

High Throughput Computing

対話的動作を考慮した蛋白質立体構造予測のHTCグリッドシステム

蟻川浩 増田慎吾 古田忠臣 金文珍 朴聖俊

高田彰二、藤川和利、砂原秀樹

奈良先端、神戸大

イントロ

蛋白質立体構造予測システムROKKY

構造予測における試行錯誤

対話的動作－柔軟なワークフロー制御

特定のアプリケーションをターゲットに ユーザーフレンドリーな グリッドコンピューティングシステムをつくる

蛋白質立体構造予測(モデリング)の最前線では、

かなり多数の計算ツールを使う

ワークフローによる制御

易ターゲットは、全自動で処理したい

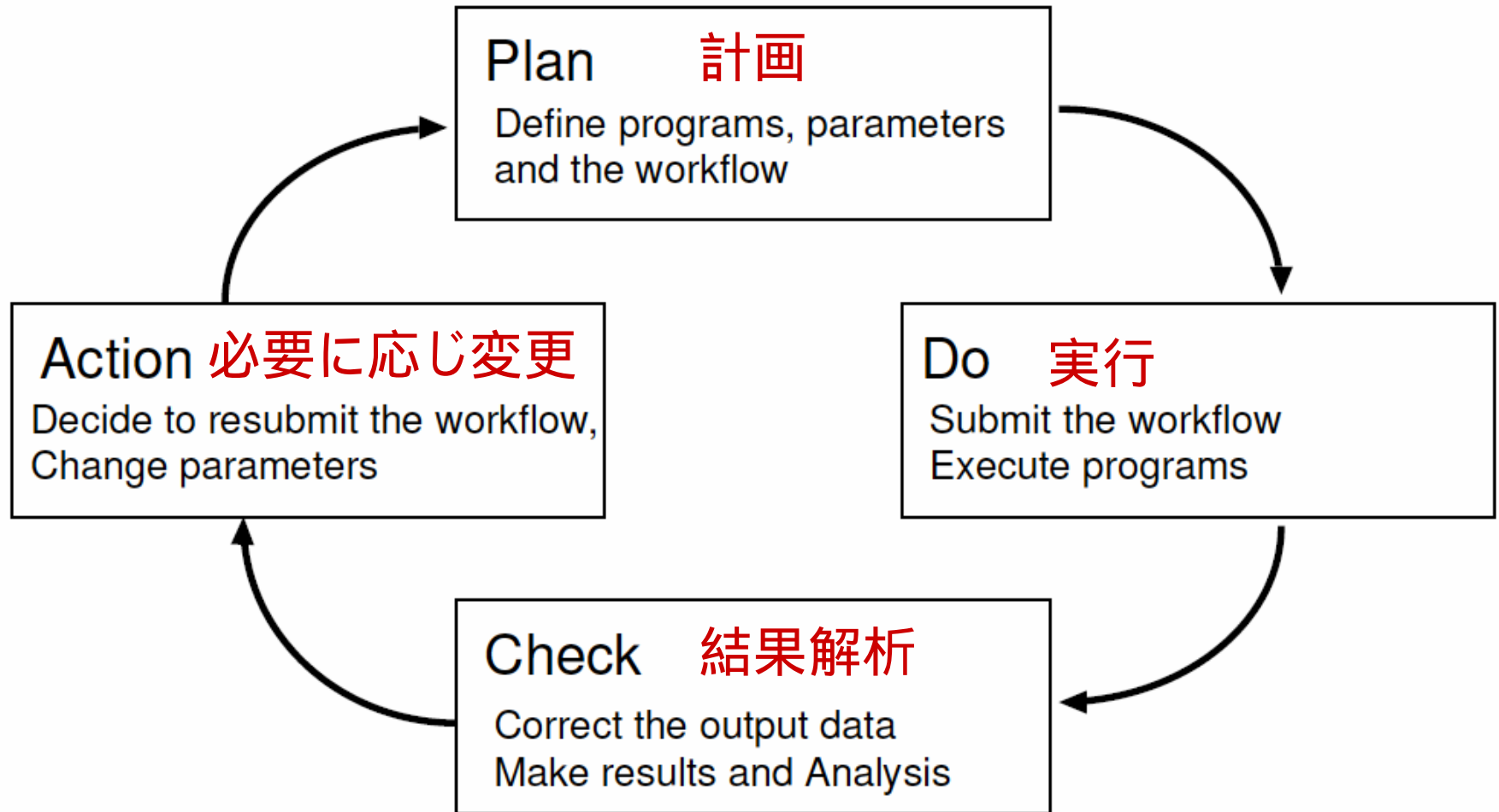
難ターゲットでは、乱数を変え独立MCシミュレーションが必要

High Throughput Computing向き

難ターゲットでは、エキスパートによる試行錯誤
(パラメータ、モデル選別)が必須

対話的動作

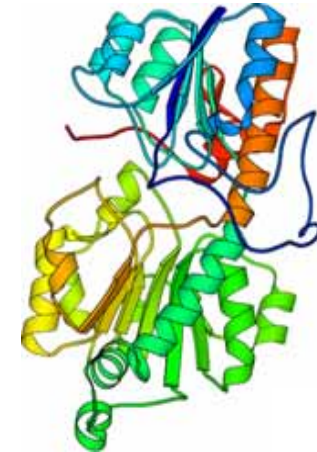
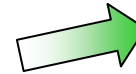
PDCAサイクル: 品質向上のための手順



蛋白質立体構造予測(モデリング)の最前線

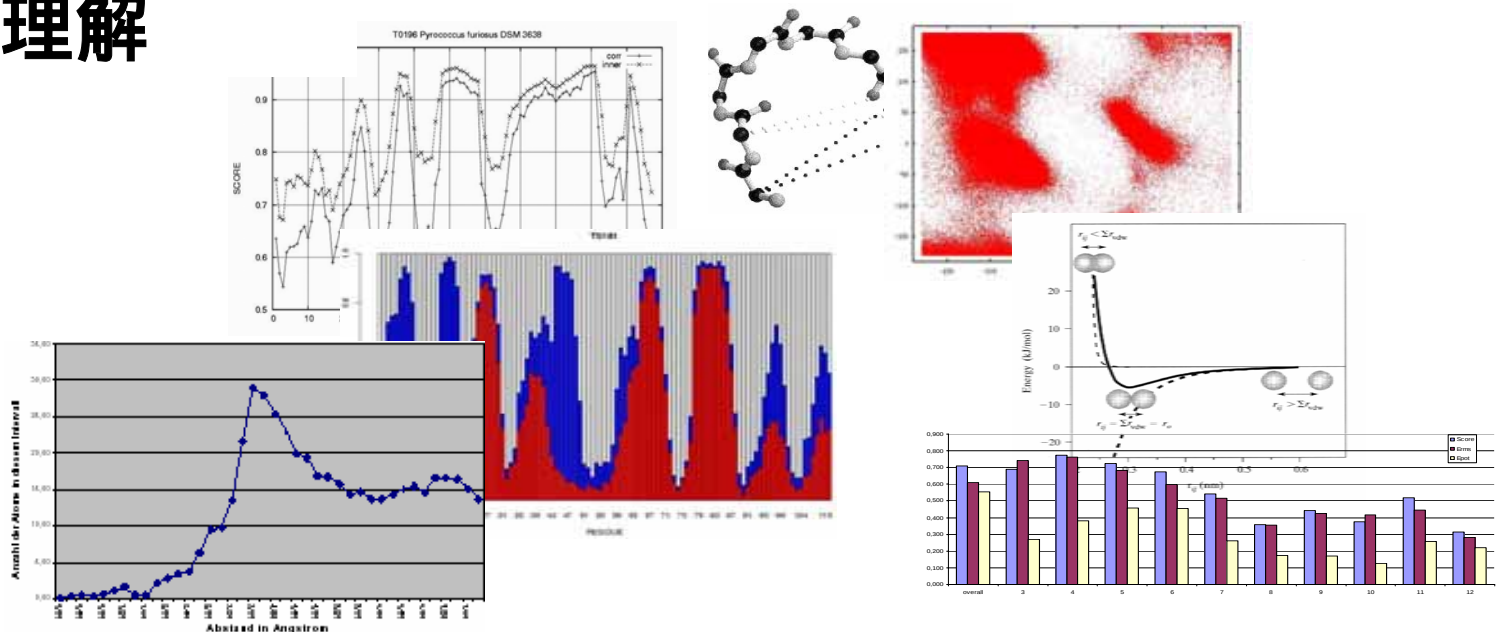
タンパク質立体構造予測(モデリング)

