

博士論文

分子動力学シミュレーションおよび機械学習を用いた
抗原抗体界面に存在する塩橋に関する研究

岡島 亮

目次

1 序論	5
1.1 抗体	5
1.2 塩橋	7
1.3 本研究の目的	11
参考文献	18
2 HEL-HyHEL-10 複合体	23
2.1 序論	23
2.2 HEL-HyHEL-10 複合体系における塩橋の振る舞い	27
2.2.1 方法	27
2.2.2 結果と考察	29
2.3 2 残基モデル系	31
2.3.1 方法	31
2.3.2 結果と考察	34
2.4 少数アミノ酸モデル系	38
2.4.1 方法	38
2.4.2 結果と考察	42
2.5 Ala 変異が塩橋に及ぼす影響	46
2.5.1 方法	46
2.5.2 結果と考察	48
2.6 結論	55
参考文献	58

3 機械学習に基づく Na-Cl イオン対の挙動解析	63
3.1 序論	63
3.2 方法	67
3.2.1 MD シミュレーションの条件	67
3.2.2 Na-Cl 対の遷移状態の位置の特定	67
3.2.3 初期構造の作成	68
3.2.4 機械学習に用いたデータ	69
3.2.5 機械学習手法	70
3.2.6 モデルの検証	84
3.2.7 機械学習の詳細	85
3.3 結果と考察	85
3.3.1 k 近傍法	85
3.3.2 ランダムフォレスト	89
3.3.3 サポートベクターマシン	93
3.3.4 ニューラルネットワーク	95
3.3.5 各特徴量の重要性	107
3.4 結論	117
参考文献	118
4 機械学習に基づく抗原-抗体界面の塩橋の挙動解析	120
4.1 序論	120
4.2 方法	120

4.2.1 塩橋形成/解離プロセスにおける $K97^Y N_{\zeta}$ - $D32^H C_{\gamma}$ 間距離の鞍点の位置 の特定	120
4.2.2 初期構造・速度の作成	121
4.2.3 機械学習に用いたデータ	122
4.2.4 機械学習手法	123
4.3 結果と考察	123
4.4 結論	126
参考文献	127
5 総括	128
謝辞	132

1 序論

1.1 抗体

抗体は免疫機構において、非常に重要な役割を果たしている^{1,2}。抗体が抗原（異物）と特異的に結合することで免疫系が活性化され、抗原が体内から排除される（図 1.1）。この抗原と特異的に結合できる性質により、抗体は医薬品としても利用もされている³⁻¹⁰。例えば、ホジキンリンパ腫や全身性未分化リンパ腫の治療薬として、**brentuximab vedotin** という抗体医薬品が使用されている³。また、HER2 陽性乳がんに対して、**ado-trastuzumab emtansine (T-DM1)** という抗体医薬品が使用されている^{4,5}。これらの抗体医薬品は、腫瘍細胞に特異的に発現している抗原を標的として結合する抗体と、細胞殺傷効果を持つ薬剤、抗体と薬剤を繋ぐリンカーからなる。抗体部分が標的とする抗原に対して選択的に結合し、抗原を発現している腫瘍細胞のみに薬剤を送り届ける。この抗体部分が抗原と高い特異性を持って結合する働きにより、正常細胞に薬剤の影響を与えない、副作用の低い薬となる。

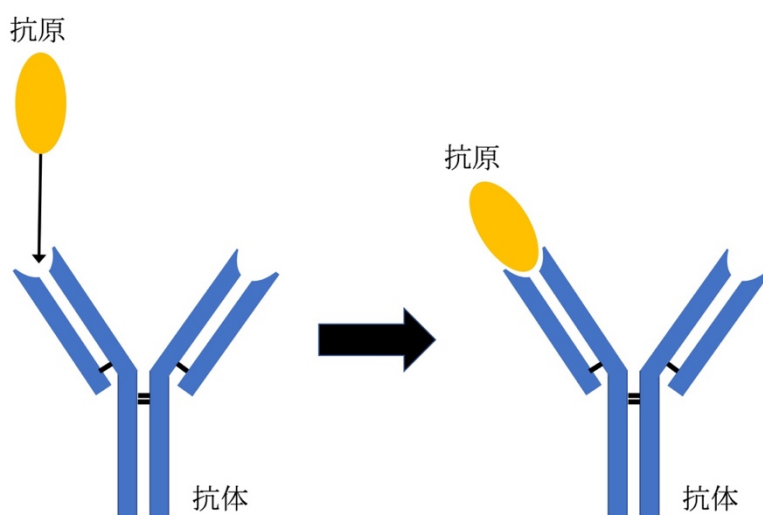


図 1.1：抗原と抗体との結合

抗体を医薬品に応用するためには、病気の治療に対して高い効果を出すために、抗原との高い親和性が必須である。しかしながら、医薬品として使用するには抗体の親和性が足りず、抗体に変異を導入することで抗原との親和性を向上させる必要がある場合がある^{11,12}。このような変異体を設計することは、タンパク質工学分野において挑戦的な課題の一つである。親和性を向上させる変異を設計するには、抗体の抗原認識のメカニズムを分子レベルで理解することが重要であると考えられる。

抗体の抗原認識メカニズムを知るために、多くの X 線結晶構造解析が実施されている¹³⁻¹⁷。例えば、肝臓ガンに特異的に発現しているヒト ROBO1 の 3 番目のフィブロネクチンドメイン(Fn3)とマウスモノクローナル抗体 B2212A の Fab フラグメントの複合体の結晶構造(PDB ID: 3WIH)が報告されている¹⁸。この結晶構造では、B2212A 軽鎖の 50 番目のチロシン (Y50^L) 残基と ROBO1 Fn3 の 68 番目のフェニルアラニン(F68)残基が水素結合を形成していた。ここで、上付き文字 L は抗体の軽鎖を示す。また、B2212A Fab 単体の結晶構造と比較すると、Y50^L の側鎖の配向が大きく変化しており、この Y50^L は抗原認識に重要な役割があるのではないかと考えられた。そこで、実際に Y50^L をアラニンに置換したところ、ROBO1 Fn3 と B2212A Fab の結合の親和性が 10 倍程度弱くなることが示された。そのため、Y50^L-F68 の水素結合が ROBO1 Fn3 と B2212A Fab の結合に重要な要素であることが分かった。

また、血管内皮増殖因子(VEGF)と抗 VEGF 抗体の Fab フラグメントとの複合体についても結晶構造が取られている¹⁹。この結晶構造より、抗体の各アミノ酸残基が抗原-抗体結合界面中で占めている面積が見積もられる。さらに、相互

性決定領域(CDR)のアミノ酸残基それぞれに対してアラニン置換を施し、抗原-抗体間結合の親和性の変化が調べられている。その結果、抗原-抗体間の親和性が 150 倍以上低下するアラニン置換変異は、N31^HA、Y32^HA、W50^HA、N52^HA、Y95^HA、S100^HA、W100^HA、W96^LA のみであった。ここで、上付き文字 H は抗体の重鎖、L は軽鎖を示す。これらのアミノ酸残基は抗原-抗体結合界面に存在しており、抗原-抗体の結合に大きく寄与していることが示された。

1.2 塩橋

塩橋とは、タンパク質内において正電荷を持つアミノ酸残基と負電荷を持つアミノ酸残基の間に形成されるイオン対である。塩橋は、様々なタンパク質機能の鍵となっていることが知られている。例えば、細胞表面において細胞外のアドレナリンを検知するβ2 アドレナリン受容体(β2AR)は心臓血管や肺の生理学に重要な役割を果たすタンパク質である²⁰。β2AR が不活性状態のとき、R131 は E268 と塩橋を形成している²¹。アドレナリンが結合すると、この塩橋が開裂しβ2AR が活性化状態となり、細胞内部にシグナルを伝達する。

また、超好熱菌のタンパク質は熱安定性が非常に高いが、この熱安定性の要因の 1 つはタンパク質表面にある塩橋の存在と考えられている。実際、超好熱古細菌 *T. celer* のリボソームタンパク質 L30e には、E6-R92 と E62-K46 の塩橋が形成されている²²。これらの塩橋がアラニン変異によってなくなると、タンパク質の熱安定性が低くなることが確かめられている。

また、抗原-抗体間の相互作用においても、塩橋が重要な役割を果たす場合もある。例えば、インフルエンザウイルス A/Memphis/31/98 (H3N2) のノイラミニダーゼ (NA) とその抗体 Mem5 の界面には、E199^{NA} と R52^H の塩橋が形成されて

いる²³。ここで、上付き文字 H は抗体の重鎖を示す。この E199^{NA} を Ala に置換することによって、NA と Mem5 の親和性が 22 倍くらい低下すること実験的に明らかにされており、E199^{NA} と R52^H の塩橋が NA と Mem5 の抗原-抗体間結合に対するホットスポットであることが示された。

また、インターフェロンガンマ受容体 IFN γ R とその抗体 A6 の複合体の結晶構造より、K52-D54^H および K52-D56^H の塩橋が形成されていることが示されている²⁴。ここで、上付き文字 H は抗体の重鎖を示す。この IFN γ R の K52 にアラニン変異を行うと、IFN γ R と A6 が結合しなくなることが分かっている。そのため、K52 が IFN γ R と A6 の相互作用のホットスポットであり、K52-D54^H および K52-D56^H の塩橋が重要な役割を果たしていると考えられる。

一方で、金属塩など低分子イオン対の研究から、イオン結合の物理的理解も進んできた。一般的に、水中におけるイオン対形成はイオンの脱水和によるエントロピーの増大により促進されると考えられている。イオン対形成に伴うクーロン相互作用の利得は、脱水和によるエンタルピー損失により大きく相殺される^{25,26}。例えば、常温(298 K)の MgSO₄ イオン対の形成に対してエンタルピー変化は+5.8 kJ/mol、エントロピー成分の変化(-T Δ S)は-18.4 kJ/mol であり、合計の自由エネルギー変化は-12.6 kJ/mol となる。また、LiB(OH)₄ イオン対の場合、エンタルピー変化は+1.0 kJ/mol、エントロピー成分の変化は-7.0 kJ/mol であり、合計の自由エネルギー変化は-6.0 kJ/mol となる。さらに、GaCl²⁺ イオン対の場合、エンタルピー変化は+4.8 kJ/mol、エントロピー成分の変化は-10.7 kJ/mol となり、合計の自由エネルギー変化は-5.9 kJ/mol となる。しかし、タリウム-ハロゲン化物イオン対に対しては、イオン対形成に大きなエンタルピー利得も伴っていることが示されている。例えば、TlCl の場合、エンタルピー変化は-26.3 kJ/mol、

エントロピー成分の変化は+5.0 kJ/mol で、自由エネルギー変化は-21.3 kJ/mol だった。

タンパク質表面において塩橋が形成される場合、塩橋を形成するアミノ酸の種類に依らず、塩橋形成による自由エネルギーへの寄与は-5 kJ/mol 程度と考えられている^{25, 27-28}。例えば、超好熱菌タンパク質 Ssh10b には E36-K68 の塩橋がタンパク質表面に確認されている²⁹。この塩橋形成による Ssh10b のフォールディング自由エネルギーへの寄与を二重変異サイクル(DMC)分析により確認したところ、298 K で-6.0 kJ/mol であることが示されている。これのエンタルピー成分への寄与についても DMC 分析により調べると、-4.8 kJ/mol で主要な成分であった。ここで、DMC 分析とは、タンパク質の分子内または分子間に働く 2 残基間の相互作用の強度を測定する方法であり、相互作用している 2 残基についての二重変異と二つの単一変異による自由エネルギー変異の差を計算する手法である。

タンパク質内部に埋もれている塩橋の場合、塩橋形成に伴う自由エネルギーの寄与はさらに小さくなると考えられている³⁰。これは塩橋がタンパク質内部に埋もれてさらなる脱水和することに由来している。例えば、超好熱菌タンパク質 Ssh10b の内部にある E54-R57 の塩橋に対する DMC 分析により、塩橋形成のフォールディング自由エネルギーへの寄与は 298 K で-2.1 kJ/mol であることが分かっている²⁹。

また、近年の研究によると、塩橋の安定性が親和性に寄与する可能性があることが示されている。ROBO1 とその抗体(B5209B)の複合体において、抗原-抗体結合エンタルピーの向上が抗原-抗体界面周囲の塩橋の安定化に起因するということが示されている³¹。B5209B と ROBO1 の結合界面には、E97^H-R477^L

と D31^H-R477^I の二つの塩橋が形成されていることが結晶構造より見えている (図 1.3)。ここで、上付き文字 H は抗体の重鎖、I は抗原を示す。さらに、B5209B に P103^HA の変異を導入すると、ROBO1 との親和性が 30 倍近く向上することが示されている。MD シミュレーションを行い、E97^H-R477^I と D31^H-R477^I の側鎖の酸素-窒素間の最短距離の分布をそれぞれ調べると、ともに P103^HA の変異体の方が WT よりも頻繁に E97^H-R477^I と D31^H-R477^I の塩橋が形成されていることが示されている。したがって、E97^H-R477^I と D31^H-R477^I の塩橋の安定性が ROBO1-B5209B の結合親和性に影響を与えており、抗体の親和性向上を狙う上で鍵になるものだと考えられる。

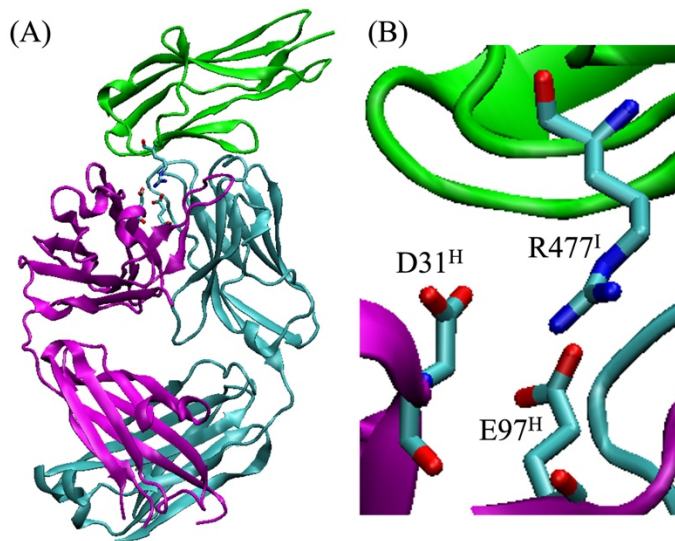


図 1.3 : ROBO1-B5209B Fab の複合体の結晶構造 (A) 全体図 (B) E97^H-R477^I と D31^H-R477^I の塩橋周りの拡大図。ROBO1 は緑で、B5209B の重鎖はマゼンタ、軽鎖はシアンで表している。また、D31^H、E97^H、R477^I の残基を棒で表している。結晶構造中に含まれない水素原子は表示していない。

1.3 本研究の目的

上記の ROBO1 抗体 B5209B の研究によると、抗体の親和性向上の戦略として、抗原-抗体複合体の形を大きく変えることなく、抗原-抗体間の塩橋を安定化させることが考えられる³¹。実際に、P103^HA 変異はこのことを実現しており、結合自由エネルギーを -8.4 kJ/mol 向上させている。ROBO1 抗体 B5209B 野生型 (WT)の結合自由エネルギー -37.8 kJ/mol に比べると小さく感じられる数値であるが、これによって親和性は 30 倍に改良することができる。また、抗原-抗体界面の中心部に位置し、すでに抗原-抗体間相互作用に重要な役割を持っている部位の改変は、複合体構造に大きく影響を与え、親和性を大きく低下させる可能性が非常に高い。一方で、今回のような抗原-抗体界面の辺縁部に位置する塩橋に注目し、大きな構造変化を引き起こさない変異を考える戦略はこのようリスクを回避することができるという利点がある。

しかし、現時点において、野生型の情報のみから、どのような環境因子によって塩橋の安定化が決まっているのかを特定する手法や理論は確立されていない。野生型の抗原-抗体複合体の MD シミュレーションから、どのように塩橋の安定化に関わる因子を探し出すのかを考察する。これまでの先行研究では、野生型と変異型の違いを見出すことに MD シミュレーションが使われてきたが、本研究では野生型の MD シミュレーション情報のみから情報を抜き出す事を試みる。つまり、野生型の MD シミュレーションのみから後述する少数アミノ酸モデルや機械学習を用いて情報を取り出す戦略を採用する。変異体を 1 つ 1 つ計算する負荷がないことが、変異体設計の観点においての本研究の戦略の利点になっている。

この論文の研究対象として、卵白リゾチーム(HEL)とその抗体である

HyHEL-10 を選ぶこととする(図 1.2)³²。その理由は、ROBO1-B5209B 系の場合と同じく、抗原-抗体界面の辺縁部に 2 つの塩橋(D32^H-K97^Y と D99^H-K97^Y)を持ち、また ROBO1-B5209B 系よりも早くから研究されて多くの実験結果がすでに存在するからである。(ここで、上付き文字 H は抗体の重鎖、Y は抗原を示す。) 例えば、D32^H を Ala 置換すると結合の親和性は 4 倍弱くなり、D32^H を Ala 置換すると結合の親和性は 2 倍弱くなることが実験により示されている³²。さらに、D32^HA/D99^HA の二重変異を加えると抗原-抗体間の親和性は 10 倍弱まることが分かった³²。個々の変異 (D32^HA と D99^HA) の効果の単純な掛け合わせでなく、さらに大きな親和性低下となっている。

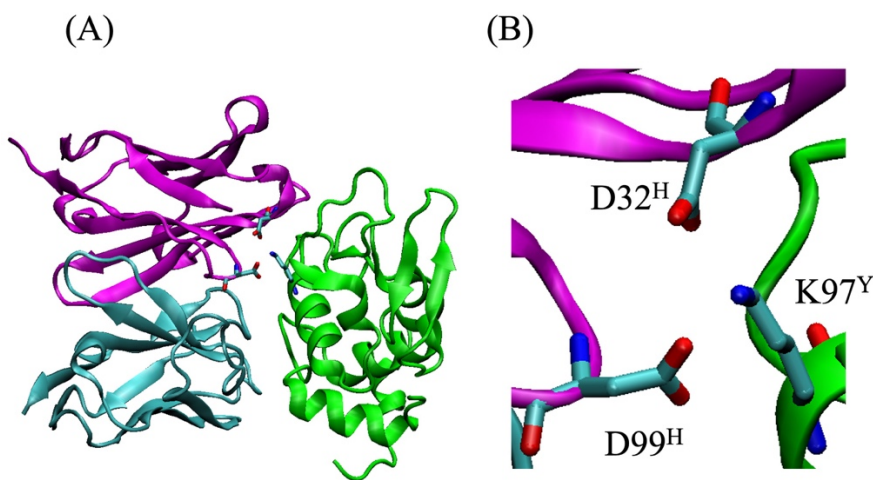


図 1.2 : HEL-HyHEL-10 複合体の結晶構造 (A) 全体図 (B) D32^H-K97^Y と D99^H-K97^Y の塩橋周りの拡大図。HEL は緑で、HyHEL-10 の重鎖はマゼンタ、軽鎖はシアンで表している。また、D32^H、D99^H、K97^Y の残基を棒で表している。結晶構造中に含まれない水素原子は表示していない。

本研究では、まず、この塩橋の安定性がどのような要素によって決定されているのかを明らかにすることを目的にする。抗原-抗体界面のどのような因子が塩橋の安定性に寄与するのかを同定することができれば、抗体の親和性を向上させるアミノ酸変異を設計するヒントとすることができると考えた。本論文の第 2 章では、HEL-HyHEL-10 複合体界面に存在する塩橋に着目する。HEL-HyHEL-10 系には、X 線結晶構造解析だけでなく、塩橋アミノ酸のアラニン置換実験などもされており、実験的な情報が多く知られている³²⁻³⁴。塩橋の安定性を見積もるには、分子動力学 (MD) シミュレーションを用いて正電荷を持つアミノ酸と負電荷を持つアミノ酸が、どれくらい塩橋を形成し解離していくのかを評価する。ここから、塩橋の安定性に対して重要な因子を調べるために、抗原-抗体間の界面の環境を部分的に再現した少数アミノ酸モデル系による解析を提案する(図 1.4)。この少数アミノ酸モデル系に、HEL-HyHEL-10 複合体に含まれる因子を一つ一つ追加していき、どれくらい HEL-HyHEL-10 複合体における塩橋の安定性が再現されるかを評価する。この手続きによって、塩橋の安定性に寄与している重要な環境因子を見出だしていく。

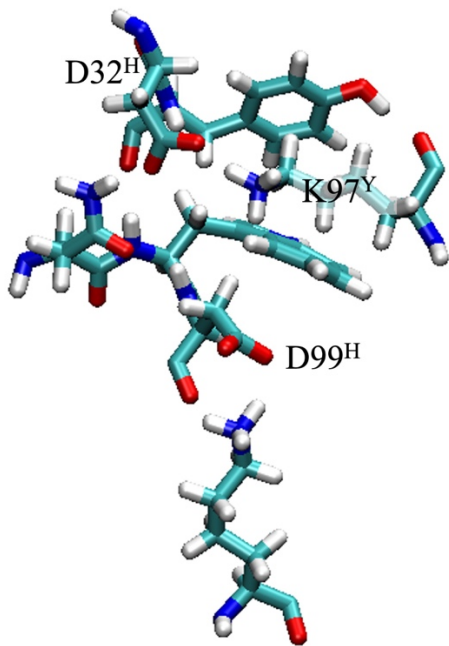


図 1.4: 少数アミノ酸モデルの系の例。この図では、塩橋を形成する D32^H、D99^H、K97^Y 以外に、4 つのアミノ酸残基を環境因子として含めたモデル系を示している。

また、塩橋の安定性は、塩橋アミノ酸ペアが会合状態である割合と解離状態である割合の比率として評価するが、この量は、遷移状態を通して解離側から会合側に流れ込む反応量と会合側から解離側に流れ込む反応量のバランスで決まっている(図 1.5)。この観点から、塩橋の安定化を考える上で遷移状態付近でのダイナミクスが周囲の環境からどのような影響を受けているのかという視点も重要であり、抗体の親和性改良のためのヒントになると思われる。よって、本研究の後半において、塩橋の形成/崩壊ダイナミクスに影響を与えている環境因子を MD 計算から明らかにすることを目的とする。しかしながら、反応を制御する因子を系統的に取り出す手法は一般的に確立されていない。

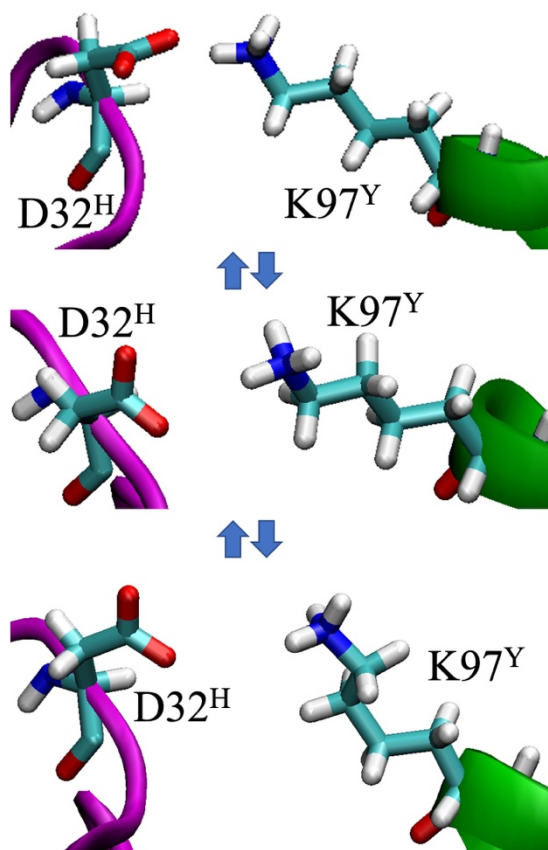


図 1.5 : HEL-HyHEL-10 界面の D32^H-K97^Y 塩橋。上段: 塩橋形成している状態。中段: 遷移状態。下段: 解離状態。

ここでは具体的に、機械学習をダイナミクスの解析に応用することを検討する。機械学習は与えられたデータを使用してデータのルールやパターンを学習し、未知のデータに対して予測を行う方法である³⁵。学習はアルゴリズムに沿って自動で行われるため、先入観なしに学習を行うことができるという利点がある。機械学習を用いたデータ分析の例としてよく使われるものの一つに、画像から手書き数字を認識するという問題がある³⁵⁻³⁶。手書き文字は人によって様々な書き方があり、人力で識別ルールを構築することは非常に困難である。このような例に対して機械学習を用いると、良い精度で分類を行うことができる。すな

わち、十分に学習した機械学習モデルは複雑に関係する多数の特徴量から重要な要素を認識できていると考えられる。こうした性質を持つ機械学習法を用いることで、複雑な抗原-抗体界面において塩橋の安定性を支配する因子を系統的に取り出すことができるのではないかと考えられる。また、野生型の塩橋安定性など平衡熱力学量に注目すると学習するデータは少なく機械学習と相性が悪い。したがって、一旦、遷移状態付近のダイナミクスに注目することで、野生型の情報のみから巨大なデータベースを構築することを可能としている点がここでのアプローチの利点となっている。

第3章においては、方法論を提案しその有効性を議論する。ここでは、テスト系として、塩橋よりも単純な Na-Cl イオン対を使用する。Na-Cl イオン対の会合/解離は Na-Cl 間の距離だけで記述できる単純な反応ではない。例えば、Geissler、Dellago および Chandler は³⁷、水中の Na-Cl イオン対について、単一の初期構造にランダムな初期運動量を割り振った MD シミュレーションを複数行った。このシミュレーションの結果、Na-Cl 対が会合状態に向かう割合と解離状態に向かう割合が半々であれば、その初期構造は遷移状態であると判定した。この解析を様々な初期構造に対して行い、Na-Cl 間距離に対する遷移状態と判定された初期構造の分布を調べた。その結果、Na-Cl イオン対の会合/解離反応の遷移状態の構造は、Na-Cl 間距離に対して幅広く分布することが明らかになった。つまり、Na-Cl イオン対の会合/解離反応の反応座標には、Na-Cl 間距離以外の要素（すなわち、水分子の座標）が含まれていることが示された。こうした環境因子を機械学習によって取り出せるかどうかで機械学習の性能を判断することが可能であると考えられる。

第4章においては、第3章で提案した機械学習を用いた解析手法を HEL-

HyHEL-10 界面の D32^H-K97^Y の塩橋の形成/崩壊ダイナミクスの解析に適用する。D32^H-K97^Y の塩橋が遷移状態にある構造を初期構造にとって、そこから塩橋が形成するか崩壊するかを MD シミュレーションで調べてデータベース化し、それを機械学習により解析をおこなう。これにより、遷移状態付近での塩橋の形成崩壊ダイナミクスに影響する環境因子を明らかにすることを目的とする。この結果は抗体の親和性改良のためのヒントになると思われる。この塩橋の役割は実験的に検証されている^{32,31}。D32^H を Ala に置換することで HEL-HyHEL-10 間の親和性が 4 倍弱くなる。また本論文の 2 章においては MD シミュレーションを用いて、D32^H-K97^Y の塩橋の安定性と、安定性に関わる環境因子を調べる解析も実施する。よって、本章で得られた結論を別の角度から比較し議論を深めることが可能になると考えられる。

本論文の構成は、本章を第 1 章として、全 5 章からなる。第 2 章では、HEL-HyHEL-10 複合体界面に存在する D32^H-K97^Y および D99^H-K97^Y の二つの塩橋の安定性について述べる。さらに、少数アミノ酸モデルを用いた解析により、これらの塩橋の安定性に重要な影響を与えている環境因子を明らかにする。第 3 章では、Na-Cl イオン対をテスト系に選び、機械学習を用いた解析手法を開発する。続く第 4 章では、第 3 章で開発した機械学習を用いた解析手法を、HEL-HyHEL-10 複合体界面の D32^H-K97^Y の塩橋の形成/崩壊ダイナミクスの解析に適用する。これによって、塩橋の形成/崩壊ダイナミクスに影響を及ぼす環境因子を特定する。最後に第 5 章において、本論文のまとめを行う。

参考文献

1. Klein, J.; Nikolaidis, N., The descent of the antibody-based immune system by gradual evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (1), 169-174.
2. Řihová, B.; Kopečková, P.; Strohalm, J.; Rossmann, P.; Větvička, V.; Kopeček, J., Antibody-directed affinity therapy applied to the immune system: in vivo effectiveness and limited toxicity of daunomycin conjugated to HPMA copolymers and targeting antibody. *Clin. Immunol. Immunopathol.* **1988**, *46* (1), 100-114.
3. Senter, P. D.; Sievers, E. L., The discovery and development of brentuximab vedotin for use in relapsed Hodgkin lymphoma and systemic anaplastic large cell lymphoma. *Nat. Biotechnol.* **2012**, *30* (7), 631-637.
4. Phillips, G. D. L.; Li, G.; Dugger, D. L.; Crocker, L. M.; Parsons, K. L.; Mai, E.; Blättler, W. A.; Lambert, J. M.; Chari, R. V.; Lutz, R. J., Targeting HER2-positive breast cancer with trastuzumab-DM1, an antibody–cytotoxic drug conjugate. *Cancer Res.* **2008**, *68* (22), 9280-9290.
5. Lambert, J. M.; Chari, R. V., Ado-trastuzumab Emtansine (T-DM1): an antibody–drug conjugate (ADC) for HER2-positive breast cancer. *J. Med. Chem.* **2014**, *57* (16), 6949-6964.
6. Junutula, J. R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D. D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S. P.; Dennis, M. S., Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol.* **2008**, *26* (8), 925.
7. Han, T. H.; Zhao, B., Absorption, distribution, metabolism, and excretion considerations for the development of antibody-drug conjugates. *Drug Metab. Dispos.* **2014**, *42* (11), 1914-1920.
8. Hamblett, K. J.; Senter, P. D.; Chace, D. F.; Sun, M. M.; Lenox, J.; Cervený, C. G.; Kissler, K. M.; Bernhardt, S. X.; Kopcha, A. K.; Zabinski, R. F., Effects of drug loading on the antitumor activity of a monoclonal antibody drug conjugate. *Clin. Cancer Res.* **2004**, *10* (20), 7063-7070.

9. Donaghy, H., Effects of antibody, drug and linker on the preclinical and clinical toxicities of antibody-drug conjugates. *MAbs* **2016**, *8* (4), 659-671.
10. Adams, G. P.; Weiner, L. M., Monoclonal antibody therapy of cancer. *Nat. Biotechnol.* **2005**, *23* (9), 1147-1157.
11. Lakayan, D.; Haselberg, R.; Gahoual, R.; Somsen, G. W.; Kool, J., Affinity profiling of monoclonal antibody and antibody-drug-conjugate preparations by coupled liquid chromatography-surface plasmon resonance biosensing. *Anal. Bioanal. Chem.* **2018**, *410* (30), 7837-7848.
12. Yamada, K.; Shikida, N.; Shimbo, K.; Ito, Y.; Khedri, Z.; Matsuda, Y.; Mendelsohn, B. A., AJICAP: affinity peptide mediated regiodivergent functionalization of native antibodies. *Angew. Chem. Int. Ed.* **2019**, *131* (17), 5648-5653.
13. Yokota, A.; Tsumoto, K.; Shiroishi, M.; Nakanishi, T.; Kondo, H.; Kumagai, I., Contribution of asparagine residues to the stabilization of a proteinaceous antigen-antibody complex, HyHEL-10-hen egg white lysozyme. *J. Biol. Chem.* **2010**, *285* (10), 7686-7696.
14. Yokota, A.; Tsumoto, K.; Shiroishi, M.; Kondo, H.; Kumagai, I., The Role of Hydrogen Bonding via Interfacial Water Molecules in Antigen-Antibody Complexation THE HyHEL-10-HEL INTERACTION. *J. Biol. Chem.* **2003**, *278* (7), 5410-5418.
15. Murase, T.; Zheng, R. B.; Joe, M.; Bai, Y.; Marcus, S. L.; Lowary, T. L.; Ng, K. K., Structural insights into antibody recognition of mycobacterial polysaccharides. *J. Mol. Biol.* **2009**, *392* (2), 381-392.
16. McLellan, J. S.; Chen, M.; Leung, S.; Graepel, K. W.; Du, X.; Yang, Y.; Zhou, T.; Baxa, U.; Yasuda, E.; Beaumont, T., Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **2013**, *340* (6136), 1113-1117.
17. Garces, F.; Lee, J. H.; de Val, N.; de la Pena, A. T.; Kong, L.; Puchades, C.; Hua, Y.; Stanfield, R. L.; Burton, D. R.; Moore, J. P., Affinity maturation of a potent family of HIV antibodies is primarily

focused on accommodating or avoiding glycans. *Immunity* **2015**, *43* (6), 1053-1063.

