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エア 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エア 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		
		・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	鷲尾	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		
		・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	五十屆	嵐 潤(理研)
		 Multi-scale simulation of subcortical visual systems with real-time robotic visuo-motor feedback 	Jan I	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 科大)	支朗 (理研 / 豊橋技)
		・「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		
		・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エア 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エア 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総括報告	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
$\label{lem:multi-scale} \mbox{Multi-scale simulation of subcortical visual systems with real-time robotic visuo-motor feedback}$	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 -背景と現状と将来への展開-	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-大規模バーチャルライブラリの開発-	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的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 次世代シークエ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月慶應義塾大学理工学部教授

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エアの研究開発」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エ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る。

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、3 4本中1 3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、cpmd, 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。

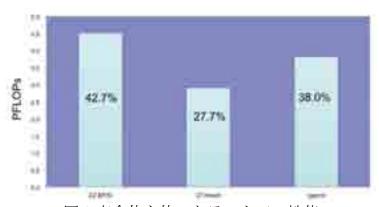


図3京全体を使ったテストでの性能

表1開発ソフトウェア一覧(全34本)

	コンピラスアあ株	No. of Contract of
	Francisco Address	マルチンと・・フルチスケール分字して11~ シェンは開発の意味上でもフラスライブラリ
	Phone NA	ACTUARROYMETERSTON-フェイス
N	MARKELA	美雄子班子教力学 联集
Ī	Editor.	程度をする利用 密度の開発通信表示(するべき対象電子対象 開発計算
	Fernan GACINE 19	THOU WHITE BETT BE - DIE
	Paris GM NA	はするでは ますながは、サラねかがり 東 利用をもずみがは、サテルののでは
*	HEE	棚舎シミュルーションブラットフォーム
	Z-011	を育めなかシモムレーテルシ(※サゼル場を正 はまな料をプログラム/
H	H-mt	株小部間にない。 ** ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・
8	E2:008E	常用手能力を500×502
3	HWY!	株型製造器から30~500 (キラセニ経済金を超プログラム)
	12-0.00E	部門者・技術権 2011年17月2
	A/MANAGE	プロテステール・マルチフィジリテスの繋がモル キートのコ



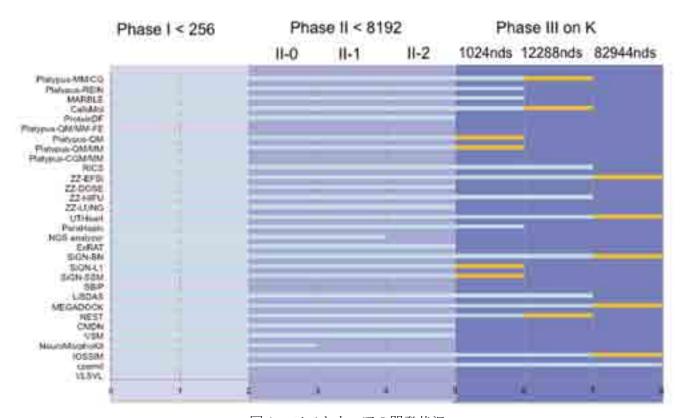


図1 ソフトウェアの開発状況

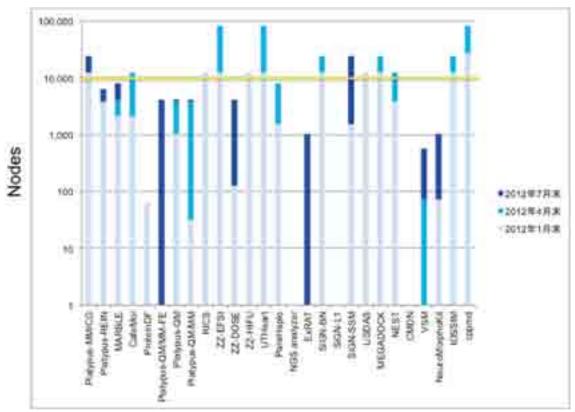
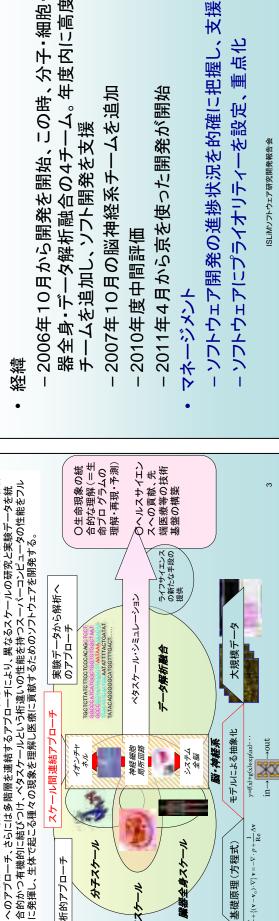


図2それぞれのソフトの京での利用可能ノード数



C'





モデルによる抽象化

留·神経系

システム 会談

かナスケーラ

 $\text{in} \rightarrow \bigcirc$ \longrightarrow out $y=f(x)+g(x)exp(az)\cdots$

 $\frac{\partial \mathbf{v}}{\partial t} + \left\{ (\mathbf{v} - \mathbf{v}_0) \cdot \nabla \right\} \mathbf{v} = -\nabla \cdot p + \frac{1}{Re} \Delta \mathbf{v}$

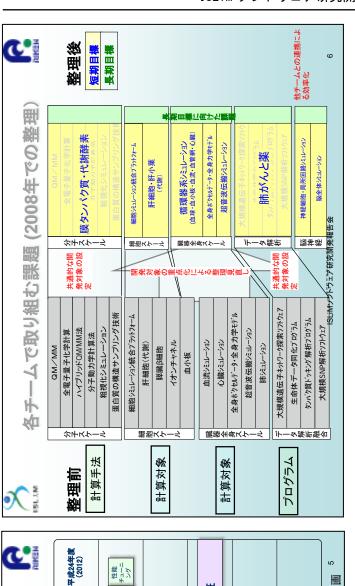
基礎原理(方程式)

研究開発の概要と達成目標

スケール間連結アプローチ

解析的アプローチ

Œ.



性能 チューニ ング

試作·評価·製造

詳細設計

概念設計

システム

実証

中間評価委員会

開発·評価

次世代生命存能合シミュレーション

第1回中間評価委員会

開発·製作·評価

ソフトウェア (グランドキャリンジ アプリケーション)

建設

設計

計算機棟

平成23年度 (2011)

平成22年度 (2010)

平成21年度 (2009)

平成20年度 (2008)

平成19年度 (2007)

平成18年度 (2006)

拖

次世代スパコン開発プロジェクトの一

この研究開発の位置づけ



C'H

註:2010年度時点での計画

ISLIMソフトウェア研究開発報告会

建設

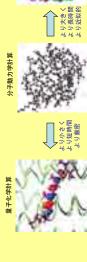
設計

研究棟

分子スケール研究開発

合化する技術を開発、タンパク質や細胞の機能発現過程のシミュレーションにつなげ、細胞スケールとの有機的な連携を図る。ターゲット・多剤排出トランスポータ、脂肪酸代謝酵素反応 開発目標: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

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 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 するためのプログラムを開発する。



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

ISLIMソフトウェア研究開発報告会

2011.1.23時点

2008年度に設定したチーム間連携課題

ISLAM

細胞から高次 機能までの一 買 変性・イオンチ 神経可塑性. 定(モデル化) 分析,高速化 **笛晒の 救**形 - 4年給 脳神経系 ケイプ 複数細胞から 臓器へ・血栓 形成 臓器全身スケ ール 画像処理パラ メータ推定 分析,高速化 運動制御 職器制御 循環器系 白栓形成 生体分子ネットワークモデル 膜での輸送モ デル **維胞スケール** 成長·変形モ デル 分析,高速化 肝細胞 ISLIMソフトウェア研究開発報告会 データ提供・ド ラッグデリバリ 可塑性や神経 変性に関する データ 代謝反応・膜 での物質輸送 分子スケール -夕同化技 代謝・膜 卜怎 データ解析融合 神経細胞デー 夕·局所回路デ 生体分子ネットワークデータ タンパク質-リ ガンド 医療画像デー タ 分析,高速化 肺がんと薬 データ解析融 合 臓器全身スケ ール 分子スケール **御覧スケール** 闵 脳神経系 高度化 提供

9

Œ



御胞スケール研究開



包括的な実証データを組み込んだ肝細胞・肝小葉のシミュレーションの実現を目指す。薬 そに :ペタフロップス級の計算力を活かして細胞を100万のボクセルに区画、 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。

<u>開発計画:</u>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。



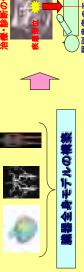
程式による物質移動をシミュレーションできる細胞シミュレーションプラットフォームを開発 E-CELL3をベースにした代謝反応がオルガネラ間も含めてシミュレーションできるようにたった。 相本 1004世別 までの性能 テストを終了 2011.1.23時点 。 進捗状況:細胞を約3万個のボクセルに分割、複数のオルガネラを表現し、移流拡散方 なった.現在、1024並列までの性能テストを終了

全身スケール研究 器



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 : ペタフロップス級の計算量を用いて、数時間で実行可能な臓器全身シミュレー 開発目標 つなげる。

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 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 期発計画:血管網、各種臓器、全身を3次元的に再現した臓器全身モデルを構築し、 射線・集東超音波など)につなげる。

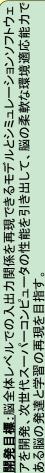


超音波伝搬シミュレ 音波を用いた悪性腫瘍の治 **固別患者のモデルを作成**

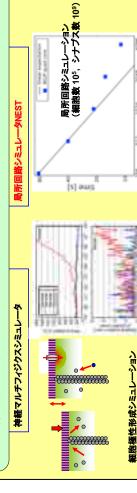
進捗状況:1mm分解能の全身ボクセルモデルを開発すると同時に新しい構造流体連成手法と多媒質内の超音波伝搬解析手法を開発した。これらにより、外科手術シミュレーションや収束超音波・重イオンビームによるガンの治療等へつなげる基盤を構築した。 2041 4 23 **在 5** 2011.1.23時点

ISLIMソフトウェア研究開発報告会

慰神絳系研究開発



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句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



細胞の極性形成について、分子レベルからのシミュレーションができた。また、局所回路シミュレータNESTでは105個の神経細胞、109個のシナプスからなる大規模局所回路のシミュレーションができた。 並列性能に関しては32000並列までのテストを終了 2011.1.23時点 **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ックスシミュレータでは神経 きた。並列性能に関しては32000並列までのテストを終了

C

夕解析融合研究開発

讣

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。

:ペタフロップス級の計算によって、創薬ターゲット探索や個人差を考慮した医

達成目標

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。 **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

ドる医療開発イノベーション 予測とエラーの発見 病気のシステム的理解に 仮説の自動生材 (関連たんぱく質の時系列データ) 関連するデータをモデルに同化 モデル 癌などの病気に関わっているモデル)

ネットワーク探索及びタンパク質構造に基づく相互作用予測の研究を行い、新たな技術開発の成功と新たな並列プログラムの開発により、これらのプログラムが8000を超えるcore **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 2011.1.23時点 ISLiMソフトウェア研究開発報告会 で稼動するようになった。



生命体基盤ソフトウェア開発・高度化

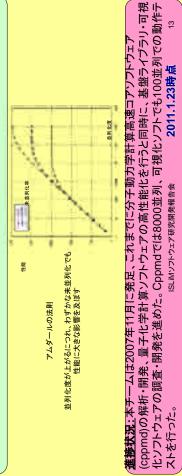
C'

開発ソフトウェア(全34本)

C.

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

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

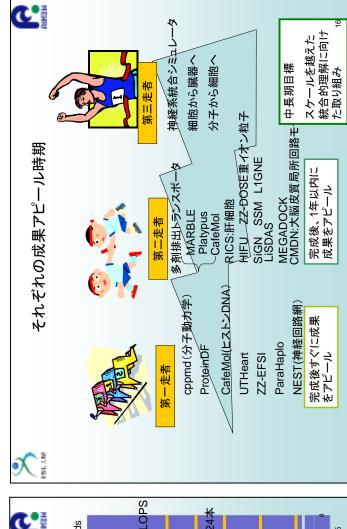


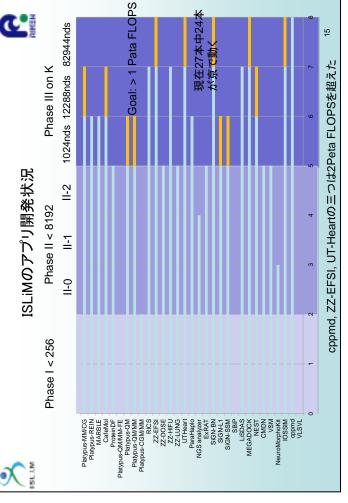
フェア名称	說明		ソフトウェア名称	1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1
ns-MM/CG	マルチコピー・マルチスケール分子シミュレー・シェン・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・		ParaHaplo	ハプロタイプ関連解析に於ける統計検定を行うた めのソフトウェア
	ノコノスにおり、別国になるノンスノーノンフ		NGS analyzer	次世代シークエンス解析プログラム
us-REIN	レプリカ交換分子動力学計算インターフェイス	ì	ExRAT	拡張RAT法による2SNP組合せの全ゲノム関連解析ソフトウェア
LE	全原子分子動力学計算			- 川が月代日本1 - 一日が月代日本1 - 一日が月代日本1 - 一日が月代日本1 - 一日が月代日本1 - 一日が月代日本1 - 一日が日本1 - 一日本1 - 日本1 - 一日本1 - 一日本1 - 日本1 - 一日本1 - 日本1 - 一日本1 - 日本1 -
ol	粗視化モデル計算		SIGN-BN (SIGN)	入規模退伍十制御ネットノーク推定ノロクフム
1DF	密度汎関数法に基づくタンパク質全電子波動 関数計算	群だ唱	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
us-QM/MM-FE	/ イブリッドQM/MM反応自由エネルギー計算	₫ ⟨ □(o	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
ns-QM	量子化学計算	n)		ネット・フーク指 圧ノロク フム
us-QM/MM	量子化学計算/分子動力学計算		SBiP	データ解析融合プラットフォーム
ns-CGM/MM	粗視化モデル計算/分子動力学計算		Lisdas	生命体データ同化プログラム
	雒 汚シミュワーションプラシャフォーム		MEGADOCK	網羅的タンパク質ドッキング解析プログラム
SI	全身ボクセルシミュレーション(ボクセル構造流体連成軽析プログラム)		NEST	Neural Simulation Tool
		溋		or tical microcil cut. Developed of 14E3 1
Ö	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)	神経(5	WSA	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ブラットフォーム)
SE	重粒子線治療シミュレーション	·)	NeuroMorphoKit	体統
	低侵襲治療シミュワーション		IOSSIM	昆虫嗅覚系全脳シミュレータ
	(ボクセル超音波伝播プログラム)		pwddo	大規模並列用MDコアプログラム
NG	肝呼吸・肺循環シミュフーション			分散並列大規模データ可視化システム
t	マルチスケール・マルチフィジックスで職シミュ	(4	SPHERE	アプリケーションミドルウェア
	フーション	.)	NLSVL .	大規模仮想化合物ライブラリ
	ISLiMソフトウェア研究開発報告会	研究開	発報告会	<u>t</u>

ZZ-EF ZZ-RB(ZZ-DO 댎) Т Неа

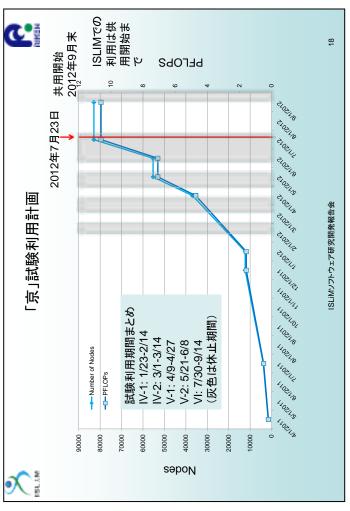
(8)草字器類

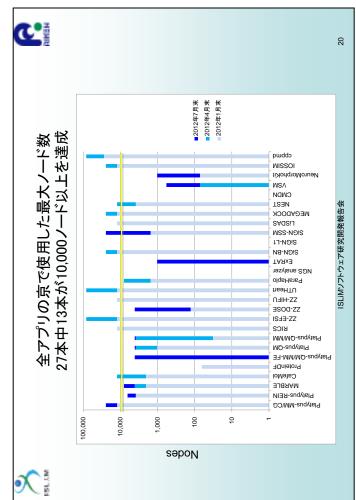
RICS

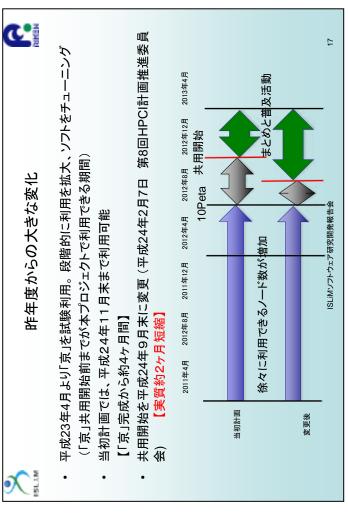


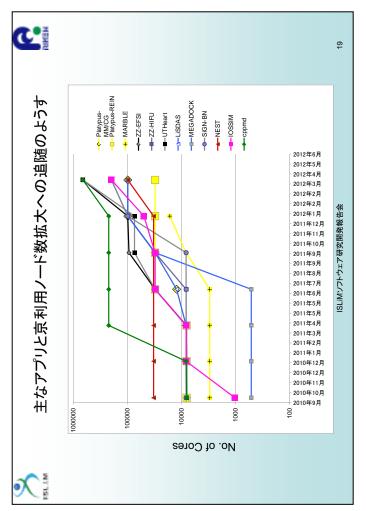


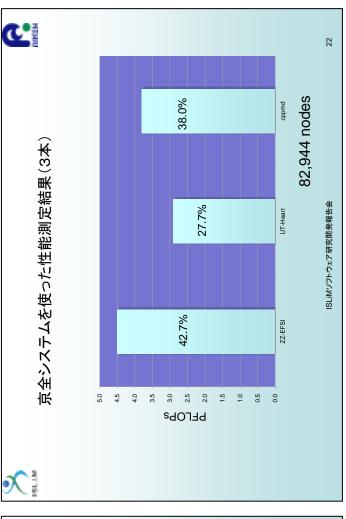
10

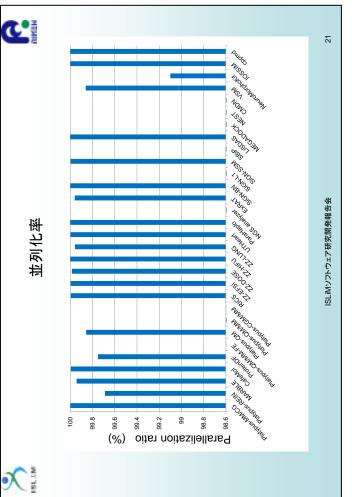


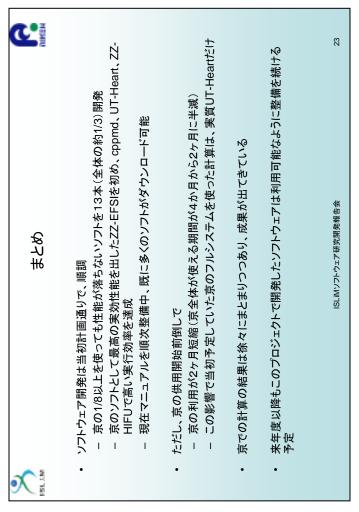


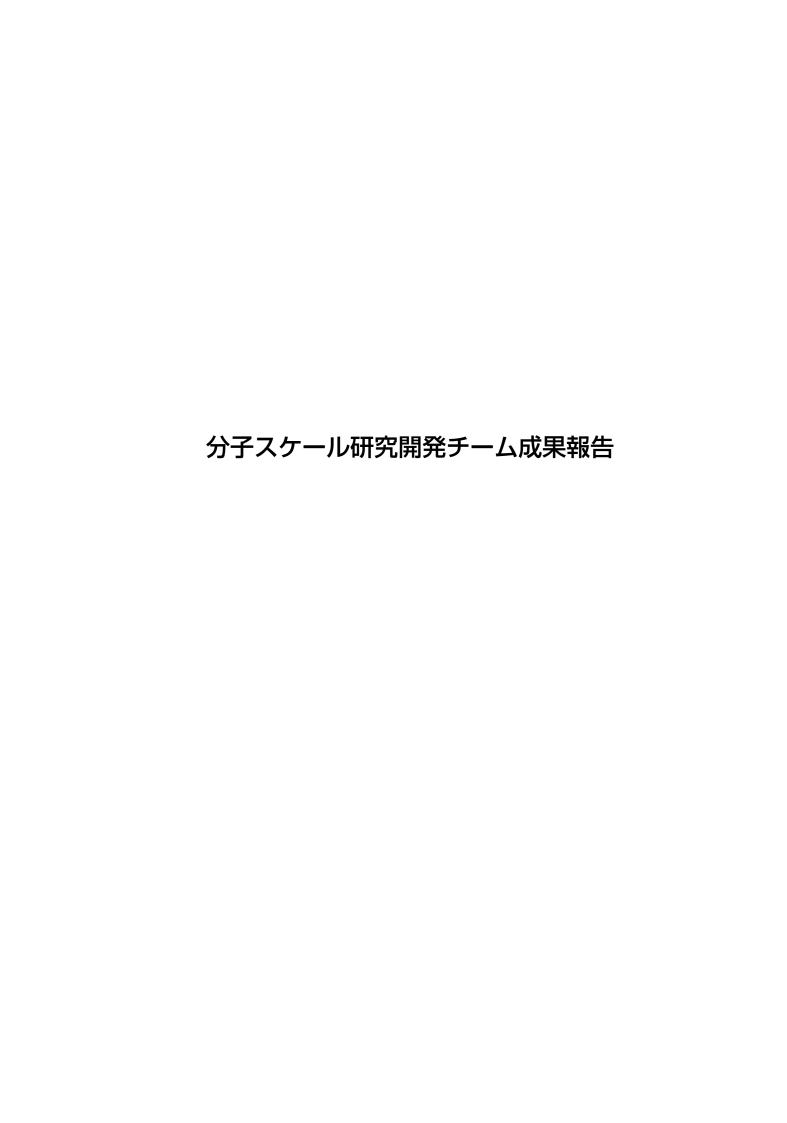












分子スケールチーム

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、そ 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、脂質代謝多 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 った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、通常の分子動力学計算 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

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)

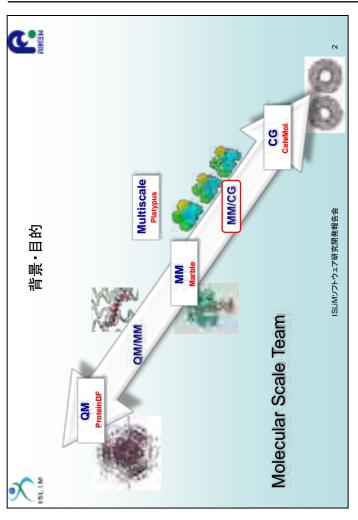
Platypus-MM/CG

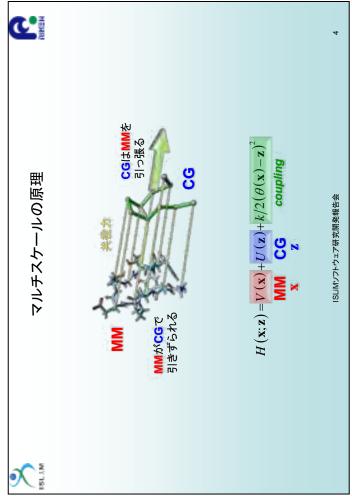
多剤排出トランスポーターの構造変化経路計算 脂質β酸化多酵素複合体の構造変化計算

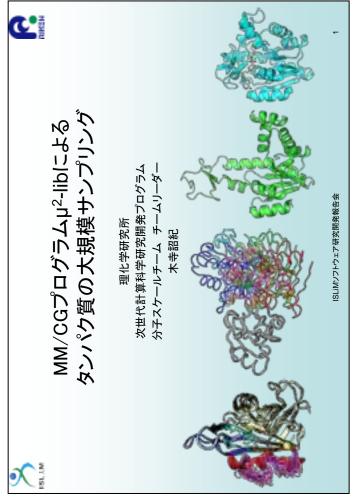
CafeMol

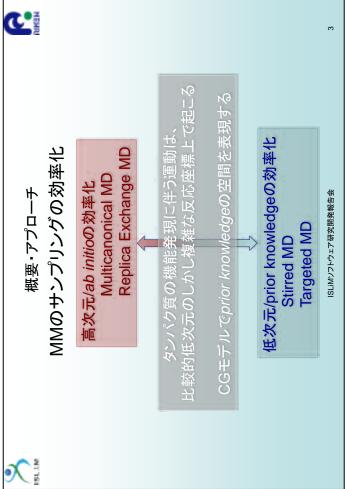
ヌクレオソーム (DNA ヒストン複合体) 動態の シミュレーション

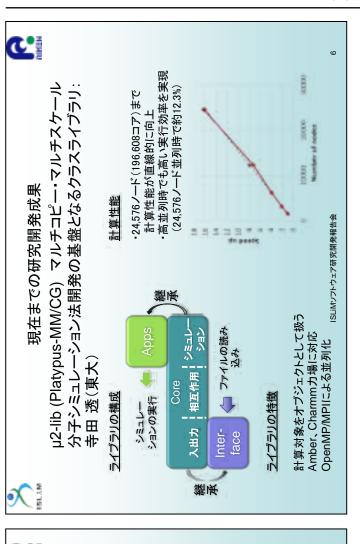
以上

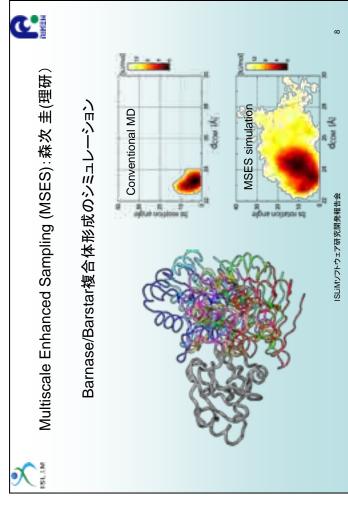


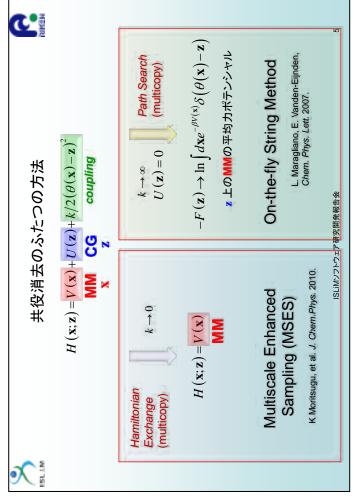


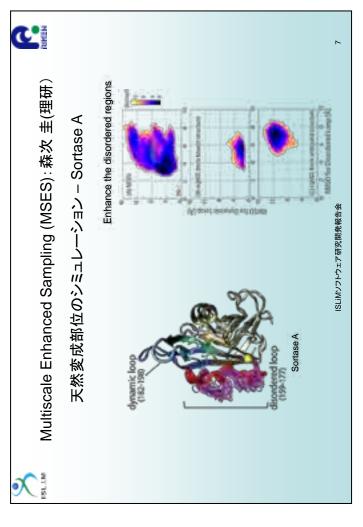


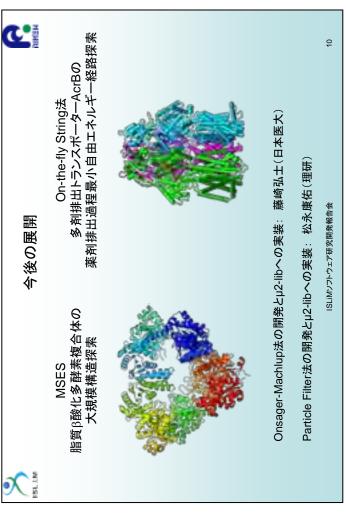


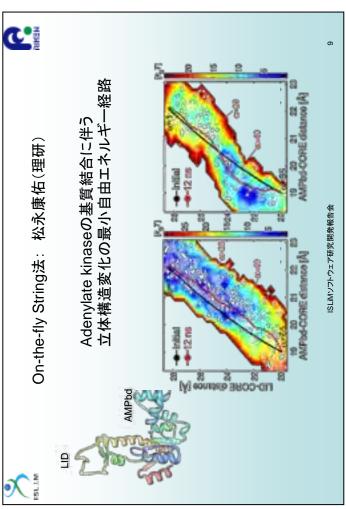












Replica-exchange interface program: REIN

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

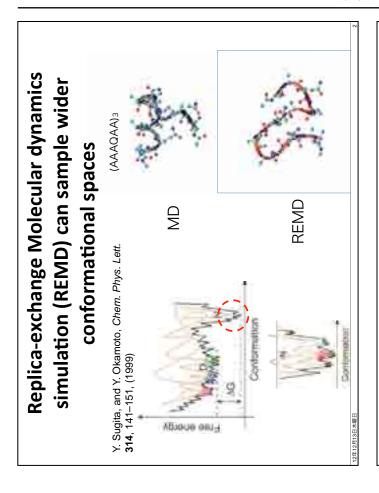
2011 年 4 月 理化学研究所 生命システム研究センター 研究員

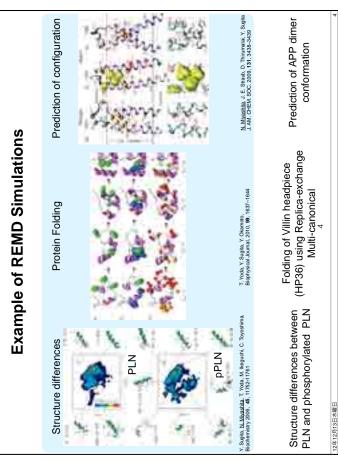
2011 年 4 月 理化学研究所 次世代計算科学研究開発プログラム 兼務

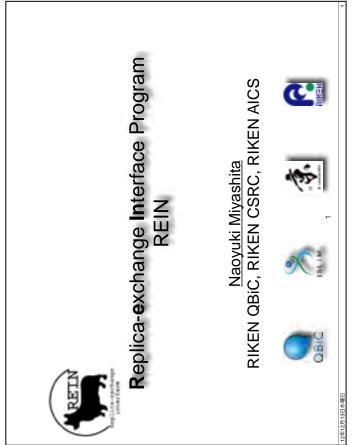
2011 年 4 月 理化学研究所 計算科学研究機構 兼務

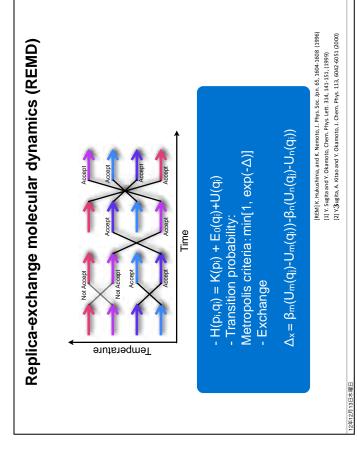
研究分野

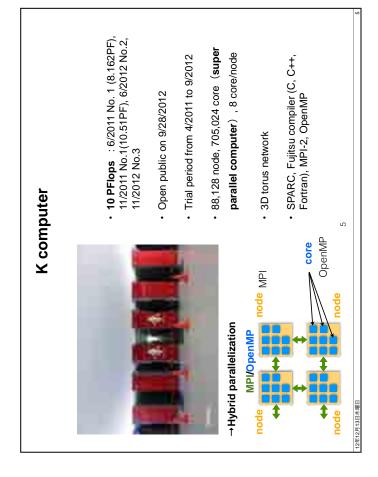
物理(理論)・生物物理・分子シミュレーション











Replica-exchange molecular dynamics (REMD)	olecu	ılar dyna	mics (R	EMD)	
Moraphi Accept A	• Sire	Replica-exchange molecular dynamics simulation [1] has been used widely as enhanced sampling method.	ge molecular as been used iing method	Replica-exchange molecular dynamics simulation [1] has been used widely as useful enhanced sampling method.	
Accept	• Be	tter Method f o	or hybrid pa	Better Method for hybrid parallelization	
Time		Layer	REMD	Parallel	
		ဇ	REM	MPI	
PEM (MPI) / MD (MPI/OpenMP)		2	MD	MPI	
replica replica		-	MD	OpenMP	
ON ON	• 2	plica-exchang : Not only tem	ge umbrella nperature bu	Replica-exchange umbrella sampling (REUS) [2]: Not only temperature but also distance,	
Duce	an	angle, dihedral and so on.	and so on.		
Bital	∑	Multi-dimensional REMD (MREM) [2]	al REMD (M	IREM) [2]	
node	[REM] K. P [1] Y. Sugi	[REM] K. Hukushima, and K. Nemoto, J. Phys. Soc. Ipn. 65, 1604-1608 (1996) [1] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. 314, 141-151, (1999)	.o, J. Phys. Soc. Jpn. 65, n. Phys. Lett. 314, 141-1	, 1604-1608 (1996) 151, (1999)	
Temperature	[2] Y. G ugi	[2] Y.&ugita, A. Kitao and Y. Okamoto, J. Chem. Phys. 113, 6042-6051 (2000)	to, J. Chem. Phys. 113,	, 6042-6051 (2000)	
12年12月13日木曜日					9

~			on	×	×
ethoc		lations.	MREM	0	0
nge m	MD program	REMD simu	Recompile of MD	ou	Necessary
a-excha	MD	rograms for F	Modification of Recompile MD program of MD	ou	Necessary
interface software for Replica-exchange method	Control	Interface software can use several conventional MD programs for REMD simulations.	style	binary run (fork, exec function)	Plugin style (It needs recompile with patch)
tware fo	Interface Software	e several conv	Language/ parallelization	Perl/ TCP socket	C/ (depend on each MD)
ace sof	Interface	tware can use	MD	CHARMM AMBER NAMD	GROWACS NAMD DLPOLY AMBER ACEMD LAMMPS QUANTUM- ESPRESSO CPMD
interf		Interface sof	Packages	aarex.pl in MMTSB toolsets [1]	PLUMED [2]

Ā

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

(Native)

×

C/MPI/OpenMP

MARBLE

TINKER

Fortran/MPI

×

×

× × No perfect MREM package on K

[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

[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

Advantage:

[1] M. Feig et al., Journal of Molecular Graphics and Modelling 22 (2004) 377-395 • [2] M. Boromi, D. Bradlustrid, E. Bussi, C. Camilloni, D. Povasti, P. Raiteri, D. Donadio, F. Marninelli, F. Petrucci, R.A. Brogila 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

The several functions on each MD package

Easy to perform REMD or MREM (simple input file)

Not support on K

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

Conventional MD packages and REMD module

e X

REMD 0 0 0 0 0 × ×

Hybrid

Language/ parallelization

MD Packages

×

0

Slow

× × 0

Fortran/MPI Fortran/MPI

Ξ 2 3 4 2 [9] ⊵

CHARMM

AMBER

 $\stackrel{*}{\triangleleft}$ ×

> 0 0

Fast Fast Fast Slow Fast

Fortran/MPI/pthread

0

C/MPI/OpenMP C/MPI/pthread

GROMACS DESMOND

NAMD

0 × 0

×

Medium

 \triangleleft

rank3

rankl

SPAWN version

レプリカ交換インターフェースプログラムの コンセプト

●外部プログラム(既存のMDプログラム)を用いる

Interface programである

(外部プログラムの修正はしない)

●外部プログラムはBinary形式である

- ●MPIのルールの範囲
- ●PCクラスタでも京コンピュータでも動く
- ●プログラム実行が非常に簡単である

12年12月13日木曜日

MREM Recompile Modification Control style Language/ parallelization NAMD Ð REIN version 1.0 Packages

will be opened on next week

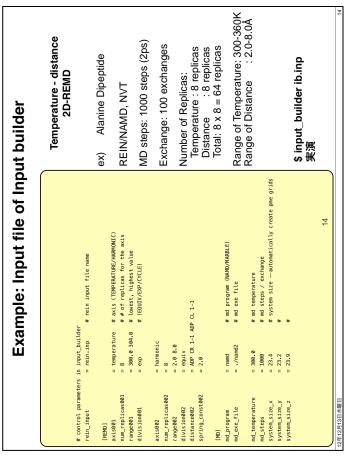
Fortran (Intel, gfortran, fujitsu) MPI-2 (Open MPI, Fujitsu MPI) K computer (sparc), FX10 (sparc), PC cluster (x86_64) REIN: Specification