TIDQWLLKNAKEDIAELKKAGITSDFY
FNAINKAKTVEEVNALKNEILKAHATID
QWLLKNAKEDIAELKKAGITSDFYFNA
INKAKTVEEVNALKNEILKAHATIDQWL
LKNAKEDIAELKKAGITSDFYFNAINKA
KTVEEVNALKNEILKAHATIDQWLLKNA
KEDIAELKKAGITSDFYFNAINKAKTVE
EVNALKNEILKAHA



立体構造構築原理 の理解

創薬



タンパク質立体構造予測の必要性

- 1,796,422 sequences
 - ~ 28,000 structures,
 - Gap in numbers
- Experiments X-ray or NMR
 - Very expensive,
 - time consuming
- Computational methods
 - Fast (minutes or hours)
 - Inexpensive (desktop PC)
 - Correct solution on average in appx. 60% of the cases (new seq. → known fold)

ROKKY: 蛋白質立体構造予測サーバー

By ACT-JST program

www.proteinsilico.org

Input

ROKKY Protein Predictor - Microsoft Internet Explorer

File (F) Edit (E) View (V) Options (O) Help (H) Address (A) http://predict.chem.sci.kobe-u.ac.jp

ROKKY **TAKADA Lab.**
Department of Chemistry Faculty of Science
Kobe University
1050-1015 predict.chem.sci.kobe-u.ac.jp

Queuing

ROKKY Protein Predictor - Microsoft Internet Explorer

File (F) Edit (E) View (V) Options (O) Help (H) Address (A) http://predict.chem.sci.kobe-u.ac.jp/task/casp6

ROKKY **TAKADA Lab.**
Department of Chemistry Faculty of Science
Kobe University
1050-1015 predict.chem.sci.kobe-u.ac.jp

[TASK] [CASP6:ROKKY] [CASP6:SERVERS]

Note: Server predictions available [here](#)

ID	Target	PDB	Length	Status	Method	Date(JST)	User	IP
94	T0196	1a1	116	complete	pd-blast (1a1)	09-06-2004 00:15:19	casp6	156.111.8.48
99	T0197	179	179	complete	3d_jury (179)	09-06-2004 03:21:26	casp6	156.111.8.48
100	T0199	1aun	235	complete	3d_jury (1aun)	09-06-2004 03:37:43	casp6	156.111.8.48

Results

ROKKY Result - 00221-10278 - Microsoft Internet Explorer

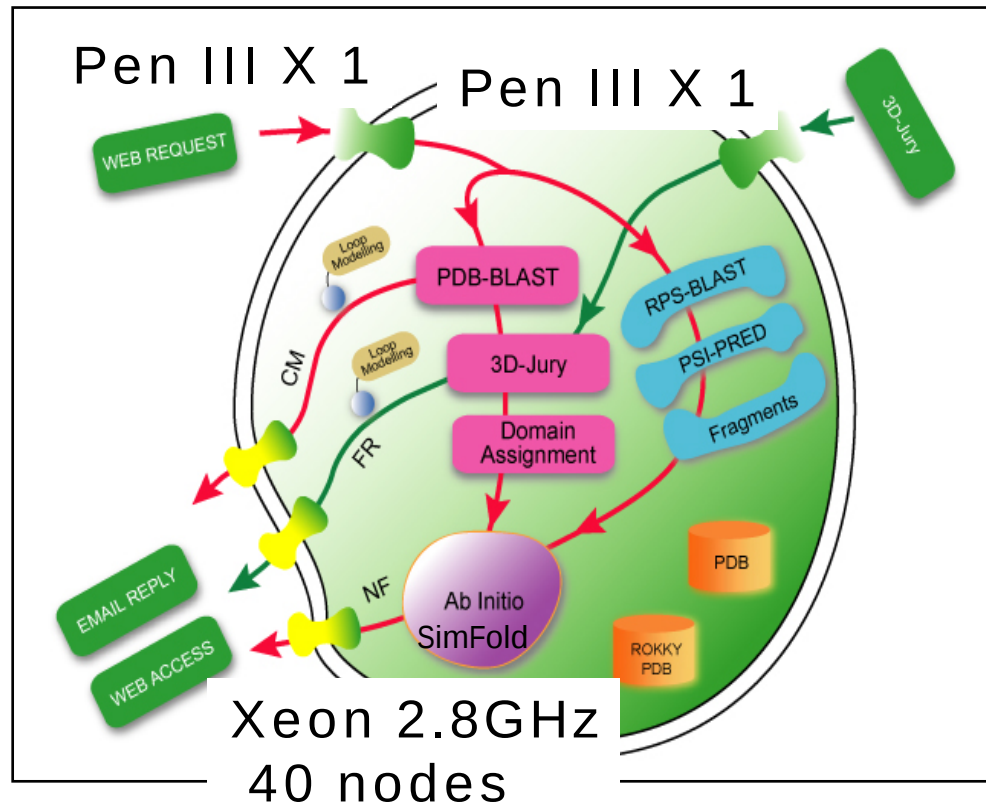
File (F) Edit (E) View (V) Options (O) Help (H) Address (A) http://predict.chem.sci.kobe-u.ac.jp/task/result.cgi?id=00221-10278&task=casp6

Pred: **1F02 A** E-value 4e-59 [PDB SUM] & **1F02 A** E-value 2e-58 [PDB SUM] & **1F05 A** E-value 1e-87 [PDB SUM] & **1F05 A** E-value 1e-84 [PDB SUM] & **1F05 A** E-value 1e-84 [PDB SUM] &

Ab Initio SimFold

Model 1 Model 2 Model 3 Model 4 Model 5

For CM targets, use PDB-BLAST
For FR targets, use 3D-Jury, &
For NF targets, use SimFold
“Large scale folding simulation”





What is CASP?

