

地球シミュレータ利用報告

研究課題: バイオシミュレーション

平成16年1月11日

郷信広

日本原子力研究所
計算科学技術推進センター

バイオシミュレーション

サブテーマ1:「密度汎関数法による超大型タンパク質の全電子計算」

東京大学生産技術研究所客員助教授 佐藤文俊

サブテーマ2:「蛋白質の高次構造変化のリアルなシミュレーション」

弘前大学理工学部教授 斎藤稔

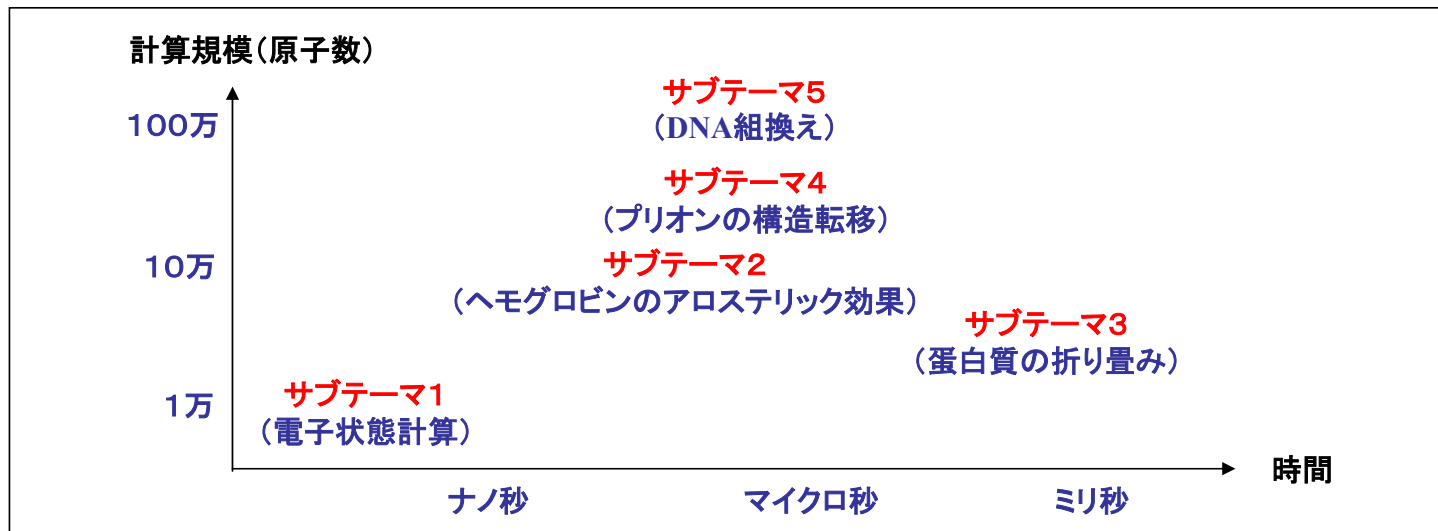
サブテーマ3:「第一原理からのタンパク質の折り畳みシミュレーション」

分子科学研究所助教授 岡本祐幸

サブテーマ4:「正常プリオンタンパク質から異常プリオンタンパク質への構造転移プロセスの解明に関する研究」産業技術総合研究所生命情報科学研究センター長 秋山泰

サブテーマ5:「分子動力学シミュレーションを用いた大規模生体超分子系の機能解析」

日本原子力研究所計算科学技術推進センター 研究員 石田恒



