

修士論文

蛋白質-蛋白質結合過程における
長距離相互作用の役割

2008年3月修了

東京大学大学院
新領域創成科学研究科 情報生命科学専攻
学籍番号：66953

有馬 聡史

指導教員： 北尾 彰朗 准教授

論文題目 : 蛋白質-蛋白質結合過程における長距離相互作用の役割

著者 : 有馬 聡史 (学生証番号 : 66953)

指導教員 : 北尾 彰朗 准教授

所属 : 東京大学大学院 新領域創成科学研究科 情報生命科学専攻

修了年月 : 2008 年 3 月

キーワード : 蛋白質、結合構造予測、静電相互作用

論文要旨 :

本研究では蛋白質の結合過程において蛋白質間に働く長距離相互作用が結合構造探索の手掛かりになっているという仮説を立て、それを検証した。手法としては結合構造が既知の蛋白質ペアを一定の距離に離し、それぞれ回転させた構造(回転後のペアの構造を配向と呼ぶ)を生成し、各配向の相互作用ポテンシャルエネルギーを計算した。得られたエネルギーから各配向の存在確率を計算し、解析を行った。その結果、結合構造をそのまま離れた配向(真の配向と呼ぶ)は長距離において最安定ではないものの、多くのペアで確率的に有利になっていることが示された。また、真の配向の存在確率と蛋白質の様々な性質との相関係数を計算したところ、蛋白質の形状および蛋白質内の双極子モーメント同士の向きに有意な相関があることを見出した。

目次

1. 序論.....	1
2. 方法.....	4
2.1 手順.....	4
2.1.1 蛋白質の選択.....	4
2.1.2 計算方法.....	5
2.1.3 データ処理と解析.....	6
2.2 原理.....	6
2.2.1 エネルギー計算.....	6
2.2.2 Generalized Born model (GB model).....	7
2.2.3 静電場の多重極展開.....	8
2.2.4 相関係数.....	10
3. 結果と考察.....	11
3.1 回転によるエネルギー変化.....	11
3.2 距離に依存する確率分布の変化.....	11
3.3 真の配向は長距離における最安定構造ではない.....	12
3.4 真の配向は長距離において確率的に有利である.....	13
3.5 真の配向の有利性は塩の有無に依存する.....	14
3.6 真の配向になる確率と静的な性質との相関.....	14
4. 結論.....	17
謝辞.....	18
参考文献.....	19
図.....	22
表.....	27
追記.....	32

1. 序論

結合構造予測とは、蛋白質単体の構造から複合体の結合構造を予測することをさす。生体内の多くの反応には蛋白質結合が関わっているため、どの蛋白質同士が結合するのか、どのように結合するのかの情報は、生物に関するあらゆる研究にとって非常に重要である。蛋白質と小分子リガンドの結合予測はある程度の成功をおさめている[1]。しかし、蛋白質同士の結合予測については多数のグループが研究しているが、いまだ発展段階にある[2,3,4]。

蛋白質同士の結合予測には二つの困難がある。一つ目は、結合構造の評価の難しさである。天然の結合構造を真の結合、それ以外の全ての結合構造を偽の結合であると呼ぶことができるが、結合予測の際には真の結合と偽の結合を見分ける評価関数を作る必要がある。しかし、実際の結合過程において真と偽が何によって見分けられているか完全には理解されていない[5]。蛋白質-小分子リガンド結合は形状の相補性のみによってもある程度予測が可能であることが知られている[6]。しかし蛋白質同士の結合は広い面で接触するため、様々な相互作用が総合的に作用して決まっていると考えられ、形状の相補性のみでは予測が難しい。二つ目は、計算コストが大きいことである。蛋白質そのものが計算に用いるには大きいのに加えて、結合予測では二つの蛋白質の三次元上における配向の組み合わせとして構造を考えなければならない。例えば半径 30 Å 程度の蛋白質同士の結合を考える。表面を球とみなしてそれぞれの表面を 5 Å² の小面積に分割し、小面積同士の全組み合わせの結合構造を計算するには、

$$\left(\frac{4 \times \pi \times 30^2}{5}\right)^2 \approx 5 \times 10^6 \text{ 通り} \quad (1.1)$$

の計算が必要になる。1 通りの評価計算に 10 秒掛かるとすると、全通りの計算を行うには約 600 日掛かることになる。

現状の結合予測アルゴリズムではランダムまたは総当り的に結合構造を生成し、それぞれに対して独自の評価関数により順位付けするものが多い[7,8,9]。その場合には十分な数の配向を試行しなければならず、限られた計算資源の下では詳細な評価関数

を用いることができなくなる。Gray らの方法では最初は粗い評価関数を用いて数を絞り、その後詳細な評価関数を用いる工夫をしている[7]。また、Fernández-Recio らの方法のように、少ない試行で済ませられるように Monte Carlo 法により開始構造の周辺を探索することで、開始構造の数を減らしている場合もある[8]。いずれの方法も最初の総当り的な結合構造の生成には特別な情報を使わず、ランダムもしくは一様に行っている。真の結合構造の周辺以外を計算することは無駄となってしまうので、何らかの方法で真の結合構造の場所を大まかに絞り、初期段階の計算範囲を減らすことで結合構造予測の効率を向上させることができる。

天然の蛋白質同士の結合過程に着目すると、接触時の構造だけをもとに結合構造探索が行われているとは考えにくい。接触時には原子間の静電力や van der Waals 力に加えて、水素結合や疎水性相互作用などが総合的に働いて特異的な結合をもたらす。しかし特に水素結合と疎水性相互作用は局所的な効果であり、総合的な評価は完全に接触しなければわからない。天然において完全に接触することで真の結合と偽の結合を見分けていると仮定すると、上述の結合構造予測アルゴリズムと同様に表面の全探索が必要になり、膨大な数の組み合わせを試行しなければいけなくなる。これを見積もるために、以下のモデルを考える。

半径 30 Å、50 kDa の蛋白質同士の真空中での結合を考える。それぞれを球と見なしたとき、平均自由行程の考え方から単位時間あたりの衝突頻度 f は以下の式で表せる。

$$f = \sigma vn \quad (1.2)$$

σ は衝突断面積、 v は分子の平均速度、 n は分子の濃度である。 σ は分子の二倍の半径を持つ円の面積として、 v は運動エネルギーの等分配則から求めることができ、温度を 298K、濃度は 1 μ M とすると、

$$f \approx 8.3 \times 10^2 \text{ (ms}^{-1}\text{)} \quad (1.3)$$

と見積もることが出来る。この頻度では 1ms につき各蛋白質上の $\sqrt{8.3 \times 10^2} \approx 29$ 点しか探索できない。したがって、一般的な蛋白質結合のタイムスケールと言われている

ms のオーダーで全探索が行われるとは考えにくい。水分子の影響を考慮すると、さらに大きく頻度が低下する。

このように、接触時の構造だけを手掛かりに結合構造探索が行われることは現実的でないと考えられる。本研究ではこの考えを基に、長距離において蛋白質の配向を決める力が働いており、その結果天然における結合構造探索が可能になっている、という仮説を立てた。長距離で配向を決める力が働けば、それだけ結合構造探索の効率上がる。

長距離において配向が決まる、もしくは探索範囲が限定されることがわかれば、その過程を結合予測アルゴリズムに応用することが期待できる。真の構造の周辺のみを限定して探索を行えば結合予測アルゴリズムの効率を向上させることができる可能性がある。例えば静電相互作用は長距離において各蛋白質内の多重極子モーメントとして近似できることが知られており、これを利用できれば非常に小さい計算コストで予測効率を向上させられるだろう[10]。

特定の蛋白質と DNA の結合において、接触面以外での長距離静電相互作用が配向を決める一因となっていることが実験によって示唆された[11]。しかしながら、様々な蛋白質において長距離の相互作用を定量的・統計的に計算する研究はなされていない。そこで本研究では、まず既知の結合構造を用いて長距離において真の結合構造がエネルギー的に有利になっているかどうかを計算により検証した。さらに蛋白質のどのような性質が長距離において配向を決める元となっているかを調べるために、多重極子モーメントを含む様々な静的性質と配向の相関を検証した。

2. 方法

2.1 手順

本研究は蛋白質同士の結合過程において、長距離で正しい結合構造を決めるための力が働いているという仮説を検証する。結合状態の結晶構造からは、結合前に長距離でそれぞれの蛋白質がどのような向きになっているかを直接知ることはできない。そのため蛋白質の向きがガノニカル分布に従って存在すると仮定し、長距離における蛋白質構造のエネルギーから確率を求めることにより議論を行う。

具体的には、結合構造が既知である蛋白質ペアを特定の距離に離してそれぞれ回転させた構造（回転させた後の蛋白質ペアの構造を配向と呼ぶ）を生成し、各配向のポテンシャルエネルギーを計算する。それによって真の結合構造をそのまま離れた配向（真の配向と呼ぶ）と、その他の配向のエネルギーの違いを測定できる。その結果から真の配向がエネルギー的・確率的に有利であるかどうかを検証し、さらに長距離での配向と関係のある蛋白質の静的性質を検討する。研究の手順を以下に示す。

2.1.1 蛋白質の選択

蛋白質ペアの選択には Mintseris らの作成したテストセットを用いた[12]。このテストセットは結合構造予測アルゴリズムのベンチマークとして用いられることを目的としている。これは様々な種類の結合方式を網羅するために、SCOP(Structural Classification of Proteins)データベースの分類から非冗長のセットとしている[13]。SCOP はドメインを単位とし、アミノ酸配列および立体構造の類似性を基準として分類を行うデータベースである。蛋白質同士の結合には蛋白質の性質と立体構造の両方が関わっているため、この分類法は結合構造の研究に適している。本研究の目的は蛋白質の結合構造と蛋白質間相互作用およびその元となる性質の関連を明らかにすることにあるため、このデータベースから構築されたテストセットは適していると判断した。

Mintseris によると、このテストセットは以下の手順で構築されている。(1) Protein

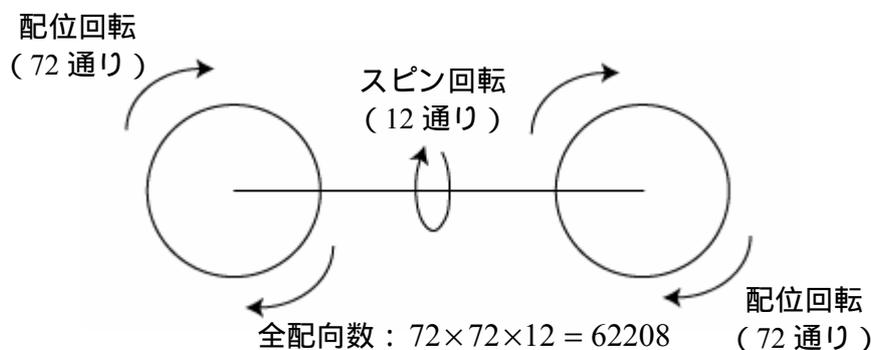
Data Bank (PDB)に結合構造が登録されている蛋白質ペアの中で、30 アミノ酸以上で構成され、単独の蛋白質構造との RMSD が 3.25 Å を超えないものを選ぶ。(2)サブユニット間結合などの恒久的な結合を棄却し、一時的に起こる結合だけを選ぶ。(3)結合部位を含むドメインについて SCOP における FAMILY 分類を取得し、FAMILY-FAMILY ペアが等しい場合に冗長であるとして絞る。なお、一時的な結合の識別は、結合構造から atomic contact vector (ACVs)を計算することにより行われている[14]。

