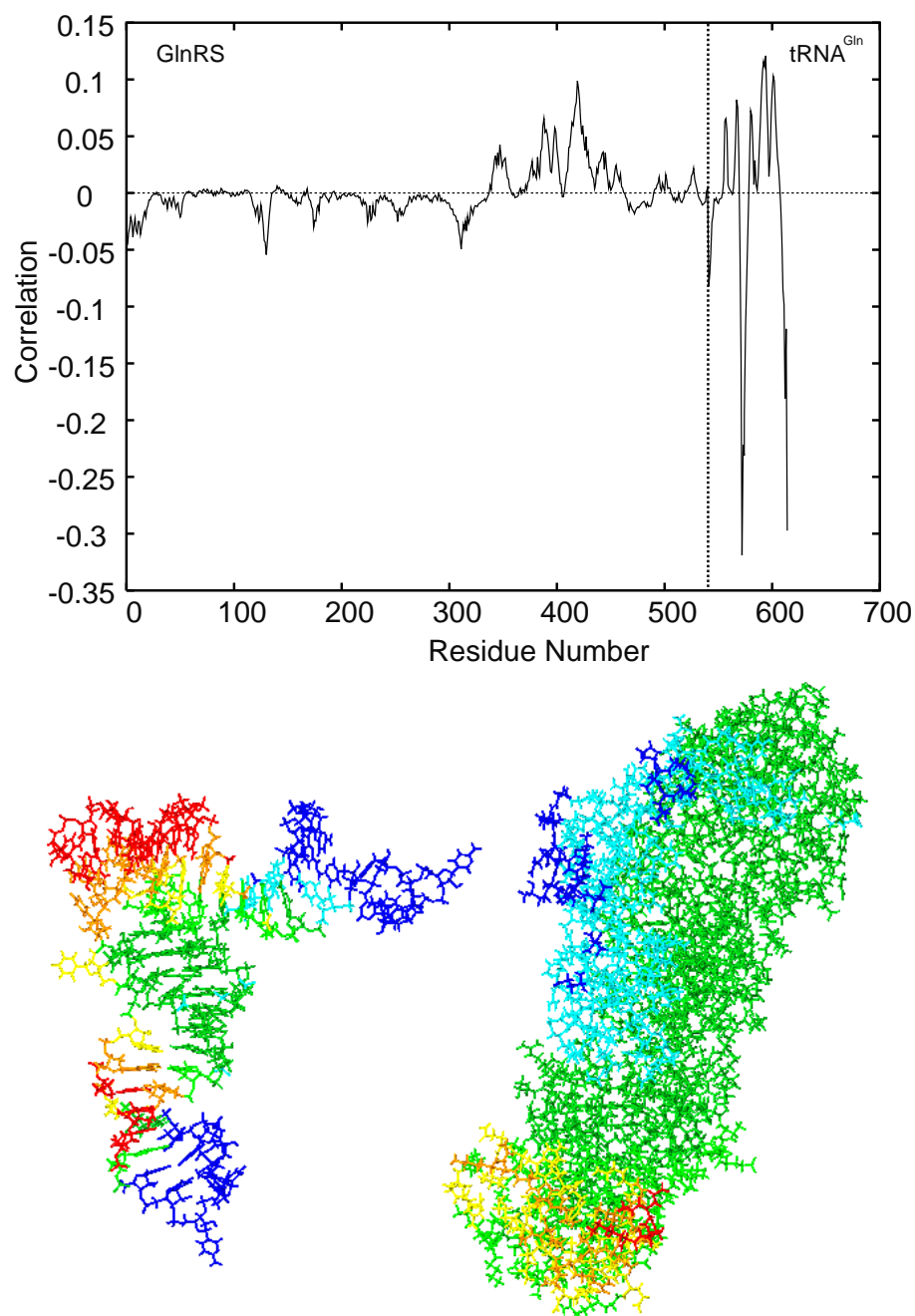
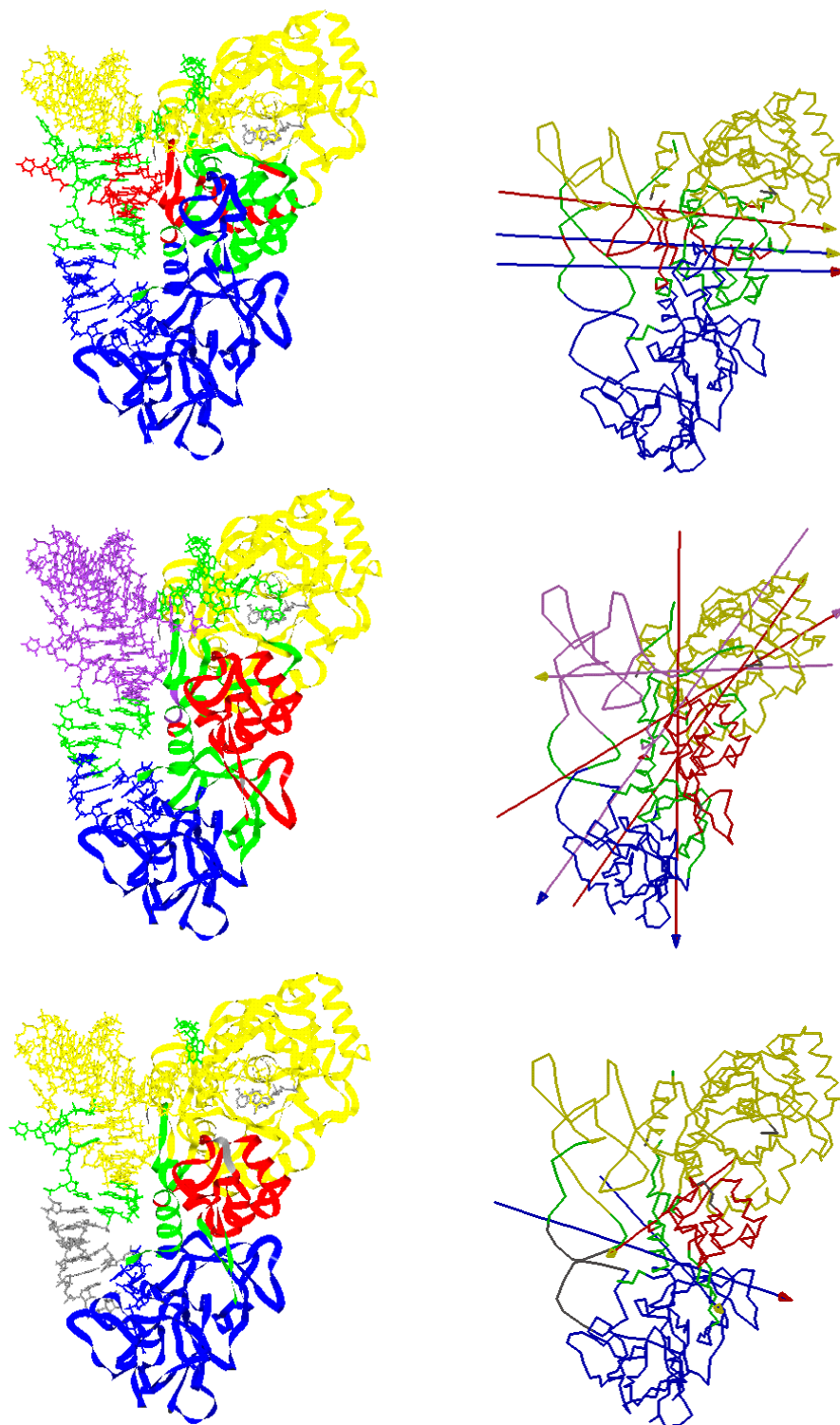


Figure 14.8 内部運動と外部運動の相関．上は，残基・ヌクレオチドの番号にしたがってプロットしたもの．下は，相関値を立体構造の中で色づけして表示したもの．赤い色ほど正の相関が大きく，青い色ほど負の相関が大きいことを示す．見やすいように，tRNA<sup>Gln</sup>とGlnRSを離して表示している．



**Figure 14.9** tRNA<sup>Gln</sup>-GlnRS 複合体の低周波数領域のモード運動を DynDom で解析したもの。最低周波数モードから3つのモードについて示す。左の図では、ワイヤーフレーム表示が tRNA<sup>Gln</sup>、リボン表示が GlnRS である。一体となって運動している領域は同じ色となっている。またそのような領域同士の相対運動の軸が、右の図で双方の色をもつ矢印で表されている。なお、緑色はそのような領域の間をつなぐ部分と判定されたところであり、灰色は内部変形が大きかった部分である。

メイン内の変形はそれほど大きくない。一方 tRNA<sup>Gln</sup> は、Fig. 13.6 および Fig. 13.7 でフリー状態の tRNA<sup>Phe</sup> および tRNA<sup>Asp</sup> でみられたのと同様、アクセプタステム、アンチコドンアーム、それ以外の 3 つの領域に分かれるような運動をしている。このことと、Table 14.1 での tRNA<sup>Gln</sup>, GlnRS 両分子の周波数分布の違いを考え合わせると、低周波数領域では、GlnRS は「固い」ドメイン同士の相対運動が主であるのに対して、tRNA<sup>Gln</sup> は二重らせん部が柔らかく、ゆっくりとした変形をするという、核酸とタンパク質のダイナミクスの特徴の違いを端的に表していると考えられる。

#### 14.5 複合体の基準振動モード運動のフリー状態との比較

低周波数領域において、複合体状態での基準振動モードとフリー状態での基準振動モードの類似性を解析することで、2 つの状態でのダイナミクスの相違を調べることができると考えられる。そこで、tRNA<sup>Gln</sup> および GlnRS について、以下の式で表される値を 2 つのモード運動の類似度  $S$  と定義して解析を行った。

$$S = \sum_{\alpha=1}^N \frac{\mathbf{d}_{\alpha}^f \cdot \mathbf{d}_{\alpha}^c}{|\mathbf{d}_{\alpha}^f| |\mathbf{d}_{\alpha}^c|} / N \quad (14.3)$$

ただし、 $\mathbf{d}_{\alpha}^f$  は原子  $\alpha$  のフリー状態でのモード運動を表すベクトル (矢印)、 $\mathbf{d}_{\alpha}^c$  は原子  $\alpha$  の複合体状態でのモード運動を表すベクトル (矢印)、 $N$  は分子を構成する原子の個数である。なお、フリー状態でのモード運動には外部運動は存在しないため、この場合の複合体のモード運動とは、外部運動の効果を除いた内部運動のみを考慮したものである。

式 (14.3) の右辺は、各原子のフリー状態と複合体状態のモード運動を表すベクトルの間の角度を  $\theta_{\alpha}$  とすると、 $\cos \theta_{\alpha}$  の平均をとったものに相当する。モード運動ベクトル  $\mathbf{d}_{\alpha}$  は原子  $\alpha$  の振動の片側振幅を示すため、式 (14.3) の計算では  $\theta_{\alpha}$  が 90 度よりも大きくなる場合はベクトルの向きを反転させて常に  $\theta_{\alpha}$  が 90 度よりも小さくなるように考慮しながら計算する必要がある。すべての原子についてベクトルの方向 (長さは問わない) が完全に一致すると、類似度  $S$  は 1 となり、すべての原子についてベクトルが直交していると類似度  $S$  は 0 になる。

この  $S$  の値を tRNA<sup>Gln</sup>, GlnRS について色で表示したものが、Fig. 14.10 である。上が tRNA<sup>Gln</sup>, 下が GlnRS であり、縦軸は複合体状態、横軸はフリー状態でのモード番号である。赤い色ほど類似度が高く、青い色ほど類似度が低いことを表す。両者の中間付近が緑色である。なお、モード番号は「0」が最低周波数モードを

表し、周波数の昇順となっている。複合体状態・フリー状態の双方でのモード番号 50 までの相互の相関を示す。

tRNA<sup>Gln</sup> の複合体中のモード運動に着目すると、モード番号 0 ~11 の 12 個のうち 6 つについて、類似度  $S$  が 0.65 を越えるフリー状態の対応モードが存在していた。また、Fig. 14.10 をみると、たとえば、tRNA<sup>Gln</sup> の複合体状態のモード番号 2, 6, 8 などは、複合体中には存在するがフリー状態では対応するモードが存在しない、ということがわかる。逆にフリー状態のモード番号 0, 3, 7 などは、複合体状態では存在しない。またフリー状態のモード番号 2 は、複合体状態の複数の周波数モード 0, 4, 9 に現われている。

これらのモード運動の対応を表にまとめたものが Table 14.2 である。また、モード周波数の対応関係をプロットしたものを Fig. 14.11 に示す。Figure. 14.11 でのプロットは対角線から大きくはずれており、対応するモード周波数の大小関係がフリー状態と複合体状態で保存していないことがわかる。これらの結果は、tRNA<sup>Gln</sup> の低周波モード運動が複合体形成によって大きな影響を受けていることを示している。

これに対して GlnRS は、モード番号 0 ~11 の 12 個のモード運動のうち、モード 6, 10 を除く 10 個の複合体状態でのモードについて、類似度  $S$  が 0.65 を越えるフリー状態の対応モードが存在していた (Table 14.3)。また対応するモード周波数の大小関係も保存していた (Fig. 14.12)。このことから、Fig. 14.9 でみられた GlnRS で支配的なドメイン間運動は、周波数や運動の方向を含めて複合体状態とフリー状態の両方でみられるということがいえる。つまり GlnRS のダイナミクスは tRNA<sup>Gln</sup> との複合体形成において、それほど大きな影響を受けていないということである。

Table 14.2 で挙げたのは、フリー状態と複合体状態に共通に存在するモードである。つまり、フリー状態にも、複合体状態での運動成分が内在している。tRNA-ARS の分子認識に tRNA のダイナミクスが関わっているとすれば、これらのような、フリー状態の tRNA に内在している複合体状態でのモード運動が認識に重要な役割を果たす可能性があると考えられる。Fig. 14.13 および Fig. 14.14 に、Table 14.2 で挙げたモードでの各原子の熱ゆらぎを示す。左の 2 列が、複合体中のモード運動で、右の 2 列がそれに対応するフリー状態のモード運動である。それぞれのモード運動の図では、右図は左図において、紙面に向かって左側から見たものとなっている。対応するモード運動同士は、ゆらぎの大きさに差がある場合もあるが、各原子の運動方向は似通っていることが確認された。

現在用いている類似度の式 (14.3) による定義は、各原子のモード運動を表すベクトルの方向を直接比較するため、モード運動を比較している分子同士の構造が異なっ

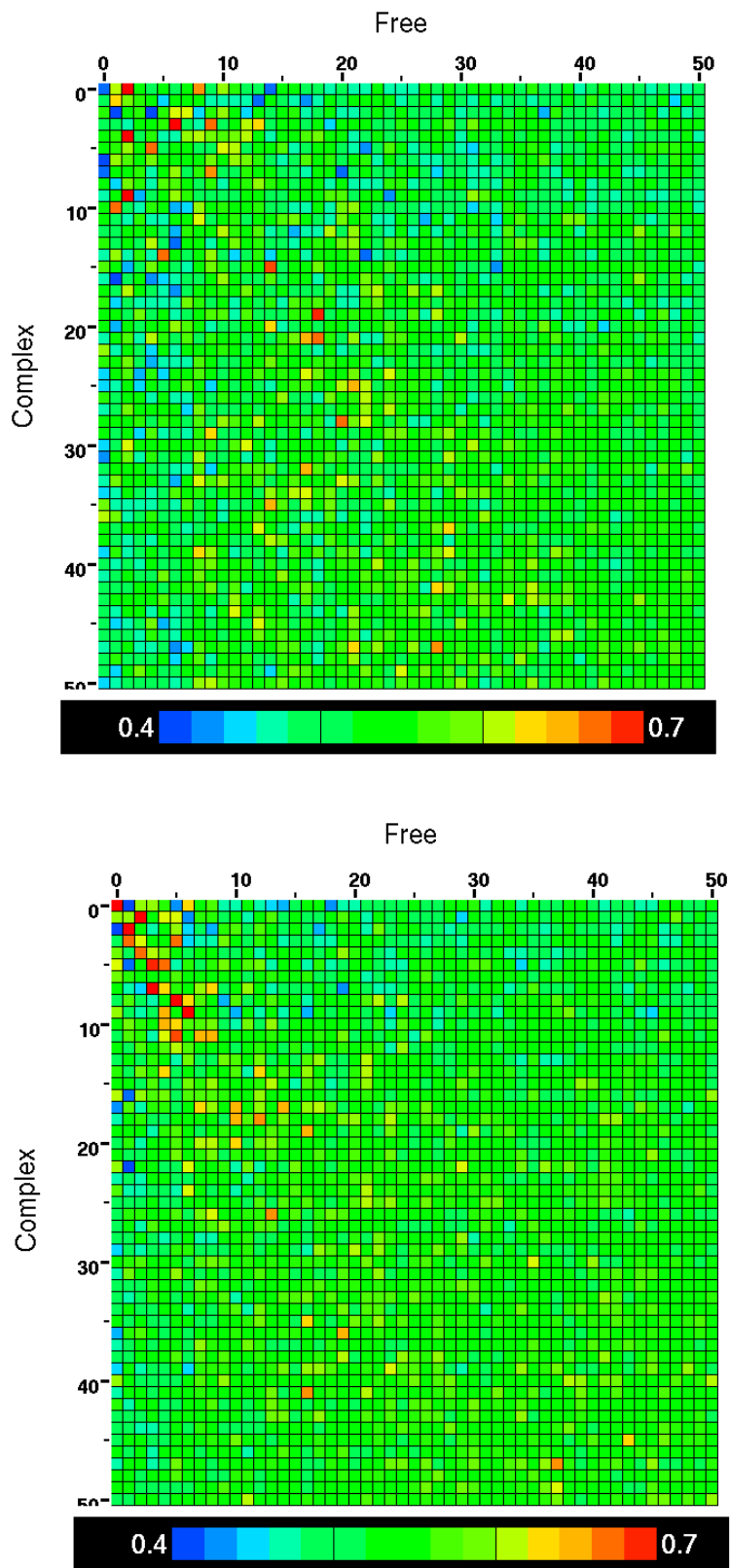


Figure 14.10 複合体状態とフリー状態でのモード運動の相関．縦軸に複合体状態，横軸にフリー状態でのモード番号を示す．モード番号は「0」が最低周波数モードで，周波数の昇順となっている．(上) tRNA<sup>Gln</sup>．(下) GlnRS．

