

生体分子シミュレーションに用いる
新規力場パラメータの開発と
タンパク質立体構造予測への応用

2006 年

片桐 大輔

千葉大学大学院 医学薬学府 創薬生命科学専攻

ゲノム機能学講座 薬品物理化学研究室

[目次]

概要	1
序論	2
方法	6
(1) 既存力場パラメータ ff03 force field を用いた タンパク質立体構造予測	6
(2) 予測構造と Native 構造のポテンシャルエネルギー比較	8
(3) 既存力場パラメータ ff03 force field のアミノ酸 コンフォメーション安定性と量子化学的手法に よるアミノ酸コンフォメーション安定性の比較	9
(4) 高精度新規力場パラメータの開発	11
(5) 高精度新規力場パラメータを用いた タンパク質立体構造予測	12
(6) β -sheet 構造における新規力場パラメータの信頼性評価	12
結果	15
(1) 既存力場パラメータ ff03 force field を用いた タンパク質立体構造予測	15
(2) 予測構造と Native 構造のポテンシャルエネルギー比較	27
(3) 既存力場パラメータ ff03 force field のアミノ酸 コンフォメーション安定性と量子化学的手法に よるアミノ酸コンフォメーション安定性の比較	36
(4) 高精度新規力場パラメータの開発	52
(5) 高精度新規力場パラメータを用いた	56
(6) β -sheet 構造における新規力場パラメータの信頼性評価	73
考察	75

既存力場の helix 構造過剰評価と	
タンパク質立体構造予測の力場依存性	75
量子化学計算の妥当性と水の溶媒効果の重要性	78
既存力場を用いた MD シミュレーションによる	
コンフォメーション安定性と溶媒効果	79
既存力場パラメータによるタンパク質立体構造に対する、	
MMPBSA(GBSA)計算を使った計算条件検討	80
新規力場パラメータの妥当性	83
タンパク質立体構造予測と MD シミュレーション時間	87
結論	89
謝辞	90
参考文献	91
MD シミュレーション変更使用オプション説明一覧	97
MD シミュレーション変更使用オプション中の引用文献	103
主論文目録	105
学位論文審査委員	106
(付録) 研究業績発表会要旨	107

[図表目次]

Figure

1. Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=1, gbsa=0. Four TRP residues in 1LE3 and TYR and TRP residues in 1L2Y were represented as a *stick* expression. ... **16**
2. Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=1, gbsa=1. Four TRP residues in 1LE3 and TYR and TRP residues in 1L2Y were represented as a *stick* expression. ... **18**
3. Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=2, gbsa=0. Four TRP residues in 1LE3 and TYR and TRP residues in 1L2Y were represented as a *stick* expression. ... **20**
4. Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=2, gbsa=1. Four TRP residues in 1LE3 and TYR and TRP residues in 1L2Y were represented as a *stick* expression. ... **22**
5. Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=5, gbsa=0. Four TRP residues in 1LE3 and TYR and TRP residues in 1L2Y were represented as a *stick* expression. ... **24**
6. Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=5, gbsa=1. Four TRP residues in 1LE3 and TYR and TRP residues in 1L2Y were represented as a *stick* expression. ... **26**
7. Comparison between QM and MM energies of the all dipeptide models. (IGB=1 (gbsa=0,1)). ... **48**
8. Comparison between QM and MM energies of the all dipeptide models. (IGB=2 (gbsa=0,1)). ... **49**
9. Comparison between QM and MM energies of the all dipeptide models. (IGB=5 (gbsa=0,1)). ... **50**
10. Conformational stabilities of ACE-ALA-NME dipeptide model in the presence of explicit water molecules. ... **51**
11. Comparison between the QM and MM energies of all dipeptide models using the original developed force field. ... **55**
12. About 1LE3 protein. (a) The potential energy plot for all simulation structures (80-

- ps + 30-ns) in 375K. (b) The RMSD($C\alpha$) plot between PDB structure and prediction structure for all simulation structures. (c) The principal component analysis for all simulation structures. The start point is [-89.704, 23.147], and the end point is [114.997, 27.753]. ... **59-60**
13. About 1L2Y protein. (a) The potential energy plot for all simulation structures (80-ps + 30-ns) in 375K. (b) The RMSD($C\alpha$) plot between PDB structure and prediction structure for all simulation structures. (c) The principal component analysis for all simulation structures. The start point is [5.104, 60.290], and the end point is [110.083, 21.398]. ... **61-62**
14. About 1VII protein. (a) The potential energy plot for all simulation structures (80-ps + 30-ns) in 375K. (b) The RMSD($C\alpha$) plot between PDB structure and prediction structure for all simulation structures. (c) The principal component analysis for all simulation structures. The start point is [-347.833, 22.757], and the end point is [38.370, 170.579]. ... **63-64**
15. About 1PGB protein. (a) The potential energy plot for all simulation structures (80-ps + 30-ns) in 375K. (b) The RMSD($C\alpha$) plot between PDB structure and prediction structure for all simulation structures. (c) The principal component analysis for all simulation structures. The start point is [-509.793, -52.367], and the end point is [331.422, -209.840]. ... **65-66**
16. Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=5, gbsa=1 by using the *original* force field. ... **70**
17. (a) 4 TRP residues can not be formed π - π interactions. (b) TRP forms the hydrophobic interaction with PRO17 instead of PRO12. (c) PHE18 makes π - π interaction with PHE11, but cannot interact with PHE7. (d) The helix region of PDB structure and prediction structure is conformed to be highly accurate. ... **71**
18. Sequences (black) and secondary structures (blue) of 1CE4 and 1NIZ are expressed by the one-letter code. These two structures have only one different amino acid (represented by black *box* emphasis). ... **85**

Table

1. Energy and RMSD(C α) comparisons between the PDB structure and the prediction structure in case of IGB=1 (gbsa=0, 1).	... 29
2. Energy and RMSD(C α) comparisons between the PDB structure and the prediction structure in case of IGB=2 (gbsa=0, 1).	... 32
3. Energy and RMSD(C α) comparisons between the PDB structure and the prediction structure in case of IGB=5 (gbsa=0, 1).	... 35
4. Conformational stabilities of all amino acids in case of IGB=1 (gbsa=0).	... 39
5. Conformational stabilities of all amino acids in case of IGB=1 (gbsa=1).	... 40
6. Conformational stabilities of all amino acids in case of IGB=2 (gbsa=0).	... 41
7. Conformational stabilities of all amino acids in case of IGB=2 (gbsa=1).	... 42
8. Conformational stabilities of all amino acids in case of IGB=5 (gbsa=0).	... 43
9. Conformational stabilities of all amino acids in case of IGB=5 (gbsa=1).	... 44
10. The main-chain torsion angles and the energy differences of all amino acids.	... 47
11. Atomic partial charges of standard amino acid.	... 53
12. Conformational stabilities of all amino acids in case of IGB=5 (gbsa=1) using the original developed force field.	... 54
13. Comparisons about 4 viewpoints (potential energies, principal component analysis, structural differences between the initial structure and the prediction structure, structural differences between the PDB structure and the prediction structure) for 5 stable structures.	... 67
14. Energy and RMSD(C α) comparisons between the PDB structure and the prediction structure in case of IGB=5 (gbsa=1) by using the original developed force field.	... 72
15. Comparisons of energy differences between the native structure and the all_helix structure. Two different force fields, ff03 force field and the original developed force field, are compared.	... 74

概要

Molecular dynamics (MD) シミュレーションはタンパク質の構造、機能解明において大変重要な役割を担っている。MD シミュレーションの結果は使用する力場パラメータに大きく依存する。ところが、力場パラメータの精度が不十分であるために実験的事実を再現できない場合が少なくなく、現在も深刻な問題となっている。本研究ではまず、MD シミュレーションプログラムである AMBER8 と、現在提供されている力場パラメータの中で最新の ff03 force field を使ってタンパク質立体構造予測に取り組んだ。この取り組みを通して、ff03 force field の問題提示、溶媒効果の重要性と最適な溶媒効果の選択、タンパク質立体構造予測法の考察を行った。次いで、ff03 force field の問題点を解消した新規高精度力場パラメータの開発を行った。最後に新規力場パラメータを使った MD シミュレーションを実行することで、(1) コンフォメーションレベルでの新規力場パラメータの精度評価、(2) タンパク質立体構造予測から、タンパク質構造レベルでの新規力場パラメータの精度評価、(3) β -sheet 構造を持つタンパク質に対する MD シミュレーションにより、タンパク質二次構造レベルでの新規力場パラメータの精度評価を行った。また、タンパク質立体構造折り畳みシミュレーションの効果的なシミュレーション方法についてもあわせて考察を行った。

