公募研究：2005 ～ 2009 年度

タンパク質相互作用のネットワーク予測から原子レベルの結合予測までの統合的研究

●清水 謙多郎 [1)2)]　◆角越 和也 [1)]　◆寺田 透 [1)]　◆中村 周吾 [1)]
1) 東京大学大学院農学生命科学研究科 2) 東京大学大学院情報理工学系研究科

＜研究の目的と進め方＞

タンパク質間相互作用は様々な生命現象の鍵を握る重要な反応であるが、生化学的解析や立体構造解析には複雑な手順・高額な装置・研究者の熟練等が必要であり、膨大な数の相互作用の全てをそれらの手法で解析するのは不可能である。本研究では、タンパク質間相互作用について、以下の 4 つのアプローチから予測・解析手法の開発を行う。(1) タンパク質 - タンパク質間相互作用予測および相互作用ネットワーク予測、(2) タンパク質 - タンパク質、リガンド、糖鎖、DNA 間相互作用部位予測、(3) タンパク質 - タンパク質間ドッキング予測（複合体構造予測）、(4) タンパク質と他の分子の間の物理的な相互作用解析。(1) については、配列情報のみから機械学習を利用して予測する手法を開発する。(2) については、配列情報のみから予測する手法と、タンパク質の構造が既知の場合は、さらに精度の高い予測ができるよう、配列情報と構造情報の両方を利用して予測する手法の 2 つを開発する。(3) については、球面調和関数と直交動径関数による関数展開を併用した高速ドッキングアルゴリズムを開発し、さらに構造精密化のため手法を開発する。(4) については、溶媒効果を採り入れた分子動力学シミュレーションにより、原子レベルの詳細な解析を行えるようにする。以上、(1) ～ (4) を実現する統合化されたシステムを構築する。

＜研究開始時の研究計画＞

タンパク質 - タンパク質間相互作用予測については、配列情報のみから Support Vector Machine（SVM）を利用して予測する手法を開発し、学習させる配列特徴、カーネル関数について詳しく検討する。タンパク質 - タンパク質間相互作用部位予測では、タンパク質を構成する各アミノ酸残基が相互作用するかどうかを予測するというもので、SVM を利用して、配列情報のみから予測する手法と、構造既知のタンパク質が利用できる場合は、さらに精度の高い予測ができるよう、配列情報と構造情報の両方を利用して予測する手法を開発する。配列・構造特徴の抽出法、カーネル関数の改良を行い、構造情報を利用する場合は、隣接残基のプロファイル、溶媒露出表面積、凹凸のパターンなどの構造特徴を学習させる手法を開発する。そのほか、Support Vector Regression（SVR）を利用し、各相互作用残基の周辺残基数を予測する手法を開発する。これにより、相互作用部位の詳細な構成や特性を予測・解析できるようにする。また、本研究では、タンパク質 - タンパク質の相互作用部位予測に加えて、タンパク質 - 糖鎖、タンパク質 -DNA 相互作用部位予測の手法を開発する。これらの相互作用部位の配列・構造上の特徴を調査し、それをもとに、SVM を用いて、これらの相互作用部位を統一的な手法で予測することを目指す。タンパク質 - タンパク質間ドッキング予測については、球面調和関数と直交動径関数による関数展開を併用した高速ドッキングアルゴリズムの開発を行う。2 つのタンパク質の相互作用エネルギーの計算では、タンパク質の形状だけでなく、残基間の経験的ペアポテンシャルや van der Waals ポテンシャルなどを取り入れる予定であり、それらの効果を詳細に解析する。また、予測精度をさらに精密化するための手法（相互作用エネルギーの再計算、予測構造のクラスタリング、側鎖モデリングなど）を開発する。本研究では、また、溶媒分子の存在下での精密なモデリングおよび詳細な相互作用の解析を行うための分子動力学シミュレーション手法の開発を行う。以上の予測手法を実現する統合化されたシステムを構築する。