Table 14.2 tRNA<sup>Gln</sup> において、フリー状態にも類似度  $S$  が 0.65 を越えるモードが存在する複合体状態の低周波数領域モードのモード番号．周波数が低い方から 12 個分について示す．

複合体状態	フリー状態
0 ( $1.42\text{cm}^{-1}$ )	2 ( $1.04\text{cm}^{-1}$ ), 8( $2.77\text{cm}^{-1}$ )
3 ( $2.67\text{cm}^{-1}$ )	6 ( $2.12\text{cm}^{-1}$ ), 9( $3.18\text{cm}^{-1}$ )
4 ( $3.32\text{cm}^{-1}$ )	2 ( $1.04\text{cm}^{-1}$ )
5 ( $3.62\text{cm}^{-1}$ )	4 ( $1.51\text{cm}^{-1}$ )
9 ( $4.73\text{cm}^{-1}$ )	2 ( $1.04\text{cm}^{-1}$ )
10 ( $5.61\text{cm}^{-1}$ )	1 ( $0.86\text{cm}^{-1}$ )

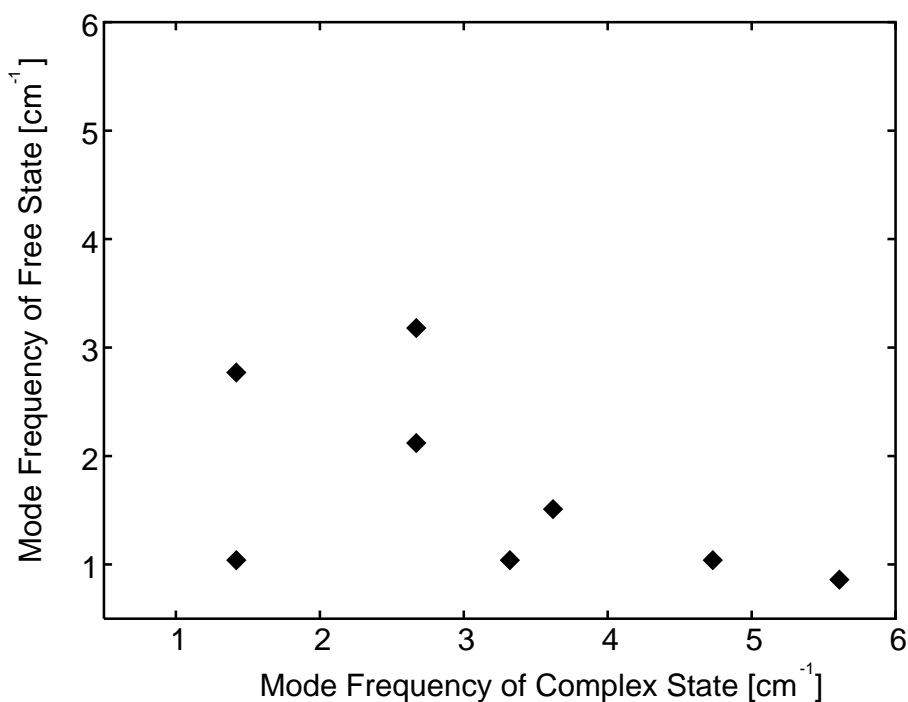


Figure 14.11 tRNA<sup>Gln</sup> において、フリー状態にも類似度  $S$  が 0.65 を越えるモード運動が存在する複合体状態の低周波数領域モードについて、複合体状態とフリー状態の周波数の対応をプロットしたもの．横軸が複合体状態の周波数、縦軸がフリー状態の周波数．周波数が低い方から 12 個分について示す．



Table 14.3 GlnRS において，複合体状態とフリー状態の両方に存在する低周波数領域モードのモード番号．周波数が低い方から 12 個分について示す．

複合体状態	フリー状態
0 (1.42cm <sup>-1</sup> )	0 (1.54cm <sup>-1</sup> )
1 (1.88cm <sup>-1</sup> )	2 (2.08cm <sup>-1</sup> )
2 (2.47cm <sup>-1</sup> )	1 (1.54cm <sup>-1</sup> )
3 (2.67cm <sup>-1</sup> )	1 (1.54cm <sup>-1</sup> ), 5 (3.89cm <sup>-1</sup> )
4 (3.32cm <sup>-1</sup> )	2 (2.08cm <sup>-1</sup> )
5 (3.62cm <sup>-1</sup> )	3 (3.20cm <sup>-1</sup> ), 4 (3.51cm <sup>-1</sup> )
7 (4.33cm <sup>-1</sup> )	3 (3.20cm <sup>-1</sup> )
8 (4.59cm <sup>-1</sup> )	5 (3.89cm <sup>-1</sup> )
9 (4.73cm <sup>-1</sup> )	6 (4.42cm <sup>-1</sup> )
11 (6.00cm <sup>-1</sup> )	5 (3.89cm <sup>-1</sup> )

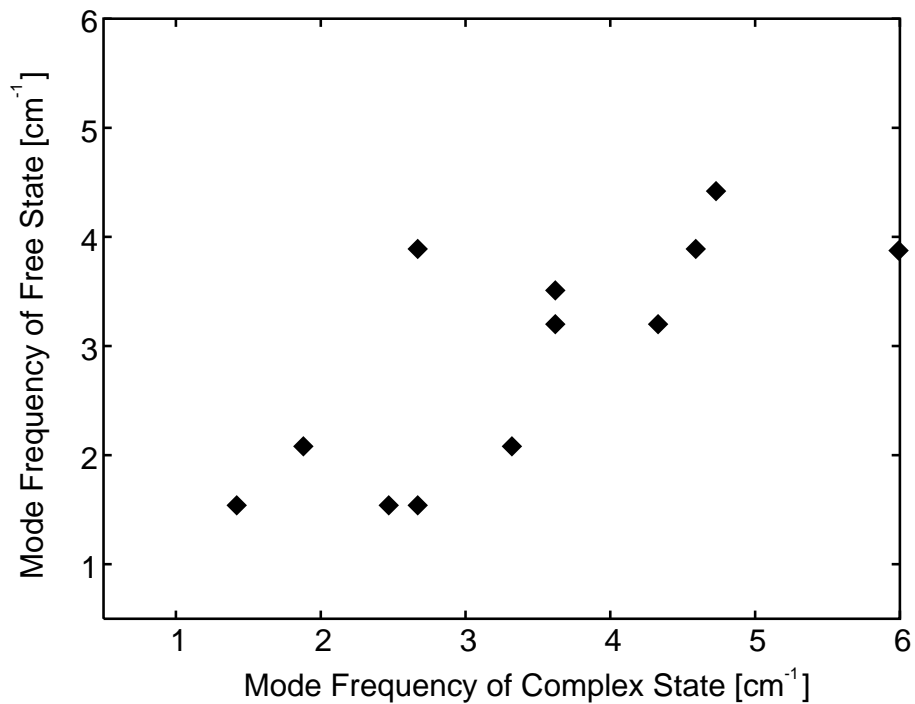


Figure 14.12 GlnRS において，フリー状態にも類似度  $S$  が 0.65 を越えるモード運動が存在する複合体状態の低周波数領域モードについて，複合体状態とフリー状態の周波数の対応をプロットしたもの．横軸が複合体状態の周波数，縦軸がフリー状態の周波数．周波数が低い方から 12 個分について示す．

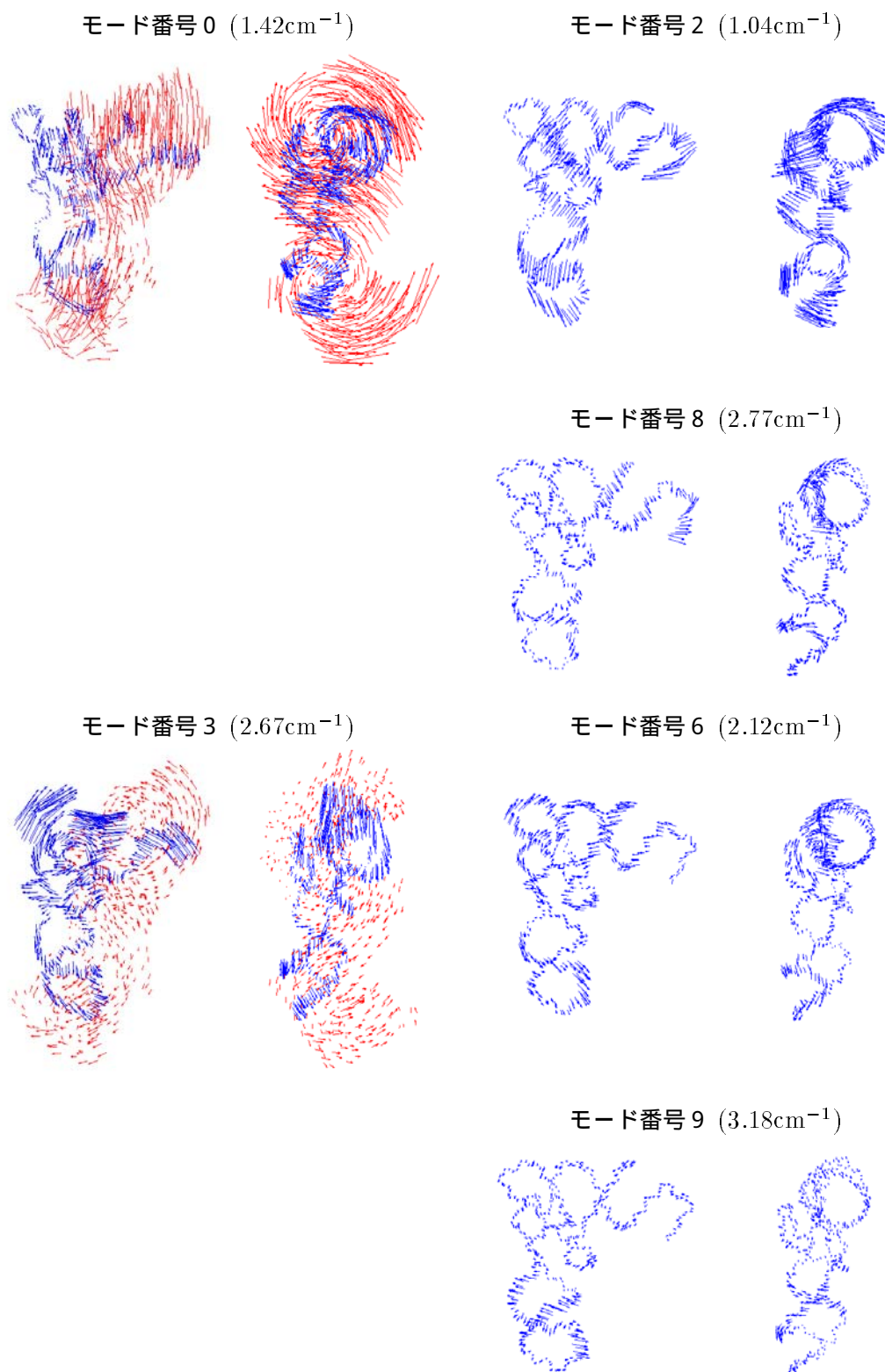


Figure 14.13  $\text{tRNA}^{\text{Gln}}$  において、フリー状態にも類似度  $S$  が 0.65 を越えるモードが存在する複合体状態の低周波数モード運動 (モード番号: 0, 3). それぞれにおいて、右図は左図において紙面に向かって左側から見たもの. 左の列は複合体状態のモード運動 (内部運動のみ), 右の列はフリー状態のモード運動. 複合体状態では、青が  $\text{tRNA}^{\text{Gln}}$ , 赤が  $\text{GlnRS}$  の各原子の運動である. 見やすいように、矢印の大きさは、複合体状態ではすべて 30 倍、フリー状態ではすべて 7 倍にしてある.

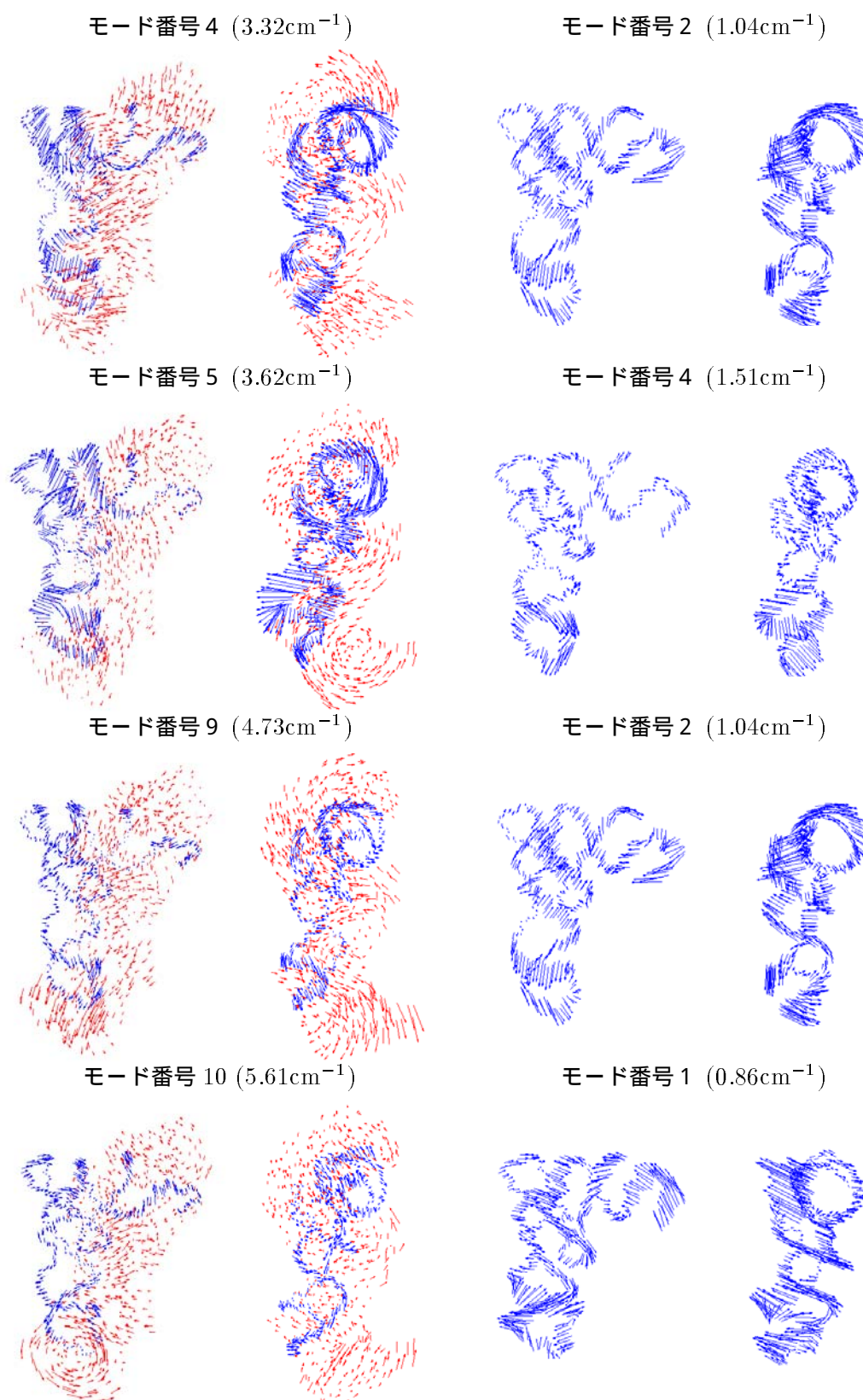


Figure 14.14 tRNA<sup>Gln</sup> において，フリー状態にも類似度  $S$  が 0.65 を越えるモードが存在する複合体状態の低周波数モード運動 (モード番号: 4, 5, 9, 10)．それぞれにおいて，右図は左図において紙面に向かって左側から見たもの．左の列は複合体状態のモード運動 (内部運動のみ)，右の列はフリー状態のモード運動．複合体状態では，青が tRNA<sup>Gln</sup>，赤が GlnRS の各原子の運動である．見やすいように，矢印の大きさは，複合体状態ではすべて 30 倍，フリー状態ではすべて 7 倍にしてある．

