

ボツリヌス毒素関連タンパク質のホモロジーモデリング

近藤 洋介¹ 山井 太成¹ 権 娟大² 宮崎 智²

概要：ボツリヌス菌が産生する非毒非血球凝集素 (nontoxic nonhemagglutinin, NTNHA) は 2012 年 3 月にその結晶構造が初めて公開された。しかし、その構造の一部は欠損しており、NTNHA の全体構造の分析のためには、計算機を用いて NTNHA の構造を扱う手法を確立する必要がある。そこで、本研究では、その結晶構造が正解構造であるとして構造予測の精度の検証を行う。NTNHA は配列長が 1000 を超える巨大タンパク質のため、構造予測の手法としてホモロジーモデリングが妥当であると考えられる。実験的に構造が決定されたタンパク質の中で、ボツリヌス菌が産生する神経毒素は NTNHA の全長に対して約 40% の配列類似性を持っている。そこでそのボツリヌス毒素を基にして NTNHA の全体構造の予測を行ったが、NTNHA のドメイン構造間のフレキシビリティのために精度が低いと考えられた。そのため、NTNHA を 3 つのドメインへ分割しホモロジーモデリングを行ったところ精度が向上した。NTNHA の構造予測は、第 1 段階として、ドメイン構造への分割が有効になると考えられる。

キーワード：非毒非血球凝集素, タンパク質構造予測, 巨大タンパク質, ドメイン構造

Homology modeling of botulinum toxin associating protein

YOSUKE KONDO¹ TAKASHIGE YAMAI¹ YEONDAE KWON² SATORU MIYAZAKI²

Abstract: The crystal structure of nontoxic nonhemagglutinin (NTNHA) produced by *Clostridium botulinum* was released in March 2012 for the first time. However, because of the loss of a part of the structure, establishment of computational methods handling NTNHA structure is necessary to extend the analysis of a whole structure of NTNHA. Therefore, we conducted the verification of structure prediction method by the crystal structure mentioned above. Homology modeling is reasonable as a method of structure prediction because NTNHA is a large protein which exceeds 1000 in its sequence length. Among proteins whose structures are experimental-determined, a botulinum toxin has approximately 40% similarity with NTNHA across its whole sequence. Although prediction of the overall structure of NTNHA was based on the botulinum toxin, it was considered that flexibility between domain structures caused bad prediction accuracy. Then, the accuracy was improved when the homology modeling of NTNHA was performed by dividing into three domains. Division into domain structures can be valid for structure prediction of NTNHA as a first step.

Keywords: nontoxic nonhemagglutinin, protein structure prediction, large protein, domain structure

1. はじめに

1.1 タンパク質の配列と構造

タンパク質は複数のアミノ酸が連なる配列から成り、折

り畳まれて立体構造を形成する。そのため、各アミノ酸の配列上の距離と各アミノ酸の高次構造時の空間上の距離の間には、必ずしも正の相関があるわけではない。また、タンパク質間の相互作用を見積もるときには、タンパク質の折り畳みによって生じるアミノ酸残基の相対的な位置関係が重要である。したがって、タンパク質の立体構造上のアミノ酸間の距離や配置は、配列から得られる直線的な距離とは別の距離であることを十分に考慮する必要がある。

¹ 東京理科大学大学院薬学研究科
Graduate School of Pharmaceutical Sciences, Tokyo University of Science, Noda, Chiba 278-8510, Japan

² 東京理科大学薬学部生命創薬科学科
Department of Medical and Life Science, Faculty of Pharmaceutical Sciences, Tokyo University of Science

1.2 構造と配列のデータの統計

2012年10月のProtein Data Bank (PDB)の冗長なタンパク質の鎖に識別子が付けられたデータセット

`ftp://ftp.ncbi.nih.gov/mmdb/nrtable`

を取得し、その統計をとった。鎖長が20以上かつ100%の配列一致度を持つ鎖を1つにまとめると、鎖の総数は5万5597であった。一方、100%の配列一致度を持つタンパク質配列が1つにまとめられたUniRefデータベース[1]の、配列長が20以上の配列の総数は1871万8140であった。つまり、2012年10月現在、長さが20以上の観測されているタンパク質の配列の内、0.297%の立体構造が決定されているということになる。さらに、立体構造の決定手法も合わせて調べた。PDBのデータセットをX線結晶構造解析、核磁気共鳴(nuclear magnetic resonance, NMR)およびその他の手法にあらかじめ分割し、鎖長が20以上かつ100%の配列一致度を持つ鎖を1つにまとめると、鎖の総数はそれぞれ5万569, 5208および518であった。PDBの非冗長配列の立体構造の決定手法の占める割合は、X線結晶構造解析、NMRおよびその他の手法において、それぞれ9割弱、1割弱および1分弱ということになる。また、これらを足し合わせると5万6295となるため、同一の配列であっても、異なる手法によって決定された構造もPDBに存在していることがわかる。

