

タンパク質間の形状相補性に関する 統計量を用いた親和性予測法の高精度化

吉川 達也^{†, ††} 塚本 弘毅[†] 蓬来 祐一郎[†] 福井 一彦[†]

タンパク質とは、生命活動を支える重要な物質の一つである。これらタンパク質固有の機能は、他のタンパク質との相互作用 (PPI: Protein-Protein Interaction) によって発現することが解っている。この PPI に関する研究は、生命現象の解明はもとより、最近では医薬品開発の分野でも注目されている。我々が提案したタンパク質間の親和性予測法は、データセットの構成や規模、親和性スコアの計算に重要なクラスタリング処理によって、結果精度が異なることが予備実験などから解っている。特に、データセットが大規模となる場合に、クラスタリングパラメータ値の予測精度に与える影響が無視できなくなる傾向がある。本研究では、全 7056 個中に 84 個の生物学的有意性を持ったタンパク質ペア (BSPPs: Biological Significant Protein Pairs) を含むデータセットを用いて、prevalence=1.19%の高難易度な BSPPs 予測問題を設定した。また、親和性スコアの定義から予測精度に影響を及ぼすと考えられる 4 つのパラメータを抽出し、計 144 個のパラメータセットを作成した。評価実験から、F-measure 最大となるパラメータセットにおいて、F-measure 最大となるパラメータセットにおいて、sensitivity(=recall)=27.4%、specificity=91.0%、precision=3.53%、accuracy=90.2%、F-measure=6.25%の結果を得た。これは BSPPs を無作為に抽出した場合 (F-measure=2.32%) と比較して、約 2.69 倍の予測精度を獲得しており、タンパク質間親和性予測法の精度向上におけるパラメータ最適化の有効性を示している。

Improvement in Accuracy of Affinity Prediction Method Using Statistics for Shape Complementarity between Proteins

TATSUYA YOSHIKAWA^{†, ††} KOKI TSUKAMOTO[†]
YUICHIRO HOURAI[†] KAZUHIKO FUKUI[†]

The protein is one of the important materials that support a vital activity in vivo. It is understood that the function of these protein endemics appears by the interaction with other proteins. One of the purposes of PPI is to clarify the life phenomenon, and is paid to attention also in the field of the drug development recently. As results of several preliminary experiments, an affinity prediction method that we proposed tends to give the different accuracy of results corresponding to the data formation, the data scale and a clustering that influences the calculation of affinity scores. Especially, in the case of increasing the data scale, the influence of the parameter values of a clustering method on the prediction accuracy cannot be disregarded. In this paper, we set a BSPPs (Biological Significant Protein Pairs) prediction problem of high difficulty of which the prevalence ratio is 1.19%. And we made the 144 parameter sets by detecting the 4 key parameters related to the definition of an affinity score. As a result of evaluation experiment, our affinity prediction method obtained 27.4% sensitivity, 91.0% specificity, 27.4% recall, 3.53% precision, 90.2% accuracy and 6.25% F-measure in a parameter set that F-measure is maximum. The prediction accuracy of this result was about 2.69 times higher than that of 2.32% F-measure in the case of extracting BSPPs at random. It shows the effectiveness of optimizing the clustering parameters in the improvement in accuracy of affinity prediction method.

1. はじめに *

タンパク質とは、核酸 (DNA/RNA)、糖質、脂質とともに生体の主要な構成成分の一つであり、生命活動を支える重要な物質である。構成アミノ酸の種類、数、結合順序などによって、固有の立体構造と機能を有している。これらタンパク質固有の機能は、他のタンパク質との相互作用によって発現することが解っている。このタンパク質間相互作用 (PPI: Protein-Protein Interaction) は、細胞内シグナル伝達、酵素反応、遺伝子発現調節と言った様々な生命現象において中心的な役割を担っている。最近では、創薬科学分野でも PPI に基づく医薬品開発が注目されており、PPI 研究は我々の身近な暮らしにとっても非常に有用であると考えられる。

