

# NetSolve を用いた PSA/GAc によるタンパク質立体構造予測システム

## Prediction of Protein Tertiary Structure with PSA/GAc using NetSolve System

宇野 尚子

Naoko UNO

### Abstract:

Prediction of protein tertiary structure attracts attention in recent years. However, much repeated calculation is needed for prediction of protein tertiary structure by the simulation. Then, in this paper, we created the protein tertiary structure prediction system in the Grid environment using NetSolve which is one of the GridRPC systems. we used PSA/GAc as the optimization technique. In the current model, since server has only the protein tertiary structure prediction library, versatility was low. Then, in this paper, we proposed the model that divided into Server which performs SA, and Server which performs evaluation. It became the high versatile system applicable to various problems by this.

## 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言われている。したがってタンパク質の立体構造を予測することにより、病理の解明や新薬の開発につながることを期待されている。しかし、シミュレーションによるタンパク質の立体構造予測は非常に多くの反復計算が必要になる。そこで本稿では Grid ミドルウェアの 1 つである NetSolve<sup>1</sup> 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、タンパク質の立体構造予測には最適化手法の 1 つである PSA/GAc (Parallel Simulated Annealing using Genetic Crossover) 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。そこで本稿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 Grid - NetSolve System-

### 2.1 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) システムがある。

<sup>1</sup>Tennessee 大学の Innovative Computing Laboratory の Jack Dongarra 等によって開発された Grid RPC System

### 2.2 NetSolve System

本研究では、Grid 環境で PSA/GAc を実装するため、Grid 上で RPC 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。Fig. 1 に NetSolve の模式図を示す。

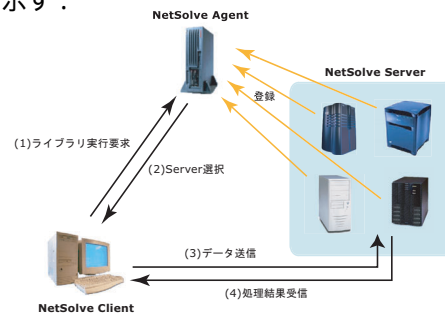


Fig. 1 NetSolve System

NetSolve システムは以下の 3 つから構成される。

- Server : 計算資源を提供する
- Client : 計算資源を利用する
- Agent : Server と Client の間を取り持つ

まず各 Server は Agent に登録する。Client が Agent に対してライブラリの要求を行うと、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 が処理結果を Client に返す。このため、Client は Server の情報を持っていなくてもライブラリを実行することができる。

## 3 PSA/GAc

PSA/GAc は並列に実行している SA の解の伝達時に遺伝的アルゴリズムのオペレータである遺伝的交叉を用いたものである。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、設計変数間交叉は各設計変数

の間でのみ交叉を行う。親個体と生成された子個体を合わせた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。

PSA/GAc ではSAを並列化することで探索点が増えるため収束が早くなる。また遺伝的交叉を行うため、部分解がある問題に対して有効な手法であるといえる。

#### 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Fig. 2に示す。

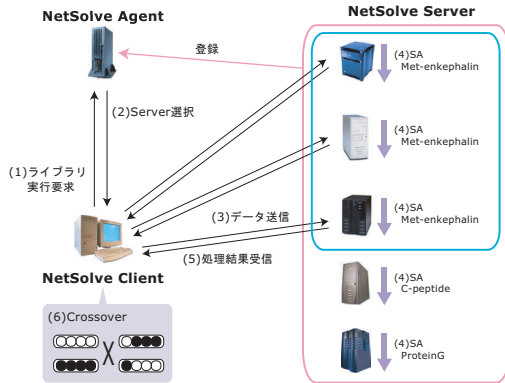


Fig. 2 現在のモデル

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。
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、Fig. 3のようにSA部分と評価部分を分離したモデルを提案する。

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。

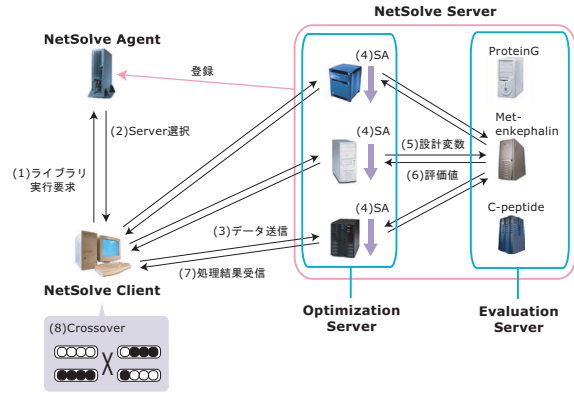


Fig. 3 提案モデル

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) 青井桂子. Grid環境におけるタンパク質立体構造予測. 同志社大学工学部知識工学科卒業論文, 2002.