

## 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	169
M-2: Platypus-REIN レプリカ交換インターフェースプログラム	170
M-3: MARBLE 全原子分子動力学計算	171
M-4: CafeMol 粗視化モデル計算	172
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	173
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	174
M-7: Platypus-QM 量子化学計算	175
M-8: Platypus-QM/MM 量子化学計算/分子動力学連成計算	176
M-9: Platypus-CGM/MM 粗視化モデル計算/分子動力学連成計算	177
C-1: RICS 細胞シミュレーションプラットフォーム	178
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	179
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定的変形についての検討	180
C-4: RICS-Insulin Granule Dynamics インスリン分泌シミュレーターの研究開発	181
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)	182
O-2: TK-Thrombocyte 細胞シミュレーションプラットフォームを血栓止血の主要細胞である血小板に応用することによる統合的な血栓止血シミュレーターの基盤モデルの開発研究	183
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	184
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	185
O-5: ZZ-LUNG 肺呼吸・肺循環シミュレーション	186
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	187
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	188
D-2: NGS analyzer 次世代シーケンサーの出力データを高速に解析し、個人間の遺伝的差異やがんゲノムの突然変異を高い正確さで同定する	189
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	190
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	191
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	192
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	193
D-7: SBiP データ解析融合プラットフォーム	194
D-8: LiSDAS 生命体データ同化プログラム	195
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	196
B-1: NEST Towards brain simulations	197
B-2: CMDN 大規模並列計算で探る大脳局所神経回路の機能	198
B-3: VMS 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)	199
B-4: NeuroMorphoKit 神経細胞形態シミュレーションキット	200
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	201
B-5: IOSSIM 昆虫嗅覚系全脳シミュレータ	202
H-1: cppmd 大規模並列用 MD コアプログラム	203
H-2: LSV 大規模データ解析ソフトウェア LSV の開発	204
H-3: SPHERE Simulation Framework for Large scale Coupled Analysis	205
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	206



## 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

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

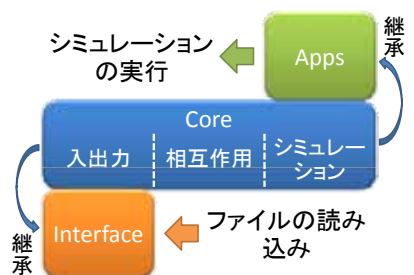
#### 問題点

- 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
  - 生体高分子の内部自由度が大きく、相互作用が複雑なため
- 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

#### 解決法

- マルチコピー・マルチスケール
  - 計算対象のコピーを複数(粗視化レベルが異なるものも含む)用意
  - コピー同士を相互作用させることで、立体構造空間探索効率を高める
- 統計力学
  - コピー間の相互作用によるバイアスを取り除く
  - 自由エネルギー計算

#### 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/MPPIによる並列化

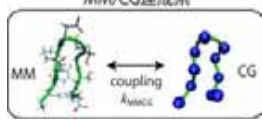
マルチコピー・マルチスケール法の開発を支援するライブラリを新たに開発

#### マルチスケール拡張サンプリング(MSES)法

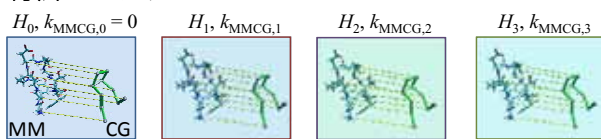
目的: 高精度な全原子モデル(MM)と高速に運動する粗視化モデル(CG)との連成シミュレーションにより、生体高分子の全原子構造サンプリングを効率的に実現

方法1: マルチスケール

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



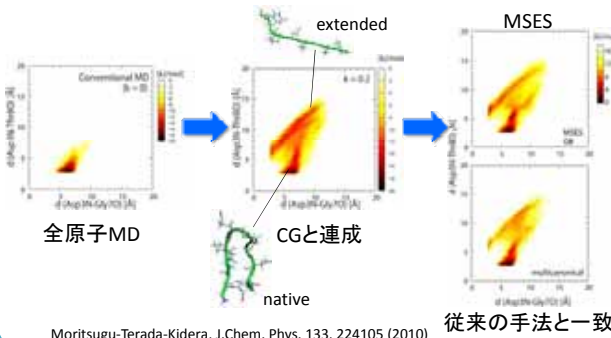
方法2: マルチコピー



Exchange  
ハミルトニアンの交換法  
Fukunishi-Watanabe-Takada, J. Chem. Phys. 116, 9058 (2002)

CGバイアスのないMM構造アンサンブル

応用例: シニョリンフォールディングの構造サンプリング



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

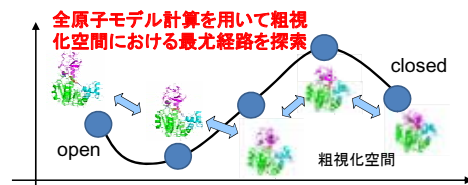
従来の手法と一致

#### 有限温度istring法

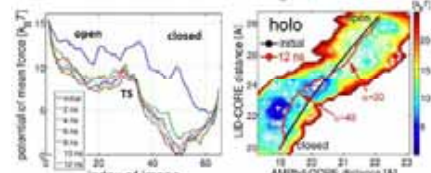
目的: タンパク質が構造変化する際の尤も確率の高い経路を探索する。経路上の遷移状態を特定することにより、構造変化のボトルネックとなる残基を同定する

方法: マルチコピー・マルチスケール

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

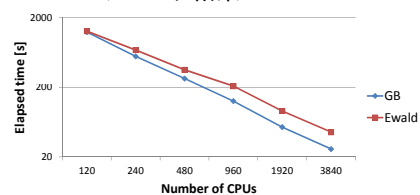


応用例: アデニル酸キナーゼの構造変化



Matsunaga et al., submitted.

性能: 京でのベンチマーク結果



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。性能値については、整備中のシステムによる暫定的な数値です。



## M-2: Platypus-REIN

# レプリカ交換インターフェースプログラム



ISLiM

開発責任者: 杉田有治<sup>1,2,3</sup> (1) 理化学研究所 基幹研 杉田理論生物化学研究室  
 (2) 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)  
 開発者: 宮下尚之<sup>2,3,4</sup> (3) 理化学研究所 計算科学研究機構 粒子系生物物理研究チーム)  
 協力者: 李 秀栄<sup>1</sup> (4) 理化学研究所 次世代計算科学研究開発プログラム 分子スケール研究チーム)

### Introduction

複雑な形を持つタンパク質の構造や物質の状態などの予測は、通常の分子動力学法(MD)やモンテカルロ法(MC)計算では、構造(状態)空間上の局所最小状態に捕まり、多くの場合、正しい構造や状態を見つける事が困難である。

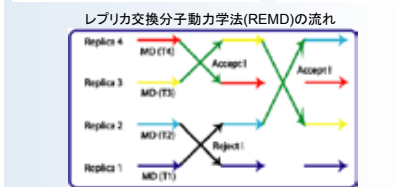
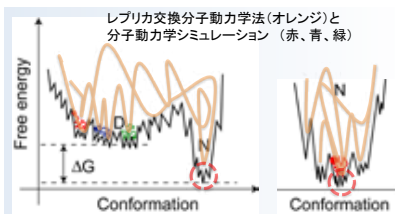
レプリカ交換分子動力学法(REMD)は、1)系の複数のコピー(レプリカ)を用いた異なる温度や距離拘束などのパラメータを持つMDを実行し、2)隣り合うパラメータを持つレプリカをある周期でメトロポリスクリテリアに従って交換させる方法である。温度 $T_m$ と状態 $\lambda_m$ を持つ $i$ 番目のレプリカのハミルトニアンを $H_m(p_i, q_i) = K_m(p_i) + U_{\lambda_m}(q_i)$ とすると、 $i$ 番目と $j$ 番目のレプリカの交換確率は以下の様になる。

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

ただし、 $\beta_m = 1/(k_B T_m)$ ,  $T_m$ : 温度,  $k_B$ : ボルツマン因子である。

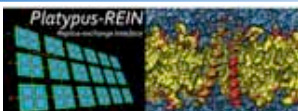
これにより構造は局所最小状態の束縛から解放され、様々な状態(構造)へ遷移が可能となる。統計処理を行う事で最適な構造や状態予測ができる。また高精度の自由エネルギープロファイルが得られるので、タンパク質間やタンパク質-薬(小分子)間などの相互作用の理解にも役立つ。

Platypus-REINは、既存の分子動力学プログラムなどをそのまま利用して、京コンピュータやPCクラスターにて、レプリカ交換シミュレーションを行うソフトウェアである。



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

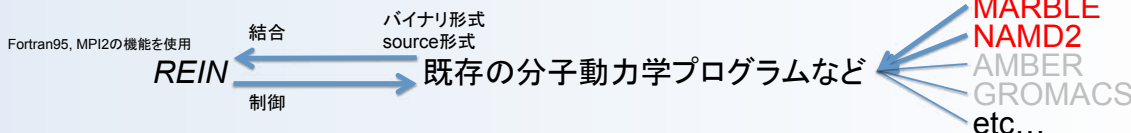
### Replica-exchange interface (REIN)



タンパク質や分子の構造予測・結合予測・構造変化予測が簡便にできる  
 京コンピュータを用いた場合、全原子を用いた大きな系や複雑な環境下にある系の予測シミュレーションが期待できる

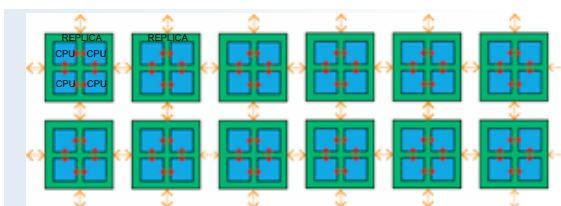
特徴1: 既存の分子動力学(MD)などのプログラムをバイナリ形式(もしくはsource形式)で組込める

→インポート形式のよく知っている既存のMDプログラムを使ったレプリカ交換分子動力学(REMD)シミュレーションができる  
 現状、NAMD2とMARBLEに対応。versionを上げるに従い機能を増やし、拡張性も上げる予定。MARBLEは京コンピュータに対応する為、京コンピュータではREIN/MARBLEを用いる。



特徴2: 多次元レプリカ交換(温度、距離拘束など)に対応

→超並列コンピュータのリソースを有効活用した予測計算ができる  
 右図にて、赤色の矢印は各々のレプリカのMD計算で用いるノード間通信。オレンジ色の矢印はレプリカ交換の際のレプリカ間通信。MD内の通信とレプリカ間の通信を分けている。MD内通信は密であるが、レプリカ間通信は疎である。多次元レプリカ交換(MREM)では非常に多くのレプリカを取り扱うため、京コンピュータなどの超並列コンピュータが必要となる。



特徴3: REMDシミュレーションに必要なsupporting software

→セットアップなどREMDシミュレーションのハードルを下げる  
 REIN\_pack: レプリカ交換に必要なinputを作成。温度の自動setup、初期ファイル設定等。  
 RETUNE (Replica-exchange tunneling time estimator): 解析計算によりシミュレーション時間を予測  
 REMAP (Replica-exchange multi-histogram analysis program): 2次元のWHAM解析プログラム

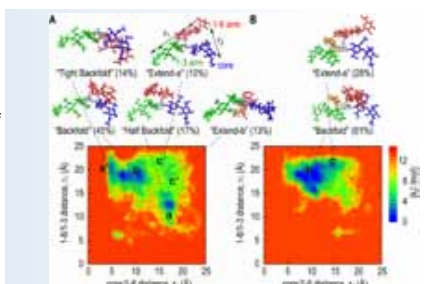
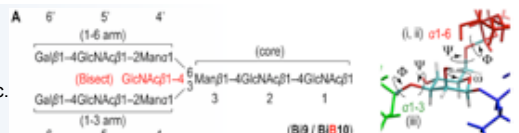


### REINの使用例 水中糖鎖の立体構造予測

N型糖鎖と受容体との結合は、免疫応答など重要な生命現象と深く関わっている。グリコド結合の柔軟性のため、実験で糖鎖の立体構造を捉えるのは極めて困難である。そこで、レプリカ交換分子動力学シミュレーションにより水中N型糖鎖の立体構造予測を行った。その結果、糖鎖-受容体結合の特異性や親和性の違いが、多様な配座異性体の平衡分布とその変化に起因していることが示唆された。

REIN/NAMD2 on RICC  
 260 hours/512 core  
 ~10,000 atoms, 300K-500K  
 64 replica x 52 ns = 3.3 micro sec.

S. Re, N. Miyashita, Y. Yamaguchi, Y. Sugita, Biophys. J. 101, L44-L46, (2011)



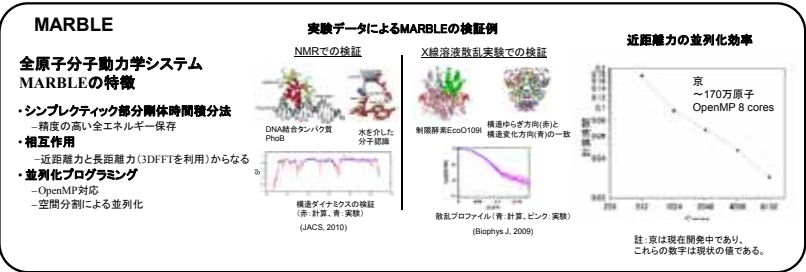
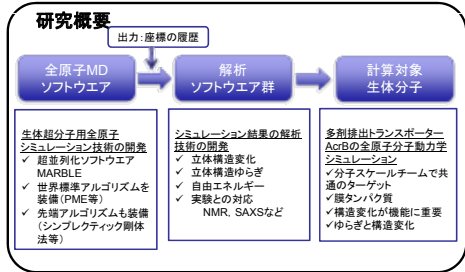




# M-3: MARBLE

## 全原子分子動力学計算

開発責任者: 池口満徳 (横浜市立大学)



## 多剤排出トランスポーターの全原子分子動力学シミュレーション

### 背景

**多剤排出トランスポーター**  
 細胞に有害な薬物の輸送(輸出)を細胞外へ排出する機能を有するタンパク質

水面上に存在する  
 水分子の空気に存在している  
 水一層のトランスポーターで、様々な薬物に対応できる。  
 水層によって有用なものは外に出さない  
 水層内薬物(有害な薬物)の輸送(輸出)を促進する  
 水分子の輸送が促進される

(図1) 大腸菌における多剤排出システム (AcrB/AcrA/TolC)

(図2) 薬剤の経路(青)

(図3) 構造的回転

### AcrB

大腸菌などのグラム陰性細菌に存在するプロトン駆動の多剤排出トランスポーターの一員

- 大腸菌の持つ薬剤耐性の主要因
- TolC, AcrA と呼ばれる蛋白質と複合体を形成して機能しており(図1)、薬剤排出の中心的存在となる。
- 排出するもの: 抗生物質、有機溶剤、染料、色素、色素、非腐食性腐食剤
- 薬剤の構造が異なる。結合型、排出型、取込型と呼ばれるモノマーからなるホモ二量体を形成する(図2)
- プロトン駆動力で異なる3状態(結合型)を構造変化をすることで薬剤を排出する(図3)

AcrBのドメイン構成  
 DN, DO, TolCドメイン (PO1, PO2), ポータードメイン (TM7, TM6, TM5, TM4, TM3, TM2, TM1), 膜貫通ドメイン

プロトンの細胞内への取り込みにかかわるアミノ酸(プロトン転移サイト)  
 結合型: Thr978, Asp407, Lys940  
 排出型: Thr978, Asp407, Asp408, Lys940  
 取込型: Thr978, Asp407, Asp408, Lys940

排出型のLys940のコンフォメーションが結合型・取込型と異なる

### 目的

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

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

### 計算方法

Program: MARBLE (Ikeguchi)  
 Force field: CHARMM 27  
 Electrostatic interaction: Non cut-off (Particle-mesh Ewald)  
 Ensemble: NPT, Temperature: 300K, Time step: 2fs

System  
 Protein: (PDB ID: 2DHH) (Non drug containing structure)  
 Missing part region(499-512) were added from another crystal structure (PDB ID: 2J8S) (47,724 atoms)  
 Lipid bilayer: POPE 809 molecules (101,125 atoms)  
 Water: TIP3P model 106,365 molecules (325,155 atoms)  
 Ion: Na<sup>+</sup> (36 atoms)  
 Total: ~474,000 atoms  
 Box size: 160Åx160Åx180Å

### 本研究で行ったシミュレーション

Asp407	Asp408	Initial state	Final state	Time
407-408-I	407-408-I	407-408-I	407-408-I	500ns
407-408-II	407-408-II	407-408-II	407-408-II	100ns
407-408-III	407-408-III	407-408-III	407-408-III	100ns

抽出型Asp407およびAsp408のプロトン化状態の異なる3種類のシミュレーションを行った。

各シミュレーションとも、結合型と取込型のモノマー(Asp407, Asp408)は脱プロトン化状態で行った。

### 結果

#### プロトン転移サイトのコンフォメーション

結合型・取込型(Asp407, 408を脱プロトン化)

構造変化のチェックに用いた原子間距離  
 (Lys940-Thr978) > (Lys940-Asp407) = (Lys940-Asp408)  
 (Lys940-Thr978) > (Lys940-Asp407) > (Lys940-Asp408)  
 (Lys940-Thr978) > (Lys940-Asp408) > (Lys940-Thr978)

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

抽出型(Asp407, 408を様々なプロトン化状態で計算)

Asp408を脱プロトン化  
 Asp408のみをプロトン化したシミュレーション(407-408<sup>II</sup>)では、抽出型の状態を安定に保持

薬剤の経路(PC2サブドメインのC末領域-TM8)

抽出型では閉じている薬剤の取り込み口が開く様子が観察された

### プロトン転移のプロセス

AcrBで設定される水の透過メカニズム - Alternative Gating mechanism -

水分子の輸送が促進される

結合型・取込型  
 薬剤の取り込み口が開く様子が観察された

抽出型  
 薬剤の取り込み口が開く様子が観察された

水の媒介によりプロトンが移動すると考えられる

### まとめ

AcrB3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、取込型の各状態のプロトマーは、Asp407, 408が脱プロトン化した状態が安定であり、抽出型では、Asp408がプロトン化した状態が安定である。
- ★ 抽出型の薬剤取り込み口が開く様子を観察することができた(407-408-I)
- ★ ここで示された、各状態のプロトン化状態から考察されるプロトン転移のプロセスは、結晶構造に見られる水のパスと矛盾しない。

### 謝辞

本研究の結果の一部は、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。



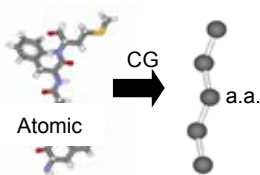


## M-4: CafeMol 粗視化モデル計算



開発責任者: 高田彰二 (京都大学理学研究科)

- 概要
  - 粗視化分子モデル計算により大規模生体分子の長時間シミュレーションを行う。
- 離散化(計算モデル化)の方法
  - 粗視化された分子モデルによる古典分子動力学法



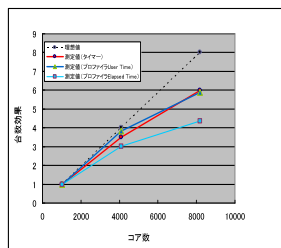
- 計算方法
  - Langevin方程式の時間発展を数値的に積分する。
- 並列化の方法
  - Neighbor list方式、レプリカ交換法
- 開発言語とライブラリ
  - Fortran90, MPI, OpenMP
- コードの公開状況
  - <http://www.cafemol.org> からダウンロード可。ただし生体膜、核酸については要相談。



Version 1.0 (only for proteins), available at [www.cafemol.org](http://www.cafemol.org)

Kenzaki et al  
J. Chem Theo Comp  
2011

- 現状での計算規模
  - 1万残基タンパク質のミリ秒相当のシミュレーションが可能
  - PCクラスタで8192コア並列
  - メモリ容量 2 GB/コア、ディスク容量 1 TB



- 次世代機「京」での計算規模
  - 10万粒子(100万原子)系の秒相当のシミュレーション
  - 反復回数 $10^{10}$ 回以上
  - メモリ容量 2 GB/コア、ディスク容量 1 PB

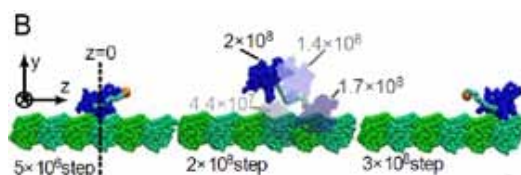
- どんなことができるか
  - X線回折やNMRによる構造情報が存在するタンパク質、核酸などの動態の分子動力学シミュレーション。とくに構造変化を伴うタンパク質ドッキング、分子モーターやトランスポーターの大規模構造変化などをシミュレーションできる。
  - 例として、キネシンの動態、多剤排出トランスポーターの動態、DNAヒストン複合系(ヌクレオソーム)など。

### キネシンの動態

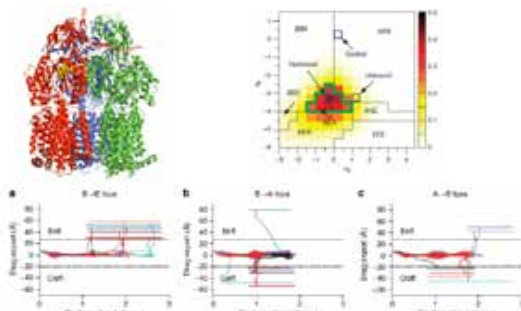


双頭キネシン:  
高精度モデリング

単頭キネシン:  
運動シミュレーション

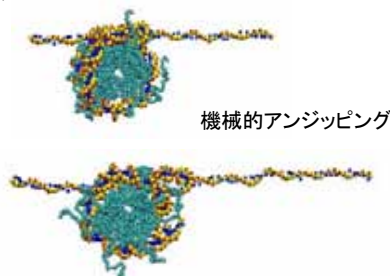


### 多剤排出トランスポーターAcrBの動態



Yao et al Nature Comm 2010

### ヌクレオソーム



機械的アンジッピング



## 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

開発責任者: 佐藤文俊(東京大学生産技術研究所)

開発者: 平野敏行、上村典子、恒川直樹、松田潤一(東京大学生産技術研究所)



ProteinDFとは:

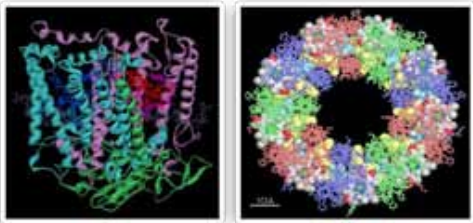
- ・タンパク質が持つ全ての電子の分子軌道が計算できる量子化学計算ソフトウェア

本プロジェクトは:

- ・「京」への実装がミッション\*
- ・光合成反応系がターゲット\*\*
- ・ストレスの少ない基礎・応用研究展開に貢献

\* 京速コンピュータ京の試験利用を利用いたしました。

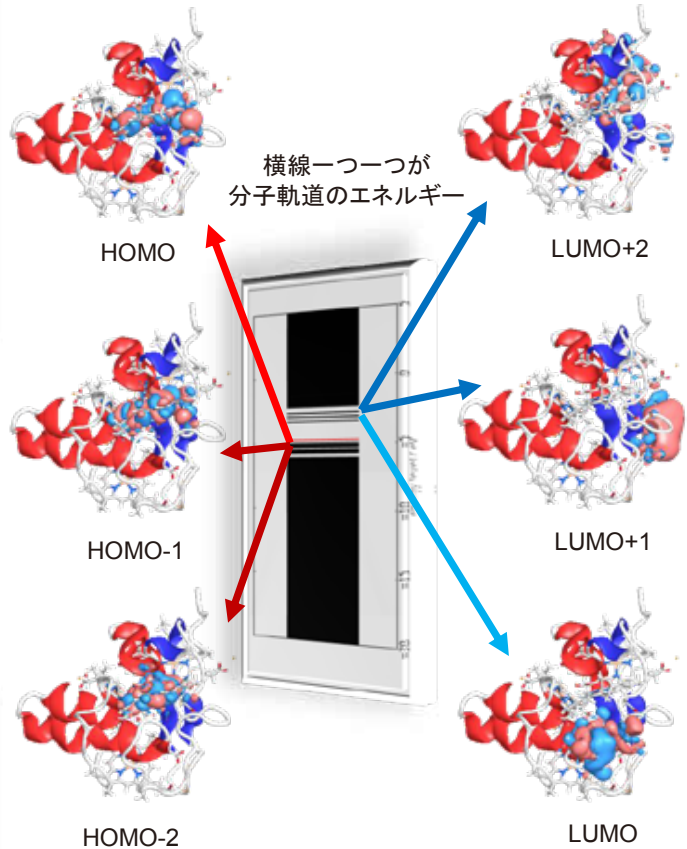
\*\* 京速コンピュータ京を利用する予定です。



光合成反応中心・アンテナタンパク質

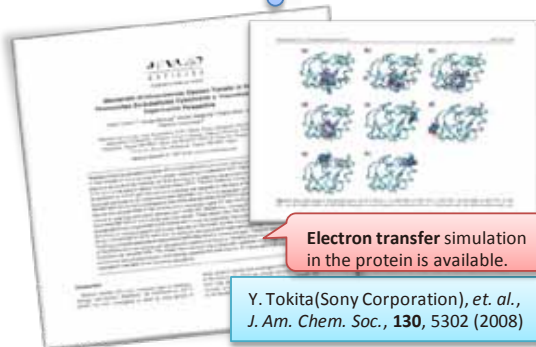
本プロジェクトでは、プログラムを「京」に対応させるだけでなく、正しい大規模DFT計算を実行するための研究も同時に行っています。

- ① 難しいSCF計算を収束に導く技術
- ② 超大規模DFT計算でも信頼できる交換相関項計算法の開発
- ③ 量子化学計算結果を分子力学計算へ還元する方法の研究



計算事例  
~複雑タンパク質全電子計算達成のために~  
ヘム(鉄-ポルフィリン)を  
4つ持つcyt.  $c_3$ のB3LYP分子軌道計算

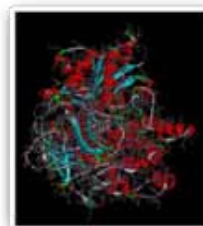
タンパク質量子化学計算による応用研究も進んでいます



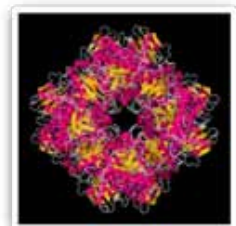
Electron transfer simulation in the protein is available.