Mintseris らのテストセットは 84 ペアからなる。本研究では計算時間の都合により、分子量の小さい 34 ペアを本研究でのテストセットとし、計算を行った。

2.1.2 計算方法

蛋白質ペアの結合構造をそれぞれの蛋白質に分離して、ソフトウェア AMBER を用いて不足原子を補充し、それぞれ別個に構造の最適化を行った。

最適化した構造を特定の距離に離し、それぞれ回転させて配向を作成した。回転は以下の図のようにそれぞれの蛋白質の幾何中心に関して行う。



まず回転と配位回転は既に生成されている単位球面上の一様分布(72点)を利用して、自分の中心から相手の中心へのベクトルをその一様分布に合うように回転させることで行った(配位回転と呼ぶ)[15]。その後、中心同士を結ぶ軸に関して30°ずつ回転させて生成した(スピン回転と呼ぶ)。また、距離は蛋白質同士を回転させた際にすき間が最短となるときのギャップ長として定義した。

それぞれの配向について AMBER を用いてエネルギー計算を行った。距離は 5 Å ~

15 Å を用い、それぞれについて生理的な塩濃度の場合と塩無しの場合を条件とした。条件をまとめると以下の通りである。

全配向数	62208個
距離	5 Å, 7 Å, 10 Å, 15 Å
塩濃度	0.0 M, 0.2 M

2.1.3 データ処理と解析

各条件下の各配向について結合エネルギー、結合角エネルギー、van der Waals エネルギー、静電エネルギーなどが別々に得られる。このエネルギーの値から各配向の存在確率を計算して主に議論を行う。

また、長距離において真の配向をとる確率と相関をとるために、蛋白質ペアの様々な性質を計算した。相互作用については多重極展開により総電荷相互作用、双極子相互作用などを求めた。形状については溶媒接触表面積 (SASA)、分子量、分子量の比、主軸長の比を求めた。SASA はソフトウェア SURF[16]を用いた。分子量の比 M_{rate} は、

$$M_{rate} = (\text{小さい方の分子量} / \text{大きい方の分子量})$$

で定義し、ペアにおける分子量の違いが少ないほど 1 に近づく。主軸長の比は各蛋白質について慣性主軸 L_{rate} を計算し、

$$L_{rate} = (\text{第三主軸の長さ} / \text{第一主軸の長さ})$$

で定義し、蛋白質の形状が球に近いほど 1 に近づく。

アミノ酸の割合については各蛋白質について全残基数に対する極性残基数の割合 R_{polar} 、酸性残基数の割合 R_{acid} 、塩基性残基数の割合 R_{base} を求めた。

2.2 原理

2.2.1 エネルギー計算

蛋白質のエネルギー計算および構造の最適化は AMBER[17]を用いて行った。パラメータには parm99 を用いた。溶媒の効果を考慮したエネルギーを計算するため、Generalized Born model (GB model)を利用した。塩濃度は 0.0M および 0.2M に設定し、

水の比誘電率は 78.5 とした。

配向は空間上で一様に分布しているとし、また各配向の存在確率はカノニカル分布にしたがうと仮定して、配向 i に対して得られたエネルギー E_i から以下の式で存在確率 P_i を求めた。

$$P_i = \frac{\exp\left(-\frac{E_i}{k_B T}\right)}{\sum_{i'} \exp\left(-\frac{E_{i'}}{k_B T}\right)} \quad (i, i': \text{配向}) \quad (2.1)$$

2.2.2 Generalized Born model (GB model)

蛋白質は細胞内において溶媒である水分子に囲まれる状態で存在しているが、エネルギー計算を行う際に十分な数の水分子を配置すると計算コストが非常に大きくなってしまふ。そこで、GB model を用いることで、水分子を露に記述せずに近似的な水中のポテンシャルエネルギーを計算することができる[18]。GB model は溶質分子を全原子モデルで表し、溶媒は連続体として考え、その影響を関数で近似するモデルである。

溶媒の効果は溶質と溶媒の間で働く van der Waals 力の影響、溶媒の内部に溶質の分子を挿入する空間を作る影響、溶質の各原子が持つ電荷が溶媒に及ぼす影響に分けて考えることができる。この考えに基づき、GB model では溶質の真空中でのポテンシャルエネルギー U_{vac} に蛋白質の形状を反映する自由エネルギー項を加える形で、以下のように溶媒中でのポテンシャルエネルギー U_{tot} を得る[19]。

$$U_{tot} = U_{vac} + \Delta G_{sa} + \Delta G_{el} \quad (2.2)$$

ΔG_{sa} は van der Waals 項と溶媒中に空間を作る効果の項を合わせた電荷によらない自由エネルギー項、 ΔG_{el} は電荷を溶媒中に移行するための自由エネルギー項である。 ΔG_{sa} は溶質の表面積に比例することが知られているため、単純な比例式として近似できる。GB model では ΔG_{el} を以下の式で計算する。

$$4\pi\epsilon_0\Delta G_{el} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_i\sum_j\frac{q_iq_j}{f_{GB}} \quad (2.3)$$

$$f_{GB} = \left[r_{ij}^2 + R_iR_j \exp\left(-\frac{r_{ij}^2}{4R_iR_j}\right) \right]^{\frac{1}{2}} \quad (i, j: \text{原子}) \quad (2.4)$$

r_{ij} は溶質原子 i 、 j 間の距離である。 R_i 、 R_j は有効ボルン半径と呼ばれる変数で、系の形を反映するように任意に定義する。本研究においては Onufriev らが考案したモデルで、幅広い種類の蛋白質で良い近似を示す以下の定義を用いた[20,21]。

$$R_i^{-1} = \tilde{\rho}_i^{-1} - \rho_i^{-1} \tanh(1.0\psi_i - 0.8\psi_i^2 + 4.85\psi_i^3)$$

$$\psi_i = \frac{\tilde{\rho}_i}{4\pi} \int_{solute} \theta(|r| - \tilde{\rho}_i) \frac{1}{r^4} d^3r \quad \tilde{\rho}_i = \rho_i - 0.09\text{\AA} \quad (i: \text{原子}) \quad (2.5)$$

ρ_i は溶質原子 i の van der Waals 半径、 $\theta(X)$ はステップ関数で、 X が 0 より大きければ 1、そうでなければ 0 となる関数である。積分は溶質原子 i の周囲を取り囲む溶質原子群の van der Waals 体積に対して行われる。そのため周りを取り囲む溶質原子群の数が多いほど、つまり溶質原子 i が奥に埋もれているほど R_i の値は大きくなり、表面付近では R_i の値は van der Waals 半径に近くなる。

さらに、塩の影響を考慮する際には式(2.3)の代わりに以下の式を用いる[22]。

$$4\pi\epsilon_0\Delta G_{el} = -\frac{1}{2}\sum_i\sum_j\frac{q_iq_j}{f_{GB}} \left[1 - \frac{\exp(-\kappa f_{GB})}{\epsilon} \right] \quad (i, j: \text{原子}) \quad (2.6)$$

κ は Debye-Hückel screening 変数であり、AMBER においては

$$\kappa = 0.316\sqrt{[salt]} \quad (2.7)$$

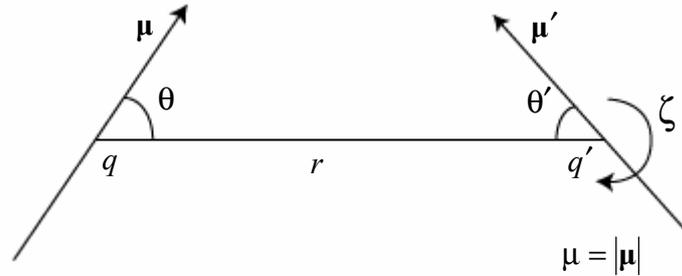
が用いられる。

2.2.3 静電場の多重極展開

電荷分布による静電場は点電荷、双極子モーメント、四重極子モーメント、さらに高次モーメントの重ね合わせとして近似的に記述できることが知られており、この近

似は静電場の多重極展開と呼ばれる。これを用いて、電荷分布同士の相互作用による静電エネルギー V を以下の式で表すことができる[10]。

$$\begin{aligned}
 V &= \frac{1}{4\pi\epsilon} [V_{c-c} + V_{c-d} + V_{d-d} + \dots] \\
 V_{c-c} &= \frac{qq'}{r} \\
 V_{c-d} &= \frac{1}{r^2} (q\mu' \cos\theta + q'\mu \cos\theta') \\
 V_{d-d} &= \frac{\mu\mu'}{r^3} V_{d-d(scaled)} \\
 V_{d-d(scaled)} &= 2 \cos\theta \cos\theta' + \sin\theta \sin\theta' \cos\zeta
 \end{aligned} \tag{2.8}$$



q は総電荷、 μ は双極子モーメント、 θ は双極子モーメントと中心同士を結ぶ線がなす角度、 ζ は二つの双極子モーメントのねじれ角を表す（上図）。この展開により、静電エネルギーを総電荷相互作用 V_{c-c} 、総電荷-双極子相互作用 V_{c-d} 、双極子相互作用 V_{d-d} 、あるいはさらに高次の多重極子の組み合わせの寄与として分けて考えることができる。 V_{d-d} の係数を除いた $V_{d-d(scaled)}$ （スケールされた双極子相互作用と呼ぶ）は $-2 \sim +2$ の間の値を取り、双極子モーメント同士の立体的な向きを表す指標となる。この値が低いほどエネルギー的に安定な向きであり、 -2 で最も安定な向きとなる。天然において総電荷相互作用や双極子相互作用は水分子やイオンに遮蔽されてしまうため評価が行いにくいだが、電荷の偏りからもたらされる双極子モーメントの向きは遮蔽によらないため、配向と相関をとるのに適した指標であると考えられる。

点電荷および双極子モーメントは以下のように計算した。

$$\begin{aligned}
 q &= \sum q_i \\
 \boldsymbol{\mu} &= \sum q_i \mathbf{r}_i \quad (i: \text{原子})
 \end{aligned}
 \tag{2.9}$$

q_i には AMBER の parm99 から得られる電荷を用いた。一般に双極子モーメントは \mathbf{r}_i の原点の取り方に依存するが、今回は蛋白質を幾何中心に関して回転させてエネルギーを計算するので、双極子モーメントも同様に各蛋白質の幾何中心を原点とした。