＜研究期間の成果＞

1. タンパク質-タンパク質間相互作用予測

与えられた 2 つのタンパク質が相互作用するかどうかを、アミノ酸配列情報のみから機械学習 SVM を用いて学習・予測する手法を開発した。配列特徴としては、(1) アミノ酸の隣接ペアの出現頻度（400 × 2 次元）、(2) (1) の 3 つ組の出現頻度（8000 × 2 次元）、(3) アミノ酸の物理化学特性に基づく分類の隣接ペアの出現頻度（49 × 2 次元）（具体的には、側鎖の双極子と体積をもとに 7 つに分類、側鎖の双極子は、密度汎関数理論 B3LYP/6-31G*、GAUSSIAN03 に組み込みの機能を使用して計算、側鎖の体積は Sybyl6.8 を使用）、(4) (3) の 3 つ組の出現頻度（343 × 2 次元）の 4 通りを試したところ、(2) が最も良い結果を示した。カーネル関数については、タンパク質のペア *A, B* が相互作用するかどうかを判定するとき、相互作用するタンパク質 *C, D* がすでにわかっているとすると、*A* と *C*、*B* と *D*、または *A* と *D*、*C* と *B* が類似した配列パターンをもつとき、これらが相互作用すると判定するような関数を新たに開発した。データセットとして HPRD（Human Protein Interaction Database）に登録されたタンパク質のペアを用い、5-fold cross validation で評価した結果、AUC 0.885、MCC（Matthews correlation coefficient）0.625 となり、Martin らの結果（AUC 0.862、MCC 0.581）、Shen らの結果（AUC 0.825、MCC 0.526）に比べて高い性能を得た。

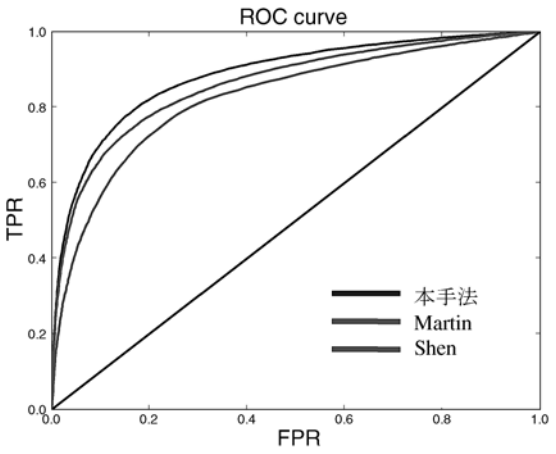

図1 タンパク質-タンパク質間相互作用予測の性能

図1に、本課題で開発した手法、Martinらの手法、Shenらの手法のROC曲線を示す。この図からも、我々の手法が最も良い結果を得ていることがわかる。(なお、Martinらの手法、Shenらの手法は、それぞれの論文で主張している複数の手法の中で、最良のものを示す。)また、HPRDより取得したMAP kinase (Ras-Raf1-MEK1-ERK1-ELK1-SRF) のタンパク質間相互作用ネットワークを対象に予測を行った結果、372個の相互作用のうち347個 (93.3%) を予測することができた。

2.タンパク質間相互作用部位予測

(a) アミノ酸配列情報のみを用いた予測

アミノ酸配列情報のみを用いた予測では、予測対象のタンパク質の配列を、非冗長データベースに対してPSI-BLASTにより検索をかけ、類似の配列のマルチプルアラインメントを求め、これをもとにプロファイルを作成し、さらにSVMの入力として予測を行う手法を開発した。カーネル関数としては、Radial Basis Function (RBF) を用いた。予測に用いる特徴量としては、残基の出現頻度とPSSM(位置特異的スコア行列)の2つを検討した。残基の出現頻度とPSSMからSVMの特徴ベクトルを作る際に、予測対象となっている残基のみでベクトルを構成するのではなく、予測対象の残基を中心とした配列上近隣の11残基分のデータを結合してベクトルを構成した。また、SVMを2段で適用する手法 (1段目で得られた結果をさらに2段目のSVMの入力とする手法) を開発した。これは、相互作用部位が配列上連続していることを利用して、孤立して相互作用部位と予測された結果を修正することを意図した手法である。

