

ライフサイエンス・グランドチャレンジの 現状と今後

次世代生命体統合シミュレーション研究開発グループ

姫野 龍太郎

分子スケール研究開発チーム

木寺 詔紀

細胞スケール研究開発チーム

横田 秀夫

臓器全身スケール研究開発チーム

高木 周

データ解析融合研究開発チーム

宮野 悟

脳神経系研究開発チーム

石井 信

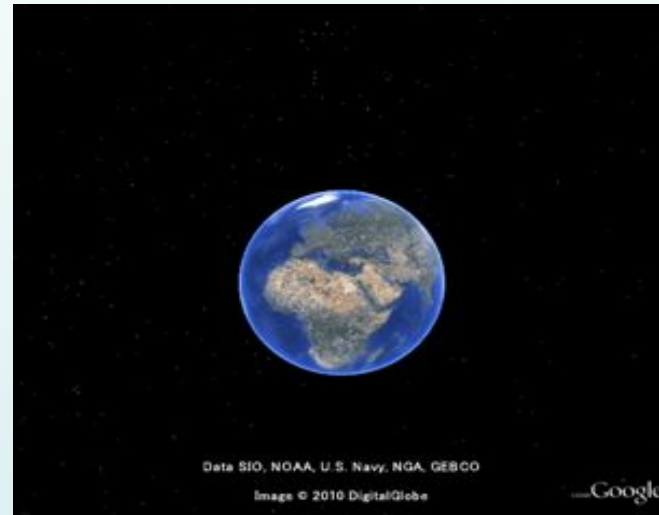
生命体基盤ソフトウェア開発・高度化チーム

泰地 真弘人

そもそもの動機

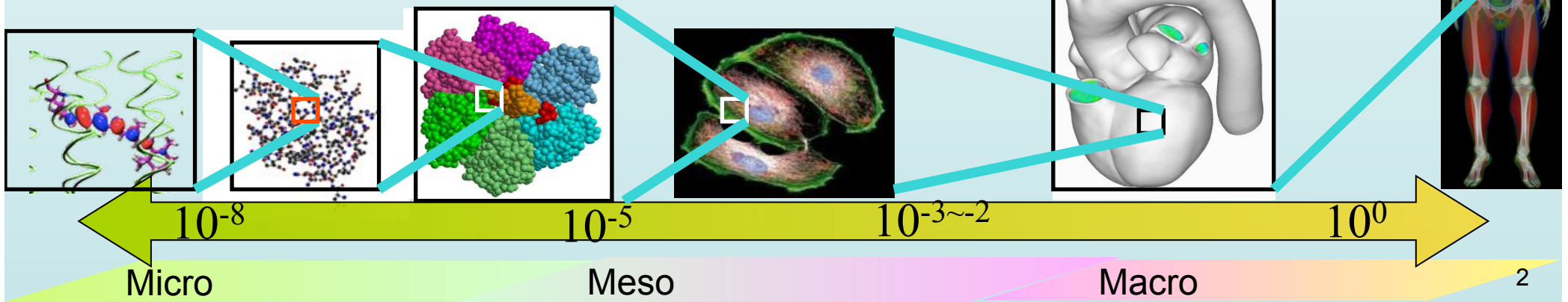
生命現象は最も複雑で難しい問題

複雑で美しい振舞いを示す**超**多体系多階層問題



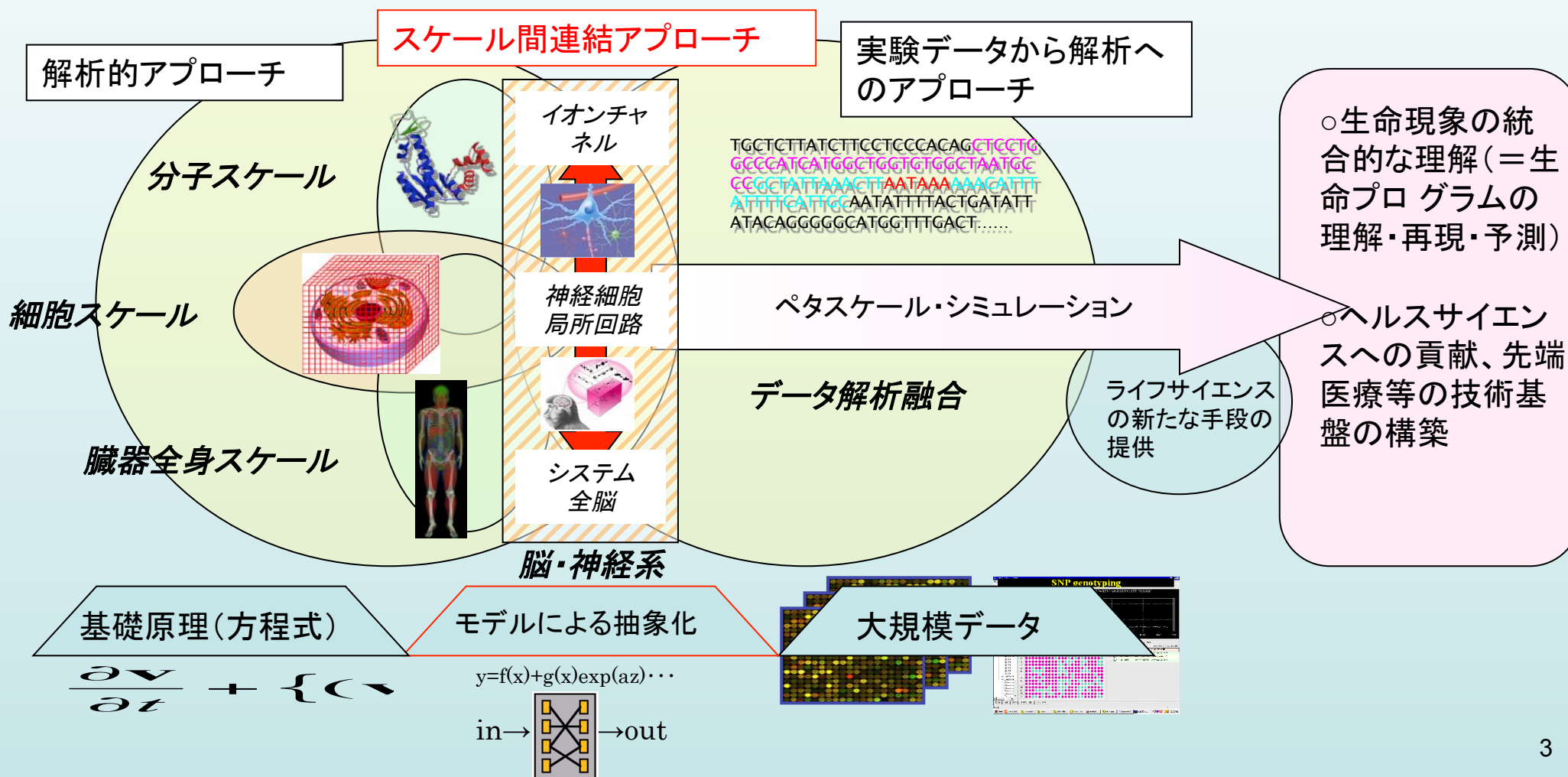
スーパーコンピュータを使って、この複雑な生命現象を解析

記述する生物学から、予測する生物学へ



研究開発の概要と達成目標

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



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

ISLiMのアプリ開発状況

Phase I < 256

Phase II < 8192

Phase III on K

II-0

II-1

II-2

1024nds

12288nds

82944nds

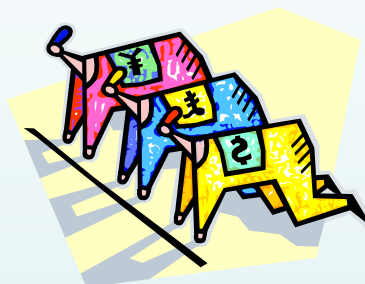


Goal: > 1 Peta FLOPS

現在27本中24本
が京で動く

cppmd, ZZ-EFSI, UT-Heartの三つは3Peta FLOPSを超えた

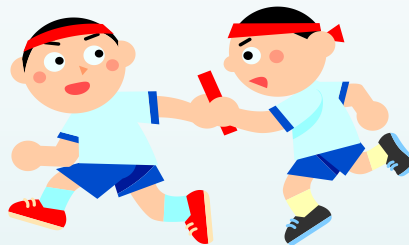
それぞれの成果アピール時期



第一走者

cppmd (分子動力学)
 ProteinDF
 CafeMol (ヒストンDNA)
 UHeart
 ZZ-EFSI
 ParaHaplo
 NEST (神経回路網)

完成後すぐに成果
 をアピール



第二走者

多剤排出トランスポータ
 MARBLE
 Platypus
 CafeMol
 RICS: 肝細胞
 HIFU ZZ-DOSE 重イオン粒子
 SIGN SSM L1GNE
 LiSDAS
 MEGADOCK
 CMDN: 大脳皮質局所回路モ

