

B-016

NetSolveを用いたPSA/GAcによるタンパク質立体構造予測 Prediction of Protein Tertiary Structure with PSA/GAc using NetSolve System

廣安 知之[†] 三木 光範[†] 宇野 尚子[‡] 岡本 祐幸[§]
Tomoyuki HIROYASU Mitsunori MIKI Naoko UNO Yuko OKAMOTO

1. はじめに

近年タンパク質の立体構造予測が注目されている。タンパク質は生物の体を形成する重要な物質で、その機能は構造に大きく関わっているといわれている。したがってタンパク質の立体構造を知ることにより、病理の解明や新薬の開発につながる事が期待されている。タンパク質はエネルギーが低く安定した状態に折り畳まれるため、シミュレーションによるタンパク質の立体構造予測は、エネルギー最小化問題として捉えることができる。シミュレーションによるタンパク質の立体構造予測では非常に多くの反復計算が必要になる。そこで本研究ではGridミドルウェアの1つであるNetSolve[1]を用いて、Grid環境におけるタンパク質立体構造予測システムを構築する。また、最適化手法として遺伝的交叉を用いた並列シミュレーテッドアニーリング(Parallel Simulated Annealing using Genetic Crossover: PSA/GAc)[2]を用いる。PSA/GAcの探索の大部分はSAであり、この部分は独立に実行できる。また一定間隔で解交換のための交叉を行う部分は2つ以上の個体で同期をとる必要がある。このため、RPC機能を用いてGrid上にPSA/GAcを実装する場合、SAの部分をServerが行い、交叉処理をClientが行うシステムが考えられる。しかし、この方法ではServerが処理するSAはある対象問題に特化したSAとなり汎用性に欠ける。そこで本研究では最適化部分と評価部分を分離したシステムを構築する。

2. 遺伝的交叉を用いた並列シミュレーテッドアニーリング

PSA/GAcは並列に実行しているSAの解の伝達時に遺伝的交叉のオペレータである遺伝的交叉を用いたものである。図1にPSA/GAcの模式図を示す。

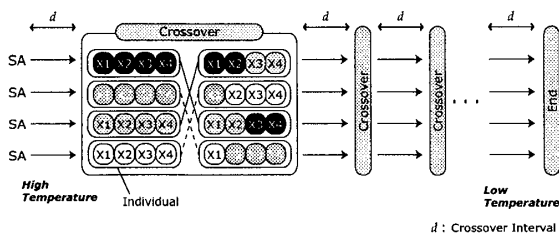


図1: PSA/GAcの仕組み

このモデルでは、解の伝達時に並列に実行しているSAから親としてランダムに2個体を選択し、設計変数交叉を行う。設計変数間交叉とは各設計変数の間でのみ交叉を行うことである。親個体と生成された子個体を合わせ

[†]同志社大学 工学部

[‡]同志社大学大学院 工学研究科

[§]岡崎国立共同研究機構 分子科学研究所

た4個体の中から良好な2個体を選択し、次の探索点とする。以上の「並列SA」「遺伝的交叉」の処理を終了条件まで繰り返す。

3. Grid - NetSolve System -

3.1 Gridとは

Gridとは計算資源の場所やネットワークの接続方法などを意識せずに、世界各地の計算資源を利用できるようにする技術である。遠隔地にある計算資源のライブラリを呼び出す方法として、RPC(Remote Procedure Call)機能をGrid上で実現したGridRPCがある。

3.2 NetSolve System

本研究では、Grid環境でPSA/GAcを実装するため、GridRPC機能を使用できるGridミドルウェアの1つであるNetSolveを用いた。図2にNetSolveの模式図を示す。

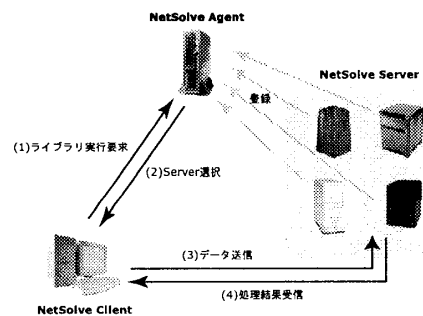


図2: NetSolveの構成

NetSolveは、計算資源を提供するServer、計算資源を利用するClient、ServerとClientの間を取り持つAgentの3つのコンポーネントから構成されている。

あらかじめ各ServerはAgentに登録されている。まず、ClientがAgentに対してライブラリの実行要求を行うと、AgentはClientの要求するライブラリが実行できるServerをClientに紹介する。ClientがAgentから紹介されたServerにデータを送ると、Serverはライブラリを実行し、処理結果をClientに返す。このように、NetSolveではClientがServerの情報を持っていないでもライブラリを実行することができる。

