

平成 22 年 5 月 1 日現在

研究種目： 基盤研究(B)
 研究期間： 2007 ～ 2009
 課題番号： 19300101
 研究課題名(和文) タンパク質フォールディングの大規模シミュレーションと大量軌道データ解析法の開発
 研究課題名(英文) Protein folding simulation on the large scale computer system

研究代表者

太田 元規 (OTA MOTONORI)
 名古屋大学・大学院情報科学研究科・教授
 研究者番号： 40290895

研究成果の概要(和文): コンピュータの性能は指数的に向上しており、少し前までは不可能だと思われていた大規模計算も数年たつと実現可能となっている。本研究ではタンパク質について大規模なフォールディングシミュレーションを実施した。その結果得られる大量データの処理についても新しい手法開発を行った。対象としたタンパク質は Trp-Cage と λ リプレッサである。開発した手法は、マルチプル軌道アラインメント法、部分オーダ軌道アラインメント法、遍歴プロフィール法、などである

研究成果の概要(英文): Because of the progress of computational technology, a large scale calculation that would be unrealistic a couple of years ago, becomes realistic nowadays. In this study, we conducted protein folding simulations on the large scale computers, and analyzed large amount of data produced by the simulations, using newly developed techniques. We simulated folding of Trp-Cage and λ repressor, and developed the multiple trajectory alignment, the partial order trajectory alignment, the itinerary profile.

交付決定額

(金額単位:円)

	直接経費	間接経費	合計
2007 年度	4,600,000	1,380,000	5,980,000
2008 年度	4,900,000	1,470,000	6,370,000
2009 年度	4,900,000	1,470,000	6,370,000
年度			
年度			
総計	14,400,000	4,320,000	18,720,000

研究分野: 総合領域

科研費の分科・細目: 情報学・生体生命情報学

キーワード: 分子動力学, シミュレーション, スーパーコンピュータ, 軌道解析技術

1. 研究開始当初の背景

タンパク質は実際にはミリ秒から秒でいどでフォールドし、固有の立体構造を保持するが、内部自由度が大きいため計算量は莫大となるので、そのフォールディング過程をコンピュータでシミュレーションすることは不可能だと思われていた。しかし、小さいが安定な立体構造を保持する人工タンパク質が設計され、必ずしも大きなタンパク質を扱う必要がなくなってきたこと、および、コ

ンピュータの処理能力が数年単位で倍加した結果、大規模計算が可能となったことで、系を選べばコンピュータシミュレーションが現実的なものとなってきた。

2. 研究の目的

いくつかのタンパク質について、スーパーコンピュータを利用した大規模フォールディングシミュレーションを実施し、フォールドする軌道を得る。

シミュレーションから得られた大量の軌道データを効率良く解析する新規手法を開発し、データに適用することでタンパク質のフォールディング機構についての知見を抽出する。

3. 研究の方法

(1) スーパーコンピュータを利用して大規模なタンパク質のフォールディングシミュレーションを実施する。シミュレーションのターゲットとしては Trp-Cage(20 残基)とスリプレッサー(80 残基)を選定した。Trp-Cage についてはこれまでに、水を露わには含まないシミュレーションを実施して、軌道の解析を行い、論文としてまとめ発表した。しかし、水を露わに含んだシミュレーションについては、実行していなかったため、これを行う。分子動力学のプログラムとしては、連携研究者の池口准教授が作成した MARBLE を利用する。スリプレッサーは80残基と大きなタンパク質であるが、マイクロ秒単位でフォールドする置換体が作成されており、シミュレーションでも速いフォールディングが実現するのであればフォールドする軌道が得られる可能性がある。このタンパク質については、水を露わに置いて計算を行うのは現実的ではないので、水を露わには扱わない方法でシミュレーションを行う。プログラムは AMBER を利用する。

(2) 軌道の解析を行う。大量軌道の大量情報処理に適した、新しい方法論を開発し適用する。マルチプル軌道アラインメント法、部分オーダー軌道アラインメント法、遍歴プロフィール法、などを考案し適用した。