ブラインド構造予測コンテスト

2年に一度

春、構造未知蛋白質のアミノ酸配列がウェブ上で公開される

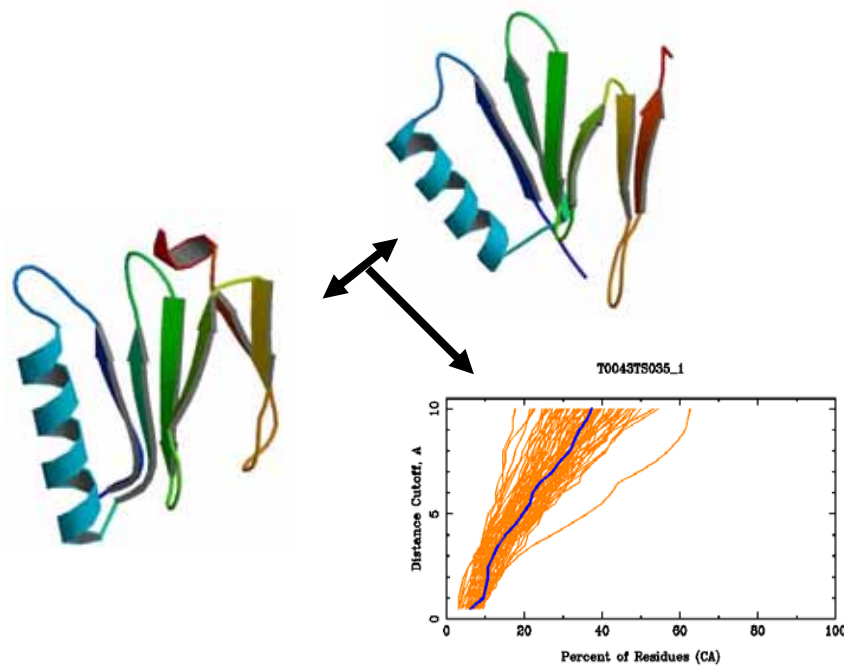
夏、予測屋が構造予測

実験による構造決定

秋、予測構造の数値評価

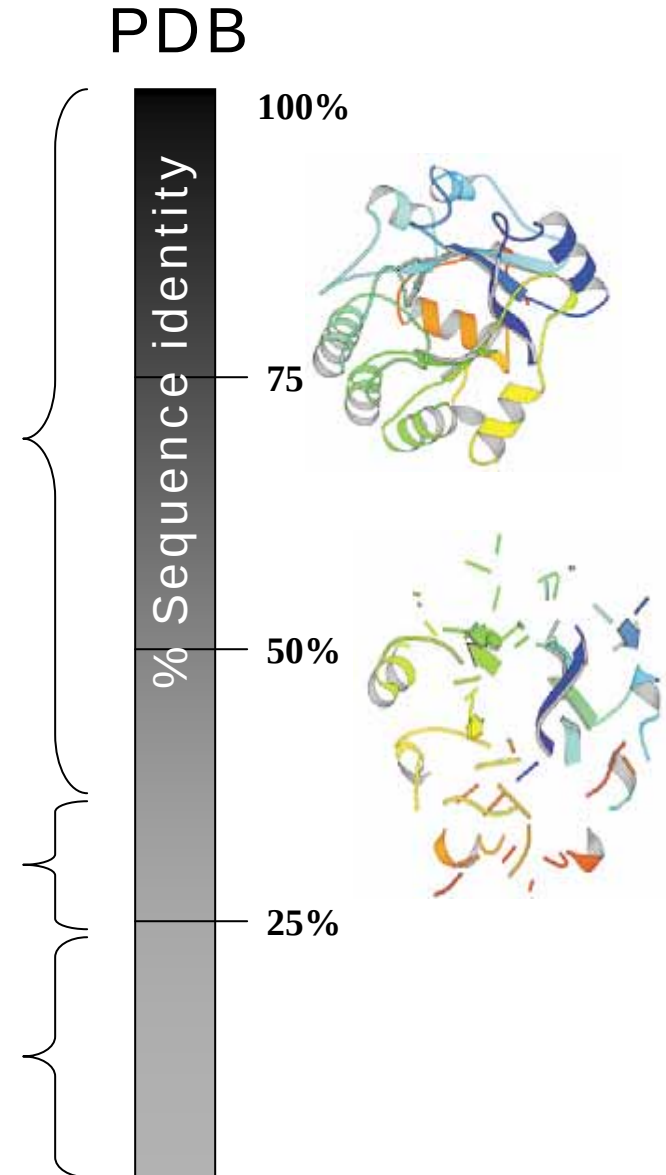
12月、結果発表の会議

	#Groups	#Predictions
1 (1994)	35	
2 (1996)	72	947
3 (1998)	98	3807
4 (2000)	163	4922
5 (2002)	215	28727
6 (2004)	228	38806



ターゲットの分類

- CM(Comparative Modeling)
 - 配列比較
 - 側鎖高精度モデリング
- FR (Fold Recognition)
 - 弱い配列相同性の検出
 - シミュレーション
- NF (New Fold)
 - フォールディングシミュレーション



サーバーとヒューマンの予測

サーバー

ヒューマン

予測する
時間

48時間

約一ヶ月

方法

全自動

なんでもあり

サーバーの予測結果が
公開されているので参照可

CASP 6 で、日本の貢献は今までになく大きかった

国別予測チーム数、2位

理研構造ゲノムプロジェクトが、沢山問題をくれた

CBRC FR/Hで3位 (Talk)

Chimera (北里大) FR/H部門で4位 (Talk)

Rokky (神戸大) NFサーバー部門で2位

Rokko & Rokky (神戸大) in NF部門のエリートクラブ入り

CBRC ドメイン予測で好成績

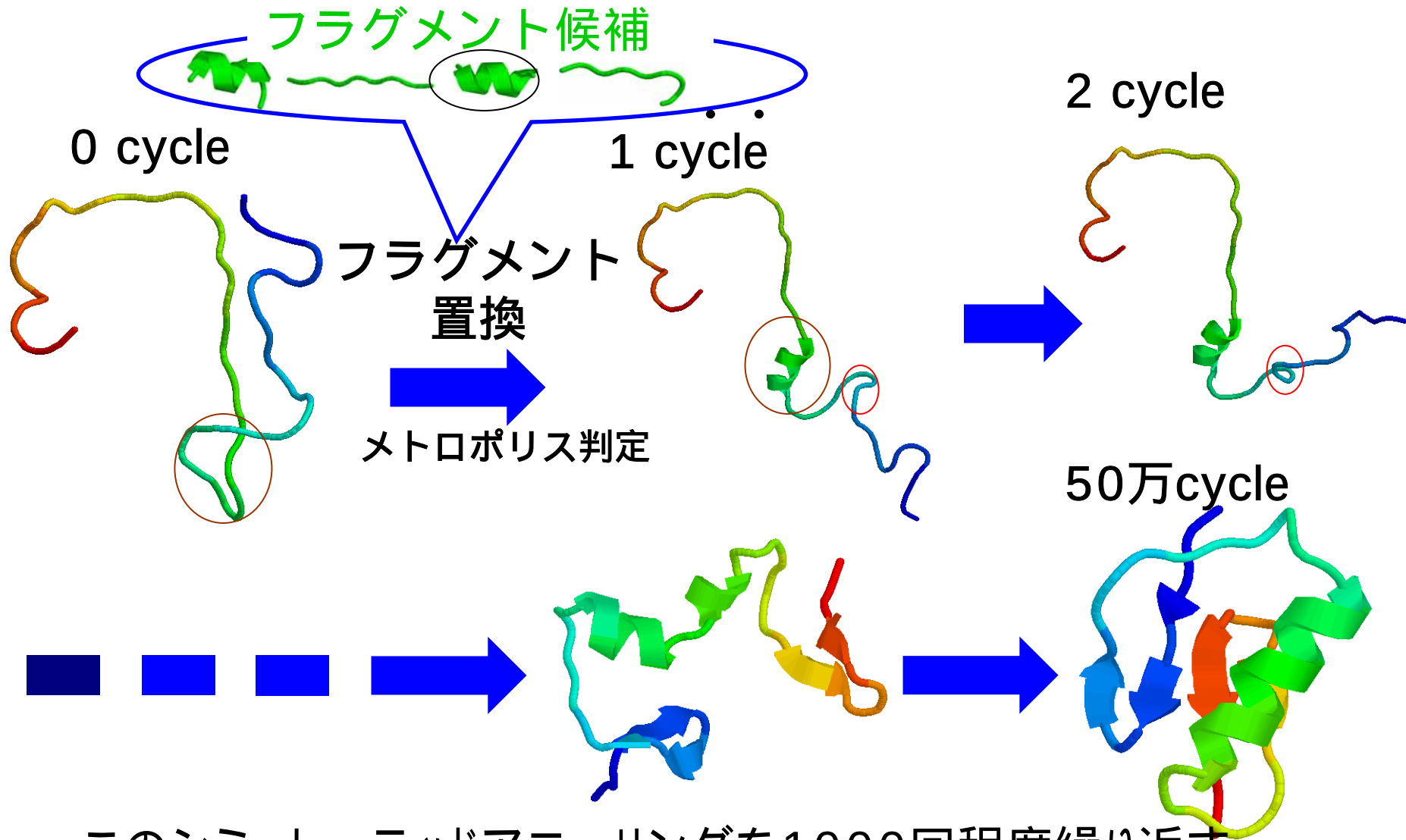
エリートクラブメンバー入り(~ 200チーム中)