序論

近年、コンピュータ技術の著しい発展により、これまで計算機シミュレーションが困難であった巨大タンパク質や複雑な生命現象について、簡便に且つ高速でシミュレーションを行うことが可能になってきた^[1,2]。これにともない、生命科学分野における計算科学の役割は重要なものとなってきている。特に創薬分野においては、新薬を迅速且つ簡便に設計するための手法の一つとして計算科学に多くの期待が寄せられている。これまでも、薬と受容体の結合エネルギーをシミュレーションすることによって薬の効果を予測する試み^[3,4]、薬と受容体の結合によって引き起こされる受容体の構造変化をシミュレーションする試み^[5,6]、受容体の構造から薬の構造を予測する試み^[7,8]など、数多くの計算科学を用いた創薬研究がなされてきた。しかし、これらの予測結果と実験的結果の相関は必ずしも高いとは言えない。コンピュータが高速化し、計算科学が多くの創薬研究に用いられていく一方で、生体分子シミュレーション精度向上のための技術がまだ成熟していない点が問題である。

創薬研究において、タンパク質立体構造予測も計算科学に期待されていることのひとつである。SNP 解析により、病気の直接的な原因である DNA の識別が明確化されるようになった^[9]。しかし、薬の直接のターゲットとなる受容体の立体構造まで生化学的実験により決定することは難しい。現在 NCBI の GenBank にはタンパク質について 2,135,138 種の Sequence データが登録されているが^[10]、一方で Protein Data Bank (PDB) には、同じくタンパク質について 31444 種のタンパク質立体構造が登録されるに留まっている^[11]。タンパク質立体構造登録件数は近年急激に増加しているが、全てのタンパク質立体構造を生化学的実験により決定することはまだまだ難しいのが現状である。そこで、理論的方法によりタンパク質立体構造を予測する取り組みがこれまでに数多くなされてきた。その方法の一つであるホモロジー法は、近年のコンピュータ技術の発展とタンパク質立体構造登録件数の増加により、タンパク質立体構造を一通り予測することを可能にした^[12]。しかし、新たに構造を知りたいアミノ酸配列と、すでに構造が分かっているタンパク質のアミノ酸配列の相同性が低い場合、予測精度は著しく低下してしまう。将来的には鋳型となるタンパク質立体構造の登録数の増加に従って、予測精度は向上していくと考えられているが、全く同じ配列を持っていても違う二次構造を示す部分構造も存在するため、予

測精度が飛躍的に向上するとは考えにくい。実際、PDB code 1AMP^[13]の ALA63~SER69 と PDB code 1GKY^[14]の ALA83~SER89 はともに ALA-SER-VAL-LYS-GLN-VAL-SER の 7 残基の配列を持つが、1AMP は β -sheet 構造をとり、1GKY は α -helix 構造をとる。これは、配列相同性だけに依存した構造予測法には限界があることを示している。この限界を打開するためには、「Molecular dynamics (MD) シミュレーションを用いた、配列相同性に依存しないタンパク質立体構造予測法」が必要である。このような MD シミュレーションを用いたタンパク質立体構造予測法は、配列相同性が低いタンパク質立体構造予測に有効であるだけでなく、配列相同性が高く、これを重視するタンパク質立体構造予測においても最終的な構造の最適化に有効であり、大変応用が利く手法である。そのため、ホモロジー法を用いてタンパク質立体構造予測を行っている研究者からも期待されている方法である。また、MD シミュレーションを用いることにより、タンパク質構造に関する基礎物理的理解の進展や、タンパク質立体構造構成原理の解明に大きく寄与するものと期待できる。このように MD シミュレーションを用いる立体構造予測法は多くの期待が寄せられているが、MD シミュレーションにより立体構造予測に成功した例は、現在までに PDB code 1L2Y で示される 20 残基の mini-protein の成功例だけである^[15]。しかも 1L2Y は、タンパク質を構成する 2 次構造が helix 構造と strand 構造だけの単純な立体構造である。さらに strand 領域は PRO の連続体であるためにとりうる構造の自由度が低く、比較的予測しやすい。世界的に見て、MD シミュレーションだけを使い、物理化学的法則に基づき、helix 構造と β -sheet と turn 構造が絡み合った複雑な立体構造を高精度で予測した例はない。

一方、これまで多くの研究者が MD シミュレーション精度の向上を目指し、様々な取り組みを行ってきた。MD シミュレーション中の溶媒項計算に用いる scoring function を改良する取り組み^[16]、MD シミュレーションに Langevin Dynamics を導入する取り組み^[17]、MD シミュレーションに実際に水分子を配置し MD シミュレーションを実行する取り組み^[18]、溶媒項計算そのものを改良する取り組み^[19-21]などである。またごく少数ではあるが、MD シミュレーションに用いる力場パラメータを改良する取り組み^[22]も行われてきた。その結果、比較的分子量の小さな分子や、ごく一部のタンパク質で生化学的実験事実との相関が得られるようになってきた。しかし、タンパク質立体構造予測の分野が要求するほどの高いシミュレーション精度は得られていない。また、これまでの

低い予測精度は「何」が原因で「どうすればよい」のかが明らかにされていない。我々はこの「何を」、「どうすればよいのか」を明らかにし、タンパク質立体構造予測精度向上に取り組んできた。

最初の取り組みとして、タンパク質立体構造の構成原理を解明する目的で、 α -helix 構造や β -sheet 構造のような二次構造の構成原理の詳細な研究を行った。局所的二次構造が後のタンパク質全体の構造形成を引き起こすことが考えられるからだ。

重要な問題の一つとして、「特定のアミノ酸配列に対する局所的二次構造の差異は何によって生じているのか」ということが挙げられる。この構造形成の疑問に対する妥当な説明として以下のことが考えられる。まず、タンパク質合成の初期過程において、適当な長さのペプチドは helix 構造のような二次構造で安定化する。そして、エネルギー的に安定な方向へとタンパク質合成を進め、立体構造を形成していく。この考えに基づけば β -sheet 構造も一度 helix 様構造をとり、その helix 構造の構造変化から形成されることになる。実際に、 β -lactoglobulin のタンパク質立体構造形成の過程において、helix 構造から β -sheet 構造への構造変化が確認されている^[23]。また、分子動力的研究により、 β -sheet 構造はタンパク質立体構造形成の後期過程に形成されることが確認されている^[24]。もう一つの重要な問題は、タンパク質合成過程で、なぜ特定のアミノ酸配列が来ると α -helix 構造の形成が終わり、他の二次構造が形成されていくのか？ また、そのようにして、なぜ特定部位に turn 構造や β -sheet 構造が現れるのか？ ということである。これらの問題に対する答えが局所的二次構造の形成過程を解明するであろうと考えられる。これらの観点からすると、 β -sheet 構造の構成メカニズムを理論的に調査することが重要である。そこで我々は、HIV-1 protease と myosin を例にとり、タンパク質の特定部位で β -sheet 構造が形成される要因を量子化学的に解明した。

その結果、 β -sheet 構造の安定化には、水分子の存在と水素結合による安定化エネルギーが最も主要な役割を果たすことが分かった。この研究は β -sheet 構造を形成している一部分の構造を抜き出して研究対象としており、タンパク質全体を考慮していないにも関わらず、実際に存在する β -sheet 構造の安定性を正当に評価することができた。そこで、「 β -sheet 構造は近傍のわずかな残基と水分子までを含めた局所的相互作用によって安定化される」、「 β -sheet 構造の安定性は水分子の存在を考慮すると、potential energy で評価可能である」と

いう2つの結論を導き出した^[25]。一方 α -helix構造についても構造安定性の要因を量子化学的に解明した。 α -helixの場合も β sheet構造の場合と同様に、タンパク質だけではなく周囲の環境までを含めた局所的相互作用によって安定化することが分かった^[26]。

ここまでの研究で、局所的二次構造の形成要因について核心に迫る理解を得ることができた。特に β -sheet構造の安定性には水分子の存在が必要不可欠であることが分かった。

一般的にMDシミュレーション実行時には、水分子を実際に配置し、水の溶媒効果を表現している。また、シミュレーションモデルが巨大な場合やタンパク質立体構造予測シミュレーションのような長いシミュレーション時間が必要な場合、計算コスト削減を目的として、連続溶媒モデルにより水分子の溶媒効果を近似する方法が用いられる。このような連続溶媒モデルにはGeneralized-Born (GB)法^[27-28]とPoisson-Boltzmann (PB)法^[29-30]がある。PB法はGB法に比べ精度の高い方法であるが計算量が膨大になる欠点があるため、生体分子シミュレーション実行時にはGB法を用いるのが一般的である。PB法はシミュレーション後のある目的の構造に対して、ポテンシャルエネルギーの再評価に多く用いられている。

