

バクテリアの複製起点開裂機構解明を目指した  
分子シミュレーション研究

清水 将裕



## 要旨

細菌の DNA 複製ではまず、複製起点の特定の領域で DNA 二重らせんがほどける。そして開裂した領域に DNA ヘリカーゼが結合し、DNA 複製が進行する。複製起点上ではじめにほどける DNA 領域は DNA unwinding element (DUE) と呼ばれる。複製起点に DNA-タンパク質複合体（複製開始複合体）を形成することが DUE の開裂に必要なものである。しかし、複製開始複合体が二重らせん開裂を促進する機構の詳細はまだ明らかになっていない。そこで大腸菌 *Escherichia coli* の複製開始複合体をモデルとし、分子動力学シミュレーションを用いてメカニズムの解明を目指した。複製開始複合体は多数のタンパク質を含む巨大な系であるため、主に粗視化分子動力学計算によるアプローチを試みた。

まず大腸菌 *E. coli* の複製開始複合体の全体構造をモデリングした。*E. coli* の複製開始複合体は、複製起点 *oriC* に 11 分子の複製開始タンパク質 DnaA や、DNA を強く曲げる integration host factor (IHF) が結合したものである。共同研究者である九州大学の片山勉教授らの生化学実験と筆者の分子動力学計算を組み合わせ、信頼性の高い構造モデルを近原子解像度で得た。特に *E. coli* の複製開始複合体は 3 領域に分かれていることが分かった。また、*oriC* 上の DnaA 結合サイトと IHF 結合サイトの間隔が複製開始複合体の形成に重要であることを見出した。さらに、左側サブ複合体と右側サブ複合体の位置関係が、DNA ヘリカーゼの一本鎖 DNA 結合効率に影響することを明らかにした。

複製開始複合体では、粗視化モデルから原子解像度モデルを構築することが可能であった。原子モデルを構築する手法が確立していれば、より高度な解析が可能になる。そこで次に、DNA-タンパク質複合体の粗視化モデルから原子モデルを構築する、より一般的な方法を検討した。はじめに粗視化 DNA モデルから原子モデルを構築する方法を考案した。これは各粗視化ヌクレオチド単位に対し、合致する原子モデルを構造ライブラリから探索し、当てはめる方法である。続いてタンパク質の原子モデル構築法を検討した。既存のソフトウェアを利用しつつ、DNA 結合面のアミノ酸側鎖に対しては実験で得られた構造情報を利用した。開発した手法はワトソン-クリック塩基対構造をよく再現できた。また、既存のソフトウェアのみを利用した場合に比べ、既知構造でみられる DNA-タンパク相互作用を高い精度で再現した。

DUE 開裂の分子機構を解明するためには一本鎖 DNA の性質を再現できる粗視化 DNA モデルが必要である。今後は、本研究で開発した粗視化 DNA モデルから原子モデルを構築する手法を駆使し、DNA 粗視化力場を開発する予定である。それらを利用し、DUE 開裂の機構を明らかにしたい。



目次	
要旨	3
序論	7
第一章 バクテリア DNA 複製開始複合体の近原子解像度の構造モデリングおよびその機能の洞察	8
1.1 研究の背景・目的	9
1.1.1 バクテリアの DNA 複製開始	9
1.1.2 大腸菌 <i>E. coli</i> の複製起点 <i>oriC</i>	9
1.1.3 大腸菌 <i>E. coli</i> の複製開始タンパク質 DnaA	10
1.1.4 大腸菌 <i>E. coli</i> の複製開始複合体	11
1.1.5 複製開始複合体に関する未解決の課題	11
1.1.6 研究目的と問題解決の方針	12
1.2 研究手法	13
1.2.1 粗視化分子動力学シミュレーション	13
1.2.2 粗視化モデルからの原子解像度モデルの構築	18
1.2.3 全原子分子動力学シミュレーション	18
1.3 結果	20
1.3.1 大腸菌 <i>E. coli</i> 複製開始複合体の形成：左側サブ複合体複製のモデリング	20
1.3.2 大腸菌 <i>E. coli</i> 複製開始複合体の形成：中間-右側サブ複合体のモデリング	23
1.3.3 複製開始複合体の原子モデル構築とその安定性の再評価	24
1.3.4 $\tau 2$ ボックスへ結合する DnaA ドメイン IV の向きの解析	25
1.3.5 複合体形成に伴う DnaA ドメイン III-IV の構造変化	26
1.3.6 IHF サイトと R1 ボックスの間隔の重要性	28
1.3.7 R2 ボックスと C3 ボックス間の距離の重要性の解析	28
1.4 考察	32
第二章 粗視化 DNA-タンパク質複合体モデルからの原子モデルの再構築	34
2.1 研究の背景と目的	35
2.1.1 粗視化計算の利点	35
2.1.2 粗視化モデルからの原子モデル再構築の必要性	35
2.1.3 原子モデル再構築に関する先行研究	35
2.1.4 本章の研究目的	36
2.2 研究手法	38
2.2.1 モデリング手法の概略	38
2.2.2 DNA フラグメントライブラリの構築	38
2.2.3 粗視化構造に基づいたワトソン-クリック塩基対形成の判定	40
2.2.4 DNA 原子モデルの構築	40

2.2.5 DNA-タンパク質複合体中のタンパク質の原子モデル構築 .....	42
2.2.6 粗視化分子動力学計算 .....	43
2.2.7 全原子分子動力学計算 .....	44
2.2.8 開発したモデリング手法の評価 .....	44
2.2.9 既存の原子モデル構築ツールによる 3SPN.2C DNA モデルからの原子モデル構築 .....	47
2.3 結果と考察 .....	48
2.3.1 自己再構築テストによる DNA 再構築手法の評価 .....	48
2.3.2 CGMD スナップショットから構築した DNA 原子モデルの評価 .....	50
2.3.3 DNA-タンパク質相互作用面のアミノ酸の構築 .....	52
2.3.4 エネルギー最小化計算による DNA 構造の改善 .....	55
2.3.5 ほかの原子モデル構築法との比較 .....	57
2.4 まとめ .....	59
結論 .....	60
発表論文 .....	61
謝辞 .....	62
付録 .....	63
付録 A 第一章の研究手法の詳細 .....	63
付録 B 複製開始複合体の詳細な構造解析と共同研究者の実験結果 .....	73
付録 C 第二章の研究手法の詳細/条件検討の手法 .....	93
付録 D 第二章の条件検討の結果と考察 .....	94
付録 E 粗視化モデルからの原子モデル構築法の検討の詳細 .....	99
付録 F 第二章で利用した構造ファイルのリスト(PDB ID) .....	118
参考文献 .....	123

## 序論

DNA 二重らせんは転写や複製といった様々な過程でほどこ、一本鎖の状態となる。多くのプロセスでは、二重鎖開裂に関わる因子は既に特定されている<sup>1-3</sup>。しかしこれらの開裂機構は、物理的な観点からは未解決の問題を含んでいる。熱による DNA の構造ゆらぎ、塩基間スタッキング相互作用の強さ、そして DNA-タンパク質相互作用といった要素が、どのように連携して二重鎖開裂を実現させているのかは、まだ詳しくわかっていない。

本研究では大腸菌 *Escherichia coli* の DNA 複製時に起こる複製起点の開裂現象を対象とし、その分子機構を研究した。

DNA 複製の開始は、複製起点の二重鎖開裂から始まる。この開裂には、複製開始複合体と呼ばれる DNA-タンパク質複合体の形成が必要である<sup>1</sup>。多くのバクテリアでは、DnaA タンパク質が開始複合体の構成要素である。大腸菌 *E. coli* では複製起点 *oriC* 上で DnaA がホモ多量体を形成し、さらに integration host factor (IHF) が特定の塩基配列に結合したものが複製開始複合体である<sup>1,4,5</sup>。

DNA の開裂機構は、バクテリアの種によって異なるモデルが提案されている<sup>6-10</sup>。大腸菌 *E. coli* では、開始複合体と DUE の相互作用により、この領域の開裂が促進される<sup>6-8</sup>。開始複合体に結合するのは DUE の特定の塩基配列であるが<sup>11</sup>、その選択性の原因は分かっていない。X 線結晶構造では DnaA は一本鎖 DNA との結合時、塩基と接触が少ない<sup>10</sup>。従って、DnaA タンパク質が特定の塩基配列へ強く結合するという機構では説明できない可能性が高い。開裂領域の位置を決める要因は、DNA の物性や、DnaA タンパク質と DNA の位置関係であるかもしれない。

筆者は、開始複合体が開裂する DNA 領域が何によって決まるか、その要因を探るべく研究を進めた。本論文は 2 章から構成される。まず第一章では、大腸菌 *E. coli* 複製開始複合体の高解像度構造モデルを構築した。また、*oriC* 上のタンパク質結合配列間隔の機能的意義についても検討した。

第一章では複製開始複合体の粗視化モデルから原子解像度モデルを構築することに成功した。原子モデルを構築する手法が確立していれば、より高度な解析が可能になる。そこで第二章では、DNA-タンパク質複合体の粗視化モデルから原子モデルを構築するより一般的な手法を開発・検討した。

結論では、第一章、第二章の結果をまとめ、今後の展望について述べる。

# 第一章 バクテリア DNA 複製開始複合体の近原子解像度の構造モデリングおよびその機能の洞察



## 1.1 研究の背景・目的

まず、大腸菌 *Escherichia coli* の DNA 複製開始複合体について解説する。続いて複合体構造に関する未解決の問題を挙げた後、本章の目的を述べる。

### 1.1.1 バクテリアの DNA 複製開始

バクテリアの DNA 複製は複製起点の二重鎖 DNA がほどけることから始まる。DNA がほどけるためには、その付近に複製開始複合体と呼ばれる DNA-タンパク質複合体を形成する必要がある<sup>1</sup>。多くのバクテリアがその構成因子として DnaA タンパク質を利用している。大腸菌 *E. coli* では複製起点 *oriC* に 11 分子の DnaA や IHF が結合し、複製開始複合体を形成する。大腸菌 *E. coli* の開始複合体は、DNA unwinding element (DUE) と呼ばれる DNA 領域と相互作用し、この領域の開裂を促進する<sup>6-8</sup>。開始複合体の働きで複製起点に一本鎖領域が生じると、DnaB ヘリカーゼがそこへ結合する。続いてレプリソームの形成が起こり、DNA 複製が進行してゆく<sup>4,12-14</sup>。

### 1.1.2 大腸菌 *E. coli* の複製起点 *oriC*

大腸菌 *E. coli* の概略図を図 1-1A にまとめた。まず大腸菌の *oriC* には DUE と呼ばれる領域がある<sup>15</sup>。この領域が複製開始時にはじめにほどける部分である。次に *oriC* 上には多数の DnaA 結合配列が存在する<sup>1,4,12-14,16-18</sup>。DnaA ボックスは 9 塩基対からなり、コンセンサス配列は 5'-TTATNCACA-3' である(N はどの塩基でもよい)<sup>19</sup>。複製開始複合体では DnaA ボックスのうち 11 箇所が DnaA で占められている。結合配列にはそれぞれ名前が付けられている (R1-2, R4, R5M, I1-3, C1-3,  $\tau$ 1-2)。また、それぞれ DnaA ボックスとの相同性や結合配列の向きが異なる(付録 B 表 SB-1)。コンセンサス配列と完全に一致している R1 ボックスや R4 ボックスは DnaA ドメイン IV との結合が強く(解離定数は R1 ボックスで 1-6 nM, R4 ボックスで 1 nM)、R2 以外のほかの DnaA ボックスでは DnaA の結合は弱い(解離定数は 200 nM より大きい)<sup>16-23</sup>。R2 での結合の強さは両者の中間程度である。最後に、*oriC* 上には IHF 結合配列がある<sup>4,20</sup>。IHF はここに結合し、DNA を強く曲げると考えられている<sup>24</sup>。

大腸菌 *E. coli* の *oriC* は左側領域と右側領域の 2 つに分けられる。右側領域に属する DnaA ボックス(R4, C1, I3, C2, C3, R2)はすべて同じ向きである。左側領域に含まれる DnaA ボックス(R1, R5M,  $\tau$ 2, I1, I2)のうち、 $\tau$ 2 以外は向きが共通している。それらは右側領域の DnaA ボックスとは逆向きである。一方、 $\tau$ 2 ボックスはコンセンサス配列との相同性が低く、その向きは不明瞭であった<sup>17,18</sup>。

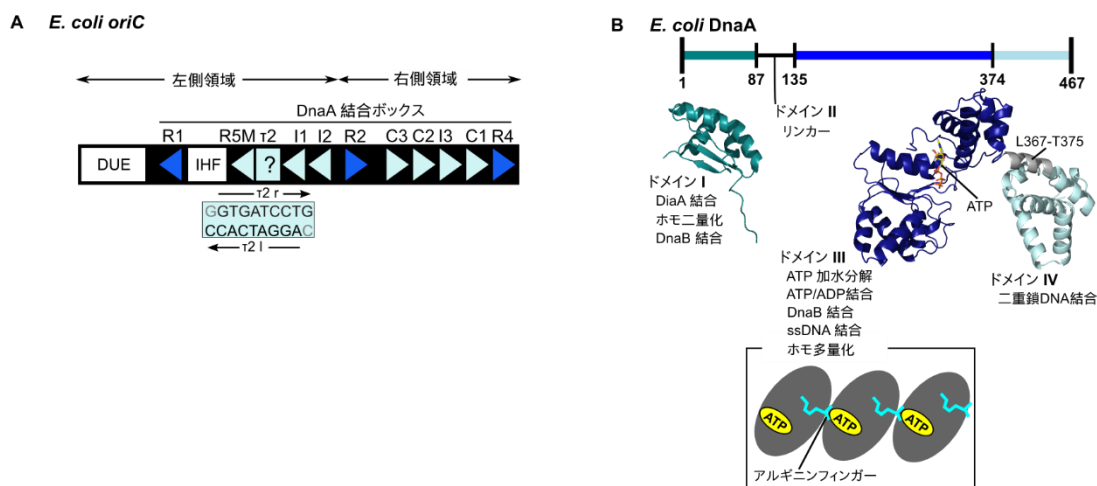


図 1-1. *oriC* と大腸菌 *E. coli* の DnaA タンパク質 A) *oriC* 領域には DUE, IHF 結合サイト(IHF)、そして 11 の DnaA ボックスがある。それぞれの DnaA ボックスについて、結合モチーフの向きが三角形で区別してある ( $\tau 2$  ボックスは除く)。青色と水色の三角形はそれぞれ結合の強い DnaA ボックス・結合の弱い DnaA ボックスを示す。 $\tau 2$  ボックスには、結合モチーフに 2 通りの可能性がある。 $\tau 2l$  (AGGATCACC)と  $\tau 2r$  (GTGATCCTG)である<sup>17,18</sup>。B) DnaA タンパク質の構造。(上段)DnaA は 4 つのドメインそれぞれについて、構造と機能を要約した。ドメイン I は、NMR によって決定された構造(PDB ID 2E0G)<sup>31</sup>を示した。ドメイン III-IV はホモロジーモデルを示した<sup>45</sup>。(下段)DnaA ドメイン III のホモ多量化形成。ATP 側の相互作用面が、別の DnaA ドメイン III のアルギニンフィンガー側の相互作用面に結合する。

なお、 $\tau 1$  と呼ばれる DnaA ボックスが R5M ボックスと IHF サイトの間にある。線状の *oriC* DNA と DnaA を混合すると、このサイトに DnaA が弱く結合する。しかし、複製開始活性な超らせん構造の DNA では、ここに DnaA は結合しない<sup>17,18</sup>。従って本研究ではこの DnaA ボックスを考慮しないことにした。

### 1.1.3 大腸菌 *E. coli* の複製開始タンパク質 DnaA

大腸菌 *E. coli* の DnaA は、4 つのドメインからなるタンパク質である(図 1-1B)<sup>4</sup>。開始複合体を形作るのは、特にドメイン III と IV の部分である。ドメイン IV は二本鎖 DNA 結合ドメインである<sup>1,4,12-14,19,25,26</sup>。9 塩基対のコンセンサス配列 (TTATNCACA) を認識する。(N は A, T, C, G のいずれでもよい。) 実際には、このモチーフと完全には一致しない配列にも結合できる。その際の結合の強さは DnaA ボックスとの相同性によって決まる。大腸菌 *E. coli* の DnaA ドメイン IV と二本鎖 DNA の複合体の構造が実験によって明らかにされている<sup>25</sup>。

ドメイン III は複数の機能をもつ AAA+ドメインであり、ATP または ADP を強く結合する<sup>1,4,12-14,27-29</sup>。ATP 結合型の DnaA(ATP-DnaA)ドメイン III は、らせん型の安定なホモ多量体を形成できる<sup>17,28</sup>。片方の DnaA ドメイン III の ATP と、他方のドメイン III のアルギニンフィンガー(R285)の相互作用をはじめとした、いくつかの特異的相互作用が重要な役割を果たす<sup>7,8,17,18,30</sup>。開始複合体中の DnaA ホモ多量体は、ATP-DnaA で構成されていると考えられている。さらに、多量化した DnaA ドメイン III は一本鎖 DNA 結合能

を有する<sup>10,11</sup>。ドメイン III はまた、ATP 加水分解能や DnaB ヘリカーゼ結合能を持つことが知られている<sup>13</sup>。好熱性真正細菌 *Aquifex aeolicus* の DnaA ドメイン III-IV のホモ多量体は、一本鎖 DNA 結合状態・非結合状態共に X 線結晶構造解析によって構造が決定されている<sup>10,28</sup>。

ドメイン II は 50 残基程度の柔軟なリンカー領域である<sup>31</sup>。ドメイン I は DnaB ヘリカーゼを結合するドメインである。また、結合は弱いですが、ドメイン I はホモ二量体を形成できることが知られている<sup>12,31</sup>。

#### 1.1.4 大腸菌 *E. coli* の複製開始複合体

*oriC* の左側領域と右側領域に、それぞれ少なくとも一ずつの DnaA ドメイン III ホモ多量体を形成する<sup>7,8,18</sup>。左側領域に形成した DnaA ホモ多量体(左側サブ複合体)が、DUE の開裂を担っている。*oriC* 右側領域の DnaA ボックスを除去しても、DUE 開裂活性は損なわれない<sup>8</sup>。左側サブ複合体と右側サブ複合体はそれぞれ DnaB ヘリカーゼ 6 量体を結合し、続いてヘリカーゼが一本鎖 DNA に結合する<sup>7,8</sup>。

開始複合体の形成機構として、次のような逐次的結合機構が知られている<sup>18,22,30</sup>。DnaA ドメイン IV は、強く結合できる R1, R2, R4 ボックスにのみ単独で安定に結合できる。ほかの DnaA ボックスは DnaA ドメイン IV との相互作用が弱く、一度結合した DnaA は容易に解離してしまう。一方、ATP-DnaA が R5M や C1 ボックスに結合した場合、ドメイン III によって隣接するボックス(R1 または R4)上の DnaA に結合できるため(図 1-1A)、解離が抑制される。こうして、DnaA が R5M や C1 ボックスに安定に結合した後は、続いて  $\tau 2$  や I3 ボックスにも同じ要領で ATP-DnaA が結合する。*oriC* 左側領域では R1 から I2 に向かって、右側領域では R4 から R2 に向かって DnaA ホモ多量体の形成が起こる。

なお、細胞周期では複製開始時のみ ATP-DnaA の比率が高く、それ以外は ADP-DnaA の比率が高い<sup>32</sup>。逐次的結合機構は ATP-DnaA でしか起こらないため、ATP-DnaA の比率の変化が複合体形成を制御している可能性がある。

#### 1.1.5 複製開始複合体に関する未解決の課題

複製開始複合体の構造を実験により原子解像度で決定したという報告はない。従って、複製開始複合体の全体像の解明が未解決の問題である。

ただ、DnaA ドメイン III-IV ホモ多量体、DnaA ドメイン IV と DNA の複合体、IHF と DNA の複合体など、各構成因子は X 線結晶構造解析によって構造が決定されている。複製開始複合体の構造で、これらの部分構造からは明らかでない箇所は以下の 3 つである。

A) DnaA ドメイン III と IV の位置関係。DnaA ドメイン III-IV は結晶構造のまま

は二本鎖 DNA を結合するスペースがない。また、構造を変えなければ DNA を結合しながら DnaA ドメイン III で多量化することができない。従って、開始複合体ではドメイン III と IV の位置関係が結晶構造とは異なっているはずである<sup>9,27,28</sup>。

- B)  $\tau 2$  ボックスでの DnaA ドメイン IV の向き。なお本研究では、ほかの *oriC* 左側領域の DnaA ボックスと同じ向きだとした場合の  $\tau 2$  ボックスを  $\tau 2l$  ボックスと呼ぶことにする。また、*oriC* 右側領域の DnaA ボックスと同じ向きだとした場合の  $\tau 2$  ボックスを  $\tau 2r$  ボックスと呼ぶことにする。
- C) R2 ボックス上の DnaA のドメイン III がホモ多量体に含まれているかどうか。ほかのほとんどの DnaA ボックスの間隔が 2 塩基対なのに対し、*oriC* 右側領域では R2 ボックスの隣の C3 ボックスは 20 塩基対離れている。なお、DnaA ボックスの間隔が 2 塩基対の時、DnaA-DnaA 相互作用は最も起こりやすい<sup>18</sup>。

### 1.1.6 研究目的と問題解決の方針

そこで本章では、複製開始複合体に関する以上の 3 つの問題 (A-C) を解決し、複製開始複合体の全体構造を明らかにすることを第一の目的とした。開始複合体は多数のタンパク質を含む巨大な系であるため、主に粗視化シミュレーションを使い研究を進める。粗視化シミュレーションは、様々な構造を効率よくモデリングできる利点がある<sup>33-37</sup>。筆者は主に先行研究で確立された方法に従う<sup>37,38</sup>。シミュレーションモデルには、X 線結晶構造解析による構造情報をはじめ、先行研究の実験結果を取り入れる<sup>10,19,24,25,30</sup>。最後に粗視化計算で得たモデルから原子解像度のモデルを構築し、安定性を評価する。

併せて、モデリングした構造を利用しながら、タンパク質結合サイトの間隔の機能的な意義についても検討する。

本研究は筆者の分子シミュレーションと、共同研究者である九州大学の片山勉教授のグループの生化学実験を組み合わせで行った。シミュレーション結果を生化学実験により検証したほか、逆に構造モデルに基づき生化学実験結果を解釈した。本論文では、主に筆者が担当したシミュレーション研究についてまとめるが、必要に応じて共同研究者の生化学実験結果についても説明する。

## 1.2 研究手法

ここでは筆者が担当したシミュレーション研究の手法をまとめる。ただしこの節が煩雑になるのを防ぐため、分子モデルの構築・シミュレーション方法の細部は付録 A にまとめた。解析に用いた指標の厳密な定義も付録 A にまとめた。

### 1.2.1 粗視化分子動力学シミュレーション

粗視化分子動力学シミュレーションには、ソフトウェア CafeMol を使った<sup>37</sup>。

#### 1.2.1.1 タンパク質モデル

使用したタンパク質のシミュレーションモデルは原子間相互作用依拠粗視化モデル (AICG2+) である<sup>39,40</sup>。このモデルでは各アミノ酸は一粒子で表現される。粒子は C $\alpha$  原子の位置に置かれる。エネルギー関数は構造依拠ポテンシャルと呼ばれるものである。そこでは、指定した参照構造を再現するポテンシャルが働く。AICG2+モデルでは、X線結晶構造解析などで得られた原子解像度の三次元構造を参照構造とし、ポテンシャルのパラメータを決定する。本研究ではタンパク質に結合したリガンドは陽に扱わなかった。つまり、大腸菌 *E. coli* の ATP-DnaA のタンパク質部分を参照構造としたものを ATP-DnaA の粗視化モデルとみなした。エネルギー関数を概略すると以下のとおりである。

$$V_{\text{AICG2+}} = V_{\text{結合長}} + V_{\text{フレキシブル・ローカル}} + V_{\text{結合角}} + V_{\text{二面角}} + V_{\text{接触粒子間}} + V_{\text{排除体積}}$$

$V_{\text{結合長}}$  はアミノ酸配列上で隣接するアミノ酸間に働くポテンシャルである。両者が仮想的なばねによって結ばれているとみなす。ポテンシャル関数は以下のものである。

$$V_{\text{結合長}}^{\text{AICG2+}} = \sum_{I=1}^{N-1} k_b (r^I - r_0^I)^2$$

ここで  $r^I$  は、I 番目と I+1 番目のアミノ酸粒子間距離である。 $r_0^I$  は参照構造中での対応する距離である。パラメータである  $k_b$  は 110.40 kcal/mol/Å<sup>2</sup> とした。

続く 2 番目の項はフレキシブル・ローカルポテンシャルと呼ばれるものである<sup>41</sup>。二次構造をとらないポリペプチド鎖の柔軟さを表現するためのものである。タンパク質の構造データベースを利用した統計ポテンシャルである。詳細は付録 A”フレキシブル・ローカルポテンシャル”にまとめた。

第 3 項から第 5 項は参照構造を再現するためのポテンシャルである。そのうち、第 3 項と第 4 項はアミノ酸配列上で近くにあるアミノ酸粒子間に働くポテンシャルである。第 3 項は次の式で表される。

$$V_{\text{結合角}} = \sum_{J=I+2} \varepsilon_{loc}^{IJ} \exp\left(-\frac{(r^{IJ} - r_0^{IJ})^2}{2W_{IJ}^2}\right)$$

ここで、 $\varepsilon$  は  $I$  と  $J$  の組ごとに異なるパラメータである。 $W_{IJ}$  はガウス型ポテンシャル関数の幅を決めるパラメータで、こちらは  $I$  と  $J$  の組に依存しない。 $r^{IJ}$  は  $I$  番目の粒子と  $J$  番目の粒子間の距離である。また、 $r_0^{IJ}$  は対応する距離の参照構造中での値である。

第 4 項は次の式で与えられる。

$$V_{\text{二面角}} = \sum_{J=I+3} \varepsilon_{loc}^{IJ} \exp\left(-\frac{(\varphi^{IJ} - \varphi_0^{IJ})^2}{2W_{\varphi,IJ}^2}\right)$$

第 3 項と同様、 $\varepsilon$  は  $I$  と  $J$  の組ごとに異なるパラメータである。また、パラメータ  $W_{IJ}$  は  $I$  と  $J$  の組に依存しない。 $\varphi^{IJ}$  は  $I$  番目の粒子から  $J$  番目の粒子の 4 粒子がなす二面角である。また、 $\varphi_0^{IJ}$  は参照構造中での値である。

第 5 項は参照構造中で接触しているアミノ酸間の相互作用である。すでに第 3,4 項に含まれているアミノ酸ペアにはこのポテンシャルは働かない。参照構造中で 2 つの重原子間距離が 6.5 Å 以内の時、その原子が属しているアミノ酸ペアを接触しているとみなした。これらのアミノ酸ペアに対し、次の式で表されるポテンシャルが働く。

$$V_{\text{接触粒子間}} = \sum_{I>J+3}^{\text{接触アミノ酸ペア}} \varepsilon_{nloc}^{IJ} \left[ 5 \left( \frac{r_0^{IJ}}{r^{IJ}} \right)^{12} - 6 \left( \frac{r_0^{IJ}}{r^{IJ}} \right)^{10} \right]$$

ここで  $\varepsilon_{nloc}^{IJ}$  は領域によって異なる値をとるパラメータである。デフォルト値は Li らによって決定されている<sup>39,40</sup>。本研究では DnaA ドメイン III 同士のホモ多量化部分のみ、パラメータはデフォルト値の 2 倍に設定した。これは、DnaA ドメイン III 同士のホモ多量化を促進する目的である。なお、 $r^{IJ}$  などの定義は第 3 項と同じである。

最後の項は排除体積効果である。この項は第 3, 4, 5 項でポテンシャルが与えられていないアミノ酸ペア（非接触ペアとする）に働く。

$$V_{\text{排除体積}} = \sum_{I>J+3}^{\text{非接触ペア}} \varepsilon_{ex} \left( \frac{C}{r^{IJ}} \right)^{12}$$

パラメータは、 $\varepsilon_{ex} = 0.6 \text{ kcal/mol}$  で、 $C = 4.0 \text{ \AA}$  である。

本研究では 2 種類の ATP-DnaA ドメイン III-IV モデルをテストしている。片方のモデルをヘリックス境界モデルとよび、もう片方をループ境界モデルと呼ぶことにする。前者は参照構造を再現する構造依拠ポテンシャル(第 3-5 項)をすべての領域にわたって与えたモデルである。後者は DnaA ドメイン III-IV 境界(L367-T375)の構造依拠ポテンシャルを消し、この領域をフレキシブル・ローカルポテンシャルと排除体積効果のみの柔軟な領域にしたものである。なお、参照構造中で L367-T375 と接触している部分と

L367-T375 との接触粒子間ポテンシャルも消してある。接触粒子間ポテンシャルが働かなくなったペアには、代わりに排除体積ポテンシャルが定義されている。

### 1.2.1.2 粗視化シミュレーションの DNA モデル

第一章では De Pablo らのグループが開発した 3SPN.1 モデルを利用した<sup>42</sup>。このモデルでは、各ヌクレオチドが 3 つの粗視化粒子で表現される。この 3 粒子はそれぞれ、糖・リン酸・塩基に対応する。糖とリン酸では、対応する原子団の重心にそれぞれ粒子を置く。プリン塩基は N1 原子の位置に、ピリミジン塩基は N3 原子の位置に粒子を置く。このモデルは塩基配列に依存した DNA の融解温度や、持続長をよく再現する。ポテンシャルは次の式に要約できる。このモデルは B 型 DNA を参照構造とし、それを再現する構造依拠ポテンシャルの一種である。

$$V_{DNA,3SPN.1} = V_{結合長}^{DNA} + V_{結合角}^{DNA} + V_{二面角}^{DNA} + V_{分子内塩基スタッキング}^{DNA} + V_{塩基対}^{DNA} + V_{排除体積}^{DNA} + V_{静電相互作用}^{DNA} + V_{排除体積}^{DNA} \quad (2)$$

粗視化粒子間の仮想的な結合の長さ、結合角、二面角に関しては以下のエネルギー関数が定義されている。

$$V_{結合長}^{DNA} = \sum_I [k_1^{DNA}(r^I - r_0^I)^2 + k_2^{DNA}(r^I - r_0^I)^4]$$

$$V_{結合角}^{DNA} = \sum_I k_a^{DNA}(\theta^I - \theta_0^I)^2$$

$$V_{二面角}^{DNA} = \sum_I k_\varphi^{DNA}[1 - \cos(\varphi^I - \varphi_0^I)]$$

まず  $r^I$ ,  $\theta^I$ ,  $\varphi^I$  はそれぞれ、結合しているとみなした粗視化粒子間の距離、結合角、二面角の値である。 $r_0^I$  などは、対応する変数の B 型 DNA の状態での値である。 $k_1^{DNA}$  などはパラメータである。

残りの項はそれぞれ以下の式によって表される。分子内の塩基同士のスタッキングエネルギーは次の式によって与えられる。

$$V_{分子内塩基スタッキング}^{DNA} = \sum_{I < J} 4\varepsilon \left[ \left( \frac{\sigma_{IJ}^{DNA}}{r^{IJ}} \right)^{12} - \left( \frac{\sigma_{IJ}^{DNA}}{r^{IJ}} \right)^6 \right]$$

式中の  $\varepsilon$  はレナードジョーンズ型ポテンシャル関数の、ポテンシャルの強さを決めるパラメータである。 $r^{IJ}$  は I 番目の粒子と J 番目の粒子の距離を表す。 $r^{IJ}$  の最安定値はパラメータ  $\sigma_{IJ}^{DNA}$  が決めている。

ワトソン-クリック塩基対形成に関するポテンシャル  $V_{塩基対}^{DNA}$  はアデニン粒子とチミン

粒子ペア、グアニン粒子とシトシン粒子ペアに働く。次の式によって表される。

$$V_{\text{塩基対}}^{DNA} = \sum_{I < J} 4\varepsilon_{bij}^{DNA} \left[ 5 \left( \frac{\sigma_{bij}^{DNA}}{r^{IJ}} \right)^{12} - 6 \left( \frac{\sigma_{bij}^{DNA}}{r^{IJ}} \right)^{10} \right]$$

式中の $\varepsilon_{bij}^{DNA}$ ,  $\sigma_{bij}^{DNA}$ はアデニン-チミンペアとグアニン-シトシンペアで異なる定数である。

排除体積効果は下の式で表される。これは、塩基スタッキングも塩基対粗以後作用も働かない粒子ペアに適用される。

$$V_{\text{排除体積}}^{DNA} = \sum_{I < J} \begin{cases} 4\varepsilon \left[ \left( \frac{\sigma_0^{DNA}}{r^{IJ}} \right)^{12} - \left( \frac{\sigma_0^{DNA}}{r^{IJ}} \right)^6 \right] + \varepsilon (r^{IJ} < r_{\text{coff}}) \\ 0 (r^{IJ} \geq r_{\text{coff}}) \end{cases}$$

ここで、 $r_{\text{coff}}$ はワトソン-クリック塩基対を形成しない塩基ペアでは 1.00 Å, それ以外の粒子ペアでは 6.86 Å である。

静電相互作用は次のデバイ-ヒュッケルの式に従う。

$$V_{\text{静電相互作用}} = \sum_{I < J} \frac{q_I q_J}{4\pi\varepsilon_0\varepsilon_k r^{IJ}} e^{-r^{IJ}/\kappa_D}$$

水の比誘電率 $\varepsilon_k = 78.0$ である。 $q_I$ と $q_J$ は粒子の電荷である。 $\kappa_D$ はデバイ長であり、次の式により与えられる。

$$\kappa_D = \left( \frac{\varepsilon_0 \varepsilon_k k_B T}{2N_A e^2 I} \right)^2$$

温度は 300 K とした。 $k_B$ はボルツマン定数、 $N_A$ はアボガドロ数である。分母の  $e$  は素電荷であり、 $I$  は溶液のイオン強度を表す。本研究では 200 mM NaCl の溶液条件( $I = 0.2$ )でシミュレーションを行った。

最後の項は溶媒からの影響である。DNA 変性に伴う水分子の再配向に関する多体効果の項である。この効果は引力として寄与している。

$$V_{\text{溶媒効果}}^{DNA} = \sum_{I < J} \varepsilon_s^{DNA} \left[ 1 - e^{-\alpha(r^{IJ} - r_s^{DNA})} \right]^2$$

パラメータ $\varepsilon_s^{DNA}$ はイオン強度に依存する値である。残りの $\alpha$ と $r_s^{DNA}$ は定数である

### 1.1.2.3 DNA とタンパク質の間に働く相互作用

エネルギー関数は次の式で表される。

$V_{DNA\text{-タンパク質}}$

$$= \sum_{I \in DNA, J \in \text{タンパク質}}^{\text{接触ペア}} \varepsilon^{IJ} \left[ 5 \left( \frac{r_0^{IJ}}{r^{IJ}} \right)^{12} - 6 \left( \frac{r_0^{IJ}}{r^{IJ}} \right)^{10} \right] + V_{\text{排除体積}} + V_{\text{静電相互作用}}$$



第 1 項はタンパク質モデルの $V_{\text{接触粒子間}}$ と同様のものである。これは非格子型郷モデル

に類似したポテンシャルでもある<sup>43</sup>。DNA-タンパク質複合体の参照構造に基づいて $r_0^{IJ}$ を定義した。このポテンシャルは参照構造中で DNA とタンパク質が接触している部分の構造を再現させる。先行研究に従い、糖または塩基粒子とアミノ酸の間のみこのポテンシャルを定義した<sup>44</sup>。本研究では DnaA ドメイン IV や IHF とその結合サイトの間にこのポテンシャルを定義した。本研究では DnaA が一箇所の結合サイトに留まり続ける状況を想定している。そのため各 DnaA ドメイン IV は、それぞれ一つの DnaA ボックスとの間のみ、このポテンシャルが定義されている。

DNA と IHF の相互作用には $\epsilon^{IJ} = 1.2 \text{ kcal/mol}$ を用いた。ここでは、十分な強さの引力を与えることで参照構造に近い構造を維持させた。DNA と DnaA ドメイン IV に関しては解離定数の実験値を再現するよう、パラメータを調節した<sup>19</sup>。その詳細は付録 A ”粗視化分子シミュレーションの詳細”にまとめた。パラメータを調節した結果、DnaA との結合が強い R1, R4 と R2 ボックスと DnaA ドメイン IV 間では、 $\epsilon = 0.36 \text{ kcal/mol}$ を用いることにした。残りの結合が弱い DnaA ボックスと DnaA ドメイン IV の間では $\epsilon = 0.315 \text{ kcal/mol}$ とした。

第二項は排除体積効果であり、タンパク質モデルの $V_{\text{排除体積}}$ と同様の関数である。第 3 項は静電相互作用についてのポテンシャルで、DNA モデルの $V_{\text{静電相互作用}}$ と同様のものである。リン酸粒子の電荷は DNA 内相互作用と同様に-1 である。タンパク質ではリジンとアルギニンが+1 の電荷をもち、アスパラギン酸とグルタミン酸には-1 の電荷を与えた。それ以外の粒子の電荷は 0 である。

#### 1.2.1.4 タンパク質モデル及び DNA-タンパク質相互作用の参照構造

ATP 結合型の DnaA ドメイン III-IV は I-TASSER を用いたホモロジーモデリングによって構築した<sup>45-47</sup>。完成したモデルは好熱菌 *Aquifex aeolicus* の ADP 結合型 DnaA ドメイン III-IV 結晶構造に類似していた<sup>27</sup>。DnaA ドメイン IV-DNA 複合体、IHF-DNA 複合体については既知構造<sup>24,25</sup>を部分的に編集し、参照構造とした。操作の詳細は付録 A “粗視化シミュレーションの参照構造の準備”にまとめた。

#### 1.2.1.5 シミュレーション

摩擦係数  $\gamma = 0.02$  の条件で、ランジュバン方程式に従いシミュレーションを行った。系の温度は 300K に設定した。時間発展アルゴリズムの詳細は CafeMol のマニュアルに記載されたものに従っている<sup>48</sup>。各シミュレーションの内容の詳細は付録 A ”粗視化分子シミュレーションの詳細”にまとめてある。

## 1.2.2 粗視化モデルからの原子解像度モデルの構築

ここでは、第一章で行った原子解像度モデル構築の概略を示す。詳細は付録 A “全原子シミュレーションの初期構造の作成”にまとめてある。なお、ここで用いた方法は第二章で確立した手法とは異なっている。DNA に対しては、ソフトウェア NAB を利用し、B 型 DNA でのヌクレオチド単位(糖・リン酸・塩基の 3 粗視化粒子分に相当)の原子モデルを作成した<sup>49,50</sup>。そしてこの構造単位を粗視化モデルに重ね合わせた。この操作をすべての粗視化ヌクレオチド(糖・リン酸・塩基の組)について実行した。タンパク質は、参照構造を重ね合わせる事が可能な場合には参照構造を重ね合わせた。DnaA ドメイン III-IV 間のリンカーなど、参照構造が適用できない部分は Phyre2 モデリングサーバーの one to one threading mode を利用して原子モデルを構築した<sup>51</sup>。立体反発が大きいと考えられる領域のアミノ酸側鎖は SCWRL4<sup>52</sup>や PyMOL<sup>53</sup>を利用して構築し直した。

## 1.2.3 全原子分子動力学シミュレーション

全原子分子動力学計算はシミュレーションソフトウェア GROMACS 4.5.5 または GROMACS 5.1.1 を利用して行った<sup>54-61</sup>。DNA とタンパク質には、AMBER99SB 力場に ParmBSC0 核酸パラメータを適用したものをを用いた<sup>62-65</sup>。ATP の力場は GROMACS のウェブサイト (<http://redmine.gromacs.org/issues/721>) からダウンロードした。水のモデルは TIP3P を使用した<sup>66</sup>。周期境界条件を持つ三斜晶系のシミュレーションボックスで計算を行った。左側サブ複合体のシミュレーションボックスは 5 組の ATP, Mg<sup>2+</sup>, DnaA ドメイン III-IV のセット、990 個のナトリウムイオン Na<sup>+</sup>, 785 個の塩化物イオン Cl<sup>-</sup>, そして 199,080 分子の水を含む。中間-右側サブ複合体のシミュレーションボックスは 6 組の ATP, Mg<sup>2+</sup>, DnaA ドメイン III-IV のセット、1,435 個のナトリウムイオン Na<sup>+</sup>, 1,255 個の塩化物イオン Cl<sup>-</sup>, そして 336,379 分子の水を含む。DnaA ドメイン III を 6 量化させた右側複合体の系では 6 組の ATP, Mg<sup>2+</sup>, DnaA ドメイン III-IV のセット、1,088 個のナトリウムイオン Na<sup>+</sup>, 922 個の塩化物イオン Cl<sup>-</sup>, そして 235,717 分子の水を含む。ナトリウムイオンと塩化物イオンは、溶液が 0.2 M NaCl 条件となるように添加してある。静電相互作用の計算には粒子メッシュエヴァルド法を用いた<sup>67</sup>。

はじめに最急降下法によりエネルギー最小化を行なった。続いて 300K で粒子数・体積・温度一定条件(NVT)で平衡化を 100 ピコ秒実行した。さらに圧力を 1bar に設定し、粒子数・圧力・温度一定条件化(NPT)で平衡化を 1 ナノ秒実施した。圧力の制御にはパリネロ-ラマンの方法を用いた<sup>68</sup>。

NVT 平衡化、NPT 平衡化の際、水分子、ナトリウムイオン、塩化物イオン以外の重原子の座標を拘束した。シミュレーションの時間発展には確率的ダイナミクスを用いた<sup>69</sup>。これは熱浴としての働きも担っている。摩擦係数の逆数を 2.0 ピコ秒に設定した。シミュレーションの 1 ステップは 2 フェムト秒とした。また、すべての共有結合長は LINCS によって拘束した<sup>70</sup>。ただし、水分子については SETTLE により拘束した

71。

平衡化の後、続いて重原子の位置に関する拘束を外し、10 ナノ秒の MD シミュレーションを行った。シミュレーションの設定は NPT 平衡化と同様である。結合長に関する拘束は引き続き適用した。

## 1.3 結果

### 1.3.1 大腸菌 *E. coli* 複製開始複合体の形成：左側サブ複合体複製のモデリング

本研究の複合体形成シミュレーションは複合体の構造を解明することを目的としている。複合体の形成過程を観察するシミュレーションではない。DNA にすべてのタンパク質を結合させた状態を初期構造とし、シミュレーションで ATP-DnaA ドメイン III 同士を結合させる<sup>17,28</sup>。これは[背景と目的]にまとめた逐次的結合機構とは異なるが、複合体の構造解明にはこの設定で十分である。なお *oriC* 左側領域、右側領域共に、隣接する DnaA 同士が結合することと、R1 ボックスと R4 ボックス上の DnaA は複合体中で ATP 側を露出している実験事実に基づき、DnaA ホモ多量化の相互作用を定義した<sup>30</sup>。

まず初めに、ATP-DnaA ドメイン III-IV 単量体の溶液中での挙動をシミュレーションした。大腸菌 *E. coli* の DnaA ドメイン III-IV 構造はまだ決定されていないため、ホモロジーモデルを作成し、シミュレーションではそれを参照構造(天然構造)とした。このモデルは、*Aquifex aeolicus* の DnaA ドメイン III-IV 結晶構造 (PDB ID : 1A1Q)によく似ていた<sup>27</sup>。全領域にわたって参照構造依拠ポテンシャル(1.2.1.1 節)を与えたモデル(ヘリックス境界モデル)では、構造の揺らぎはあるものの、参照構造に近い構造しか見られなかった。ドメイン III-IV 境界領域(L367-T375)を柔軟なループにしたモデル(ループ境界モデル、1.2.1.1 節参照)では、設定通りドメイン III と IV の位置関係が様々に変化した。続いて左側サブ複合体のモデリングシミュレーションを行った。シミュレーションの系は DUE を除く *oriC* 左側領域、IHF, 5 分子の ATP-DnaA ドメイン III-IV からなる。DnaA 分子は実際には 4 ドメインからなるが、複製開始複合体の構造を決定するのはこの 2 ドメインである。シミュレーションではまず、2 種類の指標によって複合体形成をモニターした。一つ目は  $N_{pp}$  で、DnaA ドメイン III 同士の結合が何か所で達成されているかを表す整数値である。左側では DnaA は 5 量体を形成するから、計 4 か所のドメイン III 間相互作用面がある。従ってこの値は 0 以上 4 以下となる。もう一つの指標は  $N_{pd}$  で、DnaA ドメイン IV が結合している DnaA ボックスの数を示す。すべての DnaA が結合した  $N_{pd} = 5$  の状態からシミュレーションを始めるが、途中で DnaA が解離すればこの値が減少する。 $N_{pp} = 4$  かつ  $N_{pd} = 5$  を満たしたとき複合体が完成したとみなす。なお、 $N_{pp}$  の厳密な定義は付録 A ”シミュレーション結果の解析” にまとめている。

シミュレーションは 4 種類の設定で行った。まず、DnaA のモデルにヘリックス境界モデルを用いるか、ループ境界モデルを用いるかの 2 通りがある。さらに、 $\tau 2$  ボックスへの結合の仕方が  $\tau 2l$  (向きがほかの左側 DnaA ボックスと同じ、図 1-1A)と仮定して相互作用を定義した設定と、 $\tau 2r$  (向きが右側 DnaA ボックスと同じ、図 1-1A) と仮定して相互作用を定義した場合の 2 通りがある。組み合わせはこの  $2 \times 2 = 4$  通りである。なお、 $\tau 2$  ボックスは  $\tau 2l$  または  $\tau 2r$  の片方向のみの結合ポテンシャルを設定した。どちらも、

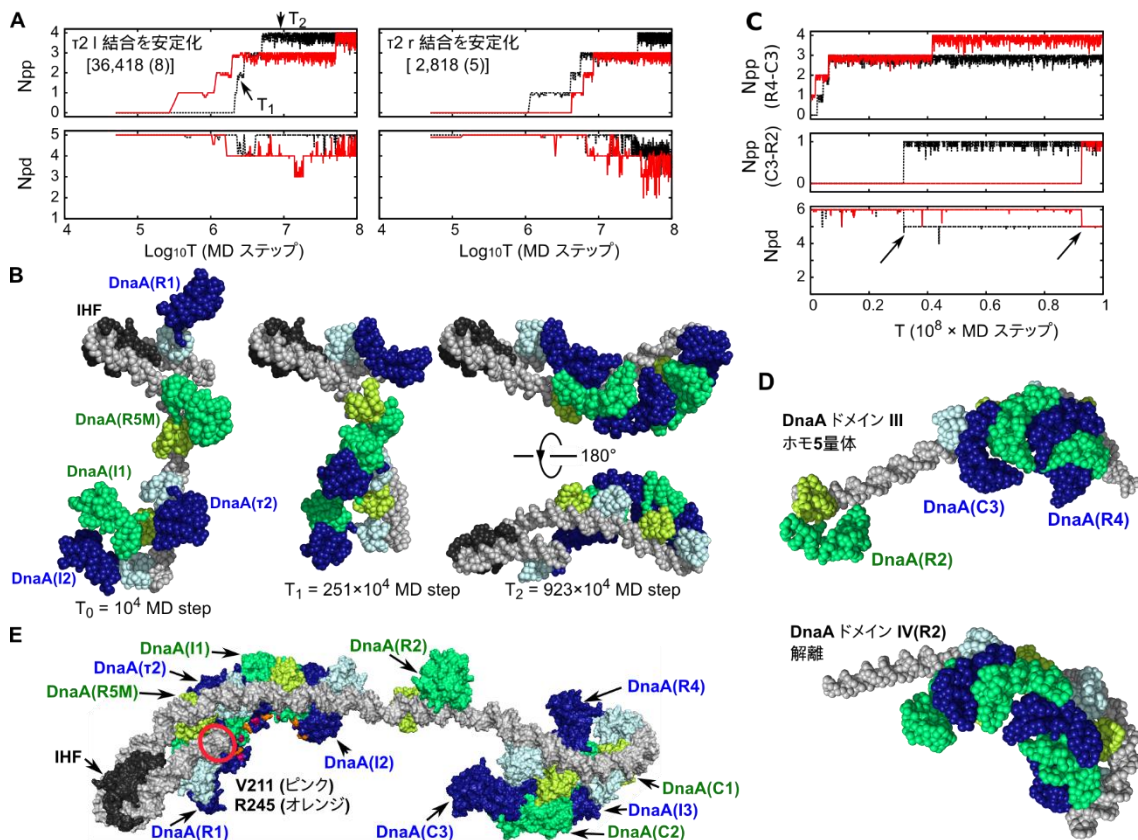


図1-2. 既知構造に基づいた複製開始複合体の構造モデリング A) *oriC* 左側領域の *oriC*-DnaA-IHF 複合体形成の軌跡を示した。τ2 l 結合を安定化 [36,418 (8)]、τ2 r 結合を安定化 [2,818 (5)] の設定ごとに2本ずつのシミュレーション結果を示した。DnaA ドメイン III 同士が結合した数(*Npp*, 上段)、DnaA ボックスに結合している DnaA ドメイン IV の数(*Npd*, 下段)を示した。それぞれ、前後を含めた9構造の平均値がプロットしてある。B) A (左側、黒)のトラジェクトリの分子構造。このシミュレーションでは DnaA を τ2l ボックスに結合させた。DNA は灰色、IHF は黒で示した。ATP 結合型 DnaA は2通りの着色法を交互に用いた。ドメイン III を濃い青色、ドメイン IV を水色で着色した分子と、ドメイン III を緑、ドメイン IV を黄緑で着色した分子がある。スナップショットを取り出した時刻も示した。右の図では異なる2方向から見た同じ複合体を示した。左側、中央、右上上段の3構造は IHF の向きを統一して示した。C) *oriC* 右側領域の *oriC*-DnaA サブ複合体形成の軌跡。代表的な2つのシミュレーション結果を示した。この図では *Npp*(R4-C3)は5分子の DnaA(R4, C1, I3, C2, C3 ボックスの DnaA)間でのドメイン III ホモ多量化の数である。*Npp*(C3-R2)は C3 ボックスの DnaA と R2 ボックスの DnaA が結合しているかを示している。*Npp*(R4-C3)、*Npp*(C3-R2)、*Npd* はそれぞれ、前後を含めた9構造の平均値をプロットした。D) Cの赤色のトラジェクトリで現れた構造を示した。(上段)DnaA ドメイン III がホモ5量体を形成した状態。(下段)シミュレーション終了後の構造。両構造は、R2 側の DNA 末端をそろえて図示してある。E) 全原子分子動力学計算後、左側サブ複合体と中間-右側サブ複合体と接続したモデル。分子の色は B と同様である。V211 と R245 は一本鎖 DNA 結合に重要な残基である<sup>11</sup>。左側サブ複合体のみ両アミノ酸を示した。*oriC* 左側の DNA 末端を赤い丸で囲んで示した。なおカッコ内の記号は DnaA が結合している DnaA ボックスの名前を示す。

