

次世代生命体統合シミュレーションソフトウェアの研究開発
Next-Generation Integrated Simulation of Living Matter

ISLiM ソフトウェア研究開発報告会

2013 年 1 月 10 日(木) - 11 日(金)

東京大学 武田ホール

文部科学省委託事業

「次世代生命体統合シミュレーションソフトウェアの研究開発」

次世代生命体統合シミュレーションソフトウェアの研究開発
Next-Generation Integrated Simulation of Living Matter (ISLiM)

ISLiMソフトウェア研究開発報告会

開催日時: 2013年1月10日(木)-11日(金) (10:00開始) (受付:9:30開始)

会場: 東京大学 武田ホール (本郷キャンパス、武田先端知ビル 5F)

主催:

文部科学省委託事業「次世代生命体統合シミュレーションソフトウェアの研究開発」

協賛:

次世代スパコンの創薬産業利用促進研究会

バイオスーパーコンピューティング研究会

プログラム

時間	(分)	DAY1 (1/10)	
10:00-10:15	(15)	●オープニング	
		茅 幸二 プログラムディレクター	
10:15-10:45	(30)	プロジェクト総合報告	
		姫野龍太郎 副プログラムディレクター	
10:45-11:55	(70)	分子スケール研究開発チーム成果報告 (1/2)	
		・ MM/CG プログラム μ 2-lib によるタンパク質の大規模サンプリング	木寺 詔紀 (理研/横浜市大)
		・ Replica-exchange interface program: REIN	宮下 尚之 (理研)
		・ 全原子分子動力学計算ソフトウェア MARBLE の開発と多剤排出トランスポーター AcrB への応用	池口 満徳 (横浜市大)
		・ タンパク質カノニカル分子軌道法 ProteinDF の京対応版プログラムの公開	佐藤 文俊 (東大)
11:55-12:50	(55)	□ランチタイム	
12:50-13:50	(60)	◆ポスター展示説明および休憩	
		講演会場前 (5F ホワイエ)	
13:50-14:45	(55)	分子スケール研究開発チーム成果報告 (2/2)	
		・ 粗視化分子シミュレーションソフトウェア CafeMol の開発と遺伝子動態研究	高田 彰二 (京大)
		・ 酵素分子の柔らかさと触媒活性：ハイブリッド QM/MM 自由エネルギー法の開発	林 重彦 (京大)
		・ Platypus-QM/MM の開発と応用	鷹野 優 (阪大)
14:45-15:35	(50)	細胞スケール研究開発チーム成果報告	
		・ 細胞スケール研究開発チーム成果統括報告	横田 秀夫 (理研)
		・ 血小板細胞シミュレーターの開発	後藤 信哉 (東海大)
15:35-16:20	(45)	◆ポスター展示説明および休憩	
		講演会場前 (5F ホワイエ)	
16:20-18:05	(105)	データ解析融合研究開発チーム成果報告	
		・ データ解析融合研究開発チーム チーム成果統括報告	宮野 悟 (東大)
		・ 大規模遺伝子ネットワーク推定とその応用	宮野 悟 (東大)
		・ 大規模ゲノム多型データと表現型データを関連付ける新規アルゴリズムの開発と、妥当性、有用性の検討	角田 達彦 (理研)
		・ 大規模タンパク質ネットワーク推定とその応用	秋山 泰 (東工大)
		・ 生命体シミュレーションのためのデータ同化技術の開発	樋口 知之 (統数研)
18:10-19:30		○懇親会	
		講演会場前 (5F ホワイエ)	

プログラム

時間	(分)	DAY2 (1/11)
10:00-10:15	(15)	● DAY2 オープニング
10:15-12:00	(105)	臓器全身スケール研究開発チーム成果報告 <ul style="list-style-type: none">臓器全身スケール研究開発チーム成果総括報告 高木 周 (理研 / 東大)次世代超音波治療機器開発のための生体内超音波伝播シミュレータ (ZZ-HIFU) 沖田 浩平 (日大)大規模並列計算に適した流体構造 / 膜連成手法 (ZZ-EFSI) 杉山 和靖 (理研)UT-heart によるサルコメア力学から心筋細胞構造を経て心拍動にいたるマルチスケール解析 鷺尾 巧 (東大)
12:00-13:00	(60)	□ ランチタイム
13:00-14:00	(60)	◆ ポスター展示説明および休憩 講演会場前 (5F ホワイエ)
14:00-15:45	(105)	脳神経系研究開発チーム成果報告 <ul style="list-style-type: none">神経系の大規模シミュレーション 石井 信 (京大)NEST による局所回路シミュレーション 五十嵐 潤 (理研)Multi-scale simulation of subcortical visual systems with real-time robotic visuo-motor feedback Jan Moren (沖縄科技大)大規模数理モデル構築プラットフォームによる視覚系数理モデルの構築 臼井 支朗 (理研 / 豊橋技科大)「京」による無脊椎動物嗅覚系シミュレーション - 背景と現状と将来への展開 - 神崎 亮平 (東大)
15:45-16:30	(45)	◆ ポスター展示説明および休憩 講演会場前 (5F ホワイエ)
16:30-17:45	(75)	生命基盤ソフトウェア開発・高度化チーム成果報告 <ul style="list-style-type: none">ISLiM ソフトウェアの高度化 泰地 真弘人 (理研)MD コアプログラムの最適化と性能評価 大野 洋介 (理研)合成可能な化学構造および反応スキームからなる大規模仮想ライブラリの構築 船津 公人 (東大)
17:45-18:00	(15)	● クロージング

プログラムは都合により変更されることがあります。

目次

ご挨拶

茅 幸二 プログラムディレクター	1
------------------	---

プロジェクト総合報告 ISLiM：ソフトウェア開発の現状と今後

姫野龍太郎 副プログラムディレクター	3
--------------------	---

分子スケール研究開発チーム成果報告

MM/CG プログラム μ 2-lib によるタンパク質の大規模サンプリング	木寺詔紀 (理研/横浜市大)	15
Replica-exchange interface program: REIN	宮下尚之 (理研)	21
全原子分子動力学計算ソフトウェア MARBLE の開発と多剤排出トランスポーター AcrB への応用	池口満徳 (横浜市大)	29
タンパク質カノニカル分子軌道法 ProteinDF の京対応版プログラムの公開	佐藤文俊 (東大)	33
粗視化分子シミュレーションソフトウェア CafeMol の開発と遺伝子動態研究	高田彰二 (京大)	37
酵素分子の柔らかさと触媒活性：ハイブリッド QM/MM 自由エネルギー法の開発	林 重彦 (京大)	41
Platypus-QM/MM の開発と応用	鷹野 優 (阪大)	47

細胞スケール研究開発チーム成果報告

細胞スケール研究開発チーム成果統括報告	横田 秀夫 (理研)	53
血小板細胞シミュレーターの開発	後藤 信哉 (東海大)	61

データ解析融合研究開発チーム成果報告

データ解析融合研究開発チーム チーム成果統括報告	宮野 悟 (東大)	65
大規模遺伝子ネットワーク推定とその応用	宮野 悟 (東大)	71
大規模ゲノム多型データと表現型データを関連付ける新規アルゴリズムの開発と、妥当性、有用性の検討	角田達彦 (理研)	77
大規模タンパク質ネットワーク推定とその応用	秋山 泰 (東工大)	83
生命体シミュレーションのためのデータ同化技術の開発	樋口知之 (統数研)	89

臓器全身スケール研究開発チーム成果報告

臓器全身スケール研究開発チーム成果統括報告	高木 周 (理研/東大)	97
次世代超音波治療機器開発のための生体内超音波伝播シミュレータ (ZZ-HIFU)	沖田浩平 (日大)	103
大規模並列計算に適した流体構造/膜連成手法 (ZZ-EFSI)	杉山和靖 (理研)	109
UT-heart によるサルコメア力学から心筋細胞構造を経て心拍動にいたるマルチスケール解析	鷲尾 巧 (東大)	115

脳神経系研究開発チーム成果報告

神経系の大規模シミュレーション	石井 信 (京大)	121
NEST による局所回路シミュレーション	五十嵐潤 (理研)	133
Multi-scale simulation of subcortical visual systems with real-time robotic visuo-motor feedback	Jan Moren (沖縄科技大)	139
大規模数理モデル構築プラットフォームによる視覚系数理モデルの構築	臼井支朗 (理研/豊橋技科大)	145
「京」による無脊椎動物嗅覚系シミュレーションー背景と現状と将来への展開ー	神崎亮平 (東大)	151

生命基盤ソフトウェア開発・高度化チーム成果報告

ISLiM ソフトウェアの高度化	泰地真弘人 (理研)	159
MD コアプログラムの最適化と性能評価	大野洋介 (理研)	165
創薬プラットフォームー大規模バーチャルライブラリの開発ー	船津公人 (東大)	169

ポスター一覧

M-1: Platypus-MM/CG 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	177
M-2: Platypus-REIN レプリカ交換インターフェースプログラム	178
M-3: MARBLE 全原子分子動力学計算	179
M-4: CafeMol 粗視化モデル計算	180
M-5: ProteinDF 密度汎関数法に基づくタンパク質全電子波動関数計算	181
M-6: Platypus-QM/MM-FE ハイブリッド QM/MM 反応自由エネルギー計算	182
M-7: Platypus-QM 量子化学計算	183
M-8: Platypus-QM/MM 量子化学計算/分子動力学連成計算	184
C-1: RICS 細胞シミュレーションプラットフォーム	185
C-2: RICS-Pre 細胞シミュレーションプラットフォーム (RICS) プリシステム	186
C-3: RICS-platelets RICS による血小板細胞シミュレーションと力学的作用による偽足の変形についての検討	187
C-4: 血小板活性化モデルのための細胞内カルシウムイオン実証実験	188
C-5: 分子動力学法による平衡状態におけるコフィリン修飾アクチンフィラメントのエネルギー解析	189
O-1: ZZ-EFSI 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)	190
O-2: 血栓形成のマルチスケールシミュレーション - 血小板の血管壁への接着に関するタンパク質間相互作用の解析 -	191
O-3: ZZ-DOSE 全身ボクセルモンテカルロ重粒子線計算プログラム	192
O-4: ZZ-HIFU 集束超音波による低侵襲治療のための HIFU シミュレータ	193
O-6: UT Heart マルチスケール・マルチフィジックス心臓シミュレータ	194
O-7: 血管内の血流シミュレーションソフトウェア開発・検証研究	195
O-8: 筋骨格系 3次元有限要素シミュレータの開発	196
O-9: 血管網シミュレーター 全身循環モデルを用いた頸動脈ステント留置術に関する血流シミュレーション	197
D-1: ParaHaplo 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	198
D-2: NGS analyzer 次世代シーケンサーの出力データを高速に解析し、個人間の遺伝的差異やがんゲノムの突然変異を高い正確さで同定する	199
D-3: ExRAT 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	200
D-4: SiGN-BN ベイジアンネットワークによる大規模遺伝子ネットワーク推定プログラム	201
D-5: SiGN-L1 L1 正則化法による大規模遺伝子ネットワーク推定プログラム	202
D-6: SiGN-SSM 状態空間モデルによる大規模遺伝子ネットワーク推定プログラム	203
D-7: SBiP データ解析融合プラットフォーム	204
D-8: LiSDAS 生命体データ同化プログラム	205
D-9: MEGADOCK 網羅的タンパク質ドッキング解析プログラム MEGADOCK の開発と応用	206
B-1/B-2: NEST による局所回路シミュレーション	207
B-3: VSM 全視覚系モデルによる視覚情報処理の解析 (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)	208
B-4: NeuroMorphoKit 神経細胞形態シミュレーションキット	209
B-5: IOSSIM カイコガ運動指令形成回路のシミュレーションの構築の試行	210
B-5: IOSSIM 昆虫嗅覚系全脳シミュレータ	211
H-1: cppmd 大規模並列用 MD コアプログラム	212
H-2: LSV 大規模データ可視化ソフトウェア LSV の開発	213
H-3: SPHERE Simulation Framework for Large scale Coupled Analysis	214
H-4: VLSVL 合成可能な化学構造および反応スキームからなる大規模仮想ライブラリ	215

ご挨拶

茅 幸二

次世代計算科学研究開発プログラム
プログラムディレクター



発表者紹介

昭和 41 年 3 月 東京大学大学院理学研究科博士課程修了
昭和 45 年 9 月 東北大学理学部化学科助教授
（昭和 48 年 9 月～昭和 49 年 9 月 米国ベル電話研究所研究員）
昭和 56 年 4 月 慶應義塾大学工学部教授
平成 11 年 4 月 岡崎国立共同研究機構分子科学研究所長
平成 11 年 4 月 慶應義塾大学名誉教授
平成 16 年 4 月 独立行政法人理化学研究所和光研究所・中央研究所長
平成 18 年 10 月 次世代計算科学研究開発プログラム プログラムディレクター(兼)
平成 20 年 4 月 次世代スーパーコンピュータ開発実施本部 副本部長
平成 20 年 5 月 基幹研究所 特別顧問(兼)
平成 22 年 4 月 社会知創成事業 次世代計算科学研究開発プログラム
プログラムディレクター(兼)

研究分野

ナノ物質化学

ご挨拶

茅 幸二

次世代計算科学研究開発プログラム プログラムディレクター

京コンピュータが世界最速の記録を達成し、一年あまりとなり、すでに米国では計算速度で京を超す計算機が完成している。この現況で、もっとも重要なのは、大型計算機時代における、計算機利用をいかに深く突き詰めるか、そしてその成果を社会に還元できるかという点にある。

「次世代生命体総合シミュレーションソフトウェアの研究開発」は、発足以来 7 年目となり、2013 年 3 月に終了する。生命科学の広範な分野の我が国を代表する計算科学の研究者が集い、京の利用による、生命科学における、医療、あるいは生命機能の複雑なシステムの理解に向けた日夜を問わない熱い挑戦をされた結果が、30 本あまりのソフトウェアの大半がペタスケールの計算を可能とする結果となった。本プログラムに参加された多くの研究者そして、とくに学術論文への投稿の機会を逸しながら、きわめて高度なソフト開発に従事された若手研究者の熱意と努力に心からの敬意と感謝を表す。

本報告会では、それらのソフト開発の成果を纏め、これからの高度計算によって、生命科学における定量化された成果がいかに意義あるものであるかを報告し、計算科学でのライフサイエンス分野の今後の寄与に、生命科学のみならず、多くの分野の方々に御理解と支持をいただくことを目的としている。

ISLiM : ソフトウェア開発の現状と今後

姫野 龍太郎

次世代計算科学研究開発プログラム
副プログラム・ディレクタ



発表者紹介

1979年3月 京都大学大学院工学研究科電気工学第二専攻修士課程修了

1979年4月 日産自動車株式会社入社、中央研究所、車両研究所勤務

主に車の空力シミュレーションの研究開発に従事

1998年1月 理化学研究所入所 情報環境室長

2003年10月 情報基盤センター長

2006年4月-2008年8月 次世代スーパーコンピュータ開発実施本部開発グループ、グループ・ディレクタ

2006年10月から、次世代計算科学研究開発プログラム、副プログラム・ディレクタ

研究分野

高性能計算、計算生体力学、計算流体力学

ISLiM: ソフトウェア開発の現状と今後

姫野 龍太郎

次世代計算科学研究開発プログラム
副プログラム・ディレクタ

1. はじめに

次世代生命体統合シミュレーションソフトウェアの研究開発プロジェクト(ISLiM)は次世代スーパーコンピュータ開発実施プロジェクトに含まれるライフ分野のグランドチャレンジとして2006年10月から開始した。このプロジェクトは理化学研究所を中核拠点として全国13の機関から、およそ200名の研究者が参加した。(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)

ソフトウェア開発は、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、HPC の6チームに別れ、表1に示すように34本のソフトウェアを開発した。HPC チームは他の5チームで開発しているソフトウェアの高速化を援助、特に2011年4月からの京へのチューニングでは大きな役割を担った。

2. 開発の現状

我々の京の利用可能期間は京の供用開始までということになっていた。この供用開始は当初2012年11月末を予定していた。しかしながら、この供用開始が9月中旬に早められた結果、我々の利用可能期間が2か月短縮されてしまった。大部分のソフトウェアでは京での大規模計算が可能にはなったが、当初予定していた研究成果を出す

ための計算を行う時間がとれなくなってしまった。非常に残念である。

図1は2006年からの開発を通して設定したマイルストーンで管理した進捗状況である。開発を3期に分け、第一期を計算アルゴリズムのチェックと並列化のテスト、第二期を理研で所有する1028ノード8192コアのクラスタを利用した高並列のテスト、第三期は京を使ったテストに位置づけている。このようにして3ヶ月毎に進捗をチェックしてきた。その結果、34本中13本のソフトウェアで1万ノードを超えるところまで性能が向上する高性能さを実現できた(図2)。また、cppmd, UT-Heart, ZZ-EFSI の3本は京全体を使ってテストを行い、約3 PetaFLOPS、ZZ-EFSI では 4.5PetaFLOPS という性能なソフトウェアに仕上がった(図3)。

3. 今後

現在、各ソフトウェアの公開に向けてマニュアル(日本語・英語)やダウンロードサイト等の整備を進めている。一部ソフトウェアに関しては一般課題等として京の計算時間を確保し、引き続き京での計算を行い、研究を継続している。今年度末、このプロジェクトは終了するが、来年度以降も引き続き、開発したソフトウェアが利用できるよう、整備し続ける予定である。

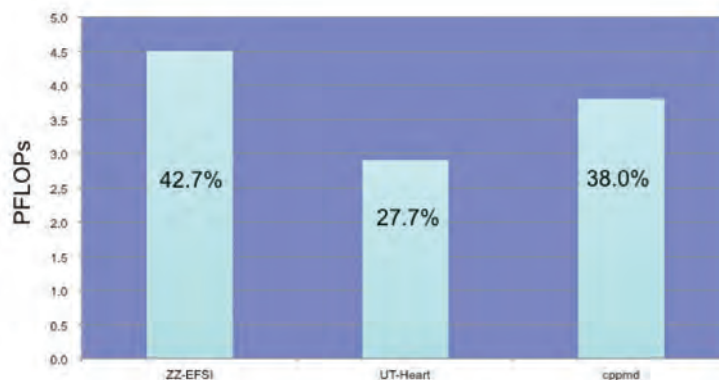


図3 京全体を使ったテストでの性能

表 1 開発ソフトウェア一覧 (全 34 本)

	ソフトウェア名称	説明
分子 (9)	Platypus-MM/CG	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-REIN	レプリカ交換分子動力学計算インターフェイス
	MARBLE	全原子分子動力学計算
	CafeMol	粗視化モデル計算
	ProteinDF	密度汎関数法に基づくタンパク質全電子波動関数計算
	Platypus-QM/MM-FE	ハイブリッドQM/MM反応自由エネルギー計算
	Platypus-QM	量子化学計算
	Platypus-QM/MM	量子化学計算/分子動力学計算
	Platypus-CGM/MM	粗視化モデル計算/分子動力学計算
細胞	RICS	細胞シミュレーションプラットフォーム
臓器全身 (6)	ZZ-EFSI	全身ボクセルシミュレーション(ボクセル構造流体力学解析プログラム)
	ZZ-RBC	微小循環シミュレータ (埋め込み境界法による微小循環プログラム)
	ZZ-DOSE	重粒子線治療シミュレーション
	HIFU	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)
	ZZ-LUNG	肺呼吸・肺循環シミュレーション
	UTHheart	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

	ソフトウェア名称	説明
データ解析融合 (9)	ParaHaplo	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
	NGS analyzer	次世代シーケンス解析プログラム
	ExRAT	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
	SiGN-BN (SiGN)	大規模遺伝子制御ネットワーク推定プログラム
	SiGN-L1 (L1GN)	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
	SiGN-SSM (SSM)	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
	SBiP	データ解析融合プラットフォーム
	LiSDAS	生命体データ同化プログラム
	MEGADOCK	網羅的タンパク質ドッキング解析プログラム
脳神経 (5)	NEST	Neural Simulation Tool
	CMDN	Cortical Microcircuit Developed on NEST
	VSM	全視覚系モデルによる視覚情報処理の解析(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)
	NeuroMorphoKit	神経細胞形態シミュレーションキット
基盤 (4)	IOSSIM	昆虫嗅覚系全脳シミュレータ
	cppmd	大規模並列用MDコアプログラム
	LSV	分散並列大規模データ可視化システム
	SPHERE	アプリケーションモデルウェア
	VLSVL	大規模仮想化合物ライブラリ

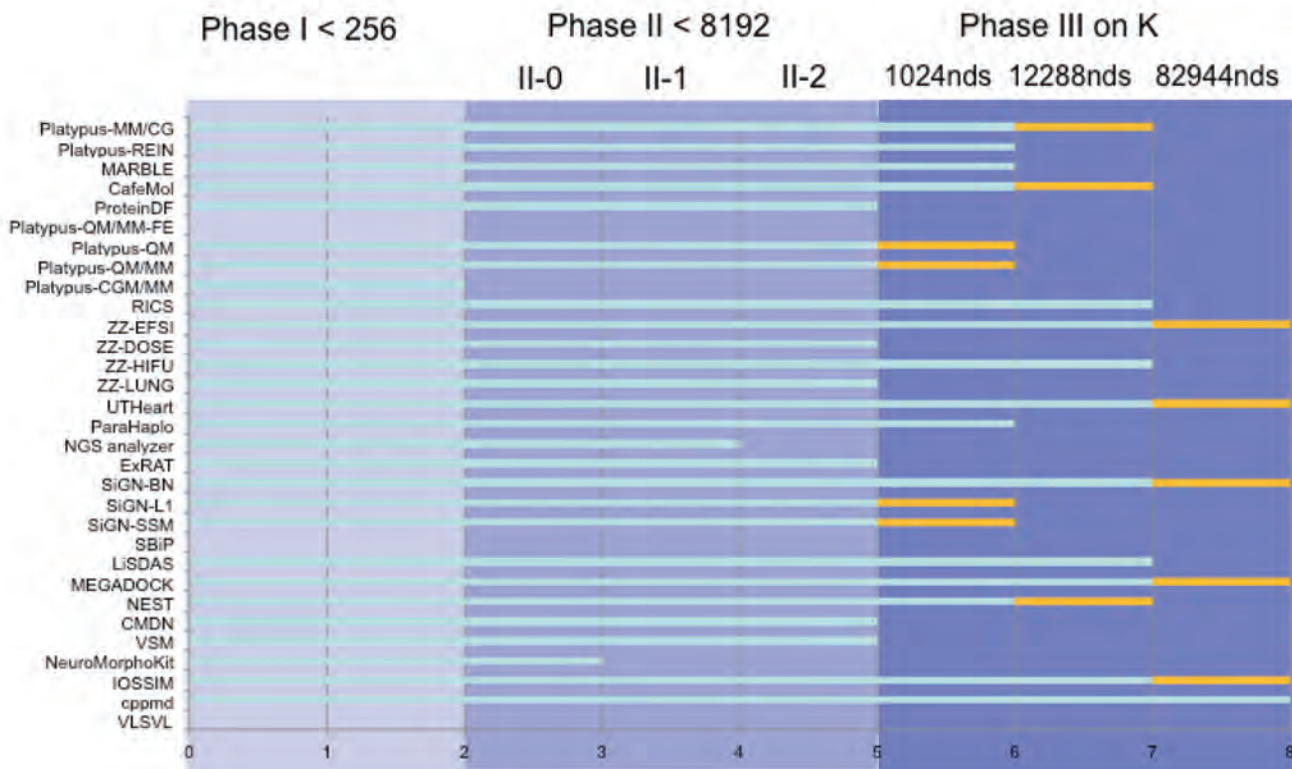


図 1 ソフトウェアの開発状況

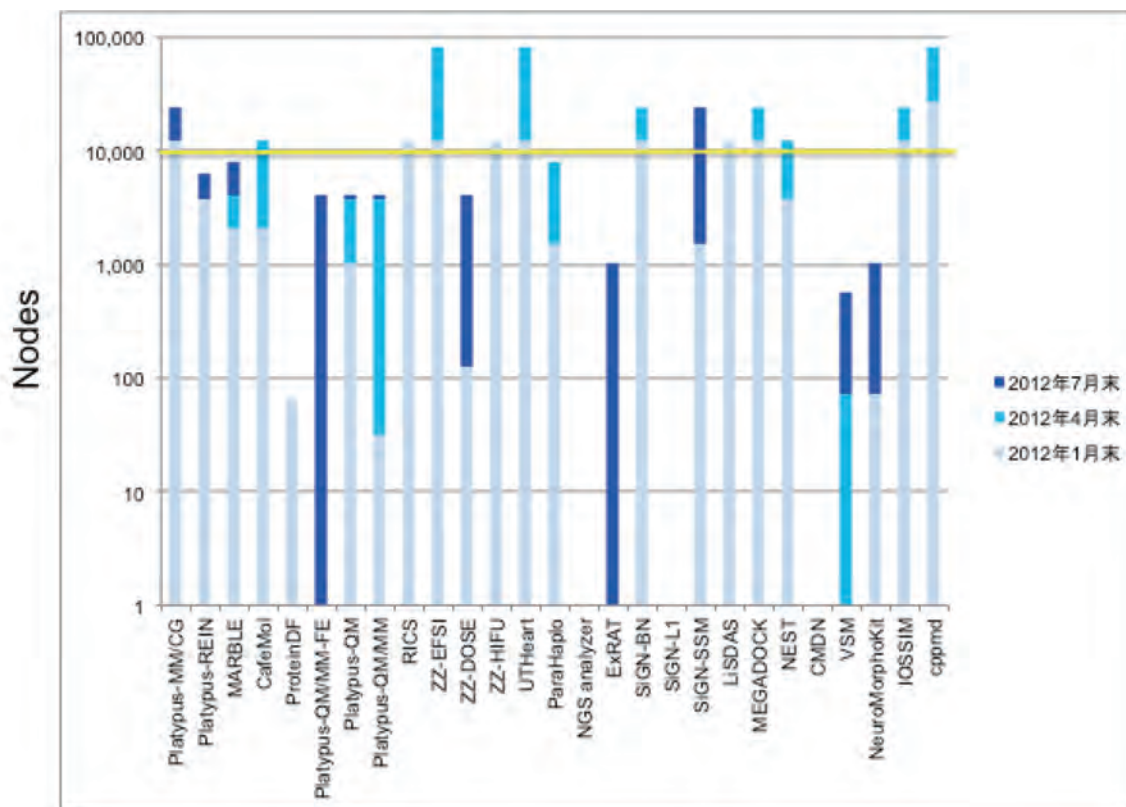


図2 それぞれのソフトの京での利用可能ノード数






ISLiM:ソフトウェア開発の現状と今後

理化学研究所
次世代計算科学研究開発プログラム
副プログラムディレクター
姫野龍太郎

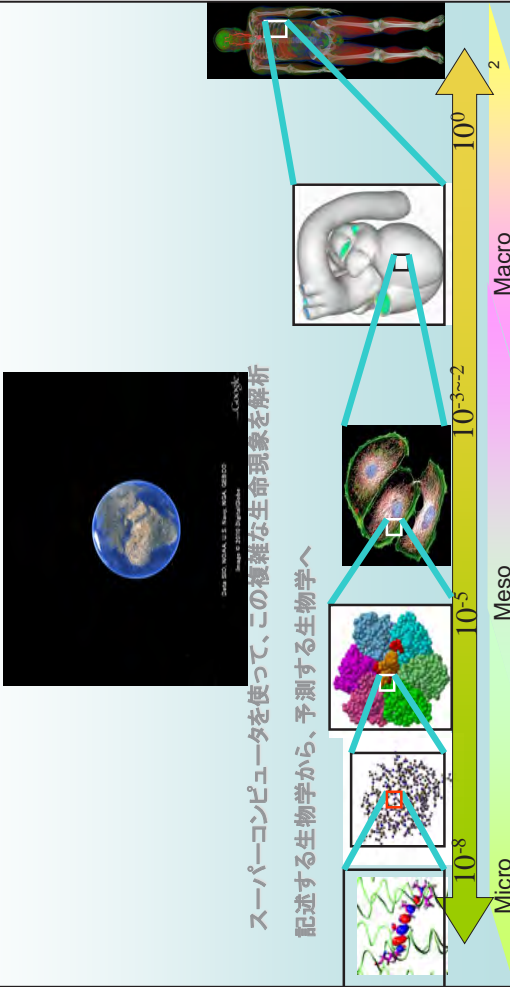



ISLiMソフトウェア研究開発報告会 1






そもそもの動機

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

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。

解析的アプローチ

分子スケール

細胞スケール

臓器全身スケール

スケール間連結アプローチ

イオンチャネル

神経細胞

神経回路

システム

脳・神経系

実験データから解析へのアプローチ

ベタスケール・シミュレーション

データ解析融合

大規模データ

基礎原理(方程式)

$$\frac{\partial v}{\partial t} + (v \cdot \nabla)v = -\nabla \cdot p + \frac{1}{Re} \Delta v$$

$$y = f(x) + g(x) \exp(xz) \dots$$



in → out

○生命現象の統合的な理解(=生命プログラムの理解・再現・予測)

○ヘルスサイエンスへの貢献、先端医療等の技術基盤の構築

ライフサイエンスの新たな手段の提供

3

これまでの開発の概要

- 経緯
 - 2006年10月から開発を開始、この時、分子・細胞・臓器全身・データ解析融合の4チーム。年度内に高度化チームを追加し、ソフト開発を支援
 - 2007年10月の脳神経系チームを追加
 - 2010年度中間評価
 - 2011年4月から京を使った開発が開始
- マネージメント
 - ソフトウェア開発の進捗状況を的確に把握し、支援
 - ソフトウェアにプライオリティーを設定、重点化

ISLiMソフトウェア研究開発報告会 4

ISLiM RIKEN

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

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)
概念設計	詳細設計	試作・評価・製造	性能チューニング			
システム	開発・製作・評価	実証				
次世代ナノ統合シミュレーション	第1回中間評価委員会	第2回中間評価委員会	実証			
次世代生命体統合シミュレーション	設計	建設				
計算機棟	設計	建設				
研究棟	設計	建設				

ISLiMソフトウェア研究開発報告会 註: 2010年度時点での計画 5

ISLiM RIKEN

各チームで取り組む課題 (2008年での整理)

整理前	整理後
<p>計算手法</p> <p>分子スケール</p> <p>細胞スケール</p> <p>臓器全身スケール</p> <p>データ解析融合</p>	<p>分子スケール</p> <p>細胞スケール</p> <p>臓器全身スケール</p> <p>データ解析</p> <p>脳神経</p>
<p>QM/MM</p> <p>全電子量子化学計算</p> <p>ハブアブリットQM/MM法</p> <p>分子動力学計算法</p> <p>粗視化シミュレーション</p> <p>蛋白質の構造サンプリング技術</p> <p>細胞シミュレーション統合プラットフォーム</p> <p>肝細胞 (代謝)</p> <p>膵臓β細胞</p> <p>イオンチャネル</p> <p>血小版</p> <p>血流シミュレーション</p> <p>心臓シミュレーション</p> <p>全身ホメオスタシス</p> <p>超音波伝播シミュレーション</p> <p>肺シミュレーション</p> <p>大規模遺伝子ネットワーク探索ソフトウェア</p> <p>生命体データ同化ソフトウェア</p> <p>タンパク質ドッキング解析ソフトウェア</p> <p>大規模SNP解析ソフトウェア</p>	<p>QM/MM</p> <p>全電子量子化学計算</p> <p>膜タンパク質・代謝酵素</p> <p>粗視化シミュレーション</p> <p>蛋白質の構造サンプリング技術</p> <p>細胞シミュレーション統合プラットフォーム</p> <p>肝細胞・肝小葉 (代謝)</p> <p>循環器系シミュレーション (血球・血小板・血流・血管網・心臓)</p> <p>全身ホメオスタシス</p> <p>超音波伝播シミュレーション</p> <p>大規模遺伝子ネットワーク探索ソフトウェア</p> <p>肺がん治療薬</p> <p>大規模SNP解析ソフトウェア</p> <p>神経細胞・高所回路シミュレーション</p> <p>脳全体シミュレーション</p>
<p>共通的な開発対象の認定</p> <p>開発対象の認定による整理見直し</p> <p>共通的な開発対象の認定</p>	<p>他チームとの連携による効率化</p>

6

ISLiM RIKEN

2008年度に設定したチーム間連携課題

提供	受合	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	肺がん治療薬	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	タンパク質ドッキング	代謝・膜	膜での輸送モデル	血液形成	血液形成	神経可塑性・変性・イオンチャネル
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	複雑細胞から臓器へ・血栓形成	細胞プラットフォーム供給	細胞の変形
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	細胞の変形	細胞から高次機能までの一貫
脳神経系	神経細胞データ・高所回路データ	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	運動制御臓器制御	
高度化	解析・高速化	解析・高速化	解析・高速化	解析・高速化	解析・高速化	解析・高速化

ISLiMソフトウェア研究開発報告会 7

ISLiM RIKEN

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ー、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。

量子化学計算 (より小さく、より長時間、より高密度)

分子動力学計算

粗視化モデル計算 (より大きく、より長時間、より近似的)

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。

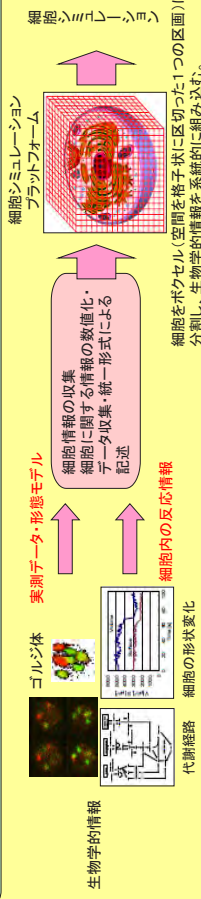
2011.1.23時点

ISLiMソフトウェア研究開発報告会 8

細胞スケール研究開発

開発目標: ペタフlops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・肝細胞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した。現在、1024並列までの性能テストを終了

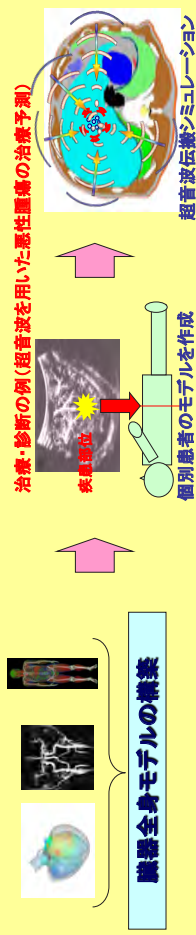
2011.1.23時点

9

臓器全身スケール研究開発

開発目標: ペタフlops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治療への基盤を構築した。

2011.1.23時点

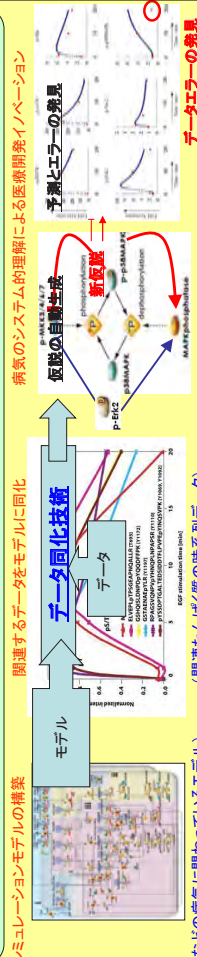
ISLiMソフトウェア研究開発報告会

10

データ解析融合研究開発

達成目標: ペタフlops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。



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の開発が8000を超えるcoreで稼動するようになった。

ISLiMソフトウェア研究開発報告会

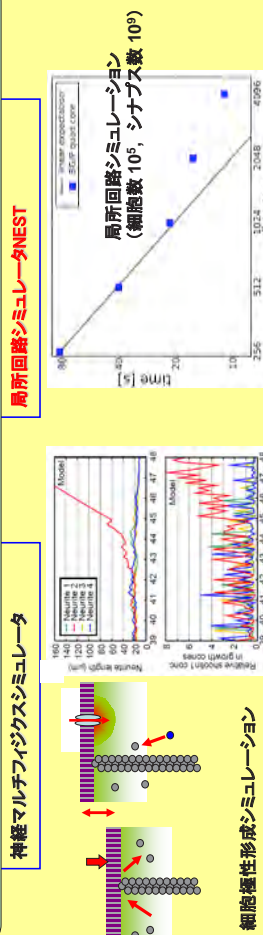
11

2011.1.23時点

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とを開発、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。



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10⁹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

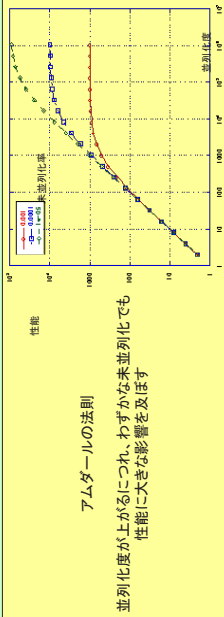
2011.1.23時点

12

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ッ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。



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。Cpmd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。

2011.1.23時点

13

ISLIMソフトウェア研究開発報告会

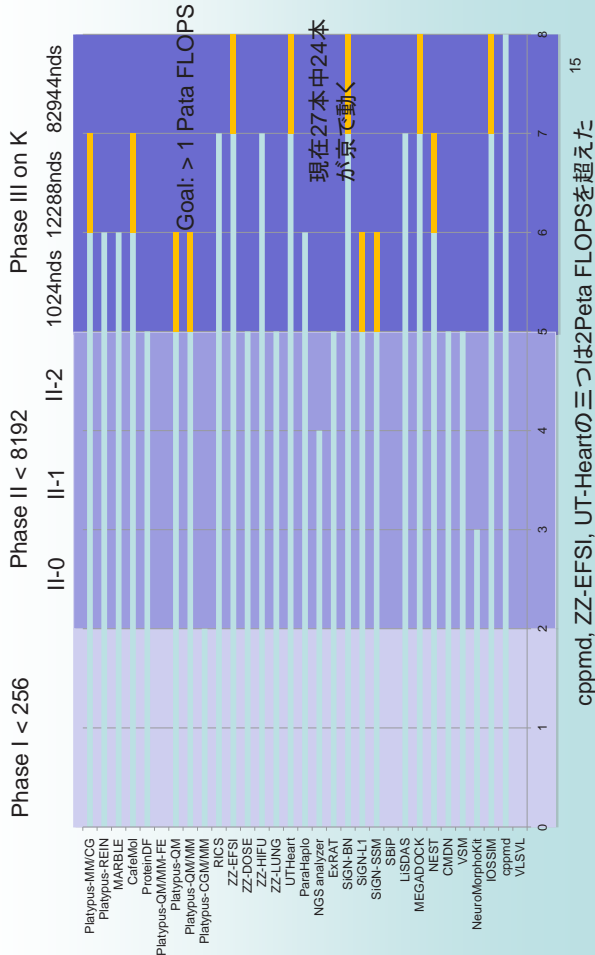
開発ソフトウェア (全34本)

ソフトウェア名称	説明
Platyus-MM/CG	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
Platyus-REIN	レプリカ交換分子動力学計算インターフェイス
MARBLE	全原子分子動力学計算
CafeMol	粗粒度モデル計算
ProteinDF	密度汎関数法に基づくタンパク質全電子波動関数計算
Platyus-OM/MM-FE	ハイブリッドQM/MM反応自由エネルギー計算
Platyus-OM	量子化学計算
Platyus-OM/MM	量子化学計算/分子動力学計算
Platyus-CGM/MM	粗視化モデル計算/分子動力学計算
RICS	細胞シミュレーションプラットフォーム
ZZ-EFSI	全身スケールシミュレーション (ボクセル構造流体力学解析ソフトウェア)
ZZ-RBC	微小循環シミュレータ (組み込み境界法による微小循環ソフトウェア)
ZZ-DOSE	重粒子線治療シミュレーション
HIFU	低周波超音波シミュレーション
HIFU	ボクセル超音波伝播シミュレーション
ZZ-LUNG	肺呼吸・肺循環シミュレーション
UTHeart	マルチスケール・マルチフィジクスの心臓シミュレーション

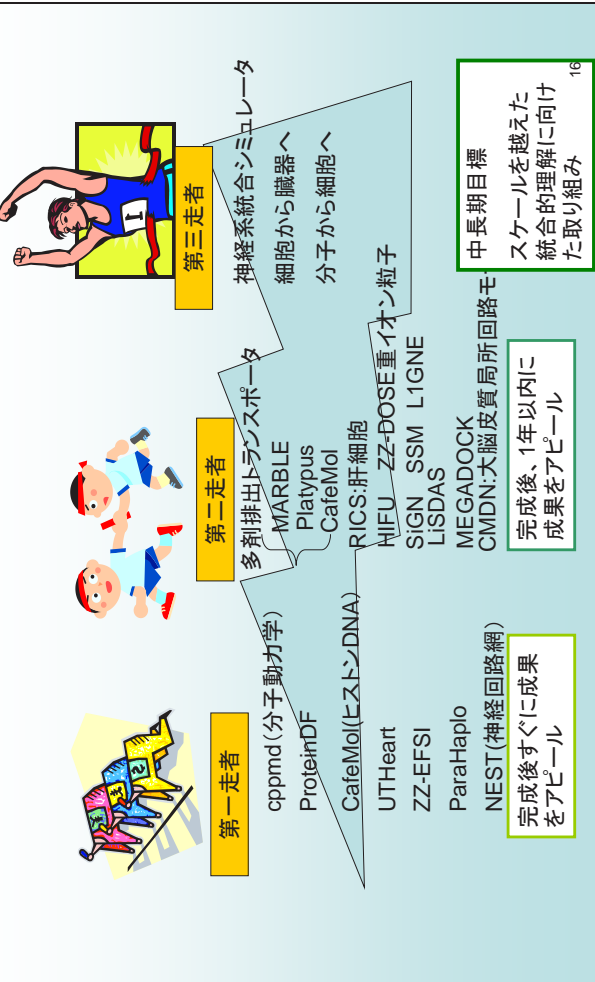
ソフトウェア名称	説明
ParaHaplo	ハプロタイプ関連解析に於ける統計検定を行うためのソフトウェア
NGS_analyzer	次世代シーケンズ解析ソフトウェア
EX-RAT	拡張PAI法によるSNP相寄せの全ゲノム関連解析ソフトウェア
SIGN-BN (SIGN)	大規模遺伝子制御ネットワーク推定プログラム
SIGN-L1 (LIGN)	再帰的正則化法による生体内分子の大規模ネットワーク推定プログラム
SIGN-SSM (SSM)	状態空間モデルによる時系列データからの遺伝子ネットワーク推定プログラム
SBIP	データ解析統合プラットフォーム
LSDAS	生命体データ同化ソフトウェア
MEGADOCK	複雑なタンパク質ドッキング解析プログラム
NEST	Neural Simulation Tool
GMDN	Cortical Microcircuit Developed on NEST
VSM	全視覚系モデルによる視覚情報処理の解析 (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)
NeuroMorphokit	神経細胞形態シミュレーションキット
IOSSIM	自由呼吸系全脳シミュレータ
cppmd	大規模並列用MDコアソフトウェア
LSV	分散並列大規模データ可視化システム
SPHERE	アプリケーションソフトウェア
VLSVL	大規模脳活性化物ラブラリ

ISLIMソフトウェア研究開発報告会

ISLIMのアプリ開発状況



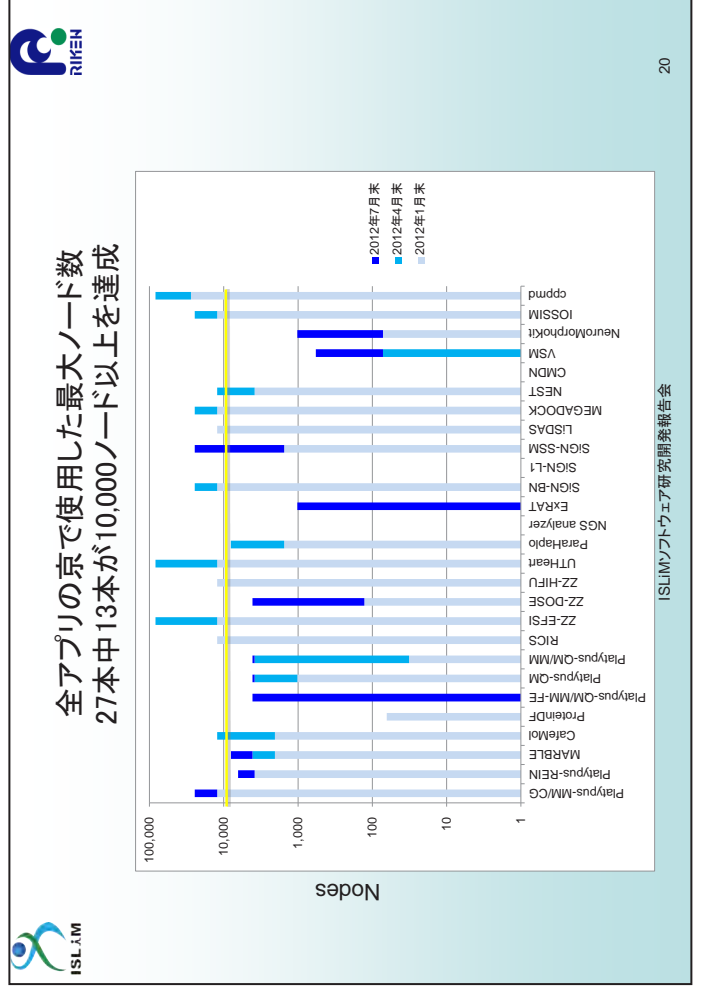
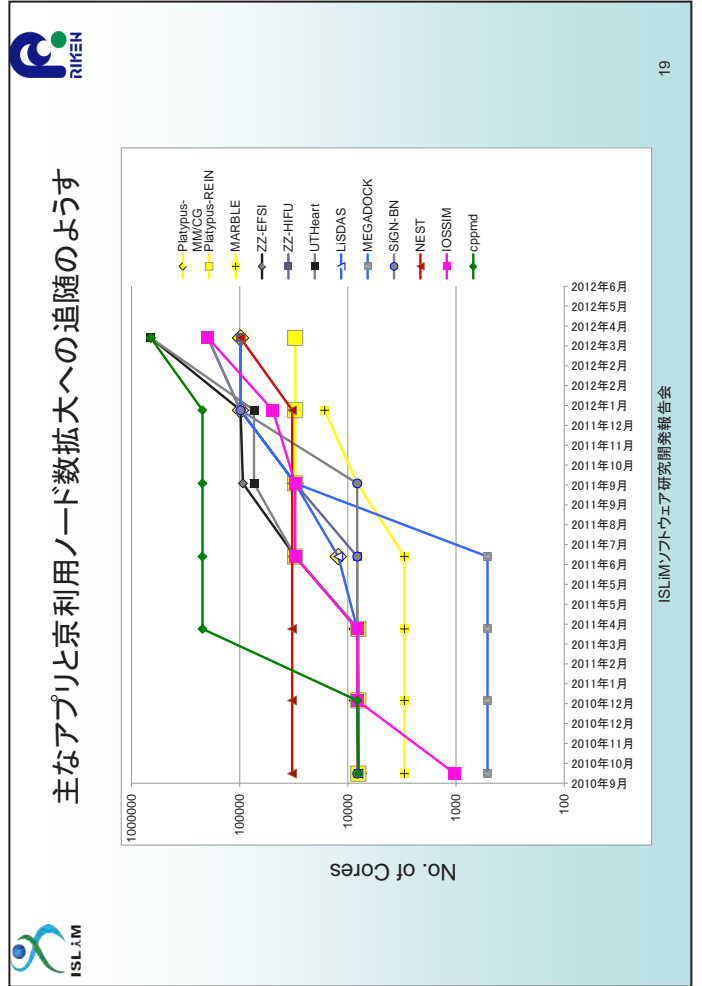
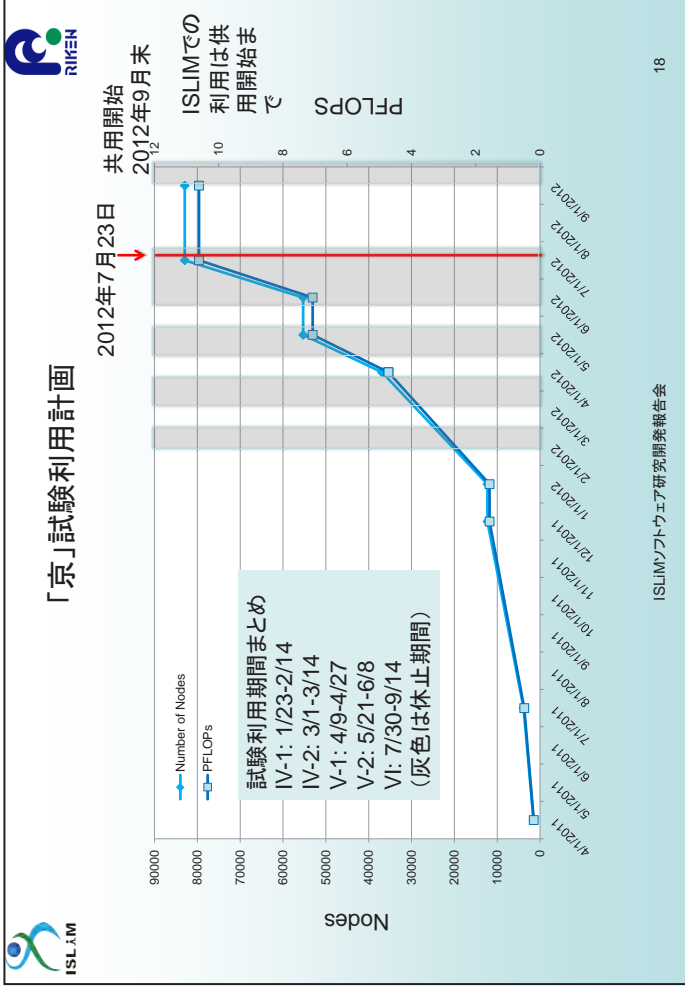
それぞれの成果アピール時期



昨年度からの大きな変化

- 平成23年4月より「京」を試験利用。段階的に利用を拡大、ソフトをチューニング（「京」共用開始前までが本プロジェクトで利用できる期間）
- 当初計画では、平成24年11月末まで利用可能
- 【「京」完成から約4ヶ月間】
- 共用開始を平成24年9月末に変更（平成24年2月7日 第8回HPC計画推進委員会）
【実質約2ヶ月短縮】

ISLiMソフトウェア研究開発報告会 17



分子スケール研究開発チーム成果報告

分子スケールチーム

木寺 詔紀

次世代計算科学研究開発プログラム
分子スケール研究開発チーム チームリーダー



発表者紹介

- 1982年3月 京都大学工学研究科高分子化学専攻博士課程退学
- 1982年8月 Cornell大学 化学科 博士研究員
- 1986年2月 京都工芸繊維大学 工芸学部 助手
- 1988年4月 蛋白工学研究所 主任研究員
- 1996年4月 京都大学大学院 理学研究科 助教授
- 2001年4月 横浜市立大学大学院 総合理学研究科 教授
- 2006年10月 理化学研究所次世代計算機科学研究開発プログラム チームリーダー (兼任)
- 2011年4月 理化学研究所 HPCI 計算生命科学推進プログラム 副プログラムディレクター (兼任)

研究分野

計算生物学、生物物理学

分子スケールチーム

木寺詔紀

次世代計算科学研究開発プログラム
分子スケール研究開発チーム チームリーダー

1. 目的

生命現象の最下層にある生体分子の機能発現過程のシミュレーションは、それ自体として時間、空間の階層性を持っており、それぞれの階層に対応したシミュレーション方法を必要とする。短時間の比較的少数の原子が関わる化学反応を含む高精度の相互作用のレベルでは量子化学計算 (QM) を用いる。より長時間のタンパク質の全原子の動きのシミュレーションは、環境として存在する水や脂質分子をあらわに取り入れた系で、経験的な力場 (MM) に基づいた古典分子動力学計算を用いる。生理学的なレベルに近づく時間スケールでの現象をより巨大な系においてシミュレーションする場合は、MM をさらに粗視化したモデル (CG) に基づいて、環境を陰的に考慮する方法によって行う。分子スケールでは、生体分子に関わるあらゆる現象を対象としたシミュレーションを可能とするために QM-MM-CG という 3 階層にわたる方法論を開発する。さらに、これらの階層を接続し、それぞれの方法の限界を超えるマルチスケール法として、QM と MM とを連成した QM/MM 法、さらに MM と CG を練成した MM/CG 法を開発する。これらの方法論によるシミュレーションをスーパーコンピュータ京の性能を十分に生かして行うことを可能とするソフトウェアを開発する。さらに、実証研究の対象として、多剤排出トランスポーター、脂質代謝酵素複合体を分子スケールの共通ターゲットとして選び、スーパーコンピュータ京上で多階層でのシミュレーションを行うことで、これらのソフトウェアの有用性を示す。

2. 現時点での成果

2.1 ProteinDF (QM)

巨大複雑系であるタンパク質の状態を最も精密なレベルでシミュレートする目的で、タンパク質全体を含んだ全電子カノニカル分子軌道計算を行う。計算法には基底関数にガウス型関数を使用する密度汎関数法 (DFT) を採用した。実績のある汎関数 B3LYP を実装し、高度並列化を行っている。現時点で 8,000 並列を達成している。適用例として、インシュリン 6 量体、ミオグロビン、

チトクローム c などのカノニカル分子軌道を求め、カノニカル分子軌道の重要性を示し、分子軌道と分子機能との関わりを明らかにした。現在、光合成アンテナ蛋白質部分構造の計算を行っている。

2.2 Platypus-QM/MM (QM/MM)

QM と MM を連成することで、QM/MM 分子動力学シミュレーションを高精度で高速に実行することができる。*ab initio* 量子化学計算として RHF, UHF, R-DFT, U-DFT, CIS, CIS(D), CIS-DFT, MP2, CASSCF を組み込み、基底状態だけでなく励起状態解析にも対応し、電子移動や生化学反応を on-the-fly 計算によりシミュレートできる。さらに、Chain-of-state 法を利用して、反応の最小自由エネルギー経路が得られる。QM 計算部は最大 32,768 コアの並列(並列化率 99.9888%)を、Chain-of-state 法による QM/MM 計算では 98,304 コアの並列(Weak Scaling 効率 93.36%)を達成している。また、遠距離力を高精度で計算できる Zero-Dipole Summation 法を高度化した。適用例として発光蛋白質 Sirius の発色団の励起状態解析による発光機構の解明と RNase HI の加水分解反応機構の解析を行っている。

2.3 Platypus-QM/MM-FE (QM/MM)

生体分子機能に関わる化学反応をハイブリッド QM/MM 法を用いて解析し機能発現の分子機構の解明を可能にする方法論を開発した。QM/MM-FE 法では、MM 分子力場を用いた MD 計算による MM 領域の構造サンプリングを行い、その MM 分布によって定義される自由エネルギー曲面上で QM 領域の反応経路探索を行うことで、MM 領域のゆらぎの中で起こる酵素反応を定量的に評価することができる。これまでに 700 ns の MD サンプルによる自由エネルギー曲面上での QM 構造最適化に成功している。現在、京で 4000 並列を達成している。 α -アミラーゼの酵素反応機構に適用し、タンパク質の構造揺らぎの反応自由エネルギーへの寄与を明らかにした。

2.3 Marble (MM)

全原子分子動力学法のソフトウェアである。空間

分割、PME 法、シンプレクティック剛体法などの高効率・高精度のアルゴリズムを用いており、現在、京で 8000 並列超を実現している。応用として MD と NMR、X 線溶液散乱実験との連携技術の開発を行った。また、共通ターゲットである多剤排出トランスポーターの分子動力学シミュレーションを行い、プロトン移動による構造変化誘導、トランスポーターに対する薬剤の結合・排出機構についての全原子シミュレーションを実施した。

2.4 Platypus-REIN (MM/CG)

レプリカ交換分子動力学法を行うためのソフトウェアである。レプリカ交換分子動力学法は異なる環境変数で行われる分子動力学計算で駆動される多数のレプリカが確率的に他の環境変数に移行することによって、通常分子動力学計算では到達できない広範囲の構造アンサンブルを精密に求めることができる。ここではさらに、多次元の環境変数を選択することでさらなるサンプリング空間の拡張をめざす多次元レプリカ交換法を採用している。京で 3 万並列超を実現するとともに、溶液中の糖鎖構造のモデリングや蛋白質によるペプチド認識機構の解明などに適用した。さらに細胞膜中でのアミロイド前駆体のシミュレーションを行うことで、アミロイド形成過程の初期のモデリングを行っている。「京」に資するソフトウェアとして、ISLiM ホームページ上でソースコードも含めて公開を行った。

2.5 Platypus-MM/CG (MM/CG)

MM と CG の二つのモデルを共役項で制約させながらひとつのシミュレーションとして行う。それによって、MM の精度と CG の効率を両立させることを目指している。主要な方法論は、マルチスケール拡張サンプリング(MSES)法、On-the-fly ストリング法である。MSES 法は、上記のシミュレーションをそのまま分子動力学計算として実施し、最終的に CG からの影響のない MM の構造アンサンブルを得るために、共役項の強度を変数とする Hamiltonian 交換法を用いている。On-the-fly ストリング法は、同様なシミュレーションを経路探索の問題として行う。CG のポテンシャルを flat にして、共役項の強度を大きくすることによって CG 自由度の上での MM ポテンシャルの平均力場を求めることができ、最小自由エネルギー経路を得ることができる。MSES 法は天然変性タンパク質 Sortase A の disorder-order 転移、Barnase-Barstar のタンパク質間相互作用に、On-the-fly ストリング法は Adenylate kinase の

構造変化に適用した。

2.5 CafeMol (CG)

粗視化ポテンシャルを用いて、タンパク質、核酸、脂質を含んだ系の大規模、長時間の分子動力学シミュレーションを行った。これまでに、蛋白質、DNA および脂質それぞれの単体系および核酸-蛋白質、脂質-蛋白質の複合体の粗視化モデルの開発および実装を完了した。現在、京への実装と、レプリカ交換法によって 98000 コアの並列を達成している。これまでに、タンパク質のフォールディング、キネシンの微小管上の歩行、多剤排出トランスポーターの薬剤排出、ヌクレオソームの動態などのシミュレーションに適用し、実験または既存のシミュレーション手法では困難であった、それぞれの運動の機構を明らかにしてきた。特に、多剤排出トランスポーターでは、機能的回転の機構および薬剤取り込みの経路についての知見を与えた。

2. プロジェクト終了時の達成目標

それぞれの方法論によるシミュレーションを京の性能を十分に生かして以下の実証研究の目標を以下にあげる。

ProteinDF

光合成反応中心、光合成アンテナ蛋白質のカノニカル全電子計算

Platypus-QM

数万電子系の基底状態/励起状態の電子構造の計算

Platypus-QM/MM

数千電子系とそれを取り巻く数 10 万古典原子による反応自由エネルギー地形の計算

Platypus-QM/MM-FE

多剤排出トランスポーターのプロトン移動の自由エネルギー計算

MARBLE

多剤排出トランスポーターの薬剤排出過程のシミュレーション

Platypus-REIN

タンパク質の構造変化、構造予測および、タンパク質間の結合自由エネルギープロファイル (数万~十万粒子系)

Platypus-MM/CG

多剤排出トランスポーターの構造変化経路計算
脂質 β 酸化多酵素複合体の構造変化計算

CafeMol

ヌクレオソーム (DNA ヒストン複合体) 動態のシミュレーション

以上

ISLiM RIKEN

MM/CGプログラムμ²-libによる タンパク質の大規模サンプリング

理化学研究所
次世代計算科学研究開発プログラム
分子スケールチーム チームリーダー
木寺詔紀

ISLiMソフトウェア研究開発報告会 1

ISLiM RIKEN

背景・目的

Molecular Scale Team

QM ProteinDF
QMM
MM Marble
MM/CG
CG CafeMol

ISLiMソフトウェア研究開発報告会 2

ISLiM RIKEN

概要・アプローチ

MMのサンプリングの効率化

高次元/*ab initio*の効率化
Multicanonical MD
Replica Exchange MD

タンパク質の機能発現に伴う運動は、
比較的次元のしかし複雑な反応座標上で起こる
CGモデルで *prior knowledge* の空間を表現する

低次元/*prior knowledge*の効率化
Stirred MD
Targeted MD

ISLiMソフトウェア研究開発報告会 3

ISLiM RIKEN

マルチスケールの原理

MMがCGで
引きずられる



共役力

CGはMMを
引っ張る

$$H(\mathbf{x}; \mathbf{z}) = V(\mathbf{x}) + U(\mathbf{z}) + k/2(\theta(\mathbf{x}) - \mathbf{z})^2$$

MM \mathbf{x} CG \mathbf{z} coupling

ISLiMソフトウェア研究開発報告会 4

共役消去のふたつの方法

Hamiltonian Exchange (multicoopy)

$k \rightarrow 0$

$H(\mathbf{x}; \mathbf{z}) = V(\mathbf{x})$ **MM**

Path Search (multicoopy)

$k \rightarrow \infty$

$U(\mathbf{z}) = 0$

$-F(\mathbf{z}) \rightarrow \ln \int d\mathbf{x} e^{-\beta V(\mathbf{x})} \delta(\theta(\mathbf{x}) - \mathbf{z})$

\mathbf{z} 上の **MM** の平均力ポテンシャル

coupling

\mathbf{x} **CG** \mathbf{z}



Multiscale Enhanced Sampling (MSES)

K Moritsugu, et al. *J. Chem. Phys.* 2010.

On-the-fly String Method

L. Maragliano, E. Vanden-Eijnden, *Chem. Phys. Lett.* 2007.


ISLiMソフトウェア研究開発報告会 5

現在までの研究開発成果

μ2-lib (Platyus-MM/CG) 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:
寺田 透 (東大)

ライブラリの構成

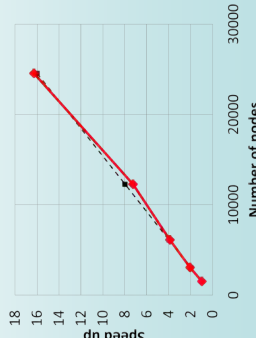


ライブラリの特徴



計算対象をオブジェクトとして扱う
Amber, Charmm力場に対応
OpenMP/MPIによる並列化

計算性能

- 24,576ノード(196,608コア)まで計算性能が直線的に向上
- 高並列時でも高い実行効率を実現 (24,576ノード並列時で約12.3%)

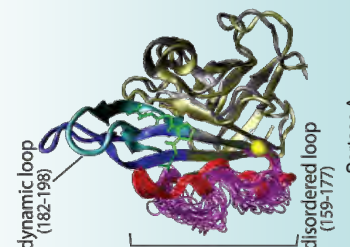


ISLiMソフトウェア研究開発報告会 6

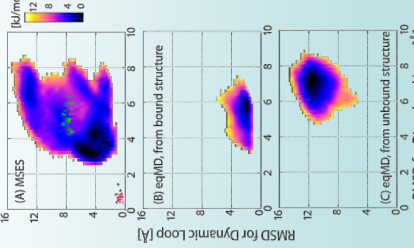
Multiscale Enhanced Sampling (MSES): 森次 圭(理研)

天然変成部位のシミュレーション - Sortase A





dynamic loop (182-198)
disordered loop (159-177)
Sortase A

Enhance the disordered regions



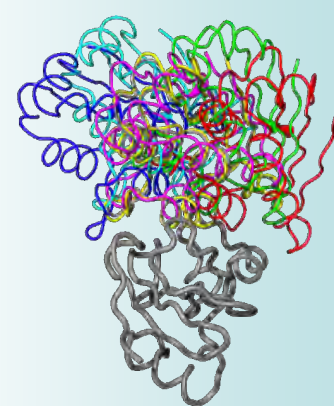
(A) MSES
(B) eqMD, from bound structure
(C) eqMD, from unbound structure

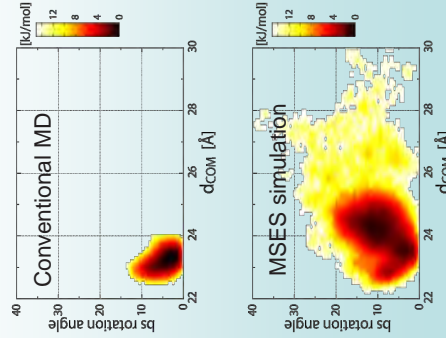
ISLiMソフトウェア研究開発報告会 7

Multiscale Enhanced Sampling (MSES): 森次 圭(理研)


Barnase/Barstar複合体形成のシミュレーション






Conventional MD
MSES simulation

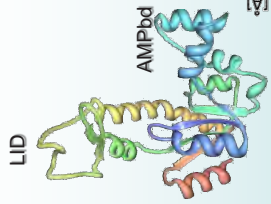
ISLiMソフトウェア研究開発報告会 8



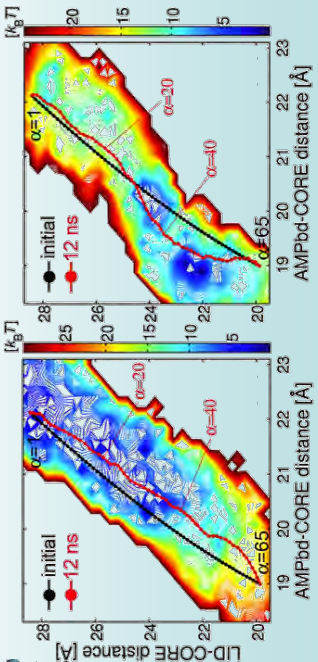


On-the-fly String法: 松永康佑 (理研)


Adenylate kinaseの基質結合に伴う
立体構造変化の最小自由エネルギー経路




LID
AMPbd



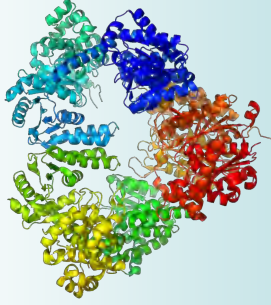
ISLiMソフトウェア研究開発報告会 9



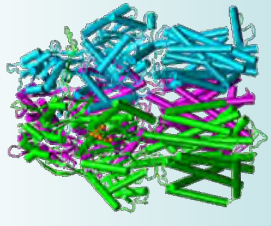


今後の展開

MSES
脂質β酸化多酵素複合体の
大規模構造探索



On-the-fly String法
多剤排出トランスポーターAcrBの
薬剤排出過程最小自由エネルギー経路探索



Onsager-Machlup法の開発とμ2-libへの実装: 藤崎弘士 (日本医大)
Particle Filter法の開発とμ2-libへの実装: 松永康佑 (理研)

ISLiMソフトウェア研究開発報告会 10

Replica-exchange interface program: REIN

宮下 尚之

理化学研究所 生命システム研究センター
分子機能シミュレーション研究チーム 研究員




発表者紹介

2003年 3月 総合研究大学院大学数物科学研究科機能分子科学専攻博士課程修了
2003年 4月 東京大学 分子細胞生物学研究所 学術研究支援員
2005年 4月 ボストン大学 化学科 Research Associate
2007年 4月 岐阜大学 人獣感染防御研究センター 助教
2007年 12月 理化学研究所 次世代計算科学研究開発プログラム 研究員
2011年 4月 理化学研究所 生命システム研究センター 研究員
2011年 4月 理化学研究所 次世代計算科学研究開発プログラム 兼務
2011年 4月 理化学研究所 計算科学研究機構 兼務




研究分野

物理（理論）・生物物理・分子シミュレーション



Replica-exchange Interface Program REIN

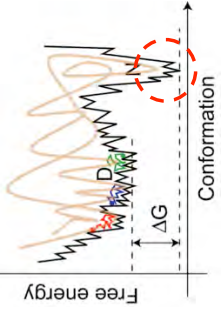
Naoyuki Miyashita
RIKEN QBIC, RIKEN CSRC, RIKEN AICS

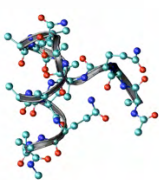
12年12月13日本曜日 1

Replica-exchange Molecular dynamics simulation (REMD) can sample wider conformational spaces

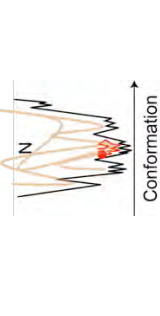
Y. Sugita, and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141–151, (1999)



MD



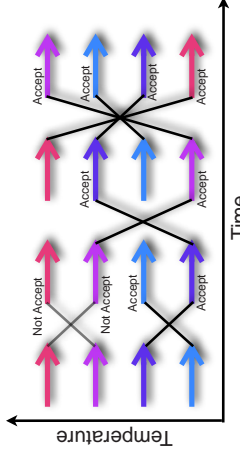
(AAQAQA)₃



REMD

12年12月13日本曜日 2

Replica-exchange molecular dynamics (REMD)

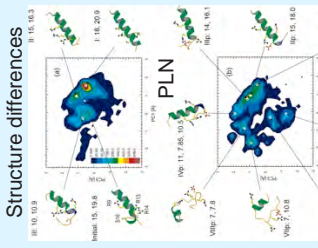


- $H(p_i, q_i) = K(p_i) + E_0(q_i) + U(q_i)$
- Transition probability:
- Metropolis criteria: $\min[1, \exp(-\Delta)]$
- Exchange
- $\Delta_x = \beta_m(U_m(q_i) - U_m(q_j)) - \beta_n(U_n(q_i) - U_n(q_j))$

12年12月13日本曜日 3

Example of REMD Simulations

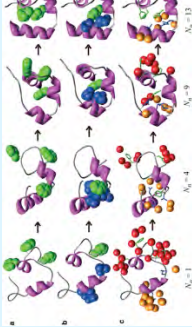
Structure differences



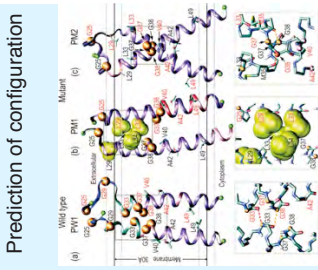
PLN

pPLN

Protein Folding

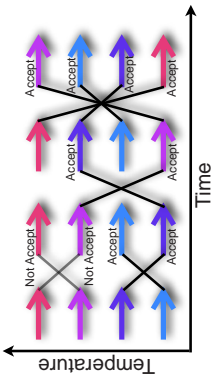


Prediction of configuration



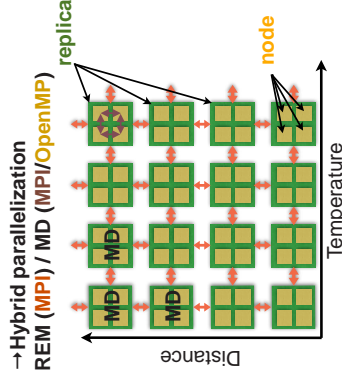
12年12月13日本曜日 4

Replica-exchange molecular dynamics (REMD)



- Replica-exchange molecular dynamics simulation [1] has been used widely as useful enhanced sampling method.
- Better Method for hybrid parallelization

Layer	REMD	Parallel
3	REM	MPI
2	MD	MPI
1	MD	OpenMP



→ Hybrid parallelization
REM (MPI) / MD (MPI/OpenMP)

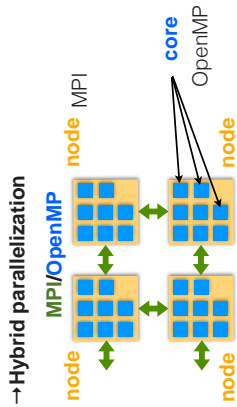
- Replica-exchange umbrella sampling (REUS) [2]: Not only temperature but also distance, angle, dihedral and so on.
- Multi-dimensional REMD (MREM) [2]

[REM] K. Fukushima, and K. Nemoto, J. Phys. Soc. Jpn. 65, 1604-1608 (1996)
 [1] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. 314, 141-151, (1999)
 [2] Y. Sugita, A. Kitao and Y. Okamoto, J. Chem. Phys. 113, 6042-6051 (2000)

12年12月13日水曜日

6

K computer



→ Hybrid parallelization
MPI/OpenMP

- 10 PFlops : 6/2011 No. 1 (8.162PF), 11/2011 No.1(10.51PF), 6/2012 No.2, 11/2012 No.3
- Open public on 9/28/2012
- Trial period from 4/2011 to 9/2012
- 88,128 node, 705,024 core (super parallel computer) , 8 core/node
- 3D torus network
- SPARC, Fujitsu compiler (C, C++, Fortran), MPI-2, OpenMP

12年12月13日水曜日

5

Conventional MD packages and REMD module

MD Packages	Language/parallelization	Hybrid	Speed	REMD	MREM	on K
CHARMM [1]	Fortran/MPI	X	Slow	○	○	X
AMBER [2]	Fortran/MPI	X	Medium	○	X	X
NAMD [3]	Fortran/MPI/ptthread	○	Fast	○	○	△*
GROMACS [4]	C/MPI/OpenMP	○	Fast	○	○	△
DESMOND [5]	C/MPI/ptthread	○	Fast	○	X	X
TINKER [6]	Fortran/MPI	X	Slow	X	X	X
MARBLE [7]	C/MPI/OpenMP	○	Fast	X	X	○ (Native)

- Complex input files and batch file for REMD/MREM in each package

- No perfect MREM package on K

[1] B. R. Brooks et al., J Comput. Chem 30: 1545-1614, 2009
 [2] D. A. Pearlman et al., Comput. Phys. Comm. 91, 1-41, 1995
 [3] J. C. Phillips et al., J. Comput. Chem. 26:1781-1802, 2005
 [4] H.J.C. Berendsen et al., Comput. Phys. Comm. 91:43-56, 1995
 [5] K. J. Bowers et al., in SC '06: Proceedings of the 2006 ACM/IEEE conference on Supercomputing, New York, NY, USA, 2006. ACM Press.
 [6] J. W. Ponder and F. M. Richards, J. Comput. Chem., 8, 1016-1024, 1987.
 [7] M. Ikeguchi, J Comput Chem 25: 529-541, 2004

* We are trying to install NAMD into K-computer with Prof. Kamada@kyobe U.

12年12月13日水曜日

7

interface software for Replica-exchange method



Interface software can use several conventional MD programs for REMD simulations.

Packages	MD	Language/parallelization	style	Modification of MD program	Recompile of MD	on K
aarex.pl in MMTSB toolsets [1]	CHARMM AMBER NAMD	Perl/ TCP socket	binary run (fork, exec function)	no	no	○ X
PLUMED [2]	GROMACS NAMD DLPOLY AMBER ACEMD LAMMPS QUANTUM-ESPRESSO CPMD	C/ (depend on each MD)	Plugin style (It needs recompile with patch)	Necessary	Necessary	○ X

[1] M. Feig et al., Journal of Molecular Graphics and Modelling 22 (2004) 377-395

[2] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia and M. Parrinello, PLUMED: a portable plugin for free energy calculations with molecular dynamics, Comp. Phys. Comm. 180, 1961 (2009), Available also on [arXiv:0902.0874](https://arxiv.org/abs/0902.0874)

Advantage :

- The several functions on each MD package
- Easy to perform REMD or MREM (simple input file)
- Not support on K

12年12月13日水曜日

8

Very easy execution -- three steps

```

1) Input_builder → rein.inp
#input_builder ib.inp

2) Batch_builder → rein.sh
#batch_builder bb.inp

3) submitting of job
# pjsub rein.sh

Restart
#batch_builder bb.inp restart

rein_converter → creating the trajectories series
#rein_converter rc.inp
    
```

12年12月13日本曜日

13

Example: input file of Batch builder

```

# control parameters in batch_builder
rein_input = rein.inp # REIN input file name

[BATCH]
batchfile_name = rein.sh # batch filename
job_scheduler = K # job scheduler type (KI/GE/ATCC)
queue_name = small # queue name, k:(SMALL/LARGE), fx10: REGULAR.
num_threads_md = 8 # number of threads in md
num_mpiproc_md = 4 # # of mpi process for md
cpu_time = 01:00:00 # cpu time
num_core_cpu = 8 # number of core in cpu, for k = 8, for fx10 = 16, ...

[REMO]
replica_exchange = yes # (YES/NO): replica exchange / umbrella sampling
num_of_exchange_steps = 100 # number of exchange steps
    
```

12年12月13日本曜日

15

Example: Input file of Input builder

```

# control parameters in input_builder
rein_input = rein.inp # rein input file name

[REMO]
axis001 = temperature # axis (TEMPERATURE/ANGLE/IC)
num_replicas001 = 8 # # of replicas for the axis
range001 = 300.0 304.0 # lowest, highest value
division001 = exp # (EQUIV/EXP/CYCLE)

axis002 = harmonic
num_replicas002 = 8
range002 = 2.0 8.0
division002 = equiv
distance002 = AUP CN 1-1 AUP CL 1-1
spring_const002 = 2.0

[MD]
md_program = md program (NAMD/MARBLE)
md_exe_file = ./namd2 # md exe file
md_temperature = 300.0 # md temperature
md_steps = 1000 # md steps / exchange
system_size_x = 23.4 # system size --automatically create pme grids
system_size_y = 23.2 #
system_size_z = 23.9 #
    
```

12年12月13日本曜日

14

Temperature - distance
2D-REMD

ex) Alanine Dipeptide
REIN/NAMD, NVT
MD steps: 1000 steps (2ps)
Exchange: 100 exchanges
Number of Replicas:
Temperature : 8 replicas
Distance : 8 replicas
Total: 8 x 8 = 64 replicas
Range of Temperature: 300-360K
Range of Distance : 2.0-8.0Å

\$ input_builder ib.inp
実行

Output format

Replica#, condition#, Temperature, (Distance etc..), Total, Kinetic, Potential, Extra

```

##### EXCHANGE_STEP = 1
1 1 290.14150 -1624.33470 -1994.49040
2 2 309.04970 -1677.37840 -2071.65680
3 3 314.56290 -1658.91060 -2060.22250
4 4 313.83160 -1657.91540 -2058.29250
5 5 304.91160 -1642.47860 -2031.47760
6 6 309.84760 -1635.80170 -2031.09000
7 7 289.44280 -1685.76920 -2067.79130
8 8 285.62440 -1683.47140 -2051.09170
##### EXCHANGE_STEP = 2
1 2 306.76840 -1664.28540 391.36800
2 1 317.18100 -1651.98420 404.65210
3 4 310.83750 -1629.92330 396.55920
4 3 284.34980 -1650.03960 362.76680
5 6 294.21380 -1680.86470 375.31110
6 5 336.64260 -1606.86420 429.48070
7 8 296.55280 -1698.13390 378.33510
8 7 299.56080 -1662.84080 382.11260
    
```

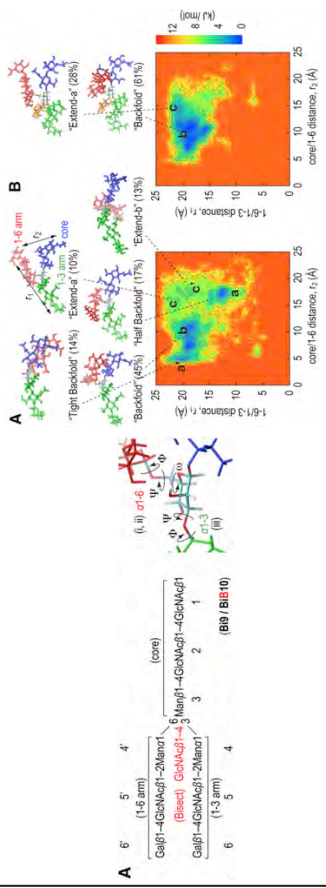
12年12月13日本曜日

16

Applications of REIN (version 0.1)

Structural Diversity and Changes in Conformational Equilibria of Biantennary Complex-Type N-Glycans in Water Revealed by Replica-Exchange Molecular Dynamics Simulation

Suyong Re, Naoyuki Miyasahita, Yoshiaki Yamaguchi, Yuji Sugita, Biophys. J. 101, Pages L44-L46 (2011)



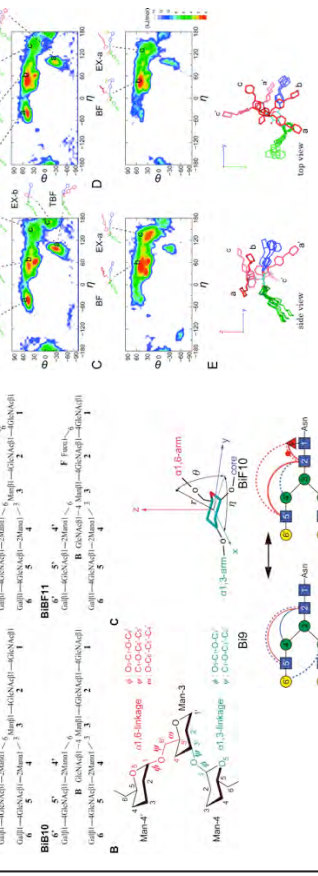
18 12年12月13日本曜日

17 12年12月13日本曜日

Applications of REIN (version 0.1)

Effect of Bisecting GlcNAc and Core Fucosylation on Conformational Properties of Biantennary Complex-Type N-Glycans in Solution

Wataru Nishima, Naoyuki Miyasahita, Yoshiaki Yamaguchi, and Suyong Re, J. Phys. Chem. B, 116 (23), pp. 8504-8512, (2012)



19 12年12月13日本曜日

REIN

Purpose: Structure prediction of a protein, Conformational prediction of proteins, Interactions among molecules.

Target: More larger system and Complex system

→ Very easy interface, and advanced-use is available too.

	*Current (REIN on RICC)	**REIN on K
System size	15,000 atoms	40,000 or more
MD steps	1,000 steps	10,000 steps or more
Number of cores for each MD	8 cores	32 cores
Number of replicas	64 replicas	640 replicas = 32 x 20
Total Number of cores	512 cores	20480 cores
Simulation time	52 ns x 64 replica = 3.3 μs	500 ns x 640 replica = 320 μs or more

* N-Glycans 20 ** Membrane proteins

20 12年12月13日本曜日

Summary

- Replica-Exchange Interface Program (REIN)
- REIN version 1.01 open (islim web page)
http://www.islim.org/islim-dl_j.html
E-mail to : yukimya+reinmiva@riken.jp
- Free software (GPL version 3)
- Interface program of Multi-dimensional Replica-exchange molecular dynamics (REMD) simulation
 - Temperature, Distance, Angle, Dihedral
- Binary MD currently NAMD2, near future MARBLE v0.6
- Three supporting softwares
 - Input builder
 - Batch builder
 - rein converter
- Fortran95/MPi/OpenMP
- PC cluster, K computer, FX10 super computer etc..

