



生命科学のグランドチャレンジで開発中のソフトウェアの現状とそのターゲット

次世代生命体統合シミュレーション研究開発グループ
グループディレクタ

姫野龍太郎



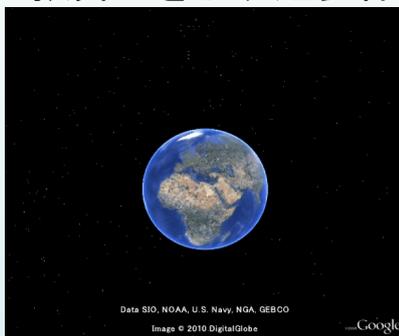
この研究開発の位置づけ 次世代スパコン開発プロジェクトの一部

現在

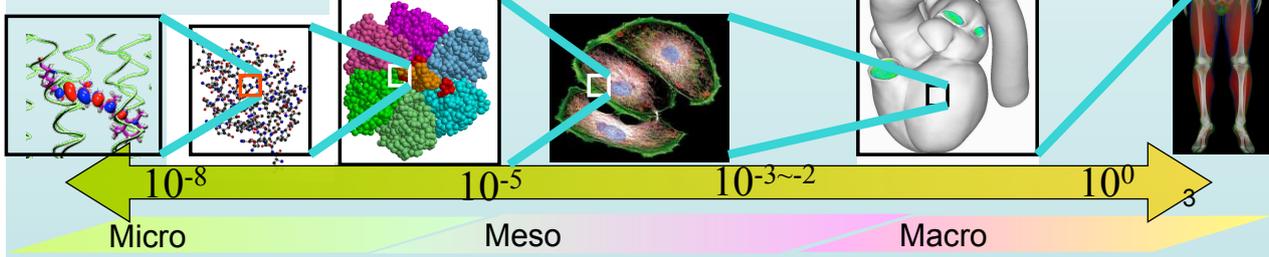
		平成18年度 (2006)	平成19年度 (2007)	平成20年度 (2008)	平成21年度 (2009)	平成22年度 (2010)	平成23年度 (2011)	平成24年度 (2012)
システム		概念設計		詳細設計		試作・評価・製造		性能チューニング
ソフトウェア (グランドチャレンジ) アプリケーション	次世代ナノ統合シミュレーション	開発・製作・評価					実証	↑ 供用開始
	次世代生命体統合シミュレーション	開発・製作・評価					実証	
施設	計算機棟	設計		建設				
	研究棟	設計		建設				
								2

生命現象は最も複雑で難しい問題

複雑で美しい振舞いを示す **超**多体系多階層問題

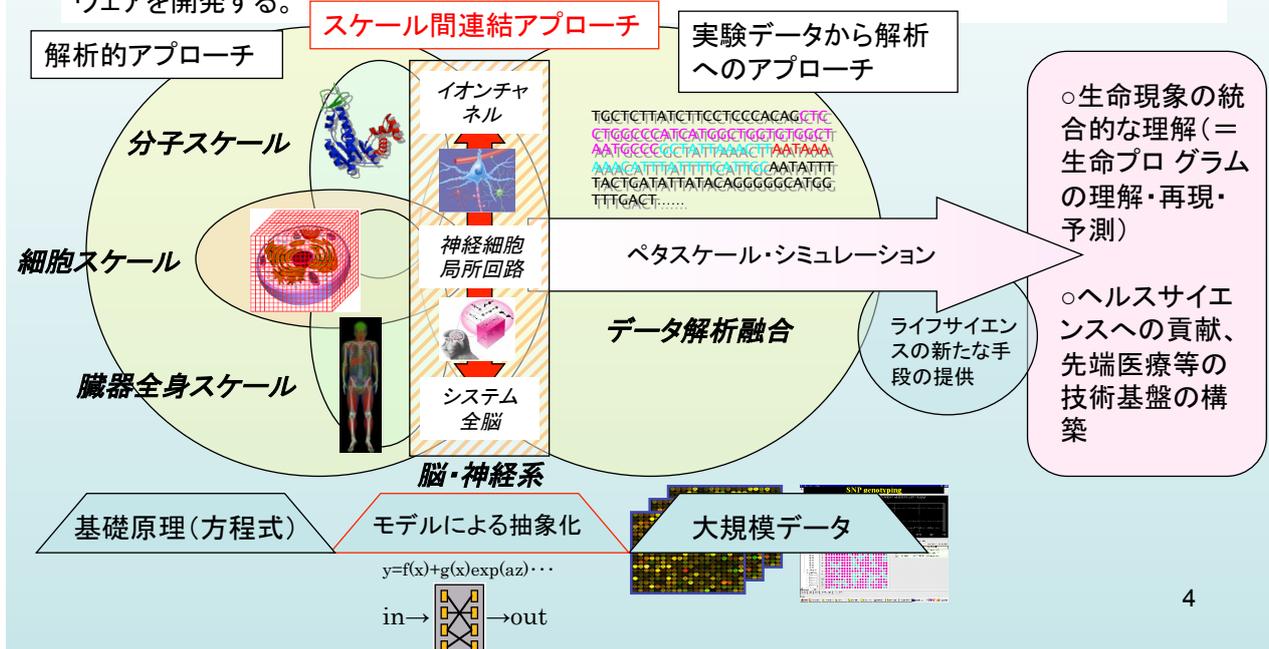


スーパーコンピュータを使って、この複雑な生命現象を解析
記述する生物学から、予測する生物学へ



研究開発の概要と達成目標 (2008年10月以降)

基礎方程式に基づく解析的アプローチと、大量の実験データから未知の法則に迫る実験データから解析へのアプローチ、さらには多階層を連結するアプローチにより、異なるスケールの研究と実験データを統合的かつ有機的に結びつけ、ペタスケールという桁違いの性能を持つスーパーコンピュータの性能をフルに発揮し、生体で起こる種々の現象を理解し医療に貢献するためのソフトウェアを開発する。





各チームの主ターゲットと チーム間連携

赤枠は大きく進展した
もの、血栓形成は必要
性が高まり、今後加速



ISL&M

受 提供	データ解析融 合T	分子スケ ールT	細胞スケ ールT	臓器全身 スケールT	脳神経系T
データ解析融 合T	肺がんと薬	データ同化 技術	生体分子ネ ットワー クモデル	画像処理 パラメータ 推定	パラメータ推 定(モデル化)
分子スケ ールT	タンパク質- リガンド	代謝・膜	膜での輸 送モデル	血栓形成	神経可塑性 ・変性・イオ ンチャネル
細胞スケ ールT	生体分子ネ ットワー クデータ	代謝反応・ 膜での物 質輸送データ	肝細胞	複数細胞 から臓器へ ・血栓形成	細胞プラッ トフォーム 供給
臓器全身 スケールT	医療画 像データ	データ提供 ・ドラッグデ リバリ	血流・分泌 物輸送	循環器系	細胞の変形
脳神経系T	神経細 胞データ・ 局所回 路データ	可塑性や 神経変性 に関するデ ータ	成長・変形 モデル	運動制御 臓器制御	細胞から 高次機能 までの一貫
高度化T	分析・高速化	分析・高速化	分析・高速化	分析・高速化	分析・高速化

5



2008年の見直し 開発するソフトウェア、短期目標



ISL&M
整理前

計算手法

計算対象

計算対象

プログラム

分子スケ ール	QM/MM 全電子量子化学計算 ハイブリッドQM/MM法 分子動力学計算法 粗視化シミュレーション 蛋白質の構造サンプリング技術	共通的な 開発対象 の設定	分子スケ ール	QM/MM 全電子量子化学計算 膜タンパク質・代謝酵素 分子動力学計算法 粗視化シミュレーション 蛋白質の構造サンプリング技	長期目標 に向けた課題
細胞スケ ール	細胞シミュレーション統合プラットフォーム 肝細胞(代謝) 膵臓β細胞 イオンチャネル 血小板		細胞スケ ール	細胞シミュレーション統合プラットフォーム 肝細胞・肝小葉 (代謝)	
臓器全身 スケール	血流シミュレーション 心臓シミュレーション 全身ホクセルデータ・全身力学モデル 超音波伝搬シミュレーション 肺シミュレーション		臓器全身 スケール	循環器系シミュレーション (血球・血小板・血流・血管網・心臓) 全身ホクセルデータ・全身力学モデル 超音波伝搬シミュレーション	
データ解 析融合	大規模遺伝子ネットワーク探索ソフトウェア 生命体データ同化プログラム タンパク質ドッキング解析プログラム 大規模SNP解析ソフトウェア		データ解 析	大規模遺伝子ネットワーク探索ソフト 生命体同化プログラム 肺がんと薬 タンパク質ドッキング解析プログラム 大規模SNP解析ソフトウェア	
		共通的な 開発対象 の設定	脳神経	神経細胞・局所回路シミュレーション 脳全体シミュレーション	

短期目標

長期目標

他チームとの連携に
よる効率化

2010.12.6

6



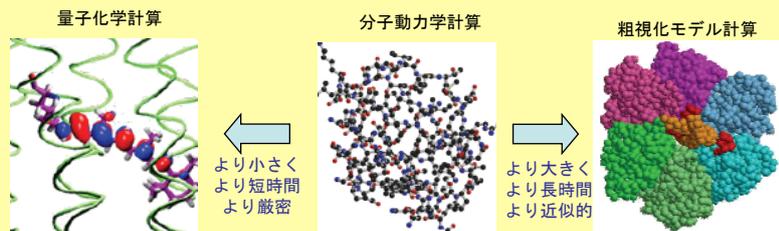
分子スケール研究開発



開発目標:量子化学計算(QM)・分子動力学計算(MM)・粗視化モデル計算(CG)を総合化する技術を開発、タンパク質や細胞の機能発現過程のシミュレーションにつなげ、細胞スケールとの有機的な連携を図る。

ターゲット:多剤排出トランスポータ、脂肪酸代謝酵素反応

開発計画:QM、MM、CGの各スケールでプログラムの開発を進めるとともに、それらの手法を結合したQM/MM、MM/CG法によってマルチスケールシミュレーションを実現するためのプログラムを開発する。



進捗状況:QM/MM/CGの3階層それぞれで計算方法をプログラム化し、テストを続けながら開発、8000並列程度まで性能が出るまでになっている。

7

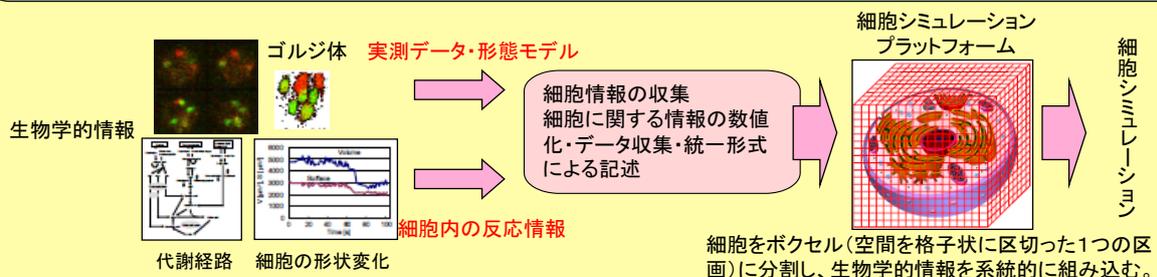


細胞スケール研究開発



開発目標:ペタフロップス級の計算力を活かして細胞を100万のボクセルに区画、そこに包括的な実証データを組み込んだ**肝細胞・肝小葉のシミュレーションの実現**を目指す。薬剤等の作用や副作用の予測を行うシミュレーションの開発につなげる。

開発計画:生物の空間的情報を入れ込むことが出来るプラットフォームの開発と平行して、プラットフォームに実装するためのデータ収集とモデルの構築を進める。また、プラットフォームと各種データ・モデルとの有機的な連携を図る。