完成後、1年以内に
 成果をアピール



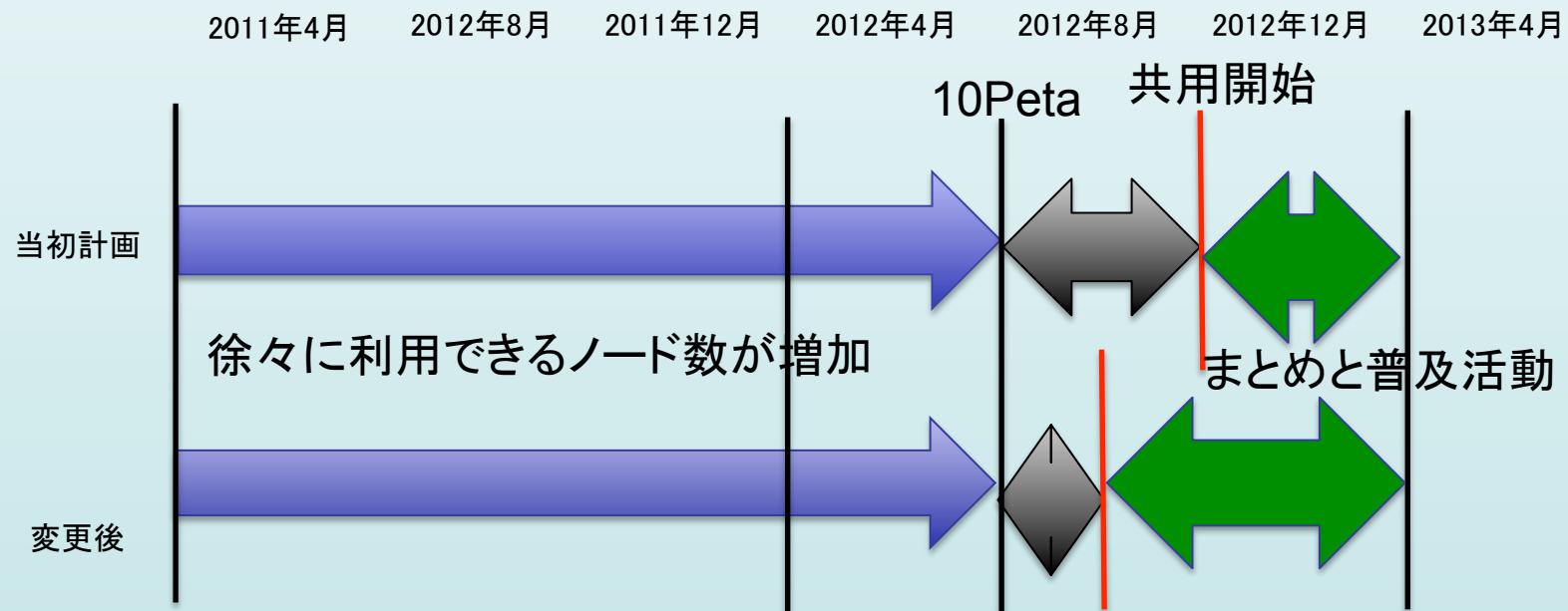
第三走者

神経系統合シミュレータ
 細胞から臓器へ
 分子から細胞へ

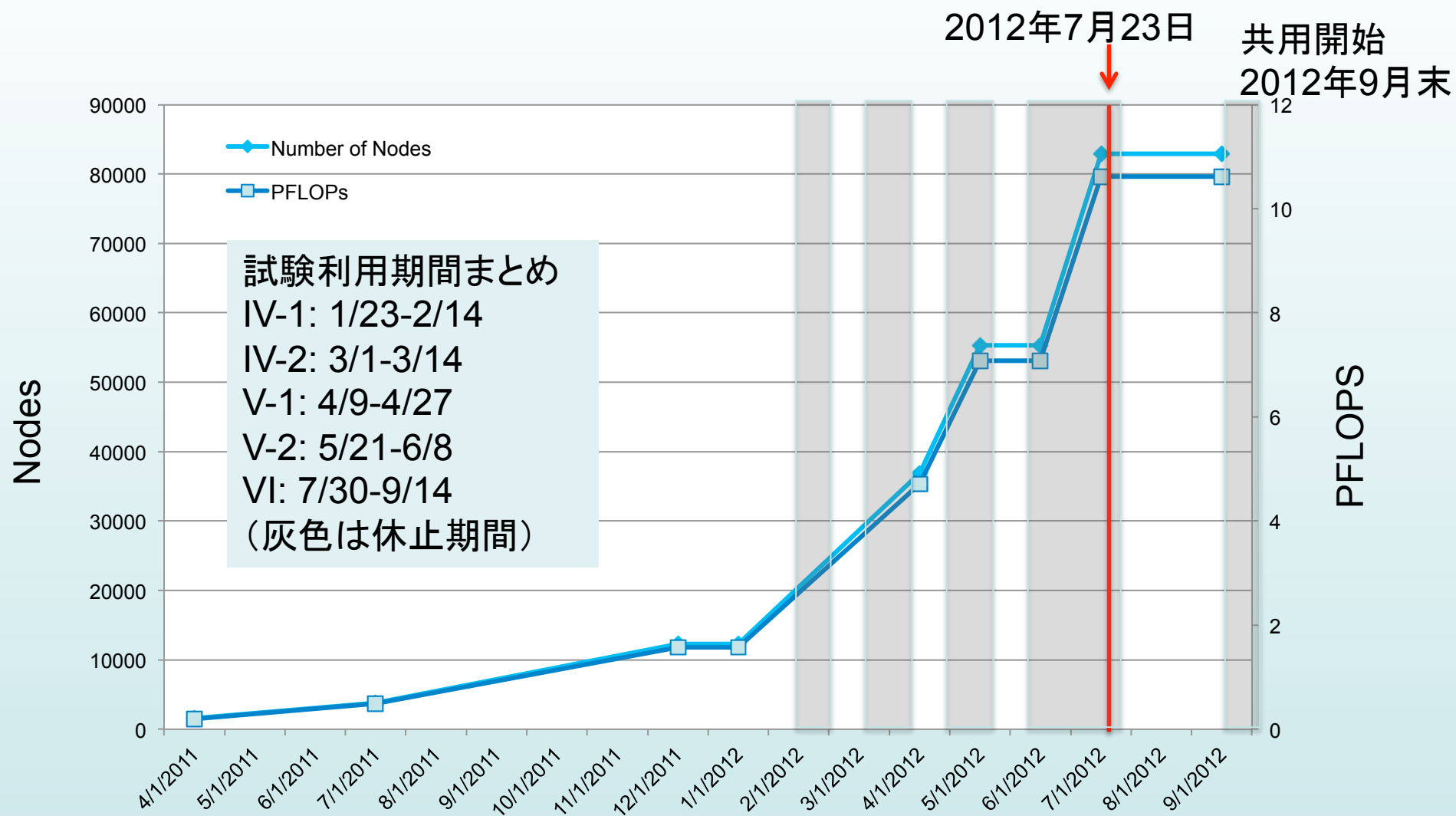
中長期目標
 スケールを越えた
 統合的理解に向け
 た取り組み

昨年度からの大きな変化

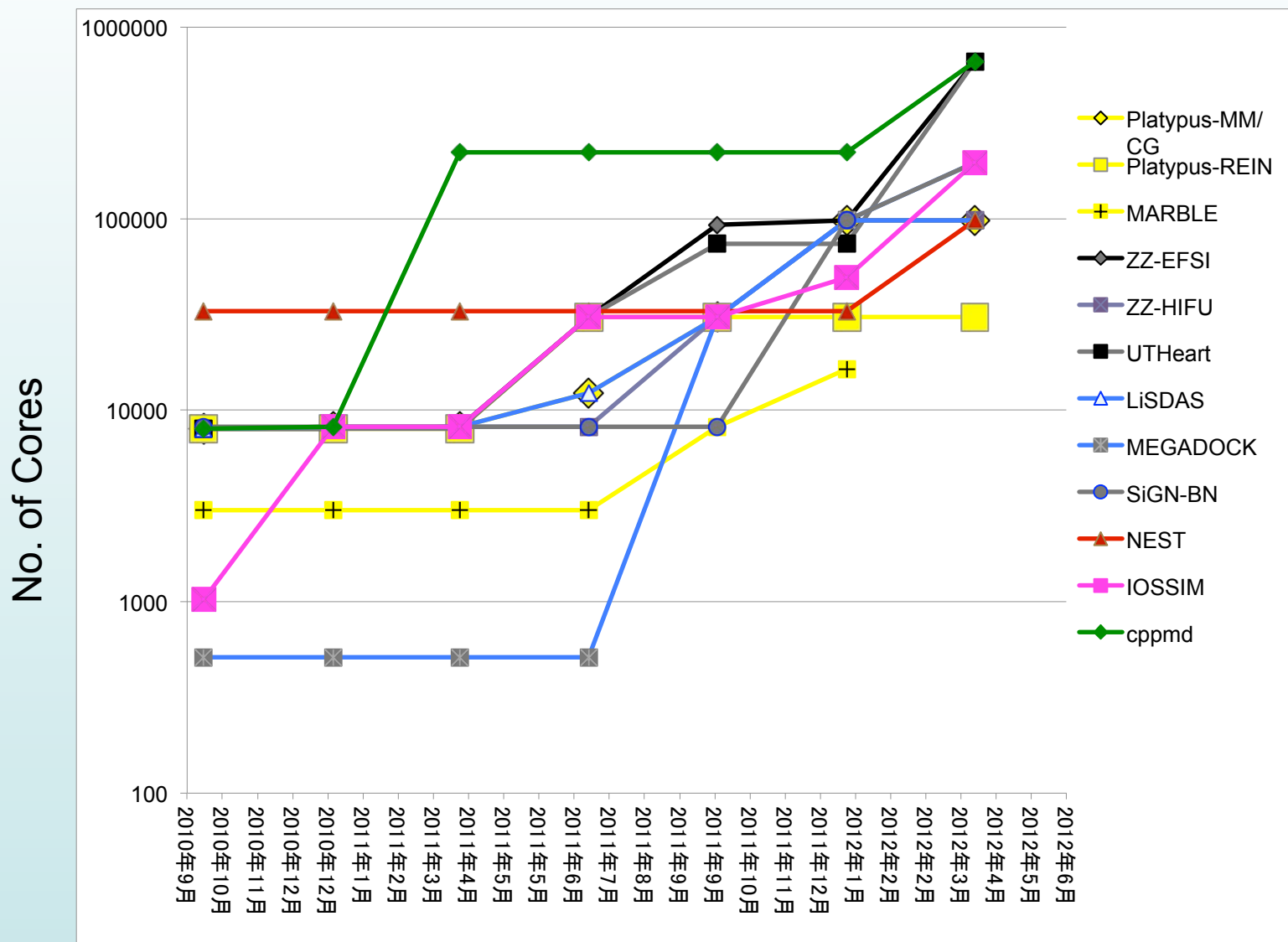
- 平成23年4月より「京」を試験利用。段階的に利用を拡大し、ソフトウェアをチューニング。
(「京」共用開始前までが本プロジェクトで利用できる期間)
- 当初計画では、平成24年11月末まで利用可能 **【「京」完成から約4ヶ月間】**
- 共用開始を平成24年9月末に変更 (平成24年2月7日 第8回HPCI計画推進委員会) **【実質約2ヶ月短縮】**
 - 衆議院決算行政監視委員会(国会版事業仕分けの影響) *何とかしてプロダクト・ランの計算時間確保が必要!*



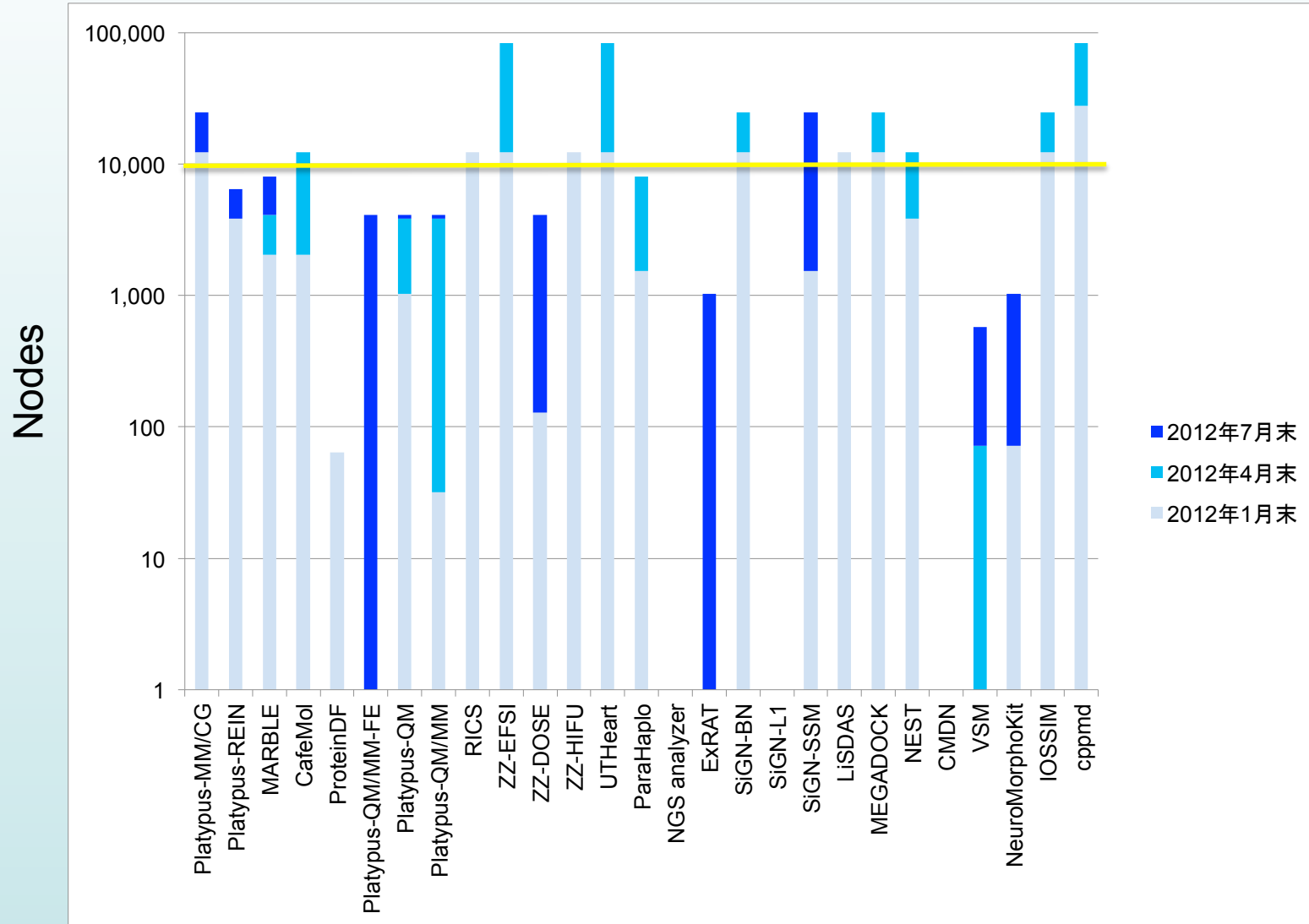
「京」試験利用計画



主なアプリと京利用ノード数拡大への追従のようす



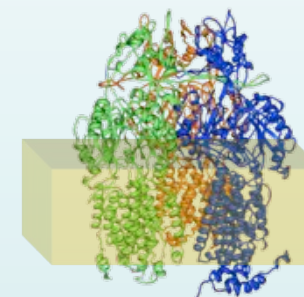
全アプリの京で使用した最大ノード数 27本中13本が10,000ノード以上を達成



