

博士学位論文

蛋白質の二次構造帰属法 ProSSA の応用と
アプリケーションの開発

平成 20 年 3 月

九州工業大学大学院生命体工学研究科

小副川 博也

【目次】

第1章 序論	3
第2章 二次構造帰属法 ProSSA のデータベースの開発	9
2-1 はじめに	9
2-2 主鎖の水素結合様式の連続性パターンとしての DSSP 法とデータベース	10
2-3 主鎖二面体角の連続性パターンとしての ProSSA 法	14
2-4 ProSSA 法のデータベース	20
第3章 α ヘリックス構造の解析と配列デザイン	23
3-1 はじめに	23
3-2 α ヘリックス構造の境界領域の予測	24
3-3 α ヘリックス構造の不安定領域の検索とその手法	29
3-4 種々の蛋白質への適用	32
3-5 α ヘリックス構造に対する変異導入部位の検索プログラムの開発	34
第4章 蛋白質 CD スペクトルからの二次構造含量の評価	38
4-1 はじめに	38
4-2 CD スペクトルと二次構造の関係	40
4-3 CD スペクトルと ProSSA 法による二次構造含量の評価	43
4-4 参照 CD スペクトルの算出	49
4-5 CD スペクトルから二次構造含量を推定するプログラム	55
4-6 推定した二次構造含量の検証	56
4-6-1 CD データのセットからの二次構造含量の推定	56
4-6-2 CD スペクトルからの二次構造含量の推定	59
4-7 蛋白質 CD スペクトルのデータベース	62
第5章 プロテアーゼ消化配列の ProSSA 法による解析	66
5-1 はじめに	66
5-2 プロテアーゼ消化部位と二次構造の相関	67

第6章 β ストランド構造の解析	78
6-1 はじめに	78
6-2 ProSSA 法によって帰属される β ストランド構造の解析	80
6-3 シート形成 β ストランド構造の解析	87
6-4 一本鎖 β ストランド構造の解析	97
6-5 主鎖二面体角による構造安定化エネルギーの算出	106
第7章 ジペプチドを単位パラメータとする二次構造の統計的解析	109
7-1 蛋白質解析の際の基本単位の意味	109
7-2 二次構造にみられるジペプチド単位構造の分布	111
7-3 ジペプチド単位でみる Chou-Fasman パラメータ	112
7-3-1 α ヘリックス構造の統計解析	112
7-3-2 β ストランド構造の統計解析	117
7-3-3 シート形成 β ストランド構造の統計解析	120
7-3-4 一本鎖 β ストランド構造の統計解析	122
7-4 ジペプチド単位でみる Richardson パラメータ	124
7-4-1 α ヘリックス構造の統計解析	124
7-4-2 β ストランド構造の統計解析	149
7-5 ジペプチド単位パラメータの自由エネルギー計算	165
7-5-1 はじめに	165
7-5-2 Ala-Ala の自由エネルギー計算	165
第8章 結論	168
参考文献	172
謝辞	179

第1章 序論

バイオインフォマティクスは、1990年代から、情報科学と生命科学の融合分野で使われ始めた用語である。このバイオインフォマティクスの成果の一つに、ヒトゲノムのすべての配列の読み取りがある¹⁾。ヒトゲノムの配列の読み取りが完了したことにより、蛋白質の研究はポストゲノムシーケンスの視点が加わり、これまで以上の注目を浴びている。蛋白質は、フォールディングや立体構造解明、機能向上や構造安定化、新規蛋白質の作製などさまざまな研究が理論と実験の両方から取り組まれている。これらは、すべて別々の研究ではなく、互いに関連しあっている。その中でも、立体構造の解明は情報科学の発展によって進展し、その成果は立体構造データを蓄積するプロテイン・データバンクの登録数として現れている (Figure 1-1)^{2,3)}。

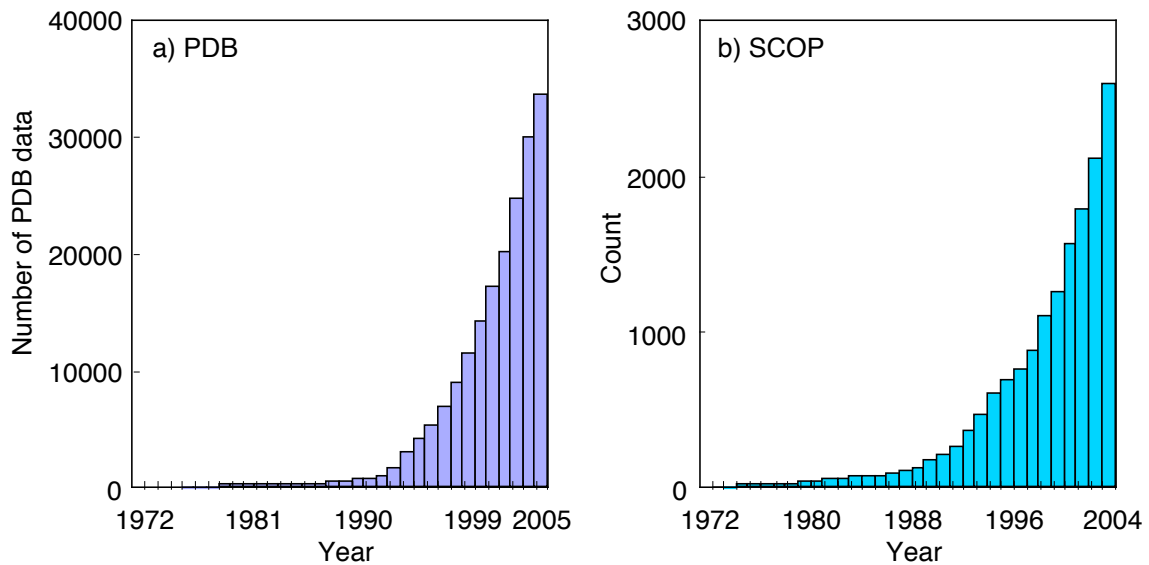


Figure 1-1. 蛋白質データとトポロジカルな分類の推移

蛋白質立体構造データの増加とともに、蛋白質の二次構造の空間的配置の違いによる立体構造の分類、つまりトポロジカルな分類の数も増加した⁴⁸⁾。Pauling と Corey⁹⁾が提唱した蛋白質の二次構造は、蛋白質の階層構造の構造の一つであり、多くの研究者が解析の対象にしてきた。二次構造は、蛋白質の立体構造の予測や設計を行なう場合の基本骨格として重要な役割をもつ

ている。歴史的には、蛋白質のアミノ酸配列の中で、どの領域を二次構造と見なすかについて取り組まれてきた。その中でも、水素結合様式のパターンによる二次構造帰属法 (Definition of Secondary Structure of Protein; DSSP) が普及した¹⁰⁾。しかしながら、DSSP 法は、潜在的な問題を含んでいる。とくに主鎖のコンホメーションが同一でありながら異なる二次構造として分類される点については、二次構造として帰属されるべき範囲を狭めてしまう。金光は、この問題を指摘するとともに解消する方法として、主鎖二面体角の連続性パターンによる二次構造帰属法 Protein Secondary Structure Assignment (ProSSA) を提案した¹¹⁾。ProSSA 法は DSSP 法の帰属結果を包括している。この論文の第 2 章では、ProSSA 法から得られる二次構造帰属の結果をアーカイブしたデータベースを構築した。このデータベースは第 3 章以降で使用した。

ポスト・ゲノムシーケンスの時代を迎えて、新たに蛋白質の研究分野へ参入した研究者・技術者がいる。そして、その多くの人たちが蛋白質を解析・設計するため、蛋白質に関連するアプリケーションを利用している。しかしながら、特に蛋白質の設計についてはアプリケーションから得られた結果を利用者が見て、経験的に設計するポイントを探し出すことが多い。つまり、ほとんどのアプリケーションは、設計の指針を示すというよりも、得られた結果を利用者が設計するため、あるいは設計した後の単なる裏付けのデータとして読み替えているに過ぎない。たとえば、蛋白質のコンホメーションの分布を示すラマチャンドラン・マップは、あくまでも、コンホメーションの分布を示すだけのデータであり、そのデータを予測や設計に用いるのは、利用者の見方しだいである。この論文の第 3 章では、 α ヘリックス構造を予測するだけでなく、設計のための指針も示せるアプリケーションを開発した。つまり、アプリケーションを使うユーザがデータをデザイン用として解釈するのではなく、設計のための指針を示すデータをユーザへ提供することを目指した。

また、二次構造を設計して蛋白質を発現させた後は、立体構造が野生型と同じ構造を形成しているかを確認する必要がある。そのための手段として、よく用いられる方法は、分光学的手法の一つである円偏光二色性 (Circular Dichroism; CD) を使った解析である。CD スペクトルの観測は容易であり時間もかからないため、多くの研究者が利用している。その CD 測定装置によって蛋白質の CD スペクトルを測定するときに使われる波長領域は、180 nm から長波長側の

である。その領域の中でも、遠紫外領域である 190～260 nm はアミドの吸収帯である。つまり、蛋白質の主鎖のコンホメーションを反映したスペクトルが得られる波長領域である。そして、この遠紫外領域のスペクトルは、蛋白質の二次構造の含量を推定する際に使われる波長領域でもある。そして、多くの研究者たちが、さまざまなアルゴリズムを用いて取り組んできた^{12,21)}。遠紫外領域よりも短波長側の真空紫外領域（160～190 nm）も含めた CD スペクトルを使うと、CD スペクトルと二次構造含量の相関が得られるという報告もある^{22,25)}。しかしながら、真空紫外領域を測定するためには専用の装置が必要であり、多くの研究者・技術者は容易に利用できない。しかしながら、現在普及している CD 測定装置を使った遠紫外領域の CD スペクトルと二次構造含量には相関が見られない場合がある。この相関が見られないのはアルゴリズムや波長領域に原因があるのではなく、二次構造帰属法に問題があると考えられる。この論文の第4章では、二次構造帰属法として主鎖二面体角の連続性パターンで二次構造を帰属する ProSSA 法を使って、遠紫外領域の CD スペクトルと二次構造含量の関係を詳細に調べた。さらに、ProSSA 法に基づいて、蛋白質 CD スペクトルから二次構造含量を推定するためのアプリケーションを開発した。

蛋白質の中でも、プロテアーゼはアミノ酸配列の特定の配列を認識し、ペプチド結合を切断して基質蛋白質を不活性型から活性型へ、またはその逆へ変化させる働きがある。この切断されるペプチド結合は、認識配列の中のうち基質蛋白質の表面に存在する。なぜなら、酵素と基質は鍵と鍵穴の関係にあり、蛋白質内部に認識配列が存在したとしても、内部の認識配列を切断するには、いったん基質蛋白質が変性する必要があるからである。また、蛋白質表面以外にも、ループ領域に存在する認識配列、あるいは、 α ヘリックスの内部を切断するなど、切断配列に関する統一的な見解ができていない。とくに、二次構造との関係を示す結果は得られていない。二次構造帰属法 DSSP では、プロテアーゼによって切断される配列と二次構造の関係には、相関を見ることができない。この論文の第5章では、ProSSA 法を使って切断配列と二次構造の関係を調べた。

この論文で用いる ProSSA 法は、前述したように主鎖二面体角の連続性パターンによって二次構造を帰属する方法であるが、現在広く採用されている水素結合様式のパターンで二次構造

を帰属する DSSP 法の帰属結果を包括している (Figure 1-2) ¹¹⁾。

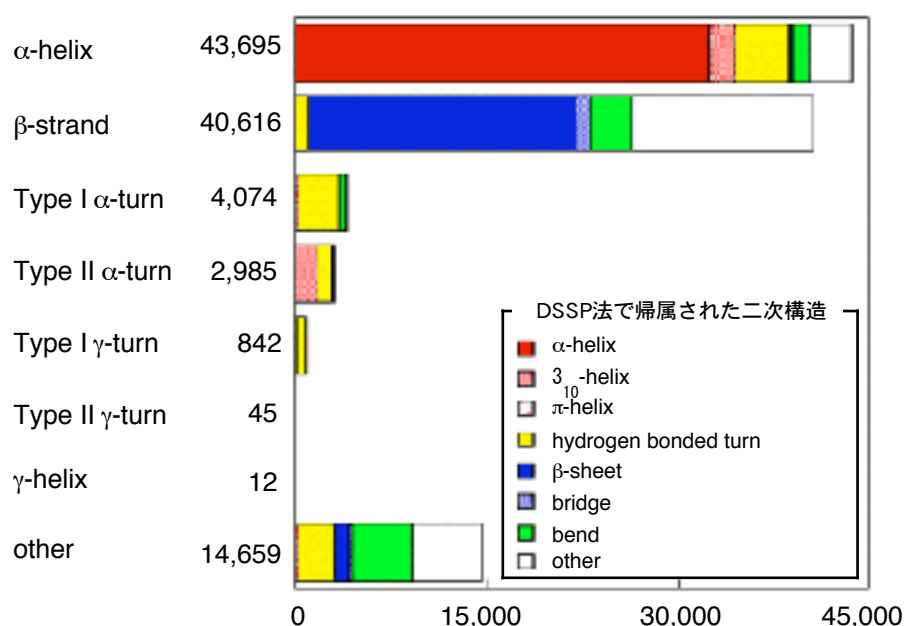


Figure 1-2. ProSSA法による二次構造の分類

ProSSA 法はコンホメーションの規則性で二次構造を定義しているため、 β シート構造の属する β ストランド構造 (シート形成 β ストランド構造) と、 β シート構造と同じコンホメーションであり水素結合が関わっていない伸びた構造 (一本鎖 β ストランド構造) の区別がつかないという問題点がある。逆に言えば、それらの2種類の β ストランド構造を区別することも可能である。そこで、第6章ではシート形成と一本鎖の二つの β ストランド構造を分類し、それらの違いを明らかにした。

蛋白質の二次構造に共通する普遍的な原理・原則を導き出すための方法として、二次構造に含まれるアミノ酸残基について統計的に解析する取り組みがある。この解析方法は蛋白質の構成成分であるアミノ酸残基を基本単位とするのが特徴である ²⁶⁻³⁵⁾。一方、一次配列から二次構造を予測するには配列情報を加える必要がある。この配列情報を得るために、数残基を一組として取り扱う方法や、二次構造の末端領域に対してポジションを決めて末端配列を特徴づけるなどして配列情報の代替としている。しかしながら、アミノ酸残基を基本単位としているため、

あくまでも個々のアミノ酸残基の特徴が積算されたにすぎない。蛋白質の二次構造は、主鎖のアミド間の水素結合が構造形成に大きく関わっている。この主鎖のアミド結合は、一見するとどれも同じであり、特徴づけられないように見えるが、アミドの両側にあるアミノ酸側鎖を含めれば、アミド結合を特徴づけられると考えられる。実際に、野生型と変異型の構造を確認するための手法として、円偏光二色性測定法や赤外分光法がよく用いられるが、これらの分光法はアミド結合の励起状態や振動状態を観測している。すなわち、配列情報に加えてアミド結合の特徴をもつ基本単位としてジペプチドを考えた。ジペプチドは最小の配列情報を有しアミド結合を含むため、二次構造の特徴を抽出できる最小単位と考えられる。第7章では二次構造解析として、アミノ酸残基によらないジペプチドを基本単位とした統計的な解析に取り組んだ。

情報科学の進展にともなって、インターネット利用者数は年々増加している (Figure 1-3) ³⁶⁾。

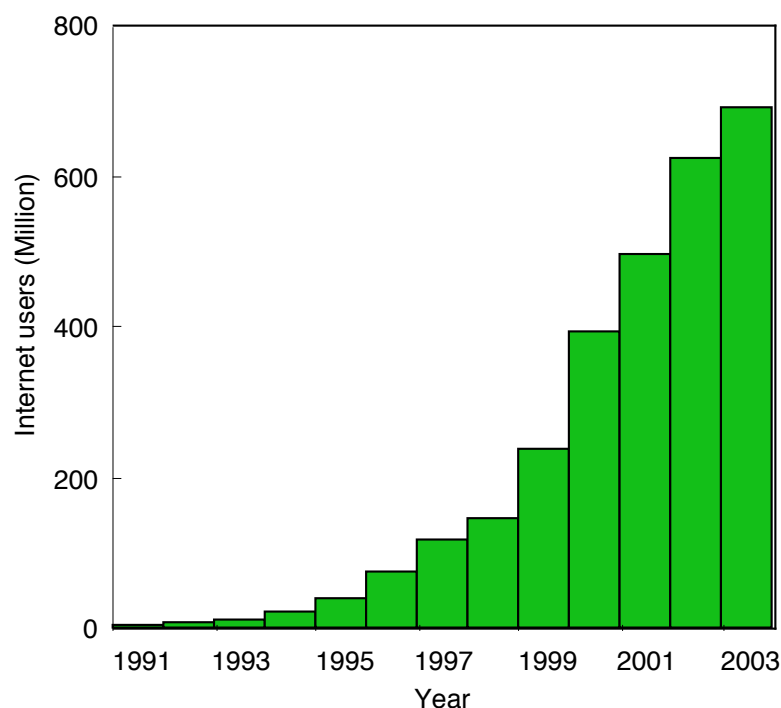


Figure 1-3. 世界のインターネット利用者数の推移

つまり、情報を発信・配信すれば多くの人たちがその情報を利用できる。バイオインフォマティクスの分野では、プロテイン・データバンクをはじめとして、さまざまなサーバが稼働し利

用されている。この論文で使っている二次構造帰属法 ProSSA も多くの研究者や技術者に提供できるように、第2章では ProSSA 法のデータベースを構築し Web 上に公開することを目指した。さらに、第4章では実際に測定した CD データや論文に掲載された CD データの情報をデータベースとしてまとめ、CD 測定の際の参考となることを目指した。

以上、この研究は、蛋白質の二次構造帰属法 ProSSA がどのような研究に応用できるかを調べ、得られた結果をアプリケーションに反映させ提供することを目的としている。

第2章 二次構造帰属法 ProSSA のデータベースの開発

2-1 はじめに

蛋白質の立体構造がX線結晶解析によって明らかにされてから、蛋白質の立体構造データをデータベース化し管理するようになった。そのデータベースがプロテイン・データバンク (PDB) であり、多くの研究者が利用している²³⁾。一時期、蛋白質に関連するデータベースは濫立したものの、現在は統合再編を繰り返し淘汰の時代を迎えている。有用なデータベースが残っていくであろうが、そのうちの一つとして、PDB に登録された蛋白質を二次構造の空間的配置、すなわち、トポロジカルな構造に分類している Structural Classification of Proteins (SCOP) がある⁴⁸⁾。また、蛋白質の構造を考える上で、もっとも基本的な繰り返し構造である二次構造を帰属するために、主鎖の水素結合様式パターンで二次構造を帰属する Definition of Secondary Structure of Proteins (DSSP) が提案され、その帰属ファイルをアーカイブした DSSP データベースが構築されている¹⁰⁾。

遺伝子工学の普及により、変異蛋白質の作製が容易になり変異蛋白質を使った研究が数多く報告されてきた。それにとまって、蛋白質の構造データではなく論文の情報を基にデータベースが構築されている。Protein Mutant Database (PMD) は、変異蛋白質の情報を論文より集めて構築されたデータベースである³⁷⁾。また、Thermodynamic Database for Proteins and Mutants (ProTherm) は、論文の中の熱力学的データに注目したデータベースである^{38,41)}。Table 2-1 には、これらのデータベースのアドレスをまとめている。

Table 2-1. 蛋白質に関するデータベース

Protein Data Bank (PDB)	http://www.rcsb.org/pdb/
Structural Classification of Proteins (SCOP)	http://scop.mrc-lmb.cam.ac.uk/scop/
The Protein Mutant Database (PMD)	http://pmd.ddbj.nig.ac.jp/~pmd/pmd-j.html
Thermodynamic Database for Proteins and Mutants (ProTherm)	http://gibk26.bse.kyutech.ac.jp/jouhou/Protherm/protherm.html
DSSP database	http://swift.cmbi.ru.nl/gv/dssp/

このように蛋白質の立体構造をまとめた多くのデータベースが存在する中で、蛋白質の二次構造のデータベースとしては、DSSP 法のデータベースが広く使われている。しかしながら、DSSP 法には、潜在的な問題が存在する。主鎖のコンホメーションが同一でありながら異なる二次構造として分類される点については、二次構造として帰属されるべき範囲を狭めてしまう。金光は、この問題を指摘するとともに解消する方法として、主鎖二面体角の連続性パターンによる二次構造帰属法 Protein Secondary Structure Assignment (ProSSA) を提案した¹¹⁾。こ ProSSA 法は、DSSP 法の問題点を解決し、なおかつ DSSP 法の帰属結果を包括した手法である。

この章では、二次構造についての研究で広く使われている DSSP のアルゴリズムとファイルフォーマット、そして、ProSSA のアルゴリズムとファイルフォーマットについて述べる。

2—2 主鎖の水素結合様式の連続性パターンとしての DSSP 法とデータベース

Define Secondary Structure of Proteins (DSSP) 法は、蛋白質の二次構造を帰属する方法の中でも、多くの研究者に使われている¹⁰⁾。その理由の一つは、二次構造を帰属したファイルを早くからデータベースとして公開したためである。この DSSP 法は、蛋白質の主鎖のアミド結合の水素とカルボニルの酸素の水素結合様式のパターンで二次構造を帰属している。以下に、まず水素結合について述べ、つぎに α ヘリックスの定義で使われる n-ターン、そして β シートで使われるラダーについて解説する。

(1) 水素結合

DSSP 法では、水素結合を静電相互作用エネルギー E が -0.5 kcal/mol 未満と定義する。記号では、

$$\text{Hbond}(i, j) = [E < -0.5 \text{ kcal/mol}]$$

で表わす。この水素結合を使って α ヘリックスや β シートを定義する。

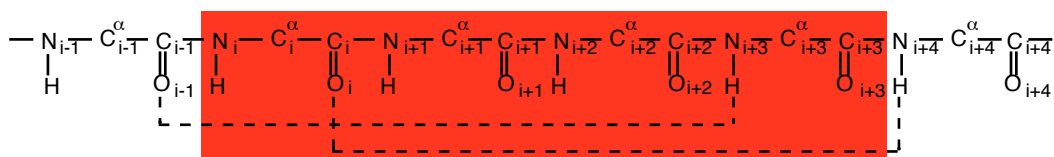
(2) n-ターン

n-ターンは、 i 番目と $i+n$ ($n=3, 4, 5$) 番目のアミノ酸残基同士の水素結合で定義する。記号では、

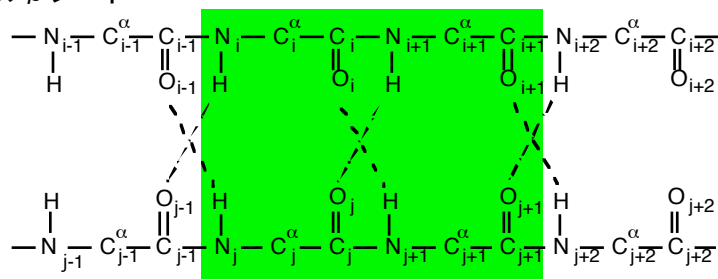
$$n\text{-turn}(i) =: \text{Hbond}(i, i+n), n=3,4,5.$$

で表わす。

(a) α ヘリックス



(b) パラレル β シート



(c) アンチパラレル β シート

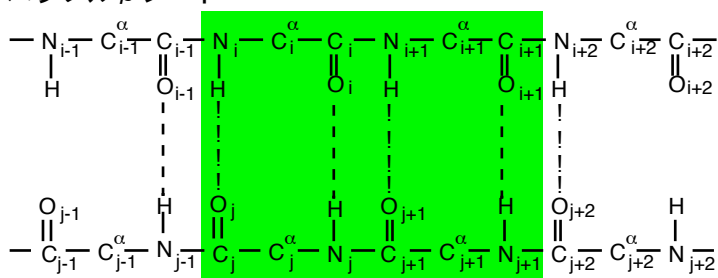


Figure 2-1. 最も短い α ヘリックスと β シート

(3) ブリッジ

ブリッジは、 $i-1, i, i+1$ 番目のアミノ酸残基と $j-1, j, j+1$ 番目のアミノ酸残基の二つの重なり合っていない伸びた構造であり、互いの水素結合の相手の違いにより、パラレルブリッジとアンチパラレルブリッジに分類される。つまり、

$$\begin{aligned} \text{Parallel Bridge}(i, j) &= [\text{Hbond}(i-1, j) \text{ and } \text{Hbond}(j, i+1)] \text{ or} \\ &\quad [\text{Hbond}(j-1, i) \text{ and } \text{Hbond}(i, j+1)], \\ \text{Antiparallel Bridge}(i, j) &= [\text{Hbond}(i, j) \text{ and } \text{Hbond}(j, i)] \text{ or} \\ &\quad [\text{Hbond}(i-1, j+1) \text{ and } \text{Hbond}(j-1, i+1)] \end{aligned}$$

である。

(4) α ヘリックス

α ヘリックスは、4-turn が2回以上連続する構造と定義する。たとえば、i-1 番目と i+3 番目のアミノ酸残基、i 番目と i+4 番目のアミノ酸残基がともに 4-turn ならば、最小の長さ 4 の α ヘリックスである (Figure 2-1 (a))。記号では、

$$4\text{-helix}(i, i+3) = [4\text{-turn}(i-1) \text{ and } 4\text{-turn}(i)]$$

で表わす。

(5) ラダーと β シート

ラダーは、連続したブリッジ構造である。さらに、 β シートは、連続したラダーである。つまり、 β シートは、最小2残基で構成している連続したブリッジ構造である (Figure 2-1 (b), (c))。

DSSP データベースは、このようなアルゴリズムを組み込んだ DSSP プログラムを使って作成された帰属ファイルをアーカイブしている。DSSP のファイルフォーマットは、大きく三つの情報をもっている。つまり、PDB ファイルから抽出した情報、DSSP プログラムが求めた統計情報、二次構造の帰属情報である。ここでは、1DPM.dssp ファイルを例に説明する (Figure 2-2)。

【PDB ファイルから抽出した情報】

3 行目から 6 行目までが PDB ファイルから抽出した情報である。3 行目から「HEADER」、
「COMPND」、
「SOURCE」、
「AUTHOR」の順に並んでいる。HEADER は、PDB の HEADER レコード、つまり、蛋白質を機能で分類したときの名称、PDB に登録したときの日付、PDB コードをそのまま抽出した行である。COMPND は、PDB の COMPND レコードの中でも蛋白質

の名前が書かれてある MOLECULE の行のみを抽出した行である。SOURCE は、PDB の SOURCE レコードの中でも、生物種名が書かれた ORGANISM_SCIENTIFIC の行のみを抽出した行である。AUTHOR は、PDB ファイルの AUTHOR レコードを抽出した行である。

```

1:  Secondary Structure Definition by the program DSSP, updated CMBI version by ElmK / April 1,2000 DATE=9-MAR-2006
2: REFERENCE W. KABSCH AND C.SANDER, BIOPOLYMERS 22 (1983) 2577-2637
3: HEADER HYDROLASE 2 MOLECULE: PHOSPHOTRIESTERASE ; 13-FEB-96 10PM
4: COMPND 2 MOLECULE: PHOSPHOTRIESTERASE ; 13-FEB-96 10PM
5: SOURCE 2 ORGANISM_SCIENTIFIC: PSEUDOMONAS DIMINUTA
6: AUTHOR J.L.VANHOOGHE,M.M.BENNING,F.M.RAUSHEL,H.M.HOLDEN
7: 658 2 0 0 0 TOTAL NUMBER OF RESIDUES, NUMBER OF CHAINS, NUMBER OF SS-BRIDGES (TOTAL,INTRACHAIN,INTERCHAIN )
8: 24001.0 ACCESSIBLE SURFACE OF PROTEIN (ANGSTROM**2)
9: 474 72.0 TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(CI)-->H-NC(I) SAME NUMBER PER 100 RESIDUES
10: 62 9.4 TOTAL NUMBER OF HYDROGEN BONDS IN PARALLEL BRIDGES, SAME NUMBER PER 100 RESIDUES
11: 10 1.5 TOTAL NUMBER OF HYDROGEN BONDS IN ANTIPARALLEL BRIDGES, SAME NUMBER PER 100 RESIDUES
12: 2 0.3 TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(CI)-->H-NC(I+5), SAME NUMBER PER 100 RESIDUES
13: 0 0.0 TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(CI)-->H-NC(I+4), SAME NUMBER PER 100 RESIDUES
14: 2 0.3 TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(CI)-->H-NC(I+3), SAME NUMBER PER 100 RESIDUES
15: 0 0.0 TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(CI)-->H-NC(I+2), SAME NUMBER PER 100 RESIDUES
16: 2 0.3 TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(CI)-->H-NC(I+1), SAME NUMBER PER 100 RESIDUES
17: 0 0.0 TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(CI)-->H-NC(I+0), SAME NUMBER PER 100 RESIDUES
18: 0 0.0 TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(CI)-->H-NC(I+1), SAME NUMBER PER 100 RESIDUES
19: 30 4.6 TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(CI)-->H-NC(I+2), SAME NUMBER PER 100 RESIDUES
20: 82 12.5 TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(CI)-->H-NC(I+3), SAME NUMBER PER 100 RESIDUES
21: 225 34.2 TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(CI)-->H-NC(I+4), SAME NUMBER PER 100 RESIDUES
22: 28 4.3 TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(CI)-->H-NC(I+5), SAME NUMBER PER 100 RESIDUES
23: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 *** HISTOGRAMS OF ***
24: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 RESIDUES PER ALPHA HELIX
25: 2 2 10 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 PARALLEL BRIDGES PER LADDER
26: 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 ANTIPARALLEL BRIDGES PER LADDER
27: 2 2 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 LADDERS PER SHEET
28: # RESIDUE AA STRUCTURE BP1 BP2 ACC N-H-->O O-->H-N N-H-->O O-->H-N TCO KAPPA ALPHA PHI PSI X-CA Y-CA Z-CA
29: 1 35 A A 0 0 103 0, 0.0 11,-1.2 0, 0.0 2,-0.3 0.000 360.0 360.0 360.0 -56.5 29.6 14.3 18.9
30: 2 36 A R E -A 11 0A 160 9,-0.2 2,-0.4 10,-0.1 9,-0.2 -0.917 360.0-142.1-124.8 150.1 28.7 10.7 19.9
31: 3 37 A I E -A 10 0A 4 7,-3.5 7,-2.4 -2,-0.3 2,-0.5 -0.913 23.1-119.3-112.6 137.2 30.7 7.8 21.1
32: 4 38 A N E +A 9 0A 44 -2,-0.4 88,-0.3 5,-0.2 2,-0.2 -0.684 36.7 172.7 -81.0 122.6 30.2 4.2 20.0
33: 5 39 A T E -A 8 0A 5 3,-2.1 3,-1.4 -2,-0.5 125,-0.4 -0.720 49.0-102.7-119.9 169.9 29.3 1.7 22.8

```

Figure 2-2. DSSPファイルのフォーマット

【DSSP プログラムが求めた統計情報】

7 行目から 27 行目は、DSSP プログラムが PDB ファイルから求めた統計情報である。7 行目は、PDB ファイルの X 線結晶データから得られた全残基数、鎖数、全 SS 結合の数、チェーン内の SS 結合の数、チェーン間 SS 結合の数を書いている。8 行目は、蛋白質の表面積で単位は、Å²である。9 行目から 22 行目は、水素結合のタイプごとの数と残基ごとの数である。23 行目から 27 行目は、ヒストグラムの結果であり、24 行目から、αヘリックスごとの残基数、パラレルブリッジの数、アンチパラレルブリッジの数、シートごとのラダーの数の順に並んでいる。

【二次構造の帰属情報】

29 行目以降に DSSP プログラムが算出した二次構造の帰属情報を書いている。「#」は通し番号、「RESIDUE」は残基番号とチェーンナンバー、「AA」はアミノ酸残基の一文字表記、「STRUCTURE」は二次構造の情報、「BP1」と「BP2」はβブリッジの相手になっているアミ

ノ酸残基の通し番号, 「ACC」は溶媒接触表面積, 「N-H \rightarrow O」 と 「O \rightarrow H-N」 は, i 番目のアミノ酸残基と $i+n$ 番目のアミノ酸残基の水素結合エネルギー, 「TCO」は i 番目のアミノ酸残基の C=O と $i-1$ 番目のアミノ酸残基の C=O からなるコサインの値, 「KAPPA」は $i-2$ と i と $i+2$ 番目のアミノ酸残基の α 炭素がとる角度 (ベンド), 「ALPHA」は $i-1$ と i と $i+1$ と $i+2$ 番目のアミノ酸残基の α 炭素から作った二面体角 (キラリティー), 「PHI」 と 「PSI」 はアミノ酸残基の α 炭素の主鎖二面体角, 「X-CA」 と 「Y-CA」 と 「Z-CA」 はアミノ酸残基の α 炭素の X, Y, Z の座標である。なお, TCO は, 計算で値を求められているが, 二次構造に関わることには使われていない。

しかしながら, 水素結合の様式で二次構造を帰属すると, 主鎖のコンホメーションに規則性が見られない場合が生じたり, また, β シート構造の属する β スtrand構造と同じコンホメーション, つまり単に伸びた構造は, 水素結合の相手がいないため帰属できない¹⁰⁾。これらを解決した二次構造帰属法が ProSSA 法である。

2—3 主鎖二面体角の連続性パターンとしての ProSSA 法

ProSSA 法は, 蛋白質の主鎖二面体角の連続性パターンで二次構造を帰属する方法である¹¹⁾。主鎖の形の規則性に着目した二次構造帰属法であるため, 二次構造で不規則なコンホメーションは排除される。なお, ProSSA 法は, DSSP 法で帰属した二次構造を包括した方法でもある。ここでは, ProSSA 法で使われる二面体角の算出方法とアルゴリズムについて述べる。

【主鎖二面体角のアルゴリズム】

まず, 二次構造を帰属するために使う主鎖二面体角 (ϕ_i , ψ_i) を PDB ファイルに書いてある原子の三次元座標データから算出する (Figure 2-3)。算出のときに使う原子の原子座標は, 次の6つである: C_{i-1} の原子座標 (X_1, Y_1, Z_1), N_i の原子座標 (X_2, Y_2, Z_2), C_i^α の原子座標 (X_3, Y_3, Z_3), C_i の原子座標 (X_4, Y_4, Z_4), N_{i+1} の原子座標 (X_5, Y_5, Z_5)。 ϕ_i は, C_{i-1} , N_i , C_i^α が作る平面 J と, N_i , C_i^α , C_i が作る平面 K のなす角である。 ψ_i は, N_i , C_i^α , C_i が作る平面 K と, C_i^α , C_i ,

N_{i+1} が作る平面 L のなす角である。 ω は、 C_i^α , C_i , N_{i+1} が作る平面 L と、 C_i , N_{i+1} , C_{i+1}^α が作る平面 M のなす角である (Figure 2-4)。 $C_i^\alpha-C_i$ が N_i-H_i とトランスの位置であれば、 $\phi_i=0$, $C_i^\alpha-N_i$ が C_i-O_i とトランスの位置であれば $\psi_i=0$, $C_i^\alpha-C_i$ が $N_{i+1}-C_{i+1}^\alpha$ とトランスの位置であれば $\omega_i=0$ とする。また、 ϕ_i , ψ_i , ω_i の回転は、結合軸を N 末端から C 末端側を見て右回りを正の方向とする⁴²⁾。

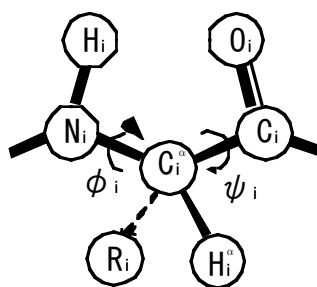


Figure 2-3. 主鎖二面体角 (ϕ_i , ψ_i)

つぎに、二面体角について述べる (Figure 2-4)。二面体角は、二つの平面 H, I からなる角度のことである (Figure 2-4(a))。その二面体角を求める計算式を示す。三つの原子座標 $C_{i+1}(X_1, Y_1, Z_1)$, $N_i(X_2, Y_2, Z_2)$, $C_i^\alpha(X_3, Y_3, Z_3)$ で作られる平面 J は、

$$\begin{vmatrix} X & Y & Z & 1 \\ X_1 & Y_1 & Z_1 & 1 \\ X_2 & Y_2 & Z_2 & 1 \\ X_3 & Y_3 & Z_3 & 1 \end{vmatrix} = 0,$$

すなわち、

$$A_1X + B_1Y + C_1Z + D_1 = 0$$

である。ただし、

$$\begin{aligned} A_1 &= Y_1Z_2 + Y_2Z_3 + Y_3Z_1 - Y_3Z_2 - Y_1Z_3 - Y_2Z_1 \\ B_1 &= X_3Z_2 + X_1Z_3 + X_2Z_1 - X_1Z_2 - X_2Z_3 - X_3Z_1 \\ C_1 &= X_1Y_2 + X_2Y_3 + X_3Y_1 - X_3Y_2 - X_1Y_3 - X_2Y_1 \\ D_1 &= X_3Y_2Z_1 + X_1Y_3Z_2 + X_2Y_1Z_3 - X_1Y_2Z_3 - X_2Y_3Z_1 - X_3Y_1Z_2 \end{aligned}$$

とする。

さらに, $N_i(X_2, Y_2, Z_2)$, $C_i^\alpha(X_3, Y_3, Z_3)$, $C_i(X_4, Y_4, Z_4)$ で作られる平面Kは,

$$A_2X + B_2Y + C_2Z + D_2 = 0$$

である。ただし,

$$A_2 = Y_2Z_3 + Y_3Z_4 + Y_4Z_2 - Y_4Z_3 - Y_2Z_4 - Y_3Z_2$$

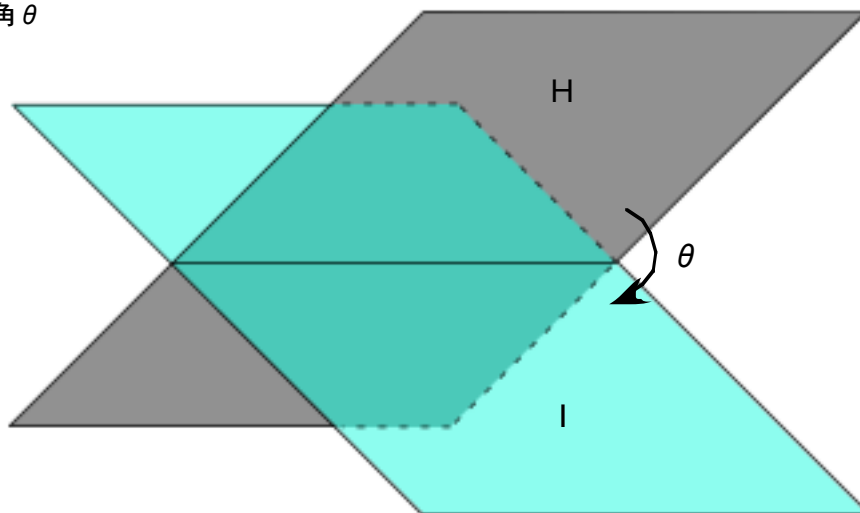
$$B_2 = X_4Z_3 + X_2Z_4 + X_3Z_2 - X_2Z_3 - X_3Z_4 - X_4Z_2$$

$$C_2 = X_2Y_3 + X_3Y_4 + X_4Y_2 - X_4Y_3 - X_2Y_4 - X_3Y_2$$

$$D_2 = X_4Y_3Z_2 + X_2Y_4Z_3 + X_3Y_2Z_4 - X_2Y_3Z_4 - X_3Y_4Z_2 - X_4Y_2Z_3$$

とする。

(a)二面体角 θ



(b)主鎖二面体角 (ϕ_i, ψ_i, ω_i)

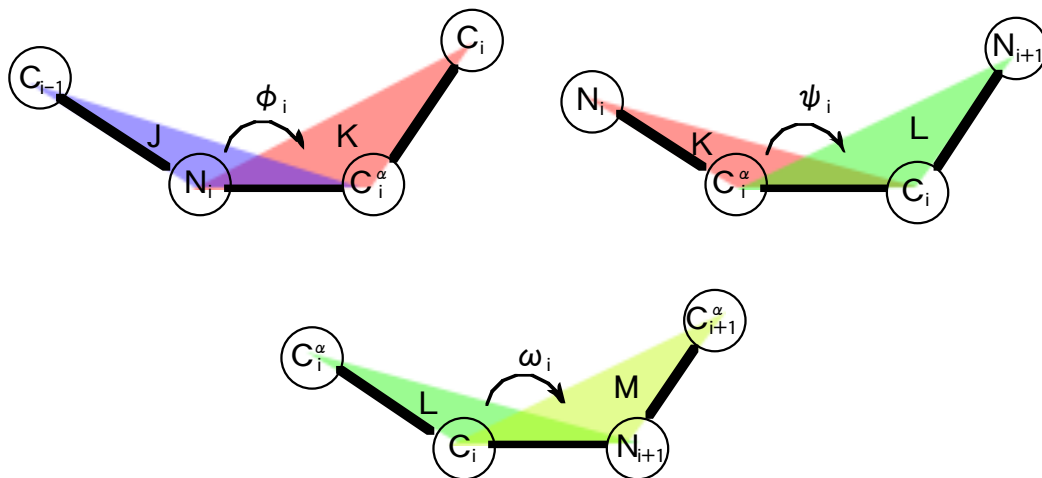


Figure 2-4. 2つの平面がなす角

したがって、二つの平面 J, K のなす角 ϕ_i は,

$$\cos \phi_i = \pm \frac{A_1 A_2 + B_1 B_2 + C_1 C_2}{\sqrt{A_1^2 + B_1^2 + C_1^2} \sqrt{A_2^2 + B_2^2 + C_2^2}}$$

より,

$$\phi_i = \pm \cos^{-1} \frac{A_1 A_2 + B_1 B_2 + C_1 C_2}{\sqrt{A_1^2 + B_1^2 + C_1^2} \sqrt{A_2^2 + B_2^2 + C_2^2}}$$

が得られる。このとき、符号は,

$$\begin{aligned} A_3 &= B_2(Z_3 - Z_2) - C_2(Y_3 - Y_2) \\ B_3 &= C_2(X_3 - X_2) - A_2(Z_3 - Z_2) \\ C_3 &= A_2(Y_3 - Y_2) - B_2(X_3 - X_2) \end{aligned}$$

より,

$$S = A_1 A_3 + B_1 B_3 + C_1 C_3$$

を求め、 $S > 0$ ならば $+\phi_i$, $S < 0$ ならば $-\phi_i$ とする。同様に、 ψ_i は平面 KL, ω_i は平面 LM から求めることができる。

つぎに、二次構造の定義に使う主鎖二面体角 (ϕ , ψ) の領域は、以下の三つである (Figure 2-5) :

- (1) $-128^\circ < \phi < -34^\circ$ かつ $-73^\circ < \varphi < 18^\circ$ (α_R 領域),
- (2) $+30^\circ < \phi < +120^\circ$ かつ $-30^\circ < \varphi < +70^\circ$ (α_L 領域),
- (3) $-180^\circ < \phi < -40^\circ$ あるいは $+120^\circ < \phi < +180^\circ$, かつ $-130^\circ < \varphi < -180^\circ$ あるいは $+60^\circ < \varphi < +180^\circ$ (β 領域)。

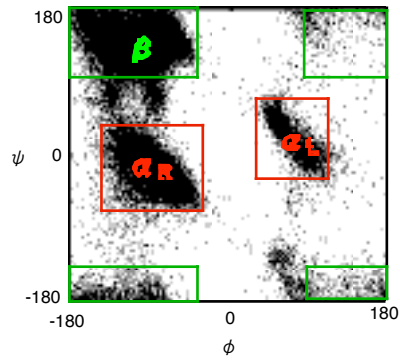


Figure 2-5. ProSSA法で使われる領域

これら三つの領域の中で、アミノ酸残基の α 炭素の二面体角 (ϕ , ψ) が、つぎのような連続性がみられるとき、二次構造と帰属する (Figure 2-6) :

- (1) α_R 領域内で4残基以上連続して現れ、その領域から1残基分外れた両末端にあるアミノ酸残基をそれぞれ N-cap, C-cap と名付け、N-cap から C-cap までを α_R ヘリックス構造,
- (2) α_L 領域内で4残基以上連続して現れ、その領域から1残基分外れた両末端にあるアミノ酸残基をそれぞれ N-cap, C-cap と名付け、N-cap から C-cap までを α_L ヘリックス構造,
- (3) β 領域内で2残基以上連続して現れたとき、 β スtrand構造。

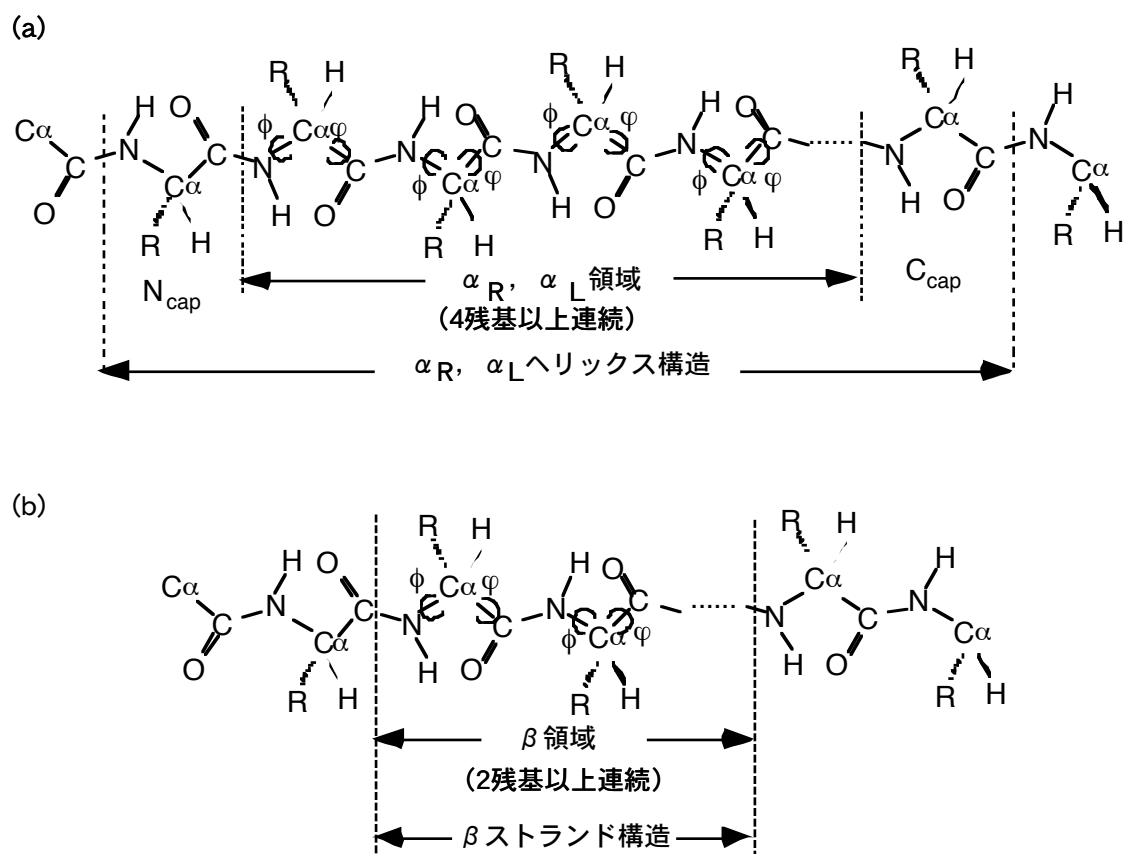


Figure 2-6. アミノ酸残基の連続性

このようなアルゴリズムによって PDB ファイルから主鎖二面体角の連続性パターンにより二次構造を帰属する。そして、ProSSA 法は、蛋白質の約 90%を二次構造と帰属できた (Figure 2-7) ¹¹⁾。

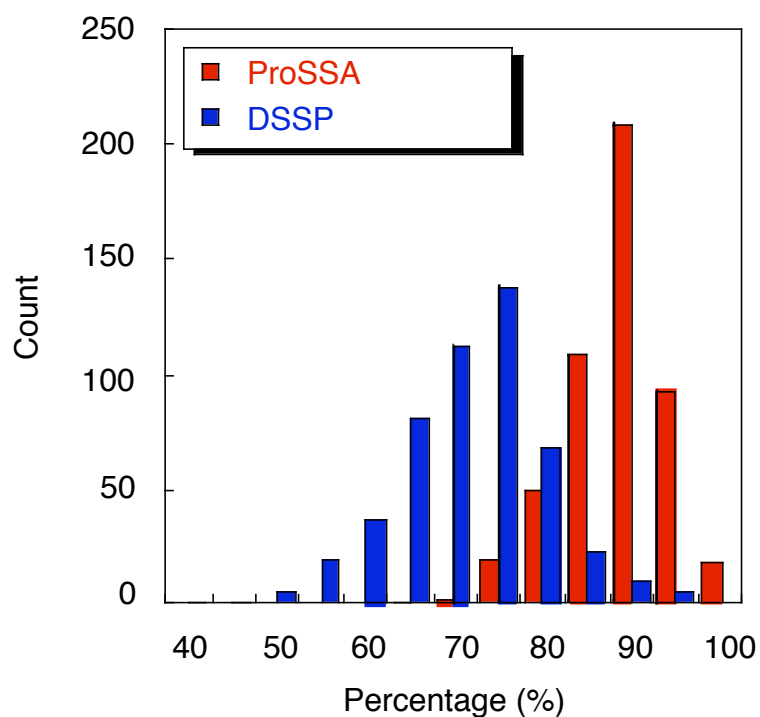


Figure 2-7. 二次構造の割合

2-4 ProSSA 法のデータベース

ProSSA 法によって得られた蛋白質の二次構造情報を集めたデータベースを構築するため、サーバーマシンを用意した (Table 2-2)。

Table 2-2. サーバマシンの構成

Name	IBM xSeries 206 8482-2EJ
CPU	Intel Pentium4 Processor 3.0GHz
Memory	1GHz (PC3200 ECC DDR DIMM)
OS	Vine Linux 3.2 ftp版
Web Application	Apache2-2.0.55-0v1.2

OS は、Vine Linux 3.2 ftp 版、Web アプリケーションは、Apache 2.x 系列 (Vine Linux のオリジナルパッケージ Apache2-2.0.55-0v1.2) を使った。このサーバーマシン上に構築した ProSSA デ

ータベースには、<http://www.ohuchi-lab.bio.kyutech.ac.jp/database/prossa/>からアクセスすることができる (Figure 2-8)。

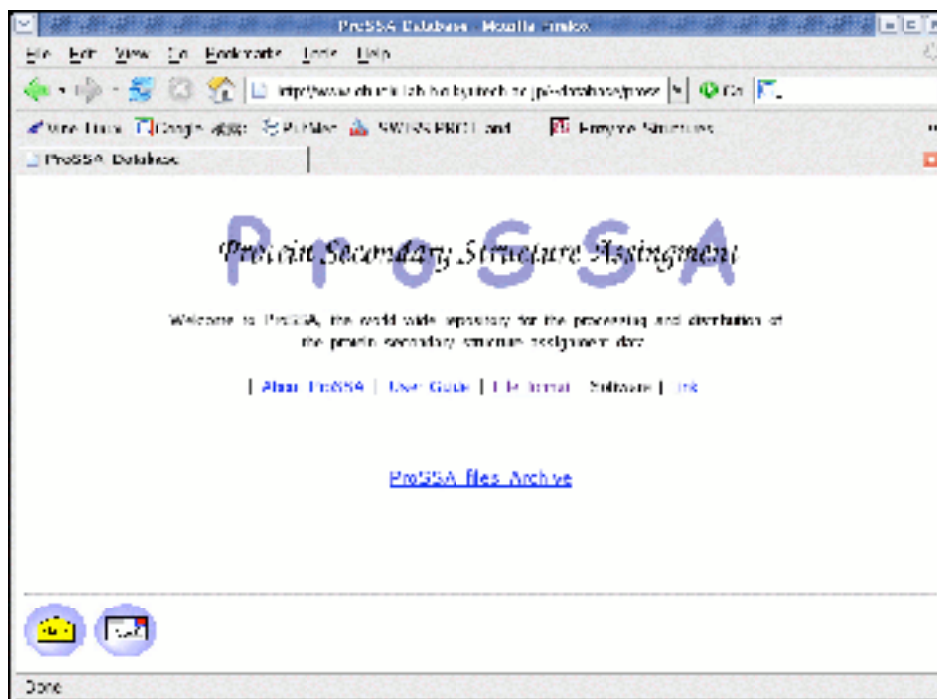


Figure 2-8. ProSSAデータベースのトップページ

このサーバマシン上に、ProSSA 法によって得られる帰属ファイルを蓄積していくが、そのファイルのフォーマットを Figure 2-9 に示す。

帰属データは、大きく二つの項からなる。一つは、PDB の HEADER 情報を抽出した部分、もう一つは、二次構造を帰属した結果の部分である。PDB の HEADER 情報は、PDB ファイルから HEADER, TITLE, EXPDTA, AUTHOR, JRNL の 5 つのレコードを抜き出した。二次構造の帰属結果は、次のような情報が左の列から書かれてある：1) アミノ酸配列情報 (通し番号 (#), アミノ酸残基番号 (seqNum), サブユニット番号, アミノ酸残基の一文字表記 (AA)), 2) 主鎖二面体角情報 (ϕ 角 (PHI), ψ 角 (PSI), ω 角 (OMEGA), 拡張 ϕ 角 (E_PHI), 拡張 ψ 角 (E_PSI)), 3) 二次構造帰属の帰属データ (α_R ヘリックス (ALPHA_R), α_L ヘリックス (ALPHA_L), β スtrand (BETA)), 6) α 炭素の三次元座標 (x, y, z) (ca_X, ca_Y, ca_Z),

7)温度因子 (B_value)。

***** PROTEIN SECONDARY STRUCTURE DEFINITION BY THE PROGRAM PSSA, VERSION 1.0 ***** DATE=2003-06-27
REFERENCE K. KONKO AND S. OHUCHI (2003) XXXX,Y-Z.

HEADER HYDROLASE 13-FEB-96 1DPM
TITLE THREE-DIMENSIONAL STRUCTURE OF THE ZINC-CONTAINING
TITLE 2 PHOSPHOTRIESTERASE WITH BOUND SUBSTRATE ANALOG
TITLE 3 DIETHYL 4-METHYLBENZYLPHOSPHONATE
EXPDTA X-RAY DIFFRACTION
AUTHOR J.L.VANHOOKE,M.M.BENNING,F.M.RAUSHEL,H.M.HOLDEN
JRNL AUTH J.L.VANHOOKE,M.M.BENNING,F.M.RAUSHEL,H.M.HOLDEN
JRNL TITL THREE-DIMENSIONAL STRUCTURE OF THE ZINC-CONTAINING
JRNL TITL 2 PHOSPHOTRIESTERASE WITH THE BOUND SUBSTRATE ANALOG
JRNL TITL 3 DIETHYL 4-METHYLBENZYLPHOSPHONATE
JRNL REF BIOCHEMISTRY V. 35 6020 1996

PDB header information

#	Amino Acid Sequence		Dihedral Angle					Assignment Data			a-carbon atom co-ordinates			B value
	seqNum	AA	PHI	PSI	OMEGA	E_PHI	E_PSI	ALPHA_R	ALPHA_L	BETA	ca_X	ca_Y	ca_Z	B_val
1	35	A A	360.0	-56.5	179.5	360.0	303.5	0	0	0	29.6	14.3	18.9	97.4
2	36	A R	-124.8	150.1	178.2	235.2	150.1	0	0	E	28.7	10.7	19.9	40.7
3	37	A I	-112.6	137.2	-179.6	247.4	137.2	0	0	E	30.7	7.8	21.1	17.8
4	38	A N	-81.0	122.6	178.6	279.0	122.6	0	0	E	30.2	4.2	20.0	19.7
5	39	A T	-119.9	169.9	-179.6	240.1	169.9	0	0	E	29.3	1.7	22.8	13.9
6	40	A V	-71.7	-0.4	178.9	288.3	359.6	ZRT	0	0	28.2	-1.9	22.7	13.1
7	41	A R	-114.2	-0.0	-178.8	245.8	360.0	ZRT	0	0	24.7	-0.7	23.3	23.9
8	42	A G	126.5	-169.6	-179.3	126.5	190.4	0	0	E	24.9	2.0	20.7	16.5
9	43	A P	-86.0	144.5	176.8	274.0	144.5	0	0	E	26.1	5.6	20.6	32.8
10	44	A I	-130.2	158.7	178.6	229.8	158.7	0	0	E	26.0	7.9	23.5	14.3
11	45	A T	-81.1	156.0	179.7	278.9	156.0	Nc	0	E	26.6	11.6	23.8	18.2
12	46	A I	-53.3	-40.3	-179.4	306.7	319.7	H	0	0	29.8	12.7	25.3	23.3
13	47	A S	-61.7	-33.8	-179.9	298.3	326.2	H	0	0	27.9	14.0	28.4	31.2
14	48	A E	-68.0	-22.8	-178.9	292.0	337.2	H	0	0	26.5	10.7	29.0	20.4
15	49	A G	-65.4	-31.6	-179.7	294.6	328.4	H	0	0	29.9	9.0	29.5	21.1
16	50	A G	49.1	-148.6	178.9	49.1	211.4	Cc	0	0	30.6	10.6	33.0	13.6
17	51	A F	-55.4	118.5	-177.5	304.6	118.5	0	0	E	33.3	8.9	35.1	19.5

Figure 2-9. ProSSAの帰属データのフォーマット (1DPM)

ProSSA のファイルで出力されていて、DSSP のファイルでは出力されていない情報として、温度因子 (B-value) がある。B-value は原子の揺らぎを表わすパラメータである。このパラメータは、蛋白質の二次構造のゆらぎ^{11,43)}、蛋白質の揺らぎの予測⁴⁴⁾など、さまざまな研究に使われている。したがって、二次構造帰属ファイルに B-value を書き出すのは有用である。

PrSSA データベースを構築したサーバマシンは、第4章の CD スペクトルのデータベース、第6章の α ヘリックス構造の評価とデザインのためのプログラムにも使うマシンでもある。

第3章 α ヘリックス構造の解析と配列デザイン

3-1 はじめに

蛋白質の立体構造は、一次構造、二次構造、三次構造、四次構造の順に階層的にとらえられる。この階層性は立体構造の形成の順番とも考えられている。つまり、蛋白質は一次構造（アミノ酸配列）から始まり、局所的な繰り返し構造である二次構造を形成し、それらが空間上のさまざまな位置に配置されることで、三次構造、場合によっては、四次構造を形成する。このような階層性の関係をみても、二次構造は立体構造を形成する上で礎の構造といえる。したがって、これまでも多くの研究者が一次配列から二次構造の予測をしてきた^{31,45,47)}。初期の研究として、Chou と Fasman は、球状蛋白質の二次構造の中に含まれるアミノ酸残基の出現頻度である Chou-Fasman パラメータを使って二次構造を予測した。しかしながら、Chou-Fasman パラメータは、統計的なパラメータを算出するのに用いた蛋白質の数が少ないために問題視されている⁴⁸⁾。くわえて、Chou-Fasman 法は二次構造の核を中心に予測を行なっているため、二次構造の境界領域が不明確となる問題もある。Chou-Fasman パラメータは二次構造の領域に対するアミノ酸残基の傾向指数でありながら、予測を二次構造の核としている矛盾な論理があることも予測精度に影響を与えていると考えられる。一方、二次構造に対するアミノ酸残基の傾向指数の中でも、二次構造末端の境界領域をターゲットにしたパラメータ（Richardson パラメータ）がある^{11,30,34,49)}。Richardson らは、二次構造の末端境界領域にポジションをさだめ、各ポジションでのアミノ酸残基の傾向指数を求めた。これは、ポジションごとの Chou-Fasman パラメータを算出した手法とも言える。

この章では、二次構造の境界を明確にする ProSSA 法によって蛋白質二次構造を帰属し、そこから Richardson パラメータを算出し、二次構造の境界領域の予測を行なった。また、予測結果が外れた領域に対しては、蛋白質の二次構造設計へ応用するための方法を示した。

3—2 α ヘリックス構造の境界領域の予測

金光は蛋白質の立体構造をトポロジカルに分類しているデータベース SCOP から、1081 種類の立体構造上、偏りのない球状蛋白質を選び、ProSSA 法によっても二次構造を帰属し Richardson パラメータを算出した (Figure 3-1) ¹¹⁾。

	N''	N'	Ncap	N1	N2	N3	N4	C4	C3	C2	C1	Ccap	C'	C''
A	0.9	0.7	0.4	1.1	1.3	1.2	1.4	1.7	1.4	1.4	1.2	0.6	0.8	0.8
C	1.1	1.1	0.9	0.6	0.6	0.8	1.0	1.4	0.9	0.8	1.2	1.2	0.8	0.7
D	0.9	0.9	2.7	0.9	1.7	1.5	0.5	0.6	0.7	1.0	0.6	0.9	1.1	1.6
E	1.0	0.8	0.6	1.3	2.5	2.0	0.6	0.9	1.2	1.6	1.0	0.6	0.8	1.1
F	1.0	1.1	0.4	0.9	0.7	1.0	1.4	1.3	1.1	0.7	1.1	1.1	0.9	0.8
G	1.5	1.1	1.7	0.5	0.7	0.6	0.3	0.4	0.3	0.3	0.4	3.7	1.1	1.1
H	1.1	1.0	1.2	0.8	1.0	1.0	0.7	0.8	1.0	0.9	1.4	1.3	0.8	1.2
I	0.9	1.2	0.2	0.8	0.5	0.7	1.6	1.2	1.1	0.8	0.7	0.7	0.9	0.8
K	0.9	0.8	0.5	0.9	1.0	0.9	0.9	0.8	1.4	1.8	1.3	0.9	1.2	1.2
L	0.8	1.2	0.3	1.0	0.6	0.9	1.8	1.9	1.6	1.1	1.5	0.8	0.8	0.7
M	0.8	1.4	0.3	0.9	0.6	0.9	1.4	1.7	1.3	1.2	1.3	0.8	0.8	0.6
N	1.1	0.9	2.3	0.5	0.9	0.7	0.5	0.5	0.7	0.9	1.3	1.6	1.0	1.1
P	1.3	1.3	1.4	3.2	0.5	0.3	0.2	0.4	0.2	0.3	0.0	0.0	3.0	1.9
Q	0.9	0.9	0.5	0.9	1.4	1.7	1.1	1.0	1.2	1.3	1.3	0.9	0.8	0.9
R	1.0	0.8	0.5	0.9	1.0	0.9	1.4	0.9	1.4	1.5	1.2	0.8	0.9	1.0
S	1.1	0.9	2.6	0.7	1.3	1.0	0.5	0.7	0.8	1.1	1.1	0.8	0.9	1.1
T	1.0	1.0	2.0	0.8	0.8	1.0	0.7	0.6	0.6	0.7	1.1	0.7	0.9	1.0
V	0.9	1.1	0.2	0.9	0.5	0.9	1.4	0.9	0.8	0.6	0.6	0.6	0.7	0.6
W	0.7	0.9	0.4	1.5	1.2	1.0	1.4	1.7	1.1	0.8	0.8	0.5	0.7	0.8
Y	1.0	1.0	0.6	1.0	0.8	0.9	1.2	1.1	0.9	0.7	1.2	1.0	0.9	0.9

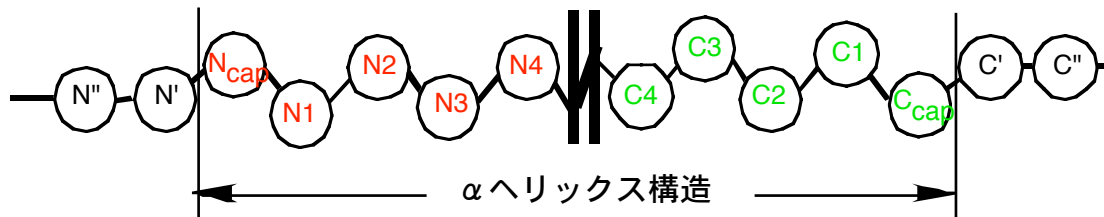
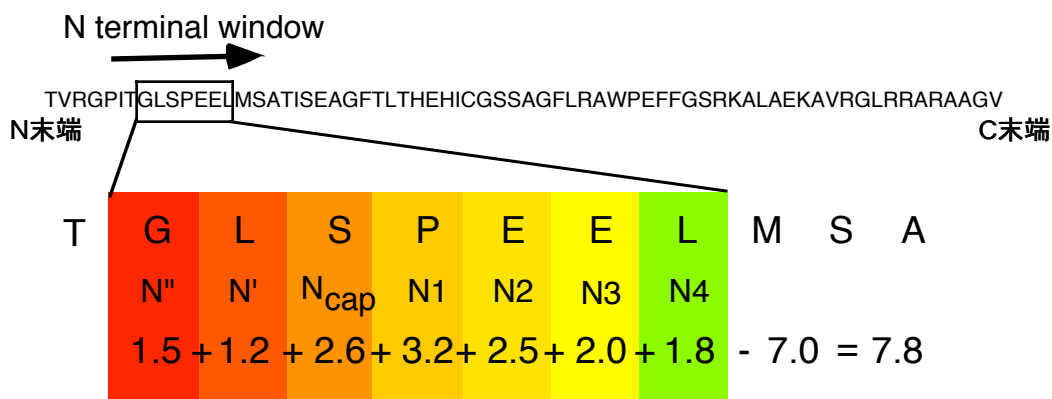


Figure 3-1. α ヘリックス構造の末端領域とその領域のアミノ酸残基の傾向指数

Figure 3-1 には、Ncap から N4、C4 から Ccap までを α ヘリックス構造の中の両末端のポジションでのアミノ酸残基の傾向指数、また、N'', N' と C', C'' は α ヘリックス構造から外れた 2 残基分のポジションでのアミノ酸残基の傾向指数を示している。赤色の傾向指数は、1.5 より高い値、青色は、0.5 より低い値を示す。Figure 3-1 から、両末端の境界領域のアミノ酸残基の傾

向指数は、1.0 に収束しているものの依然として偏りがあった¹¹⁾。この偏りがヘリックス末端の特徴と考え、アミノ酸配列から α ヘリックスの両末端がどの位置に現れるかという予測ができると考えた。つまり、 α ヘリックス構造の形成傾向をみるためのスコアを算出し、末端領域を予測した。つぎに、 α ヘリックス構造の形成傾向のスコアを算出するためのアルゴリズムについて述べる (Figure 3-2)。

(a) N α スコア



(b) C α スコア

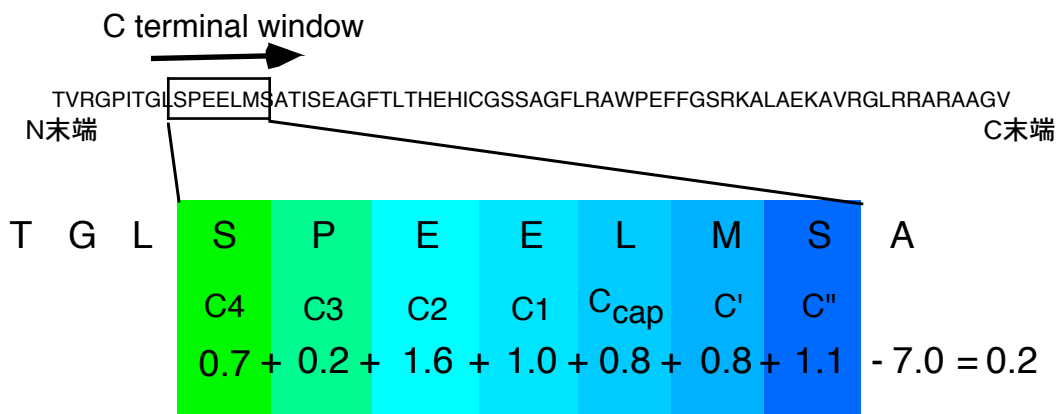


Figure 3-2. α ヘリックス構造の形成傾向スコアの算出法

まず、 α ヘリックス構造の N 末端の形成傾向を表すスコア (N α スコア) について述べる (Figure 3-2(a))。予測に使うアミノ酸配列で、7 残基で一組となるウインドウを作り、そのウインドウの中のアミノ酸残基に対して、N 末端側から記号 N', N'', N_{cap}, N1, N2, N3, N4 を

順に割り当てた。そして、割り当てた際のアミノ酸残基に対応した傾向指数を Table 3-1 から取得し、それらの合計から 7 を引いた値を Ncap の位置にあるアミノ酸残基のスコアとして、Na スコアと名付けた。たとえば、Figure 3-1(a)の場合、7 残基のウインドウは GLSPEEL である。この配列に対して、N'', N', Ncap, N1, N2, N3, N4 を順に当てはめていく。そして、Table 3-1 から、グリシンは 1.5、ロイシンは 1.2、セリンは 2.6 のように各ポジションを対応づけ、ウインドウ内の傾向指数の総和から 7 を引いた値 7.8 を、記号 Ncap のスコアとする。このウインドウはアミノ酸配列に沿って 1 残基ずつスライドさせ、Ncap のスコアの変化を見ることができる。C 末端側のスコア (Ca スコア) も同様に算出した (Figure 3-2(b))。なお、ウインドウ内の配列に対する傾向指数の総和から 7 を引くのは、基準を 0.0 にするためである。これによって、0 以上のスコアがもつ領域が α ヘリックス構造を形成しやすい領域とみなせる。

7 残基を一組のウインドウにした理由は、 α ヘリックス構造の末端領域で形成される特異的な構造形成の可能性を考慮している。まず、N 末端側のウインドウについて述べる：1) Ncap と N4 は、 α ヘリックス構造の内部で水素結合を形成している i 番目と $i+4$ 番目のアミノ酸残基に対応している、2) N3 と Ncap は、側鎖と主鎖の間で capping box を形成する^{28,50)}、3) N1, N2 は、ほかのポジションのアミノ酸残基とも水素結合を形成しないが、 α ヘリックス構造の一巻きが 3.6 残基であるから、スコアを算出するにあたり、N1, N2 を除いて算出するより、含めて算出した方が合理的である。同様に、C 末端側のウインドウについて述べる：1) C4 と Ccap は、 α ヘリックス構造の内部で水素結合を形成している i 番目と $i+4$ 番目のアミノ酸残基に対応する、2) C'' と C3 と C' と C2 は、主鎖間の水素結合を形成する³⁰⁾、3) C1 は、ほかのポジションのアミノ酸残基とも水素結合をしないが、N2, N3 のときと同様に、C1 のみを除いてスコアを算出するよりも、C1 を含めた方が合理的と考えられる。ウインドウを使ってスコアを算出することで配列上の近距離相互作用を反映できる。

実際に、この方法を PDB コード 1D2Z の Chain B に対して適用した (Figure 3-3)。横軸は、アミノ酸配列、縦軸はスコアである。アミノ酸配列のところには、PDB ファイルに対して ProSSA 法によって帰属して得られた α ヘリックス構造も描いた。また、青線は、Na スコア、赤線は Ca スコアである。波線は、二次構造予測ソフトウェア GENETYX に組み込まれている

Chou-Fasman 法から得られた α ヘリックス構造予測の結果である^{26,51)}。Chou-Fasman 法の結果は、もともと 1.0 を基準としているが、 $N\alpha$, $C\alpha$ スコアの基準値 0.0 に合わせるために、算出結果から 1 を引いて、10 をかけた値を用いた。たとえば、Chou-Fasman 法で得られた結果が 0.9 ならば、変換後は $(0.9-1) \times 10 = -1$ である。