PDBのデータセットについて、今度は、鎖長が1000以上かつ100%の配列一致度を持つ鎖を1つにまとめると、鎖の総数は283であった。一方、UniRefデータベースの配列長が1000以上の配列の総数は65万310であった。つまり、長さが1000以上の観測されているタンパク質の配列の内、0.044%の立体構造が決定されているということになる。したがって、配列長が1000以上のタンパク質の配列の観測はなされているが、その立体構造の決定率は上記の配列長が20以上の場合に比べて小さいことがわかる。また、配列長が20以上の立体構造の決定数に対する1000以上の立体構造の決定数は0.509%であるため、配列の決定数のその割合である3.474%よりも小さい。この理由の1つとして、巨大タンパク質の立体構造の決定は配列の決定よりも困難である可能性が考えられる。さらに、立体構造の決定手法も合わせて、先程と同様に調べた。PDBのデータセットをX線結晶構造解析、NMRおよびその他の手法にあらかじめ分割し、鎖長が1000以上かつ100%の配列一致度を持つ鎖を1つにまとめると、鎖の総数はそれぞれ261, 0および25であった。ここで、配列長が1000以上のタンパク質を巨大タンパク質と呼ぶことにすると、巨大タンパク質の立体構造の決定は、ほぼX線結晶構造解析により行われているのが現状であることがわかる。また、NMRは配列長が828以下のタンパク質に関しては強力な立体構造の決定手法として用いられているが、巨大タンパク質には応用されていない。

1.3 立体構造の決定と予測

ところで、タンパク質は、他分子との空間上の距離が小さくなることによってその機能を発揮する。したがって、タンパク質の機能発現には各アミノ酸の空間上の位置の方が配列上の位置よりも重要である。そこで、タンパク質の機能の分析には、タンパク質の配列を立体構造へ変換する手法が有効となってくる。その手法には、立体構造を実験的に決定する手法と計算機を用いて予測する手法の2つがある。巨大タンパク質に関しては、立体構造の決定率が配列の決定率に比べて小さいこと、決定手法がX線結晶構造解析にほぼ限られていることの2点を考えると、計算機を用いた構造予測が特に必要となってくるのが想定される。X線結晶構造解析によりタンパク質の結晶構造を得るには、高純度かつ大量のタンパク質の精製が必要である。このとき、タンパク質の結晶化の条件を変えて何度も繰り返すことが必要であるが、良質なタンパク質の結晶が得られれば、高精度のタンパク質の立体構造を得ることができる。その一方で、計算機を用いた構造予測は、計算機内に原子を配置することによって行われる。このとき、タンパク質の立体構造をほとんどの場合得ることができるので、予測精度の検証が必要となる。

2. 立体構造の予測手法

タンパク質の構造予測の手法には多くの手法があるが、本稿ではその中の3つを挙げる。

2.1 ab initio 法

タンパク質を構成する各原子の間には電気的な力が働いている。原子はその力を受けることで、原子の存在する位置によって原子の持つエネルギーは変化し、原子の持つ位置エネルギーが低くなるように原子は配置していく。この考え方を基に構造予測を行う手法がab initio法である。Ab initio法では、タンパク質を構成する全ての原子の間に働く電気的な力からタンパク質の持つ位置エネルギーを計算し、そのエネルギーが低くなる原子の配置を得る。原子の配置は計算機内のシミュレーションによっていくつも生成することができるが、原子数が多くなるほど最適な原子の配置を見つけることが難しくなる。そのため、構造予測のターゲットは基本的に小さなペプチドに限られる。また、計算量の削減のため、タンパク質の持つ位置エネルギーの計算には各原子の位置に関する経験的な関数が用いられる。そのため、構造予測の精度はその関数に依存することになる。

2.2 フラグメントアセンブリ法

長さが9くらいのフラグメントの構造を用いて、タンパク質の構造を組み上げる手法が開発されている[2]。この手法では、まず、構造予測のターゲットの配列のフラグメン

トから、構造データベース中の構造のフラグメントが探索される。次に、計算機内で伸びた鎖として定義されたターゲットの初期構造が、フラグメントの構造的な情報によって組み替えられる。さらに、配列-構造関数等のスコアリング関数によってターゲットの構造は数量化され、その関数を最適化するように原子を配置していくという手順で構造予測が行われる。この手法はフラグメントアセンブリ法と呼ばれ、ターゲットが新規のフォールドを持つ場合でも構造予測が可能とされている。しかし、複雑なフォールドを持つタンパク質では、フラグメントの探索法やスコアリング関数が不十分である等の理由で構造予測が困難になるという問題点がある。そのため、ターゲットが巨大タンパク質の場合は、考慮するフラグメントの数が増えることになりその全体構造の予測は精度が低下する可能性がある。

