学位論文

計算科学による
タンパク質複合体の構造決定法に
関する研究

金森　英司

2011 年

要旨

タンパク質の多くは生体内で他のタンパク質と相互作用することで機能しており、タンパク質間の相互作用も立体構造の観点から理解することが重要である。しかしながら、タンパク質が相互作用して形成する複合体の立体構造の決定は実験手法の制約から難しいことも多い。そこで本研究では、単体のタンパク質立体構造がそれぞれ既知である場合に、計算科学のアプローチを用いて、それらが形成するタンパク質複合体の構造を決定するための 2 つの新規解析手法を開発した。

1 つ目の解析手法は、核磁気共鳴(NMR)のアミノ酸選択的交差飽和法の実験データと分子動力学計算を組み合わせ高い精度の複合体構造を構築する手法である。NMR による実験法の 1 つである交差飽和法は、相互作用する 2 つのタンパク質のうちの一方を同位体標識し、スピン緩和実験によりタンパク質―タンパク質相互作用面にあるアミノ酸残基の同定を可能にする手法である。高分子量のタンパク質にも適用できる手法だが、複合体の立体構造を決定するためには得られる情報が不十分であった。

アミノ酸選択的交差飽和法は交差飽和法を発展させた手法で、アミノ酸の種類ごとに標識が可能なため、交差飽和源となるアミノ酸残基についても情報が得られる。実験で観測されるシグナル強度の変化は、タンパク質複合体モデルからブロッホ方程式を解くことで計算できる。よって、観測値と計算値の差をペナルティ関数として取り入れ、分子動力学計算による Simulated Annealing（焼きなまし法）により構造を探索することで実験値を再現するような精度の高い複合体構造を得ることができる。

本手法を酵母ユビキチンと酵母ユビキチンヒドロラーゼ 1 の複合体に適用したところ、初期構造として結合状態のユビキチンを用いた場合でも、非結合状態のユビキチンを用いた場合でも X 線結晶構造により得られている既知複合体構造と非常に高い精度で一致する構造を得ることができた。得られた構造は、顕な水中での分子動力学計算によりさらに精度を高めることができた。さらに、得られた複合体が実験結果を満たすためのペナルティ関数の基準について検討し本手法の適用可能性について考察した。

2 つ目の解析手法は、タンパク質間相互作用についての実験データを使用せずに、2 つの単体の立体構造からタンパク質複合体の構造を予測する手法である（タンパク質－タンパク質ドッキング）。複合体構造の予測には構造を評価するための目的関数が重要となる。そこで、タンパク質の相互作用部位を Evolutionary Trace 法を用いてアミノ酸配列の保存度により予測し、その値と分子表面の相補性により構造を評価する目的関数を定義した。265 個の既知のタンパク質相互作用面について解析したところ、酵素タンパク質とシグナル伝

達タンパク質において、配列保存度の高いアミノ酸残基ほど相互作用部位になりやすいという相関がみられた。

本手法を評価するため、タンパク質複合体構造予測法の評価のための国際的コミュニティである CAPRI のブラインドテストを用いたところ、TolB/Pal 複合体などの例において非結合状態のタンパク質を初期構造として用いて天然の複合体構造と類似の複合体構造を得ることができた。さらに開発した目的関数を基に複合体構造予測を自動化するサーバを構築し、いくつかのブラインドテストにおいて CAPRI コミュニティで最も精度の高い構造の構築に成功した。

目次

序論 7

1　アミノ酸選択的交差飽和法と分子動力学計算を用いたタンパク質複合体の構造決定法 14

1-1　要旨 14

1-2　序論 14

1-3　方法 19

1-3-1　CS 源となる残基の組み合わせ候補の作成 19

1-3-2　Ub と YUH1 の初期構造の準備 19

1-3-3　3 次元構造の構築 20

1-3-4　Simulated Annealing（焼きなまし法）によるモデル構築 22

1-4　結果と考察 23

1-4-1　ASCS 実験結果による CS 源残基の組み合わせ候補の作成 23

1-4-2　Ub-YUH1 タンパク質複合体のモデル構築 25

1-4-3　E_{ASCS} と RMSD の関係 33

1-4-4　ASCS 法の適用可能性 35

1-5　結論 36

参考文献 38

2　Evolutionary Trace 法を用いた分子表面によるタンパク質複合体構造の予測 41

2-1　要旨 41

2-2　序論 41

2-3　方法 42

2-3-1　概観.....42

2-3-2　目的関数.....44

2-3-3　最適化.....45

2-4　結果と考察.....48

2-4-1　既知の複合体構造を学習データセットに用いたパラメータの決定.....48

2-4-2　Evolutionary Trace 法による予測部位とタンパク質間相互作用部位の相関.....49

2-4-3　CAPRI ターゲットを用いたテスト.....52

2-4-4　ターゲット 12:cohesin と dockerin のドッキング.....53

2-4-5　ターゲット 18:xylanase と TAXI xylanase inhibitor のドッキング.....54

2-4-6　ターゲット 21:Orc1p と Sir1p のドッキング.....55

2-4-7　ターゲット 24 および 25:ARF1 と ArfBD のドッキング.....56

2-4-8　ターゲット 26:TolB と Pal のドッキング.....57

2-4-9　ターゲット 27:E2-25K と Ubc9 のドッキング.....60

2-4-10　ターゲット 28:NEDD4L のホモ二量体構造の予測.....61

2-4-11　ターゲット 40:Trypsin と API-A のドッキング.....62

2-4-12　ターゲット 42:TPR モチーフのつくるホモ二量体構造の予測.....64

2-5　結論.....65

参考文献.....66

Appendix.....69

A　アミノ酸選択的交差飽和実験の方法と結果.....69

B　CAPRI(Critical Assessment of PRedicted Interactions) 73

C　Evolutionary Trace 法.....77

謝辞 80

序論

タンパク質の機能をその立体構造から解明する取り組みは 1958 年のジョン・ケンドリューらによるミオグロビンの X 線結晶構造解析[1]以来続けられており、近年のタンパク質の発現、精製、結晶化の技術の発展、シンクロトロン放射光の利用、計算機の高性能化によってますます加速され、2011 年 11 月現在では 77,243 件の構造が PDB(Protein Data Bank)[2]に登録されている。

しかしながら、多くのタンパク質は、他のタンパク質を含む生体分子と相互作用して機能を発揮しており、タンパク質単体の立体構造が明らかになっても、タンパク質の機能を理解するには十分でない。タンパク質－タンパク質の相互作用を立体構造の観点から理解することが必要である。

一方、近年 two-hybrid 法等の技術によりタンパク質－タンパク質の相互作用が大規模に解析されるようになり[3,4]、生体内でのタンパク質間のネットワークが解明されるようになった。

タンパク質－タンパク質相互作用の理解は応用面からも重要性が増している。タンパク質－タンパク質相互作用面に特異的に結合し相互作用を阻害する低分子や抗体は医薬品としての利用が期待できるためである。図 1 は 2005 年の種類別の薬の標的の数であり、タンパク質－タンパク質相互作用を標的としたものが 20%を占めるようになっている[5]。GPCR(G protein-coupled receptor)などの創薬に適した標的が今後枯渇した場合、これまで困難とされてきたタンパク質－タンパク質相互作用を標的とした新薬開発の重要度がいっそう増していくと考えられる。

タンパク質－タンパク質相互作用を創薬標的とした例としてはがん抑制タンパク質 p53 とタンパク質 MDM2 の例が良く知られている。MDM2 は p53 と相互作用し p53 を抑制的に調節する。多くの悪性腫瘍で MDM2 の過剰発現がみられており p53 と MDM2 の相互作用の阻害はがん治療法として期待されている[6]。p53 と MDM2 の複合体の立体構造は X 線結晶構造解析により決定され[7]、明らかにされた相互作用に基づいて薬物の探索が行われた。2004 年に Nutlin と呼ばれる強い阻害活性を持つ低分子化合物が報告された[6]。MDM2 と Nutlin の複合体構造が X 線結晶構造解析により解かれ、Nutlin は p53 をミミックする形で MDM2 に結合していた。このように、タンパク質－タンパク質複合体構造を決定し、タンパク質－タンパク質相互作用部位と相互作用様式を明らかにすることは応用面からも極めて重要である。

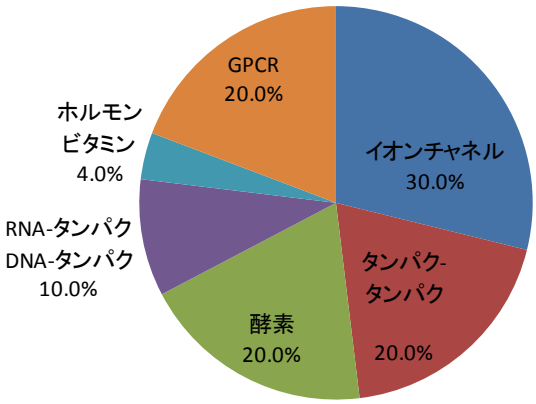

図１　種類別の薬の標的の数(2005 年)

以上のような状況から、タンパク質が相互作用して形成する複合体の立体構造を決定することが重要な課題となってきたが、タンパク質複合体を実験的に決定することは容易ではない。

タンパク質複合体の構造決定法としては、通常の単体のタンパク質と同じように X 線結晶構造解析を用いることが最も高精度で信頼性の高い構造を得る方法であり、原子分解能の構造を決定することができる。しかし、そのためには相互作用する複数のタンパク質を精製し、複合体を形成した状態で結晶化する条件を見つけなければならず極めて困難な課題となる。また、このような特徴から安定な複合体には比較的適用可能であるが、細胞内でのシグナル伝達のために生じるような一時的なタンパク質複合体の構造決定には適していない。

このような事情から、タンパク質単体の構造がそれぞれ既知であり、それらが直接的な相互作用をする場合に、既知の単体の構造を活用して複合体の立体構造を決定する手法が開発されるようになった。

そのような手法は大きく２つに分けられる。実験の情報を基に計算機による解析で複合体構造を構築する手法[8]と主に計算機を用いる予測法である。

前者で用いられる実験手法は多岐にわたり、得られる構造情報の種類、空間分解能、精度、適用可能な対象が手法によって異なる。実験手法の１つは、低温電子顕微鏡による単粒子解析等の手法で得られた 10-20Å程度の低分解能の像に単体の構造を当てはめる手法である[9]。この手法は多くのタンパク質から構成される対称性が高く分子量の大きい複合体の構築に適しており複合体全体の形状と分子の大まかな配置を決定するのに適している。

他の実験手法としては、X 線小角散乱（small-angle X-ray scattering、SAXS）が知られている、この手法からも複合体の形状についての低分解能の構造情報が得られる。この情報と既知の単体の構造を組み合わせることで、複合体モデルを作成することができる[10]。この手法の利点は、低温電子顕微鏡や核磁気共鳴を適用するのが難しい 50-250kDa の分子量の複合体に適していることである。また本手法は溶液条件を変えることが容易なため、溶液条件の変化により異なった状態の複合体構造を捉えることが可能である[8]。

その他、蛍光共鳴エネルギー移動（fluorescence resonance energy transfer, FRET）[11,12]法では相互作用するタンパク質の蛍光標識した部位間の距離情報が得られる他、質量分析[13,14]や部位特異的アミノ酸変異実験[15]といった手法によってもタンパク質複合体の構造情報を得ることができる。

以上のように様々な実験手法がある中、タンパク質分子間の相互作用面を同定するのに最も適しているのが核磁気共鳴（NMR）による手法である。本手法は生理的条件に近い溶液中で測定が可能である点が特徴である。NMR の手法としては、ケミカルシフトパータベーション法[16]や水素-重水素交換実験[17]が知られているが、予測される相互作用面の精度は高くなかった。2000 年に高橋らによって報告された交差飽和法は、タンパク質間のスピン緩和を測定することで、より信頼性の高い相互作用面の情報を得ることが可能であることを示した[18,19]。

しかし、これらの手法は相互作用面の同定はできるが、同定された相互作用面を用いた複合体構造は複数考えられるため、一意の複合体構造を高い信頼度で決定するためにはさらに他の実験結果を要する。

そこで、交差飽和法を発展させた手法として、アミノ酸選択的交差飽和法が提案された[20]。この手法はアミノ酸選択的な同位体標識を行うことで単なる相互作用面の情報ではなくアミノ酸残基の相互作用ペアに近い情報を与える。

本論文の第 1 章では、このアミノ酸選択的交差飽和法で得られた実験データを利用し、分子動力学計算による Simulated Annealing（焼きなまし法）により、実験データを再現する高い精度の複合体構造を決定する手法について述べる。

一方、計算機による複合体構造の予測法としては、1978 年に Wodak らにより先駆的な研究がなされた[21]。その後、タンパク質間相互作用が注目を集めるようになったことと、高速な計算機が使用できるようになったことにより、近年多くのアルゴリズムが提案されるようになった。

ZDOCK[22]や PIPER[23]などのプログラムは高速フーリエ変換を用いることで構造探索を高速化している。また、ICM-DISCO[24]や RosettaDock[25]などのプログラムはモン

テカルロ法による構造探索を行う。一方、PatchDock[26]は幾何学的ハッシングと呼ばれる技法を用いる。このように各プログラムでは分子の形状や特性の相補性を評価するために多様な手法が用いられている。

本論文の第 2 章では、木下らがタンパク質分子の表面形状をもとにタンパク質の機能推定を行うデータベースを作成した[27,28]点から着想を得て開発した、タンパク質分子の形状相補性の評価のために分子表面の表現を用いた複合体構造予測法について述べる。相互作用部位の予測法としては、Evolutionary Trace 法（進化追跡法）[29]によるアミノ酸配列の保存度を適用した。

以上のようにタンパク質複合体の構造決定法には実験の情報を用いる手法と計算機による予測法があり、それぞれ新しい手法を発展させている。本論文ではそれぞれについて新しいアプローチの解析手法を述べるが、これらは相補的な関係にあり、実際は厳密な区別はできない。すなわち、前者は実験情報に基づく手法であるが、実験で得られる情報が部分的で、一意な構造を決定するのに不十分である場合は、他の実験情報も含め、後者による計算機による予測法を統合的に用いて精度を向上させるように利用される。一方、後者の手法はケースによっては複数の候補の中に正しい構造を提示することが可能であるが構造の正しさを定量的に評価することが難しい。よって、文献情報等から得られる実験の情報が精度を高めるために用いられる。また、後者の方法は常に正しい答えを提示することは難しいが、更なる実験のための仮説となる複合体モデルを提供することで実験の負担を削減することに利用されている。

参考文献

1. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature. 1958;181:662–666.

2. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nat Struct Biol. 2003;10:980.

3. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature. 2000;403:623–627.

4. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A. 2001;98:4569–4574.

5. Barfai T, Lees GV. Drug discovery from bedside to wall street. Elsevier academic press 2006

6. Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z, Kong N, Kammlott U, Lukacs C, Klein C, Fotouhi N, Liu EA. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science. 2004;303:844–848

7. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, Pavletich NP. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science. 1996;274:948–953.

8. Robinson CV, Sali A, Baumeister W. The molecular sociology of the cell. Nature. 2007;450:973–982.

9. Fotin A, Cheng Y, Grigorieff N, Walz T, Harrison SC, Kirchhausen T. Structure of an auxilin-bound clathrin coat and its implications for the mechanism of uncoating. Nature. 2004;432:649–653.

10. Förster F, Webb B, Krukenberg KA, Tsuruta H, Agard DA, Sali A. Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies. J Mol Biol. 2008;382:1089–1106.

11. Yan Y, Marriott G. Analysis of protein interactions using fluorescence technologies. Curr Opin Chem Biol. 2003;7:635–640.

12. Muller EG, Snydsman BE, Novik I, Hailey DW, Gestaut DR, Niemann CA, O'Toole ET, Giddings TH Jr, Sundin BA, Davis TN. The organization of the core proteins of the yeast spindle pole body. Mol Biol Cell. 2005;16:3341–3352.

13. Back JW, de Jong L, Muijsers AO, de Koster CG. Chemical cross-linking and mass spectrometry for protein structural modeling. J Mol Biol. 2003;331:303–313.

14. Vasilescu J, Figeys D. Mapping protein-protein interactions by mass spectrometry. Curr Opin Biotechnol. 2006;17:394–399.

15. Sivasubramanian A, Chao G, Pressler HM, Wittrup KD, Gray JJ. Structural model of the mAb 806-EGFR complex using computational docking followed by computational and experimental mutagenesis. Structure. 2006;14:401–414.

16. Foster MP, Wuttke DS, Clemens KR, Jahnke W, Radhakrishnan I, Tennant L,

Reymond M, Chung J, Wright PE. Chemical shift as a probe of molecular interfaces: NMR studies of DNA binding by the three amino-terminal zinc finger domains from transcription factor IIIA. J Biomol NMR. 1998;12:51–71.

17. Paterson Y, Englander SW, Roder H. An antibody binding site on cytochrome c defined by hydrogen exchange and two-dimensional NMR. Science. 1990;249:755–759.

18. Shimada I. NMR techniques for identifying the interface of a larger protein–protein complex: cross-saturation and transferred cross-saturation experiments. Methods Enzymol. 2005;394:483–506.

19. Takahashi H, Nakanishi T, Kami K, Arata Y, Shimada I. A novel NMR method for determining the interfaces of large protein–protein complexes. Nat Struct Biol. 2000;7:220–223.

20. Igarashi S, Osawa M, Takeuchi K, Ozawa S, Shimada I. Amino acid selective cross-saturation method for identification of proximal residue pairs in a protein–protein complex. J Am Chem Soc. 2008;130:12168–12176.

21. Wodak SJ, Janin J. Computer analysis of protein-protein interaction. J Mol Biol. 1978;124:323–342.

22. Hwang H, Vreven T, Pierce BG, Hung JH, Weng Z. Performance of ZDOCK and ZRANK in CAPRI rounds 13-19. Proteins. 2010;78:3104–3110.

23. Kozakov D, Hall DR, Beglov D, Brenke R, Comeau SR, Shen Y, Li K, Zheng J, Vakili P, Paschalidis ICh, Vajda S. Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13-19. Proteins. 2010;78:3124–3130.

24. Fernández-Recio J, Abagyan R, Totrov M. Improving CAPRI predictions: optimized desolvation for rigid-body docking. Proteins. 2005;60:308–313.

25. Wang C, Schueler-Furman O, Andre I, London N, Fleishman SJ, Bradley P, Qian B, Baker D. RosettaDock in CAPRI rounds 6-12. Proteins. 2007;69:758–763.

26. Schneidman-Duhovny D, Nussinov R, Wolfson HJ. Automatic prediction of protein interactions with large scale motion. Proteins. 2007;69:764–773.

27. Kinoshita K, Furui J, Nakamura H. Identification of protein functions from a molecular surface database, eF-site. J Struct Funct Genomics. 2002;2:9–22.

28. Kinoshita K, Nakamura H. Identification of protein biochemical functions by similarity search using the molecular surface database eF-site. Protein Sci.

2003;12:1589–1595.

29. Mihalek I, Res I, Lichtarge O. A family of evolution-entropy hybrid methods for ranking protein residues by importance. J Mol Biol. 2004;336:1265–1282.

1　アミノ酸選択的交差飽和法と分子動力学計算を用いたタンパク質複合体の構造決定法

1-1　要旨

核磁気共鳴（Nuclear Magnetic Resonance、以下 NMR とする）による実験法の１つである交差飽和(Cross-Saturation、以下 CS とする)法はスピン緩和実験によりタンパク質複合体の相互作用面を同定することを可能にする。CS 法をさらに発展させた、アミノ酸選択的交差飽和法（Amino Acid Selective Cross-Saturation、以下 ASCS とする）法では、CS ドナー側をアミノ酸選択的に同位体ラベルすることが可能になったことにより、それら選択的にラベルされたアミノ酸の近傍にある CS アクセプター側アミノ酸残基を同定することが可能となった。

これらの ASCS 法の実験データと分子動力学計算による Simulated Annealing（焼きなまし法）を組み合わせて、ASCS 法の実験値を再現する正確なタンパク質複合体の構造を決定する新規の手法を開発した。

この手法を酵母ユビキチン（Ubiquitin、以下 Ub とする）と酵母ユビキチンヒドロラーゼ１（Yeast Ubiquitin Hydrolase 1、以下 YUH1 とする）の複合体に適用したところ、X 線結晶構造により得られた既知複合体構造と非常に高い精度で一致する構造を得ることができた。

1-2　序論

近年の構造生物学の手法の発展により、タンパク質の立体構造の決定はおおいに加速し、ますます多くのタンパク質立体構造が公共データベース Protein Data Bank（以下、PDB とする）[1]に登録され公開されるようになった。

しかしながら、多くのタンパク質は他のタンパク質をはじめとする生体分子と相互作用することで機能しており、単体のタンパク質の立体構造が得られただけでは相互作用によって実現される機能およびそのメカニズムを十分に理解することができない。

よって、タンパク質間の相互作用についても立体構造の観点から明らかにすることが重要である[2-6]。特に、タンパク質が相互作用し複合体を形成したときの構造を決定することができれば、大きな手掛かりとなる。

これまでに X 線結晶構造解析、電子顕微鏡、X 線小角散乱、そして核磁気共鳴（Nuclear

Magnetic Resonance、以下 NMR とする）といったいくつかの実験手法がタンパク質間の相互作用の解析やタンパク質複合体の構造決定に適用されている[7]。

各手法は、それぞれ長所と短所を有しており場合により使い分けられる。X 線結晶構造解析は、タンパク質立体構造を決定する手法として最も広く適用され、分解能の高い構造を決定することができる。タンパク質複合体の場合でもタンパク質単体の場合と同じように適用可能である。しかし、X 線結晶構造解析において最も困難となるのはタンパク質の結晶化であるが、タンパク質複合体の場合は、タンパク質が複合体を形成した状態で結晶化する結晶化条件を見つけることが必要となり、単体のタンパク質の構造決定よりもさらに困難となる。また、タンパク質間相互作用は結合力が弱く、形成される複合体が安定ではないものも多いためこういった場合複合体の結晶化はより困難となる。

一方、NMR によるタンパク質立体構造の決定法もよく利用される手法である。溶液の状態で測定ができるためタンパク質の結晶化の必要がなく、生理的条件下でのタンパク質の構造情報が取得できる点に特徴がある。NMR によるタンパク質立体構造決定における問題点は適用できるタンパク質の大きさに上限があることである。洗練された NMR のパルステクニック[8]と Stereoarray Isotope Labeling (SAIL) [9]のような特異的な同位体ラベリングにより、現在では分子量 30kDa のタンパク質の立体構造を決定することができるようになった[10]。しかしながら、分子量の大きいタンパク質複合体の構造決定はいまだ困難な課題である。

NMR は立体構造（原子座標）を決定するだけの手法ではなく、ケミカルシフトパータベーション法(Chemical Shift Perturbation 法)[11]や水素-重水素交換実験[12]により、タンパク質複合体の相互作用面を見積もることが可能であり現在では一般的な手法となっている。