本研究では、以上の知見を基に、水の溶媒効果に注目して、これまでのMDシミュレーションを用いた立体構造予測法の問題点を精査した。次いで、この精査から導かれた既存力場パラメータの問題点を明らかにし、この問題点を改善するために新規力場パラメータの開発を行った。最後に、新たに開発した力場パラメータにより、タンパク質立体構造予測を再び行い、新規力場パラメータの評価、立体構造予測精度評価を行った。このとき、主に β -sheet構造からなるいくつかのタンパク質について、新規力場パラメータの β -sheet構造に対する信頼性評価も行った。

MDシミュレーションに用いる高精度力場パラメータを開発したことにより、タンパク質立体構造予測分野の発展だけではなく、生体分子シミュレーションを使う全ての分野に大きな発展があることを強く確信している。

方法

(1) 既存力場パラメータ ff03 force field を用いたタンパク質立体構造予測

全てのタンパク質立体構造予測は CPU : Intel® Xeon™ 3.20GHz 4CPU、OS : Red Hat Linux 7.3.2 のコンピュータ上で行われた。構造予測に用いる Molecular dynamics (MD) シミュレーションは AMBER8 ソフトウェアパッケージ^[31]の sander モジュールを用いて実行された。この AMBER8 ソフトウェアは Intel® Fortran Compiler for IA-32 Version 8.1 により、コンパイルされた。またコンパイルの際、Intel® Math Kernel Library Version 7.2 を使用した。MD シミュレーションの並列化計算には MPICH Version 1.2.6 を使用した。

予測対象として、PDB code 1LE3^[32]で示される β -sheet 構造のみからなる 17 残基のタンパク質、PDB code 1L2Y^[33]で示される 2 つの α -helix と 1 つの strand 構造からなる 20 残基のタンパク質、PDB code 1VII^[34]で示される 3 つの α -helix 構造からなる 36 残基のタンパク質、PDB code 1PGB^[35]で示される 1 つの α -helix と 4 本の β -sheet 構造からなる 56 残基のタンパク質、の合計 4 つのタンパク質を選択した。1LE3 は β -sheet 構造を形成するタンパク質の中で残基数が少なく、MD シミュレーションモデルとして優れたモデルである。1L2Y は helix 構造と strand 構造の 2 つの二次構造を持つタンパク質の中で最も残基数が少ない。また、このタンパク質についてののみ MD シミュレーションによるタンパク質立体構造予測が成功している。1VII は helix 構造のみからなり、且つ、2 本以上の helix 構造を持つタンパク質の中で最も残基数の少ないタンパク質である。1PGB は helix 構造と sheet 構造の両方を持ち、多くの研究者がこのタンパク質立体構造予測を目指している。

立体構造予測に用いられる全ての初期構造は直鎖状構造であり、AMBER8 プログラムパッケージの Leap モジュールを使って作成された。100 step のエネルギー極小化計算の後、DT=1-fs^{*1} の time step を使い、80-ps で 500K まで昇温し、5-ns で 300K まで徐々に温度を落とした。さらに 1LE3 は 288K で NMR によって測定された構造であること、1L2Y は 282K で NMR によって測定された構造であること、1VII は 303K で NMR によって測定された構造であること、1PGB は X 線結晶解析構造であることから、1-ns

かけて 300K から 1LE3 では 288K まで、1L2Y では 282K まで、1VII では 303K まで、1PGB では 5K まで温度変化を行った。

溶媒項計算に関して、Generalized-Born 法を使用した。AMBER8 に IGB=1^{*2}, IGB=2^{*3}, IGB=5^{*4} として用意されている 3 種類の GB 法全てについて MD シミュレーションを行った。また、各溶媒条件について、タンパク質の表面積計算 (SA 計算) を導入しない場合 (gbsa=0^{*5}) と導入する場合 (gbsa=1^{*6}) について MD シミュレーションを行った。よって、合計 6 パターンの条件により、タンパク質立体構造予測を行った。

遠距離相互作用項について、CUT=200 Å^{*7}、RGBMAX=200 Å^{*8} のキーワードを用い、すべての原子間の相互作用計算に近似を用いない、No Cut Off による MD シミュレーションを実行した。温度変化に関して nmropt=1^{*9}、IGB キーワードの実行に関して nrespa=2^{*10} キーワードを用いて MD シミュレーションを実行した。これら 2 つのキーワードは AMBER8 ユーザーズマニュアルにより使用が推奨されている。

シミュレーション中の温度の取り扱いについて、IGB=1、2 の全てのパターンでは ntt=1^{*11} キーワードを用いた。一方、IGB=5 の全てのパターンでは ntt=3^{*12} キーワードを用いて Langevin dynamics を実行した。このとき、溶媒である水分子の衝突頻度 gamma_ln^{*13} は 1-ps⁻¹ と定義した。ここで、300K における実際の水分子の衝突頻度は 50-ps⁻¹ 程度であるが、AMBER8 開発者等はこれよりもかなり小さな 5~10-ps⁻¹ を推奨値としている。また、これより小さい値も良いとされ、ユーザーズマニュアルの Example には 1-ps⁻¹ が記載されている。

溶媒項計算を実行する際の各原子の Born 半径は IGB=1 の全ての場合でデフォルトの半径を使った。IGB=2、5 については mbondi2 半径を使った。これらの半径についても AMBER8 ユーザーズマニュアルにより推奨されている半径である。

力場パラメータには最新の力場パラメータとして提供されている ff03 force field を使用した。これまでの研究でそれ以前の力場パラメータ (ff94、ff96、ff98、ff99、ff02、ff02EP) よりも、最も良くタンパク質構造を表現できることが分かっている^[22]。

その他のキーワードについてはすべてデフォルト値を使用した。

全ての MD シミュレーション終了後、MD シミュレーションの最後の snap

shot の構造を予測構造とした。最後に PDB 構造と予測構造の主鎖 C α の RMSD を比較した。

(2) 予測構造と Native 構造のポテンシャルエネルギー比較

(1) のときと同じマシン、ソフトウェアを使用した。

1LE3、1L2Y、1VII、1PGB の 4 つの対象タンパク質の PDB 構造を初期構造とした。構造を緩和させる目的で 5 step のエネルギー極小化計算の後、DT=1-fs の time step を使い、5-ps で 300K まで昇温し、300K の状態を保ち、わずか 5-ps の MD シミュレーションを行った。1-ps おきに構造を得、最後の snap shot を Native 構造とした。各シミュレーション条件、力場パラメータは (1) の場合と同じである。よって 4 つのタンパク質に対して合計 6 パターンの MD シミュレーションを実行した。

一方予測構造に関して、タンパク質立体構造予測シミュレーション中の 300K 付近の 5 つの構造を取りだした。

予測構造の GB-ポテンシャルエネルギーについて、5 つの構造の各ポテンシャルエネルギーの平均をとり、予測構造の GB-ポテンシャルエネルギーとした。Native 構造の GB-ポテンシャルエネルギーは、300K における MD シミュレーションの 5 つの snap shot 由来のエネルギーを平均して得た。このようにして得られた両者の GB-ポテンシャルエネルギーを比較した。この比較により、予測構造が PDB 構造と比べ、ポテンシャルエネルギー的に十分に安定な構造であるかの確認ができ、シミュレーション時間とシミュレーション方法の妥当性を評価できる。

PB 法によるポテンシャルエネルギーについて、Native 構造では 300K における MD シミュレーションの 5 つの snap shot 構造を使った。AMBER8 の pbsa プログラムを使用し、この構造の PB-ポテンシャルエネルギーを再計算し、その平均をとって Native 構造の PB-ポテンシャルエネルギーとした。予測構造では GB-ポテンシャルエネルギー計算に用いたときと同じ 5 つの構造を使った。まず 5 つの PB-ポテンシャルエネルギーを計算し、その平均をとって予測構造の PB-ポテンシャルエネルギーとした。MM-PBSA 計算に用いた条件は以下の通りである。Reaction field energy は (溶媒比誘電率) / (分子の比誘電率) によって計算された。このとき溶媒比誘電率は 80.0、分子の比誘電率は 1.0 であった。PB 計算に用いる格子点は 2.0Åお

きに配置された。PB 計算の繰り返し数は最大 500 回までを許容した。溶媒接触半径は 1.6 Å とした。分子の cavity 半径は pbsa 計算の初期化過程で最適化された。表面張力は $0.0050 \text{ kcal mol}^{-1} \text{ \AA}^2$ とした。