2.2.4 相関係数

相関係数 ρ には以下の式で定義される相関係数を用いた[23]。

$$\rho = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{\sum (x_i - \bar{x})^2}{N}} \sqrt{\frac{\sum (y_i - \bar{y})^2}{N}}} \quad (i: \text{標本})
 \tag{2.10}$$

N は標本数、 x と y は相関を見たいデータである。標本から得られた相関係数が母集団、つまり一般的な蛋白質ペアに対して有意であるかを検定するため、 t 検定を行った。母集団の相関係数 ρ_m に対して以下の帰無仮説を立てる。

帰無仮説：母集団の相関係数 ρ_m は 0 である。

ρ_m が 0 であるなら、標本の相関係数 ρ から以下の式で求められる t の値は自由度 $(N-2)$ の t 分布にしたがう。

$$t = \frac{\rho \sqrt{N-2}}{\sqrt{1-\rho^2}}
 \tag{2.11}$$

この t の両側確率 P が α より小さければ、帰無仮説は有意水準 α で棄却される。このとき母集団の相関係数 ρ_m は 0 ではないと言え、統計的に有意であることが示される。

3. 結果と考察

各蛋白質ペアの静的性質を表 3.1 に、テストセット全体の性質を表 3.2 に示した。

3.1 回転によるエネルギー変化

各条件、各配向についてポテンシャルエネルギーを計算した。それぞれの蛋白質の中心に関して回転を行ったので、蛋白質内での原子間の結合長や結合角のエネルギーは変化せず、蛋白質間に働く van der Waals エネルギーと静電エネルギーのみが変化する。それらの標準偏差を表 3.3 に示した。標準偏差は同じ蛋白質ペアのエネルギーについて求めているので、回転によってもたらされるエネルギー変化を表す。

本研究で用いた蛋白質間距離 5 Å から 15 Å の範囲では、van der Waals エネルギー変化に比べて静電エネルギー変化が 5 倍から 20 倍大きく、エネルギー変化への寄与は静電エネルギーが主となっていることがわかる。以下ではエネルギーとして van der Waals エネルギーと静電エネルギーの合計を用いる。

表 3.3 よりエネルギー変化の標準偏差が 0.1 kcal/mol 程度になっており、この値は熱ゆらぎで埋もれてしまう程度の値である。しかしながら、今回の計算によって真の配向においてランダムに回転している場合とは異なる結果が得られているため、長距離での影響が及んでいると考えることができる。これは長距離での配向探索に対してエネルギー変化の値そのものが効いているだけではなく、それに起因するその他の要素が同時に効いているためであると予想できる。例えば Higo らの示したように、蛋白質間で水分子が双極子モーメントが揃うように配向する現象があり、これが蛋白質間の長距離相互作用の効果を増幅させている可能性がある[24]。

3.2 距離に依存する確率分布の変化

図 3.1 にエネルギーから計算した存在確率の分布を表すヒストグラムを示した。ただし存在確率 P に全配向数 N を掛けて $P' = NP$ とスケールした P' を横軸としている。 P' は蛋白質間相互作用の働かない、つまり配向に依存するエネルギー変化が起こらな

い無限遠において 1 となる。この図より、距離が近いほどヒストグラムが幅広化していることがわかる。三つの蛋白質ペアについて示したが、他のペアでも同様の変化が見られる。これは、距離が近づくにつれて高エネルギーと低エネルギーの差が大きくなっていることを表している。

確率分布の変化を定量的に見るため、下式のように全配向 (62208 個) のエネルギーに対して、エネルギーの低い順に 50% の配向 (31104 個) を取る確率 P_{half} を求めた。

$$P_{half} = \frac{\sum_{i=1}^{N/2} \exp(-\beta E_i)}{\sum_i \exp(-\beta E_i)} \quad (i: \text{配向}) \quad (3.1)$$

図 3.2 は全てのペアについて P_{half} を求め、グラフにしたものである。縦軸は P_{half} 、横軸は蛋白質間の距離である。 P_{half} は大きいほど低エネルギーの配向をとる確率が高いということになる。この図より、全てのペアにおいて距離が近づくほど P_{half} の値が大きくなることが確認できる。つまり、蛋白質同士が遠距離にあるとそれぞれは自由に回転するが、距離が近づくにつれて徐々に低エネルギーの安定配向に拘束されていくことを表しており、このことは蛋白質の結合過程において長距離で探索範囲が限定されていくことを示唆している。

3.3 真の配向は長距離における最安定構造ではない

各ペアについて全配向エネルギーを低い順に並べたとき、真の配向のエネルギーが上位から何番目にいるかを調べた。それを全配向数に対するパーセンテージとして表したものを、図 3.3 に距離別の棒グラフで上位からの割合として示した。これにより真の配向は長距離において最安定ではなく、平均では上位約 30% 程度となっていることがわかる。また、多くのペア (15 と 5 を比較したとき、34 ペア中 28 ペア) において距離が近づくとき真の配向の順位が上昇していることがわかる。最安定ではないということと、確率的に不利であることは一致しない。例えば真の配向が中央の順

位にいても、それより下の順位の配向が大きく不利なエネルギーを持っていれば、式 (2.1)より真の配向をとる確率は相対的に上昇する（次項）。

3.4 真の配向は長距離において確率的に有利である

長距離相互作用が効いているとき、ランダムな回転をしている場合に比べて真の配向およびその付近がどの程度有利になっているかを調べた（図 3.4）。上位 1 個とは蛋白質の真の配向の配位回転がお互いの相手の中心に向いた状態とし、上位 10%は真の配向の配位回転に近い 10%（72 個中 7 個）のどれかがお互いの相手の中心に向いた状態とする。配位回転の近い 10%とは、各配位回転を単位ベクトルに対して適用した結果、真の配向に対応する配位回転と距離が近くなる 7つの配位回転として定義した。同じ配位回転におけるスピン回転に関してはエネルギーを足し合わせて確率を計算した（配位回転とスピン回転の定義は 2.1.2 項参照）。つまり、上位 1 個では蛋白質同士が真の結合部位同士を向き合せている状態、上位 10%では真の結合部位の周辺同士を向き合せている状態に対応する。図 3.4 では下式で表されるように、エネルギーから求めた確率をランダム回転としたときの確率でスケールしている。この値を真の配向の有利性 A_{true} と呼ぶ。

$$A_{true} = \frac{\frac{\sum_j^n \exp(-\beta E_j)}{j}}{\frac{\sum_i^N \exp(-\beta E_i)}{i}} \cdot \frac{n}{N} \quad \begin{array}{l} N : \text{全配向数} \\ n : \text{上位の配向数} \\ E_i : \text{各配向 } i \text{ のエネルギー} \\ E_j : \text{上位配向 } j \text{ のエネルギー} \end{array} \quad (3.2)$$

A_{true} は 1 でランダム回転と等しく、1 より大きいほどランダム回転より確率が高いこととなる。

表 3.4 に示したように、多くの場合（5 Å において上位 1 個では 34 個中 28 個、上位 10%では 34 個中 27 個）で A_{true} は 1 以上になっており、ランダム回転よりも確率が上昇しているといえる。平均値では約 9%上昇している（5 Å・上位 1 個の場合）。この結果より、接触前の長距離において真の配向をとる確率が高いことが示された。

蛋白質の結合は一回の試行で行われているわけではなく、蛋白質同士が複数いる状態で複数回の試行により真の結合を探索していると考えられる。わずかに 9%の上昇であっても、例えば 10 回試行した場合にはランダムな試行に比べて確率が $(1.09)^{10} \approx 2.4$ 倍にも上昇するため、結合構造探索に効いている可能性がある。

ランダム回転より確率が低いペアは 34 個中 1AK4、1D6R、1EAW、1FQJ、1ML0、2SIC の 6 個である。表 3.1 よりこれらに共通する性質を探すと、1EAW を除いてはスケールされた双極子相互作用 $V_{d-d(scaled)}$ が高い (1EAW 以外は全て上位 9 位以内) ことが挙げられる。これは双極子モーメント同士が不安定な向きになっていることに対応する。つまり、真の配向において双極子モーメント同士の向きが不安定になるペアは、長距離において真の配向を探索できる確率が低いことを意味する。このことは 3.6 項において示す相関係数としても確認できる。

3.5 真の配向の有利性は塩の有無に依存する

前項では生理的な塩濃度 0.2M での結果を示したが、塩を考慮しない場合についての真の配向の有利性 A_{true} の計算結果を表 3.5 に示した。表 3.4 と比較すると、蛋白質間距離 5 の場合、ランダム回転より真の配向が有利になるペアの割合が塩濃度 0.2M では約 80%に対して、塩無しでは約 55%に減少している。式(2.6)より塩濃度は電荷による静電エネルギーの値に影響を与えるため、この結果は多くの蛋白質ペアにおいて生理的な塩濃度で蛋白質同士の配向が決まりやすい電荷分布になっていることを表す。このことは生体内で長距離での静電相互作用が真の配向の探索効率の向上に効いていることを示唆している。

3.6 真の配向になる確率と静的な性質との相関

表 3.3 に示した静的性質と真の配向の有利性 A_{true} に関連があるかどうかを調べるため、表 3.4 の塩濃度 0.2M、上位 1 個の結果に対して相関係数を計算した(表 3.6)。相関係数の有意検定により 5%の有意水準で有意であるとされたものは橙地に赤字、10%

の有意水準のものは赤字で示している（ただし、検定により有意であるかどうかは実際の相関の有無を表しているのではなく、あくまで今回の標本数から得られた相関係数の大きさが母集団の相関を推定するために統計的に足りるかということを表す）。

まず、テストセットの全てを用いて相関係数を求めた(表中の ALL 列)。この結果から特にスケールされた双極子相互作用 $V_{d-d(scaled)}$ 、ペアのうち小さい方の主軸長の比 L_{rate} 、小さい方の酸性残基の割合 R_{acid} 、大きい方の塩基性残基の割合 R_{base} に有意な相関があることを見出した。これらの実際の分布を確認するため、分布図を示した(図 3.5)。7CEI、1UDI、2PCC が大きく離れた位置にいるため、今度はこれらを外れ値と見なして残りのセットで相関係数を取ったものを計算した(表中の ALL-3 列)。この操作でも $V_{d-d(scaled)}$ 、小さい方の L_{rate} 、小さい方の R_{acid} の相関は崩れなかったため、これらの性質は外れ値に引きずられた見かけの相関ではなく、真の相関である可能性が高いといえる。

次に、酵素-阻害剤のペア(表中の E 列)およびその他のペア(表中の O 列)に分類されるものだけのセットで相関係数を求めた。抗体-抗原ペアについては数が少ないので計算を行わない。酵素-阻害剤のペアに関しては、総電荷相互作用、双極子相互作用、小さい方の R_{acid} 、大きい方の R_{base} に有意な相関が見られる。その他のペアに関しては $V_{d-d(scaled)}$ に有意な相関が見られる。よって、酸性/塩基性残基の割合は特に酵素-阻害剤のペアと関連のある指標であることがわかる。酵素-阻害剤の結合は酵素の本来の結合相手である基質の結合よりも強固である場合が多い。そのため、より強力に静電相互作用が働いている傾向があると考えられ、それが相互作用の値や酸性・塩基性残基の割合での有意な相関として表れているのであろう。これは通常の蛋白質間結合過程を反映していない危険性がある。