2000 年に高橋らによって報告された、交差飽和（Cross-Saturation、以下 CS という）法は、CS アクセプター側タンパク質におけるスピン緩和を測定することで、より信頼性の高い相互作用面の情報を与えることが可能な方法である[13,14]（図 1-1）。

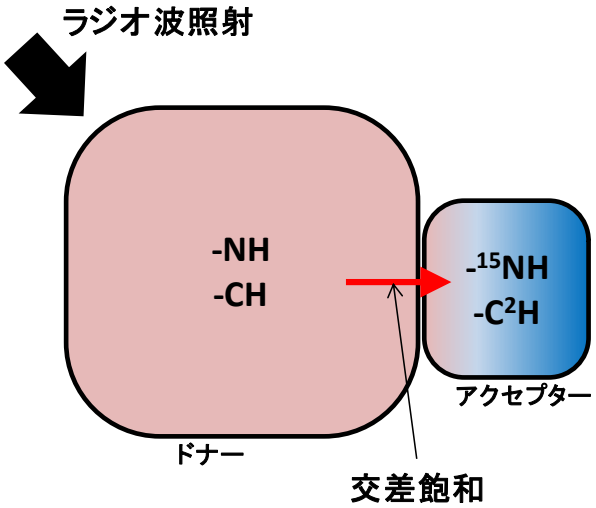

図 1-1

CS 法の原理（模式図）

アクセプター側タンパク質は ^{2}H、^{15}N で標識されており、ドナー側タンパク質は標識されていない。炭化水素プロトンをラジオ波照射するとドナー側炭化水素プロトンの磁化が飽和される。飽和は双極子 - 双極子相互作用により直接照射されていない他のプロトンにも伝播し、相互作用面をとおしてアクセプター側の近傍のプロトンにも及ぶ。アクセプター側のプロトン磁化の飽和は ^{1}H-^{15}N HSQC スペクトルのシグナル強度の減少として観測される。

CS 法では、CS はアクセプター側タンパク質の主鎖の窒素原子とアミドプロトンの ^{1}H-^{15}N HSQC（Hetero-nuclear Single Quantum Coherence：異種核単量子コヒーレンス）スペクトル[15]のシグナルの減少として観測される。^{1}H-^{15}N HSQC は共有結合した水素原子と窒素原子の相関シグナルを検出する方法で、プロトンの磁化が飽和するとシグナルが減少する。よって、ラジオ波を当てた場合と当てない場合のシグナル強度比からプロトン磁化の飽和が検出できる。CS の強さはドナーからの距離に依存しており相互作用面近傍のみに及ぶ。よって、HSQC シグナルのアミノ酸ごとの帰属があれば、シグナル強度の減少したアミノ酸（すなわち相互作用面近傍にあるアミノ酸）がどのアミノ酸残基であるかが同定できることになる。

CS 法では CS アクセプター側タンパク質のシグナルの帰属は必要であるが、CS ドナー側タンパク質のシグナルの帰属は要さないため、CS ドナー側としては高分子量のタンパク質を用いることができる。

従来の CS 法は CS アクセプター側タンパク質のタンパク質—タンパク質相互作用面を同

定することはできたが、CS ドナー側に関する情報はほとんど得ることができなかった。よって、そこから予測されるタンパク質複合体構造としては多様な候補が考えられ、他に十分な情報を与えなければ、候補を絞り込むことができなかった[16,17,18]。

一方、CS 法を発展させた手法である、アミノ酸選択的交差飽和法（Amino acid selective cross-saturation、以下 ASCS とする）法では、CS ドナー側のタンパク質を、アミノ酸選択的に ^{1}H ラベルし、バックグラウンドは ^{2}H ラベルとする[19]。また CS アクセプター側は ^{2}H、^{15}N で均一にラベルする。このラベル法を用いると、CS ドナータンパク質の選択的に ^{1}H ラベルされたアミノ酸のみから交差飽和が起こる。

図 1-2 A)の 3 つの図は、それぞれアミノ酸種 X、Y、Z を選択的にラベルした際の交差飽和実験の結果を模式的に示している。それぞれのラベル実験に対し、CS アクセプタータンパク質のアミドプロトン H1、H2、H3、H4 が交差飽和している。これらの結果をそれぞれ単独で見た場合、CS アクセプター側で交差飽和が観測されても CS ドナー側で交差飽和源となりうるアミノ酸は通常複数あり特定することができない。

しかし図 1-2 B)のようにこれらの実験結果を組み合わせると、CS 源となるアミノ酸残基 X、Y、Z の分布と CS アクセプター側で交差飽和したアミノ酸残基の分布が空間上で対応する組み合わせは極めて限定されることが分かる。

このように、CS ドナータンパク質と CS アクセプタータンパク質の各立体構造と、異なったアミノ酸をラベリングした複数の ASCS 実験の結果を組み合わせた解析を行えば、CS ドナータンパク質中の CS の原因となったアミノ酸残基（以後 CS 源と呼ぶ）となったアミノ酸残基と CS アクセプタータンパク質中で交差飽和を受けたアミノ酸残基の組み合わせをある程度特定することが可能となる。

しかし、組み合わせが絞られただけでは複合体構造は決定できない。ASCS の実験結果を再現しながらも、タンパク質複合体構造として妥当なコンフォメーションエネルギーをもつ構造を構築する必要がある。そこで、次にこれらの CS 源と CS アクセプター側の残基の組み合わせ候補に対応する初期複合体構造から開始して、分子動力学計算による Simulated Annealing（焼きなまし法）を用いて ASCS の実験値を定量的に満たすような複合体構造を構築した。ここで、CS アクセプタータンパク質の各アミノ酸残基の ASCS の値（シグナル強度比）は、複合体構造モデルに基づいてブロッホ方程式を解くことにより精密に求めることができる。よって、実験値と計算値の ASCS の値の差の 2 乗の和をペナルティ関数として分子動力学の系におけるコンフォメーションエネルギーとして加えることで実験値を再現するような構造を探索でき、得られたモデルを定量的に評価することができる。

ASCS 法と分子動力学計算による本構造決定手法を検証するため、タンパク質複合体とし

て酵母ユビキチン(ubiquitin、以下 Ub とする)と酵母ユビキチンヒドロラーゼ 1 の C90S 変異体（以下、単に YUH1 とする。）の複合体を用いた[20]。CS 源となるアミノ酸残基と交差飽和が観測されたアミドプロトンの可能な組み合わせを探索したところ、考えられる全組み合わせのうち、3 次元空間上で相補的な対応となるいくつかの妥当な組み合わせが得られた。これらの組み合わせから得られる Ub と YUH1 の複合体構造候補から分子動力学計算を開始して、最終的に ASCS の実験結果とも合致する複合体構造を得ることができた。得られた複合体構造の精度を X 線結晶構造解析で得られている既知の複合体構造と比較して議論する。

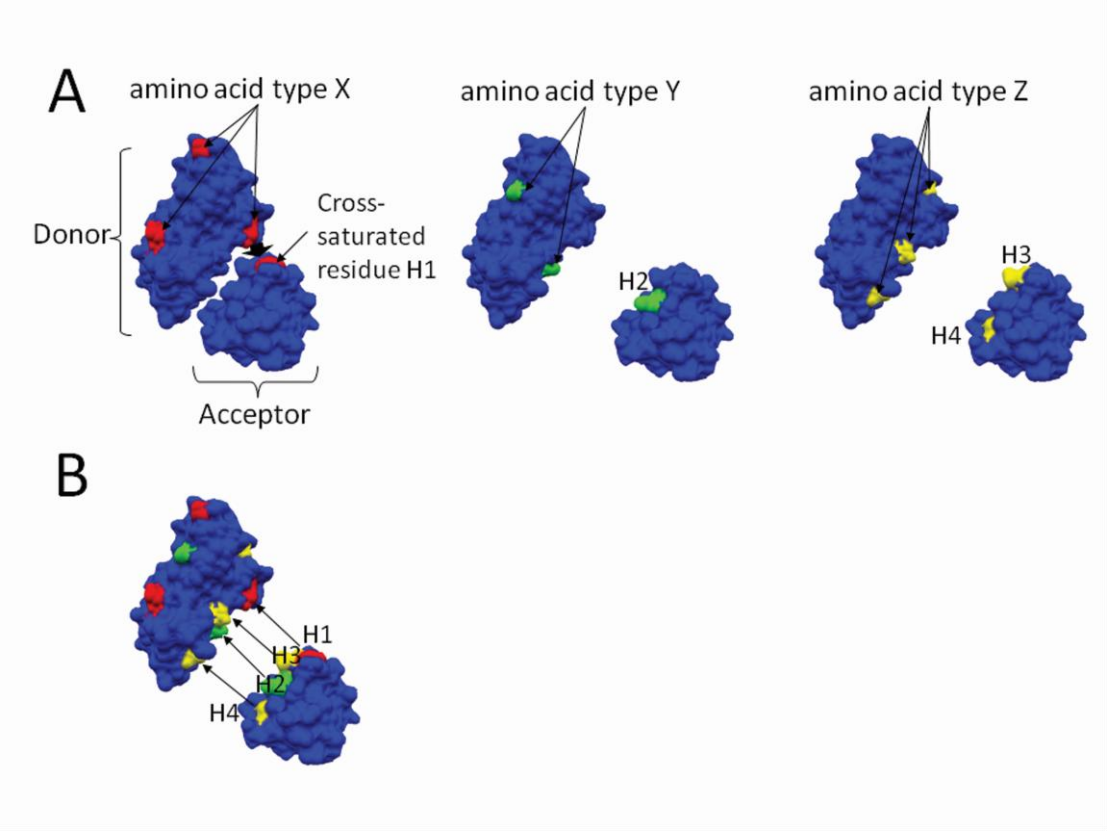

図 1-2

A） ASCS 実験の模式図

各図において、CS ドナー分子側のアミノ酸種 X、Y、Z がそれぞれ選択的に ^{1}H ラベルされている。その他のアミノ酸は ^{2}H となっている。

CS アクセプター側では選択的にラベルされた X、Y、Z の近傍のアミドプロトン H1、H2、H3、H4 のみで交差飽和が観測される。

B) ASCS 実験結果の組み合わせによる近接する残基ペアの推定

CS ドナー分子中で実際に CS の原因となったアミノ酸残基（CS 源）は、複数の ASCS 実験結果を組み合わせ、CS 源候補となるアミノ酸残基と交差飽和したアミドプロトンの空間的な相補性を評価することにより推定できる。

1-3 方法

1-3-1 CS 源となる残基の組み合わせ候補の作成

CS ドナー側タンパク質にアミノ酸種 X、Y、Z がそれぞれ x、y、z 個含まれ、これらを選択的にラベルして ASCS 実験を行った結果、各実験について CS アクセプター側タンパク質で a、b、c 個の残基から顕著な交差飽和が観測されたとする。ドナー側の 1 つの CS 源に対してアクセプター側の複数のアミノ酸残基が近傍にあって交差飽和を受けることもあるので重複を許して計算すると、a+b+c 個の残基に対する CS 源の可能な組み合わせは $x^a \times y^b \times z^c$ 個となる。それらの組み合わせを空間的配置から妥当な順に並べる必要がある。

CS 源となる残基の組み合わせは「最大距離」を用いて次のように評価した。まず、CS ドナータンパク質の立体構造を用いて CS 源候補の各アミノ酸残基に対し、その側鎖の全プロトンの重心を代表点と定義する。一方 CS アクセプタータンパク質側については顕著な交差飽和が観測されたアミドプロトンの位置を用いる。

CS 源となる残基の組み合わせに対して、CS 源代表点とアミドプロトンのペアの距離が計算できる。例えば、CS を受けたアミノ酸残基 a_1、a_2、a_3 について、それぞれ X_1、X_2、X_3 が CS 源として対応する組み合わせでは、a1-X1、a2-X2、a3-X3 の各距離が計算できる。そのうちの最大の距離が最小になるような最適な相対配置を探す。最適配置での「最大距離」をその CS 源となる残基の組み合わせに対する「最大距離」と定義する。

可能な CS 源の組み合わせのうち、より小さい「最大距離」を与える CS 源の組み合わせがもっともらしいと考えることができる。

1-3-2 Ub と YUH1 の初期構造の準備

Ub と YUH1 の立体構造（原子座標）は X 線結晶構造解析により決定された Ub と YUH1 の複合体構造（PDB code:1cmx）を用いて準備した[20]。本研究で検証に用いたタンパク質のアミノ酸配列は 1cmx のものとは僅かに異なる。実際に用いた実験系に合わせて、Ub の 19 番目の Pro、24 番目の Glu、28 番目の Ala の各アミノ酸残基をそれぞれ Ser、Asp、Ser に立体構造モデル上で変異させた。一方、YUH1 は 90 番目のアミノ酸残基 Cys を Ser

に変異させた。Ub の C 末端と YUH1 の 90 番目の Cys の S 原子に形成されていた共有結合は切断した。その後、水素原子を myPresto/tplgene プログラムにより発生させた。AMBER param 96 力場[21]を用いて、この構造を myPresto/cosgene プログラム(myPresto: http://presto.protein.osaka-u.ac.jp/myPresto4/)[22]によりエネルギー極小化した。この複合体構造を本研究中で以後、参照構造として比較の対象とする。

本参照構造は残基番号 1-76 の Ub と残基番号 6-234 の YUH1 からなる。YUH1 のうち残基番号 63-77 の領域は 1cmx の構造には含まれていなかった。よって、YUH1 のモデルはこれらの領域を含まない 2 つのチェーンとして取り扱った。次に、本 Ub-YUH1 複合体構造を Ub と YUH1 の二つに切り離す。

複合体の初期構造は単純な構造重ね合わせによる粗いドッキングで次のように作成する。ASCS 実験は YUH1 のラベルされたアミノ酸残基と Ub の交差飽和が起きたアミノ酸の多くの組み合わせ候補を提示する。それらの組み合わせは、1-3-1 節で述べた方法によってリストアップされる。

Ub と YUH1 の初期複合体構造の候補はそれらの組み合わせに基づいて、Ub の構造と YUH1 の構造のペアとなるアミノ酸残基を重ね合わせることで得られる。その際 YUH1 の CS 源候補残基の側鎖原子と Ub の主鎖プロトンの位置を重ね合わせる。YUH1 の側鎖原子としては、Ala、Leu、Tyr、Pro のアミノ酸種について、それぞれ Cβ、Cγ、Cζ、Cγ原子を代表として用いた。

これらの粗いドッキングで得た複合体には原子間の衝突がみられる。よって、YUH1 分子の重心と Ub 分子の重心を結ぶ方向に、Ub 分子を 15Å平行移動して YUH1 分子から引き離す。これらの複合体をエネルギー極小化した構造を以降の節において初期構造として用いる。

1-3-3　3 次元構造の構築

ASCS の実験値（ラジオ波を照射した時としていない時の HSQC スペクトルのシグナル強度比）を再現する三次元複合体構造を構築するために、以前報告されたアルゴリズム[16,17]を改良し、ASCS の実験値と計算値の差の 2 乗を足し合わせ、ペナルティ関数として分子動力学計算におけるコンフォメーションエネルギーに加えた。ここでその手続きを簡単に述べる。

CS ドナータンパク質（タンパク質 D）の j 番目のプロトン($m+1 \leqq j \leqq n$)がラジオ波照射された時、CS アクセプタータンパク質（タンパク質 A）の i 番目のプロトン($1 \leqq i \leqq m$)に対するブロッホ方程式は次のように書ける[17]。

$$\frac{\mathrm{d}M_z^i}{\mathrm{d}t} = -\rho_i\left(M_z^i - M_0^i\right) - \sum_{k=1,k\neq i}^{m} \sigma_{ik}\left(M_z^k - M_0^k\right) + \sum_{j=m+1}^{n} \sigma_{ij} M_0^j \quad \dots (1)$$

ここで M_z および M_0 はそれぞれ磁化の垂直成分の値およびその平衡値を示す。パラメータσとρはそれぞれ交差緩和速度定数（cross-relaxation rate constant）と自己緩和速度定数（auto-relaxation rate constant）を表す。2 スピン系においてパラメータσとρは次のように与えられる[23]。

$$\sigma_{ij} = \frac{\hbar^2\gamma_i^2\gamma_j^2}{10r^6}\left\{\frac{6\tau_c}{1+\left(\omega_i+\omega_j\right)^2\tau_c^2} - \frac{\tau_c}{1+\left(\omega_i-\omega_j\right)^2\tau_c^2}\right\} \quad \dots (2)$$

$$\rho_{ij} = \frac{\hbar^2\gamma_i^2\gamma_j^2}{10r^6}\left\{\frac{\tau_c}{1+\left(\omega_i-\omega_j\right)^2\tau_c^2} + \frac{3\tau_c}{1+\omega_i^2\tau_c^2} + \frac{6\tau_c}{1+\left(\omega_i+\omega_j\right)^2\tau_c^2}\right\}, \qquad \rho_i = \sum_{i\neq j}\rho_{ij} \quad \dots (3)$$

$$\hbar = h/2\pi$$

これらの等式において、τ_c、γ、ω、h、r はそれぞれ分子相関時間、核磁気回転比、ラーモア周波数、プランク定数、原子間距離をあらわす。上記の式を行列形式で表すと、M_0^iは全スピンについて同じであるため[23]、

$$\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}t} = -\mathbf{R}\cdot\boldsymbol{\eta} + \boldsymbol{\Xi} \quad \dots (4)$$

となる。ここで、

$$\boldsymbol{\eta} = \begin{bmatrix} \dfrac{M_z^1 - M_0}{M_0} \\ \vdots \\ \dfrac{M_z^m - M_0}{M_0} \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} \rho_1 & \sigma_{12} & \cdots & \sigma_{1(m-1)} & \sigma_{1m} \\ \sigma_{21} & \rho_2 & \cdots & \sigma_{2(m-1)} & \sigma_{2m} \\ \vdots & & \ddots & & \vdots \\ \sigma_{(m-1)1} & \sigma_{(m-1)2} & \cdots & \rho_{m-1} & \sigma_{(m-1)m} \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{m(m-1)} & \rho_m \end{bmatrix},$$

$$\boldsymbol{\Xi} = \begin{bmatrix} \displaystyle\sum_{j=m+1}^{n} \sigma_{1j} \\ \vdots \\ \displaystyle\sum_{j=m+1}^{n} \sigma_{mj} \end{bmatrix} \quad \dots (5)$$

である。微分方程式(4)を解くとηは次のようになる[23]。

$$\boldsymbol{\eta} = \{\mathbf{E} - \mathbf{T}\cdot\exp(-\boldsymbol{\Lambda}\cdot t)\cdot\mathbf{T}^{-1}\}\cdot\mathbf{R}^{-1}\cdot\boldsymbol{\Xi} \quad \dots (6)$$

ここで、ΛはR の対角化で得られる対角化された固有値行列である。

$$\boldsymbol{\Lambda} = \mathbf{T}^{-1}\cdot\mathbf{R}\cdot\mathbf{T} \quad \dots (7)$$

一方、M_z^i/M_0はラジオ波を照射した時としていない時の HSQC スペクトルのシグナル強度比（intensity ratio）と等しいためηは実験的にも得ることができる。（η=(intensity

ratio)-1))。よって、ペナルティ関数は擬似ポテンシャルエネルギーE_{ASCS}として、ηの理論値($\eta_{calc.}$)と実験的に観測された値($\eta_{exp.}$)の差の 2 乗の和として計算できる。

$$E_{ASCS} = \frac{1}{2}\sum_{h=1}^{H} W_h \sum_{\substack{\text{acceptor}\\ \text{residues}}} \left(\eta_{calc.} - \eta_{exp.}\right)^2 \quad \dots (8)$$

ここで h は 1 から H までの異なったラベリングによる ASCS の実験を示す。W_hは h 番目の実験の重みとする。

あるタンパク質複合体モデルに対する系の全エネルギーE_{total}は次のように書ける。

$$E_{total} = E_{MM} + E_{ASCS} \quad \dots (9)$$

ここで E_{MM}は通常のコンフォメーションエネルギーである。本研究では、これは AMBER param 96 力場[21]を用いて計算される。分子動力学計算は式(9)を用いて、機能追加した myPresto/cosgene プログラム[22]で実行された。

1-3-4 Simulated Annealing（焼きなまし法）によるモデル構築

この段階での第一の過程は、前述の E_{ASCS}を含んだ E_{total}を用いて、候補複合体モデルの分子動力学計算による Simulated Annealing（焼きなまし法）を行うことである。顕に水分子を考慮しない真空中の計算で溶媒の効果を考慮するため距離依存の誘電率 4R を用いた（R は原子間距離（Å））。1-3-1 節で得られた CS 源の組み合わせに対して 1-3-2 節の手順により作成した初期候補構造それぞれに対して、初期速度を変えることで各 2 つの Simulated Annealing 計算を実行した。冷却のスケジュールは、500 K から 100 K までを 150 psec の間に一定の割合で冷却することとした。静電相互作用とファンデルワールス相互作用の計算は 20 Åのカットオフ計算とした。

タイムステップは 1.0 fs とし、水素原子には SHAKE メソッドを適用して結合長を拘束した。

高温真空中の計算で YUH1 と Ub のそれぞれの構造が大きく壊れることのないよう、それぞれの分子内で原子間の距離を一定の範囲内に保つように拘束ポテンシャルを適用した。この距離による拘束法は神谷らによって開発されたもので[24]、残基間距離拘束と残基内距離拘束の 2 種類からなる。残基間距離拘束は異なった残基間の 6Å以内の主鎖原子（N、Cα、C、O、Cβ）間の拘束である、ただしアミノ酸配列で隣り合う残基間には適用しない。残基内距離拘束は同一残基内で側鎖重原子と主鎖重原子（N、Cα、C）間の距離拘束である。この拘束方法は、適度な柔軟性を許容するため、深いポケットへ分子をドッキングさせるシミュレーションに適している。分子の構造変化を許さないと原子間の衝突によりドッキン

グできないためである。

Ub 分子の C 末端領域は、YUH1 のポケットに結合する重要な領域だが柔軟性が高い。高温中の計算で大きく変化しないよう C 末端 6 残基の主鎖の2面角(ϕおよびφ)に角度の変異を10度以内に押さえるような追加の拘束ポテンシャルを与えた。

Ala、Arg、Leu、Tyr、Pro をそれぞれラベルした 5 つの実験データを用いて、式(8)により E_{ASCS}を計算した。式(8)における重み W_hはすべての実験 h に対して 50 kcal/mol とした。また、シグナル強度比が 0.85 よりも大きいものについては実験ノイズとして計算に加えない。これらの条件は以前に報告された研究[16]と同じ条件である。

1-4　結果と考察

1-4-1　ASCS 実験結果による CS 源残基の組み合わせ候補の作成

YUH1 の 5 種類のアミノ酸（Ala、Arg、Leu、Tyr、Pro）をそれぞれ選択的に ^{1}H 標識した ASCS 実験（Appendix A を参照）の結果、Ub の各アミノ酸のアミドプロトンのシグナル強度比（図 1-3、赤丸）が得られた。