21

12年12月13日本曜日

Acknowledgement

- Dr. Yuji Sugita, RIKEN ASI, RIKEN QBIC, RIKEN AICS
- Dr. Yoshiki Yamaguchi, RIKEN ASI
- Prof. Mitsunori Ikeguchi, Yokohama City University
- Dr. Suyong Re, RIKEN ASI
- Dr. Wataru Nishima, RIKEN ASI
- Mr. Norio Takase, Isogo Soft
- This research was supported by
 - Research and Development of the Next-Generation Integrated Simulation of Living Matter, a part of the Development and Use of the Next-Generation Supercomputer Project of the Ministry of Education, Culture, Sports, Science and Technology (MEXT)
 - RIKEN Quantitative Biology Center (QBIC)
 - K-computer, RIKEN Advanced Institute for Computational Science (AICS)
 - JSPS KAKENHI Grant-in-Aid for Young Scientists (B) Number 24700299.
 - We thank the RIKEN Integrated Cluster of Clusters (RICC) at RIKEN for the computer resources used for the calculation

22

12年12月13日本曜日



22

12年12月13日本曜日

23

全原子分子動力学計算ソフトウェア MARBLE の開発と 多剤排出トランスポーターAcrB への応用

池口 満徳

横浜市立大学 大学院生命ナノシステム研究科
分子スケール研究開発チーム



発表者紹介

- 1994 年 3 月 東京大学大学院農学系研究科応用生命工学専攻博士課程修了
- 1994 年 4 月 東京大学農学部助手
- 2001 年 4 月 横浜市立大学大学院総合理学研究科助教授
- 2009 年 4 月 横浜市立大学大学院生命ナノシステム科学研究科准教授

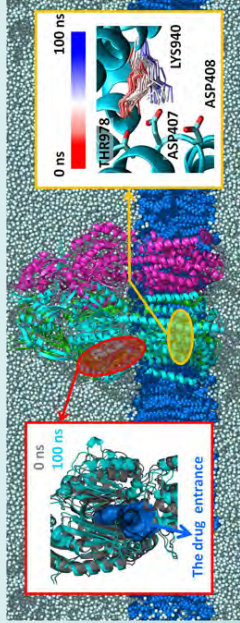
研究分野

生物物理学, 分子シミュレーション

ISLiM

全原子分子動力学計算ソフトウェア MARBLEの開発と 多剤排出トランスポーターAcrBへの応用

横浜市立大学 大学院生命ナノシステム研究科
分子スケール研究開発チーム
池口 満徳



ISLiMソフトウェア研究開発報告会 1

ISLiM

背景・目的

- 背景
 - タンパク質やDNAなどの生体高分子の立体構造を原子レベルで明らかにする
構造生物学の進展
 - 実験的に決定される立体構造は、静的な構造であることがほとんど
 - しかし、生体分子が機能を果たすときには、動的であることが重要
 - 例：分子モーター(ATP等のエネルギーを使って構造変化することで力を出す)
 - 計算機中で、物理学法則に基づき、生体分子の動的構造をシミュレーションする
方法が有効
 - 特に、超並列化時代に突入して、計算対象が、方法開発に重きを置いたモデル
タンパク質から、生物学的に注目されるタンパク質に移行したことが重要な点
- 目的
 - 「京」で効率的に動作する全原子分子動力学(MD)ソフトウェアMARBLEを作製
 - 実用性検証のため、多剤排出トランスポーターAcrBの分子シミュレーションを行い、
その機能メカニズムを解明

ISLiMソフトウェア研究開発報告会 2

ISLiM


概要・アプローチ

- 対象モデル：
 - 水素まで含んだ全原子モデルを対象とする。
 - 周囲の水分子、イオンなどの溶媒は露わに扱い、誘電率一定などの近似は
用いない。
 - 膜タンパク質の計算時には、膜分子も原子レベルで扱う。
- シンプレクティック部分剛体時間積分法
 - 精度の高い全エネルギー保存
- 並列化アプローチ
 - 空間分割を用いて、通信量の削減を図る
- 相互作用計算
 - 近距離力： 空間分割により、近接ノード通信を用いる
 - 遠距離力： Particle Mesh Ewald法によりFFTを用いて計算
- 実験データによる検証
 - NMR, SAXS

ISLiMソフトウェア研究開発報告会 3

ISLiM

研究・開発のステップチャート





- 生体超分子用全原子
シミュレーション技術の開発
 - ✓ 超並列化ソフトウェア
MARBLE
 - ✓ 世界標準アルゴリズムを装
備 (PMIE等)
 - ✓ 先端アルゴリズムも装備
(シンプレクティック剛体法
等)
- シミュレーション結果の解析
技術の開発
 - ✓ 立体構造変化
✓ 立体構造ゆらぎ
 - ✓ 自由エネルギー
✓ 実験との対応
NMR, SAXSなど
- 多剤排出トランスポーター
AcrBの全原子分子動力学シ
ミュレーション
 - ✓ 分子スケールチームで共
通のターゲット
 - ✓ 膜タンパク質
 - ✓ 構造変化が機能に重要
✓ ゆらぎと構造変化

出力：座標の履歴

全原子MD
ソフトウェア

解析
ソフトウェア群



計算対象
生体分子

現在までの研究開発成果

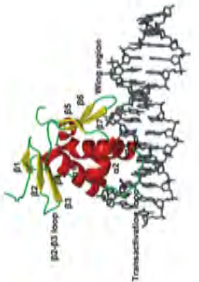
- 通信の最適化
 - 3Dトラーランスネットワークに合わせて、分割法や通信法を抜本的に見直した
 - 分割・通信には、Mid Point法を利用
 - ほぼすべてで、隣接通信を利用 (エネルギーの和、FFTを除く)。
- OpenMP対応
 - ノード内は共有メモリ利用、ノード間はMPI利用することで、MPIプロセスの削減、通信の効率化を目指した。
 - 短距離力、遠距離力とも対応。
- 短距離部分: カーネルを用いて、計算効率向上 (高度化チーム・富士通の助言)
- 最初: ピーク比性能: ~5%, SIMD 利用率: 0.02% □ ピーク比性能: ~29%, SIMD 61%
- 遠距離部分
 - 長距離力には3DFFTを利用、3DFFTの2次元分割
- 短距離部分と遠距離部分のオーバーラップ
- 3DFFTは通信が主なので、短距離部分の計算とオーバーラップをさせる
- 多剤排出トランスポーターActBでの検証

ISLiMソフトウェア研究開発報告会 5

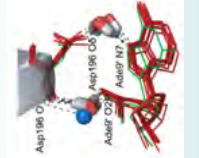



実験データによるMARBLEの検証例

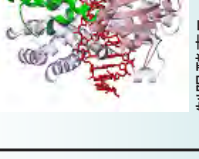
NMRでの検証



DNA結合タンパク質
PhoB



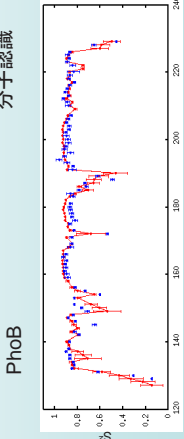
水を介した
分子認識



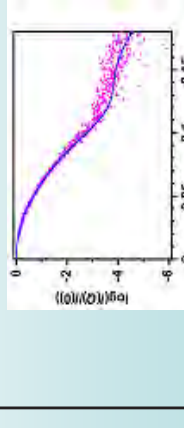
制限酵素EcoO109I

構造ゆらぎ方向(赤)と
構造変化方向(青)の一致

X線澄液散乱実験での検証





構造ダイナミクスの検証
(赤: 計算、青: 実験)

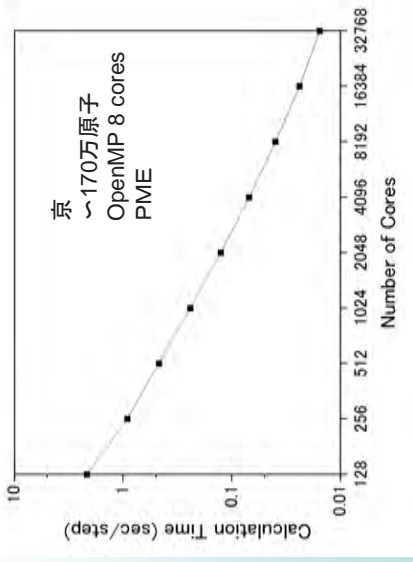


散乱プロファイル(青: 計算、ピンク: 実験)



(Biophys J, 2009)

PMEでの計算時間



ISLiMソフトウェア研究開発報告会 7

AcrBの全原子分子動力学シミュレーション

System

Lipid: POPE 809 molecules (101,125 atoms)

water: TIP3P model 108,385 molecules (325,155 atoms)

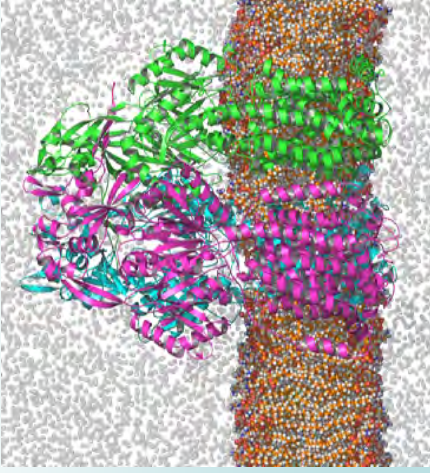
protein: (PDB ID: 2DHH) (No drug containing structure)

Missing part (499-512) was added.
(47,724 atoms)

Ion: Na⁺ (36 atoms)

Total: 474,040 atoms

Box size: 160Å X 160Å X 180Å



Program : MARBLE

Force field: CHARMM27

Electrostatic interaction: Non cut-off
(Particle-mesh Ewald)

Ensemble: NPT

Temperature: 300K,

Time step: 2fs

膜貫通部位のプロトン化状態の決定

プロトン濃度勾配のエネルギー利用には、

- ✓ 水分子がそれぞれの側からアクセス可能
- ✓ 同時には両方からアクセスは不可
- ✓ プロトン着脱可能な基(カルボキシル基等)を経由して、プロトンは透過

膜貫通部位で、プロトン着脱可能な基は、**Asp407, Asp408**のカルボキシル基
それぞれの構造でのプロトン化状態を決定

Binding

Extrusion

Access

Extrusion構造のLys940のコンフォメーションが異なる

ポータードメインの構造変化

挿入型のAsp407, Asp408が脱プロトン化した状態のシミュレーション(407-408-I)において、ポータードメインの構造が、取込型の構造へと近付いた。

ポータードメインの運動

PCA

● 407-408-Iの埋没100nsの平均
● 407-408-I100nsの変化

Buried surface area (Å²)

Time (ns)

407-408-Iの排出型は、407-408Hの排出型と比べて、Buried surface Areaが減少した(各ドメイン間が離れた)

ペリプラズム側から見て、ポータードメインの中心を軸とした、時計回りの運動が、構造変化の主要な成分である

排出型から取込型への構造変化のメカニズム

ポータードメイン

膜貫通ドメイン

取込型

排出型

水の入り口

薬物の入り口

プロトン転移サイト

ペリプラズム側へプロトンが移動
電荷:-1

ペリプラズム側からプロトンが移動
電荷:0

Asp408がプロトン化
電荷:-1

Asp407
電荷:-1

Asp408
電荷:-1

Asp407
電荷:-1

Asp408
電荷:-1

★ 挿入型から取込型への構造変化において、ポータードメインの中心を軸とした、スクリーモーニング(上面の赤矢印)による、PC2サブドメインの動きにより、薬物の入り口および、水の入り口が形成される。

★ これらの構造変化は、プロトン化状態の変化に伴う電荷の変化により生じる静電引力が駆動力となって生じていると考えられる。

タンパク質カノニカル分子軌道法 ProteinDF の京対応版プログラムの公開

佐藤 文俊

東京大学生産技術研究所
革新的シミュレーション研究センター 教授



発表者紹介

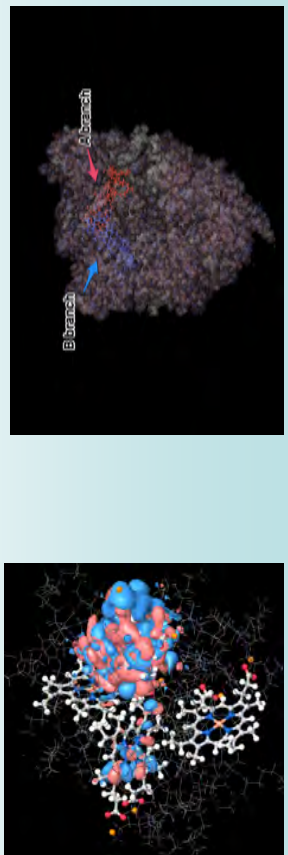
- 1990年 3月 慶應義塾大学大学院理工学研究科物理学専攻博士課程修了
- 1990年 4月 科学技術庁特別研究員
- 1991年 3月 九州工業大学情報工学部助手
- 2002年 11月 東京大学生産技術研究所客員助教授
- 2004年 12月 東京大学生産技術研究所助教授
- 2008年 4月 東京大学生産技術研究所教授

研究分野

計算生体分子科学、量子分子生物学

ISLiM 東京大学 東京大学大学院工学系研究科 革新的シミュレーション研究センター 教授 佐藤文俊

タンパク質カノニカル分子軌道法ProteinDFの京対応版プログラムの公開

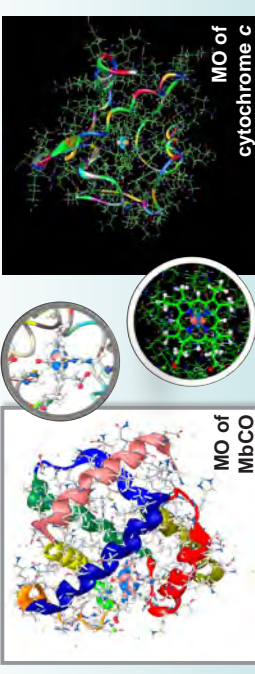


ISLiMソフトウェア研究開発報告会 1

ISLiM 東京大学 東京大学大学院工学系研究科 革新的シミュレーション研究センター 教授 佐藤文俊

背景・目的

背景
ProteinDFによるタンパク質の分子軌道計算



MO of MbCO
94.7%: Feのd軌道
99.9%: Heme+近位His Fe(II)で失活

MO of cytochrome c
85.2%: Feのd軌道
Fe(II)⇌Fe(III)で電子移動

目的
ProteinDFの京対応版プログラムの公開

ISLiMソフトウェア研究開発報告会 2

ISLiM 東京大学 東京大学大学院工学系研究科 革新的シミュレーション研究センター 教授 佐藤文俊

概要・アプローチ

ProteinDFとは

- タンパク質のカノニカル分子軌道を計算する超大規模量子化学計算プログラム
- C++でコーディング

主な機能と特徴

- HF, DFT(含ハイブリッド汎関数)
- ガウス型基底関数
- 分子積分にRI法やRT法アルゴリズムなどが選択可能
- 大規模SCF計算を成功させるための高精度初期値作成法
- エネルギー勾配計算



光合成反応中心タンパク質の分子軌道
・スペシャルペア
・アクセサリー-バクテリオクロロフィル
・バクテリオフィオフィチン
・Pから4Å以内のアミノ酸残基




ISLiMソフトウェア研究開発報告会 3

ISLiM 東京大学 東京大学大学院工学系研究科 革新的シミュレーション研究センター 教授 佐藤文俊

機能・性能がどこまで達成しているか

- 機能
 - 基本的な計算機能は「イノベーション基盤シミュレーションソフトウェアの研究開発」プロジェクトで開発(済み)
 - バージョン管理
 - プランチを作らずに、様々なハードウェアに対応
 - 本プロジェクトでは、「京」対応のために新しい動作形式を研究開発
- 性能
 - 分散並列計算機における数千ハイブリッド並列計算
 - 「京」での正常動作
 - MPI/OpenMPでのハイブリッド並列で動作確認
 - 1024x8並列まで動作確認
 - 「京」での演算性能向上・チューニング作業中
 - 分子積分・数値積分の単体性能の向上




ISLiMソフトウェア研究開発報告会 4

その際に克服すべき問題は何か

- 「京」で正常に動作させること
 - コンパイラの仕様に沿ったC++コードの記述を行う
 - 分散ローカルファイルシステムの対応
- 効率の良い大規模並列分子軌道計算アルゴリズムの開発
 - 理論上 $O(N^4)$ の分子積分アルゴリズムの計算量の削減
 - ロードバランスの均等化
 - 1つ1つの分子積分の演算量が、基底関数の組み合わせにより異なる
 - 総演算量を前もって見積もることが困難
 - 通信コストの削減
 - 分子軌道計算において、全通信は避けられない
- 巨大行列の取り扱い
 - 1,000残基規模タンパク質の軌道数 $\approx 100,000$
 - 行列1つに必要なメモリ量 $\approx 10^5 \times 10^5 \times 8 \text{ byte} = 80 \text{ GB}$
 - 行列演算に最低限必要なメモリ量 $\approx 80 \text{ GB} \times 3 = 240 \text{ GB}$
 - 最低限必要なメモリ量がノードに収まりきれない

ISLiMソフトウェア研究開発報告会 5

どのように解決したか (最適化など)

- 「京」で正常に動作させる対策
 - OpenMPコードの書き換え
 - 新しいコンパイラでOpenMP3.0仕様に対応
 - 分散ローカルファイルシステム用I/Oルーチンの構築
 - 単体性能向上
 - 富士通さんとチューニング → 継続中
- 大規模分子・大規模並列用分子積分アルゴリズムの開発
 - 高精度・高効率カットオフ法採用による計算コストの低減
 - シェルの組み合わせの並び替えによるタスク均等化
 - 動的ロードバランスの採用
- 巨大行列の取り扱い
 - 大域分散行列の採用
 - ScaLAPACKによる高速密行列演算
 - 大域分散行列用分子積分・数値積分アルゴリズムの実装

7 x 8 行列を2 x 2小行列に分割

Re(1,1)	Re(1,2)	Re(1,3)	Re(1,4)	Re(1,5)	Re(1,6)	Re(1,7)	Re(1,8)
Re(2,1)	Re(2,2)	Re(2,3)	Re(2,4)	Re(2,5)	Re(2,6)	Re(2,7)	Re(2,8)
Re(3,1)	Re(3,2)	Re(3,3)	Re(3,4)	Re(3,5)	Re(3,6)	Re(3,7)	Re(3,8)
Re(4,1)	Re(4,2)	Re(4,3)	Re(4,4)	Re(4,5)	Re(4,6)	Re(4,7)	Re(4,8)
Re(5,1)	Re(5,2)	Re(5,3)	Re(5,4)	Re(5,5)	Re(5,6)	Re(5,7)	Re(5,8)
Re(6,1)	Re(6,2)	Re(6,3)	Re(6,4)	Re(6,5)	Re(6,6)	Re(6,7)	Re(6,8)
Re(7,1)	Re(7,2)	Re(7,3)	Re(7,4)	Re(7,5)	Re(7,6)	Re(7,7)	Re(7,8)
Re(8,1)	Re(8,2)	Re(8,3)	Re(8,4)	Re(8,5)	Re(8,6)	Re(8,7)	Re(8,8)

node 2 node 3 node 4

ISLiMソフトウェア研究開発報告会 6








計算ターゲット

- Light-harvesting protein B-800/850
 - 18,207 atoms
 - 846 residues
 - 18 chains
 - ヘテロダイマー × 9量体
 - 27 BCLs
 - c.a. 90,000 orbitals




ISLiMソフトウェア研究開発報告会 8

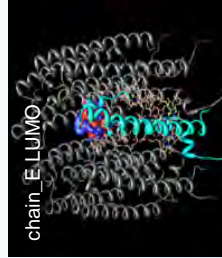




計算事例結果


- Light-harvesting protein B-800/850




chain_E_HOMO




chain_E_LUMO




chain_F_HOMO



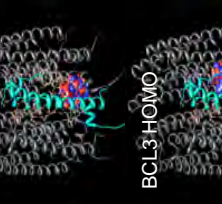
chain_F_LUMO



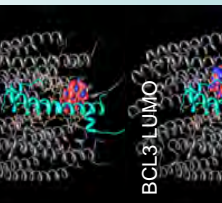
BCL1_HOMO




BCL1_LUMO



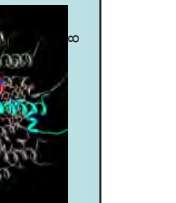
BCL2_HOMO



BCL2_LUMO



BCL3_HOMO

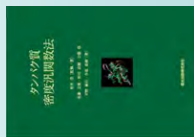


BCL3_LUMO

ISLiMソフトウェア研究開発報告会 8

参加者	所属	期間	担当
平野 敏行	東大生研	H19-	超々並列
西村 康幸	東大生研 → CEIソフト	H19-20	プレ・ポスト処理
上村 典子	東大生研 → RIST	H19-23	複雑計算収束
恒川 直樹	東大生研 → 就活中	H20-23	自由エネルギー・構造作成
松田 潤一	東大生研	H21-	超大規模用交換相関積分

- 京速コンピュータ「京」を利用いたしました。



粗視化分子シミュレーションソフトウェア CafeMol の開発と遺伝子動態研究

高田 彰二

京都大学大学院
理学研究科生物学専攻




発表者紹介

- 1990年3月 京都大学理学研究科化学専攻修士課程修了
- 1991年3月 岡崎国立共同研究機構分子科学研究所技官
- 1998年4月 神戸大学理学部化学科講師・助教授
- 2007年3月 京都大学理学研究科生物学専攻准教授（現在に至る）

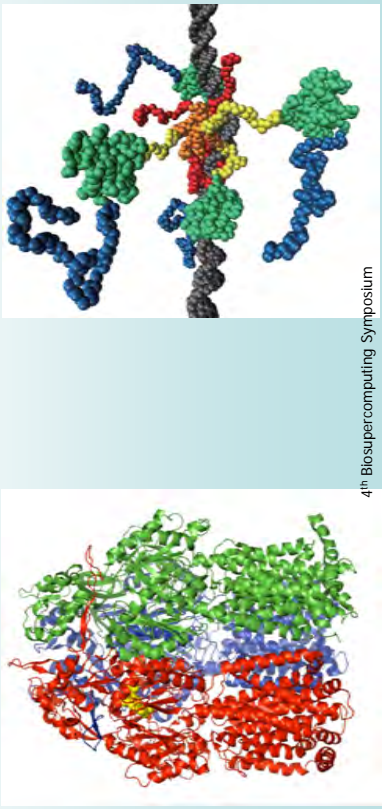
研究分野

理論生物物理学




Molecular machines and nuclear processes studied by coarse-grained molecular simulations


Shoji Takada
Dept Biophysics, Grad School of Science, Kyoto University



4th Biosupercomputing Symposium

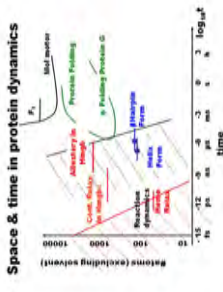
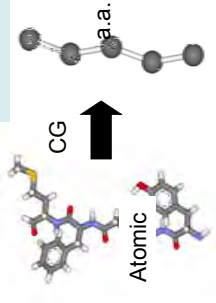


1




Background & Purpose


- Background
 - Living matter is hierarchic
 - In molecular scales, atomic structures made revolution in the field, but a gap between atomic resolution and cellular scale is still large. Large fluctuation makes the structural analysis harder.
 - For long-time simulations of huge systems, coarse-graining is useful.
- Purpose
 - On the basis of atomic resolution structural data and simulation technology, we
 - 1) systematically construct CG models,
 - 2) develop a software **CafeMol** that implements the CG models, and,
 - 3) apply them to various biological phenomena at molecular scale important in cellular biology.

4th Biosupercomputing Symposium




2




CafeMol

- Summary
 - By CG models, it simulate long-time dynamics of huge biomolecular systems.
- Approach
- Method
 - Classical molecular dynamics with CG model
- Simulations
 - Time propagation based on Langevin dynamics and others
- Parallel programming
 - Neighbor list, replica exchange
- Computer language and library
 - Fortran90, MPI, OpenMP
- Code release
 - Freely downloadable from <http://www.cafemol.org>
- Target size
 - Simulations that correspond to milliseconds and longer.

4th Biosupercomputing Symposium




3



CafeMol: Current status

- Current status of CafeMol
 - Completed
 - Implementation of CG models for protein, NA, lipid
 - Parallel computation up to 98000 cores on K
 - Single node efficiency up to 33% on K
 - To do
 - Accurate derivation of and more accuracy in CG models → Multiscale algorithm for CG parameter derivation from atomic level
- Applications
 - Multidrug transporter, kinesin, nucleosome, transcription factor (p53 etc)
 - Experimental verification
 - Multidrug transporter, collaboration with S. Murakami (Tokyo Inst Tech)
 - Kinesin, collaboration with M. Tomishige (Univ Tokyo)

4th Biosupercomputing Symposium



4

Atomic interaction based CG model

Deriving CG para from AA via a multiscale method

AICG1; Li, Wolynes, & Takada PNAS 2011
 AICG2; Li, Terakawa, Wang, & Takada PNAS 2012

AICG2 model

$$V = \sum_I k_b (r^I - r_0^I)^2 + \sum_J V_a^J(\theta_J) + \sum_I V_{dih}^I(\phi_I) + \sum_{J+2 \leq I \leq J+3} \epsilon_{loc}^I \exp\left(-\frac{(r^I - r_0^I)^2}{2w^2}\right) + \sum_{I>J+3} \epsilon_{nlb}^I [5(r_0^I / r^I)^{12} - 6(r_0^I / r^I)^{10}] + \sum_{I>J+3} \epsilon(C / r^I)^{12}$$

Flexible local potential (Statistical potential)

Energy decomposition

pairwise contact energies

Fluctuation matching

rmsf bet. AA and CG

Gohike et al., 2003.
 Chiu and Voith, 2006
 Li et al., 2010.



CafeMol, software for CGMD

Simulating "protein at work"

Also for nucleic acids, membrane, and their complex

Fortran 90, MPI & openMP

Running on K-computer

1-node efficiency 33%

Source available for ver 2.0 (proteins, DNA, & RNA), at www.cafemol.org

Protein: Structure-based, switching
 DNA: Structure-based+ P-Chem
 Membrane: P-Chem

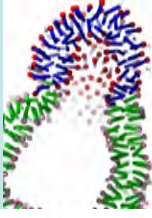
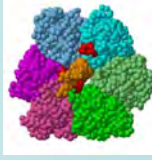
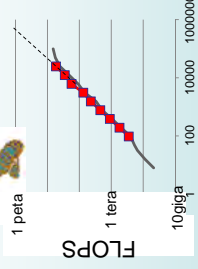
Parallel performance REMD @RICC

parallel 99.995%

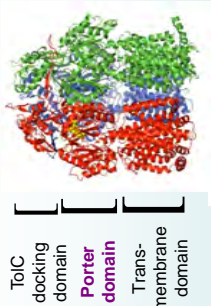
Protein-DNA complex



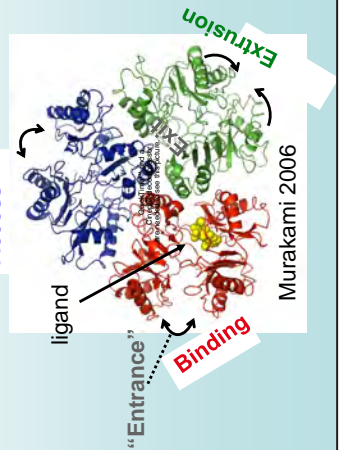
Kenzaki
 J. Comp. Theo. Chem. 2011



Application 1) Multidrug transporter



Access



Murakami 2006



Yao

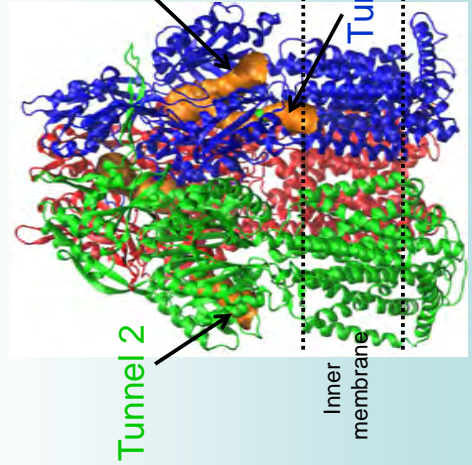


Yao et al, Nature Comm 2010

Drug uptake pathways

Tunnel 1, Long path from membrane surface

Tunnel 2, Short path from periplasm

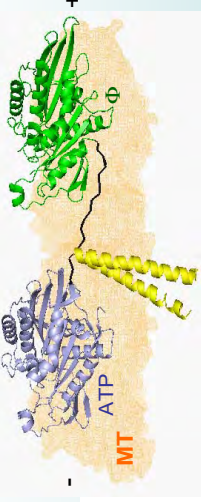
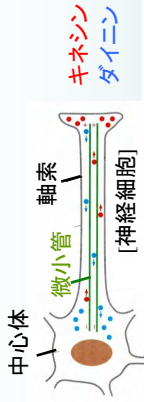


A B E

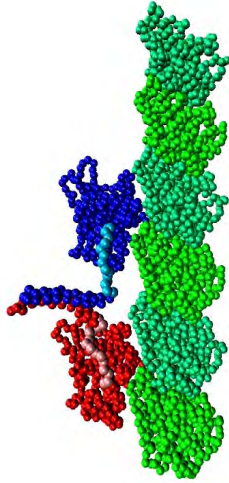
Application 2) Kinesin



Kanada



CG MD simulations together with high-resolution modeling of K-MT complex
In collaboration with experimentalists (Tomishige, Kikkawa)

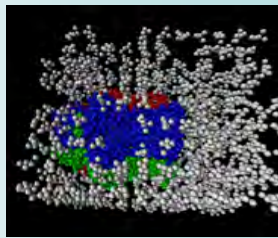
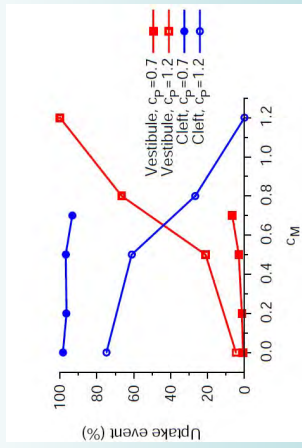


Dynamic simulation of drug uptake and export

- 273 ligand molecules in a 160x160x300 (Å) box
- 300K Langevin
- 100 trajectories
- Switching **ABE** → **BEA** → **EAB**

We found

drug-dependent uptake pathways

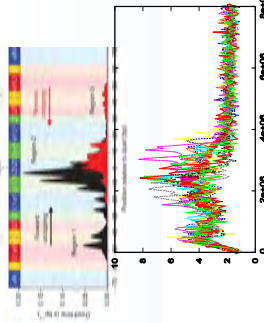


4th Biosupercomputing Symposium

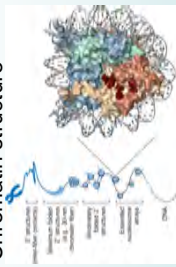
Application 3) Nucleosome dynamics



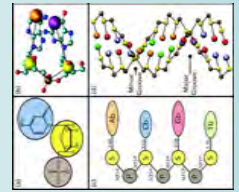
Kenzaki



Towards modeling of Chromatin structure

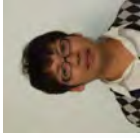


Assembly & disassembly of nucleosome



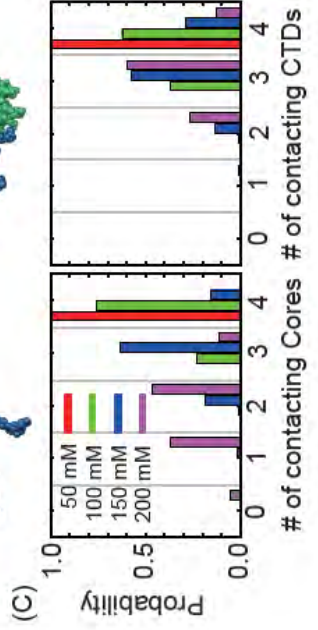
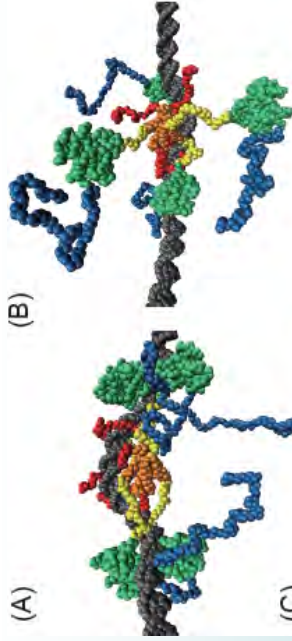
CG model of DNA de Pablo et al

Application 4) full-length p53



Terakawa

BPJ 2011
JACS 2012



酵素分子の柔らかさと触媒活性： ハイブリッド QM/MM 自由エネルギー法の開発

林 重彦

京都大学大学院理学研究科
分子スケール研究開発チーム 開発責任者



発表者紹介

- 1997 年 11 月 京都大学大学院理学研究科化学専攻博士課程修了
- 1998 年 4 月 日本学術振興会特別研究員 (名古屋大学大学院理学研究科)
- 2000 年 4 月 米国イリノイ大学アーバナシャンペーン校 博士研究員
- 2003 年 6 月 京都大学福井謙一記念研究センター 研究フェロー (博士研究員)
- 2003 年 10 月 科学技術振興機構 PRESTO さきがけ研究者
- 2005 年 5 月 京都大学大学院理学研究科化学専攻 准教授 現在に至る

研究分野

理論化学、生物物理学

ISLiM

酵素分子の柔らかさと触媒活性: ハイブリッド QM/MM 自由エネルギー法の開発

京都大学
大学院理学研究科
林 重彦

reactant TS product

L1 L2 L1 L2 L1 L2

ISLiMソフトウェア研究開発報告会

1

ISLiM

タンパク質の機能は酵素活性により 制御されている

F₁F₀ ATP synthase

ADP + P_i → ATP + ~10 kcal/mol

local

reversible!

motor ↔ **generator**

Noji et al.

Noji, Takeuchi, et al.

global

ISLiMソフトウェア研究開発報告会

2

ISLiM

機能に織り込まれている マルチスケールな現象

Causality of molecular motor

Chemical reaction

Conformational changes of subunits

Conformational changes of complexes

Slow dynamics!

Causality of synthesis

Quantum mechanics

Classical mechanics

of atoms involved

化学反応と遅いタンパク質ダイナミクスの相関?
(酵素の高い触媒活性の秘密?)

ISLiMソフトウェア研究開発報告会

3

ISLiM

酵素反応シミュレーションの方法論

- 1. Molecular dynamics (MD) simulations**
 - Large protein systems
 - Empirical force fields (MM)
 - No explicit description of electronic states
- 2. Molecular orbital or density functional theories (QM)**
 - Explicit description of electronic states
 - Huge computational cost
- 3. QM/MM simulations**
 - Reactive moieties are described by *ab initio* QM methods.
 - Protein environment is treated by MM.

SH and Ohmine (2000) (QM/MM-opt)

SH, Tajkhorshid, Schuiten (2003) (excited state QM/MM-MD)

ISLiMソフトウェア研究開発報告会

4

ISLiM QM と MM の計算時間の大きなギャップ

100 days with 1 fs / step = 8.64 Ms

Trajectory time	Computational time/ step
10 ps	864 s = 14.4 m
QM/MM	86 s
-MD	15 min / step
	Computational cost
	QM >>> MM
MM-MD	0.86 s
GPU	12 step / s
Anton	120 step / s
	120k step / s

A long-time direct QM/MM-MD is difficult.

ISLiMソフトウェア研究開発報告会

5

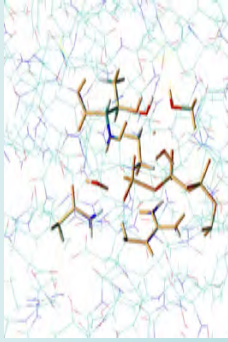
解決のための基本的アイデア

Complete separation between

- QM/MM calculation of a chemical reaction and
- MD simulation for protein conformational sampling

QM/MM free energy (FE) geometry optimization

Geometry optimization of the QM molecule on a **free energy surface** defined with the MM conformational thermal distribution obtained by MD simulations



Converged at 600 ns

ISLiMソフトウェア研究開発報告会

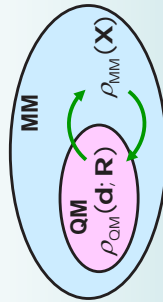
6

ISLiM QM/MM RWFE-SCF Method

Kosugi and SH, JCTC (2012)

Problem of MF-QM/MM

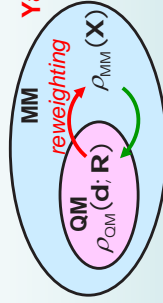
Conventional



Frequent iteration of QM/MM opt. and MD sampling is necessary.

Convergence problem!

Present method



MM distribution is reweighted.

Much quicker convergence of iterative MD sampling

$$\rho_{MM}(d, R; X) = \frac{\exp[-\beta \{E^{OM-MM}(d, R; X) - E^{OM-MM}(d_0, R_0; X)\}]}{\langle \exp[-\beta \{E^{OM-MM}(d, R; X) - E^{OM-MM}(d_0, R_0; X)\}] \rangle_0} \times \rho_{MM}(d_0, R_0; X)$$

ISLiMソフトウェア研究開発報告会

7

ISLiM QM/MM Long-Range Interaction

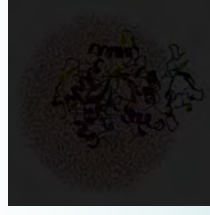
Ewald summation technique

- **Conventional:** hard to implement for continuum electron density
- **Present method: RESP operator treatment** (Hayashi and Ohmine, 2000) makes its implementation possible.

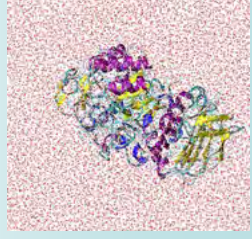
The method allows one to use directly MD samples obtained by sophisticated MD program packages using PME method.

Complete separation of QM/MM and MD sampling

Conventional: sphere cluster

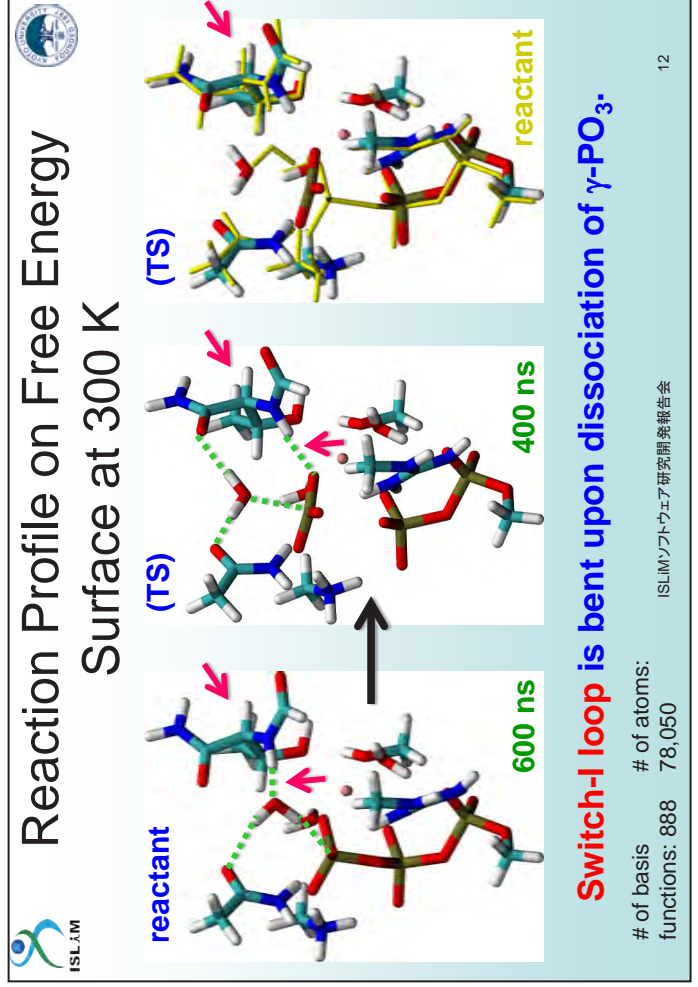
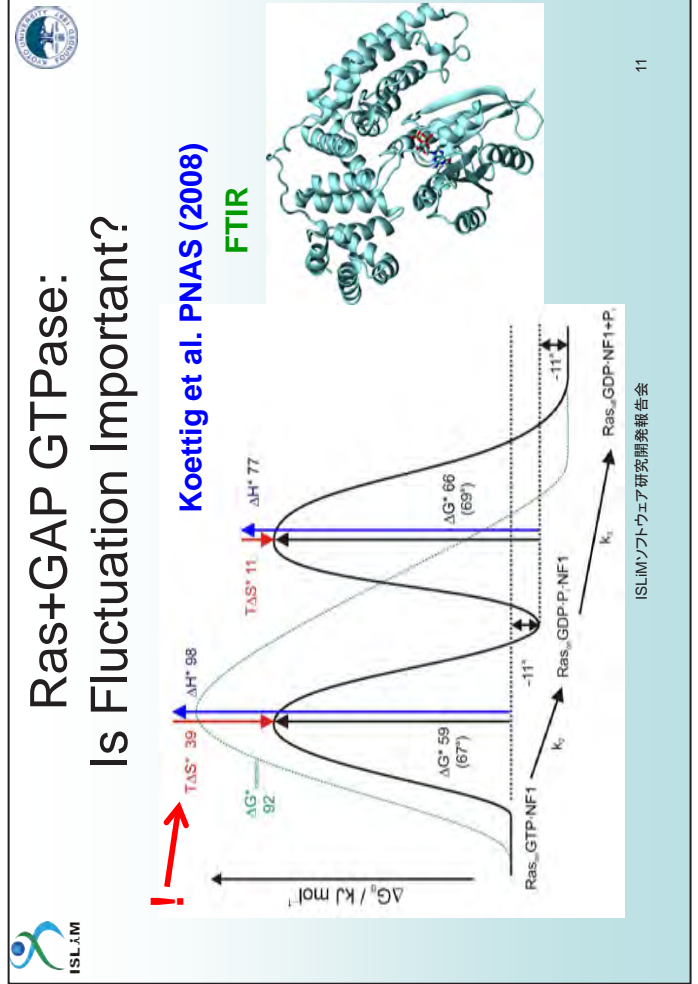
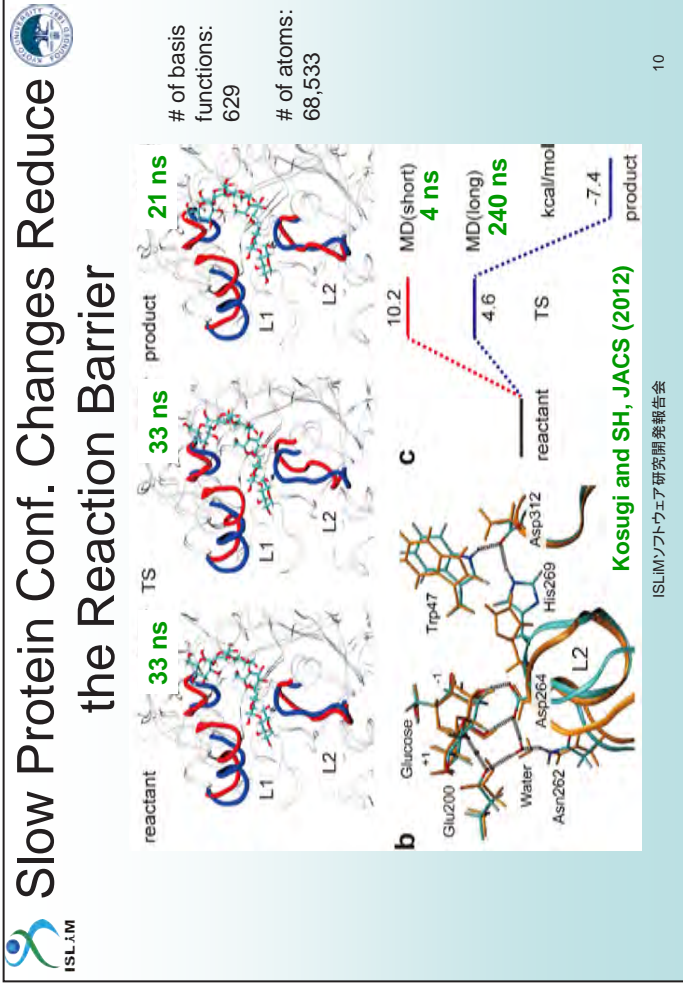
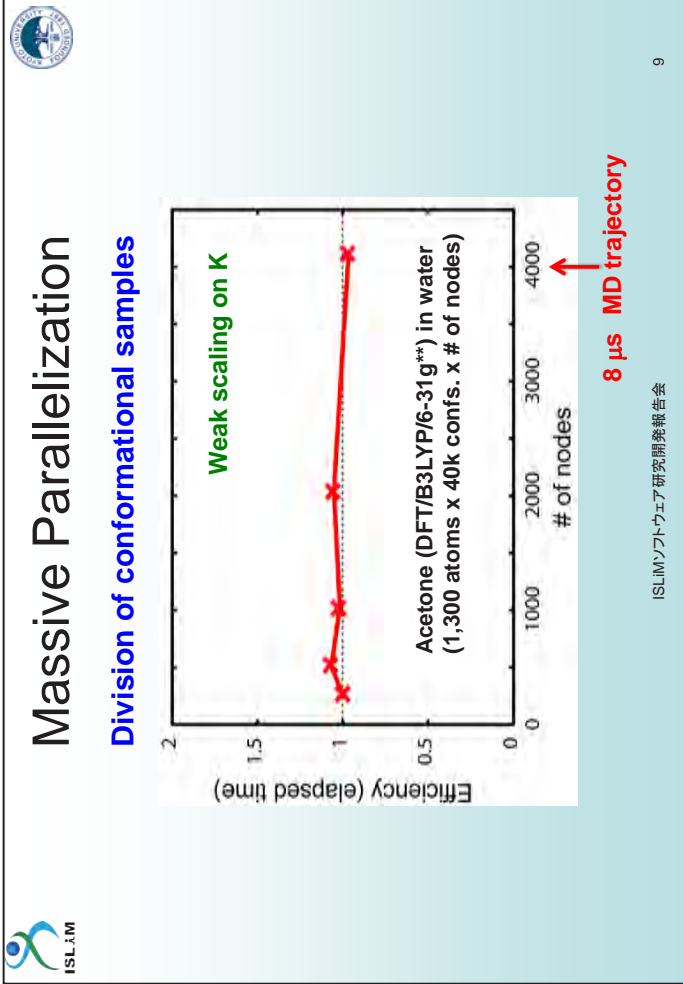


Present treatment: periodic boundary condition



ISLiMソフトウェア研究開発報告会

8

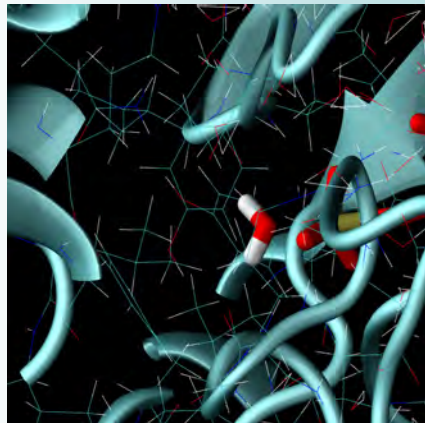
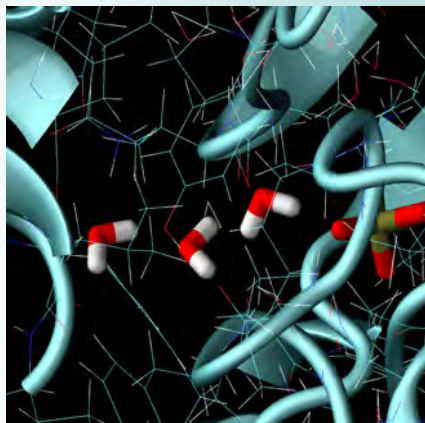




Water Molecules Are Squeezed Out Upon TS Formation

reactant

(TS)



ISLiMソフトウェア研究開発報告会

13



まとめ

- QM/MM 自由エネルギー法である QM/MM RWFE-SCF 法を開発した。
- MD シミュレーションによって得られる MM サンプルによる自由エネルギー面上での反応経路解析が可能となった。
- QM/MM 計算と MD サンプリングが完全に分離: MD シミュレーションにどんなプログラムを使っても良い (AMBER, CHARMM (クラスタ版) 分子力場使用可)。
- 「京」上で 4000 並列まで動作検証済み。現在、もう少し大きな系へのチューニングを実施中。
- タンパク質の特徴的な揺らぎは、酵素活性に重要 (かも?)

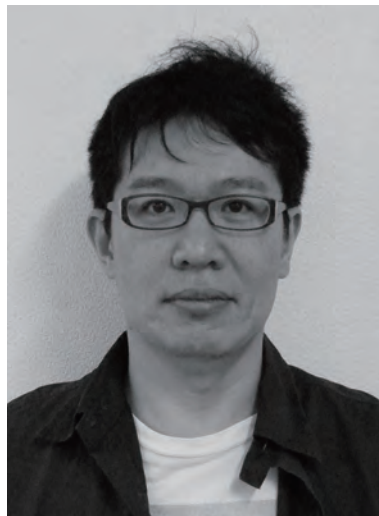
ISLiMソフトウェア研究開発報告会

14

Platypus-QM/MM の開発と応用

鷹野 優

次世代計算科学研究開発プログラム
分子スケール研究開発チーム



発表者紹介

2002 年 3 月 大阪大学大学院理学研究科化学専攻博士後期課程修了 博士 (理学)
2002 年 4 月 日本学術振興会海外特別研究員
2004 年 4 月 日本学術振興会特別研究員
2005 年 6 月 大阪大学蛋白質研究所 助手
2007 年 4 月 大阪大学蛋白質研究所 助教

研究分野

計算化学、量子化学、生物物理学、蛋白質計算科学






Platypus-QM/MMの開発と応用

大阪大学蛋白質研究所
附属蛋白質解析先端研究センター
助教 鷹野 優





ISLiMソフトウェア研究開発報告会 1

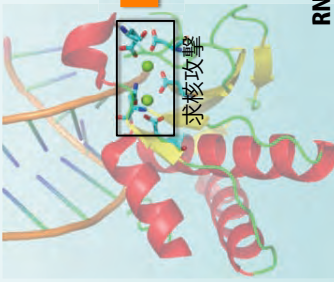
背景・目的

- 1) 生体高分子は巨大で、ヘテロで、ダイナミックな系
- 2) 生体高分子の機能の発揮は局所部位の電子状態に依存する

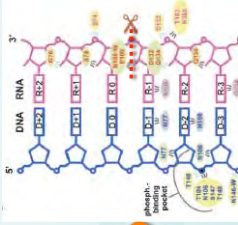
上記2つの特徴をもつ生体高分子をリアルなモデルによって扱い、高速に計算できる「新規QM/MD連成計算手法」を開発して機能を解明し、分子設計へ



発光





発色団の励起状態



DNA/RNA hybridのRNA部分の加水分解

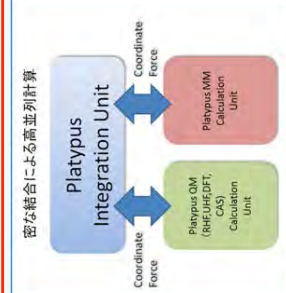
Matsuda et al. *Nature Methods* 2008, 5, 339. RNaseHI
Nowotny et al. *Cell* 2005, 121, 1005. 2

超並列 *ab initio* QM/MD連成計算プログラム Platypusの開発

(PLATform for dYnamics Protein Unified Simulation)

密な結合による高並列計算



$$E = E_{QM} + E_{MM} + E_{QM/MM}$$

Platypus-MM
並列性能に優れた新規
長距離ポテンシャルの開発
(zero-multipole法)

+

Chain-of-State法 (疎結合並列計算)

最小自由エネルギー経路！

ISLiMソフトウェア研究開発報告会 3




現在までの研究開発成果

- RHF, UHF, R-DFT, U-DFT, CIS, CIS(D), CIS-DFT, MP2, CASSCFの実装
- エネルギー・力計算における積分計算のハイブリッド並列化とSIMD化
- 2電子積分計算の原子軌道基底から分子軌道基底への変換部分の高度化
- CASSCF計算におけるdirect CI計算の高速化 (CAS(16,16)まで可能)

↑ QM部分の高速化・超並列化

- 新規の静電相互作用計算法の開発 (Zero-dipole法)
- 粗結合 (chain-of-state法)を利用した超並列化

↑ MM部分の高速化・超並列化

- 線形応答関数解析によるQMMM境界の判定

↑ QM/MM境界の高精度化

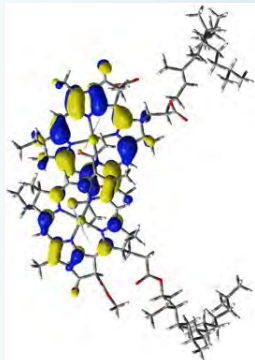
ISLiMソフトウェア研究開発報告会 4

「京」での並列性能(QM)

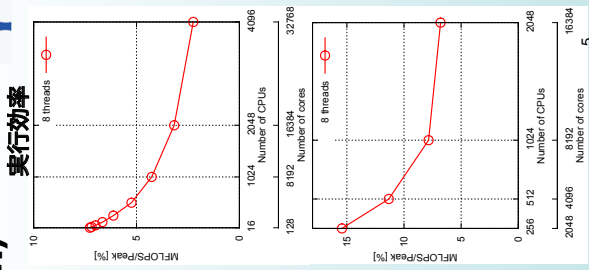
並列化率 99.9888%

RHF/cc-pVDZ
 総数32,768コア
 並列化率 **99.9888%**
 実行効率 7.27~2.24%
 (8192コア : 4.26%)

CASCI(16,16)/6-31G**
 総数16,384コア
 並列化率 **99.9728%**
 実行効率 15.43~6.80%
 (8192コア : 7.85%)



光合成反応中心
 スペシャルペア
 (280原子)
 +
 点電荷(30,904個)



ISLiMソフトウェア研究開発報告会

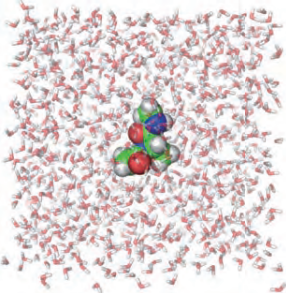
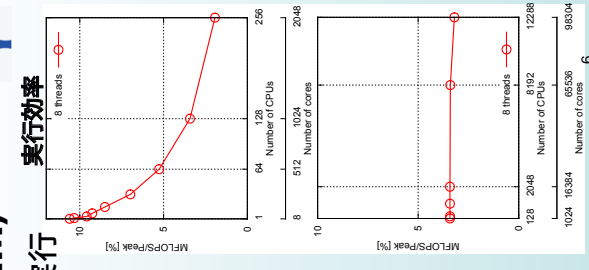
「京」での並列性能(QM/MM)

98,304コアのハイブリッド並列の実行

RHF/cc-pVDZ
 Ace-ALA-NME
 +
 1,080 水分子

Strong scaling
 総数2,048コア
 並列化率 99.7893%
 実行効率 10.60~1.94%

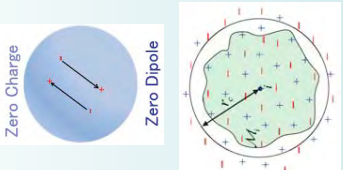
Weak scaling (Chain of State)
 総数98,304コア
 128プロセス x 768レプリカ
 実行効率 3.42~3.20%
Weak scaling効率 : 93.36%

ISLiMソフトウェア研究開発報告会

Zero-Dipole 法

電荷と dipole の双方がゼロとなる中性条件と静電遮蔽効果を考慮
 遠距離の静電相互作用の効果を近距離に繰り込む



Accuracy of Zero-Dipole method

$$E_{\text{Zero-Dipole}} = \sum_{\substack{i \in \text{MM} \\ j \in \text{QM}}} q_i q_j \left[\frac{\text{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\text{erfc}(\alpha r_c)}{r_c} + b \left(\frac{1}{r_{ij}^2} - \frac{1}{r_c^2} \right) \right]$$

$$- \left\{ \frac{3 \text{erfc}(\alpha r_c)}{2 r_c} + \frac{\alpha}{\sqrt{\pi}} \exp(-\alpha^2 r_c^2) \right\} \sum_{i \in \text{MM}} q_i^2 - \frac{\alpha}{\sqrt{\pi}} \sum_{i \in \text{MM}} q_i^2$$

$$b = \frac{\text{erfc}(\alpha r_c)}{2 r_c^3} + \frac{\alpha \exp(-\alpha^2 r_c^2)}{\sqrt{\pi} r_c^2}$$

従来の particle mesh Ewald 法で必要となる
 逆空間FFTの All-to-All 通信が不要

プログラムの超並列化が可能

Fukuda et al. J. Chem. Phys. 2011, 134, 164107, Fukuda et al. J. Chem. Phys. 2012, 137, 054314.

ISLiMソフトウェア研究開発報告会

QM/MM境界問題

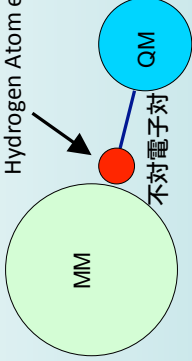
QM/MM境界に化学結合が存在する場合

QM空間の切断 → 不對電子対の発生 → QM (分子軌道) の歪み

正しいエネルギー及び力が得られない

1. どのように切るか?
 Link atom method
2. どこを切るか?
 線形応答関数解析

Hydrogen Atom etc...
 QM/MMモデル化に起因する
 QM領域の電子構造の誤差の最小化

$$\delta\rho(\mathbf{r}) = \int \frac{\delta\rho(\mathbf{r}')}{\delta v(\mathbf{r}')} d\mathbf{r}'$$


ISLiMソフトウェア研究開発報告会

線形応答関数解析

従来：QM/MMモデルによる誤差 $\delta v(\mathbf{r}')$ の最小化

QM/MMモデル由来のQM領域の誤差 $\delta\rho(\mathbf{r})$ の最小化

[ガイドライン]

- (i) 共有結合では電子トランスファーが大きい平衡核間距離ではなく、結合不安定が起こる(結合が切れかけた)状態で最大となる
- (ii) トランスファーブロックがない状況ではサイト数が多くなるにつれフリーデル振動型の振動・減衰となる。
- (iii) アラニンジペプチド, グルタミン酸, アラニントリペプチド系ではsp³接合点が伝播をブロックする。水素結合経由の伝播は約0.001(ex. 0.1 Hartreeの $\delta v(\mathbf{r})$ に対し0.0001の誤差伝播)。 α ヘリックス系では明白に誤差伝播が構造化される(右図)。

従来のsp³接合点での切断が有効

QM/MM境界問題の定量化が可能に!

ISLiMソフトウェア研究開発報告会

Ueda et al. *Int. J. Quantum Chem.* in press.

9

RNase HI

QM/MMを利用した疎結合(chain of state)超並列シミュレーションによる酵素反応機構研究

QM (B3LYP) + on the fly QM/MM chain of state

活性中心のみのモデル計算

約17 kcal/molの安定化

TS1 28.7, INT1 22.2, TS2 27.4, INT2 23.6, TS3 24.9, Product 0.2, Reactant 0.0

AG (kcal/mol)

Substrate, Water, Nucleophile, D132, D71, D192, E109, O4, O1, O2, O3, O4, O5, O6, O7, O8, O9, O10, O11, O12, O13, O14, O15, O16, O17, O18, O19, O20, O21, O22, O23, O24, O25, O26, O27, O28, O29, O30, O31, O32, O33, O34, O35, O36, O37, O38, O39, O40, O41, O42, O43, O44, O45, O46, O47, O48, O49, O50, O51, O52, O53, O54, O55, O56, O57, O58, O59, O60, O61, O62, O63, O64, O65, O66, O67, O68, O69, O70, O71, O72, O73, O74, O75, O76, O77, O78, O79, O80, O81, O82, O83, O84, O85, O86, O87, O88, O89, O90, O91, O92, O93, O94, O95, O96, O97, O98, O99, O100, O101, O102, O103, O104, O105, O106, O107, O108, O109, O110, O111, O112, O113, O114, O115, O116, O117, O118, O119, O120, O121, O122, O123, O124, O125, O126, O127, O128, O129, O130, O131, O132, O133, O134, O135, O136, O137, O138, O139, O140

RNAseH + DNA/RNA + 4786水分子 (16,872原子)

DNA/RNA hybridのRNA部分の加水分解の酵素反応機構 (最小自由エネルギー経路)

ISLiMソフトウェア研究開発報告会

10

発光蛋白質sirius

QM/MMによる励起状態ダイナミクス

水中

励起状態ダイナミクス

水中

励起エネルギー計算 CIS/6-31+G(d)

真空中 276.42 nm (実験値: 345 nm)

水中 347.64 nm

protein + 449水分子 (1,374原子)

protein + 7365水分子 (23,900原子)

siriusの発光機構の解明

siriusの発光強度の向上

ISLiMソフトウェア研究開発報告会

11

共同研究者

中村 春木 (阪大蛋白研)	→	現・近畿大先端研
米澤 康滋 (阪大蛋白研)	→	現・阪大院理
山中 秀介 (阪大蛋白研)	→	現・北里大薬
下山 紘充 (阪大蛋白研)	→	現・阪大蛋白研
福田 育夫 (理研)	→	現・阪大蛋白研
山崎 秀樹 (阪大蛋白研)	→	
大島 勘二 (阪大蛋白研)	→	
中田 一人 (NEC・阪大蛋白研)	→	

謝辞

本発表の一部は、京速コンピュータ「京」の試験利用および一般利用によるものです。

ISLiMソフトウェア研究開発報告会

12

細胞スケール研究開発チーム成果報告

細胞スケール研究開発チーム 成果統括報告

横田 秀夫

次世代計算科学研究開発プログラム
細胞スケール研究開発チーム チームリーダー



発表者紹介

- 1993年 3月 日本大学大学院農学研究科畜産学専攻修士課程修了
- 1999年 3月 東京大学大学院工学研究科精密工学専攻 博士授与
- 1999年 4月 理化学研究所素形材工学研究室協力研究員
- 2003年 10月 理化学研究所 VCAT 開発チームリーダー
- 2006年 10月 理化学研究所細胞スケール研究開発チームリーダー兼務 (現在に至る)
- 2011年 10月 理化学研究所基幹研究所生物情報基盤構築チームヘッド (現在に至る)
- 2012年 4月 理化学研究所計測情報処理研究チーム副チームリーダー (現在に至る)
- 北海道大学大学院 工学研究科 客員教授
- 神戸大学 大学院工学研究科 システム情報学研究科 客員教授
- 東海大学 総合医学研究所 循環器内科学 客員教授
- 東京農工大学大学院 工学府 客員教授

研究分野

- バイオイメーjing
- 細胞内画像処理
- バイオメディカルシミュレーション

細胞スケール研究開発チーム成果統括報告

横田 秀夫

次世代計算科学研究開発プログラム
細胞スケール研究開発チーム チームリーダー

1. 目的

細胞や臓器の部位特異的な生物学的情報を系統的に実装させて生命現象の予測や実験仮説構築に利用できるモデル開発に必要なソフトである細胞シミュレーションプラットフォームを完成させ、これを超高速計算機で稼働させることにより、細胞や臓器単位の機能的形態学的応答の再現や病態メカニズムの理解や制御法の開発に資するシミュレータを開発することを目的とする。細胞生物学の目的は細胞内の現象を理解し説明することにあるが、細胞内の現象は未だ理解されていない現象が多く残されており、定量的な解析が可能な現象はごくわずかである。一方、シミュレーションはある規則に則り、数値解析して現象を再現することに他ならず、現象の理解、数理モデルの構築無しに細胞のシミュレーションの実現はあり得ない。この圧倒的に情報が少ない『細胞』を対象としたシミュレーションに於いては、計算科学としてのアプリケーションソフトウェアの開発だけでは、意味のある細胞現象のシミュレータを構築する事は不可能である。そこで、シミュレータの開発と同時にシミュレーションのための実験を行う事により、新たな細胞シミュレータの開発を目的とした。さらに、近年進展が著しい各種イメージング技術により取得した細胞内における物質の局在の定量データを用いることにより、細胞シミュレーションの解析結果の検証を行い、実際の細胞内で生じている現象を再現出来るシミュレータの開発を目指す事とした。

これまでの細胞シミュレーションでは、細胞内生じている現象のモデル化と個別の現象のシミュレーションが行われてきた。これらの研究成果により、細胞の膜機能(物質の透過)、物質の拡散、生化学反応について、その数理モデルの構築が実施されてきた。しかしながら、従来の研究では、細胞内や膜の場所を考慮することは不可能であった。一方、細胞の中ではさまざまな現象が複雑に影響し生命が成立している。細胞の中では、物質の濃度勾配やオルガネラ等によってその機能の役割分担がされていると共に、臓器内部における環境によって同じ細胞で

も挙動が異なっている。疾患時においては、細胞の応答やその周囲の環境の異常を再現することが必要であり、これらの現象の再現のためには、複数の事象と空間(場)を考慮したシミュレーションを連成して解析することが必要である。我々は、正常時の細胞の挙動や異常な挙動である疾患の再現、さらにはその治療法に関するシミュレータの開発を目指して研究を推進している。なお、具体的課題として、ターゲットを赤血球と肝臓細胞、血小板、膵細胞に限定し、細胞内外の代謝、膜透過、物質の拡散の現象を連成して解析することを目的とした。また、これらの対象は、細胞内のオルガネラを考慮した細胞内現象のシミュレーションと複数細胞計算が必要なことに加えて、将来的な計算機能力の向上をにらんだ臓器や全身レベルでのシミュレーションの開発を目指している。

1. ソフト研究開発の概要・アプローチ

2.1 細胞シミュレーション統合プラットフォーム(RICS)の概要

細胞シミュレーション統合プラットフォーム(Riken Integrated Cell Simulator: RICS)は、細胞内の場や複数のシミュレーションの連成を考慮した共通基盤である。固定格子の空間内に細胞の形状モデルや反応モデルを構築し、様々な細胞反応のシミュレーションを可能にする。具体的には細胞を100万のボクセルに区画し、区画したボクセルに実測データより得られた細胞内の物質・移動量などの情報を取り込み、代謝のパスウェイを設定することで細胞内現象をシミュレーションする。

2.2 RICSのアプローチ

RICSは、細胞内の生化学反応と物質拡散の連成解析、物質の膜透過、膜電位計算、移流拡散を弱連成の形式で連成解析する。RICSで計算可能な形状を作成するには2つの方法があり、一つはCADを用いてポリゴンの細胞・オルガネラ形状を作成し、RICSプリシステムを使用してボ

クセル化する方法、もう一つは実際の顕微鏡画像から得られる 3 次元断層画像を再構築し、オルガネラ等の領域のラベル化を行った後、ボクセル化する方法である。CAD と顕微鏡画像を組み合わせて形状を作成することも可能で、特殊な状況を作り出し、シミュレーションすることが可能である。細胞の反応にはこれまでの細胞シミュレーションにおいて実績のある E-Cell3 システムを使用し既存の生化学反応モデルを使用できるだけでなく、SBML2.0 などの他の言語で書かれたものも既存のコンバータを介して利用することを可能とした。RICS は大規模並列計算機による実行を考慮して作成されており、PC クラスタから、「京」を含んだ多くの Linux ベースの高速計算機で実行可能である。これまでに 8192 並列までのスケールビリティを確認しており、現在「京」での超高並列計算に向けたチューニングを進めている。

2. 現時点でのソフト研究開発成果

3.1 RICS の開発状況

作成したシステムを用いて実際の顕微鏡データから得た細胞形状を用い、内部でのカルシウムイオン(Ca^{2+})の移動、反応、膜輸送担体をシミュレーションした。細胞形状はヒト肝臓由来細胞である HepG2 細胞株を用い共焦点レーザー顕微鏡を使用して取得した細胞形状および核とミトコンドリアの形状からなる。顕微鏡画像から 3 次元 Volume データを作成し、細胞内の反応として 24 種類の物質と、11 種類の生化学反応を設定した (図 1)。

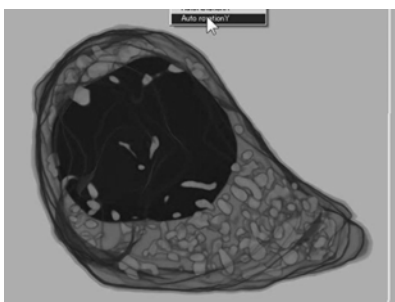


図 1: 共焦点レーザー顕微鏡から RICS プリシステムを使用して再構築したヒト肝臓細胞の形状

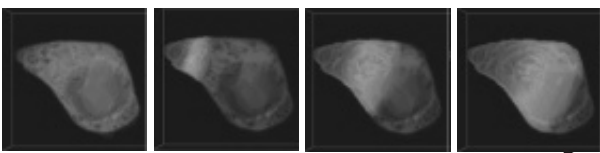


図 2: RICS を用いて計算した細胞内 Ca^{2+} の移動

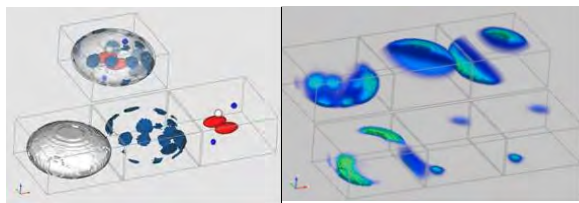
Ca^{2+} だけを通過させるチャネルを細胞膜の一部に局在化させて Ca^{2+} を流入させて解析した (図 2)。細胞内の Ca^{2+} の緩衝反応を設定した場合は、生化学反応が無い場合に比べ細胞内 Ca^{2+} のみかけの拡散速度の低下がみられ、実際の細胞内に近い Ca^{2+} の動態が再現された。

3.2 肝細胞シミュレーション

肝細胞の解糖系、糖新生、グリコーゲン生成と消費、フルクトース代謝、尿素回路、ミトコンドリアクエン酸回路、酸化リン酸化、脂肪酸合成と分解、アルコール代謝、グルタミン/グルタミン酸代謝、メチオニン代謝、プリン核酸代謝など、糖とアミノ酸を中心とした代謝経路を調査し、個々の酵素反応速度に基づいて大規模な代謝モデルを構築した。実証実験により、さまざまな環境下での代謝物の定量計測を行い、その数値情報を基に代謝モデルを構築した。現在、485 変数、582 の反応からなる詳細な肝細胞の代謝モデルを構築することに成功した。肝手術における虚血とその後の再灌流では、クランプと再灌流による代謝機能低下が重篤な不良を起さない限度として、15 分というクランプ時間が慣習的に使用されている。そこで、細胞外の酸素濃度と栄養としての代謝物質濃度、細胞膜を介した膜輸送が起こる代謝物質濃度について一時的に 15 分間 50%、10%、1% に低下させる摂動を設定し、代謝変動を計算した。また肝臓の低酸素・低グルコース応答に必要な代謝システムから、糖代謝、ミトコンドリア機能を実装すると共に、血管の上流、下流の細胞モデルを構築して、酸素やエネルギー代謝の heterogeneity を勘案した類洞モデルをほぼ完成させた。門脈型 (P 型) と中心静脈型 (C 型) の細胞シミュレーションを連結することにより、虚血および再灌流に対する主要代謝物の濃度変動が顕著に buffering されて robustness が獲得できるようになることが明らかになった。さらに、細胞レベルの精緻さを保持した臓器代謝シミュレーションの妥当性を検証するために先端的バイオイメージングを利用した肝細胞・類洞・肝小葉レベルの代謝情報収集を研究室で独自に開発した 2 光子レーザー顕微鏡を用いて実施し、肝細胞素の画像情報を収集し、RICS によるモデル構築に利活用した。また肝臓微小循環系の上流と下流にあたる門脈終末枝と中心静脈における血流速度計測と酸素分圧計測を大腸がん肝臓転移モデルにおいて実施し、門脈域に形成された微小転移病巣の大きさと酸素消費の間の相関関係を明らかにした。

3.3 血小板シミュレーション

細胞シミュレータープラットフォームに、実証実験により取得した血小板細胞内のミトコンドリア、ゴルジ体、小胞体、濃染顆粒、血小板膜特異的受容体蛋白などの局在情報を取り入れた血小板細胞基盤モデルを作成した。刺激を受け活性化した血小板にて起こる濃染顆粒の放出機構の再現を目指した。作成したモデルにより、直径 $2\ \mu\text{m}$ の血小板細胞表面の数十 nm^2 程度の受容体刺激の血小板細胞局所の細胞応答（細胞内シグナル産生）の細胞内拡散による血小板細胞全体の活性化過程（ GPIIb/IIIa の活性化構造転化）と、活性化の途中にて濃染顆粒の放出が起こり、放出された ADP による P2Y_{12} ADP 受容体刺激が加算された時の血小板細胞全体の活性化過程の差異を見いだした。このことにより、放出反応を介する活性化刺激伝達促進効果の論理的予測が可能となった。血管壁の損傷部位における、止血に向かう血小板細胞の positive feedback 機構を論理的、定量的に予測、説明する初めてのモデルとなった。



3D血小板モデル 血小板内物質の局在

3.4 小胞動態モデル

全反射蛍光顕微鏡による膵臓 β 細胞のインスリン顆粒動態の画像解析から、グルコース刺激によるインスリン分泌では、restless newcomer が主様式であることが判明した。その動態を元にインスリン顆粒が細胞膜に融合する過程をモデル化した。1) 細胞内部では顆粒がゴルジ体から分離し、細胞膜に接近するまでは細胞骨格に沿って移動する。2) 細胞膜付近ではブラウン運動に近い挙動を示し、細胞膜に融合した後、インスリンを分泌する。これらの顆粒動態の違いから、細胞内を外部層と内部層の2つに分割してモデルを構築した。構築したモデルは、全反射顕微鏡 (TIRF) による分泌とほぼ同様の傾向を示した。現在、このモデルを RICS に搭載することで小胞の3次元解析を計画している。

3. ソフト研究開発の達成目標



基本アプリケーションである RICS の機能開発は終了し、現在「京」での動作確認と計算性能向上を図っており、平成 24 年度には、当初目的の計算を実現できる見込みである。さらに、開

発したアプリケーションを用いて、生命科学の実問題を計算することに注力している。具体的な課題は、上記に述べた肝臓、血小板、膵臓の細胞内外の物質の反応と挙動である。これらのモデル化を完成すると共に「京」での大規模計算に適した計算モデルに改変する予定である。この作成したモデルにより、臓器内における細胞の役割、機能分担を再現することが出来ると共に、疾患時における細胞内の状況、薬物投与時の反応を再現することが可能となり、治療や薬物開発に貢献出来ると考える。さらに、血小板や膵臓 β 細胞、腎臓細胞、脳神経細胞など局所的な反応が全体に強い影響を及ぼす計算を行いたいと考えている。

開発したソフトウェアとモデルは、プログラム終了までに無償公開を計画している。現状の大規模計算システムでは、情報科学の研究者が生命現象を解析しているが、研究の裾野を広げるためには、生命科学の研究者が「京」を使いこなして細胞のシミュレーションを実施する環境を整える必要がある。そこで、細胞スケール研究開発チームとして、RICS の利用範囲を拡大することを目的に周辺アプリケーションの開発を進めている。具体的には、シミュレーションの形状モデルの構築、物質分布の配置、オルガネラの場所と反応式の対応などのシミュレーションセッティングを設定するプリポストシステムを開発している。このシステムは、関連する外部のアプリケーションである、VCAT (形状処理)、E-CELL IDE (反応式)、V-Sphere::CBC (流体解析)、V-Isio (可視化) を連結してシミュレーションのモデル構築を GUI のみで実現するものである。これらのアプリケーション開発と共に、細胞シミュレーションモデルの整備と公開に加えて、RICS ユーザー会を設立して、細胞シミュレーションに関連する研究者コミュニティ作りを実施する予定である。プロジェクト終了後には、開発し公開する基本細胞シミュレータを元に、個別の臓器に分化したシミュレーションに関する研究が開花することを期待する。

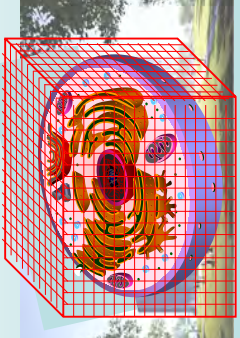
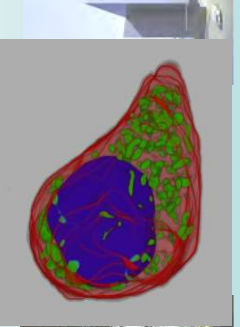
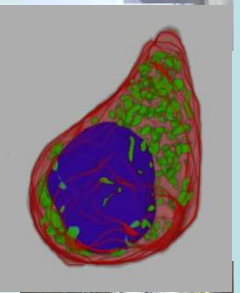
5. 謝辞

本発表の結果は、理化学研究所が実施している京速コンピュータの試験利用と理化学研究所情報基盤センター (RICC) RIKEN Integrated Cluster of Clusters の計算による。





Research activity of Cell Scale Simulation Team

Hideo YOKOTA
RIKEN






ISLiMソフトウェア研究開発報告会





RICS RIKEN Integrated Cell Simulation Platform

nm





mm

- Biochemical reaction (Metabolism)
 - E-cell
- Diffusion (Passive transport)
 - concentration gradient
- Membrane transport (Active transport)
 - Membrane transport
- Membrane Functions (Selective permeability)
 - ion channels, transporters, receptor, et al.
- Blood flow (Computational Fluid dynamics)
- Membrane potential (Hodgkin-Huxley)

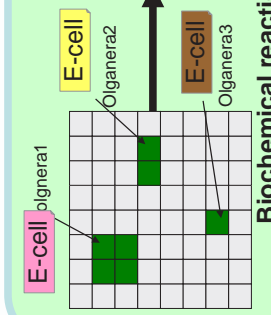
3

- Simulation of cells cluster
- Integrated simulation above phenomenon.

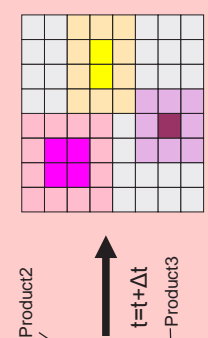



Coupled simulation in RICS

Biochemical reaction



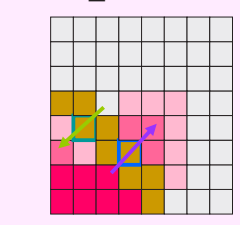
Diffusion



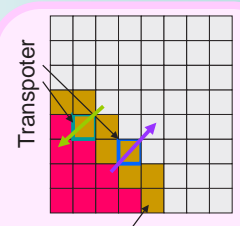
Product1 → Product2 → Product3

$t = t + \Delta t$



Membrane



Transport

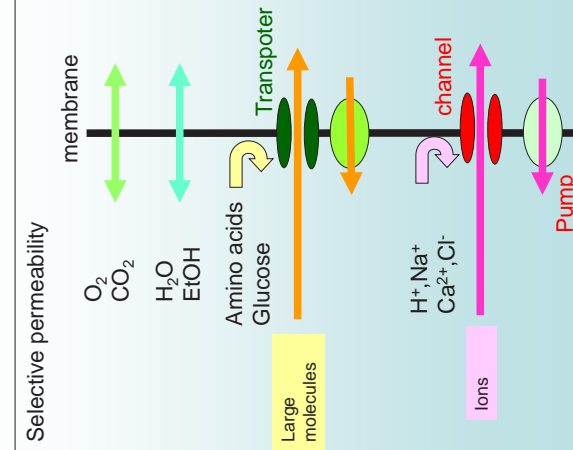


$t = t + \Delta t$

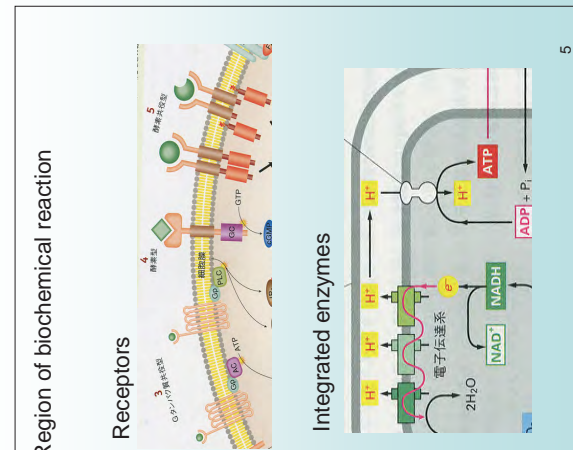





Functions of membrane

Selective permeability



Region of biochemical reaction



Flux of membrane transporters

$$\frac{\partial \phi}{\partial t} = \nabla \cdot (C \nabla \phi) - \nabla \cdot f_m + [\text{BioChemical Reaction}]$$



$$f_m = \sum_{Func} SW_{Func} f_{m, Func}$$

Func = {penetrate, channels, pumps, transporters, gap junctions}

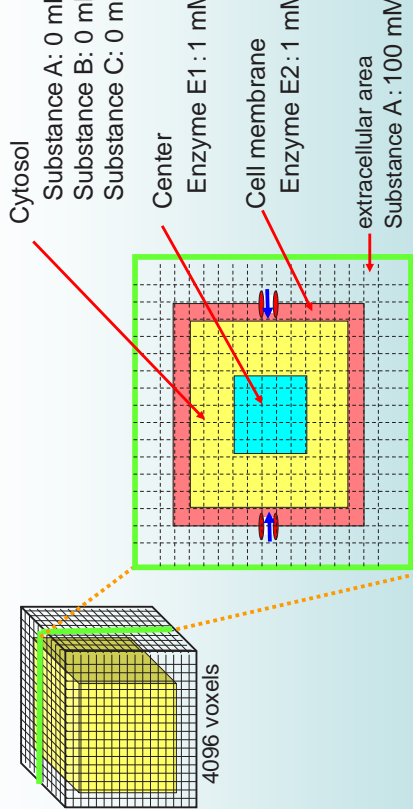
SW = regulatory function of each transporter

Transport of the substance through the membrane is the sum of the flow velocity for each transporter.

6

Simulation of membrane transport with bioreaction



Cytosol
Substance A: 0 mM
Substance B: 0 mM
Substance C: 0 mM

Center
Enzyme E1: 1 mM

Cell membrane
Enzyme E2: 1 mM

extracellular area
Substance A: 100 mM



4096 voxels

A $\xrightarrow{Ez1}$ B $\xrightarrow{Ez2}$ C

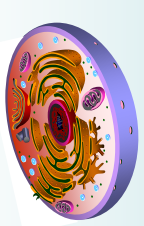
$V_{max} = 1.0$ $V_{max} = 2.0$

$K_m = 0.5$ $K_m = 1.2$

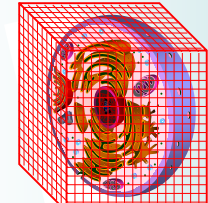
7

Section of a cell in voxels

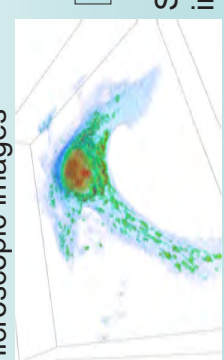


Virtual cell shape (CAD, CG)




RICS cell model

↑

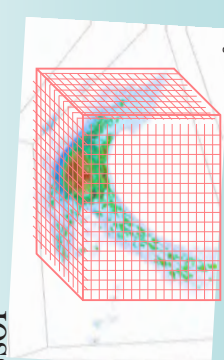


3D rebuilding Cell data from microscopic images

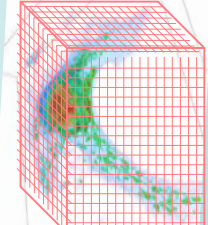


RICS Preprocessor

↑





Sectional images

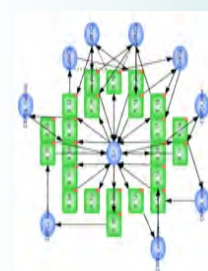


RICS cell model

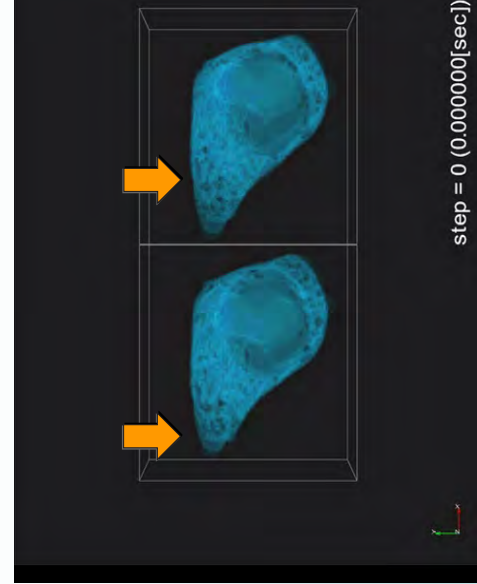
©RIKEN: Live Cell Modeling Project

Ca²⁺ kinetics simulation (HepG2)



Ca²⁺reaction model
reaction: 24
material: 10



diffusion+reaction
(apparent diffusion rate 28μm/s)

step = 0 (0.000000[sec])

Only diffusion
(diffusion rate 32μm/s)

9

Liver cell simulation

The diagram illustrates the multi-scale simulation of liver cells. It starts with a whole liver, zooms into hepatic lobules, then hepatic cords, and finally shows microscopic images of the real shape and simulated cells. Labels include: Perisinusoid, Capillary, Bile duct, Portal vein, Hepatic artery, Central vein, and Microscopic image.