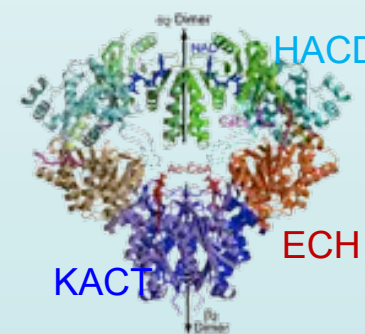
分子スケールチーム：概要

- チーム全体目標
 - 分子スケールチームでは、量子化学(QM)、古典力学(MM)、粗視化モデル(CG)の方法とそれらを連成するQM/MM、MM/CGの方法を確立し、共通ターゲットとして選んだ多剤排出トランスポーターと脂肪酸代謝酵素の機能を解析する
- 主要な成果
 - 多剤排出トランスポーター: CGで薬剤の結合、排出メカニズムを明らかにした。MMでは薬剤の取り込み機構を明らかにした。
 - 脂肪酸代謝酵素: 大規模なドメイン運動によりチャネリング機構があり得ることを示した。
 - MM/CGの新規法を確立し、タンパク質系での精密な自由エネルギー地形を明らかにした。
- 今後の予定

戦略プログラムへ引き継がれる課題と一般課題として採択された課題で、ほぼ今後の研究の継続が担保された。年度末までに、共通ターゲットにおけるさらなる成果を創出する。



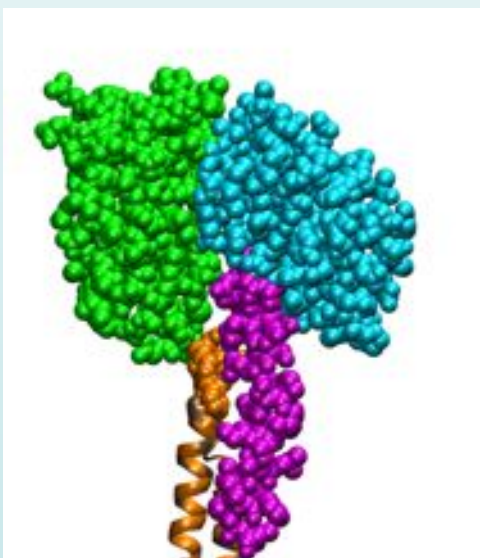
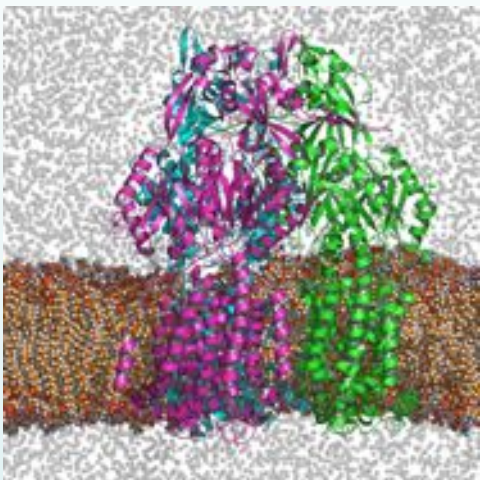
多剤排出トランスポーターAcrB



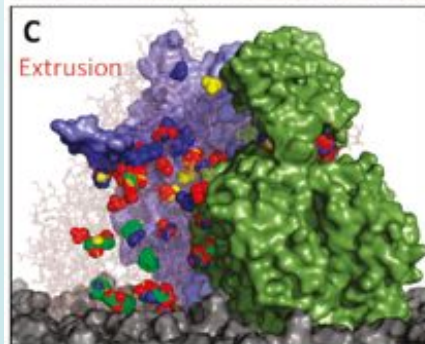
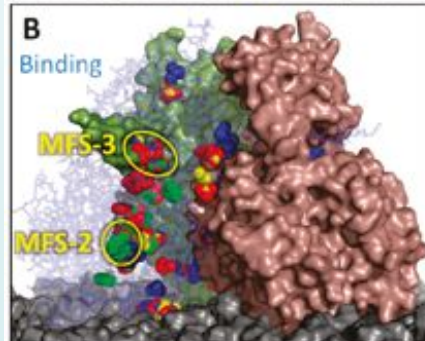
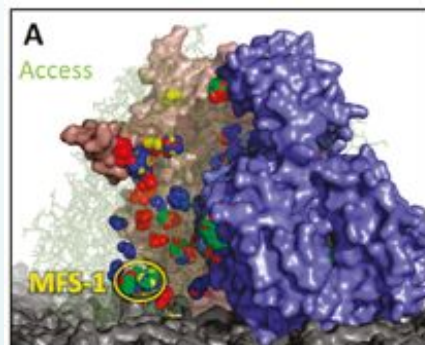
脂肪酸代謝酵素

多剤排出トランスポータの薬剤排出と取込

薬剤取込/MM

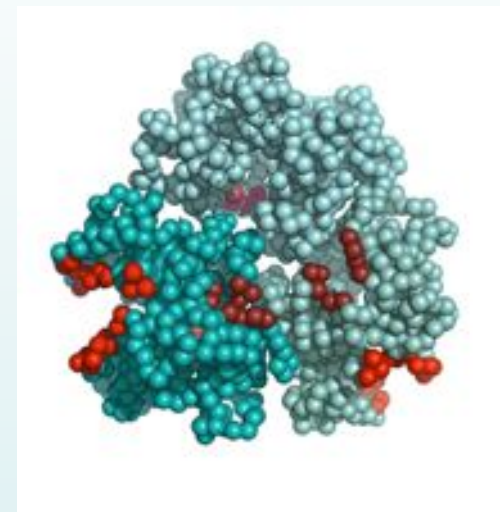


薬剤結合/CG

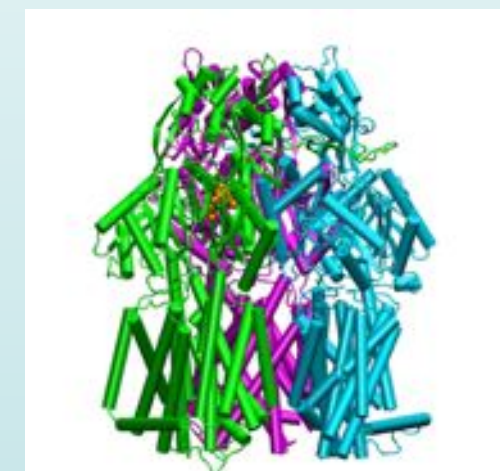


3D-RISM

薬剤排出/CG



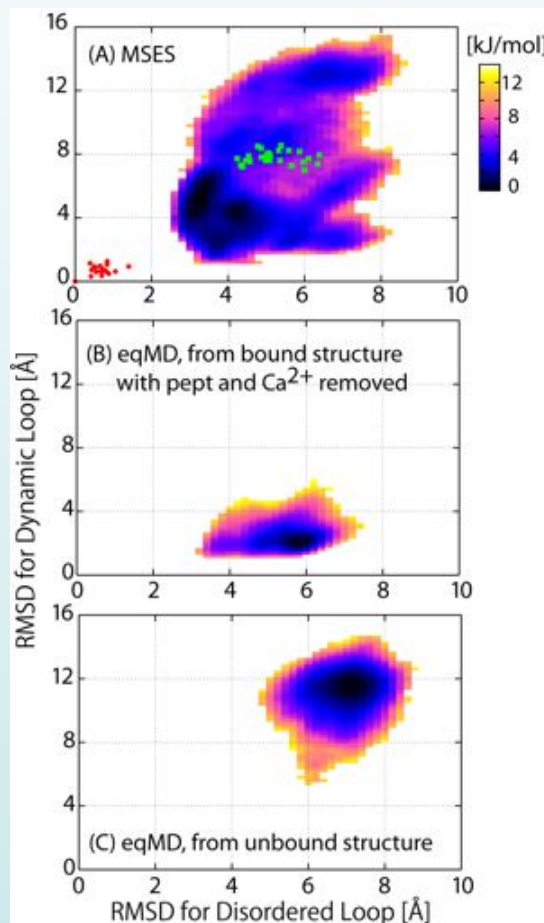
薬剤排出/MM



MM/CG

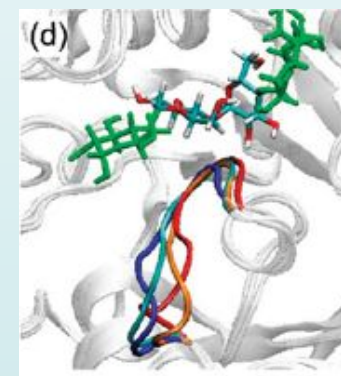
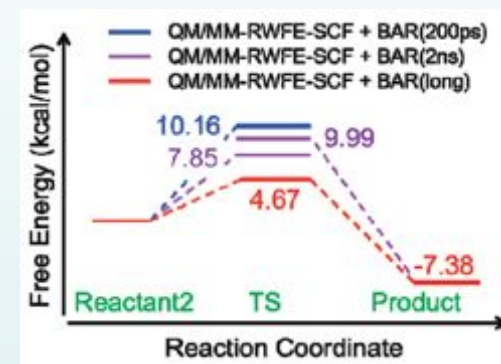


On-the-fly String法
アデニル酸キナーゼの基質結合に伴う立体構造変化



MSES法
Sortaseの天然変成ドメインの構造分布

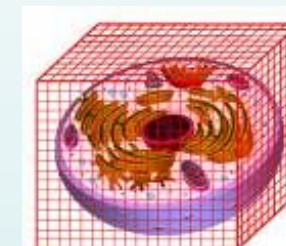
QM/MM



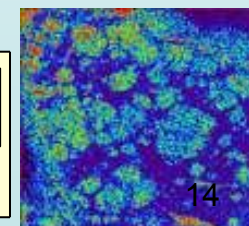
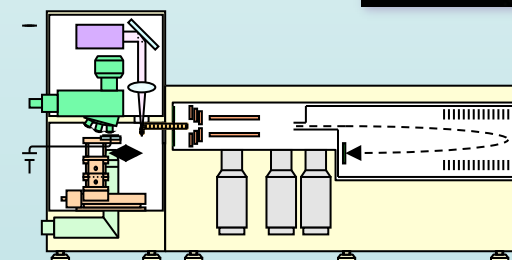
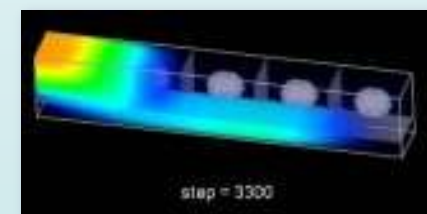
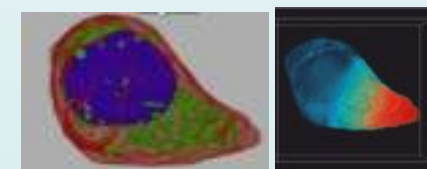
QM/MM-RWFE-SCF法
 α -アミラーゼの大規模長時間運動と共役した酵素反応