小さい方の L_{rate} は全てのセットで一定の相関を見せている。よって、これは一般に採用できる静的性質となりうることを示唆している。 L_{rate} は大きいほど蛋白質の形状が球に近いことを表すので、小さい方の蛋白質の形状が丸いほど長距離相互作用エネルギーによる真の配向の確率が高くなっているということである。蛋白質の形状が丸

いということは形状の特異性による結合構造探索がしにくいということを意味しており、そのような蛋白質では長距離相互作用が配向が決める性質として重要になっているのかもしれない。したがって、 L_{rate} は長距離相互作用が結合構造探索に効くペアと効かないペアを見分ける静的性質となりうる。

$V_{d-d(scaled)}$ は酵素-阻害剤のペアで有意水準から外れているものの、各セットにおいて相関を見せている。そのため、これも一般の蛋白質に採用できる静的性質となりうる。 $V_{d-d(scaled)}$ は各蛋白質内の双極子モーメント同士の相対的な向きを表し、マイナスに大きいほどエネルギー的に安定になる指標なので、負の相関になっている。つまり双極子モーメント同士が安定な向きになっているほど、長距離において真の配向になる確率が高くなる。ただし $V_{d-d(scaled)}$ は L_{rate} とは異なり単体から得られる静的性質ではなく、蛋白質同士の相対的な向きが計算に必要な。そのため、長距離で配向が決まると予測された蛋白質ペアに対して、 $V_{d-d(scaled)}$ が有利になる配向が真の結合配向である確率が高いと見なす利用法が考えられる。 $V_{d-d(scaled)}$ と長距離が負の線形関係になっているということは、長距離で真の配向をとっている条件のもとでは双極子モーメント同士が安定な向きになっている確率が高いことを意味する。ただし、 $V_{d-d(scaled)}$ の値が最小の場所が真の配向であるとは限らないので、今後の研究で数値的な関係として記述することが必要になる。双極子モーメントの計算は原子数 N に対して $O(N)$ で行えるので、全原子を用いてエネルギー計算を行う場合の $O(N^2)$ よりも計算時間を大きく短縮できる。例えば結合構造探索アルゴリズムにおいて、双極子モーメント同士の向きを静電エネルギーの代替として用いる、もしくは向きが合っている場所に絞って探索を行うなどの利用が期待できる。

L_{rate} と $V_{d-d(scaled)}$ が種類を問わず相関を示す指標であることを確認するためには、本研究で用いたようなアミノ酸配列と立体構造から選んだセットではなく、着目している L_{rate} 、 $V_{d-d(scaled)}$ の値に対して広い範囲をカバーしたセットで試行する必要がある。その結果としてこれらの性質と長距離における真の配向の有利性の線形な分布が得られれば、種類を問わず適用できる指標であることを確認できる。

4. 結論

全体のまとめ

本研究では、まず蛋白質同士の距離が近づくにつれて各配向の確率分布が変化することを示した。この結果から、結合過程において蛋白質間の距離が近づくにつれて徐々に結合構造の探索範囲が狭まっていくことが推測できる。そして真の結合配向は長距離においてエネルギー的に最安定にはなっていないが、特に生理的な塩濃度下では確率的に多くのペアで有利になっていることを示した。このことは長距離においても真の配向を探索できることを意味する。さらに長距離での真の配向の有利性と各蛋白質の静的性質の相関係数を計算した。その結果、特に双極子モーメント同士が安定な向きを向いていることを表す「スケールされた双極子相互作用」と蛋白質の形が球に近いことを表す「主軸長の比」に対して有意な相関が見られた。これらは種類別のサブセットに対しても相関を見せたため、蛋白質の種類によらず結合構造予測に利用できる指標となる可能性がある。

今後の課題

本研究では 34 個の蛋白質ペアのテストセットを用いたが、特定の性質に着目したサブセットを用いて相関を見る際にはデータ数および値の分布が十分であるとはいえない。有効であると予想された指標に対して、その値に着目したサブセットと新たに相関をとることで、実際に相関があることの確実性を高めることができる。今後はさらに標本の数を増やして相関係数自体の信頼性を向上させると共に、様々な指標でのサブセットに対して相関を調べ、結合構造予測に利用可能な確実性のある指標を探索していくことが必要である。またそうして得られた有効な指標を数値的な形で利用できるようにして、今後の結合予測アルゴリズムで利用できるようにすることも重要な課題である。

謝辞

本研究にあたって北尾彰朗准教授に度重なる相談、議論を通して研究内容、研究方法について助言をいただきました。城地保昌助教には研究に関する様々な物理理論についてご教授いただきました。論文執筆にあたって竹村和浩博士、桜庭俊さんには内容に対する多大なご指導をいただきました。研究室のメンバーである Lee-Wei Yang 博士、Choon-Peng Chng さん、二島渉さん、田口淳子さん、原田隆平さん、原真佐夫さん、Yuexi Ding さんには、セミナーや日頃の会話を通して助けていただきました。また、秘書の方々、過去に研究室を訪れた方々にも研究を支えていただきました。皆様に心からお礼を申し上げます。

参考文献

- [1] Sousa SF, Fernandes PA, Ramos MJ. Protein–Ligand Docking: Current Status and Future Challenges. *Proteins*. 2006; 65(1):15-26.
- [2] Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, Wodak SJ. CAPRI: A Critical Assessment of PRedicted Interactions. *Proteins*. 2003; 52(1):2-9.
- [3] Gray JJ. High-resolution protein–protein docking. *Curr Opin Struct Biol*. 2006; 16(2):183-93.
- [4] Bonvin AM. Flexible protein–protein docking. *Curr Opin Struct Biol*. 2006; 16(2):194-200.
- [5] Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins*. 2002; 47(4):409-43.
- [6] Gabb HA, Jackson RM, Sternberg MJ. Modelling Protein Docking using Shape Complementarity, Electrostatics and Biochemical Information. *J Mol Biol*. 1997; 272(1):106-20.
- [7] Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations. *J Mol Biol*. 2003; 331(1):281-99.
- [8] Fernández-Recio J, Totrov M, Abagyan R. ICM-DISCO Docking by Global Energy Optimization With Fully Flexible Side-Chains Proteins. 2003; 52(1):113-7.
- [9] Dominguez C, Boelens R, Bonvin AM. HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information *J Am Chem Soc*. 2003; 125(7):1731-7.
- [10] Buckingham AD. MOLECULAR QUADRUPOLE MOMENTS. *Quarterly Reviews of the Chemical Society* 1959; 13:183-214.
- [11] Ramirez-Carrozzi VR, Kerppola TK. Long-range electrostatic interactions influence the orientation of Fos-Jun binding at AP-1 sites. *J Mol Biol*. 2001; 305(3):411-27.

- [12] Mintseris J, Wiehe K, Pierce B, Anderson R, Chen R, Janin J, Weng Z. Protein-Protein Docking Benchmark 2.0: an update. *Proteins*. 2005; 60(2):214-6.
- [13] Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol*. 1995; 247(4):536-40.
- [14] Mintseris J, Weng Z. Atomic contact vectors in protein-protein recognition. *Proteins*. 2003; 53(3):629-39.
- [15] Hardin RH, Sloane NJA, Smith WD. Spherical Coverings.
<http://www.research.att.com/~njas/coverings/>
- [16] Gaudio AC, Takahata Y. SURF.
http://server.ccl.net/cca/software/SOURCES/FORTRAN/molecular_surface/surf/
- [17] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman (2006), AMBER 9, University of California, San Francisco.
- [18] Still WC, Tempezyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics *J Am Chem Soc*. 1990; 112:6127-9.
- [19] Berendsen HJC. *Simulating the Physical World*. 2007
- [20] Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins*. 2004; 55(2):383-94.
- [21] Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL 3rd. Performance comparison of the generalized Born and Poisson methods in the calculation of the electrostatic solvation energies for protein structures *J Comput Chem*. 2004;

25(2):265-84.

- [22] Srinivasan J, Trevathan MW, Beroza P, Case DA. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects *Theor Chem Acc.* 1999; 101:426-34.
- [23] 小寺平治. 新統計入門. 1996.
- [24] Higo J, Sasai M, Shirai H, Nakamura H, Kugimiya T. Large vortex-like structure of dipole field in computer models of liquid water and dipole-bridge between biomolecules. *Proc Natl Acad Sci U S A.* 2001; 98(11):5961-4.