サブテーマ1:「密度汎関数法による超大型タンパク質の全電子計算」

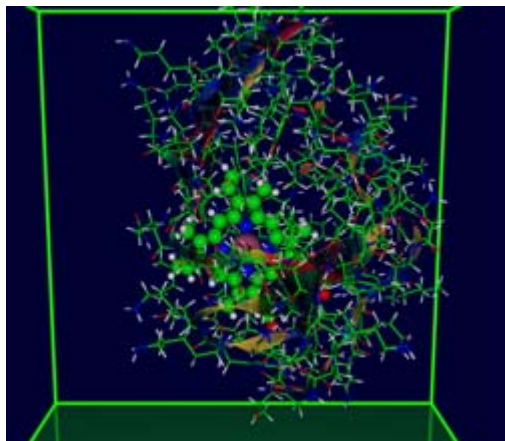
タンパク質全電子計算の生物学的意義

背景

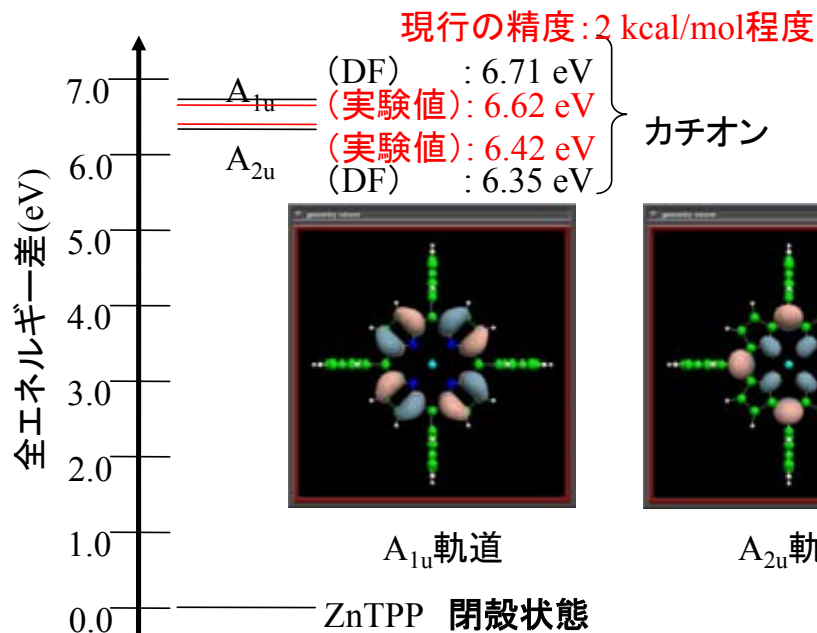
- ・タンパク質固有の化学反応は精度良く求める必要がある
- ・現在のところ、ProteinDFはそれに応える唯一の手段

電子を抜くのに・・・

鉄イオン	700 kcal/mol
ヘム	100
ヘムタンパク質	10



密度汎関数法による計算精度
ZnTPPのイオン化ポテンシャル



タンパク質全電子計算プログラムの工夫

タンパク質全電子計算前提条件:
1構造当たり...

残基数 : 原子数 : 電子数 : パラメータ数*
= 100 : 2,000 : 7,000 : (10,000)² → 1億

*DZ基底関数相当

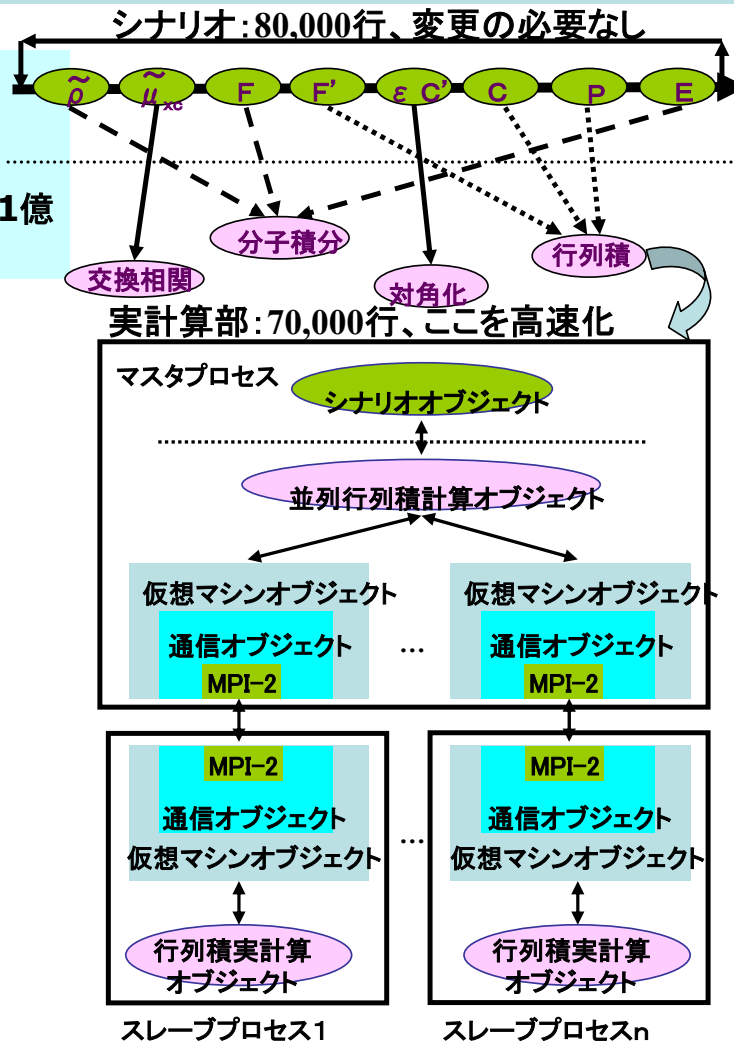
考慮点:

