博士論文

粗視化分子動力学シミュレーションを用いたタンパク質-リガンド結合過程の網羅的解析

指導教官　清水 謙多郎 教授

応用生命工学専攻

生物情報工学研究室

根上　樹

目次

第 1 章　序論 **3**

第 2 章　MARTINI 粗視化力場 **5**

2.1　MARTINI における粗視化モデル 5

2.2　相互作用ポテンシャル 6

2.3　時間スケール 7

2.4　タンパク質への適用 7

第 3 章　MARTINI 粗視化力場を用いたタンパク質-リガンド結合過程の比較シミュレーション **10**

3.1　タンパク質-リガンド複合体の分類および計算対象 10

3.1.1　タンパク質-リガンドペアの分類手法 10

3.1.2　分類結果 12

3.1.3　計算対象 13

3.2　計算手法 15

3.2.1　シミュレーション条件 15

3.2.2　粗視化モデルのマッピングおよびパラメータ 17

3.2.2.1　タンパク質 17

3.2.2.2　リガンド 17

3.3　結果 20

3.3.1　リガンド結合状態の再現 20

3.3.2　リガンドの密度分布および結合親和性 24

3.3.3　リガンドのパラメータの修正 33

3.3.3.1　LinB 33

3.3.3.2　PAL 35

3.3.4　結合経路の解析 37

3.3.4.1　分子流束密度 37

3.3.4.2　分子流束密度 37

3.3.4.3　タンパク質表面の性質と結合経路の関係の数値的な評価　42

3.3.5 水とリガンドの交換 49
3.4 まとめ 51

第 4 章 Dual-basin potential による MARTINI 粗視化シミュレーションの拡張 53
4.1 Dual-basin potential 53
4.2 タンパク質の構造変化シミュレーション 55
4.2.1 計算対象 55
4.2.2 計算手法 56
4.2.3 結果 57
4.3 タンパク質-リガンド結合シミュレーションへの適用 58
4.3.1 計算手法 58
4.3.2 結果 59
4.4 まとめ 61

第 5 章 総括 62

付録 A 化合物の MARTINI 粗視化力場パラメータの決定 64
A.1 力場決定の手順 64
A.2 結合パラメータの比較 65
A.3 水-オクタノール分配係数の比較 69

付録 B 結合、解離速度定数の推定 71
B.1 ベイズ推定による速度定数の推定 71
B.2 結合、解離の定義による速度定数の依存性 73

参考文献 79

第1章　序論

生体内で起こる代謝、免疫、運動などのさまざまな現象は、タンパク質による反応がその多くを担っており、生命現象を解明する上ではタンパク質の構造、機能の理解は不可欠なものとなる。近年、分子生物学の分野におけるさまざまな実験技術の発達によりタンパク質構造の解明は加速度的に進んでいる。こうした知見の蓄積が進むことにより、タンパク質構造をもとに薬剤設計を行う sructure based drug design[1, 2] などの情報科学的手法は応用の幅が広まり、さらに重要性を増すことが予想される。このような創薬へ応用を行う上で、薬剤分子 (リガンド) とタンパク質の相互作用の理解は不可欠であり、これまでに様々な研究が行われてきた。こうした研究の多くはエネルギー的に安定な結合状態を理解することに主眼が置かれており、リガンドの結合過程に着目した研究は少ない。しかしながら、リガンドの受容体への結合速度が薬効に大きく寄与することが明らかになってきており [3, 4, 5]、結合メカニズムの解明も重要な研究課題のひとつとなる。

タンパク質-リガンド結合過程を原子レベルで詳細に解明することは実験的手法ではきわめて困難であり、分子動力学シミュレーションを用いた計算科学的手法によりリガンドの結合過程を直接見ることが結合メカニズムを解明する上で非常に有効な方法となると考えられる。しかしながら、タンパク質とリガンドの結合は巨大な系に対する長時間スケールの解析が必要となるため、通常の分子動力学シミュレーションの適用は計算コストが高いことが問題となる。これまでタンパク質-リガンド結合に関する分子動力学シミュレーションを用いた研究としては、実験では見出せなかった thioflavin-T と β-リッチペプチドの結合をシミュレーションにより明らかにした例 [6] や、nicotinic acetylcholine receptor への isoflurane の結合に関して複数の結合状態を見出した研究 [7] などが報告されているが、過去のこうした研究では結合状態の予測にとどまっていた。最近になって、複数のグループにより、trypsin への benzamide の結合過程 [8] や β-adrenergic receptor に対する agonist や antagonist の結合過程 [9] を長時間のシミュレーションにより再現した研究が報告されている。しかしながら、これらの研究においても、結合過程の追跡は μs のオーダーであるため、リガンドの結合過程の時間スケールがミリ秒に及ぶ場合もあることを考慮すると、依然として計算コストの問題が存在するといえる。また、リガンドの結合過程はリガンド結合ポケットやリガンドの形状、物理化学的性質によって異なると考えられ、結合過程を一般的に理解するためには様々な性質を持つ多数のタンパク質-リガンドのペアにつ

いて計算を行う必要があり、計算の効率化が望まれる。

分子動力学シミュレーションにおいて計算時間を削減するための方法として、粗視化と呼ばれるものがある。粗視化とは、いくつかの原子をひとまとめにすることで系を単純化し、巨大な系に対する長時間のダイナミクスを解析する手法である。生体分子に関しては、脂質二重膜の系に対する研究が数多くなされているほか [10, 11]、タンパク質においてもフォールディングなどのダイナミクスに関する研究が行われており [12, 13]、粗視化手法は様々な対象に対して適用されている。タンパク質-リガンド結合に関しては、HIV-1 Protease に対して結合過程を解析した研究などがある [14] が、このような例で用いられている力場は特定の対象に対して最適化されたものであり、一般的なタンパク質-リガンド結合への適用はまだなされていない。

本研究では、タンパク質-リガンド結合における計算コストの問題を克服し幅広い対象について調べるため、熱力学パラメータに基づく半定量的な力場である MARTINI 粗視化力場を用いたシミュレーションによりタンパク質-リガンド結合過程を解析した。様々な性質をもつ多数のタンパク質-リガンドのペアに対して粗視化シミュレーションによる解析を行い、結合経路とそれぞれの性質の間の関係性を調べることで結合過程について一般的な知見を得ることを試みた。また一方で、MARTINI 粗視化力場を用いたシミュレーションにおいてはタンパク質の構造を保持するために束縛をかける必要があることから、大きな構造変化を扱えないという制約が存在し、この問題を克服するためのシミュレーション手法の拡張を行った。

本論文では第 2 章において MARTINI の概要に触れた後、第 3 章においてタンパク質-リガンド結合過程の比較シミュレーションについて記述する。タンパク質-リガンドの結合過程について一般的な知見を得るため、タンパク質-リガンドペアを物理化学的性質に基づいて分類し、異なる性質を持つ多数のタンパク質-リガンドペアについてリガンド結合シミュレーションを実行し、結合過程の傾向に関する解析を行った。第 3 章ではタンパク質の大きな構造変化を扱うための MARTINI 粗視化シミュレーションの拡張について記述する。タンパク質の構造変化を可能とするため、束縛ポテンシャルに対して dual-basin potential と呼ばれるポテンシャルを適用し、タンパク質の構造変化を再現することを可能とした。さらに、この手法をタンパク質-リガンドの系に適用し、構造変化を伴うタンパク質-リガンド結合過程を解析した。

第2章　MARTINI粗視化力場

粗視化とは、巨大な系においていくつかの原子をまとめて一つの粒子とみなすことにより系を単純化し計算時間を削減する手法であり、脂質膜をはじめとしてタンパク質のフォールディングなど、生体分子のさまざまな系に関して粗視化シミュレーションを適用した研究が行われている [10, 13]。本研究では、Marrink らによって開発された MARTINI 粗視化力場 [15, 16, 17] を用いた。従来用いられてきた粗視化モデルには様々なものが存在するが、その多くは特定の対象とする系に対し、運動や相互作用を適切に近似するようにモデル化しパラメータを決定するものである。これに対し、MARTINI 力場では分配自由エネルギーなどの熱力学的データを基準とした共通のパラメータを設定することで一般的な適用が可能となっている。MARTINI 力場では脂質やアミノ酸といった生体分子についてある程度力場が整備されており、様々なリガンドとタンパク質の相互作用を調べるという本研究の目的に対し、この力場は適していると考えられる。以下では MARTINI 力場の概要について触れる。

2.1　MARTINIにおける粗視化モデル

MARTINI による粗視化シミュレーションでは、原則として 4 つの重原子を一つの粒子 (粗視化原子) とみなして計算を行う。水分子に関しては 4 分子を一つの粗視化粒子として扱い、分子として実際に系に含めて計算を行う。ただし、ベンゼンなどの小さい環構造は 2 原子を一つの粗視化原子とすることにより、3 員環のような環状構造とする。各粗視化原子はそれぞれの部分構造の極性によって大まかに charged (Q)、polar (P)、nonpolar (N)、apolar (C) の 4 種類に分類され、さらにその中で極性の強さによる分類が行われ、1 (非極性) から 5 (極性) の 5 種類、あるいは水素結合形成能に基づいて d (donor)、a (accepter)、da (donor/accepter)、0 (none) の 4 種類に分類される。一例として、図 2.1 にペプチド (Val-Phe-Arg) および水の MARTINI 粗視化モデルを示した。

図 2.1: MARTINI 粗視化モデルの一例：ペプチド (Val-Phe-Arg) および水の粗視化マッピング [15, 17]。

2.2 相互作用ポテンシャル

非結合粗視化原子間の相互作用は Lennard-Jones ポテンシャルおよび静電相互作用によって決まる。粗視化原子 i、j 間の LJ ポテンシャルは 12-6 ポテンシャルにより以下のように定義される。

$$V_{\mathrm{LJ}}(r) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] \tag{2.2.1}$$

σ_{ij} は粗視化原子間の距離に関するパラメータであり、通常 σ_{ij} = 0.47 nm として計算される。ε_{ij} は相互作用の強さに関するパラメータであり、相互作用する粗視化原子種の組み合わせに対応して 2.0 kJ mol^{-1} から 5.6 kJ mol^{-1} の異なる値をとる。ただし、環状構造同士の相互作用においては、細かいマッピングを行っているため、小さい原子間距離 (σ_{ij} = 0.43 nm) およびスケールされた相互作用パラメータ (75 %) が用いられる。

荷電している原子 (Q) に関しては静電相互作用がクーロンポテンシャルにより定義される。

$$V_{\mathrm{el}}(r) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_{\mathrm{r}} r} \tag{2.2.2}$$

q_i、q_j は粗視化原子 i、j それぞれの電荷であり、ε_0 は真空の誘電率を表す。比誘電率は ε_{r} = 15 として計算される。LJ ポテンシャルおよび静電相互作用のカットオフ半径は 1.2 nm であり、LJ ポテンシャルでは 0.9-1.2 nm、静電相互作用では 0-1.2 nm でシフトするポテンシャルが用いられる。

結合原子間の相互作用は結合長、結合角、二面角に関する以下のような調和ポテンシャ

ルにより計算される。

$$V_{\mathrm{bond}}(R) = \frac{1}{2}K_{\mathrm{bond}}(R-R_0)^2 \tag{2.2.3}$$

$$V_{\mathrm{angle}}(\theta) = \frac{1}{2}K_{\mathrm{angle}}\left\{\cos(\theta)-\cos(\theta_0)\right\}^2 \tag{2.2.4}$$

$$V_{\mathrm{dih}}(\psi) = K_{\mathrm{dih}}(\psi-\psi_0)^2 \tag{2.2.5}$$

ここで、$R-R_0$、$\cos(\theta)-\cos(\theta_0)$、$\psi-\psi_0$ は結合長、結合角、二面角の平衡状態からのずれを表し、K_{bond}、K_{angle}、K_{dih} は力の定数を表す。これらの値は、全原子シミュレーションにおける構造、運動との比較により決定される。

2.3 時間スケール

分子の運動には、並進、回転、振動の 3 つの自由度が存在し、回転、振動は並進に比べ非常に速い時間スケールでの運動となる。粗視化力場では、複数の原子を一つの粗視化原子とみなしており、分子内の振動や回転などの自由度は小さくなる。したがって、粗視化シミュレーションでは運動の時間スケールが遅くなると考えられ、これによって大きな時間刻みを用いることができる。MARTINI 力場では通常 20 fs から 40 fs 程度の時間刻みが用いられ、これは全原子シミュレーションのおよそ 10 倍程度の値となる。一方、粗視化により分子内運動の自由度が失われることから、その自由エネルギー地形は滑らかとなる。これにより、粗視化シミュレーションにおける分子の運動は実際よりも速くなるという性質を持つ。全原子シミュレーションとの拡散係数の比較から、粗視化におけるダイナミクスはおよそ 2-10 倍速くなることが知られている。したがって、粗視化におけるシミュレーション時間は実際の時間の数倍に対応することになり、サンプリング効率が上昇することが期待される。また、結合速度定数などの運動速度に関する値は実験値に比べ大きく見積もられることになる。標準的には、水の拡散速度に基づき時間を 4 倍にスケーリングすることで実際の系と対応させる。以下、本論文ではシミュレーション時間の記述は粗視化における時間をそのまま表すものとする。

2.4 タンパク質への適用

ペプチドの MARTINI 力場は Montecelli らにより整備されている [17]。アミノ酸分子は主鎖に対応する一つの骨格原子と側鎖に対応するいくつかの原子によりモデル化されており、各アミノ酸側鎖に対してパラメータが決められている。ペプチド分子はアミノ酸分子が連なったものとして扱われ (図 2.1)、全原子構造から機械的に構築することができる。しかしながら、タンパク質のような複雑な構造を持つものに対してはシミュレーション

において構造が不安定となり、通常の MARTINI 力場のみでは計算ができない。したがって、一定の距離に存在する原子同士を調和ポテンシャルで繋げる elastic network モデルを用いることにより構造を束縛することが必要となる。

水中の HIV-1 Protease に対してシミュレーションを行った例を図 2.2 に示す。このタンパク質は対照的な二量体を形成しており、中央に結合ポケットを有する。このタンパク質に対し、通常の MARTINI 力場のみにより 200 ns のシミュレーションを行ったものが図 2.2 の下段左の図であり、中央のポケットが崩れ、タンパク質構造が壊れていることがわかる。これに対し、elastic network model により主鎖原子に束縛を行ったものが図 2.2 の下段右の図であり、初期構造が保存されていることがわかる。

図 2.2: 水中の HIV-1 Protease に対する MARTINI 粗視化シミュレーションにおけるタンパク質の粗視化構造。ベージュおよび黄色の部分はそれぞれタンパク質の主鎖、側鎖に対応する。上段は初期構造を表し、下段は 200 ns のシミュレーション後における構造を表す。左は通常の MARTINI 力場のみを用いた場合、右はそれに加えて elastic network model により構造の束縛をかけたときの構造を表す。

このように現段階では、MARTINI 力場によるシミュレーションにおいてはタンパク質構造の束縛を行う必要があり、そのためタンパク質の大きな構造変化を伴う現象を取り扱うことができないという制約が存在する。

第3章　MARTINI粗視化力場を用いたタンパク質-リガンド結合過程の比較シミュレーション

タンパク質-リガンド結合過程に関する一般的な知見を得るために、様々な種類のタンパク質-リガンドのペアについて粗視化シミュレーションを行い、結合過程の傾向の解析を行った。本研究では、結合ポケットおよびリガンドについて異なる性質を持つ様々なタンパク質-リガンドペアを選んでシミュレーションを行い、それらの性質と結合過程の関係について見出すというアプローチをとることとした。はじめにタンパク質-リガンドペアをリガンド結合ポケットおよびリガンドの性質に基づいて分類し、その後各分類について具体例を選択してリガンド結合シミュレーションを行い、結合過程を解析した。

本章ではまずはじめに3.1節においてタンパク質-リガンドペアの分類について触れた後、3.2節でリガンド結合シミュレーションの方法論について述べる。その後、3.3節において結合経路や速度定数などに関する解析結果、考察を述べる。

