

課題名 (タイトル) :

マルチスケール分子動力学シミュレーション法の開発

利用者氏名 : ○森貴治, 松永康佑, 今井隆志, 宮下尚之, Jaewoon Jung,

Michael Feig, 原田隆平

所属 : 神戸研究所 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

計算科学研究機構 粒子系生物物理研究チーム

次世代計算科学研究開発プログラム 分子スケール研究開発チーム

近年、X 線結晶構造解析や核磁気共鳴分光法、一分子測定などの様々な技術の発展により、細胞内で起こるありとあらゆる現象が、核酸やタンパク質、糖鎖などの生体分子から構成されるネットワークにより制御されていることが明らかになってきた。生命現象の素過程の中で、タンパク質はしばしばミリ秒～秒スケールでダイナミックに構造変化を起こし、一方で、酵素反応や物質認識、物質輸送といったピコ秒～ナノ秒スケールの生化学反応を厳密に制御している。生体分子のこのようなマルチスケールな現象を、原子・分子レベルで理解することは非常にチャレンジングな問題である。本研究グループでは、主に新しい分子動力学法シミュレーション法を開発し、それらを用いてタンパク質の動的構造の理解と機能メカニズムの解明を目指す。特に、細胞環境を強く意識したシミュレーションや ATP などのリガンド結合がキーとなるタンパク質のミリ秒スケールのダイナミクス解析を行う。本研究では、京コンピュータの利用も視野に入れて、高並列分子動力学計算プログラムの開発も行う。

課題 1 : Solvation in cellular environments: How does the interaction with aqueous solvent change in crowded environments? (Feig・原田)**1. 本課題の研究の背景、目的、関係するプロジェクトとの関係**

Macromolecular crowding has been shown to alter biomolecular structure and dynamics compared to dilute environments. Past studies of crowding have focused largely on volume exclusion by crowder molecules with the main conclusion that such crowding effects lead to more compact states. Most discussions of crowding so far have focused on the nature of protein-crowder interactions, while

relatively little is known about the effect of crowding on the structure and dynamics of water. However, the effects of crowding on water properties are a key part in developing a full understanding of the behavior of biomolecules in cellular environments. In order to obtain a better understanding of water in crowded biological environments, we performed a series of fully-atomistic molecular dynamics simulations of highly concentrated protein solutions to mimic biological crowding environments.

2. 具体的な利用内容、計算方法

To mimic crowded protein environments, two types of systems were set up. The first system consisted of eight protein G molecules; the second system consisted of four protein G molecules and eight villin headpiece sub-domains. All three systems were solvated in explicit solvent under periodic boundary conditions. Furthermore, to examine the effect of concentration, five different box sizes were considered for each system. The protein concentrations range from 144 g/L to 619 g/L for the mixed protein G/villin system corresponding to protein crowder volume fractions between 10 and 43%. All the systems were then solvated with explicit TIP3P water molecules in cubic boxes. The initial systems were minimized and subsequently heated to 298 K. Production simulations were then continued in the NPT ensemble at a temperature of 298 K and a pressure of 1 bar. All simulations were performed using the molecular dynamics simulation package NAMD version 2.7b2. The CHARMM 27 all-atom

force field was used in combination with the CMAP correction term. Each simulation was carried out for 300 ns.

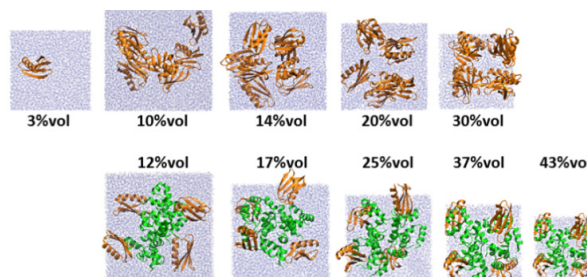


Figure 1: Simulated systems after 100 ns. Protein G is shown in brown and villin headpiece subdomain in green.

3. 結果

Available volume as a function of distance from the closest solute heavy atom was calculated to determine what fraction of water molecules is far enough from any solute to retain bulk properties. In the most crowded cases, the available volume reduces to zero at 10 Å with 30%vol protein G crowders and at 8 Å with 43%vol protein G/villin crowders. For crowder volume fractions of 30% and more, the bulk volume is reduced to below 10% while most of the available volume is found within the first solvation shell (≤ 4 Å) from the protein. This means that in the most crowded cases there is theoretically almost no room for bulk water and overall hydration properties are therefore expected to be altered significantly.