(3) 既存力場パラメータ ff03 force field のアミノ酸コンフォメーション安定性と量子化学的手法によるアミノ酸コンフォメーション安定性の比較

全ての MD シミュレーションは CPU : Intel® Xeon™ 1.70GHz 2CPU、OS : Red Hat Linux 2.4.18 のコンピュータ上で行われた。MD シミュレーションは AMBER8 ソフトウェアパッケージの sander モジュールを用いて実行された。この AMBER8 ソフトウェアは Intel® Fortran Compiler for IA-32 Version 8.1 により、コンパイルされた。またコンパイルの際、Intel® Math Kernel Library Version 7.2 を使用した。MD シミュレーションの並列化計算には MPICH Version 1.2.6 を使用した。

ALA、ARG、ASN、ASP、CYS、GLN、GLU、GLY、HID (δ 位に水素がついた HIS)、ILE、LEU、LYS、MET、PHE、PRO、SER、THR、TRP、TYR、VAL の 20 種類のアミノ酸について、N 末端側に ACE (acetyl 基) を、C 末端側に NME (N-methyl 基) を結合し、ACE-XXX-NME (XXX は任意のアミノ酸) ジペプチドモデルを作成した。これらのモデルは AMBER8 プログラムパッケージの Leap モジュールを使用し、sequence コマンドにより作成された。10000 step のエネルギー極小化計算の後、DT=1-fs の time step を使い、20-ps で 300K まで昇温し、300K の状態を保ち 6-ns の MD シミュレーションを行った。各シミュレーション条件、力場パラメータは (1) の場合と同じである。よって 20 種類のアミノ酸に対して合計 6 パターンの MD シミュレーションを実行した。

6-ns のシミュレーションの間 1-ps ずつ構造を得、合計 6000 個の構造を得た。これらの構造に対し、phi-psi 角をそれぞれ求め、(phi : psi) = (-110 ~ -30 : -90 ~ 40)、(-180 ~ -110, 160 ~ 180 : -180 ~ -150, 110 ~ 180)、(-110 ~ -50 : 40 ~ 110)、(40 ~ 90 : -140 ~ -30)、(-110 ~ -40 : 110 ~ 180, -180 ~ -150)、(-180 ~ -110 : -20 ~ 50)、(30 ~ 80 : 10 ~ 80) に存在した構造をそれぞれ αR 、C5、C7eq、C7ax、 β 、 $\beta 2$ 、 αL のコンフォメーションとみなした。Boltzmann 分布の式よりそれぞれのコンフォメーションの存在比から、C7eq コンフォメーションに対する各コンフォメーションのエネルギー差を求めた。

ACE-ALA-NME モデルについてのみ、実際に水分子を発生させたモデルのコンフォメーション安定性解析を行った。AMBER8 の leap モジュールを使い、タンパク質の周囲 x、y、z 方向に 10 Å の厚さの水分子を発生させた。タンパク質と水分子の原子間の衝突を避けるため、タンパク質と水分子の間に 1.0 Å の空間を作った。合計 630 個の水分子を発生させた。遠距離相互作用項について、CUT=8.0 Å、キーワードを用い、原子間距離が 8.0 Å 以上の原子間の相互作用計算を省略する Cut Off による MD シミュレーションを実行した。なお、この MD シミュレーションは周期的境界条件下で行われた。連続溶媒モデルの場合と同様にして、10000 step のエネルギー極小化計算の後、DT=1-fs の time step を使い、20-ps で 300K まで昇温し、300K の状態を保ち 6-ns の MD シミュレーションを行った。昇温過程について、ntb=1^{*14}、ntp=0^{*15} のキーワードを使い一定体積のもと MD シミュレーションを行った。次に、ntb=2^{*14}、ntp=1^{*15} のキーワードを使い、一定圧力のもと MD シミュレーションを行った。その他のキーワードについてはすべてデフォルト値を使用した。6-ns のシミュレーションの間 1-ps ずつ構造を得、合計 6000 個の構造を得た。これらの構造に対し、連続溶媒モデルと同様の方法を使ってコンフォメーションを分類した。また、Boltzmann 分布の式よりそれぞれのコンフォメーションの存在比から、C7eq コンフォメーションに対する各コンフォメーションのエネルギー差を求めた。

より精度の高いコンフォメーション安定性解析を行う目的で、量子化学計算を使った安定性解析を行った。タンパク質構造の主要な二次構造の α -helix 構造と β -sheet 構造を形成している α R コンフォメーションと C5 コンフォメーションに注目して量子化学計算を実行した。全ての量子化学計算は CPU : Intel 1.6GHz 2CPU、IA-64、Itanium2、OS : Red Hat Linux 2.4.29 のコンピュータ上で行われた。量子化学計算は Gaussian03 ソフトウェアパッケージを用いて実行された。

α R コンフォメーションとして、(ϕ : ψ) = (-60 : -40) の主鎖構造を持つものを初期構造とし、C5 コンフォメーションとして、(ϕ : ψ) = (-180 : -180) の主鎖構造を持つものを初期構造とした。PRO についてのみ主鎖が側鎖とつながった構造をしているため、 α R コンフォメーションで (-61 : -35)、C5 コンフォメーションで (-65 : 150) の初期構造を使用した。^[36]

量子化学計算の計算条件は以下の通りである。各モデルの構造最適化計算には HF 法を使い、基底系は 6-31G**を用いた。水の溶媒効果を取り入れるため、SCRF 法の IEFPCM 法を使用し、Solvent=Water キーワードを使用した。分子の Cavity 半径には UAHF 半径を使用した。次いで、最適化された構造に対して、B3LYP/cc-pVTZ で Single Point Energy 計算を行いポテンシャルエネルギーの再評価を行った。このとき、SCF 計算については収束条件を最も厳しくするために SCF=Tight キーワードを用いた。水の溶媒効果は構造最適化計算の時と同じものを用いた。こうして求めた α R-ポテンシャルエネルギーと C5-ポテンシャルエネルギーの差をとり、MD シミュレーションのコンフォメーション安定性と比較した。

(4) 高精度新規力場パラメータの開発

(3) で求めた、量子化学的アミノ酸コンフォメーション安定性を MD シミュレーションが再現するようにするため、高精度新規力場パラメータの開発を行った。特に、静電的相互作用項が MD シミュレーションに最も大きな影響を与えるため、新しい電荷パラメータを作成した。

(3) の量子化学計算で使用したのと同じマシン、ソフトウェア、ACE-XXX-NME モデルを使った。(3) の量子化学計算実行時と同じ計算条件で構造最適化を行った後、最適化された構造に対して、B3LYP/cc-pVTZ で静電ポテンシャル計算を行った。このとき、SCF 計算については (3) と同様に、収束条件を最も厳しくするために SCF=Tight キーワードを用いた。水の溶媒効果は (3) と同じく構造最適化計算の時と同じものを用いた。

こうして得られた静電ポテンシャルから、RESP 法^[37]を使い各アミノ酸に電荷が割り当てられた。その際、2-stage-fitting 法を使用した。1st-stage では α R コンフォメーションと C5 コンフォメーションの 2 つのコンフォメーションを組み合わせて電荷が割り当てられた。この際、ACE、NME の Blocking group は合計の電荷が 0 になるように調整した。2nd-stage では化学的に同等である原子に同じ電荷を割り振るように指定した。この時、Blocking group と重原子は 1st-stage の電荷を保存するように固定した。こうして各アミノ酸の水の溶媒効果を加味した電荷を得た。次に、全てのアミノ酸に対するジペプチドモデルの ACE と NME が同じ電荷を持つように

操作し、ACE、NME の電荷を得た。最後に、Piotr Cieplak 等と同様の方法^[37]を使い、N 末端、C 末端のアミノ酸についても水の溶媒効果を加味した電荷を得た。

こうして新たに作成された力場パラメータを使って、(3) のときと同様に ACE-XXX-NME モデルの MD シミュレーションを行った。(3) と同じく、MD シミュレーションの結果から各アミノ酸のコンフォメーション安定性を求め、量子化学的アミノ酸コンフォメーション安定性と比較した。

(5) 高精度新規力場パラメータを用いたタンパク質立体構造予測

(1) と同じ 1LE3、1L2Y、1VII、1PGB の 4 つのタンパク質を対象にタンパク質立体構造予測を行った。使用したマシン、ソフトウェアは (1) と同じである。(1) と同じく、立体構造予測に用いられる全ての初期構造は直鎖状構造であり、AMBER8 プログラムパッケージの Leap モジュールを使って作成された。100 step のエネルギー極小化計算の後、80-ps で 5K から 375K まで昇温した。その後 375K で 30-ns の MD シミュレーションを行った。通常、タンパク質の折り畳みは秒オーダーでおこる現象であるため、加速的に折り畳みシミュレーションを行う目的で、やや高い温度で MD シミュレーションを行った。1-ps ごとに構造を得、合計 30000 個の予測構造を得た。このうち、MD simulation 中におけるポテンシャルエネルギーが最も低い方から、5 つの予測構造を抜き出した。この 5 つの予測構造が集団を形成する場合は、その集団において最も初期構造から離れている構造を最終的な予測構造として採用した。5 つの予測構造が全く集団を形成しない場合は、5 つの予測構造の中で最も初期構造から離れた構造を最終的な予測構造として採用した。初期構造からの予測構造の離れ具合は、30000 個の予測構造に対する主成分解析の結果から求めた。

