

研究課題別評価

1 研究課題名:

特異的な DNA 配列に結合する蛋白質の設計システム開発

2 研究者氏名:

河野秀俊

3 研究のねらい:

本研究は、計算機シミュレーションにもとづき、新規ナノ機能分子、DNA配列特異的に結合する蛋白質の設計システムを構築することを目的とする。DNA配列に対する特異性を持った蛋白質ができれば、DNAの切断機能など異なる分子機能を持つ蛋白質を融合することによって、さまざまな分子機能を特定のDNA配列に作用させることができる。DNA結合蛋白質とDNA切断機能を融合した蛋白質は、すでに培養細胞において欠陥遺伝子を正常遺伝子に置き換えることに成功したという報告もある。また、DNA結合蛋白質自体が特定のDNA配列に結合することによって、各遺伝子のオン・オフ状態(転写)を制御することができる。本研究では、このDNA結合蛋白質がどのようにDNAと配列特異的に結合しているかを天然蛋白質に学び、それにもとづいて特定のDNA配列に対して結合する蛋白質の創製を目指す。

4 研究成果:

(1) 蛋白質-DNA 認識における直接認識の定量化

転写因子などDNA結合蛋白質は、配列特異的にDNAに結合し、遺伝子の発現を制御している。その認識機構を解明するために、500以上の蛋白質とDNAの複合体の立体構造が決定されているが、個々の複合体構造を眺めただけでは特異性を定量的に評価することができない。また、どのようなDNA配列を蛋白質が認識するか調べるためには、蛋白質ひとつひとつに対して膨大な量の結合実験を行なう必要がある。そこで、構造バイオインフォマティクスのアプローチにより、蛋白質-DNA複合体の立体構造の特徴を解析し、どのDNA結合蛋白質にも共通に使えるアミノ酸残基-塩基対(20x4通り)のポテンシャル関数(直接認識ポテンシャル)を構築した。このポテンシャルを用いて、蛋白質-DNAの複合体を評価すると、図1に示すように特異的結合の複合体(図1の左)では好まれるアミノ酸残基と塩基のペア、つまり、エネルギーの低い赤い部分の領域が複数存在するのに対し、非特異的結合の複合体(図1の右)では大部分がエネルギーの高い青い面になっている。つまり、DNAの配列によって直接認識エネルギーが異なる。特異性の定量化は、複合体構造にランダムDNA配列を載せたときのエネルギー値の分布に対して認識配列を載せたときのエネルギー値がどの程度ずれるか(Z-score)で評価した。

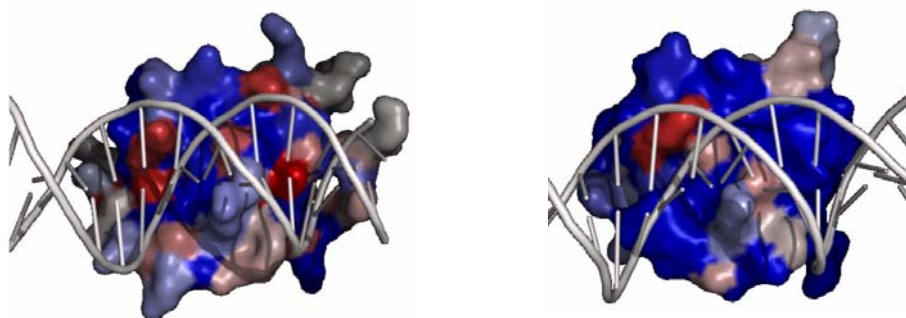


図 1. 蛋白質-DNAの直接認識エネルギー。蛋白質の表面にエネルギーをマップ。低エネルギーが赤、高エネルギーが青。左: 特異的結合。右: 非特異的結合。

(2) 蛋白質-DNA 認識における間接認識の定量化

水素結合など蛋白質と直接的な相互作用をしていない塩基対を変えることによって蛋白質とDNAの結合親和性が大きく変化することが数多く報告されている。これは、蛋白質-DNA認識