- 分散並列化されている
- 計算機資源を十分に利用する
- 様々な計算機種に対応しやすい構造である
- 複数人によるコーディングに耐えられる
- 各種計算ルーチン間におけるデータの従属性が取り払われている
- ...



工夫点:

- オブジェクト指向プログラミングの採用
- 各ルーチンがメモリを最大限利用するための動的メモリの制御
- 計算シナリオ・実計算(律速)部・並列ライブラリの独立化
- ...



チューニング作業:ベクトル・並列化効率について

・ES用並列ライブラリへの変換(達成率100%)

ProteinDFはMPMD方式。ESのMPI-2の動的プロセス生成の実装が異なっていたため、動的プロセス生成を使用せずにMPMDを実現するようソースを大幅に変更。

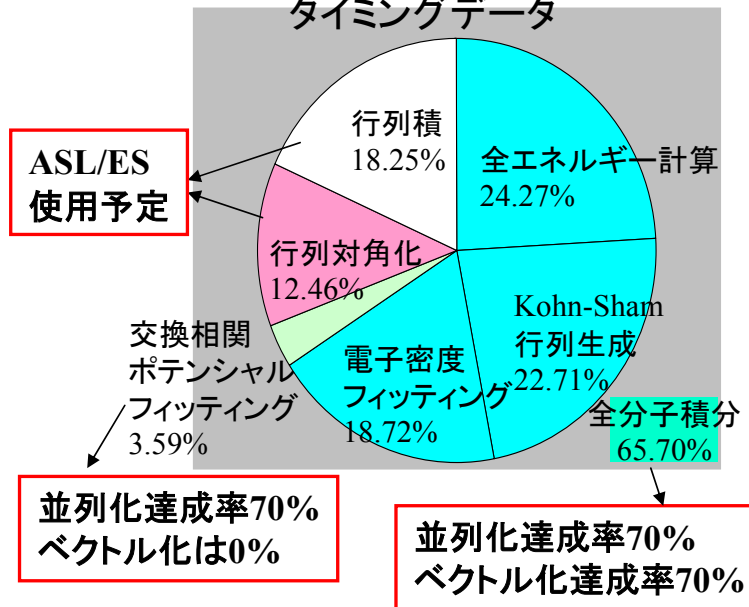
・並列化チューニング(達成率70%)

ベンチマーク計算を行いながらチューニング中。15台のヘテロワークステーションクラスターで95%の並列化効率を達成していたルーチンが、ESでは効率が落ちてしまう原因を追求中。

・ベクトル化チューニング(達成率50%)