2.3 ホモロジーモデリング法

タンパク質のフォールドの数は配列の数よりも少ないことが推定されているため [3], 配列の多様性に対するフォールドの多様性は少ないと考えられる。タンパク質の進化を考えると、タンパク質は時間と共にその配列が変化していく。しかし、配列の変化に比べてフォールドの変化は少なく、タンパク質の配列が類似していれば構造も類似するという仮定が成立する。ホモロジーモデリングではこの仮定のもとに構造予測が行われ、配列類似性が高くかつ立体構造が実験的に決定されているタンパク質が鋳型として用いられる。このとき、鋳型が存在するという条件であれば、構造予測のターゲットが巨大タンパク質の場合もホモロジーモデリングの適用は可能である。しかし、配列類似性の高い鋳型が存在しない場合は構造予測の精度が低下する場合がある [4]。また、新規のフォールドを予測することは困難であり、予測構造は鋳型構造の持つフォールドとほとんど変わらない。

3. 研究材料

3.1 構造予測の精度の検証

本研究では、ボツリヌス毒素複合体のサブユニットの 1 つである非毒非血球凝集素 (nontoxic nonhemagglutinin, NTNHA) の構造予測の精度を検証する。NTNHA は、2012 年 3 月に、X 線結晶構造解析による解像度 2.70 Å の構造が公開されている [5]。そこで、この構造を NTNHA の正解の構造であるとして予測精度の検証を行う。

3.2 ボツリヌス毒素

ボツリヌス菌が産生する神経毒素は ~150 kDa の 1 本鎖のタンパク質として合成される [6]。その後、ボツリヌス菌が持つタンパク質分解酵素によって ~50 kDa の軽鎖と ~100 kDa の重鎖へと分解され、ジスルフィド結合によって連結した 2 本鎖構造となる。その 2 本鎖の分子は、各々

が ~50 kDa である結合ドメイン、チャネル形成ドメインおよび酵素活性ドメインの 3 つの機能ドメインから成る。結合ドメインは重鎖の C 末端側 (H_C) のドメインであり、神経細胞の膜受容体への結合に関わる。この結合により、毒素はエンドサイトーシスで細胞内に取り込まれ、エンドソーム内に入る。ここで、毒素はジスルフィド結合を切り離し、軽鎖 (light chain, LC) の酵素活性ドメインが解離する。一方、チャネル形成ドメインは重鎖の N 末端側 (H_N) のドメインであり、エンドソームの膜に孔を開ける。この孔からエンドペプチダーゼ活性を持っている酵素活性ドメインが細胞質内に移行し、神経伝達物質を内包した小胞の細胞膜への融合に関わるタンパク質を分解する。これによって神経の伝達は遮断され、ボツリヌス毒素は毒性を発揮する。

3.3 ボツリヌス毒素複合体

ボツリヌス毒素はボツリヌス菌が産生する他のタンパク質と複合体を形成している [7]。NTNHA と分子量の異なる 3 種類の血球凝集素がそのサブユニットである。さらに、その複合体には、各サブユニットの構成が異なる 300 kDa の M 毒素、500 kDa の L 毒素および 900 kDa の LL 毒素が観測されている。ボツリヌス菌は産生する毒素の抗原性の相違により A~G の 7 つの血清型に分類されており、血清型によって産生するボツリヌス毒素複合体は異なる。M 毒素はボツリヌス毒素と NTNHA の複合体であり、A~F 型が産生する。L 毒素は M 毒素と血球凝集素の複合体であり、A~D と G 型が産生する。LL 毒素は L 毒素が重合したものと考えられており、A 型が産生する。

ところで、経口的に摂取したタンパク質は、ペプシンを含む強酸性の胃液、続いて十二指腸、小腸に流れ込む過程で様々なタンパク質分解酵素による分解や変性を受ける。ボツリヌス毒素もその複合体もタンパク質であり、経口的に摂取した場合、同様に分解や変性を受けると考えられる。ところが、経口摂取の際に、ボツリヌス毒素複合体ではボツリヌス毒素のみのときよりも、ボツリヌス毒素の毒性が高いことがマウスを用いた実験によってわかっている [8]。ここから考えられることは、ボツリヌス毒素に結合するタンパク質はボツリヌス毒素の運搬体として働くということであるが、その運搬のメカニズムはまだよくわかっていない。

3.4 ボツリヌス毒素複合体の構造

ボツリヌス毒素複合体の電子顕微鏡像から、L 毒素は 3 本腕の 14 量体の構造であるという仮説が立てられている [9]。そのため、この特異的な複合体の構造がボツリヌス毒素の運搬に関わると考えられており、この複合体の立体構造の解明が望まれている。各サブユニットの構造は、異なる血清型のボツリヌス菌において決定された構造があ