において、アミノ酸残基と塩基との直接的な相互作用(直接認識)のみならず、DNA自体の構造と柔軟性(間接認識)の効果が重要であることを意味する。この効果を定量的に評価するために、DNAの構造変形のしやすさに配列依存性があるか調べた。これまでの研究により、DNAの立体構造データやゲル中のDNAの流速測定からAT配列に代表されるプリン-ピリミジンと続く配列は硬く、TA配列のようなピリミジン-プリンと続く配列は柔らかいことが知られている。しかし、このようなダイマー配列(10配列パターン)の特徴でDNAとの結合親和性の変化を説明しきれない。そこで、より長い配列のテトラマー配列(136配列パターン)の構造特徴を調べた。既知のDNAの立体構造ではこのテトラマー配列パターンを調べるだけの十分な統計量がないため、分子動力学計算により各テトラマー配列の構造案サンプルを得た。塩基対を剛体として扱うことにより、DNAを6つの構造パラメータで表現した。

まず、X線結晶構造解析によって得られた立体構造既知のDNAとMDの結果を比較した。構造既知のDNAの数が少ないため、テトラマー配列をその真ん中のダイマー配列にまとめて比較を行った。結晶構造とMDの結果は、平均構造、構造変形のしやすさ(S_{xy} =パラメータの分布の大きさ、つまり、多次元分布の体積で評価)はともにより相関を示し、特に、構造変形のしやすさは高い相関を示した(相関係数 0.9)(図2)。また、これまでの研究によって明らかにされているように、ATなどプリン-ピリミジン配列は最も硬く、AAなどのプリン-プリン配列がその次に硬く、TAなどのピリミジン-プリン配列は最も柔らかいという結果が再現されていた。しかし、テトラマー配列では、真ん中の塩基対の前後の配列の影響を強く受け、柔軟性は著しく変化した(図3)。特に、ピリミジン-プリン配列では、前後の配列によってプリン-ピリミジンと同等の硬さを持つ配列から非常に柔らかい配列まであり、単にダイマー配列だけで硬い、柔らかいを判断できないことがわかった。6つのパラメータの値は多次元ガウス分布に近い分布を示したので、調和関数型のポテンシャルを各テトラマー配列に対して構築した。このポテンシャルを用い、遺伝子の転写制御に重要なヌクレオソーム構造をとりやすいDNA配列とそうでない配列を明確に区別することができることを示した。これは、DNAの構造特性に遺伝子制御の情報が刷り込まれていることを意味する。以上の結果は、論文1, 5で発表した。また、複合体の立体構造から直接認識、間接認識の特異性を評価するweb server を公開(論文3)し、蛋白質のDNA認識についての総説(論文4)をまとめた。

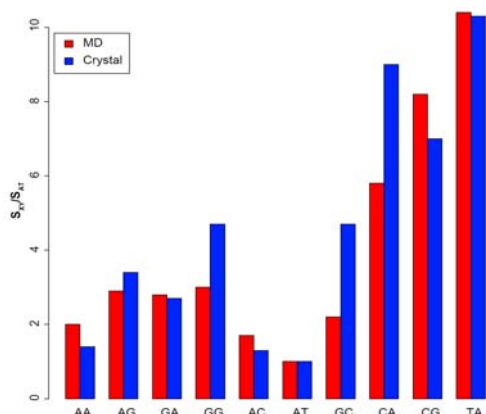


図2. DNAの構造変形のしやすさ(S_{xy})とDNA配列。テトラマーを真中のダイマーごとにまとめた。結晶構造とMDの結果がよく一致している。

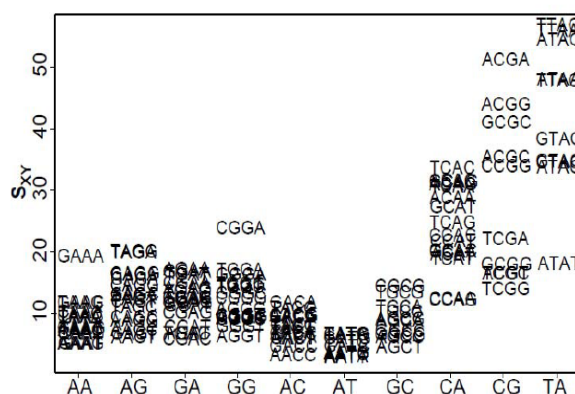


図3. テトラマー配列の構造変形のしやすさ。真ん中の配列が同じであっても両側の配列が変わることによって、配列の構造変形のしやすさが大きく変わる。

(3) 新規亜鉛結合フィンガー型蛋白質の設計と実験による検証

(1) および(2)で開発したポテンシャルをもとに、亜鉛結合フィンガー型蛋白質のひとつ Zif268 を設計のテンプレート構造(図4)としてアミノ酸配列の設計を行った。亜鉛結合フィンガー型蛋白質は、ひとつのフィンガーが3塩基対を認識する。それをn個つなぐことにより、3n長のDNA配列を特異的に認識できると期待されること、ヒトなど真核生物の転写因子に最も多く