(6) β -sheet 構造における新規力場パラメータの信頼性評価

PDB code 1B03^[38]、1J4M^[39]、1LE0^[40]、1LE1^[40]、1LE3、1NIZ^[41]で示されるの 6 つの β -sheet 構造に対して、既存力場パラメータである ff03 force field と本研究で開発した新規力場パラメータの 2 つの力場パラメータを使い、MD シミュレーションを実行した。全ての MD シミュレーションは CPU: Intel® Xeon™ 1.70GHz 2CPU、OS: Red Hat Linux 7.3.2 のコンピュータ上で

行われた。MD シミュレーションに使用したソフトウェアは (1) と同じである。

1B03 は残基数 18、1J4M は残基数 14、1LE0 は残基数 13、1LE1 は残基数 13、1LE3 は残基数 17、1NIZ は残基数 16 である。この 6 つのタンパク質は、PDB に登録されている構造の中で以下の 8 つの条件を全て満たしたものである。①Protein であること。(DNA や RNA やウイルス等の構造ではない。) ②残基数が 30 残基以下であること。③1 量体であること。④タンパク質構造の組成が β -sheet 構造 30%以上 100%以下となるもの。⑤タンパク質構造組成が helix 構造 0%となるもの。⑥ β -sheet 構造は 1 本以上 2 本以下となるもの。⑦複合体構造などの他の構造を含まず、単独で PDB に登録されているもの。⑧CYS によるジスルフィド結合を含まないもの。

ここで、1NIZ は"YNKRKRIHIGPGRAFYTTKNIIG"の配列を持つが N 末側の YNKR と C 末側の KNIIG の構造が無い。そこで、1NIZ のみ、N 末に ACE、C 末に NME を付加した。また、NMR での測定条件が"pH = 5"、"Ionic strength = 10mM"であったことから、HIS 残基は δ 位と ϵ 位に水素がついた構造 (HIP) とし、イオン濃度に関しては saltcon=0.01^{*16} オプションを使用した。

まず、タンパク質の主鎖 Φ , Ψ 角を全てのアミノ酸に対して (Φ , Ψ) = (-60, -40) としたものを"all_helix 構造"の初期構造とし、PDB 構造のままのものを"Native 構造"の初期構造とした。この計算モデルに対し、MD シミュレーションを実行した。両者ともに 500-step の minimize 後、5-ps で 100K から 300K まで昇温した。その後 300K で 5-ps の MD シミュレーションを行い、1-ps ずつ構造を得、合計 5 つの構造を得た。1LE1 の場合のみ、Native 構造に対して 5-ps の MD シミュレーションでは、NMR 測定構造の側鎖の異常接近が解消されなかったため、50-ps の MD シミュレーションを行い最後 5 つの構造を得た。

このようにして得られた各々の 5 つの構造に対して MMPBSA(GBSA)計算を行った。MMPBSA(GBSA)計算に用いたマシンは MD シミュレーションに用いたものと同じである。GBSA 計算は MD シミュレーションと同じ計算条件で実行された。また、MD シミュレーションに用いた計算 Option は、IGB=5、ntt=3、gamma_ln=1.0、cut=200.0、rgbmax=200.0、nrespa=2、gbsa=1 であり、残りは default 値を使用した。PBSA 計算に用いた計算条件は (2)

のときと同じものを用いた。

最後に、各々の方法の Native 構造のポテンシャルエネルギーと all_helix 構造のポテンシャルエネルギーの差をとり、ff03 force field のときのポテンシャルエネルギー差と比べることで、新規力場パラメータの信頼性評価を行った。

結果

(1) 既存力場パラメータ ff03 force field を用いたタンパク質立体構造予測

IGB=1 (gbsa=0), IGB=1 (gbsa=1), IGB=2 (gbsa=0), IGB=2 (gbsa=1), IGB=5 (gbsa=0), IGB=5 (gbsa=1) の6つの溶媒条件パターンで、1LE3、1L2Y、1VII、1PGB の4つのタンパク質について立体構造予測を行った。

IGB=1 (gbsa=0) の時の予測構造と PDB 構造を、主鎖 C α 原子について最も距離変位が小さくなるように重ねあわせた図を Fig.1 に示す。

1LE3 について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 6.374 Å であった。PDB 構造では4つの TRP の側鎖の環構造が交互に並ぶことで、 π - π 相互作用をとり、 β -sheet 構造を安定化している。一方予測構造では、4つの TRP 側鎖の環構造が交互に並ぶことは無く、2つの helix 構造を形成した。また、PDB 構造では 7ASP、8ALA、9THR、10LYS で turn 構造をとるが、予測構造では turn 構造が予測できていなかった。構造予測には完全に失敗した。1L2Y について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 1.567 Å であった。PDB 構造では1つの helix 構造、1つの 3-10helix 構造、1つの strand 構造を形成している。予測構造では、2つの helix 構造と1つの strand 構造が形成された。3-10helix 構造を予測することはできなかったが、全体的に良く似ている構造が予測できた。また、1L2Y の構造形成に重要な PRO 側鎖と疎水的相互作用を形成する TYR、TRP の2つのアミノ酸側鎖について、側鎖の向きもほぼ一致していた。1VII について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 5.574 Å であった。PDB 構造では3本の helix 構造を形成している。また、N 末端側から数えて1本目の helix 構造と2本目の helix 構造は 7PHE、11PHE、18PHE の3つの側鎖環構造の作る π - π 相互作用により構造が安定化されている。予測構造では、4本の helix 構造が形成された。7PHE、11PHE、18PHE の3つの側鎖環構造の作る π - π 相互作用は形成されなかったため、1本目と2本目の helix 構造の距離が PDB 構造に比べると大きくなっていた。全体的な構造は似ているが、予測は完全には成功していなかった。1PGB について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 11.520 Å であった。PDB 構造では1本の helix 構造と4つの β -sheet 構造を形成している。一方予測構造では、4本の helix 構造を形成し、 β -sheet 構造を1つも予測できなかった。