Current available functions of REIN (MREM) Distance NAMD MARBLE Axis

rankl

rankl

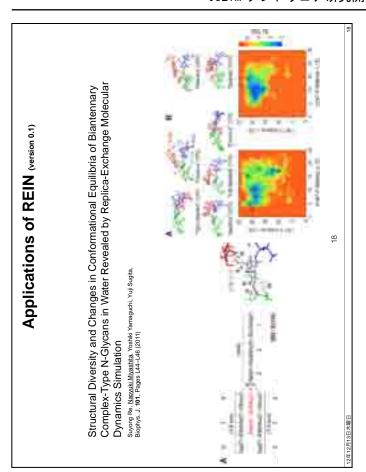
(script 言語での通信をサポートしていない) MPI/TCP socket by perl and python K computer fork/exe functions (複製禁止) system functionでmpiを起動 mpi_comm_spawn (MPI-2) Not supports: Prohibited: Allowed: 2年12月13日木曜日

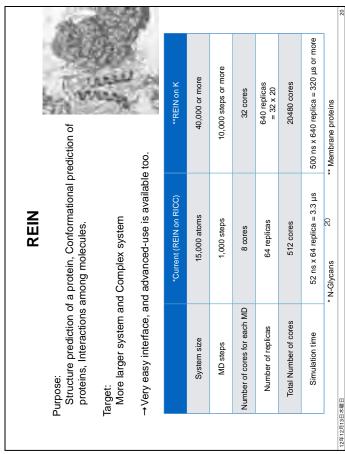


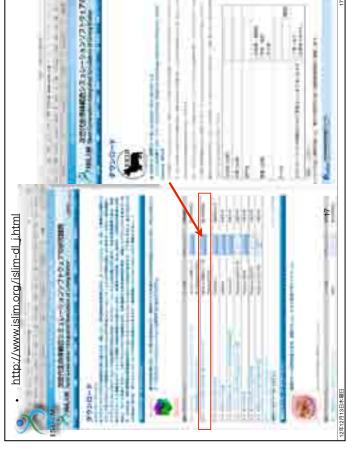
Replica#, condition#, Temperature, (Distance etc), Total, Kinetic, Potential, Extra					
#	EXCHANGE_STEP =	TEP = 1			
п	н	290.14150	-1624.33470	370.15570	-1994.49040
7	73	309.04970	-1677.37840	394.27840	-2071.65680
٣	٣	314.56290	-1658.91060	401.31190	-2060.22250
4	4	313.83160	-1657.91540	400.37900	-2058.29450
ľ	S	304.91160	-1642.47860	388.99900	-2031.47760
9	9	309.84760	-1635.80170	395.29630	-2031.09800
7	7	299.44280	-1685.76920	382.02210	-2067.79130
00	œ	288.62440	-1683.47140	368.22020	-2051.69170
#	= EXCHANGE_STEP	STEP = 2			
п	2	306.76840	-1664.28540	391.36800	-2055.65340
2	П	317.18100	-1651.98420	404.65210	-2056.63630
m	4	310.83750	-1629.92330	396.55920	-2026.48240
4	ю	284.34980	-1655.03960	362.76680	-2017.80640
S	9	294.21380	-1680.86470	375.35110	-2056.21580
9	ıs	336.64260	-1606.88210	429.48070	-2036.36280
7	00	296.55280	-1698.13390	378.33510	-2076.46900
					0000000000

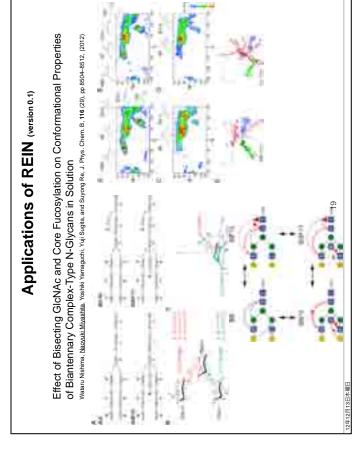
16

K (fx10) or SGE(UGE) type batch Batch_builder create batch script \$ batch_builder bb.inp
\$ batch_builder bb.inp restart Exchange steps: 100 exchanges for submission the job to job REIN: 32 core (4 node) Total node: 66 node Thread: 1 (flat MPI) script can be selected. Example: input file of Batch builder scheduler. rein_converter → creating the trajectories series REGULAR. replica_exchange = yes # ([YES]/NO): replica exchange / umbrella sampling num_of_exchange_steps = 1900 # number of exchange steps of core in cpu, for k=8, for fx10 = 16, Very easy execution -- three steps # batch filename
job scheduler type ([K]/GE/RICC)
queue name, k:([SMqLL]/LARGE),fx10: 15 13 # number of threads in md
of mpi process for md
cpu time # REIN input file name Restart #batch_builder bb.inp restart 1) Input_builder > rein.inp
#input_builder ib.inp 2) Batch_builder → rein.sh
#batch_builder bb.inp #rein_converter rc.inp 3) submitting of job
pjsub rein.sh # control parameters in batch_builder = rein.inp = rein.sh = 01:00:00 small num_mpiproc_md cpu_time [BATCH] batchfile_name job_scheduler num_threads_md ndo_core_cpu rein_input queue_name









Dr. Yuji Sugita, RIKEN ASI, RIKEN QBIC, RIKEN AICS 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 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 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

PC cluster, K computer, FX10 super Three supporting softwares Fortran95/MPI/OpenMP computer etc.. rein converter Input builder Batch builder Summary 2 Temperature, Distance, Angle, Dihedral Replica-Exchange Interface Program (REIN) currently NAMD2, near future MARBLE v0.6 REIN version 1.01 open (islim web page) Interface program of Multi-dimensional E-mail to : yukimya+reinmiya@riken.jp Replica-exchange molecular dynamics (REMD) simulation http://www.islim.org/islim-dl_j.html Free software (GPL version 3) Binary MD • • •

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



全原子分子動力学計算ソフトウェア MARBLEの開発と

横浜市立大学 大学院生命ナノシステム研究科

多剤排出トランスポーターAcrBへの応用

分子スケール研究開発チーム 光口 強衙



ISLIMソフトウェア研究開発報告会

背景·日的



- 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、<u>動的である</u>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<u>動的構造をシミュレーションする</u> 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

- 「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
- その機能メカニズムを解明
- 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、

ISLiMソフトウェア研究開発報告会



研究・開発のステップチャート

出力:座標の履歴



全原子MD ソフトウェア

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

- 水素まで含んだ全原子モデルを対象とする.

対象モデル

膜タンパク質の計算時には、膜分子も原子レベルで扱う

用いない。

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

ソフトウェア群

計算対象 生体分子

AcrBの全原子分子動力学シ / 分子スケールチームで共 多剤排出トランスポーター

✓構造変化が機能に重要 **/膜タンパク**質 通のターゲット

/ゆらぎと構造変化

超並列化ソフトウェア MARBLE

/ 立体構造ゆらぎ / 立体構造変化

/ 自由エネルギ-/ 実験との対応

先端アルゴリズムも装備

NMR, SAXSなど

生体超分子用全原子

概要・アプローチ

- 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

- 近距離力: 空間分割により,近接ノード通信を用いる

- 空間分割を用いて,通信量の削減を図る

相互作用計算

並列化アプローチ

シミュレーション技術の開発

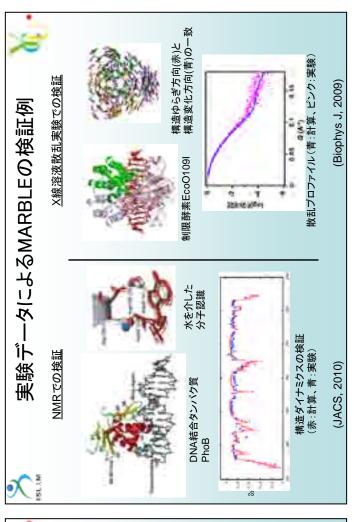
シミュレーション結果の解析 技術の開発

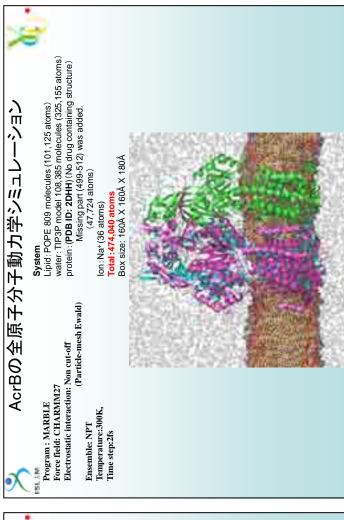
世界標準アルゴリズムを装 備(PME等)

(シンプレクティック圏体法等)

30

ISLAM





16394 32788

100A 3048 4096 819E

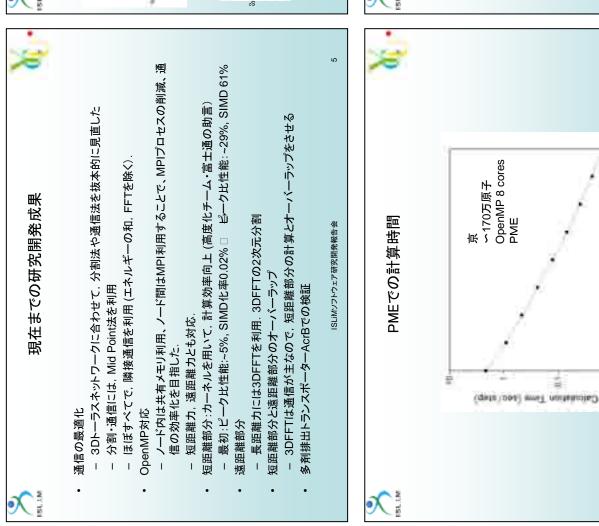
ž

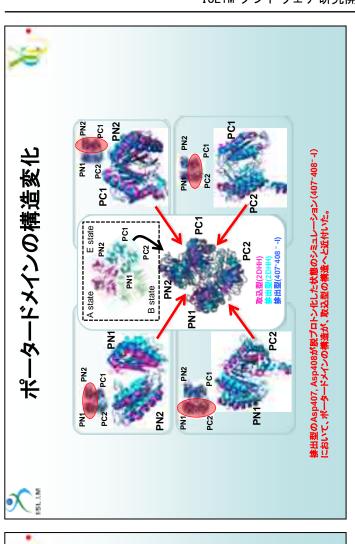
ä

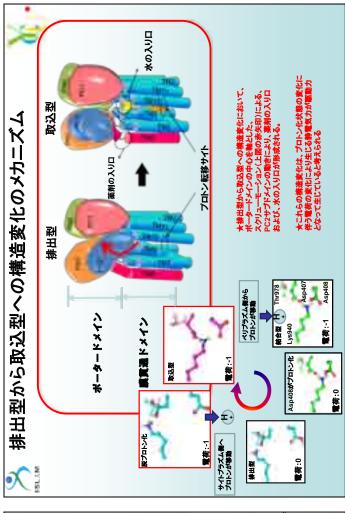
900

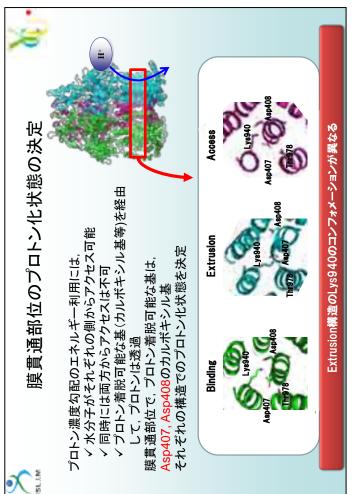
Number of Occess

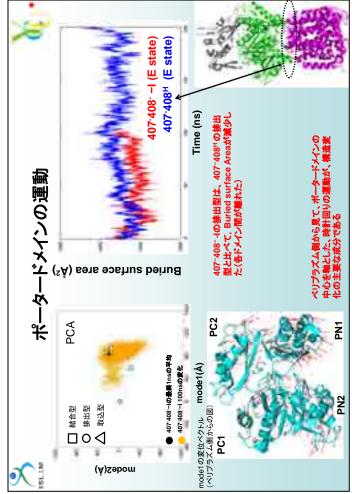
ISLIMソフトウェア研究開発報告会











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

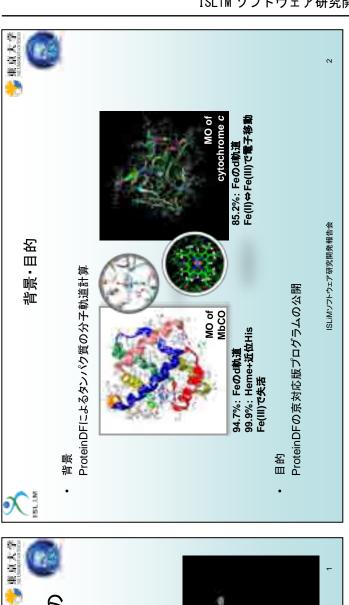
2002 年 11 月 東京大学生産技術研究所客員助教授

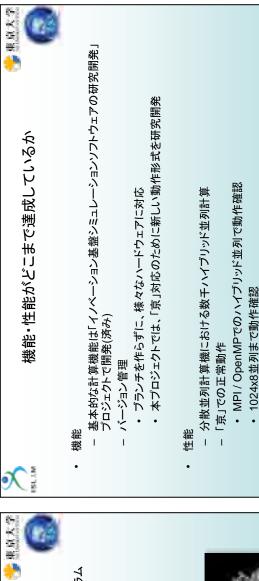
2004年 12月 東京大学生産技術研究所助教授

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







大規模SCF計算を成功させるための高精度初期値作成法 ISLIMソフトウェア研究開発報告会 光合成反応中心タンパク質の分子軌道 ・アクセサリーバクテリオクロロフィル •Pから4Å以内のアミノ酸残基 ・バクテリオフィオフィチン エネルギー勾配計算 ・スペンャンペア

ISLIMソフトウェア研究開発報告会

- 「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

- 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

ProteinDFとは

EST YM

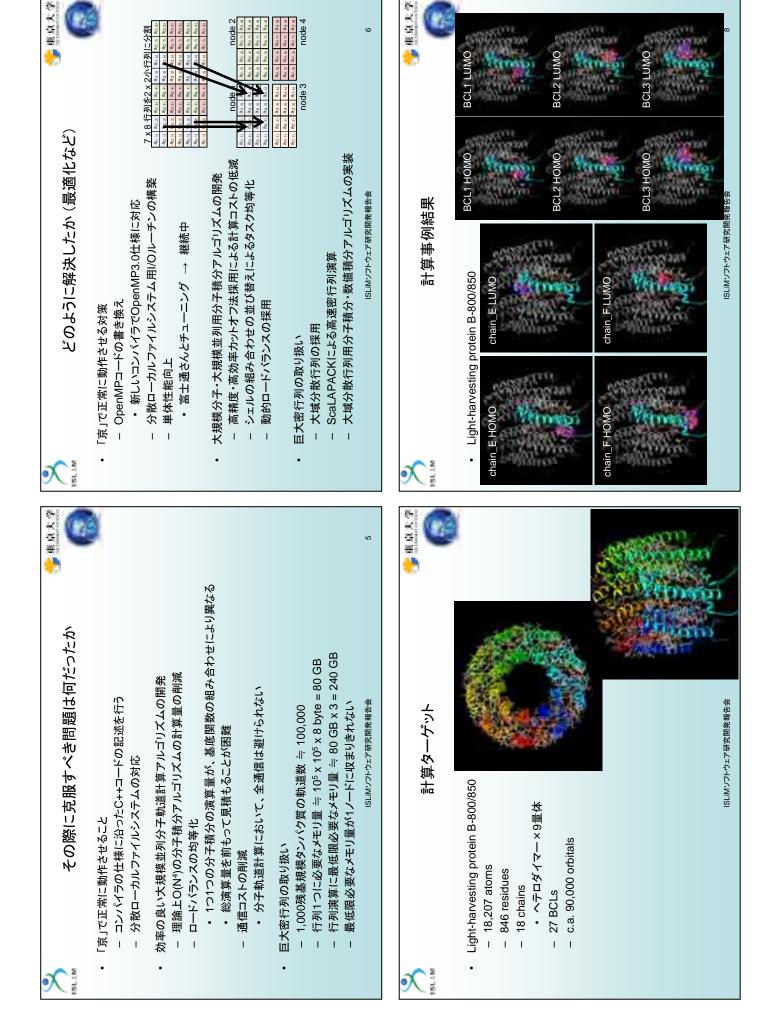
概要・アプローチ

分子積分にRI法やRT法アルゴリズムなどが選択可能

HF, DFT(含ハイブリッド汎関数)

主な機能と特徴

- ガウス型基底関数





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

高田彰二 京都大学大学院 理学研究科生物科学専攻



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

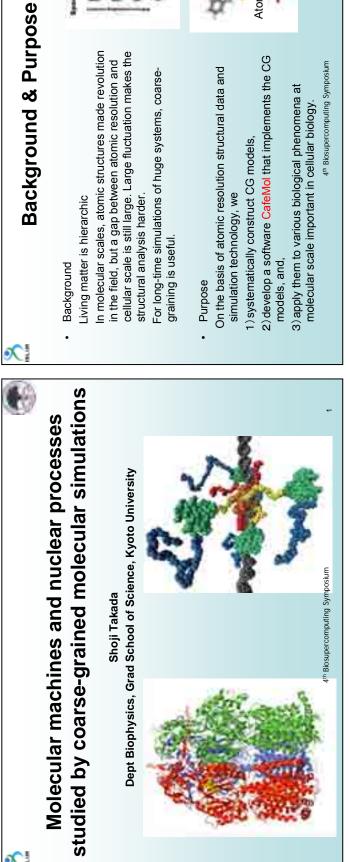
1991年3月 岡崎国立共同研究機構分子科学研究所技官

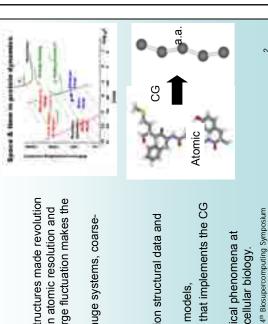
1998 年 4 月 神戸大学理学部化学科講師·助教授

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





CafeMol: Current status

- Completed

Current status of CafeMol

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

Multiscale algorithm for CG parameter derivation from atomic level Accurate derivation of and more accuracy in CG models →

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



CafeMol

Approach

By CG models, it simulate long-time dynamics of huge biomolecular systems.

Summary

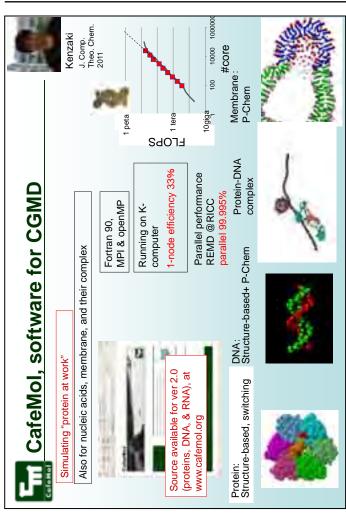
- Clasical molecular dynamics with CG model

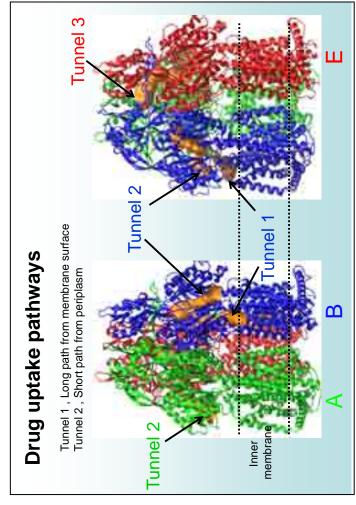
- Simulations
- Time propagation based on Langevin dynamics and others
- Neighbor list, replica exchange Parallel programming
- Computer language and library - Fortran90, MPI, OpenMP
- Code release
- Freely downloadable from http://www.cafemol.org

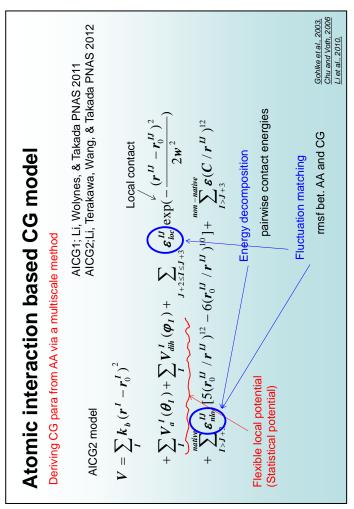
Simulations that correspond to milliseconds and longer.

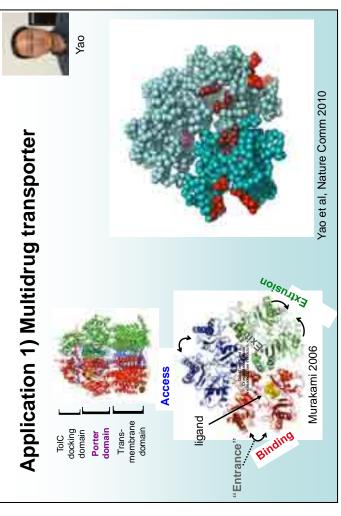
4th Biosupercomputing Symposium

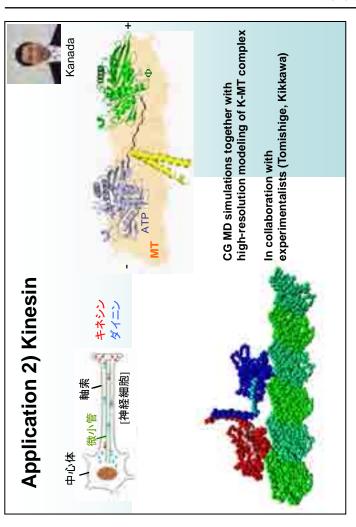
38

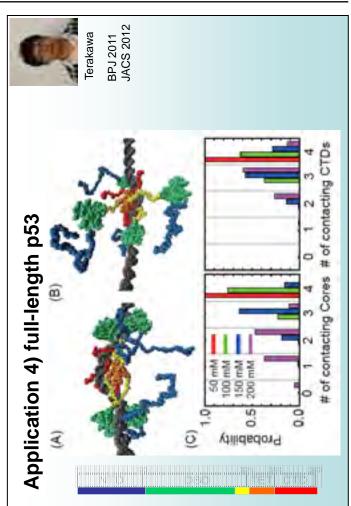


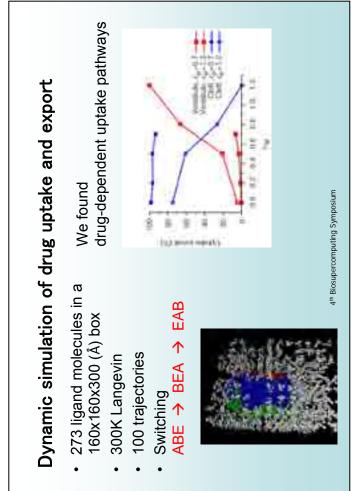


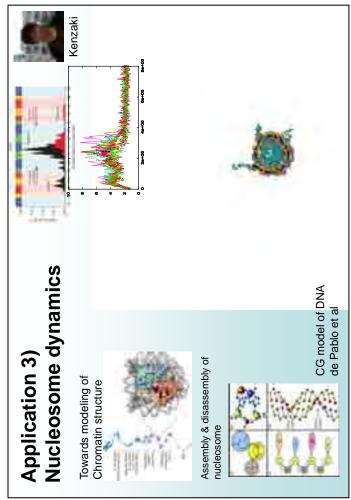












酵素分子の柔らかさと触媒活性: ハイブリッド QM/MM 自由エネルギー法の開発

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

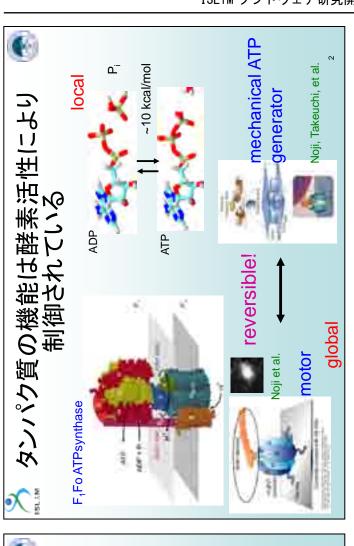
2003年6月 京都大学福井謙一記念研究センター 研究フェロー (博士研究員)

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



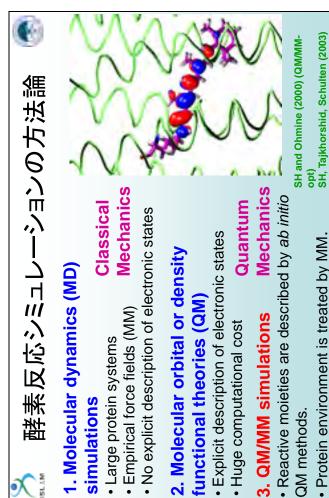
ハイブリッド QM/MM 自由エネル ギー法の開発

大学院理学研究科 京都大学

林 重彦

芦

酵素分子の柔らかさと触媒活性



excited state QM/MM-MD)

ISLIMソフトウェア研究開発報告会

Protein environment is treated by MM.

ISLiMソフトウェア研究開発報告会

【QM と MM の計算時間の大きなギャップ】 **Computational cost** α-amylase QM 15 min / step QM>>>MM A long-time direct QM/MM-MD is difficult. Computational time/ step 100 days with 1 fs / step = 8.64 Ms 864 s = 14.4 m120k step / s 120 step / s SLiMソフトウェア研究開発報告会 12 step / s 0.86 s 8.6 s 86 s Trajectory time 1 ms 10 ps 1 µs 100 ps 1 ns 10 ns 100 ns QM/MM MM-MD Anton GPU

Geometry optimization of the QM molecule on a free energy MD simulation for protein conformational sampling surface defined with the MM conformational thermal 解決のための基本的アイデア QM/MM calculation of a chemical reaction and QM/MM free energy (FE) geometry optimization distribution obtained by MD simulations Converged at 600 ns ISLIMソフトウェア研究開発報告会 Complete separation between





Kosugi and SH, JCTC (2012)

Problem of MF-QM/MM

Conventional

Present method

QM/MM RWFE-SCF Method

QM/MM Long-Range Interaction

Conventional: sphere cluster

 Conventional: hard to implement for **Ewald summation technique** continuum electron density

treatment (Hayashi and Ohmine, 2000) makes Present method: RESP operator its implementation possible.

Yang et al.

 $\rho_{MM}(\mathbf{X})$

 $\rho_{QM}(\mathbf{d}; \mathbf{R})$

 $\rho_{MM}(\mathbf{X})$

 $\rho_{\text{OM}}(\mathbf{d};\mathbf{R})$

Σ

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

Much quicker convergence of

iterative MD sampling

MM distribution is reweighted.

and MD sampling

 $- \times \rho_{MM} (\mathbf{d}_{o}, \mathbf{R}_{o}; \mathbf{X})$

 $\left\{ \exp \left[-eta \left\{ \mathsf{E}^{\mathsf{OM}-\mathsf{MM}}\left(\mathbf{d},\mathbf{R};\mathbf{X}
ight) - \mathsf{E}^{\mathsf{OM}-\mathsf{MM}}\left(\mathbf{d}_{\mathsf{O}},\mathbf{R}_{\mathsf{O}};\mathbf{X}
ight)
ight\}
ight]
ight
angle$

 $\exp\left[-eta\left\{\mathsf{E}^{\mathrm{OM-MM}}(\mathbf{d},\mathbf{R};\mathbf{X})-\mathsf{E}^{\mathrm{OM-MM}}(\mathbf{d}_{\mathrm{o}},\mathbf{R}_{\mathrm{o}};\mathbf{X})\right\}
ight]$

 $\rho_{MM}(\mathbf{d},\mathbf{R};\mathbf{X}) = -$

Convergence problem!

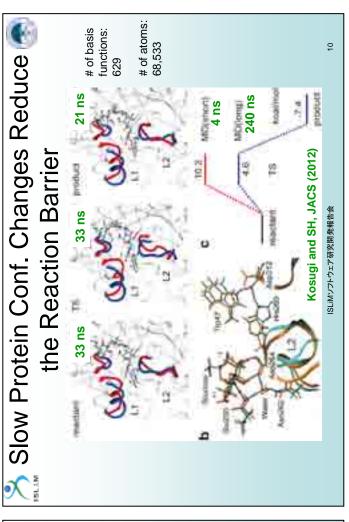
necessary.

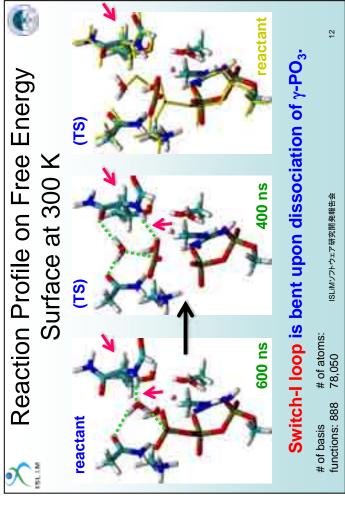


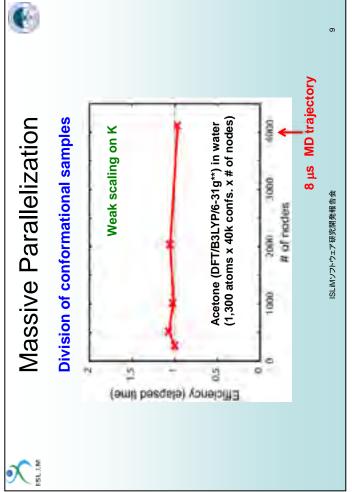
Present treatment: periodic boundary condition

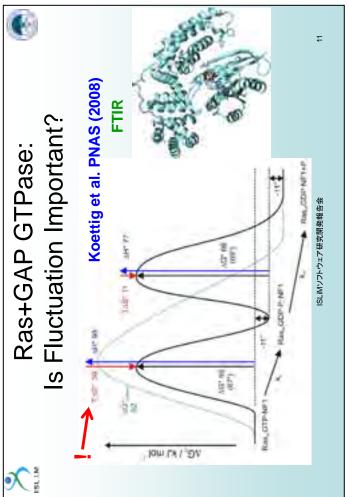
Complete separation of QM/MM

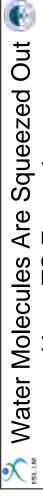
Frequent iteration of QM/MM opt. and MD sampling is

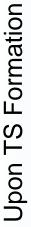








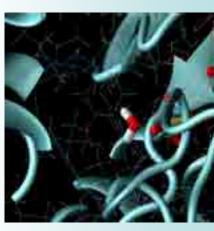












ISLiMソフトウェア研究開発報告会

13

まとめ



- 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

4



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

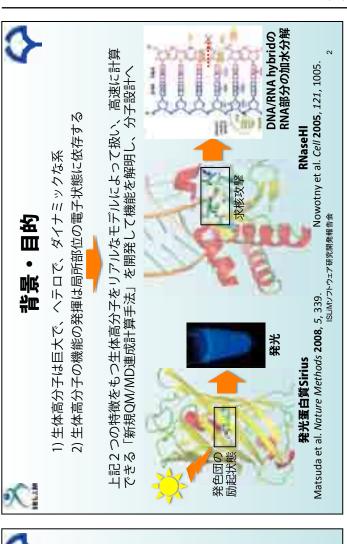
2004年4月 日本学術振興会特別研究員

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



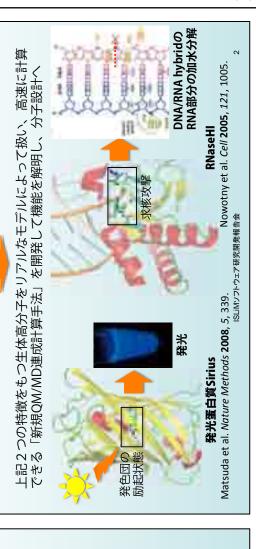
Platypus-QM/MMの開発と応用

附属蛋白質解析先端研究センター

助教 鷹野

裛

大阪大学蛋白質研究所





現在までの研究開発成果



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

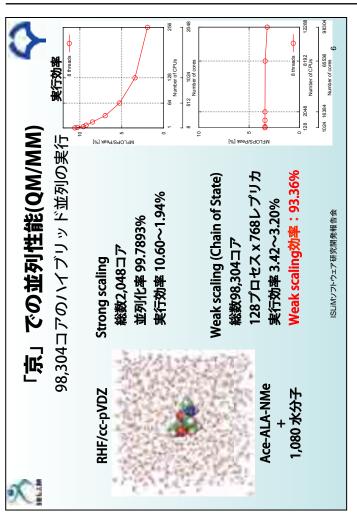
►QM/MM境界の高精度化

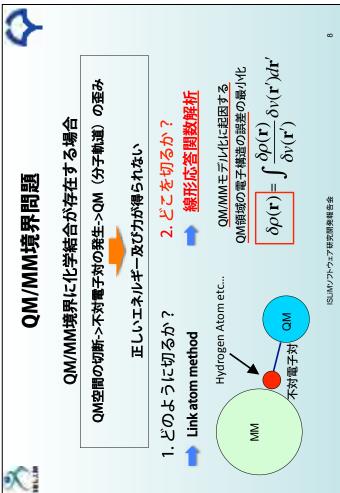
SLiMソフトウェア研究開発報告会

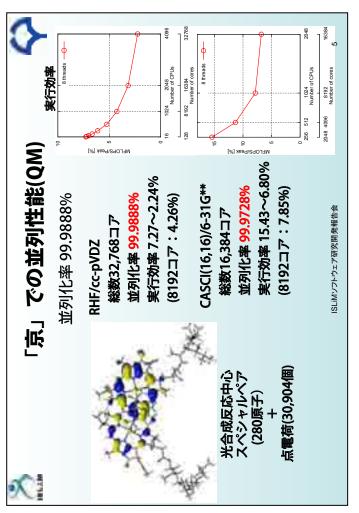
最小自由エネルギー経路 ISLiMソフトウェア研究開発報告会

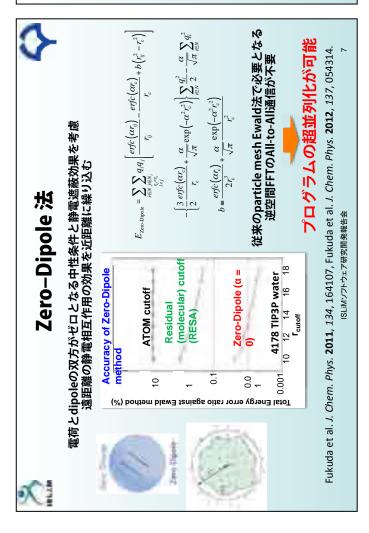


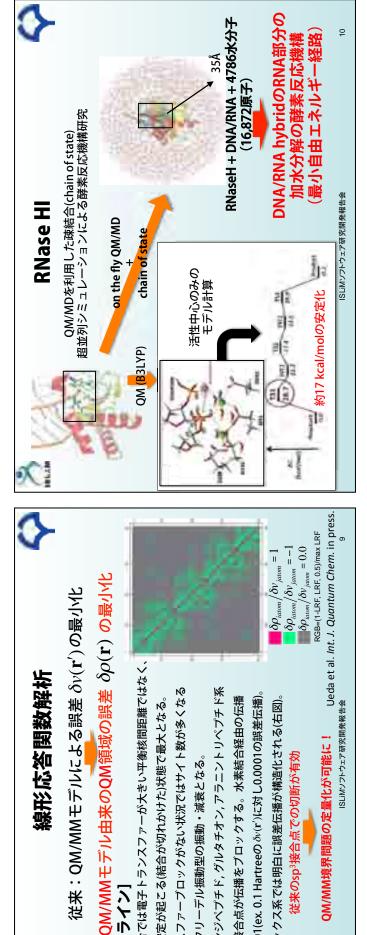
ISLiMソフトウェア研究開発報告会











(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、

[ガイドライン]

(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。

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。

(iii) アラニンジペプチド, グルタチオン, アラニントリペプチド系

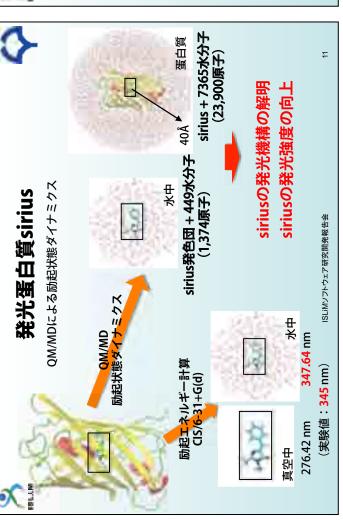
は約0.001(ex. 0.1 Hartreeの か(r)に対し0.0001の誤差伝播) ではsp3接合点が伝播をブロックする。水素結合経由の伝播

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

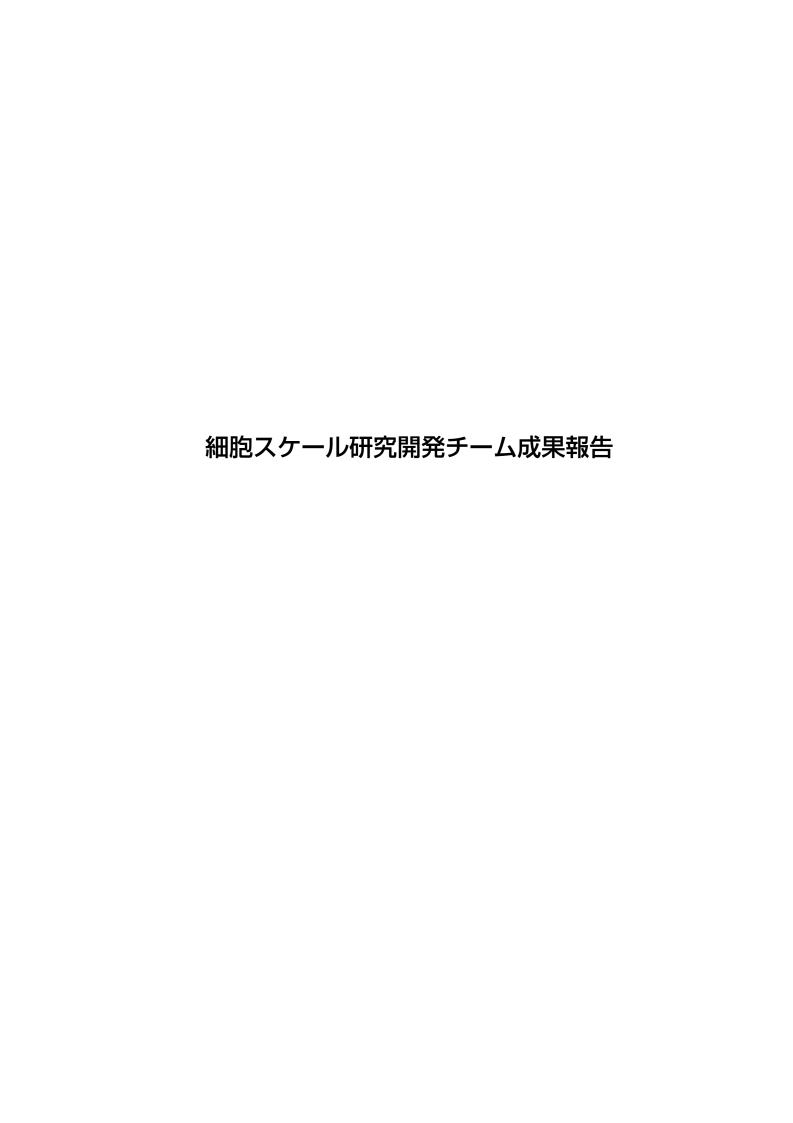
線形応答関数解析





ISLIMソフトウェア研究開発報告会

QM/MM境界問題の定量化が可能に!