18. Nakayama, T.; Mizohata, E.; Yamashita, T.; Nagatoishi, S.; Nakakido, M.; Iwanari, H.; Mochizuki, Y.; Kado, Y.; Yokota, Y.; Satoh, R., Structural features of interfacial tyrosine residue in ROBO1 fibronectin domain-antibody complex: Crystallographic, thermodynamic, and molecular dynamic analyses. *Protein Sci.* **2015**, *24* (3), 328-340.

19. Muller, Y. A.; Chen, Y.; Christinger, H. W.; Li, B.; Cunningham, B. C.; Lowman, H. B.; de Vos, A. M., VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 Å resolution and mutational analysis of the interface. *Structure* **1998**, *6* (9), 1153-1167.

20. Hanson, M. A.; Cherezov, V.; Griffith, M. T.; Roth, C. B.; Jaakola, V.-P.; Chien, E. Y.; Velasquez, J.; Kuhn, P.; Stevens, R. C., A specific cholesterol binding site is established by the 2.8 Å structure of the human β 2-adrenergic receptor. *Structure* **2008**, *16* (6), 897-905.

21. Ballesteros, J. A.; Jensen, A. D.; Liapakis, G.; Rasmussen, S. G.; Shi, L.; Gether, U.; Javitch, J. A., MECHANISMS OF SIGNAL TRANSDUCTION-Activation of the β 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J. Biol. Chem.* **2001**, *276* (31), 29171-29177.

22. Chan, C.-H.; Yu, T.-H.; Wong, K.-B., Stabilizing salt-bridge enhances protein thermostability by reducing the heat capacity change of unfolding. *PLoS One* **2011**, *6* (6), e21624.

23. Lee, J. T.; Air, G. M., Interaction between a 1998 human influenza virus N2 neuraminidase and monoclonal antibody Mem5. *Virology* **2006**, *345* (2), 424-433.

24. Yoshida, K.; Kuroda, D.; Kiyoshi, M.; Nakakido, M.; Nagatoishi, S.; Soga, S.; Shirai, H.; Tsumoto, K., Exploring designability of electrostatic complementarity at an antigen-antibody interface directed by mutagenesis, biophysical analysis, and molecular dynamics simulations. *Sci. Rep.* **2019**, *9* (1), 4482.

25. Anslyn, E. V.; Dougherty, D. A., *Modern physical organic chemistry*. University science books: 2006.
26. Marcus, Y.; Hefter, G., Ion pairing. *Chem. Rev.* **2006**, *106* (11), 4585-4621.
27. Meyer, E. A.; Castellano, R. K.; Diederich, F., Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem. Int. Ed.* **2003**, *42* (11), 1210-1250.
28. Biedermann, F.; Schneider, H.-J. r., Experimental binding energies in supramolecular complexes. *Chem. Rev.* **2016**, *116* (9), 5216-5300.
29. Ge, M.; Xia, X.-Y.; Pan, X.-M., Salt bridges in the hyperthermophilic protein Ssh10b are resilient to temperature increases. *J. Biol. Chem.* **2008**, *283* (46), 31690-31696.
30. Bosshard, H. R.; Marti, D. N.; Jelesarov, I., Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *J. Mol. Recognit.* **2004**, *17* (1), 1-16.
31. Yamashita, T.; Mizohata, E.; Nagatoishi, S.; Watanabe, T.; Nakakido, M.; Iwanari, H.; Mochizuki, Y.; Nakayama, T.; Kado, Y.; Yokota, Y., Affinity Improvement of a Cancer-Targeted Antibody through Alanine-Induced Adjustment of Antigen-Antibody Interface. *Structure* **2019**, *27* (3), 519-527. e5.
32. Shiroishi, M.; Yokota, A.; Tsumoto, K.; Kondo, H.; Nishimiya, Y.; Horii, K.; Matsushima, M.; Ogasahara, K.; Yutani, K.; Kumagai, I., Structural Evidence for Entropic Contribution of Salt Bridge Formation to a Protein Antigen-Antibody Interaction THE CASE OF HEN LYSOZYME-HyHEL-10 Fv COMPLEX. *J. Biol. Chem.* **2001**, *276* (25), 23042-23050.
33. Pons, J.; Rajpal, A.; Kirsch, J. F., Energetic analysis of an antigen/antibody interface: alanine scanning mutagenesis and double mutant cycles on the HyHEL-10/lysozyme interaction. *Protein Sci.* **1999**, *8* (5), 958-968.
34. Tsumoto, K.; Ogasahara, K.; Ueda, Y.; Watanabe, K.; Yutani, K.; Kumagai, I., Role of Salt Bridge

Formation in Antigen-Antibody Interaction ENTROPIC CONTRIBUTION TO THE COMPLEX BETWEEN HEN EGG WHITE LYSOZYME AND ITS MONOCLONAL ANTIBODY HyHEL10. *J. Biol. Chem.* **1996**, 271 (51), 32612-32616.

35. Bishop, C. M., パターン認識と機械学習. ベイズ理論による統計的予測 **2008**.

36. 中井悦司, *IT エンジニアのための機械学習理論入門: 機械学習のしくみを学びデータサイエンスの本質を理解する*. 技術評論社: 2015.

37. Geissler, P. L.; Dellago, C.; Chandler, D., Kinetic pathways of ion pair dissociation in water. *J. Phys. Chem. B* **1999**, 103 (18), 3706-3710.

2 HEL-HyHEL-10 複合体

2.1 序論

免疫機構において抗体は抗原と結合し、その働きを阻害する、抗体と結合した抗原を体内から排除するなどの役割をする¹⁻²。この働きにおいて重要な抗体の性質が、抗原に対して特異的・強固に結合するという性質である。そのため、抗体の抗原に対する親和性と特異性は免疫機構において重要であり、抗体医薬品としても利用されている³⁻¹⁰。そのため、特定の抗原を対象として高い親和性を持つ抗体を設計するためにも、免疫系における抗体の抗原認識のメカニズムを理解することは重要である。抗体の抗原認識のメカニズムを理解するとは、抗体が抗原とどのような相互作用を持って結合するのかを知ることであり、抗体の性質を分子レベルで知ることである。これを実現するために、様々な研究がなされている。例えば、X線結晶構造解析により、様々な抗体や抗原-抗体複合体の立体構造が得られている。このようにして得られた立体構造から、抗体と抗原がどのように相互作用しているのかを調べられている¹¹⁻¹⁵。これらの研究の結果、抗体と抗原の結合界面には、構造的・静電的に高い相補性があることが示されている。そのほかにも、抗体のアミノ酸残基にAla置換を行い、置換による抗原との親和性の変化を調べる方法もある¹⁶⁻¹⁸。置換による親和性の変化の大きさから、Ala置換を行なった残基が抗原-抗体間の結合にどれほど寄与しているのかを調べる。また、抗原と抗体の結合の際の熱量測定を行い、エンタルピー的な寄与なのかエントロピー的な寄与なのかを調べることも行われている¹⁹⁻²⁰。このような実験的な手法の他にも、分子動力学(MD)シミュレーションを用いた研究もある²¹⁻²⁵。この手法により、抗体の動的な性質などの実験では解析しにくい特徴を調べる。

抗原-抗体間の結合界面の中で、静電的に非常に重要な構造として塩橋がある。塩橋は正に荷電したアミノ酸残基と負に荷電したアミノ酸が静電的に引きあった構造である。塩橋は強い静電相互作用によって形成されるため、抗原-抗体間で形成された場合、抗原と抗体の結合親和性に強い影響を及ぼすと考えられる。実際、インフルエンザウイルス A/Memphis/31/98 (H3N2)のノイラミニダーゼ (NA)とその抗体 Mem5 は、E199^{NA} と R52^Hの残基が塩橋を形成しており、この部分がホットスポットの一つとなっている²⁶。更に、E199^{NA}に変異を加えると NA と Mem5 との親和性が低下し、これにより、Mem5 による NA の無力化はされなくなる²⁷⁻²⁸。また、他の例として、卵白リゾチーム(HEL)とその抗体 HyHEL-10 Fv との界面では D32^H-K97^Y と D99^H-K97^Yの二つの塩橋が確認されている²⁰。ここで、上付き文字 H、L、Y はそれぞれ抗体の重鎖、軽鎖、抗原を示す。熱力学的な研究によると、D32^HA および D99^HA の変異をそれぞれ加えて塩橋を形成できなくした場合、抗原-抗体間の結合親和性が著しく低下することが示されている²⁰。WT の結合自由エネルギーが-50.2 kJ/mol だったのに対し、D32^HA および D99^HA の変異体系における結合自由エネルギーはそれぞれ-46.4 および-48.2 kJ/mol となっている。その他にも、多数の抗原-抗体界面において塩橋の存在は確認されている。

また、実験だけでなく計算の研究によっても塩橋の安定性が抗原-抗体間の強い結合に対して重要な役割を果たしていることが示されている。ROBO1 とその抗体(B5209B)の MD シミュレーションによって、塩橋の安定化が結合エンタルピーの向上を導いているという報告がある²²。この研究では、塩橋を形成するアミノ酸残基への直接的な変異だけでなく、塩橋から離れた残基への Ala 置換でさえも、塩橋の安定性に影響を及ぼし、結合親和性を変化させる可能性がある

ということが示された。また、Yoshida らは²¹、インターフェロンガンマ受容体 (IFN γ R)とその抗体である A6 との界面に、電荷を入れ替えるように、いくつかの変異を加えることで、抗原-抗体の結合親和性と塩橋との関係性を調べた。電荷を入れ替えるというのは、R84^{Ag}E-E27^{VL}R のように、二つの残基を入れ替えるように変異を導入している。この研究によると、R84^{Ag}E-E27^{VL}R と K98^{Ag}D-D28^{VL}K の変異を加えても IFN γ R-A6 間の高い親和性を保持するが、K52^{Ag}D-D54^{VH}K/D56^{VH}K という変異を加えた場合には表面プラズモン共鳴 (SPR) 信号は観測されなかった。MD シミュレーションでは、E84^{Ag}-R27^{VL} と D98^{Ag}-K28^{VL} の塩橋は、それぞれの変異体で形成されていたが、一方で K52^{Ag}D-D54^{VH}K/D56^{VH}K の変異は複合体の構造に大きな変化をもたらし、結果として塩橋が形成されなくなった。また、一部の塩橋は X 線結晶構造においては安定に形成されていても、生体中などにおいては熱ゆらぎによって解離している事があるという事が、MD シミュレーションによって示されている²¹⁻²²。塩橋の重要性については広く認識されている。しかし、どのようにして塩橋が安定化、不安定化しているのかについては未だ不明瞭である。

卵白リゾチーム (HEL) とその抗体 HyHEL-10 については、実験・計算両面から幾つもの研究がなされている^{11-12, 17, 19-20, 23-24, 29}。HEL と HyHEL-10 の結合界面には 2 つの塩橋が形成されていることが結晶構造から見えている (図 2.1)。これらの塩橋の存在は、HEL と HyHEL-10 の結合親和性に対して大きな寄与をしている¹⁹⁻²⁰。

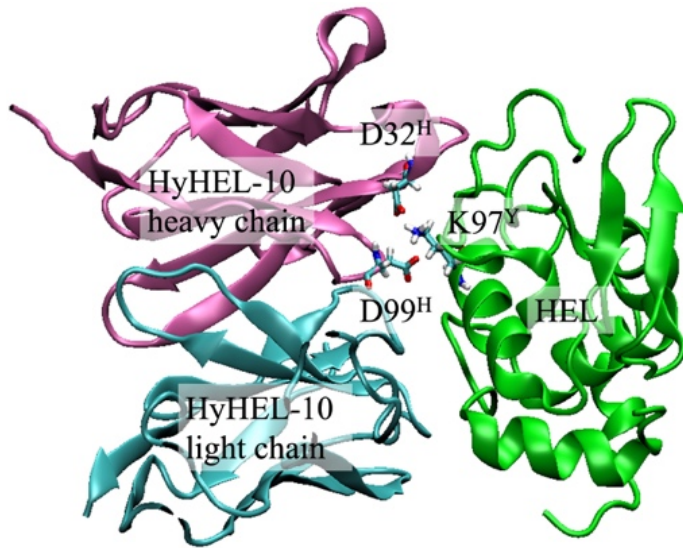


図 2.1 : HEL-HyHEL-10 複合体の図。緑で示しているのが HEL。赤で示しているのが HyHEL-10 の重鎖。水色で示しているのが HyHEL-10 の軽鎖。また、HEL-HyHEL-10 界面で塩橋を形成しているリシン残基とアスパラギン酸残基を棒で表した。

そこで、本研究では HEL-HyHEL-10 複合体の界面に形成されている二つの塩橋(D32^H-K97^Y、D99^H-K97^Y)に着目し、どのような要因によって複合体界面での塩橋の安定性が決められているのかを調べた。これらの塩橋の動的な性質を調べるために、我々は HEL-HyHEL-10 複合体の系の MD シミュレーションを行った。後述するが、この MD シミュレーション結果、水中では D32^H-K97^Y の塩橋は安定に形成されていたが、D99^H-K97^Y の塩橋はほとんど形成されていなかった。また、リシンとアスパラギン酸間の距離を調べるために Ca-Ca 間距離を計算すると興味深いことに、結晶構造中でも MD シミュレーション中でも D32^H-K97^Y の組み合わせの方が D99^H-K97^Y の組み合わせの方よりも長くなっていた。直感的にはリシンとアスパラギン酸がより接近している方が塩橋は形成されや

すそうに思えるのに、それに反するような結果となったため、塩橋の安定性を決めるメカニズムはあまり単純ではないという事が分かった。したがって本研究では、少数のアミノ酸のみを含み、抗原-抗体間の界面の構造を部分的に再現する、塩橋周りの環境を簡略化したモデルを構築し、このモデルを使って塩橋の安定性に影響を与える主要な要因を調査することを目的とした。さらに同様のアプローチによって、D32^HA と D99^HA の変異がそれぞれ塩橋にもたらす影響についても調べた。

2.2 HEL-HyHEL-10 複合体における塩橋の振る舞い

2.2.1 方法

我々は、HEL-HyHEL-10 複合体の水中での挙動を調べるため、MD シミュレーションを行った。MD シミュレーションを行うにあたって、および、エネルギーなどの種々の物理量の計算には *GROMACS 4.6.7* を用いた³⁰。MD シミュレーションを行うにあたって、HEL-HyHEL-10 Fv 複合体の結晶構造(PDB ID: 2DQJ)を初期構造として利用した。結晶構造においては、水素原子が不足していたため、不足していた水素原子の座標を追加した。MD シミュレーションにおいては周期境界条件を用い、単位セルは 12 面体セルとし、その中に 1 つの HEL-HyHEL-10 複合体と 18,464 個の水分子、55 の Na⁺イオン、60 の Cl⁻イオンを含めた。Na⁺イオン、および Cl⁻イオンの個数は、その濃度が生理食塩水濃度(0.15 M)となるように決めた。この系を、後述する他の系と比較する為に、HEL-HyHEL-10 複合体系または単純に、複合体系と記述する。MD シミュレーションは 298 K, 1 atm の *NpT* 条件下で行なった。この系は the low-memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS)を用いてエネルギー最小化を行なった。力場はタンパク質につ

いては AMBER 力場を改良した FUJI 力場を用い^{21, 31}、水については TIP3P 力場を用いた。長距離静電相互作用の計算には Particle-Mesh Ewald (PME)法を用い、カットオフ距離は 9 Å とした。温度と圧力を一定に保つ為に、Nose-Hoover 法および Berendsen 法を用いた。また、linear constraint solver (LINCS)を用いて、共有結合の結合長は一定に保たれるようにした。シミュレーションの時間間隔は 3 fs とし、構造のデータは 15 ps 間隔で保存した。各原子の初期速度は、298 K における Maxwell 分布に従ってランダムに設定し、初期速度の違いから 5 本のトラジェクトリーを作成した。MD シミュレーションは 300 ns 流した。ポテンシャルエネルギーと初期構造の HEL-HyHEL-10 複合体の骨格構造を基準とした、MD シミュレーション中の RMSD の変化を調べた (図 2.2)。この結果から、HEL-HyHEL-10 複合体系は 150 ns で十分に平衡化していると判断し、後半の 150 ns のデータを塩橋の解析のために使用した。

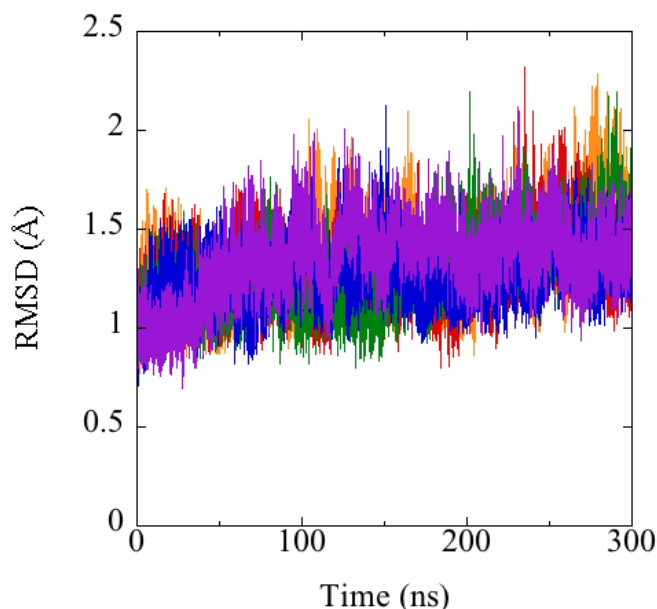


図 2.2: HEL-HyHEL-10 複合体の RMSD の時間変化。初期構造を基準として計算している。初期速度をランダムに与えて 5 つのシミュレーションを行っており、各色の線はそれぞれのシミュレーションでのデータである。

塩橋の解析を行うために、本研究ではアスパラギン酸側鎖の CO₂ とリシン側鎖の NH₃ 間の最短距離が 2.5 Å 未満となっていることを塩橋形成の条件とした。また、塩橋の安定性を測る尺度として、塩橋の形成割合を用いた。塩橋の形成割合は保存した構造データの中から、保存した全構造のうち、塩橋が形成している構造の割合として定義した。

塩橋の安定性をエネルギーの尺度から議論するために、我々は塩橋の形成割合から、塩橋形成時と解離時の自由エネルギー差を次のように定義した。

$$F = -k_B T \ln \frac{p}{1-p}$$

ここで、 p は塩橋の形成割合、 k_B はボルツマン定数、 T は温度である。自由エネルギー差と塩橋の形成割合は本質的に等価であり、自由エネルギー差の上昇・減少は形成割合の減少・上昇とそれぞれ直接的に相関し、それは塩橋の不安定化・安定化をそれぞれ示す。しかしながら、塩橋の安定性を自由エネルギー差の尺度から評価することは、塩橋の安定性と後述する塩橋形成による種々のエネルギー変化の関係を議論する上でより有用なため、こちらを用いた。

2.2.2 結果と考察

結晶構造において、HEL と HyHEL-10 の界面には二組の塩橋 D32^H-K97^Y と D99^H-K97^Y が存在した¹⁹⁻²⁰。しかし、MD シミュレーション中では D99^H-K97^Y の塩橋は非常に不安定で、シミュレーション中はほとんど解離していた。塩橋形成割合はわずか 0.02 だった。それに対して、D32^H-K97^Y の塩橋はシミュレーション中でも非常に安定で、塩橋形成割合は 0.90 だった (図 2.3)。

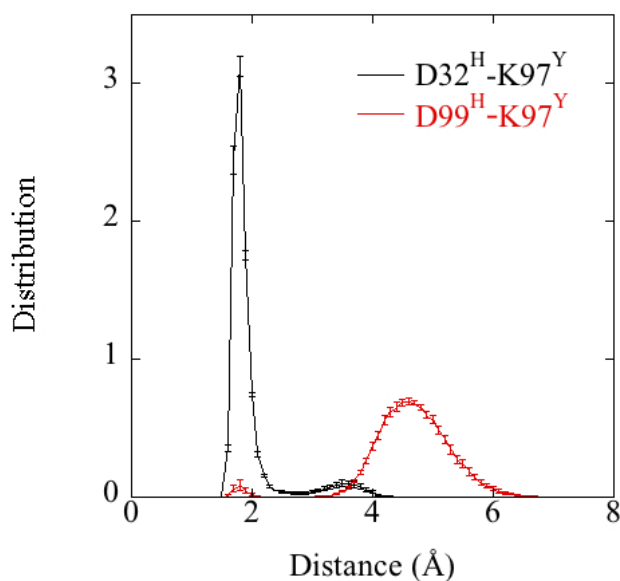


図 2.3 : HEL-HyHEL-10 複合体系において、シミュレーション中でのリシン側鎖の NH₃ とアスパラギン酸側鎖の CO₂ の最短距離の分布。黒線が D32^H-K97^Y についての分布で、赤線が D99^H-K97^Y についての分布である。

この二組の塩橋の形成割合の差に最も影響していそうな要素として、リシンとアスパラギン酸間の距離が考えられる。そこで、リシンとアスパラギン酸の Ca-Ca 間の距離に着目した。タンパク複合体の三次元的な主鎖構造は水中ではあまり大きく動くことができず、実質的に Ca-Ca 間距離は束縛されていると考えられる。我々は Ca-Ca 間距離が増加した場合、塩橋はより不安定化するのではないかと考えた。しかし実際は、MD シミュレーション中で塩橋形成割合が高かった D32^H-K97^Y の塩橋の Ca-Ca 間距離が約 10.2 Å であり、塩橋形成割合が低かった D99^H-K97^Y の塩橋の Ca-Ca 間距離が約 8.6 Å と、塩橋が安定な方が Ca-Ca 間距離が長くなるという結果になった (図 2.4)。塩橋の安定性を決定づけるのに、どのような要因が寄与しているのかを調べるため、我々は少数アミノ酸モデルと名付けた数残基のモデル系を構築した。

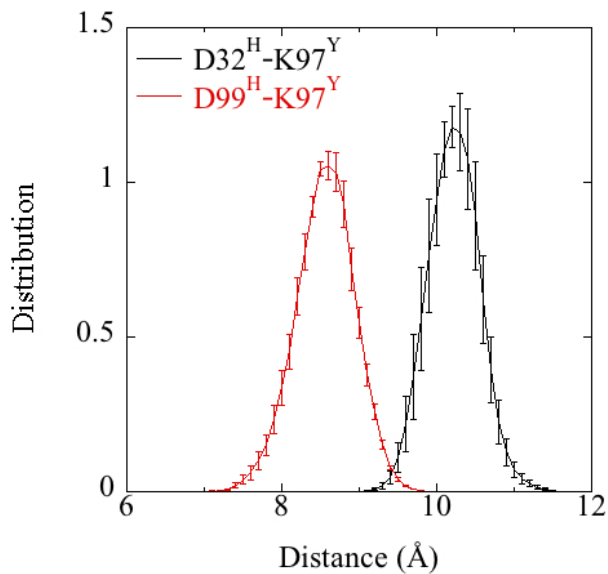


図 2.4 : HEL-HyHEL-10 複合体系において、シミュレーション中でのリシンとアスパラギン酸の Ca-Ca 間距離の分布。黒線が D32^H-K97^Y についての分布で、赤線が D99^H-K97^Y についての分布である。

2.3 2 残基モデル系

2.3.1 方法

より詳細に塩橋が安定化・不安定化するメカニズムを調べるために、少数アミノ酸モデルを導入した。少数アミノ酸モデルは HEL-HyHEL-10 界面の塩橋を作るアスパラギン酸、リシン残基および、周囲の数残基のみを含めることで、界面環境の一部のみを再現したモデル系である。第一に、アスパラギン酸とリシン残基のみの相互作用からなる塩橋安定性を調べるため、2 残基系を構築し、互いの位置関係の影響を調べるため、主鎖の原子の座標を束縛した。その後、周囲の界面環境の変化を調べるため、塩橋周囲のアミノ酸残基を順次含めたモデル系を構築した。そうして、HEL-HyHEL-10 複合体系における塩橋の安定性と同じ塩

橋安定性を保つモデル系を作ることで、どのアミノ酸残基が塩橋の安定性に対して重要だったのかを調べた。このような手法は建設的アプローチとして、経済傾向の分析からセル内でのドメイン形成まで、幅広い分野で複雑な系の挙動を理解するために用いられている³²⁻³⁵。

第一に、塩橋を形成するリシンとアスパラギン酸の相対的な配置のみで、塩橋の安定性がどの程度影響を受けるのかを調べるため、2残基モデルを構築した。2残基系の単位セルにはリシン、アスパラギン酸残基を1個ずつと水分子を1,555個、Na⁺イオンを4個にCl⁻イオンを4個含めた。アミノ酸主鎖の末端の電荷の効果を除外するため、アセチル(ACE)及び*N*-メチルアミノ(NME)基でアミノ酸の末端をキャップした。このキャップは、以下のすべてのモデル系に対して行った。最初に、塩橋安定性へのリシンとアスパラギン酸の距離のみの影響を調べるため、MDシミュレーション中、C α 原子のみを、ばね定数10 kJ/mol/Å²の調和振動子ポテンシャルで初期座標に束縛した。このモデルはRes2Dモデルと以下呼称する。束縛するC α -C α 間距離を5.0, 5.5, 6.0, ..., 11.5 Åと変化させ、比較することで、リシンとアスパラギン酸の距離の影響を調べた。また、初期構造はHEL-HyHEL-10複合体系のMDシミュレーションの最終構造から対応するアミノ酸残基(リシン、およびアスパラギン酸)を抜き出してきており、C α -C α 間距離が束縛する距離となるように各原子を平行移動させている。HEL-HyHEL-10複合体系のMDシミュレーションは初期速度を変えて5本行ったため、初期構造はそれぞれの系で5つずつ作っている。また、タンパク中ではアミノ酸残基の主鎖はあまり大きく動くことができず、結果、アミノ酸側鎖の伸びる向きなどの配向はあまり変わらない。そのため、次にC α -C α 間距離に加えてリシンとアスパラギン酸の配向(側鎖の向き)の影響を調べるため、主鎖の重原子の座標を束縛

した MD シミュレーションを行った。この系を、D32^H-K97^Y の塩橋に対応する系を Res2O32、D99^H-K97^Y の塩橋に対応する系を Res2O99 と呼ぶ。

塩橋の安定性 (塩橋形成時と解離時の自由エネルギー差) がどのような要因によってなっているのかを調べるために、自由エネルギーをいくつかのエネルギー項に分割して計算した。第一に、塩橋を形成するリシンとアスパラギン酸側鎖間で働くエネルギー項 (E_p) で、塩橋の形成によって最も直接的に相互作用エネルギーが変化する要素である。次に、塩橋を形成するリシンの側鎖とアスパラギン酸の主鎖間、およびアスパラギン酸側鎖とリシン主鎖間での相互作用エネルギー (E_m) である。第三に、塩橋を形成するリシン、アスパラギン酸側鎖と、モデル系内に含まれるその他のアミノ酸残基、すなわち環境、との相互作用 (E_e)。最後に、塩橋を形成するリシンとアスパラギン酸側鎖のコンフォメーションエントロピー項 (E_s) であり、この項は、以下のように定義した。

$$E_s = - T \Delta S_p$$

ここで、 T は温度、 ΔS_p は塩橋形成状態と解離状態でのコンフォメーションエントロピーの差である。コンフォメーションエントロピー S_p は以下の式に沿って計算した。

$$S_p = - k_B \sum_i \int d\chi_i p_i(\chi_i) \ln p_i(\chi_i).$$

ここで、 k_B はボルツマン定数、 $p_i(\chi_i)$ は塩橋を形成するリシン、アスパラギン酸側鎖の i 番目のねじれ角 χ_i の確率分布。内部座標表現はカーテシアン座標表現よりもより正確なコンフォメーションエントロピーの値を得られるという報告があった³⁶。重要なこととして、我々は統計誤差によって引き起こされた確率分布関数の乱雑な振る舞いが、エントロピー計算に深刻な誤差を引き起こすということに気づいた。したがって、エントロピーについての誤差を小さくするた

め、確率分布関数 $p_i(\chi_i)$ は標準偏差 8 度のガウス関数を用いてスムージングをかけた。

2.3.2 結果と考察

最も単純な 2 残基系(Res2D)を用いることで、塩橋の安定性に対する Ca-Ca 間距離、すなわちリシンとアスパラギン酸の距離の影響を調べた (図 2.5)。興味深いことに、 Ca-Ca 間距離が 5.5–10.0 Å の範囲では、塩橋形成割合はほとんど変わっていない。また、塩橋の安定性がどのような相互作用エネルギー項によって決定されているのかを調べた (図 2.6(a))。その結果、5.5–8.0 Å の範囲ではリシンとアスパラギン酸の側鎖間の相互作用エネルギー (E_p) は Ca-Ca 間距離に対して単調減少しているが、環境からの相互作用によるエネルギー E_e の増加によって、 E_p の減少分が打ち消されている。8.0–10.0 Å の範囲では、 E_e も E_p もほとんど変化していない。これは、塩橋の形成割合 (塩橋形成状態と解離状態の自由エネルギー差) が Ca-Ca 間距離に対してほとんど相関していない原因と思われる。 Ca-Ca 間距離にかかわらず、Res2D モデルにおける塩橋形成割合は 0.15 程度であり、この値は HEL-HyHEL-10 複合体系における D32^H-K97^Y の塩橋とも D99^H-K97^Y の塩橋とも異なっている。なので、リシンとアスパラギン酸間の距離だけでは塩橋の安定性を決定づけるには不十分であり、次にその相対的な配向の影響を調べた。