図4. 亜鉛結合フィンガー型構造。

見られる構造であることから実験的によく研究されている。

望むDNA配列を認識する蛋白質の設計に成功するには、次の2つの条件をクリアしなければならない。ひとつは、想定どおりの立体構造をとること、もうひとつは望む配列を選択的に認識することである。一つ目の条件はさきがけ研究課題として採択される前に開発した、与えられた立体構造に矛盾しないアミノ酸配列を計算する方法を用い、二つ目の条件は直接、間接認識ポテンシャルを用いて配列選択性を設計した。設計の正しさを検証するために、蛋白質を合成もしくは大腸菌で発現させ、その構造とDNA結合能を調べた。実験による検証は、同志社女子大学、根木博士、京大科研の今西助教の協力を得た。

a) テンプレート構造をとるアミノ酸配列およびDNA配列選択性を持つアミノ酸配列の設計

まず、ポテンシャルの精度を知るため、Zif268蛋白質のさまざまなDNA配列に対する結合自由エネルギーの実験値と直接認識エネルギーを比較した。図5に示すように、両者は高い相関を示し、直接認識エネルギーが結合親和性の指標になることがわかった。次に、さきがけ研究開始以前に開発した方法を用いて、テンプレート構造に矛盾しないアミノ酸配列群を計算した。その結果をもとに、DNAを認識するヘリックス部分のアミノ酸配列以外は計算した構造に最適なアミノ酸配列にし、DNAを認識するヘリックス部分の7つのアミノ酸配列を設計した。設計は、全配列パターン $20^7 (= 1.3 \times 10^9)$ の中からターゲットDNA配列を選択的に認識する配列を探しだすことで行った。ターゲット配列は、野生型と同じTGG、これまで認識に成功したという報告がないTAG、TCG、TTGとした。蛋白質を配列選択的に結合させるため、認識させたいDNA配列とそれ以外の配列のエネルギー差が大きくなるようなアミノ酸配列を計算した(図6)。つまり、 1.3×10^9 のアミノ酸配列に対して、64通りの3塩基対に対する親和性を計算し、ターゲット配列とそれ以外の配列に対するエネルギー差が大きいアミノ酸配列を探した。その結果、図7に示すように、TGGをターゲットにした場合、TGGとの親和性が最も高い(直接認識エネルギーが最も低い)、かつ、次に親和性の高いDNA配列との間に明確なエネルギー差のあるアミノ酸配列が見つかった。その他の配列に対して、明確なエネルギー差をもったアミノ酸配列は見つからなかった。