DnaA ボックスのコンセンサス配列とある程度一致している<sup>17,18</sup>。すなわち τ2l に結合させた設定では、τ2r 方向特異的な結合ポテンシャルはない。ただし、DnaA の向きに関わらず DNA のリン酸粒子とアルギニン・リジンとの静電相互作用による引力は存在する。ヘリックス境界モデルの DnaA を用いて複合体形成のシミュレーションを行った場合、τ2 ボックスの向きに関わらずサブ複合体は完成しなかった。多くのトラジェクトリで、DnaA ドメイン III-IV の DNA からの解離が見られた。先行研究では、DnaA ドメイン IV

が DNA を結合するとき、DnaA ドメイン III と IV の位置関係が変化することを指摘している<sup>9,27,28</sup>。つまり、DnaA ドメイン III と IV の位置関係が変化しにくいヘリックス境界モデルでは、DNA を安定に結合できない可能性があった。シミュレーション結果は先行研究の指摘と合致しており、DNA 結合、または複製開始複合体形成に伴う DnaA ドメイン III-IV の構造変化の必要性を示している。

そこでループ境界モデルを用いたシミュレーションを行った。ドメイン III-IV 境界のうち、L366-T375 を柔軟なループにしたことは、次の根拠に拠っている。A) ホモロジーモデリングの際、この領域の二次構造予測結果がループであったことと<sup>45</sup>、B) この領域にはトリプシンによる分解を受けやすい箇所が存在することである<sup>72</sup>。

図 1-2A には、 $N_{pp}$  と  $N_{pd}$  の時間発展を示した。シミュレーションは各条件で 10 本ずつ行っているが、そのうち 2 本ずつの結果を示した。 $\tau 2l$  の場合には、両方とも最終的には複合体が完成している。図 1-2B には、黒で示したトラジェクトリでの  $N_{pp} = 0$  ( $t = 0$ ), 2 ( $t = T_1$ ), 4 ( $t = T_2$ ) の状態の 3 構造を示した。複合体を形成したときには、 $t = T_2$  の図が示す通り、DnaA ドメイン III の多量体に沿う形で DNA は少し曲がった。計 10 回のシミュレーションのうち 8 回で複合体の完成が確認できた。シミュレーションの全スナップショットのうちで複合体を形成していたものの割合は 36% である(付録 B 表 SB-2)。図 1-2A で見られるように、一時的に複合体完成条件を満たさない構造に戻った時にも、すぐに相互作用が復活した。

他方、 $\tau 2$  への結合を  $\tau 2r$  に設定した場合には、複合体完成条件を満たすスナップショットの割合は著しく低下し、 $\tau 2l$  の場合の 1/10 程度(2.8%)であった(付録 B 表 SB-2)。また計 10 回のシミュレーションのうち複合体の完成が見られたのは 5 回であった。したがって、*oriC* 左側領域の DnaA ボックスの向きがすべてそろっているほうがサブ複合体を形成しやすいといえる。実際、DnaA ボックスの向きがそろっているほうが DnaA 多量体を形成しやすいという先行研究がある<sup>18</sup>。筆者の計算はそれに合致する。また、ループ境界モデルでのみ複合体完成が見られたことは、複製開始複合体の形成時、このループを介した構造変化が重要であることを示している。

得られた左側サブ複合体の中から代表構造を選出し、残りの構造を代表構造に重ね合わせた。まず、 $\tau 2$  ボックスを  $\tau 2l$  方向に設定したときには RMSD (根平均二乗偏差)は 6 Å 付近に分布していた (付録 B 図 SB-1)。RMSD の計算には DnaA タンパク質のみを利用した。ただし L366-T375 部分も除外した。複合体の分子量が大きいことや、ドメイン III-IV 間が柔軟であることを考慮するとこの値は小さいと考えられる。実際いくつかの構造を比べるとどれも類似している(付録 B 図 SB-2)。

参照(天然)構造中の DnaA ドメイン IV-DNA 相互作用面の再現率である  $Q_{\text{box}}$  も DnaA ボックスごとに調べた(付録 B 図 SB-3)。完成した複合体のみを対象に分布を図示している。 $\tau 2r$  方向に設定した  $\tau 2$  ボックスを除き、 $Q_{\text{box}}$  は主に 0.5 から 0.9 の範囲に分布していた。これは DnaA ドメイン III-IV の一分子が R1 ボックスに結合している場合の  $Q_{\text{box}}$

の分布に類似していた。一方、 $\tau 2r$  方向に設定した  $\tau 2$  ボックスの値はほぼ 0.2 未満であった。ほかの DnaA ボックスに DnaA ドメイン IV が結合しているとき、 $\tau 2r$  ボックスと DnaA ドメイン IV の結合はあまり安定ではないことが示唆された。この傾向は図 1-2A 右パネルの  $N_{pd}$  の変化とも合致している。実際、 $\tau 2$  ボックスを  $\tau 2r$  方向に設定したときには構造はよりばらついていた(付録 B 図 SB-1, 図 SB-4)。

### 1.3.2 大腸菌 *E. coli* 複製開始複合体の形成：中間-右側サブ複合体のモデリング

続いて *oriC* 右側部分の複合体をモデリングした。この系には 6 つの DnaA ボックス(R2, C3, C2, I3, C1, R4)と 6 分子の DnaA ドメイン III-IV を含む。*oriC* 左側部分と同様、すべての DnaA ボックスに DnaA が結合した状態からシミュレーションをはじめ、DnaA ドメイン III をホモ多量化させた。このシミュレーションの目的は、DnaA の 6 量体を形成できるかどうかを調べることである。ほかの DnaA ボックス間隔の多くが 1-3 塩基対なのに対し、C3 ボックスと R2 ボックス(図 1-1A)の間隔は 20 塩基対ある。DNA 塩基配列上では、R2 ボックスがほかの 5 つの DnaA ボックスから孤立しているように見える。DnaA ボックスの向きが同じとはいえ、C3 ボックスと R2 ボックスにそれぞれ結合した DnaA 同士が結合できるかは明らかではなかった。

筆者は DnaA が 6 量化できるようドメイン III 間に引力ポテンシャルを与え、複合体の形成をシミュレーションした。このシミュレーションでは  $N_{pp}$  の最大値は 5 であり、 $N_{pd}$  の最大値は 6 である。R2 ボックス上の DnaA の挙動を調べるため、 $N_{pp}$  をさらに 2 つに分割した。 $N_{pp}$  (R4-C3)は、R4 から C3 ボックスまでの 5 分子の DnaA のホモ多量化を追跡する指標である。ホモ 5 量体になったときに最大値 4 をとる。 $N_{pp}$  (C3-R2)は R2 上の DnaA ドメイン III が C3 上の DnaA のドメイン III に結合したとき 1, それ以外では 0 である。先ほどと同様、10 回のシミュレーションのうち 2 回のデータを図 1-2C に示した。シミュレーションでは  $N_{pp}$  (R4-C3)は  $N_{pp}$  (R4-C3)=4, つまり R2 上の DnaA 以外で 5 量体を形成できることが分かった(図 1-2C,付録 B 表 SB-3)。しかしその一方、 $N_{pp}$  (C3-R2)が 1 になったときには毎回  $N_{pd}$  の低下が伴った。構造を確認したところ、これは DnaA ドメイン IV の R2 ボックスからの解離であった。したがって *oriC* 右側領域と DnaA ドメイン III-IV のシステムでは DnaA のホモ 5 量体を形成できるが、ホモ 6 量体は形成できないことが示唆された。R2 ボックス上の DnaA はドメイン III のホモ多量体には加わらず単独で存在していると考えられる。

この複合体から代表構造を選び図 1-2D に示した。左側サブ複合体の場合と同様に、DnaA ホモ多量体の部分で DNA が曲がっている。なお、10 回のシミュレーションのうちで R2 ボックス以外の DnaA が 5 量体を形成したものは 7 本あった。その一方、6 量体は極度に不安定であり、その完成条件を満たした構造はわずかであった。(付録 B 表 SB-3)

以上のシミュレーションでは R2 ボックス上 DnaA と C3 ボックス上 DnaA 間に引力を

定義して行った。この設定が R2 ボックス以外での DnaA ホモ 5 量体形成に影響した可能性を排除するため、この引力なしでのシミュレーションを行った。このシミュレーションでは、記録した全スナップショットのうち 62% で DnaA の 5 量体が完成していた(付録 B 表 SB-3)。全スナップショットには複合体が形成するまでの過程が含まれていることを踏まえると、これは高い値である。よって DnaA の 5 量体は安定だといえる。左側サブ複合体と同様、代表構造とそれ以外の構造とは互いに類似していた(RMSD が 6 Å 付近, 付録 B 図 SB-5)。DnaA ボックスと DnaA ドメイン IV の Q<sub>box</sub> の分布は、DnaA ドメイン III-IV 単量体が R1 ボックスに結合している場合の分布に類似していた(付録 B 図 SB-6)。

次に共同研究者の生化学実験結果の結果を要約する。R2 ボックス上の DnaA ドメイン III が C3 ボックス上の DnaA ドメイン III と結合するならば、R2 ボックスへの DnaA 結合は DnaA のヌクレオチド状態(ATP 型か、ADP 型か)に依存して変化するはずである。ATP-DnaA を *oriC* と混合した場合のみ、C3 ボックスに結合した DnaA ドメイン III が R2 ボックスへの ATP-DnaA の結合を促進するためである。背景で説明した通り、DnaA ドメイン III のホモ多量化は ATP-DnaA でのみ起こる。共同研究者の実験の結果、ヌクレオチド依存的な違いは現れなかった。ただし共同研究者の実験からは、R2 上の DnaA 単量体と右側サブ複合体を架橋する DnaA 分子がわずかに現れる可能性が示唆された。

ここまでの結果と、R2 ボックスへの DnaA の結合にはほかの DnaA 分子の協同性がないという先行研究に基づき<sup>16,17</sup>、サブ複合体を再定義した。R2 ボックスに結合した DnaA を中間サブ複合体と呼ぶことにし、R4-C3 ボックスに結合した DnaA のホモ 5 量体を右側サブ複合体と呼ぶことにする。左側サブ複合体と右側サブ複合体を比較すると、右側サブ複合体のほうで DNA は強く曲げられていた。

### 1.3.3 複製開始複合体の原子モデル構築とその安定性の再評価

粗視化シミュレーションでモデリングした複合体から原子モデルを再構築し、短時間の全原子分子動力学シミュレーションを行った。シミュレーションは左側サブ複合体と、中間-右側サブ複合体それぞれで行った。シミュレーション中に大きな構造変化が見られなかったことから、粗視化計算で得たモデルは不安定ではないことが分かった(付録 B 図 SB-7)。シミュレーション後の 2 構造をつなぎ合わせ *oriC*-DnaA-IHF 複合体の全体を構築したものが図 1-2E である。中間サブ複合体は、右側サブ複合体よりも左側サブ複合体の近くにある。図 1-2E に円で示した DnaA 末端は、DnaA の V211 や R245 など一本鎖 DNA 結合領域のほうを向いている<sup>11</sup>。この DNA 末端の先には DUE がある。この構造モデルは、DUE の開裂が左側サブ複合体と DUE との相互作用によって促進されるという実験事実と合致している。

一方 *oriC* 右側領域に DnaA ドメイン III-IV の 6 量体を形成した状態からシミュレーションを開始すると複合体は壊れてしまった。具体的には、C2 ボックスの DnaA と C3 ボ



ックスの DnaA の結合が維持できず、解離が起こった。粗視化シミュレーションと同様、*oriC* 右側領域に結合した DnaA が 6 量体を形成することは困難であることが示された(付録 B 図 SB-7)。

### 1.3.4 $\tau 2$ ボックスへ結合する DnaA ドメイン IV の向きの解析

続いて、 $\tau 2$  ボックスでの DnaA ドメイン IV 結合方向を決定する分子メカニズムをさらに解析した。前述のとおり、DnaA を  $\tau 2r$  に結合させる設定にした場合には、左側サブ複合体で DnaA と  $\tau 2$  ボックスとの相互作用は不十分であった。一方、 $\tau 2l$  に結合させたときには、DnaA は  $\tau 2l$  ボックスに安定に結合していた。複合体の構造を観察すると、DnaA を  $\tau 2r$  に結合させた場合でも、DnaA ドメイン IV の向きが反転し  $\tau 2l$  結合時と同じ向きになっているものが見つかった(付録 B 図 SB-4)。つまりこれらの複合体では  $\tau 2r$  ボックスに基づいた  $Q_{\text{box}}$  は低いが見かけ上は  $\tau 2l$  に DnaA を結合させたときと似た構造になっていた。そこで  $\tau 2$  ボックス上の DnaA ドメイン IV の向きを定量した。筆者は  $\theta_{\text{DnaA-}\tau 2}$  ( $0^\circ \leq \theta_{\text{DnaA-}\tau 2} \leq 180^\circ$ ) を定義し、これを DnaA ドメイン IV の向きの指標とした(図 1-3A)。 $\theta_{\text{DnaA-}\tau 2}$  が  $0^\circ$  付近の時、DnaA ドメイン IV の向きは  $\tau 2l$  結合時と同じである。逆に  $180^\circ$  付近の時は  $\tau 2r$  結合時の向きである。 $\tau 2$  ボックスの向きを  $\tau 2r$  に設定したシミュレーションでは、サブ複合体の 70%程度で  $\tau 2$  ボックスの DnaA ドメイン IV が反転していた(図 1-3B)。

シミュレーションから、DnaA ドメイン IV は  $\tau 2l$  に結合するほうが  $\tau 2r$  に結合するよりも安定であることが分かった。また、左側サブ複合体での DnaA ドメイン III ホモ多量化が DnaA ドメイン IV と  $\tau 2r$  ボックスとの相互作用の障害になることも分かった。DnaA ドメイン III がホモ 5 量体を形成したことで DnaA ドメイン IV が反転した、という計算結果は、DnaA ドメイン III が DnaA ドメイン IV-DNA 間の適切な相互作用を補助しているとも解釈できる。

続いて共同研究者の生化学実験の内容を要約する。 $\tau 2$  ボックスでの DnaA の結合の向きを決定するため、P1 ヌクレアーゼアッセイにより左側サブ複合体の DUE 開裂活性を測定した。DUE を開裂する役割は、左側サブ複合体が持っている<sup>8,23</sup>。中間-右側サブ複合体は DUE 開裂にはかかわらない。従って左側サブ複合体の形成能が低下した場合は、DUE 開裂活性の低下として現れると考えられる。

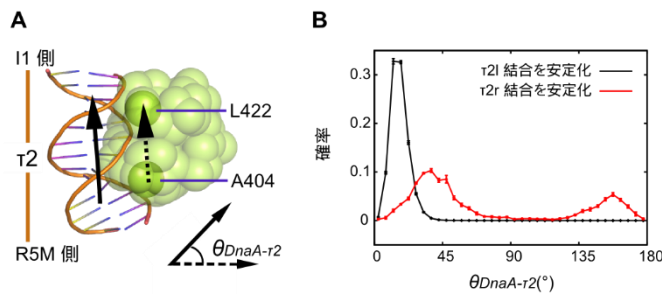


図 1-3. DnaA ドメイン IV の  $\tau 2$  ボックスへの結合の向き A) 角度  $\theta_{DnaA-\tau 2}$  を 2 つのベクトルによって定義した。ひとつは  $\tau 2$  ボックスの DNA 軸に沿ったベクトル(実線矢印)である。もう一つは DnaA ドメイン IV 内で定義したベクトル(点線矢印)である。B) 形成した複製開始複合体モデル中の角度  $\theta_{DnaA-\tau 2}$  の分布を、 $\tau 2$  ボックスの結合の設定ごとにプロットした。エラーバーは標本標準偏差を示す。

$\tau 21$  ボックスと DnaA との結合を低下させた変異体と、 $\tau 2r$  ボックスと DnaA と結合を弱めた変異体プラスミドをそれぞれ作成し、DUE 開裂活性を測定した。前者の変異体のみで活性低下がみられた(付録 B 図 SB-8)。これは  $\tau 2$  ボックスの方向が  $\tau 21$  であることを示している。ただしこの変異体での活性低下は緩やかであり、DUE 開裂活性は十分に残っていた。これは  $\tau 21$  ボックスへの DnaA ドメイン IV の結合が、単に両者の結合の強さだけで決まるのではないことを示唆している。複合体全体としてより安定な構造をとろうとする効果が、 $\tau 21$  への DnaA 結合促進として現れたと考えられる。この結果はシミュレーションと一致している。

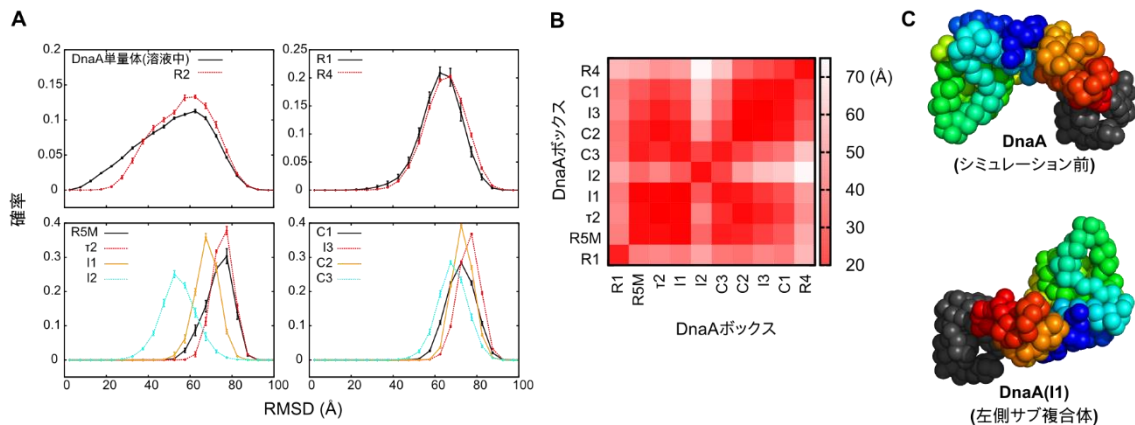
### 1.3.5 複合体形成に伴う DnaA ドメイン III-IV の構造変化

X 線結晶構造解析の結果から、DnaA のドメイン III とドメイン IV の位置関係は、二本鎖 DNA 結合時と複製開始複合体形成時にそれぞれ変化すると考えられてきた<sup>27,28</sup>。筆者のシミュレーションでも、ループ境界モデルの DnaA を用いた場合のみサブ複合体を形成した。シミュレーション結果から、複製開始複合体形成時にドメイン III-IV の構造がどう変化したかを解析した。

はじめに二本鎖 DNA 結合に伴う DnaA ドメイン III-IV の構造変化を調べた。これは DnaA ドメイン III がホモ多量化する前の状態である。しかし、以下の理由で構造変化の詳細を解析することはできなかった。

*E.coli* DnaA ホモロジーモデルに類似している *A. aeolicus* DnaA ドメイン III-IV の結晶構造では、二重鎖 DNA を結合できるスペースがない<sup>27</sup>。ドメイン III と IV 位置関係を少し変えることで、DNA を結合できるスペースができる<sup>27</sup>。*E.coli* DnaA モデルを用いた筆者のシミュレーションでも、DnaA ドメイン IV を DNA に結合させるためには、少しの構造変化が必要であった(付録 B 図 SB-9A)。しかし、本研究の粗視化モデルでは側鎖の情報不十分であるため、立体障害の効果を過小評価してしまう。また、DnaA ドメイン III-IV 間のリンカーのアミノ酸配列は *E. coli* と *A. aeolicus* では異なっているが、I-TASSER のホモロジーモデルでは *A. aeolicus* の DnaA 結晶構造を参照する結果になっていた。そのため、この部分のホモロジーモデルの精度は信頼できない。

一方、DnaA ドメイン III のホモ多量化に伴うドメイン III-IV の構造変化は、シミュレ



**図 1-4. 複合体中の DnaA 分子の構造解析** 複製開始複合体での DnaA の構造変化を DnaA ボックスごとに調べた。A) 各スナップショットの DnaA ドメイン III と、参照構造の DnaA ドメイン III の間の RMSD を求めた。RMSD は分子のドメイン IV 同士を重ね合わせた後で計算した。エラーバーは標本標準偏差を表す。B) DnaA ボックスごとの DnaA 分子の構造比較。それぞれの DnaA ボックスペアで、DnaA ドメイン IV から見た DnaA ドメイン III の相対位置を比較した。赤色が濃いほど 2 つの DnaA ボックスに結合した DnaA の構造が互いに類似している。C) シミュレーション前の参照構造(上)と DnaA ホモ 5 量体中での代表的な DnaA ドメイン III-IV の構造(下)の比較。複合体中の DnaA 構造は I1 ボックスの DnaA の例を示した。この 2 構造はドメイン IV(灰色)の向きをそろえて図示した。ドメイン III は K135(青色)から L373(赤色)へと、虹色に変化するよう着色した。

ーション中にはっきりと観測された。複合体中の DnaA 分子それぞれについて、ドメイン IV を参照構造と重ね合わせた状態で、ドメイン III と、参照構造のドメイン III との RMSD を計算した(図 1-4A)。コントロール実験として、溶液中で DNA に結合していない DnaA 単量体(ループ境界モデル)についても同様の解析をした。また、DNA 上で単量体である R2 ボックスの DnaA に対しても、同様に RMSD を計算した。DnaA のドメイン III-IV 間の構造変化は、DnaA ホモ 5 量体中のもののほうが明らかに大きかった。また、RMSD の分布には興味深い傾向が見られた。左側サブ複合体と右側サブ複合体で、対応する位置にある DnaA の RMSD の分布は互いに似ていた。たとえば、ATP 側の結合面を露出した R1 上 DnaA と R4 上 DnaA の分布は類似している。ほかにも  $\tau 2$  と I3, I1 と C2, I2 と C3 のペアで、結合している DnaA の RMSD の分布がよく似ていた。RMSD はホモ 5 量体の中間にある 3 つの DnaA で、より大きかった。さらに、DnaA ドメイン IV を基準としたときの DnaA ドメイン III の幾何重心の位置を、DnaA ボックス間で比較した(図 1-4B)。その結果、DnaA ドメイン IV から見たドメイン III の方向は、それぞれのホモ 5 量体の中央の 3 分子では類似していた。一方で、末端に位置する R1, I2, C3 上の DnaA ドメイン III は他とは違う方向を向いていた。

代表的なスナップショットを図 1-4C と付録 B 図 SB-9B に示した。ホモ 5 量体の中央の 3 分子の構造変化は、*A. aeolicus* の複製開始複合体形成時に起こると予想されたものと同様であった<sup>9</sup>。一方でホモ 5 量体の末端では固有の構造をとる。これは、左側サブ複合体と右側サブ複合体では DnaA の構造が違うことを示唆する、共同研究者らの過去の実験結果と合致する<sup>7</sup>。

### 1.3.6 IHF サイトと R1 ボックスの間隔の重要性

続いて IHF サイトと R1 ボックスの間隔の重要性を生化学実験とシミュレーションの双方によって調べた。

はじめに共同研究者の生化学実験の結果を要約する。まず IHF サイトと R1 ボックスの間の欠失変異体を持つプラスミドを作成した。全部で 4 種類、それぞれ 5, 6, 8, 10 塩基対の欠失がある。このプラスミドを用いて複製開始複合体を形成させ、DUE の開裂活性を測定した。するといずれの場合にも DUE 開裂は不活性であった(図 1-6A, 付録 B 図 SB-10)。従ってこの領域が DUE 開裂活性に重要であることが示された。*In vivo* では、この領域に 10 塩基対挿入した変異プラスミドは複製されないことを示唆する報告があり、この実験結果はそれに矛盾しない<sup>73</sup>。

つづいて筆者の計算モデリングについて説明する。まず、野生型 *oriC* の左側サブ複合体モデルで、R1 ボックス上に結合している DnaA の位置を 5, 6, 8, 10 塩基対分 IHF サイト側にずらしてみた。図 1-6B では 5 塩基対ずらしたときと 8 塩基対ずらしたときの構造を示す。8 または 10 塩基対ずらしたものでは DnaA ドメイン IV の位置が IHF と重なった。ここから、これらの変異体では IHF と R1 ボックスの DnaA は同時に結合できないことが示唆された。残りの変異体ではこうした立体障害は起こらなかった。したがってこの 2 つの変異体ではサブ複合体を形成する可能性が残されている。そこで次に、IHF サイトと R1 ボックス間の 5 塩基対欠失変異体で DnaA ホモ 5 量体を形成できるか、シミュレーションで調べた。計 10 回のシミュレーションの中で複合体の完成基準を満たすものは現れなかった。この変異体では R1 ボックス上の DnaA が R5M ボックス上の DnaA に十分に接近できなかった。従って、IHF サイトと R1 ボックスの間隔を変更した変異体の DUE 開裂活性低下は、左側サブ複合体を形成できないことに起因すると考えられる。従って、IHF 結合サイトと R1 ボックスの間隔は DUE 開裂のために厳密に設計されていることが示唆された。

### 1.3.7 R2 ボックスと C3 ボックス間の距離の重要性の解析

本節ではまず共同研究者の生化学実験結果の要点をまとめる。

*oriC* の左側サブ複合体は DUE 開裂活性に加え、一本鎖 DNA へ DnaB ヘリカーゼを装着させる活性を有している<sup>8</sup>。ただし *oriC* 全長が存在する場合には DnaB 装着活性はより高いため、中間-右側サブ複合体も一本鎖 DNA に DnaB ヘリカーゼを装着させる活性を持つと考えられる。双方向に DNA 複製を行うため、*oriC* の一本鎖 DNA 領域には 2 組の DnaB ヘリカーゼ 6 量体がそれぞれ逆向きに結合する<sup>74</sup>。そこで共同研究者は *oriC* に形成したサブ複合体が協同し、2 組のヘリカーゼを一本鎖 DNA へ装着させているのではないかと考えた<sup>8,30</sup>。そして *oriC* の R2 ボックスと C3 ボックスの間は 20 塩基対の間隔があるが、この間隔に機能的意義があるのではないかと考え、検証した。

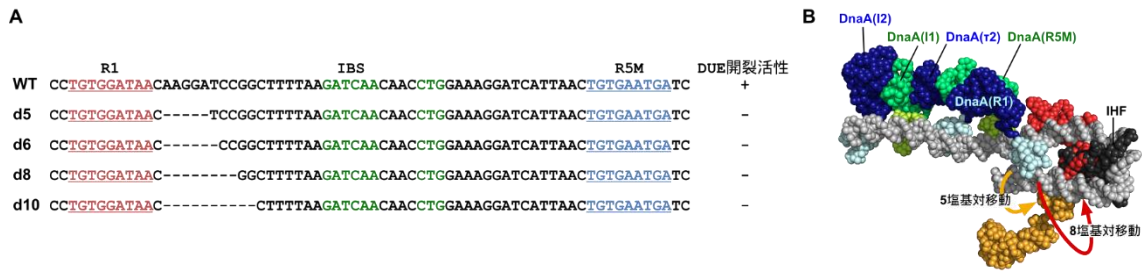
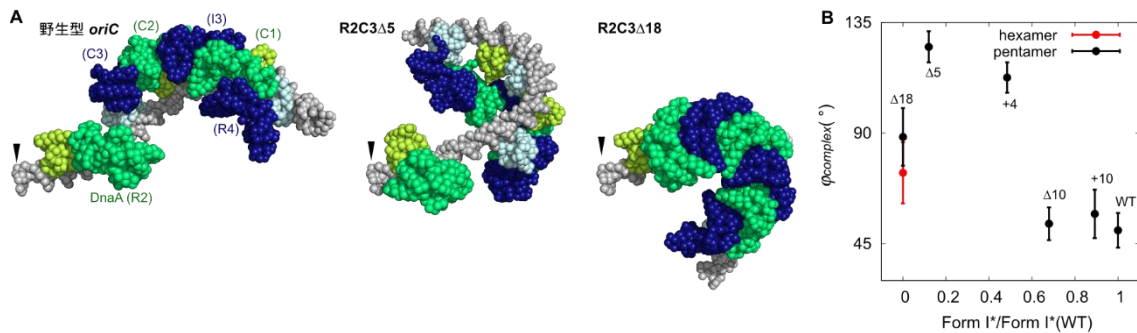


図 1-5. R1 ボックス-IHF サイト間領域に欠失変異がある *oriC* の DUE 開裂活性 A) R1 ボックスから R5M ボックスまでの *oriC* 塩基配列。野生型の配列とともに、それぞれの変異体で欠失させた領域(ハイフン)を示した。R1 ボックスは赤色、R5M ボックスは青色、IHF 結合サイトは緑色で示した。また DUE 開裂活性を右側に示した。+が野生型と同程度、-が不活性である。このパネルは共同研究者の実験データである。B) R1-IBS 領域に欠失変異を導入した場合の、R1 ボックス上の DnaA ドメイン III-IV 位置の推定。左側サブ複合体の代表構造を用い、R1 ボックス上の DnaA を R5M ボックス側に 5 塩基対(橙)または 8 塩基対(赤)、DNA のらせんに沿って移動させた。

共同研究者はこの領域に変異をもつプラスミドを作成した。5 塩基対、10 塩基対、18 塩基対欠失させた変異体プラスミドをそれぞれ PR2C3 $\Delta$ 5, PR2C3 $\Delta$ 10, PR2C3 $\Delta$ 18 と呼ぶことにする。そして 4 塩基対、10 塩基対挿入した変異体プラスミドを PR2C3+4, PR2C3+10 と呼ぶことにする。共同研究者の実験では、変異体の DUE 開裂活性は野生型と同程度であった。しかし DnaB ヘリカーゼの一本鎖 DNA 装着活性は、変異体によって違いが現れた(付録 B 図 SB-11)。PR2C3 $\Delta$ 10, PR2C3+10 では活性の低下は小さかった。一方 PR2C3 $\Delta$ 5, PR2C3 $\Delta$ 18 では活性が大幅に低下した。また PR2C3+4 を用いた場合にも活性低下がみられた。この実験より、R2 ボックスと C3 ボックスの間隔が一本鎖 DNA への DnaB 装着に重要であることが分かった。また DNA が約 10 塩基対で一周であることと、10 塩基対の挿入/欠損変異体での活性低下が少なかったことから、左側サブ複合体から見た右側サブ複合体の向きの変化が関わっていることが示唆された。他方、DnaB ヘリカーゼや DnaA の *oriC* への結合量は変異によってほとんど変わらなかった。このことから、*oriC* 変異体では一本鎖 DNA に DnaB が接近しづらくなっていると考えられた。

次に筆者のシミュレーションの内容を説明する。分子の構造の視点から R2 と C3 ボックス間隔の機能を理解するため、生化学実験で検討した *oriC* 右側領域変異体の DnaA 複合体をモデリングした。野生型の場合と同様、DnaA が 6 量体を形成できるかを調べるシミュレーションと DnaA ホモ 5 量体のみを形成させるシミュレーションを行った。ここでは PR2C3 $\Delta$ 5, PR2C3 $\Delta$ 10 などに対応する変異体をそれぞれ  $\Delta$ 5 変異体、 $\Delta$ 10 変異体などと呼ぶことにする。まず、 $\Delta$ 10,  $\Delta$ 18 変異体ではホモ 6 量体の形成が見られた。全シミュレーション時間の 50%以上で 6 量体を形成していたことから、シミュレーションモデルではこれらの複合体は安定であるといえる(付録 B 表 SB-3)。一方  $\Delta$ 5 変異体ではホモ 6 量体が見られたものの全シミュレーション時間の 1%未満であった。そして +4 変異体でホモ 6 量体が見られたトラジェクトリは 10 本中 1 本であり、+10 変異体ではホモ 6 量体は形成しなかった。なお、共同研究者のゲルシフトアッセイでは  $\Delta$ 10 変異体の R2



**図 1-6. *oriC* 右側領域の挿入/欠失変異体上の DnaA ホモ多量体構造** A. 野生型 *oriC* の中間-右側複合体・変異体の複合体の代表構造を示した。R2C3Δ5 変異体では、DnaA ホモ 5 量体の構造を示した。R2C3Δ18 変異体では DnaA ホモ 6 量体構造を示した。分子の着色は図 1-2 に従う。B) 実験による DnaB 装着活性(Form I\*活性)と、シミュレーション中の複合体構造の相関プロット。Form I\*実験では、24 nM DnaA の条件でのデータを参照した。DnaA のシミュレーション構造モデルでは  $\phi_{complex}$  を定義した。変異導入によって DnaA ホモ多量体の向きがどれだけ変わったかを示す指標である。詳細は付録 A”シミュレーション結果の解析”に示した。エラーバーは 9 回のシミュレーションの標準偏差である。

上 DnaA はほかの DnaA ボックスの DnaA とは結合しないことが示唆された。よって、シミュレーションの設定のいずれかが、Δ10 変異体のホモ 6 量体を不適切に安定化したのだと考えられる。例えばループ境界モデルのループ領域が現実の分子よりも柔軟になってしまっている可能性などが考えられる。続いて DnaA のホモ 5 量体のみを形成させるシミュレーションを行った場合は、どの変異体でもホモ 5 量体を形成し、できた 5 量体はそのまま維持された。

形成した複合体から代表構造を選び、R2 側の DNA 末端をそろえた状態で構造を比較した(図 1-6A, すべての変異体については付録 B 図 SB-12)。R2 側の末端をそろえた理由は左側サブ複合体からの右側サブ複体の相対位置を比較するためである。R2 側の末端は左側サブ複合体へと続いている。Δ10, +10 変異体の右側サブ複体の位置、向きは野生型に類似していた。ほかの変異体の DnaA ホモ 5 量体の向きは野生型とは異なっていた。つまり DnaB ヘリケース装着活性が高いものは野生型に類似した構造で、低いものは野生型と異なる構造であった。また Δ18 変異体に形成した DnaA6 量体では他と異なり DNA が強く曲げられていた。この違いがこの変異体の DnaB 装着活性に反映しているかもしれない。シミュレーションで得られた構造モデルは、視覚的には DnaB ヘリケース装着アッセイの結果ときれいに対応していた。

DnaA 多量体の野生型との構造の違いをより定量的に評価するため、角度  $\phi_{complex}$  を定義した。DnaA ホモ多量体の向きを定義しておき、野生型を基準として変異体では向きがどれだけ変わったかを示す角度が  $\phi_{complex}$  である。結果を図 1-6B に示した。コントロール実験として、野生型同士の  $\phi_{complex}$  も示してある。横軸には DnaB ヘリケース装着活性(Form I\*アッセイの活性値)、縦軸に  $\phi_{complex}$  をプロットしてある。Δ18 変異体以外では両者はよく相関している。つまり、左側サブ複合体から見た右側サブ複体の向きが野生型に近ければ近いほど DnaB を装着しやすいことが分かる。DnaB は右側サブ複合体

に結合すること、一本鎖 DNA 領域はまず左側で生じることから、右側サブ複合体の向きが野生型での向きから大きく変わったならば、DnaB ヘリカーゼが一本鎖 DNA を装着しづらくなると解釈できる。

左側サブ複合体と右側サブ複合体の位置関係に機能的な意味があることが分かった。言い換えれば2つのサブ複合体の間隔は、一本鎖 DNA に DnaB を装着しやすいようデザインされているともいえる。

## 1.4 考察

バクテリアの複製開始複合体では、DnaA がそれぞれ DnaA ボックスに結合し、かつドメイン III によってホモ多量体を形成する<sup>1,4,12-14</sup>。この複合体の個々の構成要素の構造は解明されていたが、全体構造はまだ明らかにされていなかった。本研究では分子シミュレーションと共同研究者の生化学実験を組み合わせ、この構造解明に取り組んだ。大腸菌 *E. coli* ではこの複合体は左側サブ複合体、中央の DnaA 単量体、右側サブ複合体に分かれていることが明らかになった。左側サブ複合体は DNA, IHF, DnaA ホモ 5 量体からなる。右側サブ複合体は DNA と DnaA ホモ 5 量体からなる。ただし、中央の DnaA 単量体と右側サブ複合体を架橋する DnaA 分子が現れる場合がある可能性も示された。

左側サブ複合体と右側サブ複合体の相対的な位置関係が、DnaB ヘリカーゼの一本鎖 DNA 装着に重要であることが明らかになった。複製開始複合体が開裂する DUE 領域が狭いことから、限られた一本鎖領域に DnaB を 2 組、異なる向きに結合させるために、両複合体の位置関係が精密に設計されているのかもしれない。共同研究者と筆者は、左側サブ複合体と右側サブ複合体それぞれが DnaB ヘリカーゼを一組ずつ結合することで、高度に組織だった、2 組の DnaA-DnaB-DnaC 複合体を形成するという分子機構を提案する。そしてこの複合体が DnaB の一本鎖 DNA 装着を可能にしていると考えている(付録 B 図 SB-13)。DnaA のドメイン I は主要な DnaB 結合領域を有している。複数の DnaA ドメイン I が DnaB ホモ 6 量体に結合することで、DnaB は DnaA ホモ多量体に安定に結合できる<sup>31,75</sup>。一本鎖 DNA を装着する前には DnaB ホモ 6 量体は DnaC ヘリカーゼローダーホモ 6 量体と強く結合している<sup>12</sup>。この 2 組の DnaA-DnaB-DnaC 複合体同士の距離や向きが、DnaB の一本鎖 DNA 装着に重要であると考えられる。

シミュレーションからは、R2 ボックスに結合した DnaA は単量体で存在することが示された。この結果は、共同研究者による R2-C3 領域の 10 塩基対挿入変異体を使った Form I\*アッセイ結果が野生型とあまり変わらなかったことと合致する。また本章では結果の概要をまとめたのにとどめたが、共同研究者の実験は R2 上の DnaA と C3 上の DnaA はドメイン III で結合しないことを示唆している。R1 ボックスを削除した変異体では、R2 に結合した DnaA が左側サブ複合体の形成を誘導しているとの報告がある<sup>8,76</sup>。ここでは、DnaA ドメイン I のホモ 2 量化が、DnaA ドメイン III の本多量化よりも重要な役割を果たしているのかもしれない<sup>77,78</sup>。R2 ボックスと I2 ボックスは逆向きであり、かつ離れていることから、この 2 つの DnaA ボックスに結合した DnaA 同士がドメイン III ホモ多量化を起こすことは考えにくい。

DnaA ドメイン III の R227 や L290 をアラニンに置換した変異体では、左側サブ複合体の形成能のみが低下し、右側では低下しない<sup>7</sup>。シミュレーションからは、左側サブ複合体の末端の DnaA のドメイン III-IV 位置関係が右側サブ複合体のものとは異なっていた。複製開始複合体形成のために必要な構造の違いが、おそらく DnaA ドメイン III



間の相互作用を弱める変異に対する応答への違いを説明しているのかもしれない。

研究の結果、 $\tau 2$  ボックスでの DnaA の結合の向きは周囲の DnaA ボックスと同じであることが確かめられた。これは先行研究の DMS フットプリントアッセイの結果と合致していた<sup>18</sup>。この先行研究では、 $\tau 2$  ボックスへの DnaA の結合は他の DnaA 分子との相互作用の結果促進されることを示している。DnaA ドメイン III がホモ多量体を形成する結果、 $\tau 2$  ボックス近傍で DnaA ドメイン IV が結合できる位置と結合の向きが制限されるのだと考えられる。この DnaA ドメイン III ホモ多量化を介した間接的な DnaA-DNA 結合の補助は、複製開始複合体形成の際に重要な役割を果たしているのかもしれない。

最後に、シミュレーションモデルの改善の余地について述べる。本研究では DnaA ドメイン III と IV の境界領域を柔軟なループにした<sup>41</sup>。しかし、柔軟にしすぎた可能性がある。ただ、複製開始複合体の構造はどれも類似していたことから、DnaA ドメイン IV-DNA 間相互作用と DnaA ドメイン III ホモ多量化によって複合体の概形はほぼ決まるといえる。従って 10 アミノ酸の領域の柔軟性が高くとも全体構造にはそれほど影響はないと考えられる。本研究では DnaA ドメイン IV-DNA 複合体の参照構造として DnaA ドメイン IV-R1 ボックス複合体の結晶構造を利用した。実際 DnaA ドメイン IV と、結合が弱い DnaA ボックスとの複合体構造はまだ解明されていない。本研究では同じ参照構造を用い、結合の強さを変えることで DnaA ボックスの違いを表現した。しかし実際には結合が強いサイトと結合が弱いサイトとでは、DnaA の結合の仕方が少し異なるかもしれない。

## 第二章 粗視化 DNA-タンパク質複合体モデルからの原子 モデルの再構築

## 2.1 研究の背景と目的

### 2.1.1 粗視化計算の利点

今までに多くの生体分子や生体分子複合体の構造が実験によって解明されてきた。しかしそのほとんどは安定な構造に限られている。実際の生体分子は動的であり、互いに結合や解離を繰り返している。この動的過程の中で現れる構造を直接観測することは困難である。こうした背景から、生体分子の挙動を”観察する”手段として分子動力学シミュレーション(molecular dynamics simulation, MD)は重要である。ただし現状ではシミュレーションできる時間はマイクロ秒程度の時間スケールに限られている。これは多くの場合、観察したい生命現象の時間スケールよりも短い。例えば核内では多くのタンパク質が柔軟に構造を変え、寿命の短い DNA-タンパク質複合体を様々に形成している。この過程は遺伝子制御の分子機構に深く関わっている<sup>79-81</sup>。しかし、全原子 MD で研究するには膨大なシミュレーション時間を要する。こうした、大きな生体分子複合体の一時的な準安定状態や挙動を調べるのには粗視化分子動力学計算(coarse-grained MD, CGMD)が役立つ<sup>34,36,38,82-85</sup>。CGMD では、分子の解像度を粗くすることでシミュレーションに要する時間を大幅に削減する。その結果、全原子 MD よりも精度は低い、生体分子複合体のより多様な構造を探索することが可能になる。先行研究には、CGMD により p53 タンパク質の DNA 上の運動を解析したというものがある<sup>38</sup>。また本論文の第一章では、多数のタンパク質と DNA からなる大腸菌 *E. coli* の複製開始複合体を CGMD でモデリングした。

### 2.1.2 粗視化モデルからの原子モデル再構築の必要性

CGMD 中に新たな構造が現れた時は、それが現実的かどうかを慎重に検討する必要がある。粗視化(coarse-grained, CG)分子構造の妥当性・安定性を評価する方法の一つに、CG モデルからの原子モデル再構築がある。原子モデルを構築し、それを用いて全原子 MD を行えば、より高い精度で安定性を評価できる。あるいは全原子 MD によってその構造に重要な塩橋・水素結合を予測できる。そうすれば、変異体を用いた生化学実験による検証の可能性も開ける。MD は本質的に近似であるため、その結果を実験と比較することが必要である<sup>86</sup>。CGMD の結果を様々な実験事実と比べるためには、原子解像度モデルの構築が必須である。

### 2.1.3 原子モデル再構築に関する先行研究

CG モデルからの原子モデル再構築手法は、今までにいくつも提案されてきた<sup>52,87-117</sup>。それらはホモロジーモデリング、タンパク質デザイン、マルチスケール分子シミュレーションなどの目的で開発された。主鎖だけが与えられたポリペプチド鎖に側鎖を構築す

る操作は、よく知られた再構築の例である。側鎖の配座の構造ライブラリから最適なものを選択し、配置するアプローチが主流である<sup>52,105-113</sup>。また、C $\alpha$  原子の座標のみからポリペプチドの主鎖を構築する研究も盛んにおこなわれてきた。この場合にも、多くは主鎖の断片の構造ライブラリを利用して構築される<sup>98,104,114-117</sup>。その一方で、構造ライブラリを用いずにモデリングする方法も提案されている<sup>88</sup>。

高解像度モデルの再構築手法を分子シミュレーション研究に取り入れる場合、次の2つのステップからなる。最初のステップは原子をCG粒子の周囲に配置することである。次のステップは全原子力場を用いたエネルギー最小化を実行し、より妥当な構造にすることである<sup>87-96,99-101,118</sup>。原子を配置するステップとエネルギー最小化のどちらに重点を置くかは、それぞれの手法で異なる。Rzepiela が提案した手法では、はじめに原子をCG粒子の周りにランダムに配置する。その後、全原子力場を用いた徐冷法により、構造を緩和してゆく<sup>87</sup>。Wassenaar らが提案した手法も原子の配置は厳密ではない。ただし、不適切な構造異性体が現れるのを避ける工夫をしている。続いて全原子力場を用いた構造緩和を繰り返し、原子モデル構築を達成している<sup>88</sup>。手法として独特なものは、Liu らが提案した **configuration-bias Monte Carlo** 法を用いたものが挙げられる<sup>97</sup>。

基本的には、CGモデルの一構造が対応するのは原子モデルの構造の集合であり、対応する原子モデルを一意には決定できない。CGモデルに沿って原子を配置するだけで安定な構造を得られる保証はないので、原子力場を用いた構造緩和を利用するのが妥当である。一方で、十分に解像度が高いCGモデルに対しては、幾何学的なアプローチにより妥当な原子モデルを直接得られることがある。この場合には構造緩和はそれほど重要ではない<sup>102</sup>。

#### 2.1.4 本章の研究目的

本章の目的はDNA-タンパク質複合体のCGモデルから原子モデルを構築する新しい方法を検討し、確立することである。様々なCGモデルに適用できる手法を提案するが、煩雑さを避けるため一種類のCGモデルに絞って性能の評価を行う。用いるCGモデルは、アミノ酸がC $\alpha$ 原子の位置に配置した粒子で表されるタンパク質モデルと、一ヌクレオチドが糖・リン酸・塩基の3粒子で表現されているDNAモデルである<sup>40,119,120</sup>。このモデルは分子動力学計算ソフトウェアCafeMol<sup>37</sup>で主に用いているものである。従って本研究で検討する手法はCafeMolのシミュレーション結果に直ちに適用できる。

本章で用いたCGモデルと同程度の解像度のCGモデルを用い、原子モデルの再構築を行った報告は他にもある。本研究では以下の2点に重点を置く。

- 1) DNAの再構築を高精度で高速に行うこと。
- 2) DNA-タンパク質相互作用面を精度よく再現すること。

DNA の原子モデル構築では、予め DNA 原子モデルの構造ライブラリを作成しておき、その中から CG モデルに合致するヌクレオチドを選び出し、当てはめる方法を検討する。モデリングの性能は A) 既知の結晶構造を一度粗視化し、原子モデルを再構築するテストおよび、B) DNA-タンパク質複合体の CGMD 後の構造から原子モデルを構築するテストによって評価する。B) では特に、DNA に特徴的な構造であるワトソン-クリック塩基対の再現性を評価する。CG モデルでワトソン-クリック塩基対様の構造をとっている場合、再構築した原子モデルでもワトソン-クリック塩基対の水素結合を形成していることが望ましい。