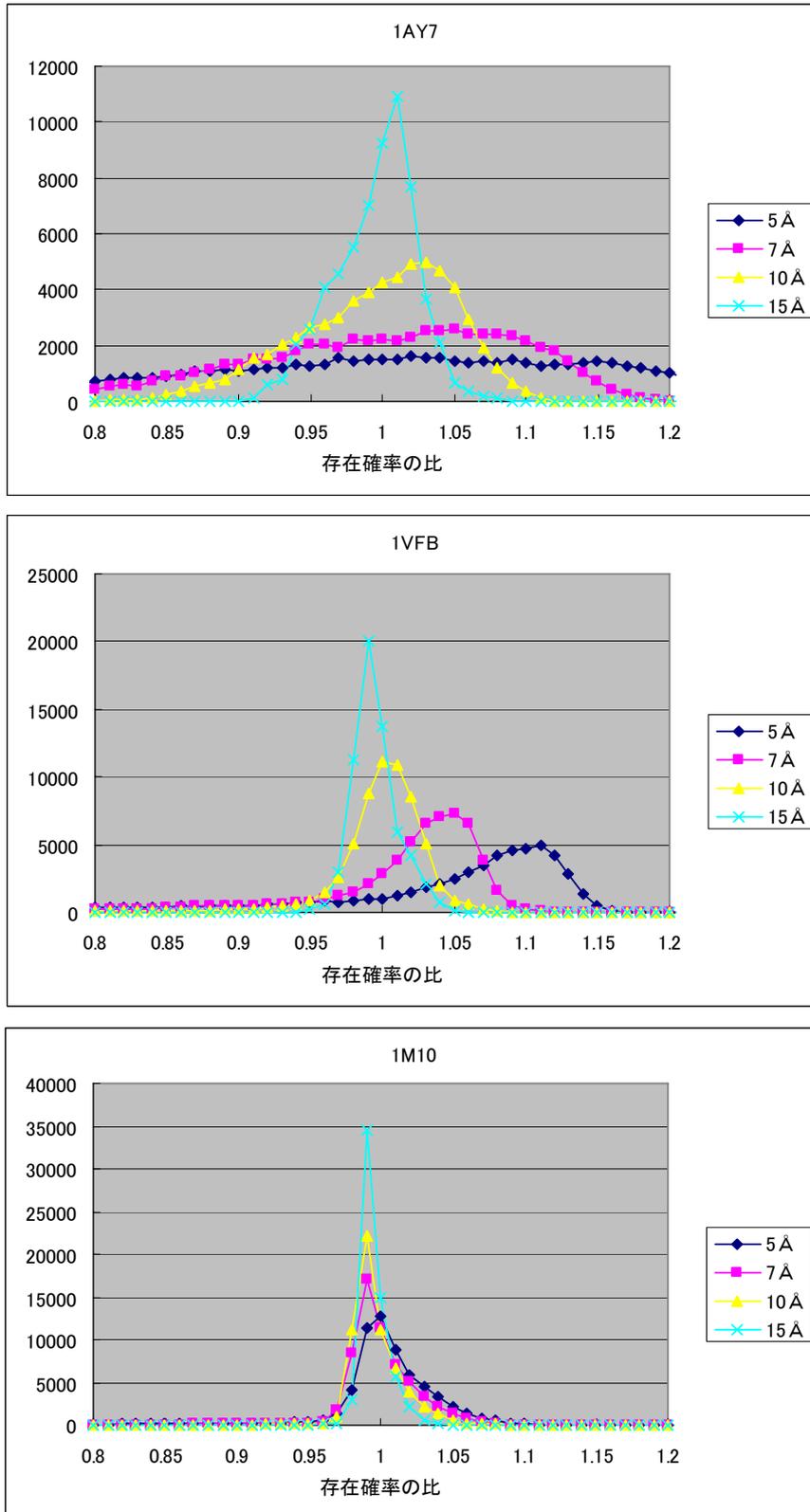


図 3.1 各配向の存在確率のヒストグラム。テストセットのうち 1AY7、1VFB、1M10 のペアについて計算結果を示した。存在確率 P に全配向数 N を掛けて $P' = NP$ とスケールした P' を横軸としている。 P' は蛋白質間相互作用の働かない、つまり配向に依存するエネルギー変化が起こらない無限遠において 1 となる。蛋白質間距離が近くなるほど分布が幅広くなっているのがわかる。

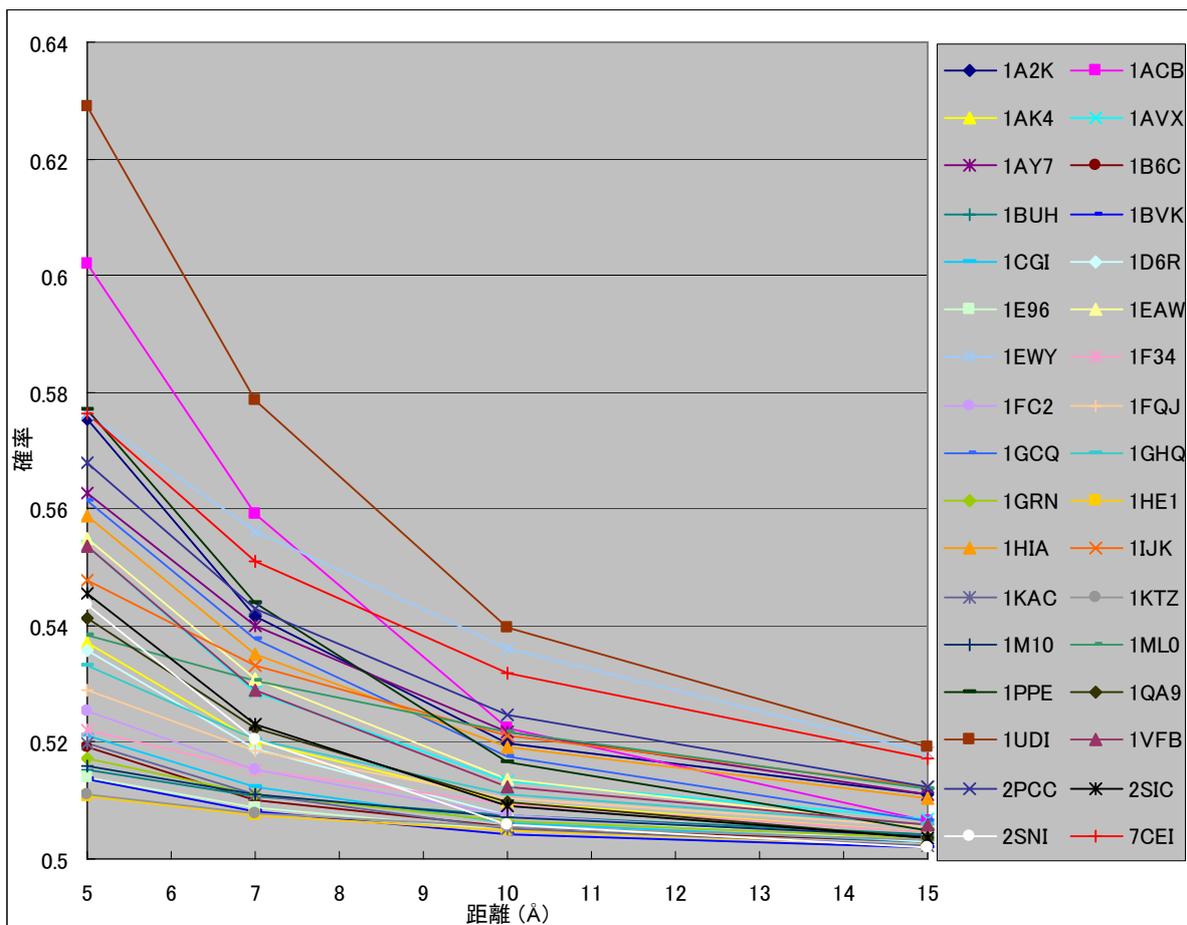


図 3.2 距離に依存する確率 P_{half} の変化。横軸に蛋白質間距離、縦軸に全配向の 50% のエネルギーを低い順に足した確率をとった。確率が 0.5 のときは蛋白質同士を無限遠に置いて、回転によって各配向のエネルギー分布に変化がない場合に等しい。全ての蛋白質ペアにおいて距離が近づくにつれて確率の値が大きくなっている。

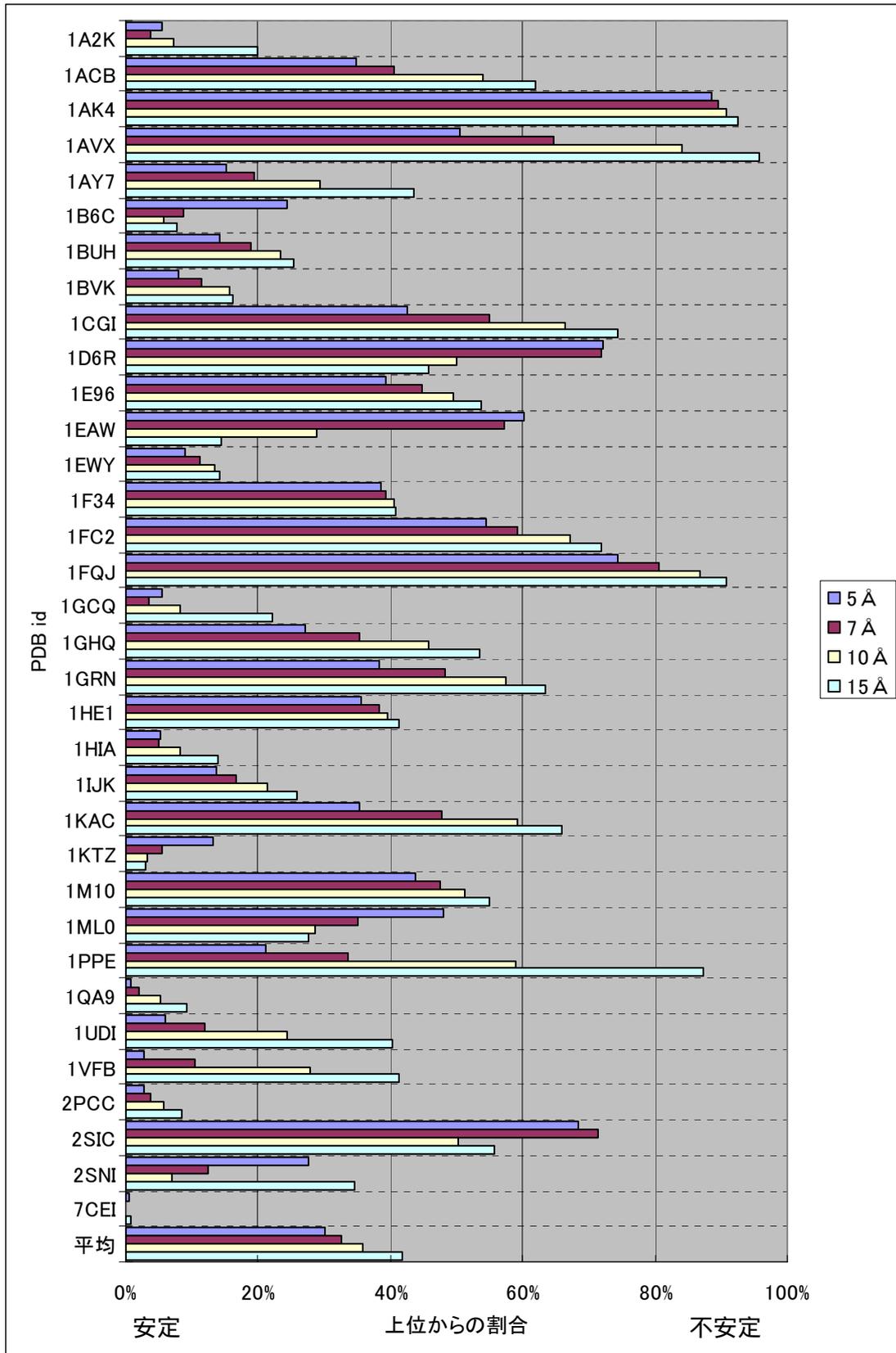


図 3.3 真の配向の順位。全配向 (62208 個) のエネルギーを低い順に並べたときの真の配向 (1 個) の順位を上位からのパーセンテージとして表し、横軸とした (塩濃度 0.2 M)。平均は全ペアの値を平均にしたもので、上位から約 30%程度となっている。