4. 研究成果

(1) シミュレーションの実施

①20残基からなる小さなタンパク質、TrpCage について、水を露わに取り入れた分子動力学シミュレーションを、東工大のスパコングリッドマシン:TSUBAME 上で実行した。まず伸びた形のタンパク質をコンピュータの中に作製した。このタンパク質の末端から8Åの位置まで水を付加したところ、タンパク質全体は一边を88Åとする立方体の中にすっぽりと入った。系の総原子数は64,000個となった。この系を構造最適化した後に、圧力一定の分子動力学計算を80ps行い、単位ボックス周辺の水を緩和させた。この立体構造から初期速度をランダムに変更し、温度一定の分子動力学計算を60本、5ns実行した。タンパク質が十分まわり小さくなった軌道52本についてはボックスサイズを調整し64Åと小さくし、水も配置し直した。残りの8本は更に5nsの計算を行ったところサイズが小さくなったので、ボックスサイズを変更した。これらについては、合計が35nsになるまでシミュレーションを続けた。多くの軌道ではタンパク質の慣性半径が8Å単位までコンパクトになったが、天然構造には至らなかった。

②80残基からなる置換型スリプレッサーのフォー

ルディングシミュレーションを実施した。水は露わには扱わず、Generalized Born モデルで取り込んだ。50nsの軌道を400本得た。1軌道あたり2,500スナップショットからなるので、構造の総数は100万個となる。PDBに登録されているダイマー構造のプロトマーを天然構造だと仮定し、それと類似の構造が100万スナップショットの中にあるかを確認したところ、RMSD値が4.72Åの構造が含まれていた(図1)。

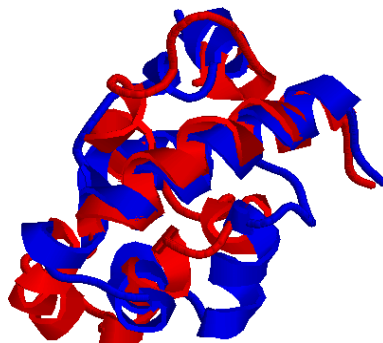


図1:シミュレーションで得られた構造(赤)と天然構造(青)の重ね合わせ。RMSD値は4.72Å

(2) 軌道の解析

①マルチプル軌道アラインメント

天然の山羊のαラクトアルブミン(123残基)のアミノ酸配列はグルタミン酸から始まるが、これを人工的に大腸菌で発現させるとN末端に開始コドンに相当するメチオニンが付加される。このたった1つのメチオニンの付加が、構造安定性を3.5kcal/mol劣化させ、アンフォールディング速度を5.7倍も速くする。メチオニンの付加が構造安定性に与える影響を調べるために、東京大学(当時)の荳口らにより、天然型、人工型のαラクトアルブミンについて、天然構造からのアンフォールディングシミュレーションが実施されていた(天然型、人工型ともに5nsの計算を10本、498K)。得られた合計20本の軌道を比較し、安定性の違いを生じる天然型、人工型の運動の特徴を把握することが課題となる。これを遂行するために、マルチプル軌道アラインメント法を開発、適用した。マルチプル軌道アラインメントを実施するには、以下のような手順を踏む。まずペアワイズ軌道アラインメントを作成し、類似性スコアから樹形図を描く。次に樹形図を参考にして類似している軌道ペアから順次アラインメントを作成していく。アラインメントができた時にはそれに関与した複数の軌道を合成して平均的な軌道(プロフィール)を導いておく。アラインメントされた軌道群と、別の軌道をアラインメントする場合は、平均的軌道とのアラインメントを実行する。このようにして全ての軌道のアラインメントが終わればマルチプル軌道アラインメントが完成する。これを基にして樹形図を作成し、アラインメントを作成し、ということを何度か繰り返し、樹形図とアラインメントが不変になったところで作業を止める。