Radial distribution functions were calculated for the distances between water oxygen atoms and the nearest heavy atom of any of the protein crowder molecules. The radial distribution functions are normalized by bulk water density and the theoretically accessible volumes described above. It can be seen that the first solvation peak remains largely unaffected even under highly crowded conditions. The second solvation peak is also present in all cases but for highly crowded systems with

volume fractions of 30% and above the water density is reduced significantly relative to bulk densities with increasing distance from the solute. The reduction in bulk density is presumed to be a result of an increasing number of small solvent cavities at highly crowded conditions that are too small to accommodate water molecules. Interestingly this effect is negligible for less than 30%vol but rapidly dominates for crowder fractions of more than 30%vol. Diffusion coefficients were calculated from mean-square displacements of water oxygen atoms. Dielectric constants were calculated for water only in each of the simulations from the overall water box dipole fluctuations. Diffusion rates and dielectric constants are significantly reduced as a function of crowding shown in the following figures. The reduced dielectric constant has implications for the stability of biomolecules in crowded environments and suggests a prescription for modeling solvation in simulations of cellular environments.

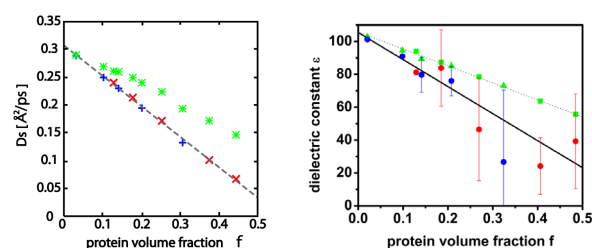


Figure 2: The linear relationships of diffusion coefficient and dielectric constant with respect to protein volume fraction.

4. まとめ

In this study, we investigated physical properties of water in the presence of protein crowders at different concentrations. We find that under highly crowded conditions, hydration properties change significantly from bulk solvent. Both, hydration structure and dynamics are altered as a function of crowding. The most dramatic change is a significant decrease in self-diffusion and dielectric response. The reduced diffusion rates are expected to affect hydrodynamic properties in cellular environments

while a reduced dielectric response alters the thermodynamics of folded proteins.

5. 今後の計画・展望

While the focus of this study has been solely on the structure and dynamics of water, future studies will investigate the effect of altered hydration properties on biomolecular solutes in more detail. The results from this work suggest a prescription for developing mean-field models of solvation in cellular environments, for example by developing implicit descriptions of cellular environments using the reduced dielectric response reported here. It is our hope that such models will facilitate physically realistic studies of biomolecular dynamics on cellular scales.

課題 2 : パスサンプリングによるタンパク質構造変化解析 (松永)

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

分子動力学計算の並列計算では、一般に長距離相互作用項を評価するための通信がボトルネックとなり、並列度を上げることは難しい。並列度を上げて、有効に超並列計算(>1024 コア)を実現するための方法として、我々はパスサンプリング(有限温度ストリング法)の開発に取り組んでいる。パスサンプリングとは、単一の分子動力学計算の代わりに、多数のレプリカをタンパク質の構造変化パス上に配置し、それらレプリカを並列に計算することで構造変化パス空間をサンプルする手法である。レプリカ間の相互作用は弱く、通信は疎であるために並列度の高いアルゴリズムである。

2. 具体的な利用内容、計算方法

これまで我々は、ストリング法を Adenylate kinase (溶媒を含めて 65,000 原子)という構造変化と機能が密接に関わる酵素へ応用し、自由エネルギーの低い物理的に尤もな構造変化パスを捉えることに成功した。また遷移状態の厳密な評価のため、遷移状態候補周りを初期条件として多数の分子動力学計算を行い、トラジェクトリが 2 つの状態に別れるか否かをテストした。

今年度は更に、幾つかの異なる初期パスを準備し、異なる初期パスに対してストリング法を適用し、同じ構造変化パスへ収束するかをテストした。そのために、初期パスの生成法として、従来の 2 構造間を線形につないだ steered MD の他に、タンパク質のドメイン部に異なる重みを用いて中間構造を生成し、それらをスプライン補間により繋いだパス上を steered MD する方法を開発した。