細胞スケール研究開発チーム：概要

- チーム全体目標
 - 臓器、組織内での細胞機能のうち、特に実験で観測される肝臓の代謝機能を中心にその再現を目指す。
 - 細胞内の物質の時空間変化に対応するため、細胞内を微細な格子に切り分け(ボクセルモデル)、細胞内成分の濃度などの偏りと変化を再現可能とする。
 - チーム間連携課題として、上記手法を臓器全身スケールでの血小板代謝、神経細胞チームとの膜電位機能の再現に展開する。

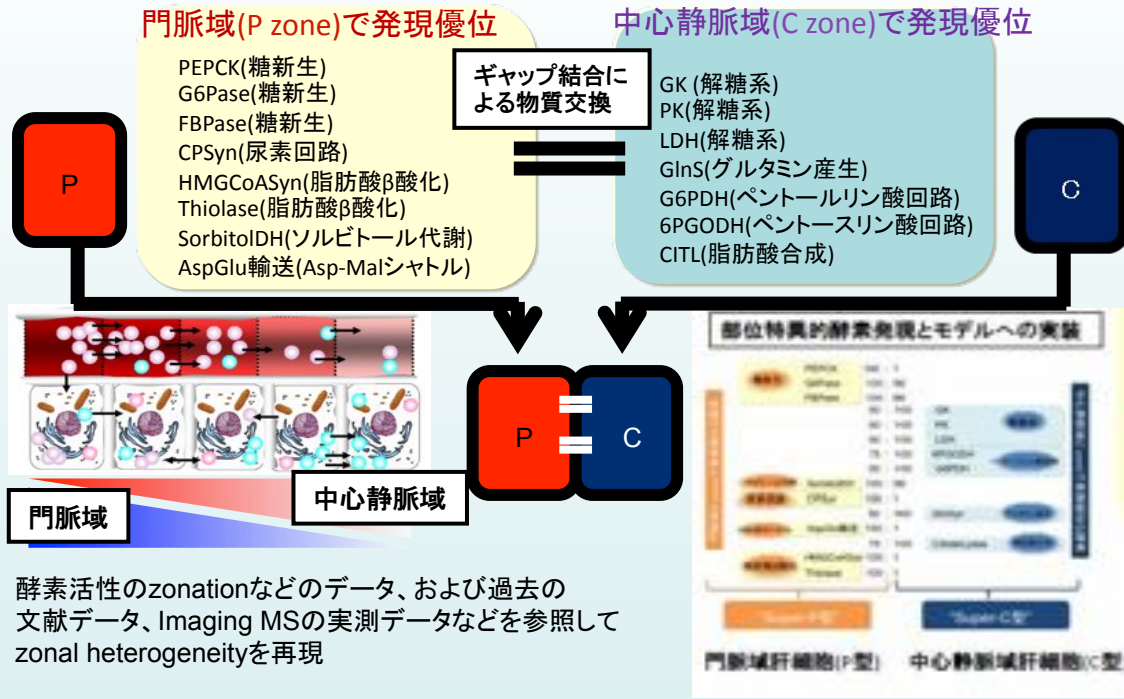


- 個別テーマで際立ったもの
 - 肝細胞代謝モデルの構築と機能分担を解明
 - 実測情報・文献情報を元に代謝ネットワークを開発し、肝細胞代謝シミュレーションを実現することにより、肝細胞の機能分担による恒常性を再現
 - 血小板活性と血流連成シミュレーションを開発
 - 神経細胞膜電位シミュレーションを実現
 - 細胞の時空間シミュレータRICSの開発

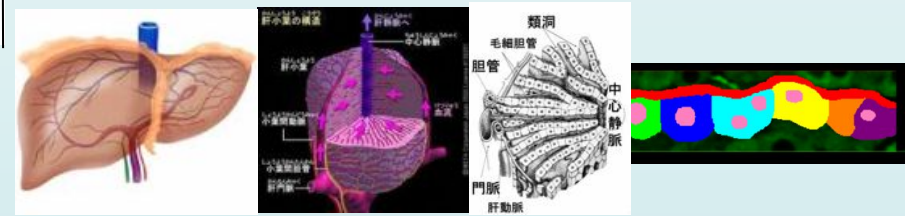
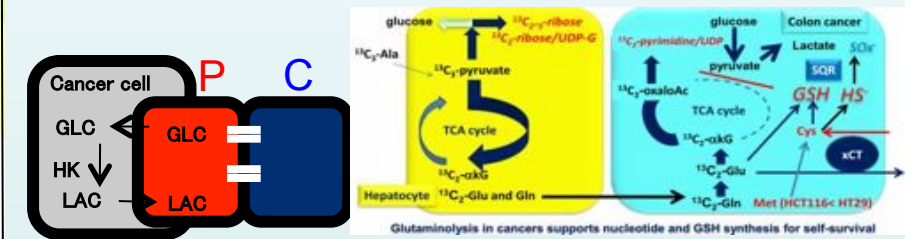
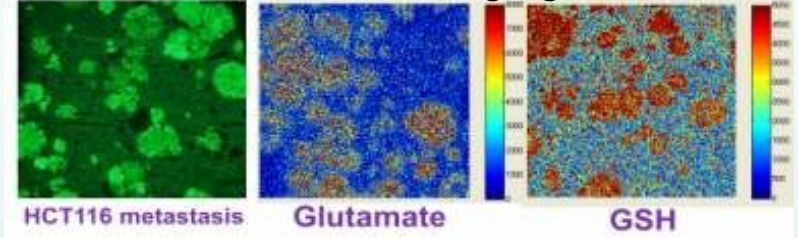




肝細胞シミュレーション



肝細胞・大腸癌代謝シミュレーションを用いた代謝システム応答予測とImaging MSによる実証



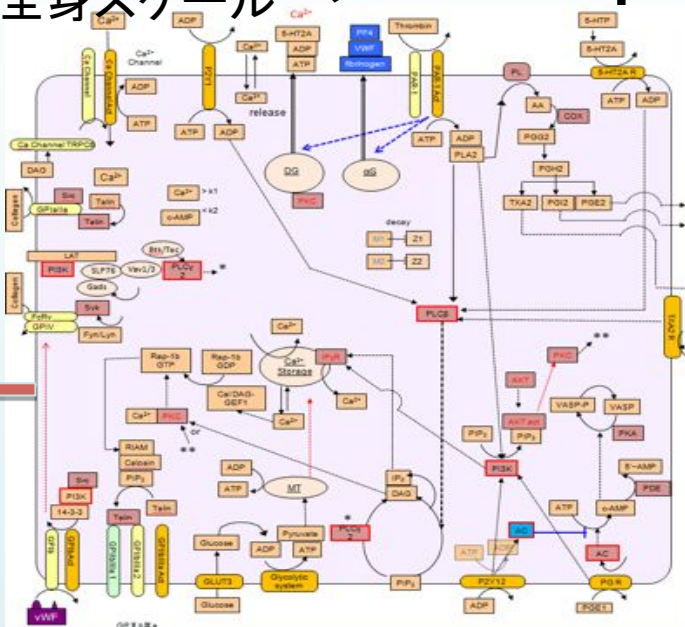
「京」にてRICSによる肝細胞詳細代謝モデルが動作！
RICCにて小組織の時空間シミュレーションが動作

計算リソース等から、小規模計算にて結果が望める小組織(複数細胞+血流)解析に集中

- ・肝臓における細胞の機能分担を解析
- ・血流との連成解析により、肝組織における細胞機能を解析
- ・転移性肝がん(大腸がんの肝転移)における肝臓予備能の予測と実証

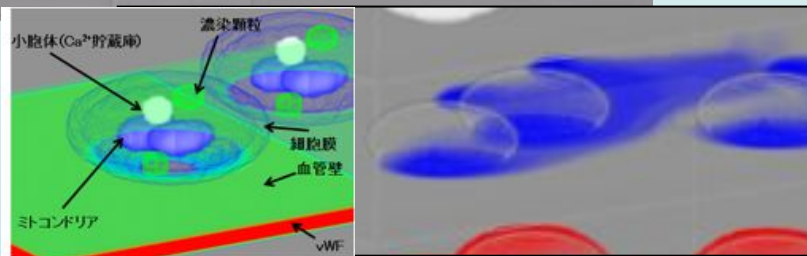
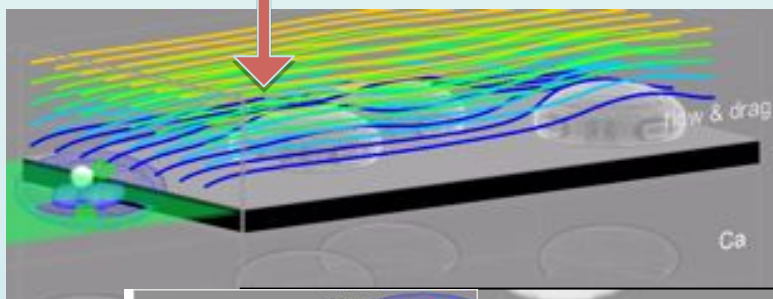
チーム間連携課題