Y. Tokita(Sony Corporation), et. al.,  
*J. Am. Chem. Soc.*, **130**, 5302 (2008)

産業界での使用例



ヒドロゲナーゼによるH<sub>2</sub>の生産  
田村隆先生(岡山大学)との共同研究



RuBisCOによる炭素固定

研究室独自研究例



## 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。

### 理論

$$\text{QM/MM ハミルトニアン} \quad \hat{H}(\mathbf{r}, \mathbf{R}, \mathbf{X}) = \hat{H}^0(\mathbf{r}, \mathbf{R}) + \hat{H}^{\text{QM-MM}}(\mathbf{r}, \mathbf{R}, \mathbf{X}) + E^{\text{MM}}(\mathbf{X})$$

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

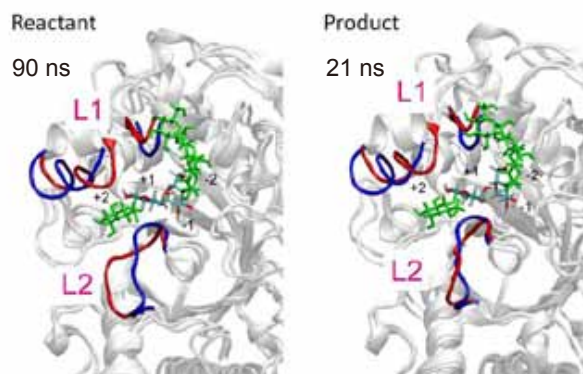
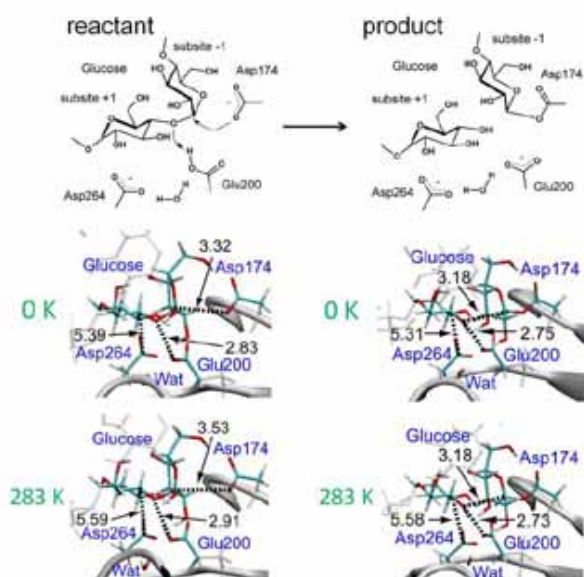
$$\text{Fock or Kohn-Sham 方程式} \quad \hat{f}^{\text{QM/MM}}(\mathbf{r}; \mathbf{R}) = \hat{f}^0(\mathbf{r}; \mathbf{R}) + \sum_A^{N_{\text{QM}}} \hat{q}_A(\mathbf{r}; \mathbf{R}) \langle V_A(\mathbf{R}, \mathbf{X}) \rangle_{\mathbf{X}, E[\mathbf{q}(\mathbf{d}, \mathbf{R}); \mathbf{R}, \mathbf{X}]}$$

$$\text{静電平均場} \quad \langle V_A(\mathbf{R}, \mathbf{X}) \rangle_{\mathbf{X}, E[\mathbf{q}(\mathbf{d}, \mathbf{R}); \mathbf{R}, \mathbf{X}]} = \langle V_A(\mathbf{R}, \mathbf{X}) \omega(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X}) \rangle_{\mathbf{X}, E[\mathbf{q}(\mathbf{d}_{\text{ref}}, \mathbf{R}_{\text{ref}}); \mathbf{R}_{\text{ref}}, \mathbf{X}]}$$

$$\text{Reweighting factor} \quad \omega(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X}) = \frac{\exp(-\beta \Delta E^{\text{QM-MM}}(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X}))}{\langle \exp(-\beta \Delta E^{\text{QM-MM}}(\mathbf{d}, \mathbf{d}_{\text{ref}}; \mathbf{R}, \mathbf{R}_{\text{ref}}; \mathbf{X})) \rangle_{\mathbf{X}, E[\mathbf{q}(\mathbf{d}_{\text{ref}}, \mathbf{R}_{\text{ref}}); \mathbf{R}_{\text{ref}}, \mathbf{X}]}}$$

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

### 計算結果



左図: QM 領域の最適化構造。熱揺らぎの効果により化学的状態に応じた顕著な構造変化が見られる。

上図: 活性部位付近のタンパク質構造変化。初期構造 (青) から大きな変化が見られる。また、化学反応に関連する L2 ループの大きな構造変化が得られた。

### プログラム開発

- 平均場 QM-MM 相互作用の並列化はサンプル分割。
- QM 部分は検討中。

### 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

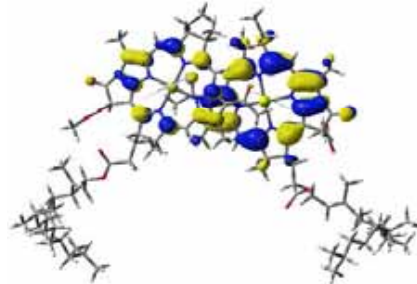
開発責任者: 中村春木 (大阪大学蛋白質研究所)



## 1. 概要

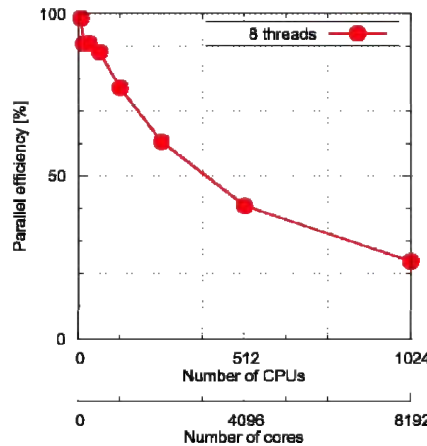
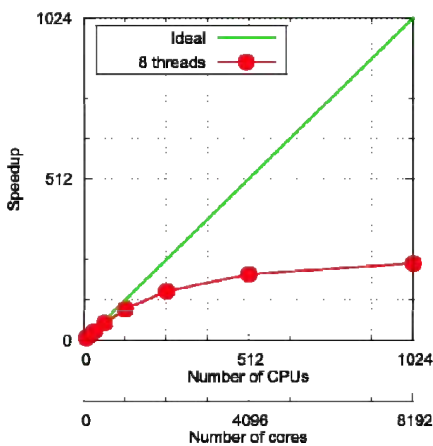
分子軌道法および密度汎関数理論(DFT)を用いた蛋白質など生体分子系の全電子第一原理計算を行う。本研究では生体中での励起状態等、化学結合の結合解離、分子間力の記述を行うため、Quantum Mechanics(QM)計算の高度化を行ってきた。

具体的には、Platypus-QMへのCIS、CIS-DFT、CIS(D)、CASSCF、MP2の実装を行った。CISは生体色素などの励起状態を記述する為の基礎理論であり、CIS-DFT、CIS(D)はその高精度理論に該当する。これらの実装により生体色素による光吸収(光合成系)などの高精度計算が可能になった。またCASSCFは化学結合の結合解離過程を記述する為の理論であり、今回の実装、高効率化により、化学反応過程が記述可能となった。又、MP2の実装により、分子内水素結合等の分子間力に起因する生化学過程も計算可能となった。



Phodobacter sphaeroides由来の光合成活性中心のバクテリオクロロフィルa二量体(スペシャルペア)のHOMO

## 2. 京での性能評価



京コンピュータ上においてハイブリッド並列とSIMD化を進めており、エネルギー計算における積分計算およびFock行列の2電子項計算のカーネルコードを作成した。AO数2728の系において、並列化率99.96% (8CPUx1スレッドと1024CPUx8スレッドから算出)が得られた。512CPUx8スレッドでは40%を超える並列化効率を観測できた。

系: スペシャルペア(原子数280)のcc-pVDZ基底関数系(AO数2728)

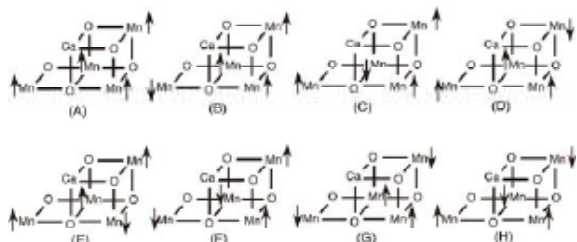
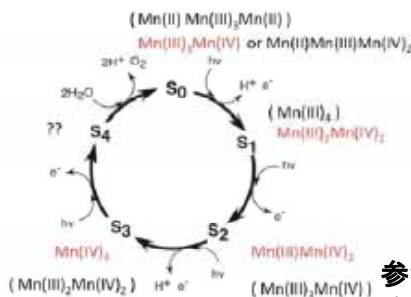
## 3. Platypus-DFTによる応用計算

### 光合成の水分解反応中心の計算

可視光を用いて水を分解し酸素と水素(プロトン+電子)を発生させるこの反応は現在のエネルギー問題・地球温暖化問題を解決する鍵となる反応である。

その触媒中心は酸素架橋マンガン四核よりなるクラスターである。重原子の位置を固定し、架橋アミノ酸の側鎖中のアセテートを酢酸にモデル化して、Kokサイクルを想定して酸化状態を変えながら、コア部分の最安定電子状態を調べた。

Kokサイクルとマンガンクラスターの酸化数(赤が想定した状態)



PS IIのマンガンクラスターの可能な全てのスピン構造  
マンガンイオンがスピン分極率から見れば完全な磁性状態にあるのにも関わらず、スピンや電荷が各サイト間を容易に転移するLabileな電子構造を持つ事、またサイト間の磁氣的相互作用解析の結果からは、クラスター全体の磁氣的なカップリングの結果全系のスピン状態が決定している事が明らかになった[1-2]。

### 参考文献

1. Kanda et al., *Chem. Phys. Lett.* 506, 98-103 (2011)
2. Yamanaka et al., *Chem. Phys. Lett.* 511, 138-145 (2011)



M-8: Platypus-QM/MM

# 量子化学計算／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. 概要

Platypus-QM/MM(Quantum mechanics/ Molecular mechanics)では、生体高分子を経験的分子力場エネルギー関数と非経験的分子軌道法で記述して分子動力学シミュレーションを遂行できる。経験的分子力場エネルギー関数では各原子は質量と電荷を持ち周辺原子との結合関係を経験的に予め定められたパラメーターによって計算する。一方、非経験的な分子軌道法では、各原子は原子核の電荷と電子波動関数によってのみ記述される。我々はボルン-オッペンハイマー近似の範囲内で、原子核が電子構造とは断熱的に分離して取り扱えると仮定したうえで、ハイブリッドエネルギー関数で記述された系のハミルトニアンを古典的に積分して系を時間発展させている。Platypus-QM/MMはこの経験的分子力場ユニットと非経験的ユニットを連成させてQM/MM分子動力学シミュレーションを実施する。この際、MM空間からの静電的寄与を従来法に比べてはるかに正確に取り込む方法として、新たに開発したZero-Dipole Summationのアルゴリズムを取り入れた。さらに、Chain-of-State法に基づく疎結合超並列計算によって、最小自由エネルギー経路決定するスキームを導入した。

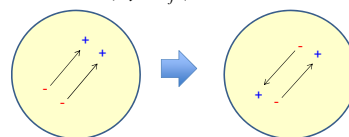
## 2. Zero-dipole summation アルゴリズムのMD計算への適用

Wolfらによって提案された周期系において人工的な中性系を作ることによって静電相互作用エネルギーの精度を保つ工夫[1]は、福田らによって双極子モーメントもゼロとする系とするアルゴリズムに発展された[2]。このアルゴリズムでは、周期系に必須の補正関数におけるダンピングパラメータを限りなく小さくすることが可能であり、非周期的な系への拡張も可能である。

このアルゴリズムを水分子の系および膜蛋白質の系に応用し、Ewald法によるエネルギーおよび力の値と比較したところ、極めて良い精度で一致することが確認された。さらに、このアルゴリズムによるMD計算の結果、それらの系のダイナミックな性質も保たれていることが確認された。

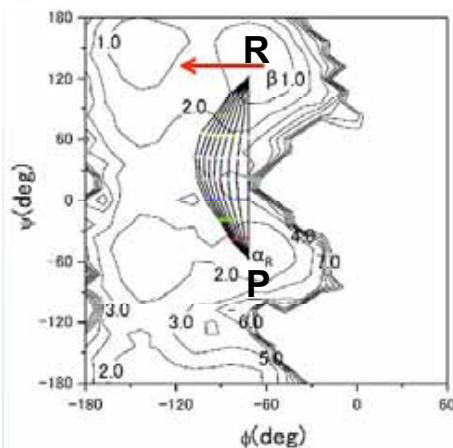
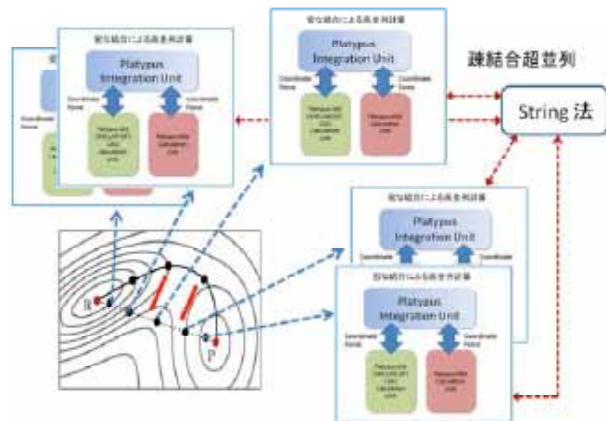
$$U \approx \frac{1}{2} \sum_{i \in N} \sum_{\substack{j \in N \\ (r_{ij} < r_c, \\ j \neq i)}} q_i q_j \left\{ V(r_{ij}) - V(r_c) + \frac{F(r_c)}{2r_c} (r_{ij}^2 - r_c^2) \right\} - \left\{ \frac{V(r_c)}{2} + \frac{r_c}{4} F(r_c) + \frac{\alpha}{\sqrt{\pi}} \right\} \sum_{i \in N} q_i^2$$

$$V(r_{ij}) = \frac{\text{erfc}(\alpha |\vec{r}_i - \vec{r}_j|)}{|\vec{r}_i - \vec{r}_j|} \quad F(r) = -\frac{\partial V}{\partial r}$$



## 3. 最適自由エネルギー経路探索法: Chain-of-State + hybrid-QM/MDシミュレーション

酵素などの複雑な分子過程の自由エネルギー計算に対する高効率の計算のため、Chain-of-State法による疎結合並列計算を用いた、複数のパラメータ値が同期して類似の計算を行う二重化した並列化スキームを導入した。Chain-of-State法の初期配置として、ある平衡状態RとP間を結ぶ試行経路(下記の例では直線)上に等間隔に計算シミュレーション点を配置する。各計算シミュレーション点ではお互い他とは異なる位相空間上のパラメーター(反応座標)を持ち定められた時間間隔中(フェーズ)に各々位相空間点近傍の自由エネルギー面に関する計算を実行する。各々の計算が収束した後に、ホストにこれらの計算情報を集約して評価することで次の時間間隔に各シミュレーションが使用するパラメーターを決定してブロードキャストし次のフェーズの計算を続行する。パラメーターが収束するまでこの計算とパラメーター集約の過程を繰り返すことで最適なパラメーターを決定し、RとPを結ぶ最適経路が決定されることになる。従来のChain-of-State法は、ほとんどが古典力学の範疇でなされてきたが、我々は、Platypus-QMで開発した電子状態計算を導入することによって、電子構造が変化する化学反応の自由エネルギー経路の計算を可能とした。



水溶液中でのアラニン・ジペプチドの構造変化について、R(β構造)からP(α-helix)への変化の最適自由エネルギー経路の探索例。水はMM、ペプチドはQMとして計算。

[1] Wolf et al. *J. Chem. Phys.*, **110**, 8254-8282 (1999), [2] Fukuda et al. *J. Chem. Phys.* **134**, 164107 (2011)





## M-9: Platypus-CGM/MM

# 粗視化モデル計算 / 分子動力学連成計算

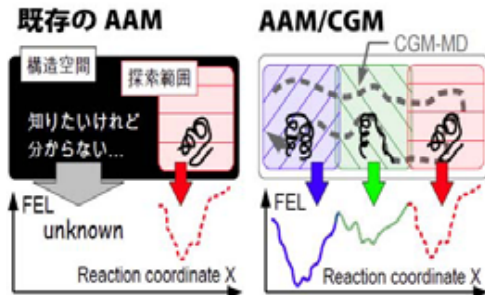


開発責任者: 中村春木 (大阪大学蛋白質研究所)

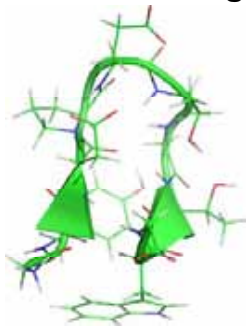
### 1. 概要

蛋白質の分子動力学に置いてもっとも一般的に使用されている全原子モデル(AAM)は、構造情報が詳細であるがゆえに構造空間の探索に時間がかかる。特にドメイン運動のような現象を扱うには不向きである。

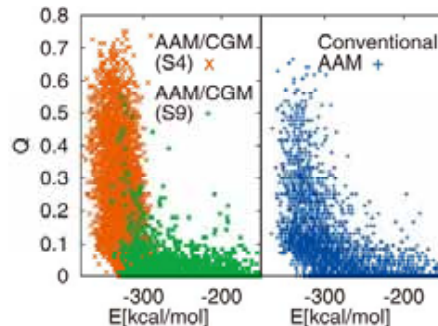
一方で粗視化モデル(CGM)は長時間のシミュレーションに向いてはいるが、原子レベルの相互作用を記述することはできない。このため蛋白質の構造予測や低分子化合物との結合を取り扱うことはできない。そこでAAMの信頼性とCGMの効率性の双方を活用するマルチスケールシミュレーション手法を提案し、その手法で用いる為のCGMシミュレーター(platypus-CGM)を開発してきた。



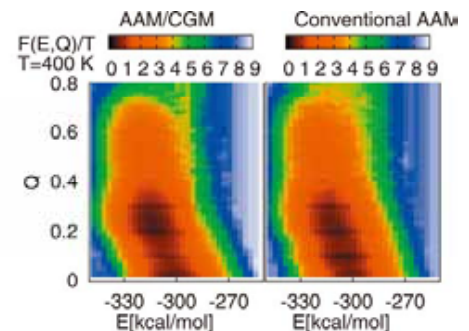
### 2. 手法の検証 (Chignolin)\*1



Chignolin分子(PDBID:1uao)  
配列: GYDPETGTWG



エネルギーとQ値の散布図  
(左) AAM/CGM、  
(右) 従来のall-atom multicanonical



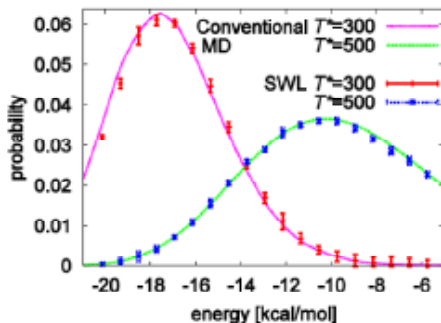
エネルギーとQ値のFEL  
(左) AAM/CGM、  
(右) 従来のall-atom multicanonical

**AAM/CGMの結果とAAMシミュレーションの結果がよく一致**

### 3. 構造サンプリングの効率化\*2

Wang-Landau法ではエントロピー $S$ をエネルギー $E$ と時間 $t$ の関数で表し、 $S=S+f$  ( $f$ は分布関数 $P(E)$ が平らになるたびに減少するパラメータ)と変更する。サンプリング初期の段階の $S$ の凸凹をのぞくため、 $S$ の曲率を最小化する変分原理を導入し、自動的に精度よく構造をサンプリングする方法を確立した(Smoothing Wang-Landau法:SWL法)。

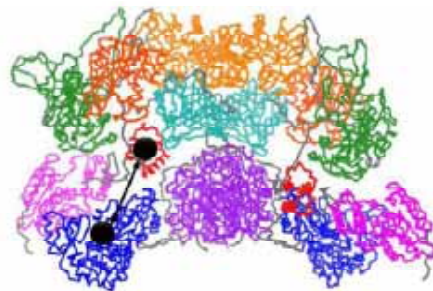
$$f = \omega \int_{E_{min}}^{E_{max}} dE \left( \frac{\partial^2 S}{\partial E^2} \right)^2 + (1 - \omega) \int_{E_{min}}^{E_{max}} dE (S - S^*)^2$$



真空中のAlanine-dipeptideをSWL法を用いて計算したカノニカル分布(赤・青)とカノニカルシミュレーションの結果(ピンク・緑) => 一致

### 4. 今後の計画

- SWL法の脂肪酸合成酵素シミュレーションへの応用



黒丸はACPとMATドメインで、矢印の長さが $R$ 、 $RI$ に対してSWL法を応用することで、拡散エントロピーを効率的に求めるのが狙い。

- プログラムのさらなる高速化: 相互作用テーブルの通信の効率化

### 参考文献

1. Shimoyama et al., *J.Chem.Phys.* 133, 135101 (2010)
2. Shimoyama et al., *J.Chem.Phys.* 134, 024109 (2011)





# C-1: RICS 細胞シミュレーションプラットフォーム (RICS:RIKEN Integrated Cell Simulator)



開発責任者: 横田秀夫 須永泰弘 (理化学研究所細胞スケール研究開発チーム)

**Motivation:** 細胞は複雑な外形形状や内部構造を持ち場所によって代謝反応が異なる。形状を考慮して細胞内反応を計算したい

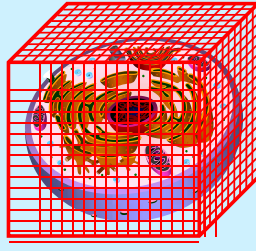
**Our Approach:** 細胞シミュレーション統合シミュレーションプラットフォームの開発

## 細胞シミュレーションプラットフォーム RIKEN Cell Simulator (RICS)

～細胞内の場を考慮した生化学反応シミュレーター～

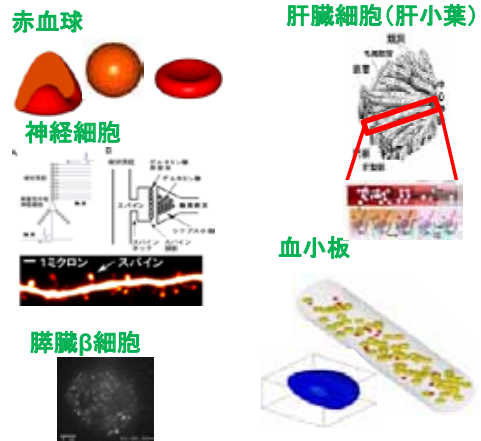
膜透過 + 拡散 + 生化学反応(代謝)

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

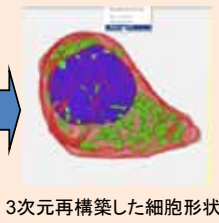


臓器と細胞をつなぐシミュレーションを目指しています。

## RICSのターゲット



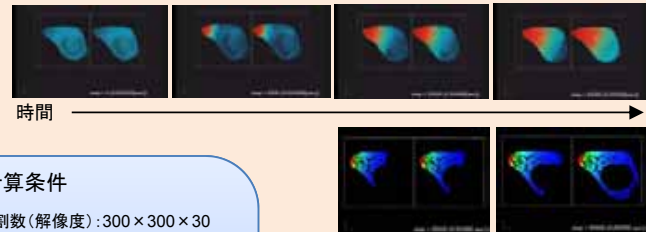
## 顕微鏡画像から作成した3次元形状内での計算



RICS-Solver  
で計算

反応: 24反応  
物質数: 10個

計算結果 (左: 反応あり、右: 反応なし)