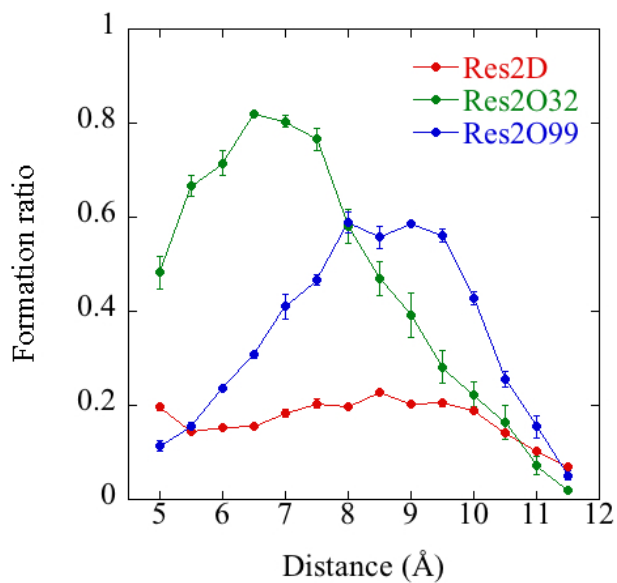


図 2.5 : C α -C α 間距離に対する塩橋の形成割合の変化。赤線が Res2D の系、緑線が Res2O32 の系、青線が Res2O99 の系。

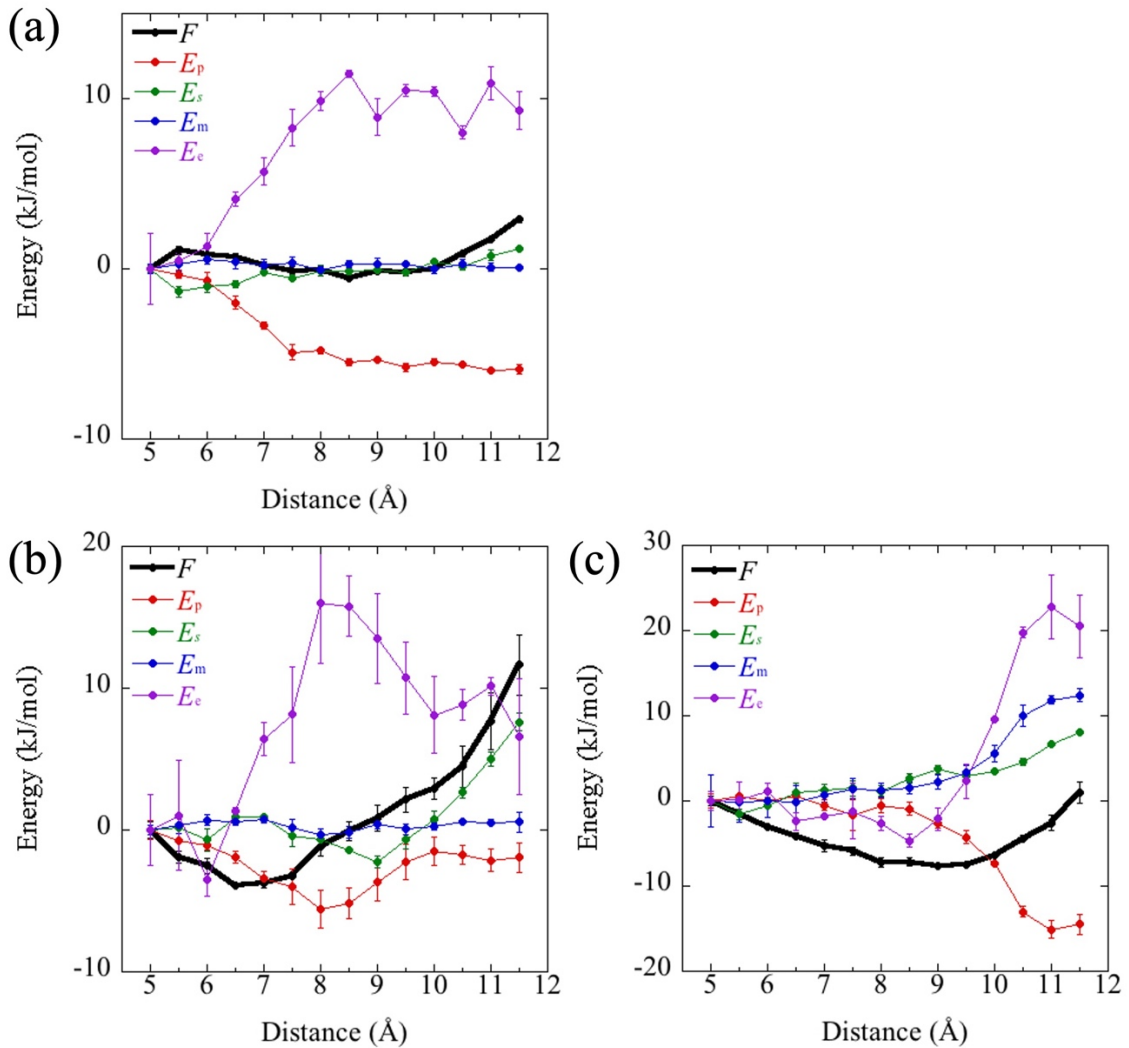


図 2.6 : Ca-Ca 間距離に対する塩橋形成時と解離時の種々のエネルギーの差。黒線が塩橋形成率から計算した自由エネルギー(F)差。赤線がリシン側鎖とアスパラギン酸側鎖の相互作用エネルギー(E_p)差。緑線がリシン側鎖とアスパラギン酸側鎖のエントロピー項のエネルギー差。青線がリシン側鎖とアスパラギン酸主鎖およびアスパラギン酸側鎖とリシン主鎖の相互作用エネルギーの和の差。紫線は塩橋を形成するリシンおよびアスパラギン酸の側鎖とその他のアミノ酸残基との相互作用エネルギー差。(a)は Res2D の系、(b)は Res2O32 の系、(c)は Res2O99 の系。

HEL-HyHEL-10 複合体の三次構造、四次構造の影響を調べるためには、リシンとアスパラギン酸の Ca-Ca 間距離だけでなく、主鎖から側鎖の伸びる向きなども束縛する必要があった。そこで、Res2O32 および Res2O99 のモデル系を構築し、側鎖の伸びる向きを束縛した影響を調べた。図 2.5 に示した通り、側鎖の向きまで束縛したことは、塩橋の安定性に対して大きな影響を及ぼした。Res2O32 の場合、塩橋の形成割合は Ca-Ca 間距離が 6.5 Å で最高となり、 Ca-Ca 間距離が 7.5 Å を過ぎたあたりから大きく減少する。それに応じた、塩橋形成状態と解離状態の自由エネルギー差は 6.5 Å で最小となり、7.5 Å から急激に増加する。この傾向は、 E_p 、 E_e 、 E_s によるものである (図 2.6(b))。Res2O99 の場合、塩橋は Ca-Ca 間距離が 8.0 – 9.5 Å の範囲で非常に安定となっており、自由エネルギー差は最小となっている。 Ca-Ca 間距離が 8.0 Å 以下の範囲では主に E_e によって自由エネルギーが減少しており、10 Å 以上の領域では E_s 、 E_m 、 E_e によって自由エネルギーが増加することでこのような傾向になっている (図 2.6(c))。

このように、リシンとアスパラギン酸の相対的な配向によって、塩橋の安定性が大きく影響を受けることが分かったが、Res2O32 のモデル系においても、Res2O99 のモデル系においても、HEL-HyHEL-10 複合体系における塩橋の安定性は再現できなかった。D32^H-K97^Y の塩橋については、複合体系では塩橋形成割合が 0.9 あり、非常に安定だったが、Res2O32 モデルでは Ca-Ca 間距離を 10.0 Å と複合体系と同様にしても、塩橋形成割合が 0.21 とあまり安定にならなかった。D99^H-K97^Y の塩橋については、複合体系では塩橋形成割合が 0.02 あり、非常に不安定だったが、Res2O99 モデルでは塩橋形成割合が 0.49 と複合体系に比べて大きく安定となった。したがって、リシンとアスパラギン酸の配向の影響を

含めても塩橋の安定性は再現できない、すなわちリシンとアスパラギン酸の2残基のみでは複合体系の塩橋の安定性は決定されていないということが分かり、塩橋周囲の界面環境の影響を調べるため、より大きな少数アミノ酸モデル (Res3-Res7)を調べた。

2.4 少数アミノ酸モデル系

2.4.1 方法

HEL-HyHEL-10 複合体系の界面の塩橋の安定性を再現するためには、リシンとアスパラギン酸の2残基(C α -C α 間距離やリシンとアスパラギン酸の配向)だけでは不十分であった。そのため、塩橋の安定性を決めているそのほかの要因を探するため、HEL-HyHEL-10 界面のアミノ酸残基を複数含めたモデル系を構築した。HEL-HyHEL-10 結晶構造において、K97^YはD32^H及びD99^H双方と塩橋を形成しており、三角形の配置で塩橋を形成している。そこで、二つの塩橋間での相互作用、すなわち、この三角形の配向が重要である可能性を考え3残基系を構築した。3残基系にはK97^Y、D32^H、D99^Hと1,629個の水分子、6個のNa⁺イオン、5個のCl⁻イオンを含めた。この系をRes3と記述した。

次に、界面残基のうちどのアミノ酸残基が塩橋に強く影響するのかを調べるため、塩橋を形成するリシン、アスパラギン酸残基側鎖と周囲のアミノ酸残基の相互作用エネルギーを調べた(図2.7(a))。そこで、K49^LとN97^Hの側鎖が塩橋を作るアミノ酸側鎖に対して非常に大きな静電相互作用を持つことが分かった。そのため、K97^Y、D32^H、D99^H、K49^L及び2,630個の水分子、7個のNa⁺イオン、7個のCl⁻イオンを含む4残基系、K97^Y、D32^H、D99^H、K49^L、N97^H及び2,667個の水分子、7個のNa⁺イオン、7個のCl⁻イオンを含む5残基系を作った。

これらの系をそれぞれ Res4、Res5 と名付けた。次に、Y33^H 及び W98^H の側鎖が塩橋を作るアミノ酸側鎖に対して大きな Lennard-Jones (LJ) 相互作用を持つため、K97^Y、D32^H、D99^H、K49^L、Y33^H、N97^H、W98^H 及び 2,794 個の水分子、8 個の Na⁺ イオン、8 個の Cl⁻ イオンを含む 7 残基系を作り、Res7M と呼称した。これらのモデル系の初期座標は全て、HEL-HyHEL-10 複合体系の MD シミュレーションの最終構造から取られている。各モデル系について初期構造は 5 つずつあり、そのそれぞれの初期構造に対して、異なる初期構造を振った 3 本の MD シミュレーションを行った。また、HEL-HyHEL-10 界面の環境を再現するため、全てのアミノ酸主鎖の重原子の座標は MD シミュレーションの初期構造の座標に束縛している。複合体系では、Y33^H 及び W98^H の側鎖の配向はあまり大きく変化しないが、Res7M のモデルでは MD シミュレーション中に大きく動いていることが分かった (図 2.8)。そのため、Y33^H 及び W98^H の側鎖の配向を再現するため、7 残基系で Y33^H 及び W98^H の側鎖の重原子の座標を束縛したモデル系、Res7S を作った。各初期構造に対して 3 本の MD シミュレーションを 120 ns に渡って行った。その他のシミュレーション条件は複合体系の場合と同様である。

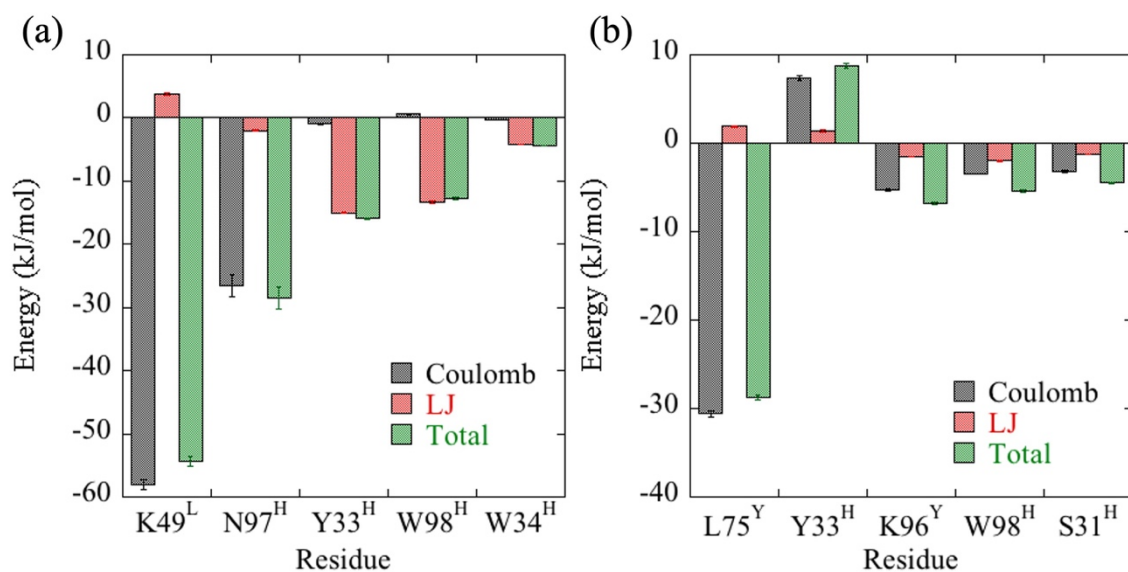


図 2.7 : 塩橋を形成するリシンとアスパラギン酸 (K97^Y、D32^H、D99^H) の側鎖に対する周囲のアミノ酸残基の相互作用エネルギー。黒棒がクーロン相互作用エネルギー、赤棒が LJ 相互作用エネルギー。緑棒は合計。また、相互作用エネルギーが大きかったアミノ酸残基のみ図に載せている。(a)は周囲のアミノ酸の側鎖との相互作用エネルギーを、(b)は主鎖との相互作用エネルギーをプロットした。

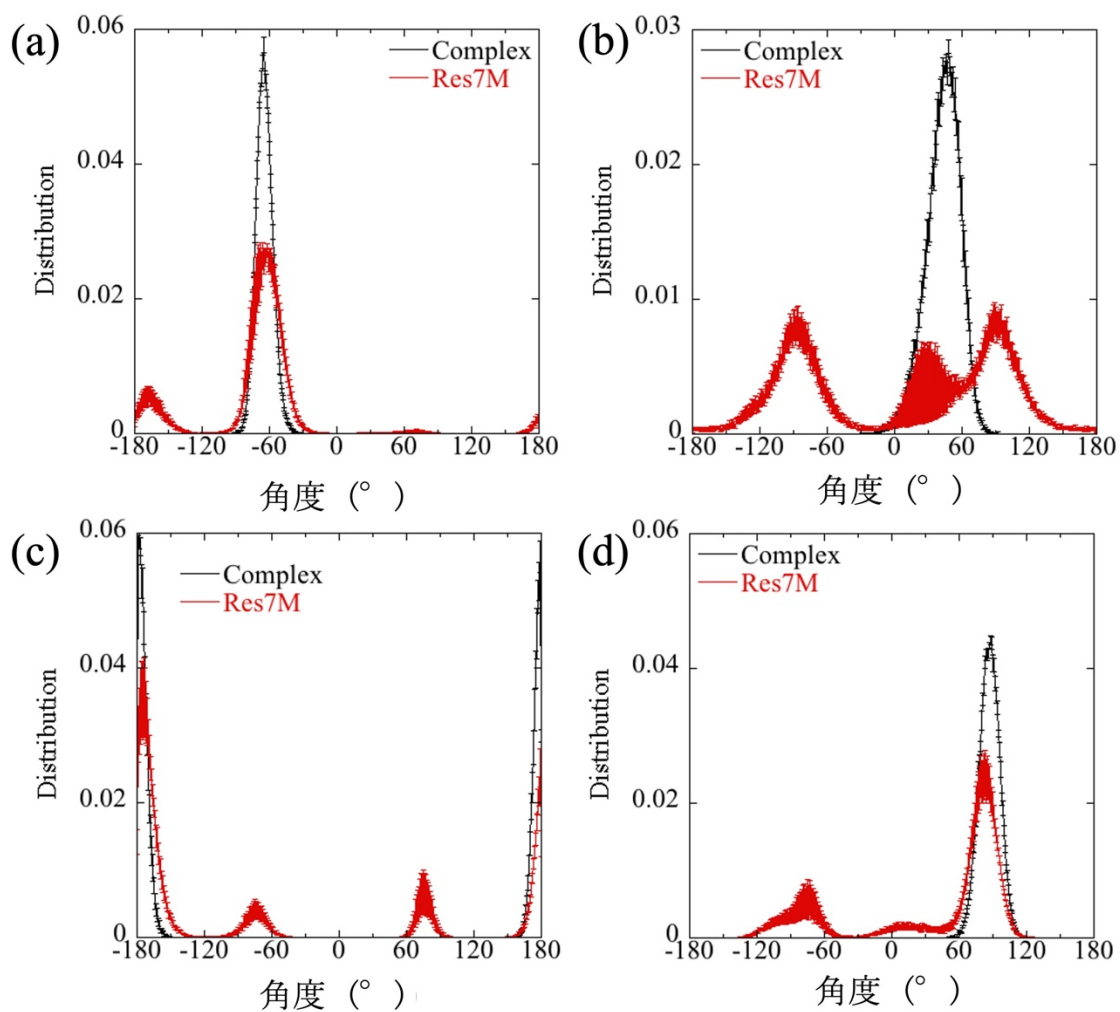


図 2.8: Y33^H 及び W98^H の側鎖の χ 角度の分布。黒線は複合体の系、赤線は Res7M の系についての分布。(a)は Y33^H の χ_1 角度の分布、(b)は Y33^H の χ_2 角度の分布、(c)は W98^H の χ_1 角度の分布、(d)は W98^H の χ_2 角度の分布。

2.4.2 結果と考察

HEL-HyHEL-10 の界面には二組の塩橋が三角形の形に配置されている。二組の塩橋がこのような形に配置されていると、互いの塩橋が強く相関している事例がある²²。そこで、これらの塩橋同士の相関による影響を調べるために、Res3 モデルを導入した。Res3 モデルにおいて D99^H-K97^Y の塩橋の形成割合は Res2 モデルよりも若干小さかったが、D32^H-K97^Y の塩橋の形成割合は Res2 モデルよりもかなり大きかった (図 2.9)。追加した 3 残基目が塩橋の安定性に影響を及ぼしていることは明確だが、Res3 モデルは HEL-HyHEL-10 複合体系の塩橋の安定性を再現できていなかった。したがって、塩橋の安定性は二組の塩橋の三角形の構造だけでなく、塩橋周囲の界面環境によっても強い影響を受けていると考えられる。

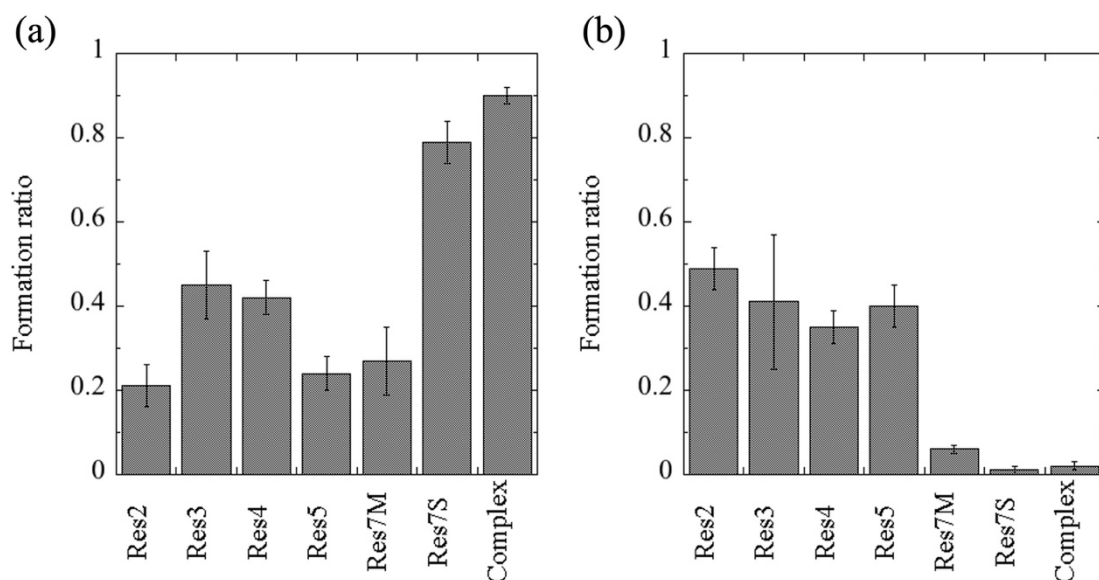


図 2.9 : 各モデル系と複合体系における塩橋の形成割合。(a)は D32^H-K97^Y について、(b)は D99^H-K97^Y についての図。

塩橋周囲の強い曲性を持つ残基の影響を調べるため、Res4 および Res5 モデルを利用した。Res4 モデルは Res3 モデルに K49^L の残基を追加しており、Res5 モデルは Res4 モデルに更に N97^H の残基を追加している。K49^L と N97^H の残基は塩橋を作るリシンとアスパラギン酸残基にクーロン相互作用で直接的に、強く相互作用を及ぼすため、これらの残基が塩橋安定性をコントロールすると考えた。しかしながら、これらのモデルは Res3 モデルと比較して、HEL-HyHEL-10 複合体系における塩橋安定性との再現度を向上させなかった。D32^H-K97^Y の塩橋の Res4 モデルにおける形成割合は Res3 モデルとほぼ同じだった。複合体系における D32^H-K97^Y の塩橋の形成割合は Res3 モデルと比較して非常に高いが、Res5 モデルにおいては逆に Res3 モデルよりも不安定になってしまった。D99^H-K97^Y の塩橋については、K49^L と N97^H の残基を追加したことは、塩橋形成割合に対して大きな影響を及ぼさず、Res4 や Res5 モデルにおいても複合体系よりもずっと高くなっている。

次に、LJ 相互作用によって二つの塩橋と強く相互作用する、塩橋周囲の芳香族残基の影響について、Res7M モデルを構築して調べた。Res7M モデルは Res5 モデルに Y33^H と W98^H の二つの残基を追加して構築した。これらの残基は塩橋を形成するリシン、アスパラギン酸に対して強い静電相互作用は及ぼさないが、LJ 相互作用を通して塩橋安定性に影響を及ぼしている可能性がある。実際、Res7M モデルにおける D99^H-K97^Y の塩橋の形成割合は HEL-HyHEL-10 複合体系とほぼ一致している (図 2.9(b))。Res2-5 モデルと比べると、D99^H-K97^Y の塩橋を形成するリシン、アスパラギン酸残基と周囲の残期間での相互作用エネルギー (E_c) が大きく増加しており、これによって塩橋形成状態と解離状態の自由エネルギー差が増加することを説明できる (図 2.10(b))。

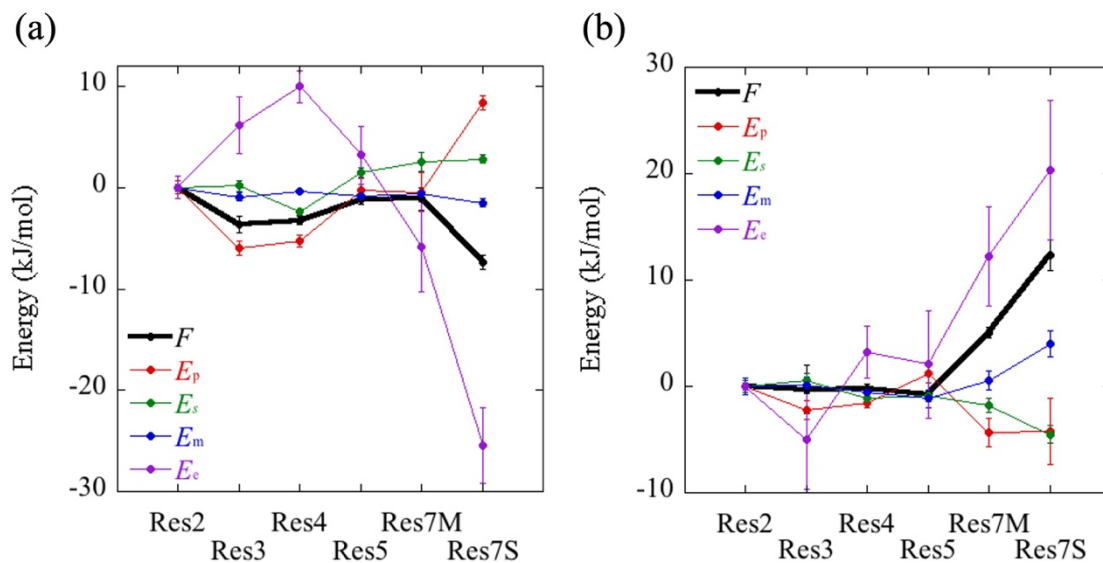


図 2.10 : 各モデル系においての塩橋形成時と解離時の種々のエネルギーの差。黒線が塩橋形成率から計算した自由エネルギー(F)差。赤線がリシン側鎖とアスパラギン酸側鎖の相互作用エネルギー(E_p)差。緑線がリシン側鎖とアスパラギン酸側鎖のエントロピー項のエネルギー差。青線がリシン側鎖とアスパラギン酸主鎖およびアスパラギン酸側鎖とリシン主鎖の相互作用エネルギーの和の差。紫線は塩橋を形成するリシンおよびアスパラギン酸の側鎖とその他のアミノ酸残基との相互作用エネルギー差。(a)は D32^H-K97^Y について、(b)は D99^H-K97^Y についての図。

しかしながら、Res7M モデルでも D32^H-K97^Y の塩橋の安定性は説明できない。そこで、Y33^H と W98^H の残基の側鎖の動きを調べた。これらの残基の χ_1 、 χ_2 角の分布を調べると、HEL-HyHEL-10 複合体系においてはこれらの残基の側鎖は大きく動くことはできないのに対し、Res7M モデルでは大きく動くことが分かった (図 2.8)。複合体系においては Y33^H と W98^H の残基はこれらを取り囲むより大きな周囲の環境としての残基によって大きく動くことを制限されている

と考えられる。そこで、この効果を再現するために、Y33^HとW98^Hの側鎖の動きの束縛をRes7Mモデルに導入し、Res7Sモデルを構築した。

Res7Sモデルの導入によって、D32^H-K97^Yの塩橋の形成割合はHEL-HyHEL-10複合体系と同程度になった(図2.9(a))。Res5モデルと比較すると、D32^H-K97^Yの塩橋は形成割合で0.55、自由エネルギー差で-6.3 kJ/molだけ安定化している。この結果は自由エネルギー変化のうち、 E_e の部分の挙動から(図2.10(a))、Y33^HとW98^Hの側鎖の環の配向がD32^H-K97^Yの塩橋を形成するリシン、アスパラギン酸と周囲の環境との相互作用に影響し、塩橋を安定化させていると考えられる。D99^H-K97^Yの塩橋については、Res7Sモデルは複合体系における塩橋形成割合をRes7Mモデルよりも少しだけよく再現している。図2.9(a)によると、D99^H-K97^Yの塩橋の形成割合はRes7S、Res7Mモデルと複合体系でそれぞれ 0.01 ± 0.01 、 0.06 ± 0.01 、および 0.02 ± 0.01 だった。Res5モデルと比較した場合だと、Res7SモデルではD99^H-K97^Yの塩橋の形成割合は0.39だけ不安定化しており、自由エネルギー差だと13.1 kJ/molだけ上昇している。この結果から、Y33^HとW98^Hの側鎖の環の配向によってはD99^H-K97^Yの塩橋を少し不安定化させる影響を及ぼしていることがわかる。以上の結果により、我々はRes7Sモデルが複合体系における二組の塩橋の挙動を説明する十分かつ最小のモデルであると結論づけた。

HELとHyHEL-10の界面において存在した二組の塩橋のうち、D32^H-K97^Yの塩橋はシミュレーション中でも安定だったが、D99^H-K97^Yの塩橋は不安定だった。塩橋は負に荷電した残基と正に荷電した残基の間での静電相互作用によって形成される構造であるため、塩橋を形成するリシンとアスパラギン酸の配向と周囲の残基との静電相互作用が塩橋の安定性を決定づける支配的な要因だと予測した。しかしながら、Res2-5モデルではHEL-HyHEL-10複合体系におけ

る D32^H-K97^Y と D99^H-K97^Y の塩橋の形成割合は再現できなかった。D32^H-K97^Y の塩橋を安定化し、なおかつ D99^H-K97^Y の塩橋を不安定化するモデル系を構築するには、Y33^H と W98^H の残基の存在が不可欠だった。さらに、Y33^H と W98^H の側鎖の環の配向が極めて重要だった。

2.5 Ala 変異が塩橋に及ぼす影響

2.5.1 方法

D32^HA および D99^HA 変異を用いた実験研究により、D32^H-K97^Y と D99^H-K97^Y の二組の塩橋が HEL-HyHEL-10 複合体の結合親和性に重要な役割を果たしていることが判明している¹⁹⁻²⁰。しかしながら、分子レベルで何が起きているのかについてはいまだに不鮮明である。これらの変異の影響を調べるために、D32^HA および D99^HA 変異をした HEL-HyHEL-10 複合体についての MD シミュレーションを行った。MD シミュレーションは結晶構造(PDB ID 1IC4 : D32^HA、1IC5 : D99^HA)を元に水素を追加した構造を初期構造として、初期速度を変えて5本のシミュレーションをそれぞれ 300 ns 行った。単位セルには HEL-HyHEL-10 変異複合体一つと、18,464 個の水分子、55 の Na⁺イオン、61 の Cl⁻イオンを含めた。そのほかのシミュレーション条件、および手順は WT の HEL-HyHEL-10 複合体系の際と同様に行った。

さらに、HEL-HyHEL-10 複合体に変異を導入したことが、塩橋の安定性に対してどのような影響を及ぼしているのかをより詳細に調べるために、WT で行ったのと同様に、変異体についての少数アミノ酸モデルを構築した。最初に、WT において HEL-HyHEL-10 複合体系における塩橋安定性を説明する十分かつ最小のモデルと考えた Res7S モデルについて、D32^HA および D99^HA 変異体について

も構築し、調べた。これらの初期構造は、WTの複合体から作成した Res7S モデルに対して、D32^HA および D99^HA の変異を導入することによって作成した。水分子の一つを Na⁺イオンに置き換えることで、変異を導入したことによる系の電荷の変化を打ち消し、中性の電荷を保った。MD シミュレーションの条件は WT に関する少数アミノ酸モデル系と同様とした。D32^HA 変異によって引き起こされる環境の効果の影響を詳細に調べるために、WT と D32^HA 変異体についての 8 残基モデル(Res8S)および 9 残基モデル(Res9S)を構築した。WT の Res8S モデルは K97^Y、D32^H、D99^H、K49^L、Y33^H、N97^H、W98^H、L75^Y の 8 残基と 2,756 個の水分子、8 個の Na⁺イオン、8 個の Cl⁻イオンを一つの単位セルに含めた。Res9S モデルでは、K97^Y、D32^H、D99^H、K49^L、Y33^H、N97^H、W98^H、L75^Y、D101^H の 9 残基と 2,725 個の水分子、9 個の Na⁺イオン、8 個の Cl⁻イオンを一つの単位セルに含めた。L75^Y の残基は WT の HEL-HyHEL-10 複合体中での塩橋を作るリシン、アスパラギン酸残基との相互作用エネルギーの解析の結果から選んだが、D101^H の残基は D32^HA 変異を導入した HEL-HyHEL-10 複合体中での相互作用エネルギーの解析を元に導入した。D32^HA 変異体についての Res8S モデルおよび Res9S モデルを構築する際には、Res7S モデルの場合と同様に、WT 複合体についての Res8S モデル、Res9S モデルにそれぞれ D32^HA 変異を導入し、電気的中性を保つ為に水分子の一つを Na⁺イオンに置き換えた。また、Res7S モデルの場合と同様に、Res8S モデル、Res9S モデルでも主鎖だけでなく Y33^H と W98^H の側鎖の重原子の座標を束縛した。

後述するが、アラニン変異が主鎖の骨格構造の変化を引き起こし、さらにこの主鎖構造の変化が塩橋安定性に影響を及ぼしうるということが示唆された。この主鎖構造に対する影響について検証する為、Res7S、Res8S、および Res9S モ

デルの主鎖および Y33^H と W98^H の側鎖を束縛する基準となる構造を変えたモデル系を構築した。Res7S-9S のモデル系では、束縛の基準となる構造は、WT の HEL-HyHEL-10 複合体についての MD シミュレーションの最終構造を用いていたが、新たに構築したモデル系では D32^HA 変異を施した複合体についての MD シミュレーションの最終構造を用いた。D32^HA 変異体についても、初期速度の違いにより 5 本おシミュレーションを行った為、各モデル系について 5 つの初期構造ができ、それぞれの初期構造について初期速度を変えて 3 本の MD シミュレーションを 120 ns にわたって行なった。D32^HA 変異体の構造を束縛の基準に使ったモデル系については、それぞれ Res7Sm、Res8Sm、Res9Sm モデルと呼称した。Res7Sm-9Sm のモデル系についてのシミュレーション条件、操作については、その他のモデル系の場合と同様にした。

2.5.2 結果と考察

実験的に抗原-抗体間の親和性に対する塩橋の効果についての研究はなされており、Ala 変異を導入することはよく用いられている。HEL-HyHEL-10 複合体については、D32^H-K97^Y と D99^H-K97^Y の二つの塩橋について、D32^HA および D99^HA の変異を導入することで、これらの塩橋が形成されなくなることによる結合親和性の減少が確認されている。特に D32^HA による結合親和性の減少は D99^HA によるものよりも大きい。K_a の値は、WT で 4.2×10^8 、D32^HA 変異体で 1.0×10^8 、D99^HA 変異体で 2.1×10^8 M であった。結合自由エネルギーはそれぞれ -50.2、-46.4、-48.2 kJ/mol であり、WT からの変化は D32^HA 変異体で +3.8、D99^HA 変異体で +2.1 kJ/mol であった。D32^HA 変異が D99^HA 変異よりも大きな影響を及ぼしていることは、WT の HEL-HyHEL-10 複合体において D32^H-K97^Y の塩橋が