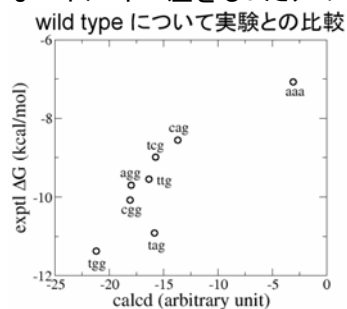


図5. 結合自由エネルギーと直接認識エネルギーの関係。

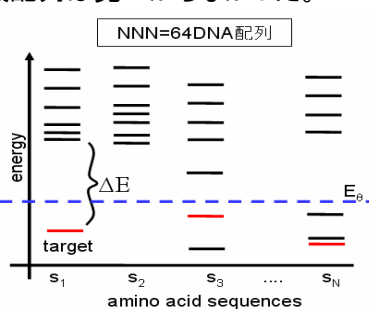


図6. アミノ酸配列設計の概念図

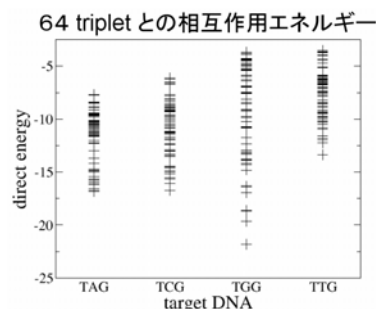


図7. ターゲットDNAとの相互作用エネルギーを最小とするアミノ酸配列の64のトリマーDNA配列に対する相互作用エネルギーの分布。

b) 実験による構造、結合能の検証。

a)で設計した4つの蛋白質の立体構造(2次構造形成)を調べるため、円二色性スペクトルを測定した。スペクトル曲線はどれも野生型によく類似しており、設計したアミノ酸配列は野生型と同様な立体構造をとることが示唆された。ゲルシフトアッセイによるターゲットに対する結合実験では、4つのアミノ酸配列とも数nMから1000 nMの解離定数でターゲットDNAに結合をした。特に、TGGに対しては、設計した蛋白質は約5倍親和性が落ちるものの、野生型より2から6倍高い配列選択性を示した(図8)。残りの3つの蛋白質の

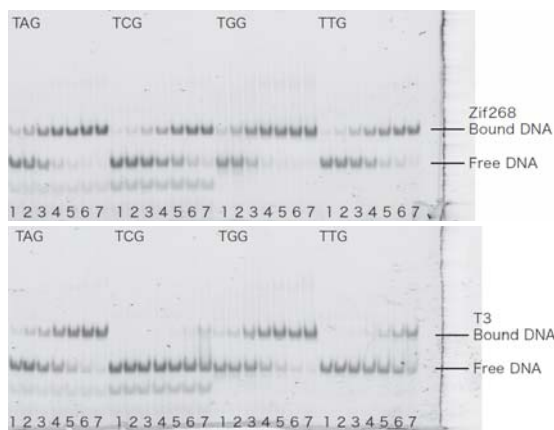


図8. ゲルシフトによる結合実験結果。上段: wild type. 下段: 設計した蛋白質。TGGへの選択性が向上。

配列選択性は低く、TNNパターンに同程度の親和性で結合した。この結果は、設計において最も親和性の高い配列と次に高い配列にエネルギー差がなかった計算結果と矛盾しない。さらに、TGG結合蛋白質の配列親和性の序列を計算と実験で比較してみた。計算では、TGG > TTG > TAG > TCG > CAG > AAA の順であるのに対し、実験では TGG > TAG > TTG > TCG > CAG > AAA であり、TTGとTAGの順序の反転を除いて両者は一致した(論文投稿準備中)。

設計したTGG結合蛋白質の配列選択性が野生型に対して向上した要因は、2つ考えられる。ひとつはDNA認識ヘリックスのアミノ酸配列を変えたこと、もうひとつはフィンガー全体のアミノ酸配列を変えたことである。どちらの要因が効いているか明らかにするために、野生型に設計した蛋白質のヘリックスを入れたもの、設計した蛋白質に野生型のヘリックスを入れたものを作成し、結合実験を行った。結果、どちらの場合も選択性が少し上がった。つまり、2つの要因が相まって選択性を上げていることが示唆された。この要因を立体構造の見地から明らかにするために、設計した蛋白質とDNAの複合体の立体構造を決定すべく、結晶化を試みている(東大、西山教授との共同研究)。同時に、モデル構造による分子動力学シミュレーション計算を行い、蛋白質-DNAのダイナミクスの見地から解析を進めている。

(4) DNA の構造変形のしやすさと DNA の水和の関係

(2)にまとめたように、136 のテトラマー配列をもつ DNAの構造変形のしやすさを分子動力学計算によって調べ、その配列依存性を明らかにした。では、その配列依存性はどこから生じてくるのであろうか。X線結晶構造解析によって、AATT配列はDNAの副溝に水のスパイン構造をもつことが知られている。そこで、副溝における水の水和率(水分子を介した異なる2本鎖間の塩基の水素結合ブリッジ形成率)で評価した。安定な水和構造には、塩基間に水1分子が介在するもの、水2分子が介在するものが観察され、図9に示すようにこのふたつのブリッジの形成率と構造変形のしやすさに強い相関があることがわかった(論文投稿準備中)。また、副溝の幅はブリッジ形成率が高い(すなわち、硬い)配列は小さく、水和率が低い(すなわち、柔らかい)配列は大きいことがわかった(論文2)。この相関からは、ブリッジを形成したからDNAが変形しにくいのか、それとも、DNAが変形しにくいから安定にブリッジを形成するのかわからない。それを明らかにするために、ブリッジ形成率の最も高いAATT配列に対して、仮想的に水がブリッジしている塩基の原子の電荷をゼロにし水素結合を形成できない条件で分子動力学計算を行い、ブリッジ形成率と構造変形のしやすさを調べた。結果、水のブリッジが形成されないにも関わらず、DNAは硬い構造(S_{xy} が小さい値)をとった。これは、DNAの硬さは水和によってもたらされた結果ではなく、DNAの揺らぎが小さいことが結果的に水のブリッジ形成をうながしていると解釈される。すなわち、DNA構造の変形のしやすさは、おもに塩基対のスタッキングエネルギーによって決まることを示唆している。