<i>Group</i>	<i>NF+FR/A</i>	<i>NF</i>	<i>GDT_TS</i> (< 25%)
<i>BAKER</i>	<i>6</i>	<i>6</i>	<i>6</i>
<i>KOLINSKI&BUJNICKI</i>	<i>6</i>	<i>6</i>	<i>6</i>
<i>Ginalski</i>	<i>5</i>	<i>4</i>	<i>5</i>
<i>SAM-T04-hand</i>	<i>3</i>	<i>4</i>	<i>3</i>
<i>Skolnick-Zhang</i>	<i>3</i>	<i>0</i>	<i>5</i>
<i>Jones-UCL</i>	<i>3</i>	<i>2</i>	<i>1</i>
<i>BAKER-ROBETTA</i>	<i>3</i>	<i>3</i>	<i>1</i>
<i>BAKER-ROBETTA_04</i>	<i>0</i>	<i>1</i>	<i>1</i>
<i>Rokko</i>	<i>0</i>	<i>2</i>	<i>0</i>
<i>Rokky</i>	<i>0</i>	<i>2</i>	<i>0</i>
<i>KIAS</i>	<i>1</i>	<i>0</i>	<i>1</i>
<i>MCon</i>	<i>0</i>	<i>0</i>	<i>1</i>

by
B.K.Lee

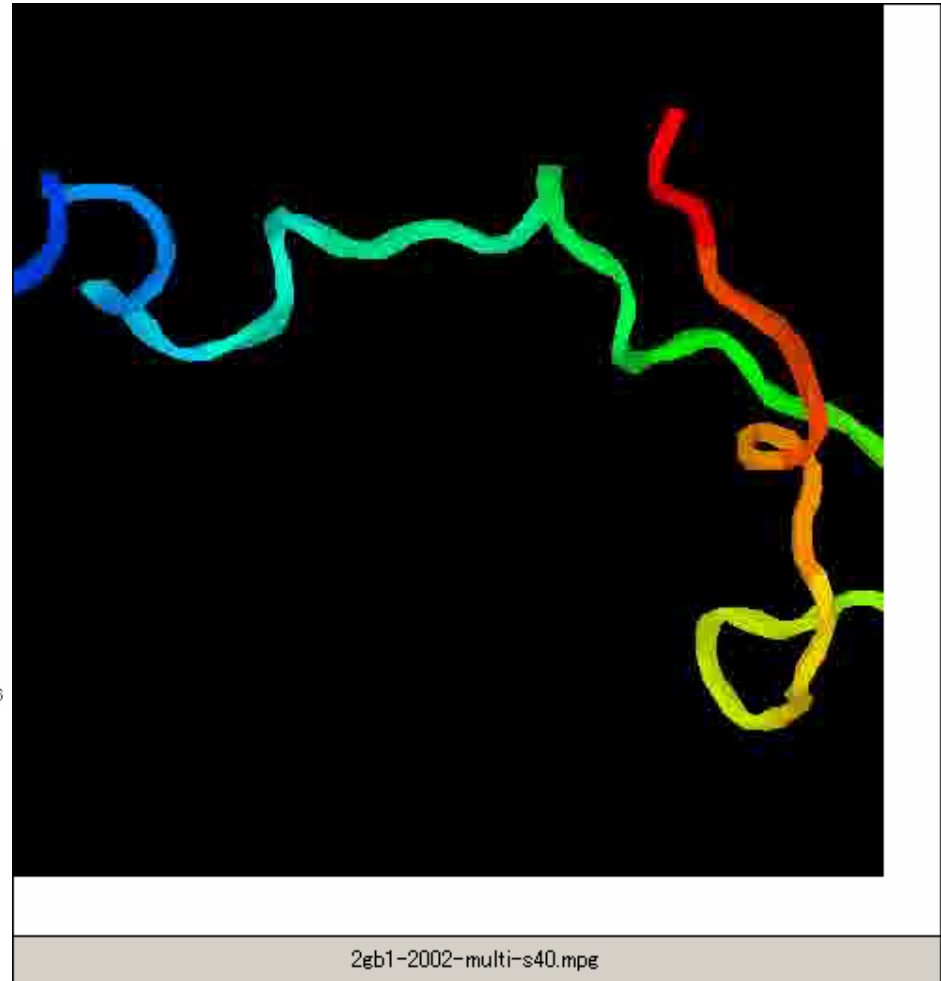
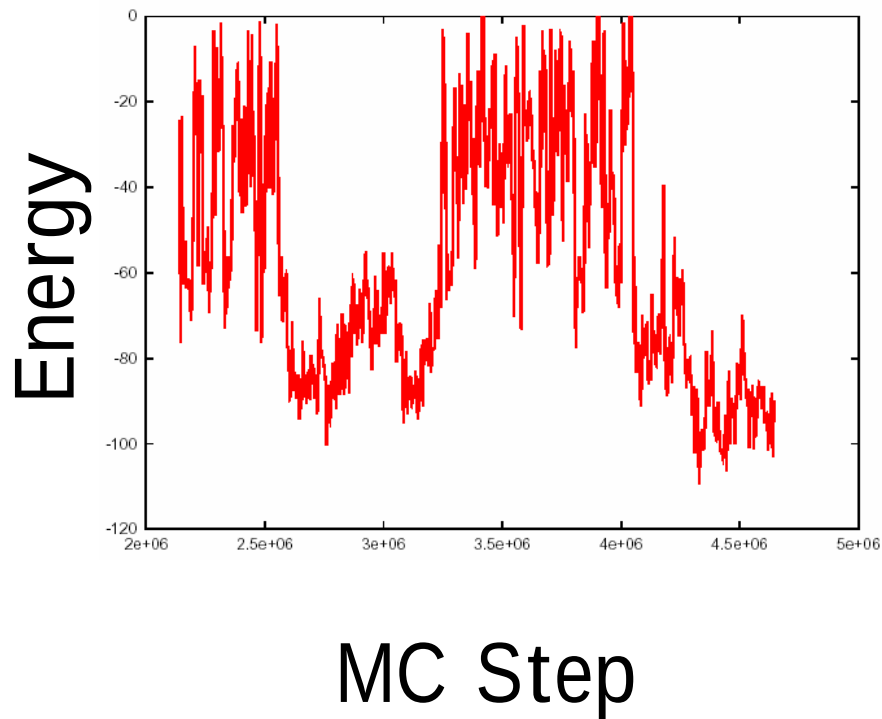
Rokky (神戸大) NFサーバー部門で2位