Summary

- RICS (Riken Integrated Cell Simulation Platform)
 - : Spatiotemporal simulation of the cell
- Couple simulations of multiple phenomena
 - : biochemical reactions, diffusion, transport, membrane function, membrane potential, and advection
- Real cell shape simulation
 - : verify simulation results against actual observations
- 京(kei) available
 - : large-scale parallel more than 24,000nodes

Open source release : December 2012

Acknowledgement

Keio Univ.	RIKEN
Ayako Yachie	Yasuhiro Sunaga
Makoto Suematsu	Shin Yoshizawa
Tokai Univ.	Shigeo Noda
Yohei Nagasawa	Takashi Koyama
Shinya Goto	Yuki Tsujimura
Osaka Univ.	Koichi Takahashi
Shingo Murakami	Tajji Adachi
Hiroshi Hibino	
Yoshihisa Kurachi	
Kobe Univ.	
Tadao Shibasaki	
Susumu Seino	

血小板細胞シミュレーターの開発と血栓シミュレーターへの応用

後藤信哉

東海大学医学部内科学系循環器内科学 教授
東海大学バイオ研究医療センター代謝疾患研究センター長



発表者紹介

- 1986年3月 慶應義塾大学医学部卒業
- 1990年3月 慶應義塾大学大学院博士課程内科学専攻単位取得退学
- 1992年3月 博士（医学：慶應義塾大学）
- 1992年7月 スクリプス研究所分子実験医学部門博士研究員
- 1996年4月 東海大学医学部内科学I教室助手
- 1998年4月 東海大学医学部循環器内科講師
- 2002年4月 東海大学医学部内科学系循環器内科部門助教授
- 2007年4月 東海大学医学部内科学系（循環器内科学）教授
- 2008年4月 東海大学大学院バイオ研究医療センター代謝疾患研究センター長兼任
東海大学総合医学研究所代謝システム医学部門長兼任

研究分野

循環器内科学、血小板生物学、システム医学、薬剤開発




血小板細胞シミュレーターの開発と血栓シミュレーターへの応用

学校法人東海大学
医学部内科学系循環器内科学 教授
バイオ研究医療センター代謝疾患研究センター センター長
後藤信哉



細胞スケール



臓器全身スケール







血小板細胞代謝シグナルモデル



血小板細胞接着活性化モデル



ISLIMソフトウェア研究開発報告会 1

背景・目的

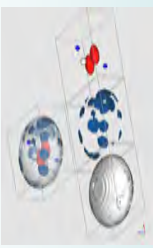
- 背景
 - 「**血小板細胞**」は、分裂、増殖せず、生理機能が止血、血栓形成に特化している。血小板細胞の活性化は、細胞膜上の受容体刺激開始後数分の時間経過にて完結する。シミュレーションすべき生理機能は単純、かつモデル化が容易である。直径-5 μmの血小板細胞数個を対象としたマイクロメートルスケールの実証実験により精緻化できる。
 - しかし、「**血小板細胞**」は心筋梗塞、脳梗塞などの動脈硬化血栓性疾患の発症において 重要な役割を演じる。アスピリン、クロピドグレルなどのナノメートルスケールの分子としての抗血小板薬により心筋梗塞再発予防効果は臨床的に証明されている。
 - 「**血小板細胞**」は血球細胞として、血管内を常に全身循環している。血管壁損傷部位への接着、集積は数cm以上を対象とする臓器全身スケールのイベントである。
- 目的
 - マイクロメートルスケールの「**血小板細胞**」に注目し、血小板細胞を構成するナノメートルスケールの分子の細胞内局在変化により、活性化に伴う機能変化を再現する「**血小板細胞**」代謝、シグナルシミュレーションを作成する。
 - 微小循環内における血管壁損傷部位への「**血小板細胞**」の接着、集積動態によりマイクロメートルスケールからミリメートルスケールを再構成する「**血小板細胞**」接着シミュレーターを作成する。
 - 「**血小板細胞**」を中心に、血管機能の調節機構の構成的理解を目指す

ISLIMソフトウェア研究開発報告会 2





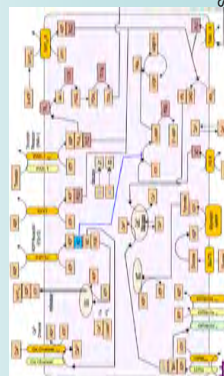
概要・アプローチ

- 基盤プログラムとしてRicken Cell Simulator Platform (RICS)を用いる
- アプローチ
 - 実証実験にて取得した血小板細胞の構造情報をRICSに入れ込み、「**血小板細胞**」形状モデルを作成する
- 「**血小板細胞**」を構成する各ボクセル内に、代謝、活性化に寄与する化学反応式を入れ込む
 - 単一血小板接着モデルを作成し、複数血小板の接着、凝集に拡大する
- 「**血小板細胞**」代謝、活性化モデルと「**血小板細胞**」モデルの連成計算を行なう



RICS上に表現した血小板。
|血小板モデル|
細胞膜| GPIIb/IIIa | 細胞内オルガネラ





ISLIMソフトウェア研究開発報告会 3




現在までの研究開発成果

- 現時点の開発ソフトウェアについて
 - オルガネラの局在を組み込んだ「**血小板細胞**」形状モデルは確立できた。また、実証実験により得られた精緻な生物学的情報により精緻化した「**血小板細胞**」活性化基盤モデルも確立できた。実証実験を反映する「**血小板細胞**」接着基盤モデルも構成できた。
 - RICSにおいて、細胞内の現象を同時に再現するために、RICSにおいて移流計算を行なうべく工夫を行なっている。特殊な条件において計算は可能であるが、生物学的に妥当な条件では明らかな計算エラーがでる。
- RICSはe-cellを基盤とした、細胞内の代謝、活性化シグナルの基盤ソフトである。細胞外の移流拡散計算のために組み込んだサブレーションは血小板細胞には最適化されていない。
 - 移流計算のサブレーションの組み込みにはさらに工夫が必要
 - 計算例として、実証実験とシミュレーションによる細胞内カルシウムイオン濃度変化を示す





ISLIMソフトウェア研究開発報告会 4

データ解析融合研究開発チーム成果報告

データ解析融合研究開発チーム

宮野 悟

次世代計算科学研究開発プログラム
データ解析融合研究開発チーム チームリーダー



発表者紹介

- 1977年 3月 九州大学理学部数学科卒業
- 1979年 3月 九州大学大学院理学研究科修士課程数学専攻修了
- 1979年 6月 九州大学理学部助手
- 1985年 4月 Alexander von Humboldt 財団研究員
- 1987年 4月 Universität GH Paderborn 助手
- 1987年 12月 九州大学理学部助教授
- 1993年 3月 九州大学理学部教授
- 1996年 4月 東京大学医科学研究所教授

研究分野

計算システム生物学、バイオインフォマティクス、メディカルインフォマティクス

データ解析融合研究開発チーム

宮野 悟

次世代計算科学研究開発プログラム
データ解析融合研究開発チーム チームリーダー

1. 目的

「計測技術の大規模化・精緻化・簡便化により、遺伝子から環境因子にわたる生命システムに関するデータが、超高次元化・超ヘテロ化・超膨大化した。しかも、不観測性・欠損の問題も同時に有している。こうした中、データ解析はデータの多様化と増大化に急速に引き離され、シミュレーションは生命体個別の現実データを反映できず予測能力に限界がある。方法論にパラダイムシフトが必要となった。」これが、2006年、研究の準備段階での研究推進のメッセージであった。それから7年を経過した現在、次世代シーケンサーの発達や遺伝子発現データ解析の高精度化・低コスト化が大きく進み、がんなどの病態の分子システムの理解と病気の原因遺伝子の解明に向けたデータ解析及びシミュレーション技術の研究は全世界的潮流になった。本チームの目的設定が極めて的確であったといえる。ただ、京の利用期間が想定していたよりも短かったこと、及び、現時点での京の運用がライフサイエンス研究に最適になっていないことが残念なところである。

本研究チームは、ペタスケールの計算を前提にして、超高次元大規模ヘテロデータ解析と生命体シミュレーションを融合し、生命体システムに対する予測と発見の基盤情報技術を構築することを目的として研究開発を行ってきた。4人のPIが、図1の文脈で、次の3つの技術開発を目標に研究を行った。

① ゲノムワイド関連解析により疾患や薬物反応

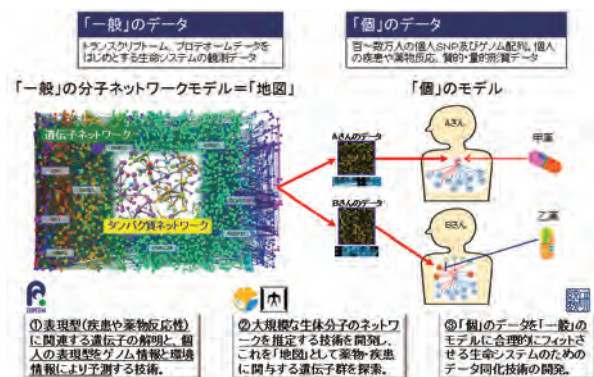


図1

性などの表現型に関連する遺伝子の解明と、個人の表現型をゲノム情報と環境情報により予測する技術

- 理化学研究所ゲノム医科学研究センター・角田達彦(2011.4~2013.3)；鎌谷直之(2006.10~2011.3)
- ② 大規模な生体分子のネットワークを推定する技術を開発し、これを「地図」として薬物・疾患に関与する遺伝子群を探索する技術
 - 東大医科学研究所・宮野悟
 - 東京工業大学情報理工学研究科・秋山泰
- ③ データ同化技術の活用による一般のモデルから個のモデルを創出する技術
 - 統計数理研究所・樋口知之

2. 成果

2.1 「肺がんと薬」を共通軸とした研究と、その広がり

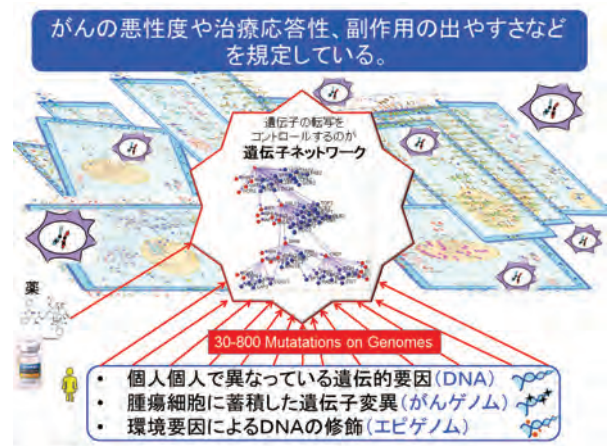


図2

本チームの研究開発は2006年に開始したが、2008年に実施された研究の進捗状況についての中間評価において、チームとして共通の研究軸をもつことによりチーム内での相乗効果を発揮さ

せることが指示された。そこで、本チームのメンバーが実験データの取得も含め対応可能なものとして「肺がんと薬」を共通のテーマとしたが、優れた技術が開発されたため、社会的・科学的インパクトの大きな成果を出すこともできるよう、多少研究に広がりを持たせた。図2のもと、図3のように技術開発の方向を明確化したが、肺がんとその薬に限らず、本チームが開発したソフトウェアアプリケーション群は有効であることが示された。

2.2 開発ソフトウェアと開発責任者

チームの構成と役割を図4に示している。チームリーダーの宮野悟は「大規模遺伝子ネットワーク推定とその応用」を業務委託課題として担当し、ヒト全遺伝子規模の遺伝子ネットワークを推定・探索するソフトウェア群を開発した。この全体パッケージを SiGN と称している。ベジアンネットワークにノンパラメトリック回帰を組み合わせた方法(SiGN-BN)、状態空間モデルに次元圧縮を組み合わせた方法、構造方程式もでるなどに L1 正則化法を駆使したネットワーク推定法群(SiGN-L1)などのアプリケーション群がパッケージに含まれており、すべて京コンピュータでの高並列化を達成した。トランスクリプトームのネットワークを推定することに用いることができるが、ソフトウェアとしては生命科学に限らず、汎用的なものである。ノード間の通信量が京をもちいても効率よくスケールさせることが困難であることが判明した「最適ベジアンネットワーク推定ソフトウェア」では、32 ノードの最適ベジアンネットワークの推定を実現し、これは世界記録となっている。

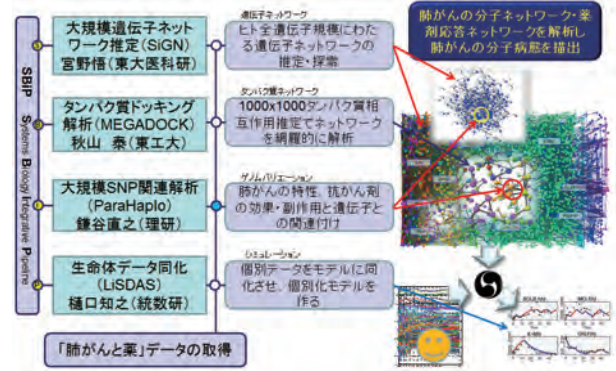


図4

秋山泰は、1000×1000 規模のタンパク質相互作用を網羅的に推定することでタンパク質ネットワークを構築できるソフトウェア MEGADOCK を開発し、京コンピュータでの高並列化を達成した。その能力は 1000×1000 を超えるものとなった。網羅的なタンパク質間相互作用予測システムを京コンピュータの上で効率的に動作する並列ソフトウェアパッケージとしてまとめ、前処理・後処理のツール群も京コンピュータ周辺環境で動作することを実現した。MEGADOCK の開発により、予測されたタンパク質相互作用情報をトランスクリプトームネットワークと合わせてシステムを理解することが京コンピュータの計算リソースが十分に利用することができれば、様々な疾患においても、分子ネットワーク・薬剤応答ネットワークを解析し、その分子病態を描出することを狙うことが可能になり、今後に期待が大きい。

角田達彦は、前任者の鎌谷直之の開発を引き継ぎ、ParaHaplo というソフトウェアを開発した。京コンピュータでの高並列化を達成し、Wellcome Trust のデータを用いてその優れた点を確認している。これは肺がんに限ったものではなく、患者群と対照群の全ゲノム上の SNP を用いハプロタイプ単位のゲノムワイド関連解析を行い、疾患関連遺伝子を網羅的に探索するためのソフトウェアである。Nature Genetics に掲載されている論文の 55% が GWAS を用いて、疾患遺伝子を探索しており、GWAS は理化学研究所の中村祐輔（現・シカゴ大学教授）の率いる遺伝子多型研究センター（当時の名称。その後、ゲノム医科学研究センターに改組）が世界で最初に開発し、その有効性を証明したものである。これから時間をかけて世界において HaraHaplo の有効性が確認実証されることを期待したい。また、角田はこれに関連して、拡張 RAT 法による 2 SNP 組合せの全

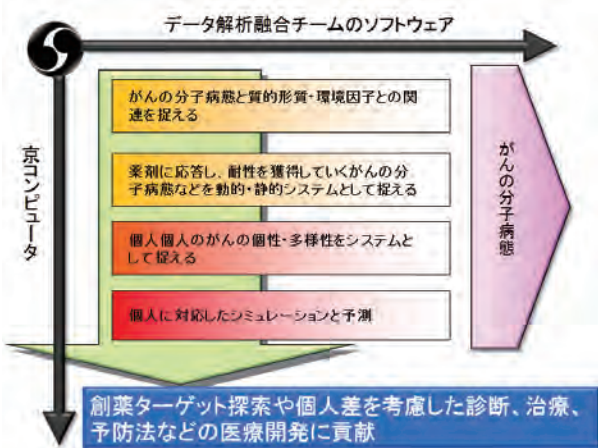


図3

ゲノム関連解析ソフトウェア (ExRAT) を開発している。これは、病気へのリスクを上げる原因となるような複数の遺伝子による相乗効果を大規模体系的網羅的に見つけ出すソフトウェアであり、京コンピュータのアーキテクチャに合わせ並列度を上げ、京コンピュータ上で実データを用いて実行させて、未知の複数要因による新たな疾患発症機序の発見と、未だ解決していない遺伝力の問題を解く成果を狙っている。京コンピュータにおいてかなりの高並列化を達成している。また、次世代シーケンサーデータ解析プログラム (NGSanalyzer) は、がんを体細胞変異のゲノム学と遺伝統計学により解明するために、次世代シーケンサーからの人間のがん細胞と正常細胞の全ゲノムの膨大なデータを網羅的・正確・高速に解析するためのソフトウェアである。がんの種類や対象とする病気により、解析パイプラインのパラメータの調整が実験による検証のプロセスを経ながら調整する必要があるため、パイプラインの並列度を上げ、オープンソースでコードを配布し、様々のゲノム解析研究において京コンピュータ上で利用されていくことが期待される。

データ同化技術は、個別データをモデルに同化させ、個別化モデルを作り、個に対するシミュレーションと予測をするものである。LiSDAS は、階層性を考慮した粒子フィルタアルゴリズムなどを京コンピュータ上で超高並列化したソフトウェアであり、京コンピュータにおいて高並列化を達成している。オミックス計測技術から生成されるデータを利用して、生体内分子相互作用ネットワークのシミュレーションモデルを構築することに応用している。

最後に、開発したソフトウェアを連携して使うことができる SBiP (Systems Biology integrative Pipeline) というデータ解析プラットフォームを宮野のグループで開発した。SBiP は京コンピュータで走らせるソフトウェアではなく、ユーザのコンピュータにインストールし、高機能 GUI により、SBiP から京コンピュータの上で走るソフトウェアをつないで解析パイプラインを作成できるものである。京コンピュータの運用が開始されてから歯科開発に着手できない部分があり、また、東大医科学研究所ヒトゲノム解析センターのスーパーコンピュータを使って考案した解決策も京コンピュータの運用方針に合わないとのこ

とで、苦難と膨大な努力の末、SBiP から SiGN 関係のソフトウェアを京コンピュータで実行しその結果をもどし、人手では実行不可能な手順の大規模データ解析 (遺伝子ネットワーク解析を中心) ができるものになりつつある。ドキュメンテーションもきちんとしており、オープンソースですでに公開している。

3. プロジェクト終了時の達成目標

ネットワーク解析による薬のターゲット探索やがんなどの病態を理解する情報技術、並びにゲノム情報に基づいた個人差を考慮した医療のための基盤情報技術を創出することができ、目標を達成することができた。本プログラム期間内に得られた個々の成果については、各 PI から報告がある。

2012 年になり、半導体チップで DNA を読む革新的なシーケンサーの実用化された。さらに、2013 年には、高集積半導体チップがリリースされ、誰もが自分の全 DNA 情報を 10 万円程度で丸ごと手にいれられることが確実になった。さらに、1 万円ゲノムを実現するナノポア半導体シーケンサー技術も実用化されている。これまで主に生物や病気の「研究」のために行っていたシーケンスに対し、がんや患者さんの全 DNA 情報や遺伝子発現情報を得て、臨床的に翻訳・解釈し、治療として「患者さんに戻す」臨床シーケンスの研究が米国やカナダでは既に始まり、国を上げて個別化医療を推し進めている。そして、医療・生命科学のビッグデータが誕生する。エピゲノム、トランスクリプトーム (シーケンス、及び定量データ) もこの次世代のシーケンサーで解析できるものである。これまで遺伝子発現プロファイルデータの取得には、マイクロアレイを使ってきたが、これが次世代シーケンサーによる RNA-seq という方法に置き換わり始めている。

本チームが開発したソフトウェアアプリケーションは、まさにこの医療・生命ビッグデータ時代に挑戦できるものであり、京コンピュータ、並びにその他のスーパーコンピュータを用いていた活用に大きな期待を寄せている。本チームの研究開発は、単に目標達成という「終わり」ではなく、個別化医療新時代の幕を開けてくれるものと考えている。

データ解析融合研究開発チーム チーム成果統括報告



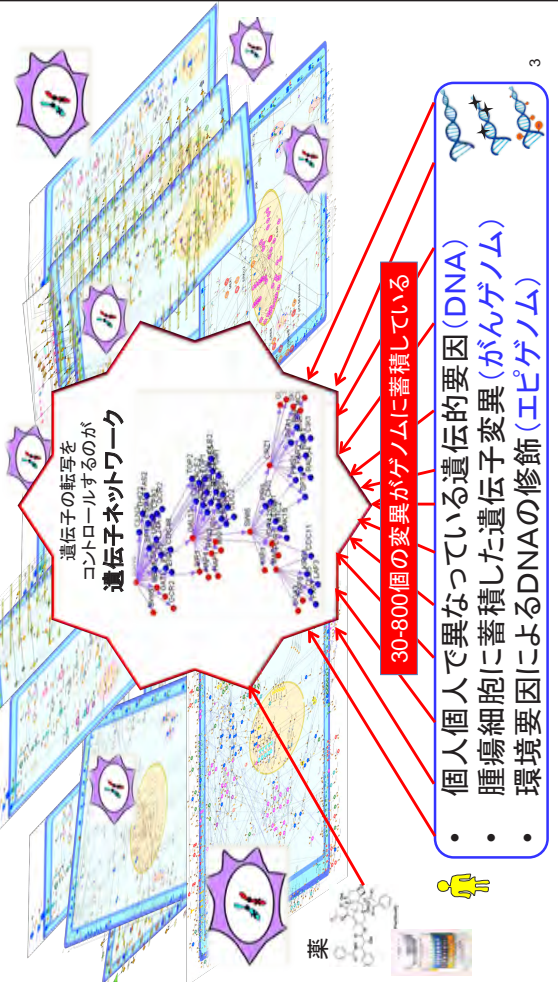
理化学研究所
次世代計算科学研究開発プログラム
データ解析融合研究開発チーム チームリーダー
宮野 悟
2013年1月10日

 東京大学 (宮野 悟) 大規模遺伝子ネットワーク推定とその応用 	 統計数理研究所 (樋口 知之) 生命体シミュレーションのためのデータ同 化技術の開発
 理化学研究所 (角田 達彦) 大規模ゲノム多型データと表現型 データを関連付ける新規アルゴリス ムの開発と、妥当性、有用性の検討 	 東京工業大学 (秋山 泰) 大規模タンパク質ネットワーク推定とその 応用

ISLiMソフトウェア研究開発報告会

1

がんの悪性度や治療応答性、副作用の出やす さなどを規定している。



3

データ解析融合チームの背景・目的



「一般」のデータ

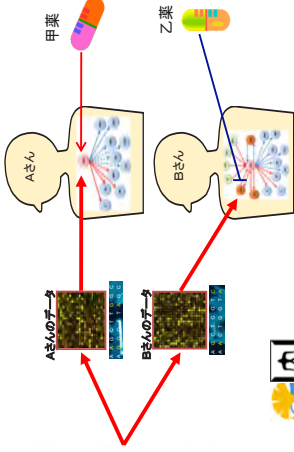
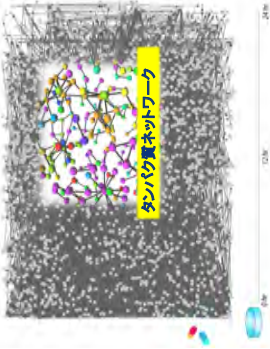
トランスクリプトーム、プロテオームデータをはじめとする生命システムの観測データ

「個」のデータ

百〜数万人の個人SNP及びゲノム配列、個人の疾患や薬物反応、質的・量的形質データ

「一般」の分子ネットワークモデル≒「地図」

薬剤に反応して動く遺伝子ネットワーク



① 薬剤型 (疾患や薬物反応性) に
関連する遺伝子情報と環境情報に
より予測する技術。

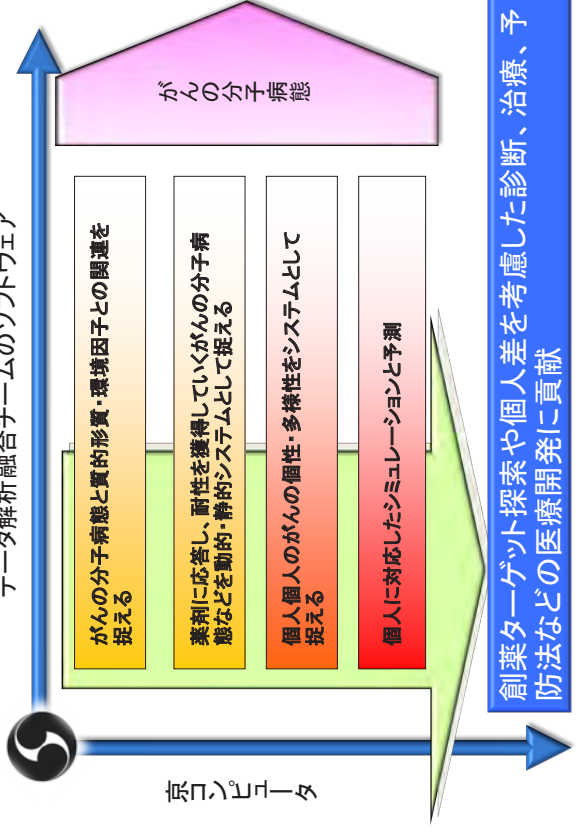
② 大規模な生体分子のネットワー
クを推定する技術を開発し、これを
「地図」として薬物・疾患に関与
する遺伝子群を探索。

③ 「個」のデータを「一般」のモ
デルに合理的にフィットさせる
生命システムのためのデータ
同化技術の開発。

2

「肺がんと薬」のシステムを理解するための課題

データ解析融合チームのソフトウェア



4

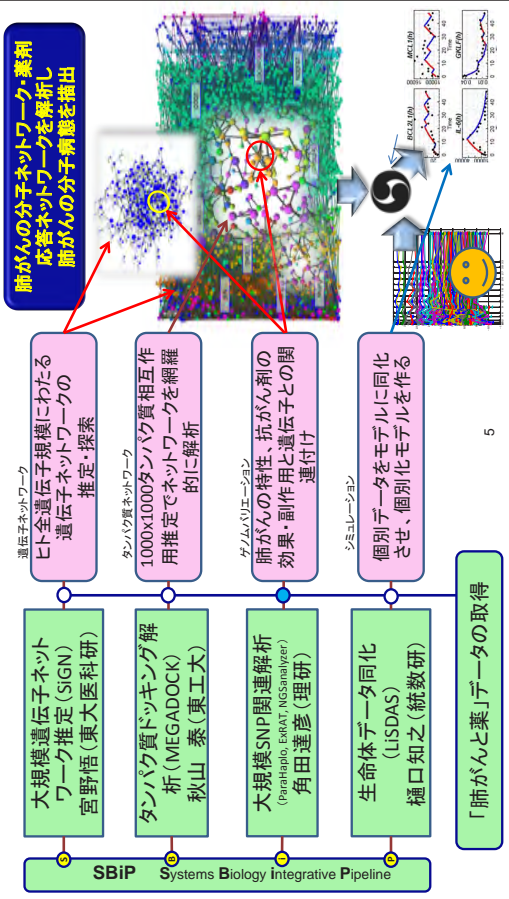


開発アプリケーション

アプリケーション名	略称	担当PI
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	角田達彦
次世代シーケンサーデータ解析プログラム	NGS analyzer	角田達彦
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	角田達彦
大規模遺伝子制御ネットワーク推定プログラム	SIGN SIGN-BN, SIGN-SSM, SIGN-L1	宮野 悟
データ解析融合プラットフォーム	SBiP	宮野 悟 スバロンには裏表しない
生命体データ同化プログラム	LISDAS	樋口知之
網羅的タンパク質ドッキング解析プログラム	MEGADOCK	秋山 泰



チームの構成と役割・開発ソフトウェア





大規模遺伝子ネットワーク推定とその応用

データ解析融合研究開発チーム
東京大学医科学研究所ヒトゲノム解析センター
宮野 悟

■業務参加者

東大医科研ヒトゲノム解析センター

新井田厚司(特任助教・本委託事業による雇用)

鳥村徹平(助教) 斉藤あゆむ(技術職員) 関谷弥生(教務補佐員)

山口 類(講師) 井元清哉(准教授)

東大情報理工・コンピュータ科学専攻

玉田 嘉紀(助教)

□業務協力者 長崎正朗(東北大学)(平成23年度まで業務参加者)

ISLiMソフトウェア研究開発報告会

1



大規模遺伝子ネットワーク推定ソフトウェアSIGNの開発目的

肺がんなどの分子ネットワーク・薬剤応答ネットワークを解析し、その分子病態を描出するため、DNAチップ等(これからは次世代シーケンサー)によるRNA-Seqによる定量)で計測された遺伝子発現データから遺伝子間の発現の依存関係を表す「遺伝子ネットワーク」を推定・予測するソフトウェアを開発すること。

ISLiMソフトウェア研究開発報告会

2

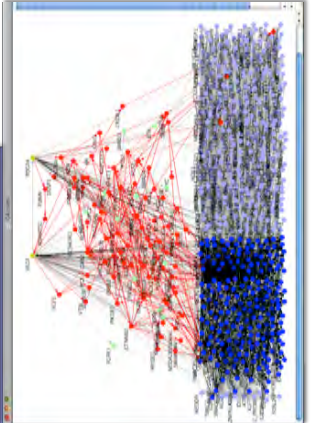


SIGNの達成したこと

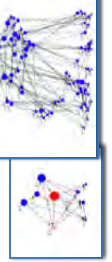
- 最適な数十遺伝子規模のネットワークからゲノムワイド(2万遺伝子)ネットワークまで、動的・静的・パーソナルネットワークの推定を可能にした。

- SIGN-BN: Bayesian N+ Nonparametric Regression
- SIGN-SSM: State Space Model + Dimension Reduction
- SIGN-L1: Statistical Graphical Modeling + L1 Regularization

ゲノムワイドネットワーク



最適ネットワーク



SIGNで、たとえば、次のようなことが可能になってきた

- 抗がん剤に応答して動的に変化する大規模な遺伝子ネットワークを描出すること
- がんの予後・再発を予測する遺伝子群をネットワーク解析から同定
- 数百のがんサンプルの遺伝子発現プロフィールデータから個々のがんの個性・多様性を分子ネットワークとして抽出すること(再発・生存・死、抗がん剤が効く人・効かない人の違い)
- これらを地図として用いて、新たな分子標的探索の可能性がでてきた

ISLiMソフトウェア研究開発報告会

4

SiGN-L1による大規模遺伝子ネットワーク解析で一挙に捉えた上皮間葉転換の鍵遺伝子群

上皮間葉転換(Epithelial-Mesenchymal Transition EMT)

- 上皮系細胞が間葉系細胞に形態変化する現象
- がんの悪性度、がんの浸潤性、線維症に深く関連
- ゲファイチニブなどの抗がん剤を効かなくする原因

EMTを引き起こすいくつかの遺伝子は見つかっているが、そのメカニズムはよくわかっていない

【共同研究】
高橋 隆(名大医学研究科)

http://gamoku3.urin.jp/topics/saito.html ISLIMソフトウェア研究開発報告会

i番目のサンプルの構造方程式モデルの係数を推定

Sample Weightingを使う

モジュレータ: 遺伝子AとBの条件付き従属性に影響を与え因子を導入

モジュレータ i-th sample

Low High

ネットワーク

ISLIMソフトウェア研究開発報告会

SiGN-L1

- 方法: 3種類の統計的グラフィカルモデル(グラフィカルガウシアンモデル、ベクトル自己回帰モデル、構造方程式モデル)により、遺伝子発現データから遺伝子ネットワークモデルを構築
- L1正則化法を用いてスパースなモデルのパラメータ空間を探索することにより、100,000ノードまでの大規模遺伝子ネットワーク推定が可能
 - 最大並列数(コア数): 98,304 (12,288ノード)
 - 実行効率: 1コアで11.98%を達成。
 - ハイブリッド並列化に対応。

ISLIMソフトウェア研究開発報告会

構造方程式モデルで遺伝子ネットワークをモデル化

構造方程式



遺伝子ネットワーク

サンプル

共発現ネットワーク

ISLIMソフトウェア研究開発報告会

○: 遺伝子(トランスクリプト)
→: 条件付き依存関係(相関ではない)

大規模遺伝子ネットワーク解析で一挙に捉えた 上皮間葉転換 (EMT) の鍵遺伝子群 (SiGN-L1 の応用)

入力: 762 個のがん細胞株遺伝子発現データ (Epithelial/Mesenchymal のラベル付き株)
13,508 transcripts = 13,006 miRNA + 502 mRNA



出力: 13,508 個の遺伝子からなる遺伝子ネットワーク762個
EMT の度合で遺伝子ネットワークの構造がどのように変化するかを描き出された。

L1 正則化法に基づく大規模遺伝子
ネットワーク推定 (SiGN-L1)

- スバコンが EMT 抑制因子として予測したトップ24 遺伝子のうち12 遺伝子について「当たっていいえ」。残り12 遺伝子については EMT に関する直接的な結果は無し。
- 高橋隆 (名古屋) が KLF5 のノックダウンで EMT が起こることを実験で証明 (Shimamura et al. PLOS ONE 2011)
- GRHL2 が EMT 誘導することが発表される (Cieply B et al. Cancer Res 72(9):2440-53, 2012)
Shimamura I, Imoto S, Shimada Y, Hosono Y, Niida A, Nagasaki M, Yamauchi R, Takahashi T, Miyano S. (2011) A Novel Network Profiling Analysis Reveals System Changes in Epithelial-Mesenchymal Transition. PLOS ONE 6(6): e20804.

regulator	type	regulatory effect change	Evidence
IRF6	A	101.04	Nat Cell Biol 10(6): 693-691, 2008
GRHL2	A	87.08	Cancer Res 72(9):2440-53, 2012
SLIT1	A	61.13	Cancer Res 72(9):2440-53, 2012
SLIT (SIP1)	I	26.72	Nat Cell Biol 10(6): 1207-18, 2008
miR-200a	A	31.55	Nat Cell Biol 10(6): 693-691, 2008
KLF4	A	26.28	Nat Cell Biol 10(6): 693-691, 2008
miR-200a	A	22.08	Nat Cell Biol 10(6): 693-691, 2008
OVOL2	A	17.70	Cancer Res 70(5):2115-26, 2010
FOXJ2	A	17.26	Cancer Res 70(5):2115-26, 2010
TCF4 (E22)	I	14.15	J Cell Sci 122(Pt 7): 1014-24, 2009
KLF3	A	13.58	
miR-200b	A	13.58	
miR-200c	A	12.50	
miR-200d	A	12.50	
KLF5	A	12.42	PLOS ONE 6(6): e20804, 2011
miR-192	A	12.20	PNAS 104(9): 3432-7, 2007
FOXJ1	A	11.69	Cancer Res 70(5):2115-26, 2010
ZNF105	A	11.39	Cancer Res 68(7):2765-91, 2008
SNX2-1	A	11.21	
HNF1B	A	11.08	
ZEB2 (EZF)	I	10.01	Oncogene 24(14):2375-86, 2005
THM29	I	9.87	
SNAIL	I	9.74	Cancer Res 62(6): 1618-81, 2002



ISLiM ソフトウェア研究開発報告会

SiGN-BN

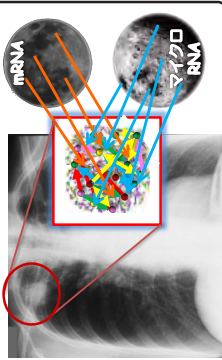
- 方法: ベイジアンネットワーク + 非線形回帰
- 高効率並列ブートストラップアルゴリズム
 - 最大動作コア数: 196,608 コア (24,576 ノード)
 - 並列化効率: 0.72 (196,608 vs 98,304 コア時)
- 全遺伝子規模のベイジアンネットワーク推定を実現
- 高並列化は無理だが、32ノードの最適ベイジアンネットワーク推定も達成 (世界記録) (京には導入せず)

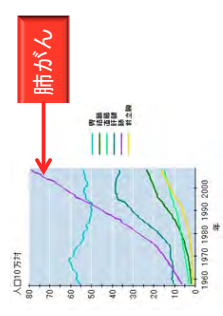
ISLiM ソフトウェア研究開発報告会

SiGN-BN が炙り出した 肺がんのマイクロRNA 遺伝子ネットワークに 映し出された「再発」のスイッチを入れている 黒幕ネットワーク達

肺がん患者の腫瘍中で実際に働いている遺伝子発現制御ネットワークを炙り出す







肺がん

がん死亡第1位
2010年 69,813人

【共同研究】
高橋 隆 (名古屋大学医学研究科)

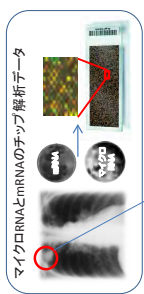
ISLiM ソフトウェア研究開発報告会


肺がんのマイクロRNA 遺伝子ネットワーク

【共同研究】
高橋 隆 (名古屋大学医学研究科)


マイクロRNA と mRNA のチップ解析データ



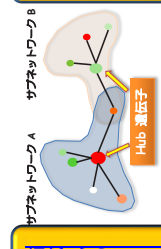
124 人の肺がん検体遺伝子発現解析データ



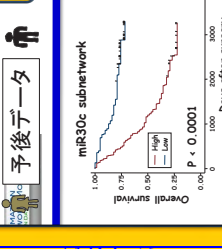
非線形回帰ベジアンネットワーク



肺がんの予後の良・不良のスイッチの同定にシステマ的方法論の有効性が示された



再発・死亡と順位に関連する 14 個のサブネットワークと、そのサブ遺伝子が見つかる



予後予測がブランク・マイヤー・タイヤグラムのp値が小さくなるサブネットワークを網羅的に探索
P < 0.0001

Computational Systems Biology
ISLiM ソフトウェア研究開発報告会

これが肺腺がんのマイクロRNA/mRNA遺伝子ネットワークといっても、訳がわからない

400 mRNA+32 microRNAのネットワーク

ISLiMソフトウェア研究開発報告会

13

NKX2-1とROR1
実験で確かめてみると、まさにそのとおり

NKX2-1を導入

肺癌検体

Expression ratio

- High in poor prognosis
- Low in poor prognosis

Expression ratio (TTF-1-introduced / Vector control)

- High in TTF-1/NKX2-1-introduced
- Low in TTF-1/NKX2-1-introduced
- N.A. (miRNA) : Positive binding in E19.5 mouse lung

CHIP-chip assay data (Tagne et al., PLoS ONE, 2012)

ISLiMソフトウェア研究開発報告会

The Cancer Gene Network Galaxy
がん遺伝子ネットワークデータベース

- 京コンピュータを259CPU年使って、256の様々ながんの遺伝子ネットワークを推定し、データベース化した
- 各遺伝子ネットワークは8000個の遺伝子からなる大規模なもの

Where is "My Cancer" in this galaxy?

ISLiMソフトウェア研究開発報告会

"The Cancer Gene Network Galaxy" のどこに KLF5が現れるかを探索してみると...



- KLF5 activates **Perp** (TP53 apoptosis effector) in **Glioma** and **Breast Cancer**.
- Until now, it is only known that direct targets of KLF5 contribute to the **maintenance of embryonic stem cell undifferentiated state**. Especially, **Perp** is one of KLF5 target. (Tommaso Russo's Group. BMC Biology, 2010)

!?

KLF5 might contribute to the maintenance of **cancer stem cell undifferentiated state** in Glioma and Breast Cancer

ISLiMソフトウェア研究開発報告会



16

SIGN-SSM

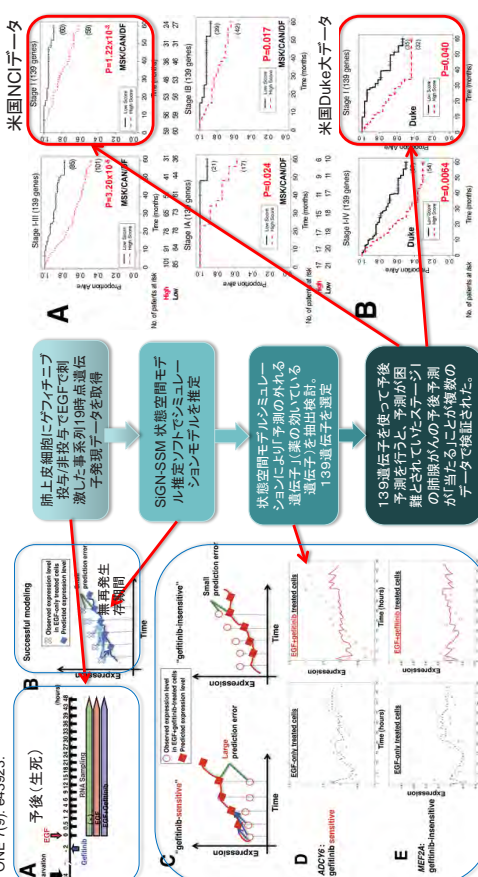
- 状態空間モデルにより時系列遺伝子発現データから、モジュールネットワーク、遺伝子ネットワークおよび動的発現予測モデルを構築
- 次元縮約および一意性制約により高次元短時系列である遺伝子発現データから動的システムの推定可能
- 高並列化・高効率化
 - 最大動作コア数: 196,608 コア (24,576 ノード)
 - 並列化効率 0.92 (24576 vs 12288)

ISLIMソフトウェア研究開発報告会 17






抗がん剤ゲフィチニブ応答EGFRパスウェイ遺伝子群の同定によるStage I の肺腺がんの予後予測の成功 (SIGN-SSM)

Yamauchi M*, Yamaguchi R*, Nakata A*, Kohno T*, Nagasaki M, Shimamura I, Imoto S, Salto A, Ueno K, Hatanaka Y, Yoshida R, Hiyachi I, Nomura M, Beer DG, Yokota S, Gotoh N (2012) Epidermal Growth Factor Receptor Tyrosine Kinase Defines Critical Prognostic Genes of Stage I Lung Adenocarcinoma. PLOS ONE 7(9): e43923.



ISLIMソフトウェア研究開発報告会






SBiP

Systems Biology integrative Pipeline

- SIGNを解析パイプラインのコンポーネントとしてユーザが簡単に利用できる高機能GUIを備えたソフトウェアプラットフォーム
- 京にて各解析パイプラインの処理の一部を実行し、その結果を、SBiPの視覚化コンポーネント群を用いて保存できるようにしている。
- ユーザはSBiPに用意されているさまざまな解析コンポーネントを組み合わせてカスタマイズした解析フローを実行可能。

ISLIMソフトウェア研究開発報告会 19

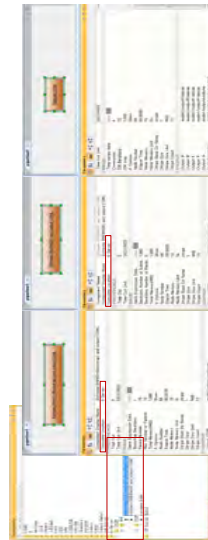
SBiP を使用した京での SIGN の実行

Webサービス機能を利用、


- SIGN-BN (HC+Bootstrap)
- SIGN-BN (NNSR)
- SIGN-SSM

を京の計算ノード上で実行するためのコンポーネントを実装した。

実行中



実行後



(H)C+Bootstrap
SIGN-BN

パイプライン上で京コンポーネントと既存コンポーネントを組み合わせることで、京での出力データをシームレスに処理できるようにした。

ISLIMソフトウェア研究開発報告会 20

大規模ゲノム多型データと表現型データを関連付ける新規アルゴリズムの開発と、妥当性、有用性の検討

角田 達彦

理化学研究所 ゲノム医科学研究センター
統計解析・技術開発グループ グループディレクター



発表者紹介


1995年3月 東京大学大学院 工学系研究科 博士課程修了
1995年4月 京都大学大学院 工学研究科 助手
1997年4月 東京大学医科学研究所 ヒトゲノム解析センター リサーチアソシエイト
1998年4月 東京大学医科学研究所 ヒトゲノム解析センター 助手
2000年4月 理化学研究所 遺伝子多型研究センター チームリーダー
2008年4月 理化学研究所 ゲノム医科学研究センター チームリーダー
2011年4月 理化学研究所 ゲノム医科学研究センター グループディレクター
医学博士・工学博士

研究分野

ゲノム医科学, 医科学数理

受賞歴


1993年 International Joint Conference on Neural Networks Student Award
1994年 情報処理学会第48回全国大会奨励賞受賞
2004年 The 13th Takeda Science Foundation Symposium on Bioscience on Genome Analysis and Medicine Poster Award
2009年 日本人類遺伝学会奨励賞受賞



大規模ゲノム多型データと表現型データを関連付ける新規アルゴリズムの開発と、妥当性、有用性の検討

理化学研究所
ゲノム医学研究センター
グループディレクター
角田 達彦


ISLiMソフトウェア研究開発報告会 1



背景・目的

- 背景: 全くの未知のものも含め、疾患の原因を探索するには、全ゲノム上で患者群と対照群との間でゲノムDNA配列を比較する、ゲノムワイド関連解析 (GWAS) が極めて有効であり、われわれ理化学研究所ゲノム医学研究センターでは、2002年に世界に先駆け初めてのGWASを実現してから、世界を牽引してきた。この方法をさらに劇的に推進するために、単点の解析だけでなく、近傍の、あるいは遠距離にある複数点の解析を行ったり(それぞれParaHaploとExRAT)、対象とするマーカーを、これまでの固定されたセットから拡張し、次世代シーケンサーにより全ゲノム配列への探索を行ったり (NGS analyzer) するなどの新たな方法を導入する必要がある。これらには、スーパーコンピュータリングを駆使することが必要不可欠になる。
- 目的:
 - ParaHaplo: GWASの方法論を拡張し、単点ではなく連続する複数点に対して超並列に行うことによって、飛躍的な検出力の向上と高速性能を狙う。
 - ExRAT: 複数の遺伝子の相互作用により発症する可能性を、全ゲノム上で網羅的に超並列に探索する。
 - NGS analyzer: 次世代シーケンサーの爆発的データを超並列に処理し、ゲノム多様性を高精度かつ包括的に解析する。

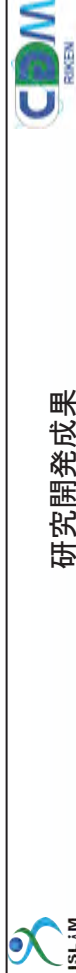
ISLiMソフトウェア研究開発報告会 2



概要・アプローチ

- 研究開発コードの概要
 - ParaHaplo: 人のゲノム全体にわたる遺伝的な相違点を、近傍にある複数点を同時にみたハプロタイプという単位で、患者群と対照群とで比較することにより、疾患の遺伝的原因を探るための統計検定計算
 - ExRAT: 遺伝子間相互作用が発症リスクを変化させる疾患関連遺伝子の組合せを全ゲノムで探索する。2SNP間の全組合せを超並列に行う方法と、SNP間の連鎖不平衡(相関)も考慮した、より精密な方法の2種類を実装。前者で全組合せをスクリーニングし、後者で経験的p値を求める手順を想定している。
 - NGS analyzer: 次世代シーケンサーの出力データを高速に解析し、個人間の遺伝的差異やがんゲノムの突然変異を高い正確さで同定する。
- アプローチ
 - ParaHaplo: Haplotype頻度を用いたType I errorの確率計算を、ハイブリッド並列で実装したマルチプロセス並列モンテカルロ法で行う。
 - ExRAT: RAT (Rapid Association Test) を、データ分割法で実装したインポートランスサンプリング法で行う。
 - NGS analyzer: ヒト標準ゲノム配列に対するマッピングと確率計算に基づいた多様性検出を、領域分割で実装した直接法による密度行列の対角化により行う。

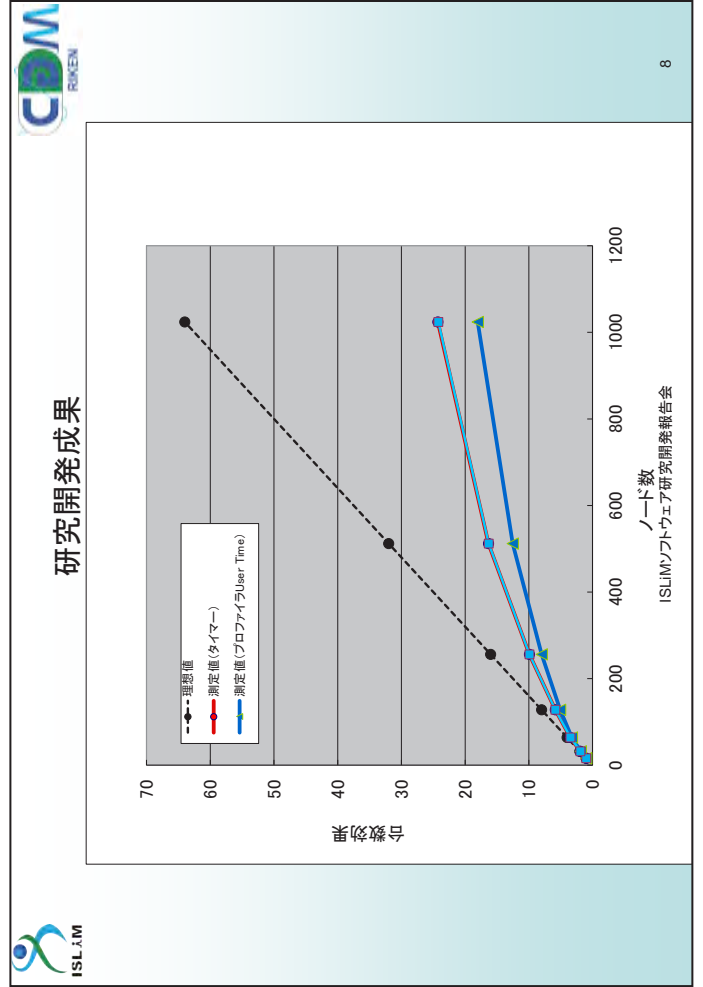
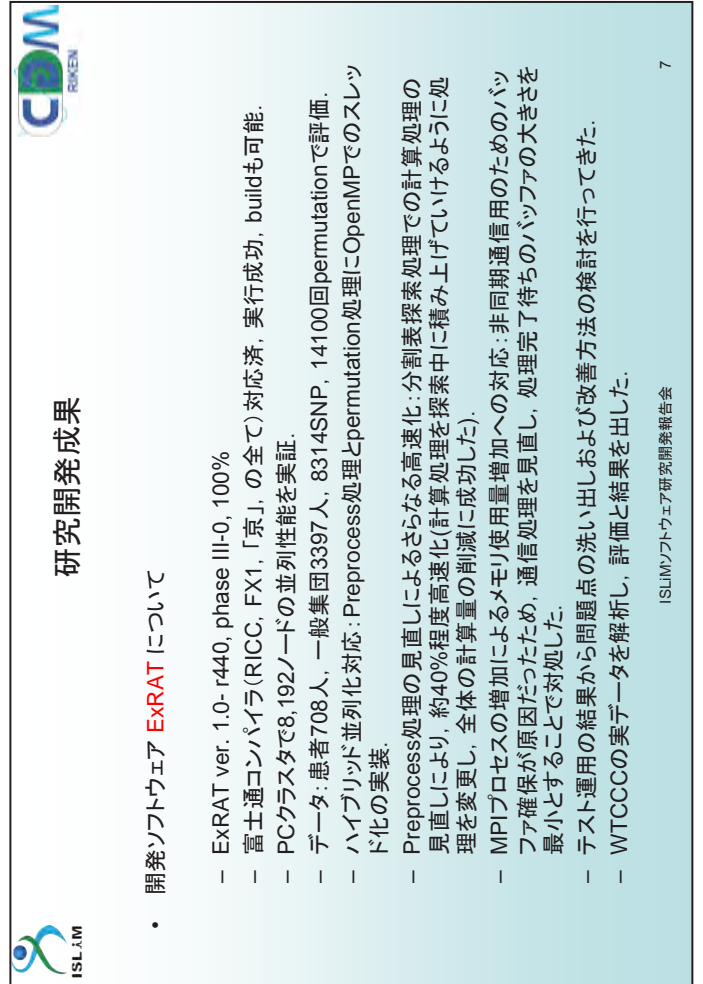
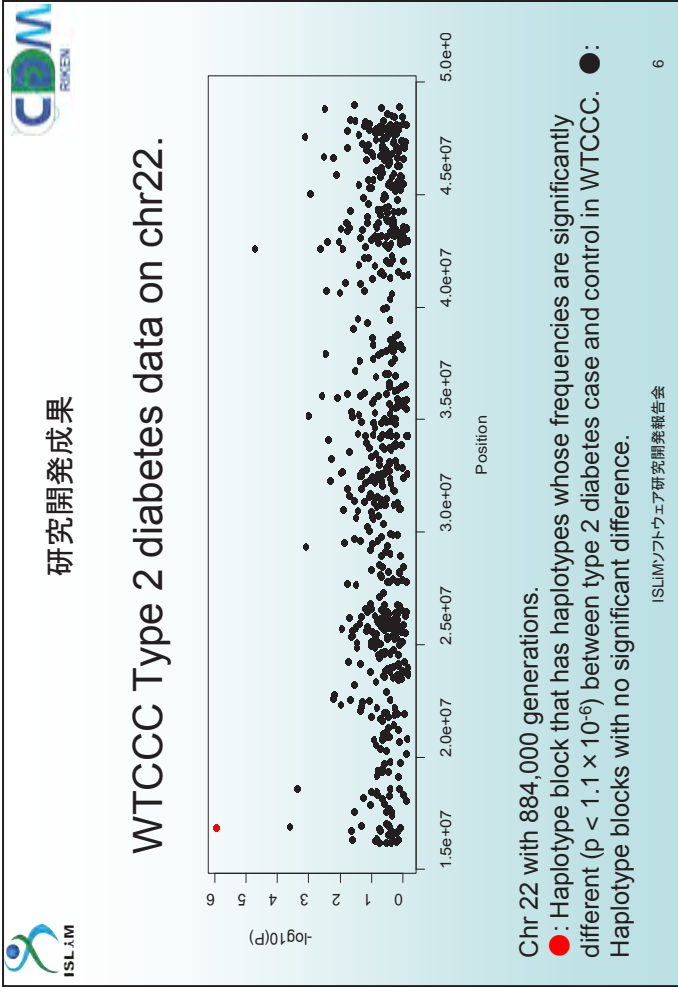
ISLiMソフトウェア研究開発報告会 3

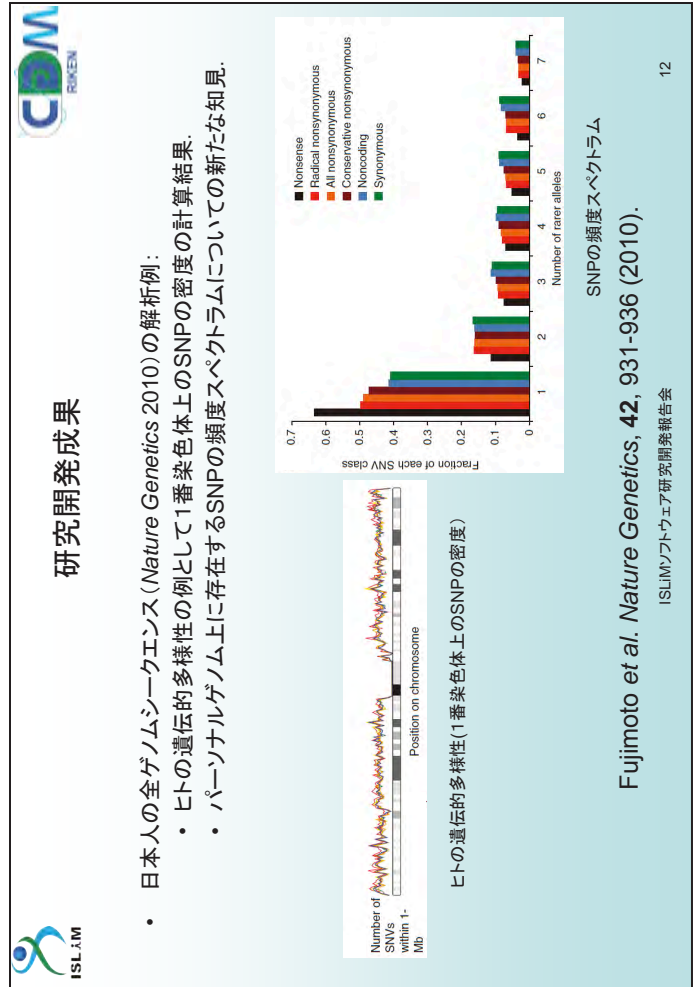
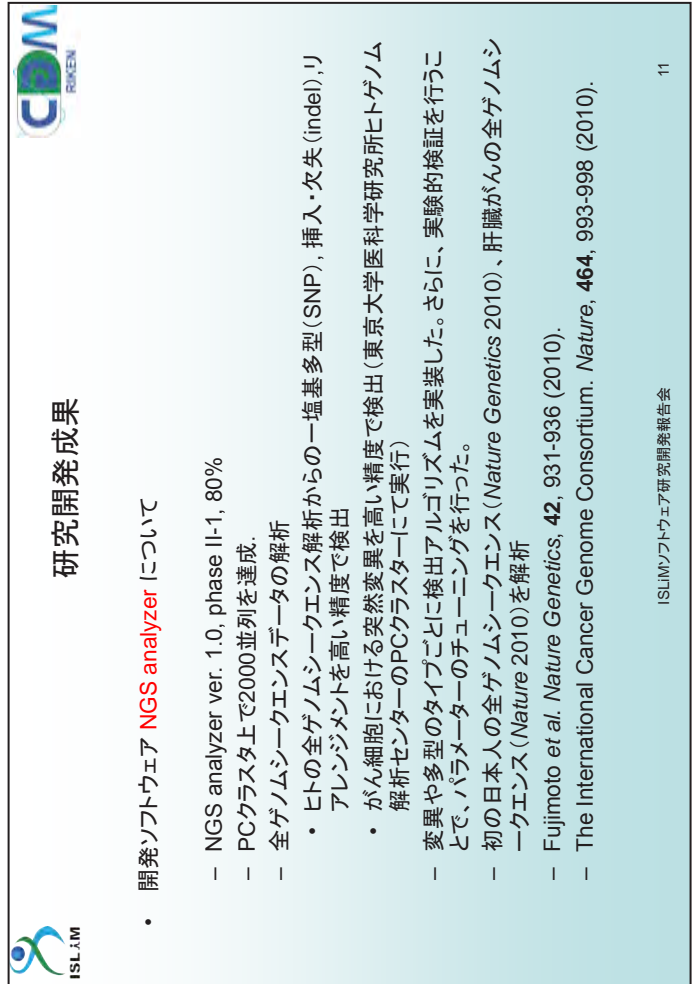
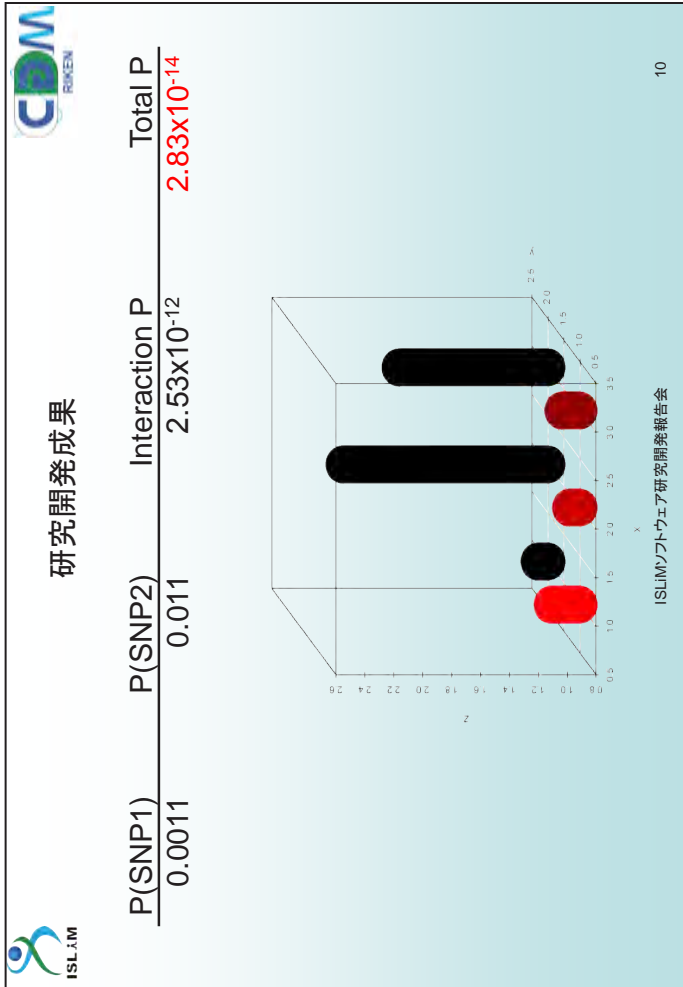
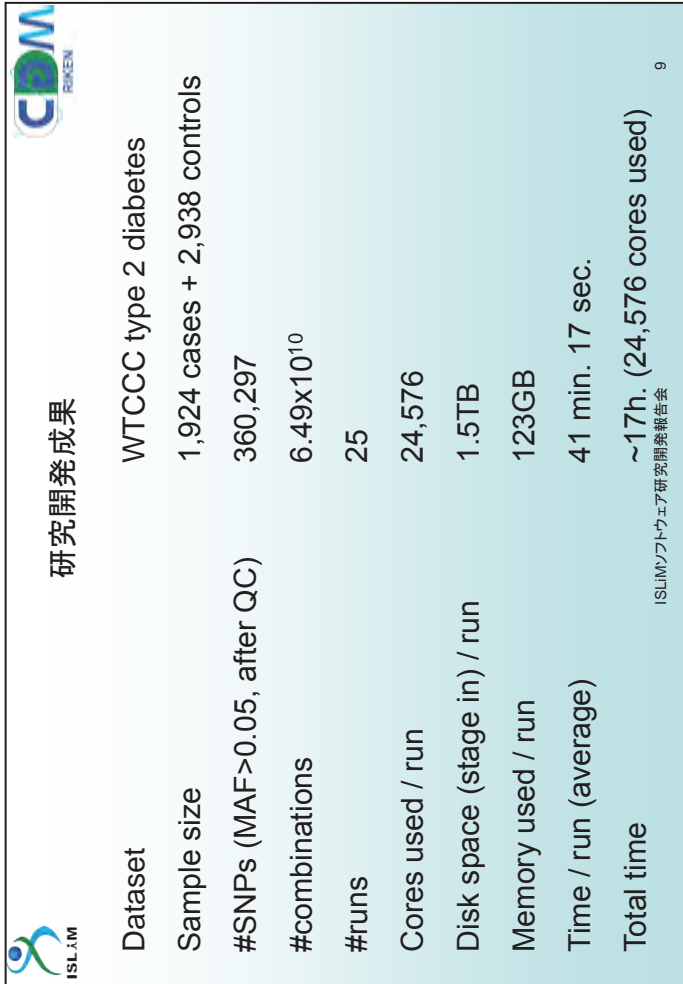


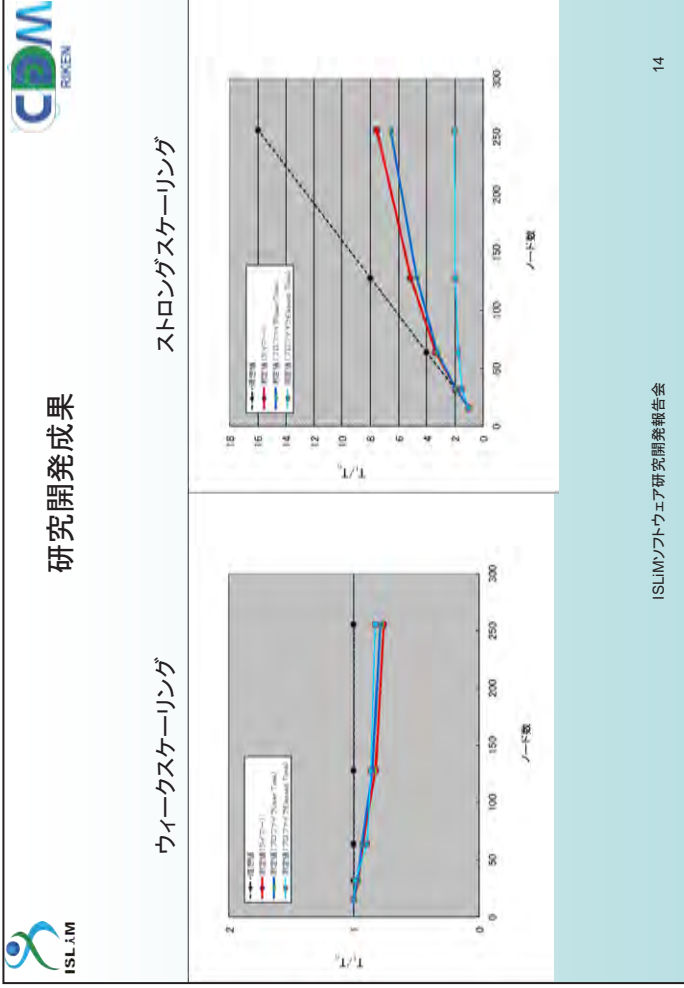
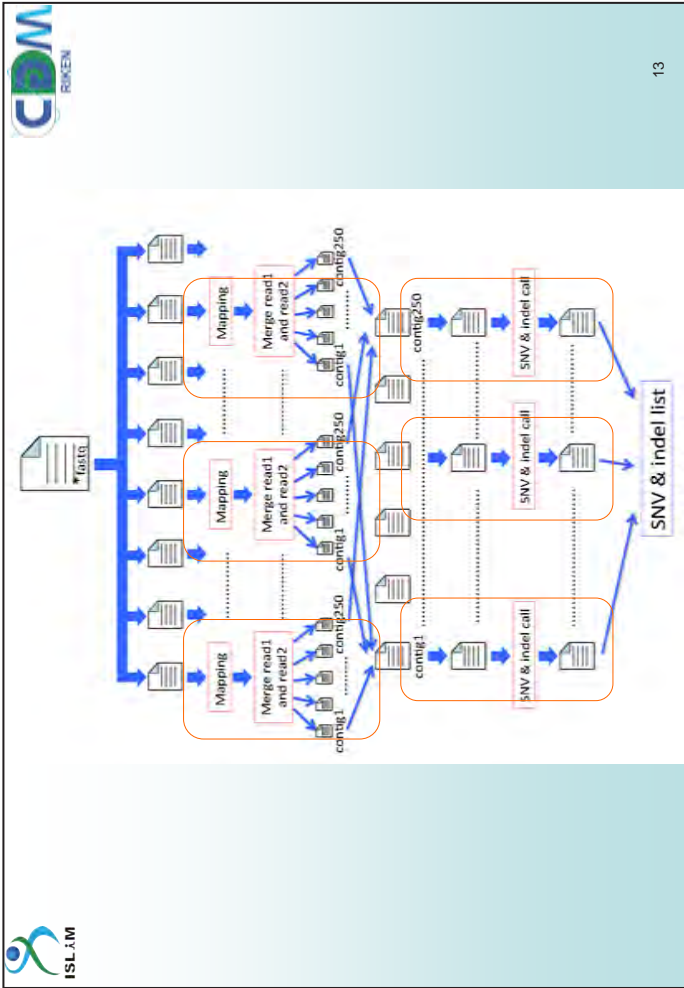
研究開発成果

- 開発ソフトウェア **ParaHaplo** について
 - ParaHaplo ver. 4.0 β , phase IV, 50%
 - 富士通コンパイラ (RICC, FX1, 「京」, の全て) 対応済、実行成功
 - PCクラスターで8,000ノードの並列性能を実証。
 - 「京」でも並列性能が出ることを確認。
 - 「京」ではコア並列 (ハイブリッド並列) も実装している。
 - 現在は、一層の高速化を目指しSIMD化を行っている。
 - ソースコード公開済。
 - 論文出版: Misawa K, Kamatani N (2011) ParaHaplo 3.0: a program package for imputation and haplotype-based whole-genome association study using parallel computing. Source Code Biol Med.

ISLiMソフトウェア研究開発報告会 4







ISLiM

CDM
ROKEN

研究開発成果

ウィークスケーリング

ストロングスケーリング

ISLiMソフトウェア研究開発報告会

14

ISLiM

nature genetics

研究開発成果

ARTICLES

Nature Genetics, 42, 931-936 (2010).

Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing

Akihiro Fujimoto^{1,2}, Hidewaki Nakagawa¹, Naoya Hosono¹, Kaoru Nakano¹, Tetsuo Abe¹, Keith A Boroevich¹, Masao Nagasaki^{1,2}, Rui Yamaguchi^{1,2}, Tetsuo Shibuya³, Michiaki Kubo¹, Satoru Miyano^{2,3}, Yusuke Nakamura^{1,3} & Tatsuhiko Tsunoda^{1,2}

We report the analysis of a Japanese male using high-throughput sequencing to x40 coverage. More than 99% of the sequence reads were mapped to the reference human genome. Using a Bayesian decision method, we identified 3,132,608 single nucleotide variations (SNVs). Comparison with six previously reported genomes revealed an excess of singleton nonsense and nonsynonymous SNVs, as well as singleton SNVs in conserved non-coding regions. We also identified 5,319 deletions smaller than 10 kb with high accuracy, in addition to copy number variations and rearrangements. *De novo* assembly of the unmapped sequence reads generated around 3 Mb of novel sequence, which showed high similarity to non-reference human genomes and the human herpesvirus 4 genome. Our analysis suggests that considerable variation remains undiscovered in the human genome and that whole-genome sequencing is an invaluable tool for obtaining a complete understanding of human genetic variation.

merica, Inc. All rights reserved.

ISLiM

CDM
ROKEN

研究開発成果

LETTERS

Nature Genetics, 44, 760-764 (2012).

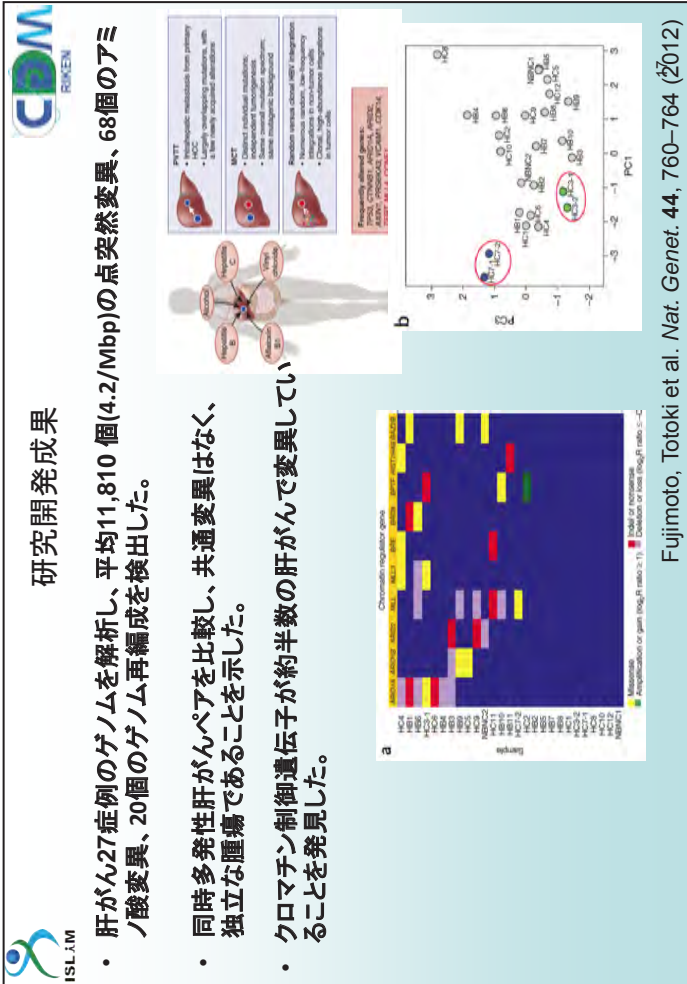
Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators

Akihiro Fujimoto^{1,16}, Yasushi Totoki^{2,16}, Tetsuo Abe¹, Keith A Boroevich¹, Fumie Hosoda², Ha Hai Nguyen¹, Masayuki Aoki¹, Naoya Hosono¹, Michiaki Kubo¹, Fuyuki Miya¹, Yasuhito Arat², Hiroyuki Takahashi², Takuya Shirakihara², Masao Nagasaki³, Tetsuo Shibuya³, Kaoru Nakano¹, Kumiko Watanabe-Makino¹, Hiroko Tanaka¹, Hiromi Nakamura², Jun Kusuda⁴, Hidenori Ojima⁵, Kazuaki Shimada⁶, Takuji Okusaka⁷, Masaki Ueno⁸, Yoshinobu Shigekawa⁸, Yoshihiko Kawakami⁹, Koji Arihiro¹⁰, Hideki Ohdan¹¹, Kunihito Gotoh¹², Osamu Ishikawa¹², Shun-ichi Aritaumi¹³, Masakazu Yamamoto¹³, Terumasa Yamada¹², Kazuaki Chayama^{14,9}, Tomoo Kostege⁶, Hiroki Yamaue⁸, Naoyuki Kamatani¹, Satoru Miyano³, Hitoshi Nakagata^{5,14}, Yusuke Nakamura^{1,15}, Tatsuhiko Tsunoda¹, & Hidevaki Nakagawa¹

ights reserved.

研究開発成果

- 肝がん27症例のゲノムを解析し、平均11,810個(4.2/Mbp)の点突然変異、68個のアミノ酸変異、20個のゲノム再編成を検出した。
- 同時多発性肝がんペアを比較し、共通変異はなく、独立な腫瘍であることを示した。
- クロマチン制御遺伝子が約半数の肝がんで変異していることを発見した。



Fujimoto, Totoki et al. *Nat. Genet.* 44, 760-764 (2012)

タンパク質間相互作用ネットワークの推定と その応用に関する研究

秋山 泰

東京工業大学 大学院情報理工学研究科
計算工学専攻 教授



発表者紹介

- 1990年 3月 慶應義塾大学大学院理工学研究科電気工学専攻博士課程修了(工学博士)
- 1990年 4月 通商産業省工業技術院電子技術総合研究所 研究官
- 1992年 4月 京都大学 助教授(化学研究所)
- 1996年 4月 技術研究組合新情報処理開発機構 研究室長(並列応用つくば研究室)
- 2000年 4月 通商産業省工業技術院電子技術総合研究所 主任研究官(生命情報科学研究センター検討チーム長)
- 2001年 4月 独立行政法人産業技術総合研究所 生命情報科学研究センター長
- 2007年 4月 東京工業大学大学院情報理工学研究科計算工学専攻 教授
- 2011年 12月 東京工業大学情報生命博士教育院 教育院長(兼任)

研究分野

バイオインフォマティクス、並列処理応用

タンパク質間相互作用ネットワークの推定とその応用に関する研究
Large-scale protein-protein interaction network prediction by an exhaustive rigid docking system MEGADOCK

秋山 泰
 東京工業大学 大学院情報理工学研究所

■ 業務参加者
 松崎由理 (産学官連携研究者・本委託事業による雇用)、石田貴士 (助教)
 □ 業務協力者
 内古爾伸之 (現所属: 中央大学) (平成23年度まで業務参加者)

What we do:
PPI network estimation from proteins tertiary structure data

Input: Protein structures

Output: PPI networks

Target: Millions of combinations of proteins
Strategy: Exhaustive rigid docking among target proteins

- Feasible assuming massively parallel computing environments
- Prediction of complex structures can also be provided

ISLiMソフトウェア研究開発報告会 2

MEGADOCK: An ultra-fast and parallelized protein docking tool

- Questions asked by protein docking
 - 1:1 docking: How given proteins interact?
 - Our challenge by exhaustive docking: **Which protein pairs interact?**
- Developing an ultra-fast docking tool **MEGADOCK** feasible for exhaustive docking of million-order pairs

Need large number of docking calculations

Where? 1:1 docking

Our problem

Exhaustive rigid docking and post-docking analysis

Exhaustive docking & post-docking analysis

	P1	P2	P3	P4	P5
P1	○	×	×	×	×
P2	×	○	×	×	×
P3	×	×	○	×	×
P4	×	×	×	○	×
P5	×	×	×	×	○

PPI network prediction

- Search **novel PPIs** in lung cancer pathway proteins and cancer **drug related proteins**
- The first large-scale PPI prediction challenge (2000 x 2000 = **4 million** scale)
- To contribute finding target proteins of cancer drugs

EGFR signaling network

A compact score function of MEGADOCK

Compress three terms into **one complex number**:

- Shape complementarity
- Hydrophobic interaction
- Electrostatic interaction

$$S(\alpha, \beta, \gamma) = \Re \left[\sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N R(l, m, n) L(l + \alpha, m + \beta, n + \gamma) \right]$$

$$R(l, m, n) = G_R(l, m, n) + w_h H(l, m, n) + i\phi(l, m, n)$$

$$L(l, m, n) = G_L(l, m, n) + iw_e q(l, m, n)$$

$$S(\alpha, \beta, \gamma) = \Re \left[\text{IFT} \left[\text{DFT} [R(l, m, n)]^* \text{DFT} [L(l, m, n)] \right] \right]$$

Convolution can be calculated fast by FFT (Katchalski-Katzir model)

Ligand protein

Receptor protein

1+H +iφ	2+H +iφ	3+H +iφ	3+H +iφ	1+H +iφ
-45	-45	-45	-45	-45
2+H +iφ	-45	-45	2+H +iφ	2+H +iφ
-45	-45	-45	-45	-45
3+H +iφ	-45	-45	3+H +iφ	3+H +iφ
-45	-45	-45	-45	-45
1+H +iφ	2+H +iφ	3+H +iφ	3+H +iφ	1+H +iφ
-45	-45	-45	-45	-45
2+H +iφ	-45	2+H +iφ	2+H +iφ	2+H +iφ
-45	-45	-45	-45	-45
3+H +iφ	3+H +iφ	1+H +iφ	1+H +iφ	3+H +iφ
-45	-45	-45	-45	-45

Ohue M, Matsuzaki Y, Ishida T, Akiyama Y.
Lecture Note in Bioinformatics 7632,
178-187, Springer, 2012.

ISLIMソフトウェア研究開発報告会

5

Implementation with hybrid parallelization

MPI parallelization

Thread parallelization

OpenMP

Receptor voxelization

Receptor FFT

Loop for Ligand rotations

Ligand voxelization

Ligand FFT

Convolution

Score calculation (Inverse FFT)

Save high scoring docking poses

Worker Node

Master Node

Receptor

Ligand

Post-docking analysis to predict interacting pairs

- Get 2,000 high scoring decoys by docking of each protein pairs
- Conduct clustering based on structure similarity
- The highest docking score (normalized) of the data included in the cluster $C_i : s_i$
- Cluster population (normalized) : m_i
- Select populated clusters C^* with threshold m^* of population of the cluster $C^* = \{C_i | m_i > m^*\}$
- Decide PPI score E
- Evaluate each pair of protein combination as interacting if E is higher than the threshold E^*

Receptor

Ligand

C_1

C_2

S_1 (Normalized docking score)

S_2

Number of data

cluster

$$E = \begin{cases} \max s_i, i \in C^* & \text{if } C^* \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

$$\text{interaction} = \begin{cases} \text{true} & \text{if } E > E^* \\ \text{false} & \text{otherwise} \end{cases}$$

Performance of protein docking by MEGADOCK

Docking accuracy was comparable to conventional tools (Benchmark 4.0, 176 bound complexes)

8.8 times speedup compared to ZDOCK 3.0 (single node, single thread)

Success Rate

Number of Predictions

Total time for 100 docking [hr]

MEGADOCK

ZDOCK 3.0

ZDOCK 2.3

ZDOCK 3.0

ISLIMソフトウェア研究開発報告会

8

Scalability on K computer

Thread parallelization

One docking of proteins, FFT size: 140

MPI parallelization

Strong scaling > 0.95 on 24576 nodes

An exhausting docking of 220 proteins
Size of FFT: 140

ISLIMソフトウェア研究開発報告会

9

PPI prediction using general benchmark data

Binding partner prediction from 44x44=1936 dockings and post-processing based on highest docking score (Diagonal: interacting pairs)

F-measure : 0.42

$$F\text{-measure} = \frac{2 \cdot TP}{(TP + FP) + (TP + FN)}$$

PPI prediction by MEGADOCK achieved better than random performance on general benchmark dataset (monomer pair from protein-protein docking benchmark 2.0, Mintseris *et al*, *Proteins*, 2005.)

Matsuzaki Y, *et al.*, *J Bioinform Comput Biol*, 7:991-1012 (2009).
Ohue M, *et al.*, submitted.

10

Application to bacterial chemotaxis pathway

13 proteins (101 structures)

- True Positive (Red line)
- False Negative (Blue dashed line)
- False Positive (Purple dashed line)

F-measure : 0.44

Acceptable performance was shown on a real biology pathway reconstruction problem.

A suggestion of CheY-CheC interaction by using "False-positive" pair CheY-CheD as a mediator

Matsuzaki Y, *et al.*, *J Bioinform Comput Biol*, 7:991-1012 (2009).

ISLIMソフトウェア研究開発報告会

11

PPI prediction results (Apoptosis)

F-measure : 0.28

PPI prediction by docking without any other knowledge showed comparable results to template-based search of interaction partners (F-measure 0.30, Ozbabacan *et al.*, *J Struct Biol*, 2012).

	Interacting	No Interaction
Positive	88	364
Negative	96	1105

Ohue M, *et al.*, *Tech. Rep. IPSJ SIG*, 2012-BIO-32(13), 1-8, 2012.

12

Application to non-small cell lung cancer pathway

- Completed large-scale exhaustive docking
 - 497 structures, all-to-all docking = 247,009 structure pairs
- Achieved high PPI prediction performance
 - Precision 0.29
 - Recall 0.47
 - F-measure **0.36**

PPI prediction of about **250 thousand** structure pairs showed comparable performance to the application to bacterial chemotaxis (10 thousand pairs).

	Interacting	No Interaction
Positive	53	131
Negative	59	747

Counts are based on protein species

44 proteins (497 structures)

hsa05223

ISLiM

13

Interaction of cancer pathway and proteins related to Gefitinib

EGFR pathway related to non-small cell lung cancer

2,000 x 2,000 = 4 million docking

44 proteins (497 structures)

294 proteins (1424 structures)

Proteins related to Gefitinib estimated by Miyano lab., The Univ. of Tokyo, from microarray analysis

Searching novel cancer related PPIs

ISLiMソフトウェア研究開発報告会

ISLiM

References

(Papers contributed by MEGADOCK group)

- Ohue M, Matsuzaki Y, Ishida T, Akiyama Y. Improvement of the protein-protein docking prediction by introducing a simple hydrophobic interaction model: an application to interaction pathway analysis., *Lecture Note in Bioinformatics*, **7632**:178-187, 2012.
- Ohue M, Matsuzaki Y, Akiyama Y. Docking-calculation-based method for predicting protein-RNA interactions., *Genome Informatics*, **25**:25-39, 2011.
- Fleishman SJ, et al. Community-wide assessment of protein-interface modeling suggests improvements to design methodology., *Journal of Molecular Biology*, **414**:289-302, 2011.
- Uchikoga N, Hirokawa T. Analysis of protein-protein docking decoys using interaction fingerprints: application to the reconstruction of CaM-ligand complexes., *BMC Bioinformatics*, **11**:236, 2010.
- Ohue M, Matsuzaki Y, Matsuzaki Y, Sato T, Akiyama Y, MEGADOCK: an all-to-all protein-protein interaction prediction system using tertiary structure data and its application to systems biology study., *IPSJ Transactions on Mathematical Modeling and Its Applications*, **3**: 91-106, 2010.
- Matsuzaki Y, Matsuzaki Y, Sato T, Akiyama Y. *In silico* screening of protein-protein interactions with all-to-all rigid docking and clustering: an application to pathway analysis., *Journal of Bioinformatics and Computational Biology*, **7**:991-1012, 2009.

(Submitted)

- Matsuzaki Y, Ohue M, Uchikoga N, Akiyama Y. Protein-protein interaction network prediction by using rigid-body docking tools: application to bacterial chemotaxis., *submitted*.
- Uchikoga N, Matsuzaki Y, Ohue M, Hirokawa T, Akiyama Y. Improved post-processing of protein-protein docking data using profiles of interaction fingerprints., *submitted*.
- Ohue M, Matsuzaki Y, Uchikoga N, Ishida T, Akiyama Y, MEGADOCK: An all-to-all protein-protein interaction prediction system using tertiary structure data., *submitted*.

ISLiMソフトウェア研究開発報告会

ISLiM

15

Acknowledgement

- This work was supported in part by a Grant-in-Aid for Research and Development of The Next-Generation Integrated Life Simulation Software from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT).
- Part of the result was obtained by early access to the K computer at the RIKEN Advanced Institute for Computational Science.

Masahito Ohue
Kohei Yamamoto
Takayuki Fujiwara
Takehiro Shimoda

ISLiMソフトウェア研究開発報告会

ISLiM

16

生命体シミュレーションのための データ同化技術の開発

樋口知之

次世代計算科学研究開発プログラム
データ解析融合研究開発チーム



発表者紹介

- 1989年 3月 東京大学理学系研究科地球物理学博士課程修了
- 1989年 4月 統計数理研究所予測制御研究系予測理論研究部門助手
- 1994年 12月 統計数理研究所予測制御研究系予測理論研究部門助教授
- 2002年 7月 統計数理研究所予測制御研究系システム解析研究部門教授
- 2004年 4月 統計数理研究所モデリング研究系教授, 予測発見戦略研究センター副所長(兼務)(~2010.3)
- 2011年 4月 統計数理研究所 所長

研究分野

ベイジアンモデリング, データ同化



ISLiM

統計数理研究所
The Institute of Statistical Mathematics

生命体シミュレーションのためのデータ同化技術の開発


統計数理研究所・所長
樋口知之

統計数理研究所・データ同化研究開発センター
吉田亮 中野慎也 長尾大道 斎藤正也





データ同化
研究開発センター



ISLiM

統計数理研究所
The Institute of Statistical Mathematics

I. ソフト研究開発の背景・目的


モデリングの制約

- タンパク質相互作用や転写因子データベースにもとづく反応経路列挙型モデリングの限界(組み合わせ爆発)
- モデリングの対象である分子リストが不完全なため、閉じた系として記述できない


実験の制約

- 細胞の環境依存性や生体内分子のバリエーションならびに個体間の遺伝的差異等がもたらす生化学反応の多様性
- 個体ごとに異なる生化学パラメータやネットワークポロジ-

実験とモデリングの知識循環から永続的なモデル改良が必要である



データ同化
研究開発センター




ISLiM


統計数理研究所
The Institute of Statistical Mathematics

II. ソフト研究開発の概要・アプローチ

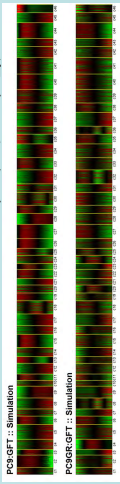
生命体内分子の計測データをもとに、シミュレーションの再現性・予測力を改善するためのパラメータチューニングやモデルの改良をベイズ推定により行う。



実験・計測




モデル設計
評価と選択
モデルの再構成




シミュレーション

仮説構築
薬剤応答性の予測
実験のデザイン



データ同化
研究開発センター



ISLiM

統計数理研究所
The Institute of Statistical Mathematics

状態空間モデル

● システムモデル - シミュレーションモデルのマルコフ過程としての表現

$$\mathbf{x}_n = \mathbf{f}_G(\mathbf{x}_{n-1}, \mathbf{q}) + \text{noise}$$

● 観測モデル

- 状態変数と観測データの関連付け

$$\mathbf{y}_n = \mathbf{g}(\mathbf{x}_n) + \text{noise}$$

G

モデル

→

$f_G(\cdot | \mathbf{q})$

パラメータ

→


$\mathbf{x} = \{\mathbf{x}_n\}$

状態変数

→

$\mathbf{D} = \{\mathbf{y}_n\}$

データ



データ同化
研究開発センター

ISLIM 統計数理研究所 The Institute of Statistical Mathematics

バイズ推定 - 逆問題を解く

事後分布 - データDを観測したもとの ω の確からしさ

$$p(w|D) \propto p(D|w)p(w)$$

ω : モデルG, パラメータ θ , 状態 x のどれでも良い

データDから未知の ω を推定する

(国立遺伝学研究所・木村暁氏より提供)

データ同化 研究開発センター

ISLIM LiSDASにおける細胞内生化学反応のモデル

タンパク質相互作用 ~ 100 sec

タンパク質の拡散 ~ 10分 ~ 10時間

タンパク質の分解 ~ 10分 ~ 10時間

翻訳 ~ 30 min

転写 ~ 30 min

mRNAの寿命 ~ 10 min to over 10 h

転写因子の結合 (転写を促進または抑制する)

遺伝子

ISLIM 統計数理研究所 The Institute of Statistical Mathematics

定式化

RNA・タンパク質等の物質量の時間変化に関する常微分方程式

$$\frac{dx_j}{dt} = -d_j k_j x_j + \sum_{a \in A(j)} h^A(x_d | c_a, p_a) \times \prod_{i \in I(j)} h^I(x_i | d_i, q_i) + b^k \prod_{b \in B(j)} x_b$$

- $A(j)$: 物質 j の発現を促進する物質 (Activators)
- $B(j)$: 物質 j の発現を抑制する物質 (Inhibitors)
- $I(j)$: 物質 j の発現を抑制する物質 (Inhibitors)
- A と B は物質 j がmRNAのときに空でない。
- ヒル関数 $h^A(x | c, p) = \frac{c^A}{c^A + \sum_i x_i^p}$, $h^I(x | d, q) = \frac{d^I}{d^I + \sum_i x_i^q}$
- $B(j)$: 物質 j が複合体である場合に、それを構成する物質(Binding)
- 推定するモデルパラメータ $\theta = (x_j(0); a^k, c^k, b^k, d^k, p^k, q^k)$
- 広い範囲の常微分方程式系に適用可能

ISLIM 統計数理研究所 The Institute of Statistical Mathematics

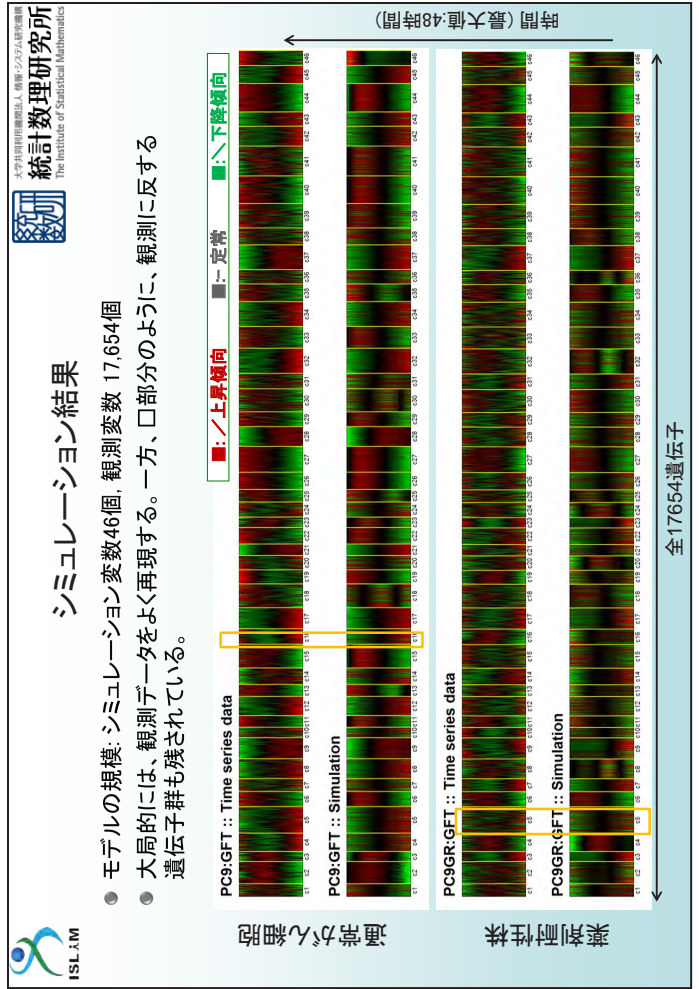
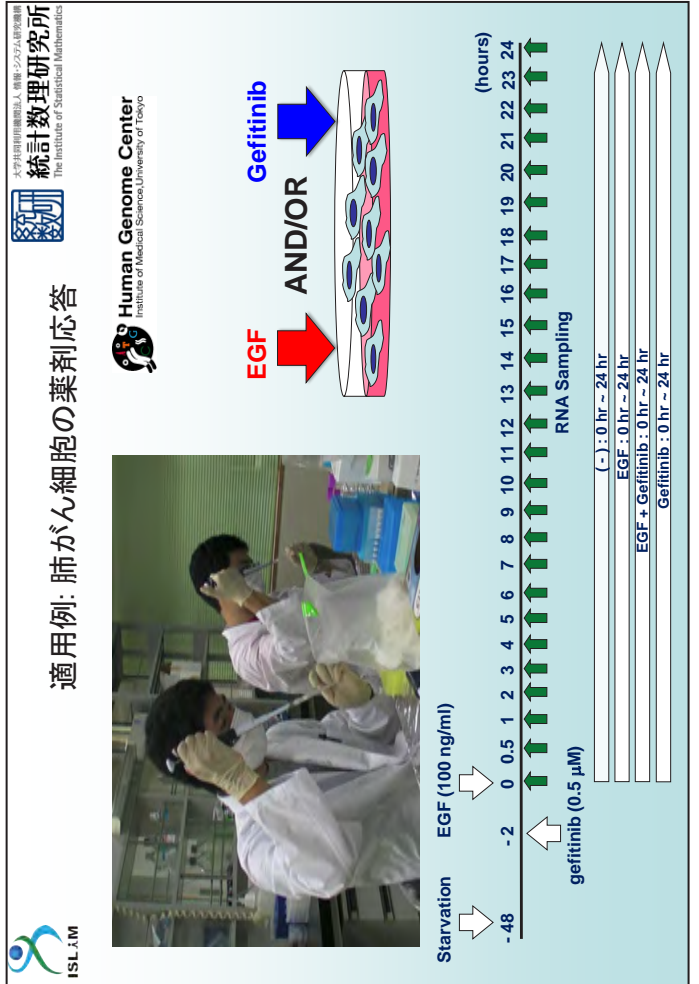
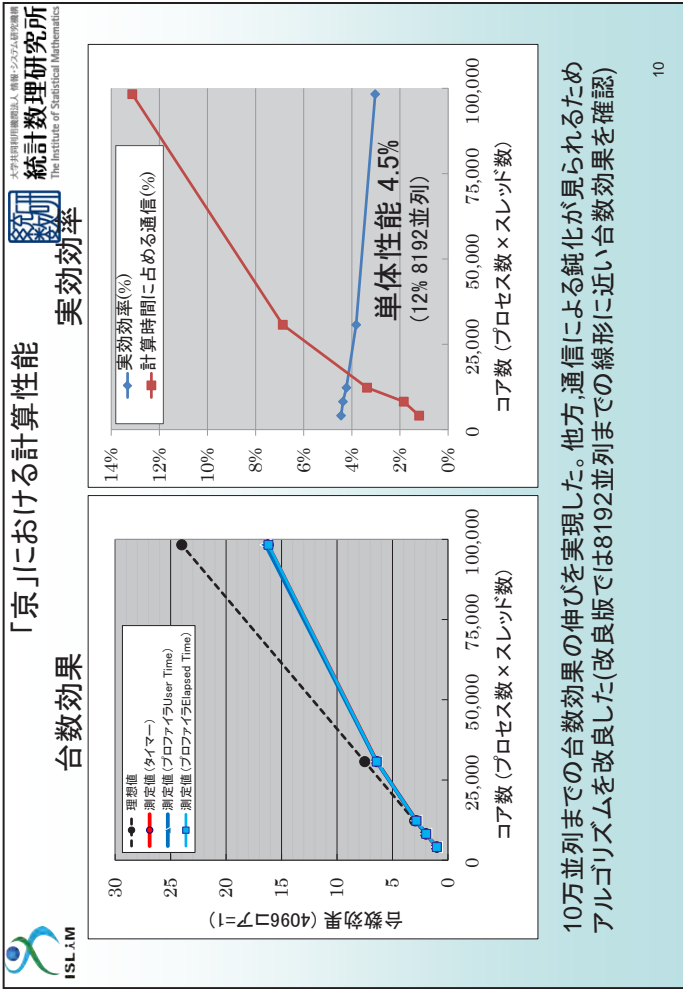
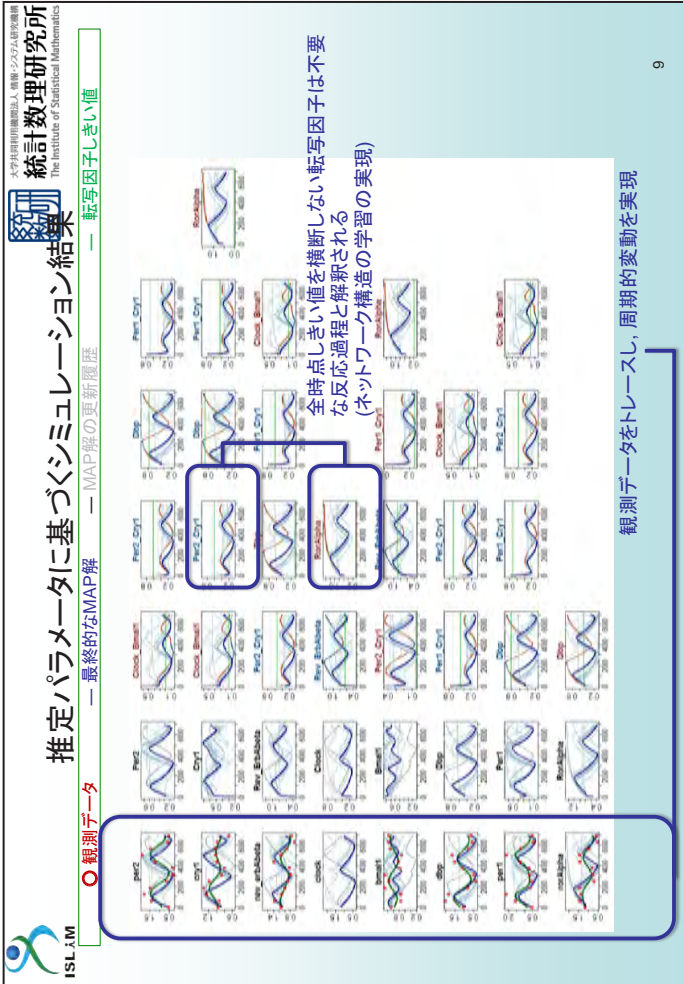
III. 現時点でのソフト研究開発成果

テスト問題: 哺乳動物の概日周期転写ネットワーク

- モデルの規模
- システム変数の個数: 19
- 観測変数の個数: 7
- モデルパラメータ数: 116

Reagent/par2: Description: Transcription; ReagentID: 1
Regulatory elements in promoter: I/F box, D box
Activators: Bim1/Clack, Zip1/Nyct2
Inhibitors: Per1/Cy1, Per2/Cy2, Per3, Dec1

Reagent/par1: Description: Transcription; ReagentID: 2
Regulatory elements in promoter: I/F box, D box
Activators: Bim1/Clack, Zip1/Nyct2
Inhibitors: Per1/Cy1, Per2/Cy2, Per3, Dec1



ISLIM

データ同化ライブラリ適用事例: 線虫初期胚における細胞内流動データ同化 (国立遺伝研 木村研究室(細胞建築学)との共同研究)

prior $p(\theta)$ → likelihood $p(y_{1:T} | \theta)$ → posterior $p(\theta | y_{1:T})$

Model parameters
 • velocity of the cell wall $u_{wall}(L, t)$
 • viscosity coefficient
 • density

EnKF
 • search model parameter maximizing the likelihood
 • observed velocity field

MCMC
 • obtain posterior distribution (multi peaked pattern?)