D99^H-K97^Y の塩橋よりもずっと安定であったという MD シミュレーションの結果とも一致している。しかしながら、WT の HEL-HyHEL-10 複合体における塩橋の形成割合は D32^H-K97^Y の塩橋が 0.90、D99^H-K97^Y の塩橋が 0.05 であり、この差の大きさは実験研究における D32^H-K97^Y と D99^H-K97^Y の塩橋の HEL-HyHEL-10 の結合親和性に対する寄与の大きさの差と一致していない。すなわち、D99^HA 変異の導入による結合自由エネルギーの変化の大きさが相対的に大きすぎるように思える。

D32^H-K97^Y と D99^H-K97^Y の二つの塩橋が Ala 変異によってそれぞれ形成できなくなった時の影響を調べるために、HEL-HyHEL-10 複合体の D32^HA および D99^HA 変異体の MD シミュレーションを行った。WT の系における D32^H-K97^Y の塩橋の形成割合は 0.90 ± 0.02 であったのに対し、D99^HA 変異体の系においては、 0.99 ± 0.00 であった。この結果は K97^Y と相互作用する D99^H の残基がなくなることで、D32^H-K97^Y の塩橋が安定化しているように思える。これに対して D99^H-K97^Y の塩橋は WT の系ではほとんど形成されていなかったが（形成割合 0.02 ± 0.01 ）、D32^HA 変異を入れることによって大きく安定化するようになった（形成割合 0.42 ± 0.19 ）。これは、D32^H の残基の存在によって、D99^H-K97^Y の塩橋が形成されることが強く抑制されるということを示している。エネルギー的な観点からは、Ala 変異により静電相互作用の減少が引き起こされたが、三角構造をとっていた塩橋のうち、残った組の塩橋が安定化されることによって、減少した静電相互作用エネルギーが部分的に補填された、すなわち残った組の塩橋の安定化が緩衝効果として働いた。これが、D32^H-K97^Y の塩橋は D99^H-K97^Y の塩橋よりもずっと安定なのに、D32^HA と D99^HA の変異が実験的に同程度の $\Delta\Delta G$ の値である理由であると考えられる。

より詳細に Ala 変異の影響を調べるために、D32^HA および D99^HA 変異体の系について、Res7S モデルを構築した。これらのモデルにおいて、主鎖構造は WT の Res7S モデルのものと同じ参照座標に束縛した。この条件での束縛でも、D99^HA 変異を導入することによって D32^H-K97^Y の塩橋の形成割合は 0.08 上がった (図 2.11)。これは複合体の WT と D99^HA 変異体の系とよく似た傾向となっている。そのため、我々は D99^HA 変異複合体系における D32^H-K97^Y の塩橋の安定性を決定する主要な要素は Res7S モデルに含まれていると結論づけた。エネルギー分解解析では、D32^H-K97^Y の塩橋は E_p および E_e によって安定化していると示されている (図 2.10(a))。

D32^HA 変異体については、Res7S モデルでは D99^H-K97^Y の塩橋の形成割合は 0.03 ± 0.01 と、WT の Res7S モデルにおける 0.01 ± 0.01 からほとんど向上しなかった (図 2.12)。しかしながら、複合体の系では D32^HA 変異体系は塩橋形成割合 0.42 ± 0.19 と、D99^H-K97^Y の塩橋は WT よりもずっと安定化している。Res7S モデルで D32^HA 変異による D99^H-K97^Y の塩橋の安定化が再現できないということは、D32^HA 変異によって引き起こされた塩橋の周囲での環境の変化が大きいということを示している。

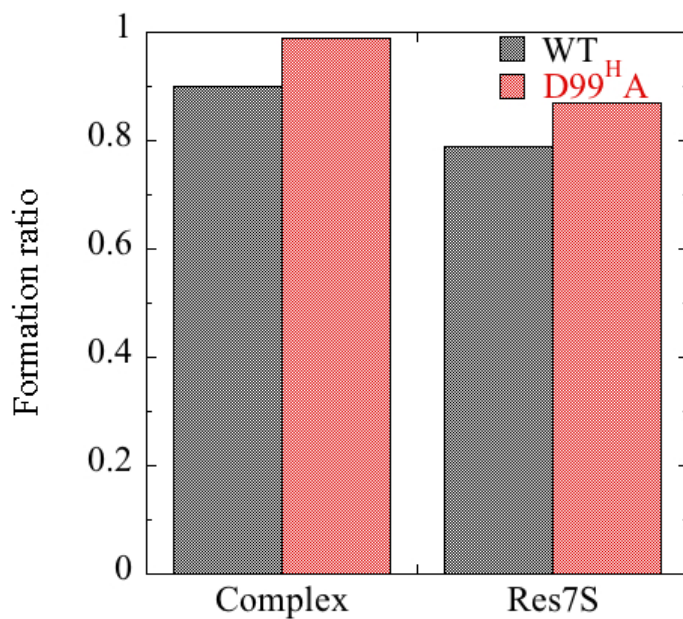


図 2.11 : 塩橋の形成割合。黒棒が WT、赤棒が D99^HA の変異を加えた複合体系および Res7S モデル系。

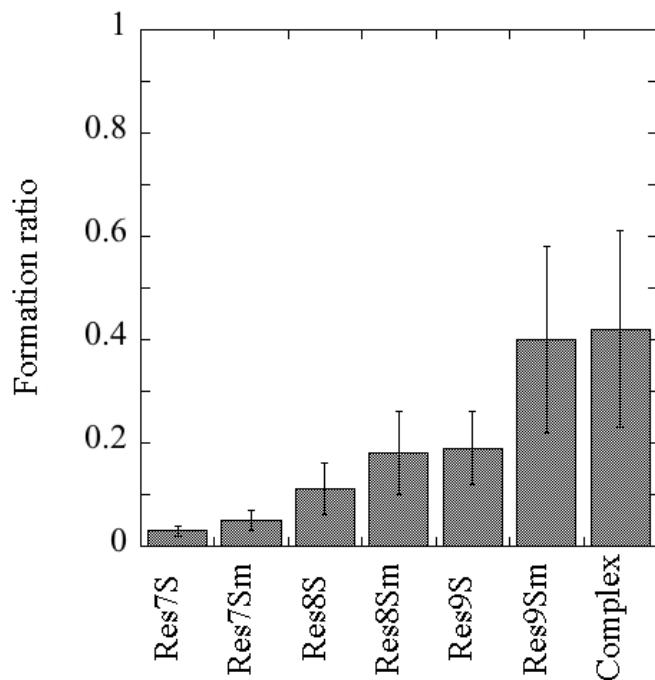


図 2.12 : D32^HA の変異を加えた各モデル系および複合体系における塩橋の形成割合。

Res7S モデルにおける塩橋周りでの環境をより複合体系に近づけるために、Les7S モデルに L75^Y の残基を加えて Res8S モデルを構築した。L75^Y は主鎖が K97^Y の側鎖と強く相互作用している。塩橋の形成割合は 0.11 ± 0.05 に上がったが、複合体系の 0.42 ± 0.19 よりはいまだに低い (図 2.12)。そのため、さらに、Res8S モデルに D101^H の残基を加えて Res9S モデルを構築した。D101^H は WT の複合体系では、二つの塩橋に対してほとんど相互作用を及ぼしていなかったが、D32^HA 変異体の複合体系では D99^H 残基と LJ 相互作用をしている (-3.2 ± 0.3 kJ/mol)。結果として、Res9S モデルにおける塩橋の形成割合は Res8S モデルからさらに 0.08 だけ向上したが (図 2.12)、複合体系における形成割合よりもいまだに低かった。それに対して、WT の Res8S および Res9S モデルにおける塩橋の形成割合は Res7S モデルのものと非常に近かった。これは、塩橋の形成割合を決める主要な環境の効果が Res7S モデルにすでに含まれており、WT においては L75^Y と D101^H の残基は塩橋にほとんど影響を及ぼさないことを示している。

次に、変異を導入したことによって、主鎖の立体構造が変化していないかを調べた。立体構造の変化を調べるために、WT と変異複合体に対して、Res7S モデルに含めた 7 つのアミノ酸残基の骨格構造についての RMSD を計算した (図 2.13)。WT 複合体系の結晶構造を基準の構造とした。結果は、D32^HA 変異体についての RMSD が WT および D99^HA 変異体の RMSD よりも大きかった。そのため、WT や D99^HA 変異体と比べ、D99^HA の Res7S モデルに入れたアミノ酸残基の主鎖の立体構造は、束縛した構造に対して大きくゆがんでいたことが分かった。この D32^HA 変異によって引き起こされる主鎖の立体構造の歪みが、塩橋

の安定性に対して影響を及ぼしているのではないかと考えた。この仮定について検証するために、主鎖構造を束縛するための基準となる構造を WT ではなく D32^HA 変異複合体として、Res7Sm、Res8Sm、Res9Sm の三つのモデルを構築した。これらのモデルはそれぞれ Res7S、Res8S、Res9S のモデルと含まれているアミノ酸などの個数は同一で、主鎖構造を束縛する基準が D32^HA 複合体の構造か、WT 複合体の構造かのみが異なる。

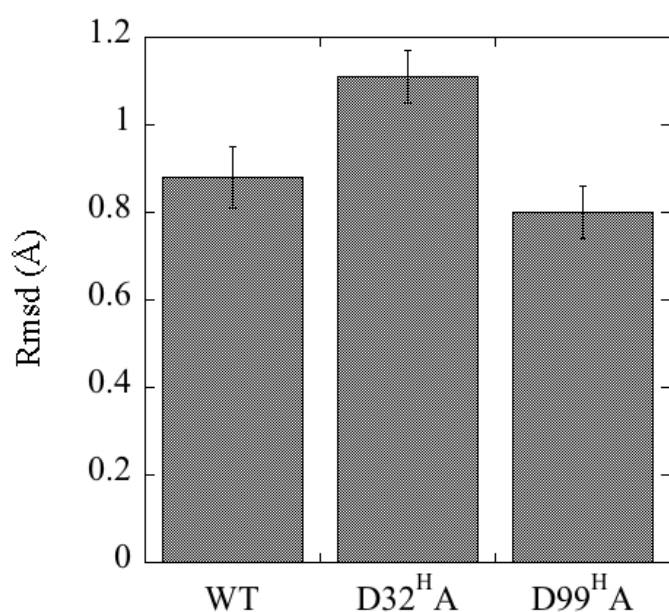


図 2.13 : Res7S モデルに含めた 7 つのアミノ酸残基の骨格構造についての RMSD。WT 複合体系の結晶構造を基準の構造とした。

Res7Sm および Res8Sm モデルについては、塩橋の形成割合はそれぞれ 0.05 ± 0.02 、 0.18 ± 0.08 と Res7S(0.03 ± 0.01)、Res8S(0.11 ± 0.05)モデルからわずかに上昇したが、D32^HA 複合体系の 0.42 ± 0.19 にはまだ届かなかった。しかし、Res9Sm モデルになると、塩橋の形成割合が 0.40 ± 0.18 と Res9S モデル(0.19 ± 0.07)から倍程度上昇し、D32^HA 複合体系における形成割合とほぼ同一となった (図 2.12)。これらの結果から、塩橋の安定性に対して主鎖の立体構造が大きく

影響を与えていたことが分かった。したがって D99^HA 変異が主鎖の立体構造に対して大きな歪みを生じさせなかったのに対して、D32^HA 変異は主鎖の立体構造に大きな歪みを引き起こすことによって、結果として D99^H-K97^Y の塩橋の安定性を大きく上げた結論づけた。エネルギーの観点からは、主鎖の立体構造の変化および L75^Y と D101^H の二つの残基が追加されたことの双方により、 E_p の成分が減少した (図 2.14)。この効果が D99^H-K97^Y の塩橋を安定化させていると考えられる。

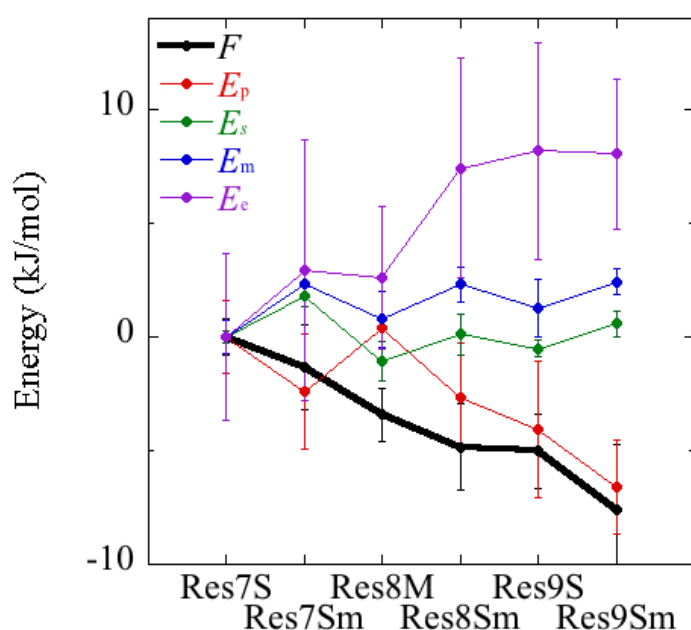


図 2.14 : D32^HA の変異を加えた各モデル系においての塩橋形成時と解離時の種々のエネルギーの差。黒線が塩橋形成率から計算した自由エネルギー(F)差。赤線がリシン側鎖とアスパラギン酸側鎖の相互作用エネルギー(E_p)差。緑線がリシン側鎖とアスパラギン酸側鎖のエントロピー項のエネルギー差。青線がリシン側鎖とアスパラギン酸主鎖およびアスパラギン酸側鎖とリシン主鎖の相互作用エネルギーの和の差。紫線は塩橋を形成するリシンおよびアスパラギン酸の側鎖とその他のアミノ酸残基との相互作用エネルギー差。

2.6 結論

塩橋は多くの抗原-抗体界面に存在し、塩橋が形成されることは抗原-抗体相互作用に対して重要な役割をする^{20,26-28}。抗体の設計をする際、新たに導入した塩橋が不安定で、結果として結合親和性が向上しないということがしばしばある。したがって、塩橋の安定性について理解することは抗体の設計を行う上でも助けとなる。本研究では、塩橋の安定性についての解析に焦点を当てた。卵白リゾチーム(HEL)とその抗体HyHEL-10の界面に形成される二つの塩橋の安定性を調べるためにMDシミュレーションを行った。二つの塩橋のうちD32^H-K97^Yの塩橋は安定に存在していたが、D99^H-K97^Yの塩橋はMDシミュレーション中ほとんど解離していた。本研究の主目的はHEL-HyHEL-10複合体系のMDシミュレーション中で観測された塩橋の安定性について理解することである。

HEL-HyHEL-10複合体系においてどのような要素が塩橋の安定性を決定づけているのかを明らかにするために、いくつかの少数アミノ酸モデルを導入した。モデルに塩橋周りの環境としてのアミノ酸を含めなかった場合、すなわちリシンとアスパラギン酸の2残基のみのモデル系の場合、複合体系において観測された塩橋安定性を再現することはできなかった。塩橋は強い静電相互作用によって形成された構造であるため、塩橋周りの環境としては、親水性のアミノ酸残基が重要であろうと考えられる³⁷。しかしながら、少数アミノ酸モデル系を使った解析の結果、疎水性残基であるチロシンとトリプトファン(Y33^H、W98^H)が塩橋の安定性を決めるのに重要であるということが分かった。とりわけ、D32^H-K97^Yの塩橋に対しては、Y33^HとW98^Hの側鎖の環の動きを束縛することが複合体系における安定性を再現するために必要であった。実際、側鎖の束縛を導入す

ることによって、D32^H-K97^Yの塩橋の形成割合は上昇した。その際、塩橋と環境との間の相互作用エネルギーは低下した。したがって、Res7Sモデルに含まれるK49^L、N97^H、Y33^H、W98^Hのアミノ酸残基が、塩橋周囲の環境として塩橋の安定性を決める主要な役割を果たしていると我々は結論づけた。塩橋の安定性を決めるのには静電的な作用が主要な役割を果たすだろうと直感的には思えるが、リシンとアスパラギン酸の直接的な相互作用は塩橋の安定性を決めるのに主要な役割とはなっていなかった。この結果は、塩橋のエンタルピー的な関連性が少ないという最近の研究とも一致している。

更に、D32^HA と D99^HA の変異を加えた HEL-HyHEL-10 系における D32^H-K97^Y と D99^H-K97^Y の塩橋の挙動について調べた。WT 複合体系における MD シミュレーション中では、D32^H-K97^Y の塩橋は D99^H-K97^Y の塩橋よりもずっと安定であったが、過去に行われた実験研究では、D32^HA と D99^HA の変異を加えることによって、どちらも同じように HEL-HyHEL-10 間の結合親和性が低下することが報告されている¹⁹⁻²⁰。それぞれの変異体系における塩橋の役割を理解するために、D32^HA と D99^HA の変異体系についての MD シミュレーションを行った。その結果、WT の複合体系においてもともと安定に存在していた D32^H-K97^Y の塩橋は D99^HA の変異体系においても同様に存在していただけだったが、WT の系においてほとんど形成されていなかった D99^H-K97^Y の塩橋は D99^HA の変異体系においてしばしば形成されるようになっていた。したがって、Ala 置換によって塩橋の三角形の構造が取れなくなり、生じたエネルギー的な損失が残ったもう片方の塩橋の安定化によって緩和されていると結論づけた。先行研究の中で、ROBO1 とその抗体 B5209B の間に形成される塩橋の三角構造について、同様の効果が観測された²²。塩橋に対する変異の効果をより詳細に調べるため、

WT の系に対して構築した Res7S モデルを適用した。モデル系の主鎖構造を WT の複合体の主鎖構造に束縛した場合、すなわち WT のモデル系に変異を施した系だと、D99^HA 変異体の系における D32^H-K97^Y の塩橋の安定性はよく再現できていた。しかし、D32^HA 変異体の系における D99^H-K97^Y の塩橋の安定性は Res7S モデル系では再現できず、変異複合体系の主鎖構造に束縛した上で L75^Y と D101^H のアミノ酸残基を加えた Res9Sm モデル系を導入しなければならなかった。この結果が示すこととして、D99^HA 変異は塩橋周囲の環境に対してほとんど影響を及ぼさないが、D32^HA 変異は WT の構造からの主鎖の歪みを引き起こし、L75^Y と D101^H のアミノ酸残基との強い相互作用を発生させる。

我々は MD シミュレーションを用いて HEL-HyHEL-10 界面に存在する二つの塩橋の安定性を決める要因について調べ、特定してきた。少数アミノ酸モデルを使った解析により、塩橋の安定性を決めるために重要な塩橋周囲の残基や立体構造を特定することに成功した。D32^H-K97^Y と D99^H-K97^Y の塩橋の安定性を決めるのに、どちらも Y33^H と W98^H のアミノ酸残基の存在が主要な役割を果たした。方法論的には、ここで使った構成論的アプローチは、主鎖などの立体構造の束縛や変異の導入、主鎖構造の歪みなど、塩橋の安定性に対する様々な影響を体系的に定量化できる。しかしながら、統計誤差が大きすぎるため、どのエネルギー項の変化が塩橋安定性を決める要因となったのかについては断定することはできなかった。加えて、水分子のエントロピー的な寄与が抗原-抗体間の結合自由エネルギーに対して影響を与えているが^{20, 22}、本研究ではこの効果は含めていなかった。これらのテーマは抗体医薬品の設計をする上で重要な課題として残っている。

参考文献

1. Parkin, J.; Cohen, B., An overview of the immune system. *Lancet* **2001**, *357* (9270), 1777-1789.
2. Delves, P. J.; Roitt, I. M., The immune system. *N. Engl. J. Med.* **2000**, *343* (1), 37-49.
3. Senter, P. D.; Sievers, E. L., The discovery and development of brentuximab vedotin for use in relapsed Hodgkin lymphoma and systemic anaplastic large cell lymphoma. *Nat. Biotechnol.* **2012**, *30* (7), 631-637.
4. Phillips, G. D. L.; Li, G.; Dugger, D. L.; Crocker, L. M.; Parsons, K. L.; Mai, E.; Blättler, W. A.; Lambert, J. M.; Chari, R. V.; Lutz, R. J., Targeting HER2-positive breast cancer with trastuzumab-DM1, an antibody–cytotoxic drug conjugate. *Cancer Res.* **2008**, *68* (22), 9280-9290.
5. Lambert, J. M.; Chari, R. V., Ado-trastuzumab Emtansine (T-DM1): an antibody–drug conjugate (ADC) for HER2-positive breast cancer. *J. Med. Chem.* **2014**, *57* (16), 6949-6964.
6. Junutula, J. R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D. D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S. P.; Dennis, M. S., Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol.* **2008**, *26* (8), 925.
7. Han, T. H.; Zhao, B., Absorption, distribution, metabolism, and excretion considerations for the development of antibody-drug conjugates. *Drug Metab. Dispos.* **2014**, *42* (11), 1914-1920.
8. Hamblett, K. J.; Senter, P. D.; Chace, D. F.; Sun, M. M.; Lenox, J.; Cerveny, C. G.; Kissler, K. M.; Bernhardt, S. X.; Kopcha, A. K.; Zabinski, R. F., Effects of drug loading on the antitumor activity of a monoclonal antibody drug conjugate. *Clin. Cancer Res.* **2004**, *10* (20), 7063-7070.
9. Donaghy, H., Effects of antibody, drug and linker on the preclinical and clinical toxicities of antibody-drug conjugates. *MAbs* **2016**, *8* (4), 659-671.
10. Adams, G. P.; Weiner, L. M., Monoclonal antibody therapy of cancer. *Nat. Biotechnol.* **2005**, *23* (9), 1147-1157.

11. Yokota, A.; Tsumoto, K.; Shiroishi, M.; Nakanishi, T.; Kondo, H.; Kumagai, I., Contribution of asparagine residues to the stabilization of a proteinaceous antigen-antibody complex, HyHEL-10-hen egg white lysozyme. *J. Biol. Chem.* **2010**, *285* (10), 7686-7696.
12. Yokota, A.; Tsumoto, K.; Shiroishi, M.; Kondo, H.; Kumagai, I., The Role of Hydrogen Bonding via Interfacial Water Molecules in Antigen-Antibody Complexation THE HyHEL-10-HEL INTERACTION. *J. Biol. Chem.* **2003**, *278* (7), 5410-5418.
13. Murase, T.; Zheng, R. B.; Joe, M.; Bai, Y.; Marcus, S. L.; Lowary, T. L.; Ng, K. K., Structural insights into antibody recognition of mycobacterial polysaccharides. *J. Mol. Biol.* **2009**, *392* (2), 381-392.
14. McLellan, J. S.; Chen, M.; Leung, S.; Graepel, K. W.; Du, X.; Yang, Y.; Zhou, T.; Baxa, U.; Yasuda, E.; Beaumont, T., Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **2013**, *340* (6136), 1113-1117.
15. Garces, F.; Lee, J. H.; de Val, N.; de la Pena, A. T.; Kong, L.; Puchades, C.; Hua, Y.; Stanfield, R. L.; Burton, D. R.; Moore, J. P., Affinity maturation of a potent family of HIV antibodies is primarily focused on accommodating or avoiding glycans. *Immunity* **2015**, *43* (6), 1053-1063.
16. Rajpal, A.; Taylor, M. G.; Kirsch, J. F., Quantitative evaluation of the chicken lysozyme epitope in the HyHEL-10 fab complex: Free energies and kinetics. *Protein Sci.* **1998**, *7* (9), 1868-1874.
17. Pons, J.; Rajpal, A.; Kirsch, J. F., Energetic analysis of an antigen/antibody interface: alanine scanning mutagenesis and double mutant cycles on the HyHEL-10/lysozyme interaction. *Protein Sci.* **1999**, *8* (5), 958-968.
18. Cunningham, B. C.; Wells, J. A., High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **1989**, *244* (4908), 1081-1085.
19. Tsumoto, K.; Ogasahara, K.; Ueda, Y.; Watanabe, K.; Yutani, K.; Kumagai, I., Role of Salt Bridge Formation in Antigen-Antibody Interaction ENTROPIC CONTRIBUTION TO THE COMPLEX

BETWEEN HEN EGG WHITE LYSOZYME AND ITS MONOCLONAL ANTIBODY HyHEL10. *J. Biol. Chem.* **1996**, *271* (51), 32612-32616.

20. Shiroishi, M.; Yokota, A.; Tsumoto, K.; Kondo, H.; Nishimiya, Y.; Horii, K.; Matsushima, M.; Ogasahara, K.; Yutani, K.; Kumagai, I., Structural Evidence for Entropic Contribution of Salt Bridge Formation to a Protein Antigen-Antibody Interaction THE CASE OF HEN LYSOZYME-HyHEL-10 Fv COMPLEX. *J. Biol. Chem.* **2001**, *276* (25), 23042-23050.

21. Yoshida, K.; Kuroda, D.; Kiyoshi, M.; Nakakido, M.; Nagatoishi, S.; Soga, S.; Shirai, H.; Tsumoto, K., Exploring designability of electrostatic complementarity at an antigen-antibody interface directed by mutagenesis, biophysical analysis, and molecular dynamics simulations. *Sci. Rep.* **2019**, *9* (1), 4482.

22. Yamashita, T.; Mizohata, E.; Nagatoishi, S.; Watanabe, T.; Nakakido, M.; Iwanari, H.; Mochizuki, Y.; Nakayama, T.; Kado, Y.; Yokota, Y., Affinity Improvement of a Cancer-Targeted Antibody through Alanine-Induced Adjustment of Antigen-Antibody Interface. *Structure* **2019**, *27* (3), 519-527. e5.

23. Sinha, N.; Li, Y.; Lipschultz, C. A.; Smith-Gill, S. J., Understanding antibody-antigen associations by molecular dynamics simulations: Detection of important intra-and inter-molecular salt bridges. *Cell Biochem. Biophys.* **2007**, *47* (3), 361-375.

24. Moreira, I.; Fernandes, P.; Ramos, M., Hot spot computational identification: Application to the complex formed between the hen egg white lysozyme (HEL) and the antibody HyHEL-10. *Int. J. Quantum Chem.* **2007**, *107* (2), 299-310.

25. Liu, W.; Liu, G.; Zhou, H.; Fang, X.; Fang, Y.; Wu, J., Computer prediction of paratope on antithrombotic antibody 10B12 and epitope on platelet glycoprotein VI via molecular dynamics simulation. *Biomed. Eng. Online* **2016**, *15* (2), 647-658.

26. Venkatramani, L.; Bochkareva, E.; Lee, J. T.; Gulati, U.; Laver, W. G.; Bochkarev, A.; Air, G.

M., An epidemiologically significant epitope of a 1998 human influenza virus neuraminidase forms a highly hydrated interface in the NA–antibody complex. *J. Mol. Biol.* **2006**, *356* (3), 651-663.

27. Lee, J. T.; Air, G. M., Interaction between a 1998 human influenza virus N2 neuraminidase and monoclonal antibody Mem5. *Virology* **2006**, *345* (2), 424-433.

28. Gulati, U.; Hwang, C.-C.; Venkatramani, L.; Gulati, S.; Stray, S. J.; Lee, J. T.; Laver, W. G.; Bochkarev, A.; Zlotnick, A.; Air, G. M., Antibody epitopes on the neuraminidase of a recent H3N2 influenza virus (A/Memphis/31/98). *J. Virol.* **2002**, *76* (23), 12274-12280.

29. Padlan, E. A.; Silverton, E. W.; Sheriff, S.; Cohen, G. H.; Smith-Gill, S. J.; Davies, D. R., Structure of an antibody-antigen complex: crystal structure of the HyHEL-10 Fab-lysozyme complex. *Proc. Natl. Acad. Sci. U.S.A* **1989**, *86* (15), 5938-5942.

30. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E., GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435-447.

31. Fujitani, H.; Matsuura, A.; Sakai, S.; Sato, H.; Tanida, Y., High-level ab initio calculations to improve protein backbone dihedral parameters. *J. Chem. Theory Comput.* **2009**, *5* (4), 1155-1165.

32. Tesfatsion, L., Agent-based computational economics: A constructive approach to economic theory. *Handbook of computational economics* **2006**, *2*, 831-880.

33. Murakami, S.; Matsumoto, R.; Kanamori, T., Constructive approach for synthesis of a functional IgG using a reconstituted cell-free protein synthesis system. *Sci. Rep.* **2019**, *9* (1), 1-13.

34. Kaneko, K.; Tsuda, I., *Complex systems: chaos and beyond: a constructive approach with applications in life sciences*. Springer Science & Business Media: 2011.

35. Kaneko, K., Complex systems biology. *Genome Informatics* **2004**, *15* (2), 302-303.

36. Hikiri, S.; Yoshidome, T.; Ikeguchi, M., Computational methods for configurational entropy using internal and Cartesian coordinates. *J. Chem. Theory Comput.* **2016**, *12* (12), 5990-6000.

37. Sinha, N.; Mohan, S.; Lipschultz, C. A.; Smith-Gill, S. J., Differences in electrostatic properties at antibody–antigen binding sites: implications for specificity and cross-reactivity. *Biophys. J.* **2002**, *83* (6), 2946-2968.