3.1　タンパク質-リガンド複合体の分類および計算対象

3.1.1　タンパク質-リガンドペアの分類手法

様々な種類のタンパク質-リガンドの系について調べるため、はじめにタンパク質-リガンド複合体をリガンド結合ポケットおよびリガンドの性質に基づいて分類した。タンパク質-リガンドペアのデータセットは、Binding MOAD [18, 19] を元に作成した。Binding MOADは溶媒やイオン等を除く生物学的に意味のあるリガンドを含む構造だけを集めたタンパク質-リガンド複合体結晶構造データベースであり、23269個の構造が登録されている。このデータベースから、リガンドの結合していない状態のタンパク質構造が得られており、かつ分子量550以下のリガンドを持つものを抽出してデータセットとした。

次に、データセット中のタンパク質およびリガンドの性質について分類を行った。タンパク質におけるリガンド結合ポケットの形状および疎水性、リガンドの疎水性、親水性リ

ガンドにおける電荷相互作用の有無について、以下に示したような判断基準に基づき、それぞれの特徴が強いタンパク質を選択した。

リガンド結合ポケットの形状： 最初に、CAVER 3.0[20, 21] を用いて結合ポケットの検出を行った。CAVER 3.0 はタンパク質構造とリガンドの座標位置を入力として、リガンドの存在する位置からタンパク質の外部に到達するまでの経路を検出し、その経路を連続した座標として出力する。結合経路上の各座標点については、検出開始位置 (リガンドの位置) からの経路上の距離 l に対し、タンパク質原子との最小距離 $r(l)$ が与えられており、結合のしやすさを表現するコスト関数が以下のように定義される。

$$Cost = \int_0^L r(l)^{-2} dl \tag{3.1.1}$$

ここで、L は経路の全長である。この式は $r(l)^{-2}$ の全経路における積分であり、半径が小さく経路長が長いほど大きな値をとり、リガンドが奥まった位置にあることを示す。本研究では、$Cost > 0.2$ のときを狭いポケット、$Cost < 0.12$ かつ $r(l)$ の平均が 3.5 Å以上のときを広いポケットとして、タンパク質-リガンドペアを選択した。

図 3.1: ポケット検出結果の一例 (Haloalkane dehalogenase LinB)。左はポケットの検出結果を青い球の連続したものとして表した図であり、タンパク質構造を茶色のリボンモデル、リガンドをスティックモデルで表している。右図は検出開始位置 (リガンドの位置) からの経路上の距離 l に対してそれぞれの球の半径 $r(l)$ をプロットしたものを表す。

リガンド結合ポケットの親水性： 複合体構造において、リガンドの周囲 5 Å以内に存在するアミノ酸残基について Kyte-Doolittle の疎水性指標 [22] の平均を計算し、-0.8 以下のものを親水性、0 以上のものを疎水性ポケットと判定した。

リガンドの疎水性： リガンドの水-オクタノール分配係数 (LogP) を用いた分類を行った。EPI Suite 4.11[23] を用いて PHYSPROP データベースにおける LogP の実験値を取得

し、実験値のないものについては ALogPS 3.0[24, 25] を用いて予測した。LogP の値が-1 以下のものを親水性、1 以上のものを疎水性と判定した。

静電相互作用： 最初に Openbabel[26] を用いて pH = 7 におけるリガンドの荷電状態を予測した。ただし共鳴構造を持つ酸、塩基に関しては、いずれかの共鳴構造において形式電荷を持つ原子は、全て形式電荷を持つものと見なした (図 3.2)。複合体構造において、電荷を持つリガンド原子から 3.5 Å 以内に存在する電荷を持つタンパク質原子 (アスパラギン酸、グルタミン酸のカルボン酸およびアルギニン、リジン、ヒスチジンのアミン) について、反対の符号の電荷を持つ原子数＞同じ符号の電荷を持つ原子数である場合に静電相互作用していると見なした。

図 3.2: 共鳴構造を持つ酸 (上段) および塩基 (下段)。形式電荷を持つものと判定した原子を色付き (赤：負電荷、青：正電荷) で示している。

3.1.2 分類結果

データセットの分類結果は表 3.1 のようになった。分類結果から、リガンド結合ポケットの疎水性、親水性のリガンドは親水性の基質結合ポケットに結合し、疎水性のリガンドは疎水性の基質結合ポケットに結合する傾向を見出すことができた。タンパク質-リガンドペアの数が多く一般的なものと考えられるこれらの分類を表に示したように分類 I～VI とし、これらの分類に属するものに対して粗視化シミュレーションによる解析を行うこととした。

表 3.1: タンパク質-リガンド複合体の分類結果。表の数は、90 パーセント以上の相同性を持つものに関して BLASTClust[27] によるクラスタリングを行い、冗長性を排除した後の値を示している。

		リガンド		
		親水性・	親水性・	疎水性
リガンド結合ポケット		静電相互作用あり	静電相互作用なし	
親水性	狭い	179 …I	46 …III	11
	広い	260 …II	132 …IV	37
疎水性	狭い	26	12	142 …V
	広い	39	21	105 …VI

3.1.3 計算対象

3.1 における分類 I〜VI から 2 組ずつタンパク質-リガンドペアを選び、それぞれについてタンパク質-リガンド結合シミュレーションを行った。表 3.2 に選択したタンパク質を示す。

表 3.2: 計算対象に選択したタンパク質

Class	Protein	PDB ID	
		ligand-free	complex (ligand)
I	3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase from *Saccharomyces cerevisiae* (DAHPS)	1OFB[28]	1HFB (phosphoenol pyruvate)[28]
	Aspartate aminotransferase from *Escherichia coli* (AST)	1AAW[29]	1ASM (maleic acid)[29]
II	6-phosphogluconolactonase from *Trypanosoma brucei* (6PGL)	2J0E [30]	3EB9 (citric acid)[31]
	Histidine acid phosphatase from *Francisella tularensis* subsp. holarctica LVS (HAP)	3IT2 [32]	3IT1 (tartaric acid)[32]
III	Purine nucleoside phosphorylase from *Thermus thermophilus* HB8 (PNP)	1ODL[33]	1ODJ (guanosine)[33]
	Nucleoside 2-deoxyribosyltransferase from *Lactobacillus leichmannii* (NdRT)	1F8X [34]	1F8Y (5-methyl-2'-deoxypseudouridine)[34]
IV	Levansucrase from *Bacillus subtilis* (Lev)	1OYG [35]	1PT2 (sucrose)[35]
	Pterocarpus angolensis Lectin from *Pterocarpus angolensis* (PAL)	1S1A [36]	2ARE (mannose)[37]
V	Haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* (LinB)	1IZ7 [38]	1G5F (dichroloethane)[38]
	Penicillin G acylase from *Escherichia coli* (PGA)	1PNK [39]	1AI7 (phenol)[40]
VI	Ketosteroid isomerase from *Pseudomonas putida* (KSI)	3VSY[41]	2PZV (phenol) [42]
	Catalytic domain of human phosphodiesterase 4D (PDE4D)	3SL3 [43]	1Y2B (Ethyl 3,5-dimethyl-1H-pyrazole-4-carboxylate)[44]

MARTINI においてタンパク質の大きな構造変化を扱うことができないという制約のため、以上の計算対象はいずれもリガンド結合状態と非結合状態において構造変化の小さいもの (主鎖構造の RMSD < 1 Å) を選び、リガンドの運動に着目して解析を進めることとした。

3.2 計算手法

3.2.1 シミュレーション条件

MARTINI 粗視化力場を用いたシミュレーションにより、リガンドが溶液中からタンパク質の結合部位に進入する様子を追跡することを試みた。一般に、粗視化シミュレーションは通常の全原子シミュレーションに比べ精度が劣るため、多数のシミュレーションを行ってそれらの結果を統計的に解析するというアプローチをとることとした。

シミュレーションの系の構成およびシミュレーション時間はそれぞれの計算対象について以下の表 3.3 のような条件で行った。これらの系はタンパク質分子の周りに少数のリガンド分子をランダムに配置し、水溶媒中に溶解することにより構築した (図 3.3)。シミュレーションは、それぞれ異なるリガンドの初期配置に対し数十回から百回のシミュレーションを行った。時間刻みは 20 fs を用い、温度および圧力は 1 bar、310 K で一定とした。シミュレーションにはプログラムパッケージ GROMACS 4.0.7[45] を用いた。

表 3.3: シミュレーション条件

計算対象	リガンド		濃度 (c_{lig}) / mM	シミュレーション時間
DAHPS	Phosphoenolpyruvate	×10	15	5 μs ×50
AST	Maleic acid	×50	45	2 μs ×50
6PGL	Citric acid	×10	17	5 μs ×50
HAP	Tartaric acid	×30	33	2 μs ×50
PNP	Guanosine	×10	9.4	5 μs ×50
NdRT	2 ’-deoxycytidine	×15	23	5 μs ×50
Lev	Sucrose	×5	14	4 μs ×50
PAL	Glucose	×5	18	1 μs ×100
LinB	Dichloroethane	×10	41	1 μs ×100
PGA	Phenol	×30	39	5 μs ×50
KSI	Phenol	×10	10	2 μs ×50
PDE4D	Ethyl 3,5-dimethyl-1H-pyrazole-4-carboxylate	×5	12	5 μs ×50

図 3.3: Levansucrase(Lev)-sucrose のシミュレーションにおける初期配置の例。茶色はタンパク質分子の主鎖構造、緑色はリガンド分子、水色は粗視化水分子 (通常の水分子 4 分子に相当) を表す。

3.2.2 粗視化モデルのマッピングおよびパラメータ

3.2.2.1 タンパク質

タンパク質構造の MARTINI モデルは既存のプロトコルにより X 線結晶構造から生成し、タンパク質形状は ELNEDYN[46] と呼ばれる elastic network モデルを用いて束縛した。束縛は 0.9 nm 以内にある主鎖原子同士に対して力の定数 500 kJ mol^{-1} nm^{-2}、0.43 nm 以内の距離にある側鎖原子同士に対して力の定数 50 kJ mol^{-1} nm^{-2} の調和ポテンシャルを用いた。

3.2.2.2 リガンド

結合シミュレーションに用いたリガンドのマッピングおよび結合パラメータを図 3.4 および表 3.4 に示した。Glucose および sucrose の粗視化モデルは Lpez らによって設計されたものを用いた [47]。また、phenol については tyrosine の側鎖の類縁体と見なし、tyrosine の粗視化構造の側鎖部分を phenol として用いることとした。アミノ酸側鎖のパラメータの決定は、側鎖部分のみの構造に対して、全原子シミュレーションと粗視化シミュレーションにおける物理化学的性質の比較を行うことによりなされている [17] ため、phenol の性質を再現できるものと考えられる。Dichloroethane は 4 つの重原子からなるため、一つの

粗視化原子により表すこととした。粗視化原子の種類は、dichloroethane の水-オクタノール分配自由エネルギーの実験値 $\Delta G_{\mathrm{part}} = -8.5\,\mathrm{kJ\,mol^{-1}}$[48] に基づき、それに最も近い値を持つ C4 粗視化原子 ($\Delta G_{\mathrm{part}} = 9\,\mathrm{kJ\,mol^{-1}}$[16]) を選んだ。それ以外の分子に関しては、付録 A のプロトコルの通りに決定した。

図 3.4: リガンドの MARTINI 粗視化マッピング。水素原子以外の原子に対して 3~5 個程度のグループに分け、粗視化粒子 (B1, B2, ...) を割り当てた。

表 3.4: リガンドの結合パラメータ。K_{bond} が const. であるものについては、距離を一定と見なし、シミュレーション時に LINCS アルゴリズム [49] により結合長を固定した。

リガンド	bonds	R_0 / nm	K_{bond} / kJ mol^{-1}nm^{-2}	angles	θ_0 / deg	K_{angle} / kJ mol^{-1}	dihedrals	φ_0 / deg	K_{dih} / kJ mol^{-1}
Phosphoenol-pyruvate	B1-B2	0.290	const.						
Maleic acid	B1-B2	0.323	const.						
Citric acid	B1-B2	0.300	18000	B1-B2-B3	95	170			
	B2-B3	0.300	18000						
Tartaric acid	B1-B2	0.313	const.						
Guanosine	B1-B2	0.353	30000	B1-B3-B4	150	200	B1-B2-B3-B4	155	60
	B1-B3	0.257	30000	B2-B3-B4	103	300	B2-B3-B4-B5	78	25
	B2-B3	0.241	const.	B3-B4-B5	128	300	B4-B5-B7-B6	180	200
	B3-B4	0.244	const.	B3-B4-B7	120	400			
	B4-B5	0.295	const.	B4-B5-B6	69	200			
	B4-B7	0.161	const.	B5-B4-B7	125	200			
	B5-B6	0.294	const.	B5-B6-B7	84	200			
	B5-B7	0.390	const.	B6-B7-B4	94	200			
	B6-B7	0.290	const.						
2-deoxycytidine	B1-B2	0.317	const.	B1-B3-B4	123	400	B1-B3-B4-B5	-145	30
	B1-B3	0.280	const.	B2-B3-B4	150	400	B2-B1-B3-B4	-157	30
	B2-B3	0.244	const.	B3-B4-B5	104	500	B3-B4-B5-B6	-177	150
	B3-B4	0.219	const.						
	B4-B5	0.220	const.						
	B4-B6	0.268	const.						
	B5-B6	0.285	const.						
Glucose	B1-B2	0.375	35000						
	B1-B3	0.331	35000						
	B2-B3	0.322	50000						
Sucrose	B1-B2	0.222	30000	B1-B2-B4	130	10	B1-B3-B4-B5	130	25
	B2-B3	0.247	30000	B3-B2-B4	110	150	B2-B1-B3-B4	80	2
	B2-B4	0.429	30000	B5-B4-B2	20	50	B3-B4-B5-B6	-70	20
	B4-B5	0.293	30000	B6-B4-B2	85	150			
	B4-B6	0.372	30000						
Phenol	B1-B2	0.270	const.						
	B1-B3	0.270	const.						
	B2-B3	0.270	const.						
Ethyl 3,5-dimethyl-1H-pyrazole-4-carboxylate	B1-B2	0.303	const.	B1-B2-B4	136	500			
	B1-B3	0.355	const.	B1-B3-B4	110	500			
	B2-B3	0.244	const.						
	B2-B4	0.206	const.						
	B3-B4	0.213	const.						