Fig. 1 線虫初期胚の細胞質流動の模式図

ISLIM

ソフトウェア公開状況

- ISLIM サイト内で5月7日にβ版を公開
<https://www.islim.org/FS-APL/FS-Form/form.cgi?Code=Code24J>
- 年度内に完成版を公開予定
- 開発中の付属ツール
 - SBML(to/from)LiSDAS
 - データ同化ライブラリ

ISLIMソフトウェア研究開発報告会

データ同化
研究開発センター

ISLIM

データ同化による細胞壁における力分布推定

細胞質流動を駆動するメカニズムの仮説:
細胞壁層におけるミオシンが後ろから前へ移動する時細胞質を引きずる。この引きずりは細胞壁層でのみ起こるが、細胞質の流体力学的性質によって細胞全体に流れが波及して細胞質における流れが生じる。

仮説の検証方法:
仮説をもとにしたシミュレーションに実験で計測された流動の速度場を再現するかどうかを検証する。
番号1~3072の各シミュレーションで、速度ベクトルの差の2乗の和 $(V_{i,t} - V_{sim})^2$ を計算
→ $Pos(1), Pos(2), \dots, Pos(3072)$ が得られる; $Pos(i)$ は、番目の力分布の事後確率

実験

データ同化

ポックス:
黒線: 実験値
赤線: データ同化

速度方向の速さ ($\mu\text{m/s}$)

距離方向の速さ (距離 from the posterior pole (x , μm))

長軸方向の位置

ISLIMソフトウェア研究開発報告会

データ同化
研究開発センター

ISLIM

まとめ

- 転写制御回路モデルのパラメータ推定を対象にLiSDASを開発した。年度内の周辺ツールを含めた完成版の公開に向けての作業を継続中である。
- 常微分方程式で与えられたモデルであれば、そのまま若しくは少しの拡張で適用できる。つぎのようなLiSDASを活用する課題に取り組み予定である。

インフルエンザ感染ネットワーク
井元・山口・上・宮野 (東大医科研)
斎藤・樋口 (統数研)

化合物分子設計
山下(統研大)
吉田・伊庭・樋口(統数研)

細胞体流動
長尾・斎藤 (統数研)
木村(国立遺伝研)

神経細胞ネットワーク
吉田 (統数研)
飯野 (東京大)
石原 (九州大)
岩崎 (茨城大)

ISLIMソフトウェア研究開発報告会

データ同化
研究開発センター

臓器全身スケール研究開発チーム成果報告

臓器全身スケール研究開発チーム 成果総括報告

高木 周

次世代計算科学研究開発プログラム
臓器全身スケール研究開発チーム チームリーダー



発表者紹介

1995年3月 東京大学大学院工学系研究科機械工学専攻博士課程修了
1998年4月 東京大学大学院工学系研究科 講師
2001年4月 東京大学大学院工学系研究科 助教授
2007年4月～現在 理化学研究所次世代計算科学, チームリーダー
20010年4月～現在 東京大学大学院工学系研究科 教授

研究分野

計算生体力学, 流体力学, マルチスケール計算力学

臓器全身スケール研究開発チーム成果総括報告

高木 周

次世代計算科学研究開発プログラム

臓器全身スケール研究開発チーム チームリーダー

1. 目的

臓器全身スケール研究開発チームでは、CT、MRI、超音波画像装置等により取得された各個人の人体ボクセルデータを基にして、コンピュータ上に人体を再構築し、各種病態の予測や治療の支援を行なうためのシミュレーションツールの開発を行うことを目的としている。特に、日本人の死因で大きな比重を占める癌の治療、循環器系疾患に関連する計算手法の開発に重点を置き、病態のメカニズムの解明と薬効の評価、低侵襲治療法の支援を行なうためのシミュレーションツールを開発している。

2. 現時点での成果

2.1 医用画像データに適した計算手法の開発（オイラー型流体構造連成手法: ZZ-EFSI）

本研究では、従来は境界適合メッシュを用いた有限要素法による解析が主流であった流体・構造連成問題に対して、メッシュ生成のプロセスなしに画像データから直接計算を行なうオイラー型流体構造連成手法を新たに開発した。本手法は、医用画像データのみならず、一般に画像データを基にして時々刻々形状が変化していくような問題全般に適用可能であり、固定された構造格子の上を自由に變形する物体の運動を計算することができ超並列の計算に特に適した定式化を行っている。すでに、フルコアを用いた計算で実効性能 4.5Peta Flops の計算速度を記録している。

2.2 血小板上と血管壁のタンパク質間相互作用と流れの計算を連成させた血球細胞・流体連成手法（マルチスケール血栓シミュレータ: ZZ-Throm）

ZZ-EFSI 法を膜と流体の連成問題へまで発展させ、赤血球や血小板などの血球細胞を大量に含む血流の計算を可能にさせた手法である。さらに、血小板上の糖タンパク GP-1b α と損傷血管壁上の vWF 分子との分子間相互作用である Receptor-Ligand 反応をモンテカルロ法により計算することにより、分子間結合による吸着力と、流体運動により引き剥がす流体力を連成させながら解く計算手法を考案した。赤血球の有無が、血小板吸着に与える影響について調べ、血小板吸着にとって赤血球がもたらす血漿側のゆらぎが本質的な

影響を与えることを再現するのに成功している。さらに、赤血球の体積率を変化させ、実験データとの定量的比較を行っている段階。

2.3 心筋細胞の詳細な挙動から心臓全体を再現する心臓シミュレータ（マルチスケール・マルチフィジックス心臓シミュレータ: UT-Heart）

ISLiM プロジェクトで開発されたプログラムの中で唯一、「京」の優先課題として選ばれたソフトウェア。東京大学、久田らにより 10 年以上の年月をかけて開発されてきた心臓シミュレータである。このシミュレータでは、細胞膜のイオンチャンネル、ポンプ、トランスポーターなどの作用に基づく細胞内イオン環境のダイナミクスを記述する電気生理学モデル、並びにイオンによって制御される細胞内収縮タンパクであるアクチン・ミオシン間のクロスブリッジ運動を記述する興奮収縮連関モデルから出発し、細胞、組織、臓器を経て最終的な血液拍出に至る心臓機能の全過程を各レベルにおける要素間の相互作用を含めて合理的に数理化し、心臓の振る舞いを再現できる。本年度は、サルコメ \leftrightarrow 心筋細胞 \leftrightarrow 心臓全体の 3 レベル階層統合に成功し、「京」での超大規模計算に成功した。

2.4 国産初の治療器開発に向けた超音波治療シミュレーション（強力集束超音波シミュレータ: ZZ-HIFU）

患者個々の人体ボクセルデータを用いて強力集束超音波(HIFU)により腫瘍焼灼を行なうシミュレーションを行う方法の開発を終えた。現時点において、98304 コアを用いた計算で、実行性能 21%を達成している。現在、国産初の超音波治療器を目指して、国内の医療機器メーカーと連携して乳腺治療を対象として実機的设计支援のためのシミュレーションを開始している。具体的には、実際の乳房の画像データを基にし、焦点制御のシミュレーションを実施し、位相制御の重要性を確認している。

2.5（モンテカルロ重粒子線治療シミュレータ: ZZ.DOSE）

重粒子線治療の手術計画に利用可能なモンテ

カルロ計算によるシミュレータを開発している。現時点では、プログラムのハイブリッド並列化が済み、大規模計算の準備をしている。また、理研の仁科加速器センターとの共同研究により、計算結果の検証用の実験データを取得している。現時点では、計算負荷の軽い系に対して、実験との比較を行い、モデルの検証を進めている。

2.6 その他 (0D-1D-3D 全身血管網シミュレータ: ZZ-Vasc)

大規模3次元血流計算に対する境界条件を与えるための計算を行なうソフトウェアである。現時点では、東京大学の島まりの研究グループと連携をとり、動脈硬化を起こした総頸動脈および脳血管系であるウィルス動脈輪に対して、患者の画像データから得られた情報を基に3次元計算と0D-1D計算の連成を行い、術後の血流量の分配予測のシミュレーションなどを行っている。

3. プロジェクト終了時の達成目標

1) ZZ-EFSI

上述の通り、ZZ-EFSIはすでに京のフルノード約65万コアを用いた計算で実行性能4.5Peta Flopsを達成している。プロジェクト終了時までには、ソフトウェアの最新版の公開と随時更新を行なっていく。本手法は、医療応用だけでなく、「ものづくり」分野も含めた幅広い分野への適用が可能なソフトであり、公開を通して多くのユーザーを募り、プロジェクトの成果の有効活用を図る。

2) ZZ-Throm

損傷血管壁への血小板の吸着に伴う、血栓成長の初期過程を再現するシミュレータを完成させる。これにより、分子レベルでのReceptor-Ligand結合から、多数の変形赤血球・血小板を含む大規模な流動構造までを連成して解くマルチスケールシミュレーションを達成し、Receptor-Ligandの分子間結合が血栓成長に与える影響を調べる。また、東海大学医学部、後藤らにより取得される実験データとの比較により、計算結果の妥当性を検証し、さらに実験データと計算結果の相補的な利用により実験で解析が困難な膜上における糖タンパクGP1b- α の動的振る舞いなどについて知見を得る。

3) 心臓シミュレータ (UT-Heart)

筋原線維を構成するアクチン・ミオシンフィラ

メントにおける機能タンパク個々の運動やそれを支えるZ帯などの微小構造を具体的にモデル化し、そこから心臓の拍動までをシームレスに接続し、ミクロとマクロの因果関係を明らかにするためのシミュレータ開発をさらに進めていく。このようにタンパク質レベルの確率的運動、複合体としての細胞レベルの収縮運動、および臓器レベルの拍動運動の3つのスケールを合理的に結び付けるシミュレーションを達成することにより、心筋細胞内で起こる分子レベルの物質輸送から心筋の収縮を再現し、新たな医学的・生理学的知見を得ることが可能となる。

4) ZZ-HIFU

国産初の治療器の設計に向けた大規模計算を引き続き行う。本年度は実際に乳房のデータを基にして計算を実施しており、このデータに対してさらにパラメータを振った計算を進める。特に、治療の対象とする部位への超音波集束の制御を考慮した場合に必要な超音波パネル(アレイトランスデューサー)の数と配置、リアルタイムモニタリングの方法について、実験データとの比較を行いながら機器設計を進めて行くシステムを構築する。これにより、大規模計算の結果を利用して、機器の設計および実機へ搭載していくためのソフトウェアの開発を進めることができる。



5) ZZ-DOSE

仁科加速器センターで新たに取得される実験データにより、モンテカルロ法による計算結果の比較を行い、開発された計算手法の妥当性の検証を行う。加えて、「京」での計算に関して、ハイブリッドプログラムの改良を行い、並列化効率を上げる。

6) ZZ-Vasc


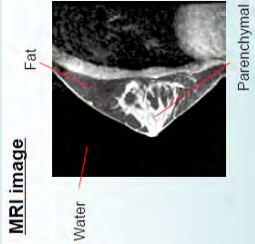
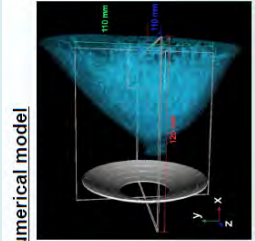
本年度構築された0D-1D-3Dの連成手法に関して、患者の血管データを使用したシミュレーションを引き続き実施し、計算結果の検討および連成手法の安定性の検証を行う。ZZ-VASCは、3次元医用画像データを用いた計算との連成以外にも、0D-1D連成の部分を用いてさまざまな共同研究を行ってきた。その一つとして、心臓周りの動脈硬化を調べる新たな電子血圧計の開発に協力し製品化に成功したが、さらに、シミュレーションの結果を用いて、新たな動脈硬化判定指標の提案を行う。



以上

臓器全身スケール研究開発チーム 成果総括報告

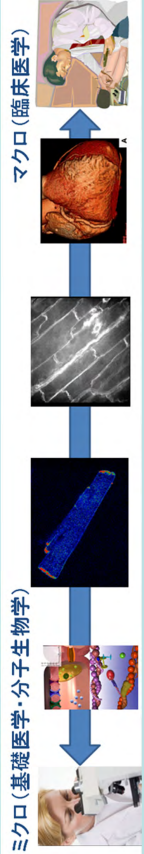
理化学研究所
次世代計算科学研究開発プログラム
臓器全身スケール研究開発チーム チームリーダー
高木 周



臓器全身スケールチーム：概要

- チーム全体目標
 - 細胞スケール、分子スケールの効果を取り入れたマルチスケール生体力学シミュレーションにより、病態の予測や治療の支援を可能とするソフトウェアの開発を行う。
- 個別テーマでの主立った成果
 - 心臓シミュレータ UT-Heart: 「京」での計算により「サルコメア ↔ 心筋細胞 ↔ 心臓全体」の3階層統合に世界で初めて成功。
 - 流体構造連成 ZZ-EFSI, 血栓シミュレータ ZZ-Throm: 新しい流体構造連成手法を開発。「京」での計算により世界最速 4.5ペタフロップスを達成。血栓の初期過程のマルチスケールシミュレーションを実施。赤血球の存在が血小板の壁面粘着を促進させることを再現。
 - 超音波治療シミュレータ ZZ-HIFU: 国産初の超音波治療器の開発に向けて、乳がん治療器設計のシミュレーションを開始。
- 今後の予定
 - ソフトウェアの整備と「京」での本格的計算の継続。



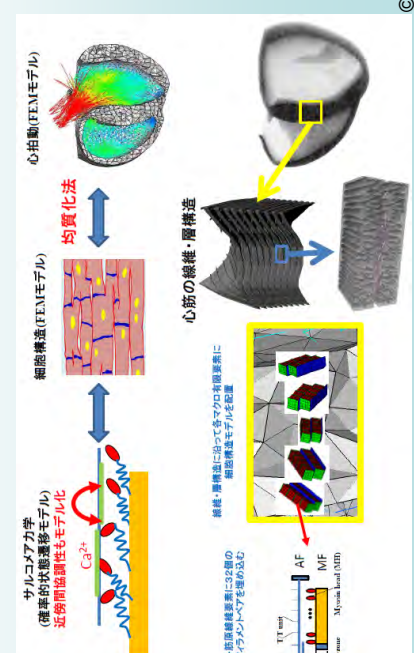
マイクロ(基礎医学・分子生物学) マクロ(臨床医学)

© 久田(東大)

マルチスケール・マルチフィジックス心臓シミュレータ (UT-Heart)



- UT-Heartは既に臨床研究に用いられている。
- グラウンドチャレンジで開発されたソフトのうち、唯一、「京」利用の優先課題となる。
- 「京」モデルでは、サルコメアレベルの運動から心筋細胞の機能再現、さらに心臓全体と3段階の異なるスケールの階層統合に成功。



心拍動(FEMモデル)
均質化法
細胞構造(FEMモデル)
心筋の繊維・層構造

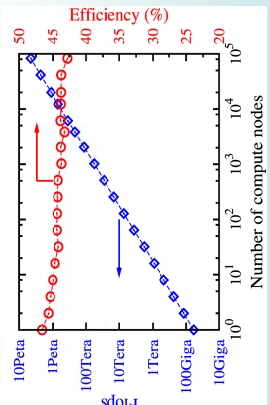
各階層要素数: 32階のフラクタルペンを組み込む
UT 100k, AF 100k, WF 100k, Myocardium 100k

© 鷲尾, 久田(東大)

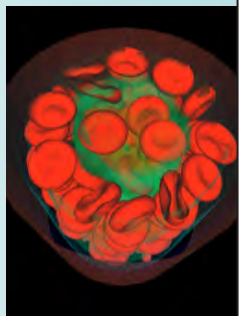
医用画像データに適した流体構造連成手法(ZZ-EFSI)

- 「京」のフルノード計算で、実効速度 4.5ペタフロップスを達成(世界最速の流体構造連成手法ZZ-EFSIの開発に成功)



Efficiency (%)

Number of compute nodes



マルチスケール血栓シミュレータ(ZZ-Throm) (1/4)

ISLAM RIKEN

血小坂表面のGP1b α 分子と血管壁上のvWF分子のタンパク質相互作用に状態遷移理論を適用したモデルにより、血流計算との連成に成功。血栓接着のよりリアルスティックな計算が可能となり、赤血球の存在が血小板接着には重要とする実験的事実をシミュレーションにより実証。

血流 (連続体力学)

スケール間相互作用

Ligand-Receptor結合 (Stochastic Process)

分子間相互作用 (分子動力学)

- 物質拡散
- 代謝反応・活性
- 形態変化

マルチスケール血栓シミュレータ(ZZ-Throm) (2/4)

ISLAM RIKEN

PRESENT SIMULATION

EXPERIMENT: Gaetgens *et al.* (1980) *Blood Cells.*, 6, 799.

- axial migration of RBCs \rightarrow cell free layer near vessel wall
- parachute/slipper shape

without RBCs

with RBCs ($Ht = 20\%$)

壁 ← axis →

血小板位置の時間変化(半径方向) (赤血球ありで血小板のゆらぎ大)

マルチスケール血栓シミュレータ(ZZ-Throm) (3/4)

ISLAM RIKEN

With RBC (WLC Model)

Without RBC

マルチスケール血栓シミュレータ(ZZ-Throm) (4/4)

ISLAM RIKEN

狭窄部の形状の影響

超音波治療シミュレータ : ZZ-HIFU (1/3)

ISLIM

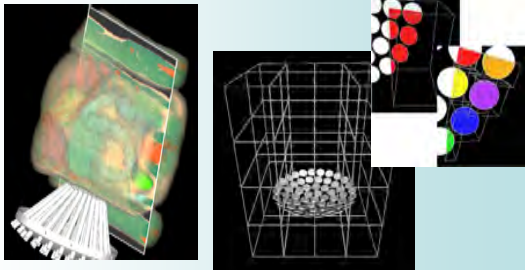
概要

- 強力集束超音波 (HIFU) 治療機器から照射された超音波が人体内(骨や臓器など音響特性の異なる多媒質体内)を伝播する様子をシミュレーションするためのアプリケーション
- 2HMzの超音波の1波長0.75mmを15格子($\Delta x=0.05\text{mm}$)で解像する場合、450x300x3000mmの領域に対して9000x6000x6000=340G格子点が必要であり、そのような大規模入力データの作成から超大規模並列計算が可能

・ アプローチ

- 多媒質に対する質量、運動量の保存式を粘性流体もしくは粘弾性体に対する構成方程式および媒質の状態方程式と伴に解くことで、不均質媒体中の超音波伝播を再現し、粘性散逸を熱源とする温度の時間発展式により温度場を得る。
- 有限差分法によって基礎方程式を近似
 - ・ 空間2m次の中心差分(問題によって精度を選択)
 - ・ FDTD法に準じた時間発展(1or2次精度を選択)
- 並列化手法
 - ・ 領域分割とMPIによるプロセス並列
 - ・ OpenMPを用いたスレッド並列によってマルチコアに対応

*Okita et al., Int. J. Numer. Meth. Fluids 2011, 65:43-66



9

超音波治療シミュレータ : ZZ-HIFU (2/3)

ISLIM

乳腺疾患治療用256chアレイトランスデューサの開発支援
(国産初の超音波治療の開発に向けて)

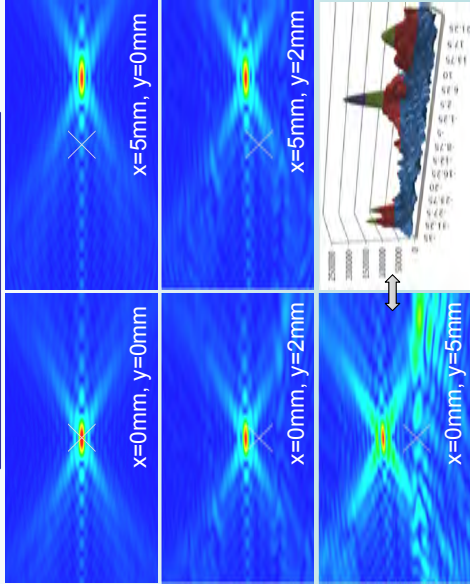
TSBMI*における低侵襲診断治療機器の開発との共同

焦点位置による超音波音場の変化

- ・ 超音波で診ながら超音波で治療する治療器の設計に向けたシミュレーション
- ・ 中央部に診断プローブ用ホールを準備する必要あり。
- ・ チャネル数、配置をシミュレーションにより検討

256ch, 2MHz, F=0.8

*システム疾患生命科学による先端医療技術開発拠点

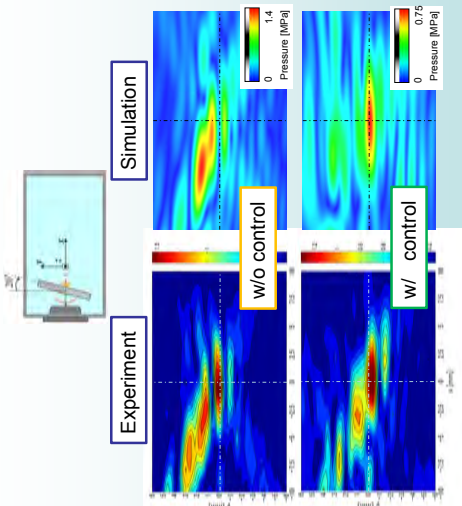


10

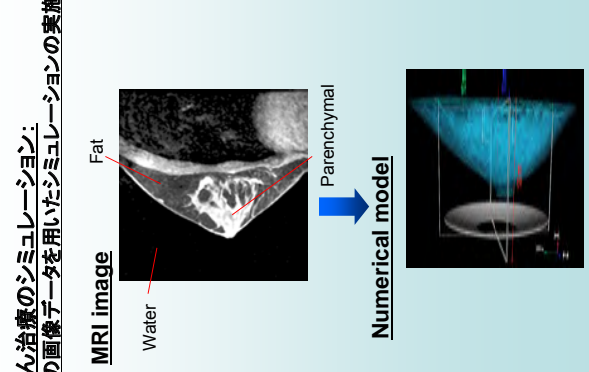
超音波治療シミュレータ : ZZ-HIFU (3/3)

ISLIM

シミュレーション結果を用いた実験による検証:
形状と物性が既知のものに対するシミュレーションの信頼性を確認



**乳がん治療のシミュレーション:
乳房の画像データを用いたシミュレーションの実施**

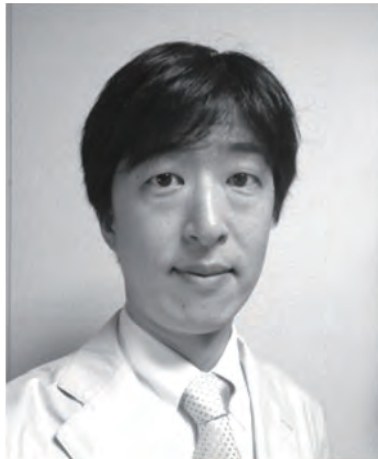


11

次世代超音波治療機器開発のための 生体内超音波伝播シミュレータ (ZZ-HIFU)

沖田 浩平

日本大学生産工学部機械工学科 准教授
理化学研究所 次世代計算科学研究開発プログラム
臓器全身スケール研究開発チーム 客員研究員





発表者紹介

2002年3月 大阪大学工学研究科機械物理工学専攻博士課程修了
2002年4月 東京大学 IML 研究員
2004年4月 東京大学 21世紀 COE 機械システムイノベーション 特任研究員
2006年4月 理化学研究所 VCAD システム研究プログラム 研究員
2007年4月 理化学研究所 VCAD システム研究プログラム 上級研究員
2011年4月 東京大学大学院工学系研究科機械工学専攻 特任研究員
2011年9月 日本大学生産工学部機械工学科 准教授

研究分野

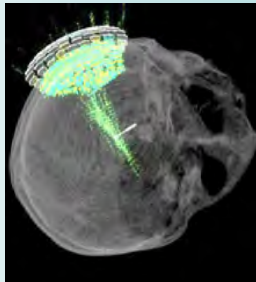
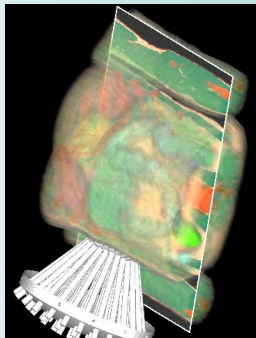
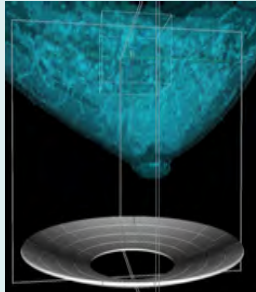
流体工学, 数値流体力学, 混相流







次世代超音波治療機器開発のための 生体内超音波伝播シミュレータ (ZZ-HIFU)

日本大学 生産工学部 機械工学科 准教授
 理化学研究所 次世代計算科学研究開発プログラム
 臓器全身スケール研究開発チーム 客員研究員
 沖田浩平

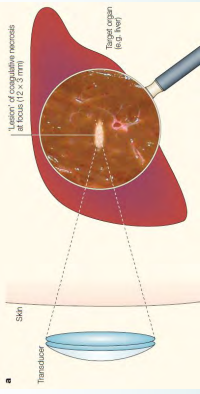
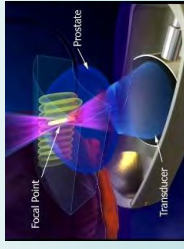
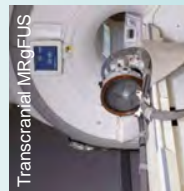
鳴見竜太(東大工), 東隆(東大工), 高木周(東大工), 松本洋一郎(東大工)



High-Intensity Focused Ultrasound therapy


Kennedy JE., Nat. Rev. Cancer, 5:321-27, 2005



MRI-guided Focused Ultrasound Surgery

has been developed for various tumors.

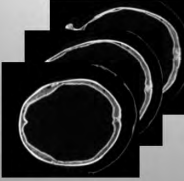
- Uterine fibroid
- Cancer in Brain, Breast, Liver etc.
- Sonothrombolysis, Neuromodulation, BBB, Parkinson's disease
- DDS with US
- Pre and Post contrast imaging for treatment validation
- Accurate targeting
- Real-time monitoring of temperature

国産の超音波治療機器開発を促進するために超音波伝播シミュレーションによる支援




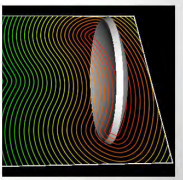
HIFU simulation based on digital human model



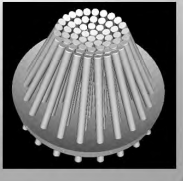
CT/MRI



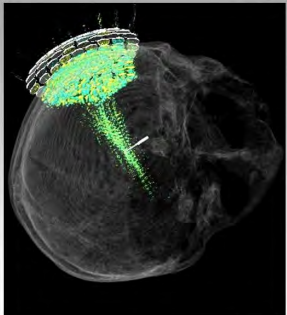
Voxel Phantom



SDF






CAD



HIFU Simulation

生体情報および設計形状を基にした入力データ作成から、それらを利用した超大規模並列計算による生体内超音波伝播シミュレーション

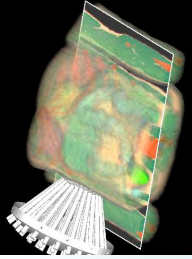







ZZ-HIFU

• 概要

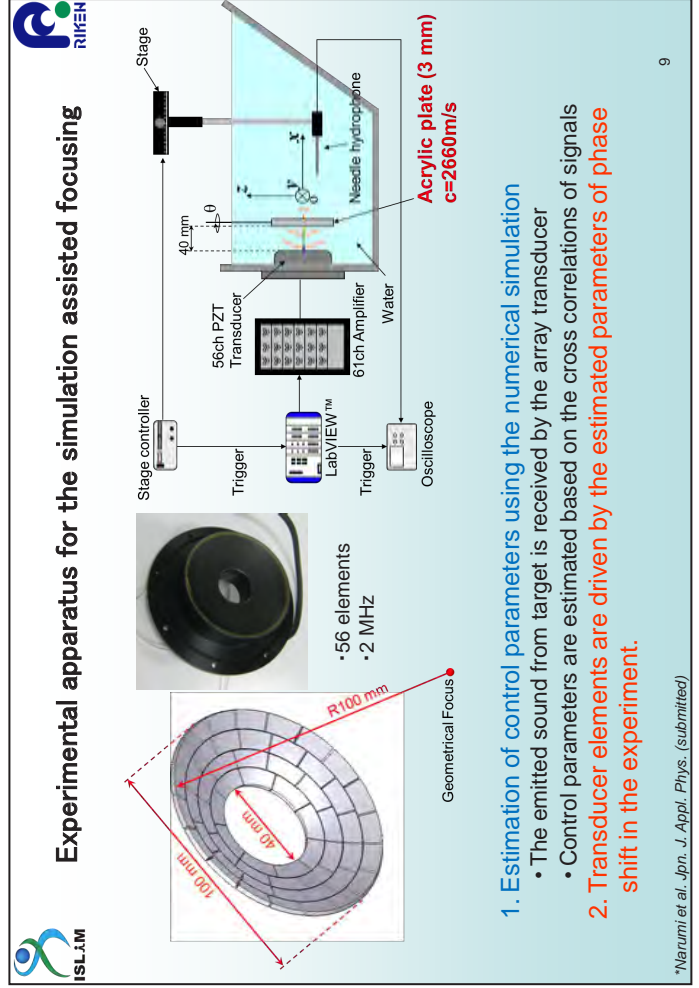
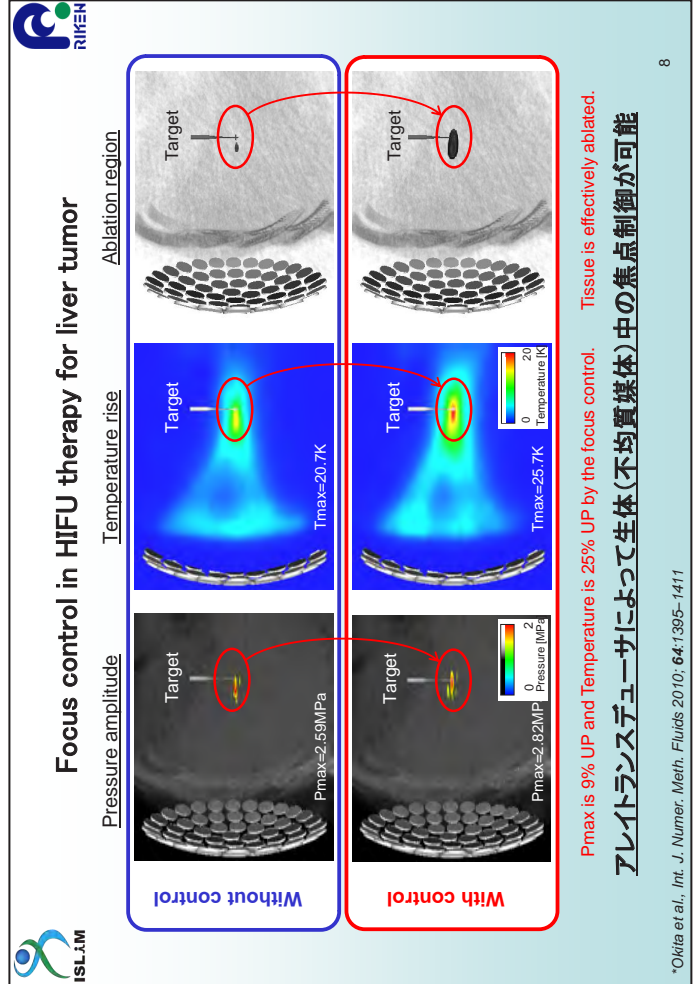
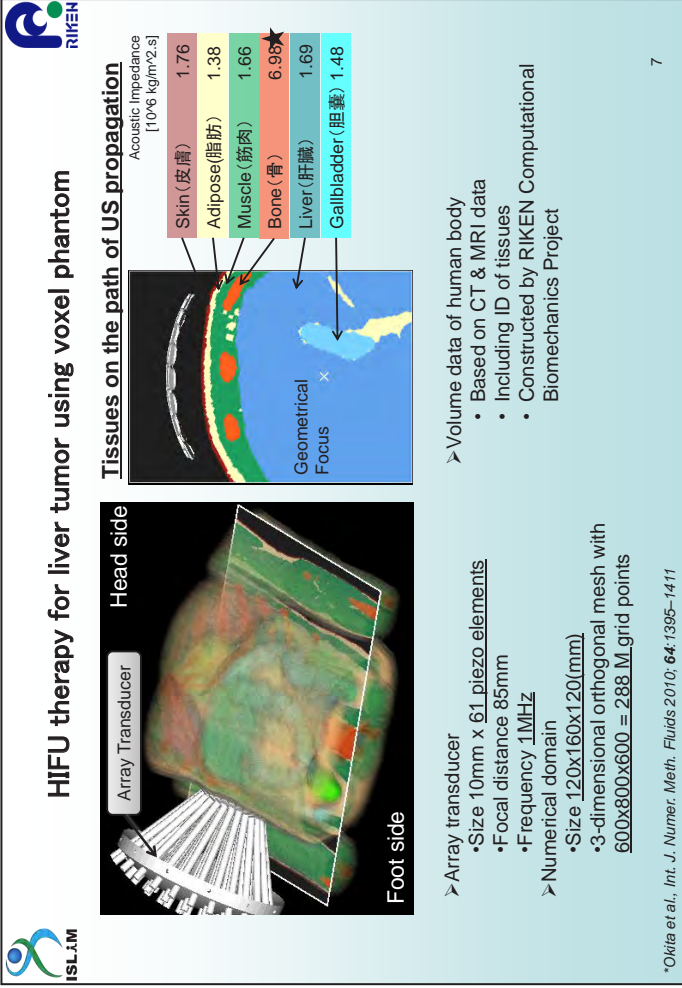
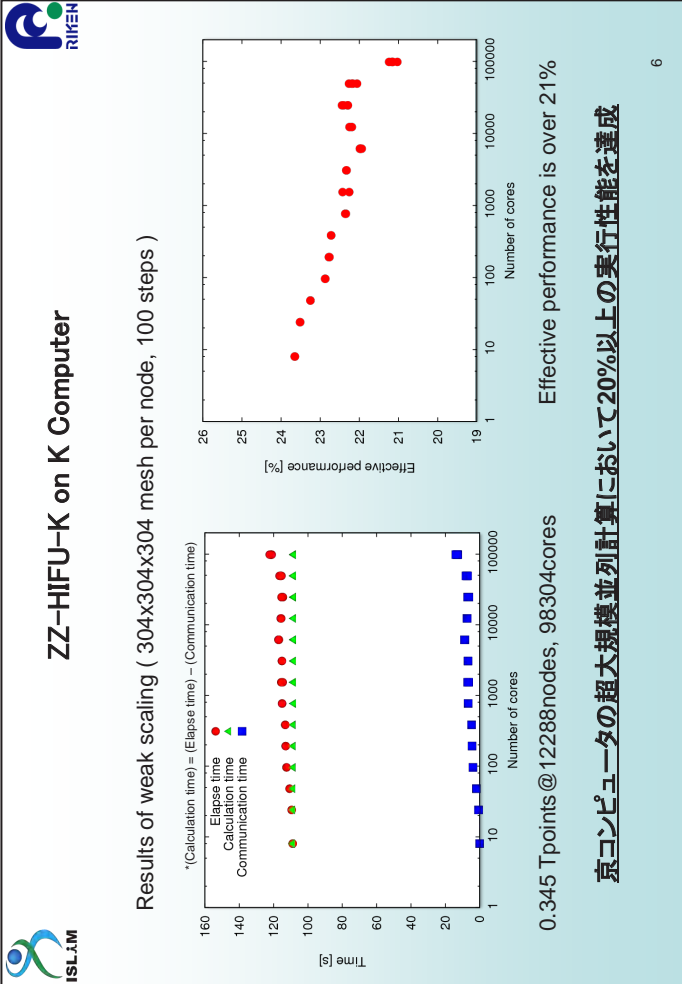
- 強力集束超音波 (HIFU) 治療機器から照射された超音波が人体内(骨や臓器など音響特性の異なる多媒質体内)を伝播する様子をシミュレーションするためのアプリケーション
- 2MHzの超音波の1波長0.75mmを15格子($\Delta x=0.05mm$)で解像する場合、 $450 \times 300 \times 300mm$ の領域に対して $9000 \times 6000 \times 6000 = 340G$ 格子点が必要であり、そのような大規模入力データの作成から超大規模並列計算が可能
- アプリケーチ
- 多媒質に対する質量、運動量の保存式を粘性流体もしくは粘弾性体に対する構成方程式および媒質の状態方程式と併に解くことで、不均質媒体中の超音波伝播を再現し、粘性散逸を熱源とする温度の時間発展式により温度場を得る。
- 有限差分法によって基礎方程式を近似
 - 空間2m次の中心差分(問題によって精度を選択)
 - FDTD法に準じた時間発展(1or2次精度を選択)
- 並列化手法
 - 領域分割とMPIによるプロセス並列
 - OpenMPを用いたスレッド並列によってマルチコアに対応







*Okita et al., Int. J. Numer. Meth. Fluids 2011; 65:43-66



Results of the simulation assisted focusing

Simulation assisted focusing results showing pressure distribution and focal spot control.

形状と物性が既知のものに対するシミュレーションの信頼性を確認

シミュレーションで得られた制御パラメータを実際にトランスデューサで実験した結果、弾性体の影響を考慮したシミュレーションによって、焦点制御の精度が向上

*Narumi et al. Jpn. J. Appl. Phys. (submitted)

乳腺疾患治療用次世代超音波治療機器の開発支援

TSBIMI[®]における低侵襲診断治療機器の開発との共同

- 女性の罹患率では乳がんがトップ
- 皮膚表面も傷をつけない温存療法
- 治療機器開発のはじめのターゲットとして、超音波的にそれ程複雑ではない部位(反射, 屈折など)
- MRgHIFUに代わる世界初のUS-CTgHIFUの開発
 - 3D画像でのモニタリング
 - ビームパス表示
 - リアルタイム温度モニタ
 - 治療直後の治療効果判定

MRIとUST (US-CT)との画像比較

MRガイド下集束超音波手術
FUS: MR guided Focused Ultrasound Surgery
<http://fus.breastcancer.jp/>

TSBIMI
Therapeutic Simulation by Image Matching
*システム疾患生命科学による先端医療技術開発拠点

新たな技術開発の課題に対するシミュレーションの採用

乳房のMRI画像を利用したHIFUシミュレーション

Numerical model

- 56ch transducer
- Focal distance 100 mm
- Frequency 2.1MHz
- 3D orthogonal mesh
- Size 120x110x110 mm
- 1200x1100x1100 grid
- 8x4x4 sub domains

MRI image

成分	音速 [m/s]	密度 [kg/m ³]
脂肪	1465 ^[1]	985 ^[1]
乳腺	1547 ^[2]	1032 ^[2]
結合組織	1615 ^[1]	1090 ^[1]

[1] ICU REPORT 61, [2] T. D. Mast., 2000

w/o control

w/ control

*嶋尾ら, 日本超音波治療研究会, 2012

Mechanical Clot Damage from Cavitation during Sonothrombolysis

Szeri@UC Berkeley, Hoelscher@UC San Diegoらのグループと共同研究

- Numerical domain size: 120x160x120mm
- Grid points: 600x800x600
- Frequency: 220KHz.
- 128 cores on Franklin (Cray XT4)@ NERSC

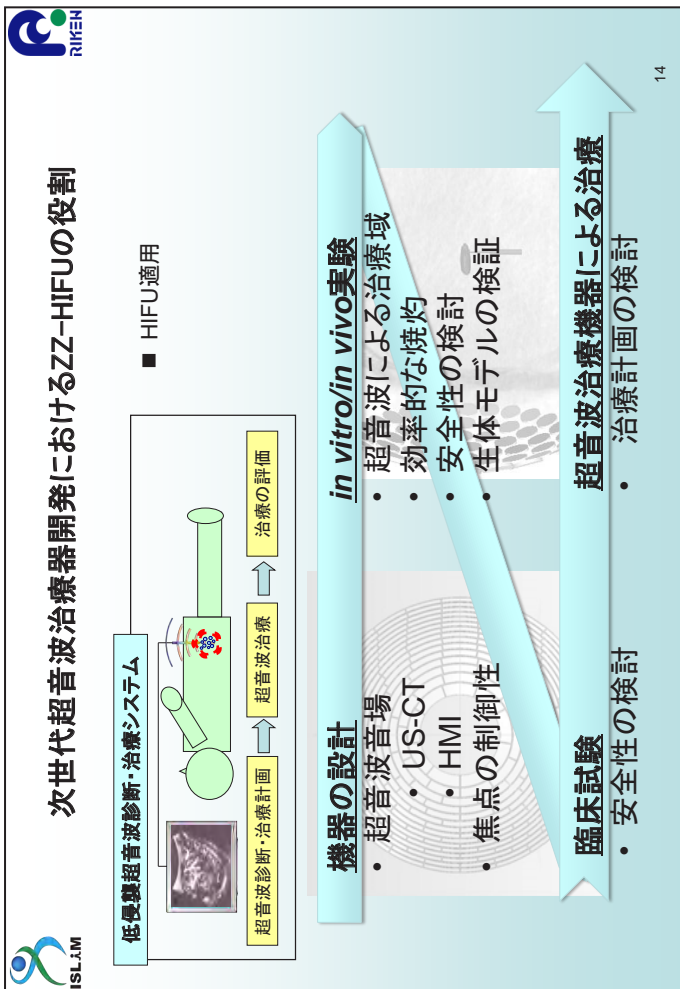
頭蓋骨のファントムを透過した超音波の焦点における圧力波形から、焦点における気泡の挙動と血栓溶解に関する研究を行っている。

Pressure [MPa]

Bubble radius [μm]

time [μs]

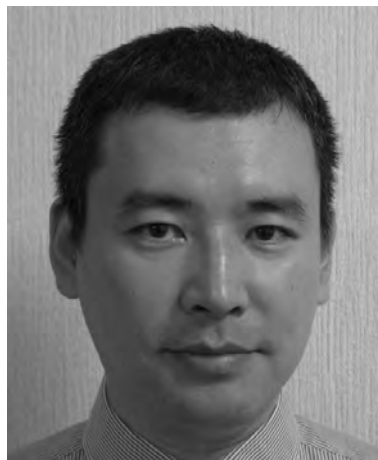
*Hope et al. J. Acoust. Soc. Am. (accepted)



大規模並列計算に適した 流体構造/膜連成手法 (ZZ-EFSI)

杉山 和靖

独立行政法人 理化学研究所 情報基盤センター
技術開発ユニット ユニットリーダー



発表者紹介

2000年3月 東京大学 大学院工学系研究科 機械工学専攻 博士課程 修了
2000年4月～2001年3月 東京大学 IML 研究員
2001年4月～2005年3月 独立行政法人 海上技術安全研究所 研究員
2005年4月～2007年9月 Twente University (オランダ) 研究員
2007年10月～2008年3月 東京大学 大学院工学系研究科 リサーチフェロー
2008年4月～2012年9月 東京大学 大学院工学系研究科 特任准教授
2012年10月～現在 理化学研究所 情報基盤センター

研究分野

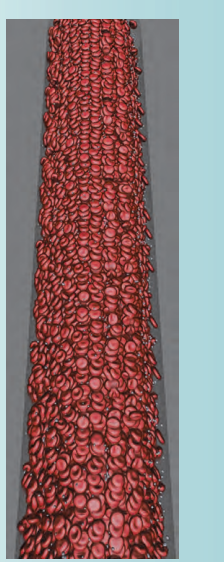
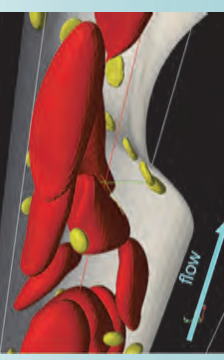
流体力学, 計算力学, 生体力学

ISLAM RIKEN

大規模並列計算に適した 流体構造/膜連成手法 (ZZ-EFSI)

理学研究所 情報基盤センター 技術開発ユニット
杉山和靖

伊井 仁志 (阪大・基礎工), 高木 周 (東大・工, 理研), 松本 洋一郎 (東大・工)

1

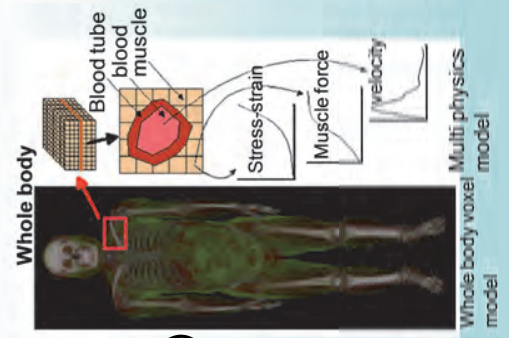
ISLAM RIKEN

背景・医療画像と相性の良い流体構造解析を実現したい
生体: 機械部品とは異なり, 元々, 設計図が存在せず

医療画像 (CT/MRI)

Voxel data (VOF; Volume Of Fluid)
→ 多媒質幾何・材質

(有限差分・固定格子)
Euler法に基づくシミュレーション



Whole body model

Whole body voxel model

Blood tube blood muscle

Stress-strain

Muscle force

velocity

Multi physics model

2

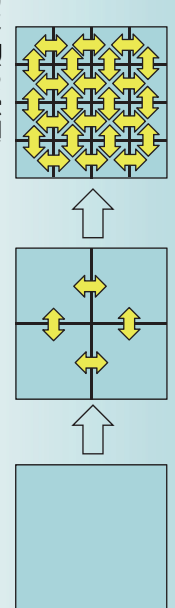
ISLAM RIKEN

背景・「京」の性能を有効に活用したい
スカラー機・超大規模並列, 並列処理の階層性

1秒間に1京(1・10¹⁶)回の演算
≐ 8・10⁴ (ノード) × 8 (コア) × 8 (演算) × 2・10⁹ (Hz)

・MPI (or 自動並列) ・OpenMP ・パイプライン・メモリアクセス量 に対する演算量

和積演算量は多いほど良いが, 差分は避けたい。



1並列 → 4並列 → 16並列 → ...

通信(ノード間の情報のやりとり)

ノード毎の演算量を均一にしたい → Euler法が有効
反復の多い処理を避けたい → 非圧縮性媒体を解析対象としても
圧力Poisson式を解きたくない

3

ISLAM RIKEN

目的・方策

- 大規模並列計算に適した流体構造/膜連成手法 (ZZ-EFSI) の開発.
- 既存の手法のチューニングではなく, 新たに, 実装すべき式を定め, スキーム, アルゴリズムを作る.
(計算規模の拡張が容易で, 実行効率の高い解析法を1から作る)

Euler型有限差分法

- 固定メッシュ上で, 流体力学, 構造力学 の式を離散化.

・ Sugiyama, Ii et al. (2011) *J. Comput. Phys.*, **230**, 596.
・ Ii, Gong et al. (2012) *Comm. Comput. Phys.*, **12**, 544.

動的パラメータを持つ擬似圧縮性法

- 数値的に安定で陽的な時間積分.

4

ISLIM **RIKEN**

概要・アプローチ (Lagrange法 vs. Euler法)

Lagrange法 (物質点によって方程式を記述)

Euler法 (空間的に固定した点で方程式を記述)

二相の区別は?

Fluid Solid (境界適合メッシュ)

Fluid Solid (空間的に固定した点で方程式を記述)

固体の記述は?

固体の変形メッシュ

境界適合メッシュ

固体の記述は?

左Cauchy-Green変形テンソル

物質点の相対位置の変化

固体体積率 c.f. 混相流のVOF法

• Sugiyama, Ii et al. (2011) *J. Comput. Phys.*, **230**, 596.
 • Ii, Sugiyama et al. (2011) *Int. J. Numer. Meth. Fluids*, **65**, 150.

5

ISLIM **RIKEN**

概要・アプローチ (SMAC法 vs. Present ACM)

SMAC法 反復計算あり

Present ACM 反復計算なし

予測ステップ

$$\mathbf{v}^* = \mathbf{v}^N + (\Delta t) \left(-\nabla p^N + \mathbf{F}(\mathbf{v}^*, \mathbf{v}^{N-1}, \dots) \right),$$

反復

Poisson式

$$\nabla^2(\delta p) = \frac{\rho \nabla \cdot \mathbf{v}^*}{\Delta t},$$

圧力発展式

$$\beta^{-2} \frac{\delta p}{(\Delta t)} + \gamma p^N + \rho \nabla \cdot \mathbf{v}^* = 0.$$

微圧縮性流体と見立てる

修正ステップ

$$p^{N+1} = p^N + \delta p,$$

$$\mathbf{v}^{N+1} = \mathbf{v}^* - (\Delta t) \frac{\nabla(\delta p)}{\rho}.$$

Ohwada & Asinari (2010) *J. Comput. Phys.* **229**

パラメータ β, γ を動的に決定
 → 擬似マッハ数の上界が解析的に:
 $Ma \leq 2\sqrt{3} \times (\text{CFL number})$

6

ISLIM **RIKEN**

計算の妥当性 (流体構造連成機能の実装)

キャビティ流れ中のneo-Hooke体粒子の運動

十分に検証された計算結果を良く再現

--- Zhao, Freund & Moser (2009, *J. Comput. Phys.* **227**)
 SMAC+ Lagrangian for solid
 --- Present (ACM + full Eulerian)

• Sugiyama, Ii et al. (2011) *J. Comput. Phys.*, **230**, 596.

7

ISLIM **RIKEN**

計算の妥当性 (流体膜連成機能の実装)

せん断流中の膜カプセル

十分に検証された計算結果を良く再現

キャピラリ数 $Ca = \frac{\mu a \dot{\gamma}}{E_s}$

$Ca = 0.2$

$Ca = 0.0125$

変形パラメータDの時間変化

● Eggleton et al. (1998) *Phys. Fluids*, **10**, 1834.
 △ Pozrikidis (1995) *J. Fluid Mech.*, **297**, 123.

Full Eulerian ($\Delta x=d/16$)
 Full Eulerian ($\Delta x=d/32$)
 Full Eulerian ($\Delta x=d/64$)

$D = \frac{L-B}{L+B}$

$Ca=0.2$
 $Ca=0.1$
 $Ca=0.05$
 $Ca=0.025$
 $Ca=0.0125$

• Ii, Gong et al. (2012) *Comm. Comput. Phys.*, **12**, 544.
 • Ii, Sugiyama et al. (2012) *J. Comput. Phys.*, **231**, 2328.
 • Takagi, Sugiyama et al. (2012) *J. Appl. Mech.*, **79**, 010911.

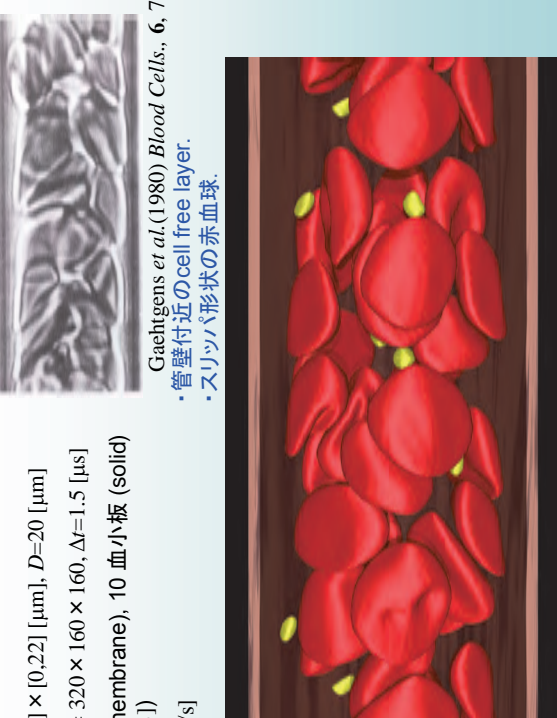
8

計算の妥当性 (赤血球「群」の振る舞い、実験観測との比較)

- $[0.44] \times [0, 22] \times [0, 22]$ [μm], $D=20$ [μm]
- $N_x \times N_y \times N_z = 320 \times 160 \times 160$, $\Delta t=1.5$ [μs]
- 30 赤血球 (membrane), 10 血小板 (solid) ($Ht \approx 20$ [%])
- $\langle u \rangle \approx 2$ [mm/s]

Gaehgans *et al.* (1980) *Blood Cells.*, 6, 799.

- 管壁付近の cell free layer.
- スリッパ形状の赤血球.



ISLAM RIKEN

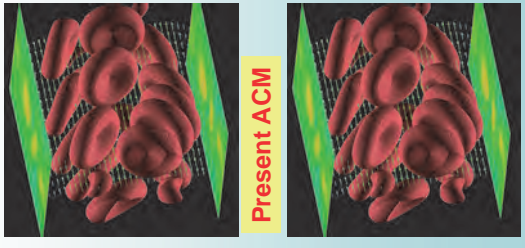
• Ii, Sugiyama *et al.* (2012) *J. Biomech. Sci. Eng.*, 7, 72.

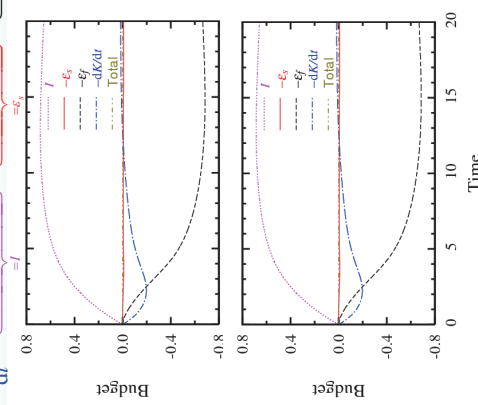
計算の妥当性 (運動エネルギー輸送の収支)

チャネル流れ内biconcave neo-Hooke体粒子

SMAC+FFT

Present ACM



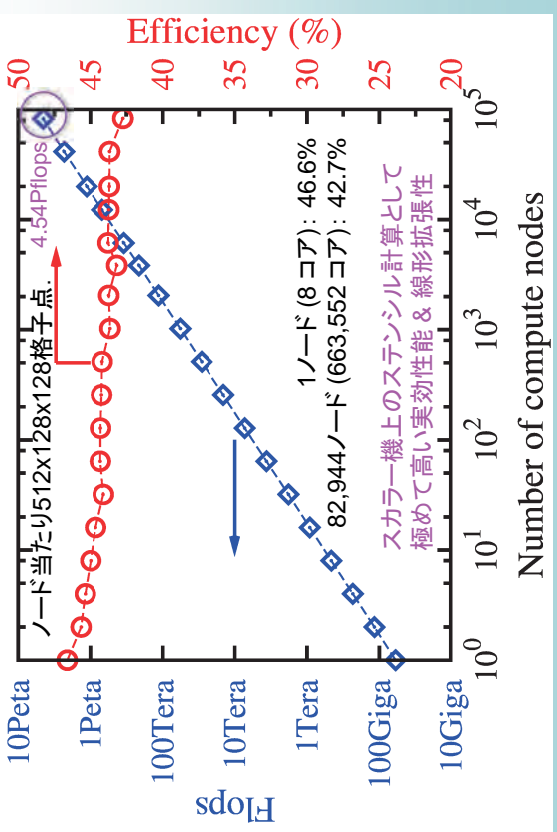
$$-\frac{dK}{dt} + \underbrace{(v_x(-\Delta P/L))_\Omega}_{=s_c} - \underbrace{(\phi_s GB' \cdot D')_\Omega}_{=s_e} - \underbrace{(2\mu D' : D')_\Omega}_{=s_v} = 0.$$


エネルギー保存性を満足

ISLAM RIKEN

並列性能@「京」 (実効性能と並列台数. weak scalability)

チャネル流れ内biconcave neo-Hooke体粒子



ノード当たり512x128x128格子点. 4.54Plopts

1ノード (8コア): 46.6%

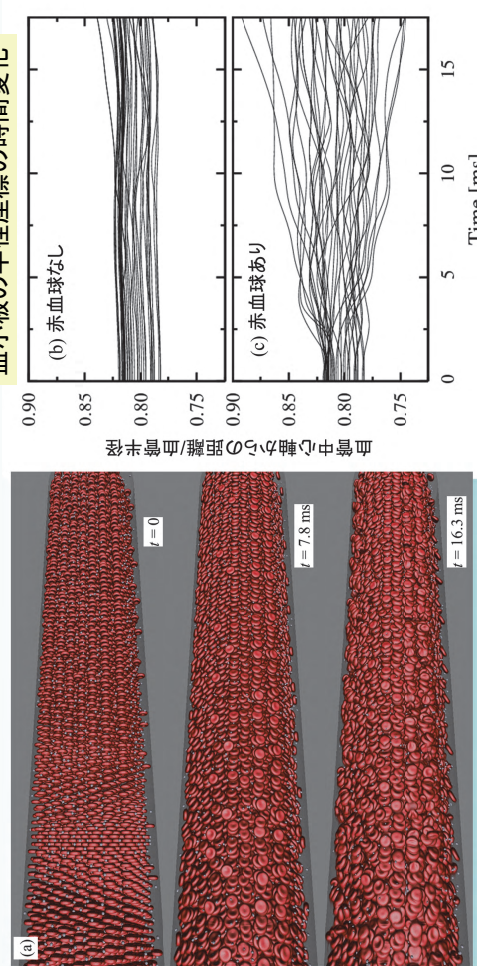
82,944ノード (663,552コア): 42.7%

スカラー機上のスタンシル計算として極めて高い実効性能 & 線形拡張性

ISLAM RIKEN

計算例・大規模血流シミュレーション 血管径約100 μm

血小板の半径座標の時間変化



(b) 赤血球なし

(c) 赤血球あり

ISLAM RIKEN

ISLIM RIKEN

血栓シミュレータの開発 血小板血栓に至る過程を数値予測したい

マルチスケール・マルチフィジックス

血流 (連続体力学)

スケール間相互作用

Ligand-Receptor結合 (確率論的)

GPIIb/IIIa-vWF 結合・切断 (反応速度, 確率)
伸縮力

分子的相互作用 (分子動力学)

0 (ms) 5 (ms) 10 (ms) 15 (ms) 20 (ms)

13

ISLIM RIKEN

血小板付着の様子, 赤血球の影響

赤血球なし

Vertical velocity at $t = 16$ [ms]

赤血球あり ($Ht = 14\%$)

Vertical velocity at $t = 16$ [ms]

14

ISLIM RIKEN

血小板付着の様子, 赤血球の影響

狭窄部

flow

GPIIb α -vWF 結合数の時間変化

Total number of bonds

t [ms]

- $Ht = 14.0\%$
- $Ht = 10.5\%$
- $Ht = 7.0\%$
- $Ht = 3.5\%$
- $Ht = 0\%$

15

ISLIM RIKEN

まとめ (ZZ-EFSI)

新たな方法論の確立

Euler型流体・構造/膜連成解析法

- 十分に検証された計算結果を再現.
- 複雑な境界形状/多数の分散体を含む系の解析を容易に.

動的パラメータを持つ擬似圧縮性法

- $Ma \leq 2\sqrt{3} \times (CFL \text{ number})$.
- 完全陽的な時間積分を実現.

実効性能・線形拡張性の高い超大規模並列計算を可能に.

血栓シミュレータへの応用・拡張

血流とLigand-Receptor結合のマルチスケール/フィジックス解析

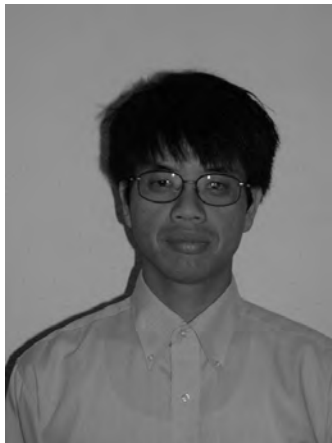
- 血栓形成の第一段階における, 赤血球の存在の重要性を示す.

16

UT-heart によるサルコメア力学から 心筋細胞構造を経て心拍動にいたる マルチスケール解析

鷺尾 巧

東京大学大学院 新領域創成科学研究科



発表者紹介

- 1990 年 3 月 大阪大学理学部数学科修士課程修了
- 1991 年 4 月 日本電気(株)入社 コンピュータ技術本部にて HW 開発に従事
- 1993 年 4 月 同社 C&C 研究所にて高速計算アルゴリズムの研究開発に従事
- 2004 年 4 月 東京大学新領域創成科学研究科・(独) 科学技術振興機構 GREST 研究員として心臓シミュレータの開発に従事
- 2009 年 4 月 同研究科 特任研究員として同シミュレータの開発を継続

研究分野

数値解析, 計算生物学

東京大学
THE UNIVERSITY OF TOKYO

ISLIM

UT-heartによるサルコメア力学から 心筋細胞構造を経て心拍動にいたる マルチスケール解析

東京大学新領域創成科学研究科
特任研究員
鷲尾巧




ISLIMソフトウェア研究開発報告書

1

東京大学
THE UNIVERSITY OF TOKYO

ISLIM

背景・目的

- **背景**
 - 「ヒトゲノムプロジェクト」の完成を受けて生命科学の焦点は遺伝情報の解明からたんぱく質および高次の生命単位の機能の解明へと移行しつつある。
 - 「フィジオーム」とは、たんぱく質から細胞内小器官、細胞、組織を経て臓器、そして個体までの各階層に属する膨大な数の機能モジュール間の相互作用として表現される生命現象を計算機シミュレーションによってモデル論的に再現しようとする研究領域であり、ポストゲノムの医学・生物学の研究の中心課題のひとつとなっている。
- **目的**
 - 近年発展の著しい分子生物学により次々と得られるミクロな知見をマクロな臨床医学へと合理的に接続し、医療へ貢献すると共に、新たな学術領域 [計算医学] を創出する。
 - マクロな心機能を細胞下レベルの電気化学・力学的なミクロ事象からありのままに再現するマルチスケール・マルチフィジックス心臓シミュレータ UT-Heart を実現し、ミクロレベルの異常と心疾患の関係を合理的に解明し、医学・医療に貢献する。

ISLIMソフトウェア研究開発報告書

2

東京大学
THE UNIVERSITY OF TOKYO

ISLIM

概要・アプローチ

- **研究開発コードの概要**
 - 本シミュレータは、イオンチャネルや収縮タンパクの電気化学・力学現象を再現した心筋細胞有限要素モデル、並びに医療画像(CT)から精密に作られた全心臓有限要素モデルから成る。後者の各有限要素に前者の心筋細胞有限要素モデルを埋め込み、両者を独自に考案した非線形均質化アルゴリズムにより連成させることにより、細胞下のミクロ事象から心臓の拍動や血液の拍出などのマクロ事象を再現することを可能とする。
 - 以上により、これまで推測の域を出なかったミクロ事象(例えばデスミンなど機能タンパクの異常)とマクロ事象(例えば拡張型心筋症、心肥大症など)の関係を合理的に解明し、従来にはない観点から医学・医療に貢献することを旨とする。
- **アプローチ**
 - 大規模な電気・力学・力学現象のマルチフィジックス問題を、流体構造連成解析などの有限要素法をベースとする最先端の科学計算手法により正面からシミュレートする。
 - 形ある細胞モデル(有限要素法により内部構造を再現した3次元数値細胞)を開発し、均質化法をベースにミクロ現象とマクロ現象をシームレスにつなぐマルチスケールシミュレーションを達成する。

ISLIMソフトウェア研究開発報告書

3

東京大学
THE UNIVERSITY OF TOKYO

ISLIM

現在までの研究開発成果

- 現時点の開発ソフトウェアについて
 - ミクロスケール(心筋細胞複合体)とマクロスケール(心室)のマルチスケール解析を均質化法により実現。
 - 克服すべきであった問題点
 - 均質化法における非圧縮性の取り扱い
 - 分子モデルの確率的振る舞いと連続体モデルの有限要素解析とを結び付ける計算手法の確立
 - 解決策
 - 非圧縮材料を扱うことのできる均質化法を開発
 - ミクロとマクロでエネルギー的に矛盾の生じないロバスタな連成法を開発
 - 計算事例
 - 問題サイズ: 5万自由度細胞モデル × 63万細胞 (計: 約315億自由度)
 - 京のフルノードを使用し、正常な一心拍を再現

ISLIMソフトウェア研究開発報告書

4

東京大学
THE UNIVERSITY OF TOKYO

プロジェクト終了時の達成目標



- 約束
 - 筋原線維サルコメア内のたんぱく質レベルの確率的運動、複合体としての細胞モデル収縮運動、および臓器レベルの拍動運動の3つのスケールをシームレスに結び付けるシミュレーションを実現し、たんぱく質レベルの機能と臓器レベルの拍動性能の関連について新たな知見を得る。特に、心疾患の事例として肥大型心筋症を取り上げる。

脳神経系研究開発チーム成果報告

神経系の大規模シミュレーション

石井 信

次世代計算科学研究開発プログラム
脳神経系研究開発チーム チームリーダー



発表者紹介

- 1986年3月 東京大学工学部 卒業
- 1988年3月 東京大学大学院工学系研究科 修士課程 修了
- 1988年3月 (株)リコー中央研究所 研究員
- 1997年3月 工学博士(東京大学)
- 1997年4月 奈良先端科学技術大学院大学情報科学研究科 助教授
- 2001年4月 同上 教授
- 2007年7月 京都大学大学院情報学研究科システム科学専攻 教授
- 2008年10月 理化学研究所次世代計算科学研究開発プログラム チームリーダー

研究分野

システム神経生物学、バイオインフォマティクス、統計的学習

神経系の大規模シミュレーション

石井 信

次世代計算科学研究開発プログラム
脳神経系研究開発チーム チームリーダー

1. 目的

脳神経系は、物質基盤の上に情報基盤を構築した点で特異な生命システムである。また、遺伝子・分子から脳全体、個体集団に至る生命の階層の中で情報処理基盤が構築されるため、必然的に多階層をなす複雑システムである。脳神経系の情報処理は、二つの異なる時定数を持つ様式に大別される。第一は、ミリ秒から秒の時定数を持ち、外部からの入力刺激に基づき外部への働きかけ（運動）を実施する刺激運動変換過程であり、第二は、分から時、日といった長い時定数を持ち、環境に依存して、刺激運動変換自体を変化させる適応（学習）過程である。前者は、神経回路レベルでの電気生理学的な現象（神経発火）がメインプレーヤであり、後者は、神経発火のみならず神経細胞内のイオンや酵素、受容体などの分子がメインプレーヤとなる。脳神経系の情報処理は、常に脳をとりまく物理世界の時定数（すなわち実時間）において実現される。したがって、入力から出力までをつないだ刺激運動変換の実時間での再現、また、実環境において外的刺激にさらすことで適応過程を再現する必要がある。以上のことから、脳全体レベルでの入出力変換、およびその学習をシミュレーションのターゲットとし、そのための構成要素である神経細胞および神経回路の研究（神経細胞シミュレーション、神経回路シミュレーション）を並行して実施している。前者について、200億個を超える細胞からなるヒトの脳全体のシミュレーションは、「京」であっても困難であるので、対象を限るものとし、ほ乳類の視覚系と無脊椎動物（特に昆虫）の嗅覚系を対象として、入出力にわたる回路のシミュレーション研究を実施している（視覚系シミュレーション、昆虫嗅覚系シミュレーション）。

2008年10月に発足した脳神経系研究開発チームでは、石井（京大、理研）が神経細胞シミュレーション、深井、Diesmann（理研）が神経回路シミュレーション、黒田（東大）、臼井（理研）、銅谷（沖縄科学技術大）が視覚系シミュレーション、神崎（東大）が昆虫嗅覚系シミュレーションを主担当することとして、7PI体制でスタートした。2010年4月からは、神経細胞シミュレーション

に市川（東大医科研）を加えた8PI体制になっている。また、Diesmannは2011年9月よりドイツJuelich Research Centerに転出したため、現在は、国際協力体制で研究開発を行っている。

2. 現時点での成果

2.1 神経細胞シミュレーション

石井（京大、理研）：発達期の神経細胞における特徴的な局所構造の獲得や、軸索伸長などの動的な形態変化の理解を目指し、神経細胞形態形成シミュレーションプラットフォーム NeuroMorphoKitを開発してきた。細胞形態形成に関わるシグナル分子の反応拡散現象、細胞骨格の動態（主にアクチン系）、細胞膜の変形を含むマルチフィジックスモデルにより汎用的な細胞形態変化のシミュレーションを実現し、これまでに、アクチン線維により駆動される葉状仮足の自己組織化、化学走性などを再現した(Nonaka, et al., *Neural Networks*, 2011)。NeuroMorphoKitの高度化のため、大規模並列環境におけるパフォーマンスチューニング、神経系への適用を可能にするためのシミュレーションモデルの拡張を行った。前者について、MPIによる並列化アルゴリズムの改善を推し進め、並列実行時に特にボトルネックとなっていた、重合・脱重合・分岐・切断・キャッピングなどアクチンキネティックスの確率シミュレーションにおける分子濃度情報の更新、および細胞膜の変形シミュレーションにおけるエネルギー計算で発生するルートノードへの通信量を大幅に削減した。これにより、京上での初期アクチン繊維数約120万本のシミュレーションにおけるストロングスケールは、従来の $\alpha=0.91(512)$ から $\alpha=0.99(8192)$ に改善した。後者について、力学的な曲げと結合タンパク質による結束を考慮したアクチンフィラメントのパネネットワークモデル化、および膜モデルの三次元化を実施し、現在統合を進めている。これらのモデル拡張により、三次元でのアクチンフィラメントの挙動を考慮したシミュレーションを行うことが可能となった。このことは、軸索伸長時に軸索先

端部に見られる特徴的な構造である成長円錐のシミュレーションなどにおいて特に重要である。

市川 (東大医科研) : A-Cell は形態を有する神経細胞について、電気生理学反応と細胞内分子反応を同時に計算できるシミュレータである。これまでの A-Cell は細胞の形態変化に対応できなかったため、NeuroMorphoKit との間で形態変化、局所膜電位やコンパートメント内分子濃度などを共有できるためのデータ交換形式 netCDF-neuron を規定しその基本動作を確認した。また、記憶形成を修飾する転写因子 NF- κ B の振動について約 40 万コンパートメントを有する 3D モデルを構築(Ohshima, et al., PLoS ONE, 2012)し、MPI 化された A-Cell によって 99.99% の並列化率を達成した。

2.2 神経回路シミュレーション

Diesmann (Juelich Research Center) : 神経回路の京上での大規模シミュレーションを可能とするソフトウェア環境 NEST(NEural Simulation Tool)を開発してきた。NEST は主に点神経細胞 (形態を有しない神経細胞) からなる回路のシミュレーションを得意としているが、形態のある細胞も対応可能である。神経の回路内情報伝達は神経発火による。発火ごとにそのイベント情報を全細胞に伝達するイベント駆動方式は、正確であるものの効率が悪い。NEST では、時間軸を一定の間隔に区切り、その間隔内では予測に基づき細胞ごとに積分を行った上で、間隔内で発生した発火を全細胞に伝え、発火については時間をさかのぼって各細胞の積分の更新を行うという時間駆動型のアプローチをとることで、効率の良い計算を行っている。この際にカルマンフィルタなどの精度の良い予測法を用いることで精度を確保する(Plesser, et al., *Neural Computation*, 2009; Hanuschkin, et al., *Frontiers in Neuroinformatics*, 2010)。神経回路シミュレーションの難しさは、原理的には全細胞間が接続し得るため、空間局所性など並列化できるてがかりが少ないことである。しかし、これまでに NEST は、京の上で、 5×10^8 個の細胞、 3×10^{12} 個のシナプスからなる、小型のほ乳類の脳全体に匹敵する巨大回路のシミュレーションを実施した。24,576 ノード(196,608 コア)までの hybrid 並列実装による実行確認ができており、上記回路でのシミュレーションを行ったところ、196,608 コアを用いて、1 秒の生物時間を 1,544 秒でシミュレーションできている。各コアの 1 神経細胞あたりの計算速度で比較すると、Juelich センターの Blue Gene P(JUGENE)よりも数倍程度高速である。このような、また、さらに

規模を拡大した回路シミュレーションを行うため、新たにデータ構造の見直しを行い、効率的なメモリ利用を実現した (Helias, et al., *Frontiers in Neuroinformatics*, 2012)。NEST が目指すのは、大型のほ乳類の脳全体レベルとなる大規模回路のシミュレーションである。現在、京の全ノードが使用できれば、 10^9 個の細胞からなる大規模回路のシミュレーションをテストできるまで開発が進んでいる。今のところ京の利用機会が得られていないが、今後機会を得て、世界最大規模の回路シミュレーションを行い、ボトルネックとなる可能性のあるコミュニケーション性能や回路規模についての性能評価を行うことを目標としている。また、NEST において部位局所的に遅い時定数で作用する神経修飾物質の効果を取り入れ、ドーパミンによる時間差分学習が可能であることを示した(Potjans, et al., *PLoS Computational Biology*, 2011)。

深井 (理研) : 大脳新皮質で見られる特徴的な階層構造が神経活動とどのような関係を持つのか、いかなる情報処理を実現しているかを明らかにすることを目指して、神経回路シミュレーションの研究を行っている。視覚野では、外界から得られる膨大な視覚情報から、視野内にある注目すべき情報に焦点を当て、それを重点的に処理している (視覚的注意)。電気生理学的・解剖学的知見に基づく詳細なネットワークを持つ視覚皮質の層構造局所回路モデルを NEST により構築 (Potjans and Diesmann, *Cerebral Cortex*, 2012) し、視覚的注意が視覚皮質層の動態に与える影響をシミュレーションにより検討した。Bottom-up 的な視覚情報と Top-down 的な注意入力局所回路内で作用しあうことを示し、視覚的注意が皮質層の視覚処理に果たす役割を予測した (Wagatsuma, et al., *Frontiers in Computational Neuroscience*, 2011)。また、局所回路の数を増やし、視覚野内の局所領域をより忠実に再現するシミュレーションを行った結果、空間的注意と特徴に基づく注意による、神経細胞応答の変調効果が、定性的かつ定量的に良く説明されることがわかった。

2.3 視覚系シミュレーション

黒田 (東大) : これまでに、大脳皮質の錐体細胞を対象として、カルシウムシグナル関連分子を含む詳細学習 (Spike Timing Dependent Plasticity; STDP) モデルを構築、また、カルシウムの上下流の分子を中心とした簡素化モデルを構築、それを用いて、オタマジャクシの視蓋における方向選択性機構の獲得をシミュレーションにより明らか

にした(Honda, et al., *Journal of Neuroscience*, 2011)。この簡易 STDP モデルを NEST のプラグインとして実装し、京上で動くようにした。さらに、STDP の制御モデルとしての再モデル化を行い、このモデルにより、カルシウムダイナミクスとリン酸化ダイナミクスの異なる役割が明白となった。

白井 (理研) : ヒトの視覚系は、眼光学系、網膜、視覚皮質、脳幹(眼球運動駆動部)などで構成される多階層複雑系であり、その機能の理解には、構成要素のモデル化とともに、それらを統合した大規模シミュレーションが必要である。これまでに、計算機環境に依存せずモデルの構築や統合が可能なモデル開発共有プラットフォーム環境 (PLATO) の開発(Kannon, et al., *Neural Networks*, 2011)を進め、また、モデル結合に用いる共有データフォーマットならびにその利用支援ライブラリを RICC ならびに京上で MPI 動作可能なように実装を進めた。視覚系を構成する各構成要素モデルの状況は以下のものである。(1) 眼球運動系は、コンダクタンススペースの細胞モデルとして記述し、物体注視時に生じる微小眼球運動ならびに視野上の物体を反射的に捉える急速眼球運動の生成を可能とした (Inagaki, et al., *Neural Networks*, 2011; 京上で 30 並列を実現)。(2) 眼光学系は、眼球形状の特性、瞳孔系の特性などの知見を追加して構築を進め、1000x1000 の網膜像の生成を可能とした(400 並列を実現)。(3) 網膜は、視細胞、水平細胞、双極細胞、ならびに神経節細胞から構成されている。視細胞について、約 3600 万個の細胞から構成されるモデルの構築を行い、それらのうち 150 万個の錐体細胞からなる Cone Mosaic と呼ばれる L 型、M 型、S 型の視細胞配置による応答を再現し、1,504 並列での動作(ストロングスケール $\alpha=0.89$)を確認した。双極細胞層および神経節細胞層は、それぞれ 150 万個、40 万個の細胞から構成されるモデルを構築し、それぞれの細胞応答を再現するとともに 1,024 並列での動作を確認した。また、これまでに開発した共有データフォーマットならびにその利用支援ライブラリ PLATONIC を用いて、眼球運動モデル、眼光学モデル、網膜視細胞モデル(Cone Mosaic)ならびに双極細胞層・神経節細胞層モデルの統合を進め、自然画像、錯視画像を用いたシミュレーションを RICC ならびに京上で並列実施することに成功した。

銅谷 (沖縄科学技術大、京大、理研) : 視覚系の出力を司る眼球運動系、すなわち、網膜-上丘-眼球運動制御系のモデルを NEST 上に実装した。特に上丘は多階層から成るスパインキングニュー

ロン回路モデルとして実装した。網膜入力に基づきサッカド眼球運動が誘発される際に、上丘回路の各層では特徴的な時空間パターンが形成される。例えば、サル神経生理学実験によれば、上丘の出力を与える中間層のバースト細胞は、ベル型の発火パターンを持ち、深層では活動の時空間的な拡張が生じることが分かっているが、シミュレーションはこれらを良く再現し、その新たな機能的な意味付けを与えた。また、上丘により駆動される眼球運動は、サッカド角度に対してピーク速度は特徴的な非線形カーブを描くが、この性質もシミュレーションにより良く再現できた。シミュレータは現在 RICC 上で並列動作しているが、今後京上で動作確認を行う。また、視覚運動系を中心とした閉回路シミュレーションを実施した。このシミュレーションは ATR (京都) に置かれたロボット頭部、RICC (和光) 上の上丘シミュレータ、それらをネットワークでつなぐクラスタモジュール(MUSIC を用いて実装)からなる。ロボットの眼球にあたるカメラ画像から特徴点を抽出し、その入力に対する上丘モデル (RICC の 256 並列で動作) の回路動作に基づき、ロボットカメラが特徴点位置に跳躍眼球運動を実行する。システム全体の遅れ時間は 120ms であるが、これは主にネットワークの遅れによる。

2.4 昆虫嗅覚系シミュレーション

神崎 (東大) : モデル生物であるカイコガを対象に 10^5 程度の神経細胞からなる脳全体レベルの嗅覚情報処理、すなわち、脳における入力 (匂い受容) から出力 (匂い源探索行動) までの情報処理のシミュレーションを目指している。これまでの成果は以下のものである。(1) カイコガ神経データベース (BoND) の整備を進め、約 1,600 個のカイコガの脳内神経の三次元形態・生理データを収集、解析などに利用可能にした (Kazawa et al. 2008)。(2) 嗅覚系一次中枢である触角葉 (哺乳類脳の嗅球に対応)、および匂い源探索の運動パターンを生成する前運動中枢である側副葉 (LAL-VPC) において、各領域また各領域を連結する神経細胞について細胞レベルで構造と機能のデータを解析した (Namiki and Kanzaki 2008, Kazawa et al. 2009, Iwano et al. 2010, Namiki and Kanzaki 2011)。側副葉においては、持続的なステアリング信号を出力する側副葉からの出力である下降性神経が側副葉の特定の領域に分岐していること、側副葉の入力細胞の嗅覚応答が一過的であること、また側副葉は両側の触角情報を統合すること、さらには、行動指令信号の長期的応答が側副葉で生成されること等を明らかにし

た。(3) 脳から胸部運動系に下降する神経細胞を選択的にラベルし、選別したニューロンをスライドガラス上に展開し培養することで、電位依存性の電流を計測する手法を確立した。同時に触角葉や側副葉を構成する神経細胞の電位や電流クランプ実験を進め、触角葉から上位中枢への出力神経や側副葉神経細胞のスパイク発火の特性情報を得た。(4) 側副葉を構成する神経細胞の膜電位データと形態データから、側副葉の神経回路を構成する両極性神経 (BLN) や局所介在神経 (LN) をさらに分類し (Iwano et al. 2010)、神経細胞の平均応答波形を作成した。この平均応答を作り出す側副葉神経回路のシナプス結合強度の推定を行い、BLN が抑制性であるという条件下では一意に近い解を得た (Yamagishi et al. 2012)。現在この結果を基に、側副葉回路モデルを構築中である。(5) 多数の脳の形状に対して一種の平均操作を加えることで、カイコガの「標準脳」を作成した。神経細胞をこの標準脳座標系に $10\mu\text{m}$ 以下の誤差でマッピングするプロトコルを確立した (Ikeno et al. 2012)。(6) 神経細胞の三次元形態から神経形態を抽出する技術を試行し (Nakajima et al. 2008, 2009)、神経細胞の形態を効率よく取得するソフトウェアを開発した (Ikeno et al. 2012)。(7) 神経細胞プラットフォームシミュレータである NEURON を京に移植し、チューニングを行い、単体性能で 1.54GFLOPS/core を達成した。これは $10,000$ コンパートメントの神経細胞を 32 コア並列でシミュレーションした際にリアルタイム実行するのに必要な速度であり、シミュレーション時のコンパートメント数を節減すればリアルタイムシミュレーションが可能である。さらにシナプス間の通信をノード配置最適化したポイント間通信に改良することによって通信コストを 2 倍程度以上に改良した。現在は $196,608$ コアの並列実行で 187TFLOPS の計算性能を達成している。(8) 実数値遺伝的アルゴリズム (RCGA) を用いて、単一細胞の膜特性を推定するソフトウェアを作成した。現在までに、 13 万コア未満の並列実行について良好な並列性能を得た。このプログラムを触角葉出力神経に適用し、入力信号 (電流注入) として適当な正弦+三角波を入力することで早い収束が得られることを示した。

2.5 学術的成果とアウトリーチ

本研究に関わる国際学術専門雑誌掲載論文の数は、2008年 (10月から12月) は5件、2009年は10件、2010年は9件、2011年は24件、2012年は22件と順調である。その一部は、Nature Cell Biology, Molecular Systems Biology, Nature

Neuroscience, Neuron, Journal of Neuroscience, PNAS など分野を代表する一流雑誌に掲載された。これらはシステム生物学研究、大規模データ解析などによるもの (いわゆる計算神経科学に関するもの) であり、スパコン利用を前提としたものとは限らないが、本プロジェクトに関わるメンバーの学術活動のレベルの高さを示している。京によるシミュレーションが含まれるものとして、Helias, M., et al. Supercomputers ready for use as discovery machines for neuroscience. Frontiers in Neuroinformatics, 6:26 (2012) が誌上発表され、また、国際会議 SC12 においても、Miyamoto, D. et al. Neural circuit simulation of Hodgkin-Huxley type neurons toward peta scale computers が発表された。また、プロジェクト期間内で、関係研究者および学生の受賞は、稲垣の CNS2011 poster award など 16 件、メディアに取り上げられたのは、NHK サイエンスゼロ (神崎グループ、2009年9月19日)、NHK スペシャル (石井、2012年6月3日) を含め 30 件である。このように、研究活動およびアウトリーチ活動は良好である。

3. プロジェクト終了時の達成目標

3.1 神経細胞シミュレーション

神経細胞におけるマルチフィジックス (骨格系、細胞膜、細胞内分子の反応拡散) を統合的にシミュレーションするためのプラットフォーム NeuroMorphoKit により、神経伸長の方向決定性に関するフィラメントベースの三次元シミュレーションを実行する。京上での $10,000$ コア並列を目標とする。並行して、膜電位および分子の反応拡散の計算を可能とし、可視化など各種インターフェースを備える A-Cell の MPI 化を進め、NeuroMorphoKit との動作統合のためのデータ交換法を開発する。

3.2 神経回路シミュレーション

NEST のコミュニケーションボトルネックについてアルゴリズムの改良をさらに進め、京上での実行時間が確保されれば、世界最高水準となる、 10^9 個の細胞、 10^{13} 個程度のシナプスからなる回路シミュレーションを行うことで、小型の霊長類の脳全体レベルの回路動態の再現を目指す。NEST による複数カラムモデルシミュレーションを実施し、特徴注意と空間注意の違いの説明を試みる。また、電気シナプスを考慮した並列計算技術を NEST に実装し、既存のシミュレータの技術的な困難を排除することで、より現実的な局所回路モデルの高速シミュレーションを可能にする。

3.3 視覚系シミュレーション

視覚系の入力部に相当する眼光学-初期視覚系に関する統合シミュレーション環境 **VSM** に対して、脳の情報処理の実時間性を考慮した並列化を進め、**RICC** 上で高並列実行するとともに、**FX** (京) 上でのチューニングを進める。**FX** 上では **2000** コア並列以上を目指す。また、**VSM** を用いて、錯視現象のモデルに基づく機構解明を目指す。網膜モデル、上丘モデル、眼球運動モデル、視覚空間モデルを **MUSIC** により結合した視覚系の入出力閉回路シミュレーションを、**FX** 上で、実時間と同程度の時間で動作させる。

3.4 昆虫嗅覚系シミュレーション



詳細な生物学実験に基づいた昆虫の嗅覚系神経回路および細胞に関するデータベースと、それに基づいて形態抽出や標準脳マッピングを行うことでシミュレーションにつなげるシミュレータ **IOSSIM** の開発を完了する。京上の **80,000** ノード実行で **1** ペタに近い実効性能の達成を目指す。神経回路レベル・単一細胞レベルの生物学的パラメータ探索とその評価を行う。匂い情報入力から匂い源探索行動の行動指令を生成する中枢回路 (**LAL-VPC**) のシミュレーションを、京上で高並列実施する。大規模マルチコンパートメントモデルについては、京上で **2~3** 万コア並列程度により **LAL-VPC** 神経回路の特性の時間計算を達成する。実時間動作させた昆虫脳シミュレーションを移動ロボットとリアルタイムで通信させる。

3.5 将来に向けて

脳神経系研究開発チームの研究は、長期的には、脳神経系における発達障害の機序の解明、最適な学習 (教育) 過程の設計、脳を模した知的情報処理デバイスの開発などの社会・産業応用、また、ブレインマシンインターフェースの高精細化などの医療応用に貢献することを目指している。欧米では、脳神経系シミュレーション研究はこれまでスパコン開発と足並みを揃えて発展しており、**Blue Gene/Q** のターゲットの一つは神経回路シミュレーションとされている。また、欧州では **Human Brain Project** というメガサイエンスの動きがあるが、ここでもスパコン開発が重要となっている。われわれは、次世代の **Exa** スケール計算機と **NEST** の拡張により、ヒトの脳全体レベルの回路シミュレーションが可能と考えているが、そのためには脳のマクロ、ミクロのレベルでの回路パラメータを調べる基礎研究も重要である。また、大脳皮質の運動などの機能の理解には、複雑な形

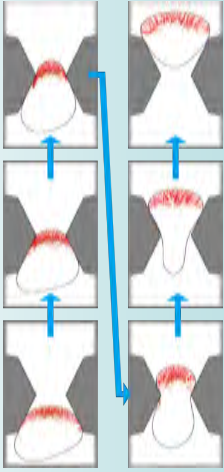
態を持つ細胞からなる大規模回路の超高並列シミュレーションが必要となる。後者について、**ISLiM** プロジェクトでは回路 (通信) を得意とする **NEST** と、精緻な細胞を得意とする **NEURON** の両者について、京上でチューニングを行ってきたので、今後はそれらを統合することのできるソフトウェアの開発も必要であると思われる。

なお、京上の性能に関しては京速コンピュータ京の試験利用および特別運用による。**RICC** 上の性能に関しては理化学研究所情報基盤センターの **PC** クラスタ **RICC** による。京上の実装については **ISLiM** 高度化チームの強力なサポートも得た。関係各位に感謝申し上げます。






神経系の大規模シミュレーション 脳神経系研究開発チームの成果

京都大学 大学院情報学研究所
理化学研究所 次世代計算科学研究開発プログラム
石井 信




1


脳神経系チームの目標

脳神経系シミュレーションで解明すべき生命機能
脳神経系にとって最も重要な機能は、情報処理（知覚処理）と学習（適応）である

①秒以下の時間スケールで情報を処理



②分、時間という時間スケールで学習



次世代スーパーコンピュータで脳神経系のリアルなシミュレーションが初めて可能に

今まで


- スケールを小さくした局所回路のシミュレーション
- 実際の脳の複雑度を再現できるリアルな局所回路のシミュレーション
- 脳神経系の実際の入出力関係を再現するシミュレーション

↑

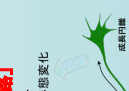
次世代スーパーコンピュータでは

- 実際の脳の複雑度を再現できるリアルな局所回路のシミュレーション
- 脳神経系の実際の入出力関係を再現するシミュレーション


局所回路




神経細胞と局所神経回路




シナプス結合
スパイクの形態変化



神経細胞





局所回路



昆虫脳 (約10,000ニューロン) 全網膜 (約1,000,000ニューロン) 2

①脳神経系の基本的な情報処理単位の役割解明
【神経細胞と局所神経回路】

②脳全体レベルでの情報処理機構の解明
【無脊椎動物視覚系とほ乳類視覚系】

脳神経系チームの構成

神経細胞・局所回路シミュレーション

Shin Ishii
Kazuhiisa Ichikawa
NeuroMorphoKit-A-Cell

Tomoki Fukai
Markus Diesmann
NEST-CMDN



脳全体レベルシミュレーション

Kenji Doya
Shiro Usui
Shinya Kuroda
VSM (SC-Retina-STDP)

Ryohei Kanzaki
IOSSIM-NEURON-K+

ロボット応用 医療応用

3

神経形態シミュレーション(NeuroMorphoKit)

Nonaka, Honda, Ishii, Neural Networks, 24(9), 2011

細胞膜(エネルギー緩和)

移動境界条件の設定

シグナル伝達(反応拡散)

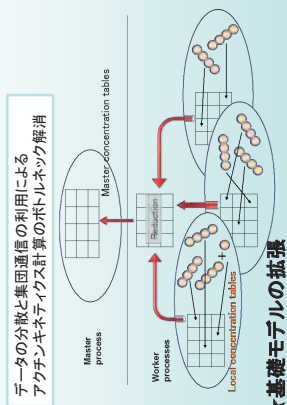
機能的な作用

細胞骨格(確率イベント)

キネティック制御

★京上での並列実装のパフォーマンスチューニング


データの分散と集団通信の利用による
アクチンキネティック計算のボトルネック解消



Master process Worker processes Local concentration tables Master concentration tables

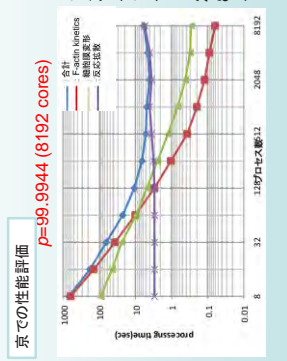
★基礎モデルの拡張

ファイメントの力学的曲げ・結束の導入
シグナル伝達系 (F-actin, Myosin, Formin)



膜モデルの三次元化

3D Membrane representation (M) Membrane (M) Actin filament (A) Add (New) (A) 4



京での性能評価

$p=99.9944$ (8192 cores)

処理時間 (秒)

1000 100 10 1 0.1 0.01

8 32 128 512 2048 8192

合計 F-actin filaments 細胞膜形状 反応拡散

1. 1,120万アクチンフィラメント、二次元200x200コンパートメント、膜ノード数100種類の細胞形態変化シミュレーションを実現

2. 大規模並列環境(京8192並列、RICC 4096並列)での動作を実現

A-Cellによる転写因子NF-κBシミュレーションの高速化

背景と目的

- NF-κBはCNSに発現しており、活性化がシナプスや樹状突起で進行。
- GultamateでNF-κBが活性化され、また細胞内Ca²⁺濃度上昇はその活性化に十分。
- NF-κBの活性化は細胞接着タンパク質、NOS、BDNF、CamKIIαの遺伝子が発現。
- NF-κB欠損マウスでは放射線治療学習に異常(図1)。
- 核内NF-κBが振動することが知られるが、これまで3Dシミュレーションは存在しなかった。そこで空間パラメータの影響を調べるためにA-Cellによる3Dシミュレーションを行う。

方法

- 核周辺の空間パラメータを明らかにするため、神経細胞の細胞体を模した球形細胞で検討(コンパートメント数: 約400,000、図2)。
- IKK活性化→IκB分解→NF-κB活性化→NF-κB核内移行
- A-Cell高速化のためにこれまでのMPIバージョンをモデル化(図2)。
- かつモデル非依存なMPI高速化を実施。

結果

- 核細胞質体積比、拡散定数、核膜輸送量、IκB翻訳場所の各空間パラメータすべてが振動パターンに影響した(図3)。

今後の展開

- MPI parallelization of A-Cell>99.99% (Ohshima, D., et al., PLOS ONE, 2012, e46811)
- より現実的な細胞形態やオルガネラの影響を明らかにし、かつ並列化率のさらなる向上を図る。

局所回路シミュレータ(NEST)

guided by neuroscientific goal to combine microscopic and macroscopic structure

Step	Description	Network type	Neurons	Synapses	Status Dec 2012
1	reduced primate visual cortex (32 areas, each 1 local microcircuit)	random at micro and macro scale	3·10 ⁶	3·10 ¹⁰	technically achieved on K on 28. Aug 2011
2	reduced primate visual cortex respecting relative size of areas (32)	spatially organized at micro scale	10 ⁸	10 ¹²	random network of 10 ⁶ neurons and 10 ¹² synapses achieved on K on 15. Sep. 2012
3	human brain model (approx. 100 areas)	spatially organized	max 10 ¹¹	max 10 ¹⁵	

Expected results:

- Step 1: investigation of reduced visual cortex model now possible on K
 - characterization of stationary state (rates, irregularity)
 - comparison of low-frequency EEG and LFP to experiments
- Step 2: models with closed functional circuits
 - capability to simulate full scale model of a functional unit at microscopic resolution
- Step 3: full human brain simulation
 - assessment of capability to simulate full scale brain model

NESTの京上での性能

(1) Runtime improvements

- speedup $\alpha = 0.59$ @ 65536 cores (strong scaling)
- hybrid MPI + OpenMP code (simulation and wiring)
- achieved phase V1 on K

(2) Memory improvements

- storage of neurons and synapses with heterogeneous plasticity rules employing **sparse representations** enables efficient distribution on M≥100k cores

```

1 used synapse type and <3 local synapses of this type
sub-pl_synapse_hom: 40 B
target_pointer: 8 B
rport: 8 B
weight: 8 B
delay: 8 B
STDP_K_plus: 8 B
            
```

(3) production network 5·10⁸

- 1 second biological time
- 196608 cores
- 26 minute simulation time

(4) Next milestone: N_{max}=10⁹

- theory of memory usage
- possible with present code's memory consumption on full machine size of K (end of red broken line)

回路シミュレーションによる大脳皮質情報処理の解明(CMDN)

1) 層構造をもつ視覚野局所回路の構築
 Potjans and Diesmann and Fukai (2012) Cereb Cortex

2) それを用いた視覚的注意の2カラムモデル
 Wagatsuma, Potjans, Diesmann and Fukai (2011) Front Comput Neurosci

3) 空間注意(上)と特徴注意(下)の8カラムモデル

実験

Bottom-up Sensory Input
 Top-down Spatial Attention
 Top-down Feature-based Attention
 Selective Propagation to the Visual Column

シミュレーション

Layer 5
 Layer 5

4) 局所回路モデルのためのデータ解析

- 大脳皮質回路と海馬(Colein et al., Nature 2009)の計算原理の共通性を発見(シナプス-カンマ波の結合)

全視覚系シミュレーション(VSM)

視覚系サブモデルの構築
視覚系を構成する各サブシステムレベルでの視覚機能解析

モデル統合環境の構築(PLATO)
言語やシステムによらないモデル統合ライブラリと実行Agent

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
全視覚系モデルは、それを構成する各部位の詳細モデルの統合による構築を目指すことから、それに必要なプログラム言語に依存しないモデル統合環境(PLATO)を構築

大規模視覚系モデルは、視覚系を構成するサブシステムモデルをPLATO上で統合して構築

1. 統合は、MPI/netCDFの共有フォーマットにより実現
2. 統合したモデルはAgentにより、入出力などの管理のもとに実行される。

全視覚系モデルの構築
全視覚系モデルシミュレーションによる包括的な視覚機能解析

Sub-model (parts)

- We should share
 - data format
 - source code
 - I/O data
 - parameters
 - hardware

integrate

Large scale model

Retinal image Cone mosaic Response of BC-GC

各視覚系サブモデルの出力(応答)

眼球運動系

網膜-上丘-眼球運動制御ループのスパイクニューロンモデルをNEST上で実装

- 眼球運動について行動レベル、神経活動レベルで性質の再現に成功 (論文 in revision)
- 視覚刺激により駆動されるサッカドのシミュレーションを、RICCとATR (京都)にあるヒューマノイドロボットを接続して実施

上丘内神経細胞の活動

SC Model, RICC, ATR, Eye Motor Driver, Image Grabber, Internet, Output Colony, Input Colony

神経活動・サルのバースト細胞の活動(右)をモデルは良く再現できる(左)

行動: サッカドの角度に対して、速度ピークは非線形となり(右)、運動時間は線形となる(左)

無脊椎動物嗅覚系シミュレーション(IOSSIM)

目標: モデル生物であるカイコガの脳を対象とし、嗅覚入力から行動出力に至る神経ネットワークをニューロンレベルの詳細モデルとして構築し、実験と比較しながら脳機能を予測し、理解するためのシミュレーションをおこなう

分析(実験)

カイコガの脳。約10万個のニューロンからなり、約1万個が嗅覚系に關与する。

Bilateral Neuron

Local interneuron

平均応答率 (BIN: 10ms)

統合(モデル化)

単一ニューロンから神経ネットワークの構築

脳内への嗅覚入力から行動出力までの経路

シミュレーション

単一ニューロン活動のシミュレーション

神経ネットワーク活動のシミュレーション

IOSSIMシミュレータの構築

京上のチューニング

Single core performance (N = 256)

- Tuning with completions
- Substitution
- change number of core
- Cache Optimize
- Cache size tuning

細胞膜情報推定

遺伝子

Multi Compartment

マルチコンパートメントモデル

マルチコンパートメントモデル等面回路方程式

電位依存性イオンチャネルのコンダクタンス

形態の円柱近似

本構造により接続関係を表現




ソフトウェアの達成状況

	RICC上の達成並列(コア)数	京上の達成並列(コア)数	京上のストロングスケール(測定コア数)	ソフトウェアの最大スケール
神経細胞シミュレーション NeuroMorphoKit	4096	8192	0.994(8192) Hybrid	アクチン線維数 1.2×10^6
局所回路シミュレーション NEST	8000	196608	0.59(65536) Hybrid	細胞数 5×10^8 シナプス数 3×10^{12}
全視覚系シミュレーション VSM	2460 (whole system 1024)	1504	0.89 (RICC 1024) MPI	視細胞(錐体)数 1.5×10^6
昆虫嗅覚系シミュレーション IOSSIM	8192	196608	0.83(196603) Hybrid	コンパートメント数 10^4 細胞数 8×10^5

13




学術的成果



国際学術専門雑誌掲載論文(予定を含む)件数

	2008(OCT-)	2009	2010	2011	2012~
神経細胞シミュレーション	2	3	2	9	5
局所回路シミュレーション	0	4	6	8	11
全視覚系シミュレーション	0	0	0	2	2
昆虫嗅覚系シミュレーション	3	3	1	5	4

論文掲載ハイライト

Helias, et al. (2012), Frontiers in Neuroinformatics, 6(26).
 Nishiyama, et al. (2011). Nature Cell Biology, 13(6).
 Linden, et al. (2011). Neuron, doi:10.1016/j.neuron.2011.11.006.
 Honda, et al. (2011). Journal of Neuroscience, 31(4).
 Fujii, et al. (2011). Proceedings of National Academy of Science USA, 108(44).
 Toriyama, et al. (2010). Molecular Systems Biology, 6(394).
 Isomura, et al. (2009). Nature Neuroscience, 12.