ベンチマーク計算を行いながらチューニング中。プロファイルを取れる計算サイズが小規模で、達成度はまだ低い。

WSクラスター上での100残基計算
タイミングデータ



1残基計算における分子積分のベクトル化率

計算ルーチン	ベクトル化率(%)	性能(MFLOPS)
DfEri::auxSet	98.60	2613.3
DfEri::ericalc	91.48	104.6
DfEri::fmtRecursive	99.17	2978.9
DfOverlap::ovpqqcalc	92.05	109.7
DfDensityfitting::calcnSRou	96.60	606.1
DfDensityfitting::calcRamda	95.40	146.6
DfDensityfitting::calcRouAlp	96.87	646.3
DfDensityfitting::calctplusRr	96.96	396.8

31残基タンパク質における 分子積分と交換関連計算の並列化効率

	serial	2	4	8	32
交換関連ポテンシャル	5776	2967	1512	775	235
		0.97	0.96	0.93	0.77
分子積分(DfEri, DfOverlap)	21935	11839	6450	3607	964
		0.93	0.85	0.76	0.71
分子積分(DfDensityfitting)	13607	5699	3007	1722	442
		1.19	1.13	0.99	0.96

サブテーマ3:「第一原理からのタンパク質の折り畳みシミュレーション」

生物学的意義

2003年4月 ヒトゲノム解読完了宣言
約3万個の「遺伝子」(ORF)

ポストゲノム時代の最重要課題

ヒトゲノム計画によって決定された、DNAの塩基配列
(すなわち、蛋白質のアミノ酸配列)の情報から、
それぞれの蛋白質がどういう立体構造(形)をして
いて、どういう機能(働き)をするかを探ること。

本研究課題では、まず、蛋白質の立体構造をそのアミノ酸配列
の情報のみから計算機シミュレーションによって予測することを
目指す。

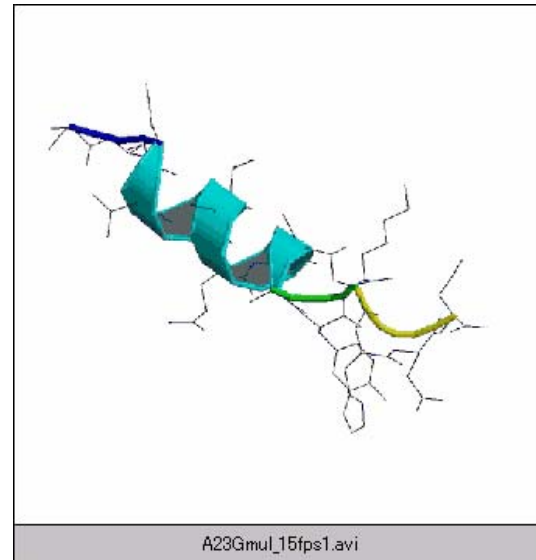
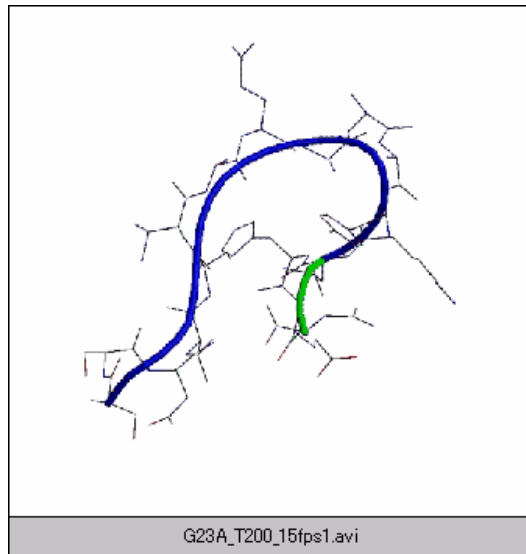
溶媒分子も含めて、原子レベルの詳細を取り入れた
分子シミュレーションに基づいているのが特徴。

計算上の工夫

拡張アンサンブル法という強力なシミュレーション手法を導入したこと。

拡張アンサンブル法 (Generalized-Ensemble Algorithms) エネルギー空間上のランダムウォークを実現することにより、エネルギー極小状態に留まるのを避ける。

[拡張アンサンブル法の例: マルチカノニカル法、レプリカ交換法など]



