

グリッドによる大規模高速計算技術 を用いた蛋白質構造予測 - High Throughput Computing -

高田 彰二 (stakada@kobe-u.ac.jp)

神戸大学 理学部

藤川 和利 (fujikawa@itc.aist-nara.ac.jp)

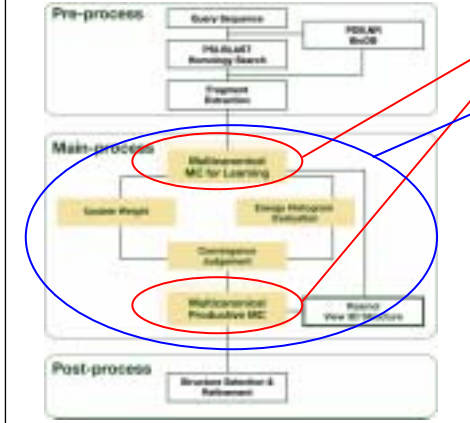
奈良先端科学技術大学院大学 情報科学センター

HTCグループの概要

- 目的
 - 大学等の遊休計算機を利用したグリッドサービスの実現
 - 繰り返し計算により、徐々に結果の精度を高めるような科学技術計算処理の支援
 - 単独の計算処理(ジョブ)だけでなく、計算処理群全体に対してのスループット向上
 - ユーザの戦略を考慮
- ターゲットアプリケーション
 - *Ab initio*蛋白質構造予測

蛋白質構造予測の流れ

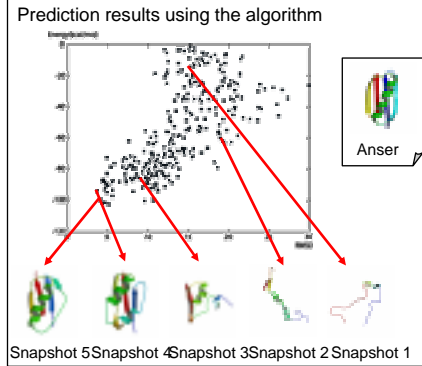
Workflow of *Ab initio* Protein Structure Prediction