PPI に関する従来研究の多くは、Giotら[1]が行った *Drosophila melanogaster* に関する相互作用マップの研究に代表されるように、ゲノム規模やアミノ配列解析に基づいたアプローチがなされてきた[2-4]。しかしながら、タンパク質の一次構造のみを用いて得られたこれらの相互作用マップには、タンパク質の立体構造に基づく相互作用性が見積もられておらず、得られた結果から機能と構造の関係にまで広げて解析を行うことは難しかった。このような理由から、創薬分野な

どへの応用を目的とした PPI 研究には、タンパク質の立体構造に基づいたアプローチが考えられる。一方、タンパク質の立体構造に基づいた PPI の研究は、Smithら[5]によって議論されているが、具体的な解析結果や精度などについては言及されていない。さらに、このようなタンパク質の立体構造情報に基づく大規模な PPI 解析システムの実現には、計算速度と解析精度のトレードオフを克服するための改善策が必要となる。たとえば、巨大分子同士の組合せによって探索空間が膨大となるタンパク質間ドッキング (PPD: Protein-Protein Docking) の計算においては、分子動力学シミュレーションなどによる複合体予測は困難を極める。このような PPD に関する従来研究の多くは、参照実験データの不足や探索空間の恣意的縮小による最適解保障の観点などから問題があった[6-9]。このため、タンパク質の立体構造情報を用いた PPI 研究は、これまで本格的に行われてこなかった。

そこで我々は、この立体構造に基づいた PPI を予測するために、PPD (Protein-Protein Docking) の評価結果を用いる方法を提案した。PPD の計算には、タンパク質間の形状相補性を網羅的に探索する方法を採用した。この網羅的な計算には前述した膨大な計算時間を要するが、我々は、(1) タンパク質表面を構成する原子種に応じて、表面の厚さを調整してモデル化することにより、従来研究と同等以上の高精度を維持し、(2) 形状相補性の計算に、共著者の蓬来らが開発した高性能 FFT ライブラリ (CONV3D)[10]を用いて高速化を図り、さらに、超並列計算機 IBM Blue Gene (8192 cores, 2TB memory) を用いて超並列化を実現して、高速かつ高精度な形

*[†] 産業技術総合研究所 生命情報工学研究センター
Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST)

^{††} 大阪大学大学院 情報科学研究科
Graduate School of Information Science and Technology, Osaka University

状相補性評価システムを開発した[11].

我々が提案したPPI予測法は、このPPD計算で得られた複合体の予測構造群に対して、Groupingと称する独自の統計的手法を用いることにより、タンパク質間の親和性を高精度に予測する手法である。ところが、この親和性予測法は、データセットの構成や規模、タンパク質間の結合可能性の評価値（親和性スコア）の計算に重要なクラスタリング処理によって、結果精度が異なることが予備実験などから解ってきた。特に、データセットが大規模となる場合に、クラスタリングパラメータ値の予測精度に与える影響が無視できなくなる傾向がある。そこで本研究では、親和性スコアの定義から、親和性予測法の結果精度に影響を及ぼすと考えられる4つのパラメータを最適化することによって、タンパク質間親和性予測法の予測精度を向上させることを目的とする。

2. タンパク質間親和性予測法

2.1 親和性予測法の手順

本研究における親和性予測法とは、Groupingと称する独自の統計的手法を用いて、タンパク質間の形状相補性に関する特徴量から、タンパク質間の生物学的な有意性を予測することである。以下にその詳細な手順を示す。

STEP 1 形状相補性評価

2つのタンパク質の立体構造情報を用いて、タンパク質間の形状相補性の評価を行う。まず、対象となるタンパク質を3次元格子($I \times m \times n$)状に離散化し、その各格子点に対して、タンパク質の幾何学的な特性に基づく構造属性値を割り当てる。次に、一方のタンパク質に対して、もう一方のタンパク質を回転・並進移動させ、その際に重なり合う構造属性値の積和（以下、形状相補性スコア S_{PSC} とする）を計算する。たとえば、回転角の刻み： $\Delta\theta=15^\circ$ とした場合、回転の対称性から回転体の数は3600個となる[12]。各回転体の3次元格子全点において評価が行われるので、全体では、計 $(I \times m \times n) \times 3600$ 個の評価結果を得ることができる。本研究では、この中から形状相補性スコア S_{PSC} の降順に $N_c=512$ 個を抽出し、これを形状相補性評価の候補解集合 X とする。