ていると、分子にとって本質的には同じ運動であっても、ベクトルの向きが異なってしまうことで類似度が低く算出されてしまうということがある。これに対応するために、tRNA の二重らせん構造からの相対的な向きや位置を基準としてベクトルの方向を評価する方法を開発することが今後検討が必要な課題である。

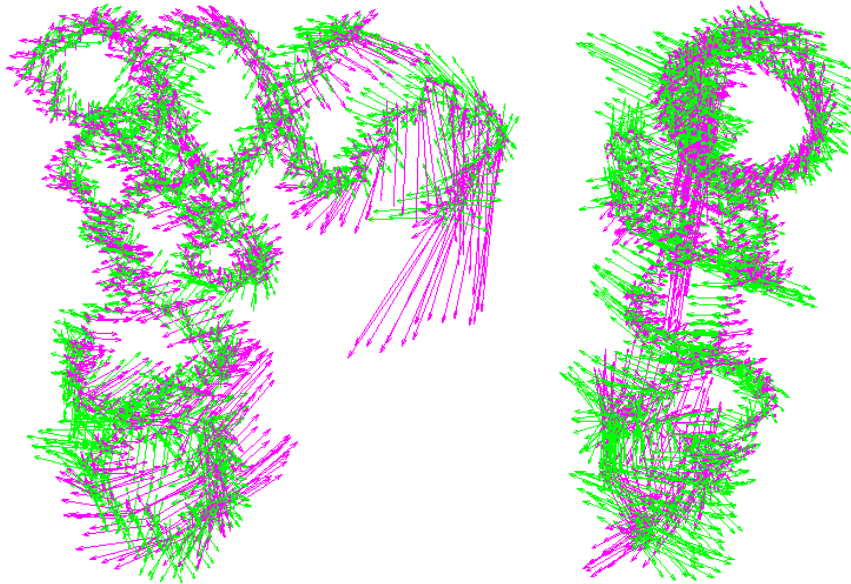


Figure 14.15 モード番号 10 までの保存・非保存モードの重ね合わせ．保存モード（モード番号 1, 2, 4, 6, 8, 9）を緑，非保存モード（モード番号 0, 3, 5, 7）をマゼンタで表す．矢印の大きさはすべて 7 倍にしてある．

フリー状態の tRNA<sup>Gln</sup> において、低周波数領域の 10 個のモードを複合体中でも保存しているモード運動と保存していないモードに色分けして重ね合わせたものを Fig. 14.15 に示す．保存しているモード（モード番号 1, 2, 4, 6, 8, 9）を緑で、非保存モード（モード番号 0, 3, 5, 7）をマゼンタで表す．矢印の大きさはすべて 7 倍にしてある．

L 字型が開いたり閉じたりする方向のゆらぎ成分が抑えられ、アクセプタステム、アンチコドンアームがそれとは直交する方向に swing するゆらぎ成分が保存していることがわかる．他のアミノ酸系についてもフリー状態および複合体状態の基準振動モード解析を行い、両者の対応関係に Gln 系と同様の特徴がみられるかどうかを検証することで、Gln 系で得られた知見が tRNA-ARS 複合体において普遍的かどうかを調べることができる．また第 11 章で述べた Bullock らの aptamer など、ARS との結合能やアミノアシル化効率に大きな差がある置換体のモード運動を、wild type のそれと比較することで、分子認識に関わる重要な運動成分を抽出できるのではないかと考える．

また Sarai らのグループは、DNA 結合タンパク質において、ワトソクリック結合している塩基対周辺に各アミノ酸が存在するときの自由エネルギーマップを、構造データベース検索から、あるいは Multicanonical Monte Carlo 法などの自由エネルギー計算によって網羅的に求め、複合体中の塩基とアミノ酸の相互作用などを詳細に解析・評価することができる手法を開発している<sup>163-165)</sup>。tRNA-ARS 複合体にも同様の方法を適用することで、似た立体構造をとる tRNA の塩基配列の違いが、複合体形成自由エネルギーに与える影響を定量的に評価することができると考えられる。本研究で用いた手法で tRNA 同士のダイナミクスの違いを解析できれば、Sarai らの手法による結果と併せることで、tRNA-ARS 複合体形成のアフィニティを静的・動的の両面から解析することができると考えられる。

## 第 15 章

### 第 III 部のまとめ

第 III 部では、まずフリー状態の tRNA<sup>Phe</sup> および tRNA<sup>Asp</sup> に基準振動モード解析を適用し、低周波数領域の分子全体にわたるモード運動の解析を行った。基準振動モード解析によって得られた各原子の熱ゆらぎの大きさは、X 線結晶解析の温度因子データによる熱ゆらぎの大きさと高い相関があり、本研究でのシミュレーションの妥当性を示していた。最低モード周波数は、球状タンパク質の  $1/4 \sim 1/10$  と小さく、tRNA の特異的な L 字型構造がもつゆっくりとした大きな特徴的なダイナミクスが明らかになった。低周波数領域では、tRNA 分子のアクセプタステム、アンチコドンアームなどの領域が一体となって、大きな運動をしていた。一体となって運動する領域は tRNA<sup>Phe</sup> と tRNA<sup>Asp</sup> で異なっており、似たような L 字型立体構造をとる両分子でもダイナミクスに微妙な差異があることが明らかになった。

また、*E. coli* 由来の tRNA<sup>Gln</sup>-GlnRS 複合体について、フリー状態のモード運動と複合体状態のモード運動の詳細な解析を試みた。基準振動モード解析による各原子の熱ゆらぎの大きさは、X 線結晶解析の温度因子データと非常に高い相関を示した。複合体のモード周波数分布は、フリー状態の GlnRS の周波数分布に似ているが、最低周波数は、若干低周波数寄りにシフトした。各原子の熱ゆらぎの大きさは、フリー状態に比べて  $1/2 \sim 1/4$  と小さくなった。tRNA<sup>Gln</sup> では GlnRS との接触面でゆらぎが大きく抑えられている一方で、接触面から遠い側の D アーム、T アーム、アンチコドンアームなどでゆらぎが大きい部分がみられた。複合体形成による熱ゆらぎの大きさの変化は、tRNA<sup>Gln</sup> と GlnRS が接触している広い範囲でゆらぎの大きな減少がみられた。これに対して、tRNA<sup>Gln</sup> の T アーム、D ループ、エキストラループ、GlnRS のアンチコドン認識ドメインなどでは、ゆらぎの減少の割合は比較的小さかった。

低周波数領域のモード運動では、tRNA<sup>Gln</sup> と GlnRS が大きくカップリングして運動していることが明らかになった。tRNA<sup>Gln</sup> のアクセプタステム、アンチコドンアーム

ム，中央部分の各領域と，GlnRS の触媒ドメイン，アンチコドン認識ドメインなどの各ドメインが，モードによって一体となって協調的に運動していた．これらのモード運動のうち，各分子の相対運動（外部運動）は，各分子の内部変形に伴う運動（内部運動）の  $1/2 \sim 1/4$  程度で，これは従来報告されていた球状タンパク質同士の低周波数領域の，外部運動を中心とするモード運動とは大きく異なる特徴的なものであった．外部運動と内部運動は，接触面付近で絶対値の大きな負の相関をもち，外部運動と内部運動が打ち消しあった結果として，接触面でのゆらぎが小さくなっていることが明らかになった．また低周波数領域の主要な運動は，GlnRS はドメイン同士の相対運動であり， $\text{tRNA}^{\text{Gln}}$  は二重らせん部分のゆっくりとした変形であるという，タンパク質と核酸のダイナミクスの特徴の違いもみられた．

複合体状態とフリー状態でのモード運動を相互に比較して，その類似度を解析した結果，GlnRS は複合体形成によっても低周波数領域でのモード運動の保存性が高く，複合体形成によって GlnRS のダイナミクスは大きな影響を受けていないことがわかった． $\text{tRNA}^{\text{Gln}}$  では，複合体状態とフリー状態の対応モードの周波数の大小関係は保存しておらず， $\text{tRNA}^{\text{Gln}}$  のダイナミクスは複合体形成によって大きな影響を受けているといえる．一方で，フリー状態のダイナミクスに，いくつかの複合体状態のモード運動が内在していることも明らかになった． $\text{tRNA-ARS}$  の分子認識に  $\text{tRNA}$  のダイナミクスが関わっているとすれば，これらのモード運動は認識に重要な役割を果たす可能性がある． $\text{tRNA-ARS}$  の分子認識へのダイナミクスの関与が明らかになれば，RNA-タンパク質複合体において，接触部位ではなく RNA 自身の安定性の違いが RNA-タンパク質の結合に影響を及ぼす例としても重要な意味をもつ．

また，本研究で開発した，基準振動モード運動の比較法は， $\text{tRNA-ARS}$  以外の核酸-タンパク質，タンパク質-タンパク質複合体などでも応用が可能な汎用的なものであり，幅広い応用が期待される．

今後，さらに Gln 以外のアミノ酸系について  $\text{tRNA-ARS}$  複合体のダイナミクスを計算し，相互に結果を比較することで， $\text{tRNA-ARS}$  の分子認識のアミノ酸種特異性について，これまでの静的な観点からの研究結果に加えて，ダイナミクスの視点に立った新たな知見が得られると思われる．

そのためには，並列計算システムの，エネルギー極小化，Hessian 計算，基準振動モード計算などの計算能力を高めて，さらに大きな複合体について，多数のモードの網羅的解析を行えるようにすることが必要である．また，基準振動モード解析法に加えて，非調和成分も考慮することができる分子動力学シミュレーションによる解析も行うことが必要である．





おわりに



## 第 16 章

### おわりに

本論文では、核酸および核酸タンパク質複合体を対象に、コンピュータ・シミュレーションによって、構造モデリング、自由エネルギー解析、ダイナミクス解析を行った研究について述べた。

シミュレーション要素技術の開発として、生体分子の構造モデリング、Poisson-Boltzmann 分子動力学法による構造サンプリングと自由エネルギー解析、二面角基準振動モード解析法によるダイナミクス解析などを、並列計算技術を用いて統合的に、かつ大規模分子に対しても高速に行うことができるシステムを構築した。

Hessian の並列計算アルゴリズムでは、遺伝的アルゴリズムを用いた通信コストとプロセッサの負荷分散の最適化によって、高い並列化効率を実現できた。シミュレーションの要素技術については、今後ますます「並列」がキーワードとなるであろう。シングルプロセッサの能力向上は、光や電子の速度が律速となって、ここ 10 年以内に頭打ちになるといわれている。その一方で、並列技術とそれを支える通信技術はさらに大きく発達すると思われる。大規模な並列化の効果を最大限に取り込めるような並列計算アルゴリズムを開発してゆくことが必要である。

DNA ミニヘアピン分子については、熱安定な配列とそうでない配列について、構造モデリングと自由エネルギー解析で両者の違いを解析した。モデリングした構造に対して、Poisson-Boltzmann 分子動力学法と WHAM 法を適用することで、自由エネルギープロファイルを直接比較して、両者を区別できることを示した。構造モデリングについては、核酸の構造データが今後増えてくれば、エネルギー最小化による方法とデータベースから配列 - 構造相関情報統計ポテンシャルとして抽出してそれを用いる方法を組み合わせ、精度の高い構造モデリングを行うことができるようになると思われる。核酸の場合にとくに計算量を減らせるという二面角系の特徴を活かしつつ、より現実に近いモデルで溶媒効果を取り込む方法を取り入れていくことが今後の課題である。また、多数の構造をサンプリングして反応座標に対する自

由エネルギープロファイルを求める方法は、現在次々と新しいアルゴリズムが開発されている。これらを適用することで、より網羅的に核酸分子の熱安定性とその配列依存性の解析を進めることができると思われる。

tRNA-ARS 複合体については、両分子のフリー状態、複合体状態のダイナミクスを解析した。フリー状態の tRNA では、低周波数領域のモード運動の様子とそのアミノ酸種による違いを明らかにした。複合体状態では、複合体中での tRNA・ARS の密にカップリングした特徴的な運動と、フリー状態のモード運動に内在する複合体状態でのモード運動、複合体形成によっても ARS のダイナミクスが大きな影響を受けていない一方で、tRNA には複合体形成によるダイナミクスの変化がみられることなどを示した。X 線結晶解析技術の進歩で、今後さらに多くの生物種・アミノ酸系で複合体の立体構造が解かれるであろう。アミノ酸系の違いによるダイナミクスの違いを詳細に解析することで、tRNA-ARS の分子認識メカニズムにダイナミクスの観点からアプローチできる期待がある。これらの複合体の解析にはさらに大規模な計算コストが必要となるので、そのような系も扱えるようにするためのシステムの改良、基準振動モード解析では追えないような大きな構造変化を伴うダイナミクスを考慮する工夫などを取り入れることを考えている。

生体分子とくに核酸のコンピュータ・シミュレーションの重要性はこれからますます高まるとと思われる。本研究で開発した計算技術は汎用的な応用が可能であり、本研究で扱った DNA ミニヘアピン分子や tRNA-ARS 複合体に限らず、一般の核酸、核酸タンパク質複合体への適用が可能であると思われる。本研究の手法とその応用結果は、今後の核酸・核酸タンパク質複合体の研究において少なからず貢献することが期待される。

## 謝辞

本研究全般にわたって御指導頂いた東京大学大学院農学生命科学研究科 応用生命工学専攻の清水謙多郎 教授に心から感謝致します。

研究を進めるにあたり，数多くの助言を与えて下さった池口満徳 元助手・現横浜市立大学助教授に心から感謝致します。

学生時代から御指導頂いた土井淳多 東大名誉教授・現千葉工業大学教授に感謝致します。

DNA ミニヘアピン分子の立体構造座標を提供していただき，御議論頂いた理化学研究所の平尾一郎博士，千葉工業大学の河合剛太助教授，スタンフォード大学の吉澤聡子博士に感謝致します。

毎日の研究生生活の中で，いつもあたかかくはげましてくださった生物情報工学研究室の皆さん，卒業生の皆さんに感謝致します。

最後に，日々の疲れを癒し，心をなごませてくれた妻 美弥と長女 瑠々花に感謝致します。



# Appendix

## A Two-Stage 法の理論

### A.1 ポテンシャルのガウス関数での近似

Two-Stage 法<sup>21)</sup> では、拡散方程式 (8.2) の解析解を求められるようにするために、ポテンシャルをガウス関数の和で近似する。対象とするのは、van der Waals ポテンシャル、静電ポテンシャル、水素結合ポテンシャルである。二面角ポテンシャルは、拡散方程式の解を求めるのが困難であるので、計算には含めない。

原子  $i, j$  の間にはたらく van der Waals ポテンシャル

$$E_{vdw}^{ij} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \quad (\text{A.1})$$

水素結合ポテンシャル

$$E_{hb}^{ij} = \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \quad (\text{A.2})$$

および、静電ポテンシャル

$$E_{el}^{ij} = \frac{q_i q_j}{\epsilon r_{ij}} \quad (\text{A.3})$$

をガウス関数で近似する。このうち、van der Waals ポテンシャルと水素結合ポテンシャルは、以下のように近似する。

$$E_{vdw,hb}^{ij}(r_{ij}) = \sum_{k=1}^L a_{ij}^k G_{\gamma_{ij}, b_{ij}^k}(r_{ij}) \quad (\text{A.4})$$

ただし

$$G_{\gamma,b}(r) = \begin{cases} \exp(-br^2), & \text{for } r > \gamma \\ \exp(-b\gamma^2), & \text{for } r \leq \gamma \end{cases} \quad (\text{A.5})$$

ここで、 $L$  は近似に用いるガンマ関数  $G$  の数で、本計算では  $L = 3$  としている。 $a_{ij}^k$  および  $b_{ij}^k$  は、ポテンシャルにガウス関数をフィッティングすることにより求める。 $\gamma_{ij}$  は  $\gamma_{ij} = \alpha d_{ij}$  より得られる。 $d_{ij}$  は2原子間のエネルギーがゼロになるときの距離で、 $\alpha$  は van der Waals ポテンシャルのときは 0.75、水素結合ポテンシャル