DNA-タンパク質相互作用面を正確に構築することは、その後の機能解析に重要である。分子間相互作用面は原子密度が高く、構造最適化計算で不適切な方向を向いたアミノ酸側鎖を矯正することは難しい。再構築対象の構造情報がすでに得られている場合には、それを最大限利用することで再構築モデルの精度が高まる。そこで筆者は既知の X 線結晶構造解析の構造情報を DNA-タンパク質相互作用面のアミノ酸に適用する。ほかの領域には既存のタンパク質モデル構築ソフトウェア (PD2 CA2MAIN<sup>104</sup>, SCWRL4<sup>52</sup>) を用いる。結晶構造と既存のツールを組み合わせる手法は筆者の知る限り今まで検討されたことはない。この手法を用いることで、DNA-タンパク質相互作用面の構造の正確さと、様々な構造の CG モデルに対して原子モデル構築が可能であるという汎用性の両立が期待できる。

さらに、全原子力場のエネルギー最小化計算の必要性についても検討する。

## 2.2 研究手法

### 2.2.1 モデリング手法の概略

粗視化分子動力学計算(CGMD)で得られた DNA-タンパク質複合体の粗視化(CG)モデルから原子モデルを構築する方法を検討する。CGMD から得られた構造には情報が少なく、また使う力場によっては原子レベルの構造との対応をつけづらい場合がある。例えば、CG 粒子間の仮想結合長が自然長から大きく外れている時には、CG モデルに対応し、かつ安定な原子モデルを構築することは難しくなる。そうした状況も考慮したうえで、CG モデルの構造とできるだけ矛盾せず、かつ高精度の原子モデルの構築を目指す。特に DNA-タンパク質の相互作用面を正確にモデリングすることを目指した。これは再構築した原子モデルを用いて全原子 MD を行うためには重要な点である。原子モデル構築は以下の 4 つのステップからなる。概略図を図 2-1 にも示した。

- 1) CG モデルの局所構造（結合長、結合角など）を、1 K 下の CGMD で緩和する。
- 2) DNA の原子モデルを構築する
- 3) タンパク質の原子モデルを構築する
- 4) 全原子力場を用いたエネルギー最小化により不安定な局所構造を取り除く。

議論を分かりやすくするため、本研究では特定の CG モデルを使って構築法を評価する。まず DNA では de Pablo らによって開発された 3SPN.2C モデルを使う。このモデルでは、DNA の各ヌクレオチドはそれぞれ糖・リン酸・塩基に対応する 3 粒子からなる(図 2-2A)。粒子は対応する原子団の重心に置かれている。タンパク質では、各アミノ酸が C $\alpha$  原子の位置の一粒で表現される AICG2+モデルを使う<sup>40</sup>。ただし、本研究のアプローチはほかの CG モデルにも適用可能である。

### 2.2.2 DNA フラグメントライブラリの構築

DNA 原子モデル構築の前段階として、DNA の断片の構造ライブラリを構築した。まず初めにプロテインデータベース (PDB) から、DNA を含み解像度が 2.2 Å 以内の構造をすべて(2017 年 3 月 28 日時点)ダウンロードした。続いてダウンロードした構造ファイルを 3 タイプに分類した。一つ目は DNA とタンパク質の複合体のうち、DNA が構造ファイル中に一分子しか存在しないものである。二つ目は DNA とタンパク質の複合体で、複数の DNA 鎖が存在するものである。三つめはそれ以外である。それぞれのグループの構造ファイルをランダムに半分に分割した。片方を DNA フラグメントライブラリ構築に利用した。残りの半分は再構築手法のテストに利用した。本章では、原子の占有度が 1.0 ではなく、複数のモデルが記載されている場合には、占有度が高いほうの構

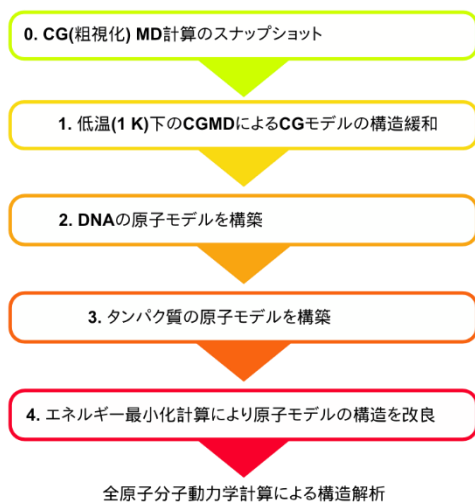


図 2-1. DNA-タンパク質複合体粗視化モデルから原子モデルを構築する手順

造を利用した。

それぞれの構造ファイルのから、デオキシリボヌクレオチド単位が 3 つ連続した構造を探し、そのうち 5' 端のヌクレオチドの糖、中央のヌクレオチド、3' 端のヌクレオチドの糖とリン酸を構造単位として抽出した。これは CG モデルの 6 粒子分に相当する。実際の原子モデル構築時には、主に 3' 側ヌクレオチドの糖を除く CG 粒子 5 つ分の構造を利用する。これらをそれぞれ糖  $S_{i-1}$ 、リン酸  $P_i$ 、糖  $S_i$ 、塩基  $B_i$ 、およびリン酸  $P_{i+1}$  と呼ぶことにする(図 2-2B)。ヌクレオチドの番号は 5' 側から 3' 側へ増大するように振ってある。残りの 3' 側の糖  $S_{i+1}$  は 5' 末端のヌクレオチドをモデリングする際に用いる。

抽出した構造単位を DNA フラグメントと呼ぶことにする。構造ファイルの中で欠損原子があるものは DNA フラグメントとみなさない。次にそれぞれの構造フラグメントに対し、CG 粒子に対応する座標(糖・塩基・リン酸の重原子の重心座標)を求めた。

続いて以下の手順に従って直交座標系を定義した。まず  $P_i$  を始点とし  $P_{i+1}$  を終点とするベクトルを X 軸に平行でかつ X 軸の正の向きとした。三点  $S_i, P_i, P_{i+1}$  はすべて XY 平面上にあるとした。そして  $S_i, P_i, P_{i+1}$  の幾何重心が原点に一致するとした(図 2-2B 参照)。この直交座標系での各 DNA フラグメントの座標をフラグメントライブラリの内容とした。フラグメントライブラリ構築に利用した構造ファイルの PDB ID は付録 F にまとめた。

作成したフラグメントライブラリは総計 22,347 構造の DNA フラグメントを含んでいる。内訳はチミン塩基のものが 5,118 構造、シトシン塩基のものが 5,940 構造、アデニン塩基のものが 4,980 構造、そしてグアニン塩基のものが 6,309 構造である。ライブラリ中のフラグメントを、ワトソン-クリック塩基対を形成していたものとそれ以外に分類した。ここではワトソン-クリック塩基対は CG 粒子の構造から判別する方法をとった(ワトソン-クリック塩基対形成の基準は後述)。ワトソン-クリック塩基対を形成していたグループをグループ A と呼ぶことにした。

### 2.2.3 粗視化構造に基づいたワトソン-クリック塩基対形成の判定

塩基の粒子  $B_i$  のペアが以下の 4 条件をすべて満たすとき、その塩基はワトソン-クリック塩基対を形成しているとみなした。

- A) アデニンとチミンのペア、またはシトシンとグアニンのペアである。
- B) 両者の距離は  $6.3 \text{ \AA}$  より小さい。
- C) 典型的な B 型 DNA で塩基対を形成しているヌクレオチドペアの糖・塩基の重心座標(4 点)を参照構造として控えておいた。B 型 DNA は NAB により生成した<sup>4950</sup>。塩基対形成を調べたい CG ヌクレオチドペアの糖・塩基の 4 点を参照構造に重ね合わせたとき、RMSD (root mean square deviation, 根平均二乗偏差)が  $0.65 \text{ \AA}$  より小さい。
- D) ピリミジン塩基が複数のプリン塩基との間で上記の 3 条件を満たした場合は、C で計算した RMSD が最小のペアのみが塩基対形成しているとみなす。

CGMD での粒子の揺らぎは、全原子 MD での対応する座標の揺らぎよりも大きい傾向にある。結果、塩基粒子間の距離が現実の塩基対形成塩基ペアの重心間距離よりも大きいにもかかわらず、塩基対を形成しているとみなすのが妥当な場合が生じる。そういった塩基ペアが排除されるのを防ぐ目的で、上記の 4 条件はいくらか緩い条件となっている (付録 E 図 SE-1)。

### 2.2.4 DNA 原子モデルの構築

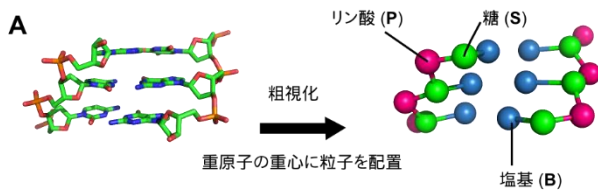
DNA の CG モデルに基づき、デオキシリボヌクレオチド単位ごとに原子モデルを構築した。各ヌクレオチドと、その周囲の CG 粒子 2 つ(5'側の隣接ヌクレオチドの糖、3'側の隣接ヌクレオチドのリン酸)を用い、ライブラリ構築時と同様の直交座標系を定義した。そしてこの座標系での CG5 粒子の座標を求めた。続いて、この CG5 粒子に最も合致する DNA フラグメントをライブラリから探した。以下の評価式 F によって構造の近さを定量し、最小の F を与えるものを探した。

$$F = \sum_{j \in \{S_{i-1}, P_i, B_i, S_i, P_{i+1}\}} w_j (\mathbf{r}_j - \mathbf{r}_{j, \text{template}})^2 \quad (1)$$

ここでパラメータ  $w_j$  が F でのそれぞれの粒子の重みを示している。 $\mathbf{r}_j$  は定義した座標系での各粒子の座標である。 $\mathbf{r}_{j, \text{template}}$  は、ライブラリ中の DNA フラグメントの対応する粒子の座標である。F を最小にするフラグメントの座標を、再構築した原子モデルのこの座標系での座標とした。最後に元の座標系に戻した。

他方 5'末端と 3'末端のヌクレオチドでは上記の方法を行うための粒子が不足している。そこで、以下の方法でモデリングをした。まず 5'端のヌクレオチドは以下の方法で





**B DNA原子モデルの構築法 (ヌクレオチド単位毎にモデリング)**

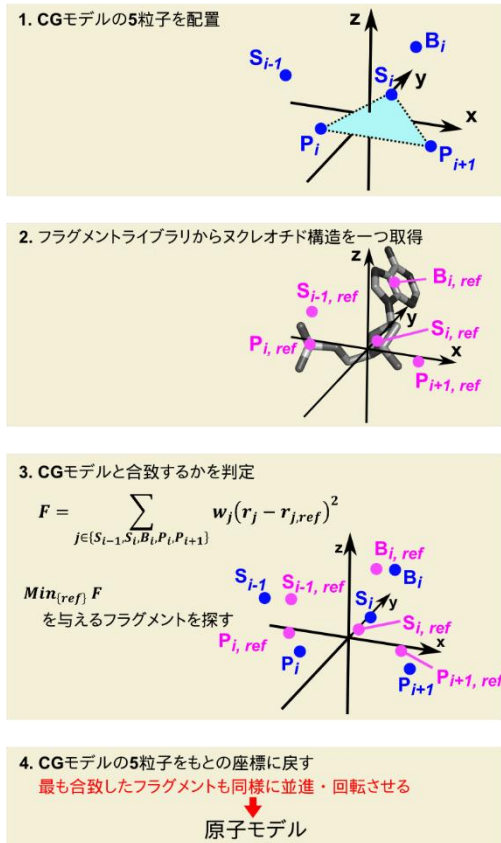


図 2-2. DNA 粗視化モデルから原子モデルを構築する手法の概略 A) 本章で用いた粗視化モデルと DNA の重原子の対応。B) DNA の原子モデル再構築の手順。

モデリングした。はじめに  $S_i$  と  $B_i$  の距離を計算した。これを  $D_{S_i-B_i}$  と呼ぶことにする。以下同様に、 $D_{S_i-P_{i+1}}$ ,  $D_{S_i-S_{i+1}}$ ,  $D_{B_i-P_{i+1}}$ ,  $D_{B_i-S_{i+1}}$ ,  $D_{P_{i+1}-S_{i+1}}$  を求めた。フラグメントライブラリから、次の  $F_5'$  を最小にする構造を求めた。

$$F_{5'} = \sum_{j \in \{S_i-B_i, S_i-P_{i+1}, S_i-S_{i+1}, B_i-P_{i+1}, B_i-S_{i+1}, P_{i+1}-S_{i+1}\}} (D_j - D_{j,ref})^2 \quad (2)$$

CG モデルの  $S_i$ ,  $B_i$ ,  $S_{i+1}$ ,  $P_{i+1}$  の 4 点に、 $F_5'$  を最小にするフラグメントの対応する CG 粒子を重ね合わせ、この末端のヌクレオチドの原子モデルとした。他方 3'末端のヌクレオチドでは  $P_i$ ,  $S_i$ ,  $B_i$ ,  $S_{i-1}$  の 4 点を使い、5'と同じ要領で  $F_3'$  を定義し、原子モデルを構築した。

なお、 $F$ ,  $F_5'$ ,  $F_3'$  を計算するのに先立ち、モデリングする CG ヌクレオチドがワトソン-クリック塩基対に関与しているかを上述 2.2.3 節の判定条件で調べた。関与してい

る場合には、グループ A のフラグメントのみを使い、モデリングを行った。それ以外の場合にはすべてのフラグメントを使って原子モデル構築を行った。

### 2.2.5 DNA-タンパク質複合体中のタンパク質の原子モデル構築

続いて CGMD 後の CG タンパク質モデルから原子モデルを構築した。タンパク質モデルでは各アミノ酸は  $C\alpha$  原子に置かれた一粒子で表現される。ここでは対象のタンパク質と DNA との複合体構造が原子レベルで解明されていることを仮定した。筆者は三つの異なる方法をテストした。一つ目は CG モデルの座標と既知構造の  $C\alpha$  原子の座標を用い、既知構造を重ね合わせる方法である。このとき既知構造は剛体として取り扱い、タンパク質全体を回転・並進させることで CG モデルに重ね合わせた。この方法を「重ね合わせ」と呼ぶ。二つ目の方法は既存のソフトウェアを利用する方法である。まず  $C\alpha$  原子の座標からタンパク質の主鎖を構築するプログラム PD2 CA2MAIN を使い、主鎖を構築した。PD2 CA2MAIN<sup>104</sup> ではエネルギー最小化を行わない方法を使った。なおこのプログラムは同時に  $C\beta$  原子の座標も与える。続いて側鎖構築ソフトウェア SCWRL4 を用いて側鎖を構築した<sup>52</sup>。SCWRL4 には、側鎖モデリング時にリガンドとの空間的重なりを避ける機能がある。そこですでに構築した DNA 原子モデルの座標を入力として与え、DNA-タンパク質間で重なりがないモデルを得た。なお、フラグメントライブラリで構築したのは DNA の重原子部分だけであるため、SCWRL4 の入力に使う前に GROMACS に含まれる gmx pdb2gmx プログラムで水素原子を生成した<sup>59,60</sup>。この方法を“PD2+SCWRL4”と呼ぶ。

最後に、両者を組み合わせたアプローチを試した。(図 2-3)。DNA-タンパク質相互作用面近傍のアミノ酸には、既知の DNA-タンパク質複合体中の側鎖の構造を適用した。アミノ酸の重原子の半径  $7.0 \text{ \AA}$  の範囲内に DNA の重原子が存在するとき、そのアミノ酸は DNA との相互作用面近傍にあると定義した。それ以外の領域に対しては PD2 CA2MAIN と SCWRL4 によってモデリングした。詳細は以下のとおりである。まず PD2 CA2MAIN でタンパク質の主鎖を構築した。PD2 CA2MAIN ではエネルギー最小化を行わない方法を使った。続いて、DNA-タンパク質複合体の既知構造のタンパク質部分を、 $C\alpha$  原子の座標を使い、CG モデルに重ね合わせた。タンパク質全体を剛体とし、並進、回転によって重ね合わせた。DNA-タンパク質相互作用面近傍と定義した各アミノ酸では、重ね合わせた既知構造の側鎖を平行移動した。それにより既知構造由来の  $C\beta$  原子の座標と、PD2 CA2MAIN で構築した  $C\beta$  原子の座標を一致させた。この操作で付加された側鎖を、CG モデルから構築したアミノ酸側鎖とした。最後に相互作用面近傍以外のアミノ酸側鎖を SCWRL4 によって構築した。この時、DNA-タンパク質相互作用面近傍のアミノ酸をすべてグリシンとして取り扱った。また、すでにモデリングした DNA・DNA-タンパク質相互作用面近傍のアミノ酸側鎖を立体的境界条件(リガンドの位置情報)として入力した。これにより構築する側鎖と既存の原子モデルとの空間的な重なり

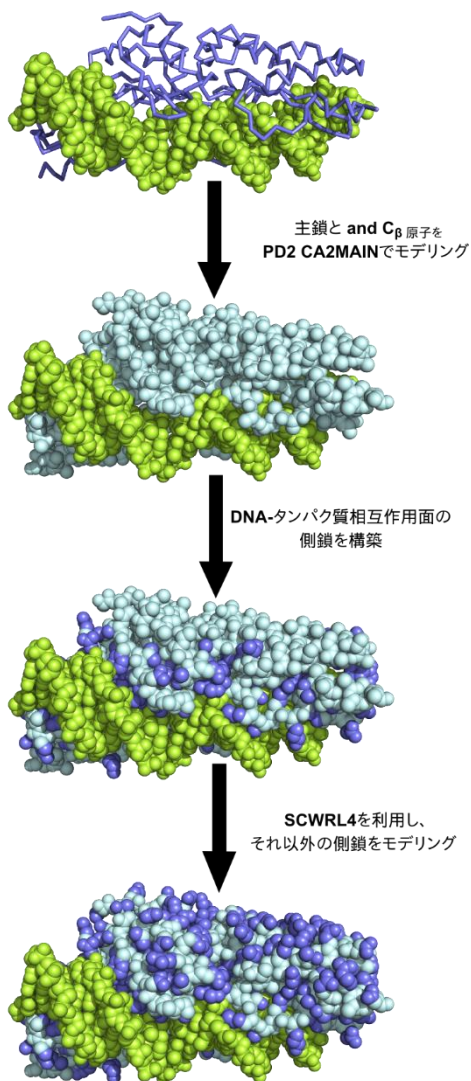


図 2-3. タンパク質の原子モデル構築の流れ 粗視化モデルはワイヤーモデルで表現した。原子モデル(重原子)は球で表現した。タンパク質の主鎖および  $C_{\beta}$  原子は明るい水色で着色した。ほかの部分は紫色で示した。DNA は黄緑色で示した。

を避けた。なおこの原子団には、リガンドの位置情報として利用する前に水素原子を付加した。水素原子の付加には `gmx pdb2gmx` プログラムを利用した。 $C_{\beta}$  原子とそれに結合した水素原子はリガンドの位置情報に含めない。

## 2.2.6 粗視化分子動力学計算

分子計算ソフトウェア CafeMol 3.0 を用いた<sup>37</sup>。タンパク質には AICG2+モデルを用い、DNA には 3SPN.2C モデルを用いた<sup>40,120,121</sup>。モデルの詳細は付録 C に記した。ランジュバン方程式に従い、シミュレーションを行った。

まず、テストセット A(後述)の 180 構造で CGMD をそれぞれ行った。まず 300 K 下で  $10^4$  MD ステップの等温 CGMD を行った。この計算後の分子構造を再構築のテストに使った。

次に再構築の最初の過程として、1 K 下で  $10^4$  MD ステップの等温 CGMD を行った(図 2-1)。それにより、CG モデルの粒子間の不自然な仮想結合長、仮想結合角を改善した。ランジュバンダイナミクスの乱数の種を変えながら各複合体のシミュレーション(上述の 2 ステップ)を 20 回ずつ行った。

それ以外にも、CG DNA モデルの構造緩和方法を検討している。この条件検討については付録 C, 付録 E 表 SE-1 にまとめてある。ここでは種々の条件の徐冷法を試した。

## 2.2.7 全原子分子動力学計算

計算はソフトウェア GROMACS 2016.3 を用いて行った<sup>59-61</sup>。行った計算はエネルギー最小化計算とそれに続く粒子数・体積・温度一定条件下(NVT)での短時間の平衡化計算である。本研究で主に用いたのは Amber parmbsc1 力場<sup>122</sup>と TIP3P 水モデル<sup>66</sup>の組み合わせである。計算結果の力場依存性を調べるため、CHARMM27 力場<sup>123-127</sup>と CHARMM TIP 3P モデル<sup>127</sup>の組み合わせも用いた。シミュレーションの系は三斜晶系の三次元領域であり、周期境界条件がある。DNA-蛋白質複合体を系に配置した後は、系を水分子で満たした。さらに一部の水分子をナトリウムイオンや塩化物イオンと置換し、系の電荷をゼロにするとともに水溶液を 0.1 M NaCl の条件に設定した。静電相互作用の計算には粒子メッシュエヴァルド法を用いた<sup>67</sup>。実空間の静電相互作用のカットオフは 1.0 nm とした。レナード-ジョーンズ型のポテンシャルのカットオフも 1.0 nm とした。エネルギー最小化は最急降下法を用いて行った。その後、300 K の温度条件で NVT 平衡化計算を 10 ピコ秒実施した。シミュレーションの一ステップは 2 フェムト秒とした。すべての共有結合は LINCS によって拘束した<sup>70</sup>。水については SETTLE を用いて拘束した<sup>71</sup>。熱浴には Bussi らが開発した方法を用いた<sup>128</sup>。温度カップリングのパラメータ  $\tau$  は 0.1 ピコ秒とした。水分子、ナトリウムイオン、塩化物イオン以外の重原子の座標は拘束した。本研究ではテストセット A の 180 種の DNA-タンパク質複合体に対して上記の操作を行った。各複合体に対し、異なる 5 構造を試した。従って各条件で 900 構造について上記の操作を実施した。CHARMM27 力場ではテストセット C のみを用い、条件ごとに 125 構造 (25 種類を 5 構造ずつ)のエネルギー最小化・等温計算を実施した。

## 2.2.8 開発したモデリング手法の評価

### 2.2.8.1 評価に使った構造ファイルのセット

原子モデル構築手法を評価するため、3 つのデータセットを用意した。これらは X 線結晶構造解析による結晶構造である。それぞれの PDB ID は付録 F にまとめてある。

テストセット A: これが本研究のテストで主に利用した構造ファイルのセットである。全部で 180 構造の DNA-タンパク質複合体構造ファイルである。図 2-4, エネルギー最小化の力場依存性検討、CGDNA モデルの構造緩和条件検討以外ではこのテストセットすべてを使った。

テストセット B: DNA の結晶構造を粗視化し、再び原子モデルを構築するテストに利用した(図 2-4)。一分子の DNA のみを含む DNA-タンパク質の構造が 30 構造、それ以外の DNA-タンパク質複合体が 642 構造、A 型 DNA が 39 構造、Z 型 DNA が 29 構造、そして 4 重鎖 DNA が 22 構造含まれている。

テストセット C: この構造はテストセット A の部分構造である。全部で 25 構造の DNA-タンパク質複合体で構成される。エネルギー最小化の力場依存性の検討、CG モデルの構造緩和条件検討で使用した。テストセット A から 30 構造をランダムに選んだ。計算時間短縮のため、その中から分子量が特に大きい構造は除外した。またタンパク質のサブユニット間の接触面積が非常に小さい構造も除外した。

すべての構造は解像度が  $2.2 \text{ \AA}$  以内の結晶構造である。DNA のフラグメントライブラリに用いた構造ファイルとの重なりはない。再構築と CGMD を簡便に行う目的で、構造ファイルは事前に選別してある。DNA の内部に欠損原子を含むヌクレオチドや、アデニン・チミン・シトシン・グアニン以外の塩基を持つヌクレオチドが存在する構造ファイルは除外した。タンパク質についても、内部に欠損原子を含むアミノ酸がある場合や、20 種類の標準的なアミノ酸以外を内部に含む構造ファイルは除外した。ただし、セレノメチオニンについてはセレン原子の座標を硫黄原子の座標とみなしたメチオニンとして取り扱った。テストセット B では、ヌクレオチドのレベルで構造を選別した。欠損原子を含むデオキシリボヌクレオチドや、アデニン・チミン・シトシン・グアニン以外のヌクレオチドは解析に用いなかった。逆に、欠損原子を含むヌクレオチドが存在する構造ファイルでも、欠損がないヌクレオチドは解析に用いた。ヌクレオチドを除外した後に末端となったヌクレオチドは、再構築プログラムが末端として認識する事情で、末端のヌクレオチドとして取り扱った。

テストセット A のシミュレーションでは、実際の機能単位にかかわらず、PDB ファイルに含まれている構造でシミュレーション、再構築を行った。

### 2.2.8.2 根平均二乗偏差

最初に、結晶構造を CG モデル(3SPN.2C モデル)に変換し、開発した手法で再び原子モデルに戻すテストを行った。このテストを自己再構築テストと呼ぶことにする。このテストでの再構築の精度を見るため、ヌクレオチドごとに結晶構造と再構築モデルの間

の根平均二乗偏差(root mean square deviation, RMSD)を調べた。自己再構築テストでは、事前に分子の重ね合わせを行わずに RMSD を計算した。

### 2.2.8.3 ワトソン-クリック塩基対の再現性

ワトソン-クリック塩基対を形成しているとみなせる CG モデルから構築した原子モデルは、ワトソン-クリック塩基対の水素結合を再現することが望ましい。構築した原子モデルでのワトソン-クリック塩基対の再現度を評価の指標に利用した。CG モデルでワトソン-クリック塩基対を形成しているとみなした塩基ペアのうち、再構築後の原子モデルでもワトソン-クリック塩基対の水素結合を再現している塩基ペアの数を数え、割合を求めた。この操作は DNA-タンパク質複合体ごとに行った。CG モデルでのワトソン-クリック塩基対形成の判定基準は 2.2.3 節で既に述べてある。ただし、この基準で同じプリン塩基が複数の塩基対に共有された場合には、再構築後の原子モデルでワトソン-クリック塩基対の再現度がより高くなるような塩基ペア一つを選び、それ以外は除外した。実際にはプリン塩基が共有されていたために除外された塩基対は 0.6% 未満である。従ってこの塩基対二重カウントの問題は、結果に影響を及ぼさないとみなせる。なお原子モデルでは、ワトソン-クリック塩基対の 2 ないし 3 つの水素結合のドナー-アクセプター距離がすべて 3.5 Å より小さかった時、ワトソン-クリック塩基対の水素結合を再現しているとした。

### 2.2.8.4 DNA の連結部のモデリング精度

DNA 再構築手法ではヌクレオチドごとに原子モデルを当てはめる。従って、ヌクレオチド間の接続部での原子間距離、角度の精度は保証されていない。そこで、ヌクレオチド間の連結部にある原子間の結合距離、結合角が期待された値であるかを調べた。P 原子と O3' 原子の結合距離の分布と、O3' 原子、P 原子、O5' 原子からなる結合角の分布を調べた。

### 2.2.8.5 DNA-タンパク質間相互作用の再現性

結晶構造中の DNA-タンパク質間相互作用が CGMD、原子モデル再構築を経た後での程度再現されているかを調べた。既知の DNA-タンパク質複合体の結晶構造中で、互いに接近している頻度が高い原子の組み合わせについて (リジンの N<sub>ε</sub> 原子と DNA のリン酸の OP1 原子など)、それぞれ原子間接触の再現度を定義し、解析した。CGMD を行う前の初期構造(結晶構造)群(テストセット A または C)から、該当する原子ペアで距離が 7.0 Å 以内のものをすべて選んだ。そして CGMD・原子モデル構築を経た後、選んだ原子ペアの距離を再度測定した。もし原子ペアの距離が結晶構造での距離 $\pm 1$  Å の範囲にあったとき、その原子ペアの相互作用を再現できたと定義した。再現した相互作用の数を数え、再現した割合を計算した。原子の組み合わせの種類ごとに計算した。

エネルギー最小化の構造でこの解析を行う際には、構造が極度に不安定で最急降下法による構造緩和ができなかった場合や、構造が不安定でその後の MD が止まってしまった場合には、その複合体に含まれるすべての接触を”再現できなかった”とした。

#### 2.2.8.6 DNA のらせん軸の構造

DNA のらせん軸はソフトウェア Curves+を用いて行った<sup>129,130</sup>。

#### 2.2.9 既存の原子モデル構築ツールによる 3SPN.2C DNA モデルからの原子モデル構築

本研究で開発した手法を既存のモデリングツールと比較するため、Wassenaar らが開発した原子モデル再構築ツールを用いて、3SPN.2C モデルからの原子モデル構築を試みた<sup>88</sup>。モデリングには GROMACS 5.1.1 を併せて利用した<sup>54-61</sup>。まず 3 つの結晶構造の DNA(PDB ID : 1A1F, 1BNA, 1CDW)を粗視化し、Wassenaar らのツールを使った。このツールを 3SPN.C モデルに適用するため、トポロジーファイルの粗視化粒子と原子の対応を定義しなおした。それ以外はデフォルトの設定で利用した。O3'原子を隣のヌクレオチドのリン酸グループに含めることができるかが不明瞭だったため、この原子がリン酸に含まれるとした場合と、糖に含まれるとした場合の 2 通りを試した。

## 2.3 結果と考察

### 2.3.1 自己再構築テストによる DNA 再構築手法の評価

はじめに DNA の原子モデル構築法の性能を調べた。結晶構造を粗視化したモデルを、開発した手法で再び原子モデルに戻すテストを行った(自己再構築テスト)。このテストでは再構築法の性能を明確に評価できる。再構築法が優れていれば、元の結晶構造との構造の違いは小さい。ただし、このテストで用いる粗視化(CG)モデルはある意味で“優れたモデル”である。つまり、CG モデルに対応する安定な原子モデルが現実存在しているため、原子モデルを構築しやすい。実用上は、このテストの成績が良いだけでは不十分である。CGMD やその他のモデリングで得た分子構造は、原子モデルとの対応づけが難しい場合がある。CG 粒子の配置の定義に素直に従うだけでは、不自然・不安定な構造が現れる可能性がある。しかし、この自己再構築テストの実施は有益である。このテストで高い再現性を出すことは再構築法の満たすべき必要条件である。

まず 5 種類の構造セットについて自己再構築テストを行った。ここでは 2.2.8.1 節でテストセット B と名付けた構造セットを用いた。ここには一分子の DNA とタンパク質の複合体構造が 30 構造、それ以外の DNA-タンパク質複合体が 642 構造、A 型 DNA が 39 構造、Z 型 DNA が 29 構造、そして 4 重鎖 DNA が 22 構造含まれている。重みづけパラメータ(2.2.4 節参照)は  $w_{Bi} = 1.0$ ,  $w_{Si-1} = 0.1$ ,  $w_{Pi} = 0.75$ ,  $w_{Si} = 1.0$ , and  $w_{Pi+1} = 0.1$  とした。この値は、事前に検討したもので、詳細は付録 D、付録 E 図 SE-2, 図 SE-3, 表 SE-2, SE-3 にまとめてある。自己再構築テストの結果の 4 構造をを図 2-4A に示した。元の結晶構造の DNA がピンク色、再構築後の構造が緑色である。ほとんどの場合にこの再構築法は結晶構造をよく再現している。B 型の DNA だけでなく、一本鎖状態の DNA、Z 型 DNA、四重鎖 DNA の構造のすべてで、ほとんどのヌクレオチドを精度よく再現できている。結晶構造と再構築モデルの構造のずれを RMSD で比較した。どの構造セットでも RMSD は 0.1-0.2 Å の周囲で最も頻度が高かった。ここからも、本研究の再構築法が様々な構造の DNA に適用できることが分かる(図 2-4B)。

続いて、より少ない構造セットでより詳細に原子モデルの再構築を検討した。ここで用いたテストセット A には 180 の DNA-タンパク質複合体の構造ファイルが含まれる。テストセット B と同様の自己再構築テストでは、RMSD の分布は 0.14 Å 付近にピークを持っていた。RMSD が 0.25 Å 以内の範囲には全体の 70% のヌクレオチドが存在した。そして 91% のヌクレオチドが 0.5 Å 以内の RMSD であった。このテストセットに対しても、実験で決定された DNA 構造を高精度で再現できた。構造を観察した結果、DNA の塩基の面の向きをよく再現できることが確認できた。エネルギーを直接計算しなくてもこの結果が得られたことは、塩基の面の向きは多くの場合に制限されていることを示唆している。



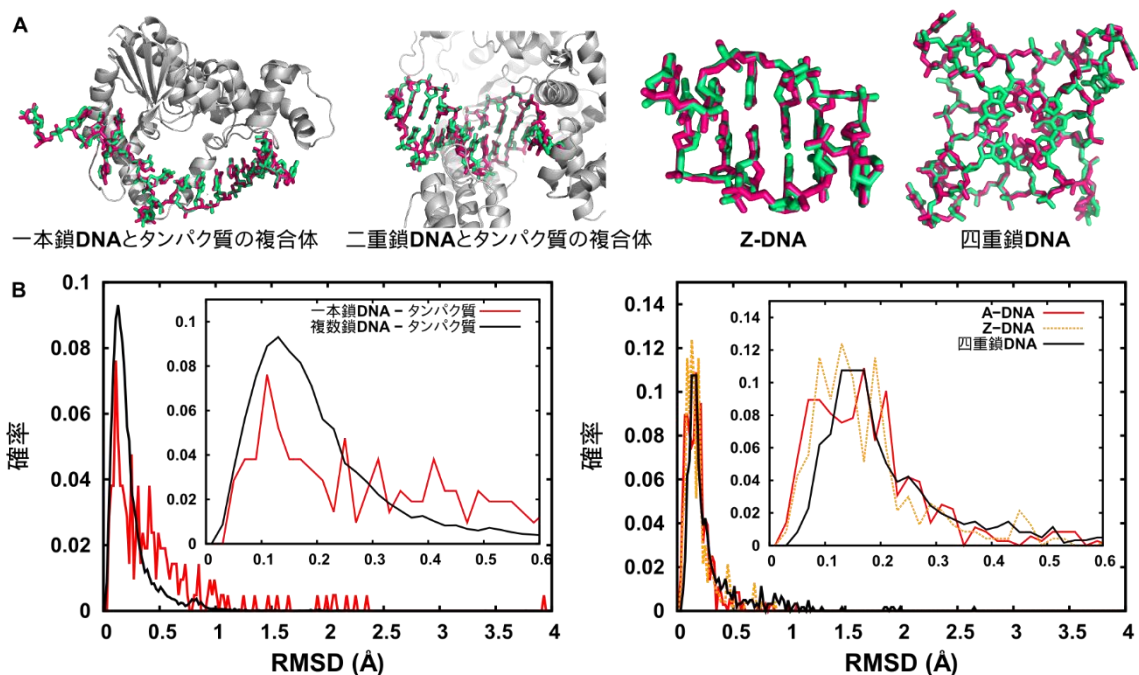


図 2-4. DNA 構造の自己再構築テスト 複合体中に DNA を一分子のみ含む DNA-タンパク質複合体、二本以上の DNA 鎖を含む DNA-タンパク質複合体、A 型 DNA、Z 型 DNA、DNA 四重鎖それぞれについて開発した手法を評価した。A) X 線結晶解析により決定された構造 (ピンク色) と再構築したモデル (緑色) の比較。B) ヌクレオチドごとに算出した、結晶構造と再構築構造の RMSD の分布。ただし、末端のヌクレオチドのデータは含まない。

テストセット A の自己再構築テストでモデリングがうまくいかなかった構造として、RMSD が  $2.0 \text{ \AA}$  より大きい構造が 10 ヌクレオチドあった。結晶構造と再構築後の構造を比較したところ、塩基の面がおよそ  $180^\circ$  回転していた(付録 E 図 SE-4)。こういったヌクレオチドは、結晶構造中で他のヌクレオチドと塩基間スタッキングをしていないか、またはワトソン-クリック塩基対を形成していなかった。従って、二重鎖 DNA でない構造では原子モデルの再構築が困難な場合があることが示唆された。付録 E 表 SE-4 には、モデリングパラメータ  $w_i$  を変更してもモデリングに失敗した構造をまとめてある。そこでも、二重鎖 DNA でない構造は原子モデル構築に失敗することがあるという、同様の結果が得られた。

続いて、テストセット A の末端のヌクレオチドの再現性について調べた。先に 2.2.4 節で示した通り、末端ではそれ以外とは異なる方法でモデリングしている。自己再構築テストの結果は、5'末端ヌクレオチドの RMSD は内部や 3'末端のヌクレオチドの RMSD よりもやや大きかった(付録 E 図 SE-5)。ただし、末端に位置しているヌクレオチドは全原子 MD を用いれば比較的容易に構造を緩和できると期待される。従って DNA 末端のモデリングに関して、さらに操作を追加することは検討しなかった。

加えて、DNA 再構築法にかかる時間のテストを行った。計 145 塩基対からなる(PDB ID: 3UT9 より) DNACG モデルから 1,000 回、原子モデルを構築した。計算機は 3.40GHz の Intel(R) Core(TM) i7-3770 CPU を用いた。所要時間はおよそ 140 から 153 秒程度であ

った。従って、一塩基対あたり 0.001 秒程度でモデリングできた。内訳はフラグメントライブラリを読み込む 1.5-1.6 秒の時間と、1,000 構造のモデリングに要する時間からなる。CGMD のトラジェクトリを網羅的に原子モデルに変換するのに十分な速度があるといえる。

### 2.3.2 CGMD スナップショットから構築した DNA 原子モデルの評価

続いて、原子モデル構築法が CGMD 後の分子構造にも適用できるかを調べた。自己再構築テストとは異なり、CGMD 後の構造は原子モデルに対応付けられない粒子間距離、粒子間仮想結合角をとっている可能性がある。そこでまず構造緩和を行うことにした。筆者は 1 K 下で短時間の CGMD を行った。そしてその後で DNA 原子モデルを構築した。この 1 K の CGMD は CG モデルの局所構造を改善するが、複合体全体のおおきな構造変化は許容しない。構造緩和手法を選択するに当たり、様々な条件の徐冷法等を試している。その詳細は付録 D にまとめた。また付録 E 図 SE-6, SE-7, SE-8, SE-9, 付録 E 表 SE-1)にも内容を示してある。

テストセット A の 180 の DNA-タンパク質複合体それぞれを初期構造とし、300 K で短い CGMD を実行した。この CGMD スナップショットが再構築法のテストの対象である。

はじめに CG モデルのヌクレオチド内の糖粒子-塩基粒子仮想結合長を調べた。まずは結晶構造で CG 粒子に対応する 2 座標間の距離を調べた。するとそれぞれ 4.3Å と 4.9Å の付近に鋭いピークが現れた。これはそれぞれ、糖とピリミジン塩基の重心間距離、糖とプリン塩基の重心間距離である。一方温度 300 K の CGMD 直後の CG モデルでは、CG 分子の揺らぎにより分布は滑らかになった。結晶構造でみられた 2 つのピークは消失し、4.6Å 付近にピークがあるだけとなった(図 2-5B “CGMD”)。続く 1 K の構造緩和シミュレーション後には、二つのピークが再び現れた(図 2-5B “1 K 緩和計算”)。従って、1 K の構造緩和計算により CG モデルの糖-塩基間距離は改善され、より原子モデルに対応付けやすくなった。さらに、同一ヌクレオチド内の糖-リン酸の仮想結合長の分布も調べた(付録 E 図 SE-10)。この 2 粒子の距離も、1 K の構造緩和シミュレーションによって改善された(付録 E 図 SE-10)。CG モデルの仮想結合角に関しては、同一ヌクレオチド内のリン酸・糖・塩基がなす角を調べた(付録 E 図 SE-10)。緩和シミュレーションによる構造改善はほぼ見られなかった。しかしこの場合には、300 K の CGMD 後の角度が結晶構造の対応する角度よりも狭い範囲に分布していた。

続いてワトソン-クリック塩基対構造の再現性を調べた。CG モデルでワトソン-クリック塩基対様の構造をとっていた場合には、原子モデル構築後には同じ塩基ペアでワトソン-クリック塩基対の水素結合を再現していることが望ましい。CG モデルからワトソン-クリック塩基対を判定する基準および、原子モデル構築後のワトソン-クリック塩基対再現度の定義はそれぞれ 2.2.3 節、2.2.8.3 節にまとめてある。

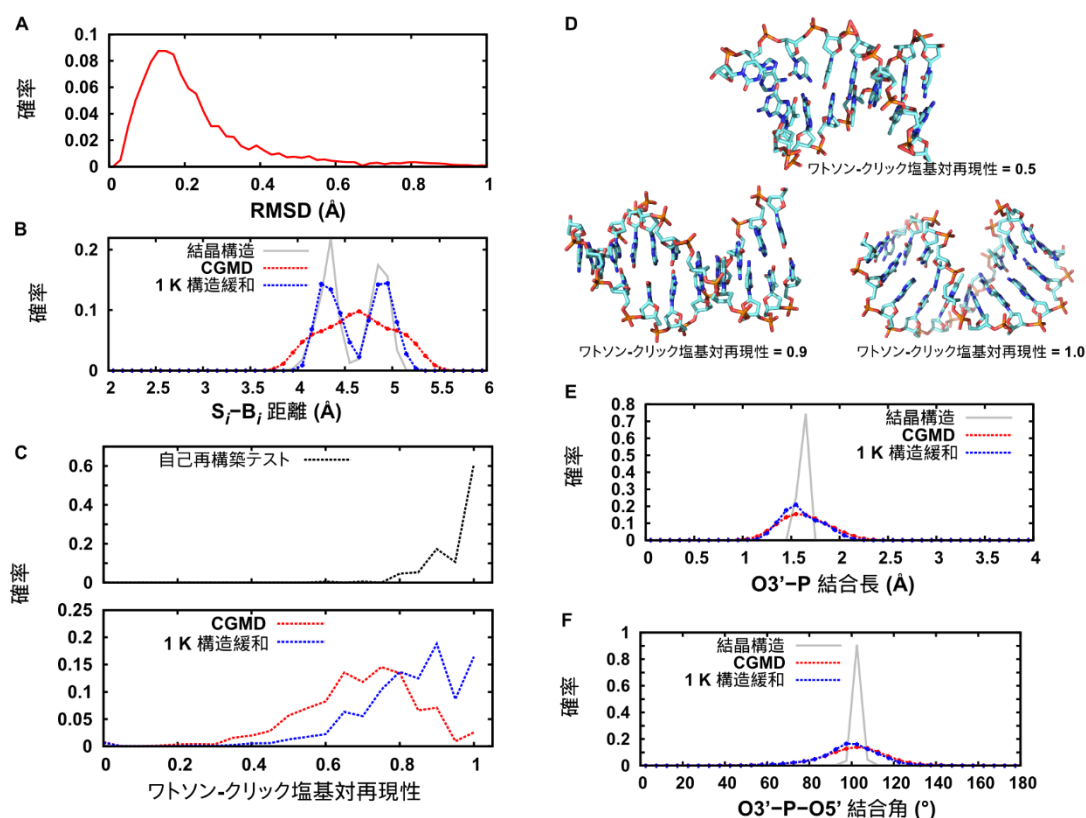


図2-5. テストセットAを用いたDNA再構成手法の評価 A) 自己再構築テストでの結晶構造と再構築モデル間のRMSD。RMSDはヌクレオチドごとに計算した。末端のヌクレオチド以外についての結果を示した。B-E) CGMD後の構造からのDNA原子モデル構築 B) 粗視化モデルの糖粒子と、それに結合している塩基粒子の距離の分布。C) ワトソン-クリック塩基対再現性の分布 D) 再構築したDNAの例。ワトソン-クリック塩基対再現性も併記した。E) O3'-P結合距離の分布。F) O3'-P-O5'結合角の分布。

なお、図で“CGMD”と表記したものは300 KでCGMDを行った後、再構築したDNA構造のデータを示す。“1 K 構造緩和”は、1 KのCGMDによって構造緩和を行った後のDNA構造のデータを表す。“結晶構造”は、結晶構造で粗視化粒子の位置に対応する座標から求めたデータを示した。エラーバーは20回のシミュレーションの標準誤差である。

自己再構築テストの構造でワトソン-クリック塩基対再現度を調べると、高い再現性が見られた。ここではワトソン-クリック塩基対を再現すべき塩基ペアは、結晶構造をCGモデルに変換した構造を参照して決めた。テストセットAの180の結晶構造の内61%ですべてのワトソン-クリック塩基対を再現できた(図2-5C, 黒点線)。一方、300 KのCGMD直後には、それぞれの塩基対再現性は0.3から1.0の幅広い領域に分布していた。CGMDスナップショットの原子モデル構築では、再現すべきワトソン-クリック塩基対は300 KのCGMD直後の構造を参照し決定した。構造緩和を行った後にDNA原子モデルを構築するとワトソン-クリック塩基対再現性は大きく上昇した(図2-5C, 青点線)。再現性が0.8を超える構造の割合が大幅に増加し、すべての塩基対を再現できた構造も増加した。従って1 Kの構造緩和計算は、再構築した原子モデルのワトソン-クリック塩

基対再現性を大きく改善するといえる。ただ、塩基対の再現性が 0.5 程度の構造でも塩基の面の向きは大まかには塩基対形成時と類似していた (図 2-5D)。

加えて、ワトソン-クリック塩基対再現性が、DNA フラグメントライブラリのフラグメント数の増加によって改善されうるかを調べた。DNA フラグメントの数が異なるいくつかのライブラリを使って DNA の原子モデルを構築し(付録 E 表 SE-5)、結果を比較した(付録 D, 付録 E 図 SE-11)。フラグメントの多いライブラリを用いた場合にも、ワトソン-クリック塩基対再現性の改善は少しであった。自己再構築テストでの塩基対再現性との間には、依然として大きな差がある。自己再構築テストの RMSD など、ほかの指標のフラグメントライブラリ依存性も付録 E 図 SE-11 にまとめてある。

最後に隣接するヌクレオチド間の構造を調べた。DNA の原子モデル構築では一ヌクレオチドごとに原子構造を配置する。そのため、同じヌクレオチド内の構造が正確であっても、2ヌクレオチドの連結部分の構造の精度は保証されない。そこで、共有結合を形成しているべき O3'原子-P 原子間の距離を測定した。また、O5'原子-P 原子-O3'原子の結合角を調べた。末端の塩基が関わらない連結部についてのデータは図 2-5E,F に、末端が関わる連結部のデータは付録 E 図 SE-5 にまとめてある。これらの分布のピークは結晶構造のピークと一致しているものの、この原子間距離や角度の値は結晶構造よりもばらついていた。温度 1 K で構造を緩和してから原子モデル構築した場合も、そうでもない場合も結果に違いは見られなかった。ただ、O3'原子-P 原子間距離と O5'原子-P 原子-O3'原子結合角は全原子力場を用いてエネルギー最小化を行うことで改善されると考えられる。実際、あとで述べるようにエネルギー最小化によってこれらの局所分子構造は大幅に改善された。

### 2.3.3 DNA-タンパク質相互作用面のアミノ酸の構築

タンパク質の原子モデル構築では、DNA-タンパク質相互作用面のアミノ酸側鎖の方向を正しくモデリングすることが重要である。ここでは 2 つの状況を想定している。I. DNA-タンパク質複合体の構造は既に解かれている。そしてシミュレーション中で複合体の DNA-タンパク質相互作用面以外の構造が変化した場合。II. シミュレーション中に、タンパク質と DNA が今まで知られていなかった結合の仕方をした場合。この I. の状況では、可能な限り既知構造を利用したほうが精度の良いモデルが構築できる。他方 II. の場合には、参考にできる実験データがなければ、CGMD だけからは側鎖の配向の情報は得られない。適切な側鎖の構造は、原子モデルを構築し、全原子 MD を長時間行うことで検討できるかもしれない。

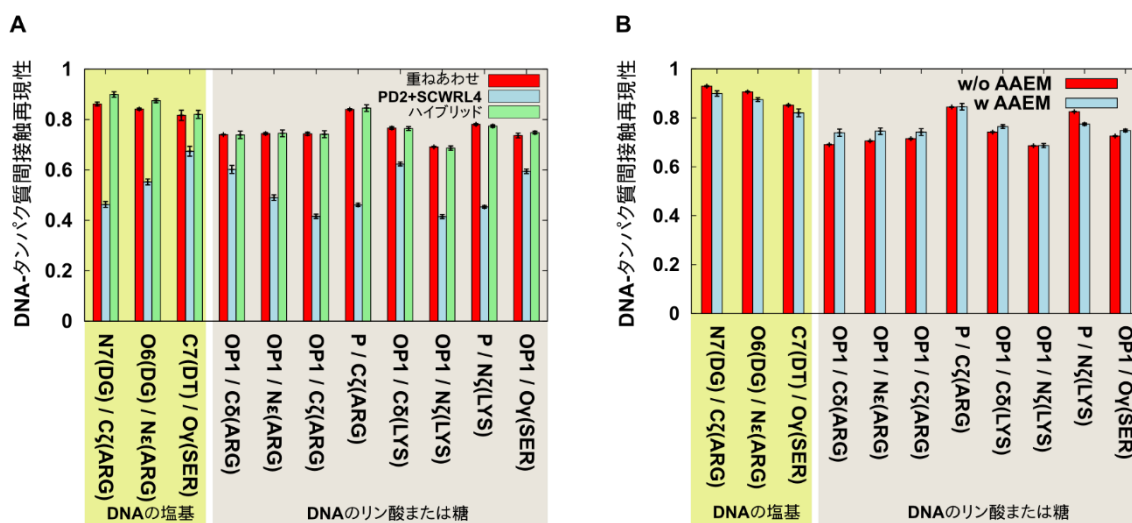


図 2-6. DNA-タンパク質相互作用の再現性のテスト DNA-タンパク質間の接触の再現性。A) 異なる 3 つのタンパク質モデル構築法での再現性の違いを示した。“重ね合わせ”では、実験によって決定されたタンパク質結晶構造を、粗視化モデルに重ね合わせた。“PD2+SCWRL4”では、モデリングソフトウェア PD2 CA2MAIN を用いて主鎖を構築した後で、ソフトウェア SCWRL4 で側鎖を構築した。“ハイブリッド”では、主鎖の構築に PD2 CA2MAIN を利用し、DNA-タンパク質相互作用面の側鎖構築に既知の結晶構造情報を利用した。その他の領域の側鎖は SCWRL4 を用いてモデリングした。DG がグアニン塩基を、DT はチミン塩基を示す。エラーバーは 5 種類の異なる CGMD スナップショットを用いたモデリング結果の標準誤差を示す。B) 全原子力場を用いたエネルギー最小化前後での、原子間接触再現性の比較。グラフ中“w/o AAEM”はエネルギー最小化前、“AAEM”はエネルギー最小化後の値を示す。ここでは、ハイブリッド手法で構築した構造の結果を示した。エラーバーは異なる CGMD スナップショットを用いたモデリング結果の標準誤差を示す。エネルギー最小化前は N = 20, 最小化後は N = 5 である。