ここで、CS アクセプタータンパク質から、シグナル強度比<0.80 となる顕著な交差飽和が起きたアミノ酸残基を取り出し、それらに対して CS 源となるアミノ酸残基の可能な組み合わせを作成した。Ala、Leu、Tyr、Pro を選択的に標識した実験データについて、それぞれ、2 個（L71 と L73）、3 個(L73、R74、G75)、1 個(G75)、1 個(G75)の計 7 個のアミノ酸残基を取り出した。Arg を選択的に標識したデータについては顕著な交差飽和が起きたアミノ酸残基は観測されなかった。

次に、1-3-1 節で述べた手法を用いて、CS 源アミノ酸残基の可能な組み合わせを発生させた。

YUH1 には 10 個の Ala、17 個の Leu、4 個の Tyr、そして 11 個の Pro があるため、CS 源候補の組み合わせの数は $10^2 \times 17^3 \times 4^1 \times 11^1$=21,617,200 個となる。

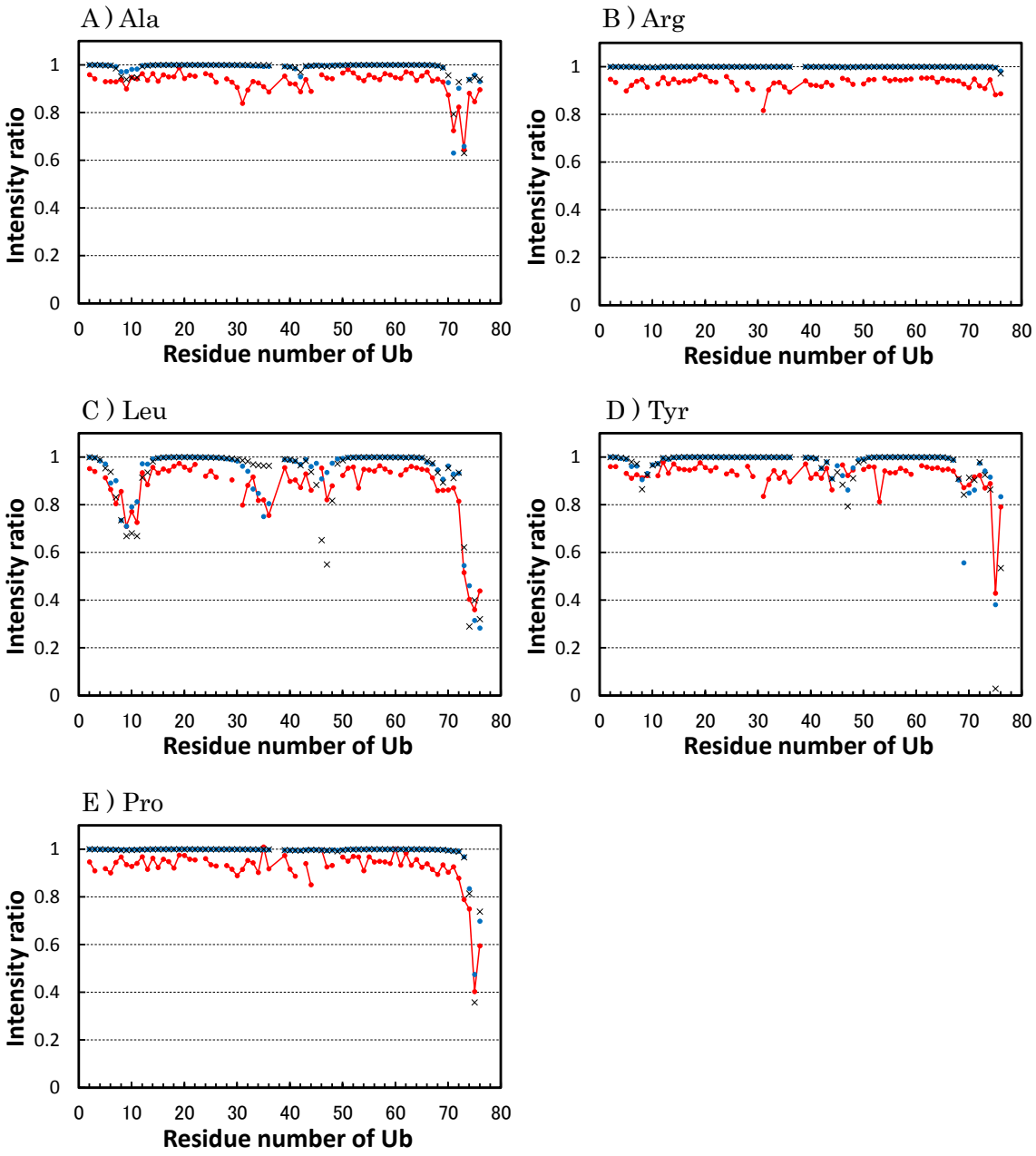

図 1-3

5 種類のアミノ酸を選択的に標識した ASCS 実験で観測されたシグナル強度比。

CS 源として、それぞれ A)Ala、B)Arg、C)Leu、D)Tyr、E)Pro を用いた。

赤丸と実線は実験で観測された値、青丸は最小の E_{ASCS} を示した構造から計算した値、×は参照構造から計算した値を示す。

1-4-2　Ub-YUH1 タンパク質複合体のモデル構築

CS 源候補の組み合わせを「最大距離」の小さいものから並べた。アミノ酸残基のペアのリストに基づいて、1-3-2 節で述べたように、Ub と YUH1 のモデルを粗くドッキングすることで Ub-YUH1 複合体の初期構造が得られる。「最大距離」が 6.0Å未満となる 76 個のペアリストについて初期構造を作成した。

次に、これら 76 個の候補には似た構造も多く含まれるため RMSD（Root Mean Square Deviation：平均二乗偏差、Appendix B-2 を参照）によってクラスタリングすると、図 1-4 および表 1-1 に示すような 42 個の候補に絞られた。

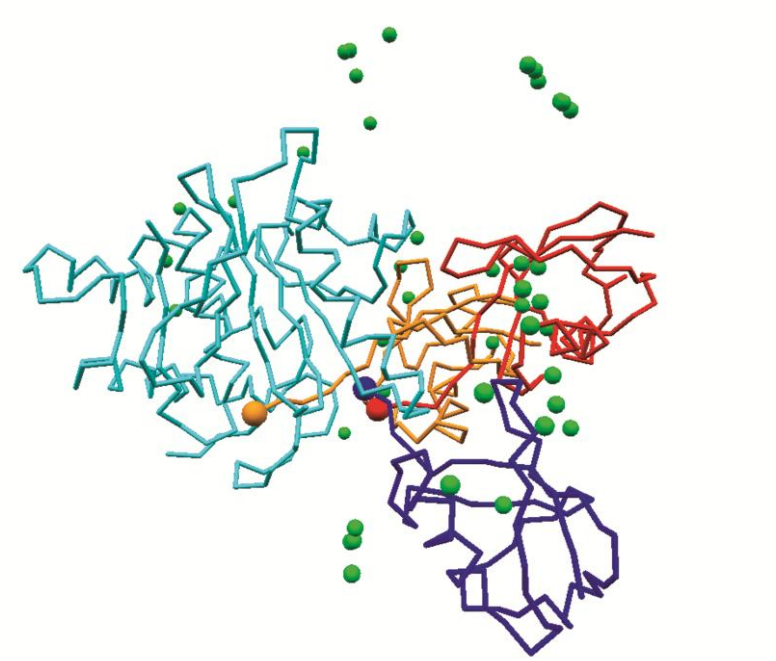

図 1-4

分子動力学計算の初期構造。YUH1 側を固定とし Ub 分子の相対位置を示す。緑の球は 42 個の初期構造における Ub 分子の重心を示す。残基ペアリストの 1 番目および 34 番目から発生させた Ub の初期構造をそれぞれ赤および青で示した。参照構造における YUH1 および Ub 分子をそれぞれ薄青とオレンジで示した。Ub 分子の C 末端の Cα原子をスペースフィルドモデルで示している。

表 1-1

交差飽和源候補のアミノ酸残基(YUH1)と顕著な交差飽和が観測されたアミノ酸残基(Ub)の組み合わせリスト。「最大距離」が小さい上位 42 個。

アミドプロトンが交差飽和されたアミノ酸残基（Ub）	モデル番号	L71	L73	L73	R74	G75	G75	G75	最大距離（Å）
交差飽和源候補のアミノ酸残基(YUH1)	1	A227	A227	L58	L165	L58	Y167	P10	4.809
	2	A227	A227	L165	L58	L58	Y167	P10	4.892
	3	A227	A227	L58	L58	L165	Y167	P10	5.065
	4	A218	A227	L165	L165	L165	Y167	P10	5.365
	5	A218	A227	L165	L58	L165	Y167	P10	5.372
	6	A218	A227	L165	L165	L58	Y167	P10	5.372
	7	A46	A46	L45	L48	L48	Y37	P49	5.393
	8	A46	A46	L48	L45	L48	Y37	P49	5.393
	9	A227	A227	L229	L165	L58	Y167	P10	5.405
	10	A46	A46	L44	L44	L48	Y37	P49	5.408
	11	A46	A46	L45	L44	L48	Y37	P49	5.409
	12	A46	A46	L44	L48	L48	Y37	P49	5.410
	13	A46	A46	L44	L45	L48	Y37	P49	5.415
	14	A218	A227	L58	L58	L165	Y167	P10	5.469
	15	A218	A227	L165	L58	L58	Y167	P10	5.469
	16	A227	A227	L229	L48	L48	Y37	P49	5.596
	17	A227	A227	L58	L229	L58	Y167	P10	5.656
	18	A32	A32	L229	L229	L48	Y37	P49	5.660
	19	A32	A32	L48	L48	L48	Y37	P49	5.828
	20	A227	A227	L165	L229	L58	Y167	P10	5.839
	21	A227	A227	L229	L229	L229	Y37	P15	5.840
	22	A227	A227	L229	L58	L58	Y167	P60	5.845
	23	A227	A227	L229	L165	L58	Y167	P60	5.846
	24	A227	A227	L229	L229	L48	Y37	P15	5.847

25	A32	A32	L229	L229	L229	Y167	P15	5.858
26	A156	A227	L58	L58	L58	Y167	P60	5.884
27	A218	A227	L58	L229	L229	Y167	P15	5.886
28	A227	A227	L229	L39	L48	Y37	P49	5.887
29	A46	A46	L48	L48	L44	Y37	P42	5.888
30	A218	A227	L229	L229	L229	Y167	P15	5.898
31	A46	A46	L45	L45	L45	Y37	P42	5.906
32	A46	A46	L48	L45	L45	Y37	P42	5.906
33	A32	A32	L229	L48	L229	Y37	P15	5.916
34	A227	A227	L58	L58	L229	Y167	P15	5.925
35	A46	A46	L48	L45	L44	Y37	P42	5.934
36	A46	A46	L48	L44	L45	Y37	P42	5.952
37	A32	A32	L229	L229	L48	Y37	P15	5.984
38	A227	A227	L165	L165	L58	Y214	P60	5.992
39	A227	A227	L58	L58	L165	Y167	P60	5.996
40	A227	A227	L165	L58	L58	Y167	P60	5.996
41	A227	A227	L58	L58	L58	Y214	P60	5.996
42	A218	A218	L165	L165	L58	Y214	P60	5.996

図 1-4 をみると、42 個の候補の Ub 分子の重心は YUH1 の周りに幅広く分布していることが分かる。

次に、これらの構造に分子動力学計算による Simulated Annealing を適用して 84 個(42 個の候補について各 2 個ずつ)の構造を作成した。この分子動力学計算において式(8)の E_{ASCS} を計算する際に、ノイズを除くためシグナル強度比<0.85 であった計 28 個のアミノ酸残基を用いた。その内訳は Ala、Arg、Leu、Tyr、Pro を選択的にラベルしたデータについてそれぞれ 5 個、1 個、14 個、4 個、4 個であった。

計算の結果、この 84 個の計算のうち 9 個は参照構造と比較して $RMSD_{total}$ が 5.0Å未満の構造に至った。ここで $RMSD_{total}$ とは計算で得た複合体と参照構造の複合体を、YUH1 分子側で重ね合わせたときに、Ub 分子側の主鎖原子の RMSD の値として定義した。

これらの 9 個の構造は、表 1-1 のリストのうち 1 番目、2 番目、9 番目、22 番目、34 番目、41 番目、42 番目の残基ペアのリストから作成した初期構造から得られた。特に 1 番目と 34 番目の初期構造からは初期速度を変えた 2 つの計算が両方とも $RMSD_{total}$ が 5.0Å未

満の構造に至った。しかしながら、図 1-4 をみるとこれらは初期構造の時点では非常に異なっていることが分かる。この結果は今回の Simulated Annealing の手順が構造空間を十分に探索できるため、最終的に得られるモデルは初期構造にはそれほど強く依存しないことを示唆する。

150 psec の Simulated Annealing 計算終了時、これらの構造の $RMSD_{total}$ はそれぞれ 2.47、4.28、2.53、3.86、2.33、2.73、2.73 Å であった。

図 1-5 は、84 個（42×2 個）の 150 psec のトラジェクトリーから 1.0 psec ずつ構造のスナップショットを取り出し、参照構造に対する $RMSD_{total}$ と、E_{ASCS} 値をプロットしたものである。

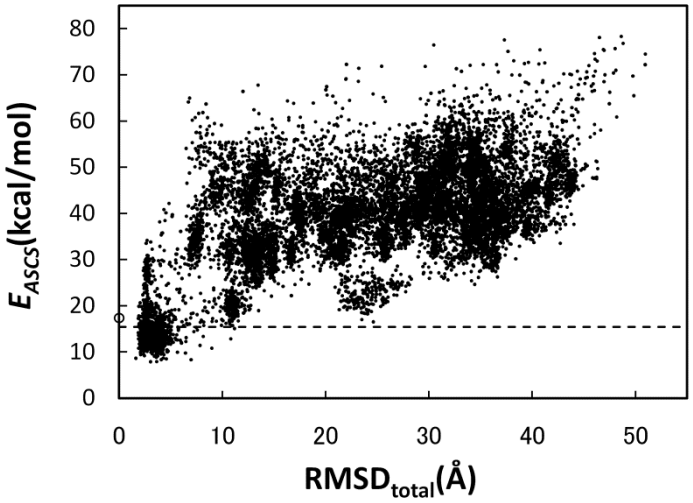

図 1-5

分子動力学計算によりサンプルされた構造群の E_{ASCS} 値と Ub の $RMSD_{total}$ 値の関係。$RMSD_{total}$=0 の位置の○は参照構造から計算された E_{ASCS} 値を示す。破線は E_{ASCS}=15.4kcal/mol を示す。

$RMSD_{total}$ と E_{ASCS} 値の相関係数は 0.61 であり、E_{ASCS} 値が小さい複合体構造は E_{ASCS} 値が大きい複合体構造よりももっともらしい構造であることを示す。図 1-5 中には $RMSD_{total}$ が 2.0–5.0Å で E_{ASCS} 値が 10–20 kcal/mol に相当する位置に 1 つの構造クラスターがあることも分かる。図 1-5 中、最小の E_{ASCS} 値(E_{ASCS} = 7.88 kcal/mol)を持つ構造の $RMSD_{total}$

は 3.05 Åで、図 1-6 にその立体構造を参照構造に重ね合わせて示した。参照構造のタンパク質複合体構造がよく再構成されているのが分かる。

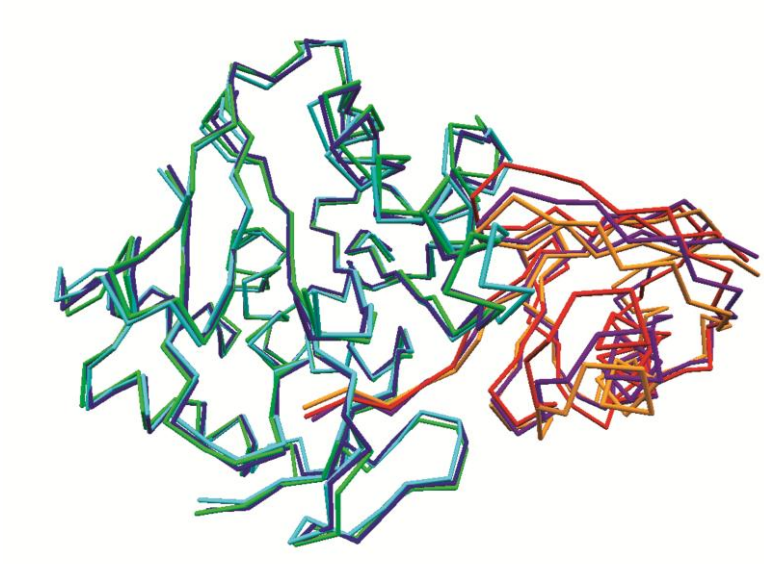

図 1-6

YUH1 と Ub の複合体構造。最小の E_{ASCS} をとった構造の YUH1 と Ub をそれぞれ青および赤で示した。顕わな水分子中で平衡化を行った後の YUH1 と Ub をそれぞれ緑および紫で示した。さらに参照構造の YUH1 と Ub をそれぞれ薄青およびオレンジで示した。全ての複合体は YUH1 を参照構造の YUH1 に重ね合わせてある。

この構造は表 1-1 において 34 番目のペアリストから初期構造を作成し計算を行ったときに、温度 345 K、時間 54 psec の時にサンプルされた構造であった。本構造の、相互作用面の RMSD 値（$RMSD_{interface}$ とする）は 2.81Åで C 末端の RMSD 値（$RMSD_{c\text{-}ter}$ とする）は 0.77Åであった。

ここで、$RMSD_{interface}$ は Ub 分子のアミノ酸残基のうち YUH1 分子から 5Å以内にある 32 アミノ酸残基の主鎖原子の RMSD 値である。また、$RMSD_{c\text{-}ter}$ は Ub の C 末端 8 残基の主鎖原子の RMSD 値である。$RMSD_{interface}$ と $RMSD_{c\text{-}ter}$ はともに YUH1 を参照構造の YUH1 に重ね合わせて計算した。図 1-7 は E_{ASCS} 値と $RMSD_{total}$ の時間変化を示している。E_{ASCS} 値も $RMSD_{total}$ も 100 psec 以内に十分に収束していることが分かる。

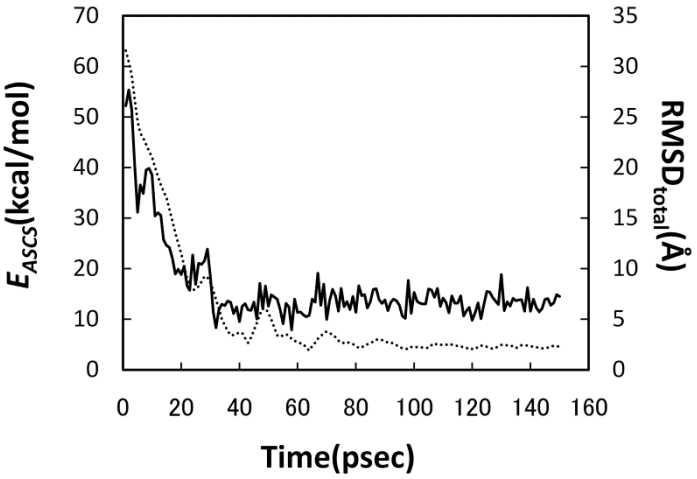

図 1-7

分子動力学計算時の E_{ASCS} 値（実線）と $RMSD_{total}$ 値（破線）の時間変化

最小の $RMSD_{total}$ を持つタンパク質複合体構造は、表 1-1 の 1 番目のペアリストから初期構造を作成して計算をした時に温度 356 K、時間 54 psec の時にサンプルされた。この構造の E_{ASCS} 値は 8.64 kcal/mol で、$RMSD_{total}$ は 1.63Å、$RMSD_{interface}$ は 1.60Å、$RMSD_{c\text{-}ter}$ は 0.56Åであった。

図 1-3 中の青丸は最小の E_{ASCS} 値をとった構造から計算される、Ub 分子の各アミノ酸残基のシグナル強度比を示す。比較のため併せて参照構造から計算されるシグナル強度比も図中×で示す。参照構造から計算される E_{ASCS} 値は、17.35 kcal/mol であった。よって、図 1-5 中に示したように E_{ASCS} 値の理論的な下限は 17.35 kcal/mol 程度であると考えられる。このとき式（8）を用いると残基 1 つあたりのシグナル強度比の平均の誤差は0.16程度であると見積もられる。

顕わな水中でさらに分子動力学計算を行うことで、複合体構造の精度をさらに改善することができる。最小の E_{ASCS} 値をとった複合体構造を、半径 42Åの水分子のドロップの中に入れ、温度 300 K で 300 psec の分子動力学計算を行った。この計算において原子間距離による拘束は適用したが E_{ASCS} 項は計算に含めていない。ただし原子間距離による拘束は、より分子の柔軟性を許容するため残基間の拘束のみとし、柔軟な構造をとると考えられる Ub 分子の C 末端領域には拘束を適用していない。

図 1-6 中の緑および紫のモデルがこの分子動力学計算の最終構造を示す。$RMSD_{total}$ は

1.44Å、$RMSD_{interface}$は1.32Å、$RMSD_{c\text{-}ter}$は0.87Åであった。

このように顕わな水中で行った計算により、真空中でSimulated Annealingにより得た構造の精度をさらに高めることができた。これは、溶媒モデルが改善されたためであると考えられる。

しかしながら、同じ顕わな水中での計算において、E_{ASCS}項も計算に含めた場合には、最終構造は、$RMSD_{total}$は3.00Å、$RMSD_{interface}$は2.68Å、$RMSD_{c\text{-}ter}$は0.99Åとなり、構造を改善することはできなかった。

結果として、E_{ASCS}項の導入は、計算において複合体形成の過程を加速することができるが、通常の力場に加えE_{ASCS}項による力が重ねて働くことで複合体構造を歪めてしまうことがあると考えられる。

以上の計算は、複合体構造をいわゆるbound-boundドッキングにより構築した。つまり、両タンパク質分子の初期モデルとしてそれぞれ結合時の構造を用いた。それらはインデュースドフィット(誘導適合)によりすでに構造変化したものである。本研究で提案する手法を実用上使用するためには、unbound(非結合状態)構造を用いて、複合体構造を構築する必要がある。本手法は分子動力学計算を使用することで、タンパク質分子の柔軟性を許容する手法であるため、インデュースドフィットの効果もうまく表現されるであろうことが期待できる。

そこでunbound-boundドッキングの場合についても有効性のテストを行った。Ubの初期構造としてunbound構造(PDB code:1ubq)を使用して[25]、YUH1のbound構造へのドッキングを試みた。Ubのunbound構造とbound構造の違いは、重原子のRMSD値、Cα原子のRMSD値、C末端6残基の重原子のRMSD値、C末端6残基のCα原子のRMSD値がそれぞれ1.68Å、1.02Å、2.30Å、1.81Åである。