結果は、特徴量として、残基の出現頻度を利用した場合とPSSMを利用した場合でのRecallが、それぞれ、53.2%、62.3%で、PSSMを用いた場合の予測性能の方が高かった (Precisionは30.0%に揃えた)。また、SVMの段数については、1段の場合は53.2%、2段の場合は54.2%と性能差は小さかった。

(b) アミノ酸配列情報と既知の構造情報を用いた予測

構造情報の利用については、分子表面上、隣接しているアミノ酸残基に対して、構造類似のタンパク質のプロファイルを作成し、これらの残基を構成する疎水性・非疎水性原子の溶媒露出表面積と合わせてSVMの入力とし、予測を行う手法を開発した。予測に用いる特徴量としては、配列情報としては残基の出現頻度 (freq) とPSSM、構造情報としては残基単位の溶媒露出度 (rASA) と、残基内の極性原子および非極性原子の溶媒露出表面積 (aASA) を検討した。特徴ベクトルの作成については、予測対象の残基と空間的に近隣にある残基、計15残基分のデータを利用した。また、1段と2段のSVMの適用も検討した。

予測に用いたデータセットおよび相互作用部位の定義は、上と同じであり、結果は、残基の出現頻度とPSSMを比較した場合、Recallは、それぞれ、71.4%、66.2%とPSSMを用いた場合の予測性能が高かった。また、残基単位の溶媒露出度と、残基の極性原子および非極性原子の溶媒露出表面積の比較では、66.2%、68.7%と、後者の方が高かった。また、1段と2段のSVMの比較では、66.2%、69.2%と、配列情報のみを利用した予測よりも大幅な性能向上が見られた。

また、我々は、SVR (Support Vector Regression) を用いて、各残基の周辺の相互作用残基数を予測する手法を新たに開発した。これは、注目している残基が相互作用部位のどの程度中心にあるかを予測することを意図したものである。その手順は以下の通りである。ターゲットのタンパク質の配列に対し、PSI-BLASTにより、類似の配列のマルチプルアラインメントを求め、プロファイルを作成する。それと合わせて、各表面残基の極性・非極性原子の溶媒露出表面積を計算し、残基ごとに、空間的に隣接する14残基を加えた15残基分のプロファイルと溶媒露出表面積を取り出して機械学習SVRの入力とする。

予測に用いたデータセットは、配列一致度30%で冗長性を除いた168個のタンパク質からなるデータセットで、複合体の構造はProtein Quaternary Structure file sever (PQS) から取得した。予測結果を5-fold cross validationで評価したところ、全体の相関係数は0.59であった。

構造既知のタンパク質について、リガンドが結合する空間上の位置を予測する手法を開発した。本手法は、タンパク質表面にメタン分子を格子状にプローブさせ、タンパク質分子とのvan der Waals相互作用エネルギーを計算するというものである。プローブの生成はdouble cubic lattice method (DCLM) により、力場パラメータとしてはAmber parm94を使用した。エネルギー値が小さいものをクラスタリングし、さらにそれをseedとして、より緩いエネルギー値の条件でクラスタを広げるという手法を用いた。35個のタンパク質 - リガンド複合体 (bound) 構造と、35個の単体のタンパク質 (unbound) 構造からなるLaurieとJacksonのデータセットを使用し、予測を行った結果を表1に示す。

表1に示すように、PocketFinderやQ-siteFinderなど現在広く用いられている手法より高い精度で予測でき、とくにunbound予測における予測精度の向上が大きいという結果を得ている。**図2**は、ストレプタビディン (PDB ID: 2RTA) のリガンド結合部位予測の例を示したものである。予測順位1位の部位が黄色で、順位が下がるにつれ青色に近づく。黄色の部位が実際のリガンド (ビオチン) の位置と一致していることがわかる。