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

バイオイメージング 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動させること 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エアの開発だけでは、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。これらの研究の成果により、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、では、細胞内や膜の場所を考慮することでは不可能であった。一方、細胞の中では、物質の濃度勾配やオルガネラキ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量・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クラスターから、「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 並列までのスケービラリティを確認 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)。

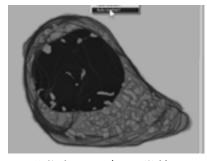


図1: 共焦点レーザー顕微鏡から RICS プリシステムを使用して再構築したヒト肝臓細胞の形状









図 2: RICS を用いて計算した細胞内 Ca²⁺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。

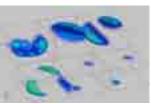
3.2 肝細胞シミュレーション

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、酸化的リン酸化、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、肝細胞索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 µm の血小板細胞表面の数十 nm2 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、活性化の中途にて濃染顆粒の放出が起 こり、放出された ADP による P2Y12 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。





3 D 血小板モデル 血小板内物質の局在 3.4 小胞動態モデル

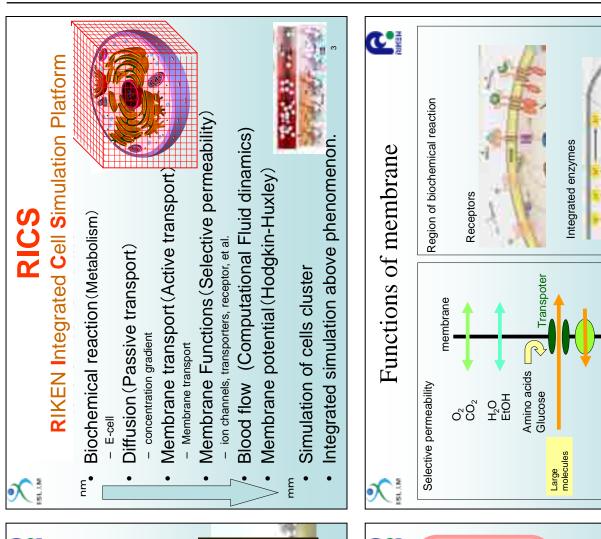
全反射蛍光顕微鏡による膵臓 β 細胞のインスリン類動態の画像解析から、グルコース刺激によるインスリン分泌では、restless newcomerが主様式であることが判明した。その動態ををにインスリン類粒が細胞膜に融合するが当期に融合するが判明した。1) 細胞膜には近いないは、細胞膜に接近する時間ではは地では、細胞膜に接近するでは、個点では、一次のでは、一次のでは、一次のでは、一次のでは、一次のでは、一次のででは、一次のででは、一次のででは、一次のででは、一次のででは、一次のででは、一次のでででいる。現在、このモデルをRICSに搭載するにとで小胞の3次元解析を計画している。

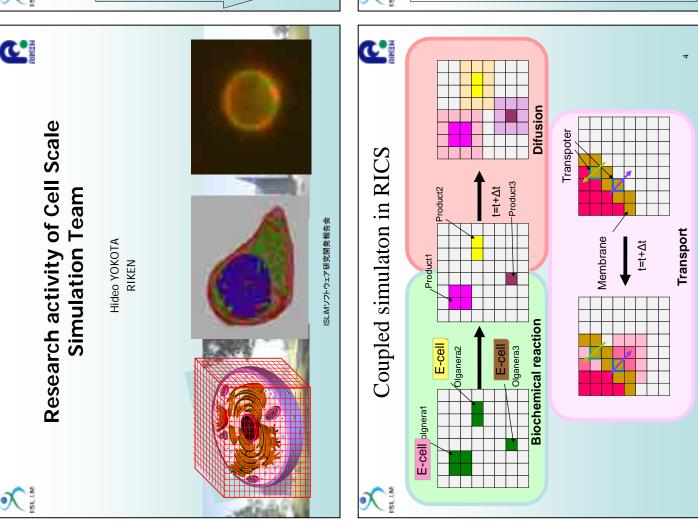
3. ソフト研究開発の達成目標

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 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 を整えることが必要である。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花開くことを期待する。

5. 謝辞

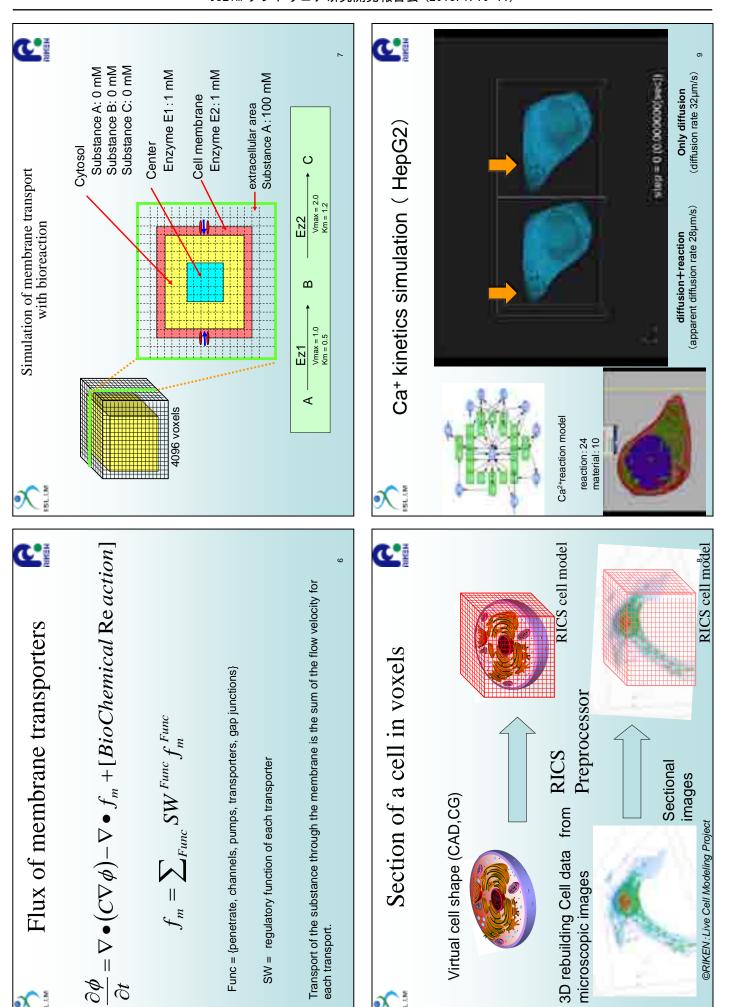
本発表の結果は、理化学研究所が実施している京速コンピュータの試験利用と理化学研究所情報基盤センター(RICC) RIKEN Integrated Cluster of Clustersの計算による。

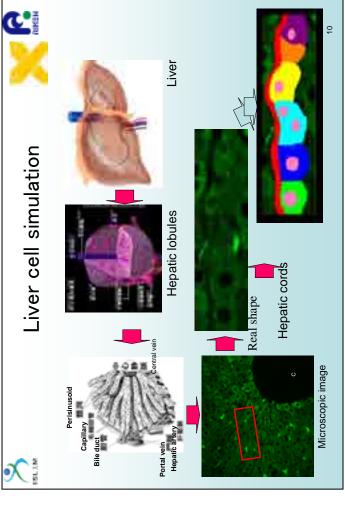


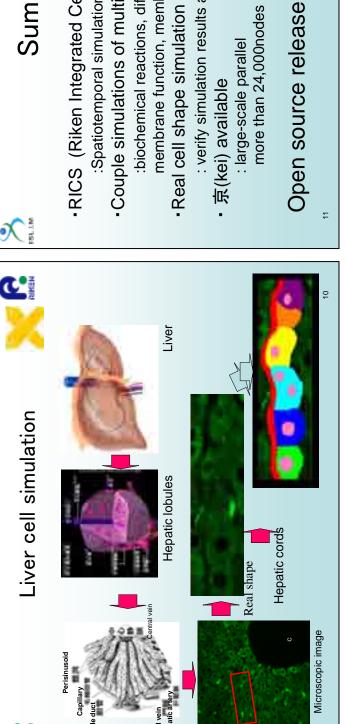


H⁺,Na⁺ Ca²⁺,Cl⁻

lons







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

Summary

C.



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

verify simulation results against actual observations

more than 24,000nodes : large-scale parallel

Open source release: December 2012

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



卣小板昝뿬シ≒ュワータ−の開発と自柱シ≒ュ|

ターへの 応用

学校法人東海大学

センター長 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 後藤信哉



臓器全身スケール



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

「<mark>血小板細胞</mark>」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。

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

メートルスケールの実証実験により精緻化できる。

「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

淵

背景·目的

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-2 mの血小板細胞数個を対象としたマイクロ

/עי

割を演じる。アスピリン、クロピドグレルなどのナノメー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。

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 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 シミュワ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「<mark>血小板細胞</mark>」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論的理解を目指す

ISLIMソフトウェア研究開発報告会

עצון

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

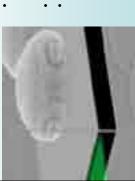
オルガネラの局在を組み込んだ「血小板細胞」形状モデルは確立できた。また、実証実 RICSにおいて、細胞内の現象と細胞外の現象を同時に再現するために、RICSにおい デルも確立できた。実証実験を反映する「<mark>血小板細胞</mark>」接着基盤モデルも構成できた。 験により得られた精緻な生物学的情報により精緻化した「血小板細胞」活性化基盤モ て移流計算を行なうべく工夫を行なっている。特殊な条件において計算は可能である が、生物学的に妥当な条件では明らかな計算エラーがでる。

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、「<mark>血小板細胞</mark>」形状モデルを

基盤プログラムとしてRicken Cell Simulator Platform (RICS)を用いる

アプローチ

作成する



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

維胞膜| GPIbα | 維胞内オルガネラ

RICS上に表現した由小板 |白小板モデル| 「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

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 RICSIte-cellを基盤とした、細胞内の代謝、活性化シグナルの サブルーチンは血小板細胞には最適化されていない。

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 移流計算のサブルーチンの組み込みにはさらに工夫が必要

ウムイオン濃度変化を示す



「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

SLiMソフトウェア研究開発報告会

ISLIMソフトウェア研究開発報告会

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あった。それか ら7年を経過した現在、次世代シークエンサーの 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、超高次元大規模へ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

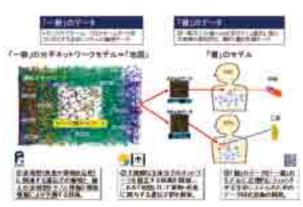


図 1

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

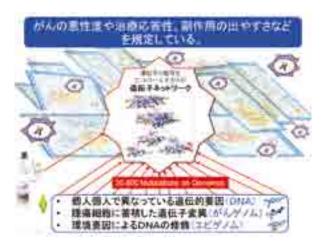


図 2

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。ベイジアン 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「最適ベイジアンネットワーク 推定ソフトウェア」では、32 ノードの最適ベイジ 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。

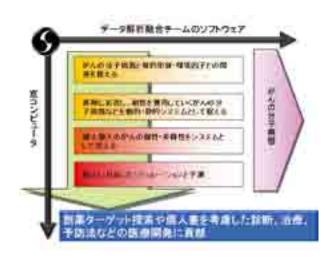


図3



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を 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 というソフトウエァを開発した。 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かめて 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 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、次世代シークエ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

(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、次世代シークエ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有り、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

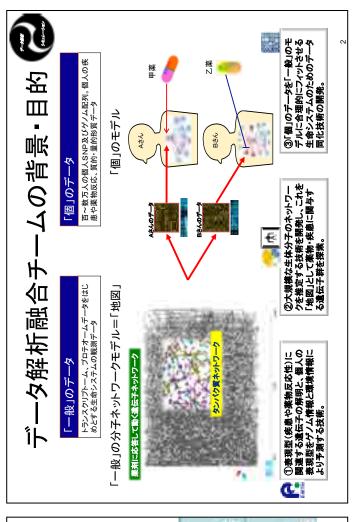
とで、苦難と膨大な努力の末、SBiP から SiGN 関係のソフトウェアを京コンピュータで実行し その結果をもどし、人手では実行不可能な手順の大規模データ解析(遺伝子ネットワーク解析を中心)ができるものになりつつある。ドキュメンテーションもきちんとしており、オープンソースですでに公開している。

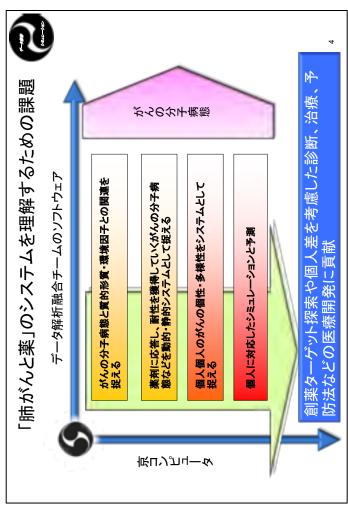
3. プロジェクト終了時の達成目標

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。

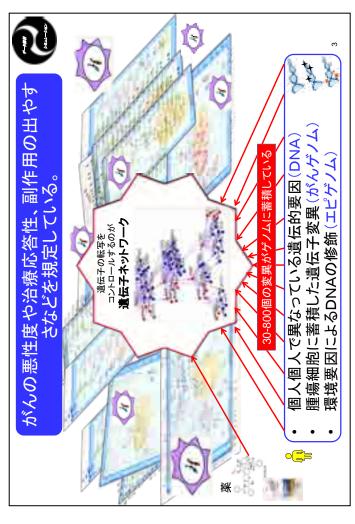
2012 年になり、半導体チップで DNA を読む 革新的なシークエンサーの実用化された。さらに、 2013 年には、高集積半導体チップがリリースさ れ、誰もが自分の全 DNA 情報を 10 万円程度で 丸ごと手にいれられることが確実になった。さら に、1万円ゲノムを実現するナノポア半導体シー クエンサー技術も実用化されている。これまで主 に生物や病気の「研究」のために行っていたシー クエンスに対し、がんや患者さんの全 DNA 情報 や遺伝子発現情報を得て、臨床的に翻訳・解釈し、 治療として「患者さんに戻す」臨床シークエンス の研究が米国やカナダでは既に始まり、国を上げ て個別化医療を推し進めている。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(シークエンス、及び定量 データ) もこの次世代のシークエンサーで解析で 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、これが次世代シークエンサーによる 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。

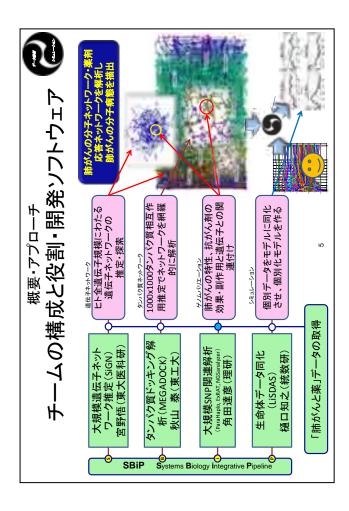














▶大規模遺伝子ネットワーク推定とその応用

大規模遺伝子ネットワーク推定ソフ

ウェアSiGNの開発目的

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

■業務参加者

アゲノム 解 柞 センター

斉藤あゆむ(技術職員) 関谷弥生(教務補佐員) 新井田厚司(特任助教・本委託事業による雇用) 島村徹平(助教)

井元清哉(准教授) 山口 類(講師)

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

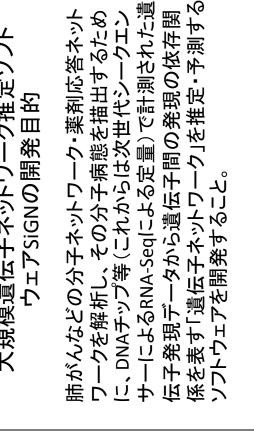


SiGNで、たとえば、次のようなことが可能になってきた

ISLIMソフトウェア研究開発報告会

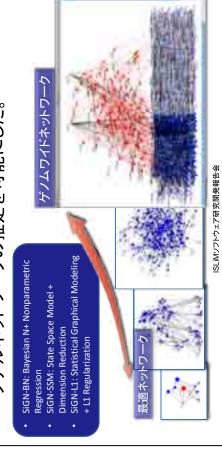
- 抗がん剤に応答して動的に変化する大規模 な遺伝子ネットワークを描出すること
 - がんの予後・再発を予測する遺伝子群を ネットワーク解析から同定
- 数百のがんサンプルの遺伝子発現プロファ イルデータから個々のがんの個性・多様性 発・生存・死、抗がん剤が効く人・効かない を分子ネットワークとして抽出すること(再 人の違い)
 - これらを地図として用いて、新たな分子標 的探索の可能性がでてきた

ISLiMソフトウェア研究開発報告会



SiGNの達成したこと

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



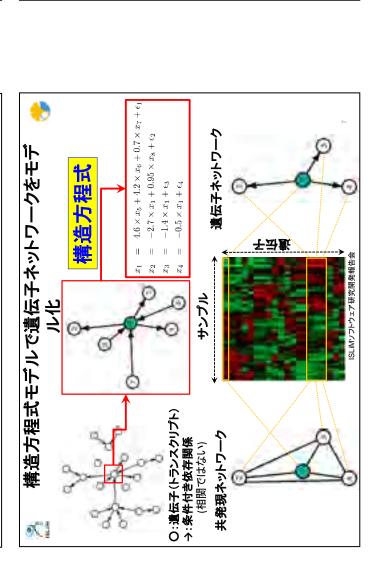
Sign-L1

eX.

- 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(12,288/一ド)
- 実行効率:1コアで11.98%を達成。
- ハイブリッド掛列化に対応。

ISLIMソフトウェア研究開発報告会



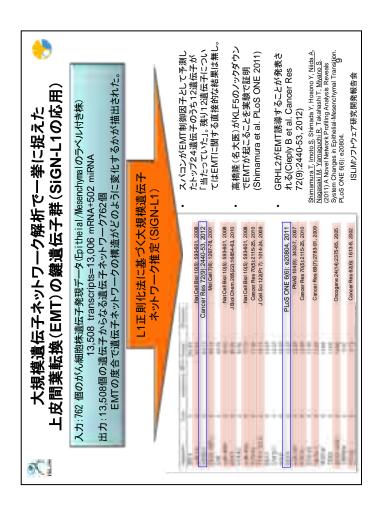


モジュレータ i-th sample

I ≥ ISLiMソフトウェア研究開発報告会

ネットワーク

ISLiMソフトウェア研究開発報告会



全遺伝子規模のベイジアンネットワーク推定 を実現

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)

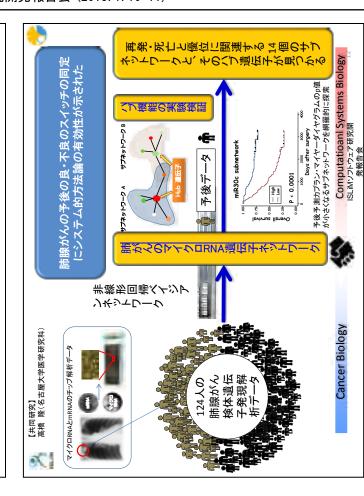
高効率並列ブートストラップアルゴリズム

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)

ジアンキットロ

方法:ベイジアンネットワーク+ 非線形 回帰

SIGN-BN



肺がん

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

m.

映し出された「再発」のスイッチを入れていた 肺がんのマイクロRNA遺伝子ネットワークに

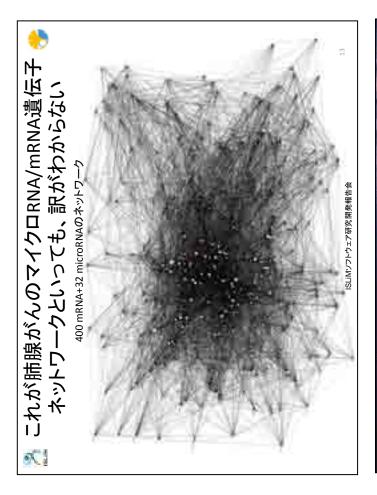
黒幕ネットワーク達

SiGN-BNが炙りだした

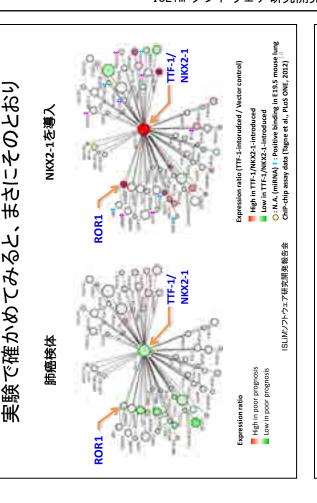
【共同研究】 高橋 隆(名古屋大学医学研究科)

ISLiMソフトウェア研究開発報告会

がん死亡第1位 2010年69,813人







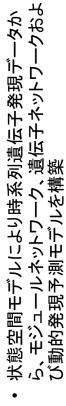
NKX2-1 CROR1

- 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



0



- 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 口
- 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)

SLiMソフトウェア研究開発報告会

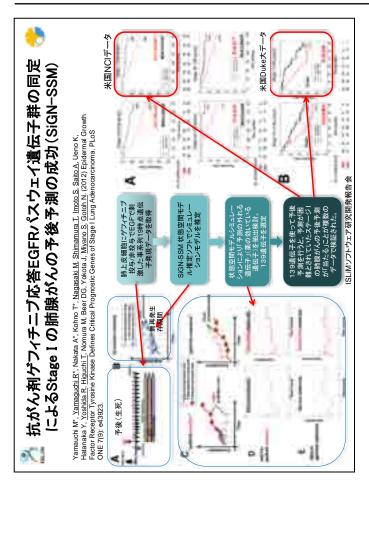


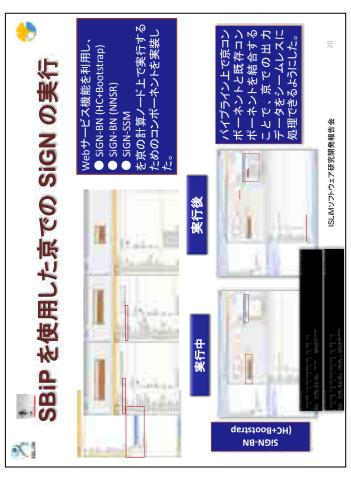
SBiP

Systems Biology integrative Pipeline

- Signを解析 ペイプレインのコンポーキントとしてユー 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
- 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。 河
- ポーネントを組み合わせてカスタマイズした解析フ ユーザはSBiPに用意されているさまざまな解析コン 一を実行可能。

ISLIMソフトウェア研究開発報告会





00

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

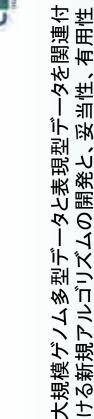
1993 年 International Joint Conference on Neural Networks Student Award

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





の検討

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 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

(NGS analyzer)するなどの新たな方法論を導入する必要がある. これらには, スー たセットから拡張し、次世代シークエンサーにより全ゲノム配列への探索を行ったり

パーコンピューティングを駆使することが必要不可欠になる。

効であり、われわれ理化学研究所ゲノム医科学研究センターでは、2002年に世界に先

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 駆け初めてのGWASを実現してから,世界を牽引してきた. この方法論をさらに劇的に

背景·目的

ゲノム医科学研究センター グラープディフクター 理化学研究所 角田 達彦

ISLIMソフトウェア研究開発報告会

概要・アプローチ









開発ソフトウェア ParaHaplo について

NGS analyzer: 次世代シークエ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

SLiMソフトウェア研究開発報告会

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

- 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

四8:

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.



富士通コンパイラ(RICC, FX1,「京」, の全て)対応済, 実行成功 PCクラスタで8,000ノードの並列性能を実証 ParaHaplo ver. 4.0 β, phase IV, 50% 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 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 こ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

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

研究開発コードの概要

遺伝的原因を探るための統計検定計算

- 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: 次世代シークエ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
- 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 で実装したマルコフ連鎖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タンス 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

ISLIMソフトウェア研究開発報告会

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.

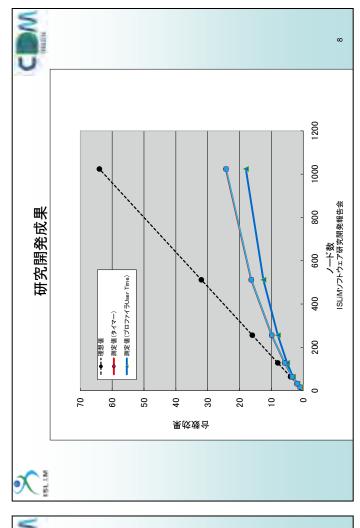
現在は、一層の高速化を目指しSIMD化を行っている. - 「京」ではコア並列(ハイブリッド並列)も実装している.

- ソースコード公開済

- 「京」でも並列性能が出ることを確認

	研究開発成果	Was D	
Dataset	WTCCC type 2 diabetes	betes	
Sample size	1,924 cases + 2,938 controls	38 controls	
# of SNPs	365,947 (autosomes)	es)	
# of haplotype blocks	42,226		
# of genarations (max)	291,200 (chromosome 22)	ome 22)	
Cores used (max)	64,000		
Disk space (stage in) / run	300 TB		
Memory used / run	560 MB		
Total time (Elapsed)	> 1 WGek ISLIMソフトウェア研究開発報告会	Ŋ	

研究開発成果 WTCCC Type 2 diabetes data on chr22. Chr 22 with 884,000 generations. Chr 22 with 884,000 generations. Chr 22 with 884,000 generations. Haplotype block that has haplotypes whose frequencies are significantly different (p < 1.1 × 10-6) between type 2 diabetes case and control in WTCCC. Haplotype blocks with no significant difference.	NOTES OF		 9.00 9.00 9.00 9.00 9.00 9.00 9.00	9
Chr 22 with different (p × Haplotype b	研究開発成果	TCCC Type 2 diabetes data on chr22.		ISLIMソフトウェア研究開発報告会
	SL.IM	>	Chr 22 with different (p	



€ Yes

研究開発成果

開発ソフトウェア ExRAT について

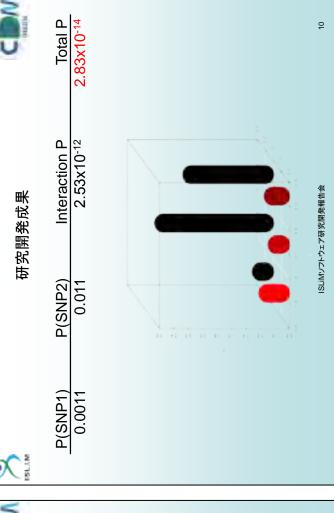
- ExRAT ver. 1.0- r440, phase III-0, 100%
- 富士通コンパイラ(RICC, FX1,「京」, の全て)対応済, 実行成功, buildも可能.
- PCクラスタで8,192ノードの並列性能を実証.
- データ: 患者708人, 一般集団3397人, 8314SNP, 14100回permutationで評価.
- ハイブリッド並列化対応: Preprocess処理とpermutation処理にOpenMPでのスレッド化の実装.
- Preprocess処理の見直しによるさらなる高速化:分割表探索処理での計算処理の 見直しにより、約40%程度高速化(計算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:非同期通信用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
- 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。

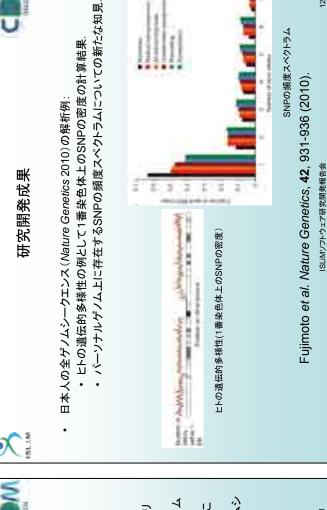
最小とすることで対処した。

- WTCCCの実データを解析し、評価と結果を出した.

ISLiMソフトウェア研究開発報告会









開発ソフトウェア NGS analyzer について

- NGS analyzer ver. 1.0, phase II-1, 80%
 - PCクラスタ上で2000並列を達成.
- 全ゲノムシークエンスデータの解析
- ヒトの全ゲノムシークエンス解析からの一塩基多型(SNP), 挿入・欠失(indel), J アレンジメントを高い精度で検出
- がん細胞における突然変異を高い精度で検出(東京大学医科学研究所ヒトゲノム 変異や多型のタイプごとに検出アルゴリズムを実装した。さらに、実験的検証を行うこ 解析センターのPCクラスターにて実行)
- 初の日本人の全ゲノムシークエンス(Nature Genetics 2010)、肝臓がんの全ゲノムシ とで、パラメーターのチューニングを行った。 **一クエンス (Nature 2010)を解析**
- Fujimoto et al. Nature Genetics, 42, 931-936 (2010).
- The International Cancer Genome Consortium. Nature, 464, 993-998 (2010).

SLiMソフトウェア研究開発報告会

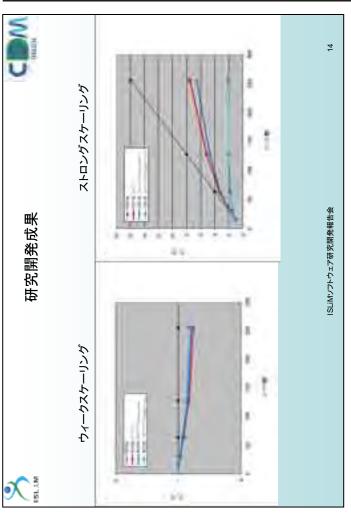
12

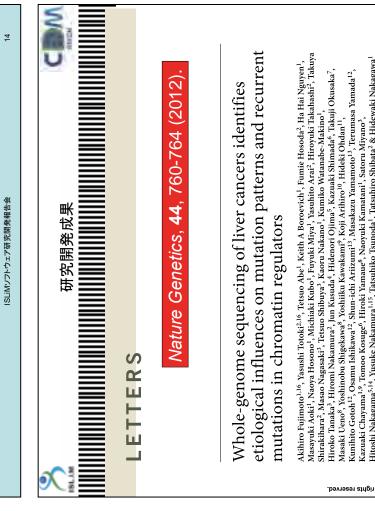
~17h. (24,576 cores used)

41 min. 17 sec.

Time / run (average)

Total time



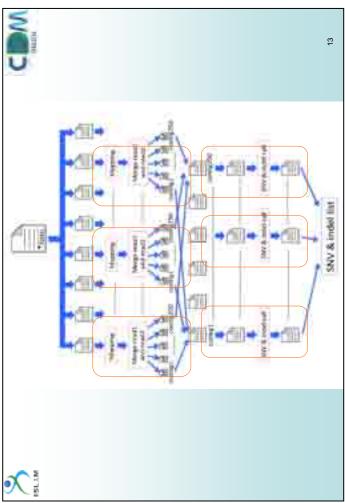


ARTICLES

研究開発成果

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

nature genetics



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 single nucleotide variations (SNVs). Comparison with six previously reported genomes revealed an excess of singleton nonsens: and nonsynonymous SNVs, as well as singleton SNVs in conserved non-coding regions. We also identified 5,319 deletions Masao Nagasaki³, Rui Yamaguchi³, Tetsuo Shibuya³, Michiaki Kubo¹, Satoru Miyano²²³, Yusuke Nakamura¹¹³ & Akihiro Fujimoto^{1,2}, Hidewaki Nakagawa¹, Naoya Hosono¹, Kaoru Nakano¹, Tetsuo Abe¹, Keith A Boroevich¹, We report the analysis of a Japanese male using high-throughput sequencing to ×40 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 analysis of a Japanese individual using massively parallel sequencing Fatsuhiko Tsunoda^{1,2}

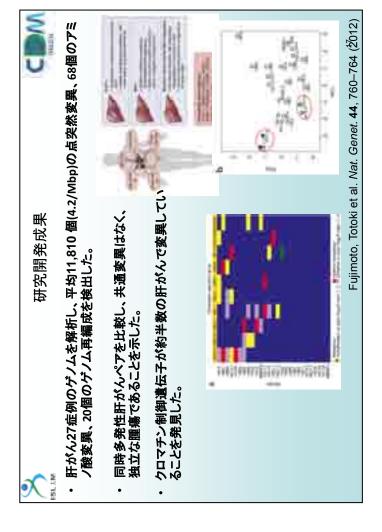
genetic variation.

nerica, Inc. All rights reserved.