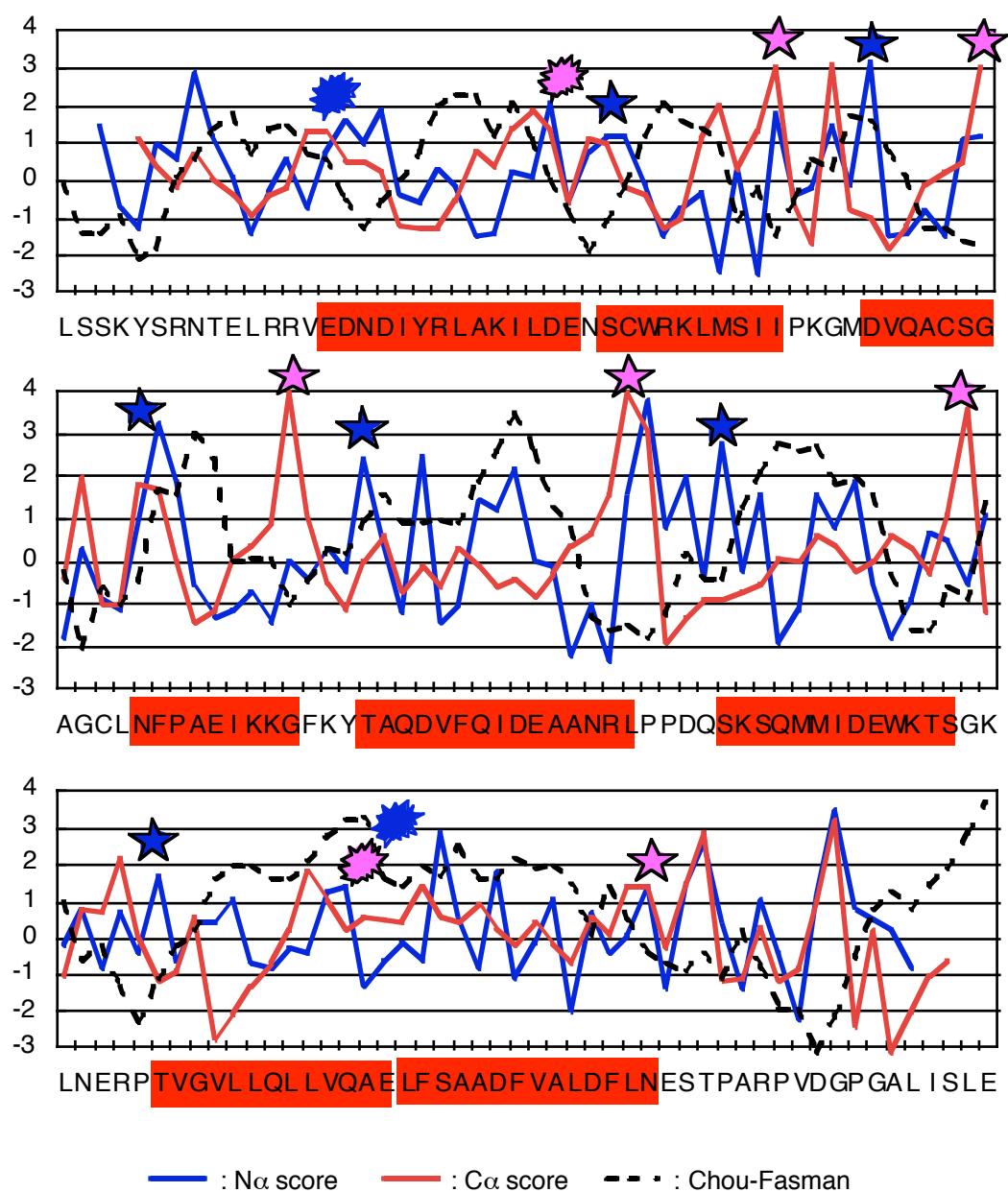


Figure 3-3. α ヘリックス構造の末端領域の予測 (1D2Z chain B)

$N\alpha$, $C\alpha$ スコアを予測スコアととらえると、実際の α ヘリックス構造の末端位置の前後にス

コアの最大点があった。そこで、実際の末端位置とスコアの最大値に相関があった部分に星印をつけた。星印ではないマークは、相関がなかった位置である。

Chou-Fasman 法による二次構造予測と、実際の二次構造の帰属結果を照らし合わせると、Chou-Fasman 法は、二次構造の核の部分を予測した結果であった。つまり、境界領域については予測できなかった。ところが、 $\text{N}\alpha$, $\text{C}\alpha$ スコアを見ると、Chou-Fasman 法で予測した α ヘリックス領域の前後に最大点があった。そして、その最大点は、実際の二次構造の帰属結果の末端の部分にあった。つぎに、他の蛋白質に対して、 α ヘリックス構造の末端領域を予測した (Table 3-1)。

Table 3-1. 6種類の蛋白質に対する予測

	PDB code	ヘリックス 全残基数	ヘリックス 含量 (%)	実際の末端の 残基数	相関個数	$\frac{\text{相関個数}}{\text{実際の末端残基数}} (\%)$
N末端側	1D2Z	96/150	64	8	6	75
	1A43	55/72	76	5	3	60
	3UBP	61/100	61	4	3	75
	1AGI	35/125	28	3	3	100
	1ASX	64/152	42	4	3	75
	1FCD	112/174	64	8	4	50
C末端側	1D2Z	96/150	64	8	6	75
	1A43	55/72	76	5	3	60
	3UBP	61/100	61	4	2	50
	1AGI	35/125	28	3	1	33
	1ASX	64/152	42	4	3	75
	1FCD	112/174	64	8	4	50

Table 3-1 は、6 種類の蛋白質に対する予測結果である。上から二つずつ all alpha クラス, alpha+beta クラス, alpha/beta クラスの蛋白質を示す。左から、PDB コード、蛋白質の全残基数に対する α ヘリックス構造中の残基数とヘリックス含量、PDB ファイルに対して二次構造を帰

属したときの α ヘリックス構造の末端の残基の数、末端の残基と予測の相関の数、実際の末端の残基の数に対する相関の数の割合を並べた。

N 末端側の結果をみると、ヘリックス含量に多少の影響を受けているものの、実際の末端の残基数に対する相関の数の割合は、それぞれの蛋白質に大きな差がないと判断される。一方、C 末端側も、それぞれの蛋白質に大きな差がない。しかしながら、alpha+beta クラスの 1AGI のみが、他の蛋白質に比べて、N 末端側が 100%、C 末端側が 33%と極端に異なる値を示した。これは、1AGI の N 末端側は、他の蛋白質よりも、アミノ酸残基の特異性の影響、つまり近距離の相互作用による影響が大きいと考えられる。逆に、C 末端側は、アミノ酸残基の特異性の影響が小さく、二次構造同士の相互作用が関与している。

以上のことから、 α ヘリックス構造の末端領域でのアミノ酸残基の傾向指数を使うことで、構造の形成傾向を算出し従来の予測ではあいまいだった末端残基を精密に特定できた。

3—3 α ヘリックス構造の不安定領域の検索とその手法

前節までは、アミノ酸配列に対して二次構造の形成傾向スコアを算出し、二次構造の末端領域を予測した。その結果、従来の予測法では得られなかった末端部位を予測できた。ところが、予測から外れた領域が存在した。この外れた領域を、今度は蛋白質設計の観点にたって考える。つまり、構造が明らかになっている蛋白質に対してスコア算出法を適用し、実際の蛋白質の末端であるにも関わらずスコアが低いとき、どのような解釈をすればよいだろうか。この研究ではスコアが二次構造の形成傾向を示すことから、実際には、二次構造を形成しているにも関わらず、潜在的に二次構造の末端が不安定の可能性があると考え。つまり、末端領域に、不安定の原因を作っている因子が存在する。これは、二次構造の末端を安定化させることで、二次構造全体が安定化し、ひいては、高次構造も安定化するという考えに基づいている。たとえば、Figure 3-4 は、実際には、 α ヘリックス構造の末端でありながら、Ncap 部位の Na スコアが低い例である。アミノ酸配列の黄色で囲んでいる領域は、7 残基一組のウィンドウであり、配列の下の赤色の線は、実際の α ヘリックス領域である。Ncap 部位の Na スコアが 0 よりも低い値を

示している。スコアが低い原因は、ウインドウ内の配列で傾向指数が低いバリン残基（Ncap 部位）、あるいはアスパラギン酸残基（N4 部位）であると考えた。これらの残基を Figure 3-1 の傾向指数表より、それぞれの部位で傾向指数が高いアミノ酸残基に置き換えることで、スコアを高くすることができる。このように、スコアを高くすることで二次構造の末端領域を安定化させると考えられる。実際に、所属研究室において部位特異的変異実験を行ない、さまざまな変異蛋白質を作製し構造解析をしてきた。これらの手順を Figure 3-5 に示す。

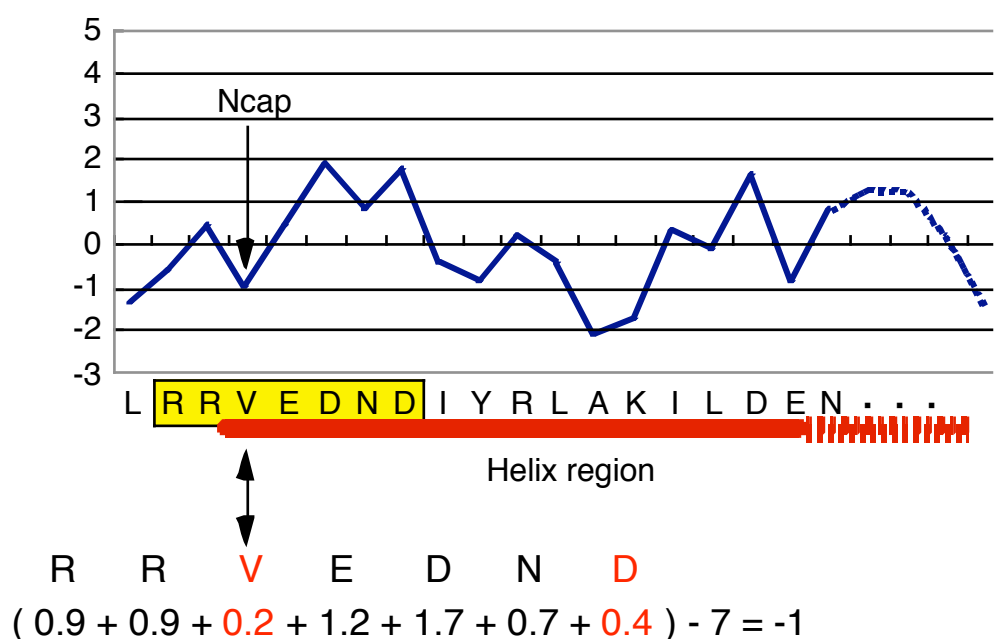


Figure 3-4. 構造末端の不安定領域の検索

このように、従来は、経験などによって変異を導入するポイントを探したり、変異を導入する際のアミノ酸残基の候補として 19 種類を考える必要があったりしたものが、この手法により、変異導入のポイントを絞り込み、さらには、変異導入するアミノ酸残基の候補も絞り込むことができる。

なお、Figure 3-5 の「構造解析」は、CD スペクトルや IR スペクトルなど分光学的手法によって、デザインした蛋白質がどのような変化を生じているのかを調べることを意味する。

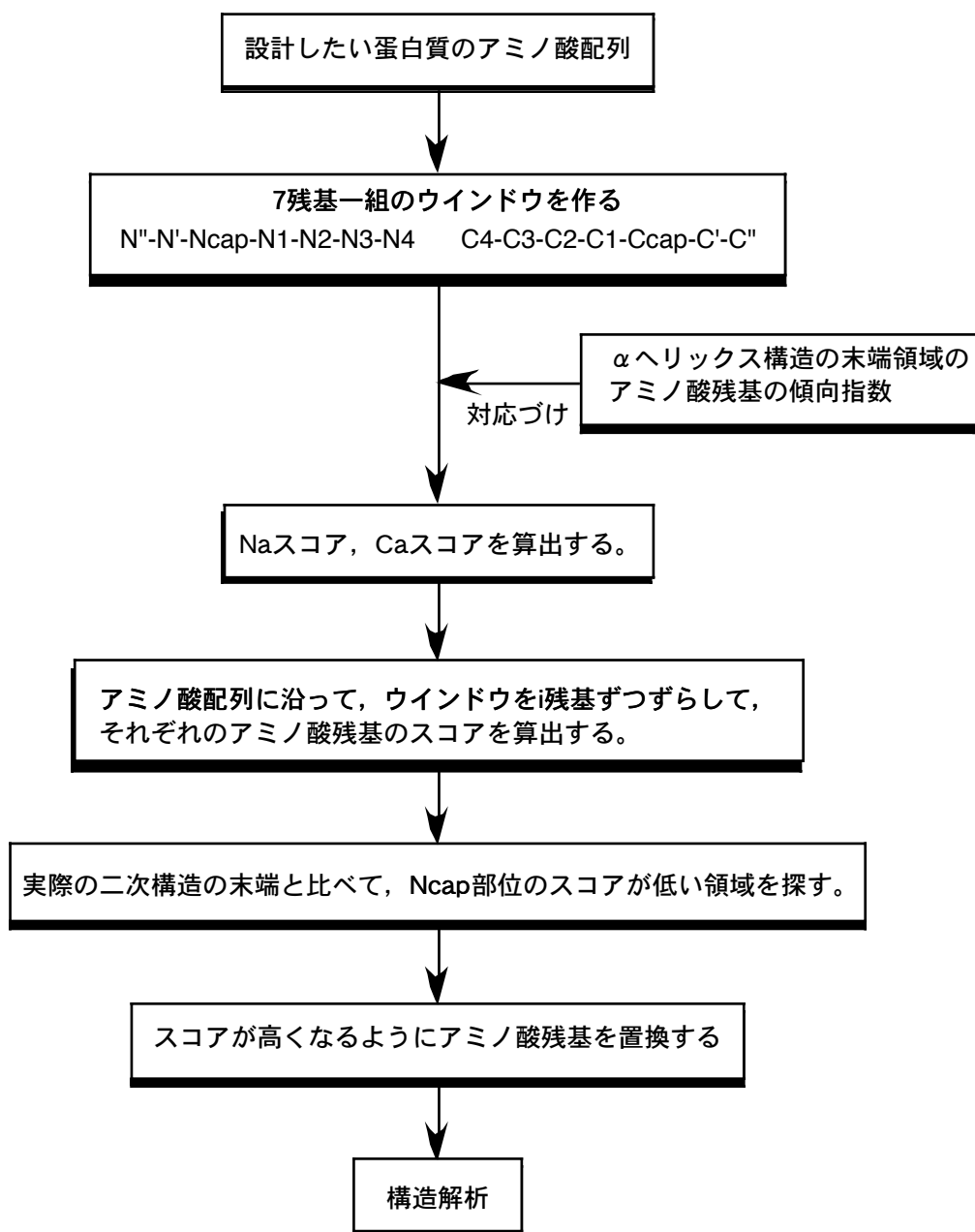


Figure 3-5. 蛋白質をデザインするための手順

3—4 種々の蛋白質への適用

3—3節で述べた蛋白質デザインの手法を，他の研究者によって報告された 19 種類の変異蛋白質の中で， α ヘリックス構造の末端部位に変異を導入している変異蛋白質を評価した (Table 3-2)。変異蛋白質のデータは，熱動力学的データベース ProTherm から取得した⁴¹⁾。

Table 3-2. 種々の変異蛋白質への適用

PDB code	mutation	helix末端領域の ポジション	Ncap, Ccap残基の スコア(wild type)	Ncap, Ccap残基の スコア(mutant)	ΔT_m ($^{\circ}\text{C}$)
1STN	V66L	C3	1.4	2.2	+3.9
2RN2	Q113P	C1	0.8	-0.5	-2.1
	H114A	Ccap	0.8	0.1	-7.7
2LZM	K60P	N1	2.1	5.4	+0.3
	T115E	N1	-0.1	0.4	-1.7
	N116D	N2	-0.1	0.7	-0.2
	G113A	C1	-0.9	-0.1	+0.4
	A82P	C'	1.4	1.4	+0.8
1ARR	E17A	N2	2.4	1.2	-2.0
1CTS	D375G	N2	1.8	0.5	+4.6
1BVC	G23A	N3	0.9	1.5	-2.2
1TTP	D112G	N2	1.8	0.8	-0.6
	C81G	N4	4.8	4.1	-4.8
	S33L	N4	2.4	3.7	-7.6
	C81S	N4	4.8	4.3	-4.4
	C118S	C4	3.5	2.8	-8.5
	C118V	C4	3.5	3.0	-4.3
1DPM	R207S	Ncap	0.2	2.3	30以上
	G209D	N2	0.2	1.2	20以上

Table 3-2 は，左の列から，野生型の蛋白質の PDB コード，変異導入部位，変異を導入したときのヘリックス末端領域のポジション，野生型のヘリックス末端部位のスコア，変異を導入

したあとのヘリックス末端部位のスコア、野生型と変異体の変性温度の差を表している。赤文字は、アミノ酸残基を置換する前のスコアよりも、置換したあとのスコアの方が高いとき、 ΔT_m が正の値を示したことを意味する。あるいは、置換前のスコアよりも、置換後のスコアの方が低いとき、 ΔT_m が負の値を示したことを意味する。青文字は、赤文字とは反する現象が生じた場合を意味する。たとえば、PDB コード 1STN のとき、C3 部位で 0.8 の傾向指数を示すバリンから、1.6 の傾向指数をもつリジンに置換したとき、スコアが 1.4 から 2.2 へ増えて、そのときの ΔT_m は 3.9 なので、結果として、熱に対する安定化を図る方向へアミノ酸置換をしたと考えることができる。PDB コード 1DPM は金光の結果である¹¹⁾。野生型の蛋白質に対して、変異部位 R207S と G209D は、それぞれ 30 °C 以上、20 °C 以上の熱安定化が達成されたことを示す。

Table 3-2 から、約 70% の蛋白質について、 α ヘリックス構造の末端部位のスコアが野生型より変異体の方が高いとき、 ΔT_m は正の値を示した。さらに、末端部位のスコアが野生型よりも変異体の方が低いとき、 ΔT_m は負の値を示した。これらの結果から、スコアが高ければ ΔT_m は高い場合が多く占めた。しかしながら、1DPM 以外の蛋白質ではスコアや ΔT_m に大きな変化は見られなかった。これは、今回の手法を使って、構造の安定化を目指した実験ではないからである。しかしながら、19 種類の変異蛋白質のうち、70% の蛋白質において相関があったため、この手法は熱安定化を図るための一つの手法として使える。

これまでに報告されている変異導入による構造安定化は、さまざまな着眼点がある。Hiraga と Yutani は、蛋白質内のフリーのシステイン残基が与える安定性への役割を調べるためにシステイン残基を他のアミノ酸残基へ置換した⁵²⁾。Pinker らは、蛋白質表面に露出している α ヘリックスの領域に注目して変異を導入しているが、変異導入に使う蛋白質に対して、明らかにされた相互作用の結果を使った⁵³⁾。Zhi らは、蛋白質の活性部位へ変異を導入して、活性と構造の安定性について調べた⁵⁴⁾。Nicholson らは、二次構造に着目しているものの、ヘリックス双極子に基づいて変異を導入した⁵⁵⁾。このように、変異を導入する箇所を決定するためには、ある程度の目星をつけて取り組んではいるものの、定量的な手法を使って実験に取り組んでいるわけではない。場合によっては、変異導入後理由付けをしている研究もある。変異導入のためには

実験の経験がそれなりに必要である。それに対して、この研究のスコア算出法を使った変異導入部位の検索法は、 α ヘリックスの末端領域の特異性を使って、数値化し、どのような球状蛋白質に対しても、設計に必要なデータを系統立てて提供できる方法である。つまり、変異導入の実験に取り組む研究者や技術者にとって、実験の経験があまりなくても、この研究の手法を使えば、変異導入部位のポイントが絞り込め、すぐに実験に取り掛かることができるといえる。

3—5 α ヘリックス構造に対する変異導入部位の検索プログラムの開発

3—3節において、スコア算出法を利用した変異導入部位の検索とアミノ酸残基の置換候補の提示をする方法について述べた。この手法が、変異導入実験に慣れていない研究者や技術者にとって、有用であることも述べた。しかしながら、手法としては便利であるものの、実験者が、配列に対してスコアを算出していくには時間がかかる懸念がある。そこで、この節では、手法を自動化するためのアプリケーションの開発について述べる。アプリケーションは、Webブラウザで使えるようにした。

利用者が、あらかじめ用意するものは、設計したい構造既知の蛋白質のアミノ酸配列と、 α ヘリックス構造の領域である。 α ヘリックス構造の領域は、ProSSA法の帰属結果を使う必要がある。なぜならば、スコアを求めるのに使っているRichardsonパラメータは、ProSSA法の α ヘリックス構造の帰属結果に基づいているためである。

アプリケーションのインターフェースは、二つのフレームに分かれている (Figure 3-6)。上のフレームには、設計したい蛋白質の名前を入力する入力フォーム、アミノ酸配列を入力する入力フォーム、 α ヘリックス構造の領域を指定するプルダウン形式のメニューがある。これらは、すべて必須の項目である。Figure 3-6は、蛋白質として、1FCDを使った例である。PDB codeの入力フォームに「1FCD」を、Sequenceの入力フォームに1FCDの一文字表記で書いたアミノ酸配列を、helix regionのところにデザインしたいヘリックスの領域をN末端の残基番号とC末端の残基番号で指定し、検索する。すると、下のフレームにアミノ酸配列、ヘリックス領域、スコアの結果が表示される。下のフレーム内の一行目にはアミノ酸配列に対する通し番号、二

行目に入力したアミノ酸配列，三行目と四行目にそれぞれ $N\alpha$ スコアと $C\alpha$ スコアが出力される。なお，アミノ酸配列の中の赤色の部分は，指定したヘリックス領域である。灰色は，スコアが負の値のときである。

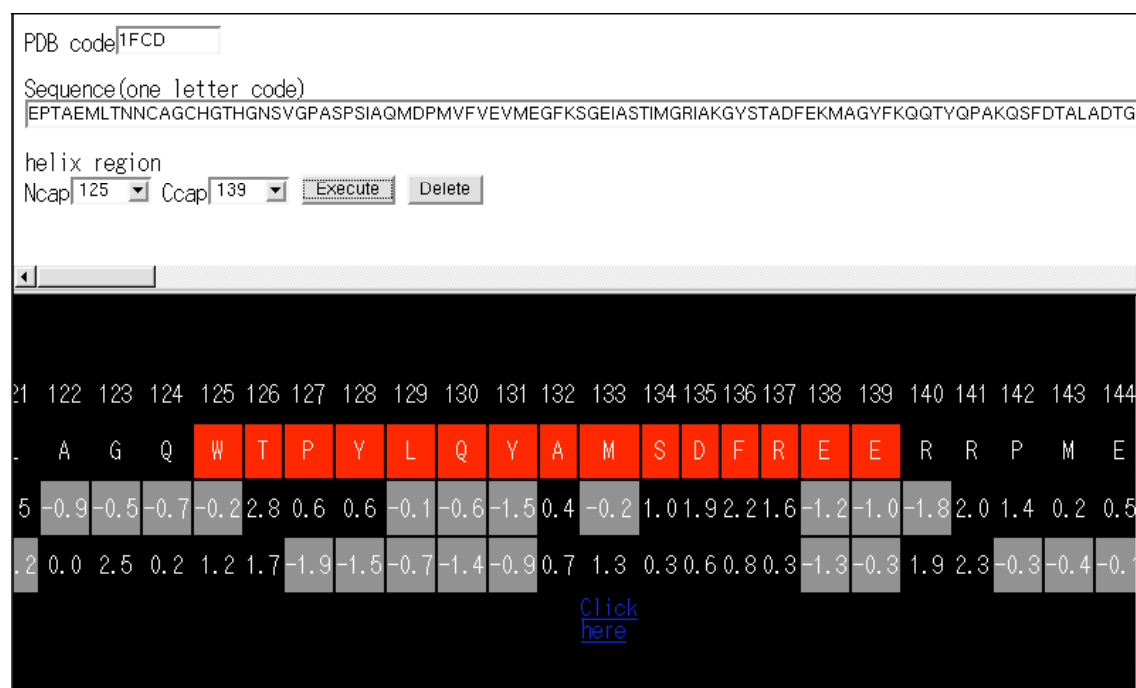


Figure 3-6. 蛋白質の二次構造を設計するためのプログラム

スコアを見ると，N 末端にある 125 番目のトリプトファン残基の $N\alpha$ スコアが-0.2 であり，0 より低いスコアである。また，C 末端にある 139 番目のグルタミン酸残基も-0.3 のように 0 より低いスコアである。つまり，トリプトファン残基の $N\alpha$ スコアやグルタミン酸残基の $C\alpha$ スコアを 0 より高くするように，それぞれの境界領域に含まれるアミノ酸残基のどれかを別のアミノ酸残基に置換する。この例では，N 末端は，配列 GQWTPYL のうちの，Ncap 部位のトリプトファン残基の $N\alpha$ スコアを低くしている原因のアミノ酸残基を置換の対象とし，C 末端は，配列 DFREERR のうちの，Ccap 部位のアルギニン残基の $C\alpha$ スコアを低くしている原因のアミノ酸残基を置換の対象とする。つぎに，置換の対象であるアミノ酸残基を見るために，青文字の「Click here」をクリックする。すると，Richardson パラメータを表示した別ウインド

ウが開く (Figure 3-7)。

	N ^α	N ¹	Ncap	N1	N2	N3	N4		C4	C3	C2	C1	Ccap	C ¹	C ^α
A	0.9	0.7	0.4	1.1	1.3	1.2	1.4		1.7	1.4	1.4	1.2	0.6	0.8	0.8
C	1.1	1.1	0.9	0.6	0.6	0.8	1.0		1.4	0.9	0.8	1.2	1.2	0.8	0.7
D	0.9	0.9	2.7	0.9	1.7	1.5	0.5		0.6	0.7	1.0	0.6	0.9	1.1	1.6
E	1.0	0.8	0.6	1.3	2.5	2.0	0.6		0.9	1.2	1.6	1.0	0.6	0.8	1.1
F	1.0	1.1	0.4	0.9	0.7	1.0	1.4		1.3	1.1	0.7	1.1	1.1	0.9	0.8
G	1.5	1.1	1.7	0.5	0.7	0.6	0.8		0.4	0.8	0.3	0.4	3.7	1.1	1.1
H	1.1	1.0	1.2	0.8	1.0	1.0	0.7		0.8	1.0	0.9	1.4	1.3	0.8	1.2
I	0.9	1.2	0.2	0.8	0.5	0.7	1.6		1.2	1.1	0.8	0.7	0.7	0.9	0.8
K	0.9	0.8	0.5	0.9	1.0	0.9	0.9		0.8	1.4	1.8	1.3	0.9	1.2	1.2
L	0.8	1.2	0.3	1.0	0.6	0.9	1.8		1.9	1.6	1.1	1.5	0.8	0.8	0.7
M	0.8	1.4	0.3	0.9	0.6	0.9	1.4		1.7	1.3	1.2	1.3	0.8	0.8	0.6
N	1.1	0.9	2.3	0.5	0.9	0.7	0.5		0.5	0.7	0.9	1.3	1.6	1.0	1.1
P	1.3	1.3	1.4	3.2	0.5	0.3	0.2		0.4	0.2	0.3	0.0	0.0	3.0	1.9
Q	0.9	0.9	0.5	0.9	1.4	1.7	1.1		1.0	1.2	1.3	1.3	0.9	0.8	0.9
R	1.0	0.8	0.5	0.9	1.0	0.9	1.4		0.9	1.4	1.5	1.2	0.8	0.9	1.0
S	1.1	0.9	2.6	0.7	1.3	1.0	0.5		0.7	0.8	1.1	1.1	0.8	0.9	1.1
T	1.0	1.0	2.0	0.8	0.8	1.0	0.7		0.6	0.6	0.7	1.1	0.7	0.9	1.0
V	0.9	1.1	0.2	0.9	0.5	0.9	1.4		0.9	0.8	0.6	0.6	0.6	0.7	0.6
W	0.7	0.9	0.4	1.5	1.2	1.0	1.4		1.7	1.1	0.8	0.8	0.5	0.7	0.8
Y	1.0	1.0	0.6	1.0	0.8	0.9	1.2		1.1	0.9	0.7	1.2	1.0	0.9	0.9

Figure 3-7. アミノ酸置換候補

Figure 3-7 の横の列にアミノ酸残基の傾向指数を、縦の列にヘリックス末端領域のポジションを示す。つまり、Richardson パラメータを出力している。青色の傾向指数は、各ポジションにおいて、もっとも小さい傾向指数、赤色の傾向指数は各ポジションにおいて、もっとも大きい傾向指数、黄色は実際のヘリックスの末端領域に存在したアミノ酸残基の傾向指数である。緑色は、実際のヘリックスの末端領域に存在したアミノ酸残基が、そのポジションの中でもっとも高い傾向指数であったときの値である。オレンジ色は、実際のヘリックスの末端領域に存在したアミノ酸残基が、そのポジションの中で最も低い傾向指数であったときの値である。円で囲っている部分は、実際のアミノ酸残基の傾向指数である。まず、N 末端領域の方から見ると、N 末端領域の中で Ncap 部位の N α スコアを低下させている原因のアミノ酸残基は、1.0 より小さい値をもつ N¹ 部位のグルタミン酸残基と N2 部位のメチオニン残基と N3 部位のバリン残基

の三つである。これらの中でも、N2 部位のメチオニン、もっとも低い傾向指数 0.6 であるため、N2 部位でもっとも高い傾向指数 2.5 をもつグルタミン酸残基へ置換すると、Ncap 部位の N α スコアは、高い値 (=1.7) を示す。一方で、C 末端領域の中で Ccap 部位の C α スコアを低下させている原因のアミノ酸残基は、1.0 より小さな値をもつ C'部位のイソロイシン残基、C'部位のグルタミン酸残基、C4 部位のグリシン残基の三つである。これらの中でも、C4 部位のグリシン残基は、もっとも低い傾向指数 0.4 であるので、C4 部位で最も高い傾向指数 1.9 をもつロイシン残基へ置換することで、Ccap 部位の C α スコアを高くすることができる。ただし、ここでは、アミノ酸残基の置換候補の中から、傾向指数がもっとも小さな値をもつアミノ酸残基を選んだが、必ずもっとも低い値を選ぶ必要はない。

以上述べてきたように、この章では、従来の二次構造予測法では曖昧だった α ヘリックス構造の末端部位の予測をするために、ProSSA 法で帰属した α ヘリックス構造の末端領域のアミノ酸残基の傾向指数である Richardson パラメータを使って、予測したいアミノ酸配列に対して、7 残基一組のウィンドウによるスコアを算出し、そのスコアを α ヘリックス構造の形成傾向スコアと見なし、末端部位の予測を可能にした。さらに、 α ヘリックス構造末端の予測法を実際の蛋白質の配列に対して実施し、得られたスコアを不安定領域を検索するためのスコアと見なし、 α ヘリックス構造の末端の不安定因子の候補を探すための方法として使った。不安定領域の検索方法としては、研究者や技術者の誰でも使えるようにアプリケーションを開発した。この章でも、ProSSA 法から得られた結果が、蛋白質設計の分野への応用が可能であることを示した。

第4章 蛋白質 CD スペクトルからの二次構造含量の評価

4-1 はじめに

円偏光二色性 (Circular Dichroism; CD) 測定は、X 線結晶解析法や NMR 解析法よりも、容易に蛋白質の構造を解析できる。特に自然界や生体内に存在する実際の蛋白質の環境に近い状態で観測できるのが特徴である。もっとも、CD 測定法は X 線結晶解析法や NMR 解析法と異なり、蛋白質を構成している原子の座標データを得ることはできない。しかしながら、CD 測定法は、蛋白質のそれぞれの二次構造が占める割合、野生型と変異体の構造の違い、温度変化による構造変化、有機溶媒に対する蛋白質の影響などを、時間をかけることなく調べることができる。このような手軽さもあって、CD 測定法は蛋白質の構造を調べるのに多くの研究者が利用している。

CD を使って蛋白質のを解析するのに用いる波長領域は、蛋白質のどの部分に着目して解析するかによって異なるが、この章では、蛋白質の主鎖のコンホメーションに着目し、CD スペクトルと二次構造含量の関係を調べるため、遠紫外 (far-UV) 領域 190 から 260 nm を使った。なぜならば、far-UV 領域は、主鎖を構成しているアミドの吸収帯であるため、コンホメーションがとり得る形がスペクトルに反映されるためである。

蛋白質の CD スペクトルから二次構造含量を推定する研究は、CD スペクトルの極値に注目した方法と、蛋白質 CD スペクトル全体を使った方法に分けられる。まず、これらの研究の概要について述べる。

(1) CD スペクトルの極値を使う方法

Greenfield と Fasman は、ポリ-L-リジンを使って、三つの基本となる CD スペクトルを測定し、そのデータを用いて二次構造含量を推定した。そのために、まず、ヘリックス含量を、CD スペクトルの 208 nm の極値に注目して、次の式から導き出した¹²⁾。

$$\% \alpha \text{ helix} = \frac{[\theta]_{208 \text{ nm}} - 4,000}{33,000 - 4,000} \quad (4-1)$$

このとき得られたヘリックス含量を固定して、ベータ含量やランダム含量を変えていき、実際に測定して得られた CD スペクトルと一致するような値を探していった。この方法は、あとに述べる式(4-2)を前提としている。しかしながら、この手法は、合成ポリペプチドを使っている点に問題があり、X 線結晶データとは異なる結果を与えた。

Chen と Yang は、CD スペクトルの 222 nm の極値に注目して、次の式を考案した⁵⁶⁾。

$$[\theta]_{222} = -30300f_H - 2340$$

しかしながら、彼らは、ベータ構造については関係式を見つけきれなかった。ただし、これらの含量推定法は、ヘリックス含量を推定するときに使われる場合がある。

(2) 蛋白質の CD スペクトル全体を使った方法

CD スペクトルの特定の波長に着目せずに、CD スペクトルで測定した波長領域全体を使って二次構造含量を推定する方法がある^{13-25,57)}。この方法は、構造既知の蛋白質の CD スペクトルと X 線結晶データから得られた二次構造含量を使って、それぞれの二次構造含量が 100% のときの CD スペクトル、つまり参照 CD スペクトルを算出し測定した CD スペクトルから二次構造含量を推定している。くわえて、X 線結晶解析データから二次構造含量を得るための二次構造帰属法としては、水素結合の様式パターンで帰属する DSSP 法を利用している。DSSP 法は第 1 章で述べたように主鎖のコンホメーションに規則性がみられない部分も二次構造と帰属するという問題がある。また、波長領域として、真空紫外領域まで拡張して解析する方法があるが、真空紫外領域まで測定できる装置が特殊であるため、誰でも使えるわけではない。そして、一般に普及している CD 測定装置は遠紫外領域まで測定でき、しかも、主鎖を構成しているアミドの吸収帯であることを考えると、波長領域としては、遠紫外領域までの測定したデータを使う場合が多い。したがって、この章では、遠紫外領域のデータと二次構造含量の関係を調べることとした。

4—2 CD スペクトルと二次構造の関係

蛋白質の CD スペクトルと二次構造含量の相関をみると、二つの問題があげられる。

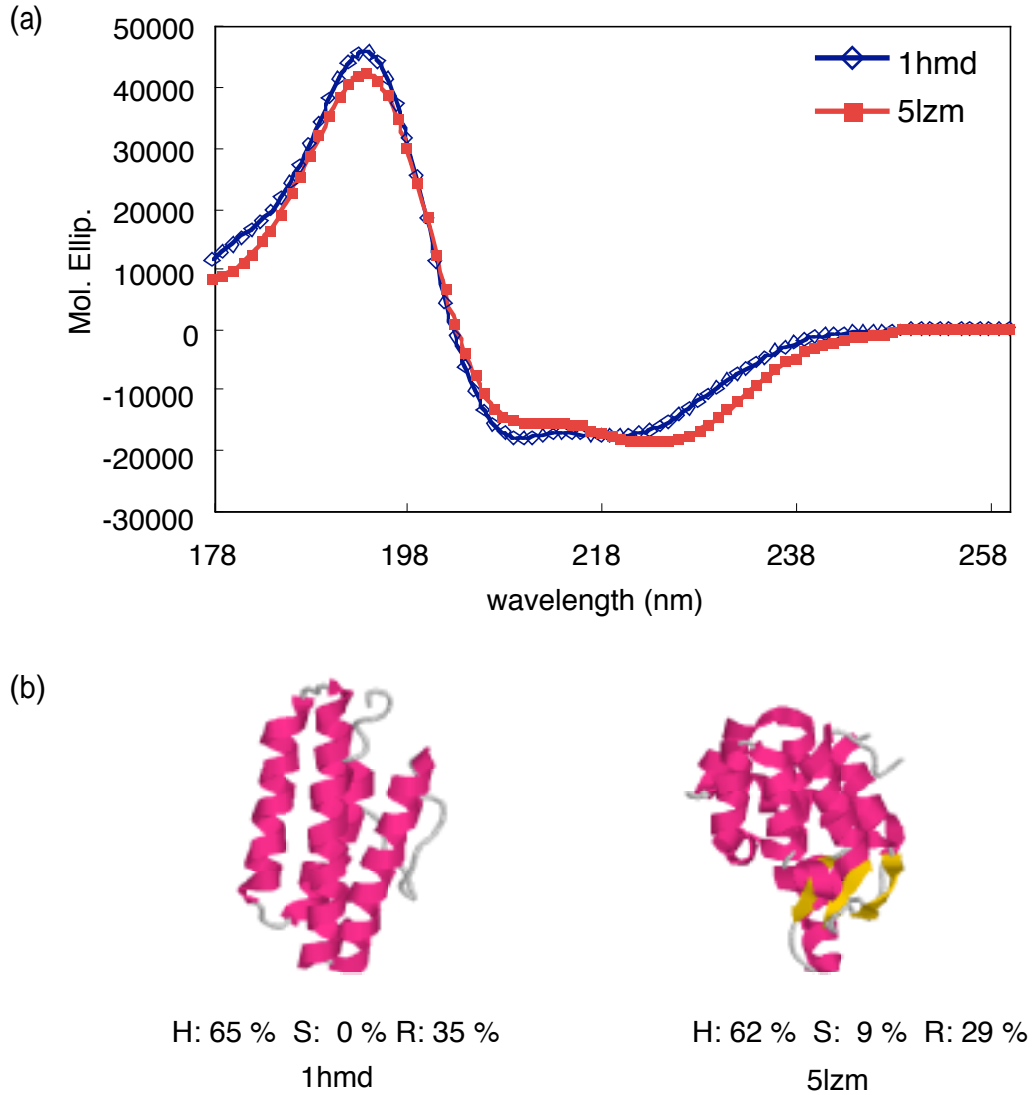


Figure 4-1. ヘムエリトリン (1hmd) とT4リゾチーム (5lzm)の CDスペクトルと二次構造含量

一つは、同じ CD スペクトルでありながら二次構造含量が異なるという問題である (Figure 4-1)。Figure 4-1 は、ヘムエリトリン (1hmd) と T4 リゾチーム (5lzm) の CD スペクトル (Figure 4-1 (a)) と、PDB ファイルから DSSP 法によって得られた二次構造含量 (Figure 4-1 (b)) である。

「H」が α ヘリックス含量, 「S」が β ストランド含量, 「R」がその他 (α ヘリックスでも β ス

トランドでもない構造) の含量である。これら二つの蛋白質の CD スペクトルは、同じスペクトルを描いている。しかしながら、二次構造含量は、1hmd で α ヘリックス含量が 65%, β ストランド含量が 0%, その他の含量が 35%の値を得ている。5lzm では、それぞれ 62%, 9%, 29% である。つまり、二次構造含量は、たがいに異なる値を示しており奇異な印象を与える。

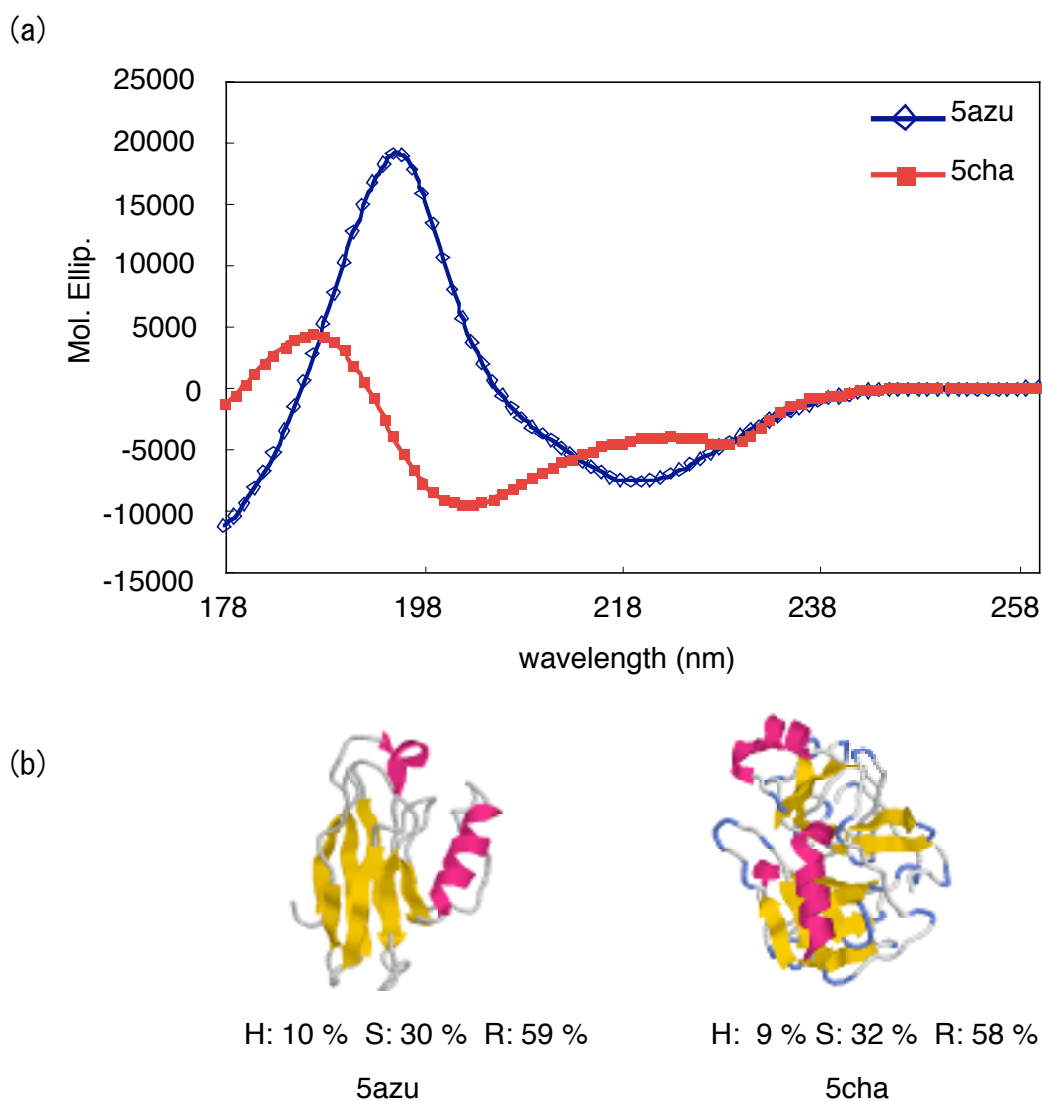


Figure 4-2. アズリン (5azu) と α キモトリプシン (5cha) の CD スペクトルと二次構造含量

二つ目は、異なる CD スペクトルでありながら二次構造含量が同じという問題である (Figure 4-2)。Figure 4-2 は、アズリン (5azu) と α キモトリプシン (5cha) の CD スペクトル (Figure 4-2

(a) と PDB ファイルから DSSP 法を使って得られた二次構造含量 (Figure 4-2 (b)) である。二つの蛋白質の CD スペクトルは異なるスペクトルを描いている (Figure 4-2 (a))。しかしながら、二次構造含量の値は、5azu では α ヘリックス含量が 10%, β スtrand 含量が 30%, その他の含量が 59% であり、5cha は、それぞれ 9%, 32%, 58% である。つまり、二次構造含量は同じ数値を示しているものの、異なる CD スペクトルを与えている。

蛋白質の CD スペクトルと二次構造含量は、次のような関係がある：

$$P(\lambda) = f_H \cdot H(\lambda) + f_B \cdot B(\lambda) + f_R \cdot R(\lambda) \quad (4-2)$$

$P(\lambda)$ ：波長 λ のときの蛋白質の平均残基モル橢円率

$H(\lambda)$ ：波長 λ のときの α ヘリックス構造の平均残基モル橢円率

$B(\lambda)$ ：波長 λ のときの β ストランド構造の平均残基モル橢円率

$R(\lambda)$ ：波長 λ のときのその他の構造の平均残基モル橢円率

f_H, f_B, f_R ：二次構造成分比 ($0 \leq f_H \leq 1, \sum f_i = 1$)

式 (4-2) は、ある波長 λ での蛋白質の残基モル橢円率 $P(\lambda)$ は、その波長における α ヘリックス構造の残基モル橢円率 $H(\lambda)$ と α ヘリックス構造の構成成分比 f_H の積、 β スtrand 構造の平均残基モル橢円率 $B(\lambda)$ と β スtrand 構造の構成成分比 f_B の積、その他の構造の平均残基モル橢円率 $R(\lambda)$ とその他の構造の構成成分比 f_R の積の総和である。つまり、4-1 節の蛋白質 4enl と 6ldh, 1fec と 1ova に対する CD スペクトルと二次構造含量を式 (4-2) にあてはめると、蛋白質 CD スペクトルと二次構造含量はマッチングしない。この原因として、二次構造帰属法に問題があると考えられる。水素結合様式のパターンで二次構造を帰属する DSSP 法は、適切ではないと考えられるため、ProSSA を使って蛋白質の CD スペクトルと二次構造の関係を調べた。

4—3 CD スペクトルと ProSSA 法による二次構造含量の評価

蛋白質 CD スペクトルと二次構造含量の関係をみるためには、すでに立体構造がわかっている、すなわち PDB に登録された蛋白質の CD スペクトルデータが必要である。この節では、インターネット上で公開されている二次構造含量推定プログラム CDsstr に含まれている 26 種類の蛋白質を使った⁵⁸⁾。なぜならば、CDsstr で使われているデータは、CD スペクトルと PDB コードの対応づけがなされている。さらに、モル吸光係数の差 $\Delta \epsilon$ を平均残基モル楕円率 $[\theta]$ に変換するために、オリジナルデータに 3300 を積算した。この 26 種類の蛋白質に対して ProSSA 法によって二次構造を帰属し、総残基数、 α ヘリックス構造の残基数と含量、 β ストランド構造の残基数と含量、その他の構造 (α ヘリックス構造でも β ストランド構造でもない構造)の残基数と含量を算出した (Table 4-1)。平均値は、全残基数が 228 個、 α ヘリックス含量が 38%、 β ストランド含量が 41%、その他の含量が 23%であった。また、CD スペクトルと二次構造含量の相関を見るには、対照蛋白質が必要である。そこで、二次構造含量の平均値に近いトリパノチオンリダクターゼ (1fec) (α ヘリックス含量: 40%、 β ストランド含量: 43%、その他の含量: 20%) を対照蛋白質として、残りの 25 種類の蛋白質と比較した。その結果、190 nm 近傍にて、四つの特徴が見られた。その特徴を、ミオグロビン (1vxf)、コンカナバリン (1nls)、エノラーゼ (4enl)、アズリン (5azu) の四つの蛋白質を比較蛋白質として使用する (Figure 4-3 ~4-6)。

Figure 4-3 には、対照蛋白質 1fec と比較蛋白質 1vxf の CD スペクトルと、PDB ファイルから ProSAA 法を使って得られた二次構造含量を示している。1vxf の CD スペクトルは、190 nm 付近の正の極値、208 nm と 222 nm 付近の負の極値に特徴がある。そして、1fec は、190 nm 付近の正の極値に特徴がある。そして、1vxf の正の極値の方が、1fec よりも高い位置にある。一方、二次構造含量を比べると、ヘリックス含量は、1vxf の方が 1fec よりも 2 倍以上大きく、ストランド含量は、逆に、1fec の方が 1vxf よりも 10 倍以上大きい。以上のことから、ヘリックス含量が、対照蛋白質 1fec よりも、比較蛋白質の方が大きければ、190 nm 付近の正の極値は、1fec よりも比較蛋白質の方が高かった。なお、ストランド含量が、対照蛋白質の方が大きいときの CD スペクトルの条件は、見つからなかった。

Table 4-1. ProSSAによる26種類の蛋白質の二次構造情報

	全残基数	ヘリックス		ストランド		その他	
		残基数	含量 (%)	残基数	含量 (%)	残基数	含量 (%)
1vxf	153	144	94	4	3	11	7
2mhb	287	253	88	14	5	32	11
1hmd	113	89	79	13	12	13	12
5cyt	103	59	57	32	31	18	17
1beb	156	34	22	88	56	40	26
3psg	321	64	20	184	57	81	25
5azu	128	22	17	69	54	39	30
1tld	220	25	11	130	59	67	30
5cha	236	26	11	133	56	79	33
1esa	240	23	10	142	59	77	32
2pab	114	10	9	76	67	29	25
1sxn	151	13	9	83	55	55	36
1nls	237	6	3	169	71	64	27
1rei	107	0	0	76	71	31	29
1ypi	247	130	53	81	33	45	18
5fx2	147	67	46	52	35	29	20
1oya	399	173	43	121	30	111	28
5lzm	162	133	82	22	14	16	10
4lzt	129	70	54	28	22	37	29
1lnf	316	140	44	90	28	88	28
9pap	212	75	35	85	40	58	27
1rbx	124	41	33	68	55	21	17
6ldh	329	167	51	102	31	66	20
4enl	436	208	48	124	28	109	25
1fec	485	193	40	207	43	98	20
1ova	366	138	38	165	45	72	20
平均値	228	89	38	91	41	53	23

つぎに、対照蛋白質 1fec と比較蛋白質 1nls を使って CD スペクトルと二次構造含量の関係を
示す (Figure 4-4)

対照蛋白質 1fec の CD スペクトルの特徴は、190 nm 付近の正の極値である。1nls の CD スペ
クトルの特徴も 190 nm 付近の正の極値である。そして、1nls の CD スペクトルは、190 nm 付

近の正の極値が 1fec よりも低い。一方、二次構造含量は、ストランド含量において、1nls の方が、1fec よりも多い。つまり、比較蛋白質のストランド含量が対照蛋白質 1fec よりも多ければ、190 nm 付近の正の極値は対照蛋白質よりも低くなる。すなわち、対照蛋白質の CD スペクトルよりも 190 nm 付近の強度が増える。同様に、対照蛋白質 1fec と比較蛋白質 5azu について調べた (Figure 4-5)。CD スペクトルは、1fec、5azu とともに、190 nm 付近に正の極値をもち、強度は、5azu の方が減少していた。一方、二次構造含量において、ストランド含量は、5azu の方が 1fec よりも多く、その他の含量も 5azu の方が多い。つまり、対照蛋白質に対して、ストランド含量が多く、その他の含量也多ければ、190 nm 付近の正の極値は、減少する。つぎに、対照蛋白質 1fec と比較蛋白質 4enl について CD スペクトルと二次構造の関係を調べた (Figure 4-6)。1fec と 4enl は、ともに CD スペクトルが 190 nm 付近に正の極値をもち、強度は、4enl の方が強い。一方、二次構造含量に対して、ストランド含量は、4enl の方が少なく、その他の含量は 4enl の方が多い。つまり、対照蛋白質に対して、比較蛋白質のストランド含量の方が少なく、その他の含量の方が多い場合、190 nm 付近の正の極値は高くなるということが言える。

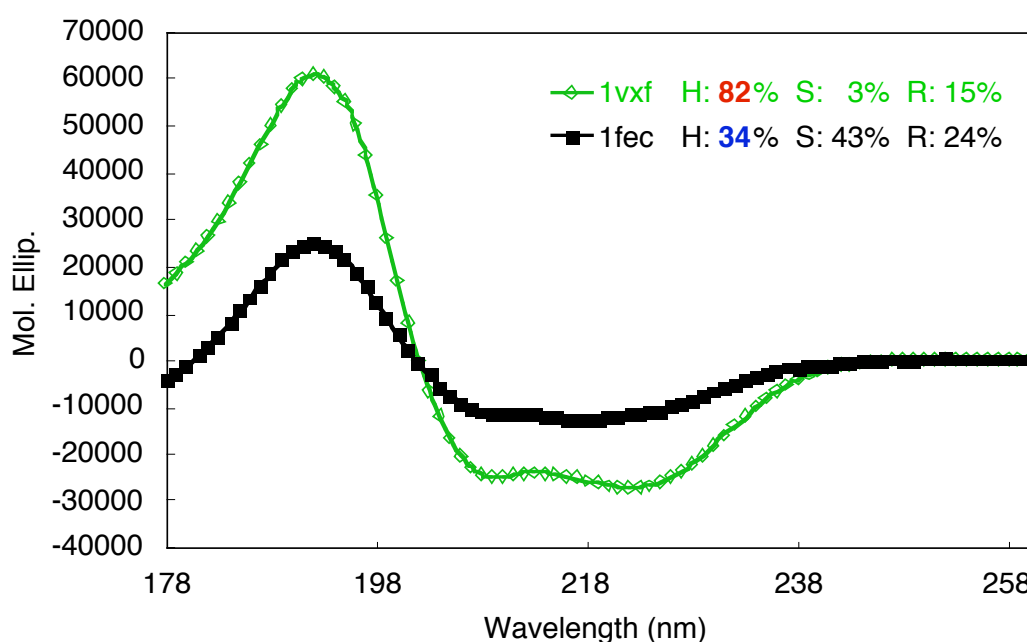


Figure 4-3. 対照蛋白質1fecと比較蛋白質1vxfの関係

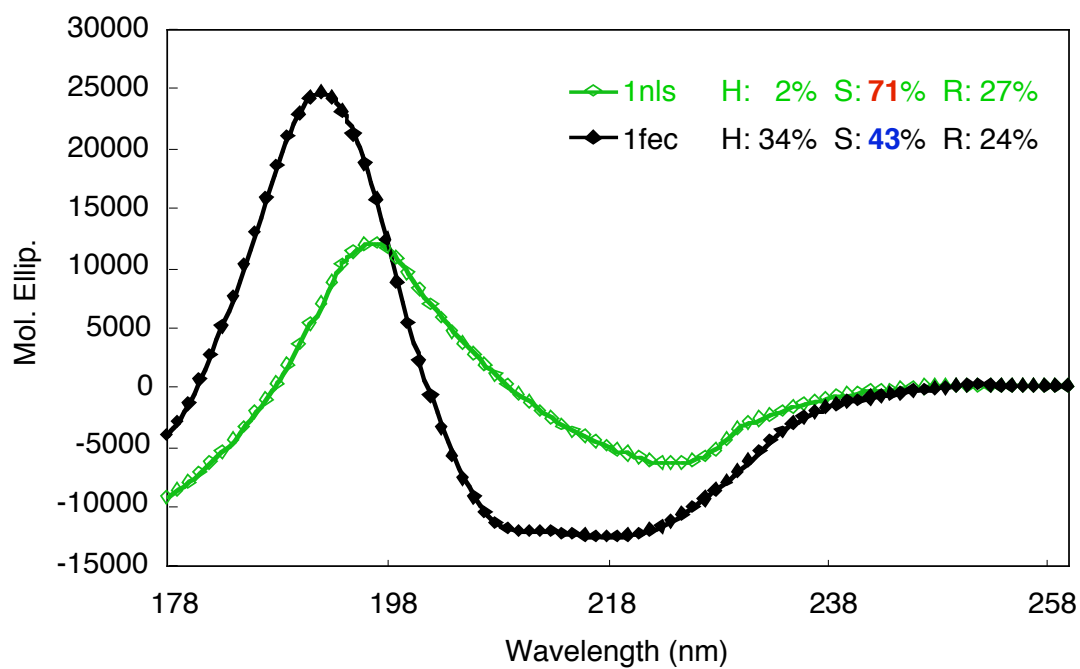


Figure 4-4. 対照蛋白質1fecと比較蛋白質1nlsの関係

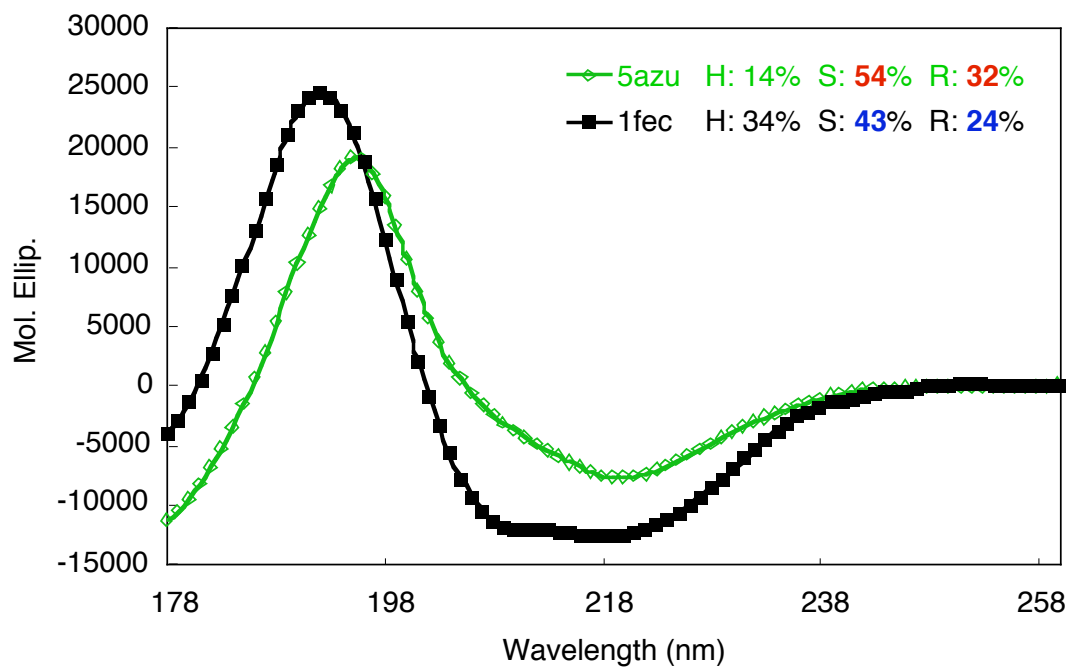


Figure 4-5. 対照蛋白質1fecと比較蛋白質5azuの関係

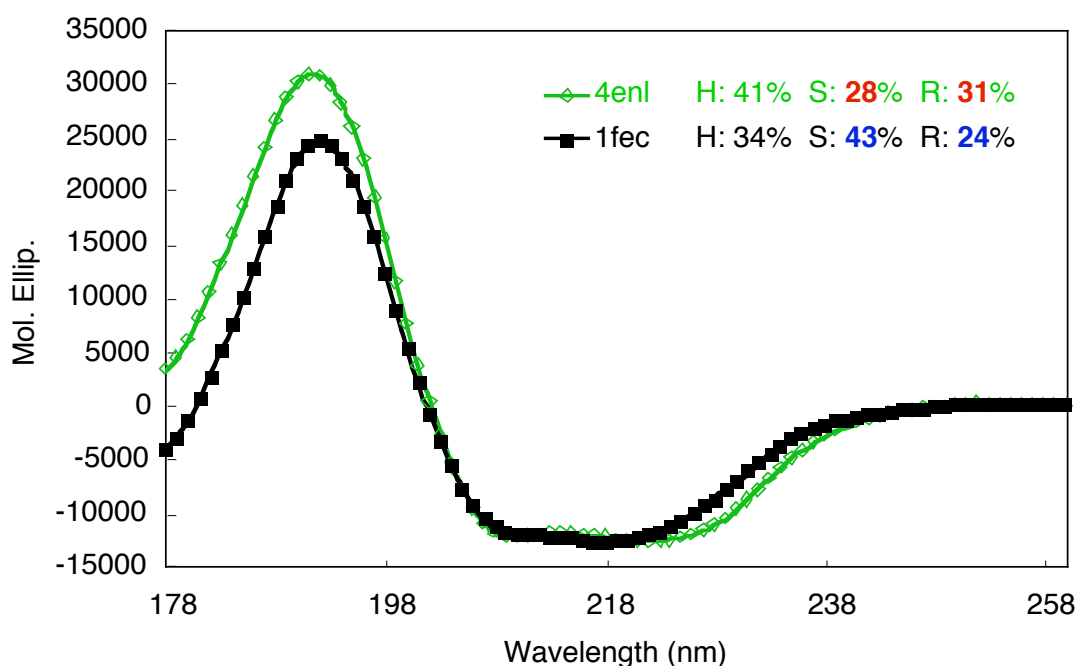


Figure 4-6. 対照蛋白質1fecと比較蛋白質4enlの関係

以上をまとめると、対照蛋白質 1fec の CD スペクトルと二次構造含量に対して、一つ目の特徴は、ヘリックス含量が多ければ 190 nm 付近の極値が高くなる。二つ目の特徴は、ストランド含量が多ければ、190 nm 付近の極値が低くなる。三つ目の特徴は、ヘリックス含量とその他の含量が多ければ、190 nm 付近の極値が高くなる。四つ目の特徴は、ストランド含量とその他の含量が多ければ、190 nm 付近の極値が低くなる。これら四つの特徴をもつ蛋白質は、25 蛋白質のうち 20 個であった。つまり、80%の蛋白質に相関があった。

二次構造帰属法として DSSP を使って構造既知の 26 種類の蛋白質の二次構造の含量を求め、先ほどと同じように CD スペクトルと二次構造含量の関係を調べた。すると、14 種類の蛋白質と相関があった。つまり、56%の蛋白質に相関があった。

ところで、相関が見られなかった蛋白質は 5 つであった (Figure 4-7)。相関が見られなかった理由は、X 線結晶構造解析のときの構造と水中での構造が異なるためと考えられる。

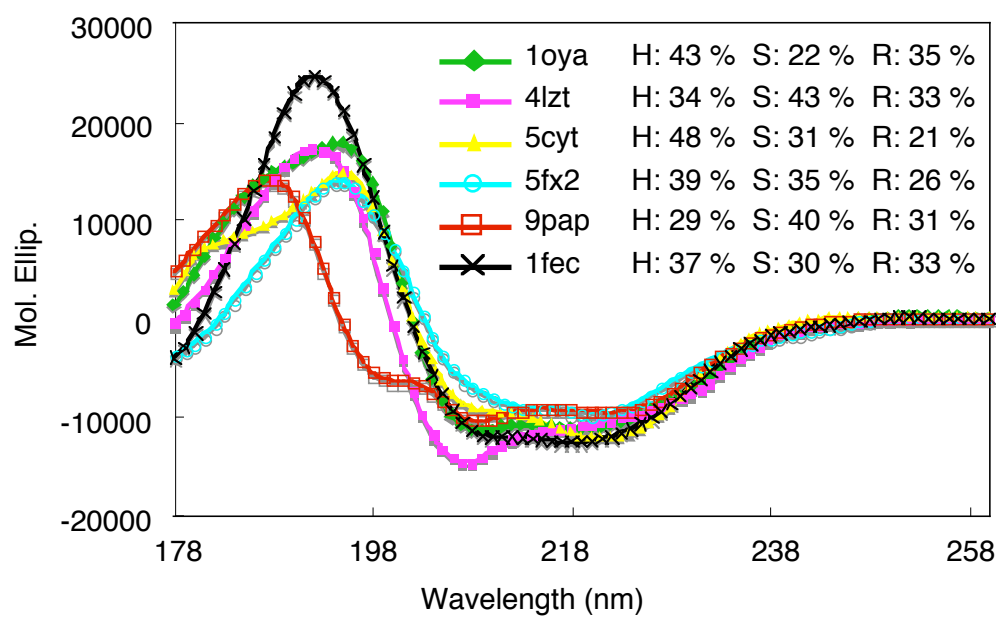


Figure 4-7. 相関が認められなかった五つのCDスペクトル

4-4 参照 CD スペクトルの算出

4-3節で相関が見られた 21 種類の蛋白質は PDB ファイルが存在する。したがって、CD スペクトルデータと PDB ファイルから得られた二次構造含量、そして式 (4-2) を使ってそれぞれの二次構造の CD スペクトルを最小二乗法を使って算出した (Figure 4-8)。

■で示す赤線はヘリックス含量を 100%, ●で示す青線はストランド含量を 100%, ×で示す黒線はその他の含量を 100%とみなした CD スペクトルを意味する。この CD スペクトルを参照 CD スペクトルとし、蛋白質 CD スペクトルから二次構造含量を推定するときの基礎データとして用いられる。

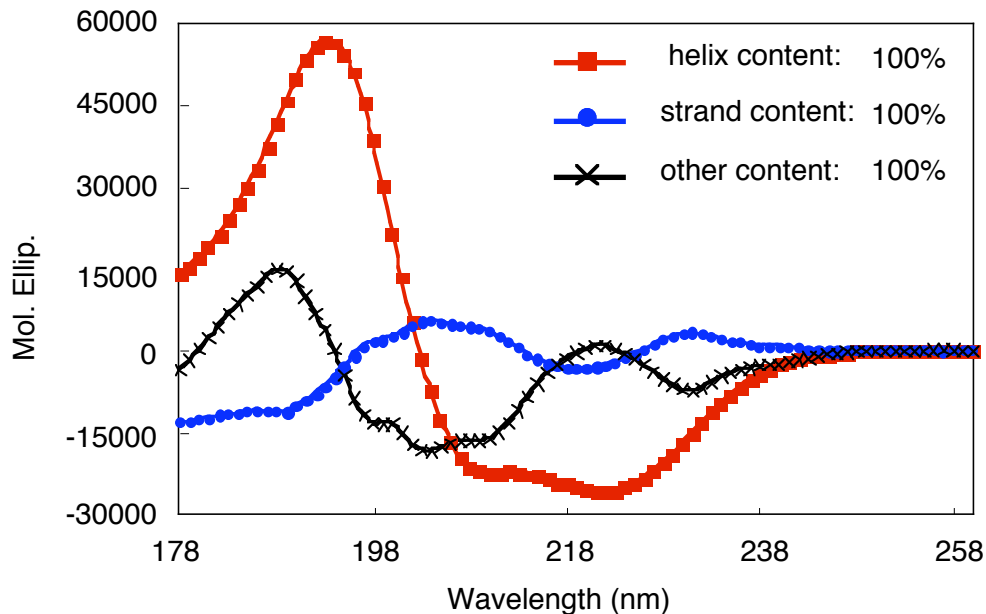


Figure 4-8. 21蛋白質のCDスペクトルから算出した参照CDスペクトル

つぎに、今回の研究で得た参照 CD スペクトルと構造がわかっている蛋白質の二次構造含量を使って、実際に測定した蛋白質の CD スペクトルとの整合性について調べた。

対象の蛋白質は、X 線結晶データが存在するホスホトリエステラーゼ (PTE) であり、この蛋白質は、 α/β 型の蛋白質である (Figure 4-9)。

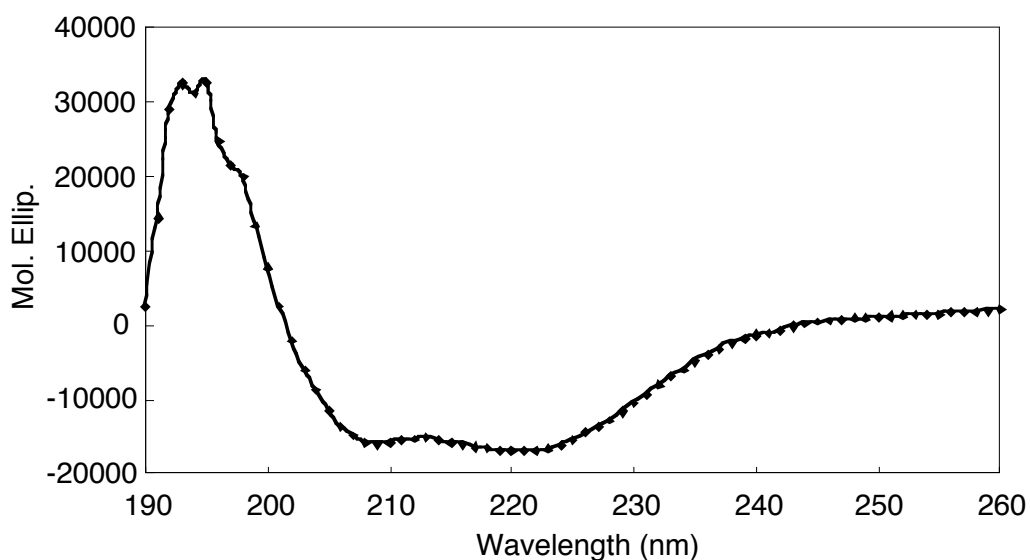
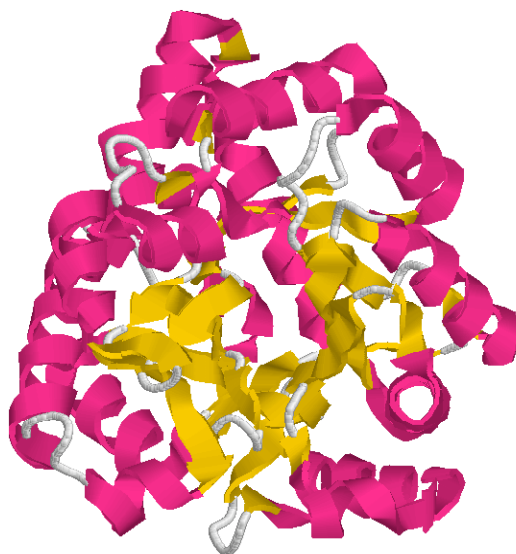


Figure 4-9. PTEのCDスペクトルと立体構造（1DPM）

PTE の PDB ファイル（1DPM）から二次構造帰属法 ProSSA を用い二次構造含量を求め、この値と参照 CD スペクトル、つまり式 (4-1) によって再構成した CD スペクトルを得た。これを実際に測定された CD スペクトルと比較した (Table 4-2; Figure 4-10)。なお、比較のため、DSSP 法による二次構造含量も求めた。

■ で示す赤色の線「Obs」は、実際の蛋白質の CD 測定から得たスペクトルであり、○で示す青色の線「ProSSA」と×で示す緑色の線「DSSP」は、PDB ファイルから二次構造帰属法 ProSSA

と DSSP を使って算出した二次構造含量を参照 CD スペクトルと組み合わせて再構成したスペクトルである。Obs と ProSSA の CD スペクトルは、ほとんど同じスペクトルである。そして、Obs と ProSSA の平均二乗偏差の平方根は 0.09 であり、相関係数は 1.00 を得た。この結果から、この研究で算出した参照 CD スペクトルは有用であると言える。

Table 4-2. PDBファイルから算出したPTEの二次構造含量

Methods	Secondary Structure Contents (%)			
	alpha	beta	turn	other
ProSSA	57	29	-	17
DSSP	43	13	-	45

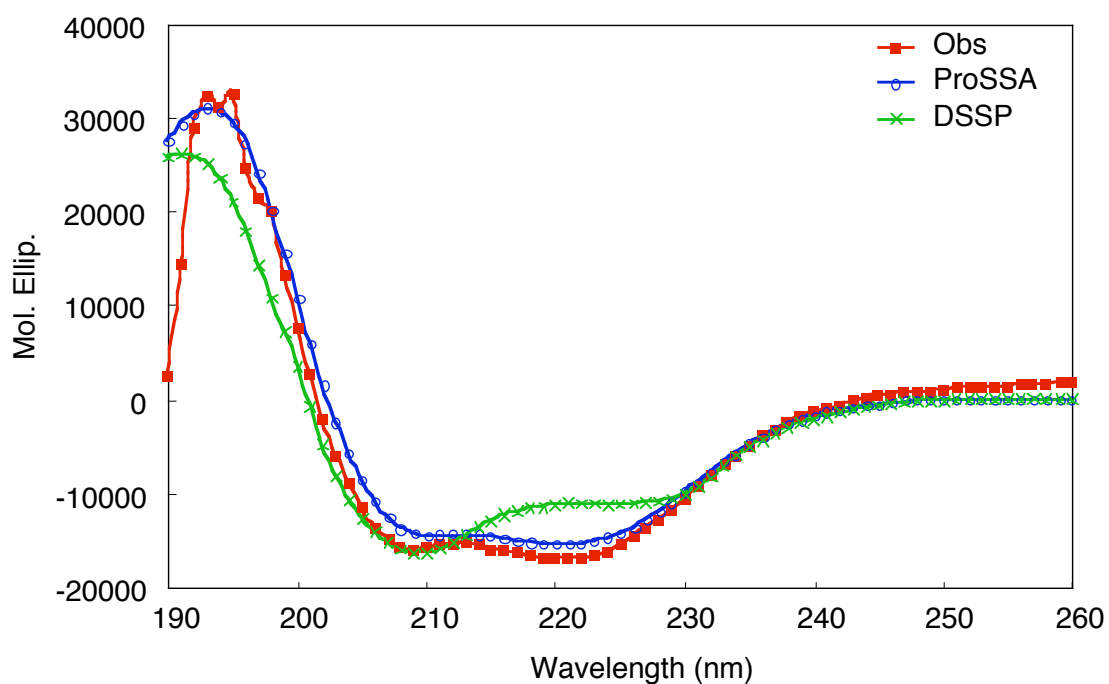


Figure 4-10. PDBファイルから得た二次構造含量を使って再現したCDスペクトルと実際に観測したCDスペクトル (PTE)