Ub分子のunbound状態の構造モデルは、1-3-2節で述べたbound状態の構造モデルと同様の手順で用意した。

つまり、三か所の変異(P19S、E24D、A28S)を立体構造モデルに手動で導入して、原子間距離拘束も同様とした。このunbound状態のUb分子モデルを、bound-boundのケースで用意したYUH1から15Å引き離した初期構造に重ね合わせた。初期構造は表1-1の34番目のペアリストのものを用いた。その後、bound-bound ドッキングのケースで説明したものと同じ条件で分子動力学計算によるSimulated Annealingを実行した。

図1-8はE_{ASCS}と$RMSD_{total}$の時間変化を示す。分子動力学の最終構造は図1-9に示したようになり、$RMSD_{total}$は4.06Å、$RMSD_{interface}$は3.82Å、$RMSD_{c\text{-}ter}$は2.03Åとなった。

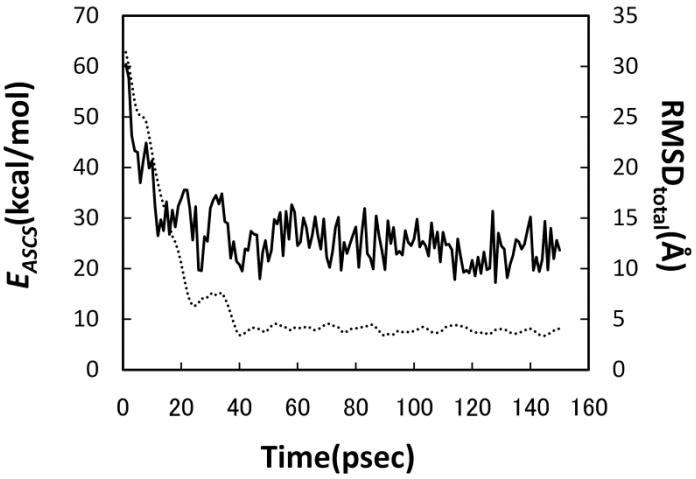

図 1-8

bound-unbound モデルを初期構造に用いた分子動力学計算の E_{ASCS} 値（実線）と $RMSD_{total}$ 値（破線）の時間変化

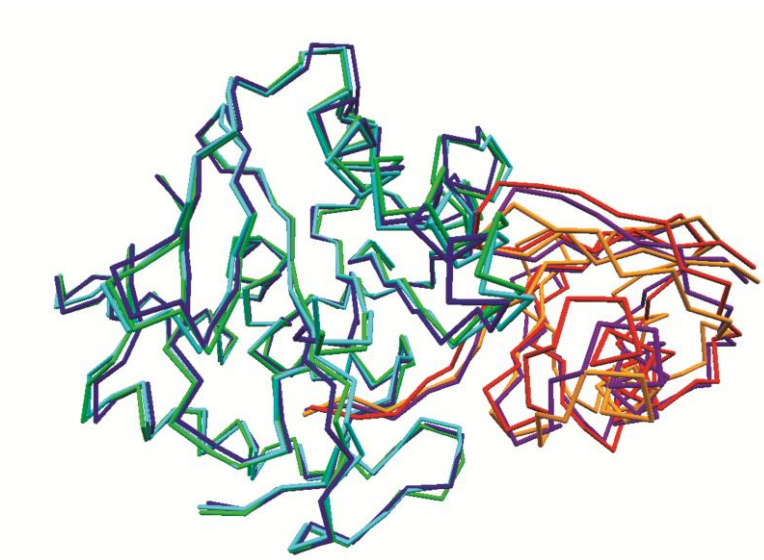

図 1-9

bound-unbound モデルから作成した YUH1 と Ub の複合体構造。YUH1 と Ub をそれぞれ青および赤で示した。顕わな水分子中で平衡化を行った後の YUH1 と Ub をそれぞれ緑および紫で示した。さらに参照構造の YUH1 と Ub をそれぞれ薄青およびオレンジで示した。全ての複合体は YUH1 を参照構造の YUH1 に重ね合わせてある。

さらに、bound-bound のケースで述べた手順と同じようにして、顕わな水中での分子動力学計算を実行すると、$RMSD_{total}$ が 2.67Å、$RMSD_{interface}$ が 2.43Å、$RMSD_{c\text{-}ter}$ が 0.83 Åとなり、構造がさらに改善された。この最終構造も図 1-9 中に示した。

この結果は、今回の Simulated Annealing の手順が構造の柔軟性を許容しているため、bound 状態の構造を使った場合でも、unbound 状態の構造を使った場合でも正確にタンパク質複合体構造を予測できることを示す。

この研究では、タンパク質の柔軟性を許容する新規手法を用いて、タンパク質複合体構造を高い精度で予測することができた。Ub の C 末端部分にインデュースドフィットによる構造変化が見られるため、剛体モデルによるドッキング手法では十分な精度で複合体構造が予測できない可能性がある。

ASCS のシグナル(η)は原子間距離に強く依存する。すなわち、η 値は r^{-6} に比例するため原子が衝突するほど接近すると $\eta_{calc.}$値に大きな誤差をもたらす。今回は分子動力学計算を使っているため、分子力場によりこの誤差を避けることができた。

タンパク質―タンパク質ドッキングプログラムとしては HADDOCK[26]というプログラムが知られている。このプログラムでは、CS 実験や、化学シフトパータベーション、変異体実験などにより得られたタンパク質―タンパク質相互作用のデータを ambiguous interaction restraint (AIR) として利用することができる。AIR とは、原子間距離の拘束方法の一種である。しかしながら、HADDOCK はηそのものを計算するものではない。一方、本研究の手法はタンパク質複合体モデルから直接η値の正確な計算を行う。それゆえ、Simulated Annealing 計算時に、$\eta_{calc.}$と$\eta_{exp.}$を直接比較することができる。これによって、計算した構造が実験結果を再現しているかどうかの定量的評価が可能である。

1-4-3 E_{ASCS} と RMSD の関係

今回の Ub-YUH1 複合体の系においては、E_{ASCS}値が最小となる構造は天然状態の複合体構造をよく再現した。

しかし、本手法を適用して、未知の複合体構造を構築するためには、予測された複合体構造が、ASCS 実験の結果と十分一致していることを示すための E_{ASCS}の基準値が必要である。そこで、図 1-5 で示した構造の $RMSD_{interface}$ と E_{ASCS}の関係をさらに解析した。

ここで、E_{ASCS}の大きさは式(8)の定義から分かるように計算に用いる残基数に依存するため、次式(10)であらわされるηの2乗誤差の残基平均である$<(\eta_{calc.}-\eta_{exp.})^2>$を計算することとする。

$$\langle(\eta_{calc.}-\eta_{exp.})^2\rangle \equiv \frac{2E_{ASCS}}{(\text{Number of residues})\times \text{W}} \quad \ldots (10)$$

ここで、YUH1-Ub の系では、Number of residues = 28、重み W = 50 kcal/mol であった。サンプルされた構造を $\text{RMSD}_{\text{interface}}$ の値で 1Åから 7Åの 1Å区切りのクラスに分けると、各クラスの$<(\eta_{\text{calc.}}-\eta_{\text{exp.}})^2>$の分布は図1-10に示すようになった。1–2Å、2–3Å、3–4Åのクラスの分布のピークは 0.02 のほぼ同一の位置に観測された。その他のクラス、特に5–6Å、6–7Åのクラスのピークは有意に 0.04 より$<(\eta_{\text{calc.}}-\eta_{\text{exp.}})^2>$が大きい位置に観測された。同一の位置にピークが観測されたクラス1–2Å、2–3Å、3–4Åを統合すると、クラス1–4Åの$<(\eta_{calc.}-\eta_{exp.})^2>$の平均値(m)は 0.022、標準偏差(σ)は 0.007 となった。

結果として、サンプルされた構造の$<(\eta_{calc.}-\eta_{exp.})^2>$がこの平均値 m(=0.022)よりも小さいか、あるいは大きくても m + σ(=0.029)程度であれば、その構造は ASCS の実験結果と矛盾がないと考えることができる。

今回の系では、図 1-5 に示したように平均値 m(=0.022)は E_{ASCS} = 15.4 kcal/mol に対応する。ここで、E_{ASCS} が 15.4 kcal/mol を下回るような構造の集団は天然構造に近いもの以外には観測されていない。

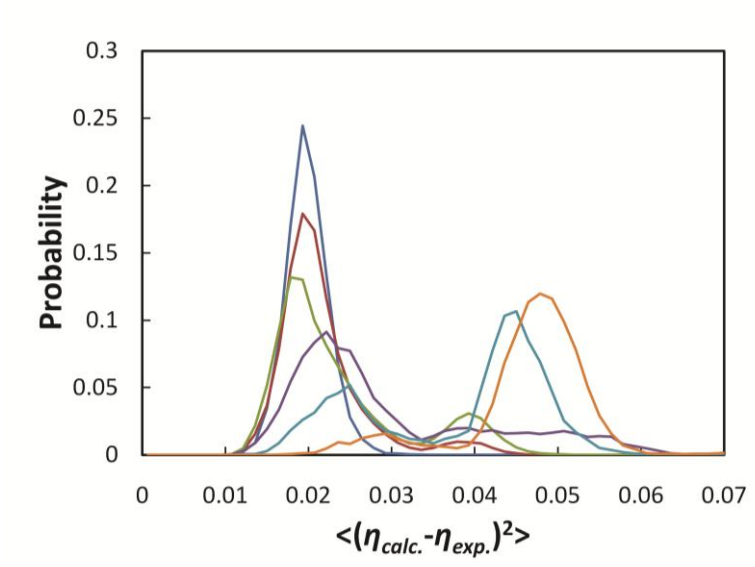

図 1-10

$\text{RMSD}_{\text{interface}}$ により区分した時の$<(\eta_{calc.}-\eta_{exp.})^2>$の分布。

$\text{RMSD}_{\text{interface}}$ による区分 1-2Å、2-3Å、3-4Å、4-5Å、5-6Å、6-7Åにおける分布をそれぞれ青、赤、緑、紫、シアン、オレンジの線で示した。

1-4-4　ASCS 法の適用可能性

ASCS 法は、タンパク質複合体において、分子間の近接するアミノ酸残基ペアの情報を得るための一般的な手法である。この手法を用いるための条件として(1)CS ドナータンパク質のアミノ酸選択的安定同位体標識および ^{2}H、^{15}N で一様に標識した CS アクセプタータンパク質。(2)CS ドナータンパク質との複合体形成時の CS アクセプタータンパク質の主鎖の NMR シグナルの帰属。(3)CS アクセプタータンパク質と CS ドナータンパク質の単体の立体構造の 3 点を必要とする[19]。

アミノ酸選択的ラベリングは、11 種類のアミノ酸(Ala、Arg、His、Ile、Leu、Lys、Met、Phe、Pro、Tyr、Trp)については高効率で行うことができるため[19]、ほとんどのタンパク質複合体のケースに適用し、CS ドナータンパク質の接触面を同定することが可能であると考えられる。実際に、PDB に登録されている既知のタンパク質複合体構造で、タンパク質ドッキング評価のベンチマークセット[27]に含まれる 88 個のタンパク質複合体について調べたところ、少なくとも 5 残基以上の CS 源となりうるアミノ酸残基が CS アクセプター分子との相互作用面近傍に存在することが分かった。ここで相互作用面近傍に存在する CS 源となりうるアミノ酸残基の平均数は 8.9 であった。この手法は上記の(1)-(3)の要件を満たせば多くのタンパク質―タンパク質相互作用に適用できる手法であるが、本質的に 2 つの制限がある。分子のサイズが大きい場合と複合体形成時の構造変化が大きい場合である。

一つ目の問題については、構造データベース中のタンパク質―タンパク質複合体についての解析により、タンパク質の表面全体に存在するアミノ酸残基数に対する、相互作用の面に存在するアミノ酸残基数の割合は、タンパク質のサイズが大きくなるにつれて減少する傾向がある[28]。よって、タンパク質の分子量が大きい場合、立体構造上同じようなアミノ酸残基の分布が現れる可能性が増すことになり、複数の相互作用部位候補から正しいものを一つに限定することが難しくなる。

解析の結果、タンパク質のサイズが 40 kDa より大きいと似たアミノ酸配置が複数現れる傾向があることが分かった。そういった場合、これらアミノ酸残基の配置が似た相互作用部位候補も ASCS 実験の結果を満たすようになるため、唯一のものとして適切なモデルを構築することが難しくなると考えられる。

そういった場合には、さらにサイト特異的なラベリング手法や変異実験による実験データなどを使うことにより、複数の相互作用部位候補から適切な部位を限定するか、複数の複合体構造候補を得た後に、実験結果を満たす唯一の候補を選択する必要がある。

二つ目の問題について、今回提案する手法は 1-3-3 節でも述べたように、疑似ポテンシャルエネルギーE_{ASCS}を使った分子動力学計算によるため、複合体形成時にインデュースドフ

ィットによりいくらかの構造変化がある場合にも適用できる。

しかしながら、intrinsically disordered protein [29]においてよく見られる、インデュースドフォールディングと呼ばれる二次構造の変化を伴うような大きな構造変化がある場合にはこの手法を用いることができない。そのような場合には相互作用面のアミノ酸残基の化学シフトが大きく変わってしまうと思われる。さらなる実験を行い相互作用部位の新たな構造モデルを構築する必要がある。

1-5　結論

アミノ酸選択的交差飽和(ASCS)法の実験データと分子動力学計算を組み合わせた新しいタンパク質—タンパク質複合体構造構築法を開発した。

タンパク質—タンパク質複合体において、一方のタンパク質中の交差飽和ドナーとなるアミノ酸残基を選択的にラベルする。NMR の交差飽和実験によって、相互作用するタンパク質中の交差飽和アクセプターとなるアミノ酸残基を観測することができる。

分子動力学計算において構造モデルからアミノ酸残基の ASCS の値ηを直接計算することができる。実験で得られた ASCS の値ηと構造モデルから計算によって得られた ASCS の値ηの差をペナルティ関数として分子動力学計算を行う。そのペナルティ関数によるエネルギーと通常の力場によるエネルギーの総和が最小化されることにより最適なタンパク質—タンパク質複合体モデルを得ることができる。

ASCS 実験データに基づいた荒いドッキングにより、複数のタンパク質複合体構造の候補を作成できる。その後これらの構造をクラスタリングすることにより候補が絞られる。これらの候補構造をそれぞれ一定の距離引き離して配置し、分子動力学計算による Simulated Annealing の初期構造として用いる。分子動力学計算の結果、多数のタンパク質複合体構造が得られる。これらの構造のうち実験値に最も近い ASCS の値ηを示す構造をもっともらしい構造として選択する。

この手法をユビキチンと酵母ユビキチンヒドロラーゼ 1 (Ub-YUH1)複合体構造に適用したところ、ASCS 実験データを用いることで高い精度で構造を決定することができた。

実行した 84 個の分子動力学計算のうち 9 個は $RMSD_{total}$ の値が 5Å未満となり、X 線結晶構造解析で得られた構造に類似した構造に至った。さらにその後、顕わな水分子中で分子動力学計算を行うことにより $RMSD_{total}$ 値が 1.44Åとなり、より正確な構造を得ることができた。

初期構造として unbound-bound 構造を用いた場合でも高精度で正しい複合体が計算可

能であった。予測された構造の $RMSD_{total}$ 値は 2.67 Å であった。

参考文献

1. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nat Struct Biol. 2003;10:980.

2. Pons C, Grosdidier S, Solernou A, Pérez-Cano L, Fernández-Recio J. Present and future challenges and limitations in protein–protein docking. Proteins. 2010;78:95–108.

3. Smith GR, Sternberg MJ. Prediction of protein–protein interactions by docking methods. Curr Opin Struct Biol. 2002;12:28–35.

4. Janin J. Assessing predictions of protein–protein interaction: the CAPRI experiment. Protein Sci. 2005;14:278–283.

5. Lensink MF, Méndez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd Edition. Proteins. 2007;69:704–718.

6. Kanamori E, Murakami Y, Tsuchiya Y, Standley DM, Nakamura H, Kinoshita K. Docking of protein molecular surfaces with evolutionary trace analysis. Proteins. 2007;69:832–838.

7. van Dijk AD, Boelens R, Bonvin AM. Data-driven docking for the study of biomolecular complexes. FEBS J. 2005;272:293–312.

8. Pervushin K, Riek R, Wider G, Wüthrich K. Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA. 1997;94:12366–12371

9. Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A, Güntert P. Optimal isotope labelling for NMR protein structure determinations. Nature. 2006;440:52–57.

10. Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature. 2007;445:618–622.

11. Foster MP, Wuttke DS, Clemens KR, Jahnke W, Radhakrishnan I, Tennant L, Reymond M, Chung J, Wright PE. Chemical shift as a probe of molecular interfaces: NMR studies of DNA binding by the three amino-terminal zinc finger domains from transcription factor IIIA. J Biomol NMR. 1998;12:51–71.

12. Paterson Y, Englander SW, Roder H. An antibody binding site on cytochrome c defined by hydrogen exchange and two-dimensional NMR. Science. 1990;249:755–759.

13. Shimada I. NMR techniques for identifying the interface of a larger protein–protein complex: cross-saturation and transferred cross-saturation experiments.

Methods Enzymol. 2005;394:483–506.

14. Takahashi H, Nakanishi T, Kami K, Arata Y, Shimada I. A novel NMR method for determining the interfaces of large protein–protein complexes. Nat Struct Biol. 2000;7:220–223.

15. Bax A, Ikura M, Kay LE, Torchia DA, Tschudin R. Comparison of different modes of two-dimensional reverse-correlation NMR for the study of proteins. J Magn Reson. 1990;86:304–318.

16. Matsuda T, Ikegami T, Nakajima N, Yamazaki T, Nakamura H. Model building of a protein–protein complexed structure using saturation transfer and residual dipolar coupling without paired intermolecular NOE. J Biomol NMR. 2004;29:325–338.

17. Matsuda T, Nakajima N, Yamazaki T, Nakamura H. CAD-ICAD complex structure derived from saturation transfer experiment and simulated annealing without using pairwise NOE information. J Mol Recognit. 2004; 17:41–50.

18. Quadt-Akabayov SR, Chill JH, Levy R, Kessler N, Anglister J. Determination of the human type I interferon receptor binding site on human interferon-alpha2 by cross saturation and an NMR-based model of the complex. Protein Sci. 2006;15:2656–2668.

19. Igarashi S, Osawa M, Takeuchi K, Ozawa S, Shimada I. Amino acid selective cross-saturation method for identification of proximal residue pairs in a protein–protein complex. J Am Chem Soc. 2008;130:12168–12176.

20. Johnston SC, Riddle SM, Cohen RE, Hill CP. Structural basis for the specificity of ubiquitin C-terminal hydrolases. EMBO J. 1999;18:3877–3887.

21. Kollman P, Dixon R, Cornell W, Fox T, Chipot C, Pohorille A. The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In: van Gunsteren WF, Weiner PK, Wilkinson AJ, editors. Computer simulations of biomolecular systems. The Netherlands: Kluwer/Escom; 1997. pp 83–96.

22. Fukunishi Y, Mikami Y, Nakamura H. The filling potential method: a method for estimating the free energy surface for protein-ligand docking. J Phys Chem B. 2003;107:13201–13210.

23. Dobson CM, Olejniczak ET, Poulsen FM, Ratcliffe RG. Time development of proton nuclear overhauser effects in proteins. J Magn Reson. 1982;48:97–110

24. Kamiya N, Yonezawa Y, Nakamura H, Higo J. Protein-inhibitor flexible docking

by a multicanonical sampling: native complex structure with the lowest free energy and a free-energy barrier distinguishing the native complex from the others. Proteins. 2008;70:41–53.

25. Vijay-Kumar S, Bugg CE, Cook WJ. Structure of ubiquitin refined at 1.8 Å resolution. J Mol Biol. 1987;194:531–544.

26. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc. 2003;125:1731–1737.

27. Hwang H, Pierce B, Mintseris J, Janin J, Weng Z. Protein–protein docking benchmark version 3.0. Proteins. 2008;73:705–709.

28. Chen H, Zhou HX. Prediction of interface residues in protein–protein complexes by a consensus neural network method: test against NMR data. Proteins. 2005;61:21–35.

29. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol. 2005;6:197–208.