3.3　結果

3.3.1　リガンド結合状態の再現

シミュレーションの結果について、はじめに複合体構造における結合位置への結合を調べた。タンパク質-リガンド複合体構造におけるタンパク質の Cα 部分をシミュレーションにおける主鎖粒子と重ね合わせて構造の比較を行ったところ図 3.5 のようになり、いずれのタンパク質-リガンドペアの系でも結晶構造におけるリガンド結合ポケットへの結合を確認することができた。続いて、シミュレーション中におけるリガンドの運動の様子を調べるため、結晶構造における結合位置からの距離の時間変化を調べた。タンパク質構造と複合体結晶構造を Cα 原子について重ね合わせた後、複合体構造におけるリガンドの重心位置と、シミュレーションにおけるリガンドの重心との間の距離を計算することにより、本来の結合位置からの距離とした。各系において、シミュレーション中におけるリガンドの結合位置からの距離をプロットしたものを図 3.6 に示した。プロットはそれぞれ各系について多数シミュレーションを行ったうちの 1 回の結果を示している。いずれのタンパク質-リガンドペアの系でも、シミュレーションの中で遠くに存在したリガンドが接近し、リガンドが結晶構造におけるリガンドの重心位置から 3 Å以内に接近できることが確認された。また、シミュレーションによっては、リガンドが結合した状態から解離していく過程も見出すことができた。

図 3.5: MARTINI 粗視化力場を用いたリガンド結合シミュレーションにおけるリガンド結合状態のスナップショット (DAHPS-NdRT)。茶色のスティックモデルと青のリボンはそれぞれ粗視化シミュレーションにおけるタンパク質の主鎖構造、リガンド複合体構造における主鎖構造を表し、緑およびマゼンタはそれぞれ粗視化シミュレーションにおけるリガンドと複合体構造におけるリガンドを表す。

図 3.5: MARTINI 粗視化力場を用いたリガンド結合シミュレーションにおけるリガンド結合状態のスナップショット (続き：Lev-PDE4D)。

図 3.6: シミュレーション中におけるリガンドの結合位置からの距離に関する時間プロット (DAHPS-NdRT)。グラフはそれぞれの系について複数回行ったシミュレーション結果のうちの一つについて、3 Å以内に接近したリガンドのみを表示している。異なる色のプロットは異なるリガンドに対応する。

図 3.6: リガンドの結合位置からの距離に関する時間プロット (続き:Lev-PDE4D)。

3.3.2 リガンドの密度分布および結合親和性

リガンドが結合状態を再現できることが確認できたことを受け、続いて基質結合ポケット付近におけるリガンドの安定性を調べるため、シミュレーション中におけるリガンドの 3 次元密度分布および結合親和性を調べた。

リガンドの 3 次元密度分布は、空間を一辺 6 Å のグリッドに分割した後、各グリッドに対してシミュレーション中に存在したリガンドののべ数を数え上げ、スナップショットの

数とグリッドの体積で割ることによって計算した。図 3.7 に 3 次元密度分布の計算結果を示す。いずれの系でも、複合体構造におけるリガンド結合位置の付近に密度の高い領域が確認できた。また、リガンドの結合位置以外にもタンパク質表面にリガンド密度の高い領域が複数確認された。

図 3.7: シミュレーション中におけるリガンドの三次元密度分布 (DAHPS-NdRT)。緑のメッシュはある密度での等位面を表し、マゼンタは複合体結晶構造におけるリガンドを表す。密度の高い領域のうち、複合体構造における結合位置に最も近いサイトを S1 とし、その他をを S2, S3 . . . などとして示している。等位面の密度はそれぞれ、$50c_{\mathrm{lig}}$ (DAHPS)、$50c_{\mathrm{lig}}$ (AST)、$50c_{\mathrm{lig}}$ (HAP)、$150c_{\mathrm{lig}}$ (6PGL)、$100c_{\mathrm{lig}}$ (PNP)、$150c_{\mathrm{lig}}$ (NdRT)、$70c_{\mathrm{lig}}$ (Lev)、$40c_{\mathrm{lig}}$ (PAL)、$8c_{\mathrm{lig}}$ (LinB)、$20c_{\mathrm{lig}}$ (PGA)、$100c_{\mathrm{lig}}$ (KSI)、$150c_{\mathrm{lig}}$ (PDE4D) とした。

図 3.8: シミュレーション中におけるリガンドの三次元密度分布 (Lev-NdRT)。

次に、リガンドの三次元密度分布において観測された各ピークに関して、結合および解離の速度定数を計算した。結合速度の推定はベイズ推定によるアプローチに基づいて行い、以下のように結合、解離の速度定数 k_{on}、k_{off} を計算した (付録 B.1)。

$$k_{\mathrm{on}} = \frac{\sum n_i + 1}{c_{\mathrm{lig}} \sum T_i} \tag{3.3.1}$$

$$k_{\mathrm{off}} = \frac{\sum n_i + 1}{\sum T_i} \tag{3.3.2}$$

また、それぞれの標準誤差 σ_{on} および σ_{off} について、

$$\sigma_{\mathrm{on}}^2 = \frac{\sum n_i + 1}{(c_{\mathrm{lig}} \sum T_i)^2} \tag{3.3.3}$$

$$\sigma_{\mathrm{off}}^2 = \frac{\sum n_i + 1}{(\sum T_i)^2} \tag{3.3.4}$$

とすることにより算出した。

リガンドの結合および解離状態の判定は 3 次元密度分布におけるそれぞれのピークからの距離に基づいて行った。シミュレーションにおける各リガンド分子の重心位置と、3 次元密度分布の各領域について密度が極大となるセルの中心との距離のうち、最小のものを d_{min} と定義し、この値により結合および解離を以下のように判定した。

結合状態： d_{min} が r_{unbound} 以上となった時点から、次に r_{bound} 以下になるまでの間。

解離状態： d_{min} が r_{bound} 以下となった時点から、次に r_{unbound} 以上になるまでの間。

解離と結合の判定に用いた r_{bound} および r_{unbound} は、ここではそれぞれ 3 Å、10 Å とした (付録 B.2)。判定した結合および解離状態を元に、シミュレーションにおいて結合および解離にかかった時間と結合、解離の生じた回数を数え上げることで速度定数を計算した。

さらに、実験値がわかっているものについて比較を行う上で、リガンド結合位置付近の結合サイト S1 についてより詳細に調べるために、リガンドの配座を考慮した速度定数の推定を行った。最初に、全シミュレーションにおけるリガンドのスナップショットのうち、ピークの周囲 10 Å以内のものについて、RMSD を指標としたクラスタリングを GROMOS アルゴリズム [50] を用いて行った。アルゴリズムの概要を以下に示す。

GROMOS アルゴリズム：

1. 各構造の間の RMSD を計算。

2. 各構造について、その構造からのRMSDがカットオフパラメータ以下である構造のリストを作成。

3. 2で計算したリストのうち、構造の数が最大のものをクラスタとして取り出す。RMSDの基準となった構造をクラスタの代表構造とする。

4. 3で選んだクラスタをデータセットから除外し、残りについて再度2に戻って繰り返す。

クラスタリングのカットオフパラメータは1 Åを用いた。得られたクラスタにおいてランクが10以上のものの代表構造のうち、結晶構造におけるリガンドとのRMSDが最も小さいものを選択し、その構造を結合状態の基準とした。ここで、結晶構造におけるリガンドとのRMSDは、結晶構造を粗視化モデルにマッピングし、対応する粒子同士のRMSDを計算することにより計算した。シミュレーションにおける各リガンド分子と基準構造とのRMSDのうち最小のものを$\mathrm{RMSD_{min}}$とし、結合および解離を以下のように定義した。

結合状態： $\mathrm{RMSD_{min}}$が$\mathrm{RMSD_{unbound}}$以上となった時点から、次に$\mathrm{RMSD_{bound}}$以下になるまでの間。

解離状態： $\mathrm{RMSD_{min}}$が$\mathrm{RMSD_{bound}}$以下となった時点から、次に$\mathrm{RMSD_{unbound}}$以上になるまでの間。

ここでは、$\mathrm{RMSD_{bound}}$および$\mathrm{RMSD_{unbound}}$は、それぞれ2 Å、15 Åとした(付録B.2)。

それぞれの系について、各ピークにおける結合、解離の速度定数k_{on}、k_{off}および解離定数$K_{\mathrm{d}}(=k_{\mathrm{on}}/k_{\mathrm{off}})$を計算した結果を表3.5にまとめた。また、結合サイトS1への結合、解離の観測数を表3.6に示した。なお、これらの表において、RMSDを指標とした結合判定を行った場合の結果をS1’として記した。

表 3.5: 各シミュレーションより推定された速度パラメータ。Experimental data については、PAL を除きシミュレーションに用いたものと同じ種類のリガンドに対するデータとなっている。

	Experimental data	Site	k_{on} / $10^7 M^{-1} s^{-1}$	k_{off} / $10^6 s^{-1}$	K_d / mM ($= k_{off}/k_{on}$)
DAHPS		S1	6.18 ± 0.55	0.151 ± 0.021	2.44 ± 0.40
		S2	391 ± 6	9.77 ± 0.15	2.50 ± 0.10
		S3	739 ± 6	43.2 ± 0.4	5.84 ± 0.07
		S4	574 ± 13	3.85 ± 0.09	0.672 ± 0.022
		S5	164 ± 3	14.1 ± 0.2	8.59 ± 0.18
AST	$K_d = 19$ mM [51]	S1	4.39 ± 0.25	1.11 ± 0.07	25.3 ± 2.1
		S1'	2.54±0.19	0.551±0.048	21.7 ± 2.5
		S2	530 ± 0	45.1 ± 0.3	8.51 ± 0.09
		S3	439 ± 3	45.1 ± 0.4	10.3 ± 0.1
		S4	1490 ± 10	89.3 ± 0.5	6.00 ± 0.00
6PGL		S1	6.43 ±0.95	0.0173 ± 0.0170	0.269 ± 0.270
		S2	19.8 ± 1.2	10.9 ± 0.7	54.9 ± 4.9
		S3	45.6 ± 2.1	12.5 ± 0.6	27.3 ± 1.8
		S4	130 ± 10	2.22 ± 0.16	1.71 ± 0.16
HAP	$K_i = 0.2$ mM[32]	S1	33.6 ± 1.2	5.44 ± 0.21	16.2 ± 0.9
		S1'	16.6 ± 0.08	2.71 ± 0.15	16.3 ± 1.2
		S2	234 ± 2	174 ± 2	74.3 ± 1.0
		S3	276 ± 3	57.9 ± 0.7	21.0 ± 0.4
		S4	75.1 ± 17	9.72 ± 0.23	1.30 ± 0.04
PNP	$K_d = 0.8$ mM[33]	S1	2.62 ± 0.23	0.181 ± 0.027	6.91 ± 1.20
		S1'	0.405 ± 0.071	0.012 ± 0.012	2.97 ± 3.00
		S2	6.00 ± 0.44	0.274 ± 0.028	4.56±0.58
		S3	7.31 ± 0.41	0.970 ± 0.060	13.3 ± 1.1
		S4	2.88 ± 0.30	0.0840 ± 0.0160	2.91 ± 0.63
NdRT		S1	1.76 ± 0.18	0.0778 ± 0.0170	4.43 ± 1.10
		S2	10.1 ± 0.6	0.485 ± 0.036	4.80 ± 0.50
		S3	43.5 ± 1.7	1.25 ± 0.05	2.88 ± 0.17
		S4	12.7 ± 0.5	2.48 ± 0.1	19.6 ± 1.1
		S5	36.1 ± 1.0	3.15 ± 0.09	8.73 ± 0.36
		S6	54.3 ± 1.9	1.74 ± 0.06	3.21 ± 0.16
Lev	$k_{on} = 6.6 \times 10^5 M^{-1} s^{-1}$	S1	3.04 ± 0.43	0.243 ± 0.056	8.02 ± 2.16
	$k_{off} = 8.3 \times 10^3 s^{-1}$	S1'	0.743 ± 0.180	0.0654 ± 0.0462	8.80 ± 6.58
	($K_d = 13$ mM)[52]	S2	4.43 ± 0.47	1.43 ± 0.17	32.4 ± 5.1
		S3	3.19 ± 0.46	0.297 ± 0.061	9.31 ± 2.32
		S4	2.61 ± 0.35	1.28 ± 0.18	49.2 ± 9.6
		S5	2.82 ± 0.38	0.611 ± 0.105	21.7 ± 4.7
		S6	29.3 ± 1.2	13.5 ± 0.6	46.2 ± 2.7
		S7	4.90 ± 0.48	2.01 ± 0.21	41.1 ± 6.0
		S8	13.1 ± 0.8	8.57 ± 0.51	65.2 ± 5.5
PAL	$K_i = 0.53$ mM[37]	S1	7.17 ± 0.22	64.8 ± 2.0	90.4 ± 3.9
	(Ligand: mannose)	S1'	4.09 ± 0.17	40.1±1.6	98.1 ± 5.6
		S2	0.233 ± 0.040	1.41 ± 0.29	60.7 ± 16.0
LinB	$K_i = 2.31$ mM[38]	S1	2.32 ± 0.26	4.55 ± 0.58	196 ± 33
		S2	255 ± 3	1230 ± 13	484 ± 7
PGA	$K_i = 0.82$ mM[40]	S1	0.286 ± 0.060	0.0932 ± 0.0470	32.6 ± 18.0
		S1'	0.261 ± 0.057	0.0468 ± 0.0330	17.9 ± 13.0
		S2	523 ± 5	55.7 ± 0.5	10.7 ± 0.1
		S3	181 ± 9	1.62 ± 0.08	0.892 ± 0.062
		S4	125 ± 2	26.4 ± 0.4	21.2 ± 0.5
		S5	161 ± 3	20.3 ± 0.3	12.6 ± 0.3
		S6	83.5 ± 2.1	7.63 ± 0.19	9.13 ± 0.33
KSI	$K_i = 0.135$ mM[42]	S1	100 ± 6	0.496 ± 0.041	0.495 ± 0.052
		S1'	54.4 ± 4.8	0.102 ± 0.019	0.187 ± 0.038
		S2	39.4 ± 1.3	11.7±0.4	29.8 ± 1.4
PDE4D	$IC_{50} = 0.082$ mM[44]	S1	80.8 ± 7.6	0.273 ± 0.034	0.338 ± 0.053
		S1'	6.70 ± 1.00	0.0189 ± 0.0110	0.282 ± 0.170
		S2	7.87 ± 0.93	0.289 ± 0.048	3.67 ± 0.75
		S3	24.7 ± 2.5	0.266 ± 0.037	1.08 ± 0.19

表 3.6: 結合位置付近の結合サイト S1 への結合、解離の観測数

	S1		S1'	
	n_{bind}	n_{unbind}	n_{bind}	n_{unbind}
DAHPS	127	52		
AST	313	262	181	129
6PGL	45	0		
HAP	769	693	405	332
PNP	128	44	32	0
NdRT	94	20		
Lev	49	18	16	1
PAL	1054	1045	631	623
LinB	80	60		
PGA	22	3	20	1
KSI	245	145	129	29
PDE4D	111	62	43	2

12 個の系のうち、6PGL、Lev、LinB、KSI、PDE4D の 5 つの系については S1 における親和性が最も高かった。またその他の系においても、S1 における k_{off} は他の領域に比べ小さい値をとる傾向にあった。このことから、リガンドが結合ポケットにおいて安定、あるいは 1 度結合した後長い時間滞在する傾向にあると考えられる。

また、実験データが明らかになっている 9 つの系について、実験値との比較を行った。AST、PNP、Lev、KSI、PDE4D の 5 つの系については、S1 および S1' における Kd の計算値と実験値の差は 10 倍以内であり、同程度のオーダーの値が得られた。したがって、これらの系においては、実験結果を再現できることが示唆される。ただし、PNP、Lev、PDE4D の S1' においては観測された解離の数が少ないため誤差が大きく、さらに長時間のシミュレーションが必要になると考えられる。リガンドの解離が遅い場合、解離が起こる前に有限の時間でシミュレーションが打ち切られてしまうことにより解離にかかる時間が小さく見積もられるため、さらに長時間のシミュレーションを行うことにより値が小さくなると考えられる。特に PNP は解離が 1 度も確認されていないため、シミュレーション時間を伸ばすことで実験の値に近くなることが予想される。

その他の系については実験値に比べ 10 倍〜100 倍程度大きい値が得られた。PGA については S1、S1' ともに解離の数が非常に少ないため、PNP 等と同様、長時間のシミュレーションが必要となる。ただし、この例に関しては結合、解離ともに速度定数が小さく、非常に遅い時間スケールとなることが予想されるため、効率的なサンプリングを行うための手法の開発が必要になると考えられる。一方、HAP、PAL、LinB については、ある程

度の数の結合、解離が観測されているが、K_d は実験値に比べ大きく異なった。これは粗視化モデルにおいて、全原子レベルでの相互作用を正確に表現できないことや、粗視化することによって粒子半径が大きくなることで、結合状態が安定でなくなるためと考えられる。HAP、PAL について、複合体構造を粗視化マッピングすると、図 3.9 のようになり、MARTINI の粒子の衝突距離 4.7 Å以内の距離にアミノ酸残基が存在する。一方全原子レベルではリガンドはこれらのアミノ酸残基と塩橋を形成し、強く相互作用しているため、この違いが安定性の違いに関わっていると考えられる。また、LinB に関しては、リガンドが chloride 原子を有しており、これが負電荷を持つ原子とハロゲン結合により相互作用する可能性がある [53, 54]。MARTINI では主に疎水性などの性質によって相互作用パラメータが決定されるため、こうした相互作用は考慮されない。実際に、LinB におけるタンパク質-リガンド複合体構造においては、リガンド付近に His 272 の Nε2 原子や Asp 108 の Oδ 原子が存在しており、こうした相互作用が考慮されないことによって差が生じたと考えられる。したがって、実験的な性質を再現するには、これらの影響を考慮したパラメータ設定が必要になる。3.3.3 節では、LinB と PAL についてこれらの影響を考慮したリガンドの粗視化力場パラメータの修正に関して記述した。