Speculative execution
+
Iterative operation



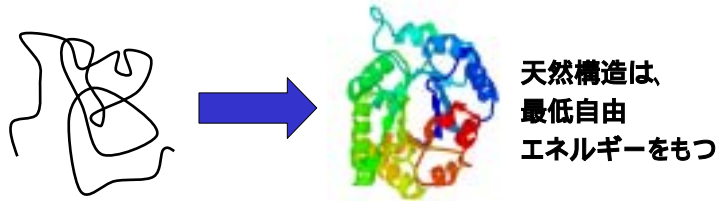
High Throughput Computing



蛋白質構造予測

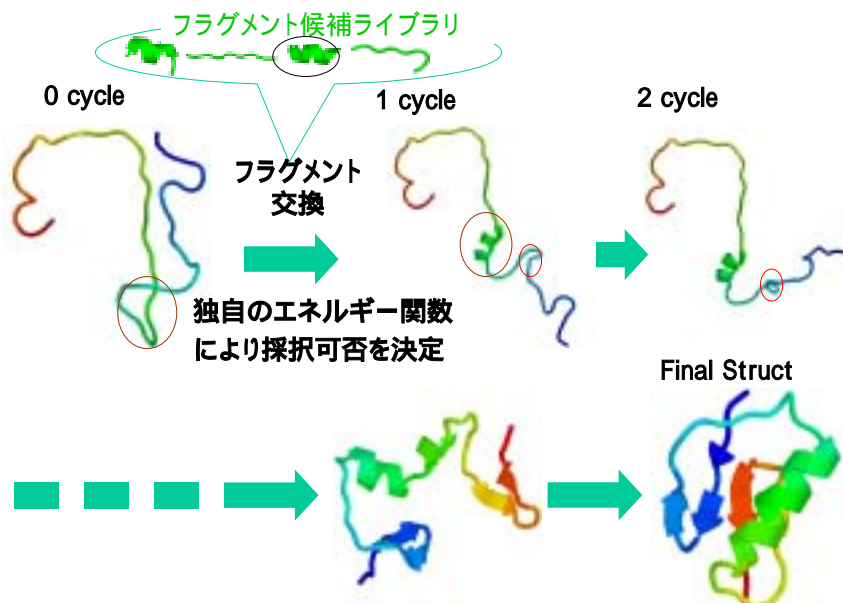
蛋白質構造予測の特徴

- アミノ酸配列からの *ab initio* 蛋白質構造予測
- 経験的なエネルギー関数を使って、低エネルギー構造を探索



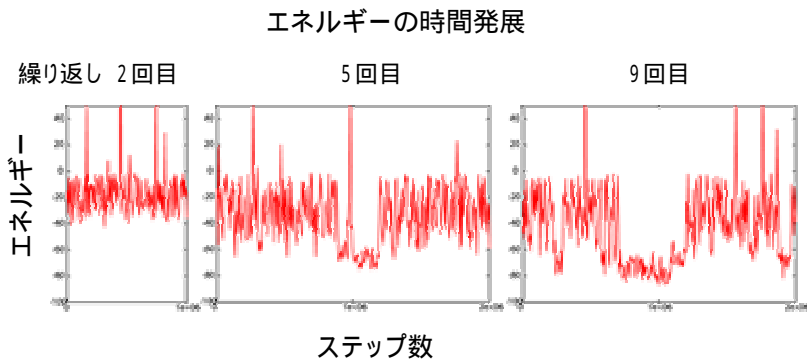
- 独自に作成した粗視化エネルギー関数
- フラグメント組み合わせ法による構造離散化
- マルチカノニカルアンサンブルモンテカルロ法による構造探索

構造構築法: フラグメント組み合わせ

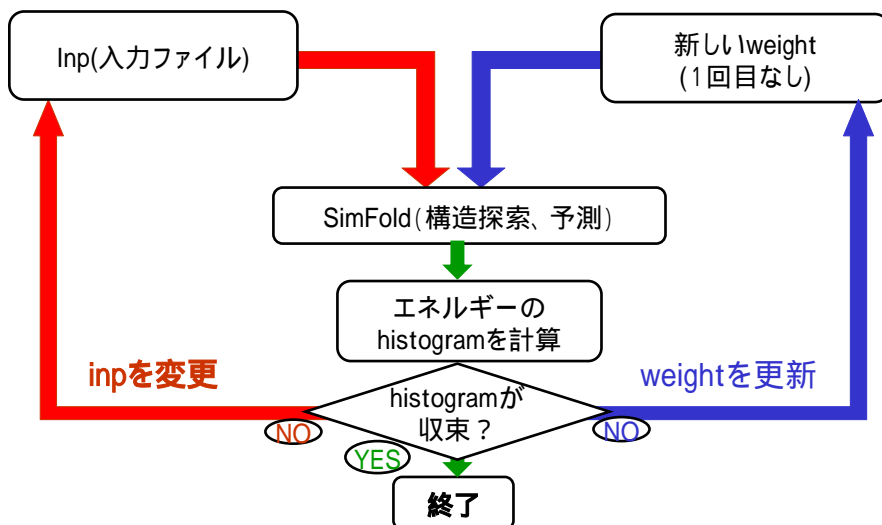


蛋白質の構造探索: マルチカノニカル法

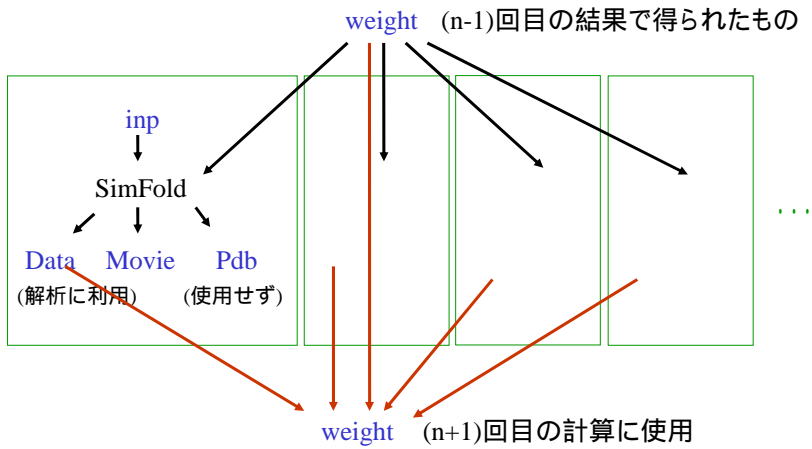
- マルチカノニカルアンサンブル法による繰り返し構造探索
 - weightを繰り返し更新することによって低エネルギー構造を探索できる



基本のフローチャート



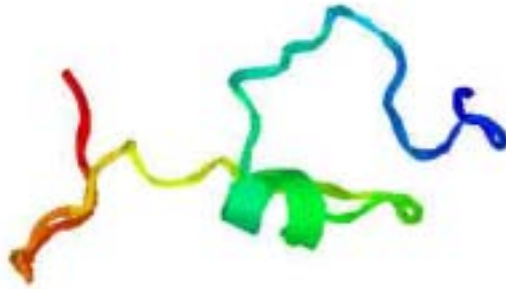
n回目の繰り返しのイメージ



Step数の例

- Protein G(56残基)の場合

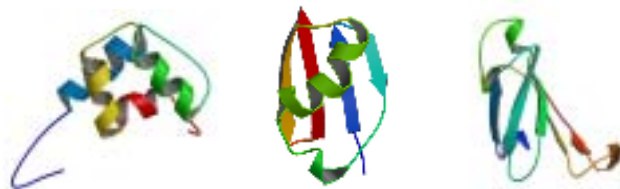
繰り返し	総Step数 (100万単位)
1	1
2	10
3	30
4	120
5	240
6	240
7	480
8	480