上述した I. の状況に焦点を当て、DNA-タンパク質複合体の既知の構造情報を取り入れることにした。DNA-タンパク質相互作用面をできる限り正確にモデリングし、かつ分子の構造変化にも対応するため、筆者はハイブリッド手法を考案した。この方法では DNA-タンパク質相互作用面のみ既知構造の情報を利用する。残りのアミノ酸側鎖は既存のモデリングソフトウェアである SCWRL4 によって構築する。

CGMD 後の構造から、タンパク質の原子モデルを構築した。ハイブリッド手法を試す前に、2 つのアミノ酸粒子の距離が実験値と合致するかを調べた。隣接する C $\alpha$  原子間の距離は cis-プロリン以外では 3.8 Å に鋭いピークを持っている。温度 1 K の構造緩和計算によって構造を緩和すると、C $\alpha$  原子間の距離は結晶構造で見られる距離に近づいた(付録 E 図 SE-12)。タンパク質の原子モデルを構築した後、原子モデルのエネルギー最小化計算を行った。

エネルギー最小化計算後、結晶構造で見られた DNA-タンパク質相互作用の再現度を調べた。温度が 300 K の条件下、計 10,000 MD ステップの CGMD は短時間のシミュレーションである。従ってこの計算の後も初期(結晶)構造の DNA-タンパク質相互作用は維持されたままだと考えられる。このような状況では、単純に結晶構造を CG モデルに重ね合わせると、DNA-タンパク質相互作用をよく再現できると考えられる。したがって、“重ね合わせ”のアプローチをポジティブコントロールとみなした。解析には 15 種

類の原子の組み合わせを用いた。それぞれについて、原子モデル構築後に、結晶構造で見られた原子間接触がどれだけ再現されているかを調べた。DNA-タンパク質間の原子間接触の再現性の定義は2.2.8.5節にまとめてある。結果は図2-6Aと付録E表SE-6にまとめた。図2-6Aでは、タンパク質とDNAの塩基の相互作用と、タンパク質とDNAの骨格部分の相互作用とで背景を塗り分けている。アミノ酸とDNA塩基の相互作用は、塩基配列の認識に影響する。アミノ酸とDNAの骨格の相互作用は、塩基配列非特異的なDNA-タンパク質相互作用の強さに関わっている。本研究の”ハイブリッド”法では”重ね合わせ”と同程度、DNA-タンパク質間の原子の接触を再現できた。よって、ハイブリッド手法はDNA-タンパク質相互作用面のアミノ酸側鎖を精度よく構築できるといえる。解析した15の原子ペアでは、タンパク質とDNAの塩基の原子の接触の再現性は、タンパク質とDNAの骨格部分の原子の接触の再現性よりも高かった。また多くの場合に”PD2+SCWRL4”の方法で構築した原子モデルではこの再現性が他2つの方法よりも低かった。DNAと接触している原子がアミノ酸主鎖から離れているほど、”PD2+SCWRL4”の方法でのDNA-タンパク質間の原子の接触の再現性が低下した。例えば、リジンのN<sub>ε</sub>原子とDNAのOP1原子ペアの接触の再現度は40%程度と低かったが、セリンのO<sub>γ</sub>原子とDNAのOP1原子ペアの接触の再現度は60%であった。おそらくこれは、アミノ酸側鎖のロータマーの選択時、既知構造と異なる構造を選択する可能性を反映している。例えば、主鎖の構造が決まると同時に、O<sub>γ</sub>原子の位置はかなり絞り込まれる。以上で検討した状況では、アミノ酸側鎖の方向と、CGモデルからのC<sub>α</sub>原子の座標を共に再現するハイブリッド法が有利である。つまりハイブリッドアプローチはI.の状況のDNA-タンパク質複合体のモデリングに有効である。

次にエネルギー最小化前と最小化後で、この再現性の比較を行った。再現性はほぼ同程度であった。ハイブリッドアプローチに対しての結果は図2-6Bに示した。それ以外は付録E表SE-6、表SE-7、図SE-13にまとめてある。タンパク質を”PD2+SCWRL4”アプローチでモデリングした場合は、エネルギー最小化前、最小化後共にほかの2つの方法よりも再現性が低かった。つまり、エネルギー最小化の初期構造が悪いとき、最急降下法による構造改善は期待できないといえる。

また、DNA-タンパク質間の接触の再現性の力場依存性を調べた。テストセットCを用い、AMBER99 bsc1力場<sup>122</sup>を用いた場合と、CHARMM27力場<sup>123-127</sup>を用いた場合の再現性を比較した。少なくとも両者の間には差異は確認できなかった(付録E表SE-8)。

さらにDNA-タンパク質相互作用面のアミノ酸側鎖の構造を確認した。C<sub>β</sub>原子は既知構造から構築した側鎖とPD2 CA2MAINで構築した主鎖の境界に当たる。本研究のハイブリッド構築法では、C<sub>β</sub>原子を頂点とする結合角が必ずしも適切であるとは限らない。そこでこの結合角の分布を調べた。エネルギー最小化前の構造でも、結晶構造中の角度周りに分布していた。全原子力場でエネルギー最小化を行うと、角度の分布は実験値に近いものに改善された(付録E図SE-14)。

エネルギー最小化後に、ソフトウェア GROMACS 2016.3 を使い、短時間の全原子 MD を行った<sup>59-61</sup>。このシミュレーションは、95%の複合体で無事終了した。しかし約 5%の複合体では LINCS の警告が多数生じ、計算が止まってしまった。シミュレーションの初期構造に不自然な構造が存在するとき、このような状況が起こる。従ってほとんどの場合に、構築した DNA-タンパク質複合体の原子モデルを使った全原子 MD が可能である。計算が停止した構造の多くは B 型ではない DNA 構造を含んでいた。B 型 DNA 構造を形成し易い 3SPN.2C モデルとうまく整合しない結果、不安定な構造を生じたのかもしれない。

### 2.3.4 エネルギー最小化計算による DNA 構造の改善

はじめにワトソン-クリック塩基対の再現性を調べた。エネルギー最小化後には、ほとんどの構造で 0.8 以上の再現性となった。そして 60%以上の構造で 1.0 となった(図 2-7A)。エネルギー最小化前よりも明らかに高い再現度であり、自己再構築テストの場合と同程度である。従ってこれは、原子モデル構築がより困難な CGMD 中の構造に対しても本研究の手法が有効であることを示している。ワトソン-クリック塩基対を再現すべき塩基ペアのドナー-アクセプター間距離を調べると、2.7 から 3.2 Å の範囲に収まっていた(図 2-7B)。この分布の幅は自己再構築テストやエネルギー最小化前分布よりも明らかに狭い。以上よりワトソン-クリック塩基対の再現性の観点からは、エネルギー最小化計算による構造の改善が確認された。

全原子力場の効果により、DNA 骨格の構造も明確に改善された。O3'-P 結合長、O5'-P-O3' 結合角はともに、結晶構造で見られる範囲にほぼ収まった(図 2-7C, D)。

エネルギー最小化による DNA 構造の改善が全原子力場の影響を受けるのかを調べた(付録 E 図 SE-15)。AMBER99 bsc1 力場を用いた場合に比べ、CHARMM27 力場ではワトソン-クリック塩基対の再現度がやや低かった。ただし、それぞれの力場を用いたエネルギー最小化後の構造を比較すると、違いはわずかであった。全原子力場のパラメータの違いが原子の位置にわずかなずれを生み出し、それが反映されたのかもしれない。結局、どちらの力場を用いた場合にも問題はないといえる。ヌクレオチド間の接続部の構造も、どちらの力場を用いても改善された。

以上の解析では、主に局所構造を見てきた。それと合わせ、構築後のモデルはより大域的な DNA の構造変化も再現できる必要がある。タンパク質との相互作用による DNA の変形を再現できるかどうかは、CGMD 力場に大きく依存する。ただし、筆者の原子モデル再構築法が、CGMD 中に起こった DNA の変形を反映できることは確かめなければならない。図 2-8A, B では、TATA 結合タンパク質について、(PDB ID: 1CDW<sup>131</sup>) 結晶構造と、エネルギー最小化を含めた原子モデル構築後の構造を比較して図示した。この結晶構造はテストセット A に含まれている。エネルギー最小化を行った 5 構造の内一つでは、原子モデルがうまく構築できなかつたために続く全原子 MD ができなかつた

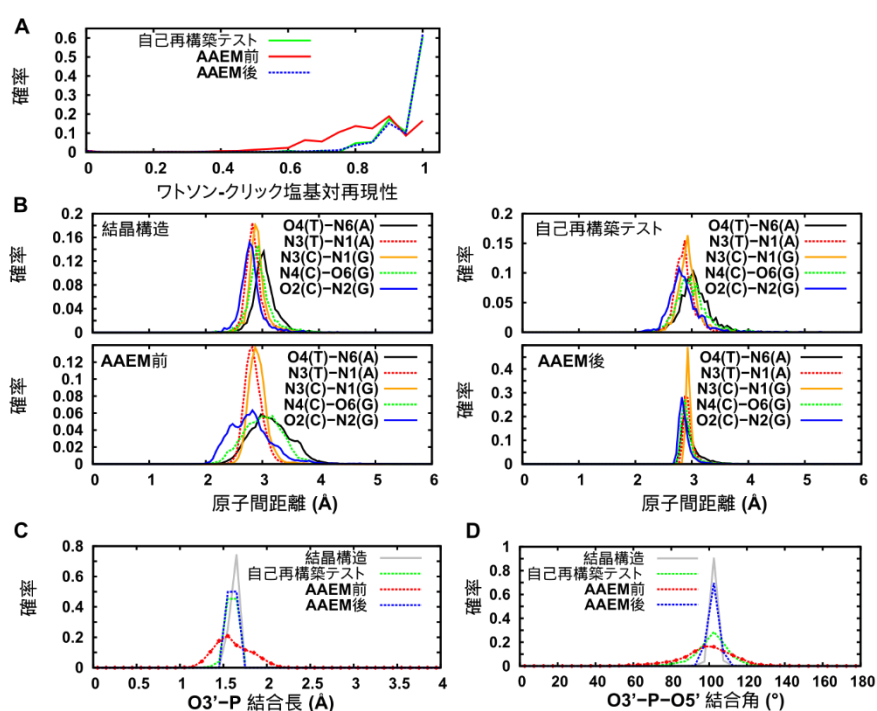
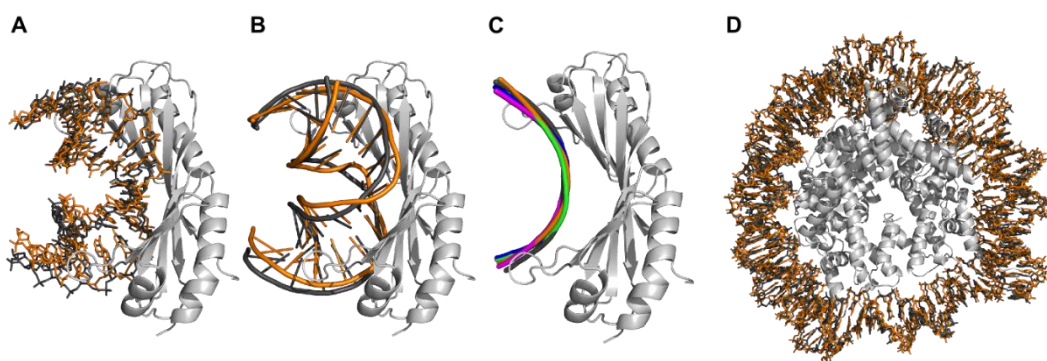


図 2-7. 全原子力場を用いたエネルギー最小化による DNA 局所構造の改善 A) ワトソン-クリック塩基対再現性。B) ワトソン-クリック塩基対のドナー-アクセプター間の距離の分布。粗視化モデルに基づき、ワトソン-クリック塩基対を形成しているべき塩基ペアでのドナー-アクセプター間距離を調べた。C) O3'原子と P 原子の間の共有結合距離。D) O3'原子、P 原子、O5'原子のなす角の分布。“結晶構造”は、結晶構造での値の分布を示す。“自己再構築テスト”は自己再構築テスト時の分布を示す。“AAEM 前”はエネルギー最小化前の値を示す。エネルギー最小化前の値のエラーバーは、20 回のシミュレーション結果の標準誤差を示す。“AAEM 後”はエネルギー最小化後の分布を示す。エネルギー最小化後の結果は、引き続き 10 ピコ秒の全原子計算を実行できた構造のみを使って算出した。

が、残りの 4 構造では全原子 MD を問題なく実行できた。もとの結晶構造のタンパク質は薄い灰色、結晶構造の DNA は黒よりの灰色、再構築に成功したモデルの内一つの DNA 部分をオレンジ色で示した。再構築に成功したモデルでは、TATA 結合タンパク質に結合した DNA の構造をよく再現できた。図 2-8C には DNA 構造解析ソフト Curves+ で計算した DNA のらせん軸が描かれている。暗い灰色で着色したものが結晶構造での DNA のらせん軸である。それ以外の色で着色したものは、エネルギー最小化後の安定な 4 構造のものである。この 4 構造では、DNA のらせん軸の曲がりや結晶構造のらせん軸とよく一致していた。

別の例としてヌクレオソームに結合した DNA の構造の再現性も確認した(図 2-8D)。CGMD に使った構造の PDB ID は 3UT9<sup>132</sup>で、この構造もテストセット A に含まれていたものである。ヌクレオソームの CGMD 後の構造から、結晶構造に近い DNA 構造を再現できた。





**図 2-8. 変形した DNA の構造再現性** 大きく変形した DNA をモデリングできることを確認した。結晶構造と、エネルギー最小化を含めた再構築後の構造を比較した。A-C) TATA 結合タンパク質による DNA の変形を調べた。PDB ID : 1CDW の構造について解析した。A-B) 再構築した構造のうち一つの DNA 部分を結晶構造に重ね合わせた。A と B では、同じ構造を二種類の異なる方法で描画している。描画には PyMOL を用いた<sup>53</sup>。C) DNA のらせん軸をソフトウェア Curves+によって解析し、比較した。C では結晶構造と、4 つの再構築後の構造のらせん軸を比較している。D) ヌクレオソームの結晶構造中の曲がった DNA を再構築できるかを調べた。PDB ID : 3UT9 の構造について解析した。図で、結晶構造中のタンパク質は白色に近い灰色で示した。結晶構造中の DNA は黒に近い灰色で示した。A, B, D では再構築した構造を橙色で示した。C では結晶構造の DNA のらせん軸を黒に近い灰色、再構築したモデルのらせん軸をそれぞれピンク、緑、青、橙色で示した。

### 2.3.5 ほかの原子モデル構築法との比較

CG モデルから原子モデルを構築する方法は、CG モデルの性質・解像度に強く依存する。Rzpieira らや Wassenaar らは生体分子一般に向けた原子モデル構築法を提案している<sup>87,88</sup>。彼らの方法では、最初に原子を配置する際には構造のゆがみを許している。彼らはタンパク質と脂質膜からなる系の CG モデルを対象に、原子モデルを構築した。また、Wassenaar らは解像度が低い CG 脂質膜モデルから、より解像度が高い CG 脂質膜モデルへの変換にも成功している。ただ、彼らが原子モデル構築を行ったタンパク質 CG モデルは、本研究で用いたものよりも解像度が高い。特にアミノ酸側鎖の構造情報を含む。この点で今回筆者がテストしたモデルとは異なる。

DNA の CG モデルから原子モデルを構築する方法はいくつかのグループで試されている。Naômé らは各ヌクレオチドが一粒で表現されるモデルからの原子モデル構築を行っている<sup>103</sup>。このグループでは Rzpieira らの方法に基づき、そこに変更を加えて原子モデル構築を達成している。ほかのグループでは 3SPN.2C モデルよりも解像度が高いモデルから再構築を行っている<sup>89,100,102,118</sup>。彼らの場合は、CG モデルの構造のみから対応する原子モデルの一つをほぼ直接構成できる。筆者の構造ライブラリによるアプローチは、エネルギー最小化を行わなくとも精度の高い構造を出力できる利点がある。MD で構造を最適化する方法よりも圧倒的に速い。本研究の手法を用いれば、1,000 から 1,000,000 とした膨大な構造を短時間で原子モデルに変換できる。それを応用すれば、例えば X 線小角散乱の実験データとの比較ができる<sup>99</sup>。

筆者は Wassenaar らが開発した構築法を使い、3SPN.2C モデル(粗視化 DNA モデル)からの原子モデル構築を試みた。粗視化粒子と原子の対応関係を変更する以外は初期設定のままでモデリングを行った。しかし、このプログラムはエネルギー最小化の段階で

止まるか、その後の MD で多数の LINCIS の警告とともに途中で停止してしまった。こうした GROMACS のエラーは構造が非常に不安定な場合に見られる。原子を配置する段階でより詳細に原子間の位置関係を制御する必要があったのだと推測される。この観点では、構造ライブラリに基づく筆者の方法は、彼らの手法の原子を配置するステップの代わりとして利用できる。

アミノ酸側鎖の情報を持たない低解像度 CG モデルから原子モデルを構築する際には、特有の問題点がある。不適切な方向に側鎖が配置された場合、原子の密度が高い環境ではそれを矯正することは難しい。その点で、原子を比較的大雑把に配置し、MD で構造を緩和する方法にはリスクがある。本研究では、実験によって既に決定された構造を用いることでこのリスクを回避している。この方法により、既知の DNA-タンパク質相互作用面の再現率は既存のツールのみを用いた場合よりも大幅によくなっている。既知の 3 次元構造を CG モデルに重ね合わせるアイデアは既に提案されている<sup>92</sup>。本研究の方法では、既知構造を利用する領域を DNA-タンパク質相互作用面のアミノ酸側鎖のみに絞ることで、CGMD 中に起こったほかの部分の構造変化に対応できるようになっている。

## 2.4 まとめ

本研究では、DNA-タンパク質複合体の粗視化(CG)モデルから原子解像度モデルを構築する新たな方法を開発した。主なターゲットは各アミノ酸が一粒子で表現され、DNAのヌクレオチドが3粒子で表現されたDNA-タンパク質複合体のCGモデルである。

はじめにDNA再構築の新たな手法を考案した。この方法で、DNAの塩基の面の向きをほとんどの場合に再現できた。さらにB型の二重らせんDNA以外にも適用できることが分かった。原子モデル構築に先立ち、粗視化計算(CGMD)後の構造に、さらに1K下で短時間のCGMDを行うと、モデルの構造が改善された。この1Kの緩和計算後にDNA再構築を行うことで、ワトソン-クリック塩基対構造の再現性が大幅に高まった。

続いて、C $\alpha$ 原子の位置に粒子を置いたタンパク質CGモデルからの原子解像度モデル構築法を開発した。筆者は既知の結晶構造の重ね合わせと、既存のモデリングツールを組み合わせ、両者の利点をともに得られるようにした。これによりDNAとタンパク質の相互作用を高い精度で再現することができた。考案したハイブリッドアプローチは、タンパク質多量体やタンパク質-リガンド複合体にも適用できると考えられる。

CGMDで得られたDNA-タンパク質複合体からの原子解像度モデルの構築法を提案した。最後にエネルギー最小化計算を行うことにより、原子モデルの局所構造は実験で決定された既知構造の付近に収束した。さらに、ほとんどの場合で続けて全原子MDを実行できた。従って、CGモデルから再構築した原子モデルは、全原子MDでさらに解析できるだけの精度を持っているといえる。

## 結論

本研究では、大腸菌 *E. coli* の複製開始複合体が特定の DNA 領域のみを開裂する機構の解明を目指した。筆者はまず複製開始複合体の構造をモデリングした。筆者の構造モデルの妥当性は、共同研究者の生化学実験により支持された。巨大な複合体の構造をシミュレーションによってモデリングできたことから、構造と機能の関係も解析した。*oriC* 上の DnaA 結合サイトと IHF 結合サイトの間隔、そして左側サブ複合体と右側サブ複合体の位置関係の機能的意義を見出した。

二重らせん開裂機構解明の視点からは、複製開始複合体の構造が明らかにできたこと、特に DUE 領域と DnaA ドメイン III ホモ多量体の位置関係を大まかに予想できるモデルを得たことが意義深い。

粗視化 DNA-タンパク質複合体モデルから原子モデルを構築する方法を確立したことで、二重らせん開裂のメカニズムにより詳細に迫ることができると期待される。DnaA ドメイン III が DNA をとらえる過程を詳しく調べるには、アミノ酸を一粒子で表現したモデルでは不十分である。その一方で、複製開始複合体の構造をサンプリングするには、現状では粗視化計算が必須である。筆者が開発した方法は双方の利点をともに活かすことを可能とする。

今後の課題として、粗視化 DNA モデルの開発が重要である。現在、二重鎖 DNA の開裂過程をシミュレーションするのにふさわしい粗視化モデルが存在しない。この粗視化モデルは、一本鎖 DNA と二重鎖 DNA の性質をともに再現するものでなければならない。特に二重らせん開裂領域が何によって決まるのかを知るためには、一本鎖 DNA 構造の塩基配列依存性を組み込む必要がある。こうした、一本鎖 DNA 構造の塩基配列依存性に重点を置いた粗視化モデルはまだない。

近年、X 線小角散乱実験データを用いて、一本鎖 DNA の構造アンサンブルをモデリングした研究が報告された<sup>133,134</sup>。そこでは、塩基配列の違いに由来する一本鎖 DNA の構造の違いが明確に現れていた。このことより、X 線小角散乱実験のデータを丁寧に再現することが、高性能な粗視化 DNA モデル構築の手掛かりになると考えられる。実験で測定された慣性半径を粗視化モデルの評価に使った例はすでにある。しかし、そこで用いられた慣性半径は粗視化粒子の半径に基づく見積もりであった。それでは厳密性に欠ける。それに対して、本研究で開発した方法で原子解像度モデルを構築すれば、より厳密に実験値と比較できる。つまり、原子解像度モデルの構築法は粗視化 DNA モデルの開発時にも大いに役立つと考えられる。DNA モデルを開発した後で、複製開始複合体のシミュレーションを再開し、DUE 開裂過程を明らかにする予定である。

## 発表論文

### 第一章

Masahiro Shimizu, Yasunori Noguchi, Yukari Sakiyama, Hironori Kawakami, Tsutomu Katayama, and Shoji Takada; Near-atomic structural model for bacterial DNA replication initiation complex and its functional insights; *Proc. Natl. Acad. Sci. U. S. A.*; 113(50); E8021-E8030; 2016.

doi: 10.1073/pnas.1609649113

PNAS ウェブサイトのリンク

<http://www.pnas.org/content/113/50/E8021.short>

### 第二章

The English version of chapter 2 of this thesis is the Accepted Manuscript version of a Published Work that appeared in the Journal of Chemical Theory and Computation, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see: Masahiro Shimizu, and Shoji Takada; *J. Chem. Theory Comput.*, 2018, 14 (3), pp 1682–1694; DOI: 10.1021/acs.jctc.7b00954

## 謝辞

修士課程から 5 年間にわたりご指導いただきました高田彰二教授に心より感謝いたします。また、筆者との共同研究を進めてくださり、貴重な実験データを提供いただいた九州大学の片山勉教授、川上広宣助教、野口泰徳博士、崎山友香里様に心より御礼申し上げます。さらに、片山研究室の皆様には、生化学実験に理解のない筆者の意見を辛抱強く聞いてくださった上に、たくさんの有益な助言をいただきました。ありがとうございました。

最後に、コンピュータの使い方をはじめ様々な場面で助けていただいた研究室の皆様がこの場を借りてお礼申し上げます。

## 付録

### 付録 A 第一章の研究手法の詳細

#### フレキシブル・ローカルポテンシャル

このポテンシャルは、二次構造をとっていないポリペプチド鎖の柔軟さを表現するモデルである<sup>41</sup>。このポテンシャルはアミノ酸配列上で連続した3アミノ酸粒子の結合角ポテンシャルと、連続した4粒子の二面角ポテンシャルからなる。いずれも統計ポテンシャルであり、以下の式で与えられる。

$$V_{loc}^{flp} = - \sum_{l=1}^{N-2} k_B T \ln \left( \frac{P_{ba}(\theta_l)}{\sin \theta_l} \right) - \sum_{l=1}^{N-3} k_B T \ln(P_{dih}(\varphi_l))$$

ここで、 $k_B$ はボルツマン定数である。温度  $T$  は 300.0 K としている。 $\theta_l$  は  $l$  から  $l+2$  番目までの3アミノ酸粒子のなす角である。 $\varphi_l$  は  $l$  から  $l+3$  番目までの4アミノ酸粒子のなす二面角である。 $P_{ba}(\theta)$  や  $P_{dih}(\varphi)$  は対応する角度の確率分布のヒストグラムである。これはプロテインデータバンク (<http://www.rcsb.org/pdb/>) に登録されているタンパク質構造データのループ領域での確率分布である。また、この確率分布はアミノ酸の種類ごとに計算されている。

まず  $\theta$  に関しては、このヒストグラムを3次のスプライン補完することで連続な  $P_{ba}(\theta)$  のポテンシャル関数を構築している。十分なサンプルがない領域では、サンプルがある領域から離れるにつれ、ポテンシャルは直線状に増大する。

つぎに  $\varphi$  に関しては、ヒストグラムは次のフーリエ級数への当てはめがされている。

$$f_{dih}^{flp}(\varphi) = \sum_m^3 k s_m \sin(m\varphi) + \sum_n^3 k c_n \cos(n\varphi) + C$$

ここで、 $k s_m$ ,  $k c_n$ ,  $C$  は定数である。

#### 粗視化シミュレーションの参照構造の準備

ホモロジーモデリングを利用することで、粗視化計算の参照構造を用意した。分子の重ね合わせなどにはソフトウェア PyMOL を利用した<sup>53</sup>。また、エネルギー最小化は最急降下法によって行った。ここでは GROMACS 4.5.5 を用いた<sup>54-58</sup>。この時シミュレーションの溶液条件は、200 mM の NaCl とした。

#### 大腸菌 *E. coli* の DnaA ドメイン III-IV

まず大腸菌 *E. coli* の ATP-DnaA ドメイン III-IV の参照構造を用意した。ドメイン IV の構造は既知であるため、まず I-TASSER ウェブサーバを使い、ドメイン III のホモロジーモデリングを行った<sup>45-47</sup>。このとき、構造テンプレートとして *T. maritima* DnaA ドメイン III (PDB ID : 2Z4S) を指定し、モデリング時にこの構造をより参照するようにした<sup>11</sup>。続いて I-TASSER で出力された *E. coli* DnaA ドメイン III ホモロジーモデルと DnaA ドメイン IV の結晶構造<sup>25</sup> をテンプレートに指定し、今度は I-TASSER で *E. coli* DnaA ドメイン III-IV のホモロジーモデルを作成した。ホモロジーモデリングの結果、*Aquifex aeolicus* の ADP 結合型 DnaA ドメイン III-IV によく似た構造が得られた。I-TASSER のモデリング結果にはリガンド結合位置予測が含まれている。筆者は ACP 結合 DnaA ドメイン III-IV の ACP の炭素原子を酸素原子に置換し、ATP 結合型 DnaA ドメイン III-IV モデルとした。ACP は、ATP の 2 番目のリン酸の酸素が炭素に置き換わった分子である。さらに、*Aquifex aeolicus* DnaA ドメイン III-IV の結晶構造(PDB ID : 3R8F)を参考に<sup>10</sup>、ATP 近傍にマグネシウムイオンを配置した。

こうして出来上がった *E. coli* DnaA ドメイン III-IV 単量体から、DnaA 四量体の構造を作成した。ここでは *A. aeolicus* DnaA ドメイン III-IV と一本鎖 DNA の X 線結晶構造 (PDB ID : 3R8F) を利用した。この結晶構造(四量体)の *A. aeolicus* DnaA に、作成した *E. coli* DnaA モデルを重ね合わせた後で、*A. aeolicus* DnaA を除去した(つまり置換した)。最後にエネルギー最小化を行った。得られた構造のうち内部 2 分子の DnaA ドメイン III の結合状態を DnaA 間相互作用の参照構造とした。内部の DnaA の片方を ATP-DnaA ドメイン III-IV 単量体の参照構造とした。

### 大腸菌 *E. coli* DnaA ドメイン IV と二本鎖 DNA 複合体の参照構造

続いて *E. coli* DnaA ドメイン IV と二本鎖 DNA 複合体の参照構造を用意した。この複合体の X 線結晶構造は既に報告されている<sup>25</sup>。シミュレーションでは *E. coli* DnaA ドメイン III-IV の参照構造にホモロジーモデルを用いる。そこでホモロジーモデルのドメイン IV を結晶構造に重ね合わせ、結晶構造をホモロジーモデルに置換した。

参照構造は DnaA ボックスごとに用意した。各 DnaA ボックスと一致するように DNA の塩基配列を変更した。塩基配列の変更は以下の手順で行った。

まず AmberTools13 のソフトウェア NAB を用いて塩基 (アデニン、シトシン、グアニン、チミン) の構造モデルを生成した<sup>49,50</sup>。続いて、塩基の 3 原子を使った重ね合わせを行った。この 3 点はプリン塩基では N9, C4, C8 原子、ピリミジン塩基では N1, C2, C6 原子である。結晶構造の 3 原子と NAB のモデルの 3 原子を重ね合わせた後、結晶構造の塩基を除去した。

最後にエネルギー最小化を行った。この複合体の DNA のうち、DNA-DnaA ドメイン IV 結合部の参照構造には 9 塩基対の DnaA ボックス部分のみ用いた。



## 大腸菌 *E. coli* IHF と DNA 複合体の参照構造

最後に *E. coli* IHF と二重鎖 DNA の複合体構造を用意した。既知の結晶構造(PDB ID : 1IHF)を利用し、それを編集した。まずタンパク質を編集した。I-TASSER サーバを用い、既知構造 (PDB ID : 1IHF<sup>24</sup>)の欠損アミノ酸が補完されている構造を得た<sup>45-47</sup>。I-TASSER では、*E. coli* IHF の結晶構造 (PDB ID : 1IHF)を鋳型に指定した。I-TASSER から出力されたモデルを結晶構造(PDB ID : 1IHF)の IHF タンパク質に重ね合わせ、置換した。

続いて DNA を編集した。結晶構造に含まれるニックの部分を変更した。この結晶構造にはニックが存在するが、*oriC* の該当箇所にはニックを仮定しない。結晶構造中、ID が C の DNA 鎖の-38 から-34 の残基番号の領域、ID が E の DNA 鎖の 34 から 38 の残基番号の領域を参照し、ニックのない構造を作成した。また、DnaA ドメイン IV の時と同じ要領で塩基配列を *oriC* の対応する領域のものに変更した<sup>4,24,135</sup>。

最後に結晶構造中のカドミウムイオンをマグネシウムイオンに変更し、エネルギー最小化を行い、参照構造を得た。

## 粗視化分子シミュレーションの詳細

### DnaA ドメイン IV と DNA の結合強さのパラメータ決定

解離定数の実験値に基づき、DnaA ドメイン IV-DNA の相互作用パラメータを決定した。検討したのは接触粒子間ポテンシャルの  $\epsilon^{IJ}$  の値である。 $\epsilon^{IJ}$  は R1, R4, R2 ボックスと DnaA ドメイン IV の  $\epsilon^{IJ}$  (以降  $\epsilon_{strong}$  と呼ぶ。) と、それ以外の DnaA ボックスと DnaA ドメイン IV の  $\epsilon^{IJ}$  (以降  $\epsilon_{weak}$  と呼ぶ。) の 2 種を考え、それぞれを検討した。

DnaA ドメイン IV (L373-S467) と 21 塩基対の DNA (5'-TCCTTGTTATCCACAGGGCAG-3')を系に含めた。R1 ボックスには下線が引いてある。分子を半径 100 Å の球内に配置し、各分子には幾何重心がそこから出るのを抑えるポテンシャルを与えた。チミン塩基を持つ DNA 末端の糖粒子には球の中心にとどまるポテンシャルを与えた。ここでは  $10^9$  MD ステップのシミュレーションを実施した。0.015 kcal/mol の幅で  $\epsilon^{IJ}$  の値を変更した。各条件で 2 回ずつシミュレーションを実施した。座標とエネルギーは  $10^4$  MD ステップごとに出力した。

シミュレーション中での DNA と DnaA ドメイン IV の解離定数を見積もった。初期構造の影響を考慮し、最初の  $10^7$  MD ステップのデータは解析対象から除外した。DNA と DnaA ドメイン III 相互作用のエネルギーの和が -0.3 kcal/mol 以下となったとき、両者が結合しているとみなした。解離定数  $K_d$  を以下の式によって定義した。

$$r_{結合} = N_{結合} / N_{総計}$$

$$c_{\text{結合}} = \frac{r_{\text{結合}}}{N_A} \times \frac{1}{\frac{4}{3} \times \pi \times (R)^3}$$

$$c_{\text{非結合}} = \frac{(1 - r_{\text{結合}})}{N_A} \times \frac{1}{\frac{4}{3} \times \pi \times (R)^3}$$

$$K_d = \frac{c_{\text{非結合}}^2}{c_{\text{結合}}}$$

上式で、 $N_{\text{結合}}$ は DnaA ドメイン IV と DNA が結合していたスナップショットの数を表す。 $N_{\text{総計}}$ は 9,9000 構造である。球の半径  $R$  は 100 Å とした。 $N_A$  はアボガドロ数である。結果は付録 B 表 SB-4 にまとめた。報告されている解離定数<sup>19</sup>と比較し、 $\epsilon_{\text{strong}} = 0.36$  kcal/mol,  $\epsilon_{\text{weak}} = 0.315$  kcal/mol に決定した。

## 複製開始複合体のシミュレーション

### シミュレーション 1 : DnaA 単量体のダイナミクス

ヘリックス境界モデル、ループ境界モデルそれぞれで、 $10^8$  MD ステップのシミュレーションを行った。ランジュバンダイナミクスの乱数の種を変えながら、それぞれ 5 回のシミュレーションを行った。シミュレーションの初期構造は ATP-DnaA ドメイン III-IV の参照構造とした。

### シミュレーション 2 : DnaA ドメイン IV と DnaA ボックスの相互作用解析シミュレーション

ループ境界モデルの DnaA ドメイン III-IV と 13 塩基対の DNA (5'-CCTGTGTGGATAA-3'; 下線部が R1 ボックス)を系に含めたシミュレーションを行った。シミュレーションの初期構造では、DnaA ドメイン IV は R1 ボックスに結合している。この構造は ATP-DnaA ドメイン III-IV の参照構造を DnaA ドメイン IV-二本鎖 DNA 複合体構造 (PDB ID : 1J1V)に重ね合わせ、既知構造の DnaA ドメイン IV を除去したものである。DnaA ドメイン IV と DNA の接触相互作用ポテンシャルの  $\epsilon^{IJ}$  は  $\epsilon_{\text{strong}}$  とした。ランジュバンダイナミクスの乱数の種を変えながら、5 回シミュレーションを行った。 $10^4$  MD ステップごとに座標とエネルギーを記録した。

### シミュレーション 3 : 左側サブ複合体の形成シミュレーション

左側サブ複合体のシミュレーションでは ATP-DnaA を 5 分子、IHF を一組、102 塩基

対の *oriC* 左側領域を含む (付録 B 表 SB-5 参照)。複合体形成シミュレーションの初期構造は以下の 2 ステップで作成した。

#### [ステップ 1]

ステップ 1 では DNA にタンパク質を結合させた。

まず *oriC* 左側領域に IHF が結合した構造モデルを用意した。粗視化 DNA 断片と、粗視化した IHF-DNA 複合体の参照構造をつなげて作成した。DNA 断片はソフトウェア NAB を利用し作成した<sup>49,50</sup>。粗視化した DNA 断片と粗視化 IHF-DNA 複合体の共通部分を重ね合わせ、後から余分な粗視化粒子を除去した。重ね合わせには PyMOL を利用した<sup>53</sup>。

続いて DnaA ドメイン III-IV を R1 ボックスに結合させた。DnaA 分子を R1 ボックスの近くに配置し、 $5 \times 10^6$  MD ステップの MD シミュレーションにより R1 ボックスに結合させた。この際、 $\epsilon^H = 3.0$  kcal/mol とした。DnaA はヘリックス境界モデルを用いた。ランジュバンダイナミクスの乱数の種を変え、10 回の独立したシミュレーションを行った。その中から DnaA ドメイン IV が R1 ボックスに結合しているものを探した。続いて R1 ボックスに DnaA が結合した構造を初期構造とし、同じ要領で R5M ボックスに DnaA を結合させた。同様にして、 $\tau 2$  ( $\tau 2l$  または  $\tau 2r$ ), I1, I2 の順に DnaA ボックスに DnaA を結合させた。R5M から I2 ボックスに DnaA を結合させる過程でも、5 回または 10 回の独立したシミュレーションを行い、その中で DnaA が結合している構造を選んだ。このステップでは IHF のすべての粒子は固定されており、動かない。またこの段階では DnaA ドメイン III 同士に引力は働かない。

#### [ステップ 2]

続くステップ 2 では、本計算の初期構造として複数の異なる構造を用意した。ステップ 1 で得た構造を初期構造とし、 $10^7$  MD ステップのシミュレーションを行った。独立したシミュレーションを各設定で 10 回ずつ、ランダム力の乱数の種を変えながら行った。DnaA の DNA からの解離を抑える目的で  $\epsilon^H$  はすべて  $\epsilon_{strong}$  に設定した。ループ境界モデルまたはヘリックス境界モデル、 $\tau 2l$  に結合させた場合か  $\tau 2r$  に結合させた場合かの  $2 \times 2 = 4$  種の設定についてそれぞれシミュレーションした。系の並進運動を抑えるため、IHF の 3 つのアミノ酸 (IHF $\alpha$  P65, IHF $\beta$  H16 and R62) を固定した。この段階でも DnaA ドメイン III 間の引力は切っている。ここで得た終構造を複合体形成シミュレーションの初期構造とした。

#### [複合体形成シミュレーション]

シミュレーションを各設定で 10 回ずつ、ランダム力の乱数の種を変えながら行った。各シミュレーションは  $10^8$  MD ステップ行った。DnaA ドメイン III 同士の接触ポテンシ

ヤルを与え、多量体を形成させた。DnaA ドメイン IV-DNA 相互作用の  $\epsilon^{IJ}$  は、DnaA ボックスごとに  $\epsilon_{strong}$  または  $\epsilon_{weak}$  を用いた。系の並進運動を抑えるため、IHF の 3 つのアミノ酸(IHF $\alpha$  P65, IHF $\beta$  H16 and R62)を固定した。先行研究に基づき<sup>30</sup>、DnaA ドメイン III のホモ多量化は以下の組み合わせで形成するよう設定した。

	アルギニンフィンガー側で結合	ATP 側で結合
ドメイン III ペア 1	R1 ボックス上 DnaA	R5M ボックス上 DnaA
ドメイン III ペア 2	R5M ボックス上 DnaA	$\tau$ 2 ボックス上 DnaA
ドメイン III ペア 3	$\tau$ 2 ボックス上 DnaA	I1 ボックス上 DnaA
ドメイン III ペア 4	I1 ボックス上 DnaA	I2 ボックス上 DnaA

#### シミュレーション 4: 中間-右側サブコンプレックスの形成シミュレーション

初期構造は左側サブ複合体の時と同様に作成した。使用した DNA は R2 ボックスから R4 ボックスまでの領域である(付録 B 表 SB-5)。ステップ 1 では DnaA を R4, C1, I3, C2, C3, R2 ボックスの順に結合させた。本計算はループ境界モデルでのみ行うため、ステップ 2 はループ境界モデルのみで実施した。ステップ 1,2 ともに R2 側の DNA 末端塩基対を固定した。

複合体形成シミュレーションも左側サブ複合体と同様、 $10^8$  MD ステップのシミュレーションを 10 回行った。互いに結合する DnaA ドメイン III のペアは以下にまとめた。

	アルギニンフィンガー側で結合	ATP 側で結合
ドメイン III ペア 1	R4 ボックス上 DnaA	C1 ボックス上 DnaA
ドメイン III ペア 2	C1 ボックス上 DnaA	I3 ボックス上 DnaA
ドメイン III ペア 3	I3 ボックス上 DnaA	C2 ボックス上 DnaA
ドメイン III ペア 4	C2 ボックス上 DnaA	C3 ボックス上 DnaA
ドメイン III ペア 5	C3 ボックス上 DnaA	R2 ボックス上 DnaA

ここでは 2 種類のシミュレーションを行った。片方の設定ではすべてのドメイン III ペアで引力が働くが、もう片方のシミュレーションではドメイン III ペア 5 では引力が働かない。よって両ドメイン III は結合しない。

#### シミュレーション 5: *oriC* 左側領域変異体の複合体形成シミュレーション

*oriC* 左側領域変異体の DnaA ホモ多量体形成をシミュレーションした。用いた変異体は R1 ボックスと IHF サイトの間の 5 塩基対欠失変異体である(付録 B 表 SB-5, R1IBS-d5 変異体)。系にはこの *oriC* 左側領域変異体、IHF, 5 分子の DnaA ドメイン III-IV を含め

た。はじめ、DnaA ドメイン III 同士は互いに離れている。まず、初期構造作成のステップ 2(シミュレーション 3 参照)を行い、10 種類の異なる構造を用意した。続いてシミュレーション 3 と同様の条件で、DnaA のホモ多量体形成を調べた。DnaA ドメイン III-IV はすべてループ境界モデルを用いた。

#### シミュレーション 6: *oriC* 右側変異体の DnaA 複合体形成シミュレーション

ここでは 5 種類の *oriC* 右側領域変異体を用意し、複合体形成シミュレーションを行った。変異体はそれぞれ R2C3Δ5, R2C3Δ10, R2C3Δ18, R2C3+4, R2C3+10 と名付けてあり、塩基配列は付録 B 表 SB-5 にまとめてある。設定はシミュレーション 4 と同様であり、2 タイプの複合体形成シミュレーションを行った。

#### シミュレーション 7: 中間-右側サブ複合体の挙動解析

シミュレーション 4 と 6 で得たサブ複合体を初期構造にして、さらに  $2 \times 10^8$  MD ステップのシミュレーションを行った。ランダム力の乱数の種を変えながら野生型 *oriC*, あるいは変異体ごとに 9 回ずつシミュレーションを行った。設定はシミュレーション 4 や 6 と同様である。このシミュレーションの解析の際は、初めの  $5 \times 10^7$  MD ステップのデータは除外した。

#### シミュレーション 8: 全原子シミュレーションの初期構造の準備 (ステップ 1)

シミュレーション 3 と 4 で得た左側サブ複合体と中間-右側サブ複合体を接続した。シミュレーション 3 や 4 と同じ設定で、この複合体に対し  $5 \times 10^5$  MD ステップのシミュレーションを実施した。

#### シミュレーション 9: 全原子シミュレーションの初期構造の準備 (ステップ 2)

徐冷法を行い、粗視化粒子の結合長や結合角を緩和した。計算は  $10^4$  MD ステップ実施した。系の温度を 1.495 K ずつ、同じ時間間隔で低下させた。計 200 回温度を下げることで、はじめ 300 K であった系を 1 K まで冷却した。それ以外の設定は複合体形成のシミュレーションと同じである。シミュレーションは、左側サブ複合体と中間-右側サブ複合体とをつなげた複合体全体に対して行った。

あわせて、*oriC* 右側領域で DnaA ドメイン III ホモ 6 量体を形成した複合体を一つ選び、同様の徐冷法を実施した。

徐冷法の終構造から全原子モデルを構築した。

#### 全原子シミュレーションの初期構造の作成

粗視化モデルから原子モデルを構築した。DNA では、各粗視化ヌクレオチド(糖・リン酸・塩基の組)ごとに原子モデルを重ね合わせた。重ね合わせには粗視化粒子の定義

に基づき、糖の重心、リン酸の重心、塩基の N1 または N5 原子の位置を用いた。使った原子モデルは NAB で生成した B-型 DNA である<sup>49,50</sup>。また、5'末端のヌクレオチドでは、糖の重心、塩基の N1 または N5 原子、隣接するヌクレオチドのリン酸の重心を重ね合わせに用いた。

ATP-DnaA ドメイン III-IV の原子モデルを構築するにあたっては、以下の 3 種を用意した。

- ・大腸菌 *E. coli* DnaA ドメイン IV の結晶構造 (PDB ID : 1J1V)
- ・Phyre2 サーバの one to one threading mode によって生成した DnaA ドメイン III-IV の原子モデル<sup>51</sup>。計 11 分子の DnaA それぞれについて、粗視化モデルを鋳型とし、生成した。
- ・ATP・Mg イオン結合 DnaA ドメイン III(K135-L366)の 5 または 6 量体。これらは DnaA ドメイン III-IV 参照構造作成時の ATP・Mg イオン結合 DnaA ドメイン III-IV の 4 量体を利用して作成した。

これらを組み合わせ、モデルを構築した。まず Phyre2 のモデルを、C $\alpha$  原子の座標を使い粗視化モデルに重ねた。次にドメイン III の 5 (6)量体をこの Phyre2 モデルに重ね合わせた。L366 付近のアミノ酸のなかで、重ね合わせた 2 構造のずれが小さいアミノ酸を第一の境界とした。DnaA ドメイン III の K135 からその境界残基までは、ドメイン III の 5(6)量体モデルを原子モデルとした。

続いて Phyre2 モデルに DnaA ドメイン IV (V374-S467)結晶構造<sup>25</sup>を重ね合わせた。V374 付近にあり、重ね合わせた 2 構造のずれが小さいアミノ酸を第二の境界に定めた。第一の境界残基の次のアミノ酸から、第二の境界残基までは Phyre2 のモデルを採用した。残りの残基はドメイン IV の結晶構造を採用した。結晶構造中のセレノメチオニンは PyMOL を用いてメチオニンに変更した。また、立体障害により全原子シミュレーションに支障をきたすアミノ酸側鎖は、SCWRL4<sup>52</sup>または PyMOL を用いて構築しなおした。

IHF は、粗視化モデルに参照構造を重ね合わせることで原子モデルに変換した。

結局、複合体は全原子計算を行うには大きすぎたため、左側部分と中間-右側部分の 2 領域に分割し、それぞれで全原子シミュレーションを実行した。左側部分は R1 ボックスから I2 ボックスまでが含まれている。中間-右側部分は R2 ボックスから R4 ボックスまでの部分を含む (付録 B 表 SB-5)。

### 全原子シミュレーション後の左側サブ複合体と中間-右側サブ複合体の接続

左側サブ複合体と中間-右側サブ複合体の DNA には 5 塩基対の共通部分を用意していた。二つの複合体をこの共通部分で重ね合わせた。重ね合わせにはソフトウェア PyMOL を利用した。

## シミュレーション結果の解析

分子構造は VMD<sup>136</sup> や PyMOL<sup>53</sup> を用いて可視化した。図の作成には PyMOL, inkscape<sup>137</sup>, gnuplot<sup>138</sup> を用いた。

### 複合体形成の達成度

複製開始複合体の分子間相互作用の達成度の基準として、 $Q_{III-III}$ ,  $Q_{box}$  を定義した。これらは参照構造で見られる DnaA ドメイン III 間、または DnaA ドメイン IV-DNA の相互作用の再現率である。参照構造中で、異なる DnaA ドメイン III に属する 2 つの重原子間の距離が 6.5 Å 以内である場合に、重原子の属するアミノ酸ペアは”天然構造で接触している”と定義した。同様に、参照構造で DNA の重原子と DnaA ドメイン IV の重原子の距離が 6.5 Å 以内のとき、対応する粗視化粒子ペアは”天然構造で接触している”とした。シミュレーション中に、このペアの距離が天然(参照)構造での距離の 1.2 倍以内であったとき、接触を再現したとみなす。すべての接触ペアのうち、再現しているペアの割合が Q 値である。それぞれの DnaA ドメイン III 間相互作用面では  $Q_{III-III}$  と呼ぶ。またそれぞれの DnaA ボックスの DnaA ドメイン IV-DNA の Q 値を  $Q_{box}$  と呼ぶことにする。以下の条件を満たすとき、サブ複合体を形成したとみなした。