一方、DSSP は、210 nm 付近の強度が他の二つのスペクトルに比べて低く、そして、220 nm 付近に正の極値があり、明らかに異なるスペクトルパターンである。なお、Obs と DSSP の平均二乗偏差の平方根は 0.23、相関係数は 0.98 の値を得た。

この研究で得られた参照 CD スペクトルに対して、SSE-338W 型の蛋白質二次構造解析プログラム⁵⁹⁾で使われている参照 CD スペクトルは、Yang らの参照 CD スペクトルであり、ヘリックス CD スペクトルの構成アミノ酸残基を n=10 としている (Figure 4-11)⁵⁷⁾。これらの参照 CD スペクトルを比べると、ヘリックス含量が 100% のときの CD スペクトル (ヘリックス CD スペクトル) は、190 nm 付近の強度が異なっていた (Figure 4-12)。

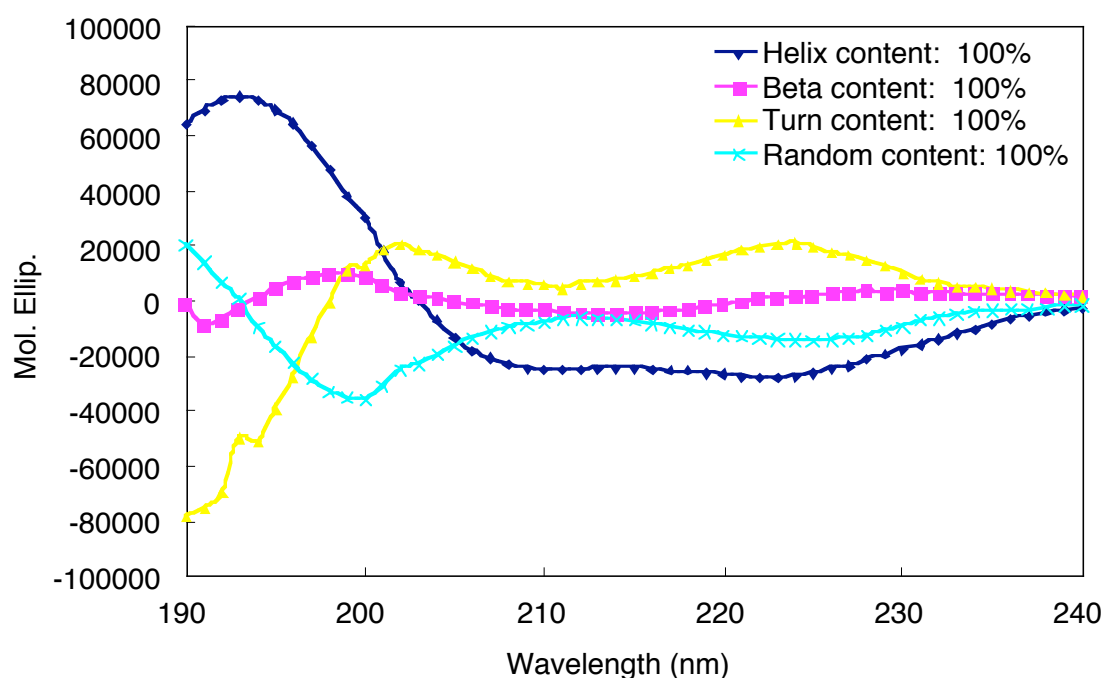


Figure 4-.11 SSEプログラムで使われているYangらの参照CDスペクトル

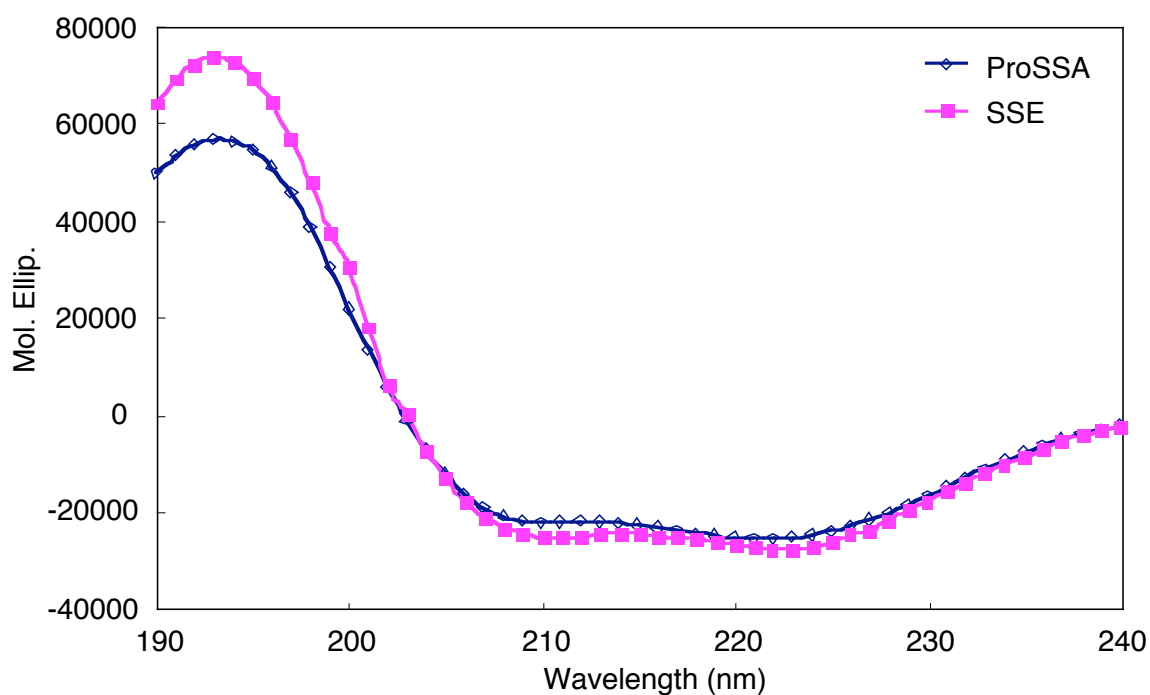


Figure 4-12. ProSSAによるヘリックスCDスペクトルと
SSEに組み込まれているヘリックスCDスペクトル

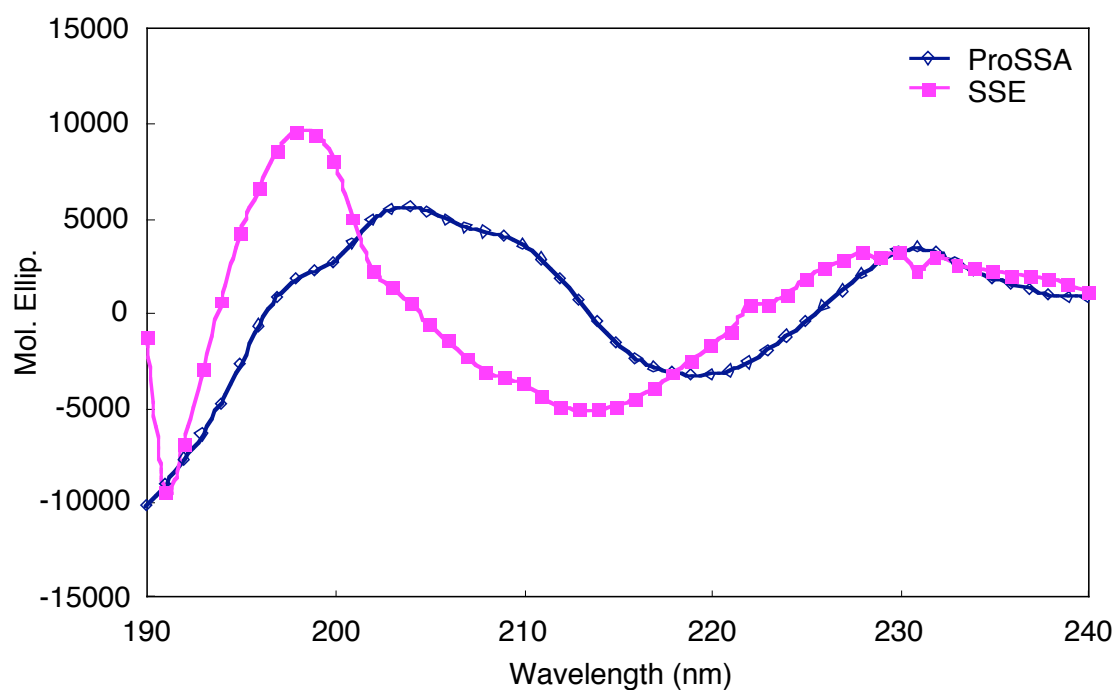


Figure 4-13. ProSSAによるストランドCDスペクトルと
SSEに組み込まれているストランドCDスペクトル

さらに、ストランド含量が100%のときのCDスペクトル（ストランドCDスペクトル）を比べると、ProSSAのストランドCDスペクトルが低い値を示していて、全体的に右にシフトしている（Figure 4-13）。

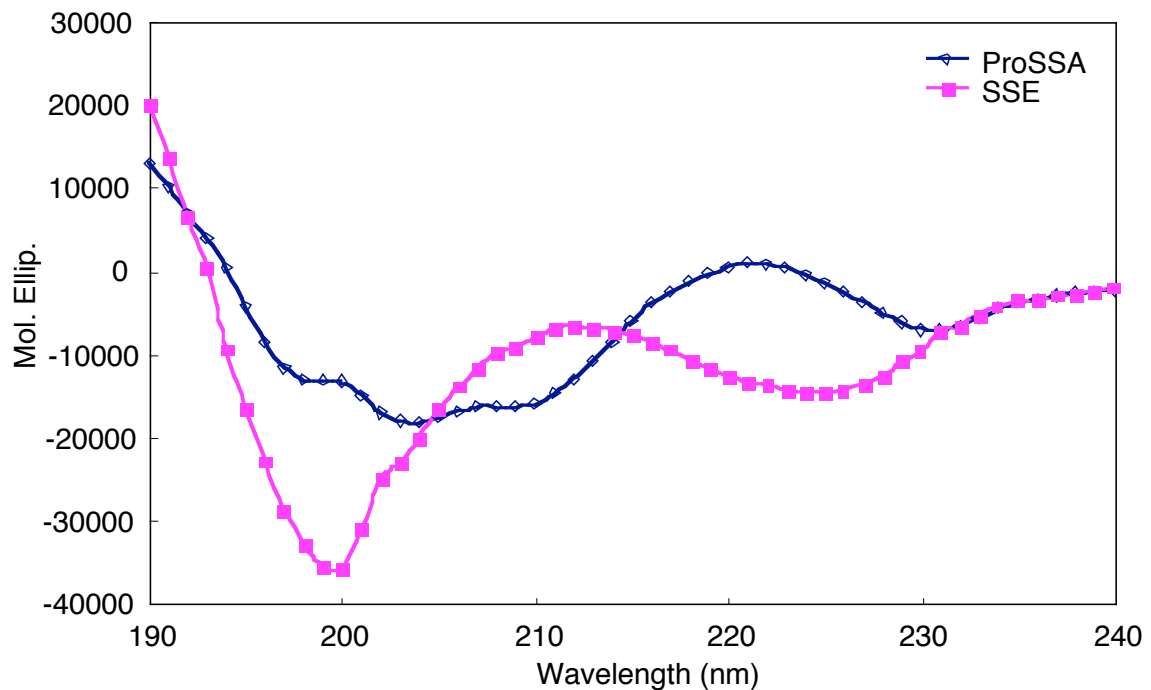


Figure 4-14. ProSSAによるその他のCDスペクトルと
SSEに組み込まれているその他のCDスペクトル

そして、その他の含量が100%のときのCDスペクトル（その他のCDスペクトル）は、全体的に右にシフトしており、強度も200 nm付近ではProSSAの方が低い（Figure 4-14）。これらの違いは、X線結晶データから二次構造含量を算出するときに使う二次構造帰属法の違いによるものと考えられる。

4—5 CD スペクトルから二次構造含量を推定するプログラム

4—3節で求めた参照 CD スペクトルを使って、蛋白質 CD スペクトルから二次構造含量を推定するプログラムを開発した。アルゴリズムは、参照 CD スペクトルから作った $100 \times 100 \times 100$ 個の CD スペクトルと、二次構造含量を推定したい CD スペクトルの平均二乗偏差の平方根

$$RMSD = \sqrt{\frac{\sum_{\lambda} \{P_{obs}(\lambda) - P_{calc}(\lambda)\}^2}{\sum_{\lambda} \{P_{obs}(\lambda)\}^2}}$$

の値が最小になる CD スペクトルを探すようにした。ただし、 $P_{obs}(\lambda)$ は波長 λ のときの測定した蛋白質 CD スペクトル、 $P_{calc}(\lambda)$ は波長 λ のときの参照 CD スペクトルから作った CD スペクトルである。プログラム言語は C 言語を使い、キャラクタ・ユーザ・インターフェース (CUI) とグラフィカル・ユーザ・インターフェース (GUI) の二つを用意した。入力ファイルは、一行目に「#」から始まるコメント行、二行目以降に CD データを 260 nm から 190 nm まで書いたファイルである。出力ファイルは、一行目に二次構造含量の推定結果、2行目以降は、1列目に波長、2列目に実際の CD スペクトルデータ、3列目に推定した二次構造含量と参照 CD スペクトルから算出した CD スペクトルである。各列の区切りは、タブ、もしくはカンマを用いており、表計算ソフトで容易に読み込めるようにした。

4—6 推定した二次構造含量の検証

蛋白質 CD スペクトルの研究で、二次構造含量を推定する方法の検証方法として、(1)蛋白質 CD データのセットから一つ取り除いた蛋白質の二次構造含量を推定して検証する方法、(2)参照 CD スペクトルを使って、参照 CD スペクトルを算出するときに使われていない蛋白質の CD スペクトルを使って検証する方法、の二つがある。それぞれをつぎに述べる。

4—6—1 CD データのセットからの二次構造含量の推定

この方法は、相関が見られた 21 種類の蛋白質の CD スペクトルの中から、一つの蛋白質の CD スペクトルを取り除き、残りの 20 種類の蛋白質の CD スペクトルから参照 CD スペクトルを算出し、取り除いた蛋白質の CD スペクトルから二次構造含量を推定する方法である。その結果をつぎに示す (Table 4-3)。

X-ray contents は、X 線結晶解析データから ProSSA 法を使って求めた二次構造含量であり、CD contents は、参照 CD スペクトルを使って推定した二次構造含量である。つぎに、X 線結晶解析データのヘリックス含量とストランド含量、CD から推定したヘリックス含量とストランド含量のそれぞれの相関を調べた (Figure 4-15, 4-16)。

この結果から、ヘリックス含量は、相関係数 $r=0.96$ であり、1.00 に限りなく近い値なので、相関があるといえる。一方、ストランド含量は、多少広がっているものの、相関係数 $r=0.86$ であり、相関があるといえる。これらの結果は、Matsuo らの報告に一致する²⁴⁾。ただし、彼らは、真空紫外領域 (160 nm~260 nm) まで測定し、二次構造帰属法として DSSP 法を使い、しかもアルゴリズムとして優れているとされているプログラムを使った結果である。それに対して、今回の研究で使った波長領域は、178 nm から 260 nm までであり、二次構造帰属法として ProSSA 法を使った。そしてアルゴリズムは、参照 CD スペクトルを算出するのに最小二乗法を、二次構造含量を推定するのにフィッティングを使った。それにも関わらず、相関係数が、彼らの値 (ヘリックスの場合 $r=0.937$ 、ストランドの場合 $r=0.826$) よりもよい相関を得られたということは、やはり、二次構造を帰属する方法に問題があったといえるし、真空紫外領域まで測定する必要はないと考える。

Table 4-3. 21種類の蛋白質の二次構造含量

PDB code	X-ray contents (%)			CD contents (%)		
	helix	strand	other	helix	strand	other
1vxf	94	3	7	100	0	0
2mhb	88	5	11	84	16	0
1hmd	79	12	12	78	19	3
5lzm	82	14	10	73	27	0
4enl	48	28	25	54	24	22
6ldh	51	31	20	53	33	14
1lnf	44	28	28	50	41	9
1ypi	53	33	18	46	18	36
1fec	40	43	20	42	33	25
1ova	38	45	20	36	38	26
5azu	17	54	30	29	71	0
1nls	3	71	27	28	72	0
1rbx	33	55	17	22	40	38
2pab	9	67	25	21	64	15
1beb	22	56	26	20	48	32
3psg	20	57	25	13	62	25
1sxn	9	55	36	9	55	36
1rei	0	71	29	6	94	0
1esa	10	59	32	0	50	50
1tld	11	59	30	0	55	45
5cha	11	56	33	0	34	66

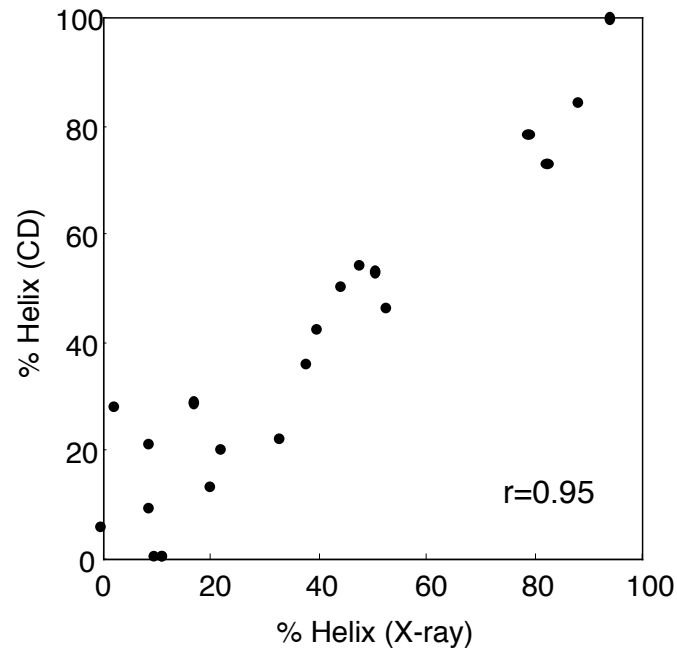


Figure 4-15. X線結晶解析データからProSSAを使って算出した二次構造含量とCDスペクトルから推定したヘリックス含量の相関

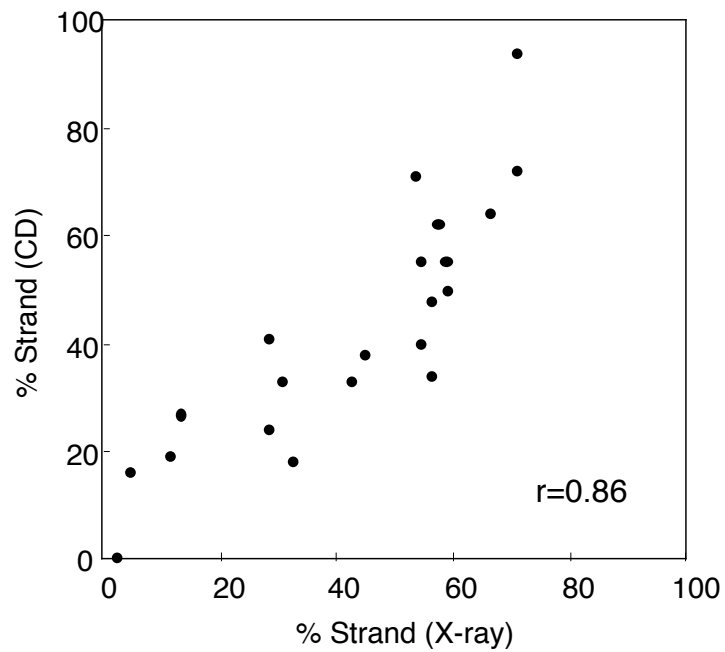


Figure 4-16. X線結晶解析データからProSSAを使って算出した二次構造含量とCDスペクトルから推定したストランド含量の相関

4—6—2 CD スペクトルからの二次構造含量の推定

蛋白質 CD スペクトルから二次構造含量を推定するプログラムは、いくつか存在する。その中で、Yang らの方法と K2D がある。Yang らの方法は、最小二乗法を使って二次構造含量を推定する⁵⁷⁾。K2D は、ニューラルネットワークを使って二次構造含量を推定する¹⁹⁾。それぞれ異なるアルゴリズムを実装している。これらのプログラムと、今回、開発したプログラムを使って、実際に測定された蛋白質 CD スペクトルから二次構造含量を推定した。使った CD スペクトルは、 α/β 構造のホスホトリエステラーゼ (PTE) である (Figure 4-17)。

つぎに、PTE の CD スペクトルから二次構造含量を推定した結果をつぎに示す (Table 4-4)。

Table 4-4. PTEのCDスペクトルから推定した二次構造含量

	Secondary Structure Contents (%)			
	alpha	beta	turn	other
X-ray (ProSSA)	57	29	-	17
X-ray (DSSP)	43	13	-	45
This study	52	29	-	19
SSE	31.2	48.4	0.0	20.3
K2D	59	7	-	33

X-ray は、PDB ファイルから二次構造帰属法 ProSSA と DSSP を使って算出した二次構造含量である。これは、Table 4-3 と同じものである。This study は、今回、開発したプログラムから得た二次構造含量、SSE は、Yang らの参照 CD スペクトルを使っているプログラムから求めた二次構造含量、K2D は、ニューラルネットワークによって得られた二次構造含量である。これらの中で、もっとも X-ray (PDB)に近い値は、今回、開発したプログラムから得た二次構造含量である。その結果、この研究で開発したプログラムでは、ヘリックス含量 52%、ストランド含量 28%、その他の含量 20%であったのに対し、Yang の方法では、ヘリックス含量 31.2%、ストランド含量 48.4%、ターン含量 0.0%、その他の含量 20.3%であり、K2D においては、ヘリックス含量 59%、ストランド含量 7%、その他の含量 33%であった。この研究と Yang の方法のそ

の他の含量について、ほぼ同じ結果となった。また、この研究と K2D は、ともにヘリックス含量が多いと判定されているが、K2D は、ストランド含量が 7% とかなり少なく見積もられた。PTE の PDB ファイル (1DPM) から求めた二次構造含量は、57%, 29%, 17% であった。これは、今回の研究の結果とほぼ一致する。したがって、PTE の CD スペクトルからの二次構造含量の推定結果は妥当な値と言える。

つぎに、実際に測定した CD スペクトルと、推定した二次構造含量と参照 CD スペクトルによって再構成したスペクトルの誤差を調べた (Figure 4-17)。その結果、平均二乗偏差の平方根は 0.28、相関係数は 0.96 だった。したがって、推定した二次構造含量は、かなりの再現性があるといえる。

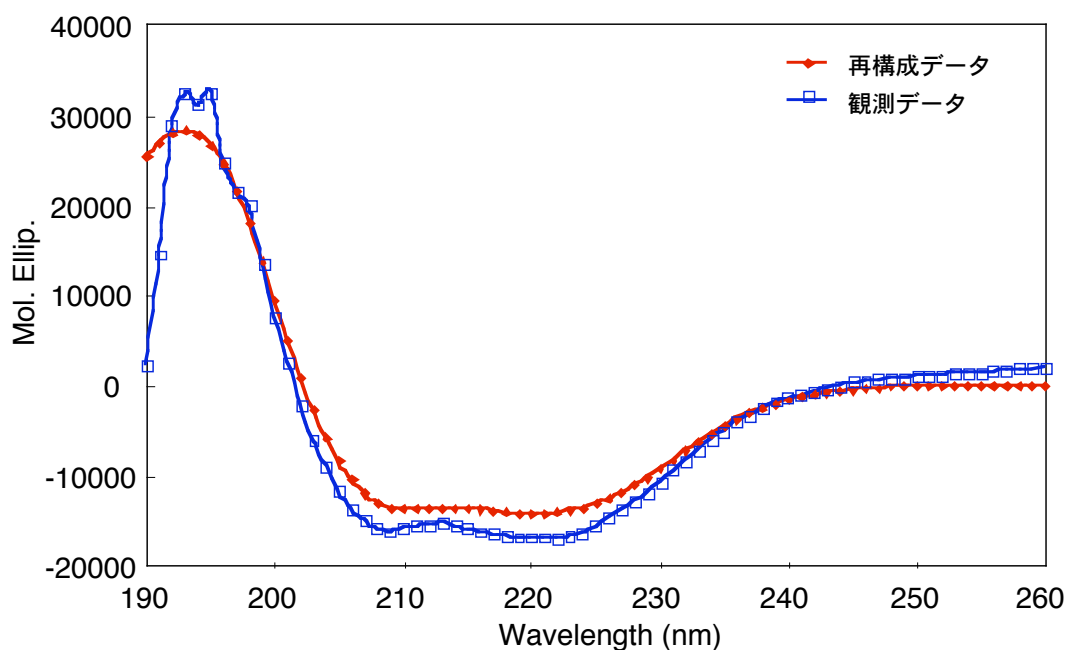


Figure 4-17. 推定した二次構造含量から再構成したCDスペクトルと実際に観測したCDスペクトル (PTE)

また、二次構造含量を推定する CD スペクトルとして、8 回くり返し配列の蛋白質の CD スペクトルも調べた (Figure 4-18)。

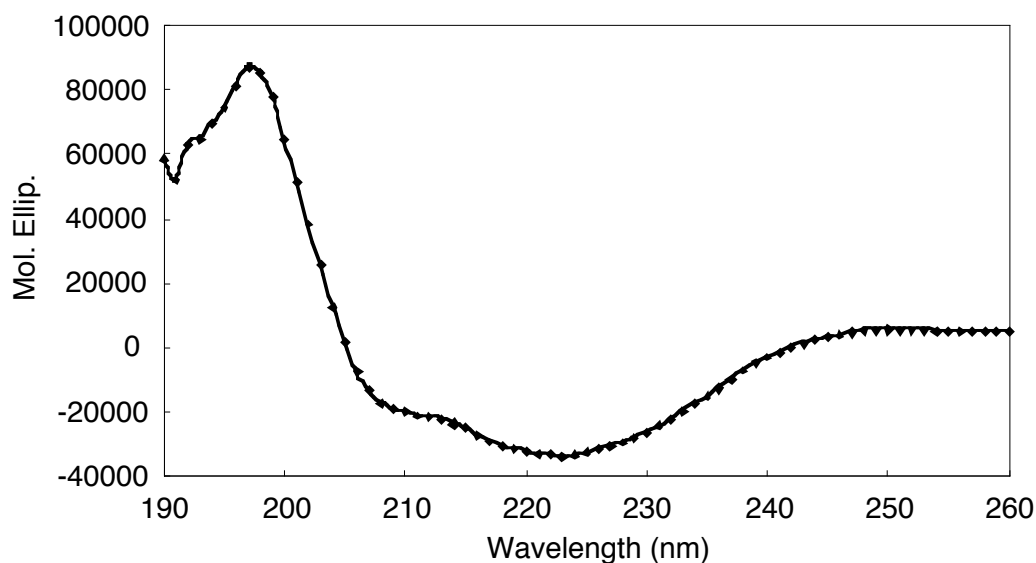


Figure 4-18. 8回くり返し配列の蛋白質のCDスペクトル

この CD スペクトルから推定した二次構造含量の結果をつぎに示す (Table 4-5)。その結果、この研究では、ヘリックス含量 100%, その他の含量 0%であり, SSE では, ヘリックス含量 65.8%, ストランド含量 11.4%, ターン含量 22.7%, その他の含量 0.0%であり, K2D は, ヘリックス含量 70%, ストランド含量 10%, その他の含量 17%であった。どの方法も、ヘリックス含量がどの含量よりも多かった。その中でも、CD スペクトルのパターンから、今回の研究の結果は、ヘリックス含量 100%と見なした参照 CD スペクトルに似ている。したがって、PTE 同様、二次構造含量推定結果は妥当な値と考える。なお、8回繰り返し配列の蛋白質の X 線結晶解析データはない。

Table 4-5. 8回くり返し配列の蛋白質のCDスペクトルから推定した二次構造含量

	Secondary Structure Contents (%)			
	alpha	beta	turn	other
This study	100	0	-	0
SSE	65.8	11.4	22.7	0.0
K2D	70	10	-	17

4-7 蛋白質 CD スペクトルのデータベース

蛋白質の CD スペクトルのデータベースは、蛋白質の立体構造や機能を研究している研究者にとって有益な情報を提供する。そして、蛋白質の CD スペクトルに関するオンラインサイトとして、二次構造含量を推定するプログラムを Web 上から実行できる DICHROWEB^(60,61)や、CD スペクトルのデータベース The Protein Circular Dichroism Data Bank (PCDDDB)^(62,63)がある。PCDDDB は、CD スペクトル測定装置で実際に蛋白質の CD スペクトルを測定し、蛋白質の情報、測定条件、測定結果、推定した二次構造含量の結果を提供しているサイトである。しかしながら、PCDDDB はさまざまな蛋白質の CD スペクトルを測定せねばならず、時間と労力が必要である。この節では、ProSSA 法をベースに開発した二次構造推定プログラムの結果や実際に測定した蛋白質の CD スペクトルの測定条件・測定結果、過去の論文に掲載された CD スペクトルの測定条件やスペクトルを含むデータベースについて述べる。

この研究で構築したデータベースは二つのカテゴリに分類した。一つは、CD スペクトル測定装置で測定した CD スペクトルデータを使ったもの、もう一つは、論文に掲載された蛋白質 CD スペクトルを数値化したものである。これらのスペクトルを、それぞれ Measured CD Spectra と Paper CD Spectra とした (Figure 4-19)。

これらはどちらも次の項目をもつ (Figure 4-20)。

- (1) CD スペクトルの波形を描いた図
- (2) CD スペクトルの測定条件
- (3) 測定した蛋白質の X 線結晶解析から得られる立体構造
- (4) 二次構造含量
- (5) 注釈

つぎに、(1) ～ (5) の項目について述べる。

(1) CD スペクトルの波形図

蛋白質の CD スペクトルの波形は、実際に測定した CD スペクトルから描いた波形と論文から CD スペクトルを抽出し数値化したデータを使って描いた波形の 2 種類に分ける。なお、数

値化するために使ったソフトウェアは, Spectrum Converter Version 1.00 (<http://www.vector.co.jp/soft/win95/edu/se349903.html>) である。また, 波長領域は, 190 nm から 260 nm を基本とするが, 論文から抽出した CD スペクトルは, 論文に掲載された波長領域を採用する。

(2) CD スペクトルの測定条件

蛋白質を実際に測定したときの測定条件をのる。また, 論文から得た測定条件のうち, 記述されていない条件は, 空欄にする。

(3) 測定した蛋白質の X 線結晶解析から得られた立体構造

X 線結晶解析データは, PDB から取得する。さらに, 立体構造を描画したときの二次構造は, ProSSA 法を使って帰属した二次構造の結果を反映させる。なお, PDB ファイルが存在しない蛋白質はのせない。



Figure 4-19. 蛋白質CDスペクトルのデータベース

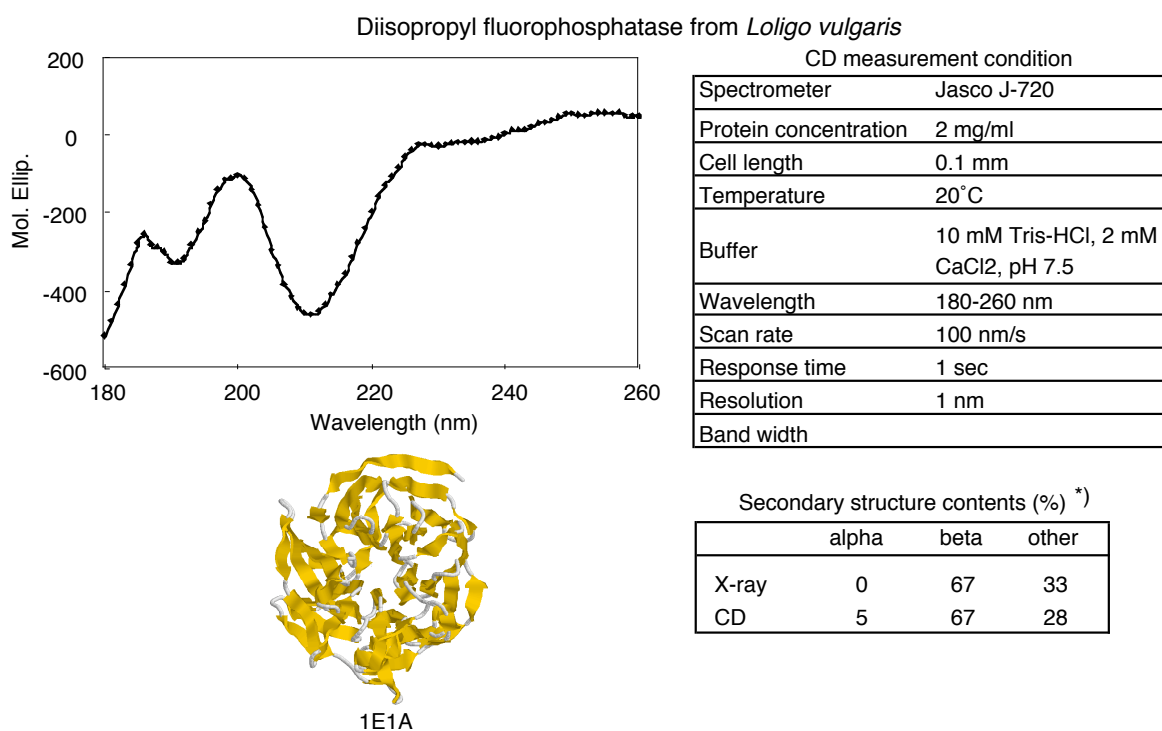
(4) 二次構造含量

二次構造含量は PDB ファイルから ProSSA 法を使って得られた二次構造含量と、CD スペクトルから新しく開発した二次構造含量推定プログラムを使って算出した二次構造含量をのせる。なお、PDB ファイルが存在しない場合は空欄にする。

(5) 注釈

CD スペクトルや測定条件を引用した参考文献, CD スペクトルの波形を描いたときに使ったファイル, ProSSA 法で帰属したファイルへのリンク, 立体構造を Rasmol で描画したときに使ったファイルへのリンクなどを記入する。また、存在しない項目はのせない。

つぎに、データベースのフォーマットを Figure 4-20 に示す。



- Note: Mol. Ellip. was converted into mean residue molar ellipticity
- Reference: *Biochim. Biophys. Acta*, **1546**, 312-324 (2001).
- Extracted CD data: [DFPase-scan.csv](#)
- ProSSA format: [1E1A.dsp](#)
- Rasmol drawing file (PDB format) applied to ProSSA: [1E1A_prossa.pdb](#)

^{*)} X-ray and CD are the protein secondary structure contents calculated using ProSSA method from PDB file and estimated from CD spectrum, respectively. The alpha, beta and other are α -helix, β -strand and other structure contents, respectively.

Figure 4-20. 蛋白質CDスペクトルのデータベースのフォーマット

以上、この章では、蛋白質の CD スペクトルと二次構造含量の関係を調べるために、二次構造帰属法として、より多くの二次構造を帰属する ProSSA を使った。そして、構造が知られている CD スペクトルと、PDB ファイルから ProSSA 法によって算出した二次構造含量との相関を調べ、相関が見られた 21 種類の蛋白質の CD スペクトルを使って参照 CD スペクトルを求めた。この参照 CD スペクトルと、PDB ファイルから ProSSA によって得た二次構造含量を組み合わせ再構成した CD スペクトルは、実際に測定した CD スペクトルに一致した。また、二次構造含量を推定する方法の有用性を調べるために、21 種類の蛋白質の CD スペクトルから 1 種類を取り除いた 20 種類の CD スペクトルを使って参照 CD スペクトルを算出し、取り除いた CD スペクトルから二次構造含量を推定した。その結果、Matsuo らの結果と一致するデータを得た。二次構造帰属法として ProSSA は、真空紫外領域まで測定する必要がなく CD の研究に適しているといえる。さらに、この参照 CD スペクトルを使って実際に測定された CD スペクトルから二次構造含量を推定した。その結果、X 線結晶解析から算出した二次構造含量に極めて近い値を得た。したがって、今回の研究で求めた参照 CD スペクトルを使った二次構造含量を推定する方法は、構造がまだわかっていない蛋白質の CD スペクトルから二次構造含量を推定したり、温度によって構造が変化する様子を二次構造含量で定量的にみる研究に使える。

第5章 プロテアーゼ消化配列の ProSSA 法による解析

5-1 はじめに

プロテアーゼには、基質となる蛋白質の一次配列の特定の部位を認識、切断することで、その蛋白質を不活性型から活性型へ、または、その逆へ変化させる働きがある。このようなプロテアーゼの研究は、基質のどの配列を加水分解するのか、プロテアーゼの働きを阻害する阻害剤としてはどのようなものがあるのか、プロテアーゼ活性をもつにはどのような過程を経るのか、プロテアーゼの機能を向上させるにはどうすればよいか、などを対象としている。つまり、プロテアーゼ自身の立体構造、機能、認識する配列などに焦点が向いている。その中でも、基質となる蛋白質は、プロテアーゼによって一次配列上どのような配列が切断されるかに注目されている。基質蛋白質の中の消化を受ける配列が、どのような二次構造の中に存在するのかについて調べた例は少ない。たとえば、切断される配列はヘリックスの真ん中やループ領域という報告がある^{64,65)}。しかしながら、これらの研究は、二次構造を帰属する方法が統一されていない状況での比較に過ぎない。PDB ファイルの二次構造情報を参考にするのも一つの方法であるが、すべての PDB ファイルの二次構造情報が、PDB が推奨する DSSP 法に沿っているわけではない。そこで、まず、二次構造帰属法として、多くの研究者に利用されている水素結合様式パターンによって二次構造を帰属する DSSP 法を使って、プロテアーゼが切断する配列と二次構造の関係を調べた。(Figure 5-1)。

Figure 5-1 の左側の列は、基質蛋白質、右側は、基質蛋白質の一次配列と消化配列である。緑色で囲っているのが消化配列、赤色と黄色で囲っている配列は、水素結合様式パターンで二次構造を帰属する DSSP 法の α ヘリックス構造と β ストランド構造である。ほとんどの基質蛋白質は、消化配列から離れたところに二次構造が存在するか、もしくは、二次構造が存在しなかった。このように、プロテアーゼの消化する配列と二次構造とは全く相関が内容に見えるが、二次構造の帰属法を変えることで、まったく異なる様相が見えてきた。具体的には、蛋白質の約 70%を二次構造と帰属する DSSP 法から、約 90%を二次構造と帰属する ProSSA 法を使ったときの切断配列と二次構造の関係を検証した。

プラスミノーゲン	557 K	558 C	559 P	560 G	561 R	562 V	563 V	564 G	565 G	566 C	567 V	568 A
キモトリプシノーゲン	10 L	11 S	12 G	13 L	14 S	15 R	16 I	17 V	18 N	19 G	20 E	21 E
コアグローゲン	14 G	15 V	16 L	17 G	18 R	19 T	20 Q	21 I	22 V	23 T	24 T	
	41 S	42 G	43 V	44 S	45 G	46 R	47 G	48 F	49 S	50 I	51 F	
プロスロンピン	14J Y	14K I	14L D	14M G	15 R	16 I	17 V	18 E	19 G	20 S	21 D	22 A
	266 A	267 A	268 I	269 E	270 G	271 R	272 T	273 S	274 E	275 D	276 H	
トリプシノーゲン	141 W	142 G	143 N	144 T	145 K	146 S	147 S	148 G	149 T	150 S	151 Y	
DFPase	141 P	142 A	143 D	144 Y	145 T	146 R	147 S	148 M	149 Q	150 E	151 K	152 F
	303 K	304 Q	305 Y	306 C	307 E	308 T	309 L	310 K	311 F	312 G	313 I	314 F
BPTI	11 T	12 G	13 P	14 C	15 K	16 A	17 R	18 I	19 I	20 R	21 Y	22 F

：切断配列： α ヘリックス構造： β ストランド構造

Figure 5-1. DSSP法による二次構造と切断配列の関係

5—2 プロテアーゼ消化部位と二次構造の相関

SWISS-PLOT⁶⁶⁾より、プロテアーゼとその基質となる蛋白質と切断配列の情報を検索し 46 個のデータを選び出した。これらのデータの中から、PDB ファイルを有するもので消化を受ける前の立体構造を探したところ、5種類の蛋白質を得た。さらに、二つの論文からプロテアーゼによって消化を受ける蛋白質の情報を取得した^{64,65)}。結果として、合計で 7 種類の消化を受ける前の基質蛋白質の PDB ファイルを取得した。つぎに、これら 7 個のデータを使って、消化配列と二次構造の関係を調べた。

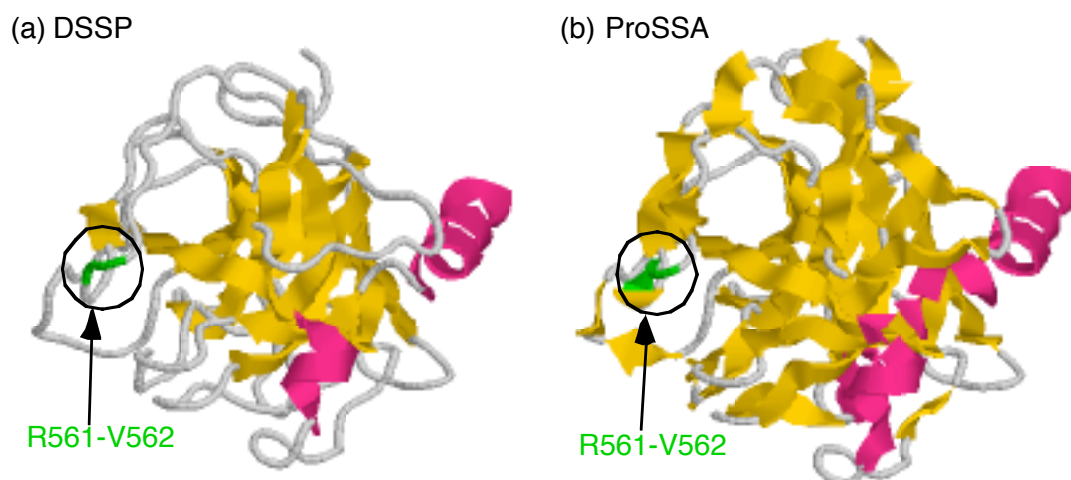


Figure 5-2. プラスミノーゲンの立体構造 (1QRZ) と切断配列

プラスミノーゲンは、ウロキナーゼの基質蛋白質である (Figure 5-2)。なお、ウロキナーゼは、プラスミノーゲンアクチベータの一つであり、プラスミノーゲンのアルギニン-バリンの結合を切断して、プラスミンにする活性がある。Figure 5-2 は、赤色のリボンで α ヘリックスを、黄色のリボンで β ストランド構造を表わしたプラスミノーゲンの立体構造である。そして、左側は、DSSP 法で二次構造を帰属したときの立体構造のリボン表示、右側は、ProSSA 法のリボン表示である。プラスミノーゲンは、ウロキナーゼによって、561 番目のアルギニンと562 番目のバリンの配列が認識され、切断される。DSSP 法で二次構造を帰属した場合、消化配列の前後に二次構造は見当たらないが、ProSSA 法では、前後に二次構造が見当たる。これは、ProSSA 法は、DSSP 法よりも、多くの二次構造を帰属することができるためである。したがって、DSSP 法では、消化配列と二次構造に関係が見られなかったが、ProSSA 法を使うことで、何らかの相関性を見つけることができると考えた。

キモトリプシノーゲンは、トリプシンの基質蛋白質であり、トリプシンによって得られた中間体 π -キモトリプシンは、キモトリプシンの基質蛋白質である (Figure 5-3)。トリプシンは、キモトリプシンなどの前駆体を活性化させる働きがあり、リジンやアルギニンのカルボニル側を切断する。そして、キモトリプシンは、ロイシン、アスパラギンなどのカルボニル側を切断

する。つまり、キモトリプシノーゲンは、トリプシンによって、アルギニンが認識され、15 番目のアルギニンのカルボニル側の結合、つまり Arg15-Ile16 の結合が切断される。この切断によって得られた中間体 π -キモトリプシンは、キモトリプシンによって 13 番目のロイシンのカルボニル側の結合、つまり Leu13-Ser14 の結合と、146 番目のチロシンと 148 番目のアスパラギンのカルボニル側の結合、つまり、Tyr146-Thr147 と Asn148-Ala149 の結合が切断され、キモトリプシンが得られる。ただし、1CHG からは、Tyr146-Thr147 と Asn148-Ala149 を含む配列や構造は得られなかったため、Figure 5-3 には、記していない。

Figure 5-3 から、DSSP 法では、消化配列の前後に二次構造は見られない。しかし、ProSSA 法では、消化配列の前に β ストランド構造が存在する。

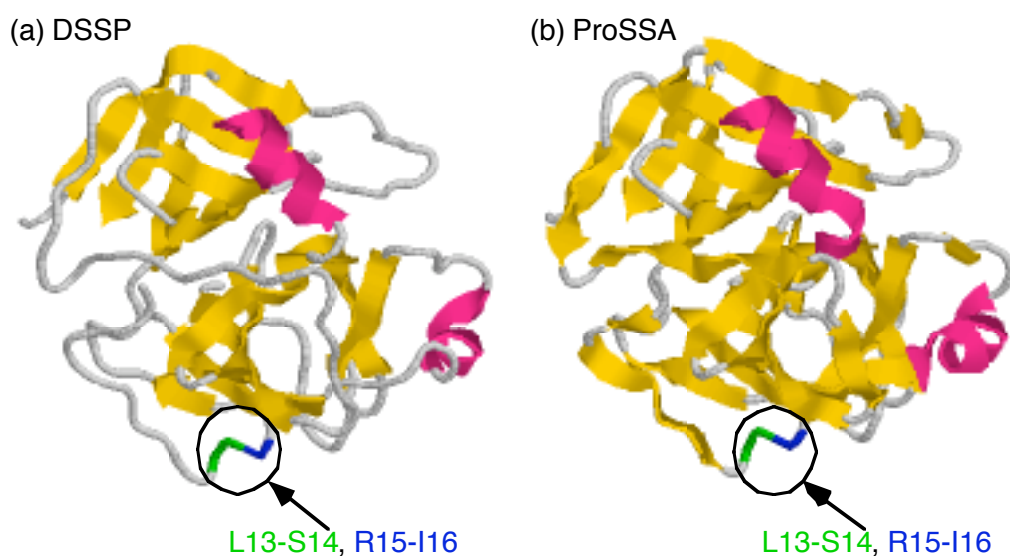


Figure 5-3. キモトリプシノーゲンの立体構造 (1CHG) と切断配列

コアグロゲンは、カプトガニ凝固酵素の基質蛋白質である (Figure 5-4)。なお、カプトガニ凝固酵素は、コアグロゲンの 18 番目のアルギニンと 47 番目のアルギニンのカルボニル側を選択的に切断する。つまり、コアグロゲンの Arg18-Thr19 と Arg47-Gly48 のペプチド結合を切断する。Figure 5-4 から、DSSP 法では、二つの切断配列の前後には、二次構造は存在しなかったが、ProSSA 法では、 β ストランド構造や α ヘリックス構造が存在した。なお、PDB フ

ファイルでは、47 番目のアルギニンと 48 番目のグリシンの残基番号が、それぞれ 46 番目と 47 番目であったので、Figure 5-4 では、PDB ファイルの残基番号を使った。

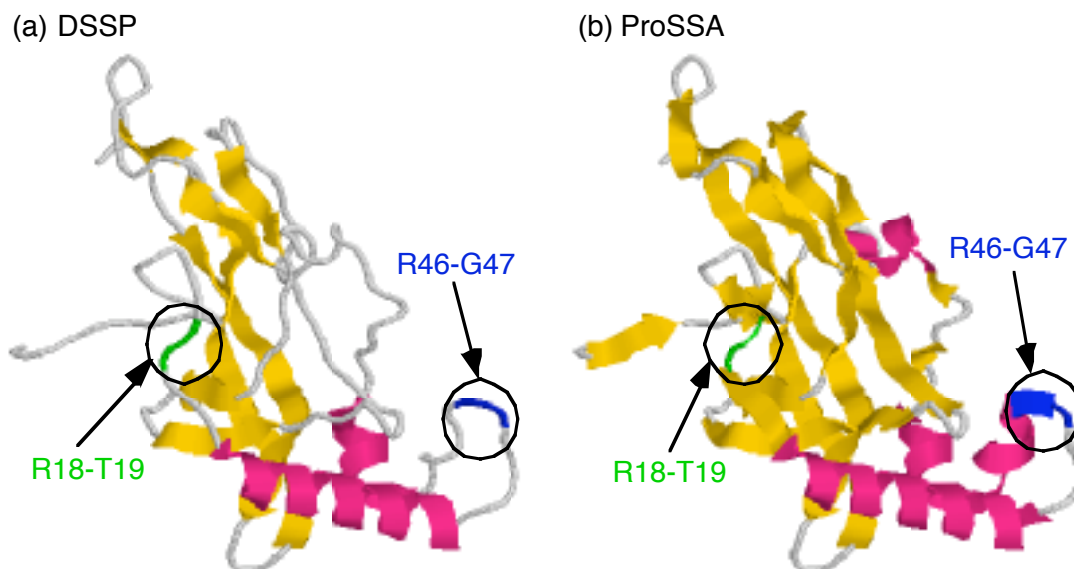


Figure 5-4. コアグロゲンの立体構造 (1AOC) と切断配列

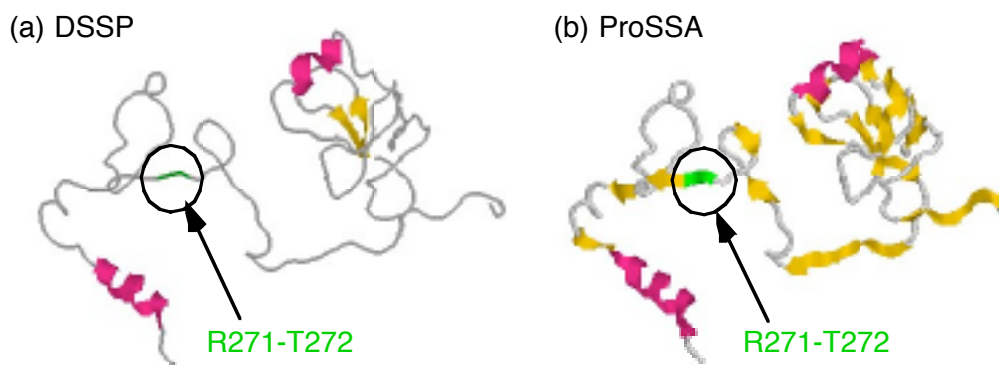


Figure 5-5. プロスロンビンの立体構造 (1A0H) と切断配列 (その1)

プロスロンビンは凝固因子 Xa の基質蛋白質である (Figure 5-5, 5-6)。凝固因子 Xa は、プロスロンビンの Arg-Thr, Arg-Ile を選択的に切断する。Figure 5-5 から凝固因子 Xa は、プロスロンビンの 271 番目のアルギニン残基と 272 番目のスレオニン残基の結合を切断し、Figure 5-6

から凝固因子 Xa は、プロスロンビンの 15 番目のアルギニン残基と 16 番目のイソロイシン残基の間の結合を切断する。Figure 5-5 において、DSSP 法では消化配列の近辺には二次構造は見られず、ProSSA 法では、消化配列は β ストランド上に存在した。また、Figure 5-6 において、DSSP 法と ProSSA 法の両方では消化配列の前後に二次構造があった。

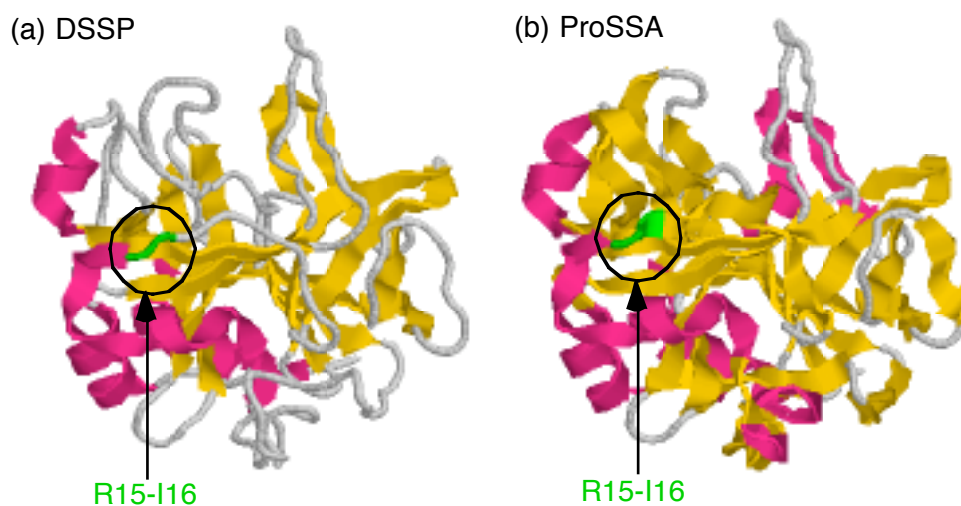


Figure 5-6. プロスロンビンの立体構造 (1HAG) と切断配列 (その2)

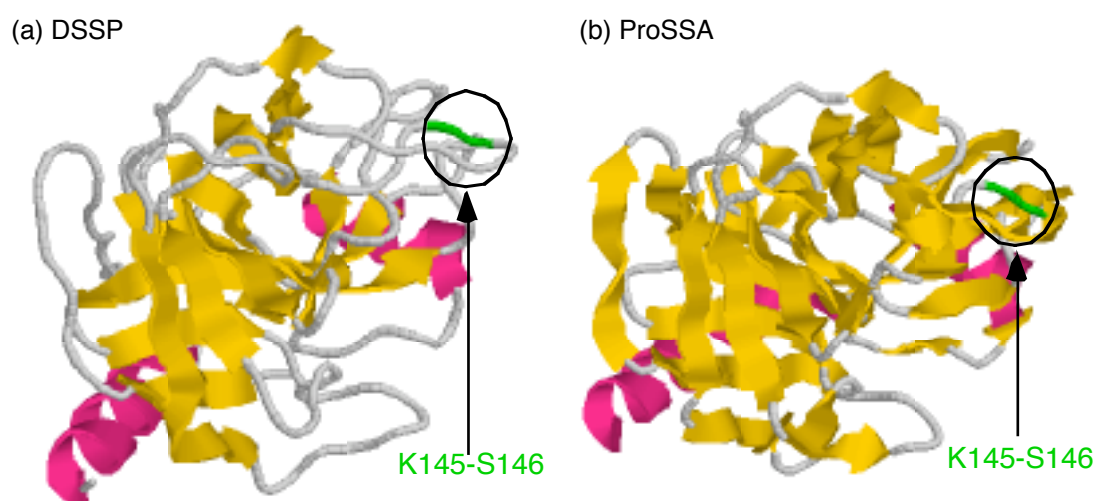


Figure 5-7. トリプシノーゲンの立体構造 (1GQI) と切断配列

トリプシノーゲンは α トリプシンの基質蛋白質である (Figure 5-7)。 α トリプシンはアルギニン残基やリジン残基のカルボニル側を切断する。Figure 5-7 では、 α トリプシンによって、トリプシノーゲンの 145 番目のリジン残基と 146 番目のセリン残基のペプチド結合を切断する。これまでと同様に、DSSP 法では切断配列の前後に二次構造は存在せず、ProSSA 法では β ストランド構造が存在した。

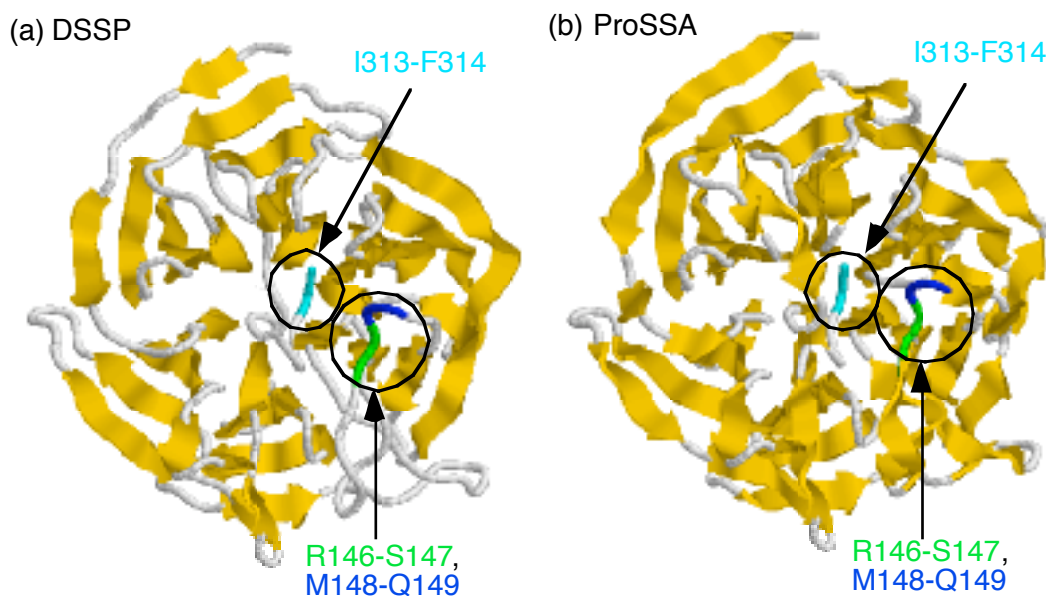


Figure 5-8. DFPaseの立体構造 (1E1A) と切断配列

DFPase は、146 番目のアルギニンと 147 番目のセリン、148 番目のメチオニンと 149 番目のグルタミン、313 番目のイソロイシンと 314 番目のフェニルアラニンの三ヶ所のペプチド結合をプロテイナーゼ K によって切断される⁶⁴⁾ (Figure 5-8)。Figure 5-8 の左側の DSSP 法を使ったリボン表示では、三つの消化配列の前後に二次構造が見られない。しかしながら、右側の ProSSA 法を使ったリボン表示では、Arg146-Ser147 と Met148-Gln149 の前後に β ストランド構造が見られたが、Ile313-Phe314 の近くには二次構造は見られなかった。なお、Hartleib と Ruterjans は、Arg146-Ser146 がループ領域に存在することを指摘した。

牛膵臓トリプシン・インヒビター (BPTI) は、15 番目のリジンと 16 番目のアラニンのペプチド結合をトリプシンによって消化される⁶⁵⁾ (Figure 5-9)。これは、トリプシンがリジンのカ

ルボニル側のペプチド結合を選択的に切断するためである。Figure 5-9 の左側は、DSSP 法で二次構造を帰属したときのリボン表示、右側は ProSSA 法のためのリボン表示である。DSSP 法と ProSSA 法の両方で、消化される配列の前後に β ストランド構造が見られる。

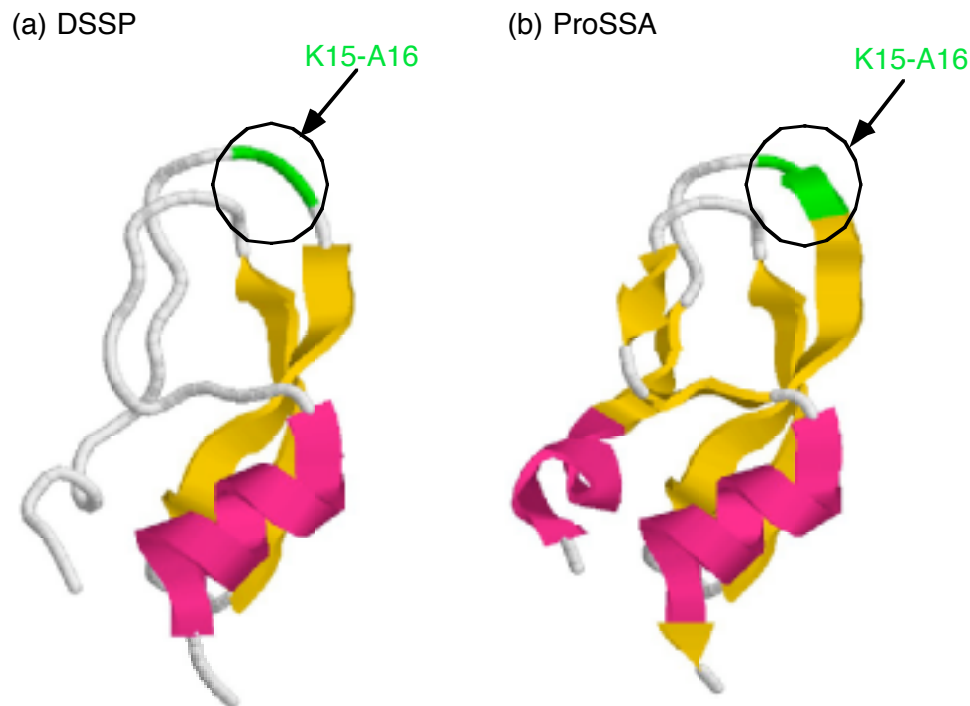


Figure 5-9. BPTIの立体構造（5PTI）と切断配列

これらの結果より、DSSP 法で二次構造を帰属したとき、ほとんどの基質蛋白質において、消化を受ける配列と二次構造に関係が見られなかった (Figure 5-1)。しかしながら、ProSSA 法で二次構造を帰属したとき、消化を受ける配列の前後に二次構造が存在した。そこで、ProSSA 法で二次構造を帰属したときの基質蛋白質の消化部位と二次構造の関係をつぎに示す (Figure 5-10)。

Figure 5-10 は、左にプロテアーゼの基質、右のアミノ酸配列を表し、緑色の囲いがプロテアーゼによる消化配列、赤色と黄色がそれぞれ ProSSA 法による α ヘリックス構造、 β ストランド構造である。消化配列と二次構造の関係をみると、消化配列のほとんどが β ストランド構造の N 末端の境界領域に存在した。

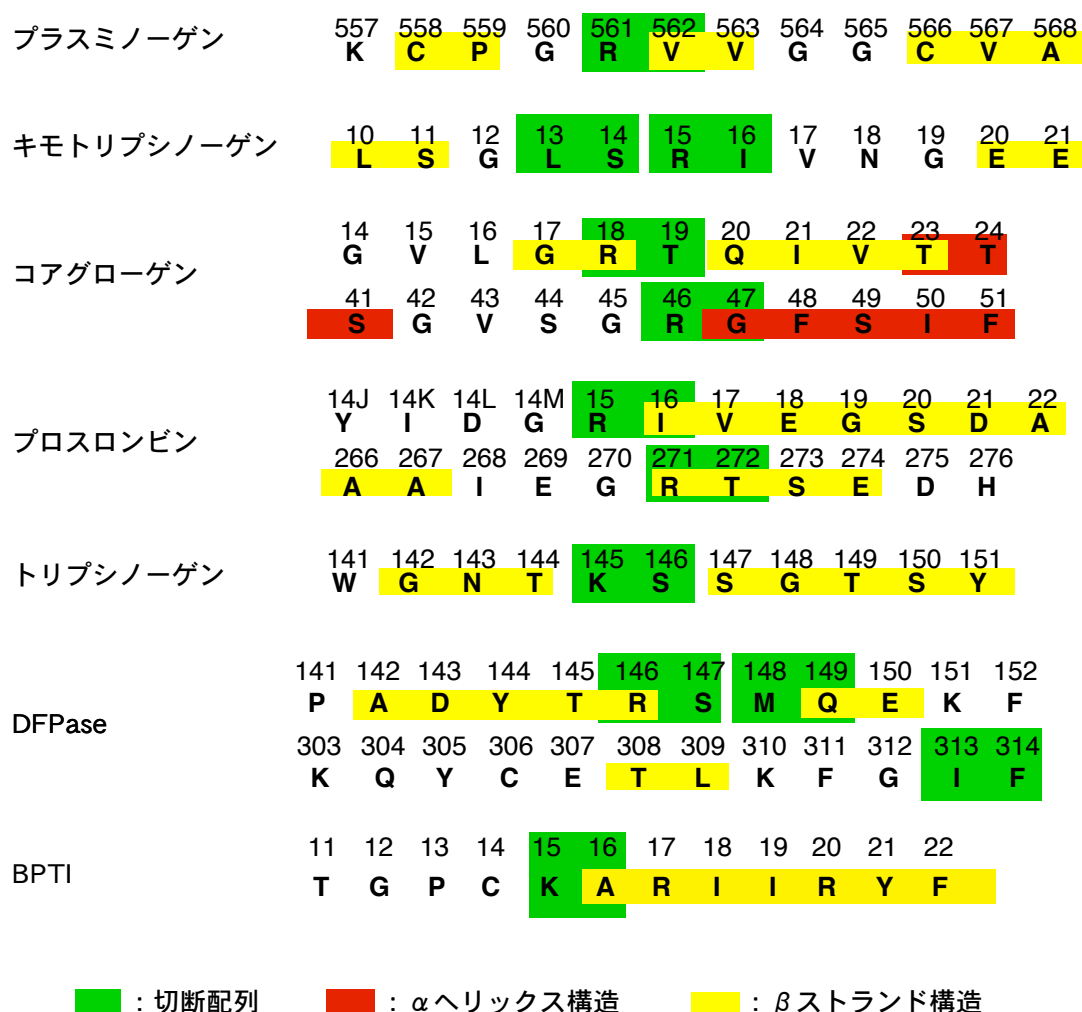


Figure 5-10. ProSSA法による二次構造と切断配列の関係

基質の蛋白質の中には、プロテアーゼの認識配列でありながら消化されない配列が存在する場合があった。このような配列についても、二次構造とどのような関係にあるのかを調べた。例として、プロプロテイナーゼ E の消化配列と同じ配列が、二次構造とどのような関係をもつかを示す (Figure 5-11)。

黒色で囲っている配列は、認識配列でありながら、消化を受けない配列である。ほとんどの基質で、消化を受けない配列は、βストランド構造の C 末端の境界領域に存在した。これは、プロテアーゼが二次構造の制限を受けていることを示すものと理解できる。

しかしながら、わずかな例ではあるが、消化される配列と二次構造との相関が見られない場

合もあった (Figure 5-12)。

キモトリプシノーゲンでは、 β ストランド構造の N 末端の境界領域ではない部分に切断配列があった。また、コアグロゲンでは、 α ヘリックス構造の N 末端の境界領域にあった。

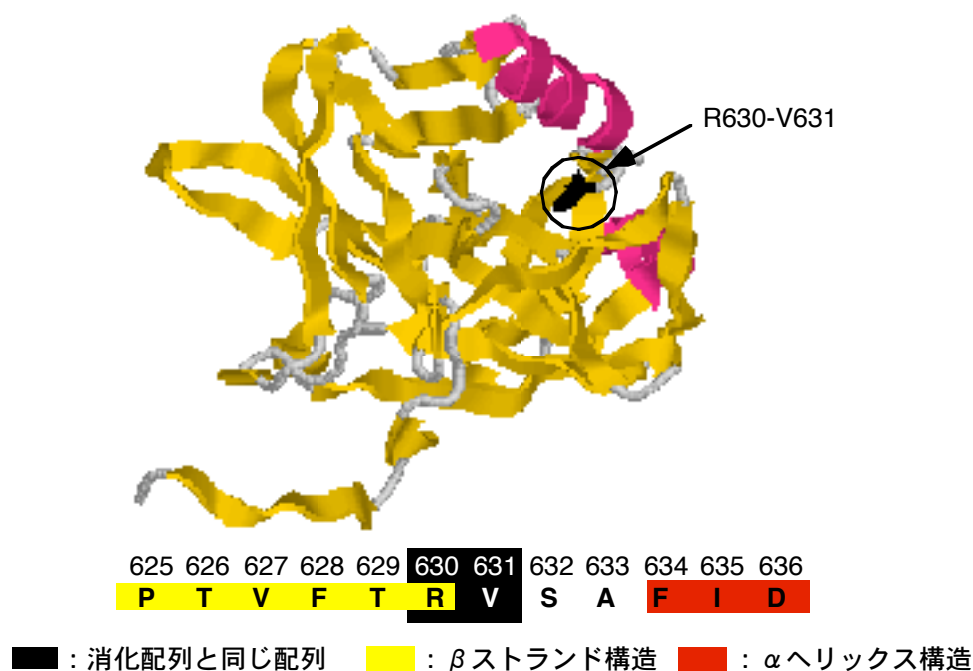
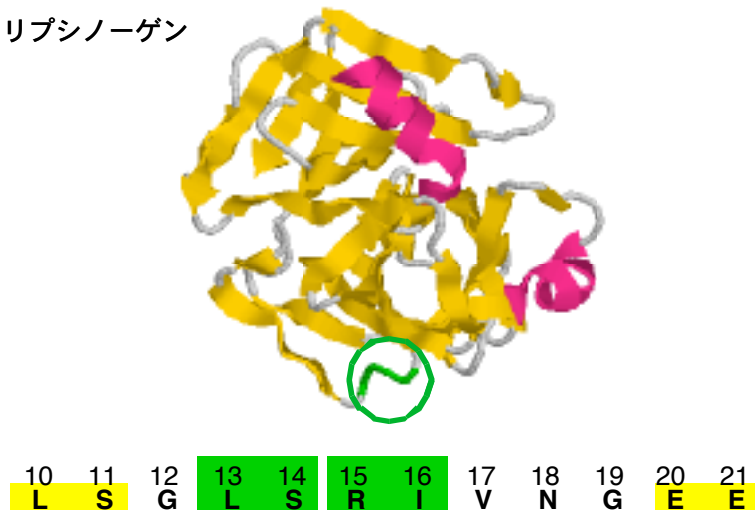


Figure 5-11. 切断配列と同じ配列 (プロプロテイナーゼE)

これまで、プロテアーゼの研究は、プロテアーゼ自身の構造安定化や機能の向上を目指してきた。そして、プロテアーゼが消化する配列は、ループ領域やヘリックスの真ん中などに存在するといわれていたが、その時に使われている二次構造帰属法は、統一されていなかった。そこで、まず、DSSP 法で二次構造を帰属し、切断配列と二次構造の関係を調べた。ところが、DSSP 法では、相関が得られなかった。しかしながら、主鎖二面体角の連続性パターンによる二次構造帰属法 ProSSA を用いると、消化される配列は、 β ストランド構造の N 末端の境界領域に現れ、消化されない配列は、それ以外の領域に存在した。すなわち、プロテアーゼは、認識配列の中でも、その配列が、 β ストランド構造の N 末端の境界領域に存在する配列を消化の対象とすることがわかった。これは、プロテアーゼが、配列だけでなく、二次構造も認識して

いることに他ならない。

(a) キモトリプシノーゲン



(b) コアグロゲン

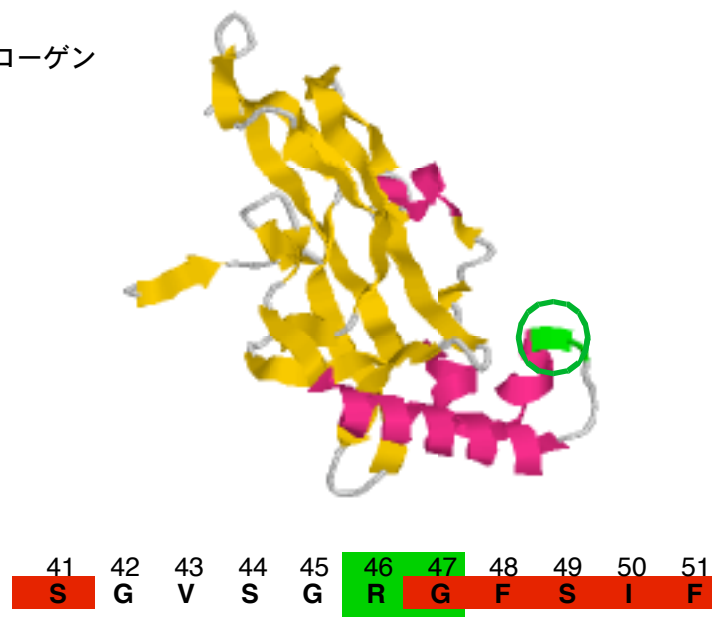


Figure 5-12. 相関が見られない消化される配列

今回の解析結果を踏まえると、プロテアーゼごとの認識配列は、ほとんどわかっており、基質の蛋白質の表面にきている配列を情報に加え、二次構造情報を追加することで、プロテアー