進捗状況:細胞を約3万個のボクセルに分割、複数のオルガネラを表現し、移流拡散方程式による物質移動をシミュレーションできる細胞シミュレーションプラットフォームを開発、E-CELL3をベースにした代謝反応がオルガネラ間も含めてシミュレーションできるようになった。現在、8192並列までの性能テストを終了

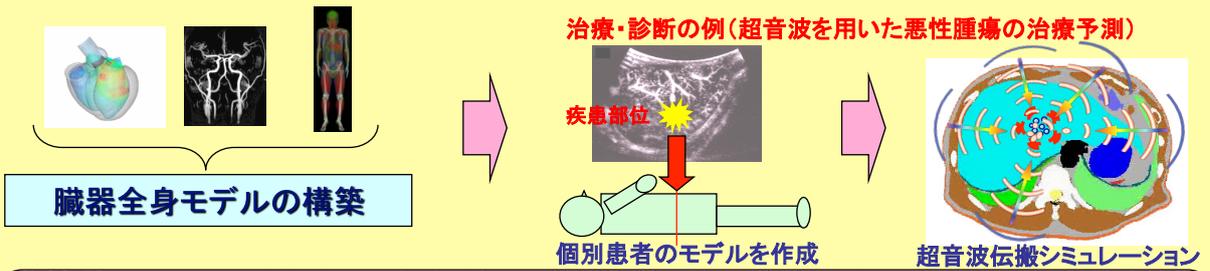


臓器全身スケール研究開発



開発目標:ペタフロップス級の計算量を用いて、数時間で実行可能な臓器全身シミュレーションの構築を目指す。これにより、病態予測や治療予測等の医療支援ツールの開発につなげる。

開発計画:血管網、各種臓器、全身を3次的に再現した臓器全身モデルを構築し、病態予測や治療予測を行う臓器全身モデルを次世代スパコンに実装、**循環器系疾患**を始めとする種々の病態のシミュレーションと、**ガンなどの治療支援シミュレーション**(放射線・集束超音波など)につなげる。



進捗状況:1mm分解能の全身ボクセルモデルを開発すると同時に新しい構造流体連成手法と多媒質内の超音波伝搬解析手法を開発した。これらにより、外科手術シミュレーションや収束超音波・重イオンビームによるガンの治療等へつなげる基盤を構築した。

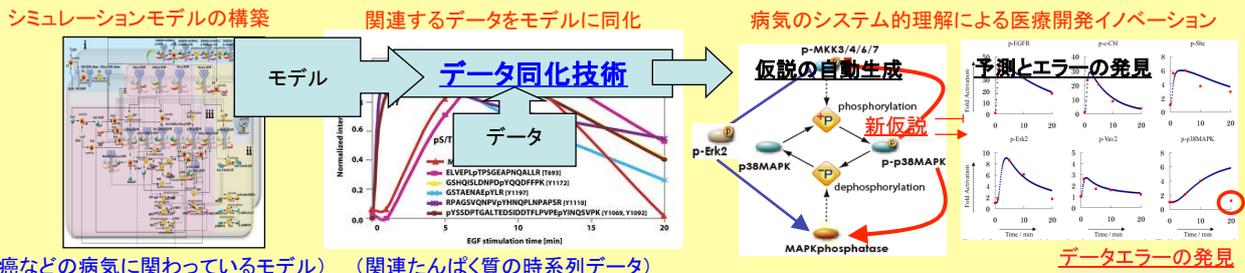


データ解析融合研究開発



達成目標:ペタフロップス級の計算によって、創薬ターゲット探索や個人差を考慮した医療のための基盤情報技術の構築を目指す。これにより、ヒト全遺伝子を対象とした創薬ターゲット遺伝子探索の実現につなげる。

開発計画:遺伝子ネットワーク推定・タンパク質ネットワーク推定・ゲノム多型解析技術等の研究、ならびにデータ同化技術の開発を進め、開発した解析法やデータ同化技術を次世代スパコンに実装し、データ解析とシミュレーションを融合した技術を開発する。**ターゲット:肺ガンと薬**



進捗状況:分子のネットワークを地図として抽出するための方式として、大規模遺伝子ネットワーク探索及びタンパク質構造に基づく相互作用予測の研究を行い、新たな技術開発の成功と新たな並列プログラムの開発により、これらのプログラムが8000を超えるcoreで稼動するようになった。



脳神経系研究開発

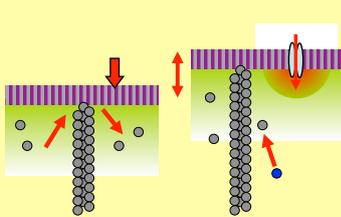


開発目標: 脳全体レベルでの入出力関係を再現できるモデルとシミュレーションソフトウェアを開発、次世代スーパーコンピュータの性能を引き出して、脳の柔軟な環境適応能力である脳の発達と学習の再現を目指す。

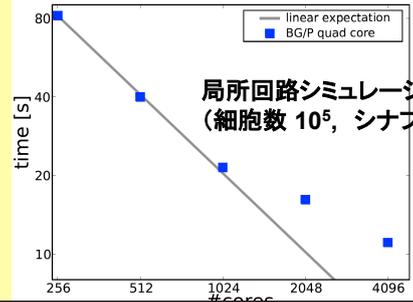
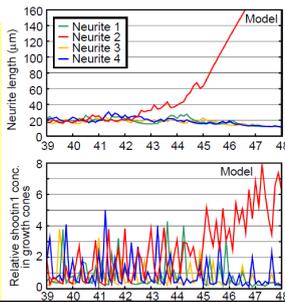
開発計画: 神経細胞と局所回路のシミュレータ、および、脳のシミュレータとして昆虫の匂い情報処理回路モデル、ほ乳類の網膜モデルを開発、次世代スーパーコンピュータに適したプログラムとすることで、昆虫の嗅覚系情報処理では実時間で、ほ乳類の視覚系情報処理は60倍の時間で計算可能とする。

神経マルチフィジクスシミュレータ

局所回路シミュレータNEST



細胞極性形成シミュレーション



進捗状況: 本チームは2008年10月に発足、昨年度は神経マルチフィジクスシミュレータでは神経細胞の極性形成について、分子レベルからのシミュレーションができた。また、局所回路シミュレータNESTでは10⁵個の神経細胞、10⁹個のシナプスからなる大規模局所回路のシミュレーションができた。並列性能に関しては32000並列までのテストを終了



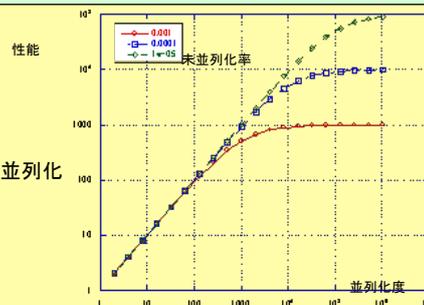
生命体基盤ソフトウェア開発・高度化



開発目標: プロジェクトで開発される各種ソフトウェアの高性能化、特に次世代スーパーコンピュータ向けの最適化を行い、次世代スーパーコンピュータの生命科学への応用を促進すると同時にグランドチャレンジを実現する。

開発計画: コアソフトウェア・基盤ライブラリの開発と他チームアプリケーションの調査・高速化に向けた検討を進め、次世代スーパーコンピュータ実機でのアプリケーションの高速化、チューニングを行う。

アムダールの法則
並列化度が上がるにつれ、わずかな未並列化でも性能に大きな影響を及ぼす



進捗状況: 本チームは2007年11月に発足、これまでに分子動力学計算高速コアソフトウェア(cppmd)の解析・開発、量子化学計算ソフトウェアの高性能化を行うと同時に、基盤ライブラリ・可視化ソフトウェアの調査・開発を進めた。Cppmdでは8000並列、可視化ソフトでも100並列での動作テストを行った。

- 分子動力学や量子力学用ソフトの高速コアライブラリーの開発
- 並列ミドルウェア・超大規模データ処理可視化システム開発(領域型解法、粒子ベースの解法用。ライフサイエンス分野を主な対象としているが、汎用的に使える設計)
- 高並列化・高速化対応(各チームは1,000並列まで自助努力)

ハードウェアに寄り添った最適化

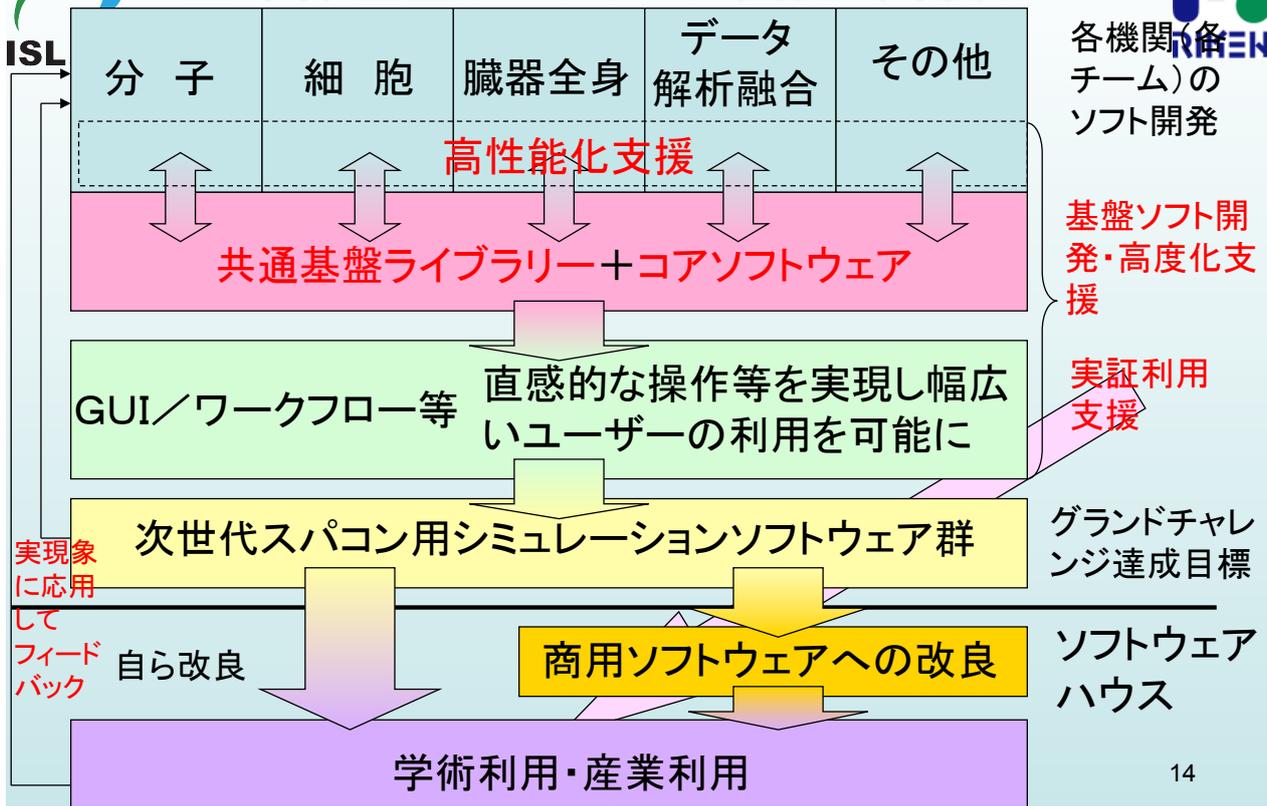
- SIMD機構の最大利用:アセンブラ・コーディング
- マルチスレッド並列とMPI並列の両方の活用
- MPIよりも低レベルの通信ライブラリの活用
- オンキャッシュ計算(キャッシュブロッキング、プリフェッチ)
- 計算と通信のオーバーラッピング
- ロードバランシング