図 3.9: PAL (左) および AST (右) のタンパク質-リガンド複合体における MARTINI 粗視化マッピング。複合体構造とともに粗視化粒子を半透明で表している。緑の粒子はリガンドをマッピングしたものであり、青の半透明の粒子はリガンドの粗視化粒子の 4.7 Å以内にある粒子を表している。各粒子の数字はリガンド粗視化粒子との距離のうち最小のものを表す。

また、Lev に関しては速度定数 k_{on}、k_{off} の実験値も明らかになっており、これらの値と計算結果との比較を行うと、S1 の結果における k_{on}、k_{off} の結果は実験値に比べ 100 倍程度大きい値となった。一方、リガンドの結合様式を考慮した S1' の結果においては、k_{on}、k_{off} は 10 倍程度の値となった。2.3 節において記述したとおり、MARTINI におけるダイナミクスはおよそ 3-8 倍速くなるため、この影響を考慮すると、S1' における結果は実験

値と比較して妥当な値であるといえる。このことから、Lev の系においては、速度的にも実際の性質を再現できているといえる。

3.3.3 リガンドのパラメータの修正

3.3.3.1 LinB

LinB の複合体において、リガンドは結合サイトの His272 の Nε2 および Asp108 の Oδ に近い位置に存在し、これらとリガンドの chloride 原子の間にハロゲン相互作用が働いていると考えられる。したがって、結合サイトのアミノ酸残基とリガンドの間の相互作用を強く評価することにより本来の性質を再現できると考え、アミノ酸残基に対応する粒子とリガンドの粒子の間の相互作用を強めて解析を行うことにより速度定数の実験値の再現を試みた。

リガンドとして用いていた粒子 C4 の代わりに新たにパラメータを設定した特殊な粒子 LC4 を定義し、用いて新たにリガンド結合シミュレーションを行った。LC4 粒子の相互作用は、histidine 残基において Nε2 原子を含む粒子に対応する SP1 粒子および aspargine 残基において Oδ 原子を含む粒子に対応する Qa 粒子との間の相互作用について、Lennard-Jones potential (式 2.2.1) におけるパラメータ ε を 6-14 kJ mol^{-1} の間で設定し、それ以外の相互作用は従来の C4 粒子を同じパラメータを用いた。シミュレーションは通常のパラメータで行ったものと同様の条件を用い、各 ε について 1μs のシミュレーションを 50 回行った。

ε を変化させたときの k_{on}、k_{off}、K_d の値の変化に関するプロットを図 3.10 に示す。k_{off} および K_d は ε が 9 kJ mol^{-1} になるまで減少し、$\varepsilon > 9$ kJ mol^{-1} においてほぼ一定となった。

図 3.10: ε に対する速度パラメータの依存性 (左：結合、解離の速度定数。右：解離定数)。

次に、$\varepsilon = 9$kJ mol^{-1} についてさらに 50 回追加のシミュレーションを行い、計 100 回のシミュレーション結果についてリガンドの密度分布および速度定数に関する解析を行っ

た。図 3.11 および表 3.7 に結果を示す、

図 3.11: LinB についてパラメータを修正した場合のリガンド結合シミュレーションにおける 3 次元密度分布。緑のメッシュはリガンド密度 $80c_{\mathrm{lig}}$ における等位面を表し、マゼンタは複合体結晶構造におけるリガンドを表す。密度の高い領域のうち、複合体構造における結合位置に最も近いサイトを S1 とし、その他をを S2, S3, S4 として示している。

表 3.7: LinB についてパラメータを修正した場合のリガンド結合シミュレーション結果から計算した速度パラメータ。

	Site	k_{on} / $10^7\mathrm{M}^{-1}\,\mathrm{s}^{-1}$	k_{off} / $10^6\mathrm{s}^{-1}$	K_{d} / mM ($=k_{\mathrm{off}}/k_{\mathrm{on}}$)
LinB	S1	8.98 ± 0.86	0.271 ± 0.062	3.02 ± 0.75
	S2	5.40 ± 0.47	1.86 ± 0.21	34.4 ± 5.0
	S3	244 ± 3	115 ± 2	47.4 ± 0.9
	S4	477 ± 5	142 ± 2	29.8 ± 0.5

リガンドの 3 次元密度分布において、結晶構造におけるリガンド付近で密度が高い領域 S1 が存在した。また、相互作用を変化させたことにより、それに対応してタンパク質原子の histidine 残基の Nε2 原子および aspargine 残基の Oδ 原子に近い位置にいくつかの安定な領域が新たに見出された。これらの領域に対する解離定数を計算すると、S1 において最も小さい値が得られ、ポケット内が最も安定となった。解離定数の値は 3.02 mM であり、実験値 ($K_{\mathrm{i}} = 2.31\,\mathrm{mM}$) と非常に近い値をとったことから、パラメータの修正によりエネルギー的な性質を実際の系に近づけることができたといえる。なお、ハロゲン結合の強さは $10\,\mathrm{kJ\,mol}^{-1}$ 程度のオーダーであることが過去の研究から示唆されており [53]、今回設定したパラメータはこれに近い値であるため妥当なものであると考えられる。

3.3.3.2 PAL

PAL について複合体構造を粗視化原子にマッピングすると、タンパク質粒子とリガンド粒子の間の距離が衝突半径以下になっており、立体反発のため結晶構造が安定とならないことが考えられる。したがって、粗視化に伴う立体反発を弱めることで結合サイトにおいてリガンドが安定となると考え、アミノ酸残基とリガンド内の粒子の間の衝突半径を小さく修正して解析を行った。

はじめに、結合状態におけるリガンドとその周囲のタンパク質原子の距離を調べるため、複合体構造 (PDBID:2ARE) を初期構造とした全原子シミュレーションを行った。シミュレーションは、複合体構造を約 23000 分子の TIP3P 水分子 [55] に溶解し、さらに Na^+ および Cl^- を塩濃度が 50 mM で系中の電荷の総和が 0 になるように加えた系について行った。タンパク質の力場は Amber99SB-ildn[56] を用い、糖の力場には GLYCAM06[57] を用いた。310 K、1 bar の定温定圧シミュレーションを 100 ns 行い、シミュレーション中におけるリガンドと周囲のタンパク質原子との距離を解析した。

シミュレーション結果のトラジェクトリに対し、MARTINI 粗視化原子に対応する各部分構造の重心位置を計算し、MARTINI 粗視化モデルのマッピングを行った。このようにして得られた仮想的な粗視化トラジェクトリに対し、リガンド粒子とその周囲のタンパク質粒子の間の距離を計算した。表にシミュレーション中におけるリガンドからの平均距離が MARTINI における衝突半径 0.47 nm 未満であった粒子の種類と、その平均距離を記した。

表 3.8: PAL の複合体の全原子シミュレーション結果から算出したリガンド-タンパク質粗視化粒子間の平均距離。0.47 nm 未満であったタンパク質粒子の種類と、その平均距離を表記している。

Ligand atom	Protein CG atom (Atom type)	Distance / nm
B1	Phe 132 SC3 (SC5)	0.444
B1	Asp 86 SC1 (Qa)	0.421
B2	Gly 105 BB (P5)	0.397
B2	Gly 106 BB (P5)	0.391
B3	Gly 220 BB (P5)	0.380
B3	Gly 221 BB (P5)	0.417

これらのタンパク質粗視化粒子との間の相互作用の調整を　と同様にリガンドに特殊な粒子を割り当てることで実現した。新たに LP1、LP4 粒子を定義し、Glucose の B1 において P1 の代わりに LP1、B2 および B3 に P4 の代わりに LP4 粒子を用いた。表 3.8 における結果に基づいて、これらの距離が安定になるように LP1、LP4 とその他の粒子の間

の相互作用における衝突距離 σ の値を設定した。衝突距離は、Lennard-jones potential (式 2.2.1) において上で求めた平均距離が最も安定となるように、平均距離の $2^{-1/6}$ 倍の値を用いた。また、原子種について同じ組み合わせが存在した場合には、それらのうち小さいものを用いることとした。最終的に設定したパラメータ σ の値を表 3.9 に示す。表に示していない相互作用については、従来の P1 および P4 と同じパラメータを用いた。

表 3.9: 新たに設定した粗視化粒子 LC1、LC4 における相互作用パラメータ。表に示していない相互作用については、従来の P1 および P4 と同じパラメータを用いた。

Atom type (Ligand)	Atom type	σ / nm
LP1	Qa	0.38
LP1	P5	0.34
LP1	SC5	0.40
LP4	Qa	0.38
LP4	P5	0.35

新たに設定した粒子種を用いて、再度タンパク質-リガンド結合シミュレーションを行った。シミュレーション時間は 5μs×20 とし、それ以外の条件については通常のパラメータにおいて行ったものと同様に実行した。シミュレーション結果について 3.3.2 節と同様にリガンドの 3 次元密度および速度定数の解析を行った結果を図 3.12 および表 3.10 に示す。

図 3.12: PAL において修正したパラメータを用いた場合のリガンド結合シミュレーションにおける 3 次元密度分布。緑のメッシュはリガンド密度 $150c_{\mathrm{lig}}$ における等位面を表し、マゼンタは複合体結晶構造におけるリガンドを表す。密度の高い領域のうち、複合体構造における結合位置に最も近いサイトを S1 とし、その他を S2 として示している。

表 3.10: PAL において修正したパラメータを用いた場合のリガンド結合シミュレーション結果から計算した速度パラメータ。

Site	k_{on} / $10^7\mathrm{M}^{-1}\,\mathrm{s}^{-1}$	k_{off} / $10^6\mathrm{s}^{-1}$	K_{d} / mM ($= k_{\mathrm{off}}/k_{\mathrm{on}}$)
S1	16.6 ± 2.2	0.524 ± 0.081	3.15 ± 0.64
S1'	10.4 ± 1.7	0.263 ± 0.057	2.53 ± 0.69
S2	231 ± 9	7.56 ± 0.30	3.28 ± 0.18

ポケット以外の箇所でほぼ同程度に安定な領域が存在したものの、リガンドは複合体構造における結合位置で最も安定であった。また、S1 および S1' における解離定数は実験値のおよそ 5 倍程度であった。通常のポテンシャルを用いた場合の結果がおよそ 90 mM であり、実験値の 180 倍近い値であったことを考えると、大きく改善されたといえる。

3.3.4 結合経路の解析

3.3.4.1 分子流束密度

リガンドが結合過程においてどのような経路を通って結合しているのかを調べるための解析を行った。はじめに、シミュレーション中におけるリガンド結合過程に関するトラジェクトリを抽出した。トラジェクトリの抽出は、リガンドが結合位置から半径 r_{unbound} 以内に到達してから、その後 r_{unbound} 以上離れることなく r_{bound} 以内に接近した場合に、r_{unbound} から r_{bound} に接近するまでの間をリガンド結合過程として考え、その間のリガンドの座標を考慮することとした。ここでは、r_{bound} を 3 Å とし、r_{unbound} については電荷を持つリガンドが結合する DAHPS、AST、6PGL、HAP およびとくにリガンドが深い位置に結合する PGA について 30 Å とし、それ以外の系については 20 Å として抽出した。また、結合位置はそれぞれの系におけるサイト S1 の中心とした。結合、解離の判定に関して、全体のシミュレーションより得られたリガンド結合過程のトラジェクトリに対し、以下の解析を行った。

3.3.4.2 分子流束密度

リガンドの結合過程における分子流束密度の計算を行った。空間を一辺 2.5 Å のグリッドに分割し、各グリッドに対して、分子流束密度 $\mathbf{j}_i$ を以下の式により計算した。

$$\mathbf{j}_i = \frac{1}{TV}\sum_{t=1}^{T}\delta(\mathbf{x}_t - \mathbf{c}_i)\mathbf{v}_t \tag{3.3.5}$$

ここで、$\mathbf{x}_t$ および $\mathbf{v}_i$ はそれぞれ t 番目のスナップショットにおける重心座標位置および重心速度を表し、T は結合過程のトラジェクトリの長さ、V は各セルの体積距離を表す。$\mathbf{c}_i$ はセル i の中心の座標であり、$\delta(\mathbf{x}_t - \mathbf{c}_i)$ は $\mathbf{x}_t$ がセル i の内部にあるとき 1、それ以外の時 0 となる。この式は、各セルに対して、結合リガンドの速度の和をとり、トラジェクトリ全体の長さとグリッドの体積で割り平均化したものを意味する。このようにして計算したベクトルを複合体結晶構造の重ね合わせて図示した結果を図 3.13〜3.15 に示す。

図 3.13: 分子流束密度のベクトルマップ (分類 I (DAHPS、AST) および分類 II (6PGL、HAP))。図は分子流束密度と複合体構造のタンパク質表面を併せて描いており、黒い点線で囲った部分が基質結合ポケットを示す。タンパク質表面は赤が負電荷、青は正電荷を持つアミノ酸を表し、その他のアミノ酸については Kyte-Doolittle の疎水性指標に基づき、疎水性の高いものから橙～白～水色と色分けされている。緑および黄色の矢印は分子流束密度の計算結果をベクトルで表したものであり、緑色に近いほど流束密度の値が高いことを示す。値の範囲はそれぞれ 0.002～0.005 molecules $nm^{-2}ns^{-1}$(DAHPS)、0.01～0.05 molecules $nm^{-2}ns^{-1}$(AST)、0.002～0.02 molecules $nm^{-2}ns^{-1}$(6PGL)、0.05～0.1 molecules $nm^{-2}ns^{-1}$(HAP) であり、図には下限の値より大きいものについてのみ表示している。

図 3.14: 分子流束密度のベクトルマップ (分類 III (PNP、NdRT) および分類 IV (Lev、PAL))。黒い点線で囲った部分は基質結合ポケットを示す。タンパク質表面は Kyte-Doolittle の疎水性指標に基づき、疎水性の高いものから橙～白～水色と色分けされている。緑および黄色の矢印は分子流束密度の計算結果をベクトルで表したものであり、緑色に近いほど流束密度の値が高いことを示す。値の範囲はそれぞれ 0.001～0.005 molecules $nm^{-2}ns^{-1}$(PNP)、0.002～0.02 molecules $nm^{-2}ns^{-1}$(NdRT)、0.02～0.05 molecules $nm^{-2}ns^{-1}$(Lev)、0.005～0.01 molecules $nm^{-2}ns^{-1}$(PAL) であり、図には下限の値より大きいものについてのみ表示している。

図 3.15: 分子流束密度のベクトルマップ (分類 V (LinB、PGA) および分類 VI(KSI、PDE4D))。黒い点線で囲った部分は基質結合ポケットを示す。タンパク質表面は Kyte-Doolittle の疎水性指標に基づき、疎水性の高いものから橙 ~ 白 ~ 水色と色分けされている。緑および黄色の矢印は分子流束密度の計算結果をベクトルで表したものであり、緑色に近いほど流束密度の値が高いことを示す。値の範囲はそれぞれ 0.02~0.05 molecules $nm^{-2}ns^{-1}$(LinB)、0.002~0.01 molecules $nm^{-2}ns^{-1}$(PGA)、0.01~0.1 molecules $nm^{-2}ns^{-1}$(KSI)、0.004~0.02 molecules $nm^{-2}ns^{-1}$(PDE4D) であり、図には下限の値より大きいものについてのみ表示している。

これらの結果において、リガンドの流束密度が周囲に比べて高い箇所を見出すことができた。このことから、リガンドは特定の経路を通って基質結合ポケットへ侵入していく傾向が高いことが分かった。はじめに、電荷を持たないリガンドが結合する分類 III~VI の