フラグメント組合わせによる 構造形成シミュレーション

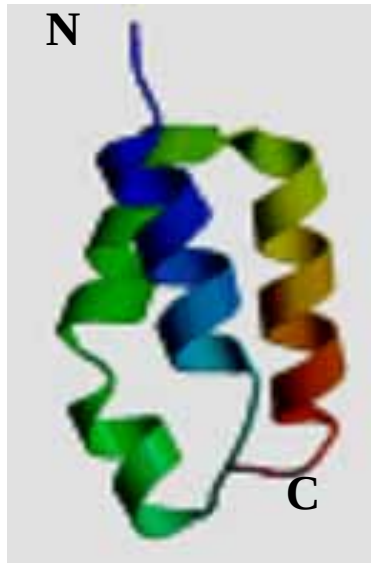


このシミュレーテッドアニーリングを1000回程度繰り返す。

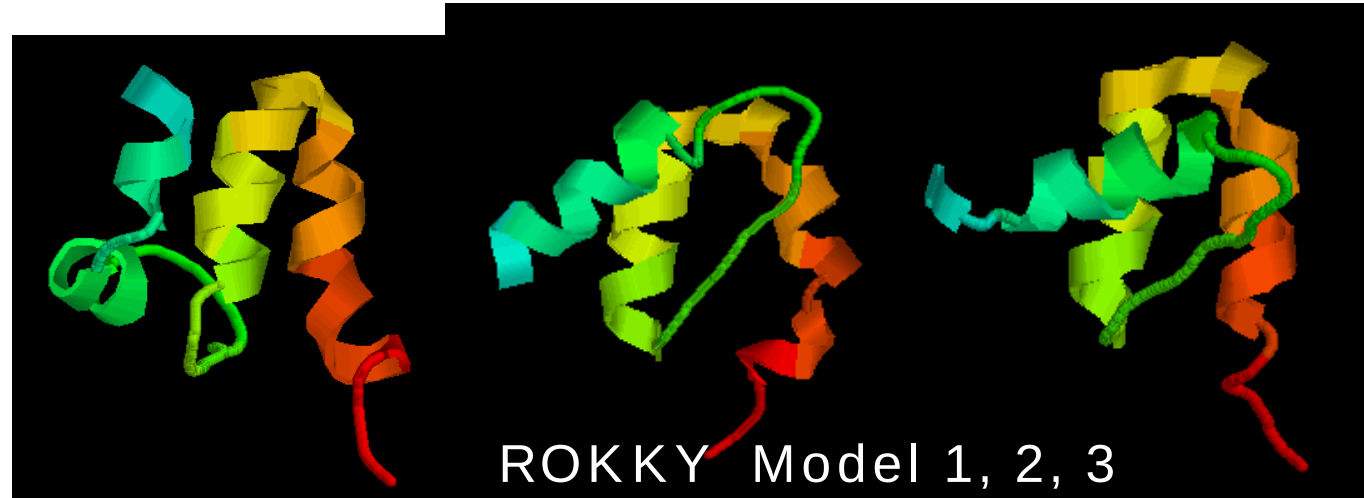
Protein G folding with multicanonical fragment assembly