ハードウェアに向けた計算アルゴリズム

- 生体高分子の原子間の長距離力を求めるポアソン方程式
→ FFTから実空間マルチグリッド計算
- 酵素反応の機能解析・薬剤の正確な結合自由エネルギー予測:量子化学計算でのオーダーN手法

Knowledge Data Sharing Cooperative Work ツール群のスクリプト結合

「生命体基盤ソフトウェア開発・高度化チーム」

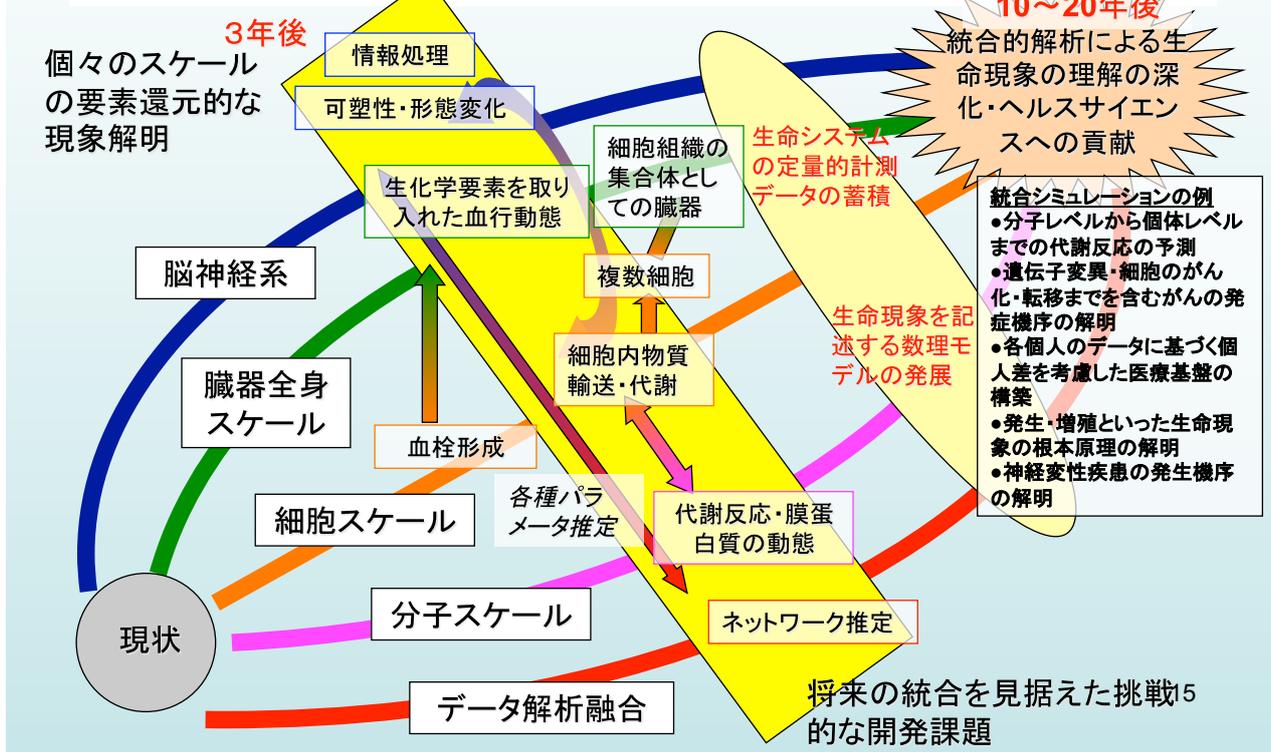




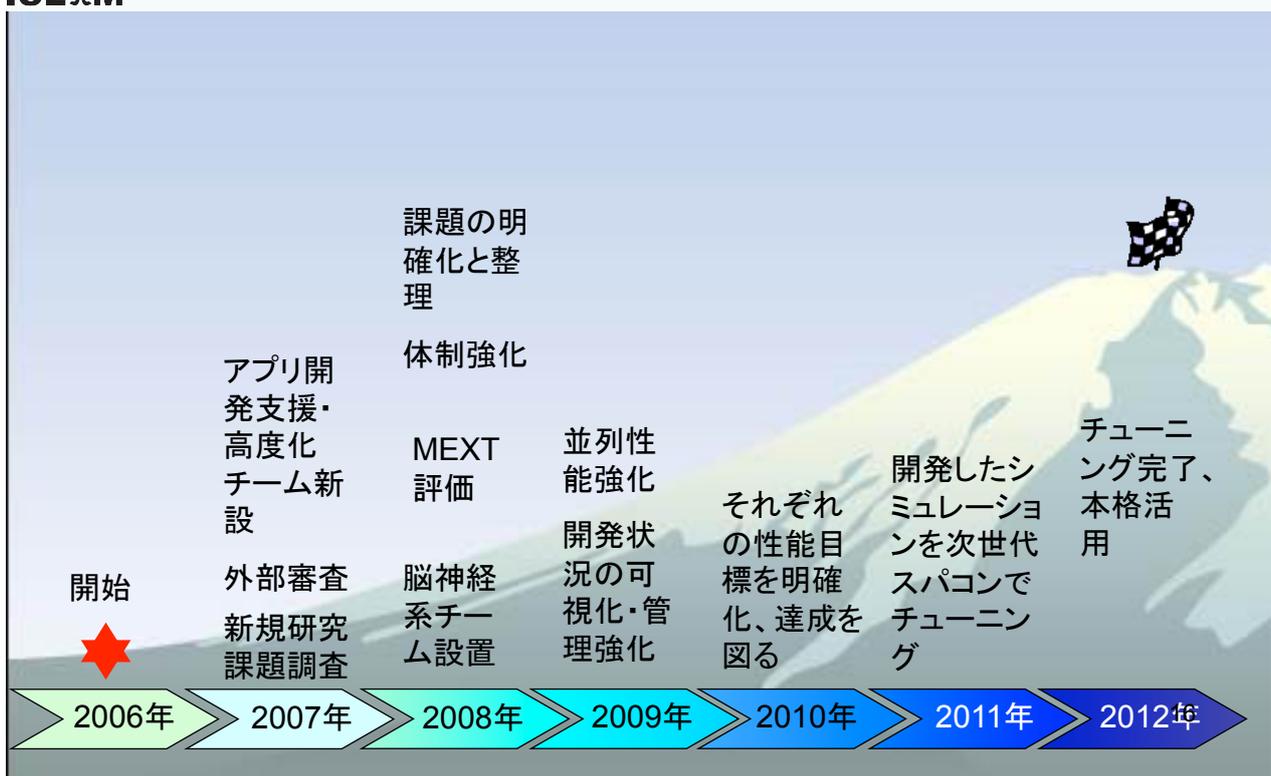
チーム目標と共通目標(モデル化を目指す現象)



「生命現象の統合的解明」に向け、スケールを超えた連携を視野に入れつつサイエンスとしてインパクトのある研究開発課題を検討



これまでの経緯と今後の予定



2. ライフサイエンス分野における 計算科学研究開発の世界の現状

17

17

1) 世界的な研究の中心は

- 三極：アメリカと欧州（+豪州、ニュージーランド）、日本

アメリカ：DoE、NIH、DoD、民間企業（ベンチャー、製薬企業）が主、DoEが基礎科学（生物学＋計算科学）、NIHが医療を担当するため、互いの協力は希薄。計算科学と医療をつなぐ動きがない。

欧州：EUがスポンサーとなり、IT技術を医療に生かす取り組みを推進。一例：VPH(Virtual Physiological Human)、個別患者の治療が目標。大学・医療機関・製薬企業・医療機器メーカーが参加、マッチングファンド、5年で数百億円規模。脳科学ではスイスがBlueBrainプロジェクトを推進。世界をリード。



SciDAC

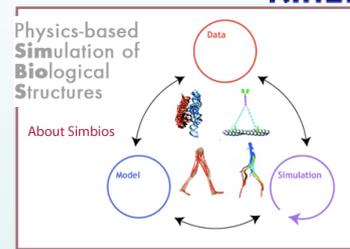
Scientific Discovery through Advanced Computing

18

18

2) 海外での研究開発の今後の方向

- 世界的に取り組まれているのは
 - タンパクの構造や変化などの解析
 - 大量の実験データに基づく多次元因子解析など
- 現在の研究開発でホットな話題はマルチスケール
 - NIHもEUでも二つ以上のスケールをつなぐ研究開発にファンディング
 - 日本は**グランドチャレンジ**として既にマルチスケールに挑戦
 - ・方程式に忠実にモデル化し、スケールを超える
 - ・膨大な計算量は**次世代スパコン**で解消



<http://simbios.stanford.edu/>

欧米では医療機関で使うことを前提に**簡略化したモデル化**。
(脳科学では例外的にスパコンを指向)

戦略分野化で、日本が世界に貢献



http://ec.europa.eu/information_society/newsroom/cf/itemdetail.cfm?item_id=3956



<http://www.nibib.nih.gov/Research/ProgramAreas/MathModeling> 19

3) 国際比較

分子スケール

- アメリカは超並列計算で先行
- 粗視化モデルによる計算、全電子を入れたタンパク質の量子化学計算では日本がリード

細胞スケール

- システムバイオロジー(E-cell等)は日本発
- 細胞群から臓器へ、分子から細胞へのアプローチはどこもこれから

臓器・全身スケール

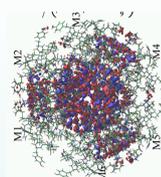
- 国際的な研究開発体制が構築されつつある
- 日本は心臓モデル、健康者からの詳細人体モデルでリード

バイオインフォマティクス

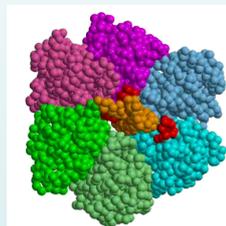
- 個人の多様性と医療をつないでゆくところでは日本は優位
- 次世代超高速シーケンサー等による巨大データに直面、世界的な課題

脳科学

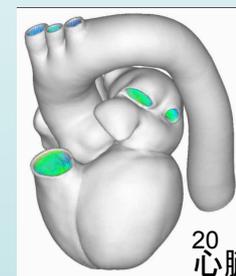
- スーパーコンピュータによるシミュレーションではBlueBrainなどが先行
- しかし、計算論的神経科学では日本が世界をリード、BlueBrainではこの観点が今は抜けている



全電子計算では世界をリード(ProteinDF)



粗視化モデルによる分子モータ



3) 国際比較(続き)

数万並列規模の計算能力をもったソフトウェア開発を推進

今後、次世代スパコンに向けた並列性能向上を加速

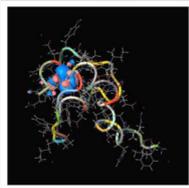
⇒分子・細胞・臓器全身の各スケールと実験データ解析、脳神経系を統合して研究開発

★世界的にも全くない試み

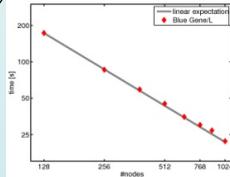
同時に行うことにより、手法やソフトを互いに利用し、開発を加速

⇒次世代スパコンで世界をリードする

↓ チーム間の連携



全電子計算での並列性能 (ProteinDF) 2500コアまでの並列を確認



脳の局所回路シミュレーションでの並列性能 (BlueGeneLで1000並列までの速度向上を確認)

受提供	データ解析融合	分子スケール	細胞スケール	臓器全身スケール	脳神経系
データ解析融合	肺ガンと薬	ターゲットタンパクの特定	生体分子ネットワークモデル	画像処理パラメータ推定	パラメータ推定(モデル化)
分子スケール	タンパク-リガンド	代謝・膜	膜での輸送モデル	血栓形成	神経可塑性・変性・イオンチャンネル
細胞スケール	生体分子ネットワークデータ	代謝反応・膜での物質輸送データ	肝細胞	複数細胞から臓器へ・血栓形成	細胞プラットフォーム供給
臓器全身スケール	医療画像データ	データ提供・ドラッグデリバリー	血流・分泌物輸送	循環器系	全身モデル
脳神経系	神経細胞データ・局所回路データ	可塑性や神経変性に関するデータ	成長・変形モデル	運動制御臓器制御	細胞から高次機能までの一言