系について傾向を見ると、タンパク質は分子表面に存在する溝の近くを通過する傾向にあった。また、親水性のリガンドは比較的親水性の表面付近を通過する傾向が見られ、疎水性のリガンドは疎水性の表面付近を通過する傾向が見られた。一方、電荷を持つリガンドが結合する分類 I および II の例では、タンパク質表面の起伏によらず、矢印は正電荷を持つアミノ酸の付近に存在する傾向があった。今回のシミュレーションにおいて用いたリガンドは全て負電荷を持つものであり、電荷の相互作用が結合過程に大きく影響することが示唆される。

3.3.4.3 タンパク質表面の性質と結合経路の関係の数値的な評価

次に、結合経路がどの程度タンパク質表面の形状や疎水性と関係しているかを数値的に理解するため、以下の指標を計算し、リガンド結合過程との関係を調べた。計算は全てタンパク質-リガンド複合体をシミュレーションにおけるタンパク質構造に重ね合わせた上で行った。取得したリガンド結合経路の軌跡に対して以下のような指標の計算を行った。

リガンド結合経路周囲のタンパク質表面の疎水性： 結合経路の周囲におけるタンパク質表面の疎水性に関する指標を計算した。最初に、タンパク質-リガンド複合体構造におけるアミノ酸残基について、溶媒接触表面積が 1 Å^2 以上のものを表面残基として抽出した。その後、リガンド結合経路のトラジェクトリにおける各重心座標において、周囲 7 Å以内に存在するアミノ酸残基に対し、Kyte-Doolittle 疎水性指標の平均値を計算した。

タンパク質表面における溝とリガンドとの距離： 結合経路において、リガンドがどの程度タンパク質表面の溝の近くを通過しているかの指標を計算した。最初に、タンパク質-リガンド複合体構造を用いて、タンパク質表面における溝の検出を行った。検出方法はタンパク質のポケット検出プログラム GHECOM[58, 59] による手法に倣い、小さい球が侵入でき、かつ大きい球が侵入できない領域を探索し、それを溝とした 3.16。大きい球および小さい球の半径をそれぞれ r_{large}、r_{small} としたときの具体的な計算の手順を以下に示す。

1. タンパク質の各原子について $r_{\mathrm{large}} + r_{\mathrm{atom}}$ 以内の領域を探索し、それを全て合わせた領域を取得する。ここで、r_{atom} は各原子の半径とする。

2. 1 で得られた領域について、その領域の表面から距離 $r_{\mathrm{large}} + r_{\mathrm{small}}$ 以内の領域を除外する。

3. 2 で得られた領域について、各タンパク質原子から $r_{\mathrm{small}} + r_{\mathrm{atom}}$ 以内の領域を除外する。

4. 3 で得られた領域について、その領域の表面から r_{small} 以内の領域を加える。

図 3.16: タンパク質表面の溝の検出。半径 r_{small} の小さい球が侵入でき、半径 r_{large} の大きい球が侵入できない領域をポケットとして検出した。

ここでは、r_{large} および r_{small} はそれぞれ 10 Å、1 Å とした。リガンド結合経路のトラジェクトリにおける各重心座標について、上記のように得られたタンパク質の溝の領域との間の最小距離を計算した。

このように決定した 2 つの指標について、リガンドが結合ポケットに近づくにつれてどのように変化するかを調べた。結合過程の各時点におけるリガンド結合位置からの距離を横軸とし、それぞれの指標を縦軸としたプロットをそれぞれ作成した。例として、6PGL および KSI について 5 回のリガンド結合イベントに対してプロットを作成した結果を図 3.16、3.18 に示す。疎水性に関するプロットに着目すると、6PGL については、リガンドの結合位置からの距離が 10 Å 以上離れているようなタンパク質表面付近において、親水性の残基の近くを通過する傾向にあることが分かる。一方、KSI においては、リガンドが結合位置から遠い位置でも疎水性残基の近くを通過していることが確認できる。また、溝からの距離に着目すると、6PGL では溝からの距離が 4~6 Å 離れたところにおいても結合経路が集中していたのに対し、KSI ではリガンドの結合位置から離れたタンパク質表面付近においても比較的溝に近い位置 (<2 Å) に結合経路が集中する傾向がみられた。このことから、それぞれの系において、リガンドの結合経路は異なった特徴を持っていることが考えられる。

図 3.17: リガンド結合経路周囲の疎水性に関するプロットの一部。横軸がリガンドの結合サイト S1 からの距離、縦軸はリガンド重心座標の周囲 7 Å以内のタンパク質表面残基における Kyte-Doolittle 疎水性指標の平均値とした。ただし、経路の連続性を表現するため、周囲にアミノ酸残基が存在しない場合を-5 としてプロットした。プロットは 5 回の結合イベントについてのみ表示しており、異なる色のプロットはそれぞれ異なる結合イベントを表す。

図 3.18: リガンド結合経路とタンパク質表面の溝との位置関係に関するプロットの一部。横軸がリガンドの結合サイト S1 からの距離、縦軸はリガンド結合過程における重心座標のタンパク質の溝からの距離を表す。プロットは 5 回の結合イベントについてのみ表示しており、異なる色のプロットはそれぞれ異なる結合イベントを表す。

続いて、これらの二次元プロット平面上においてリガンド結合経路がどの位置を通過している比率が高いかを全結合イベントから計算した。疎水性に関するプロットに対して、縦軸を 0.5、横軸を 0.5 Å ずつの区間に分割した後、各グリッドについて、全体の結合イベントのうち、そのグリッドの内部を通過して結合した結合イベント数の割合を計算した。また、タンパク質表面の溝からの距離に関するプロットに対しても同様に、縦軸を 0.5 Å、横軸を 0.5 Å ずつの区間に分割した後、各グリッドについて同様の値を求めた。疎水性に関する結果を 3.19~3.21 に、タンパク質表面の溝からの距離に関する結果を図 3.22~3.24

にそれぞれ示した。

図 3.19: リガンド結合経路周囲の疎水性に関するカラーマップ (分類 I (DAHPS、AST) および分類 II (6PGL、HAP))。横軸がリガンドの結合サイト S1 からの距離、縦軸はリガンド重心座標の周囲 7 Å以内のタンパク質表面残基における Kyte-doolittle 疎水性指標の平均値を表す。各グリッドの色は、シミュレーションから得られた結合イベント全てのうち、そのグリッドの内部を通過して結合した結合イベント数の割合を表す。また、点線は表面残基全体における Kyte-doolittle 疎水性指標の平均値を表しており、プロット右上に平均値を記載している。

図 3.20: リガンド結合経路周囲の疎水性に関するカラーマップ (分類 III (PNP、NdRT) および分類 IV (Lev、PAL))。

図 3.21: リガンド結合経路周囲の疎水性に関するカラーマップ (分類 V (LinB、PGA) および分類 IV (KSI、PDE4D))。

図 3.22: リガンド結合経路とタンパク質表面の溝との位置関係に関するカラーマップ (分類 I (DAHPS、AST) および分類 II (6PGL、HAP))。横軸がリガンドの結合サイト S1 からの距離、縦軸はリガンド結合過程における重心座標のタンパク質の溝からの距離を表す。各グリッドの色は、シミュレーションから得られた結合イベント全てのうち、そのグリッドの内部を通過して結合した結合イベント数の割合を表す。

図 3.23: リガンド結合経路とタンパク質表面の溝との位置関係に関するカラーマップ (分類 III (PNP、NdRT)、分類 IV (Lev、PAL))。

図 3.24: リガンド結合経路とタンパク質表面の溝との位置関係に関するカラーマップ (分類 V (LinB、PGA) および分類 IV (KSI、PDE4D))。

結合経路周囲の疎水性に対する傾向を見ると、親水性のリガンドが結合するような分類 I~IV の系では、リガンド結合位置から 10 Å程度以上離れたタンパク質表面付近において、表面残基全体の平均よりも親水性の高い部分を通過する傾向が高いことが明らかとなった。リガンド結合位置付近では疎水性が強くなる傾向がありこれはこれはポケットに進入するにつれて疎水コアに近づくためであると考えられ、特に狭い結合ポケットを持つ AST や PNP、NdRT などは表面残基全体の平均よりも疎水性が高くなった。一方、疎水性ガンドが結合するような分類 V および VI では、リガンド結合位置から 10 Å程度以上離れてもポケット内部と同程度の疎水性の領域を通る確率が高く、タンパク質表面における疎水性部分を通過する傾向にあることが見出された。

また、結合経路とタンパク質表面の溝との位置関係について考えると、電荷を持たないリガンドが結合する分類 III~VI の系では、溝からの距離が 2 Å以下の箇所に経路が集中する傾向が見られた。全体の確率の傾向から、リガンドは、結合位置からの距離が PGA において 25 Å程度、その他のケースにおいて 15~20 Å程度の距離でタンパク質表面に着地し、その後、溝を通過しながら結合ポケットに侵入していく性質があると考えられる。このことから、タンパク質表面における起伏は結合経路を決定づける要因の一つになることが予想される。一方、リガンドが電荷を持つような分類 I および II の系では、溝から 4~6 Å程度離れた領域においても 0.3 以上の値を持つグリッドが多く見られ、電荷を持たないリガンドの場合に比べ、タンパク質の形状の影響が小さい傾向があった。電荷を持つリガンドにおいては、タンパク質表面残基との静電相互作用が存在し、これが強く影響していると考えられる。

以上の結果は、分子流束密度において見られた傾向に一致しており、タンパク質表面の形状や疎水性などの数値的な指標とリガンドの結合経路の傾向を関連づけられる可能性を示唆している。

3.3.5 水とリガンドの交換

今回の研究では、リガンドのダイナミクスに焦点を置いた解析を行ったが、最後にリガンドと水の交換について軽く触れる。リガンドの結合は多くの場合、水との交換を伴う。したがって、リガンドの安定性や結合速度は水の挙動に影響されるため、ポケットにおける水との相互作用を理解することは重要となる。

今回のシミュレーション結果における例として、phenol が疎水性の狭い結合ポケットに結合する例である PGA、および phenol が疎水性の広い結合ポケットに結合する例である KSI に関して水とリガンドが交換する様子を図 3.25 および図 3.26 に示した。これらの結果を比較すると、PGA では最初にリガンド結合サイトに存在した水が解離し、そこに生じた空間にリガンドが入るという形での交換が見られた。一方、KSI ではリガンドが結合

するとき、同時に水がリガンドをよけるように抜けていく様子が見られた。これらのメカニズムの違いは、ポケットの広さによるものであり、特に狭いポケットではポケット内における水の挙動がリガンドの結合に大きく影響すると考えられる。実際に、狭い疎水性ポケットを有するタンパク質においては、ポケット内に水分子が少ない状態をとりやすい傾向にあることがシミュレーションおよび実験により示唆されている [60, 61, 62]。また、節における速度定数の解析結果から、PGA の結合、解離の速度定数は KSI に比べて小さく、水とリガンドの交換のメカニズムの違いがリガンドの結合、解離の時間スケールに影響するものと予想される。

こうした結果から、基質結合ポケットの広さによって水とリガンドの交換が異なるメカニズムで起こることが示唆され、またこのようなメカニズムの違いがリガンドの結合速度などに大きく影響するものと予想される。今後は、こうしたポケットの広さによる水の挙動の違いや、それが速度定数に与える影響などについても詳細に解析していきたいと考えている。

図 3.25: PGA における水とリガンドの交換。茶色のスティックモデルがタンパク質の主鎖構造、緑のスティックモデルがリガンド、水色が粗視化水分子を表す。黒の点線で囲った部分はリガンド結合サイトを示す。

図 3.26: KSI における水とリガンドの交換。茶色のスティックモデルがタンパク質の主鎖構造、緑のスティックモデルがリガンド、水色が粗視化水分子を表す。黒の点線で囲った部分はリガンド結合サイトを示す。

3.4 まとめ

それぞれ異なる性質を持つ多数のタンパク質-リガンドペアの系に対して MARTINI 粗視化力場を用いたリガンド結合シミュレーションを行った結果、いずれの系においてもリガンドが溶液中に溶解している状態から基質結合ポケットへ結合した状態を再現することに成功した。その後、3 次元密度分布による解析および結合親和性に関する解析から、リガンドは基質結合ポケット内部で比較的安定に存在することを見出した。また、結合、解離の速度や結合親和性について実験値との比較を行った結果、エネルギー的、速度的な性質を概ね正しく表現できていることが示された。一部の系については従来の MARTINI の力場のままでは実験結果と一致しないものが存在したが、MARTINI では扱われない相互作用や、全原子シミュレーションにおける結果を考慮してパラメータを修正することにより、MARTINI の枠組みの中で実験結果の再現が可能であることを示した。これらのことから、MARTINI 粗視化力場によるリガンド結合シミュレーションによって、実験結果を再現できる妥当性の高い解析を行うことができることが示唆された。ただ一方では、解

離に関するサンプリングが十分でなく、親和性が正しく評価できない系も存在した。このような例に対してはさらに長時間のシミュレーションが必要になることが考えられるが、タンパク質-リガンドの系によっては結合、解離時間スケールが非常に遅いことが考えられ、効率的なサンプリング手法の開発が必要になると考えられる。

さらに、結合経路に関して解析を行った結果、リガンドが特異的な経路を通過して基質結合ポケットに侵入することが見出された。さらに、タンパク質表面における形状や物理化学的性質との関係を見ることで、リガンドの性質によって異なる特徴を持つ結合経路を通過していることが明らかとなった。親水性のリガンドは親水性のタンパク質表面、疎水性のリガンドは疎水性の表面付近を通過する傾向にあることを見出し、リガンドの疎水性は結合位置周辺のみで無くタンパク表面付近におけるリガンドの経路についても影響していることが分かった。また、タンパク質における溝との位置関係を解析することで、リガンドはタンパク質表面における溝に沿って結合する傾向にあることが示唆された。ただし、電荷を持ったリガンドについては電荷による相互作用が強く働くため、電荷を持つアミノ酸残基に強く影響を受けることが示唆された。これらの結果より、タンパク質の表面における溝や疎水性といった一般的な性質とリガンドの結合経路の関係性を結びつけることができる可能性が示唆された。

また、今回のシミュレーション結果から、基質結合ポケットにおける水とリガンドの交換がタンパク質によって異なるメカニズムで生じていることが確認された。これは基質結合ポケットの広さによる違いによる生じるものであると予想され、また速度定数の解析結果から水とリガンドの交換メカニズムの違いが結合、解離速度の時間スケールに影響していると考えられる。今後は、こうしたポケットの広さによる水の挙動の違いや、それが速度定数に与える影響などについても詳しく解析していきたいと考えている。

第4章　Dual-basin potentialによるMARTINI粗視化シミュレーションの拡張

2.4 節において述べたように、MARTINI におけるシミュレーションにおいてはタンパク質の構造を束縛する必要があり、そのためにタンパク質の大きな構造変化を伴う系については適用不可能となる。一方、タンパク質とリガンドの結合の過程でタンパク質の大きな構造変化を伴うような例は数多く知られており、一般的にタンパク質-リガンド結合過程を調べる上ではタンパク質の構造変化を扱うことが必要となる。この問題を克服するため、dual-basin potential[63, 64] の手法を MARTINI 粗視化力場と組み合わせた拡張を行うことで、構造変化を表現することを試みた。この章では、dual-basin potential による MARTINI 粗視化シミュレーションの拡張方法について触れた後、この手法を用いたタンパク質の構造変化シミュレーションおよび構造変化を伴うタンパク質-リガンド結合の解析について述べる。

4.1　Dual-basin potential

通常、粗視化力場を用いたタンパク質のシミュレーションを行う場合には基準となる構造が存在し、その構造が安定となるような構造ベースの力場に基づいて計算が行われるため、大きな構造変化を表現することが困難である。これに対し、タンパク質があらかじめ異なる複数の立体構造をとることがわかっている場合に、複数の構造ベースのポテンシャルをつなぎ合わせることでそれらの間の構造変化を扱う手法が開発されている。本研究ではこの考え方を MARTINI 粗視化力場におけるシミュレーションに適用し、構造変化の表現を試みた。