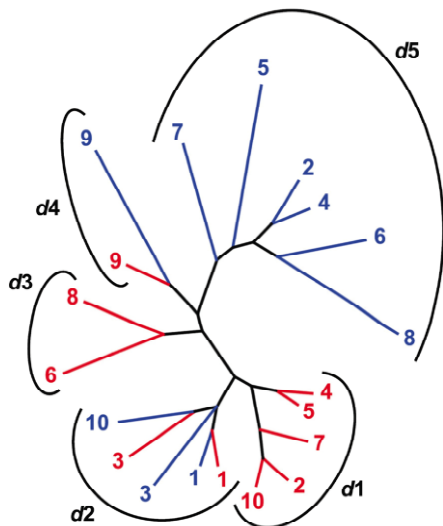


図2: 山羊 α ラクトアルブミンのアンフォールディング軌道の樹形図. 赤は天然型. 青は人工型

図2はマルチプル軌道アラインメントから得られた樹形図である. 樹形図の d1 クラスタは全て天然型, d5 は全て人工型からなっている. つまり, これらに分類された軌道が典型的な天然型, 人工型であり, その特徴を調べることで安定性の要因が理解できる. アラインメントを見ると, d1 では軌道は全て構造クラスタ E1 から E5 への直接転移を含んでいるが, d5 では E2-E4 といった中間構造を経由している. つまり, 天然型のアンフォールディングは all or nothing 型の典型的な2状態転移を示すが, 人工型は多くの中間状態を含み緩やかに転移することがわかった.

②部分オーダアラインメント

一般的なマルチプルアラインメントでは, その構築過程で平均配列(軌道)にあたるプロフィールを作成するが, 保存していない箇所があると配列や軌道の個性が相殺されてしまう. そこで, 平均化をせずに配列(軌道)の束の情報を保持しておき, もっともスコアの良いものとのアラインメントをとるような方法: 部分オーダアラインメントが提案されている. この方法を, 以前にシミュレーションをした Trp-Cage の軌道に適用した. アミノ酸配列は 20 種類のアミノ酸の列になるが, 軌道は様々な構造の時系列なので部分オーダアラインメントを実施する場合に構造クラスタの同定も同時に行う必要があった. そこで, 軌道が加わる時にクラスタリングを行い, 場合によってはスナップショットの属する軌道のクラスタ属性を変更するようにして, アラインメントを実施した. 図3では, e クラスタが新たに加わったので(上パネル), 軌道1の2番目のスナップショット(T1S2)が b クラスタから移動し, クラスタネットワーク構造が変化したことを示している(下パネル).

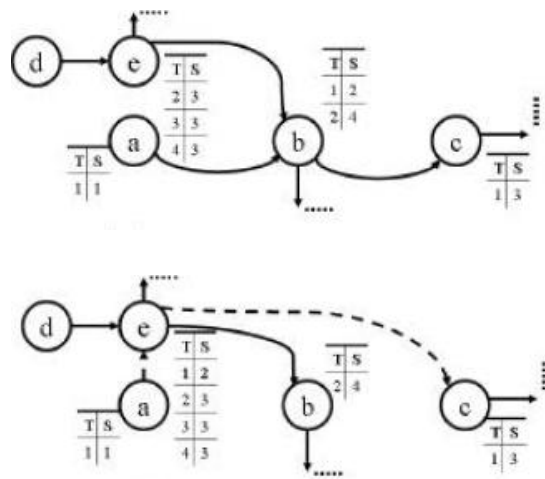


図3: 新しいクラスタが加わる際にスナップショットが帰属するクラスタが変更される様子.

このようにして部分オーダアラインメントを実施した結果, Trp-Cage がフォールドする際に N 末と C 末が近づいた形(リング構造)の出現が重要であるという, 以前, 目視により発見した知見を, 自動で得ることができた.

③遍歴プロフィール

Trp-Cage の軌道は, 以前に, フォールドする直前のものについては軌道アラインメントを実施し, 詳細な分類を行った. しかし, 全くの伸びた状態から, フォールドする直前のかかなりコンパクトな状態までどのようにして至るのかについては手つかずであった. 軌道アラインメントは精密な解析ができるという強みがあるが, 重ね合わせができるほど軌道が類似していないと比較に利用できないので, これに変わる, もっと大雑把な軌道類似性を把握する方法が必要となる. そこで, 遍歴プロフィールを考案した. 200 本ある Trp-Cage の軌道の各スナップショットをクラスタリングし, 軌道をクラスタ列に変換する. 次に, 各軌道について, 各クラスタを含むか含まないかを調べ, デジタルデータにする. これを遍歴プロフィールという. 遍歴プロフィールを比較することで, 軌道の大雑把な類似性がわかる. 類似性スコアに基づいて UPGMA 法で軌道の樹形図を書くと, 軌道は天然構造にフォールドする群, 疑似天然構造にフォールドする群, フォールドしない群に分離された. 更にフォールドしない群は2つの群に分離された. それぞれの群から, 属性が一致している軌道を選択し, それらについてクラスタのネットワーク図を書き, フォールド過程を調べた. その結果, フォールディングの初期過程で, 主鎖が左巻きとなるか, 右巻きとなるかが, その後の軌道が天然構造にフォールドするか, 疑似天然構造にフォールドするかを決定づけていることがわかった(図4).