- 1) すべての DnaA ドメイン III 間相互作用面について  $Q_{III-III} > 0.99$
- 2) すべての DnaA ボックスで  $Q_{box} > 0.1$

筆者たちはまた、 $N_{pp}$  と  $N_{pd}$  を定義した。まず  $N_{pp}$  は  $Q_{III-III} > 0.99$  となっている DnaA ドメイン III 相互作用面の数である。そして、 $N_{pd}$  は  $Q_{box} > 0.1$  となっている DnaA ボックスの数である。

シミュレーション 4 で DnaA ドメイン III を 6 量化させる設定の際は、以下の条件を満たしたとき R2 ボックス以外の DnaA が 5 量体を形成しているとみなした。

- 1) R2 上の DnaA と C3 上の DnaA の相互作用面では  $Q_{III-III} = 0$
- 2) それ以外の DnaA ドメイン III 間相互作用面で  $Q_{III-III} > 0.99$
- 3) すべての DnaA ボックスで  $Q_{box} > 0.1$

図 1-3B, 1-4, 1-6, 付録 B 図 SB-1 から SB-6 ではサブ複合体形成条件を満たしたスナップショットのみを解析の対象とした。シミュレーション 2 のトラジェクトリでは、 $Q_{box} > 0.1$  を満たすものを解析の対象にした。

### DnaA ボックス上の DnaA ドメイン IV の向き

DnaA ドメイン IV と  $\tau 2$  ボックスの結合の向きの指標として  $\theta_{DnaA-\tau 2}$  ( $0^\circ < \theta_{DnaA-\tau 2} < 180^\circ$ ) を定義した。これは 2 つのベクトルのなす角である。片方は DnaA の L422 から A404 に向かうベクトルである。もう片方は DnaA の  $\tau 2I$  ボックスに沿ったベクトルである。

後者のベクトルの始点は  $\tau 21$  ボックス(5'-AGgatccc-3')の端の 2 塩基対(大文字で示した AG)の糖・塩基粒子の幾何重心である。終点は  $\tau 21$  ボックスの 2 塩基対(下線を引いた cc)の糖・塩基の幾何重心である。

同様に  $\tau 2$  ボックス以外でも  $\theta_{DnaA-box}$  を定義し、付録 B 図 SB-4 で用いた。この  $\theta_{DnaA-box}$  が  $90^\circ$  を超えるか否かで DnaA ドメイン IV の向きを区別した。

### DnaA ボックス間の DnaA タンパク質の構造比較

シミュレーション 3, 4 で現れたサブ複合体中での、DnaA の構造を解析した。異なる DnaA ボックスの DnaA は、互いに異なる構造をとっているのかを調べた。はじめに、すべての DnaA ボックスについて以下の操作を行い、DnaA の構造セットを作成した。

1. 複合体を形成しているすべてのスナップショットから、該当する DnaA ボックスに結合している DnaA ドメイン III-IV の構造を集めた。
2. 集めた構造それぞれを、ドメイン IV の I376-S467 部分で参照構造に重ね合わせた。

続いて DnaA ボックスのペアを一つ決め、構造を比較した。両者の DnaA 構造セットから一構造ずつ取り出し、ドメイン III の幾何重心間の距離を計算した。これを 2 つの構造セット間のすべてのペアで行い、平均値を出した。この指標は図 1-4B で利用した。

### 変異導入による右側サブ複合体の向きの変化

*oriC* 右側領域に変異を導入したことによる DnaA 多量体の向きの変化を、指標  $\varphi_{complex}$  ( $0^\circ < \varphi_{complex} < 180^\circ$ ) を定義し、定量した。この解析にはシミュレーション 7 の結果を用いた。

まず次のようにして複合体の向きを定義した。最初にベクトル  $\mathbf{r4}$  を定義した。このベクトルの始点は I3 ボックスに結合した DnaA の K135-L366 領域の幾何重心である。終点は R4 ボックスに結合した DnaA の K135-L366 の幾何重心である。同様に、ベクトル  $\mathbf{c3}$  を定義した。最後に  $\mathbf{r4}$  と  $\mathbf{c3}$  の外積を求めた。これを複合体の向きと定義した。

それぞれの変異体に対し、角度  $\varphi_{complex}$  を以下の操作で求めた。変異体の右側複合体と野生型の右側サブ複合体で、二つの複合体の向きがなす角を求めた。シミュレーションで現れたスナップショットをすべて使い、変異体と野生型の間のすべての複合体ペアについて同様の計算をした。その加重平均を  $\varphi_{complex}$  と定義した。コントロールとして、野生型同士の比較も行った。

### 複合体の代表構造の選出

シミュレーション 3, 4, 5 では、次のようにして代表構造を選出した。まずトラジェク



トリごとに候補を一構造ずつ出した。複合体形成条件を満たしたスナップショットのペアで重ね合わせを行い、RMSD を計算した。これをすべてのペアについて行った。次にそれぞれの構造ごとに、ほかの構造との RMSD を足し合わせ、平均値を算出した。この平均値が最小となるものを、そのトラジェクトリからの代表構造候補とした。RMSD 計算には DnaA の K135-L366 と I376-S467 領域を用いた。

続いて、代表構造候補の中で同様の手順を踏み、代表構造を一つ選出した。

付録 B 図 SB-4 でも、同様の方法で代表構造を決めた。ただし構造をトラジェクトリごとに分けて、代表構造を直接選出した。

シミュレーション 7 では、DnaA ドメイン III ホモ 5 量体の K135-L366 部分の重心座標が、その加重平均に最も近いものを代表構造とした。

### グラフのエラーバー

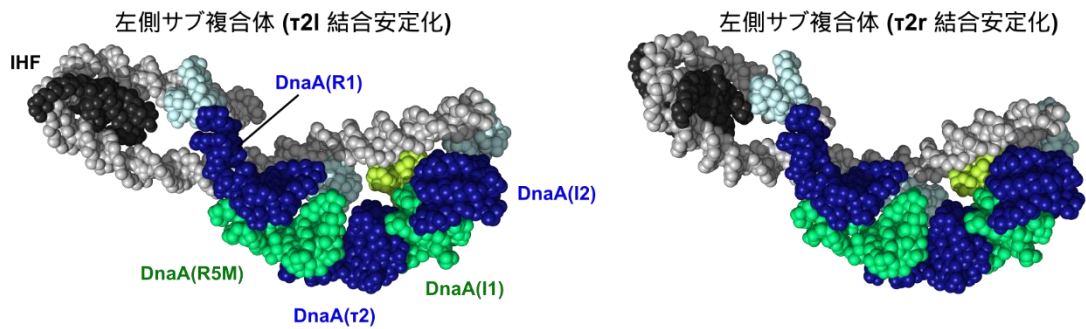
ヒストグラムのエラーバーはトラジェクトリごとのデータの標本標準偏差を示す。(図 1-3B, 1-4A, 付録 B 図 SB-3, SB-6)。標準偏差の計算の際には、サブ複合体形成条件を満たしたスナップショットの数によって、それぞれのトラジェクトリに重みづけをした。

図 1-6B では、エラーバーは 9 回の独立した試行の標本標準偏差である。

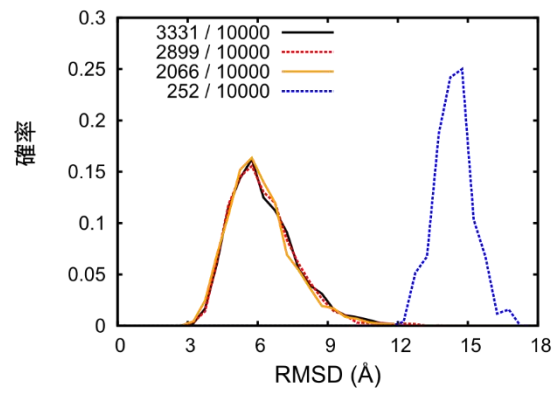
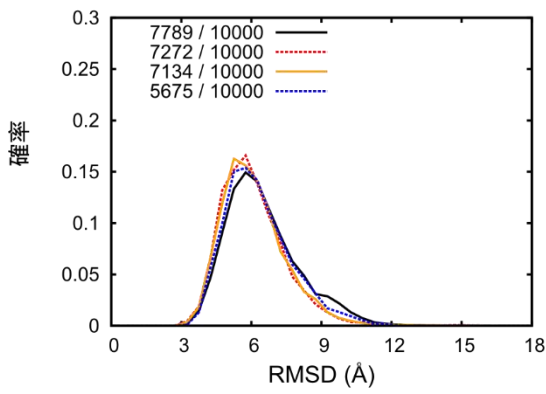
## 付録 B 複製開始複合体の詳細な構造解析と共同研究者の実験結果

ここでは、第一章のシミュレーション結果の詳細と、共同研究者の生化学実験結果をまとめた。生化学実験については、第一章でシミュレーション結果と直接比較したデータだけを示した。

A



B 左側サブ複合体 (τ2l 結合安定化)



左側サブ複合体 (τ2r 結合安定化)

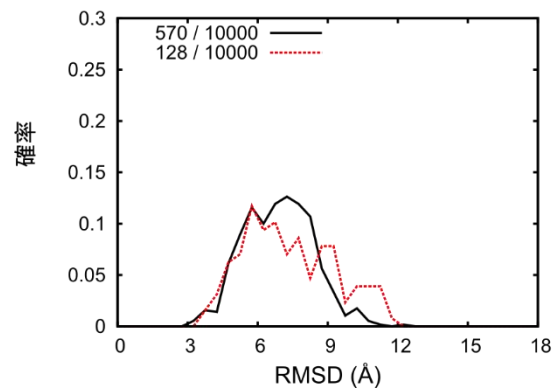
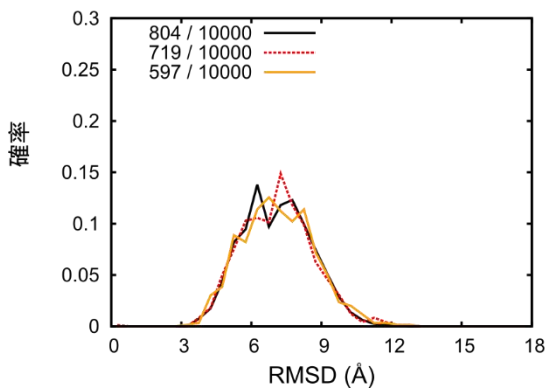
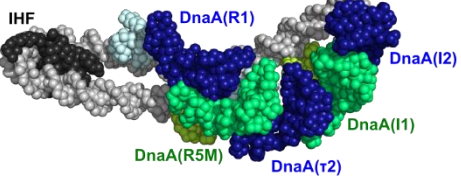


図 SB-1. 左側サブ複合体の構造多様性

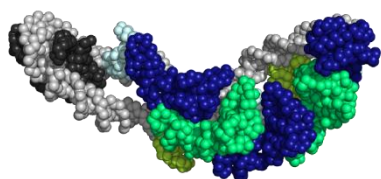
A) 左側サブ複合体の代表構造を示した。τ2 ボックスへの結合を τ2l に設定したものを左側、τ2r に設定したものを右側に示した。DNA は灰色で、IHF は黒で示した。ATP 結合型 DnaA は 2 通りの着色法を交互に使った。ドメイン III を濃い青色、ドメイン IV を水色で着色した場合と、ドメイン III を緑、ドメイン IV を黄緑で着色した場合である。B) 各複合体構造を代表構造に重ね合わせた後で、RMSD を計算した。トラジェクトリごとに RMSD の分布を示した。この操作は、複合体の形成条件を満たしたのものについてのみ行った。重ね合わせ、RMSD 計算は DnaA の K135-L366 と I376-S467 部分のみを用いて行った。すなわち、DnaA の柔軟なリンカー部分は除外した。上段の 2 つのパネルは τ2 ボックスへの結合を τ2l に設定したものであり、下段が τ2r に設定したものである。複合体形成が確認できたそれぞれ 8 つ、5 つのトラジェクトリは、視認性のために 2 つのパネルに分割して示してある。パネルの左上部には複合体形成条件を満たしたスナップショット数をグラフの色の横に示した。それぞれのシミュレーションで 10,000 スナップショットずつ構造を記録した。

A

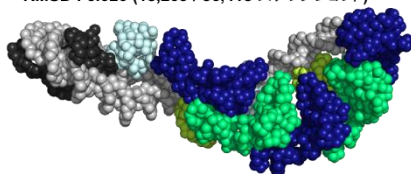
代表構造



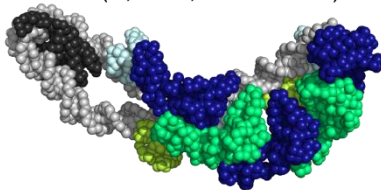
RMSD : 5.216 (9,105 / 36,418 スナップショット)



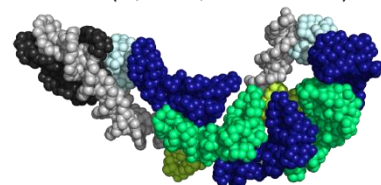
RMSD : 6.020 (18,209 / 36,418 スナップショット)



RMSD : 7.047 (27,314 / 36,418 スナップショット)

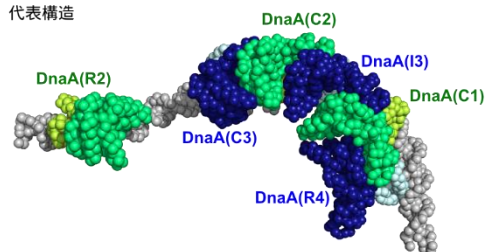


RMSD : 10.268 (35,690 / 36,418 スナップショット)

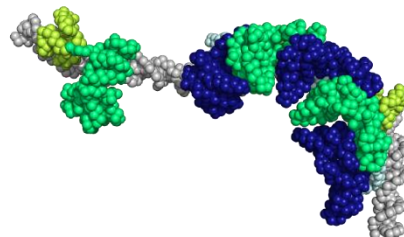


B

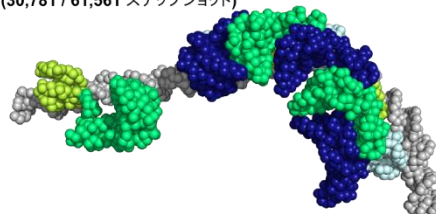
代表構造



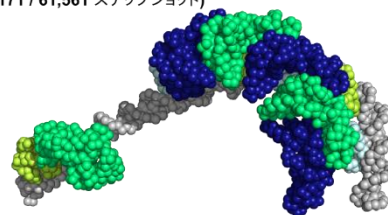
RMSD : 4.844 (15,391 / 61,561 スナップショット)



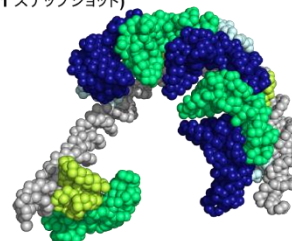
RMSD : 5.543  
(30,781 / 61,561 スナップショット)



RMSD : 6.419  
(46,171 / 61,561 スナップショット)



RMSD : 8.852  
(60,330 / 61,561 スナップショット)



**図 SB-2. 左側および中間-右側複合体のスナップショット。**複合体形成条件を満たした構造のいくつかを示した。左側サブ複合体は A, 中間-右側複合体は B に示してある。代表構造を最上段に示した。それ以外の構造については、代表構造に DnaA ホモ 5 量体の K135-L366, I376-S467 で重ね合わせた場合の RMSD を示した。複合体は RMSD の小さい側から順位を付けた。構造モデルにはこの RMSD と順位を併記した。構造モデルは DnaA ホモ 5 量体の K135-L366, I376-S467 の向きをそろえ、示してある。着色法は図 SB-1 に従う。シミュレーションで現れたサブ複合体の構造は互いに類似していた。

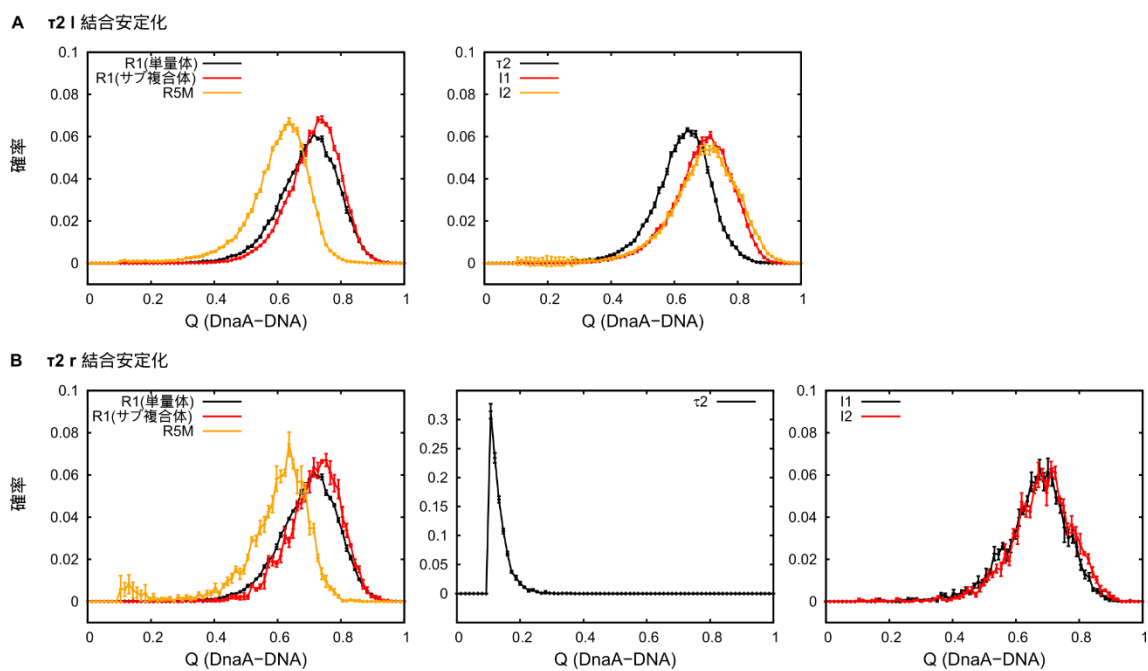
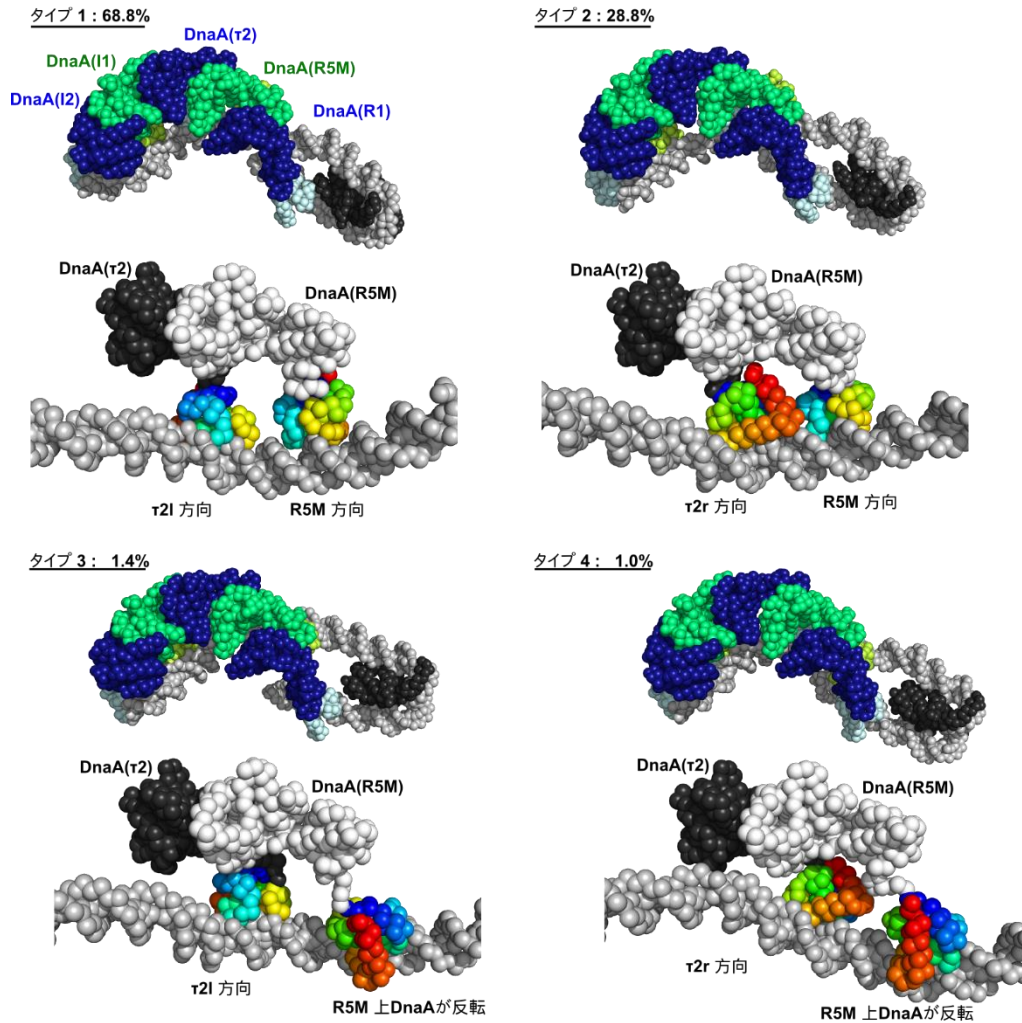
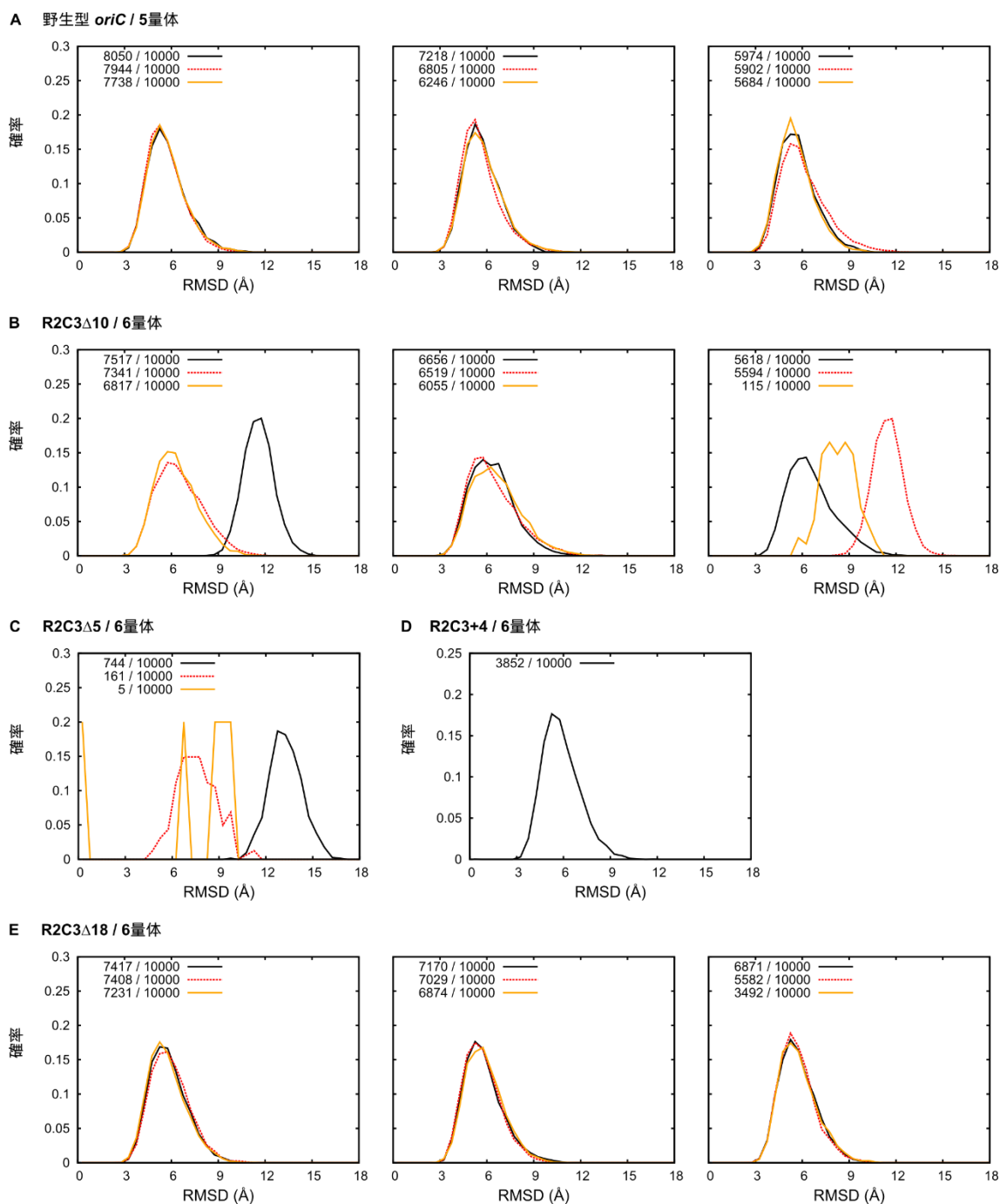


図 SB-3. 形成した左側サブ複合体中での DnaA ドメイン IV と DnaA ボックス結合の  $Q_{\text{box}}$ 。それぞれの DnaA ボックスと DnaA ドメイン IV 相互作用面の  $Q_{\text{box}}$  の分布を示した。比較のため、一分子の DnaA ドメイン III-IV と、R1 ボックスの間の  $Q_{\text{box}}$  の分布も示した(R1(単量体))。この時用いた DNA は 13 塩基対の長さである。A) DnaA の  $\tau 2$  ボックスへの結合を  $\tau 2$ I に設定した場合。B) DnaA の  $\tau 2$  ボックスへの結合を  $\tau 2$ r に設定した場合。エラーバーは標本標準偏差を示した。

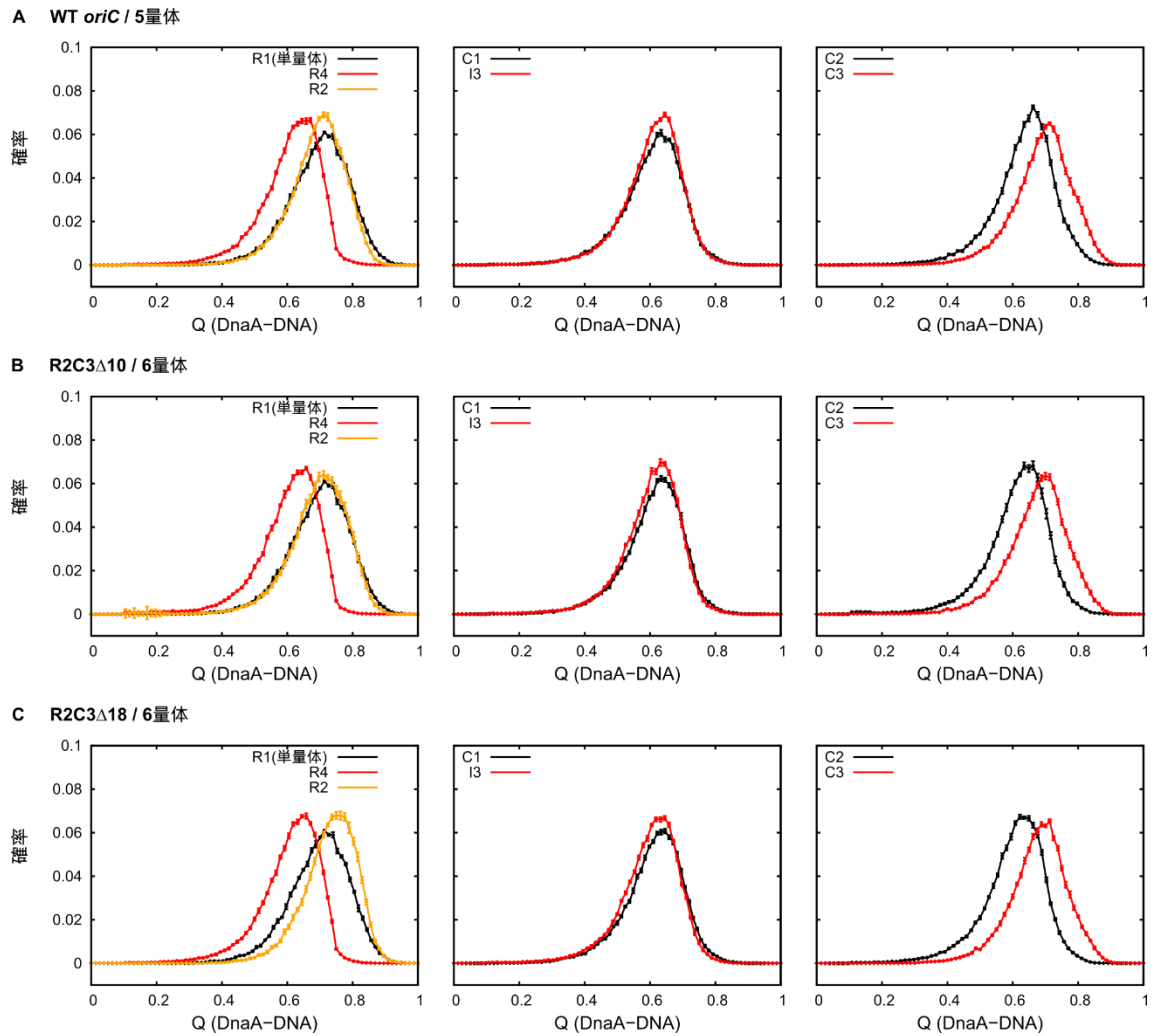


**図 SB-4. DnaA ドメイン IV 結合向きの変転** DnaA ドメイン IV の結合向きの変転から、DnaA の  $\tau 2$  ボックスへの結合を  $\tau 2r$  に設定した場合に現れた構造を 4 種類に分類した。**タイプ 1:** DnaA ドメイン IV の R5M ボックスへの結合の向きは本来の R5M の向き<sup>25</sup>であり、DnaA ドメイン IV の  $\tau 2$  ボックスへの結合向きは  $\tau 2l$  と同じである。これは総計 2,818 構造のうちの 68.8%であった。**タイプ 2:** DnaA ドメイン IV の R5M ボックスへの結合の向きは本来の R5M の向きであり、DnaA ドメイン IV の  $\tau 2$  ボックスへの結合向きは  $\tau 2r$  と同じである。これは総計 2,818 構造のうちの 28.8%であった。**タイプ 3:** DnaA ドメイン IV の R5M ボックスへの結合の向きは R5M から反転しており、DnaA ドメイン IV の  $\tau 2$  ボックスへの結合向きは  $\tau 2l$  と同じである。これは総計 2,818 構造のうちの 1.4%であった。**タイプ 4:** DnaA ドメイン IV の R5M ボックスへの結合の向きは R5M から反転しており、DnaA ドメイン IV の  $\tau 2$  ボックスへの結合向きは  $\tau 2r$  と同じである。これは総計 2,818 構造のうちの 1.0%であった。

それぞれのタイプの代表構造を上段に示した。着色の方法は図 SB-1 と同様である。R5M ボックスと  $\tau 2$  ボックスに結合した DnaA 部分を拡大したものを下段に示した。ここでは DNA を灰色、R5M ボックスの DnaA ドメイン III を黒、 $\tau 2$  ボックスの DnaA ドメイン III を白で着色した。DnaA ドメイン IV はそれぞれ V374(青)から S467(赤)に向かう虹色で示した。R5M ボックスの DnaA ドメイン IV が反転した場合や、DnaA ドメイン IV の  $\tau 2$  ボックスへ結合向きが  $\tau 2r$  と同じ場合には、DnaA ドメイン IV の C 末端領域が手前を向く(朱色-赤色の部分)。R5M、 $\tau 2$  以外の結合ボックスの DnaA ドメイン IV が反転した構造や、中間-右側サブ複合体で DnaA ドメイン IV が反転した構造は全体の 0.3%未満であった。

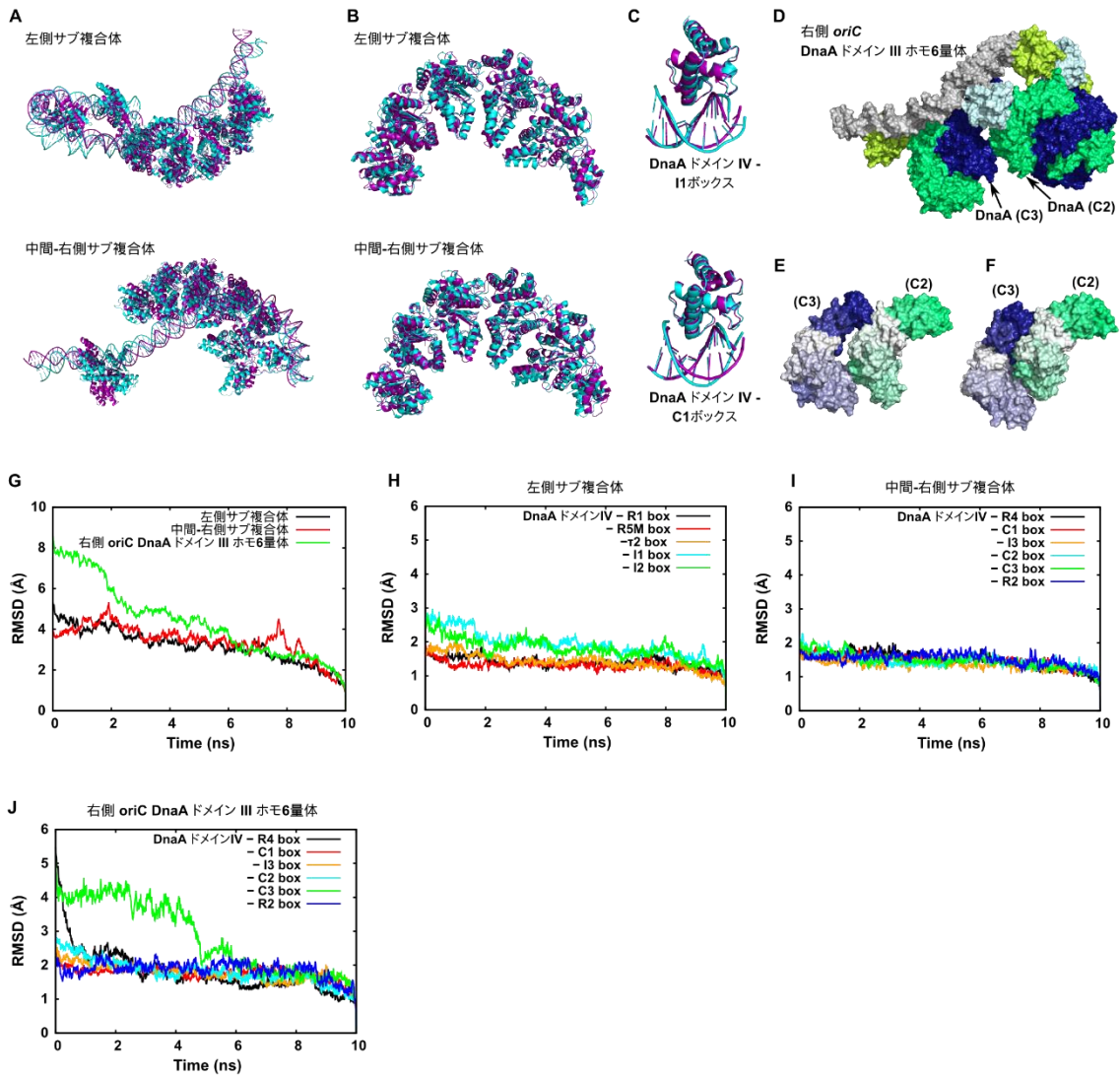


**図 SB-5.中間-右側サブ複合体の構造多様性** A-E) DnaA ホモ多量体形成シミュレーションの解析。各複合体構造を代表構造に重ね合わせた後で、RMSD を計算した。トラジェクトリごとに RMSD の分布を示した。この操作は、複合体の形成条件を満たしたものについてのみ行った。それぞれのヒストグラムは、複合体形成が見られたトラジェクトリごとにプロットしたものである。野生型、R2C3Δ10 (B), R2C3Δ18 (C)変異体では、視認性のために3つのパネルに分けて示した。パネルの左上には複合体形成条件を満たしたスナップショット数をグラフの色の横に示した。それぞれのシミュレーションでは 10,000 スナップショットずつ記録した。重ね合わせ、RMSD 計算は DnaA の K135-L366 と I376-S467 部分のみを用いて行った。すなわち、DnaA の柔軟なリンカー部分は除外した。



図SB-6. 中間-右側サブ複合体中での DnaA ドメイン IV と DnaA ボックス結合の  $Q_{\text{box}}$ 。野生型 *oriC*, *oriC* 変異体それぞれで、DnaA ボックスと DnaA ドメイン IV 相互作用面の  $Q_{\text{box}}$  の分布を示した。比較のため、一分子の DnaA ドメイン III-IV と、R1 ボックスの間の  $Q_{\text{box}}$  の分布も示した(R1(単量体))。この時用いた DNA は 13 塩基対の長さである。エラーバーは標本標準偏差を示した。





**図SB-7. 全原子分子動力学計算の間に起こったサブ複合体の構造変化** 全原子分子動力学計算は10 ナノ秒行った。A-C) シミュレーション前の構造(紫色)とシミュレーション後の構造(水色)の比較。A) 左側サブ複合体(上段)と中間-右側サブ複合体(下段)。DNA とドメイン III-IV 間のループ領域が柔軟に構造変化できるため、サブ複合体の構造はある程度は変化した。ただ、複合体全体の概形には大きな変化は見られなかった。B) ATP-DnaA ホモ 5 量体の K135-L366 部分を図示した。上段が左側サブ複合体で、下段が中間-右側サブ複合体である。この領域は、柔軟なループ領域に設定しなかったドメイン III 領域である。いずれの場合にも、構造はほとんど変化しなかった。C) DnaA ドメイン IV の I376-S467 と DnaA ボックス。DnaA のこの領域は柔軟なループ領域に設定しなかったドメイン IV 領域である。上段には DnaA ドメイン IV と II ボックスの複合体を図示した。下段には DnaA ドメイン IV と C1 ボックスの複合体を図示した。この場合も、全原子シミュレーション前後でほとんど構造は変わらなかった。構造が変わらないことは、この複合体がある程度は安定な構造であることを示唆している。D-E) *oriC* 右側領域に形成した DnaA ドメイン III ホモ 6 量体の全原子シミュレーション。この複合体は粗視化シミュレーション中にはほぼ現れなかったが、安定性を全原子分子計算によっても検証した。D) シミュレーション後の複合体の全体像。C3 ボックスに結合した DnaA と C2 ボックスに結合した DnaA を矢印で示した。この 2 分子の DnaA ドメイン III は、シミュレーション中にほとんど解離してしまった。着色の方法は図 SB-1 に従う。E) D と同じ構造の、C3 ボックスに結合した DnaA K135-L366 と C2 ボックスに結合した DnaA K135-L366 部分の拡大。C3 ボックス上の DnaA は、K135 から L366 に向かって白から徐々に青に変わるグラデーションで着色した。C2 ボックス上の DnaA は、K135 から L366 に向かって白から徐々に緑色に変わるグラデーションで着色した。F) DnaA ホモ 5 量体で存在する中間-右側複合体のパネル E に対応する領域。こちらも全原子計算後の構造である。この複合体では、DnaA ドメイン III 同士の間は維持されている。G-J) シミュレーションの終構造と、各時点の構造とを重ね合わせたときの RMSD。サブ複合体の ATP-DnaA ホモ多量体 K135-L366 部分については G に示した。左側サブ複合体の DnaA ドメイン IV-DnaA ボックス部分を H に示した。中間-右側サブ複合体の DnaA ドメイン IV-DnaA ボックス部分を I に示した。そして *oriC* 右側領域に形成した DnaA ドメイン III ホモ 6 量体の DnaA ドメイン IV-DnaA ボックス部分を J に示した。G では左側サブ複合体、中間-右側サブ複合体ともに RMSD は 4Å 付近に分布していた。これは、この大きさの複合体にしては小さいといえる。一方、DnaA ドメイン III ホモ 6 量体ではシミュレーションの開始時に 7Å 付近になっている。これはシミュレーションの初期構造が不安定であることを示している。パネル H と I では RMSD は 2Å 付近に分布しており、構造が安定であることを示唆している。一方 J では DnaA ドメイン IV と R3 ボックスの領域や、DnaA ドメイン IV と R4 ボックスの領域が大きく構造を変化させている。これは不適切な 6 量体構造のために起こったと考えられる。

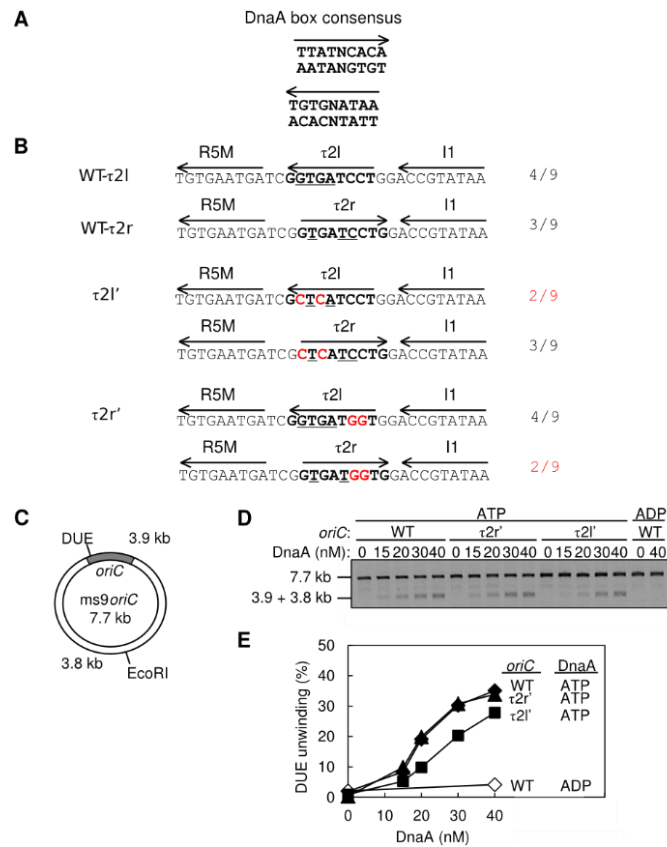
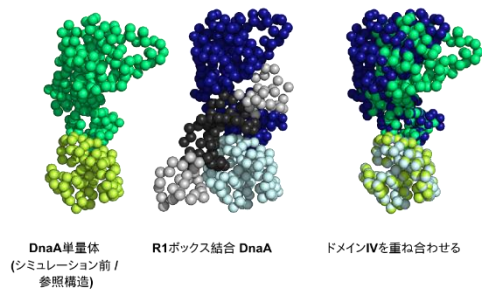


図 SB-8. τ2 ボックスに変異を含む *oriC* を用いた DUE 開裂アッセイ これは、共同研究者の実験データである。 A) DnaA ボックスのコンセンサス配列。それぞれの方向の 9 塩基対のコンセンサス配列を示した。B) 野生型と変異体の塩基配列。R5M から I1 までの領域が示してある。上の 2 段には野生型 *oriC* の塩基配列。τ2 ボックスの 2 つの定義(τ2l, τ2r, 黒太字)が向き(矢印)とともに示してある。コンセンサス配列と一致する部分に下線を引いてある。真ん中の 2 段には τ2l' 変異体の塩基配列を示してある。下の 2 段には τ2r' 変異体の塩基配列を示してある。変異を導入した部分は赤色で示した。C) M13*oriC*MS9 プラスミドの構造。oriC, DUE, EcoRI サイトの位置を示してある。D-E) DUE 開裂アッセイ。D) 図に示した量の ATP-または ADP-DnaA と、M13*oriC*MS9 またはその変異体プラスミドとを混合した。続けて一本鎖 DNA 特異的なヌクレアーゼ P1, EcoRI, アガロースゲル電気泳動を用いて解析を行った。E) ゲルの画像のバンドの強度を定量し、3.8 kb と 3.9 kb のバンドの 7.7kb のバンドに対する相対量を”DUE unwinding (%)”で示した。

A



B

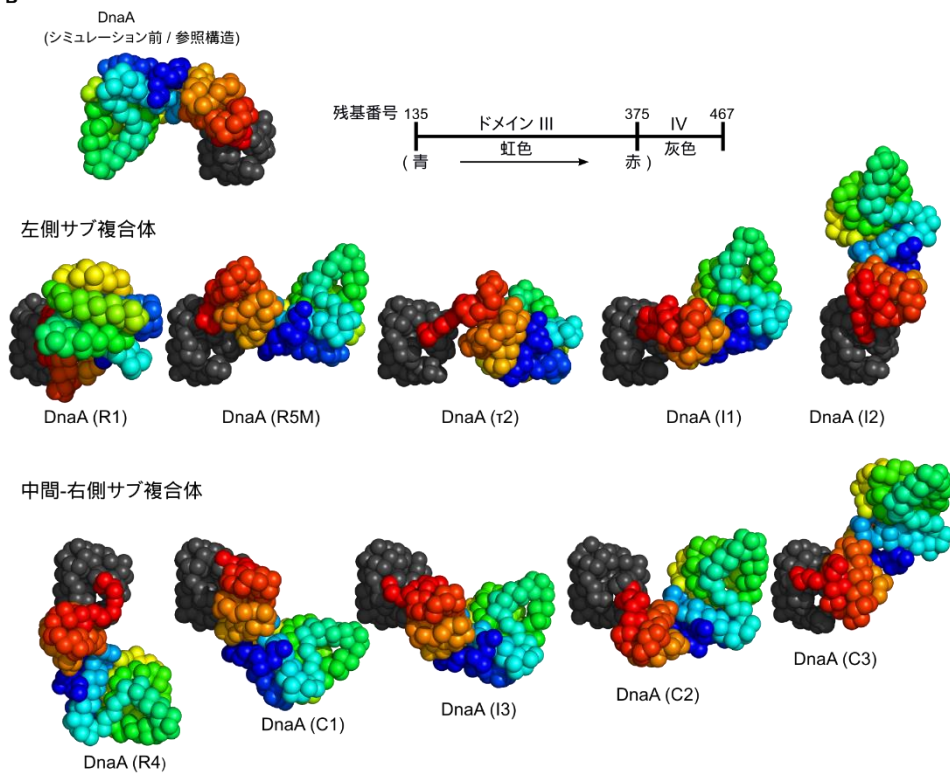


図 SB-9. 複製開始複合体中の DnaA 分子の構造比較 A) 二本鎖 DNA 結合に伴う DnaA ドメイン III と IV の位置関係の変化。左側が DnaA ドメイン III-IV 単量体の参照(天然)構造である。二本鎖 DNA に結合した DnaA ドメイン III-IV の構造を中央に示した。これら 2 構造のドメイン IV を重ね合わせ、ドメイン III の向きを比較したものが右図である。DNA は灰色、ただし DnaA ボックスは黒で示した。参照構造の DnaA ドメイン III は緑、ドメイン IV は黄緑色で示した。二本鎖 DNA に結合した DnaA ドメイン III-IV のドメイン III 部分は濃い青色、ドメイン IV 部分は水色で示した。B) それぞれの DnaA ボックスに結合した DnaA ドメイン IV の代表構造を示した。すべての構造で、ドメイン IV(灰色)の向きをそろえた。ドメイン III は K135(青色)から L373(赤色)へと、虹色に変化するように着色した。

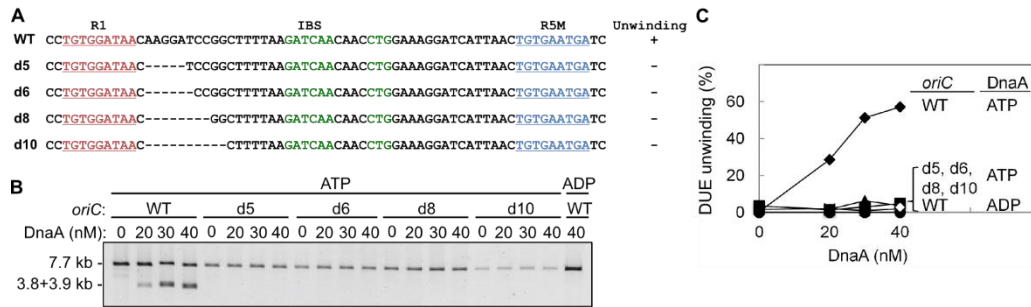


図 SB-10. R1 ボックス-IHF サイト間領域に欠失変異がある *oriC* の DUE 開裂活性 これは、共同研究者の実験データである。A) R1 ボックスから R5M ボックスまでの *oriC* 塩基配列。野生型の配列とともに、それぞれの変異体で欠損させた領域(ハイフン)を示した。R1 ボックスは赤色、R5M ボックスは青色、IHF 結合サイトは緑色で示した。DUE 開裂活性を右側に示した。+が野生型と同程度、-が不活性である。これは本文の図 1-5A と同じものである。B) 図に示した量の ADP または ATP-DnaA を M13*oriC*M59 プラスミド(*oriC* 領域を持つプラスミド)またはその欠失変異体とともにインキュベートした。続けて一本鎖 DNA 特異的なスクレーパー P1, EcoRI, 1%アガロースゲル電気泳動を用いて解析した。C) P1 スクレーパーによって分解された *oriC* DNA 分子の、インプットに対する割合を”DUE unwinding (%)”で示した。