ぜによる消化部位の予測プログラムを作成することが可能となる。

以上のことから、第4章のCDに関する研究に続いて、ProSSA法は、プロテアーゼの研究にも有用な方法であることを証明した。

第6章 β ストランド構造の解析

6-1 はじめに

Anfinsen のドグマとして知られるように、蛋白質の立体構造は一次構造によって決まると認識されてきたが、必ずしもそうではないことが判明した。例えば、スクレイピーや、アルツハイマー病などは、正常に機能しているプリオン蛋白質（正常型蛋白質）から正常に機能しない蛋白質（異常型蛋白質）へ変化することで引き起こされる病気がある。プリオンの場合、正常型蛋白質の立体構造は α ヘリックス構造であるが、異常型は β シート構造であり、正常型から異常型へ構造変化することで病気が発症する。このような α ヘリックス構造から β シート構造への構造変化は、蛋白質自身の立体構造の安定化エネルギーが変化の前後の立体構造に対して非常に近い状態の場合に区別がつかない。多くはその蛋白質が置かれている環境が深く影響していると考えられる。

β シート構造への変化は、 β シート構造体の集積によってなされる場合が多い。伸びた構造としての β ストランド構造が2本、3本と、平面上の β シート構造を形成し、その平面構造体が積層し巨大分子マトリックスの構造体として現れる。この会合性構造体は、不溶性である場合が多く、ときに病気の原因物質として認められる。先のプリオンやアミロイド繊維も β シート構造が積層して巨大分子マトリックスを形成している。このように、 β シート構造の解析は、これらの病気などの解明にもつながる。これまでの蛋白質二次構造解析は α ヘリックス構造が中心であったが、スクレイピーやアルツハイマーの原因が明らかになるにつれて、 β シート構造の研究にも注目が集まってきている^{11,32,33,67)}。しかし、そのほとんどは β シート構造に属する β ストランド構造の解析であった。

金光は、主鎖二面体角の連続性パターンによる二次構造帰属法 ProSSA を開発した¹⁰⁾。ProSSA 法は、 β シート構造と同じコンホメーションに対しては β ストランド構造と分類帰属される。しかしながら、ProSSA 法は水素結合の情報をもたないため、 β ストランド構造が β シート構造に関わっているものなのか、もしくは、関わっていないものなのかの区別がつかない (Figure 6-1)。一方で、DSSP 法は β シート構造のみを帰属していることから、DSSP 法と ProSSA 法を

組み合わせれば、 β シート構造形成ストランド構造と、 β 構造非形成ストランド構造を区別することが可能となる。よって、この章では、ProSSA法で帰属した β ストランド構造をシート構造に属している β ストランド構造と単に伸びた β ストランド構造の二つの β ストランド構造に分けて β ストランド構造の解析を行なった。

繰り返すが、単に伸びた β ストランド構造とは β ストランド構造の主鎖のアミドと他の β ストランド構造のアミドが水素結合していない構造のことであり、これを一本鎖 β ストランド構造とよぶ。また、 β シート構造に属している β ストランド構造とは、 β ストランド構造の主鎖のアミドと他の β ストランド構造のアミドが水素結合している構造のことであり、これをシート形成 β ストランド構造とよぶ。具体的な振り分け方法は、ProSSA法で帰属した β ストランド構造が、DSSP法で帰属した β シート構造と重なっていれば、シート形成 β ストランド、重なっていなければ、一本鎖 β ストランド構造と分類した。(Figure 6-2)。

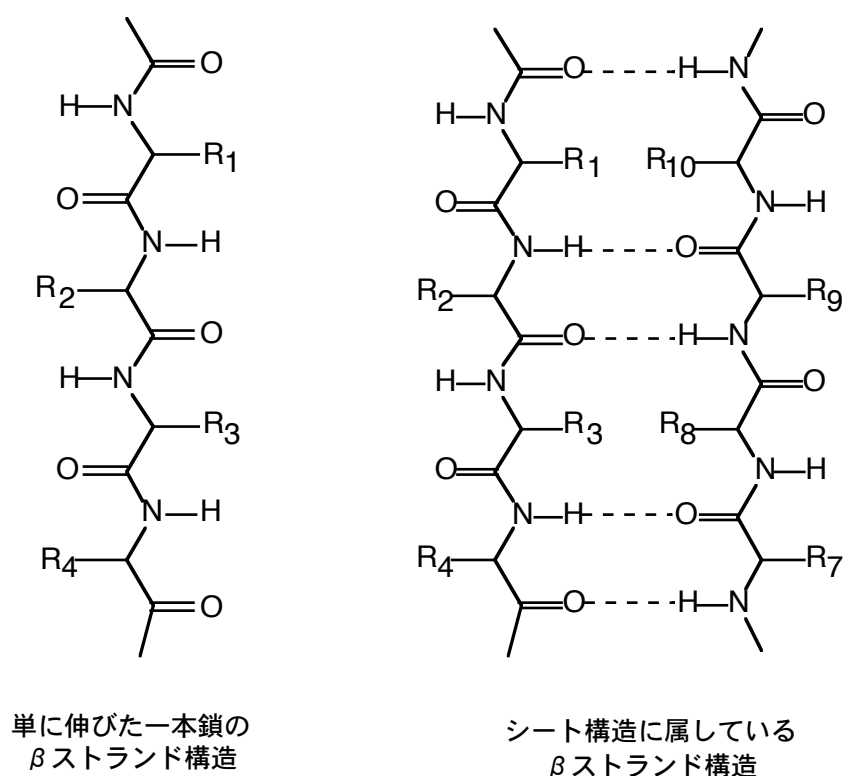


Figure 6-1. ProSSA法の β ストランド構造

Figure 6-2 は、シート形成 β ストランド構造が四本、一本鎖 β ストランド構造が一本の例である。これらの β ストランド構造は、ProSSA 法で使う β 領域で2 残基以上連続している必要がある。なお、単に β 構造、または、 β ストランド構造と書いたときは、シート形成と一本鎖を区別していないときの構造のことをさす。

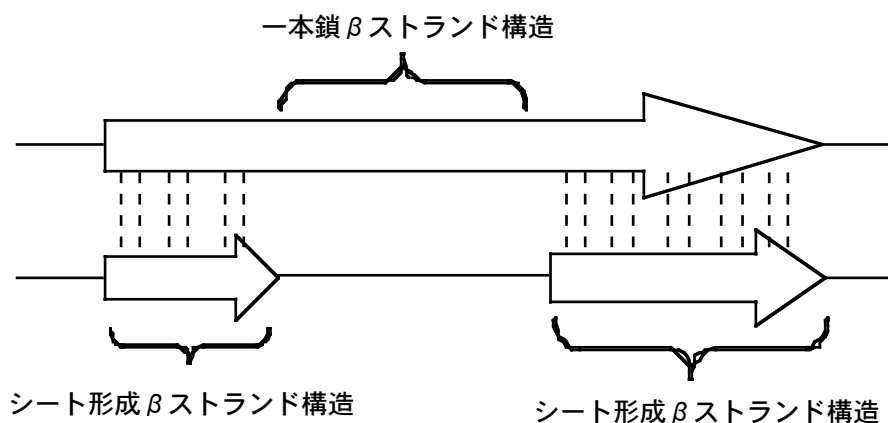


Figure 6-2. シート形成と一本鎖の β ストランド構造の例

6—2 ProSSA 法によって帰属される β ストランド構造の解析

まず、1081 蛋白質に対して ProSSA 法を適用し二次構造を帰属した。そして、 β ストランド構造の長さに対する出現頻度を調べた (Figure 6-3)。抽出した β ストランド構造のアミノ酸残基数は、最小で2 残基、最大で27 残基であった。平均残基数は、4.7 残基であった。Figure 6-2 の結果を見ると、鎖長が長くなるにつれ出現頻度は減少した。これは Penel らの結果と一致する^{11,32,33)}。また、もっとも多く出現した鎖長は2 残基であり、この結果は Eswar らと異なる³²⁾。この理由は β ストランド構造を4 残基以上にしたためである。

つぎに、 β ストランド構造の中のアミノ酸残基の主鎖二面体角の平均値と標準偏差を調べた (Table 6-1)。

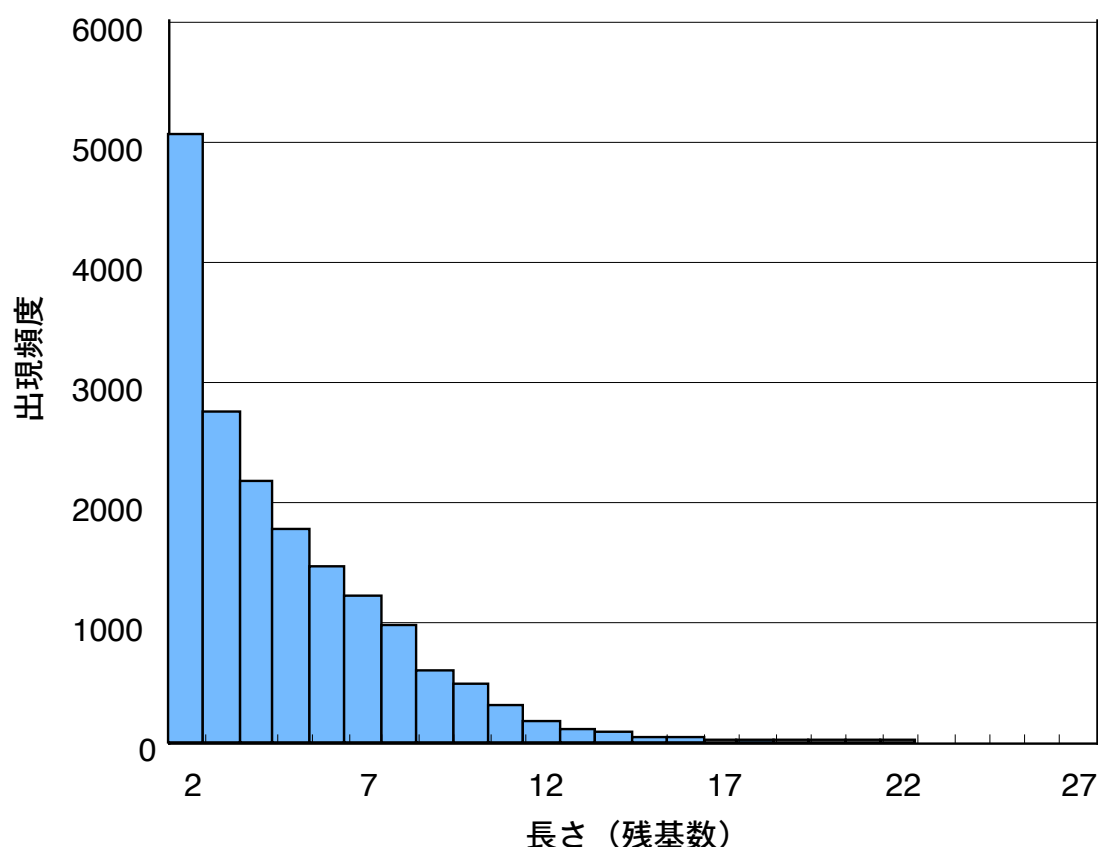


Figure 6-3. β ストランド構造の長さに対する出現頻度

ほとんどのアミノ酸残基において、 ϕ 角が -110° 、 ψ 角が 130° 前後の値を示した。ただし、グリシン残基は側鎖がなく自由度が高いため、ほかのアミノ酸残基とは異なる角度を示した。プロリン残基は側鎖が主鎖と結合し自由度がほとんどないため、ほかのアミノ酸残基とは異なる角度をもった。標準偏差は、ほとんどのアミノ酸残基は 40° 前後を示した。しかしながら、システイン、アスパラギン酸、イソロイシン、アスパラギン、セリン、スレオニン、バリンの ψ 値は、 40° 前後の値をもたなかった。また、イソロイシンとバリンは、標準偏差が 30° より低い値を示し、特定の角度に収束していた。

主鎖二面体角 (ϕ , ψ) の平均値は、分布の広がりの中から重心を求めた値である。つまり、 β ストランド構造の中それぞれのアミノ酸残基が、どのような角度に集まっているのかはわからない。そこで、主鎖二面体角の分布をヒストグラムで表現し最頻値を調べた (Figure 6-4)。

ヒストグラムでは、-180°から 180°を 5°間隔で区切ったときの出現頻度を調べた。

Table 6-1. β ストランド構造の中のアミノ酸残基の
二面体角の平均値と標準偏差

	Average		Standard deviation	
	ϕ	ψ	ϕ	ψ
Ala	-107.0	136.3	39.4	44.3
Cys	-113.9	128.6	33.8	58.4
Asp	-97.3	106.8	32.1	81.9
Glu	-104.2	133.0	33.7	38.3
Phe	-113.8	134.1	29.6	38.7
Gly	-52.8	34.0	116.7	154.3
His	-112.2	130.5	36.1	45.7
Ile	-110.0	128.6	22.1	24.8
Lys	-104.9	132.1	31.5	41.2
Leu	-103.1	131.8	25.8	31.7
Met	-110.7	132.3	31.5	43.5
Asn	-103.5	111.9	32.5	74.8
Pro	-66.5	138.9	10.9	42.3
Gln	-108.8	132.6	30.6	43.6
Arg	-109.5	133.6	33.5	40.1
Ser	-108.8	132.2	36.9	66.1
Thr	-111.4	129.8	27.7	65.9
Val	-112.2	130.5	23.1	29.8
Trp	-110.5	133.7	32.2	41.5
Tyr	-114.9	136.7	29.7	36.8

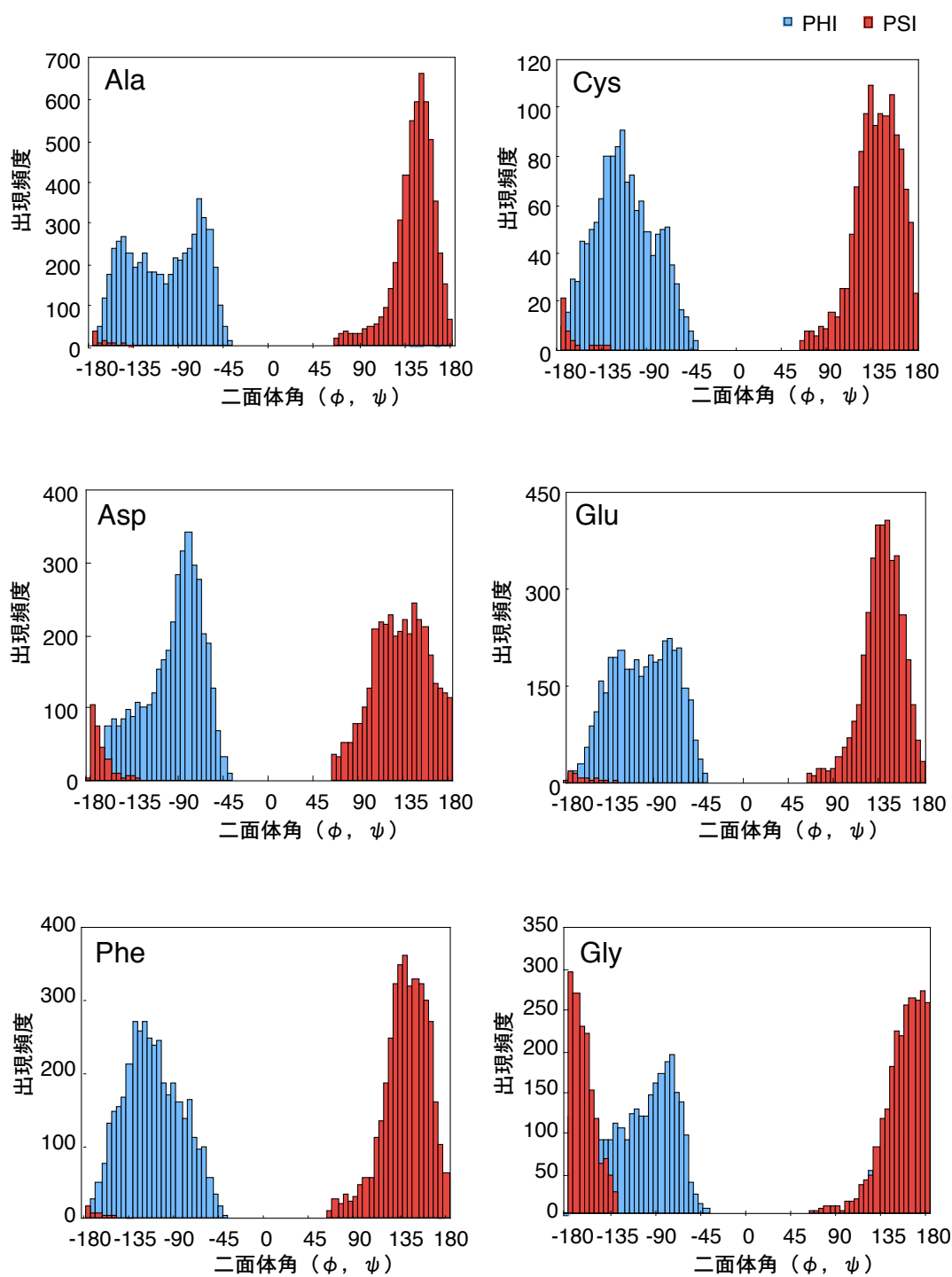


Figure 6-4. β ストランド構造の中の各アミノ酸残基の主鎖二面体角の分布 (その1)

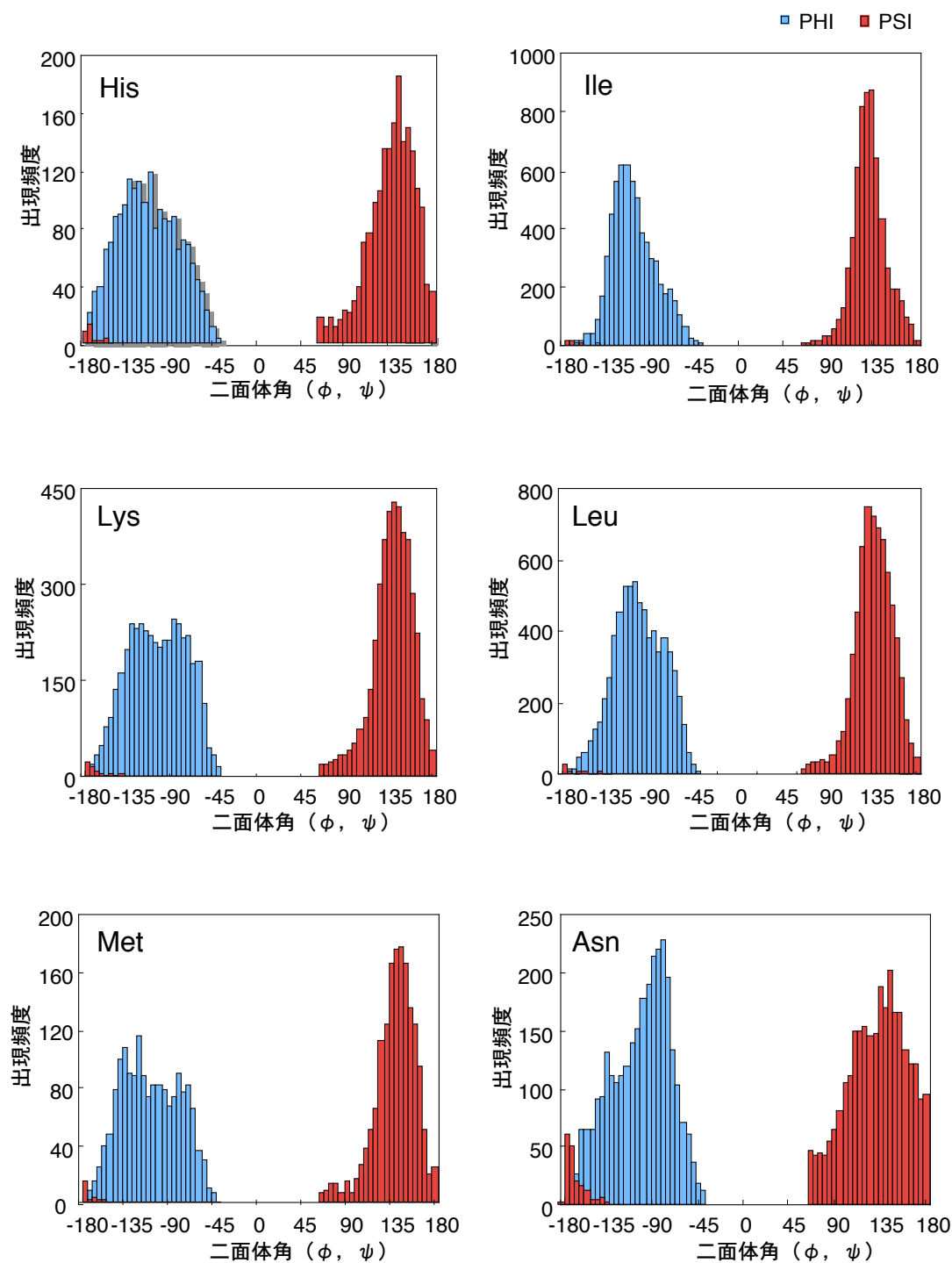


Figure 6-4. β ストランド構造の中の各アミノ酸残基の主鎖二面体角の分布 (その2)

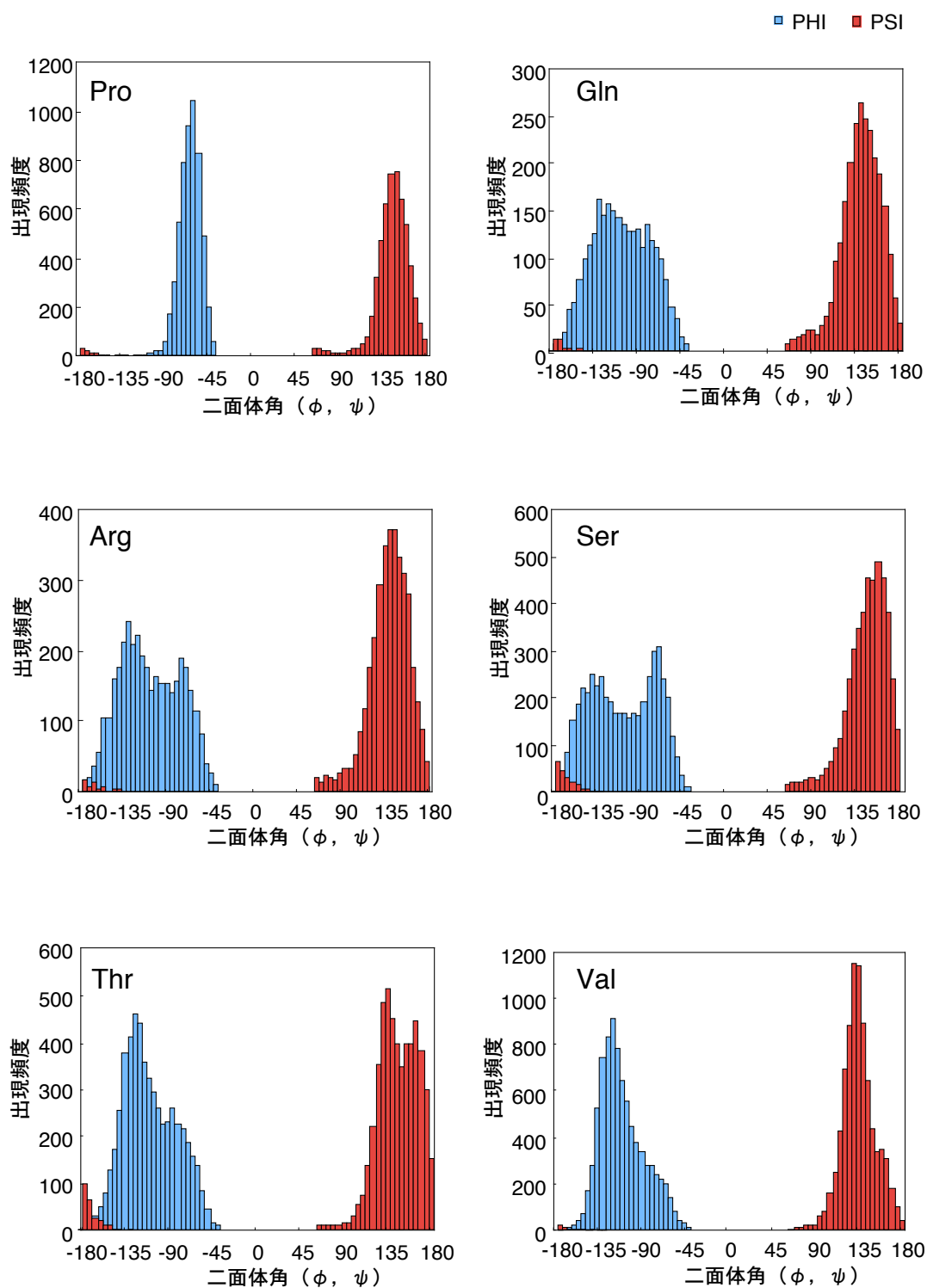


Figure 6-4. β ストランド構造の中の各アミノ酸残基の主鎖二面体角の分布 (その3)

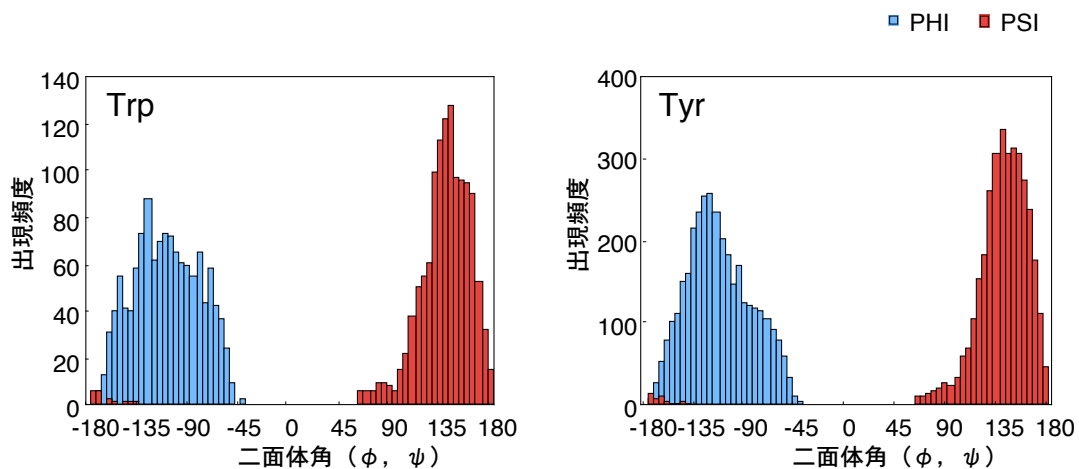


Figure 6-4. β ストランド構造の中の各アミノ酸残基の主鎖二面体角の分布（その4）

Figure 6-4 をみると大きく二つに分類できる。主鎖二面体角の最大値が一つの場合と、二つ以上の場合である。最大値が一つのアミノ酸残基は、イソロイシン、プロリン、バリンで、二つ以上はこれら三つの残基以外である。二つ以上の最大値を示すのアミノ酸残基は、もともと β ストランド構造が、シート形成と一本鎖の二種類を含んでいることから、これら二種類のコンホメーションが β ストランド構造の分布に現れていると考えられる。最大値が一つだけのアミノ酸残基は、シート形成でも一本鎖でも、ほとんど同じ角度をもつことを意味する。これらの解釈は、6-3節、6-4節で行なう。

6—3 シート形成 β ストランド構造の解析

3—2節の β ストランド構造の長さに対する出現頻度と同様に、シート形成 β ストランド構造の長さに対する出現頻度を調べた (Figure 6-5)。

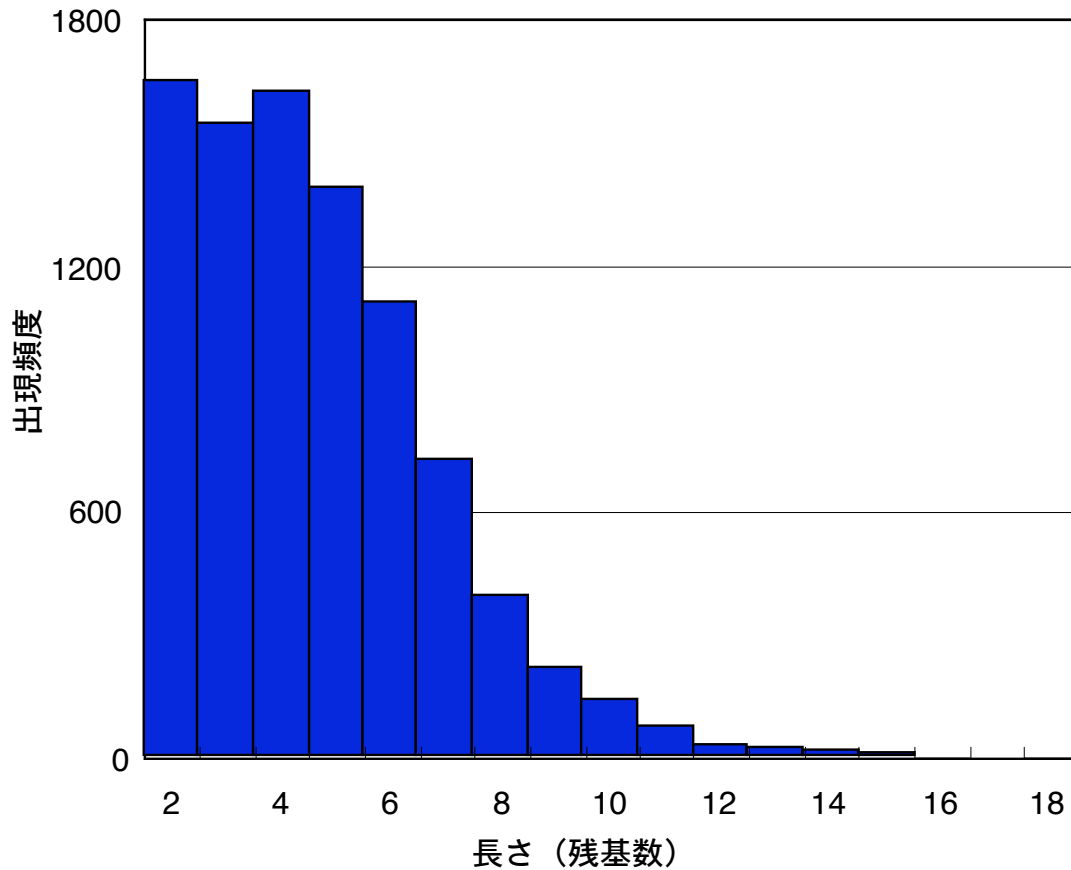


Figure 6-5. シート形成 β ストランド構造の長さに対する出現頻度

シート形成 β ストランド構造のアミノ酸残基は、最小で2残基、最大で18残基であった。また、平均残基数は4.6残基であった。 β ストランド構造の長さに対する出現頻度と比べて、アミノ酸残基が増えるほど出現頻度は減少しない。そのかわり、偶数残基ごとに最大値を示す傾向が見える。この結果はSternbergとThorntonの結果とは異なるが、ZhuとBlundellの結果と一致する^{29,67)}。偶数残基ごとに最大値が現れる要因は、 β シート構造の側鎖の向きから説明できる。 β シート構造の側鎖は、主鎖の進行方向に対して垂直方向に向いている。側鎖の向きは1残基ごとに上下に現れる。このことから、偶数残基ごとに最大値が見られるのは側鎖同士の分

子間相互作用の可能性が考えられる。奇数残基、たとえば、3残基の場合、N末端からC末端方向に対して、上に2残基分、下に1残基分の側鎖の構造を形成する。ところが、偶数残基、たとえば、4残基の場合、上に2残基分、下に2残基分の側鎖があるため、それら2残基同士で分子間相互作用を形成し、安定化していると考えられる。よって、偶数残基の長さをもつシート形成 β ストランド構造の組み合わせが見えていいると考えられる。この考えを明らかにするには、2残基の組み合わせの中に、分子間相互作用を見出す必要がある。

β シート構造の側鎖は、主鎖に進行方向に対して垂直方向に向いている。側鎖の向きは、2残基ごとに上下に現れるが、片面に同じ性質のアミノ酸残基が現れる場合、いわゆる両親媒性 β シート構造を形成し (Figure 6-6)、シート形成 β ストランド構造のヒストグラム (Figure 6-5) での特徴と関係していると考えられる。つまり、シート形成 β ストランド構造は、比較的両親媒性の構造が多い。

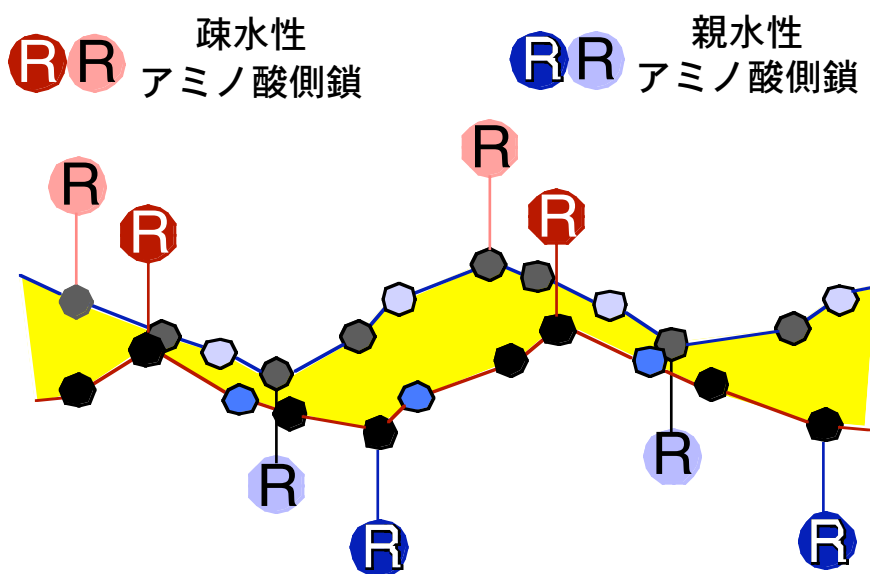


Figure 6-6. 両親媒性のシート構造

つぎに，シート形成 β ストランド構造の中に現れるアミノ酸残基の傾向指数について調べた (Table 6-2)。

Table 6-2. シート形成 β ストランド構造のアミノ酸残基の傾向指数

AA	Total	β -sheet	P _{sheet}
A	17871	2679	0.78
C	2872	726	1.31
D	12802	1303	0.53
E	13952	1863	0.69
F	8507	2432	1.48
G	16012	1871	0.60
H	4971	972	1.01
I	12316	4303	1.81
K	12685	2013	0.82
L	19113	4094	1.11
M	4744	964	1.05
N	9556	1105	0.60
P	10090	819	0.42
Q	8367	1252	0.77
R	10689	1830	0.89
S	12535	1959	0.81
T	12026	2784	1.20
V	15012	5631	1.94
W	2973	739	1.29
Y	7641	2141	1.45
Total	214734	41480	

赤色の数は，1.45 以上，青色の数は，0.45 以下の傾向指数である。Table 6-2 の結果は，Zhu

と Blundell の結果とほぼ同じ傾向指数を得ている²⁹⁾。主にアミノ酸残基の側鎖が β 分岐している残基ほど傾向指数が高い^{26,29,32)}。

つぎに、シート形成 β ストランド構造に対して、それぞれのアミノ酸残基の主鎖二面体角の平均値と標準偏差を調べた (Table 6-3)。

Table 6-3. シート形成 β ストランド構造のアミノ酸残基の二面体角の平均値と標準偏差

	Average		Standard deviation	
	ϕ	ψ	ϕ	ψ
Ala	-124.5	139.3	31.7	33.1
Cys	-122.4	129.9	26.4	54.1
Asp	-109.8	118.6	27.9	59.9
Glu	-115.9	133.6	28.5	28.2
Phe	-120.5	136.6	25.3	29.4
Gly	-58.0	49.9	127.9	148.3
His	-121.4	133.7	26.6	29.9
Ile	-115.0	128.8	18.0	20.1
Lys	-114.6	134.1	28.0	27.8
Leu	-111.0	130.0	21.0	24.4
Met	-119.7	135.2	27.6	30.3
Asn	-113.0	125.4	25.1	41.1
Pro	-70.2	138.4	10.4	26.3
Gln	-118.7	135.4	24.3	24.9
Arg	-118.2	135.1	26.3	25.0
Ser	-123.9	134.8	28.2	54.8
Thr	-120.1	131.5	20.8	52.0
Val	-117.4	130.9	18.1	23.6
Trp	-117.9	136.5	24.8	35.3
Tyr	-122.5	138.1	22.7	30.0

ここでは、二面体角の分布を、平均値と標準偏差であらわす。その結果、どのアミノ酸残基も、 ϕ 角は、おおよそ 120° 、 ψ 角は、 130° の周辺に集まっていて、標準偏差も、おおよそ 30° の範囲内であった。ただし、グリシンは、自由度が高く、プロリンは側鎖が主鎖と結合しているので主鎖二面体角の角度が限定され、異なる分布を示した。

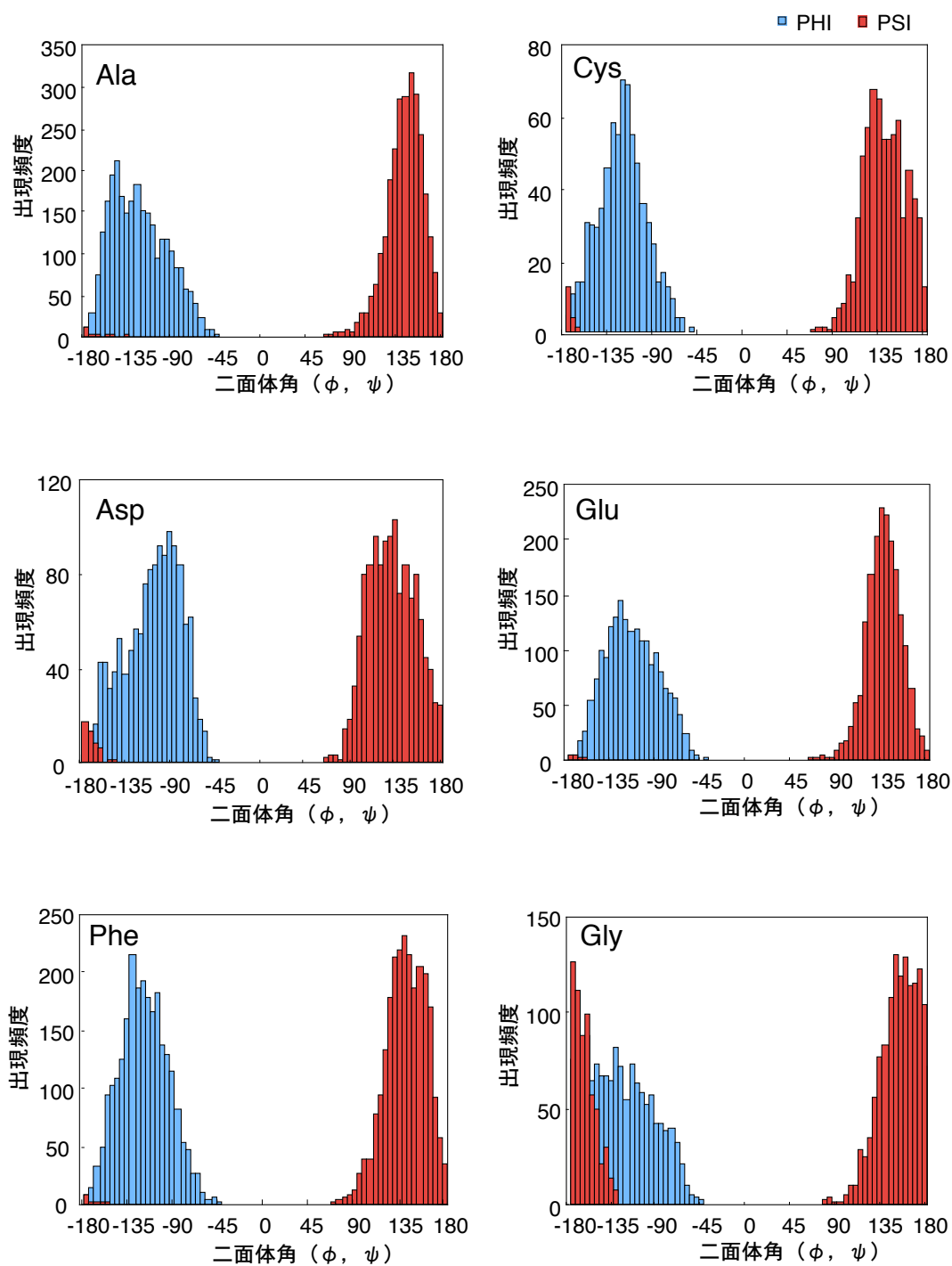


Figure 6-7. シート形成 β ストランド構造の中の各アミノ酸残基の主鎖二面体角の分布 (その1)

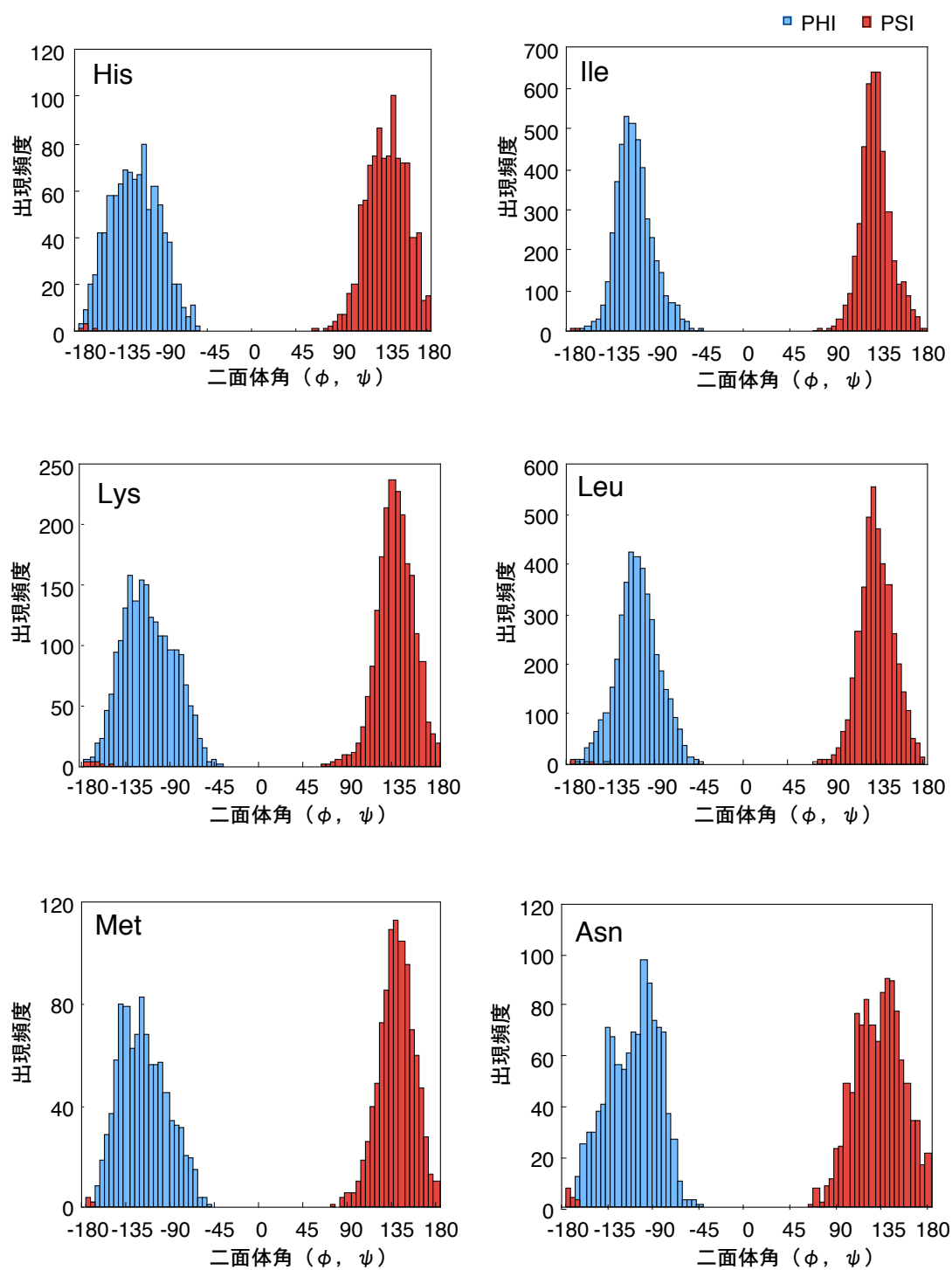


Figure 6-7. シート形成 β ストランド構造の中の各アミノ酸残基の主鎖二面体角の分布 (その2)

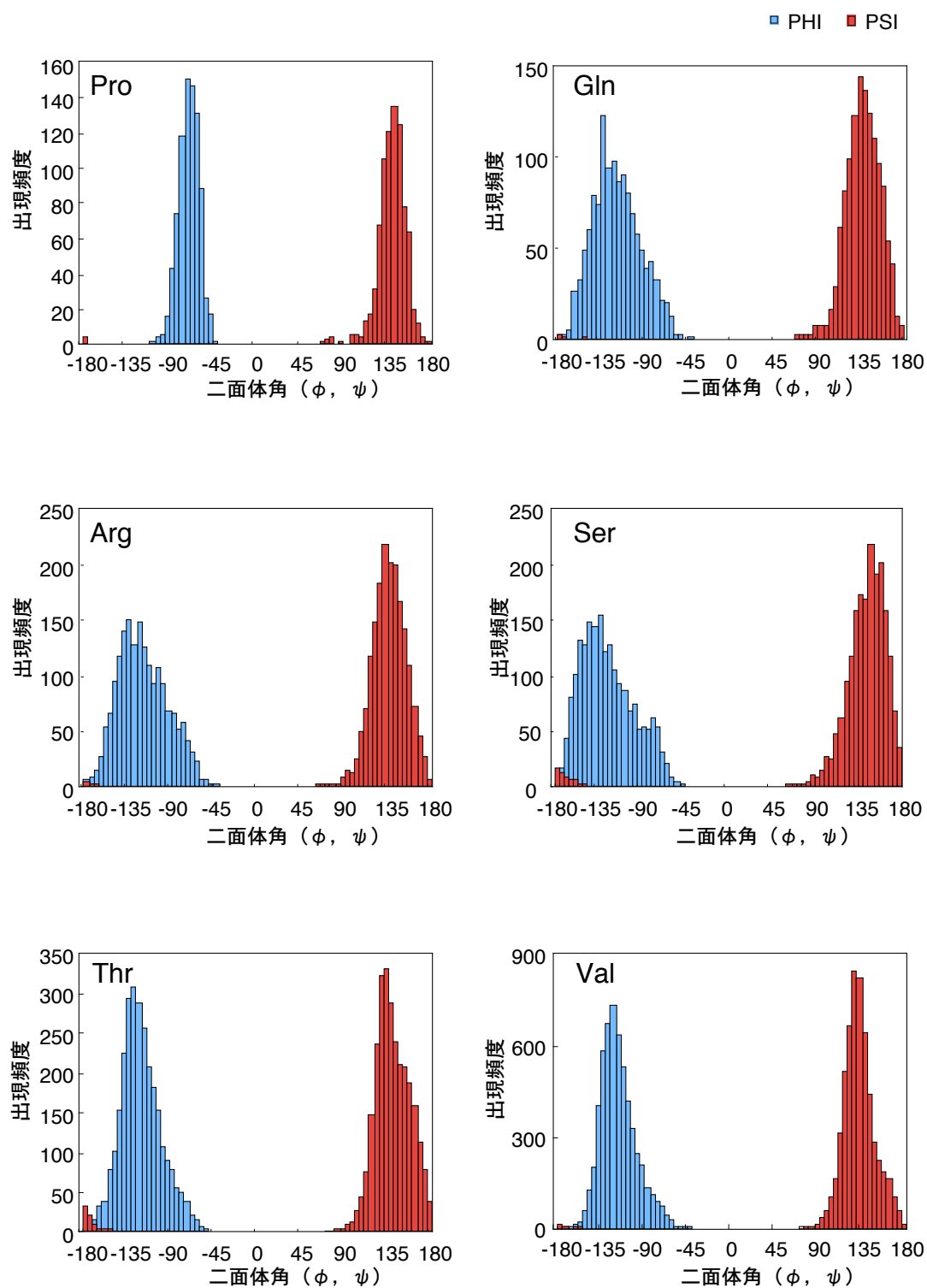


Figure 6-7. シート形成ストランド構造の中の各アミノ酸残基の主鎖二面体角の分布 (その3)

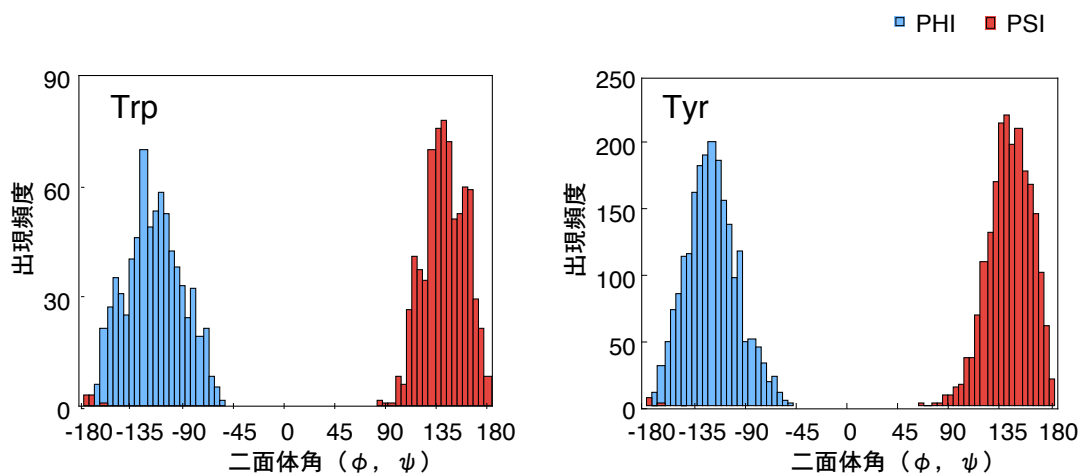


Figure 6-7. シート形成 β ストランド構造の中の各アミノ酸残基の主鎖二面体角の分布（その4）

統計的な見方として、平均値は全体の散らばり具合から判断する評価方法であり、どの角度が構造中によく見られるのかについて知ることはできない。そこで、 β ストランド構造と同様に、シート形成 β ストランド構造の中のアミノ酸残基の分布を調べた（Figure 6-7）。

シート形成 β ストランド構造の中に現れるアミノ酸残基の主鎖二面体角の分布（Figure 6-7）と β ストランド構造の中に現れるアミノ酸残基の分布（Figure 6-4）を比べると、 β ストランド構造で見られた二面体角の複数の最大値は、シート形成 β ストランド構造の最大値と一致する。また、 β ストランド構造で二面体角の最大値が一つだけのものに対しては、シート形成 β ストランド構造の二面体角の最大値と一致する。このことから、 β ストランド構造で見られた二面体角の最大値は、シート形成 β ストランド構造の最大値を含んでいるといえる。

つぎに、主鎖二面体角の最頻値を求めた（Table 6-4）。ただし、角度の間隔は 1° である。この結果から、ほとんどのアミノ酸残基の主鎖二面体角は、平均値に近い値だった。しかし、それでも、いくつかのアミノ酸残基については、平均値とは異なる値だった。その中でも、アスパラギン酸残基の主鎖二面体角の ϕ 角が、平均値からも離れていて、他のアミノ酸残基の値と比べても、異なる値だった。アスパラギン酸残基は、極性アミノ酸残基であるが、他の極性アミノ酸残基は、平均値に近い値を示している。したがって、アスパラギン酸残基が、なぜ他のア

ミノ酸残基と異なる値を示したかについてはわからない。

Table 6-4. シート形成 β ストランド構造のアミノ酸残基の二面体角の最頻値

	ϕ	ψ
Ala	-147	146
Cys	-121	127
	-116	140
Asp	-88	125
Glu	-134	133
	-123	
Phe	-130	130
Gly	-173	150
His	-113	141
Ile	-121	129
Lys	-120	139
Leu	-118	128
Met	-135	137
Asn	-104	140
Pro	-70	136
Gln	-130	132
Arg	-131	141
Ser	-141	149
Thr	-127	135
Val	-124	131
Trp	-126	134
	-112	
	-106	
Tyr	-120	137

6—4 一本鎖 β ストランド構造の解析

一本鎖 β ストランド構造のアミノ酸残基数に対する出現頻度を調べた (Figure 6-8)。一本鎖 β ストランド構造のアミノ酸残基数は、 β ストランド構造やシート形成 β ストランド構造と同様に、最小で2残基、最大で18残基であった。しかしながら、平均残基数は、2.8残基だった。シート形成 β ストランド構造のような特徴的なヒストグラムではなく、長くなるにつれて出現頻度が減少している。これは、Eswar らの結果と一致する³²⁾。また、最大の出現頻度は、2残基であり、シート形成 β ストランド構造と同じ残基数だった。この結果は、Eswar らの結果と異なるが、3—3節でも述べたように、彼らは、4残基以上連続した構造を β 構造としている。したがって、Figure 6-6 から長さが長くなるにつれ出現頻度が減少していることを考えると、もし、4残基以上であれば、出現頻度が最大の長さは彼らと同じ結果となる。

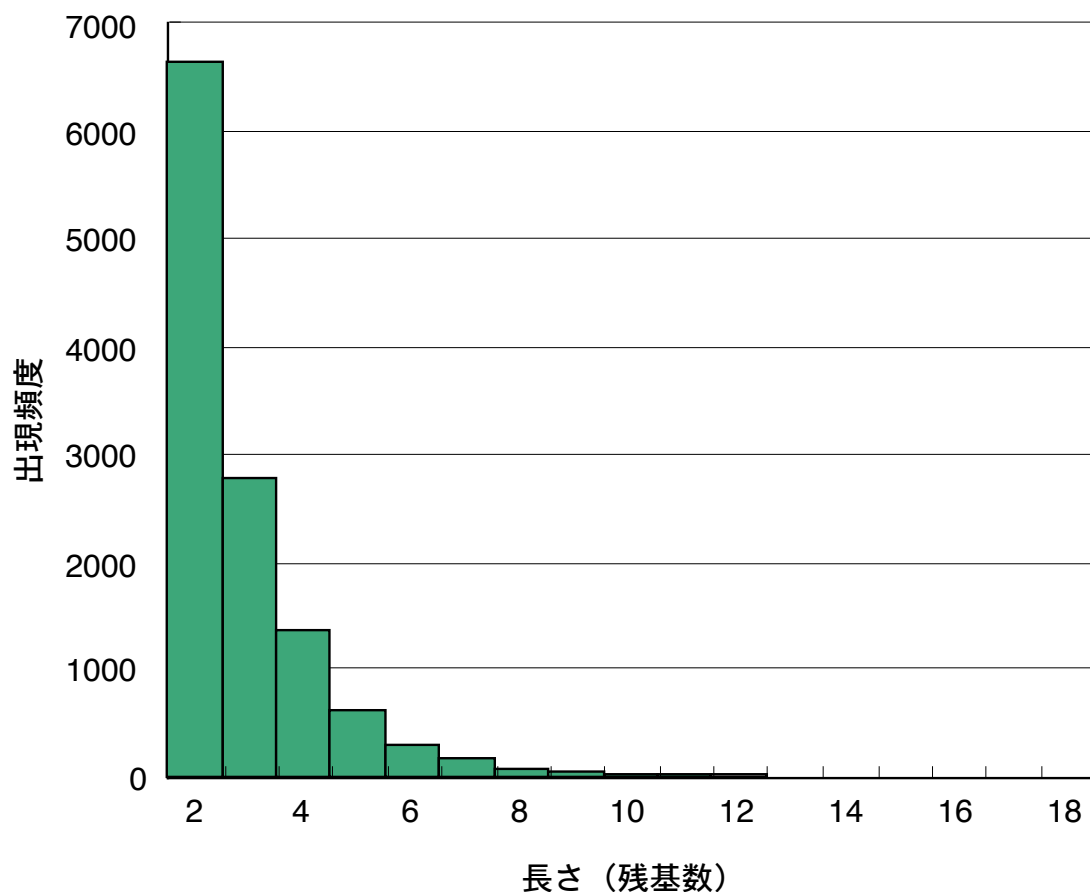


Figure 6-8. 一本鎖 β ストランド構造の長さに対する出現頻度

つぎに，シート形成 β ストランド構造と同様に，一本鎖 β ストランド構造の中のアミノ酸残基の傾向指数について調べた（Table 6-5）。

Table 6-5. 一本鎖 β ストランド構造のアミノ酸残基の傾向指数

AA	Total	Single	P _{single}
A	17871	2190	0.78
C	2872	434	0.96
D	12802	2079	1.03
E	13952	1754	0.80
F	8507	1288	0.96
G	16012	1712	0.68
H	4971	768	0.98
I	12316	1654	0.85
K	12685	1924	0.96
L	19113	2425	0.81
M	4744	614	0.82
N	9556	1526	1.02
P	10090	4473	2.82
Q	8367	1132	0.86
R	10689	1529	0.91
S	12535	2297	1.17
T	12026	2223	1.18
V	15012	2229	0.94
W	2973	421	0.90
Y	7641	1101	0.92
Total	214734	33773	

シート形成 β ストランド構造では，バリン，イソロイシン，フェニルアラニンの三つのアミ

ノ酸残基の傾向指数が大きい値を示した (Table 6-2)。しかしながら、一本鎖 β ストランド構造では、それらのアミノ酸残基に特徴が見られない (Table 6-5)。プロリンの傾向指数が高いという特徴を示した^{32,68)}。

つぎに、シート形成 β ストランド構造のときと同じように、一本鎖 β ストランド構造の中のアミノ酸残基に対する主鎖二面体角の平均値と標準偏差を調べた (Table 6-6)。

Table 6-6. 一本鎖 β ストランド構造のアミノ酸残基の二面体角の平均値と標準偏差

	Average		Standard deviation	
	ϕ	ψ	ϕ	ψ
Ala	-88.5	132.1	38.3	54.1
Cys	-102.5	127.0	40.3	61.3
Asp	-90.9	104.9	33.3	82.9
Glu	-93.2	133.0	34.8	45.0
Phe	-102.0	129.6	32.9	48.4
Gly	-47.0	19.5	106.5	158.0
His	-102.6	125.5	42.3	61.0
Ile	-98.9	128.6	26.1	33.1
Lys	-95.8	129.6	32.1	51.3
Leu	-91.3	134.0	28.1	40.9
Met	-98.9	127.8	33.0	57.2
Asn	-98.5	105.3	35.6	84.1
Pro	-65.8	139.3	10.8	43.4
Gln	-99.6	130.9	32.3	51.4
Arg	-100.8	131.7	37.9	50.8
Ser	-97.6	129.6	39.0	73.3
Thr	-102.8	128.6	31.1	76.7
Val	-101.5	130.1	28.5	38.3
Trp	-101.2	128.7	39.5	49.8
Tyr	-101.5	134.4	35.4	43.4

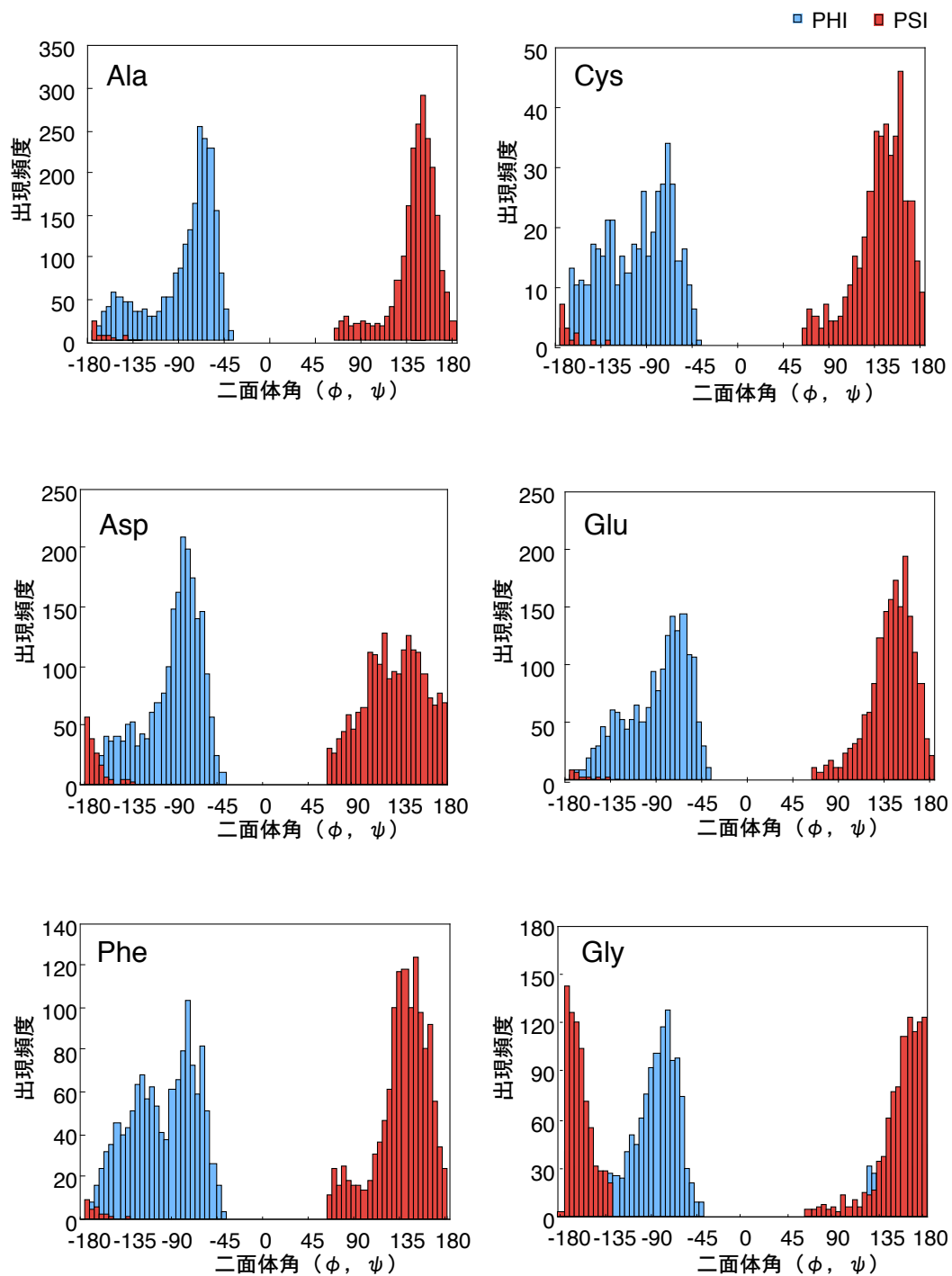


Figure 6-9. 一本鎖 β ストランド構造の中の各アミノ酸残基の主鎖二面体角の分布 (その1)

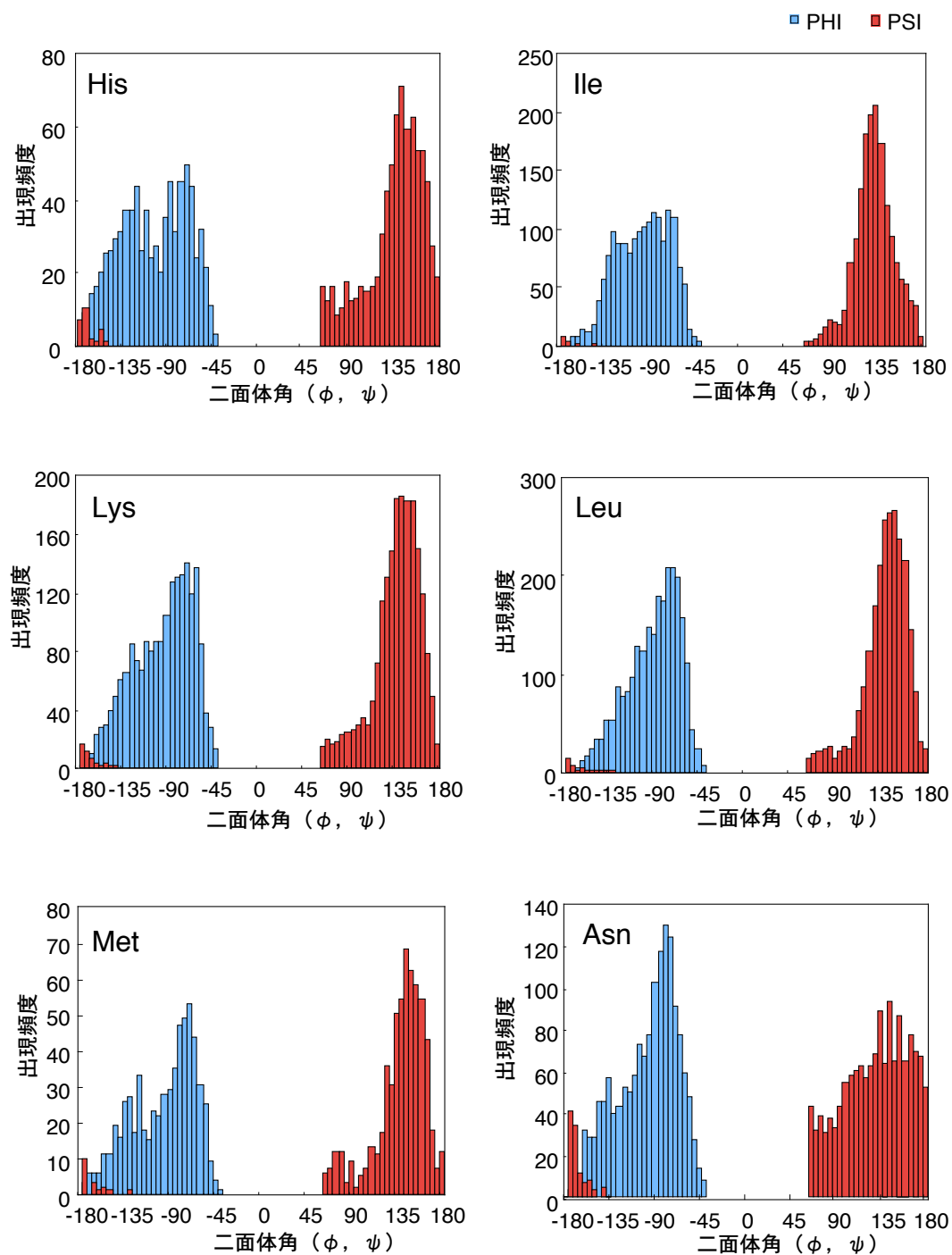


Figure 6-9. 一本鎖 β ストランド構造の中の各アミノ酸残基の主鎖二面体角の分布 (その2)

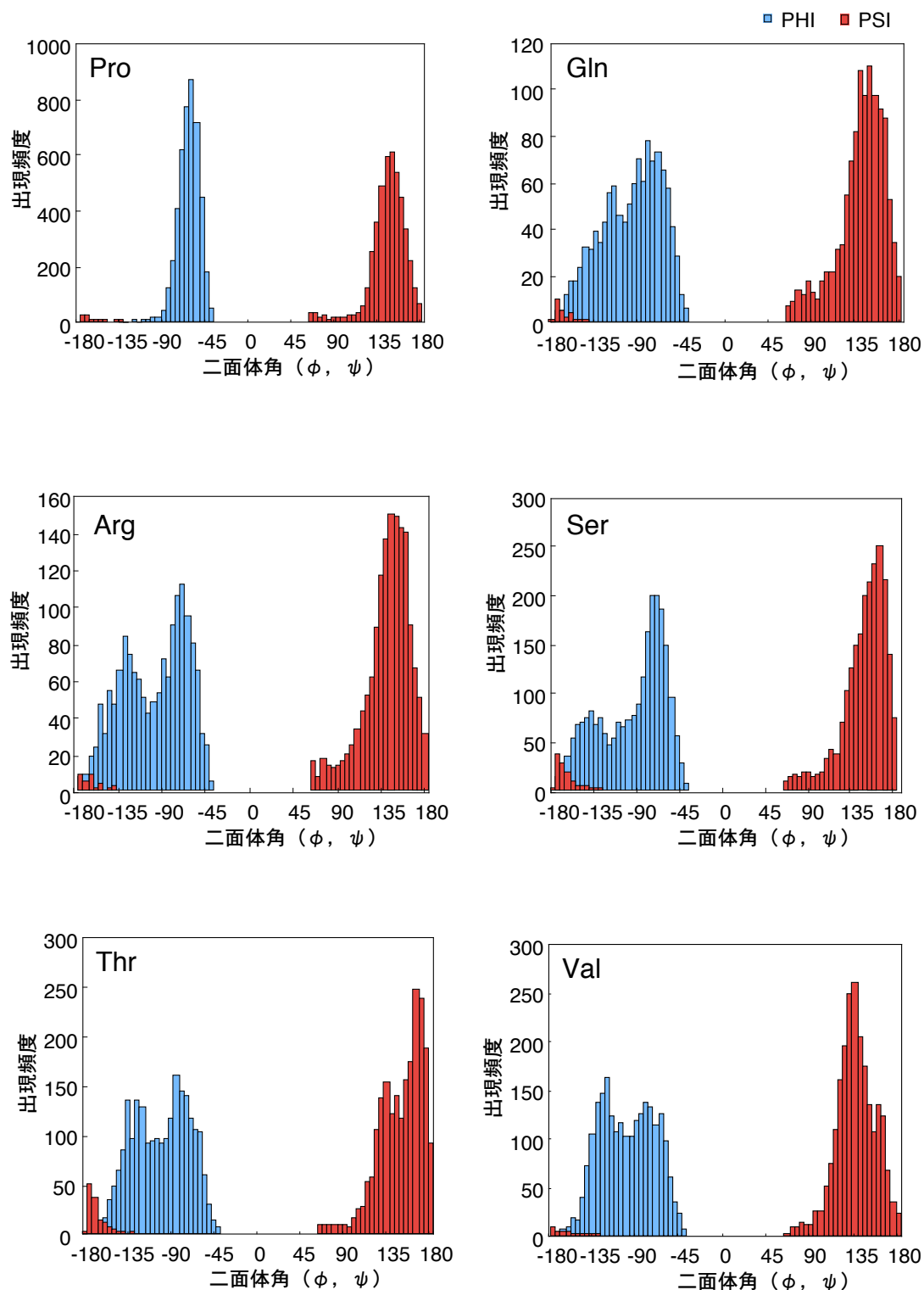


Figure 6-9. 一本鎖 β ストランド構造の中の各アミノ酸残基の主鎖二面体角の分布 (その3)