### 計算条件

分割数(解像度): 300 × 300 × 30  
並列方法: 領域分割  
並列数: 256並列  
 $\Delta t = 0.0005$  (秒)  
計算Step数: 20000Step  
使用した計算機:  
RICC超多目的クラスター  
計算実時間:  
反応なし: 34分  
反応あり: 24時間  
 $Ca^{2+}$ の拡散係数:  $1 \times 10^{-9} (m^2/s)$

反応があると見かけの拡散速度が遅くなる

細胞内の局所的な反応が細胞内の物質の濃度変化に影響を与えていることを示唆した。

## RICCの構成方程式

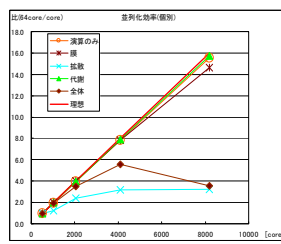
構成方程式:  $\frac{\partial \phi_{i,m}}{\partial t} = \nabla_C \nabla \phi_{i,m} + E_{i,m} + Q_{i,m}$

拡散 反応 膜透過

物質生成項:  $E_{i,m} = \lim_{\Delta t \rightarrow 0} \frac{EC_{i,m}(\phi_{i,m}, \Delta t) - \phi_{i,m}}{\Delta t}$

膜透過項:  $Q_{i,m} = \sum_n a_{m,n} (-v_{m,n} (\phi_{i,m} - \phi_{n,m}))$

$\phi_{i,m}$ : 媒質 i 内の物質 m の濃度  
 $EC_{i,m}$ : 媒質 i 内の物質 m の反応後の濃度  
 $a_{m,n}$ : 媒質 m と n の膜面積  
 $C_i$ : 物質 i の拡散係数  
 $V_{i,m}$ : 媒質 i のボクセル内体積率  
 $v_{m,n}$ : 物質 m の媒質 n の透過率



RICCを用いた大規模並列

## 謝辞

本発表の結果の一部は、理化学研究所が実施している京速コンピュータ“京”の試験利用によるものです。



## C-2: RICS-Pre

### 細胞シミュレーションプラットフォーム(RICS)プリシステム

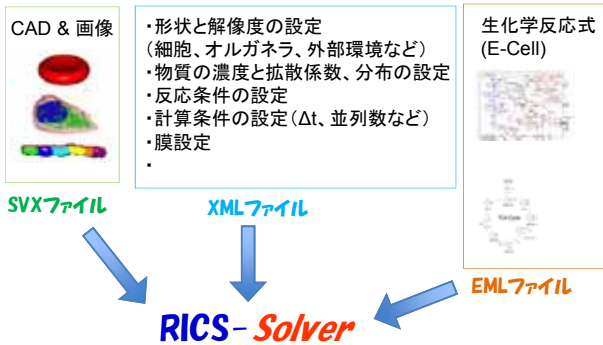
開発責任者: 横田秀夫 須永泰弘 (理化学研究所細胞スケール研究開発チーム)



**Motivation:** 実際の細胞は複雑な内部構造を持ち場所によって代謝反応が異なるため複雑な形状や初期条件を効率よくRICSに設定したい。

**Our Approach:** 細胞シミュレーション統合プラットフォームプリシステム (RICS-Pre) の開発

#### RICS Solver で必要な設定ファイル



#### RICS-Preがあると

スペシャリストではなくても、細胞シミュレーションのモデル作成が可能になる。

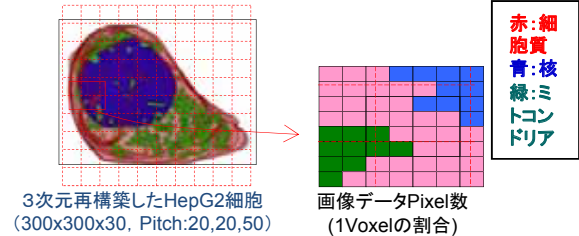
- ✓ 基本形状の作成が容易  
単純な基本形状はCADを用いずRICS-Preで作成
- ✓ 解析条件の設定が容易  
入力フォームに従って必要な情報を設定

#### RICS-Preの機能

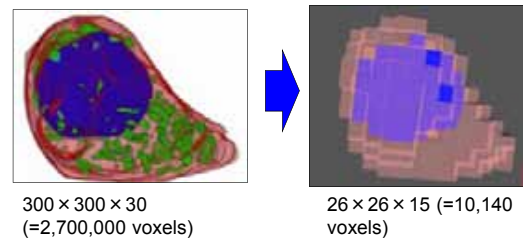
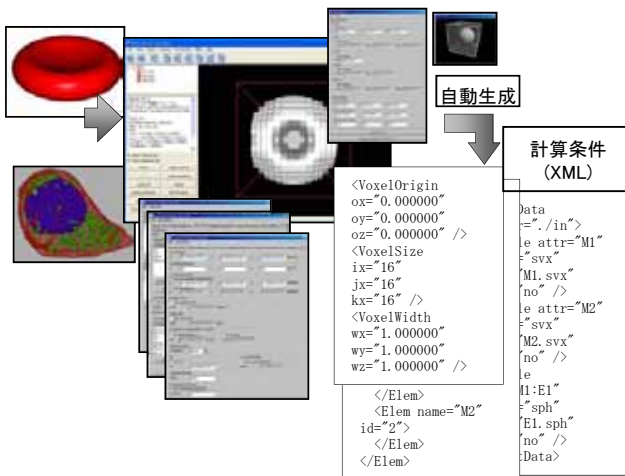
- ✓ 実細胞形状及びCAD形状の取り込み
- ✓ 基本形状の生成
- ✓ 形状の修正
- ✓ 細胞内の媒質, 物質パラメータ等の設定
- ✓ 反応経路及び膜機能を三次元データへ設定
- ✓ 計算条件の設定

#### Voxelデータの解像度変更

細かい顕微鏡データから精度よく粗い解像度の形状モデルを作成



#### RICS-Preのフロー



#### 謝辞

本発表の結果の一部は、理化学研究所が実施している京速コンピュータ“京”の試験利用によるものです。



### C-3: RICS-platelets

## RICSによる血小板細胞シミュレーションと力学的作用による偽足的変形についての検討

七澤 洋平、田村 典子、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)



本研究では、血小板細胞について ①細胞の生化学的反応 ②力学的作用による部分的な変形の2点に着目し、研究を進めている。

### RICSを用いた血小板細胞による複数細胞によるシミュレーションと内部反応代謝マップの精緻化

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

前年度までに、RICSを用いた血小板細胞モデルを構築し(図1)、細胞外からの細胞膜への刺激に応じた、血小板細胞内オルガネラである濃染顆粒からのADP、ATP、セロトニンなどの内包物質放出を各種生化学反応の積み重ねにより表現を可能とした。

本年度の成果として、複数の血小板細胞を計算機中に設定し、ある細胞にはGP1baへと局所的にvWFによる刺激を与え、その箇所から始まる活性化プロセスを経て濃染顆粒からの内包物質放出、更に別細胞へと放出物質による活性化の誘起を確認した(図2)。

細胞内代謝反応マップの精緻化を進め、実験により測定可能な物質を組み込み、計算結果と実験結果の比較を可能なものとした(図3)。

### 血小板膜上蛋白質GPIbaとvWFとの接着時に生じる偽足を考慮した血小板運動モデル

PGE1によりGPIIb/IIIaの活性化を阻害した血小板細胞が、フローチャンバーのvWF付着面へとGPIba-vWF結合による接着時に生じる糸状の存在(偽足)の性質検証を、血小板細胞を楕円体形状の剛体と仮定し、運動モデルを用いて計算機シミュレーションにて行った。

本年度は、血流から生じる流体力の計算を双平面とせん断速度の影響を考慮し、運動モデルの精緻化を行った。1つの血小板細胞について種々計算条件を変えることで、よりセンシティブに偽足の長さ分布に影響を与える要因の検証を行った。

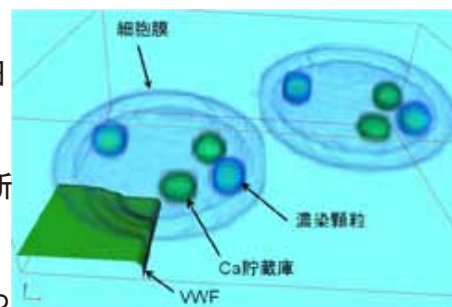


Fig. 1. RICSにて、複数細胞を考慮しての血小板細胞モデルの構築

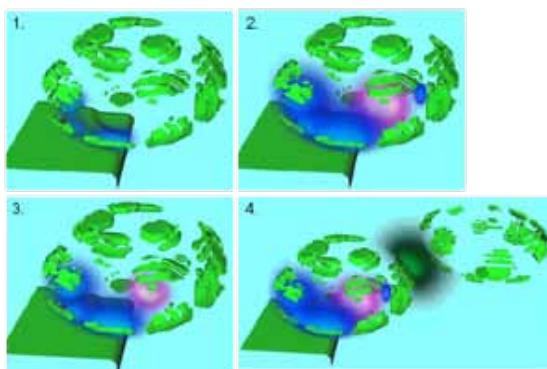


Fig. 2.

1. 局在するGPIIb/IIIaがvWFにより刺激される
2. GPIIb/IIIa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。

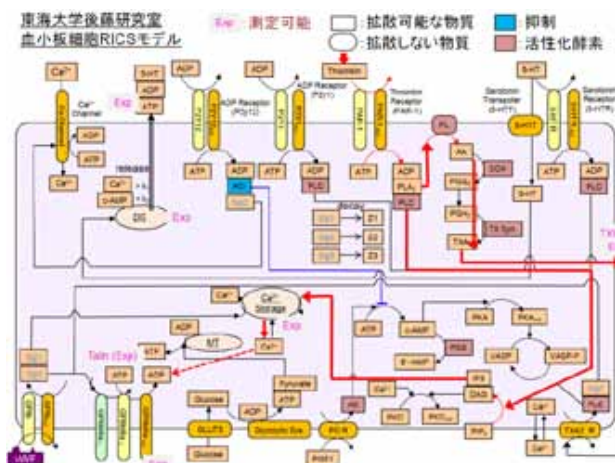


Fig. 3. 血小板細胞内代謝反応マップの拡張、及び、測定可能物質との比較





### C-4 RICS-Insulin Granule Dynamics

## インスリン分泌シミュレーターの研究開発

開発責任者: 清野 進 (神戸大学大学院医学研究科細胞分子医学、細胞スケール研究開発チーム)



### 背景と目的

日本では成人の約6人に1人が糖尿病もしくは糖尿病予備軍と言われている(厚生労働省平成19年国民健康・栄養調査)。日本人の糖尿病の主な原因の1つは、膵臓β細胞からのインスリン分泌能の低下である。しかし、膵臓β細胞内でのインスリン顆粒の動態に関しては不明な点が多い。

本研究では膵臓β細胞内でのインスリン顆粒動態のメカニズムを解明するために、全反射型蛍光顕微鏡(TIRFM)で取得した動画よりインスリン顆粒の動態モデルを、またβ細胞の代謝動態モデルをRICSを用いて作成し、インスリン分泌シミュレーターを開発することを目的とする。



図1. 本研究の概略

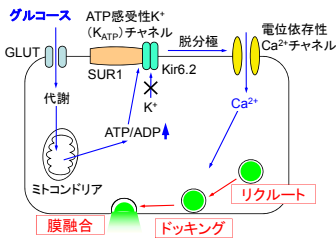


図2. 膵臓β細胞におけるグルコース応答性インスリン分泌

青線: グルコース代謝  
赤線: インスリン顆粒の小胞輸送

### 方法

1. TIRFMを用いて、膵β細胞のグルコース応答性インスリン顆粒動態を解析する。
2. インスリン顆粒動態をモデル化する。
3. 膵β細胞のグルコース代謝を細胞シミュレーション統合プラットフォーム(RICS)でモデル化する。

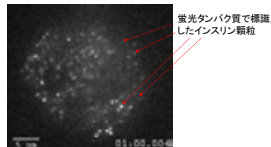


図3. マウス膵β細胞のTIRFM画像

### 結果とまとめ

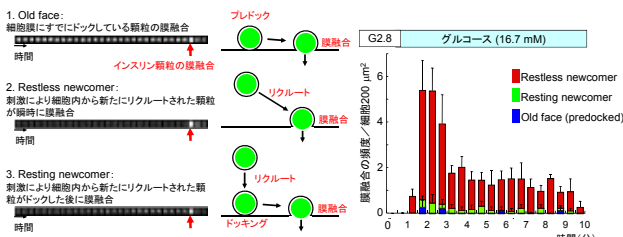


図4. インスリン顆粒の開口分泌様式(左)とインスリン顆粒動態(右)

TIRFM解析により、グルコースで刺激された膵β細胞では、インスリン顆粒は2相性の膜融合を示し、ほとんどがrestless newcomer、すなわち細胞内部から新たにリクルートされた顆粒によるものであった。

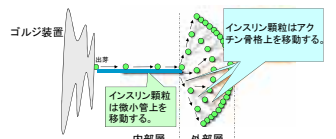


図5. インスリン顆粒動態モデルの概略

本モデルはグルコース刺激されたβ細胞におけるインスリン顆粒の膜融合過程だけでなく、インスリン顆粒の補充過程を取り入れたモデルである。

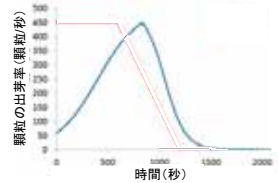


図6. ゴルジ装置でのインスリン顆粒の出芽数の設定

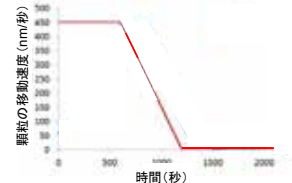


図7. 顆粒の微小管上での移動速度の設定

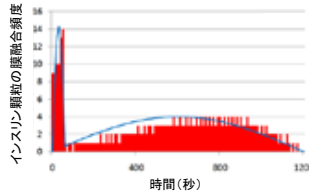


図8. シミュレーション結果

- グルコース応答性インスリン分泌の特徴である急峻な第1相と持続性の第2相が表現できた。
- 細胞膜に融合するインスリン顆粒数も実験結果に近いシミュレーション結果が得られた。

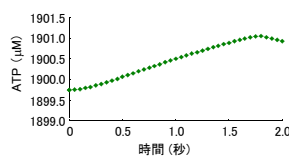


図9. 膵β細胞におけるATP濃度(左)とK<sub>ATP</sub>チャネル電流(右)のモデル化

- 膜電位の伝播機能: 電位を物質として扱い、膜電位の伝播を拡散で表現する。
- Hodgkin-Huxleyモデルの実装: 神経興奮による膜電位の変化を同モデルで計算する。(現段階では神経細胞のモデルを利用)

$$\frac{\partial V}{\partial t} = \nabla \cdot CV + H$$

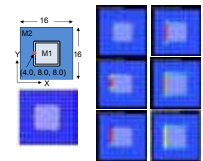


図10. 細胞膜上での膜電位の拡散

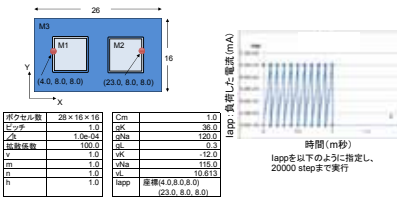


図11. Hodgkin-Huxleyの連続刺激テストの設定

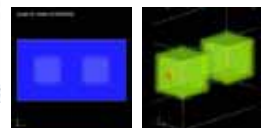


図12. 膜電位連続刺激テスト

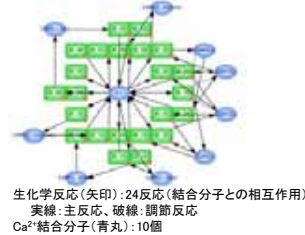


図13. 細胞内でのCa<sup>2+</sup>反応と拡散のモデル化

- RICS上でATP濃度変化、K<sub>ATP</sub>チャネル電流をモデル化した。
- RICS上で膜電位を表現可能なシステムを開発することで、膜電位伝播計算機能を実装した。
- 細胞内のCa<sup>2+</sup>緩衝作用を設定した場合、生化学反応が無かった時に比べ、Ca<sup>2+</sup>の見かけの拡散速度の低下が認められた。
- 実際の膵β細胞でみられるCa<sup>2+</sup>動態をRICS上でモデル化できることが示唆された。

### 研究参加者

神戸大学大学院工学研究科教授 玉置 久  
神戸大学大学院医学研究科客員准教授 福島光夫  
神戸大学大学院医学研究科講師 柴崎忠雄  
神戸大学大学院工学研究科修士課程1年 田畑直紀  
理化学研究所 横田秀夫  
理化学研究所 須永泰弘



図14. Ca<sup>2+</sup>の反応と拡散の連成 (ボリュームレンダリング)



# O-1: ZZ-EFSI

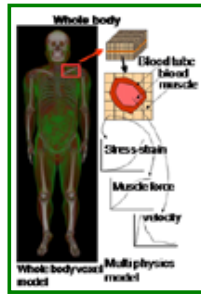
## 全身ボクセルシミュレーション(ボクセル構造流体連成解析プログラム)

開発責任者: 高木周 (理化学研究所 次世代計算科学研究開発プログラム)



### 開発に至る背景

- 医療画像(CTスキャン, MRI等)と相性の良い流体構造連成解析により, 個人毎の診断に即した治療支援を実現したい.
- 赤血球・血小板を含む血流を対象とする超大規模並列計算を容易に実現し, 血小板血栓に至る機構を明らかにしたい.



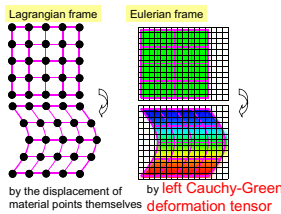
### 開発ソフトウェアの特徴

大規模並列計算に適した流体構造/膜連成解析を実現

#### Euler型流体・構造/膜連成解析法

- 固定メッシュ上で, 全ての物理量を更新できるように定式化.
- voxelデータの利用により, メッシュ生成・再構成を不要.
- 複雑な境界形状を持つ/多数の分散体を含む問題の扱いを容易に.

How is the solid deformation described?



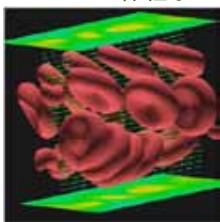
#### 動的パラメータを持つ擬似圧縮性法 (Artificial Compressibility Method; ACM)

- 圧力修正に際して, 速度発散を最小化する処理を導入.
- 数値的に安定で完全陽的な時間積分.
- 高い実効性能, 並列性能を実現.

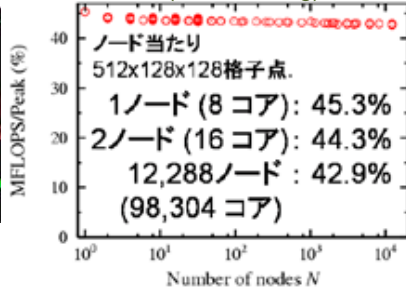
### 現時点での開発状況

- 開発言語とライブラリ: FORTRAN, C++, MPI, OpenMP, VSphere
- 「京」のマニュアルに従い, プロファイラ, PAツールを使って性能チューニング.
- 「京」で, ステンシル計算として極めて高い実効性能, 台数効果を達成.

チャンネル流れ内biconcave neo-Hooke体粒子



実効性能と並列台数 (weak scaling)



### 参考文献

Sugiyama, K., Ii, S., Takeuchi, S., Takagi, S. and Matsumoto, Y. (2010) Full Eulerian simulations of biconcave neo-Hookean particles in a Poiseuille flow, *Comput. Mech.*, **46**: 147-157.

Nagano, N., Sugiyama, K., Takeuchi, S., Ii, S., Takagi, S. and Matsumoto, Y. (2010) Full-Eulerian finite-difference simulation of fluid flow in hyperelastic wavy channel, *J. Fluid Sci. Tech.*, **5**: 475-490.

Sugiyama, K., Ii, S., Takeuchi, S., Takagi, S. and Matsumoto, Y. (2011) A full Eulerian finite difference approach for solving fluid-structure coupling problems, *J. Comput. Phys.*, **230**: 596-627.

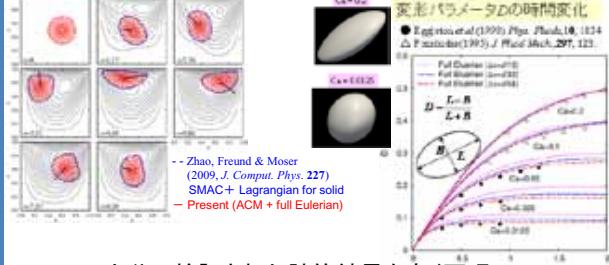
Ii, S., Sugiyama, K., Takeuchi, S., Takagi, S. and Matsumoto, Y. (2011) An implicit full Eulerian method for the fluid-structure interaction problem, *Int. J. Numer. Meth. Fluids*, **65**: 150-165.

Ii, S., Gong, X., Sugiyama, K., Wu, J., Huang, H., and Takagi, S. (2011) A full Eulerian fluid-membrane coupling method with a smoothed volume-of-fluid approach, *Comm. Comput. Phys.* (accepted).

Takagi, S., Sugiyama, K., Ii, S., and Matsumoto, Y. (2011) A review of full Eulerian methods for fluid-structure interaction problems, *J. Appl. Mech.* (accepted).

### 計算の妥当性 (流体構造/膜連成機能の実装)

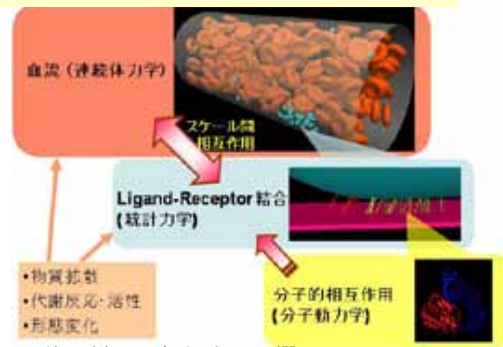
キャピティ流れ中の超弾性体粒子 せん断流中の膜カプセル



### 十分に検証された計算結果を良く再現 現在開発/実施中の計算事例 (血栓シミュレータ)

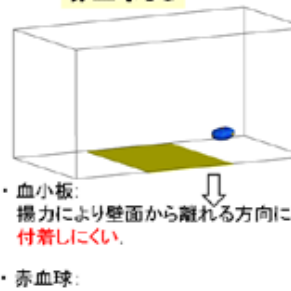
血小板血栓に至る機構解明を目指す

### 血流とLigand-Receptor結合のマルチスケール/フィジックス解析法の概要



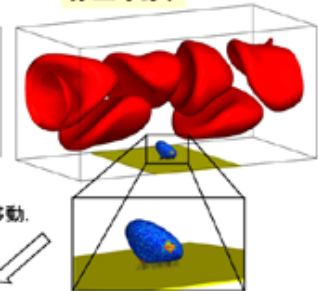
血小板付着の様子, 赤血球の影響

#### 赤血球なし



- 血小板: 摩擦力により壁面から離れる方向に移動. 付着しにくい.

#### 赤血球あり



- 赤血球: チャンネル流コア部に集積. 壁面垂直方向の速度成分の変動を誘起.
- 血小板: 壁面に接近する機会が増え, 付着しやすくなる.

血栓形成の第一段階における, 赤血球の存在の重要性.