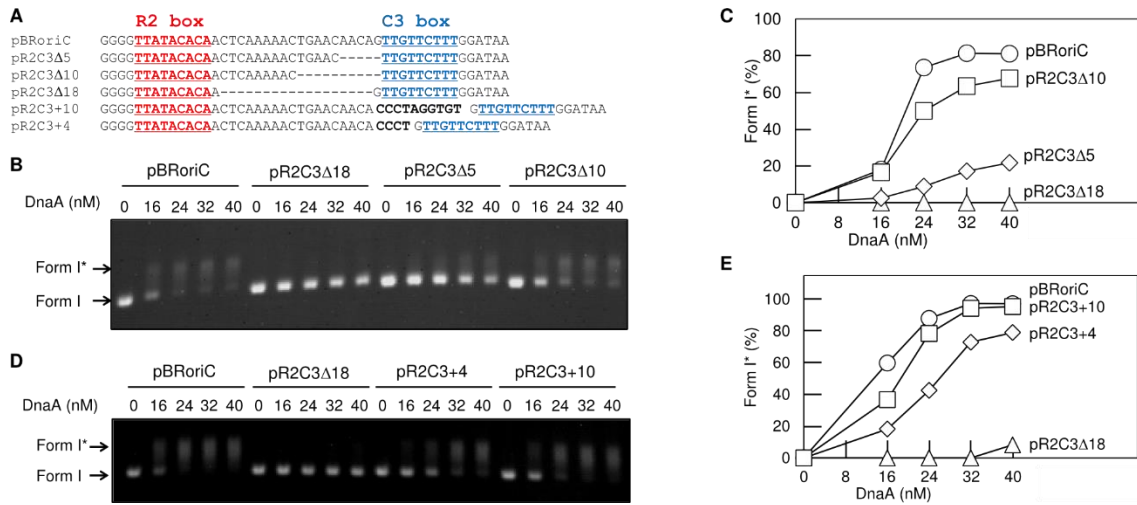


図 SB-11. *oriC* の R2-C3 領域の間隔を変えた場合の DnaB 装着活性 これは、共同研究者の実験データである。 A) R2 から C3 までの DNA 塩基配列。pBRoriC プラスミドの塩基配列と変異体の塩基配列が示してある。R2 ボックスと C3 ボックスはそれぞれ赤と青で着色してある。欠失させた塩基配列はハイフンによって示し、挿入した塩基配列は黒の太字で示してある。B-E) Form I\* アッセイ。このアッセイは DnaB ヘリカーゼの一本鎖 DNA 装着活性を調べるものである。図に示した量の DnaA を、pBRoriC プラスミド(またはその変異体), DnaB, DnaC, SSB, IHF, ジャイレースとともにインキュベートした。その後、アガロースゲル電気泳動で解析した(B, D)。電気泳動ゲルのバンドの強度を定量し、FormI\*状態の占める割合をプロットした(C, E)。

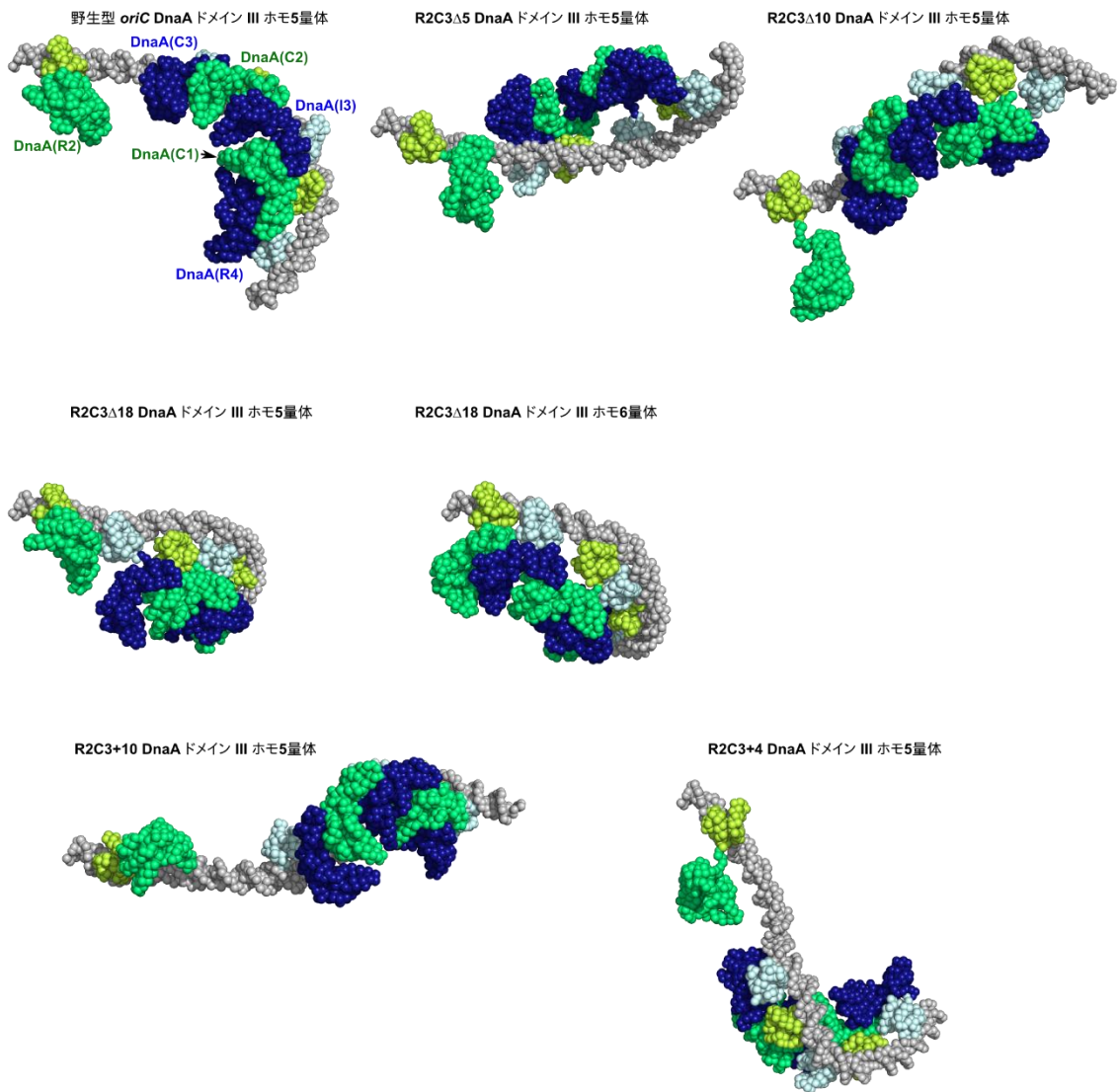
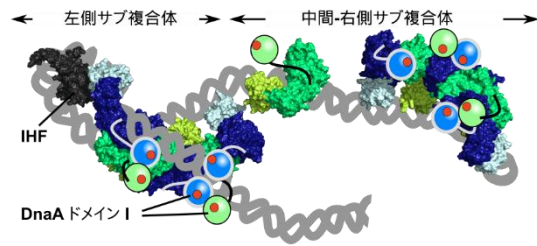
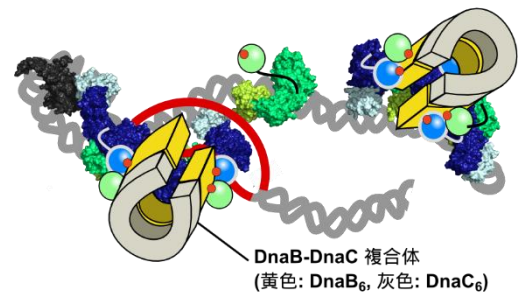


図 SB-12. *oriC* 変異体の複合体形成シミュレーション R2-C3 領域の DNA の長さを変えたそれぞれの変異体で、形成した複合体の代表構造を示した。それぞれの複合体は左側の DNA 末端の向きをそろえ、図示した。着色の方法は図 SB-1 と同じである。図 1-6 でも同じ代表構造を図示しているが、図 SB-12 とは視点が異なっている。

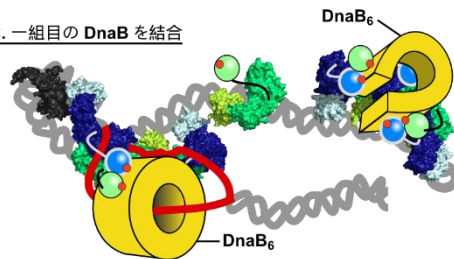
**A. 大腸菌 *E. coli* の複製開始複合体構造モデル**



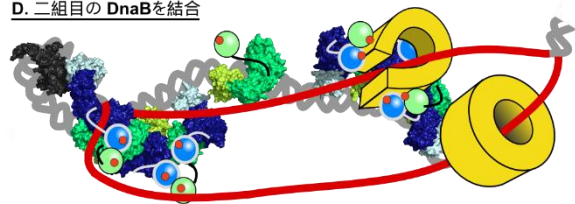
**B. DUE 開裂と DnaB-DnaC 複合体の結合**



**C. 一組目の DnaB を結合**



**D. 二組目の DnaB を結合**



**図 SB-13. DnaB ヘリカーゼ結合メカニズムのモデル** A) 大腸菌 *E. coli* の複製開始複合体を図示した。DnaA ドメイン III-IV 部分にはシミュレーションでモデリングした構造を利用した。DnaA ドメイン I-II 部分は単純化して図示してある。DnaA ドメイン I は二量体を形成できる<sup>78</sup>。B) DUE 開裂後、らせん型の DnaB<sub>6</sub>-DnaC<sub>6</sub> 複合体<sup>139</sup> が左側サブ複合体と中間-右側サブ複合体にそれぞれ結合する<sup>30</sup>。DnaA ドメイン I が主要な DnaB 結合サイトを有している<sup>31,75</sup>。複数の DnaA ドメイン I が DnaB<sub>6</sub> に結合する。C) 左側サブ複合体に結合した DnaB<sub>6</sub> が、一本鎖 DNA 領域を結合する。この過程では、DnaA ドメイン III の N 末端領域も DnaB と結合する可能性がある<sup>12</sup>。D) 結合した DnaB<sub>6</sub> が一本鎖 DNA 領域を延長する。それにより、もう一組の DnaB<sub>6</sub> が結合できるようになる。IHF と DnaA ドメイン III-IV 部分の着色は図 S1-1 と同じである。DnaA ドメイン I-II は青い球と灰色の曲線、または緑の球と黒の曲線で示した。DnaA ドメイン I の DnaB 結合サイトをオレンジ色の点で示した。DNA の二重鎖領域は灰色のらせんで示した。一本鎖領域は赤色の線で示した。DnaB ヘリカーゼは黄色、DnaC ヘリカーゼローダは灰色で示した。



表 SB-1. DnaA ボックス

	塩基配列 (5' - 3')
R1	<b>TTATCCACA</b>
R5M	<b>TCATTCACA</b>
$\tau$ 2 l	AGGAT <b>TCACC</b>
$\tau$ 2 r	GTGAT <b>TCCTG</b>
I1	<b>TTATACGGT</b>
I2	CTGAT <b>TCCA</b>
R2	<b>TTATACACA</b>
C3	<b>TTGTTCTTT</b>
C2	ATA <b>ACTACC</b>
I3	<b>TTGATCCAA</b>
C1	<b>TTCCCTGACA</b>
R4	<b>TTATCCACA</b>

コンセンサス配列と一致する部分を赤色で示した。

表 SB-2. 左側サブ複合体形成シミュレーションの統計

$\tau$ 2 結合設定	ドメイン III-IV 境界	時間の総計(10 <sup>4</sup> MD ステップ)
$\tau$ 2 l	ヘリックス	0
$\tau$ 2 l	ループ	36,418
$\tau$ 2 r	ヘリックス	0
$\tau$ 2 r	ループ	2,818

スナップショットは 10<sup>4</sup>MD ステップごとに記録した。複製開始複合体形成条件を満たしたスナップショットの数を示した。

表 SB-3. 中間-右側サブ複合体形成シミュレーションの統計

DNA	DnaA 多量化	トラジェクトリ 本数	時間の総計 ( $10^4$ MD steps)
野生型 <i>oriC</i>	5	9	61,561
(C3 ボックスの DnaA と R2 ボックスの DnaA 間に引力なし)			
野生型 <i>oriC</i>	5	7	19,827
(C3 ボックスの DnaA と R2 ボックスの DnaA 間に引力あり)			
野生型 <i>oriC</i>	6	3	19
R2C3 $\Delta$ 5	5	10	68,707
R2C3 $\Delta$ 5	6	3	910
R2C3 $\Delta$ 10	5	10	68,630
R2C3 $\Delta$ 10	6	9	52,232
R2C3 $\Delta$ 18	5	10	69,070
R2C3 $\Delta$ 18	6	9	59,074
R2C3+4	5	10	66,484
R2C3+4	6	1	3,852
R2C3+10	5	9	56,725
R2C3+10	6	0	0

スナップショットは  $10^4$ MD ステップごとに記録した。複製開始複合体形成条件を満たしたトラジェクトリの数と、スナップショットの総数を示した。

表 SB-4. DnaA ドメイン IV と 21 塩基対のシステムの DnaA-DNA 解離定数の検討

$\epsilon_{DnaA-DNA\ interaction}^{IJ}$ (kcal/mol)	$k_d$ (M)
0	$3.98 \times 10^{-4}$
	$4.18 \times 10^{-4}$
0.270	$3.08 \times 10^{-5}$
	$3.01 \times 10^{-5}$
0.285	$6.15 \times 10^{-6}$
	$1.16 \times 10^{-5}$
0.300	$2.54 \times 10^{-6}$
	$7.37 \times 10^{-7}$
0.315	$1.51 \times 10^{-7}$
	$4.08 \times 10^{-7}$
0.330	$3.10 \times 10^{-8}$
	$2.18 \times 10^{-7}$
0.345	$2.98 \times 10^{-9}$
	$2.49 \times 10^{-8}$
0.360	—
	$1.27 \times 10^{-8}$
0.375	$1.14 \times 10^{-8}$
	—
0.390	—
	—

シミュレーション内容の詳細は付録 A ”粗視化分子シミュレーションの詳細”にまとめられた。各 $\epsilon_{DnaA-DNA\ interaction}^{IJ}$ に対して、2回ずつシミュレーションを行った。DnaA 分子が DNA から解離しなかった場合には、ハイフンを記した。

表 SB-5. シミュレーションで用いた DNA の塩基配列

	塩基配列 (5'-3')
<i>oriC</i> 左側 (R1-12)	TGCCCTGTGGATAACAAGGATCCGGCTTTTAAGATCAACAACCTGGAAAAGGATCATTAACTGTGAATGATCGGTGATCCTGGACCGTATAAGCTGGGATCA G
R1IBSd5	TGCCCTGTGGATAAC-----TCCGGCTTTTAAGATCAACAACCTGGAAGAAGGATCATTAACTGTGAATGATCGGTGATCCTGGACCGTATAAGCTGGGATCAG
野生型 <i>oriC</i> 中間-右側	AGGGGTTATACACAACCTCAAAAACTGAACAACAGTTGTTCTTTGGATAACTACC GGTTGATCCAAAGCTTCTGACAGAGTTATCCACAGTAGA
R2C3Δ5	AGGGGTTATACACAACCTCAAAAACTGAAC-----TTGTTCTTTGGATAACTACC GGTTGATCCAAAGCTTCTGACAGAGTTATCCACAGTAGA
R2C3Δ10	AGGGGTTATACACAACCTCAAAAAAC-----TTGTTCTTTGGATAACTACC GGTTGATCCAAAGCTTCTGACAGAGTTATCCACAGTAGA
R2C3Δ18	AGGGGTTATACACAA-----GTTGTTCTTTGGATAACTACC GGTTGATCCAAAGCTTCTGACAGAGTTATCCACAGTAGA
R2C3+10	AGGGGTTATACACAACCTCAAAAACTGAACAACAGTTGTTCTTTGGATAACTACC GGTTGATCCAAAGCTTCTGACAGAGTTATCCACAGTAGA AGA
R2C3+4	AGGGGTTATACACAACCTCAAAAACTGAACAACAGTTGTTCTTTGGATAACTACC GGTTGATCCAAAGCTTCTGACAGAGTTATCCACAGTAGA
<i>oriC</i> 左側領域 (全原子 MD)	TGCCCTGTGGATAACAAGGATCCGGCTTTTAAGATCAACAACCTGGAAAAGGATCATTAACTGTGAATGATCGGTGATCCTGGACCGTATAAGCTGGGATCA GAATGAGGGGTTA
<i>oriC</i> 中間-右側領域 (全原子 MD)	CAGAATGAGGGGTTATACACAACCTCAAAAACTGAACAACAGTTGTTCTTTGGATAACTACC GGTTGATCCAAAGCTTCTGACAGAGTTATCCACAGTAGA

## 付録 C 第二章の研究手法の詳細/条件検討の手法

### 結果の可視化・図の作成

シミュレーション後の分子構造は VMD<sup>136</sup> や PyMOL<sup>53</sup> を用いて可視化した。図の作成には PyMOL, inkscape<sup>137</sup>, gnuplot<sup>138</sup>, Microsoft Powerpoint 2013 を用いた。

### 第二章で使用した粗視化モデルの力場

タンパク質モデルは第一章と同じものを用いた<sup>40</sup>。  
DNA モデルは 3SPN.2C モデルを用いた<sup>120,121</sup>。これは第一章で使ったモデルを(3SPN.1)、モデルを開発したグループが改良したものである(3SPN.2C)。粒子は糖・塩基・リン酸の重心に置く。第二章では力場の詳細はそれほど重要ではないため、エネルギー関数の概略を以下に示す。

$$V_{3SPN.2C} = V_{結合長} + V_{結合角} + V_{二面角} + V_{二面角,周期} + V_{塩基スタッキング} \\ + V_{クロス-スタッキング} + V_{塩基対} + V_{排除体積} + V_{静電相互作用}$$

ここで  $V_{静電相互作用}$  はデバイ-ヒュッケル型エネルギー関数である。

DNA-タンパク質相互作用も第一章と同様のポテンシャルを利用した。ここでは、粒子間接触ポテンシャルのパラメータ  $\epsilon$  は 0.6 kcal/mol とした。

### ソフトウェア CafeMol を用いた徐冷法

CafeMol の徐冷法では、入力として初期温度、最終温度、温度変化の回数を指定する。計算では等温分子動力学計算が行われ、その間に、一定の頻度で一定の温度幅ずつ温度が下がる。温度が下がる頻度は入力した温度低下の回数で決まる。最後の温度低下以降は、設定した最終温度での等温分子動力学計算が実施される。

### 付録 D で用いた解析方法

#### RMSD を用いた指標 / RMSD 計算

DNA 構築法の条件を検討する目的で、 $P(\text{RMSD} < 0.2)$  を定義した。これは自己再構築テストを行った際の RMSD が 0.2 より小さいヌクレオチドの、全体に占める割合である。ただしこのとき、末端のヌクレオチド以外を対象とした。

徐冷法や 1 K での CGMD で粗視化モデルの構造を緩和した際、緩和の前後で複合体の構造がどの程度変わったかは RMSD によって評価した。または、緩和計算の再現性を確認する目的で、同じ構造について計 11 回の緩和計算を行い、構造の違いを RMSD によって評価した。これらの場合は、RMSD 計算前に 2 つの粗視化構造の重ね合わせを

行っている。事前の重ね合わせの有無が自己再構築テストでの RMSD 算出と異なる部分である。

### ワトソン-クリック塩基対再現性のテスト

本文に示した方法でワトソン-クリック塩基対の再現性を計算した後、この値が 0.98 を上回る構造の数を数えた。全構造に対するこの割合を”WC over0.98”と名付け、解析の指標とした。また同様に”WC over0.8”も定義した。

### ヌクレオチド連結部分の再現性

テストセット A の結晶構造から、DNA の P 原子と O3'原子の距離の平均値と、標準偏差を求めておいた。粗視化モデルから再構築した原子モデルで、P-O3'間距離が(平均値 $\pm 2 \times$ 標準偏差)に含まれるとき、再構築したモデルでの P-O3'間距離は自然な距離であるとみなした。これを実現する確率を  $P_{P-O3', realistic}$  と名付けた。同じ要領で  $P_{O3'-P-O5', realistic}$  を定義した。これらの確率を求める際、末端ヌクレオチドを含む連結部のデータは除外した。

## 付録 D 第二章の条件検討の結果と考察

### DNA 再構築時の重みパラメータ $w_i$ の検討

DNA 再構築に用いる重みパラメータ  $w_i$  の値を検討した。DNA 再構築でははじめに粗視化 DNA の 5 粒子の座標を用い直交座標系を定義する。その後、この 5 粒子がフラグメントライブラリの各構造にどれだけ近いかを以下の F によって評価する。詳細は本文第二章にある。

$$F = \sum_{j \in \{S_{i-1}, P_i, B_i, S_i, P_{i+1}\}} w_j \left( \mathbf{r}_j - \mathbf{r}_{j, DNA \text{ フラグメント}} \right)^2$$

ここで  $\mathbf{r}_j$  は定義した直交座標系での各粒子の座標である。F を計算する際、各粒子の寄与の大きさをパラメータ  $w_i$  が決定する。そこで、このパラメータの値を検討した。

5 つの独立したパラメータを連続変数として解析するには計算コストが高すぎるため、各  $w_i$  がとる値をいくつかに絞り、離散的な変数として検討した。調べた組み合わせの中でより良いものを一組選択した。本研究の再構築法では粗視化計算後のスナップショットには 1 K で  $10^4$  MD ステップの等温 MD を行い、構造を改善している。一方、パラメータの検討では構造の悪い粗視化モデルからも精度の高い原子モデルを構築できることを目指した。そこで粗視化計算後のスナップショットに  $10^3$  MD ステップの徐冷法を行った構造をテストに用いた。詳細なスケジュールは付録 E 表 SE-1 の最上段に示したものである。この条件による粗視化モデルの構造改善の程度は後述(付録 D 粗視化モ

デルの結合長・結合角を改善する方法の検討)する。データは付録 E 図 SE-6 パネル ABC 上段の “10<sup>3</sup> (300 K)” に示してある。パラメータのチューニングにはテストセット A を利用した。はじめに、すべてのパラメータ  $w_i$  を 1.0 にした組と、 $w_i$  の内一つを 0, 0.1, 0.25, 0.5, 0.75, 1.333, 1.5, 2.0, 4.0, 10.0 のどれかに変えた組を試した。そして P(RMSD<0.2) と WC over0.98,  $P_{P-O3', realistic}$ , and  $P_{O5'-P-O3', realistic}$  の 4 つの指標をプロットした。これらの定義は付録 C “付録 D で用いた解析方法” 項に記してある。筆者はワトソン・クリック塩基対の再現度(WC over0.98) と、 $P_{O5'-P-O3', realistic}$  にトレードオフがあることを見つけた。同様にワトソン・クリック塩基対の再現度(WC over0.98) と、 $P_{P-O3', realistic}$  にもトレードオフがあることを見つけた 付録 E 図 SE-2。そのため、4 つの指標がどれもまんべんなく良いパラメータの組を探すことにした。これら 4 つの指標の相対的な重要度を定めることは困難であるため、この 4 指標を組み合わせた一つの指標を使った評価はできない。

続いて、 $w_{Si-1}$  を 0.1 または 4.0 とした条件で残り一つのパラメータを 0, 0.1, 0.25, 0.5, 0.75, 1.333, 1.5, 2.0, 4.0, 10.0 のいずれかに変更した。付録 E 図 SE-2 右側には  $w_{Si-1}$  を 0.1 とした場合の結果を示してある。結局、最適なパラメータを求める指針はここからは出てこなかった。

今度は 5 つのパラメータ  $w_j$  をそれぞれ 0.1, 1.0, 4.0 のどれかにし、自己再構築テストの P(RMSD<0.2) と WC over0.98 を比較した。重複を考慮すると全部で  $3^5-2=241$  通りのパラメータの組を試した。この 2 つの再現性の指標が大きい順にパラメータの組を並べたものを付録 E 表 SE-2, 表 SE-3 にまとめた。ここから、

$$w_{Si-1} = 0.1, w_{Pi} \leq 1, w_{Si} \geq 1, w_{Bi} \geq 1, \text{ and } w_{Pi+1} \leq 1.$$

の場合に、自己再構築テストの RMSD が小さく、ワトソン・クリック塩基対の再現性が高くなる傾向にあることが分かった。

この結果を踏まえ、最後に  $w_{Si-1} = 0.1$ ,  $w_{Si}$  と  $w_{Bi}$  は 1.0, 1.333, 1.5, 2.0, 4.0 のいずれか、 $w_{Pi}$  と  $w_{Pi+1}$  は 0.1, 0.25, 0.5, 0.75, 1.0 のいずれかとして DNA を再構築し、その結果を比較した。再び P(RMSD<0.2) と WC over0.98,  $P_{P-O3', realistic}$ , and  $P_{O5'-P-O3', realistic}$  の 4 つの指標を計算した(付録 E 図 SE-3)。最終的に  $w_{Si-1} = 0.1$ ,  $w_{Pi} = 0.75$ ,  $w_{Si} = 1.0$ ,  $w_{Bi} = 1.0$ , and  $w_{Pi+1} = 0.1$  を最適なパラメータの組とした。この組では、4 つの指標いずれもがまんべんなく良い。具体的には P(RMSD<0.2) = 0.559, WC over0.98 = 0.103,  $P_{P-O3', realistic} = 0.0931 (\pm 0.00085)$ , and  $P_{O5'-P-O3', realistic} = 0.205 (\pm 0.0013)$  である。なおカッコ内は標準誤差を示す(N = 20)。

本論文の第二章及び、付録 D, E の以下の解析ではすべてこのパラメータの組を用いた。

また、どのような構造がパラメータに関係なく原子モデル構築に失敗するかを調べた。付録 E 表 SE-4 には試したすべて(932 通り)のパラメータセットのすべてで、自己再構築テストの RMSD が 1.5 Å を超えた構造を並べた。またそのいくつかを付録 E 図 SE-4 に示した。これらはワトソン・クリック塩基対形成に関わっていないものや、塩基間スタ

ッキングをしていないものであった。そしてこれらの結晶構造と再構築モデルでは塩基がおよそ  $180^\circ$  回転していた。つまり、塩基の面の向きが通常と異なる構造ではモデリングに失敗する場合がある。

### 粗視化モデルの結合長・結合角を改善する方法の検討

この検討ではテストセット C を用いた。粗視化計算後の CG モデルの局所構造、つまり結合長や結合角を改善する方法を検討した。第 2 章で原子モデル構築法を開発する目的は、粗視化計算でサンプリングした一時的な構造や不安定な構造を詳しく調べることである。従って粗視化モデルの仮想的な結合長さや結合角を改善する一方で、複合体全体が大きく構造を変えることを防ぐ必要がある。様々なスケジュールの徐冷法と、シミュレーション時間を変えた 1 K の等温 CGMD を行った。徐冷法のスケジュールは付録 E 表 SE-1 にまとめてある。はじめに DNA の糖粒子と塩基粒子の仮想結合長の改善を調べた(付録 E 図 SE-6)。

最初に構造緩和後の  $S_i-B_i$  および  $S_i-P_i$  の仮想結合長を調べた。(付録 E 図 SE-6A, B)  $10^3$  MD ステップの徐冷法と  $10^2$  MD ステップの 1 K の CGMD 以外では構造は大きく改善された。リン酸粒子・糖粒子・塩基粒子からなる仮想結合長の分布は、徐冷法を行ってもそれほど変化しなかった。(付録 E 図 SE-6C)

次に、徐冷法/1 K CGMD による複合体全体の構造変化の程度を比較した。徐冷法(または 1 K の MD)前後の構造を重ね合わせ、RMSD を計算したものを付録 E 図 SE-7 に示してある。計 20 回の試行の平均と標準偏差を示した。徐冷法のシミュレーションステップ数が大きいほど、RMSD のばらつきが大きくなった。また徐冷法の初期温度を低く設定するほど、RMSD のばらつきは小さかった。RMSD の標準偏差が小さいことは、大きな構造変化が起こりにくいことを示している。

最後に、DNA の原子モデル構築後の構造を評価・比較した。筆者は WC over0.98, WC over0.8, P<sub>P-03'</sub>, realistic, and P<sub>O5'-P-03'</sub>, realistic を計算した。

すると、 $10^3$  MD ステップの徐冷法と  $10^2$  MD ステップの 1 K MD ではこの値は他より低く、それ以外は明確な差異は現れなかった(付録 E 図 SE-8)。

以上の結果に基づき、粗視化モデルの構造緩和の手段として 1 K 下で  $10^4$  MD ステップの等温 CGMD を行うことに決定した。DNA 粗視化モデルの局所構造の改善という観点からは、仮想結合長や結合角の改善の度合いが長時間( $10^6$  MD ステップ)の徐冷法を行った場合と同程度であった。さらに、10,000 MD ステップは 10,000 構造程度をすべて緩和してもそれほどの計算時間はかからない。筆者は多くの粗視化モデルから原子モデルを構築することを視野に入れているため、計算コストも重要な要素である。

最後に、この 1 K の緩和計算の再現性を調べた。この目的のため、テストセット C を用い、追加の緩和計算を行った。同じ粗視化計算のスナップショットを初期構造として、1 K の緩和計算を 10 回行った。それぞれでランジュバンダイナミクスの乱数の種



を変えた。筆者たちはシミュレーション後の 11 構造(もともとの緩和計算+10 回の追加計算)のすべてのペアについて、構造を重ね合わせた後に RMSD を計算した。そして、RMSD の平均値と標本標準偏差を計算した。多くの場合に RMSD は 0.2 Å 付近に分布していた(付録 E 図 SE-9A, B)。この値は非常に小さいといえる。したがって、ほとんどの場合に構造緩和計算の再現性が確認できた。なお、いくつかの DNA-タンパク質複合体では、DNA 部分の RMSD は大きな値となった。こうした構造をいくつか調べると、それらは B 型 DNA とは異なる構造をとっていた(付録 E 図 SE-9C)。本研究で使った DNA 粗視化モデル 3SPN.2C は B 型 DNA を安定化させる効果が強いため、これは力場の影響かもしれない。つまり B 型を好む力場のために、緩和計算中に徐々に B 型 DNA に近づいていった可能性がある。このような場合にはむしろ 3SPN.2C 以外の力場を用いることが適切だと考えられる。

### **DNA 再構築テストのフラグメントライブラリ依存性**

フラグメントライブラリのサイズや内容が DNA 再構築テストの結果に与える影響を調べた。第 2 章で主に使ったライブラリのほかに 7 種類のフラグメントライブラリを用意し、DNA 再構築の結果を比較した。用意したフラグメントライブラリは以下のとおりである。各ライブラリのフラグメント数は付録 E 表 SE-5 に示した。

A: 2,085 枚の構造ファイルから生成したライブラリ。これは第 2 章のためにダウンロードしたファイルの内、テストセット A を除いたすべてである。

B: 全部で 953 枚の構造ファイルから生成したライブラリ。これは第二章のためにダウンロードしたファイルの内、テストセット A と、第 2 章でフラグメントライブラリに用いたファイルを除いた残りすべてである。

C: 全部で 565 枚の構造ファイルから作成したライブラリ。これは第 2 章でフラグメントライブラリ構築に使ったファイルのおよそ半分の構造ファイルを使って作成したものである。PDB ファイルを 3 つのグループ (DNA-タンパク質複合体で DNA 鎖を一本のみ含むもの、それ以外の DNA-タンパク質複合体、さらにそれ以外) に分けた内のそれぞれから、ランダムに半分を選びとった。

D: 全部で 282 枚の構造ファイルから精製したライブラリ。上述のフラグメントライブラリ C で使った PDB ファイルに対し、C で行ったものと同様の操作を行い、282 構造を取り出した。

E: 全部で 140 枚の構造ファイルから精製したライブラリ。上述のフラグメントライブラリ D で使った PDB ファイルに対し、C で行ったものと同様な操作を行い、140 構造を取り出した。

F: 第 2 章のためにダウンロードした DNA-タンパク質複合体の構造ファイルすべてを使って生成したライブラリ。ただしテストセット A の構造ファイルは除外してある。

G: 第 2 章のためにダウンロードした構造ファイルの中でタンパク質を含まないものすべてを使って生成したライブラリ。

データセット依存性のテストには、テストセット A を用いた。自己再構築テストを行い、また構造緩和後の CGMD スナップショットから構築した原子モデルの構造を評価した(付録 E 図 SE-11)。フラグメント数が少ないライブラリ(D,E)やライブラリ G を除けば、フラグメントライブラリが結果に与える影響はわずかであった。ライブラリ D, E, G では、第二章で主に使っていたライブラリと比べ、自己再構築テストの RMSD は明らかに大きく、また構築した原子モデルのワトソン-クリック塩基対再現性は、明らかに低下していた。ヌクレオチド連結部の構造はライブラリの違いの影響をほとんど受けなかった。(付録 E 図 SE-11C, D)

## 付録 E 粗視化モデルからの原子モデル構築法の検討の詳細

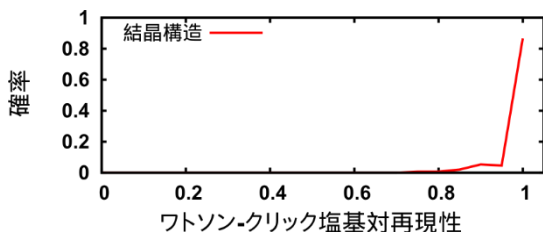


図 SE-1. 結晶構造を用いたワトソン-クリック塩基対再現性のテスト テストセット A の結晶構造におけるワトソン-クリック塩基対の再現性を示した。結晶構造を粗視化モデルに変換した後で、ワトソン-クリック塩基対を形成しているとみなせる塩基ペアを探し出した。粗視化モデルでワトソン-クリック塩基対を形成しているとみなした塩基ペアのうち、実際の結晶構造ではいくつが塩基対を形成しているかの割合を算出した。この割合をワトソン-クリック塩基対の再現性と定義し、その分布を求めた。この図では、ワトソン-クリック塩基対再現性が1になるべき状況で、実際に1になるかどうかを試験している。期待通り、ほとんどすべての場合にこの値は1.0となった。ただし、すべての構造で1になるわけではなかった。粗視化モデルでの塩基対形成の基準が、いくらか緩いものになっている。ワトソン-クリック塩基対形成基準のより詳細な定義に関して、以下(\*)に示した。

\*粗視化 DNA モデルのワトソン-クリック塩基対再現性は本文 2.2.3 節に示した。ただし、複数の塩基ペアが同じプリン塩基を共有してしまった場合には、結晶構造を参照し、塩基ペアごとにドナー-アクセプター距離の最大値を求めた。この最大値を最小にする塩基ペアのみがワトソン-クリック塩基対を形成しているとした。

結晶構造ではワトソン-クリック塩基対のドナー-アクセプター間の距離がどれも  $3.5 \text{ \AA}$  より小さい場合に、ワトソン-クリック塩基対を実現しているとみなした。

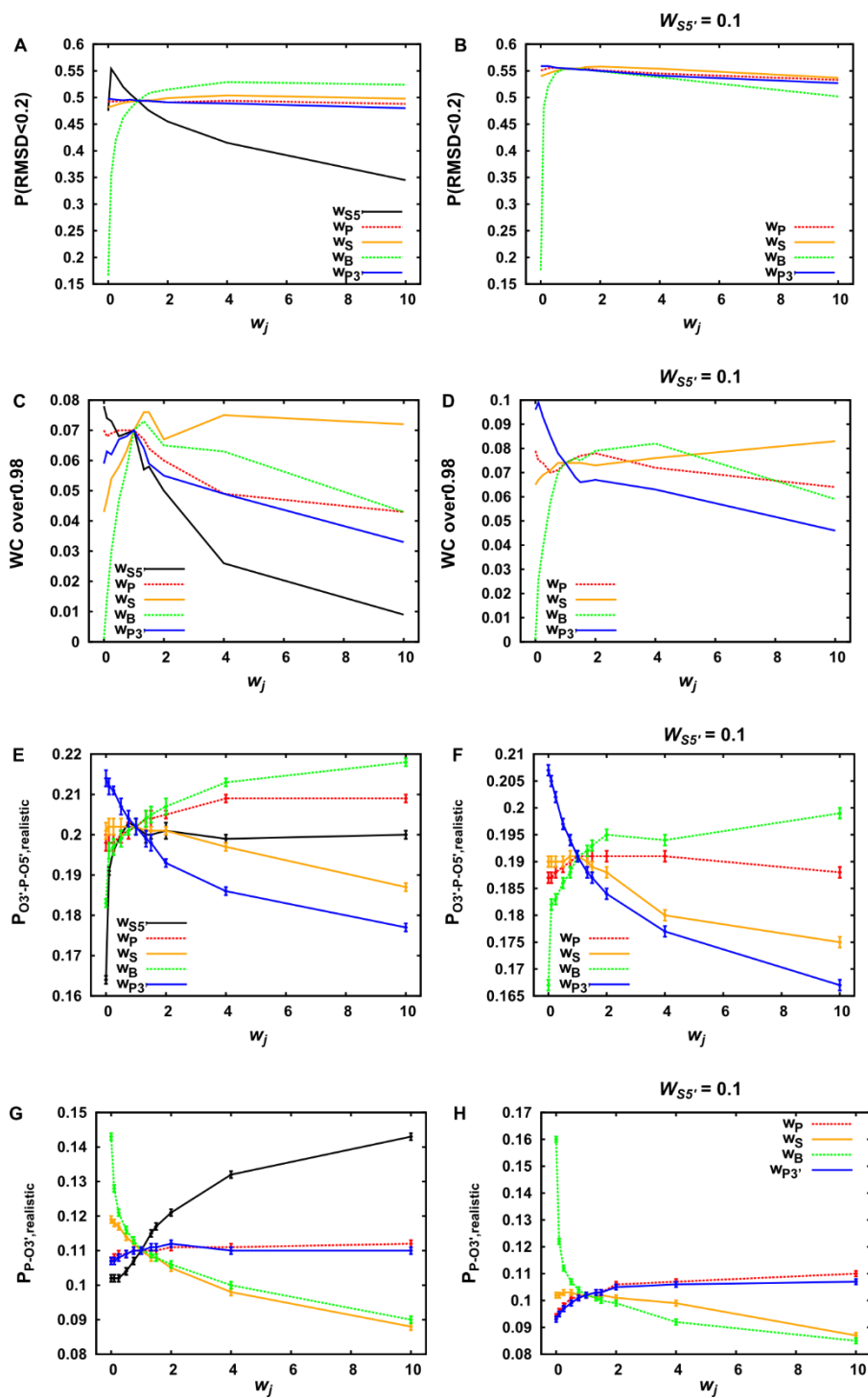


図 SE-2. DNA 構築時の重みづけパラメータ  $w_j$  の検討 A, C, E, G) まず  $w_j$  ( $w_{Si+1}, w_{Pi}, w_{Si}, w_{Bi}, w_{Pi+1}$ ) のうち一つを変更した。B, D, F, H)  $w_{SS'}$  を 0.1 に設定し、それ以外の  $w_j$  うち一つを変更した。  $w_j$  は 0.0, 0.1, 0.25, 0.5, 0.75, 1.0, 1.333, 1.5, 2.0, 4.0, 10.0 のいずれかとした。 A, B)  $P(\text{RMSD} < 0.2)$  をプロットした。 C, D). WC over 0.98 をプロットした。原子モデル構築時にワトソン-クリック塩基対構造を再現すべき塩基ペアは、徐冷法後の粗視化構造に基づいて定義した。 E, F)  $P_{O3'-P-O5', \text{realistic}}$  を示した。 G, H).  $P_{P-O3', \text{realistic}}$  を示した。それぞれの指標の定義については、付録 C にまとめている。 E-G ではエラーバーは 20 回の徐冷法のそれぞれの終構造からモデリングした結果の標準誤差を示す

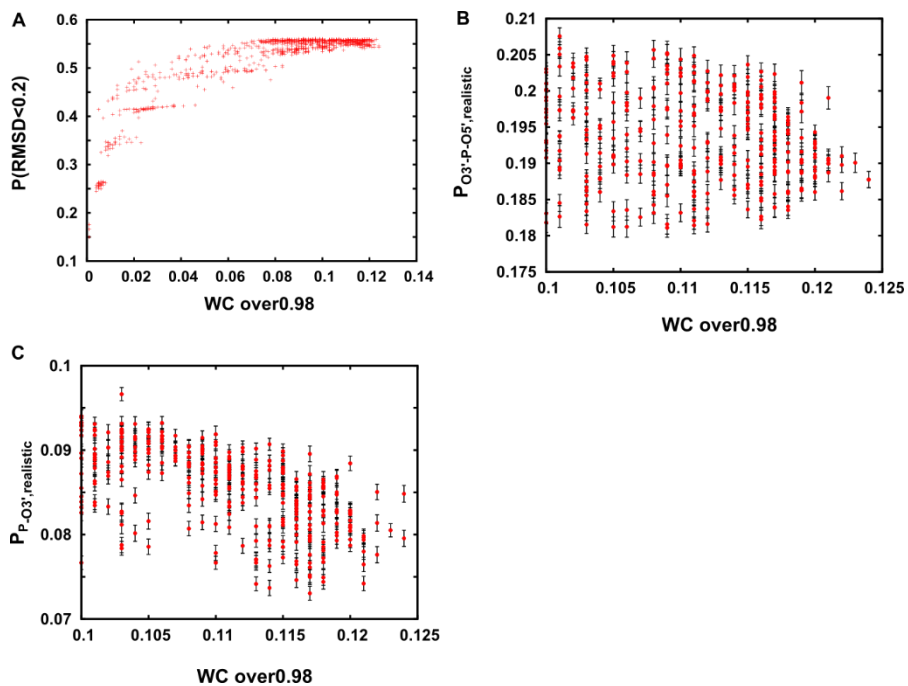


図 SE-3. DNA 構築時の重みづけパラメータ  $w_j$  の検討 A)パラメータ検討時に試した 932 のパラメータセットについて、 $P(\text{RMSD} < 0.2)$  と  $\text{WC over } 0.98$  の関係をプロットした。 B)  $\text{WC over } 0.98$  が 0.1 以上になったパラメータセットについて、 $P_{0.03, P-0.05, \text{realistic}}$  と  $\text{WC over } 0.98$  の関係を示した。 C)  $\text{WC over } 0.98$  が 0.1 以上になったパラメータセットについて、 $P_{P-0.03, \text{realistic}}$  と  $\text{WC over } 0.98$  の関係を示した。それぞれの指標の定義は付録 C にまとめてある。B, C では、エラーバーは 20 回の徐冷法のそれぞれの終構造からモデリングした結果の標準誤差を示す。

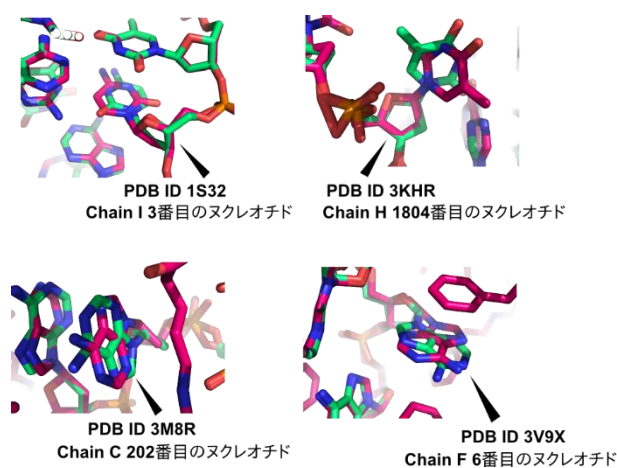


図 SE-4. モデリングに失敗したヌクレオチドの例 パラメータ検討時に試した 932 通りのパラメータ  $w_j$  のセットすべてでモデリングに失敗した構造を示した。結晶構造が炭素原子をピンク色で示したモデル、再構築したモデルが炭素原子を緑色で示したモデルである。窒素原子は青、酸素原子は赤、リン原子は橙色で示した。

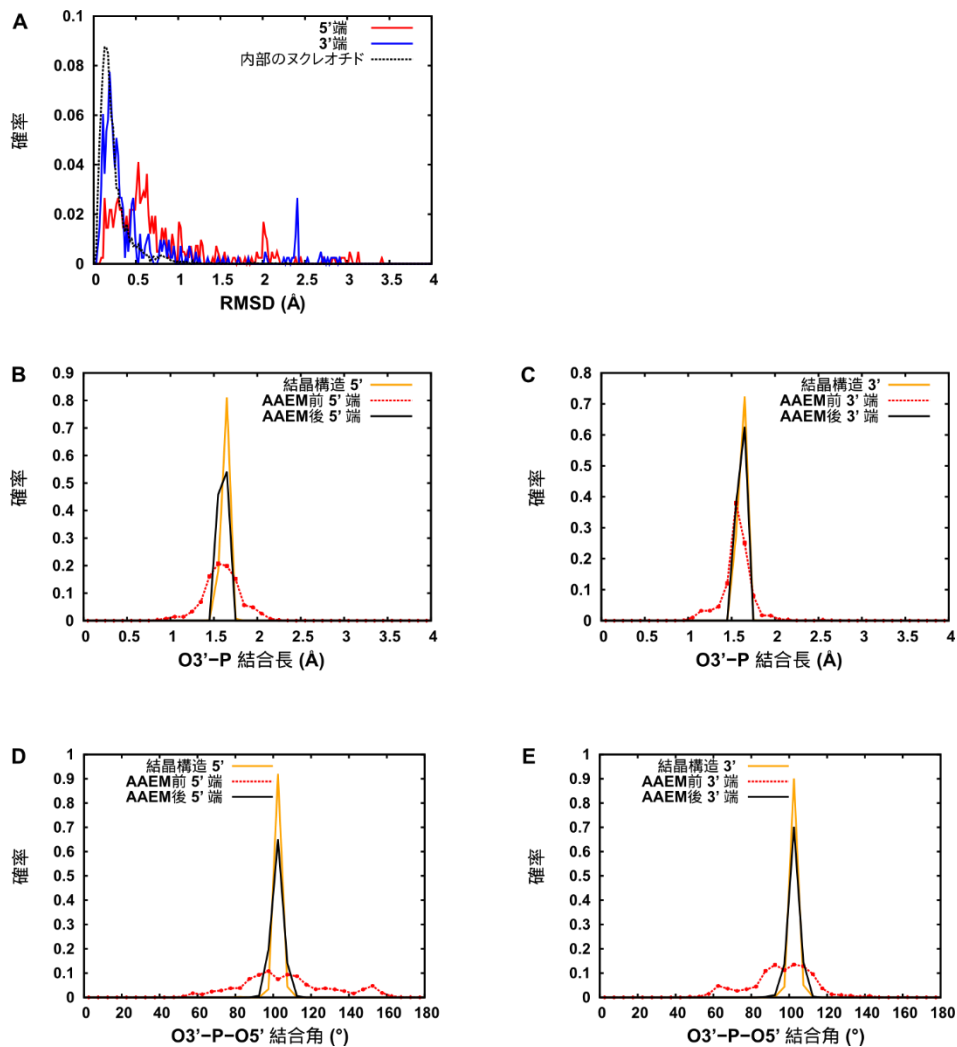


図 SE-5. 末端ヌクレオチドの原子モデル構築 A) 自己再構築テスト時の結晶構造との間の RMSD. B-E) CGMD 後の構造からの原子モデル構築。O3'-P 共有結合長 (B, C) と O3'-P-O5' 結合角 (D, E)を示した。“結晶構造”は結晶構造での値を示す。“AAEM 前”は全原子力場を用いたエネルギー最小化計算前の分布を示す。これは CGMD 後に 1K の緩和 MD 計算を行った後で構築した原子モデルでの分布である。エラーバーは 20 回のシミュレーションの標準誤差を示す。“AAEM 後”は DNA-タンパク質複合体にエネルギー最小化計算を行った後の分布を示す。ここでは、それに続く 10 ピコ秒の MD を実施できた構造のみを解析に用いた。