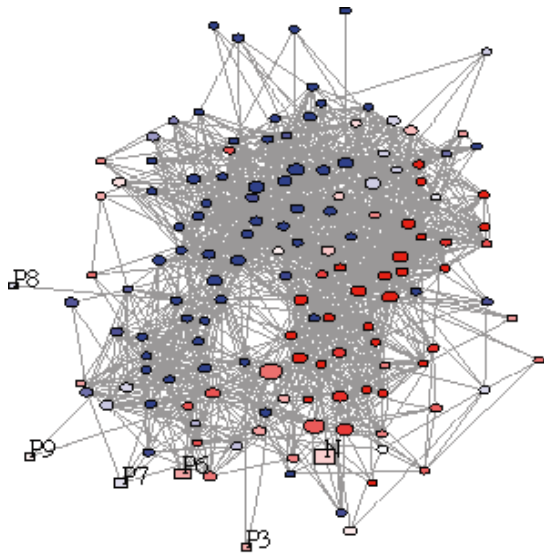


図4: Trp-Cage の軌道のネットワーク図. 赤が左巻キリング, 青が右巻キリング. 上部が初期過程. 左巻キリングのクラスタが連なり, N(天然構造)クラスタに接続している.

④ λリプレッサのクラスタ解析

λリプレッサがどのような経路を辿ってフォールドするかを調べるため, 距離マップを用いて構造を階層的にクラスタリングし, その構造遷移図を作成した. 全体像を把握するために閾値を 7Å に設定してクラスタリングを実施し, 顕著なクラスタ遷移だけに着目したところ, おおよそ5種類のクラスタ集団が確認されたので, このクラスタ集団の遷移課程を詳しく調べた. その結果, もっとも初期過程であらわれるクラスタ集団ではランダム構造からヘリックスが形成されるが, この時, 1番目と4番目, そして2番目と3番目のαヘリックスの接触で特徴付けられるジッパーとよばれる構造となり, 次のクラスタ遷移がおきることがわかった. λリプレッサの場合, 天然構造よりもコンパクトになるクラスタ集団も確認された. これらはαヘリックスが一部壊れることで強いパッキングを達成していた. また, データベースに登録されている構造はダイマーだが, その相互作用面にあたるヘリックスは構造形成能が低いこともわかった(シミュレーションはモノマー).

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 9 件)

- ① H. Sun, H. Ferhatosmanoglu, M. Ota, Y. Wang, An enhanced partial order curve comparison algorithm and its application to analyzing protein folding trajectories. *BMC Bioinformatics* **9** (2008) 344 (査読あり)

- ② H. Sun, H. Ferhatosmanoglu, M. Ota, Y. Wang, Enhanced partial order curve comparison over multiple protein folding trajectories. *Comput. Syst. Bioinformatics. Conf.* **6** (2007) 299-310 (査読あり)
- ③ T. Oroguchi, M. Ikeguchi, M. Ota, K. Kuwajima and A. Kidera. Unfolding pathways of goat α-lactalbumin as revealed in multiple alignment of molecular dynamics trajectories. *J. Mol. Biol.* **371** (2007)1354-64 (査読あり)

[学会発表] (計22件)

- ① 太田元規, 大きなタンパク質のフォールディングシミュレーション, 2009 年度「タンパク質・ゲノム構造」勉強会, 2009 年 9 月, 諏訪
- ② M. Ota, M. Ikeguchi, A. Kidera, Annotating protein-folding funnel by landscape profile and network, VII European Symposium of the Protein Society, 2007, May, Stockholm

6. 研究組織

(1)研究代表者

太田 元規 (OTA MOTONORI)
名古屋大学・大学院情報科学研究科・教授
研究者番号: 40290895

(2)研究分担者

ナシ

(3)連携研究者

池口 満徳 (IKEGUCHI MITSUNORI)
横浜市立大学・大学院生命ナノシステム科学研究科・准教授
研究者番号: 61261995