STEP 2 Grouping

STEP 1で求めた候補解集合 X を立体構造の類似度にしたがって階層的に分類する。ここで、クラスタリングの結果得られる各クラスタ C_k の代表解 y_k の集合を Y と定義する。詳細は、2.2節に示した。

STEP 3 Zスコアの計算

STEP 2で求めた Y の形状相補性スコア S_{PSC} の分布(Z)と、各クラスタの類似構造分布(Z_c)を計算する。ここで、 Z_c^k は X を母集団とする $y_k \in Y$ のZスコアであり、 Z_c^k は C_k のメンバ数集合 $M_c^k = \{M_c^k | k \in N_c, k \geq 0\}$ を母集団とする C_k のメンバ数 M_c^k のZスコアである。

STEP 4 親和性スコア S_{AFF} の計算

STEP 3で求めた各クラスタ C_k に Z_c^k と Z_c^k をそれぞれ重み付けして線形結合した値 Z^k を計算し、 $Z^k > 3.0\sigma$ となる l 個の Z^l の平均値を親和性スコア S_{AFF} と定義する。ここで、 w_0 と w_c は、それぞれ Z_c^0 と Z_c^k の重み係数を表している：

$$Z^k = w_0 \cdot Z_c^0 + w_c \cdot Z_c^k, Z^l = \{Z^k | Z^k > 3.0\sigma\} \quad (1)$$

$$S_{AFF} = \frac{1}{l} \sum_l Z^l$$

2.2 Grouping

Groupingとは、形状相補性評価に基づく親和性予測法をより高精度化するために提案した独自の手法であり、形状相補性評価の結果得られる512個の候補解群を広義の最短距離法に基づいた階層的なクラスタリングによって、10個のクラスタに分割する処理のことを表している。複数の候補解の形状相補性スコアと立体構造を統計的に解析することによって、1

つの形状相補性スコアのみを用いて親和性を評価する場合と比べて、より高精度なタンパク質間の親和性を予測することができる。以下にその詳細な手順を示す。

STEP 1 対象データ集合 X の決定

形状相補性評価で得られる512個の候補解群をクラスタリングの対象データ集合 $X = \{x_1, \dots, x_{512}\}$ に設定する。各データ x_i は、 $x_i = (S_{PSC}^i, T_i(T_x, T_y, T_z), R_i(R_\theta, R_\phi, R_\psi))$ で表される。ここで、 S_{PSC}^i は形状相補性スコア、 $T_i(T_x, T_y, T_z)$ 、 $R_i(R_\theta, R_\phi, R_\psi)$ は、3.4節に後述するLigandタンパク質の並進量 $T_i(T_x, T_y, T_z)$ 、回転量 $R_i(R_\theta, R_\phi, R_\psi)$ を示している。はじめに、全データ x を形状相補性スコア S_{PSC}^i の降順に並び替える。

STEP 2 代表データ y_k の決定

次に、形状相補性スコア S_{PSC}^i が最大となる x_i をクラスタ1(C_1)の代表データ $y_1 \in Y$ とする。

STEP 3 データ x_i と y_k 間の非類似度計算

未分類データ $x_i \in C_k$ と代表データ y_j 間の距離(非類似度) $D_1(y_j, x_i)$ 及び $D_2(y_j, x_i)$ を計算する。ここで、 D_1 は、データ x_i と代表データ y_j のLigand並進移動による構造の違いを並進量のユークリッド距離で定義している。また、 D_2 は、データ x_i と代表データ y_j 間のLigand回転移動による構造の違いを法線ベクトル \mathbf{n}_i と \mathbf{n}_j の余弦差で定義する。ここで、ベクトル \mathbf{n} は、点G、C、Nを含む平面 α に垂直な法線ベクトルを示している。点GはLigandタンパク質の重心、点C及び点NはそれぞれC末端とN末端のC α 原子の座標を表している。