プログラムのチューニング

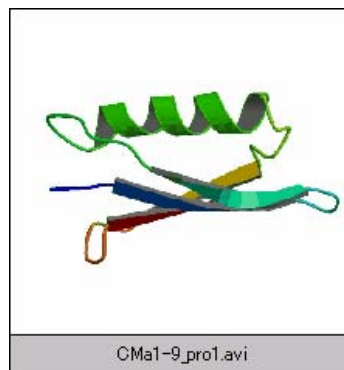
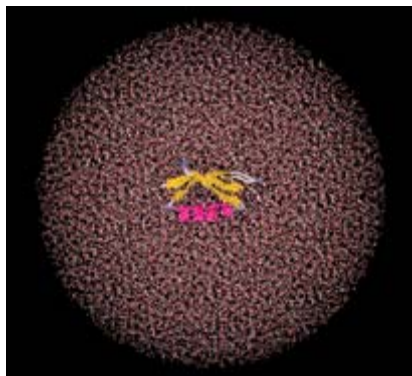
水中(水分子17,187個)のProtein G(全原子数52,416 個の系)において、**レプリカ交換分子動力学法(REMD)**によるシミュレーションのプログラムを、地球シミュレータ向けにチューニングし、現在、112ノードの本申請中。

ノード数	ベクトル化率	並列化率	並列化効率
14	98.1	99.9179	91.65
28	98.0	99.9179	84.53
56	98.0	99.9996	84.87
112	97.8	99.9996	73.70

最終目標のテストラン

水分子17,187個を含むProtein Gの系
(全原子数52,416個)

180 psec のレプリカ交換
分子動力学シミュレーション:(タンパク質のみ描画)



CMa1-9_pro1.avi

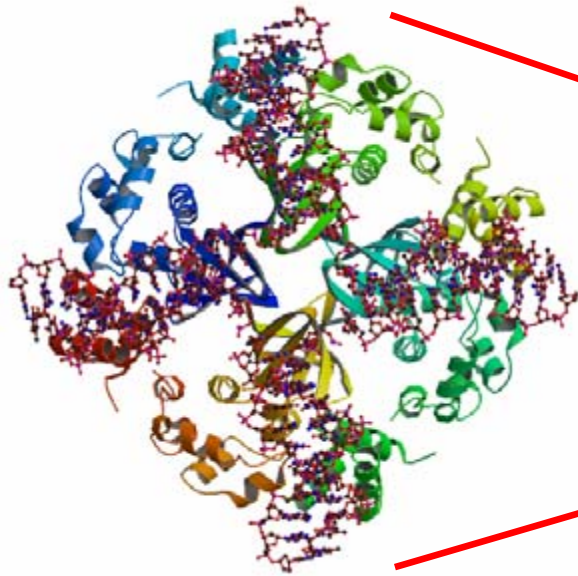
サブテーマ5:

分子動力学シミュレーションを用いた大規模生体超分子系の機能解析

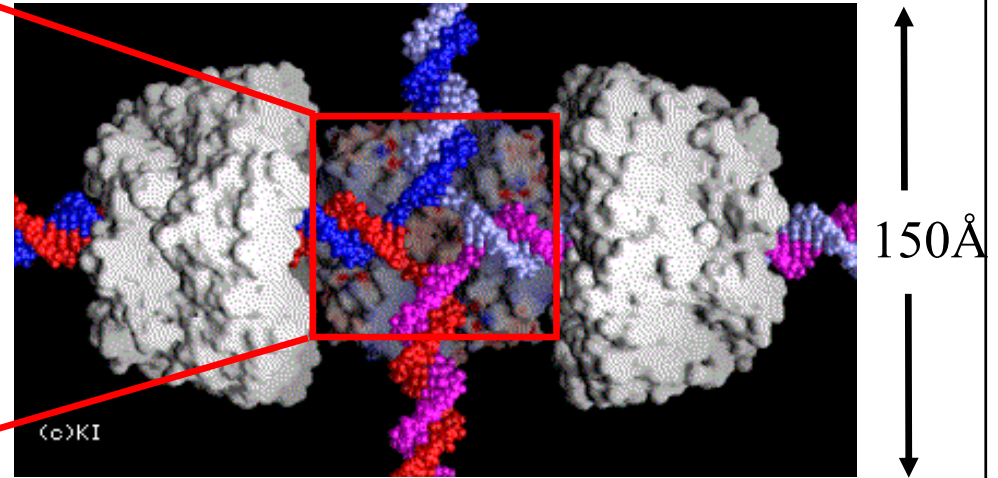
生物学的意義: DNA組換えのダイナミクス解明

DNA組換えのモデル

300Å



DNAに結合したRuvA蛋白質



DNAに結合したRuvAB蛋白質

West, SC. Cell, (1998)

巨大生体高分子のシミュレーションシステム **PABIOS** (PArallel BIOSimulator) の開発

[シミュレーションの現状]
生体高分子系 (1~2分子)
 10^4 粒子、 10^6 ステップ
のシミュレーション

大規模化、高速化

スーパーコンピュータ

生体高分子複合体系
 10^6 粒子、 10^6 ステップ以上
のシミュレーション

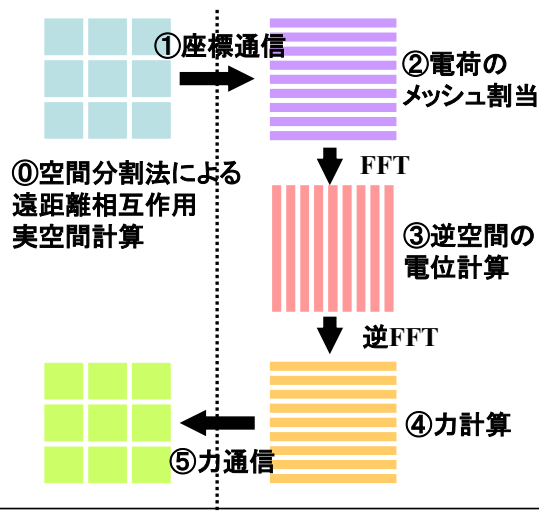
PABIOSの高速化・高精度化

並列化 .. 多数(100以上)のPEを使用
空間分割法による通信コスト削減

長距離相互作用 .. 計算量の多い部分を高速化
PPPM (Particle-Particle Particle-Mesh)法

高精度の時間積分法を採用
Multi Time Step法による時間ステップ幅の増大

PABIOSにおける遠距離相互作用計算
(実空間計算① + 逆空間計算①-⑤)



地球シミュレータ上におけるPABIOSの計算性能

並列化及びベクトル化チューニング

計算負荷が高い長距離静電相互作用計算ルーチンを重点的に並列ベクトルチューニング

18ノード(144CPU)計算

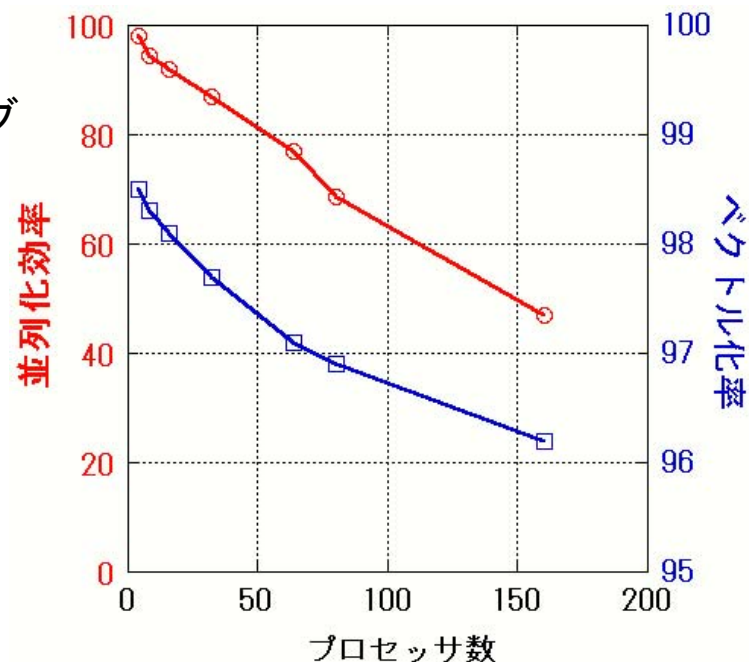
並列化効率50.3%

ベクトル化率96.4%

(現在もチューニング中)