4. 現在のモデル

PSA/GAcのSAの部分をNetSolveのAPIを用いてServerで実行するモデルを図3に示す。

以下に現在のモデルの処理を示す。

1. ClientはAgentに「SAによるタンパク質立体構造予測」ライブラリの実行要求を行う。

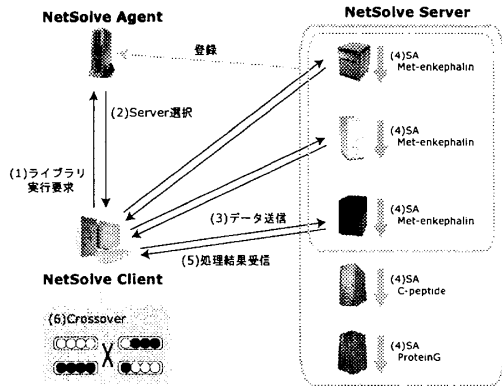


図 3: 現在のモデル

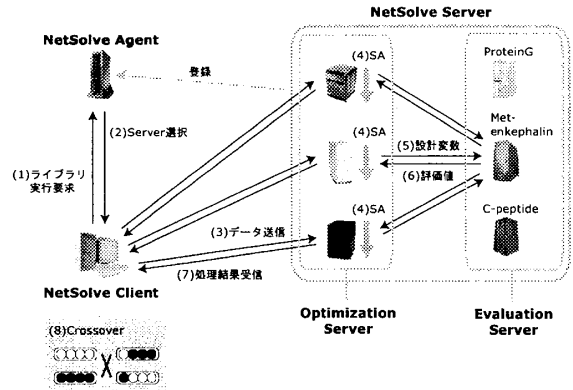


図 4: 提案モデル

2. Agent は「SA によるタンパク質立体構造予測」ライブラリを持つ Server を Client に紹介する。
3. Client は Agent から紹介された Server に個体情報と SA のパラメータを送信する。
4. Server は Client から受け取ったデータを元に SA を行う。
5. Server は SA の探索が終了すると、結果 (個体情報やパラメータ) を Client に返す。
6. Client は Server から個体情報を受け取り、ランダムに 2 個体選択し、設計変数間交叉を行う。
7. 1~6 の処理を終了条件まで繰り返す。

しかしこのモデルでは特定のタンパク質のエネルギー値が計算できる SA のライブラリを各タンパク質で作成する必要があり、汎用性に欠ける。

5. 提案モデル

現在のモデルでは汎用性に欠けるという問題を解消するため、図 4 のように最適化処理を行う Server (Optimization Server : O-Server) と評価計算を行う Server (Evaluation Server : E-Server) を分離したモデルを提案する。

以下に提案モデルの処理を示す。

1. Client は Agent に SA のライブラリの実行要求を行う。
2. Agent は SA のライブラリを持つ O-Server を Client に紹介する。
3. Client は Agent から紹介された O-Server に個体情報、SA のパラメータ、対象タンパク質の情報を送信する。
4. O-Server では SA の探索が行われる。
 - (a) O-Server では評価部で Client から指定されたタンパク質のエネルギー計算を行うライブラリの実行要求を行う。

- (b) Agent は特定のタンパク質の評価計算ができる E-Server を O-Server に紹介する。
- (c) O-Server は E-Server に設計変数 (二面角) を送信する。
- (d) E-Server は O-Server から送られた設計変数をもとに評価計算を行い、結果を O-Server に返す。
5. O-Server で SA の探索が終了するまで、4 の (a)~(d) の処理を繰り返す。
6. O-Server は SA の探索が終了すると、結果 (個体情報やパラメータ) を Client に返す。
7. Client は O-Server から個体情報を受け取り、ランダムに 2 個体選択し設計変数間交叉を行う。
8. 1~7 の処理を終了条件まで繰り返す。

このように SA を行う Server と評価を行う Server に分けることによって、ある問題に特化した SA のライブラリを持つ Server は不必要になる。Client は「SA を行うライブラリ」を持つ Server に問題の種類と SA の実行要求を行えばよい。

6. おわりに

本稿では、多くの反復計算が必要であるシミュレーションによるタンパク質立体構造予測システムとして、NetSolve を用いた PSA/GAc によるシステムを構築した。現在はある問題に特化したライブラリを持つ Server を用意していたが、それでは汎用性が低いため、SA を行う Server と評価を行う Server に分けたモデルを提案した。これにより様々な問題に適用できるシステムとなった。

参考文献

[1] NetSolve <http://icl.cs.utk.edu/netsolve/>
 [2] 廣安 知之, 三木 光範, 小塚 真貴, 岡本 祐幸. 遺伝的交叉を用いた並列シミュレーテッドアニーリングの検討. 情報処理学会論文誌: 数値モデルと応用, Vol.43, No.SIG10(TOM7), 2002.