のときは0.77であるとする．また，静電ポテンシャルについては，誘電率  $\varepsilon$  を定数とするときは，

$$E_{el}^{ij}(r_{ij}) = q_i q_j H_{\gamma_{ij}}(r_{ij}) \quad (\text{A.6})$$

ただし，

$$H_{\gamma}(r) = \begin{cases} 1/r, & \text{for } r > \gamma \\ 1/\gamma, & \text{for } r \leq \gamma \end{cases} \quad (\text{A.7})$$

とする． $\varepsilon = 4r_{ij}$  とするときには，関数  $1/4r^2$  を  $0.4 \leq r \leq 60$  の範囲で10個のガウス関数でフィッティングすることで，以下のようにする．

$$E_{ij}^{el}(r_{ij}) = q_i q_j \sum_{k=1}^{10} a_k \exp(-b_k r_{ij}^2) \quad (\text{A.8})$$

上の各式において  $r_{ij}$  が零の付近において関数を定数にしているのは， $r_{ij} = 0$  付近で関数が発散することを避けるためである．

## A.2 ガウス関数に対する拡散方程式の解

van der Waals ポテンシャルが

$$g(\mathbf{r}_1, \dots, \mathbf{r}_n) = \sum_{i < j} \sum_{k=1}^L a_{ij}^k G_{\gamma_{ij}, b_{ij}^k}(\|\mathbf{r}_i - \mathbf{r}_j\|) \quad (\text{A.9})$$

であるとき，境界条件が，

$$g(\mathbf{r}_1, \dots, \mathbf{r}_n, 0) = g(\mathbf{r}_1, \dots, \mathbf{r}_n) \quad (\text{A.10})$$

である関数  $g$  の拡散方程式

$$\sum_{k=1}^3 \sum_{i=1}^n \frac{\partial^2}{(\partial \mathbf{r}_i^k)^2} g(\mathbf{r}_1, \dots, \mathbf{r}_n, t) = \frac{\partial}{\partial t} g(\mathbf{r}_1, \dots, \mathbf{r}_n, t) \quad (\text{A.11})$$

を満たす解は，

$$g(\mathbf{r}_1, \dots, \mathbf{r}_n, t) = \sum_{i < j} \sum_{k=1}^L a_{ij}^k \Gamma_{\gamma_{ij}, b_{ij}^k}(\|\mathbf{r}_i - \mathbf{r}_j\|, 2t) \quad (\text{A.12})$$

となる．ただし，

$$\begin{aligned} \Gamma_{\gamma, b}(r, t) = & \frac{e^{br^2/(1+4bt)}}{2[(1+4bt)^{1/2}]^3} \left[ \operatorname{erfc} \left( \frac{\gamma(1+4bt)-r}{2[t(1+4bt)]^{1/2}} \right) + \operatorname{erfc} \left( \frac{\gamma(1+4bt)+r}{2[t(1+4bt)]^{1/2}} \right) \right] + \\ & \left( \frac{t}{\pi} \right)^{1/2} \frac{1}{r} \frac{(-4bt)e^{-b\gamma^2}}{1+4bt} \left[ e^{-(\gamma-r)^2/4t} - e^{-(\gamma+r)^2/4t} \right] + \\ & \frac{1}{2} e^{-b\gamma^2} \left[ \operatorname{erf} \left( \frac{\gamma-r}{2t^{1/2}} \right) + \operatorname{erf} \left( \frac{\gamma+r}{2t^{1/2}} \right) \right] \end{aligned} \quad (\text{A.13})$$



である． erf は error function と呼ばれる関数で，次式で定義される．

$$\operatorname{erf}(x) \equiv (2/\pi^{1/2}) \times \int_0^x e^{-u^2} \quad (\text{A.14})$$

また，erfc は  $1 - \operatorname{erf}$  である．

静電ポテンシャルでは， $\varepsilon$  が定数の場合は，以下のようになる．

$$h(\mathbf{r}_1, \dots, \mathbf{r}_n) = \sum_{i < j} q_i q_j H_{\gamma_{ij}}(\|\mathbf{r}_i - \mathbf{r}_j\|) \quad (\text{A.15})$$

であるとき，境界条件が，

$$h(\mathbf{r}_1, \dots, \mathbf{r}_n, 0) = h(\mathbf{r}_1, \dots, \mathbf{r}_n) \quad (\text{A.16})$$

である関数  $h$  の拡散方程式

$$\sum_{k=1}^3 \sum_{i=1}^n \frac{\partial^2}{(\partial \mathbf{r}_i^k)^2} h(\mathbf{r}_1, \dots, \mathbf{r}_n, t) = \frac{\partial}{\partial t} h(\mathbf{r}_1, \dots, \mathbf{r}_n, t) \quad (\text{A.17})$$

を満たす解は，

$$h(\mathbf{r}_1, \dots, \mathbf{r}_n, t) = \sum_{i < j} q_i q_j \Theta_{\gamma_{ij}}(\|\mathbf{r}_i - \mathbf{r}_j\|, 2t) \quad (\text{A.18})$$

となる．ただし，

$$\begin{aligned} \Theta_{\gamma}(r, t) = & \frac{1}{2} \left[ \left( \frac{1}{\gamma} + \frac{1}{r} \right) \operatorname{erf} \left( \frac{\gamma+r}{2t^{1/2}} \right) + \left( \frac{1}{\gamma} - \frac{1}{r} \right) \operatorname{erf} \left( \frac{\gamma-r}{2t^{1/2}} \right) \right] - \\ & \left( \frac{t}{\pi} \right)^{1/2} \frac{1}{\gamma r} \left[ e^{-(\gamma-r)^2/4t} - e^{-(\gamma+r)^2/4t} \right] + \\ & \frac{1}{2} e^{-b\gamma^2} \left[ \operatorname{erf} \left( \frac{\gamma-r}{2t^{1/2}} \right) + \operatorname{erf} \left( \frac{\gamma+r}{2t^{1/2}} \right) \right] \end{aligned} \quad (\text{A.19})$$

である．また  $\varepsilon = 4r_{ij}$  の場合は，拡散方程式を満たす解は以下のようになる．

$$U_{ij}^{el}(r_{ij}, t) = q_i q_j \sum_{k=1}^{10} a_k (1 + 8b_k t)^{-3/2} \exp \left( \frac{-b_k r_{ij}^2}{1 + 8b_k t} \right) \quad (\text{A.20})$$

### A.3 Time-reversing procedure

Time-reversing procedure での時刻のスケジューリングは，以下の式で行う．

$$t_q = t_{max} \left( 1 - \frac{q}{Q} \right)^m \quad (x = 0, 1, \dots, Q) \quad (\text{A.21})$$

ただし， $t_q$  は現在の時刻， $Q$  は総ステップ数である．計算は  $m = 2$  として行う．

また，時刻の初期値  $t_{max}$  の値は，以下の式で見積もることができる．

$$t_{max} > \frac{1}{6} d^2 m(m+1)(2m+1) \quad (\text{A.22})$$

ただし  $d$  は第 5.1 節における  $k_{max}$ ， $m+2$  はボンド数である．

Two-Stage 法での time-reversing procedure の手順は，以下のようになる．

1. 時刻  $t$  を十分に大きな値 ( $t = t_{max}$ ) とする。「現在の点」(初期値) はランダムに選んでよい。
2. 現在の点からはじめて、最急降下法とニュートン・ラフソン法の組み合わせにより、Hessian の固有値がすべて正 (正定値) となるまで  $F(\boldsymbol{x}, t)$  に対してエネルギー極小化を行う。
3. もし時刻  $t$  があらかじめ設定した値  $t_{local}$  よりも小さい場合は、Hessian の固有値と固有ベクトルを求める。そうでない場合は、手順 5 へ進む。
4. Hessian の固有値を大きい方から  $N_v$  個選び出し、対応する固有ベクトル方向にそれぞれに  $N_e$  個の点を探索する。両隣よりもエネルギー値の低い「谷」の点があれば、そこから手順 2 と同様にエネルギー極小化を行う。到達した極小点のエネルギー値が現在のエネルギー値よりも小さい場合は、この極小点を現在の点とする。
5.  $t = 0$  ならば終了する。そうでなければ  $t$  を  $\delta t$  だけ減らして、手順 2 に戻る。
6. 最終的に求まる  $F(\boldsymbol{x}, 0) = f(\boldsymbol{x})$  での極小点が、エネルギー最小点 (グローバルミニマム) になるとする。

## B Poisson-Boltzmann 分子動力学法の理論

### B.1 Poisson-Boltzmann 方程式

Debye-Hückel 理論<sup>166)</sup> は、水溶液中で完全にイオンに解離する強電解質の溶液系におけるイオン - 水間の相互作用を与える理論である。溶液を誘電率  $\epsilon$  の連続体とする。溶液中のあるイオン  $i$  に着目すると、そのイオンのまわりの電場  $\phi_i$  の中に置かれた粒子  $j$  のポテンシャルエネルギーは、粒子の電荷を  $z_j e$  とすると、 $z_j e \phi_i$  である。ただし  $z$  はイオン価数、 $e$  は電子電荷である。Boltzmann 分布則を用いると、イオン  $j$  の分布数密度  $n_j$  は以下の式で与えられる。

$$n_j = n_j^0 \exp(-\beta z_j e \phi_i) \quad (\text{B.23})$$

ただし、

$$\beta = \frac{1}{k_B T} \quad (\text{B.24})$$

である。 $n_j^0$  は単位体積中のイオン  $j$  の平均数であり、イオン  $j$  のモル濃度を  $C_j$ 、アボガドロ数を  $N$  とすると、 $NC_j/1000$  で与えられる。複数のイオンを含んだ系全体

の静電場を求めるための方程式は，以下の Poisson-Boltzmann 方程式

$$\nabla \cdot \varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) = - \left( \rho(\mathbf{r}) + \sum_j z_j e n_j^0 \exp(-\beta z_j e \phi(\mathbf{r})) \right) \quad (\text{B.25})$$

で与えられる．希薄溶液では，右辺の指数関数を  $\beta$  について展開し，1 次の項までで近似することができる．これと溶液が全体として中性であるという条件から，

$$\nabla \cdot \varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) = -\rho(\mathbf{r}) + \kappa(\mathbf{r})^2 \varepsilon(\mathbf{r}) \phi(\mathbf{r}) \quad (\text{B.26})$$

が導かれる．ただし，

$$\kappa(\mathbf{r})^2 = \frac{\beta}{\varepsilon(\mathbf{r})} \sum_j z_j^2 e^2 n_j^0 = \frac{2\beta e^2 N}{1000\varepsilon(\mathbf{r})} I \quad (\text{B.27})$$

である． $I$  はイオン強度， $\kappa$  は Debye-Hückel のパラメータと呼ばれる，長さの逆数のディメンジョンをもつ量である．定性的には， $1/\kappa$  はイオンがまわりに形成するイオン雲の大きさと考えられる．

この式 (B.26) を用いて，連続体近似した溶媒が作り出す静電場を求めることができる．

## B.2 有限差分 Poisson-Boltzmann 法

系の空間を格子に分割することで，Poisson-Boltzmann 方程式は以下のように有限差分法で解くことができる．Poisson-Boltzmann 方程式に体積積分を実行すると，以下のように変形できる．

$$\int \nabla \cdot \varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) dV = - \int \rho(\mathbf{r}) dV + \int \varepsilon(\mathbf{r}) \kappa(\mathbf{r})^2 \phi(\mathbf{r}) dV \quad (\text{B.28})$$

ガウスの定理を用いると，

$$\int_{\Omega} \mathbf{n} \cdot \{ \varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) \} dS = -Q + \int \varepsilon(\mathbf{r}) \kappa(\mathbf{r})^2 \phi(\mathbf{r}) dV \quad (\text{B.29})$$

ただし， $Q$  は体積積分を行った領域内部に含まれる全電荷， $\Omega$  は体積積分を行った領域の表面， $\mathbf{n}$  はその表面外側を向く単位ベクトルである．体積積分を点  $(x, y, z)$  を囲む立方体とすると，式 (B.29) の左辺は離散空間では以下のように与えられる．

$$h \left( \sum_{i,j,k} \varepsilon(x+i/2, y+j/2, z+k/2) \phi(x+i, y+j, z+k) \right) - \phi(x, y, z) \sum_{i,j,k} \varepsilon(x+i/2, y+j/2, z+k/2) \quad (\text{B.30})$$

ただし  $\sum_{i,j,k}$  は， $i = \pm h$  のときには  $j, k = 0$  などとなり，6 つの項の和を表す．

ここで，点  $(x, y, z)$  における電荷密度を  $\rho_0$ ，遮蔽定数を  $\kappa_0$ ，静電ポテンシャルを  $\phi_0$  とおき， $\phi_{ijk}$  を点  $(x, y, z)$  から  $(i, j, k)$  だけずらした位置の静電ポテンシャル，また  $\varepsilon_{ijk}$  を点  $(x, y, z)$  から  $(i/2, j/2, k/2)$  だけずらした位置の誘電率の分布であると定義する．また， $\kappa_0$  が非ゼロ，すなわちイオンが存在できるのは溶媒領域であるため，右辺の  $\varepsilon$  は  $\varepsilon_s$  と考えてよい．このとき式 (B.28) の点  $(x, y, z)$  のまわりの離散表現は次式のように表せる．

$$h \left( \sum_{i,j,k=\pm h} \varepsilon_{ijk} \phi_{ijk} - \phi_0 \sum_{i,j,k=\pm h} \varepsilon_{ijk} \right) = -h^3 \rho_0 + h^3 \kappa_0^2 \varepsilon_s \phi_0 \quad (\text{B.31})$$

この式を  $\phi_0$  について解くことで次式を得る．

$$\phi_0 = \frac{\sum_{i,j,k=\pm h} \varepsilon_{ijk} \phi_{ijk} + h^2 \rho_0}{\sum_{i,j,k=\pm h} \varepsilon_{ijk} + h^2 \kappa_0^2 \varepsilon_s} \quad (\text{B.32})$$

式 (B.32) を格子に切った空間全てについて立式すると最終的には連立一次方程式を解くという問題に帰着される．

有限差分 Poisson-Boltzmann 法では，空間を格子に切るため，格子要素に原子電荷を割り振ることが必要である．各頂点には， $X, Y, Z$  軸それぞれについて向かい合う頂点の電荷との比が，単位格子中の点電荷の座標と等しくなるように配分する (Fig. B.1)．つまり，以下の式で表される関係を満たすように  $q_1 \sim q_8$  を決定する (Rogers の方法<sup>167</sup>) ．

$$\sum_{i=1}^8 q_i = q \quad (\text{B.33})$$

$$q_1 : q_2 = q_4 : q_3 = q_5 : q_6 = q_8 : q_7 = (1 - a) : a \quad (\text{B.34})$$

$$q_1 : q_4 = q_2 : q_3 = q_5 : q_8 = q_6 : q_7 = (1 - b) : b \quad (\text{B.35})$$

$$q_1 : q_5 = q_2 : q_6 = q_3 : q_7 = q_4 : q_8 = (1 - c) : c \quad (\text{B.36})$$

これは以下のように解ける．

$$q_1 = (1 - a)(1 - b)(1 - c)q \quad (\text{B.37})$$

$$q_2 = a(1 - b)(1 - c)q \quad (\text{B.38})$$

$$q_3 = (1 - a)b(1 - c)q \quad (\text{B.39})$$

$$q_4 = ab(1 - c)q \quad (\text{B.40})$$

$$q_5 = (1 - a)(1 - b)cq \quad (\text{B.41})$$

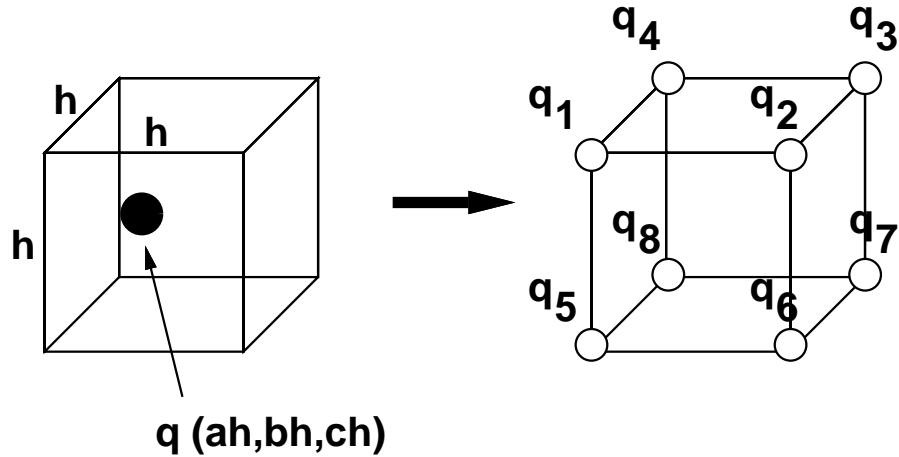


Figure B.1 格子頂点への電荷のマッピングを示す概念図．

$$q_6 = a(1-b)cq \quad (\text{B.42})$$

$$q_7 = (1-a)bcq \quad (\text{B.43})$$

$$q_8 = abcq \quad (\text{B.44})$$

このようにすることで，単位格子中の総電荷量と電気双極子モーメントが保存される．

一方，誘電率は Malcolm の方法<sup>168)</sup> にしたがって，単位格子の表面上に以下の式で与える (Fig. B.2) ．

$$\varepsilon = \frac{\varepsilon_A \varepsilon_B}{(1.5-a)\varepsilon_A + (a-0.5)\varepsilon_B} \quad (\text{B.45})$$