3 機械学習に基づく Na-Cl イオン対の挙動解析

3.1 序論

生体中で、静電相互作用は塩橋や水素結合といった形で影響を及ぼしている。塩橋や水素結合は抗原-抗体間の界面にも多々存在する¹⁻²。Ala 置換などを用いた研究により、これらの静電相互作用を基とした構造が抗原-抗体間の親和性に大きな影響を与えていることが分かっている³⁻⁴。前章では HEL-HyHEL-10 界面に存在する塩橋について述べたが、これらの塩橋は生体中では形成と解離を繰り返していることが分かった。

水中における Na-Cl 対は、イオン相互作用について考える上で、最も単純なモデルの一つである。Na-Cl 間距離を反応座標として自由エネルギー曲線を描くと会合状態と解離状態を分ける鞍点が存在することが分かる。しかしながら、近年、Na-Cl 対周囲の水分子のイオン相互作用に及ぼす影響が大きく、単純なイオン間距離は反応座標として最適でないことが分かってきた。例えば Geissler と Dellago と Chandler は⁵、水中における Na-Cl イオン対の解離ダイナミクスに対して、溶媒、すなわち水分子の動きが重要な役割を果たすと述べている。彼らは、水中での Na-Cl 対について、多数の構造サンプリングを行った。この各構造について、その構造を初期構造として運動量をランダムで割り振った、シミュレーション時間が 150 fs の複数の分子動力学 (MD) シミュレーションを行った。このシミュレーションの結果、会合状態に向かうトラジェクトリーと解離状態に向かうトラジェクトリーの割合が同じである構造を、遷移状態上の構造として定義した。得られた遷移状態の構造から、Na-Cl 解離反応の遷移状態の Na-Cl 間距離上での分布を調べると、自由エネルギー曲線上の鞍点を中心にして、幅広い分布をしていることが明らかになった。これにより、Na-Cl 解離の真の反応座標に

は、Na-Cl 間距離と直交する方向の座標成分が含まれていることが示されたことになる。さらに、Na-Cl 対の解離は Na⁺イオンの第一水和圏に 6 番目の水分子が追加されることで起こり、遷移状態ではこの水分子が入るための隙間が作成されるというメカニズムが提唱された。

Mullen と Shea と Peters もまた⁶、Na-Cl 対の解離ダイナミクスの記述において、溶媒座標が重要であると指摘している。彼らは、水中での Na-Cl 対の解離について、構造サンプリングを行った。次に、得られた各構造について、20 本のトラジェクトリーを使用して解離する確率 p_B を計算した。さらに、最尤法を用いて、 p_B を予測するのに最適な反応座標を導いた。その反応座標には、イオン座標以外に溶媒に関する座標 (Na-Cl 間の水分子の密度や水の配位数) が含まれていた。

Yonetani は⁷、水中での Na-Cl 対の解離反応について、Na⁺イオンと Cl⁻イオンそれぞれに対する水分子の配位数と、Na⁺イオンと Cl⁻イオンを橋渡しする水分子の存在の重要性を示した。ここでは、Na⁺イオンへの水の配位数を以下のよう
に定義している。

$$n_{Na} = \sum_{i \in \omega} f_i^{Na}$$

$$f_i^{Na} = \frac{1 - \tanh [a(R_i^{Na-O} - b)]}{2}$$

ここで、 i は水分子のインデックスで、 R_i^{Na-O} は Na イオンと水分子の O 原子の距離である。また、 b はイオン-水分子間距離で水和しているかどうかの判定基準、 a は f_i^{Na} がどれだけ急激に変化するかを制御するパラメータである。同様に、Cl⁻イオンへの水の配位数は

$$n_{Cl} = \sum_{i \in \omega} f_i^{Cl}$$

$$f_i^{cl} = \frac{1 - \tanh [a(R_i^{cl-H} - b)]}{2}$$

ここで、 R_i^{cl-O} はClイオンと水のH原子の距離で、残りのパラメータはNaイオンの場合と同様である。このように、水分子の配位数は連続値として定義している。Na⁺イオンとClイオンを橋渡しする水分子が存在するかどうか(ω_B)については、

$$\omega_B = \max M_i$$

$$M_i = \min (f_i^{Na}, f_i^{Cl})$$

というように定義している。すなわち、水分子 i が Na⁺イオンと Clイオンに同時に配位している場合のみ、 f_i^{Na} および f_i^{Cl} の値がどちらも大きくなり (1 に近くなり)、その結果、 M_i も大きく 1 に近い値になる。したがって、 ω_B も 1 に近い大きな値を取ることになる。 f_i^{Na} および f_i^{Cl} は共に連続的であるので、橋渡しする水分子の存在を示す変数 ω_B も連続的である。

上記のように定義された座標の重要度を調べるために、まず、MD シミュレーションを使用し、イオン間距離と Na イオンに対する水分子の配位数 (n_{Na}) を座標軸に取った 2 次元自由エネルギー面、およびイオン間距離と Cl イオンに対する水分子の配位数 (n_{Cl}) を座標軸に取った 2 次元自由エネルギー面を計算した。その結果、Na イオンに対する水分子の配位数 (n_{Na}) に対しては、会合状態に対応する井戸が自由エネルギー曲面上に 2 つ見られた ($n_{Na} = 4, 5$) が、解離反応経路に直接的に接続しているのは $n_{Na} = 5$ の方のみであった。同様に、Cl イオンに対する水分子の配位数 (n_{Cl}) に対しては、会合状態に対応する井戸は自由エネルギー曲面上に 2 つ見られた ($n_{Cl} = 5, 6$) が、遷移状態へとつながる解離反応経路は $n_{Cl} = 6$ の方にのみ存在していた。したがって、Na-Cl 対の水和状態の変化が解離を始めるきっかけであると考えることができる。

さらに、Yonetani⁷はNa⁺イオンとCl⁻イオンを橋渡しする水分子が存在することが、Na-Cl 対の解離ダイナミクスにどのように影響しているかを調べるために、イオン間距離と n_{Na} に ω_B を加えた3次元自由エネルギー面を計算している。この自由エネルギー面の、 $\omega_B = 0$ の2次元断面と $\omega_B = 1$ の2次元断面を比較すると、橋渡しする水分子の存在する方($\omega_B = 1$)が解離への障壁が低いことが分かった。同様に、イオン間距離と n_{Cl} に ω_B を加えた3次元自由エネルギー面を計算し、 $\omega_B = 0$ の2次元断面と $\omega_B = 1$ の2次元断面を比較した。こちらの結果も、 n_{Na} についての結果と同様の傾向が見られた。どちらの座標系で見ても、Na⁺イオンとCl⁻イオンを橋渡しする水分子の存在は、Na-Cl 対の解離ダイナミクスに対して、自由エネルギー障壁を低下させることが分かった。

このように、Na-Cl 対の会合/解離プロセスに対して、溶媒座標が重要であることが分かってきた。したがって、多くの溶媒原子の座標から真に重要な座標を選び出すことは、イオンの会合/解離ダイナミクスを理解する上で重要である。しかし現状では、物理化学的な直感によって最初に選ばれた溶媒座標の範囲でしか検討されておらず、系統的に原子座標の重要性を評価し選び出すことはなされていない。本研究では、どの溶媒原子座標がNa-Cl 対の会合/解離プロセスにおいて、どの程度重要なのかを系統的・直接的に評価する手法を開発することを目標にする。

本研究では、個々の溶媒原子座標がどの程度Na-Cl 対の会合/解離に影響を及ぼすのかを見出すために、近年注目を集めている機械学習的な方法を適用することを考える。機械学習は、与えられたデータから系統的にルールを構築し、未知のデータを予測することを可能にする手法である。Na-Cl 対の会合/解離の問題に対しても、機械学習法を用いることで、MD シミュレーションのデータか

ら直接的に会合/解離を制御している座標を取り出せる可能性がある。

3.2 方法

3.2.1 MD シミュレーションの条件

本研究では、MD シミュレーションは *GROMACS 4.6.7* を用いて行った。また、一辺が約 25 Å の立方体周期境界セルに Na⁺イオンと Cl⁻イオンがそれぞれ 1 個ずつと水分子が 510 個入った系を対象として取り扱った。力場は OPLS 力場を用い、水は TIP3P を使用した。シミュレーション条件は 1 ステップの時間刻みを 3 fs とし、298 K, 1 atm の *NpT* アンサンブルとした。温度と圧力の調整には Nose-Hoover および Berendsen 法を用いた。長距離静電相互作用の計算には Particle-Mesh Ewald (PME)法を用い、カットオフ距離は 9 Å とした。上記のシミュレーション条件は、この章で書かれている全ての MD シミュレーションについて、共通して用いている。

3.2.2 Na-Cl 対の遷移状態の位置の特定

はじめに、Na-Cl 対会合/解離プロセスにおける Na-Cl 間距離の鞍点がどこにあるのかを調べるため、Na-Cl 対の系について 600 ns の MD シミュレーションを行った。このシミュレーション中での Na-Cl 間距離の分布をとった。次に、Na-Cl 間距離を 2.50, 3.00, 3.50, 4.00, 4.50, 5.00, 5.50, 6.00 Å で、バネ定数をそれぞれ 4000 kJ/mol nm² で拘束をかけ、また、Na-Cl 間距離 3.70 Å で 8000 kJ/mol nm² で拘束をかけてそれぞれ 120 ns の MD シミュレーションを行い、アンブレラサンプリングを行った。このシミュレーション結果より、*GROMACS 4.6.7* の *g_wham* プログラムを用いることで、Na-Cl 間距離に対する平均力ポテンシャル

(PMF)曲線を作成した (図 3.1)。この結果、Na-Cl 間距離に対する PMF の鞍点は 3.70 Å であると分かった。

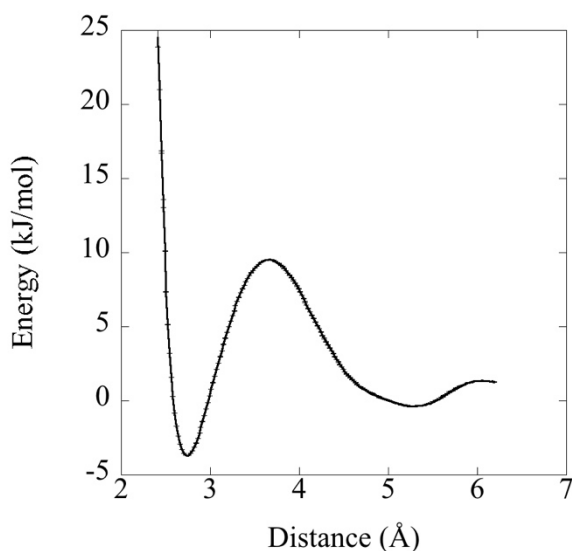


図 3.1 : Na-Cl 間距離に対する自由エネルギー曲線。横軸は Na-Cl 間距離で、縦軸は平均力ポテンシャル。

3.2.3 初期構造の作成

続いて、鞍点付近での系の構造を得るため、Na-Cl 間距離 3.70 Å でバネ定数 8000 kJ/molnm² の拘束をかけて行った MD シミュレーションから、Na-Cl 間距離が 3.70 Å に近い時刻の構造をサンプリングした。これらの構造に対し、Na イオンおよび Cl イオンの座標に束縛をかけて MD シミュレーションを行った。300 ps ごとに座標および速度の出力を行い、その中から Na-Cl 間距離が 3.65 ~ 3.75 Å である時刻の座標および速度を初期構造として取得した。これによって得られた点は 19 万点であり、以後、機械学習を行う際にはテストデータとして扱った。また、座標および速度の出力を 30 ps ごとにして同様に初期構造を作った点が 100 万点あり、こちらはトレーニングデータとして利用した。

3.2.4 機械学習に用いたデータ

得られた初期構造・初期速度に対し、束縛をかけずに 150 fs の MD シミュレーションを行った。シミュレーションの後、150 fs の時点で Na-Cl 間距離が 3.70 Å 以下ならば会合(True)とし、逆に 3.70 Å より大きければ解離(False)とした。この結果を機械学習の出力とし、以後、二値分類でこれを予測できるかどうかを議論した。

また、特徴量として用いたデータは初期構造の状態でのデータを用いた。特徴量としてはまず、Na-Cl 間の距離を用いた。次に、Na-Cl 対から最も近い水分子から 5 番目までに近い水分子について、その水分子の O 原子から Na イオンおよび Cl イオンまでの距離をそれぞれ特徴量として用いた。水分子の Na-Cl 対からの距離については、Na イオンから O 原子までの距離と Cl イオンから O 原子までの距離の和で定義した。さらに、Cl イオンの Na イオンに対する相対速度について、その Na イオン方向の成分を用いた。ここで、相対速度の動径方向成分 (v_{rad}) とは、Cl イオンの Na イオンに対する相対速度のベクトルを計算し、Cl イオンの Na イオン方向への単位ベクトルと内積を取ることで定義した。

$$v_{rad} = \vec{v} \cdot \frac{\vec{r}}{|\vec{r}|}$$

ここで、 \vec{r} は Na イオンを基準とした Cl イオンの位置ベクトルであり、 \vec{v} は Na イオンに対する Cl イオンの相対速度ベクトルである。同様に、Na-Cl 対から 5 番目までに近い水分子の O 原子の速度について、Na イオン方向と Cl イオン方向の速度成分を計算してそれぞれ特徴量とした。また、水の H 原子の Na イオンおよび Cl イオンからの距離と速度についても同様に計算し、特徴量とした。また、

各特徴量については、データの平均値が 0、分散が 1 となるように、標準化している。

特徴量データセットは、Na-Cl 間の距離および Cl イオンの Na イオンに対する速度の 2 次元のデータセットを作成した。次に、このデータセットに Na-Cl 対から最も近い水分子の Na イオン-O 原子間および Cl イオン-O 原子間距離を含めた 4 次元の特徴量データセットを作成した。さらに、この 4 次元のデータセットに Na-Cl 対から 2 番目に近い水分子の Na イオン-O 原子間および Cl イオン-O 原子間距離を含めた特徴量データセットを作った。以下同様に、Na-Cl 対から 5 番目に近い水分子についてまで、1 組ずつ含める特徴量を増やしたデータセットを作成した。また、水分子に関する特徴量について、Na イオン-O 原子間および Cl イオン-O 原子間距離だけでなく、O 原子の Na イオンおよび Cl イオン方向への速度について含めたデータセットも、同様に作成した。

3.2.5 機械学習手法

機械学習の手法には教師あり学習と教師なし学習がある。教師あり学習とは、正解値を持つトレーニングデータに基づいて未知のデータを予測する手法である。それに対して、教師なし学習は、与えられたデータのクラスタリングなどにより、既知の正解データを使わずデータから情報を取り出す手法である。また、教師あり学習には回帰問題と分類問題が存在する。回帰問題は連続的な値を予測する問題である。分類問題は、与えられた特徴量データから、複数の離散カテゴリの一つに分類する問題である。分類問題を扱う機械学習手法には k 近傍法やサポートベクターマシン、決定木、ランダムフォレスト、ニューラルネットワークなどの多くの手法がある。

本研究では、Na-Cl 対が会合するか解離するかの二値分類問題を取り扱う。二値分類問題を扱う機械学習では、与えられたデータを特徴量により分類する。特徴量とは、インプットとして扱う情報（原子の座標など）のことである。それに対して予測結果とは、Na-Cl 対が会合するか解離するかなどの、与えられた特徴量から予測した情報のことである。機械学習の基本的なアルゴリズムは、与えられた特徴量 x から結果 y を出力する関数 $y(x)$ としてモデルを構築する。このモデルの関数 $y(x)$ の種々のパラメータを、トレーニングデータを用いた学習によって改善していくという方法をとる。どのようなモデル（関数）を構築するかは、使用する機械学習手法や、与えるハイパーパラメータによって決まる。

(a) k 近傍法

k 近傍法では、シンプルに未知の点は近傍の既知のデータ点と同様の特徴を持つと推定することで分類を実施する。データ点間の距離は、特徴量空間上での距離（一般的にはユークリッド距離を用いる）として計算される。したがって、この手法は前もってモデルのパラメータを最適化する過程（トレーニング）を含まない手法となる。未知のデータに対する予測では、データ点に含まれるノイズの影響を抑えるために、最近傍の k 個のデータの多数決により分類を判断する。 k はモデルの性質を決めるハイパーパラメータである。 k が小さい場合、局所的で小さな分類領域を表現することが可能であるが、例外的な点やノイズの影響による誤判定をする可能性が高くなる。逆に k が大きい場合、例外的な点やノイズの影響による誤判定の可能性は小さくなるが、データ周囲の平均的な様相に予測が左右されるため、細かい構造を持つ分類境界を表現できなくなる可能性がある。また、各特徴量のスケールが異なると、スケールの大きな特徴量のみ

でデータ間の距離が決まってしまうため、各特徴量を標準化することが重要である。

k 近傍法は単純であるが、参照点となる既知のデータが特徴量空間上に十分に分布していれば高い正答率が期待できる。しかしながら、参照データが十分に存在しない場合、予測を行いたいデータと関連の薄いデータと比較する可能性が高くなり、正答率は下がる。また、この手法のモデルは与えられたデータセットを参照点としてそのまま用いるため、参照データが増えるとモデルのデータ容量も増える。またそのほかにも、未知のデータが与えられるたびに、その都度各参照データとの距離を計算するため、データが増えると計算量も膨大なものとなる。

本研究では、18 万点のデータを参照用データとして、1 万点のテストデータに対して k 近傍法による予測を行った。予測を行う際に、テストデータの点からの距離などによって重み付けはせずに多数決を行なっている。テストデータから何番目に近い参照データまでを予測に用いるかについては、その値を変数 k として定義した。

k の値を決定するために、Na-Cl 対から 3 番目までに近い水分子に関する特徴量までを含めたデータセットに対し、 k の値の変化に対する正解率の変化を調べた。 k の値は 1 から 9 まで調べた。また、多数決を取る際に会合と解離のデータが同数とならないように、奇数の値のみを調べた。詳細な結果は 3.3.1(a) で述べるが、 $k = 9$ での精度が最も高かったので、以下の k 近傍法による予測では $k = 9$ を採用した。

次に、どれくらい遠くの水分子までが Na-Cl 対の会合/解離ダイナミクスに有効な影響を及ぼしているかを調べるために、会合/解離ダイナミクス予測の正

解率を計算した。最初は Na-Cl 間距離のみで予測をし、その後 Na-Cl 対に近い水分子に関する座標を特徴量として加えながら予測をする。追加した座標によって正解率が向上すれば、その座標は会合/解離ダイナミクスを制御する重要な因子であると考えることができる。

(b) 決定木分析

決定木 (分析) は与えられたデータを、ある特徴量が閾値以上かどうかなどの条件で分割し、それを段階的に繰り返してデータを各クラスに分類する手法である。学習において各分割の条件を設定され、この段階的な分割を行う予測器を決定木と呼ぶ。決定木の特徴としては、データの分割のされ方が直接的であるため、他の機械学習手法と比べて解釈が容易である。しかしながら、データを際限なく分割してしまうと、トレーニングデータに対して過学習を起こしやすい。過学習に対する対策としては、(1)決定木の深さに上限を設ける、(2)データが一定以上細かくなったら分割を終了する、(3)もしくは下記のランダムフォレストのようにアンサンブル学習を行うなどの方法がある。

(c) ランダムフォレスト

ランダムフォレストは、複数の決定木を用いたアンサンブル学習である。各決定木で独立して予測を行い、分類問題の場合、多数決で多くの予測器が予測した結果を、最終的な予測として出力する。

ランダムフォレストの学習を行うには、まず、トレーニングデータセットからブートストラップサンプリングにより複数のデータセットを作成する。ブートストラップサンプリングとは、母集団から重複を許してランダムに複数のデ

ータを抽出し、新たなデータセットとする手法である。次に、作成した各データセットに対してそれぞれ決定木を作成する。ただし、決定木作成におけるデータの分割で、分割毎に全ての特徴量の中からランダムに選び出し、それらの特徴量を分割条件の候補としている。このプロセスによって、作成された決定木同士が似たような分類器とならないように調整しアンサンブル学習の有効性を高めている。

本研究では、ブートストラップサンプリングにより新たなデータセットを作成する際、元のトレーニングデータと同じ大きさのデータセットを作成した。決定木の各分割において、ランダムに選び出す特徴量の個数は、全特徴量の個数の平方根を超えない自然数とした。また、ランダムフォレストにおいては、データを標準化しても結果は変わらないが、他の機械学習手法との比較のため、学習と予測の前にデータの標準化をした。

決定木の本数と許容する深さはモデルの精度や学習の効率を決める重要なパラメータである。これらを適切に決めるために、各パラメータの変化に対する正解率の変化を調べた。対象として、Na-Cl 対から 3 番目までに近い水分子に関する特徴量までを含めたデータセットを用意し、これに対する予測性能からパラメータの値を決定した。

次に、Na-Cl 対から何番目に遠い水分子までが Na-Cl 対の会合/解離ダイナミクスに影響を及ぼしているかを調べた。特徴量として考慮する水分子の座標を、Na-Cl 対から近い順に水分子 1 個分ずつ増やしながらかそれぞれ予測を行い、正解率を計算した。加えた水分子の座標に対する正解率の変化から、Na-Cl 対の会合/解離ダイナミクスに影響を及ぼしている水分子を調べた。

さらに、特徴量の重要性を feature importance 法と permutation importance 法

の2つの指標を用いて評価した。feature importance 法では、特徴量の重要性を評価するために、ジニ不純度という指標を導入する。ジニ不純度とは単一のノードに対して定義される値であり、そのノードにおいてデータセットがどれくらい分類できているかを測る指標である。二値分類をしたノード k におけるジニ不純度は、

$$G(k) = \sum_{i=1}^2 p(i) \times (1 - p(i))$$

という式で表される。 i はターゲットラベルである。 $p(i)$ はそのノードに保持されるデータセット中での、ターゲットラベル i のデータの割合である。ジニ不純度は、データセットが一方のターゲットラベルのデータに分類されているほど小さくなる。特徴量の重要度は、ある特徴量で分割した時に、ジニ不純度がどれくらい下がるかで定義される。まず、ノード k が特徴量 j によって分割されている場合、その分割における特徴量 j の重要度を以下の式で計算する。

$$I(j, k) = N_{parent} \times G_{parent} - (N_{left_child} \times G_{left_child} + N_{right_child} \times G_{right_child})$$

ここで、 $parent$ は分割元となるノード、 $left_child$ および $right_child$ は分割によってできた子ノードの左側と右側を表す。 N はノード内でのデータセットのデータ数、 G はそのノードのジニ不純度を表す。すなわち、各項ではデータ数によって重み付けされたジニ不純度を計算しており、 $I(j, k)$ は分割前と分割後の重み付けされたジニ不純度の差となっている。特徴量 j の重要度は、特徴量 j によって行われる全ての分割についての、 $I(j, k)$ の和となる。

$$I(j) = \sum_{k \in F(j)} I(j, k)$$

$F(j)$ は特徴量 j によって分割されるノードの集合である。この重要度 $I(j)$ を正規化したものを特徴量重要度 (feature importance) と呼び、決定木やランダムフォレスト

ストの学習における特徴量の重要度を表すものとして使用する。

次に、permutation importance 法は⁸、重要度を調べたい特徴量の値をランダムに並び換えることで、その特徴量の重要度を評価する方法である。こちらの方法では、あるデータセットに対して学習済みのモデルにより正解率を計算し、重要度の計算に用いる。使用するデータセットはトレーニングデータセットに限らず、任意に設定できる。この方法での特徴量 j に対する重要度の計算は、以下の式のように行う。

$$I(j) = s - \frac{1}{K} \sum_{k=1}^K s_{k,j}$$

s は与えられたデータセットに対する正解率である。 $s_{k,j}$ は、与えられたデータセットの特徴量 j の値をランダムに並び替えたデータセットに対する正解率である。データの並び替えは複数回にわたって行っており、 K はその回数である。特徴量 j の値をランダムに並び替えた際の正解率の減少値が、その特徴量の重要度として定義される。この permutation importance 法は決定木やランダムフォレスト以外の機械学習法に対しても適用できるし、学習過程ではなく予測過程に対しても適用できる。

本研究では、特徴量の重要性の計算は、4つのモデルに対して行った。1つめは、Na-Cl間の距離とClイオンの速度、およびNa-Cl対から5番目までに近い水分子のO原子のNaイオンおよびClイオンとの距離とNaイオンおよびClイオン方向への速度について特徴量に含めたモデルである。2つめは、Na-Cl間の距離とClイオンの速度、およびNa-Cl対から3番目までに近い水分子のO原子のNaイオンおよびClイオンとの距離とNaイオンおよびClイオン方向への速度、それらの水分子同士のO原子間距離について特徴量に含めたモデルであ

る。3つめは、Na-Cl間の距離とClイオンの速度、およびNa-Cl対から5番目までに近い水分子のO原子と2つのH原子についてのNaイオンおよびClイオンとの距離について特徴量に含めたモデルである。4つめは、3つ目のモデルの特徴量に、水分子のO原子およびH原子のNaイオンおよびClイオン方向への速度について加えたモデルである。

(d) サポートベクターマシンによる分類

サポートベクターマシン (SVM) は、特徴量空間上でデータの分類境界

$$y = w\phi(x) + b$$

を決定する。ここで、 $\phi(x)$ は特徴量空間変換関数と呼ばれるもので、非線形な分類境界を擬似的に線形な分類境界に変換するものである。また、 w は重みパラメータであり、 b はバイアスと呼ばれる。これらのパラメータは、マージンを最大化するように決定される。マージンは分類境界と最も近くのデータ点までの距離として定義される。すなわち、SVMは分類境界とデータ間の距離を大きくすることで、分類の予測精度を高める戦略を取っている。しかし、ノイズに当たるデータまで分類するように学習を行ってしまうと、汎化性能が低くなってしまふ。そのため、ある程度の誤分類を許容する必要がある(ソフトマージン)。SVMでは、正則化係数 C を導入することで誤分類の許容度を制御している。 C が大きいほど誤分類を許容せず、 $C \rightarrow \infty$ の極限では全てのデータを完全に分類する。

SVMは、データの次元が増加しても分類精度が高く、最適化すべきパラメータが少ないという利点がある。しかしながら、トレーニングデータが増えると計算コストが膨大になるという問題点がある。

(e) ニューラルネットワーク

ニューラルネットワークは、複数のパーセプトロンが多層に組み合わさった構造をしている。多層に組み合わさったパーセプトロンの層のうち外部からの入力を受け取る層を入力層と呼び、最終的な出力を計算する層を出力層と呼ぶ。入力層と出力層の間の層は、中間層もしくは隠れ層と呼ばれる。k番目のパーセプトロンが単純に0か1の信号を出力する場合、このパーセプトロンにおける出力 y_k は以下のように計算される。

$$y_k = h(a_k) = \begin{cases} 0 & (a_k \leq 0) \\ 1 & (a_k > 0) \end{cases}$$
$$a_k = b_k + \sum_i w_{ik} x_i$$

ここで、 $x_i (i = 0, 1, \dots, n)$ はパーセプトロンへの入力値である。 $w_{ik} (i = 0, 1, \dots, n)$ は重みと呼ばれ、 b_k はバイアスと呼ばれる。このように、入力値の線型結合として a_k を計算し、一般に活性化関数と呼ばれる関数 $h(a_k)$ を用いて出力 y_k に変換する

活性化関数の性質は学習結果に大きな影響を与える。一般的に、ニューラルネットワークでは連続値を出力するシグモイド関数や ReLU 関数が活性化関数としてよく用いられる。シグモイド関数は

$$h_{sig}(a_k) = \frac{1}{1 + \exp(-a_k)}$$

ReLU 関数は

$$h_{ReLU}(a_k) = \begin{cases} 0 & (a_k \leq 0) \\ a_k & (a_k > 0) \end{cases}$$

と定義される。

最終的な出力は直接的な分類の予測結果であるため、出力層では隠れ層と異なる活性化関数を使用する。一般的に、出力層の活性化関数としてソフトマッ

クス関数を用いられる。ソフトマックス関数による出力は、以下のように表される。

$$y_k = h_{softmax}(a_1, a_2, \dots, a_n) = \frac{\exp(a_k)}{\sum_{i=1}^n \exp(a_i)}$$

ここでは、出力層のノード数は n 個としている。この出力 y_k の値は 0 から 1 の間であり出力層の全ノードからの出力値の総和は 1 となる。したがって、ソフトマックス関数による出力結果を確率として解釈することができる。最も確率の高い分類の結果を、最終的な結果として採用する。ニューラルネットワークの学習を行う上では、このように確率として結果を出力することが必要であるが、予測においては結果を確率として出力する必要はない。また、ソフトマックス関数による値の変換は大小関係には影響を及ぼさない。よって、実際には、出力層では計算コストの高いソフトマックス関数は省略し、 a_k を参照して予測を行う。

ニューラルネットワークの学習には損失関数を用いる。一般的に用いられる損失関数には、二乗和誤差

$$E = \frac{1}{2} \sum_k (y_k - t_k)^2$$

や、交差エントロピー誤差

$$E = - \sum_k t_k \log y_k$$

がある。ここで、 t_k は正解ラベルである。分類問題の場合、 y_k や t_k は確率として表される。 t_k の値は、正解のクラスの確率は 1、外れのクラスの確率は 0 として定義される。これらの損失関数をモデル性能の指標として、最小化するように最適な重みパラメータを探索する。

ニューラルネットワークモデルにおいて、損失関数を最小化するアルゴリズムは重要である。一般的に、ニューラルネットワークモデルのパラメータ空間

における損失関数の形状は複雑であり、大域的な最小値を見いだすのは難しい。アルゴリズムの選択を誤ると、パラメータ更新の手法によっては学習が収束しなかったり、大幅な時間がかかったりすることもある。そこで、パラメータを効率的に最適化するための様々なアルゴリズムが提案されてきた。

例えば、パラメータを更新するための基本的なアルゴリズムとして、最急降下法がある。この方法は各パラメータに対する損失関数の勾配（微分）を用い、勾配方向にパラメータを更新するというステップを繰り返すアルゴリズムである。

$$\mathbf{W} \leftarrow \mathbf{W} - \eta \frac{\partial L}{\partial \mathbf{W}}$$

その式は上記のように表される。 \mathbf{W} はパラメータ w_{ik} を要素に持つ行列であり、 L は損失関数、 η は学習係数である。しかしながらこの方法では、損失関数の極小値にパラメータが更新された場合、それ以上学習が進まないという問題点がある。

この問題点を解決するアルゴリズムとして、確率的勾配降下 (SGD) 法が提案された。SGD 法は、ランダムに並べかえたトレーニングデータを頭から1つ選び、その1点に対して計算した勾配を $\frac{\partial L}{\partial \mathbf{W}}$ の近似としてパラメータ更新をする方法である。逐次的に全てのトレーニングデータに対してこのパラメータの更新を行う。(全てのトレーニングデータに対して1回ずつ更新を行った回数を、エポックという単位で表現する。) トレーニングデータは、1エポック毎にランダムに並べかえる。ランダムにデータの順番が入れ替わるので、極小値に囚われにくくなると考えられる。SGD 法にはこのような利点がある一方で、 η が大きすぎると最小値を飛び越して何度も無駄な更新を行ってしまう欠点がある。この欠点について改善したアルゴリズムとして、Momentum 法や AdaGrad 法がある。

Momentum 法は過去の勾配の変動を考慮することで、SGD 法における振動

的な更新を抑える。具体的には、

$$\mathbf{v} \leftarrow \alpha \mathbf{v} - \eta \frac{\partial L}{\partial \mathbf{W}}$$

$$\mathbf{W} \leftarrow \mathbf{W} + \mathbf{v}$$

というように、変数 \mathbf{v} を導入し、これをパラメータに足すことで、パラメータの更新を行う。 \mathbf{v} は勾配 $\frac{\partial L}{\partial \mathbf{W}}$ についての指数平滑移動平均となっており、 α は0から1の値を取るハイパーパラメータである。 $\alpha = 0$ の場合、 \mathbf{v} は現時点での勾配のみを扱っており、SGDと一致する。 α が大きくなると、ノイズ的な勾配の振動が抑えられ、より滑らかに更新が行われる。

AdaGrad法は、学習率を調整することで、SGD法の振動的な更新を抑える。学習が進み最小値に近づくと、最小値に至るために必要なパラメータの変化量は小さくなる。そこで、学習が進むにつれて、学習係数を小さくするという方法をとる。AdaGrad法では、これを一つ一つのパラメータに対して個別に行う。更新の式は以下のように定義される。

$$\mathbf{h} \leftarrow \mathbf{h} + \frac{\partial L}{\partial \mathbf{W}} \odot \frac{\partial L}{\partial \mathbf{W}}$$

$$\mathbf{W} \leftarrow \mathbf{W} - \eta \frac{1}{\sqrt{\mathbf{h}}} \frac{\partial L}{\partial \mathbf{W}}$$

ここで、 \odot は行列の要素毎にかけ算を行う演算子として用いている。この手法では、更新毎のパラメータに対する勾配 $\frac{\partial L}{\partial \mathbf{W}}$ の累積二乗和となる変数、 \mathbf{h} を導入し、学習係数を \mathbf{h} の平方根で割ることで、各パラメータの学習率を小さくする。ただしこの方法では、 \mathbf{h} は学習が進むに従って大きくなる一方であるため、学習率は小さくなる一方となる。そのため、学習開始直後に大きな勾配となっていた場合、その後ほとんどそのパラメータは更新されないという問題点がある。

そこで、この問題点を修正した RMSprop 法というアルゴリズムが提案され

た。

$$\mathbf{h} \leftarrow \alpha \mathbf{h} + \frac{\partial L}{\partial \mathbf{W}} \odot \frac{\partial L}{\partial \mathbf{W}}$$
$$\mathbf{W} \leftarrow \mathbf{W} - \eta \frac{1}{\sqrt{\mathbf{h}}} \frac{\partial L}{\partial \mathbf{W}}$$

α は0から1の値を取るハイパーパラメータである。この方法では、過去の \mathbf{h} の値に α をかけることで、古い勾配の影響が徐々に減っていくようになっている。これにより、学習開始直後に大きな勾配があっても、学習が進むようにしている。

Adam 法は Momentum 法と RMSprop 法を組み合わせることで、より効率的なパラメータ空間の探索を行う手法である。

$$\mathbf{v} \leftarrow \alpha_1 \mathbf{v} - \frac{\partial L}{\partial \mathbf{W}}$$
$$\mathbf{h} \leftarrow \alpha_2 \mathbf{h} + \frac{\partial L}{\partial \mathbf{W}} \odot \frac{\partial L}{\partial \mathbf{W}}$$
$$\mathbf{W} \leftarrow \mathbf{W} - \eta \frac{1}{\sqrt{\mathbf{h}}} \mathbf{v}$$

α_1 および α_2 はそれぞれ0から1の値を取るハイパーパラメータである。Adam 法は第一式で Momentum 法のように勾配の瞬間的な変化の影響を弱め、第二式で RMSprop 法のように学習率の調整を行う。これにより、勾配のノイズ的な振動を抑えたパラメータの計算を行う。