14

細胞の形態形成におけるマルチフィジックス ~シグナル伝達/骨格/膜~

細胞内シグナル伝達

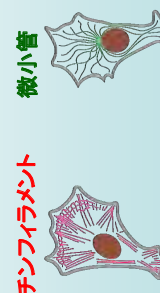
- 細胞外シグナルを情報処理
- 骨格系を制御

細胞骨格

- アクチンフィラメント・微小管
- 機械的に膜を支持

細胞膜

- 細胞骨格の再編に依存して、柔軟に変化
- 細胞内化学反応の境界条件





細胞膜直下に多く存在 中心体から放射状に存在

各要素を統合してシミュレーションする必要がある

by Dr. Kaoru Karoh

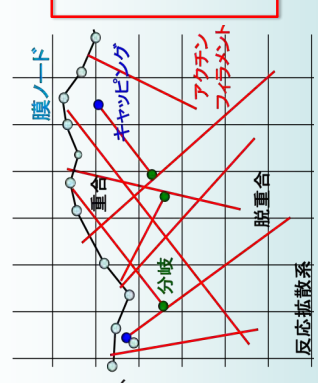
15

NeuroMorphoKit の概要・アプローチ

計算モデル

- 細胞内シグナル伝達は、二次元格子化された空間で化学反応拡散計算を行う。
- アクチンフィラメントは線分で表現、コンパートメントに存在する各種分子の濃度に応じて確率的に重合・脱重合、キャッピング、分岐を含むイベントを生成する。
- 細胞膜は膜ノードを頂点としたポリゴンで表現、エネルギー最適化により形態を求める。



計算内容

拡散 および 反応

重合・脱重合、キャッピング、分岐を含む

アクチン繊維の更新

重合・脱重合、キャッピング、分岐による

細胞膜ノード

アクチンフィラメント活性、膜張力、体積活性、表面積変化に関するエネルギーの最小化

並列化の方法

- 膜エネルギー最適化、アクチンフィラメントのキネティクス計算
- OpenMP, MPIによるハイブリッド並列化

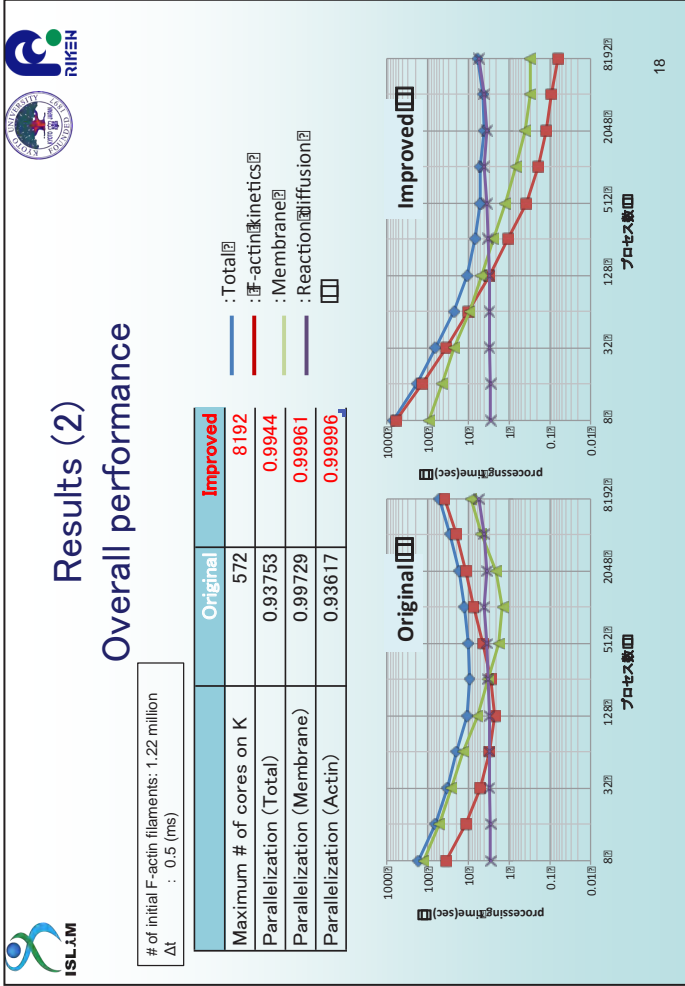
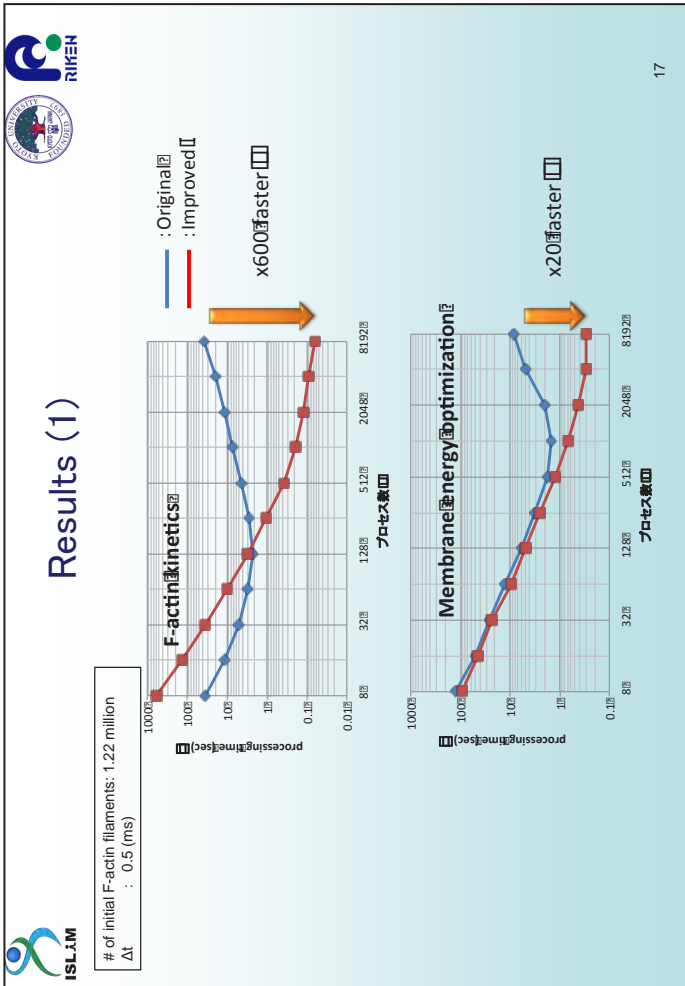
使用する計算法

- 最急降下法
- Runge-Kutta法

開発言語と主なライブラリ

- C/C++
- MPI, OpenMP, GSL, NetCDF

16



Filopodia formation

Filopodia in the growth cone
Actin (red), Microtubule (green)
Dehmelt et al., 2004

Question
How does the growth cone produce filopodia?

in vitro reconstruction of filopodia
Filopodia-like structure appears in a membrane-free environment only with actin, Arp2/3 and fascin.
Ideses et al., PLoS ONE, 2008

Question
How is the filopodia-like structure self-organized by minimum components?

Simulation results

Simulation:

Simulation:

Experiment in vitro:

Network:

Fascin [μM] vs Arp 2/3 [nM] plot showing the relationship between these two proteins.

Ideses et al., PLoS ONE, 2008

Our model successfully generated various types of filamentous structures.

NEST による局所回路シミュレーション

五十嵐 潤

次世代計算科学研究開発プログラム
脳神経系研究開発チーム 特別研究員






発表者紹介

2007年3月 九州工業大学 生命体工学研究科 脳情報専攻 博士課程修了
2007年4月 九州工業大学 生命体工学研究科 脳情報専攻 研究員
2008年7月 ホンダ・リサーチ・インスティテュート・ジャパン 客員研究員
2009年4月 次世代計算科学研究開発プログラム 特別研究員

研究分野

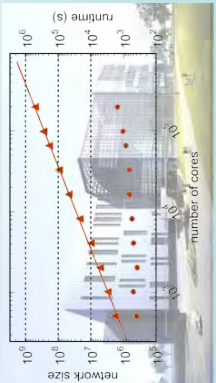
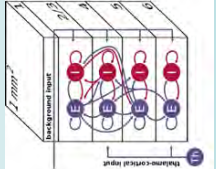
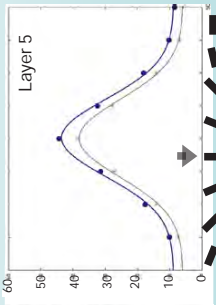
計算論的神経科学、大規模神経回路シミュレーション、神経科学

NESTによる局所回路シミュレーション




五十嵐 潤¹、Moritz Helias^{1,2}、舩本 現³、Tobias C Potjans^{1,2}、Markus Diesmann^{1,2}、深井 朋樹¹

¹理化学研究所 次世代計算科学研究開発プログラム 脳神経系研究開発チーム、² Inst of Neuroscience and Medicine (INM-6), Computational and Systems Neuroscience, Jülich Research Centre, ³ 理化学研究所 次世代計算科学研究開発プログラム 高度化チーム

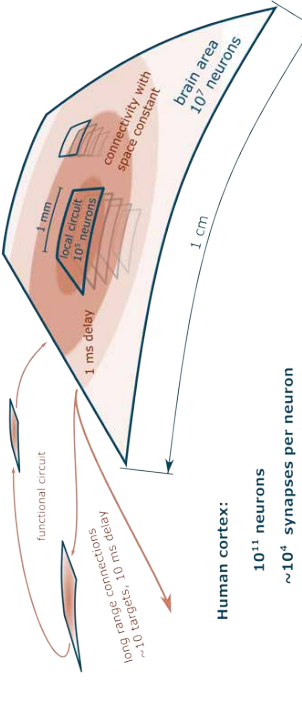




ISLIMソフトウェア研究開発報告会

1

人間の脳の構造






Human cortex:
10¹¹ neurons
~10⁴ synapses per neuron

- 脳はネットワークのネットワーク: 局所回路の結合、領域内の結合、領域間の長距離結合
- 近接領域の神経細胞からのシナプス入力は50%、長距離結合が50%
- 脳機能は多くの領域に分布している
- 部分的な領域のモデルでは完全に入力を表現できず、予測性能が制限される。また、脳の機能を再現するには全規模のモデルが必要

ISLIMソフトウェア研究開発報告会

2

全規模のシミュレーションの問題

- メモリオーバーヘッドが並列数とともに上昇する
- 計算時間ではなく、メモリ量がネットワークサイズの上限を決める
- 全メモリ資源を使う: 最大限にメモリを使用し規模を拡大
- メモリー消費の数理モデルをベースとした解析
 - Kunkel S, Potjans TC, Eppler JM, Plesser HE, Morrison A and Diesmann M (2012) Meeting the memory challenges of brain-scale network simulation. *Front Neuroinform* 5:35
 - 異なるスケールでは、メモリ消費占有する部分が、ソフトウェアの異なる構成部分になる
- 共通のプラットフォームの必要性
 - 脳のモデルは非常に複雑で、作成、デバッグには大きな労力
 - 他者によって発表されたモデルの再現は容易ではない

ISLIMソフトウェア研究開発報告会

3





NEST シミュレータ




重要な目標:

- 体系的に新しいシミュレーション技術を公開する
- GPLの一般なリリースで複製


- 多数の研究室との共同研究 (2001年~)
- 国際的な高等課程おける教育
 - 沖縄計算論的神経科学コース
 - Advanced Course in Computational Neuroscience ACCN, ヨーロッパ

ISLIMソフトウェア研究開発報告会


4



ISLAM



nest::
initiative




JULICH
FORSCHUNGSZENTRUM
RIKEN

10⁹ の神経細胞からなる局所回路モデルにむけて


- 京における大規模シミュレーションの3つのマイルストーン
 1. 京へのNESTの移植 第2世代 * 2011年9月 NEST 2.1
 2. 10⁸ の神経細胞の実現 第3世代 2012年5月 NEST 2.2 (次期リリース)
 3. 全脳規模 (10⁹) にむけて 第4世代 2012年9月

*第2世代 = 第2世代 シミュレーションカーネル


- 10⁸ に相当する規模:
 - 最大視野 (第一視覚野) よりも大きい規模
 - 視覚皮質のモデルが可能
 - より大きなネットワーク: 大きな遅延、疎な巨視的結合



ISLAM




nest::
initiative




JULICH
FORSCHUNGSZENTRUM
RIKEN

ISLIMソフトウェア研究開発報告会


5



ISLAM



nest::
initiative



JULICH
FORSCHUNGSZENTRUM
RIKEN


第3世代、第4世代カーネルのメモリー配置

第3世代メモリーレイアウト

- 神経細胞やシナプス結合のデータ構造におけるスパースさに対応する

第4世代メモリーレイアウト


- シナプスの性質の"不均一さ"の表現に対応する
- 10,000コアでは、数個のターゲットを持つ神経細胞は大きなオーバーヘッドを生じる
- 新規のデータ構造によって、シナプスの一般的な性質は保ちながら、短いターゲットのリストに効率的に対応



used synapse type and target neurons

neurons with local targets


→ 3 local synapses of this type




used synapse type and target neurons

neurons with local targets


→ 3 local synapses of this type



ISLAM




nest::
initiative




JULICH
FORSCHUNGSZENTRUM
RIKEN

ISLIMソフトウェア研究開発報告会


6



ISLAM

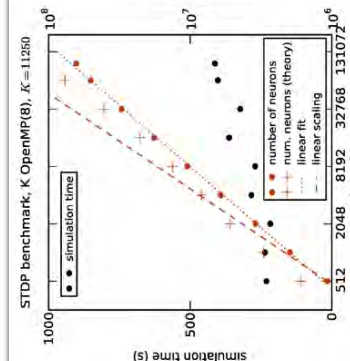


nest::
initiative



JULICH
FORSCHUNGSZENTRUM
RIKEN

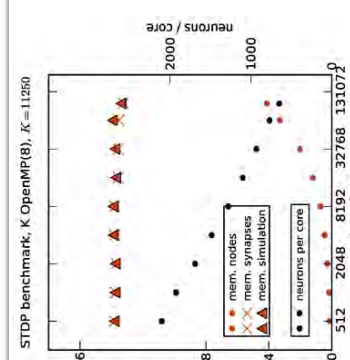
ソフト研究開発成果(NEST): 京における第3世代のモデル規模の拡大



STDP benchmark, K OpenMP(8), K=11250

simulation time (s)

number of cores




STDP benchmark, K OpenMP(8), K=11250


memory / node (GB)

neurons / core


- 神経科学者は利用可能な資源の最大使用量が知りたい
- ノードあたり利用可能なメモリ量 (16GB) を最大限に用いる
- 1コアあたりの神経細胞数
- シミュレーション時間は、通信時間の増加が原因で、増加する。



ISLAM




nest::
initiative




JULICH
FORSCHUNGSZENTRUM
RIKEN

ISLIMソフトウェア研究開発報告会


7



ISLAM

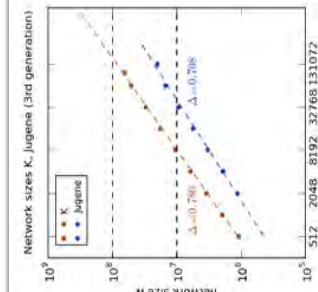


nest::
initiative



JULICH
FORSCHUNGSZENTRUM
RIKEN

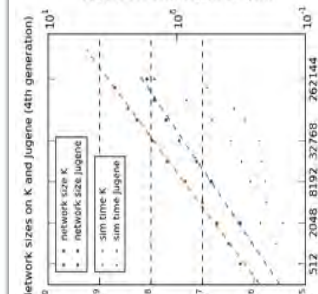
ソフト研究開発成果(NEST): 京とJugeneiにおける第3、4世代カーネルの性能比較



Network sizes on K, Jugenei (3rd generation)

network size N

number of cores




Network sizes on K and Jugenei (4th generation)


network size N

number of cores


- 波線の右端は、京の可能なマシンサイズを示す
- JUGENEを用いて、MPIジョブの最大使用数、65,536 = 262,144 coresまでを試験
 - 京の最大構成の74%の MPIジョブに相当
- 京を用いて、MPIジョブの最大使用数、24,576 = 196,608 coresまでを試験を
- 京で第4世代カーネルを用いて、生物学的1秒のシミュレーションに、1520秒=25分で達成



ISLAM



nest::
initiative




JULICH
FORSCHUNGSZENTRUM
RIKEN

ISLIMソフトウェア研究開発報告会

8



ISLIM



nest::
ANALYZE



JÜLICH
FORSCHUNGSZENTRUM



RIKEN

ソフト研究開発成果(NEST) : 大脳皮質の最小構成単位となる階層型局所回路モデル



in vivo anatomy




in vitro physiology


- 1 mm³に相当
- 10万神経細胞、10億シナプス
- 各層は興奮性(80%)と抑制性の神経細胞集団(20%)からなる
- 興奮と抑制は同一な力学的性質の神経細胞モデル
- 水平方向に一樣な結合
- 電気生理学と解剖学データに基づき、各層の細胞種ごとに結合確率を設定

ISLIMソフトウェア研究開発報告会


9




ISLIM



nest::
ANALYZE

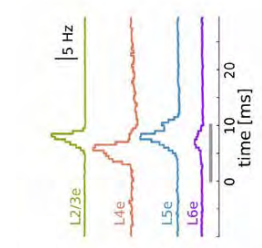


JÜLICH
FORSCHUNGSZENTRUM

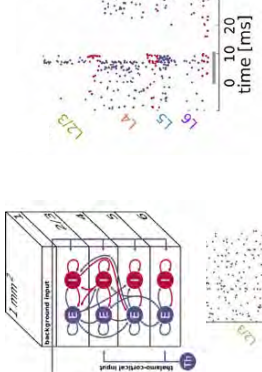


RIKEN

ソフト研究開発成果(NEST) : 各層の神経細胞の振る舞いを再現

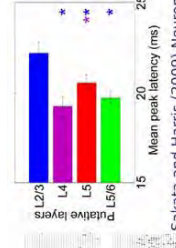


5 Hz

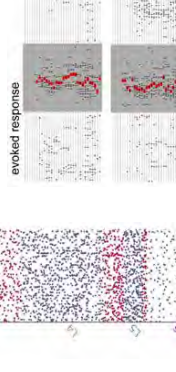


以下のような実験結果を再現

- 神経細胞の非同期的で不規則な発火活動
- 各層の正確な発火頻度
- 刺激に対する層間の発火順序



Mean peak latency (ms)




evoked response

Potjans TC & Diesmann M (2012)
The cell-type specific connectivity of the local cortical network explains prominent features of neuronal activity. *Cerebral Cortex* 10:1093/cercor/ahs358


Sakata and Harris (2009) Neuron

ISLIMソフトウェア研究開発報告会


10




ISLIM



nest::
ANALYZE




JÜLICH
FORSCHUNGSZENTRUM



RIKEN

ソフト研究開発成果(CMDN) : 視覚的、特徴的注意の神経活動への影響



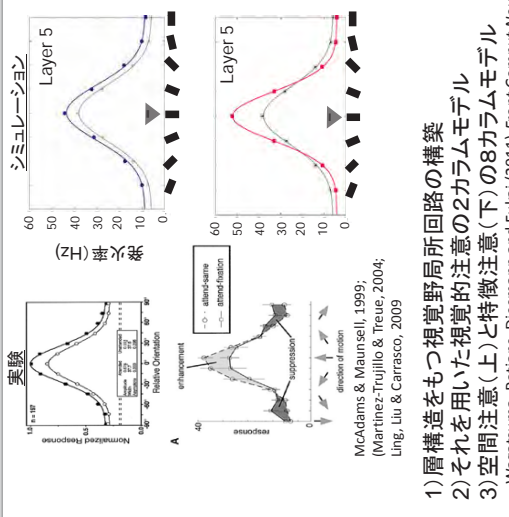
Bottom-up Sensory Input
Sensory input is mainly projected to a receptive field (RF) corresponding visual column.

Top-down Spatial Attention
Top-down Attention

Top-down Feature-based Attention
Homogenous Projection to the Subpopulation

Top-down Attention
My target is II

Selective Projection to the Visual Column



シミュレーション

Layer 5

Layer 5


(上) 樹文紙

- 1) 層構造をもつ視覚野局所回路の構築
- 2) それを用いた視覚的注意の2カラムモデル
- 3) 空間注意(上)と特徴注意(下)の8カラムモデル


Wagatsuma, Potjans, Diesmann and Fukai (2011) *Front Comput Neurosci*

ISLIMソフトウェア研究開発報告会


11




ISLIM



nest::
ANALYZE




JÜLICH
FORSCHUNGSZENTRUM



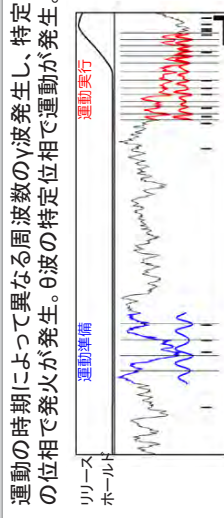
RIKEN

ソフト研究開発のための実験解析(CMDN) : 運動皮質のθ、γ帯の神経的な振動現象



ラットの自発運動時の運動皮質を測定

- ・脳波
- ・神経発火



運動準備

運動実行

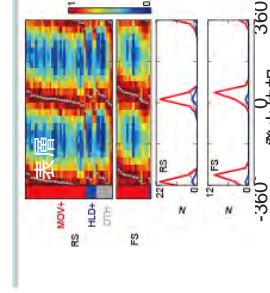
リリース

ホールト

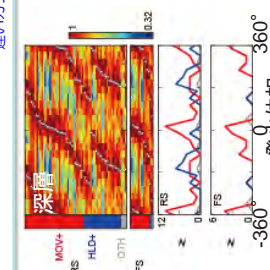
遅いガンマ波

速いガンマ波

運動の時期によって異なる周波数のγ波発生し、特定の位相で発火が発生。θ波の特定位相で運動が発生。



表面



深層

表面、深層の神経細胞では、γ、θ波に対して、発火の位相固定に顕著な違い

神経基盤と情報処理での役割は何か？

ISLIMソフトウェア研究開発報告会

12

Multi-scale simulation of subcortical visual systems with real-time robotic visuo-motor feedback

Jan Morén

Brain and Neural Systems Team,
Integrated Simulation of Living Matter Group,
RIKEN Computational Science Research Program



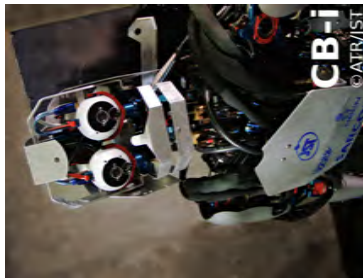
発表者紹介

1995-11	MSc Computer Science, Lund, Sweden
2002-09	PhD Cognitive Science, Lund, Sweden
2002-06	Research Assistant, Lund, Sweden
2003-11	Post-Doc, CRL/NICT, Kyoto, Japan
2005-05	Visiting Researcher, NICT, Kyoto, Japan
2006-09	Researcher, ATR, Kyoto, Japan
2009-02	Researcher, Kyoto University, Kyoto, Japan
2012-11	Researcher, OIST, Okinawa, Japan

研究分野

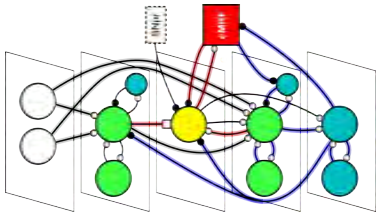
Integrated Systems Biology Laboratory, Department of Systems Science, Graduate School of Informatics, Kyoto University

Multi-scale simulation of subcortical visual systems with real-time robotic visuo-motor feedback



Jan Morén
Tomohiro Shibata
Norikazu Sugimoto
Kenji Doya

Kyoto University
NAIST
ATR
OIST

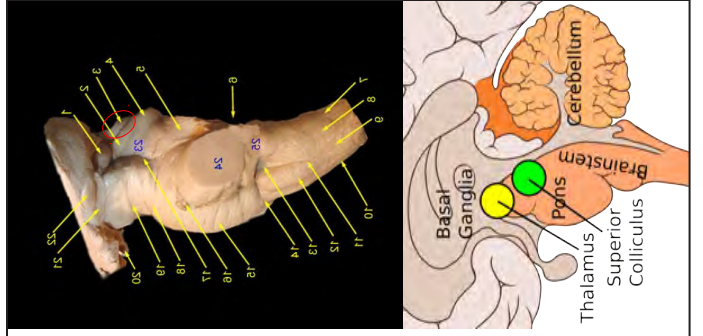
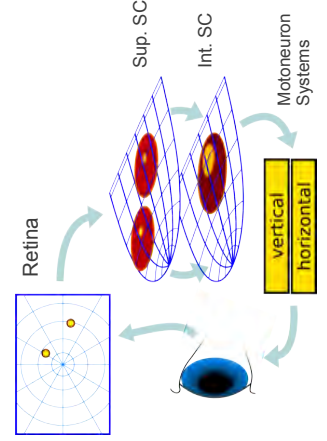


Outline

- Goals, Introduction
- The Superior Colliculus
- Model Simulation
- Connected to CB-I robot

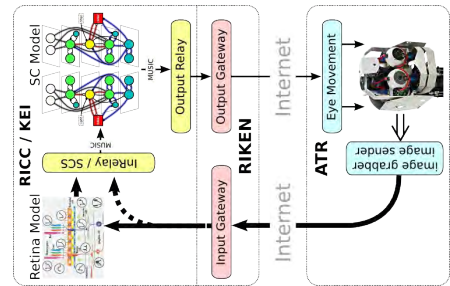
Superior Colliculus

- Subcortical Saccade Generation Loop
- Retinotopic
- Superficial SC: Sensor Processing
- Intermediate SC: Motor Command
- Burst output give eye-velocity profile



Project Goals

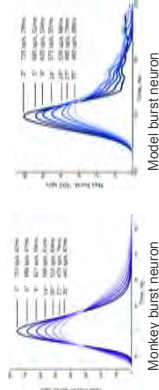
- Close the Perception-Action loop
- Large-Scale neuron-level Simulation
- Use a Humanoid Robot to ground Model in the World



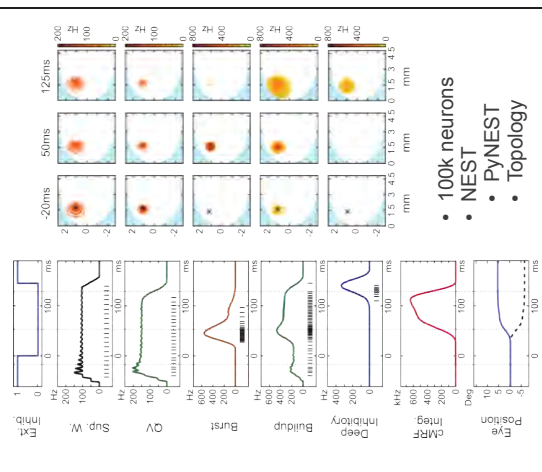
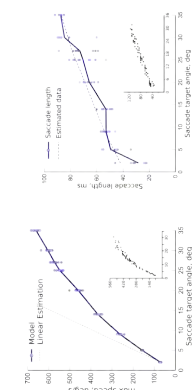
Simulation Results

- Tune Burst neuron connections

- Primate burst neuron data
- Constant spikes

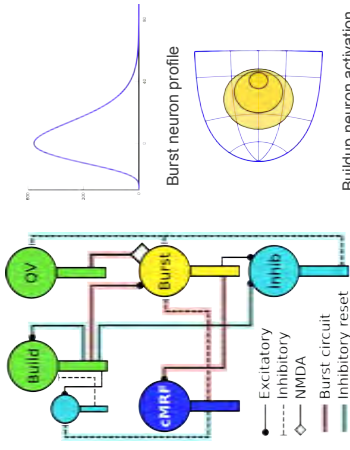


- Complex log output weights give Main Sequence relationship

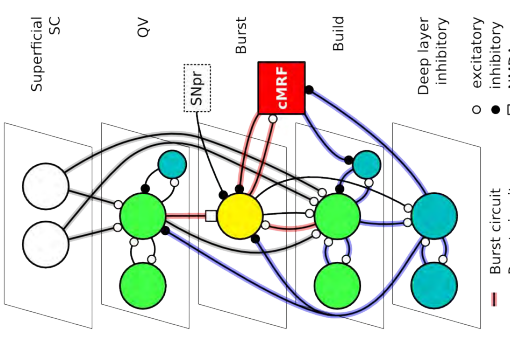


- 100k neurons
- NEST
- PyNEST
- Topology

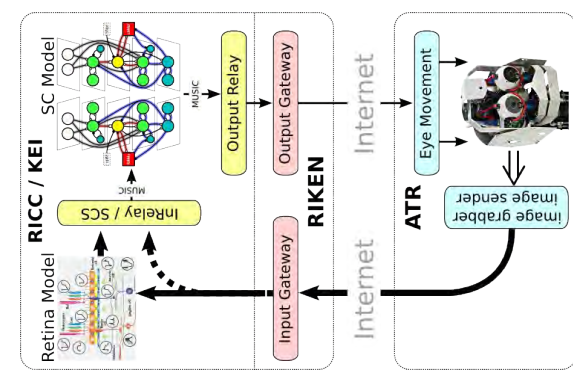
Model



- NMDA receptor bursting activation
- cMRF spike integrator reciprocal inhibition
- Buildup neuron disinhibition causes spreading activation
- Activation-triggered deep layer system-wide inhibition

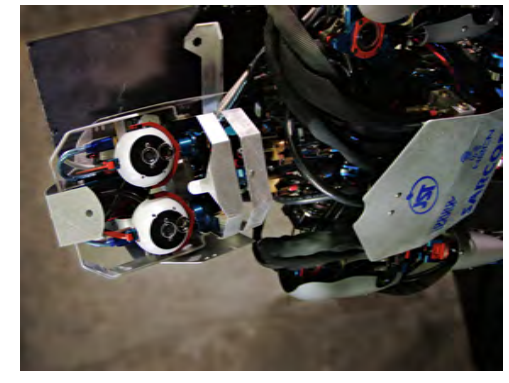


Closing the Loop



- Connect CB-i Robot to Model As Physical Eyes
 - Camera — Visual Input
 - Eyeball — Motor Output
- Connect Over Distance
 - Wako-Shi ←→ Kyoto
- Near Real-Time control
 - React Directly to Input

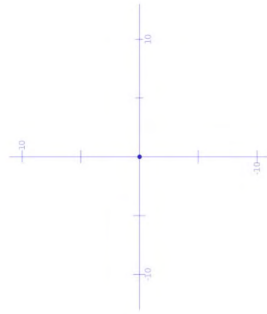
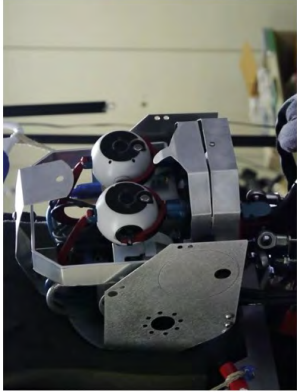
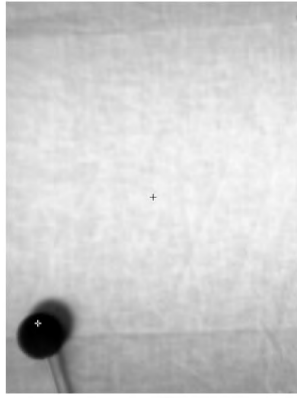
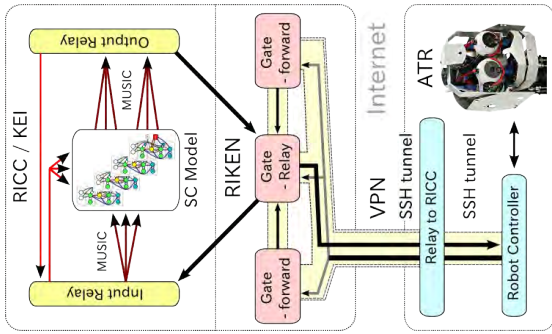
CB-i Humanoid Robot



- Height: 155cm
- Weight: 85kg
- DOF: 51
- Power: Electric, Pneumatic, Hydraulic
- Eyes: 2x Wide-field, 2x Foveal, 2x 2 DOF, Firewire
- Maker: SARCOS & ATR

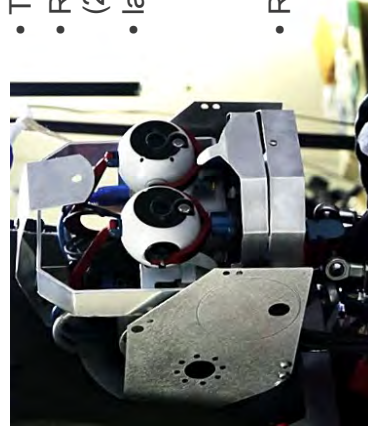
Closing the Loop

- **RICC**: 256-1024 Nodes
 - Scaled down, double-sided model implementation
 - MUSIC Interconnections (MULTI Simulation Coordinator)
- **Internet Connection**
 - VPN connection
 - SSH over TCP/IP
- **ATR**
 - CB-i in test harness
 - Motor system
 - Video Saliency Extraction



Results

- Track, saccade towards target
- Run at ~2x real time (256 nodes)
- latency ~150ms roundtrip ~110-140ms without intra-model delays
- Reactive, Usable Behaviour



Robots and Clusters are useful together for Real-Time, Real-World interaction

Conclusion

- We have built a large-scale, spiking neuron-level model of the Superior Colliculus
 - Reproduces burst neuron output profiles
 - Saccade *Main Sequence* behaviour
 - Suggest spreading activity amount (but not direction) tracks saccade end time
- Real-time distributed visuomotor control of robot by large-scale cluster model
 - Practical approach to test computationally intensive models in the real world

大規模脳数理モデル構築プラットフォーム による視覚系数理モデルの構築

臼井支朗

豊橋技術科学大学・エレクトロニクス先端融合研究所 特任教授
理化学研究所・脳科学総合研究センター
神経情報基盤センター (客員主幹研究員)





発表者紹介

- 1974年1月 カリフォルニア大学大学院博士課程単位取得満期退学
- 1974年3月 名古屋大学工学部助手
- 1974年6月 Ph.D (カリフォルニア大学)
- 1979年4月 豊橋技術科学大学講師
- 1980年4月 豊橋技術科学大学助教授
- 1986年4月 豊橋技術科学大学教授
- 2002年4月 理化学研究所脳科学総合研究センター・ニューロインフォマティクス
技術開発チーム チームリーダー
- 2003年3月 豊橋技術科学大学退職
- 2003年4月 豊橋技術科学大学 名誉教授 客員教授
- 2005年4月 理化学研究所脳科学総合研究センター・神経情報基盤センター
副センター長兼任
- 2007年4月 同センター長兼任
- 2012年10月 豊橋技術科学大学エレクトロニクス先端融合研究所 特任教授
理化学研究所脳科学総合研究センター・神経情報基盤センター
客員主幹研究員兼任

研究分野


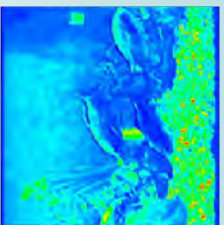
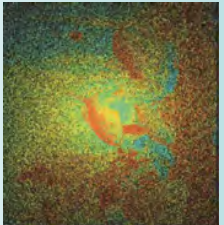
ニューロインフォマティクス、神経情報工学、視覚生理工学

大規模脳数理モデル構築プラットフォームによる 視覚系数理モデルの構築



臼井支朗^{1,2,3}, 稲垣圭一郎¹, 観音隆幸², 神山齊己⁴, 石原彰人⁵

1. 理化学研究所・次世代計算科学研究開発プログラム
2. 理化学研究所・脳科学総合研究センター
3. 豊橋技術科学大学・エレクトロニクス先端融合研究所
4. 愛知県立大学
5. 中京大学


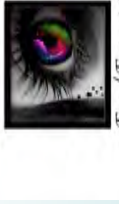

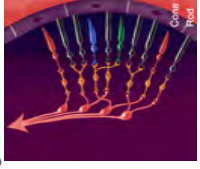
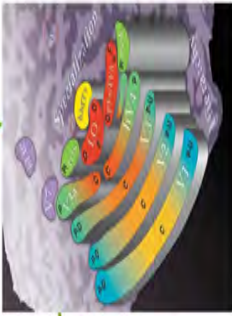




ISLIM成果報告会2012

1



背景・目的

大規模視覚系数理モデル構築
錯視や注意など視覚機能を
シミュレーションにより評価


ISLIM成果報告会2012

2


概要・アプローチ

サブシステムモデル

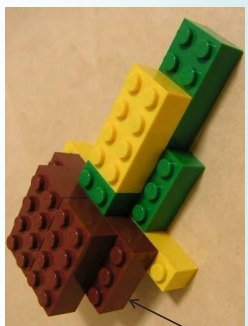


parts

integrate!



大規模モデル



We should share

- data format
- source code
- I/O data
- parameters
- hardware



plug in
module

- realistic level
- computational level
- functional level
- abstract level

Platform for collaborative brain system modeling : **PLATO**
 モデル統合環境を構築し、サブシステムモデル統合による大規模モデル開発を実現
 → IOを共通化することで、統合に用いるサブシステムモデルの、言語や記述レベルを吸収し
 統合ライブラリと実行エージェントによる統合シミュレーション管理を実現

ISLIM成果報告会2012

3

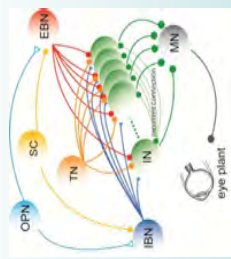



研究開発成果：眼球運動・眼光学モデル

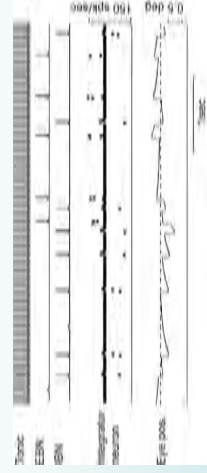
- 物体注視時の眼球運動の再現とそれを考慮した詳細な網膜像の再現

- 眼球運動モデル：微小眼球運動ならびにサッカードの再現


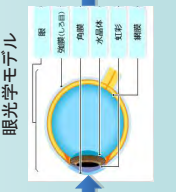
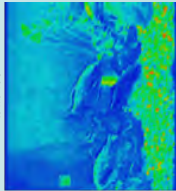
解剖・生理データに基づく脳幹の
眼球運動制御神経回路の記述



電気生理データに基づく細胞パラメータの推定
と微小眼球運動の再現



Inagaki et al., Neural Networks, 2011

- 眼光学モデル：網膜像の生成

波長感度、波面収差、瞳孔など
眼球システム特性を考慮して、
詳細な網膜像を計算
光の各波長に対して眼球レンズの
屈折率が異なることから、10nm毎
に網膜像を計算・統合

ISLIM成果報告会2012

研究開発成果: 眼球運動・眼光学モデル

Response of brain stem neuron and microsaccade
Inagaki et al., Neural Networks, 2011

Retinal image
Cone Mosaic
Response of V1 mode

眼球運動モデル:

- サッカードに関係した脳幹の神経回路網をスパイクニューロンモデルで記述、サッカー、マイクロサッカド、ドリフト眼球運動を再現
- 京、RICCにて30並列(モデル規模とステップ刻みによる実行時間の制限の為)。

眼光学系モデル

- 1000x1000(pix)の自然画像を入力
- モデルは京、RICCで400並列(OpenMP+MPI)。
- 並列化効率(parallel portion)は0.99305

ISLiM 成果報告会2012

研究開発成果: 網膜モデル

- 網膜の活動電位・神経スパイクを再現
- 各細胞層のモデルを詳細に構築した後、PLATOにて他のサブシステムと統合

神経節細胞 Y-type X-type
双極細胞
水平細胞
視細胞

123x123 (2460x2460 μm , approx. 10deg) | 6000x6000 modeling

研究開発成果: 視細胞モデル(Cone Mosaic)

網膜視細胞(錐体モザイク)モデル

Input Image
Cone Mosaic

- ✓ 3600万個の視細胞により直径30度視野角を表現
- ✓ 同視細胞層に錐体は150万個存在
- ✓ 実験データに基づいた χ^2 -fittingにより錐体、桿体の空間配置を再現
- ✓ L.M.S錐体の空間存在密度に関する実験報告に基づいて存在割合を再現。

ISLiM 成果報告会2012

研究開発成果: 双極細胞層-神経節細胞層モデル

双極細胞層-神経節細胞層のモデル

錐体モデル (Kamiyama & Sone, 2010) V_{cone}

双極細胞 C_m , g_m , g_m , E_s , E_m , E_s

Parallel conductance modelで記述 V_{mbr}

イオン電流モデルで記述 (Fohlmeister & Miller, 1997) V_{ngc}

視細胞層 (Cone mosaic) 150万
双極細胞層 150万
神経節細胞層 40万

± 30 deg

ISLiM 成果報告会2012

研究開発成果: 双極細胞層-神経節細胞層モデル

双極細胞層-神経節細胞層のモデル

継体モデル (Kamiyama & Sone, 2010)

双極細胞

$$C_m \frac{dV_{cone}}{dt} = g_m(V_{bc} - E_m) + g_s(V_{bc} - E_s) - g_m(V_{bc} - E_m)$$

Parallel conductance modelで記述

イオン電流モデルで記述 (Fohlmeister & Miller, 1997)

Synaptic transmission

$$T_i = \frac{(V_{cone} - a)^n}{(V_{cone} - a)^n + K_m^n}$$

glutamate receptor

$$\frac{dR}{dt} = \alpha \cdot (1 - R) \cdot \sum_i w_i T_i - \beta \cdot R$$

membrane potential

$$g_s = \bar{g}_s \cdot R$$

$$C_m \frac{dV_{bc}}{dt} = g_s(V_{bc} - E_s) - g_m(V_{bc} - E_m)$$

研究開発成果: 網膜モデル

視細胞層(コーンモザイク)モデル

- 3600万の視細胞モデルのうち150万の錐体細胞から構成される
- 波長感受特性の異なる3種(L:long, M:middle, S:short cone)の分布を再現
- 生理データに基づいて各錐体細胞の光感受度応答特性を再現

コーンモザイク

Fovea high cone density high resolution

Peripheral low cone density sparse resolution

網膜像に対する視細胞層の応答再現

解剖・生理データに基づく網膜神経回路構造

- 視細胞、水平細胞、双極細胞および神経節細胞層からなる網膜モデル構築
- 生理データに基づいて、各細胞層の電位応答を再現・可視化

研究開発成果: 網膜モデル

錐体モザイク

Realistic retinal image

Cone Mosaic

Response of BC-GC

Response of brain stem neuron and microcaccade (Inagaki et al., Neural Networks, 2011)

継体モザイク:

- 1000x1000(pixel)の網膜像(眼光学モデル)の出力を入力として利用
- 京で1504(node制限の為)までOpenMPとMPIのハイブリッドにて並列化
- 並列化効率(parallel portion)は0.99909
- 双極細胞-神経節細胞モデル(チューニング中)
- 1000X1000の錐体出力を入力
- 京にて1024並列までmpi並列化

Process time (sec) vs num of proc

モデル統合とシミュレーション

眼球運動-眼光学-網膜の統合モデル

external image & gaze position (30 core)

eye movement model

gaze position

eye optics model (400 core)

retinal image

cone mosaic model (1504 core)

Cone mosaic (L, M, S)



bipolar, ganglion model (1024 core)

cone mosaic (response in L, M, S cone)

e.g. cone mosaic with Platonic

眼球運動-眼光学-網膜(視細胞-双極細胞-神経節細胞)によりモデルを構成

- 各モデルを並列化し、PLATO上で統合
- 共有ファイルフォーマット(インタフェース)とPlatonic Libraryにより、IOレベルで結合
- 統合モデルはAgentにより、実行管理のもとシミュレーション
- 現状の問題点として、Agentとインタフェース間の通信により各モデルの並列性能が向上しても、統合モデル上で性能が向上しない(改善中)

まとめ

- モデル統合環境PLATOの構築
 - サブシステムレベルでモデル構築・並列化を行い、共有フォーマットにてIOレベルで結合することで大規模モデル構築を可能にした
- 眼球運動、眼光学、網膜からなる視覚系のモデルをPLATOにて構築
- サブシステムモデルレベルでは、眼球運動にて微小サッカード生成のメカニズムを Computational Levelで明らかにした
- 視細胞モデルならびに双極細胞－神経節細胞層における信号処理(細胞応答)を情報量を指標として評価中

本プロジェクトにて、上記成果を学術論文5本、国際会議13件、国内会議1件で報告した

今後の課題

- Platonicインタフェースにおける速度低下の改善
 - 統合モデルの速度向上
- 錐体モザイクで実施した情報量評価の他領域への適用
 - e.g. 微小眼球運動による視知覚の向上などの評価を進める

13




Fin

ISLiM成果報告会2012

14

「京」による 無脊椎動物嗅覚系シミュレーション ー背景と現状と将来への展開ー

神崎 亮平

次世代計算科学研究開発プログラム
脳神経系研究開発チーム



発表者紹介

- 1986年3月 筑波大学大学院生物科学研究科生物物理化学専攻博士課程修了
- 1987年1月 アリゾナ大学神経生物学部門博士研究員
- 1991年7月 筑波大学生物科学系助手、講師、助教授
- 2003年4月 筑波大学生物科学系教授
- 2004年4月 東京大学大学院情報理工学系研究科知能機械情報学専攻教授
- 2006年4月 東京大学 先端科学技術研究センター生命知能システム分野教授

その間

- 1996年9月-1999年8月 科学技術振興事業団さきがけ研究21研究員（兼任）
- 2003年4月-現在 アリゾナ大学神経科学部 Adjunct Professor

研究分野

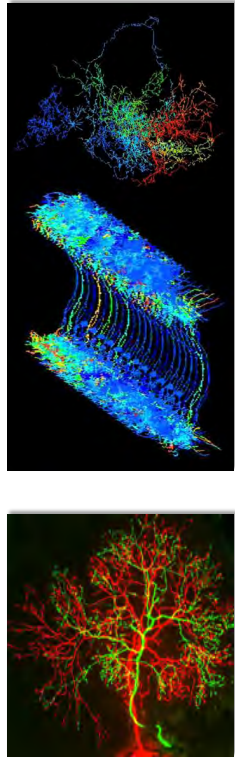
神経行動学、感覚生理学、生物-ロボット融合学



「京」による無脊椎動物嗅覚系シミュレーション — 背景と現状と将来への展開 —

次世代計算科学研究開発プログラム
脳神経系研究開発チーム

神崎亮平 (東京大学 先端科学技術研究センター, 教授)



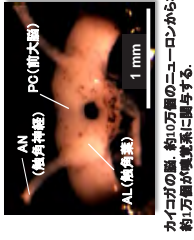
ISLiM成果報告会2012



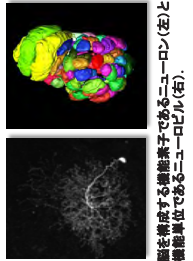
目的

モデル生物であるカイコガの脳を対象とし、嗅覚入力から行動出力に至る神経ネットワークをニューロンレベルの詳細モデルとして構築し、実験と比較しながら脳機能能を予測し、理解するためのシミュレーションをおこなう。方法論の確立。

分析 (実験)

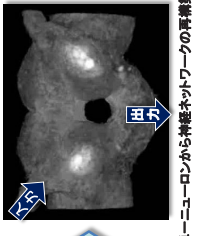


カイコガの脳、約10万個のニューロンからなり、約1万個が嗅覚系に属する。

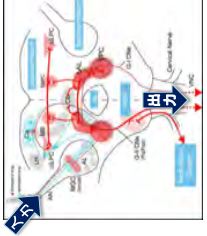


脳を構成する神経素子であるニューロン(左)と機能単位であるニューロビル(右)。

統合 (モデル化)



単一ニューロンから神経ネットワークの再構築



脳内への嗅覚入力から行動出力までの経路

シミュレーション



単一ニューロン活動のシミュレーション



神経ネットワーク活動のシミュレーション



全脳リアルタイムシミュレーションに 要する計算資源

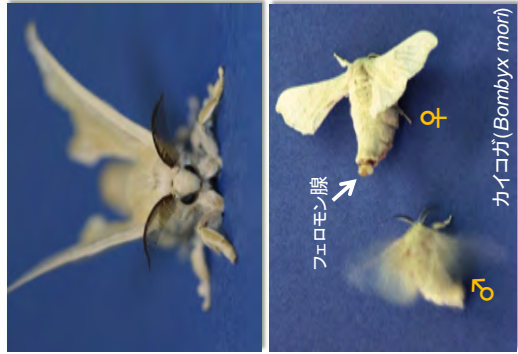
項目	ヒトの全脳	カイコガ全脳	カイコガ嗅覚・運動系
細胞数	10 ¹¹	10 ⁵	10 ⁴
コンパートメント数	10 ⁴	10 ⁴	10 ⁴
必要計算量 (FLOP/1秒)	10 ²¹	10 ¹⁶	10 ¹⁵

計算科学研究ロードマップ白書
<http://open-supercomputer.org/wp-content/uploads/2012/03/science-roadmap.pdf>

ムーアの法則を適用すれば、2030年から2040年ごろにヒトの全脳シミュレーション(10¹¹ニューロン)が可能レベルに到達すると予測できるが、そこに至る過程として昆虫クラスの脳(10⁴ニューロン)を対象にすることは現実的なアプローチとして重要である。



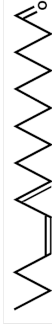
モデル:カイコガの匂い源探索行動



ボンビコール (bombykol)



ボンビカール (bombykal)



カイコガは、文科省「ナショナルバイオリソースプロジェクト(NBRP)」のバイオリソースの1つで、ライフサイエンス研究の基礎・基盤のモデル生物。

嗅覚系全脳シミュレーター(仮想脳)の構築

昆虫嗅覚系全脳シミュレーター (Insect Olfactory System **SIM**ulator)

神経細胞データベースの構築と整備

BoND/IVBPF

細胞形態抽出

KNEWRITE

標準脳・マッピング

Brain Composer

細胞膜情報推定・結合強度推定

大規模シミュレータ

NEURON_K+

脳

脳内の神経叢 銀染色像(水平切片)

主要神経叢の三次元再構築

昆虫脳の階層的構造

神経細胞 (ニューロン)

触角葉 (匂い識別)

キノコ体 (記憶学習)

神経細胞データベースBoNDの構築と整備

神経細胞データベース(BoND)

形態情報

生理情報

>1581 Identified Neurons Recorded

- Morphology: Single-Neuron-3D-Structure, Cell Clusters
- Electrophysiology: Olfactory, Visual, Mechanosensory Responses, Channels
- Imaging: Voltage-Sensitive Dye, Ca-Indicator (dye or transgenic (GCaMP))
- Immunohistochemistry: GABA, 5HT, FMRFamide, tyramine, histamine, proteins

INCF Japan Node

ニューロインフォマティクス国際統合機構 (INCF) 日本ノード: 理研 神経情報基盤センター (臼井支助センター長) に構築するプラットフォームの1つとして情報公開している (IVB-PF 委員長: 荻真 隆)

細胞形態抽出ソフトウェアの開発

KNEWRITE
Manual/auto/semi-auto mode
Ray Burst Sampling Algorithm

三次元の細胞形態抽出

抽出時間の短縮

手動抽出

自動抽出

領域を限定し再自動抽出

半自動による正確かつ高速な形態抽出

手動, 自動, 半自動での抽出が可能 (自動抽出には Ray Burst Sampling Algorithm (Rodriguez et al., 2006) を採用)

Ikeno et al. (2012) Computational Intelligence and Neuroscience (in press)

三次元画像として得られた神経細胞の形態を, シミュレーション用のマルチプラットフォームモデルに自動・半自動・手動で変換するためのソフトウェアを開発した。その結果, 半自動で行うことで, 高速かつばらつきが少い抽出を行うことができた。

単一細胞シミュレーション

細胞形態を円柱近似で表現する
(マルチコンパートメントモデル)

形態の円柱近似
水構図により節間距離を表現

膜特性に関する実測データ

単一ニューロン活動のシミュレーション

マルチコンパートメントモデル等価回路方程式

$$I = C_m \frac{dV}{dt} + g_{Na} \cdot m^3 \cdot h \cdot (V - V_{Na}) + g_K \cdot n^4 \cdot (V - V_K) + g_{Ca} \cdot (V - V_{Ca}) + g_{leak} \cdot (V - V_{leak})$$

$$\frac{dm}{dt} = \alpha_m(V) \cdot (1 - m) - \beta_m(V) \cdot m$$

$$\frac{dh}{dt} = \alpha_h(V) \cdot (1 - h) - \beta_h(V) \cdot h$$

$$\frac{dn}{dt} = \alpha_n(V) \cdot (1 - n) - \beta_n(V) \cdot n$$

V: 細胞内の電位差 (膜電位), *C_m*: 膜容量, *V_{Na}*, *V_K*, *V_{Ca}*, *V_{leak}*: リークイオンの平衡電位
m, *h*: Na⁺チャネルの活性化, 不活性化ゲートの確率, *n*: K⁺チャネル活性化ゲートの確率
 α , β : ゲートの開閉状態の遷移速度

標準脳の構築と神経細胞マッピング

標準脳の構築

平均化処理

標準脳にマッピング

ファイナリゼーション
位置合わせ (剛体変形)

位置合わせ (非剛体変形)

標準脳への神経細胞の登録

個体差を吸収し、統一したシミュレーションを行うため、カイコが標準脳を構築した。また、神経細胞を標準脳上にマッピングしていくことで、形態的特徴を反映させた神経回路の再構築を行った。

標準脳上の神経細胞マッピングと誤差評価

Differences (μm)

Legend: Original (blue), Registered (red)

N = 6 * * P < 0.05 * * * P < 0.01

単体性能の高速化

Single core performance (N = 256)

MFGPS/core

通信高速化

Calculation time [sec]

Number of Cores

ベンチマーク用回路の構築

細胞一つ当たり
20-100個のシナプスを作成

約800,000個のシナプスにより回路を構成

約8,000個の細胞と

神経細胞シミュレータ「NEURON」(www.neuron.yale.edu)をベースに、SIMD化やOpenMP対応など京向けの高速化を施した「NEURON_K+」を構築した。これにより、自動最適化時の約2倍の単体性能と大規模実行に耐える並列性能が得られた。

京を用いた神経回路大規模シミュレーション

標準脳への神経細胞の登録

ベンチマーク用回路の構築

通信高速化

単体性能の高速化

神経細胞シミュレータ「NEURON」(www.neuron.yale.edu)をベースに、SIMD化やOpenMP対応など京向けの高速化を施した「NEURON_K+」を構築した。これにより、自動最適化時の約2倍の単体性能と大規模実行に耐える並列性能が得られた。

細胞膜情報推定ソフトウェアの開発

実数値GAIによるパラメータ推定

遺伝子
 $(g_{ij}, g_{ij}, g_{ij}, \dots)$
 $(0.02, 0.111, 0.0020, \dots)$
 $(0.12, 0.124, 0.0024, \dots)$
 $(0.038, 0.151, 0.0064, \dots)$
 $(0.066, 0.111, 0.0011, \dots)$
 $(0.002, 0.256, 0.0023, \dots)$
 $(0.005, 0.162, 0.0014, \dots)$

シミュレーションと生理実験結果との比較

$$F = \frac{1}{\int_0^{T(t)} dt} \int_0^{T(t)} |P(t) - T(t)| dt$$

選択・変更・交叉

パラメータ探索の並列性能の評価

コア数

シミュレーションによる仮想的な電流クランプ応答波形を用いた、推定ソフトウェアの評価

$$E = \sum_{i=1}^n \frac{|Result_i - Target_i|}{Range_i}$$

Single Compartment

Multi Compartment

シナプス強度推定

前運動中枢の5領域モデル

領域ごとの平均活動度を元にすべてのパターン遷移をニューラルネットワークとして表現

結合強度推定 (生理応答を再現する結合を推定)

推定されたシナプス結合強度

生理実験結果

領域(左脳) 領域(右脳)

Bilateral Neuron
Local Interneuron

形態的な情報による結合可能性の推定

日本神経回路学会2012年度最優秀研究賞

結合推定の解

M-モデルの頑健性から得られた主要回路

両者の共通部分より求めた主要回路

BLN = 抑制性の仮定で求めた入力から出力へのフィードフォワードネットワーク

結合度の強さから得られた主要回路

嗅覚系シミュレーション

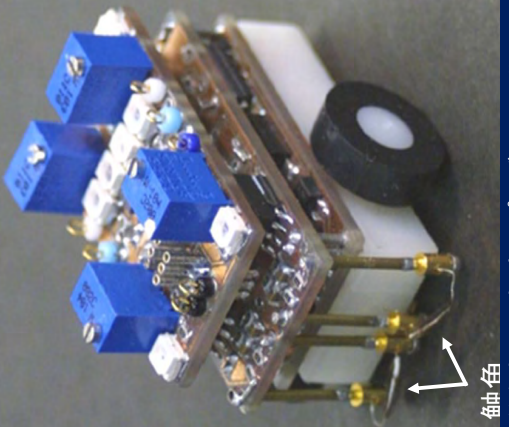
前運動中枢の神経活動の京によるシミュレーション

行動指令ニューロン

前運動中枢を構成するニューロン

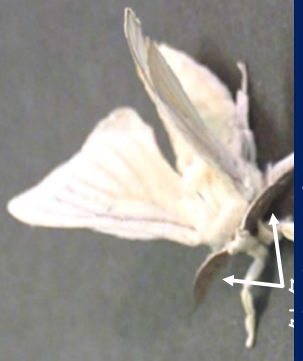
12ニューロンx2径
192ニューロンに拡張
Simulation time: 1000step(50msec)
86nodesで実行

昆虫脳モデル-操縦型ロボット フェロモンの匂い源に定位する移動ロボット



触覚

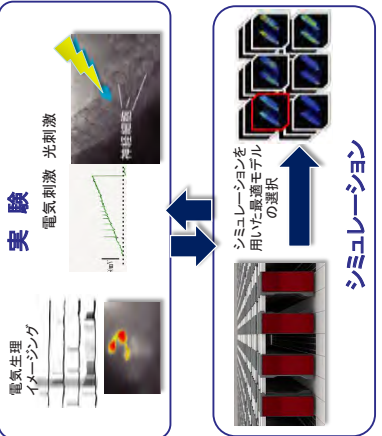
基本的な匂い源探索のプログラム行動により、複雑に変化する匂い環境下で
定位行動が実現された。



将来の目標:IOSSIMの活用

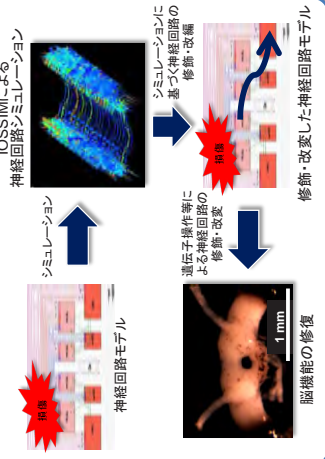
リアルタイムシミュレーションと生体計測の 融合による脳回路の探索

- 脳内状態の観測まで含めたリアルタイムシミュレーションを行い、オンタイムで実験と比較することによって脳内の状態を推定する。
- 脳の制御や、パラメータを推定するための刺激を行う。



脳シミュレーションに基づく脳機能の修復 (ニューロリハビリテーション)

- ニューロンレベルでの精緻な神経回路のシミュレーションを用いて神経回路の一部に損傷や欠損が生じた場合でも、損傷・欠損部位を予測する。
- 特定のニューロン(群)の操作による機能修復のシミュレーションを可能とし、
- その上で特定のニューロン(群)の活動を遺伝子操作技術(たとえばオプトジェネティクス)により実施し、機能回復を支援する新しい技術の創出を目指す。



生命基盤ソフトウェア開発・高度化チーム成果報告

ISLiM ソフトウェアの高度化

泰地 真弘人

次世代計算科学研究開発プログラム

生命体基盤ソフトウェア開発・高度化チーム チームリーダー



発表者紹介

- 1992年3月 東京大学理学系研究科物理学専攻博士課程修了
- 1992年4月 東京大学教養学部助手
- 1997年1月 統計数理研究所助教授
- 2002年4月 理化学研究所ゲノム科学総合研究センターチームリーダー
- 2008年4月 理化学研究所基幹研究所グループディレクター
- 2011年4月 理化学研究所生命システム研究センター生命モデリングコアコア長

研究分野

高性能計算・計算機アーキテクチャ

ISLIM ソフトウェアの高度化

泰地 真弘人

次世代計算科学研究開発プログラム

生命体基盤ソフトウェア開発・高度化チーム チームリーダー

1. 目的

京コンピュータは、80,000 プロセッサ、640,000 コア以上を有する大規模な超並列計算機です。このような規模の計算機を使いこなすためには、並列処理、特に京コンピュータに通曉した研究者による支援体制が必要です。そのために、分子動力学計算コアソフトウェアの開発などの独自開発を通じた人員育成を行うと同時に、他チームで開発されたソフトウェアの性能評価・開発支援を行っています。さらに、開発容易化のためのミドルウェア開発・可視化ソフトウェア開発を行い、産業利用に向けた大規模仮想化合物ライブラリの開発を進めています。

2. 現時点での成果

2.1 分子動力学コアソフトウェア

「京」用アプリケーションの最適化ノウハウの蓄積・提供および最適化の参考コードの提供のため、分子動力学コアプログラムの開発・最適化を行ないました。主要な計算部分であるカットオフ法カーネルでは 60%以上の演算効率を達成しました。カーネル部分の SIMD 化率は全命令の 57%、浮動小数点演算命令に関しては 99%以上となり、「京」への最適化が十分行なわれています。最適化阻害要因(特に C++コンパイラの場合)に関する知見は、他チームの支援業務にも活かされています。ノードあたり 6000 原子程度の計算規模に対し、全ノード並列で演算速度 4Pflops・効率 40%を達成しており、ノードあたり 1000 原子以上の計算規模であれば「京」全ノードまでの並列性があることを確認しました。

長距離相互作用計算においては、現在主流の PME 法は FFT が大域通信を必要とするため「京」の全ノード規模での並列性は得られないと考えられるので、高速多重極展開法(FMM)を実装中です。単純な実装での評価ではありますが、「京」全ノード規模での並列性を確認できています。

開発したコードは、「京」での最適化・分子動力

学計算の参照コードとして自由に再利用できるように公開する予定です。

2.2 ソフトウェア高度化

ソフトウェアの性能評価・チューニング支援を行っています。第一走者アプリケーションを中心に、性能評価報告を行いました。チューニング支援については、特に神戸拠点における京に向けての開発支援を中心に、開発実施本部との窓口・京に向けたチューニング外注の窓口機能を担っています。

2.3 可視化ソフトウェア

大規模データを扱える並列可視化システム ISL-LSV を開発しました。

2.4 並列処理ミドルウェア

ソフトウェア開発容易化のためのミドルウェア Sphere を開発しました。連続体シミュレーションを中心に、ソルバ群の組み合わせにより容易に並列化されたソフトウェア開発を可能にします。本ソフトウェアは臓器全身スケール研究チームで開発されている ZZ-EFSI/ZZ-HIFU、細胞スケール研究チームで開発されている RICS で用いられています。



2.5 大規模仮想化合物ライブラリ

10 億化合物規模の合成可能性をもつ仮想化合物ライブラリを、東大工学部船津研究室を中心に開発しています。

2. プロジェクト終了時の達成目標



プロジェクト終了時に、より多くのアプリケーションが京コンピュータ上で十分な性能を発揮していることが我々の目標です。また、創薬応用に向けた 10 億化合物規模の大規模化合物ライブラリを完成させます。



以上

ISLiMソフトウェアの高度化



理化学研究所
次世代計算科学研究開発プログラム
生命体基盤ソフトウェア開発・高度化チーム
チームリーダー
泰地真弘人






背景・目的



- 背景
 - 京 = 64万コア: 大規模並列化の要請
 - ソフトウェア開発の困難・高性能計算技術の必要性
 - 京利用上のノウハウの蓄積
- 目的
 - 各チームのソフトウェアの高度化
 - 需要の多いコアソフトウェアの開発
 - MDコアソフトウェア





アプローチ

- 京コンピュータへの移植
 - コンパイルオプション
 - クロスコンパイル
- 京コンピュータ上での並列目標設定
 - 到達ノード数
- 京コンピュータ上での性能測定
 - ストロンク・スケーリング
 - ウィーク・スケーリング
 - ピークパフォーマンス
 - ホットスポット同定
- 測定結果に基づく性能改善
 - 並列化アルゴリズム、データ分割方法の見直し
 - MPI / OpenMP ハイブリッド並列化
 - SIMD化、パイプライン化、キャッシュヒット率改善、メモリ利用効率改善、並列I/O
 - コード変更の提案、実施




高度化支援対象コード一覧

- 分子スケール研究開発チーム
 - Platyplus-MM/CG, Platyplus-REIN, Platyplus-QM/MM-FE, Platyplus-QM, Platyplus-QM/MM, Platyplus-CGM/MM, MARBLE, CafeMol
- 細胞スケール研究開発チーム
 - RICS
- 臓器全身スケール研究開発チーム
 - ZZ-EFSI, ZZ-RBC, ZZ-DOSE, ZZ-HIFU, ZZ-LUNG, UT-HEART
- データ解析融合研究開発チーム
 - ParaHaplo, NGS analyzer, ExRAT, SIGN-BN, SIGN-L1, SIGN-SSM, SBIP, LISDAS, MEGADOCK
- 脳神経系研究開発チーム
 - NEST, CMDN, VSM, NeuroMorphKit, IOSSIM
- 生命体基盤ソフトウェア開発・高度化チーム
 - cppmd, LSV, SPHERE, VLSVL






MARBLEの高度化支援

- 主要コストのdirect_atomモジュールの改善
 - ピーク比の向上
 - 10% → 27%
 - 最内ループ回転数を増加(10回転 → 500~600回転)させ、ループ最適化の効果を向上
 - 6.0秒 → 4.8秒
- キャッシュの有効利用化
 - コンパイルオプションや最適化指示子の利用
 - 浮動小数点キャッシュアクセス待ち時間が改善
- ループ最適化
 - 最内ループのif文による最適化阻害の指摘

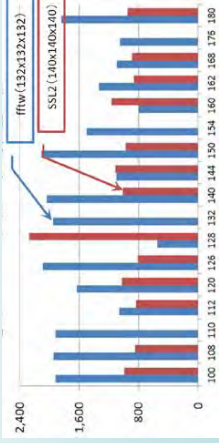


ISLiMソフトウェア研究開発報告会



MEGADOCKの高度化支援

- FFT計算の高性能化
 - FFTW版
 - プラン作成・消去を初回のみに限定
 - CSSL2版
 - スレッド並列実行化
 - 1次元複素FFTルーチンを用いた3次元複素FFTの実装
 - FFT長の2のべき乗化
 - FFT処理作業配列のパディング
 - 198秒 → 159秒
- コンパイルオプションの見直し
 - 浮動小数点演算関連オプション

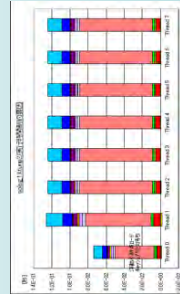
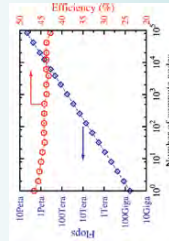


ISLiMソフトウェア研究開発報告会



ZZ-EFSIOの高度化支援

- 最内ループ回転数が少ない場合のループ最適化
 - 3重ループ → 1重ループ化、及びループ分割
 - 実行時間: 4.0秒 → 2.8秒; ピーク比: 32.1% → 46.8%
 - 浮動小数点演算待ち: 6.5% 減少
- MPI_Allreduce性能の改善
 - 6要素 → 1要素ずつに分割によるTofu高機能バリア通信化
 - 実行時間: 30% 短縮
- 隣接通信の隠蔽
 - 内部領域と袖領域への分割
 - 通信・演算を各領域毎にオーバーラップ
 - 通信をmasterスレッド、計算をその他のスレッドに割当
 - 実行時間: 0.93秒 → 0.78秒
- 単体性能の改善
 - ループ分割と変数コピー数削減
 - 実行時間: 49.6秒 → 38.9秒

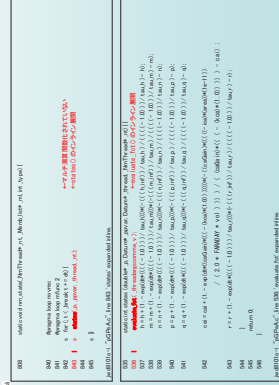


ISLiMソフトウェア研究開発報告会




IOSSIMの高度化支援


- 京コンピュータへの移植
 - クロスコンパイル環境への対処
 - 多段階ビルド
- システム関数の最適化
 - exp関数, log関数, pow関数をマルチ演算関数に置換
 - ピーク比 2倍に



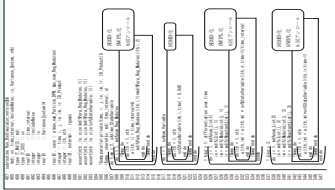
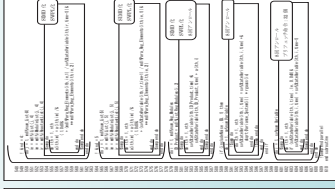
ISLiMソフトウェア研究開発報告会




LiSDASの高度化支援




- SIMD化、ソフトウェアパイプライン化の促進
 - デバッグコードの除去 (assert文, if文)
- 浮動小数点除算、SQRT関数の最適化
 - 逆数近似演算命令とFloatin-Point Multiply-Add/Subtract命令の利用
- `-fp_relaxed` オプション
 - 除算例外とアンダーフローの検出
 - 割り込みイベントの検出


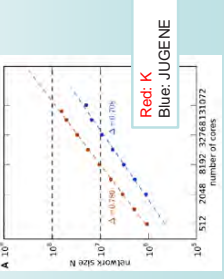
ISLiMソフトウェア研究開発報告会 9




その他の高度化支援




- ParaHaplo
 - 数万人 x 数百万SNPのゲノムデータの入力
 - 並列I/Oに向けたデータ分割ツールの開発
- NEST
 - ニューロン・シナプスモデルのメモリ使用量の低減
 - ニューロン数/ノードの測定と見積り

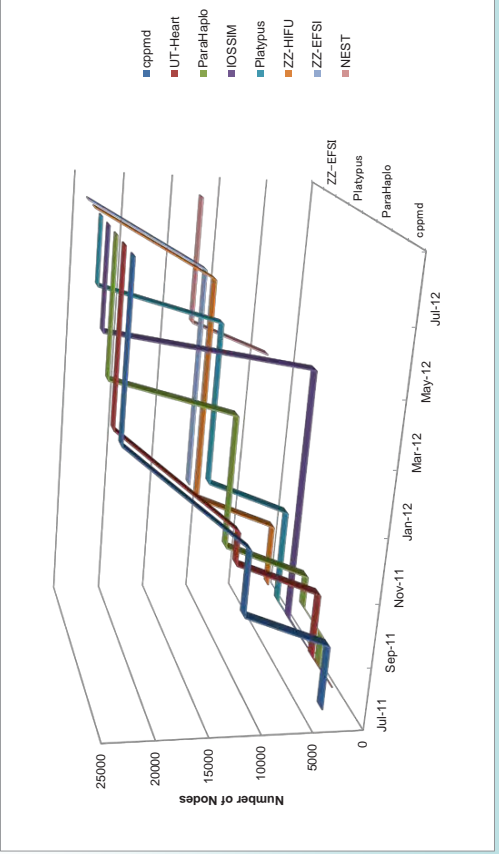



ISLiMソフトウェア研究開発報告会 10




到達ノード数推移






ISLiMソフトウェア研究開発報告会 11



ISLiMコードの高度化支援



	移植	測定	単体チューニング	並列チューニング	その他
分子	○ 富士通コンパイラ対応 (Catemol)	○	○ SIMD, Pipeline (Catemol) Loop, Cache (MARBLE)		
細胞	○ Sphere(RICS)	○			
臓器全身	○ Vsphere(ZZ-EFSI)	○	○ Loop (ZZ-EFSI)	○ 通信(ZZ-EFSI)	
脳	○ ビルド方法の構築 (IOSSIM) 富士通コンパイラ対応 (NEST)	○	○ thread化 (NEST) SIMD化、マルチ関数 (IOSSIM)		○ メモリ消費削減 (NEST)
データ解析 融合	○	○	○ SIMD, Pipeline (LiSDAS)		○ 分散I/O (Parahaplo), FFT (MegaDock)
高度化	○	○	○	○	

ISLiMソフトウェア研究開発報告会 12

MD コアプログラムの最適化と性能評価

大野 洋介

次世代計算科学研究開発プログラム
次世代生命体統合シミュレーション研究推進グループ
生命体基盤ソフトウェア開発・高度化チーム




発表者紹介

- 1996 年 3 月 東京大学大学院理学系研究科天文学専攻博士課程修了
- 1996 年 4 月 理化学研究所 基礎科学特別研究員
- 1999 年 4 月 理化学研究所 計算科学研究室 協力研究員
- 2001 年 4 月 理化学研究所 ものづくり情報技術統合化研究プログラム 研究員
- 2004 年 4 月 理化学研究所 ゲノム科学総合研究センター(GSC) 研究員
- 2008 年 4 月 理化学研究所 次世代計算科学研究開発プログラム 上級研究員 (現兼務)
- 2012 年 4 月 理化学研究所 生命システム研究センター(QBiC) 上級研究員

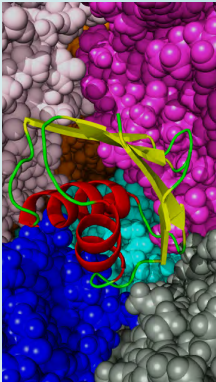

研究分野

計算科学、専用計算機




MDコアプログラムの最適化と性能評価

理化学研究所
次世代計算科学研究開発プログラム
次世代生命体統合シミュレーション研究推進グループ
生命体基盤ソフトウェア開発・高度化チーム 兼務
大野 洋介


ISLiMソフトウェア研究開発報告会 1



目的

- 「京」での最適化ノウハウを他チームへ提供する。
 - コンパイラ・ハードウェアの能力を把握する。
 - ・ 特に、最適化の挙動
 - 具体的な最適化の知見を蓄積する
 - 得られた最適化の知見を他のチームへ提供する。
 - 再利用可能なコードとして提供する。
 - ・ 最適化事例
 - ・ 最適化済み分子動力学コード


ISLiMソフトウェア研究開発報告会 2



概要・アプローチ

- 研究開発コードの概要
 - 基本的な分子動力学計算
 - ・ アンサンブル：NVE, NVT, NPT
 - ・ 時間積分：velocity Verlet
 - ・ クーロン相互作用：cutoff, PME, FMM
 - ・ 並列化：直方体空間分割, cell index method
- アプローチ
 - 主にcutoff カーネルの最適化
 - ・ どの手法でも主要な計算部分
 - 高速多重展開法(FMM)
 - ・ 大規模並列での本命

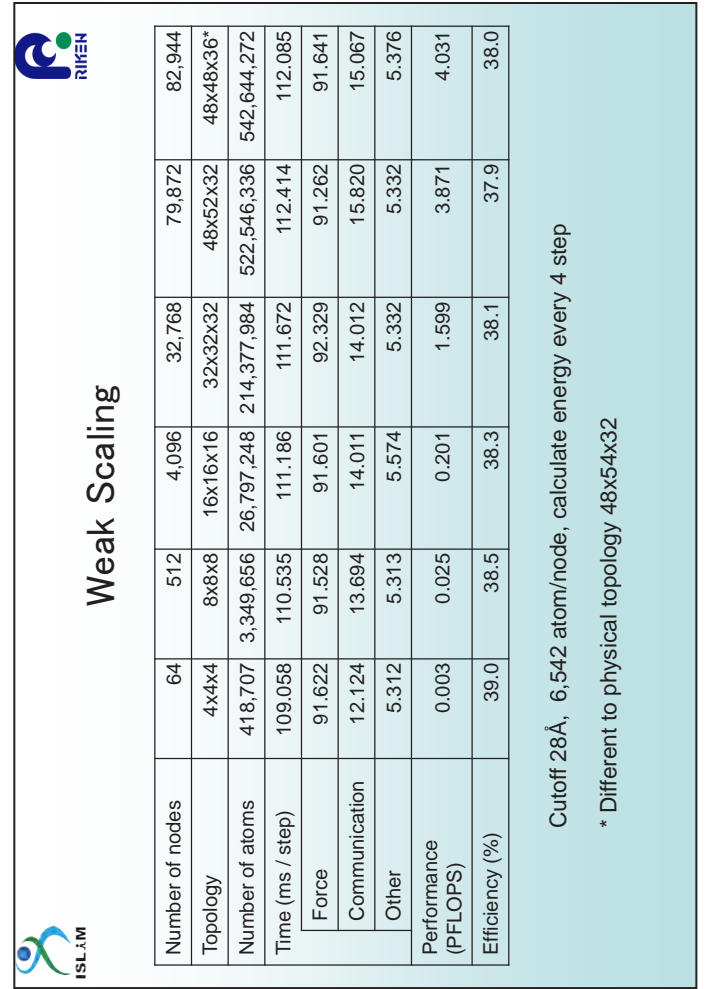
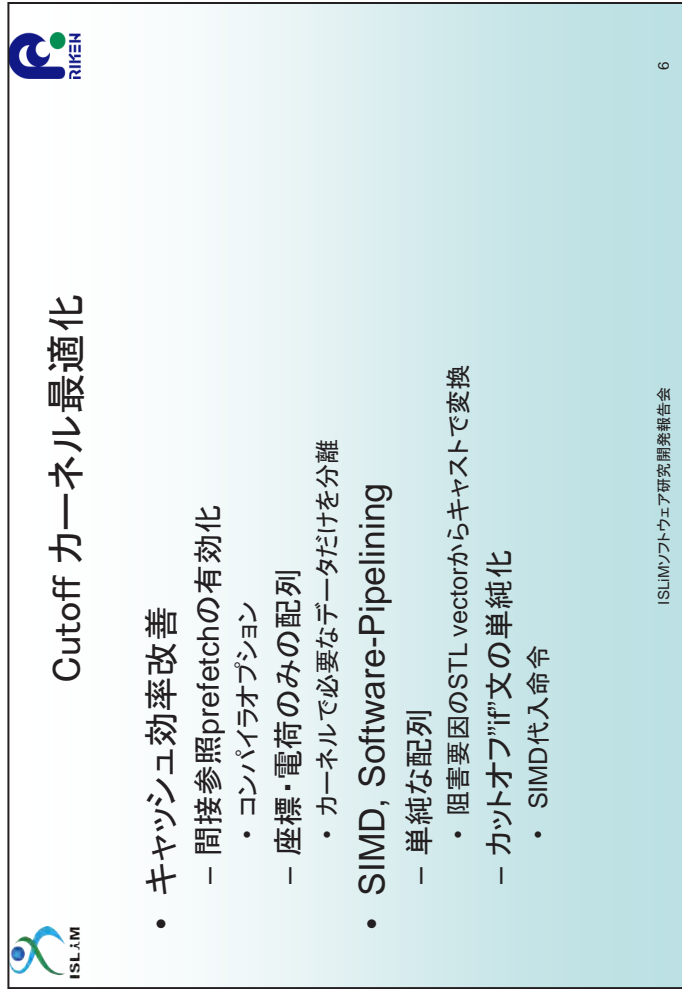
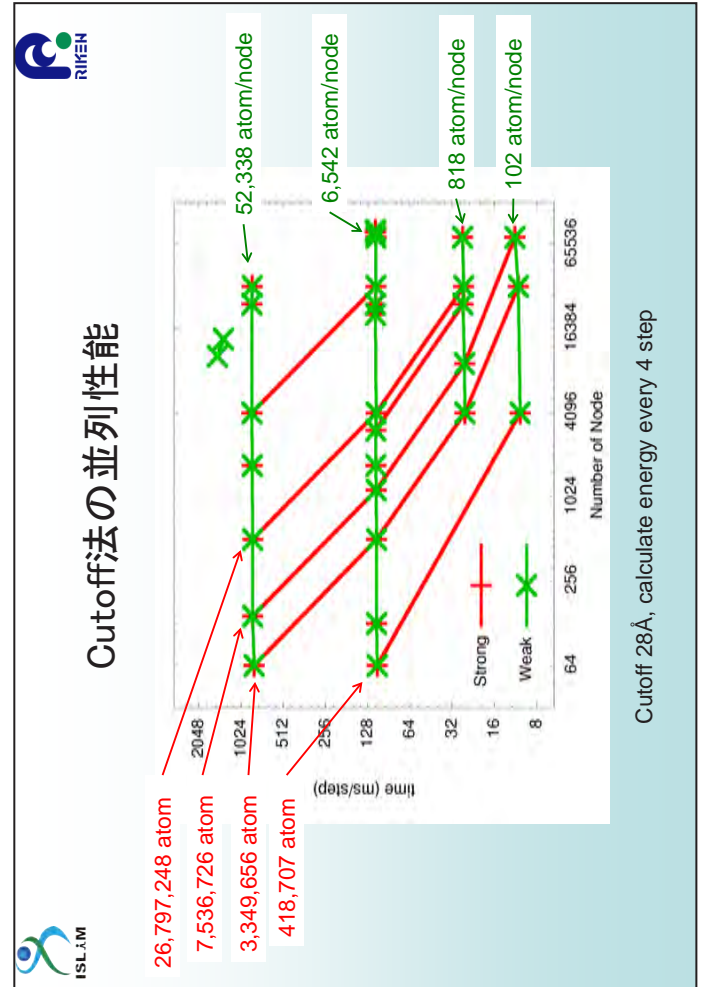
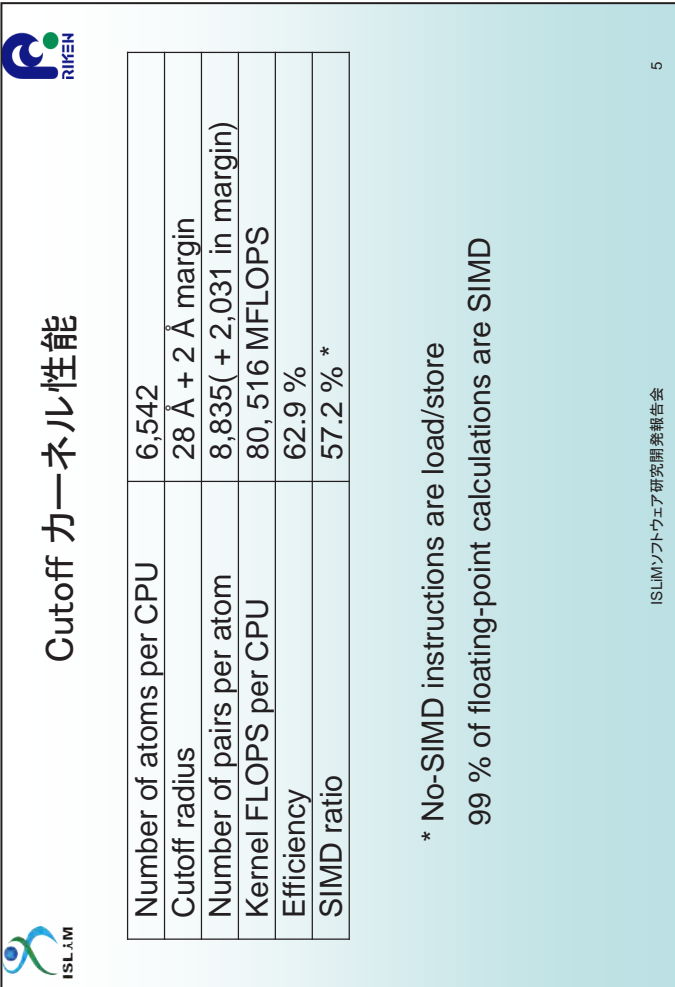
ISLiMソフトウェア研究開発報告会 3

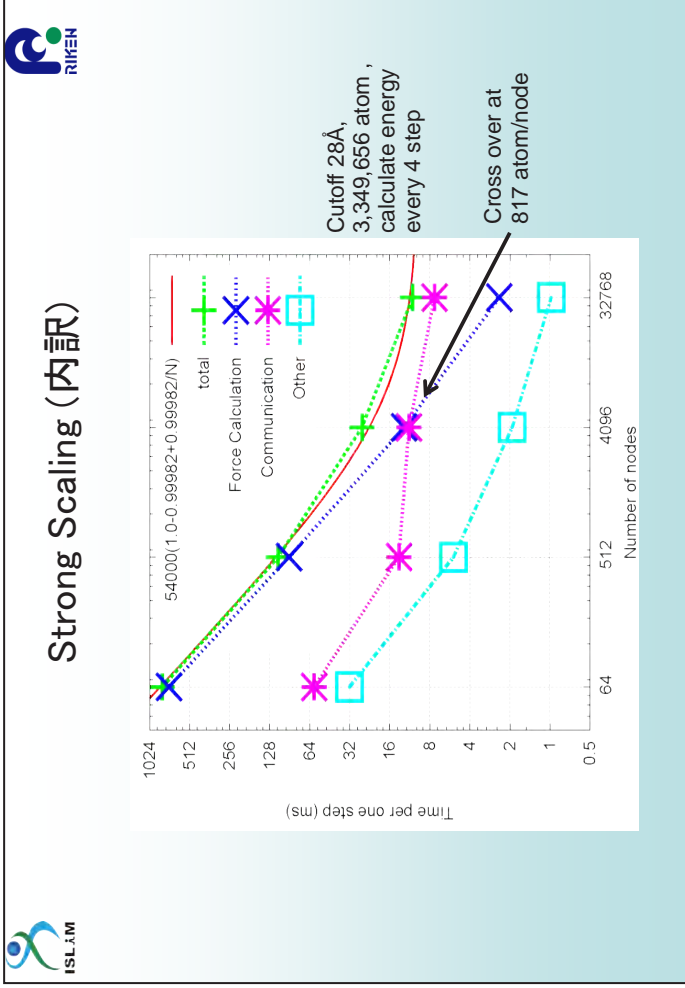


現在までの研究開発成果

- Cutoffカーネルの演算効率60%以上
 - SIMD, Software-pipelining
 - ・ データ形式の変更
 - 単純配列
 - ・ "if"文のマスク演算化
- Cutoff法 4Pflops 40%
 - TOFUネットワークに適した通信
 - ・ ほぼ近接通信、3次元空間分割
- 高並列FMM
 - PMEで使われるFFTに比べ大域通信が少ない

ISLiMソフトウェア研究開発報告会 4





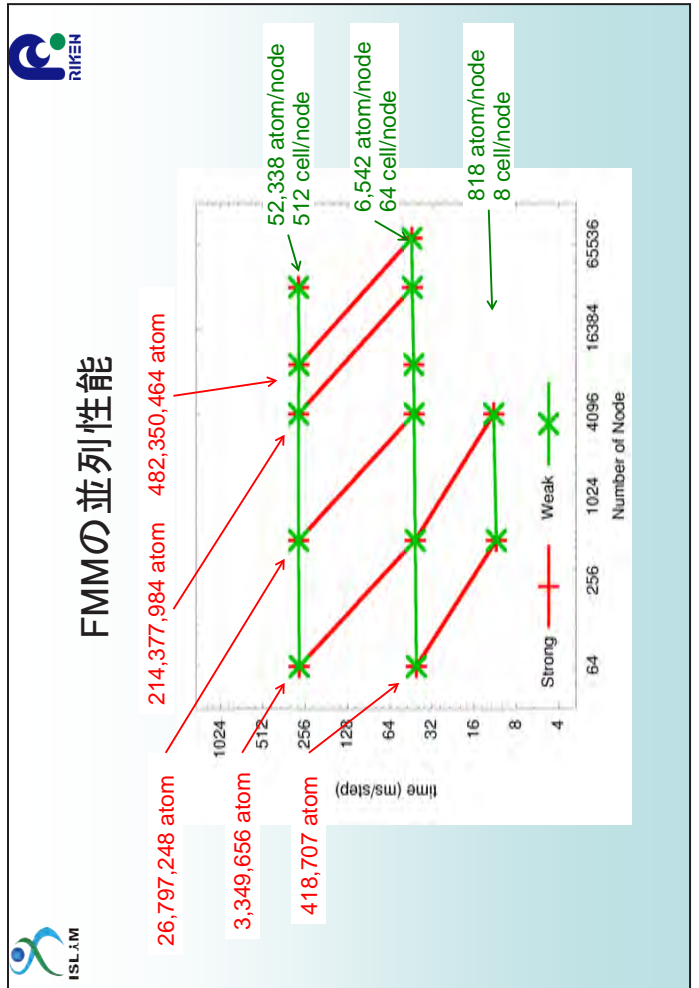
Strong Scaling

Number of nodes	64	512	4,096	32,768
Number of atoms per node	52,338	6,542	817	102
Number of transfer cells	2,232	936	504	342
Number of communication nodes	26	26	124	342
Time (ms / step)	823.07	110.54	25.90	10.77
Force	731.66	91.53	12.50	2.41
Communication	59.28	13.69	11.48	7.38
Other	32.13	5.31	1.92	0.98
Scalability (to 64 nodes)	1.00	0.93	0.50	0.15

Cutoff 28Å, 3,349,656 atom, calculate energy every 4 step

Conclusion

- 「京」での最適化の調査・ノウハウの蓄積・提供のため MDコアプログラムの開発・最適化を実施した。
- Cutoff カーネルの演算効率60%以上を達成した。
- 全ノードCutoff法計算で、演算速度4Pflops・効率40%を達成した。
- FMMを実装し、全ノード規模での並列性を確認した。
- 最適化事例、最適化済みMDコードとして再利用できるように公開準備を進めている。



創薬プラットフォーム —大規模バーチャルライブラリの開発—

船津公人

VLSVL 合成可能な化学構造および反応スキームからなる
大規模仮想ライブラリ 発表者



発表者紹介

1983年3月 九州大学大学院理学研究科化学専攻博士課程修了
1984年3月 豊橋技術科学大学工学部物質工学系 助手
1988年3月 豊橋技術科学大学工学部知識情報工学系 助手
1992年4月 豊橋技術科学大学工学部知識情報工学系 助教授
2004年4月 東京大学大学院工学系研究科 教授

研究分野

ケモインフォマティクス (化学情報学)
関連する研究分野としてプロセスシステム工学、有機合成化学、分析化学、触媒化学、
創薬化学など

発表創薬プラットフォーム — 大規模バーチャライブラリの開発 —

東京大学
大学院工学系研究科化学システム工学専攻
教授
船津公人

ISLiM

ISLiMソフトウェア研究開発報告会

1

新薬開発の特質

(03年~07年)

薬物標的の同定
リード化合物発見
リード化合物最適化
非臨床試験
臨床試験(治験)
申請承認

563,589化合物
83化合物
26化合物

成功確率 1 / 6,790 (0.0147%)
1 / 21,670 (0.005%)

できずとぶつく 製薬産業2009

● R&Dリスクが高い
● 長期、投資額、成功確率...
● 漏れの懸念

たまたま手に入った化合物の中から選択している..