NF Target (T0215) 注:色は気にしないでください



Native



“フォールディング”によるNFターゲットの構造予測

非常に計算機パワーを必要とする

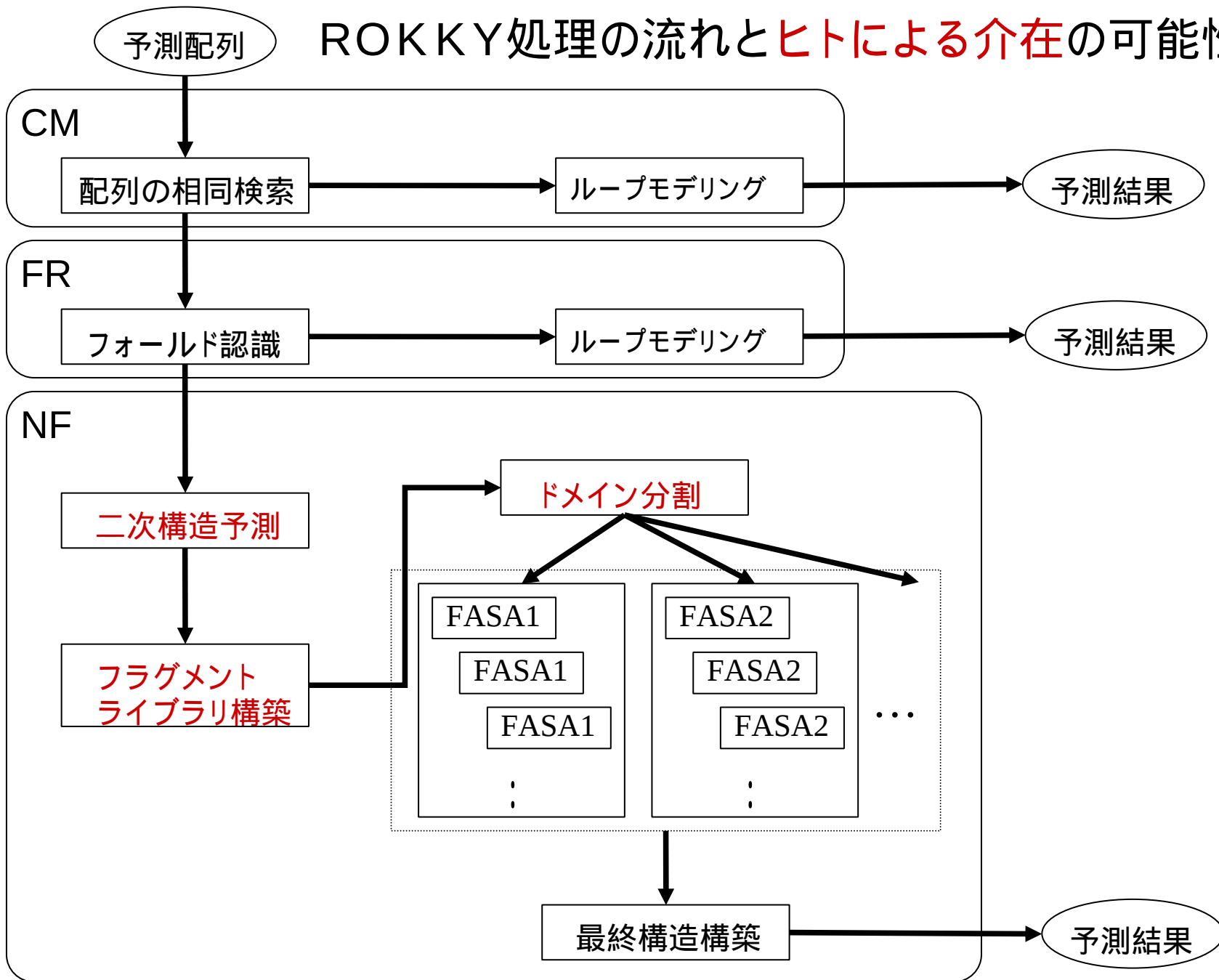
お互いに疎結合な多数の計算を行う

まさにHigh Throughput Computing向き課題

構造予測における試行錯誤

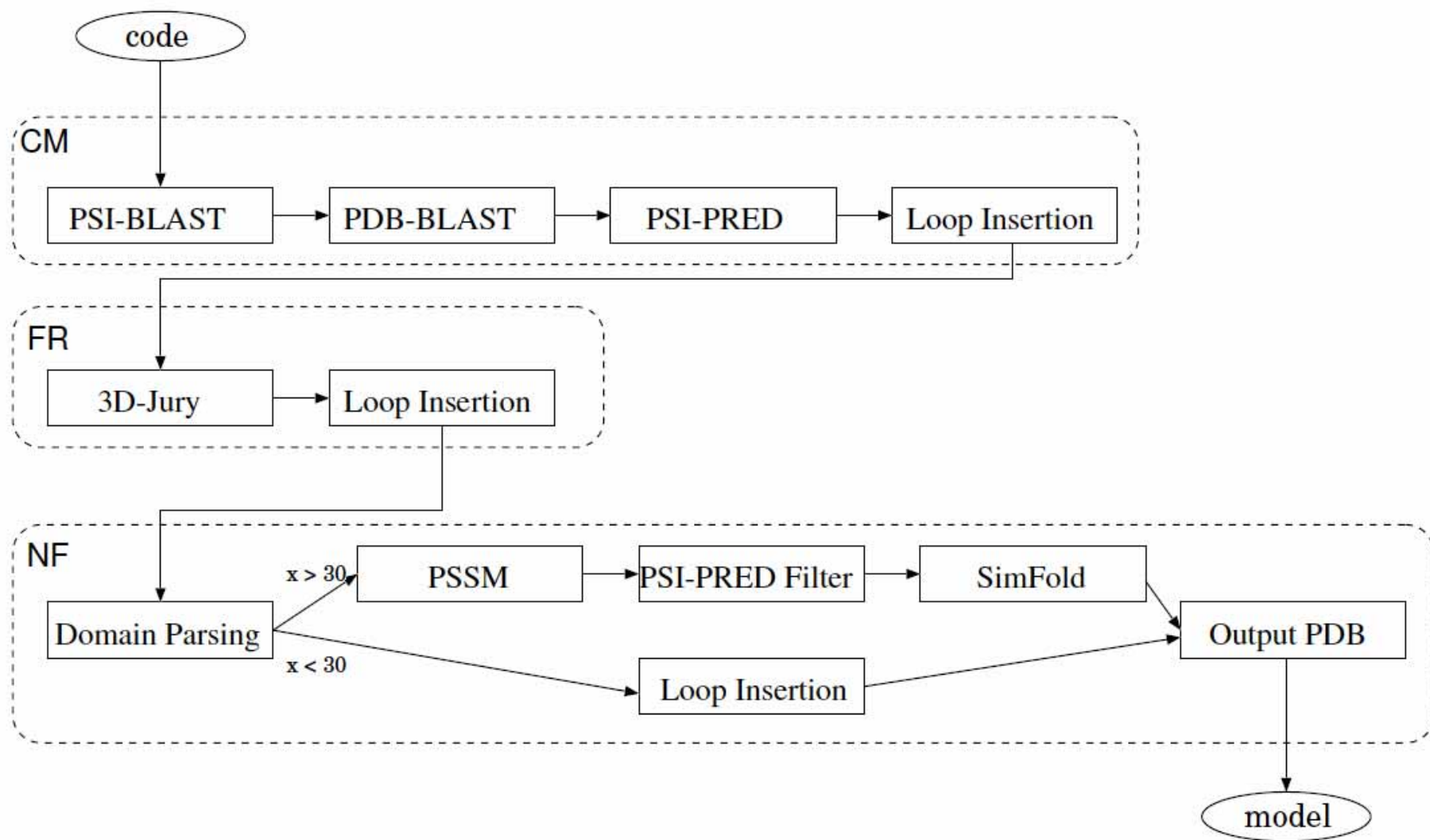
NFターゲットの構造予測には、現時点で、
全自動には限界有り
エキスパートの経験が要る。