3. 結果

4 本の異なる初期パスに対し、ストリング法を適用した結果、そのうち 3 本は同じ特徴をもつ構造変化パスへ収束することを確認した。残りの 1 本は、違うパスへとトラップされた。現在この一本の自由エネルギープロファイルを計算中であり、他の 3 本に比べてエネルギーが高いかを調べている。

4. まとめ

有限温度ストリング法を使ったパスサンプリングの弱点として、初期パスに対する結果の依存性が挙げられる。これを解決するために我々はシステマティックに初期パスを複数生成する方法を開発した。複数の初期パスに対してストリング法を適用した結果、現在のところ、多くのパスが同じ特徴をもつ構造変化パスに収束することが確認されている。

5. 今後の計画・展望

より大きな系の、複雑な自由エネルギー地形を持つ構造変化に対しては、ひとつの構造変化パスではなく、複数の安定なパスが同時に存在することが予想される。このためには、ストリング法で得られる複数パスをまとめて解析し、実効的なエネルギーバリアを評価する統計手法が必要となる。今後はこのための方法論の構築を行う。

課題 3 : 分子動力学シミュレーションソフトウェアの開発 (森)

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

イオンチャネルやポンプなどの膜タンパク質は機能を発揮するとき、ドメイン運動や開閉運動といったダイナミックな構造変化をミリ秒～秒スケールで起こし、一方で、酵素反応における化学結合の開裂やイオン配位などの生化学反応をピコ秒～ナノ秒スケールで厳密に制御する。タンパク質の構造変化と機能の相関を調べる手法として、広く分子動力学シミュレーションが行われてきた。しかしながら、従来の全原子モデルのみを用いた方法では、せいぜいマイクロ秒オーダーの計算が限界であり、生体反応スケールのミリ秒には遠く及ばない。このような問題を解決するためには、マルチスケールなアプローチが必要であるが、既存の分子動力学シミュレーションソフトウェアではマルチスケールな現象を扱うことができない。

我々のグループでは、新規分子動力学計算プログラム GENESIS を開発している。本プログラムは CHARMM 力場を用いた生体系の分子動力学計算が行うことができるソフトウェアで、「京」での利用を視野に入れて高並列化などのチューニングを行っている。

本研究では、マルチスケール分子動力学シミュレーションのプラットフォームとなる GENESIS を開発し、粗視化モデルと全原子モデルを連成させる新しいシミュレーション法の開発を目指す。

2. 具体的な利用内容、計算方法

生体分子系の全原子シミュレーションで一般的に用いられているアルゴリズムを GENESIS に導入した。導入したアルゴリズムは以下の通りである。

- 結合拘束法 (SHAKE, SETTLE, LINCS 法)
- 温度・圧力制御 (Langevin, Berendsen, Nose-Hoover 法)
- NPT アンサンブル (isotropic, anisotropic, semi-isotropic), NPAT アンサンブル
- レプリカ交換分子動力学法
- 粗視化分子モデル

導入したアルゴリズムが正しく動作するかどうかを、CHARMM や NAMD などの他の分子動力学計算ソフトウェアから得られた結果と比較しながらチェックを行った。本テスト計算で扱った系は POPE 脂質二重膜で、原子数は 14168 個からなり、64 個の POPE 脂質分子と約 2000 個の水分子が含まれる (Figure 3)。シミュレーション時間は 50 ns (1step = 2 fs)、静電相互作用計算には本研究グループの今井が導入したパーティクルメッシュエwald法、力場には CHARMM36、温度・圧力制御には Langevin thermostat, Langevin position 法を用い、NPT (1atm, 310K) および NPAT (1atm, 310K) アンサンブル下で計算を行った。GENESIS の並列計算には 128CPU(32MPI/4OpenMP)、NAMD の並列計算には 64CPU(64MPI)を用いた。