2　Evolutionary Trace 法を用いた分子表面によるタンパク質複合体構造の予測

2-1　要旨

タンパク質分子表面の形状相補性と Evolutionary Trace 法（進化追跡法。以下、ET 法とする）を用いて得られたアミノ酸配列の保存度を用いた新規のタンパク質―タンパク質複合体の構造予測法を開発した。

複合体構造の予測は保存度の重みをつけた形状相補性を評価する目的関数を最適化することによって行う。目的関数の最適化は、遺伝的アルゴリズムとモンテカルロ法による構造サンプリングにより行った。

この手法をタンパク質複合体予測の国際的なコミュニティである CAPRI のターゲット（ブラインドテストの課題となる複合体構造）に適用して性能を評価したところ、いくつかのターゲットについて天然状態に近い構造を予測することができた。

さらに、複合体構造予測における、ET 法の有効性を調べたところ、アミノ酸配列の保存度の情報は限られた種類のタンパク質（シグナル伝達関連と酵素タンパク質）について有用であることがわかった。

2-2　序論

タンパク質構造の情報はタンパク質の分子機能を理解するために極めて重要である。タンパク質の機能は通常他のタンパク質または低分子との相互作用を通して実現されるため、タンパク質と他の分子の結合様式を予測することは、タンパク質の生物学的機能を理解するための第一段階となる。

この目的のため、これまで多くの手法が提案されてきた[1,2]。ひとつの方法は、タンパク質立体構造の予測に threading 法が用いられているように、すでに知られた複合体に対して類似性探索を行うことである。

しかしながら、単体のタンパク質立体構造と比べ、既知のタンパク質―タンパク質複合体の数は少なく、特に PDB に登録されているヘテロ複合体構造は 8654 件である[3]。

よって、タンパク質―タンパク質複合体予測の別の有望な方法として、タンパク質―タンパク質ドッキング計算が期待されている。

これまでにタンパク質―タンパク質ドッキングのためのいくつかの手法が提案されてお

り、ZDOCK[4]、Haddock[5]、Molfit[6]、RosettaDock[7]などのプログラムソフトが利用可能となっている。

これらの手法はいくつかのケースにおいて正しい構造を予測することに成功しており、精度は徐々に向上してきているが、これまでの CAPRI（Critical Assessment of PRedicted Interactions、ブラインドテストによりタンパク質複合体構造予測法の評価を行う国際的なコミュニティ。詳細は Appendix B を参照）の結果[8,9]によるとどのソフトもすべてのケースについて正しい予測ができるわけではない。

本研究では、タンパク質の分子表面に基づいた新しいタンパク質―タンパク質複合体の予測法について述べる。

分子表面の形状を用いることで低分子の結合の予測が可能であることはこれまでの研究で報告されている[10,11]。タンパク質―タンパク質ドッキングに応用するために、分子表面の形状、物理化学的特性と進化的保存度の情報を組み合わせた方法を開発した。進化的保存度は Real-valued Evolutionary Trace 法（以下単に ET 法とする）[12]（詳細は Appendix C を参照）により計算され、分子表面のうちのホットスポットを推定するために利用できる。

予測された複合体構造はエネルギー極小化を行い、さらに相互作用面の静電ポテンシャル、疎水性、形状の相補性により評価した。

2-3　方法

2-3-1　概観

タンパク質表面のペアのドッキングは、目的関数の最適化によって行う。目的関数は Connolly のアルゴリズム[13]で作成された分子表面の頂点（図 2-1）を用いた関数で、頂点に最近接するアミノ酸残基の配列保存度と形状の相補性を評価する関数として定義した。

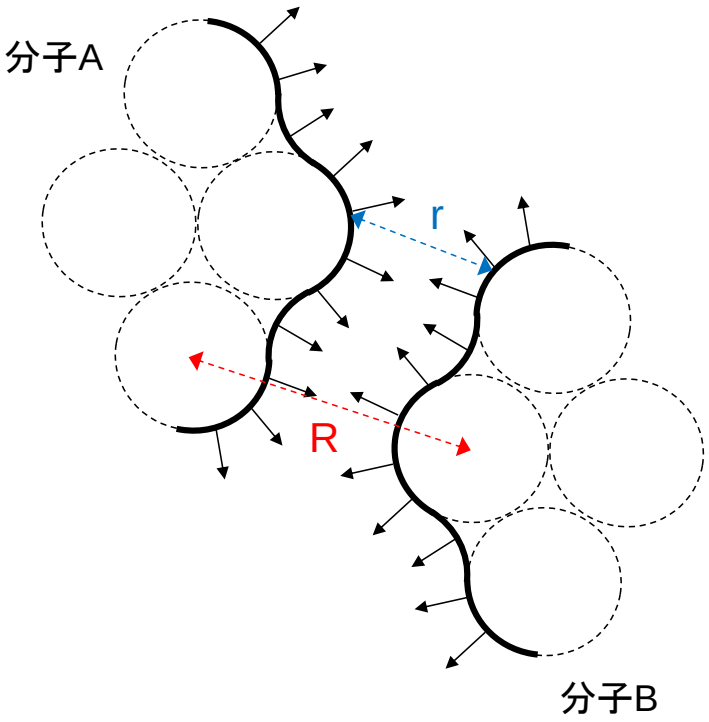

図 2-1　分子表面を用いた目的関数の定義

相互作用する 2 つのタンパク質を分子 A、分子 B とする。各分子の原子を vdW（van der Waals）半径の破線の円で示し、分子表面を曲線で示した。分子表面は分子の表面を原子と衝突しないように探査球を転がすことによって作成される。分子表面は離散化された頂点の集まりで表現され、目的関数の定義はこれらの頂点を用いる。図中で各頂点を矢印で示した。各頂点は位置座標（図中矢印の始点）と単位法線ベクトル（図中矢印の向き）の値をもつ。目的関数における頂点間の距離および原子間の距離をそれぞれ r および R で示す。

目的関数の最適化は遺伝的アルゴリズムとモンテカルロ法によるサンプリングを用いて行った。すなわち前者による大域的な探索と後者による局所探索を組み合わせて行う。

まず、遺伝的アルゴリズムの過程で相補的な表面の組み合わせを探索して構造の初期候補を与え、その後、モンテカルロ法の過程ではその構造に対して、局所的な配置の最適化を行う。

このアルゴリズムを目的関数のスコアの差が 10 サイクル続けて 0.01 未満になるまで繰り返し適用し複数の候補複合体構造を得た。

さらにそれらの候補を静電ポテンシャル、疎水性、形状の相補性により、さらに評価した。

2-3-2　目的関数

相互作用する 2 つのタンパク質分子 A および B からそれぞれ作成した分子表面をドッキングさせるとする。2 つの項からなる目的関数を次式のように定義し、値 F を最大化する構造を探索するものとする。ここで a と b は適切な重みとする。

$$\mathrm{F} = \mathrm{aF_1} + \mathrm{bF_2} \quad \ldots (1)$$

関数の第一項 F_1 は本手法の主要な項で、タンパク質分子表面を表す頂点（図 2-1 中の矢印）を用いて分子表面の形状相補性を評価する。その際、頂点が由来するアミノ酸残基の保存度と頂点ペアの距離によるダンピングファクターを考慮する。

$$\mathrm{F_1} = \sum_{vertex\ pairs} I^2 \cdot (c_A c_B)^2 \cdot D \quad \ldots (2)$$

ここで、

$$\mathrm{I} = \frac{-\mathbf{n}_\mathrm{A} \cdot \mathbf{n}_\mathrm{B} + 1}{2} \quad \ldots (3)$$

とする。 $\mathbf{n}_A$ および $\mathbf{n}_B$ はそれぞれ頂点ペアをつくる分子表面 A、B 上の頂点の法線ベクトルとする。c_A、c_B はそれぞれ分子表面 A、B 上の頂点の保存度スコアとする。D はダンピングファクターで、

$$\mathrm{D} = \begin{cases} \dfrac{1}{4\mathrm{r}_\mathrm{AB}^2}, & (\mathrm{r_{AB}} > 0.5) \\ 1.0, & (\mathrm{r_{AB}} \leq 0.5) \end{cases} \quad \ldots (4)$$

とする。r_{AB}（図 2-1 中の r）は頂点ペア間の距離（Å）である。

式(2)は分子表面 A、B から 1 つずつ頂点を取り出してできる全ての頂点ペアについての和となっている。ただし、モンテカルロ法による最適化時には、計算時間を短縮するため必要に応じて、カットオフテーブルを用いてカットオフ計算を行う。

それぞれの頂点の保存度スコアの値 c はその頂点にもっとも近い原子の属するアミノ酸残基の保存度スコアを用いる。アミノ酸残基の保存度スコアはマルチプルアライメントされたアミノ酸配列群に対して real-valued ET 法[12]のスコアを用いて下記のように求める。

$$c_i = 1 - \frac{\int_{-\infty}^{s_i} \exp(-(t-m)^2/2\sigma^2)\mathrm{d}t}{\int_{-\infty}^{+\infty} \exp(-(t-m)^2/2\sigma^2)\mathrm{d}t} = 1 - \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{s_i} \exp(-(t-m)^2/2\sigma^2)\mathrm{d}t \quad \ldots (5)$$

ここで s_i は i 番目のアミノ酸残基の ET 法によるスコアであり、m および σ はそれぞれスコアの平均および標準偏差である。よって保存度のスコア c_i は最も変異の起こりやすいアミノ酸残基ほど 0.0 に近く、保存度の高いアミノ酸残基ほど 1.0 に近い値をとる。

ET 法に使用するアミノ酸配列のマルチプルアライメントは Swiss-Prot[14]や nr といった配列データベースに対する BLAST[15]検索により集めた配列を clustalW[16]でアライメ

ントすることにより得た。

本研究の応用例においては基本的に Swiss-Prot データベースを用いたが、Swiss-Prot にほとんど相同な配列がみつからないときは nr データベースを用いた。各適用例についてのデータベース選択については 2-4 節で述べる。

第二項はタンパク質 A、B の原子間の衝突を避けるために導入した。基本的に van der Waals エネルギーと同じ形の関数で符号を逆にしたものである。

$$F_2 = \sum_{atom\ pairs} \max\left[4.0\left(\left(\frac{\sigma}{R_{AB}}\right)^6 - \left(\frac{\sigma}{R_{AB}}\right)^{12}\right), -10.0\right] \quad \dots (6)$$

ここで R_{AB}（図 2-1 中の R）はタンパク質 A および B の原子間の距離であり、

$$\sigma = \frac{\rho_A + \rho_B}{2^{\frac{1}{6}}} \quad \dots (7)$$

とする(ρは原子の van der Waals 半径)。F_2 には原子ペアに対する下限（-10.0）を定義してある。これは最適化の際に大きな負の値になるのを避けるために導入した。和はタンパク質 A、B 間のすべての原子ペアについてとる。これについても適宜カットオフを適用して、計算を高速化する。

2-3-3　最適化

目的関数の最適化による構造の探索は次の 4 段階の過程で行う（図 2-2)。

Step　1:　はじめにタンパク質 A および B の分子表面をそれぞれ 3Åの 3 次元のグリッドで分割する。そして、グリッドで分割された各部分表面において頂点の平均の法線ベクトルと平均の位置を計算する。ここで、頂点の数が 5 に満たない部分表面については計算の効率化のために以後の過程で扱わないことにする。

Step　2:　一方の部分表面を表面 A から、もう一方の部分表面を表面 B からランダムに選ぶことにより部分表面のペアを作成する。200 個の部分表面のペアを作成し初期候補群とする。次に、以下のようにして部分表面の各ペアを相補的になるよう配置することで、分子 A、分子 B の相対位置が決まり、複合体の初期構造を発生させることができる。

（1）部分表面 A、B の重心が原点に一致するように平行移動する。

（2）次に、部分表面 A、B の平均の法線ベクトルが x 軸上で、反対を向くように配置する。

（3）x 軸周りの回転の自由度が残るため、部分表面の第一主軸を y 軸方向に向くようにする。

Step　3:　それぞれの部分表面の組み合わせから作成される複合体の初期構造に対して、Step 2 で用意した座標系を用いて、モンテカルロ法により式（1）で示した目的関数 F を

最適化する。

表面 A は固定し、表面 B にランダムな微小変位（回転および並進移動）を与える。ランダムな変位の大きさは以下の範囲とする。回転移動については、x 軸まわりの回転を-0.125 から 0.125 ラジアン、y 軸および z 軸まわりの回転を-0.05 から 0.05 ラジアンとする。並進移動については x 軸方向の移動を-0.1 から 0.1Å、y 軸および z 軸方向の移動を-0.25 から 0.25Åとする。ここで、最初に x 軸と部分表面の法線ベクトルの向きを平行にそろえたため x 軸は、y 軸および z 軸と異なった取り扱いをしている。y 軸および z 軸まわりの大きな回転、および x 軸方向の大きな並進移動は分子表面が衝突する結果になり易いためである。

微小変位により目的関数 F が増大した場合には、その変位を受け入れさらに次のランダムな微小変位を発生させる。F が減少した場合にはその微小変位を却下する。ランダムな変位と目的関数の評価は目的関数 F が 10 回続けて改善しなくなるまで繰り返し実行する。

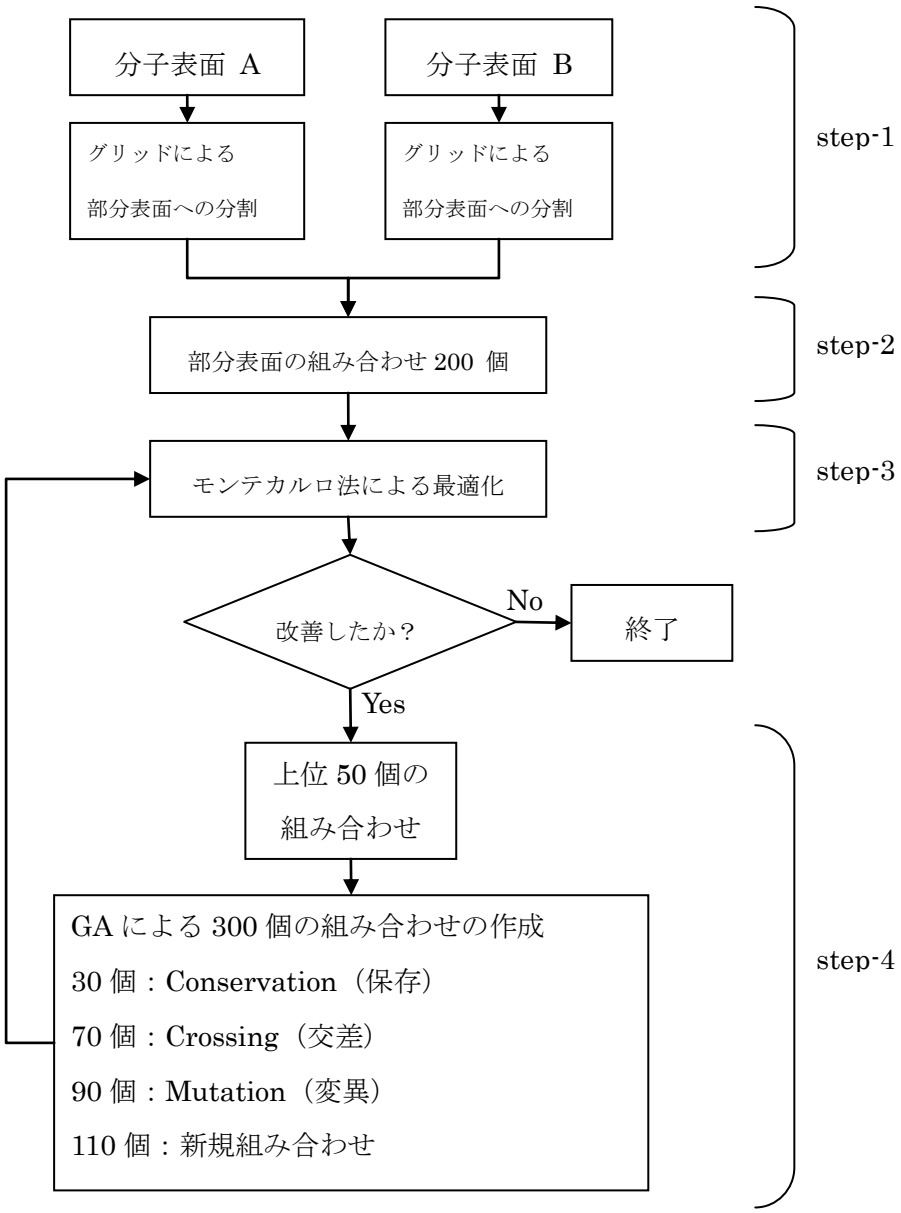

図 2-2

ドッキングにおける構造探索過程のフローチャート

Step　4:　前のステップで目的関数 F を最適化した部分表面の組み合わせを基にして、次の世代となる 300 個の新しい組み合わせを作成する。ここで部分表面の組み合わせを探索するために、次のように遺伝的アルゴリズムの方法を用いて新しい組み合わせを作成した。まず目的関数の上位 30 個は次の世代でもそのまま残す（Conservation：保存)。次に上位 50 個の組み合わせから、ランダムに二つの組み合わせを選びそのうちの一方の組み合わせの部分表面 A と他方の組み合わせの部分表面 B を組み合わせる(Crossing：交差)。これを繰り返して、70 個の新しい組み合わせを発生させる。次に、上位 50 個の組み合わせから一つのペアをランダムに選び、部分表面 A および B のどちらかを、3 次元のグリッドにおいて隣接するグリッドの部分表面と入れ替える(Mutation：突然変異)。これにより、50 個の新しい組み合わせを発生させる。同様にして、ランダムに選んだ組み合わせに対して、A、B のうちの一方の部分表面を、全表面からランダムに選択した部分表面と入れ替える。これにより、40 個の新しい組み合わせを発生させる。さらに、全く新しい候補を取り入れるため、全表面から完全にランダムに選択した部分表面 A、B の組み合わせを 110 個追加する。上記の過程で、同じ部分表面の組み合わせが現れたときは却下し新しい組み合わせが得られるまで乱数を発生させる。

結果として遺伝的アルゴリズムの過程で 300 個の新しい組み合わせが得られる。それらは再び Step 3 のモンテカルロ法による目的関数 F の最適化に供される。

CAPRI コンテストでは、最終的に得られた上位の候補を目視により検討し、さらに表面の相補性を評価した。相補性は静電ポテンシャル、疎水性、分子表面の点から相補性の度合いを統計的に評価した[17]。

計算時間は、タンパク質のサイズと予測された相互作用面の大きさに依存する。典型的な例で、保存度の高い部位として相互作用面を検出できる場合は予め可能性の高い部位の表面のみを使うことで 2-3 時間で計算できる。一方、結合サイトに関して全く情報が得られない場合にはタンパク質の表面全体を使用することになり 3、4 日の計算時間を要する。

同様に必要となる構造探索もそれぞれのタンパク質によって異なるが、一般的な例では、モンテカルロ法による探索は 200-500 回、遺伝的アルゴリズムによる候補の発生は 20-30 世代必要となる。

2-4　結果と考察

2-4-1　既知の複合体構造を学習データセットに用いたパラメータの決定

調節可能なパラメータを決定するため、複合体構造が既知で PDB に登録されている以下

の 7 個のタンパク質複合体構造を使用した。それぞれの PDB code と相互作用する分子のチェーン ID を示す。Rap1A/CRaf1(PDB code:1c1y、A-B)、Rac1/SptP(PDB code:1g4u、R-S)、RhoGAP/RhoA(PDB code:1tx4、A-B)、Rac/RhoGDI(PDB code: 1ds6:A-B)、RalGDS/Ras(PDB code:1lfd、A-B)、Ras/Ras-GAP(PDB code:1wq1、R-G)、Rab3A/rabphilin-3A (PDB code:1zbd、A-B)。

これらの構造は、small G タンパク質の関与するタンパク質複合体[18]のうち、Swiss-Prot または nr データベースに 100 以上の相同なアミノ酸配列が存在するものである。

これらのタンパク質について、構造探索によって得られた目的関数の最大値を持つ構造と X 線結晶構造との RMSD が 1.0Å程度になるように、式（1）の目的関数の変数 a と b を調整した。その結果、後述の CAPRI ターゲット 27 以降のタンパク質を除き、その他のタンパク質へ適用する際は a=2.0、b=3.0 を用いた。ターゲット 27 以降については、分子表面の作成プログラムとして、MSMS ではなく msroll というプログラムを使用した。これらは共に Connolly の方法に基づいたプログラムであるが、プログラムで使用されているパラメータが若干異なり、msroll で作成した表面の方がポリゴンを表現する頂点の密度が高いため、msroll で作成した分子表面を使用する際は a=1.3、b=3.0 を使用した。

2-4-2 Evolutionary Trace 法による予測部位とタンパク質間相互作用部位の相関

本研究のタンパク質ータンパク質ドッキング手法は、式（1）の目的関数 F における分子表面の頂点の重みとして、アミノ酸配列の保存度だけを用いることを目的としたものでは無く、後ほど CAPRI ターゲット 21 への適用結果として述べるように、重みとして例えばアミノ酸変異実験の結果などあらゆるものを取り入れることが可能である。

しかしながら、アミノ酸配列の保存度の利用は計算機による方法としては最も使用し易くかつ強力である[12]。さらに、今後も配列データベースは急速な拡大を続け、機能に関するアノテーションも進むと考えられるため、将来的にもさらに役に立つ方法になると期待される。

よってここで、どのような種類のタンパク質複合体が、アミノ酸配列の保存度を利用した相互作用部位の推定に適しているのかを検討する。

この目的のため、209 個の立体構造既知のタンパク質複合体を使用して 265 個のタンパク質—タンパク質相互作用面についてアミノ酸配列の保存度の解析を行った。これらは Mintseris と Weng らの研究[18]で用いられたタンパク質複合体のセットである。

アミノ酸配列の保存度と、タンパク質—タンパク質相互作用面に含まれることの相関を調べるために、265 個の相互作用面をそのタンパク質の機能に従って次のように大まかに分

類した。シグナル伝達(signal transduction、72 個)、酵素(enzyme、43 個)、インヒビター(inhibitor、19 個)、抗体(antibody、36 個)、抗原(antigen、31 個)、その他(64 個)である。

各分類のタンパク質複合体の PDB code を表 2-1 に示す。

表 2-1　ET 法の評価に用いたタンパク質複合体の分類

各分類の複合体を(PDB code):(解析対象のチェーン ID):(相互作用の相手のチェーン ID)の形式で記述した。

signal transduction
1A2K:D:A 1AGR:A:E 1AGR:E:A 1AVZ:B:C 1AVZ:C:B 1B6C:A:B 1B6C:B:A 1BDJ:A:B 1BLX:A:B 1BLX:B:A 1BUH:A:B 1C1Y:A:B 1C4Z:A:D 1C4Z:D:A 1CDM:A:B 1CMX:B:A 1DS6:A:B 1DU3:D:A 1DU3:F:A 1EAY:A:C 1ES7:A:D 1ES7:C:D 1EV2:A:E 1EV2:E:A 1F51:E:A 1F5Q:A:B 1F5Q:B:A 1FLT:V:Y 1FOE:A:B 1FOE:B:A 1FQ1:B:A 1FQK:A:B 1FQK:B:A 1G3N:A:C 1G3N:C:A 1G4U:R:S 1G4Y:R:B 1GCQ:B:C 1GOT:A:B 1GOT:B:A 1HCF:A:Y 1HCF:B:Y 1HE1:C:A 1HE8:A:B 1HE8:B:A 1I2M:A:B 1I2M:B:A 1I4D:D:A 1IB1:A:E 1IBR:A:B 1JDP:A:H 1JMA:B:A 1K9O:D:A 1LFD:B:A 1LFD:B:C 1QAV:A:B 1QAV:B:A 1QBK:C:B 1QMZ:A:B 1QMZ:B:A 1RRP:A:B 1TNR:A:R 1TNR:R:A 1TX4:A:B 1TX4:B:A 1VRK:A:B 1WQ1:G:R 1WQ1:R:G 1WWW:V:Y 1WWW:W:Y 1ZBD:A:B 1ZBD:B:A
enzymes
1A4Y:B:A 1ACB:E:I 1AVA:A:C 1AVW:A:B 1AZZ:A:C 1AZZ:A:D 1BQQ:M:T 1BVN:P:T 1CDK:A:I 1CHO:E:I 1CLV:A:I 1CSE:E:I 1D6R:A:I 1DFJ:E:I 1DHK:A:B 1DPJ:A:B 1DTD:A:B 1EAI:A:C 1EJA:A:B 1F34:A:B 1F7Z:A:I 1FAK:H:I 1FLE:E:I 1GLO:E:I 1HIA:B:I 1HX1:A:B 1I4O:A:C 1JIW:P:I 1JLT:B:A 1JTD:A:B 1JTG:A:B 1KIG:H:I 1PPF:E:I 1SLU:B:A 1STF:E:I 1TBR:H:R 1TMQ:A:B 1TOC:B:R 1UGH:E:I 2BTC:E:I 2SIC:E:I 3BTH:E:I 4HTC:H:I
inhibitors
1A4Y:A:B 1AVA:C:A 1AVW:B:A 1BQQ:T:M 1CHO:I:E 1D6R:I:A 1DF9:C:B 1DFJ:I:E 1DHK:B:A 1F7Z:I:A 1FAK:I:H 1G73:D:A 1JLT:A:B 1K9O:I:E 1PPF:I:E 1SGP:I:E 1TBR:R:H 2SIC:I:E 3BTH:I:E
antibody
1A2Y:A:C 1A2Y:B:C 1BGX:H:T 1BGX:L:T 1BJ1:H:W 1BZQ:L:A 1DEE:D:G 1DQJ:A:C 1DQJ:B:C 1E6J:H:P 1EGJ:H:A 1EO8:H:A 1FBI:H:X 1FC2:D:C 1FE8:H:A 1FE8:L:A 1FJ1:A:F 1FJ1:B:F 1FNS:H:A 1FSK:C:A 1G9M:H:G 1IIS:C:B 1JHL:H:A 1JPS:H:T 1JTP:A:L 1MLC:B:E 1NCC:H:N 1NCC:L:N 1OSP:H:O 1OSP:L:O 1QKZ:H:A 1WEJ:H:F 1WEJ:L:F 2HMI:D:B 2JEL:H:P 2VIR:B:C
antigen