ISLiMソフトウェア研究開発報告会

2

化合物ライブラリの現状

■ 実物

大学	化合物数
Broad Institute	> 250,000
UCSF	> 150,000
Stanford	> 130,000
Kalamazoo Valley Community College	> 100,000
UCLA	> 75,000

- ・米国
- ・韓国: ナショナルバンク 17万化合物
- ・東京大学: 2010年3月 196,970化合物

■ DB

- REAL Database 1500万化合物
- Drug-like selection 400万化合物
- ChemNavigator* 5530万化合物

■ VL

- everything 2800万化合物
- purchasable 1700万化合物
- drug-like 1000万化合物

ISLiM

ISLiMソフトウェア研究開発報告会

3

化合物ライブラリの問題点

■ 規模と多様性

創薬の研究対象となる化合物
C, N, O, S からなる30原子まで^[1]: 10⁶⁰個
スクリーニング用実在化合物^[2]: 5.5 × 10⁰⁷個

1 / 10⁵³

全体のごく僅かな部分からヒット確率を上げるには、
全体空間からできるだけ万遍なくサンプルを揃えるべき...

■ 入手可能性

- ・メガファーマが保有する化合物(1990年代): 10⁰⁶個
- ・バーチャルライブラリ
バーチャルスクリーニングで高スコアを得た化合物であっても
合成検討にコストがかかる

ISLiMソフトウェア研究開発報告会

4

ISLiM

目的

- リード構造探索のための
大規模な化合物仮想ライブラリ (VL) の構築
- VL に含まれる仮想化合物の多様性の確保、および
仮想化合物の合成ルート情報の提供

ISLiMソフトウェア研究開発報告会

5

ISLiM

システム概要

Transform automatic extraction module

Reaction database → Transform Generator → Transform Library

VL Generator

Chemical Library → Target → Precursors → Product → Virtual Library

Reactant

ISLiMソフトウェア研究開発報告会

6

ISLiM

Transformの抽出例

Reaction example

Extracted Transform

- : Decrease bond order of double bond
- △: Make new single bond or increase bond order

ISLiMソフトウェア研究開発報告会

7

ISLiM

バーチャルライブラリの構成と目標

10億構造 Virtual Library

1億構造

1億3000万スキーム

Chemical Library 42万分子

5億構造

5億構造

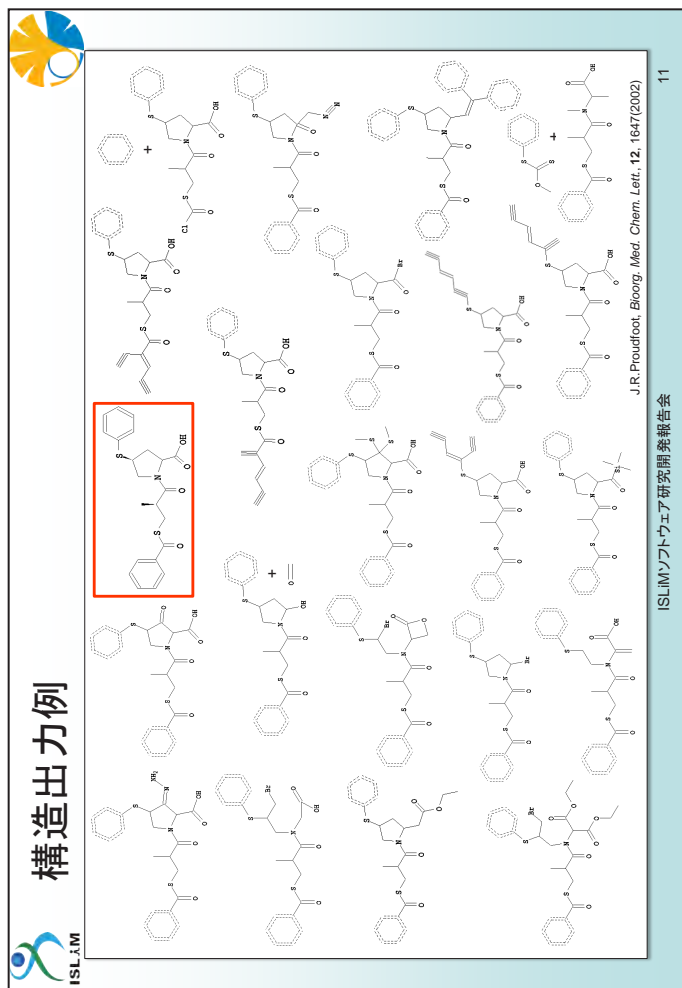
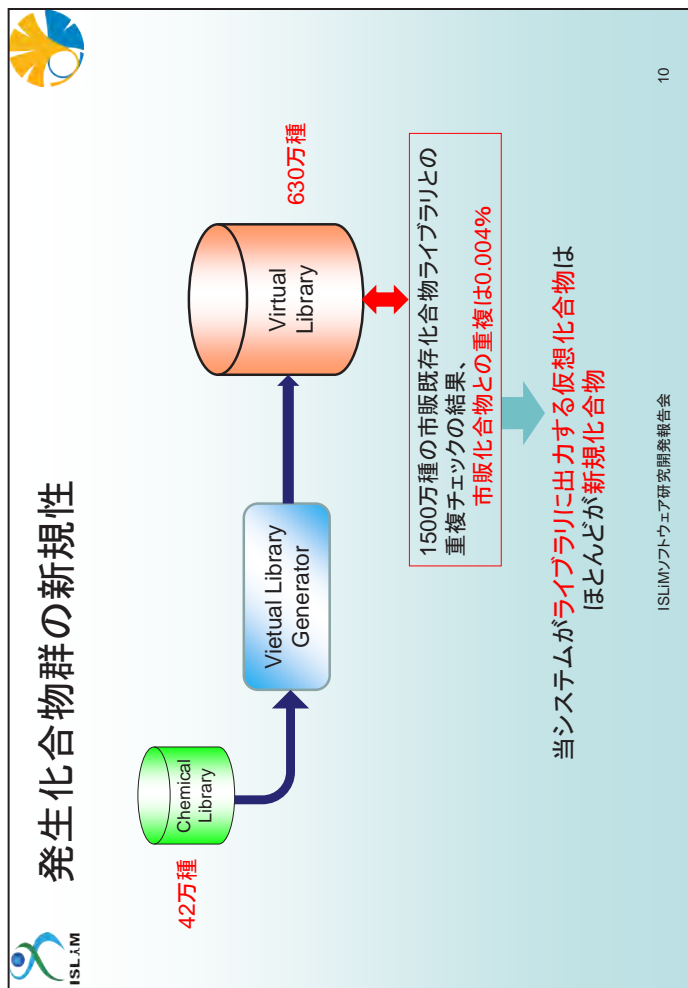
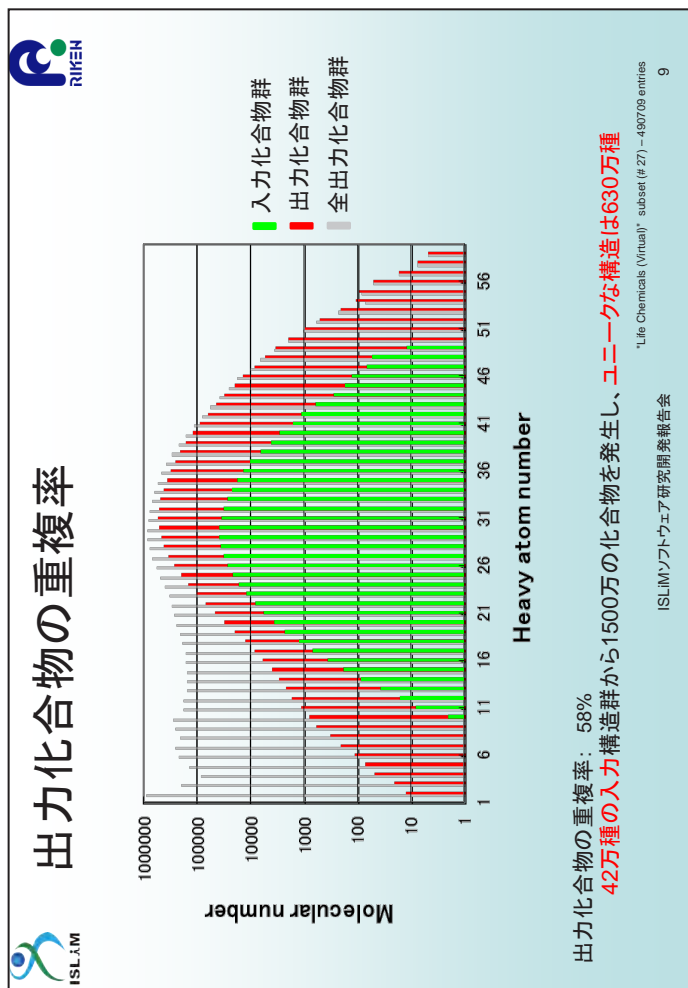
○ → ○ : Component of Reaction Scheme

● : Not Lead compound

○ : Likely lead compound

ISLiMソフトウェア研究開発報告会

8



Lipinski's rule of 5, rule of 3

	rule of 5	rule of 3
分子量	< 500	< 300
水素結合供与体数	≤ 5	≤ 3
水素結合受容体数	≤ 10	≤ 3
分配係数計算値 (cLogP)	≤ 5	≤ 3
回転可能結合数	≤ 10	≤ 3
極性表面積 (A ²)	≤ 140	≤ 60

- Rule of 5 : 経口吸収性の高い構造特性
- Rule of 3 : リード化合物 (ドラッグメント化合物) として相応しい構造特性

[1] Lipinski, C., et al., *Adv. Drug. Deliv. Rev.* 23, 3(1997)
[2] Veber et al., *J. Med. Chem.* 45, 2615-2623 (2002)

ISLiMソフトウェア研究開発報告会 12

Lipinski's rule of 5 どのくらい満たす?

rule of 5	input	output
分子量	<500	85.6% 75.3%
水素結合供与体数	≤5	100.0% 99.9%
水素結合受容体数	≤10	99.9% 99.7%
分配係数計算値 (cLogP)	≤5	79.9% 79.9%
回転可能結合数	≤10	98.2% 93.0%
極性表面積 (A ²)	≤140	99.0% 96.4%

"Life Chemicals (Virtual)" subset (# 27)

ISLiMソフトウェア研究開発報告会 13

分布の異なる種構造群の活用検討

ISLiM RIKEN

ISLiMソフトウェア研究開発報告会 "Life Chemicals (Virtual)" subset (# 27) - 490709 entries 14

入力化合物群の特性分布の影響

ISLiM RIKEN

ISLiMソフトウェア研究開発報告会 15

極性表面積と回転可能結合数の関係

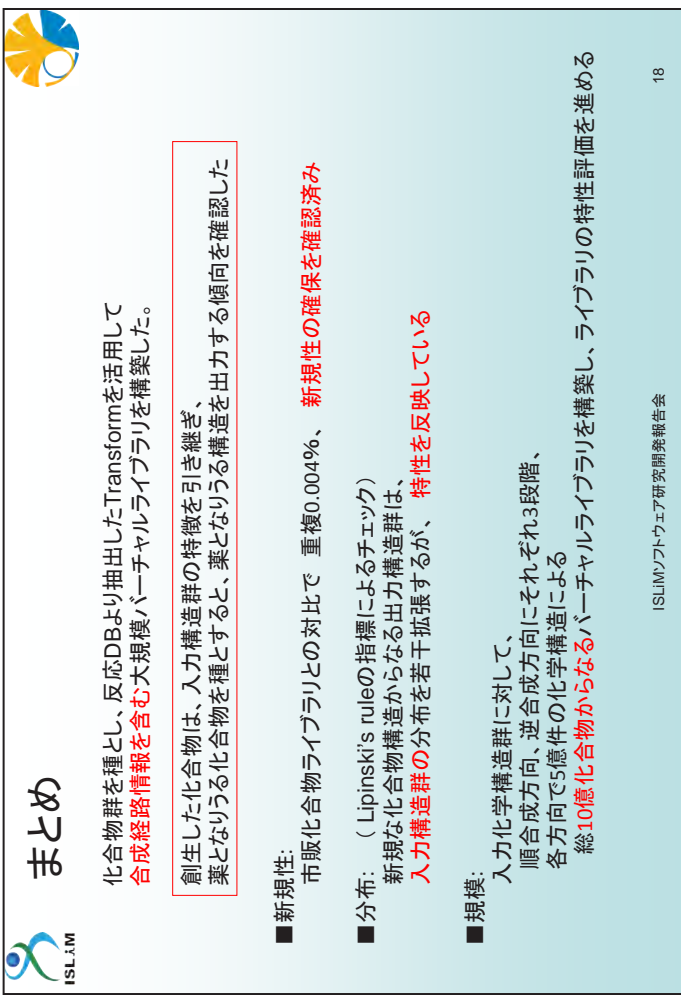
ISLiM RIKEN

ISLiMソフトウェア研究開発報告会 16



ISLiMソフトウェア研究開発報告会

17



ISLiMソフトウェア研究開発報告会

18

ポスター一覧



M-1: Platypus-MM/CG

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



ISLiM 開発責任者: 木寺詔紀 (理化学研究所分子スケール研究開発チーム)

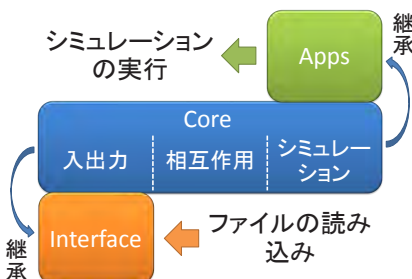
分子動力学法の目的

- 生体高分子が担う重要な生命現象をコンピュータ上に再現する
 - 自発的な立体構造・高次構造形成
 - 分子認識・分子間相互作用
 - 立体構造変化

問題点と解決法

- 多くの生命現象の時間スケールは、ミリ秒から秒のオーダー
- 通常の分子動力学法では、計算時間がかかりすぎて再現できない
- マルチコピー・マルチスケール法と統計力学を用いてこれらの問題を解決

ライブラリの構成

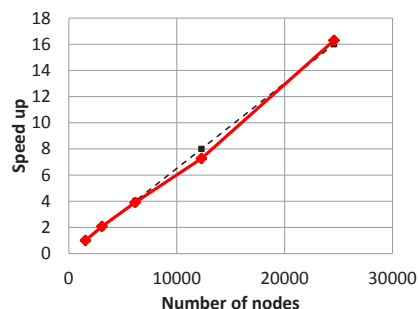


ライブラリの特徴

- 計算対象をオブジェクトとして扱う
- Amber、Charmm力場に対応
- OpenMP/MPIによる並列化

計算性能

- 24,576ノード(196,608コア)まで計算性能が直線的に向上
- 高並列時でも高い実行効率を実現(24,576ノード並列時で約12.3%)

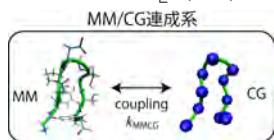


マルチスケール拡張サンプリング(MSES)法

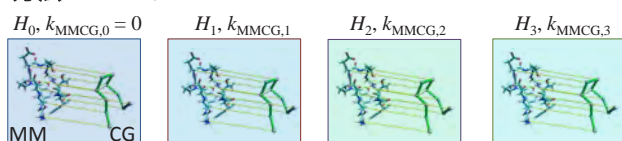
目的: 高精度な全原子モデル(MM)と高速に運動する粗視化モデル(CG)との連成シミュレーションにより、生体高分子の全原子構造サンプリングを効率的に実現

方法1: マルチスケール

$$H = H_{MM} + H_{CG} + k_{MMCG} [\theta(\mathbf{r}_{MM}) - \mathbf{r}_{CG}]^2$$



方法2: マルチコピー

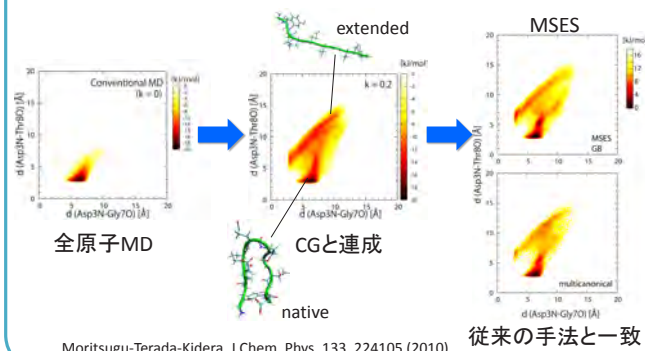


ハミルトニアンレプリカ交換法

Fukunishi-Watanabe-Takada, J. Chem. Phys. 116, 9058 (2002)

CGバイアスのないMM構造アンサンブル

応用例: シニョリンフォールディングの構造サンプリング



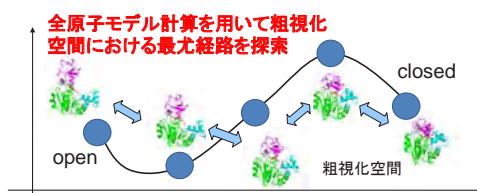
Moritsugu-Terada-Kidera, J. Chem. Phys. 133, 224105 (2010)

有限温度ストリング法

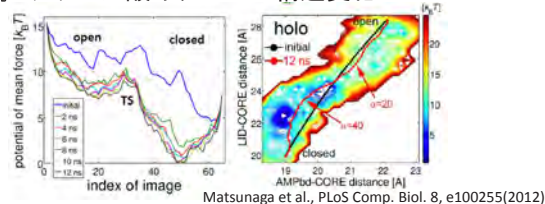
目的: タンパク質が構造変化する際の尤も確率の高い経路を探索する。経路上の遷移状態を特定することにより、構造変化のボトルネックとなる残基を同定する。

方法: マルチコピー・マルチスケール

構造変化を記述する粗視化空間内において、構造変化経路を定義し、経路を離散化した点周りの全原子マルチコピーMDにより最尤経路を探索する。



応用例: アデニル酸キナーゼの構造変化



Matsunaga et al., PLoS Comput. Biol. 8, e100255(2012)

Publications

- K. Moritsugu, T. Terada, A. Kidera. "Scalable free energy calculation of proteins via multiscale essential sampling" *J. Chem. Phys.* **133**, 224105 (2010).
- K. Moritsugu, T. Terada, A. Kidera. "Disorder-to-Order Transition of an Intrinsically Disordered Region of Sortase Revealed by Multiscale Enhanced Sampling" *J. Am. Chem. Soc.* **134**, 7094-7101 (2012).
- Y. Matsunaga, H. Fujisaki, T. Terada, T. Furuta, K. Moritsugu, A. Kidera. "Minimum Free Energy Path of Ligand-Induced Transition in Adenylate Kinase" *PLoS Comput. Biol.* **8**, e1002555 (2012).

謝辞: 本ポスターの結果の一部は、理化学研究所が実施している京速コンピュータ「京」の試験利用によるものです。



M-2: Replica-exchange Interface program: REIN



Naoyuki Miyashita^{1,2,3}, Suyong Re⁴, and Yuji Sugita^{1,3,4}

¹ RIKEN QBiC, ² RIKEN CSR P ISLiM, ³ RIKEN AICS, ⁴ RIKEN ASI

Introduction Replica-exchange Molecular Dynamics (REMD)

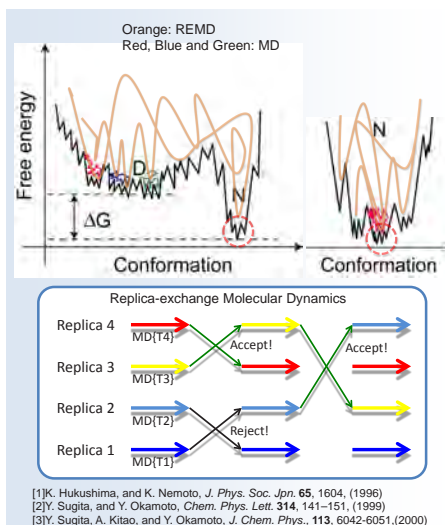
Replica-exchange molecular dynamics (REMD) simulation [2] is one of popular enhanced conformational sampling methods, and it has been widely used to predict the protein conformation, configuration and so on. In the original REMD simulation, the dynamics of noninteracting replicas of the system, each at a specified temperature, are performed by molecular dynamics (MD) simulation. Pairs of replicas that have nearest neighbor temperatures are exchanged every few steps using the Metropolis criterion to ensure detailed balance.

Multi-dimensional replica-exchange method (MREM) [3] is one of the extensions. It exchanges not only nearest neighbor temperatures but also the nearest neighbor parameter sets that relate to a potential energy. Exchange probability can be written:

$$\min[1, \exp(-\Delta)], \quad \Delta = \beta_m(U_{Am}(q_i) - U_{Am}(q_j)) - \beta_n(U_{An}(q_j) - U_{An}(q_i))$$

where, $\beta_m = 1/(k_B T_m)$, T_m : temperature, k_B : Boltzmann constant, and $U_{Am}(q_i)$ is the potential energy of m -th state at i -th replica.

MREM needs huge computer resources, and is promising method for the future enhanced sampling. We have developed Replica-exchange Interface (REIN) program package to perform MREM simulation using conventional MD program packages, such as NAMD2, MARBLE and so on.



[1]K. Hukushima, and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604, (1996)
 [2]Y. Sugita, and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141-151, (1999)
 [3]Y. Sugita, A. Kitao, and Y. Okamoto, *J. Chem. Phys.*, **113**, 6042-6051, (2000)

Replica-exchange interface program (REIN)



- Easy setup and execution of Multi-dimensional REMD for Bio-molecules.
- Utilizing the conventional MD packages.
- Run on PC clusters and K computer.

Feature 1) REIN controls conventional MD program package, which is in a binary format.

→ NAMD2 and MARBLE are available.

Fortran95/03, MPI 2

REIN

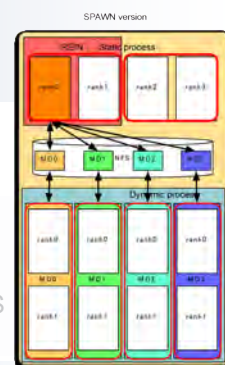
Control

MPI_COMM_SPAWN

Binary format

Conventional MD packages

MARBLE
 NAMD2
 AMBER
 GROMACS
 etc...



Feature 2) Multi-dimensional REMD (MREM)

→ It can exchange not only nearest neighbor temperatures, but also nearest neighbor distances etc... The parameters depend on the function of "constraint" in the used MD package program.

Feature 3) Three supportive software are included

→ Input builder, Batch builder, REMD converter

Now available



http://www.islim.org/islim-dl_j.htm

Twitter: @rein_devel

E-mail: yukimya+rein@gmail.com

One of the applications

Structure prediction of N-glycans in solution

We performed REMD simulation to predict the structures of N-glycans with and without bisecting GlcNAc using REIN [4]. These simulations show that the N-glycan modification like the bisecting GlcNAc can select a certain "key" in the framework of "bunch of keys" mechanism.

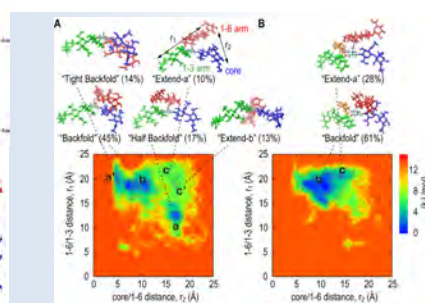
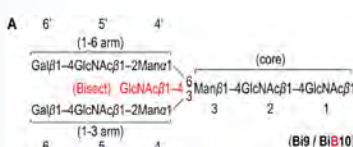
REIN/NAMD2 on RICC

260 hours/512 core

~10,000 atoms, 300K-500K

64 replica x 52 ns = 3.3 micro sec.

[4] S. Re, N. Miyashita, Y. Yamaguchi, Y. Sugita, *Biophys. J.* **101**, L44-L46, (2011)
 [5] Wataru Nishima, Naoyuki Miyashita, Yoshiki Yamaguchi, Yuji Sugita, and Suyong Re, *J. Phys. Chem. B*, **116** (29), pp 8504-8512, (2012)





M-3: MARBLE 全原子分子動力学計算



開発責任者: 池口満徳 (横浜市立大学)

研究概要

出力:産種の履歴

**全原子MD
ソフトウェア**

生体超分子用全原子シミュレーション技術の開発
✓超並列化ソフトウェア
MARBLE
✓世界標準アルゴリズムを装備 (PME等)
✓先端アルゴリズムも装備 (シンプレクティック剛体法等)

**解析
ソフトウェア群**

シミュレーション結果の解析技術の開発
✓立体構造変化
✓自由エネルギー
✓実験との対応
NMR, SAXSなど

**計算対象
生体分子**

多量排出トランスポーター
AcrBの全原子分子動力学シミュレーション
✓分子スケールチームで共通のターゲット
✓膜タンパク質
✓構造変化が機能に重要
✓ゆらぎと構造変化

MARBLE

全原子分子動力学システム MARBLEの特徴

- シンプレクティック部分剛体時間積分法
 - 精度の高い全エネルギー保存
- 相互作用
 - 近距離力と長距離力 (3DFFTを利用) からの分離
- 並列化プログラミング
 - OpenMP対応
 - 空間分割による並列化

実験データによるMARBLEの検証例

NMRでの検証

DNA結合タンパク質 PhoB
水を介した分子認識
構造ダイナミクス検証 (非-計算, 青; 実験) (JACS, 2010)

X線溶液散乱実験での検証

制限酵素 Eco109I
構造中から方向(赤)と構造変化方向(青)の一致
散乱プロファイル(青:計算, ピンク:実験) (Biophys J, 2009)

並列化効率

赤 ~170万原子
OpenMP 8 cores
PME

多剤排出トランスポーターの全原子分子動力学シミュレーション

背景

多剤排出トランスポーター
細胞に有害な多剤(薬剤)を細胞外へ排出する働きを持つタンパク質

★膜質膜上に存在する
★多数の生薬に存在している
★一層のラメラ構造で、様々な薬剤に対応。
★細胞にとって有用なものは外に出さない
★膜内結合(膜貫通膜の膜貫通性質)
★がん細胞の抗がん剤排出

(図1) 大腸菌における多剤排出システム (AcrB/AcrA/TolC)

AcrB
大腸菌などのグラム陰性菌に存在するプロトン駆動の多剤排出トランスポーターの一種

- 大腸菌のもつ薬剤耐性の主要因
- TolC, AcrA とともに環状複合体を形成して細胞膜を貫通し、薬剤排出の中心的役割をなす。
- 排出するもの、排出速度、排出方向が異なる。
- 薬剤の薬剤の種類が異なる。結合型、排出型、排出型(膜貫通膜の膜貫通性質)の異なるモノマーからなるホモ二量体を形成する(図2)
- プロトン駆動力で異なる3状態間を順番に構造変化をすることで薬剤を排出する(図3)

(図2) 薬剤の結合(特) 結合型 排出型

(図3) 薬剤の排出(特) 結合型 排出型

AcrBのドメイン構成

PN1, PN2, PC1, PC2, TM1-TM8, TA, TD, PC1, PC2, PN1, PN2

ポータードメイン

結合型 排出型

結合型 Lys940-Thr978, Asp407, Asp408, Lys940

排出型 Lys940-Thr978, Asp407, Asp408, Lys940

排出型のLys940のコンフォメーションが結合型・排出型と異なる

目的

全原子分子動力学シミュレーションにより、AcrBの多剤排出のメカニズムを解明する。

とくに、各状態のプロトン転移サイトのプロトン化状態を調べ、
→ 各状態の構造を安定に維持することのできるプロトン化状態を調べる

計算方法

System
Protein: (PDB ID: 2DHH) (Non drug containing structure)
Missing part region(499-512) were added from another crystal structure (PDB ID: 2J8S) (47,724 atoms)
Lipid bilayer: POPE 809 molecules (101,125 atoms)
Water: TIP3P model 108,385 molecules (325,155 atoms)
Ion: Na⁺ (36 atoms)
Total ~474,000 atoms
Box size: 160ÅX160ÅX180Å
Program : MARBLE (Ikeguchi)
Force field: CHARMM 27
Electrostatic interaction:
Non cut-off (Particle-mesh Ewald)
Ensemble: NPT,
Temperature: 300K,
Time step: 2fs

結果

構造変化のチェックに用いた原子間距離

結合型
K940-D407, K940-D408, K940-T978

排出型
K940-D407, K940-D408, K940-T978

(Lys940-Thr978 > Lys940-Asp407 = Lys940-Asp408)

排出型
K940-D407, K940-D408, K940-T978

(Lys940-Asp408 > Lys940-Thr978 > Lys940-Asp407)

プロトン転移サイトのコンフォメーション

結合型・排出型 (Asp407, 408を脱プロトン化)

Asp407およびAsp408を脱プロトン状態で、プロトン転移サイトの構造を安定に維持

排出型 (Asp407, 408を様々なプロトン化状態で計算)

Asp408のみをプロトン化したシミュレーション(407-408⁺)では、排出型の状態を安定に保持

Asp408を脱プロトン化

本研究で行ったシミュレーション

Asp407	Asp408	Initial number	Simulation Time
407-408 ⁻	-	2000	300 ns
407-408 ⁺	-	2000	300 ns
407-408 ⁻	+	2000	300 ns
407-408 ⁺	+	2000	300 ns
407-408 ⁻	-	407-408 ⁺ 100 ns	300 ns
407-408 ⁺	-	407-408 ⁻ 100 ns	300 ns

★排出型のAsp407およびAsp408のプロトン化状態の異なる二種類のシミュレーションを行った。

★各シミュレーションとも、結合型と排出型のモノマーのAsp407, Asp408は脱プロトン化状態で行った。

ポータードメインの構造変化

PN1, PN2, PC1, PC2

結合型 (A state, B state)

排出型 (2DHH) 膜貫通ドメイン

排出型(407-408⁻)

★排出型のAsp407, Asp408が脱プロトン化した状態のシミュレーション(407-408⁻)において、ポータードメインの構造が、排出型へと変化した。

ポータードメインの主成分解析(膜貫通ドメインでフィット)

ベリプラズム側から見て、ポータードメインの中心を軸とした、時計回りの運動が、構造変化の主要成分である。

ポータードメイン-膜貫通ドメイン間の Buried surface Area の比較

407-408⁻の排出型は、407-408⁺の排出型と比べて Buried surface Area が減少した(各ドメイン間が離れた)

407-408⁻ (E state) 407-408⁺ (E state)

排出型から結合型への構造変化のメカニズム

ポータードメイン 膜貫通ドメイン

薬剤の入り口 水の入り口

プロトン転移サイト

電荷: -1 脱プロトン化

電荷: -1

電荷: 0

電荷: 0

電荷: -1

電荷: 0

★排出型から結合型への構造変化において、ポータードメインの中心を軸とした、スクリューモーショ(上図の赤矢印)による、PC2サブドメインの動きにより、薬剤の入り口および、水の入り口が形成される。

★これらの構造変化は、プロトン化状態の変化に伴う電荷の変化により生じる静電引力が駆動力となって生じていると考えられる。

★ベリプラズム側からプロトンが移動

★水分子の媒介によりプロトンが移動すると考えられる

謝辞

本研究を実施するに当たり、東京工業大学大学院生命理工学研究科 村上 聡 教授に協力をいただきました。ここに、感謝申し上げます。

まとめ

AcrB量体の膜貫通領域に存在する、プロトン転移サイトのプロトン化状態について、分子動力学シミュレーションにより調べた。

- ★結合型、排出型のプロトンは、Asp407, 408が脱プロトン化した状態が安定であり、排出型では、Asp408がプロトン化した状態が安定である。
- ★排出型のポータードメインの構造が排出型に近づく過程が観察された(407-408⁻)。
- ★排出型から結合型へのポータードメインの構造変化は、膜貫通ドメインに対して、ベリプラズム側へ移動するスクルーモーションであり、この運動による、PC2サブドメインの配置の変化により、薬剤の取り込み口と、水の入り口が開くことが示された。また、この構造変化の駆動力は、プロトン転移サイトの 電荷の変化による静電引力であると考えられる

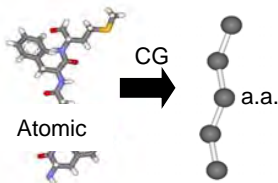


M-4: CafeMol 粗視化モデル計算



開発責任者: 高田彰二 (京都大学理学研究科)

- 概要
 - 粗視化分子モデル計算により大規模生体分子の長時間シミュレーションを行う。
- 離散化(計算モデル化)の方法
 - 粗視化された分子モデルによる古典分子動力学法



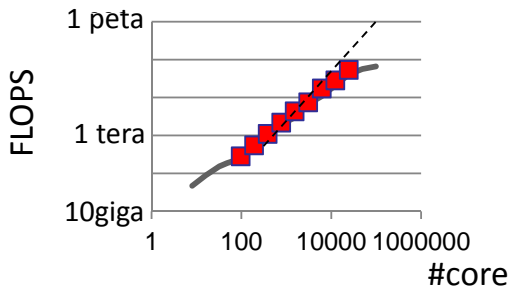
- 計算方法
 - Langevin方程式の時間発展を数値的に積分する。
- 並列化の方法
 - Neighbor list方式、レプリカ交換法
- 開発言語とライブラリ
 - Fortran90, MPI, OpenMP
- コードの公開状況
 - <http://www.cafemol.org> からダウンロード可。ただし生体膜、核酸については要相談。



Version 2.0 (only for proteins), available at www.cafemol.org

Kenzaki et al
J. Chem Theo Comp
2011

- 現状での計算規模
 - 1万残基タンパク質のミリ秒相当のシミュレーションが可能
 - PCクラスターで98000コア並列
 - メモリ容量 2 GB/コア、ディスク容量 1 TB



- 次世代機「京」での計算規模
 - 10万粒子(100万原子)系の秒相当のシミュレーション
 - 反復回数 10^{10} 回以上
 - メモリ容量 2 GB/コア、ディスク容量 1 PB

理論構築: 粗視化モデルの導出

Atomic interaction-based CG model,
Li, Wolynes, Takada, PNAS 2011
Li et al, PNAS 2012

全原子モデルのアミノ酸対エネルギー
全原子モデルとのゆらぎ適合

$$V = \sum_i k_i^r (r^i - r_0^i)^2 + \sum_i k_i^\theta (\theta^i - \theta_0^i)^2 + \sum_i \{ e_{\phi,1}^i [1 - \cos(\phi^i - \phi_0^i)] + e_{\phi,3}^i [1 - \cos 3(\phi^i - \phi_0^i)] \} + \sum_{i>j=3}^{native} e_{ij}^\beta [5(r_0^\beta / r^\beta)^{12} - 6(r_0^\beta / r^\beta)^{10}] + \sum_{i>j=3}^{non-native} e(C / r^\beta)^{12}$$

どんなことができるか

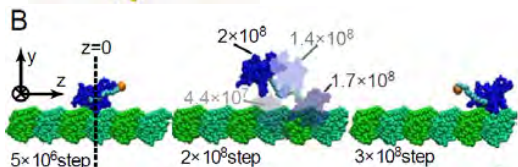
- X線回折やNMRによる構造情報が存在するタンパク質、核酸などの動態の分子動力学シミュレーション。とくに構造変化を伴うタンパク質ドッキング、分子モーターやトランスポーターの大規模構造変化などをシミュレーションできる。
- 例として、キネシンの動態、多剤排出トランスポーターの動態、DNAヒストン複合系(ヌクレオソーム)など。

キネシンの動態



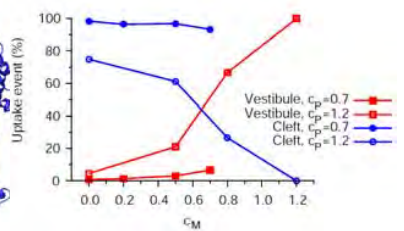
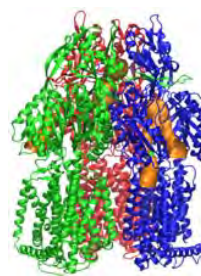
双頭キネシン:
高精度モデリング

単頭キネシン:
運動シミュレーション

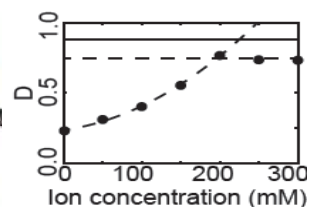
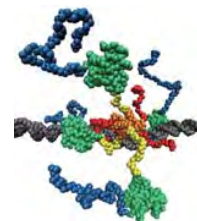


多剤排出トランスポーターAcrBの動態

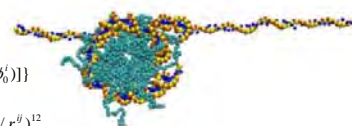
Yao et al, J Comp Chem 2010
Yao et al Nature Comm 2010



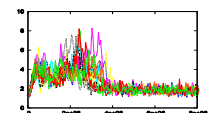
遺伝子動態: p53



遺伝子動態: ヌクレオソーム



機械的アンジッピング





M-5: ProteinDF

球面グリッドの回転によるDFT交換相関数値積分の精度向上

開発責任者: 佐藤文俊 (東京大学生産技術研究所)
 ○松田潤一, 平野敏行



背景

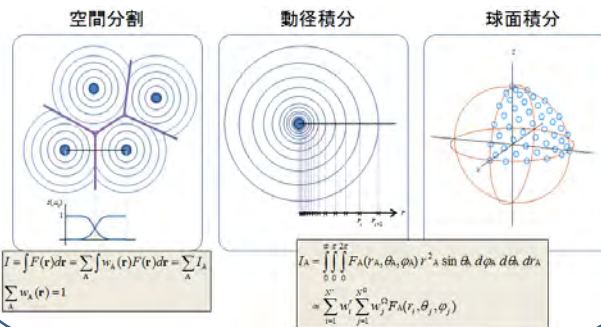
密度汎関数 (DF) 計算プログラム ProteinDF[1] を京で実行することで、これまでにない規模の生体分子全電子計算が可能になる。DF 計算における交換相関 (XC) 項は被積分関数が複雑なため、グリッドによる数値積分を行うのが一般的だが、その精度と演算量が問題になる。

$$\mu_{pq}^{XC} = \int g_p(\mathbf{r}) \mu_{XC}(\mathbf{r}) g_q(\mathbf{r}) d\mathbf{r}$$

化学反応や振動の解析では全エネルギーで約 0.03 kcal/mol の計算精度が要求される。グリッド点数を増やせば精度が向上するが、演算時間が増大し、大規模分子の計算は京をもってしても非現実的なものになってしまう。なるべく少ないグリッド数でより高い精度を実現できることが望ましい。

グリッド法による多中心数値積分

多中心積分を実現するグリッド法は3つの方法の組合せであり、それぞれの手法に関してこれまでに多くの報告がある[2-10]



精度の評価方法

評価にあたっては、被積分関数を電子密度とすることで、全空間積分が電子数となることを利用した。

$$N_{\text{integ}} = \int F(\mathbf{r}) d\mathbf{r} = \int \rho(\mathbf{r}) d\mathbf{r}$$

このとき数値積分は

$$N_{\text{integ}} = \sum_{pq} W^A W^T W^S \rho_p = \sum_{pq} W^A W^T W^S \sum_{pq} P_{pq} g_p g_q$$

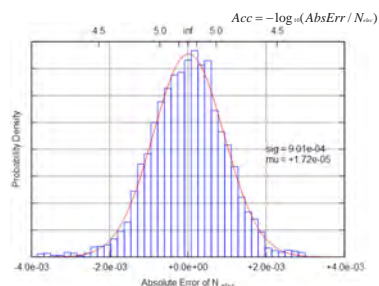
となり、電子数 (整数) が真値となる。

誤差の分布

これまでの我々の研究から、方法の組み合わせよりも、グリッドと分子の角度が、精度により影響すると分かっている。

その詳細を確認するため、アミノ酸1~数残基について、1度刻みでグリッドを回転させて数値積分計算を行い、誤差の分布を調べた。

SG-1相当のグリッドにおいて、報告されている精度[2]よりも1桁低い、4桁程度の精度となる場合が確認された。



グリッド全体を回転したときの誤差分布の例
 対象分子: イノロイン 電子数: 72
 グリッド: Beckeの方法[2] 点数(50,194)

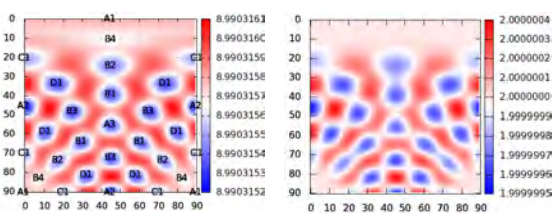
球面グリッドの回転

各球面積分において、回転に対する値の変化を調べたところ、被積分関数を1としたとき、つまり

$$I = \sum_{N_\Omega} w_A w_\Omega$$

の分布が、電子数の誤差分布と類似しており、誤差の桁も近いことが分かった。これより「空間分割重みの球面積分誤差が支配的であり、これを減らせば全体誤差も減る」との仮説を立てた。

I の値を指標とし、リファレンスグリッドの値と近くなるよう、各球面グリッドの角度をニュートン法によってそれぞれ決定するプログラムを実装した。



グリッド回転による誤差の分布
 左: 等核2原子分子モデルでのIの例 右: 水素分子の電子数の分布
 グリッド: Beckeの方法[2] サイズ(50,194) x, y軸: 回転角θ, φ

実験結果

アミノ酸1および3残基について、球面グリッドの回転による積分精度向上の効果を確認した。

いずれも、worst値に対して、平均して1桁以上精度が向上している。

これらは報告されている5~6桁[2]もしくはそれ以上の精度にあたる。

アミノ酸3残基

	this work	worst	ΔAcc
Gly-Ile-Val	6.36	4.42	+1.94
Ile-Val-Glu	7.74	4.52	+3.21
Val-Glu-Gln	5.52	4.55	+0.97
Glu-Gln-Cys	5.78	4.54	+1.24
Gln-Cys-Cys	6.86	4.66	+2.20
Cys-Cys-Thr	6.14	4.66	+1.48

アミノ酸1残基

	this work	worst	ΔAcc
Gly	6.00	4.66	+1.34
Ala	5.55	4.51	+1.04
Ser	6.47	4.49	+1.98
Pro	4.96	4.49	+0.47
Thr	5.35	4.57	+0.78
Val	6.10	4.34	+1.76
Asp	5.07	4.53	+0.54
Asn	5.42	4.51	+0.91
Leu	5.56	4.36	+1.20
Ile	5.61	4.27	+1.34
Gln	5.41	4.51	+0.89
Glu	5.86	4.53	+1.34
Lys	5.75	4.26	+1.50
Met	6.16	4.44	+1.72
His	6.08	4.54	+1.54
Phe	5.27	4.50	+0.76
Arg	5.30	4.27	+1.03
Tyr	5.02	4.52	+0.50
Trp	5.81	4.55	+1.26
Cys-Cys	5.29	4.44	+0.84

数値は Acc = -log10(AbsErr / N_acc)

まとめ

DF計算向けの数値積分について

- ・SG-1相当のグリッドでは4桁の精度となる場合があることを明らかにした
- ・球面グリッドの回転により、グリッド点数を増やすことなく、精度を上げる方法を提案し、その有用性が示唆された

参考文献

- [1] T. Inaba, F. Sato, et al., *Chem. Phys. Lett.*, 2007, **434**, 331
- [2] A. D. Becke, *J. Chem. Phys.*, 1988, **88**, 2547
- [3] C. W. Murray, et al., *Mol. Phys.*, 1993, **78**, 997
- [4] P. M. W. Gill, et al., *Chem. Phys. Lett.*, 1993, **209**, 506
- [5] O. Treutler, et al., *J. Chem. Phys.*, 1995, **102**, 346
- [6] R. E. Stratmann, et al., *Chem. Phys. Lett.*, 1996, **257**, 213
- [7] M. E. Mura, et al., *J. Chem. Phys.*, 1996, **104**, 9848
- [8] H. Ishikawa, et al., *Int. J. Quantum. Chem.*, 1999, **72**, 509
- [9] V. I. Lebedev, *Zh. Vychisl. Mat. Mat. Fiz.*, 1975, **15**, 48
- [10] V. I. Lebedev, *Dokl. Mat.*, 1999, **59**, 477



M-6: Platypus-QM/MM-FE ハイブリッドQM/MM反応自由エネルギー計算



開発責任者: 林 重彦(京都大学大学院理学研究科分子スケール研究開発チーム)

概要

新規な QM/MM 自由エネルギー法 (QM/MM-RWFE-SCF 法) を開発した。QM/MM 自由エネルギー法では、MD シミュレーションによりサンプルされた MM 部分の構造分布により定義される自由エネルギー曲面上で、QM 法によって取り扱われる活性部位分子の最適自由エネルギー構造が決定される。我々は、平均場近似と統計的 reweighting の手法を組み合わせ、さらに QM-MM 間の長距離クーロン相互作用を Ewald 法により適切に考慮することにより、非常に精度の高い効率的な手法を開発した。特に、本手法のスキームでは、QM/MM 法部分の計算と MM 構造サンプリングのための MD シミュレーション部分が完全に分離される。従って、MD シミュレーションに既存の洗練された MD プログラムを使用することが可能になり、計算実行上の大きな柔軟性を達成している。(Ref. Kosugi and Hayashi, JCTC, 8:322-334, 2012.)

理論

QMMM ハミルトニアン $\hat{H}(\mathbf{r}, \mathbf{R}, \mathbf{X}) = \hat{H}^0(\mathbf{r}, \mathbf{R}) + \hat{H}^{QM/MM}(\mathbf{r}, \mathbf{R}, \mathbf{X}) + E^{MM}(\mathbf{X})$

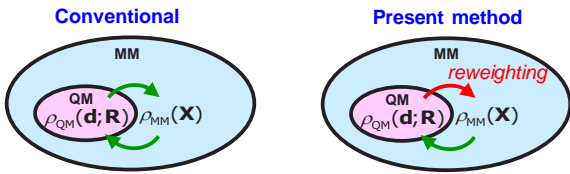
自由エネルギー汎関数 $F[\Psi_{MF}; \mathbf{R}] = -\beta^{-1} \ln \int d\mathbf{X} \exp(-\beta E[\Psi_{MF}(\mathbf{r}; \mathbf{R}); \mathbf{R}, \mathbf{X}])$

Fock or Kohn-Sham 方程式 $\hat{f}^{QMMM}(\mathbf{r}; \mathbf{R}) = \hat{f}^0(\mathbf{r}; \mathbf{R}) + \sum_A^{N_{QM}} \hat{q}_A(\mathbf{r}; \mathbf{R}) \langle V_A(\mathbf{R}, \mathbf{X}) \rangle_{\mathbf{X}, \mathcal{E}[\mathbf{q}(d, \mathbf{R}); \mathbf{R}, \mathbf{X}]}$

静電平均場 $\langle V_A(\mathbf{R}, \mathbf{X}) \rangle_{\mathbf{X}, \mathcal{E}[\mathbf{q}(d, \mathbf{R}); \mathbf{R}, \mathbf{X}]} = \langle V_A(\mathbf{R}, \mathbf{X}) \omega(d, \mathbf{d}_{ref}; \mathbf{R}, \mathbf{R}_{ref}; \mathbf{X}) \rangle_{\mathbf{X}, \mathcal{E}[\mathbf{q}(d_{ref}, \mathbf{R}_{ref}); \mathbf{R}_{ref}, \mathbf{X}]}$

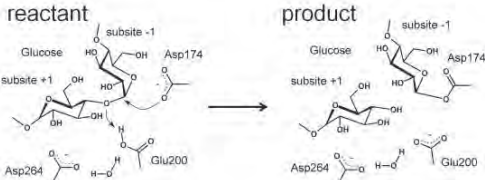
Rewighting factor $\omega(d, \mathbf{d}_{ref}; \mathbf{R}, \mathbf{R}_{ref}; \mathbf{X}) = \frac{\exp(-\beta \Delta E^{QMMM}(d, \mathbf{d}_{ref}; \mathbf{R}, \mathbf{R}_{ref}; \mathbf{X}))}{\langle \exp(-\beta \Delta E^{QMMM}(d, \mathbf{d}_{ref}; \mathbf{R}, \mathbf{R}_{ref}; \mathbf{X})) \rangle_{\mathbf{X}, \mathcal{E}[\mathbf{q}(d_{ref}, \mathbf{R}_{ref}); \mathbf{R}_{ref}, \mathbf{X}]}}$

自由エネルギーの核座標微分 $\frac{\partial F[\Psi_{MF}; \mathbf{R}]}{\partial \mathbf{R}} = \left\langle \frac{\partial E[\Psi_{MF}(\mathbf{r}; \mathbf{R}); \mathbf{R}, \mathbf{X}]}{\partial \mathbf{R}} \right\rangle_{\mathbf{X}, \mathcal{E}[\mathbf{q}, \mathbf{R}, \mathbf{X}]}$



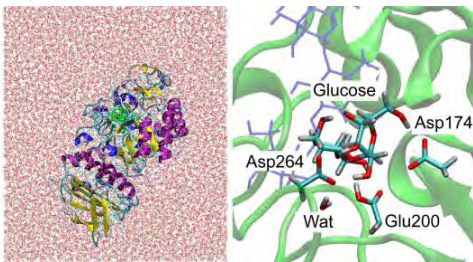
テスト系

α -amylase のグリコシド結合加水分解反応 (の前半)

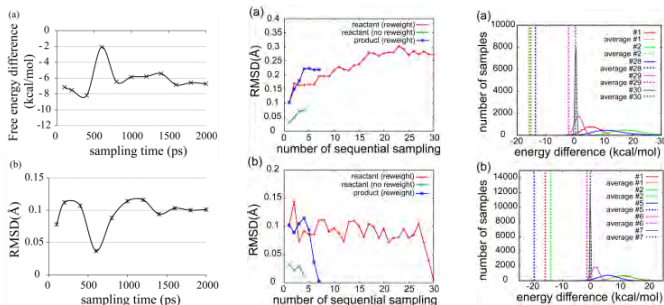


シミュレーション系: 68,533 原子、周期境界条件、PME、Amber

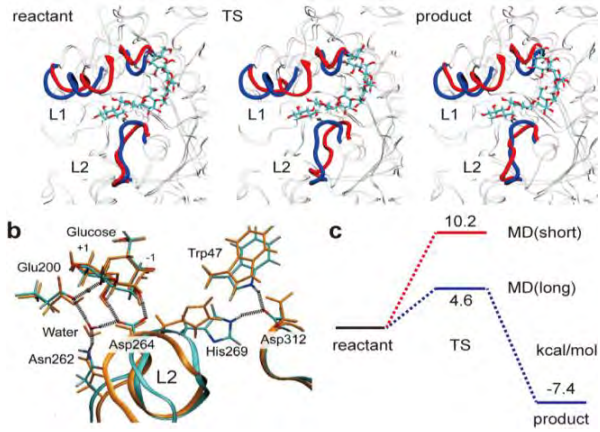
QM 領域: 69 原子、6-31(+)*、629 基底関数、DFT/B3LYP



統計アンサンプルの収束

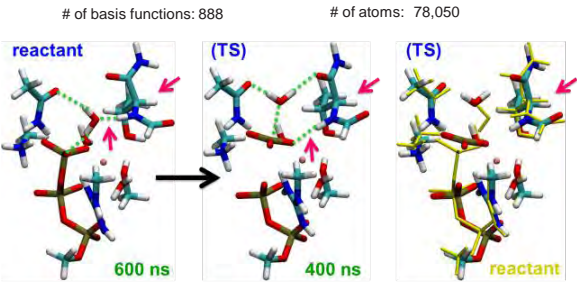


タンパク質の遅い構造変化と酵素化学反応の相関

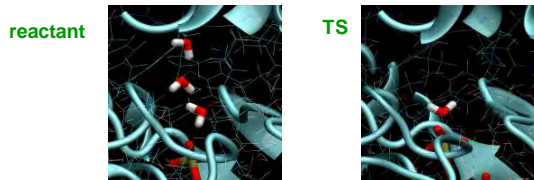


L2 ループの大きな動きが反応の活性化自由エネルギーを大きく下げる

Ras-GAP Gタンパク質 活性化エントロピーによる大きな安定化

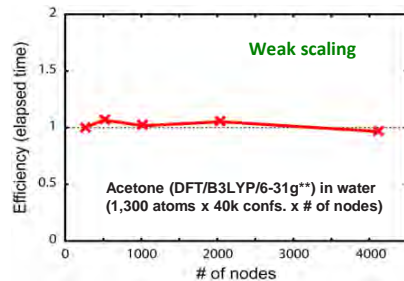


TS 生成に伴い switch-I ループが大きく曲がる



TS 生成に伴い活性部位近傍の水分子が排除される

京上での並列化 MD サンプルで分割



プログラム開発

- ARMCI を導入中 (GAMESS)
- 大きな系に対してチューニング中

謝辞

本プログラムの開発は、大学院生の小杉貴洋さんと井上雄介さん、博士研究員の神谷基司博士との共同研究である。



M-7: Platypus-QM 量子化学計算

開発責任者: 中村春木 (大阪大学蛋白質研究所)



1. Platypus-QM

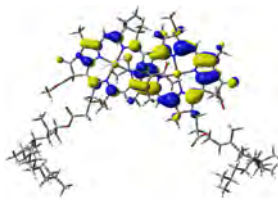
生体高分子の機能の発揮は**局所部位の電子状態**に依存する

基底状態・励起状態の電子構造や化学反応での電子構造変化の信頼性の高いシミュレーションの実施可能な量子化学計算プログラム

- 特徴・ **RHF, UHF, R-DFT, U-DFT, CIS, CIS(D), CIS-DFT, MP2, CASSCF**の実装
- エネルギー・力計算における**積分計算のハイブリッド並列化とSIMD化**
 - 2電子積分計算の**原子軌道基底から分子軌道基底への変換部分の高度化**
 - CASSCF計算における**direct CI計算の高速化** (CAS(16,16)まで可能)

量子化学計算の高速化・超並列化

2. 「京」での並列性能



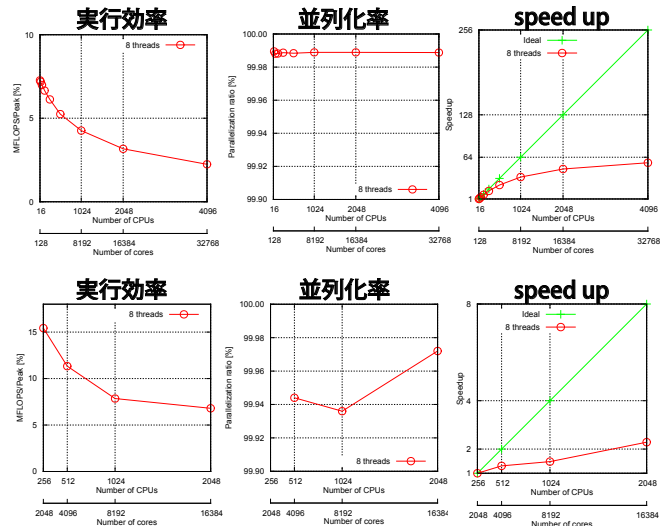
光合成反応中心
スペシャルペア
(280原子)
+
点電荷(30,904個)

RHF/cc-pVDZ

総数**32,768コア**
並列化率 **99.9888%**
実行効率 **7.27~2.24%**
(8192コア : 4.26%)

CASCI(16,16)/6-31G**

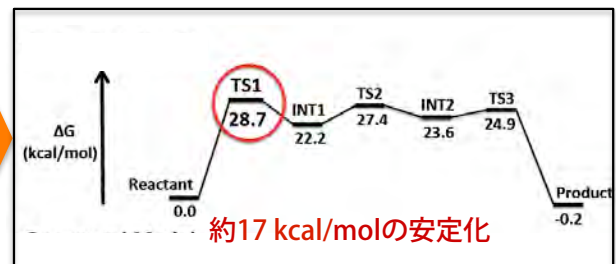
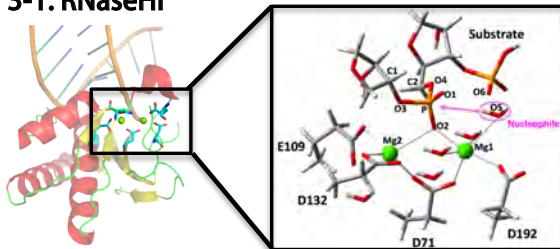
総数**16,384コア**
並列化率 **99.9728%**
実行効率 **15.43~6.80%**
(8192コア : 7.85%)



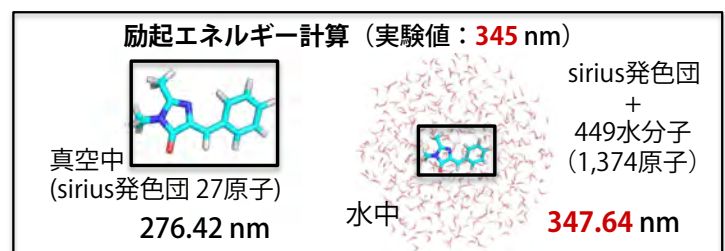
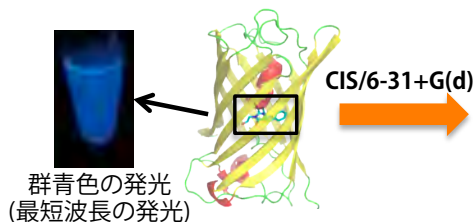
3. Platypus-QMの適用

3-1. RNaseHI

活性中心のみのモデル



3-2. 発光蛋白質sirius





M-8: Platypus-QM/MM

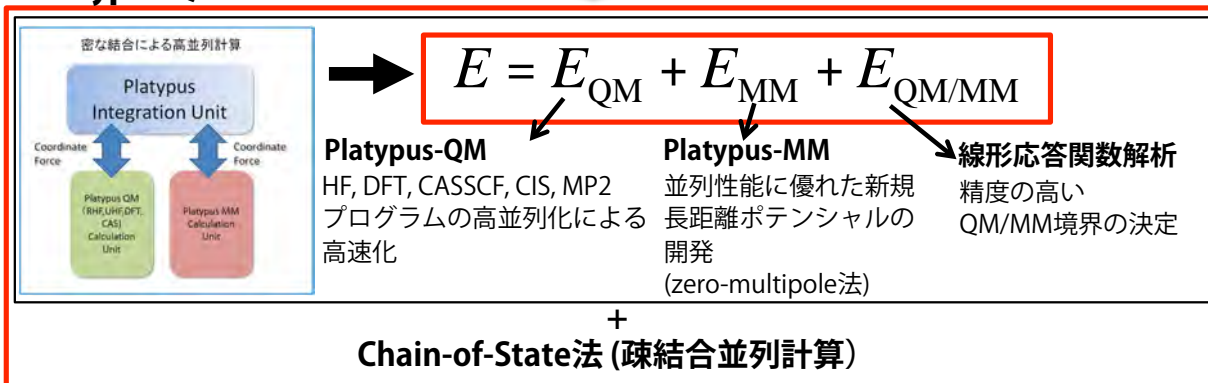
量子化学／分子動力学連成計算

開発責任者: 中村春木 (大阪大学蛋白質研究所)

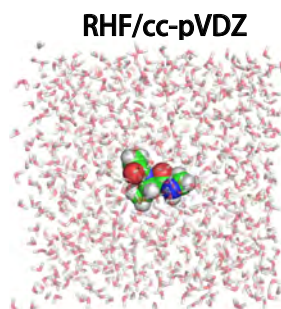
1. 超並列 *ab initio* QM/MD 連成計算プログラム Platypus-QM/MM の開発

- 1) 生体高分子は巨大で、ヘテロで、ダイナミックな系
- 2) 生体高分子の機能の発揮は局所部位の電子状態に依存する

Platypus-QM/MM



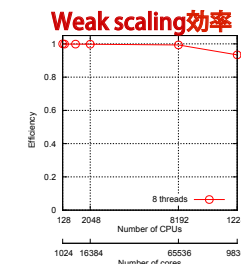
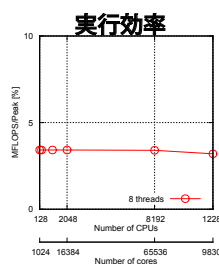
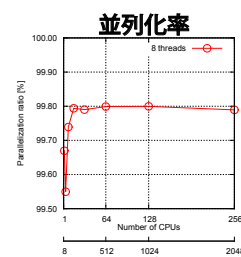
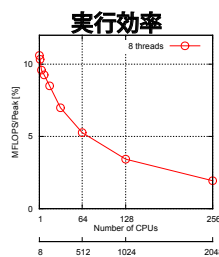
2. 「京」での並列性能



Ace-ALA-NMe
+
1,080 水分子

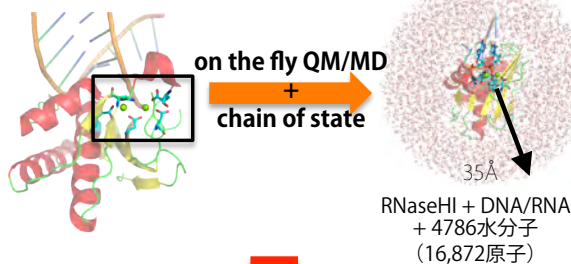
Strong scaling
総数2,048コア
並列化率 99.7893%
実行効率 10.60~1.94%

Weak scaling (Chain of State)
総数98,304コア
128プロセス x 768レプリカ
実行効率 3.42~3.20%
Weak scaling効率: 93.36%



3. Platypus-QM/MMの適用

3-1. RNaseHI



DNA/RNA hybridのRNA部分の加水分解の
酵素反応機構
(最小自由エネルギー経路)

3-2. 発光蛋白質sirius



siriusの発光機構の解明・siriusの発光強度の向上



C-1: RICS 細胞シミュレーションプラットフォーム (RICS: RIKEN Integrated Cell Simulator)



開発責任者: 横田秀夫 須永泰弘 (理化学研究所細胞スケール研究開発チーム)

Motivation: 細胞は複雑な外形形状や内部構造を持ち場所によって代謝反応が異なる。形状を考慮して細胞内反応を計算したい

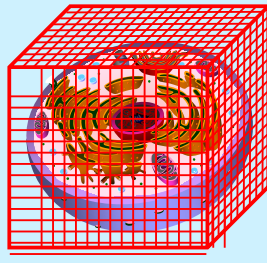
Our Approach: 細胞シミュレーション統合シミュレーションプラットフォームの開発

細胞シミュレーションプラットフォーム RIKEN Cell Simulator (RICS)

～細胞内の場を考慮した生化学反応シミュレーター～

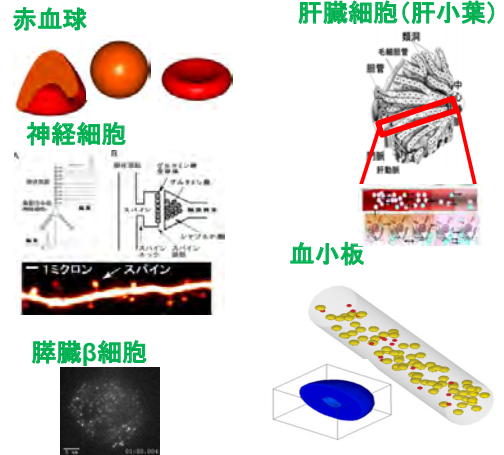
膜透過 + 拡散 + 生化学反応(代謝)

- ・連続体として細胞内を表現
- ・空間をボクセル空間で表現
- ・細胞内小器官などの複雑な空間構造にも対応
- ・複数細胞を同時に計算可能
- ・細胞膜の機能を表現可能

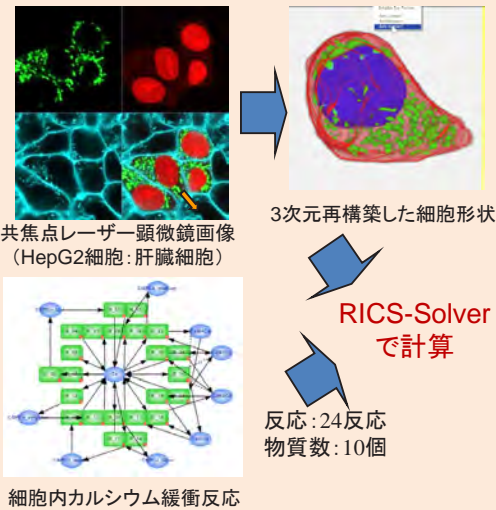


臓器と細胞をつなぐシミュレーションを目指しています。

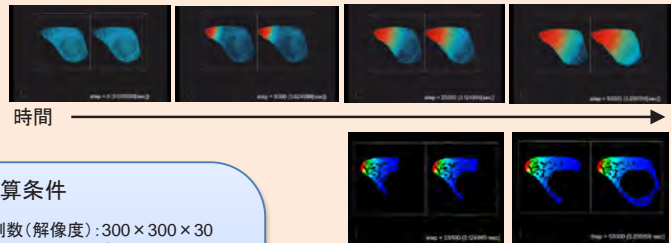
RICSのターゲット



顕微鏡画像から作成した3次元形状内での計算



計算結果 (左: 反応あり、右: 反応なし)



計算条件

分割数(解像度): 300 × 300 × 30
 並列方法: 領域分割
 並列数: 256並列
 $\Delta t = 0.0005$ (秒)
 計算Step数: 20000Step
 使用した計算機:
 RICC超多目的クラスター
 計算実時間:
 反応なし: 34分
 反応あり: 24時間
 Ca^{2+} の拡散係数: 1×10^{-9} (m²/s)

反応があると見かけの拡散速度が遅くなる

細胞内の局所的な反応が細胞内の物質の濃度変化に影響を与えていることを示唆した。

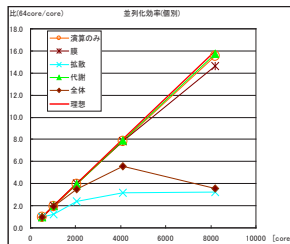
RICCの構成方程式

構成方程式:
$$\frac{\partial \phi_{i,m}}{\partial t} = \nabla C_i \nabla \phi_{i,m} + E_{i,m} + Q_{i,m}$$

物質生成項:
$$E_{i,m} = \lim_{\Delta t \rightarrow 0} \frac{EC_{i,m}(\phi_{i,m}, \Delta t) - \phi_{i,m}}{\Delta t}$$

膜透過項:
$$Q_{i,m} = \sum_n \frac{a_{m,n} (-v_{i,m,n} (\phi_{i,m} - \phi_{i,n}))}{Vol \cdot V_{i,m,n}}$$

$\phi_{i,m}$: 媒質 内の物質 の濃度 $V_{i,m}$: 媒質 のボクセル内体積率
 $EC_{i,m}$: 媒質 内の物質 の反応後の濃度 $V_{i,m}$: 1ボクセルの体積
 $a_{m,n}$: 媒質 間の膜面積 $v_{i,m,n}$: 物質 の媒質 間の膜透過率
 C_i : 物質 の拡散係数



謝辞

本発表の結果の一部は、理化学研究所が実施している京速コンピュータ”京”の試験利用によるものです。



C-2: RICS-Pre

細胞シミュレーションプラットフォーム(RICS)プリシステム

開発責任者: 横田秀夫 須永泰弘 (理化学研究所細胞スケール研究開発チーム)



Motivation: 実際の細胞は複雑な内部構造を持ち場所によって代謝反応が異なるため複雑な形状や初期条件を効率よくRICSに設定したい。

Our Approach: 細胞シミュレーション統合プラットフォームプリシステム (RICS-Pre) の開発

RICS Solver で必要な設定ファイル



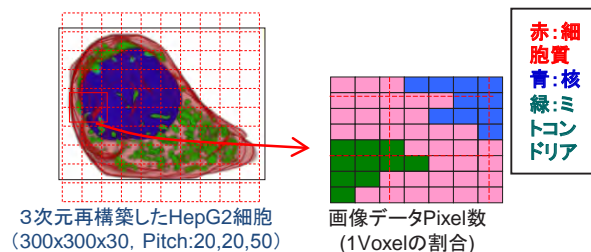
RICS-Preがあると

スペシャリストではなくても、細胞シミュレーションのモデル作成が可能になる。

- ✓ 基本形状の作成が容易
単純な基本形状はCADを用いずRICS-Preで作成
- ✓ 解析条件の設定が容易
入力フォームに従って必要な情報を設定

Voxelデータの解像度変更

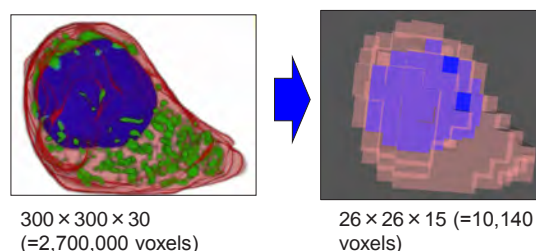
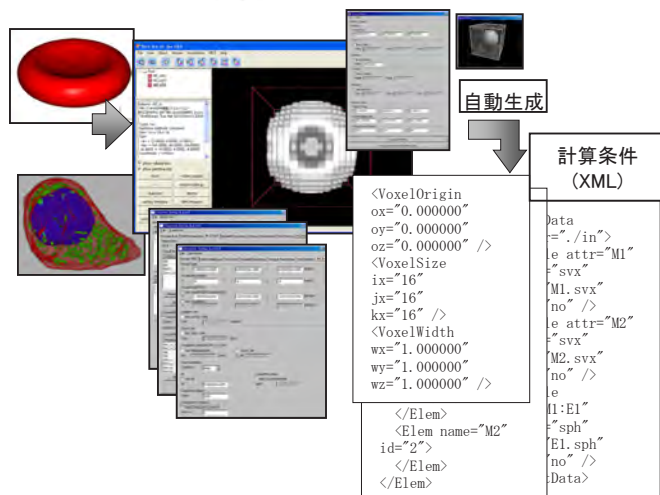
細かい顕微鏡データから精度よく粗い解像度の形状モデルを作成



RICS-Preの機能

- ✓ 実細胞形状及びCAD形状の取り込み
- ✓ 基本形状の生成
- ✓ 形状の修正
- ✓ 細胞内の媒質、物質パラメータ等の設定
- ✓ 反応経路及び膜機能を三次元データへ設定
- ✓ 計算条件の設定

RICS-Preのフロー



謝辞

本発表の結果の一部は、理化学研究所が実施している京速コンピュータ"京"の試験利用によるものです。



C-3: RICS-platelets

RICSによる血小板細胞シミュレーションの検討

ISLiM

七澤 洋平、田村 典子、塩崎 聖治、岡 秀樹、後藤 信哉 (東海大学医学部 臓器/細胞スケール研究開発チーム)



細胞内部の代謝反応/シグナル拡散に着目し、細胞内の物質/オルガネラの位置情報を考慮した血小板細胞内での代謝拡散反応について、理研細胞シミュレーション統合プラットフォーム(RICS)を用いての計算機によるシミュレーションを行なった。RICSとは、細胞を複数のボクセルに区画し、そこに実測データより得られた細胞内の物質/移動量などの情報を取り込み、細胞内の現象を表現可能とする細胞シミュレーション統合プラットフォームである。

前年度までに、RICSを用いた血小板細胞モデルを構築し(図1)、細胞外からの細胞膜への刺激に応じた、血小板細胞内オルガネラである濃染顆粒からのADP、ATP、セロトニンなどの内包物質放出を各種生化学反応の積み重ねにより表現を可能とした。また、複数の血小板細胞を計算機中に設定し、ある細胞にはGP1baへと局所的にvWFによる刺激を与え、その箇所から始まる活性化プロセスを経て濃染顆粒からの内包物質放出、更に別細胞へと放出物質による活性化の誘起を確認した(図2)。

本年度の成果として、代謝反応マップの精緻化を進めるため、血小板内外の代謝反応系をECELLを用いてシステム化し、計算結果と実験結果の比較を可能なものとした(図3)。また、RICSでは移流拡散中での生化学反応の計算が可能である。血小板細胞外における凝固系を単純化したモデルの計算を行った(図4,5,6)。

これらの個々に研究を進めた血小板モデルを統合することで、より実現現象を再現することが可能な血小板モデルの構築を目指す。

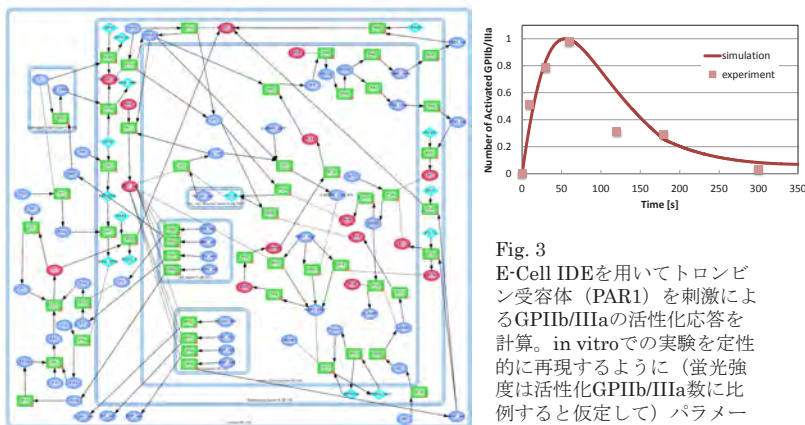


Fig. 3 E-Cell IDEを用いてトロンビン受容体 (PAR1) を刺激によるGPIIb/IIIaの活性化応答を計算。in vitroでの実験を定性的に再現するように(蛍光強度は活性化GPIIb/IIIa数に比例すると仮定して)パラメータを設定。

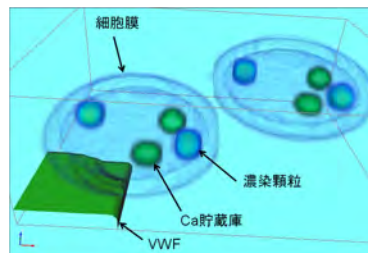


Fig. 1. RICSにて、複数細胞を考慮しての血小板細胞モデルの構築

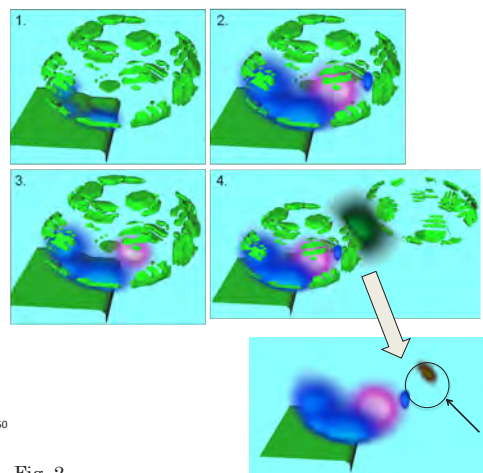


Fig. 2. 1. 局在するGP1baがvWFにより刺激される
2. GP1baから放出されるシグナル物質による刺激によりCa貯蔵庫からCaが放出
3. 細胞内Ca濃度上昇によって濃染顆粒が刺激される
4. 濃染顆粒に含まれる物質の放出。周囲細胞へと活性化に繋がる物質を伝える。また、自らのレセプターを刺激することで活性化を促進させるフィードバック反応も引き起こす。矢印で示した箇所が濃染顆粒からの放出箇所周囲のADPレセプターP2Y12の反応である。

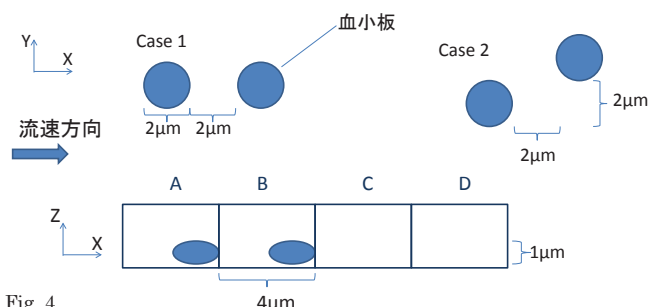


Fig. 4 血小板Likeの構造体を設定し、流れに対して直列に配置する場合とずらして配置する場合での移流による影響について計算を行った。細胞外において2種の反応を定義した。PT+AP->T+AP...(1), FG+T->F+T...(2)(PT:プロトロンビン, T:トロンビン, AP:血小板活性部位, FG:フィブリノーゲン, F:フィブリン)。なお、血小板の全体の初期状態は活性化状態とし、式2についてはTの濃度が一定値以上の場合のみ機能するものとする。拡散係数はPT, FG, T, F = 500, 500, 50, 10[µm²/s]とし、初期状態で細胞外においてPT, FGは1µMが存在し、血小板付近の流速は最大100µm/sとした。