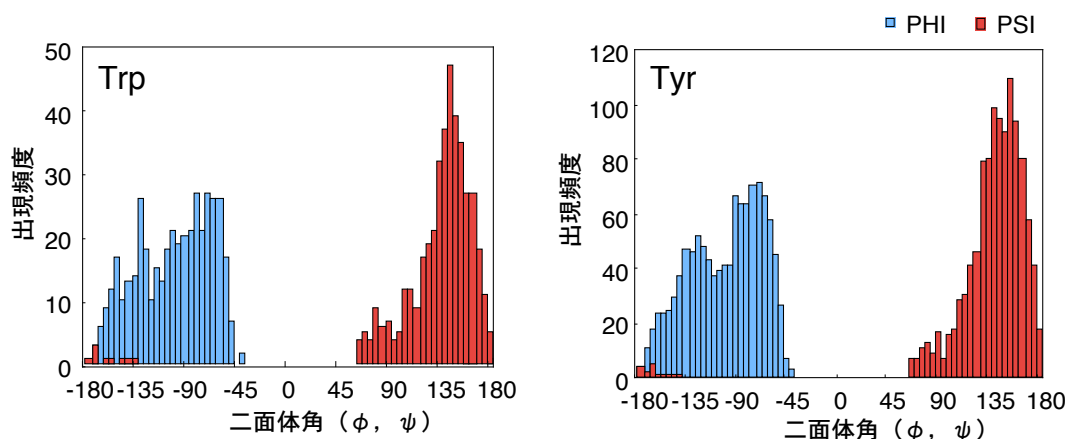
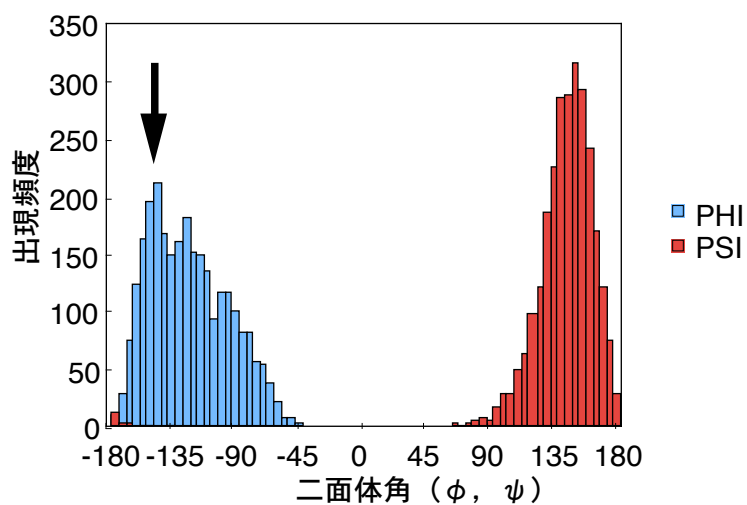


Figure 6-9. 一本鎖 β ストランド構造の中の各アミノ酸残基の主鎖二面体角の分布（その4）

ϕ 角は、おおよそ -100° 、 ψ 角は、おおよそ 125° だった。シート形成 β ストランド構造の主鎖二面体角と比べると、 ϕ 角に違いが見られるものの、 ψ 角は、ほぼ同じ値である。また、 β ストランド構造やシート形成 β ストランド構造と同様に、一本鎖 β ストランド構造の中のアミノ酸残基に対しての主鎖二面体角の分布を調べた（Figure 6-9）。

6-3節と同様に、 β ストランド構造の二面体角の分布（Figure 6-4）と、一本鎖 β ストランド構造（Figure 6-9）の二面体角の分布を比べると、 β ストランド構造の分布の最大値は、一本鎖 β ストランド構造の最大値とすべて重なる。つまり、 β ストランド構造の二面体角の分布には、シート形成だけでなく、一本鎖 β ストランド構造の二面体角の分布も含んでいる。また、Figure 6-7のシート形成 β ストランド構造とも比較した。その結果、もっとも多く現れた二面体角は違っていた。しかしながら、グリシンとロイシン以外のアミノ酸残基において、同じような角度に最大値が存在した。その例をアラニンで示す（Figure 6-10）。

(a) シート形成



(b) 一本鎖

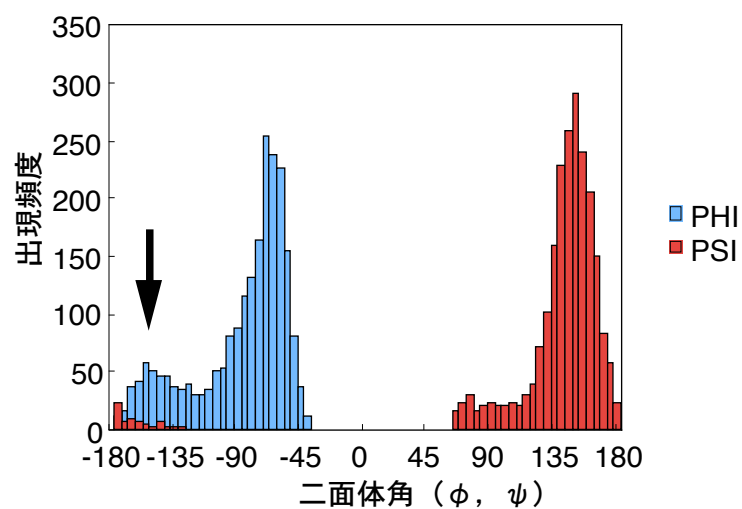


Figure 6-10. シート形成と一本鎖のベータストランド構造の中のAlaのピークの一致

以上の結果から、シート形成でも一本鎖でも、伸びた構造の中のアミノ酸残基の主鎖二面体角は、共通する角度を潜在的に有している。

つぎに、主鎖二面体角の分布から、最頻値を求めた (Table 6-7)。ただし、二面体角の角度の間隔は、 1° 刻みとした。

Table 6-7. 一本鎖 β スtrand構造のアミノ酸残基の二面体角の最頻値

	ϕ	ψ
Ala	-71	147
Cys	-73	136
Asp	-80	144
Glu	-78	152
Phe	-78	134
Gly	-78	-173
His	-71	131
Ile	-125	129
Lys	-73	143
Leu	-67	145
Met	-72	144
Asn	-79	153
Pro	-65	145
Gln	-95	138
Arg	-70	149
Ser	-71	163
Thr	-86	163
Val	-123	133
Trp	-85	144
Tyr	-70	147
		153
		154

その結果、シート形成 β スtrand構造とは異なる最頻値だった。これは、シート形成 β スtrand構造は、 β スtrand構造の主鎖のアミド間の水素結合が関わっているため、その結合が主鎖二面体角に反映された結果と考える。一方、一本鎖 β スtrand構造は、ほかのセグメントと水素結合していないため、シート形成とは異なる主鎖二面体角をとったと考えられる。

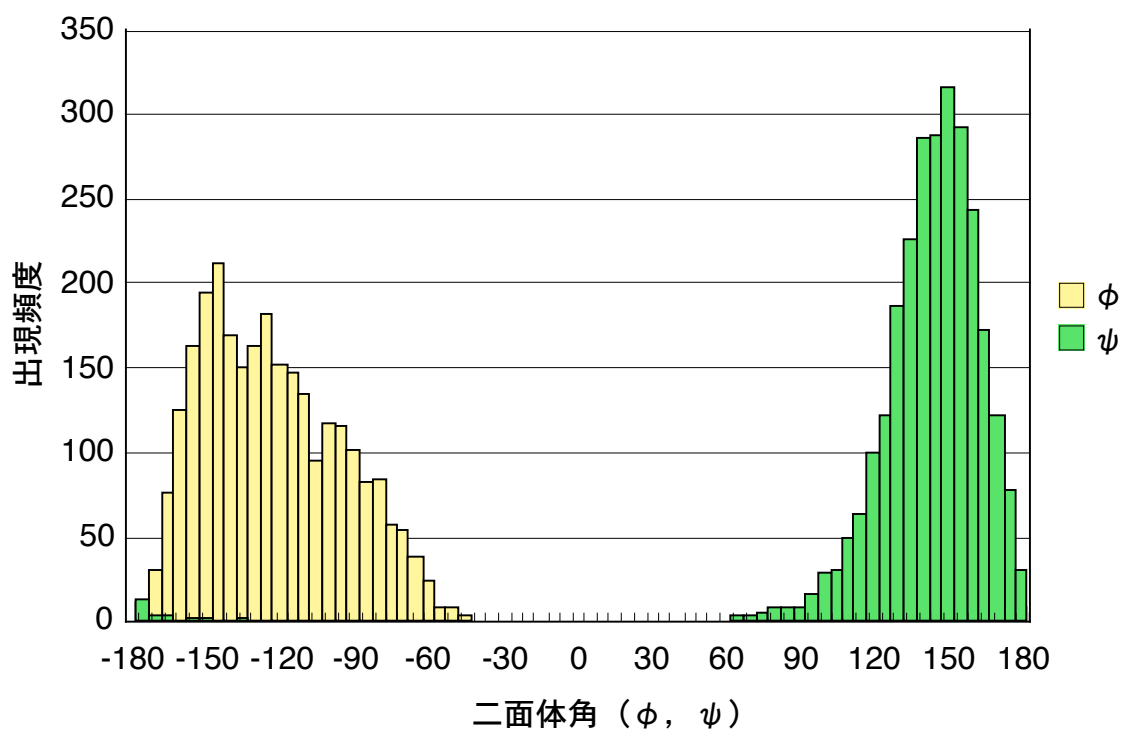
この最頻値の違いは、Figure 6-7 と Figure 6-9 の分布からも判断できる。ところで、一本鎖 β ストランド構造のバリン、グリシン、プロリン、イソロイシンは、他のアミノ酸残基とは異なる最頻値だった。しかしながら、バリン、プロリン、イソロイシンは、シート形成における最頻値に近い値だった。これは、Figure 6-7 と Figure 6-9 の分布からも判断できる。したがって、バリン、プロリン、イソロイシンは、 β ストランド構造が、シート形成でも一本鎖でも関係なく、伸びた構造に対して、特定の主鎖二面体角をもっている。また、プロリンは、傾向指数でもっとも高い値をもつアミノ酸残基でもある。つまり、特定の角度に制限を受けるプロリンは、他の二次構造と水素結合していない一本鎖 β ストランド構造にとって、必要なアミノ酸残基であると考えられる。Eswar らは、一本鎖 β ストランド構造は、そのコンホメーションを保つために、まわりの水分子や蛋白質の極性残基の側鎖と水素結合していることを指摘した³²⁾。

6—5 主鎖二面体角による構造安定化エネルギーの算出

6—3、6—4節で、シート形成と一本鎖の β ストランド構造に対するアミノ酸残基の主鎖二面体角の分布を調べた。その結果、それぞれのアミノ酸残基の ψ 角は、ほぼ同じ分布であったにもかかわらず ϕ 角に違いが見られた (Figure 6-11)。このような分布の違いは、立体構造エネルギーに違いがあると考えられる。そこで、これらの主鎖二面体角でもっとも出現頻度が高い値を使って、それぞれの安定化エネルギーを計算した (Figure 6-12)。

Figure 6-12 は、アラニンのジペプチド Ala-Ala をモデル分子として、シート形成と (-147°, 146°)、一本鎖を (-71°, 147°) 表わした。シート形成 β ストランド構造の一本鎖 β ストランド構造の安定化エネルギーの差は、0.76 kcal/mol で、一本鎖 β ストランド構造の方が安定であった。おそらく、 β シート構造ではストランド間での水素結合がエネルギー安定化に寄与しているため、シート形成の β ストランド構造の方は不安定であると考えられる。

(a) シート形成 β ストランド構造



(b) 一本鎖 β ストランド構造

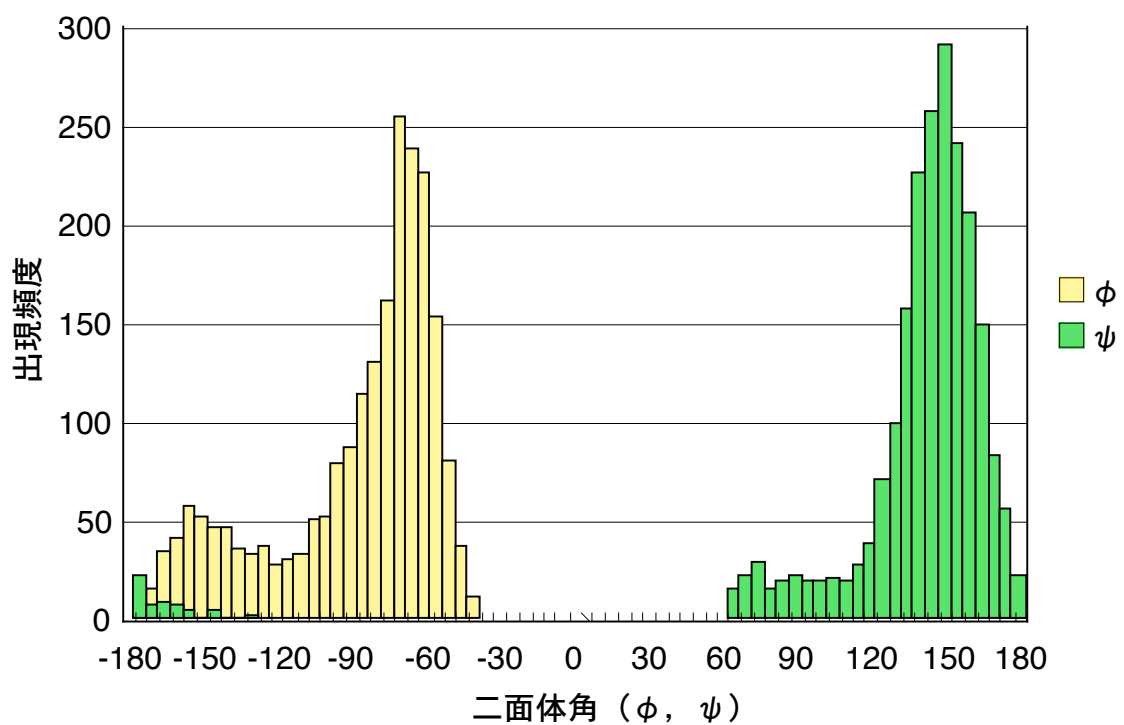


Figure 6-11. Alaの主鎖二面体角の分布と出現頻度

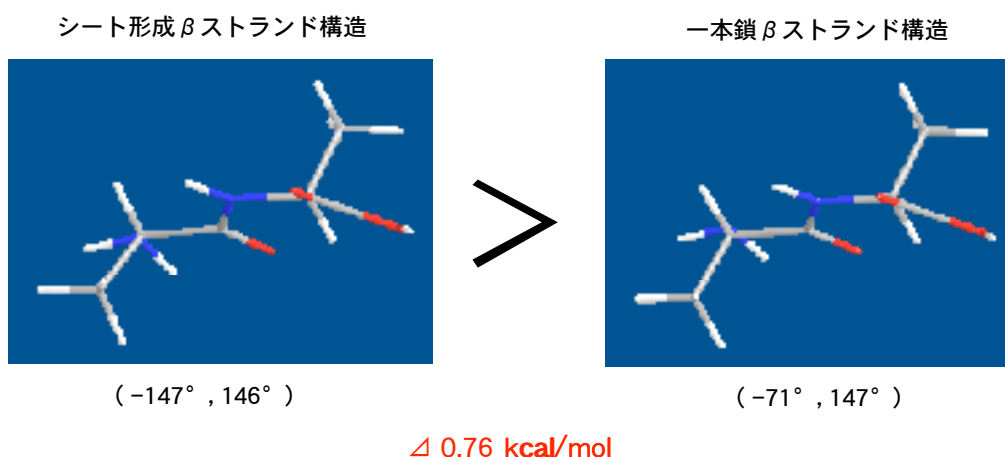


Figure 6-12. ジペプチドAla-Alaのモデル分子の分子軌道計算

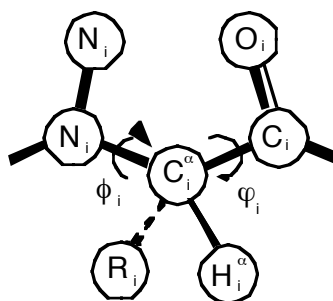
以上, この章では, ProSSA 法の β ストランド構造をシート形成と一本鎖の二種類の β ストランド構造に分け, それぞれの構造に関する長さや主鎖二面体角の分布について調べた。その結果, シート形成 β ストランド構造は偶数残基ごとに長さの最大値が存在することがわかった。両親媒性の構造を形成している場合が多いといえる。また, シート形成と一本鎖の主鎖二面体角の分布から, 互いに出現頻度をもっとも多い角度は異なるものの, 最大値の場所が同じような角度のところに現れたアミノ酸残基が存在した。このことから, シート形成でも一本鎖でも伸びた構造の中のアミノ酸残基は, ある特定の角度を潜在的に有していると考えられる。さらに, シート形成と一本鎖の主鎖二面体角の分布の中で, 最頻値を使ってモデルペプチド Ala-Ala の分子軌道計算から, シート形成よりも一本鎖の方が, エネルギー的に安定であることを明らかにした。シート形成 β ストランド構造と一本鎖 β ストランド構造について, このような結果は, これまで例がなく, 今後は, 実際の蛋白質分子をデザインする際の構造情報として役に立つといえる。

第7章 ジペプチドを単位パラメータとする二次構造の統計的解析

7-1 蛋白質解析の際の基本単位の意味

遺伝子は四つの塩基によって構成されており、配列上の基本単位は3塩基一組のコドンとして現れ、遺伝子を設計するのは当然のことながらコドン単位である。蛋白質は基本成分である20種類のアミノ酸残基一つ一つを基本単位としている。よって、蛋白質をアミノ酸配列から設計する際は、アミノ酸残基を単位として考えるのが当然のこととされている。蛋白質の基本構成成分が20種類のアミノ酸残基なので、解析の対象としてアミノ酸残基に注目するのは、理にかなっている。しかしながら、この常識的な取り組みに反して、アミノ酸残基を基本単位とするのではなく、ジペプチド^(69,70)、トリペプチド⁽⁷¹⁻⁷⁴⁾、ペンタペプチド⁽⁴⁷⁾などを基本単位として解析した研究もある。この章では、蛋白質の分子設計を指向するには基本単位としてジペプチドを単位とする場合の合理性を述べ、ジペプチド単位での二次構造解析を行なった (Figure 7-1)。

(a) アミノ酸残基 AA_i



(b) ジペプチド $X_i X_{i+1}$

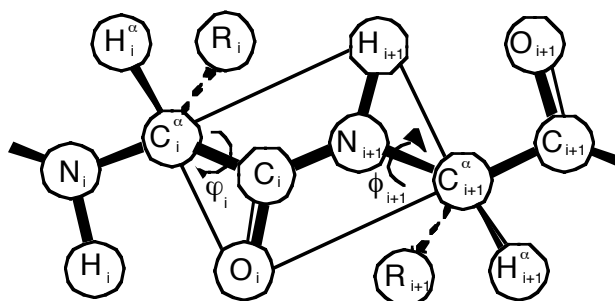


Figure 7-1. アミノ酸残基とジペプチド

ジペプチドを基本単位として提案する理由は三つである。一つは、二次構造形成に関わっている水素結合はアミド間の相互作用である。そして、このアミドの性質は両側のアミノ酸側鎖で特徴づけられる。二つ目は、コンホメーション解析で使われる円偏光二色性 (CD) 解析や赤外線 (IR) 解析は、それぞれアミドの吸収と振動を見ている。したがって、アミドを含めたジペプチドで解析した方が理にかなっている。三つ目は、ジペプチドは最小の配列情報をもつ。これらのことから、ジペプチドは二次構造を解析する上での最小の基本単位と考えるのが妥当である。

この章では、まず、ジペプチドに関する Chou-Fasman パラメータを算出し、さらに、Richardson パラメータを算出した。さらに、主鎖二面体角についても解析した。ただし、主鎖二面体角は、アミノ酸残基 AA_i のときは (ϕ_i, ψ_i) であったが、ジペプチド X_iX_{i+1} の場合は、 (ψ_i, ϕ_{i+1}) で解析した。また、この章で解析に用いた蛋白質は、二次構造の空間的配置に片寄りのない 1081 種類の蛋白質を使った¹¹⁾。

7—2 二次構造にみられるジペプチド単位構造の分布

まず，1081 蛋白質の二次構造に対するアミノ酸残基の傾向指数を示す（Table 7-1）。

Table 7-1. 蛋白質の二次構造に対するアミノ酸残基の傾向指数

AA	Total	Helix	Strand	P _{helix}	P _{strand}
A	17871	10022	5296	1.26	0.79
C	2872	1170	1260	0.91	1.17
D	12802	5663	3849	0.99	0.80
E	13952	7671	3859	1.23	0.74
F	8507	3577	3886	0.94	1.22
G	16012	5486	3967	0.77	0.66
H	4971	2200	1880	0.99	1.01
I	12316	4999	6190	0.91	1.34
K	12685	6121	4227	1.08	0.89
L	19113	10225	6858	1.20	0.96
M	4744	2428	1663	1.15	0.94
N	9556	4094	2942	0.96	0.82
P	10090	2363	5466	0.52	1.45
Q	8367	4525	2568	1.21	0.82
R	10689	5405	3592	1.13	0.90
S	12535	5355	4711	0.96	1.00
T	12026	4613	5424	0.86	1.20
V	15012	5453	8200	0.81	1.46
W	2973	1338	1234	1.01	1.11
Y	7641	3181	3398	0.93	1.19
Total	214734	95889	80470		

二次構造の中のアミノ酸残基に対する傾向指数は，Chou-Fasman パラメータとよばれ，1.0 を境に，1.0 より高ければ二次構造の中に現れやすく，1.0 より低ければ現れにくいことを意味する。金光は，二次構造に対するアミノ酸残基の傾向指数は， α ヘリックス構造では，1.0 へ近づいたものの， β ストランド構造は，依然として，傾向の高いアミノ酸残基の連続から構成されることを指摘した¹¹⁾。

つぎに，1081 種類の蛋白質に対するジペプチドの数を示す（Table 7-2）。

Table 7-2. 1081種類の蛋白質の中のジペプチドの数

$i+1$ i	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	1728	230	1007	1184	652	1294	405	1019	1047	1683	437	723	685	718	941	963	944	1256	257	589	17762
C	218	53	164	153	115	234	94	147	131	243	57	140	152	124	152	188	171	178	32	101	2847
D	1085	155	623	840	605	1044	259	859	693	1142	255	511	685	409	576	689	667	909	209	512	12727
E	1183	148	778	1015	536	905	268	830	982	1355	367	637	484	554	708	683	730	1024	215	449	13851
F	628	126	554	547	299	627	208	456	530	740	177	402	409	305	386	557	514	565	118	306	8454
G	1250	232	907	886	676	1241	369	995	1055	1316	344	671	573	616	788	1003	977	1149	209	638	15895
H	363	87	270	279	247	400	144	309	223	464	109	186	313	182	236	268	258	327	82	189	4936
I	1042	157	822	804	397	887	295	656	773	960	222	619	629	458	609	769	795	817	130	397	12238
K	1038	130	783	926	484	824	253	740	813	1149	294	582	574	473	566	706	676	926	159	447	12543
L	1655	240	1149	1250	615	1331	452	904	1257	1630	412	906	1032	798	1100	1198	1092	1153	205	612	18991
M	414	56	291	297	184	341	96	274	330	396	117	233	247	181	250	237	269	320	37	145	4715
N	751	133	525	509	377	789	217	604	527	837	214	432	580	348	412	512	516	701	148	355	9487
P	814	129	689	818	421	821	220	519	482	814	193	455	388	400	430	640	524	735	140	361	9993
Q	792	105	442	575	335	586	203	480	481	835	174	352	346	389	415	417	406	547	119	299	8298
R	827	153	667	753	492	744	252	638	532	1044	197	457	426	419	539	554	522	813	157	422	10608
S	957	193	701	723	554	1059	322	704	650	1113	249	514	586	506	593	829	679	834	199	470	12435
T	920	181	691	702	512	991	285	710	601	1170	228	467	662	427	520	673	681	880	202	437	11940
V	1296	201	963	967	560	969	349	882	913	1224	317	671	724	532	799	870	911	1127	189	465	14929
W	207	42	188	193	117	217	85	150	185	262	64	142	107	154	174	166	176	191	37	104	2961
Y	555	106	495	451	293	589	166	386	410	646	150	389	392	320	424	475	432	481	117	315	7592
Total	17723	2857	12709	13872	8471	15893	4942	12262	12615	19023	4577	9489	9994	8313	10618	12397	11940	14933	2961	7613	213202

Table 7-2 で、ジペプチドの数が1,000 以上に関わっているアミノ酸残基は、アラニン、アスパラギン酸、グルタミン酸、グリシン、イソロイシン、リジン、ロイシン、プロリン、アルギニン、セリン、チロシン、バリンである。これら 11 種類のアミノ酸残基は、Table 7-1 の各アミノ酸残基の数が 10,000 以上のものである。つまり、ジペプチドでは、これらのアミノ酸残基の組み合わせが 1,000 以上として現れた。また、アミノ酸残基の数が 5,000 未満のものは、ジペプチドでは、100 未満として現れた。

7—3 ジペプチド単位でみる Chou-Fasman パラメータ

7—3—1 α ヘリックス構造の統計解析

蛋白質の二次構造をジペプチド単位で見た場合のパラメータの求め方は、二通り考えられる。一つは、アミノ酸残基による Chou-Fasman パラメータを使って順列の考えに基づいた算出方法と、もう一つは、ジペプチドを単位パラメータとして算出する方法である。まず、順列の考えにそったジペプチドパラメータは、二次構造の含まれる i 番目のアミノ酸残基 X_i の傾向指数と

$i+1$ 番目のアミノ酸残基 X_{i+1} の傾向指数の積で表わせる。たとえば、 i 番目のアミノ酸残基として Ala, $i+1$ 番目のアミノ酸残基として Cys であれば、順列に基づくジペプチドのパラメータは、 $1.26 \times 0.91 = 1.14$ である。しかしながら、アミノ酸残基の Chou-Fasman パラメータは、個々のアミノ酸残基の傾向指数である。つまり、AC を CA としても同じ計算結果を得る。したがって、順列を使ってジペプチドを算出するのは適切ではない。したがって、ここでは配列情報をもつことの優位性を考慮して、ジペプチドを基本単位とする二次構造の Chou-Fasman パラメータを算出した。まず、1081 種類の蛋白質の中の α ヘリックス構造に出現するジペプチドの数を調べた (Table 7-3)。

Table 7-3. α ヘリックス構造に出現するジペプチドの数

$i+1$ i	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	1193	124	419	761	353	533	205	547	638	1168	279	323	108	477	617	444	382	636	151	289	9647
C	98	21	53	79	43	69	37	49	57	129	20	44	23	56	61	58	56	57	13	34	1057
D	574	53	238	441	256	192	98	372	317	606	119	168	199	202	265	245	238	363	107	233	5286
E	830	73	326	659	269	302	142	445	592	863	230	263	103	385	438	323	331	472	126	230	7402
F	316	46	166	262	119	198	77	182	226	399	92	139	60	136	170	164	126	202	68	127	3275
G	341	59	181	234	165	190	82	227	202	416	110	108	113	156	183	173	162	249	47	124	3522
H	179	27	102	151	97	106	55	114	94	246	59	68	65	87	108	100	89	128	34	68	1977
I	554	61	267	384	152	258	94	233	336	486	97	238	93	239	284	267	233	266	49	134	4725
K	646	54	331	569	227	238	126	298	424	616	153	247	87	288	300	283	256	322	69	213	5747
L	1151	116	528	783	289	641	214	431	755	1001	267	421	177	478	665	515	411	486	101	287	9717
M	272	28	119	170	83	139	43	138	189	249	65	98	50	102	135	98	103	146	17	60	2304
N	387	41	163	280	161	177	92	208	223	394	99	126	135	154	181	153	170	225	65	152	3586
P	256	17	140	318	109	75	49	126	126	256	65	61	42	136	106	120	94	131	32	93	2352
Q	519	42	199	384	162	195	112	238	273	520	117	149	61	273	250	166	162	221	62	154	4259
R	504	68	327	493	215	242	133	302	300	564	116	211	60	256	307	232	198	301	75	202	5106
S	489	65	245	351	217	271	127	302	301	577	133	170	170	252	285	243	218	318	89	183	5006
T	461	60	208	350	166	260	97	253	234	538	101	135	171	203	238	201	184	272	70	138	4340
V	650	74	278	443	186	250	110	246	366	500	138	211	100	259	355	257	234	315	53	142	5167
W	116	18	65	93	59	56	45	71	87	162	32	44	20	79	90	59	58	71	19	41	1285
Y	260	38	156	193	129	203	55	147	176	349	81	134	69	160	179	155	103	165	48	125	2925
Total	9796	1085	4511	7398	3457	4595	1993	4929	5916	10039	2373	3358	1906	4378	5217	4256	3808	5346	1295	3029	88685

この表から、蛋白質中のジペプチドのうち、41.6%が α ヘリックス構造のジペプチドであることがわかった。

つぎに、Table 7-2 と Table 7-3 より、 α ヘリックス構造のジペプチドの傾向指数を算出した。算出方法は、Chou-Fasman パラメータの算出方法をジペプチドに適用した。すなわち、全蛋白質中における各ジペプチドの割合と、 α ヘリックス構造中における各ジペプチドの割合の比を

とった；

$$\text{ジペプチドの傾向指数} = \frac{\frac{\alpha \text{ヘリックス構造の中の特定のジペプチドの数}}{\alpha \text{ヘリックス構造におけるすべてのジペプチドの数}}}{\frac{\text{蛋白質の中の特定のジペプチドの数}}{\text{蛋白質の中のすべてのジペプチドの数}}}$$

たとえば、i 番目のアミノ酸残基を Ala, i+1 番目のアミノ酸残基を Cys, すなわち、ジペプチド AC の場合、全蛋白質中に含まれる割合は、230/213202 より約 0.11 であり、 α ヘリックス構造中に含まれる割合は 124/88685 より約 0.14 であるため、AC の傾向指数は、0.14/0.11 より 1.27 である。この計算方法をすべてのジペプチドの適用した (Table 7-4)。

Table 7-4. α ヘリックス構造のジペプチドの傾向指数

i+1 i	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	1.7	1.3	1.0	1.5	1.3	1.0	1.2	1.3	1.5	1.7	1.5	1.1	0.4	1.6	1.6	1.1	1.0	1.2	1.4	1.2
C	1.1	1.0	0.8	1.2	0.9	0.7	0.9	0.8	1.0	1.3	0.8	0.8	0.4	1.1	1.0	0.7	0.8	0.8	1.0	0.8
D	1.3	0.8	0.9	1.3	1.0	0.4	0.9	1.0	1.1	1.3	1.1	0.8	0.7	1.2	1.1	0.9	0.9	1.0	1.2	1.1
E	1.7	1.2	1.0	1.6	1.2	0.8	1.3	1.3	1.4	1.5	1.5	1.0	0.5	1.7	1.5	1.1	1.1	1.1	1.4	1.2
F	1.2	0.9	0.7	1.2	1.0	0.8	0.9	1.0	1.0	1.3	1.2	0.8	0.4	1.1	1.1	0.7	0.6	0.9	1.4	1.0
G	0.7	0.6	0.5	0.6	0.6	0.4	0.5	0.5	0.5	0.8	0.8	0.4	0.5	0.6	0.6	0.4	0.4	0.5	0.5	0.5
H	1.2	0.7	0.9	1.3	0.9	0.6	0.9	0.9	1.0	1.3	1.3	0.9	0.5	1.1	1.1	0.9	0.8	0.9	1.0	0.9
I	1.3	0.9	0.8	1.1	0.9	0.7	0.8	0.9	1.0	1.2	1.1	0.9	0.4	1.3	1.1	0.8	0.7	0.8	0.9	0.8
K	1.5	1.0	1.0	1.5	1.1	0.7	1.2	1.0	1.3	1.3	1.3	1.0	0.4	1.5	1.3	1.0	0.9	0.8	1.0	1.1
L	1.7	1.2	1.1	1.5	1.1	1.2	1.1	1.1	1.4	1.5	1.6	1.1	0.4	1.4	1.5	1.0	0.9	1.0	1.2	1.1
M	1.6	1.2	1.0	1.4	1.1	1.0	1.1	1.2	1.4	1.5	1.3	1.0	0.5	1.4	1.3	1.0	0.9	1.1	1.1	1.0
N	1.2	0.7	0.7	1.3	1.0	0.5	1.0	0.8	1.0	1.1	1.1	0.7	0.6	1.1	1.1	0.7	0.8	0.8	1.1	1.0
P	0.8	0.3	0.5	0.9	0.6	0.2	0.5	0.6	0.6	0.8	0.8	0.3	0.3	0.8	0.6	0.5	0.4	0.4	0.5	0.6
Q	1.6	1.0	1.1	1.6	1.2	0.8	1.3	1.2	1.4	1.5	1.6	1.0	0.4	1.7	1.4	1.0	1.0	1.0	1.3	1.2
R	1.5	1.1	1.2	1.6	1.1	0.8	1.3	1.1	1.4	1.3	1.4	1.1	0.3	1.5	1.4	1.0	0.9	0.9	1.1	1.2
S	1.2	0.8	0.8	1.2	0.9	0.6	0.9	1.0	1.1	1.2	1.3	0.8	0.7	1.2	1.2	0.7	0.8	0.9	1.1	0.9
T	1.2	0.8	0.7	1.2	0.8	0.6	0.8	0.9	0.9	1.1	1.1	0.7	0.6	1.1	1.1	0.7	0.6	0.7	0.8	0.8
V	1.2	0.9	0.7	1.1	0.8	0.6	0.8	0.7	1.0	1.0	1.0	0.8	0.3	1.2	1.1	0.7	0.6	0.7	0.7	0.7
W	1.3	1.0	0.8	1.2	1.2	0.6	1.3	1.1	1.1	1.5	1.2	0.7	0.4	1.2	1.2	0.9	0.8	0.9	1.2	0.9
Y	1.1	0.9	0.8	1.0	1.1	0.8	0.8	0.9	1.0	1.3	1.3	0.8	0.4	1.2	1.0	0.8	0.6	0.8	1.0	1.0

赤色の数値は 1.5 以上の傾向指数であり、青色の数値は 0.5 以下である。Table 7-4 から、たとえば、AC と CA の傾向指数は、それぞれ 1.3 と 1.1 だった。つまり、アミドを考慮したジペ

プチドを基本単位として扱ったため、アミノ酸残基の組み合わせとしては、アラニンとシステインを使っているにも関わらず異なる傾向指数を示した。そして、アラニン、グルタミン酸、リジン、ロイシン、メチオニン、グルタミン、アルギニンは、Table 7-1 のアミノ酸残基の α ヘリックス構造の傾向指数 P_{helix} において、1.0 より高い傾向指数をもつアミノ酸残基であったが、これらの組み合わせは、1.5 以上の傾向指数として現れた。

傾向指数が高いジペプチドは、そのジペプチドがヘリックス構造の中で頻繁に現れることを意味するのみで、コンホメーションについての情報を得ることができない。そこで、ジペプチドの二面角 (ψ_i, ϕ_{i+1}) の分布を調べた (Figure 7-2)。ここでは、ジペプチドの N 単側がアラニンで、傾向指数が 1.5 以上のジペプチド、つまり、AA, AE, AK, AL, AM, AQ, AR を調べた。Figure 7-2 より、ヒストグラムの形状が、鋭利な形のものと、幅広い形のものに分類できる。鋭利な形は、AA, AE, AK, AR であり、幅広い形は、AL, AM, AQ である。つまり、傾向指数が高くても、特定の角度に集中している場合と、幅広い角度をとっているジペプチドに分類できた。

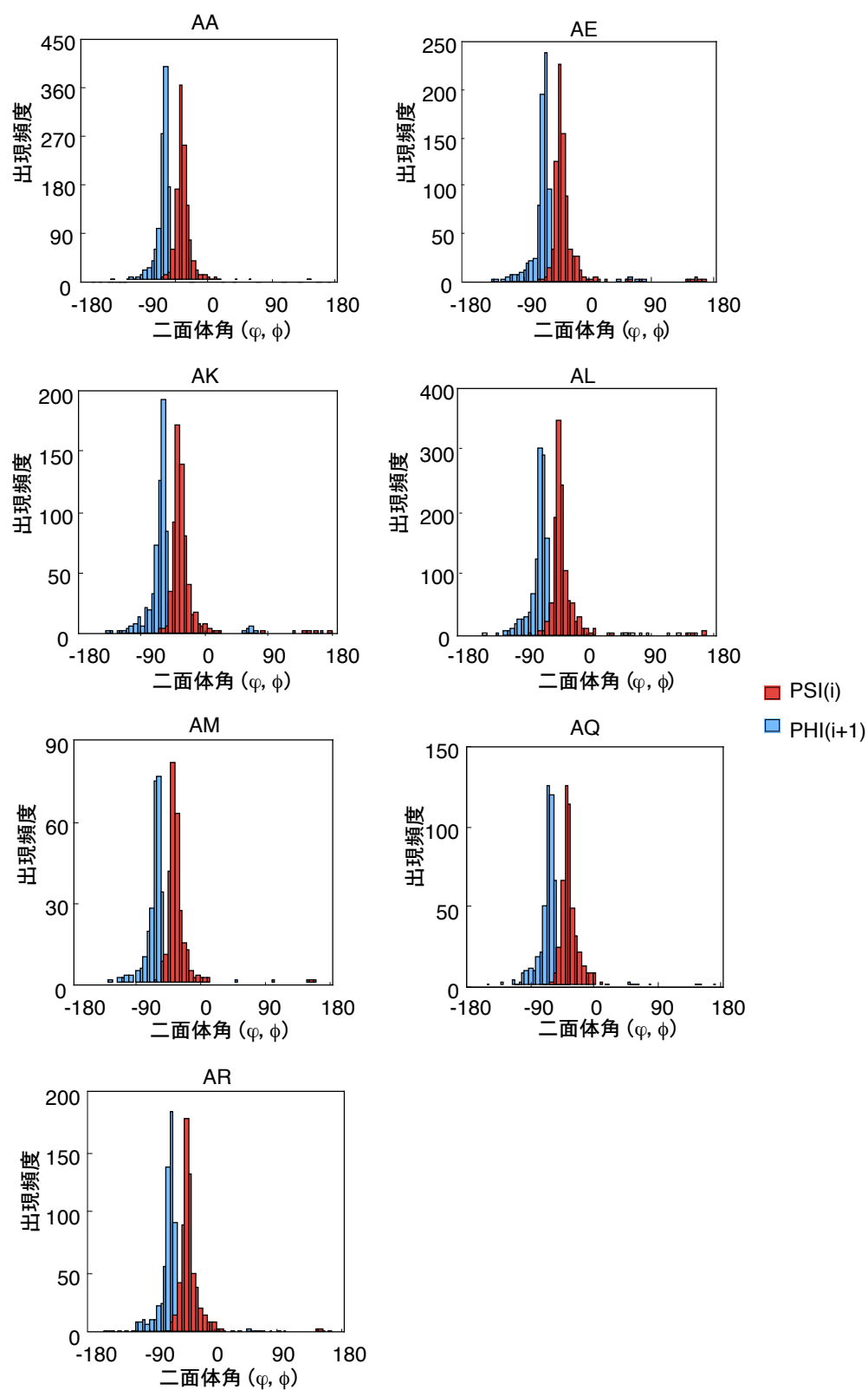


Figure 7-2. α ヘリックス構造の中のジペプチドの主鎖二面体角の分布

7-3-2 β ストランド構造の統計解析

7-3-1 節の α ヘリックス構造と同様に、 β ストランド構造に対してジペプチドを基本単位として統計的に解析した。まず、1081 種類の蛋白質の中の β ストランド構造に現れるジペプチドの数を調べた (Table 7-5)。

Table 7-5. β ストランド構造に出現するジペプチドの数

$i+1$ i	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	271	67	193	197	200	197	106	357	192	348	96	145	409	120	187	259	302	467	60	175	4348
C	70	15	66	31	42	57	27	66	40	77	23	43	71	37	48	71	70	84	17	44	999
D	129	35	46	70	127	84	31	211	65	165	34	51	111	41	63	95	107	259	27	109	1860
E	181	41	142	136	169	96	43	283	172	308	91	98	277	72	116	135	198	426	55	116	3155
F	190	57	221	200	130	166	91	222	185	233	54	153	236	110	147	251	270	303	36	126	3381
G	230	48	138	144	160	151	68	261	145	262	57	119	189	95	124	189	217	290	51	153	3091
H	75	36	66	43	86	75	39	138	62	122	24	41	107	45	57	83	93	153	27	68	1440
I	361	77	368	301	185	304	158	351	317	379	99	253	453	161	238	356	435	469	59	202	5526
K	228	43	125	138	170	140	52	324	176	334	87	112	354	87	141	197	211	452	55	112	3538
L	313	82	359	277	213	223	158	369	312	428	96	271	670	207	288	419	461	529	80	233	5988
M	88	15	97	69	59	73	38	96	66	101	33	66	155	40	64	83	116	137	15	53	1464
N	121	37	66	56	86	76	33	187	73	170	38	40	161	44	58	78	96	214	33	102	1769
P	268	59	144	143	153	153	74	237	149	320	77	107	230	122	152	219	206	426	46	126	3411
Q	139	32	92	93	115	103	41	174	103	183	37	69	205	51	92	130	133	234	34	78	2138
R	191	55	130	115	170	137	51	250	118	315	47	77	270	79	108	148	173	395	53	125	3007
S	194	60	98	104	204	163	64	274	110	279	56	85	181	81	91	154	175	342	55	157	2927
T	231	67	152	126	252	165	75	356	157	420	86	112	352	109	136	180	269	467	93	208	4013
V	463	94	424	378	312	332	171	548	397	590	138	290	512	206	345	439	513	705	114	264	7235
W	53	17	60	56	41	65	26	65	60	65	18	52	50	53	58	65	84	95	8	48	1039
Y	169	45	203	149	106	141	71	195	135	201	43	132	208	100	157	192	230	262	46	133	2918
Total	3965	982	3190	2826	2980	2901	1417	4964	3034	5300	1234	2316	5201	1860	2670	3743	4359	6709	964	2632	63247

その結果、蛋白質の中のジペプチドのうち 32%が β ストランド構造のジペプチドだった。さらに、Table 7-3 と Table 7-5 より、 α ヘリックス構造と同様に Chou-Fasman パラメータの算出方法を使って、 β ストランド構造のジペプチドの傾向指数を算出した (Table 7-6)。傾向指数が 1.5 以上を赤色で、0.5 以下を青色で色分けした。傾向指数が 1.5 以上の値だったジペプチドは、Table 3-1 の β ストランド構造の中のアミノ酸残基の傾向指数 P_{strand} において、1.0 より大きい値をもつアミノ酸残基が関わっていた。

Table 7-6. β ストランド構造のジペプチドの傾向指数

i \ i+1	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.5	1.0	0.6	0.6	1.0	0.5	0.9	1.2	0.6	0.7	0.7	0.7	2.0	0.6	0.7	0.9	1.1	1.3	0.8	1.0
C	1.1	1.0	1.4	0.7	1.2	0.8	1.0	1.5	1.0	1.1	1.4	1.0	1.6	1.0	1.1	1.3	1.4	1.6	1.8	1.5
D	0.4	0.8	0.2	0.3	0.7	0.3	0.4	0.8	0.3	0.5	0.4	0.3	0.5	0.3	0.4	0.5	0.5	1.0	0.4	0.7
E	0.5	0.9	0.6	0.5	1.1	0.4	0.5	1.1	0.6	0.8	0.8	0.5	1.9	0.4	0.6	0.7	0.9	1.4	0.9	0.9
F	1.0	1.5	1.3	1.2	1.5	0.9	1.5	1.6	1.2	1.1	1.0	1.3	1.9	1.2	1.3	1.5	1.8	1.8	1.0	1.4
G	0.6	0.7	0.5	0.5	0.8	0.4	0.6	0.9	0.5	0.7	0.6	0.6	1.1	0.5	0.5	0.6	0.7	0.9	0.8	0.8
H	0.7	1.4	0.8	0.5	1.2	0.6	0.9	1.5	0.9	0.9	0.7	0.7	1.2	0.8	0.8	1.0	1.2	1.6	1.1	1.2
I	1.2	1.7	1.5	1.3	1.6	1.2	1.8	1.8	1.4	1.3	1.5	1.4	2.4	1.2	1.3	1.6	1.8	1.9	1.5	1.7
K	0.7	1.1	0.5	0.5	1.2	0.6	0.7	1.5	0.7	1.0	1.0	0.6	2.1	0.6	0.8	0.9	1.1	1.6	1.2	0.8
L	0.6	1.2	1.1	0.7	1.2	0.6	1.2	1.4	0.8	0.9	0.8	1.0	2.2	0.9	0.9	1.2	1.4	1.5	1.3	1.3
M	0.7	0.9	1.1	0.8	1.1	0.7	1.3	1.2	0.7	0.9	1.0	1.0	2.1	0.7	0.9	1.2	1.5	1.4	1.4	1.2
N	0.5	0.9	0.4	0.4	0.8	0.3	0.5	1.0	0.5	0.7	0.6	0.3	0.9	0.4	0.5	0.5	0.6	1.0	0.8	1.0
P	1.1	1.5	0.7	0.6	1.2	0.6	1.1	1.5	1.0	1.3	1.3	0.8	2.0	1.0	1.2	1.2	1.3	2.0	1.1	1.2
Q	0.6	1.0	0.7	0.5	1.2	0.6	0.7	1.2	0.7	0.7	0.7	0.7	2.0	0.4	0.7	1.1	1.1	1.4	1.0	0.9
R	0.8	1.2	0.7	0.5	1.2	0.6	0.7	1.3	0.7	1.0	0.8	0.6	2.1	0.6	0.7	0.9	1.1	1.6	1.1	1.0
S	0.7	1.0	0.5	0.5	1.2	0.5	0.7	1.3	0.6	0.8	0.8	0.6	1.0	0.5	0.5	0.6	0.9	1.4	0.9	1.1
T	0.8	1.2	0.7	0.6	1.7	0.6	0.9	1.7	0.9	1.2	1.3	0.8	1.8	0.9	0.9	0.9	1.3	1.8	1.6	1.6
V	1.2	1.6	1.5	1.3	1.9	1.2	1.7	2.1	1.5	1.6	1.5	1.5	2.4	1.3	1.5	1.7	1.9	2.1	2.0	1.9
W	0.9	1.4	1.1	1.0	1.2	1.0	1.0	1.5	1.1	0.8	0.9	1.2	1.6	1.2	1.1	1.3	1.6	1.7	0.7	1.6
Y	1.0	1.4	1.4	1.1	1.2	0.8	1.4	1.7	1.1	1.0	1.0	1.1	1.8	1.1	1.2	1.4	1.8	1.8	1.3	1.4

つぎに、前節と同様に、 β ストランド構造の中のジペプチドでも、主鎖二面体角 (ψ_i, ϕ_{i+1}) に特徴が見える可能性があると考え、主鎖二面体角の分布を調べた (Figure 7-3)。

Figure 7-3 に使ったジペプチドは、 α ヘリックス構造において傾向指数が高くて、 β ストランド構造では低い AA, AE, AK, そして、 β ストランド構造において、傾向指数が比較的高い AP, AV である。この結果から、ジペプチド AP, AV は、鋭利な形をしたヒストグラムであり、それ以外は、広がった形をしたヒストグラムだった。つまり、 β ストランド構造の中のジペプチドは、取り得る主鎖二面体角が特定の角度に集中したものと、集中しないものの二つに分類できた。なお、 ψ 角は、どのジペプチドでもある特定の角度に集中していた。これは、 α ヘリックス構造でも同じだった。

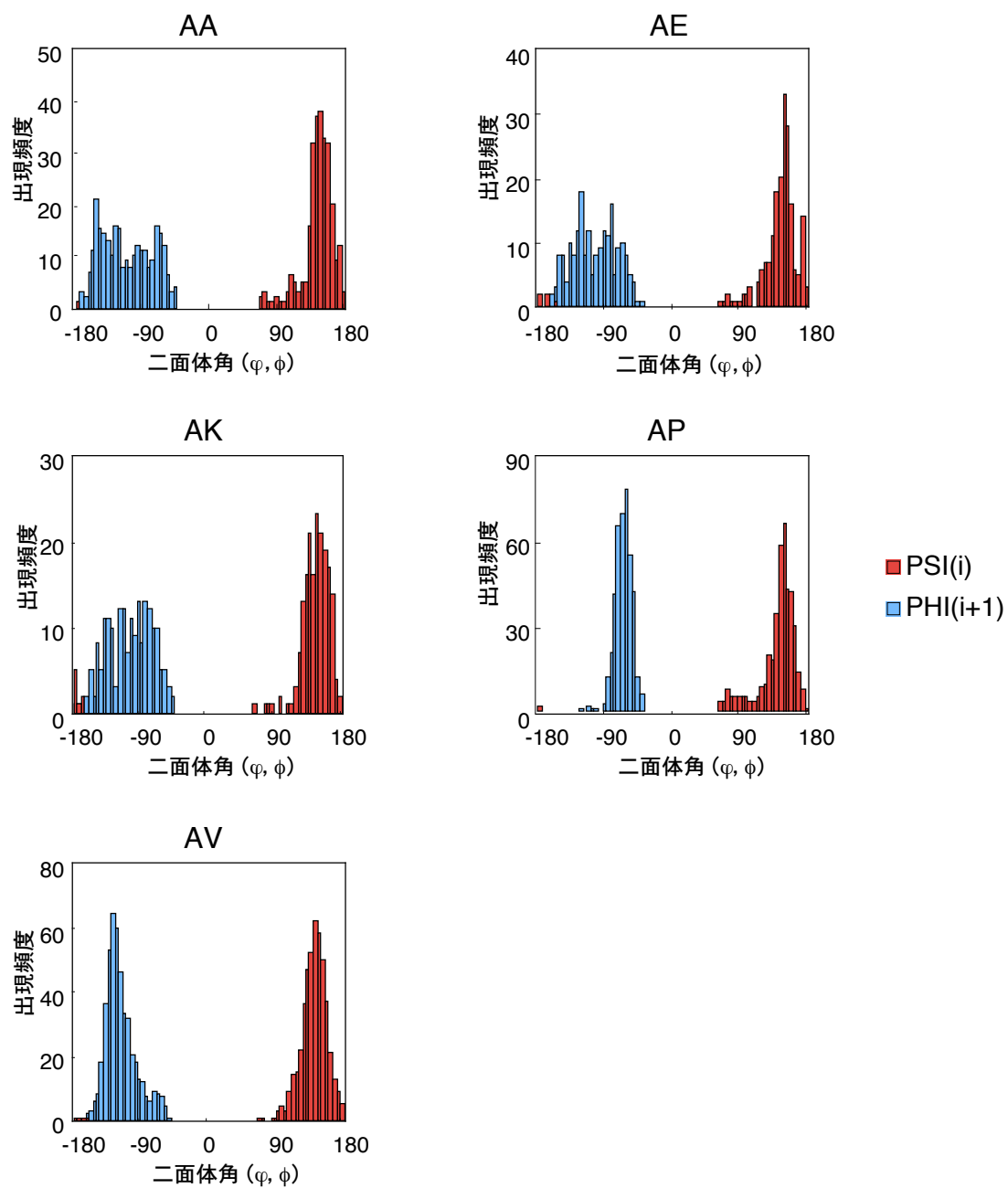


Figure 7-3. β ストランド構造の中のジペプチドの主鎖二面体角の分布

7—3—3 シート形成 β ストランド構造の統計解析

前節では、ProSSA 法で定義された β ストランド構造に対しての Chou-Fasman パラメータを算出した。ところが、第2章でも述べたように、ProSSA 法の β ストランド構造は、シート形成と一本鎖の β ストランドを区別できない。そこで、この節では、シート形成 β ストランド構造でのジペプチドに対して、統計的に解析した。

まず、1081 種類の蛋白質の β ストランド構造におけるジペプチドの数を算出した (Table 7-7)。

Table 7-7. シート形成 β ストランド構造に出現するジペプチドの数

$\begin{smallmatrix} i+1 \\ i \end{smallmatrix}$	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	149	37	49	84	127	91	54	270	72	226	61	49	43	63	75	93	131	342	34	108	2158
C	44	12	25	22	29	25	11	51	19	58	17	20	11	19	31	28	32	58	10	34	556
D	44	14	7	21	57	26	9	122	8	76	15	11	9	13	20	21	29	137	8	47	694
E	73	30	29	48	99	35	8	199	66	188	47	29	38	34	56	41	87	273	28	67	1475
F	122	45	88	119	91	101	55	167	105	164	40	66	34	63	82	119	159	227	26	95	1968
G	118	27	29	72	115	48	33	179	58	157	33	41	21	40	65	68	105	189	31	94	1523
H	40	13	19	14	55	28	21	104	30	85	14	10	14	21	29	30	43	103	18	46	737
I	259	64	166	207	154	185	98	303	181	309	78	121	94	101	161	203	266	388	49	164	3551
K	99	24	28	49	107	49	21	202	65	197	54	36	46	32	56	66	79	316	27	59	1612
L	182	60	141	162	173	109	92	299	149	323	68	116	79	111	148	176	242	413	59	196	3298
M	59	10	33	35	46	34	20	75	32	67	22	25	24	22	34	38	48	108	12	40	784
N	46	17	10	19	53	20	6	101	22	88	19	5	14	17	26	25	38	117	18	48	709
P	23	11	7	10	23	12	9	46	10	44	20	3	4	11	8	14	17	91	5	19	387
Q	62	21	18	35	78	32	20	124	44	110	18	24	29	22	35	46	62	155	19	43	997
R	94	27	43	45	106	47	19	169	43	184	26	21	42	35	39	60	82	278	26	82	1468
S	99	30	16	43	132	64	29	168	40	174	35	22	11	37	43	55	84	232	32	76	1422
T	136	39	44	64	174	76	39	254	73	290	57	38	55	52	68	79	140	337	58	123	2196
V	303	67	191	235	250	187	115	473	242	452	116	129	113	132	213	213	307	579	92	217	4626
W	32	11	20	33	30	30	19	54	33	42	11	27	6	33	35	27	42	72	4	37	598
Y	112	27	91	89	90	89	41	158	66	157	29	55	36	60	90	99	131	203	33	100	1756
Total	2096	586	1054	1406	1989	1288	719	3518	1358	3391	780	848	723	918	1314	1501	2124	4618	589	1695	32515

その結果、 β ストランド構造の中のジペプチドのうち約51%が、シート形成 β ストランド構造に属していた。

つぎに、シート形成 β ストランド構造のジペプチドに対する傾向指数を算出した (Table 7-8)。

Table 7-8. シート形成 β ストランド構造のジペプチドの傾向指数

i+1 \ i	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.6	1.1	0.3	0.5	1.3	0.5	0.9	1.7	0.5	0.9	0.9	0.4	0.4	0.6	0.5	0.6	0.9	1.8	0.9	1.2
C	1.3	1.5	1.0	0.9	1.7	0.7	0.8	2.3	1.0	1.6	2.0	0.9	0.5	1.0	1.3	1.0	1.2	2.1	2.0	2.2
D	0.3	0.6	0.1	0.2	0.6	0.2	0.2	0.9	0.1	0.4	0.4	0.1	0.1	0.2	0.2	0.2	0.3	1.0	0.3	0.6
E	0.4	1.3	0.2	0.3	1.2	0.3	0.2	1.6	0.4	0.9	0.8	0.3	0.5	0.4	0.5	0.4	0.8	1.7	0.9	1.0
F	1.3	2.3	1.0	1.4	2.0	1.1	1.7	2.4	1.3	1.5	1.5	1.1	0.5	1.4	1.4	1.4	2.0	2.6	1.4	2.0
G	0.6	0.8	0.2	0.5	1.1	0.3	0.6	1.2	0.4	0.8	0.6	0.4	0.2	0.4	0.5	0.4	0.7	1.1	1.0	1.0
H	0.7	1.0	0.5	0.3	1.5	0.5	1.0	2.2	0.9	1.2	0.8	0.4	0.3	0.8	0.8	0.7	1.1	2.1	1.4	1.6
I	1.6	2.7	1.3	1.7	2.5	1.4	2.2	3.0	1.5	2.1	2.3	1.3	1.0	1.4	1.7	1.7	2.2	3.1	2.5	2.7
K	0.6	1.2	0.2	0.3	1.4	0.4	0.5	1.8	0.5	1.1	1.2	0.4	0.5	0.4	0.6	0.6	0.8	2.2	1.1	0.9
L	0.7	1.6	0.8	0.8	1.8	0.5	1.3	2.2	0.8	1.3	1.1	0.8	0.5	0.9	0.9	1.0	1.5	2.3	1.9	2.1
M	0.9	1.2	0.7	0.8	1.6	0.7	1.4	1.8	0.6	1.1	1.2	0.7	0.6	0.8	0.9	1.1	1.2	2.2	2.1	1.8
N	0.4	0.8	0.1	0.2	0.9	0.2	0.2	1.1	0.3	0.7	0.6	0.1	0.2	0.3	0.4	0.3	0.5	1.1	0.8	0.9
P	0.2	0.6	0.1	0.1	0.4	0.1	0.3	0.6	0.1	0.4	0.7	0.0	0.1	0.2	0.1	0.1	0.2	0.8	0.2	0.3
Q	0.5	1.3	0.3	0.4	1.5	0.4	0.6	1.7	0.6	0.9	0.7	0.4	0.5	0.4	0.6	0.7	1.0	1.9	1.0	0.9
R	0.7	1.2	0.4	0.4	1.4	0.4	0.5	1.7	0.5	1.2	0.9	0.3	0.6	0.5	0.5	0.7	1.0	2.2	1.1	1.3
S	0.7	1.0	0.1	0.4	1.6	0.4	0.6	1.6	0.4	1.0	0.9	0.3	0.1	0.5	0.5	0.4	0.8	1.8	1.1	1.1
T	1.0	1.4	0.4	0.6	2.2	0.5	0.9	2.3	0.8	1.6	1.6	0.5	0.5	0.8	0.9	0.8	1.3	2.5	1.9	1.8
V	1.5	2.2	1.3	1.6	2.9	1.3	2.2	3.5	1.7	2.4	2.4	1.3	1.0	1.6	1.7	1.6	2.2	3.4	3.2	3.1
W	1.0	1.7	0.7	1.1	1.7	0.9	1.5	2.4	1.2	1.1	1.1	1.2	0.4	1.4	1.3	1.1	1.6	2.5	0.7	2.3
Y	1.3	1.7	1.2	1.3	2.0	1.0	1.6	2.7	1.1	1.6	1.3	0.9	0.6	1.2	1.4	1.4	2.0	2.8	1.8	2.1

これまでと同様、傾向指数が1.5以上を青色で、0.5以下を赤色で色分けした。シート形成 β ストランド構造のアミノ酸残基に対する Chou-Fasman パラメータ (Table 3-2) と比較すると、Chou-Fasman パラメータで、1.3以上の傾向指数をもつアミノ酸残基同士のジペプチドは、すべて1.5以上の傾向指数を示した。また、ジペプチドの組み合わせの中には、片方のアミノ酸残基のパラメータが1.8を超えている場合は、ほとんどのペプチドのパラメータが1.5より高い値を示した。1.5未満の値を示したジペプチドは、いずれもアミノ酸残基の傾向指数が0.6以下を示したアスパラギン酸、グリシン、イソロイシン、プロリンが関わるジペプチドとIQだった。一方、アミノ酸残基のパラメータが1.0よりも低い値をもつものの同士のジペプチドの組み合わせは、ジペプチドのパラメータでも、1.0を超えた組み合わせは存在しなかった。また、 α ヘリックス構造や β ストランド構造のジペプチド解析と同様に、アミノ酸残基の順番によってジペプチドの傾向指数が異なった。たとえば、アラニンとシステインの組み合わせの傾向指数は、ACが1.1、CAは1.3だった。

7—3—4 一本鎖 β ストランド構造の統計解析

まず、1081 種類の蛋白質の中の本鎖 β ストランド構造に現れるジペプチドの数を調べた (Table 7-9)。

Table 7-9. 一本鎖 β ストランド構造に出現するジペプチドの数

$i \backslash i+1$	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	95	18	103	75	50	74	33	48	72	75	21	75	350	33	75	110	109	77	11	43	1547
C	21	1	31	3	5	17	15	9	13	11	4	16	58	10	9	30	23	15	4	6	301
D	68	11	33	38	47	39	16	42	35	60	10	34	96	17	24	62	53	65	18	35	803
E	75	8	93	70	47	38	28	60	87	85	30	48	225	29	43	74	78	98	18	34	1268
F	47	7	98	54	24	45	25	39	63	40	11	59	185	37	32	91	74	53	6	21	1011
G	83	16	95	59	35	63	26	42	59	64	18	60	158	35	43	85	69	66	17	36	1129
H	22	13	35	23	21	34	12	22	21	23	6	23	86	16	20	37	31	26	7	17	495
I	64	7	142	60	23	63	28	25	84	42	14	91	321	38	52	110	109	50	4	25	1352
K	98	11	74	70	42	56	24	65	75	92	19	54	283	38	61	105	91	75	20	33	1386
L	86	15	166	83	24	64	45	31	106	65	16	109	541	68	91	204	169	80	13	24	2000
M	16	3	48	27	8	30	9	8	23	16	7	30	120	11	28	39	49	21	1	10	504
N	43	15	45	28	22	31	17	54	40	57	14	28	138	19	23	40	34	55	11	28	742
P	191	35	120	114	95	95	49	96	106	177	39	77	225	90	102	163	131	175	28	83	2191
Q	60	8	58	39	26	43	11	32	41	55	15	38	165	18	43	61	47	43	9	24	836
R	73	16	67	53	41	59	21	56	57	86	14	39	210	34	45	65	65	59	19	24	1103
S	71	20	63	44	44	73	25	58	53	66	15	51	160	31	35	76	55	57	16	50	1063
T	65	16	91	44	55	47	21	49	61	70	15	54	275	33	39	70	82	78	19	57	1241
V	109	13	165	104	36	73	34	43	102	88	14	116	349	49	83	163	144	65	14	29	1793
W	13	5	32	15	11	22	3	7	15	12	2	19	41	14	13	22	31	8	3	7	295
Y	33	11	83	36	11	35	19	21	56	23	6	56	157	29	49	61	68	37	8	23	822
Total	1333	249	1642	1039	667	1001	461	807	1169	1207	290	1077	4143	649	910	1668	1512	1203	246	609	21882

その結果、 β ストランド構造のジペプチドのうち、約 35%のジペプチドが一本鎖 β ストランド構造に属してした。

つぎに、ジペプチドに対する傾向指数を算出した (Table 7-10)。

Table 7-10. 一本鎖 β ストランド構造のジペプチドの傾向指数

i+1 \ i	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.5	0.8	1.0	0.6	0.7	0.6	0.8	0.5	0.7	0.4	0.5	1.0	5.0	0.4	0.8	1.1	1.1	0.6	0.4	0.7
C	0.9	0.2	1.8	0.2	0.4	0.7	1.6	0.6	1.0	0.4	0.7	1.1	3.7	0.8	0.6	1.6	1.3	0.8	1.2	0.6
D	0.6	0.7	0.5	0.4	0.8	0.4	0.6	0.5	0.5	0.5	0.4	0.6	1.4	0.4	0.4	0.9	0.8	0.7	0.8	0.7
E	0.6	0.5	1.2	0.7	0.9	0.4	1.0	0.7	0.9	0.6	0.8	0.7	4.5	0.5	0.6	1.1	1.0	0.9	0.8	0.7
F	0.7	0.5	1.7	1.0	0.8	0.7	1.2	0.8	1.2	0.5	0.6	1.4	4.4	1.2	0.8	1.6	1.4	0.9	0.5	0.7
G	0.6	0.7	1.0	0.6	0.5	0.5	0.7	0.4	0.5	0.5	0.5	0.9	2.7	0.6	0.5	0.8	0.7	0.6	0.8	0.5
H	0.6	1.5	1.3	0.8	0.8	0.8	0.8	0.7	0.9	0.5	0.5	1.2	2.7	0.9	0.8	1.3	1.2	0.8	0.8	0.9
I	0.6	0.4	1.7	0.7	0.6	0.7	0.9	0.4	1.1	0.4	0.6	1.4	5.0	0.8	0.8	1.4	1.3	0.6	0.3	0.6
K	0.9	0.8	0.9	0.7	0.8	0.7	0.9	0.9	0.9	0.8	0.6	0.9	4.8	0.8	1.1	1.4	1.3	0.8	1.2	0.7
L	0.5	0.6	1.4	0.6	0.4	0.5	1.0	0.3	0.8	0.4	0.4	1.2	5.1	0.8	0.8	1.7	1.5	0.7	0.6	0.4
M	0.4	0.5	1.6	0.9	0.4	0.9	0.9	0.3	0.7	0.4	0.6	1.3	4.7	0.6	1.1	1.6	1.8	0.6	0.3	0.7
N	0.6	1.1	0.8	0.5	0.6	0.4	0.8	0.9	0.7	0.7	0.6	0.6	2.3	0.5	0.5	0.8	0.6	0.8	0.7	0.8
P	2.3	2.6	1.7	1.4	2.2	1.1	2.2	1.8	2.1	2.1	2.0	1.6	5.7	2.2	2.3	2.5	2.4	2.3	1.9	2.2
Q	0.7	0.7	1.3	0.7	0.8	0.7	0.5	0.6	0.8	0.6	0.8	1.1	4.6	0.5	1.0	1.4	1.1	0.8	0.7	0.8
R	0.9	1.0	1.0	0.7	0.8	0.8	0.8	0.9	1.0	0.8	0.7	0.8	4.8	0.8	0.8	1.1	1.2	0.7	1.2	0.6
S	0.7	1.0	0.9	0.6	0.8	0.7	0.8	0.8	0.8	0.6	0.6	1.0	2.7	0.6	0.6	0.9	0.8	0.7	0.8	1.0
T	0.7	0.9	1.3	0.6	1.0	0.5	0.7	0.7	1.0	0.6	0.6	1.1	4.0	0.8	0.7	1.0	1.2	0.9	0.9	1.3
V	0.8	0.6	1.7	1.0	0.6	0.7	0.9	0.5	1.1	0.7	0.4	1.7	4.7	0.9	1.0	1.8	1.5	0.6	0.7	0.6
W	0.6	1.2	1.7	0.8	0.9	1.0	0.3	0.5	0.8	0.4	0.3	1.3	3.7	0.9	0.7	1.3	1.7	0.4	0.8	0.7
Y	0.6	1.0	1.6	0.8	0.4	0.6	1.1	0.5	1.3	0.3	0.4	1.4	3.9	0.9	1.1	1.3	1.5	0.7	0.7	0.7

この結果、これまでの二次構造に対するジペプチド解析と同様に、順番を考慮した傾向指数を得た。その中でも、一本鎖 β ストランド構造のアミノ酸残基に対する Chou-Fasman パラメータ (Table 3-5) から、プロリン残基の傾向指数は 2.5 を超える高い値であったため、プロリンを含むジペプチドは 1.5 以上の高い傾向指数を示した (PE, PG, DP は除く)。そして、プロリンとの組み合わせにおいても、順番によって傾向指数が異なった。特に、二つのジペプチド GP と PG は、それぞれ傾向指数が 2.7 と 1.1 のように大きく傾向指数が異なった。また、プロリンが関わらないジペプチドの傾向指数が 1.5 以上の値を示したものは、すべて 1.0 前後のアミノ酸残基の組み合わせだった。つまり、出現傾向がわずかながら高い、あるいは低いアミノ酸残基同士を、一本鎖 β ストランド構造の中でジペプチドとして見た場合、出現傾向が比較的、高かった。それらの中でも、システインとヒスチジンは、アミノ酸残基の傾向指数がそれぞれ 0.97 と 0.98 のように 1.0 より低いにもかかわらず、CH は 1.6、HC は 1.5 だった。

7—4 ジペプチド単位でみる Richardson パラメータ

Richardson らは、蛋白質の二次構造に対する統計的解析において、二次構造の領域を対象にして解析する Chou-Fasman 解析を拡張し、二次構造の末端領域に対して、ポジションをさだめ、それぞれのポジションにおけるアミノ酸残基の傾向指数を求めた^{27,29,30,34,49}。Richardson 解析は、あたかも、配列情報を保持しているかのような解析方法であるが、本質的には、個々のアミノ酸残基に対する Chou-Fasman 解析と同じである。つまり、あるポジションで、特定のアミノ酸残基の傾向指数はわかるが、その隣接するポジションにくるアミノ酸残基がどのようなアミノ酸残基がくるのかという情報は含まれない。あくまでも、隣接するポジションにどのようなアミノ酸残基がきやすいのかという情報のみである。

ところで、7—3 節では、ジペプチドを基本単位とし、Chou-Fasman パラメータを算出した。その結果、個々のアミノ酸残基を基本単位としたときの Chou-Fasman 解析からは、得られない有用な情報を得ることができた。つまり、配列を考慮した傾向指数が得られた。しかしながら、ジペプチドの Chou-Fasman 解析は、二次構造の内部に対する解析である。

この章では、 α ヘリックス構造と β ストランド構造の末端領域に対して、ジペプチドを基本単位とした Richardson パラメータを算出し、調べることにした。

7—4—1 α ヘリックス構造の統計解析

1081 蛋白質に対して、ProSSA 法で α ヘリックス構造を帰属した。その α ヘリックス構造の末端境界領域に対して、次のようにポジションを定義した； α ヘリックス構造の N 末端のアミノ酸残基の位置を Ncap ポジションと名付け、このポジションから、 α ヘリックスの C 末端の方向へ四つのアミノ酸残基に対して、ポジション N1, N2, N3, N4 をラベリングした。同様に、 α ヘリックスの構造の C 末端のアミノ酸残基の位置を Ccap ポジションと名付け、 α ヘリックス構造の N 末端の方向へ四つのアミノ酸残基に対して、ポジション C1, C2, C3, C4 をラベリングした。(Figure 7-4)。

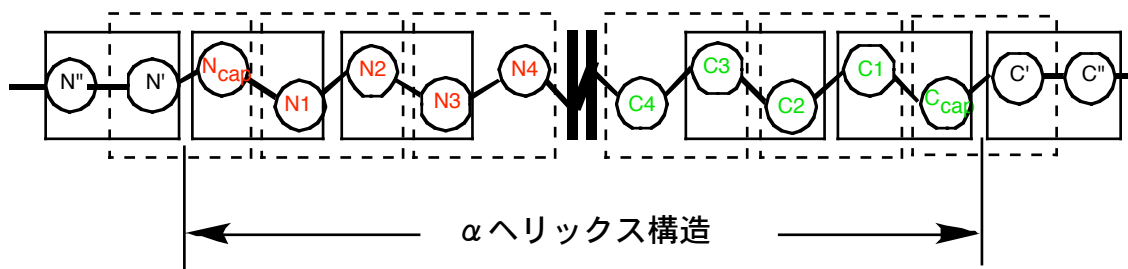


Figure 7-4. α ヘリックス構造の末端領域に対するポジション

このポジションでのアミノ酸残基の傾向指数は、次の通りである (Table 7-11)¹¹⁾。

Table 7-11. α ヘリックス構造の末端領域でのアミノ酸残基の傾向指数

	N''	N'	Ncap	N1	N2	N3	N4	C4	C3	C2	C1	Ccap	C'	C''
A	0.9	0.7	0.4	1.1	1.3	1.2	1.4	1.7	1.4	1.4	1.2	0.6	0.8	0.8
C	1.1	1.1	0.9	0.6	0.6	0.8	1.0	1.4	0.9	0.8	1.2	1.2	0.8	0.7
D	0.9	0.9	2.7	0.9	1.7	1.5	0.5	0.6	0.7	1.0	0.6	0.9	1.1	1.6
E	1.0	0.8	0.6	1.3	2.5	2.0	0.6	0.9	1.2	1.6	1.0	0.6	0.8	1.1
F	1.0	1.1	0.4	0.9	0.7	1.0	1.4	1.3	1.1	0.7	1.1	1.1	0.9	0.8
G	1.5	1.1	1.7	0.5	0.7	0.6	0.3	0.4	0.3	0.3	0.4	3.7	1.1	1.1
H	1.1	1.0	1.2	0.8	1.0	1.0	0.7	0.8	1.0	0.9	1.4	1.3	0.8	1.2
I	0.9	1.2	0.2	0.8	0.5	0.7	1.6	1.2	1.1	0.8	0.7	0.7	0.9	0.8
K	0.9	0.8	0.5	0.9	1.0	0.9	0.9	0.8	1.4	1.8	1.3	0.9	1.2	1.2
L	0.8	1.2	0.3	1.0	0.6	0.9	1.8	1.9	1.6	1.1	1.5	0.8	0.8	0.7
M	0.8	1.4	0.3	0.9	0.6	0.9	1.4	1.7	1.3	1.2	1.3	0.8	0.8	0.6
N	1.1	0.9	2.3	0.5	0.9	0.7	0.5	0.5	0.7	0.9	1.3	1.6	1.0	1.1
P	1.3	1.3	1.4	3.2	0.5	0.3	0.2	0.4	0.2	0.3	0.0	0.0	3.0	1.9
Q	0.9	0.9	0.5	0.9	1.4	1.7	1.1	1.0	1.2	1.3	1.3	0.9	0.8	0.9
R	1.0	0.8	0.5	0.9	1.0	0.9	1.4	0.9	1.4	1.5	1.2	0.8	0.9	1.0
S	1.1	0.9	2.6	0.7	1.3	1.0	0.5	0.7	0.8	1.1	1.1	0.8	0.9	1.1
T	1.0	1.0	2.0	0.8	0.8	1.0	0.7	0.6	0.6	0.7	1.1	0.7	0.9	1.0
V	0.9	1.1	0.2	0.9	0.5	0.9	1.4	0.9	0.8	0.6	0.6	0.6	0.7	0.6
W	0.7	0.9	0.4	1.5	1.2	1.0	1.4	1.7	1.1	0.8	0.8	0.5	0.7	0.8
Y	1.0	1.0	0.6	1.0	0.8	0.9	1.2	1.1	0.9	0.7	1.2	1.0	0.9	0.9