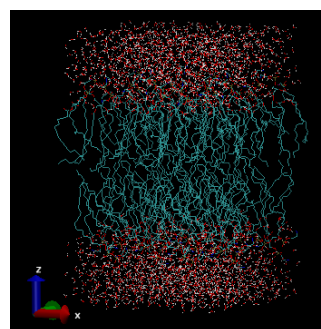


Figure 3. POPE 脂質二重膜

3. 結果

Figure 4 に NPT 条件下での脂質一分子あたりの表面積の時間変化を、同じ条件下で行った GENESIS、NAMD の結果と実験値との比較を示す。GENESIS で得られた結果は、実験データをよく再現している。また、NAMD との比較に関しても、トラジェクトリーは一致していないが、非常によく似た挙動を示している。GENESIS を用いて信頼性の高い膜タンパク質系のシミュレーションを行うことが出来ると期待される。

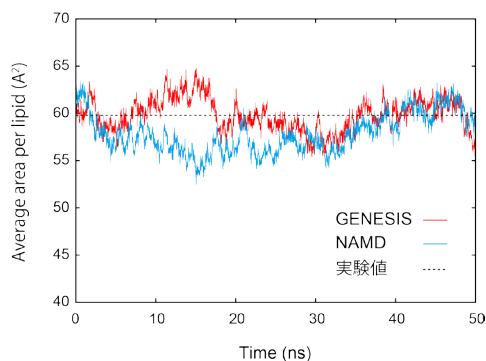


Figure 4. POPE 脂質一分子あたりの時間変化

4. まとめ

本研究室で開発中の分子動力学シミュレーションソフトウェア GENESIS に、膜タンパク質系のシミュレーションでよく用いられているアルゴリズムを導入し、脂質二重膜を対象にしてプログラムが正しく動作することを確認した。さらにマルチスケール分子動力学シミュレーション開発のために粗視化分子モデルおよびレプリカ交換分子動力学法を GENESIS に導入した。

5. 今後の計画・展望

膜タンパク質や脂質二重膜のような大規模系のシミュレーションでは、平衡状態に達した後、注目する物理量の平均値を得られるまでに計算時間が膨大にかかるため、64CPU 以上を用いても一ヶ月以上を要する場合がある。今後はより高速な並列化アルゴリズムの導入や、今年度導入した粗視化モデルと全原子モデルを連成させる新しいマルチスケールシミュレーション法の開発を行う。

6. これまで利用した状況、継続して利用する際に行う具体的な内容

今年度は、GENESIS を信頼性の高いソフトウェアに仕上げるという観点から、並列化効率のチェックや

並列計算が正しく行われているかどうかのテスト計算のために RICC を用いてきた。今後は GENESIS を用いて膜タンパク質系のシミュレーションのプロダクションランを行う予定である。

課題 4 : Development of parallelization for Molecular dynamics (Jung)

1. Background and purpose of the project, relationship of the project with other projects

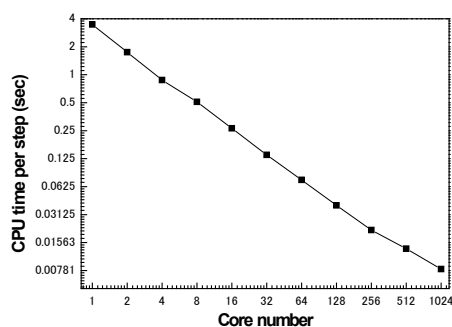
My past research was to develop new QM/MM methods for biomolecules and to introduce parallelization in QM/MM. Current purpose is the development of the efficient parallelization scheme for molecular dynamics program spatial decomposition scheme and the understanding of enzyme reaction using QM/MM.

2. Specific usage status of the system and calculation method

Our scheme of parallelization of MD uses the so-called midpoint method in which two particles interact on a particular box if and only if the midpoint of the segment connecting them falls within the region of space associated with that box. We also use a modified constrained optimization with QM/MM where both Hartree Fock and MP2 are available as QM.

3. Result

With the newly developed path optimization method, we investigated the phosphorylation reaction by cAMP-dependent protein kinase (PKA). We found that the transition state is formed before the movement of proton. We confirmed our method by the accurate energy barrier compared with the experimental result. As for the parallelization of the molecular dynamics, we confirmed the parallel efficiency using RICC clusters. The CPU speed result of seca proteins with cutoff is as following figure:



4. Conclusion

With the recently developed method (QM/MM with the modified constrained optimization), we investigated the phosphorylation reaction by PKA. We also developed parallelization using spatial decomposition scheme on molecular dynamics.