ROKKY処理の流れとヒトによる介入の可能性



対話的動作を考慮した 柔軟なワークフロー制御

作成したROKKYのワークフロー

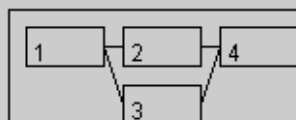


ワークフロー設計 & 制御支援ツール



ワーク投入、停止等

プログラムの実行順序



Work Number :: 4

コマンド

パラメータ1

パラメータ2

パラメータ3

パラメータ4

パラメータ5

ループX

ループY

ループZ

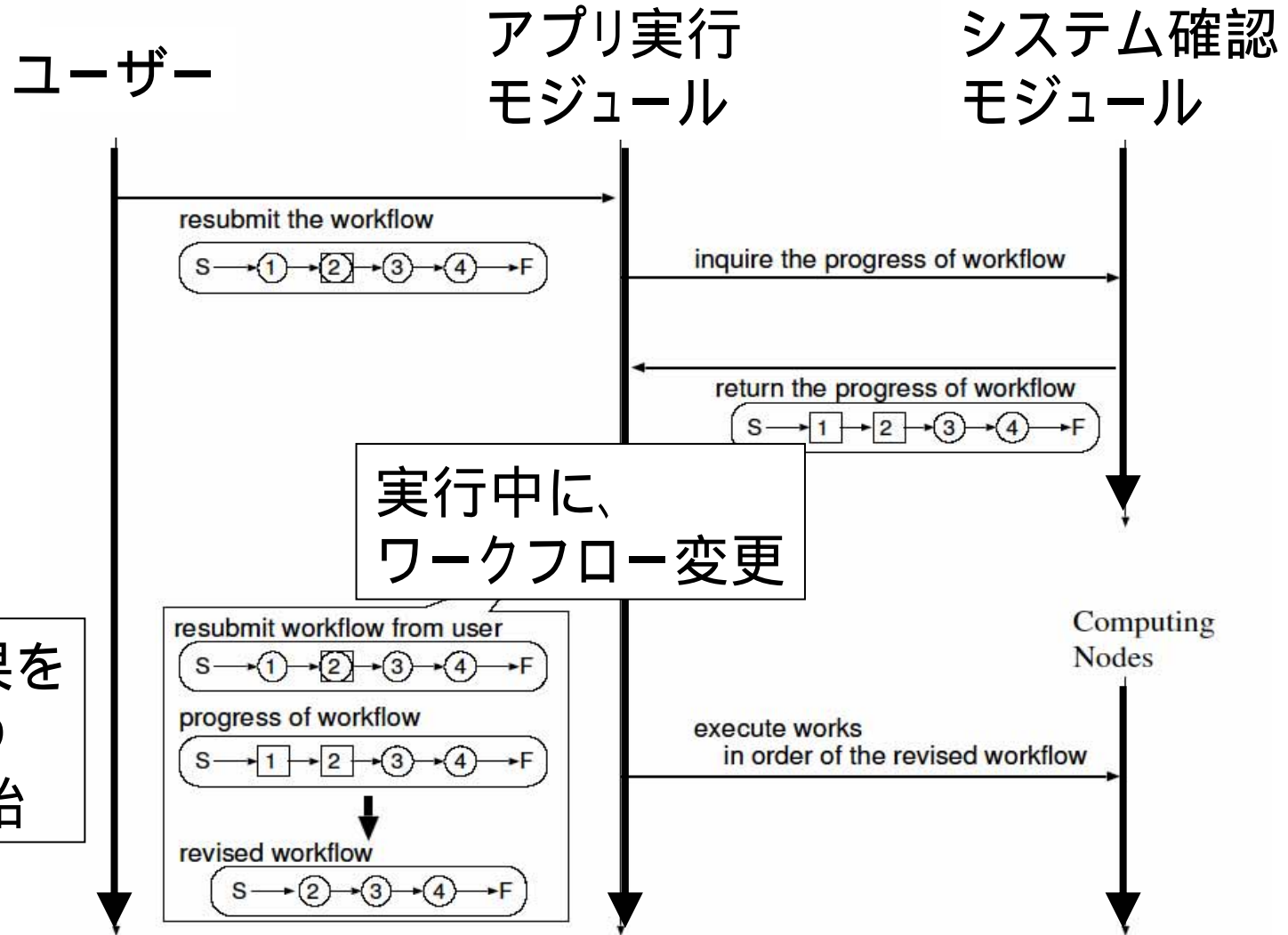
☐ jRadioBut... ☐ jRadioBut... ☐ jRadioBut...

次のワーク

OK Cancel

入力パラメータ

ワークフロー差分実行機構



- ② a work on standby
- ② a work with changed parameters
- ② a finished work

モジュール構成

ユーザーインターフェイス

プログラム
実行

システム
状況監視

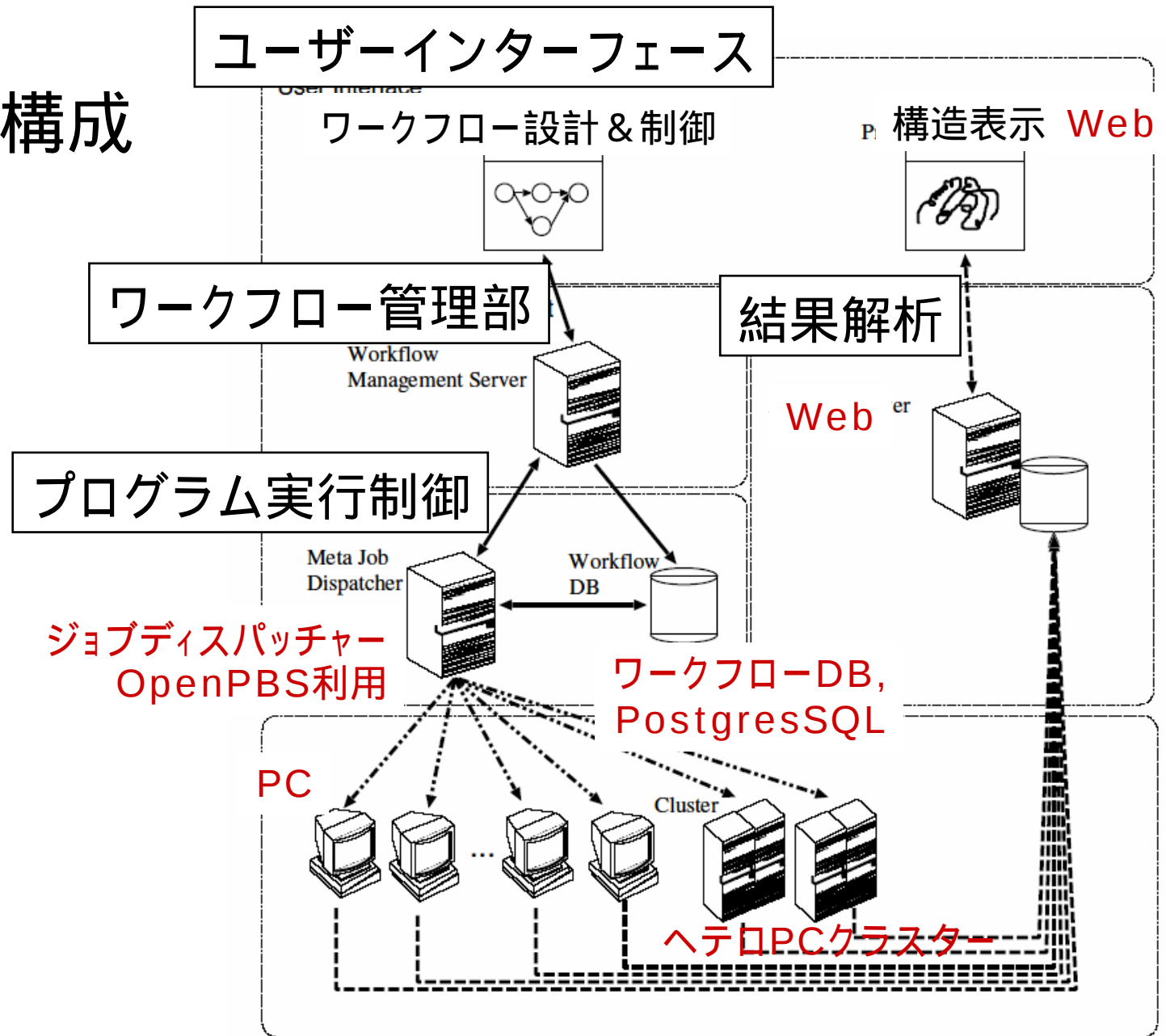
計算結果
解析 & 表示

Grid Environment

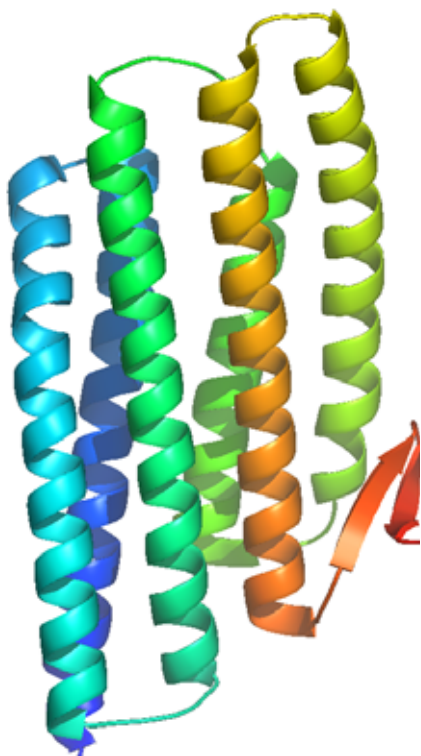
Grid Middleware (Condor, OpenPBS,...)