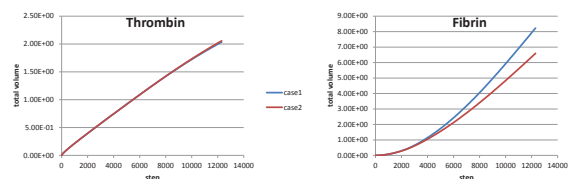


Fig. 5 それぞれの物質の産生量について、トロンビンはほぼ同量、フィブリンはCase1が高濃度となった。

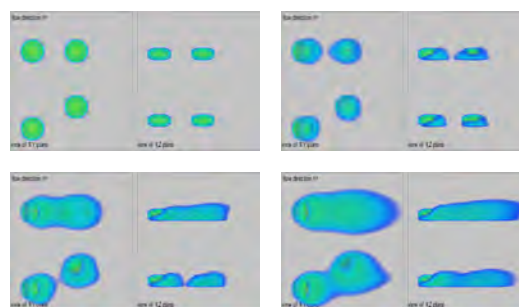


Fig.6 領域を4分割し、式2のトロンビン閾値を超えるボクセルをカウントした。全体としては、Case1>Case2となり、特にBの領域ではもっとも多くのボクセルが閾値を超えていた。本研究の条件においては、Case1が、血小板が流れ方向に対して直列に並ぶことで血小板間に作用する流速が低下し、そこにトロンビンが蓄積された結果だと判断できる。



C-4

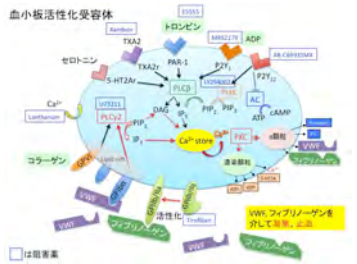
血小板活性化モデルのための細胞内カルシウムイオン実証実験

田村典子¹⁾ 七澤洋平¹⁾ 後藤信哉¹⁾
 1) 東海大学医学部内科学系循環器内科



背景

カルシウムイオン(Ca²⁺)は細胞内のシグナル伝達では、セカンドメッセンジャーとして働く重要な因子である。血小板では活性化物質により刺激を受けると、初期活性化として、カルシウムストアである小胞体からCa²⁺が放出される。細胞内Ca²⁺濃度([Ca²⁺]_i)の上昇により、血小板の活性化シグナルは惹起される。なかでも血小板凝集を引き起こすADP, ATP, Ca²⁺は濃染顆粒から、血栓形成に必要なフィブリンノーゲンやVon Willebrand因子(VWF)、各種凝固因子などはα顆粒から細胞外へ放出され凝固を伴う血栓形成は細胞内外のCa²⁺により促進される。



目的

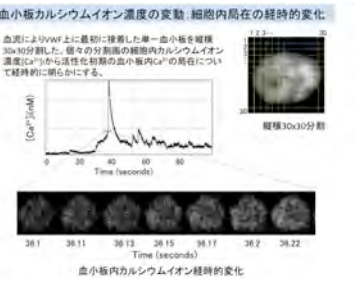
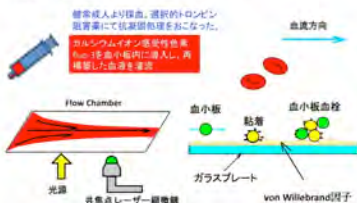
血流条件下、VWF上血小板の初期活性化にともなうカルシウムイオンの上昇について

- 1) 反応時間と濃度
 - 2) カルシウムイオン局在の経時的変化
 - 3) 血小板の[Ca²⁺]_i拡散速度
- を定量化し、RISCを用いた血小板活性化シミュレーションモデルの実証実験として明らかにする。

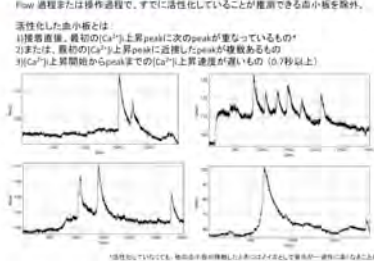
方法

- ・ フローチャンバーを用いた血流条件下のVWF上の血小板内カルシウムイオンの検出を共焦点レーザー顕微鏡にて行ったリアルタイムイメージングを行った。
- ・ 健康人より採血し、抗トロンビン薬(アルガトロン)にて抗凝固処理を行った。全血中の血小板にカルシウム蛍光指示薬Fluo-3 AMを導入し血液サンプルとした。
- ・ ヒトの血漿タンパクから精製したVon Willebrand 因子(VWF)を固相化したガラスプレート上に、ヒトの動脈血流に相当する速度1500 s⁻¹にて全血を灌流した。
- ・ VWF上に接着する血小板イメージを30x30の格子に分割し各格子のカルシウム濃度の経時的変化より細胞内のカルシウムイオンの局在変化を解析した。

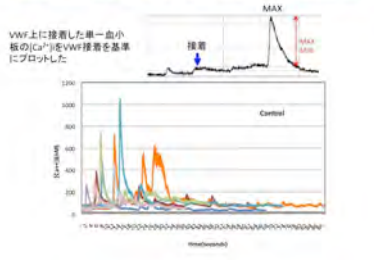
フローチャンバーシステムを用いた実証実験
 血流条件下、マリス上に接着した血小板の血液凝固活性に由来するカルシウムイオンをリアルタイムにイメージングするシステム



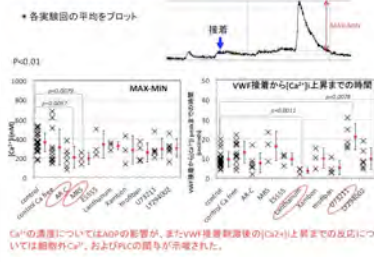
血小板[Ca²⁺]_iサンプルの選択



実験1: VWF接着後の血小板[Ca²⁺]_i上昇時の反応時間と濃度



結果1: 血流条件下、VWF接着直後の[Ca²⁺]_i上昇におよぼす因子の検出



実験2: 血小板の[Ca²⁺]_i上昇時の局在について、VWF接着直後およびVWFに接着後時間が経過した活性化後の[Ca²⁺]_i上昇と局在の比較

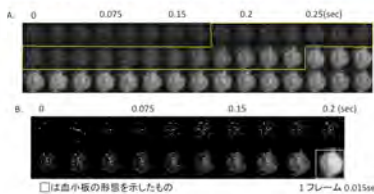
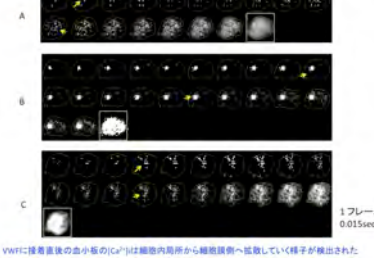


図1 血流条件下、VWF上単一血小板Ca²⁺の経時的局在

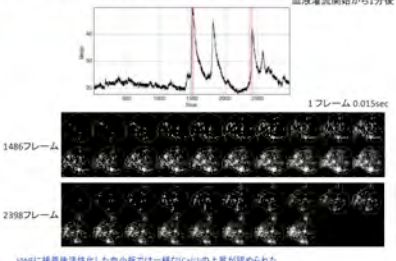
A. Fluo3 AMによる血小板内Ca²⁺のイメージング
 B. 黄色で囲んだ部分の各画素の蛍光強度が高いシグナルを経時的に積算したイメージ

VWF接着直後の血小板Ca²⁺の局在



VWFに接着直後の血小板のCa²⁺は細胞内局所から細胞膜へと拡散していく様子が見出された。

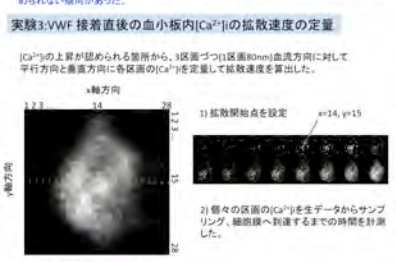
VWF上に接着後、活性化のすんだ血小板Ca²⁺の局在



結果2: 血小板[Ca²⁺]_i上昇時の局在について、VWF接着直後およびVWF接着後の活性化した血小板の[Ca²⁺]_i上昇と局在の比較

	control	Activated platelet	PLC阻害時	細胞外Ca ²⁺ Free
細胞内の局所から[Ca ²⁺] _i 上昇	41/54例 (75.9%)	9/30例 (30%)	6/25例 (24%)	32/36例 (88.9%)
細胞質より一様に[Ca ²⁺] _i 上昇	13/54例 (23.6%)	21/30例 (70%)	19/25例 (76%)	4/36 (11.1%)

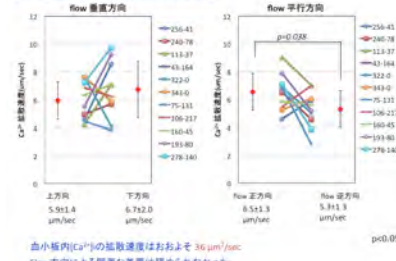
実験3: VWF接着直後の血小板内[Ca²⁺]_iの拡散速度の定量



実際のExcelデータ

Flow direction	Flow speed (μm/sec)	Ca ²⁺ diffusion rate (μm/sec)
上方向	5.9±1.8	5.9±1.8
下方向	6.7±2.0	6.7±2.0
flow 垂直方向	5.5±1.3	5.5±1.3
flow 平行方向	5.3±1.3	5.3±1.3

結果3: 血小板内[Ca²⁺]_iの拡散速度の定量



総括

血流条件下、VWFに起因する血小板の初期活性化にともなうカルシウムイオンの上昇には、活性化動態によるカルシウムイオンの局在変化が認められた。

分子動力学法による平衡状態におけるコフィリン修飾アクチンフィラメントのエネルギー解析

藤井徹矢¹・井上康博^{1,2}・安達泰治^{1,2} (1京大院・工・マイクロエンジニアリング専攻 2京大・再生医科学研究所・バイオメカニクス研究領域)

背景と目的

アクチン細胞骨格は、細胞内で絶えず動的な再構築を行うことにより、細胞の力学的環境変化にตอบสนองすることが示唆されている。アクチン細胞骨格の構成要素であるアクチンフィラメントは、単量体のG-アクチンが多数重合した二重らせん構造を有しており、種々のアクチン関連タンパク質(ARP)と結合することで、その振る舞いに変化する。本研究では、ARPの一つであるコフィリン(Cofilin)に着目した。コフィリンはアクチンフィラメントの脱重合・切断に関与し、アクチン細胞骨格の動的な再構築を促進する。アクチン-コフィリン間の結合親和性は張力に依存することが報告されているが、その分子レベルでのメカニズムは未だに明らかにされていない。これを解明するためには、実験的手法では観察の難しいミクロスケールの立体構造をシミュレーションできる分子動力学(Molecular Dynamics: MD)法が有用である。そこで本研究ではMDシミュレーションを用いて、張力無負荷時の平衡状態におけるコフィリン修飾アクチンフィラメント(Cofil AF)の立体構造を計算し、アクチン-コフィリン間・アクチン-アクチン間の相互作用エネルギーとフィラメントの立体構造の関係に着目して解析を行った。

方法

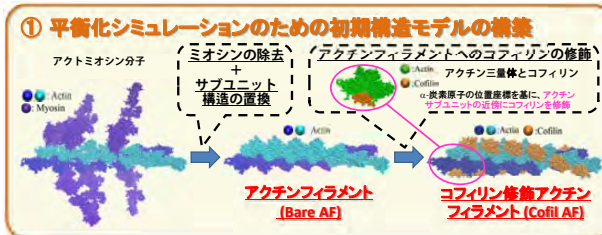
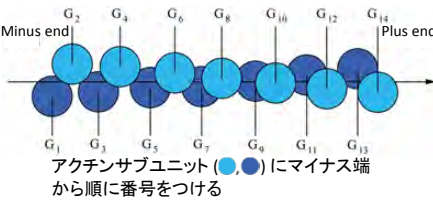
分子動力学(MD)法

一つ一つの原子に対しニュートンの運動方程式を適用することで、分子の振る舞いを観察

Newton's equation of motion

$$m_i \frac{d^2 x_i}{dt^2} = - \sum_{j \neq i} \nabla_i \phi(x_{ij})$$

アクチンフィラメントの模式図



平衡化の評価

平均二乗偏差(RMSD)

システムの平衡化が十分行われていれば、RMSD値は収束する

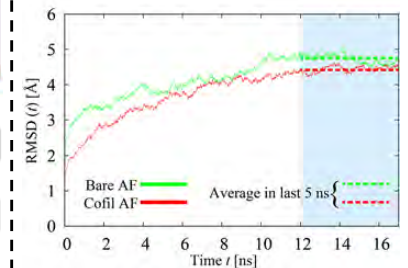


Fig. 1 Bare AFおよびCofil AFのRMSDの時間変化

両フィラメントは平衡化されたと考えられる

結果

ねじれ角

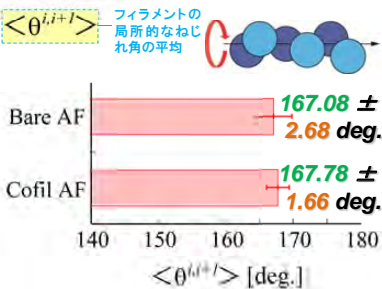


Fig. 2 Bare AFおよびCofil AFの平衡状態におけるねじれ角

- 平均値に有意差はない → 平衡構造のねじれ角は変化しない
- 標準偏差の減少 → 揺らぎの減少

コフィリン修飾アクチンフィラメントのエネルギー解析

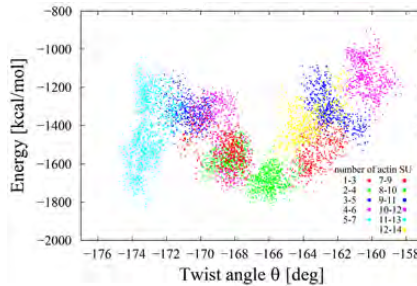


Fig. 3 Cofil AFにおけるねじれ角とコフィリン-アクチン間エネルギーの関係

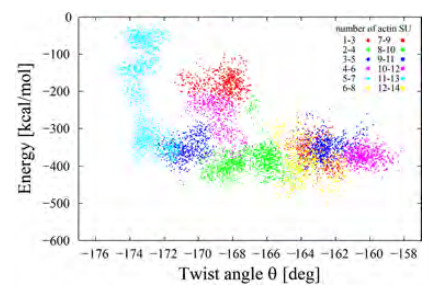


Fig. 4 Cofil AFにおけるねじれ角とアクチン-アクチン間エネルギーの関係

- エネルギーはねじれ角に依存している(Fig. 3, Fig. 4)
- コフィリン-アクチン間のエネルギーに2つの極小値が見られた(Fig. 3)

考察

ねじれ角-174度付近でアクチン-コフィリン間エネルギーの極小値が見られたサブユニット11-13には、次の特徴が見られた。

- アクチン同士の間相互作用エネルギーは0付近に分布している。(Fig. 4)
- 重心間距離は増大していて、12~17 nsにおいては、他の5つのアクチンサブユニット間の距離よりも長い。(Fig. 5)
- 平衡化計算初期の5 nsでは、両サブユニット間に形成される水素結合の数は、平均で約3個なのに対し、12~17 nsではほとんど観測されなかった。(Fig. 6)

以上3点は、コフィリン修飾アクチンフィラメントがサブユニット11-13間で切断されたことを示唆している。このようなアクチンフィラメントの立体構造変化は、コフィリン-アクチン間で相互作用する特定のアミノ酸残基によって引き起こされた可能性がある。

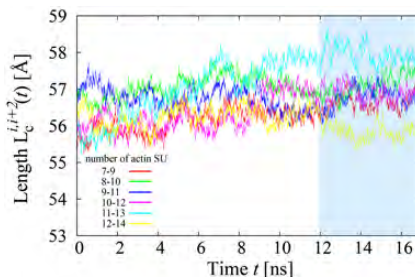


Fig. 5 Cofil AFにおけるアクチンサブユニット間距離の時間変化

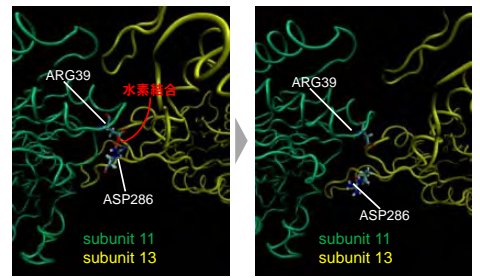


Fig. 6 Cofil AFのサブユニット11-13間において、初期状態からの5ns間で水素結合の形成頻度が最も高かったアミノ酸残基のペア (a) t = 2 ns (b) t = 17 ns

本研究の一部は、日本学術振興会最先端・次世代研究開発支援プログラム「生体システムの構造・機能適応ダイナミクスの力学的理解」(LR017)および文部科学省科学研究費補助金(22616602,23127506)による援助を受けた。記して謝意を表す。



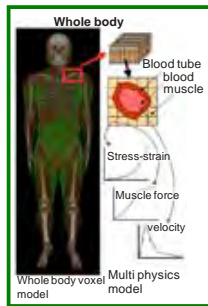
O-1: ZZ-EFSI 全身ボクセルシミュレーション (ボクセル構造流体連成解析プログラム)

開発責任者: 高木周 (理化学研究所 次世代計算科学研究開発プログラム)



開発に至る背景

- 医療画像(CTスキャン, MRI等)と相性の良い流体構造連成解析により, 個人毎の診断に即した治療支援を実現したい.
- 赤血球・血小板を含む血流を対象とする超大規模並列計算を容易に実現し, 血小板血栓に至る機構を明らかにしたい.



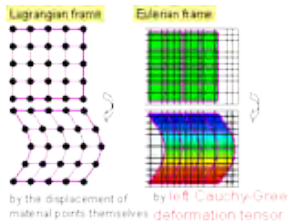
開発ソフトウェアの特徴

大規模並列計算に適した流体構造/膜連成解析を実現

Euler型流体・構造/膜連成解析法

- 固定メッシュ上で, 全ての物理量を更新できるように定式化.
- voxelデータの利用により, メッシュ生成・再構成を不要.
- 複雑な境界形状を持つ/多数の分散体を含む問題の扱いを容易に.

How is the solid deformation described?



動的パラメータを持つ擬似圧縮性法 (Artificial Compressibility Method; ACM)

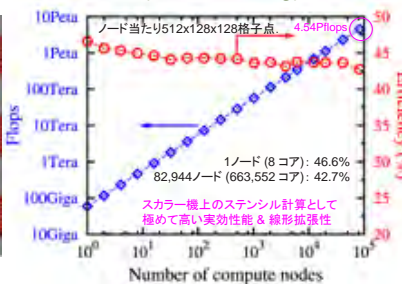
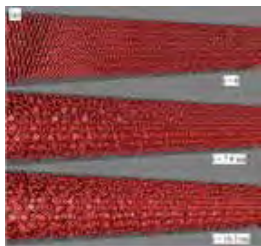
- 圧力修正に際して, 速度発散を最小化する処理を導入.
- 数値的に安定で完全陽的な時間積分.
- 高い実効性能, 並列性能を実現.

現時点での開発状況

- 開発言語とライブラリ
FORTRAN, C++, MPI, OpenMP, VSphere
- 「京」のマニュアルに従い, プロファイラ, PAツールを使って性能チューニング.
- 「京」で, ステンシル計算として極めて高い実効性能, 台数効果を達成.

血管内における多数の分散体

実効性能と並列台数 (weak scaling)



謝辞

本研究は「次世代計算科学研究開発プログラム」の助成を受けて行なわれた。
本研究の計算結果の一部は, RIKEN Integrated Cluster of Clusters (RICC) システムを利用して得られた。

参考文献

Sugiyama, K., Ii, S., Takeuchi, S., Takagi, S. and Matsumoto, Y. (2010) Full Eulerian simulations of biconcave neo-Hookean particles in a Poiseuille flow, *Comput. Mech.*, **46**: 147-157.

Nagano, N., Sugiyama, K., Takeuchi, S., Ii, S., Takagi, S. and Matsumoto, Y. (2010) Full-Eulerian finite-difference simulation of fluid flow in hyperelastic wavy channel, *J. Fluid Sci. Tech.*, **5**: 475-490.

Sugiyama, K., Ii, S., Takeuchi, S., Takagi, S. and Matsumoto, Y. (2011) A full Eulerian finite difference approach for solving fluid-structure coupling problems, *J. Comput. Phys.*, **230**: 596-627.

Ii, S., Sugiyama, K., Takeuchi, S., Takagi, S. and Matsumoto, Y. (2011) An implicit full Eulerian method for the fluid-structure interaction problem, *Int. J. Numer. Meth. Fluids*, **65**: 150-165.

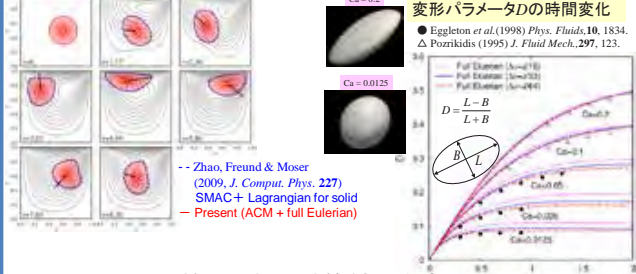
Ii, S., Gong, X., Sugiyama, K., Wu, J., Huang, H., and Takagi, S. (2012) A full Eulerian fluid-membrane coupling method with a smoothed volume-of-fluid approach, *Comm. Comput. Phys.*, **12**: 544-576.

Takagi, S., Sugiyama, K., Ii, S., and Matsumoto, Y. (2012) A review of full Eulerian methods for fluid-structure interaction problems, *J. Appl. Mech.*, **79**: 010911.

Ii, S., Sugiyama, K., Takagi, S. and Matsumoto, Y. (2012) A computational blood flow analysis in a capillary vessel including multiple red blood cells and platelets, *J. Biomech. Sci. Engrg.*, **7**: 72-83.

計算の妥当性 (流体構造/膜連成機能の実装)

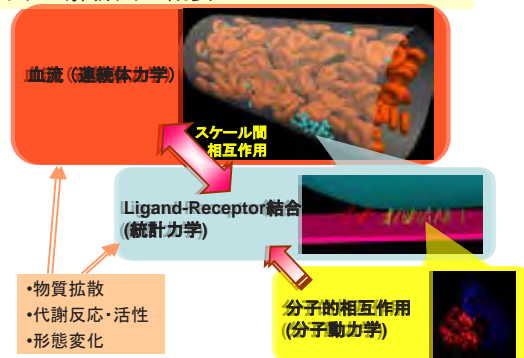
キャビティ流れ中の超弾性体粒子 せん断流中の膜カプセル



十分に検証された計算結果を良く再現
現在開発/実施中の計算事例 (血栓シミュレータ)

血小板血栓に至る機構解明を目指す

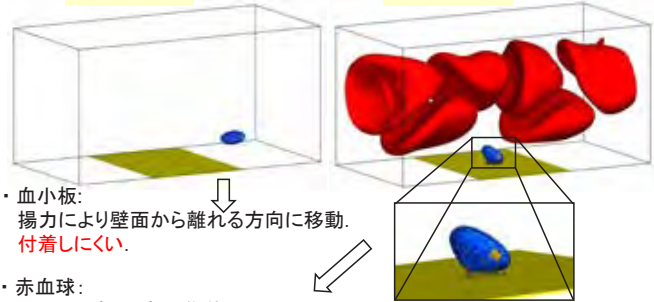
血流とLigand-Receptor結合のマルチスケール/
フィジクス解析法の概要



血小板付着の様子, 赤血球の影響

赤血球なし

赤血球あり



- 血小板: 揚力により壁面から離れる方向に移動。付着しにくい。
- 赤血球: チャンネル流コア部に集積。壁面垂直方向の速度成分の変動を誘起。
- 血小板: 壁面に接近する機会が増え, 付着しやすくなる。

血栓形成の第一段階における, 赤血球の存在の重要性。

⇒ 今後は, 多数の赤血球が存在する条件で, 血小板の付着から血栓の初期過程を「京」で解析



O-2

血栓形成のマルチスケールシミュレーション

- 血小板の血管壁への接着に関するタンパク質間相互作用の解析 -

塩崎聖治 (東海大), 伊井仁志 (大阪大), 杉山和靖 (理研), 後藤信哉 (東海大), 高木周 (東大)



研究の背景

本研究では、血小板の凝集による血栓の形成についてマルチスケール解析を行っています。血栓の形成の初期過程の概要は次の通りです。

- 血管壁が損傷するとその表面は血管内皮細胞下のコラーゲンが露出し、血漿中の von Willebrand 因子 (vWF) が結合する。
- vWF が血小板を介して結合する (一次凝集)。
- 粘着した血小板は活性化され、さらにその周囲の血小板も二次的に活性化される。
- 二次的に活性化された表面には血小板膜糖蛋白の GPIIb/IIIa が発現し、GPIIb/IIIa とうし vWF やフィブリンが架橋し、血小板が凝固する (二次凝集)。
- 二次凝集によって形成された血小板血栓上で血液凝固因子が活性化され、フィブリン繊維が形成され血小板血栓を覆い、より強固な血栓が形成される。

日本人の死因 (2009年) 厚生労働省

- 悪性新生物: 34%
- 心疾患: 18%
- 脳血管疾患: 12%
- 肺炎: 11%
- その他: 8%
- 肝疾患: 2%
- 腎不全: 2%
- 自殺: 2%
- 不慮の事故: 3%
- 老衰: 4%

血栓症: 30%

左図: 心筋梗塞患者から採取した血栓の SEM 画像
白: 血小板
黄: フィブリン
赤: 赤血球
緑: 白血球

血小形成初期過程の概要

血小板表面には 1 万 5 千から 2 万個の GPIIb/IIIa が存在している。

血小板表面において GPIIb/IIIa は 5 力所への局在が観察されており、その局在が血小板-血管壁間相互作用に大きな影響を与えている可能性が指摘されている。

WVF binding site

GPIIb (glycoprotein IIb)
Shi-Zhong Luo, et al., Blood, 109, pp.603-609, 2007.

WVF A1 domain

D1 D3 A1 A2 A3 D4 B1 B2 B C1 C2 WVF subunit

GPIIb binding site collagen binding site WVF multimer

collagen subendothelial tissue

血小板の吸着
Reininger, A. J., et al., 2006

計算結果

局在密度の制御

Calculation area: 1500nm x 1500nm
4000 GPIIb, 3600 vWF
Localization: 5 places
Localization area: 0.09µm
Density ratio: 2-10

密度制御前後の細胞膜上 GPIIb 分布

局在域, 非局在域の GPIIb 密度

平均的な血小板表面の 1/5 に相当する計算領域の中央部に拡散速度の小さな領域を作成し、GPIIb の局在を模擬した。

計算モデルの概要

計算結果のスナップショット

● GPIIb molecules bound to vWFA1
● unbound GPIIb molecules

血小-血管壁間の吸着力 (0.09µm², 平衡状態)

Parameter	Value	Reference
k (pN/nm)	0.1	J. Kim, 2010
k ₀ (pN/nm)	0.09	
k ₀ ² (s ⁻¹)	0.0027	J. Kim, 2010
K _{eq} = k ₀ /k ₀ ² (-)	10 ¹ - 10 ⁵	
D _{eq} (cm ² /s)	10 ⁻¹²	A. Kusumi, 1993.

血小-血管壁間の吸着力の動的応答 (0.09µm², ratio = 10, K_{eq}=10)

血小-血管壁間の吸着力の動的応答 (0.09µm², ratio = 10, K_{eq}=10)

← GPIIb-vWF 結合の最大値

血小-血管壁間の距離が大きくなるに従って、GPIIb-vWF の 1 本 1 本の吸着力は大きくなるが、結合の数が減少していくため、ある時点で急激に吸着力が減少する。

平均定数が小さいほど GPIIb の局在の影響が大きくなるが、これは GPIIb と vWF のペアが結合を形成しうる範囲内に存在する頻度の違いの影響が大きくなるためである。

計算方法

本研究では、血小板凝集について以下の 3 つのスケールに分けて解析を行い、タンパク質スケールから血流スケールまでを繋ぐマルチスケールシミュレーションを行っています。

- 血管内流れ
- 血小板-血管壁間相互作用モデルを組み込んだ血流のシミュレーションを行っている。
- 血小板-血管壁間相互作用
- 血小板細胞膜上に存在する GPIIb の挙動について、血小板表面の分子動力学モデルを構築し、動的モンテカルロ法による解析を行っている。動的モンテカルロ法を用いる事によって血小板表面に存在する GPIIb の拡散、vWF との結合の形成、及び切断の現象を統計的に取り扱うことが可能となり、各時刻における結合の数を求めることができる。
- GPIIb-vWF 相互作用
- 分子動力学法を用いてタンパク質間相互作用の解析を行います。ここでは自由エネルギーについての計算を行い、細胞表面スケールの解析を念頭に置き、タンパク質の結合、分離の rate のモデル化を行う。

Cell adhesion model

$$R + L \xrightleftharpoons[k_r]{k_f} C$$

R: receptor
L: ligand
C: bond complex
k_f: association rate const.
k_r: dissociation rate const.

Binding force

$$f_b = k(\delta - \delta_0)$$

inner wall of a blood vessel

この発表では「血小板-血管壁間相互作用」について述べる。

Events for KMC (spring model) (M. Dembo, 1987)

- bond formation
- bond breakage
- GPIIb diffusion

$$k_f = k_0^f \exp\left[-\frac{k_b(t-t_0)}{2k_b T}\right]$$

$$k_r = k_0^r \exp\left[\frac{(k-t_0)(t-t_0)}{2k_b T}\right]$$

$$k_f^0 = k_0^f \exp\left[-\frac{\Delta G_{b0}^f}{2k_b T}\right]$$

$$k_r^0 = k_0^r \exp\left[-\frac{\Delta G_{b0}^r + \Delta G_{b0}^f}{2k_b T}\right]$$

Initialize Lattice

Identify All Processes and Rates R_i

Choose a Process α

$$P(\alpha) = \frac{R_\alpha}{\sum_i R_i}$$

Do Process α, Increment Time

$$\Delta t = \frac{1}{\sum_i R_i \ln(r_i)}$$

$$r_i \in (0, 1)$$

KMC Algorithm

自由エネルギー

反応座標

Unstretched state

Forward (結合の形成)
reverse (結合の切断)

Stretched state

自由エネルギー

反応座標

まとめ

血小板一次凝集について以下の知見を得た。

- 血小板表面格子モデルを構築し、GPIIb の挙動について動的モンテカルロ法を用いた解析を行い、タンパク質スケールから血流スケールまでを繋ぐマルチスケールシミュレーションの枠組みを作成した。
- 実験により得られたパラメータを用いた解析から、血小板の吸着力を見積もり、血小板細胞膜上での GPIIb の局在化が血小板-血管壁間相互作用に有意な影響を与えている可能性を示した。

今後の予定

- タンパク質結合部として linear spring でなく、エントロピー弾性を考慮した worm-like chain モデルを採用した解析を行う
- 分子動力学法を用いて GPIIb-vWF 相互作用を解析し、rate をモデル化する

Worm-like chain (WLC) モデル

- l_p: persistence length (持続長)
- L: contour length (伸びきり鎖長)
- x: 伸び

$$F_{ext}(x) = \frac{k_B T}{l_p} \left[\frac{x}{L} + \frac{1}{2} \left(\frac{x}{L} \right)^3 \right]$$

研究体制

次世代計算科学研究開発プログラム
次世代生命体統合シミュレーション研究推進グループ

理化学研究所 高木周、杉山和靖、横田秀夫、野田茂徳、須永泰弘
東大 石川順一、島本憲夫、清水和弥
大阪大学 伊井仁志
東海大学 後藤信哉、岡野樹、田村典子、七澤洋平

敬称略



O-3: ZZ-DOSE based on PHITS 全身ボクセルモンテカルロ重粒子線計算プログラム



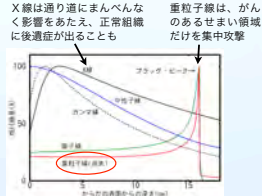
開発責任者: 石川 顕一 (理化学研究所臓器全身スケール研究開発チーム、東京大学大学院工学系研究科)

背景・目的

がんの重粒子線治療

重粒子線 (加速器から得られる炭素イオンのビーム) で、非侵襲にがんを治す治療法

- ▶ 線量の集中性が高い (ピンポイントに照射可能)
- ▶ 強い生物効果 (細胞を死滅させる力が高い)
- ▶ 他の治療法では治らないがん (メラノーマ、肝癌、手術できない肉腫) が治る
- ▶ 同じ治るにしてもより短期間で安全に治りQOLが高い (X線なら20~40回照射、重粒子線なら1回照射)
- ▶ 高度先進医療として、近年、国民の注目が高まっている



1984年 重粒子線治療センター建設計画スタート
1994年 炭素イオン線臨床試験開始
2003年 高度先進医療として承認
放射線医学総合研究所での患者数
1994.6-2009.2の間に4504人 (年々増え2008年度は684人)

重粒子線は目に見えないので、3D-CT画像を使って線量を計算し、治療計画を立てる

→ 信頼性の高い線量計算法が必要

- 既存の計算手法: ベンシルビーム法等
- 長所: 簡便で計算時間が短い (数分)
 - 短所: 不均質な媒質中 (骨の近くなど) で誤差大 → がん病巣へは不十分な線量、正常組織に不要な被ばく

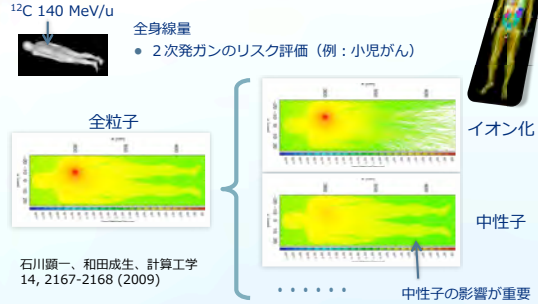
- 様々な組織から構成される不均質な体系についても高精度で線量計算できる手法が必要
- がん病巣への線量を十分に、正常組織の被ばくを低減

モンテカルロ法を用いた線量計算の実用化が切望されている

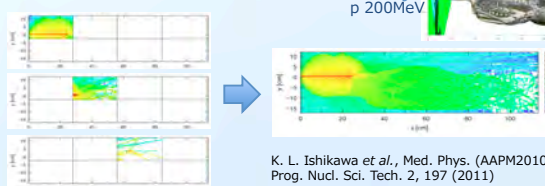
モンテカルロ法を用いた重粒子線線量計算シミュレーターを開発

研究開発成果

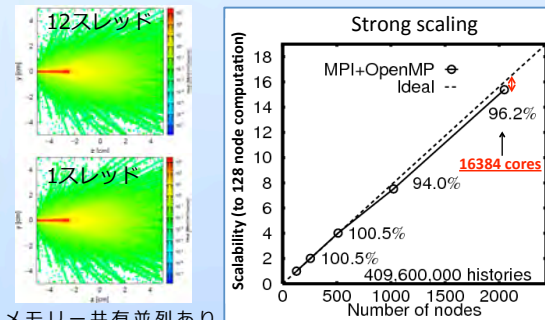
ボクセルファントムを用いた全身線量計算



領域分割モンテカルロ法



OpenMP-MPIハイブリッド並列



メモリー共有並列あり (上) およびなし (下) で計算した水ファントム中の線量分布。全く同じ結果が得られている。

京で良好なストロングスケーリングを得ている。

T. Furuta et al., Proceedings of 6th Japan-Korea Joint Meeting on Medical Physics and 11th Asia-Oceania Congress of Medical Physics (2011)
T. Furuta et al., IFMBE Proceedings 39, 2099 (2012)

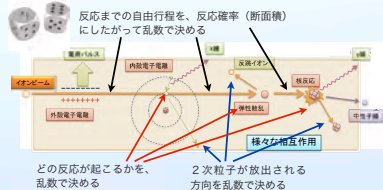
京で131,072コアまでの動作を確認

概要・アプローチ

モンテカルロ法

粒子の輸送や一連の反応を、乱数を用いてそっくりそのままシミュレーション

粒子・重イオン輸送計算コードPHITS [http://phits.jaea.go.jp/] をベースに



- フォノンイオン化が発見し、加速器実験・検出器・放射線遮蔽の計算に広く使われる
- 複雑な体系に対する精度の高い計算が可能
- 入射粒子数を増やすことによって、統計的精度を上げられる
- 並列化効率が高い

モンテカルロ粒子輸送計算では、通常、ヒストリー (入射粒子) を並列化し、すべてのProcessing Elementが全体系の情報を保持。

→ 全身線量計算では膨大なメモリーが必要 (コア当たり数~数十ギガバイト) 京のスペック超え!

→ 領域分割モンテカルロ法

→ OpenMP-MPIハイブリッド並列

現状と展望

OpenMP-MPIハイブリッド並列をPHITSに実装し、version 2.50として公開済み (http://phits.jaea.go.jp/getj.html)

京での全身ボクセル線量計算

展望

2次発がんリスクの評価

がん治療施設との共同研究 (国立がん研究センター東病院、順天堂大学)

本ポスターの結果の一部は、理化学研究所が実施している京速コンピュータ「京」の試験利用によるものです。



O-4: ZZ-HIFU

集束超音波による低侵襲治療のためのHIFUシミュレータ

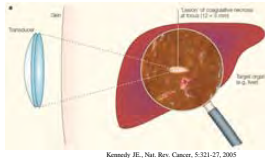


開発責任者: 松本洋一郎 (東京大学工学系研究科機械工学専攻)

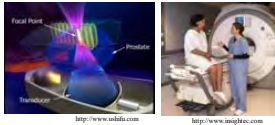
1. Background and Objectives

High-Intensity Focused Ultrasound therapy

- Minimally-invasive treatment, which provides the heat coagulation of tissue around the focal point of ultrasound.
- Benefits
 - Short hospital stay
 - Radiation free
 - Repeatable treatment
- Approved for tumors such as the prostate hyperplasia and uterine fibroids.



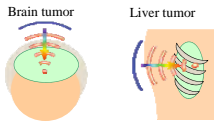
Kennedy JE, Nat. Rev. Cancer, 5:321-27, 2005



MRI-guided Focused Ultrasound Surgery has

- been developed for various tumors.
- Pre and Post contrast imaging for treatment validation
- Accurate targeting
- Real-time monitoring of temperature

HIFU for tumor behind bones



Displacement and diffusion of focal point due to the reflection and refraction of ultrasound at the interface of bones



Focus control by array transducer

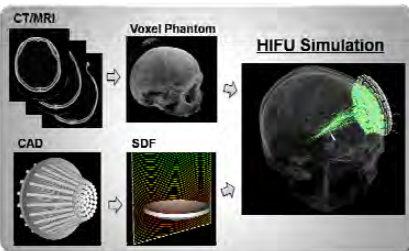
Simulation assisted focusing methods

- A layered wavevector-frequency domain model using input from CT scans (Clement & Hynynen, 2002)
- A time-reversal process based on prior CT scans using the wave propagation simulation (Aubry et al., 2003)
 - Digital human model is constructed based on CT images
 - Control parameters of array transducer is obtained by a simulation

Objectives

- Realization of the appropriate focus control in the body using array transducer
- Prediction of the treatment region for the preoperative planning
- Support of the development of HIFU device from design to approval

2. HIFU Simulator



The media of a volume model of human body (voxel phantom) is discretized into cubic elements with the voxel data for a living human body collected by CT/MRI technologies. On the other hand, the shape of the transducer defined by CAD is represented by the volume data of the signed distance function. Then, the ultrasound propagation from the transducer through the voxel phantom is performed by large-scale parallel computing.

Basic equations

The momentum conservation equation for mixture:

$$\rho \frac{\partial \mathbf{u}}{\partial t} = -\nabla p + \nabla \cdot \left[\zeta_{\text{eff}} \nabla (\epsilon) \mathbf{I} + 2\mu_{\text{eff}} \left(\epsilon - \frac{1}{3} \nabla \cdot (\epsilon) \mathbf{I} \right) \right] + \nabla \cdot \mathbf{a}'$$

The constitutive equation for bulk elasticity:

$$\frac{\partial p}{\partial t} = -\zeta_{\text{eff}} \nabla \cdot (\epsilon) - \left(\rho_{\text{eff}} \zeta_{\text{eff}} - \frac{2}{3} \mu_{\text{eff}} \right) \nabla \cdot \mathbf{u} \quad \text{where } \rho_{\text{eff}} = \sum_i f_i \rho_i \quad \text{and} \quad \frac{1}{\rho_{\text{eff}} \zeta_{\text{eff}}} = \sum_i \frac{f_i}{\rho_i \zeta_i}$$

The constitutive equation for shear elasticity:

$$\frac{\partial \zeta_{\text{eff}}}{\partial t} = 2\mu_{\text{eff}} \left(\epsilon - \frac{1}{3} \nabla \cdot (\epsilon) \mathbf{I} \right)$$

The equation of state of media (Tait's equation)

$$p = \rho \left(\frac{p + A}{\rho_0 + A} \right)^\gamma \quad \text{where } A = \frac{\rho_0 c_0^2}{\gamma} - p_0 \quad \text{and} \quad \gamma = \frac{B}{A} + 1$$

Non-linear coefficient

The heat equation for mixture:

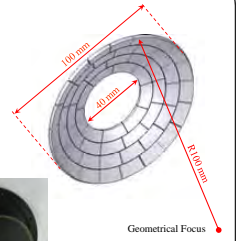
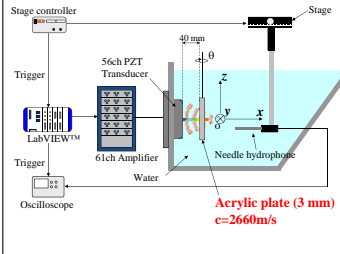
$$\rho_0 c_{\text{eff}} \frac{\partial T}{\partial t} = \frac{\partial}{\partial x_j} \left(\kappa_{\text{eff}} \frac{\partial T}{\partial x_j} \right) + \zeta_{\text{eff}} (\epsilon_{ij})^2 + 2\mu_{\text{eff}} \left(\epsilon_{ij} \epsilon_{ij} - \frac{1}{3} (\epsilon_{ii})^2 \right)$$

Numerical method

- Basic equations are discretized by the Finite Difference Method of 6-th order central difference in space and are explicitly integrated in time based on FDTD method.
- Perfectly Matched Layer is employed for the non-reflecting boundary.
- Parallel computing is performed by the domain decomposition with MPI + OpenMP.

3. Simulation Assisted Focus Control

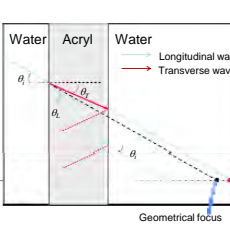
Experimental Apparatus



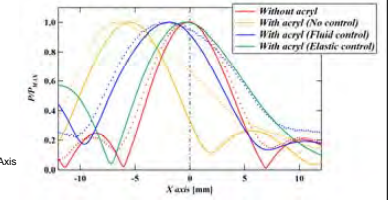
- 56 elements
- Frequency 2 MHz
- Focal distance: 100 mm

The acrylic plate as a phantom of bone is inserted between the transducer and target. The angle of the acrylic plate can be changed. The transducer is driven by 61ch amplifier under the control of LabVIEW. The acoustic field is measured by using needle hydrophone 3-dimensionally

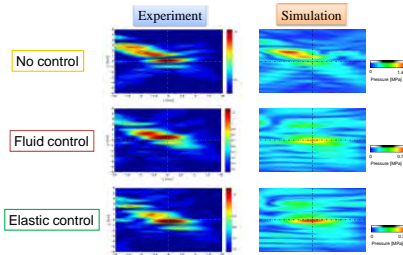
Influence of elasticity on focusing



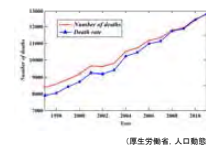
Pressure profiles (dot: experiment, line: simulation)



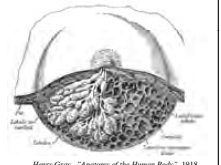
The results of experiment and simulation are in relatively good agreement.



HIFU for breast cancer



MRI 下で行った集束超音波治療
PDS: MRI-guided Focused Ultrasound Surgery
<http://iis.breastcancer.or.jp/>

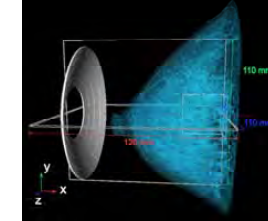


Henry Gray, "Anatomy of the Human Body", 1918

MRI image



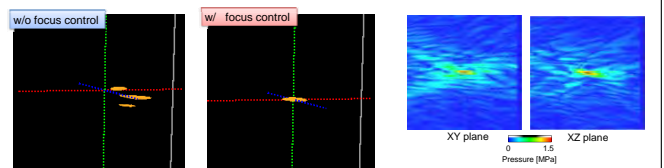
Numerical model



	Sound speed [m/s]	Density [kg/m ³]
Fat	1465 ^[1]	985 ^[1]
Parenchymal (乳腺)	1547 ^[2]	1032 ^[2]
Connective	1615 ^[2]	1090 ^[2]

[1] ICRU REPORT 61, [2] T. D. Mast, "Empirical relationships between acoustic parameters in human soft tissues", 2000

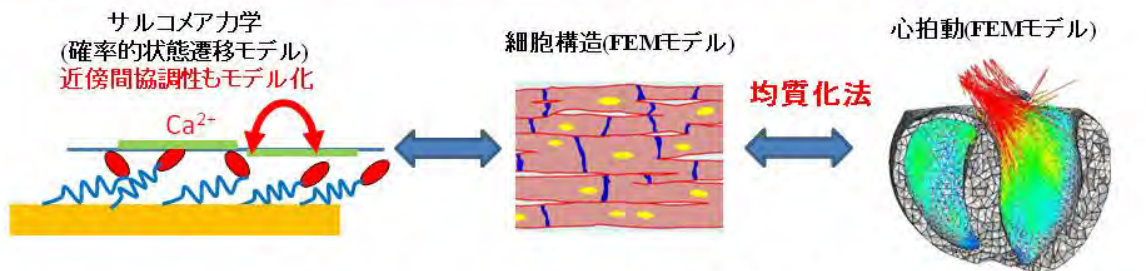
- 56 piezo elements
- Focal distance 100 mm
- Frequency 2 MHz
- 3D orthogonal mesh
- Size: 120x110x10 mm
- 1200x1100x100 grid
- 8x4x4 sub domains



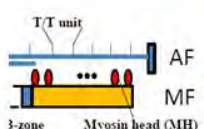
Focus is converged at the target by phase control.

マルチスケール解析手法

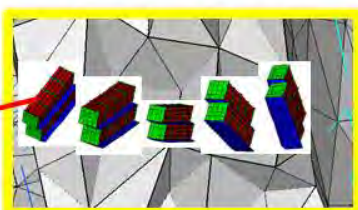
- ◆ サルコメアモデルとFEMモデルの時間刻み幅のギャップを克服
- ◆ 心筋の異方性構造を細胞構造モデルを適切に配置することで反映



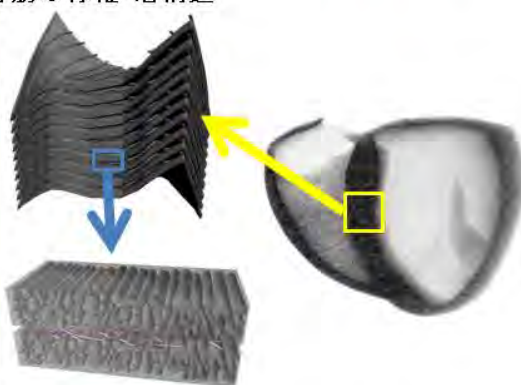
各筋原線維要素に複数の
フィラメントペアを埋め込む



線維・層構造に沿って各マクロ有限要素に
細胞構造モデルを配置

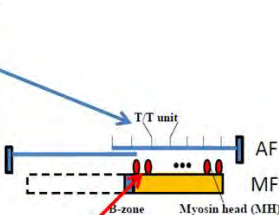
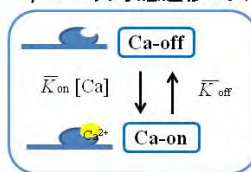


心筋の線維・層構造

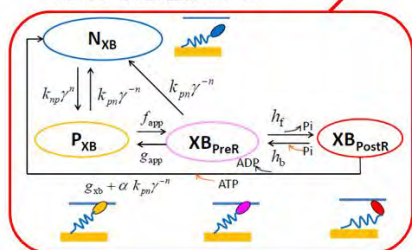


サルコメア力学モデル

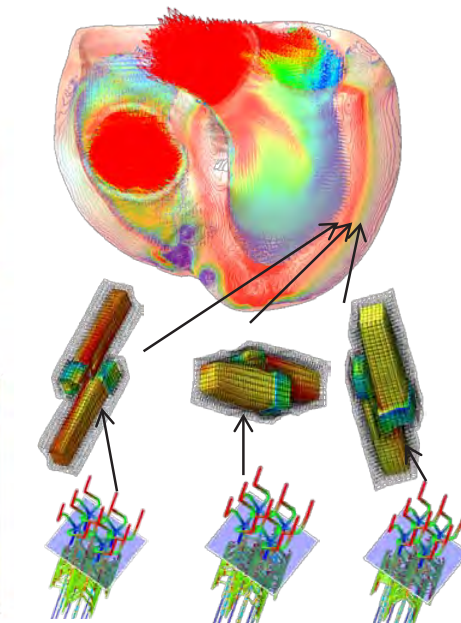
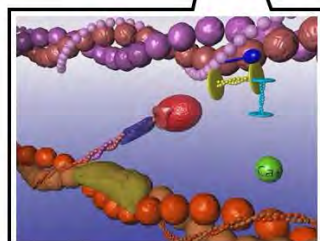
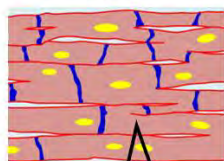
T/Tユニット状態遷移モデル



ミオシン状態遷移モデル



心筋細胞



京で実施したシミュレーション例
60万個の心筋要素に2.3万自由度の細胞モデル
を配置. 1.5拍動を60万コアを使って17時間で計算.

現在取り組んでいる課題: 肥大型心筋症のメカニズム解明



O-7

血管内の血流シミュレーションソフトウェア開発・検証研究



李 慶恩 (理化学研究所臓器全身スケール研究開発チーム)、石峯 康浩 (理化学研究所臓器全身スケール研究開発チーム)、野田 茂穂 (理化学研究所ACCC)、高木 周 (理化学研究所臓器全身スケール研究開発チーム、東京大学機械工学)

Introductions

• Many computational simulations of blood flow have been done by using a body-fitted mesh and a fixed mesh to understand correlation between hemodynamics and the development of vascular diseases.

Aim

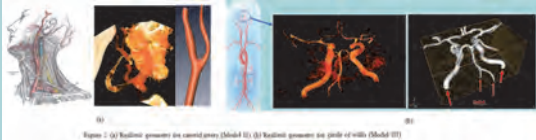
- To introduce a supporting tool for hemodynamics analysis on voxel system of Eulerian frame to reduce costs (especially man-based time) and to produce user-friendly tools for non-experts (i.e. medical doctors).
- Long-term purpose is to study the blood flow in a very complicated vessel, a compliant vessel and a deformable organ without the body-fitted mesh and a re-mesh process.

Numerical Models

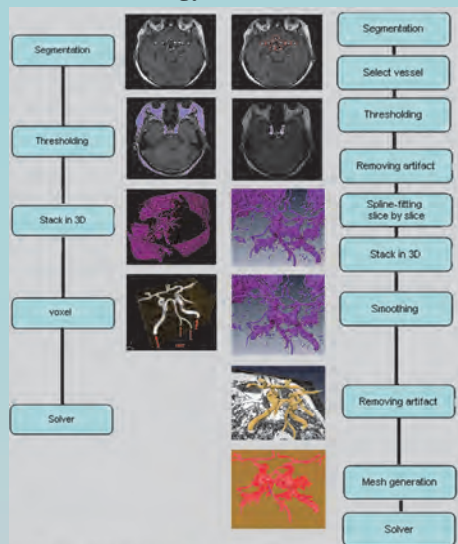
- Simplified s-shaped geometry for femoral artery (Model I)



- Realistic geometry for carotid artery (Model II & III)



Numerical Methodology



(a) fixed mesh Eulerian system, (b) body-fitted mesh Lagrangian system
Figure 3. comparison of procedures to reconstruct geometry from medical images for Eulerian system and Lagrangian system.

- Volume integral over a control volumes in Navier-Stokes equations is calculated on staggered grids in several spatial finite difference discretizations and time discretization with FAVOR (fraction area/volume obstacle representation) method and HSMAC (Highly Simplified MAC) Method.
- FAVOR method on fixed grid and voxel size play a role in the definition of contour of vascular surface as well as results in computational simulations in Figure 4.
- Boundary conditions (Parabolic steady inlet flow ($Re_{mean} = 500$), no-slip wall and free velocity gradient outlet condition)

Results

- The comparisons of velocity of combination method of both FVM and FDM in this study (Method I) and spectral/hp element method (Method II)[1].

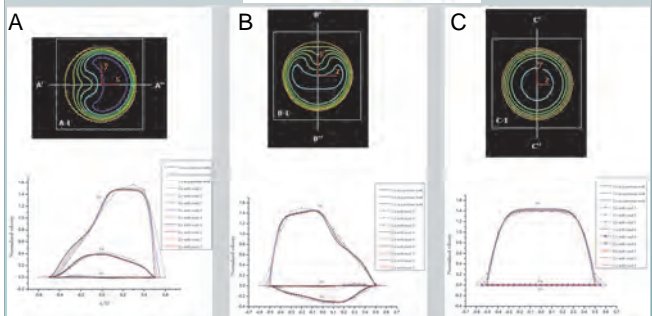
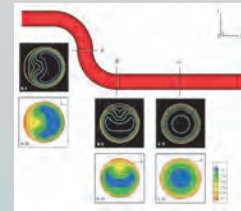


Figure 4. Comparisons of velocity of flow at A, B and C sections in the double bend by modified FDM in the present study (panel I) and by spectral/hp element method (panel II) [1].

Spatial Convergence test with the double bend model

Method I. Estimation of accuracy order by substituting the equations as below.

$$f_{h_{m+1}} = f_{error} - C_h(h_m)^p + H.O.T = f_{h_{m-1}} + C_h(h_m)^p - H.O.T \text{ at } N = (i, j, k)$$

at the size of voxel m ($h_m = \frac{2}{3} \times \frac{1}{2^{m+1}}$, at $m = 1, 2, \dots, 5$)

Voxel 1	Voxel 2	Voxel 3	Voxel 4	Voxel 5
		0.88	0.87	0.84
Voxel 2	0.88		0.84	1.0
Voxel 3	0.87	0.84		1.0
Voxel 4	0.84	1.0	1.0	

Method II. Grid Convergence Index (GCI) method is based on estimated fractional error.

	1,2&3	2,3&4	3,4&5
Error order, p	0.80	0.90	1.0

Method III. Estimation of accuracy order with an assumption that the results on the finest voxel is very close to exact solution.

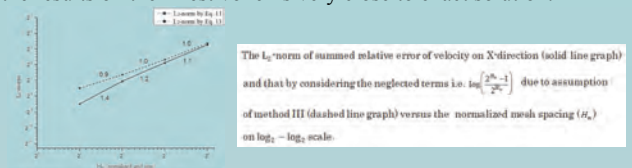


Figure 5. The L_2 -norm of summed relative error of velocity on X-direction versus the normalized mesh spacing (h_m) on the level of grid on $\log_2 - \log_2$ scale.

Conclusions

Our findings confirm the voxel based simulation software can be applied to blood flow simulations for biomedical analysis without a surface reconstruction procedure and a complicate mesh generation.

Reference

[1] K. E. Lee, K. H. Parker, C. G. Caro, S. J. Sherwin, The spectral/hp element modelling of steady flow in nonplanar double bends, International Journal for Numerical Methods in Fluids 57 (2008) 519-529

Acknowledgements

This research was supported by Research and Development of the Next-Generation Integrated Simulation of Living Matter, a part of the Development and Use of the Next-Generation Supercomputer Project of the Ministry of Education, Culture, Sports, Science and Technology.

筋骨格系3次元有限要素シミュレータの開発

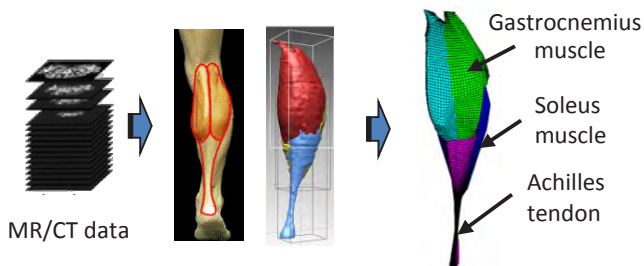
3D FE Simulator for Mechanical Analysis of Human Musculo-skeletal System

責任者: 山村直人 (理化学研究所 臓器全身スケール研究開発チーム)

We are developing a computational system that accounts for the combined behavior of cerebral nervous system and musculo-skeletal system, aiming to provide a useful tool for investigating the mechanisms underlying motor dysfunctions (e.g., Parkinson's disease). The musculo-skeletal system was modeled based on a finite element description of soft-tissues, in which the material behavior of the muscle is separated into an active and a passive part. In the description of muscle activation, the motor-unit activity model was introduced to take into account the effects of neural signals on the behavior of motoneurons. The simulation results by the developed model showed the inhomogeneous distributions of the activated muscle fibers and motor-unit forces.

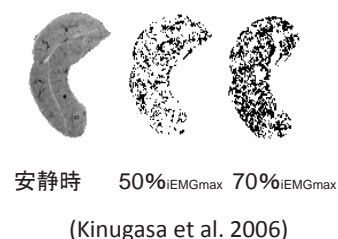
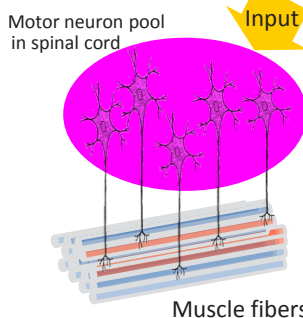
3D FE model of musculo-skeletal system

- 3D geometry and architecture of musculo-skeletal system reconstructed from CT/MR images
- Nonlinear FE method for incompressible hyperelastic materials
- Phenomenological muscle contraction model, Hill model and Motor neuron model



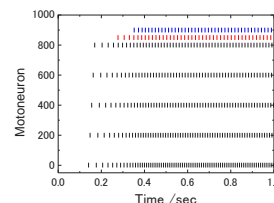
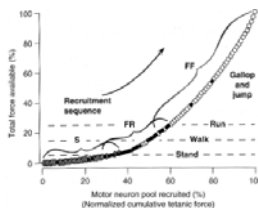
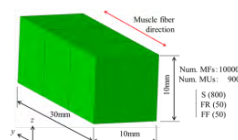
Micro-mechanical model of muscle contraction

Motor unit



Motor neuron model

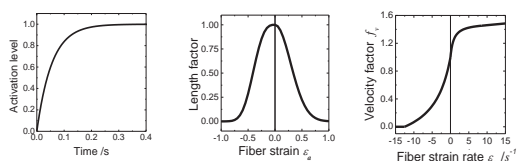
- Recruitment
- Rate coding



Muscle contraction model (Hill model)

$$\sigma_{fiber}^{act} = \sigma_0^{isom} \cdot \frac{A_0}{A} \cdot f_t \cdot f_l \cdot f_v$$

where $\left\{ \begin{array}{l} \sigma_0^{isom} : \text{Maximum isometric stress} \\ A_0, A : \text{Cross-sectional area of a muscle fiber before and after deformation} \end{array} \right.$



f_t : Activation f_l : Length factor f_v : Velocity factor

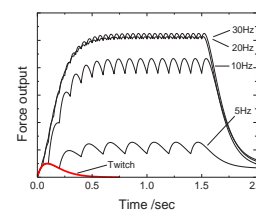
Impulse response for each discharge

$$f_{i,j}(t) = \frac{P_i t}{T_i} e^{-t/T_i} \cdot g_{i,j}$$

$\left\{ \begin{array}{l} T_i : \text{Twitch contraction time} \\ P_i : \text{Maximum twitch force} \\ g_{i,j} : \text{Discharge frequency dependence factor} \end{array} \right.$

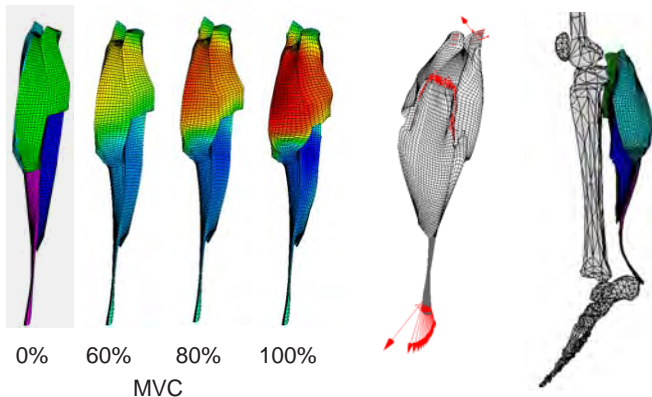
Response of continuous discharges

$$F_i^{motor}(t) = \sum_{j=1}^k f_{i,j}(t) \quad (t - t_{i,j} \geq 0)$$



Simulation results of a triceps surae muscle

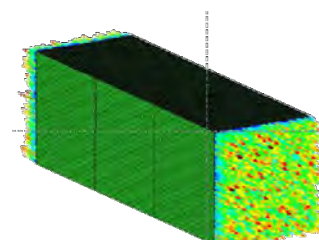
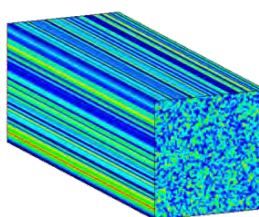
Displacement distributions Nodal forces Joint movement



Simulation results of isometric contraction

Cauchy stress distributions

Nodal forces





0-9:血管網シミュレーター

全身循環モデルを用いた頸動脈ステント留置術に関する血流シミュレーション

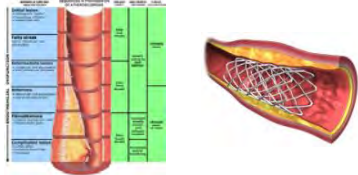
開発責任者: Fuyou Liang 高木周
 共同研究: 大島まり(東京大学情報学環) 藤澤慶(東京大学大学院学際情報学府学際情報学専攻)
 所属: 理化学研究所 次世代計算科学研究開発プログラム 臓器全身スケール研究開発チーム



研究目的 全身の血液循環を考慮した血流数値解析

□動脈硬化

動脈壁が肥厚・硬化し、血流を遮断。
 →酸素や栄養が重要組織に到達できなくなる
 (脳梗塞や心筋梗塞の原因)



□ステント留置術

術後に過灌流症候群の危険性
 →痙攣や意識障害、半身麻痺、脳出血の発生

□過灌流症候群

過灌流症候群を発症した患者
 →術前と術後で、中大脳動脈における平均流速の増加

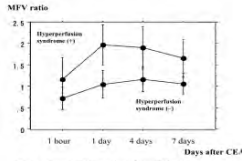


Fig. 2. Changes in MCA mean flow velocity (MFV) ratio in 14 patients with hyperperfusion syndrome. The MFV ratio ↑ from 1.0 (day 1, day 1) and 7 days after CEA were higher than those in the remaining 16 patients without hyperperfusion syndrome.

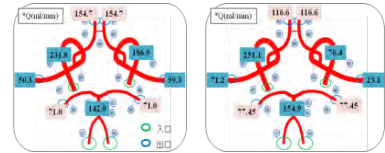
1) 藤本ら 経頭蓋カドプラーによる過灌流症候群の早期診断 独立行政法人国立病院機構九州医療センター 脳血管センター-脳卒中研究部脳血管内科

□最終目標

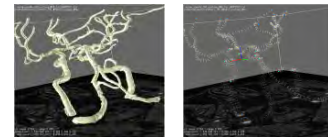
手術計画を目標とした術前術後の血行動態の予測のための全身循環を考慮した血流シミュレーションの開発

□患者データ

PC-MRIによる流量計測
 →末梢血管パラメータの推定

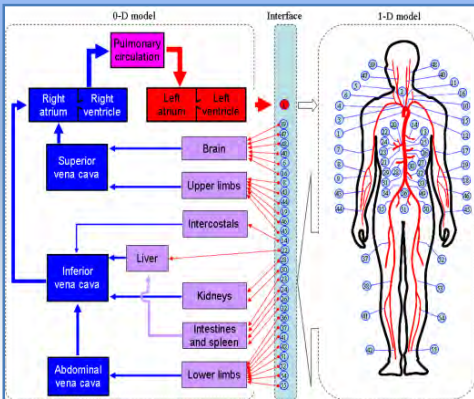


CTによる患者個人の形状データ
 →血管半径、血管長さの計測

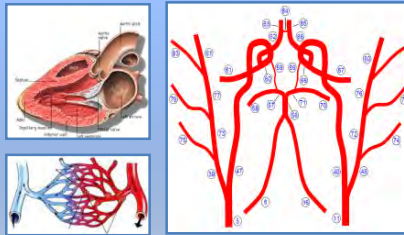


マルチスケール血流解析の概要

0次元解析



1次元解析



【1次元解析】
 ・頭頸部を含む全身83本の主要な動脈のモデル化
 ・軸方向のみの流れ解析
 ・血管壁の変形による脈波伝播を考慮

【0次元解析】
 ・毛細血管、静脈、心臓をモデル化
 ・Lumped parameter modelによる物理現象の再現

【1次元解析】

・連続の式

$$\frac{\partial A}{\partial t} + \frac{\partial q}{\partial x} = 0$$

・ナビエ-ストークス式

$$\frac{\partial q}{\partial t} + \frac{\partial}{\partial x} \left(\frac{q^2}{A} \right) + \frac{A}{q} \frac{\partial p}{\partial x} = -8\pi\nu \frac{q}{A}$$

・圧力の状態方程式

$$p(x,t) - p_0 = \frac{4}{3} \frac{Eh}{r_0} \left(1 - \sqrt{\frac{A_0}{A}} \right)$$

【0次元解析】

$$C \frac{dP_i}{dt} + Q_{i+1} - Q_i = 0$$

$$L \frac{dQ_{i+1}}{dt} + P_{i+1} - P_i = RQ_{i+1}$$

解析結果

【解析対象】

左中大脳動脈

【解析条件 術前】

手術前の患者の血管形状/末梢データ
 総頸動脈に狭窄あり
 平均流速: 15.35 cm/s

【解析条件 術後】

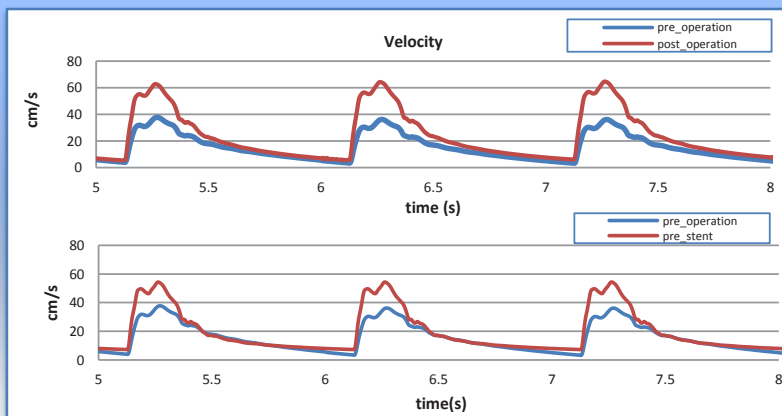
手術後の患者の血管形状/末梢データ
 総頸動脈に狭窄なし
 平均流速: 24.19 cm/s

【解析条件 術後予測】

手術前の患者の血管形状データ/末梢データ
 総頸動脈に狭窄なし
 平均流速: 20.24 cm/s

【結果】

左中大脳動脈において、平均流速の増加が確認され、過灌流症候群を発症した計測例と同様の傾向見られた。



三澤計治¹, 長谷川亜樹¹, 角田達彦^{1,2}(理化学研究所 次世代計算科学研究開発プログラム¹, 理化学研究所 ゲノム医科学研究センター²)

研究目的

- 近年のDNA配列決定技術の進歩により、大量のゲノム配列が解析されるようになりました。ゲノム配列は、個人の間には違いがあります。これを利用し、数万人に対して、ゲノム全体に渡り、患者群とコントロール群との遺伝子頻度を統計的に解析することによって、疾患関連遺伝子を発見する手法がゲノムワイド関連解析(GWAS)です。
- GWASでは多重検定によるtype I errorが問題になります。多重検定の問題に対しては、Bonferroni correctionが広く行われていますが、SNP間でlinkage disequilibriumがあるため、疾患関連遺伝子を見落とす恐れが出てきます。
- 我々は、多項分布を用いて、SNP間にlinkがある場合でも正確にtype I errorを計算する式を導きました。それに基づき、genome wideのP値を適切なthresholdを設定する方法を開発しました(Misawa et al. 2008)。また、高速計算のため、マルコフ連鎖モンテカルロ法(MCMC法)でtype I errorを計算する方法を開発しました。これらの方法の他にpermutationに基づいたRAT法も開発されています(Kimmel and Shamir 2006)。これらの方法は時間がかかるため、ゲノムワイドの解析を行う際には、高速に計算する必要があります。

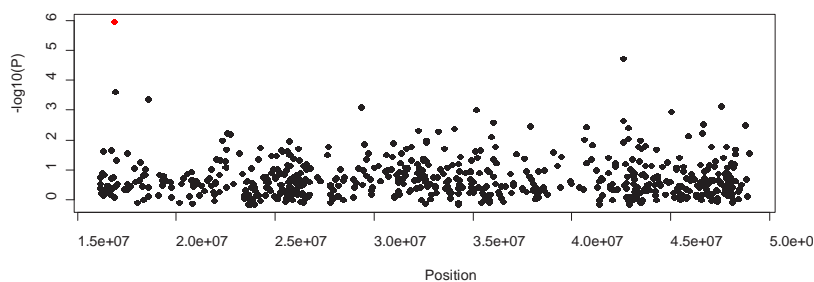
ParaHaploの開発

- 我々は京速コンピュータ「京」(右図)の上で、ゲノムワイド関連解析のためのプログラムParaHaploを開発しました。「京」は8万ノード・64万コアの計算ユニットからなる非常に速いスーパーコンピュータです。
- ParaHaploでは、ゲノムワイドのhaplotype多型データを分割し、データを分散入力させた上で、ユニット間・コア間のハイブリッド並列により、Misawa et al. (2008)の方法とRAT法の両方に関して、高速計算を実現しました (Misawa and Kamatani 2009, 2010, 2011)。
- ParaHaploのソースコードは以下のサイトで公開されています。
<http://sourceforge.jp/projects/parallelgwas/releases/51104>



実際のデータをParaHaploで解析した例

- Wellcome Trust Case-Control Consortium (WTCCC)から、2型糖尿病の患者1,924人とコントロール2,938人の遺伝子型データの提供を受けました。右図は22番染色体に関してparaHaploを用いたハプロタイプ関連解析を行った結果です。赤い点がparaHaploによって初めて2型糖尿病と関連があると推測された領域になります。



謝辞

- 京での計算に関しては京速コンピュータ「京」の試験利用での結果です。
- また、PCクラスタでの性能計測に関しては理化学研究所情報基盤センターのRICCを使用しています。

参考文献

- Kimmel, G. and Shamir, R. (2006) A fast method for computing high-significance disease association in large population-based studies. *Am J Hum Genet*, **79**: 481-92.
- Misawa K, Fujii S, Yamazaki T, Takahashi A, Takasaki J, Yanagisawa M, Ohnishi Y, Nakamura Y, Kamatani N (2008) New correction algorithms for multiple comparisons in case-control multilocus association studies based on haplotypes and diplotype configurations. *J Hum Genet* **53**:789-801
- Misawa K, Kamatani N (2009) ParaHaplo: A program package for haplotype-based whole-genome association study using parallel computing. *Source Code Biol Med* **4**:7
- Misawa K, Kamatani N (2010) ParaHaplo 2.0: a program package for haplotype-estimation and haplotype-based whole-genome association study using parallel computing. *Source Code Biol Med* **5**:5
- Misawa K, Kamatani N (2011) ParaHaplo 3.0: A program package for imputation and a haplotype-based whole-genome association study using hybrid parallel computing. *Source Code Biol Med* **6**:10



D-2: NGS analyzer

次世代シーケンサーの出力データを高速に解析し、個人間の遺伝的差異やがんゲノムの突然変異を高い正確さで同定する

開発責任者: 角田達彦 (理化学研究所ゲノム医科学研究センター)
開発担当者: 藤本明洋 (理化学研究所ゲノム医科学研究センター)



● 概要

次世代シーケンサーの出力データを高速に解析し、ヒト個人間の遺伝的差異やがんゲノムの突然変異を高い正確さで同定する。

● 離散化(計算モデル化)の方法

ヒト標準ゲノム配列に対するマッピングと確率計算に基づいた多様性検出
計算方法直接法による密行列の対角化並列化の方法領域分割

● 開発言語とライブラリ

C

● 現状での計算規模

東京大学医科学研究所ヒトゲノム解析センター Shirokane2 2000コア
「京」250ノード

● 実際のデータを用いNGS analyzerを動かした例

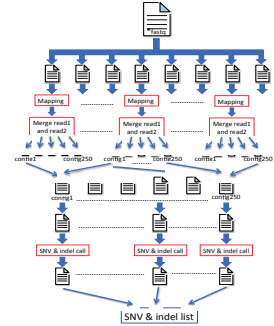
(1) 初の日本人ゲノムシーケンス解析 (Nat Genet 2010 42: 931-6)

- 正確な多型検出法を開発した。
- 約300万の一塩基多様性を検出した。
- 他の個体との比較により、アミノ酸多型にはレアなものが多く、負の自然選択の影響を受けたことが示唆された。
- 中間サイズの欠失 (300bp-10kbp) を検出する手法を開発した。
- 標準ゲノム配列に存在しない新規配列を約3.4Mbp検出した。

(2) 肝がんゲノム解析 (Nature 2010 464, 993-8, Nat Genet 2012 44:760-4)

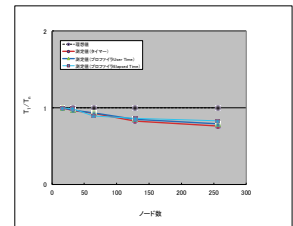
- がんの体細胞突然変異を高精度に検出する手法を開発した。
- がんの体細胞突然変異のパターンを解析した。
- がん細胞に感染しているウイルスのゲノムを検出した。
- がん細胞の遺伝的多様性を解析した。
- 肝がんでは、クロマチン制御遺伝子が高頻度に変異していた。

データの分割と解析

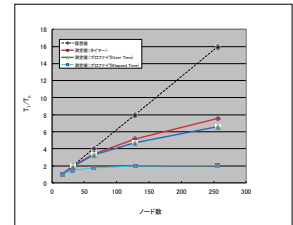


「京」における並列化効率

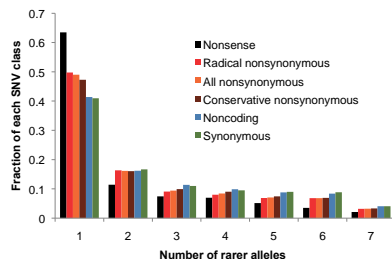
ウィークスケーリング



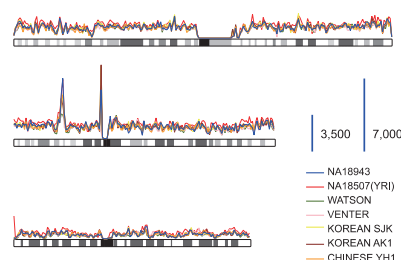
ストロングスケーリング



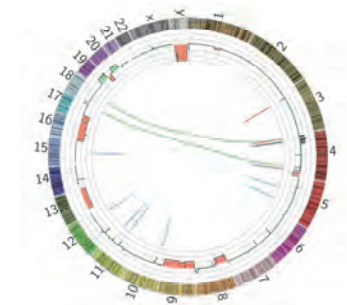
SNPの頻度スペクトラム



ヒトの遺伝的多様性 (1Mbp当たりのSNPの密度)



がんの突然変異 (染色体異常とコピー数異常)



参考文献

- * Fujimoto A, Nakagawa H, Hosono N, Nakano K, Abe T, Boroevich KA, Nagasaki M, Yamaguchi R, Shibuya T, Kubo M, Miyano S, Nakamura Y, Tsunoda T (2010) Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing. *Nat Genet* 42: 931-6
- The International Cancer Genome Consortium (2010) International network of cancer genome projects. *Nature* 464, 993-8.
- Fujimoto A, Totoki Y (equally contributed) et al. (2012) Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nature Genetics* 44: 760-764



D-3: ExRAT

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

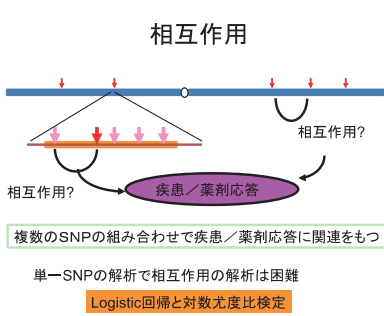


開発責任者: 角田達彦 (理化学研究所ゲノム医科学研究センター)

概要:

遺伝子間相互作用が発症リスクを変化させる疾患関連遺伝子の組合せを全ゲノムで探索する。2SNP間の全組合せを超並列に行う方法と、SNP間の連鎖不平衡(相関)も考慮した、より精密な方法の2種類を実装。前者で全組合せをスクリーニングし、後者で経験的p値を求める手順を想定。

遺伝子どうしが相互作用を起こして疾患への発症リスクを上昇する現象とそれらの遺伝子を新たに疾患関連遺伝子として発見することができる



パーミュテーションテスト

ゲノムワイド関連解析の基盤構築 → 解析対象SNPは50万
ボンフェローニ補正は保守的すぎる

Permutation Test

統計量に関する事前の仮定を必要とせず現実的な補正
しかし通常のランダムパーミュテーションテスト(SPT)は極めて計算負荷が高い

高速化 RAT (rapid association test)

Kimmel and Shamir (2006)

Importance Samplingによる近似

分割表作成 + χ^2 検定 → 尤度比検定に拡張

- ExRAT ver. 1.0- r440, phase III-0, 100%
- 富士通コンパイラ(RICC, FX1, 「京」, の全て)対応済, 実行成功, buildも可能.
- PCクラスターで8,192ノードの並列性能を実証.
- データ: 患者708人, 一般集団3397人, 8314SNP, 14100回 permutation で評価.
- ハイブリッド並列化対応: Preprocess処理とpermutation処理にOpenMPでのスレッド化の実装.
- Preprocess処理の見直しによるさらなる高速化: 分割表探索処理での計算処理の見直しにより, 約40%程度高速化(計算処理を探索中に積み上げていけるように処理を変更し, 全体の計算量の削減に成功した).
- MPIプロセスの増加によるメモリ使用量増加への対応: 非同期通信のためのバッファ確保が原因だったため, 通信処理を見直し, 処理完了待ちのバッファの大きさを最小とすることで対処した.
- テスト運用の結果から問題点の洗い出しおよび改善方法の検討を行ってきた.
- WTCCCの実データを解析し, 評価と結果を出した.

目的と方法のまとめ

- 目的: 2つの遺伝子(SNP)が同時に作用して疾患の原因となるものを探索する.
- データ: 全染色体上の50万SNP × 数千人
→ 500000 × 500000 / 2 = 10¹¹組合せ

解析方法1: 単純組合せのロジスティック解析

(all_comb)

解析方法2: Permutation test に基づく経験分布から計算し, より厳密なp値を出す

(RAT)

2つのプログラムの適用方法

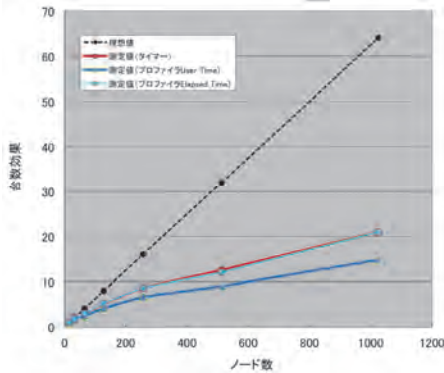
ステップ1: 単純組合せロジスティック回帰 all_comb

- 全体の計算量は比較的軽い
- 概念がわかりやすいので, 多くの人が使うはず
- データ並列に適している
- 最初のスクリーニングに使う



ステップ2: permutation test による RAT

- ゲノム上の性質(連鎖不平衡)も考慮した計算
- 計算量が重いので, スクリーニング後の精査へ



「京」上のall_comb

Importance Sampling

パーミュテーションによって得られる統計量の分布



赤い部分のみをサンプリングできる分布を事前に定義

赤に含まれるある1個のパーミュテーションの生起確率

サンプリング数を十分にとって足し合わせる → 赤の比率

RAT法の拡張

より複雑な解析に対応するために

各種線形回帰モデルと尤度比検定への拡張

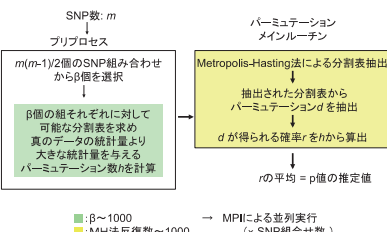
2SNPの相互作用付きロジスティック回帰モデル

ケースとなる確率P

$$\logit(P) = \mu + \alpha X_A + \beta X_B + \gamma X_A X_B$$

↑ SNP A の効果 ↑ SNP B の効果 ↑ SNP間の相互作用

ExRAT: 並列化

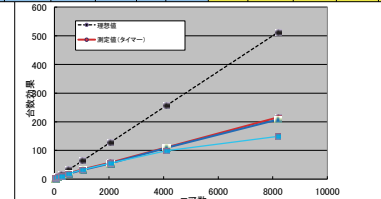


■: β~1000 ■: MH法反復数~1000 → MPIによる並列実行 (x SNP組合せ数)

Dataset	WTCCC type 2 diabetes
Sample size	1,924 cases + 2,938 controls
#SNPs (MAF>0.05, after QC)	360,297
#combinations	6.49x10 ¹⁰
#runs	25
Cores used / run	24,576
Disk space (stage in) / run	1.5TB
Memory used / run	123GB
Time / run (average)	41 min. 17 sec.
Total time	~17h. (24,576 cores used)

MPIによる並列化

プログラムの種類	プログラムの名前	プログラムのバージョン	プログラムの実行時間 (秒)	プログラムの最大メモリ使用量 (MB)	プログラムの最大コア数	プログラムの最大計算数	プログラムの最大メモリ使用率 (%)
all_comb	all_comb	1.0	22451.76	22451.76	1	22451.76	2.1%
all_comb	all_comb	1.0	1193.58	1193.58	1	1193.58	0.2%
all_comb	all_comb	1.0	6376.80	6376.80	1	6376.80	0.3%
all_comb	all_comb	1.0	2045.18	2045.18	1	2045.18	0.2%
all_comb	all_comb	1.0	1137.59	1137.59	1	1137.59	0.1%
all_comb	all_comb	1.0	1868.14	1868.14	1	1868.14	0.2%
all_comb	all_comb	1.0	381.18	381.18	1	381.18	0.0%
all_comb	all_comb	1.0	4025.11	4025.11	1	4025.11	0.4%
all_comb	all_comb	1.0	193.25	193.25	1	193.25	0.0%
all_comb	all_comb	1.0	10381.92	10381.92	1	10381.92	1.0%



「京」上importance sampling



D-4: SiGN-BN

ベイジアンネットワークによる大規模遺伝子ネットワーク推定プログラム



開発責任者: 宮野 悟 (東京大学 医科学研究所 ヒトゲノム解析センター)

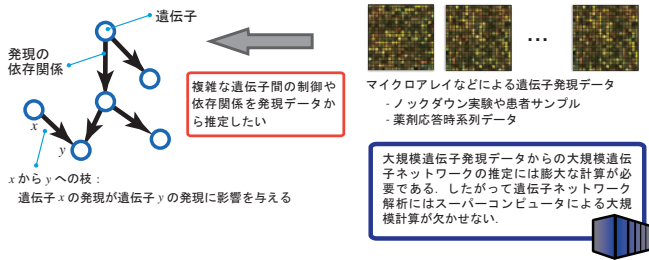
SiGN 開発メンバー: 玉田 嘉紀, 島村 徹平, 山口 類, 新井田 厚司, 斉藤 あゆむ, 長崎 正朗, 井元 清哉, 宮野 悟

概要

SiGN (サイン) は遺伝子発現データからスーパーコンピュータを用いて遺伝子ネットワークを予測・推定するソフトウェア群である。遺伝子ネットワーク解析により、観測データから薬剤標的遺伝子の同定、副作用の予測・回避、創薬ターゲット・毒性関与とパスウェイ探索などが容易になることが期待される。**SiGN** では様々な実験データの解析に対応するためベイジアンネットワーク、状態空間モデル、グラフィカルガイシアンモデル、ベクトル自己回帰モデルが遺伝子ネットワークモデルとして利用可能である。**SiGN-BN** はこれらのうちベイジアンネットワークを用いた遺伝子ネットワーク推定ソフトウェアで、スーパーコンピュータを用いた並列計算によりヒト全ゲノムを網羅した数万遺伝子の遺伝子ネットワークを短時間で推定可能である。

Visit SiGN Web Site at <http://sign.hgc.jp>

遺伝子ネットワーク = 遺伝子発現の依存関係をネットワークで表現したモデル



ノンパラメトリック回帰によるベイジアンネットワーク

Bayesian Network Model Imoto et al. (2002)

Representation of the joint probability distribution by the DAG (Directed Acyclic Graph) structure

Gene Expression = Random Variable
Node = Gene
Directed Edge = Dependency

$$\pi(G|X) \propto \pi(G) \prod_{i=1}^n \prod_{j \in \text{pa}(i)} f(x_{ij} | p_{ij}, \theta_j) \pi(\theta_j | \lambda) d\theta_j$$

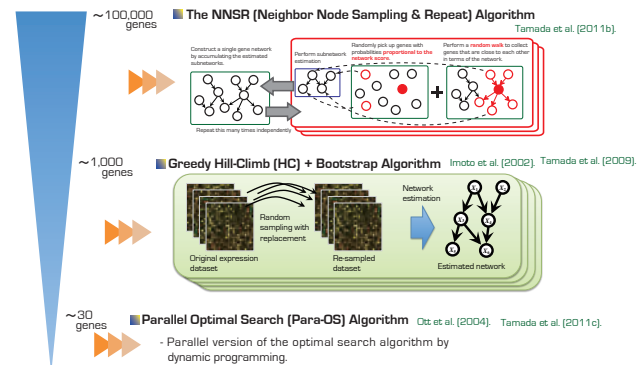
G: gene network X: gene expression data

The parent-child relationship is modeled by the B-spline nonparametric regression

$$x_{ij} = m_{ij}(p_{ij}^{(1)}) + \dots + m_{ij}(p_{ij}^{(k)}) + \varepsilon_{ij}$$

$$m_{ij}(p_{ij}^{(k)}) = \sum_{r=1}^k \gamma_{r,ij} b_r^{(k)}(p_{ij}^{(k)})$$

Parallel Algorithms for Estimating Gene Network Structures



Reference

Affara, M. et al. (2007). Understanding endothelial cell apoptosis: What can the transcriptome glycome and proteome reveal? *Philosophical Transactions of Royal Society B* 362 (1484), 1469-1487.

Araki, H. et al. (2009). Analysis of PPARα-dependent and PPARα-independent transcription regulation following fenofibrate treatment of human endothelial cells. *Angiogenesis* 12 (3), 221-229.

Imoto, S. et al. (2002). Estimation of genetic networks and functional structures between genes by using Bayesian network and nonparametric regression. *Pac. Symp. Biocomput.* 7, 175-186.

Imoto, S. et al. (2006). Computational strategy for discovering druggable gene networks from genome-wide DNA expression profiles. *Pacific Symposium on Biocomputing (PSB 2006)*, 11, 559-571.