3. ソフトウェア開発状況



現在の開発状況



項番	アプリケーション名	略称	開発責任者						2010年4月末	2010年6月14日	
A1-1	マルチコピー・マルチスケール分子シミュレーション法開発の基盤となるクラスライブラリ	Platypus-MM/CG	木寺								
A1-2	レプリカ交換分子動力学計算インターフェイス	Platypus-REIN	杉田								
A2-1	全原子分子動力学計算		現Phase	到達並列数	進み(月)	ハイブリッド並列化	Fjコンパイラ対応(RICC)	Fjコンパイラ対応(FX1)	実行時間の測定	演算量の測定	到達並列数
A3-1	粗視化モデル計算		II-1	4000	-7	×	○	○	×	×	8000
A4-1	密度汎関数法に基づくタンパク質全電		II-1	1024	5	×	○	×	×	×	1024
A5-1											2048
A6-1	量子										8000
A6-2	量子										8000
A6-3	粗視										128
B1-1	細胞										8000
C1-1	全身										2000
C1-2	微視(埋)										128
C1-3	重粒子線治療シミュレーション		II-1	1024	-5	×	×	×	×	×	1024
C2-1	低侵襲治療シミュレーション(ボクセル超音波伝播プログラム)		II-2	8000	6	×	○	○	○	○	8000
C3-1	肺呼吸・肺循環シミュレーション		II-0								512
C6-1	マルチスケール・マルチフィジックス心		II-1								1024
			II-2								8000
			II-0								23
			II-2	8000	4	×	○	○	×	×	8000

管理: 開発フェーズを4つに分類

Phase I: プロトタイプ開発と並列化のテスト(想定2006年~2008年)

Phase II: 大規模並列化(想定2009年~2010年)

Phase III: 超大規模並列+実機チューニング(想定2011年)

Phase IV: 次世代スパコン上での本格的な実行とPR普及(想定2012年)

各I-IIIのフェーズの中に、さらに3つのマイルストーンを設定、10段階に

開発ソフトウェア: 全34本

20本: 1024並列以上、15本: 8000並列以上

可視化ソフトも100GPGPU並列

3ヶ月毎にアップデート、問題があれば機動的に対処



開発ソフトの優先順位とリソースの有効活用



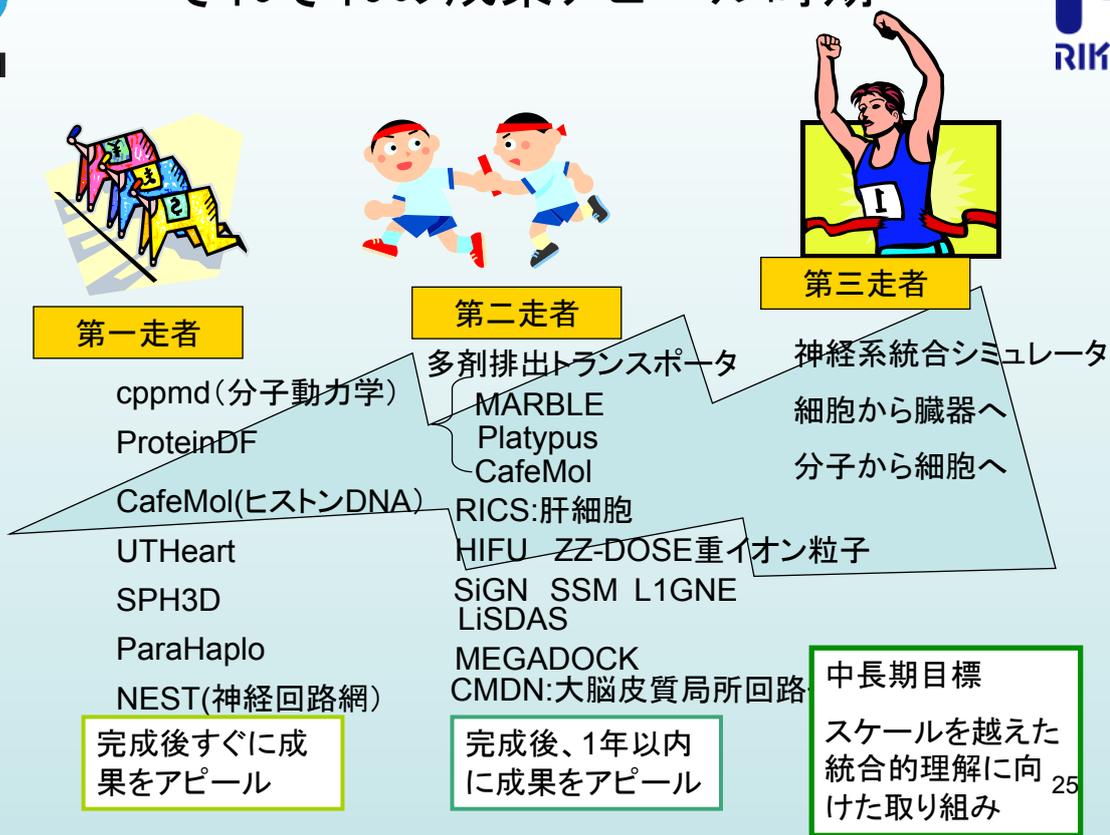
背景

- ・稼働初期(部分稼働~本格稼働の間)にはユーザーが使えるマシン時間は少ない
- ・高度化チームが同時に取り組めるソフトも限られる

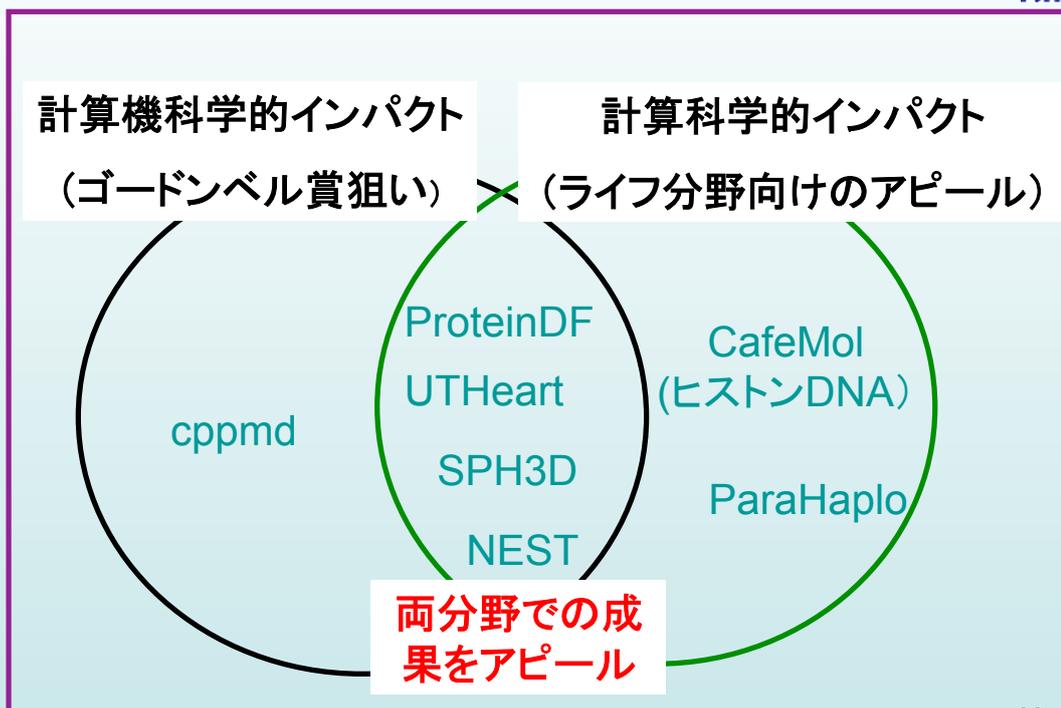
- ・全ソフトの開発進捗をうまくコントロール、フェーズをずらして、リソースを有効利用できるように図る。
- ・開発中の**アプリに優先順位**をつける。
- ・稼働後**短期間に成果を計画的かつ波状的に**出してゆ

く

それぞれの成果アピール時期



第一走者の分類

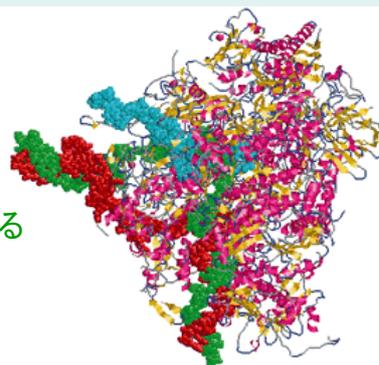


世界最速の分子動力学計算でGordonBell賞を目指す

— タンパク質の機能を高速で長時間シミュレーションするための計算技術 —

ソフトウェアの特徴

- 分子動力学計算ソフトウェア: 10万コア以上の超並列化 (100万-1000万原子系で) 10Petaflopsで1億原子級のシミュレーションを実現し、実効性能で最速を目指す
- 生体分子の運動の長時間シミュレーション
100万原子系のサブミリ秒シミュレーションを目指す
(対象例) 細胞膜構造、核膜輸送、ウイルスなど
【従来の100倍以上の時間スケール】
- これまで不可能だった時間スケールで、生理的に意味のあるタンパク質の大きな構造変化を捉えることにより生命システムの原子レベルでの構成原理を理解すると共に創薬に必要なレベルの計算精度を実現する



大規模分子動力学シミュレーション²⁷

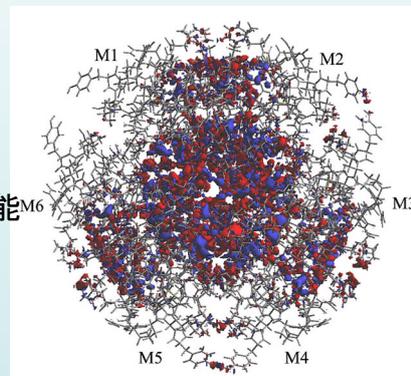
密度汎関数法に基づく タンパク質全電子波動関数計算 (ProteinDF)

世界最大のタンパク質全電子計算を目指す

— タンパク質の電子レベルでの反応を正確に解明するための計算科学技術 —

ソフトウェアの特徴

- 密度汎関数法による全電子波動関数計算ソフトウェア
タンパク質の励起状態を1万並列超で計算
- 巨大なタンパク質の励起状態や電子移動の計算が可能
(1千残基、10万軌道、励起状態計算)
→ Protein Data Bankの99%以上が計算可能に
- ミネラルを含むタンパク質 (酵素やヘムタンパク質) の働きを解明するのに有効
- 電子が主役の高速反応が詳細に解明可能になり、生体模倣光電子素子、人工光合成などへの応用が期待される



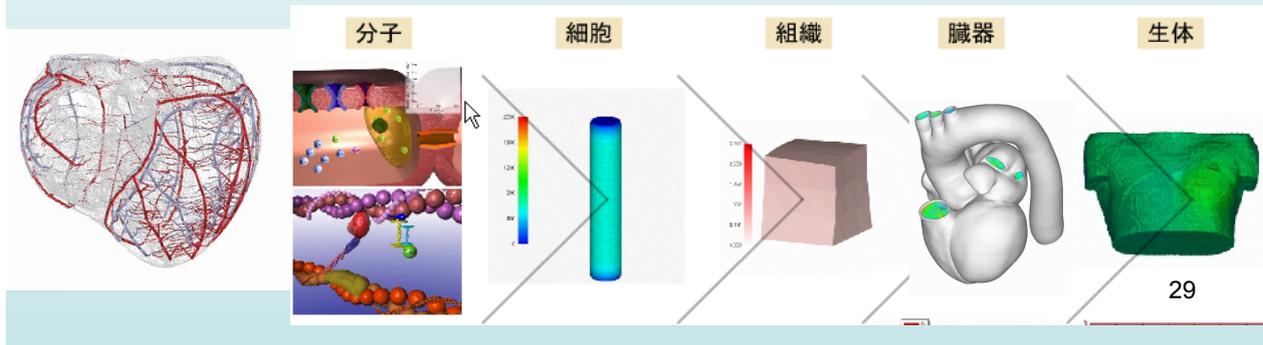
(成果の例)
インスリンの凝集に伴う電子の再分布
6量体と単量体の全電子分布の差を描いたもの
10Petaでは光合成に関係するアンテナタンパクなどがターゲット²⁸