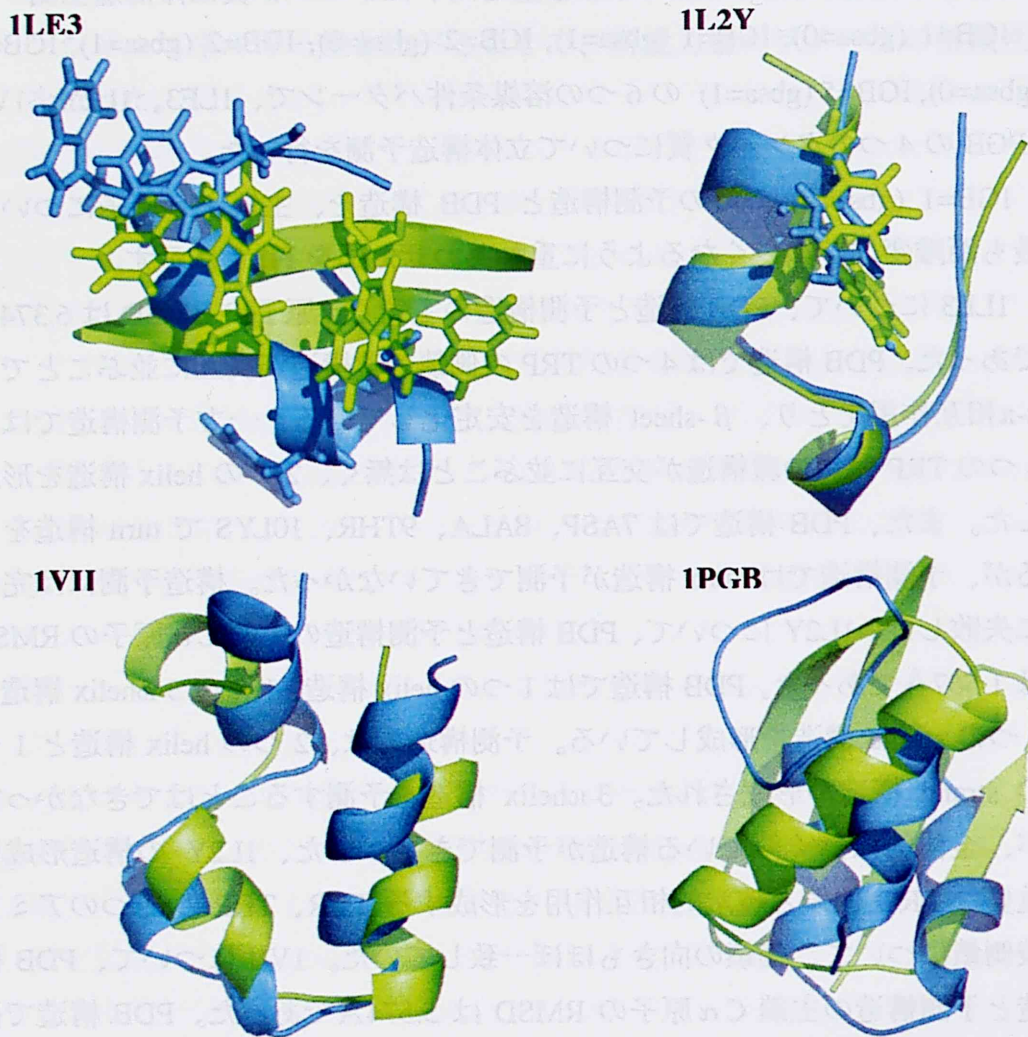


Fig. 1 : Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=1, gbsa=0. Four TRP residues in 1LE3 and TYR and TRP residues in 1L2Y were represented as a *stick* expression.

IGB=1 (gbsa=1) の時の予測構造と PDB 構造を、主鎖 C α 原子について最も距離変位が小さくなるように重ねあわせた図を Fig. 2 に示す。

1LE3 について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 5.604 Å であった。IGB=1 (gbsa=0) の時と同様に、予測構造は 2 つの helix 構造を形成した。また、4 つの TRP 側鎖の環構造が交互に並ぶ π - π 相互作用はみられなかった。さらに、turn 構造をとるべき領域についても予測に失敗した。1L2Y について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 3.299 Å であった。IGB=1 (gbsa=0) の時と同様に、予測構造は 2 つの helix 構造と 1 つの strand 構造を形成した。3- α helix 構造を予測することはできなかった。TYR、TRP の 2 つのアミノ酸側鎖の向きは PDB に近い向きを示したが全体構造としてはあまり似ていなかった。1VII について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 5.085 Å であった。予測構造は 4 本の helix 構造を形成した。7PHE、11PHE、18PHE の 3 つの側鎖環構造の作る π - π 相互作用は形成されなかったため、1 本目と 2 本目の helix 構造の距離が PDB 構造に比べると大きくなっていた。また、C 末端側の 3 本目の helix 構造をとるべき領域で、helix 構造は 2 本にわかれてしまった。全体的な構造は似ているが、予測は完全には成功していなかった。1PGB について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 10.894 Å であった。予測構造は 4 本の helix 構造を形成した。 β -sheet 構造は一つも予測できなかった。

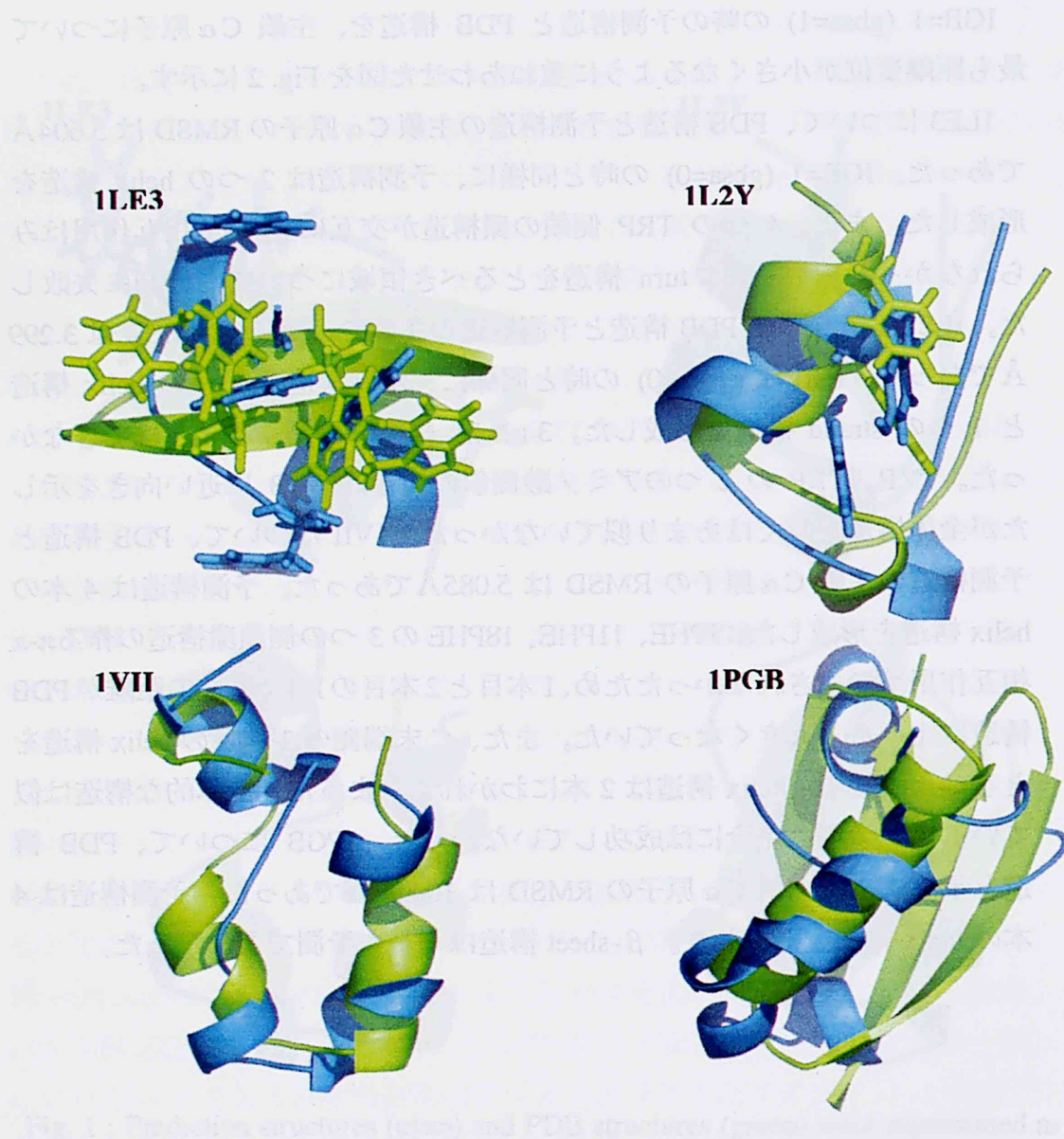


Fig. 2 : Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=1, gbsa=1. Four TRP residues in 1LE3 and TYR and TRP residues in 1L2Y were represented as a *stick* expression.

IGB=2 (gbsa=0) の時の予測構造と PDB 構造を、主鎖 C α 原子について最も距離変位が小さくなるように重ねあわせた図を Fig. 3 に示す。

1LE3 について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 6.823 Å であった。IGB=1 (gbsa=0, 1) の 2 つの場合と同様に、予測構造は 2 つの helix 構造を形成した。また、4 つの TRP 側鎖の環構造が交互に並ぶ π - π 相互作用はみられなかった。さらに、turn 構造をとるべき領域についても予測に失敗した。1L2Y について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 6.030 Å であった。予測構造は helix 構造を形成せず、全体が strand 様構造になっていた。TYR, TRP の位置も PDB 構造とは大きくずれていた。IGB=1 (gbsa=0, 1) の 2 つの場合とは異なり、予測には大きく失敗した。1VII について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 9.799 Å であった。予測構造は PDB 構造と同じ 3 本の helix 構造を形成した。しかし、7PHE、11PHE、18PHE の 3 つの側鎖環構造の作る π - π 相互作用は形成されなかった。二次構造的には PDB 構造に近い構造を予測できたが、三次構造的には完全に予測に失敗した。1PGB について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 10.586 Å であった。予測構造は 6 本の helix 構造を形成した。helix 構造を形成するべき領域では正しく helix 構造が予測できたが、 β -sheet 構造を形成するべき領域では小さな 5 本の helix 構造が形成された。 β -sheet 構造は一つも予測できなかった。