る。また、各サブユニットの配列は、異なる血清型間において相同であると考えることができる。このため、ホモロジーモデリングにより理論的には全ての血清型の各サブユニットの構造を得ることが可能である。各サブユニットの中で、今までどの血清型においても決定されていなかった NTNHA の構造は、2012 年 3 月に A 型 [5]、2012 年 9 月に D 型 [10] が公開された。しかし、その両方の NTNHA は一部の構造が欠損しており、その部位の中には血球凝集素との結合に関わる部位が含まれていると推定されている。このため、ボツリヌス毒素複合体の構造の分析には、計算機を用いて NTNHA の構造を扱う手法の確立が必要となる。そこで、本研究では、2012 年に決定された NTNHA の構造がないという前提で、NTNHA の構造予測を行う。NTNHA は配列長が 1150~1200 の巨大タンパク質であるため、前述のことを考慮すると、構造予測の手法はホモロジーモデリングが妥当であると考えられる。

4. NTNHA のホモロジーモデリング

4.1 方法

本研究には東京理科大学高速並列計算機システムを利用し、タンパク質のモデリングのソフトウェアには、グラフィカルユーザインターフェースの操作を可能とする Discovery Studio 3.1 (以下 DS と記載) [11] を用いた。ホモロジーモデリングの精度の検証のため、立体構造が 2012 年 3 月に公開された A 型のボツリヌス菌の NTNHA が含まれる配列データを PDB から取得した (PDBID: 3V0A)。その配列データから NTNHA の鎖である B 鎖の残基番号 3~1196 を抽出し、その配列をクエリとして PSI-BLAST 検索 [12] を行った。このとき、DS のローカルデータベースである PDB_nr95 を参照データベースとした。PDB_nr95 は、PDB の 95% 以上の配列一致度を持つ配列を 1 つにまとめ、X 線結晶構造解析の解像度等に基づいて代表の配列を選定したデータベースである。また、PDB_nr95 は更新されていないため、2012 年になってから公表された NTNHA のエントリーは含まれていない。PSI-BLAST 検索の結果、B 型のボツリヌス菌が持つ神経毒素 (PDBID: 1S0E) [13] のみが NTNHA の全長に渡ってアライメントされていた。そのため、この B 型のボツリヌス毒素を鋳型として用いた。この PSI-BLAST 検索のターゲットと鋳型の配列一致度は 19.2% であり、配列類似性は 42.0% であった。ホモロジーモデリングの精度の向上のためには、構造予測のターゲットと鋳型のペアワイズアライメントを正確に行うことが必要である。ところが、このときの配列類似性は低く、両配列のみを比較しただけでは正確なペアワイズアライメントを得ることができない可能性が考えられた。そこで、ターゲットの A 型の NTNHA と配列類似性の高い配列を収集しマルチプルアライメントを行った。PDB から取得している A 型の NTNHA の配列をクエリと

して、米国の国立生物工学情報センター (national center for biotechnology information) の非冗長タンパク質配列データベースに BLAST 検索をかけた。その結果、E 値が 10^{-5} 以下かつアライメント長が 1000 以上であるヒットは 92 配列であった。これらの配列にターゲットと鋳型の配列を合わせ、合計 94 配列に対して DS のマルチプルアライメントプロトコルを実行した。このマルチプルアライメントからターゲットと鋳型の 2 つの配列を抽出し、両配列のペアワイズアライメントを得た。このときの配列一致度は 17.3% であり、配列類似性は 40.2% であった。このペアワイズアライメントと B 型のボツリヌス毒素の構造を用いて、NTNHA のホモロジーモデリングを行った。DS はターゲットと鋳型のペアワイズアライメントに基づいて原子の空間上の位置を決める。アライメント上でターゲットと鋳型の原子が同様である場合は、鋳型構造の対応する原子の位置にターゲットの原子を置く。それ以外の原子は CHARMm 残基トポロジーライブラリーの内部座標を当てはめていき、ターゲットの初期構造を作成する。さらに、4 Å の間でランダムに原子を動かし、確率密度関数 (probability density function, PDF) を計算する。PDF は分子の中の原子の幾何学的な配置、例えば、結合長、結合角、二面角や非結合距離を記述する関数のことである [14]。PDF の最適化を行うように原子を配置していくことによって、ターゲットの構造予測を行う。

4.2 結果と考察

NTNHA の予測構造と結晶構造の比較のため、DS の構造アライメントプロトコルを用いて両構造を重ね合わせた。このとき、両構造の $C\alpha$ 原子の重心を合わせ、平均二乗偏差 (root mean square deviation, RMSD) が最小となるように回転させた。RMSD は式 (1) により定義される。

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n \|a_i - b_i\|^2} \quad (1)$$