Ott, S. et al. (2004). Finding optimal models for small gene

ソフトウェア仕様

- 概要・特徴**
 - ノンパラメトリック回帰によるベイジアンネットワークを用いて遺伝子発現データから遺伝子ネットワークを推定するソフトウェア
 - 「京」上にて超並列で動作し高速に遺伝子ネットワーク (有向グラフ) を計算可能
 - NNSR アルゴリズムを用いるとこれまで必要だった「遺伝子選択」をせずに、全ゲノムによる大規模遺伝子ネットワークの推定が可能
 - HC+Bootstrap 法を用いることにより高信頼ネットワークを計算可能
 - マイクロアレイだけでなく様々な事前知識をネットワーク推定に利用可能
- 出力**
 - 出力データ: 遺伝子ネットワーク (テキスト/CSML)
 - 推定された枝に信頼度等の情報付き
 - Cell Illustrator で解析可能
- 並列性能 (HC+Bootstrap)**
 - 最大動作コア数: 196,608 コア (24,576 ノード)
 - 並列化効率: 0.72 (196,608 vs 98,304 コア時) (ストロングスケーリング)
 - ※ 2012/12/14 現在
 - 「京」は現在開発中であり、これらの数字は整備中のシステムによる暫定的な値である
- 「京」以外でのターゲット環境**
 - 東京大学 医科学研究所 ヒトゲノム解析センター (HGC) スーパーコンピュータシステム (Xeon/Opteron PC クラス)
- ソフトウェア配布形態**
 - 「京」ユーザに実行バイナリを配布
 - HGC ではシステムにインストール済み

これまでの応用例

肺癌細胞株 PC-9 + Gefitinib 投与による薬剤応答パスウェイ解析

■ 使用データ

- 細胞: 肺癌細胞株 PC-9
- 薬: Gefitinib (イレッサ: 抗がん剤)
- マイクロアレイ: Agilent 44K Whole Human Genome Oligo Microarray
- 計測データ: 26 時点 × 1 回計測 = 26 アレイデータ
- EGF, Gefitinib 投与・非投与の 4 条件

■ 目的

- Gefitinib を投与した肺癌細胞株 PC-9 での遺伝子ネットワークの動的変化を見ることで、未知作用機序の同定

■ 利用手法

- 動的ベイジアンネットワーク (HC) による Node-Set Separation 法

■ 結果

- Node-Set Separation 法により時点毎のネットワークが得られ時間による変化が解析可能

ヒト臍帯静脈内皮細胞 (HUVEC) によるアポトーシス関与遺伝子同定 Affara et al. (2007)

■ 使用データ

- 細胞: ヒト臍帯静脈内皮細胞 (HUVEC)
- 刺激: 血清除去 & siRNA ノックダウン
- マイクロアレイ: CodeLink UniSet Human 20K gene chip
- 計測データ: 8 時点 × 3 回計測 時系列データ
- siRNA ノックダウンアレイ
- ※ 2 アレイデータ

■ 目的

- アポトーシスを誘導した HUVEC を時系列に観測しネットワーク解析をすることによりアポトーシス制御に関わる関連遺伝子の同定

■ 利用手法

- HC+Bootstrap による動的ベイジアンネットワーク

■ 結果

- ネットワーク推定および解析の結果 GABARAP がハブ遺伝子 (子の多いネットワーク中のノード) として抽出され実験的にも HUVEC においてアポトーシスへの関与を確認

- その他の事例**
- HUVEC/Fenofibrate を用いた高脂血症薬新規ターゲット遺伝子の同定 Imoto et al. (2006).
 - HUVEC/Fenofibrate による非脂質代謝系作用パスウェイの同定 Araki et al. (2009).
 - メラノーマ細胞を用いた予後予測因子・新規関連遺伝子の同定 Wang et al. (2012).

PR 文部科学省 新学術領域研究「システムがん」

SiGN is also used in the **Systems Cancer Project** that started newly in 2010. Visit our web site at <http://systemscancer.hgc.jp> and follow our tweets @SystemsCancer in Twitter for updates.

Acknowledgments

Computational time was provided by the Super Computer System, Human Genome Center, Institute of Medical Science, The University of Tokyo, RIKEN Integrated Cluster of Clusters (RICC) system, and early access to the K computer at the RIKEN Advanced Institute for Computational Science.

本発表の結果の一部は、理化学研究所が実施している京連コンピュータ「京」の試験利用によるものです。



D-5: SiGN-L1

L1正則化法による大規模遺伝子ネットワーク推定プログラム



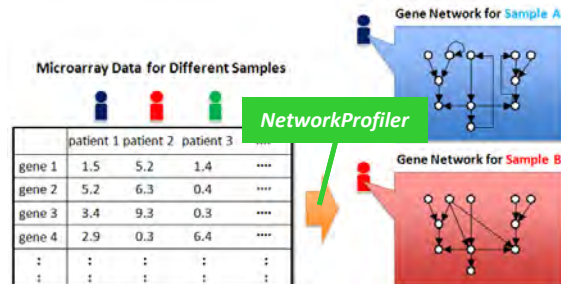
開発責任者: 宮野 悟 (東京大学 医科学研究所 ヒトゲノム解析センター)

SiGN 開発メンバー: 玉田 嘉紀、島村 徹平、山口 類、新井田 厚司、関谷 弥生、斉藤 あゆむ、井元 清哉、宮野 悟

概要

SiGN (サイン) は遺伝子ネットワーク発現データからスーパーコンピュータを用いて遺伝子ネットワークを予測・推定するソフトウェア群 (SiGN-BN, SiGN-L1, SiGN-SSM) である。このうち、SiGN-L1 は L1 正則化法によるスパースな統計的グラフィカルモデル (グラフィカルガウシアンモデル、ベクトル自己回帰モデル、構造方程式モデル) を用いた遺伝子ネットワーク推定ソフトウェアであり、スーパーコンピュータを用いた並列計算により、ヒト全遺伝子を対象とした大規模遺伝子ネットワークを推定することが可能である。推定された遺伝子ネットワークを解析することにより、観測データから薬剤標的遺伝子の同定、創薬ターゲット・毒性関与パスウェイの探索などの応用が期待される。

Case 4. 異なるサンプルに対する遺伝子発現データからサンプル各々のネットワークを推定



SiGN-L1の特徴

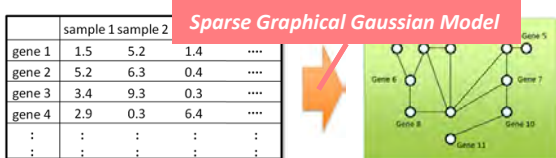
- 3 種類の統計的グラフィカルモデル (グラフィカルガウシアンモデル、ベクトル自己回帰モデル、構造方程式モデル) により、遺伝子発現データから遺伝子ネットワークモデルを構築
- L1 正則化法を用いてスパースなモデルのパラメータ空間を探索することにより、100,000 ノードまでの大規模遺伝子ネットワーク推定が可能
- 次世代スパコン「京」上で、最大並列数 (コア数): 98,304、最大ノード数: 12,288、実行効率: 1コアで11.98%を達成。なお、ハイブリッド並列化には対応済み。
- ベータ版のバイナリを公開開始 <http://sign.hgc.jp/sign1/index.html>

これまでの応用事例

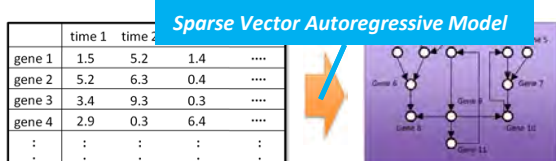
遺伝子ネットワーク = 遺伝子発現間の依存関係を表したグラフ構造

入力と出力

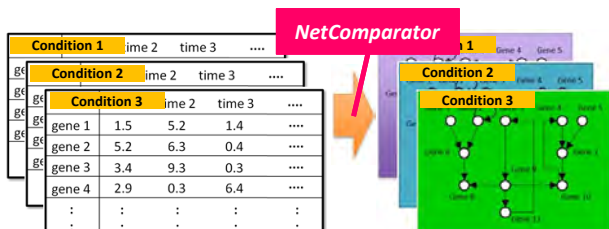
Case 1. Sample A に対する遺伝子破壊株発現データから遺伝子ネットワークを推定



Case 2. Sample A に対する時系列遺伝子発現データから遺伝子ネットワークを推定



Case 3. Sample A に対する異なる実験条件下の時系列遺伝子発現データから各々の実験条件下での遺伝子ネットワークを推定



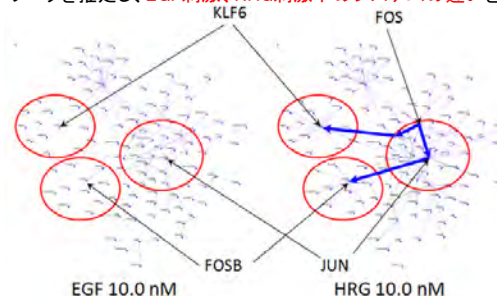
参考文献

- Shimamura et al. (2007) Weighted lasso in graphical Gaussian modeling for large gene network estimation based on microarray data. *Genome Informatics*, 19, 142-153.
- Shimamura et al. (2009) Recursive regularization for inferring gene networks from time-course gene expression profiles. *BMC Systems Biology*, 3, e41.
- Shimamura et al. (2010) Inferring dynamic gene networks under varying conditions for transcriptomic network comparison. *Bioinformatics*, 26(8), 1064-1072.
- Shimamura et al. (2010) Collocation-based sparse estimation for inferring continuous-time dynamic gene networks. *Genome Informatics*, 24, 164-178.
- Shimamura et al. (2011) A novel network profiling analysis reveals system changes in epithelial-mesenchymal transition. *PLoS ONE* 6(6): e20804, doi:10.1371/journal.pone.0020804, 2011.

二つの異なる成長因子による細胞内の情報伝達パスウェイの比較

データ: 8実験条件下において8時点で計測した時系列遺伝子発現データ (Nagashima et al., 2007)

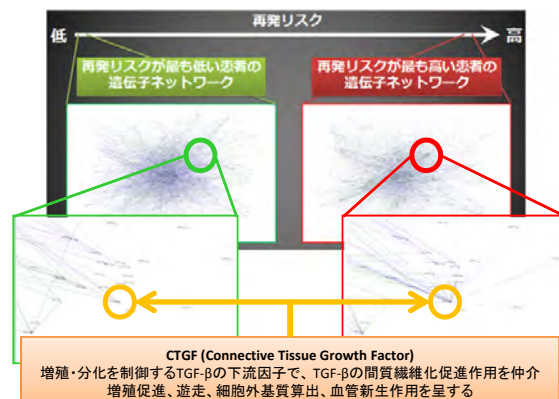
目的: EGF、HRG各々の成長因子によって誘導される 8 実験条件下での遺伝子ネットワークを推定し、EGF刺激、HRG刺激下のシステムの違いを比較



再発リスクに関わる日本人肺癌患者の遺伝子ネットワークの比較

データ: 226症例の肺癌がん患者の遺伝子発現データ (国立がんセンター横田先生、河野先生との共同研究)

目的: 再発リスクをモジュレーターとしたときの 226 症例各々の遺伝子ネットワークを推定し、再発リスクが高い患者と低い患者のシステムの違いを比較





D-6: SiGN-SSM

状態空間モデルによる大規模遺伝子ネットワーク推定プログラム

開発責任者: 宮野 悟 (東京大学 医科学研究所 ヒトゲノム解析センター)



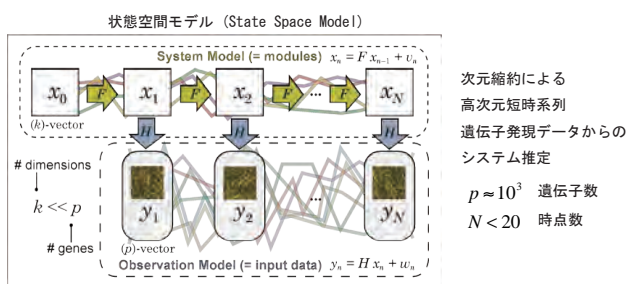
SiGN 開発メンバー: 玉田 嘉紀, 島村 徹平, 山口 類, 新井田 厚司, 斉藤 あゆむ, 長崎 正朗, 井元 清哉, 宮野 悟

概要

SiGN (サイン) は遺伝子発現データからスーパーコンピュータを用いて遺伝子ネットワークを予測・推定するソフトウェア群である。遺伝子ネットワーク解析により、観測データから薬剤標的遺伝子の同定、副作用の予測・回避、創薬ターゲット・毒性関与とパスウェイ探索などが容易になることが期待される。**SiGN** では様々な実験データの解析に対応するためベイジアンネットワーク、状態空間モデル、グラフィカルガウシアンモデル、ベクトル自己回帰モデルが遺伝子ネットワークモデルとして利用可能である。**SiGN-SSM** はこれらのうち状態空間モデルを用いたネットワーク推定ソフトウェアで、スーパーコンピュータを用いた並列計算により、時系列遺伝子発現データから数千遺伝子のモジュールおよび遺伝子ネットワークの推定、および予測モデルの構築が可能である。

Visit SiGN Web Site at <http://sign.hgc.jp>

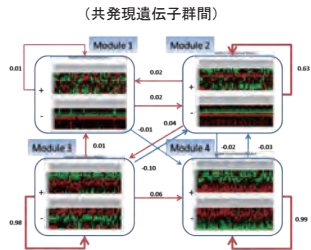
状態空間モデルによる時系列遺伝子発現データからの情報抽出



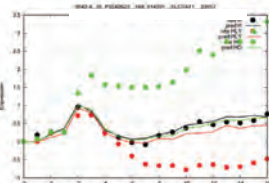
システム構築パラメータ推定

$$\theta = \{H, F, R, x_0\}$$

モジュールネットワーク (共発現遺伝子群間)



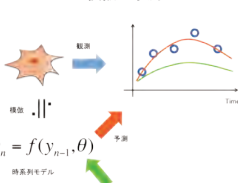
動的発現予測モデル (シミュレーション)



遺伝子ネットワーク (個別遺伝子間)



システムの模倣と予測



Reference

Yamaguchi and Higuchi. State-space approach with the maximum likelihood principle to identify the system-generating time course gene expression data of yeast. *Int J of Data Mining and Bioinformatics*, 1(1): 77-87, (2006).
 Yamaguchi et al. Finding module-based gene networks with state-space models - Mining high-dimensional and short time-course gene expression data. *IEEE Signal Processing Magazine*, 24(1): 37-46 (2007).
 Hirose et al. Statistical inference of transcriptional module-based gene networks from time course gene expression profiles by using state space models. *Bioinformatics*, 24: 932-942, (2008).
 Yamaguchi et al. Predicting differences in gene regulatory systems by state space models. *Genome Informatics*, 21: 101-113, (2008).

ソフトウェア仕様

- 概要・特徴**
 - 状態空間モデルにより時系列遺伝子発現データから、モジュールネットワーク、遺伝子ネットワークおよび動的発現予測モデルを構築
 - 次元縮約および一意性制約により高次元時系列データから動的システムからの推定可能
 - 並べ替え検定によるネットワーク推定
 - 繰り返し計測を持つ時系列、欠測および不均衡計測を含む時系列の情報も適切に扱うことが可能
 - 動的予測に基づくケース・コントロール時系列の比較 (制御構造に差異のある遺伝子群の予測)
 - データ解析統合プラットフォーム SBIP との連携
- 出力**
 - 出力データ: システム構造パラメータ、状態推定値 (テキスト) (遺伝子ネットワーク、モジュールネットワーク、予測モデル構築に使用可)
- 並列性能**
 - 最大動作コア数: 196,608 コア [24,576 ノード]
 - 並列化効率: 0.92 (196,608 vs 98,304 コア時) (ストロングスケーリング)
- 「京」以外でのターゲット環境**
 - 東京大学 医科学研究所 ヒトゲノム解析センター (HGC) スーパーコンピュータシステム (Xeon/Opteron PC クラスタ)
- ソフトウェア配布形態**
 - オープンソースとして公開
 - GNU Affero General Public License
 - <http://sign.hgc.jp/signssm/index.html>
- 入力**
 - 入力データ: マイクロアレイなどの遺伝子発現時系列データ
 - 例: 薬剤応答時系列データ
 - サンプル (アレイ) 数: 10 ~ 30 アレイ
 - 時系列データは複数回の繰り返し計測を推奨
 - ノード (遺伝子) 数: 100 ~ 2,000 ノード

これまでの応用例

正常肺上皮細胞 SAEC+EGF(+Gefitinib) 投与下 EGF 応答パスウェイ解析

使用データ
 細胞: 肺正常細胞 SAEC
 刺激 / 薬: EGF, Gefitinib (イレッサ: 抗がん剤)
 マイクロアレイ: Agilent 44K Whole Human Genome Oligo Microarray

計測データ: 19 時点 x 1 回計測 in 48 時間
 EGF, Gefitinib 投与・非投与の 4 条件

結果 遺伝子間制御構造に差異のある遺伝子群の予測および肺がん予後予測シグネチャーへの応用に成功
 Yamauchi et al. *PLoS ONE*, 2012.

動的発現予測 → **発現制御構造推定** → **制御構造に差異のある遺伝子群の発現パターンに基づく肺がん患者群予後予測**

肺がん細胞株 PC9/PC9GR に対する Gefitinib 動的効果予測

使用データ
 細胞: 肺がん細胞株 PC9 および PC9GR
 刺激 / 薬: EGF, Gefitinib (イレッサ: 抗がん剤)
 マイクロアレイ: Agilent 44K Whole Human Genome Oligo Microarray

計測データ: 26 時点 x 1 回計測 in 26 時間

結果 薬剤効果プロファイル

その他の事例

- 乳がん細胞 (MCF7) 薬剤刺激データからの薬剤応答パスウェイ推定
- HUVEC データからの遺伝子/モジュールネットワーク推定
- メラノーマ細胞 / Paclitaxel による新規抗癌剤ターゲットの同定

Acknowledgments

Computational time was provided by the Super Computer System, Human Genome Center, the Institute of Medical Science, The University of Tokyo, RIKEN Integrated Cluster of Clusters (RIICC) system, and early access to the K computer at the RIKEN Advanced Institute for Computational Science.



D-7: SBiP

データ解析融合プラットフォーム



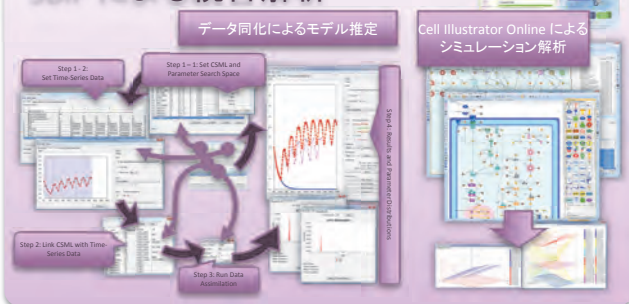
開発責任者: 宮野 悟 (東京大学 医科学研究所 ヒトゲノム解析センター)

SBiP 開発メンバー: 関谷 弥生, 斉藤 あゆむ, 山口 類, 井元 清哉, 島村 徹平, 新井田 厚, 玉田 嘉紀, 宮野 悟

概要

本研究は、データ解析融合チームで開発した技術とプログラムを、相互に連携させるソフトウェアプラットフォームとしての SBiP を構築することを目標としている。平成 24 年度は、平成 23 年度の開発で達成した SiGN を東京大学ヒトゲノム解析センターをテスト環境として実装した経験を踏まえ、京のゲートウェイノードから SiGN などの解析コンポーネントが実行できるように改良を行い、ソフトウェアとしての完成度を高めた。ユーザは SBiP に用意されているさまざまな解析コンポーネントを組み合わせながら、カスタマイズした解析フローを実行可能である。例えば SiGN-SSM、SiGN-L1、SiGN-BN などの遺伝子の発現情報からネットワークを推定するコンポーネント群を京の上で実行し、それらの結果をユーザに整理し視覚化して表示することが可能である。オンラインドキュメントの整備と改良は継続して行っている。また、平成 22 年度までに取得した肺がん及び乳がん関連データについて、SBiP 上の解析パイプラインで解析処理を行う上で必要となった解析コンポーネントの改良を行った。

SBiP による統合解析



SBiP オンラインドキュメント整備

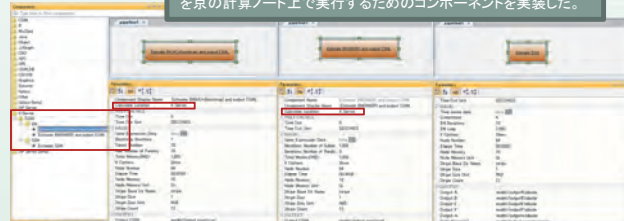


<http://sbip.hgc.jp>

SBiP オンラインドキュメントにおいて、各コンポーネントの解説ページの更新を行いつつ、SBiP アプリケーションの起動のほか、目的別の解析フローの作成例やサンプルフローの作成に関するビデオチュートリアルを作成した。

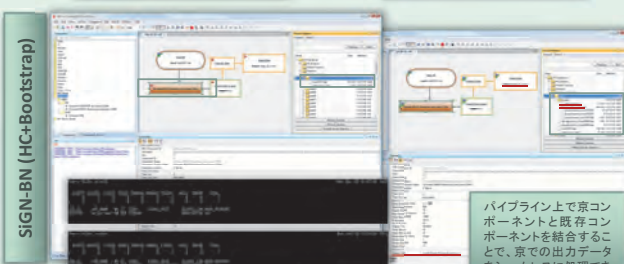
SBiP を使用した京での SiGN の実行

Webサービス機能を利用し、
● SiGN-BN (HC+Bootstrap) ● SiGN-BN (NNSR) ● SiGN-SSM
を京の計算ノード上で実行するためのコンポーネントを実装した。



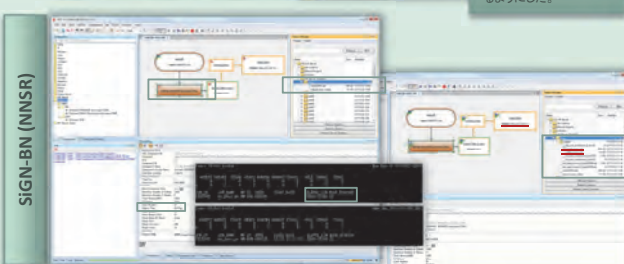
実行中

実行後



SiGN-BN (HC+Bootstrap)

パイプライン上で京コンポーネントと既存コンポーネントを結合することで、京での出力データをシームレスに処理できるようにした。

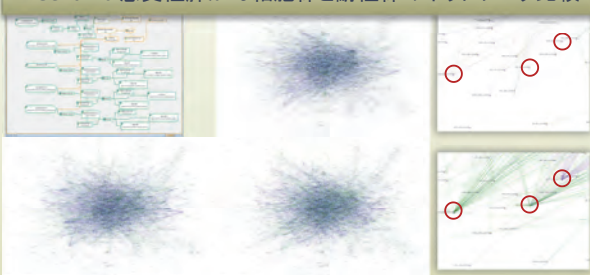


SiGN-BN (NNSR)



SiGN-SSM

Gefitinib 感受性肺がん細胞株と耐性株のネットワーク比較



肺がん・乳がん関連データ解析

乳がん細胞株のリガンド応答時系列データを使ったネットワーク推定法の比較





D-8: LiSDAS

生命体シミュレーションのためのデータ同化技術の開発

開発責任者: 樋口知之 (統計数理研究所データ解析融合研究開発チーム)

○斎藤正也 吉田亮 中野慎也 長尾大道 (統計数理研究所・データ同化研究開発センター)



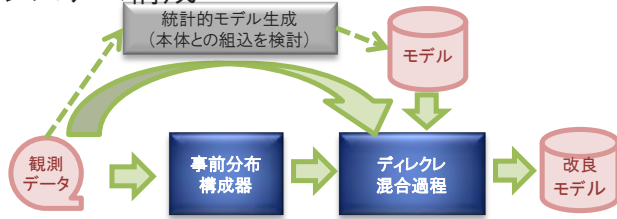
統計数理研究所
The Institute of Statistical Mathematics

データ同化
研究開発センター

1. 背景

- 実験にもとづくタンパク質相互作用や転写制御因子の働きについての知識のみから、現象を再現する力学モデルを構成することは未だ困難である。
- この困難の根本的は、細胞間の個体差のために計測値の誤差が大きいことや、おそらく実験で検証されていない反応過程が存在するためであると考えられる。
- そこで、LiSDASでは、生体内分子の計測データを参照値としてあたえることで、シミュレーションの再現性・予測力を改善するためのパラメータチューニングやモデルの改良をベイズ統計の枠組みで実行する。

システム構成



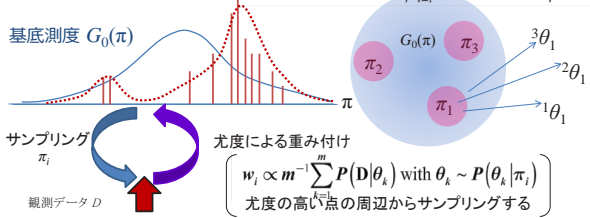
モデル

- 転写制御回路
 - 分解 $\frac{dx_j}{dt} = -d_j k_j + {}^c k_j A(x_{a(j)} | c_j, p_j) R(x_{r(j)} | d_j, q_j) + {}^b k_j x_{b(j,1)} x_{b(j,2)}$
 - 転写活性 $A(x | c, p) = x^p / (x^p + c^p)$
 - 転写抑制 $R(x | d, q) = c^q / (x^q + d^q)$
 - 複合体生成
- モデルパラメータ $\theta = (x_j(0), d_j k_j, c_j k_j, b_j k_j, c_j, p_j)$ の平均・分散
- 観測時系列 $D = (y_{i,n})$

2. 計算手続き

ディレクレ混合過程を利用したパラメータ推定

$$\text{改良された分布 } P(\pi | \Pi_n, D, \gamma) = \frac{1}{n + \gamma} \sum_{i=1}^n w_i \delta_{\pi_i}(\pi) + \frac{\gamma}{n + \gamma} G_0(\pi)$$



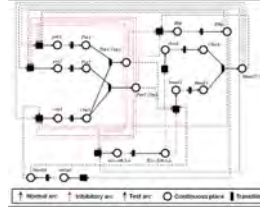
- D : 観測データ
- θ, π : モデルパラメータ (反応速度, 転写因子しきい値)
- " $a \sim b$ ": 分布 b からサンプリングして a に代入

```

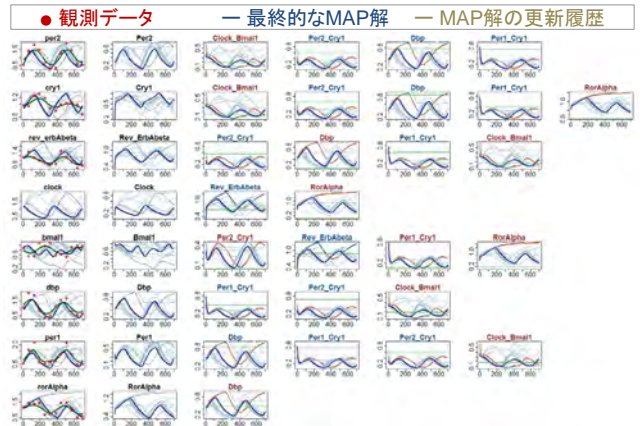
For i = 1, ..., N [N ~ 10^6]
  • For j = 1, ..., m:  $\theta^{(j)} \sim p(\cdot | \pi^*)$ . [m = 128]
  •  $w^* = \sum_j p(D | \theta^{(j)}) / m$ . [ $p(D | \theta^{(i)}) \approx w^*$  の近似に使う]
  • If  $i \leq n$  then [保持する  $\pi$  の個数  $n = 128$ ]
    •  $\pi^{(i)} := \pi^*$ ,  $w^{(i)} := w^*$ .
    •  $\pi^* \sim G_0(\cdot)$ . [LiSDASが前処理で設計]
  • Else
    •  $\theta^* \sim \sum_j p(D | \theta^{(j)}) \delta(\theta - \theta^{(j)}) / m$ .
    •  $\pi^* \sim p(\pi | \pi^{(1)}, \dots, \pi^{(n)}, \theta^*)$ 
    •  $\propto 1 / (\gamma + n) \times \sum_i p(\theta^* | \pi^{(i)}) \delta(\pi - \pi^{(i)})$ 
    •  $+ \gamma / (\gamma + n) \times p(\theta^* | \pi) G_0(\pi)$ .
    • If  $\exists k \in \{1, \dots, n\}$ .  $p(D | \pi^{(k)}) < p(D | \pi^*)$ 
      then
        •  $\pi^{(k)} := \pi^*$ ,  $w^{(k)} := w^*$ .
    
```

3. 性能測定

哺乳動物の概日周期転写ネットワーク



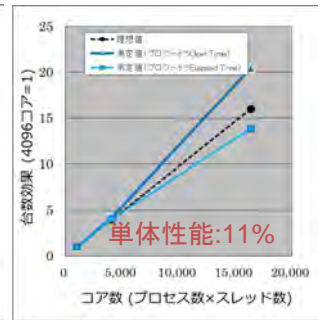
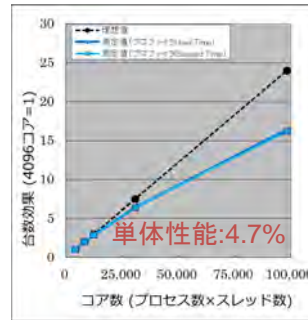
- モデルの規模
- システム変数の個数: 29
- 観測変数の個数: 7
- モデルパラメータ数: 116
- 観測をトレースするパラメータが発見できることを確認。



台数効果

評価法(通常)

評価法(京用)



5. ソフトウェア配布

- 配布先
 - ISLiM公開ページ http://www.islim.org/islim-dl_j.html
 - LiSDAS <https://www.islim.org/FS-APL/FS-Form/form.cgi?Code=Code24J>
- 2012年5月に試作版を公開
- 年度内に周辺ツールを追加した完成版を公開予定
- データ同化ライブラリ
- SBML(from/to)LiSDAS

参考文献

Yoshida R, et al., (2010) Bayesian experts in exploring reaction kinetics of transcription circuits, *Bioinformatics*, 26(18), i589-595.
 Yoshida R, et al. (2010) Bayesian learning in sparse graphical factor models via variational mean-field annealing, *Journal of Machine Learning Research*, 11:1771-1798.
 Yoshida R, et al. (2008) Bayesian learning of biological pathways on genomic data assimilation, *Bioinformatics*, 24(22):2592-2601.

D-9: MEGADOCK 網羅的タンパク質ドッキング解析プログラムMEGADOCKの開発と応用

開発責任者: 秋山 泰 (東京工業大学 大学院情報理工学専攻 計算工学専攻)
 開発担当者: 松崎由理, 内古閑伸之, 石田貴士, 大上雅史

MEGADOCKは、タンパク質立体構造データを入力として、細胞内のタンパク質を網羅的にドッキングすることで、タンパク質間相互作用 (PPI) ネットワークを予測するプログラムである。

我々は、空間をボクセル表現した近似的な剛体ドッキング計算について、形状相補性の評価モデル (rPSCモデル) を新規に提案し、従来法より高速な計算を可能とした。これに加えて、静電相互作用、疎水性に関する近似項も導入した評価関数を、高速フーリエ変換 (FFT) により効率的に計算するアプリケーションを開発した。また、OpenMPとMPIを用いたハイブリッド並列化を行うことで、大規模並列計算を可能にした。PPIネットワーク予測については、A) 標準的なベンチマークデータセット、B) 細菌走化性系、C) ヒトアポトーシス系、D) 非小細胞肺がんバスキュール の問題に適用し、タンパク質間ネットワーク予測精度としてそれぞれ F値 0.42, 0.44, 0.28, 0.36 を得た。

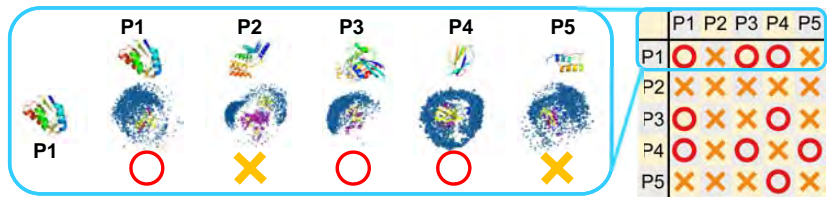
また、肺がんに関連の深い EGF 受容体シグナル伝達系を対象に 2,000 x 2,000 級の組み合わせのドッキング問題にチャレンジし、検討すべき未知タンパク質間相互作用を得た。
 京速コンピュータ「京」を用いたスケーラビリティ測定では、24,576ノードで良好な並列化効率 (ストロングスケーリング 0.95) を得た。

入力: タンパク質立体構造

出力: タンパク質間相互作用ネットワーク



網羅的ドッキングとポストドッキング解析



計算を効率的にする工夫① スコア関数の新規提案

FFTによる高速な剛体ドッキング計算

1回転角のスコア (3600通り計算) $L(l, m, n)$ $R(r, m, n)$ $O(N^6)$

$$S(\alpha, \beta, \gamma) = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N R(l, m, n) L(l + \alpha, m + \beta, n + \gamma)$$

FFTによる畳み込み計算 $O(N^3 \log N)$ (Katchalski-Katzir model)

$$S(\alpha, \beta, \gamma) = \text{IFFT} [\text{FFT}[R(l, m, n)] \text{FFT}[L(l, m, n)]]$$

新規に提案したスコア関数

三つの項一つの複素数で表現

- 形状相補性
- 疎水性相互作用
- 静電相互作用

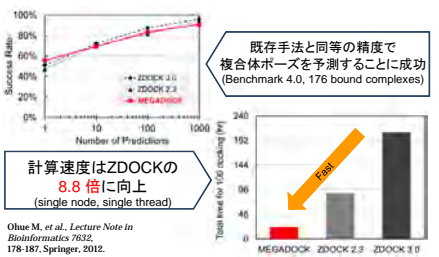
$$S(\alpha, \beta, \gamma) = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N R(l, m, n) L(l + \alpha, m + \beta, n + \gamma)$$

$$R(l, m, n) = \text{Re}[L(l, m, n)] + \text{Im}[L(l, m, n)]$$

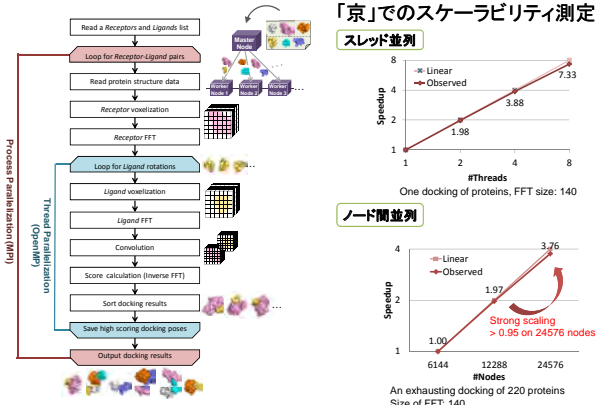
$$L(l, m, n) = \text{Re}[L(l, m, n)] + \text{Im}[L(l, m, n)]$$

FFTによる畳み込み計算が一面のみ (先行研究例: ZDOCK3.0では8回)

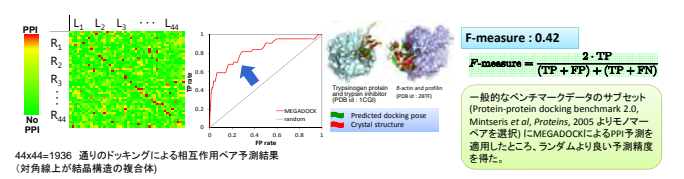
ドッキング精度と計算速度の検証



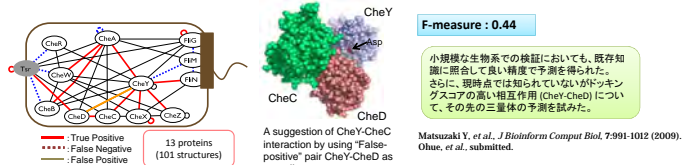
計算を効率的にする工夫② ハイブリッド並列化



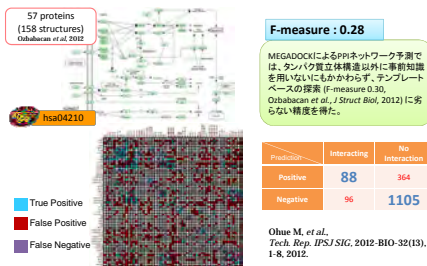
A) ベンチマークデータによるタンパク質間相互作用予測の検証



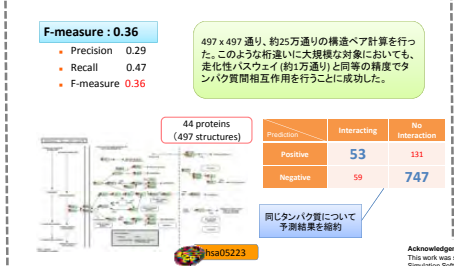
B) 細菌走化性バスキュールへの応用



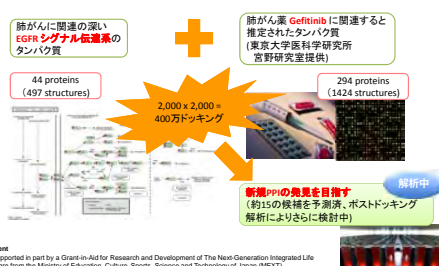
C) アポトーシスバスキュールへの応用



D) 非小細胞肺がんバスキュールへの応用



肺がん薬関連タンパク質間相互作用の探索



Acknowledgement
 This work was supported in part by a Grant-in-Aid for Research and Development of The Next-Generation Integrated Life Simulation Software from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT). Part of the result was obtained by early access to the K computer at the RIKEN Advanced Institute for Computational Science.



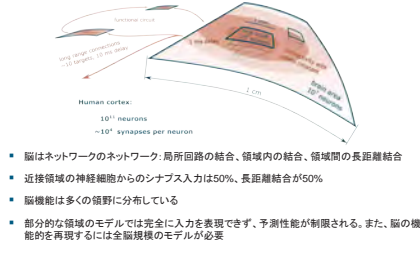
B-1/B-2: NESTによる局所回路シミュレーション



五十嵐 潤¹、Moritz Helias^{1,2}、舛本 現³、Tobias C Potjans^{1,2}、Markus Diesmann^{1,2}、深井朋樹¹

1 理化学研究所 次世代計算科学研究開発プログラム 脳神経系研究開発チーム、2 Inst of Neuroscience and Medicine (INM-6), Computational and Systems Neuroscience, Jülich Research Centre, 3 理化学研究所 次世代計算科学研究開発プログラム 高度化チーム

人間の脳の構造



全脳規模のシミュレーションの問題

- メモリオーバーヘッドが並列数とともに上昇する
- 計算時間ではなく、メモリ量がネットワークサイズの上限を決める
- 全メモリ資源を使う、最大限にメモリを使用して規模を拡大
- メモリー消費の数理モデルをベースとした解析
 - Kunkel S, Potjans TC, Eppler JM, Plesser HE, Morrison A and Diesmann M (2012) Meeting the memory challenges of brain-scale network simulation. *Front Neuroinform* 5:35
 - 異なるスケールでは、メモリ消費占有する部分が、ソフトウェアの異なる構成部分になる
- 共通のプラットフォームの必要性
 - 脳のモデルは非常に複雑で、作成、デバッグには大きな努力
 - 他者によって発表されたモデルの再現は容易ではない

NEST シミュレータ



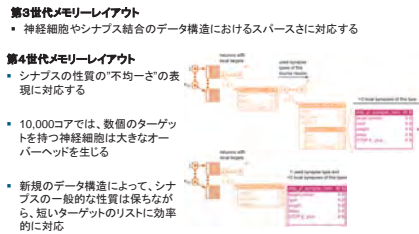
重要な目標:
体系的に新しいシミュレーション技術を公開する
GPLの一般的なリリースで
作製

- 多数の研究室との共同研究 (2001年~)
- 国際的な高等課程における教育
 - 沖縄計算論の神経科学コース
 - Advanced Course in Computational Neuroscience ACCN, ヨーロッパ

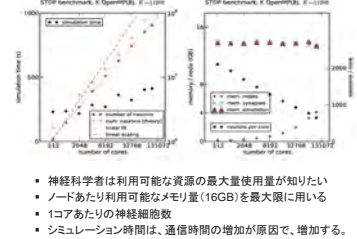
10⁹の神経細胞からなる局所回路モデルにむけて

- 京における大規模シミュレーションの3つのマイルストーン**
 - 京へのNESTの移植 第2世代 * 2011年9月 NEST 2.1
 - 10⁸の神経細胞の実現 第3世代 2012年5月 NEST 2.2(次期リリース)
 - 全脳規模(10⁹)にむけて 第4世代 2012年9月
- *第2世代 = 第2世代 シミュレーションカーネル
- 10⁸に相当する規模:**
 - 最大視野(第一視覚野)よりも大きい規模
 - 視覚皮質のモデルが可能
 - より大きなネットワーク、大きな遅延、疎な巨視的結合

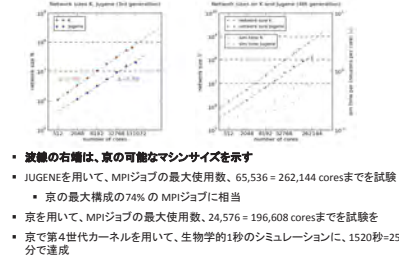
第3世代、第4世代カーネルのメモリー配置



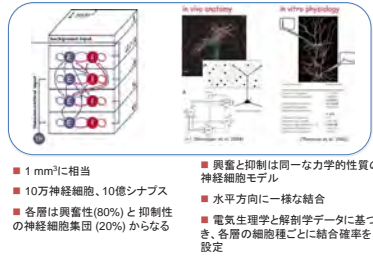
ソフト研究開発成果(NEST): 京における第3世代のモデル規模の拡大



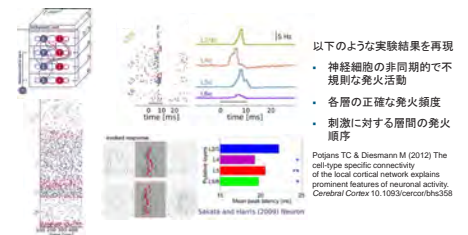
ソフト研究開発成果(NEST): 京とJUEGENEにおける第3、4世代カーネルの性能比較



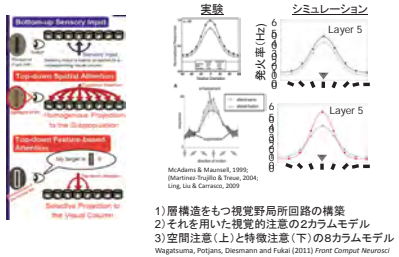
ソフト研究開発成果(NEST): 大脳皮質の最小構成単位となる階層型局所回路モデル



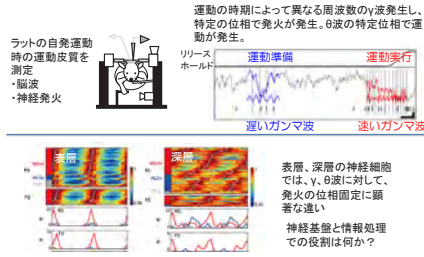
ソフト研究開発成果(NEST): 各層の神経細胞の振る舞いを再現



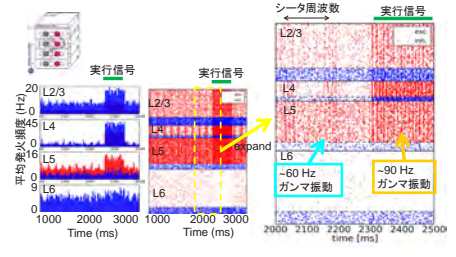
ソフト研究開発成果(CMDN): 視覚的、特徴的注意の神経活動への影響



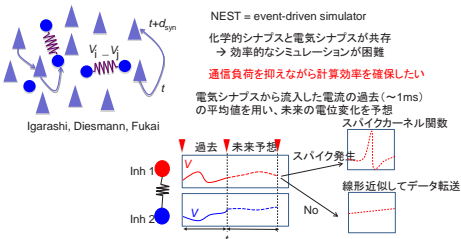
ソフト研究開発のための実験解析(CMDN): 運動皮質のθ、γ帯の神経的な振動現象



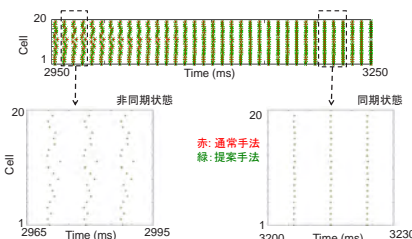
ソフト研究開発成果(CMDN): 運動皮質モデルによる振動現象の再現



ソフト研究開発成果(CMDN): 少ない通信回数での電気結合の計算



ソフト研究開発成果(CMDN): 電気シナプスで結合した20個の抑制細胞



謝辞

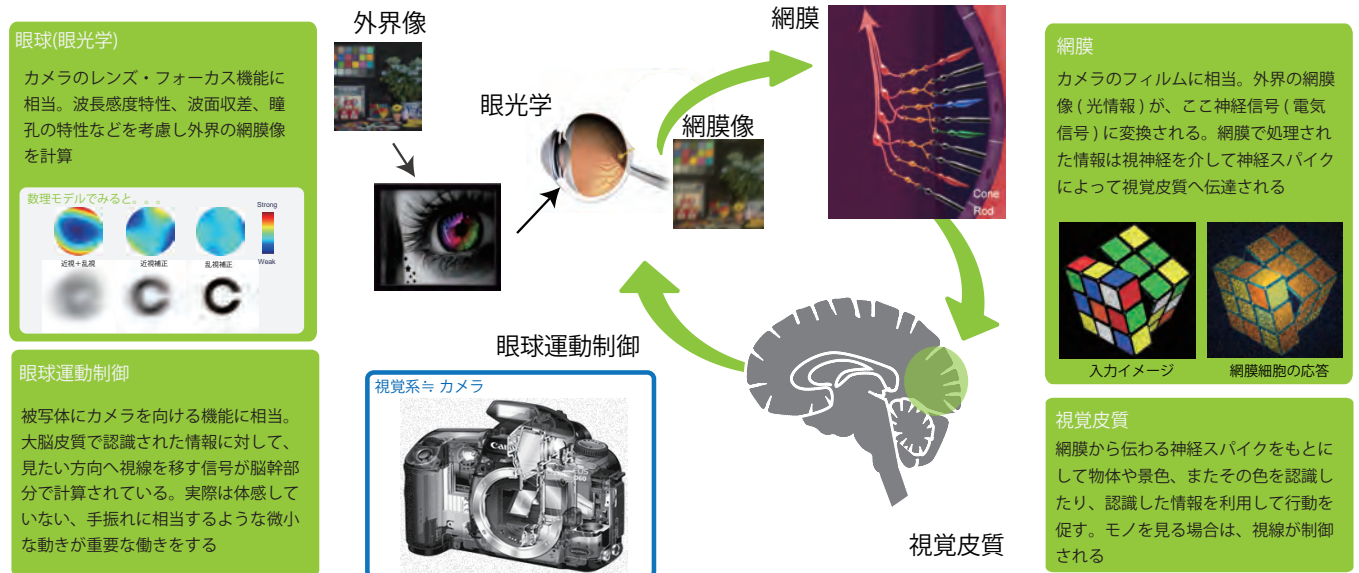
共同研究者
CMDN 磯村 宜和 (玉川大学 脳科学センター/ 局所回路実験データ収集)
半田 高史 (次世代計算科学/ 局所回路実験データ収集)
我妻 伸彦 (学振PD/ 視覚注意の局所回路モデル)
NEST Susanne Kunkel

- 京は現在開発中であり、スライド中の数字は現状の値である。
- 京上の性能に関しては京コンピュータの試験利用、および特別運用による。開発および運用にあたる関係各位に感謝申し上げます。とくに京上の実験については、ISLiM高度化チームの協力無くしては困難であった。感謝申し上げます。

B-3: VSM

全視覚系モデルによる視覚情報処理の解析
(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)

臼井支朗, 稲垣圭一郎, 観音隆幸, 神山齊己 (愛知県立大), 石原彰人 (中京大学)



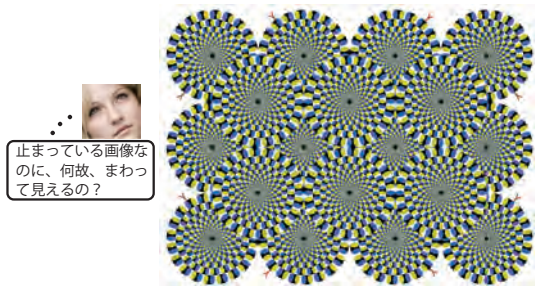
全視覚系モデルの構築

脳機能の1つである視覚、「ものを見る」ことをターゲットに、視覚機能がどのように脳で実現されているかを解明する。

具体的に何をやるの？

眼球運動、眼光学、網膜ならびに視覚皮質からなる全視覚系を詳細な大規模モデルとして構築し、種々の視覚現象をシミュレーションにより解析する

→ 各サブシステムレベル (Micro) と全視覚系レベル (Macro) の観点から視覚現象における視覚系の役割を解明する



PLATO(共有モデル統合環境)

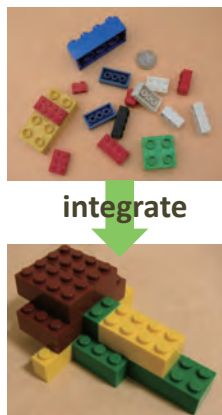
Sub-models (parts)

- 詳細サブモデルの構築と並列化
- 共有 IO フォーマットによるモデル接続
- Agent による統合モデルの実行・管理

We should share

- data format
- source code
- I/O data
- parameters
- hardware

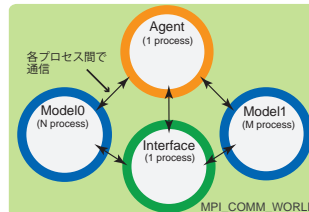
Large-scale model



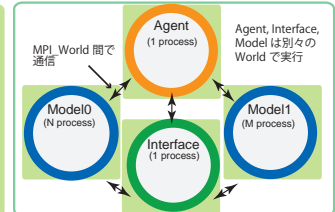
PLATO によるモデル統合

PLATO では、2つ (MPI World 共有型 (MPMD)、複数 MPI World 統合型) のモデル統合方法を提供している。モデルは Interface によって接続され、モデル実行・データ交換のタイミングを制御する Agent により統合モデルのシミュレーションが管理される

MPI World を共有する方法 (MPMD)



独立した複数 MPI World を統合する方法

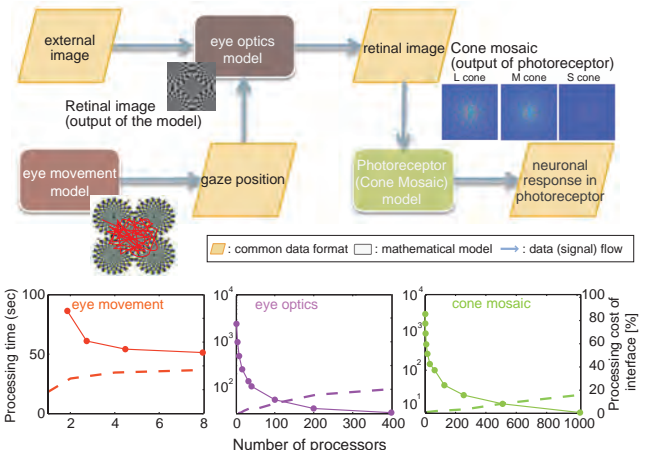


スタンドアロン型のスパコン (京など) におけるモデル統合に有用

高速ネットワーク経由で複数スパコンで個別に処理される高並列モデルの統合といった応用も可能

視覚系のモデル

ヒトの視覚は、眼球、眼光学系、網膜、脳皮質により構成されており、外界の情報は、これらの部位で多階層的に処理されることで、物の認識や視線の移動といった機能が実現されている。シミュレーションでは、入力イメージに対して、それぞれの部位の出力 (神経応答等) の可視化、情報量による解析を行い、視覚機能の解明を進める。



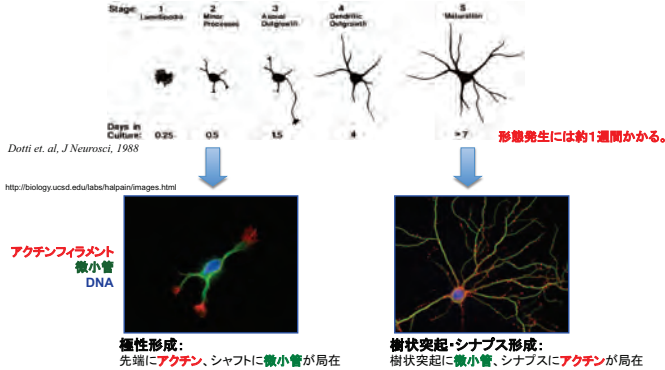


B-4: NeuroMorphoKit 神経細胞形態シミュレーションキット

開発責任者: 行縄直人 (京都大学 大学院情報学研究所)
 本田直樹 (京都大学 大学院情報学研究所)
 石井 信 (京都大学 大学院情報学研究所, 理化学研究所 次世代計算科学研究開発プログラム)



神経細胞の形態と細胞骨格



生物学的なゴール

- 形態形成における細胞骨格の動的な再構成の理解
- 以下の因子に制御されるニューロンの形態変化のダイナミクスのシミュレーション
 - 細胞骨格 (アクチンフィラメント および 微小管)
 - 細胞膜のダイナミクス
 - シグナル伝達

ソフトウェア開発のゴール

- 個々のフィラメントを基本とした
- 細胞内シグナル伝達・細胞骨格・細胞膜を統合した
- 全細胞スケール(数十万~数百万本のアクチンフィラメントを含む)の大規模な形態変化を記述する

モデルおよびシミュレータを開発する

基礎方程式

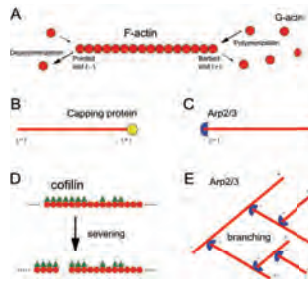
- 反応拡散方程式
- 細胞骨格・膜に関して新たに方程式を構築

連続シミュレーション

- 細胞内シグナル伝達系(反応拡散系)
- 細胞骨格系(アクチンフィラメント)
- 細胞膜

アクチンフィラメント

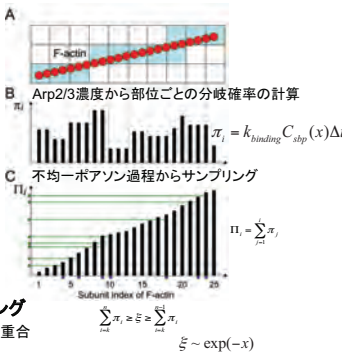
F-actinのキネティクス



キネティクスイベントは確率的にサンプリング

- アクチンモノマー分子の濃度に比例した確率で重合
- Arp2/3の濃度に比例した確率で分岐

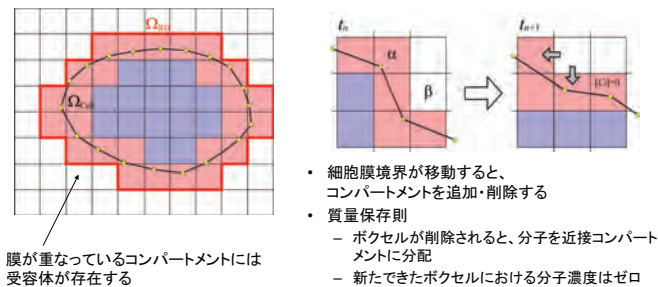
分岐計算の効率化



細胞内シグナル伝達

$$\frac{dC_i}{dt} = D_i \frac{\partial^2 C_i}{\partial x^2} + f_{rec}(C_j) + f_{actin}(C_i, F)$$

移動境界条件



膜のダイナミクスのエネルギーモデル

膜のダイナミクスは準安定状態を想定する

$$M(t) = \arg \min E[M, F(t)] \quad E = E_{volume} + E_{surface} + E_{bending} + E_{F-actin}$$

体積を一定に保つ

$$E_{volume} = \alpha_{volume} (V - V_0)^2$$

表面積を一定に保つ

$$E_{surface} = \alpha_{surface} (S - S_0)^2$$

細胞膜を滑らかにする

$$E_{bending} = \alpha_{bending} \int \frac{1}{R(s)^2} ds$$

R: 曲率半径

アクチンフィラメントを真っ直ぐにする

$$E_{F-actin} = \sum_i \frac{1}{2} \tan^2 \theta_i \Delta x_i^2$$

Mogkner, et al., Biophys J, 1996

膜ポリゴンの頂点の位置は時間 Δt おきに以下の最急降下法により、Eを最小化するように最適化する

$$x_i^{t+1} = x_i^t - \eta^t \text{grad}_i E^t(x^t, F)$$

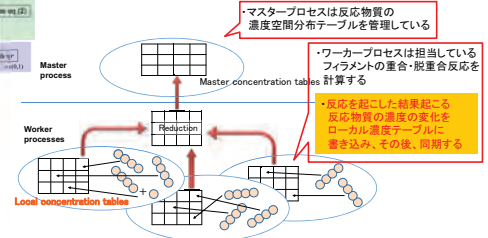
ソフトウェアの実装と京での性能評価

アルゴリズム全体



並列化モデル

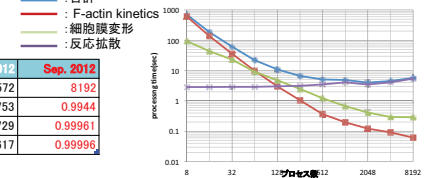
- 1個の管理プロセス(マスタープロセス)とN個のワーカプロセス
- ワーカプロセスはフィラメントを分散して保持し、フィラメントに関する計算に専念する



性能評価

総初期フィラメント数: 122万本
 タイムステップ: 0.5 (ms)
 計算ステップ数: 5ステップ

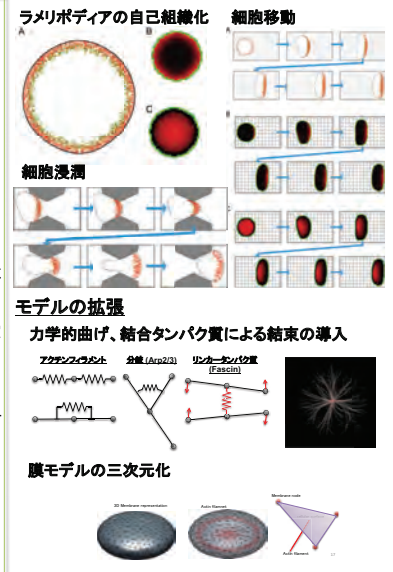
	Mar. 2012	Sep. 2012
Maximum # of cores on K	572	8192
Parallel portion (Total)	0.93753	0.9944
Parallel portion (Membrane)	0.99729	0.99961
Parallel portion (Actin)	0.93617	0.99996



ソフトウェアの現状と今後

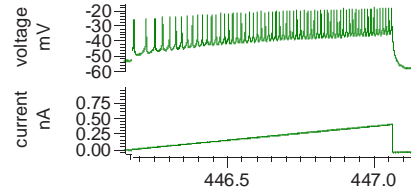
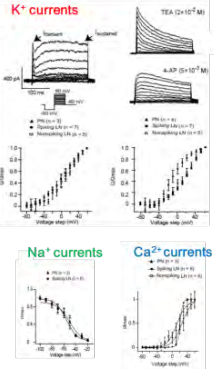
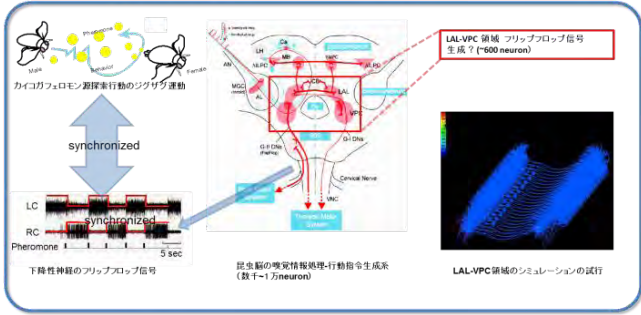
- 120万アクチンフィラメント、二次元200x200コンパートメント、膜ノード数100規模の細胞形態変化シミュレーションを実現
- 本シミュレータを用いて葉状仮足の自己組織化、細胞移動および浸潤における細胞の形態変化を再現 (Honda, Nonaka, and Ishii, *Neural Networks*, 2011)
- 大規模並列環境(京 8192並列、RICC 4096並列)での動作を実現(ストロングスケール $\alpha=0.99$ (8192))
- 膜電位シミュレーション機能を持つ神経細胞シミュレータA-Cell (京都大学・市川一寿教授による開発)との連携動作のためのデータ交換フォーマット(NetCDF-neuron)を策定
- アクチンフィラメント計算の並列化法を改善中、京での1万コア程度でのシミュレーションを実現を目指す
- パネネットワークモデルによるアクチンフィラメントの曲げ、結束のモデルへの導入、三次元化によるモデル拡張と統合を実施中
- 糸状仮足成長円錐の動的な勾配検知、方向転換についての高並列シミュレーションを実施、イメージングデータの再現を図る

シミュレーション結果



開発責任者: 神崎亮平 (東京大学先端科学技術研究センター 脳神経研究開発チーム)

目標: 昆虫脳の嗅覚情報処理から行動信号生成までの神経ネットワークを実験と比較できる詳細モデルを使ってシミュレーションする



触角葉出力神経からの current clamp

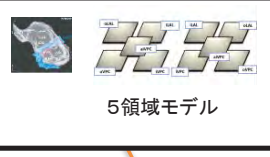
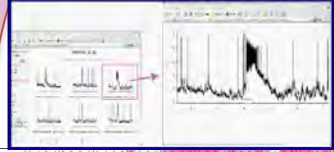
カイコガ細胞には 2つ以上のK⁺電流
1つのNa⁺電流 一つ以上のCa²⁺電流が存在する

触角葉細胞体からの voltage clamp

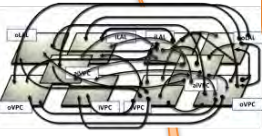
Morphology



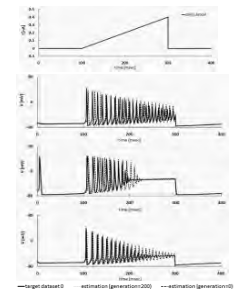
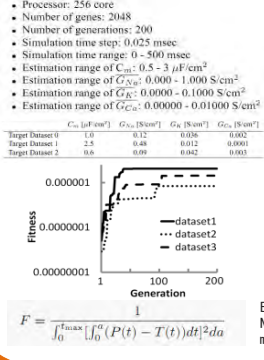
Physiology



可能な結合を推定

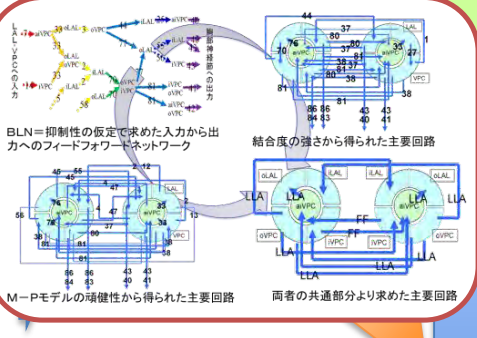
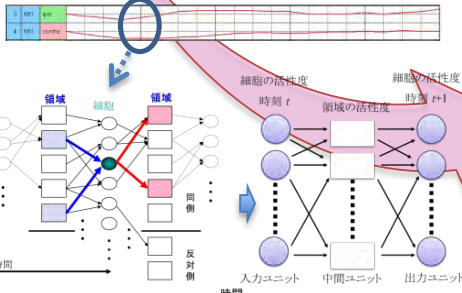


Parameter estimator using NEURON



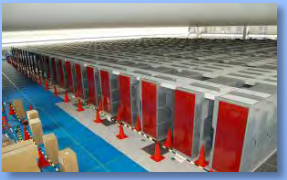
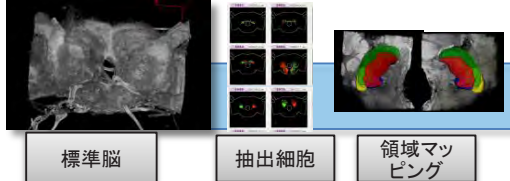
Estimator by Genetic Algorithm using NEURON is implemented, fitting for multicompartment model is ongoing.

平均応答波形BN_ID661

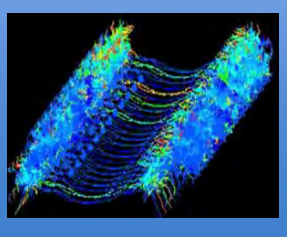


シナプス結合強度を推定

領域を介した結合間のパターンの変化と見なす
すべてのパターン遷移を縮約してニューラルネットワークとして表現



Simulation



結果は当日に発表します

※本研究は、京速コンピュータ京の試験利用を用いている。



B-5: IOSSIM

昆虫嗅覚系全脳シミュレータ

開発責任者: 神崎亮平 (東京大学先端科学技術研究センター 脳神経研究開発チーム)



背景

ヒト(1000億個の神経細胞)
昆虫(10万個の神経細胞)

脳をつくる神経細胞(ニューロン)の形やはたらきは昆虫・哺乳類をとおして共通だが、神経系の規模は小さい。
異種感覚統合 連合学習 定位行動など環境に適応するため充分な複雑な行動を行う。
そのため 昆虫脳は感覚から行動までの脳の情報処理をシミュレーションするためのモデル系をして優れている。

シミュレーションの構築

形態抽出 → mapping → シナプス結合可能性の推定
シナプス結合強度推定
神経の生理・形態記録

神経細胞データベース BoND
細胞形態抽出 NEWRITE
サイコが標準編 BrainComposer
細胞膜情報推定・結合強度推定
大規模シミュレータ NEURON K+

カイコガのニューロンデータベース(BoND)とマルチスケール計測

Database Morphology Physiology Gene Neuron Networks Behavior Robot

形態抽出ソフト(KNEWRITE)

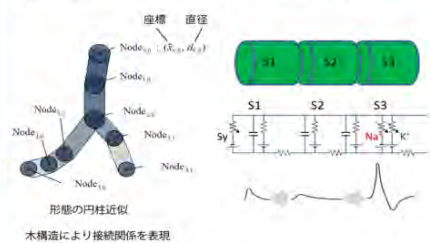
抽出時間の評価
形態的な変換精度の評価
シミュレーションによる変換精度の評価

標準脳構築ソフト(BrainComposer)

抽出・変形による標準脳への登録
標準脳に各種テンプレート画像
標準脳スタック画像
標準脳テクスチャ画像

基礎方程式

マルチコンパートメント・等価回路モデル



マルチコンパートメントモデル等価回路方程式

$$C_m \frac{dv_n}{dt} = \sum_i g_{in}(v-v_i) + I_{n-1,n} + I_{n+1,n}$$

電位依存性イオンチャネルのコンダクタンス

$$\frac{dg_i}{dt} = m^{x_i} y_i (x_i, y_i = 1 \sim 4)$$

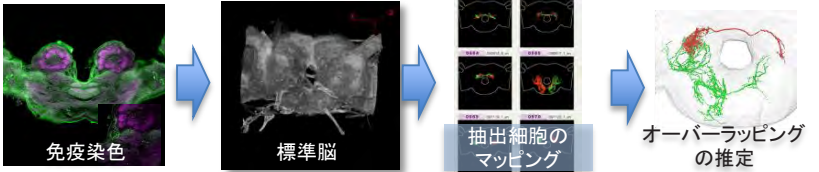
$$\frac{dgate}{dt} = \alpha_i(v)(1-gate) - \beta_i(v)gate$$

m or $n \in gate$
 α β はシグモイド関数

シミュレーションを用いた神経パラメータ推定

実数値GAIによるパラメータ推定
パラメータ探索の並列性能の評価
シミュレーションによる仮想的な電流クランプ応答波形を用いた。推定ソフトウェアの評価

シミュレーション前のデータの流れ



NEURON K+

ベンチマーク用回路の構築
単体性能の高速度化
通信高速度化
MPI_Algorithm
MPI_Isend
Split branches & allocate to each core
5000個の神経細胞(各5000コンパートメント)を0.025msec単位でリアルタイムに計算するには、約300 TFLOPS必要
現在は最大187 TFLOPS(20万コア)を達成

謝辞

本ソフトウェアの開発に際しては、京速コンピュータ京の試験利用、理化学研究所情報基盤センターのRICC、東京大学情報基盤センターのHA8000とFX10を用いた。京上の実装におけるISLiM高度化チーム、特に河本氏、大野氏のサポートに感謝する。

- 昆虫全脳シミュレーション開発チーム
神崎亮平 並木重宏 ハウプト・S・周一
加沢知毅 高嶋暁 櫻井健志 岩月知香
佐藤陽平 宮本大輔 森友亮 田淵理史
藤原輝史 峰岸諒 佐藤秀仁
池野英利 石井愛弓 新田倫子
西川都子 小林嘉太 山岸嘉彦 小杉展弘
小野島隆之 五十嵐吉輝



H-1: cppmd

大規模並列用MDコアプログラム



開発責任者: 大野 洋介 (理化学研究所生命体基盤ソフトウェア開発・高度化チーム)

Molecular Dynamics Simulation

The cell is constructed by molecules such as water, protein etc. Every molecules are constructed by atoms. These atoms influence each other and move complexly. Molecular dynamics (MD) simulation calculates the interaction and motion of atoms and reveal the structure and functions of the bio-molecules.

Because coulomb force that is force between charged atoms is affect to long distance, we must calculate force between all pair of atom. It takes very long time when the number of atom is large. Using parallel computer, it requires many communication. Therefore we use many techniques that reduce calculation and communication considering coulomb force is small at long distance.

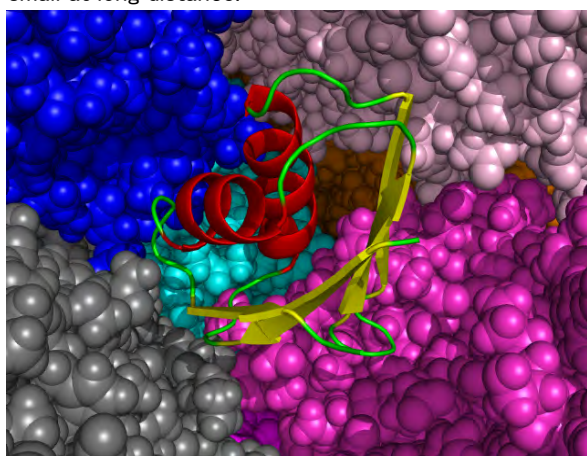


Figure1 Example of Protein(hide water)

Parallel Performance

We used spatial decomposition by rectangular for parallelization. One CPU calculate atoms in one rectangular box. When region that one CPU has is small, the communication time is larger than that of calculation. Total simulation time is decrease proportional to reciprocal of the number of CPU (N_{CPU}) when N_{CPU} is small, but this decrease is limited at large N_{CPU} . These decrease of simulation time against to the number of CPU for simulation with fixed size is “strong scaling”. We obtained the result that the limit was 100 atoms per CPU (Table 2 and Figure 2). Full K computer will perform simulation with over 10 million atoms.

The performance increase when the size of simulation per CPU is fixed is “weak scaling”. In the case of the number of atom per CPU is 6500, the performance was completely scaled (Table 3).

Table 3 Weak Scaling
Time consumption for 6,542 atom/CPU

Number of Node	64	512	4,096	32,768	79,872	82,944	
Number of Atom	418,707	3,349,656	26,797,248	214,377,984	522,546,336	542,644,272	
Time (ms / step)	Total	109.058	110.535	111.186	111.672	112.414	112.085
	Force	91.622	91.528	91.601	92.329	91.262	91.641
	Communication	12.124	13.694	14.011	14.012	15.820	15.067
	Other	5.312	5.313	5.574	5.332	5.332	5.376
Performance (PFLOPS)	0.003	0.025	0.201	1.599	3.871	4.031	
Efficiency (%)	39.0	38.5	38.3	38.1	37.9	38.0	

Molecular Dynamics Simulation on the K computer

Using the K computer, we achieved 4.39 PFLOPS, 42.9% efficiency with 522 million atoms (Table 1).

Table1 Performance of Massive MD Simulations

Number of CPU	79,872	82,944
Number of atom	522 million	543 million
Number of atom per CPU	6542	6542
Time (for 1000 step)	116.4 sec	112.1 sec
Performance	4.39 PFLOPS	4.03 PFLOPS
Efficiency	42.9 %	38.0 %

Reduce CPU to fit simulation box
Topology was not adjusted

Table 2. Strong Scaling
time consumptions for 3,349,656 atoms

Number of CPU	64	512	4,096	32,768	
Number of atom per CPU	52,338	6,542	817	102	
Amount of communication*	2,232	936	504	342	
Target of communication	26	26	124	342	
Time (ms / step)	Total	823.07	110.54	25.90	10.77
	Force	731.66	91.53	12.50	2.41
	Communication	59.28	13.69	11.48	7.38
	Other	32.13	5.31	1.92	0.98

* volume of data swap area

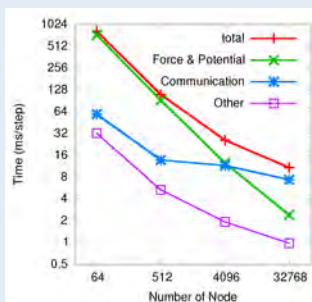


Figure 2 Strong Scaling

Table 4 Force calculation at one CPU
6,532 atom

Time consumption	77.45 sec
Performance	80.949 GFLOPS
Efficiency	63.2 %
SIMD ratio	57.2 %

SIMD ration include all instruction

Single CPU Performance

Atoms assigned to a CPU are distributed to 8 cores. When the number of atom per CPU is larger than several hundred, the load unbalance is small and 8 cores well perform.

SIMD instruction is applied to force calculation because the number of atom affect to an atom is large even if ignore far atoms.



H-2: LSV

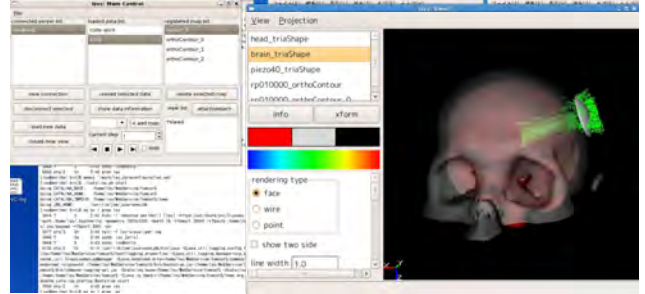
大規模データ可視化ソフトウェアLSVの開発

開発責任者: 小野謙二 (理化学研究所 計算科学研究機構)

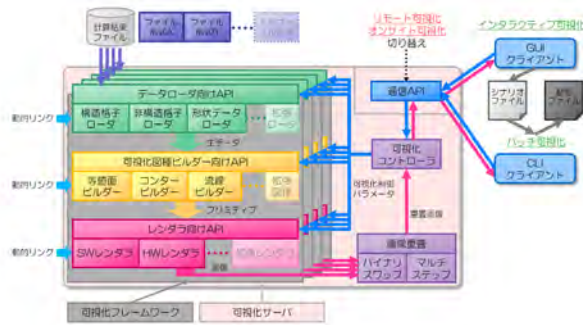


LSV “Large-Scale data Visualization software”

- 大規模データの並列可視化
- 共通UIによる統一化されたリモート/ローカル可視化環境
- インタラクティブ/バッチ可視化
- ソフトウェア/ハードウェアレンダリング
- プラグインによる機能拡張
- XMLによるスクリプト処理

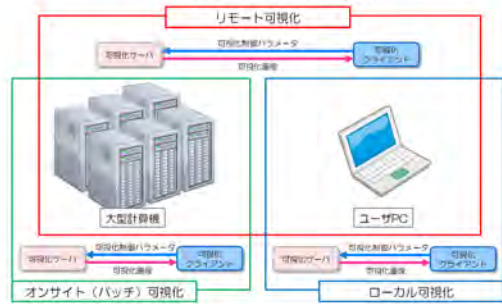


システム概念図



オペレーションモード

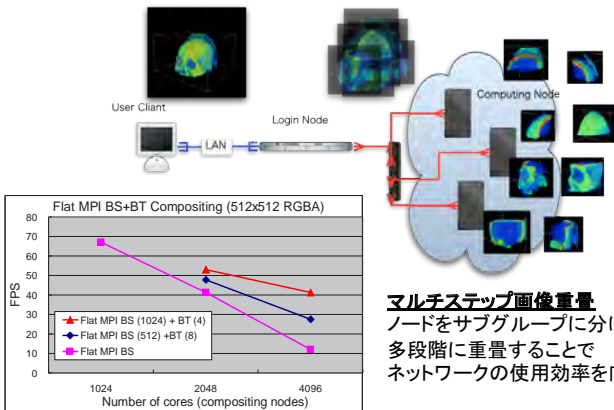
大規模データの並列可視化
インタラクティブ



定義されたシナリオによる動画作成
バッチ

小規模データの可視化
インタラクティブ・バッチ

画像重畳



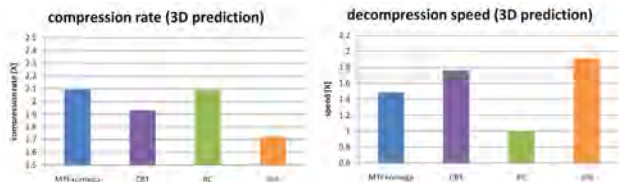
RICCでの並列可視化テスト

- データセット (XXX x YYY x ZZZ, 100GB)
- 100ノードを使用したリモートインタラクティブ可視化
- インタラクティブな視点・可視化パラメータの変更操作を達成
 - データロード: 46 sec
 - コンターライン作成: 0.14 sec
 - 画像重畳: 0.04 sec
 - レンダリング: 0.6 sec

今後の開発計画

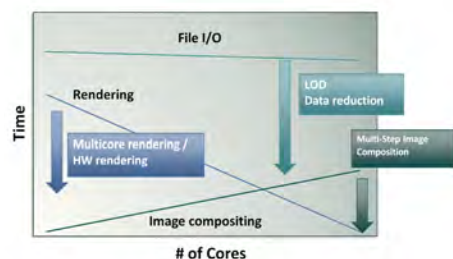
高速データ圧縮

- データロード時間の短縮
- 予測差分可逆圧縮
- 比較
 - **Move-To-Front + オメガ符号化**
 - **完全二分木符号化(CBT)**
 - Range符号化 [Lindstrom and Isenburg, 2006]
 - zlib: 汎用圧縮ライブラリ [Jean-loup Gailly and Mark Adler]



インテリジェントデータリダクション

- 特徴を保存した非可逆圧縮
- Level-Of-Detail
- Region-Of-Interest
- 特徴抽出
- データベース・メタデータの利用





H-3: SPHERE

Simulation Framework for Large scale Coupled Analysis



開発責任者: 小野謙二 (理化学研究所 計算科学研究機構)

背景

- ・シミュレーションプログラム開発の課題
 - ・シミュレーション対象がMulti-physics/Multi-scale化する事により、複雑なコードとなり、開発の難易度が高くなる
 - ・HPCアーキテクチャーが複雑化(Multi-core化、ヘテロジニアス化)し、効率的なプログラム開発が難しい
 - ・新しいアルゴリズムを用いてHPCの恩恵を得るには、物理と計算機の両方のスキルが必要
- ・解決策
 - ・アルゴリズムの研究開発に専念できるシミュレーションプログラム開発環境を提供
 - ・多様化するアーキテクチャーのローカライズを吸収することにより、今後のアーキテクチャーにも対応

SPHERE

Skeleton for PHysics and Engineering REsearch

並列シミュレーションプログラム開発のフレームワーク環境

- ・実行環境、計算条件設定インターフェース (XML)
- ・並列File I/O
- ・動的配列確保
 - 自動領域分割、ノード通信テーブル作成
 - 領域間通信用ライブラリの提供
- ・プログラムスケルトンの提供

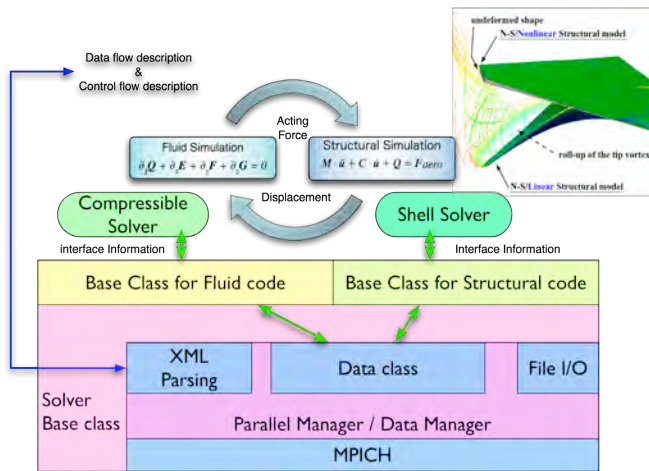
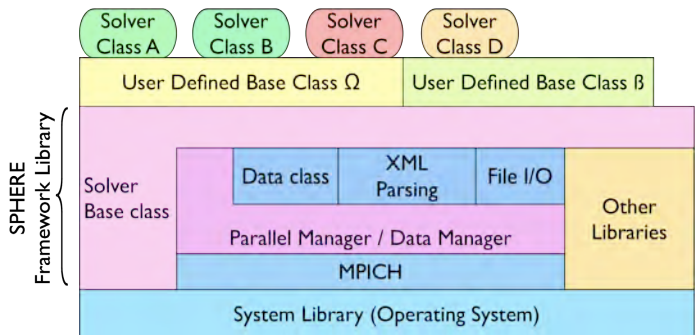


Fig.2 An example of coupled analysis on the middleware

Fig.1 SHPERE system

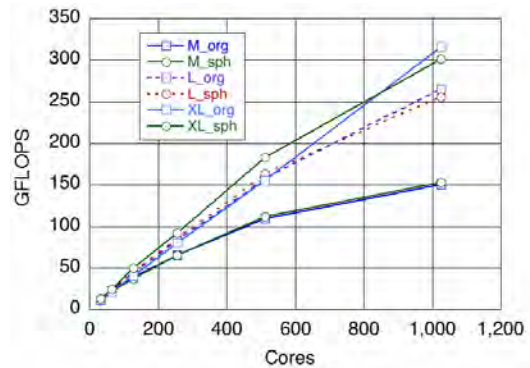


Fig.3 Scalability of SPHERE on RSCC

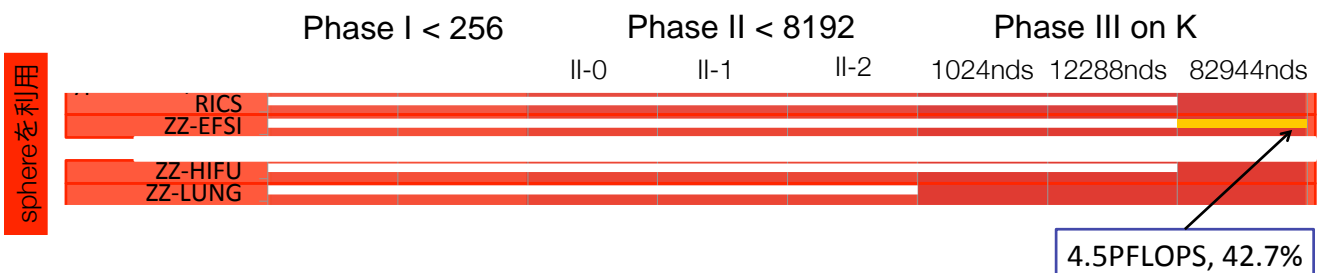


Fig.4 Performance of Software which based on SPHERE



H-4: VLSVL

合成可能な化学構造および反応スキームからなる大規模仮想ライブラリ



開発責任者: 船津公人 (東京大学 大学院工学系研究科)

1. 背景

- ・創薬の第一歩は有望なリード化合物の発見で、ハイスループットスクリーニング (HTS) が重要な役割を果たしてきた。しかし、HTSで有望なリード化合物が見いだせないケースもあり、その探索効率とコストが問題となっている。
- ・現在は、コンピュータ支援バーチャルスクリーニングによるリード化合物探索が推進されており、そのための、医薬品になりうる質の高い仮想化合物群からなる大規模なバーチャルライブラリの構築が望まれている。
- ・バーチャルスクリーニングにおいて、高スコアのリード化合物群を得られるかどうかは、バーチャルライブラリを構成する仮想化合物群の質や多様性、規模などに左右され、医薬品開発の成否に大きな影響を与える。総じて、**バーチャルライブラリの質と量が重要な意味を持つ**といえる。
- ・一方、スクリーニング結果、高スコアのリード化合物群を絞り込んで、文字通りバーチャルであるが故に、対象化合物を入手するための合成検討に多大なコストを要する。**この問題に対する解決策が必要**とされている。

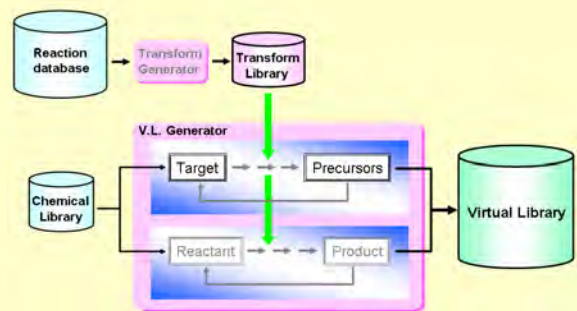
2. 目的

1. リード化合物となりえる仮想化合物群からなる、大規模な化合物バーチャルライブラリ (VL) の構築
2. VLに含まれる仮想化合物の多様性を確保し、仮想化合物の合成ルートに関する情報提供を可能とする

3. 方法

- ① 化学反応スキームに表現されるリアクタントとプロダクトについて、それぞれの反応部位情報と構造特徴の変化から記述されるtransformを反応データベースから抽出・集約する。
- ② 入力されたドラッグライクなターゲット構造に対して、このtransformを適用し、ターゲット構造を得るための合成前駆体を提案させ、蓄積する。
- ③ 出力された合成前駆体を、再度ターゲット構造としてシステムに再入力し、合成前駆体を提案させるという処理を繰り返す。
- ④ 入力構造に対して、反応予測方向にもtransformを適用し、反応生成物も提案させ、これを繰り返して、合成ルート樹を形成する化合物ライブラリとする。

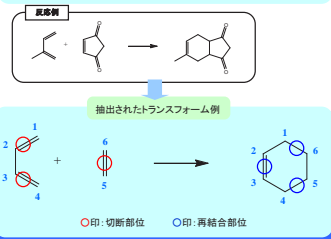
4. システムの構成



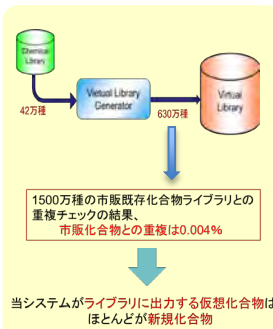
大規模な反応データベースからtransformを自動抽出することによって、多様な反応事例に適用できるtransformを多量に蓄積し、その適用によって誘導される仮想化合物群の多様性を確保する。

5. transform

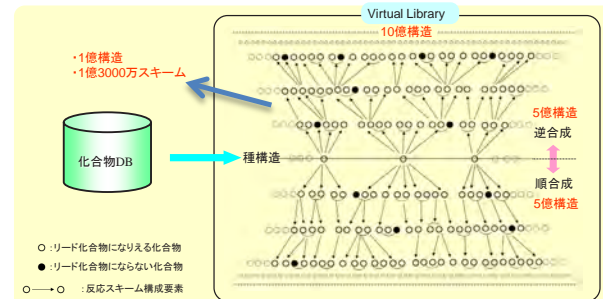
- ・transformには、反応スキームを構成する各化合物の反応中心情報が収納される。
- ・具体的には、反応中心の部分構造とその周囲の構造特徴情報、結合の切断部位、結合次数の変化した部位とその次数、新たに生成した単結合の部位、脱離した置換基の種類と部位、加わった置換基の種類とその位置など。
- ・システムに入力された化学構造に、各transformに記述された部分構造が含まれる場合は、該当部位をtransform構造に置き換えた反応スキームを提案する。



6. 発生化合物群の新規性



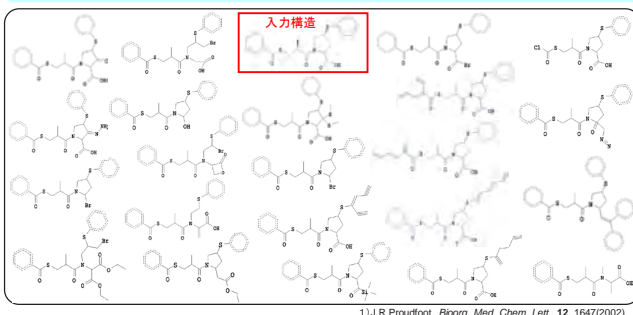
7. バーチャルライブラリの構成



- ・システムに化合物DBから種構造が入力されると、逆合成提案の場合は、種構造を合成するための前駆体が発生し、順合成提案の場合は、種構造を出発物質に含む反応生成物が発生する。
- ・そこで得られた化合物を種として、これらの処理を順次繰り返して、化合物ライブラリを構築する。
- ・矢印は反応スキームを表し、その両端の丸印は、各反応スキームを構成する化合物を示している。
- ・発生した化合物には、反応スキームの経路情報が付与され、合成経路検討の有用な情報となる。

8. 構造出力例

- ・下に、ある医薬品をtargetとした場合に、このシステムが提案する合成前駆体例を示します。
- ・これらの化合物は、入力構造 (target) の基本骨格を反映していることが確認できます。そのため、発生化合物の物理化学的、及び、薬理的な性質も程度受け継がれていると推測されます。
- ・医薬品とそのリード化合物については、構造的な差異は小さく、共通の骨格を持つ¹⁾とされています。
- ・よって、ドラッグライクな化合物を種構造として、当手法により構築されたバーチャルライブラリには、医薬品のリード化合物としての要求を満たす、質の良い化合物が含まれていると期待されます。



9. Lipinski's rule of 5 を指標とした評価例

分布の異なる種構造群の活用検討