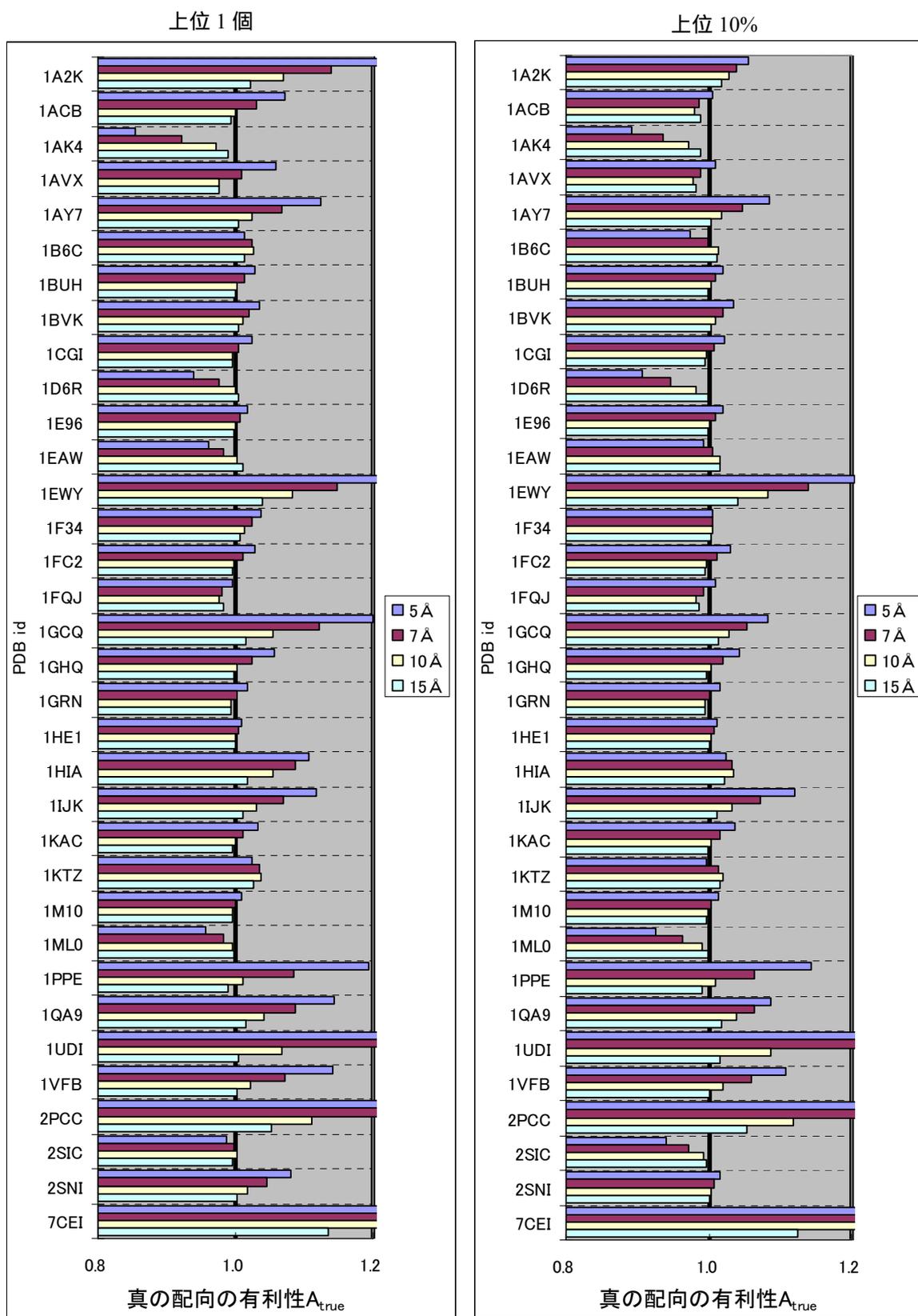
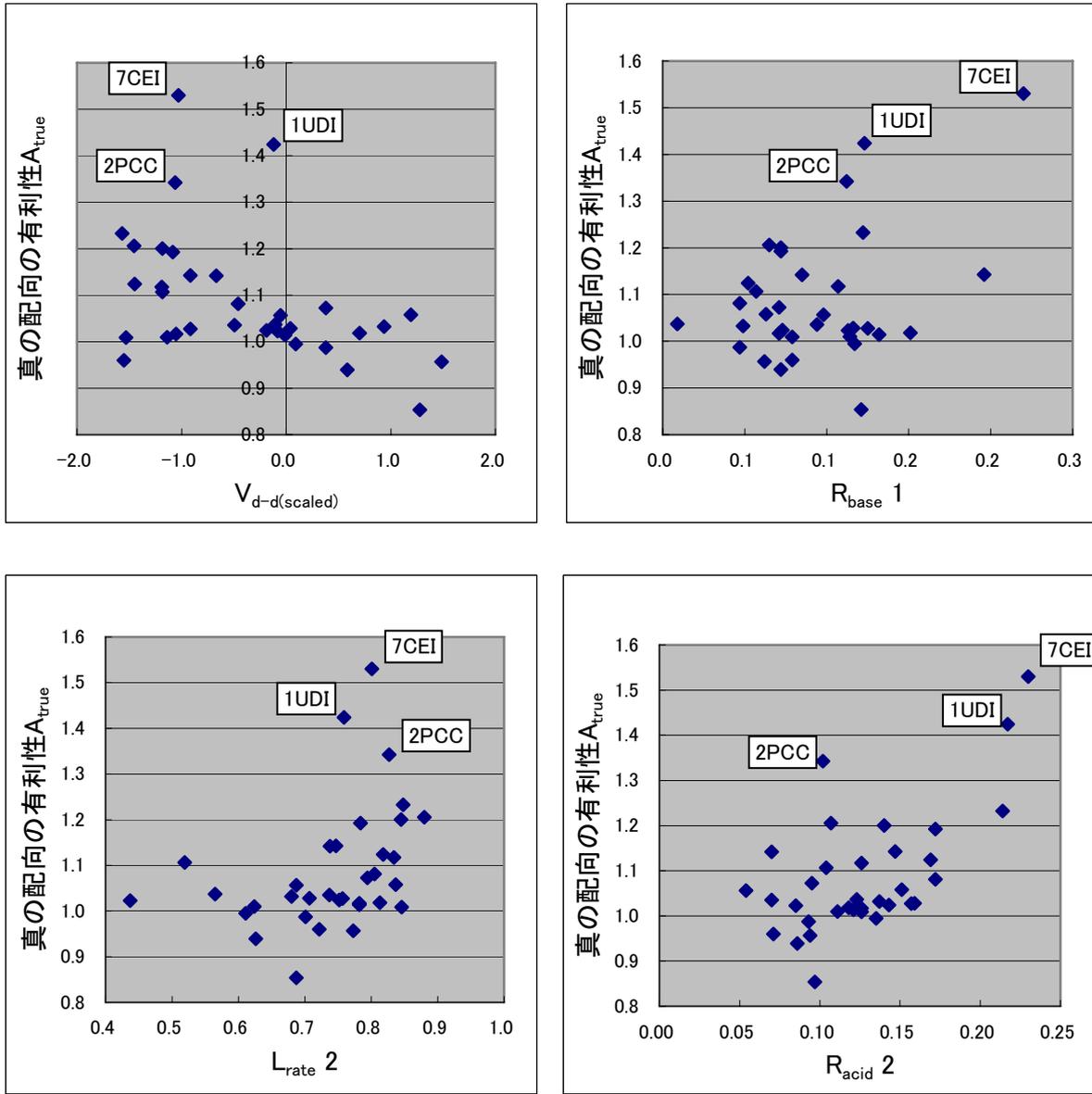


図 3.4 長距離における真の配向の有利性 A_{true} 。ランダムな回転で真の配向（もしくはその付近）となる確率に対して、エネルギーから求めた確率の比をとったもの真の配向の有利性として定義し、グラフにした。1.0 より大きければランダム回転より有利になっているといえる。



(全て単位は無次元)

図 3.5 真の配向の有利性 A_{true} に対する静的性質の分布。縦軸を真の配向の有利性、横軸を相関を見たいデータとしてプロットを行った。真の配向の有利性の上位3つ (7CEI、1UDI、2PCC) が特に外れていることがわかるため、これらを外れ値とした。各データは表 3.3 の値を用いている。

PDB	type	相互作用			形状							アミノ酸					
		V _{c-c}	V _{d-d}	V _{d-d} (scaled)	SASA 1	SASA 2	Mass 1	Mass 2	M _{rate}	L _{rate} 1	L _{rate} 2	R _{polar} 1	R _{polar} 2	R _{acid} 1	R _{acid} 2	R _{base} 1	R _{base} 2
1BVK	A	265.6	-1.3	-0.49	12	7	25	14	0.58	0.75	0.74	55%	55%	7%	7%	9%	13%
1VFB	A	358.6	-2.2	-0.67	11	7	25	14	0.58	0.75	0.74	56%	55%	7%	7%	9%	13%
1ACB	E	0.0	4.5	0.38	12	5	25	7	0.29	0.87	0.79	49%	48%	6%	10%	7%	10%
1AVX	E	-159.4	6.5	1.19	11	10	23	19	0.81	0.81	0.84	50%	48%	5%	15%	6%	11%
1AY7	E	278.9	-20.7	-1.45	6	5	11	10	0.97	0.87	0.82	56%	53%	13%	17%	5%	10%
1CGI	E	-26.6	-1.8	-0.19	12	4	26	6	0.25	0.87	0.75	49%	68%	6%	14%	7%	13%
1D6R	E	39.8	1.1	0.59	10	5	23	6	0.28	0.80	0.63	53%	69%	5%	9%	7%	10%
1EAW	E	-239.1	-4.2	-1.55	12	4	26	6	0.24	0.87	0.72	44%	57%	10%	7%	8%	18%
1EWY	E	318.8	-39.3	-1.57	17	6	34	11	0.31	0.80	0.85	49%	56%	13%	21%	12%	5%
1F34	E	1295.0	-0.7	-0.11	15	10	34	15	0.45	0.71	0.57	50%	47%	13%	12%	1%	9%
1HIA	E	-318.8	-4.3	-1.18	11	4	25	5	0.20	0.83	0.52	47%	67%	13%	10%	6%	17%
1PPE	E	-39.8	-4.1	-1.08	10	3	23	3	0.14	0.81	0.78	53%	59%	5%	17%	7%	14%
1UDI	E	-637.5	-2.0	-0.12	12	6	26	9	0.36	0.77	0.76	42%	54%	9%	22%	12%	7%
2PCC	E	-584.4	-54.0	-1.06	15	7	33	12	0.36	0.80	0.83	51%	55%	15%	10%	11%	18%
2SIC	E	53.1	1.8	0.38	11	7	28	11	0.40	0.91	0.70	42%	40%	6%	9%	5%	6%
2SNI	E	6.6	-0.6	-0.46	11	5	28	7	0.27	0.91	0.81	42%	45%	5%	17%	5%	16%
7CEI	E	-730.5	-20.8	-1.03	9	6	15	10	0.68	0.86	0.80	58%	54%	13%	23%	22%	12%
1A2K	O	-398.5	-18.5	-1.46	13	11	28	22	0.79	0.79	0.88	47%	45%	11%	11%	7%	14%
1AK4	O	66.4	20.0	1.28	9	10	18	16	0.89	0.89	0.69	46%	41%	12%	10%	12%	8%
1B6C	O	26.6	0.0	-0.01	17	7	37	12	0.32	0.78	0.78	49%	46%	12%	12%	13%	13%
1BUH	O	-99.6	-23.7	-0.92	16	5	33	9	0.26	0.71	0.76	44%	49%	11%	16%	13%	11%
1E96	O	-33.2	2.6	0.70	11	10	21	20	0.92	0.64	0.81	50%	48%	12%	12%	15%	11%
1FC2	O	-79.7	0.4	0.04	14	4	24	5	0.22	0.53	0.71	56%	57%	10%	16%	12%	7%
1FQJ	O	-511.4	0.5	0.09	17	9	36	16	0.43	0.68	0.61	54%	51%	15%	14%	12%	19%
1GCQ	O	59.8	-6.4	-1.18	5	4	8	7	0.84	0.87	0.85	39%	47%	12%	14%	7%	9%
1GHQ	O	-212.5	-1.5	-0.06	15	9	34	14	0.41	0.84	0.69	47%	54%	11%	5%	10%	12%
1GRN	O	39.8	-17.9	-1.05	11	12	22	21	0.95	0.77	0.78	45%	51%	10%	13%	7%	12%
1HE1	O	13.3	-16.4	-1.14	10	8	20	14	0.73	0.82	0.62	48%	45%	12%	11%	11%	10%
1IJK	O	-431.7	-173.0	-1.19	14	11	30	23	0.77	0.49	0.83	57%	49%	16%	13%	11%	15%
1KAC	O	0.0	1.9	0.94	10	8	20	14	0.68	0.75	0.68	51%	52%	5%	14%	5%	11%
1KTZ	O	-79.7	-0.2	-0.08	7	7	12	9	0.77	0.82	0.44	64%	55%	17%	9%	11%	11%
1M10	O	-106.3	-50.4	-1.53	15	11	30	23	0.77	0.56	0.85	47%	49%	9%	13%	8%	15%
1ML0	O	-836.8	97.2	1.49	20	5	41	7	0.18	0.69	0.77	50%	63%	12%	9%	6%	19%
1QA9	O	19.9	-5.9	-0.92	8	7	12	11	0.92	0.69	0.75	63%	62%	21%	15%	20%	12%