世界最大規模の 心筋細胞レベルからの心臓全体シミュレーションを目指す

— 世界でも前例のないバーチャル心臓を世界最大規模でコンピュータ上に再現 —

ソフトウェアの特徴

- 分子レベルの電気化学、力学現象を再現した、心筋細胞のモデルを使って心臓を組み上げ、心臓の動きと血液の拍動を再現するシミュレーション・ソフト
- 各種心臓病(不整脈や拡張型心筋症など)の基になっている因子を細胞内で起きる分子レベルの挙動から検討することが可能。医療・創薬への応用を検討。



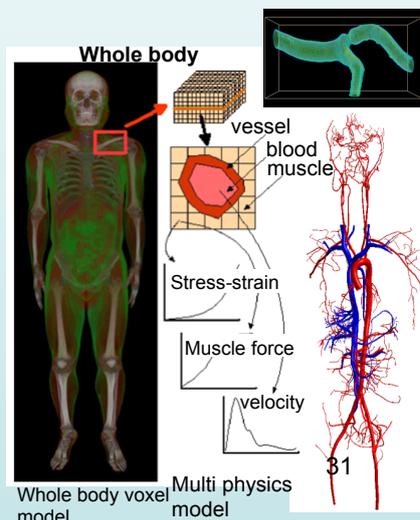
全身ボクセルシミュレーション (ボクセル構造流体連成解析プログラム:SPH3D)

世界最大規模の 血栓の形成、輸送、梗塞の同時シミュレーションを目指す

— 医療応用に向け、やわらかな人体の解析・予測に適した計算科学技術 —

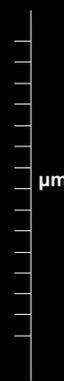
ソフトウェアの特徴

- 大きな血管から毛細血管までの赤血球、血小板を入れた計算ができる構造流体連成ソフトウェア
- 医療データ(CTスキャン,MRI等)から直接シミュレーションを行うことができ、個人毎の診断に即応してシミュレーションが利用可能に
- 血栓形成・閉塞による心筋梗塞, 脳梗塞の予測と薬効評価
- 動脈瘤・動脈硬化症等の循環器系疾患の予測と治療法の検討
- 全身血管網シミュレータとの連成による血圧波形(血圧測定)からの循環器系疾患の予測



血栓症シミュレーション





■東海大学 後藤伸哉 田村典子 七澤洋平

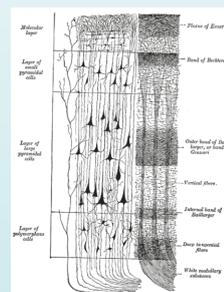
32

世界最大規模の大脳皮質神経回路の活動シミュレーションを目指す

— 単純な神経細胞モデルから実際の脳の活動が説明できるかを解明 —

ソフトウェアの特徴

- 大脳皮質の100コラムに相当する1千万神経細胞、神経細胞間を接続するシナプス1千億の信号処理を再現・予測
(従来の世界トップの100倍の規模)
- 超高並列プログラム(現時点で32,000並列)
次世代スパコン上では数十万並列を達成、**世界最速を目指す**



大脳皮質の情報処理単位
1コラム (1mm³) に
10万個の神経細胞

膨大な長さのDNAが核の中に収納され、必要なときに読み出せる謎に挑む

— 細胞核内でのDNAのヒストンへの巻き付きと、ほどくメカニズムの解明 —

ソフトウェアの特徴

○規模・時間的に巨大な生体分子の現象をシミュレーションするための粗視化モデル計算

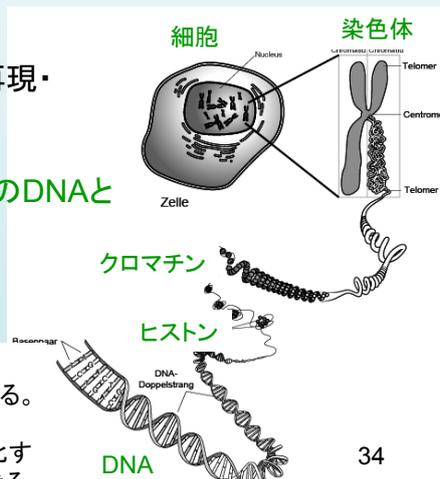
○世界最大1千万原子規模・世界最長のミリ秒の現象を再現・予測を目指す

○学術的なインパクトを与える、遺伝情報が転写される際のDNAとヒストンのダイナミクスを再現

細胞中で膨大な長さのDNAはヒストンに巻き付き、コンパクトに核に収納されている(クロマチン構造)

遺伝子が発現する際には、適切な場所のDNAをヒストンからほどき、遺伝情報を転写している。

iPS細胞の機能は、クロマチン構造による転写制御機構を初期化することによって達成されるといわれている。そのなかにも挑戦できる



47疾患の原因遺伝子を一挙に発見することを目指す

— 個人の遺伝情報と疾病、薬物反応性の関連を明らかに —

ソフトウェアの特徴

○複数の遺伝子や環境要因(身長、体重、喫煙、飲酒など)の組み合わせによって決まる疾病や薬物の効き方、副作用などの関連を網羅的に解析

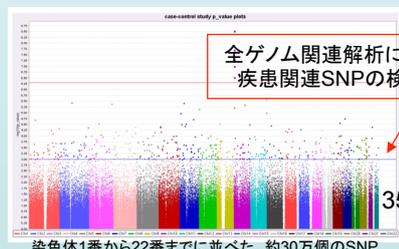
○全ゲノム(30億塩基)の個人ごとの遺伝情報の違いの中から、疾病に関連する遺伝情報を、網羅的に精度よく探し出すことが可能に

○バイオバンクジャパンに登録されている約20万人、47疾患を対象に解析

47疾患

- 悪性腫瘍(肺がん、食道がん、胃がん、大腸がん、肝がん、膵がん、胆嚢・胆管がん、前立腺がん、乳がん、子宮がん、卵巣がん、造血器腫瘍) ●バセドウ病 ●高脂血症 ●糖尿病 ●関節リウマチ ●骨粗しょう症 ●閉塞性動脈硬化症(ASO) ●狭心症 ●心筋梗塞 ●心筋症 ●心不全 ●不整脈 ●脳血管障害(脳梗塞、脳動脈瘤) ●慢性肝疾患(B型・C型慢性肝炎、肝硬変) ●子宮筋腫 ●子宮内膜炎 ●尿路結石 ●ネフローゼ症候群 ●過敏性腸症候群(薬疹) ●ケロイド ●アトピー性皮膚炎 ●花粉症 ●ぜんそく ●結核 ●間質性肺炎 ●肺線維症 ●肺気腫(COPD) ●白内障 ●緑内障 ●熱性けいれん ●てんかん ●歯周病 ●筋萎縮性側索硬化症(ALS)

SNP遺伝子型と疾患との関連を示すアソシエーション検定のP値の-log P値



4. 第一走者以外のソフトウェアの概要

アプリのリスト(1)

	アプリケーション名	略称	開発責任者	言語	並列化の方法
分子	密度汎関数法に基づくタンパク質全電子波動関数計算	ProteinDF	佐藤(東大)	C++	MPI, OpenMP
	粗視化モデル計算	CafeMol	高田(京大)	Fortran90	MPI, OpenMP
	全原子分子動力学計算	MARBLE	池口(横浜市大)	C	MPI
	マルチコピー・マルチスケール分子シミュレーション法開発の基盤となるクラスライブラリ	Platypus-MM/CG	木寺(横浜市大・理研)	C++	MPI
	ハイブリッドQM/MM反応自由エネルギー計算	Platypus-QM/MM-FE	林(京大)	FORTRAN77	GAMESSのsocksライブラリ
	レプリカ交換分子動力学計算インターフェイス	Platypus-REIN	杉田(理研)	Fortna90	MPI
	粗視化モデル計算/分子動力学計算	Platypus-CGM/MM	中村(阪大)	C++, Fortna90	MPI
	量子化学計算/分子動力学計算	Platypus-QM/MM	中村(阪大)	FORTRAN77, Fortran90	MPI
	量子化学計算	Platypus-QM	中村(阪大)	FORTRAN77, Fortran90	MPI
細胞	細胞シミュレーションプラットフォーム	RICS	横田(理研)	Fortran, C, C++	Sphere(MPI)
	全身ボクセルシミュレーション(ボクセル構造流体連成解析プログラム)	SPH3D	高木(東大・理研)	Fortran, C, C++	Sphere(MPI)
臓器全身	マルチスケール・マルチフィジックス心臓シミュレーション	UTHeart	久田(東大)	Fortran	MPI&OpenMP
	低侵襲治療シミュレーション(ボクセル超音波伝播プログラム)	HIFU	高木・松本(東大)	Fortran90	MPI&OpenMP
	微小循環シミュレーター(埋め込み境界法による微小循環プログラム)	ZZ-RBC	高木(東大・理研)	Fortran90	Sphere(MPI)
	重粒子線治療シミュレーション	ZZ-DOSE	高木(東大・理研)	Fortran90	MPI
	肺呼吸・肺循環シミュレーション	ZZ-LUNG	和田(阪大)	Fortran, C, C++	Sphere(MPI)

データ解析融合	ハプロタイプ関連解析に於ける統計検定を行うためのソフトウェア	ParaHaplo	鎌谷(理研)	C	MPI, OpenMP
	大規模遺伝子制御ネットワーク推定プログラム	SIGN	宮野(東大)	C99	MPI, EP (SGE)
	再帰的正則化法による生体内分子の大規模ネットワーク推定プログラム	LIGN	宮野(東大)	R	EP (SGE)
	状態空間モデルによる時系列データからの遺伝子ネットワーク推定プログラム	SSM	宮野(東大)	C, C++, JAVA	MPI, OpenMP
	データ解析融合プラットフォームの開発	SBIP	宮野(東大)		プラットフォーム
	網羅的タンパク質ドッキング解析プログラム	MEGADOCK	秋山(東工大)	C++	MPI
	生命体データ同化プログラム	LiSDAS	樋口(統数研)	Fortran90, C, C++	MPI, OpenMP
	次世代シーケンス解析プログラム	NGS analyzer	鎌谷(理研)	perl, shell	EP
脳神経	拡張RAT法による2SNP組合せの全ゲノム関連解析ソフトウェア	ExRAT	鎌谷(理研)	C++	MPI
	Neural Simulation Tool	NEST	Diesman(理研)		
	Cortical Microcircuit Developed on NEST - Interneuron	CMDN	深井(理研)	SLI, C++	MPI
	神経細胞形態シミュレーションキット	NeuroMorphoKit	石井(京大)	MATLAB, C/C++	逐次
	視覚系シミュレーションのための共有プラットフォーム	VSM	臼井(理研)		プラットフォーム
基盤	昆虫嗅覚系全脳シミュレータ	IOSSIM	神崎(東大)		実験系アプリ
	大規模並列用MDコアプログラム	cppmd	泰地(理研)	C++	MPI
	大規模仮想化合物ライブラリー	VLSVL	船津(東大)		EP
	分散並列大規模データ可視化システム	LSV	小野(理研)	C++, C	MPI
	アプリケーションモデルウェア	SPHERE	小野(理研)	C++,C	MPI