IGB=2 (gsa=0) の時の予測構造と PDB 構造を、主鎖 Ca 原子について
 最も類似度が小さくなるように重ねあわせて図を Fig. 3 に示す。

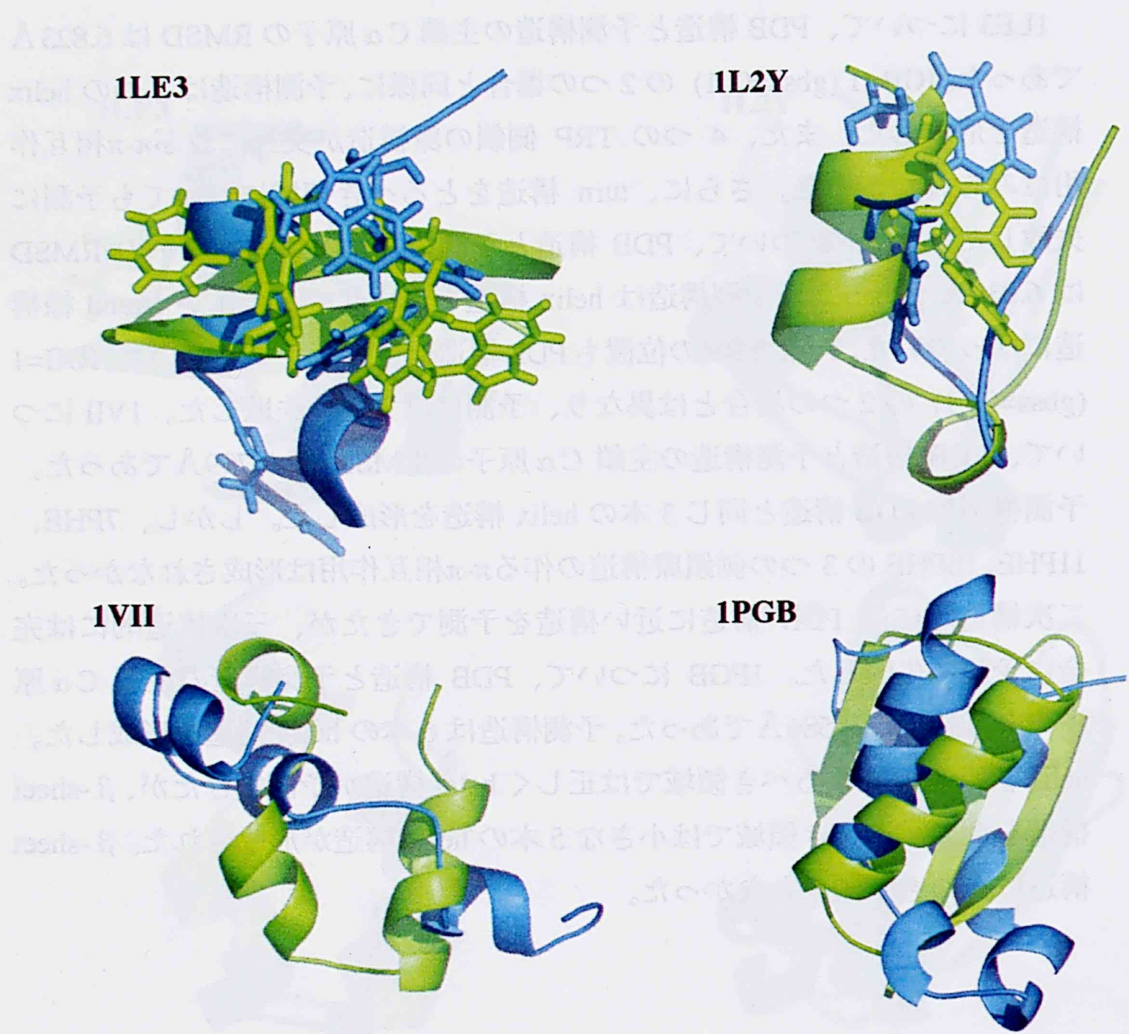


Fig. 3 : Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=2, gbsa=0 Four TRP residues in 1LE3 and TYR and TRP residues in 1L2Y were represented as a *stick* expression.

IGB=2 (gbsa=1) の時の予測構造と PDB 構造を、主鎖 C α 原子について最も距離変位が小さくなるように重ねあわせた図を Fig. 4 に示す。

1LE3 について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 6.232 Å であった。予測構造は 1 本の helix 構造を形成した。この helix 構造は turn 構造を形成するべき領域で形成された。また、4 つの TRP 側鎖の環構造が交互に並ぶ π - π 相互作用はみられなかった。これまでの場合と同様に、 β -sheet 構造は予測されなかった。1L2Y について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 3.478 Å であった。予測構造は 1 本の helix 構造を形成した。3-10helix 構造を予測することはできなかった。TYR に対する TRP の向きは PDB 構造に近い構造だった。予測にはやや成功している。1VII について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 6.701 Å であった。予測構造は 2 本の helix 構造を形成した。PDB 構造において N 末端から数えて 2 本目の helix 構造を形成するべき領域で helix 構造の予測に失敗した。また、7PHE、11PHE、18PHE の 3 つの側鎖環構造の作る π - π 相互作用は形成されなかった。1PGB について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 11.632 Å であった。予測構造は 3 本の helix 構造を形成した。これまでの場合と同様に、 β -sheet 構造は一つも予測できなかった。予測には大きく失敗した。

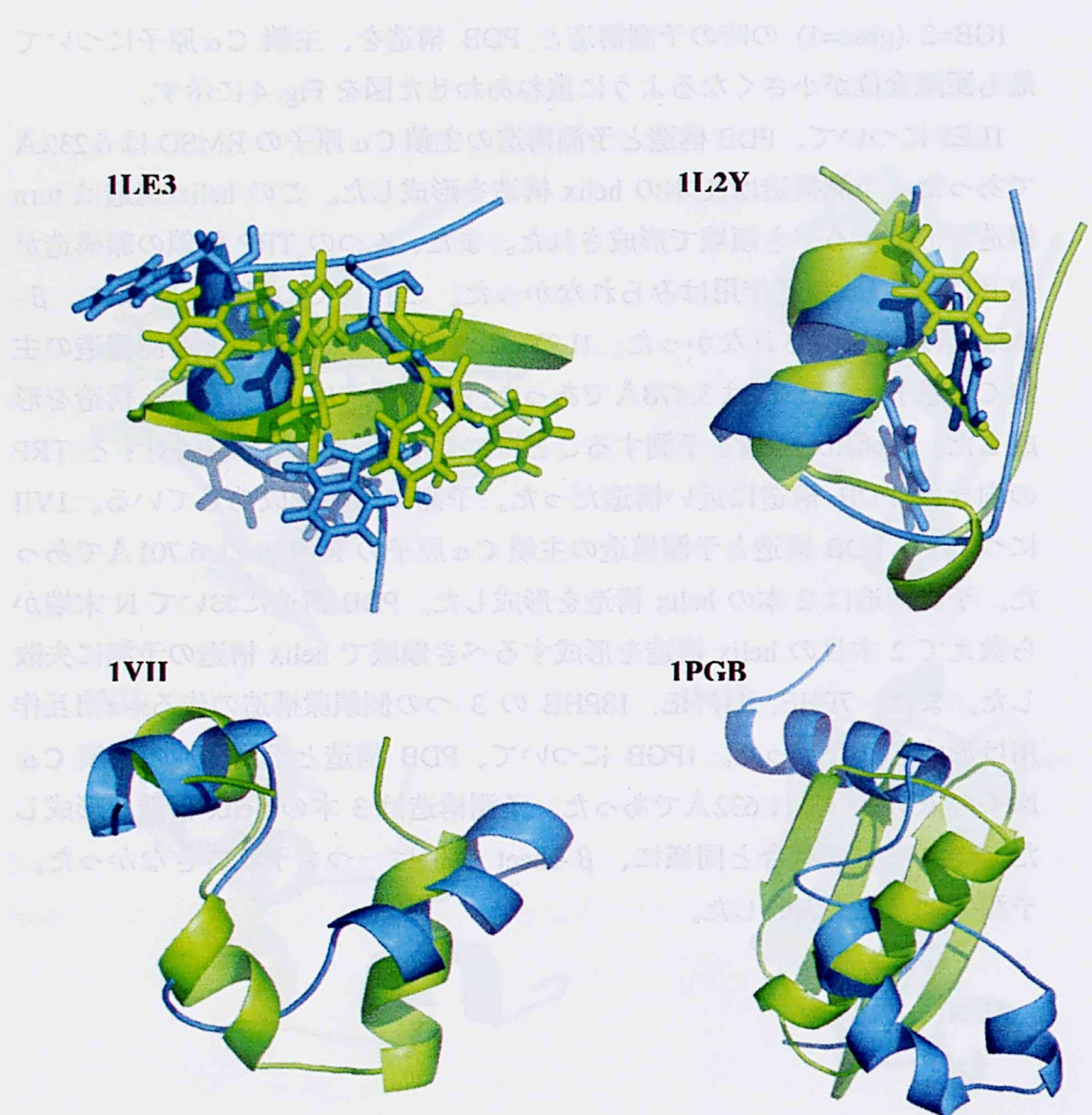
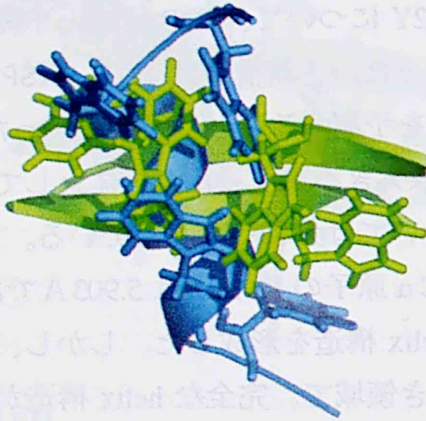


Fig. 4 : Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=2, gbsa=1 Four TRP residues in 1LE3 and TYR and TRP residues in 1L2Y were represented as a *stick* expression.