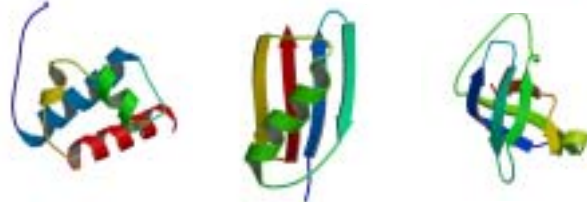
グリッド計算でゲノムワイドな構造予測へ

- 小型蛋白質(60アミノ酸)では、ある程度成功する(下図)
- 長さNに対して指数関数的に構造探索が困難に！
(150アミノ酸以上が必要)
- ゲノムワイド(10000規模の数)の蛋白質への適用が目標

計算による
最低エネルギー構造



天然構造(答え)



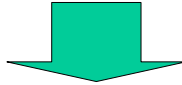
探索された構造の利用

- 最小エネルギー構造は天然構造の候補
- 全構造のクラスター分析
 - 得られる構造中心は、天然構造の候補
- 非天然構造は、
 - 変性状態アンサンブルの構造モデル
(実験では得られない情報)
 - エネルギー関数改良のために必須情報

グリッドシステムによる支援

ユーザの要求

- 単独の計算処理(ジョブ)だけでなく, 計算処理群全体に対してのスループット向上
- 複数のジョブによる繰り返し計算の待ち合わせ時間を最小にしたい
 - 負荷変動の影響を最小限に



- ジョブ制御ポリシー
 - いたるところにジョブを投入
 - ジョブを途中終了

ジョブ制御ポリシーとは

- 並列に実行されたジョブが必ずしもほぼ同時に終わるとは限らない.
 - 計算機のトラブルによるもの
 - 他のジョブによる計算機の高負荷
- 対処として
 - すべてのジョブの終了を待つ.
 - 遅れているジョブを途中終了させ, 新たなweightによるジョブを投入する.
 - 多めにジョブを投入し, 必要数の結果を収集した時点でジョブの強制終了を行う.
 - など



Contents

- Latest Data
- List of data from the lowest energy
- Energy distribution
- Protein Structure



Resource Manager

- Forecast the daily load pattern of nodes
- Make the job dispatching schedule

Seed Pool Manager

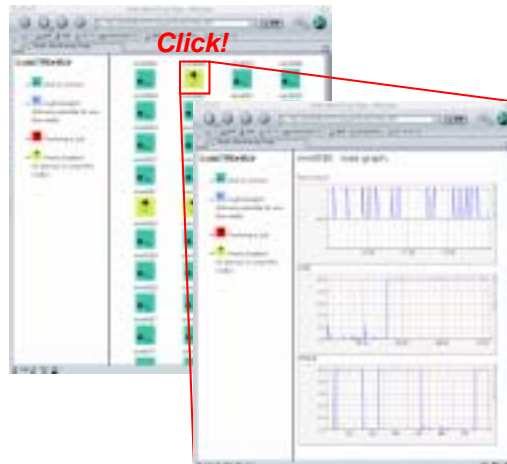
- Submit a job to a processing node according to the schedule

Webインターフェースによるジョブ投入

Job submission page



CPU Load monitor page



結果表示インターフェース

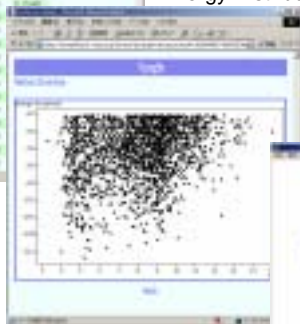
Main Page



Original Viewer System

We can display **the latest data.**
We can display **energy distribution graph.**
We can display **the folding results.**

Energy Distribution Graph



Snapshot



ジョブ投入システム

- Sun
 - Sun GridOne Engine
- Linux
 - Condor
- FreeBSD
 - DQS(Distributed Queuing System)

現在の状況

- ジョブスケジューラの組み込み
 - DQSのみ
- 負荷情報の収集・解析
 - SNMPによる収集
 - 過去1週間の情報を利用したノードの負荷予測
- Webを利用したノードの利用状況表示

今後の課題

- ジョブ制御ポリシーを記述する方法の確立
- n回目の入力パラメータの設定
 - ユーザインタラクションの受付
- データベースシステムとの連携
 - 結果の再利用
 - ノウハウの蓄積