⇒ 今後は, 多数の赤血球が存在する条件で, 血小板の付着から血栓の初期過程を「京」で解析





## O-2: TK-Thrombocyte

細胞シミュレーションプラットフォームを血栓止血の主要細胞である血小板に応用することによる統合的な血栓止血シミュレーターの基盤モデルの開発研究

開発責任者: 後藤信哉 (東海大学医学部内科学系循環器内科)

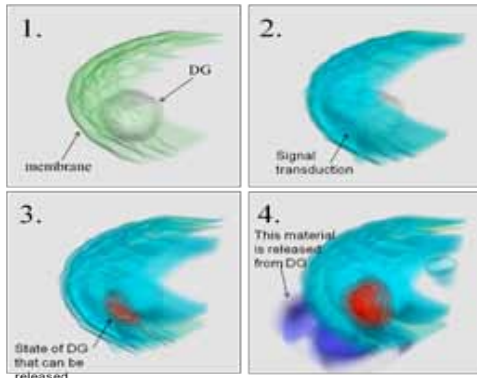


### 背景と目的

- 血小板細胞には核がなく、分裂増殖しない。生理機能は止血、血栓形成に特化している。
- 血小板細胞にはミトコンドリアなどの細胞内小器官が存在し、代謝、シグナル活力が高い。
- 血小板細胞活性化は、受容体刺激開始後数分にて完結するため、シミュレーションの対象時間が短い。また、東海大学拠点ではシミュレーションを補強する定量的実証実験の継続的施行が可能。

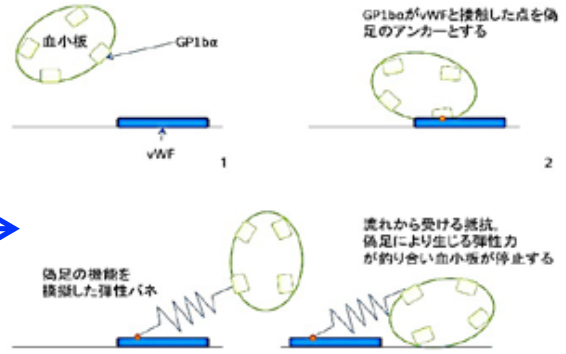
### 研究開発コードの概要とコード開発に向けたアプローチ

#### 血小板細胞放出シミュレーション



細胞スケールと臓器全身スケールの連携

#### 血小板細胞接着シミュレーション

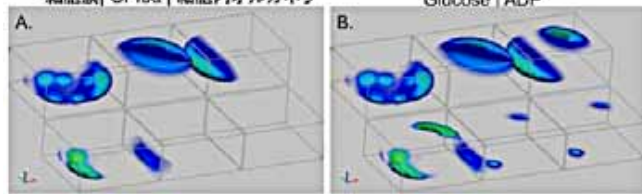
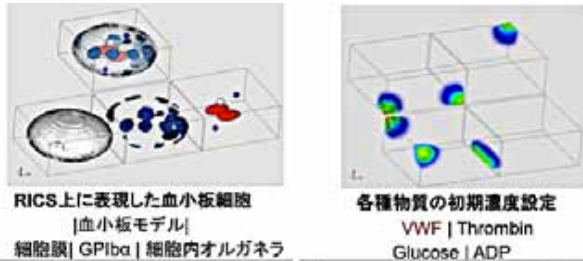


#### b. 臓器全身スケールシミュレーター (血小板細胞接着シミュレーター)

### 現時点での開発ソフトウェアとその成果

#### a. 細胞スケールシミュレーター (RICSの血小板細胞への応用)

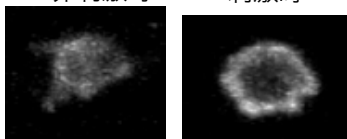
— 実証実験にて計測された細胞内小器官の局在、細胞表面上に局在するGPIIb/IIIaとvon Willebrand因子の相互作用による活性化シグナルの広がり、細胞内の濃染顆粒放出を組み込んだ。



放出反応を考慮した場合(A)には、放出反応を考慮しなかった場合(B)に活性化反応の広がりが効率的であることをシミュレーション計算により予測したシグナル1 | シグナル2 | シグナル3

活性化GPIIb/IIIa | 細胞外ADP | 細胞外ATP

リン酸化シグナル蛋白VASP-Pの局在— 実証実験による活性化刺激後のシグナル蛋白の細胞内局在を定量しシミュレーターを精緻化



— 「京」を用いたa, bの大規模連携計算により、細胞スケールと臓器全身スケールの統合的理解を目指す

— 実証実験に裏打ちされ、血小板細胞上の特異的接着蛋白(von Willebrand因子と結合するGPIIb/IIIa)の細胞表面上の局在を考慮し、接着した血小板細胞が受ける流体力による受動的形態変化(偽足形成)を組み込んだ血小板細胞接着シミュレーターを作成した

#### 計算例(血小板形態が球体と楕円体の場合の差異)

偽足長さの分布(計算値: 球体の場合)

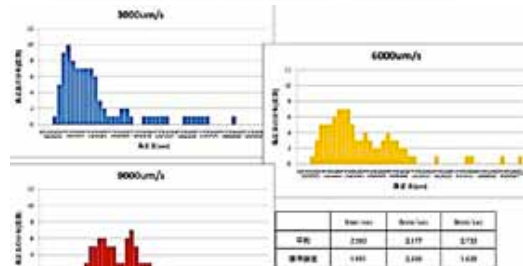
	ave.	s.d.	min.	max.
3000s <sup>-1</sup>	1.37881	0.013475	1.35500	1.40781
6000s <sup>-1</sup>	2.490625	0.026479	2.46661	2.53651
9000s <sup>-1</sup>	3.609572	0.032834	3.58564	3.65743

偽足長さの分布(計算値: 楕円体の場合)

楕円体形状x:y=1:1,x:z=2:1,y:z=2:1

	ave.	s.d.	min.	max.
3000s <sup>-1</sup>	1.315085	0.035819	1.27721	1.39528
6000s <sup>-1</sup>	2.561676	0.060653	2.51179	2.69274
9000s <sup>-1</sup>	3.802511	0.057597	3.74976	3.89441

モデルを楕円体形状にすることで偽足長にバラツキが生じる



流体の流れ場における偽足長の実測値

— 流れ場にて、von Willebrand因子に接着した血小板の偽足長を実測してシミュレーターを精緻化





# 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



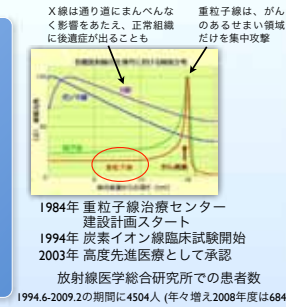
開発責任者: 石川 顕一 (理化学研究所臓器全身スケール研究開発チーム)

## 背景・目的

### がんの重粒子線治療

重粒子線 (加速器から得られる炭素イオンのビーム) で、非侵襲にがんを治す治療法

- ▶ 線量の集中性が高い (ピンポイントに照射可能)
- ▶ 強い生物効果 (細胞を死滅させる力が強い)
- ▶ 他の治療法では治らないがん (メラノーマ、肝臓がん、手術できない肉腫) が治る
- ▶ 同じ治るにしてもより短期間で安全に治りQOLが高い (X線なら20~40回照射、重粒子線なら1回照射)
- ▶ 高度先進医療として、近年、国民の注目が高まっている



重粒子線は目に見えないので、3D-CT画像を使って線量を計算し、治療計画を立てる

➡ 信頼性の高い線量計算法が必要

- 既存の計算手法: ペンシルビーム法等
- 長所: 簡便で計算時間が短い (数分)
  - 短所: 不均質な媒質中 (骨の近くなど) で誤差大 → がん病巣へは不十分な線量、正常組織に不要な被ばく

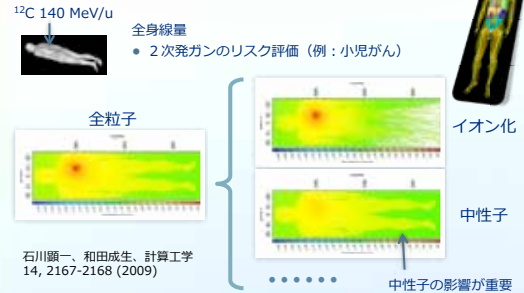
- 様々な組織から構成される不均質な体系についても高精度で線量計算できる手法が必要
- がん病巣への線量を十分に、正常組織の被ばくを低減

モンテカルロ法を用いた線量計算の実用化が切望されている

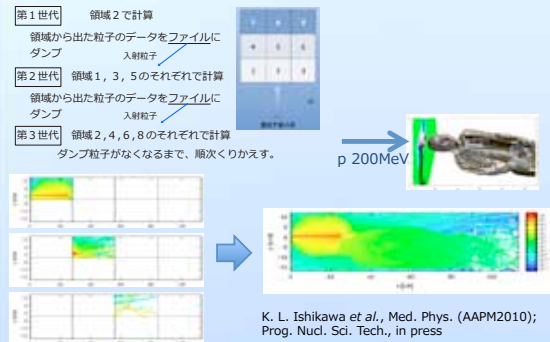
モンテカルロ法を用いた重粒子線線量計算シミュレーターを開発

## 研究開発成果

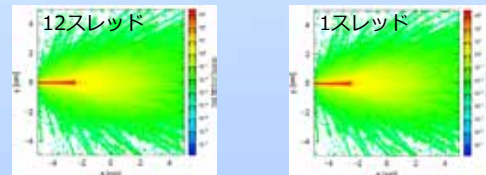
### ボクセルファントムを用いた全身線量計算



### 領域分割モンテカルロ法



### OpenMP-MPIハイブリッド並列



メモリー共有並列あり (左) およびなし (右) で計算した、水ファントム中の線量分布。全く同じ結果が得られている。

T. Furuta et al., Proceedings of 6th Japan-Korea Joint Meeting on Medical Physics and 11th Asia-Oceania Congress of Medical Physics (2011)

京でも8000コアまでの動作を確認

## 概要・アプローチ

### モンテカルロ法

粒子の輸送や一連の反応を、乱数を用いてそっくりそのままシミュレーション

粒子・重イオン輸送計算コードPHITS [http://phits.jaea.go.jp/] をベースに



- フォノンイマンとワウムが発見し、加速器実験・検出器・放射線遮蔽の計算に広く使われる
- 複雑な体系に対する精度の高い計算が可能
- 入射粒子数を増やすことによって、統計的精度を上げられる
- 並列化効率が高い

モンテカルロ粒子輸送計算では、通常、ヒストリー (入射粒子) を並列化し、すべてのProcessing Elementが全体系の情報を保持。

➡ 全身線量計算では膨大なメモリーが必要 (コア当たり数~数十ギガバイト) 京のスペック超え!

➡ 領域分割モンテカルロ法

➡ OpenMP-MPIハイブリッド並列

## 今後の展望

領域分割モンテカルロ法 } 同時実装  
OpenMP-MPIハイブリッド並列

➡ 0.3ミリ角での全身ボクセル線量計算 (1640×890×5630ボクセル)

➡ 治療計画の検証と高度化  
2次発がんリスクの評価  
低線量被ばくの長期的影響の評価 (福島原発事故で放出された放射性物質の健康影響)

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

**O-4: ZZ-HIFU**

**集束超音波による低侵襲治療のためのHIFUシミュレータ**

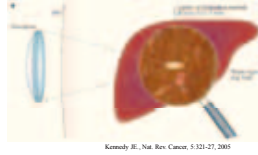


開発責任者: 松本洋一郎 (東京大学工学系研究科機械工学専攻)

**1. Background and Objectives**

**High-Intensity Focused Ultrasound therapy**

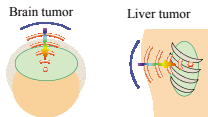
- Minimally-invasive treatment, which provides the heat coagulation of tissue around the focal point of ultrasound.
- Benefits
  - Short hospital stay
  - Radiation free
  - Repeatable treatment
- Approved for tumors such as the prostate hyperplasia and uterine fibroids.



**MRgFUS** has been developed for various tumors.

- Pre and Post contrast imaging for treatment validation
- Accurate targeting
- Real-time monitoring of temperature

**HIFU for tumor behind bones**



Displacement and diffusion of focal point due to the reflection and refraction of ultrasound at the interface of bones



Focus control by array transducer

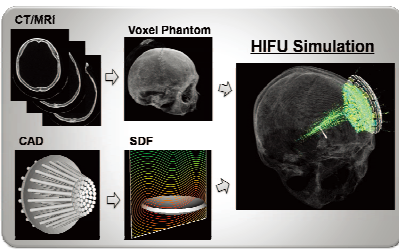
**Simulation assisted focusing methods**

- A layered wavevector-frequency domain model using input from CT scans (Clement & Hynynen, 2002)
- A time-reversal process based on prior CT scans using the wave propagation simulation (Aubry et al., 2003)
  - Digital human model is constructed based on CT images
  - Control parameters of array transducer is obtained by a simulation

**Objectives**

- Realization of the appropriate focus control in the body using array transducer
- Prediction of the treatment region for the preoperative planning
- Support of the development of HIFU device from design to approval

**2. HIFU Simulator**



The media of a volume model of human body (voxel phantom) is discretized into cubic elements with the voxel data for a living human body collected by CT/MRI technologies. On the other hand, the shape of the transducer defined by CAD is represented by the volume data of the signed distance function. Then, the ultrasound propagation from the transducer through the voxel phantom is performed by large-scale parallel computing.

**Basic equations**

■ The mass conservation equation for mixture:

$$\frac{1}{\rho_m} \frac{\partial \rho_m}{\partial t} + \frac{\partial}{\partial x_j} \left( \frac{\rho_m v_j}{\rho_m} \right) = \frac{\partial E}{\partial t}$$

where  $\rho_m = \sum_i \rho_i \rho_i$  and  $\frac{1}{\rho_m} \frac{\partial \rho_m}{\partial t} = \sum_i \frac{\partial \rho_i}{\partial t} \frac{\rho_i}{\rho_m}$  sound source due to the transducers

■ The equation of state of media (Tait's equation)

$$\rho = \rho_0 \left( \frac{p + A}{\rho_0 + A} \right)^\gamma$$

where  $A = \frac{\rho_0 c_0^2}{\gamma}$ ,  $\rho_0 = \rho_0$  and  $\gamma = \frac{B}{A} + 1$  Non-linear coefficient

■ The momentum conservation equation for mixture:

$$\rho_m \frac{\partial v_i}{\partial t} + \frac{\partial}{\partial x_j} \left( \rho_m v_i v_j + 2\mu_{ij} \left( v_i - \frac{1}{3} v_k v_k \right) \right)$$

■ The heat equation for mixture:

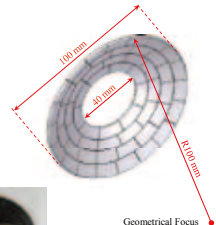
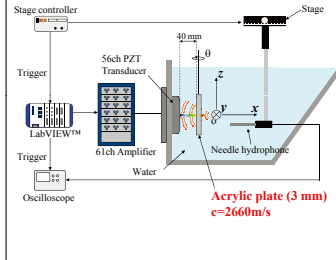
$$\rho_m C_p \frac{\partial T}{\partial t} = \frac{\partial}{\partial x_j} \left( \kappa_{ij} \frac{\partial T}{\partial x_j} \right) + \zeta_{ij} (v_i v_j) + 2\mu_{ij} \left( v_i - \frac{1}{3} v_k v_k \right)$$

**Numerical method**

- Basic equations are discretized by the **Finite Difference Method** of 6-th order central difference in space and are explicitly integrated in time based on FDTD method.
- Perfectly Matched Layer** is employed for the non-reflecting boundary.
- Parallel computing** is performed by the domain decomposition with MPI + OpenMP.

**3. Simulation Assisted Focus Control**

**Experimental Apparatus**



- 56 elements
- Frequency 2 MHz
- Focal distance: 100 mm

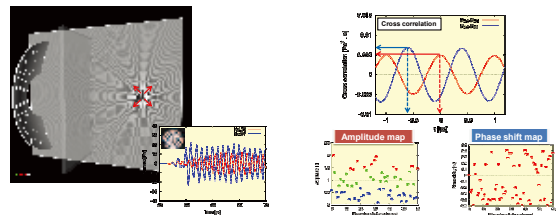
The acrylic plate as a phantom of bone is inserted between the transducer and target. The angle of the acrylic plate can be changed. The transducer is driven by 61ch amplifier under the control of LabVIEW. The acoustic field is measured by using needle hydrophone 3-dimensionally

**Method & Results**

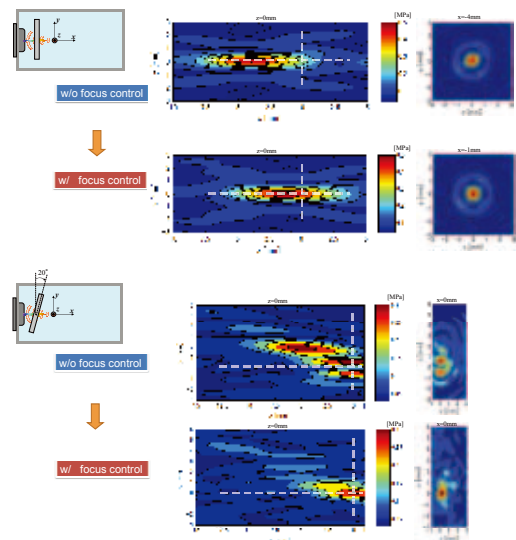
**I. Estimation of control parameters using the NUMERICAL SIMULATION**

1. Ultrasound is emitted from a target point and received by the array transducer

2. Take cross correlations of the received pressure and choose the amplitude and phase delay to make the cross correlation maximum.



**II. Transducer is driven by the estimated parameters in the EXPERIMENT**



Since the sound speed of the acrylic plate is faster than water, ultrasounds focused away from the target in both cases without focus control. By driving the array transducer with the control parameters which were previously obtained by the numerical simulation, the focal points were assigned to the target correctly in the experiments with focus control. Therefore, the simulation assisted focus control is effective and the numerical simulation is so accurate that the focus control is enabled.





## O-5: ZZ-LUNG

### 肺呼吸・肺循環シミュレーション



開発責任者: 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)

#### ソフト開発の背景

厚生労働省の人口動態統計(2010年版)によると、日本人の死因順位において肺炎が4位、慢性閉塞性肺疾患が9位を占めている(図1)。このように呼吸器系疾患は、多くの国民の生命に関わる重大な疾患であるにも関わらず、薬剤治療や外科治療の選択

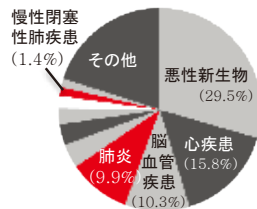


図1. 日本人の死因順位

肢が乏しく、効果的な治療法の開発が大きな課題となっている。また、重篤な疾患の際に呼吸補助に利用されることが多い人工呼吸器に関しても、外部から強制的に換気を行うことで肺の末梢組織に損傷が発生し、患者の予後を悪くする大きな要因となっている。

これらの問題の解決を阻む根本的な要因として、肺の末梢部位における空気の流れとそれに伴う組織の変形の関係が明確にされていないという事実がある。肺は、約200ミクロンの肺胞と呼ばれる房状の組織が連結した複雑な形状をしており、横隔膜の上下運動によって肺胞の集合体を受動的に変形し、それによって肺内部に空気の流れが発生する。しかし、気流の速度や組織に作用する応力を生体内で計測することは、最先端の計測技術を用いても極めて困難である。一方で、計算機性能の向上によって、複雑な形状の物体と流体の相互作用を解析するシミュレーションが近年、ようやく実用レベルで可能になりつつある。

そこで、我々は、神戸に建設中の京速コンピュータ「京」を用いた大規模計算によって、呼吸で生じる肺の末梢部位におけるダイナミクスを解明するため、シミュレーションソフトの開発を進めている。

#### 開発コードの概要

本開発コードは、有限差分法に基づく全身ボクセルシミュレーション構造流体連成解析プログラムZZ-EFSIをベースとしている(詳細は、O-1: 理化学研究所・高木周のポスターをご覧ください)。

計算領域は、空間上に固定した一様な3次元の直方体格子で全体を分割し、VOF関数によって、空気が流れる気道領域、血液が流れる毛細血管領域ならびに肺組織を区別している(図2)。このため、並列計算に当たっては、矩形領域を均等分割することで、効率的なロードバランスを実現することが可能である。

また、理化学研究所のVCADシステム研究プログラムで構築されたV-Sphereと呼ばれる開発環境を利用してMPIとOpenMPのハイブリッド並列を実現している。開発言語もC++とFortran90のハイブリッドとなっている。

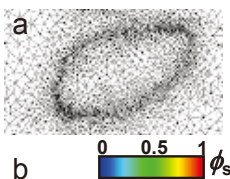


図2. (a) 流体構造線成解析に用いられることが多い非構造格子の一例(出典: Gao & Hu J. Comp. Phys., 2009)。構造体の境界付近に多くの格子を配置し、構造体の移動とともに格子も移動もしくは再定義する。(b) 空間に固定した直方体格子の一例。構造体の運動に伴って時間発展するVOF関数を用いて流体と構造体を区別する。

#### 計算事例

現段階では、コードの信頼性を確認しつつ、解析すべき現象の基本的な振る舞いを把握するため、肺末梢部の形状をシンプルな幾何形状で模擬した系において計算を行っている。すなわち、肺胞管を円柱、その周りを取り囲む肺胞を球で模擬した幾何形状を構築し、その内部に呼吸による気流が発生するようにしている。さらに、球の周りに規則的な穴の空いた球殻を配置し、血流が流れる毛細血管を模擬した(図3)。その計算例を示したものが図4である。

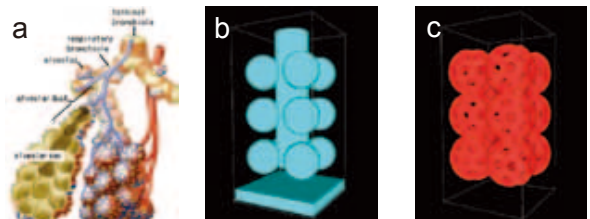


図3. (a) 肺末梢部位の模式図(出典: 牛木・小林著「人体の正常構造と機能 I 呼吸器」) (b) 計算に用いた気道形状の鳥瞰図 (c) 毛細血管形状の鳥瞰図。肺胞管と12個の肺胞から成る系を128x128x256の計算格子で作成。肺胞管ならびに肺胞の直径は50格子、毛細血管の厚さは7格子で表現されている。

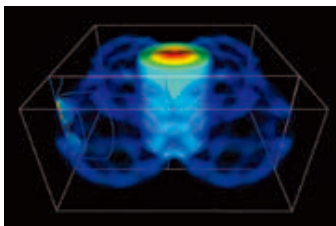


図4. 計算で得られた速度分布のポリウムレンダリング。図3の計算形状の下端を強制的に上下に振動させながら、毛細血管内部にだけ水平方向に圧力勾配が発生するように境界条件を設定して計算を行った。ただし、図で示されているのは、計算形状の一部のみ。

#### 今後の展開

今後は、より現実的な系において計算を実施し、末梢気道における気流の閉塞が肺組織の応力場にどのような影響を与え、肺胞構造の破壊を引き起こすかについて解析を進めていく予定である。さらには、この解析を通して、肺炎や慢性閉塞性肺疾患などの病態をより深く理解し、効果的な治療につながる示唆を得たいと考えている。

人工呼吸器による肺損傷を抑制できる喚起の制御手法を模索することも重要なトピックだと考えている。酸素と二酸化炭素の動態を解析するモジュールも精緻化し、循環器系の理解に重要なガス交換プロセスについても解析を進めていく方針である。

なお、ソースコードは将来的には無償で公開する予定であるが、現時点では開発途上であるため、利用希望者は、共同研究という形式で、開発者と協議しながら適用範囲を検討していくことを想定している。

#### 謝辞

本研究は、文部科学省 最先端・高性能汎用スーパーコンピュータの開発利用「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」の支援を受け行われたものです。また、理化学研究所が実施している京速コンピュータ「京」の試験利用において、開発コードのテスト計算をさせていただきました。記して感謝いたします。





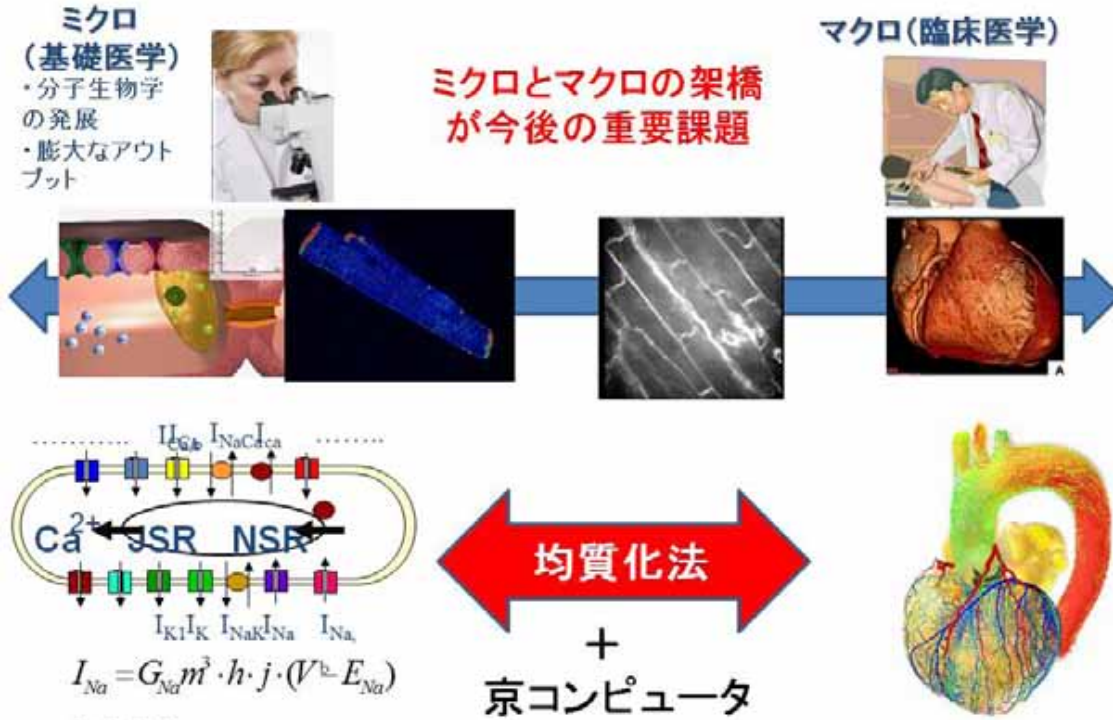
## O-6: UT-Heart

マルチスケール・マルチフィジックス心臓シミュレータ

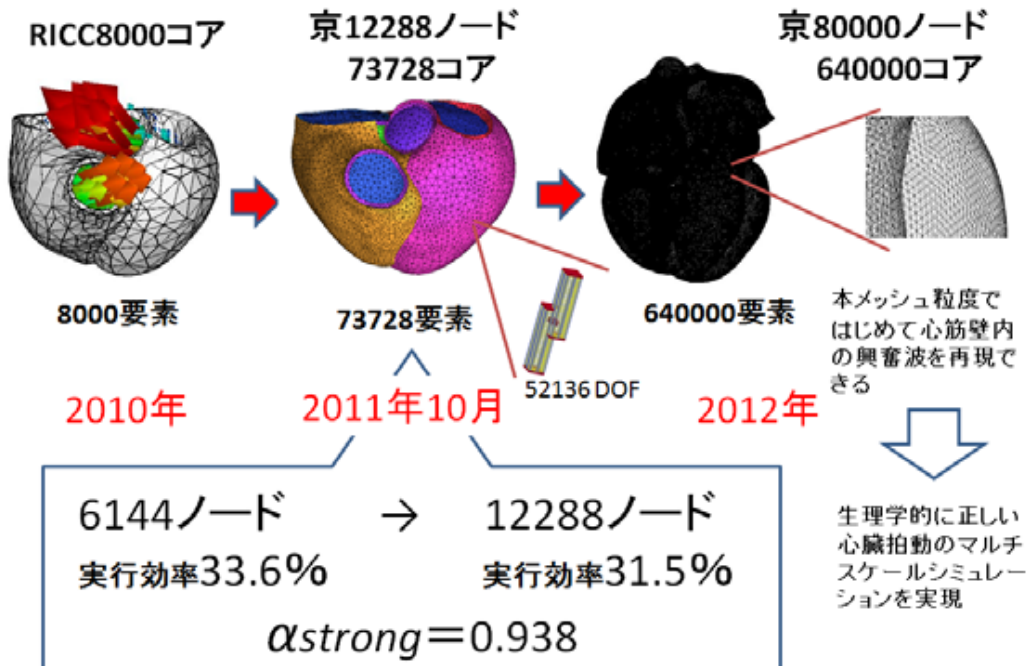


開発責任者: 久田俊明 (東京大学)