5. Schedule and prospect for the future

Currently, we finish the development of the new domain decomposition with short range. This year, we'll include long range interaction in our method with better parallel efficiency. In addition, we'll investigate many conformational changes of proteins with our existed program. We also finish the reaction path optimization with QM/MM and we'll find out the main enzyme function related with phosphorylation reaction.

6. If you wish to extend your account, provide usage situation (how far you have achieved, what calculation you have completed and what is yet to be done) and what you will do specifically in the next usage term.

I finished the enzyme reaction calculation of PKA at the Hartree Fock level and finished the implementation of parallelization with cutoff. The QM/MM calculation at the MP2 level and parallel implementation with the long range interactions are remained.

7. If you have a "General User" account and could not complete your allocated computation time, specify the reason.

I expected that I could complete the allocated computation time if I continue the computation every day. However, in many times, I should wait for long time until other group finish the calculation and I should have used other resources for those times.

8. If no research achievement was made, specify the reason.

Research achievement was made enough.

課題 5 : レプリカ交換インターフェースプログラム (REIN)の開発 (宮下)

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

レプリカ交換分子動力学法(REMD)はタンパク質の構造予測や構造変化予測、結合予測等を行う際によく用いる方法として知られている。それを発展させた多次元レプリカ交換分子動力学法は大規模なタンパク質の構造予測や構造変化予測には有効な方法であるが、必要なレプリカ数が膨大になることから非常に多くのリソースが必要となる。しかし、京コンピュータの様な超並列コンピュータが今後のスーパーコンピュータの主流になりつつあるので、将来性のある有望な方法であると考えられる。そこで、我々は、京コンピュータと PC クラスタで動く多次元レプリカ交換分子動力学プログラムを、次世代計算科学プロジェクトのプログラムの一つとして開発している。RICC では開発の為の大規模テスト計算などを行っている。

2. 具体的な利用内容、計算方法

モデル系(apoa1 や alanine dipeptide など)の大規模テスト計算をさまざまな条件で行い、その結果を REIN の開発に反映させている。

3. 結果

まだ成果物である REIN の論文はでていない。REIN のアプリケーションとして糖鎖の構造予測の論文[8]がある。

4. まとめ

今年は大規模並列計算用の開発サポート用に RICC

平成 23 年度 RICC 利用報告書

を使用した。予定では今年度は REIN の京コンピュータ上での product run の為のテスト計算で使う予定であったが、開発が行き着かなかった。この理由はリカーシブ MPI が京コンピュータでサポートされないという事と、京コンピュータのシステム側から、MPI_COMM_SPAWN を用いて欲しいとの事であったので、今年度後半は京コンピュータにあわせた開発を行っていた。

5. 今後の計画・展望

来年度早々(4月)には、今年度予定していた膜中のアミロイドの全原子モデルの REIN を用いたテスト計算を RICC にて行いたい。また、要望として、RICC で現在サポートされていない MPI_COMM_SPAWN のサポートをお願いしたい。

6. RICC の継続利用を希望の場合は、これまで利用した状況(どの程度研究が進んだか、研究においてどこまで計算出来て、何が出来ていないか)や、継続して利用する際に行う具体的な内容

京コンピュータ用のプログラムはほぼ完成した。また、本来、今年度後半に行う予定であった、京コンピュータで行う product run の大規模テスト計算の系のモデリングも今年度中に終わる予定である。

今年度行う予定であった、大規模テスト計算を、来年度早々に行いたいと考えている。具体的にはアミロイド前駆体蛋白質(APP)と脂質分子との相互作用に関する研究である。脂質分子の成分を変え、多成分膜を作成し、単一成分膜中の APP との相互作用の相違、および構造の相違に関する研究を行いたいと考えている。