1ADQ:A:H 1BGX:T:H 1BGX:T:L 1BJ1:W:H 1BZQ:A:L 1DQJ:C:A 1DQJ:C:B 1DZB:X:A 1EOO:A:B 1EOO:B:A 1E6J:P:H 1E96:A:B 1EO8:A:H 1FBI:X:H 1FE8:A:L 1FNS:A:H 1FSK:A:C 1G9M:G:C 1G9M:G:H 1HEZ:A:E 1IIS:A:C 1IIS:B:C 1JHL:A:H 1JTP:L:A 1MLC:E:A 1MLC:E:B 1NCC:N:H 1NCC:N:L 1WEJ:F:H 2JEL:P:H 2VIR:C:B
other
1AIP:A:C 1AIP:A:D 1AIP:C:A 1AIP:D:A 1AK4:A:D 1AK4:D:A 1ARO:P:L 1ATN:A:D 1AXI:A:B 1AXI:B:A 1B41:A:B 1BML:A:C 1BP3:A:B 1BP3:B:A 1CD9:B:A 1CIC:B:C 1CIC:C:B 1D5M:A:C 1DKG:A:D 1DKG:D:A 1DN1:A:B 1DN1:B:A 1DX5:I:M 1DX5:M:I 1EBD:B:C 1EFU:A:B 1EFU:B:A 1F6O:A:B 1F6O:B:A 1FAK:H:T 1FAK:L:T 1GH6:A:B 1HYR:A:C 1HYR:B:C 1HYR:C:A 1HYR:C:B 1I1R:A:B 1I5K:A:C 1I7W:B:A 1ICF:A:I 1IM3:A:D 1IM9:A:D 1IM9:D:A 1IRA:X:Y 1IRA:Y:X 1ITB:A:B 1ITB:B:A 1J7V:L:R 1KAC:B:A 1KCG:A:C 1KCG:B:C 1KCG:C:A 1KCG:C:B 1LPB:B:A 1NRN:H:R 1QO3:A:C 1QO3:C:A 1SBB:A:B 1T7P:A:B 1T7P:B:A 1YCS:A:B 1YCS:B:A 2PCC:A:B 2PCC:B:A

各複合体について相互作用部位を予測する側のタンパク質のアミノ酸配列を問い合わせ配列とし、Swiss-Prot データベースに対し BLAST 検索で相同な配列を検索した。相同性が低いと立体構造の違いが大きいため、見つかった配列のうち相同性が 25%以上のものを用いた。また、問い合わせ配列全体の長さのうち一部分とのみ相同性を有する配列を除くため、問い合わせ配列の 70%以上と相同性を有する配列のみを用いた（これをカバー率と定義する）。

タンパク質 A および B からなるタンパク質複合体がある場合、A における B の相互作用部位と B における A の相互作用部位の解析ができる。また、2 個のサブユニットからなるものだけでなく 3 個以上のサブユニットを有するものもあるため 209 個のタンパク質複合体から 455 個の解析対象を見出すことができた。これらのうち、20 個以上の相同配列を利用でき、相互作用部位に 5 個以上のアミノ酸残基を有するものだけを調べた結果 265 個となった。

次に clustalW を用いてマルチプルアライメント（多重配列）を作成した。次に 2-3-2 節の式(5)によりアミノ酸配列の各サイトの配列保存度 c_i を計算した。この c_i の値を用いて、配列保存度の高い上位 8%のサイトでかつ溶媒に露出しているアミノ酸残基を予測された相互作用部位のアミノ酸残基として取り出す。一方、複合体構造を用いることで、実際に相互作用部位に位置するアミノ酸残基を定義できる。本研究では、アミノ酸残基中の一つ以上の原子（水素を除く）がもう一方のタンパク質から 4.0Å以内にある場合に相互作用部位のアミノ酸残基として定義する。こうして相互作用部位にあると予想されたアミノ酸残

基が実際に相互作用部位に存在するかを調べることができる。

相互作用部位にあると予測したアミノ酸が実際に相互作用部位に存在するかの相関を次式の Matthews の相関係数(MCC：Matthews Correlation Coefficient)[19]を用いて評価した。

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP}) \times (\mathrm{TN} + \mathrm{FN}) \times (\mathrm{TP} + \mathrm{FN}) \times (\mathrm{TN} + \mathrm{FP})}}$$

ここで TP、TN、FP、FN はそれぞれ真陽性（相互作用部位にあると予測したアミノ酸が実際に相互作用部位にある）、真陰性（相互作用部位にあると予測していないアミノ酸が実際に相互作用部位にない）、偽陽性（相互作用部位にあると予測したアミノ酸が実際は相互作用部位にない）、偽陰性（相互作用部位にあると予測していないアミノ酸が実際は相互作用部位にある）となったアミノ酸残基の数である。

各機能分類について、平均の MCC は次のようであった。

シグナル伝達 0.14(0.22)、酵素 0.24(0.14)、インヒビター-0.01(0.14)、抗体-0.05(0.09)、抗原 0.02(0.13)、その他 0.02(0.19)であった。括弧内の値は MCC の値の標準偏差を示す。シグナル伝達の機能をもつタンパク質と酵素タンパク質において相互作用面のアミノ酸配列の保存度が他に比べて比較的高いことが分かる。注目すべき点として、シグナル伝達関連タンパク質の高い MCC 値は、21 個の small G タンパク質に由来していることが分かった。これら small G タンパク質に対する平均の MCC 値は 0.30 であった。その他の興味深い点は、酵素とインヒビターの組み合わせで複合体が構成されるが、酵素に対する MCC 値が 0.24 であるのに対して、インヒビターに対する MCC 値は-0.01 となり、保存度が逆の傾向を示している点である。これは Bradford と Westhead らによって別の方法によっても指摘されている[20]。

この解析の結果、対象のタンパク質がシグナル伝達関連か酵素の場合には ET 法を使用して潜在的な相互作用部位を推定するのがよいと考えられる。

しかしながら、各分類において、標準偏差が示すように MCC の値のばらつきは小さくないため、ET 法の有効性については、対象のタンパク質ごとに注意深く検討する必要がある。

2-4-3　CAPRI ターゲットを用いたテスト

本研究で考案したタンパク質―タンパク質ドッキング手法の性能を評価するため、本手法を CAPRI のターゲット（ブラインドテストの課題となる複合体構造をターゲットと呼ぶ）に適用した。

まず、すでにブラインドテストが終了し結果が公表されている過去のターゲットについ

ての追試を行った。過去のターゲットのうち抗原—抗体複合体（ターゲット 2-7、13 および 19)、ホモオリゴマー（ターゲット 1、9 および 10)、タンパク質単体の構造が未知であるためホモロジーモデリングによる単体構造の作成を必要とするターゲット 11 と 14。対象のタンパク質の相同なアミノ酸配列が十分得られないターゲット 8 などを除いた結果、本研究の手法が適用できるターゲット 12、18、21 に対して評価した。

ターゲット 24 以降についてはブラインドテストに参加して評価を行った。

適用においては、先に述べた small G タンパク質を学習セットに用いて決定したパラメータを用いて同じ条件で行い、目的関数 F の高い複数の構造を候補として選択し評価を行った。

CAPRI で用いられている呼び方に従って便宜的に、複合体のうち分子量の大きい側のタンパク質を”receptor”、小さい側のタンパク質を”ligand”と呼ぶこととする。

2-4-4　ターゲット 12: cohesin と dockerin のドッキング

ターゲット 12 は cohesin と dockerin のドッキングである。cohesin はセルロソーム（cellulosome）の分子骨格（scaffold）タンパク質のドメインである。dockerin は酵素活性を持つサブユニットのドメインであり、cohesin—dockerin の相互作用により、分子骨格に触媒を行うサブユニットが付け加えられる[21]。 cohesin と dockerin の複合体生成は cellulosome 構築の基礎的な過程ということになる。よって、これらのタンパク質を先述の機能によって分類した場合、その他に分類される。CAPRI では unbound 構造の cohesin（PDB code: 1anu）と bound 構造の dockerin を用いて複合体を予測することが求められた。

dockerin に対しては、BLAST 検索により、Swiss-Prot データベースから 14 個の相同性の高いアミノ酸配列が見つかった。各アミノ酸残基に対して保存度のスコアを計算し、それを重みとして dockerin の全表面をドッキング計算に用いた。一方、cohesin に対しては、BLAST 検索によって、相同性のある配列を見つけることができなかったため、全表面に対して等しく重みを 0.9 として計算を行った。このケースにおいては、目的関数 F が最大となる候補として、X 線結晶構造解析で得られた構造（PDB code: 1ohz）[22]に類似の構造を得ることができた。X 線結晶構造と比較すると、receptor 側である cohesin のタンパク質で重ね合わせを行った時、ligand 側である dockerin の全 Cα原子の RMSD が 1.03Å であり、界面のアミノ酸の Cα原子の RMSD は 0.66Å であった（図 2-3)。

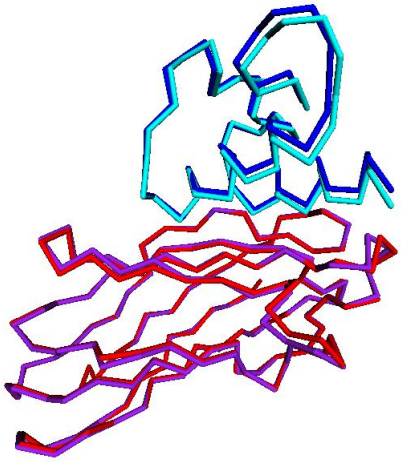

図 2-3　ターゲット 12 についての最も精度の高い予測構造と結晶構造の比較

cohesin(receptor)及び dockerin(ligand)の複合体の結晶構造(PDB code:1ohz)の Cα原子のトレースをそれぞれ赤および青で示した。それらに対応する予測構造を紫および薄青で示した。構造は receptor 側分子である cohesin で構造重ね合わせがされている。

2-4-5　ターゲット 18: xylanase と TAXI xylanase inhibitor のドッキング

ターゲット 18 は xylanase と TAXI xylanase inhibitor の複合体の予測であった。これは、酵素とそのインヒビターがつくる複合体構造の予測である。xyalnase については unbound 構造（PDB code:1ukr）、インヒビターについては単体の構造が知られていないため bound 構造を用いたドッキングを行う。xylanase に対しては、BLAST 検索により 27 個の相同タンパク質のアミノ酸配列を見つけることができた。xylanase の分子表面全体のうち相同性の高い 25%のアミノ酸残基の作る表面をドッキングに使った。一方、TAXI inhibitor については、分子表面全体を等しい重み(=0.9)としてドッキングに使用した。これは 2-4-2 節で述べたようにインヒビターにおいては、アミノ酸の保存度は相互作用部位を推定するのによい指標にならないためである。

ドッキングの結果、実験構造（PDB code:1t6g）[23]と最も近い複合体モデルの xylanase（ligand 側）の全 Cα原子の RMSD は 5.8Å であった。相互作用面の Cα原子の RMSD は 3.5Å であった。このケースでは、もっとも実験構造に近い構造は、目的関数 F が 5 番目に高い構造であった（図 2-4）。

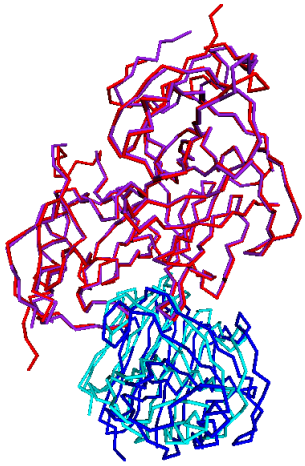

図 2-4　ターゲット 18 についての最も精度の高い予測構造と結晶構造の比較

xylanase inhibitor(receptor)及び xylanase(ligand)の複合体の結晶構造(PDB code:1t6g)の Cα原子のトレースをそれぞれ赤および青で示した。それらに対応する予測構造を紫および薄青で示した。構造は receptor 側分子である xylanase inhibitor で構造重ね合わせがされている。

2-4-6　ターゲット 21: Orc1p と Sir1p のドッキング

ターゲット 21 は Orc1p(origin recognition complex) と Sir1p (silent information regulator protein)の作る複合体構造の予測であった。

Orc1p として PDB code:1m4z、Sir1p として PDB code:1z1a を使う unbound-unbound ドッキングである。

相同性検索により、Orc1p に対してはわずか 4 つ、Sir1p に対しては全く相同なアミノ酸は配列が見つからなかった。こういった場合、前述の他のターゲットに対してはタンパク質の全表面を等しい重みで用いたが、今回は、変異実験により推定される相互作用部位の情報があったため[24-26]、これらの文献情報により推定される部位の分子表面を用いた。

Sir1p については、Y483、V484、S485、R487、A505 のアミノ酸残基の周りが相互作用部位と推定された[26]。Orc1p については H ドメインが Sir1p との相互作用部位であると報告されている[25]。

これらの情報を用いることで正しい予測モデルを得ることができた。実験で得られた複合体構造 (PDB code:1zhi) [27]に最も近いモデルは、Sir1 (ligand 側) の全 Cα原子の RMSD は 7.7Å、相互作用面の Cα原子の RMSD は 3.4Åであった。これらの値は、ターゲット 12 や 18 に対するものよりも大きいが、CAPRI の基準によると”acceptable”カテゴリーの予測精度には相当する。このケースでは、目的関数 F が最大のモデルが、最良のモデルであっ

た（図 2-5）。

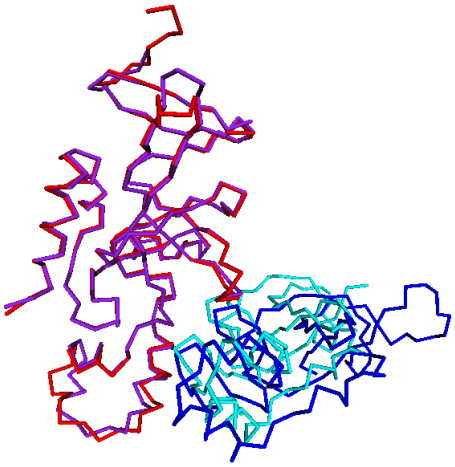

図 2-5　ターゲット 21 についての最も精度の高い予測構造と結晶構造の比較

Orc1(receptor)及び Sir1(ligand)の複合体の結晶構造(PDB code:1zhi)の Cα原子のトレースをそれぞれ赤および青で示した。それらに対応する予測構造を紫および薄青で示した。構造は receptor 側分子である Orc1 で構造重ね合わせがされている。

2-4-7　ターゲット 24 および 25: ARF1 と ArfBD のドッキング

ターゲット 24 と 25 は small GTPase である ARF1 と ARHGAP21 の Arf-binding domain(ArfBD)のつくる複合体構造の予測であった。ARHGAP21 は ARF1 に結合することでゴルジ体にリクルートされる。ArfBD の単体の立体構造が知られていなかったため、ターゲット 24 ではホモロジーモデリングにより PDB code:1btn を鋳型に ArfBD を構築して homology-unbound ドッキングを行うことが求められた。ターゲット 25 では ArfBD 構造として bound 構造が用いられた。ARF1 の構造は PDB code:1o3y を用いる。

前述のターゲットとは異なり、このターゲットでは複合体の予測にあたり、Ran と RanBD の複合体(PDB code:1k5d、チェーン A および B)についての文献情報[28]により影響を受け、かえって誤った構造を選択する結果となった。

ARF1 と Ran の配列相同性は 22%程度であるが、ARF1 と Ran を構造で比較すると、166 アミノ酸残基中 125 残基は立体構造上対応している。機能も関連しており、ともに Ras スーパーファミリーに属する small GTPase である。同様に ArfBD と RanBD の配列相同性は 9%しかないが、113 アミノ酸残基中 70 残基は立体構造上対応している。これらは共に、対応する GTPase の GTP 結合構造に特異的に結合する。このような類似性を考慮して、Ran—RanBD 複合体に類似した結合様式のモデルを ARF1—ArfBD 複合体の予測モデルとして

選択し提出した。

しかしながら、このような類似性にかかわらず、結果として ARF1-ArfBD と Ran-RanBD の結合モードは全く異なっていることが分かった。ARF1—ArfBD の複合体構造は PDB code:2j59 として公開された[29]。

ArfBD、RanBD とも対応する GTPase の”Switch I”領域と PH ドメインで相互作用していたが、ArfBD はさらに C 末端へリックスによって”Switch II”領域とも相互作用していた。

Ran-RanBD 複合体においては、”Switch II”領域はさらに別のタンパク質である ranGAP と結合する。よって、ARF1-ArfBD と Ran-RanBD の GTPase の構造を重ね合わせても、ArfBD と RanBD は PH ドメインの構造についても空間的な重なりが見られなかった。

そこで、Ran-RanBD 複合体に関する文献情報を利用しないで、ET 法による重みによって探索した構造アンサンブルから、正しい複合体に類似の構造を探したところ、ターゲット 25 については”medium”カテゴリーに相当する精度の構造をアンサンブルの中に見つけることが出来た。(図 2-6)

このモデルは、目的関数 F で見た場合に 44 番目に高い構造であった。全 Cα原子に対する RMSD=2.86Å、相互作用面の Cα原子に対する RMSD=2.59Å であった。

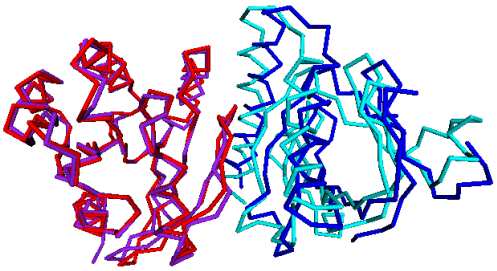

図 2-6　ターゲット 25 についての最も精度の高い予測構造と結晶構造の比較

ARF1(receptor)及び ArfBD(ligand)の複合体の結晶構造(PDB code:2j59)の Cα原子のトレースをそれぞれ赤および青で示した。それらに対応する予測構造を紫および薄青で示した。構造は receptor 側分子である ARF1 で構造重ね合わせがされている。

2-4-8　ターゲット 26：TolB と Pal のドッキング

ターゲット 26 は TolB （PDB code: 1c5k）と Pal periplasmic domain（PDB code: 1oap, 65–137）のつくる複合体構造（PDB code: 2hqs）[30]の予測である。ともに unbound 構造

を用いて行う予測である。

BLAST により、相同性が 25%以上で配列のカバー率が 70%以上の配列を検索すると、TolB に対して 46 個、Pal に対して 41 個の相同配列が見つかった。これらは殆ど同じ配列も含んでいるため、2 つの配列を選んだ時に全ての組み合わせで相同性が 95%以下となるようクラスタリングで代表配列を選ぶと、TolB に対して 28 個、Pal に対して 26 個の相同な配列が残された。これらの配列を用いて ET 法の解析を行ったあと、保存度の高い 35%のアミノ酸残基のつくる分子表面を、相互作用面の可能性のある部分として抜き出してドッキングに用いた。同じように保存度の高い 50%のアミノ酸残基の作る分子表面でもドッキングを行った。

さらに TolB 側についてはアミノ酸変異実験に関する文献情報[31]により残基番号 218、246、249、250、267、292、293、306、308、313、380 が相互作用部位になると示唆されていた。よってこれらのアミノ酸残基とその周りのアミノ酸残基 204、205、248、262、264、268、269、290、307、334、336、337 の分子表面を切り出し、目的関数の F_1 項における重みを 0.9 として用いたドッキングも行った。Pal 側は ET 法で特定した 50%の分子表面を用いた。

ドッキングの結果得られた目的関数が高い計 120 個の予測モデルからさらに目視での選別を行った。その結果、最終的な 10 個の構造に含めることはできなかったが、Scoring 部門で利用するための追加の 100 構造として、3 個の”medium”カテゴリーの精度の構造と 4 個の”acceptable”カテゴリーの精度の構造を登録することができた。3 個の”medium”カテゴリーの構造は 3 種類のドッキング条件から 1 つずつ得られた。

ここで最も精度の高かったモデルは、35%の分子表面を使った際に目的関数 F が 14 番目に高い構造であり、L_rms=4.281Å、I_rms=1.970Åであった（図 2-7）。

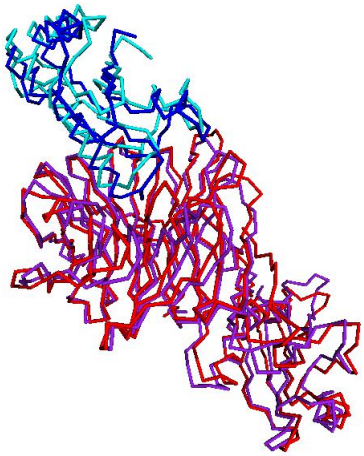

図 2-7　ターゲット 26 についての最も精度の高い予測構造と結晶構造の比較

TolB(receptor)及び Pal(ligand)の複合体の結晶構造(PDB code:2hqs)の Cα原子のトレースをそれぞれ赤および青で示した。それらに対応する予測構造を紫および薄青で示した。構造は receptor 側分子である TolB で構造重ね合わせがされている。

ET 法で予測した相互作用部位およびドッキングにより予測した複合体構造から得られる相互作用部位が、実験で得られた複合体の相互作用部位とどの程度一致しているかの評価を行った。予測構造および実験で得られた複合体については相互作用の相手側の分子から 4Å以内にあるアミノ酸残基を相互作用部位としてアミノ酸残基単位での評価を行った。

Pal を相互作用の相手側として TolB 側について解析すると図 2-8 のようになった。TolB 分子を空間充填モデルで示し、予測される相互作用部位（図 C については実験構造から得られた相互作用部位）を赤で示した。Pal 分子は青線で示した。図 A は ET 法で予測された相互作用部位を示しており Pal との現実の相互作用部位とある程度一致している。実験で得られた相互作用部位（図 C）を正解として感度=TP/(TP+FN)と定義すると 0.48 となり真の相互作用部位の半数近くを検出しているが、正しい部位以外の予測（偽陽性）も多く、特異性=TP/(TP+FP)と定義すると 0.13 であり、相関係数 MCC は 0.14 であった。図 B はドッキングによる予測構造から得られた相互作用部位であり、同じように評価すると、感度=0.55、特異性=0.71、MCC=0.60 となり、ET 法だけによる予測よりも精度が向上する。