表1 タンパク質-リガンド相互作用部位予測の結果

		1位の予測部位	3位以内の予測部位	平均precision
我々の手法	Bound	0.800	1.000	0.839
	Unbound	0.743	0.857	0.771
Q-SiteFinder	Bound	0.743	0.943	0.739
	Unbound	0.514	0.829	0.619
Pocket-Finder	Bound	0.714	0.771	0.375
	Unbound	0.514	0.657	0.354

Precisionは、予測部位と実際のリガンドの存在部位が一致している割合を示す。予測順位が1位のものと、3位以内のものについて、precision ≥ 0.25 (25%以上、予測部位と実際のリガンドの存在部位が重なっているものの割合) の結果を示す。

図2 リガンド結合部位予測の例

図中のドットは、リガンドの存在位置を予想したもの。黄色に近いほど予測順位が高いことを示す。

また、上記研究開発に関連して、リガンド結合状態・非結合状態のタンパク質のデータベースBUDDY-systemを開発し、以下のURLでサービスを一般に公開している (**図3**)。本データベースは、タンパク質、リガンドの単体からそれらの複合体、逆に複合体から単体を検索する機能をもち、リガンドに対する条件を細かく設定することが可能である。すでに16,203個の結合状態 (bound) と非結合状態 (unbound) の対を登録しており (平成21年10月現在)、現在は、結合状態と非結合状態の構造変化と

ダイナミクス、Biological Unit を考慮した結合残基と相互作用の詳細、リガンド周辺の missing residue の解析結果などを登録した新しいデータベースを開発している。

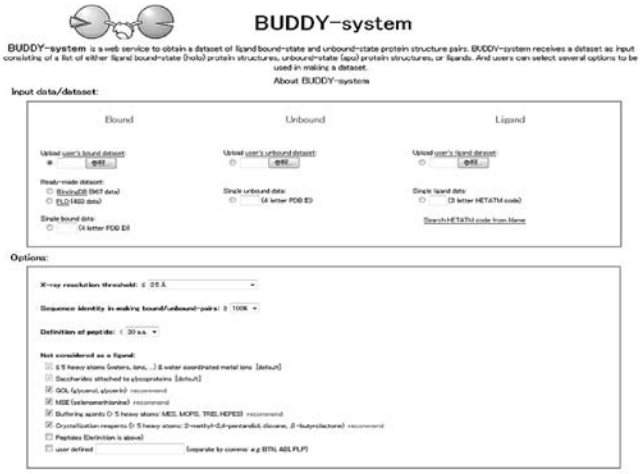

図3　タンパク質 - リガンド複合体データベース http://www.bi.a.u-tokyo.ac.jp/services/buddy/current/index.cgi

3.タンパク質間ドッキング予測

ドッキングアルゴリズムとして、球面調和関数と新規に設計した正規直交基底関数での級数展開による高速内積計算を使ったアルゴリズムを開発した。ドッキングシミュレーションでは、タンパク質の相互作用を表すスコア関数を、各分子から定義されるスカラー場の関数 f, g の内積の線形和

$$c(\mathrm{T}) = \sum_i w_i \int f_i(\boldsymbol{x}) g_i^{\mathrm{T}}(\boldsymbol{x}) d\boldsymbol{x}$$

として表し、これを最小にする変換関数Tを求める。各コンフォメーションにおいてこのスコア関数の値を計算し、その値が低いものから順に、上位のものを候補コンフォメーションとする。各スカラー場は、その内積が、表現したいエネルギーもしくは性質を反映するように柔軟に定義でき、例えば、分子形状の相補性や各種ペアポテンシャル、静電相互作用などを表現することが可能である。また、本手法では、スカラー場を上記正規直交基底関数で展開することにより、スコア関数の計算に必要な内積計算を高速に行うとともに、配座空間の探索に必要な座標変換操作も高速に行えることを示した。本課題では、展開係数によるスカラー場の表現能力が、中心からの距離rの増加に従って劣化するという、球面調和関数に基づく基底関数を用いた方式の問題点を解決するため、とくに分子の表現空間を階層的に定義し、それぞれの階層において異なる動径基底関数を適用する手法を新たに開発した。これにより、比較的少数の係数でスカラー場を効率的に表現することができるようになり、我々が調べたタンパク質の約7割で予測精度を改善することができた。