7. 一般利用で演算時間を使い切れなかった理由

京コンピュータの為のプログラム開発が遅れたため。

8. 利用研究成果が無かった場合の理由

現在、プログラムの開発中であるため。

平成 23 年度 RICC 利用研究成果リスト

【論文、学会報告・雑誌などの論文発表】

- 1) M. Feig and Y. Sugita, “Variable Interactions between Protein Crowders and Biomolecular Solutes Are Important in Understanding Cellular Crowding”, *J. Phys. Chem. B*, **116**, 599-605 (2012).
- 2) R. Harada, Y. Sugita, and M. Feig, “Protein Crowding Affects Hydration Structure and Dynamics”, *J. Am. Chem. Soc.*, (in press).
- 3) Takaharu Mori, Fumiko Ogushi, and Yuji Sugita, “Analysis of lipid surface area in protein-membrane systems combining Voronoi tessellation and Monte Carlo integration methods”, *J. Comput. Chem.*, **33**, 286-293 (2012).
- 4) Mizuki Morita, A. V. S. K. Krishna Mohan, Shandar Ahmad, Takaharu Mori, Yuji Sugita, and Kenji Mizuguchi, "Lipid recognition propensities of amino acids in membrane proteins from atomic resolution data", *BMC Biophysics*, (in press).
- 5) 森貴治, 杉田有治, “膜タンパク質が触媒する物質輸送現象の分子メカニズム(<小特集>計算化学)”, シミュレーション, **30** (2011).
- 6) S. Fuchigami, H. Fujisaki, Y. Matsunaga, A. Kidera, "Protein Functional Motions: Basic Concepts and Computational Methodologies", *Advances in Chemical Physics* **145**, 35 (2011)
- 7) D.M. Leitner, Y. Matsunaga, C.B. Li, T. Komatsuzaki, A. Shojiguchi, and M. Toda, "Non-Brownian Phase Space Dynamics of Molecules, the Nature of their Vibrational States, and non-RRKM Kinetics", *Advances in Chemical Physics* **145**, 83 (2011)
- 8) Suyong Re, Naoyuki Miyashita, Yoshiki Yamaguchi, and Yuji Sugita, "Structural Diversity and Changes in Conformational Equilibria of Biantennary Complex-Type N-Glycans in Water Revealed by Replica-Exchange Molecular Dynamics Simulation", *Biophysical Journal* (Biophysical Letter) **101** L44–L46 (2011).

【国際会議、学会などでの口頭発表】

- 1) 森貴治 『分子シミュレーションによる膜タンパク質のダイナミクス解析』 CBI/JSBi2011 合同大会(神戸) Nov. 2011
- 2) 森貴治 “Molecular dynamics simulations for the protein secretory pathway”, 第 49 回日本生物物理学会年会(姫路) Sep. 2011
- 3) 森貴治、小串典子、杉田有治, 『膜タンパク質-脂質二重膜系における脂質 1 分子表面積解析法の開発』 日本物理学会 2011 年秋季大会(富山) 2011/9
- 4) 宮下尚之 2012/01/31 先端融合科学シンポジウム『タンパク質アセンブリ ー会合、超分子化、凝集ー』 “APP の膜近傍領域の構造と膜中での会合(二量体)構造予測”, 神戸
- 5) Naoyuki Miyashita, 2011/06 Korea-Japan Symposium on Statistical Mechanics Approaches to Nano/Bio-Sciences, “The development of multi-dimensional replica-exchange interface (REIN) software and the estimation of simulation time of REMD”, Korea

【その他（ポスター発表）】

- 1) Takaharu Mori, Fumiko Ogushi, and Yuji Sugita, “Analysis of the area per lipid in protein-membrane systems”, Biophysical Society 56th Annual Meeting, San Diego (USA), 2012/2
- 2) Y. Matsunaga et al., “Conformational Transition Pathways of Adenylate Kinase Explored by the String Method”, Biophysical Society 56th Annual Meeting, Feb. 2012, San Diego, USA
- 3) N. Miyashita, S. Re and Y. Sugita, “REIN: Replica-exchange interface program”, 2nd AICS International Symposium, Mar 1-2, 2012, Kobe
- 4) 宮下尚之、李秀栄、杉田有治 “レプリカ交換インターフェースプログラム”, 文部科学省「革新的ハイパフォーマンス・コンピューティング・インフラ（HPCI）の構築」次世代ナノ統合シミュレーションソフトウェアの研究開発（ナノ）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（ライフ）公開シンポジウム, 3/5-6, 2012, 神戸
- 5) 宮下尚之、李秀栄、杉田有治 “Plutypus-REIN:レプリカ交換インターフェース”, ISLiM 成果報告会 2011, 12/21-22, 東京