STEP 4 クラスタ1(C_1)への分類

STEP 3で求めた $D_1(y_1, x_i)$ 及び $D_2(y_1, x_i)$ が以下の2つの条件を同時に満たす場合、データ x_i をクラスタ1(C_1)に分類する。

[11]においては、 $D_1=10 \text{ \AA}$ 、 $D_2=\pi/6$ とした。

(条件 1) $D_1(y_1, x_i) \leq D_1$

(条件 2) $D_2(y_1, x_i) \leq D_2$

STEP 5 クラスタ1(C_1)の決定

全ての未分類データ $x_i \in C$ に対して、STEP 3~STEP 4を繰り返し、クラスタ1(C_1)を決定する。

STEP 6 全クラスタ C の決定

STEP 2~STEP 5の手順を繰り返し、全てのクラスタ C を決定する。クラスタ数 $N=10$ となるか、全てのデータ x が何れかのクラスタ C_k に分類されたところで、処理を終了する。

3. 評価実験

3.1 親和性予測問題

本研究における親和性予測問題とは、複数のタンパク質ペアからなるデータセットの中から、生物学的有意性を持ったタンパク質ペア(BSPPs: Biological Significant Protein Pairs)を高精度に探索することである。すなわち、BSPPsを正例、生物学的有意性を持たないタンパク質ペア(BIPPs: Biological Insignificant Protein Pairs)を負例とする二項分類問題と捉えることができる。問題の大きさは、全データセットの中に含まれる正例の割合(BSPPs含有率, prev: prevalence)次式で表される。たとえば、3.4節に後述するデータセットを用いた場合のprevは、1.19% (=84/7056)である。

$$prev = n(BSPPs) / (n(BSPPs) + n(BIPPs)) \quad (2)$$

3.2 評価指標

我々が提案した親和性予測法の精度を評価する指標として、F-measureを採用した。これは、問題の規模を表すBSPPs含有率に依存した評価指標であり、トレードオフの関係にあるrecallとprecisionを1つの指標で評価することができる。本研究では、評価実験の結果得られた予測精度とBSPPsを無作為抽出した場合の精度を比較することで、精度の向上率を評価する。ここで、3.4節に後述する7056個のデータセットを用いたBSPPs予測問題に対して、無作為抽出法を適用した場合

の予測精度は、 $AUC_{md}=0.5$, $sens_{md}=0.5$, $spec_{md}=0.5$, $rec_{md}=0.5$, $prec_{md}=prev=0.0119$, $acc_{md}=0.5$, $F_{md}=2 \times 0.5 \times 0.0119 / (0.5 + 0.0119)=0.0232$ である。

3.3 パラメータセット

2.1 節に示した親和性スコアの定義式から、予測精度に影響を及ぼすと考えられる4つのパラメータ D_r , D_l , w_c , w_s を抽出して、表1に示すような値に設定した。ここで、 D_r , D_l は、Groupingで定義されたパラメータであり、クラスタリングの対象となる候補補の構造類似性を評価する際に用いる閾値である。また、 w_c , w_s は、Groupingで得られたクラスタのスコア分布とクラスタの密度分布を足し合わせて、親和性スコア S_{Aff} を計算する際に用いる重み係数を表している。評価実験では、これらの4つのパラメータを用いて、計144通りのパラメータセット $P(D_r, D_l, (w_c, w_s))$ を作成する。

表1 最適化対象の親和性予測パラメータ
Table 1 Affinity prediction parameters for optimization.