この表は、Figure 3-1 の表と同じものである。そして、この節でのポジションでのアミノ酸残基の傾向指数は、すべて、この表の結果を指している。

この節では、隣り合う二つのポジションを一組として、N 末端側から、N''N', N'Ncap,

NcapN1, N1N2, N2N3, N3N4, C 末端側から, C4C3, C3C2, C2C1, C1C', C'C''をジペプチドの解析に使うポジションとして, ジペプチドの解析をした。

まず, α ヘリックス構造の N 末端の境界領域で N''N'ポジションでのジペプチドの数を調べた (Table 7-12)。

Table 7-12. α ヘリックス構造のN''N'ポジションでのジペプチドの出現頻度

N' \ N''	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	41	5	41	23	19	57	7	25	37	49	22	24	25	13	17	35	31	37	6	13	527
C	6	4	3	5	3	8	5	4	1	9	2	6	8	3	5	9	5	8	0	2	96
D	29	7	15	17	25	26	10	35	14	46	13	7	26	14	11	15	13	27	6	16	372
E	26	4	20	14	19	31	11	29	25	48	20	18	36	12	18	14	20	46	5	17	433
F	18	4	19	11	13	27	12	23	11	31	2	7	20	8	11	18	18	14	1	5	273
G	54	10	25	42	37	56	25	62	48	84	32	17	22	21	34	33	43	71	10	27	753
H	13	6	10	10	6	13	4	6	5	24	9	6	17	4	10	7	8	9	2	3	172
I	25	0	16	22	13	30	10	19	15	36	4	21	19	11	19	21	19	25	2	15	342
K	21	2	23	25	17	22	5	19	15	50	16	15	19	12	8	21	12	31	7	21	361
L	28	10	26	21	8	40	7	22	24	50	10	28	36	20	26	26	30	37	5	9	463
M	4	1	6	5	7	8	4	8	5	12	3	6	13	4	4	6	11	15	1	1	124
N	21	5	14	14	22	23	8	31	11	36	10	11	23	15	5	14	22	30	6	16	337
P	24	4	28	21	18	36	9	21	16	49	11	17	23	22	15	24	25	37	5	16	421
Q	13	10	11	15	17	26	3	23	11	24	5	10	8	8	10	8	10	18	2	10	242
R	13	4	19	13	16	27	3	30	13	39	3	13	22	14	13	15	17	30	7	14	325
S	22	8	25	20	19	39	9	33	21	58	12	10	23	16	20	29	32	24	9	11	440
T	14	5	26	26	21	48	14	23	12	45	11	13	21	11	18	15	22	24	4	15	388
V	27	8	22	27	16	38	9	29	25	45	10	21	39	12	20	20	32	29	1	13	443
W	4	0	5	3	1	5	1	2	4	3	4	7	3	2	2	4	6	6	1	3	66
Y	15	6	7	10	11	25	8	12	10	20	2	16	20	7	16	15	17	18	6	9	250
Total	418	103	361	344	308	585	164	456	323	758	201	273	423	229	282	349	393	536	86	236	6828

全体的に, システイン, メチオニン, トリプトファンが関わるジペプチドは, 他のジペプチドの出現頻度に比べて, 一桁の値を示した。これは, 蛋白質の現れるアミノ酸残基のとして, システイン, メチオニン, トリプトファンの数が少ないため, ジペプチドでカウントしたときにもその影響が現れたと考えた。このように, もともと蛋白質内にあまり存在しないアミノ酸残基同士の組み合わせは, 出現頻度は低く。このことは, α ヘリックス構造だけでなく, 他の二次構造でも同じことがいえる。

つぎに, N''N'ポジションにおけるジペプチドの傾向指数を算出した (Table 7-13)。算出に使った計算式は, つぎの通りである;

$$\text{ポジションでのジペプチドの傾向指数} = \frac{\frac{\alpha\text{ヘリックス構造の中の特定のポジションに現れる特定のジペプチドの数}}{\alpha\text{ヘリックス構造の特定のポジションに現れるすべてのジペプチドの数}}}{\frac{\text{蛋白質の中の特定のジペプチドの数}}{\text{蛋白質の中のすべてのジペプチドの数}}}$$

Table 7-13. α ヘリックス構造のN''N'ポジションでのジペプチドの傾向指数

N' \ N''	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.7	0.7	1.3	0.6	0.9	1.4	0.5	0.8	1.1	0.9	1.6	1.0	1.1	0.6	0.6	1.1	1.0	0.9	0.7	0.7
C	0.9	2.4	0.6	1.0	0.8	1.1	1.7	0.8	0.2	1.2	1.1	1.3	1.6	0.8	1.0	1.5	0.9	1.4	0.0	0.6
D	0.8	1.4	0.8	0.6	1.3	0.8	1.2	1.3	0.6	1.3	1.6	0.4	1.2	1.1	0.6	0.7	0.6	0.9	0.9	1.0
E	0.7	0.8	0.8	0.4	1.1	1.1	1.3	1.1	0.8	1.1	1.7	0.9	2.3	0.7	0.8	0.6	0.9	1.4	0.7	1.2
F	0.9	1.0	1.1	0.6	1.4	1.3	1.8	1.6	0.6	1.3	0.4	0.5	1.5	0.8	0.9	1.0	1.1	0.8	0.3	0.5
G	1.3	1.3	0.9	1.5	1.7	1.4	2.1	1.9	1.4	2.0	2.9	0.8	1.2	1.1	1.3	1.0	1.4	1.9	1.5	1.3
H	1.1	2.2	1.2	1.1	0.8	1.0	0.9	0.6	0.7	1.6	2.6	1.0	1.7	0.7	1.3	0.8	1.0	0.9	0.8	0.5
I	0.7	0.0	0.6	0.9	1.0	1.1	1.1	0.9	0.6	1.2	0.6	1.1	0.9	0.7	1.0	0.9	0.7	1.0	0.5	1.2
K	0.6	0.5	0.9	0.8	1.1	0.8	0.6	0.8	0.6	1.4	1.7	0.8	1.0	0.8	0.4	0.9	0.6	1.0	1.4	1.5
L	0.5	1.3	0.7	0.5	0.4	0.9	0.5	0.8	0.6	1.0	0.8	1.0	1.1	0.8	0.7	0.7	0.9	1.0	0.8	0.5
M	0.3	0.6	0.6	0.5	1.2	0.7	1.3	0.9	0.5	0.9	0.8	0.8	1.6	0.7	0.5	0.8	1.3	1.5	0.8	0.2
N	0.9	1.2	0.8	0.9	1.8	0.9	1.2	1.6	0.7	1.3	1.5	0.8	1.2	1.3	0.4	0.9	1.3	1.3	1.3	1.4
P	0.9	1.0	1.3	0.8	1.3	1.4	1.3	1.3	1.0	1.9	1.8	1.2	1.9	1.7	1.1	1.2	1.5	1.6	1.1	1.4
Q	0.5	3.0	0.8	0.8	1.6	1.4	0.5	1.5	0.7	0.9	0.9	0.9	0.7	0.6	0.8	0.6	0.8	1.0	0.5	1.0
R	0.5	0.8	0.9	0.5	1.0	1.1	0.4	1.5	0.8	1.2	0.5	0.9	1.6	1.0	0.8	0.8	1.0	1.2	1.4	1.0
S	0.7	1.3	1.1	0.9	1.1	1.1	0.9	1.5	1.0	1.6	1.5	0.6	1.2	1.0	1.1	1.1	1.5	0.9	1.4	0.7
T	0.5	0.9	1.2	1.2	1.3	1.5	1.5	1.0	0.6	1.2	1.5	0.9	1.0	0.8	1.1	0.7	1.0	0.9	0.6	1.1
V	0.7	1.2	0.7	0.9	0.9	1.2	0.8	1.0	0.9	1.1	1.0	1.0	1.7	0.7	0.8	0.7	1.1	0.8	0.2	0.9
W	0.6	0.0	0.8	0.5	0.3	0.7	0.4	0.4	0.7	0.4	2.0	1.5	0.9	0.4	0.4	0.8	1.1	1.0	0.8	0.9
Y	0.8	1.8	0.4	0.7	1.2	1.3	1.5	1.0	0.8	1.0	0.4	1.3	1.6	0.7	1.2	1.0	1.2	1.2	1.6	0.9

その結果、傾向指数は、1.0 に収束する傾向だった (Figure 7-5(a))。そして、最大の傾向指数を示したジペプチドは、QC だった。また、アミノ酸残基に対する Richardson パラメータで、N''ポジションの最大傾向指数のグリシンと N'ポジションの最大傾向指数のメチオニンの組み合わせである GM の傾向指数も 2.9 の高い傾向指数を示した。最小の傾向指数を示したジペプチドは、CW, WC, IC だった。もともとシステインとトリプトファンの二つのアミノ酸残基は蛋白質中에서도出現頻度が低く、それら二つのアミノ酸残基を組み合わせの場合、 α ヘリックス構造の N''N'ポジションでは、出現頻度が0 だったためである。

ところで、N''と N'のそれぞれのポジションでのアミノ酸残基の傾向指数が 1.0 以上の組み合わせのジペプチドの中で、ジペプチドの傾向指数が 0.5 以下の値をもつものは、FM, FY, HY,

RM, YM だった。一方で、それぞれのポジションでのアミノ酸残基の傾向指数が 1.0 以下の組み合わせのジペプチドの中で、ジペプチドの傾向指数が 1.5 以上の値をもつものは、FH, YH, YW だった。ただし、YH と YW に関わっているアミノ酸残基は、傾向指数が 1.0 同士である。つまり、特徴が見られないアミノ酸残基同士の組み合わせをジペプチドでみたことで、特徴のある結果を生み出した。

つぎに、N'Ncap ポジションでのジペプチドの数を調べた (Table 7-14)。

Table 7-14. α ヘリックス構造のN'Ncapポジションでの出現頻度

N' \ Ncap	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	12	8	70	14	9	60	22	2	17	8	5	43	22	5	15	52	55	2	6	6	433
C	4	1	19	2	2	17	4	0	2	2	2	10	10	0	3	16	7	2	0	0	103
D	16	3	52	13	2	50	21	8	7	14	3	48	6	9	12	57	25	8	3	9	366
E	4	3	61	15	7	53	6	5	8	9	5	47	9	6	9	50	31	6	2	13	349
F	11	3	54	5	4	42	6	3	11	9	0	30	13	5	6	61	39	5	0	4	311
G	12	8	88	17	13	97	15	4	27	16	6	64	15	16	17	82	62	17	2	14	592
H	9	5	24	5	4	18	6	3	5	2	0	15	8	6	5	27	12	3	4	6	167
I	14	5	80	13	7	57	11	0	16	9	1	53	55	7	7	68	54	2	0	2	461
K	6	4	64	17	7	24	9	0	8	13	3	37	21	12	9	49	33	2	0	9	327
L	23	8	107	24	7	77	16	1	23	13	4	70	83	13	15	143	121	6	2	14	770
M	6	1	35	7	5	15	2	2	3	10	0	18	29	5	6	34	37	2	1	3	221
N	7	1	50	11	1	36	5	8	12	6	4	38	9	4	6	42	24	2	5	6	277
P	13	11	59	21	12	57	17	6	7	11	5	32	21	9	11	66	41	11	4	12	426
Q	7	1	44	18	4	35	4	1	5	2	0	17	17	4	9	32	19	4	1	9	233
R	5	2	48	18	7	29	9	7	7	7	0	37	16	5	4	37	33	2	3	9	285
S	14	2	57	17	8	61	9	3	9	12	3	38	8	7	10	54	37	6	1	5	361
T	11	7	64	11	4	51	17	6	5	8	2	39	26	14	8	63	48	4	4	10	402
V	23	5	86	17	9	55	13	3	14	8	1	49	63	10	15	91	65	8	3	5	543
W	3	2	19	5	0	13	3	0	1	2	0	7	4	0	1	15	9	1	0	2	87
Y	6	3	46	8	3	23	6	2	10	7	4	25	11	1	5	31	37	2	1	6	237
Total	206	83	1127	258	115	870	201	64	197	168	48	717	446	138	173	1070	789	95	42	144	6951

この結果から、システイン、メチオニン、トリプトファンは、蛋白質の中でも比較的、出現頻度が少ないため、N'Ncap ポジションでのジペプチドの出現頻度でも一桁の値として現れた。また、イソロイシンやグルタミン酸、バリントリプトファンは、 α ヘリックス構造の中でも出現頻度は低いいため、ジペプチドの出現頻度にも影響を与えた。

さらに、N'Ncap ポジションでのジペプチドの傾向指数を算出した (Table 7-15)。

Table 7-15. α ヘリックス構造のN'Ncapポジションでのジペプチドの傾向指数

Ncap \ N'	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.2	1.1	2.1	0.4	0.4	1.4	1.7	0.1	0.5	0.1	0.4	1.8	1.0	0.2	0.5	1.7	1.8	0.0	0.7	0.3
C	0.6	0.6	3.6	0.4	0.5	2.2	1.3	0.0	0.5	0.3	1.1	2.2	2.0	0.0	0.6	2.6	1.3	0.3	0.0	0.0
D	0.5	0.6	2.6	0.5	0.1	1.5	2.5	0.3	0.3	0.4	0.4	2.9	0.3	0.7	0.6	2.5	1.1	0.3	0.4	0.5
E	0.1	0.6	2.4	0.5	0.4	1.8	0.7	0.2	0.2	0.2	0.4	2.3	0.6	0.3	0.4	2.2	1.3	0.2	0.3	0.9
F	0.5	0.7	3.0	0.3	0.4	2.1	0.9	0.2	0.6	0.4	0.0	2.3	1.0	0.5	0.5	3.4	2.3	0.3	0.0	0.4
G	0.3	1.1	3.0	0.6	0.6	2.4	1.2	0.1	0.8	0.4	0.5	2.9	0.8	0.8	0.7	2.5	1.9	0.5	0.3	0.7
H	0.8	1.8	2.7	0.5	0.5	1.4	1.3	0.3	0.7	0.1	0.0	2.5	0.8	1.0	0.6	3.1	1.4	0.3	1.5	1.0
I	0.4	1.0	3.0	0.5	0.5	2.0	1.1	0.0	0.6	0.3	0.1	2.6	2.7	0.5	0.4	2.7	2.1	0.1	0.0	0.2
K	0.2	0.9	2.5	0.6	0.4	0.9	1.1	0.0	0.3	0.3	0.3	1.9	1.1	0.8	0.5	2.1	1.5	0.1	0.0	0.6
L	0.4	1.0	2.9	0.6	0.3	1.8	1.1	0.0	0.6	0.2	0.3	2.4	2.5	0.5	0.4	3.7	3.4	0.2	0.3	0.7
M	0.4	0.5	3.7	0.7	0.8	1.3	0.6	0.2	0.3	0.8	0.0	2.4	3.6	0.8	0.7	4.4	4.2	0.2	0.8	0.6
N	0.3	0.2	2.9	0.7	0.1	1.4	0.7	0.4	0.7	0.2	0.6	2.7	0.5	0.4	0.4	2.5	1.4	0.1	1.0	0.5
P	0.5	2.6	2.6	0.8	0.9	2.1	2.4	0.4	0.4	0.4	0.8	2.2	1.7	0.7	0.8	3.2	2.4	0.5	0.9	1.0
Q	0.3	0.3	3.1	1.0	0.4	1.8	0.6	0.1	0.3	0.1	0.0	1.5	1.5	0.3	0.7	2.4	1.4	0.2	0.3	0.9
R	0.2	0.4	2.2	0.7	0.4	1.2	1.1	0.3	0.4	0.2	0.0	2.5	1.2	0.4	0.2	2.0	1.9	0.1	0.6	0.7
S	0.4	0.3	2.5	0.7	0.4	1.8	0.9	0.1	0.4	0.3	0.4	2.3	0.4	0.4	0.5	2.0	1.7	0.2	0.2	0.3
T	0.4	1.2	2.8	0.5	0.2	1.6	1.8	0.3	0.3	0.2	0.3	2.6	1.2	1.0	0.5	2.9	2.2	0.1	0.6	0.7
V	0.5	0.8	2.7	0.5	0.5	1.7	1.1	0.1	0.5	0.2	0.1	2.2	2.7	0.6	0.6	3.2	2.2	0.2	0.5	0.3
W	0.4	1.5	3.1	0.8	0.0	1.8	1.1	0.0	0.2	0.2	0.0	1.5	1.1	0.0	0.2	2.8	1.6	0.2	0.0	0.6
Y	0.3	0.9	2.9	0.5	0.3	1.2	1.1	0.2	0.7	0.3	0.8	2.0	0.9	0.1	0.4	2.0	2.6	0.1	0.3	0.6

この結果から、傾向指数の分布が0.5 付近と 2.4 付近に集まっていた (Figure 7-5(b))。最大の傾向指数を示したジペプチドは、MS だった。このアミノ酸残基の組み合わせは、N'ポジションにおいて、最大の傾向指数を示すメチオニンと、Ncap ポジションで2番目に高い傾向指数を示すセリンだった。最小の傾向指数を示したジペプチドは、AV, CI, CQ, CW, CY, FM, FW, HM, II, IW, KI, KW, LI, MM, QM, RM, WF, WI, WM, WQ, WW だった。これらは、すべて出現頻度は0 だった。

ところで、Ncap ポジションにアスパラギン酸、アスパラギン、セリン、スレオニンが関わるジペプチドの傾向指数は、ほとんどが1.5 以上だった。これらのアミノ酸残基がNcap ポジションにくるのは、N capping box の形成に関わるためである^{27,30,50}。つまり、N'ポジションのアミノ酸残基の種類に関係なく、Ncap ポジションには、アスパラギン酸、アスパラギン、セリン、スレオニンは出現しやすいといえる。ただし、DT のみは、傾向指数が1.1 と他に比べて少し傾向指数が低かった。また、Ncap ポジションにグリシンが存在している場合も、ジペプチドの傾向指数のほとんどが1.5 以上だった。このことから、N'ポジションのアミノ酸残基の種類に関わらず、Ncap ポジションにグリシン残基が出現しやすいといえる。

そのほかに、N'と Ncap のそれぞれのポジションにおいて、1.0 以上の傾向指数をもつアミノ酸残基の組み合わせで、ジペプチドの傾向指数が0.5 以下は存在しなかった。一方で、1.0 以下の傾向指数をもつアミノ酸残基の組み合わせで、ジペプチドの傾向指数が 1.5 以上は HW と WC だった。

つぎに、NcapN1 ポジションでのジペプチドの出現頻度を調べた (Table 7-16)。

Table 7-16. α ヘリックス構造のNcapN1ポジションでのジペプチドの出現頻度

Ncap	N1	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	16	4	15	16	5	6	3	9	9	23	5	7	32	5	12	10	15	9	6	7	214	
C	6	0	6	1	1	5	0	5	2	8	1	4	19	1	6	5	6	4	0	4	84	
D	95	5	48	99	49	31	23	61	64	115	15	26	183	30	47	52	57	78	31	39	1148	
E	24	4	12	19	7	8	4	9	17	17	7	5	52	12	13	15	9	18	7	8	267	
F	8	0	4	8	3	8	2	2	8	8	3	2	23	6	4	10	6	5	2	6	118	
G	81	14	56	58	42	48	18	44	46	86	25	27	63	31	36	53	50	62	12	31	883	
H	17	1	7	9	6	11	3	9	7	18	3	7	60	11	9	6	4	5	5	6	204	
I	5	0	5	4	0	3	2	1	5	2	0	4	15	1	2	3	10	2	1	1	66	
K	17	1	8	16	10	8	2	3	21	16	2	4	35	6	8	11	9	16	2	7	202	
L	14	0	15	15	5	7	7	3	8	8	1	6	44	4	11	11	5	4	4	3	175	
M	2	0	3	5	0	4	1	0	2	2	0	1	18	3	3	2	3	3	0	3	55	
N	60	5	32	63	29	21	19	51	35	71	16	18	117	19	34	20	31	44	19	30	734	
P	51	3	26	41	13	6	9	30	32	52	18	8	36	29	19	15	15	30	5	16	454	
Q	21	3	4	8	4	4	0	10	4	13	1	4	28	5	6	8	3	3	3	7	139	
R	10	2	7	15	3	9	5	9	10	20	3	2	29	5	14	8	7	10	7	4	179	
S	107	9	76	99	39	33	23	39	65	96	20	20	144	56	52	40	46	77	19	37	1097	
T	75	6	50	76	32	22	8	36	44	83	14	18	132	28	29	27	30	52	13	26	801	
V	11	0	10	6	2	5	0	7	3	1	0	1	23	2	3	4	9	3	4	8	102	
W	4	0	0	2	2	2	1	0	3	2	1	1	3	3	6	1	6	3	2	1	43	
Y	15	1	6	8	3	8	5	11	10	7	1	2	33	5	3	9	6	7	5	4	149	
Total	639	58	390	568	255	249	135	339	395	648	136	167	1089	262	317	310	327	435	147	248	7114	

NcapN1 ポジションでのジペプチドの傾向指数は、次の通りである (Table 7-17)。

Table 7-17. α ヘリックス構造のNcapN1ポジションでのジペプチドの傾向指数

	N1	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Ncap																					
A		0.3	0.5	0.4	0.4	0.2	0.1	0.2	0.3	0.3	0.4	0.3	0.3	1.4	0.2	0.4	0.3	0.5	0.2	0.7	0.4
C		0.8	0.0	1.1	0.2	0.3	0.6	0.0	1.0	0.5	1.0	0.5	0.9	3.7	0.2	1.2	0.8	1.1	0.7	0.0	1.2
D		2.6	1.0	2.3	3.5	2.4	0.9	2.7	2.1	2.8	3.0	1.8	1.5	8.0	2.2	2.4	2.3	2.6	2.6	4.4	2.3
E		0.6	0.8	0.5	0.6	0.4	0.3	0.4	0.3	0.5	0.4	0.6	0.2	3.2	0.6	0.6	0.7	0.4	0.5	1.0	0.5
F		0.4	0.0	0.2	0.4	0.3	0.4	0.3	0.1	0.5	0.3	0.5	0.1	1.7	0.6	0.3	0.5	0.3	0.3	0.5	0.6
G		1.9	1.8	1.9	2.0	1.9	1.2	1.5	1.3	1.3	2.0	2.2	1.2	3.3	1.5	1.4	1.6	1.5	1.6	1.7	1.5
H		1.4	0.3	0.8	1.0	0.7	0.8	0.6	0.9	0.9	1.2	0.8	1.1	5.7	1.8	1.1	0.7	0.5	0.5	1.8	1.0
I		0.1	0.0	0.2	0.1	0.0	0.1	0.2	0.0	0.2	0.1	0.0	0.2	0.7	0.1	0.1	0.1	0.4	0.1	0.2	0.1
K		0.5	0.2	0.3	0.5	0.6	0.3	0.2	0.1	0.8	0.4	0.2	0.2	1.8	0.4	0.4	0.5	0.4	0.5	0.4	0.5
L		0.3	0.0	0.4	0.4	0.2	0.2	0.5	0.1	0.2	0.1	0.1	0.2	1.3	0.2	0.3	0.3	0.1	0.1	0.6	0.1
M		0.1	0.0	0.3	0.5	0.0	0.4	0.3	0.0	0.2	0.2	0.0	0.1	2.2	0.5	0.4	0.3	0.3	0.3	0.0	0.6
N		2.4	1.1	1.8	3.7	2.3	0.8	2.6	2.5	2.0	2.5	2.2	1.2	6.0	1.6	2.5	1.2	1.8	1.9	3.8	2.5
P		1.9	0.7	1.1	1.5	0.9	0.2	1.2	1.7	2.0	1.9	2.8	0.5	2.8	2.2	1.3	0.7	0.9	1.2	1.1	1.3
Q		0.8	0.9	0.3	0.4	0.4	0.2	0.0	0.6	0.2	0.5	0.2	0.3	2.4	0.4	0.4	0.6	0.2	0.2	0.8	0.7
R		0.4	0.4	0.3	0.6	0.2	0.4	0.6	0.4	0.6	0.6	0.5	0.1	2.0	0.4	0.8	0.4	0.4	0.4	1.3	0.3
S		3.4	1.4	3.2	4.1	2.1	0.9	2.1	1.7	3.0	2.6	2.4	1.2	7.4	3.3	2.6	1.4	2.0	2.8	2.9	2.4
T		2.4	1.0	2.2	3.2	1.9	0.7	0.8	1.5	2.2	2.1	1.8	1.2	6.0	2.0	1.7	1.2	1.3	1.8	1.9	1.8
V		0.3	0.0	0.3	0.2	0.1	0.2	0.0	0.2	0.1	0.0	0.0	0.0	1.0	0.1	0.1	0.1	0.3	0.1	0.6	0.5
W		0.6	0.0	0.0	0.3	0.5	0.3	0.4	0.0	0.5	0.2	0.5	0.2	0.8	0.6	1.0	0.2	1.0	0.5	1.6	0.3
Y		0.8	0.3	0.4	0.5	0.3	0.4	0.9	0.9	0.7	0.3	0.2	0.2	2.5	0.5	0.2	0.6	0.4	0.4	1.3	0.4

その結果、傾向指数は、0.4 と 2.0 と 2.8 に最大値が見られた (Figure 7-5(c))。最大の傾向指数を示したジペプチドは、DP だった。これは、Ncap ポジションでのアミノ酸残基の傾向指数が最大であるアスパラギン酸と、N1 ポジションでのアミノ酸残基の傾向指数が最大であるプロリンの組み合わせのためである。最小の傾向指数を示したジペプチドは、CC, CH, CW, FC, IC, IF, IL, IM, LC, MC, MF, MI, MM, MW, QH, VC, VH, VL, VM, VN, WC, WD, WI (すべて 0.0) の 23 種類だった。また、ジペプチドの中で、Ncap ポジションにアスパラギン酸、グリシン、アスパラギン、セリン、スレオニンが存在する場合、ほとんどの傾向指数は、1.5 以上を示した。これら 5 つのアミノ酸残基は、Ncap ポジションでのアミノ酸残基の傾向指数においても、1.5 以上の値であるため、その影響がジペプチドの傾向指数にも表れたと考えた。それらの中でも、アスパラギン酸、アスパラギン、セリン、スレオニンは、N'Ncap ポジションのところでも述べたように、N capping box 形成に関わるためである。ただし、DG, NG, SG, TG, TH は、傾向指数が 1.0 未満であるため、ジペプチドにおける N1 ポジションにグリシン、ヒスチジンが存在する可能性は、低いといえる。さらに、ジペプチドの中で、N1 ポジションにプロリンやトリプトファンが存在するとき、とくにプロリンの場合は、ほとんどの傾向指数は、

1.5 以上だった。これは、N1 ポジションでのプロリンやトリプトファンの傾向指数が、1.5 以上であるため、その影響がジペプチドでも現れたと考えた。それらの中でも、プロリンは、N 末端の巻き始めであることから、N1 ポジションにプロリンが関わるジペプチドの傾向指数は、1.5 以上の値を示した。ただし、IP、WP は、1.0 未満の値だった。また、傾向指数が 0.5 以下のジペプチドは、Ncap と N1 のそれぞれのポジションでのアミノ酸残基の傾向指数が 0.5 以下のものが関わっているものがほとんどだった。

ところで、Ncap と N1 のそれぞれのポジションでのアミノ酸残基の傾向指数が 1.0 以上のアミノ酸残基の組み合わせの中で、ジペプチド HT は、傾向指数が 0.5 だった。一方で、1.0 以下のアミノ酸残基の組み合わせで、ジペプチドの傾向指数が 1.5 以上のものは存在しなかった。

つぎに、N1N2 ポジションでのジペプチドの出現頻度を調べた (Table 7-18)。

Table 7-18. α ヘリックス構造のN1N2ポジションでのジペプチドの出現頻度

N2 \ N1	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	73	7	63	111	12	34	13	14	38	28	5	20	17	37	35	55	27	23	14	13	639
C	6	0	7	3	1	4	1	1	1	1	1	4	2	1	4	4	6	6	4	1	58
D	49	2	30	75	14	19	7	9	18	21	3	17	4	19	17	27	19	16	4	20	390
E	68	3	50	107	9	28	9	8	54	22	8	25	13	30	31	42	25	20	5	11	568
F	20	4	30	38	8	12	4	3	16	11	3	16	7	8	12	22	18	5	7	11	255
G	35	6	24	21	9	16	4	13	11	15	5	10	11	10	10	11	12	14	3	9	249
H	9	1	14	18	7	4	8	4	9	7	5	5	0	5	9	8	10	7	3	2	135
I	32	3	28	41	14	17	10	11	24	20	7	21	10	16	12	29	19	13	5	7	339
K	32	0	44	85	8	16	4	7	34	13	2	19	13	19	13	38	24	11	7	6	395
L	52	6	79	98	14	37	18	16	47	22	7	31	20	33	42	56	23	18	10	21	650
M	11	0	16	23	1	8	2	6	11	13	3	8	4	4	6	3	8	5	1	3	136
N	21	1	15	26	8	9	6	2	6	10	3	5	4	6	4	8	9	13	4	7	167
P	118	8	90	224	34	36	18	37	59	66	17	33	4	70	54	77	41	47	17	40	1090
Q	27	2	36	50	1	15	8	8	8	19	5	3	5	23	10	18	12	5	4	3	262
R	34	3	36	58	11	16	9	14	13	10	2	18	8	13	9	25	10	10	7	11	317
S	41	3	16	36	13	20	10	12	20	21	2	16	4	18	17	25	15	8	5	8	310
T	36	3	26	37	6	15	10	11	22	18	2	10	7	17	17	36	24	14	7	9	327
V	46	5	44	64	9	30	14	8	27	19	5	22	15	39	14	30	18	15	2	9	435
W	9	0	16	14	6	8	5	10	5	11	2	3	7	8	15	9	8	3	6	2	147
Y	25	3	37	28	5	21	5	5	11	19	5	11	11	13	10	13	9	6	1	10	248
Total	744	60	701	1157	190	365	165	199	434	366	92	297	166	389	341	536	337	259	116	203	7117

N1N2 ポジションでのジペプチドの傾向指数は、次の通りである (Table 7-19)。

Table 7-19. α ヘリックス構造のN1N2ポジションでのジペプチドの傾向指数

N1 \ N2	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	1.3	0.9	1.9	2.8	0.6	0.8	1.0	0.4	1.1	0.5	0.3	0.8	0.7	1.5	1.1	1.7	0.9	0.5	1.6	0.7
C	0.8	0.0	1.3	0.6	0.3	0.5	0.3	0.2	0.2	0.1	0.5	0.9	0.4	0.2	0.8	0.6	1.1	1.0	3.7	0.3
D	1.4	0.4	1.4	2.7	0.7	0.5	0.8	0.3	0.8	0.6	0.4	1.0	0.2	1.4	0.9	1.2	0.9	0.5	0.6	1.2
E	1.7	0.6	1.9	3.2	0.5	0.9	1.0	0.3	1.6	0.5	0.7	1.2	0.8	1.6	1.3	1.8	1.0	0.6	0.7	0.7
F	1.0	1.0	1.6	2.1	0.8	0.6	0.6	0.2	0.9	0.4	0.5	1.2	0.5	0.8	0.9	1.2	1.0	0.3	1.8	1.1
G	0.8	0.8	0.8	0.7	0.4	0.4	0.3	0.4	0.3	0.3	0.4	0.4	0.6	0.5	0.4	0.3	0.4	0.4	0.4	0.4
H	0.7	0.3	1.6	1.9	0.8	0.3	1.7	0.4	1.2	0.5	1.4	0.8	0.0	0.8	1.1	0.9	1.2	0.6	1.1	0.3
I	0.9	0.6	1.0	1.5	1.1	0.6	1.0	0.5	0.9	0.6	0.9	1.0	0.5	1.0	0.6	1.1	0.7	0.5	1.2	0.5
K	0.9	0.0	1.7	2.7	0.5	0.6	0.5	0.3	1.3	0.3	0.2	1.0	0.7	1.2	0.7	1.6	1.1	0.4	1.3	0.4
L	0.9	0.7	2.1	2.3	0.7	0.8	1.2	0.5	1.1	0.4	0.5	1.0	0.6	1.2	1.1	1.4	0.6	0.5	1.5	1.0
M	0.8	0.0	1.6	2.3	0.2	0.7	0.6	0.7	1.0	1.0	0.8	1.0	0.5	0.7	0.7	0.4	0.9	0.5	0.8	0.6
N	0.8	0.2	0.9	1.5	0.6	0.3	0.8	0.1	0.3	0.4	0.4	0.3	0.2	0.5	0.3	0.5	0.5	0.6	0.8	0.6
P	4.3	1.9	3.9	8.2	2.4	1.3	2.5	2.1	3.7	2.4	2.6	2.2	0.3	5.2	3.8	3.6	2.3	1.9	3.6	3.3
Q	1.0	0.6	2.4	2.6	0.1	0.8	1.2	0.5	0.5	0.7	0.9	0.3	0.4	1.8	0.7	1.3	0.9	0.3	1.0	0.3
R	1.2	0.6	1.6	2.3	0.7	0.6	1.1	0.7	0.7	0.3	0.3	1.2	0.6	0.9	0.5	1.4	0.6	0.4	1.3	0.8
S	1.3	0.5	0.7	1.5	0.7	0.6	0.9	0.5	0.9	0.6	0.2	0.9	0.2	1.1	0.9	0.9	0.7	0.3	0.8	0.5
T	1.2	0.5	1.1	1.6	0.4	0.5	1.1	0.5	1.1	0.5	0.3	0.6	0.3	1.2	1.0	1.6	1.1	0.5	1.0	0.6
V	1.1	0.7	1.4	2.0	0.5	0.9	1.2	0.3	0.9	0.5	0.5	1.0	0.6	2.2	0.5	1.0	0.6	0.4	0.3	0.6
W	1.3	0.0	2.5	2.2	1.5	1.1	1.8	2.0	0.8	1.3	0.9	0.6	2.0	1.6	2.6	1.6	1.4	0.5	4.9	0.6
Y	1.3	0.8	2.2	1.9	0.5	1.1	0.9	0.4	0.8	0.9	1.0	0.8	0.8	1.2	0.7	0.8	0.6	0.4	0.3	1.0

この結果から、ジペプチドの傾向指数は、0.6 と 1.0 と 1.8 に最大値が見られた (Figure 7-5(d))。最大の傾向指数を示したジペプチドは、PE だった。このジペプチドは、N1 と N2 ポジションでのアミノ酸残基の傾向指数において、最大の傾向指数をもつプロリンとグルタミン酸の組み合わせである。最小の傾向指数を示したジペプチドは、CC, HP, KC, MC, WC の 5 種類だった。これら 5 種類のジペプチドは、 α ヘリックス構造の中に現れなかったジペプチドである。また、ジペプチドの傾向指数が 1.5 以上のものは、N1 と N2 のそれぞれのポジションでのアミノ酸残基の傾向指数が 1.5 以上のものが関わっていた。

ところで、N1 と N2 のそれぞれのポジションでのアミノ酸残基の傾向指数が 1.0 以上のものの組み合わせの中で、ジペプチド YW は、傾向指数が 0.3 だった。一方で、アミノ酸残基が 1.0 以下のものの組み合わせの中で、ジペプチド HH は、傾向指数が 1.7 だった。

つぎに、N2N3 ポジションでのジペプチドの出現頻度を調べた (Table 7-20)。

Table 7-20. α ヘリックス構造のN2N3ポジションでのジペプチドの出現頻度

N3 \ N2	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	89	8	67	97	32	41	14	26	33	55	23	23	8	43	29	45	36	41	16	19	745
C	8	0	3	6	2	3	3	3	0	6	1	3	1	2	1	5	4	8	0	1	60
D	77	6	57	98	25	22	11	42	45	67	10	16	5	37	31	34	36	44	7	31	701
E	111	5	102	195	38	33	21	41	65	95	23	45	7	87	47	49	69	73	15	38	1159
F	20	4	16	24	13	8	4	10	9	10	6	5	2	9	10	9	10	12	3	7	191
G	34	8	25	32	11	27	12	25	22	36	6	5	11	23	12	20	17	20	4	15	365
H	19	7	16	15	4	4	7	11	3	16	3	7	1	9	3	11	12	11	3	3	165
I	6	2	22	24	10	15	5	4	15	13	2	7	10	10	6	14	16	9	1	8	199
K	36	2	54	69	10	19	6	15	24	33	13	14	6	31	13	27	26	17	4	14	433
L	43	5	36	41	11	19	7	13	11	27	7	13	16	16	16	19	30	22	6	8	366
M	12	0	7	3	4	8	3	5	5	8	2	4	6	4	3	6	3	7	1	1	92
N	32	0	21	25	18	9	4	17	23	32	6	5	4	17	16	19	18	20	2	9	297
P	25	1	4	21	9	9	4	6	3	17	7	3	0	14	7	10	6	12	4	4	166
Q	38	3	38	73	16	10	16	14	18	25	5	9	6	36	18	22	13	16	5	8	389
R	33	6	41	35	18	13	9	17	20	18	3	7	0	23	22	20	15	24	7	9	340
S	46	3	45	57	23	20	9	22	30	48	8	19	5	40	29	28	41	33	14	15	535
T	45	5	26	49	5	14	7	13	14	24	7	10	7	21	17	18	21	20	6	8	337
V	39	4	23	27	12	16	11	6	11	15	4	4	9	12	12	14	16	18	3	4	260
W	9	1	8	14	5	2	5	5	6	10	5	3	0	11	5	10	6	4	0	7	116
Y	18	2	17	16	11	11	7	5	5	22	7	9	5	16	9	20	8	8	2	5	203
Total	740	72	628	921	277	303	165	300	362	577	148	211	109	461	306	400	403	419	103	214	7119

さらに、N2N3 ポジションでのジペプチドの傾向指数を算出した (Table 7-21)。

Table 7-21. α ヘリックス構造のN2N3ポジションでのジペプチドの傾向指数

N3 \ N2	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	1.5	1.0	2.0	2.5	1.5	0.9	1.0	0.8	0.9	1.0	1.6	1.0	0.3	1.8	0.9	1.4	1.1	1.0	1.9	1.0
C	1.1	0.0	0.5	1.2	0.5	0.4	1.0	0.6	0.0	0.7	0.5	0.6	0.2	0.5	0.2	0.8	0.7	1.3	0.0	0.3
D	2.1	1.2	2.7	3.5	1.2	0.6	1.3	1.5	1.9	1.8	1.2	0.9	0.2	2.7	1.6	1.5	1.6	1.4	1.0	1.8
E	2.8	1.0	3.9	5.8	2.1	1.1	2.3	1.5	2.0	2.1	1.9	2.1	0.4	4.7	2.0	2.1	2.8	2.1	2.1	2.5
F	1.0	1.0	0.9	1.3	1.3	0.4	0.6	0.7	0.5	0.4	1.0	0.4	0.1	0.9	0.8	0.5	0.6	0.6	0.8	0.7
G	0.8	1.0	0.8	1.1	0.5	0.7	1.0	0.8	0.6	0.8	0.5	0.2	0.6	1.1	0.5	0.6	0.5	0.5	0.6	0.7
H	1.6	2.4	1.8	1.6	0.5	0.3	1.5	1.1	0.4	1.0	0.8	1.1	0.1	1.5	0.4	1.2	1.4	1.0	1.1	0.5
I	0.2	0.4	0.8	0.9	0.8	0.5	0.5	0.2	0.6	0.4	0.3	0.3	0.5	0.7	0.3	0.5	0.6	0.3	0.2	0.6
K	1.0	0.5	2.1	2.2	0.6	0.7	0.7	0.6	0.9	0.9	1.3	0.7	0.3	2.0	0.7	1.1	1.2	0.5	0.8	0.9
L	0.8	0.6	0.9	1.0	0.5	0.4	0.5	0.4	0.3	0.5	0.5	0.4	0.5	0.6	0.4	0.5	0.8	0.6	0.9	0.4
M	0.9	0.0	0.7	0.3	0.7	0.7	0.9	0.5	0.5	0.6	0.5	0.5	0.7	0.7	0.4	0.8	0.3	0.7	0.8	0.2
N	1.3	0.0	1.2	1.5	1.4	0.3	0.6	0.8	1.3	1.1	0.8	0.3	0.2	1.5	1.2	1.1	1.0	0.9	0.4	0.8
P	0.9	0.2	0.2	0.8	0.6	0.3	0.5	0.3	0.2	0.6	1.1	0.2	0.0	1.0	0.5	0.5	0.3	0.5	0.9	0.3
Q	1.4	0.9	2.6	3.8	1.4	0.5	2.4	0.9	1.1	0.9	0.9	0.8	0.5	2.8	1.3	1.6	1.0	0.9	1.3	0.8
R	1.2	1.2	1.8	1.4	1.1	0.5	1.1	0.8	1.1	0.5	0.5	0.5	0.0	1.6	1.2	1.1	0.9	0.9	1.3	0.6
S	1.4	0.5	1.9	2.4	1.2	0.6	0.8	0.9	1.4	1.3	1.0	1.1	0.3	2.4	1.5	1.0	1.8	1.2	2.1	1.0
T	1.5	0.8	1.1	2.1	0.3	0.4	0.7	0.5	0.7	0.6	0.9	0.6	0.3	1.5	1.0	0.8	0.9	0.7	0.9	0.5
V	0.9	0.6	0.7	0.8	0.6	0.5	0.9	0.2	0.4	0.4	0.4	0.2	0.4	0.7	0.4	0.5	0.5	0.5	0.5	0.3
W	1.3	0.7	1.3	2.2	1.3	0.3	1.8	1.0	1.0	1.1	2.3	0.6	0.0	2.1	0.9	1.8	1.0	0.6	0.0	2.0
Y	1.0	0.6	1.0	1.1	1.1	0.6	1.3	0.4	0.4	1.0	1.4	0.7	0.4	1.5	0.6	1.3	0.6	0.5	0.5	0.5

その結果、ジペプチドの傾向指数は、0.6 と 1.0 と 2.2 に最大値が見られた (Table 7-5(e))。最

大の傾向指数を示したジペプチドは、EE だった。このジペプチドは、N2 と N3 のそれぞれのポジションでのアミノ酸残基の最大の傾向指数がともにグルタミン酸（N2 ポジションで 2.5, N3 で 2.0）であることが影響していると考えた。最小の傾向指数を示したジペプチドは、CC, CK, CW, MC, NC, PP, RP, WP, WW の 9 種類だった。これら 9 種類のジペプチドは、N2N3 ポジションでの出現頻度が 0 だったためである。また、N3 ポジションに、グルタミン酸、グルタミンが存在するジペプチドの傾向指数は、ほとんどが 1.5 以上だった。これも、N capping box の形成が関与していると考え。ただし、それらの中でも、PE, VE, CR, FQ, IQ, LQ, MQ, VQ は、N3 ポジションにグルタミン酸やグルタミンが現れたとしても、傾向指数が 1.0 未満であるため、出現傾向は低いといえる。また、1.5 以上の傾向指数を示したジペプチドは、アミノ酸残基での傾向指数でも、1.5 以上のアミノ酸残基の組み合わせが多い。

ところで、アミノ酸残基の傾向指数で 1.0 以上のアミノ酸残基同士の組み合わせにおいて、ジペプチドの傾向指数が 0.5 以下のジペプチドは、HF と WW だった。一方で、アミノ酸残基の傾向指数で 1.0 以下のアミノ酸残基同士の組み合わせにおいて、ジペプチドの傾向指数が 1.5 以上のジペプチドは、HC, HH, WH, WM, WS, WY だった。

つぎに、N3N4 ポジションでのジペプチドの出現頻度を調べた (Table 7-22)。

Table 7-22. α ヘリックス構造のN3N4ポジションでのジペプチドの出現頻度

N4 \ N3	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	92	5	20	30	30	17	17	57	46	132	19	18	8	34	55	23	23	66	19	29	740
C	4	2	8	2	3	4	1	1	7	8	2	4	0	3	6	2	5	3	3	4	72
D	76	8	14	17	48	8	12	69	28	106	24	9	2	19	26	17	17	83	11	32	626
E	131	20	16	22	52	10	12	97	38	180	55	14	3	32	54	15	25	92	21	33	922
F	24	3	10	9	12	7	6	20	27	41	9	11	3	15	12	12	12	26	8	10	277
G	28	6	6	15	11	10	2	31	17	49	13	8	5	10	21	8	11	35	4	12	302
H	14	1	6	9	13	3	0	19	9	34	5	4	1	6	10	4	6	18	0	3	165
I	37	3	10	14	11	11	2	23	16	38	5	12	3	20	26	11	22	23	4	8	299
K	36	5	12	16	25	9	6	31	9	50	10	7	1	17	17	16	16	52	4	21	360
L	66	6	26	21	42	17	10	35	35	88	17	15	9	26	40	26	21	44	8	23	575
M	18	1	9	11	5	3	0	15	12	20	5	2	0	6	14	7	1	12	0	7	148
N	30	5	4	5	14	1	5	17	13	27	9	5	0	2	16	4	10	25	8	11	211
P	8	0	5	4	6	3	1	10	5	23	8	1	1	1	5	3	7	7	4	7	109
Q	59	3	4	13	24	8	7	47	27	82	12	13	1	21	37	11	15	47	12	18	461
R	33	4	16	13	22	8	8	34	12	51	5	10	1	7	19	8	9	29	2	15	306
S	45	10	12	9	26	7	7	49	14	57	7	10	5	18	18	17	15	42	4	28	400
T	52	8	7	14	18	9	7	35	23	70	9	8	7	15	32	12	17	40	11	9	403
V	41	5	16	27	21	9	5	31	26	49	9	13	7	28	46	14	18	29	7	17	418
W	8	1	6	6	6	1	6	7	8	15	1	2	0	5	9	3	7	7	1	4	103
Y	24	5	5	12	8	7	6	20	10	30	5	7	4	9	17	10	6	11	5	13	214
Total	826	101	212	269	397	152	120	648	382	1150	229	173	61	294	480	223	263	691	136	304	7111

さらに、N3N4 ポジションでのジペプチドの傾向指数を算出した (Table 7-23)。

Table 7-23. α ヘリックス構造のN3N4ポジションでのジペプチドの傾向指数

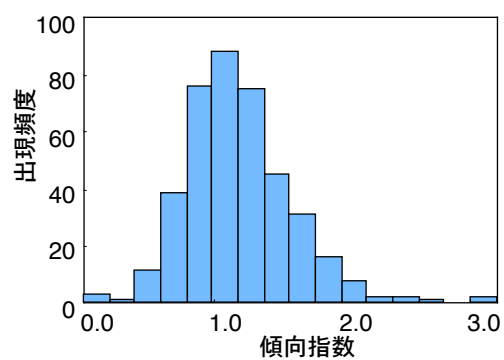
	N4	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
N3																					
A		1.6	0.7	0.6	0.8	1.4	0.4	1.3	1.7	1.3	2.4	1.3	0.7	0.4	1.4	1.8	0.7	0.7	1.6	2.2	1.5
C		0.6	1.1	1.5	0.4	0.8	0.5	0.3	0.2	1.6	1.0	1.1	0.9	0.0	0.7	1.2	0.3	0.9	0.5	2.8	1.2
D		2.1	1.5	0.7	0.6	2.4	0.2	1.4	2.4	1.2	2.8	2.8	0.5	0.1	1.4	1.4	0.7	0.8	2.7	1.6	1.9
E		3.3	4.1	0.6	0.6	2.9	0.3	1.3	3.5	1.2	4.0	4.5	0.7	0.2	1.7	2.3	0.7	1.0	2.7	2.9	2.2
F		1.1	0.7	0.5	0.5	1.2	0.3	0.9	1.3	1.5	1.7	1.5	0.8	0.2	1.5	0.9	0.6	0.7	1.4	2.0	1.0
G		0.7	0.8	0.2	0.5	0.5	0.2	0.2	0.9	0.5	1.1	1.1	0.4	0.3	0.5	0.8	0.2	0.3	0.9	0.6	0.6
H		1.2	0.3	0.7	1.0	1.6	0.2	0.0	1.8	1.2	2.2	1.4	0.6	0.1	1.0	1.3	0.4	0.7	1.7	0.0	0.5
I		1.1	0.6	0.4	0.5	0.8	0.4	0.2	1.1	0.6	1.2	0.7	0.6	0.1	1.3	1.3	0.4	0.8	0.8	0.9	0.6
K		1.0	1.2	0.5	0.5	1.5	0.3	0.7	1.3	0.3	1.3	1.0	0.4	0.1	1.1	0.9	0.7	0.7	1.7	0.8	1.4
L		1.2	0.7	0.7	0.5	2.0	0.4	0.7	1.2	0.8	1.6	1.2	0.5	0.3	1.0	1.1	0.7	0.6	1.1	1.2	1.1
M		1.3	0.5	0.9	1.1	0.8	0.3	0.0	1.6	1.1	1.5	1.3	0.3	0.0	1.0	1.7	0.9	0.1	1.1	0.0	1.4
N		1.2	1.1	0.2	0.3	1.1	0.0	0.7	0.8	0.7	1.0	1.3	0.3	0.0	0.2	1.2	0.2	0.6	1.1	1.6	0.9
P		0.3	0.0	0.2	0.1	0.4	0.1	0.1	0.6	0.3	0.8	1.2	0.1	0.1	0.1	0.3	0.1	0.4	0.3	0.9	0.6
Q		2.2	0.9	0.3	0.7	2.1	0.4	1.0	2.9	1.7	2.9	2.1	1.1	0.1	1.6	2.7	0.8	1.1	2.6	3.0	1.8
R		1.2	0.8	0.7	0.5	1.3	0.3	1.0	1.6	0.7	1.5	0.8	0.7	0.1	0.5	1.1	0.4	0.5	1.1	0.4	1.1
S		1.4	1.6	0.5	0.4	1.4	0.2	0.7	2.1	0.6	1.5	0.8	0.6	0.3	1.1	0.9	0.6	0.7	1.5	0.6	1.8
T		1.7	1.3	0.3	0.6	1.1	0.3	0.7	1.5	1.1	1.8	1.2	0.5	0.3	1.1	1.8	0.5	0.7	1.4	1.6	0.6
V		0.9	0.7	0.5	0.8	1.1	0.3	0.4	1.1	0.9	1.2	0.9	0.6	0.3	1.6	1.7	0.5	0.6	0.8	1.1	1.1
W		1.2	0.7	1.0	0.9	1.5	0.1	2.1	1.4	1.3	1.7	0.5	0.4	0.0	1.0	1.6	0.5	1.2	1.1	0.8	1.2
Y		1.3	1.4	0.3	0.8	0.8	0.4	1.1	1.6	0.7	1.4	1.0	0.5	0.3	0.8	1.2	0.6	0.4	0.7	1.3	1.2

その結果、0.8 と 1.2 と 2.2 と 3.0 に偏りが見られた (Figure 7-5(f))。最大の傾向指数を示したジペプチドは、EM だった。このジペプチドは、アミノ酸残基の傾向指数において、N3 と N4 ポジションでそれぞれ 2.0 と 1.8 のように、1.5 以上の傾向指数同士だった。最小の傾向指数を示したジペプチドは、CP, HH, HW, MH, MP, MW, NG, NP, PC, WP の 10 種類だった。これら 10 種類は、N3N4 ポジションに現れなかったジペプチドだった。

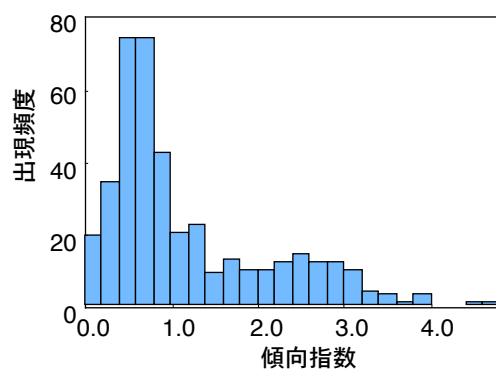
また、N3 ポジションにグルタミン酸やグルタミンが現れたときのほとんどのジペプチドの傾向指数は、1.5 以上だった。これは、N capping box 形成に関わっているためといえる。

ところで、アミノ酸残基の傾向指数が 1.0 以上のアミノ酸残基の組み合わせにおいて、ジペプチドの傾向指数が 0.5 以下のジペプチドは、HC, HW, HY, WM の 4 種類だった。一方で、アミノ酸残基の傾向指数が 1.0 以下のアミノ酸残基の組み合わせにおいて、ジペプチドの傾向指数が 1.5 以上のジペプチドは、CD, CK, FK, SC, WH の 4 種類だった。

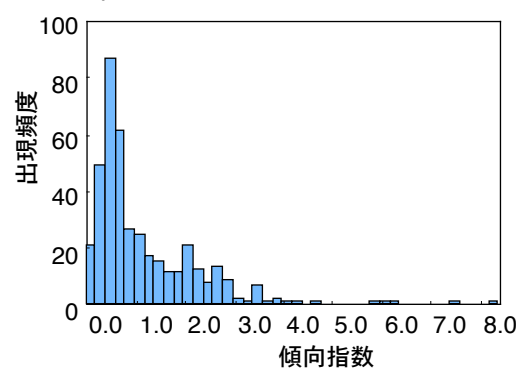
(a) N"N'ポジション



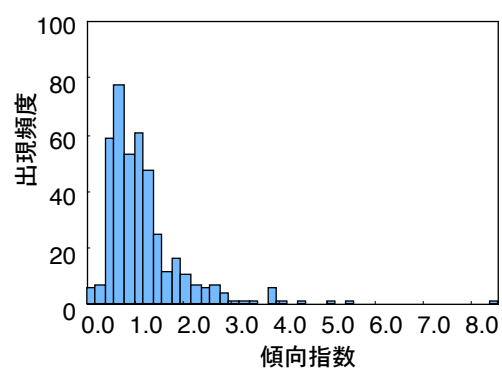
(b) N'Ncapポジション



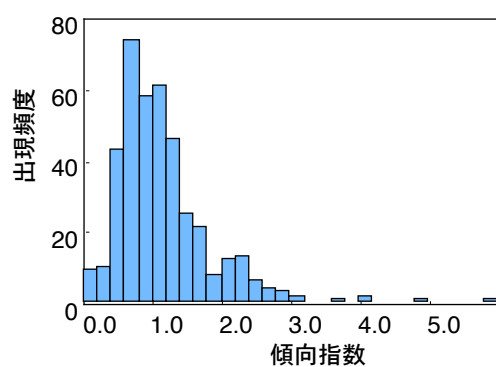
(c) NcapN1ポジション



(d) N1N2ポジション



(e) N2N3ポジション



(f) N3N4ポジション

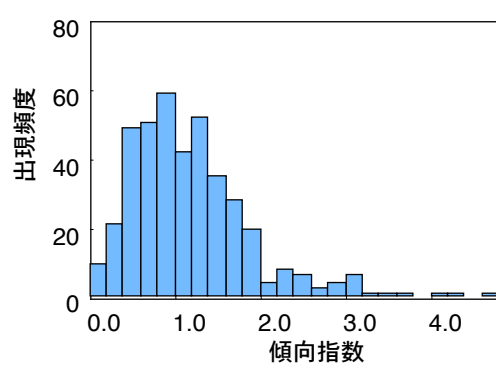


Figure 7-5. α ヘリックス構造のN末端領域でのジペプチドの傾向指数の分布

つぎに、C'C''ポジションでのジペプチドの出現頻度を調べた (Table 7-24)。

Table 7-24. α ヘリックス構造のC'C''ポジションでのジペプチドの出現頻度

C'' \ C'	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	28	2	66	34	8	34	14	17	40	20	2	17	42	20	17	28	28	13	5	11	446
C	5	2	6	9	0	5	5	2	8	2	1	2	4	2	3	4	3	4	1	4	72
D	29	3	29	38	23	18	11	32	32	36	10	18	46	14	23	19	19	24	11	16	451
E	12	3	37	32	18	35	8	12	37	20	12	14	21	11	18	20	15	12	4	14	355
F	12	0	23	16	4	25	6	11	22	9	3	15	23	10	10	15	24	9	2	8	247
G	44	10	37	18	15	51	13	44	45	37	11	21	20	26	34	26	35	22	5	18	532
H	9	0	12	8	3	6	3	4	4	8	0	4	14	2	6	7	10	7	2	7	116
I	23	1	40	21	5	18	8	12	25	21	4	19	51	10	22	22	22	14	2	12	352
K	22	4	41	37	20	28	12	22	36	32	7	29	52	12	20	30	18	17	3	16	458
L	33	4	53	25	14	20	7	18	39	25	6	25	65	20	27	40	29	28	5	10	493
M	4	0	11	4	2	13	3	6	4	4	3	9	12	5	10	6	9	4	0	1	110
N	19	4	19	12	13	24	6	24	18	23	8	12	37	8	12	18	12	23	4	5	301
P	71	5	106	104	23	92	33	27	46	58	7	55	31	53	48	66	52	50	10	25	962
Q	13	3	13	13	3	22	7	13	12	21	3	3	23	7	4	12	9	10	3	8	202
R	13	4	24	20	10	30	13	12	15	24	4	20	27	8	7	27	15	11	3	16	303
S	24	2	43	29	13	38	10	10	15	25	4	16	29	10	22	29	21	13	3	9	365
T	18	7	20	23	12	31	11	22	20	40	5	12	38	14	11	21	16	18	3	14	356
V	28	2	33	23	5	23	13	13	29	19	2	14	36	10	19	23	30	16	2	5	345
W	1	2	5	5	2	10	4	1	3	7	1	2	6	1	5	2	5	3	3	1	69
Y	17	1	27	16	8	12	5	10	11	16	4	17	11	9	11	11	15	9	0	1	211
Total	425	59	645	487	201	535	192	312	461	447	97	324	588	252	329	426	387	307	71	201	6746

さらに、C'C''ポジションでのジペプチドの傾向指数を算出した (Table 7-25)。

Table 7-25. α ヘリックス構造のC'C''ポジションでのジペプチドの傾向指数

C'' \ C'	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.5	0.3	2.1	0.9	0.4	0.8	1.1	0.5	1.2	0.4	0.1	0.7	1.9	0.9	0.6	0.9	0.9	0.3	0.6	0.6
C	0.7	1.2	1.2	1.9	0.0	0.7	1.7	0.4	1.9	0.3	0.6	0.5	0.8	0.5	0.6	0.7	0.6	0.7	1.0	1.3
D	0.8	0.6	1.5	1.4	1.2	0.5	1.3	1.2	1.5	1.0	1.2	1.1	2.1	1.1	1.3	0.9	0.9	0.8	1.7	1.0
E	0.3	0.6	1.5	1.0	1.1	1.2	0.9	0.5	1.2	0.5	1.0	0.7	1.4	0.6	0.8	0.9	0.6	0.4	0.6	1.0
F	0.6	0.0	1.3	0.9	0.4	1.3	0.9	0.8	1.3	0.4	0.5	1.2	1.8	1.0	0.8	0.9	1.5	0.5	0.5	0.8
G	1.1	1.4	1.3	0.6	0.7	1.3	1.1	1.4	1.3	0.9	1.0	1.0	1.1	1.3	1.4	0.8	1.1	0.6	0.8	0.9
H	0.8	0.0	1.4	0.9	0.4	0.5	0.7	0.4	0.6	0.5	0.0	0.7	1.4	0.3	0.8	0.8	1.2	0.7	0.8	1.2
I	0.7	0.2	1.5	0.8	0.4	0.6	0.9	0.6	1.0	0.7	0.6	1.0	2.6	0.7	1.1	0.9	0.9	0.5	0.5	1.0
K	0.7	1.0	1.7	1.3	1.3	1.1	1.5	0.9	1.4	0.9	0.8	1.6	2.9	0.8	1.1	1.3	0.8	0.6	0.6	1.1
L	0.6	0.5	1.5	0.6	0.7	0.5	0.5	0.6	1.0	0.5	0.5	0.9	2.0	0.8	0.8	1.1	0.8	0.8	0.8	0.5
M	0.3	0.0	1.2	0.4	0.3	1.2	1.0	0.7	0.4	0.3	0.8	1.2	1.5	0.9	1.3	0.8	1.1	0.4	0.0	0.2
N	0.8	1.0	1.1	0.7	1.1	1.0	0.9	1.3	1.1	0.9	1.2	0.9	2.0	0.7	0.9	1.1	0.7	1.0	0.9	0.4
P	2.8	1.2	4.9	4.0	1.7	3.5	4.7	1.6	3.0	2.3	1.1	3.8	2.5	4.2	3.5	3.3	3.1	2.1	2.3	2.2
Q	0.5	0.9	0.9	0.7	0.3	1.2	1.1	0.9	0.8	0.8	0.5	0.3	2.1	0.6	0.3	0.9	0.7	0.6	0.8	0.8
R	0.5	0.8	1.1	0.8	0.6	1.3	1.6	0.6	0.9	0.7	0.6	1.4	2.0	0.6	0.4	1.5	0.9	0.4	0.6	1.2
S	0.8	0.3	1.9	1.3	0.7	1.1	1.0	0.4	0.7	0.7	0.5	1.0	1.6	0.6	1.2	1.1	1.0	0.5	0.5	0.6
T	0.6	1.2	0.9	1.0	0.7	1.0	1.2	1.0	1.1	1.1	0.7	0.8	1.8	1.0	0.7	1.0	0.7	0.6	0.5	1.0
V	0.7	0.3	1.1	0.8	0.3	0.8	1.2	0.5	1.0	0.5	0.2	0.7	1.6	0.6	0.8	0.8	1.0	0.4	0.3	0.3
W	0.2	1.5	0.8	0.8	0.5	1.5	1.5	0.2	0.5	0.8	0.5	0.4	1.8	0.2	0.9	0.4	0.9	0.5	2.6	0.3
Y	1.0	0.3	1.7	1.1	0.9	0.6	1.0	0.8	0.8	0.8	0.8	1.4	0.9	0.9	0.8	0.7	1.1	0.6	0.0	0.1

その結果、ジペプチドの傾向指数は、1.0 へ収束する傾向が見られた (Figure 7-6(a))。ただし、N''N'ポジションのときの収束の仕方とは異なる形だった。最大の傾向指数を示したジペプチドは、PD だった。これは、ジペプチド PD の N'ポジションのプロリンが、アミノ酸残基の傾向指数で 1.6、N''ポジションのグルタミン酸が、1.6 の組み合わせだからだと考えた。最小の傾向指数を示したジペプチドは、CF, FC, HC, HM, MC, MW, YW だった。これらは、C'C''ポジションでのジペプチドの出現頻度が 0 だったためである。

C'とC''のそれぞれのポジションでのアミノ酸残基の傾向指数が1.0 以上のアミノ酸残基の組み合わせで、ジペプチドの傾向指数が0.5 以下のジペプチドは、DG だった。一方、アミノ酸残基の傾向指数が 1.0 以下のアミノ酸残基の組み合わせで、ジペプチドの傾向指数が 1.5 以上のジペプチドは、FT, WC, WW だった。

つぎに、CcapC'ポジションでのジペプチドの出現頻度を調べた (Table 7-26)。続いて、CcapC'ポジションでのジペプチドの傾向指数を算出した (Table 7-27)。

Table 7-26. α ヘリックス構造のCcapC'ポジションでのジペプチドの出現頻度

C' \ Ccap	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	19	1	39	26	9	36	7	8	25	19	4	20	44	12	15	18	23	12	3	12	352
C	8	3	6	4	1	4	4	6	4	2	1	2	31	5	4	7	7	5	0	0	104
D	23	3	21	24	8	18	6	16	25	40	3	13	75	6	17	24	9	25	3	7	366
E	5	4	12	15	5	42	4	10	17	13	4	17	34	9	11	11	10	15	3	5	246
F	16	2	20	17	6	19	5	6	23	9	3	12	71	10	15	21	14	7	3	5	284
G	171	27	83	82	129	107	40	184	136	236	46	54	48	53	78	63	94	172	28	108	1939
H	14	0	13	12	14	19	5	6	10	18	1	8	43	9	9	12	6	6	4	7	216
I	10	2	24	11	6	26	7	8	22	3	4	10	61	6	15	12	20	2	1	3	253
K	24	0	25	19	11	32	3	11	21	29	12	14	46	10	11	15	20	18	2	10	333
L	17	1	37	19	4	36	6	14	38	15	2	28	131	14	17	46	33	10	2	4	474
M	4	1	11	9	1	11	1	3	6	4	5	9	23	2	7	11	6	3	0	1	118
N	30	9	20	23	15	40	11	34	27	38	10	23	81	18	27	28	15	25	7	16	497
P	0	0	0	0	1	2	0	1	1	0	0	0	0	0	0	3	0	1	0	1	10
Q	16	5	19	10	4	20	3	8	19	17	3	17	42	5	9	10	13	8	2	4	234
R	16	2	16	15	11	33	5	15	15	24	6	15	41	5	19	6	15	10	3	2	274
S	21	4	45	19	11	27	8	6	15	17	4	15	50	19	18	22	13	9	1	6	330
T	18	4	22	16	3	24	4	4	11	7	2	16	34	9	11	23	21	9	5	8	251
V	17	1	25	15	5	24	1	5	27	6	0	14	54	8	15	22	17	7	0	7	270
W	1	0	2	2	0	5	0	0	7	0	1	2	15	1	3	3	6	4	0	1	53
Y	19	3	14	20	6	18	4	9	21	4	3	18	46	5	11	13	21	4	2	6	247
Total	449	72	454	358	250	543	124	354	470	501	114	307	970	206	312	370	363	352	69	213	6851

Table 7-27. α ヘリックス構造のCcapC'ポジションでのジペプチドの傾向指数

C' \ Ccap	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.3	0.1	1.2	0.7	0.4	0.9	0.5	0.2	0.7	0.4	0.3	0.9	2.0	0.5	0.5	0.6	0.8	0.3	0.4	0.6
C	1.1	1.8	1.1	0.8	0.3	0.5	1.3	1.3	1.0	0.3	0.5	0.4	6.3	1.3	0.8	1.2	1.3	0.9	0.0	0.0
D	0.7	0.6	1.0	0.9	0.4	0.5	0.7	0.6	1.1	1.1	0.4	0.8	3.4	0.5	0.9	1.1	0.4	0.9	0.4	0.4
E	0.1	0.8	0.5	0.5	0.3	1.4	0.5	0.4	0.5	0.3	0.3	0.8	2.2	0.5	0.5	0.5	0.4	0.5	0.4	0.3
F	0.8	0.5	1.1	1.0	0.6	0.9	0.7	0.4	1.4	0.4	0.5	0.9	5.4	1.0	1.2	1.2	0.8	0.4	0.8	0.5
G	4.3	3.6	2.8	2.9	5.9	2.7	3.4	5.8	4.0	5.6	4.2	2.5	2.6	2.7	3.1	2.0	3.0	4.7	4.2	5.3
H	1.2	0.0	1.5	1.3	1.8	1.5	1.1	0.6	1.4	1.2	0.3	1.3	4.3	1.5	1.2	1.4	0.7	0.6	1.5	1.2
I	0.3	0.4	0.9	0.4	0.5	0.9	0.7	0.4	0.9	0.1	0.6	0.5	3.0	0.4	0.8	0.5	0.8	0.1	0.2	0.2
K	0.7	0.0	1.0	0.6	0.7	1.2	0.4	0.5	0.8	0.8	1.3	0.7	2.5	0.7	0.6	0.7	0.9	0.6	0.4	0.7
L	0.3	0.1	1.0	0.5	0.2	0.8	0.4	0.5	0.9	0.3	0.2	1.0	4.0	0.5	0.5	1.2	0.9	0.3	0.3	0.2
M	0.3	0.6	1.2	0.9	0.2	1.0	0.3	0.3	0.6	0.3	1.3	1.2	2.9	0.3	0.9	1.4	0.7	0.3	0.0	0.2
N	1.2	2.1	1.2	1.4	1.2	1.6	1.6	1.8	1.6	1.4	1.5	1.7	4.3	1.6	2.0	1.7	0.9	1.1	1.5	1.4
P	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1
Q	0.6	1.5	1.3	0.5	0.4	1.1	0.5	0.5	1.2	0.6	0.5	1.5	3.8	0.4	0.7	0.7	1.0	0.5	0.5	0.4
R	0.6	0.4	0.7	0.6	0.7	1.4	0.6	0.7	0.9	0.7	0.9	1.0	3.0	0.4	1.1	0.3	0.9	0.4	0.6	0.1
S	0.7	0.6	2.0	0.8	0.6	0.8	0.8	0.3	0.7	0.5	0.5	0.9	2.7	1.2	0.9	0.8	0.6	0.3	0.2	0.4
T	0.6	0.7	1.0	0.7	0.2	0.8	0.4	0.2	0.6	0.2	0.3	1.1	1.6	0.7	0.7	1.1	1.0	0.3	0.8	0.6
V	0.4	0.2	0.8	0.5	0.3	0.8	0.1	0.2	0.9	0.2	0.0	0.6	2.3	0.5	0.6	0.8	0.6	0.2	0.0	0.5
W	0.2	0.0	0.3	0.3	0.0	0.7	0.0	0.0	1.2	0.0	0.5	0.4	4.4	0.2	0.5	0.6	1.1	0.7	0.0	0.3
Y	1.1	0.9	0.9	1.4	0.6	1.0	0.7	0.7	1.6	0.2	0.6	1.4	3.7	0.5	0.8	0.9	1.5	0.3	0.5	0.6

その結果、ジペプチドの傾向指数の分布は、0.6 に収束していたものの、それでも 1.5 以上の傾向指数が存在した (Figure 7-6(b))。最大の傾向指数を示したジペプチドは、CP だった。このジペプチドを構成しているアミノ酸残基の組み合わせは、Ncap と C' のそれぞれのポジションにおいて、最大の傾向指数であるアミノ酸残基同士ではない。つまり、アミノ酸残基による Richardson パラメータからは得られない情報である。最小の傾向指数を示したジペプチドは、CW, CY, HC, KC, MW, PX (X=A, C, D, E, H, L, M, N, P, Q, R, T, V, W), WC, WF, WH, WI, WL, WW だった。これらは、すべて CcapC' ポジションに現れなかったジペプチドである。

また、Ccap ポジションにグリシン、C' ポジションにプロリン残基が関わっているジペプチドの傾向指数は、すべて 2.0 以上だった (ただし、PP=0.0, TP=1.6)。つまり、グリシンが Ccap ポジションに現れたとき、C' ポジションにくるアミノ酸残基の種類はあまり関係ない。そして、C' ポジションにプロリンが現れたとき、Ccap ポジションのアミノ酸残基の種類はあまり関係ないといえる。もちろん、傾向指数が高いジペプチドを比べると、アミノ酸残基の種類によって、ジペプチドの傾向指数に差はあるが、1.5 以上の傾向指数という枠で考えた場合は、関係ないといえる。

それぞれのポジションでのアミノ酸残基の傾向指数が 1.0 以上であるアミノ酸残基の組み合わせにおいて、ジペプチドの傾向指数が 0.5 以下のジペプチドは、CG と CN だった。一方、1.0 以下のアミノ酸残基の組み合わせにおいて、ジペプチドの傾向指数が 1.5 以上のものは、QC, QN, YT だった。

つぎに、C1Ccap ポジションでのジペプチドの出現頻度を調べた (Table 7-28)。続いて、C1Ccap ポジションでのジペプチドの傾向指数を算出した (Table 7-29)。