式 (B.32) を，境界条件を与えて解く．境界条件には，自己無撞着境界法 (Self-Consistent Boundary Method)<sup>169)</sup> を用いる．溶質を中に含むように任意に選んだ境界面  $\Gamma$  と，無限に離れた面とで挟まれてできる空間を考える．この領域における静電ポテンシャルは，

$$\Delta \phi(\mathbf{r}) = \kappa_s^2 \phi(\mathbf{r}) \quad (\text{B.46})$$

とかける．ここで次式を満たすグリーン関数  $G$  を導入する．

$$(\nabla^2 - \kappa_s^2)G = -\frac{1}{\varepsilon_s} \delta(\mathbf{R} - \mathbf{r}) \quad (\text{B.47})$$

グリーンの公式を用いると，静電ポテンシャルは，

$$\phi(\mathbf{r}) = \int_{\Gamma} \varepsilon_s (\nabla \phi \cdot \mathbf{n}) G d\mathbf{S} + \int_{\Gamma} \varepsilon_s \phi (\nabla G \cdot \mathbf{n}) d\mathbf{S} \quad (\text{B.48})$$

とかける．ただし， $\mathbf{n}$  は面  $\Gamma$  の外向きの単位法線ベクトルを表す．概念図を Fig. B.3 に示す．

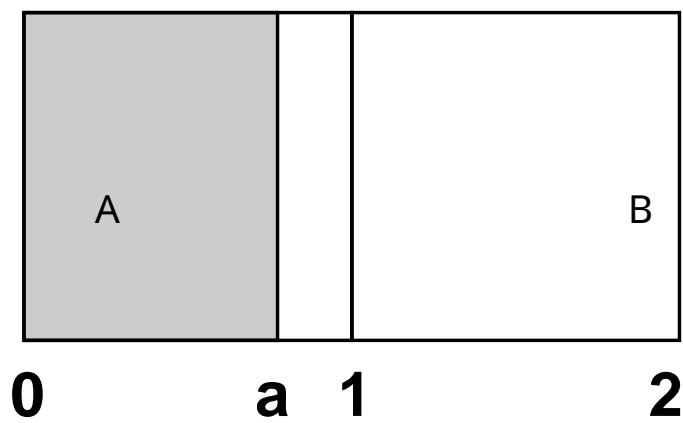


Figure B.2 格子面への誘電率のマッピングを示す概念図 .

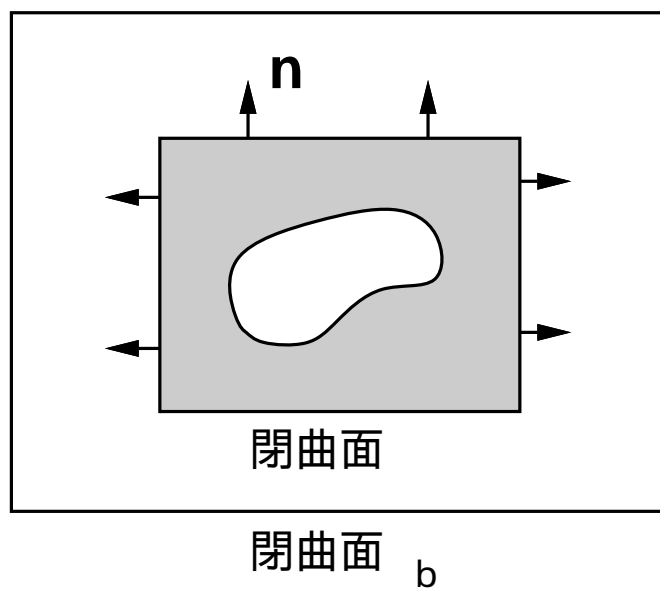


Figure B.3 自己無撞着の境界条件設定の概念図 .

## B.3 Poisson-Boltzmann 分子動力学法

誘電率と Debye 遮蔽定数が空間中で一様でない場合は、Poisson-Boltzmann 方程式を解くことは困難なので、まず誘電率  $\epsilon_c$ 、Debye 遮蔽定数  $\kappa_c$  が一様な場合を考える。このとき Poisson-Boltzmann 方程式は次式で与えられる。

$$\epsilon_c \Delta \phi(\mathbf{r}) = -\rho(\mathbf{r}) + \epsilon_c \kappa_c^2 \phi(\mathbf{r}) \quad (\text{B.49})$$

この式は、無限遠で  $\phi(\mathbf{r}) = 0$ 、 $\nabla \phi(\mathbf{r}) = \mathbf{0}$  という境界条件のもとで、以下のように解くことができる。

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_c} \int \frac{\rho(\mathbf{R}) \exp(-\kappa_s |\mathbf{R} - \mathbf{r}|)}{|\mathbf{R} - \mathbf{r}|} d\mathbf{R}^3 \quad (\text{B.50})$$

ここで、誘電率  $\epsilon_c$ 、Debye 遮蔽定数  $\kappa_c$  として、溶媒での値 ( $\epsilon_s$  および  $\kappa_s$ ) を用いると、式 (B.49) で静電ポテンシャル  $\phi(\mathbf{r})$  が求めれば、それを再現する電荷分布  $\rho^i(\mathbf{r})$  は次式のように求められる。

$$\rho^i(\mathbf{r}) = \epsilon_s \kappa_s^2 \phi(\mathbf{r}) - \epsilon_s \Delta \phi(\mathbf{r}) \quad (\text{B.51})$$

これを式 (B.50) に代入すると、

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_s} \int \frac{\rho^i(\mathbf{R}) \exp(-\kappa_s |\mathbf{R} - \mathbf{r}|)}{|\mathbf{R} - \mathbf{r}|} d\mathbf{R}^3 \quad (\text{B.52})$$

となる。式 (B.51) で求めた、誘電率・遮蔽定数が空間中で一様なときの静電ポテンシャルを再現する電荷分布  $\rho^i(\mathbf{r})$  は、空間の格子点上に与えられる。これに対して分子動力学計算などでは、電荷は各原子の中心に存在すると考える。このため  $\rho^i$  も各原子に割り振ることを考える。

- 分子表面に存在する仮想電荷は、距離がもっとも近い原子へ割り振る。
- 分子内部に位置する仮想電荷は、 $\mathbf{r}$  を内部に含むすべての原子に均等に割り振る。

このとき、仮想電荷分布は次式のように書ける。

$$\rho^i(\mathbf{r}) = \sum_j C_j^i(\mathbf{r}) \quad (\text{B.53})$$

ただし、 $C_j^i(\mathbf{r})$  は、原子  $j$  に割り振られた仮想電荷分布である。さらに静電場を原子間距離と各原子の電荷の大きさの関数で近似することを考える。

$$\phi(\mathbf{r}) \sim \sum_j q_j^e f(|\mathbf{r} - \mathbf{r}_j|) \quad (\text{B.54})$$





右辺の第 1 項は距離  $d_j$  に依存しないので  $q_j^e$  に，第 2 項は距離  $d_j$  の関数なので  $f(d_j)$  に対応する．

境界領域，溶質領域 ( $d_j \leq b_j$ ) ではイオンは存在しないので，ガウスの定理より原子  $j$  から距離  $d_j$  のところの静電ポテンシャルは次式で与えられる．

$$\phi_j^i(d_j) = \frac{1}{4\pi\epsilon_p} \frac{Q_j}{d_j^2} \quad (\text{B.59})$$

ただし， $Q_j$  は半径  $d_j$  の球殻内に存在する電荷量の合計である．

以上から，原子  $j$  によって作り出される静電ポテンシャルは，次式で近似できることになる．

$$\phi_j(d_j) = \begin{cases} q_j^e \frac{1}{4\pi\epsilon_s} \frac{\exp(-\kappa_s d_j)}{d_j} & b_j < d_j \\ Q_j \frac{1}{4\pi\epsilon_b} \frac{1}{d_j} + \alpha_j & r_j < d_j \leq b_j \\ q_j \frac{1}{4\pi\epsilon_p} \frac{1}{d_j} + \beta_j & d_j \leq r_j \end{cases} \quad (\text{B.60})$$

ただし， $r_j$  は原子  $j$  の原子半径， $b_j$  は  $r_j + r_{probe}$ ， $Q_j$  は原子  $j$  の内部と表面に存在する電荷の総量， $q_j$  は原子  $j$  の電荷である． $r_{probe}$  は水分子の半径 1.4Å に設定する．また  $\alpha_j$  と  $\beta_j$  は定数で， $\phi_j(d_j)$  が連続関数となるように定義される．

これを用いると，系全体の静電ポテンシャル  $E^{el}$  は次式で与えられる．

$$E^{el} = \frac{1}{2} \sum_k q_k \sum_j \phi_j(r_{jk}) \quad (\text{B.61})$$

ただし， $r_{jk}$  は原子  $j$  と原子  $k$  との間の距離である．

各原子に加わる溶媒からの力と真空中での静電気力の和は，実効電荷  $q^e$  の構造依存性は小さいとして実効電荷の微分項をゼロとみると，次式で表される．

$$\mathbf{F}^{el} = \frac{1}{2} \sum_{i \neq j} (q_j^e q_i + q_i^e q_j) \frac{1 + \kappa_s r_{ij} \exp(-\kappa_s r_{ij})}{4\pi\epsilon_s r_{ij}^2} (\mathbf{r}_i - \mathbf{r}_j) \quad (\text{B.62})$$

ただし  $f(r) = \exp(-\kappa_s r)/4\pi\epsilon_s r$  である．

## C WHAM の理論

系の「状態」を，エネルギー  $E$  と任意に設定した反応座標  $\xi$  (reaction coordinate) の関数であると考える．

WHAM (Weighted Histogram Analysis Method)<sup>24,95,25)</sup> は，分子動力学法や Monte Carlo 法のシミュレーションから得られたアンサンブルをもとに，系の状態密度を見積もり，任意の温度についての状態が出現する確率密度  $P_\beta(E, \xi)$  を求めることがで

きる方法である．温度の異なる複数の run から誤差が最小になるように求めることができる．また  $E$ ,  $\xi$  などに関する拘束をかけながらシミュレーションを行っても，その影響を取り除くことができる．

### C.1 1回のシミュレーションからの系の状態密度の見積もり

エネルギーが  $E \sim E + \Delta E$  かつ反応座標が  $\xi \sim \xi + \Delta\xi$  の範囲に含まれる系の状態の数を  $\Omega(E, \xi)$  個とする．カノニカルアンサンブルの場合，エネルギー値  $E$  の状態 1 個が出現する確率  $w(E)$  は，

$$w(E) \propto \exp(-\beta E) \quad (\beta \equiv 1/k_B T) \quad (C.63)$$

である．このとき，系のエネルギーが  $E$  かつ反応座標が  $\xi$  となる確率  $P(E, \xi)$  は，

$$P(E, \xi) \propto \Omega(E, \xi)w(E) = \Omega(E, \xi) \exp(-\beta E) \quad (C.64)$$

となる．ここで系の分配関数を，

$$Z(\beta) = \int \Omega(E, \xi) \exp(-\beta E) dE d\xi \quad (C.65)$$

とすると，

$$P(E, \xi) = \frac{1}{Z(\beta)} \Omega(E, \xi) \exp(-\beta E) \quad (C.66)$$

となる．

シミュレーションでは，総ステップ数  $n$  のうち，エネルギーが  $E \sim E + \Delta E$  かつ反応座標が  $\xi \sim \xi + \Delta\xi$  の状態を生成したステップ数  $N(E, \xi)$  を数えることで，

$$P(E, \xi) = \frac{N(E, \xi)}{n} \quad (C.67)$$

$$\Omega(E, \xi) = Z(\beta) \frac{N(E, \xi)}{n} \exp(\beta E) \quad (C.68)$$

と系の状態密度を見積もることができる．

系に拘束エネルギー  $E_c(E, \xi)$  を導入すると，式 (C.66) は以下ようになる．

$$P(E, \xi) = \frac{1}{Z_c(\beta)} \Omega(E, \xi) \exp(-\beta(E + E_c)) \quad (C.69)$$

ただし  $Z_c(\beta)$  は拘束のある系の分配関数であり，

$$Z_c(\beta) = \int \Omega(E, \xi) \exp(-\beta(E + E_c)) \quad (C.70)$$

とかける．

式 (C.69) と式 (C.70) を連立させることで， $Z_c(\beta)$  と拘束がない系での状態密度  $\Omega(E, \xi)$  を求めることができる．シミュレーションが理想的に行われていれば 1 回の run で  $\Omega(E, \xi)$  を求めることが原理的には可能であるが，実際には複数の run を統合して精度の向上を図る．

## C.2 複数のシミュレーションからの系の状態密度の見積もり

次に,  $R$  回のシミュレーションを行った場合を考える. 次の変数および関数を定義する.

$N_i(\xi, E)$  :  $i$  番目のシミュレーション中に状態  $(\xi, E)$  が現われた回数

$n_i$  :  $i$  番目のシミュレーションの総ステップ数

$T_i$  :  $i$  番目のシミュレーションの温度

$\beta_i$  :  $\frac{1}{k_B T_i}$

$\Omega_i(\xi, E)$  :  $i$  番目のシミュレーションより見積もられる状態密度関数

$Z_i$  :  $i$  番目のシミュレーションの分配関数

$E_i(E, \xi)$  :  $i$  番目のシミュレーションの構造拘束エネルギー.

ただし  $E_i$  は変数として  $\xi, E$  のみをとるとする.

$k$  番目のシミュレーション条件を温度  $T_k$ , 総ステップ数  $n_k$  とすると,  $k$  番目のシミュレーションから系の状態密度  $\Omega_k(E, \xi)$  を, 式 (C.68) を用いて次のように見積もることができる.

$$\Omega_k(E, \xi) = Z_k(\beta_k) \frac{N_k(E, \xi)}{n_k} \exp(\beta_k(E + E_k)) \quad (\beta_k = 1/k_B T_k) \quad (C.71)$$

拘束エネルギー  $E_k(E, \xi)$  をどのようにとっても, 拘束がない系での状態密度  $\Omega(E, \xi)$  は共通であるので, それぞれの run で見積もられた  $\Omega_k(E, \xi)$  は本来一致するはずである. しかし実際には統計誤差などにより, 複数の run で完全に一致することはない. そこで, 統計的に信頼性が高くなるように複数の run の結果を統合して,  $\Omega(E, \xi)$  を求めることを考える.