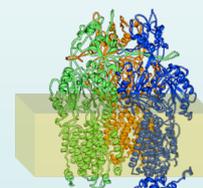
多剤排出トランスポーターにおける薬剤排出シミュレーション (MARBLE/Platypus/CafeMol)

研究代表者: 木寺 詔紀(理化学研究所)

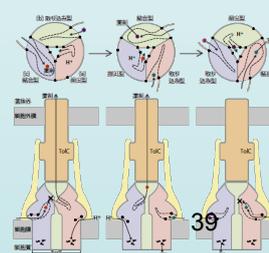
**抗生物質が効かなくなる
メカニズムの理論的解明を目指す**

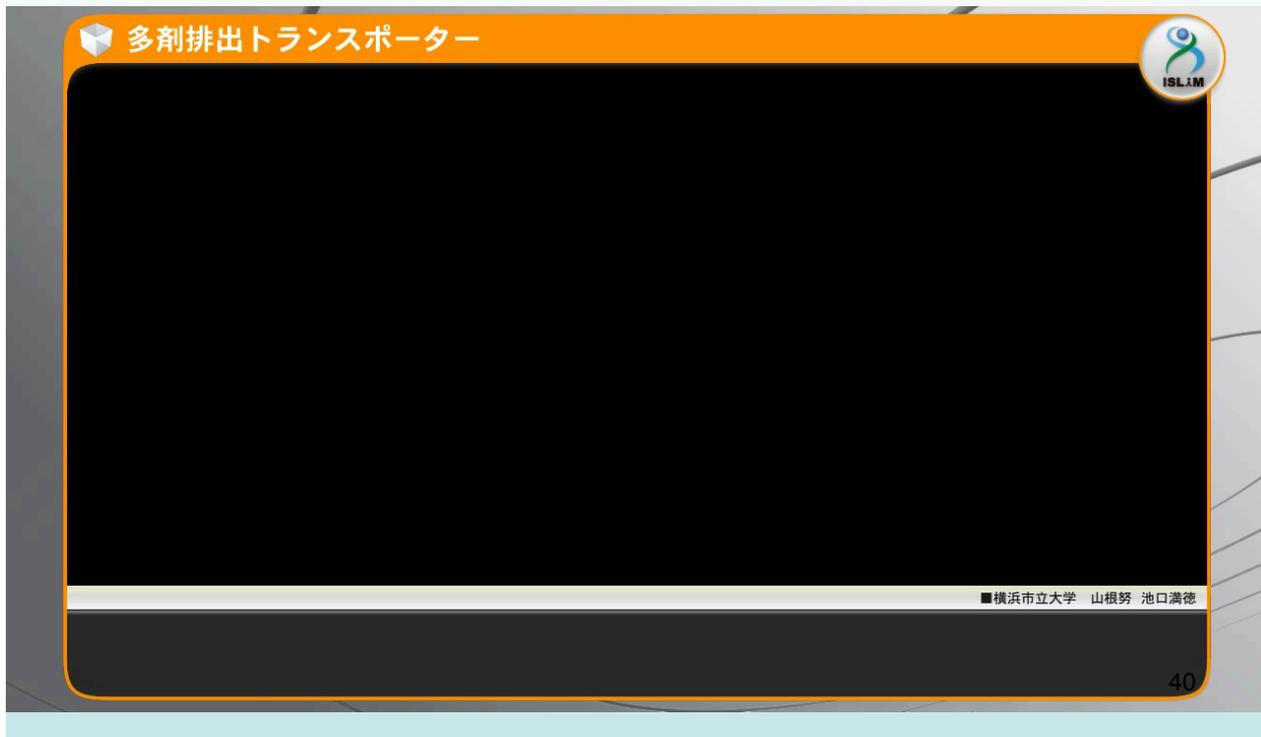
ソフトウェアの特徴

- 電子スケール～原子・分子スケール～タンパク質全体スケールの多階層(量子化学・分子動力学・粗視化モデル)の連成計算によって巨大タンパク質の動作機序を明らかにする。
- 計算対象: 細胞膜にある多剤排出トランスポーター
3つのタンパク質が結合した巨大タンパク質。
各々、取り込み型 → 結合型 → 排出型
と順繰りに構造を変えながら、薬剤化合物を排出すると考えられている。そのメカニズムを探る
- 細菌やがん細胞の薬剤耐性化の主要な原因の1つであるメカニズムの解明につながり、薬剤耐性化しにくい抗生物質や、排出を妨げる新薬の設計等への波及が期待できる。



多剤排出トランスポーター-AcrB





低侵襲治療シミュレーション:HIFU (ボクセル超音波伝播プログラム)

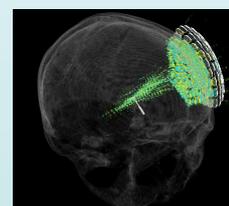
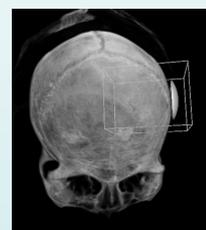
開発責任者: 松本 洋一郎 (東京大学)

超音波治療機器の設計開発・安全性の確認に貢献

— 患部だけを焼く集束超音波治療の安全な適用に向けた計算科学技術 —

ソフトウェアの特徴

- 人体内(骨や臓器など音響特性の異なる多媒質体内)の超音波伝搬をシミュレーションするソフトウェア
- 医療画像データから開発されたボクセル人体モデルを使ってシミュレーション。HIFU(強力集束超音波)を使った治療機器の超音波の集束状況の高精度な制御をシミュレーション可能
- 現在の超音波治療では対応困難な肝臓などの体内深部にある臓器や脳など骨で囲まれた部位に対して、集束超音波で腫瘍を焼灼する医療機器の開発に利用。シミュレーション結果をもとに超音波の照射方法を検討し、手術計画を立案可能



頭蓋骨越しのHIFU照射シミュレーション

低侵襲治療支援のための人体内超音波伝播シミュレータ



■理化学研究所 沖田浩平 小野謙二 高木周 ■東京大学 松本洋一郎

42

細胞シミュレータ: RICS

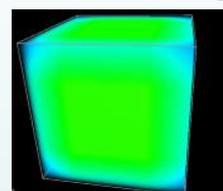
反応—拡散—膜透過の連成計算を100万Voxel空間で実現!
(100³)

小規模モデルの計算時間(赤血球)

2×2×2(=8)ボクセルを逐次(1CPU)で計算
1000steps/197秒 Δt=1ms (実時間1秒)

赤血球形状モデル

64×64×25(=102400)ボクセルを逐次(1CPU)で計算
512並列:1.36h Δt=0.05ms
実時間1分:600h 10万並列:3hr



2×2×2ボクセル
oxyHbを表示



64×64×25ボクセル

肝細胞詳細モデル

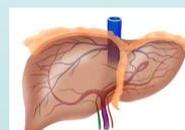
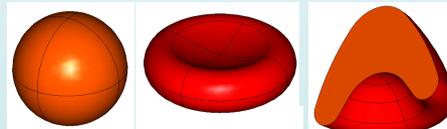
100×100×100(百万)ボクセルを10万並列)で計算

フル代謝モデルを想定 Δt=0.05ms
実時間1分 : 10万並列:30hr

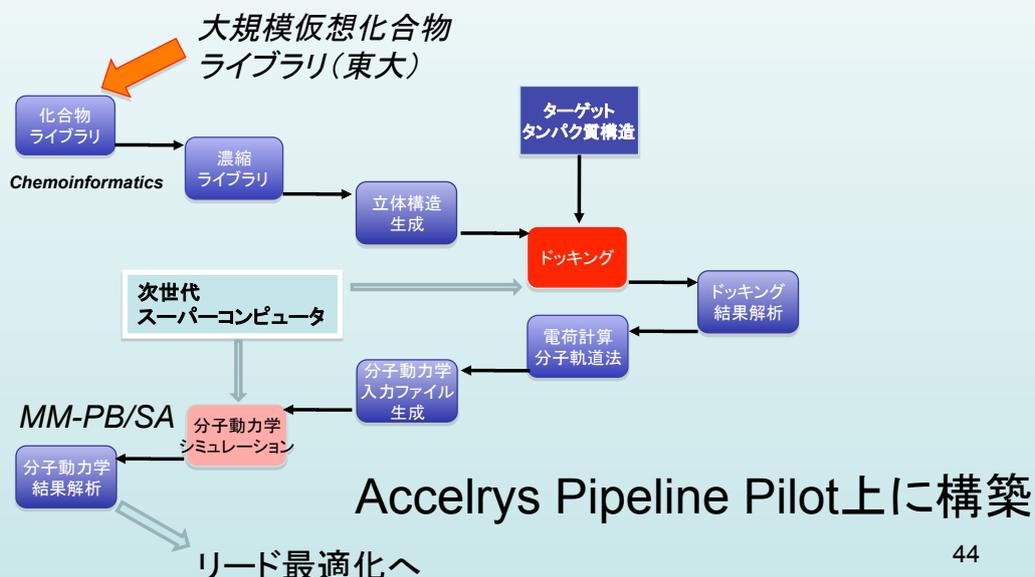
肝小葉モデル(代謝骨格モデル)

50×50×50(百万)ボクセルを10万並列)で計算

フル代謝モデルを想定 Δt=0.5ms
実時間10分 : 10万並列:40hr



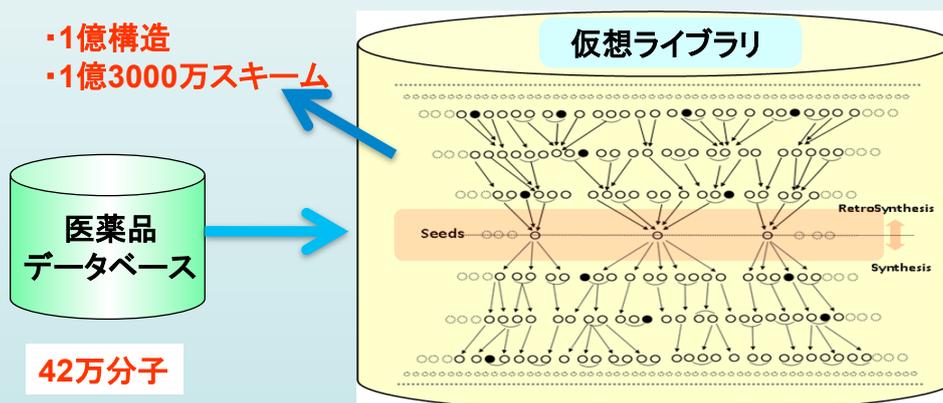
- 大規模な分子シミュレーションを用いた薬剤スクリーニング向けパイプライン



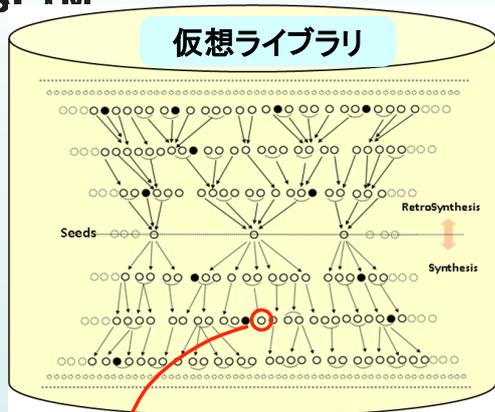
44

大規模仮想ライブラリ構築現状

- 現状での構築規模
 - 42万化合物構造を種として、逆合成スキームを2段処理
 - 1億個の新規仮想構造、1億3000万スキームを発生
 - 使用作業メモリ量 4GB、作業ディスク容量 4TB、2000時間(Xeon3.6GHz)