今、タンパク質が 2 つの異なる配座の間で構造変化する場合を考え、それらの構造が明らかとなっているとする。2 つの参照構造を元にした構造ベースのポテンシャル $V_1(\mathbf{x})$ および $V_2(\mathbf{x})$ に対して、以下のような dual-basin potential を計算することにより、構造変化を表現する。

$$V_{\mathrm{db}}(\mathbf{x}) = \frac{V_1(\mathbf{x}) + V_2(\mathbf{x}) + \Delta V}{2} - \sqrt{\left(\frac{V_1(\mathbf{x}) - (V_2(\mathbf{x}) + \Delta V)}{2}\right)^2 + \Delta^2} \tag{4.1.1}$$

ここで、ΔV はポテンシャル $V_2(\mathbf{x})$ に対してエネルギーの高さを変化して 2 つの状態の存在確率を調節するパラメータであり、正であれば構造 1 の存在確率が、負であれば構造 2 の存在確率が上昇する。Δ は 2 つの状態間の状態遷移の障壁の高さを調節するポテンシャルでありこの値を大きくすることでそれぞれの状態間の遷移の頻度が上がる。このポテンシャルは、2 つの基底を持つ図 4.1 ような形となる。それぞれの参照構造 1 および 2 付近では元の構造ベースポテンシャルに近い形となり、参照構造付近で安定となる。

図 4.1: Dual-basin potential のポテンシャル形状。

この dual-basin potential を MARTINI 粗視化シミュレーションと組み合わせるため、タンパク質の MARTINI 粗視化モデルにおいて構造に依存する部分のポテンシャルを用いて式 4.1.1 を計算することとした。構造 n に対する構造ベースのポテンシャルを $V_n(\mathbf{x})$ とする。本研究で用いた ELNEDYN モデルにおいて、$V_n(\mathbf{x})$ は隣接した残基の主鎖粒子間の結合に関するポテンシャル $V_n^{\mathrm{b}}(\mathbf{x})$、隣接していない残基の主鎖粒子間の距離を束縛する Elastic network ポテンシャル $V_n^{\mathrm{el}}(\mathbf{x})$ および、隣接した 3 つの残基の主鎖粒子がつくる結合角に関するポテンシャル $V_n^{\mathrm{a}}(\mathbf{x})$ の和で表される。

$$V_n(\mathbf{x}) = V_n^{\mathrm{b}}(\mathbf{x}) + V_n^{\mathrm{el}}(\mathbf{x}) + V_n^{\mathrm{a}}(\mathbf{x}) \tag{4.1.2}$$

$V_n^{\mathrm{b}}(\mathbf{x})$ については、構造の大きな変化はないものとし、2 つの構造における結合距離の平均を平衡距離とし、それぞれの構造で同じポテンシャルを用いた。

$$V_1^{\mathrm{el}}(\mathbf{x}) = V_2^{\mathrm{el}}(\mathbf{x}) = \sum_i \frac{k_{\mathrm{b}}}{2}\left(r_i(\mathbf{x}) - \frac{r_{i,1} + r_{i,2}}{2}\right)^2 \tag{4.1.3}$$

ここで、k_{b} は ELNEDIN における隣接主鎖粒子間の結合に関する力の定数、$r_i(\mathbf{x})$ は i 番目の結合の長さ、$r_{i,n}$ は参照構造 n における i 番目の結合の長さを表す。

束縛に関するポテンシャル $V_n^{\mathrm{el}}(\mathbf{x})$ については以下のように計算した。

$$V_n^{\mathrm{el}}(\mathbf{x}) = \sum_{(i,j)\in P_n} \frac{k_{\mathrm{el}}}{2}\left(r_{i,j}(\mathbf{x}) - r_{i,j,n}\right)^2 + \sum_{(i,j)\in Q} \frac{k_{\mathrm{el}}}{2}\left(r_{i,j}(\mathbf{x}) - \frac{r_{i,j,1}+r_{i,j,2}}{2}\right)^2 \tag{4.1.4}$$

ここで、非隣接残基の主鎖粒子のペアのうち、参照構造 n において距離がカットオフ距離 R_c 以内であり、参照構造 1 における距離と参照構造 2 における距離の差が 0.1 nm 以上であるものの集合を P_n、参照構造 1 における距離と参照構造 2 における距離の差が 0.1 nm 未満で、参照構造 1 と 2 における距離のいずれかが R_c 以内の集合を Q とした。$r_{i,j}(\mathbf{x})$ は粗視化粒子 i,j の間の距離、$r_{i,j,n}$ はそれぞれの参照構造における i,j の間の距離、k_{el} は束縛の力の定数を表す。

結合角に関するポテンシャル $V_n^{\mathrm{a}}(\mathbf{x})$ は次のように計算した。

$$V_n^{\mathrm{a}}(\mathbf{x}) = \sum_{(i,j)\in P_n} \frac{k_{\mathrm{a}}}{2}\left(\cos\theta_i(\mathbf{x}) - \cos\theta_{i,n}\right)^2 \tag{4.1.5}$$

ここで、k_{a} は ELNEDIN における隣接主鎖粒子間の結合に関する力の定数、$\theta_i(\mathbf{x})$ は i 番目の結合角の大きさ、$\theta_{i,n}$ は参照構造 n における i 番目の結合角の大きさを表す。

また、粗視化粒子の強い衝突を避けるため、2 つの参照構造の双方において側鎖の粗視化粒子間の距離が MARTINI における衝突半径より小さい場合には式 4.1.4 を用いて計算した。

k_{a} および k_{b} は ELNEDIN における規定の値を用い、束縛のカットオフ距離 R_c および強さ k_{el} については適宜設定するものとした。また、MARTINI 粗視化力場においては、二次構造 (coil,helix,helix (N/C-terminus),β-strand) により主鎖の粒子の種類が異なる [47]。そのため、ある残基において参照構造 1 と参照構造 2 で二次構造が異なる場合には、その残基の二次構造を coil とみなして粗視化原子を割り当てることとした。

束縛に関するポテンシャル $V_n^{\mathrm{el}}(\mathbf{x})$ については以下のように計算した。

4.2 タンパク質の構造変化シミュレーション

4.2.1 計算対象

最初に、構造変化を表現できるかどうかを調べるため、大きな構造の変化が起こるタンパク質について、タンパク質と水の系におけるシミュレーションを行った。計算対象には、*Thermotoga maritima* MSB8 由来の ribose binding protein および *Aquifex aeolicus* 由来

の adenylate kinase を選んだ (図 4.2)。これらはいずれも、タンパク質-リガンド複合体構造およびリガンドが結合していない構造が明らかとなっており、リガンドが結合していない状態で開いた構造、リガンドが結合した状態で閉じた構造をとる。それぞれのタンパク質に関して構造の変化を見ると、ribose binding protein においては開いた構造 (2FN9) と閉じた構造 (2FN8)[65] の間の RMSD は 3.25 Å、adenylate kinase については開いた構造 (2RH5) と閉じた構造 (2RGX)[66] の間の RMSD は 5.85 Å であり、大きく構造変化していることが分かる。

図 4.2: Ribose binding protein (左) および adenylate kinase (右) のタンパク質-リガンド複合体構造およびリガンド非結合型の構造。複合体構造を茶色、リガンド非結合型の構造を青、リガンドをマゼンタで表している。

4.2.2 計算手法

2 つのタンパク質について、束縛ポテンシャルを表 4.1 に記したパラメータにより構築した。

表 4.1: シミュレーションに用いたタンパク質構造および束縛ポテンシャル構築の際に用いたパラメータ

Protein	Reference structure (PDB)		R_c / nm	k_{el} /kJ mol^{-1} nm^{-1}	
	Open	Close		Backbone	Sidechain
ribose binding protein	2FN9	2FN8	1.4	100	100
Adenylate kinase	2RH5	2RGX	1.2	200	50

タンパク質-水の系は次のように構築し、温度 310 K、圧力 1 bar の条件下で１ μs のシミュレーションを行った。

Ribose binding protein タンパク質 1 分子、粗視化水分子 6400 分子

Adenylate kinase タンパク質 1 分子、粗視化水分子 6300 分子

計算はソフトウェア gromacs 4.0.7 を改変し、dual-basin potential の計算を可能としたものを用いた。シミュレーションは Δ および ΔV を変化させながら行い、１ μs のシミュレーション内で複数回構造の遷移が確認できるような条件を探索した。

4.2.3 結果

Δ および ΔV を調整しながらシミュレーションを行った結果、1μs のシミュレーションにおいて相互に構造変化が観測できるような Δ および ΔV を見つけることに成功した。Ribose binding protein について Δ= 450 kJ mol^{-1}、ΔV= 310 kJ mol^{-1}、adenylate kinase について Δ= 850 kJ mol^{-1}、ΔV= 1700 kJ mol^{-1} としたときのシミュレーションにおけるタンパク質構造の時間変化を図 4.3 に示す。これらの結果より、タンパク質が閉じた状態の結晶構造に近い構造と、開いた状態の結晶構造に近い構造を交互にとり、1μs のシミュレーション中において複数回の構造遷移が生じることを確認できた。したがって、MARTINI 力場と dual-basin potential を組み合わせることにより、たんぱく質の大きな構造変化を扱えることが示された。

図 4.3: 水-タンパク質の系のシミュレーションにおける構造変化の様子 (左：ribose binding protein、右：adenylate kinase)。上段の図は各タンパク質のシミュレーション中において観測された、開いた状態と閉じた状態の骨格構造をそれぞれ青と赤のモデルで示したものである。下段の図はシミュレーション中におけるタンパク質構造について、開いた状態と閉じた状態それぞれの結晶構造に対する主鎖の RMSD を計算し、時間プロットしたものである。青は開いた構造に対する RMSD、赤は閉じた構造に対する RMSD を表す。

4.3 タンパク質-リガンド結合シミュレーションへの適用

4.3.1 計算手法

Dual-basin potetial を適用した MARTINI 粗視化シミュレーションによりタンパク質の構造変化を再現することに成功したため、続いて、タンパク質の構造変化を伴うリガンド結合過程の解析を試みた。上の解析で用いた ribose binding protein について、glucose との結合シミュレーションを行った。シミュレーションの系はタンパク質の周りに 2 個のリガンド分子をランダムに配置し、6300 個の水分子に溶解することによって構築した。ここではリガンド分子の初期配置を変えながら 2μs のシミュレーションを 50 回行った。

4.3.2　結果

シミュレーションの結果、リガンドが複合体結晶構造と同様の位置への結合が全体のシミュレーション中で 13 回確認できた。このとき、リガンドが結合した状態でタンパク質構造が閉じ、複合体結晶構造と同様の構造を再現できることを確認した (図 4.4)。次に、複合体構造におけるリガンドの重心とシミュレーションにおけるリガンドの重心との間の距離の時間プロット、およびシミュレーション中における構造と閉じた構造との RMSD 値の時間プロットを作成した。10 回のシミュレーションのうちの 1 つについてプロットを行った結果を図 4.5 に示す。時間プロットから、リガンドはタンパク質構造が開いてるときに結合し、その後タンパク質構造が閉じて安定化する様子が見られた。また、リガンドの位置とタンパク質構造の関係をプロットしたところ、図の右のプロットのようになり、タンパク質が開いた構造をとっているときにリガンドが結合位置に接近する様子が見られる。

図 4.4: Ribose binding protein のリガンド結合シミュレーション中におけるリガンドの結合状態の構造と複合体結晶構造との重ね合わせ。茶色と緑色のモデルはそれぞれシミュレーションにおけるタンパク質の骨格構造とリガンドを表す。また青のリボンモデルとマゼンタのモデルは複合体構造のタンパク質構造およびリガンドを表す。

図 4.5: Ribose binding protein のリガンド結合シミュレーションにおけるタンパク質構造とリガンドの位置に関するプロット。プロットは 10 回行ったシミュレーションのうちの一回の結果について示している。左の図において、青はシミュレーションにおけるリガンドと複合体構造におけるリガンド結合位置との重心距離、赤はシミュレーション中におけるタンパク質構造と複合体構造の間の RMSD をそれぞれ時間プロットしたものである。右の図はリガンドの結合位置からの距離を横軸とし、同じ時点でのタンパク質構造の複合体構造からの RMSD を縦軸としたものであり、赤および青はそれぞれシミュレーションにおける別のリガンドを表す。

続いてリガンドの結合がタンパク質構造に及ぼす影響を理解するため、タンパク質構造変化、およびリガンドの結合状態に関して、状態変化の解析を行った。タンパク質とリガンドの状態を以下の 4 つに分類し、それぞれの間の遷移速度を計算した。

Unbound-open (UO): タンパク質が開いていてリガンドが結合していない状態

Bound-open (BO): タンパク質が開いていてリガンドが結合している状態

Unbound-close (UC): タンパク質が閉じていてリガンドが結合していない状態

Bound-close (BC): タンパク質が閉じていてリガンドが結合している状態

タンパク質の構造変化の判定は、シミュレーション中のタンパク質構造と、開いた構造 (2FN9) および閉じた構造 (2FN8) との RMSD を指標とした。開いた構造との RMSD が 1 Å以下となった時点から、閉じた構造との RMSD が 1 Å以下の状態になるまでの間を開いた状態、閉じた構造との RMSD が 1 Å以下となった時点から、開いた構造との RMSD が 1 Å以下の状態になるまでの間を閉じた状態と定義した。また、リガンド結合に関しては、シミュレーション中におけるタンパク質構造に対して複合体結晶構造を重ね合わせ、複合体のリガンドとシミュレーションにおけるリガンドとの重心距離によって判定した。リガンドが 10 Å以上離れた時点から 3 Å以内に近づくまでを解離状態とし、リガンドが 3 Å以内に到達した時点から 10 Å以上離れるまでを結合状態とした。シミュレーション

中におけるそれぞれの状態の存在する時間および状態間の遷移回数を数え上げ、付録 B.1 に準拠した方法を用い式 (B.1.5、B.1.8) により遷移速度の推定を行った。各状態間の遷移に対する速度定数の解析結果を図 4.6 に示す。解析結果において、リガンド結合の有無に対するタンパク質構造の平衡を考えると、リガンドが結合していない状態での平衡定数は k_1/k_2= 1.04 であるのに対し、結合した状態では k_3/k_4= 13.1 となった。このことから、リガンドが結合することによって、閉じた構造が安定に存在できるようになることが示唆された。

k_1= $1.81\times10^7\mathrm{s}^{-1}$
k_2= $1.88\times10^7\mathrm{s}^{-1}$
k_3= $3.20\times10^7\mathrm{s}^{-1}$
k_4= $3.99\times10^6\mathrm{s}^{-1}$
k_5= $1.13\times10^9\mathrm{M}^{-1}\,\mathrm{s}^{-1}$
k_6= $3.47\times10^7\mathrm{s}^{-1}$
k_7= $5.30\times10^8\mathrm{M}^{-1}\,\mathrm{s}^{-1}$
k_8= $8.03\times10^6\,\mathrm{s}^{-1}$

図 4.6: 速度定数の計算結果

4.4 まとめ

MARTINI 粗視化力場においてタンパク質の構造変化を扱うため、dual-basin potential と組み合わせた手法の開発を行った。MARTINI における束縛ポテンシャルを dual-basin potential に適用することで、タンパク質構造が開閉する大きな構造変化を扱うことに成功した。さらに、構造変化を伴うタンパク質とリガンドの系に適用し、リガンド結合シミュレーションを行ったところ、リガンドの結合に伴って構造が変化し、結晶構造と同様の構造を再現することに成功した。また、状態遷移に関する解析から、リガンドが結合することによりタンパク質が閉じた構造で安定となることが見出された。

第5章　総括

本研究では、タンパク質とリガンドの結合過程に対する一般的な知見を得ることを目的として、様々な性質を持つ多数のタンパク質-リガンドの系を対象にMARTINI粗視化力場を用いた分子動力学シミュレーションを行い、リガンド結合過程の比較解析を行った。さらに、MARTINI粗視化シミュレーションの適用の幅を広げることを目的とし、構造変化を扱うことができるようなシミュレーション手法の拡張にも取り組んだ。