Whole-genome sequencing and comprehensive variant

Hitoshi Nakagama^{5,14}, Yusuke Nakamura^{1,15}, Tatsuhiko Tsunoda¹, Tatsuhiro Shibata² & Hidewaki Nakagawa¹

IST YM



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)

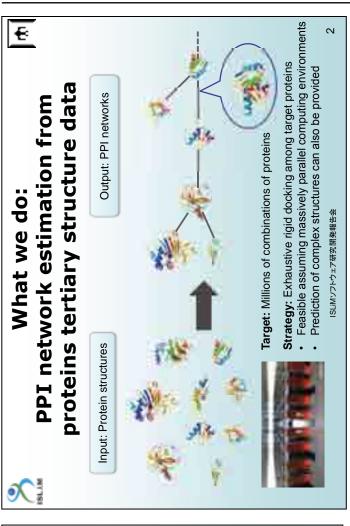
2001年4月 独立行政法人産業技術総合研究所 生命情報科学研究センター長

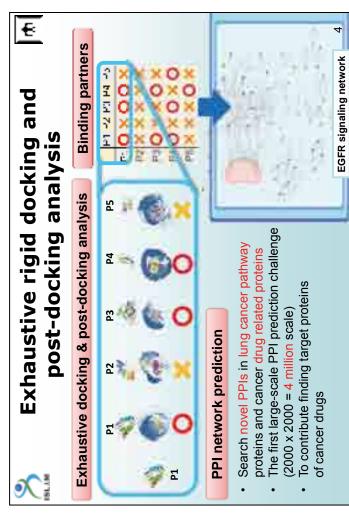
2007 年 4 月 東京工業大学大学院情報理工学研究科計算工学専攻 教授

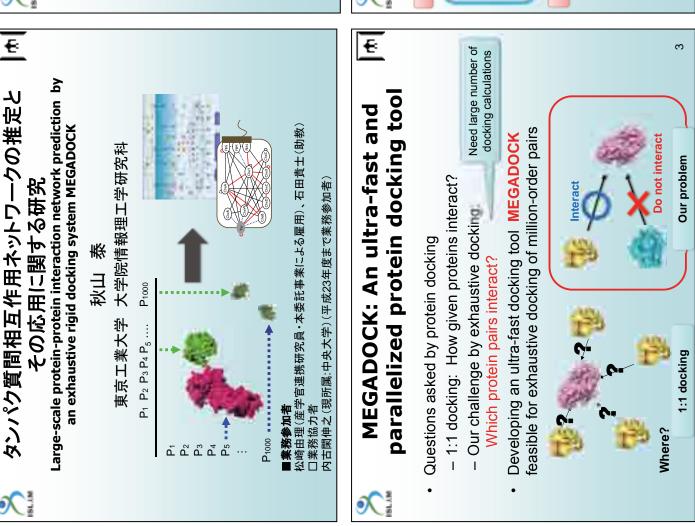
2011 年 12 月 東京工業大学情報生命博士教育院 教育院長 (兼任)

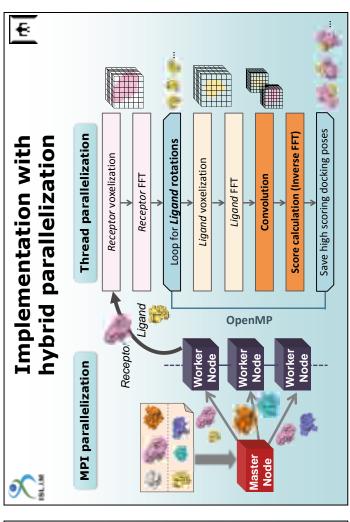
研究分野

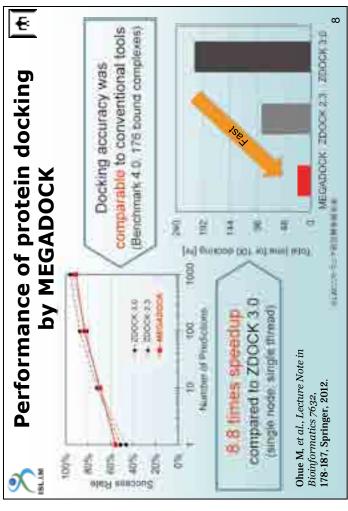
バイオインフォマティクス、並列処理応用

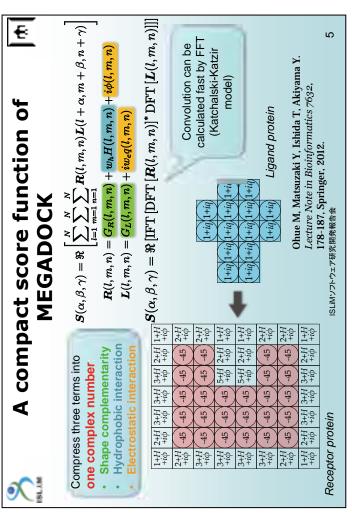


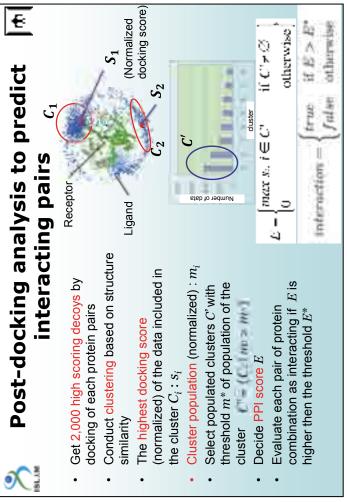


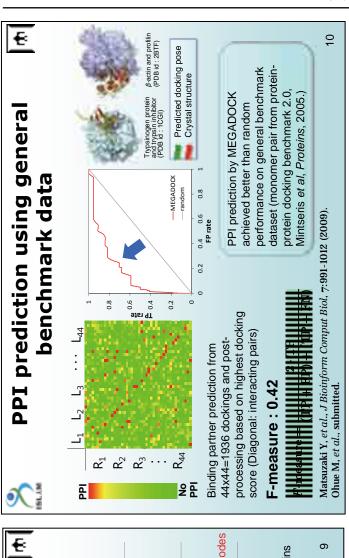


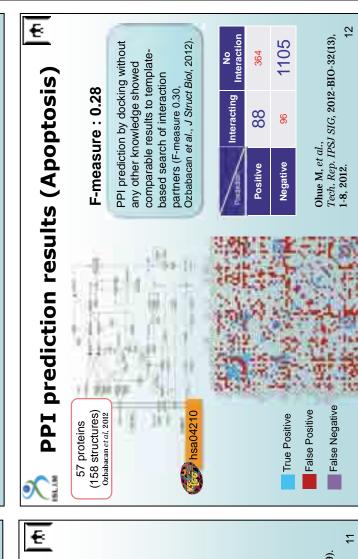


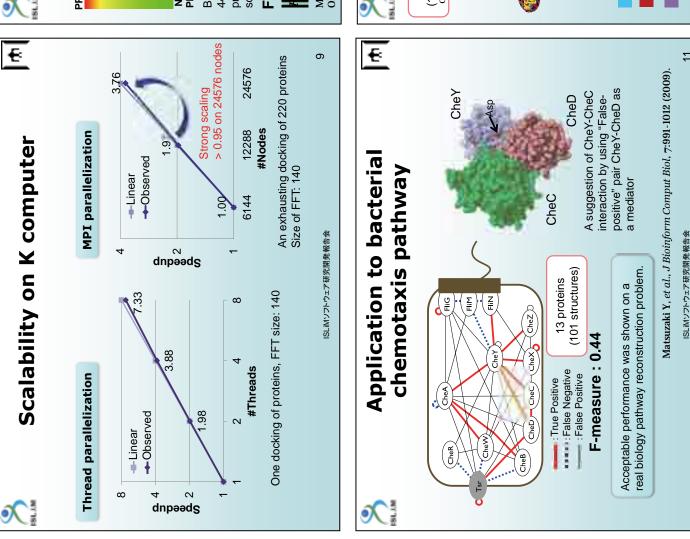












Searching novel cancer related PPIs

€

€ Application to non-small cell ung cancer pathway Completed large-scale exhaustive docking

€

and proteins related to Gefitinib

Interaction of cancer pathway

estimated by Miyano lab., The Univ.

Proteins related to Gefitinib

of Tokyo, from microarray analysis

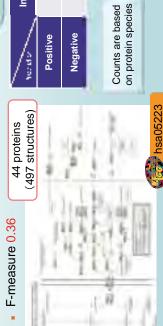
non-small cell lung cancer EGFR pathway related to

(1424 structures) 294 proteins

> $2,000 \times 2,000$ docking

(497 structures) 44 proteins

- . 497 structures, all-to-all docking =247,009 structure pairs
- Achieved high PPI prediction performance
 - Precision 0.29
- 0.47 Recall



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

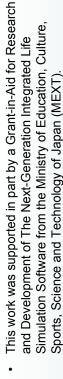
Interaction 131 nteracting 53

747 59 Counts are based €

3

Acknowledgement

ISLIMソフトウェア研究開発報告会



introducing a simple hydrophobic interaction model: an application to interaction pathway analysis., Lecture

Ohue M, Matsuzaki Y, Ishida T, Akiyama Y, Improvement of the protein-protein docking prediction by

(Papers contributed by MEGADOCK group)

References

Fleishman SJ, et al., Community-wide assessment of protein-interface modeling suggests improvements to

Ohue M, Matsuzaki Y, Akiyama Y, Docking-calculation-based method for predicting protein-RNA

interactions., Genome Informatics, 25:25-39, 2011 Note in Bioinformatics, 7632:178-187, 2012.

Uchikoga N, Hirokawa T, Analysis of protein-protein docking decoys using interaction fingerprints:

design methodology., Journal of Molecular Biology, 414:289-302, 2011

application to the reconstruction of CaM-ligand complexes., BMC Bioinformatics, 11:236, 2010. interaction prediction system using tertiary structure data and its application to systems biology

Ohue M, Matsuzaki Y, Matsuzaki Y, Sato T, Akiyama Y, MEGADOCK: an all-to-all protein-protein

Part of the result was obtained by early access to the K computer at the RIKEN Advanced Institute for Computational Science.



Takayuki Fujiwara **Fakehiro Shimoda** Kohei Yamamoto Masahito Ohue

9

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.

(Submitted)

Matsuzaki Y, Ohue M, Uchikoga N, Akiyama Y, Protein-protein interaction network prediction by using rigidbody docking tools: application to bacterial chemotaxis., submitted

Matsuzaki Y, Matsuzaki Y, Sato T, Akiyama Y, In silico screening of protein-protein interactions with all-to-

study., IPSJ Transactions on Mathematical Modeling and Its Applications, 3: 91-106, 2010.

all rigid docking and clustering: an application to pathway analysis., Journal of Bioinformatics and Computational Biology, 7:991-1012, 2009.

- Uchikoga N, Matsuzaki Y, Ohue M, Hirokawa T, Akiyama Y, Improved post-processing of protein-protein
- ISLIMソフトウェア研究開発報告会

12

87

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

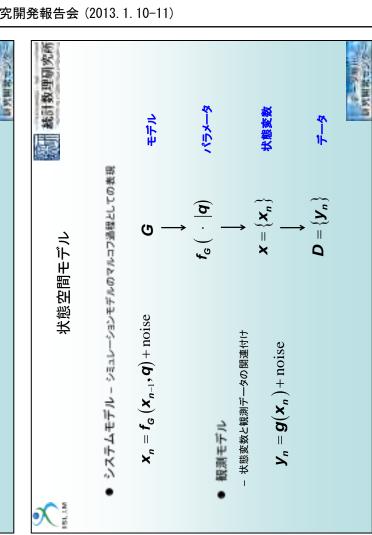
副所長(兼務)(~2010.3)

2011 年 4月 統計数理研究所 所長

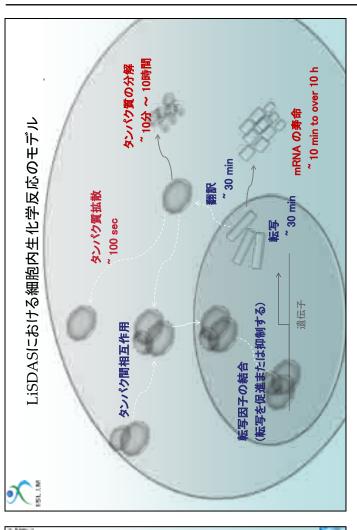
研究分野

ベイジアンモデリング、データ同化

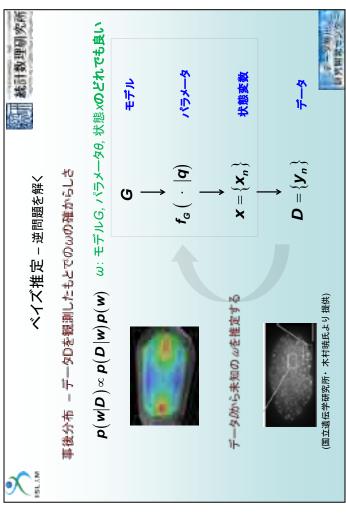




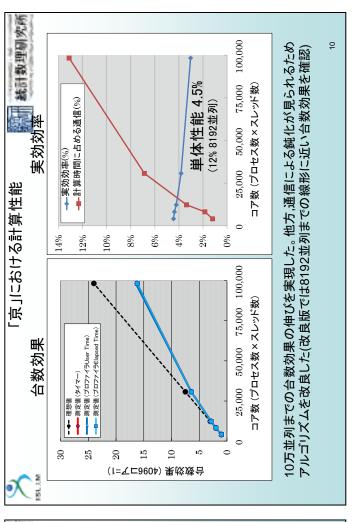


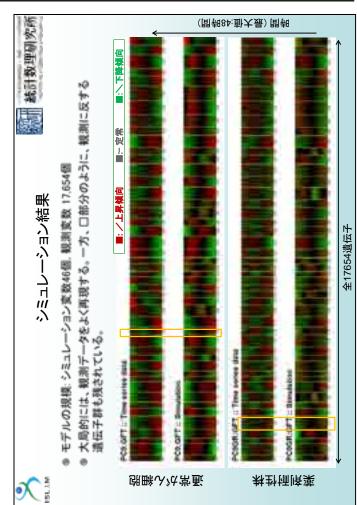


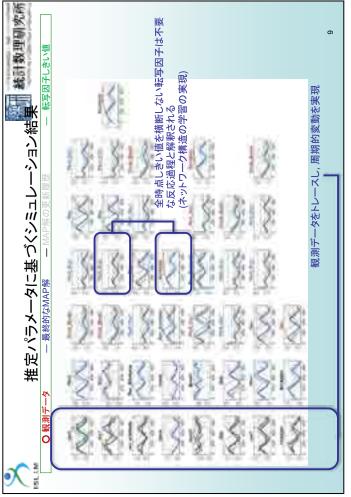




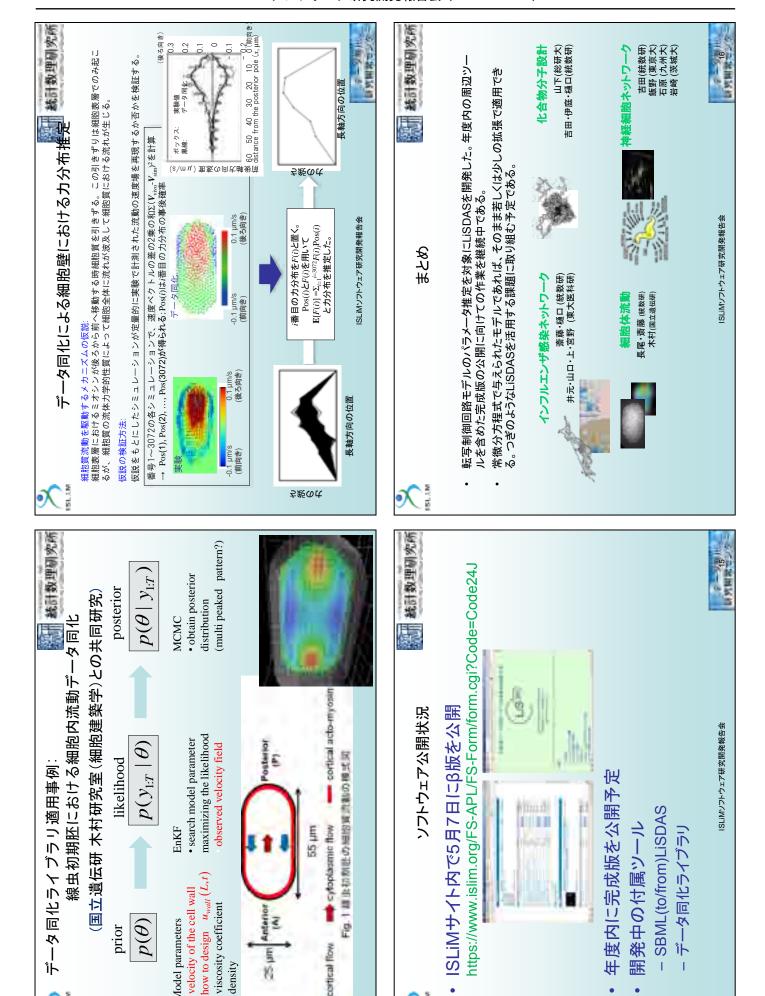












ISL AM

cortical flow

viscosity coefficient

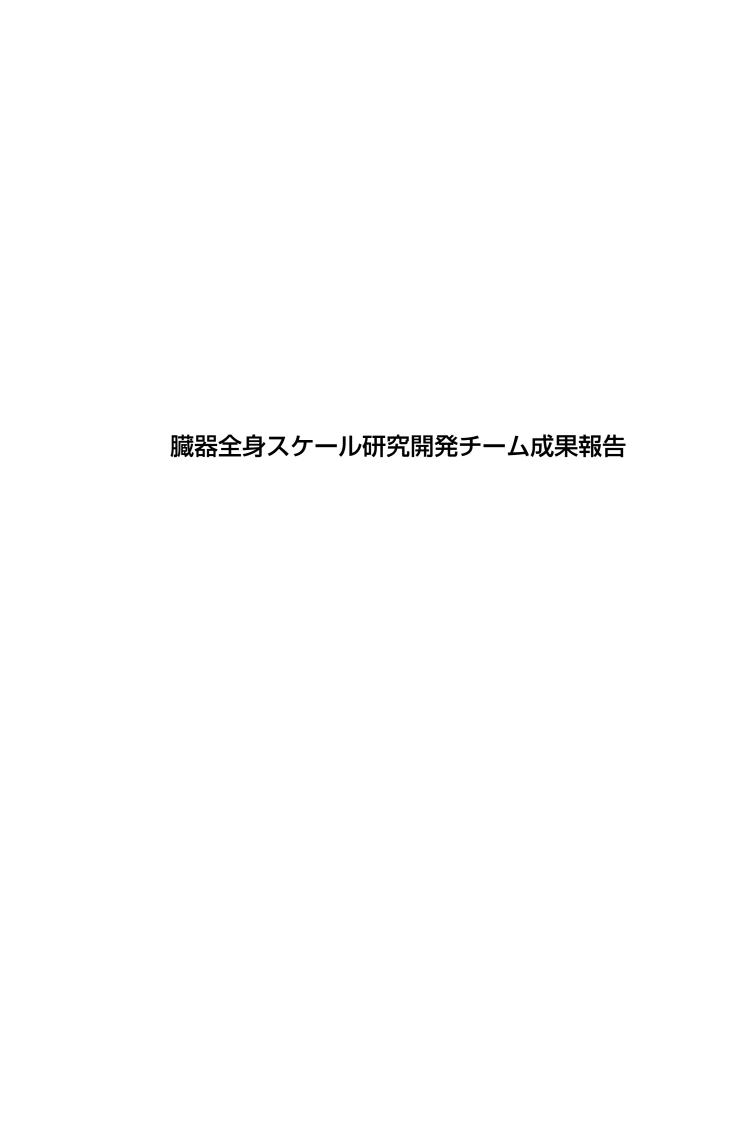
density

馬の

 $|b(\theta)|$

prior

Model parameters



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変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-1bα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ック 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**←→**心筋細胞**←→**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の設計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、東京大学の大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、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、「ものつくり」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となる超音波パネル (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、0 D-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

久田(東大)

0

ラクロ(国産農学)



臓器全身スケール研究開発チーム

理化学研究所

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

個別テーマでの主立った成果

チーム全体目標

C.

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ペタフロップを達成。 血栓症の初期過程のマルチスケ 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

ソフトウェアの整備と「京」での本格的計算の継続

スクロ (植取製料・分子弁製料

器設計のシミュレーションを開始.

今後の予定

一 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。

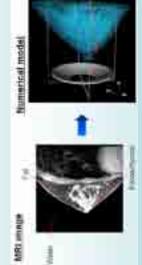
C'

臓器全身スケールチーム:概要

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





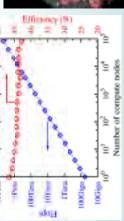


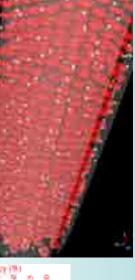




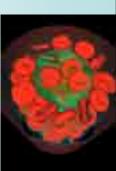
医用画像データに適した流体構造連成手法(ZZ-EFSI)

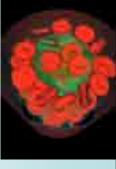
「京」のフルノード計算で、実効速度 4.5ペタフロップスを達成(世界最速の流体構造連手 法ZZ-EFSIの開発に成功)















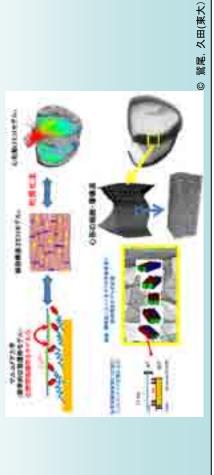
マルチスケール・マルチフィジックス心臓シミュレータ (UT-Heart)

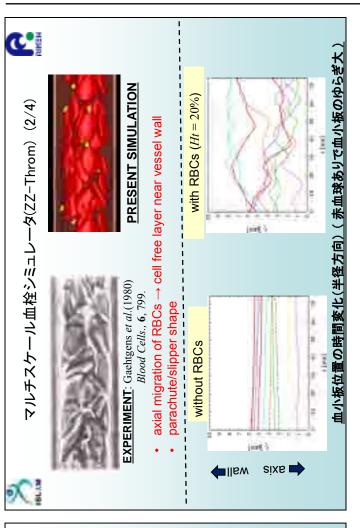
UT-Hearlは既に臨床研究に用いられている.

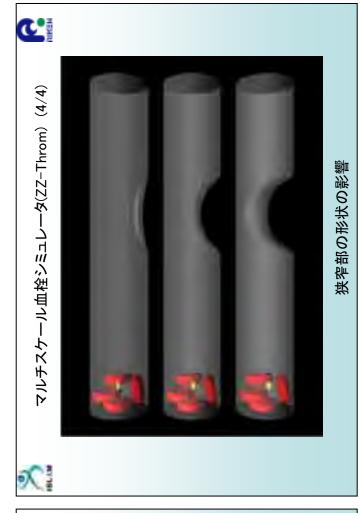
•

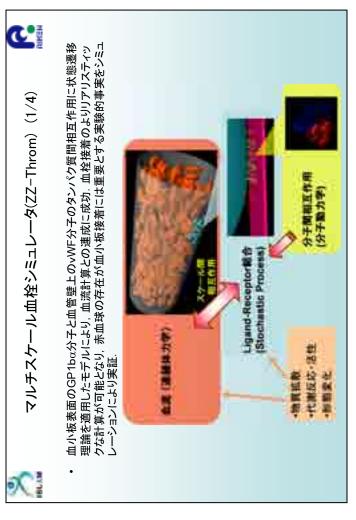
- グランドチャレンジで開発されたソフトのうち,唯一,「京」利用の優先課題となる. •
- 「京」モデルでは,サルコメアレベルの運動から心筋細胞の機能再現,さらに心臓全体 と3段階の異なるスケールの階層統合に成功

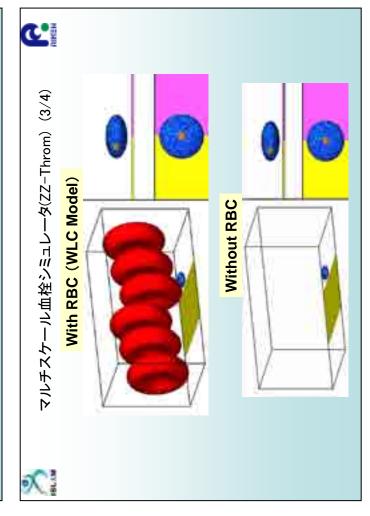


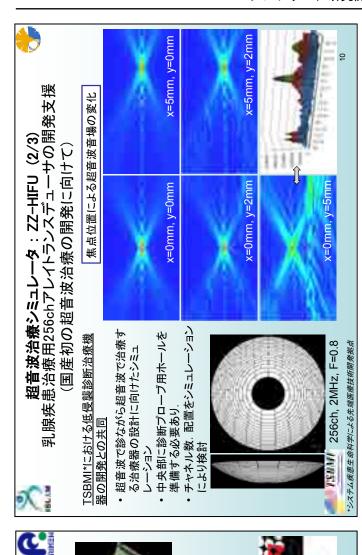












超音波治療シミュレータ:ZZ-HIFU(1/3)

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

強力集東超音波(HIFU)治療機器から照射された超音波が人

2HMzの超音波の1波長0.75mmを15格子(⊿x=0.05mm)で解

9000x6000x6000=340G格子点が必要であり、そのような大

像する場合, 450x300x300mmの領域に対して

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

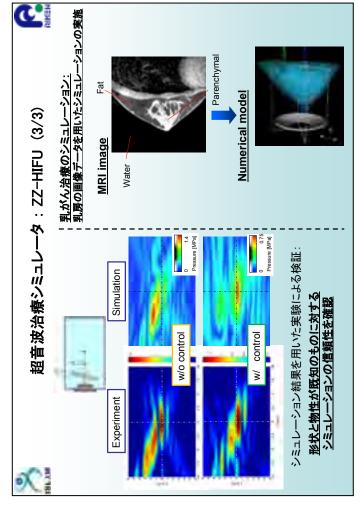
解くことで、不均質媒体中の超音波伝播を再現し、粘性散逸を 熱源とする温度の時間発展式により温度場を得る.

・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(10亿次精度を選択)

有限差分法によって基礎方程式を近似

- 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



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

• 領域分割とMPIIこよるプロセス並列

- 並列化手法

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

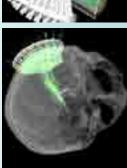
流体工学. 数值流体力学. 混相流



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

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

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)







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



provides the heat coagulation of tissue Minimally-invasive treatment, which around the focal point of ultrasound. **Energy therapy**

C.

High-Intensity Focused Ultrasound therapy

C.

- Benefits
- Short hospital stay Radiation free
- Repeatable treatment

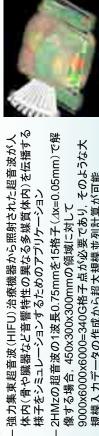


国産の超音波治療機器開発を促進するために超音波伝播シミュレーションによる支援

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

Real-time monitoring of temperature Accurate targeting



様子をショュレーションするためのアプリケーション

HIFU Simulation

xel Phanton



9000x6000x6000=340G格子点が必要であり、そのような大規模入カデータの作成から超大規模並列計算が可能 像する場合, 450x300x300mmの領域に対して

- 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 - 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 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)
- 領域分割とMPIIこよるプロセス並列

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

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

生体情報および設計形状を基にした入力データの作成から、

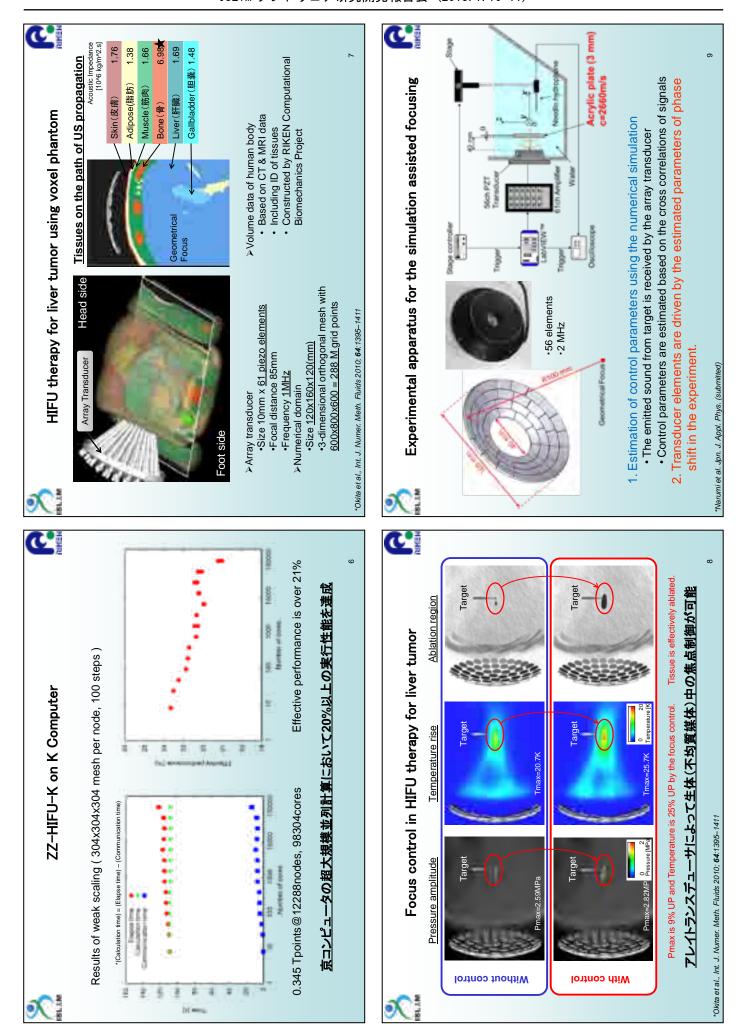
OpenMPを用いたスレッド並列によってマルチコアに対応

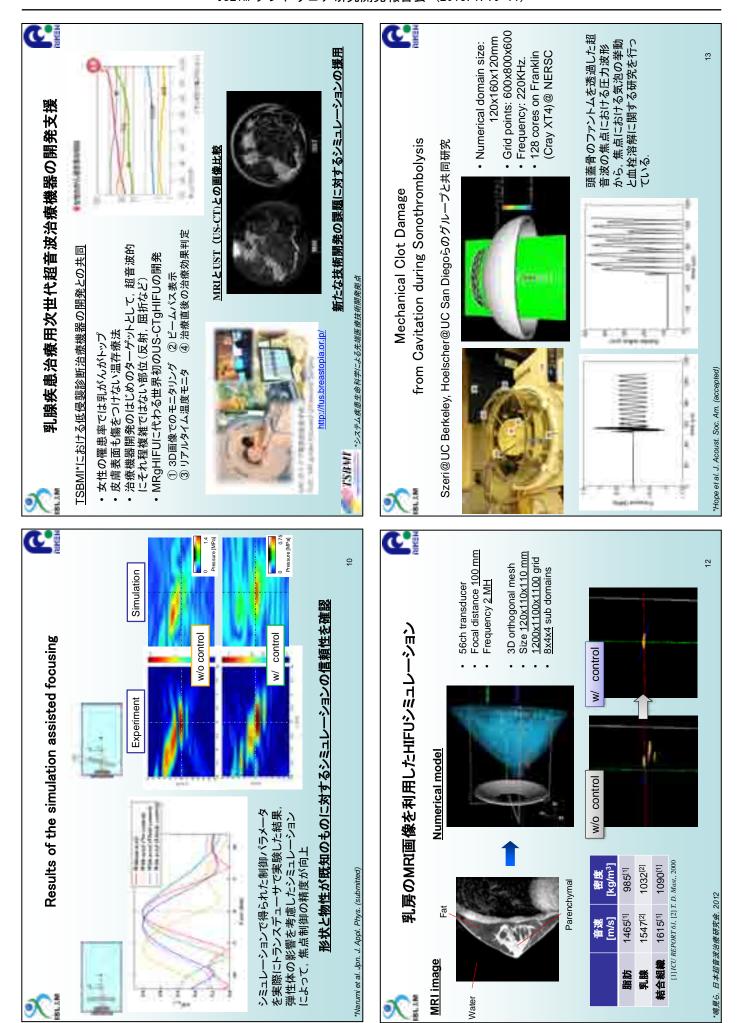
*Okita et al., Int. J. Numer. Meth. Fluids 2011; 65:43-66

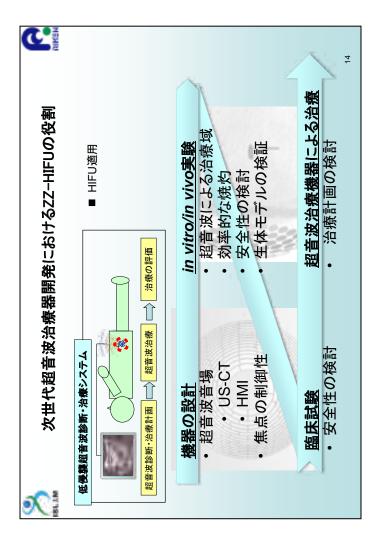
並列化手法



HIFU simulation based on digital human model







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



流体構造/膜連成手法 (ZZ-EFSI) 大規模並列計算に適した

œ

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)

背景・医療画像と相性の良い流体構造解析を実現したい

œ

plood: musede pod fube

Voxel data (VOF; Volume Of Fluid)

→ 多媒質幾何·材質

理化学研究所 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)







Euler法に基づくシミュレーション

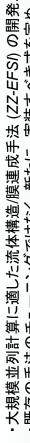
有限差分•固定格子)



œ 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







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

和積演算量は多いほど良いが

差分は避けたい。

に対する演算量

-SIMD

(or 自動並列)

-OpenMP

・パイプライン・メモリアクセス量

≒ 8・10⁴ (ノード) × 8 (コア) × 8 (寅算)× 2・109 (Hz)

1秒間に1京(1・1016)回の演算

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

ŧ

16並列

4並列

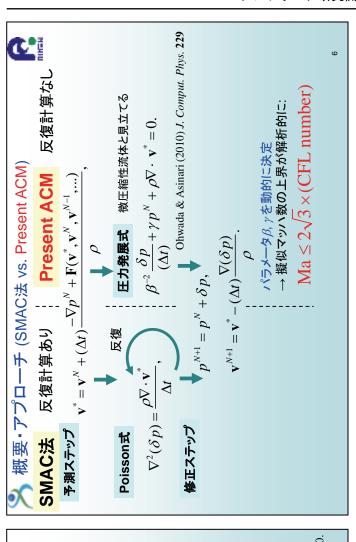
数値的に安定で陽的な時間積分

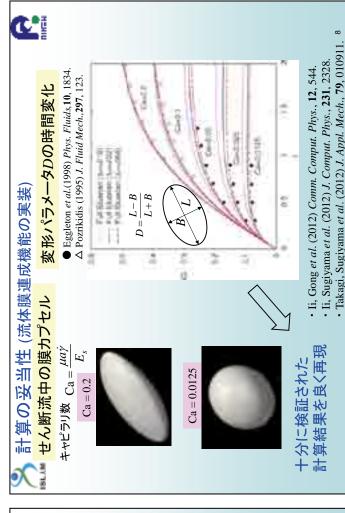
反復の多い処理を避けたい ──非圧縮性媒体を解析対象としても

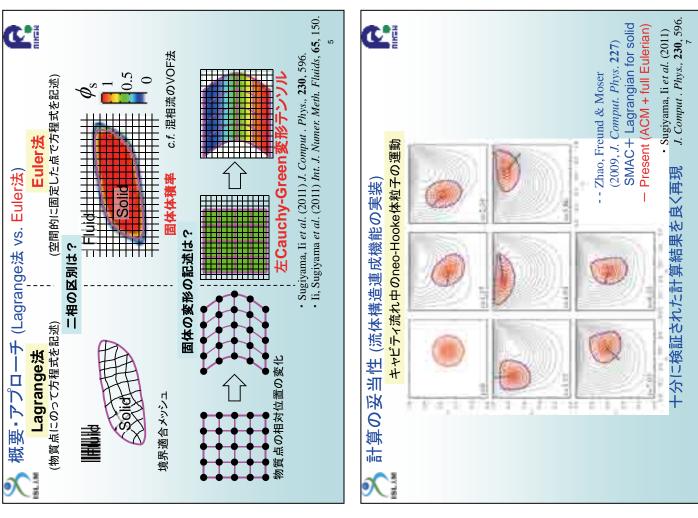
ノード毎の演算量を均一にしたい → Euler法が有効

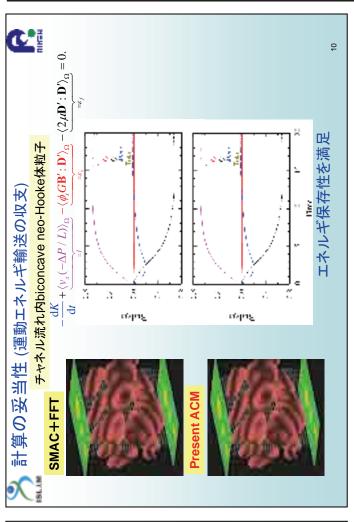
通信(ノード間の情報のやりとり)