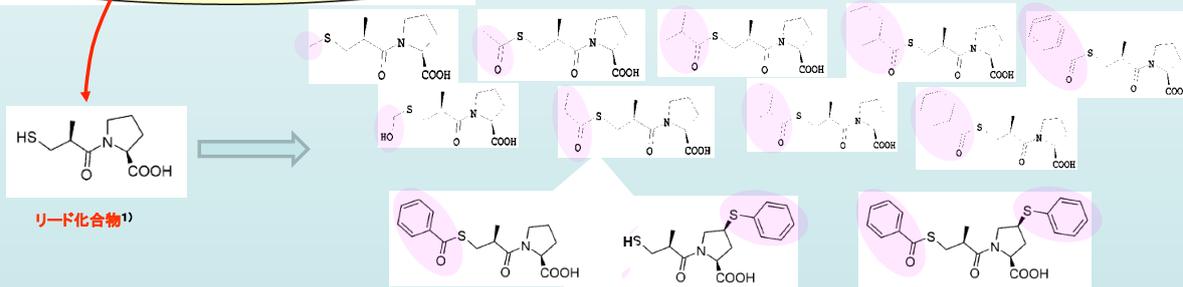
45



構造記述子を満足する周辺構造を
構造創出プログラムを利用して緻密に創出
プログラムは船津研独自のもの

効果:

リード化合物との構造類似性(薬理活性)を保ちつつ、周辺構造を可能な限り網羅的に発生させて、候補化合物群を得る



1) J.R.Proudfoot, *Bioorg. Med. Chem. Lett.*, 12, 1647(2002)



13機関: 東京大学、大阪大学、京都大学、東北大学、北陸先端大学、東海大学、千葉大学、慶応大学、横浜市大、東工大、広島大学、統計数理研究所、理化学研究所
約200名の研究者(ポスドク約80名)

- 開発ソフトウェア: 全34本
 - そのうち20本: 1024並列以上
 - **15本: 8000並列以上**
 - 可視化ソフトも100GPGPU並列
- 今後、並列性能の改善
- 次世代スパコンにフィットした最適化
- 計算時間に合わせた適正な計算規模を設定し、利用計画を立案、フル稼働に備える



Contributors



Workshop in 2009

1st Joint Workshop on Co
7-8 Jul 2008

Workshop in 2008



添付資料

50

50

ソフトウェアの高性能化

- MDコア
大規模並列化された分子動力学計算コア
- MOコア
Platypus QM/MMのためのMOの並列化
- 第一走者アプリケーションの高性能化
 - 他チームのアプリケーションを含めたチューニング
 - 性能評価を完了、結果をフィードバック
 - 今後共同で高性能化を進める
- 高度化チームとしての本番の活動は次世代スーパーコンピュータが稼働を始める今年度以降
 - 低レベル通信API
 - SIMD命令の直接的利用
- 準備状況としては、ほぼ予定通り

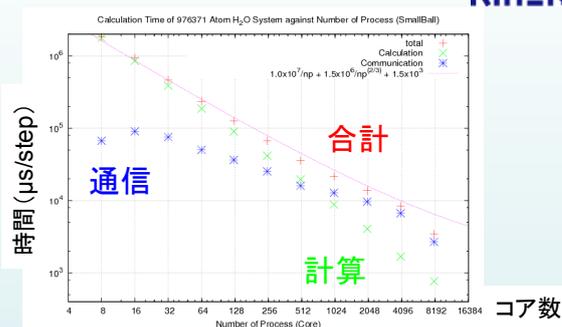
51

- 並列化の状況に関しては、アプリケーションによってまちまち。
- MDの場合：
Desmond (D. E. Shaw Research), NAMD(UIUC), Blue Matter (IBM) などが数千並列以上の並列を実現。
何れのコードも、高性能計算の専門家が中心に参加して開発を進めている。
- ライフアプリケーションでの 10^4 以上の並列度については、世界的に見てもこれまでの蓄積は非常に少ない。

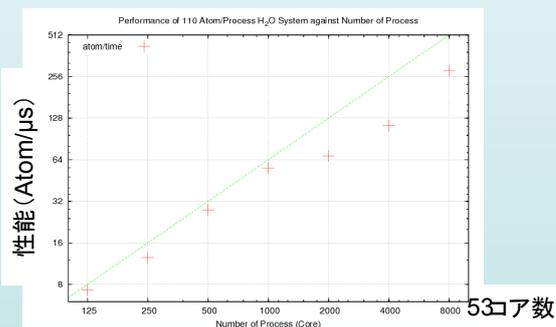
52

MDコア:cppmd(1) 開発のねらいと状況

- 大規模並列に向けて、様々な並列化手法を盛り込んだ基盤ソフトウェア
- 大規模並列化の「トレーニング」
- 開発時点から次世代スーパーコンピュータを強く意識
- 多機能版
 - 再利用性・手法比較
- 高速版
 - 二体相互作用カットオフ計算の高速化を優先
 - 高並列時の通信を考慮
 - 水 976371原子 (110原子/コア)
 - RICC 8000並列での並列化確認
 - その他の機能は多機能版から移植中

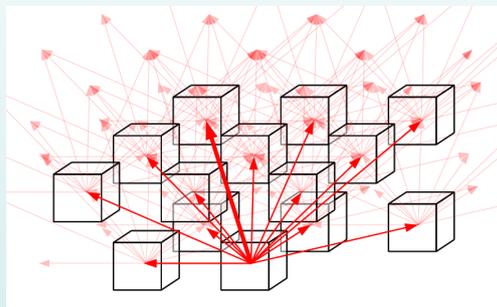


Strong scaling
976,371 atoms, smallball, XYZ通信



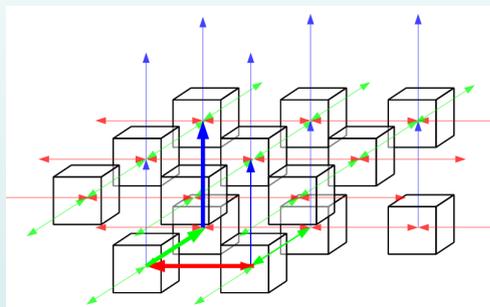
Weak scaling 110atoms/core

通常の実装



- 全方向直接
- 通信相手が多い
- 競合が激しくなる
- 次世代スーパーコンピュータの特徴を生かせない

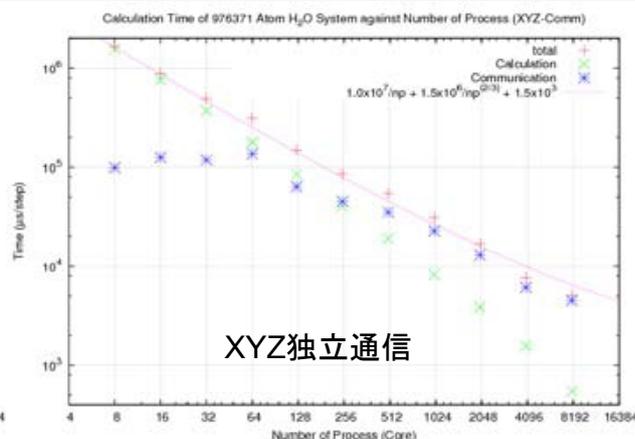
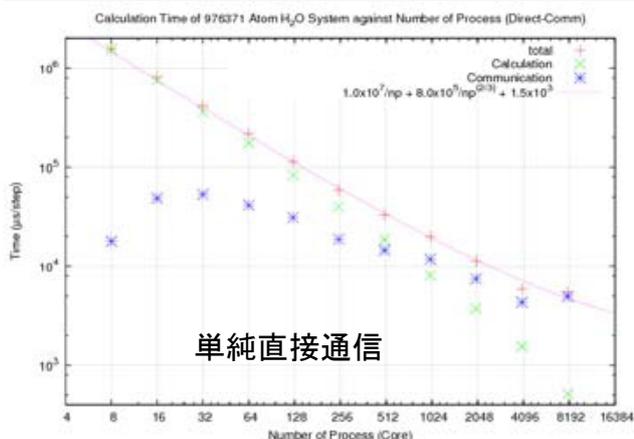
今回の実装



- x,y,z軸に逐次通信
- 通信相手が少ない
- 次世代スーパーコンピュータのネットワークポロジ(トーラス構造)に一致

8000コアでの性能向上を実現 ⁵⁴

通信の改良の効果

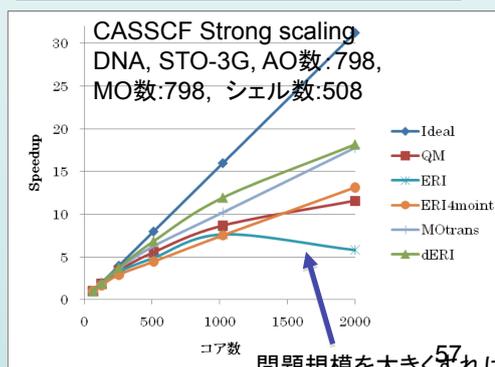
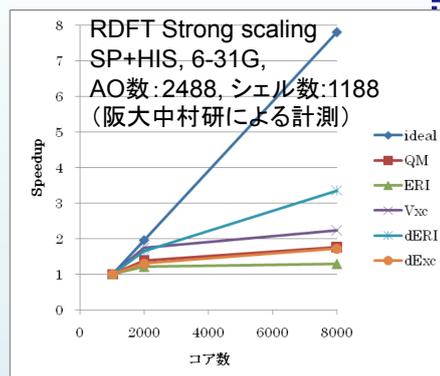


性能スケーリング(Strong scaling)。+ : 全計算時間、× : 計算時間、* : 通信時間
(RICCでの測定, fat-tree)

- 二体相互作用のカットオフ計算
B/Fの違いも考慮すると
90原子/ノード程度までスケール(11原子/コア)
- 次世代スパコン全体(80,000ノード)で
7,200,000原子
- 2×10^9 step / day (効率10%= 1Pflops)
- 2fs/stepで**4 μ sec/day**
- 今後の課題
他の計算の負荷の見つもり
単体性能チューニング
 - ✓ Integer cache missの改善(FX1での実行結果から)
 - ✓ OpenMP・SIMDの活用
 - ✓ 命令実行密度(CPI)の向上

MOコア 開発のねらいと状況

- Platypus QM/MMのための並列分子軌道法ソフトウェア
- HF/RHF/CASSCF/CASCI
- ベクトル化(当初)、SIMD化についても調査
- 基本部分の開発を完了し、阪大中村研に引き渡し
- DFT/CASDFTについては阪大中村研で開発
- CASSCFのMO変換の性能向上などに特徴
- 今後、さらに並列化率の向上に協力。
典型的な問題に対し、20,000コア以上までのスケールリング実現が目標
- その他、O(N)分子軌道法(NTChem, 有限要素基底)の高速化についても一部支援



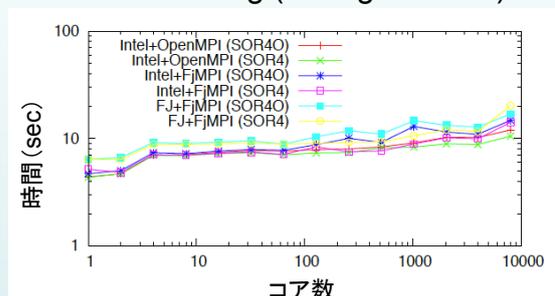
問題規模を大きくすれば
さらに改善する見込み

臓器全身スケールチームで開発中のコード

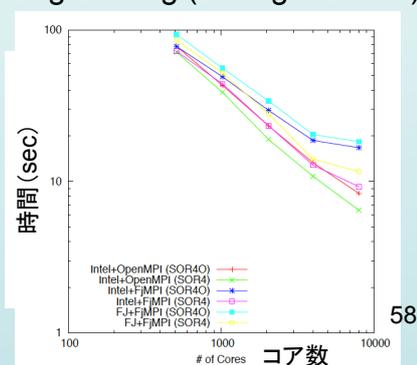
- 1メッシュ当りの演算数
 - $A=7144+827+18N_i$
 - N_i はSORのiteration回数
- Weak/Strong scaling共に良好

solphi	827flop	2.8GFLOPS@RICC (24% of peak)
solu	7144flop	1.5GFLOPS@RICC (13% of peak)
Solp_sor4	18flop/ iteration	0.71GFLOPS@RICC (6% of peak)