表 3.1 テストセットの静的性質。全蛋白質結合構造の様々な性質を示した。type は各ペアの分類で、A は抗体-抗原、E は酵素-阻害剤、O はそれ以外（酵素-基質やドメイン間結合など）を表す。

性質名	説明 ^a
V _{c-c}	真空中の総電荷相互作用の値 (kcal/mol)
V _{d-d}	真空中の双極子相互作用の値 (kcal/mol)
V _{d-d(scaled)}	スケールされた双極子相互作用
SASA 1	SASA (溶媒接触表面積) 1 (1000 Å ³)
SASA 2	SASA (溶媒接触表面積) 2 (1000 Å ³)
Mass 1	分子量 1 (kDa)
Mass 2	分子量 2 (kDa)
M _{rate}	分子量の違い: (分子量 2)/(分子量 1)
L _{rate} 1	主軸長の比 1 ^b
L _{rate} 2	主軸長の比 2 ^b
R _{polar} 1	極性残基の割合 1
R _{polar} 2	極性残基の割合 2
R _{acid} 1	酸性残基の割合 1
R _{acid} 2	酸性残基の割合 2
R _{base} 1	塩基性残基の割合 1
R _{base} 2	塩基性残基の割合 2

- ^a 1はペアのうち分子量が大きい方、
2は小さい方についてとった値。
最長の慣性主軸に対する
^b 最短の慣性主軸の長さの比。

赤字は各性質の最大値
青字は各性質の最小値

表 3.2 データセット全体の性質。セット内の最小値と最大値を示した。

データ数	34個
分子量	15 kDa ~ 52 kDa
原子数	1968個 ~ 7476個
残基数	126個 ~ 466個

表 3.3 回転によって起こるエネルギー変化の標準偏差。各蛋白質ペアに対して全配向の van der Waals エネルギーおよび静電エネルギーの標準偏差を計算し、全蛋白質ペアの平均をとった。これは回転によって蛋白質間エネルギーがどの程度変化するかを表す。

塩濃度 距離	0.0 M				0.2M			
	5 Å	7 Å	10 Å	15 Å	5 Å	7 Å	10 Å	15 Å
van der Waals エネルギー	0.029	0.018	0.010	0.004	0.029	0.018	0.010	0.004
静電エネルギー	0.533	0.483	0.431	0.367	0.128	0.073	0.032	0.011

(kcal/mol)

表 3.4 生理的な塩濃度 0.2 M の場合における真の配向の有利性 A_{true} 。ランダムな回転で真の配向もしくはその周辺となる確率に対して、エネルギーから求めた確率の比を計算した。図 3.4 のグラフに対応する。下部に平均値とランダム回転より有利（つまり 1 以上）になっているペアの割合を示した。上位 1 個・5 Å において、平均としてランダム確率より 9% 上昇している。10 Å 以内の距離では約 70% 以上のペアでランダム回転より有利になっていることがわかる。

塩濃度	0.2 M								
	配向	上位1個				上位10%			
		5 Å	7 Å	10 Å	15 Å	5 Å	7 Å	10 Å	15 Å
1A2K	1.21	1.14	1.07	1.02	1.06	1.04	1.03	1.02	
1ACB	1.07	1.03	1.00	0.99	1.01	0.99	0.98	0.99	
1AK4	0.85	0.92	0.97	0.99	0.89	0.94	0.97	0.99	
1AVX	1.06	1.01	0.98	0.98	1.01	0.99	0.98	0.98	
1AY7	1.12	1.07	1.03	1.00	1.09	1.05	1.02	1.00	
1B6C	1.01	1.03	1.03	1.01	0.97	1.00	1.01	1.01	
1BUH	1.03	1.01	1.00	1.00	1.02	1.01	1.00	1.00	
1BVK	1.04	1.02	1.01	1.00	1.04	1.02	1.01	1.00	
1CGI	1.02	1.01	1.00	0.99	1.02	1.01	1.00	1.00	
1D6R	0.94	0.98	1.00	1.00	0.91	0.95	0.98	1.00	
1E96	1.02	1.01	1.00	1.00	1.02	1.01	1.00	1.00	
1EAW	0.96	0.98	1.00	1.01	0.99	1.01	1.02	1.02	
1EWY	1.23	1.15	1.08	1.04	1.20	1.14	1.08	1.04	
1F34	1.04	1.02	1.01	1.01	1.01	1.01	1.00	1.00	
1FC2	1.03	1.01	1.00	0.99	1.03	1.01	1.00	0.99	
1FQJ	0.99	0.98	0.98	0.98	1.01	0.99	0.98	0.99	
1GCQ	1.20	1.12	1.05	1.02	1.08	1.05	1.03	1.01	
1GHQ	1.06	1.02	1.00	1.00	1.04	1.02	1.00	1.00	
1GRN	1.02	1.00	0.99	0.99	1.02	1.00	0.99	0.99	
1HE1	1.01	1.00	1.00	1.00	1.01	1.01	1.00	1.00	
1HIA	1.11	1.09	1.05	1.02	1.02	1.03	1.03	1.02	
1IJK	1.12	1.07	1.03	1.01	1.12	1.07	1.03	1.01	
1KAC	1.03	1.01	1.00	1.00	1.04	1.02	1.00	1.00	
1KTZ	1.02	1.04	1.04	1.03	1.00	1.01	1.02	1.02	
1M10	1.01	1.00	1.00	1.00	1.01	1.00	1.00	1.00	
1ML0	0.96	0.98	1.00	1.00	0.93	0.96	0.99	1.00	
1PPE	1.19	1.09	1.01	0.99	1.14	1.06	1.01	0.99	
1QA9	1.14	1.09	1.04	1.02	1.09	1.06	1.04	1.02	
1UDI	1.42	1.22	1.07	1.00	1.41	1.23	1.09	1.02	
1VFB	1.14	1.07	1.02	1.00	1.11	1.06	1.02	1.00	
2PCC	1.34	1.21	1.11	1.05	1.34	1.21	1.12	1.05	
2SIC	0.99	1.00	1.00	1.00	0.94	0.97	0.99	1.00	
2SNI	1.08	1.05	1.02	1.00	1.02	1.01	1.00	1.00	
7CEI	1.53	1.42	1.28	1.13	1.36	1.30	1.22	1.12	
平均	1.09	1.05	1.03	1.01	1.06	1.04	1.02	1.01	
ランダム回転より有利になる割合	82%	82%	71%	53%	79%	76%	68%	53%	

赤字は 1 より大、
青字は 1 以下を表す。

表 3.5 塩無しの場合における真の配向の有利性。表 3.2 と同様に真の配向の有利性 A_{true} を計算した。塩が無いと平均値としての真の配向が有利性は上昇するが、有利なペアの割合が約 55%に低下し、また蛋白質間距離に依存した変化が無くなってしまふ。

塩濃度	0.0 M							
	上位1個				上位10%			
配向								
距離	5 Å	7 Å	10 Å	15 Å	5 Å	7 Å	10 Å	15 Å
1A2K	3.93	3.54	3.16	2.80	3.01	2.90	2.76	2.55
1ACB	0.87	0.84	0.83	0.84	0.83	0.82	0.82	0.85
1AK4	0.29	0.34	0.40	0.46	0.35	0.39	0.44	0.50
1AVX	0.81	0.79	0.78	0.82	0.84	0.83	0.84	0.88
1AY7	0.55	0.57	0.60	0.66	0.60	0.60	0.63	0.68
1B6C	1.01	1.02	1.02	1.01	0.97	1.00	1.02	1.02
1BUH	0.87	0.87	0.89	0.93	0.91	0.91	0.93	0.96
1BVK	1.70	1.64	1.56	1.47	1.57	1.52	1.46	1.39
1CGI	0.84	0.83	0.84	0.86	0.85	0.85	0.85	0.87
1D6R	0.93	0.97	0.99	1.00	0.90	0.94	0.98	1.00
1E96	1.36	1.32	1.28	1.24	1.32	1.29	1.25	1.21
1EAW	2.50	2.44	2.32	2.10	2.34	2.27	2.16	1.95
1EWY	2.00	2.29	2.77	3.48	3.17	3.28	3.48	3.72
1F34	1.25	1.25	1.26	1.26	1.37	1.35	1.32	1.29
1FC2	0.66	0.67	0.69	0.72	0.69	0.70	0.72	0.75
1FQJ	0.48	0.50	0.53	0.59	0.61	0.61	0.63	0.67
1GCQ	1.30	1.22	1.14	1.09	1.15	1.12	1.10	1.07
1GHQ	0.26	0.29	0.33	0.40	0.32	0.34	0.38	0.45
1GRN	1.04	1.02	1.00	0.99	1.01	0.99	0.98	0.97
1HE1	1.17	1.15	1.14	1.11	1.16	1.14	1.12	1.10
1HIA	2.72	2.57	2.40	2.20	2.83	2.71	2.54	2.30
1IJK	0.97	1.05	1.23	1.58	1.17	1.22	1.35	1.62
1KAC	1.08	1.06	1.04	1.03	1.08	1.05	1.04	1.03
1KTZ	2.22	2.13	1.98	1.77	1.84	1.80	1.72	1.58
1M10	3.25	3.08	2.88	2.60	2.89	2.75	2.57	2.34
1ML0	0.00	0.01	0.02	0.04	0.01	0.01	0.02	0.04
1PPE	0.85	0.79	0.76	0.78	0.85	0.81	0.79	0.81
1QA9	1.99	1.83	1.68	1.53	1.78	1.69	1.59	1.46
1UDI	2.19	1.97	1.83	1.80	2.54	2.27	2.06	1.94
1VFB	0.90	0.85	0.82	0.82	0.85	0.82	0.80	0.81
2PCC	14.74	13.34	12.02	10.49	11.54	10.42	9.33	8.11
2SIC	0.55	0.58	0.61	0.66	0.57	0.61	0.65	0.70
2SNI	1.10	1.06	1.03	1.01	1.03	1.02	1.01	1.01
7CEI	19.61	17.56	14.62	10.94	13.16	12.15	10.60	8.37
平均	2.24	2.10	1.95	1.80	1.94	1.86	1.76	1.65
ランダム回転より有利になる割合	56%	59%	56%	56%	56%	56%	56%	56%