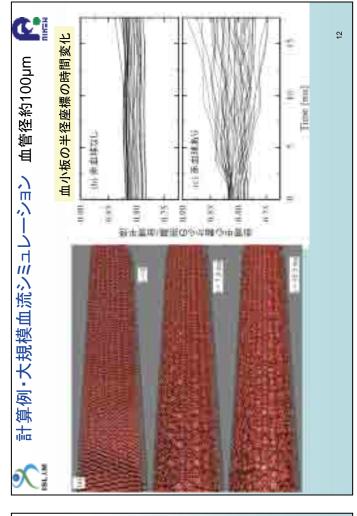
圧力Poisson式を解きたくない

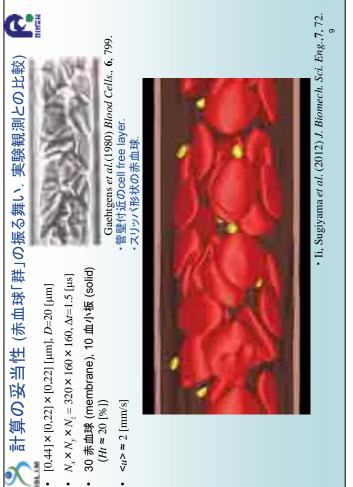


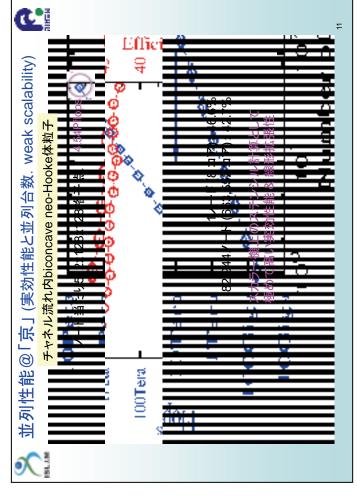


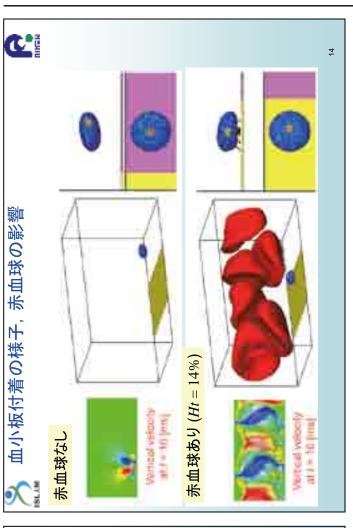


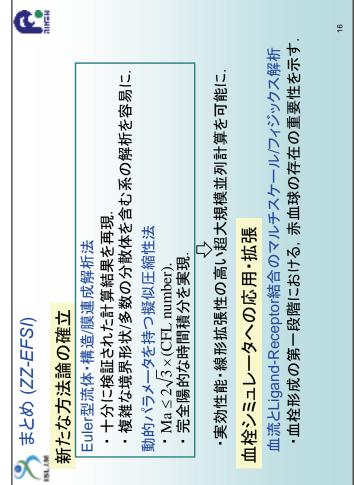


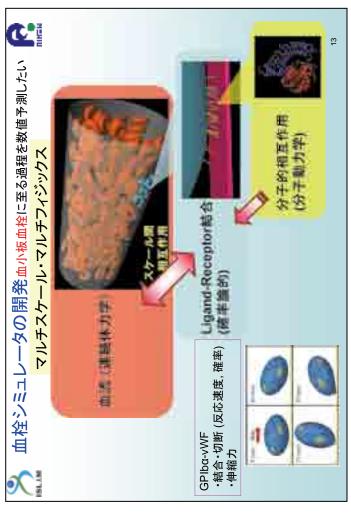


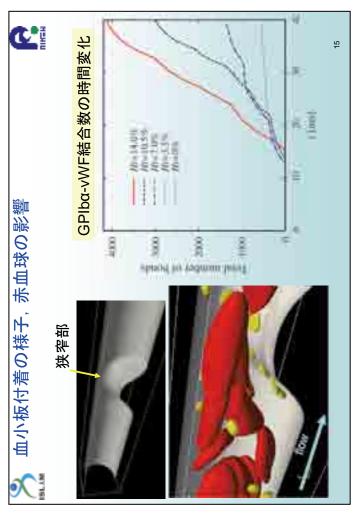












UT-heart によるサルコメア力学から 心筋細胞構造を経て心拍動にいたる マルチスケール解析

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 CREST 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

数值解析, 計算生物学



心筋細胞構造を経て心拍動にいたる UT-heartによるサルコメアカ学から 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



ISLIMソフトウェア研究開発報告会

東京大学

背景·目的

東京大学

淵

- 「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。

宏 Ш

- 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。
- 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 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 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会



現在までの研究開発成果

- 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 現時点の開発ソフトウェアについて 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

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トな連成法を開発 非圧縮材料を扱うことのできる均質化法を開発

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

計算事例

ISLIMソフトウェア研究開発報告会



東京大学

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 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 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 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、 従来にない観点から医学・医療に貢献することを目指す。
- アプローチ
- 大規模な電気・化学・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。
- 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 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、

ISLIMソフトウェア研究開発報告会





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



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

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

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 8 PI 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 O 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α =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 東を考慮したアクチンフィラメントのバネネッ 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ので、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、10⁹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 的な注意入力が局所回 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。

臼井(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ようである。(1) 眼球運動系 は、コンダクタンスベ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 α =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

ーロンの回路モデルとして実装した。網膜入力に 基づきサッカー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、サルの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、サッカード角度に対して 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 の成果は以下のようである。(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枝 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 μ m 以下の誤 差でマッピングするプロコトルを確立した (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) 神経細胞プラットホームシミュレータ 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 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、NeuroMorphKit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、10⁹ 個の細胞、10¹³ 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

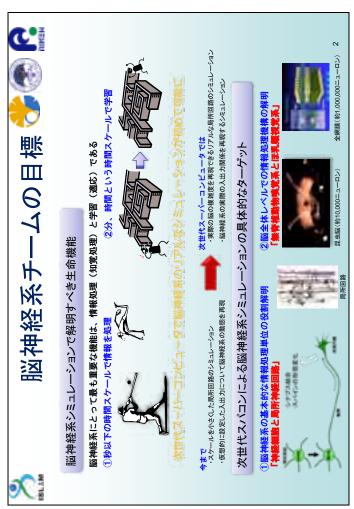
詳細な生物学実験に基づいた昆虫の嗅覚系神経回路および細胞に関するデータベースと、それに基づいて形態抽出や標準脳マッピングを行うことでシミュレーションにつなげるシミュレータ IOSSIM の開発を完了する。京上の 80,000 ノード実行で1ペタに近い実効性能の達成を目指す。神経回路レベル・単一細胞レベルの生物的パラメータ探索とその評価を行う。匂い情報入力から匂い源探索行動の行動指令を生成する中枢回路(LAL-VPC)のシミュレーションを、京上で高並列実施する。大規模マルチコンパートメントモデルについては、京上で 2~3 万コア並列程度によりLAL-VPC 神経回路の特性の時間計算を達成する。実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

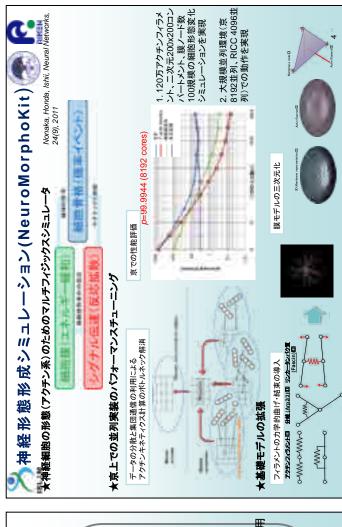
3.5 将来に向けて

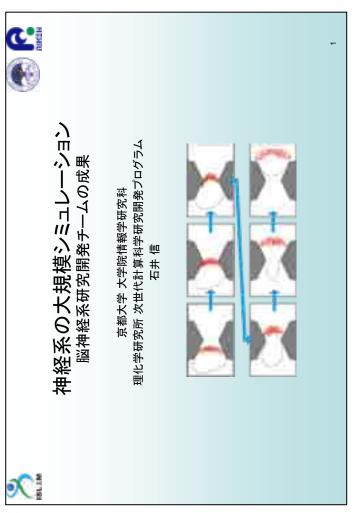
脳神経系研究開発チームの研究は、長期的には、 脳神経系における発達障害の機序の解明、最適な 学習(教育)過程の設計、脳を模した知的情報処 理デバイスの開発などの社会・産業応用、また、 ブレインマシンインターフェースの高精細化な どの医療応用に貢献することを目指している。欧 米では、脳神経系シミュレーション研究はこれま でスパコン開発と足並みを揃えて発展しており、 Blue Gene/Qのターゲットの一つは神経回路シミ ュレーションとされている。また、欧州では Human Brain Project というメガサイエンスの動 きがあるが、ここでもスパコン開発が重要となっ ている。われわれは、次世代の 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

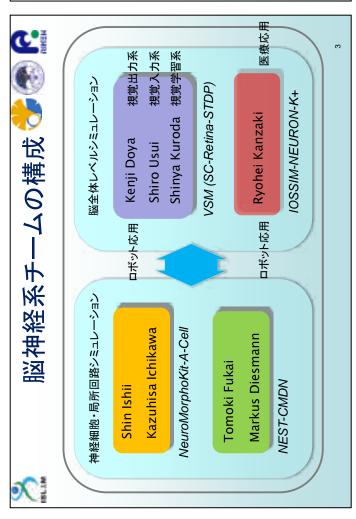
態を持つ細胞からなる大規模回路の超高並列シミュレーションが必要となる。後者について、ISLIM プロジェクトでは回路(通信)を得意とする NEST と、精緻な細胞を得意とする NEURONの両者について、京上でチューニングを行ってきたので、今後はそれらを統合することのできるソフトウェアの開発も必要であると思われる。

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る。









C.

N-Cellによる転写因子NF-xBシミュレーションの高速化 背景と目的

C.

器ショュワータ(NEST

后

匝

guided by neuroscientific goal to combine microscopic and macroscopic structure

Synapses Status Dec 2012

Neurons

Network

Description

Step

type

technically achieved on K on 28. Aug 2011

3.1010

3·10⁶

random at

reduced primate visual cortex

micro and

macro scale

(32 areas, each 1

local microcircuit)

chieved on K on

max 10¹[€]

max 10¹¹

spatially

random network 108 neurons and 1012 synapses Sep. 2012

1012

8

spatially

reduced primate

visual cortex

organized at

micro scale

especting relative

size of areas (32)

- NF・xB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・xB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-kBの活性化は細胞接着タンパク質、NOS、BDNF、CamKIIδの遺伝子が発現。 NF-kB欠損マウスでは放射迷路学習に異常(図1)。
- 核内NF-kB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。
- 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 →IkB産生→ NF-kB核外移行をモデル化(図2) 様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→IkB分解→NF-kB活性化→ NF-kB核内移行
 - A-Cell高速化のためにこれまでのMPIバージョンを改良 かつモデル非依存的なMPI高速化を実施。
- (approx. 100 areas) organized human brain model 核/細胞質体積比、拡散定数、核膜輸送量、IkB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)。









N

ы

Step 1: investigation of reduced visual cortex model now possible on K

Expected results:

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

assessment of capability to simulate full scale brain mode

Step 3: full human brain simulation

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。 図3 (Ohshima, D., et al., PLoS ONE, 2012, e46911)

MPI parallelization of A-Cell>99.99%

今後の展開

フーションによる 処理の解明(CMDN) 情報 路シニで質情報 皮 大 添

ESL LIM

C:

ᇷ

よの存

4

沿

NEST 9

speedup $\alpha = 0.59 \ @ 65536$ cores

(1) Runtime improvements

ESLAM

hybrid MPI + OpenMP code

(strong scaling)

それを用いた視覚的注意の2カラムモデル 層構造をもつ視覚野局所回路の構築 $\widehat{\Box}$ 5 3 by Potjans and Diesmann 80,000 LIF-neurons

斯尔牛理軍略子

解剖実験データ

Wagatsuma, Potjans, Diesmann and Fukai (2011) Front Comput Neurosci (F)空間注意

Potjans and Diesmann and Fukai (2012) Cereb Cortex

と特徴注意(下)の8カラムモデル

Layer 5

シミュレーション

09 20 9 8 8

(ZH)率火発

(3) production network 5·108

THE THE THE

employing sparse representations with heterogeneous plasticity rules

enables efficient distribution on

M≥100k cores

storage of neurons and synapses

(2) Memory improvements achieved phase VI on K (simulation and wiring)

26 minute simulation time 1 second biological time

196608 cores

0

McAdams & Maunsell, 1999; (Martinez-Trujillo & Treue, 2004;

Layer 5

9 20 4 3 10 0

• •

•

ø ,

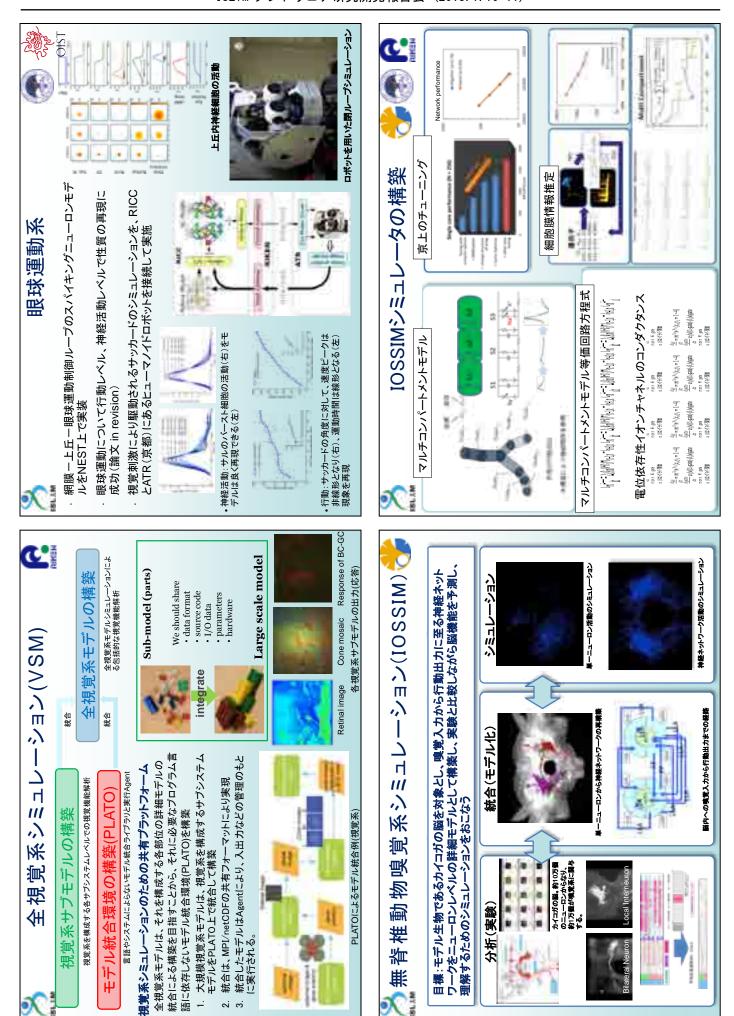
ing, Liu & Carrasco, 2009

◆大脳皮質回路と海馬 (Colgin et al., Nature 2009) の計算原理の共通性 を発見 (シータ波-ガンマ波の結合 波)

局所回路モデルの ためのデータ解析 Binzegger et al. 04

> machine size of K (end of red possible with present code's memory consumption on full (4) Next milestone: N_{max}=10⁹ theory of memory usage broken line)

128





ソフトウェアの達成状況 🔧 🦔 🖪





学術的成果

	7	0		0	ო	:	7 12). Fr (2011	7 (1) Pro	2010).
	本総額形ツェフーション	局所回路	ンドコフーション	全視覚系ショントラン	昆虫嗅覚糸ンミュアーション	# DI - 1 VI	画文稿駅ハイライト Helias, et al. (2012). Fr Nishiyama, et al. (2011	Linden, et al. (2011). No Honda, et al. (2011). Jo Fujii, et al. (2011). Proc	lonyama, et al. (2010).
原上のストロン ソフトウェアの歌			アクチン線維数	1.2 × 10 ⁶	番酌数5×108 シナプス数	3×10^{12}	視細胞(錐体)数 1.5×10 ⁶	□ソパートメント数10⁴奮 断 数 8 × 10⁵	
京上のストロング・	アインール(瀬戸コア数)		0.994(8192)	Hybrid	0.59(65536) Hybrid		0.89 (RICC 1024) MPI	0.83(196603) Hybrid	
京十の解釈は			8192		196608		1504	196608	
RICC上のMAH	第		4096		8000		2460 (whole system 1024)	8192	
			神経細胞	ション NeuroMorphoKit	同児回路 ンミュアーション	NEST	全視覚系 シニュレーション VSM	昆虫嗅覚系 シミュレーション IOSSIM	

4 ceedings of National Academy of Science USA, 108(44) 7 2 N Jeuron, doi:10.1016/j.neuron.2011.11.006. တ ω N 2 . Molecular Systems Biology, 6(394). rontiers in Neuroinformatics, 6(26) ournal of Neuroscience, 31(4) 1). Nature Cell Biology, 13(6). Isomura, et al. (2009). Nature Neuroscience, 12. 国際学術専門雑誌掲載論文(予定を含む)件数 N 9 0 က 0 က

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~

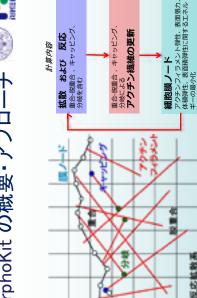
- 細胞外シグナルを情報処理

骨格系を制御

細胞内シグナル伝達

13

NeuroMorphoKit の概要・アプローチ



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

計算モデル

散計算を行う。

Song, et. al., Nat Cell Biol, 200

Cdc42-FRET imaging in fibroblast by Marsuda lab

反応執数率 アクチンフィラメントは線分で表現、 リゴンで表現、エネルギー最適化 台・脱重合、キャッピング、分岐を 細胞膜は膜ノードを頂点としたポ 分子の濃度に応じて確率的に重 コンパートメントに存在する各種

含むイベントを生成する。

Growth cone at axonal tip

アクチン重合阻害剤

数小衛

アクチンフィラメント

- アクチンフィフメント・쓇小簡

細胞骨格

- 機械的に膜を支持

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

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 並列化の方法

使用する計算法 により形態を求める。

Runge-Kutta法 • 最急降下法

by Dr. Kaoru Katoh

各要素を統合してシミュレーションする必要がある

細胞膜直下に多く存在 中心体から放射状に存在

15

OpenMP, MPIIこよるハイブリッ

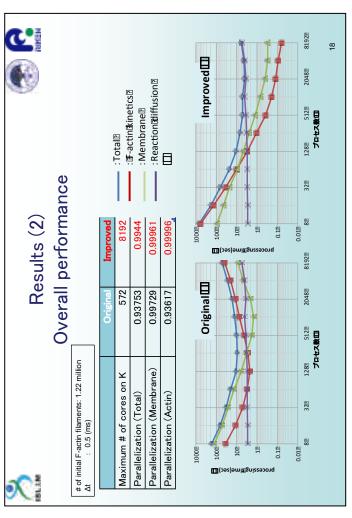
ド並列化

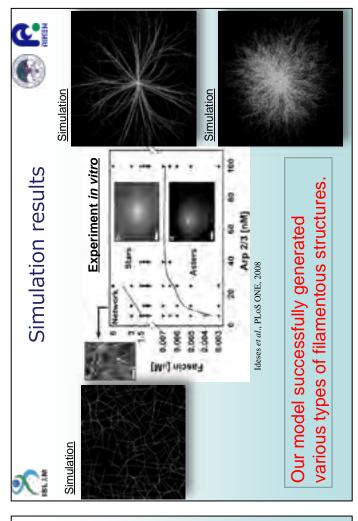
開発言語と主なライブラリ

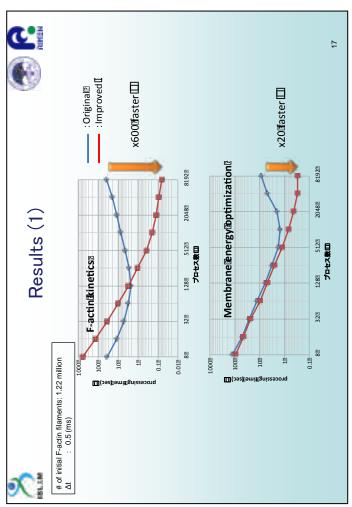
[‡]0/0

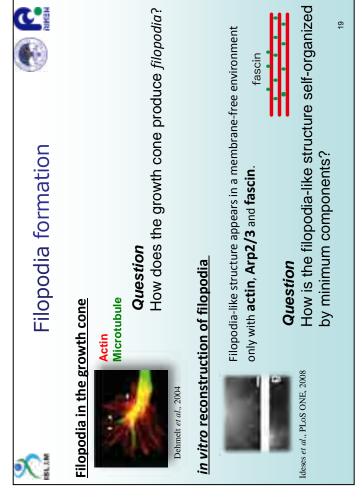
MPI, OpenMP, GSL, NetCDF

16









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・インスティチュート・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:: Julich



Advanced Course in Computational Neuroscience ACCN, ヨーロッパ

沖縄計算論的神経科学コース

ISLiMソフトウェア研究開発報告会



NEST 2.1

2011年9月 2012年5月 2012年9月

第2世代*

第4世代

全脳規模(109)にむけて 108の神経細胞の実現 京へのNESTの移植

ς.

第3世代

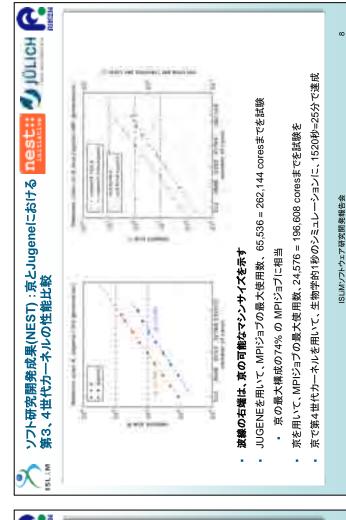
*第2世代 =第2世代 シミュレーションカーネル

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

nest::

109の神経細胞からなる局所回路モデル

こむけて



より大きなネットワーク:大きな遅延、疎な巨視的結合

最大領野(第一視覚野)よりも大きい規模

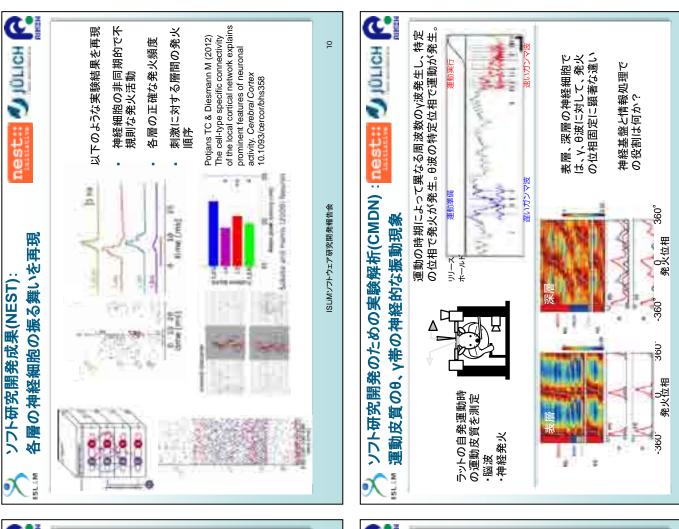
108 に相当する規模:

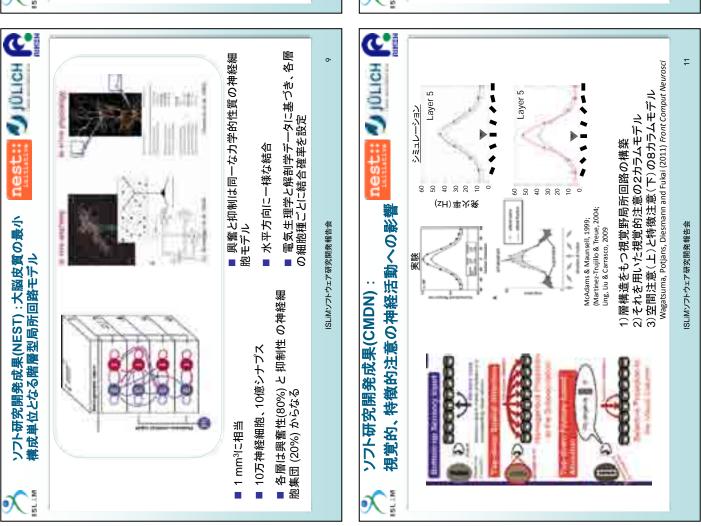
視覚皮質のモデルが可能

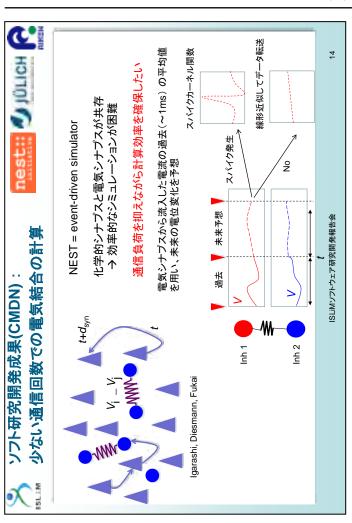
ISLIMソフトウェア研究開発報告会

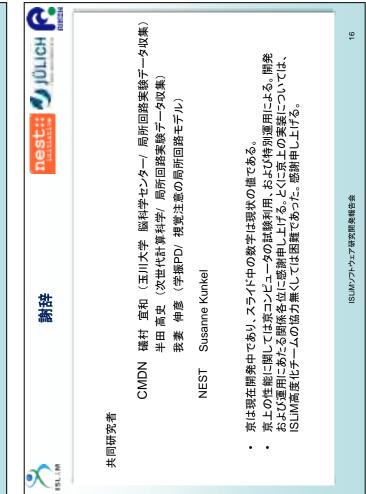
12

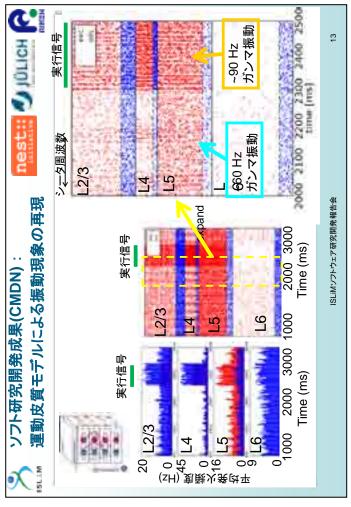
ISLIMソフトウェア研究開発報告会

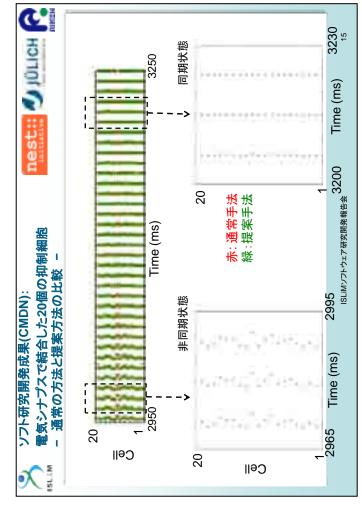












Multi-scale simulation of subcortical visual systems with real-time robotic visuo-motor feedback

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



発表者紹介

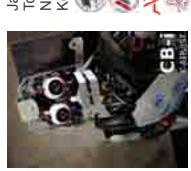
1995-11	MSc Computer Science, Lund, Sweden
2002-09	PhD Cognitive Science, Lund, Sweden
2002-06	Research Assistant, Lund, Sweden
2003-11	Post-Doc, CRL/NICT, Kyoto, Japan
2005-05	Visiting Researcher, NICT, Kyoto, Japan
2006-09	Researcher, ATR, Kyoto, Japan
2009-02	Researcher, Kyoto University, Kyoto, Japan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

研究分野

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



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



Norikazu Sugimoto **Tomohiro Shibata** Jan Morén Kenji Doya

















Project Goals

 Close the Perception-Action loop



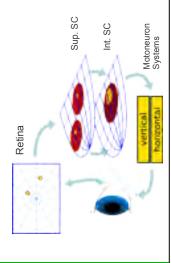
 Use a Humanoid Robot to ground Model in the World

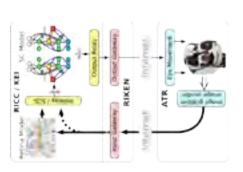


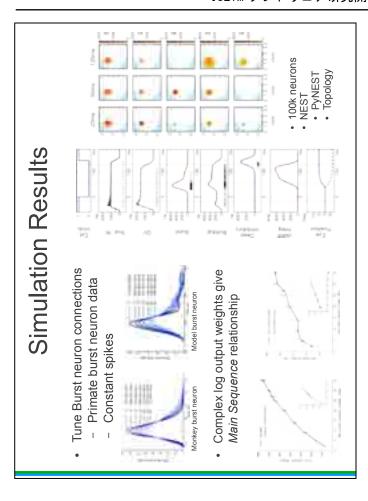
- Goals, Introduction
- The Superior Colliculus
- Model Simulation
- Connected to CB-I robot

Superior Colliculus

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







Š

Burst neuron profile

Model



CB-i Humanoid Robot

155cm 85kg Weight: Height:

• DOF:

Pneumatic Electric Power:

• Eyes:

Hydraulic

2× Wide-field 2× Foveal

2×2 DOF Firewire SARCOS & ATR Maker:

 Connect CB-i Robot to Model As Physical Eyes

— Motor Output - Visual Input - Camera - Eyeball

- Wako-Shi ←→ Kyoto Connect Over Distance

- React Directly to Input Near Real-Time control

SC Mode RICC / KEI RIKEN ATR Transport of the Parket

excitatory inhibitory NMDA

Burst circuit
 Reset circuit

Closing the Loop

heep layer inhibitiony

Buildup neuron disinhibition causes

spreading activation

Activation-triggered deep layer system-wide inhibition

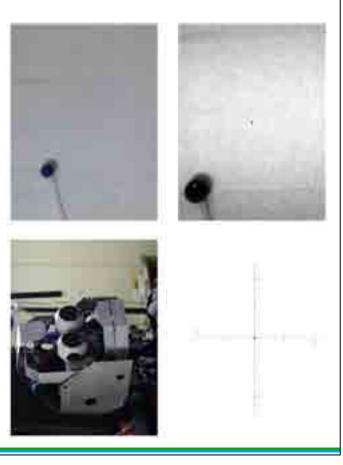
BU14

Buildup neuron activation

NMDA receptor bursting activation

cMRF spike integrator reciprocal

inhibition



RICC / KEI

- Scaled down, double-sided

• RICC: 256-1024 Nodes

(MUlti SImulation Coordinator)

Internet Connection

- VPN connection - SSH over TCP/IP

- MUSIC Interconnections

model implementation

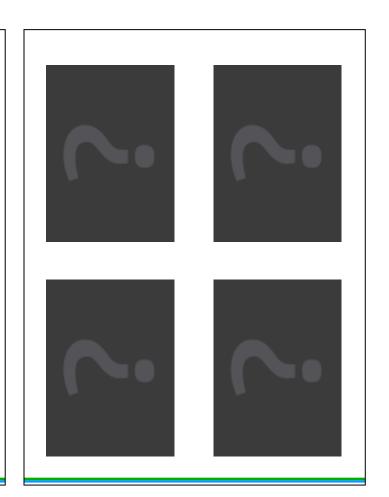
Closing the Loop



SSHum



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



- Video Saliency Extraction

- CB-i in test harness

• ATR

- Motor system

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

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

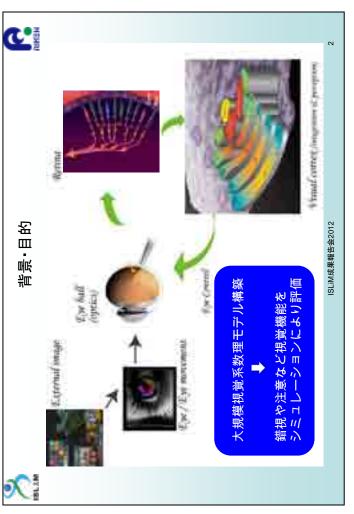
2005 年 4 月 理化学研究所脳科学総合研究センター・神経情報基盤センター 副センター長兼任

2007 年 4 月 同センター長兼任

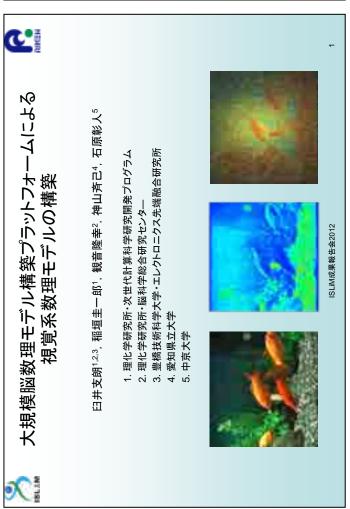
2012 年 10 月 豊橋技術科学大学エレクトロニクス先端融合研究所 特任教授 理化学研究所脳科学総合研究センター・神経情報基盤センター 客員主幹研究員兼任