本研究でニューラルネットワークを使用する際には、活性化関数は ReLU 関数を用いた。また、損失関数は交差エントロピー誤差を用いた。最適化アルゴリズムは Adam を使用した。学習の終了条件として、10 エポックで損失が 10^{-4} 未満である場合、収束しているとみなした。

ニューラルネットワークで用いるハイパーパラメータの値を決定するために、k 近傍法やランダムフォレストと同様に、各ハイパーパラメータの値に対す

る正解率の変化を調べた。調べたハイパーパラメータは、隠れ層の層数および隠れ層 1 層のノード数である。隠れ層 1 層のノード数は、全ての隠れ層に対して同一とした。隠れ層の層数を検証する際には、1 層のノード数は 50 とし、水分子についての特徴量は Na-Cl 対から 3 番目までに近い水分子について、Na イオン-O 原子間および Cl イオン-O 原子間の距離を含めた。隠れ層 1 層のノード数を検証する際には、隠れ層の層数は 2 層とし、水分子の特徴量については、層数の検証と同様とした。計算精度と計算効率を考慮してハイパーパラメータを決定した。(詳細については以下の 3.3.4(a)章で述べる。) 具体的には、隠れ層の層数は 2 層、1 層のノード数は 50 と設定した。

また、3.3.4(b)章で述べるが、ニューラルネットワークを用いて Na-Cl 対の会合/解離の予測を行う場合には、トレーニングデータの数は 18 万点では少なかつたため、100 万点のデータをトレーニングデータとして用意した。テストデータは、19 万点のデータを、1000 ずつに分割し、その平均と標準誤差をとった。

次に、Na-Cl 対から何番目に遠い水分子までが Na-Cl 対の会合/解離ダイナミクスに影響を及ぼしているかを調べた。特徴量として考慮する水分子の座標を増やしながらか予測を行い、正解率を計算した。特徴量として含めた水分子の座標に対する正解率の変化から、Na-Cl 対の会合/解離ダイナミクスに対して重要な水分子を考察した。

続いて、水分子同士の配向の Na-Cl 対の会合/解離ダイナミクスへの影響を調べた。Na-Cl 対から 3 番目までに近い水分子の座標を特徴量に含めて予測を行い、正解率を計算した。さらに、水分子の O 原子間距離を 1~3 つ特徴量として含め、それぞれに対して正解率を計算した。含める O 原子間距離は、Na-Cl 対から 1 番近い水分子と 2 番目に近い水分子、1 番目と 3 番目、2 番目と 3 番目の順

に増やした。これらの正解率の比較により、水分子同士の配向の影響を調べた。

さらに、水素原子の Na-Cl 対の会合/解離ダイナミクスへの影響を調べた。水分子の特徴量を含める際に O 原子だけでなく、2 つの H 原子についても特徴量に追加した。Na-Cl 対の会合/解離ダイナミクスへの影響の調べ方は、H 原子についても特徴量に含めた以外は、水分子の O 原子についてのみ影響を調べた際と同様に行った。

また、予測が当たったデータと外れたデータで、特徴量空間上でどのような違いがあるのかについて検証した。検証方法は、始めに、予測が当たったデータ、及び、予測が外れたデータをピックアップし、それらのデータの元になった初期状態(系の全原子の座標および速度)を取り出した。次に取り出した各初期状態の Na-Cl 対から 3 番目までに近い水分子の座標に対して乱数を加え、元の初期状態から少しだけずれた初期状態を作成した。これらのずれた初期状態から 150 fs の MD シミュレーションを行った。この MD シミュレーションにより、元の状態からのずれの大きさに対して会合状態に向かう割合を調べた。この結果から、元の状態からのずれの大きさと、各初期状態から会合/解離のどちらの状態に向かうかについて関係を考察した。ここで、元の状態からのずれの大きさは、元の状態と修正された状態の各座標成分の平均二乗差分の平方根を取ることで定義した。

3.2.6 モデルの検証

本研究では、複数の機械学習手法でモデルを構築したが、それぞれのモデルの汎化性能を評価するために、交差検証 (クロスバリデーション) を使用した。交差検証は、機械学習のモデルが過学習を起こしていないか、汎化性能を統計的

に評価する一般的な方法である。この検証法では、データを複数に分割し、そのうちの一つのデータセットをテストデータ、残りのデータセットをまとめてトレーニングデータとして扱い、正解率を算出する。分割したデータセットそれぞれが 1 回ずつテストデータなるようにこの方法で学習と正解率の算出を繰り返す。算出された正解率の平均や標準偏差を計算することで、汎化性能を評価する。本研究では、トレーニングデータを 5 つのデータセットに分割して、交差検証を行なった。

3.2.7 機械学習の詳細

本研究では、機械学習は *scikit-learn ver. 0.22.2* を用いて行った。k 近傍法の実装には、`sklearn.neighbors.KNeighborsClassifier` クラスを用いた。ランダムフォレストの実装には、`sklearn.ensemble.RandomForestClassifier` クラスを用いた。ニューラルネットワークの実装には、`sklearn.neural_network.MLPClassifier` クラスを用いた。はじめに、テストデータ用として作成した 19 万点のデータを 18 万点と 1 万点に分割し、それぞれトレーニングデータとテストデータとして使用した。このデータを用いて、k 近傍法などの複数の機械学習手法で Na-Cl 対の会合/解離の予測を行った。

3.3 結果と考察

3.3.1 k 近傍法

(a) ハイパーパラメータの決定

ハイパーパラメータ k を決定するため、 k に対する正解率の変化を調べた (図 3.2)。 $k = 1$ 、すなわち最近傍の点のみから判断した場合、正解率は 0.64 程

度で最も低く、 k を大きくすると正解率が上昇していく。すなわち、特徴量空間におけるより遠くのデータ点までを参照して予測をすることによって正解率が改善されていることが分かる。これは、水分子の速度の情報を特徴量に含めた場合も同様である。

k が小さい時に正解率が低いのは、特徴量空間上での参照データ点の密度が低いことが原因だと考えられる。この場合、未知のデータ点を分類する際に、同じカテゴリーに所属する参照データ点が近くに存在せず、異なるカテゴリーのデータ点が最近接の参照データ点として選ばれてしまう可能性がある。 k を大きくすると、より多くの弱予測器の合議で予測を行うことになるので、誤判定のリスクが下がると考えられる。

また、 k の上昇に伴う正解率の上昇は k が大きくなるにつれて徐々に緩やかになっていく。 $k=9$ で水分子速度データを含む場合において、正解率は0.71程度である。この結果は、 k がある程度大きくなることで、既に十分広域的な様相を見ているためだと考えられる。あまり k を大きくすると、あまりにも広域的な様相を見すぎて、細かい領域が見えなくなる危険性も考えられる。また、境界間際にあるデータ点についても、境界の形状によっては不正解の参照データ点が多くなる可能性もある。よって、 $k=9$ を以下の k 近傍法による予測では用いることにした。

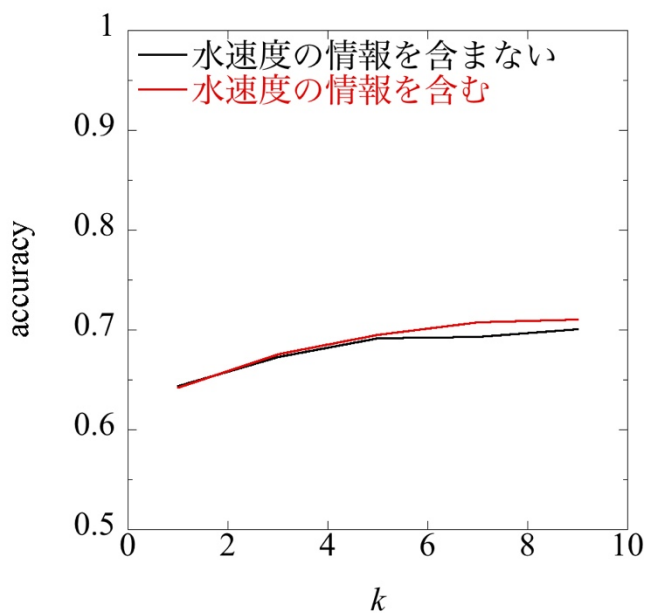


図 3.2 : k 平均法において、ハイパーパラメータ k に対する正解率の変化。横軸が k の値で、縦軸がテストデータに対する正解率である。また、黒線が予測に水分子の O 原子の速度を用いなかった場合で、赤線は用いた場合である。

(b) 反応予測

次に、入力する特徴量で表現されている水分子の個数 n に対する正解率の変化を調べた (図 3.3)。ここでは、Na-Cl 対から最も近い水分子 n 個を選択している。 $n=0$ 、すなわち、Na-Cl 間距離と Na イオンに対する Cl イオンの速度の情報のみを特徴量として含めている場合、正解率は 0.61 程度であった。Na-Cl 間の情報のみからも多少 Na-Cl 対の会合/解離を予測することができたが、あまり正解率は高くなかった。それに水分子 1 個の情報を加えると、正解率がおよそ 0.68 まで向上した。このことから、Na-Cl 対周囲の水分子が Na-Cl 対の会合/解離ダイナミクスに影響を及ぼしていることが推察できる。

しかしながら、 n が 2 から 3 程度まで大きくなると、特徴量に新たな水分子の情報を加えても正解率はほとんど向上しなくなっている。もっとも良い正解

率は 0.71 程度であった。これは、特徴量の次元の増加に対して、特徴量空間内での参照データの数が不足しているために起こっていると考えられる。

さらに水分子の速度の情報を特徴量として考慮し、学習と予測を行なった。すると、その予測の正解率は速度の情報を特徴量として考慮しなかった場合と比べて高くなった。なお、 $n = 0$ の場合は水分子の情報が特徴量に含まれていないので、水分子の速度を入れた場合も入れていない場合も同一となっている。予測の正解率が高くなったことから、水分子の速度が Na-Cl 対の会合/解離ダイナミクスに影響を及ぼしていることが分かった。一方で、 n が大きくなった時に、さらに水分子の情報を特徴量に追加しても正解率があまり向上しなくなる現象は水分子の速度の情報を特徴量に含めた場合の方が顕著だった。こちらの場合、若干 n に対して正解率が減少していることも確認できた。これは、水分子の速度まで特徴量に含めたため、より特徴量の次元が上がったので、参照データ数の不足が際立ったためと考えられる。

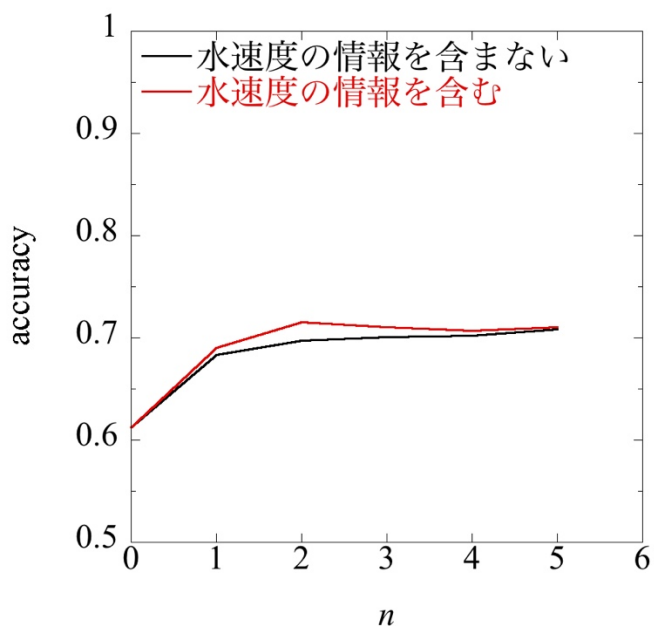


図 3.3: k 平均法において、Na-Cl 対から何番目に近い水分子までを予測に使用したかに対する正解率の変化。横軸は特徴量に加えた水分子の個数 n で、縦軸がテストデータに対する正解率である。また、黒線が予測に水分子の O 原子の速度を用いなかった場合で、赤線は用いた場合である。

3.3.2 ランダムフォレスト

(a) ハイパーパラメータの決定

ランダムフォレストにおいて使用する決定木の本数を決めるために、決定木の本数に対する正解率の変化を調べた (図 3.4)。その結果、決定木の本数は 10 本ではやや正解率が低くなるが、100 本と 1000 本ではほとんど違いは無かった。この結果は、水分子の速度の情報を特徴量に含む場合でも含まない場合でも同じであった。そのため、予測精度を維持しつつ計算コストがなるべく低くなるように、ランダムフォレストで作成する決定木の本数は 100 とした。

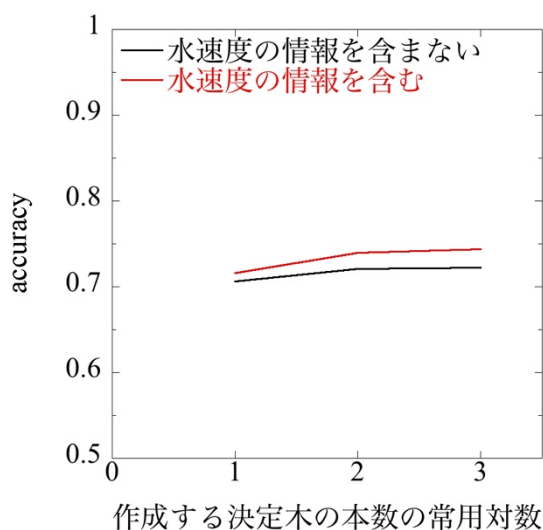


図 3.4: ランダムフォレストにおいて、何本の決定木を構築したかに対する正解率の変化。横軸は決定木の本数の対数で、縦軸がテストデータに対する正解率である。また、黒線が予測に水分子の O 原子の速度を用いなかった場合で、赤線は用いた場合である。

次に、ランダムフォレストの学習を行う際に、各決定木の深さをどこまで許容するかを決定した。そのため、決定木の深さの最大値に対する正解率の変化を調べた (図 3.5)。その結果、水分子の速度の情報を特徴量に含めなかった場合は、決定木の深さの最大値に対して正解率はほとんど変化しなかった。水分子の速度の情報を特徴量に加えた場合、決定木の深さの最大値が 20 で正解率はやや高くなったが、それ以上深さの最大値を大きくしても、正解率はほとんど変化しなかった。よって、高い正解率で計算コストが低くなるように、以下の計算では深さの最大値は 20 とした。

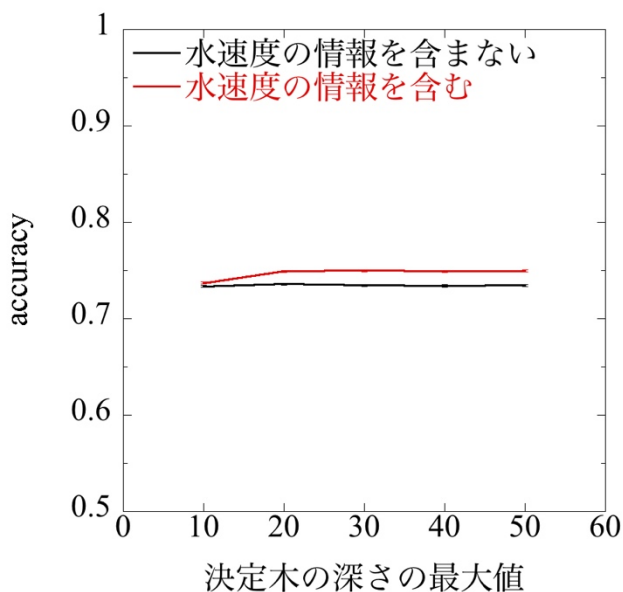


図 3.5: ランダムフォレストにおいて、各決定木の深さをどこまで許すかに対する正解率の変化。横軸は各決定木の深さの許容される最大値。縦軸がテストデータに対する正解率。また、黒線が予測に水分子の O 原子の速度を用いなかった場合で、赤線は用いた場合である。

(b) 反応予測

特徴量に水分子の情報をいくつ含めたかに対する正解率の変化も見たところ (図 3.6)、もっとも良かった正解率は 75 %程度で、 k 近傍法よりも良い予測精度をとっていた。すなわち、使用しているトレーニング (参照) データの数においては、単に特徴量空間上で近い点を見るよりも、良い分類指標が得られていると考えられる。

k 近傍法の結果と同様に、水分子の情報を特徴量に含めていない場合には正解率は 0.6 程度で、水分子の情報を特徴量として加えることで正解率が向上している。従って、Na-Cl 対周囲の水分子の重要性が示されている。また、正解率は $n = 2$ から 3 で変化しなくなっている。そのため、Na-Cl 対から 2 番目と 3 番目

に近い水分子までが Na-Cl 対の会合/解離ダイナミクスに影響を及ぼしており、それ以上遠い水分子はあまり影響を及ぼしていないと考えられる。

さらに、水分子の速度についての情報を特徴量として加えた場合、水分子の座標のみを特徴量に加えた場合よりも若干正解率が上がっていた。どちらの場合においても、Na-Cl 間距離と Cl イオンの速度の情報は特徴量に含めている。そのため、水分子の速度も多少 Na-Cl 対の会合/解離ダイナミクスに影響を及ぼしていると考えられる。ただし、正解率の上昇量から、その影響はあまり大きくないと考えられる。

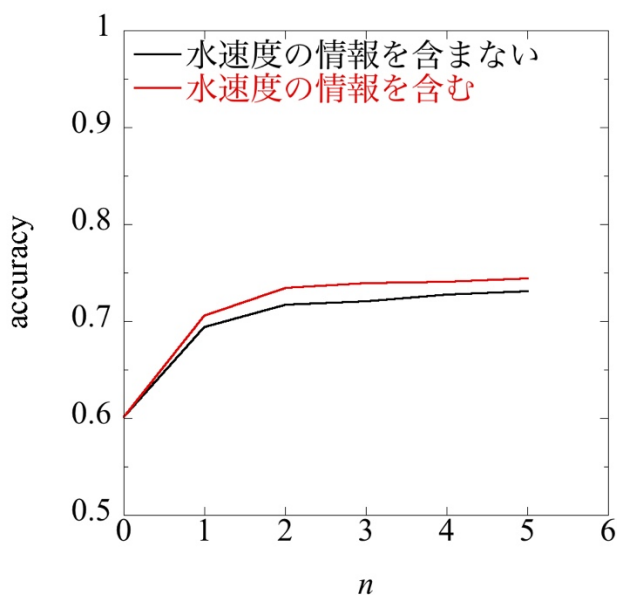


図 3.6 : ランダムフォレストにおいて、Na-Cl 対から何番目に近い水分子までを予測に使用したかに対する正解率の変化。横軸は特徴量に加えた水分子の個数 n で、縦軸がテストデータに対する正解率である。また、黒線が予測に水分子の O 原子の速度を用いなかった場合で、赤線は用いた場合である。

3.3.3 サポートベクターマシン

(a) ハイパーパラメータの決定

サポートベクターマシンでは、正則化定数 C と使用するカーネル関数の広がり指定する γ の二つのハイパーパラメータを設定する。これらのハイパーパラメータの適切な値を検証するため、いくつかの値の組合せにおいて、正解率を計算した (図 3.7)。その結果、水分子の速度を特徴量に含めた場合も含めなかった場合も、 C が 1 かつ γ が 1 の組み合わせが最も正解率が高かった。特に、水分子の速度を特徴量に含めた場合は大きく差がついている。しかしながら、ハイパーパラメータを大きくした際の正解率の低下はトレーニングデータのデータ数が少なすぎるために起こっている可能性がある。したがって、このデータベースを使った学習では、 $C=1$ かつ $\gamma=1$ と設定した。

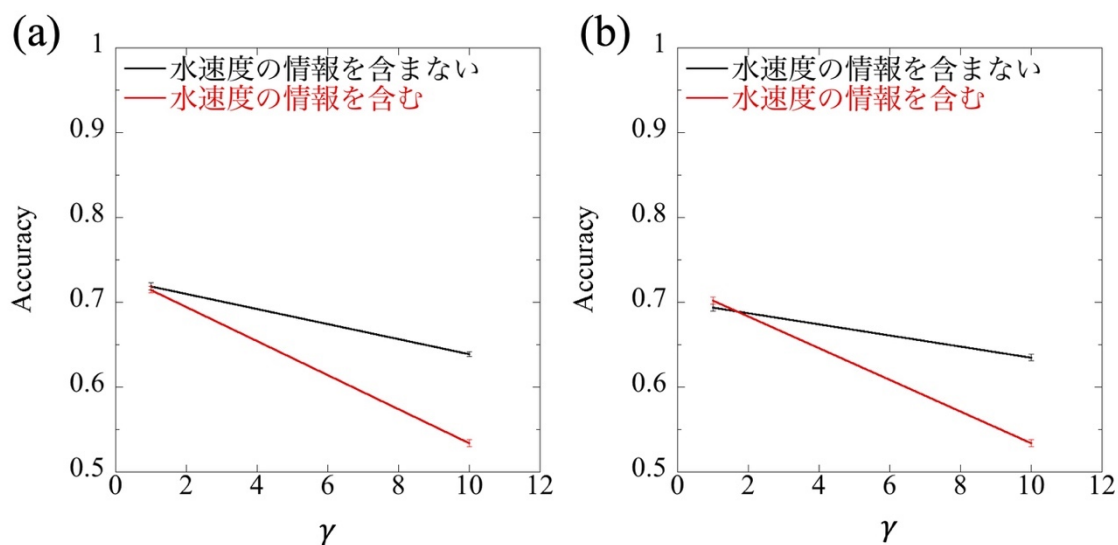


図 3.7: サポートベクターマシンの手法において、ハイパーパラメータ γ に対する正解率の変化。(a)ハイパーパラメータ $C=1$ の場合の正解率の変化。(b) $C=10$ の場合の正解率の変化。また、黒線が予測に水分子の O 原子の速度を用いなかった場合で、赤線は用いた場合である。

(b) 反応予測

次に、特徴量とする水分子の個数を変化させながら予測を行ない、正解率の変化を調べた (図 3.8)。まず、水分子の座標情報のみを特徴量として加えた場合、水分子 1 個分の情報であっても追加することで正解率の上昇が確認された。これは、水分子の会合/解離ダイナミクスへの影響を示唆する結果である。考慮する水分子を 2 個にすると正解率はさらに少し上昇する。しかし、水分子の個数をさらに増やした時の正解率は改善しなかった。さらに、水分子の速度情報も特徴量に含めた場合の予測も行ったが、結果は水分子の座標情報のみを特徴量として加えた場合とほとんど同じであった。この系においては、次元数が高くなるとサポートベクトルマシンでは良い予測モデルを作ることは難しいことを示唆していると考えられる。また、サポートベクターマシンでは、学習に要する時間はトレーニングデータのデータ数 (および特徴量の次元) に対して大きく増加する。よって、学習データを 100 万点に増やしての解析は行わなかった。

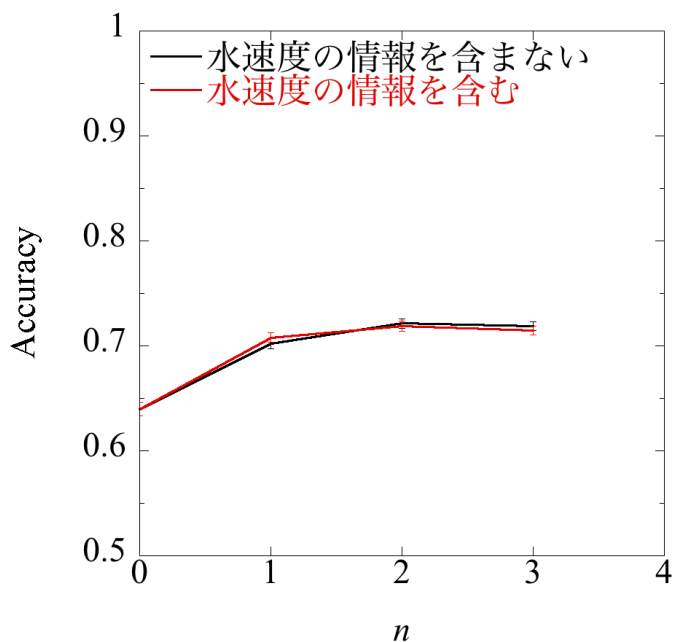


図 3.8 : サポートベクターマシンを用いた機械学習において、Na-Cl 対から何番目に近い水分子までを予測に使用したかに対する正解率の変化。横軸は特徴量に加えた水分子の個数 n で、縦軸がテストデータに対する正解率である。また、黒線が予測に水分子の O 原子の速度を用いなかった場合で、赤線は用いた場合である。

3.3.4 ニューラルネットワーク

(a) ハイパーパラメータの決定

はじめに、18 万点のトレーニングデータを用いて、ニューラルネットワークによる学習と予測を行い、隠れ層の層数に対しての正解率を計算した(図 3.9)。その結果、隠れ層の層数が増えることで正解率は低下していった。この傾向は、水分子の速度の情報まで特徴量に加えた場合の方が顕著だった。これは、特徴量の次元に対してトレーニングデータのデータ数が少ないために起こっていると考える。そのため、ニューラルネットワークを用いて Na-Cl 対の会合/解離の予

測を行うためには、トレーニングデータの数は18万点では不足していることが分かった。

そこで、学習に必要なデータ数を調べるため、トレーニングデータのデータ数に対する正解率の変化を計算した(図3.10)。この結果からは、少なくともデータ数が50万以上の場合、データ数に対して正解率はほとんど変化していなかった。水分子の速度を特徴量としてみた場合でも、これは同じである。そのため、トレーニングデータは100万点あれば、このNa-Cl対の会合/解離の予測には十分であるとした。以下では、100万点のトレーニングデータに対して学習を行った。

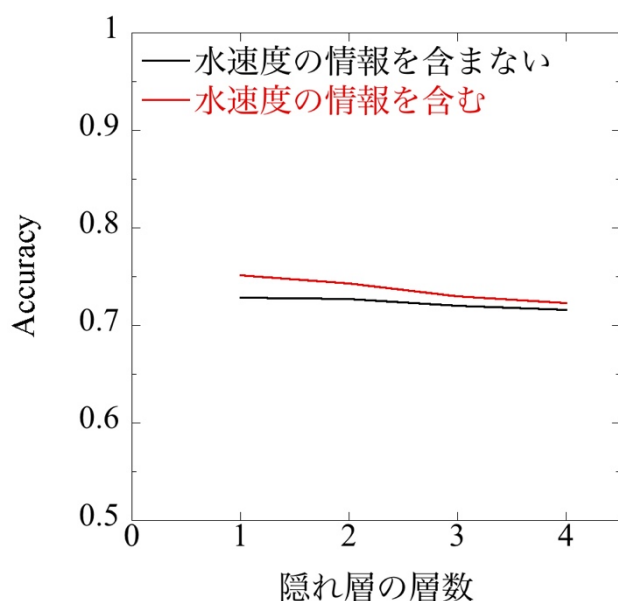


図3.9: ニューラルネットワークにおいて、隠れ層の層数に対する正解率の変化。横軸は隠れ層の層数で、縦軸がテストデータに対する正解率である。また、黒線が予測に水分子のO原子の速度を用いなかった場合で、赤線は用いた場合である。

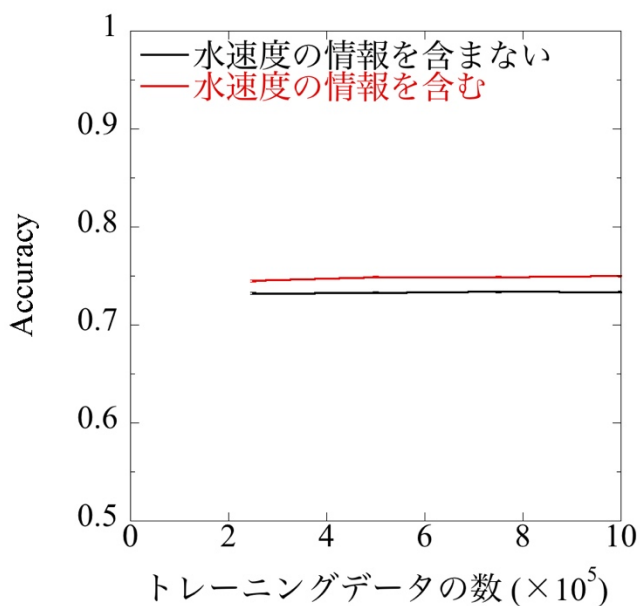


図 3.10 : ニューラルネットワークにおいて、トレーニングデータの点数に対する正解率の変化。横軸は使用したトレーニングデータの点数で、縦軸がテストデータに対する正解率である。隠れ層の深さは 2 層で、隠れ層 1 層のノード数は 50、水分子は Na-Cl 対から 3 番目に近いものまで用いた。また、黒線が予測に水分子の O 原子の速度を用いなかった場合で、赤線は用いた場合である。

隠れ層の層数は何層が適切であるかを調べるために、隠れ層の層数に対する正解率の変化を調べた (図 3.11)。この結果、隠れ層の層数の変化に対して、正解率の変動はほとんど見られなかった。水分子の速度の情報を特徴量に含めた場合も、同じく正解率の変動はほとんど見られなかった。すなわち、Na-Cl 対が会合するか解離するかの予測に隠れ層の層数はほとんど影響していなかったため、以後、隠れ層の層数は 2 層を使うことにした。

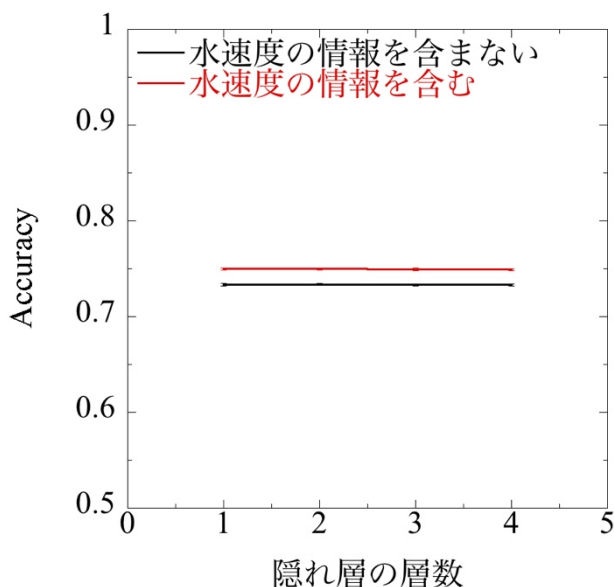


図 3.11: ニューラルネットワークにおいて、隠れ層の深さに対する正解率の変化。横軸は隠れ層の深さで、縦軸がテストデータに対する正解率である。隠れ層 1 層のノード数は 50、水分子は Na-Cl 対から 3 番目に近いものまで用いた。また、黒線が予測に水分子の O 原子の速度を用いなかった場合で、赤線は用いた場合である。

続いて、隠れ層 1 層のノード数に対する正解率の変化を調べた (図 3.12)。その結果、1 層のノード数を 10 から 50 の範囲で変えても、正解率に大きな変化はなかった。従って、1 層のノード数についてもほとんど正解率に影響していないことが分かった。水分子の速度の情報を特徴量に用いた場合も、同様に 1 層のノード数が正解率に影響していないことが分かった。また、1 層のノード数を 50、隠れ層の層数を 2 層の条件で交差検証を行なった。この際、水分子の速度も特徴量として加えている。その結果、過学習を起こしている様子はなかったため、以後 1 層のノード数は 50 とした。