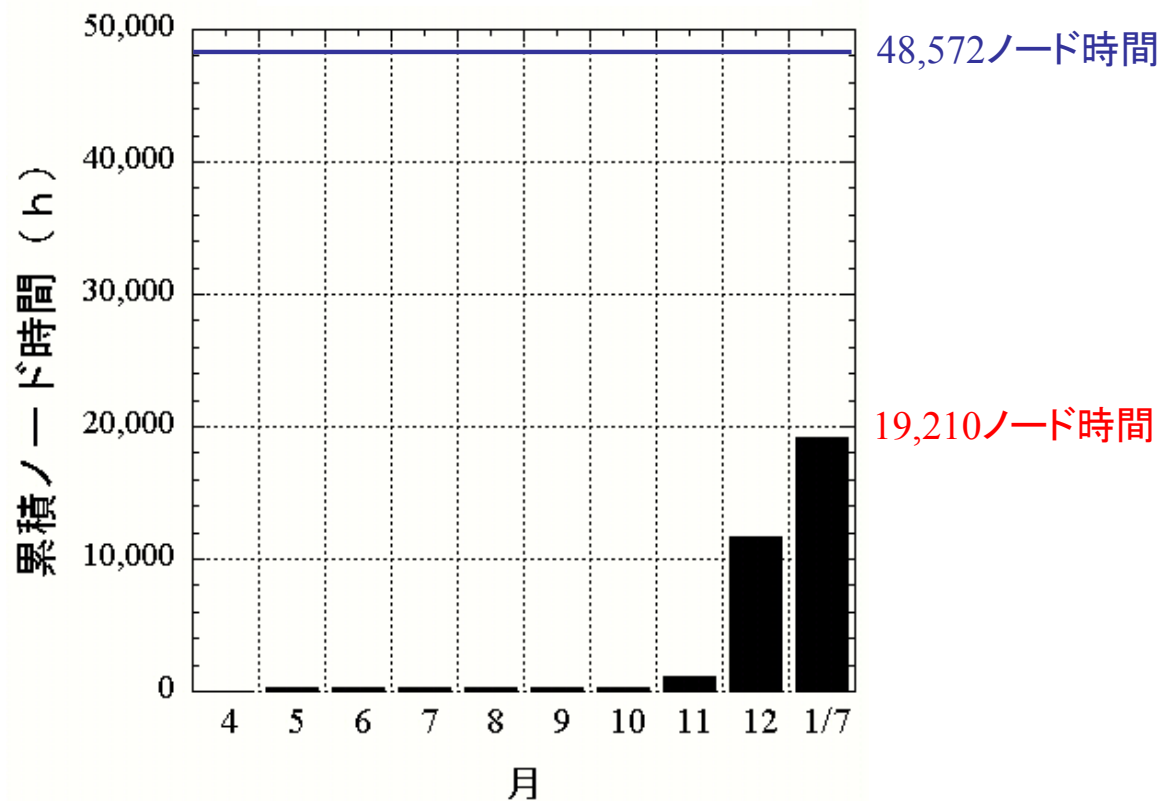


更なる大規模な系であるRuvAB-Holliday分岐DNAの系(約200万原子)でも十分計算性能を出せる



RuvA-Holliday分岐DNA複合体の系(原子数166,177原子)

地球シミュレータ利用実績(平成16年1月7日現在)



1月7日までの利用時間: 19,210ノード時間
(年度末までの割当時間: 48,572ノード時間)

今後の利用継続の必要性

これまでの成果

ベクトル化並列化チューニングによる
計算性能の向上

今後

本格的な大規模計算を開始
生体物質の機能を解明