研究分野

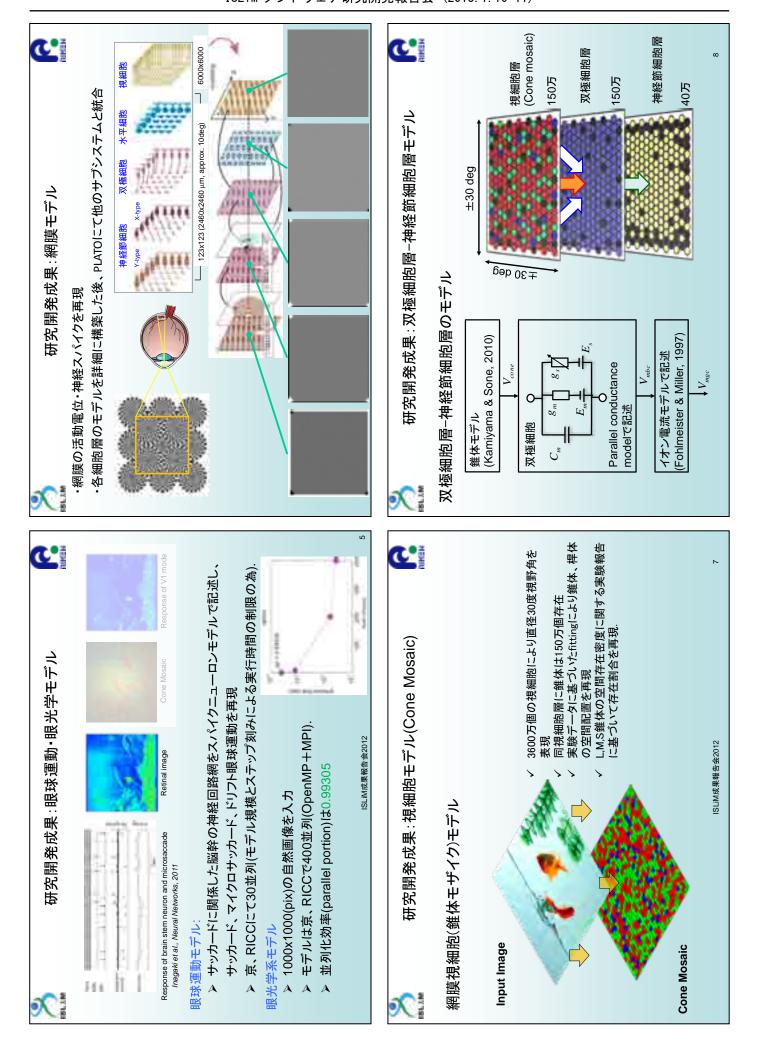
ニューロインフォマティクス、神経情報工学、視覚生理工学

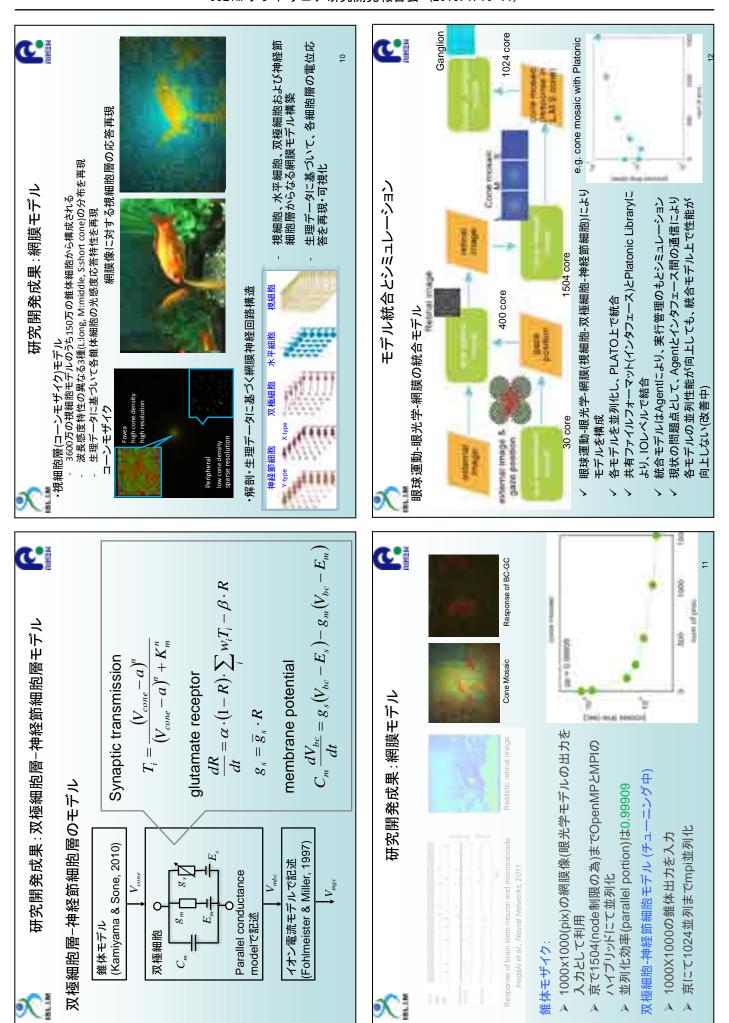


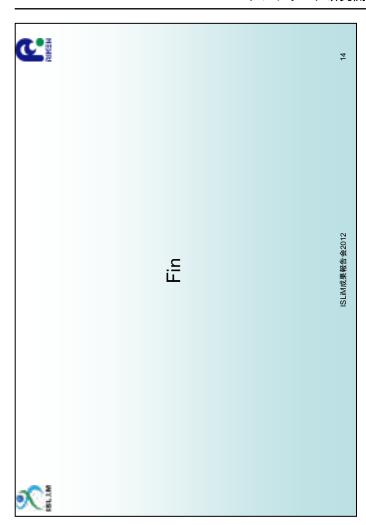
















まとめ

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

> モデル統合環境PLATOの構築

> サブシステムモデルレベルでは、眼球運部にて微小サッカード生成のメカニズムを ▶ 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

> 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
- 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 錐体モザイクで実施した情報量評価の他領域への適用

13

Computational Levelで明らかにした

「京」による 無脊椎動物嗅覚系シミュレーション 一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

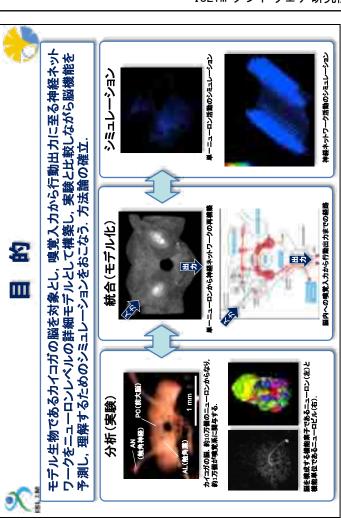
2006 年 4 月 東京大学 先端科学技術研究センター生命知能システム分野教授

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ー

神崎亮平(東京大学 先端科学技術研究センター, 教授)

次世代計算科学研究開発プログラム

脳神経系研究開発于一厶





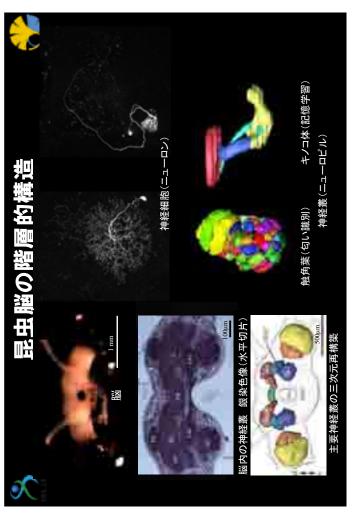
バイオリソースの1つで、ライフサイエンス研究の基礎・基盤のモデル生物.

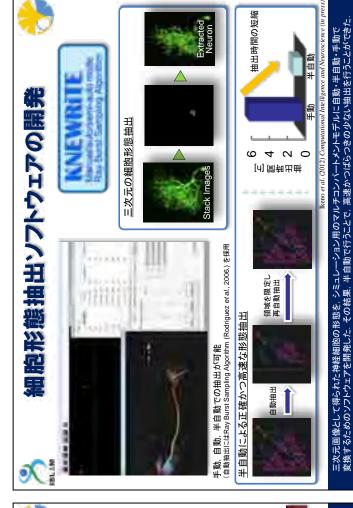
カイコガは、文科省「ナショナルバイオリソースプロジェクト(NBRP)」の

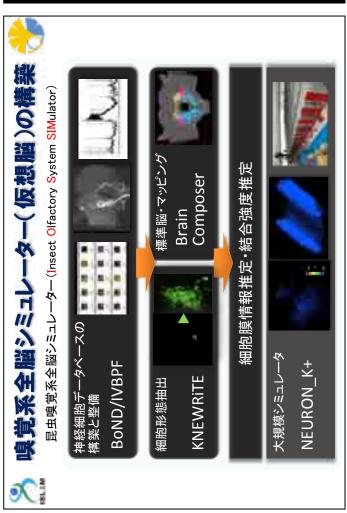
SLAM

4 踊しア ラタイムシミュレーションに

ISLIM成果報告会2012



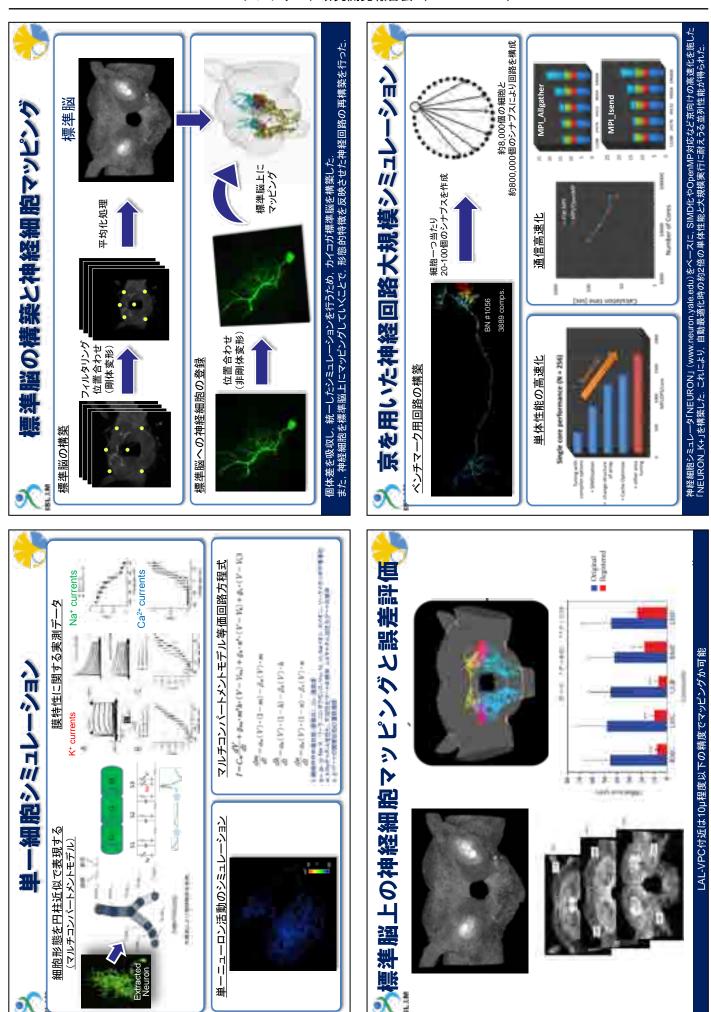




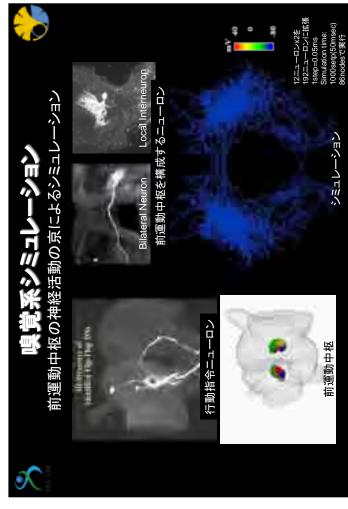


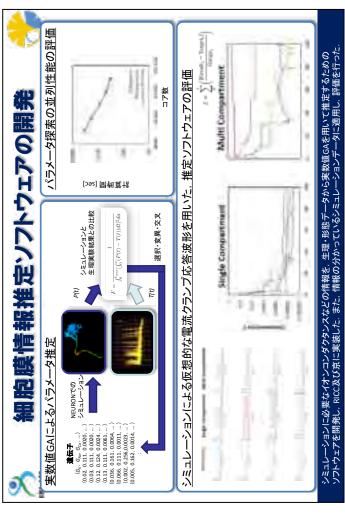
ニューロインフォマティクス国際統合機構(INCF)日本ノード:理研・神経情報基盤センター(日井支朗センター長)に構築する プラットフォームの1つとして情報公開している(IVB-PF委員会・委員長:神崎)

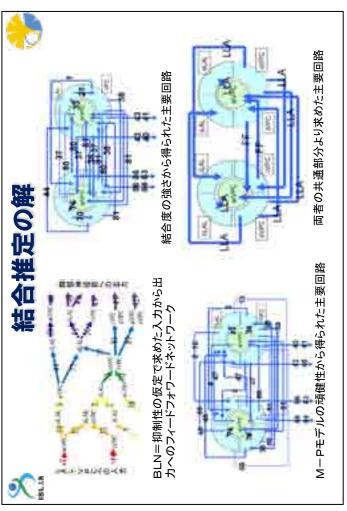
INCF Japan Node

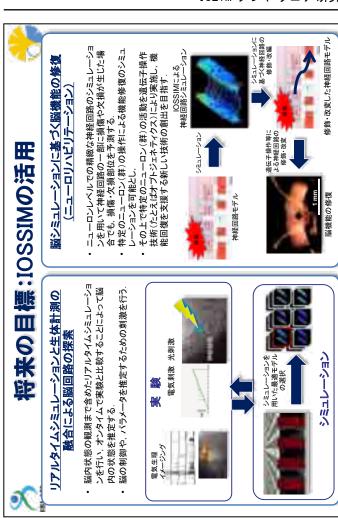














生命基盤ソフトウェア開発・高度化チーム成果報告

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暁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。

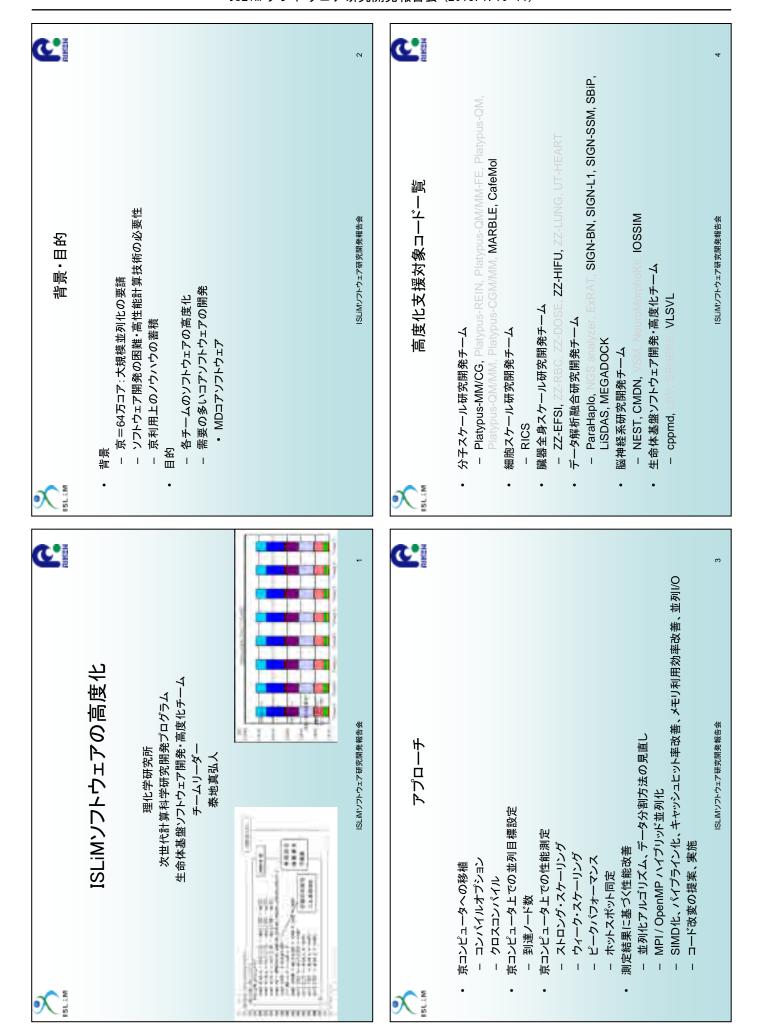
2.5 大規模仮想化合物ライブラリ

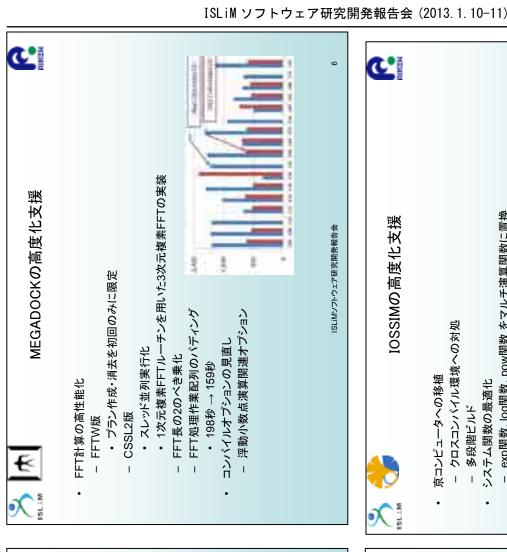
10 億化合物規模の合成可能性をもつ仮想化合物 ライブラリを、東大工学部船津研究室を中心に開発しています。

2. プロジェクト終了時の達成目標

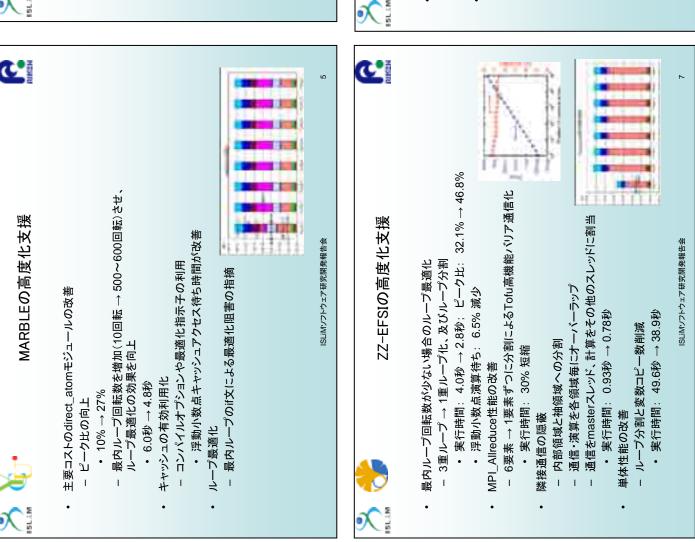
プロジェクト終了時に、より多くのアプリケーションが京コンピュータ上で十分な性能を発揮していることが我々の目標です。また、創薬応用に向けた 10 億化合物規模の大規模化合物ライブラリを完成させます。

以上

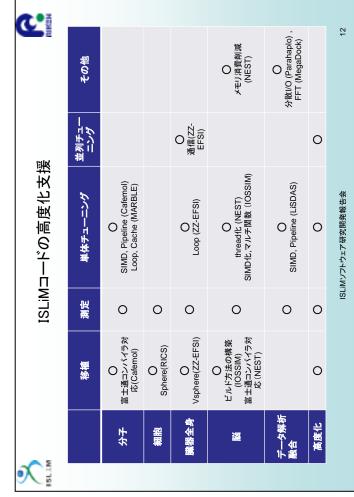


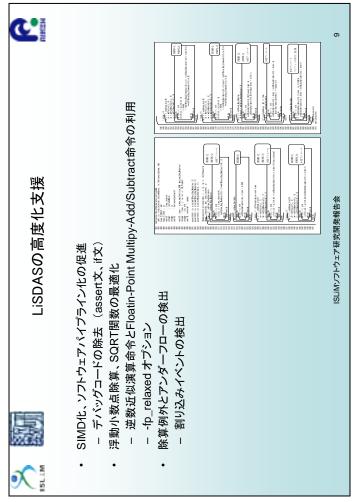


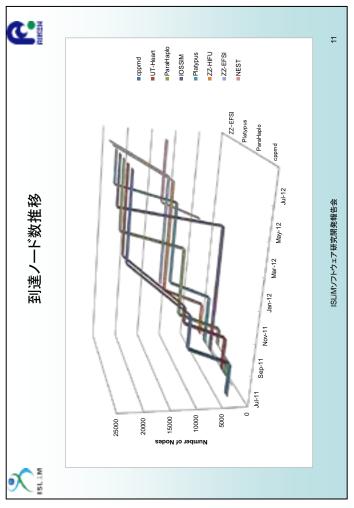












MDコアプログラムの最適化と性能評価

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 ものつ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



次世代生命体統合シミュレーション研究推進グループ 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 兼務

大野 洋介





C.

概要・アプローチ

現在までの研究開発成果

ISLIMソフトウェア研究開発報告会



データ形式の変更 - 単純配列

"I"文のマスク演算化

Cutoff法 4Pflops 40%

• 並列化: 直方体空間分割, cell index method

- 主にcutoff カーネルの最適化 ・どの手法でも主要な計算部分

アプローチ

高速多重極展開法(FMM)

・ 大規模並列での本命

クーロン相互作用: cutoff, PME, FMM

・ アンサンブル: NVE, NVT, NPT

• 時間積分: velocity Verlet

- 基本的な分子動力学計算

研究開発コードの概要

- TOFUネットワークに適した通信

高並列FMM

ほぼ近接通信、3次元空間分割

- 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

ISLIMソフトウェア研究開発報告会





「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。

- 得られた最適化の知見を他のチームへ提供する。

具体的な最適化の知見を蓄積する

・ 特に、最適化の挙動

- 再利用可能なコードとして提供する。

最適化済み分子動力学コード

• 最適化事例

C.

四





C.





Topology 4x4x4 Number of atoms 418,707 3,34 Time (ms / step) 109.058 11		16x16x16 32x32x32 26,797,248 214,377,984 111.186 111.672	48x52x32 522,546,336 112.414	48x48x36* 542,644,272 112.085
s 418,707 109.058	0.535 26,797,248	214,377,984	522,546,336	542,644,272
109.058			112.414	112.086
				1010
Force 91.622	91.528	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 0,003		1 500	3 8 7 1	1 034
(PFLOPS)		660.1	1 70:0	4.00
Efficiency (%) 39.0	38.5 38.3	38.1	37.9	38.0
	m 0	38.5	0.025 0.201 1	0.025 0.201 1.599 3 38.5 38.3 38.1

Cutoff 28Å, 6,542 atom/node, calculate energy every 4 step

60000

Cutoff 28Å, calculate energy every 4 step

Number of Node

Different to physical topology 48x54x32

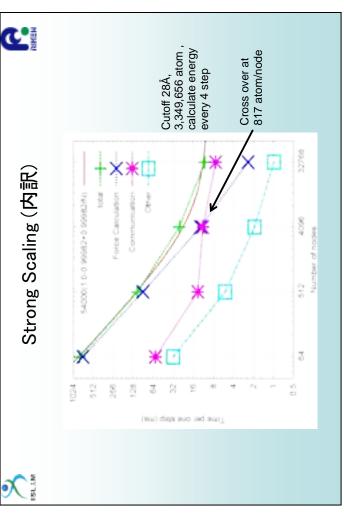
C.

Cutoff カーネル最適化

Cutoff法の並列性能

26,797,248 atom 7,536,726 atom 3,349,656 atom 418,707 atom

ISLIMソフトウェア研究開発報告会



102 342 342

817

6,542

52,338

Number of atoms per node

Number of nodes

Number of transfer cells

32,768

4,096

512

64

Strong Scaling

0.15

0.50

0.93

1.00

Scalability (to 64 nodes)

Other

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

7.38

11.48

Communication

1.92

10.77

124

26 110.54 91.53 13.69 5.31

Number of communication nodes

Time (ms / step)

Force

936

2,232 26 823.07 731.66 59.28 32.13

25.90 12.50



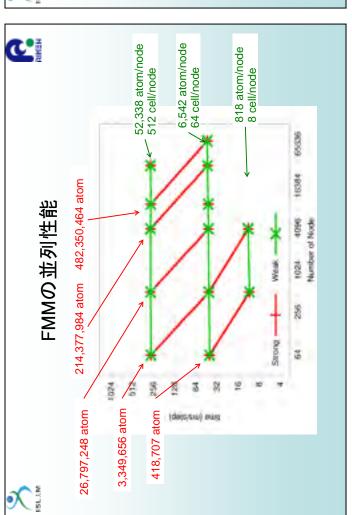


・Cutoff カーネルの演算効率60%以上を達成した。 ・全ノードCutoff法計算で、演算速度4Pflops・効率40%を達成した。

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。

ISLIMソフトウェア研究開発報告会



割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

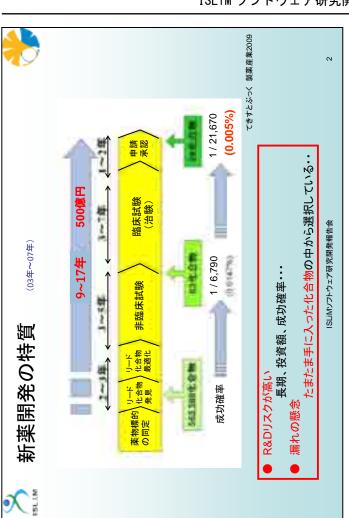
1992 年 4 月 豊橋技術科学大学工学部知識情報工学系 助教授

2004年4月 東京大学大学院工学系研究科 教授

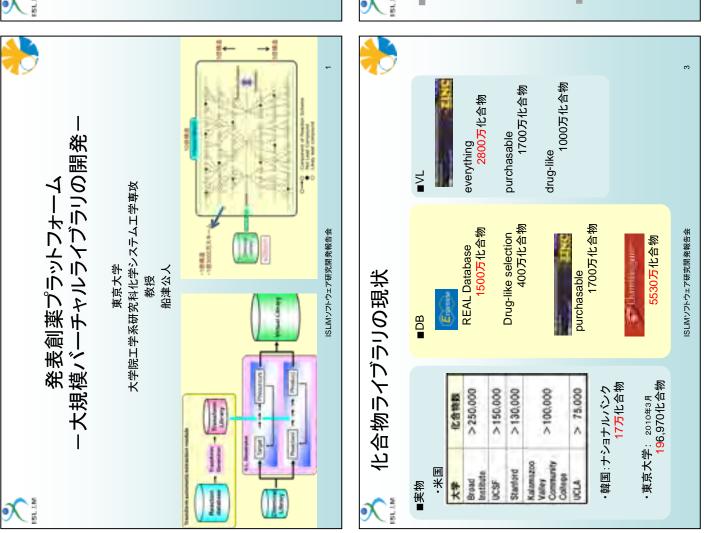
研究分野

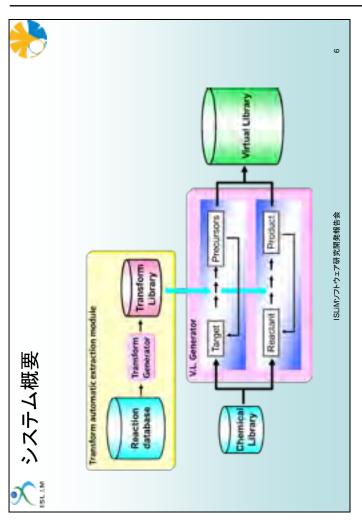
ケモインフォマティックス(化学情報学)

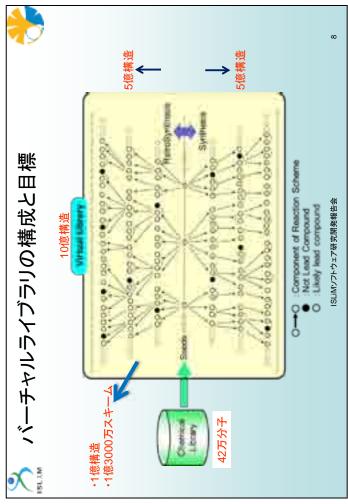
関連する研究分野としてプロセスシステム工学、有機合成化学、分析化学、触媒化学、 創薬化学など

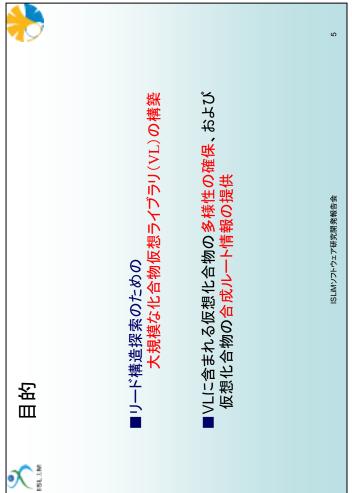


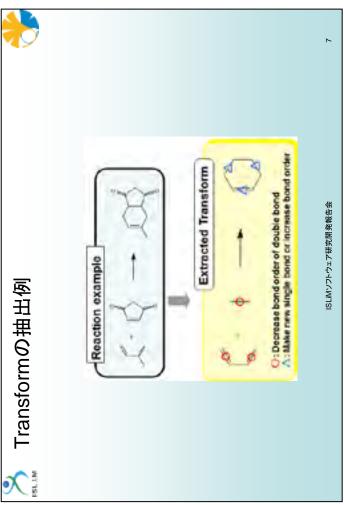


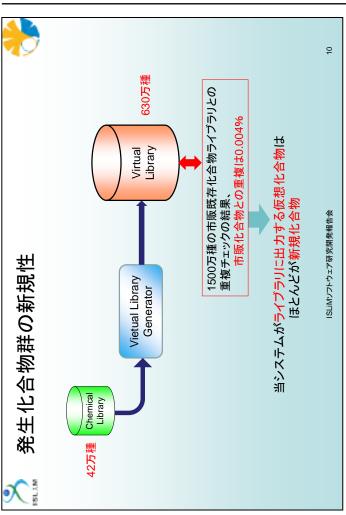


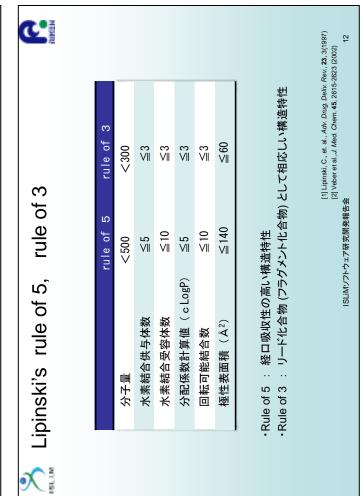


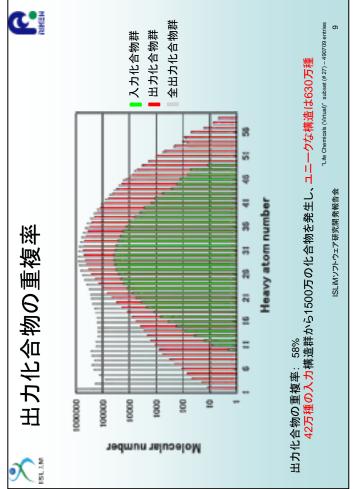


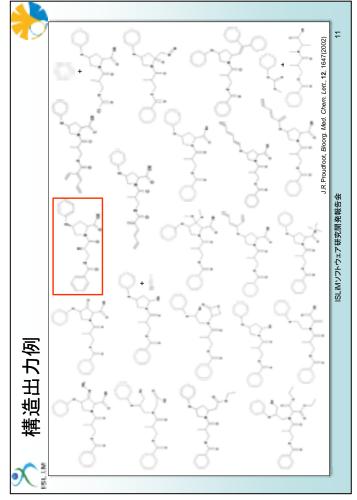


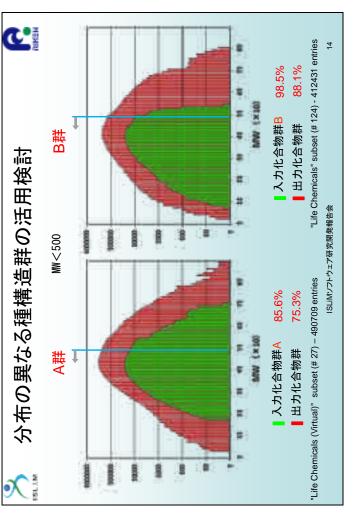


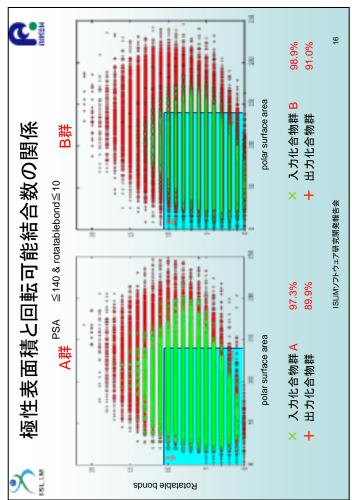




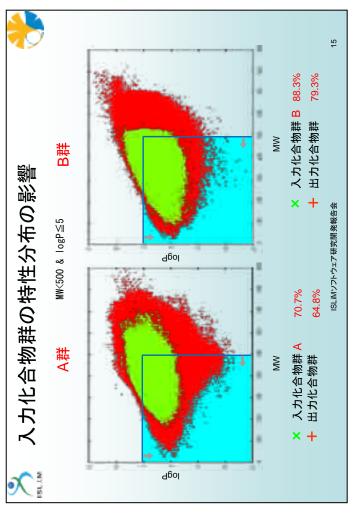














🥕 入力化合物群の特性分布の影響





まため

化合物群を種とし、反応DBより抽出したTransformを活用して 合成経路情報を含む大規模バーチャルライブラリを構築した。 創生した化合物は、入力構造群の特徴を引き継ぎ、 薬となりうる化合物を種とすると、薬となりうる構造を出力する傾向を確認した

■新規性:

市販化合物ライブラリとの対比で 重複0.004%、 新規性の確保を確認済み

入力構造群の分布を若干拡張するが、 特性を反映している 新規な化合物構造からなる出力構造群は、 ■分布: (Lipinski's ruleの指標によるチェック)

■規模:

入力化学構造群に対して、 順合成方向、逆合成方向にそれぞれ3段階 各方向で5億件の化学構造による

総10億化合物からなるバーチャルライブラリを構築し、ライブラリの特性評価を進める

ISLiMソフトウェア研究開発報告会

17

→ 適切な種構造群を活用すれば、適切な出力構造群が得られる可能性が増す

ISLIMソフトウェア研究開発報告会

出力構造の特徴分布は、入力構造群の特徴の影響を受ける

18





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



ISL:IM 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

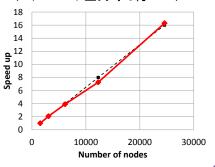


ライブラリの特徴

- 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

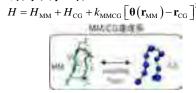
- 24,576ノード(196,608コア)まで計算 性能が直線的に向上
- ・ 高並列時でも高い実行効率を実現 (24.576ノード並列時で約12.3%)



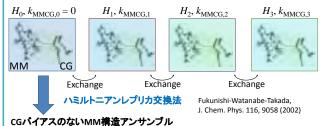
マルチスケール拡張サンプリング(MSES)法

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

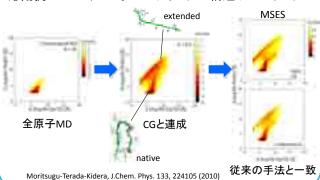
方法1: マルチスケール



方法2: マルチコピ



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

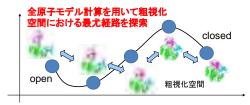


有限温度ストリング法

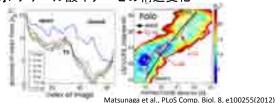
目的: タンパク質が構造変化する際の尤も確率の高い経路 を探索する。経路上の遷移状態を特定することによ り、構造変化のボトルネックとなる残基を同定する。

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



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 TRIKEN QBIC, 2 ARIKEN CSRP ISLIM, 3 RIKEN AICS, 4 RIKEN ASI



Introduction

Replica-exchange Molecular Dynamics (REMD)

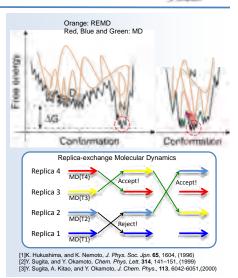
Replica-exchange molecular dynamics (REMD) simulation [2] is one of popular enhanced conformational sampling methods, and it has been widely used to predict the protein conformation, configuration and so on. In the original REMD simulation, the dynamics of noninteracting replicas of the system, each 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 extentions. 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:

$$min[1, exp(-\Delta)], \qquad \Delta = \beta_m(U_{\lambda m}(q_j) - U_{\lambda m}(q_i)) - \beta_n(U_{\lambda n}(q_j) - U_{\lambda n}(q_i))$$

where, $\beta_m = 1/(k_B T_m)$, T_m : temperature, k_B : Boltzmann constant, and , $U_{\lambda m} \left(q_i \right)$ 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.



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.



→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



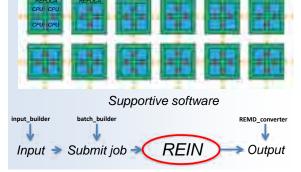
Now available Are http://www.islim.org/islim-dl_j.htm, Twitter: @rein_devel E-mail: yukimya+rein@gmail.com

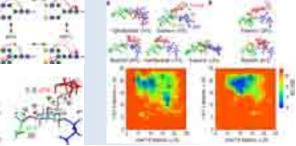
One of the applications

Structure prediction of N-glycans in solutionWe 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.