臓器/全身スケール



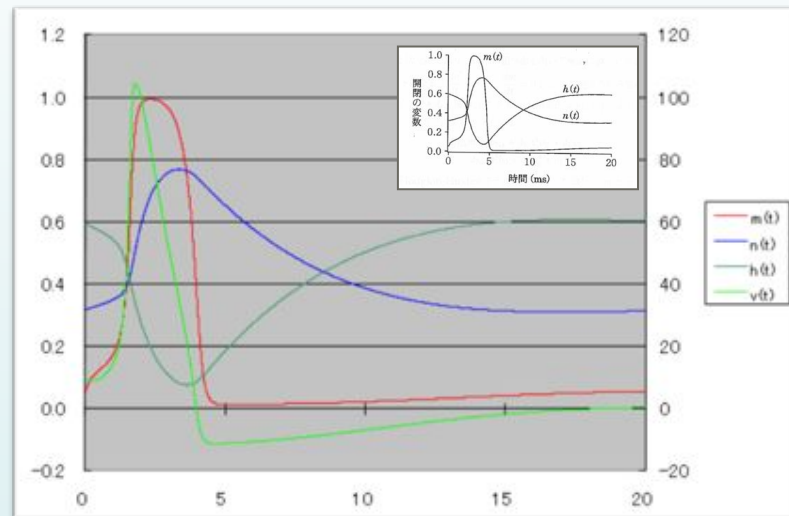
血小板細胞代謝
／活性化モデル

血小板細胞と
流体の連成
計算

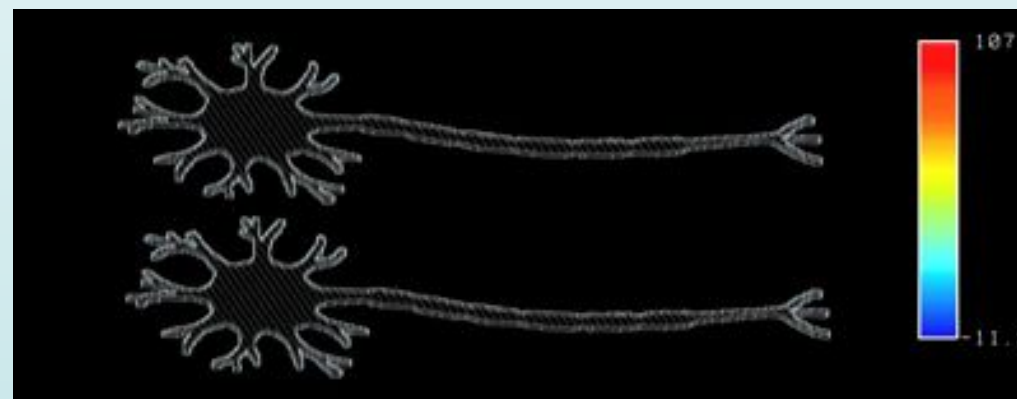


血管内の血小板活性化血流連成シミュレーション

脳神経チーム



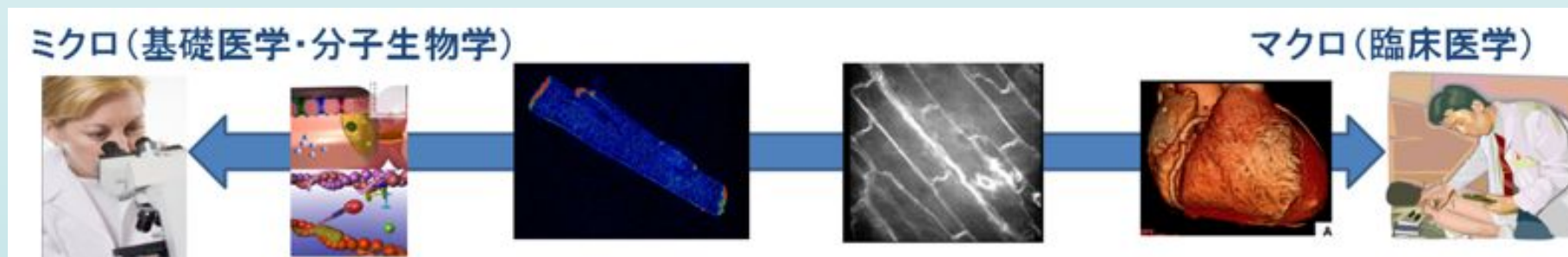
膜電位機能にHodgkin-Huxleyを実装



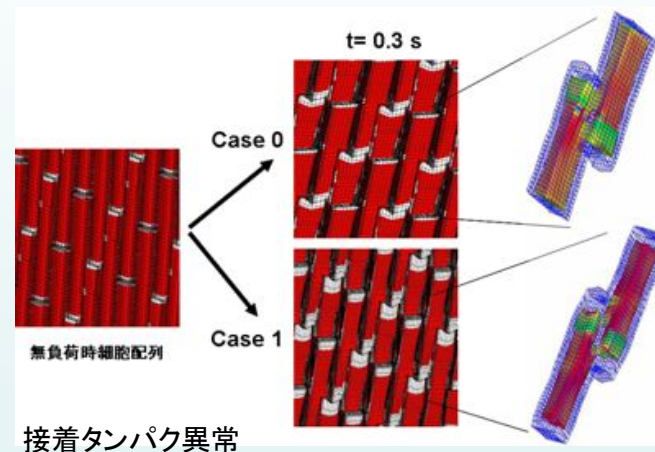
神経細胞膜電位機能シミュレーション

臓器全身スケールチーム：概要

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



- UT-Heartは既に臨床研究に使用中
- グランドチャレンジで開発されたソフトのうち、唯一、「京」利用の優先課題
- 京の全システムを使った計算を実行中

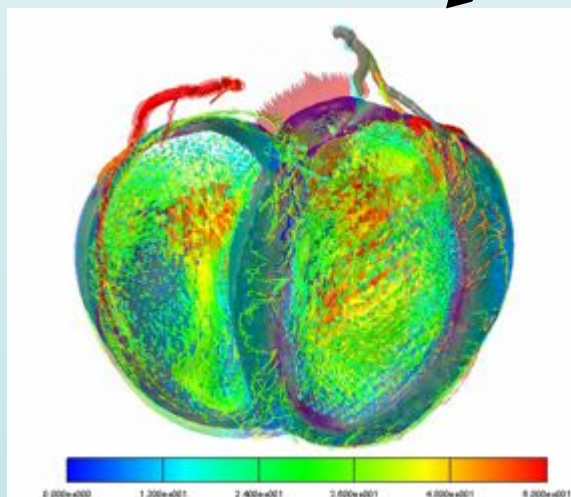


京では「毛細管まで再現した冠循環」と「細胞内の代謝機構」を導入し、虚血性心疾患など多様な病気を再現

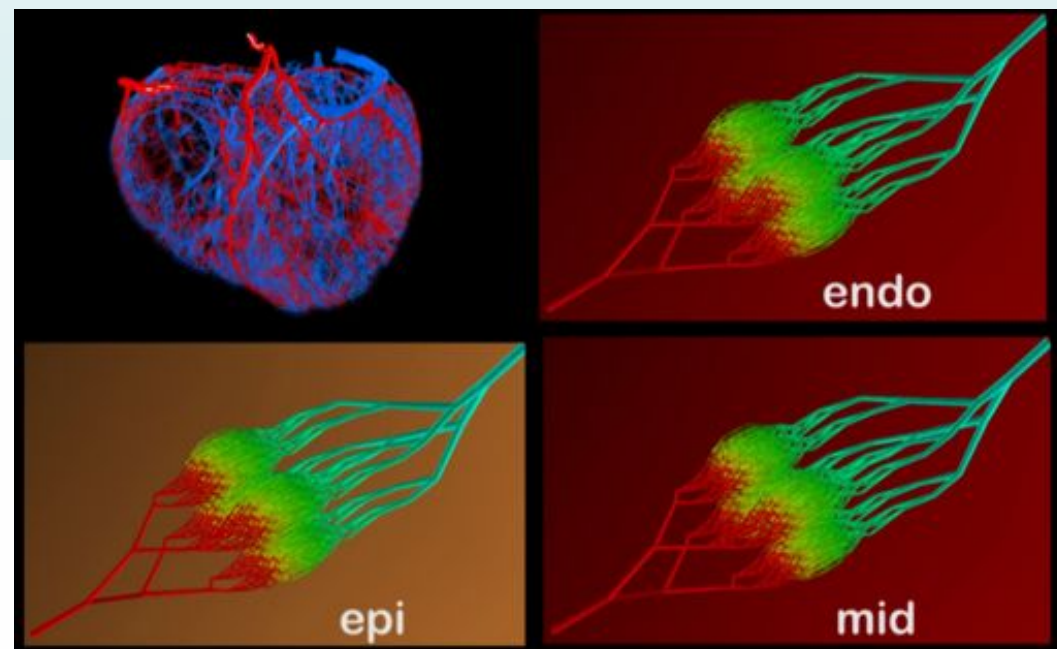
閉塞
↓

右の動画はRICCを使った予備的計算で、冠状動脈(左回旋枝)閉塞により、始めは酸素、やがてATPが欠乏する急性心筋梗塞のシミュレーション。色は収縮力を表す

可視化: 富士通(株)

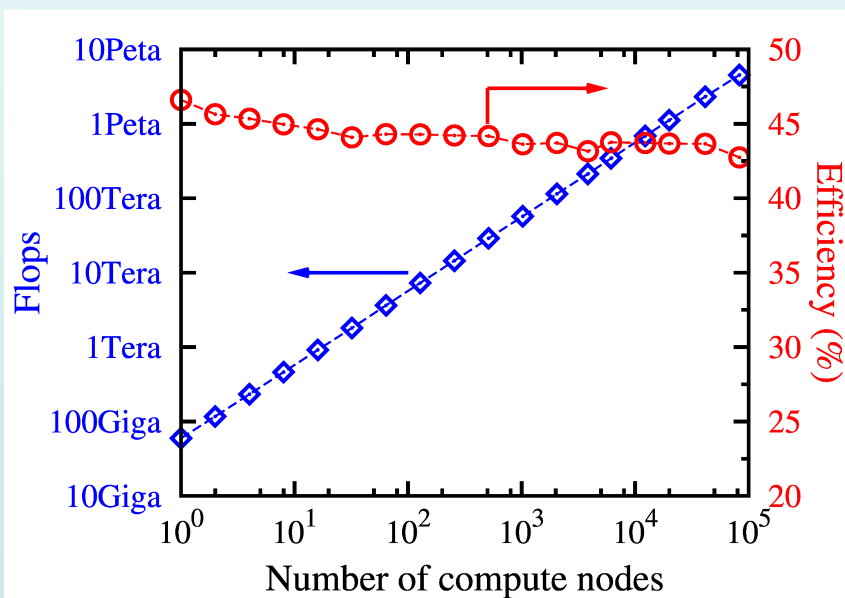


正常な冠循環: 大動脈から入った血液は4拍程で毛細血管床に至る

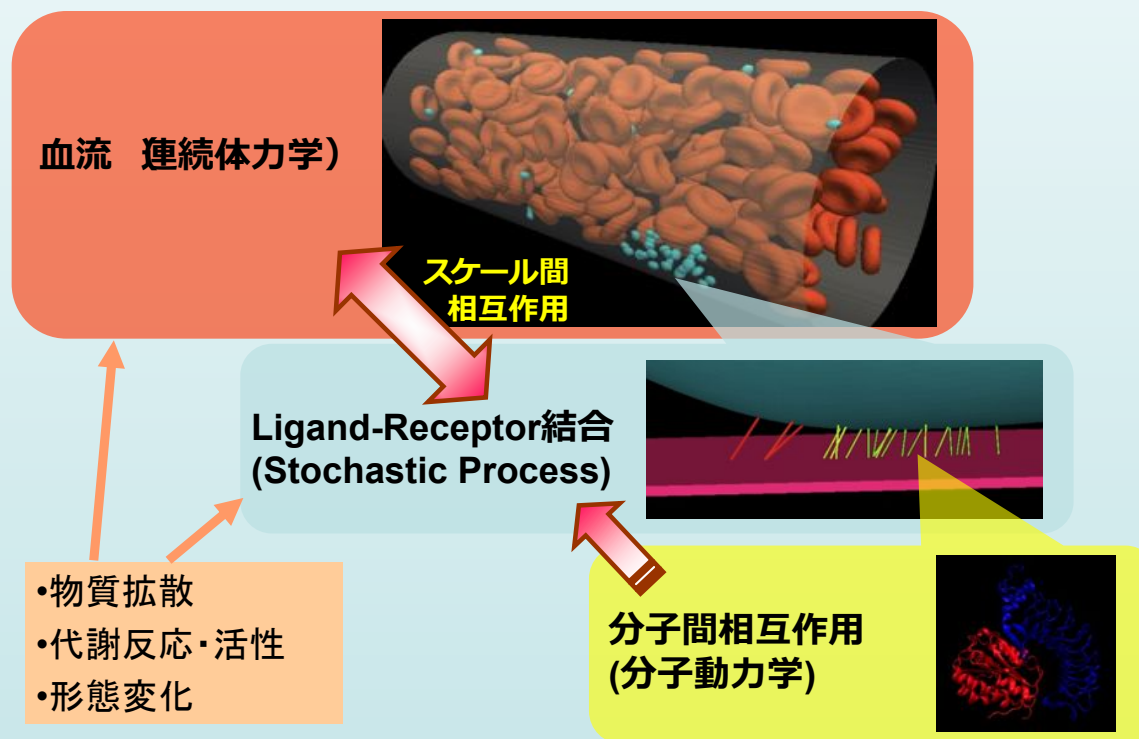


医用画像データに適した流体構造連成手法(ZZ-EFSI)と マルチスケール血栓シミュレータ(ZZ-Throm)

- 「京」のフルノード計算で、実効速度 4.5ペタフロップスを達成(世界最速の流体構造連手法ZZ-EFSIの開発に成功)
- 血小板表面のGP1b α 分子と血管壁上のvWF分子のタンパク質間相互作用に状態遷移理論を適用したモデルにより、血流計算との連成に成功. 血栓接着のよりリアリスティックな計算が可能となり、赤血球の存在が血小板接着には重要とする実験的事実をシミュレーションにより実証.