リガンド結合過程の比較解析においては、タンパク質-リガンドペアをリガンドおよび基質結合ポケットの性質に基づいて分類し、異なる性質を持つ多数の例に対して解析を行った。リガンド結合シミュレーションによる解析の結果、リガンドが結晶構造と同様の基質結合ポケットに比較的安定に結合すルことが明らかとなった。また、結合速度や結合親和性のパラメータに関する解析より、MARTINI粗視化力場の枠組みの中で実際の性質を再現することが可能であることを示した。したがって、MARTINI粗視化力場を用いた解析により妥当性の高い結果が得られるものと考えられる。しかし一方で、十分な結合、解離が確認されず、結合親和性の評価ができなかった例なども存在した。このような例は結合、解離の時間スケールが長くなると考えられるため、効率的なサンプリング手法の開発が必要となる。現在、解離過程のシミュレーションにおいて、結合位置にエネルギー的なペナルティを与えることによるサンプリングの効率化を検討している。

さらに、結合経路に関する解析から、タンパク質-リガンド結合過程において特異的な経路を通過して結合することが見出され、タンパク質表面における溝などの形状や、疎水性などの物理化学的性質がリガンドの結合過程に関係していることが示唆された。こうした解析をさらに進めて一般化し、結合経路に関する汎用性の高い知見を得ることができれば、リガンドの結合過程の予測などにもつなげることができると期待される。今回の解析からは、タンパク質の基質結合ポケットの広さによるリガンドの結合過程における大きな違いは見出されなかったが、リガンドの結合過程においては水とリガンドの交換を伴うことから、ポケットの広さが水とリガンドの交換メカニズムや結合、解離の時間スケールなどに影響する可能性が考えられ、今後の研究課題となる。

また、MARTINI粗視化力場とdual-basin potentialを組み合わせた拡張を行うことにより、タンパク質の大きな構造変化を再現することに成功した。さらに、リガンド結合シミュレーションに適用することで、タンパク質の開閉を伴うリガンド結合を再現すること

ができ、構造変化を伴うリガンド結合過程の解析への有効性が示唆された。この結果から、構造変化を伴うタンパク質-リガンドの系についてもリガンド結合過程のシミュレーションが可能になり、さらに一般的な解析を行うことができると期待される。

付録A　化合物のMARTINI粗視化力場パラメータの決定

A.1　力場決定の手順

MARTINI 力場においてはタンパク質のほか、糖や脂質などのいくつかの一般的な生体分子に関してパラメータが整備されている。しかしながら、パラメータが作成されていない分子に関してはパラメータの作成を手動で行う必要がある。

新たにパラメータを作成する場合、実験データや全原子シミュレーションの結果を再現するように決定する。本研究では、MARTINI の論文 [16] に記されている手順に従い、最初に全原子シミュレーションとの比較により粗視化粒子の割り当て (マッピング) および結合長や結合角などの構造に関するポテンシャルを決定し、その後水-オクタノール分配係数 (LogP) を実験値や全原子シミュレーションなどの結果を再現するように粗視化粒子の種類を決定した。ただし、電荷を持つリガンドについては、論文で示されている標準的な対応付けに基づいて決定することとした。以下に本研究でのパラメータ決定の手順を示した。

結合パラメータの決定：

1. 全原子力場のパラメータが存在していない場合、量子化学計算により構造最適化、電荷の計算を行い、全原子のパラメータを決定する
2. 水中での化合物単体の全原子シミュレーションを行う
3. 粗視化粒子のマッピングを行う結合パラメータを設定する
4. 水中での化合物単体の粗視化シミュレーションを行う
5. 結合距離の分布などに関して全原子の結果との比較を行い、一致していれば決定し、一致していなければ 2 から再度行う

粗視化粒子種の決定 (電荷を持たないリガンド)：

1. 実験値ががわかっていない場合、全原子シミュレーションまたは予測プログラムによる LogP の予測を行う
2. 粗視化粒子種を設定する
3. 粗視化シミュレーションにより LogP を計算する
4. LogP について実験値または 1 による推定値と粗視化シミュレーションの結果との比較を行い、一致していれば決定し、一致していなければ 2 から再度行う

A.2 結合パラメータの比較

結合パラメータは、水中の化合物のシミュレーションを全原子力場および粗視化力場それぞれで行い、シミュレーション中における分子の形状 (結合距離、結合角) を比較することで決定した。本研究では先行研究に倣い、MARTINI 粗視化パラメータを決定する際に比較に用いる力場に OPLS 力場 [67] を用いた。化合物の全原子力場パラメータが存在しない場合、はじめに全原子パラメータの作成を行った。まず、目的化合物の構造を PDBechem より取得し、Openbabel[26] を用いて PH = 7 の条件で水素を負荷した。この構造を初期構造として HF/6-31G*レベルでの量子計算による構造最適化をプログラムパッケージ gaussian09[68] を用いて行った。その後、antechamber[69] および ACPYPE[70] を用いて RESP 電荷の計算、OPLS 力場パラメータの割り当てを行った。全原子力場および粗視化力場それぞれにおけるシミュレーション条件を表 A.1 に示した。

表 A.1: 水中における化合物のシミュレーションの条件。AA は全原子シミュレーション、CG は粗視化シミュレーションを表す

	system		
Simulation	compound	SPC water	simulation time / ns
AA	1	1500	200
CG	1	490	200

全原子シミュレーションにおいては水分子に Single point charge (SPC) モデル [71] を用いた。それぞれのシミュレーションは、310 K、1 bar の定温定圧条件下で行った。

これらのシミュレーション結果から、全原子モデルと粗視化モデルの結合距離、結合角の分布の比較を行った。まず、全原子シミュレーション結果から、マッピングに従って化合物の各粗視化粒子に対応する原子のグループの重心位置を計算し、仮想的な粗視化モデ

ルのトラジェクトリを作成した。次に、このように作成した全原子のトラジェクトリと、粗視化シミュレーションにより得られたトラジェクトリについて、粗視化粒子における結合長や結合角等の分布を計算し、比較を行った。

本研究においてパラメータを作成したリガンドについて、最終的に決定した粗視化分子と全原子分子の比較を図 A.1～A.5 に示した。

図 A.1: Citrate anion 分子のダイナミクスの比較 (結合長、結合角)。

図 A.2: Guanosine 分子のダイナミクスの比較 (結合長)。

図 A.3: Guanosine 分子のダイナミクスの比較 (結合角、二面角)。

図 A.4: 2 ’-deoxycytidine 分子のダイナミクスの比較 (結合角、二面角)。

図 A.5: Ethyl 3,5-dimethyl-1H-pyrazole-4-carboxylate 分子のダイナミクスの比較 (結合角、二面角)。

A.3　水-オクタノール分配係数の比較

水-オクタノール分配係数 LogP は、水およびオクタノールに対する溶媒和自由エネルギーを ΔG_w、ΔG_{oct} として、

$$\mathrm{logP} = \log\left(\exp\left(\frac{\Delta G_w - \Delta G_{oct}}{RT}\right)\right) = \frac{\Delta G_w - \Delta G_{oct}}{2.303RT} \tag{A.3.1}$$

と求められる。本研究では、シミュレーションにより水およびオクタノールに対する化合物の溶媒和自由エネルギーを求め、そこから式に従って LogP を計算することとした。溶媒和自由エネルギーの算出は、化合物と周りの溶媒の相互作用が存在しない状態 ($\lambda = 0$: 真空中と同様) から相互作用が完全に存在する状態 ($\lambda = 1$) の間の遷移を考え、この間のエネルギー差を求めることにより行う。エネルギー差を計算するにあたって、$\lambda = 0 \sim 1$ の間を 0.05 刻みで変化させ、それぞれの中間状態についてシミュレーションを行った。その後、各段階のエネルギー差を Benett acceptance ratio (BAR) 法 [72] により計算し、その和を求めることによって最終的な値を得た。

粗視化力場について LogP を計算し、実験データを再現するように原子種の割り当てを行った。実験データが存在しない場合には、全原子についても同様の計算を行うことによ

りLogPを算出し、粗視化シミュレーションと全原子シミュレーションの結果を比較した。

全原子および粗視化モデルによるシミュレーションの系およびシミュレーション時間はそれぞれ表A.2のように設定した。オクタノールの系については水分子を飽和濃度分だけ含めた。各シミュレーションにおける温度、圧力はいずれも310 K、1 barとした。本研究においてパラメータを決定した電荷を持たないリガンドについて、最終的に決定した粗視化分子と全原子分子のLogPの比較を表A.3に示した。

表A.2: シミュレーション条件

Simulation		system			simulation time
		compound	water	octanol	
AA	octanol	1	66	199	6 ns×20
	water	1	1400		6 ns×20
CG	octanol	1	43	519	10 ns×20
	water	1	1000		10 ns×20

表A.3: 水-オクタノール分配係数の比較。CGおよびAAはそれぞれ粗視化シミュレーション、全原子シミュレーションから推定した値を表す。Exp.は実験値を表し、AlogPSはソフトウェアAlogPS 3.0による推定値を表す。

リガンド	LogP			
	CG	Exp.	AA	AlogPS
Guanosine	-2.18	-2.461[73]		
2 ’-deoxycytidine	-1.59		-1.06	-1.9
Ethyl 3,5-dimethyl-1H-pyrazole-4-carboxylate	1.58		2.89	1.1

付録B　結合、解離速度定数の推定

B.1　ベイズ推定による速度定数の推定

速度定数の解析においては、ベイズ推定により速度定数の確率分布を求め、その平均値を推定値として用いた。

リガンド結合がポアソン分布に従って生じると仮定する。結合の速度定数を k_{on}、系全体のリガンド濃度を c_{lig} とすると、時間 T の間にリガンドが n_{bind} 回だけ結合する確率は次のように表せる。

$$p(n_{\mathrm{bind}}|k_{\mathrm{on}},T) = \frac{(k_{\mathrm{on}}c_{\mathrm{lig}}T)^{n_{\mathrm{bind}}}}{n_{\mathrm{bind}}!}e^{-k_{\mathrm{on}}c_{\mathrm{lig}}T} \tag{B.1.1}$$

いま、k_{on} の事前確率分布を $p(k_{\mathrm{on}})$ とする。これに対しシミュレーションを行い、シミュレーション i において観測された結合回数を n_i、結合にかかった時間の総和を T_i とし、観測を行った後の事後確率分布 $p(k_{\mathrm{on}}|\{n_i\},\{T_i\})$ を求める。ベイズの定理 $P(X|Y) \propto P(Y|X)P(X)$ において、尤度関数 $P(Y|X)$ を $p(\{n_i\}|k_{\mathrm{on}},\{T_i\})$、とすることで、事後分布 $p(k_{\mathrm{on}}|\{n_i\},\{T_i\})$ は次のように表すことができる。

$$p(k_{\mathrm{on}}|\{n_i\},\{T_i\}) \propto p(\{n_i\}|k_{\mathrm{on}},\{T_i\})p(k_{\mathrm{on}}) \tag{B.1.2}$$

シミュレーションを全 N 回実行した後の事後分布 $p(k_{\mathrm{on}}|\{n_1,n_2,\ldots,n_N\},\{T_1,T_2,\ldots,T_N\})$ を考えると、同様の計算を繰り返すことで

$$p(k_{\mathrm{on}}|\{n_1,n_2,\ldots,n_N\},\{T_1,T_2,\ldots,T_N\}) \propto \prod_{i=1}^{N} p(\{n_i\}|k_{\mathrm{on}},\{T_i\})p(k_{\mathrm{on}}) \tag{B.1.3}$$

のように求められる。速度定数に関する事前情報は一切無く、シミュレーションを行う前の事前分布 $p(k_{\mathrm{on}})$ を一様な分布とすれば、シミュレーション後の事後確率分布は最終的に次のように表現できる。

$$p(k_{\mathrm{on}}|\{n_1,n_2,\ldots,n_N\},\{T_1,T_2,\ldots,T_N\}) = \frac{\prod_{i=1}^{N} p(\{n_i\}|k_{\mathrm{on}},\{T_i\})}{\int_0^{\infty}\prod_{i=1}^{N} p(\{n_i\}|k_{\mathrm{on}},\{T_i\})dk_{\mathrm{on}}} \tag{B.1.4}$$

したがって、この事後分布における k_{on} の平均値 $\langle k_{\mathrm{on}} \rangle$ は

$$
\begin{aligned}
\langle k_{\mathrm{on}} \rangle &= \int_0^\infty k_{\mathrm{on}} p(k_{\mathrm{on}} | \{n_1, n_2, \ldots, n_N\}, \{T_1, T_2, \ldots, T_N\}) dk_{\mathrm{on}} \\
&= \frac{\int_0^\infty k_{\mathrm{on}} \prod_{i=1}^N p(\{n_i\} | k_{\mathrm{on}}, \{T_i\}) dk_{\mathrm{on}}}{\int_0^\infty \prod_{i=1}^N p(\{n_i\} | k_{\mathrm{on}}, \{T_i\}) dk_{\mathrm{on}}} \qquad (\because \text{式} B.1.4) \\
&= \frac{\int_0^\infty k_{\mathrm{on}} \prod_{i=1}^N \frac{(k_{\mathrm{on}} c_{\mathrm{lig}} T_i)^{n_i}}{n_i!} e^{-k_{\mathrm{on}} c_{\mathrm{lig}} T_i} dk_{\mathrm{on}}}{\int_0^\infty \prod_{i=1}^N \frac{(k_{\mathrm{on}} c_{\mathrm{lig}} T_i)^{n_i}}{n_i!} e^{-k_{\mathrm{on}} c_{\mathrm{lig}} T_i} dk_{\mathrm{on}}} \qquad (\because \text{式} B.1.1) \\
&= \frac{\int_0^\infty k_{\mathrm{on}}^{(\sum n_i + 1)} e^{-k_{\mathrm{on}} c_{\mathrm{lig}} \sum T_i} dk_{\mathrm{on}}}{\int_0^\infty k_{\mathrm{on}}^{(\sum n_i)} e^{-k_{\mathrm{on}} c_{\mathrm{lig}} \sum T_i} dk_{\mathrm{on}}} \\
&= \frac{\sum n_i + 1}{c_{\mathrm{lig}} \sum T_i} \qquad \text{(B.1.5)}
\end{aligned}
$$

のように求められ、これを推定値とした。また、分散 σ_{on}^2 は

$$
\begin{aligned}
\sigma_{\mathrm{on}}^2 &= \langle k_{\mathrm{on}}^2 \rangle - \langle k_{\mathrm{on}} \rangle^2 \\
&= \int_0^\infty k_{on}^2 p(k_{\mathrm{on}} | \{n_1, n_2, \ldots, n_N\}, \{T_1, T_2, \ldots, T_N\}) dk_{on} - \langle k_{on} \rangle^2 \\
&= \frac{\int_0^\infty k_{\mathrm{on}}^{(\sum n_i + 2)} e^{-k_{\mathrm{on}} c_{\mathrm{lig}} \sum T_i} dk_{\mathrm{on}}}{\int_0^\infty k_{\mathrm{on}}^{(\sum n_i)} e^{-k_{\mathrm{on}} c_{\mathrm{lig}} \sum T_i} dk_{\mathrm{on}}} - \left(\frac{\sum n_i + 1}{c_{\mathrm{lig}} \sum T_i} \right)^2 \qquad (\because \text{式} B.1.1) \\
&= \frac{\sum n_i + 1}{(c_{\mathrm{lig}} \sum T_i)^2} \qquad \text{(B.1.6)}
\end{aligned}
$$

と求められる。

リガンドの解離についても同様にポアソン分布に従うとすると、解離の速度定数を k_{off} としたとき、時間 T の間にリガンドが n_{unbind} 回だけ解離する確率は

$$
p(n_{\mathrm{unbind}} | k_{\mathrm{off}}, T) = \frac{(k_{\mathrm{off}} T)^{n_{\mathrm{unbind}}}}{n_{\mathrm{unbind}}!} e^{-k_{\mathrm{off}} T} \qquad \text{(B.1.7)}
$$

