

## ○課題名：生命科学上の非構造化データの統合マイニング

代表研究者：馬見塚拓（京都大学化学研究所）

### S-3

#### 生命科学上の非構造化データの統合マイニング

馬見塚拓<sup>1</sup>、志賀元紀<sup>1</sup>、茅野光範<sup>1</sup>、瀧川一学<sup>1</sup>、津田宏治<sup>2</sup>

(1 京都大学化学研究所、2 産総研生命情報工学研究センター)

ビッグサイエンスの潮流と実験技術の進歩により、生命科学では、多種多様なデータが大量に蓄積されデータからの知識獲得が重要な研究課題となっている。このデータには代謝パスウェイや遺伝子配列といった「非構造化データ」が豊富であり扱いが容易でない。実際、近年の計算機科学でも非構造化データからの効率的かつ高精度知識発見が重要研究課題であり、より一般に重要である。さらに同じ情報源からの形式の異なる複数データの存在が生命科学の特徴である。このような背景の下、本研究課題では多種多様なデータを組み合わせ効率的かつ体系的な知識発見を可能にする技術の構築を行っている。本発表では特に2つの構築技術を取り上げる：1) クラスタリングは生命科学データの解析で標準的に使われる手法であるが、典型的な非構造化データのグラフと代表的な構造化データの表を組み合わせ、既存手法の性能を凌駕する効率的クラスタリング手法。2) 遺伝子型データと発現データを組み合わせ3つの異なる遺伝子（2つはSNP）間の相互作用を既存手法の10倍程度高速に検出する手法。これら手法により生命科学上の新規知識発見の効率的かつ高速な実現が期待される。

## ○課題名：ダイナミクスを考慮した膜蛋白質の構造モデリング法の開発

代表研究者：水口賢司（(独)医薬基盤研究所）

### S-4

#### 膜タンパク質のダイナミクスとタンパク質—脂質間相互作用の解析

森田瑞樹<sup>1</sup>、Krishna Mohan<sup>1</sup>、森貴治<sup>2</sup>、Shandar Ahmad<sup>1</sup>、森博幸<sup>3</sup>、岡本祐幸<sup>4</sup>、杉田有治<sup>2</sup>、水口賢司<sup>1</sup>

(1 医薬基盤研究所、2 理化学研究所、3 京都大学ウイルス研究所、4 名古屋大学大学院理学研究科)

本研究では、バイオインフォマティクス的手法と分子シミュレーションを組み合わせ、膜タンパク質のダイナミクスの理解および予測手法の確立を目指している。まず、脂質二重膜中での膜蛋白質の構造安定化機構の解明に向け、陰的溶媒モデルに基づいて短時間で膜中での膜タンパク質の配向を予測する手法の開発を行った。一方、分子動力学シミュレーションから得られたデータとX線結晶構造から、膜タンパク質—脂質間相互作用におけるアミノ酸傾向性を求めた。次に、大規模な動的構造変化の理解に向け、1) 膜タンパク質の構造特徴（溶媒露出表面積やヘリックスの曲がり具合）の同時予測およびそれらの予測誤差と分子動力学および基準振動解析に見られる膜タンパク質のダイナミクスとの関係の解析と、2) 現在解かれているX線結晶構造に見られる膜タンパク質のダイナミクスの抽出と解析を行なった。また、これらの研究成果を他の研究者が利用できるようにするために、統一したインターフェースを持つ統合Webサービスの構築を行っている。