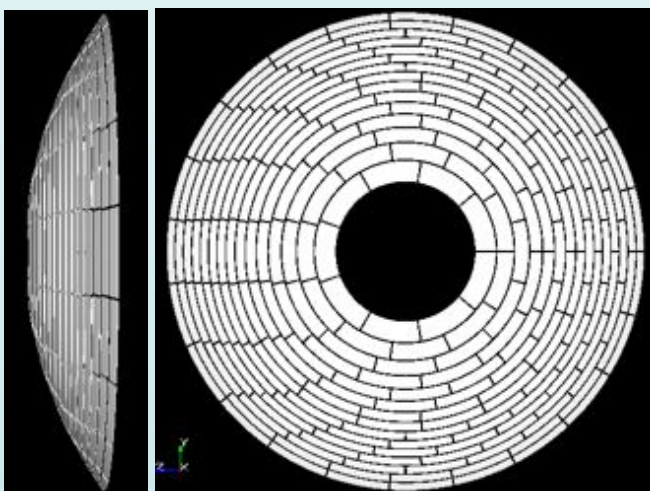
・実効速度 **4.5ペタフロップ**を達成!!
(**世界最速**の流体構造連成手法開発)



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

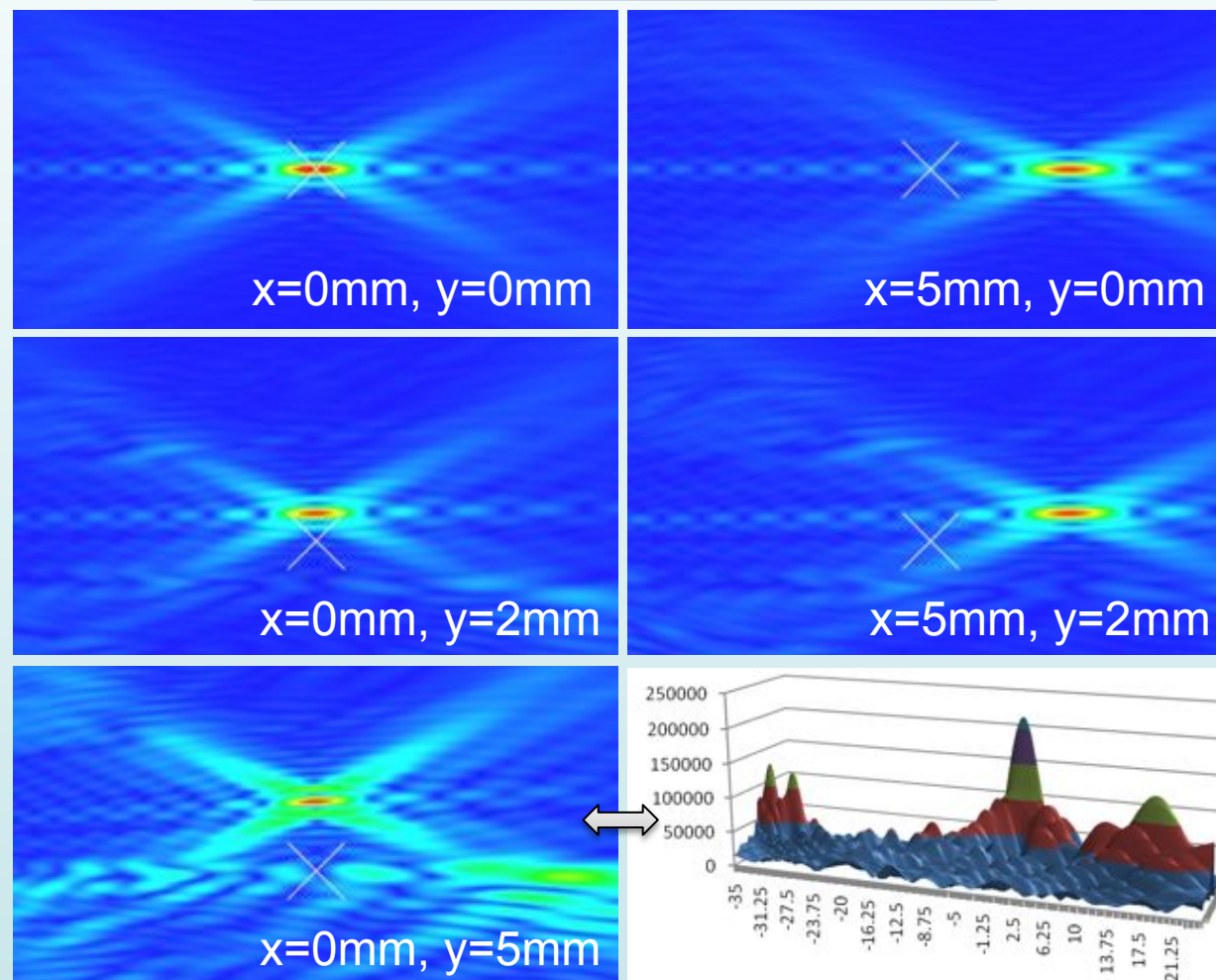
TSBMI*における低侵襲診断治療機器の開発との共同

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

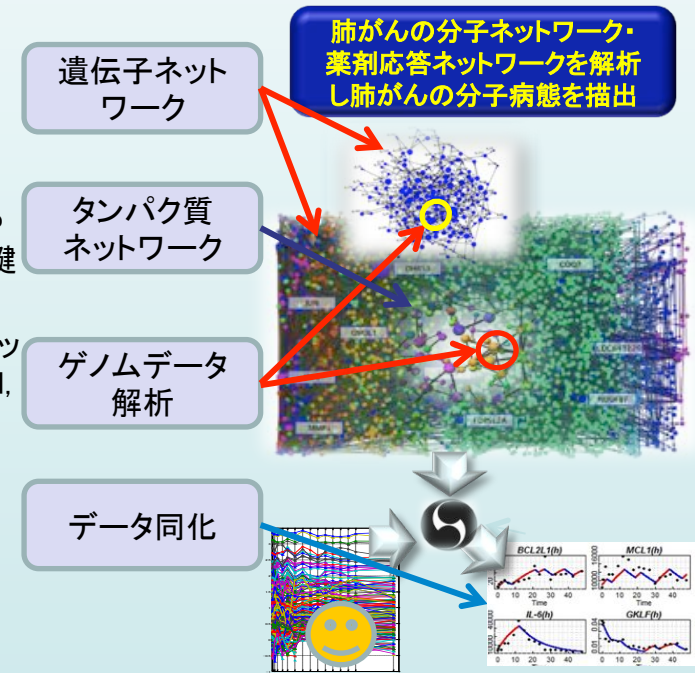


256ch, 2MHz, F=0.8

焦点位置による超音波音場の変化



- チーム全体目標
 - 「肺がんと薬」を共通のテーマにして、ゲノムと疾患の間にあるシステムと薬剤応答を解析することを可能にするソフトウェアを、ゲノムデータ解析、遺伝子ネットワーク解析、タンパク質ネットワーク解析、及びデータ同化によるシミュレーションの観点から開発
- 個別テーマで際立ったもの
 - 大規模遺伝子ネットワーク推定ソフトウェアSiGNの開発とその応用
 - ・ L1正則化法にもとづく大規模遺伝子ネットワーク解析で、がんの浸潤やゲフィチニブなどの抗がん剤耐性の原因となっている上皮間葉転換の鍵遺伝子群を一挙に捉えることに成功 (SiGN-L1)
 - ・ 大規模遺伝子ネットワーク解析により、肺腺がんの予後の良・不良スイッチとなっているサブネットワーク・遺伝子群の同定と予後予測 (SiGN-BN, SiGN-SSM)
 - 生命体シミュレーションのためのデータ同化技術及びソフトウェアLiSDASの開発
 - タンパク質ネットワーク推定ソフトMEGADOCK
 - SNP解析ソフトParaHaplo
- 今後の予定
 - 実験系プロジェクトと連携しソフトウェアを普及



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

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

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

L1正則化法に基づく大規模遺伝子ネットワーク推定 (SiGN-L1)

(1024コアを使って理論上3週間、実際は3ヶ月の計算)

regulator	type	regulatory effect change	Evidence
IRF6	A	101.04	
miR-141	A	87.58	Nat Cell Biol 10(5): 593-601, 2008
GRHL2	A	64.13	Cancer Res 72(9):2440-53, 2012
ZEB1 (SIP1)	I	50.72	Mol Cell 7(6): 1267-78, 2001
LSR	I	46.89	
miR-200b	A	31.55	Nat Cell Biol 10(5): 593-601, 2008
KLF4	A	26.28	J Biol Chem 285(22):16854-63, 2010
OVOL2	A	22.08	
miR-200a	A	17.70	Nat Cell Biol 10(5): 593-601, 2008
FOXA2	A	17.26	Cancer Res 70(5):2115-25, 2010
TCF4 (E2.2)	I	14.15	J Cell Sci 122(Pt 7): 1014-24, 2009
ELF3	A	13.58	
ZNF217	A	13.53	
MYB	A	12.50	
KLF5	A	12.42	PLoS ONE 6(6): e20804, 2011
miR-192	A	12.30	PNAS 104(9): 3432-7, 2007
FOXA1	A	11.69	Cancer Res 70(5):2115-25, 2010
ZNF165	A	11.39	
NKX2-1	A	11.21	Cancer Res 69(7):2783-91, 2009
HNF1B	A	11.08	
TFE3	A	11.01	
ZEB2 (δ EF)	I	10.66	Oncogene 24(14):2375-85, 2005
TRIM29	I	9.87	
SNAI2	I	9.74	Cancer Res 62(6): 1613-8, 2002

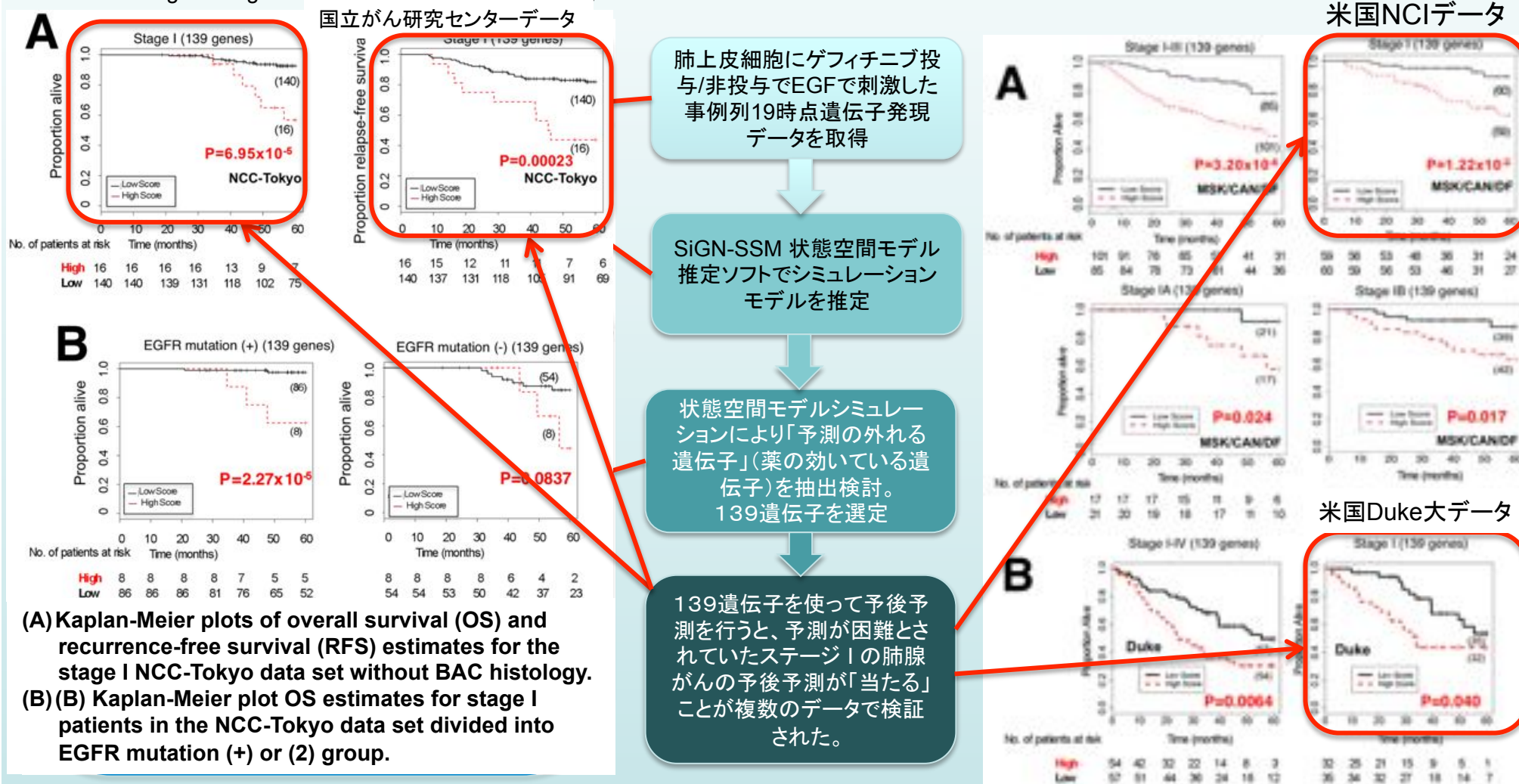
- スパコンがEMT制御因子として予測したトップ24遺伝子のうち12遺伝子が「当たっていた」。残り12遺伝子についてはEMTに関する直接的な結果は無し。
- 高橋隆 (名大医) がKLF5のノックダウンでEMTが起こることを実験で証明 (Shimamura et al. PLoS ONE 2011)
- GRHL2がEMT誘導することが発表される (Cieply B et al. Cancer Res 72(9): 2440-53, 2012)

Shimamura T, Imoto S, Shimada Y, Hosono Y, Niida A, Nagasaki M, Yamaguchi R, Takahashi T, Miyano S. (2011) A Novel Network Profiling Analysis Reveals System Changes in Epithelial-Mesenchymal Transition. PLoS ONE 6(6): e20804.