### マルチスケールシミュレーションの意義



### UT-Heart ロードマップ



上記データは京の試験利用並びに理化学研究所情報基盤センター-RICCにより得られた。京は現在開発中であり、これらの数字は現状の値である。



## 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

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、次世代計算科学研究開発プログラム)  
三澤計治 (理化学研究所次世代計算科学研究開発プログラム)

#### •ParaHaplo開発に至る背景

- ゲノムワイド関連解析とは、病気の人を集めて、病気じゃない人のDNAをゲノム全体に渡り頻度を統計検定することにより、疾患関連遺伝子を探る手法である。
- 2002年に理研で世界に先駆けて始められた
- ゲノム全体に対して統計検定を行うので、100万回以上の多重検定を行うことになる。そのため、疾患関連遺伝子でなくとも頻度に有意差があると判定される場合が出てくる。これを偽陽性という。偽陽性を減らそうとすると、真の陽性を見落とす可能性が高まる。
- ハプロタイプ単位でゲノムワイド関連解析をすることで真の陽性を見落とす可能性を減らし、より多くの疾患関連遺伝子を発見できることが期待される。そこからさらに疾患関連遺伝子ターゲットに医薬品や治療法の開発へと繋げることを目指す(Misawa et al. 2008)。

#### •現時点の開発ソフトウェアについて

- ParaHaploはPCクラスタで8,000ノードの並列性能を実証し、京速コンピュータ「京」でも並列性能が出ることが確認されている。「京」ではコア並列も実装している(Misawa and Kamata 2009, 2010, 2011)。
- 現在は、一層の高速化を目指しSIMD化を行っている

#### •実際のデータを用いParaHaploを動かした例

##### •材料と方法

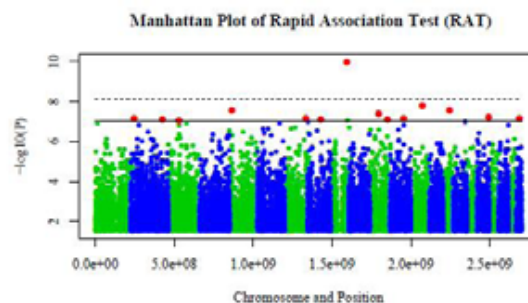
- ParaHaploにより、東京日本人44人(JPT)のサンプルと北京中国人45人(CHB)のサンプルのハプロタイプ頻度を比較し統計検定を行った

##### •結果

- ParaHaploによるハプロタイプ頻度の検定の結果がカズに示されている。
- 点線より上が従来の方法で遺伝子頻度の差が検出できた場所、実線より上がParaHaploで遺伝子頻度の差が検出できた場所である。
- 23本の染色体が濃淡で表されている。

##### •考察

- 従来法よりも多くの遺伝子頻度の差が検出できることが分かる。
- 今後は数千人規模の実際の疾患データを用い、疾患関連遺伝子を見つけることを目指す。
- 計算時間が今以上にかかるので、京速コンピュータ「京」を利用する。



#### 参考文献

- Misawa K, Fujii S, Yamazaki T, Takahashi A, Takasaki J, Yanagisawa M, Ohnishi Y, Nakamura Y, Kamatani N (2009) 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 studies using parallel computing. *Source Code Biol Med* 4:7



## D-2: NGS analyzer

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

開発責任者: 角田達彦 (理化学研究所分子スケール研究開発チーム)  
開発担当者: 藤本明洋 (理化学研究所分子スケール研究開発チーム)



- 概要

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

- 離散化(計算モデル化)の方法

ヒト標準ゲノム配列に対するマッピングと確率計算に基づいた多様性検出  
計算方法直接法による密行列の対角化並列化の方法領域分割

- 開発言語とライブラリ

Perl, C

- 現状での計算規模

初の日本人ゲノムシーケンス解析 (**Nat Genet** 42: 931-6 )

肝がんゲノム解析 (**Nature** 464, 993-8 )

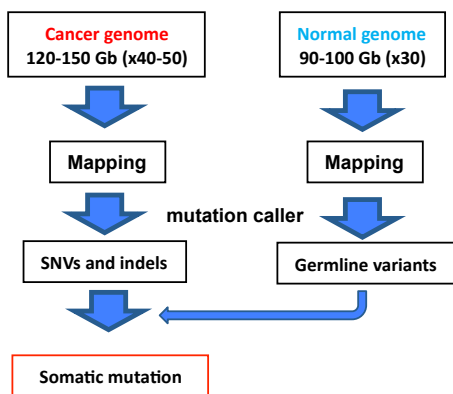
PCクラスタ2000コアで実行メモリ容量 4 TB、ディスク容量 100 TB

- 次世代機「京」での計算規模

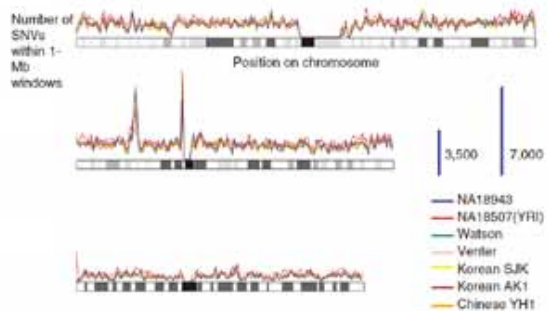
500人のがんゲノム配列解析による突然変異同定

64万コアで実行メモリ容量 2PB、ディスク容量 50 PB

### がんゲノム解析パイプライン

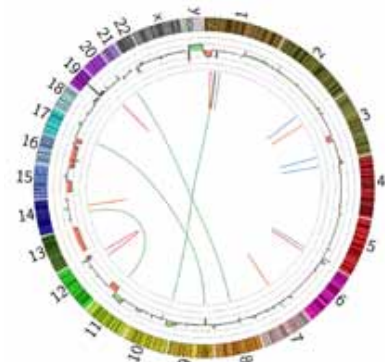
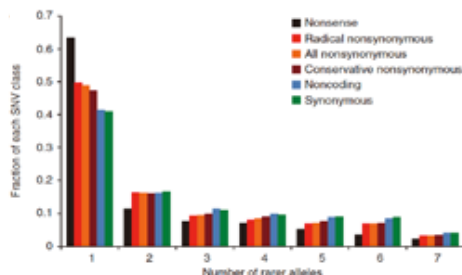


### ヒトの遺伝的多様性(1Mbp当たりのSNPの密度)



### 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







### 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

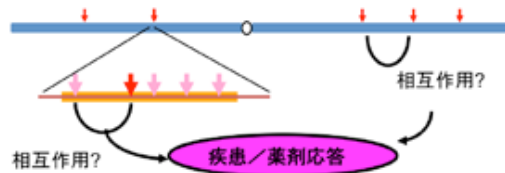


開発責任者: 角田彦彦 (理化学研究所ゲノム医科学研究センター)

#### 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している。

#### 相互作用



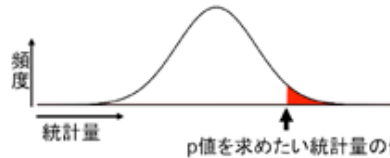
複数のSNPの組み合わせで疾患/薬剤応答に関連をもつ

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

#### Importance Sampling

パーミュテーションによって得られる統計量の分布



赤い部分のみをサンプリングできる分布を事前に定義

赤に含まれるある1個のパーミュテーションの生起確率

サンプリング数を十分にとって足しあわせる → 赤の比率

#### 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 +  $\chi^2$ 検定 → 尤度比検定に拡張

#### RAT法の拡張

より複雑な解析に対応するために

各種線形回帰モデルと尤度比検定への拡張

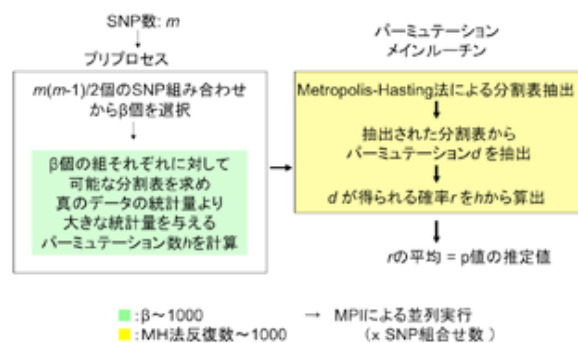
#### 2SNPの相互作用付きロジスティック回帰モデル

ケースとなる確率P

$$\text{logit}(P) = \mu + \alpha X_A + \beta X_B + \gamma X_A X_B$$

↑ SNP A の効果    ↑ SNP B の効果    ↑ SNP間の相互作用

#### ExRAT: 並列化



開発言語とライブラリ:

C++, MPI, OpenMP

現状での計算規模:

- 10万SNP、4000人分のタイピングデータ、2SNP全組合せ
- 50億組合せ × 4000人分のSNPデータ
- RICC 8192コア、メモリ容量 1.1 GB、ディスク容量 2 GB

次世代機「京」での計算規模

- 全染色体上SNPの1万人規模、2SNP全組合せで解析
- 2500億組合せ × 1万人 × 47疾患のSNPデータ
- メモリ容量 20 GB、ディスク容量 2 TB

どんなことができるか

- 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



# 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)

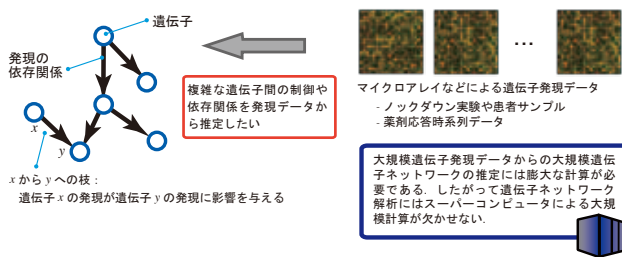
SiGN 開発メンバー: 玉田 嘉紀, 島村 徹平, 山口 類, 新井田 厚司, 齊藤 あゆむ, 長崎 正朗, 井元 清哉, 宮野 悟

### 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ベイジアンネットワーク、状態空間モデル、グラフィカルガウシアンモデル、ベクトル自己回帰モデルが遺伝子ネットワークモデルとして利用可能である。**SiGN-BN** はこれらのうちベイジアンネットワークを用いた遺伝子ネットワーク推定ソフトウェアで、スーパーコンピュータを用いた並列計算によりヒト全ゲノムを網羅した数万遺伝子の遺伝子ネットワークを短時間で推定可能である。

Visit SiGN Web Site at <http://sign.hgc.jp>

遺伝子ネットワーク = 遺伝子発現の依存関係をネットワークで表現したモデル



### ノンパラメトリック回帰によるベイジアンネットワーク

**Bayesian Network Model** Imoto et al. [2002]

Representation of the joint probability distribution by the DAG (Directed Acyclic Graph) structure

Gene Expression = Random Variable  
Node = Gene  
Directed Edge = Dependency

$$\pi(G|X) \propto \pi(G) \prod_{i=1}^n \prod_{j \in \text{pa}(i)} f(x_i | p_{ij}, \theta_{ij}) \prod_i \theta_i$$

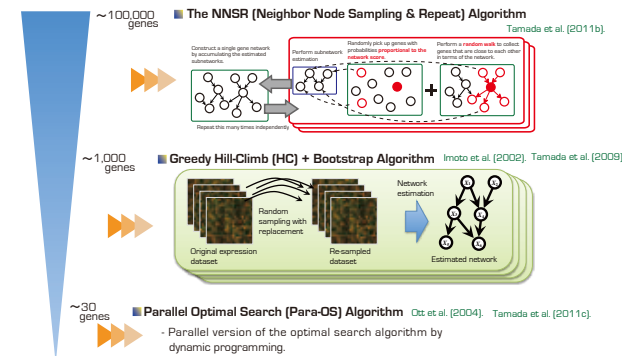
G: gene network X: gene expression data

The parent-child relationship is modeled by the B-spline nonparametric regression

$$f(x_i | X_{\text{pa}(i)}) = m_{ij}(p_{ij}^{(1)}) + \dots + m_{ij}(p_{ij}^{(k)}) + \epsilon_{ij}$$

$$m_{ij}(p_{ij}^{(k)}) = \sum_{l=1}^k \gamma_{ijl} B_l(p_{ij}^{(k)})$$

### Parallel Algorithms for Estimating Gene Network Structures



### Reference

Aifara, M. et al. (2007). Understanding endothelial cell apoptosis: What can the transcriptome glycome and proteome reveal? *Philosophical Transactions of Royal Society B* 362 (1484), 1469-1487.

Araki, H. et al. (2009). Analysis of PPARα-dependent and PPARα-independent 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. Biocomput.* 7, 175-186.

Imoto, S. et al. (2006). Computational strategy for discovering druggable gene networks from genome-wide DNA expression profiles, *Pacific Symposium on Biocomputing (PSB 2006)*, 11, 559-571.

Ott, S. et al. (2004). Finding optimal models for small gene networks, *Pac. Symp. Biocomput.* 9, 557-567.

Tamada, Y. et al. (2009). Unraveling dynamic activities of autocrine pathways that control drug-response transcriptome networks, *Pacific Symposium on Biocomputing (PSB2009)* 14, 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. bio. 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.

### 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」上で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 法を用いることでより高信頼ネットワークを計算可能
  - マイクロアレイだけでなく様々な事前知識をネットワーク推定に利用可能
- 出力**
  - 出力データ: 遺伝子ネットワーク (テキスト/CSML)
  - 推定された枝に信頼度等の情報付き
  - Cell Illustrator で解析可能
- 並列性能 (HC+Bootstrap)**
  - 最大動作コア数: 30,720 コア (3,840 ノード)
  - 並列化効率: 0.91 (30,720 vs 15,360 コア時) (ストロングスケール)
  - ※ 2011/12/1 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予定
- 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 ~ 100,000 ノード

### これまでの応用例

- 肺癌細胞株 PC-9+Gefitinib 投与による薬剤応答パスウェイ解析**
  - 使用データ**
    - 細胞: 肺癌細胞株 PC-9
    - 薬: Gefitinib (イレッサ: 抗がん剤)
    - マイクロアレイ: Agilent 44K Whole Human Genome Oligo Microarray
    - 計測データ: 26 時点 × 1 回計測 = 26 アレイデータ
    - EGF, Gefitinib 投与・非投与の 4 条件
  - 目的**
    - Gefitinib を投与した肺癌細胞株 PC-9 での遺伝子ネットワークの動的変化を見ることで、未知作用機序の同定
  - 利用手法**
    - 動的ベイジアンネットワーク (HC) による Node-Set Separation 法
  - 結果**
    - Node-Set Separation 法により時点毎のネットワークが得られ時間による変化が解析可能
- ヒト臍帯静脈内皮細胞 (HUVEC) によるアポトーシス関与遺伝子同定** Affara et al. [2007]
  - 使用データ**
    - 細胞: ヒト臍帯静脈内皮細胞 (HUVEC)
    - 刺激: 血清除去 & siRNA ノックダウン
    - マイクロアレイ: CodeLink UniSet Human 20K gene chip
    - 計測データ: 8 時点 × 3 回計測 時系列データ
    - 8 siRNA ノックダウンアレイ
    - 計 32 アレイデータ
  - 目的**
    - アポトーシスを誘導した HUVEC を時系列に観測しネットワーク解析をすることによりアポトーシス制御に関わる関連遺伝子の同定
  - 利用手法**
    - 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/Paclitaxel による新規抗癌剤ターゲットの同定

**PR** 文部科学省 新学術領域研究「システムがん」

SiGN is also used in the **Systems Cancer Project** that started newly in 2010. Visit our web site at <http://systemscancer.hgc.jp> and follow our tweets @Systemscancer in Twitter for updates.

### Acknowledgments

Computational time was provided by the Super Computer System, Human Genome Center, Institute of Medical Science, The University of Tokyo, RIKEN Integrated Cluster of Clusters (RICC) system, and early access to the K computer at the RIKEN Advanced Institute for Computational Science.

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



# D-5: SiGN-L1

## L1正則化法による大規模遺伝子ネットワーク推定プログラム



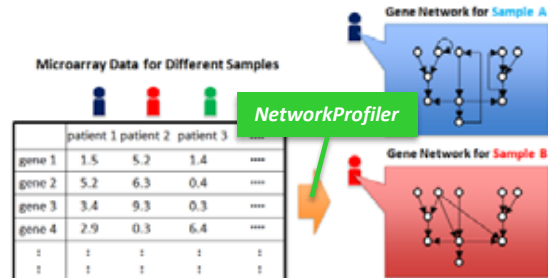
開発責任者: 宮野 悟 (東京大学 医科学研究所 ヒトゲノム解析センター)

SiGN 開発メンバー: 玉田 嘉紀、島村 徹平、山口 類、新井田 厚司、斉藤 あゆむ、長崎 正朗、井元 清哉、宮野 悟

### 概要

SiGN (サイン) は遺伝子ネットワーク発現データからスーパーコンピュータを用いて遺伝子ネットワークを予測・推定するソフトウェア群 (SiGN-BN, SiGN-L1, SiGN-SSM) である。このうち、SiGN-L1 は L1 正則化法によるスパースな統計的グラフィカルモデル (グラフィカルガウシアンモデル、ベクトル自己回帰モデル、構造方程式モデル) を用いた遺伝子ネットワーク推定ソフトウェアであり、スーパーコンピュータを用いた並列計算により、ヒト全遺伝子を対象とした大規模遺伝子ネットワークを推定することが可能である。推定された遺伝子ネットワークを解析することにより、観測データから薬剤標的遺伝子の同定、創薬ターゲット・毒性関与パスウェイの探索などの応用が期待される。

### Case 4. 異なるサンプルに対する遺伝子発現データからサンプル各々のネットワークを推定



### SiGN-L1の特徴

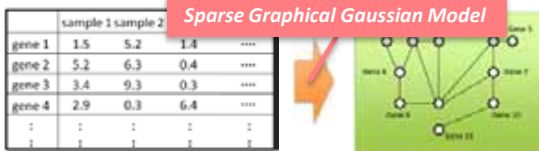
- 3種類の統計的グラフィカルモデル (グラフィカルガウシアンモデル、ベクトル自己回帰モデル、構造方程式モデル) により、遺伝子発現データから遺伝子ネットワークモデルを構築
- L1 正則化法を用いてスパースなモデルのパラメータ空間を探索することにより、100,000 ノードまでの大規模遺伝子ネットワーク推定が可能
- 東京大学医科学研究所ヒトゲノム解析センター (HGC) のスーパーコンピュータシステム (Xeon/Opteron PC クラスター) 上で、Sun Grid Engine を用いた並列化により 1,024 コアの並列化を達成。今後、RICC、RICC-FX、「京」上で動作を検証予定。

### これまでの応用事例

遺伝子ネットワーク = 遺伝子発現間の依存関係を表したグラフ構造

### 入力と出力

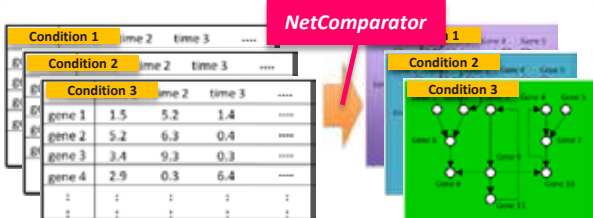
#### Case 1. Sample A に対する遺伝子破壊株発現データから遺伝子ネットワークを推定



#### Case 2. Sample A に対する時系列遺伝子発現データから遺伝子ネットワークを推定



#### Case 3. Sample A に対する異なる実験条件下の時系列遺伝子発現データから各々の実験条件下での遺伝子ネットワークを推定



### 参考文献

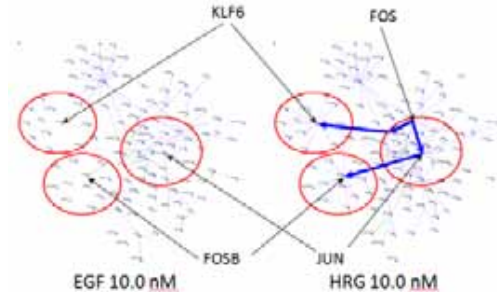
- Shimamura et al. (2007) Weighted lasso in graphical Gaussian modeling for large gene network estimation based on microarray data. *Genome Informatics*, 19, 142-153.
- Shimamura et al. (2009) Recursive regularization for inferring gene networks from time-course gene expression profiles. *BMC Systems Biology*, 3, e41.
- Shimamura et al. (2010) Inferring dynamic gene networks under varying conditions for transcriptomic network comparison. *Bioinformatics*, 26(9), 1064-1072.
- Shimamura et al. (2010) Collocation-based sparse estimation for inferring continuous-time dynamic gene networks. *Genome Informatics*, 24, 164-178.
- Shimamura et al. (2011) A novel network profiling analysis reveals system changes in epithelial-mesenchymal transition. *PLoS ONE* 6(6): e20804, doi:10.1371/journal.pone.0020804, 2011.

### 謝辞

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.

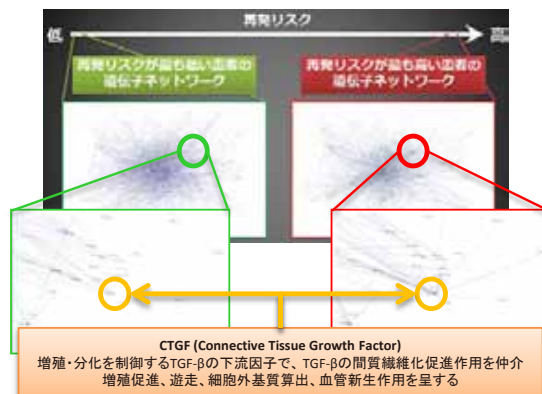
### 二つの異なる成長因子による細胞内の情報伝達パスウェイの比較

データ: 8実験条件下において8時点で計測した時系列遺伝子発現データ (Nagashima et al., 2007)  
 目的: EGF、HRG 各々の成長因子によって誘導される 8 実験条件下での遺伝子ネットワークを推定し、EGF刺激、HRG刺激下のシステムの違いを比較



### 再発リスクに関わる日本人肺癌患者の遺伝子ネットワークの比較

データ: 226症例の肺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)

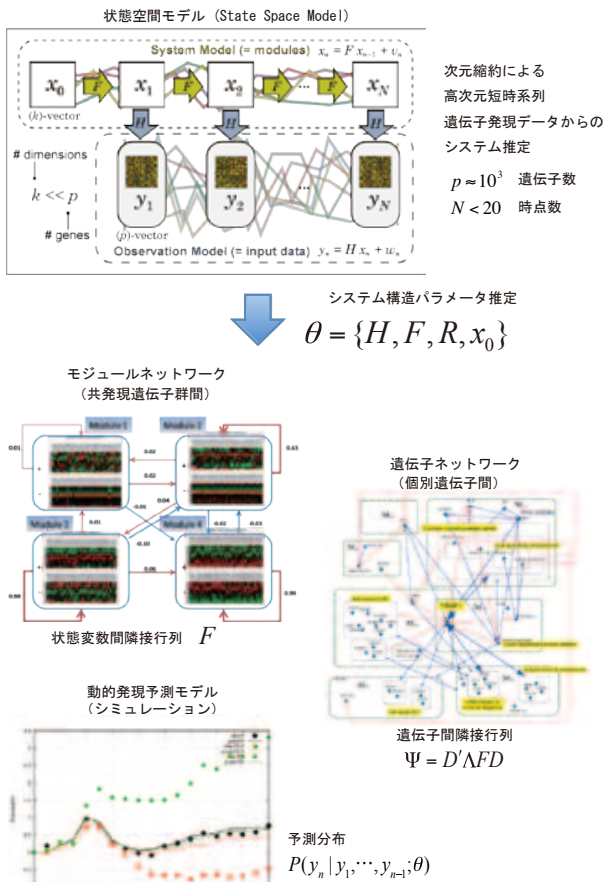
SiGN 開発メンバー: 玉田 嘉紀, 島村 徹平, 山口 類, 新井田 厚司, 斉藤 あゆむ, 長崎 正朗, 井元 清哉, 宮野 悟

#### 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ベイジアンネットワーク、状態空間モデル、グラフィカルガウシアンモデル、ベクトル自己回帰モデルが遺伝子ネットワークモデルとして利用可能である。**SiGN-SSM** はこれらのうち状態空間モデルを用いたネットワーク推定ソフトウェアで、スーパーコンピュータを用いた並列計算により、時系列遺伝子発現データから数千遺伝子のモジュールおよび遺伝子ネットワークの推定、および予測モデルの構築が可能である。

Visit SiGN Web Site at <http://sign.hgc.jp>

#### 状態空間モデルによる時系列遺伝子発現データからの情報抽出



#### Reference

Yamaguchi and Higuchi. State-space approach with the maximum likelihood principle to identify the system-generating time course gene expression data of yeast. Int J of Data Mining and Bioinformatics, 1(1): 77-87, (2006).