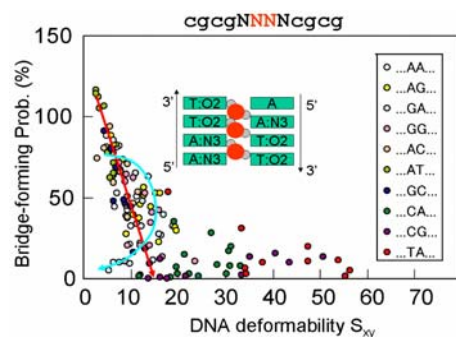


図9. 9ns 間における副溝での水のブリッジ形成率と DNA の構造変形能との関係。4 塩基対配列を真中の 2 塩基対配列ごとに色をつけた。同時に複数の水がブリッジすることがあるので 100%を超える。

5 今後の展開

(1) 今後の研究の展開

まず、野生型に比べて配列選択性の向上した蛋白質と DNA の複合体の立体構造を決定し、その選択性向上の要因を明らかにする。次に、蛋白質設計を亜鉛結合型蛋白質ではない他の蛋白質の構造をテンプレート構造に拡大し、これまで認識することの実現されていない DNA 配列に対する配列選択性と親和性の高い蛋白質を引き続き設計する。また、制限酵素や蛍光蛋白質を DNA 結合蛋白質に連結し、ゲノム中の狙った位置で機能するナノ機能分子としての蛋白質を作成する。

(2) 他の研究事業への展開

特になし。

(3) 実用化に向けた展開
特になし。

6 領域内外での活動とその効果

(1) 領域内の活動とその効果

領域会議は、毎回さまざまな刺激を受け研究遂行の励みとなった。ただ、2年目の時点で領域内の研究者がシミュレーション研究の5名と継続の1名の研究者のみになり、先輩研究者から大きな刺激や研究の方向性に関するアイデアを得ていただけに、少人数になったのは残念であった。

(2) 領域横断的活動とその効果

さきがけによる研究支援のおかげで、別領域のさきがけ研究で亜鉛結合蛋白質を用いて研究を展開している京大化研、今西助教や同志社女子の根木助教と共同研究を実施することができ、理論計算の実験的検証を行うことができた。

7 研究成果の今後の貢献について

本研究で目指した、任意の DNA 配列に対して配列選択性と親和性の高い蛋白質の創製が容易になれば、DNA 切断蛋白質や蛍光蛋白質を DNA 結合蛋白質に連結することによってさまざまな分子機能をもった蛋白質をゲノム中の狙った位置で働かせることができる。例えば、DNA を切断する蛋白質を連結することにより、特定のゲノム配列の位置に正常遺伝子の導入を行ったり異常遺伝子を除去したりできるようになると期待される。このようなツール蛋白質群の創製は、生命科学、医学、工学といった幅広い分野に資することができると思う。

8 自己評価：

蛋白質の設計は、 $20^{(\text{アミノ酸の数})}$ 乗に及ぶ実験では到達不可能な数のアミノ酸配列群の中から想定した立体構造に適合し、かつ、DNAに結合するアミノ酸配列を効率よく見つけ出す問題と捉え、その膨大な配列を瞬時かつ的確に評価できる簡単なポテンシャルを開発することからスタートした。構造バイオインフォマティクス的なアプローチにより、塩基に対するアミノ酸残基の空間分布にもとづく単純なポテンシャルで、DNA結合蛋白質のターゲットDNA配列を推定することができた。また、DNAの構造変形能にインプリントされた蛋白質-DNA認識における配列選択性への寄与を定量的に評価することができた。さらに、テトラマー配列の系統的な構造特性を世界に先駆けて発表することができた。