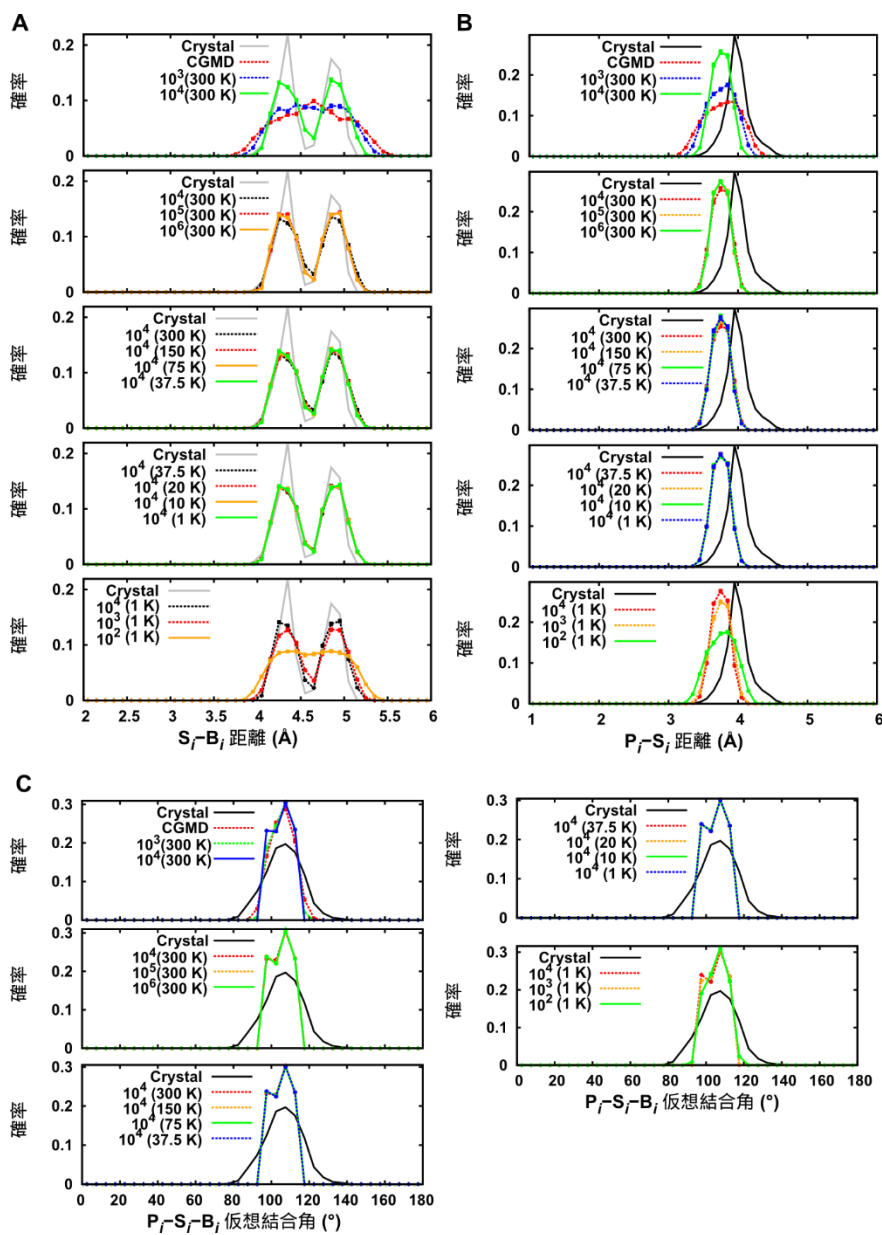


図 SE-6. 粗視化 DNA モデル構造緩和法の探索 様々な方法で粗視化モデルの構造を緩和し、局所構造の改善を調べた。この図では、テストセット C についての結果を示す。A) 同一ヌクレオチド内の塩基粒子  $B_j$  と糖粒子  $S_i$  の仮想結合長の分布。B) 同一ヌクレオチド内のリン酸粒子  $P_i$  と糖粒子  $S_j$  の仮想結合長の分布。C)  $P_i-S_i-B_j$  仮想結合角。“結晶構造”は、結晶構造の粗視化粒子に対応する座標を用いて算出した値である。それ以外の場合は、構造緩和後の結果を示した。シミュレーションのステップ数と初期温度(カッコ内)が示してある。徐冷法の最終温度は 1 K である。初期温度が 1 K と示した条件では 1 K の等温 CGMD を実施した。結晶構造以外の場合は、エラーバーは 20 回のシミュレーションの標準誤差を表す。

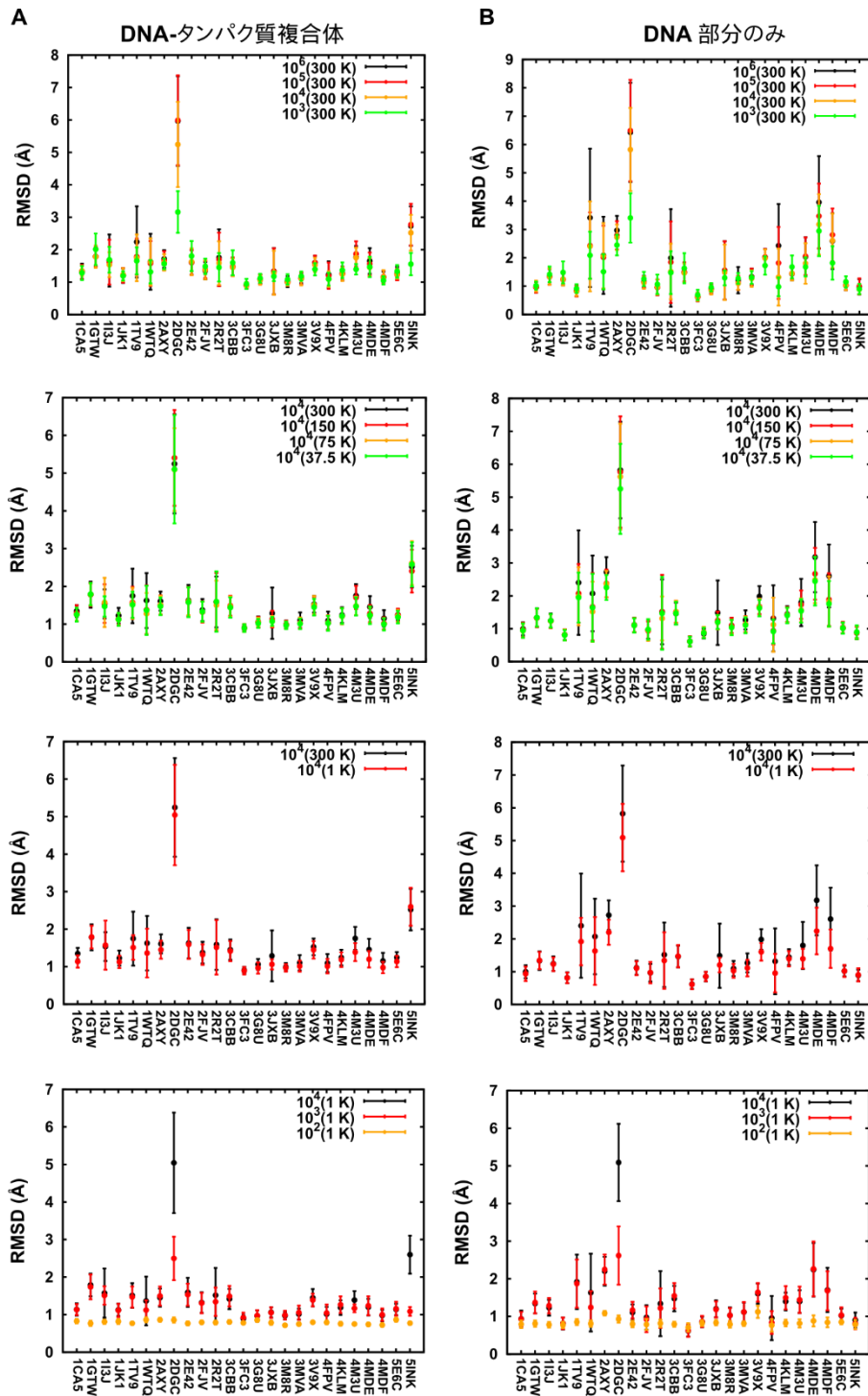


図 SE-7. 粗視化 DNA モデル構造緩和法の探索:緩和前後での構造変化 構造緩和前後の構造を重ね合わせ、RMSD を計算した。グラフにはシミュレーションのステップ数と徐冷法の初期温度(カッコ内)が示してある。徐冷法の最終温度は 1 K である。A) DNA-タンパク質複合体全体での RMSD。 B) DNA 部分のみを重ね合わせたときの RMSD。エラーバーは 20 回のシミュレーションの標準偏差を表す。



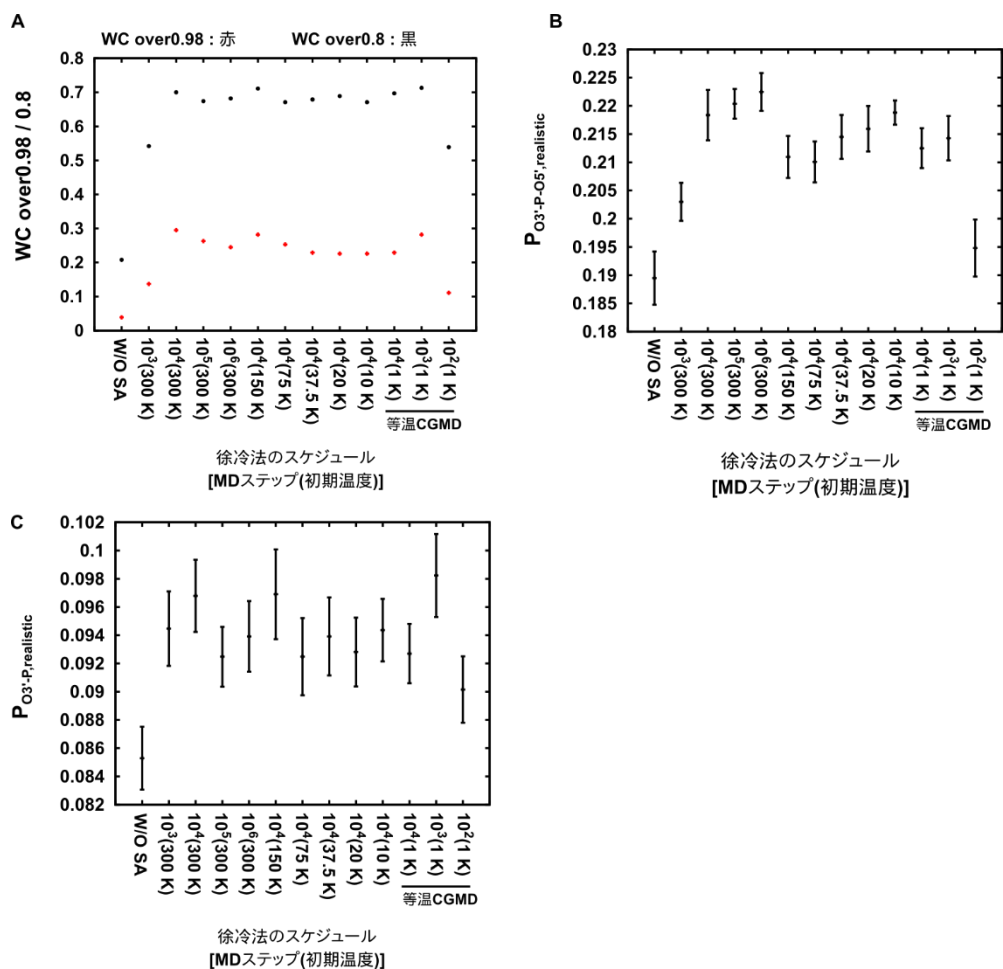


図 SE-8. 粗視化 DNA モデル構築緩和法の探索: 構築した DNA 原子モデルの構造比較 A) WC over0.8 と WC over0.98. B)  $P_{O3^3-P_{O5^5}}$ , realistic. C)  $P_{P-O3^3}$ , realistic. “W/O SA” は構造緩和を行う前の CGMD スナップショットから原子モデルを構築した場合を示す。それ以外はステップ数と徐冷法の初期温度(カッコ内)が示してある。B と C で、エラーバーは 20 回の CGMD 結果を用いたモデリングの標準誤差を表す。

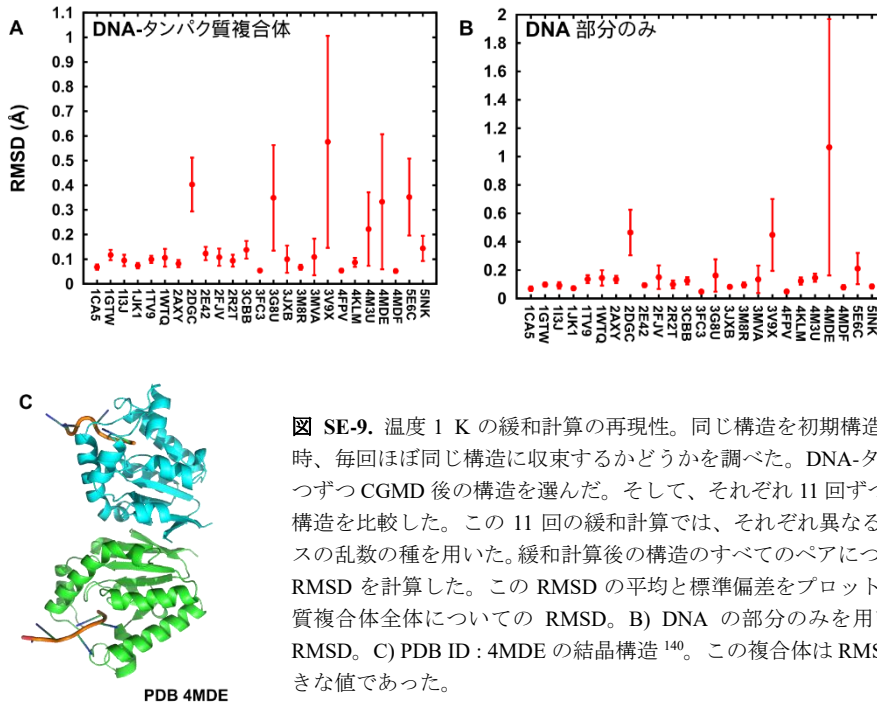


図 SE-9. 温度 1 K の緩和計算の再現性。同じ構造を初期構造として緩和計算を行った時、毎回ほぼ同じ構造に収束するかどうかを調べた。DNA-タンパク質複合体ごとに一つずつ CGMD 後の構造を選んだ。そして、それぞれ 11 回ずつ 1 K の緩和計算を行い、構造を比較した。この 11 回の緩和計算では、それぞれ異なるランジュバンダイナミクスの乱数の種を用いた。緩和計算後の構造のすべてのペアについて、構造を重ね合わせ、RMSD を計算した。この RMSD の平均と標準偏差をプロットした。A) DNA-タンパク質複合体全体についての RMSD。B) DNA の部分のみを用いて重ね合わせたときの RMSD。C) PDB ID: 4MDE の結晶構造<sup>140</sup>。この複合体は RMSD の標準偏差が比較的大きな値であった。

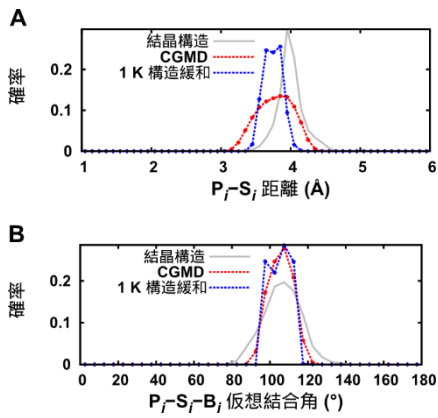
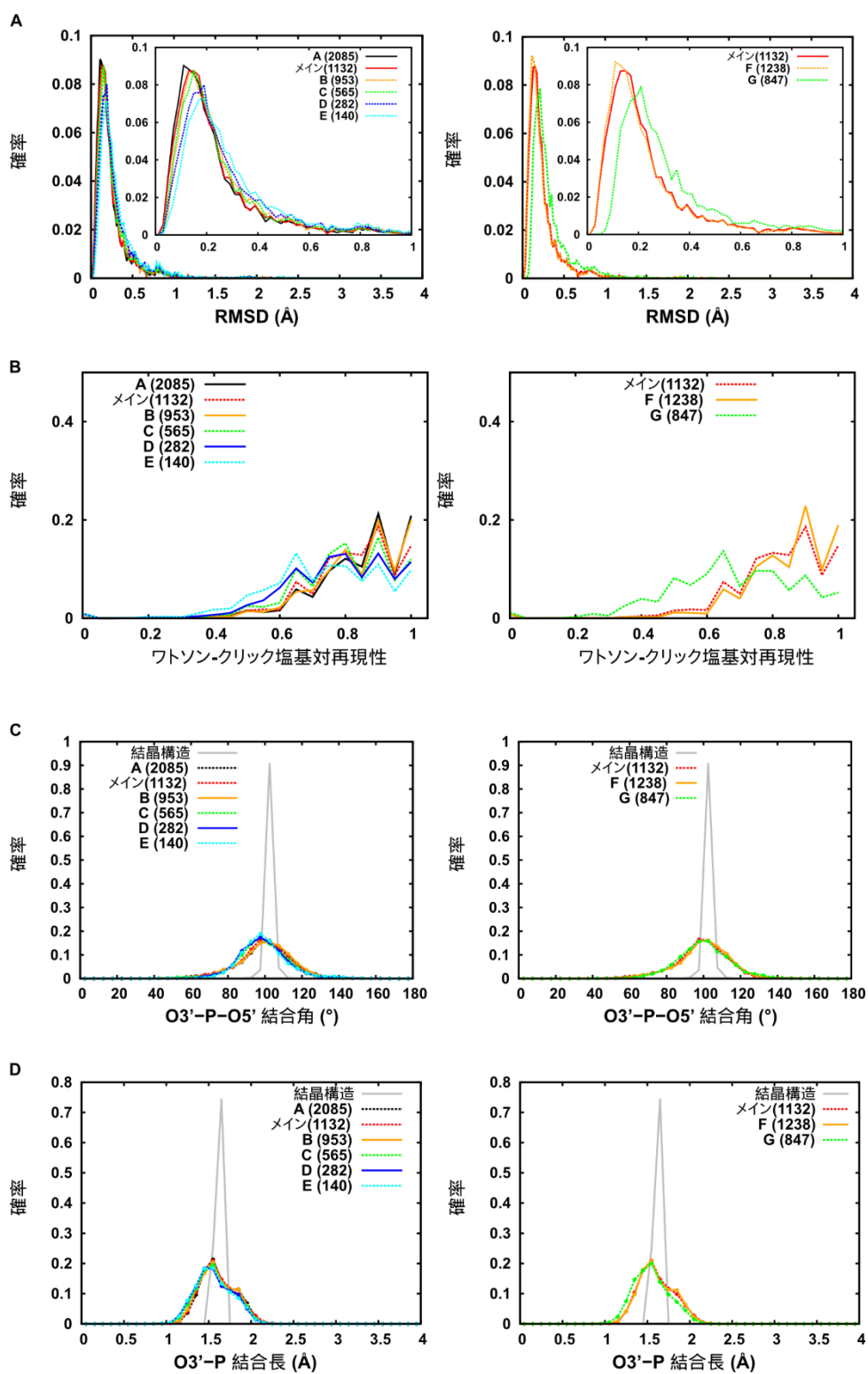


図 SE-10. 温度 1 K 下の CGMD による DNA 粗視化モデルの構造緩和 DNA 粗視化モデルの  $P_i-S_i$  仮想結合長と  $P_i-S_i-B_i$  仮想結合角の分布を結晶構造(“結晶構造”)、CGMD 直後(“CGMD”)、1 K の緩和計算後(“1 K 構造緩和”)と比較した。この図では、テストセット A についての結果を示す。A) 同一ヌクレオチド内の  $P_i-S_i$  仮想結合長。B)  $P_i-S_i-B_i$  仮想結合角。“結晶構造”は、結晶構造の粗視化粒子に対応する座標を用いて算出した値である。結晶構造以外の場合は、エラーバーは 20 回のシミュレーションの標準誤差を表す。



**図 SE-11. DNA 原子モデル再構築結果のフラグメントライブラリ依存性** A) 自己再構築テストによる評価。構築したヌクレオチドと、もとの結晶構造との RMSD。末端のヌクレオチド以外についての結果を示した。 B-D) CGMD スナップショットに 1 K の緩和計算を実施後、原子モデルを構築した。 B)ワトソン-クリック塩基対の再現性。 C) O3'-P-O5'共有結合角 D) The O3'-P 共有結合長。大文字のアルファベットはフラグメントライブラリの ID を示す。カッコ内にはライブラリ構築に用いた構造ファイル(PDB ファイル)の数を示した。"Main"は第二章で主に利用したライブラリである。それぞれのフラグメントライブラリの詳細は付録 C、付録 E 表 SE-5 に記述した。グラフ中、"結晶構造"は結晶構造での値を示す。結晶構造以外では、エラーバーはテストセット A を対象に行った 20 回のシミュレーションの標準誤差を示す。

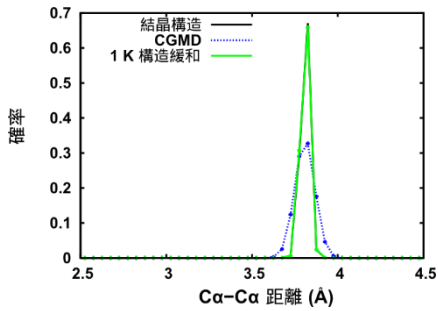


図 SE-12. 隣接するアミノ酸の Ca 原子間距離の分布 結晶構造、300 K の CGMD 直後、1 K での構造緩和計算後、の三つの状況を比較した。結晶構造以外では、エラーバーは 20 回のシミュレーションの標準誤差を表す。

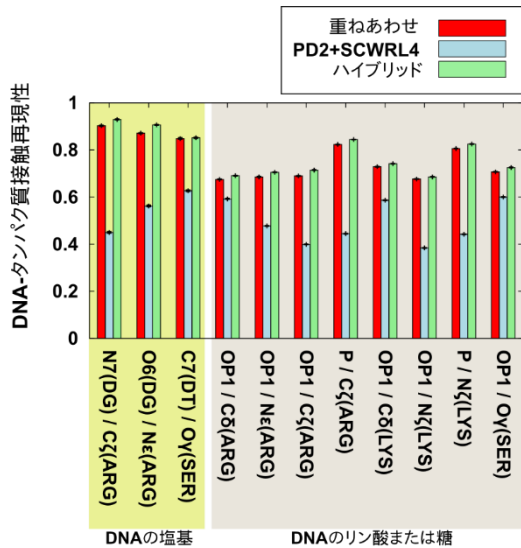


図 SE-13. エネルギー最小化計算前の DNA-タンパク質間の接触の再現性。ここでは、異なる 3 つのタンパク質モデル構築法での再現性の違いを示した。エラーバーは標準誤差を示す(N=20)。“重ね合わせ”では、実験によって決定されたタンパク質結晶構造を、粗視化モデルに重ね合わせた。“PD2+SCWRL4”では、モデリングソフトウェア PD2 CA2MAIN を用いて主鎖を構築した後で、ソフトウェア SCWRL4 で側鎖を構築した。“ハイブリッド”では、主鎖の構築に PD2 CA2MAIN を利用し、DNA-タンパク質相互作用面の側鎖構築に既知の結晶構造情報を利用した。その他の領域の側鎖は SCWRL4 を用いてモデリングした。DG がグアニン塩基を、DT はチミン塩基を示す。

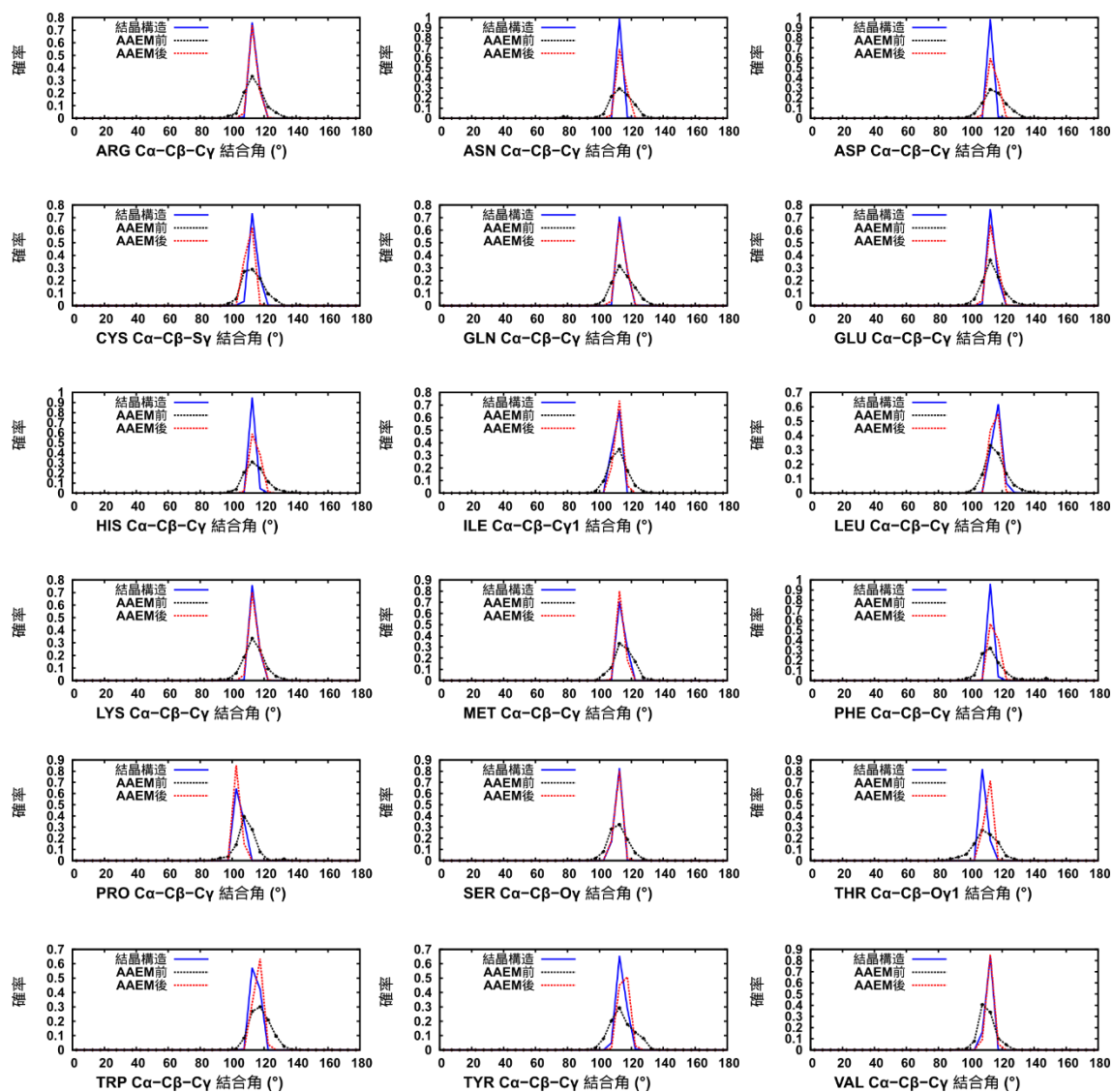


図 SE-14. タンパク質-DNA 相互作用面のアミノ酸側鎖構造の確認 タンパク質-DNA 相互作用面にあるとしてモデリングしたアミノ酸の、 $C_{\beta}$ 原子を頂点とする共有結合角の分布を調べた。グリシンとアラニンを除いた18アミノ酸について示した。結晶構造中での分布と、ハイブリッドアプローチでモデリングしたエネルギー最小化前 (AAEM前)、エネルギー最小化後 (AAEM後)の構造について示してある。エネルギー最小化前については、エラーバーは標準誤差を示す (N=20)。エネルギー最小化後の構造では、続いて10ピコ秒のMDを実行できた構造のみを使い、分布を出した。

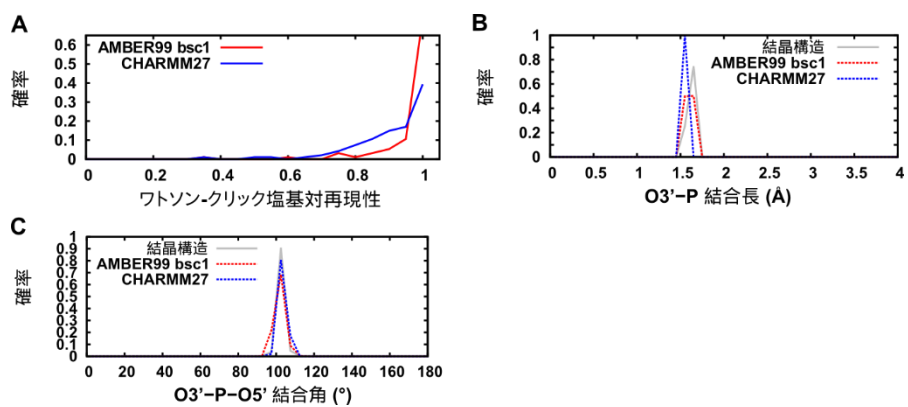


図 SE-15. DNA再構築モデルの力場依存性 A) ワトソン-クリック塩基対の再現性 B) O3'原子とP原子の共有結合角 C) O3'原子、P原子、O5'原子の共有結合角 “結晶構造”(テスト-セット A の結晶構造中での分布)とエネルギー最小化計算後のデータを示した。なお前者はエネルギー最小化計算後に 10 ピコ秒の MD を実行できた構造のみを使って分布を出した。

表 SE-1. 徐冷法のスケジュール

MD ステップ	初期温度	最終温度	温度を下げる回数
$10^3$	300 K	1 K	20
$10^4$	300 K	1 K	200
$10^5$	300 K	1 K	2000
$10^6$	300 K	1 K	20000
$10^4$	150 K	1 K	200
$10^4$	75 K	1 K	200
$10^4$	37.5 K	1 K	200
$10^4$	20 K	1 K	200
$10^4$	10 K	1 K	200

表 SE-2. 高い WC over0.98 値を実現した DNA 再構築のパラメータ  $w_j$

順位	WC over0.98	$w_{Si-1}$	$w_{Pi}$	$w_{Si}$	$w_{Bi}$	$w_{Pi+1}$
1	0.122	0.1	0.1	4.0	4.0	1.0
2	0.118	0.1	1.0	4.0	4.0	0.1
3	0.117	0.1	0.1	4.0	4.0	0.1
4	0.115	0.1	1.0	4.0	4.0	1.0
5	0.113	0.1	0.1	4.0	1.0	0.1
5	0.113	0.1	0.1	1.0	1.0	0.1
7	0.109	1.0	0.1	4.0	4.0	1.0
8	0.108	0.1	1.0	4.0	1.0	0.1

パラメータ  $w_j (w_{Si-1}, w_{Pi}, w_{Si}, w_{Bi}, w_{Pi+1})$  を 0.1, 1.0, 4.0 のいずれかにし、DNA の再構築を行った。WC over0.98 が大きい順にパラメータセットを並べ、上位 8 セットをまとめた。



表 SE-3. 高い  $P(\text{RMSD}<0.2)$  値を実現した DNA 再構築のパラメータ  $w_j$

順位	$P(\text{RMSD}<0.2)$	$w_{Si-1}$	$w_{Pi}$	$w_{Si}$	$w_{Bi}$	$w_{Pi+1}$
1	0.559	0.1	1.0	1.0	1.0	0.1
2	0.557	0.1	1.0	4.0	1.0	0.1
3	0.555	0.1	0.1	4.0	1.0	1.0
4	0.554	0.1	1.0	4.0	1.0	1.0
4	0.554	0.1	1.0	1.0	1.0	1.0
4	0.554	0.1	0.1	4.0	1.0	0.1
7	0.552	0.1	0.1	1.0	1.0	1.0
8	0.550	0.1	0.1	1.0	1.0	0.1

パラメータ  $w_j (w_{Si-1}, w_{Pi}, w_{Si}, w_{Bi}, w_{Pi+1})$  を 0.1, 1.0, 4.0 のいずれかにし、DNA の再構築を行った。  $P(\text{RMSD}<0.2)$  の大きい順にパラメータセットを並べ、上位 8 セットをまとめた。

表 SE-4. DNA の原子モデルをうまく再構築できなかった構造

PDB ID	DNA 鎖 ID	ヌクレオチド ID
3V9X	F	6
3M8R	C	202
3KHR	H	1804
2H7G	Z	517
1S32	I	3

うまく再構築できなかったヌクレオチドの PDB ID と PDB ファイル中での残基番号を示した。自己再構築テストで結晶構造との RMSD が 1.5 Å を超えたとき、モデリングに失敗したとみなした。パラメータ検討時に試した 932 通りのパラメータすべてでモデリングに失敗した構造を表にまとめた。

表 SE-5. フラグメントライブラリごとの DNA フラグメント(ヌクレオチド)数

ID	塩基タイプごとのフラグメント数			
	T	C	A	G
A	9,188	10,701	8,839	11,580
B	4,070	4,761	3,859	5,271
C	2,585	3,018	2,483	3,289
D	1,230	1,430	1,209	1,565
E	622	698	637	742
F	7,376	7,841	7,201	7,993
G	1,812	2,860	1,638	3,587

表に示したフラグメントライブラリは、DNA 再構築法の結果がフラグメントライブラリの内容にどの程度影響されるかを調べたときのものである(付録 D 参照)。第二章で主に用いたフラグメントライブラリについては本文 2.2.2 節に記載した。

表 SE-6. エネルギー最小化後のタンパク質-DNA 間接触の再現性

DNA	原子 タンパク 質	エネルギー最小化計算後 *		
		重ね合わせ	PD2+SCWRL	ハイブリッド
N7 (guanine)	C <sub>ζ</sub> (ARG)	0.862 ± 0.0074	0.463 ± 0.0115	0.899 ± 0.0111
O6 (guanine)	N <sub>ε</sub> (ARG)	0.842 ± 0.0039	0.553 ± 0.0117	0.875 ± 0.0078
C7 (thymine)	O <sub>γ</sub> (SER)	0.817 ± 0.0202	0.674 ± 0.0196	0.820 ± 0.0160
OP1	C <sub>δ</sub> (ARG)	0.740 ± 0.0012	0.602 ± 0.0160	0.739 ± 0.0148
OP1	C <sub>ζ</sub> (ARG)	0.743 ± 0.0063	0.416 ± 0.0087	0.742 ± 0.0138
OP1	N <sub>ε</sub> (ARG)	0.744 ± 0.0040	0.490 ± 0.0110	0.745 ± 0.0133
OP2	C <sub>δ</sub> (ARG)	0.764 ± 0.0017	0.597 ± 0.0096	0.757 ± 0.0064
OP2	C <sub>ζ</sub> (ARG)	0.762 ± 0.0024	0.441 ± 0.0070	0.762 ± 0.0106
P	C <sub>ζ</sub> (ARG)	0.840 ± 0.0032	0.462 ± 0.0060	0.845 ± 0.0131
OP1	C <sub>δ</sub> (LYS)	0.767 ± 0.0052	0.624 ± 0.0075	0.764 ± 0.0078
OP1	N <sub>ζ</sub> (LYS)	0.692 ± 0.0020	0.415 ± 0.0081	0.687 ± 0.0084
OP2	N <sub>ζ</sub> (LYS)	0.685 ± 0.0040	0.396 ± 0.0049	0.683 ± 0.0074
P	N <sub>ζ</sub> (LYS)	0.781 ± 0.0035	0.454 ± 0.0054	0.774 ± 0.0048
OP1	O <sub>γ</sub> (SER)	0.737 ± 0.0094	0.595 ± 0.0091	0.748 ± 0.0053
OP2	O <sub>γ</sub> (SER)	0.741 ± 0.0067	0.579 ± 0.0035	0.757 ± 0.0072

\*平均値 ± 標準誤差 (N=5).

表 SE-7. エネルギー最小化前のタンパク質-DNA 相互作用の再現性

DNA	原子 タンパク 質	エネルギー最小化計算前*		
		重ね合わせ	PD2+SCWRL	ハイブリッド
N7 (guanine)	C <sub>ζ</sub> (ARG)	0.903 ± 0.0028	0.450 ± 0.0038	0.929 ± 0.0024
O6 (guanine)	N <sub>ε</sub> (ARG)	0.871 ± 0.0020	0.562 ± 0.0035	0.906 ± 0.0019
C7 (thymine)	O <sub>γ</sub> (SER)	0.848 ± 0.0037	0.627 ± 0.0037	0.852 ± 0.0027
OP1	C <sub>δ</sub> (ARG)	0.675 ± 0.0015	0.592 ± 0.0021	0.691 ± 0.0014
OP1	C <sub>ζ</sub> (ARG)	0.690 ± 0.0021	0.399 ± 0.0015	0.715 ± 0.0017
OP1	N <sub>ε</sub> (ARG)	0.686 ± 0.0019	0.477 ± 0.0016	0.705 ± 0.0013
OP2	C <sub>δ</sub> (ARG)	0.700 ± 0.0013	0.563 ± 0.0013	0.709 ± 0.0013
OP2	C <sub>ζ</sub> (ARG)	0.711 ± 0.0016	0.400 ± 0.0013	0.723 ± 0.0009
P	C <sub>ζ</sub> (ARG)	0.823 ± 0.0015	0.445 ± 0.0011	0.844 ± 0.0010
OP1	C <sub>δ</sub> (LYS)	0.729 ± 0.0016	0.587 ± 0.0015	0.742 ± 0.0013
OP1	N <sub>ε</sub> (LYS)	0.677 ± 0.0012	0.384 ± 0.0017	0.686 ± 0.0011
OP2	N <sub>ε</sub> (LYS)	0.644 ± 0.0014	0.368 ± 0.0012	0.666 ± 0.0015
P	N <sub>ε</sub> (LYS)	0.806 ± 0.0011	0.442 ± 0.0011	0.825 ± 0.0014
OP1	O <sub>γ</sub> (SER)	0.707 ± 0.0016	0.600 ± 0.0019	0.725 ± 0.0019
OP2	O <sub>γ</sub> (SER)	0.693 ± 0.0019	0.576 ± 0.0018	0.735 ± 0.0013

\*平均値 ± 標準誤差 (N=20)

表 SE-8. タンパク質-DNA 相互作用再現性の力場依存性

DNA	タンパク質	PD2+SCWRL		ハイブリッド	
		Amber99 bsc1	Charmm27	Amber99 bsc1	Charmm27
N7 (guanine)	C <sub>ζ</sub> (ARG)	0.443 ± 0.0100	0.443 ± 0.0159	0.992 ± 0.0078	0.996 ± 0.0039
O6 (guanine)	N <sub>ε</sub> (ARG)	0.380 ± 0.0226	0.400 ± 0.0182	0.950 ± 0.0083	0.980 ± 0.0119
C7 (thymine)	O <sub>γ</sub> (SER)	0.708 ± 0.0264	0.667 ± 0.0131	0.950 ± 0.083	0.950 ± 0.0083
OP1	C <sub>δ</sub> (ARG)	0.672 ± 0.0148	0.684 ± 0.0147	0.778 ± 0.0091	0.782 ± 0.0116
OP1	C <sub>ζ</sub> (ARG)	0.452 ± 0.0113	0.463 ± 0.0190	0.774 ± 0.0106	0.785 ± 0.0110
OP1	N <sub>ε</sub> (ARG)	0.536 ± 0.0158	0.544 ± 0.0122	0.770 ± 0.0094	0.773 ± 0.0104
OP2	C <sub>δ</sub> (ARG)	0.632 ± 0.0140	0.654 ± 0.0114	0.824 ± 0.0040	0.857 ± 0.0100
OP2	C <sub>ζ</sub> (ARG)	0.435 ± 0.0117	0.441 ± 0.0134	0.813 ± 0.0098	0.836 ± 0.0145
P	C <sub>ζ</sub> (ARG)	0.506 ± 0.0148	0.503 ± 0.0171	0.882 ± 0.0083	0.903 ± 0.0093
OP1	C <sub>δ</sub> (LYS)	0.584 ± 0.0089	0.571 ± 0.0111	0.834 ± 0.0138	0.814 ± 0.0176
OP1	N <sub>ζ</sub> (LYS)	0.445 ± 0.0157	0.447 ± 0.0127	0.746 ± 0.0163	0.734 ± 0.0137
OP2	N <sub>ζ</sub> (LYS)	0.364 ± 0.0127	0.353 ± 0.0050	0.731 ± 0.0169	0.774 ± 0.0077
P	N <sub>ζ</sub> (LYS)	0.448 ± 0.0136	0.461 ± 0.0096	0.829 ± 0.0140	0.845 ± 0.0144
OP1	O <sub>γ</sub> (SER)	0.615 ± 0.0253	0.589 ± 0.0157	0.744 ± 0.0224	0.746 ± 0.0130
OP2	O <sub>γ</sub> (SER)	0.572 ± 0.0181	0.574 ± 0.0143	0.758 ± 0.0128	0.776 ± 0.0120

\*平均値 ± 標準誤差 (N=5).

## 付録 F 第二章で利用した構造ファイルのリスト(PDB ID)

### フラグメントライブラリ作成に使用した構造ファイルの PDB エントリー

1j75, 1kfs, 1krp, 1lau, 1omh, 1pfe, 1po6, 1qsl, 1rnb, 1u1k, 1u1n, 1vtg, 1xjv, 1xvk, 1xvn, 2d7d, 2fdf, 2fdh, 2fvr, 2i5s, 2oxv, 2p2r, 2qfj, 2up1, 2z70, 3hts, 3i2o, 3i3m, 3i49, 3kjp, 3kqn, 3n1i, 3q0b, 3u7f, 3u7h, 3vok, 3zdd, 4hid, 4him, 4hio, 4hj5, 4hj9, 4hqu, 4i28, 4i29, 4i2b, 4ibv, 4ibw, 4kmf, 4lz1, 4noe, 4ou6, 4qm7, 4rec, 4xzf, 4zhn, 5bua, 5haw, 5hk3, 5hre, 5hrt, 5hru, 5iff, 5td5, 5tkz, 1a1g, 1a1h, 1a1k, 1a31, 1a73, 1a74, 1aay, 1ais, 1az0, 1azp, 1azq, 1b3t, 1b94, 1b95, 1bpy, 1brn, 1bss, 1bsu, 1ca6, 1cyq, 1dfm, 1dsz, 1e3m, 1e3o, 1egw, 1eon, 1eoo, 1eyu, 1fiu, 1g2d, 1g2f, 1g9y, 1gd2, 1gu4, 1gu5, 1i3w, 1ig7, 1ipp, 1jl1, 1je8, 1jk2, 1jnm, 1jx4, 1k3w, 1k4t, 1k61, 1kbu, 1kx3, 1kx5, 1l3l, 1l3s, 1l3t, 1l3u, 1lat, 1lli, 1llm, 1lmb, 1lv5, 1m07, 1m3h, 1m5r, 1mnn, 1mus, 1n39, 1n3a, 1n3f, 1n4l, 1nh2, 1njz, 1nk7, 1nkb, 1nkc, 1nke, 1orp, 1owg, 1ph6, 1pji, 1pjj, 1pue, 1q3f, 1qbj, 1qn3, 1qn6, 1qn9, 1qna, 1qne, 1qrv, 1qum, 1r2z, 1r71, 1rff, 1rg1, 1rg2, 1rgt, 1rh6, 1rva, 1rvc, 1rxw, 1sa3, 1skm, 1ssp, 1stx, 1suz, 1sx8, 1sxq, 1tdz, 1tsr, 1tup, 1u47, 1u48, 1u49, 1ua1, 1unm, 1vto, 1w0t, 1w0u, 1wb9, 1wte, 1wtv, 1wtx, 1x9m, 1xo0, 1xsn, 1xsp, 1xvr, 1yo5, 1ytb, 1ztt, 1zzi, 2ac0, 2ahi, 2asd, 2ata, 2b0d, 2bcq, 2bcr, 2bcu, 2bcv, 2bdp, 2c7o, 2c7q, 2c7r, 2cgp, 2d5v, 2ddg, 2dp6, 2e1c, 2e43, 2e52, 2euz, 2evf, 2evh, 2evi, 2evj, 2fld, 2fmp, 2fms, 2fr4, 2g1p, 2gij, 2hax, 2heo, 2hhq, 2hht, 2hhu, 2hhv, 2hhw, 2hr1, 2hvi, 2hw3, 2i06, 2ibt, 2ih2, 2ih4, 2imw, 2irf, 2nob, 2noh, 2np7, 2o49, 2oa8, 2ofi, 2og0, 2oyt, 2p0j, 2pfn, 2pfp, 2pqu, 2pxi, 2pyl, 2r1j, 2r2r, 2uvv, 2uvw, 2vbj, 2vbo, 2ve9, 2vih, 2vjv, 2voa, 2vs7, 2w35, 2w36, 2w7n, 2xcs, 2xhi, 2xo6, 2xzf, 2y1j, 2z3x, 2zkd, 2zo0, 2zo1, 3a46, 3aaf, 3bam, 3bdp, 3bep, 3bi3, 3bie, 3bkz, 3c46, 3c58, 3cmy, 3cvu, 3cvv, 3dpg, 3dvo, 3e40, 3exl, 3eyi, 3eyz, 3ez5, 3f21, 3f8j, 3fsi, 3g6p, 3g6t, 3g6v, 3g6x, 3g6y, 3g8x, 3g99, 3g9m, 3g9o, 3g9p, 3ga6, 3gfi, 3go3, 3go8, 3gp1, 3gpp, 3gpy, 3gv5, 3gv8, 3hp6, 3ht3, 3hwt, 3i0w, 3i0x, 3iag, 3iay, 3igm, 3isb, 3isc, 3jpn, 3jpo, 3jps, 3jpt, 3jtg, 3k0s, 3k58, 3k59, 3k5m, 3khc, 3khl, 3l8b, 3lwm, 3m7k, 3m8s, 3m9o, 3mda, 3mdc, 3mfh, 3mfi, 3mr3, 3mr6, 3nci, 3ndh, 3ndk, 3ngi, 3nh1, 3o1m, 3o1p, 3o1r, 3o1s, 3o1v, 3ohb, 3oju, 3or3, 3osf, 3osg, 3oym, 3pmn, 3pv8, 3pvv, 3q23, 3q8k, 3q8s, 3qei, 3qer, 3qev, 3qew, 3qex, 3qmb, 3qmc, 3qmd, 3qmi, 3qws, 3qx3, 3qz7, 3qz8, 3rax, 3rh6, 3rjh, 3rjj, 3rkq, 3rtv, 3rzg, 3s57, 3s8q, 3si8, 3sjm, 3spl, 3spy, 3sq0, 3sq1, 3sq2, 3ssc, 3sv3, 3sv4, 3syz, 3t3f, 3tan, 3taq, 3tar, 3ted, 3tfr, 3tfs, 3thv, 3ti0, 3u6c, 3u6d, 3u6o, 3u6p, 3u6y, 3ugo, 3uk3, 3ulp, 3uta, 3v6t, 3v9z, 3vag, 3vaj, 3vd6, 3vk8, 3vke, 3wvk, 3zda, 3zh2, 3zvm, 4b21, 4b9l, 4b9n, 4b9u, 4bwj, 4bxo, 4c8l, 4c8m, 4cn2, 4cn5, 4crx, 4df8, 4dfj, 4dfk, 4dfm, 4dfp, 4dl3, 4dl7, 4doa, 4dob, 4doc, 4dqp, 4dqq, 4dqr, 4ds5, 4dse, 4dsf, 4dsi, 4dtj, 4dtm, 4dtp, 4dts, 4dtu, 4du1, 4du3, 4dwi, 4e3s, 4e9f, 4ea4, 4ea5, 4ebe, 4ecq, 4ecr, 4ecs, 4ecu, 4ecv, 4ecx, 4ecy, 4ed0, 4ed1, 4elv, 4esj, 4ez6, 4ez9, 4f2r, 4f4k, 4f5n, 4f5o, 4f5p, 4ff3, 4ff4, 4ff7, 4ffg, 4fjh, 4fji, 4fjj, 4fjl, 4fjm, 4fjn, 4fjx, 4fk0, 4fk2, 4fk4, 4flw,