Yamaguchi et al. Finding module-based gene networks with state-space models - Mining high-dimensional and short time-course gene expression data. IEEE

Signal Processing Magazine, 24(1): 37-46 (2007).

Hirose et al. Clustering with time course gene expression profiles and the mixture of state space models. Genome Informatics, 18: 258-266, (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).

#### 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。欠測および不規則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 (制御構造に差異を遺伝子群の予測)

##### 出力

- 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: 8,192 コア
- 並列化効率: 0.98 (256 コア時) (ストロングスケール)
- ※ 2011/10/31 現在 RICC, R100-FX, H60 スバコン 上で検証中

##### 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

##### 「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)

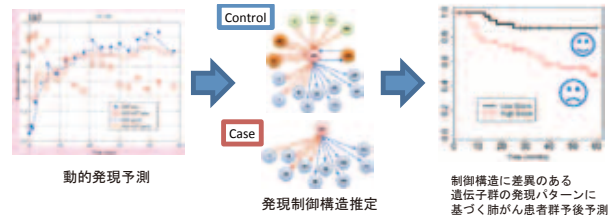
##### ソフトウェア配布形態

- 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变化があった遺伝子群の予測および、肺がん予後予測バイオマーカー探索
- 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



##### 肺がん細胞株 PC9/PC9GR に対する Gefitinib 動的効果予測

- 使用データ: 細胞: 肺がん細胞株 PC9 および PC9GR 刺激/薬: EGF, Gefitinib (イレッサ: 抗がん剤) マイクロアレイ: Agilent 44K Whole Human Genome Oligo Microarray
- 計測データ: 26 時点 × 1 回計測 in 26 時間
- 目的: 薬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
- 結果: 薬剤効果プロファイル
- 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。

##### 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
- メラノマ細胞/Paclitaxel による新規抗がん剤ターゲットの同定

#### 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.



## D-7: SBiP

### データ解析融合プラットフォーム

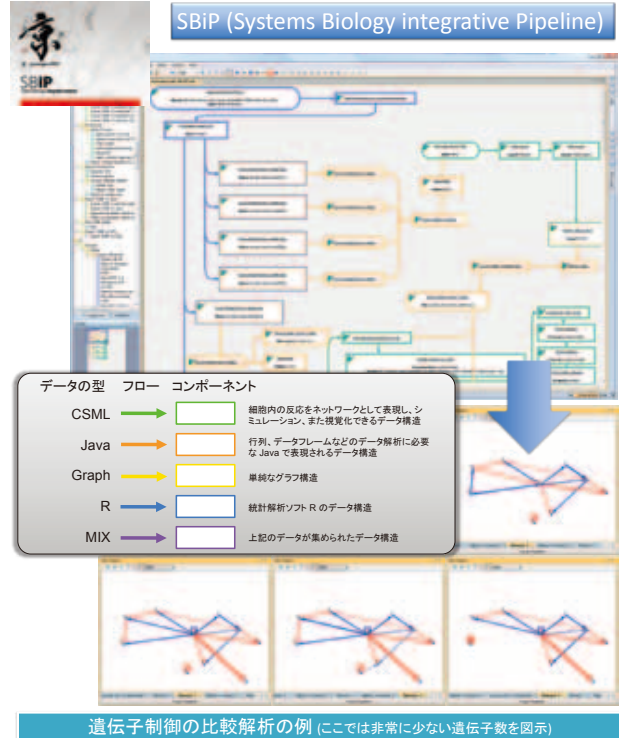
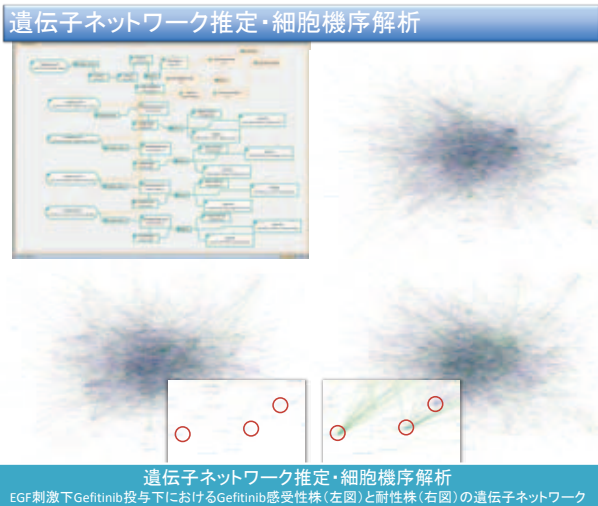
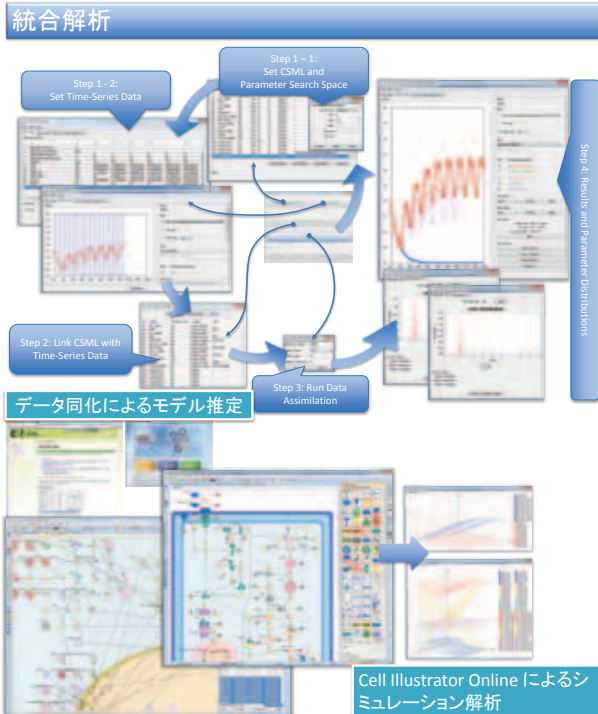


開発責任者: 宮野悟 (東京大学 医科学研究所 ヒトゲノム解析センター)

SBiP 開発メンバー: 長崎 正朗, 関谷 弥生, 齊藤 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

SBiP は、データ解析融合アプリケーション開発チームが開発した技術とプログラムのうち、特に SiGN、LiSDAS を解析パイプラインのコンポーネントとしてユーザが簡単に利用できる高機能 GUI を備えたソフトウェアプラットフォームをユーザに提供ソフトである。京に用意されているジョブスケジューリングシステムと連携し、京にて各解析パイプラインの処理の一部を実行し、その結果を、SBiP の視覚化コンポーネント群を用いて保存できるよう目指している。

ユーザは SBiP に用意されているさまざまな解析コンポーネントを組み合わせてカスタマイズした解析フローを実行可能である。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統合解析プラットフォームとなる予定である。



SBiPでは、細胞・生命プログラムの統合的理解のために新規機能を導入し、GUIと統合することで、実験データとコンピュータ解析をシームレスにつなぎ、細胞システムの解析をユーザが簡単に行うことができるソフトウェアプラットフォームを目指している。機能している部分については、ドキュメンテーション及び導入・利用法の説明 (英文・和文) を整備し、公開している。

<http://sbip.hgc.jp/>

#### 【具体的な開発内容】

新規機能の追加とユーザビリティの向上をはかった。

- SBiPの解析コンポーネントに対して、各種の表示・管理ウィンドウを開発し、実行時におけるユーザビリティを向上させた (パイプラインフロー制御スクリプト設定・実行機能、デフォルト値の preference 機能、実行時エラーレポート機能、入出力パラメータ値の実行時指定機能、等)。ワークフローの描画・可視化ライブラリー JGraph の高機能化に伴い SBiP のコードを大幅に修正し、視覚的にわかりやすいフロー作成を実現した。またコンポーネントのパッケージ化により SBiP のダウンロード時間を短縮できるようにした。SBiP は Java で開発されているが、統計計算とグラフィックスのための言語 R を不自由なく使えるようにソフトウェアを整備した。また、データ解析やネットワーク解析・比較解析などのコンポーネントを新たに追加した。シミュレーションツール Cell Illustrator の改良によりユーザビリティの向上を図った。
- SBiP アプリケーションサイト・ヘルプサイトを作成し、Java Web Start 機能を使って起動する online 版、zip ファイルを展開してローカルで実行する offline 版が利用できるようにした。
- SBiP を用いてデータ解析パイプラインを構築し、データ解析を実施した (Gefitinib 感受性肺がん細胞株と耐性株のネットワーク比較、乳がん細胞株のリガンド応答時系列データを使ったネットワーク推定法の比較、等)。





## D-8: LiSDAS

### 生命体データ同化プログラム

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)

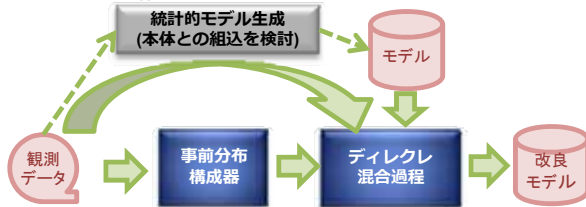
○斎藤正也 吉田亮 中野慎也 長尾大道 (統計数理研究所・データ同化研究開発センター)



## 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効する。

### システム構成

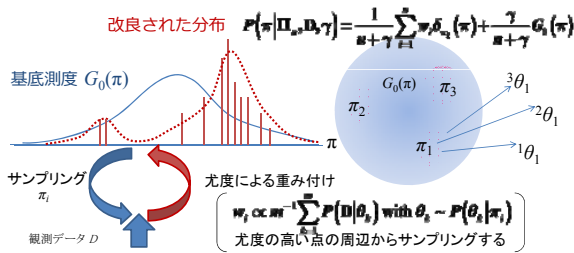


#### 転写制御回路

- $\frac{dx_j}{dt} = -d_j x_j + c_j A(x_{a(j)} | c_j, p_j) R(x_{r(j)} | d_j, q_j) + b_j \sum_{i \in N} x_{i,1} x_{i,2}$
- $A(x | c, p) = x^p / (x^p + c^p)$ ,  $R(x | d, q) = c^q / (x^q + d^q)$
  - モデルパラメタ  $\theta = (x_j(0), d_j, c_j, b_j, c_j, p_j)$  の平均・分散
  - 観測時系列  $D = (y_{i,n})$

## 2. 計算手続き

ディレクレ混合過程を利用したパラメータ推定



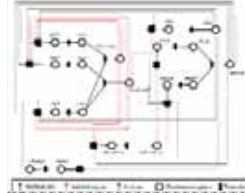
- $D$ : 観測データ
- $\theta, \pi$ : モデルパラメータ (反応速度, 転写因子しきい値)
- “ $a \sim b$ ”: 分布  $b$  からサンプリングして  $a$  に代入

For  $i = 1, \dots, N$  [ $N \sim 10^6$ ]

- For  $j = 1, \dots, m$ :  $\theta^{(j)} \sim p(\cdot | \pi^*)$ . [ $m=128$ ]
- $w^* = \sum_j p(D | \theta^{(j)}) / m$ . [ $p(D | \pi^{(j)}) \approx w^*$  の近似に使う]
- If  $i \leq n$  then [保持する  $\pi$  の個数  $n=128$ ]
- $\pi^{(i)} := \pi^*$ ,  $w^{(i)} := w^*$ .
- $\pi^* \sim G_0(\cdot)$ . [LiSDASが前処理で設計]
- Else
- $\theta^* \sim \sum_j p(D | \theta^{(j)}) \delta(\theta - \theta^{(j)}) / m$ .
- $\pi^* \sim p(\pi | \pi^{(1)}, \dots, \pi^{(n)}, \theta^*)$
- $\propto 1/(\gamma + n) \times \sum_i p(\theta^* | \pi^{(i)}) \delta(\pi - \pi^{(i)})$
- $+ \gamma/(\gamma + n) \times p(\theta^* | \pi) G_0(\pi)$ .
- If  $\exists k \in \{1, \dots, n\}$ .  $p(D | \pi^{(k)}) < p(D | \pi^*)$  then
- $\pi^{(k)} := \pi^*$ ,  $w^{(k)} := w^*$ .

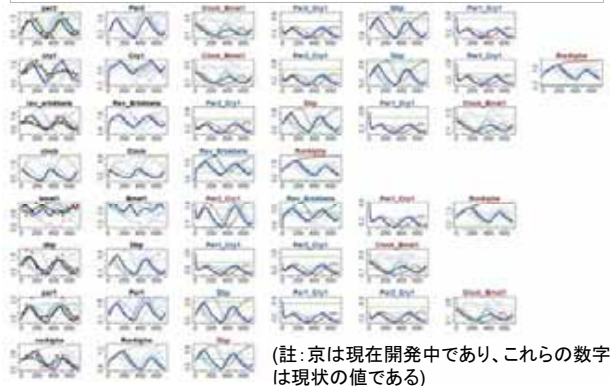
## 3. 性能測定

### 哺乳動物の概日周期転写ネットワーク



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

● 観測データ — 最終的なMAP解 — MAP解の更新履歴

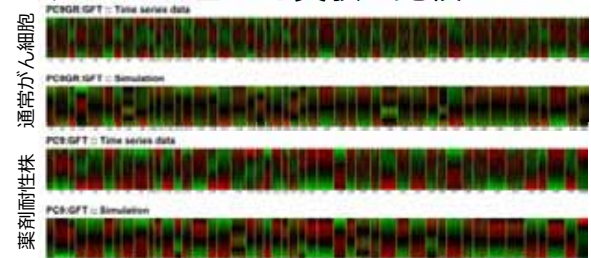


## 4. 適用事例: 肺がん細胞の薬剤応答

### シミュレーション・モデルの構成手続き

- $y_{i,n}$ : 時点  $n$  での遺伝子  $i$  の濃度 ( $i = 1, \dots, 17654$ )
- $x_{j,n}$ : 時点  $n$  での  $j$  番目の潜在変数。発現パターンが似ている遺伝子のクラスタ ( $j = 1, \dots, 46$  (任意))
- 観測行列  $H = (h_{ij})$  の推定
  - $y_{i,n} = h_{ij} x_{j,n} + \varepsilon_i$ ,  $\varepsilon_i \sim N(\cdot | 0, w_i)$ ,  $x_{j,n} \sim N(\cdot | 0, v_j)$
  - 非負対角化条件:  $\exists k. (h_{ik} > 0 \wedge \forall j. (j \neq k \Rightarrow h_{ij} = 0))$
- 潜在変数上のネットワークを決める  $a(j)$ ,  $r(j)$  の推定
  - $\frac{dx_j}{dt} = -d_j x_j + c_j A(x_{a(j)} | c_j, p_j) R(x_{r(j)} | d_j, q_j)$
  - 各  $i$  につき、正負それぞれに、トレンドの相関が最も高いものを、それぞれ  $a(j)$ ,  $r(j)$  とする。

### シミュレーションと実験の比較



#### 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、および本年3月での特別運用での結果です。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

- Yoshida R, et al. (2010) Bayesian experts in exploring reaction kinetics of transcription circuits, *Bioinformatics*, 26(18), i589-595.
- Yoshida R, et al. (2010) Bayesian learning in sparse graphical factor models via variational mean-field annealing, *Journal of Machine Learning Research*, 11:1771-1798.
- Yoshida R, et al. (2008) Bayesian learning of biological pathways on genomic data assimilation, *Bioinformatics*, 24(22):2592-2601.





# D-9: MEGADOCK

## 網羅的タンパク質ドッキング解析プログラムMEGADOCKの開発と応用

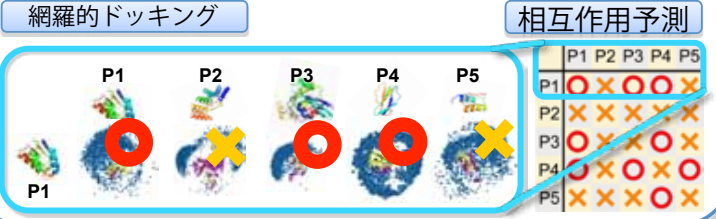
開発責任者: 秋山 泰 (東京工業大学 大学院情報理工学専攻 計算工学専攻)  
松崎由理, 内古閑伸之, 石田貴士, 大上雅史



MEGADOCKは、タンパク質立体構造に基づい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ネットワークを予測するプログラムである。我々は、剛体ドッキングの際に、形状相補性について従来と比べて単純化することで計算量を減じた評価モデル (rPSCモデル) を新規に提案した。さらに静電相互作用も導入した評価関数を実装し、高速フーリエ変換 (FFT) により効率的に計算することを可能とした。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。

標準的なベンチマークデータセットや実問題としてバクテリアの走化性の問題に適用し、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0.43, 0.45を得た。京速コンピュータ「京」において、12,288ノードで良好な並列化効率 (ストロングスケーリング0.96) を得た。今後は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する。

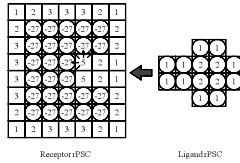
### タンパク質間相互作用予測の流れ



### 新規スコア関数rPSCの実装

real Pairwise Shape Complementarity (rPSC)

$$G_{ij}(l, m, n) = \begin{cases} \phi & \text{if } l \text{ atoms within } [i, l] \text{ fit atom } i, m, n \text{ (open space)} \\ -\phi & \text{(inside of the } l\text{)} \\ 0 & \text{surface accessible surface layer of the } l\text{;} \\ 0 & \text{surface excluding surface layer of the } l\text{;} \\ 0 & \text{inside of the } l\text{;} \\ 0 & \text{(open space).} \end{cases}$$



**MEGADOCK Score**

$$= \text{rPSC} + j \text{ Elec}$$

$$R(l, m, n) = G_R(l, m, n) + H_R(l, m, n)$$

$$L(l, m, n) = G_L(l, m, n) + H_L(l, m, n)$$

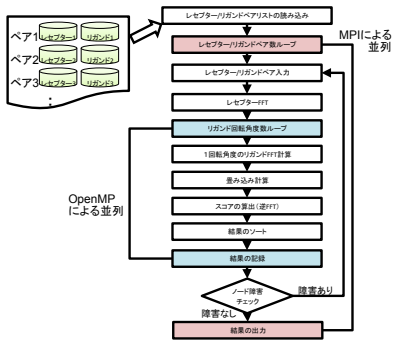
$$S(\alpha, \beta) = \sum_{l, m, n} R(l, m, n) L(l, m, n) + \alpha m + \beta n + j$$

$$S(\alpha, \beta) = \text{FFT}[\text{DFT}[R(L, m, n)] \cdot \text{DFT}[L(L, m, n)]]$$

**ZDOCK**

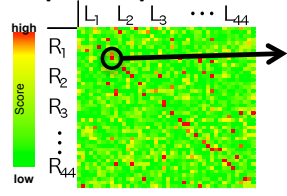
PSC = Re + j Im  
Elec = Re  
Desol = Re + j Im

### ドッキング計算のフローチャート

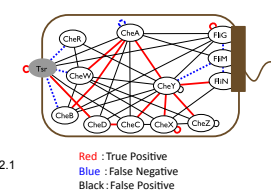


### 評価解析

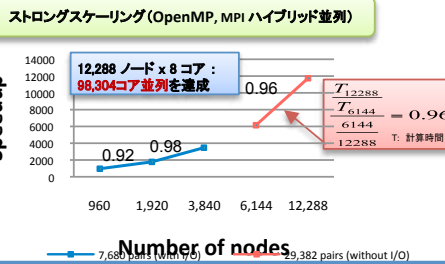
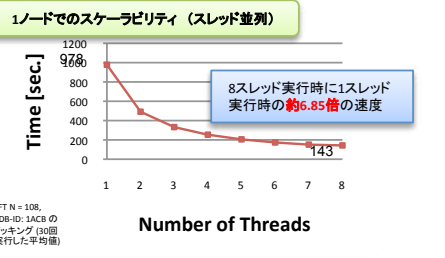
All-to-all calculation of 44 protein pairs (benchmark)



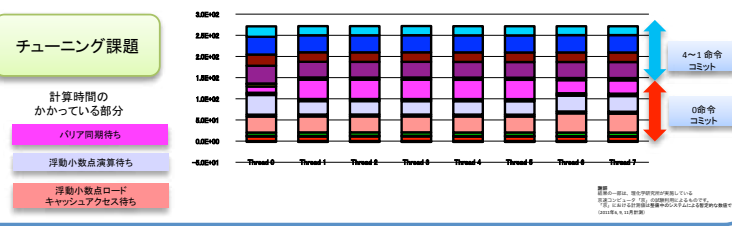
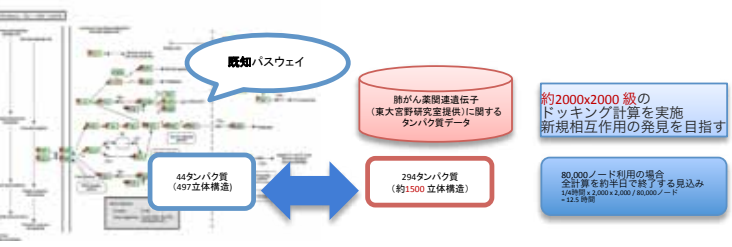
### Bacterial chemotaxis pathway



### 「京」での評価



### 応用課題 肺がん関連 (EGFR系) の遺伝子ネットワークに属するタンパク質群を対象とした網羅的PPI予測





## B-1: NEST

# Towards brain simulations

Markus Diesmann (Computational and Systems Neuroscience, INM-6, Research Center Jülich, Germany)  
RIKEN Brain Science Institute, Wako City, Saitama, Japan)

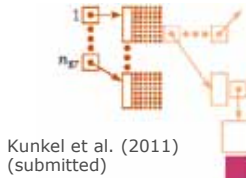
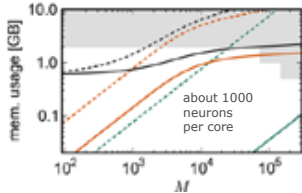


NEST is a simulator for large biological neuronal networks. Small systems of differential equations describe individual neurons which are coupled to form networks of natural size and complexity. The code runs on a variety of computer architectures: from single- and multi-core desktop computers to large clusters and HPC facilities such as K and JUGENE. NEST is subject to continuous development driven by the needs of the neuroscience community. These efforts are coordinated by the NEST Initiative ([www.nest-initiative.org](http://www.nest-initiative.org)), an international collaboration of several research institutes.