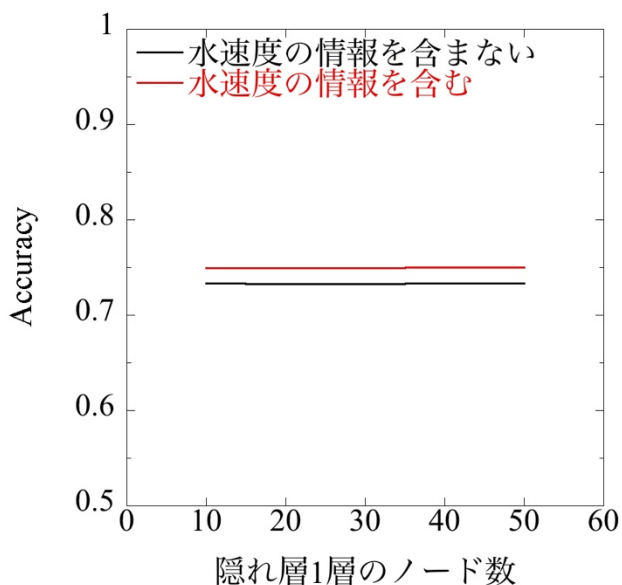


図 3.12 : ニューラルネットワークにおいて、隠れ層 1 層のノード数に対する正解率の変化。横軸は隠れ層 1 層のノード数で、縦軸がテストデータに対する正解率である。隠れ層 1 層の深さは 2 層で、水分子は Na-Cl 対から 3 番目に近いものまで用いた。また、黒線が予測に水分子の O 原子の速度を用いなかった場合で、赤線は用いた場合である。

以上の結果から、以下のニューラルネットワークの手法による解析では、隠れ層 1 層のノード数は 50、層数は 2 層として設定し、Na-Cl 対会合/解離に対する水分子の影響を調べた。

(b) 反応予測

次に、Na-Cl 対の会合/解離に対し重要な要素を調べるため、機械学習を行う際に特徴量とする水分子の個数を変化させながら予測を行ない、正解率の変化を調べた (図 3.13)。その結果、水分子に関する情報を特徴量に含めることで、正解率は向上した。しかしながら、水分子は Na-Cl 対から近いものを 2 つから 3

つ入れると、それ以上の変化はほとんどなかった。すなわち、遷移状態における Na-Cl 対の会合/解離に対し、Na-Cl 対から 2 から 3 番目までに近い水分子までが深く関わっていると考えられる。また、水分子の速度の成分を特徴量に含めた場合の方が含めなかった場合よりも正解率は高かった。そのため、水分子の速度も多少 Na-Cl 対の会合/解離ダイナミクスに影響を及ぼしていると考えられる。

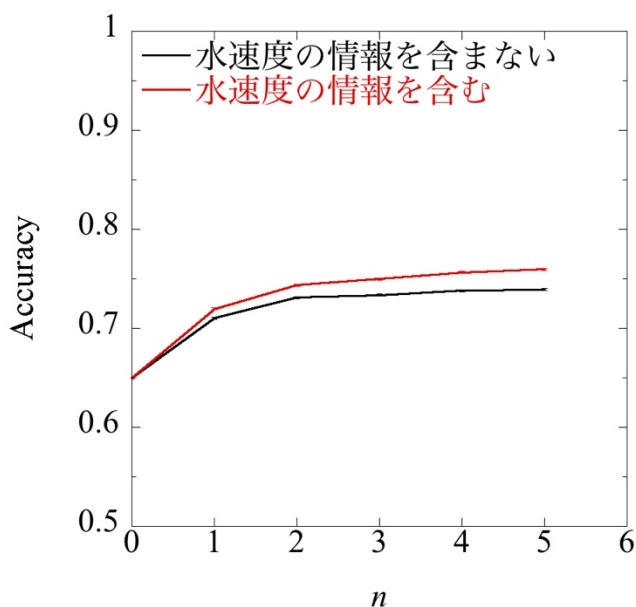


図 3.13: ニューラルネットワークにおいて、Na-Cl 対から何番目に近い水分子までを予測に使用したかに対する正解率の変化。横軸は特徴量に加えた水分子の個数 n で、縦軸がテストデータに対する正解率である。隠れ層の深さは 2 層、隠れ層 1 層のノード数は 50 で行った。また、黒線が予測に水分子の O 原子の速度を用いなかった場合で、赤線は用いた場合である。

次に、水分子同士の距離を特徴量に加えた場合についての予測精度を調べた（図 3.14）。その結果、水分子間の距離を特徴量に含めても、正解率は有意に向上していなかった。そのため、水分子間距離は Na-Cl 対の会合/解離に対してあまり重要ではないと考えられる。すなわち、水分子同士の立体的な配向はあまり Na-Cl 対の会合/解離に対して寄与していないと考えられる。水分子の速度の成分を特徴量に含めた場合、特徴量に含めなかった場合よりも正解率は高くなっている。しかしながら、水分子間の距離を特徴量に含めても含めなくても正解率にほとんど変化がない。そのため、水分子同士の立体的な配向が、Na-Cl 対の会合/解離に対する寄与している様子は検出されなかった。

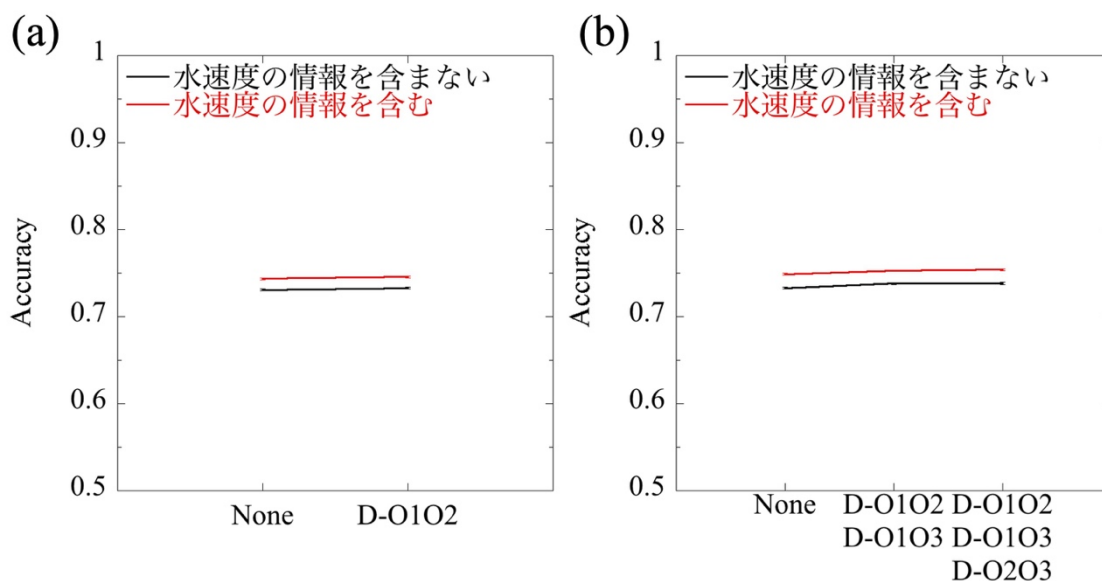


図 3.14 : ニューラルネットワークにおいて、水分子の O 原子間の距離を予測に使用した場合の正解率の変化。横軸は水分子間距離で特徴量として含めた情報、縦軸がテストデータに対する正解率である。D-O1O2 は Na-Cl 対から最も近い水分子と 2 番目に近い水分子の O 原子間距離。D-O1O3 は Na-Cl 対から最も近い水分子と 3 番目に近い水分子の O 原子間距離。D-O2O3 は Na-Cl 対から 2 番目に近い水分子と 3 番目に近い水分子の O 原子間距離。隠れ層の深さは 2 層、隠れ層 1 層のノード数は 50 で行った。また、黒線が予測に水分子の O 原子の速度を用いなかった場合で、赤線は用いた場合である。(a)は Na-Cl 対から 2 番目に近い水分子までを用いた。(b)は Na-Cl 対から 3 番目に近い水分子までを用いた。横軸の O の横の数字は、Na-Cl 対から何番目に近い水分子の O 原子であることを表している。

続いて、H 原子の情報の特徴量に加え、予測の正解率がどのように変化するかを調べた。具体的には、Na イオンおよび Cl イオンと O 原子間の距離だけでなく、二つの H 原子との距離も加えて、ニューラルネットワークによる学習お

よび予測を行った。また、水分子の速度を考慮する場合には、O 原子だけでなく H 原子の Na イオンおよび Cl イオン方向への速度成分も特徴量に加えた。図 3.15 は、H 原子の情報を特徴量に加えた場合と加えていない場合双方について、考慮した水分子の個数に対する正解率をプロットしたものである。Na イオンおよび Cl イオンと水分子の距離のみで予測した場合よりも、水分子の速度を加えた場合の方がやや正解率が高くなった。しかしながら、H 原子についての情報を特徴量に加えた場合と加えなかった場合で比較すると、あまり正解率に差はなかった。すなわち、水分子の向きは Na-Cl 対の会合/解離ダイナミクスの予測にあまり利用されていないという結果になった。

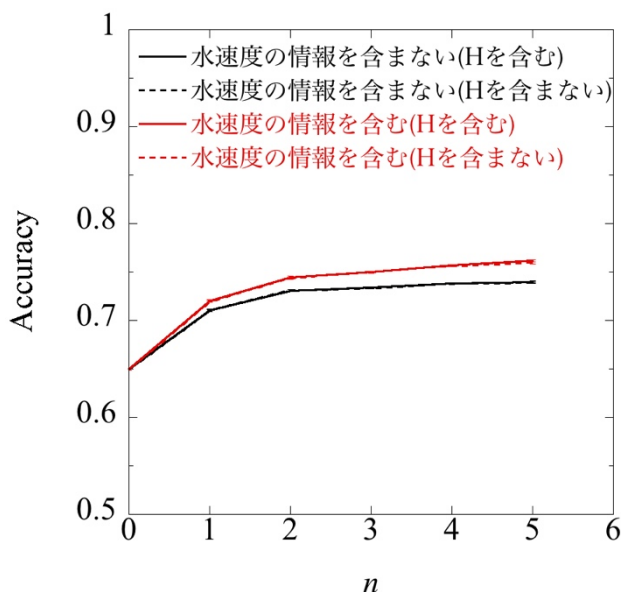


図 3.15: ニューラルネットワークにおいて、Na-Cl 対から何番目に近い水分子までを予測に使用したかに対する正解率の変化。横軸は特徴量に加えた水分子の個数 n で、縦軸がテストデータに対する正解率である。隠れ層の深さは 2 層、隠れ層 1 層のノード数は 50 で行った。実線で H 原子に関する情報を特徴量として使用した場合、点線で O 原子に関する情報のみを特徴量に使用した場合について示している。また、黒線が予測に水分子の速度を用いなかった場合で、赤線は用いた場合である。

以上の結果から、Na-Cl 対の会合/解離ダイナミクスへ大きく寄与しているのは Na-Cl 対から 3 番目に近い水分子までである。また、水分子の特徴量として重要なのは Na イオンおよび Cl イオンと水分子の距離であり、水分子同士の配向や水分子の向きはほとんど影響していないと考えられる。水分子の速度は多少影響していると考えられる。

次に、機械学習による予測が当たった点と外れた点が、それぞれ位相空間のどのような位置にあるのかについて調べた。予測が当たっていた点および外れて

いた点の中からそれぞれ 10 点ずつをランダムに選び出した。それぞれの点について、Na-Cl 対から 3 番目に近い水分子までに対して、それらの座標に小さな乱数を足し、元の点からわずかに座標をずらした点を作成した。それ以外の水分子の座標や速度は元の点のものをそのまま使用した。作成した各点が会合状態に向かうか解離状態に向かうかを MD シミュレーションで調べた。元の点からのずれの大きさを座標の RMSD として定義し、ずれの大きさごとに会合状態に向かう点の割合を調べた。ずれの大きさに対する会合状態に向かう割合の変化から、位相空間において元の点の周囲の状態を考察した。

まず、予測が当たっていた点における近傍について調べた (図 3.16)。元の点が会合状態に向かう点だった場合は(a)の図に、解離状態に向かう点だった場合は(b)の図に示した。この解析の結果、予測が当たっていた点の周りでは、元の点からある程度大きくずれるまで、会合状態に向かうかどうかは元の点と変わらなかった。実際、大半は元の点から 0.2 \AA ほどずれるまでは元の点と同様に会合もしくは解離状態に向かっていた。これは、正解した点は、位相空間において、ある程度大きな島となっている構造の内部の点であるためだと考えられる。

次に、予測が外れていた点における近傍について調べた (図 3.17)。元の点が会合状態に向かう点だった場合は(a)の図に、解離状態に向かう点だった場合は(b)の図に示した。この解析の結果、予測が外れたデータ点の周りでは二種類の傾向が見られた。一つは、元の点からのずれが大きくなるにつれて、徐々に会合状態に向かうかどうかは元の点と変わっていく傾向である。例えば、元の点が会合状態に向かう場合、元の点からのずれの大きさが 0.1 \AA にならないうちから解離状態に向かう点が増え、ずれの大きさが 0.2 \AA になると会合状態に向かう割合が 0.6 以下まで減っているものもある。この減少は、ずれの大きさに対して

ほぼ単調に起こっている。これは、位相空間において、島となっている構造の端付近にある点だと考えられる。もう一つは、元の点からのずれが小さいうちから、元の点と異なる会合/解離ダイナミクスを示すものである。少し構造がずれただけで、会合状態に向かう構造の割合が 0.4 付近になるものである。こちらの傾向のデータは、それ以上ずれの大きさが大きくなっても、会合状態に向かう割合はほぼ横ばいであった。こちらの傾向は、元の点が、会合に向かう点と解離に向かう点が狭い範囲で入り混じる、カオス的な領域にあるために起こっていると考えられる。

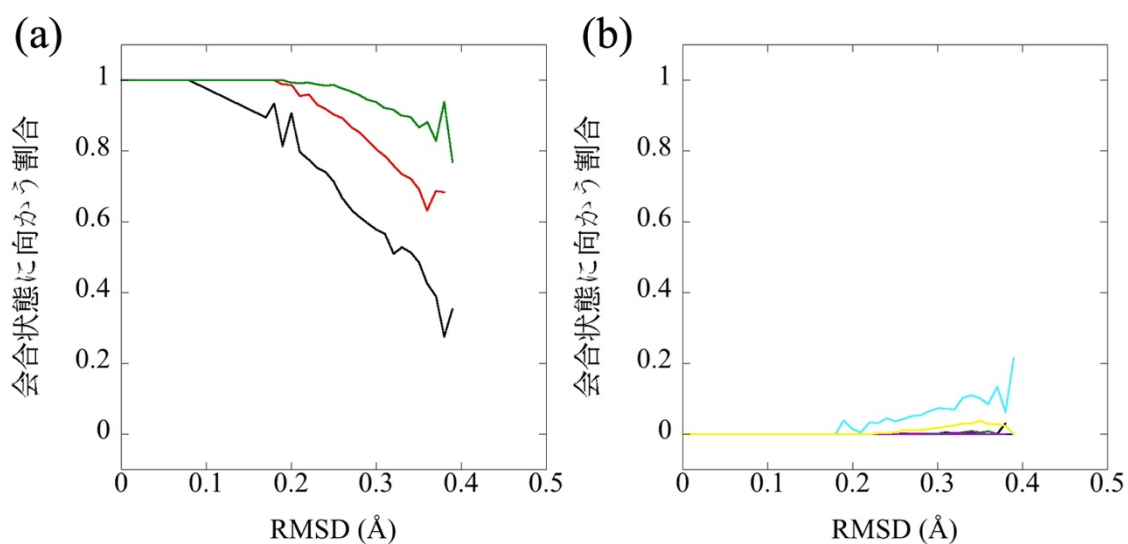


図 3.16：ニューラルネットワークによる会合/解離予測が当たっていた点について、位相空間におけるその周囲の点が会合状態に向かう割合。横軸は元の点からのずれの大きさを RMSD で示している。一本一本の線はそれぞれ元の構造に対応している。また、(a)は元の点が会合状態に向かう 3 つの点について、(b)は元の点が解離状態に向かう 7 つの点についてプロットしている。

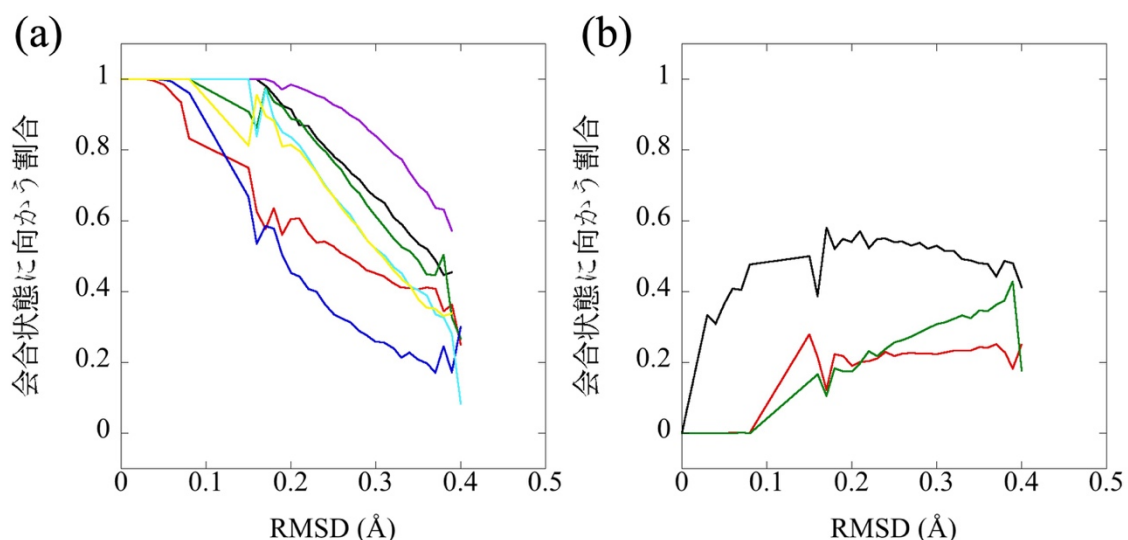


図 3.17: ニューラルネットワークによる会合/解離予測が外れていた点について、位相空間におけるその周囲の点が会合状態に向かう割合。横軸は元の点からのずれの大きさを RMSD で示している。一本一本の線はそれぞれ元の構造に対応している。また、(a)は元の点が会合状態に向かう 7 つの点について、(b)は元の点が解離状態に向かう 3 つの点についてプロットしている。

3.3.5 各特徴量の重要性

次に、Na-Cl 対の会合/解離ダイナミクスに対して大きな影響を及ぼしている物理量を調べるために、それぞれの特徴量の予測における重要性を調べた。まず、ランダムフォレストの手法で 100 万点のトレーニングデータに用いて学習を行った。この学習モデルに対して、Feature importance 法で各特徴量の重要度を評価した。また、予測テストの方に対しては、Permutation importance 法で各特徴量の重要度を評価した。

はじめに、Na-Cl 間の距離と速度、および Na-Cl 対から 5 番目までに近い水分子の距離と速度を特徴量としたデータベースを用いてランダムフォレストモ

デルを構築した。このモデルにおける各特徴量の重要度を **Permutation importance** 法により調べた (表 3.1)。この解析の結果、Na-Cl 対の会合/解離を予測するために重要な特徴量は 4 つあった。まず、Na イオンに対する Cl イオンの相対速度 (V-NaCl) である。次に、Na-Cl 対から 1 番近い水分子の O 原子と Cl イオンとの距離 (D-ClO1)、および 2 番目に近い水分子の O 原子と Cl イオンとの距離 (D-ClO2) である。最後に、4 番目に近い水分子の O 原子の Cl イオン方向への速度成分 (V-ClO4) であった。2 番目に近い水分子までが重要であるという結果は、これまでの解析で $n=2$ よりも n を大きくしてもあまり正解率が高くならなかった結果と一致している。また、同じデータベースを用いて **Feature importance** 法により各特徴量の重要度を計算した (表 3.2)。どちらの方法で計算した重要度でも、4 番目に近い水分子の Cl イオン方向への速度成分 (V-ClO4) 以外については、定性的に大きな差は見られなかった。ただし、4 番目に近い水分子の Cl イオン方向への速度成分 (V-ClO4) は、**Feature importance** 法では重要な特徴量ではないという結果だった。また、3 番目に近い水分子の O 原子と Cl イオンとの距離 (D-ClO3) は、他の特徴量の重要度との相対的な差から、あまり重要ではないと考えられる。

表 3.1 : **Permutation importance** が高い特徴量

特徴量	Permutation importance
V-NaCl	64.81± 0.28
D-ClO1	11.98± 0.56
D-ClO2	3.78± 0.72
V-ClO4	4.73±0.13

表 3.2 : Feature importance が高い特徴量

特徴量	Feature importance
V-NaCl	14.30±0.03
D-ClO1	9.83±0.07
D-ClO2	8.76±0.06
D-ClO3	4.27±0.01

Na イオンおよび Cl イオンと水分子との距離では、Cl イオンとの距離は予測に対して重要であるが、Na イオンとの距離は重要ではないという結果だった。なぜこのような結果となったのかを調べるために、データセット中の Na イオンと O 原子との距離および Cl イオンと O 原子との距離の分布を調べた (図 3.18、図 3.19)。その結果、Cl イオン-O 原子間距離に対して Na イオン-O 原子間距離は狭い (Na イオンから近い) 範囲に分布していることが分かった。すなわち、我々が定義した内容による Na-Cl 対との距離の近い水分子という条件では、Na イオンと水分子とは、Cl イオンと水分子とに比べ、ほとんどが近接している。そのため Na イオンと水分子との距離は、Na-Cl 対の会合/解離を予測する際にはあまり重要な特徴量となっていないと推察できる。

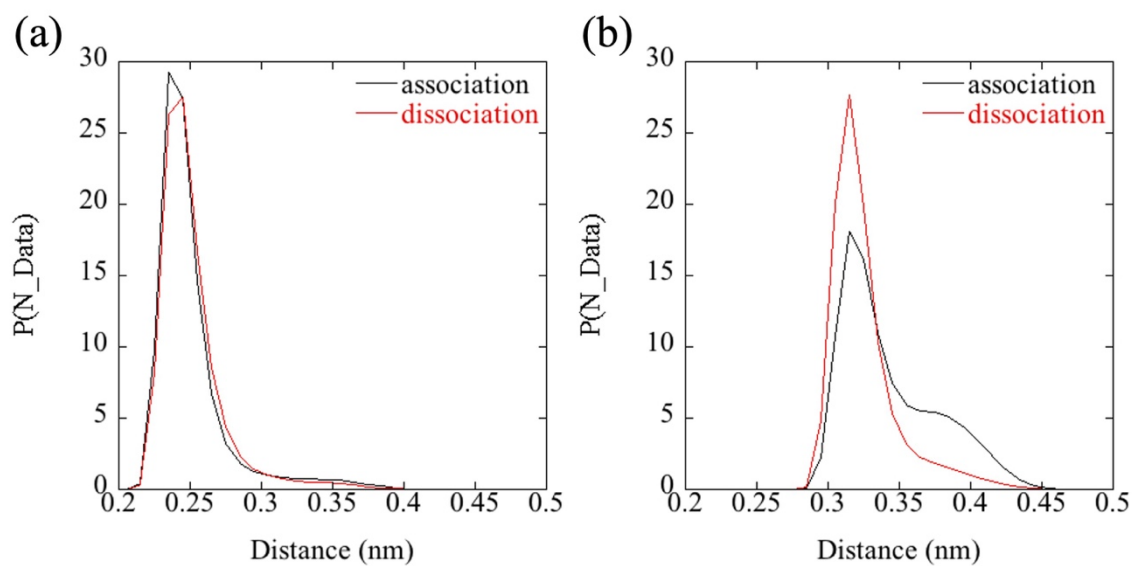


図 3.18 : Na イオンと Na-Cl 対から最も近い水分子の O 原子間の距離 (左) および、Cl イオンと O 原子間の距離 (右) の分布。横軸は Na イオン-O 原子間または Cl イオン-O 原子間の距離。縦軸は確率密度。また、黒線は会合状態に関する分布で、赤線は解離状態における分布である。

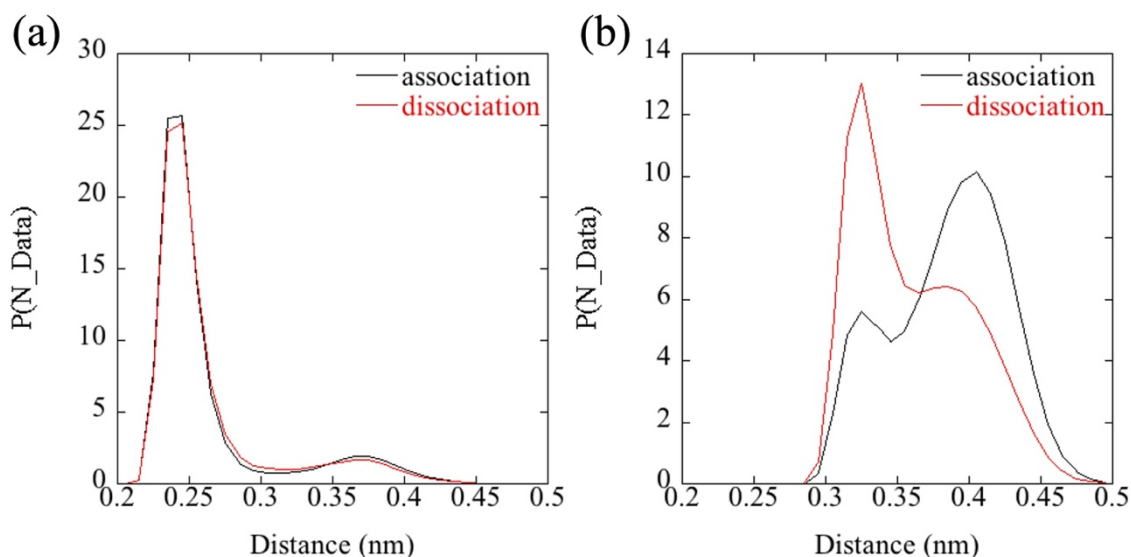


図 3.19 : Na イオンと Na-Cl 対から 2 番目に近い水分子の O 原子間の距離 (左) および、Cl イオンと O 原子間の距離 (右) の分布。横軸は Na イオン-O 原子間または Cl イオン-O 原子間の距離。縦軸は確率密度。また、黒線は会合状態に関する分布で、赤線は解離状態における分布である。

また、Na-Cl 対が会合するデータおよび解離するデータそれぞれの分布を比較すると、Na-Cl 対が解離する方向に向かうデータでは、Na-Cl 対が会合する方向に向かうデータと比較して、Cl イオン-O 原子間距離が短い領域で分布していた。Yonetani は⁷、Na-Cl 解離反応において、一つの水分子が Na イオン、Cl イオン双方と配位する構造が形成されることで、解離の自由エネルギー障壁が減少し、解離が起りやすくなると報告している。この現象は Cl イオン-O 原子間距離が短い領域で起こるために、Na-Cl 対が解離するデータの Cl イオン-O 原子間距離は短い領域で分布しているという結果となっていると考えられる。なお、我々の結果において特徴量として重要な水分子は Na-Cl 対から 2 番目までに近い水分子までであった。これは、Na-Cl 対からの近さを Na イオンからの距離と

Cl イオンからの距離の和で定義しているため、Na イオンと Cl イオン双方に近い水分子はより近い(1~2 番目)と判断されやすくなるためであると考えられる。

次に、水分子の 3 次元的な配向の重要性を調べた。Na-Cl 間の距離と速度、および Na-Cl 対から 3 番目までに近い水分子の距離および速度に、水分子の O 原子同士の距離を特徴量に加えたデータベースを作成した。このデータベースを用いてランダムフォレストモデルの学習を行い、Permutation importance 法により各特徴量の重要性を計算した (表 3.3)。V-NaO1 は Na-Cl 対から 1 番近い水分子の O 原子の Na 方向への速度成分である。その結果、水分子同士の距離は重要な特徴量でなかった。すなわち、水分子の配向はあまり重要ではないことが分かった。同じデータベースについて、Feature importance 法により各特徴量の重要度を計算した (表 3.4)。その結果、Permutation importance 法により計算された重要度と定性的に同様の結果となった。D-O1O2 は Na-Cl 対から最も近い水分子の O 原子と 2 番目に近い水分子の O 原子間の距離である。

表 3.3 : Permutation importance が高い特徴量

特徴量	Permutation importance
V-NaCl	11.67±0.04
D-ClO1	1.66±0.02
D-ClO2	0.76±0.01
V-NaO1	0.36±0.01

表 3.4 : Feature importance が高い特徴量

特徴量	Feature importance
V-NaCl	16.04±0.02
D-ClO1	10.22±0.04
D-ClO2	9.56±0.06
D-O1O2	5.07±0.03

さらに、水素についての情報を特徴量として加えた場合に、それぞれの特徴量の重要性を調べた。まず、Na-Cl 間の距離と速度、および Na-Cl 対から 5 番目までに近い水分子の O 原子と H 原子の Na イオンおよび Cl イオンへの距離について特徴量に含めたデータベースを作成した。このデータベースを用いてランダムフォレストモデルを構築し、Permutation importance 法により特徴量の重要性を調べた (表 3.5)。D-ClH1a および D-ClH1b は Na-Cl 対から最も近い水分子の H 原子と Cl イオンとの距離、D-ClH2a および D-ClH2b は Na-Cl 対から 2 番目に近い水分子の H 原子と Cl イオンとの距離である。その結果、Cl イオンの相対速度と、Cl イオンと 2 番目までの O 原子の距離が重要であるということは変わらなかった。H 原子についても、それらの Cl イオンまでの距離が、他の特徴量を比較すると若干重要であるという結果になった。しかしながら、O 原子と Cl イオンとの距離と比べると、その重要度は小さかった。この結果から、H 原子はあまり予測に対して重要でない。すなわち、水分子の向きはあまり重要でなかった。また、Na イオンおよび Cl イオン方向への O 原子および H 原子の速度成分についても特徴量に加えたデータベースを作成した。このデータベースについても同様に、Permutation importance 法により特徴量の重要性を調べた (表 3.6)。

その結果、H 原子の速度成分についての特徴量は重要な特徴量ではなかった。従って、H 原子の速度もあまり重要ではないと考えられる。

表 3.5 : Permutation importance が高い特徴量

特徴量	Permutation importance
V-NaCl	18.42±0.09
D-ClO1	2.87±0.06
D-ClO2	1.79±0.05
D-ClH1a	0.69±0.02
D-ClH1b	0.64±0.03
D-ClH2a	0.80±0.05
D-ClH2b	-0.22±0.04
D-ClO3	0.65±0.02

表 3.6 : Permutation importance が高い特徴量

特徴量	Permutation importance
V-NaCl	20.08±0.01
D-CIO1	8.62±0.25
D-CIO2	7.89±0.16
D-ClH1a	3.73±0.12
D-ClH1b	3.57±0.02
D-ClH2a	3.43±0.03
D-ClH2b	3.56±0.05
D-CIO3	2.78±0.03

また、Feature importance 法により、上記と同様の H 原子の情報を特徴量に加えたデータベースに対して、各特徴量の重要性を調べた。水分子の速度を特徴量に含めなかった場合を表 3.7 に、含めた場合を表 3.8 に示した。V-CIO3 は Na-Cl 対から 3 番目に近い水分子の O 原子の Cl イオン方向への速度成分、V-CIO1 は Na-Cl 対から最も近い水分子の O 原子の Cl イオン方向への速度成分である。Feature importance 法により計算されたものと比べると、Permutation importance 法により測った特徴量の重要性では、Cl イオンの相対速度成分が、特徴量として他の特徴量よりも重要度が大きい傾向が強く出ていた。水分子の速度を特徴量に含めた場合、Permutation importance 法による結果と定性的に同様であり、H 原子の速度は重要そうではなかった。

表 3.7 : Feature importance が高い特徴量

特徴量	Feature importance
V-NaCl	20.91±0.10
D-CIO1	3.08±0.05
V-CIO3	4.16±0.06
V-NaO1	-2.25±0.06
V-CIO1	-1.68±0.06
D-ClH2a	1.19±0.04

表 3.8 : Feature importance が高い特徴量

特徴量	Feature importance
V-NaCl	14.96±0.05
D-CIO1	6.80±0.18
D-CIO2	5.99±0.16
D-ClH1a	2.78±0.03
D-ClH1b	2.72±0.07
D-ClH2a	2.65±0.04
D-ClH2b	2.83±0.08

3.4 結論

Na-Cl 対が水中で、初期状態から会合状態に向かうか解離状態に向かうかを、MD シミュレーションを用いて調べた。初期構造の Na-Cl 間距離は PMF の鞍点である 3.7 Å 付近とした。会合状態と解離状態のどちらに向かったかについての結果を、機械学習を用いて初期状態の情報のみから予測した。機械学習予測による正解率と予測に使用した情報から、Na-Cl 対の会合/解離ダイナミクスに大きな影響を及ぼしている要素を探した。その結果、第一に、Cl イオンの Na イオンに対する相対速度が影響を及ぼしていた。そのほかに、Na-Cl 対周囲の水分子の影響が大きいことが分かった。また、会合/解離ダイナミクスに大きな影響を及ぼしている水分子は Na-Cl 対から 2~3 番目に近い水分子までであった。それ以上遠い水分子の影響は、本研究で用いた解析手法では検知されなかった。さらに、水分子の O 原子と Na イオンおよび Cl イオンとの距離だけでなく、O 原子の Na イオンおよび Cl イオン方向への速度成分も会合/解離ダイナミクスに影響を及ぼしていることが分かった。一方で、水分子同士の配置や水分子の向きは本研究の解析手法ではほとんど影響が見られなかった。

さらに、ランダムフォレストの機械学習手法を用いて学習を行ったモデルで、各特徴量の重要度を計算した。その結果、Na-Cl 対と 1 番目と 2 番目に近い水分子について、O 原子と Cl イオンとの距離の予測に対する重要性が示された。Na イオンとの距離の予測に対する重要性は、この解析からは示されなかった。実際に、Na イオン-O 原子の距離の分布は会合に導くものと解離に導くものとはほとんど変わらない。しかし、ほとんどの O 原子は Na イオンとの距離が近く、運動方程式のレベルでは O 原子と Na イオンの相互作用は大きな効果を持つ。

機械学習を分子ダイナミクスの解析に用いることで、上記のような幾つか

の発見をした。このことは、機械学習が全ての特徴量に対して偏見なく評価しながら、イオン対の振る舞いを決定する因子を見出すことができる解析ツールとなることを意味する。機械学習は、高次元のデータを機械学習は高次元のまままで解析を行うことができ、人による解析のように低次元断面に射影する必要がない点も利点となる。また、各特徴量同士の相関による影響を見いだすことも期待できる。

参考文献

1. Kumar, S.; Nussinov, R., Salt bridge stability in monomeric proteins. *J. Mol. Biol.* **1999**, *293* (5), 1241-1255.
2. Sinha, N.; Mohan, S.; Lipschultz, C. A.; Smith-Gill, S. J., Differences in electrostatic properties at antibody–antigen binding sites: implications for specificity and cross-reactivity. *BioPhys. J.* **2002**, *83* (6), 2946-2968.
3. Tsumoto, K.; Ogasahara, K.; Ueda, Y.; Watanabe, K.; Yutani, K.; Kumagai, I., Role of Salt Bridge Formation in Antigen-Antibody Interaction ENTROPIC CONTRIBUTION TO THE COMPLEX BETWEEN HEN EGG WHITE LYSOZYME AND ITS MONOCLONAL ANTIBODY HyHEL10. *J. Biol. Chem.* **1996**, *271* (51), 32612-32616.
4. Yoshida, K.; Kuroda, D.; Kiyoshi, M.; Nakakido, M.; Nagatoishi, S.; Soga, S.; Shirai, H.; Tsumoto, K., Exploring designability of electrostatic complementarity at an antigen-antibody interface directed by mutagenesis, biophysical analysis, and molecular dynamics simulations. *Sci. Rep.* **2019**, *9* (1), 4482.
5. Geissler, P. L.; Dellago, C.; Chandler, D., Kinetic pathways of ion pair dissociation in water. *J. Phys. Chem. B* **1999**, *103* (18), 3706-3710.