と求められる。結合速度との違いは濃度への依存の有無のみであり、結合定数のときと同様の計算を行うことで平均値 $\langle k_{\mathrm{off}} \rangle$ および分散 σ_{off}^2 は

$$
\langle k_{\mathrm{off}} \rangle = \frac{\sum n_i + 1}{\sum T_i} \qquad \text{(B.1.8)}
$$

$$\sigma_{\mathrm{off}}^2 = \frac{\sum n_i + 1}{(\sum T_i)^2} \tag{B.1.9}$$

と推定できる。

B.2　結合、解離の定義による速度定数の依存性

本研究では、シミュレーションにおけるリガンドの結合および解離を、2つのパラメータ r_{unbound} および r_{bound} を用いて定義した。これらの値を変化すると、観測される結合、解離の回数などが変化することになり、速度定数の推定結果が大きく変わることが考えられる。たとえば、r_{unbound} と r_{bound} の値の差を小さくすれば、シミュレーション中におけるリガンドの揺らぎの影響を受けやすくなるため結合、解離回数は大きく増加し、k_{on}、k_{off} の値は安定しなくなる。そのため、r_{unbound} および r_{bound} を適切に設定する必要がある。ここでは、r_{unbound}、r_{bound} に対する k_{on}、k_{off} の依存性を調べ、r_{unbound} および r_{bound} の変化に対して計算結果が安定するような値を探索した。

図 B.1 にそれぞれの系について r_{unbound}、r_{bound} を変化させたときの k_{on}、k_{off} の結果を示す。いずれの系においても r_{bound} が 2∼3 Å、r_{unbound} が 10 Å以上の範囲で値の変化が緩やかとなった。これらの結果を考慮し、本研究では r_{bound} および r_{unbound} をそれぞれ 3 Å、10 Å とした。

また、リガンドの結合状態からの RMSD を考慮した場合の判定についても同様の解析を行った (図 B.2)。いずれの系においても $\mathrm{RMSD}_{\mathrm{bound}}$ が 1∼2 Å、$\mathrm{RMSD}_{\mathrm{unbound}}$ が 14 Å以上の範囲で値の変化が緩やかとなったことから、本研究では $\mathrm{RMSD}_{\mathrm{bound}}$ および $\mathrm{RMSD}_{\mathrm{unbound}}$ をそれぞれ 2 Å、15 Å と設定した。

図 B.1: r_{bound} および $r_{unbound}$ に対する速度定数の推定値の依存性 (DAHPS∼HAP)。

図 B.1: r_{bound} および r_{unbound} に対する速度定数の推定値の依存性 (PNP∼PAL)。

図 B.1: r_{bound} および $r_{unbound}$ に対する速度定数の推定値の依存性 (LinB～PDE4D)。

図 B.2: RMSD$_{bound}$ および RMSD$_{unbound}$ に対する速度定数の推定値の依存性 (AST～Lev)。

図 B.2: $RMSD_{bound}$ および $RMSD_{unbound}$ に対する速度定数の推定値の依存性 (PAL～PDE4D)。

参考文献

[1] Whittle, P. J.; Blundell, T. L. Protein structure–based drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 349–75.

[2] Shailza, S.; Balwant, K. M.; Durlabh, K. S. Molecular drug targets and structure based drug design: A holistic approach. *Bioinformation* **2006**, *1*, 314–320.

[3] Swinney, D. C. Biochemical mechanisms of drug action: what does it take for success? *Nat. Rev. Drug Discov.* **2004**, *3*, 801–8.

[4] Swinney, D. C. Applications of binding kinetics to drug discovery therapeutic responses. *Pharmaceut. Med.* **2008**, *22*, 23–34.

[5] Zhang, R.; Monsma, F. Binding kinetics and mechanism of action: toward the discovery and development of better and best in class drugs. *Expert Opin. Drug. Discov.* **2010**, 1–7.

[6] Wu, C.; Biancalana, M.; Koide, S.; Shea, J.-E. Binding modes of thioflavin-T to the single-layer beta-sheet of the peptide self-assembly mimics. *J. Mol. Biol.* **2009**, *394*, 627–33.

[7] Brannigan, G.; LeBard, D. N.; Hénin, J.; Eckenhoff, R. G.; Klein, M. L. Multiple binding sites for the general anesthetic isoflurane identified in the nicotinic acetylcholine receptor transmembrane domain. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 14122–14127.

[8] Buch, I.; Giorgino, T.; De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184–10189.

[9] Vanni, S.; Neri, M.; Tavernelli, I.; Rothlisberger, U. Predicting novel binding modes of agonists to β adrenergic receptors using all-atom molecular dynamics simulations. *PLoS computational biology* **2011**, *7*, e1001053.

[10] Venturoli, M.; Maddalenasperotto, M.; Kranenburg, M.; Smit, B. Mesoscopic models of biological membranes. *Phys. Rep.* **2006**, *437*, 1–54.

[11] Bennun, S. V.; Hoopes, M. I.; Xing, C.; Faller, R. Coarse-grained modeling of lipids. *Chem. Phys. Lipids* **2009**, *159*, 59–66.

[12] Tozzini, V. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–50.

[13] Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A. Coarse-grained (multiscale) simulations in studies of biophysical and chemical systems. *Annu. Rev. Phys. Chem.* **2011**, *62*, 41–64.

[14] Li, D.; Liu, M. S.; Ji, B.; Hwang, K.; Huang, Y. Coarse-grained molecular dynamics of ligands binding into protein: The case of HIV-1 protease inhibitors. *J. Chem. Phys.* **2009**, *130*, 215102.

[15] Marrink, S. J.; De Vries, A. H.; Mark, A. E. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* **2004**, *108*, 750–760.

[16] Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–24.

[17] Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.

[18] Hu, L.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. Binding MOAD (Mother of All Databases). *Proteins* **2005**, *60*, 333–340.

[19] Benson, M. L.; Smith, R. D.; Khazanov, N. A.; Dimcheff, B.; Beaver, J.; Dresslar, P.; Nerothin, J.; Carlson, H. A. Binding MOAD, a high-quality protein-ligand database. *Nucleic Acids Research* **2008**, *36*, .

[20] Petrek, M.; Otyepka, M.; Banás, P.; Kosinová, P.; Koca, J.; Damborský, J. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC bioinformatics* **2006**, *7*, 316.

[21] Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.* **2012**, *8*, .

[22] Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132.

[23] `http://www.epa.gov/opptintr/exposure/pubs/episuite.htm`.

[24] Tetko, I. V.; Poda, G. I. Application of ALOGPS 2.1 to predict log D distribution coefficient for pfizer proprietary compounds. *J. Med. Chem.* **2004**, *47*, 5601–5604.

[25] Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A. K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V. V.; Tanchuk, V. Y.; Todeschini, R.; Varnek, A.; Marcou, G.; Ertl, P.; Potemkin, V.; Grishina, M.; Gasteiger, J.; Schwab, C.; Baskin, I. I.; Palyulin, V. A.; Radchenko, E. V.; Welsh, W. J.; Kholodovych, V.; Chekmarev, D.; Cherkasov, A.; Aires-De-Sousa, J.; Zhang, Q. Y.; Bender, A.; Nigsch, F.; Patiny, L.; Williams, A.; Tkachenko, V.; Tetko, I. V. Online chemical modeling environment (OCHEM): Web platform for data storage, model development and publishing of chemical information. *J. Comput. Aided Mol. Des.* **2011**, *25*, 533–554.

[26] O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open chemical toolbox. *J. Cheminformatics* **2011**, *3*, .

[27] `http://toolkit.tuebingen.mpg.de/blastclust`.

[28] König, V.; Pfeil, A.; Braus, G. H.; Schneider, T. R. Substrate and metal complexes of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from Saccharomyces cerevisiae provide new insights into the catalytic mechanism. *J. Mol. Biol.* **2004**, *337*, 675–690.

[29] Jäger, J.; Moser, M.; Sauder, U.; Jansonius, J. N. Crystal structures of Escherichia coli aspartate aminotransferase in two conformations. Comparison of an unliganded open and two liganded closed forms. *J. Mol. Biol.* **1994**, *239*, 285–305.

[30] Delarue, M.; Duclert-Savatier, N.; Miclet, E.; Haouz, A.; Giganti, D.; Ouazzani, J.; Lopez, P.; Nilges, M.; Stoven, V. Three Dimensional Structure and Implications for the Catalytic Mechanism of 6-Phosphogluconolactonase from Trypanosoma brucei. *J. Mol. Biol.* **2007**, *366*, 868–881.

[31] Duclert-Savatier, N.; Poggi, L.; Miclet, E.; Lopes, P.; Ouazzani, J.; Chevalier, N.; Nilges, M.; Delarue, M.; Stoven, V. Insights Into the Enzymatic Mechanism of 6-Phosphogluconolactonase from Trypanosoma brucei Using Structural Data and Molecular Dynamics Simulation. *J. Mol. Biol.* **2009**, *388*, 1009–1021.

[32] Singh, H.; Felts, R. L.; Schuermann, J. P.; Reilly, T. J.; Tanner, J. J. Crystal Structures of the Histidine Acid Phosphatase from Francisella tularensis Provide Insight into Substrate Recognition. *J. Mol. Biol.* **2009**, *394*, 893–904.

[33] Tahirov, T. H.; Inagaki, E.; Ohshima, N.; Kitao, T.; Kuroishi, C.; Ukita, Y.; Takio, K.; Kobayashi, M.; Kuramitsu, S.; Yokoyama, S.; Miyano, M. Crystal structure of purine nucleoside phosphorylase from Thermus thermophilus. *J. Mol. Biol.* **2004**, *337*, 1149–1160.

[34] Armstrong, S. R.; Cook, W. J.; Short, S. A.; Ealick, S. E. Crystal structures of nucleoside 2-deoxyribosyltransferase in native and ligand-bound forms reveal architecture of the active site. *Structure* **1996**, *4*, 97–107.

[35] Meng, G.; Fütterer, K. Structural framework of fructosyl transfer in Bacillus subtilis levansucrase. *Nat. Struct. Biol.* **2003**, *10*, 935–41.

[36] `http://www.rcsb.org/pdb/explore/explore.do?structureId=1s1a`.

[37] Buts, L.; Garcia-Pino, A.; Imberty, A.; Amiot, N.; Boons, G. J.; Beeckmans, S.; Versées, W.; Wyns, L.; Loris, R. Structural basis for the recognition of complex-type biantennary oligosaccharides by Pterocarpus angolensis lectin. *FEBS Journal* **2006**, *273*, 2407–2420.

[38] Streltsov, V. a.; Prokop, Z.; Damborský, J.; Nagata, Y.; Oakley, A.; Wilce, M. C. J. Haloalkane dehalogenase LinB from Sphingomonas paucimobilis UT26: X-ray crystallographic studies of dehalogenation of brominated substrates. *Biochemistry* **2003**, *42*, 10104–12.

[39] Duggleby, H. J.; Tolley, S. P.; Hill, C. P.; Dodson, E. J.; Dodson, G.; Moody, P. C. Penicillin acylase has a single-amino-acid catalytic centre. *Nature* **1995**, *373*, 264–268.

[40] Done, S. H.; Brannigan, J. A.; Moody, P. C.; Hubbard, R. E. Ligand-induced conformational change in penicillin acylase. *J. Mol. Biol.* **1998**, *284*, 463–475.

[41] Kobe, A.; Caaveiro, J. M. M.; Tashiro, S.; Kajihara, D.; Kikkawa, M.; Mitani, T.; Tsumoto, K. Incorporation of rapid thermodynamic data in fragment-based drug discovery. *J. Med. Chem.* **2013**, *56*, 2155–2159.

[42] Kraut, D. A.; Sigala, P. A.; Pybus, B.; Liu, C. W.; Ringe, D.; Petsko, G. A.; Herschlag, D. Testing electrostatic complementarity in enzyme catalysis: Hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLoS Biology* **2006**, *4*, 501–519.

[43] Nankervis, J. L.; Feil, S. C.; Hancock, N. C.; Zheng, Z.; Ng, H. L.; Morton, C. J.; Holien, J. K.; Ho, P. W. M.; Frazzetto, M. M.; Jennings, I. G.; Manallack, D. T.; John Martin, T.; Thompson, P. E.; Parker, M. W. Thiophene inhibitors of PDE4: Crystal structures show a second binding mode at the catalytic domain of PDE4D2. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 7089–7093.

[44] Card, G. L.; Blasdel, L.; England, B. P.; Zhang, C.; Suzuki, Y.; Gillette, S.; Fong, D.; Ibrahim, P. N.; Artis, D. R.; Bollag, G.; Milburn, M. V.; Kim, S.-H.; Schlessinger, J.; Zhang, K. Y. J. A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design. *Nature biotechnology* **2005**, *23*, 201–207.

[45] Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free 2005.

[46] Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.* **2009**, *5*, 2531–2543.

[47] LóPez, C. A.; Rzepiela, A. J.; De Vries, A. H.; Dijkhuizen, L.; Hünenberger, P. H.; Marrink, S. J. Martini coarse-grained force field: extension to carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210.

[48] Hansch, C.; Leo, A.; Hoekman, D. H. *Exploring QSAR - Hydrophobic, Electronic, and Steric Constants.*; American Chemical Society, 1995.

[49] Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

[50] Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; Van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem.* **1999**, *38*, 236–240.

[51] Onuffer, J. J.; Kirsch, J. F. Redesign of the substrate specificity of Escherichia coli aspartate aminotransferase to that of Escherichia coli tyrosine aminotransferase by homology modeling and site-directed mutagenesis. *Protein Sci.* **1995**, *4*, 1750–1757.

[52] Chambert, R.; Treboul, G.; Dedonder, R. Kinetic studies of levansucrase of Bacillus subtilis. *Eur. J. Biochem.* **1974**, *41*, 285–300.

[53] Lommerse, J. P. M.; Stone, A. J.; Taylor, R.; Allen, F. H. The nature and geometry of intermolecular interactions between halogens and oxygen or nitrogen. *J. Am. Chem. Soc.* **1996**, *118*, 3108–3116.

[54] Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16789–16794.

[55] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water 1983.

[56] Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.

[57] Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A generalizable biomolecular force field. carbohydrates. *J. Comp. Chem.* **2008**, *29*, 622–655.

[58] Kawabata, T.; Go, N. Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites. *Proteins* **2007**, *68*, 516–29.

[59] Kawabata, T. Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins* **2010**, *78*, 1195–211.

[60] Collins, M. D.; Hummer, G.; Quillin, M. L.; Matthews, B. W.; Gruner, S. M. Cooperative water filling of a nonpolar protein cavity observed by high-pressure crystallography and simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16668–71.

[61] Qvist, J.; Davidovic, M.; Hamelberg, D.; Halle, B. A dry ligand-binding cavity in a solvated protein. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6296–301.

[62] Young, T.; Hua, L.; Huang, X.; Abel, R.; Friesner, R.; Berne, B. J. Dewetting transitions in protein cavities. *Proteins* **2010**, *78*, 1856–69.

[63] Okazaki, K.-i.; Koga, N.; Takada, S.; Onuchic, J. N.; Wolynes, P. G. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11844–11849.

[64] Maragakis, P.; Karplus, M. Large amplitude conformational change in proteins explored with a plastic network model: Adenylate kinase. *J. Mol. Biol.* **2005**, *352*, 807–822.

[65] Cuneo, M. J.; Beese, L. S.; Hellinga, H. W. Ligand-induced conformational changes in a thermophilic ribose-binding protein. *BMC struct. biol.* **2008**, *8*, 50.

[66] Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Hübner, C. G.; Kern, D. Intrinsic motions along an enzymatic reaction trajectory. *Nature* **2007**, *450*, 838–844.

[67] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

[68] et al Frisch, M. Gaussian 09, Revision A.02 2009.

[69] Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

[70] Sousa, A. W.; Vranken, W. F. Open Access ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes* **2012**, 1–8.

[71] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models For Water In Relation To Protein Hydration. *Intermolecular Forces* **1981**, *31*, 331–338.

[72] Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268.

[73] `http://kaken.nii.ac.jp/pdf/2011/seika/C-19/32659/21590047seika.pdf`.