**In the course of the Next-Generation Supercomputing Project we adapted NEST to systems with more than 10,000 cores enabling the simulation of full size networks at cellular resolution:**

### Model of memory and optimizations

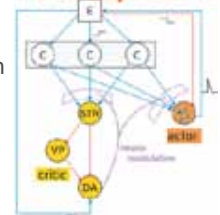
- new implementation of storage for neurons (solid blue) and synapses (solid orange) employs sparse representation
- reduced memory consumption (dashed: old impl.) enables efficient distribution on  $M \geq 10000$  cores



Kunkel et al. (2011) (submitted)

### Spike-timing and neuromodulation dependent plasticity

- NEST can simulate models of synaptic plasticity depending on
  - activity of pre- and post-synaptic neurons
  - presence of neuromodulators

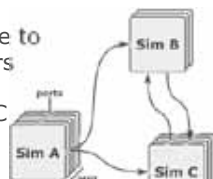


Morrison A et al. (2007) *Neural Comput.* 19, 1437-1467  
Potjans W et al. (2010) *Front Comput Neurosci* 4:141  
Potjans W et al. (2011) *PLoS CB* 7(5): e1001133

### NEST implements the MUSIC interface

NEST supports the MUSIC interface to Communicate with other simulators

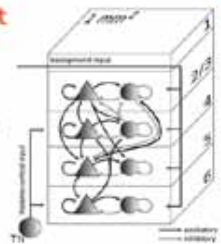
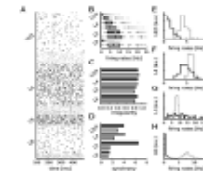
- distributed simulators A, B, and C
- exchange of data during runtime
- enables multi-simulator models



Djurfeldt M et al. (2010) *Neuroinformatics* 8(1):43-60

### Model of the local microcircuit

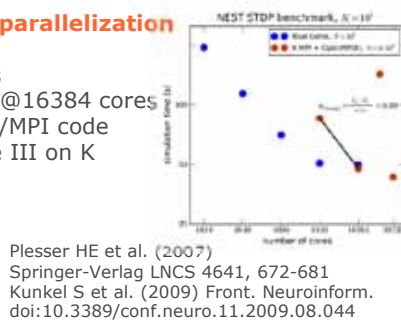
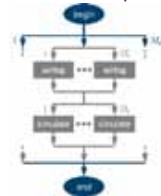
- 1 mm<sup>2</sup> cortex with layer- and cell type specific connectivity
- reproduces layer-dependent rate (A, B) and irregularity (C), and propagation



Potjans et al. (2011) (submitted)  
Potjans et al. (2011) arXiv:1106.5678v1  
Wagatsuma et al. (2011) *Front. Comput. Neurosci.* 5:31  
Lindén et al (2011) *Neuron* (in press)

### High degree of parallelization

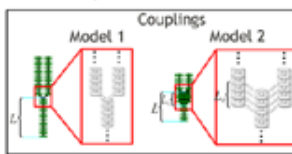
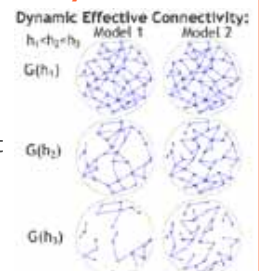
- $3 \times 10^6$  neurons
- speedup  $\alpha=0.9$  @ 16384 cores
- hybrid OpenMP/MPI code
- achieved phase III on K



Plesser HE et al. (2007) Springer-Verlag LNCS 4641, 672-681  
Kunkel S et al. (2009) *Front. Neuroinform.* doi:10.3389/conf.neuro.11.2009.08.044

### Compositionality in networks of synfire chains

- high-capacity embedding of synfire chains
- propagation probability  $p(h)$  depends on # waves  $h$
- activity settles at critical point



Trengove (2011) (submitted)

### Roadmap (22 Feb. 2010) and outlook

Step	Description	Network type	Nrms.	Syn.	Status Nov 2011	Expected results
1	reduced primate visual cortex (32 areas $\times$ 1 local microcircuit)	random at micro and macro scale	$3 \times 10^6$	$3 \times 10^{10}$	technically achieved on K on 28.Aug 2011	- characterize stationary state: rate, irregularity - comparison of low-frequency EEG and LFP to experiments
2	like 1, but respecting relative sizes of areas (32) and distance dependent connectivity	spatially organized at micro scale	$10^8$	$10^{12}$	- technically possible with current code - tests in progress	capability to simulate full scale model of a functional unit
3	human brain model (approx. 100 areas)	spatially organized	$< 10^{11}$	$< 10^{15}$		assessment of ability to simulate full scale model



# B-2: Cortical Microcircuit Developed on NEST

## 大規模並列計算で探る大脳局所神経回路の機能



開発責任者: 深井朋樹 (理化学研究所 理化学研究所 次世代計算科学研究開発プログラム 脳神経系研究チーム, 理研 脳センター、シニアチームリーダー)

・ 背景

脳機能を生み出す神経回路メカニズムを解明することが、脳研究の世界的潮流になって来つつある。そのような研究の新しい手段として、大規模並列シミュレーションが、強力な手法を与える可能性に注目されている。

例: Human Brain Project (EUの13研究機関が連携)

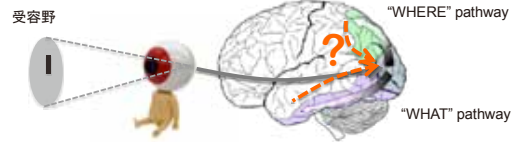
・ 目的

大規模神経回路の並列シミュレーション技術を開発し、それを用いて霊長類の脳で特に発達している大脳皮質6層局所回路のモデルを構築し、その計算原理に迫る。

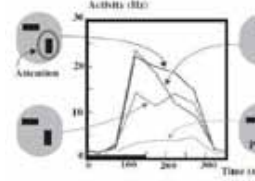
アプローチ

最新の解剖学的研究・電気生理学実験の知見を取り入れて、局所神経回路の構造を決定する。形態構造を持たないニューロンモデルを採用し、シナプスについては十分な数と可塑性を確保する。構築したモデルを脳機能の説明に適用し、科学的有効性を検証する。

— 視覚注意のシミュレーション —



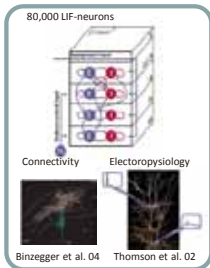
- 視覚野 (V2/V4) の細胞には、視野内の担当領域 (受容野) がある
- 各細胞ごとに好きな刺激の方向がある
- 注意を向けることで細胞の応答が変化する



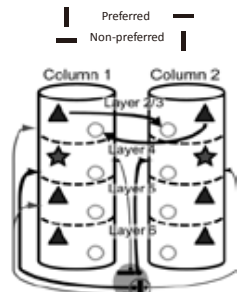
(Reynolds et al., 1999)

計算事例1

● 注意による視覚神経細胞の応答変化のシミュレーション

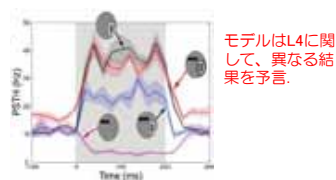


(Potjans & Diesmann, 2011)

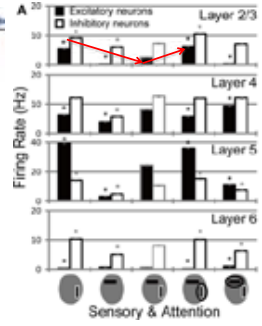


COMPUTATIONAL NEUROSCIENCE

Layer-dependent attentional processing by top-down signals in a visual cortical microcircuit model

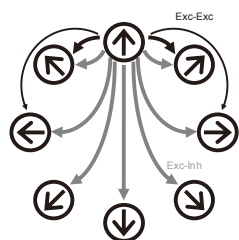


回路モデル (L2/3, L5) は、実験で見られる注意の効果をよく再現する。  
モデルはL4に関して、異なる結果を予言。

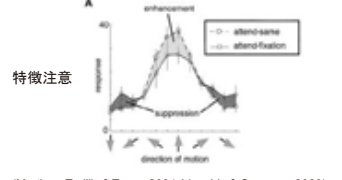


計算事例2

— 空間注意と特徴注意のシミュレーション —  
Wagatsuma, Potjans, Sakai, Diesmann, Fukai

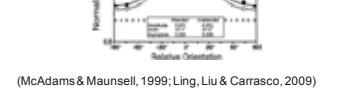


実験



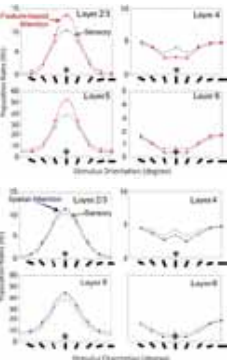
(Martinez-Trujillo & Treue, 2004; Ling, Liu & Carrasco, 2009)

空間注意



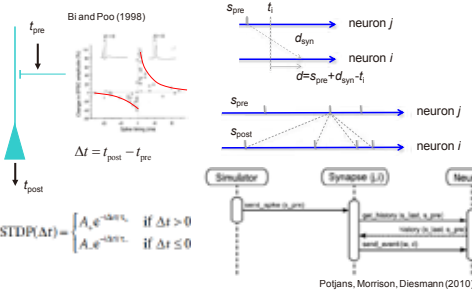
(McAdams & Maunsell, 1999; Ling, Liu & Carrasco, 2009)

シミュレーション



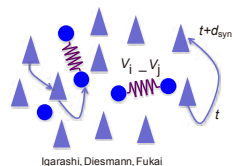
### 生物学的付加価値の付与

スパイク時間依存の可塑性 (STDP) の実装



Potjans, Morrison, Diesmann (2010)

抑制細胞間の電氣的シナプス



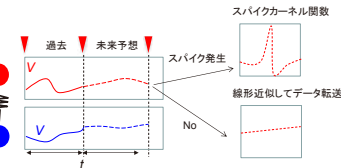
Igarashi, Diesmann, Fukai

NEST = event-driven simulator

化学的シナプスと電氣シナプスが共存  
→ 効率的なシミュレーションが困難

通信負荷を抑えながら計算効率を確保したい

電氣シナプスから流入した電流の過去 (~1ms) の平均値を用い、未来の電位変化を予想



- 共同研究者 五十嵐 潤 (次世代計算科学/ 電氣シナプスおよびNEST実装) NEST Markas Diesmann  
半田 高史 (次世代計算科学/ 局所回路実験データ収集) Tobias C Potjans  
我妻 伸彦 (学振PD/ 視覚注意の局所回路モデル) Moritz Helias  
Gen Masumoto

謝辞

- 京は現在開発中であり、スライド中の数字は現状の値である。
- 京上の性能に関しては京コンピュータの試験利用、および特別運用による。開発および運用にあたる関係各位に感謝申し上げます。とくに京上の実装については、ISLIM高度化チームの協力無くしては困難であった。感謝申し上げます。





## B-3: VMS

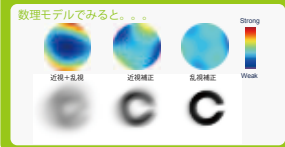
### 全視覚系モデルによる視覚情報処理の解析 (視覚系シミュレーションのための共有プラットフォーム)

開発責任者: 臼井支朗 (理化学研究所脳神経系研究開発チーム)



#### 眼球(眼光学)

カメラのレンズ・フォーカス機能に相当。波長感度特性、波面収差、瞳孔の特性などを考慮し外界の網膜像を計算



#### 眼球運動制御

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

外界像

眼光学

網膜

網膜像

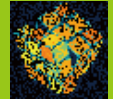
眼球運動制御

視覚系 = カメラ

視覚皮質

#### 網膜

カメラのフィルムに相当。外界の網膜像が、ここで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



入力イメージ    網膜細胞の応答

スパイクとは？  
周期的に上昇・下降する電気信号の変化でとがった波形になる。脳内では、これをもとに情報のやり取りが行われている

#### 視覚皮質

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

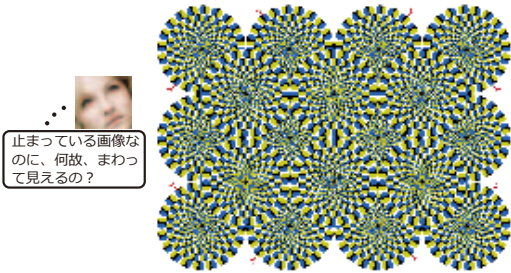
#### 全視覚系モデルの構築

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。

具体的に何をやるの？

眼球運動、眼光学、網膜ならびに視覚皮質からなる全視覚系を詳細な大規模モデルとして構築し、種々の視覚現象をシミュレーションにより解析する

→ 各サブシステムレベル (Micro) と全視覚系レベル (Macro) の観点から視覚現象における視覚系の役割を解明する



#### サブモデル構築と PLATO による統合

##### Sub-models (parts)

- 詳細サブモデルの構築と並列化
- 共有 IO フォーマットによるモデル接続
- Agent による統合モデルの実行・管理

We should share

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

Large-scale model

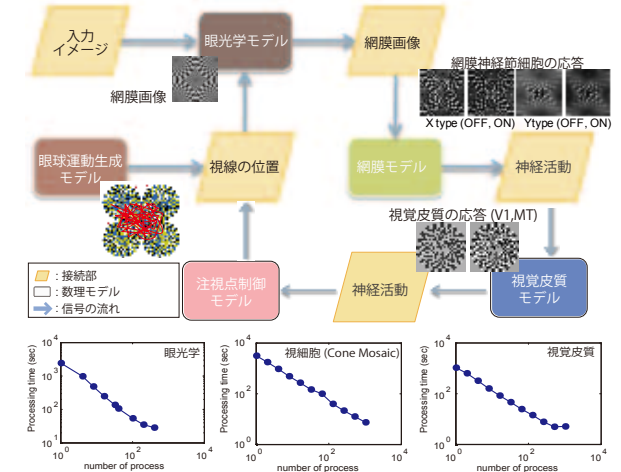


integrate

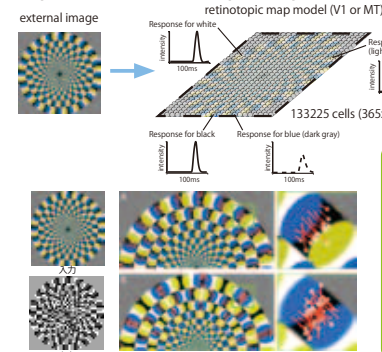


#### 視覚系のモデル

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、解析を行い、視覚機能の解明を進める。



#### 例) 視覚皮質のモデル (Demo)



大脳の神経活動を数理モデルで再現してみると、見えている動きと同じ方向に、神経活動も推移している

→ コントラストに対する視覚皮質細胞応答に錯視知覚に対応する情報が表現されていることを示唆

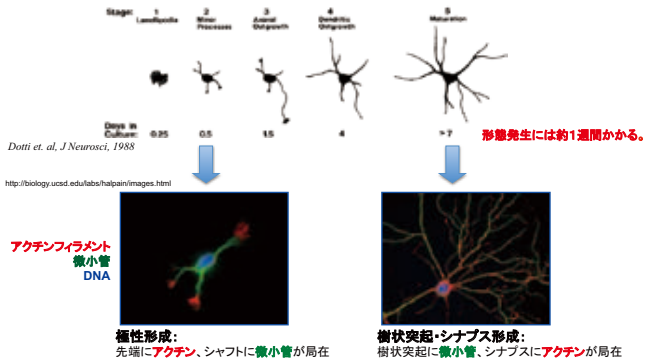


# B-4: NeuroMorphoKit 神経細胞形態シミュレーションキット

開発責任者: 行縄直人 (京都大学 大学院情報学研究所)  
 本田直樹 (京都大学 大学院情報学研究所)  
 石井 信 (京都大学 大学院情報学研究所、理化学研究所 次世代計算科学研究開発プログラム)



## 神経細胞の形態と細胞骨格



### 生物学的なゴール

- 形態形成における細胞骨格の動的な再構成の理解
- 以下の因子に制御されるニューロンの形態変化のダイナミクスのシミュレーション
  - 細胞骨格 (アクチンフィラメント および 微小管)
  - 細胞膜のダイナミクス
  - シグナル伝達

### ソフトウェア開発のゴール

- 個々のフィラメントを基本とした
- 細胞内シグナル伝達・細胞骨格・細胞膜を統合した
- 全細胞スケール (数十万~数百万本のアクチンフィラメントを含む) の大規模な形態変化を記述する

モデルおよびシミュレータを開発する

### 基礎方程式

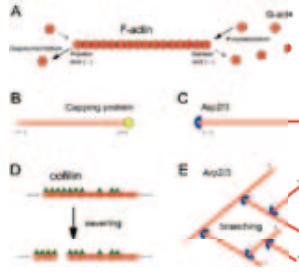
- 反応拡散方程式
- 細胞骨格・膜に関して新たに方程式を構築

### 連続シミュレーション

- 細胞内シグナル伝達系 (反応拡散系)
- 細胞骨格系 (アクチンフィラメント)
- 細胞膜

## アクチンフィラメント

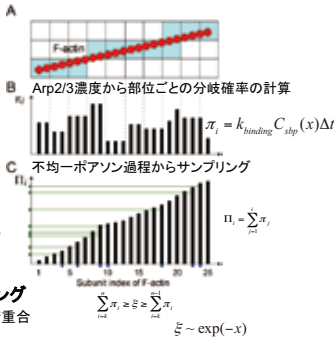
### F-actinのキネティクス



### キネティクスイベントは確率的にサンプリング

- アクチンモノマー分子の濃度に比例した確率で重合
- Arp2/3の濃度に比例した確率で分岐

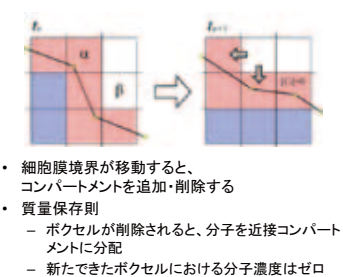
### 分岐計算の効率化



## 細胞内シグナル伝達

$$\frac{dC_i}{dt} = D_i \frac{\partial^2 C_i}{\partial x^2} + f_{rxn}(C_j) + f_{actin}(C_j, F)$$

### 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

## 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

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

体積を一定に保つ

$$E_{volume} = \alpha_{volume} (V - V_0)^2$$

表面積を一定に保つ

$$E_{surface} = \alpha_{surface} (S - S_0)^2$$

細胞膜を滑らかにする

$$E_{bending} = \alpha_{bending} \int_{\Omega_{cell}} \frac{1}{R(s)^2} ds$$

R: 曲率半径

アクチンフィラメントを真っ直ぐにする

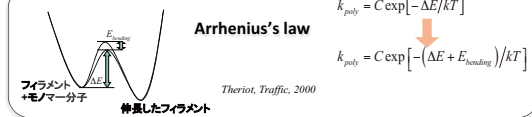
$$E_{F-actin} = \sum_i \frac{4\lambda k_B T}{l_i^3 \tan^2 \theta_i} \Delta x_i^2$$

Mogilner, et al., Biophys J, 1996

膜ポリゴンの頂点の位置は時間  $\Delta t$  おきに以下の最急降下法により、Eを最小化するように最適化する

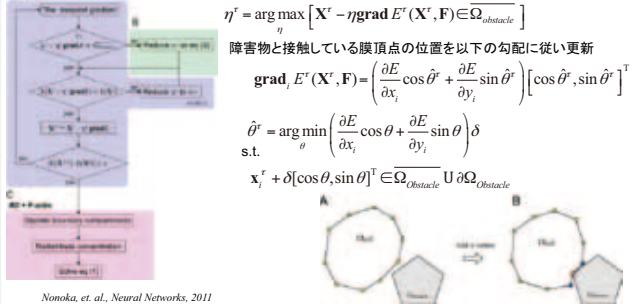
$$x_i^{\tau+1} = x_i^\tau - \eta^\tau \text{grad}_{x_i} E^\tau(x_i, F)$$

### 重合率の計算



## 全体のアルゴリズム

### 細胞と障害物の重複回避のためのヒューリスティック

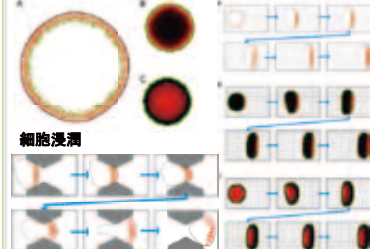


## ソフトウェアの現状と今後

1. 100万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
2. 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)
3. 大規模並列環境 (京 576 並列、RICC 4096 並列) での動作を実現 (MPI, OpenMP/ハイブリッド)
4. 膜電位シミュレーション機能を持つ神経細胞シミュレータA-Cell (京都大学・市川一寿教授による開発) との連携動作のためのデータ交換フォーマット (NetCDF-neuron) を策定
5. 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
6. パネネットワークモデルによるアクチンフィラメントの曲げ、結束のモデルへの導入、三次元化を実施中、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

#### ラメリポディアの自己組織化 細胞移動



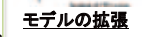
#### 京での性能評価

$\rho = 93.753\%$   
(flat MPI)

#### フィラメント計算の効率化



#### モデルの拡張





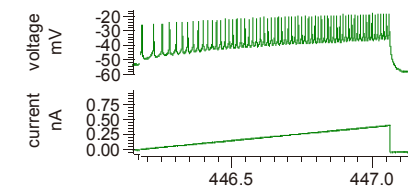
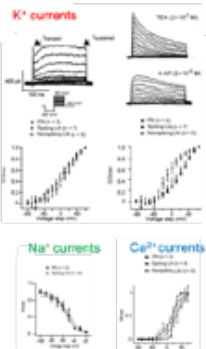
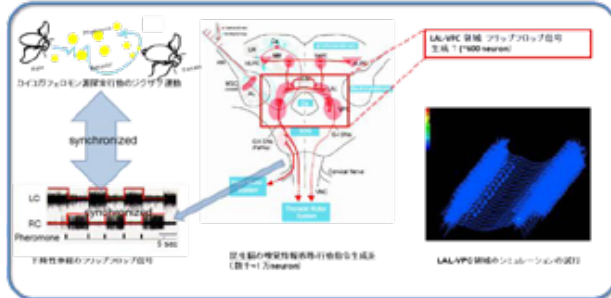
# B-5: IOSSIM

## 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)



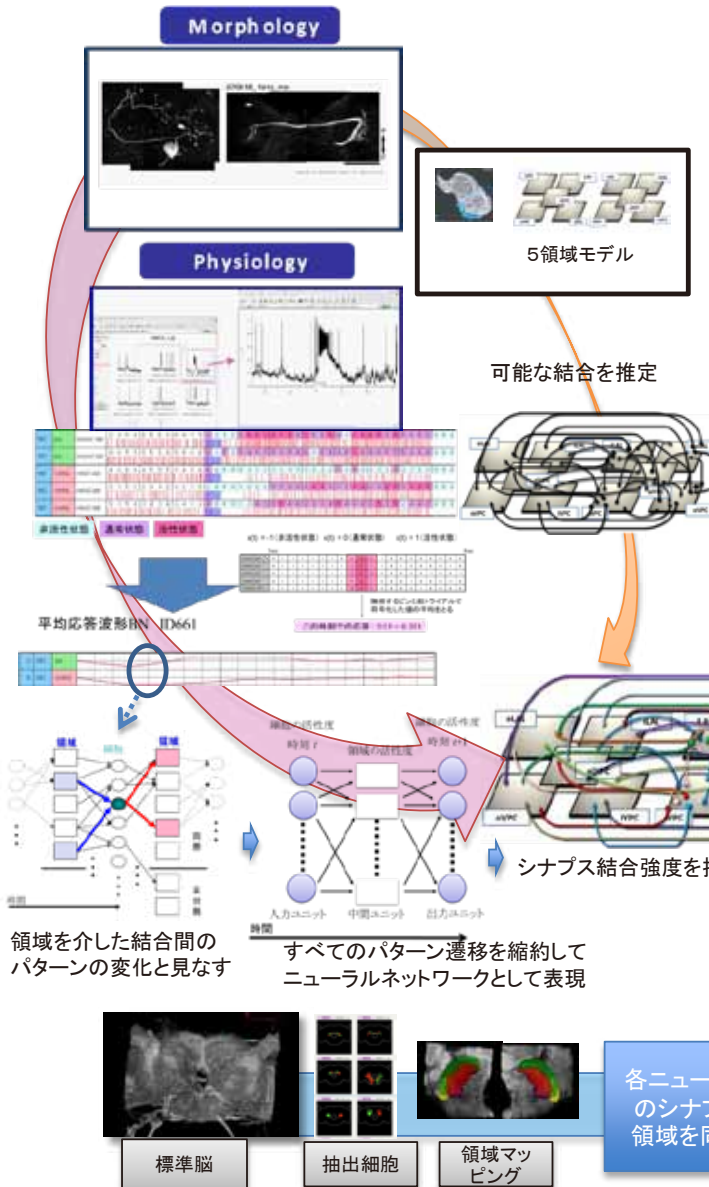
目標: 昆虫脳の嗅覚情報処理から行動信号生成までの神経ネットワークを実験と比較できる詳細モデルを使ってシミュレーションする



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 $Ca^{2+}$ 電流が存在する

触角葉細胞体からの voltage clamp



### Parameter estimator using NEURON

- Processor: 256 core
- Number of genes: 2048
- Number of generations: 200
- Simulation time step: 0.025 msec
- Simulation time range: 0 - 300 msec
- Estimation range of  $C_m$ : 0.5 - 3  $\mu\text{F}/\text{cm}^2$
- Estimation range of  $G_{Na}$ : 0.000 - 1.000  $\text{S}/\text{cm}^2$
- Estimation range of  $G_{Ca}$ : 0.0000 - 0.1000  $\text{S}/\text{cm}^2$
- Estimation range of  $G_{K}$ : 0.00000 - 0.01000  $\text{S}/\text{cm}^2$

Estimator by Genetic Algorithm using NEURON is implemented, fitting for multicompartment model is ongoing.