真の状態密度  $\Omega$  を,  $R$  個の見積もり  $\Omega_1, \dots, \Omega_R$  からできるだけ高い精度で推測するには, 各見積もりに「信頼性」にしたがった重みをつけて,

$$\Omega(E, \xi) = \sum_{j=1}^R w_j(E, \xi) \Omega_j(E, \xi) \quad (C.72)$$

$$\sum_{j=1}^R w_j(E, \xi) = 1 \quad (C.73)$$

とし, 誤差を最小にするように  $w_j$  を決めればよい. 2 乗誤差  $\delta^2 \Omega$  は, 式 (C.68)(C.72) より次のように表せる.

$$\delta^2 \Omega(E, \xi) = \sum_i w_i(E, \xi)^2 \delta^2 \Omega_i(E, \xi)$$

$$\begin{aligned}
&= \sum_i w_i(E, \xi)^2 \delta^2 \left( Z_i \frac{N_i(E, \xi)}{n_i} \exp[\beta_i(E + E_i)] \right) \\
&= \sum_i w_i(E, \xi)^2 Z_i^2 \frac{1}{n_i^2} \exp[2\beta_i(E + E_i)] \delta^2 N_i(E, \xi) \quad (C.74)
\end{aligned}$$

ここで  $\delta^2 N_k$  を，以下のように見積もる．

$$\delta^2 N_k(E, \xi) = g_k \overline{N_k(E, \xi)} \quad (C.75)$$

ただし  $g_k = 1 + 2\tau_k$  ( $\tau_k$ : integrated correlation time)，また  $\overline{N_k}$  は，総ステップ数  $n_k$  のシミュレーションを多数回行ったときの  $N_k$  の平均値である．

一般に， $N$  個の構造を分子動力学法でサンプルしても，ある構造の前後  $\tau$  ステップ間では構造に大きな相関があるとすると，サンプルした構造の数は実質的には  $N/g$  ( $g = 2\tau + 1$ ) 個となる． $g_k$  はこの補正のために導入される．

式 (C.75) は， $N_k$  の計数過程をポアソン過程であると近似している． $\overline{N_k(E, \xi)}$  を実際に求めるためには多数回の run を行う必要があるが，これを

$$\overline{N_k(E, \xi)} = n_k \frac{\Omega(E, \xi)}{Z_i} \exp(-\beta_k(E + E_i)) \quad (C.76)$$

と見積もる．以上より，最適な  $\Omega(E, \xi)$  は，式 (C.73)~(C.76) を満たすような  $w_j(E)$  を Lagrange の未定乗数法によって計算し，それを式 (C.72) に代入することによって求まる．この計算を実行すると， $w_j$  は次式のようになる．

$$w_j = \frac{\frac{n_j}{g_j Z_j} \exp(-\beta_j(E + E_j))}{\sum_{m=1}^R \frac{n_m}{g_m Z_m} \exp(-\beta_m(E + E_m))} \quad (C.77)$$

式 (C.71)(C.72)(C.77) より， $\Omega$  は次式のようになる．

$$\Omega(E, \xi) = \frac{\sum_{k=1}^R N_k(E, \xi) g_k^{-1}}{\sum_{m=1}^R \frac{n_m}{g_m Z_m} \exp(-\beta_m(E + E_m))} \quad (C.78)$$

$$Z_i = \sum_{E, \xi} \Omega(E, \xi) \exp[-\beta_i(E + E_i)] \quad (C.79)$$

式 (C.78) と式 (C.79) を連立させると， $\Omega(E, \xi)$  を求めることができる．これは非線形連立方程式であり，解析的に解を求めることが困難であるので，まず適当な  $Z_i$  を用いて式 (C.78) により  $\Omega(E, \xi)$  を求め，これを式 (C.79) に代入して  $Z_i$  を求め，これを再び式 (C.78) に代入して  $\Omega(E, \xi)$  を求める，というように  $Z_i$  が収束するまで反復して計算を行う．

$\Omega(E, \xi)$  が求めれば，これを式 (C.66) に代入することで，拘束がない系で，任意の温度において状態  $(E, \xi)$  が出現する確率密度  $P(E, \xi)$  を求めることが可能となる．

## D 基準振動モード解析の理論

分子内振動の自由度に対応した一般化座標を  $q_1, q_2, \dots, q_n$  (ベクトル表示で  $\mathbf{q}$ ) とする。本計算の場合は、共有結合まわりの二面角をパラメータにしているの、一般化座標には二面角  $\theta_1, \theta_2, \dots, \theta_n$  が対応することになる。ポテンシャルエネルギーの極小点近傍では、 $V$  を極小点  $\mathbf{q}^0$  で展開し、 $q_i^0$  からのずれの 2 乗までで近似することができる。

$$V(\mathbf{q}) = V(\mathbf{q}^0) + \sum_i \left. \frac{\partial V}{\partial q_i} \right|_{\mathbf{q}=\mathbf{q}^0} (q_i - q_i^0) + \frac{1}{2} \sum_{i,j} \left. \frac{\partial^2 V}{\partial q_i \partial q_j} \right|_{\mathbf{q}=\mathbf{q}^0} (q_i - q_i^0)(q_j - q_j^0) \quad (\text{D.80})$$

右辺第 1 項はポテンシャルエネルギーの零点を極小点におくことにより、また、第 2 項は極小点での微係数は 0 であることから、ともに 0 となる。この結果、極小点  $\mathbf{q}^0$  の近傍でのポテンシャルエネルギー  $V$  は座標の 2 次関数で表される。

$$V(\mathbf{q}) = \frac{1}{2} \sum_{i,j} F_{ij} (q_i - q_i^0)(q_j - q_j^0) \quad (\text{D.81})$$

ただし、

$$F_{ij} = \left. \frac{\partial^2 V}{\partial q_i \partial q_j} \right|_{\mathbf{q}=\mathbf{q}^0} \quad (\text{D.82})$$

とおいた。

一方、運動エネルギー  $E_k$  は、座標  $\mathbf{q}$  として直交座標  $\mathbf{r}$  を用いた場合、次式のように速度  $d\mathbf{r}_a/dt$  について定係数の 2 次の同次関数として表現される。

$$E_k = \frac{1}{2} \sum_a m_a \left( \frac{d\mathbf{r}_a}{dt} \right)^2 \quad (\text{D.83})$$

ここで  $m_a$  は、原子  $a$  の質量である。

一般化座標  $\mathbf{q}$  として二面角を用いたときには、系全体の運動を、分子内運動と分子全体の並進や回転である外部運動とに分離することが必要となる。このための条件は Eckart の条件<sup>170)</sup> といわれる。エネルギー極小構造での原子  $a$  の直交座標を  $\mathbf{r}_a^0$ 、ゆらいたときの座標を  $\mathbf{r}_a$  とする。まず分子全体の質量中心が移動しないためには、

$$\sum_a m_a \mathbf{r}_a = 0 \quad (\text{D.84})$$

が成り立つ必要がある。また質量中心まわりの回転をしないためには、

$$\sum_a m_a \mathbf{r}_a \times \dot{\mathbf{r}}_a = 0 \quad (\text{D.85})$$

が成り立つ必要がある．式 (D.85) は一般的には解析的に求めることが不可能であるが，ゆらぎが十分に小さい(すなわち  $\mathbf{r}_a$  を  $\mathbf{r}_a^0$  と置き換えてもよい) とすると，

$$\sum_a m_a \mathbf{r}_a^0 \times \dot{\mathbf{r}}_a = 0 \quad (\text{D.86})$$

とできる．式 (D.84)，(D.86) が成り立っているとき，系全体の運動は内部運動に限られるので，

$$\frac{d\mathbf{r}_a}{dt} = \sum_i \frac{\partial \mathbf{r}_a}{\partial q_i} \frac{dq_i}{dt} \quad (\text{D.87})$$

となる．これを式 (D.83) に代入すると，運動エネルギー  $E_k$  は，

$$\begin{aligned} E_k &= \frac{1}{2} \sum_a m_a \left( \sum_i \frac{\partial \mathbf{r}_a}{\partial q_i} \frac{dq_i}{dt} \right) \left( \sum_j \frac{\partial \mathbf{r}_a}{\partial q_j} \frac{dq_j}{dt} \right) \\ &= \frac{1}{2} \sum_{i,j} \left( \sum_a m_a \frac{\partial \mathbf{r}_a}{\partial q_i} \cdot \frac{\partial \mathbf{r}_a}{\partial q_j} \right) \dot{q}_i \dot{q}_j \\ &= \frac{1}{2} \sum_{i,j} H_{ij} \dot{q}_i \dot{q}_j \end{aligned} \quad (\text{D.88})$$

となる．ただし，

$$H_{ij} = \sum_a m_a \frac{\partial \mathbf{r}_a}{\partial q_i} \cdot \frac{\partial \mathbf{r}_a}{\partial q_j} \quad (\text{D.89})$$

である．微小振動を考えているので，運動エネルギーマトリクス  $H_{ij}$  は極小点  $q^0$  での値，すなわち定係数行列とみなすことになる．

式 (D.81)，(D.88) を Lagrange の運動方程式

$$\begin{aligned} \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} &= 0 \\ (L = E_k - V) \end{aligned} \quad (\text{D.90})$$

に代入すると，

$$\sum_j H_{ij} \ddot{q}_j = - \sum_j F_{ij}(q_j - q_j^0) \quad (\text{D.91})$$

となる．この式は定係数の線形微分方程式であり，その一般解  $q_i$  は

$$q_i = q_i^0 + \sum_\lambda A_{i\lambda} Q_\lambda \quad (\text{D.92})$$

ただし  $Q_\lambda$  は基準振動であり，

$$Q_\lambda = \alpha_\lambda \cos(\omega_\lambda t + \delta_\lambda) \quad (\text{D.93})$$

と，調和振動子の線形重ね合わせによって求めることができる． $q_i$  には個々の原子運動が直接対応しているのに対して，基準振動  $Q_\lambda$  は分子内の全ての原子が示す集団

運動に対応する．つまりパラメータ  $q_i$  と  $Q_\lambda$  は同一の分子の運動を異なった側面から見ていることになる．その2つのパラメータを結ぶ関係式が式 (D.92) である．

式 (D.92) を式 (D.91) に代入することにより，重ね合わせの係数  $A_{i\lambda}$  と固有振動数  $\omega_\lambda$  に関する一般化固有値方程式が得られる．

$$\left(\sum_i H_{ji}A_{i\lambda}\right)(\omega_\lambda)^2 = \sum_i F_{ji}A_{i\lambda} \quad (\text{D.94})$$

これを行列表示すると，

$$H A A = F A \quad (\text{D.95})$$

となる．基準振動モード計算法は，この一般化固有値方程式を，各固有値  $A_{ij}$  に対応する固有ベクトル間の直交条件

$$A^T H A = \{\delta_{ij}\} \quad (\text{D.96})$$

$$\delta_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (\text{otherwise}) \end{cases}$$

を満たすように解くことに帰着される． $A_{ij} = (\omega_i)^2 \delta_{ij}$  であるから，基準振動モードの計算では固有値  $A_{ij}$  が非負値であることが前提であり，そのためにはポテンシャルエネルギーの極小点  $q^\circ$  が厳密に求められている必要がある．

解析にあたっては，以下のような手順となる．

1. X 線結晶解析などのデータから，対象とする分子の3次元座標を得る．
2. 手順1で得られた構造がポテンシャルエネルギー  $V$  の極小点であることはまれであるため，手順1の構造を座標  $q$  で表し， $q$  に関するポテンシャルエネルギー  $V$  の極小化を行う．この極小化によって得られた極小点近傍の構造における微小振動が，解析対象となる．
3. 式 (D.82)，(D.89) に従って極小点での行列  $H$ ， $F$  を計算し，式 (D.95)，(D.96) の一般化固有値方程式から固有値  $(\omega_\lambda)^2$  と固有ベクトル  $A_{i\lambda}$  を求める．
4. 得られた  $\omega_\lambda$  と  $A_{i\lambda}$  を式 (D.92) に代入することによって，対象分子の分子内振動を示す原子の運動がシミュレートされる．

式 (D.93) における基準振動の位相  $\delta_\lambda$  と振幅  $\alpha_\lambda$  は，原子の位置と速度に関する初期条件から決定されなければならない．しかし熱平衡状態においては各基準振動にエネルギーが等分配されることから，振幅  $\alpha_\lambda$  については絶対温度  $T$  と対応する振動

数  $\omega_\lambda$  の関数として与えられる．すなわち，各基準振動のポテンシャルエネルギーの時間平均

$$\frac{1}{2}(\omega_\lambda)^2 \alpha_\lambda^2 \langle \cos^2(\omega_\lambda t + \delta_\lambda) \rangle = \frac{1}{4}(\omega_\lambda)^2 \alpha_\lambda^2 \quad (\text{D.97})$$

(ただし， $\langle \rangle$  は時間平均を表す．)

が  $1/2k_B T$  ( $k_B$  はボルツマン定数) に等しいことから，

$$\alpha_\lambda = \sqrt{\frac{2k_B T}{(\omega_\lambda)^2}} \quad (\text{D.98})$$

となる．

ここでは二面角  $\Delta\theta_i(t) = \sum_\lambda A_{i\lambda} \alpha_\lambda \cos(\omega_\lambda t + \delta_\lambda)$  をパラメータにしているので，以後は一般化座標  $q$  を  $\Delta\theta$  に置き換えて考えることにする．

原子運動のシミュレートには位相  $\delta_\lambda$  が必要であるが，種々の物理量の時間平均は，位相  $\delta_\lambda$  には依存せず，基準振動の振幅  $\alpha_\lambda$  と固有ベクトル  $A_{i\lambda}$  および二面角  $\Delta\theta_i$  から求められる．

二面角の対の同一時刻相関係数  $\langle \Delta\theta_i(\tau) \Delta\theta_j(\tau) \rangle$  は，

$$\langle \Delta\theta_i(\tau) \Delta\theta_j(\tau) \rangle = \frac{1}{2} \sum_\lambda A_{i\lambda} A_{j\lambda} \alpha_\lambda^2 \quad (\text{D.99})$$

となる．とくに  $i = j$  のときは，式 (D.99) は二面角  $\theta_i$  の平均 2 乗ゆらぎ

$$\langle \theta_i^2(\tau) \rangle = \frac{1}{2} \sum_\lambda A_{i\lambda}^2 \alpha_\lambda^2 \quad (\text{D.100})$$

を与える．

また， $\Delta\theta_i$  の線形関数である系の任意の量  $\Delta p_a = \sum_i P_{ai} \Delta\theta_i$  に対する相関係数は

$$\langle \Delta p_a(\tau) \Delta p_b(\tau) \rangle = \frac{1}{2} \sum_\lambda \sum_i \sum_j P_{ai} P_{bj} A_{i\lambda} A_{j\lambda} \alpha_\lambda^2 \quad (\text{D.101})$$

となる．たとえば  $p_a$  として原子の直交座標  $r_a$  を考えれば， $P_{ai} = \partial r_a / \partial \theta_i$  となり，式 (D.101) は原子  $a$  と  $b$  のゆらぎの相関を与える．さらに  $a = b$  と置けば，

$$\langle (\Delta r_a)^2 \rangle = \sum_\lambda \left( \sum_i \frac{\partial r_a}{\partial \theta_i} A_{i\lambda} \right)^2 \frac{kT}{\omega_\lambda^2} \quad (\text{D.102})$$

は，原子  $a$  の位置の変位の平均 2 乗ゆらぎである．また，モード  $\lambda$  における原子  $a$  の変位を表すベクトル  $D_{a\lambda}$  は，次の式であたえられる．

$$D_{a\lambda} = \sum_i \frac{\partial r_a}{\partial \theta_i} A_{i\lambda} \sqrt{\frac{kT}{\omega_\lambda^2}} \quad (\text{D.103})$$



## 参考文献

- 1) R. Dulbecco. A turning point in cancer research: Sequencing the human genome. *Science*, Vol. 231, pp. 1055–56, 1986.
- 2) International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, Vol. 409, pp. 860–921, 2001.
- 3) J. C. Venter, et al. The sequence of the human genome. *Science*, Vol. 291, pp. 1304–1351, 2001.
- 4) <http://www.top500.org/>
- 5) T. Murata, M. Sekijima, H. Masaki, S. Nakamura, M. Ikeguchi, and K. Shimizu. Parallel programming environment with dependence-driven sub-task scheduling: Design and application to molecular dynamics simulation. In *Proceedings of the Second IASTED International Conference on Parallel and Distributed Computer and Networks (PDCN '98)*, 1998.
- 6) S. Nakamura, M. Ikeguchi, and K. Shimizu. Parallel algorithm for efficient calculation of second derivatives of conformational energy function in internal coordinates. *J. Comput. Chem.*, Vol. 18, pp. 1716–1723, 1998.
- 7) 西村健, 中村周吾, 池口満徳, 清水謙多郎. 分子動力学シミュレーションの並列化における分割法の一考察. 情報処理学会ハイパフォーマンスコピューティング研究会 (HOKKE '99), 第 99 巻, pp. 61–66, 1999.
- 8) M. Sekijima, S. Takasaki, S. Nakamura, M. Ikeguchi, and K. Shimizu. Parsley: A scalable framework for dependence-driven task scheduling in distributed-memory multiprocessor systems. In *Proceedings of the Eleventh International Conference on Parallel and Distributed Computing and Systems (PDCS '99)*, 1999.