Parameter	Value
D_r (Å)	5, 10, 15, 20
D_l (degree)	15, 30, 45, 60
(w_c, w_s)	(0.00, 1.00), (0.05, 0.95), (0.10, 0.90), (0.15, 0.85), (0.20, 0.80), (0.25, 0.75), (0.50, 0.50), (0.75, 0.25), (1.00, 0.00)

3.4 データセット

Weng らによって作成された Protein-Protein Docking Benchmark 2.0[13-14]のうち、計168個のタンパク質 (bound 構造) を用いて、データセットを作成する。全てのタンパク質は、Receptor (以下、Rとする) と Ligand (以下、Lとする) のいずれかに分類される。RとLの数は、それぞれ84個である。ここで、1つのRと1つのLからなるタンパク質ペアをRLとすると、RLの組合せの数は7056 (=84×84) ペアである。この計7056個のRLの中には、84個のBSPPsが含まれている。ただし本研究では、結合部位の高速探索のために、全てのタンパク質を剛体モデルとして取り扱う。

4. 結果と考察

4.1 親和性予測法の評価結果

提案した親和性予測法に関する評価実験を行った結果、全てのパラメータセットにおいて、BSPPsを無作為に抽出した場合よりも高精度な結果を得ることが解った。具体的には、パラメータセット: $P(D_r, D_l, w_c, w_s)=(5, 15, 0.00, 1.00)$ において、無作為抽出法よりも最大で約2.69倍の予測精度を得た。このとき、各評価指標の値は、 $AUC=0.585$, $sensitivity(=recall)=0.274$, $specificity=0.910$, $precision=0.0353$, $accuracy=0.902$, $F\text{-measure}^{max}=0.0625$ であった (表2)。

さらに、パラメータセットと結果精度の関係について考察

表2 パラメータ最適化による親和性予測の評価結果

Table 2 Evaluation results of affinity prediction using parameter optimization.

	AUC	Sens. (=Rec.)	Spec.	Prec.	Acc.	F-measure
平均	0.576	0.500	0.646	0.0200	0.644	0.0376
標準偏差	0.00199	0.0185	0.0189	0.000533	0.0184	0.000869
中央値	0.585	0.429	0.715	0.0178	0.712	0.0338
最大 $P(D_r, D_l, w_c, w_s)$	0.614 (5,45,0.25,0.75)	0.952 (5,15,0.25,0.75)	0.926 (20,15,0.10,0.90)	0.0357 (15,45,0.10,0.90)	0.918 (20,15,0.10,0.90)	0.0625 (5,15,0.00,1.00)
最小 $P(D_r, D_l, w_c, w_s)$	0.498 (20,30,0.75,0.25)	0.167 (20,30,0.75,0.25)	0.171 (5,60,0.75,0.25)	0.0133 (5,15,1.00,0.00)	0.180 (5,60,0.75,0.25)	0.0262 (5,15,1.00,0.00)
尖度	0.319	-1.25	-1.18	-0.253	-1.18	-0.373
歪度	-1.02	0.428	-0.457	1.01	-0.457	0.951

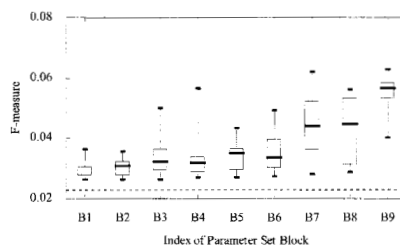


図1 ブロック毎のF-measure評価結果
Figure 1 Evaluation results of F-measure in each blocks.

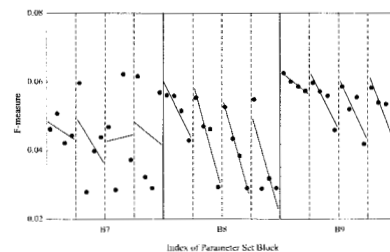


図2 B7-B9ブロックの詳細結果
Figure 2 Detailed results of B7-B9 blocks.