赤字は 1 より大、
青字は 1 未満を表す。

表 3.6 真の配向の有利性 A_{true} と静的性質の相関係数。表 3.3 の静的性質の値と表 3.4 の真の配向の有利性の相関係数をとった。セットは全てのペア (ALL 列) 外れ値を除外したペア (ALL-3 列) 酵素-阻害剤のみのペア (E 列) その他に分類されるペア (O 列) とした。相関係数の検定により有意だと判断されたものを色分けしている。

セット	ALL	ALL-3	E	O	
データ数	34	31	15	17	
塩濃度	0.2 M				
相互作用	V_{c-c}	-0.32	0.12	-0.54	0.00
	V_{d-d}	-0.29	-0.34	-0.55	-0.38
	$V_{d-d(scaled)}$	-0.44	-0.60	-0.35	-0.66
形状	SASA 1	-0.14	-0.16	0.00	-0.27
	SASA 2	-0.18	-0.14	-0.02	-0.04
	Mass 1	-0.16	-0.17	-0.16	-0.27
	Mass 2	-0.11	-0.05	0.08	0.05
	M_{rate}	0.01	0.08	0.20	0.24
	L_{rate} 1	0.10	0.00	-0.23	-0.06
アミノ酸	L_{rate} 2	0.37	0.40	0.40	0.42
	R_{polar} 1	0.11	0.10	0.31	0.00
	R_{polar} 2	0.10	0.08	-0.03	0.10
	R_{acid} 1	0.18	0.11	0.50	0.20
	R_{acid} 2	0.60	0.45	0.72	0.23
	R_{base} 1	0.39	-0.02	0.82	-0.06
	R_{base} 2	-0.06	-0.08	-0.05	0.00

5%の有意水準

10%の有意水準

それ以外

追記

・ 真の配向の有利性 A_{true} の定義においてスピン回転を足さない場合

3.4 項において真の配向の有利性 A_{true} を計算する際に、「上位 1 個」を配位回転が相手の中心の向いている状態とし、スピン回転に関しては確率を合計するという定義を行った。これは「上位 10%」を定義する際に、配位回転の近さとスピン回転の近さを総合して判断するのが困難であったために、共通してスピン回転は同一視する、としたためである。

改めて「上位 1 個」についてスピン回転を合計せず、全 62208 個中 1 個の真の配向のみを用いた場合の A_{true} を計算した（表 A）。表 3.4 および表 3.5 の「上位 1 個」の結果と比較すると、34 個中 26 個で A_{true} が上昇し、ランダム分布より有利になったペアの割合も 82% から 91% に向上した。

改めて得られた A_{true} を用いて、相関係数も再度計算した（表 B）。表 3.6 と比べて数値に変化はあるものの、傾向としては大きな変化はない。特にスケールされた双極子相互作用 $V_{d-d(scaled)}$ はほとんど相関係数に変化していないが、これは双極子モーメント同士の向きがスピン回転方向の傾向を持っていないことを意味する。このことは双極子モーメント同士の安定な向きには図 A のように二通りが考えられるが、相関に関係するのは主にスピン回転の関係しない図 A-(a)であることを示唆している。

ペアのうち小さい方の主軸長の比 L_{rate} は相関係数が全体として若干低下している。蛋白質の形状が球状の場合は細長い場合に比べて等方性が強いいため、スピン回転を足した場合（表 3.4、表 3.5）に相関が出やすい傾向があるのかもしれない。しかしながらスピン回転を足さない場合（表 A）においても相関は残っているため、やはり蛋白質の形状は長距離で真の配向をとる確率と関連があると考えられる。

表 A 上位 1 個の定義を変更して再計算した真の配向の有利性 A_{true} 。表 3.4 および表 3.5 の「上位 1 個」の結果と比較すると、34 個中 26 個で A_{true} が上昇し、ランダム分布より有利になったペアの割合も 82%から 91%に向上した。

塩濃度 距離	0.0M				0.2M			
	5 Å	7 Å	10 Å	15 Å	5 Å	7 Å	10 Å	15 Å
1A2K	3.94	3.53	3.13	2.77	1.22	1.14	1.07	1.02
1ACB	1.12	1.04	0.98	0.95	1.14	1.07	1.02	1.00
1AK4	0.27	0.31	0.37	0.44	0.86	0.92	0.97	0.99
1AVX	0.78	0.76	0.76	0.80	1.04	1.00	0.97	0.98
1AY7	0.59	0.60	0.62	0.68	1.17	1.09	1.04	1.01
1B6C	0.98	0.98	0.98	0.98	1.04	1.03	1.02	1.01
1BUH	0.82	0.82	0.84	0.87	1.04	1.02	1.01	1.00
1BVK	1.75	1.68	1.60	1.50	1.04	1.02	1.01	1.00
1CGI	0.84	0.84	0.84	0.86	1.03	1.01	1.00	1.00
1D6R	0.92	0.94	0.95	0.96	0.99	1.00	1.00	1.00
1E96	1.37	1.33	1.29	1.25	1.02	1.01	1.00	1.00
1EAW	2.66	2.54	2.37	2.12	1.01	1.01	1.02	1.01
1EWY	2.54	2.83	3.32	3.99	1.25	1.16	1.09	1.05
1F34	1.10	1.12	1.14	1.16	1.03	1.02	1.01	1.01
1FC2	0.66	0.67	0.69	0.73	1.03	1.01	1.00	0.99
1FQJ	0.55	0.56	0.59	0.63	1.00	0.99	0.98	0.98
1GCQ	1.15	1.09	1.04	1.02	1.19	1.12	1.05	1.01
1GHQ	0.23	0.25	0.29	0.36	1.06	1.02	1.00	1.00
1GRN	1.07	1.04	1.02	1.00	1.02	1.00	1.00	0.99
1HE1	1.25	1.23	1.20	1.16	1.01	1.01	1.00	1.00
1HIA	3.29	3.00	2.69	2.39	1.19	1.13	1.07	1.03
1IJK	0.45	0.53	0.67	0.95	1.11	1.07	1.03	1.01
1KAC	1.10	1.07	1.05	1.04	1.03	1.01	1.00	1.00
1KTZ	2.26	2.15	2.00	1.77	1.03	1.04	1.04	1.03
1M10	3.51	3.30	3.05	2.71	1.01	1.00	1.00	1.00
1ML0	0.00	0.01	0.02	0.04	0.98	0.99	1.00	1.00
1PPE	0.76	0.72	0.71	0.74	1.17	1.07	1.01	0.99
1QA9	1.95	1.79	1.64	1.50	1.15	1.08	1.04	1.01
1UDI	2.38	2.12	1.94	1.88	1.43	1.22	1.07	1.01
1VFB	0.86	0.81	0.79	0.80	1.14	1.07	1.02	1.00
2PCC	15.88	14.21	12.64	10.86	1.37	1.23	1.12	1.06
2SIC	0.56	0.59	0.62	0.66	1.00	1.01	1.01	1.00
2SNI	1.12	1.08	1.04	1.02	1.09	1.05	1.02	1.00
7CEI	22.19	19.31	15.55	11.32	1.61	1.47	1.29	1.14
平均	2.38	2.20	2.01	1.82	1.10	1.06	1.03	1.01
ランダム回転より 有利になる割合	56%	56%	53%	53%	91%	88%	76%	59%

表 B 改めて定義した A_{true} と静的性質の相関係数。表 3.6 と比べて若干の値の変動があるものの、傾向としては大きな変化は無い。

セット	ALL	ALL-3	E	O	
データ数	34	31	15	17	
塩濃度	0.2 M				
相互作用	V_{c-c}	-0.35	0.08	-0.58	-0.04
	V_{d-d}	-0.26	-0.30	-0.55	-0.34
	$V_{d-d(scaled)}$	-0.44	-0.61	-0.39	-0.64
形状	SASA 1	-0.14	-0.15	-0.05	-0.22
	SASA 2	-0.23	-0.25	-0.08	-0.05
	Mass 1	-0.16	-0.15	-0.23	-0.22
	Mass 2	-0.17	-0.15	0.02	0.03
	M_{rate}	-0.03	-0.02	0.20	0.19
	L_{rate} 1	0.15	0.09	-0.17	-0.05
	L_{rate} 2	0.33	0.34	0.34	0.42
アミノ酸	R_{polar} 1	0.11	0.07	0.34	0.01
	R_{polar} 2	0.14	0.16	0.01	0.12
	R_{acid} 1	0.18	0.11	0.55	0.20
	R_{acid} 2	0.58	0.39	0.66	0.22
	R_{base} 1	0.39	-0.06	0.85	-0.06
	R_{base} 2	-0.04	-0.04	-0.01	0.04

5%の有意水準
10%の有意水準
それ以外

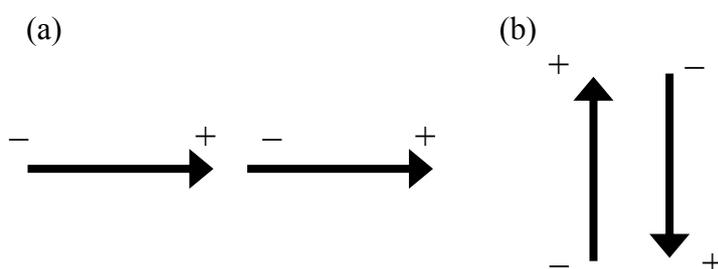


図 A 双極子モーメント同士がエネルギー的に安定となる向き。片方の双極子モーメントにスピン回転（それぞれの中心同士を結んだ軸に関する回転）を加えたとき、(a)はスピン回転の影響は無いが、(b)はスピン回転によって変化する。