当初に目論んでいた、これまで認識することが実現されていない DNA 配列を選択的に認識する新規蛋白質を設計することはできなかった。しかし、計算シミュレーションにもとづき、高々5回の試行により安定な構造をとる蛋白質を創製できること、しかも DNA に結合する蛋白質を創製できることを示すことができた。実験とシミュレーションのサイクルをもっと回す予定であったが、実験結果の解釈に手間取ったり試料の DNA 合成、ペプチド合成の発注、納品に時間がかかり、5サイクルしか回すことができなかった。シミュレーションが単なる計算結果に終わらず、実験研究者と組むことでその結果を検証することができたのは、さきがけの支援のおかげであり、非常に感謝している。今後、この共同研究を発展させていきたい。

9 研究総括の見解：

天然蛋白質では認識しない塩基配列を選択的に認識する新規蛋白質を設計するという、最も挑戦的な目標はまだ達成されていないが、天然型よりもかなり優れた配列選択性を示す TGG 結合蛋白質の設計に成功したことは評価できる。蛋白質設計に関わる複雑で広い範囲の問題に正面から取り組み、実験と計算の間のフィードバックによって両者を結びつけようとした努力は評価できるが、今後、計算手法の完成度をより高めるために、近年の電子状態計算に基づく手法との比較検討が望まれる。

10 主な論文等

(1) 論文(原著論文)発表

- 1) Fujii, S., Kono, H., Takenaka, S., Go, N. & Sarai, A. in press. Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic Acids Res.*
 - 2) Yonetani, Y., Kono, H., Fujii, S., Sarai, A. & Go, N. (2007). DNA deformability and hydration studied by molecular dynamics simulation. *Molecular Simulation* **33**, 103–107.
 - 3) Ahmad, S., Kono, H., Arauzo-Bravo, M. J. & Sarai, A. (2006). ReadOut: Structure-based Calculation of Direct and Indirect Readout Energies and Specificities for Protein-DNA Recognition. *Nucleic Acids Res.* **34**, w124–127.
 - 4) Sarai, A. & Kono, H. (2005). Protein-DNA Recognition Patterns and Predictions. *Annu Rev Biophys Biomol Struct.* **34**, 379–398.
 - 5) Arauzo-Bravo, M., Fujii, S., Kono, H., Ahmad, S. & Sarai, A. (2005). Sequence-Dependent Conformational Energy of DNA Derived from Molecular Dynamics Simulations: Toward Understanding the Indirect Readout Mechanism in Protein-DNA Recognition. *J. Am. Chem. Soc.* **127**, 16074–16089.
- その他、国際 3 報、国内 2 報

(2) 特許出願

研究期間累積件数: 0 件

(3) その他の成果

招待講演 国際 3 件、国内 3 件、その他学会等での発表 31 件

- 1) Kono, H. Protein-DNA Recognition Studied by Structural Bioinformatics. *Invited Research Seminar*, Univ. of East Anglia, Norwich, UK. (May, 2007)
- 2) Kono, H. & Sarai, A. Cryptic DNA recognition by Transcription Factors. *第44回日本生物物理学会年会*, 沖縄. (Nov. 2006).
- 3) Kono, H. Structural Bioinformatics Approach for Understanding Protein-DNA Recognition Mechanism. *2nd International Symposium: Molecular Control of Gene Expression*, Erlangen, Germany. (June 2006)
- 4) Kono, H., Solvic, A. M., Lear, J. D., Saven, J. G. & DeGrado, W. F. Computational design of water-soluble analogues of the potassium channel KcsA. *第42回日本生物物理学会年会*, 京都 (Dec. 2004).
- 5) 河野秀俊. 立体構造からみたDNA結合蛋白質のDNA配列認識機構. *日本バイオインフォマティクス学会 第3回生物情報ネットワーク研究会*, 東京. (Nov. 2004).
- 6) Kono, H. Combinatorial protein design strategies using computational method. *2nd International COE Symposium on "Large-Scale Computing Methods for Materials Chemistry and Bioscience*, Sendai, Japan. (Nov. 2004)