ここで、 a_i と b_i は構造中の i 番目の原子の位置ベクトル、 n は総原子数であり、RMSD は対応する原子間の二乗距離の相加重平均の平方根を表す。図 1 は NTNHA の予測構造と結晶構造を重ね合わせた構造を示している。図 1 の予測構造と結晶構造の間の最も大きな相違は、 nH_C ドメインの方向であるように見える。両構造の nH_C ドメインは互いにクロスしたような構造となっている。表 1 は図 1 の NTNHA の全体を重ね合わせた構造の $C\alpha$ 原子間の RMSD を示している。さらに、表 1 には、図 1 の重ね合わせた構造を各ドメインに分けたときの RMSD の計算結果も合わせて示している。表 1 の RMSD が 0 に近づくほど、予測構造と結晶構造の $C\alpha$ 原子の 3 次元空間上の配置がよく重なり合うことを示す。その一方で、RMSD は対応する $C\alpha$ 原子間の距離の中に外れ値が含まれていた場合により大きな値と

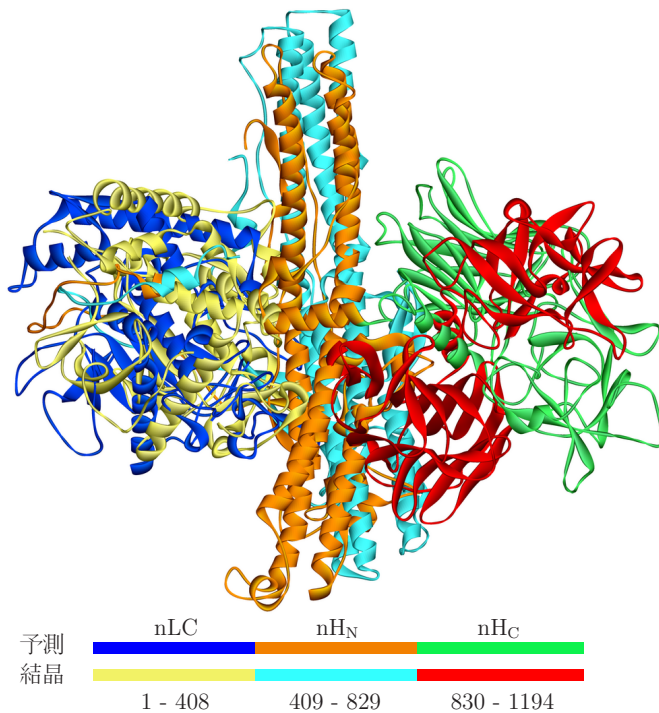


図 1 NTNHA の全体構造の重ね合わせ

表 1 全体構造の重ね合わせの RMSD

構造	RMSD (Å)
全体	25.609
nLC	17.042
nHN	19.638
nHC	36.458

なる。例えば、原子 A, B, C と原子 A', B', C' の RMSD の計算を考える。A-A', B-B', C-C' 間の距離がそれぞれ 2, 2, 2 のとき、RMSD は、 $\sqrt{\frac{2^2+2^2+2^2}{3}} = \sqrt{4} = 2.000$ である。一方、これらの距離がそれぞれ 1, 1, 4 のとき、RMSD は、 $\sqrt{\frac{1^2+1^2+4^2}{3}} = \sqrt{6} \approx 2.449$ である。つまり、全原子間の距離の合計が 6 の等距離にあるとき、全ての原子のペア間の距離が均等である場合より、距離に外れ値が含まれる原子のペアがある場合の方が RMSD は大きくなる。このため、RMSD を大きくする要因は、原子間の距離の外れ値の積み重ねや大きな外れ値がある場合が考えられる。表 1 において、nHC ドメインの RMSD が他のドメインと比較して約 2 倍となっていることから、nHC ドメインは他のドメインと比較して重なり合っていないことがわかる。NTNHA の全体構造を重ね合わせたときに、C α 原子数の多い nLC ドメインと nHN ドメインを重ね合わせた方が全体の C α 原子間の距離の外れ値が少なくなるため、nHC ドメインが不一致となったのでありと考える。

予測構造と結晶構造の nHC ドメインの不一致の理由として、ボツリヌス毒素の H_C ドメインは、NTNHA と複合体を形成しているときあるいは単体で存在するとき、その結晶構造に相違があるという報告を考慮することができ