を行うために、3.3節に定義した計144個のパラメータセットを9つのブロック $B_1 \sim B_9$ に分けて、ブロック毎に結果をまとめた (図1)。ただし、1つのブロックは、同じ (w_c, w_s) 値を持つ16個のパラメータセットから構成されており、ブロックのインデックス $B_1 \sim B_9$ は w_s 値の昇順に割り当てた。図1から、 B_1 から B_9 へと w_s の値が大きくなるにしたがって、F-measureの中央値が大きくなり、予測精度が向上していることが解った。特に、 $B_7 \sim B_9$ では、 $B_1 \sim B_6$ の場合と比較して、F-measureの中央値が約1.51倍急増している。以上の結果から、我々が提案した親和性予測法は、設定したパラメータセットが $B_7 \sim B_9$ ($w_c=0.90 \sim 1.00$, $w_s=0.10 \sim 0.00$) の時に高精度な結果を得ることが解った。

次に、 D_r と D_l の値が予測精度に及ぼす影響を考察するために、 $B_7 \sim B_9$ の各ブロック内のデータをプロットした (図2)。各ブロック内のパラメータセットの並びは、4つの並び毎に D_l の値が変わるように整理されている。全ブロックの評価結果に対して事前解析を行ったところ、前述した4つのパラメータセット毎に特徴的な直線関係が観察されたため、図2にはこれら4つのパラメータセットの評価結果に対する回帰直線を併せて図示した。図中に示した回帰直線のピークは、いずれの場合も D_l の値が15の場合であり、その傾向は B_8 と B_9 で顕著に見られた。これは、 D_r を15に設定した場合には、 D_l の

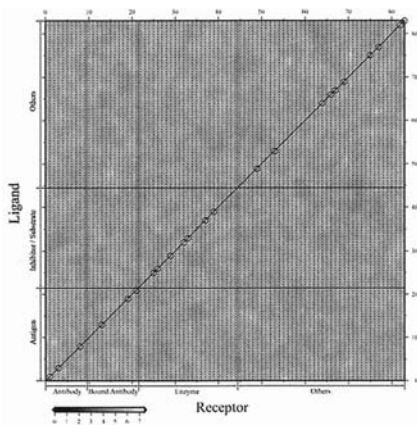


図 3 F-measure 最大のパラメータセットにおけるタンパク質間相互作用マップ

Figure 3 Protein-protein interaction map using parameter set with maximum F-measure.

値に依らず予測精度が高くなることを示している。

さらに、これまでの結果から、特に高精度となった B_3 と B_4 において D_0 を15と設定し、 D_0 の設定値によってF-measureの平均がどのように変動するかを解析した。解析結果から、 $D_0=5$ のF-measureの平均は0.0593、 $D_0=10$ では0.0574、 $D_0=15$ では0.0556、 $D_0=20$ の場合では0.0565となり、 $D_0=5$ の場合に最も高精度な結果が得られることが解った。

以上まとめると、我々が提案した親和性予測法は、本研究で設定したパラメータセットに依らず、BSPPsを無作為に抽出した場合と比較して、高精度に予測できることが解った。この時の最大予測精度は、無作為に抽出した場合の約2.69倍であった。この親和性予測法を用いて最も高精度な結果を引き出すためのパラメータセットは、 $P(D_0, D_n, w_{cs}, w_r) = (5, 15, 0.00, 1.00)$ であり、最大予測精度を得た B_3 に含まれるパラメータセットと同等の精度を得られ、かつ、パラメータ間の依存性が低くなるパラメータセットは、 $P(D_0, D_n, w_{cs}, w_r) = (5, 15, 0.05 \sim 0.00, 0.95 \sim 1.00)$ であることが解った。現在、 $400 (=20 \times 20)$ 、 $900 (=30 \times 30)$ 、 $2500 (=50 \times 50)$ の異なる規模のデータセットを用いて評価実験を行い、これらについても検証を行っている。