Table 7-28. α ヘリックス構造のC1Ccapポジションでのジペプチドの出現頻度

Ccap \ C1	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	42	6	34	28	28	261	20	19	32	55	11	43	1	27	27	33	22	23	3	16	731
C	6	4	5	3	4	32	10	3	3	9	1	6	0	9	2	6	4	3	2	0	112
D	11	6	15	8	8	60	4	18	14	17	3	29	0	5	12	16	10	13	3	12	264
E	21	9	27	10	10	118	14	22	33	28	10	35	0	13	26	27	19	21	3	16	462
F	18	3	19	16	10	94	6	16	17	19	2	20	0	13	9	21	8	16	2	9	318
G	7	2	12	6	3	37	6	5	16	24	8	18	0	7	9	16	8	9	2	6	201
H	10	4	8	7	9	59	10	9	8	22	8	12	0	12	13	11	10	12	1	9	234
I	20	9	17	7	14	81	6	7	12	18	3	17	3	10	10	19	8	11	0	9	281
K	29	8	30	17	26	110	21	20	21	49	11	46	0	21	17	23	26	24	5	35	539
L	52	13	56	52	22	314	26	29	55	50	9	73	1	30	44	44	31	27	2	25	955
M	6	2	6	7	12	63	5	9	12	13	2	16	0	4	10	7	10	7	0	8	199
N	27	5	26	17	22	82	15	14	25	23	10	30	0	9	18	24	20	11	9	20	407
P	0	0	0	0	5	2	0	0	1	4	0	1	0	0	0	1	0	2	0	0	16
Q	20	5	18	13	26	91	18	9	19	26	7	22	2	17	15	14	11	13	2	12	360
R	22	4	26	13	22	108	17	24	18	31	5	36	0	20	10	15	11	12	4	25	423
S	22	12	20	14	19	114	16	23	24	33	10	31	1	22	20	17	22	24	2	13	459
T	14	8	24	24	25	143	15	15	15	26	8	27	1	16	21	21	14	16	6	18	457
V	23	7	11	10	17	68	9	16	10	23	9	18	1	9	16	12	18	17	4	13	311
W	9	3	4	5	3	21	1	2	5	7	2	5	0	2	1	3	3	2	0	2	80
Y	12	1	14	6	15	99	3	8	21	23	5	21	0	12	14	17	10	16	3	7	307
Total	371	111	372	263	300	1957	222	268	361	500	124	506	10	258	294	347	265	279	53	255	7116

この結果から、ジペプチドの傾向指数の分布は、0.0 と 0.8 に集まっていた。それでもなお、1.5 以上の傾向指数が存在した (Figure 7-6(c))。最大の傾向指数を示したジペプチドは、LG だった。この組み合わせは、それぞれのポジションでのアミノ酸残基の傾向指数で最大の値をもつもの同士である。最小の傾向指数は、CY, IW, MW, WW, そして、プロリン残基が関わっているジペプチドのほとんどだった。

また、Ccap ポジションにグリシンが関わるジペプチドの傾向指数は、すべて 2.0 以上だった (ただし、GG=0.9, PG=0.1)。つまり、C1Ccap ポジションにおいて、Ccap ポジションにグリシン残基が関わる場合、C1 ポジションのアミノ酸残基の種類はあまり関係ないといえる。

Table 7-29. α ヘリックス構造のC1Ccapポジションでのジペプチドの傾向指数

Ccap \ C1	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.7	0.8	1.0	0.7	1.3	6.0	1.5	0.6	0.9	1.0	0.8	1.8	0.0	1.1	0.9	1.0	0.7	0.5	0.3	0.8
C	0.8	2.3	0.9	0.6	1.0	4.1	3.2	0.6	0.7	1.1	0.5	1.3	0.0	2.2	0.4	1.0	0.7	0.5	1.9	0.0
D	0.3	1.2	0.7	0.3	0.4	1.7	0.5	0.6	0.6	0.4	0.4	1.7	0.0	0.4	0.6	0.7	0.4	0.4	0.4	0.7
E	0.5	1.8	1.0	0.3	0.6	3.9	1.6	0.8	1.0	0.6	0.8	1.6	0.0	0.7	1.1	1.2	0.8	0.6	0.4	1.1
F	0.9	0.7	1.0	0.9	1.0	4.5	0.9	1.1	1.0	0.8	0.3	1.5	0.0	1.3	0.7	1.1	0.5	0.8	0.5	0.9
G	0.2	0.3	0.4	0.2	0.1	0.9	0.5	0.2	0.5	0.5	0.7	0.8	0.0	0.3	0.3	0.5	0.2	0.2	0.3	0.3
H	0.8	1.4	0.9	0.8	1.1	4.4	2.1	0.9	1.1	1.4	2.2	1.9	0.0	2.0	1.7	1.2	1.2	1.1	0.4	1.4
I	0.6	1.7	0.6	0.3	1.1	2.7	0.6	0.3	0.5	0.6	0.4	0.8	0.1	0.7	0.5	0.7	0.3	0.4	0.0	0.7
K	0.8	1.8	1.1	0.6	1.6	4.0	2.5	0.8	0.8	1.3	1.1	2.4	0.0	1.3	0.9	1.0	1.2	0.8	0.9	2.3
L	0.9	1.6	1.5	1.2	1.1	7.1	1.7	1.0	1.3	0.9	0.7	2.4	0.0	1.1	1.2	1.1	0.9	0.7	0.3	1.2
M	0.4	1.1	0.6	0.7	2.0	5.5	1.6	1.0	1.1	1.0	0.5	2.1	0.0	0.7	1.2	0.9	1.1	0.7	0.0	1.7
N	1.1	1.1	1.5	1.0	1.7	3.1	2.1	0.7	1.4	0.8	1.4	2.1	0.0	0.8	1.3	1.4	1.2	0.5	1.8	1.7
P	0.0	0.0	0.0	0.0	0.4	0.1	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Q	0.8	1.4	1.2	0.7	2.3	4.7	2.7	0.6	1.2	0.9	1.2	1.9	0.2	1.3	1.1	1.0	0.8	0.7	0.5	1.2
R	0.8	0.8	1.2	0.5	1.3	4.3	2.0	1.1	1.0	0.9	0.8	2.4	0.0	1.4	0.6	0.8	0.6	0.4	0.8	1.8
S	0.7	1.9	0.9	0.6	1.0	3.2	1.5	1.0	1.1	0.9	1.2	1.8	0.1	1.3	1.0	0.6	1.0	0.9	0.3	0.8
T	0.5	1.3	1.0	1.0	1.5	4.3	1.6	0.6	0.7	0.7	1.1	1.7	0.0	1.1	1.2	0.9	0.6	0.5	0.9	1.2
V	0.5	1.0	0.3	0.3	0.9	2.1	0.8	0.5	0.3	0.6	0.9	0.8	0.0	0.5	0.6	0.4	0.6	0.5	0.6	0.8
W	1.3	2.1	0.6	0.8	0.8	2.9	0.4	0.4	0.8	0.8	0.9	1.1	0.0	0.4	0.2	0.5	0.5	0.3	0.0	0.6
Y	0.6	0.3	0.8	0.4	1.5	5.0	0.5	0.6	1.5	1.1	1.0	1.6	0.0	1.1	1.0	1.1	0.7	1.0	0.8	0.7

C1 と Ccap のそれぞれのポジションで 1.0 以上の傾向指数をもつアミノ酸残基の組み合わせで、ジペプチドの傾向指数が 0.5 以下のものは、CY, YC, YH だった。一方、それぞれのポジションで 1.0 以下の傾向指数をもつアミノ酸残基の組み合わせで、ジペプチドが 1.5 以上のものは、存在しなかった。

つぎに、C2C1 ポジションでのジペプチドの出現頻度を調べた (Table 7-30)。続いて、C2C1 ポジションでのジペプチドの傾向指数を算出した (Table 7-31)。その結果、ジペプチドの傾向指数の分布は、0.8, 1.6 に集まっていた (Figure 7-6(d))。最大の傾向指数を示したジペプチドは、KH だった。このアミノ酸残基の組み合わせは、それぞれのポジションにおいて、最大の傾向指数をもつアミノ酸残基ではなかった。しかしながら、C1 ポジションでのアミノ酸残基の傾向指数において、ヒスチジンは 1.4 であり、最大を示すロイシンの 1.5 とさほど変わらない。したがって、アミノ酸残基の傾向指数が高いもの同士の組み合わせがジペプチドの傾向指数で最大となったと考えた。最小の傾向指数を示したジペプチドは、HW, MC, MW, PS, WD, XP (X は、ほとんどのアミノ酸残基) だった。これらは、C2C1 ポジションに現れなかったジペ

プチドだった。

Table 7-30. α ヘリックス構造のC2C1ポジションでのジペプチドの出現頻度

C1 \ C2	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	121	12	30	48	23	34	20	27	63	118	19	46	3	48	58	47	63	30	5	24	839
C	5	2	3	2	5	3	7	3	5	12	1	5	0	6	4	3	7	1	1	4	79
D	51	5	13	20	13	8	10	20	32	56	14	28	1	25	28	40	19	19	6	26	434
E	70	7	23	40	30	14	19	38	85	89	27	34	1	42	49	48	43	43	10	20	732
F	14	8	10	7	9	11	7	8	16	23	6	10	0	8	7	14	12	6	7	9	192
G	13	2	6	10	9	12	3	8	11	28	9	8	1	8	15	16	10	6	2	4	181
H	15	5	4	12	11	1	1	8	7	32	3	4	0	5	6	12	7	8	0	7	148
I	27	8	9	18	18	11	7	13	17	48	3	15	2	11	16	23	24	16	5	15	306
K	70	7	40	67	38	12	40	34	53	89	18	57	0	39	36	41	37	37	6	53	774
L	87	13	25	42	29	25	16	15	60	104	20	45	3	27	37	56	68	30	5	15	722
M	23	0	5	12	7	3	3	9	12	22	8	12	1	11	6	14	16	6	0	9	179
N	36	6	10	24	5	9	12	9	14	32	7	21	0	21	13	19	19	15	5	14	291
P	3	1	2	6	9	2	2	3	9	12	4	5	0	7	6	1	1	7	1	8	89
Q	39	8	9	29	14	9	19	13	28	41	12	27	1	21	30	18	17	15	4	18	372
R	59	4	35	45	30	13	22	23	26	58	16	29	0	26	31	28	35	24	4	24	532
S	33	7	20	31	19	15	18	26	46	64	14	15	0	27	31	29	17	19	6	14	451
T	26	6	10	16	17	3	10	9	21	46	5	12	0	11	14	16	14	10	2	15	263
V	20	5	9	16	13	3	11	9	21	50	6	18	2	8	22	18	29	10	3	13	286
W	5	1	0	3	5	2	4	4	5	10	4	4	0	1	5	5	9	5	1	1	74
Y	13	5	1	15	14	11	3	4	11	18	3	11	1	9	10	12	10	5	7	13	176
Total	730	112	264	463	318	201	234	283	542	952	199	406	16	361	424	460	457	312	80	306	7120

Table 7-31. α ヘリックス構造のC2C1ポジションでのジペプチドの傾向指数

C1 \ C2	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	2.1	1.6	0.9	1.2	1.1	0.8	1.5	0.8	1.8	2.1	1.3	1.9	0.1	2.0	1.8	1.5	2.0	0.7	0.6	1.2
C	0.7	1.1	0.5	0.4	1.3	0.4	2.2	0.6	1.1	1.5	0.5	1.1	0.0	1.4	0.8	0.5	1.2	0.2	0.9	1.2
D	1.4	1.0	0.6	0.7	0.6	0.2	1.2	0.7	1.4	1.5	1.6	1.6	0.0	1.8	1.5	1.7	0.9	0.6	0.9	1.5
E	1.8	1.4	0.9	1.2	1.7	0.5	2.1	1.4	2.6	2.0	2.2	1.6	0.1	2.3	2.1	2.1	1.8	1.3	1.4	1.3
F	0.7	1.9	0.5	0.4	0.9	0.5	1.0	0.5	0.9	0.9	1.0	0.7	0.0	0.8	0.5	0.8	0.7	0.3	1.8	0.9
G	0.3	0.3	0.2	0.3	0.4	0.3	0.2	0.2	0.3	0.6	0.8	0.4	0.1	0.4	0.6	0.5	0.3	0.2	0.3	0.2
H	1.2	1.7	0.4	1.3	1.3	0.1	0.2	0.8	0.9	2.1	0.8	0.6	0.0	0.8	0.8	1.3	0.8	0.7	0.0	1.1
I	0.8	1.5	0.3	0.7	1.4	0.4	0.7	0.6	0.7	1.5	0.4	0.7	0.1	0.7	0.8	0.9	0.9	0.6	1.2	1.1
K	2.0	1.6	1.5	2.2	2.4	0.4	4.7	1.4	2.0	2.3	1.8	2.9	0.0	2.5	1.9	1.7	1.6	1.2	1.1	3.6
L	1.6	1.6	0.7	1.0	1.4	0.6	1.1	0.5	1.4	1.9	1.5	1.5	0.1	1.0	1.0	1.4	1.9	0.8	0.7	0.7
M	1.7	0.0	0.5	1.2	1.1	0.3	0.9	1.0	1.1	1.7	2.0	1.5	0.1	1.8	0.7	1.8	1.8	0.6	0.0	1.9
N	1.4	1.4	0.6	1.4	0.4	0.3	1.7	0.4	0.8	1.1	1.0	1.5	0.0	1.8	0.9	1.1	1.1	0.6	1.0	1.2
P	0.1	0.2	0.1	0.2	0.6	0.1	0.3	0.2	0.6	0.4	0.6	0.3	0.0	0.5	0.4	0.0	0.1	0.3	0.2	0.7
Q	1.5	2.3	0.6	1.5	1.3	0.5	2.8	0.8	1.7	1.5	2.1	2.3	0.1	1.6	2.2	1.3	1.3	0.8	1.0	1.8
R	2.1	0.8	1.6	1.8	1.8	0.5	2.6	1.1	1.5	1.7	2.4	1.9	0.0	1.9	1.7	1.5	2.0	0.9	0.8	1.7
S	1.0	1.1	0.9	1.3	1.0	0.4	1.7	1.1	2.1	1.7	1.7	0.9	0.0	1.6	1.6	1.0	0.7	0.7	0.9	0.9
T	0.8	1.0	0.4	0.7	1.0	0.1	1.1	0.4	1.0	1.2	0.7	0.8	0.0	0.8	0.8	0.7	0.6	0.3	0.3	1.0
V	0.5	0.7	0.3	0.5	0.7	0.1	0.9	0.3	0.7	1.2	0.6	0.8	0.1	0.5	0.8	0.6	1.0	0.3	0.5	0.8
W	0.7	0.7	0.0	0.5	1.3	0.3	1.4	0.8	0.8	1.1	1.9	0.8	0.0	0.2	0.9	0.9	1.5	0.8	0.8	0.3
Y	0.7	1.4	0.1	1.0	1.4	0.6	0.5	0.3	0.8	0.8	0.6	0.8	0.1	0.8	0.7	0.8	0.7	0.3	1.8	1.2

C2 と C1 のそれぞれのポジションで、1.0 以上の傾向指数をもつアミノ酸残基を組み合わせ

たジペプチドの傾向指数が、0.5 以下になるものは、MC だった。一方で、1.0 以下の傾向指数をもつアミノ酸残基を組み合わせたジペプチドで、傾向指数が 1.5 異常になるものは、YW だった。

つぎに、C3C2 ポジションでのジペプチドの出現頻度を調べた (Table 7-32)。続いて、C3C2 ポジションでのジペプチドの傾向指数を算出した (Table 7-33)。その結果、ジペプチドの傾向指数の分布は、0.6 や 1.4 に集まる傾向にあった (Figure 7-6(e))。最大の傾向指数を示したジペプチドは、KK, QK, RE だった。これらのジペプチドにおいて、C3 ポジションのリジン、グルタミン、アルギニンは、アミノ酸残基の傾向指数において、最大値ではないが、1.2 以上の値を有する。したがって、配列情報を含めたジペプチドでの解析では、必ずしも、アミノ酸残基の傾向指数が最大のものの同士の組み合わせが、ジペプチドでも最大になるとは限らない。最小の傾向指数を示したジペプチドは、CP, CW, HW, NW, PP, PW, TW, WC, WG だった。これらは、C3C2 ポジションで現れなかったジペプチドだった。

Table 7-32. α ヘリックス構造のC3C2ポジションでのジペプチドの出現頻度

	C2	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
C3																						
A	122	9	39	86	9	24	18	30	103	91	30	35	9	47	81	44	22	30	14	13	856	
C	10	3	5	7	5	3	1	3	7	12	1	2	0	3	8	7	1	5	0	1	84	
D	38	3	16	30	8	6	6	25	37	30	9	9	4	10	18	14	15	12	5	11	306	
E	50	4	31	61	16	11	7	33	72	59	9	26	1	28	40	24	33	23	10	14	552	
F	36	3	21	23	12	9	9	18	25	36	7	15	4	6	25	14	7	13	6	7	296	
G	18	4	11	12	4	3	5	8	20	20	2	2	5	9	14	10	7	11	5	8	178	
H	19	2	8	17	7	1	2	10	15	13	4	4	2	4	11	14	12	10	0	3	158	
I	61	5	40	45	10	14	7	12	41	47	11	27	6	23	27	31	16	14	1	9	447	
K	81	4	38	75	12	9	12	20	82	36	17	18	9	37	41	45	12	20	8	14	590	
L	124	11	64	91	25	32	15	37	110	98	32	42	14	46	73	73	36	34	7	25	989	
M	22	3	12	16	7	5	4	13	22	21	4	10	4	5	14	15	8	12	1	5	203	
N	28	1	10	26	7	4	3	16	30	25	5	3	2	15	16	16	13	13	0	5	238	
P	2	1	3	4	4	4	4	4	4	6	11	2	1	0	4	3	2	2	2	0	1	60
Q	33	4	14	30	4	5	9	11	48	34	10	13	3	29	28	14	16	12	4	11	332	
R	59	5	35	75	19	16	18	18	37	40	9	19	9	27	35	30	14	11	4	12	492	
S	22	4	23	24	10	12	10	13	32	47	8	17	4	14	29	26	17	14	5	8	339	
T	30	3	19	31	5	4	8	9	14	32	4	6	3	19	11	22	7	14	0	7	248	
V	57	7	27	50	17	10	5	13	41	24	7	15	6	18	30	29	12	22	1	9	400	
W	7	0	6	9	4	0	2	3	8	14	1	8	1	10	12	8	5	3	1	7	109	
Y	19	3	12	20	7	9	3	10	23	31	7	19	3	16	14	12	8	10	2	6	234	
Total	838	79	434	732	192	181	148	306	773	721	179	291	89	370	530	450	263	285	74	176	7111	

Table 7-33. α ヘリックス構造のC3C2ポジションでのジペプチドの傾向指数

	C2	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
C3																					
A		2.1	1.2	1.2	2.2	0.4	0.6	1.3	0.9	2.9	1.6	2.1	1.5	0.4	2.0	2.6	1.4	0.7	0.7	1.6	0.7
C		1.4	1.7	0.9	1.4	1.3	0.4	0.3	0.6	1.6	1.5	0.5	0.4	0.0	0.7	1.6	1.1	0.2	0.8	0.0	0.3
D		1.1	0.6	0.8	1.1	0.4	0.2	0.7	0.9	1.6	0.8	1.1	0.5	0.2	0.7	0.9	0.6	0.7	0.4	0.7	0.6
E		1.3	0.8	1.2	1.8	0.9	0.4	0.8	1.2	2.2	1.3	0.7	1.2	0.1	1.5	1.7	1.1	1.4	0.7	1.4	0.9
F		1.7	0.7	1.1	1.3	1.2	0.4	1.3	1.2	1.4	1.5	1.2	1.1	0.3	0.6	1.9	0.8	0.4	0.7	1.5	0.7
G		0.4	0.5	0.4	0.4	0.2	0.1	0.4	0.2	0.6	0.5	0.2	0.1	0.3	0.4	0.5	0.3	0.2	0.3	0.7	0.4
H		1.6	0.7	0.9	1.8	0.8	0.1	0.4	1.0	2.0	0.8	1.1	0.6	0.2	0.7	1.4	1.6	1.4	0.9	0.0	0.5
I		1.8	1.0	1.5	1.7	0.8	0.5	0.7	0.5	1.6	1.5	1.5	1.3	0.3	1.5	1.3	1.2	0.6	0.5	0.2	0.7
K		2.3	0.9	1.5	2.4	0.7	0.3	1.4	0.8	3.0	0.9	1.7	0.9	0.5	2.3	2.2	1.9	0.5	0.6	1.5	0.9
L		2.2	1.4	1.7	2.2	1.2	0.7	1.0	1.2	2.6	1.8	2.3	1.4	0.4	1.7	2.0	1.8	1.0	0.9	1.0	1.2
M		1.6	1.6	1.2	1.6	1.1	0.4	1.2	1.4	2.0	1.6	1.0	1.3	0.5	0.8	1.7	1.9	0.9	1.1	0.8	1.0
N		1.1	0.2	0.6	1.5	0.6	0.2	0.4	0.8	1.7	0.9	0.7	0.2	0.1	1.3	1.2	0.9	0.8	0.6	0.0	0.4
P		0.1	0.2	0.1	0.1	0.3	0.1	0.5	0.2	0.4	0.4	0.3	0.1	0.0	0.3	0.2	0.1	0.1	0.1	0.0	0.1
Q		1.2	1.1	0.9	1.6	0.4	0.3	1.3	0.7	3.0	1.2	1.7	1.1	0.3	2.2	2.0	1.0	1.2	0.7	1.0	1.1
R		2.1	1.0	1.6	3.0	1.2	0.6	2.1	0.8	2.1	1.1	1.4	1.2	0.6	1.9	1.9	1.6	0.8	0.4	0.8	0.9
S		0.7	0.6	1.0	1.0	0.5	0.3	0.9	0.6	1.5	1.3	1.0	1.0	0.2	0.8	1.5	0.9	0.8	0.5	0.8	0.5
T		1.0	0.5	0.8	1.3	0.3	0.1	0.8	0.4	0.7	0.8	0.5	0.4	0.1	1.3	0.6	1.0	0.3	0.5	0.0	0.5
V		1.3	1.0	0.8	1.6	0.9	0.3	0.4	0.4	1.3	0.6	0.7	0.7	0.2	1.0	1.1	1.0	0.4	0.6	0.2	0.6
W		1.0	0.0	1.0	1.4	1.0	0.0	0.7	0.6	1.3	1.6	0.5	1.7	0.3	1.9	2.1	1.4	0.9	0.5	0.8	2.0
Y		1.0	0.8	0.7	1.3	0.7	0.5	0.5	0.8	1.7	1.4	1.4	1.5	0.2	1.5	1.0	0.8	0.6	0.6	0.5	0.6

C3 と C2 のそれぞれのポジションにおいて、1.0 以上の傾向指数をもつアミノ酸残基の組み合わせの中で、ジペプチドの傾向指数が0.5 以上のものは、WM だった。一方で、1.0 以下の傾向指数をもつアミノ酸残基の組み合わせにおいて、ジペプチドの傾向指数が1.5 以上のものは、CC と YN だった。

つぎに、C4C3 ポジションでのジペプチドの出現頻度を調べた (Table 7-34)。C4C3 ポジションでのジペプチドの傾向指数を算出した (Table 7-35)。その結果、ジペプチドの傾向指数は、0.4, 1.0, 2.4 に最大値が存在する分布だった (Figure 7-6(f))。最大の傾向指数を示したジペプチドは、WW だった。この組み合わせは、それぞれのポジションでのアミノ酸残基の傾向指数が1.5 以上のもの同士ではない。最小の傾向指数を示したジペプチドは、CC, CW, DP, HP, NP, PC, PP だった。これらは、C4C3 ポジションに現れなかったジペプチドだった。

ところで、C4 と C3 のそれぞれのポジションで、傾向指数が1.0 以上のアミノ酸残基の組み合わせにおいて、ジペプチドの傾向指数が0.5 以下のものは、CW だった。一方、傾向指数が1.0 以下のアミノ酸残基の組み合わせにおいて、ジペプチドの傾向指数が1.5 以上のものは、RH だった。

Table 7-34. α ヘリックス構造のC4C3ポジションでのジペプチドの出現頻度

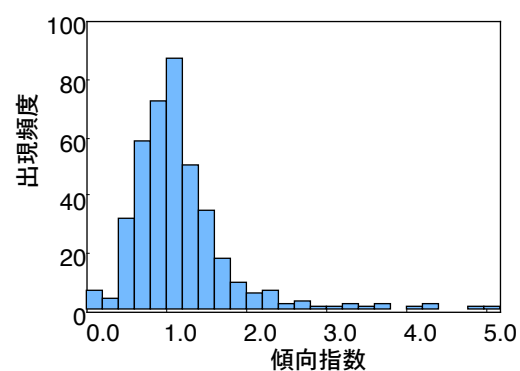
C3 \ C4	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	128	15	38	82	42	25	18	69	86	132	21	28	10	45	76	44	26	66	14	39	1004
C	17	0	4	15	7	2	6	7	8	20	5	3	1	7	5	7	8	11	0	1	134
D	21	5	13	11	24	6	7	21	10	44	12	4	0	8	10	8	7	13	9	9	242
E	58	7	13	35	18	5	9	26	37	70	12	12	3	16	30	15	22	28	6	15	437
F	33	5	11	32	19	6	10	23	34	56	13	17	3	14	24	17	19	22	4	10	372
G	22	3	9	17	9	6	6	14	15	36	4	7	4	10	14	9	4	18	3	5	215
H	15	1	8	11	5	3	2	10	7	19	3	2	0	9	4	7	6	10	4	5	131
I	65	5	18	32	17	15	7	24	52	65	13	23	8	25	29	30	10	22	7	16	483
K	42	3	21	27	14	8	2	22	22	51	12	15	3	11	14	18	11	19	6	12	333
L	152	12	51	103	28	35	27	51	122	126	28	43	9	71	116	65	43	52	12	30	1176
M	31	2	18	18	8	7	8	21	24	33	9	7	2	15	20	8	10	13	2	9	265
N	16	1	8	9	12	5	1	14	12	32	5	3	0	8	8	4	6	9	5	7	165
P	14	0	7	15	9	4	5	7	4	21	2	5	0	5	6	11	5	5	5	3	133
Q	38	4	9	24	8	7	6	15	16	49	7	10	1	13	13	11	11	14	4	9	269
R	27	5	17	30	14	9	13	27	26	42	8	17	2	16	21	13	11	10	8	13	329
S	34	2	10	13	15	10	3	32	18	47	10	6	2	10	20	12	10	20	4	15	293
T	23	4	14	15	7	4	4	15	18	38	7	7	3	8	15	10	8	11	3	9	223
V	77	4	15	40	15	11	10	26	42	46	19	17	5	22	41	25	17	31	3	7	473
W	14	1	10	8	12	4	7	9	16	27	3	3	1	11	8	7	5	10	5	5	166
Y	28	5	11	15	13	6	7	13	21	37	10	9	2	9	17	18	10	15	5	15	266
Total	855	84	305	552	296	178	158	446	590	991	203	238	59	333	491	339	249	399	109	234	7109

Table 7-35. α ヘリックス構造のC4C3ポジションでのジペプチドの傾向指数

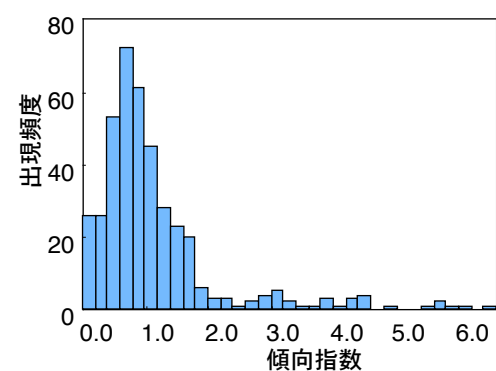
C3 \ C4	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	2.2	2.0	1.1	2.1	1.9	0.6	1.3	2.0	2.5	2.4	1.4	1.2	0.4	1.9	2.4	1.4	0.8	1.6	1.6	2.0
C	2.3	0.0	0.7	2.9	1.8	0.3	1.9	1.4	1.8	2.5	2.6	0.6	0.2	1.7	1.0	1.1	1.4	1.9	0.0	0.3
D	0.6	1.0	0.6	0.4	1.2	0.2	0.8	0.7	0.4	1.2	1.4	0.2	0.0	0.6	0.5	0.3	0.3	0.4	1.3	0.5
E	1.5	1.4	0.5	1.0	1.0	0.2	1.0	0.9	1.1	1.5	1.0	0.6	0.2	0.9	1.3	0.7	0.9	0.8	0.8	1.0
F	1.6	1.2	0.6	1.8	1.9	0.3	1.4	1.5	1.9	2.3	2.2	1.3	0.2	1.4	1.9	0.9	1.1	1.2	1.0	1.0
G	0.5	0.4	0.3	0.6	0.4	0.1	0.5	0.4	0.4	0.8	0.3	0.3	0.2	0.5	0.5	0.3	0.1	0.5	0.4	0.2
H	1.2	0.3	0.9	1.2	0.6	0.2	0.4	1.0	0.9	1.2	0.8	0.3	0.0	1.5	0.5	0.8	0.7	0.9	1.5	0.8
I	1.9	1.0	0.7	1.2	1.3	0.5	0.7	1.1	2.0	2.0	1.8	1.1	0.4	1.6	1.4	1.2	0.4	0.8	1.6	1.2
K	1.2	0.7	0.8	0.9	0.9	0.3	0.2	0.9	0.8	1.3	1.2	0.8	0.2	0.7	0.7	0.8	0.5	0.6	1.1	0.8
L	2.8	1.5	1.3	2.5	1.4	0.8	1.8	1.7	2.9	2.3	2.0	1.4	0.3	2.7	3.2	1.6	1.2	1.4	1.8	1.5
M	2.2	1.1	1.9	1.8	1.3	0.6	2.5	2.3	2.2	2.5	2.3	0.9	0.2	2.5	2.4	1.0	1.1	1.2	1.6	1.9
N	0.6	0.2	0.5	0.5	1.0	0.2	0.1	0.7	0.7	0.7	1.1	0.7	0.2	0.0	0.7	0.6	0.2	0.3	0.4	1.0
P	0.5	0.0	0.3	0.5	0.6	0.1	0.7	0.4	0.2	0.8	0.3	0.3	0.0	0.4	0.4	0.5	0.3	0.2	1.1	0.2
Q	1.4	1.1	0.6	1.3	0.7	0.4	0.9	0.9	1.0	1.8	1.2	0.9	0.1	1.0	0.9	0.8	0.8	0.8	1.0	0.9
R	1.0	1.0	0.8	1.2	0.9	0.4	1.5	1.3	1.5	1.2	1.2	1.1	0.1	1.1	1.2	0.7	0.6	0.4	1.5	0.9
S	1.1	0.3	0.4	0.5	0.8	0.3	0.3	1.4	0.8	1.3	1.2	0.4	0.1	0.6	1.0	0.4	0.4	0.7	0.6	1.0
T	0.7	0.7	0.6	0.6	0.4	0.1	0.4	0.6	0.9	1.0	0.9	0.4	0.1	0.6	0.9	0.4	0.4	0.4	0.4	0.6
V	1.8	0.6	0.5	1.2	0.8	0.3	0.9	0.9	1.4	1.1	1.8	0.8	0.2	1.2	1.5	0.9	0.6	0.8	0.5	0.5
W	2.0	0.7	1.6	1.2	3.1	0.6	2.5	1.8	2.6	3.1	1.4	0.6	0.3	2.1	1.4	1.3	0.9	1.6	4.1	1.4
Y	1.5	1.4	0.7	1.0	1.3	0.3	1.3	1.0	1.5	1.7	2.0	0.7	0.2	0.8	1.2	1.1	0.7	0.9	1.3	1.4

以上のことから、それぞれのポジションでのアミノ酸残基の傾向指数において、最大の傾向指数同士ジペプチドの傾向指数が、ほとんど最大の値だった。また、アミノ酸残基の傾向指数が、1.0 以下同士の組み合わせのジペプチドの中には、1.5 以上の傾向指数をもつものがあった。一方で、アミノ酸残基の傾向指数が、1.0 以上同士の組み合わせのジペプチドの中には、0.5 以下の傾向指数をもつものがあった。また、 α ヘリックス構造は、ヘリックス双極子を有していることから、ポジションでのアミノ酸残基の傾向指数でも N 末端側に酸性アミノ酸残基同士のジペプチドや酸性アミノ酸残基を含むジペプチド、C 末端側に塩基性アミノ酸残基同士のジペプチドや塩基性アミノ酸残基を含むジペプチドの傾向指数は高い値だった。

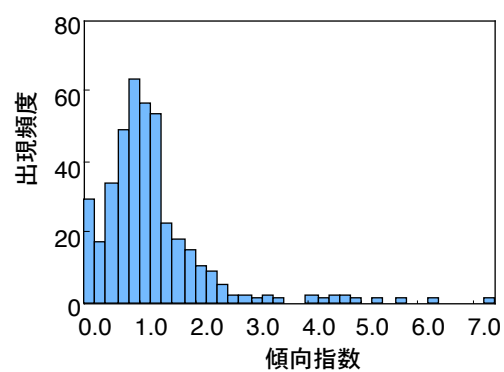
(a) C'C"ポジション



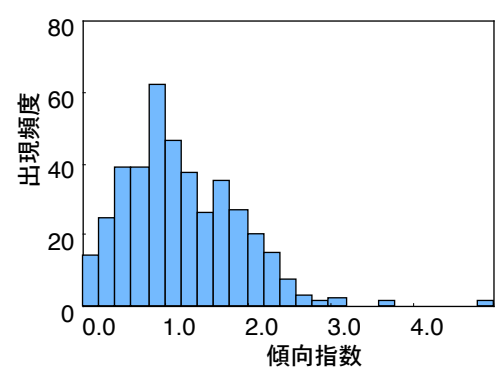
(b) CcapC'ポジション



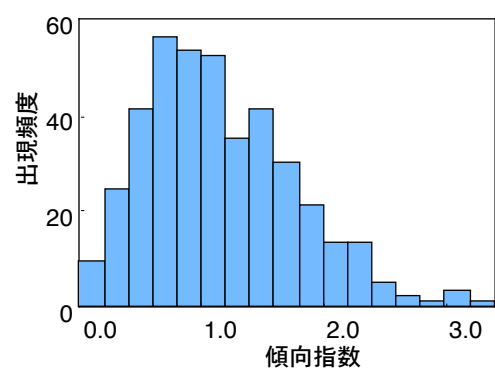
(c) C1Ccapポジション



(d) C2C1ポジション



(e) C3C2ポジション



(f) C4C3ポジション

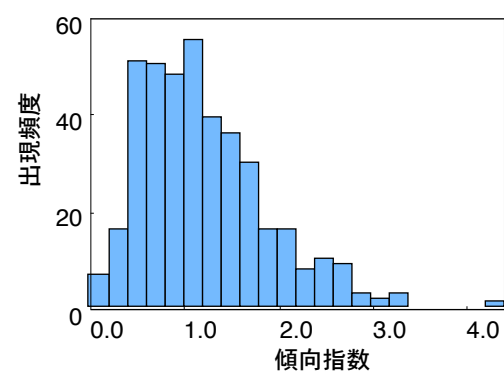


Figure 7-6. α ヘリックス構造のC末端領域でのジペプチドの傾向指数の分布

7—4—2 β ストランド構造の統計解析

α ヘリックス構造と同様に、ジペプチドを基本単位としたときの Richardson パラメータを算出するために、 β ストランド構造の中の各ポジションを次のように定義した (Figure 7-7)。

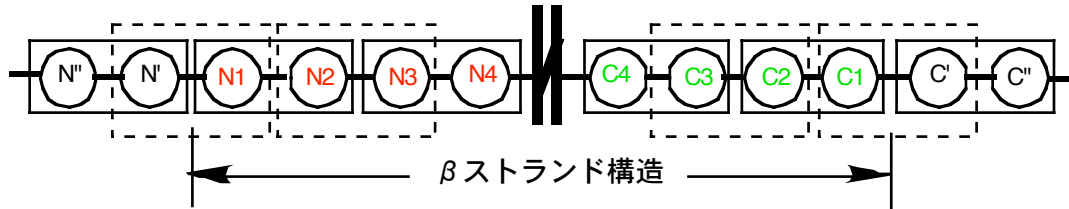


Figure 7-7. β ストランド構造の末端領域に対するポジション

まず、1081 種類の蛋白質に対して、 β ストランド構造のそれぞれのポジションでのアミノ酸残基の傾向指数は、次の通りである (Table 7-36)。

Table 7-36. β ストランド構造の末端領域でのアミノ酸残基の傾向指数

	N''	N'	N1	N2	N3	N4	C4	C3	C2	C1	C'	C''
A	1.0	0.6	0.9	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.8	0.9
C	1.1	0.6	1.2	0.9	1.1	1.2	1.2	1.1	1.2	1.1	0.7	1.0
D	1.2	1.5	0.6	1.0	0.6	0.7	0.5	0.4	0.5	1.9	1.2	1.8
E	1.0	0.8	0.9	0.7	0.6	0.7	0.7	0.9	0.7	0.6	0.9	1.4
F	0.8	0.6	1.3	0.9	1.3	1.3	1.5	1.4	1.3	0.7	0.7	0.7
G	1.1	3.3	0.8	0.5	0.6	0.5	0.6	0.7	0.6	0.7	2.0	1.4
H	1.0	1.0	1.2	0.9	0.9	1.0	1.0	1.1	0.9	1.1	0.9	1.2
I	0.7	0.4	1.2	1.1	1.6	1.6	1.7	1.6	1.5	0.7	0.7	0.5
K	0.9	1.0	1.2	1.1	0.7	0.7	0.8	0.8	1.0	0.7	0.8	1.0
L	0.9	0.5	1.0	0.7	1.1	1.0	1.1	1.1	1.2	0.6	0.8	0.5
M	0.8	0.8	1.1	0.7	1.0	0.9	1.0	0.9	1.2	0.5	0.6	0.6
N	1.2	1.7	0.8	0.9	0.7	0.7	0.6	0.6	0.6	1.5	1.3	1.5
P	1.6	0.6	0.3	2.7	1.5	1.3	1.0	1.1	1.4	2.5	2.2	0.3
Q	0.9	0.9	1.1	0.8	0.7	0.7	0.8	0.8	0.9	0.6	0.7	1.0
R	0.9	0.9	1.1	0.9	0.8	0.8	0.9	0.9	1.0	0.7	0.8	0.9
S	1.4	1.0	1.0	1.1	0.9	1.0	0.8	0.8	0.7	1.8	0.9	1.6
T	1.0	1.0	1.1	1.4	1.0	1.3	1.2	1.0	1.1	1.5	0.9	1.1
V	0.7	0.5	1.2	1.2	1.9	1.7	1.8	1.7	1.6	0.8	0.7	0.6
W	0.7	0.5	1.1	1.0	1.3	1.1	1.2	1.3	1.2	0.8	0.8	1.1
Y	0.8	0.6	1.3	0.9	1.3	1.2	1.3	1.4	1.3	0.8	0.7	0.8

Table 7-36 は、あくまでも、一つのポジションに対してのアミノ酸残基の傾向指数である。

そこで、ジペプチドの場合、二つのポジションを一組として、N 末端は、N''N', N'N1, N1N2, N2N3, N3N4, C 末端は、C'C'', C'C1, C1C2, C2C3, C3C4 として、ジペプチドを解析した。

α ヘリックス構造同様、 β スtrand構造の N 末端領域におけるジペプチドの出現頻度を調べた (Table 7-37~7-46)。

Table 7-37. β スtrand構造のN''N'ポジションでのジペプチドの出現頻度

N' \ N''	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	71	6	163	73	16	354	35	25	84	50	16	97	32	56	53	60	89	35	6	29	1350
C	18	5	18	13	3	40	6	10	13	12	0	13	13	14	14	16	17	8	1	5	239
D	47	12	71	61	24	400	28	32	75	46	15	99	25	40	63	67	55	27	9	31	1227
E	36	5	82	46	27	311	18	30	65	45	14	110	15	32	38	50	49	38	7	18	1036
F	21	5	52	25	7	118	10	9	44	27	5	44	26	18	16	45	25	16	9	10	532
G	61	7	122	65	34	259	29	38	102	69	24	111	26	45	84	103	89	48	17	30	1363
H	19	6	23	17	15	101	11	9	15	29	6	25	14	16	17	16	20	9	5	10	383
I	39	5	76	28	15	135	5	27	56	36	7	36	28	24	33	35	51	33	3	16	688
K	32	6	92	49	22	249	27	24	51	39	8	85	17	31	22	36	52	34	5	29	910
L	58	11	91	75	29	294	31	28	80	77	11	97	55	41	59	73	65	43	6	26	1250
M	12	2	21	14	5	84	0	3	14	20	7	30	3	17	9	15	14	12	1	9	292
N	29	11	75	27	14	305	24	12	51	32	8	85	29	26	35	48	35	26	6	22	900
P	63	9	136	101	38	260	23	37	76	55	15	105	31	39	45	83	61	45	12	36	1270
Q	34	8	41	16	21	144	19	13	32	23	5	54	20	28	27	25	29	14	4	13	570
R	31	4	68	39	25	180	19	18	32	28	9	70	11	20	35	35	42	20	4	22	712
S	53	15	97	57	30	322	46	26	86	59	17	86	52	58	76	93	77	35	10	21	1316
T	38	7	81	45	17	293	18	14	55	56	14	66	22	30	42	65	51	21	6	17	958
V	34	9	96	40	16	152	21	35	50	54	11	45	28	19	28	47	67	46	5	11	814
W	7	1	11	4	6	40	3	5	10	6	6	12	7	5	7	7	17	6	1	2	163
Y	27	3	49	26	15	117	13	7	27	25	6	35	26	25	25	32	15	10	3	10	496
Total	730	137	1465	821	379	4158	386	402	1018	788	204	1305	480	584	728	951	920	526	120	367	16469

Table 7-38. β ストランド構造のN'N1ポジションでのジペプチドの出現頻度

N1 \ N'	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	76	10	26	45	38	63	22	59	38	84	21	31	18	29	34	45	40	58	16	39	792
C	10	9	6	6	11	12	10	6	4	10	4	7	2	8	4	7	7	15	2	6	146
D	119	19	42	93	99	73	35	128	90	123	49	50	21	42	88	78	99	154	24	65	1491
E	55	11	45	57	40	42	25	50	70	76	22	38	13	28	63	44	55	66	13	37	850
F	27	8	17	20	24	38	13	24	28	38	8	16	5	14	16	20	20	31	5	17	389
G	303	54	172	262	186	199	94	317	403	356	100	132	70	215	241	210	254	384	58	200	4210
H	36	10	15	25	34	28	10	31	16	43	12	10	0	21	20	15	22	21	3	22	394
I	36	8	18	26	21	51	8	25	17	32	2	15	11	10	14	39	19	35	7	24	418
K	75	13	36	73	47	58	28	68	78	107	28	42	16	39	52	61	74	93	14	50	1052
L	62	10	31	52	58	22	57	49	84	20	34	13	28	35	60	43	73	3	31	817	
M	12	3	9	15	26	22	6	27	32	17	11	15	5	10	22	7	26	25	0	12	302
N	111	27	37	55	72	61	30	120	79	113	39	59	15	64	64	70	96	148	23	39	1322
P	35	12	15	40	38	22	17	39	27	39	9	21	7	13	30	26	32	44	13	34	513
Q	45	13	24	33	24	40	22	37	45	62	17	26	7	22	26	33	33	55	11	28	603
R	54	14	31	39	41	50	22	55	32	88	14	20	8	32	41	54	44	62	17	37	755
S	69	22	32	54	52	72	33	60	45	98	23	40	15	53	59	70	58	75	20	51	1001
T	76	15	54	71	43	75	41	48	69	84	15	29	12	37	54	60	68	62	13	29	955
V	75	9	20	20	24	57	12	38	31	45	20	20	6	19	26	31	34	38	7	20	552
W	18	4	5	9	5	11	4	3	10	9	4	2	1	2	5	7	5	13	3	4	124
Y	28	5	11	27	23	31	6	22	21	34	7	17	3	17	21	21	23	26	13	17	373
Total	1322	276	646	1022	900	1063	460	1214	1184	1542	425	624	248	703	915	958	1052	1478	265	762	17059

Table 7-39. β ストランド構造のN1N2ポジションでのジペプチドの出現頻度

N1 \ N2	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	76	12	63	66	44	47	31	97	84	85	29	45	154	53	67	80	117	108	17	48	1323
C	17	2	22	8	12	13	11	12	17	12	3	15	32	9	16	20	24	22	6	5	278
D	45	16	21	28	39	16	11	59	33	59	11	27	44	15	23	44	42	83	10	27	653
E	64	11	54	52	35	29	14	66	81	80	29	38	130	26	43	47	79	107	22	24	1031
F	54	9	57	47	19	45	23	54	59	45	11	45	117	36	37	79	73	65	6	24	905
G	69	10	58	45	49	46	22	85	50	84	27	49	87	42	44	70	66	87	21	53	1064
H	24	10	23	18	20	13	12	34	29	33	6	15	59	19	16	31	32	38	7	23	462
I	58	14	84	65	30	53	28	62	89	46	19	72	198	32	52	92	103	77	7	39	1220
K	62	8	46	69	42	35	16	98	83	101	23	47	167	34	44	66	98	112	15	25	1191
L	70	14	98	55	43	43	31	53	114	67	14	82	314	56	92	132	140	89	10	37	1554
M	17	3	29	27	13	25	10	23	26	21	6	20	64	11	20	30	44	28	2	9	428
N	41	16	33	25	26	18	12	60	32	50	11	12	92	17	22	30	26	61	11	30	625
P	23	5	12	10	10	7	5	18	14	20	5	8	18	10	23	22	12	29	2	7	260
Q	49	11	36	46	29	25	10	46	45	43	13	23	96	13	45	47	45	51	11	19	703
R	54	13	44	46	41	36	14	66	50	89	10	20	121	20	39	51	53	107	21	26	921
S	71	22	47	39	58	43	18	81	48	66	4	31	87	29	31	68	56	100	15	49	963
T	56	18	66	40	49	36	14	71	50	92	16	35	141	37	46	50	70	89	32	49	1057
V	86	9	115	80	37	62	30	67	97	91	19	70	187	49	82	123	123	106	16	36	1485
W	11	3	22	20	2	23	5	8	19	11	4	15	18	12	17	18	23	24	3	11	269
Y	33	10	55	34	24	27	20	44	52	43	9	38	87	27	43	50	70	63	8	29	766
Total	980	216	985	820	622	642	337	1104	1072	1138	269	707	2213	547	802	1150	1296	1446	242	570	17158

Table 7-40. β ストランド構造のN2N3ポジションでのジペプチドの出現頻度

N3 \ N2	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	51	13	21	34	33	36	21	76	33	73	16	23	81	15	27	40	34	92	9	29	757
C	11	4	8	5	5	8	2	10	5	16	3	7	12	5	6	8	14	20	3	9	161
D	28	5	5	12	30	15	6	42	11	27	9	8	22	8	12	16	17	66	7	32	378
E	30	4	21	30	40	15	10	69	32	84	19	11	47	16	17	24	27	92	7	30	625
F	27	9	26	30	23	23	16	46	32	42	9	19	24	21	18	38	38	54	7	22	524
G	28	6	14	19	19	22	10	45	19	40	7	12	23	12	22	28	27	50	9	24	436
H	13	6	8	8	20	13	5	29	9	22	4	6	12	7	9	6	7	35	3	11	233
I	62	14	52	49	38	44	23	75	57	97	24	29	58	30	41	47	74	107	12	31	964
K	63	9	23	20	52	34	17	81	26	81	22	21	68	16	37	55	29	141	12	35	842
L	47	13	38	38	43	23	21	68	39	91	12	31	86	34	39	63	66	120	21	52	945
M	13	2	15	8	12	4	10	15	9	13	4	15	24	7	15	4	16	28	5	7	226
N	14	6	7	12	26	21	4	40	11	38	7	8	22	6	6	12	19	53	3	17	332
P	114	31	48	42	61	52	30	107	56	146	32	43	78	37	46	82	81	199	20	54	1359
Q	25	4	13	13	24	14	8	41	15	60	10	8	35	13	12	17	28	55	9	23	427
R	43	12	19	14	38	31	9	59	19	70	10	16	51	15	24	24	36	104	12	23	629
S	22	10	17	19	41	33	10	61	17	54	15	14	30	16	14	20	16	76	19	28	532
T	44	19	18	20	53	31	14	89	27	101	18	16	82	22	16	27	44	112	22	47	822
V	78	10	65	53	60	55	27	123	51	132	24	48	81	32	60	61	74	151	22	53	1260
W	10	1	10	10	9	10	5	18	13	12	3	7	7	11	9	8	17	20	2	13	195
Y	29	8	20	14	17	24	10	42	19	33	12	17	34	17	28	33	37	61	10	20	485
Total	752	186	448	450	644	508	258	1136	500	1232	260	359	877	340	458	613	701	1636	214	560	12132

Table 7-41. β ストランド構造のN3N4ポジションでのジペプチドの出現頻度

N4 \ N3	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	34	11	20	28	35	21	16	60	24	47	11	15	52	13	21	35	36	77	7	33	596
C	14	3	7	6	5	5	4	17	3	10	2	4	8	6	5	9	9	10	3	8	138
D	13	3	4	3	11	10	2	25	4	21	3	3	11	7	7	10	10	25	1	13	186
E	20	7	15	14	28	9	8	41	12	29	3	12	21	7	9	11	34	54	7	12	353
F	28	11	29	31	29	27	10	48	28	44	12	24	32	19	25	35	44	55	4	20	555
G	28	3	12	17	22	15	6	38	20	31	8	9	20	7	6	28	29	46	4	13	362
H	7	4	10	4	9	5	5	26	9	15	2	7	10	5	10	8	11	27	8	6	188
I	84	10	58	47	39	48	25	71	47	78	19	50	65	28	39	68	78	100	11	35	1000
K	26	6	20	17	24	15	5	28	16	46	10	7	33	10	17	13	20	57	12	18	400
L	60	20	61	55	42	37	27	83	43	97	24	43	80	35	48	56	80	104	12	44	1051
M	18	4	12	7	5	10	7	29	1	17	3	10	19	1	5	16	20	32	2	15	233
N	17	2	6	4	1	5	4	21	6	19	6	3	8	4	11	10	16	30	3	16	192
P	39	9	20	27	19	25	14	39	17	45	14	19	41	30	22	32	45	67	4	15	543
Q	18	2	10	12	15	10	6	29	12	18	7	13	20	5	7	21	17	38	6	9	275
R	22	5	11	6	20	17	9	43	7	54	9	7	23	9	10	14	22	51	6	21	366
S	32	7	6	9	31	13	8	27	11	42	11	8	18	13	15	15	21	40	5	15	347
T	26	6	10	17	47	21	8	48	12	67	12	9	31	15	11	23	41	71	12	26	513
V	95	23	67	76	65	50	29	166	83	128	28	40	59	35	65	78	111	159	22	58	1437
W	7	6	2	5	8	11	11	15	10	14	2	7	13	8	8	11	16	15	2	7	178
Y	36	9	21	27	20	27	12	34	20	36	7	22	21	13	27	34	34	37	7	29	473
Total	624	151	401	412	475	381	216	888	385	858	193	312	585	270	368	527	694	1095	138	413	9386

つぎに、N 末端領域でのジペプチドの傾向指数を算出した (Table 7-42～7-46)。

Table 7-42. β ストランド構造のN''N'ポジションでのジペプチドの傾向指数

N' \ N''	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.5	0.3	2.1	0.8	0.3	3.5	1.1	0.3	1.0	0.4	0.5	1.7	0.6	1.0	0.7	0.8	1.2	0.4	0.3	0.6
C	1.1	1.2	1.4	1.1	0.3	2.2	0.8	0.9	1.3	0.6	0.0	1.2	1.1	1.5	1.2	1.1	1.3	0.6	0.4	0.6
D	0.6	1.0	1.5	0.9	0.5	5.0	1.4	0.5	1.4	0.5	0.8	2.5	0.5	1.3	1.4	1.3	1.1	0.4	0.6	0.8
E	0.4	0.4	1.4	0.6	0.7	4.4	0.9	0.5	0.9	0.4	0.5	2.2	0.4	0.7	0.7	0.9	0.9	0.5	0.4	0.5
F	0.4	0.5	1.2	0.6	0.3	2.4	0.6	0.3	1.1	0.5	0.4	1.4	0.8	0.8	0.5	1.0	0.6	0.4	1.0	0.4
G	0.6	0.4	1.7	0.9	0.7	2.7	1.0	0.5	1.3	0.7	0.9	2.1	0.6	0.9	1.4	1.3	1.2	0.5	1.1	0.6
H	0.7	0.9	1.1	0.8	0.8	3.3	1.0	0.4	0.9	0.8	0.7	1.7	0.6	1.1	0.9	0.8	1.0	0.4	0.8	0.7
I	0.5	0.4	1.2	0.5	0.5	2.0	0.2	0.5	0.9	0.5	0.4	0.8	0.6	0.7	0.7	0.6	0.8	0.5	0.3	0.5
K	0.4	0.6	1.5	0.7	0.6	3.9	1.4	0.4	0.8	0.4	0.4	1.9	0.4	0.8	0.5	0.7	1.0	0.5	0.4	0.8
L	0.5	0.6	1.0	0.8	0.6	2.9	0.9	0.4	0.8	0.6	0.3	1.4	0.7	0.7	0.7	0.8	0.8	0.5	0.4	0.5
M	0.4	0.5	0.9	0.6	0.4	3.2	0.0	0.1	0.5	0.7	0.8	1.7	0.2	1.2	0.5	0.8	0.7	0.5	0.3	0.8
N	0.5	1.1	1.8	0.7	0.5	5.0	1.4	0.3	1.3	0.5	0.5	2.5	0.6	1.0	1.1	1.2	0.9	0.5	0.5	0.8
P	1.0	0.9	2.6	1.6	1.2	4.1	1.4	0.9	2.0	0.9	1.0	3.0	1.0	1.3	1.4	1.7	1.5	0.8	1.1	1.3
Q	0.6	1.0	1.2	0.4	0.8	3.2	1.2	0.4	0.9	0.4	0.4	2.0	0.7	0.9	0.8	0.8	0.9	0.3	0.4	0.6
R	0.5	0.3	1.3	0.7	0.7	3.1	1.0	0.4	0.8	0.3	0.6	2.0	0.3	0.6	0.8	0.8	1.0	0.3	0.3	0.7
S	0.7	1.0	1.8	1.0	0.7	3.9	1.8	0.5	1.7	0.7	0.9	2.2	1.1	1.5	1.7	1.5	1.5	0.5	0.7	0.6
T	0.5	0.5	1.5	0.8	0.4	3.8	0.8	0.3	1.2	0.6	0.8	1.8	0.4	0.9	1.0	1.3	1.0	0.3	0.4	0.5
V	0.3	0.6	1.3	0.5	0.4	2.0	0.8	0.5	0.7	0.6	0.4	0.9	0.5	0.5	0.5	0.7	1.0	0.5	0.3	0.3
W	0.4	0.3	0.8	0.3	0.7	2.4	0.5	0.4	0.7	0.3	1.2	1.1	0.8	0.4	0.5	0.5	1.3	0.4	0.3	0.2
Y	0.6	0.4	1.3	0.7	0.7	2.6	1.0	0.2	0.9	0.5	0.5	1.2	0.9	1.0	0.8	0.9	0.4	0.3	0.3	0.4

Table 7-43. β ストランド構造のN'N1ポジションでのジペプチドの傾向指数

N' \ N1	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.5	0.5	0.3	0.5	0.7	0.6	0.7	0.7	0.5	0.6	0.6	0.5	0.3	0.5	0.5	0.6	0.5	0.6	0.8	0.8
C	0.6	2.1	0.5	0.5	1.2	0.6	1.3	0.5	0.4	0.5	0.9	0.6	0.2	0.8	0.3	0.5	0.5	1.1	0.8	0.7
D	1.4	1.5	0.8	1.4	2.0	0.9	1.7	1.9	1.6	1.3	2.4	1.2	0.4	1.3	1.9	1.4	1.9	2.1	1.4	1.6
E	0.6	0.9	0.7	0.7	0.9	0.6	1.2	0.8	0.9	0.7	0.7	0.7	0.3	0.6	1.1	0.8	0.9	0.8	0.8	1.0
F	0.5	0.8	0.4	0.5	1.0	0.8	0.8	0.7	0.7	0.6	0.6	0.5	0.2	0.6	0.5	0.4	0.5	0.7	0.5	0.7
G	3.0	2.9	2.4	3.7	3.4	2.0	3.2	4.0	4.8	3.4	3.6	2.5	1.5	4.4	3.8	2.6	3.2	4.2	3.5	3.9
H	1.2	1.4	0.7	1.1	1.7	0.9	0.9	1.3	0.9	1.2	1.4	0.7	0.0	1.4	1.1	0.7	1.1	0.8	0.5	1.5
I	0.4	0.6	0.3	0.4	0.7	0.7	0.3	0.5	0.3	0.4	0.1	0.3	0.2	0.3	0.3	0.6	0.3	0.5	0.7	0.8
K	0.9	1.2	0.6	1.0	1.2	0.9	1.4	1.1	1.2	1.2	1.2	0.9	0.3	1.0	1.1	1.1	1.4	1.3	1.1	1.4
L	0.5	0.5	0.3	0.5	1.1	0.5	0.6	0.8	0.5	0.6	0.6	0.5	0.2	0.4	0.4	0.6	0.5	0.8	0.2	0.6
M	0.4	0.7	0.4	0.6	1.8	0.8	0.8	1.2	1.2	0.5	1.2	0.8	0.3	0.7	1.1	0.4	1.2	1.0	0.0	1.0
N	1.8	2.5	0.9	1.4	2.4	1.0	1.7	2.5	1.9	1.7	2.3	1.7	0.3	2.3	1.9	1.7	2.3	2.6	1.9	1.4
P	0.5	1.2	0.3	0.6	1.1	0.3	1.0	0.9	0.7	0.6	0.6	0.6	0.2	0.4	0.9	0.5	0.8	0.7	1.2	1.2
Q	0.7	1.5	0.7	0.7	0.9	0.9	1.4	1.0	1.2	0.9	1.2	0.9	0.3	0.7	0.8	1.0	1.0	1.3	1.2	1.2
R	0.8	1.1	0.6	0.6	1.0	0.8	1.1	1.1	0.8	1.1	0.9	0.5	0.2	1.0	1.0	1.2	1.1	1.0	1.4	1.1
S	0.9	1.4	0.6	0.9	1.2	0.8	1.3	1.1	0.9	1.1	1.2	1.0	0.3	1.3	1.2	1.1	1.1	1.1	1.3	1.4
T	1.0	1.0	1.0	1.3	1.0	0.9	1.8	0.8	1.4	0.9	0.8	0.8	0.2	1.1	1.3	1.1	1.2	0.9	0.8	0.8
V	0.7	0.6	0.3	0.3	0.5	0.7	0.4	0.5	0.4	0.5	0.8	0.4	0.1	0.4	0.4	0.4	0.5	0.4	0.5	0.5
W	1.1	1.2	0.3	0.6	0.5	0.6	0.6	0.2	0.7	0.4	0.8	0.2	0.1	0.2	0.4	0.5	0.4	0.9	1.0	0.5
Y	0.6	0.6	0.3	0.7	1.0	0.7	0.5	0.7	0.6	0.7	0.6	0.5	0.1	0.7	0.6	0.6	0.7	0.7	1.4	0.7

Table 7-44. β ストランド構造のN1N2ポジションでのジペプチドの傾向指数

N2 \ N1	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.5	0.6	0.8	0.7	0.8	0.5	1.0	1.2	1.0	0.6	0.8	0.8	2.8	0.9	0.9	1.0	1.5	1.1	0.8	1.0
C	1.0	0.5	1.7	0.6	1.3	0.7	1.5	1.0	1.6	0.6	0.7	1.3	2.6	0.9	1.3	1.3	1.7	1.5	2.3	0.6
D	0.5	1.3	0.4	0.4	0.8	0.2	0.5	0.9	0.6	0.6	0.5	0.7	0.8	0.5	0.5	0.8	0.8	1.1	0.6	0.7
E	0.7	0.9	0.9	0.6	0.8	0.4	0.6	1.0	1.0	0.7	1.0	0.7	3.3	0.6	0.8	0.9	1.3	1.3	1.3	0.7
F	1.1	0.9	1.3	1.1	0.8	0.9	1.4	1.5	1.4	0.8	0.8	1.4	3.6	1.5	1.2	1.8	1.8	1.4	0.6	1.0
G	0.7	0.5	0.8	0.6	0.9	0.5	0.7	1.1	0.6	0.8	1.0	0.9	1.9	0.8	0.7	0.9	0.8	0.9	1.2	1.0
H	0.8	1.4	1.1	0.8	1.0	0.4	1.0	1.4	1.6	0.9	0.7	1.0	2.3	1.3	0.8	1.4	1.5	1.4	1.1	1.5
I	0.7	1.1	1.3	1.0	0.9	0.7	1.2	1.2	1.4	0.6	1.1	1.4	3.9	0.9	1.1	1.5	1.6	1.2	0.7	1.2
K	0.7	0.8	0.7	0.9	1.1	0.5	0.8	1.6	1.3	1.1	1.0	1.0	3.6	0.9	1.0	1.2	1.8	1.5	1.2	0.7
L	0.5	0.7	1.1	0.5	0.9	0.4	0.9	0.7	1.1	0.5	0.4	1.1	3.8	0.9	1.0	1.4	1.6	1.0	0.6	0.8
M	0.5	0.7	1.2	1.1	0.9	0.9	1.3	1.0	1.0	0.7	0.6	1.1	3.2	0.8	1.0	1.6	2.0	1.1	0.7	0.8
N	0.7	1.5	0.8	0.6	0.9	0.3	0.7	1.2	0.8	0.7	0.6	0.3	2.0	0.6	0.7	0.7	0.6	1.1	0.9	1.1
P	0.4	0.5	0.2	0.2	0.3	0.1	0.3	0.4	0.4	0.3	0.3	0.2	0.6	0.3	0.7	0.4	0.3	0.5	0.2	0.2
Q	0.8	1.3	1.0	1.0	1.1	0.5	0.6	1.2	1.2	0.6	0.9	0.8	3.4	0.4	1.3	1.4	1.4	1.2	1.1	0.8
R	0.8	1.1	0.8	0.8	1.0	0.6	0.7	1.3	1.2	1.1	0.6	0.5	3.5	0.6	0.9	1.1	1.3	1.6	1.7	0.8
S	0.9	1.4	0.8	0.7	1.3	0.5	0.7	1.4	0.9	0.7	0.2	0.7	1.8	0.7	0.6	1.0	1.0	1.5	0.9	1.3
T	0.8	1.2	1.2	0.7	1.2	0.5	0.6	1.2	1.0	1.0	0.9	0.9	2.6	1.1	1.1	0.9	1.3	1.3	2.0	1.4
V	0.8	0.6	1.5	1.0	0.8	0.8	1.1	0.9	1.3	0.9	0.7	1.3	3.2	1.1	1.3	1.8	1.7	1.2	1.1	1.0
W	0.7	0.9	1.5	1.3	0.2	1.3	0.7	0.7	1.3	0.5	0.8	1.3	2.1	1.0	1.2	1.3	1.6	1.6	1.0	1.3
Y	0.7	1.2	1.4	0.9	1.0	0.6	1.5	1.4	1.6	0.8	0.7	1.2	2.8	1.0	1.3	1.3	2.0	1.6	0.8	1.1

Table 7-45. β ストランド構造のN2N3ポジションでのジペプチドの傾向指数

N3 \ N2	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.5	1.0	0.4	0.5	0.9	0.5	0.9	1.3	0.6	0.8	0.6	0.6	2.1	0.4	0.5	0.7	0.6	1.3	0.6	0.9
C	0.9	1.3	0.9	0.6	0.8	0.6	0.4	1.2	0.7	1.2	0.9	0.9	1.4	0.7	0.7	0.7	1.4	2.0	1.6	1.6
D	0.5	0.6	0.1	0.3	0.9	0.3	0.4	0.9	0.3	0.4	0.6	0.3	0.6	0.3	0.4	0.4	0.4	1.3	0.6	1.1
E	0.4	0.5	0.5	0.5	1.3	0.3	0.7	1.5	0.6	1.1	0.9	0.3	1.7	0.5	0.4	0.6	0.6	1.6	0.6	1.2
F	0.8	1.3	0.8	1.0	1.4	0.6	1.4	1.8	1.1	1.0	0.9	0.8	1.0	1.2	0.8	1.2	1.3	1.7	1.0	1.3
G	0.4	0.5	0.3	0.4	0.5	0.3	0.5	0.8	0.3	0.5	0.4	0.3	0.7	0.3	0.5	0.5	0.5	0.8	0.8	0.7
H	0.6	1.2	0.5	0.5	1.4	0.6	0.6	1.6	0.7	0.8	0.6	0.6	0.7	0.7	0.7	0.4	0.5	1.9	0.6	1.0
I	1.0	1.6	1.1	1.1	1.7	0.9	1.4	2.0	1.3	1.8	1.9	0.8	1.6	1.2	1.2	1.1	1.6	2.3	1.6	1.4
K	1.1	1.2	0.5	0.4	1.9	0.7	1.2	1.9	0.6	1.2	1.3	0.6	2.1	0.6	1.1	1.4	0.8	2.7	1.3	1.4
L	0.5	1.0	0.6	0.5	1.2	0.3	0.8	1.3	0.5	1.0	0.5	0.6	1.5	0.7	0.6	0.9	1.1	1.8	1.8	1.5
M	0.6	0.6	0.9	0.5	1.1	0.2	1.8	1.0	0.5	0.6	0.6	1.1	1.7	0.7	1.1	0.3	1.0	1.5	2.4	0.8
N	0.3	0.8	0.2	0.4	1.2	0.5	0.3	1.2	0.4	0.8	0.6	0.3	0.7	0.3	0.3	0.4	0.6	1.3	0.4	0.8
P	2.5	4.2	1.2	0.9	2.5	1.1	2.4	3.6	2.0	3.2	2.9	1.7	3.5	1.6	1.9	2.3	2.7	4.8	2.5	2.6
Q	0.6	0.7	0.5	0.4	1.3	0.4	0.7	1.5	0.5	1.3	1.0	0.4	1.8	0.6	0.5	0.7	1.2	1.8	1.3	1.4
R	0.9	1.4	0.5	0.3	1.4	0.7	0.6	1.6	0.6	1.2	0.9	0.6	2.1	0.6	0.8	0.8	1.2	2.2	1.3	1.0
S	0.4	0.9	0.4	0.5	1.3	0.5	0.5	1.5	0.5	0.9	1.1	0.5	0.9	0.6	0.4	0.4	0.4	1.6	1.7	1.0
T	0.8	1.8	0.5	0.5	1.8	0.5	0.9	2.2	0.8	1.5	1.4	0.6	2.2	0.9	0.5	0.7	1.1	2.2	1.9	1.9
V	1.1	0.9	1.2	1.0	1.9	1.0	1.4	2.5	1.0	1.9	1.3	1.3	2.0	1.1	1.3	1.2	1.4	2.4	2.0	2.0
W	0.8	0.4	0.9	0.9	1.4	0.8	1.0	2.1	1.2	0.8	0.8	0.9	1.1	1.3	0.9	0.8	1.7	1.8	0.9	2.2
Y	0.9	1.3	0.7	0.5	1.0	0.7	1.1	1.9	0.8	0.9	1.4	0.8	1.5	0.9	1.2	1.2	1.5	2.2	1.5	1.1

Table 7-46. β ストランド構造のN3N4ポジションでのジペプチドの傾向指数

	N4	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
N3																					
A		0.4	1.1	0.5	0.5	1.2	0.4	0.9	1.3	0.5	0.6	0.6	0.5	1.7	0.4	0.5	0.8	0.9	1.4	0.6	1.3
C		1.5	1.3	1.0	0.9	1.0	0.5	1.0	2.6	0.5	0.9	0.8	0.6	1.2	1.1	0.7	1.1	1.2	1.3	2.1	1.8
D		0.3	0.4	0.1	0.1	0.4	0.2	0.2	0.7	0.1	0.4	0.3	0.1	0.4	0.4	0.3	0.3	0.3	0.6	0.1	0.6
E		0.4	1.1	0.4	0.3	1.2	0.2	0.7	1.1	0.3	0.5	0.2	0.4	1.0	0.3	0.3	0.4	1.1	1.2	0.7	0.6
F		1.0	2.0	1.2	1.3	2.2	1.0	1.1	2.4	1.2	1.4	1.5	1.4	1.8	1.4	1.5	1.4	1.9	2.2	0.8	1.5
G		0.5	0.3	0.3	0.4	0.7	0.3	0.4	0.9	0.4	0.5	0.5	0.3	0.8	0.3	0.2	0.6	0.7	0.9	0.4	0.5
H		0.4	1.0	0.8	0.3	0.8	0.3	0.8	1.9	0.9	0.7	0.4	0.9	0.7	0.6	1.0	0.7	1.0	1.9	2.2	0.7
I		1.8	1.4	1.6	1.3	2.2	1.2	1.9	2.5	1.4	1.8	1.9	1.8	2.3	1.4	1.5	2.0	2.2	2.8	1.9	2.0
K		0.6	1.0	0.6	0.4	1.1	0.4	0.4	0.9	0.4	0.9	0.8	0.3	1.3	0.5	0.7	0.4	0.7	1.4	1.7	0.9
L		0.8	1.9	1.2	1.0	1.6	0.6	1.4	2.1	0.8	1.4	1.3	1.1	1.8	1.0	1.0	1.1	1.7	2.0	1.3	1.6
M		1.0	1.6	0.9	0.5	0.6	0.7	1.7	2.4	0.1	1.0	0.6	1.0	1.7	0.1	0.5	1.5	1.7	2.3	1.2	2.3
N		0.5	0.3	0.3	0.2	0.1	0.1	0.4	0.8	0.3	0.5	0.6	0.2	0.3	0.3	0.6	0.4	0.7	1.0	0.5	1.0
P		1.1	1.6	0.7	0.7	1.0	0.7	1.4	1.7	0.8	1.3	1.6	0.9	2.4	1.7	1.2	1.1	2.0	2.1	0.6	0.9
Q		0.5	0.4	0.5	0.5	1.0	0.4	0.7	1.4	0.6	0.5	0.9	0.8	1.3	0.3	0.4	1.1	1.0	1.6	1.1	0.7
R		0.6	0.7	0.4	0.2	0.9	0.5	0.8	1.5	0.3	1.2	1.0	0.3	1.2	0.5	0.4	0.6	1.0	1.4	0.9	1.1
S		0.8	0.8	0.2	0.3	1.3	0.3	0.6	0.9	0.4	0.9	1.0	0.4	0.7	0.6	0.6	0.4	0.7	1.1	0.6	0.7
T		0.6	0.8	0.3	0.6	2.1	0.5	0.6	1.5	0.5	1.3	1.2	0.4	1.1	0.8	0.5	0.8	1.4	1.8	1.3	1.4
V		1.7	2.6	1.6	1.8	2.6	1.2	1.9	4.3	2.1	2.4	2.0	1.4	1.9	1.5	1.8	2.0	2.8	3.2	2.6	2.8
W		0.8	3.2	0.2	0.6	1.6	1.2	2.9	2.3	1.2	1.2	0.7	1.1	2.8	1.2	1.0	1.5	2.1	1.8	1.2	1.5
Y		1.5	1.9	1.0	1.4	1.6	1.0	1.6	2.0	1.1	1.3	1.1	1.3	1.2	0.9	1.4	1.6	1.8	1.7	1.4	2.1

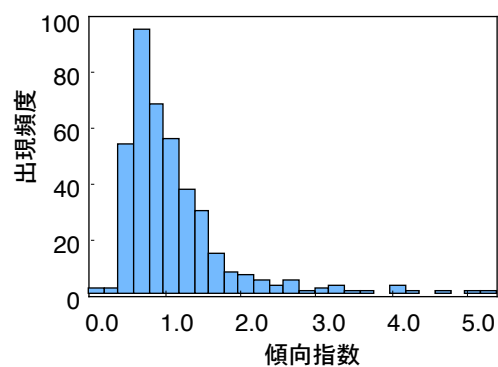
これらの結果から、ジペプチドの傾向指数は、特徴のない 1.0 への収束は見られず、特徴的な分布を示した (Figure 7-8)。

最大の傾向指数を示したジペプチドは、N''N'ポジションが DG, NG, N'N1 ポジションが GK, N1N2 ポジションが IP, N2N3 ポジションが PV, N3N4 ポジションが VI だった。これら五つのポジションにおいて、N2N3 ポジションが、N2 と N3 ポジションでの最大の傾向指数を示したアミノ酸残基の組み合わせだった。このことから、傾向指数が最大値であるアミノ酸残基の組み合わせは、ジペプチドで最大の傾向指数を示すとは限らない。たとえば、N''N'ポジションでは、最大の傾向指数をもつジペプチドは、DG と NG だった。しかしながら、N''ポジションの最大の傾向指数をもつアミノ酸残基は、プロリンである。N''N'ポジションでのジペプチド PG は、最大の傾向指数を示すと考えられるが、実際には、最大ではなかった。