る [5]。そこで、DS の構造アライメントプロトコルを用いて、B 型のボツリヌス毒素の単体のときの構造 (PDBID: 1S0E) または複合体を形成しているときの A 型のボツリヌス毒素の構造や NTNHA の構造 (PDBID: 3V0A) を重ね合わせた。図 2 と表 2 はこれらの結晶構造を重ね合わせた構造と RMSD の計算結果をそれぞれ示している。図 2 と表 2 からボツリヌス毒素は単体のときと複合体のときにおいて、両構造の構造間で大きな構造の変化があることがわかる。さらに、NTNHA とそれぞれの毒素の構造間の RMSD は、複合体の毒素と重ね合わせたときの方が、単体の毒素と重ね合わせたときよりも小さな値となっている。そのため、NTNHA の構造は、毒素が単体のときの構造よりも複合体を形成しているときの構造に近いことがわかる。本研究の正解構造として用いた NTNHA の結晶構造は、A 型のボツリヌス毒素と複合体を形成しているときの構造である。一方、鋳型として用いた構造は B 型のボツリヌス毒素の単体のときの構造である。したがって、本研究において予測された NTNHA の構造は、単体の毒素により近い構造が予測されており、複合体の NTNHA とは離れている構造であったと考えられる。さらに、X 線結晶構造解析により決定された NTNHA の構造と X 線溶液散乱法による溶液中での NTNHA の構造から、NTNHA の nHC ドメイン構造のフレキシビリティの推定がなされている [10]。これらのことから、NTNHA の予測構造と結晶構造の不一致の理由は、ボツリヌス毒素の H_C ドメインや NTNHA の nHC ドメインの構造のフレキシビリティのためであると考えられる。

5. ドメイン構造への分割

5.1 方法

上記の考察より、予測構造と結晶構造の NTNHA のドメイン構造の不一致の結果は、ドメイン構造間のフレキシビリティを考慮しなかったことが原因であると考えられた。そこで、NTNHA を各ドメインに分割しホモロジーモデリングを行った。前述の手順により得られた、ターゲットの A 型の NTNHA と鋳型の B 型のボツリヌス毒素のペアワイズアライメントを、LC ドメイン、HN ドメインおよび HC ドメインにそれぞれ分割した。さらに、鋳型の立体構造もそれぞれのドメインに分割し、DS を用いてホモロジーモデリングを行った。

表 2 結晶構造の重ね合わせの RMSD

構造 1	構造 2	RMSD (Å)
単体の毒素 (B 型)	複合体の毒素 (A 型)	26.599
単体の毒素 (B 型)	複合体の NTNHA	25.867
複合体の毒素 (A 型)	複合体の NTNHA	7.353