4.2 タンパク質間相互作用マップ

我々が提案した親和性予測法の目的は、データセットに含まれるBSPPsを高精度に予測することである。4.1節に示した評価結果において、F-measureが最大となる場合のPPIマップを図3に示す。これは、タンパク質間親和性予測法によって得られた7056個の親和性スコア S_{Aff} をプロットした図である。親和性スコア S_{Aff} が低い場合は白く、高い場合は黒くなるようにグレースケールで色付けている。横軸はReceptor (R)、縦軸はLigand (L)のインデックスを表している。また、図中の両軸に平行な直線は、タンパク質の生物学的な機能分類の境界を表し、対角線は84個のBSPPsを表している。さらに、図中の○印は、親和性予測の結果、親和性が高いと正しく予測されたBSPPsを表している。このとき、F-measure最大となるパラメータセット： $P(D_0, D_n, w_{cs}, w_r) = (5, 15, 0.00, 1.00)$ における親和性予測精度は、 $AUC=0.585$ 、 $sensitivity(=recall)=0.274$ 、 $specificity=0.910$ 、 $precision=0.0353$ 、 $accuracy=0.902$ 、 $F\text{-measure}^{max}=0.0625$ となり、図中には正しく予測された23個のBSPPsが○印で示されている。

5. まとめ

本研究では、我々が提案したタンパク質間親和性予測法の精度向上におけるクラスタリングパラメータ最適化の有効性に関する検証を行った。評価実験の結果、BSPPsを無作為に抽出した場合 ($F\text{-measure}=0.0232$) と比較して、最大で約2.69倍の予測精度を獲得しており、タンパク質間の形状相補性に関する統計量を用いた親和性予測法の精度向上における、クラスタリングパラメータ最適化の有効性を示した。

謝辞 IBM Blue Gene用に最適化した高速FFTライブラリをご提供いただいた東京大学の額田彰氏に、謹んで感謝の意を表す。

参考文献

- 1) Giot, L., et al., A protein interaction map of *Drosophila melanogaster*. *Science*, 2003. 302(5651): p. 1727-36.
- 2) Rain, J.C., et al., The protein-protein interaction map of *Helicobacter pylori*. *Nature*, 2001. 409(6817): p. 211-5.
- 3) Fields, S. and O. Song, A novel genetic system to detect protein-protein interactions. *Nature*, 1989. 340(6230): p. 245-6.
- 4) Wojcik, J. and V. Schachter, Protein-protein interaction map inference using interacting domain profile pairs. *Bioinformatics*, 2001. 17 Suppl 1: p. S296-305.
- 5) Smith, G.R. and M.J. Sternberg, Prediction of protein-protein interactions by docking methods. *Curr Opin Struct Biol*, 2002. 12(1): p. 28-35.
- 6) Palma, P.N., et al., BiGGER: a new (soft) docking algorithm for predicting protein interactions. *Proteins*, 2000. 39(4): p. 372-84.
- 7) Taylor, J.S. and R.M. Burnett, DARWIN: a program for docking flexible molecules. *Proteins*, 2000. 41(2): p. 173-91.
- 8) Norel, R., et al., Examination of shape complementarity in docking of unbound proteins. *Proteins*, 1999. 36(3): p. 307-17.
- 9) Norel, R., et al., Electrostatic contributions to protein-protein interactions: fast energetic filters for docking and their physical basis. *Protein Sci*, 2001. 10(11): p. 2147-61.
- 10) Nukada, A., et al., High Performance 3D Convolution for Protein Docking on IBM Blue Gene. In Proceedings of the Fifth International Symposium on Parallel and Distributed Processing and Applications (ISPA07), 2007. 4742: p. 958-969.
- 11) Tsukamoto, K., et al., Development of an affinity evaluation and prediction system by using the shape complementarity characteristic between proteins. *Journal of Bioinformatics and Computational Biology*, 2007(submitted).
- 12) Lattman, E., Optimal sampling of the rotation function. *Acta Crystallographica Section B*, 1972. 28(4): p. 1065-1068.
- 13) Mintseris, J., et al., Protein-Protein Docking Benchmark 2.0: an update. *Proteins*, 2005. 60(2): p. 214-6.
- 14) Chen, R., et al., A protein-protein docking benchmark. *Proteins*, 2003. 52(1): p. 88-91.