- 9) M. Sekijima, S. Takasaki, S. Nakamura, M. Ikeguchi, and K. Shimizu. A parallel programming environment with dependence-driven task scheduling in distributed-memory multiprocessor systems. In *Proceedings of the International Conference on Parallel and Distributed Computing Systems*, pp. 348–354, 2000.
- 10) M. Takasaki, S. Sekijima, S. Nakamura, M. Ikeguchi, and K. Shimizu. Scheduling policy and mechanism of Parsley parallel programming environment. In *Proceedings of the IASTED Eleventh International Conference on Parallel and Distributed Computing and System*, pp. 605–610, 2000.
- 11) J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, Vol. 267, pp. 585–590, 1977.
- 12) M. Levitt. Computer simulation of DNA double-helix dynamics. In *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 47, pp. 251–262. Cold Spring Harbor Laboratory, 1983.
- 13) B. Tidor, K. K. Irikura, B. R. Brooks, and M. Karplus. Dynamics of DNA oligomers. *J. Biomol. Str. Dyn.*, Vol. 1, pp. 231–252, 1983.
- 14) 中村周吾, 池口満徳, 廣瀬仁, 清水謙多郎. 並列構造探索システムの構築と核酸ミニヘアピン分子への適用. *生物物理*, Vol. 37, p. S37, 1997.
- 15) K. Tazaki and J. Doi. Effective atomic charge model for solvent and ion screening effect in iterative calculation and application to Monte Carlo simulation. *J. Phys. Chem.*, Vol. 100, pp. 14520–14525, 1996.
- 16) K. Tazaki and K. Shimizu. Molecular dynamics simulations in aqueous solution: Application to free energy calculation of oligopeptides. *J. Phys. Chem. B*, Vol. 102, pp. 6419–6424, 1998.
- 17) S. Nakamura, D. Kyono, M. Ikeguchi, and K. Shimizu. A new method for parallel computation of hessian matrix of conformational energy function in internal coordinates. *J. Comput. Chem.*, in press.
- 18) S. Yoshizawa, T. Ueda, Y. Ishido, K. Miura, K. Watanabe, and I. Hirao. Nuclease resistance of an extraordinarily thermostable mini-hairpin DNA fragment,

- d(GCGAAGC) and its application to *in vitro* protein synthesis. *Nuc. Acids Res.*, Vol. 22, pp. 2217–2221, 1994.
- 19) S. Yoshizawa, G. Kawai, K. Watanabe, K.-I. Miura, and I. Hirao. GNA trinucleotide loop sequences producing extraordinarily stable DNA minihairpins. *Biochemistry*, Vol. 36, pp. 4761–4767, 1997.
  - 20) I. Hirao, G. Kawai, S. Yoshizawa, Y. Nishimura, Y. Ishido, K. Watanabe, and K.-I. Miura. Most compact hairpin-turn structure exerted by a short DNA fragment, d(GCGAAGC) in solution: An extraordinarily stable structure resistant to nucleases and heat. *Nuc. Acids Res.*, Vol. 22, pp. 576–582, 1994.
  - 21) S. Nakamura, H. Hirose, M. Ikeguchi, and J. Doi. Conformational energy minimization using a two-stage method. *J. Phys. Chem.*, Vol. 99, pp. 8374–8378, 1995.
  - 22) H. Hirose, S. Nakamura, M. Ikeguchi, and J. Doi. Improved diffusion equation method for large molecules. In M. J. Geisow and R. Epton, editors, *Perspectives on Protein Engineering*, pp. 49–52. Mayflower Worldwide Ltd., 1995.
  - 23) S. Nakamura, H. Hirose, M. Ikeguchi, and K. Shimizu. Structural modeling of DNA mini-hairpin molecules with various loop sequences. *Chem. Phys. Lett.*, Vol. 308, pp. 267–273, 1999.
  - 24) S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. I. the method. *J. Comput. Chem.*, Vol. 13, pp. 1011–1021, 1992.
  - 25) S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.*, Vol. 16, pp. 1339–1350, 1995.
  - 26) 中村周吾, 田崎康一, 池口満徳, 清水謙多郎. DNA ミニヘアピン分子の熱安定性の解析. *生物物理*, Vol. 36, p. S166, 1996.
  - 27) S. Nakamura and J. Doi. Vibration analysis of transfer RNAs using normal mode calculation. In M. Doyama, et al., editors, *Computer Aided Innovation of New Materials II*, pp. 1203–1206, 1993.

- 28) S. Nakamura and J. Doi. Conformational analysis of transfer RNAs using normal mode calculation. In *Fifteenth International tRNA Workshop*, p. 300, 1993.
- 29) 中村周吾, 土井淳多. 核酸分子の基準振動モード解析. *生物物理*, Vol. 34, p. S109, 1994.
- 30) S. Nakamura and J. Doi. Dynamics of transfer RNAs analyzed by normal mode calculation. *Nuc. Acids Res.*, Vol. 22, pp. 514–521, 1994.
- 31) 中村周吾, 池口満徳, 清水謙多郎. tRNA-ARS 複合体の基準振動モード解析. 第4回「タンパク質立体構造の構築原理」ワークショップ, p. 181, 1997.
- 32) 中村周吾, 池口満徳, 清水謙多郎. アミノアシル tRNA 合成酵素・tRNA 複合体の動特性解析. 第5回「タンパク質立体構造の構築原理」ワークショップ, p. 225, 1998.
- 33) 中村周吾, 池口満徳, 清水謙多郎. tRNA-ARS 複合体の動特性解析. *生物物理*, Vol. 38, p. S118, 1998.
- 34) 中村周吾, 池口満徳, 清水謙多郎. ARS-tRNA 複合体のダイナミクス解析. *生物物理*, Vol. 39, p. S178, 1999.
- 35) 中村周吾, 池口満徳, 清水謙多郎. ARS-tRNA 複合体のドメイン運動解析. 蛋白合同年会 横浜 99, p. 163, 1999.
- 36) H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nuc. Acids Res.*, Vol. 28, pp. 235–242, 2000.
- 37) <http://www.rcsb.org/pdb/>
- 38) R. Sayle and E. J. Milner-White. RasMol: Biomolecular graphics for all. *Trends Biochem. Sci.*, Vol. 20, p. 374, 1995.
- 39) 石井隆寛, 池口満徳, 土井淳多. 生体分子のための対話型分子グラフィックスシステム LIVE の開発. *情報処理学会論文誌*, Vol. 32, pp. 590–598, 1991.

- 40) G. Nemethy, M. S. Pottle, and H. A. Scheraga. Energy parameters in polypeptides. 9. updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. *J. Phys. Chem.*, Vol. 87, pp. 1883–1887, 1983.
- 41) B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, Vol. 4, pp. 187–217, 1983.
- 42) S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, and P. Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, Vol. 106, pp. 765–784, 1984.
- 43) S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.*, Vol. 7, pp. 230–252, 1986.
- 44) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Jr. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.*, Vol. 117, pp. 5179–5197, 1995.
- 45) A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, Vol. 103, pp. 2227–2249, 1976.
- 46) B. E. Hingerty, R. H. Ritchie, T. L. Ferrell, and J. E. Turner. Dielectric effects in biopolymers: The theory of ionic saturation revisited. *Biophys. J.*, Vol. 24, pp. 427–439, 1985.
- 47) 江口至洋. タンパク質工学の物理・化学的基礎, 第1章. 共立出版, 1991.
- 48) J. Choi, J. Dongarra, R. Pozo, and D. Walker. ScaLAPACK: A scalable linear algebra library for distributed memory concurrent computers. In *Proceedings of the Fourth Symposium on the Frontiers of Massively Parallel Computation*, 1992.

- 49) L. S. Blackford, J. Choi, A. Cleary, J. Demmel, I. Dhillon, J. J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. W. Walker, and R. C. Whaley. ScaLAPACK: A portable linear algebra library for distributed memory computers - design issues and performance. In *Proceedings of Supercomputing '96*, 1996.
- 50) L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley. *ScaLAPACK Users' Guide*. Society for Industrial and Applied Mathematics, 1997.
- 51) <http://www.netlib.org/scalapack/>
- 52) T. Noguti and N. Gō. Efficient Monte Carlo method for simulation of fluctuating conformations of native proteins. *Biopolymers*, Vol. 24, pp. 527–546, 1985.
- 53) T. Noguti and N. Gō. A method of rapid calculation of a second derivative matrix of conformational energy for large molecules. *J. Phys. Soc. Jpn.*, Vol. 52, pp. 3685–3690, October 1983.
- 54) T. Noguti and N. Gō. Dynamics of native globular proteins in terms of dihedral angles. *J. Phys. Soc. Jpn.*, Vol. 52, pp. 3283–3288, September 1983.
- 55) H. Abe, W. Braun, T. Noguti, and N. Gō. Rapid calculation of first and second derivatives of conformational energy with respect to dihedral angles for proteins. general recurrent equations. *Computers & Chem.*, Vol. 8, pp. 239–247, 1984.
- 56) H. Wako and N. Gō. Algorithm for rapid calculation of hessian of conformational energy function of proteins by supercomputer. *J. Comput. Chem.*, Vol. 8, pp. 625–635, 1987.
- 57) 輪湖博, 郷信広. 蛋白質の立体構造シミュレーション. 田中實, 山本良一(編), 計算物理学と計算化学, 第11章, pp. 196–219. 海文堂出版, 東京, 1988.
- 58) R. Esser, P. Grassberger, J. Grotendorst, and M. Lewerenz, editors. *Molecular Dynamics on Parallel Computers*. World Scientific, 8–10 Feb 1999.

- 59) S. Toyoda, H. Miyagawa, K. Kitamura, T. Amisaki, E. Hashimoto, H. Ikeda, A. Kusumi, and N. Miyakawa. Development of MD engine: High-speed accelerator with parallel processor design for molecular dynamics simulations. *J. Comput. Chem.*, Vol. 20, pp. 185–199, 1999.
- 60) M. Totrov and R. Abagyan. Efficient parallelization of the energy, surface, and derivative calculations for internal coordinate mechanics. *J. Comput. Chem.*, Vol. 15, pp. 1105–1112, 1994.
- 61) G. M. Amdahl. Validity of single-processor approach to achieving large-scale computing capability. *Proc. AFIPS Conf.*, pp. 483–485, 1967.
- 62) 関嶋政和, 高崎慎也, 中村周吾, 池口満徳, 清水謙多郎. サブタスク間の依存関係に基づくスケジューリング機構を備えた並列プログラミング環境の開発. 情報処理学会論文誌プログラミング, Vol. 41, No. SIG 2 (PRO 6), pp. 65–77, 2000.
- 63) A. Amir-Aslani, O. Mauffret, P. Bittoun, F. Sourgen, M. Monnot, E. Lescot, and S. Fermandjian. Hairpins in a DNA site for topoisomerase II studied by <sup>1</sup>H- and <sup>31</sup>P-NMR. *Nuc. Acids Res.*, Vol. 23, pp. 3850–3857, 1995.
- 64) R. A. Katz, K. Gravuer, and A. M. Skalka. A preferred target DNA structure for retroviral integrase *in vitro*. *J. Biol. Chem.*, Vol. 273, pp. 24190–24195, 1996.
- 65) J. C. Connelly, L. A. Kirkham, and D. R. Leach. The SbcCD nuclease of *escherichia coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. *Proc. Nat'l Acad. Sci. U. S. A.*, Vol. 7, pp. 7969–7974, 1998.
- 66) C. Tuerk, P. Gauss, C. Thermes, D. R. Groebe, M. Gayle, N. Guild, G. Stromo, Y. d'Aubenton-Carafa, O. C. Uhlenbeck, I. Tinoco Jr., E. N. Brody, and L. Gold. CUUCGG hairpins: Extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Nat'l Acad. Sci. U. S. A.*, Vol. 85, pp. 1364–1368, 1988.
- 67) 平尾一郎, 三浦謹一郎. 1本鎖DNAの特殊構造. 蛋白質核酸酵素, Vol. 40, pp. 1583–1591, 1995.

- 68) V. P. Antao, S. Y. Lai, and I. Tinoco Jr. A thermodynamic study of unusually stable RNA and DNA hairpins. *Nuc. Acids Res.*, Vol. 21, pp. 5901–5905, 1991.
- 69) V. P. Antao and I. Tinoco Jr. Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nuc. Acids Res.*, Vol. 20, pp. 819–824, 1992.
- 70) I. Hirao, Y. Nishimura, Y.-I. Tagawa, W. Watanabe, and K.-I. Miura. Extraordinarily stable mini-hairpins: Electrophoretical and thermal properties of the various sequence variants of d(GCGAAAGC) and their effect on DNA sequencing. *Nuc. Acids Res.*, Vol. 20, pp. 3891–3896, 1992.
- 71) 吉澤聡子, 河合剛太, 上田卓也, 石戸良治, 三浦謹一郎, 渡辺公綱, 平尾一郎. 安定なヘアピン構造を形成する d(GCGAAAGC) 断片の核酸分解酵素に対する抵抗性および細胞外蛋白質合成系への応用. 第 21 回生体分子の構造と機能に関する討論会・宮澤辰雄先生追悼シンポジウム講演要旨集, pp. 21–22, 1994.
- 72) N. Sasaki, M. Izawa, Y. Sugahara, T. Tanaka, M. Watahiki, K. Ozawa, E. Ohara, H. Funaki, Y. Yoneda, S. Matsuura, M. Muramatsu, Y. Okazaki, and Y. Hayashizaki. Identification of stable RNA hairpins causing band compressions in transcriptional sequencing and their elimination by use of inosine triphosphate. *Gene*, Vol. 222, pp. 17–23, 1998.
- 73) H. Yamakawa, D. Nakajima, and O. Ohara. Identification of sequence motifs causing band compressions on human cDNA sequencing. *DNA Res.*, Vol. 3, pp. 81–86, 1996.
- 74) C. Cheong, G. Varani, and I. Tinoco Jr. Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature*, Vol. 346, pp. 680–682, 1990.
- 75) H. A. Heus and A. Pardi. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science*, Vol. 253, pp. 191–194, 1991.
- 76) P. W. Davis, W. Thurmes, and I. Tinoco Jr. Structure of a small RNA hairpin. *Nuc. Acids Res.*, Vol. 21, pp. 537–545, 1993.



- 77) E. Ennifar, A. Nikulin, S. Tishchenko, A. Serganov, N. Nevskaya, M. Garber, B. Ehresmann, C. Ehresmann, S. Nikonov, and P. Dumas. The crystal structure of UUCG tetraloop. *J. Mol. Biol.*, Vol. 304, pp. 35–42, 2000.
- 78) C. E. Amri, O. Mauffret, M. Monnot, G. Tevanian, E. Lescot, H. Porumb, and S. Femandjian. A DNA hairpin with a single residue loop closed by a strongly distorted Watson-Crick G·C base-pair. *J. Mol. Biol.*, Vol. 294, pp. 427–442, 1999.
- 79) D. A. Erie, A. K. Suri, K. J. Breslauer, R. A. Jones, and W. K. Olson. Theoretical predictions of DNA hairpin loop conformations: Correlations with thermodynamic and spectroscopic data. *Biochemistry*, Vol. 32, pp. 436–454, 1993.
- 80) D. A. Erie, K. J. Breslauer, and W. K. Olson. A Monte Carlo method for generating structures of short single-stranded DNA sequences. *Biopolymers*, Vol. 33, pp. 75–105, 1993.
- 81) H. F. Kratky. Structural investigations on small RNA molecules using different force field methods. *Pharma. Acta Helv.*, Vol. 71, pp. 57–63, 1996.
- 82) S. B. Singh and P. A. Kollman. Understanding the thermodynamic stability of an RNA hairpin and its mutant. *Biophys. J.*, Vol. 70, pp. 1940–1948, 1996.
- 83) Z. Chraïbi, M. Réfrégiers, B. Jollès, and A. Laigle. Study of the structural stability of the mini-hairpin d(GCGAAGC) by hydrogen-deuterium exchange kinetics. *J. Raman Spectrosc.*, Vol. 31, pp. 481–484, 2000.
- 84) B. Jollès, M. Réfrégiers, and A. Laigle. Opening of the extraordinarily stable mini-hairpin d(GCGAAGC). *Nuc. Acids Res.*, Vol. 25, pp. 4608–4613, 1997.
- 85) M. Zucker. Calculating nucleic acid secondary structure. *Curr. Opin. Struct. Biol.*, Vol. 10, pp. 303–310, 2000.
- 86) L. Piela, J. Kostrowicki, and H. A. Scheraga. The multiple-minima problem in the conformational analysis of molecules. deformation of the potential energy hypersurface by the diffusion equation method. *J. Phys. Chem.*, Vol. 93, pp. 3339–3346, 1989.

- 87) J. Kostrowicki, L. Piela, B. J. Cherayil, and H. A. Scheraga. Performance of the diffusion equation method in searches for optimum structures of clusters of Lennard-Jones atoms. *J. Phys. Chem.*, Vol. 95, pp. 4113–4119, 1991.
- 88) J. Kostrowicki and H. A. Scheraga. Application of the diffusion equation method for global optimization to oligopeptides. *J. Phys. Chem.*, Vol. 96, pp. 7442–7449, 1992.
- 89) Z. Li and H. A. Scheraga. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Nat'l Acad. Sci. U. S. A.*, Vol. 84, pp. 6611–6615, 1987.
- 90) R. V. Pappu, R. K. Hart, and J. W. Ponder. Analysis and application of potential energy smoothing and search methods for global optimization. *J. Phys. Chem. B*, Vol. 102, pp. 9725–42, 1998.
- 91) R. V. Pappu, G. R. Marshall, and J. W. Ponder. A potential smoothing algorithm accurately predicts transmembrane helix packing. *Nature Struct. Biol.*, Vol. 6, pp. 50–55, 1999.
- 92) S. R. Bhaumik. Conformational feasibility of a DNA hairpin with one-base loop. *Biochem. Biophys. Res. Commun.*, Vol. 220, pp. 853–857, 1996.
- 93) A. M. Ferrenberg and R. H. Swendsen. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, Vol. 61, pp. 2635–2638, 1988.
- 94) A. M. Ferrenberg and R. H. Swendsen. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, Vol. 63, pp. 1195–1198, 1989.
- 95) E. M. Boczko and C. L. Brooks, III. Constant-temperature free energy surfaces for physical and chemical processes. *J. Phys. Chem.*, Vol. 97, pp. 4509–4513, 1993.
- 96) E. M. Boczko and C. L. Brooks, III. First-principles calculation of the folding free energy of a three-helix bundle protein. *Science*, Vol. 269, pp. 393–396, 1995.
- 97) F. B. Sheinerman and C. L. Brooks. Calculations on folding of segment B1 of streptococcal protein G. *J. Mol. Biol.*, Vol. 278, pp. 439–56, 1998.