6. Mullen, R. G.; Shea, J.-E.; Peters, B., Transmission coefficients, committers, and solvent coordinates in ion-pair dissociation. *J. Chem. Theory. Comput.* **2014**, *10* (2), 659-667.
7. Yonetani, Y., Solvent-coordinate free-energy landscape view of water-mediated ion-pair dissociation. *Mol. Phys* **2017**, *115* (23), 2987-2998.
8. Breiman, L., Random forests. *Mach. Learn* **2001**, *45* (1), 5-32.
9. Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P., Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1947-1958.
10. Bartlett, P.; Freund, Y.; Lee, W. S.; Schapire, R. E., Boosting the margin: A new explanation for the effectiveness of voting methods. *Ann. Stat.* **1998**, *26* (5), 1651-1686.

4 機械学習に基づく抗原-抗体界面の塩橋の挙動解析

4.1 序論

HEL-HyHEL-10 複合体界面には、二組の塩橋、 $D32^H-K97^Y$ と $D99^H-K97^Y$ が存在する¹。これらの塩橋は、HEL-HyHEL-10 の結合親和性に対して大きな影響を及ぼしていることが分かっている¹⁻²。2章では、これらの塩橋に着目し、安定性について調べた。さらに、これらの塩橋の安定性を決めている要因を明らかにした。本章では、塩橋の動的な性質を理解するために、塩橋の形成/解離反応のダイナミクスを調べる。HEL-HyHEL-10 複合体界面に存在する塩橋の一つである $D32^H-K97^Y$ に着目する。3章で行った機械学習を用いた解析手法を適用することで、この塩橋のダイナミクスに対し、周囲のアミノ酸残基や水分子がどのように及ぼす影響を調べる。

4.2 方法

4.2.1 塩橋形成/解離プロセスにおける $K97^Y N_{\zeta}-D32^H C_{\gamma}$ 間距離の鞍点の位置の特定

はじめに、 $D32^H-K97^Y$ の塩橋形成/解離プロセスにおける $K97^Y N_{\zeta}-D32^H C_{\gamma}$ 間距離の鞍点の位置を調べた。 $K97^Y N_{\zeta}-D32^H C_{\gamma}$ 間距離に対して 3.7, 3.8, ..., 5.4 Å で、バネ定数がそれぞれ 4000 kJ/mol nm² で拘束をかけ、アンブレラサンプリングを行った。また、4.3, 4.35, ..., 4.65 Å および 4.44, 4.45, ..., 4.59 Å では 8000 kJ/mol nm² の拘束をかけてサンプリングを行った。これらのシミュレーション結果より、GROMACS 4.6.7 の g_wham プログラムを用いて、 $K97^Y N_{\zeta}-D32^H C_{\gamma}$ 間距離に対する平均力ポテンシャル (PMF) 曲線を作成した (図 4.1)。この結果より、

K97^YN_ζ-D32^HC_γ間距離に対する PMF の鞍点は 4.52 Å であると判断し、以下ではこれを閾値として塩橋の形成/解離を定義した。

4.2.2 初期構造・速度の作成

続いて、鞍点付近での構造を得るため、K97^YN_ζ-D32^HC_γ間距離が 4.52 Å でバネ定数 8000 kJ/mol nm² の拘束をかけ、MD シミュレーションを行った。座標および速度の出力を 30 ps ごとに行い、この中から K97^YN_ζ-D32^HC_γ間距離が 4.50 ~ 4.54 Å である時刻の構造を、初期構造・初期速度の点として使用した。これによって得られた点は 23.5 万点であった。

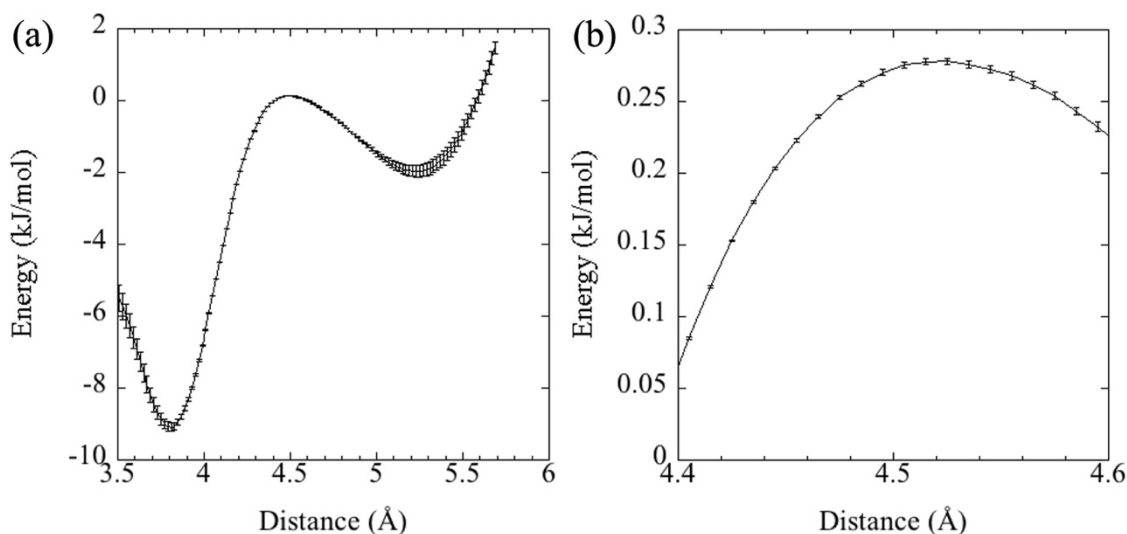


図 4.1 : K97^YN_ζ-D32^HC_γ間距離に対する自由エネルギー曲線。横軸は K97^YN_ζ-D32^HC_γ間距離で、縦軸は自由エネルギー。(a)は全体図で、(b)は鞍点付近を拡大したもの。

4.2.3 機械学習に用いたデータ

得られた初期構造・初期速度に対して、束縛をかけずにそれぞれ 150 fs の MD シミュレーションを行った。シミュレーションの後、150 fs の時点で $K97^Y N_{\zeta}$ - $D32^H C_{\gamma}$ 間距離が 4.52 Å 以下ならば会合 (True) とし、4.52 Å より大きければ解離 (False) とした。この結果を機械学習の出力とし、初期構造・初期速度の情報から機械学習 (ランダムフォレスト) で二値分類として学習/予測を行った。

機械学習に用いた特徴量は、初期構造・速度のデータのみから計算した。まず、 $K97^Y N_{\zeta}$ - $D32^H C_{\gamma}$ 間距離と $D32^H C_{\gamma}$ 原子の $K97^Y N_{\zeta}$ 原子方向への速度成分である。速度成分の計算方法は 3.2.4 で定義した方法と同じ方法で計算した。次に、周囲の水分子の O 原子の $D32^H C_{\gamma}$ 原子および $K97^Y N_{\zeta}$ 原子までの距離をそれぞれ計算し、その和が小さい順に、 $D32^H$ - $K97^Y$ の塩橋に近い水分子として定義した。この定義の元、 $D32^H$ - $K97^Y$ の塩橋から 2 番目に近い水分子までの O 原子と $D32^H C_{\gamma}$ 原子および $K97^Y N_{\zeta}$ 原子までの距離を特徴量とした。さらに、各水分子の O 原子の $D32^H C_{\gamma}$ 原子および $K97^Y N_{\zeta}$ 原子方向への速度成分についてそれぞれ特徴量とした。さらに、この他に周囲のアミノ酸残基からの影響を調べるため、2 章で作成した Res7S モデルに含まれる、 $K49^L$ 、 $Y33^H$ 、 $N97^H$ 、 $W98^H$ 、 $D99^H$ の 5 つのアミノ酸残基と $D32^H$ および $K97^Y$ の距離を特徴量として加えた。周囲のアミノ酸残基との距離の定義方法は、 $K49^L$ 、 $N97^H$ 、 $D99^H$ との距離については、 $D32^H C_{\gamma}$ 原子および $K97^Y N_{\zeta}$ 原子と $K49^L N_{\zeta}$ 原子、 $N97^H C_{\gamma}$ 原子、 $D99^H C_{\gamma}$ 原子との距離として定義した。 $Y33^H$ と $W98^H$ との距離については、 $D32^H C_{\gamma}$ 原子および $K97^Y N_{\zeta}$ 原子と $Y33^H$ 側鎖の環に含まれる重原子の重心との距離、 $W98^H$ 側鎖の二つの環に含まれる重原子の重心との距離として定義した。

4.2.4 機械学習手法

機械学習の手法はランダムフォレストを用いた。データは 23.5 万点用意した。このうち 23 万点をトレーニングデータセット、残りの 3.5 万点をテストデータセットとして用いた。また、feature importance 法を用いて、学習に使用した各特徴量の重要度を計算した。

4.3 結果と考察

遷移状態における塩橋の挙動に対する周囲のアミノ酸および水分子の影響を調べるため、ランダムフォレストを用いて機械学習を行なった。特徴量には、K97^YN_ζ-D32^HC_γ間距離と D32^HC_γ原子の K97^YN_ζ原子方向への速度成分、および 2 番目に近い水分子までの O 原子と D32^HC_γ原子および K97^YN_ζ原子との距離、K49^L、Y33^H、N97^H、W98^H、D99^Hの 5 つのアミノ酸残基と D32^Hおよび K97^Yの距離を使用した。その結果、正解率は約 0.61 だった。

次に、このモデルに対して feature importance により、使用した各特徴量の重要度を調べた (図 4.2)。その結果、D32^HC_γ原子の K97^YN_ζ原子方向への速度成分が最も重要であった。次に、K97^Yと、塩橋から 2 番目に近い水分子との距離が重要だった。続いて、K97^Yと N97^Hとの距離、D32^Hと W98^Hの距離の順で重要だった。K97^Yと、塩橋から最も近い水分子との距離も重要だった。

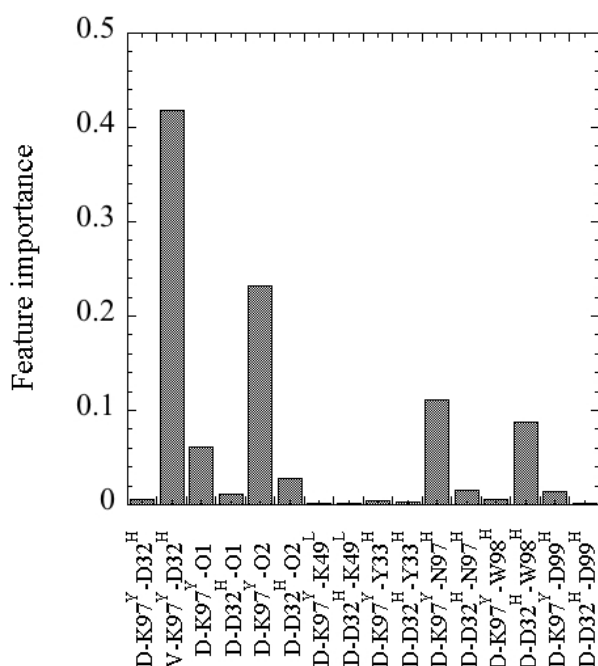


図 4.2 : Feature importance により計算された各特徴量の重要度。横軸はそれぞれの特徴量、縦軸は重要度を示している。D-K97^Y-D32^Hは K97^YN_ζ-D32^HC_γ間距離、V-K97^Y-D32^Hは D32^HC_γ原子の K97^YN_ζ原子方向への速度成分である。D-K97^Y-O1、D-D32^H-O1 はそれぞれ塩橋から最も近い水分子の O 原子と K97^YN_ζ原子および D32^HC_γ原子との距離、D-K97^Y-O2、D-D32^H-O2 はそれぞれ塩橋から 2 番目に近い水分子の O 原子と K97^YN_ζ原子および D32^HC_γ原子との距離を示している。D-K97^Y-K49^L、D-D32^H-K49^L はそれぞれ K97^YN_ζ原子および D32^HC_γ原子と K49^LN_ζ原子間距離、D-K97^Y-N97^H、D-D32^H-N97^Hはそれぞれ K97^YN_ζ原子および D32^HC_γ原子と N97^HC_γ原子間距離、D-K97^Y-D99^H、D-D32^H-D99^H はそれぞれ K97^YN_ζ原子および D32^HC_γ原子と D99^HC_γ原子間距離を示している。D-K97^Y-Y33^H、D-D32^H-Y33^Hはそれぞれ K97^YN_ζ原子および D32^HC_γ原子と Y33^H側鎖の環に含まれる重原子の重心との距離、D-K97^Y-W98^H、D-D32^H-W98^H はそれぞれ K97^YN_ζ原子および D32^HC_γ原子と W98^H側鎖の二つの環に含まれる重原子の重心との距離を示している。

続いて、D32^HC_γ原子の K97^YN_ε原子方向への速度成分以外で、特に重要そうだった 4 つの特徴量について、会合に向かうデータと解離に向かうデータに分けて、それぞれ分布を調べた。その結果、水分子については、K97^Y から近い位置にある方が、その初期構造から解離しやすいということが分かった(図 4.3)。これは特に、塩橋から 2 番目に近い水分子の方が顕著である。そのため、特徴量の重要性としては、2 番目に近い水分子の方が大きくなっていると考えられる。周囲のアミノ酸については、N97^H は K97^Y から近い位置にある方が解離しやすいが、W98^H は K97^Y から遠い位置にある場合の方が解離しやすかった。塩橋の安定性を決める上でも、W98^H 側鎖の配向は重要であり、塩橋と W98^H 側鎖の関係の重要性が示された。

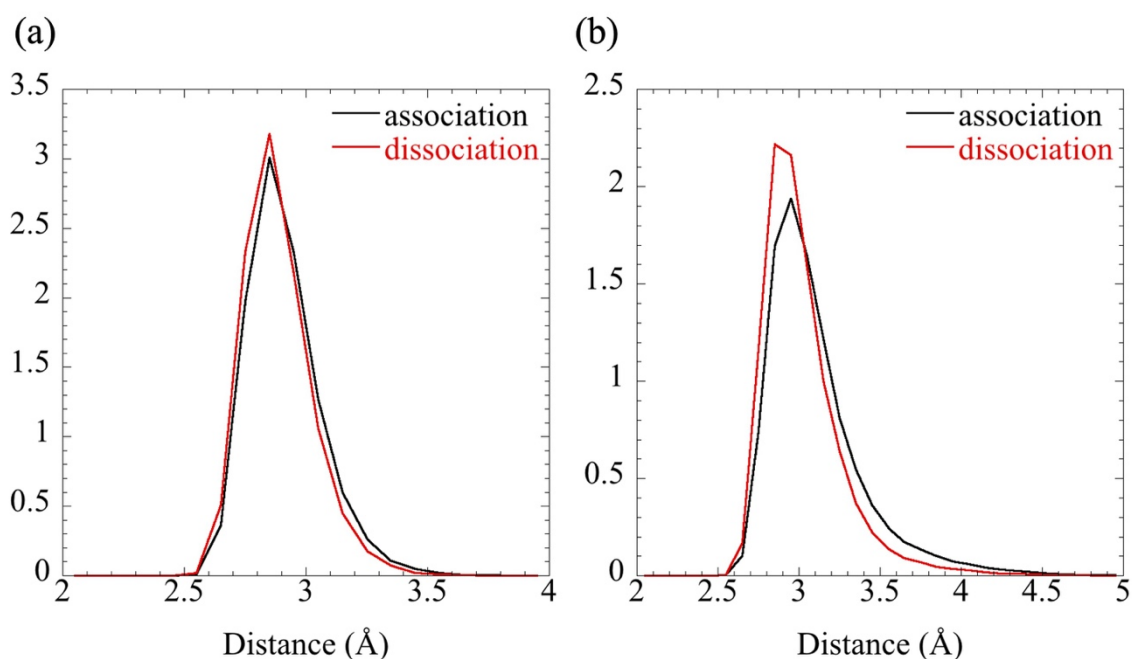


図 4.3 : (a) K97^YN_ε原子と塩橋から最も近い水分子との距離の分布。(b) K97^YN_ε原子と塩橋から 2 番目に近い水分子との距離の分布。黒線は 150 fs のシミュレーション後に会合するデータについての分布。赤線は 150 fs のシミュレーション後に解離するデータについての分布。

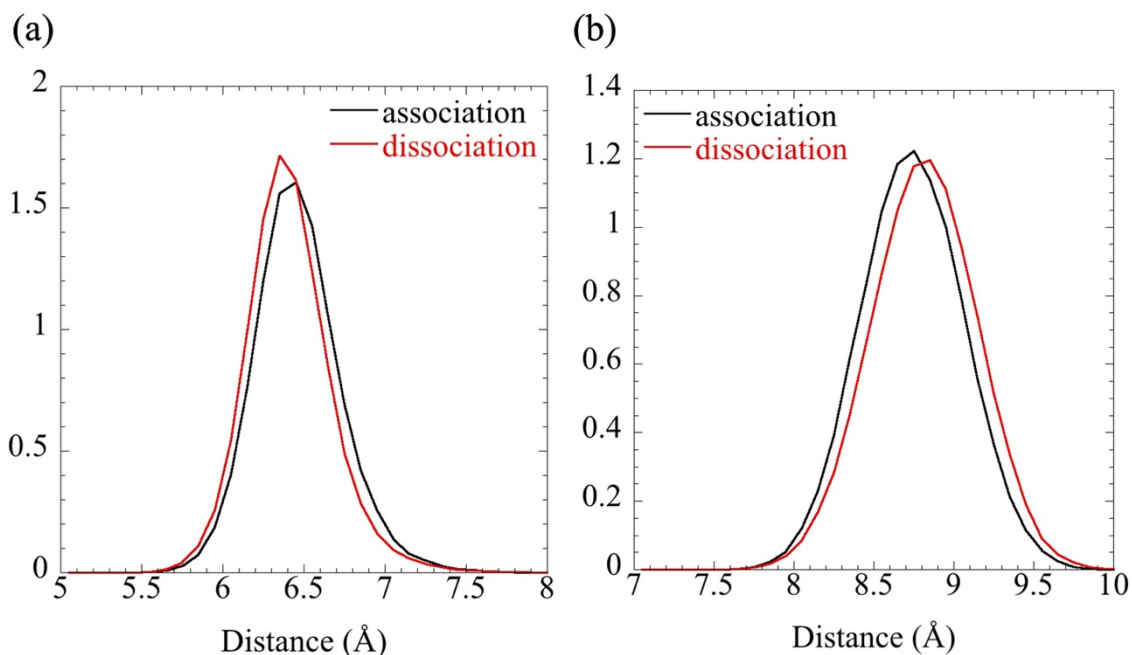


図 4.4 : (a) $K97^Y N_z$ 原子と $N97^H C_\gamma$ 原子間距離の分布。 (b) $D32^H C_\gamma$ 原子と $W98^H$ 側鎖の二つの環に含まれる重原子の重心との距離の分布。黒線は 150 fs のシミュレーション後に会合するデータについての分布。赤線は 150 fs のシミュレーション後に解離するデータについての分布。

4.4 結論

HEL-HyHEL-10 界面に存在する $D32^H-K97^Y$ の塩橋が、遷移状態から会合方向に向かうか解離方向に向かうかを、MD シミュレーションを用いて調べた。この塩橋の会合/解離プロセスにおける遷移状態は、PMF より、 $K97^Y N_z-D32^H C_\gamma$ 間距離が 4.52 Å 付近であるとした。遷移状態付近の初期構造から会合状態と解離状態のどちらに向かったかについての結果を、機械学習を用いて初期状態の情報のみから予測した。その結果、塩橋から 2 番目に近い水分子までの情報と、Res7S モデルに含めたアミノ酸に関する情報を特徴量として含めた場合、正解率

は約 0.61 となった。Feature importance 法により、このモデルの各特徴量の重要度を計算したところ、D32^HC_γ原子速度以外では、K97^Yと塩橋から 1、2 番目に近い水分子との距離が重要だった。また、K97^Yと N97^Hおよび、D32^Hと W98^Hとの距離が重要だった。特に、W98^Hは 2 章で調べた塩橋の安定性に対して重要な残基であった。会合に向かうデータと解離に向かうデータでそれぞれの特徴量の分布を取ったところ、水分子が K97^Yに近い位置に存在すると、塩橋は解離しやすくなることが分かった。これは特に 2 番目に近い水分子の方が顕著だった。周囲のアミノ酸の場合、K97^Yと N97^Hの距離が近いと塩橋は解離しやすかったが、D32^Hと W98^Hとの距離は遠い方が塩橋は解離しやすいことが分かった。

参考文献

1. Shiroishi, M.; Yokota, A.; Tsumoto, K.; Kondo, H.; Nishimiya, Y.; Horii, K.; Matsushima, M.; Ogasahara, K.; Yutani, K.; Kumagai, I., Structural Evidence for Entropic Contribution of Salt Bridge Formation to a Protein Antigen-Antibody Interaction THE CASE OF HEN LYSOZYME-HyHEL-10 Fv COMPLEX. *J. Biol. Chem.* **2001**, *276* (25), 23042-23050.
2. Tsumoto, K.; Ogasahara, K.; Ueda, Y.; Watanabe, K.; Yutani, K.; Kumagai, I., Role of Salt Bridge Formation in Antigen-Antibody Interaction ENTROPIC CONTRIBUTION TO THE COMPLEX BETWEEN HEN EGG WHITE LYSOZYME AND ITS MONOCLONAL ANTIBODY HyHEL10. *J. Biol. Chem.* **1996**, *271* (51), 32612-32616.

5 総括

本論文では、HEL-HyHEL-10 複合体界面に存在する塩橋の水中での挙動について、MD シミュレーションを用いて解析を行った。2 章では、HEL-HyHEL-10 界面に結晶構造で二つ存在した塩橋のうち、D32^H-K97^Y の塩橋は水中でも安定に存在したが、D99^H-K97^Y の塩橋は水中では不安定で、ほとんど形成されないことを示した。これらの安定性がどのような要因によって決まっているのかを調べるため、複合体界面環境を部分的に再現した少数アミノ酸モデル系を構築した。その結果、複合体界面の塩橋の安定性を再現するためには、塩橋を形成するアミノ酸残基の他に、K49^L、N97^H、Y33^H、W98^H のアミノ酸残基が存在することと、Y33^H と W98^H の側鎖の配向が重要だった。また、塩橋を形成するアスパラギン酸の一つをアラニンに置換し、一方の塩橋が形成できなくなった時の残ったもう一方の塩橋の安定性を調べた。その結果、D99^H をアラニンに置換しても、D32^H-K97^Y の塩橋安定性はほとんど変化しなかったが、D32^H をアラニンに置換した場合、D99^H-K97^Y の塩橋は大きく安定化することが分かった。なぜこのような差が出るのかを調べるために、少数アミノ酸モデル系による解析を行った。その結果、D32^HA 変異による主鎖構造の歪みが原因であったことが分かった。また、この変異複合体系の塩橋安定性を再現するためには、K49^L、N97^H、Y33^H、W98^H の他に、L75^Y と D101^H の二つのアミノ酸残基が重要であることが分かった。

続いて、塩橋のダイナミクスに着目した。塩橋周囲の多数のアミノ酸残基や水分子の中から、どの要素が塩橋のダイナミクスに大きな影響を及ぼしているのかを調べるため、新たな解析手法が必要となる。3 章では、イオン間相互作用に関する最も単純なモデルである Na-Cl イオン対の会合/解離プロセスを対象に、

機械学習を用いた解析手法の構築を試みた。Na-Cl イオン対の会合/解離プロセスにおける、解離座標に沿った自由エネルギー曲線の鞍点付近における Na-Cl 対の挙動を、MD シミュレーションを用いて調べた。シミュレーション後に Na-Cl イオン対が会合する方向に動いたか、解離する方向に動いたかを、初期構造の情報のみから機械学習で学習/予測した。機械学習手法は、k 近傍法やサポートベクターマシン、ニューラルネットワーク、ランダムフォレストを用いた。その結果、学習に必要なデータ数や計算コストから、ニューラルネットワークもしくはランダムフォレストが有用だった。この二つの機械学習手法による結果はほぼ同じで、Na-Cl イオン対の会合/解離プロセスに対して、Na-Cl 対から 2~3 番目に近い水分子が影響していることが分かった。また、水分子の速度は Na-Cl 対の動きに影響するが、水分子の向きや水分子同士の配向の影響はほとんど検出されなかった。また、機械学習による予測が正解しやすい点と外れやすい点を、位相空間における周囲の点の挙動を調べることで分析した。その結果、位相空間上で島となっている構造の内部の点は正解しやすいことが分かった。一方で、位相空間上で島となっている構造の端付近の点や、会合に向かう点と解離に向かう点が狭い範囲で入り混じるカオス的な領域の点は外れやすいことが分かった。さらに、ランダムフォレストを用いたモデルに対して **feature importance** や **permutation importance** の手法により、各特徴量の重要度を計算した。この解析の結果、水分子の情報の中でも、水分子と Cl 原子の距離が重要であることが分かった。シミュレーション後に Na-Cl 対が会合するデータと解離するデータそれぞれに対して、これらの特徴量の分布を計算し、比較したところ、水分子が Cl 原子から近い方が Na-Cl イオン対は解離しやすいことが分かった。以上の結果から、ランダムフォレストによる学習/予測を行い、各特徴量の重要度を計算す

ることが有用であることが分かった。

4 章では、Na-Cl イオン対に対して行った機械学習を用いた解析を、HEL-HyHEL-10 界面の塩橋に対して適用した。対象としたのは D32^H-K97^Y の塩橋である。この塩橋が自由エネルギー地形における鞍点の構造を初期構造として、会合方向に向かうか解離方向に向かうかを、機械学習で学習/予測し、各特徴量の重要性を計算した。その結果、塩橋周囲の水分子については、塩橋から 2 番目に近い水分子までと K97^Y との距離が重要だった。会合するデータと解離するデータで水分子と K97^Y の距離の分布を計算したところ、水分子が K97^Y に接近していると、塩橋はやや解離する方向に動きやすいということが分かった。また、この傾向は塩橋に最も近い水分子よりも、2 番目に近い水分子の方が顕著であった。周囲のアミノ酸からの影響については、N97^H と W98^H の二つのアミノ酸残基が重要だった。N97^H と K97^Y との距離が近いときに、塩橋は解離しやすかった。対して、W98^H は D32^H との距離が遠いときに塩橋は解離しやすかった。W98^H は、2 章において、塩橋の安定性に対しても重要な寄与をしていることが明らかになっているアミノ酸残基である。

本研究において、我々は HEL-HyHEL-10 複合体界面に存在する塩橋の安定性に影響している環境因子を特定することができた。本研究で提案した少数アミノ酸モデルや機械学習を用いた解析は、HEL-HyHEL-10 複合体系に限定されるものではなく他の抗原-抗体複合体系に対しても適用できるものである。こうした解析で得られる環境要因の調節を通して、塩橋の安定性のみに影響を与える変異を見出すことができれば親和性の向上につながる可能性は高い。したがって、本研究の成果は抗体の親和性を向上させる変異を設計するヒントとなると期待できる。

謝辞

本研究は、著者が東京大学大学院 総合文化研究科 広域科学専攻 博士課程の在籍中の研究成果をまとめたものです。本研究は、高度情報科学技術研究機構 (hp210172、hp210141 など) のプロジェクトの下、行いました。本研究の分子動力学計算及び機械学習は、東京工業大学 学術国際センターの TSUBAME の計算資源を利用することで、行わせていただきました。また、HEL-HyHEL-10 複合体系の分子動力学計算には、理化学研究所計算科学研究機構に設置されていた、京の計算資源も使わせていただきました。これらの施設/設備を保守、運営されているスタッフの方々に厚く御礼申し上げます。

指導教員である平岡秀一教授には、日々の研究等のスケジュールなど、様々な助言をいただきました。ここに、心より感謝の言葉を申し上げます。

東京大学先端科学研究センターの山下雄史特任准教授には、日頃より多くの御指導、御鞭撻を賜りました。また、研究活動のみならず、日々の生活の中で様々な助言をいただきました。感謝申し上げます。

修士課程において指導教員を引き受けてくださった高塚和夫先生には、研究や日々の生活の仕方についてご指導を受け、現在の研究をする上での基礎となっています。感謝致します。

当時高塚研究室に在籍しておりました、高橋聡先生や米原丈博先生、新崎康樹先生には、計算機関連や研究へのご指摘などお世話になりました。感謝致します。

また、山下研究室の佐々木皓平氏、庄司直幸氏には、日々の研究でご指摘や励ましをいただき、感謝致します。

同じく、狐塚桃子氏や宮村尚明氏には日々の研究生活の中でお世話になりました。感謝致します。

最後になりますが、研究活動を励まし、支えてくださった両親に感謝の言葉を述べさせていただきます。