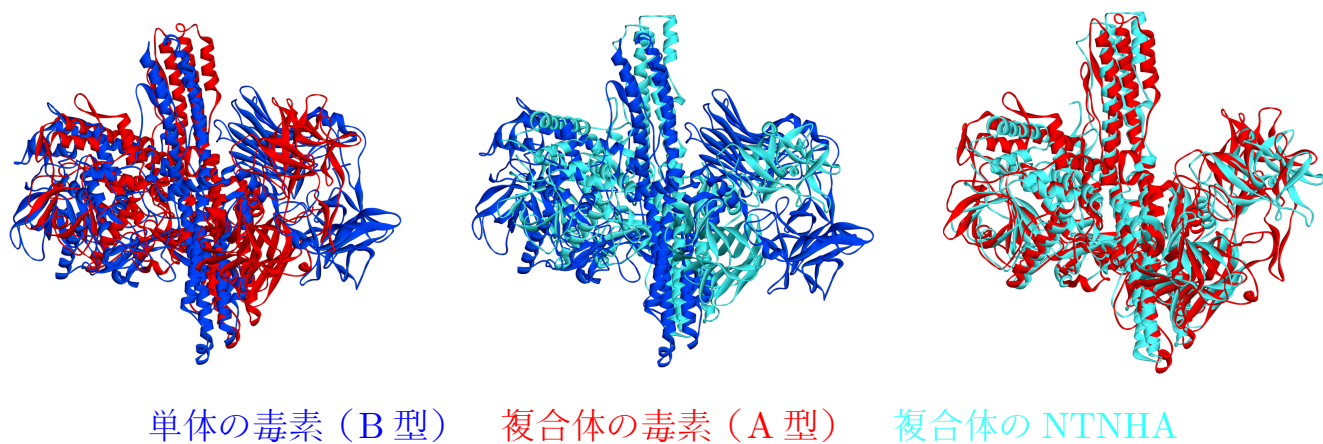


図 2 ボツリヌス毒素と NTNHA の結晶構造の重ね合わせ

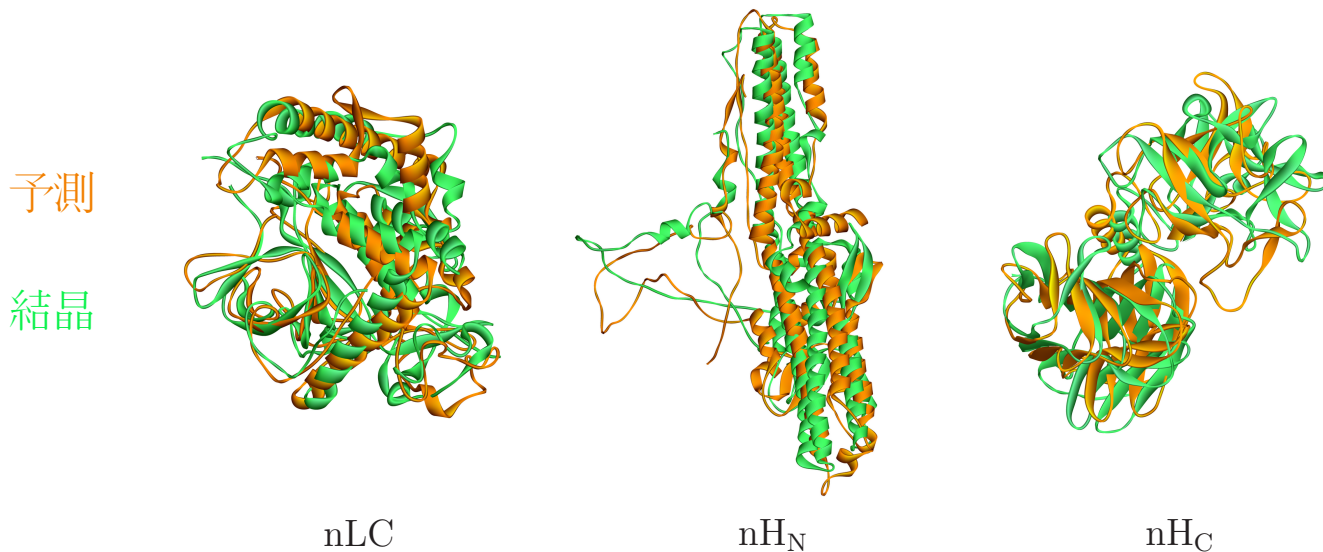


図 3 NTNHA の各ドメイン構造の重ね合わせ

表 3 各ドメイン構造の重ね合わせの RMSD

構造	RMSD (Å)
nLC	8.022
nHN	10.711
nHC	8.674

5.2 結果と考察

図 3 は DS の構造アライメントプロトコルを用いて、各ドメインに分割して予測した構造と各ドメインに分割した結晶構造を重ね合わせた構造を示している。表 3 は図 3 の RMSD の計算結果を示している。NTNHA の全体構造を重ね合わせたときの 25.609 Å と比べて、nLC、nHN および nHC ドメインの RMSD は、それぞれ 17.587 Å、14.898 Å および 16.935 Å 減少している。この結果は、予測構造と結晶構造の Cα 原子の配置は、各ドメイン構造を比較したときの方が全体構造を比較したときよりも、距離が大きく外れた Cα 原子が少なかったことを示していると考えられる。したがって、NTNHA の各ドメインのみの構造予測の方が全体の構造予測よりも、結晶構造と Cα 原子の配置

が類似した構造を予測していると考えられる。

6. 総括

6.1 NTNHA の構造予測

本研究のホモロジーモデリングのターゲットは配列長が 1194 の NTNHA であり、X 線結晶構造解析により決定された NTNHA の結晶構造を正解構造とみなして精度の検証を行った。このときの鋳型として、ターゲットの配列と全長に渡って約 40% の配列類似性が認められるボツリヌス毒素を用いた。予測構造と結晶構造を重ね合わせた結果、その RMSD は 25.609 Å であり両構造は類似していない構造であった。その理由について文献を調べると、ボツリヌス毒素や NTNHA のドメイン構造間のフレキシビリティが原因であると考えられた。そこで、NTNHA を 3 つのドメインに分割しホモロジーモデリングを行った。その結果、RMSD が各ドメインについて 14~18 Å 程度減少した。このため、NTNHA の構造予測の第 1 段階として、ドメイン構造への分割が有効な方法になり得ると考えられる。

6.2 巨大タンパク質の構造予測

ところで、本稿の冒頭において、巨大タンパク質の構造データの統計について述べた。その中で2番目に大きいタンパク質は鎖長が3367のダイニンであり、2012年3月に公開された解像度が3.80 Åの構造である(PDBID: 3VKH) [15]。ダイニンは細胞質内で微小管を形成しているタンパク質であり、その構造は8つのドメインから成る。その中のリンカードメインはさらにサブドメインに分けられており、サブドメイン間において構造の変化が推測されている。また、巨大タンパク質の構造データの中で5番目に大きいタンパク質は、鎖長が2512の脂肪酸合成酵素であり、2008年に公開された解像度が3.22 Åの構造である(PDBID: 2VZ8) [16]。ここで決定された構造は9つのドメインから成っており、5つの触媒ドメインが含まれている。このとき、このタンパク質のC末端側のドメインの構造は決定されていないが、このドメインはフレキシブルに動いて5つの触媒ドメインに脂肪酸を運搬する役割が推測されている。