4g4n, 4g4q, 4g92, 4gck, 4gjp, 4gjr, 4gxi, 4gxj, 4gzn, 4h0e, 4hcb, 4hje, 4hly, 4hn5, 4hp3, 4hue,  
 4i2o, 4j2b, 4j2e, 4j9l, 4jcy, 4jrp, 4k4h, 4k96, 4k99, 4kb0, 4klf, 4klh, 4kli, 4klt, 4klu, 4lg0, 4ljr,  
 4lmg, 4lnq, 4lvi, 4lvj, 4lvl, 4lvs, 4m0a, 4m3w, 4m3y, 4m41, 4m45, 4m8o, 4m94, 4m95, 4m9e,  
 4m9g, 4m9j, 4m9v, 4mhg, 4ndh, 4nm6, 4o0i, 4o3n, 4o3o, 4o3p, 4ofa, 4oln, 4osr, 4p2h, 4phe,  
 4q0w, 4q8e, 4qeo, 4qju, 4qr9, 4qwb, 4qwd, 4r2q, 4r55, 4r64, 4r65, 4rkh, 4rnn, 4roe, 4rpx, 4rpy,  
 4rq0, 4rq4, 4rq8, 4rtj, 4rtk, 4tu8, 4tup, 4tur, 4u0y, 4uaw, 4uay, 4ub3, 4ub4, 4ubb, 4wecg, 4wls,  
 4x9j, 4xeg, 4xo0, 4xpc, 4xpe, 4xrm, 4ycx, 4yhx, 4yov, 4yox, 4yp3, 4yr2, 4yr3, 4z47, 4z5d, 4z7z,  
 4z8f, 5a0w, 5a3d, 5av6, 5bol, 5bom, 5bpc, 5c5j, 5ciy, 5cl3, 5cl6, 5cl7, 5cla, 5cld, 5d9i, 5d9y,  
 5dba, 5deu, 5dff, 5dfh, 5dfj, 5dg0, 5dg8, 5dg9, 5dgb, 5dlf, 5dqh, 5dwa, 5e67, 5e69, 5e6a, 5eea,  
 5ewe, 5ewf, 5ewg, 5exh, 5f8a, 5ff8, 5fhd, 5g34, 5hnf, 5hr9, 5hrb, 5iii, 5iik, 5iil, 5iim, 5iin, 5inp,  
 5inq, 5j0o, 5j0p, 5j0q, 5j0r, 5j0s, 5j0u, 5j0y, 5j29, 5j2j, 5j1w, 5jre, 5jxy, 5ke6, 5kfa, 5kfb, 5kfd,  
 5kfg, 5k fj, 5kfn, 5kfr, 5kfs, 5kfu, 5kfx, 5kg0, 5kg1, 5kg2, 5kg3, 5kg4, 5kl2, 5kl3, 5kn9, 5krb,  
 5l0m, 5l1k, 5l1l, 5l9x, 5tb8, 5trd, 102d, 109d, 110d, 115d, 122d, 123d, 133d, 151d, 152d, 159d,  
 161d, 168d, 182d, 183d, 188d, 190d, 191d, 192d, 196d, 197d, 198d, 1ana, 1bd1, 1bna, 1cgc,  
 1cn0, 1cvy, 1d12, 1d15, 1d16, 1d17, 1d21, 1d23, 1d35, 1d36, 1d38, 1d39, 1d43, 1d44, 1d45,  
 1d46, 1d56, 1d60, 1d61, 1d63, 1d64, 1d65, 1d67, 1d8g, 1d90, 1d91, 1da0, 1dj6, 1dl8, 1dnt,  
 1dnx, 1dpl, 1dpn, 1eg6, 1em0, 1en3, 1en8, 1en9, 1ene, 1enn, 1evp, 1fd5, 1fdg, 1fhy, 1fms, 1fn2,  
 1g75, 1g8n, 1g8u, 1g8v, 1hq7, 1i0f, 1i0k, 1i0o, 1i0t, 1i2y, 1i47, 1i5w, 1ick, 1ih4, 1ikk, 1ilc,  
 1ims, 1j8l, 1jgr, 1jo2, 1jpp, 1jux, 1k9g, 1kci, 1kfk, 1l4j, 1l6b, 1ljx, 1m6g, 1m77, 1ma8, 1mf5,  
 1msw, 1nab, 1nqs, 1nzs, 1o56, 1ofx, 1omk, 1p4y, 1pjj, 1prp, 1pwf, 1qyk, 1qyl, 1r41, 1r68, 1rqy,  
 1s2r, 1s45, 1sgs, 1ub8, 1uhx, 1uhy, 1v3n, 1vaq, 1vj4, 1vro, 1vt8, 1vtb, 1vth, 1vtt, 1vtu, 1vtv,  
 1vtw, 1vty, 1wqy, 1xa2, 1xam, 1xcs, 1xcu, 1xjy, 1y84, 1y86, 1z5t, 1z7i, 1zbl, 1zf0, 1zfl, 1zfb,  
 1zfb, 1zfc, 1zff, 1zfm, 1zje, 1zph, 1zpi, 200d, 208d, 213d,  
 216d, 223d, 224d, 227d, 232d, 234d, 238d, 239d, 242d, 243d, 244d, 245d, 246d, 258d, 263d,  
 265d, 266d, 270d, 271d, 272d, 275d, 277d, 279d, 293d, 295d, 298d, 2b1b, 2b2b, 2b3e, 2d34,  
 2d47, 2dgc, 2dlj, 2dp7, 2dpc, 2dqp, 2dqq, 2elg, 2et0, 2fij, 2fil, 2g8f, 2g8k, 2g8v, 2g8w, 2gpx,  
 2gw0, 2gwa, 2gwq, 2hc7, 2hri, 2i2i, 2nsk, 2orh, 2pl8, 2qef, 2r7y, 2rf3, 302d, 305d, 306d, 311d,  
 314d, 315d, 317d, 318d, 320d, 324d, 331d, 334d, 336d, 337d, 338d, 346d, 352d, 355d, 360d,  
 365d, 366d, 368d, 369d, 370d, 381d, 382d, 384d, 385d, 388d, 390d, 394d, 395d, 396d, 398d,  
 3cco, 3cdm, 3co3, 3dnb, 3e9w, 3eru, 3ey2, 3f8o, 3fl6, 3fqb, 3ft6, 3ggi, 3gjl, 3gnk, 3gsj, 3gsk,  
 3hr3, 3i5e, 3iff, 3igt, 3iki, 3k18, 3kjo, 3kq8, 3ltr, 3n4n, 3nz7, 3o3f, 3omj, 3ot0, 3oz3, 3oz4,  
 3pbx, 3q61, 3qf8, 3qrn, 3sd8, 3t5e, 3t86, 3tvb, 3u08, 3u2n, 3u38, 3u89, 3ukb, 3ukc, 3uke, 3uld,  
 3uq0, 3zp8, 403d, 408d, 411d, 412d, 414d, 417d, 427d, 431d, 432d, 440d, 441d, 443d, 448d,  
 449d, 452d, 453d, 454d, 455d, 456d, 461d, 467d, 473d, 476d, 485d, 4agz, 4ah0, 4ah1, 4d9y,  
 4dqs, 4dy8, 4e2r, 4e60, 4e7y, 4e8s, 4e8x, 4e95, 4f2x, 4f2y, 4f3u, 4f4n, 4f8i, 4fs5, 4gju, 4glc,  
 4gre, 4h5a, 4hif, 4itd, 4izq, 4khu, 4kw0, 4kwx, 4l25, 4l26, 4ltf, 4ltg, 4m3i, 4mkw, 4ndf, 4o5x,

4o5z, 4ocb, 4okl, 4puq, 4qc7, 4qcl, 4qio, 4qno, 4r15, 4r45, 4r4d, 4re7, 4rhd, 4rip, 4rnk, 4ro8, 4rok, 4u5m, 4u8a, 4u8b, 4u92, 4x1a, 4xsn, 4ys5, 4z4b, 5awh, 5axe, 5axf, 5b2p, 5b2q, 5b2r, 5b2s, 5bxw, 5bz7, 5bz9, 5cdb, 5dsb, 5e36, 5ebi, 5erz, 5es0, 5et9, 5ewb, 5ezf, 5fhj, 5ihd, 5jvw, 5us2, 5usg, 5wsp, 5wsq, 7bna, 8bna, 9bna

原子モデル再構築のテストに使った構造ファイルの PDB エントリー

テストセット A	1a1f, 1ali, 1alj, 1b97, 1bc7, 1bc8, 1bnz, 1ca5, 1cdw, 1ckq, 1cz0, 1d02, 1du0, 1dux, 1f0v, 1gtw, 1h6f, 1i3j, 1i6j, 1i7d, 1jb7, 1jgg, 1jk1, 1mjm, 1mjo, 1n6j, 1oj8, 1p47, 1p71, 1qn5, 1rvb, 1s32, 1t9i, 1t9j, 1tv9, 1wd0, 1wd1, 1wto, 1wtq, 1wtr, 1wtw, 1xil, 1zaa, 1zs4, 1ztw, 2a07, 2agq, 2axy, 2bop, 2dgc, 2e42, 2ex5, 2fjv, 2fjw, 2fjx, 2h7g, 2han, 2nq9, 2o4a, 2oaa, 2r2t, 2rdj, 2vy1, 2wiw, 2xqc, 3bm3, 3c0w, 3c2p, 3cbb, 3d0p, 3d2w, 3dsd, 3fc3, 3fdq, 3fyl, 3g6u, 3g8u, 3gox, 3hdd, 3jpb, 3jxb, 3jxc, 3jxd, 3khr, 3kmd, 3kxt, 3l2c, 3ldy, 3lwh, 3m8r, 3mva, 3ngz, 3ogu, 3ojs, 3oqg, 3po4, 3pvi, 3py8, 3q22, 3q24, 3qet, 3rh4, 3s9h, 3si6, 3sm4, 3snn, 3spd, 3ut9, 3v1z, 3v9s, 3v9w, 3v9x, 3va0, 4a75, 4a76, 4aij, 4aik, 4aqx, 4b9s, 4dav, 4dlg, 4dqi, 4elt, 4elu, 4flh, 4f4w, 4f4x, 4ff2, 4fj5, 4fj8, 4fj9, 4fjk, 4fpv, 4gz0, 4hc9, 4is1, 4j2a, 4j2d, 4jwm, 4k1g, 4kb1, 4klg, 4klj, 4kl, 4klm, 4lb6, 4lup, 4m3r, 4m3t, 4m3u, 4m3x, 4m3z, 4m42, 4mde, 4mdf, 4o6a, 4ofe, 4qwa, 4qwe, 4s0n, 4uno, 4y60, 4z5h, 5b0z, 5b31, 5bnh, 5cbx, 5cby, 5cbz, 5d4r, 5d8k, 5e41, 5e6c, 5egb, 5ei9, 5hlk, 5ink, 5j2f, 5kl6, 5szt
テストセット B (一本鎖 DNA-タンパク質複合体)	1i7d, 1mw8, 1u1l, 1ulm, 1ulo, 1ulp, 1ulq, 1ulr, 2c62, 2es2, 2i0q, 3dsd, 3ngo, 3u7g, 4a15, 4dih, 4dii, 4fpv, 4hik, 4hj7, 4hj8, 4hja, 4i2a, 4i2f, 4usg, 5dlo, 5fb2, 5h8w, 5hp4, 5luw
テストセット B (複数 DNA 鎖-タンパク質複合体)	1a1f, 1ali, 1alj, 1a3q, 1awc, 1b97, 1bc7, 1bc8, 1bf4, 1bgb, 1bhm, 1bnz, 1bua, 1by4, 1c8c, 1ca5, 1cdw, 1cz0, 1d02, 1d2i, 1dc1, 1du0, 1dux, 1ebm, 1emh, 1emj, 1eo3, 1eo4, 1esg, 1ewn, 1ewq, 1f0v, 1f44, 1fjl, 1fyl, 1fym, 1g38, 1g9z, 1gtw, 1h6f, 1i3j, 1i6j, 1i8m, 1jb7, 1jgg, 1jk1, 1jxl, 1k3x, 1k82, 1l1t, 1l1z, 1l2c, 1l2d, 1l3v, 1l5u, 1lww, 1lwy, 1m3q, 1mey, 1mjm, 1mjo, 1n48, 1n6j, 1njw, 1njx, 1nfy, 1nk0, 1nk4, 1nk5, 1nk6, 1nk8, 1nk9, 1nkp, 1nlw, 1nnj, 1nvp, 1oe4, 1orn, 1owf, 1p47, 1p71, 1pm5, 1puf, 1qn4, 1qn5, 1qn8, 1qzg, 1rfi, 1rv5, 1rvb, 1rys, 1rzt, 1s0o,



<p> 1s10, 1s32, 1s9f, 1sfu, 1sl1, 1sx5, 1t7p, 1t9i, 1t9j, 1tez, 1tk5, 1tro, 1tv9,  1u45, 1u4b, 1u8b, 1ua0, 1uut, 1wd0, 1wd1, 1wto, 1wtp, 1wtq, 1wtr,  1wtw, 1xc8, 1xc9, 1xi1, 1xyi, 1y8z, 1zaa, 1zjm, 1zrf, 1zs4, 1ztw, 2a07,  2a66, 2adw, 2agq, 2aoq, 2aor, 2axy, 2b0e, 2bam, 2bcs, 2bq3, 2br0, 2c7p,  2dem, 2dnj, 2e42, 2ea0, 2etw, 2euv, 2euw, 2eux, 2evg, 2ex5, 2f5n, 2f5o,  2f5p, 2fjv, 2fjw, 2fjx, 2fmq, 2fqz, 2gb7, 2gig, 2h7g, 2han, 2hdd, 2hhs,  2hos, 2hot, 2ht0, 2i13, 2i9g, 2ih5, 2is6, 2iso, 2isp, 2itl, 2jef, 2jej, 2jg3,  2nll, 2noe, 2np6, 2nq9, 2o4a, 2oaa, 2odi, 2opf, 2pfo, 2pfq, 2pvi, 2py5,  2pyj, 2q10, 2r2t, 2rdj, 2uyc, 2uz4, 2vbl, 2vbn, 2vla, 2vs8, 2vtb, 2w42,  2wbs, 2wiw, 2wq7, 2x6v, 2xc9, 2xqc, 2xrz, 2xsd, 2xzu, 3a4k, 3bm3,  3brg, 3bs1, 3btx, 3c0w, 3c1b, 3c29, 3c2m, 3c2p, 3c5g, 3cbb, 3clz, 3co6,  3coa, 3d0a, 3d0p, 3dw9, 3e3y, 3e6c, 3eeo, 3exj, 3fc3, 3fde, 3fdq, 3fsp,  3fyl, 3g00, 3g0q, 3g6u, 3g8u, 3g97, 3g9i, 3gdx, 3gox, 3gpu, 3gpx, 3gq3,  3gq4, 3gq5, 3gv7, 3h4d, 3h8o, 3h8r, 3h8x, 3hdd, 3hpo, 3hw8, 3i8d, 3igc,  3imb, 3jpp, 3jpp, 3jpr, 3jr5, 3jx7, 3jxb, 3jxc, 3jxd, 3jxy, 3jxz, 3jy1,  3k57, 3kde, 3khr, 3kmd, 3kxt, 3l2c, 3lap, 3ldy, 3lwh, 3ljz, 3m4a, 3m8r,  3m9n, 3mby, 3mq6, 3mqy, 3mr2, 3mr5, 3mva, 3mxm, 3nae, 3ne6, 3ngz,  3o1o, 3o1t, 3o1u, 3o9x, 3ogu, 3oha, 3ojs, 3oqg, 3osn, 3p57, 3pnc, 3po4,  3pvi, 3px0, 3px4, 3px6, 3py8, 3q22, 3q24, 3q8p, 3q8q, 3qep, 3qes, 3qet,  3qi5, 3qnn, 3qno, 3rh4, 3rh5, 3rje, 3rjg, 3rjk, 3rr7, 3rrh, 3s5a, 3s9h, 3sar,  3sas, 3sat, 3sau, 3sav, 3sbj, 3si6, 3sm4, 3snn, 3spd, 3ssd, 3sz2, 3tap,  3u6e, 3u6l, 3u6m, 3u6q, 3u6s, 3uby, 3uiq, 3upq, 3ut9, 3utb, 3v1z, 3v9s,  3v9w, 3v9x, 3va0, 3vk7, 3vxx, 3vxx, 3wpc, 3zp5, 3zvn, 4a75, 4a76, 4aij,  4aik, 4aqx, 4b22, 4b23, 4b5f, 4b5j, 4b9m, 4b9s, 4b9v, 4bdp, 4c8k, 4c8n,  4c8o, 4cis, 4d6o, 4dav, 4df4, 4dkj, 4dl2, 4dl4, 4dlg, 4do9, 4dqi, 4ds4,  4dsk, 4dtn, 4dto, 4dtr, 4dtx, 4e0d, 4e0g, 4e0p, 4ect, 4ecw, 4ecz, 4ed2,  4ed3, 4ed7, 4ed8, 4ejy, 4elt, 4elu, 4f2s, 4f3o, 4f4w, 4f4x, 4f5r, 4f8r, 4fbt,  4ff2, 4fj5, 4fj8, 4fj9, 4fjk, 4fs2, 4g4o, 4g4r, 4glx, 4gxk, 4gz0, 4gz1,  4gz2, 4hc9, 4hsb, 4htu, 4huf, 4hug, 4ibu, 4is1, 4ix7, 4j2a, 4j2d, 4j9k,  4j9n, 4j9q, 4j9s, 4jwm, 4k1g, 4k4g, 4k98, 4kb1, 4khq, 4kld, 4kle, 4klg,  4klj, 4kll, 4klm, 4klo, 4klq, 4kls, 4l5r, 4lox, 4lup, 4lzg, 4m04, 4m3r,  4m3t, 4m3u, 4m3x, 4m3z, 4m42, 4m9l, 4mde, 4mdf, 4mfc, 4n56, 4n5s,  4ncb, 4nid, 4nig, 4nih, 4nii, 4o3q, 4o3r, 4o3s, 4o5k, 4o6a, 4ofe, 4opj,  4opk, 4osh, 4ost, 4osv, 4ot3, 4p4m, 4pcz, 4pd2, 4pdi, 4pgx, 4ppx, 4pw5,  4pw7, 4pzi, 4q45, 4qen, 4qtj, 4qwa, 4qwe, 4qz9, 4qza, 4qzb, 4r2a, 4r2c,  4r2d, 4r2e, 4r2p, 4r2r, 4r63, 4rdu, 4rnm, 4roc, 4rpz, 4rq2, 4rq3, 4rq7, </p>
---

	4rt2, 4rt3, 4rtl, 4rtq, 4rzt, 4s0n, 4tqr, 4tqs, 4uaz, 4ub5, 4ubc, 4uno, 4uqg, 4uqm, 4x5v, 4xa5, 4xqj, 4y60, 4yd1, 4yfu, 4ymm, 4ymo, 4yoy, 4yqw, 4yr0, 4z3a, 4z3c, 4z7b, 5av8, 5av9, 5b0z, 5b31, 5bnh, 5bt2, 5cbx, 5cby, 5cbz, 5chz, 5cl4, 5cl5, 5cl8, 5cl9, 5clb, 5clc, 5cle, 5d49, 5d4r, 5d4s, 5d8l, 5dfi, 5dpk, 5e24, 5e41, 5e6c, 5egb, 5eh2, 5ei9, 5eoz, 5g32, 5g35, 5hdn, 5hf7, 5hlk, 5hnh, 5hrd, 5hrq, 5hri, 5hto, 5ijj, 5iio, 5ink, 5inl, 5iwi, 5j0t, 5j0x, 5j2c, 5j2d, 5j2e, 5j2f, 5j2g, 5j2k, 5jh0, 5jk0, 5ke7, 5kfc, 5kfe, 5kff, 5kfh, 5kfi, 5kfk, 5kfl, 5kfm, 5kfo, 5kfp, 5kfq, 5kft, 5kfv, 5kfw, 5kfy, 5kfz, 5kg5, 5kg6, 5kg7, 5kl4, 5kl6, 5kl7, 5kn8, 5kub, 5l1j, 5l2x, 5szt, 5t01, 5t2w, 5tbc, 5tzv, 5u9h, 6mht
テストセット B (A 型 DNA)	137d, 138d, 160d, 189d, 1d13, 1d26, 1d78, 1d79, 1dns, 1dnz, 1f6e, 1y7f, 1y8l, 1y8v, 1y9f, 1y9s, 1yb9, 1ybc, 1zex, 1zey, 1zfa, 212d, 220d, 221d, 240d, 254d, 256d, 260d, 2d95, 2h05, 339d, 340d, 341d, 342d, 345d, 371d, 401d, 5fhl, 9dna
テストセット B (Z 型 DNA)	131d, 145d, 181d, 1d24, 1d48, 1d76, 1da2, 1dcg, 1dnf, 1m6r, 1woe, 1zna, 292d, 2f8w, 2hto, 2ie1, 2obz, 312d, 313d, 362d, 3p4j, 3wbo, 400d, 4dwy, 4e4o, 4fs6, 4hig, 4xqz, 5jzq
テストセット B (四重鎖 DNA)	1jrn, 1kfl, 1l1h, 1o0k, 1v3o, 2avh, 2gwe, 2hbn, 2o4f, 3eqw, 3es0, 3eui, 3eum, 3nyp, 3qxr, 3uyh, 4fxm, 4g0f, 4p1d, 4r47, 4wo2, 5ccw
テストセット C	1ca5, 1gtw, 1i3j, 1jk1, 1tv9, 1wtq, 2axy, 2dgc, 2e42, 2fjv, 2r2t, 3cbb, 3fc3, 3g8u, 3jxb, 3m8r, 3mva, 3v9x, 4fpv, 4klm, 4m3u, 4mde, 4mdf, 5e6c, 5ink

## 参考文献

- (1) Costa, A.; Hood, I. V.; Berger, J. M. Mechanisms for Initiating Cellular DNA Replication. *Annu Rev Biochem.* **2013**, *82*, 25–54.
- (2) Sainsbury, S.; Bernecky, C.; Cramer, P. Structural Basis of Transcription Initiation by RNA Polymerase II. *Nat. Rev. Mol. Cell. Biol.* **2015**, *16*, 129–143.
- (3) Burgers, P. M. J.; Kunkel, T. A. Eukaryotic DNA Replication Fork. *Annu Rev Biochem.* **2017**, *86*, 417–438.
- (4) Ozaki, S.; Katayama, T. DnaA Structure, Function, and Dynamics in the Initiation at the Chromosomal Origin. *Plasmid* **2009**, *62* (2), 71–82.
- (5) Hwang, D. S.; Kornberg, A. Opening of the Replication Origin of *Escherichia coli* by DnaA Protein with Protein HU or IHF. *J. Biol. Chem.* **1992**, *267*, 23083–23086.
- (6) Sakiyama, Y.; Kasho, K.; Noguchi, Y.; Kawakami, H.; Katayama, T. Regulatory Dynamics in the Ternary DnaA Complex for Initiation of Chromosomal Replication in *Escherichia coli*. *Nucleic Acids Res.* **2017**, *45* (21), 12354–12373.
- (7) Ozaki, S.; Noguchi, Y.; Hayashi, Y.; Miyazaki, E.; Katayama, T. Differentiation of the DnaA-oriC Subcomplex for DNA Unwinding in a Replication Initiation Complex. *J. Biol. Chem.* **2012**, *287* (44), 37458–37471.
- (8) Ozaki, S.; Katayama, T. Highly Organized DnaA-oriC Complexes Recruit the Single-Stranded DNA for Replication Initiation. *Nucleic Acids Res.* **2012**, *40* (4), 1648–1665.
- (9) Duderstadt, K. E.; Mott, M. L.; Crisona, N. J.; Chuang, K.; Yang, H.; Berger, J. M. Origin Remodeling and Opening in Bacteria Rely on Distinct Assembly States of the DnaA Initiator. *J. Biol. Chem.* **2010**, *285* (36), 28229–28239.
- (10) Duderstadt, K. E.; Chuang, K.; Berger, J. M. DNA Stretching by Bacterial Initiators Promotes Replication Origin Opening. *Nature* **2011**, *478* (7368), 209–213.
- (11) Ozaki, S.; Kawakami, H.; Nakamura, K.; Fujikawa, N.; Kagawa, W.; Park, S.-Y.; Yokoyama, S.; Kurumizaka, H.; Katayama, T. A Common Mechanism for the ATP-DnaA-Dependent Formation of Open Complexes at the Replication Origin. *J. Biol. Chem.* **2008**, *283* (13), 8351–8362.
- (12) Kaguni, J. M. Replication Initiation at the *Escherichia Coli* Chromosomal Origin. *Curr. Opin. Chem. Biol.* **2011**, *15* (5), 606–613.
- (13) Katayama, T.; Ozaki, S.; Keyamura, K.; Fujimitsu, K. Regulation of the Replication Cycle: Conserved and Diverse Regulatory Systems for DnaA and oriC. *Nat. Rev. Microbiol.* **2010**, *8* (3), 163–170.
- (14) Leonard, A. C.; Grimwade, J. E. Regulation of DnaA Assembly and Activity: Taking

- Directions from the Genome. *Annu Rev Microbiol.* **2011**, *65*, 19–35.
- (15) Bramhill, D.; Kornberg, A. Duplex Opening by DnaA Protein at Novel Sequences in Initiation of Replication at the Origin of the *E. coli* Chromosome. *Cell* **1988**, *52* (5), 743–755.
  - (16) McGarry, K. C.; Ryan, V. T.; Grimwade, J. E.; Leonard, A. C. Two Discriminatory Binding Sites in the *Escherichia coli* Replication Origin Are Required for DNA Strand Opening by Initiator DnaA-ATP. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (9), 2811–2816.
  - (17) Kawakami, H.; Keyamura, K.; Katayama, T. Formation of an ATP-DnaA-Specific Initiation Complex Requires DnaA Arginine 285, a Conserved Motif in the AAA+ Protein Family. *J. Biol. Chem.* **2005**, *280* (29), 27420–27430.
  - (18) Rozgaja, T. A.; Grimwade, J. E.; Iqbal, M.; Czerwonka, C.; Vora, M.; Leonard, A. C. Two Oppositely Oriented Arrays of Low-Affinity Recognition Sites in *oriC* Guide Progressive Binding of DnaA during *Escherichia coli* Pre-RC Assembly. *Mol. Microbiol.* **2011**, *82* (2), 475–488.
  - (19) Schaper, S.; Messer, W. Interaction of the Initiator Protein DnaA of *Escherichia coli* with Its DNA Target. *J. Biol. Chem.* **1995**, *270* (29), 17622–17626.
  - (20) Grimwade, J. E.; Ryan, V. T.; Leonard, A. C. IHF Redistributes Bound Initiator Protein, DnaA, on Supercoiled *oriC* of *Escherichia coli*. *Mol. Microbiol.* **2000**, *35* (4), 835–844.
  - (21) Blaesing, F.; Weigel, C.; Welzck, M.; Messer, W. Analysis of the DNA-Binding Domain of *Escherichia coli* DnaA Protein. *Mol. Microbiol.* **2000**, *36* (3), 557–569.
  - (22) Margulies, C.; Kaguni, J. M. Ordered and Sequential Binding of DnaA Protein to *oriC*, the Chromosomal Origin of *Escherichia coli*. *J. Biol. Chem.* **1996**, *271* (29), 17035–17040.
  - (23) Weigel, C.; Schmidt, A.; Rückert, B.; Lurz, R.; Messer, W. DnaA Protein Binding to Individual DnaA Boxes in the *Escherichia coli* Replication Origin, *oriC*. *EMBO J.* **1997**, *16* (21), 6574–6583.
  - (24) Rice, P. A.; Yang, S.; Mizuuchi, K.; Nash, H. A. Crystal Structure of an IHF-DNA Complex: A Protein-Induced DNA U-Turn. *Cell* **1996**, *87* (7), 1295–1306.
  - (25) Fujikawa, N.; Kurumizaka, H.; Nureki, O.; Terada, T.; Shirouzu, M.; Katayama, T.; Yokoyama, S. Structural Basis of Replication Origin Recognition by the DnaA Protein. *Nucleic Acids Res.* **2003**, *31* (8), 2077–2086.
  - (26) Roth, A.; Messer, W. The DNA-Binding Domain of the Initiator Protein DnaA. *EMBO J.* **1995**, *14* (9), 2106–2111.
  - (27) Erzberger, J. P.; Pirruccello, M. M.; Berger, J. M. The Structure of Bacterial DnaA: Implications for General Mechanisms Underlying DNA Replication Initiation. *EMBO J.*

- 2002, 21 (18), 4763–4773.
- (28) Erzberger, J. P.; Mott, M. L.; Berger, J. M. Structural Basis for ATP-Dependent DnaA Assembly and Replication-Origin Remodeling. *Nat. Struct. Mol. Biol.* **2006**, 13 (8), 676–683.
- (29) Sekimizu, K.; Bramhill, D.; Kornberg, A. ATP Activates DnaA Protein in Initiating Replication of Plasmids Bearing the Origin of the *E. coli* Chromosome. *Cell* **1987**, 50 (2), 259–265.
- (30) Noguchi, Y.; Sakiyama, Y.; Kawakami, H.; Katayama, T. The Arg Fingers of Key DnaA Protomers Are Oriented Inward within the Replication Origin *oriC* and Stimulate DnaA Subcomplexes in the Initiation Complex. *J. Biol. Chem.* **2015**, 290 (33), 20295–20312.
- (31) Abe, Y.; Jo, T.; Matsuda, Y.; Matsunaga, C.; Katayama, T.; Ueda, T. Structure and Function of DnaA N-Terminal Domains - Specific Sites and Mechanisms in Inter-DnaA Interaction and in DnaB Helicase Loading on *oriC*. *J. Biol. Chem.* **2007**, 282 (24), 17816–17827.
- (32) Kurokawa, K.; Nishida, S.; Emoto, A.; Sekimizu, K.; Katayama, T. Replication cycle-coordinated change of the adenine nucleotide-bound forms of DnaA protein in *Escherichia coli*. *EMBO J.* **1999**, 18 (23), 6642–6652.
- (33) Potoyan, D. A.; Zheng, W.; Komives, E. A.; Wolynes, P. G. Molecular stripping in the NF- $\kappa$ B/I $\kappa$ B/DNA genetic regulatory network. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, 113 (1), 110–115.
- (34) Mishra, G.; Levy, Y. Molecular Determinants of the Interactions between Proteins and ssDNA. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, 112 (16), 5033–5038.
- (35) Potoyan, D. A.; Savelyev, A.; Papoian, G. A. Recent Successes in Coarse-Grained Modeling of DNA. *WIREs Comput Mol Sci.* **2013**, 3 (1), 69–83.
- (36) Takada, S.; Kanada, R.; Tan, C.; Terakawa, T.; Li, W.; Kenzaki, H. Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations. *Acc. Chem. Res.* **2015**, 48 (12), 3026–3035.
- (37) Kenzaki, H.; Koga, N.; Hori, N.; Kanada, R.; Li, W.; Okazaki, K.; Yao, X.-Q.; Takada, S. CafeMol: A Coarse-Grained Biomolecular Simulator for Simulating Proteins at Work. *J. Chem. Theory Comput.* **2011**, 7 (6), 1979–1989.
- (38) Terakawa, T.; Kenzaki, H.; Takada, S. p53 Searches on DNA by Rotation-Uncoupled Sliding at C-Terminal Tails and Restricted Hopping of Core Domains. *J. Am. Chem. Soc.* **2012**, 134 (35), 14555–14562.
- (39) Li, W.; Terakawa, T.; Wang, W.; Takada, S. Energy Landscape and Multiroute Folding of Topologically Complex Proteins Adenylate Kinase and 2ouf-Knot. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, 109 (44), 17789–17794.

- (40) Li, W.; Wang, W.; Takada, S. Energy Landscape Views for Interplays among Folding, Binding, and Allostery of Calmodulin Domains. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (29), 10550–10555.
- (41) Terakawa, T.; Takada, S. Multiscale Ensemble Modeling of Intrinsically Disordered Proteins: p53 N-Terminal Domain. *Biophys. J.* **2011**, *101* (6), 1450–1458.
- (42) Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J. A Mesoscale Model of DNA and Its Renaturation. *Biophys. J.* **2009**, *96* (5), 1675–1690.
- (43) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and Energetic Factors: What Determines the Structural Details of the Transition State Ensemble And “en-Route” intermediates for Protein Folding? An Investigation for Small Globular Proteins. *J. Mol. Biol.* **2000**, *298* (5), 937–953.
- (44) Kenzaki, H.; Takada, S. Partial Unwrapping and Histone Tail Dynamics in Nucleosome Revealed by Coarse-Grained Molecular Simulations. *PLOS Comput. Biol.* **2015**, *11* (8), 1004443.
- (45) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat. Methods* **2015**, *12* (1), 7–8.
- (46) Zhang, Y. I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinformatics* **2008**, *9*:40.
- (47) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat. Protoc.* **2010**, *5*, 725–738.
- (48) Kenzaki, H; et al. *CafeMol 2.1 Manual*; **2013**.
- (49) Case, D. A.; Darden, T. A.; Cheatham, T. E., III.; Simmerling, C. L.; Wang, J.; Duke, R. ; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; et al. AMBER 13. University of California, San Francisco, **2012**.
- (50) Macke, T. J.; Case, D. A. Modeling Unusual Nucleic Acid Structures. In *MOLECULAR MODELING OF NUCLEIC ACIDS*; Leontis, NB and SantaLucia, J., Jr., Eds.; ACS SYMPOSIUM SERIES; 1998; Vol. 682, pp 379–393.
- (51) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10*, 845–858.
- (52) Krivov, G. G.; Shapovalov, M. V; Dunbrack, R. L., Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins: Struct., Funct., Bioinf.* **2009**, *77* (4), 778–795.
- (53) Schrödinger, LLC. *The PyMOL Molecular Graphics System, Version 1.7*.
- (54) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: A

- High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29* (7), 845–854.
- (55) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.
- (56) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718.
- (57) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.* **2001**, *7* (8), 306–317.
- (58) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56.
- (59) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In *SOLVING SOFTWARE CHALLENGES FOR EXASCALE*; Markidis, S., and Laure, E., Eds.; Lecture Notes in Computer Science; 2015; Vol. 8759, pp 3–27.
- (60) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (61) Bondi, A. Van Der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68* (3), 441–451.
- (62) Guy, A. T.; Piggot, T. J.; Khalid, S. Single-Stranded DNA within Nanopores: Conformational Dynamics and Implications for Sequencing; a Molecular Dynamics Simulation Study. *Biophys. J.* **2012**, *103* (5), 1028–1036.
- (63) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct. Bioinf.* **2006**, *65* (3), 712–725.
- (64) Sorin, E. J.; Pande, V. S. Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. *Biophys. J.* **2005**, *88* (4), 2472–2493.
- (65) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of  $\alpha/\gamma$  Conformers. *Biophys. J.* **2007**, *92* (11), 3817–3829.
- (66) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.
- (67) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593.

- (68) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190.
- (69) Goga, N.; Rzepiela, A. J.; de Vries, A. H.; Marrink, S. J.; Berendsen, H. J. C. Efficient Algorithms for Langevin and DPD Dynamics. *J. Chem. Theory Comput.* **2012**, *8* (10), 3637–3649.
- (70) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472.
- (71) Miyamoto, S.; Kollman, P. A. SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13* (8), 952–962.
- (72) Garner, J.; Crooke, E. Membrane Regulation of the Chromosomal Replication Activity of *E. coli* DnaA Requires a Discrete Site on the Protein. *EMBO J.* **1996**, *15* (13), 3477–3485.
- (73) Messer, W.; Hartmann-kühlein, H.; Langer, U.; Mahlow, E.; Roth, A.; Schaper, S.; Urmoneit, B.; Woelker, B. The Complex for Replication Initiation of *Escherichia coli*. *Chromosoma* **1992**, *102* (1, S), S1–S6.
- (74) Fang, L.; Davey, M. J.; O'Donnell, M. Replisome Assembly at *oriC*, the Replication Origin of *E. coli*, Reveals an Explanation for Initiation Sites Outside an Origin. *Mol. Cell* **1999**, *4* (4), 541–553.
- (75) Keyamura, K.; Abe, Y.; Higashi, M.; Ueda, T.; Katayama, T. DiaA Dynamics Are Coupled with Changes in Initial Origin Complexes Leading to Helicase Loading. *J. Biol. Chem.* **2009**, *284* (37), 25038–25050.
- (76) Kaur, G.; Vora, M. P.; Czerwonka, C. A.; Rozgaja, T. A.; Grimwade, J. E.; Leonard, A. C. Building the Bacterial Oriosome: High-Affinity DnaA Recognition Plays a Role in Setting the Conformation of *oriC* DNA. *Mol. Microbiol.* **2014**, *91* (6), 1148–1163.
- (77) Simmons, L. A.; Felczak, M.; Kaguni, J. M. DnaA Protein of *Escherichia coli*: Oligomerization at the *E. coli* Chromosomal Origin Is Required for Initiation and Involves Specific N-Terminal Amino Acids. *Mol. Microbiol.* **2003**, *49* (3), 849–858.
- (78) Felczak, M. M.; Simmons, L. A.; Kaguni, J. M. An Essential Tryptophan of *Escherichia coli* DnaA Protein Functions in Oligomerization at the *E. coli* Replication Origin. *J. Biol. Chem.* **2005**, *280* (26), 24627–24633.
- (79) Lai, W. K. M.; Pugh, B. F. Understanding Nucleosome Dynamics and Their Links to Gene Expression and DNA Replication. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 548–562.
- (80) Hustedt, N.; Durocher, D. The Control of DNA Repair by the Cell Cycle. *Nat. Cell Biol.* **2017**, *19* (1), 1–9.
- (81) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J.*



- Mol. Biol.* **2004**, *337* (3), 635–645.
- (82) Zhang, B.; Zheng, W.; Papoian, G. A.; Wolynes, P. G. Exploring the Free Energy Landscape of Nucleosomes. *J. Am. Chem. Soc.* **2016**, *138* (26), 8126–8133.
- (83) Lequieu, J.; Córdoba, A.; Schwartz, D. C.; de Pablo, J. J. Tension-Dependent Free Energies of Nucleosome Unwrapping. *ACS Cent. Sci.* **2016**, *2* (9), 660–666.
- (84) Tan, C.; Terakawa, T.; Takada, S. Dynamic Coupling among Protein Binding, Sliding, and DNA Bending Revealed by Molecular Dynamics. *J. Am. Chem. Soc.* **2016**, *138* (27), 8512–8522.
- (85) Collepardo-Guevara, R.; Portella, G.; Vendruscolo, M.; Frenkel, D.; Schlick, T.; Orozco, M. Chromatin Unfolding by Epigenetic Modifications Explained by Dramatic Impairment of Internucleosome Interactions: A Multiscale Computational Study. *J. Am. Chem. Soc.* **2015**, *137* (32), 10205–10215.
- (86) Maffeo, C.; Ngo, T. T. M.; Ha, T.; Aksimentiev, A. A Coarse-Grained Model of Unstructured Single-Stranded DNA Derived from Atomistic Simulation and Single-Molecule Experiment. *J. Chem. Theory Comput.* **2014**, *10* (8), 2891–2896.
- (87) Rzepiela, A. J.; Schäfer, L. V.; Goga, N.; Risselada, H. J.; De Vries, A. H.; Marrink, S. J. Reconstruction of atomistic details from coarse-grained structures. *J. Comput. Chem.* **2010**, *31* (6), 1333–1343.
- (88) Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *J. Chem. Theory Comput.* **2014**, *10* (2), 676–690.
- (89) Heath, A. P.; Kaviraki, L. E.; Clementi, C. From coarse-grain to all-atom: Toward multiscale analysis of protein landscapes. *Proteins: Struct., Funct., Bioinf.* **2007**, *68* (3), 646–661.
- (90) Lombardi, L. E.; Martí, M. A.; Capece, L. CG2AA: Backmapping Protein Coarse-Grained Structures. *Bioinformatics* **2016**, *32* (8), 1235–1237.
- (91) Brocos, P.; Mendoza-Espinosa, P.; Castillo, R.; Mas-Oliva, J.; Piñeiro, Á. Multiscale Molecular Dynamics Simulations of Micelles: Coarse-Grain for Self-Assembly and Atomic Resolution for Finer Details. *Soft Matter* **2012**, *8* (34), 9005–9014.
- (92) Stansfeld, P. J.; Sansom, M. S. P. From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations. *J. Chem. Theory Comput.* **2011**, *7* (4), 1157–1166.
- (93) Samiotakis, A.; Homouz, D.; Cheung, M. S. Multiscale Investigation of Chemical Interference in Proteins. *J. Chem. Phys.* **2010**, *132* (17), 175101.
- (94) Villa, A.; Peter, C.; van der Vegt, N. F. A. Self-Assembling Dipeptides: Conformational Sampling in Solvent-Free Coarse-Grained Simulation. *Phys. Chem. Chem. Phys.* **2009**,

- 11 (12), 2077–2086.
- (95) Thorpe, I. F.; Zhou, J.; Voth, G. A. Peptide Folding Using Multiscale Coarse-Grained Models. *J. Phys. Chem. B* **2008**, *112* (41), 13079–13090.
- (96) Carpenter, T.; Bond, P. J.; Khalid, S.; Sansom, M. S. P. Self-Assembly of a Simple Membrane Protein: Coarse-Grained Molecular Dynamics Simulations of the Influenza M2 Channel. *Biophys. J.* **2008**, *95* (8), 3790–3801.
- (97) Liu, P.; Shi, Q.; Lyman, E.; Voth, G. A. Reconstructing Atomistic Detail for Coarse-Grained Models with Resolution Exchange. *J. Chem. Phys.* **2008**, *129*, 114103.
- (98) Rotkiewicz, P.; Skolnick, J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.* **2008**, *29* (9), 1460–1465.
- (99) Shih, A. Y.; Freddolino, P. L.; Sligar, S. G.; Schulten, K. Disassembly of Nanodiscs with Cholate. *Nano Lett.* **2007**, *7* (6), 1692–1696.
- (100) Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. A Coarse Grained Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics. *J. Chem. Theory Comput.* **2010**, *6* (5), 1711–1725.
- (101) Gopal, S. M.; Mukherjee, S.; Cheng, Y.-M.; Feig, M. PRIMO/PRIMONA: A coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *Proteins: Struct., Funct., Bioinf.* **2010**, *78* (5), 1266–1281.
- (102) Cragolini, T.; Derreumaux, P.; Pasquali, S. Coarse-Grained Simulations of RNA and DNA Duplexes. *J. Phys. Chem. B* **2013**, *117* (27), 8047–8060.
- (103) Naômé, A.; Laaksonen, A.; Vercauteren, D. P. A Coarse-Grained Simulation Study of the Structures, Energetics, and Dynamics of Linear and Circular DNA with Its Ions. *J. Chem. Theory Comput.* **2015**, *11* (6), 2813–2826.
- (104) Moore, B. L.; Kelley, L. A.; Barber, J.; Murray, J. W.; MacDonald, J. T. High-Quality Protein Backbone Reconstruction from Alpha Carbons Using Gaussian Mixture Models. *J. Comput. Chem.* **2013**, *34* (22), 1881–1889.
- (105) Ryu, J.; Lee, M.; Cha, J.; Laskowski, R. A.; Ryu, S. E.; Kim, D.-S. BetaSCPWeb: Side-Chain Prediction for Protein Structures Using Voronoi Diagrams and Geometry Prioritization. *Nucleic Acids Res.* **2016**, *44* (W1), W416–W423.
- (106) Gaillard, T.; Panel, N.; Simonson, T. Protein Side Chain Conformation Predictions with an MMGBSA Energy Function. *Proteins: Struct., Funct., Bioinf.* **2016**, *84* (6), 803–819.
- (107) Subramaniam, S.; Senes, A. Backbone Dependency Further Improves Side Chain Prediction Efficiency in the Energy-Based Conformer Library (bECL). *Proteins: Struct., Funct., Bioinf.* **2014**, *82* (11), 3177–3187.
- (108) Lu, M.; Dousis, A. D.; Ma, J. OPUS-Rota: A Fast and Accurate Method for Side-Chain Modeling. *Protein Sci.* **2008**, *17* (9), 1576–1585.

- (109) Peterson, R. W.; Dutton, P. L.; Wand, A. J. Improved Side-Chain Prediction Accuracy Using an Ab Initio Potential Energy Function and a Very Large Rotamer Library. *Protein Sci.* **2004**, *13* (3), 735–751.
- (110) Liang, S.; Grishin, N. V. Side-Chain Modeling with an Optimized Scoring Function. *Protein Sci.* **2002**, *11* (2), 322–331.
- (111) Xiang, Z.; Honig, B. Extending the Accuracy Limits of Prediction for Side-Chain Conformations. *J. Mol. Biol.* **2001**, *311* (2), 421–430.
- (112) Jain, T.; Cerutti, D. S.; McCammon, J. A. Configurational-Bias Sampling Technique for Predicting Side-Chain Conformations in Proteins. *Protein Sci.* **2006**, *15* (9), 2029–2039.
- (113) Wang, C.; Schueler-Furman, O.; Baker, D. Improved side-chain modeling for protein–protein docking. *Protein Sci.* **2005**, *14* (5), 1328–1339.
- (114) Milik, M.; Kolinski, A.; Skolnick, J. Algorithm for Rapid Reconstruction of Protein Backbone from Alpha Carbon Coordinates. *J. Comput. Chem.* **1997**, *18* (1), 80–85.
- (115) Gront, D.; Kmiecik, S.; Kolinski, A. Backbone building from quadrilaterals: A fast and accurate algorithm for protein backbone reconstruction from alpha carbon coordinates. *J. Comput. Chem.* **2007**, *28* (9), 1593–1597.
- (116) Maupetit, J.; Gautier, R.; Tufféry, P. SABBAC: Online Structural Alphabet-Based Protein Backbone Reconstruction from Alpha-Carbon Trace. *Nucleic Acids Res.* **2006**, *34* (SI), W147–W151.
- (117) Li, Y.; Zhang, Y. REMO: A New Protocol to Refine Full Atomic Protein Models from C-Alpha Traces by Optimizing Hydrogen-Bonding Networks. *Proteins: Struct., Funct., Bioinf.* **2009**, *76* (3), 665–676.
- (118) Machado, M. R.; Pantano, S. SIRAH Tools: Mapping, Backmapping and Visualization of Coarse-Grained Models. *Bioinformatics* **2016**, *32* (10), 1568–1570.
- (119) Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; de Pablo, J. J. An Experimentally-Informed Coarse-Grained 3-Site-per-Nucleotide Model of DNA: Structure, Thermodynamics, and Dynamics of Hybridization. *J. Chem. Phys.* **2013**, *139* (14), 144903.
- (120) Freeman, G. S.; Hinckley, D. M.; Lequieu, J. P.; Whitmer, J. K.; de Pablo, J. J. Coarse-Grained Modeling of DNA Curvature. *J. Chem. Phys.* **2014**, *141* (16), 165103.
- (121) Lu, X. -J.; Olson, W. K. 3DNA: A Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic Acids Res.* **2003**, *31* (17), 5108–5121.
- (122) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; et al. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2016**, *13*, 55–58.

- (123) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **2010**, *6* (2), 459–466.
- (124) Foloppe, N.; MacKerell, A. D., Jr. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21* (2), 86–104.
- (125) MacKerell, A. D., Jr.; Banavali, N. K. All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution. *J. Comput. Chem.* **2000**, *21* (2), 105–120.
- (126) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25* (11), 1400–1415.
- (127) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616.
- (128) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.
- (129) Blanchet, C.; Pasi, M.; Zakrzewska, K.; Lavery, R. Curves+ Web Server for Analyzing and Visualizing the Helical, Backbone and Groove Parameters of Nucleic Acid Structures. *Nucleic Acids Res.* **2011**, *39*, W68–W73.
- (130) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational Analysis of Nucleic Acids Revisited: Curves+. *Nucleic Acids Res.* **2009**, *37* (17), 5917–5929.
- (131) Nikolov, D. B.; Chen, H.; Halay, E. D.; Hoffman, A.; Roeder, R. G.; Burley, S. K. Crystal Structure of a Human TATA Box-Binding protein/TATA Element Complex. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (10), 4862–4867.
- (132) Chua, E. Y. D.; Vasudevan, D.; Davey, G. E.; Wu, B.; Davey, C. A. The Mechanics behind DNA Sequence-Dependent Properties of the Nucleosome. *Nucleic Acids Res.* **2012**, *40* (13), 6338–6352.
- (133) Plumridge, A.; Meisburger, S. P.; Pollack, L. Visualizing Single-Stranded Nucleic Acids in Solution. *Nucleic Acids Res.* **2017**, *45* (9), e66.
- (134) Plumridge, A.; Meisburger, S. P.; Andresen, K.; Pollack, L. The Impact of Base Stacking on the Conformations and Electrostatics of Single-Stranded DNA. *Nucleic Acids Res.*

- 2017, 45 (7), 3932–3943.
- (135) Craig, N. L.; Nash, H. A. *E. coli* Integration Host Factor Binds to Specific Sites in DNA. *Cell* **1984**, 39 (3), 707–716.
- (136) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph. Model.* **1996**, 14 (1), 33–38.
- (137) Inkscape. <https://inkscape.org/>
- (138) Williams, T.; Kelley, C. et al. Gnuplot 4.6: An Interactive Plotting Program.
- (139) Arias-Palomo, E.; O’Shea, V. L.; Hood, I. V; Berger, J. M. The Bacterial DnaC Helicase Loader Is a DnaB Ring Breaker. *Cell* **2013**, 153 (2), 438–448.
- (140) Das, U.; Wang, L. K.; Smith, P.; Jacewicz, A.; Shuman, S. Structures of Bacterial Polynucleotide Kinase in a Michaelis Complex with GTP•Mg<sup>2+</sup> and 5'-OH Oligonucleotide and a Product Complex with GDP•Mg<sup>2+</sup> and 5'-PO<sub>4</sub> Oligonucleotide Reveal a Mechanism of General Acid-Base Catalysis and the Determinants of Phosphoacceptor Recognition. *Nucleic Acids Res.* **2014**, 42 (2), 1152–1161.