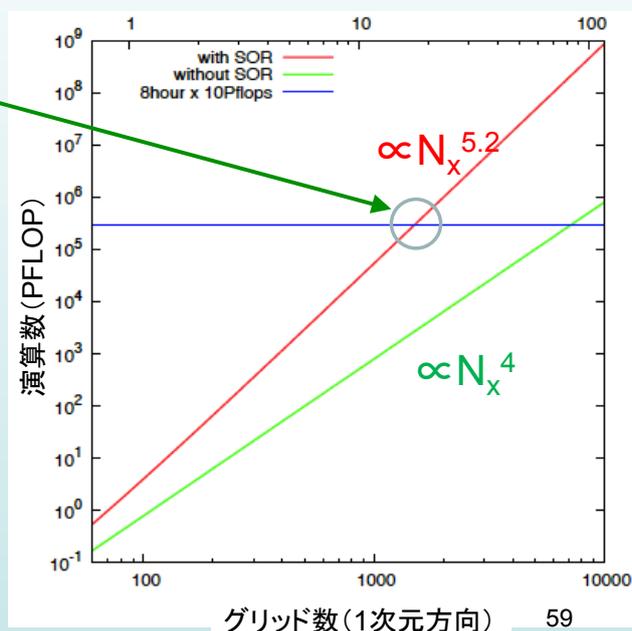
Weak Scaling (100³ grids/core)



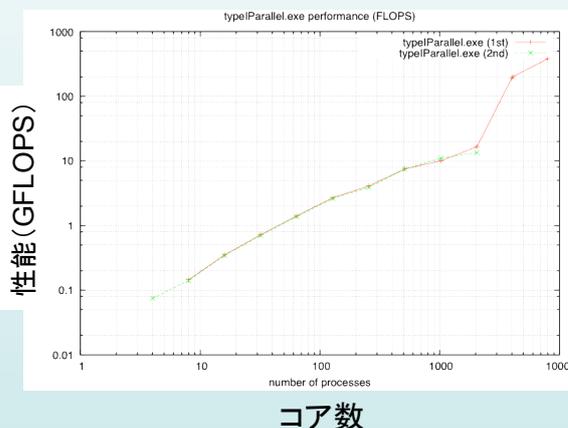
Strong Scaling (960³ grids/core)



- 次世代スパコン(10Pflopsマシン) 上でのペタスケール演算モデル
 - 現状コードでは1500³ボクセルが限界
 - このときSORの反復数は48000回
 - 並列FFTベースソルバの1000倍遅い
 - 1Core当りのボクセル数が17³: 通信ボトルネック
 - Poisson solverの改良が必須
- 課題
 - Multigrid Poisson Solverの導入 (MDコア等で開発)
 - Hybrid 並列化



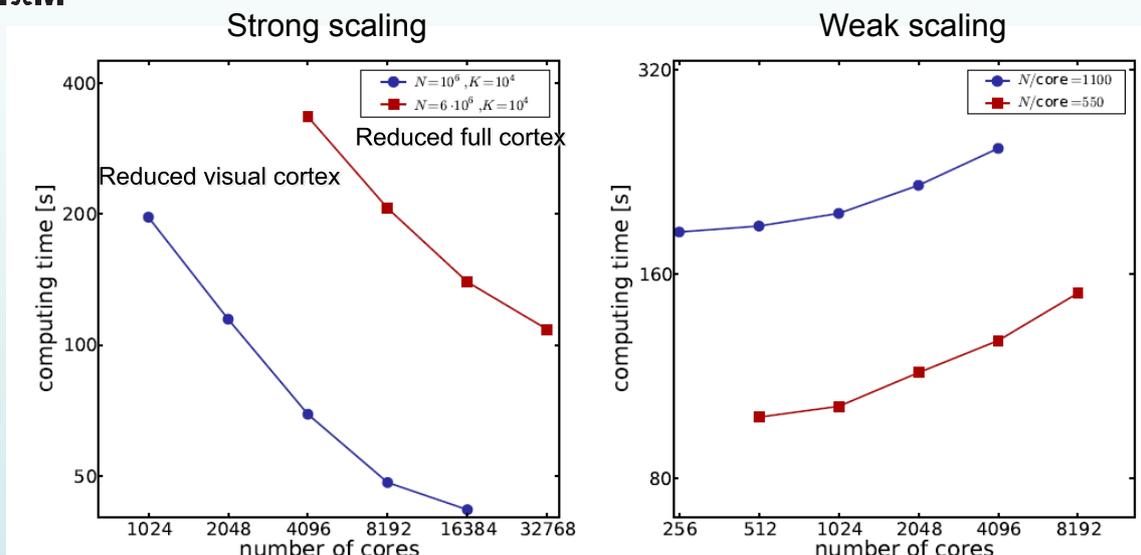
- 基本的には、Embarrassingly Parallel
- RICC 8000コアでスケーリングを確認
- ターゲット規模: 1万人集団 × 100万 SNP
次世代スパコンで15時間
- ハードウェアカウンタでは383GFLOPS
コアあたり0.05GFLOPS: 整数演算中心
→ 整数演算の浮動小数点演算への置き換えを検討
- 連鎖不平衡ブロックあたりのハプロタイプデータ数のアンバランスが、ロードバランスに直結
→ 配分の見直しによるロードバランスの改善を検討



60

<http://www.nest-initiative.org>

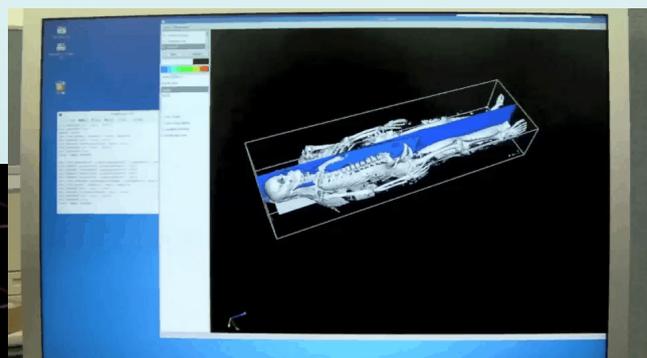
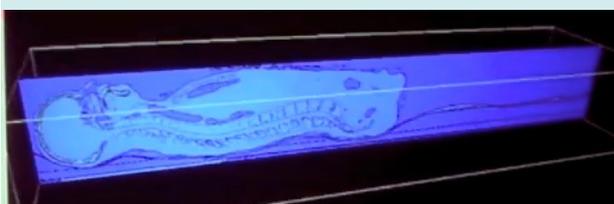
- 大規模神経回路モデルの分散シミュレーション
 - 現時点で世界最高水準の集積度を可能としている
- $O(100)$ プロセッサまで超線形、 $O(1000)$ プロセッサまで線形の性能向上
- CPU間のメッセージパッシングとマルチコアCPU内のマルチスレッディングのハイブリッドアーキテクチャ
- 連続値でのスパイクタイミングを用いた精度良い積分
- ユーザフレンドリーなPythonインターフェース**PyNEST**
- 多種のニューロンモデルとスパイクタイミング依存可塑性 (STDP) を含む多種のシナプスモデルを実装可能
- 短期可塑性 (STP)も扱うことができる



- 現在のシミュレーションは $10^6 - 10^7$ ニューロンからなるモデルで実施中

Kunkel S, Potjans T, Morrison A, Plesser HE, Diesmann M

- 大規模並列化され、大規模データに対応した可視化ソフトウェア
- リモート可視化にも対応
- 人体ボクセルデータでのテスト
(最高解像度: 0.2mm pitch, 27.5 GVoxel, 107GB)
- RICC GPGPUクラスタ(upc)を使用
 - Nehalem+
 - TESLA C1060
 - 100台



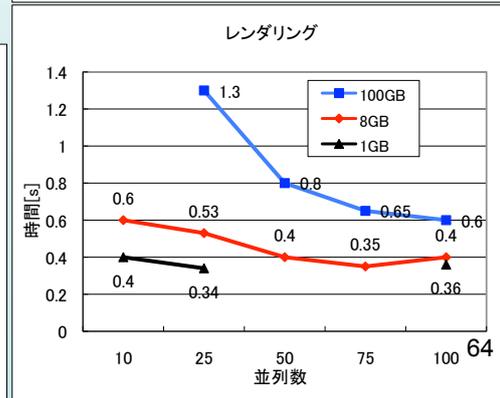
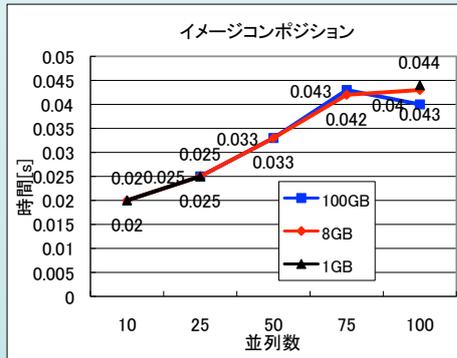
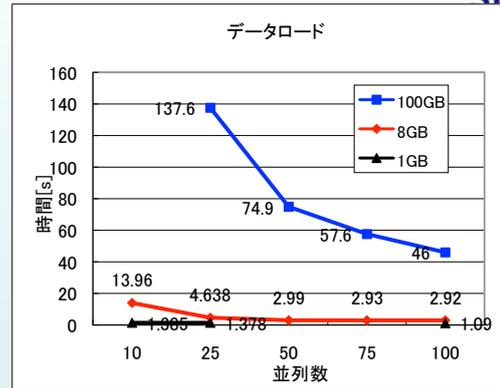


大規模可視化ソフトウェア(2) 性能測定(人体ボクセル)



ISLAM

- データ分割によりデータロードは1~3秒程度まで高速化を実現
- レンダリングは0.4秒程度
- イメージ重畳は一桁程短い時間
- 5FPS(0.2[sec/frame])の実現にはデータロードとレンダリングのチューニングが必要
 - データロード: ステージング+高速化
 - レンダリング: ノード内並列化

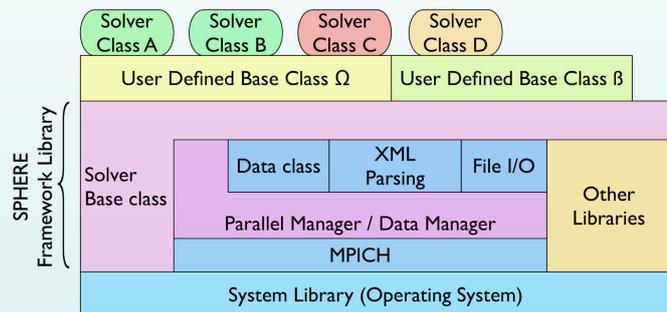


アプリケーションミドルウェアSphere 並列ソフトウェア開発加速のためのプラットフォーム



ISLAM

- 開発者向け機能
 - ラピッドプロotyping
 - ライブラリ群の利用
 - ひな形制御構造の利用
 - コード部品の再利用
 - 並列処理機能の利用
 - ソースポータビリティ
 - 混合言語 C++/F90, GPGPU
 - 連成問題のスクリプト記述



- エンドユーザ向け機能
 - 統一されたUIと実行環境
 - XMLによる制御
 - マルチプラットフォーム
 - 次世代スーパーコンピュータ, Linux cluster, Windows cluster, BG, SX, Mac, Altix, T2K...

XML file :
<SphereConfig SolverType="Incompressible_flow_solver" >

- 現在、臓器全身スケール(SPH3D, HIFU)・細胞スケールチーム(RICS)で利用