これらのことから、NTNHA以外の巨大タンパク質においてもドメイン構造間のフレキシビリティが推定されることがわかる。このため、巨大タンパク質の構造予測はドメイン構造のフレキシビリティを含めて行うことが必要である。しかし、前述の通り、巨大タンパク質は現在用いられている *ab initio* 法やフラグメントアセンブリ法によって、その全体構造を予測することは困難である。したがって、巨大タンパク質の構造予測は、ターゲットをドメイン構造へ分割し、ホモロジーモデリング等の大きな鋳型を用いることができる構造予測の手法を用いて、ドメイン毎に予測することが第1のステップであると考えられる。また、例えば、配列長が1000のタンパク質の立体構造を予測するとき、その配列の全長に対してホモロジーモデリングを一度に行うことは危険である。なぜなら、本研究のNTNHAのように予測構造と結晶構造は類似していない構造である可能性が考えられるからである。そのため、巨大タンパク質のホモロジーモデリングはドメイン構造を意識して行うことが必要である。

参考文献

- [1] Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. and Wu, C. H.: UniRef: Comprehensive and Non-redundant UniProt Reference Clusters, *Bioinformatics*, Vol. 23, No. 10, pp. 1282–1288 (2007).
- [2] Simons, K., Kooperberg, C., Huang, E. and Baker, D.: Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences Using Simulated Annealing and Bayesian Scoring Functions, *J. Mol. Biol.*, Vol. 268, No. 1, pp. 209–225 (1997).
- [3] Liu, X., Fan, K. and Wang, W.: The Number of Protein Folds and Their Distribution over Families in Nature, *Proteins*, Vol. 54, No. 3, pp. 491–499 (2004).
- [4] Baker, D. and Sali, A.: Protein Structure Prediction and

- Structural Genomics, *Science*, Vol. 294, No. 5540, pp. 93–96 (2001).
- [5] Gu, S., Rumpel, S., Zhou, J., Strotmeier, J., Bigalke, H., Perry, K., Shoemaker, C. B., Rummel, A. and Jin, R.: Botulinum Neurotoxin Is Shielded by NTNHA in an Interlocked Complex, *Science*, Vol. 335, No. 6071, pp. 977–981 (2012).
 - [6] Lacy, D. and Stevens, R.: Sequence Homology and Structural Analysis of the Clostridial Neurotoxins, *J. Mol. Biol.*, Vol. 291, No. 5, pp. 1091–1104 (1999).
 - [7] Kouguchi, H., Watanabe, T., Sagane, Y., Sunagawa, H. and Ohyama, T.: *In Vitro* Reconstitution of the *Clostridium botulinum* Type D Progenitor toxin, *J. Biol. Chem.*, Vol. 277, No. 4, pp. 2650–2656 (2002).
 - [8] Ohishi, I., Sugii, S. and Sakaguchi, G.: Oral Toxicities of *Clostridium botulinum* Toxins in Response to Molecular Size, *Infect. Immun.*, Vol. 16, No. 1, pp. 107–109 (1977).
 - [9] Hasegawa, K., Watanabe, T., Suzuki, T., Yamano, A., Oikawa, T., Sato, Y., Kouguchi, H., Yoneyama, T., Niwa, K., Ikeda, T. and Ohyama, T.: A Novel Subunit Structure of *Clostridium botulinum* Serotype D Toxin Complex with Three Extended Arms, *J. Biol. Chem.*, Vol. 282, No. 34, pp. 24777–24783 (2007).
 - [10] Sagane, Y., Miyashita, S., Miyata, K., Matsumoto, T., Inui, K., Hayashi, S., Suzuki, T., Hasegawa, K., Yajima, S., Yamano, A., Niwa, K. and Watanabe, T.: Small-angle X-ray Scattering Reveals Structural Dynamics of the Botulinum Neurotoxin Associating Protein, Nontoxic Nonhemagglutinin, *Biochem. Biophys. Res. Commun.*, Vol. 425, No. 2, pp. 256–260 (2012).
 - [11] Accelrys Software Inc., Discovery Studio Modeling Environment, Release 3.1, San Diego: Accelrys Software Inc. (2011).
 - [12] Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.: Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs, *Nucleic Acids Res.*, Vol. 25, No. 17, pp. 3389–3402 (1997).
 - [13] Eswaramoorthy, S., Kumaran, D., Keller, J. and Swaminathan, S.: Role of Metals in the Biological Activity of *Clostridium botulinum* Neurotoxins, *Biochem.*, Vol. 43, No. 8, pp. 2209–2216 (2004).
 - [14] Sali, A. and Blundell, T. L.: Comparative Protein Modeling by Satisfaction of Spatial Restraints, *J. Mol. Biol.*, Vol. 234, No. 3, pp. 779–815 (1993).
 - [15] Kon, T., Oyama, T., Shimo-Kon, R., Imamula, K., Shima, T., Sutoh, K. and Kurisu, G.: The 2.8 Å Crystal Structure of the Dynein Motor Domain, *Nature*, Vol. 484, No. 7394, pp. 345–350 (2012).
 - [16] Maier, T., Leibundgut, M. and Ban, N.: The Crystal Structure of a Mammalian Fatty Acid Synthase, *Science*, Vol. 321, No. 5894, pp. 1315–1322 (2008).