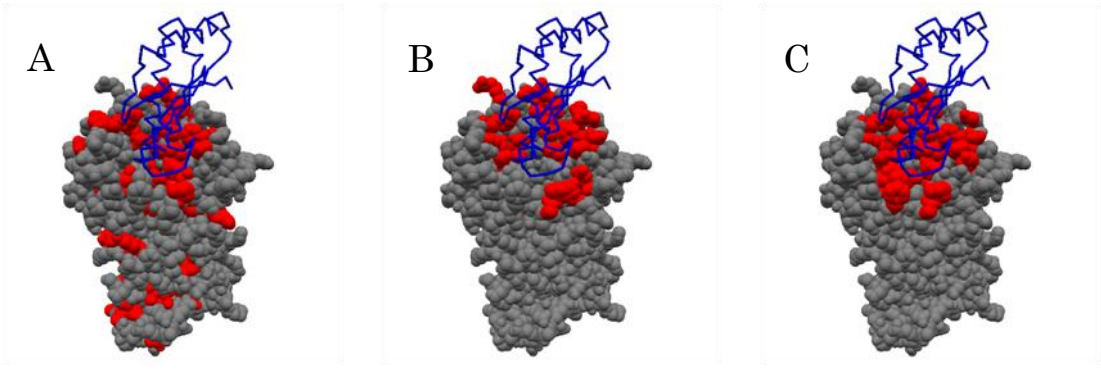

図 2-8　ターゲット 26 の TolB 分子について ET 法およびドッキング構造から予測される相互作用部位と実験で得られた構造の相互作用部位の比較。TolB 分子を空間充填モデルで示し、相互作用部位を赤で示した。Pal 分子を青で示している。比較のため実験で得られた構造を用いて表示している。A は ET 法で予測された相互作用部位（保存度の高い 35%部分）を示す。B はドッキングで得られた最も精度の高い構造から得られた相互作用部位、C は実験構造から得られた相互作用部位を比較のために示す。

同様にして TolB を相互作用の相手側として Pal 側について解析すると図 2-9 のようになった。Pal 分子を空間充填モデルで示し、予測される相互作用部位（図 C については実験構造から得られた相互作用部位）を赤で示した。TolB 分子は青線で示した。ET 法による予測結果（図 A）によると、TolB での予測と同じようにある程度真の相互作用部位を予測できているが、偽陽性が多く、感度=0.41、特異性=0.29、MCC=0.14 となった。ドッキングによる予測構造から得られた相互作用部位（図 B）は感度=0.73、特異性=1.00、MCC=0.82 となり、ET 法だけによる予測よりも精度が向上する。

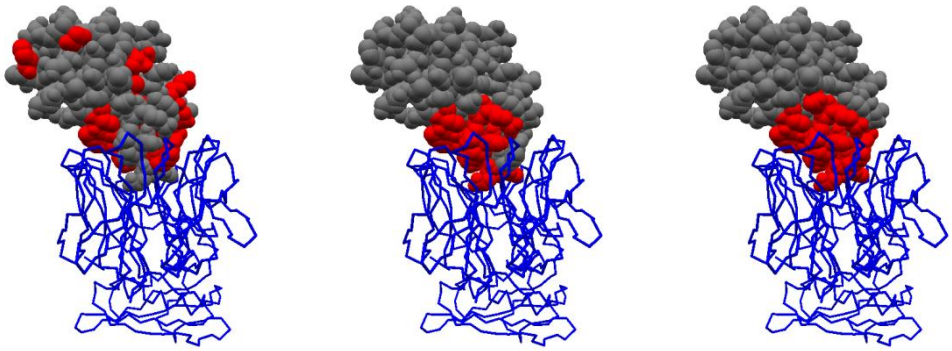

図 2-9　ターゲット 26 の Pal 分子について ET 法およびドッキング構造から予測される相互作用部位と実験で得られた構造の相互作用部位の比較。Pal 分子を空間充填モデルで示し、相互作用部位を赤で示した。TolB 分子を青で示している。比較のため実験で得られた構造を用いて表示している。A は ET 法で予測された相互作用部位（保存度の高い 35%部分）を示す。B はドッキングで得られた最も精度の高い構造から得られた相互作用部位、C は実験構造から得られた相互作用部位を比較のために示す。

このように、ET 法による予測の相関係数が 0.14 程度でも、本手法によるドッキングにより実験構造に近い構造を得ることが可能である。ドッキングの結果、ET 法で予測された相互作用部位のうち複合体構造を形成する際には妥当でないものが除かれることにより相互作用部位の予測精度を向上させることができた。

2-4-9　ターゲット 27：E2-25K と Ubc9 のドッキング

ターゲット 27 は Ubiquitin-conjugating enzyme E2-25 kDa (PDB code: 1yla)と SUMO-1-conjugating enzyme UBC9 (PDB code: 1a3s)がつくる複合体構造の予測である。

BLAST 検索から、E2-25 kDa と UBC9 に対しそれぞれ、93 個と 92 個の相同な配列が

見つかった。その後、クラスタリングによりそれぞれ 41 個と 69 個の配列に絞り込まれ、マルチプルアライメントのあと ET 法による解析を行った。

これらのタンパク質はネットワークのハブとして多くの異なったタンパク質と相互作用することが可能であるため、通常より広く、保存度の高い全体の 60%の表面を相互作用の可能性のある表面として使った。また、ET 法による解析を行わないで、目的関数 F_1 の重みを等しく 0.9 とした場合のドッキングも行った。

結果としてドッキングにより構築した 400 個の構造には、正解の結晶構造(PDB code: 2o25)に類似する構造を得ることはできなかった。

本ターゲットの結晶構造では、相互作用面の面積が 1095.2Å^2 であり、他の CAPRI ターゲットの半分程度と顕著に小さかった。本研究でほぼ正しい構造をサンプルすることができたターゲット 25 の相互作用面の面積は 1731.3Å^2、ターゲット 26 の相互作用面の面積は 2405.6Å^2 であり、この結果は本研究の手法が大きな相互作用面をもつ候補を優位に評価しやすい傾向をもつためであると考えられる。この傾向は次節のターゲット 28 で大きな相互作用面をもつホモダイマーの構造を比較的正しく予測することができたことでも確かめられた。

2-4-10　ターゲット 28：NEDD4L のホモ二量体構造の予測

ターゲット 28 は NEDD4L タンパク質の HECT ドメインが形成するホモ二量体の予測であった。単体の構造が知られておらず、約 50%のアミノ酸配列相同性を持つタンパク質の構造（PDB code:1zvd）をもとにホモロジーモデルを作成しドッキングを行った。ホモ二量体の予測の場合、通常二回対称性を有することが期待できる。よって発生させたモデルのうち、二回対称性を有するものだけを考慮した。ホモ多量体の相互作用面の予測にアミノ酸配列の保存度が有効であるか不明であったため、分子表面全体を一様な重みでドッキングに用いた。

実験で得られた構造（PDB code:2oni）を用いて予測モデルの評価が行われた結果、単体の構造において鋳型構造と実験で得られた構造にはドメイン間の相対位置に大きな違いが見られた。よって複合体構造にも大きな影響があり、参加者の全 357 モデル中に正しい構造は存在しなかった。しかしながら、本手法により提出したモデルの 1 つは、大きな構造変化のため L_rms は 23.9Å と大きいものの、相互作用部位の一致を評価する I_rms は 4.5 Å と全モデルの中で 2 番目に小さく acceptable カテゴリーに近い構造であった。このターゲットの相互作用面は 4160.3Å^2 であり、広い相互作用面を有するため本研究の手法がよく働いたと考えられる。

複合体形成時の構造変化はunbound構造を用いたドッキングにおいて主要な問題の一つである。一般的に、ターゲット24や28のように非常に大きな構造を伴う場合にはタンパク質を剛体とみなして行うドッキングで複合体構造を予測することは極めて困難である。

しかしながらターゲット 12、18、21、25、26 のように構造変化がそれほど大きくないケースでは前述のように本研究の方法で天然状態に近い構造を見つけることができている。これは恐らく、目的関数 F の F_2 項で vdW エネルギーに一定の制限をかけ、多少の原子衝突を許すよう関数のソフト化をしているためであると考えられる。

2-4-11　ターゲット 40: Trypsin と API-A のドッキング

本研究で定義した目的関数 F を用いて、自動予測のための WEB サーバを構築した（http://sysimm.ifrec.osaka-u.ac.jp/surFit/）。本サーバでは、目的関数 F における頂点の重みとして ET 法による結果のほかユーザが指定する重みを活用することができる。以降のターゲットでは本サーバにより、候補構造を生成したのちさらに検討を加えて構造候補を決定した。

ターゲット 40 は Trypsin とそのインヒビターAPI-A のドッキングである。Trypsin は unbound 構造（PDB code: 1bty）を用いる。API-A は単体の構造が未知のため bound 構造が用いられた。このインヒビターAPI-A は、Trypsin を阻害する部位が 2 つあるため、実験で得られた複合体構造（PDB code: 3e8l）[32]では、API-A（チェーン ID:C）と 2 つの Trypsin（チェーン ID:A および B）が 3 分子の複合体をつくる。よって、Trypsin と API-A の 2 種類の結合モード（チェーン A と C、チェーン B と C）の予測が求められた。

予測の結果、WEB サーバにより自動的に作成した 10 モデル中、スコアが最も高いモデルが結合モード B と C の予測において、L_rms=5.20Å、I_rms=1.29Å で medium カテゴリーに分類される良好な構造であった。またスコアが 5 番目の構造も L_rms=13.18Å、I_rms=3.39Å で acceptable カテゴリーに分類される構造であった。一方、結合モード A と C に対応する構造は含まれなかった。

次にサーバが作成した多くの構造に対して Tyrpsin の活性部位とインヒビターの阻害部位の情報を含めて検討した結果、スコア 1 位の構造と、スコア 41 位の構造がそれぞれ結合モード B と C、A と C に対応することが分かった。これらのモデルを基に、それぞれの結合モードに対して 5 つずつ最終構造を作成した。精度評価の結果、表 2-2 のように両方の結合モードに対して、high カテゴリーを含む構造の作成に成功した（図 2-10、図 2-11）。L_rms が 1.0Å以下となっており実験構造とほぼ一致するものといってよい。特に、モデル番号 3 は B-C の結合モードの予測において参加グループの全 368 構造の中で最も精度が高

い結果を示した（図 2-10）。

表 2-2　ターゲット 40 の予測構造の精度評価

	モデル番号	f_nat	L_rms（Å）	I_rms（Å）	カテゴリー
B-C の予測	1	0.95	0.84	0.44	high
	2	0.76	1.79	0.64	high
	3	0.88	0.67	0.47	high
	4	0.45	5.34	2.06	acceptable
	5	0.76	1.23	0.80	high
A-C の予測	6	0.63	5.95	1.31	medium
	7	0.88	2.82	1.26	medium
	8	0.83	4.55	1.00	medium
	9	0.89	1.00	0.45	high
	10	0.86	2.31	0.71	high

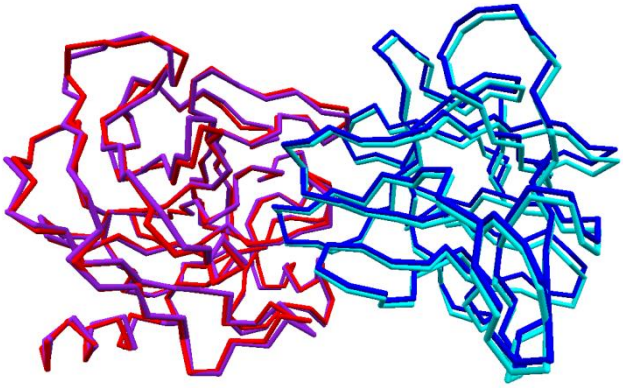

図 2-10　ターゲット 40（B-C 複合体）についての最も精度の高い予測構造（モデル番号 3）と結晶構造の比較

Trypsin と API-A 複合体の結晶構造(PDB code:3e8l2 B-C 複合体)の Cα原子のトレースをそれぞれ赤および青で示した。それらに対応する予測構造を紫および薄青で示した。構造は Trypsin で構造重ね合わせがされている。

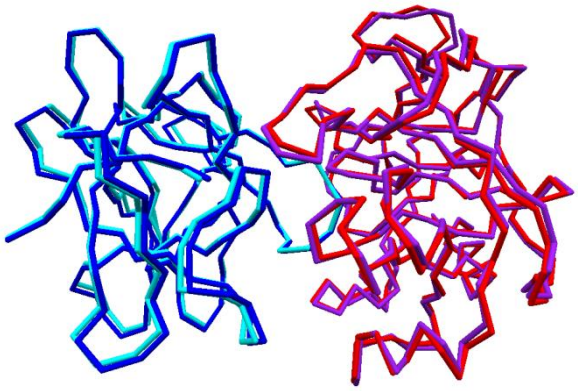

図 2-11　ターゲット 40（A-C 複合体）についての最も精度の高い予測構造（モデル番号 9）と結晶構造の比較

Trypsin と API-A 複合体の結晶構造(PDB code:3e8l2 A-C 複合体)の Cα原子のトレースをそれぞれ赤および青で示した。それらに対応する予測構造を紫および薄青で示した。構造は Trypsin で構造重ね合わせがされている。

2-4-12　ターゲット 42: TPR モチーフのつくるホモ二量体構造の予測

ターゲット 42 は TPR(tetratricopeptide repeat)モチーフがつくるホモ二量体構造の予測である。TPR はタンパク質間相互作用部位によくみられるモチーフとして知られる。本ターゲットは、TPR 同士が会合するように改変して得られた人工的なタンパク質複合体の予測である。実験で得られた構造（PDB code:2wqh）[33]には 2 種類の結合モードが存在したため、それぞれについて評価が行われた。

単体の構造は知られていないため、配列相同性の高い構造（PDB code:1na0）を基に三か所にアミノ酸変異を導入した。これはホモロジーモデリングサーバ Spanner（http://sysimm.ifrec.osaka-u.ac.jp/spanner/）により行った。

自動予測サーバにより複合体候補を作成し、2 回軸の存在するものだけを考慮した。スコア上位 10 モデル内には実験構造と一致する構造はみられなかったが、スコア上位の構造を文献 16 の方法により静電相互作用、疎水性、形状の相補性により再検討した。結果、最終候補構造の 1 つ（対称性を持つ構造でスコア 11 番目の構造）として実験構造の結合モードの 1 つを再現する構造を高精度に予測することができた。この構造は f_nat=0.658、L_rmsd=1.290Å、I_rmsd=0.732Å で high カテゴリーに分類されるものであり、参加者の全 278 構造のうちでもっとも精度の高い構造であった（図 2-12）。

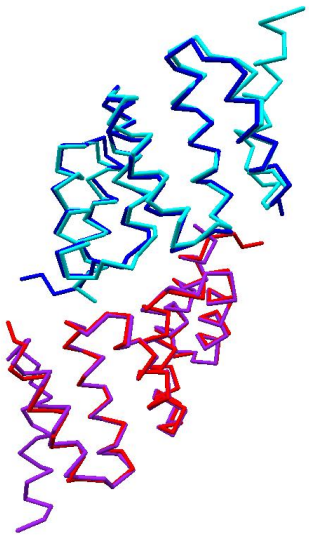

図 2-12　ターゲット 42 についての最も精度の高い予測構造と結晶構造の比較

TPR の複合体（ホモ二量体）の結晶構造(PDB code:2wqh)の Cα原子のトレースをそれぞれ赤および青で示した（結晶の対称性により作成）。それらに対応する予測構造を紫および薄青で示した。構造は赤、及び紫で示した分子で構造重ね合わせがされている。

2-5　結論

タンパク質の分子表面の相補性と ET 法によるアミノ酸配列の保存度を用いたタンパク質複合体構造の予測法（タンパク質—タンパク質ドッキング）を開発した。 265 個のタンパク質—タンパク質相互作用面を ET 法により網羅的に解析した結果、酵素タンパク質およびシグナル伝達に関係するタンパク質において、アミノ酸配列の保存度とタンパク質間の相互作用部位へのなり易さの間に相関がみられた。相関係数はそれぞれ 0.24 および 0.14 であった。

分子表面の相補性と配列保存度のスコアを用いて複合体構造を評価する目的関数を定義し、それに適した構造探索法を開発した。

開発した手法を CAPRI ターゲットに適用したところ、いくつかのケースにおいて実験で得られる複合体と類似の構造を得ることができた。

また、開発した関数を利用して複合体予測の自動化サーバが構築された。このサーバの出力する候補を用いていくつかの CAPRI ターゲットの予測を行ったところいくつかの例（ターゲット 40、42）で CAPRI コミュニティ全体において最も精度の高い複合体構造を予測することに成功した。

参考文献

1. Thornton JM, Todd AE, Milburn D, Borkakoti N, Orengo CA. From structure to function: approaches and limitations. Nat Struct Biol. 2000;7(Suppl):991–994.

2. Kinoshita K, Nakamura H. Protein informatics towards function identification. Curr Opin Struct Biol. 2003;13:396–400.

3. Hwang H, Vreven T, Janin J, Weng Z. Protein-protein docking benchmark version 4.0. Proteins. 2010;78:3111–3114.

4. Zhang C, Liu S, Zhou Y. Docking prediction using biological information. ZDOCK sampling technique, and clustering guided by the DFIRE statistical energy function Proteins. 2005;60:314–318.

5. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc. 2003;125:1731–1737.

6. Ben-Zeev E, Fux L, Amster-Choder O, Eisenstein M. Experimental and computational characterization of the dimerization of the PTSregulation domains of BglG from Escherichia coli. J Mol Biol. 2005; 347:693–706.

7. Schueler-Furman O, Wang C, Baker D. Progress in protein-protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility. Proteins. 2005;60:187–194.

8. Janin J. Assessing predictions of protein-protein interaction: the CAPRI experiment. Protein Sci. 2005;14:278–283.

9. Wodak SJ, Méndez R. Prediction of protein-protein interactions: the CAPRI experiment, its evaluation and implications. Curr Opin Struct Biol. 2004;14:242–249.

10. Kinoshita K, Furui J, Nakamura H. Identification of protein functions from a molecular surface database, eF-site. J Struct Funct Genomics. 2002;2:9–22.

11. Kinoshita K, Nakamura H. Identification of the ligand binding sites on the molecular surface of proteins. Protein Sci. 2005;14:711–718.

12. Mihalek I, Res I, Lichtarge O. A family of evolution-entropy hybrid methods for ranking protein residues by importance. J Mol Biol. 2004;336:1265–1282.

13. Connolly ML. Solvent-accessible surfaces of proteins and nucleic acids. Science. 1983;221:709–713.

14. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 2003;31:365–370.

15. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–3402.

16. Higgins DG, Thompson JD, Gibson TJ. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. 1996;266:383–402.

17. Tsuchiya Y, Kinoshita K, Nakamura H. Analyses of homo-oligomer interfaces of proteins from the complementarity of molecular surface, electrostatic potential and hydrophobicity. Protein Eng Des Sel. 2006;19:421–429.

18. Mintseris J, Weng Z. Atomic contact vectors in protein-protein recognition. Proteins. 2003;53:629–639.

19. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta. 1975;405:442–451.

20. Bradford JR, Westhead DR. Asymmetric mutation rates at enzymeinhibitor interfaces: implications for the protein-protein docking problem. Protein Sci. 2003;12:2099–2103.

21. Shimon LJ, Bayer EA, Morag E, Lamed R, Yaron S, Shoham Y, Frolow F. A cohesin domain from Clostridium thermocellum: the crystal structure provides new insights into cellulosome assembly. Structure. 1997;5:381–390.

22. Carvalho AL, Dias FM, Prates JA, Nagy T, Gilbert HJ, Davies GJ, Ferreira LM, Romão MJ, Fontes CM. Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. Proc Natl Acad Sci U S A. 2003;100:13809–13814.

23. Sansen S, De Ranter CJ, Gebruers K, Brijs K, Courtin CM, Delcour JA, Rabijns A. Structural basis for inhibition of Aspergillus niger xylanase by triticum aestivum xylanase inhibitor-I. J Biol Chem. 2004;279:36022–36028.

24. Bose ME, McConnell KH, Gardner-Aukema KA, Muller U, Weinreich M, Keck JL, Fox CA. The origin recognition complex and Sir4 protein recruit Sir1p to yeast silent chromatin through independent interactions requiring a common Sir1p domain. Mol Cell Biol. 2004;24:774–786.

25. Zhang Z, Hayashi MK, Merkel O, Stillman B, Xu RM. Structure and function of the BAH-containing domain of Orc1p in epigenetic silencing. EMBO J. 2002;21:4600–4611.

26. Gardner KA, Rine J, Fox CA. A region of the Sir1 protein dedicated to recognition of a silencer and required for interaction with the Orc1 protein in Saccharomyces cerevisiae. Genetics. 1999;151:31–44.

27. Hou Z, Bernstein DA, Fox CA, Keck JL. Structural basis of the Sir1-origin recognition complex interaction in transcriptional silencing. Proc Natl Acad Sci U S A. 2005;102:8489–8494.

28. Seewald MJ, Korner C, Wittinghofer A, Vetter IR. RanGAP mediates GTP hydrolysis without an arginine finger. Nature. 2002;415:662–666.

29. Ménétrey J, Perderiset M, Cicolari J, Dubois T, Elkhatib N, El Khadali F, Franco M, Chavrier P, Houdusse A. Structural basis for ARF1-mediated recruitment of ARHGAP21 to Golgi membranes. EMBO J. 2007;26:1953–1962.

30. Bonsor DA, Grishkovskaya I, Dodson EJ, Kleanthous C. Molecular mimicry enables competitive recruitment by a natively disordered protein. J Am Chem Soc. 2007;129:4800–4807.

31. Ray MC, Germon P, Vianney A, Portalier R, Lazzaroni JC. Identification by genetic suppression of Escherichia coli TolB residues important for TolB-Pal interaction. J Bacteriol. 2000;182:821–824.

32. Bao R, Zhou CZ, Jiang C, Lin SX, Chi CW, Chen Y. The ternary structure of the double-headed arrowhead protease inhibitor API-A complexed with two trypsins reveals a novel reactive site conformation. J Biol Chem. 2009;284:26676–26684.

33. Krachler AM, Sharma A, Kleanthous C. Self-association of TPR domains: Lessons learned from a designed, consensus-based TPR oligomer. Proteins. 2010;78:2131–2143.