ドッキングによる相互作用エネルギーの評価には、原子レベルの経験的ポテンシャルである ACE（Atomic Contact Energy）をスコア関数に用いた。また、ACE には、立体障害の効果が入っていないため、原子の立体障害を表すポテンシャル関数を新たに導入した。これは、原子の van der Waals 半径内にあると正の値をとり、表面の原子に対する値は小さくするという工夫を取り入れたもので、原子間の衝突を「ソフトに」避けるのに効果がある。

表 2 は、本手法の予測精度と計算時間を、FFT を用いた手法で現在広く利用されている FTDock と比較して示したものである。本手法は FTDock と比較して、同程度の精度で予測するのに、16 倍から 160 倍以上の高速化を達成した。

予測順位は、ネイティブに近い構造（ネイティブとの I-RMSD 値が 2.5Å 以内の構造）が出現する予測順位を表し、上位 4000 個中の予測数は、スコア上位 4000 個の中のネイティブに近い構造の数を表す。

我々は、また、候補構造から予測構造を絞り込む手法を新たに開発した。ドッキング予測では、まず剛体モデルを用いた 6 次元の自由度において粗い計算により全空間探索を行い、その後、そこで得られた上位候補について、さらに詳細な計算を行って候補構造を精密化する。精密化段階については、より詳細な物理化学的なエネルギーを使ってスコアを再計算した。具体的な手法は以下の通りである。まず、上位の候補構造に水素原子を付加し、その複合体構造について van der Waals、静電相互作用、脱溶媒和自由エネルギーの項を計算し、それらの線形和をスコア関数として再計算する。また、各エネルギーの重みは、正解が分かっている訓練データを用いて Downhill simplex minimization によって最適化を行う。表 3 に、この精密化の効果を示す。ネイティブに近い構造（Interface RMSD（I-RMSD）値が 3Å 以内の構造）が最初に現れる順位が向上していることがわかる。

表2 タンパク質-タンパク質ドッキング予測の結果

複合体構造（単体構造）	予測順位（I-RMSD値）		上位4000個中の予測数		計算時間 [分]	
	本手法	FTDock	本手法	FTDock	本手法	FTDock
1UGH (1AKZ+1UGI(A))	1 (2.02)		18		0.61	101
1BRB (1BRA+6PTI)	78 (1.86)		47		1.76	29
2SIC (1SUP+3SSI)	58 (1.67)	NA	6	0	1.91	223
2PTC (3PTN+6PTI)	33 (2.30)	502	32	8	1.76	37
1CHO (5CHA(A)+2OVO)	1697 (2.24)	127	12	7	0.63	33
2KAI (2PKA(AB)+6PTI)	24 (2.04)	223	37	25	1.78	34

表3　タンパク質-タンパク質ドッキング予測の精密化の結果

複合体構造（単体構造）	予測順位（I-RMSD値）		上位4000個中の予測数	
	本手法	精密化後の順位	本手法	精密化後の順位
1UGH (1AKZ+1UGI(A))	1 (2.02)	34 (2.09)	1 (2.02)	1 (2.70)
1BRB (1BRA+6PTI)	78 (1.86)	280 (1.75)	30 (2.57)	50 (2.81)
2SIC (1SUP+3SSI)	58 (1.67)	83 (2.34)	58 (1.67)	83 (2.34)
2PTC (3PTN+6PTI)	33 (2.30)	2 (2.22)	10 (2.78)	2 (2.22)
1CHO (5CHA(A)+2OVO)	1697 (2.24)	517 (2.09)	750 (2.73)	366 (2.55)
2KAI (2PKA(AB)+6PTI)	24 (2.04)	6 (2.47)	24 (2.04)	6 (2.47)