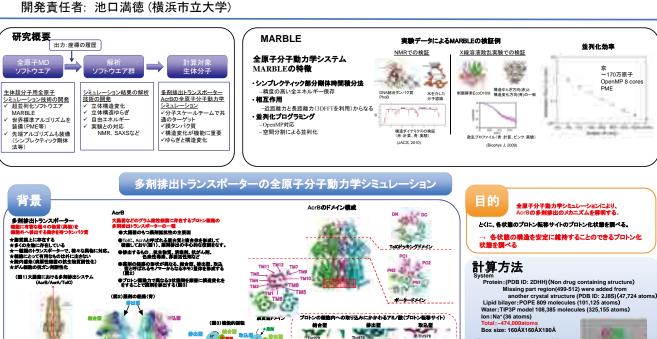




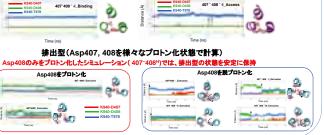
M-3: MARBLE

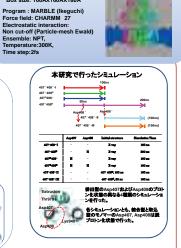
全原子分子動力学計算

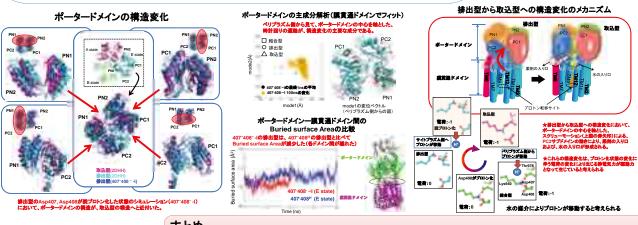












謝辞

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。

Acr63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、取込型のプロトマーは、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。

ボータードメイン

- 本作出型のポータードメインの構造が取込型に近づく過程が観察された(407・408・4)。 ★排出型から取込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、ペリプラズム側へ移動するスクリュ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気力であると考えられる



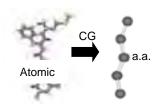
M-4: CafeMol

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

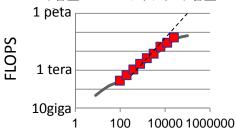


- 計算方法
 - Langevin方程式の時間発展を数値的に積分する。
- 並列化の方法
 - Neighbor list方式、レプリカ交換法
- 開発言語とライブラリ
 - Fortran90, MPI, OpenMP
- コードの公開状況
 - http://www.cafemol.org からダウンロード可。ただし生体膜、核酸については要相談。



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クラスタで98000コア並列
 - メモリ容量 2 GB/コア、ディスク容量 1 TB



#core

- 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1010回以上
 - メモリ容量 2 GB/コア、ディスク容量 1 PB

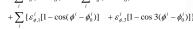
理論構築:粗視化モデルの導出

Atomic interaction-based CG model, Li, Wolynes, Takada, PNAS 2011

Li et al, PNAS 2012

$$V = \sum_{i} k_{b}^{i} (r^{i} - r_{0}^{i})^{2} + \sum_{i} k_{a}^{i} (\theta^{i} - \theta_{0}^{i})^{2}$$

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



$$+\sum_{i>i=3}^{native} \varepsilon^{ij} \left[5(r_0^{ij}/r^{ij})^{12} - 6(r_0^{ij}/r^{ij})^{10} \right] + \sum_{i>i=3}^{non-native} \varepsilon (C/r^{ij})^{12}$$

どんなことができるか

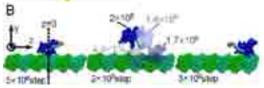
- □ X線回折やNMRによる構造情報が存在するタンパク質、核酸などの動態の分子動力学シミュレーション。とくに構造変化を伴うタンパク質ドッキング、分子モーターやトランスポーターの大規模構造変化などをシミュレーションできる。
- □ 例として、キネシンの動態、多剤排出 トランスポーターの動態、DNAヒスト ン複合系(ヌクレオソーム)など。

キネシンの動態



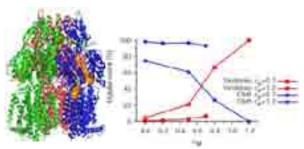
双頭キネシン: 高精度モデリング

単頭キネシン: 運動シミュレーション

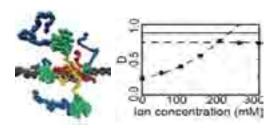


多剤排出トランスポーターAcrBの動態

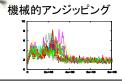
Yao et al, J Comp Chem 2010 Yao et al Nature Comm 2010



遺伝子動態:p53



遺伝子動態:ヌクレオソーム





为 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) 〇松田潤一. 平野敏行

背景

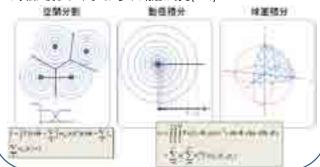
密度汎関数(DF)計算プログラムProteinDF[1]を京で 実行することで、これまでにない規模の生体分子全電 子計算が可能になる. DF計算における交換相関(XC) 項は被積分関数が複雑なため、グリッドによる数値積 分を行うのが一般的だが、その精度と演算量が問題 になる

$$\mu_{pq}^{NC} = \int g_p(r) \, \mu_{NC}(r) g_q(r) \, dr$$

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と 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.

$$N_{\text{integ}} = \int F(r)dr = \int \rho(r)dr$$

このとき数値積分は

$$M_{\text{integ}} = \sum_{G \in M} W^A W^T W^A \rho = \sum_{G \in M} W^A W^T W^A \sum_{pq} P_{pq} g_p g_p$$

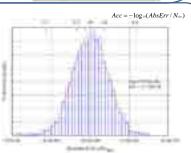
となり、電子数(整数)が真値となる。

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

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:イソロイシン 電子数:72 グリッド: Beckeの方法[2] 点数(50,194)

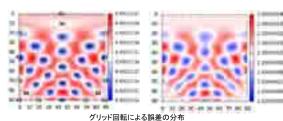
球面グリッドの回転

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

$$I = \sum w_{A} w_{C}$$

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



左: 等核2原子分子モデルでのの例 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

, 4, move						
	this work	worst	ΔΑcc			
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	ΔΑcc
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
lle	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 = -log so(AbsErr/Notes)

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、精度をあげる方法 を提案し、その有用性が示唆された

- [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、計算遂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

 $\mathsf{QM/MM} \wedge \mathsf{SJU} + \mathcal{I} \mathcal{T} \mathcal{Y} \qquad \hat{H}(\mathbf{r}, \mathbf{R}, \mathbf{X}) = \hat{H}^0(\mathbf{r}, \mathbf{R}) + \hat{H}^{\mathrm{QM-MM}}(\mathbf{r}, \mathbf{R}, \mathbf{X}) + E^{\mathrm{MM}}(\mathbf{X})$

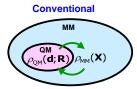
自由エネルギー汎関数 $F[\Psi_{\text{MF}}; \mathbf{R}] = -\beta^{-1} \ln \int d\mathbf{X} \exp(-\beta E[\Psi_{\text{MF}}(\mathbf{r}; \mathbf{R}); \mathbf{R}, \mathbf{X}])$

Fock or Kohn-Sham 方程式 $\hat{f}^{\mathrm{QMMM}}\left(\mathbf{r};\mathbf{R}\right) = \hat{f}^{0}\left(\mathbf{r};\mathbf{R}\right) + \sum_{n=0}^{n_{\mathrm{QM}}} \hat{q}_{A}\left(\mathbf{r};\mathbf{R}\right) \left\langle V_{A}\left(\mathbf{R},\mathbf{X}\right)\right\rangle_{\mathbf{X},E\left[\mathbf{q}(\mathbf{d},\mathbf{R});\mathbf{R},\mathbf{X}\right]}$

静電平均場 $\left\langle V_{A}\left(\mathbf{R},\mathbf{X}\right)\right\rangle_{\mathbf{X},E\left\lceil \mathbf{q}\left(\mathbf{d},\mathbf{R}\right);\mathbf{R},\mathbf{X}\right\rceil }=\left\langle V_{A}\left(\mathbf{R},\mathbf{X}\right)\omega\left(\mathbf{d},\mathbf{d}_{\mathrm{ref}};\mathbf{R},\mathbf{R}_{\mathrm{ref}};\mathbf{X}\right)\right\rangle_{\mathbf{X},E\left\lceil \mathbf{q}\left(\mathbf{d}_{\mathrm{ref}},\mathbf{R}_{\mathrm{ref}}\right);\mathbf{R}_{\mathrm{ref}},\mathbf{X}\right\rceil }$

 $\exp\!\left(\!-\!\beta\!\Delta\!E^{^{\mathrm{QM-MM}}}\!\left(\mathbf{d},\!\mathbf{d}_{_{\mathrm{ref}}};\mathbf{R},\!\mathbf{R}_{_{\mathrm{ref}}};\mathbf{X}\right)\!\right)$ $\omega\!\left(\mathbf{d},\mathbf{d}_{\mathrm{ref}};\mathbf{R},\mathbf{R}_{\mathrm{ref}};\mathbf{X}\right) = \frac{\mathrm{cap}\left(-\rho - \frac{1}{\rho - 1}\right)}{\left\langle \exp\!\left(-\beta \Delta E^{\mathrm{QM-MM}}\left(\mathbf{d},\mathbf{d}_{\mathrm{ref}};\mathbf{R},\mathbf{R}_{\mathrm{ref}};\mathbf{X}\right)\right)\right\rangle_{\mathbf{X},E\left[\eta\left(\mathbf{d}_{\mathrm{eff}},\mathbf{R}_{\mathrm{ref}},\mathbf{R}_{\mathrm{ref}},\mathbf{X}\right)\right]}$

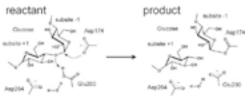
 $\frac{\partial F\left[\tilde{\Psi}_{MF};\mathbf{R}\right]}{\partial F\left[\tilde{\Psi}_{MF}(\mathbf{r};\mathbf{R});\mathbf{R},\mathbf{X}\right]} = \frac{\partial F\left[\tilde{\Psi}_{MF}(\mathbf{r};\mathbf{R});\mathbf{R},\mathbf{X}\right]}{\partial F\left[\tilde{\Psi}_{MF}(\mathbf{r};\mathbf{R});\mathbf{R},\mathbf{X}\right]}$ 自由エネルギーの核座標微分

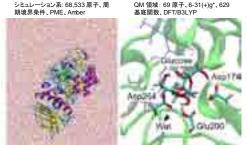


Present method ММ reweightii $\rho_{QM}(\mathbf{d};\mathbf{R})$

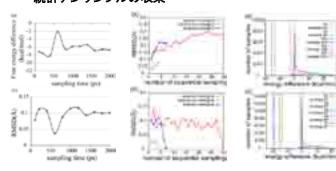
テスト系

α-amylase のグリコシド結合加水分解反応(の前半)

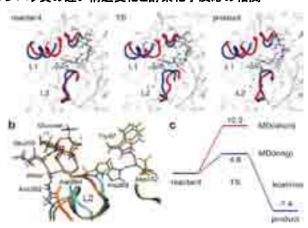




統計アンサンブルの収束



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

of basis functions: 888 # of atoms: 78,050

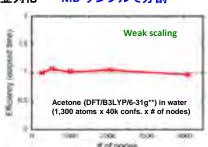
TS 生成に伴い switch-I ループが大きく曲がる





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)



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)

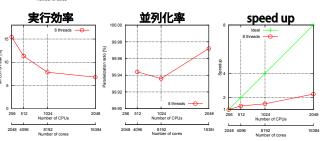
点電荷(30,904個)

RHF/cc-pVDZ 総数32,768コア 並列化率 99.9888% 実行効率 7.27~2.24%

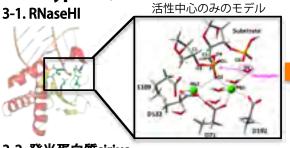
(8192コア: 4.26%)

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%)

speed up 実行効率 並列化率

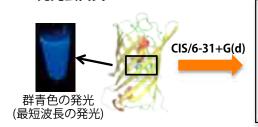


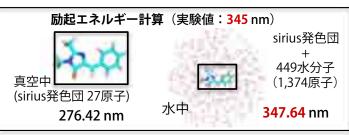
3. Platypus-QMの適用





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



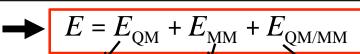
開発責任者: 中村春木(大阪大学蛋白質研究所)

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





Platypus-QM

HF, DFT, CASSCF, CIS, MP2 プログラムの高並列化による 長距離ポテンシャルの 高速化

Platypus-MM

並列性能に優れた新規 開発

▲線形応答関数解析 精度の高い

QM/MM境界の決定

(zero-multipole法)

Chain-of-State法 (疎結合並列計算)

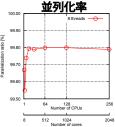
2.「京」での並列性能

RHF/cc-pVDZ

Strong scaling

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

実行効率

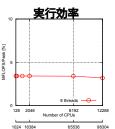


Ace-ALA-NMe 1,080 水分子

Weak scaling (Chain of State) 総数98,304コア

128プロセスx768レプリカ 実行効率 3.42~3.20%

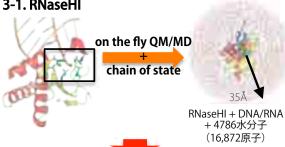
Weak scaling効率: 93.36%





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 + 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



😕 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)





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

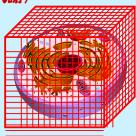
Our Approach: 細胞シミュレーション統合シミュレーションプラットフォームの開発

細胞シミュレーションフラットフォーム 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)

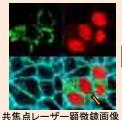
- ・連続体として細胞内を表現
- 空間をボクセル空間で表現
- ・細胞内小器官などの複雑な 空間構造にも対応
- ・複数細胞を同時に計算可能
- ・細胞膜の機能を表現可能



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。

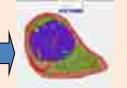


顕微鏡画像から作成した3次元形状内での計算





細胞内カルシウム緩衝反応



3次元再構築した細胞形状



反応:24反応 物質数:10個

計算条件

時間

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

Ca²⁺の拡散係数:1×10⁻⁹(m²/s)





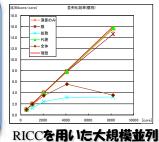
反応があると見かけの拡散速度が 遅くなる



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





計算結果(左:反応あり、右:反応なし)

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。



为 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。

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 Our Approach:

(RICS-Pre) の開発

RICS Solver で必要な設定ファイル



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
- ✓ 及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

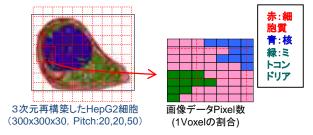
単純な基本形状はCADを用いずRICS-Pre で作成

√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

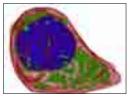
Voxelデータの解像度変更

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

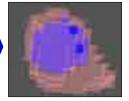


RICS-Pre07D-





 $300 \times 300 \times 30$ (=2,700,000 voxels)



26 × 26 × 15 (=10.140 voxels)

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



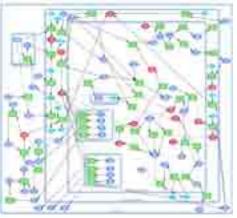
七澤 洋平、田村 典子、塩崎 聖治、岡 秀樹、後藤 信哉(東海大学医学部 臓器/細胞ス ケール研究開発チーム)

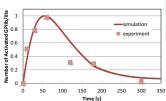
細胞内部の代謝反応/シグナル拡散に着目し、細胞内の物質/オルガネラの位 置情報を考慮した血小板細胞内での代謝拡散反応について、理研細胞シミュ レーション統合プラットフォーム(RICS)を用いての計算機によるシミュレーショ ンを行なった。RICSとは、細胞を複数のボクセルに区画し、そこに実測データよ り得られた細胞内の物質量/移動量などの情報を取り込み、細胞内の現象を表 現可能とする細胞シミュレーション統合プラットフォームである。

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GP1bα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。





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。

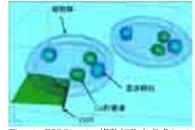


Fig. 1. RISCにて、複数細胞を考慮して の血小板細胞モデルの構築

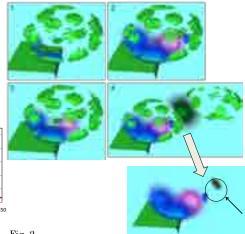
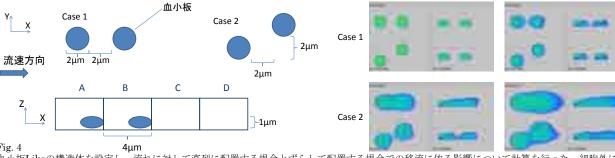
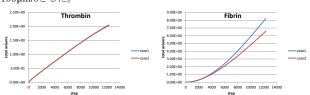


Fig. 2.

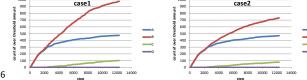
- 1. 局在するGP1baがvWFにより刺激される
- 2. GP1baから放出されるシグナル物質による刺激に よりCa貯蔵庫からCaが放出
- 3. 細胞内Ca濃度上昇によって濃染顆粒が刺激される
- 4. 濃染顆粒に含まれる物質の放出。周囲細胞へと活 性化に繋がる物質を伝える。また、自らのレセプタ-を刺激することで活性化を促進させるフィードバック 反応も引き起こす。矢印で示した箇所が濃染顆粒から 矢印で示した箇所が濃染顆粒から の放出箇所周囲のADPレセプターP2Y12の反応である。



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依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。 $PT+AP \to T+AP\dots(1)$, $FG+T \to F+T\dots(2)$ (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 1 μM が存在し、血小板付近の流速は最大 100μm/sとした



-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。



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



Ž 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

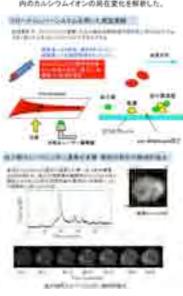


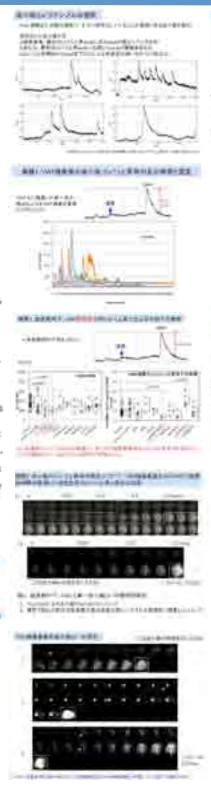
田村典子1) 七澤洋平1) 後藤信哉1)

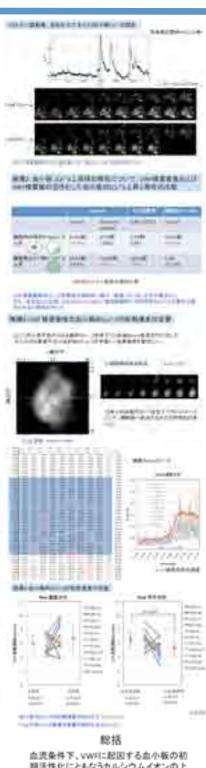
1)東海大学医学部内科学系循環器内科 背景 カルシウムイオン(Ca²)は細胞内のシグナル伝達では、 セカンドメッセンジャーとして個く重要な因子である。 血小板では活性化物質により刺激を受けると、初期活 性をとして、カルシウムストアであるが放体からなごが 放出される。繊胞内Cが湿度(Cが))の上昇により、血 小板の退性化シグナルは意思される。なかでも血小 板炭集を引き起こすARP、Cが1速は装置物から、血 栓形成に必要なフィブリノーゲンやVon Willebrand別 そいがり、各種深固肥子などは繊維から細胞が小数 セスカ 別用をはちゅかればは細胞から細胞が小数 出され凝固を伴う血栓形成は細筋内外のCs**により促 AIRCARDS 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)反応時間と遺皮 2)カルシウムイオン局在の経時的変化 3)血小板の[Ca^{3*}]/拡散速度 を定量し、RICSを用いた血小板活性化シミュレ

ションモデルの実証実験として明らかにする。

- フローチャンバーを用いた血洗条件によるvwr上の血 小板内カルシウムイギンの検力を異常点レーザー顕 機関にて行ったリアルライムイメージングを行った。 健常人より理血し、肌トロンピン薬(アルガトロ・Cン)に て飲み園気質を行った。全血中の血小板にカルシウム ム電光推示薬Pao-3 AMを導入し血染サンブルとした。 ヒトの血質タンパクから検製したvon Willebrand 窓子 (vwn)を固幅化したガラスフレート上に、ヒトの知販血 炭に相当する過度1500 %にて全血を運流した。 VWF上に指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速度の結時的変化より細胞 内のカルシウムイオンの具在変化を検討した。







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。

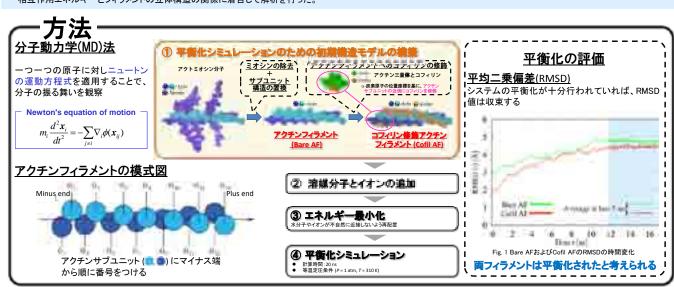
😕 C-5



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

藤井徹矢」・井上康博「2・安達泰治「、2(「京大院・エ・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)

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応答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、その振る舞いが変化する。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。



ねじれ角

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

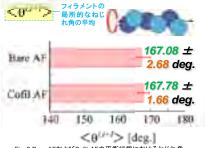
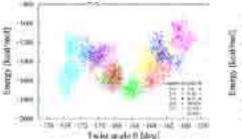


Fig. 2 Bare AFおよびCofil AFの平衡状態におけるねじれ角

- 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



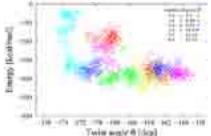


Fig. 3 Cofil AFにおけるねじれ角とコフィリン―アクチン間エネルギーの関係 Fig. 4 Cofil AFにおけるねじれ角とアクチン—アクチン間エネルギーの関係

- 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。

- 、次の特徴か見られた。 アクチン同士の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) 平衡化計算が期の5 nsでは、両サブユニット間 に取ばされるアナま社会の物は、四サブユニットに
- 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 サブュニット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。

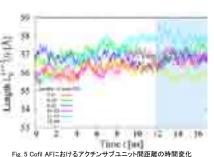






Fig. 6 Cofil AFのサブユニット11-13間において、初期状態か)の5ns間で水素 結合の形成頻度が最も高かったアミノ酸残基のペア (a) t = 2 ns (b) t = 17 ns

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) および文部科学者科学研究費補助金(22616602,23127506) による援助を受けた。記して謝意を表する



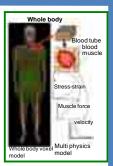
🦰 O-1: ZZ-EFSI 全身ボクセルシミュレーション (ボクセル構造流体連成解析プログラム)



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・再構成を不要に.
- 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.

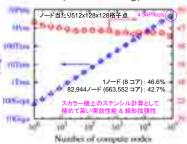
現時点での開発状況

- 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)





計算の妥当性 (流体構造/膜連成機能の実

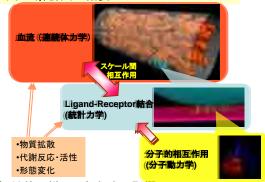
キャビティ流れ中の超弾性体粒子 せん断流中の膜カプセル 変形パラメータDの時間変化 Eggleton et al.(1998) Phys. Fluids,10, 1834
 △ Pozrikidis (1995) J. Fluid Mech.,297, 123. (2009, *J. Comput. Phys.* 227) SMAC+ Lagrangian for solid Present (ACM + full Eulerian)

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

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/ フィジックス解析法の概要



血小板付着の様子、赤血球の影響

赤血球なし 赤血球あり 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

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

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., 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-
- 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.

li, 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.

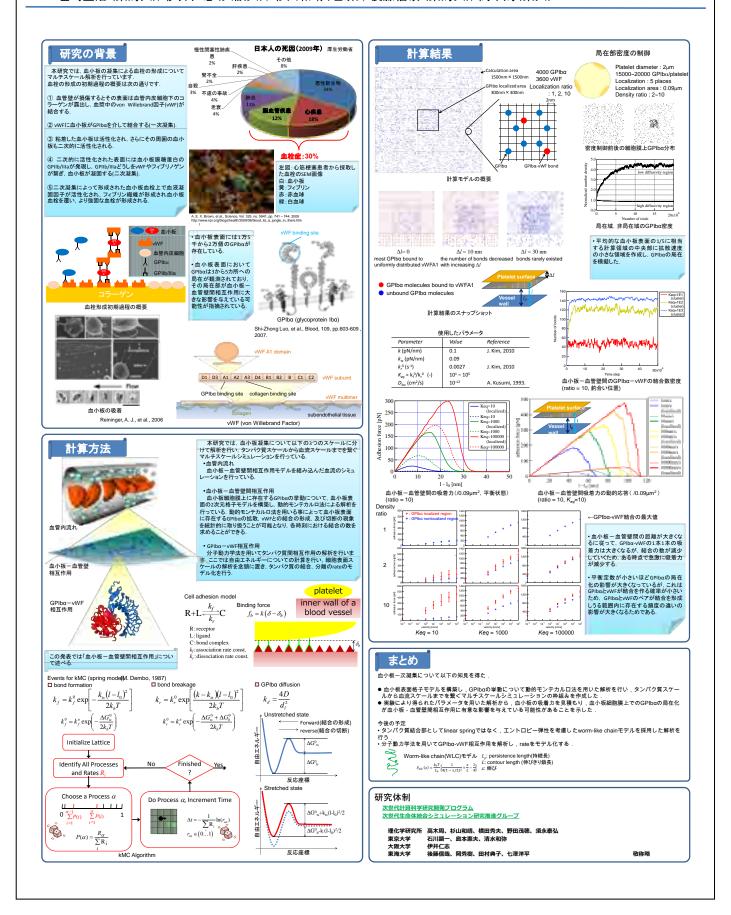




血栓形成のマルチスケールシミュレーション

- 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)



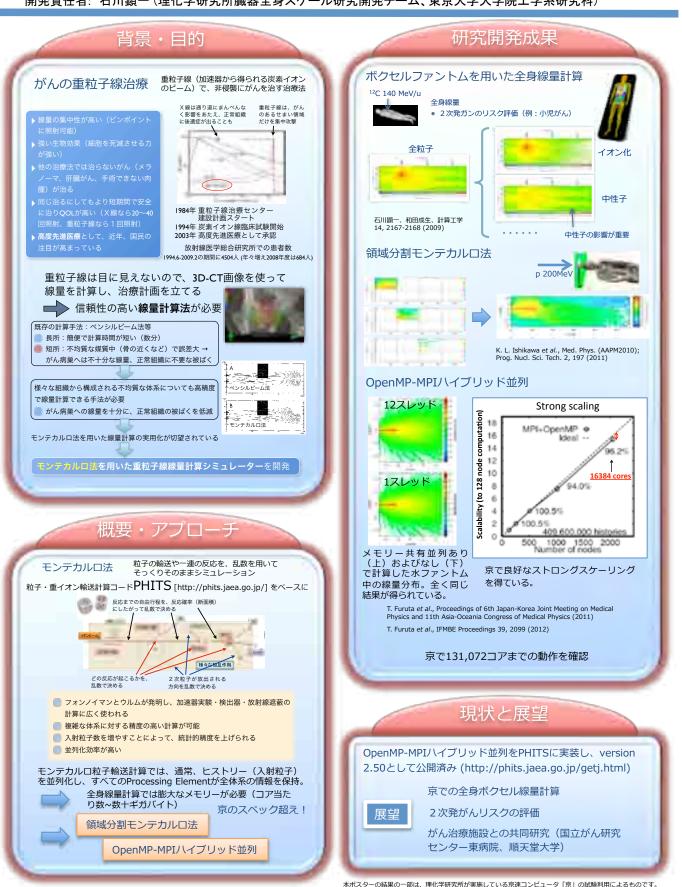


$\red{>}$ O-3: ZZ-DOSE based on PHITS



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)



◇O-4: ZZ-HIFU 集東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

- mally-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





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

- 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

 $\rho_m \frac{\partial \mathbf{u}}{\partial t} = -\nabla p + \nabla \cdot \left[\zeta_{m} \text{tr}(\mathbf{e}) \mathbf{I} + 2\mu_m \left(\mathbf{e} - \frac{1}{3} \text{tr}(\mathbf{e}) \mathbf{I} \right) \right] + \nabla \cdot \sigma_s'$

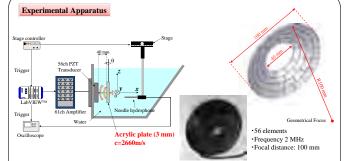
The constitutive equation for shear elasticity: $\frac{\partial \sigma'_s}{\partial t} = 2\mu_{em} \left(\mathbf{e} - \frac{1}{3} \operatorname{tr}(\mathbf{e}) \mathbf{I} \right)$

 $\frac{\partial p}{\partial t} = -\zeta_{out} \mathbf{tr}(\mathbf{e}) = -\left(\rho_{oc}c_{ou}^2 - \frac{4}{3}\mu_{ou}\right)\nabla \cdot \mathbf{u} \quad \text{where} \quad \rho_{ou} = \sum_{k} f_{k}\rho_{k} \quad \text{and} \quad \frac{1}{\rho_{oc}c_{ou}^2} = \sum_{k} \frac{f_{k}}{\rho_{k}c_{out}^2}$

 $\rho_{m}C_{m}\frac{\partial T}{\partial \tau} = \frac{\partial}{\partial x_{i}}\left(\kappa_{m}\frac{\partial T}{\partial x_{i}}\right) + \zeta_{m}\left(e_{kk}\right)^{2} + 2\mu_{m}\left\{e_{ij}e_{ij} - \frac{1}{3}\left(e_{kk}\right)^{2}\right\}$

- 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.

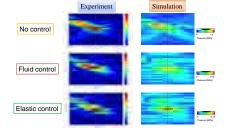
3. Simulation Assisted Focus Control -

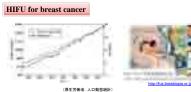


The acrylic plate as a phantom of bone is inserted between the transducer and target. The angle of the acrylic plate can be changed. The transducer is driven by 61ch amplifier under the control of LabVIEW. The acoustic field is measured by using needle hydrophone 3-dimensionally

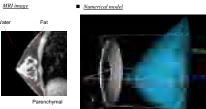
Influence of elasticity on focusing Water Water Acryl Longitudinal wave

ults of experiment and simulation are in relatively go



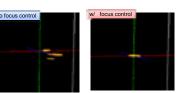


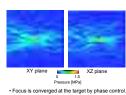






- Frequency 2 MH







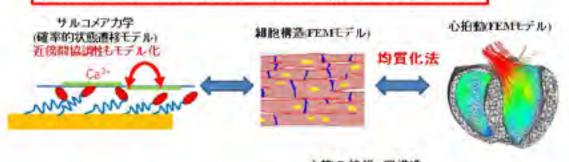
O-6: UT-Heart

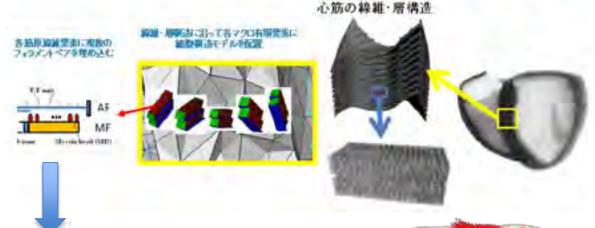


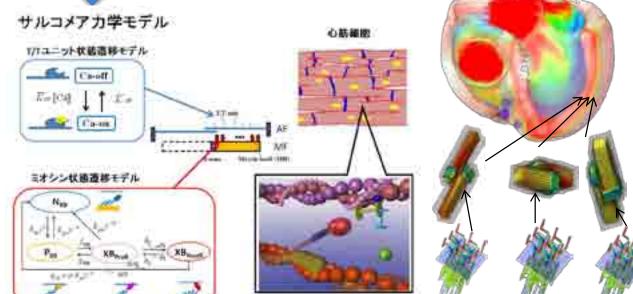
UT-heartによるサルコメアカ学から心筋細胞構造を経て心拍動にいたる

マルチスケール解析 開発責任者: 久田俊明(東京大学)

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







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

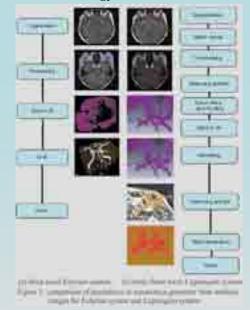
• Simplified s-shaped geometry for femoral artery (Model I)



· Realistic geometry for carotid artery (Model II & III)



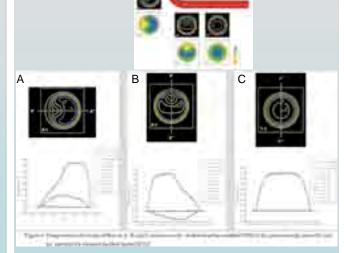
Numerical Methodology



- 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 (Remean = 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].



Spatial Convergence test with the double bend model

Method I. Estimation of accuracy order by substituting the equations as below.

Method II. Grid Convergence Index (GCI) method is based on estimated fractional error.

Method III. Estimation of accuracy order with an assumption that the results on the finest voxel is very close to exact solution.



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



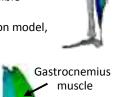
3D FE Simulator for Mechanical Analysis of Human Musculo-skeletal System

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.