Computing Nodes (Desktop PC, Cluster...)

システム構成



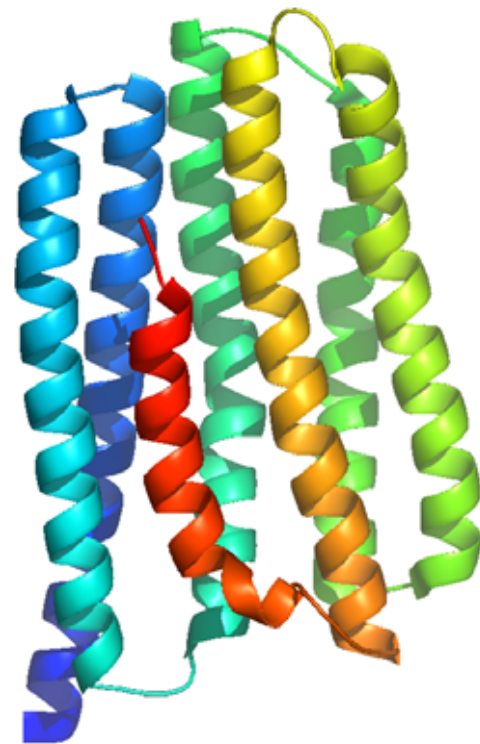
テスト計算 CASP T0198



Answer

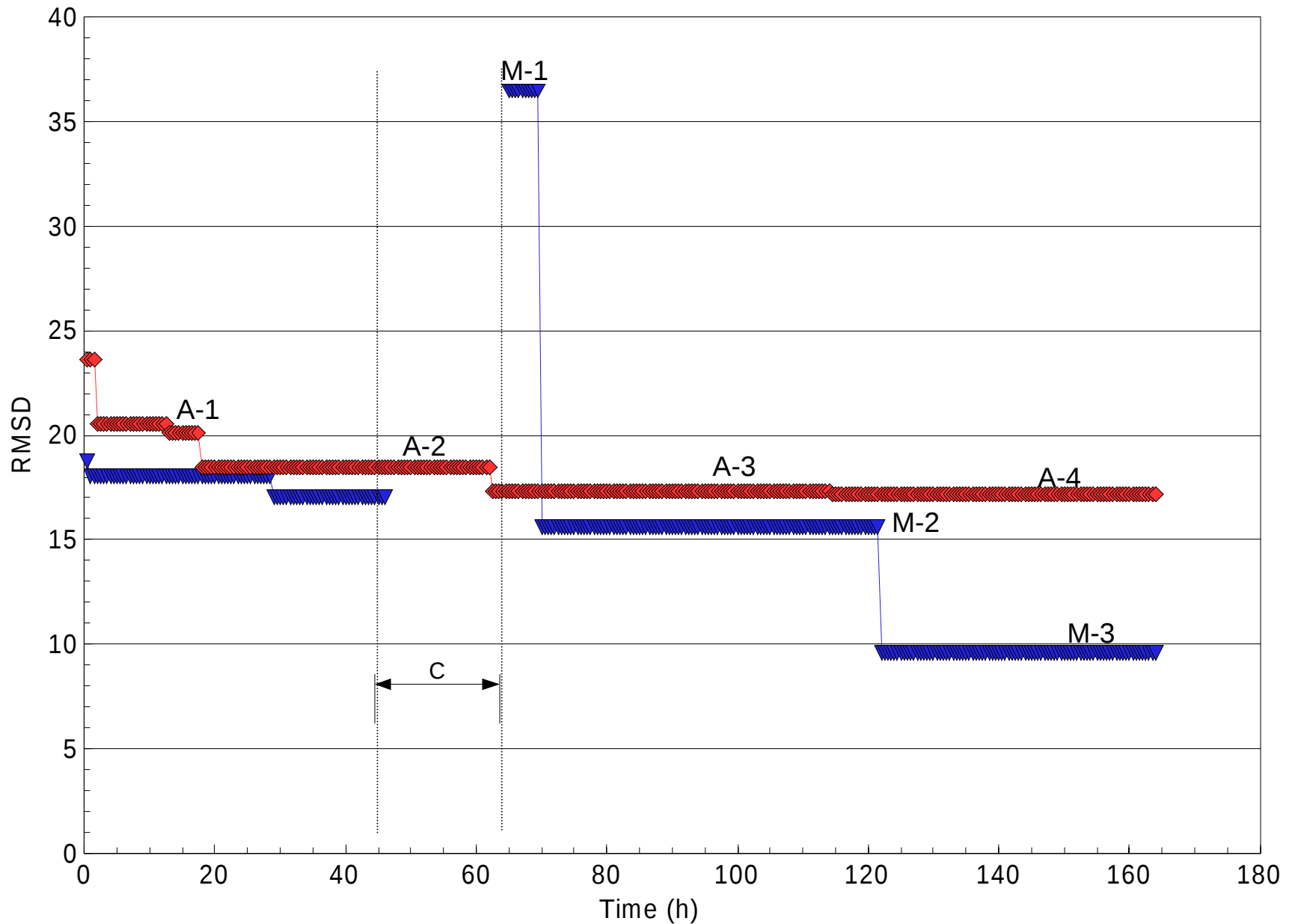


By full automation

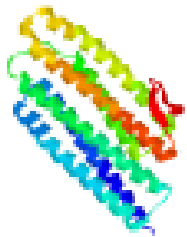


By human intervention

構造予測結果の経時変化



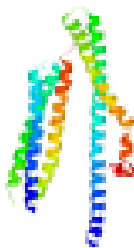
構造予測結果の経時変化



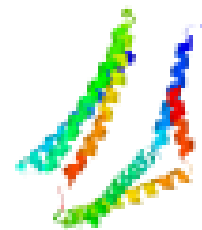
Answer



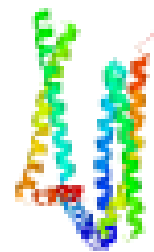
A-1
RMSD: 20.52



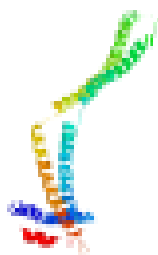
A-2
RMSD: 18.46



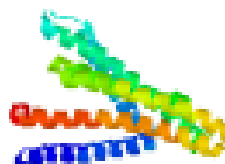
A-3
RMSD: 17.31



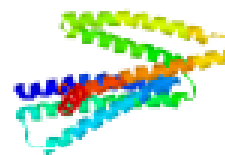
A-4
RMSD: 17.15



M-1
RMSD: 35.98



M-2
RMSD: 15.59



M-3
RMSD: 9.58

まとめと今後

- **全自動**立体構造予測サーバーROKKY
 - CASP6のNFサーバーで世界2位
- **ヒトの介在**で、しばしば予測モデルは改良できる
- 対話的動作による柔軟なワークフロー
 - 予測モデルの改良が見られた.
- より大規模グリッドに実装したい



研究参画者

藤川和利、蟻川浩、増田慎吾(奈良先端)

金文珍 朴聖俊 古田忠臣 高田彰二(神戸大)

上島豊(原研)

フラグメントアセンブリ法

既知構造データベースの**フラグメント**(構造断片)をつなぎ
合わせてターゲット構造をつくる