Appendix

A　アミノ酸選択的交差飽和実験の方法と結果

A-1　サンプル調整

アミノ酸選択的 ^{1}H ラベリング（バックグランドは ^{2}H とする）は YUH1 に対して行われた。YUH1 タンパク質は E. coli C41 strain で発現させた。細胞は 1.0 g の塩化アンモニウム、2.0 g の D-glucose-d_6、0.5 g の ^{2}H-Celtone® Base powder を含む M9 minimal medium を用いて 1.0 L の $^{2}H_2O$ (ISOTEC)中で 37℃にて培養した。

600 nm での吸光度が 0.8 に達した時、最終濃度 1.0 mM となる Isopropyl-β-D(-)-thiogalactopyranoside (IPTG)を YUH1 の発現誘導のために加えた。

同時に、^{2}H-Celtone® Base Powder に含まれる ^{2}H アミノ酸の 5 倍の濃度でラベル対象の ^{1}H アミノ酸(Ala、Arg、Leu、Tyr、Pro)を加えた(表 A-1)。細胞は発現誘導の 6 時間後に収集した。 タンパク質は DEAE Sepharose Fast Flow (GE Healthcare) 、MonoQ 5/50 GL anion exchange columns)、HiLoad 26/60 Superdex 75 pg size-exclusion column (GE Healthcare)によるカラムクロマトグラフィにより精製した。1.0 L の培養から 5 mg の収量のタンパク質 が得られた。

^{2}H、^{15}N で均一にラベルした Ub タンパク質の発現は 1.0 g の ^{15}N ラベルした塩化アンモニウム、2.0 g の D-glucose-d_6、1.0 g の ^{2}H,^{15}N-Celtone® Base powder を含む M9 minimal medium を用いて 1.0 L $^{2}H_2O$ 中で IPTG により誘導した。タンパク質の精製は以前の文献に従って行われた[1]。

表 A-1

Celtone® Base Powder(Cambridge Isotope Laboratories, Inc.)のアミノ酸組成

Celtone® Base Powder におけるアミノ酸の割合はおよそ 70%である

Ala	7.6%	Arg	5.6%	Asp	9.6%
Cys	0.7%	Glu	10.2%	Gly	6.4%
His	2.3%	Ile	3.1%	Leu	8.4%
Lys	12.0%	Met	1.6%	Phe	8.3%
Pro	5.6%	Ser	4.4%	Thr	4.8%
Trp	0.2%	Tyr	3.8%	Val	4.5%

A-2　方法

交差飽和（Cross-Saturation、以下 CS とする）実験およびアミノ酸選択的交差飽和（Amino Acid Selective Cross-Saturation、以下 ASCS とする）実験のための試料は、CS ドナー分子として異なったラベリングを施した YUH1 タンパク質 1.2 mM と CS アクセプター分子として $^{2}H,^{15}N$ で一様にラベルした Ub タンパク質 1.0 mM を混合したものが用いられた。サンプルバッファーは 50 mM のリン酸ナトリウム(pH 6.5)と 50 mM の塩化ナトリウムを含む 20% H_2O/80% $^{2}H_2O$ 水溶液とした。CS 実験及び ASCS 実験は 30℃で実行された。炭化水素プロトンへのラジオ波照射は 0.83 ppm を中心とした WURST（Wideband, Uniform Rate, and Smooth Truncation）-2 デカップリング法を用い 1.5 秒間行われた。WURST-2 のバンド幅は 4.5 kHz に設定した。 よって WURST-2 による飽和は 600 MHz の NMR 装置において-3.0～4.6 ppm であった。 recycling delay は 3.5 秒に設定した。記録時間は 1 サンプル当たり、3-12 時間であった。

A-3　結果

ASCS では CS ドナータンパク質において CS 源となる水素原子を限定するために、アミノ酸選択的な ^{1}H ラベルを使用する。図 A-1(A)は、^{2}H バックグラウンドの中で Tyr を選択的に ^{1}H ラベルした YUH1 を CS ドナータンパク質とした際の ASCS 実験の結果を示す。

ラジオ波照射がある場合(+)とない場合(-)の 2 次元 HSQC スペクトルと 1 次元 HSQC スペクトルは、Ub(CS アクセプター)の Gly 75 のアミドプロトンのシグナル強度が飽和によって減少するのを示している。しかし、Ub の Thr 9 においては、シグナル強度の減少が見られない。これらのスペクトルは明らかにタンパク質複合体中において、YUH1 中の 1 つ以上の Tyr 残基が、Ub の Gly 75 の近傍に位置する一方、Thr 9 からは遠い位置に存在することを示す。

一方、図 A-1(B)は従来の CS 実験の結果である。CS ドナーである YUH1 は一様に ^{1}H ラベルされている。この場合は、Thr 9 と Gly 75 のピーク強度はともに大きく減少しており、Thr 9 と Gly 75 がともにタンパク質複合体において相互作用面に位置することを示す。

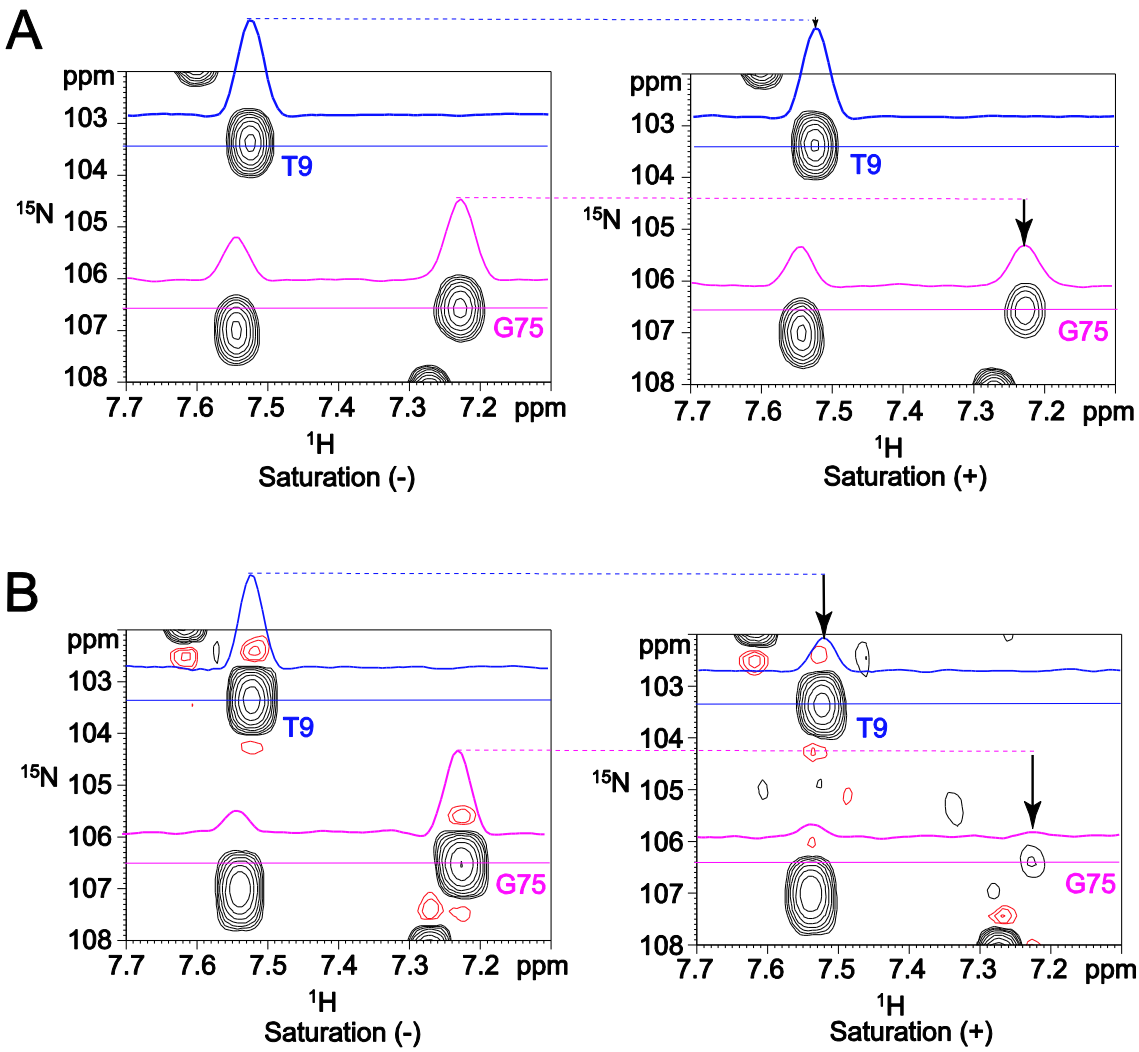

図 A-1

YUH1(C90S 変異体)からの交差飽和は Ub において HSQC シグナル強度の減衰として観測される。

YUH1 を CS ドナーとして、図 A は Tyr を選択的に ^{1}H ラベルした時、図 B は選択的ラベルをしていない（一様に ^{1}H ラベルした）時の 2 次元 HSQC スペクトルを示す。左側の図はラジオ波を照射していないとき Saturation(-)、右側の図はラジオ波を照射して交差飽和が起きているとき Saturation(+)のスペクトルを示す。

T9 と G75 のシグナルピークを通る直線上のシグナル強度（1 次元 HSQC）をそれぞれ青、マゼンタの線で示した。交差飽和によるシグナル強度の減衰を矢印で示した。

1. Sakamoto T, Tanaka T, Ito Y, Rajesh S, Iwamoto-Sugai M, Kodera Y, Tsuchida N, Shibata T, Kohno T. An NMR analysis of ubiquitin recognition by yeast ubiquitin

hydrolase: evidence for novel substrate recognition by a cysteine protease. Biochemistry. 1999;38:11634–11642.

B　CAPRI(Critical Assessment of PRedicted Interactions)

B-1　概要

CAPRI(Critical Assessment of PRedicted Interactions)はタンパク質複合体構造予測（タンパク質―タンパク質ドッキング）の性能をブラインドテストによって客観的に評価するための worldwide なコミュニティである。タンパク質立体構造の予測を目的とする CASP(Critical Assessment of methods for Structure Prediction)[1]を参考にして 2001 年に開始された。CAPRI の WEB ページは EBI（European Bioinformatics Institute）で管理されている。（http://www.ebi.ac.uk/msd-srv/capri/capri.html）。

X 線結晶構造解析などで実験的にタンパク質複合体の構造決定を行っている研究者から、未公表の複合体構造が提供されテストに用いられる（予測対象の複合体はターゲットと呼ばれる）。予測法の開発に取り組んでいる研究者は、ターゲットのタンパク質複合体を構成しているサブユニットが単体で構造決定されたときの構造(unbound 構造と呼ばれる)が与えられこれに基づいて複合体構造を予測する。

通常 2-3 週間程度の期間が与えられ、定められた期日までに各グループは予測した複合体モデルを 10 個まで事務局に登録する。そののち独立した評価チームにより各モデルの精度が評価される。

通常、基にするサブユニットの構造はともに unbound 構造であり、複合体形成時の構造の変化を考慮する必要がある(unbound-unbound ドッキングと呼ぶ)。構成要素のサブユニットの単体での構造が決定されていない場合は、便宜上、一方について複合体構成時の構造(bound 構造と呼ばれる)を使う場合もある(bound-unbound ドッキング)。また、そのものの単体構造は未知であるが、配列相同性のあるタンパク質の構造が既知の場合、それに基づいたホモロジーモデルを作成し、ドッキングを行うことが求められる。この場合は、通常の unbound-unbound ドッキングに比べ一般的に難易度が高くなる。

B-2　評価法

CAPRI での予測構造の評価は、提出された各構造について計算される 3 つの値（f_{nat}、L_{rms}、I_{rms}）によって行われ、これらにより精度の高いものから High、Medium、Acceptable、Incorrect の 4 つのカテゴリーに分類される[2]。

3 つの値は以下のように計算される。便宜的に複合体を作る二つのタンパク質のうち、分子量の大きい分子を Receptor、分子量の小さい分子を Ligand と呼ぶ。

（1）f_{nat} は分子間で接触している残基ペアをどれだけ再現できたか示す指標で、次のように定義される。Receptor と Ligand の間で、いずれかの原子の間の距離が 5Å以下であるアミノ酸残基のペアを接触しているとみなし、残基ペアと定義する。実験的に決定された天然の複合体構造と、予測された複合体構造に対して残基ペアを同定し、f_{nat} は次式により計算される。

$$f_{nat} = \frac{\text{天然構造の残基ペアのうち予測構造にもみられた残基ペアの数}}{\text{天然構造の残基ペアの数}}$$

（2）L_{rms} は天然構造と予測構造を Receptor 側の分子で重ね合わせを行った時の、Ligand 側分子の RMSD(Root mean square deviation)で定義される。重ね合わせと RMSD の計算は主鎖原子（N、Cα、C、O 原子）で行う。

RMSD は構造間の差を示すためによく用いられる指標で次式であらわされる。ここで N は原子数、$\mathbf{r}^1$、$\mathbf{r}^2$ は比較する構造 1、2 の原子の座標を示す。

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i}^{\mathrm{atoms}} (\mathbf{r}_i^1 - \mathbf{r}_i^2)^2}$$

（3）I_{rms} は天然構造と予測構造の相互作用面周辺での RMSD である。L_{rms} はモデル全体の一致度を示すためによい指標であるが分子のサイズに依存する値である。相互作用面付近で良く構造が一致していてもわずかに向きが異なれば大きな分子では相互作用面から遠い部分の寄与で値が大きくなってしまい評価に適当でない場合がある。そこで、天然の複合体構造において Receptor と Ligand の間で、いずれかの原子の間の距離が 10Å以下であるアミノ酸残基を相互作用面のアミノ酸残基と定義する。これらの主鎖原子（N、Cα、C、O 原子）をそれぞれ対応する予測構造の主鎖原子と重ね合わせを行い RMSD を計算する。これを I_{rms} と定義する。

以上のように評価は 3 種類の指標で行われる。これは複合体の予測構造を inhibitor の設計やアミノ酸変異実験の指針として用いるためには分子間の結合モードが重要になるため f_{nat} のような指標が重要である一方、全体の分子の配置が重要な場合には L_{rms} が良い指標になるといった目的の違いのためである。

上記（1）（2）（3）の値の評価の他に分子間で原子の衝突が一定数より多い複合体構造は却下される。これは、分子間で原子の衝突が多い構造は、モデルとしてふさわしくないのは明らかであるが残基ペアの数が多くなり f_{nat} の値が大きくなる場合があるためである。

f_{nat}、L_{rms}、I_{rms} が計算されると下記表 B-1 に基づいて、カテゴリーが判定される。

表 B-1　CAPRI における構造の評価基準

カテゴリー	評価基準
High	f_{nat}≧0.5 かつ（L_{rms}≦1.0 または I_{rms}≦1.0）
Medium	（f_{nat}≧0.3 かつ f_{nat}＜0.5）かつ（L_{rms}≦5.0 または I_{rms}≦2.0） または f_{nat}≧0.5 かつ L_{rms}＞1.0 かつ I_{rms}＞1.0
Acceptable	（f_{nat}≧0.1 かつ f_{nat}＜0.3）かつ（L_{rms}≦10.0 または I_{rms}≦4.0） または f_{nat}≧0.3 かつ L_{rms}＞5.0 かつ I_{rms}＞2.0
Incorrect	f_{nat}＜0.1 または（L_{rms}＞10.0 かつ I_{rms}＞4.0）

B-3　サーバ部門、Scoring Experiment

CAPRI には、通常の予測部門のほかサーバ部門がある。これは相互作用する 2 つの分子構造を入力すると複合体構造を自動的に予測して出力する機能を持つ WEB サーバの評価を目的とした部門である。サーバ部門の WEB サーバは一般に公開され利用できるものであることが求められる。また、予測開始から 24 時間以内に事務局に結果を送付する必要がある。ターゲット個別の事情や文献情報を考慮できないこと及び制限時間のため通常部門に比べて結果が劣る場合が多いが、容易なターゲットでは精度の高い構造の出力に成功するケースもみられる。特別な予備知識を必要とせず自動的に精度の高い複合体構造を出力させるサーバは利用者にとって有用であり今後のさらなる発展が望まれる。

さらに通常の予測のほかターゲット 22 以降 Scoring Experiment 部門が実施されている[3]。これは複合体構造候補を発生させること（構造サンプリング）とは独立して構造の評価法(Scoring)を発展させることを目的とした部門である。通常の予測部門の参加者は 10 個の予測構造の登録の他 100 個までの構造をこの部門のために提供することができる。各グループから提供された構造はまとめられ提供者を分からないようにして Scoring Experiment 部門の参加者に配布される。参加者はこの中から独自の評価法によって 10 個の構造を選び登録する。必要に応じて構造の最適化も許される。Scoring Experiment によって構造提供者が判別できなかった正解構造が発見されたり、提供された構造よりさらに構造が精密化されたりすることもある。

1. Moult J. Predicting protein three-dimensional structure. Curr Opin Biotechnol. 1999;10:583–588.

2. Méndez R, Leplae R, De Maria L, Wodak SJ. Assessment of blind predictions of protein–protein interactions: current status of docking methods. Proteins. 2003;52:51–67.

3. Lensink MF, Méndez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd Edition. Proteins. 2007; 69:704–718.

C　Evolutionary Trace 法

C-1　概要

Evoutionary Trace 法(進化追跡法、以下、ET 法とする)は 1986 年に Lichtarge らによって提案されたタンパク質の機能解析法である[1]。ET 法はアミノ酸配列からタンパク質間の相互作用部位や低分子の結合部位などの機能部位を推定するために用いられる。ET 法は、系統樹を用いる点とタンパク質立体構造を利用する点に特徴がある。

その方法の概要を図 C-1 に示す簡単な例を用いて説明する。まず解析対象のタンパク質のアミノ酸配列及び相同性のある配列を収集し、マルチプルアライメントを行い系統樹を作成する(図 C-1(A))。ここで解析の着眼点に基づいて配列のグループ分けをする。例えばある低分子の結合部位を調べるためには、各配列に関する知見から結合能を有するグループ、有しないグループに分ける。図では Group 1 と Group 2 の 2 つに分割した。次に各グループ内において保存しているアミノ酸残基を取り出す(図 C-1(B))。次に、各グループ内での保存残基をグループ間で比較する(図 C-1(C))。ここでグループ間でも保存されている残基は Conserved、グループ内のみで保存されている残基は Group-specific と分類される。これらは併せて TR(Trace Residue)と呼ばれる。Conserved とされた残基は Group 1、2 に共通の機能を担っており、Group-specific とされた残基は Group 間の機能の差異を担っていると解釈することができる。最後に、TR をタンパク質立体構造上にマッピングする(図 C-1(D))。タンパク質立体構造は相同配列の中のどれか 1 つについて既知であればそれを利用することができる。ここで空間的に孤立して存在する TR はノイズとして除去することができるため、タンパク質立体構造を利用して精度を向上させることができる。

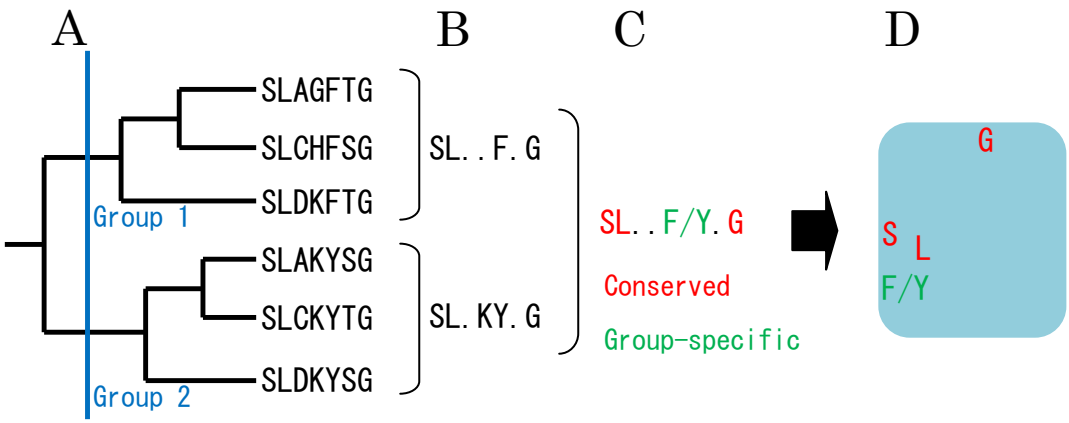

図 C-1　ET 法の概要

C-2　Real-valued ET 法

前節 C-1 で述べた ET 法では系統樹の作成後、各配列の機能に関する知見から配列をグループ分けする必要がある。本研究で用いたように多数のタンパク質について適用する場合にはグループ分けは困難であるし、得られた配列に対して十分な機能情報が得られない場合も多い。また、従来の ET 法ではグループ内で完全に保存されたアミノ酸残基しか扱わないため 1 つでも保存していない配列があると全く考慮されない点も問題であった。そこで本研究では 2004 年に Lichtarge らによって提案された real-valued ET 法[2]を用いた。real-valued ET 法の概要を下記に説明する。

系統樹の作成までの手順は従来の ET 法と同じであるが real-valued ET 法ではアミノ酸配列の各残基(残基番号 i)に対して次式により重要度の score(i)を計算する。

$$\mathrm{score(i)} = \sum_{\mathrm{node}} \frac{1}{\mathrm{num.\ of\ groups}} \sum_{\mathrm{group}} \sum_{\mathrm{residue\ type}} -\mathrm{P}\ln\mathrm{P}$$

上式においてグループの分割の方法は系統樹の node 数(=配列数-1)通り考えられるため(図 C-2 の点線の各位置)、全ての分割方法についての和とする。またアミノ酸残基の種類(20 種)について情報エントロピーの和をとる。P はグループ内でそのアミノ酸種が占めている割合を示す。これにより完全保存でない場合も考慮されることになる。図 C-2 の例の score(i)は図中に示したようになり保存度が高い位置ほどスコアは小さく完全保存の場合 0 となる。保存度の高い位置のアミノ酸を取り出し従来の ET 法と同じように立体構造上にマッピングすることができる。

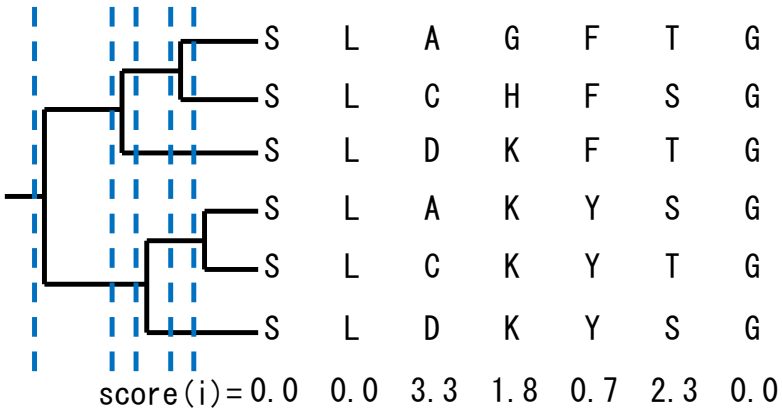

図 C-2　real valued ET 法の概要

1. Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol. 1996;257:342–358.

2. Mihalek I, Res I, Lichtarge O. A family of evolution-entropy hybrid methods for ranking protein residues by importance. J Mol Biol. 2004;336:1265–1282.

謝辞

本研究を進めるにあたりご指導頂き、種々の研究を行う機会を与えてくださった産業技術総合研究所福西快文主任研究員、大阪大学中村春木教授、名古屋大学在学中より長年にわたりご指導くださった神山勉教授に心より感謝申し上げます。

第 1 章に関する共同研究者である東京大学薬学研究科五十嵐俊介博士、大澤匡範博士、嶋田一夫教授に深く感謝いたします。

第 2 章に関してご指導くださった東北大学木下賢吾教授、CAPRI チームメンバーの村上洋一博士、土屋裕子博士、Daron M. Standley 博士に感謝申し上げます。

最後に、産業技術総合研究所バイオメディシナル情報研究センター（BIRC）、バイオ産業情報化コンソーシアム（JBIC）、日立ソリューションズ、名古屋大学神山研（J 研）の同僚、先輩、上司、そして両親にいつも温かい励ましとご助言を頂きました。謹んで御礼申し上げます。