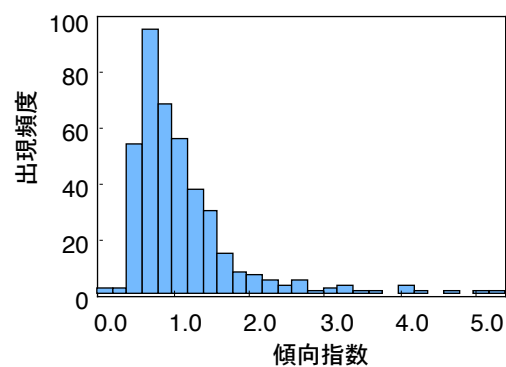
最小の傾向指数を示したジペプチドは、N''N'ポジションが CM, MH, N'N1 ポジションが HP, MW, N1N2 ポジションが PG, N2N3 ポジションが DD, N3N4 ポジションが DX (X=D, E, K, N, W), MK, MQ, NF, NG だった。その中でも、N''N'と N'N1 ポジションのジペプチドは、出現頻度が 0 だった。

また、それぞれのポジションでのアミノ酸残基の傾向指数が1.0以上のものの組み合わせで、ジペプチドの傾向指数が0.5以下のものは、N¹N¹ ポジションが HW, N²N³ ポジションが DL, DT, ST, WC だった。一方で、アミノ酸残基の傾向指数が1.0以下のものの組み合わせで、ジペプチドの傾向指数が1.5以上のものは、N¹N² ポジションが NC, N²N³ ポジションが MH, WT, YT, N³N⁴ ポジションが MH, MS だった。これらのジペプチドの中で、DT, WT, MH, MS は、アミノ酸残基の傾向指数が1.0の組み合わせであり、このことから、特徴がないアミノ酸残基でも、ジペプチド単位で調べることで、特徴が現れた。なお、ここでは、ジペプチドの傾向指数を0.5以下や1.5以上としたが、アミノ酸残基の傾向指数が1.0以上のものの組み合わせで、ジペプチドの傾向指数が1.0未満のもの、あるいは、アミノ酸残基の傾向指数が1.0以下のものの組み合わせで、ジペプチドの傾向指数が1.0より大きいものも存在した。

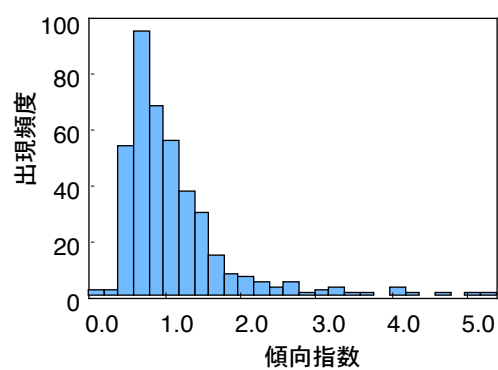
(a) N''N'ポジション



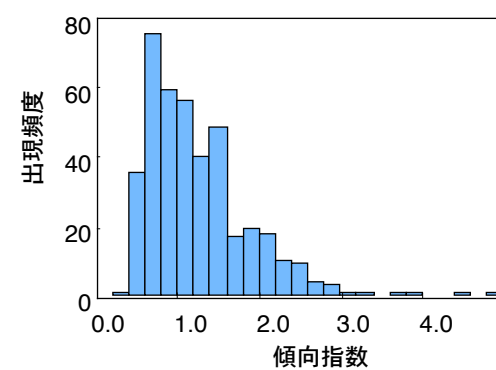
(b) N''N'ポジション



(c) N''N'ポジション



(d) N2N3ポジション



(e) N3N4ポジション

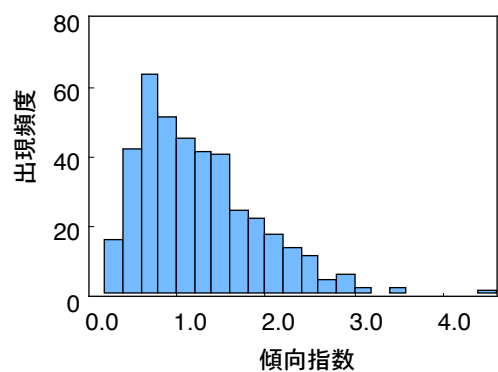


Figure 7-8. β ストランド構造のN末端領域でのジペプチドの傾向指数の分布

つぎに、 β ストランド構造のC末端領域でのジペプチドの出現頻度を調べた (Table 7-47～7-51)。

Table 7-47. β ストランド構造のC'C"ポジションでのジペプチドの出現頻度

C" \ C'	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	96	15	143	118	20	104	39	28	63	36	9	70	20	38	56	104	85	36	16	25	1121
C	15	2	5	8	5	21	7	6	4	3	2	15	4	6	9	22	10	6	3	7	160
D	80	9	122	97	43	186	37	37	72	43	13	71	12	44	44	105	73	59	13	42	1202
E	80	6	134	133	18	72	26	12	91	23	6	89	13	34	48	93	63	32	11	22	1006
F	33	6	49	47	15	55	9	13	22	16	8	29	8	14	16	45	29	16	6	14	450
G	188	43	247	164	79	265	63	102	138	139	54	153	71	116	108	222	173	151	27	69	2572
H	22	10	33	24	16	43	15	12	17	19	10	18	1	12	22	21	21	12	9	8	345
I	39	6	61	59	25	67	15	12	37	32	3	39	11	22	22	78	38	21	15	22	624
K	48	7	127	95	10	68	14	17	58	20	6	58	10	28	41	101	59	18	12	20	817
L	82	12	128	104	40	126	29	27	68	45	15	69	21	44	56	128	64	37	13	32	1140
M	21	4	25	16	9	25	3	3	19	12	4	18	6	8	11	12	14	7	3	12	232
N	69	16	69	47	25	161	26	33	50	44	18	61	24	29	30	53	61	80	13	28	937
P	151	22	207	242	66	81	37	51	87	84	22	113	5	74	64	148	90	67	37	74	1722
Q	33	5	57	45	8	45	18	13	15	20	4	34	7	28	11	51	36	18	5	11	464
R	41	6	74	68	22	67	15	17	30	23	4	46	6	32	27	56	35	20	14	27	630
S	74	18	73	68	28	103	30	21	48	31	10	74	10	48	47	102	54	26	15	33	913
T	74	13	98	62	19	104	36	19	53	32	9	50	11	26	40	109	57	33	10	15	870
V	84	10	88	58	14	94	26	20	49	31	16	50	7	47	28	72	44	19	8	13	778
W	13	0	21	18	7	17	5	5	11	10	4	17	6	10	14	16	9	4	3	6	196
Y	36	6	49	37	10	56	10	8	21	17	4	22	10	12	25	32	24	12	9	16	416
Total	1279	216	1810	1510	479	1760	460	456	953	680	221	1096	263	672	719	1570	1039	674	242	496	16595

Table 7-48. β ストランド構造のC1C'ポジションでのジペプチドの出現頻度

C' \ C1	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	66	9	88	51	11	190	14	22	44	43	13	52	89	25	28	53	62	38	9	29	936
C	16	3	16	11	10	31	6	10	13	27	5	19	22	8	18	16	14	11	1	4	261
D	154	24	89	145	79	161	40	102	88	189	35	87	265	46	74	105	103	101	40	50	1977
E	33	5	40	32	20	181	12	27	33	43	7	53	66	15	21	41	17	26	5	14	691
F	36	4	41	21	8	78	14	12	29	32	3	28	47	12	17	44	39	16	3	17	501
G	70	5	65	36	21	126	21	31	45	40	18	46	79	21	34	63	60	42	11	29	863
H	29	4	27	27	13	66	14	11	12	26	8	29	89	11	12	20	13	11	6	10	438
I	40	7	61	34	11	128	13	26	36	32	11	46	39	17	26	32	45	35	5	12	656
K	25	7	64	36	17	143	8	23	52	38	11	42	76	10	19	24	26	35	8	14	678
L	46	6	67	39	17	117	24	27	42	64	9	59	103	27	37	63	52	36	6	20	861
M	19	1	13	16	5	39	1	2	10	11	0	10	21	10	7	12	8	8	0	4	197
N	70	14	90	74	36	88	31	62	54	95	15	58	164	30	43	70	58	64	23	28	1167
P	162	20	105	133	57	417	25	94	105	156	37	88	124	73	77	90	89	107	19	50	2028
Q	30	7	27	16	17	92	7	15	21	28	1	30	43	13	19	13	18	16	4	11	428
R	29	4	43	36	16	111	12	18	25	39	7	42	50	14	31	24	30	25	6	16	578
S	146	21	129	114	49	201	45	54	93	121	26	66	199	62	72	113	101	88	22	49	1771
T	113	13	117	113	37	157	29	48	67	110	15	87	149	44	48	72	55	65	20	40	1399
V	50	10	100	60	20	168	19	43	51	52	5	68	68	19	35	45	60	56	9	16	954
W	6	1	17	14	5	28	7	5	7	12	4	17	17	8	8	13	15	7	1	2	194
Y	20	2	29	28	13	73	12	20	26	26	7	36	53	15	25	31	25	17	2	12	472
Total	1160	167	1228	1036	462	2595	354	652	853	1184	237	963	1763	480	651	944	890	804	200	427	17050

Table 7-49. β ストランド構造のC2C1ポジションでのジペプチドの出現頻度

C1 \ C2	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	56	20	120	50	32	66	30	42	38	45	17	63	161	21	43	113	98	54	13	29	1111
C	17	2	44	8	9	18	8	9	9	9	8	20	30	8	9	31	20	7	3	7	276
D	36	8	32	23	23	21	10	27	18	29	5	26	31	15	19	42	43	35	8	28	479
E	31	10	72	49	25	26	12	45	52	66	16	49	89	20	27	74	57	74	8	21	823
F	43	14	140	46	23	42	30	23	35	31	2	83	86	23	28	113	78	38	4	25	907
G	54	18	84	29	22	51	26	26	28	44	12	55	61	21	26	76	44	37	9	29	752
H	15	13	40	14	15	26	7	20	4	17	1	25	28	12	15	34	22	16	6	11	341
I	73	20	234	59	29	84	48	33	69	52	11	138	208	34	44	169	132	51	10	19	1517
K	70	11	71	36	30	45	14	71	48	63	22	54	137	23	38	113	68	71	11	31	1027
L	88	23	233	63	40	74	48	37	77	52	17	131	303	48	53	227	178	64	20	33	1809
M	20	5	63	22	11	24	8	10	15	21	4	36	74	12	13	47	47	12	5	11	460
N	26	8	40	16	16	22	15	42	23	28	1	21	48	10	8	34	31	34	8	21	452
P	81	18	95	59	42	65	24	38	43	70	18	59	87	39	48	126	82	77	17	34	1122
Q	31	9	49	29	19	35	13	26	29	36	5	25	80	14	20	71	32	38	5	10	576
R	53	24	80	28	30	35	23	38	29	60	12	52	99	18	27	65	64	75	10	28	850
S	46	7	50	16	29	49	19	41	19	51	7	44	55	15	15	64	65	44	16	36	688
T	58	14	91	32	49	41	27	51	31	77	12	66	143	25	26	77	85	78	13	43	1039
V	94	22	264	81	40	98	48	54	74	79	20	136	230	47	77	192	158	106	17	39	1876
W	16	4	43	11	4	17	8	11	8	9	1	22	22	8	10	32	33	13	1	7	280
Y	34	11	140	31	16	33	21	17	37	25	8	67	80	16	37	75	70	35	10	17	780
Total	942	261	1985	702	504	872	439	661	686	864	199	1172	2052	429	583	1775	1407	959	194	479	17165

Table 7-50. β ストランド構造のC3C2ポジションでのジペプチドの出現頻度

C2 \ C3	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	56	11	17	30	49	35	11	85	40	79	16	29	72	21	40	42	51	99	15	36	834
C	7	4	6	6	8	16	5	19	11	12	2	4	13	13	11	8	10	20	1	7	183
D	27	8	6	14	21	9	6	39	9	39	9	6	24	4	10	5	11	48	5	22	322
E	39	9	25	23	47	21	8	67	43	75	24	10	55	15	29	16	35	101	11	24	677
F	40	11	20	40	29	35	16	56	45	58	8	19	43	30	34	33	41	69	13	26	666
G	41	8	14	33	29	22	11	65	32	51	8	15	36	16	30	28	46	64	14	25	588
H	13	5	9	9	15	16	14	30	16	24	9	5	29	8	14	14	21	22	10	15	298
I	84	19	32	70	34	52	25	84	79	87	33	31	89	40	54	45	71	126	9	65	1129
K	30	7	13	24	31	19	7	57	28	77	20	17	61	18	25	17	29	76	11	16	583
L	63	8	36	54	31	49	23	88	75	86	30	35	138	53	69	47	90	104	19	42	1140
M	14	1	9	8	14	10	8	22	18	24	8	4	30	6	16	9	13	25	4	6	249
N	17	6	9	6	20	15	3	32	15	44	8	2	36	7	19	7	12	44	4	21	327
P	54	9	17	21	32	18	10	53	31	80	15	10	53	29	26	29	29	95	7	24	642
Q	27	4	7	16	35	19	3	40	13	41	6	8	35	6	23	21	18	42	6	16	386
R	33	7	13	22	28	23	7	55	27	72	8	8	61	23	21	17	21	77	10	24	557
S	36	15	10	24	35	31	13	58	29	59	15	10	29	15	20	27	24	69	12	30	561
T	35	11	16	22	48	22	10	62	40	82	23	12	63	12	17	20	51	92	15	41	694
V	94	20	36	83	77	58	31	126	111	137	29	47	87	43	85	54	104	154	23	65	1464
W	8	8	4	13	15	11	5	9	20	12	4	8	13	11	17	8	11	23	0	12	212
Y	39	14	17	35	18	26	10	49	32	46	15	22	51	20	39	33	52	55	13	31	617
Total	757	185	316	553	616	507	226	1096	714	1185	290	302	1018	390	599	480	740	1405	202	548	12129

Table 7-51. β ストランド構造のC4C3ポジションでのジペプチドの出現頻度

C3 \ C4	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
A	41	7	11	34	35	35	11	74	33	60	15	15	56	13	19	26	33	84	8	39	649
C	13	5	5	1	5	7	2	9	7	13	2	6	10	2	8	10	13	19	3	9	149
D	12	4	3	11	20	12	8	30	10	30	3	0	14	7	8	15	13	39	4	15	258
E	20	4	15	16	31	12	5	55	23	36	13	9	32	7	19	13	19	70	11	19	429
F	33	12	16	30	23	27	15	42	33	54	13	14	37	18	28	22	41	69	6	31	564
G	32	9	9	23	22	25	5	42	19	51	6	13	23	12	18	16	26	43	5	26	425
H	13	3	3	7	21	11	6	13	16	20	3	1	15	5	8	8	9	28	4	17	211
I	65	15	20	63	36	54	31	71	45	88	18	23	41	29	40	47	71	103	12	40	912
K	30	5	9	18	23	16	4	44	16	48	7	9	35	11	19	18	27	81	8	23	451
L	42	10	29	53	42	26	33	92	36	79	15	27	67	27	54	45	51	112	9	48	897
M	18	3	3	9	7	10	5	14	6	23	7	7	13	5	10	8	16	30	2	8	204
N	24	5	1	9	19	13	2	26	6	27	8	6	23	4	8	9	8	34	10	21	263
P	46	8	11	14	25	10	15	35	22	45	11	6	28	19	21	17	19	60	6	17	435
Q	21	1	12	12	18	10	8	30	21	32	5	10	23	5	12	4	22	37	8	9	300
R	29	7	8	22	35	23	5	43	17	42	6	6	30	8	14	12	21	63	7	17	415
S	32	8	9	23	28	19	6	60	12	39	7	7	33	5	8	14	25	59	8	23	425
T	44	12	8	24	35	32	12	87	26	63	12	8	34	17	26	21	24	89	24	34	632
V	89	15	29	80	57	49	25	111	49	102	32	33	61	39	59	60	80	144	24	39	1177
W	7	2	4	10	6	11	1	17	9	12	5	6	6	9	7	7	12	15	2	9	157
Y	32	5	13	22	17	21	12	40	15	30	6	14	18	25	21	24	35	44	4	30	428
Total	643	140	218	481	505	423	211	935	421	894	194	220	599	267	407	396	565	1223	165	474	9381

つぎに、これらの傾向指数を算出した (Table 7-52~7-56)。

Table 7-52. β ストランド構造のC'C"ポジションでのジペプチドの傾向指数

C" \ C'	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	0.7	0.8	1.8	1.3	0.4	1.0	1.2	0.4	0.8	0.3	0.3	1.2	0.4	0.7	0.8	1.4	1.2	0.4	0.8	0.5
C	0.9	0.5	0.4	0.7	0.6	1.2	1.0	0.5	0.4	0.2	0.5	1.4	0.3	0.6	0.8	1.5	0.8	0.4	1.2	0.9
D	0.9	0.7	2.5	1.5	0.9	2.3	1.8	0.6	1.3	0.5	0.7	1.8	0.2	1.4	1.0	2.0	1.4	0.8	0.8	1.1
E	0.9	0.5	2.2	1.7	0.4	1.0	1.2	0.2	1.2	0.2	0.2	1.8	0.3	0.8	0.9	1.7	1.1	0.4	0.7	0.6
F	0.7	0.6	1.1	1.1	0.6	1.1	0.6	0.4	0.5	0.3	0.6	0.9	0.3	0.6	0.5	1.0	0.7	0.4	0.7	0.6
G	1.9	2.4	3.5	2.4	1.5	2.7	2.2	1.3	1.7	1.4	2.0	2.9	1.6	2.4	1.8	2.8	2.3	1.7	1.7	1.4
H	0.8	1.5	1.6	1.1	0.8	1.4	1.3	0.5	1.0	0.5	1.2	1.2	0.0	0.8	1.2	1.0	1.0	0.5	1.4	0.5
I	0.5	0.5	1.0	0.9	0.8	1.0	0.7	0.2	0.6	0.4	0.2	0.8	0.2	0.6	0.5	1.3	0.6	0.3	1.5	0.7
K	0.6	0.7	2.1	1.3	0.3	1.1	0.7	0.3	0.9	0.2	0.3	1.3	0.2	0.8	0.9	1.8	1.1	0.2	1.0	0.6
L	0.6	0.6	1.4	1.1	0.8	1.2	0.8	0.4	0.7	0.4	0.5	1.0	0.3	0.7	0.7	1.4	0.8	0.4	0.8	0.7
M	0.7	0.9	1.1	0.7	0.6	0.9	0.4	0.1	0.7	0.4	0.4	1.0	0.3	0.6	0.6	0.7	0.7	0.3	1.0	1.1
N	1.2	1.5	1.7	1.2	0.9	2.6	1.5	0.7	1.2	0.7	1.1	1.8	0.5	1.1	0.9	1.3	1.5	1.5	1.1	1.0
P	2.4	2.2	3.9	3.8	2.0	1.3	2.2	1.3	2.3	1.3	1.5	3.2	0.2	2.4	1.9	3.0	2.2	1.2	3.4	2.6
Q	0.5	0.6	1.7	1.0	0.3	1.0	1.1	0.3	0.4	0.3	0.3	1.2	0.3	0.9	0.3	1.6	1.1	0.4	0.5	0.5
R	0.6	0.5	1.4	1.2	0.6	1.2	0.8	0.3	0.7	0.3	0.3	1.3	0.2	1.0	0.6	1.3	0.9	0.3	1.1	0.8
S	1.0	1.2	1.3	1.2	0.6	1.2	1.2	0.4	0.9	0.4	0.5	1.8	0.2	1.2	1.0	1.6	1.0	0.4	1.0	0.9
T	1.0	0.9	1.8	1.1	0.5	1.3	1.6	0.3	1.1	0.4	0.5	1.4	0.2	0.8	1.0	2.1	1.1	0.5	0.6	0.4
V	0.8	0.6	1.2	0.8	0.3	1.2	1.0	0.3	0.7	0.3	0.6	1.0	0.1	1.1	0.5	1.1	0.6	0.2	0.5	0.4
W	0.8	0.0	1.4	1.2	0.8	1.0	0.8	0.4	0.8	0.5	0.8	1.5	0.7	0.8	1.0	1.2	0.7	0.3	1.0	0.7
Y	0.8	0.7	1.3	1.1	0.4	1.2	0.8	0.3	0.7	0.3	0.3	0.7	0.3	0.5	0.8	0.9	0.7	0.3	1.0	0.7

Table 7-53. β ストランド構造のC1C'ポジションでのジペプチドの傾向指数

	C'	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
C1																					
A		0.5	0.5	1.1	0.5	0.2	1.8	0.4	0.3	0.5	0.3	0.4	0.9	1.6	0.4	0.4	0.7	0.8	0.4	0.4	0.6
C		0.9	0.7	1.2	0.9	1.1	1.7	0.8	0.9	1.2	1.4	1.1	1.7	1.8	0.8	1.5	1.1	1.0	0.8	0.4	0.5
D		1.8	1.9	1.8	2.2	1.6	1.9	1.9	1.5	1.6	2.1	1.7	2.1	4.8	1.4	1.6	1.9	1.9	1.4	2.4	1.2
E		0.3	0.4	0.6	0.4	0.5	2.5	0.6	0.4	0.4	0.4	0.2	1.0	1.7	0.3	0.4	0.8	0.3	0.3	0.3	0.4
F		0.7	0.4	0.9	0.5	0.3	1.6	0.8	0.3	0.7	0.5	0.2	0.9	1.4	0.5	0.6	1.0	0.9	0.4	0.3	0.7
G		0.7	0.3	0.9	0.5	0.4	1.3	0.7	0.4	0.5	0.4	0.7	0.9	1.7	0.4	0.5	0.8	0.8	0.5	0.7	0.6
H		1.0	0.6	1.3	1.2	0.7	2.1	1.2	0.4	0.7	0.7	0.9	1.9	3.6	0.8	0.6	0.9	0.6	0.4	0.9	0.7
I		0.5	0.6	0.9	0.5	0.3	1.8	0.6	0.5	0.6	0.4	0.6	0.9	0.8	0.5	0.5	0.5	0.7	0.5	0.5	0.4
K		0.3	0.7	1.0	0.5	0.4	2.2	0.4	0.4	0.8	0.4	0.5	0.9	1.7	0.3	0.4	0.4	0.5	0.5	0.6	0.4
L		0.3	0.3	0.7	0.4	0.3	1.1	0.7	0.4	0.4	0.5	0.3	0.8	1.2	0.4	0.4	0.7	0.6	0.4	0.4	0.4
M		0.6	0.2	0.6	0.7	0.3	1.4	0.1	0.1	0.4	0.3	0.0	0.5	1.1	0.7	0.4	0.6	0.4	0.3	0.0	0.3
N		1.2	1.3	2.1	1.8	1.2	1.4	1.8	1.3	1.3	1.4	0.9	1.7	3.5	1.1	1.3	1.7	1.4	1.1	1.9	1.0
P		2.5	1.9	1.9	2.0	1.7	6.4	1.4	2.3	2.7	2.4	2.4	2.4	4.0	2.3	2.2	1.8	2.1	1.8	1.7	1.7
Q		0.5	0.8	0.8	0.3	0.6	2.0	0.4	0.4	0.5	0.4	0.1	1.1	1.6	0.4	0.6	0.4	0.6	0.4	0.4	0.5
R		0.4	0.3	0.8	0.6	0.4	1.9	0.6	0.4	0.6	0.5	0.4	1.1	1.5	0.4	0.7	0.5	0.7	0.4	0.5	0.5
S		1.9	1.4	2.3	2.0	1.1	2.4	1.7	1.0	1.8	1.4	1.3	1.6	4.2	1.5	1.5	1.7	1.9	1.3	1.4	1.3
T		1.5	0.9	2.1	2.0	0.9	2.0	1.3	0.8	1.4	1.2	0.8	2.3	2.8	1.3	1.2	1.3	1.0	0.9	1.2	1.1
V		0.5	0.6	1.3	0.8	0.4	2.2	0.7	0.6	0.7	0.5	0.2	1.3	1.2	0.4	0.5	0.6	0.8	0.6	0.6	0.4
W		0.4	0.3	1.1	0.9	0.5	1.6	1.0	0.4	0.5	0.6	0.8	1.5	2.0	0.6	0.6	1.0	1.1	0.5	0.3	0.2
Y		0.5	0.2	0.7	0.8	0.6	1.5	0.9	0.6	0.8	0.5	0.6	1.2	1.7	0.6	0.7	0.8	0.7	0.4	0.2	0.5

Table 7-54. β ストランド構造のC2C1ポジションでのジペプチドの傾向指数

	C1	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
C2																					
A		0.4	1.1	1.5	0.5	0.6	0.6	0.9	0.5	0.5	0.3	0.5	1.1	2.9	0.4	0.6	1.5	1.3	0.5	0.6	0.6
C		1.0	0.5	3.3	0.6	1.0	1.0	1.1	0.8	0.9	0.5	1.7	1.8	2.5	0.8	0.7	2.0	1.5	0.5	1.2	0.9
D		0.4	0.6	0.6	0.3	0.5	0.2	0.5	0.4	0.3	0.3	0.2	0.6	0.6	0.5	0.4	0.8	0.8	0.5	0.5	0.7
E		0.3	0.8	1.1	0.6	0.6	0.4	0.6	0.7	0.7	0.6	0.5	1.0	2.3	0.4	0.5	1.3	1.0	0.9	0.5	0.6
F		0.9	1.4	3.1	1.0	1.0	0.8	1.8	0.6	0.8	0.5	0.1	2.6	2.6	0.9	0.9	2.5	1.9	0.8	0.4	1.0
G		0.5	1.0	1.2	0.4	0.4	0.5	0.9	0.3	0.3	0.4	0.4	1.0	1.3	0.4	0.4	0.9	0.6	0.4	0.5	0.6
H		0.5	1.9	1.8	0.6	0.8	0.8	0.6	0.8	0.2	0.5	0.1	1.7	1.1	0.8	0.8	1.6	1.1	0.6	0.9	0.7
I		0.9	1.6	3.5	0.9	0.9	1.2	2.0	0.6	1.1	0.7	0.6	2.8	4.1	0.9	0.9	2.7	2.1	0.8	1.0	0.6
K		0.8	1.1	1.1	0.5	0.8	0.7	0.7	1.2	0.7	0.7	0.9	1.2	3.0	0.6	0.8	2.0	1.2	1.0	0.9	0.9
L		0.7	1.2	2.5	0.6	0.8	0.7	1.3	0.5	0.8	0.4	0.5	1.8	3.6	0.7	0.6	2.4	2.0	0.7	1.2	0.7
M		0.6	1.1	2.7	0.9	0.7	0.9	1.0	0.5	0.6	0.7	0.4	1.9	3.7	0.8	0.6	2.5	2.2	0.5	1.7	0.9
N		0.4	0.7	0.9	0.4	0.5	0.3	0.9	0.9	0.5	0.4	0.1	0.6	1.0	0.4	0.2	0.8	0.7	0.6	0.7	0.7
P		1.2	1.7	1.7	0.9	1.2	1.0	1.4	0.9	1.1	1.1	1.2	1.6	2.8	1.2	1.4	2.4	1.9	1.3	1.5	1.2
Q		0.5	1.1	1.4	0.6	0.7	0.7	0.8	0.7	0.7	0.5	0.4	0.9	2.9	0.4	0.6	2.1	1.0	0.9	0.5	0.4
R		0.8	1.9	1.5	0.5	0.8	0.6	1.1	0.7	0.7	0.7	0.8	1.4	2.9	0.5	0.6	1.5	1.5	1.1	0.8	0.8
S		0.6	0.5	0.9	0.3	0.7	0.6	0.7	0.7	0.4	0.6	0.3	1.1	1.2	0.4	0.3	1.0	1.2	0.7	1.0	1.0
T		0.8	1.0	1.6	0.6	1.2	0.5	1.2	0.9	0.6	0.8	0.7	1.8	2.7	0.7	0.6	1.4	1.6	1.1	0.8	1.2
V		0.9	1.4	3.4	1.0	0.9	1.3	1.7	0.8	1.0	0.8	0.8	2.5	3.9	1.1	1.2	2.7	2.2	1.2	1.1	1.0
W		1.0	1.2	2.8	0.7	0.4	1.0	1.2	0.9	0.5	0.4	0.2	1.9	2.6	0.6	0.7	2.4	2.3	0.8	0.3	0.8
Y		0.8	1.3	3.5	0.9	0.7	0.7	1.6	0.5	1.1	0.5	0.7	2.1	2.5	0.6	1.1	2.0	2.0	0.9	1.1	0.7

Table 7-55. β ストランド構造のC3C2ポジションでのジペプチドの傾向指数

	C2	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
C3																					
A	0.6	0.8	0.3	0.4	1.3	0.5	0.5	1.5	0.7	0.8	0.6	0.7	1.8	0.5	0.7	0.8	0.9	1.4	1.0	1.1	
C	0.6	1.3	0.6	0.7	1.2	1.2	0.9	2.3	1.5	0.9	0.6	0.5	1.5	1.8	1.3	0.7	1.0	2.0	0.5	1.2	
D	0.4	0.9	0.2	0.3	0.6	0.2	0.4	0.8	0.2	0.6	0.6	0.2	0.6	0.2	0.3	0.1	0.3	0.9	0.4	0.8	
E	0.6	1.1	0.6	0.4	1.5	0.4	0.5	1.4	0.8	1.0	1.1	0.3	2.0	0.5	0.7	0.4	0.8	1.7	0.9	0.9	
F	1.1	1.5	0.6	1.3	1.7	1.0	1.4	2.2	1.5	1.4	0.8	0.8	1.8	1.7	1.5	1.0	1.4	2.1	1.9	1.5	
G	0.6	0.6	0.3	0.7	0.8	0.3	0.5	1.1	0.5	0.7	0.4	0.4	1.1	0.5	0.7	0.5	0.8	1.0	1.2	0.7	
H	0.6	1.0	0.6	0.6	1.1	0.7	1.7	1.7	1.3	0.9	1.5	0.5	1.6	0.8	1.0	0.9	1.4	1.2	2.1	1.4	
I	1.4	2.1	0.7	1.5	1.5	1.0	1.5	2.3	1.8	1.6	2.6	0.9	2.5	1.5	1.6	1.0	1.6	2.7	1.2	2.9	
K	0.5	0.9	0.3	0.5	1.1	0.4	0.5	1.4	0.6	1.2	1.2	0.5	1.9	0.7	0.8	0.4	0.8	1.4	1.2	0.6	
L	0.7	0.6	0.6	0.8	0.9	0.6	0.9	1.7	1.0	0.9	1.3	0.7	2.4	1.2	1.1	0.7	1.4	1.6	1.6	1.2	
M	0.6	0.3	0.5	0.5	1.3	0.5	1.5	1.4	1.0	1.1	1.2	0.3	2.1	0.6	1.1	0.7	0.8	1.4	1.9	0.7	
N	0.4	0.8	0.3	0.2	0.9	0.3	0.2	0.9	0.5	0.9	0.7	0.1	1.1	0.4	0.8	0.2	0.4	1.1	0.5	1.0	
P	1.2	1.2	0.4	0.5	1.3	0.4	0.8	1.8	1.1	1.7	1.4	0.4	2.4	1.3	1.1	0.8	1.0	2.3	0.9	1.2	
Q	0.6	0.7	0.3	0.5	1.8	0.6	0.3	1.5	0.5	0.9	0.6	0.4	1.8	0.3	1.0	0.9	0.8	1.3	0.9	0.9	
R	0.7	0.8	0.3	0.5	1.0	0.5	0.5	1.5	0.9	1.2	0.7	0.3	2.5	1.0	0.7	0.5	0.7	1.7	1.1	1.0	
S	0.7	1.4	0.3	0.6	1.1	0.5	0.7	1.4	0.8	0.9	1.1	0.3	0.9	0.5	0.6	0.6	0.6	1.5	1.1	1.1	
T	0.7	1.1	0.4	0.6	1.6	0.4	0.6	1.5	1.2	1.2	1.8	0.5	1.7	0.5	0.6	0.5	1.3	1.8	1.3	1.6	
V	1.3	1.7	0.7	1.5	2.4	1.1	1.6	2.5	2.1	2.0	1.6	1.2	2.1	1.4	1.9	1.1	2.0	2.4	2.1	2.5	
W	0.7	3.3	0.4	1.2	2.3	0.9	1.0	1.1	1.9	0.8	1.1	1.0	2.1	1.3	1.7	0.8	1.1	2.1	0.0	2.0	
Y	1.2	2.3	0.6	1.4	1.1	0.8	1.1	2.2	1.4	1.3	1.8	1.0	2.3	1.1	1.6	1.2	2.1	2.0	2.0	1.7	

Table 7-56. β ストランド構造のC4C3ポジションでのジペプチドの傾向指数

	C3	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
C4																					
A	0.5	0.7	0.2	0.7	1.2	0.6	0.6	1.7	0.7	0.8	0.8	0.5	1.9	0.4	0.5	0.6	0.8	1.5	0.7	1.5	
C	1.4	2.1	0.7	0.1	1.0	0.7	0.5	1.4	1.2	1.2	0.8	1.0	1.5	0.4	1.2	1.2	1.7	2.4	2.1	2.0	
D	0.3	0.6	0.1	0.3	0.8	0.3	0.7	0.8	0.3	0.6	0.3	0.0	0.5	0.4	0.3	0.5	0.4	1.0	0.4	0.7	
E	0.4	0.6	0.4	0.4	1.3	0.3	0.4	1.5	0.5	0.6	0.8	0.3	1.5	0.3	0.6	0.4	0.6	1.6	1.2	1.0	
F	1.2	2.2	0.7	1.2	1.7	1.0	1.6	2.1	1.4	1.7	1.7	0.8	2.1	1.3	1.6	0.9	1.8	2.8	1.2	2.3	
G	0.6	0.9	0.2	0.6	0.7	0.5	0.3	1.0	0.4	0.9	0.4	0.4	0.9	0.4	0.5	0.4	0.6	0.9	0.5	0.9	
H	0.8	0.8	0.3	0.6	1.9	0.6	0.9	1.0	1.6	1.0	0.6	0.1	1.1	0.6	0.8	0.7	0.8	1.9	1.1	2.0	
I	1.4	2.2	0.6	1.8	2.1	1.4	2.4	2.5	1.3	2.1	1.8	0.8	1.5	1.4	1.5	1.4	2.0	2.9	2.1	2.3	
K	0.7	0.9	0.3	0.4	1.1	0.4	0.4	1.4	0.4	0.9	0.5	0.4	1.4	0.5	0.8	0.6	0.9	2.0	1.1	1.2	
L	0.6	0.9	0.6	1.0	1.6	0.4	1.7	2.3	0.7	1.1	0.8	0.7	1.5	0.8	1.1	0.9	1.1	2.2	1.0	1.8	
M	1.0	1.2	0.2	0.7	0.9	0.7	1.2	1.2	0.4	1.3	1.4	0.7	1.2	0.6	0.9	0.8	1.4	2.1	1.2	1.3	
N	0.7	0.9	0.0	0.4	1.1	0.4	0.2	1.0	0.3	0.7	0.8	0.3	0.9	0.3	0.4	0.4	0.4	1.1	1.5	1.3	
P	1.3	1.4	0.4	0.4	1.3	0.3	1.5	1.5	1.0	1.3	1.3	0.3	1.6	1.1	1.1	0.6	0.8	1.9	1.0	1.1	
Q	0.6	0.2	0.6	0.5	1.2	0.4	0.9	1.4	1.0	0.9	0.7	0.6	1.5	0.3	0.7	0.2	1.2	1.5	1.5	0.7	
R	0.8	1.0	0.3	0.7	1.6	0.7	0.5	1.5	0.7	0.9	0.7	0.3	1.6	0.4	0.6	0.5	0.9	1.8	1.0	0.9	
S	0.8	0.9	0.3	0.7	1.1	0.4	0.4	1.9	0.4	0.8	0.6	0.3	1.3	0.2	0.3	0.4	0.8	1.6	0.9	1.1	
T	1.1	1.5	0.3	0.8	1.6	0.7	1.0	2.8	1.0	1.2	1.2	0.4	1.2	0.9	1.1	0.7	0.8	2.3	2.7	1.8	
V	1.6	1.7	0.7	1.9	2.3	1.1	1.6	2.9	1.2	1.9	2.3	1.1	1.9	1.7	1.7	1.6	2.0	2.9	2.9	1.9	
W	0.8	1.1	0.5	1.2	1.2	1.2	0.3	2.6	1.1	1.0	1.8	1.0	1.3	1.3	0.9	1.0	1.5	1.8	1.2	2.0	
Y	1.3	1.1	0.6	1.1	1.3	0.8	1.6	2.4	0.8	1.1	0.9	0.8	1.0	1.8	1.1	1.1	1.8	2.1	0.8	2.2	

これらの結果から、N 末端領域のジペプチドの傾向指数の分布と同様に、1.0 への収束ではない特徴的な分布を示した (Figure 7-9)。最大の傾向指数を示したジペプチドは、C'C''ポジションがPD, C1C'ポジションがPG, C2C1 ポジションがIP, C3C2 ポジションがWC, C4C3 ポジションがIV, VI, VV, VW だった。これら五つのポジションにおいて、C'C'と C4C3 ポジションが、それぞれのポジションでのアミノ酸残基の傾向指数で最大値だったアミノ酸残基の組み合わせだった。このことから、N 末端領域でのジペプチドの傾向指数と同じことがいえる。最小の傾向指数を示したジペプチドは、C'C''ポジションがHP, WC, C1C'ポジションがMM, MW, C2C1 ポジションがFM, HM, NM, C3C2 ポジションがWW, C4C3 ポジションがDN, ND だった。その中でも C2C1 ポジションのジペプチド以外は、出現頻度が0 だった。

それぞれのポジションでのアミノ酸残基の傾向指数が 1.0 以上のものの組み合わせで、ジペプチドの傾向指数が 0.5 以下のものは、C2C1 ポジションがCC, C3C2 ポジションがWW, C4C3 ポジションがCH, WH だった。一方で、アミノ酸残基の傾向指数が 1.0 以下のものの組み合わせで、ジペプチドの傾向指数が 1.5 以上のものは、C'C''ポジションがHC, C3C2 ポジションがMH, C4C3 ポジションがHK だった。なお、ここでも、N 末端領域のジペプチドの解析と同様、ジペプチドの傾向指数を 0.5 以下や 1.5 以上のものを対象としたが、アミノ酸残基の傾向指数が 1.0 以上のもの同士、あるいは 1.0 以下のもの同士で、ジペプチドの傾向指数が 1.0 未満、あるいは 1.0 より大きいものも存在した。

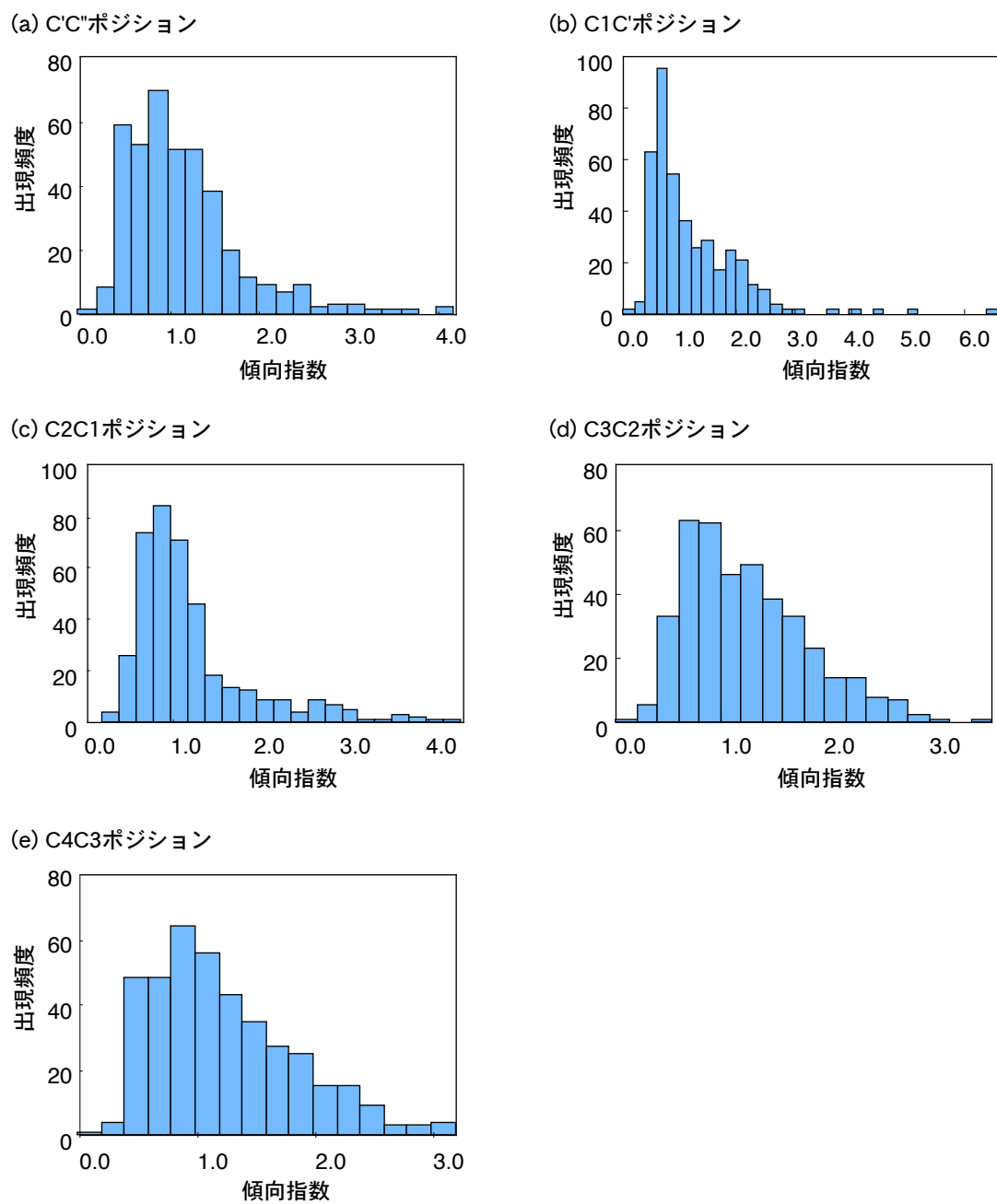


Figure 7-9. β ストランド構造のC末端領域でのジペプチドの傾向指数の分布

7—5 ジペプチド単位パラメータの自由エネルギー計算

7—5—1 はじめに

蛋白質の立体構造のエネルギーに関する研究の基本的な方針は、取り得る構造のエネルギーを計算し、最小化エネルギーを探したり、構造を比較することである。これは、構造の対象が、アミノ酸残基、ジペプチド、オリゴペプチドでも変わらない^{69,70,73)}。これらの研究は、エネルギー計算の結果から得られたデータを使ったアプローチによる。つまり、統計的な解析によって得られた二次構造の主鎖二面体角を使ったアプローチではない。この節では、二次構造としての α ヘリックス構造と β ストランド構造の中に現れるアラニンの主鎖二面体角を使って、モデルジペプチドであるアラニル・アラニンの立体構造に対する自由エネルギーを計算した。そして、それらの自由エネルギーを比べて、立体構造の安定性について検討した。ただし、ここで用いた主鎖二面体角は、アラニル・アラニンの (ψ, ϕ) ではなく、アラニンの (ϕ, ψ) を使った。

7—5—2 Ala-Ala の自由エネルギー計算

モデルジペプチドであるアラニル・アラニンの立体構造の主鎖二面体角には、二次構造の中のアラニン残基のものをを使った。つまり、 α ヘリックス構造の中におけるアラニン残基の平均的な値 $(\psi, \phi) = (-42^\circ, -62^\circ)$ と、 β ストランド構造の中におけるアラニン残基の平均的な値 $(\psi, \phi) = (142^\circ, -111^\circ)$ を使った。また、完全に伸びた構造となる主鎖二面体角 $(\psi, \phi) = (-180^\circ, -180^\circ)$ も使った。そして、これらの主鎖二面体角を使って、それぞれの構造の中におけるアラニル・アラニンの自由エネルギー（溶媒和自由エネルギーと立体構造エネルギーの和）を算出した（Figure 7-10）。溶媒和自由エネルギーは、溶媒と溶質分子との親水相互作用の寄与であるボルンエネルギーと、疎水水和の自由エネルギーの和から算出した。また立体構造エネルギーの算出には、Scheraga の ECEPP を使った。さらに、アラニル・アラニンの N 末端は NH_2 、C 末端は COOH とし、イオン化していない状態で計算した。

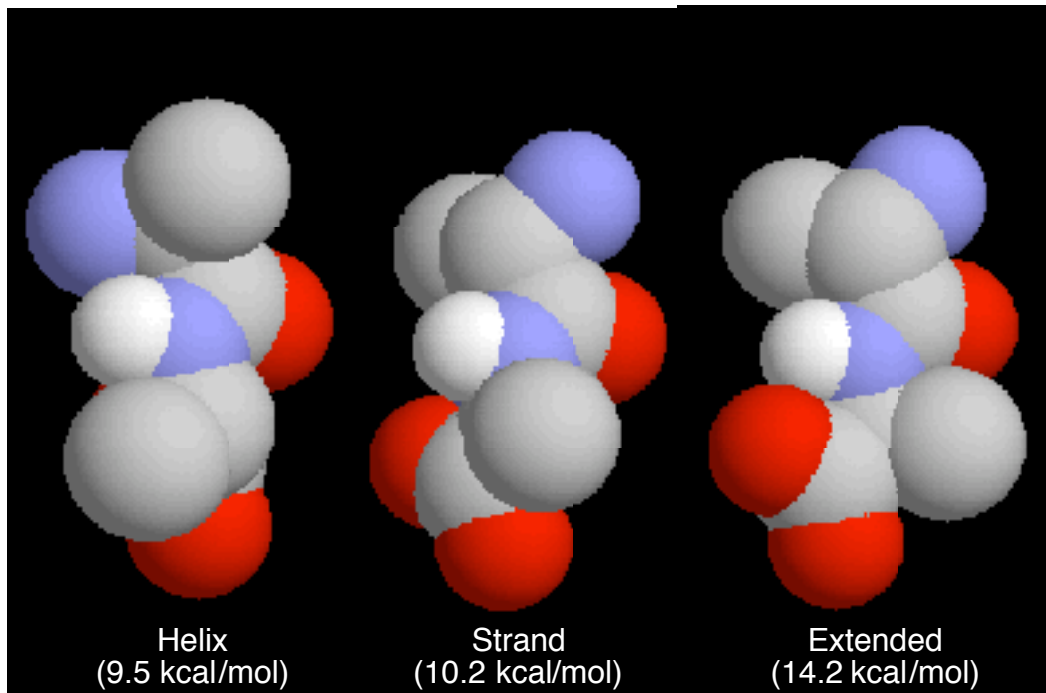


Figure 7-10. 各二次構造のアラニル・アラニンの自由エネルギー

Figure 7-10 のアラニル・アラニンは、左から α ヘリックス構造、 β ストランド構造、完全に伸びた構造であり、自由エネルギーは、 α ヘリックス構造では 9.5 kcal/mol、 β ストランド構造では 10.2 kcal/mol、完全に伸びた構造では 14.2 kcal/mol だった。ただし、それぞれの構造の中のジペプチドの自由エネルギーは、一つ一つの値だけでは、意味をもたないが、それぞれのジペプチドの自由エネルギーの差を見たときに意味をもつ。したがって、アラニル・アラニンは、 α ヘリックス構造の中では、他の構造の中よりもエネルギーが小さかった。 β ストランド構造の中での自由エネルギーは、 α ヘリックス構造より大きかったが、完全に伸びた構造より小さかった。以上のことから、モデルジペプチドであるアラニル・アラニンは、 α ヘリックス構造の中のほうが他の構造の中よりもエネルギー的に安定であり、 β ストランド構造の中では、完全に伸びた構造よりもエネルギー的に安定だった。

構造における自由エネルギーの計算は、今回、使った主鎖二面体角の周辺の角度についても調べ、構造の中でどの角度がエネルギー的に安定であるかを調べる必要がある。

この章では、蛋白質の二次構造を解析ための基本単位として、アミノ酸残基ではなく、ジペプチドを使った。その結果、ジペプチドの傾向指数が高いものは、アミノ酸残基においても傾向指数が高いもの同士の組み合わせだった。同様に、ジペプチドの傾向指数が低いものは、アミノ酸残基においても傾向指数が低いもの同士の組み合わせだった。しかしながら、アミノ酸残基の傾向指数が1.0、もしくは1.0未満のもの組み合わせの中に、ジペプチドの傾向指数が、1.5以上、または0.5以下の値を示した組み合わせが存在した。つまり、アミノ酸残基を使った解析で得られない結果を得た。また、二次構造が取り得る主鎖二面体角を使って、それぞれの二次構造の自由エネルギーを計算し、比較した。その結果、 α ヘリックス構造は、 β ストランド構造よりエネルギー的に安定であり、 β ストランド構造は、完全に伸びた構造よりも安定だった。

これらのことより、基本単位としてジペプチドを使うことにより、アミノ酸残基の解析からは得られない情報を得た。しかしながら、ジペプチドを基本単位としたことで、解析する数が、20種類から400種類へ変わったことによって情報量が増え、複雑化し、解釈しにくくなった可能性がある。

第8章 結論

この論文は、「蛋白質の二次構造帰属法 ProSSA の応用と関連アプリケーションの開発」と題して、主鎖二面体角の連続性パターンによる二次構造帰属法 ProSSA を、二次構造に関連したさまざまな研究に適用し検討した。

第2章では、主鎖二面体角の連続性パターンによって蛋白質の二次構造を帰属する方法 ProSSA とそのデータベースについて述べた。蛋白質の二次構造を帰属するのに広く使われている二次構造帰属法である DSSP 法は、規則的でないコンホメーションを二次構造と帰属したり、 β シート構造と同じコンホメーションを帰属しないなどの問題がある。その問題を解消するために、DSSP の帰属結果を包括する主鎖二面体角の連続性パターンによる二次構造帰属法 ProSSA のデータベースを構築し、Web 上に公開した。この ProSSA 法によって帰属されるファイルを使って、第3章から第7章までの研究を進めた。

第3章では、ProSSA 法を利用した Richardson パラメータを使って一次配列から α ヘリックス構造の末端の境界領域を予測した。ProSSA 法は、コンホメーションの規則性により二次構造と二次構造ではない部分を判別するため、二次構造の境界が明確である。したがって、ProSSA 法によって算出されたアミノ酸残基に対する Richardson パラメータを使って算出したスコアよりアミノ酸配列から α ヘリックス構造の末端を予測した。その結果、Chou-Fasman 法による二次構造予測では予測できなかった末端部位の予測が精密にできた。また、この末端部位の予測法を構造既知の蛋白質に適用した。そして予測と一致しなかった領域は、二次構造上、不安定な領域と考えた。つまり、二次構造の末端予測法を、構造既知の蛋白質の配列へ適用することで、二次構造の不安定領域の検索の手法に転換した。そして、不安定領域を検索した結果と Richardson パラメータを照らし合わせて、スコアが低くなった原因のアミノ酸残基に対して、スコアが高くなるようにアミノ酸置換をする手法を考えた。つまり、二次構造の不安定領域の検索結果から、変異ポイントを見つけ、アミノ酸置換をするという、二次構造の設計に必要な手法を考えた。この手法を実際の変異蛋白質のデータに対して当てはめたところ、約 70%の蛋白質が、熱に対する安定性の向上が見られた結果と一致した。つまり、この研究による二次構

造の末端予測法が、構造既知の蛋白質に適用されたとき、二次構造を設計するための手法として使える可能性があることが示唆された。つまり、ProSSA 法を使って算出された Richardson パラメータを利用することで、二次構造の末端を予測でき、なおかつ二次構造の設計にも応用できる方法を提案できた。よって、ProSSA 法は、予測だけでなく、設計にも有用であったことがわかった。

第4章では、蛋白質 CD スペクトルから二次構造含量を推定する研究に対する ProSSA 法の有用性を検証した。その結果、25 種類の蛋白質の CD スペクトルのうち、20 種類が 190 nm 付近の極値で二次構造含量との相関があった。この相関があった CD スペクトルと対照蛋白質の CD スペクトルから、二次構造含量が 100% と見なした参照 CD スペクトルを算出し、そのデータをもとに、構造未知の蛋白質の CD スペクトルから二次構造含量を推定した。その結果、推定した二次構造含量と PDB ファイルから ProSSA 法を使って算出した二次構造含量に相関が見られた。また、このときの CD スペクトルの波長領域は、遠紫外領域である。つまり、真空紫外領域で測定できる特殊な装置ではなく、普及している遠紫外領域で蛋白質の CD スペクトルを測定できる CD 測定装置を有効に活用できる。つまり、ProSSA 法は、遠紫外領域での蛋白質の CD スペクトルから二次構造含量を推定するのにも有用であった。また、CD スペクトルのデータベースを構築した。このデータベースは、実際に測定した CD スペクトルデータや測定条件だけでなく、過去に論文に掲載された CD データや測定条件も集め、多くの研究者・技術者が参考できるようにした。

第5章では、プロテアーゼが切断する配列と基質の二次構造の関係を ProSSA 法を使って調べた。その結果、プロテアーゼは、蛋白質の表面だけでなく、 β ストランド構造の N 末端の境界領域に存在する認識配列を切断していた。DSSP 法では、切断配列の前後に、二次構造が存在せず、二次構造と切断配列に相関が見られなかった。この結果は、プロテアーゼが蛋白質のどの部位を切断するかを予測するときに役立てることができる。したがって、ProSSA 法がプロテアーゼの研究に有用であることを証明した。

第6章では、ProSSA 法によって帰属された β ストランド構造を、シート形成 β ストランド構造と一本鎖 β ストランド構造の二つに分類して、統計的な解析をした。その結果、シート形成

β ストランド構造は、偶数残基ごとに長さの最大値が見られ、比較的両親媒性の構造を取ることを明らかにした。さらに、一本鎖 β ストランド構造は、シート形成 β ストランド構造のように、長さに対して特徴が見られなかったが、アラニル・アラニンは、シート形成 β ストランド構造よりも一本鎖 β ストランド構造の中の方が、エネルギー的に安定であったことを明らかにした。また、 β ストランド構造の中のアミノ酸残基の主鎖二面体角の出現頻度に見られる複数の最大値は、シート形成と一本鎖の β ストランド構造の中のアミノ酸残基の主鎖二面体角の出現頻度の最大値に対応していた。また、主鎖二面体角の出現頻度は異なるものの、シート形成と一本鎖の β ストランド構造には、共通する主鎖二面体角が存在した。さらに、バリン、プロリン、イソロイシンは、シート形成と一本鎖に関係なく、主鎖二面体角の最大値がほぼ同じだった。つまり、これら三つのアミノ酸残基は、伸びた構造に対して、特定の主鎖二面体角を取り得ると考えた。

第7章では、蛋白質の二次構造を解析するときの基本単位をジペプチドとし、統計的な解析を行なった。その結果、 α ヘリックス構造、 β ストランド構造、シート形成 β ストランド構造、一本鎖 β ストランド構造において、配列を考慮した傾向指数を得ることができた。たとえば、 α ヘリックス構造のジペプチド AC の傾向指数は、1.3 だったのに対して、CA の傾向指数は 1.1 だった。また、 α ヘリックス構造と β ストランド構造の主鎖二面体角 (ψ , ϕ) の分布において、ある角度へ収束する場合と収束しない場合の二つに分類できた。シート形成 β ストランド構造のジペプチドは、 β ストランド構造のジペプチドのうち約 51% だった。アミノ酸残基の傾向指数で 1.3 以上をもつもの同士のジペプチドの傾向指数は、すべて 1.5 以上を示した。また、ジペプチドを構成しているアミノ酸残基のうち、片方のアミノ酸残基の傾向指数が 1.8 を超えている場合は、ほとんどのジペプチドの傾向指数が 1.5 より高い値を示した。アミノ酸残基の傾向指数が 1.0 より低い値をもつもの同士のジペプチドの傾向指数は、1.0 を超えていなかった。一本鎖 β ストランド構造は、プロリンを含むジペプチドは、PE, PG, DP を除いて、1.5 以上の傾向指数を示した。また、プロリンが関わらないジペプチドの傾向指数が 1.5 以上の値を示したものは、すべて 1.0 前後のアミノ酸残基の組み合わせだった。 α ヘリックス構造や β ストランド構造のジペプチドによる Richardson パラメータは、 α ヘリックス構造の N''N' ポジション

以外は、すべて 1.0 に収束しない分布を示した。 α ヘリックス構造の N'N'ポジションも、1.0 に収束しているものの、偏りが見られた。また、ほとんどのポジションにおいて、ジペプチドの傾向指数の最大と最小は、アミノ酸残基の傾向指数でも最大や最小、もしくは、それらに近い値をもつアミノ酸残基の組み合わせだった。また、アミノ酸残基の傾向指数が 1.0 以下同士の組み合わせにおいて、ジペプチドの傾向指数が 1.5 以上のものが存在した。一方で、アミノ酸残基の傾向指数が 1.0 以上の同士の組み合わせで、ジペプチドの傾向指数が 0.5 以下のものが存在した。つまり、ジペプチド単位にしたことで、アミノ酸残基の組み合わせからは得られない情報を得たといえる。さらに、二次構造の中のアラニンの主鎖二面体角をアラニル・アラニンの主鎖二面体角 (ψ , ϕ) に使って、それぞれの二次構造に対する自由エネルギーを算出した。その結果、 α ヘリックス構造の中のアラニル・アラニンは、 β ストランド構造の中よりエネルギー的に安定であり、 β ストランド構造の中では、完全に伸びた構造の中よりエネルギー的に安定だった。

以上、この研究では主鎖二面体角の連続性パターンによる蛋白質の二次構造帰属法 ProSSA がどのような研究に適用できるかを調べた。その結果、二次構造の末端の予測と設計、蛋白質 CD スペクトルからの二次構造含量の推定、蛋白質の二次構造とプロテアーゼの切断部位の関係などの研究に応用できることを示した。また、二次構造を解析する上での単位パラメータとして、ジペプチドを提案し、アミノ酸残基の組み合わせから得られない情報を得た。ただし、ジペプチドについては、20 種類から 400 種類へ情報量が増加したため、複雑化した。しかしながら、アミノ酸残基からは得られない情報を得たのも事実である。

この研究が、蛋白質工学の発展に何らかの影響を与えることができれば幸いである。

参考文献

- 1) International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931-945.
- 2) Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer Jr, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535-542.
- 3) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Res.*, **28**, 235-242.
- 4) Murzin, A. G., Brenner S. E., Hbbard, T. and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536-540.
- 5) Hubbard, T. J., Ailey, B., Brenner, S. E., Murzin, A. G. and Chothia, C. (1999). SCOP: a structural classification of proteins database. *Nucleic Acids Res.*, **27**, 254-256.
- 6) Conte, L. L., Ailey, B., Hubbard, T. J. P., Brenner, S. E., Murzin, A. G. and Chothia, C. (2000). SCOP: a structural classification of proteins database. *Nucleic Acids Res.*, **28**, 257-259.
- 7) Conte, L. L., Brenner, S. E., Hubard, T. J. P., Chothia, C. and Murzin, A. G. (2002). SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Res.*, **30**, 264-267.
- 8) Andreeva, A., Howorth, D. Brenner, S. E., Hubbard, T. J. P., Chothia, C. and Murzin, A. G. (2004). SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res.*, **32**, D226-D229.
- 9) Pauling, L. and Corey, R. B. (1950). Two hydrogen-bonded spiral configurations of the polypeptide chain. *J. Am. Chem. Soc.*, **72**, 5349.
- 10) Kabsch, W. and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.
- 11) 金光和保. (2003). 平成 15 年度九州工業大学大学院博士論文.

- 12) Greenfield, N. and Fasman, G. D. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*, **8**, 4108-4115.
- 13) Brahms, S. and Brahms, J. (1980). Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J. Mol. Biol.*, **138**, 149-178.
- 14) Hennessey, J. P., Jr. and Johnson, W. C., Jr. (1981). Information content in the circular dichroism of proteins. *Biochemistry*, **20**, 1085-1094.
- 15) Provencher, S. W. and Glockner, J. (1981). Estimation of globular protein secondary structure from circular dichroism. *Biochemistry*, **20**, 33-37.
- 16) Compton, L. A. and Johnson, W. C., Jr. (1986). Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.*, **155**, 155-167.
- 17) Manavalan, P. and Johnson, W. C., Jr. (1987). Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.*, **167**, 76-85.
- 18) Perczel, A., Park, K. and Fasman, G. D. (1992). Analysis of the circular dichroism spectrum of proteins using the convex constraint algorithm: a practical guide. *Anal. Biochem.*, **203**, 83-93.
- 19) Andrade, M. A., Chacon, P., Merelo, J. J. and Moran, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng.*, **6**, 383-390.
- 20) Sreerama, N. and Woody, R. W. (1994). Protein secondary structure from circular dichroism spectroscopy. *J. Mol. Biol.*, **242**, 497-507.
- 21) Johnson, W. C. (1999). Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins*, **35**, 307-312.
- 22) Toumadje, A., Alcorn, S. W. and Johnson, W. C., Jr. (1992). Extending CD spectra of proteins to 168 nm improves the analysis for secondary structures. *Anal. Biochem.*, **200**, 321-331.
- 23) Matsuo, K., Yonehara, R. and Gekko, K. (2004). Secondary-structure analysis of proteins by vacuum-ultraviolet circular dichroism spectroscopy. *J. Biochem.*, **135**, 405-411.
- 24) Matsuo, K., Yonehara, R. and Gekko, K. (2005). Improved estimation of the secondary structures

- of proteins by vacuum-ultraviolet circular dichroism spectroscopy. *J. Biochem.*, **138**, 79-88.
- 25) Gekko, K. and Matsuo, K. (2006). Vacuum-ultraviolet circular dichroism analysis of biomolecules. *Chirality*, **18**, 329-334.
 - 26) Chou, P. Y. and Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.*, **47**, 45-148.
 - 27) Dasgupta, S. and Bell, J. A. (1993). Design of helix ends. *Int. J. Pept. Protein Res.*, **41**, 499-511.
 - 28) Seale J. W., Srinivasan, R. and Rose, G. D. (1994). Sequence determinants of the capping box, a stabilizing motif at the N-termini of α -helices. *Protein Sci.*, **3**, 1741-1745.
 - 29) Zhu, Z.-Y. and Blundell, T. L. (1996). The use of amino acid patterns of classified helices and strands in secondary structure prediction. *J. Mol. Biol.*, **260**, 261-276.
 - 30) Aurora, R. and Rose, G. D. (1998). Helix capping. *Protein Sci.*, **7**, 21-38.
 - 31) Steward, R. E. and Thornton, J. M. (2002). Prediction of strand pairing in antiparallel and parallel β -sheets using information theory. *Proteins*, **48**, 178-191.
 - 32) Eswar, N., Ramakrishnan, C. and Srinivasan, N. (2003). Stranded in isolation: structural role of isolated extended strands in proteins. *Protein Eng.*, **16**, 331-339.
 - 33) Penel, S., Morrison, R. G., Dobson, P. D., Mortishire-Smith, R. J. and Doig, A. J. (2003). Length preferences and periodicity in β -strands. Antiparallel edge β -sheets are more likely to finish in non-hydrogen bonded rings. *Protein Eng.*, **16**, 957-961.
 - 34) Engel, D. E. and DeGrado, W. F. (2004). Amino acid propensities are position-dependent throughout the length of α -helices. *J. Mol. Biol.*, **337**, 1195-1205.
 - 35) Costantini, S., Colonna, G. and Facchiano, A. M. (2006). Amino acid propensities for secondary structures are influenced by the protein structural class. *Biochem. Biophys. Res. Commun.*, **342**, 441-451.
 - 36) 総務省情報通信制作局. (2005). 平成17年度「情報通信白書」
 - 37) Kawabata, T., Ota, M. and Nishikawa, K. (1999). The protein mutant database. *Nucleic Acids*

- Res.*, **27**, 355-357.
- 38) Gromiha, M. M., An, J., Kono, H., Oobatake, M., Uedaira, H. and Sarai, A. (1999). ProTherm: thermodynamic database for proteins and mutants. *Nucleic Acids Res.*, **27**, 286-288.
 - 39) Gromiha, M. M., An, J., Kono, H., Oobatake, M., Uedaira, H., Prabakaran, P. and Sarai, A. (2000). ProTherm, version 2.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.*, **28**, 283-285.
 - 40) Gromiha, M. M., Uedaira, H., An, J., Selvaraj, S., Prabakaran, P. and Sarai, A. (2002). ProTherm, thermodynamic database for proteins and mutants: developments in version 3.0. *Nucleic Acids Res.*, **30**, 301-302.
 - 41) Bava, K. A., Gromiha, M. M., Uedaira, H., Kitajima, K. and Sarai, A. (2004). ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.*, **32**, D120-D121.
 - 42) IUPAC-IUB. (1971). IUPAC-IUB commission on biochemical nomenclature. Abbreviations and symbols for the description of the conformation of polypeptide chains. Tentative rules (1969). *Biochem. J.*, **121**, 577-585.
 - 43) Parthasarathy, S. and Murthy, M. R. N. (2000). Protein thermal stability: insights from atomic displacement parameters (B values). *Protein Eng.*, **13**, 9-13.
 - 44) Schlessinger, A., Yachdav, G. and Rost, B. (2006). PROFbval: predict flexible and rigid residues in proteins. *Bioinformatics*, **22**, 891-893.
 - 45) Chou, P. Y. and Fasman, G. D. (1974). Prediction of protein conformation. *Biochemistry*, **13**, 222-245.
 - 46) Argos, P., Nanei, M. and Garavito, R. M. (1978). The chou-fasman secondary structure prediction method with an extended data base. *FEBS Lett.*, **93**, 19-24.
 - 47) Figureau, A., Soto, M. A. and Toha, J. (2003). A pentapeptide-based method for protein secondary structure prediction. *Protein Eng.*, **16**, 103-107.
 - 48) Kyngas, J. and Valjakka, J. (1998). Unreliability of the Chou-Fasman parameters in predicting protein secondary structure. *Protein Eng.*, **11**, 345-348.

- 49) Richardson, J. S. and Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of a helix. *Science*, **240**, 1648-1652.
- 50) Harper, E. T. and Rose, G. D. (1993). Helix stop signals in proteins and peptides: the capping box. *Biochemistry*, **32**, 7605-7609.
- 51) Software Development Co.,Ltd. (1998). GENETYX-MAC 遺伝情報処理ソフトウェア Ver. 10.
- 52) Hiraga, K. and Yutani, K. (1996). Study of cysteine residues in the α subunit of *Escherichia coli* tryptophan synthase. 1. Role in conformational stability. *Protein Eng.*, **9**, 425-431.
- 53) Pinker, R. J., Lin, L., Rose, G. D. and Kallenbach, N. R. (1993). Effects of alanine substitutions in α -helices of sperm whale myoglobin on protein stability. *Protein Sci.*, **2**, 1099-1105.
- 54) Zhi, W., Srere, P. A. and Evans, C. T. (1991). Conformational stability of pig citrate synthase and some active-site mutants. *Biochemistry*, **30**, 9281-9286.
- 55) Nicholson, H. Anderson, D. E., Dao-pin, S. and Matthews, B. W. (1991). Analysis of the interaction between charged side chains and the α -helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry*, **30**, 9816-9828.
- 56) Chen, Y.-H. and Yang J. T. (1971). A new approach to the calculation of secondary structures of globular proteins by optical rotatory dispersion and circular dichroism. *Biochem. Biophys. Res. Commun.*, **44**, 1285-1291
- 57) Yang, J. T., Wu, C.-S. C. and Martinez, H. M. (1986). Calculation of protein conformation from circular dichroism. *Methods Enzymol.*, **130**, 208-269.
- 58) CDsstr. <http://oregonstate.edu/dept/biochem/faculty/johnson.html>
- 59) JASCO Corporation. (1994) SSE-338W 型タンパク質二次構造解析プログラム. JASCO J-700 for Windows optional software.
- 60) Lobley, L., Whitmore, L. and Wallace, B. A. (2002). DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra.. *Bioinformatics*, **18**, 211-212.

- 61) Whitmore, L. and Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.*, **32**, W668-W673.
- 62) Whitmore, L., Janes, R. W. and Wallace, B. A. (2006). Protein circular dichroism data bank (PCDDDB): data bank and website design. *Chirality*, **18**, 426-429.
- 63) Wallace, B. A., Whitmore, L. and Janes, R. W. (2006). The protein circular dichroism data bank (PCDDDB): a bioinformatics and spectroscopic resource. *Proteins*, **62**, 1-3.
- 64) Hartleib, J. and Ruterjans, H. (2001). High-yield expression, purification, and characterization of the recombinant diisopropylfluorophosphatase from *Loligo vulgaris*. *Protein Expr. Purif.*, **21**, 210-219.
- 65) Buczek, O., Kroowarsch, D. and Otlewski, J. (2002). Thermodynamics of single peptide bond cleavage in bovine pancreatic trypsin inhibitor (BPTI). *Protein Sci.*, **11**, 924-932.
- 66) Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. and Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365-370.
- 67) Sternberg, M. J. E. and Thornton, J. M. (1977). On the conformation of proteins: An analysis of β -pleated sheets. *J. Mol. Biol.*, **110**, 285-296.
- 68) Cubellis, M. V., Caillez, F., Blundell, T. L. and Lovell, S. C. (2005). Properties of polyproline II, a secondary structure element implicated in protein-protein interactions. *Proteins*, **58**, 880-892.
- 69) Kang, Y. K., Jhon, J. S. and Han, S. J. (1999). Conformational study Ac-Xaa-Pro-NHMe dipeptides: proline puckering and trans/cis imide bond. *J. Pept. Res.*, **53**, 30-40.
- 70) Grail, B. M. and Payne, J. W. (2000). Predominant torsional forms adopted by dipeptide conformers in solution: Parameters for molecular recognition. *J. Pept. Sci.*, **6**, 186-199.
- 71) Narita, M., Sode, K., Ohuchi, S., Murakawa, Y. and Hitomi, M. (1998). Statistical characterization of eleven kinds of helix elements with amino acid residues in the middle of triplets. *Bull. Chem. Soc. Jpn.*, **71**, 385-395.

- 72) Narita, M., Mochizuki, A. and Ohuchi, S. (2000). Assignments of tri- and tetrapeptide sequences in globular proteins to the 18 kinds of local structures along helices and their propensities for specific local structures. *Bull. Chem. Soc. Jpn.*, **73**, 1379-1387.
- 73) Marshall, N. J., Grail, B. M. and Payne, J. W. (2001). Predominant torsional forms adopted by oligopeptide conformers in solution: Parameters for molecular recognition. *J. Pept. Sci.*, **7**, 175-189.
- 74) Yang, J., Dong, X.-C. and Leng, Y. (2006) Conformation biases of amino acids based on tripeptide microenvironment from PDB database. *J. Theor. Biol.*, **240**, 374-384.

謝辞

この研究を進めるにあたり、終始指導してくださいました大内将吉先生には、研究を始めてから博士論文の完成までの間、研究に対する取り組み方など、さまざまな助言、激励をいただき、ありがとうございました。

また、金光和保さんには、研究やそれ以外のことについての相談を聞いてくださりありがとうございました。また、大塚康彦君は、同期として、いろいろな相談を聞いてくれてありがとうございました。

第2章では、ProSSA法のデータベースの構築に際し、ProSSA法を提案した金光和保さんにお世話になりました。第3章では、 α ヘリックス構造末端の予測とデザインの研究を進めた古賀功君に感謝いたします。第4章では、蛋白質のCDスペクトルと二次構造含量の相関に関して、阿南敬三君にお世話になりました。第7章では、ジペプチドであるアラニル・アラニンの立体構造に対する自由エネルギーの算出には、入佐正幸先生のお世話になりました。

また、同期の石川浩嗣君、梅野学君、山根岩雄君、そして、修了生の佐藤貞典さん、入江広樹さん、阿部栄信君、そして、大内研究室の現役生、卒業生、修了生には、大変お世話になりました。

また、これらの研究の一部は、笹川研究助成からの支援を受けて取り組みました。お礼申し上げます。

また、学部時代に所属した西日本工業大学では、安部晴男先生、谷口良明先生に大変お世話になりました。

最後に、長い学生生活を支え、励ましてくれた両親と兄に感謝いたします。これからもよろしく願いいたします。

2008年3月 小副川博也