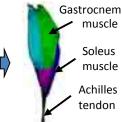
3D FE model of musculo-skeletal system

- 3D geometry and architecture of musculoskeletal system reconstructed from CT/MR images
- Nonlinear FE method for incompressible hyperelastic materials
- Phenomenological muscle contraction model, Hill model and Motor neuron model









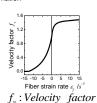
Muscle contraction model (Hill model)

$$\sigma_{\textit{fiber}}^{\textit{act}} = \sigma_0^{\textit{isom}} \cdot \frac{A_0}{A} \cdot f_t \cdot f_l \cdot f_v$$

where $egin{dcases} \sigma_0^{ ext{isom}} &: ext{Maximum isometric stress} \ A_0, A &: ext{Cross-sectional area of a muscle fiber} \end{cases}$

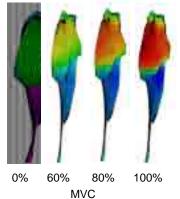






Simulation results of a triceps surae muscle

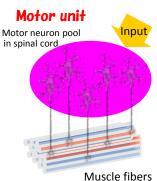
Displacement distributions Nodal forces Joint movement







Micro-mechanical model of muscle contraction

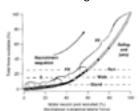


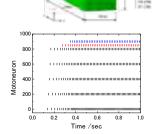


安静時 50% iEMGmax 70% iEMGmax (Kinugasa et al. 2006)

Motor neuron model

- Recruitment
 - Rate coding





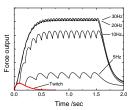
Impulse response for each discharge

$$f_{i,j}(t) = \frac{P_i t}{T_i} e^{1 - (t/T_i)} \cdot g_{i,j}$$

Twitch contraction time P_i: Maximum twitch force : Discharge frequency

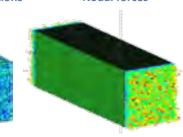
Response of continuous discharges

$$F_i^{motor}(t) = \sum_{j=1}^k f_{i,j} \quad (t - t_{i,j} \ge 0)$$



Simulation results of isometric contraction

Cauchy stress distributions



Nodal forces



○ O-9:血管網シミュレーター ○ クラ・カー・フェークー ○ クラ・カー・フェーク ○ クラ・カー・ファーク ○ クラ・カー・フェーク ○ クー・フェーク ○ クー・フェーク ○ クー・フェーク ○ クー・ファーク ○ クー・フェーク ○ クー・ファーク ○ クー・フェーク ○ クー・ファーク ○ クー・ファーク



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

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)

ロステント留置術

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

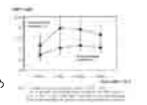
□過灌流症候群

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

□最終日標

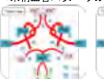
手術計画を目標とした術前術後の血行動態の予測のため の全身循環を考慮した血流シミュレーションの開発

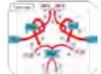


前本ら 経頭面カラードプラによる過凍流症検算の早期診断 独立行政法人国立病院機構九州医療センター 脳血管センター・臨床研究部脳血管の

PC-MRII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-→血管半径、血管長さの計測

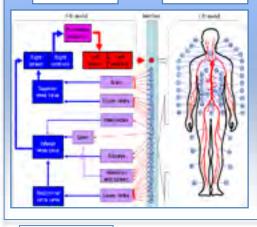




マルチスケール血流解析の概要

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

$$\frac{\partial q}{\partial t} + \frac{\partial}{\partial x} \left(\frac{q^2}{A} \right) + \frac{A}{q} \frac{\partial p}{\partial x} = -8\pi v \frac{q}{A}$$

•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】

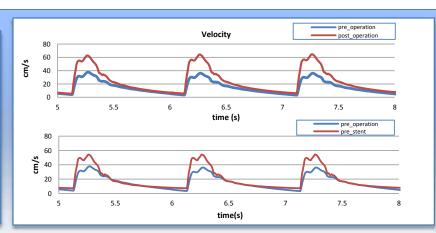
手術後の患者の血管形状/末梢データ

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



左中大脳動脈において、平均流速の増加が確認され、過潅流症候群を発症した計測例と同様の傾向見られた。



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



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。

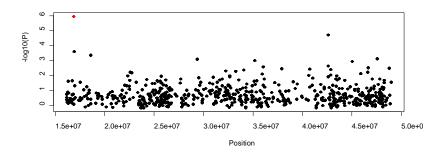
ParaHaploの開発

- 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



次世代シークエ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

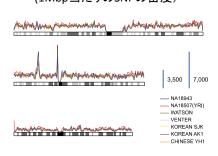
次世代シークエンサーの出力データを高速に解析し、ヒト個人間の遺伝的差異やがんゲノムの突然変異を高い 正確さで同定する。

- 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
- 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クエンス解析(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。

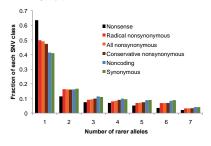
ヒトの遺伝的多様性 (1Mbp当たりのSNPの密度)



がんの突然変異 (染色体異常とコピー数異常)



SNPの頻度スペクトラム



参考文献

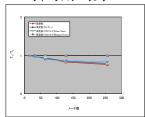
- 1. *Fujimoto A, Nakagawa H, Hosono N, Nakano K, Abe T, Boroevich KA, Nagasaki M, Yamaguchi R, Shibuya T, Kubo M, Miyano S, Nakamura Y, Tsunoda T (2010) Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing. Nat Genet 42: 931–6
- 2. The International Cancer Genome Consortium (2010) International network of cancer genome projects. Nature 464, 993-8.
- 3. 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

データの分割と解析

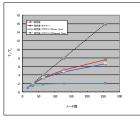
| Image: Configure ordinary | Manage: Configure ordinary | Ma

「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





D-3: ExRAT



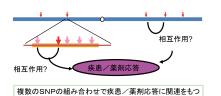
拡張RAT法による2SNP組合せの全ゲノム関連解析ソフトウェア

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

相互作用



単一SNPの解析で相互作用の解析は困難

Logistic回帰と対数尤度比検定

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 + χ²検定 → 尤, → 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クラスタで8,192ノードの並列性能を実証.
- シファスス Co, 1927 一トの业列性能を実証. 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:分割表探案処理での見直しにより、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).
- ~ 元ル・ハランバ・ソフ・い人とでと吸が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.

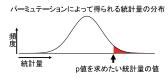
目的と方法のまとめ

- 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 x 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

Importance Sampling



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あわせる -- 赤の比率

WTCCC type 2 diabetes Sample size 1,924 cases + 2,938 controls #SNPs (MAF>0.05, after QC) 360,297 6.49x10¹⁰ #combinations #runs 25 Cores used / run 24,576 1.5TB Disk space (stage in) / run 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

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め
- 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

RAT法の拡張

より複雑な解析に対応するために

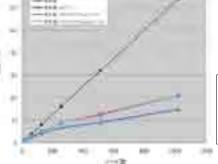
プリプロセス m-1)/2個のSNP組み合わせ

からB個を選択

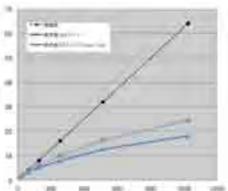
各種線形回帰モデルと尤度比検定への拡張

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 $logit(P) = \mu + \alpha x_A + \beta x_B + \gamma x_A x_B$

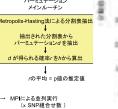




「京」上のall_comb



----「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べ 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-BN はこれらのうちべ 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



マイクロアレイなどによる遺伝子発現データ ノックダウン実験や患者サンプル 薬剤応答時系列データ

大規模遺伝子祭現データからの大規模遺伝 子ネットワークの推定には膨大な計算が必 要である。したがって遺伝子ネットワーク 解析にはスーパーコンピュータによる大規 模計算が欠かせない。

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



 $f(X_1, X_2, \cdots, X_6)$

 (X_6)

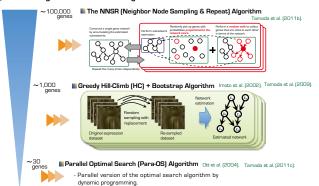
The network structure is determined by maximizing the posterior probability given the expression data

 $\pi(G \mid X) \propto \pi(G) \iiint_{i=1}^{n} f(x_{ij} \mid \boldsymbol{p} \boldsymbol{a}_{ij}, \boldsymbol{\theta}_{j}) \pi(\boldsymbol{\theta}_{j} \mid \lambda) d\boldsymbol{\theta}_{j}$

The parent-child relationship is modeled by the B-spline nonparametric regression $\lambda \ll_{m_{p}^{2}}$ $x_{ij} = m_{ji}(p_{ii}^{(j)}) + \ldots + m_{j_{q_i}}(p_{iq_j}^{(j)}) + \varepsilon_{ij}$ $m_{jk}(p_{ik}^{(j)}) = \sum_{j=1}^{M_{jk}} \gamma_{ik} b_{ik}^{(j)}(p_{ik}^{(j)})$

Parallel Algorithms for Estimating Gene Network Structures

 $= f_1(X_1) f_2(X_2) f_3(X_3 \mid X_1) \cdots f_6(X_6 \mid X_3, X_4)$



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

● 入力

入力データ:マイクロアレイなどの遺伝子発現データ 例 静的: ノックダウン実験、患者由来サンプル 動的: 薬剤応答時系列データ

サンプル (アレイ) 数:50 ~ 1,000 アレ 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:100~100,000ノード

●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 クラスタ)

● 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

これまでの応用事例

● 肺癌細胞株 PC-9+Gefitinib 投与による薬剤応答パスウェイ解析

細胞:肺癌細胞株 PC-9

: Gefitinib(イレッサ: 抗がん剤) : Agilent 44K Whole Human Genome Oligo Microarray 薬 マイクロアレイ

計測データ: 26 時点×1 回計測=26 アレイデータ FGF.Gefitinib 投与・非投与の 4 条件

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 ネットワー 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

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].

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 8 siRNA ノックダウンアレイ =+ 32 アレイデータ

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

siRNA ノックダウンアレイによる事前情報を用いた 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]

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速コンピュータ「京」の試験利用によるものです

Reference

Affara, M. et al. (2007). Understanding endothelial cell apoptosis: What can the transcriptome glycome and proteome reveal?, Philosophical Transactions of Royal Society 62 (1484), 1469-1487.

Araki, H. et al. (2009). Analysis of PPARa-dependent and PPARa-independent transcript 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*.

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. tt, S. et al. (2004). Finding optimal models for small gene

networks. Pac. Symp. Biocomput. 9, 557-567. mada, Y. et al. (2009). Unraveling dynamic activities of autocrine pathways that control drug-response transcriptome networks, Pacific Symposium on Biocomputing (PSB2009) 14,

networks, Pacific Symposium on Boschippenage 251-263.

Tamada, Y. et al. (2011a). SiGN: Large-scale gene network estimation environment for high performance computing, Genome Informatics, 25 (1), 40-52.

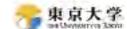
Tamada, Y. et al. (2011b). Estimating Genome-wide Gene

Networks Using Nonparametric Bayesian Network Models on Massively Parallel Computers. IEEE/ACM Trans. comput. biol. bioinformatics, 8 (3), 683-697.

Tamada, Y. et al. (2011c). Parallel algorithm for learning optimal Bayesian network structure, Journal of Machine Learning Research, 12, 2437–2459. Wang, L. et al. (2012). Cell cycle gene networks are associated with melanoma prognosis, PLoS ONE, 7(4),



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

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。

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



- 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.

- iniformatics, 25(8), 1064-1072.

 minimum et al. (2010) Collocation-based sparse estimation for inferring continuous-time dynamic gene networks.
 nonne loformatics, 24, 1564-178.

 immumura et al. (2011). A novel network profilling analysis reveals system changes in epithelial-mesenchymal transitio
 d5 ONE 6(6): e20504, doi:10.1317/journal.pone.0020804, 2011.

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



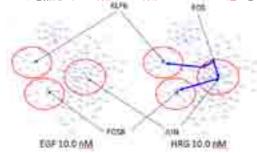
- ・3 種類の統計的グラフィカルモデル (グラフィカルガウシアンモデル、ベクトル 自己回帰モデル、構造方程式モデル)により、遺伝子発現データから遺伝子 ネットワークモデルを構築
- L1 正則化法を用いてスパースなモデルのパラメータ空間を探索することによ り、100,000 ノードまでの大規模遺伝子ネットワーク推定が可能
- ・次世代スパコン「京」上で、最大並列数(コア数):98,304、最大ノード数: 12,288、実行効率: 1コアで11.98%を達成。なお、ハイブリッド並列化には対応
- ベータ版のバイナリを公開開始 http://sign.hgc.jp/signl1/index.html

これまでの応用事例

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)

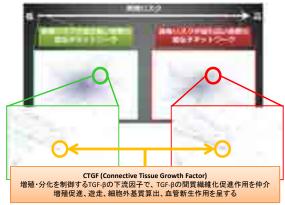
目的: EGF、HRG各々の成長因子によって誘導される8実験条件下での遺伝 子ネットワークを推定し、EGF刺激、HRG刺激下のシステムの違いを比較



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

-タ: 226 症例の肺線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

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べ 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

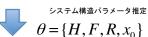
状態空間モデルによる時系列遺伝子発現データからの情報抽出

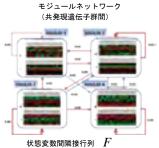
状態空間モデル (State Space Model)

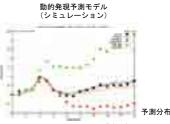
次元縮約による 高次元短時系列 遺伝子発現データからの システム推定

p≈10³ 遺伝子数 N < 20 時点数

遺伝子ネットワーク

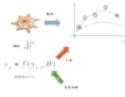






システムの模倣と予測

遺伝子間隣接行列 $\Psi = D' \Lambda FD$



 $P(y_n | y_1, \dots, y_{n-1}; \theta)$

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
- 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、欠測および不等間隔計
- 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)
- データ解析融合プラットフォーム SBiP との連携

●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

●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

これまでの応用事例

● 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×1 回計測 in 48 時間 EGF,Gefitinib 投与・非投与の 4 条件 ■目的

Gefitinib の影響により遺伝子間制御構造に 変化のあった遺伝子群の予測および、 肺がん予後予測パイオマーカ探索

■ 利用手法

コントロールデータにより構築されたモデルからの ケースデータの予測に基づくシステム異常検知

■ 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



発現制御構造推定

制御構造に差異のある 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 Gegitinib 動的効果予測

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×1 回計測 in 26 時間

■目的

葉剤耐性能の異なるがん細胞への薬剤作用点 および作用機序の予測: PC9(薬剤感受性) PC9GR(薬剤耐性)

■ 利用手法

薬剤動的効果モデル項を組み込んだモデルによる 薬剤作用下データの予測に基づく状態推定

PC9 と PC9GR の間で異なる薬剤効果プロファ イルを持つモジュールを同定. 当該モジュール支配下にある遺伝子群が 薬剤耐性能の差異に関わる可能性を示唆

薬剤効果プロファイル PC9

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 (RICC) system, and early access to the K computer at the RIKEN Advanced Institute for Computational Science.

uchi and Higuchi. State-space approach with the mum likelihood principle to identify the m-generating time course gene expression data ast. Int J of Data Mining and Bioinformatics, 1(1):

Reference

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).

Do et al. Exploring temporal transcription regulation structure of Aspergillus fumigatus in heat shock by state space model. BMC Genomics, 10: 306, (2009). Space of the state of the space models and simulations by biological state space models: search for drug mode of action, of Comp Sci and Tech, 25(1): 131-153, (2010).

Tamada et al. SiGN-SSM: open source parallel software for estimating gene networks with state space models.

Bioinformatics, 27(8): 1172-1173, (2011).

Tamada et al. SiGN: large-scale gene network estimation environment for high performance computing. Genome Informatics, 25(1): 40-52 (2011).

Yamauchi et al. Epidermal growth factor receptor tyrosine

kinase defines critical prognostic genes of stage I lung adenocarcinoma. PLoS ONE, 7(9): e43923EP-, (2012).



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すしている。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



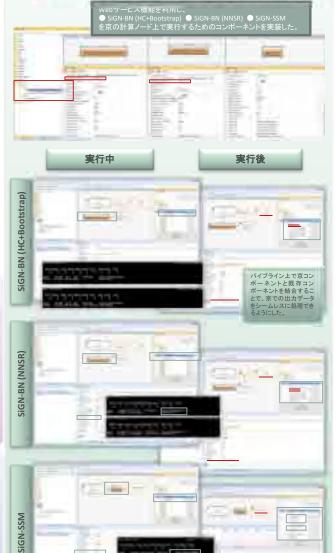
SBiP オンラインドキュメント整備



http://sbip.hgc.jp

ODIF オンフィントモュメントに 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



肺がん・乳がん関連データ解析 Gefitinib 感受性肺がん細胞株と耐性株のネットワーク比較





为 D-8: LiSDAS



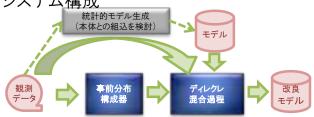
生命体シミュレーションのためのデータ同化技術の開発

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)

1. 背景

- 実験にもとづくタンパク質間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



• 転写制御回路

転写活性 転写抑制

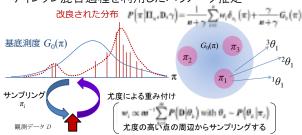
 $dx_j/dt = -{}^{d}k_j + {}^{tc}k_jA(x_{a(j)}|c_j, p_j)R(x_{r(j)}|d_j, q_j) + {}^{b}kx_{b(j,1)}x_{b(j,2)}$ • $A(x | c, p) = x^p/(x^p + c^p)$, $R(x | d, q) = c^q/(x^q + d^q)$

モデル

- モデルパラメタ $\theta = (x_i(0), {}^{d}k_i, {}^{c}k_i, {}^{b}k, c_i, p_i)$ の平均・分散
- 観測時系列 D = (y_i,)

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



- 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 ~ b": 分布bからサンプリングしてaに代入

For i = 1,..., N [N $\sim 10^{6}$]

- For j = 1,..., m: $\theta^{(j)} \sim p(\cdot | \pi^*)$. [m = 128]
- $w^* = \sum_i p(D|\theta^{(i)})/m.$ $[p(D|\pi^{(i)}) \approx w^*$ の近似に使う]
- If $i \le n$ then [保持する π 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
- $\theta^* \sim \sum_{j} p(D|\theta^{(j)}) \delta(\theta \theta^{(j)})/m$.
- $\pi^* \sim p(\pi | \pi^{(1)}, ..., \pi^{(n)}, \theta^*)$
- $\propto 1/(\gamma + n) \times \sum_{i} p(\theta^* | \pi^{(i)}) \delta(\pi \pi^{(i)})$
- $+ \gamma/(\gamma + n) \times p(\theta^*|\pi) G_0(\pi).$
- If $\exists k \in \{1,...,n\}$. $p(D|\pi^{(k)}) < p(D|\pi^*)$
 - $\pi^{(k)} := \pi^*, \ w^{(k)} := w^*.$

3. 性能測定

哺乳動物の概日周期転写ネットワーク



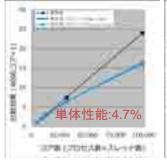
- モデルの規模
- システム変数の個数: 29
- 観測変数の個数: 7
- モデルパラメータ数: 116
- 観測をトレースするパラメータ が発見できることを確認。

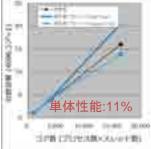


台数効果

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 大学院情報理工学研究科 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) 非小細胞肺がんパスウェイ 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 P1 P2 P3 P4 P5 **P1** P2 P4 **P5** OXOOX U 30 711 MOXXOX POXOXO P1 XXXXXX

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 新規に提案したスコア関数 ドッキング精度と計算速度の検証 R(l,m,n)つの被害数で表現 $S(\alpha,\beta,\gamma)= \Re \left[\sum_{l=1}^{K} \sum_{m=1}^{K} \sum_{n=1}^{N} R(l,m,n) L(l+\alpha,m+\beta,n+\gamma) \right]$ $R(l, m, n) = G_R(l, m, n) + \frac{\mathbf{w}_h H(l, m, n)}{\mathbf{w}_h H(l, m, n)} + \frac{i\phi(l, m, n)}{\mathbf{w}_h H(l, m, n)}$ $L(l, m, n) = G_L(l, m, n) + \frac{i\mathbf{w}_h Q(l, m, n)}{\mathbf{w}_h Q(l, m, n)}$ **藤水性相互作用** 既存手法と同等の精度で L(l,m,n)複合体ポーズを予測することに成功 (Benchmark 4.0, 176 bound complexes) 1回転角のスコア (3600通り計算) $L(l+\alpha,m+\beta,n+\gamma)$ $O(N^6)$ H 45 45 45 45 45 + $S(\alpha,\beta,\gamma) = \sum_{l=1}^{N} \sum_{l=1}^{N} \sum_{l=1}^{N} R(l,m,n) L(l+\alpha,m+\beta,n+\gamma)$ 計算速度はZDOCKの 8.8 倍に向上 $S(\alpha, \beta, \gamma) = \text{IFFT} [\text{FFT}[R(l, m, n)]^* \text{FFT} [L(l, m, n)]]$ (single node, single thread Ohue M, et al., Lecture Note in Bioinformatics 7632, 178-187, Springer, 2012. (Katchalski-Katzir model) 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 「京」でのスケーラビリティ測定 Master Node スレッド並列 F-measure : 0.42 $asure = \frac{2 \cdot TP}{(TP + FP) + (TP + FN)}$ Trypsinogen protein d-actin and profile and trypsin inhibitor (PDB id : 1CGI) ¥ Predicted docking p Crystal structure 1974年 44x44=1936 通りのドッキングによる相互作用ペア予測結果 (対角線上が結晶構造の複合体) #Threads ノード間並列 B) 細菌走化性パスウェイへの応用 F-measure: 0.44 小規模な生物系での検証においても、既存知 識に照合して良い精度で予測を得られた。 さらに、現時点では知られていないがドッキン グスコアの高い相互作用 (Chet-Chet) につい て、その先の三量体の予測を試みた。 Output docking results Matsuzaki Y, et al., J Bioinform Comput Biol, 7:991-1012 (2009) Ohue, et al., submitted. 6144 12288 24576 eraction by using "False-sitive" pair CheY-CheD as 08-934 chausting docking of 220 proteins C) アポトーシスパスウェイへの応用 D) 非小細胞肺がんパスウェイへの応用 肺がん薬関連タンパク質間相互作用の探索 肺がん薬 Gefitinib に関連すると 推定されたタンパク質 (東京大学医科学研究所 宮野研究室提供) F-measure : 0.28 F-measure: 0.36 497×497 通り、約25万通りの構造ペア計算を行った。このような桁違いに大規模な対象においても、 走化性パスウェイ(約1万通り)と同等の精度でタンパク質問相互作用を行うことに成功した。 44 proteins (497 structures) 53 88 747 **新規PPIの発見を目指す** (約15の候補を予測済、ポストドッキンク 1105 False Positiv et al., b. IPSJ SIG, 2012-BIO-32(13),



B-1/B-2: NESTによる局所回路シミュレ

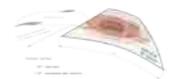




五十嵐 潤¹、Moritz Helias ¹.²、舛本 現³、Tobias C Potjans ¹.²、Markus Diesmann ¹.²、深井朋樹¹

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



- 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%
- 脳機能は多くの領野に分布している
- 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、脳の機能的を再現するには全脳規模のモデルが必要

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
- 共通のブ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

NEST シミュレータ



重要な日標: 体系的に新しいシミュレ ション技術を公開する

GPLの一般なリリースで 作製

- 多数の研究室との共同研究 (2001年~)
- 国際的な高等課程おける教育
- 沖縄計算論的神経科学コース
- nced 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

1. 京へのNESTの移植 108の神経細胞の宝現 3. 全脳規模(109)にむけて

第2世代* 2011年9月 NEST 2.1 第3世代 2012年5月 第4世代 2012年9月 NEST 2.2(次期リリース)

*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:

- 『仁相■する機構: 最大領野(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

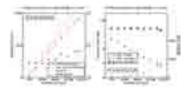
第3世代、第4世代カーネルのメモリー配置

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



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



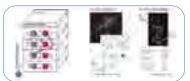
- 神経科学者は利用可能な資源の最大量使用量が知りたい
- ノードあたり利用可能なメモリ量(16GB)を最大限に用いる
- 1コアあたりの神経細胞数
- シミュレーション時間は、通信時間の増加が原因で、増加する。

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



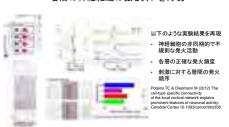
- 波線の右端は、京の可能なマシンサイズを示す
- JUGENEを用いて、MPIジョブの最大使用数、65,536 = 262,144 coresまでを試験 ■ 京の最大構成の74% の MPIジョブに相当
- 京を用いて、MPIジョブの最大使用数、24.576 = 196.608 coresまでを試験を
- 京で第4世代カーネルを用いて、生物学的1秒のシミュレーションに、1520秒=25 分で達成

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



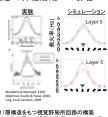
- 1 mm³に相当
- 10万神経細胞、10億シナプス
- 各層は興奮性(80%)と抑制性の神経細胞集団 (20%) からなる
- 興奮と抑制は同一な力学的性質の 神経細胞モデル
- 水平方向に一様な結合
- ■電気生理学と解剖学データに基づき、各層の細胞種ごとに結合確率を 設定

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





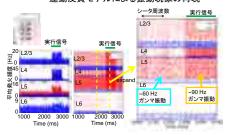
1) 層構造をもつ視覚野局所回路の構築 2) それを用いた視覚的注意の2カラムモデル 3) 空間注意(上)と特徴注意(下)の8カラムモデル

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

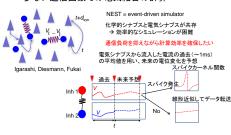
運動の時期によって異なる周波数のγ波発生し、 特定の位相で発火が発生。θ波の特定位相で運 動が発生。 運動実行 遅いガンマ法

表層、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

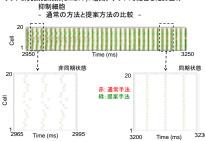
ソフト研究開発成果(CMDN): 運動皮質モデルによる振動現象の再現



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



謝辞

共同研究者

CMDN 様村 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。
- からない。 がまた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。 服券および運用にあたる関係各位に感謝申し上げる。とくに京上の 実装については、ISLM高度化チームの協力無くしては困難であった。感 謝申し上げる。



😕 B -3: VSM

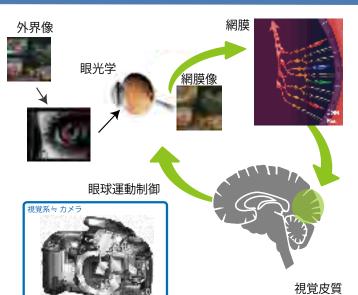


全視覚系モデルによる視覚情報処理の解析 (視覚系シミュレーションのための共有プラットフォーム)

臼井支朗, 稲垣圭一郎, 観音隆幸, 神山斉己(愛知県立大), 石原彰人(中京大学)



被写体にカメラを向ける機能に相当。 大脳皮質で認識された情報に対して、 見たい方向へ視線を移す信号が脳幹部 分で計算されている。実際は体感して いない、手振れに相当するような微小 な動きが重要な働きをする



カメラのフィルムに相当。外界の網膜 像(光情報)が、ここ神経信号(電気 信号)に変換される。網膜で処理され た情報は視神経を介して神経スパイク によって視覚皮質へ伝達される





網膜から伝わる神経スパイクをもとに して物体や景色、またその色を認識し たり、認識した情報を利用して行動を 促す。モノを見る場合は、視線が制御 される

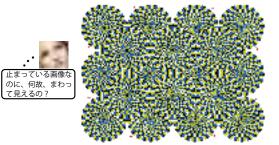
全視覚系モデルの構築

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。

具体的に何をするの?

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)

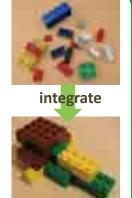
Sub-models (parts)

- 詳細サブモデルの構築と
- 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

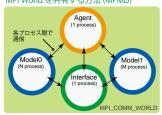
Large-scale model



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)



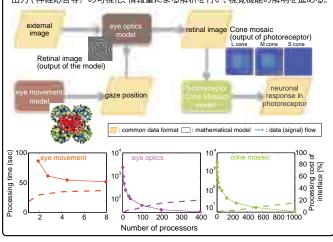
スタンドアローン型のスパコン (京など) におけるモデル統合に有用



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

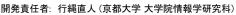
ヒトの視覚は、眼球、眼光学系、網膜、脳皮質により構成されており、外界の情報は、 これらの部位で多階層的に処理されることで、物の認識や視線の移動といった機能が実現されている。シミュレーションでは、入力イメージに対して、それぞれの部位の出力(神経応答等)の可視化、情報量による解析を行い、視覚機能の解明を進める。





为 B-4: NeuroMorphoKit

神経細胞形態シミュレーションキット



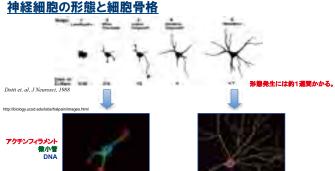
本田直樹(京都大学 大学院情報学研究科)

石井 信(京都大学 大学院情報学研究科、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

シ、シャフトに微小管が局在

- 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

極性形成:

- 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

練成シミュレーション

細胞内シグナル伝達系(反応拡散系)

Nonoka, et. al., Neural Networks, 2011

 $\xi \sim \exp(-x)$

細胞骨格系(アクチンフィラメント)

樹状突起・シナプス形成: 樹状突起に微小管、シナプスにアクチンが局在

細胞膜

アクチンフィラメント

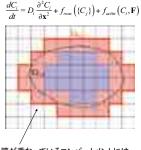
F-actinのキネティクス 分岐計算の効率化 Arp2/3濃度から部位ごとの分岐確率の計算 $\pi_{i} = k_{binding} C_{sbp}(x) \Delta t$ 不均一ポアソン過程からサンプリング

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
- Arn2/3の濃度に比例した確率で分岐

細胞内シグナル伝達

移動境界条件



膜が重なっているコンパートメントには 受容体が存在する

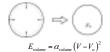
 $\sum_{i=1}^n \pi_i \geq \xi \geq \sum_{i=1}^{n-1} \pi_i$

- 細胞膜境界が移動すると、 コンパートメントを追加・削除する
- ボクセルが削除されると、分子を近接コンパート メントに分配
- 新たできたボクセルにおける分子濃度はゼロ

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

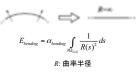
 $\hat{\mathbf{M}}(t) = \arg\min E[\mathbf{M}, \mathbf{F}(t)] \qquad E = E_{volume} + E_{surface} + E_{bending} + E_{F-actin}$

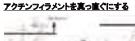


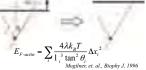
表面積を一定に保つ



細胞膜を滑らかにする







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

 $\mathbf{x}_{i}^{\tau+1} = \mathbf{x}_{i}^{\tau} - \boldsymbol{\eta}^{\tau} \mathbf{grad}_{i} E^{\tau}(\mathbf{X}^{\tau}, \mathbf{F})$

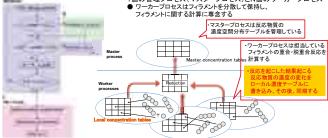
ソフトウェアの実装と京での性能評価

アルゴリズム全体

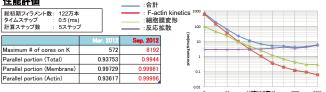
並列化モデル

1個の管理プロセス(マスタープロセス)とN個のワーカープロセス

● ワーカー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



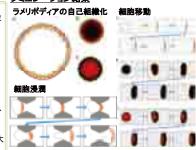
性能評価



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α=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 (東京大学・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
- バ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



モデルの拡張

力学的曲げ、結合タンパク質による結束の導入







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 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 446.5 447.0 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Ca++電流が存在する 触角葉細胞体からの voltage clamp Morph clogy Parameter estimator using NEURON 5領域モデル Physiology 可能な結合を推定 Estimator by Genetic Algorizm using NEURON is implemented. fitting for muliticompartment model is ongoing. 平均応答波形的N ID661 シナプス結合強度を推定 Simulation 領域を介した結合間の すべてのパターン遷移を縮約して パターンの変化と見なす ニューラルネットワークとして表現 のシナプス 領域を同定 領域マッ ピング 標準脳 抽出細胞 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)

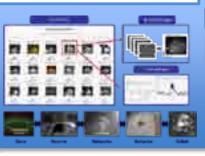


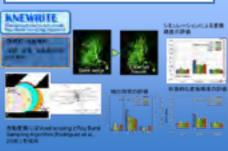


カイコガのニューロンデータベース(BoND) とマルチスケール計測

形態抽出ソフト(KNEWRITE)

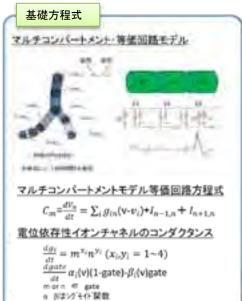
標準脳構築ソフト(BrainComposer)





シミュレーション前のデータの流れ





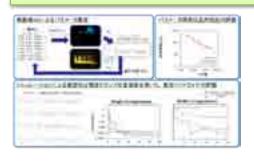




ーラッピング

の推定

シミュレーションを用いた神経パラメータ推定



謝辞

本ソフトウェアの開発に際しては、京速コンピュー 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、特に舛本氏、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 ハウブト・S・周一 加沢 知報 高鳴 職 長 井健志 岩月知香 佐藤陽平 宮本大輔 森友亮 田渕理史 藤原輝史 峰岸諒 佐藤秀仁 **迪<u>野莫利</u>**石井愛弓 新田倫子 西川部子 小林亮太 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.

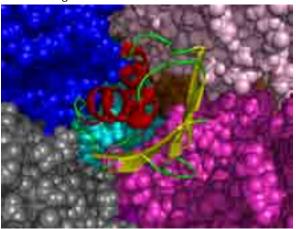


Figure 1 Example of Protein (hide water)

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).

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).

Table 1 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

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

Table 3 Weak Scaling

Tin	Time consumption for 6,542 atom/CPU							
Numb	er of Node	64	512	4,096	32,768	79,872	82,944	
Numb	er of Atom	418,707	3,349,656	26,797,248	214,377,984	522,546,336	542,644,272	
	Total	109.058	110.535	111.186	111.672	112.414	112.085	
(ms/	Force	91.622	91.528	91.601	92.329	91.262	91.641	
step)	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	
Perfo (PFL0	rmance OPS)	0.003	0.025	0.201	1.599	3.871	4.031	
Efficie	ency (%)	39.0	38.5	38.3	38.1	37.9	38.0	

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.



为 H-2: LSV



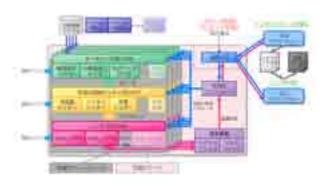
大規模データ可視化ソフトウェアLSVの開発

開発責任者: 小野謙二(理化学研究所 計算科学研究機構)

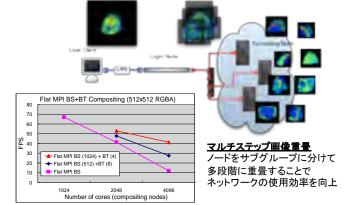
LSV "Large-Scale data Visualization software"

- 大規模データの並列可視化■ 共通UII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

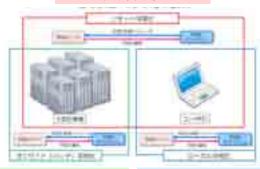


画像重畳



オペレーションモー

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

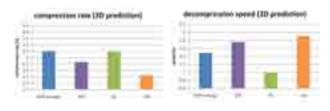
RICCでの並列可視化テスト

- 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

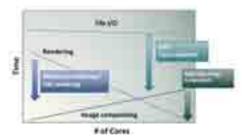
高速データ圧縮

- 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



213



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

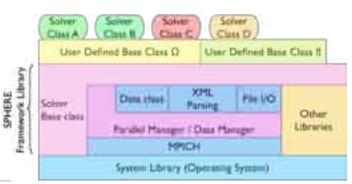
- ・シミュレーションプログラム開発の課題
- ・シミュレーション対象がMulti-physics/Multi-scale化する事により、複雑なコードとなり、開発の難易度が高くなる
- ・HPCアーキテクチャーが複雑化(Multi-core化、ヘテロジーニアス化)し、効率的なプログラム開発が難しい
- ・新しいアルゴリズムを用いてHPCの恩恵を得るには、物理と計算機の両方のスキルが必要
- •解決策
- ・アルゴリズムの研究開発に専念できるシミュレーションプログラム開発環境を提供
- ・多様化するアーキテクチャーのローカライズを吸収することにより、今後のアーキテクチャーにも対応



SPHERE

Skelton 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



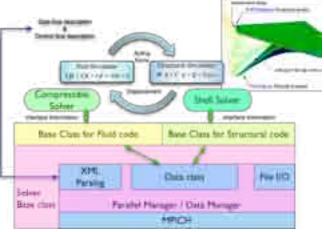


Fig.2 An example of coupled analysis on the middleware

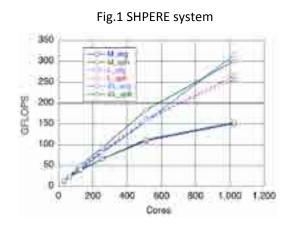


Fig.3 Scalability of SPHERE on RSCC

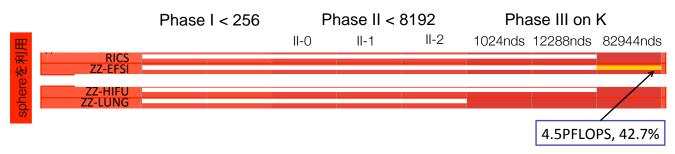


Fig.4 Performance of Software which based on SPHERE



为 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)

背景 1.

- ・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。
- この問題に対する解決策が必要とされている。

月的 2.

- 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。

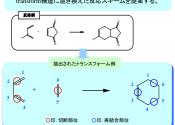
システムの構成

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応できる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。

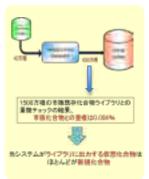
5. transform

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、新だ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。

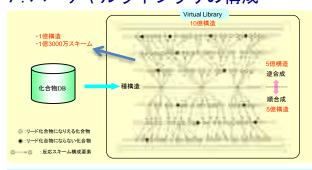
システムに入力された化学構造に、各transformに 記述された部分構造が含まれる場合は、該当部位を transform構造に置き換えた反応スキームを提案する。



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を発生し、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を発生する。 そこで得られた化合物を種として、これらの処理を順次線り返して、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、その両端の入りは、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、合成経路検討の有力な情報となる。

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、薬理学的な性質もある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。



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 MARINE medicana. MENNY L. STATES