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

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

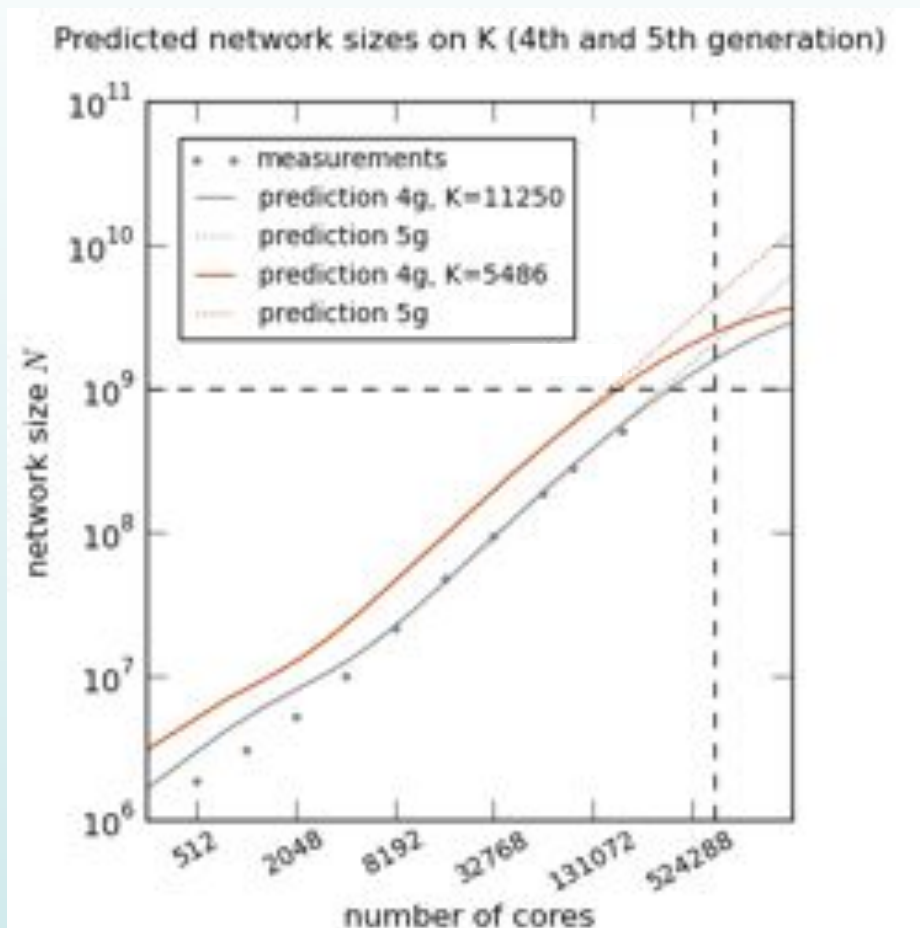
The power of survival prediction for the NCI validation data set, the Duke data set and NCC-Tokyo data set using the 139 genes.



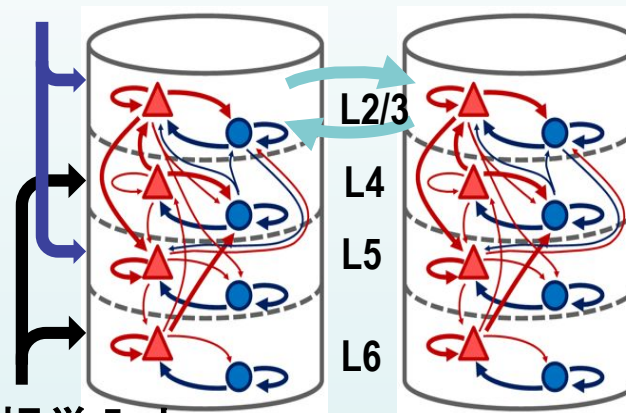
脳神経系研究開発チーム：概要

- チーム全体目標
 - ほ乳類視覚系，無脊椎動物嗅覚系における入出力関係を再現できるソフトウェアと共に，それらの構成要素である局所回路および神経細胞の動態を再現できるソフトウェアを作成する。
- 個別テーマの成果概要
 - 神経細胞シミュレーションは，京上の10000並列により細胞骨格制御を再現可能にするとした当初計画を達成，フィロポディア形成，軸索誘導などのシミュレーションに展開。
 - 局所回路シミュレーションは，約 5×10^8 個の細胞からなる回路のシミュレーションに成功，これを用いて視覚注意の影響を調べた（詳細は各スライド）。
 - ほ乳類視覚系シミュレーションは，視覚入力系の各層モデルについて京の1000並列以上により当初計画を達成，出力系は理研RICC上のシミュレータとロボットヘッドを接続することで，クローズドループオンライン制御（当初計画）を達成した。
 - 無脊椎動物嗅覚系シミュレーションは，詳細構造を有する細胞のシミュレーションについて京上の196608コア並列まで達成，実験データから回路パラメータを並列探索するアルゴリズムを組み合わせることで，嗅覚系中枢回路の高精細シミュレーションを実現した（詳細は各スライド）。
- 今後の予定
 - ソフトウェアとしての性能は当初計画を達成したので，応用研究に注力すると共に，局所回路シミュレーションは世界最大の回路シミュレーションを目指したい。

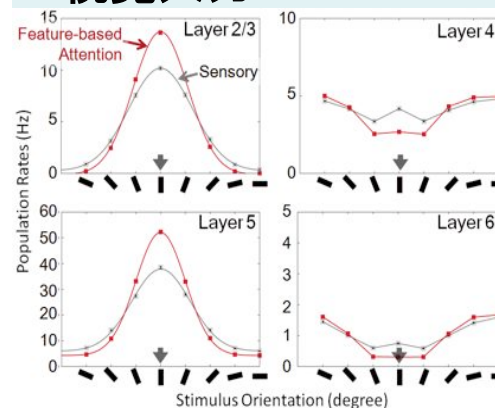
局所回路シミュレーション



注意信号



視覚入力

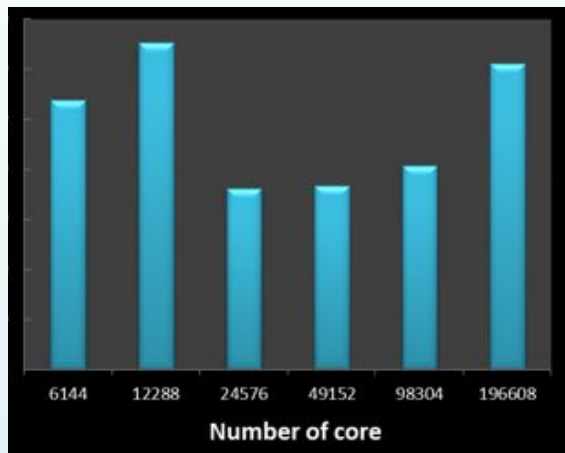


Wagatsuma et al.,
*Frontiers Comput
Neurosci* (2011)

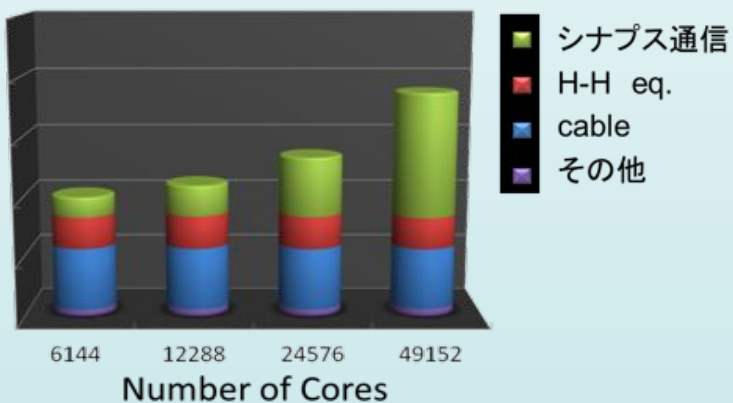
- 新しいデータ構造の導入によって、京の24576ノード(196608コア)を用いて、約 5×10^8 個の神経細胞からなる局所回路シミュレーションを達成した.
- 視覚野2カラムモデルで、トップダウン信号の各層への影響について示唆した.