4.物理的な相互作用解析

タンパク質と他の分子との間の物理的な相互作用解析については、分子シミュレーションに関する基盤技術の開発と、実際の系を対象にした解析研究の 2 つのアプローチで研究を行った。

ab initio 分子動力学（MD）とマルチカノニカル MD を統合した手法を新たに開発し、相互作用部位における化学反応の動的な

解析を可能にする基盤技術を開発した。*ab initio* MD シミュレーションは、分子の電子状態を明示的に考慮し、原子核に働く力を *ab initio* 量子化学計算から求めるというもので、化学反応を扱うことができ、従来の古典的 MD シミュレーションより原子間相互作用をより精確に計算できる可能性をもっている。本課題では、*ab initio* MD とマルチカノニカル MD を統合した手法を新たに開発し、ペプチドに適用した。古典的な定温 MD に比べて、コンフォメーションのサンプリング効率が高いことを示すとともに、ペプチドの自由エネルギー曲面を計算し、古典的なマルチカノニカル MD で得られた結果を比較した。

また、複合体モデリングの精密化のための基盤研究として、タンパク質立体構造モデルの精密化に関する研究を行った。本課題では、立体構造が多く決定されている SH2 ドメインの一つである human p56 lck（PDB ID:1LKK）の構造をもとに、同じく SH2 ドメインの、Xlp SAP の構造の予測を行った。SH2 ドメインは 100 アミノ酸残基程度からなり燐酸化チロシンを含んだペプチドを認識する機能を持つが、ペプチドを認識するループ部分がホモログ間で大きく変化している。とくに Xlp SAP は、human p56lck と比較して、ペプチド認識ループに 10 残基もの挿入配列があるため、既存の比較モデリングのみでは構造予測は困難である。本課題では、マルチカノニカル分子動力学計算を用いて、高精度モデリングを試み、結晶構造（PDB ID:1D4W）に非常に近い構造群を得ることができた。

＜国内外での成果の位置づけ＞

相互作用予測については、配列情報のみを用いた手法としては、従来の主要な手法に比べて高い予測性能を達成しており、さらにドメイン情報を用いた予測手法を合わせて用いることにより、さらなる予測精度の向上が期待される。SVR を用いて各残基の周辺の相互作用残基数を予測する手法は、これまでにない新しい手法であり、相互作用部位予測の精度向上だけでなく、その成果は、ホットスポットとの関係など、相互作用部位の性質を解析する上で重要と考えられる。タンパク質 - リガンド結合部位予測手法の開発では、Pocket Finder や Q-site Finder など現在広く用いられている手法より高い精度で予測できることを実証しており、とくに unbound 予測における予測精度の向上が大きいことを示した。ドッキングによる複合体構造予測については、FTDock など、FFT を用いた手法がよく利用されているが、我々の手法は、それらに対し 160 倍から 1700 倍の性能向上を達成した。ドッキング予測に関しては、これまでに国際会議 3 件、その他学会等で 2 件の招待講演を行った。

＜達成できなかったこと、予想外の困難、その理由＞

ほぼ計画通りの成果が得られており、糖鎖結合タンパク質予測（レクチン予測）は、当初の計画にない成果である。ドッキング予測については、候補構造の選択を行う手法を開発し、一定の効果を挙げたが、ドッキングでは考慮すべき手法およびパラメータの数が膨大で、剛体レベルでの精度向上に多くの時間と労力を要してしまい、側鎖モデリングなどを取り入れたフレキシブルドッキングの実現には至らなかった。本件については、現在、実現に向けた作業を行っている。