- 98) C. Czaplewski, S. Rodziewicz-Motowidlo, A. Liwo, D. R. Ripoll, R. J. Warak, and H. A. Scheraga. Molecular simulation study of cooperativity in hydrophobic association. *Prot. Sci.*, Vol. 9, pp. 1235–45, 2000.
- 99) X. J. Kong and C. L. Brooks. Lambda-dynamics: A new approach to free energy calculations. *J. Comput. Phys.*, Vol. 105, pp. 2414–2423, 1996.
- 100) M. Ikeguchi, S. Shimizu, K. Tazaki, S. Nakamura, and K. Shimizu. Calculation of temperature dependence of free energy caused by potential function changes. *Chem. Phys. Lett.*, Vol. 288, pp. 333–337, 1998.
- 101) N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Comput. Phys.*, Vol. 21, pp. 1087–1092, 1953.
- 102) A. Mitsutake, M. Kinoshita, Y. Okamoto, and F. Hirata. Multicanonical algorithm combined with the RISM theory for simulating peptides in aqueous solution. *Chem. Phys. Lett.*, Vol. 329, pp. 295–303, 2000.
- 103) B. A. Berg and T. Neuhaus. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett.*, Vol. 68, pp. 9–12, 1992.
- 104) N. Nakajima, J. Higo, A. Kidera, and H. Nakamura. Free energy landscapes of peptides by enhanced conformational sampling. *J. Mol. Biol.*, Vol. 296, pp. 197–216, 2000.
- 105) Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, Vol. 314, pp. 141–151, 1999.
- 106) Y. Sugita, A. Kitao, and Y. Okamoto. Multidimensional replica-exchange method for free-energy calculations. *J. Comput. Phys.*, Vol. 113, pp. 6042–6051, 2000.
- 107) Y. Sugita and Y. Okamoto. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem. Phys. Lett.*, Vol. 329, pp. 261–270, 2000.
- 108) D. Söll, J. N. Abelson, and P. R. Schimmel, editors. *TRANSFER RNA: Structure, Properties, and Recognition*. Monograph Series 9A. Cold Spring Harbor Laboratory, 1979.

- 109) W. Saenger, 西村 善文訳. 核酸構造, 第 15 章. シュプリンガー・フェアラーク東京, 東京, 1987.
- 110) S.-H. Kim, G. J. Quigley, F. L. Suddath, A. McPherson, D. Sneden, J. J. Kim, J. Weinzierl, and A. Rich. Three-dimensional structure of yeast phenylalanine transfer RNA: Folding of the polynucleotide chain. *Science*, Vol. 179, pp. 285–288, 1973.
- 111) G. Eriani, M. Delarue, O. Poch, J. Gangloff, and D. Moras. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, Vol. 347, pp. 203–206, 1990.
- 112) P. J. Beuning and K. Musier-Forsyth. Transfer RNA recognition by aminoacyl-tRNA synthetases. *Biopolymers*, Vol. 52, pp. 1–28, 1999.
- 113) 濡木理. tRNA へのアミノ酸の結合. 化学工業, Vol. 50, pp. 181–189, 1999.
- 114) 濡木理, 横山茂之. アミノアシル tRNA シンテターゼの構造と tRNA 認識機構. 志村令郎, 渡辺公綱 (編), RNA 研究の最前線, pp. 70–82. シュプリンガー・フェアラーク東京, 2000.
- 115) M. A. Rould, J. J. Perona, D. Söll, and T. A. Steitz. Structure of *E. coli* glutamyl-tRNA synthetase complexed with tRNA<sup>Gln</sup> and ATP at 2.8 Å resolution. *Science*, Vol. 246, pp. 1135–1142, 1989.
- 116) O. Nureki, D. G. Vassylyev, K. Katayanagi, T. Shimizu, S. Sekine, T. Kigawa, T. Miyazawa, S. Yokoyama, and K. Morikawa. Architectures of class-defining and specific domains of glutamyl-tRNA synthetase. *Science*, Vol. 267, pp. 1958–1965, 1995.
- 117) 横山茂之. tRNA のコドン認識とアイデンティティ. 大澤省三, 志村令郎 (編), RNA の世界, pp. 85–107. 講談社, 東京, 1990.
- 118) R. Giegé, M. Sissler, and C. Florentz. Universal rules and idiosyncratic features in tRNA identity. *Nuc. Acids Res.*, Vol. 26, pp. 5017–5035, 1998.
- 119) J. G. Arnez and D. Moras. tRNA/aminoacyl-tRNA synthetase interactions. In R. W. Simons and M. Grunberg-Manago, editors, *RNA STRUCTURE AND*

- FUNCTION*, Monograph 35, pp. 465–494. Cold Spring Harbor Laboratory Press, 1998.
- 120) S. C. Harvey, M. Prabhakaran, and J. A. McCammon. Molecular-dynamics simulation of phenylalanine transfer RNA. I. methods and general results. *Biopolymers*, Vol. 24, pp. 1169–1188, 1985.
- 121) M. Prabhakaran, S. C. Harvey, and J. A. McCammon. Molecular-dynamics simulation of phenylalanine transfer RNA. II. amplitudes, anisotropies, and anharmonicities of atomic motions. *Biopolymers*, Vol. 24, pp. 1189–1204, 1985.
- 122) A. Matsumoto, M. Tomimoto, and N. Gō. Dynamical structure of transfer RNA studied by normal mode analysis. *Eur. Biophys. J. Biophys. Lett.*, Vol. 28, pp. 369–379, 1999.
- 123) P. Auffinger, S. Louise-May, and E Westhof. Molecular dynamics simulations of the anticodon hairpin of tRNA(Asp): Structuring effects of C-H  $\cdots$  O hydrogen bonds and of long-range hydration forces. *J. Am. Chem. Soc.*, Vol. 118, pp. 1181–1189, 1996.
- 124) P. Auffinger and E Westhof. RNA hydration: Three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin. *J. Mol. Biol.*, Vol. 269, pp. 326–341, 1997.
- 125) P. Auffinger, S. Louise-May, and E Westhof. Molecular dynamics simulations of solvated yeast tRNA(Asp). *Biophys. J.*, Vol. 76, pp. 50–64, 1999.
- 126) T. Simonson, G. Archontis, and M. Karplus. A Poisson-Boltzmann study of charge insertion in an enzyme active site: The effect of dielectric relaxation. *J. Phys. Chem. B*, Vol. 103, pp. 6142–6156, 1999.
- 127) T. Simonson. Electrostatic free energy calculations for macromolecules: A hybrid molecular dynamics / continuum electrostatic approach. *J. Phys. Chem. B*, Vol. 104, pp. 6509–6513, 2000.
- 128) G. Archontis, T. Simonson, and M. Karplus. Binding free energies and free energy components from molecular dynamics and Poisson-Boltzmann calculations. application to amino acid recognition by aspartyl-tRNA synthetase. *J. Mol. Biol.*, Vol. 306, pp. 307–327, 2001.

- 129) M. Tateno, O. Nureki, K. Yura, K. Morikawa, T. Noguti, S. Yokoyama, and M. Go. Molecular recognition mechanism of tRNA(Glu) by glutamyl-tRNA synthetase: A study using molecular dynamics simulation and computer modeling. *Prot. Eng.*, Vol. 8, pp. 47–47, 1995.
- 130) M. Tateno, O. Nureki, S. Sekine, K. Kaneda, M. Go, and S. Yokoyama. A three-dimensional structure model of the complex of glutamyl-tRNA synthetase and its cognate tRNA. *FEBS Lett.*, Vol. 377, pp. 77–81, 1995.
- 131) I. Bahar and R. L. Jernigan. Vibrational dynamics of transfer RNAs: Comparison of the free and synthetase-bound forms. *J. Mol. Biol.*, Vol. 281, pp. 871–884, 1998.
- 132) Y. M. Hou, E. Westhof, and R. Giegé. An unusual RNA tertiary interaction has a role for the specific aminoacylation of a transfer RNA. *Proc. Nat'l Acad. Sci. U. S. A.*, Vol. 90, pp. 6776–6780, 1993.
- 133) Y. M. Hou. Structural elements that contribute to an unusual tertiary interaction in a transfer RNA. *Biochemistry*, Vol. 33, pp. 4677–4681, 1994.
- 134) O. Nureki, T. Niimi, T. Muramatsu, H. Kanno, T. Kohno, C. Florentz, R. Giegé, and S. Yokoyama. Molecular recognition of the identity-determinant set of isoleucine transfer RNA from *Escherichia coli*. *J. Mol. Biol.*, Vol. 236, pp. 710–724, 1994.
- 135) E. T. Peterson and O. C. Uhlenbeck. Determination of recognition nucleotides for *Escherichia coli* phenylalanyl-tRNA synthetase. *Biochemistry*, Vol. 31, pp. 10380–10389, 1992.
- 136) W. H. McClain and K. Foss. Nucleotides that contribute to the identity of *Escherichia coli* tRNA<sup>Phe</sup>. *J. Mol. Biol.*, Vol. 202, pp. 679–709, 1988.
- 137) H. Asahara, H. Himeno, K. Tamura, N. Nameki, T. Hasegawa, and M. Shimizu. *Escherichia coli* seryl-tRNA synthetase recognizes tRNA(Ser) by its characteristic tertiary structure. *J. Mol. Biol.*, Vol. 236, pp. 738–748, 1994.

- 138) J. R. Sampson, A. B. DiRenzo, L. S. Behlen, and O. C. Uhlenbeck. Nucleotides in yeast tRNA<sup>Phe</sup> required for the specific recognition by its cognate synthetase. *Science*, Vol. 269, pp. 496–512, 1989.
- 139) Y.-M. Hou, H. Motegi, R. S. A. Lipman, C. S. Hamann, and K. Shiba. Conservation of a tRNA core for aminoacylation. *Nuc. Acids Res.*, Vol. 27, pp. 4743–4750, 1999.
- 140) J.-I. Sagara, S. Shimizu, T. Kawabata, S. Nakamura, M. Ikeguchi, and K. Shimizu. The use of sequence comparison to detect identities in tRNA genes. *Nuc. Acids Res.*, Vol. 26, pp. 1974–1979, 1998.
- 141) T. Christian, R. S. A. Lipman, C. Evilia, and Y.-M. Hou. Alternative design of a tRNA core for aminoacylation. *J. Mol. Biol.*, Vol. 303, pp. 503–514, 2000.
- 142) T. L. Bullock, L. D. Sherlin, and J. J. Perona. Tertiary core rearrangements in a tight binding transfer RNA aptamer. *Nature Struct. Biol.*, Vol. 6, pp. 497–504, 2000.
- 143) O. Keskin, R. L. Jernigan, and I. Bahar. Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophys. J.*, Vol. 78, pp. 2093–2106, 2000.
- 144) A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.*, Vol. 80, pp. 505–515, 2001.
- 145) N. Gō, T. Noguti, and T. Nishikawa. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Nat'l Acad. Sci. U. S. A.*, Vol. 80, pp. 3696–3700, June 1983.
- 146) T. Noguti and N. Gō. Collective variable description of small-amplitude conformational fluctuations in a globular protein. *Nature*, Vol. 296, pp. 776–778, April 22 1982.
- 147) J.-F. Gibrat and N. Gō. Normal mode analysis of human lysozyme: Study of the relative motion of the two domains and characterization of the harmonic motion. *Proteins*, Vol. 8, pp. 258–279, 1990.

- 148) A. Kitao, S. Hayward, and N. Gō. Energy landscape of a native protein: Jumping-among-minima model. *Proteins*, Vol. 33, pp. 496–517, 1999.
- 149) A. Kitao and C. Wagner. A space-time structure determination of human CD2 reveals the CD58-binding mode. *Proc. Nat'l Acad. Sci. U. S. A.*, Vol. 97, pp. 2064–2068, 2000.
- 150) W. Braun, S. Yoshioki, and N. Gō. Formulation of static and dynamic conformational energy analysis of biopolymer systems consisting of two or more molecules. *J. Phys. Soc. Jpn.*, Vol. 53, pp. 3269–3275, 1984.
- 151) J. Higo, Y. Seno, and N. Gō. Formulation of static and dynamic conformational energy analysis of biopolymer systems consisting of two or more molecules – avoiding a singularity in the previous method. *J. Phys. Soc. Jpn.*, Vol. 54, pp. 4053–4058, 1985.
- 152) M. Zacharias and H. Sklenar. Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: Application to DNA minor groove ligand complex. *J. Comput. Chem.*, Vol. 20, pp. 287–300, 1999.
- 153) M. Tomimoto and N. Gō. Analytic theory of pseudorotation in five-membered rings. cyclopentane, tetrahydrofuran, ribose, and deoxyribose. *J. Phys. Chem.*, Vol. 99, pp. 563–577, 1995.
- 154) M. Tomimoto, H. Wako, and N. Gō. Conformational analysis of nucleic acid molecules with flexible furanose rings in dihedral angle space. *J. Comput. Chem.*, Vol. 17, pp. 910–917, 1996.
- 155) S. C. Harvey, M. Prabhakaran, and J. A. McCammon. Molecular-dynamics simulation of phenylalanine transfer RNA. I. methods and general results. *Biopolymers*, Vol. 24, pp. 1169–1188, 1985.
- 156) M. Levitt, C. Sander, and P. S. Stern. Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.*, Vol. 181, pp. 423–447, 1985.



- 157) T. Nishikawa and N. Gō. Normal modes of vibration in bovine pancreatic trypsin inhibitor and its mechanical property. *Proteins*, Vol. 2, pp. 308–329, 1987.
- 158) L. Jovine, S. Djordjevic, and D. Rhodes. The crystal structure of yeast phenylalanine tRNA at 2.0 Å resolution: Cleavage by  $Mg^{2+}$  in 15-year old crystals. *J. Mol. Biol.*, Vol. 301, pp. 401–414, 2000.
- 159) S. Hayward and H. J. C. Berendsen. Systematic analysis of domain motions in proteins from conformational change; new results on citrate synthase and T4 lysozyme. *Proteins*, Vol. 30, pp. 144–154, 1998.
- 160) M. Sprinzl, C. Horn, M. Brown, A. Ioudovitch, and S. Steinberg. Compilation of tRNA sequences and sequences of tRNA genes. *Nuc. Acids Res.*, Vol. 26, pp. 148–153, 1998.
- 161) M. Sprinzl, K. S. Vassilenko, J. Emmerich, and F. Bauer. Compilation of tRNA sequences and sequences of tRNA genes, Dec 1999. <http://www.uni-bayreuth.de/departments/biochemie/trna/>
- 162) H. Ishida, Y. Jochi, and A. Kidera. Dynamic structure of subtilisin-eglin c complex studied by normal mode analysis. *Proteins*, Vol. 32, pp. 324–333, 1998.
- 163) H. Kono and A. Sarai. Structure-based prediction of DNA target sites by regulatory proteins. *Proteins*, Vol. 35, pp. 114–131, 1999.
- 164) F. Pichierri, M. Aida, M. M. Gromiha, and A. Sarai. Free-energy maps of base-amino acid interactions for DNA-protein recognition. *J. Am. Chem. Soc.*, Vol. 121, pp. 6152–6157, 1999.
- 165) K. Sayano, H. Kono, M. M. Gromiha, and A. Sarai. Multicanonical Monte Carlo calculation of the free-energy map of the base-amino acid interaction. *J. Comput. Chem.*, Vol. 21, pp. 954–962, 2000.
- 166) Y. Marcus 著, 関集三監訳, 池田重良ほか訳. マーカス 液体化学入門. 化学同人, 1983.

- 167) N. K. Rogers and M. J. E. Sternberg. Electrostatic interactions in globular proteins. different dielectric models applied to the packing of alpha-helices. *J. Mol. Biol.*, Vol. 174, pp. 527–542, 1984.
- 168) E. D. Malcolm and J. A. McCammon. Dielectric boundary smoothing in finite difference solutions of the Poisson equation: An approach to improve accuracy and convergence. *J. Comput. Chem.*, Vol. 12, , 1991.
- 169) H. Nakamura and S. Nishida. Numerical calculations of electrostatic potentials of protein-solvent systems by the self consistent boundary method. *J. Phys. Soc. Jpn.*, Vol. 56, pp. 1609–1622, 1987.
- 170) C. Eckart. Some studies concerning rotating axes and polyatomic molecules. *Phys. Rev.*, Vol. 47, pp. 552–558, 1935.