## Simulation

結果は当日に発表します

※本研究は、京速コンピュータ京の試験利用を用いている。





B-5: IOSSIM



# 昆虫嗅覚系全脳シミュレータ

開発責任者: 神崎亮平 (東京大学先端科学技術研究センター 脳神経研究開発チーム)

背景

ヒト (10<sup>9</sup>個の神経細胞)

昆虫 (10<sup>6</sup>個の神経細胞)

脳をつくる神経細胞(ニューロン)の形やはたらかは昆虫・哺乳類をとおして共通だが、神経系の規模は小さい。  
異種感覚統合 連合学習 定位行動など環境に適応するため充分な複雑な行動を行う。  
そのため 昆虫脳は感覚から行動までの脳の情報処理をシミュレーションするためのモデル系をして使われている。

シミュレーションの構築

神経細胞データベース: BoND

細胞形態抽出: NEWRITE

カイクガ標準脳: BrainComposer

シナプス結合可能性の推定

シナプス結合強度推定

神経の生理・形態記録

カイクガのニューロンデータベース (BoND) とマルチスケール計測

Gene, Neuron, Networks, Behavior, Robot

形態抽出ソフト (KNEWRITE)

シミュレーションによる変換精度の評価

形態的な変換精度の評価

標準脳構築ソフト (BrainComposer)

標準脳構築ソフト (BrainComposer)

基礎方程式

マルチコンパートメント・等価回路モデル

マルチコンパートメントモデル等価回路方程式

$$C_m \frac{dV_n}{dt} = \sum_i g_{in}(V - v_i) + I_{n-1,n} - I_{n+1,n}$$

電位依存性イオンチャネルのH-H形式の記述

$$\frac{dg_i}{dt} = m^{\infty} n^{\infty} y_i (x_i, y_i = 1 \sim 4)$$

$$\frac{dgate}{dt} = \alpha_i(v)(1 - gate) - \beta_i(v)gate$$

$m$  or  $n \in gate$   
 $\alpha$   $\beta$  はシグモイド関数

シミュレーション前のデータの流れ

免疫染色 → 標準脳 → 抽出細胞のマッピング → オーバーラッピングの推定

NEURON K+

ベンチマーク用回路の構築

5000個の神経細胞 (各5000コンパートメント) を0.025msec単位でリアルタイムに計算するには、約300 TFLOPS必要

現在は最大22.5 TFLOPS (8192ノード時) を達成

シミュレーションを用いた神経パラメータ推定

実験値GAPによるパラメータ推定

シミュレーションによる電流クランプ応答を用いた、推定ソフトウェアの評価

謝辞

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。本ソフトウェアの開発に際しては、京速コンピュータ京の試験利用、理化学研究所情報基盤センターのRIOC、東京大学情報基盤センターのT2Kを用いた。京上の上装における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 ハウト・S・周一  
加沢知雄 高嶋暁 櫻井健志 岩月知香  
佐藤隆平 宮本大輔 森友亮 田淵理史  
藤原輝史 峰岸諒 佐藤秀仁  
津野豊利 石井愛弓 新田倫子  
西川響子 小林嘉太 山岸嘉彦 小杉辰弘  
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)

## ABSTRACT

We report all-atom simulations of molecular crowding, an early result from the "K computer," which is a 10-PFLOPS supercomputer under construction in Japan. The K compute nodes consist of 8-core SPARC64 Vllifx processors attached to dedicated routers in a six-dimensional mesh/torus. Using 27,648 such nodes, we observed 1.3 PFLOPS (37.5% of peak) during a 180 million-atom simulation. We discuss the performance and scaling of molecular dynamics simulations on "K," as well as macromolecular crowding at atomic resolution.

Living cells are crowded because macromolecules comprise ~30% of their molecular weight. Recently, the effects of crowded cellular environments on protein stability have been revealed through in-cell NMR spectroscopy. To study these effects, we performed all-atom classical molecular dynamics simulations of two systems: target proteins in solvent and target proteins in an environment of molecular crowders that mimic the conditions of a living cell.



Figure 1. Section of the K computer. Visible are four computer racks with 384 nodes.

Table 1. FLOPS efficiency on the kernel loop

Model	<i>In vitro</i>
Number of atoms	418,707
Cutoff radius	28 Å + 2 Å margin
Number of pairs per atom	8,835 (+ 2,031 in margin)
FLOP counts for 1,000 steps	469,564,792MFLOP
Kernel FLOPS per thread	7,088 MFLOPS (44.3%)
SIMD%	48.76%
FLOP counts per atom pair	103.2 FLOP

Table 4. Conditions of the simulation for the peak performance

Model	432 <i>in vitro</i>
Number of atoms	180,881,424
Cutoff radius	28 Å + 2 Å margin
Number of pairs per atom	8,835 (+ 2,031 in margin)
FLOP counts for 1,000 steps	202,851,990,144 MFLOP
Topology of computation node	24 × 24 × 48
Number of nodes	27,648
Number of cores	221,184
Theoretical peak performance	3,539 PFLOPS
Calculation time for 1,000 step	154.29 sec
Sustained performance	<b>1.315 PFLOPS</b>
Efficiency	0.3716

Table 2. Time spent in blocks in millisecond per step for weak scaling (418,707 atom/64 nodes)

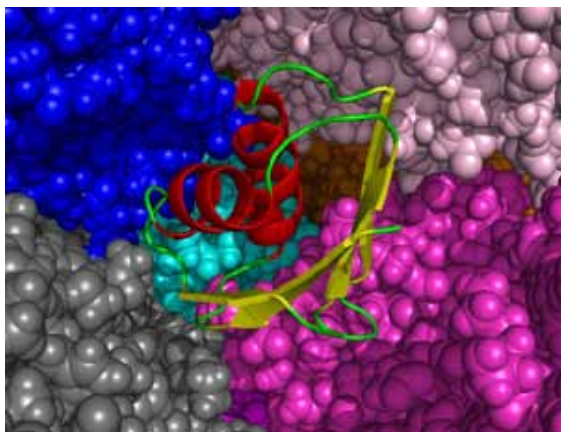
	64	256	512	4096	13824	27648
# of nodes $N_{node}$	64	256	512	4096	13824	27648
# of atoms $N_{atom}$	418,707	1,674,828	3,349,656	26,797,248	90,440,712	180,881,424
Total $T(N_{node})$	152.84	153.66	153.34	153.88	153.92	154.29
Force calculation	128.21	128.21	128.36	128.34	128.06	128.15
Communication	20.19	20.99	20.50	20.87	21.21	21.55
Other	4.45	4.47	4.48	4.67	4.64	4.60

Table 3. Time spent in blocks in millisecond per step for strong scaling (1,674,828 atoms)

# of nodes	256	512	2048	4096	16384
# of import cells	936	768	504	446	342
# of communication nodes	26	44	124	174	342
Total	153.66	84.63	28.57	18.98	13.58
Force calculation	128.21	61.78	15.42	8.33	2.68
Communication	20.99	19.96	11.75	9.55	9.96
Other	4.47	2.89	1.40	1.10	0.94

## Strong scaling

Figure 6. Vicinal area of TTHA for the *in vivo* system (viv). The TTHA molecule is drawn using the ribbon model, and ovalbumin molecules are drawn using the space-filled model. Each ovalbumin molecule is colored differently.



## Weak scaling

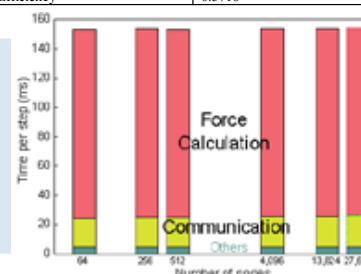


Figure 2. Weak scaling for 6,542 atoms/node. Constant wall clock times/step show the perfect scaling. Ratios of force calculation/communication/others are also constant, and force calculation is dominant.

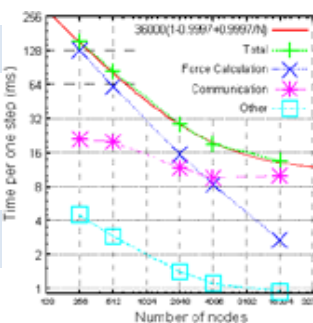


Figure 3. Strong scaling for the system of 1,674,828 atoms.

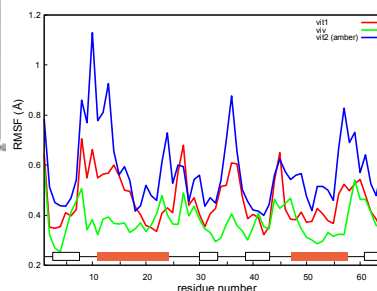


Figure 5. Conformational fluctuations of TTHA for *in vivo* (viv) and *in vitro* systems (vit1, vit2). The green, red, and blue curves indicate the RMSFs of TTHA for *in vivo* and *in vitro* systems, respectively. The abscissa axis is the residue number of TTHA and the ordinate axis is the RMSF value (in Angstrom). Red and white boxes indicate alpha-helices and beta-sheets, respectively.

## Scientific results

### ACKNOWLEDGMENTS

This work was partially supported by the Integrated Simulation of Living Matter Project, commissioned by the Ministry of Education, Culture, Sports, Science and Technology, Japan. A part of the calculations in this work were done for the performance evaluation process on the development of the Next-Generation Supercomputer. We thank the Next-Generation Supercomputer R&D Center, especially Kazuo Minami and Masaaki Terai (in the Application Development Team) for their outstanding support. We appreciate Mikio Hondou and Hikaru Inoue, from Fujitsu Co. Ltd., for their optimization support. We also thank the RIKEN Integrated Cluster of Clusters (RICC) for the computer resources used for our code development and calculations. We gratefully acknowledge the exceptional support provided by the late Takayuki Shigetani at RIKEN Advanced Center for Computing and Communication (ACCC). We also thank Toru Takinaka at NEC Informatec Systems, Ltd. for the supporting code development.



## H-2: LSV 大規模データ解析ソフトウェアLSVの開発

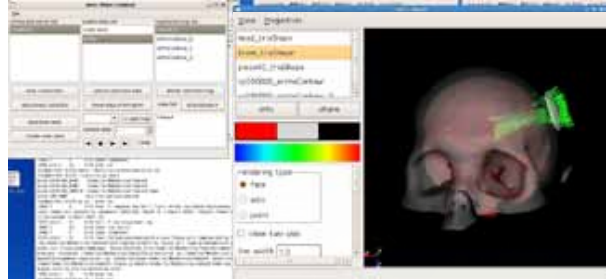


開発責任者: 小野謙二 (生命体基盤ソフトウェア開発・高度化チーム)

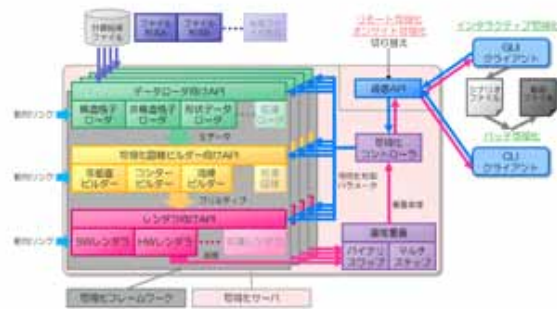
※系は現在開発中であり、性能や動作状況は現状の値です。  
※京での実行に関しては、京速コンピュータ京の試験利用での結果です。  
※GPGPUクラスタ及びPCクラスタは理化学研究所情報基盤センターのRICCを利用しています。

### LSV “Large-Scale data Visualization software”

- 大規模データの並列可視化
- 共通UIによる統一化されたりリモート/ローカル可視化環境
- インタラクティブ/バッチ可視化
- ソフトウェア/ハードウェアレンダリング
- プラグインによる機能拡張
- XMLによるスクリプト処理

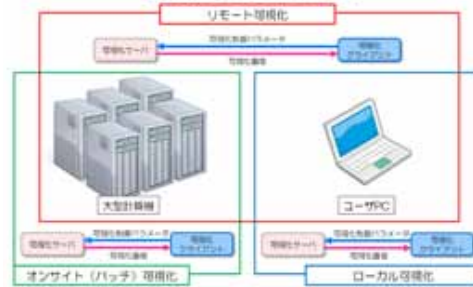


### システム概念図



### オペレーションモード

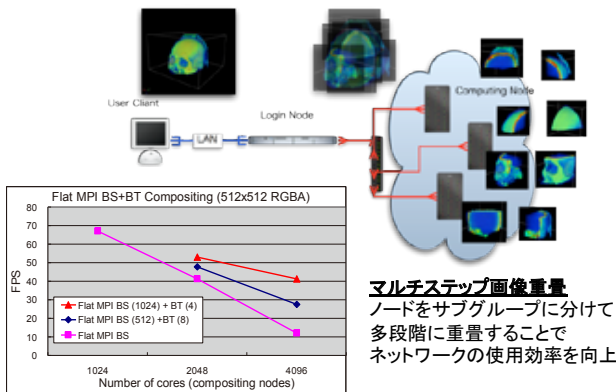
大規模データの並列可視化  
インタラクティブ



定義されたシナリオによる動画作成  
バッチ

小規模データの可視化  
インタラクティブ・バッチ

### 画像重畳



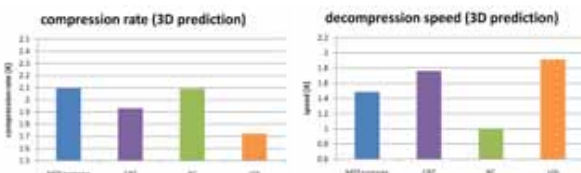
### RICCでの並列可視化テスト

- データセット (XXX x YYY x ZZZ, 100GB)
- 100ノードを使用したり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

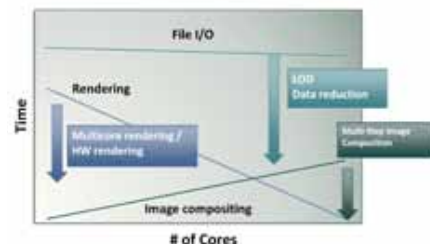
#### 高速データ圧縮

- 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 Isenbarg, 2006]
  - zlib: 汎用圧縮ライブラリ [Jean-loup Gailly and Mark Adler]



#### インテリジェントデータリダクション

- 特徴を保存した非可逆圧縮
- Level-Of-Detail
- Region-Of-Interest
- 特徴抽出
- データベース・メタデータの利用







# H-3: SPHERE

## Simulation Framework for Large scale Coupled Analysis

開発責任者: 小野謙二 (生命体基盤ソフトウェア開発・高度化チーム)

※系は現在開発中であり、性能や動作状況は現状の値です。

※京での実行に関しては、京速コンピュータ京の試験利用での結果です。

※GPGPUクラスタ及びPCクラスタは理化学研究所情報基盤センターのRICCを利用しています。



Important issues of next-generation simulation are:

- Multi-physics / Multi-scale problem
- Coupled problem

These problems require long-time analysis with small-time increment, extremely large data handling, data Interpolation through coupling surface, etc.

### SPHERE

#### Skeleton for **PH**ysics and **Eng**ineering **RE**search

An solver framework for operating CFD codes with an uniform user interface.

- Configuration parser (XML)
- File I/O
- Dynamic memory allocation
- Parallelization
- Coupling function
- Flow control

Base system

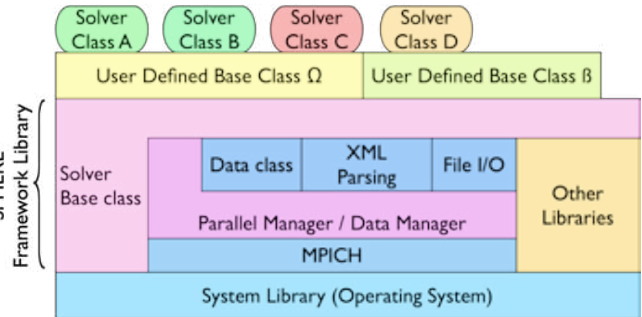


Fig.1 SHPERE system

- Unstructured data structure
- Additional information (group data)
- Pre/post processing (i.e. partitioner)
- Mathematical operations

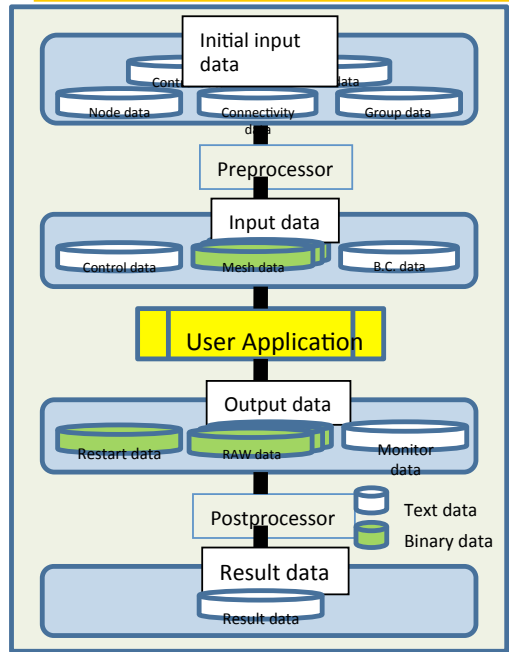


Fig.2 Concept of I/O system

### Remarks of new middleware

#### Fast I/O system for large data

- Separate mesh data for files by each component due to preprocessor convenience
- Sequential numbering for minimizing data search

#### Data structure of FEM

Element by Element style coding is employed.

- Most popular method
- Application developers easily deal with the data structure
- Definitely unified data structure has an great advantage for vectorization and optimization.

#### For life science applications

- Mesh, UNS, octree, point set are supported.

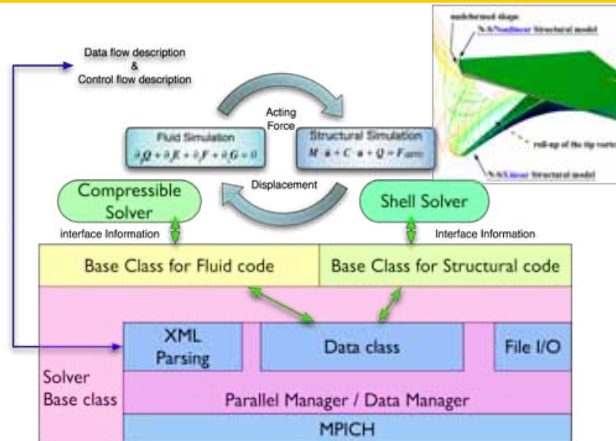


Fig.3 An example of coupled analysis on the middleware

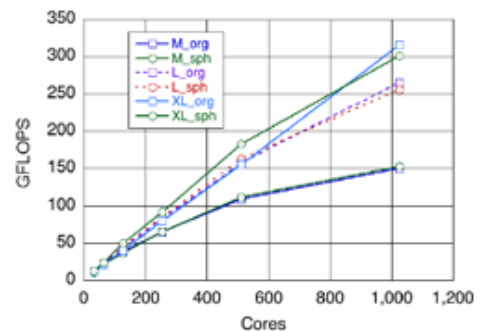


Fig.4 Scalability of SPHERE on RSCC



# H-4: VLSVL

## 合成可能な化学構造および反応スキームからなる大規模仮想ライブラリ



開発責任者: 船津公人 (東京大学 大学院工学系研究科)

### 1. 背景

- 創薬の第一歩は有望なリード化合物の発見で、ハイスループットスクリーニング (HTS) が重要な役割を果たしてきた。しかし、HTSで有望なリード化合物が見いだせないケースもあり、その探索効率とコストが問題となっている。
- 現在は、コンピュータ支援バーチャルスクリーニングによるリード化合物探索が推進されており、そのための、医薬品になりうる質の高い仮想化合物群からなる大規模なバーチャルライブラリの構築が望まれている。
- バーチャルスクリーニングにおいて、高スコアのリード化合物群を得られるかどうかは、バーチャルライブラリを構成する仮想化合物群の質や多様性、規模などに左右され、医薬品開発の成否に大きな影響を与える。総じて、**バーチャルライブラリの質と量が重要な意味を持つ**といえる。
- 一方、スクリーニング結果、高スコアのリード化合物群を絞り込んでも、文字通り**バーチャルであるが故に、対象化合物を入手するための合成検討に多大なコストを要する。この問題に対する解決策が必要**とされている。

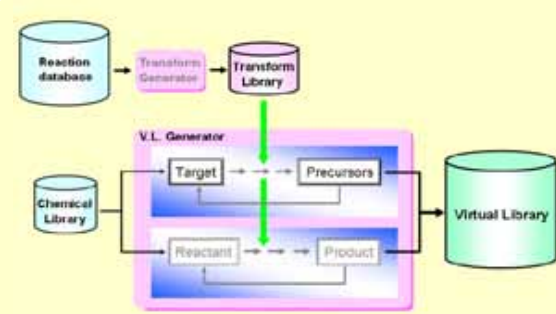
### 2. 目的

- 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
- 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

- 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。
- 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。
- 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。
- 入力構造に対して、反応予測方向にもtransformを適用し、反応生成物も提案させ、これを繰り返して、合成ルート樹を形成する化合物ライブラリとする。

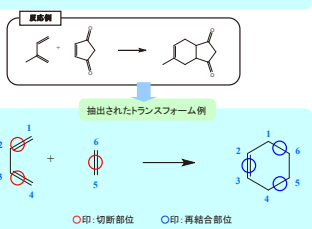
### 4. システムの構成



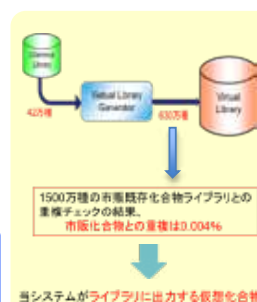
大規模な反応データベースからtransformを自動抽出することによって、多様な反応事例に適用できるtransformを多量に蓄積し、その適用によって誘導される仮想化合物群の多様性を確保する。

### 5. transform

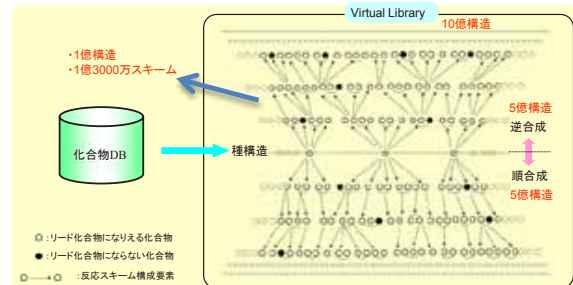
- transformには、反応スキームを構成する各化合物の反応中心情報が収納される。具体的には、反応中心の部分構造とその周囲の構造特徴情報、結合の切断部位、結合次数の変化した部位とその次数、新たに生成した単結合の部位、脱離した置換基の種類と部位、加わった置換基の種類とその位置など。
- システムに入力された化学構造に、各transformに記述された部分構造が含まれる場合は、該当部位をtransform構造に置き換えた反応スキームを提案する。



### 6. 発生化合物群の新規性



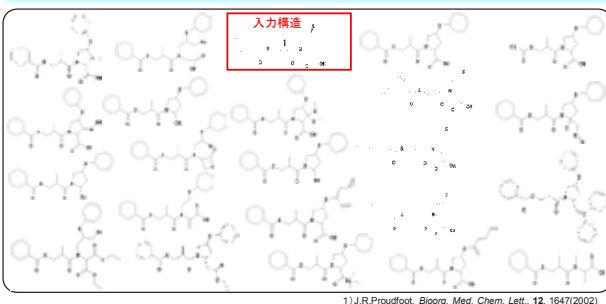
### 7. バーチャルライブラリの構成



- システムに化合物DBから種構造が入力されると、逆合成提案の場合は、種構造を合成するための前駆体を発生し、順合成提案の場合は、種構造を出発物質を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繰り返して、化合物ライブラリを構築する。
- 矢印は反応スキームを表し、その両端の丸印は、各反応スキームを構成する化合物を示している。
- 発生した化合物には、反応スキームの経路情報が付与され、合成経路検討の有力な情報となる。

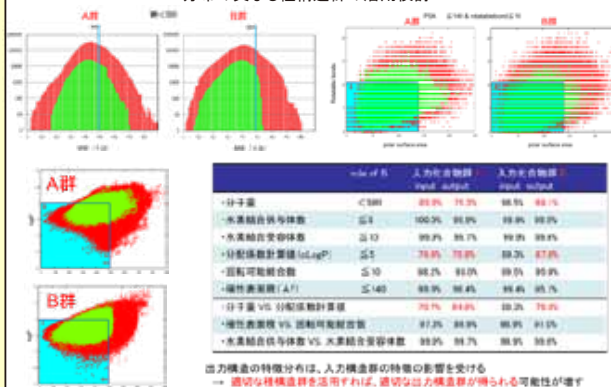
### 8. 構造出力例

- 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<sup>1)</sup>とされています。
- よって、ドラッグライクネスな化合物を種構造として、当手法により構築されたバーチャルライブラリには、医薬品のリード化合物としての要求を満たす、質の良い化合物が含まれていると期待されます。



### 9. Lipinski's rule of 5 を指標とした評価例

分布の異なる種構造群の活用検討



(謝辞)

本資料集に記載されている「京」での計算は、2011年3月の「京」の特別運用およびその後の試験利用によって行われたものです。

また、本資料集に記載されている「京」を使用した測定値は、開発整備中の「京」による、測定時点での数値です。