無脊椎動物嗅覚系シミュレーション

詳細神経シミュレーションの高並列化

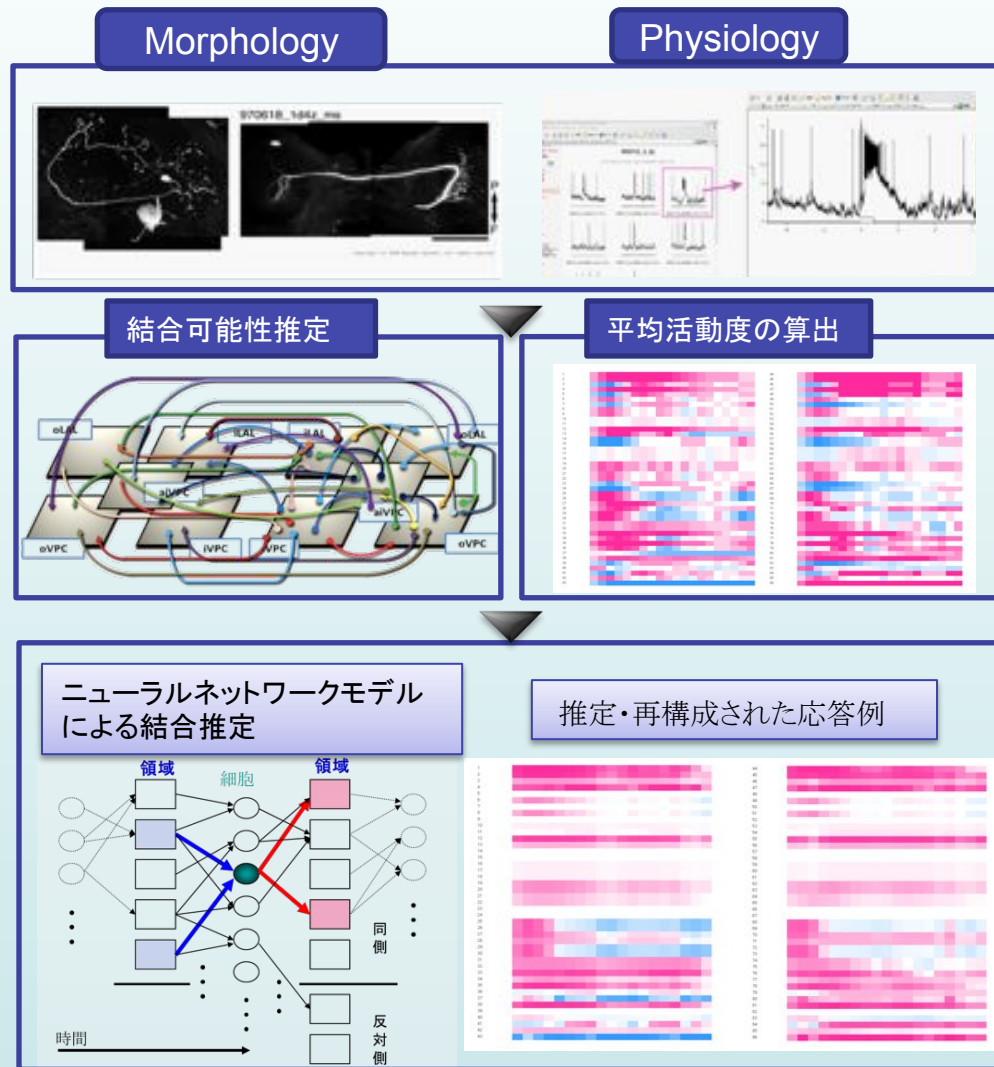


ウィークスケーリング全体通信 (Allgather)
 ・通信時間の急激な増大が1万ノード以上の領域で観測された



通信の最適化を継続実施中

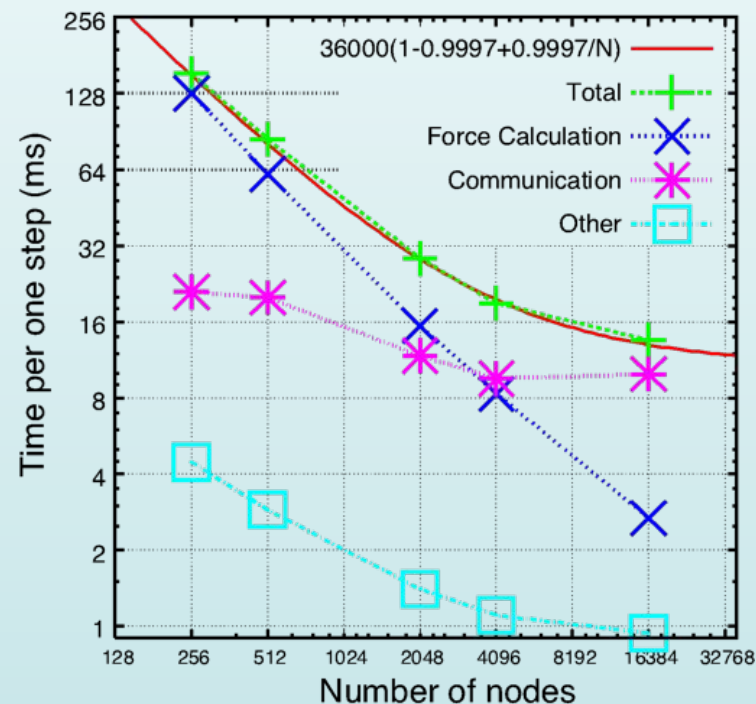
前運動中枢神経回路の結合推定



嗅覚系中枢回路の高精細シミュレーションへ

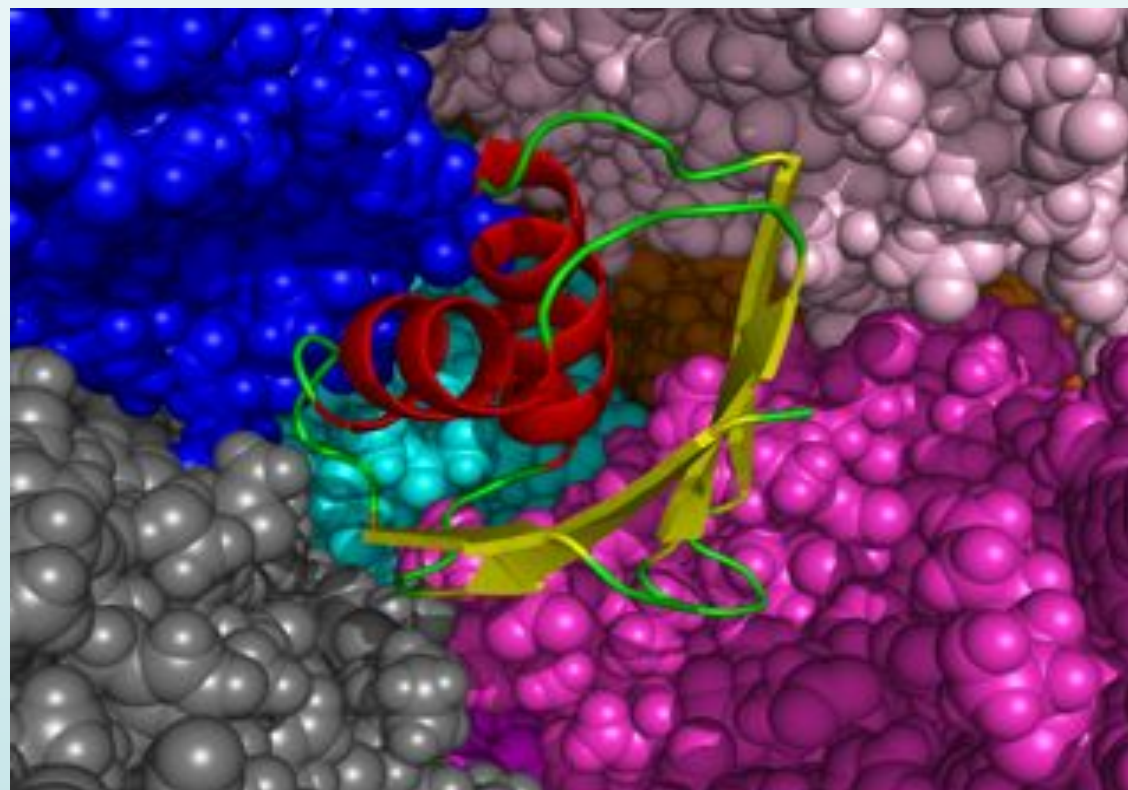
生命体基盤ソフトウェア開発・高度化チーム：概要

- チーム全体目標
 - プロジェクトで開発されたソフトウェアの京コンピュータ上での高性能化支援
 - 基盤ソフトウェア開発
 - 産業応用に向けた基盤開発(大規模仮想化合物ライブラリの開発)
- 高並列分子動力学計算コアソフトウェア
 - 京コンピュータ上での高度化のノウハウ蓄積、
またコアソフトウェアとしての他ソフトウェアの
高性能化推進のために開発
- 他ソフトウェア高性能化への支援
 - 他アプリケーションの高度化支援
 - 富士通(株)への窓口機能
- 基盤ソフトウェア開発
 - 可視化・ミドルウェア
 - 計算科学研究機構に移行(小野T)
- 大規模仮想化合物ライブラリ(東大船津研)

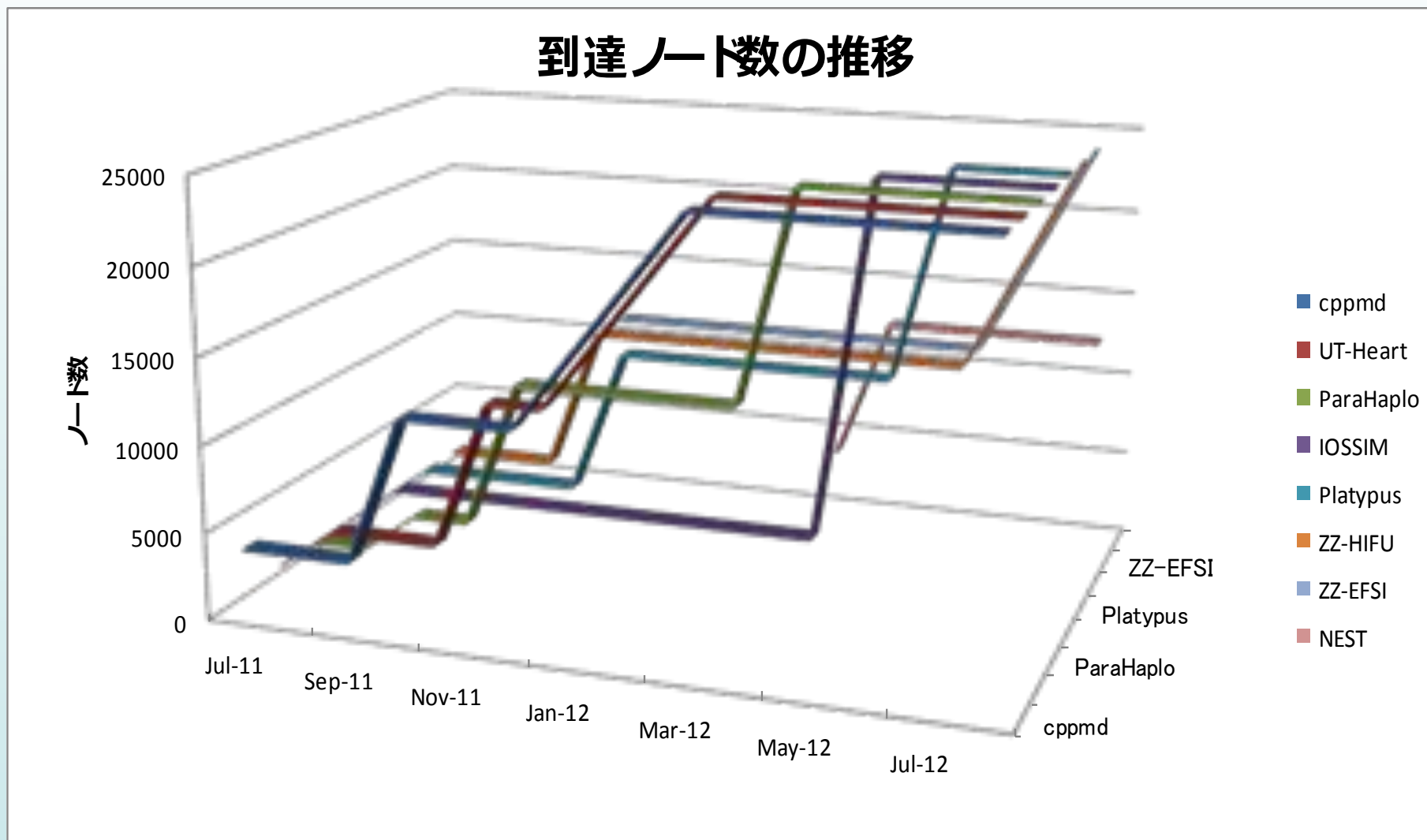


京コンピュータのフルシステムを用いて 3.8PFLOPSの実効性能を実現

- 効率 38%
- 1.8億原子、カットオフ
- カーネル効率 ~ 50%
- FMMでの効率 ~ 10%



ISLiMコードの高度化支援



ISLiMコードの高度化支援

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

- 開発は当初計画通りで、順調
 - 京の1/8以上を使っても性能が落ちないソフトを12本(全体の約1/3)開発
 - 京のソフトとして最高の実効性能を出したZZ-EFSIを初め、cppmd、UT-Heart、ZZ-HIFUで高い実行効率を達成
- ただし、京の供用開始前倒しで
 - 京の利用が2ヶ月短縮(京全体が使える期間が4か月から2ヶ月に半減)
 - この影響で当初予定していた京のフルシステムを使った計算は、実質UT-Heartだけ
- 京での計算の結果は徐々にまとまりつつあり、成果が出てきている