＜今後の課題、展望＞

タンパク質間相互作用予測は、バイオインフォマティクスにおける重要な手法であるが、本研究において、従来の手法よりも高い予測精度を実現することができた。しかしながら、タンパク質 - タンパク質相互作用予測やドッキング予測などでは、まだ改善の余地があり、今後の課題としては、これらの手法のさらなる改良が挙げられる。とくに、ドッキングについては、現在、高速性を生かしたアンサンブルドッキングなど、複合体形成時の構造変化に対応できる手法の開発に取り組んでいる。また、データベースの開発も今後の重要な課題であり、現在、タンパク質 - リガンド結合データベース、糖鎖結合タンパク質データベースの開発を行っている。今後は、タンパク質構造レベルでも、多数の対象に対する網羅的ドッキング、物理的な相互作用、ダイナミクスも含めた結果のデータベース化を手がけていきたいと考えている。従来個別に行われてきた物理化学的な相互作用の解析とデータの蓄積を通して、生命システムの理解に原子レベルからアプローチすることは今後の重要な課題と考えている。

＜研究期間の全成果公表リスト＞

1）論文／プロシーディング

1. 0702131221
   R. Jono, T. Terada, K. Shimizu: A multicanonical ab initio molecular dynamics method: application to conformation sampling of alanine tripeptide, *Chem. Phys. Lett.*, **432**, 306-312 (2006).
2. 0801161647
   M. Hirano, et al.: IgEb immune complexes activate macrophages through Fc-gamma RIV binding, *Nature Immu.*, **8**, 762-771 (2007).
3. 0702131217
   S. Yamasaki, S. Nakamura, T. Terada, K. Shimizu: Mechanism of the difference in the binding affinity of E.coli tRNAGln to glutaminyl-tRNA synthetase caused by non-interface nucleotides in variable loop, *Biophy. J.*, **92**, 192-200 (2007).
4. 0805081419
   M. Morita, S. Nakamura, K. Shimizu: Highly accurate method for ligand-binding site prediction in unbound state (apo) protein structures, *Proteins*, **73**, 468-479 (2008).
5. 0801231659
   R. Ishitani, T. Terada, K. Shimizu: Refinement of comparative models of protein structure by using multicanonical molecular dynamics simulations, *Mol. Simul.*, **34**, 327-336 (2008).
6. 0801161657
   M. Kakuta, S. Nakamura, K. Shimizu: Prediction of protein-protein interaction sites using only sequence information and using both sequence and structural information, *IPSJ Transactions on Bioinformatics,* **49**, 25-35 (2008).
7. 0912041422
   T. Terada, K. Shimizu: A comparison of generalized Born methods in folding simulations. *Chem. Phys. Lett.*, **460**, 295-299 (2008).
8. 0805081422
   T. Terada, et al.: Understanding the roles of amino acid residues in tertiary structure formation of chignolin by using molecular dynamics simulation *Proteins*, **73**, 621-631 (2008).
9. 0912041433
   W. Cao, et al.: Using a new GPI-anchored-protein identification system to mine the protein databases of Aspergillus fumigatus, Aspergillus nidulans, and Aspergillus oryzae, *J. Gen. Appl. Microbiol.* **55**, 5, 381-393 (2009).
10. 0912041436
    S. Nakamura, K. Shimizu: Comprehensive analysis of

sequence-structure relationships in the loop regions of proteins, *GIW 2009*, accepted.

11. 0912040037
S. Yamasaki, T. Terada, K. Shimizu, H. Kono, A. Sarai: A Generalized Conformational Energy Function of DNA Derived from Molecular Dynamics Simulations, *Nucleic Acids Res.*, **37**, e135 (2009).

12. 0912041442
R. Jono, Y. Watanabe, K. Shimizu, T. Terada: Multicanonical ab inito QM/MM Molecular Dynamics Simulation of a Peptide in an Aqueous Environment, *J. Comput. Chem.*, accepted.

13. 0901071711
W. Cao, et al. Computational Protocol for Screening GPI-anchored Proteins, *Proc. the First Int. Conf. Bioinformatics and Computational Biology (BICoB), Springer Lecture Notes in Bioinformatics Series*, **5462,** 164-175 (2009).

2）データベース／ソフトウェア

1. 0801171140
タンパク質 - リガンド結合状態／非結合状態ペアデータセット，
http://www.bi.a.u-tokyo.ac.jp/services/buddy/current/