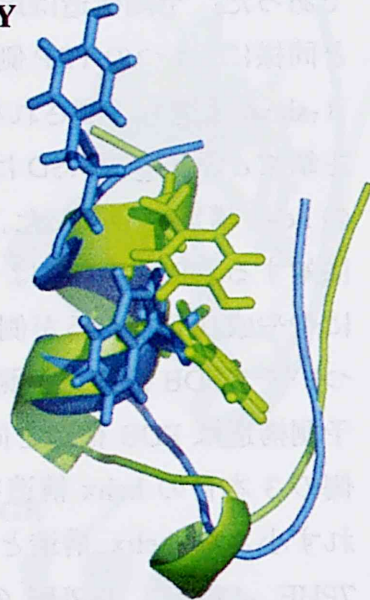
IGB=5 (gbsa=0) の時の予測構造と PDB 構造を、主鎖 C α 原子について最も距離変位が小さくなるように重ねあわせた図を Fig. 5 に示す。

1LE3 について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 8.019 Å であった。予測構造は比較的長い 1 本の helix 構造を形成した。これまでと同様に、4 つの TRP 側鎖の環構造が交互に並ぶ π - π 相互作用はみられず、 β -sheet 構造も予測されなかった。1L2Y について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 2.960 Å であった。予測構造は 4ILE~9ASP で 1 本の helix 構造を形成した。3- α helix 構造を予測することはできなかった。TYR に対する TRP の向きは PDB 構造とは大きく異なった。全体としては予測にやや成功しているが側鎖の向きについては大きく失敗している。1VII について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 5.903 Å であった。予測構造は PDB 構造と同じ 3 本の helix 構造を形成した。しかし、C 末端側の 3 本目の helix 構造を形成すべき領域で、完全な helix 構造が形成されず小さな helix 構造と長い無定形構造が予測された。これまでと同様に 7PHE、11PHE、18PHE の 3 つの側鎖環構造の作る π - π 相互作用は形成されなかった。1PGB について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 13.648 Å であった。予測構造は 3 本の helix 構造を形成した。これまでの場合と同様に β -sheet 構造は 1 つも予測できず、予測には大きく失敗した。

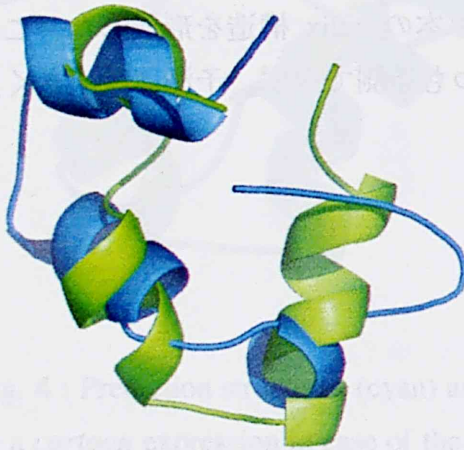
1LE3



1L2Y



1VII



1PGB

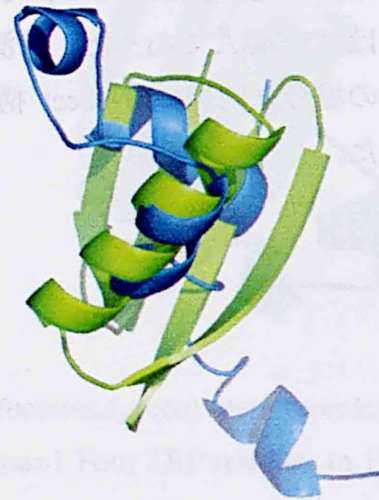


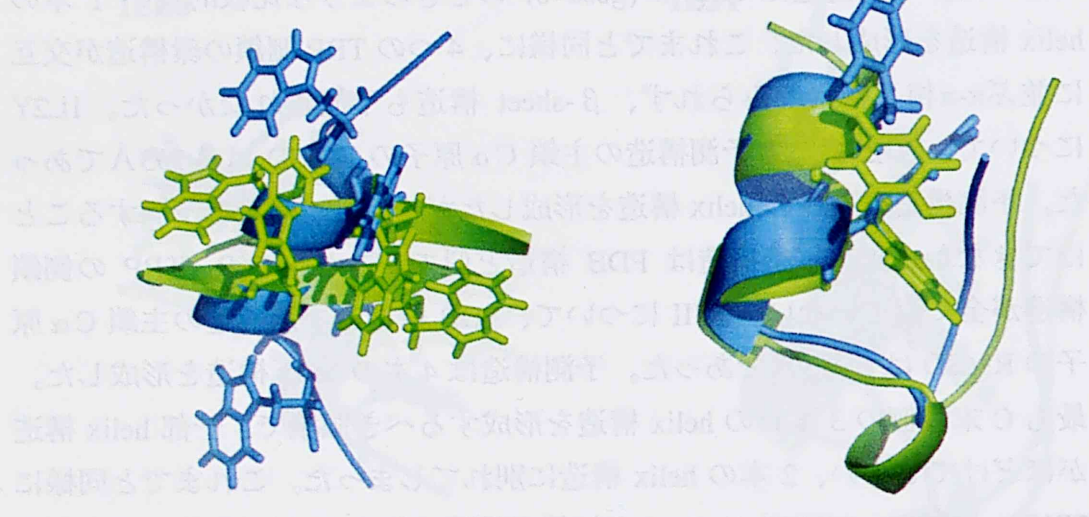
Fig. 5 : Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=5, gbsa=0 Four TRP residues in 1LE3 and TYR and TRP residues in 1L2Y were represented as a *stick* expression.

IGB=5 (gbsa=1) の時の予測構造と PDB 構造を、主鎖 C α 原子について最も距離変位が小さくなるように重ねあわせた図を Fig. 6 に示す。

1LE3 について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 7.812 Å であった。予測構造は IGB=5 (gbsa=0) のときのような比較的長い 1 本の helix 構造を形成した。これまでと同様に、4 つの TRP 側鎖の環構造が交互に並ぶ π - π 相互作用はみられず、 β -sheet 構造も予測されなかった。1L2Y について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 3.133 Å であった。予測構造は 1 本の helix 構造を形成したが 3- α helix 構造を予測することはできなかった。主鎖構造は PDB 構造と似ているが、TYR、TRP の側鎖構造が全く似ていない。1VII について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 6.463 Å であった。予測構造は 4 本の helix 構造を形成した。最も C 末端側の 3 本目の helix 構造を形成するべき領域で、一部 helix 構造がほどけてしまい、2 本の helix 構造に別れてしまった。これまでと同様に 7PHE、11PHE、18PHE の 3 つの側鎖環構造の作る π - π 相互作用は形成されなかった。1PGB について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 14.507 Å であった。予測構造は 4 本の helix 構造を形成した。これまでの場合と同様に β -sheet 構造は 1 つも予測できず、予測には大きく失敗した。

1LE3 (gbsa=1) の時の予測構造と PDB 構造を、主鎖 C α 原子にのみ
最も精度が良いように重ね合わせるように表示した図を Fig. 6 に示す。

1LE3 について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 0.17 Å
であった。PDB 構造は gbsa=0) のときの RMSD は 0.17 Å 以下に
なっており、この結果から、この構造は PDB 構造と予測構造が互
にほぼ等しいと判断される。



1LE3 の場合、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 0.17 Å
であった。PDB 構造は gbsa=0) のときの RMSD は 0.17 Å 以下に
なっており、この結果から、この構造は PDB 構造と予測構造が互
にほぼ等しいと判断される。

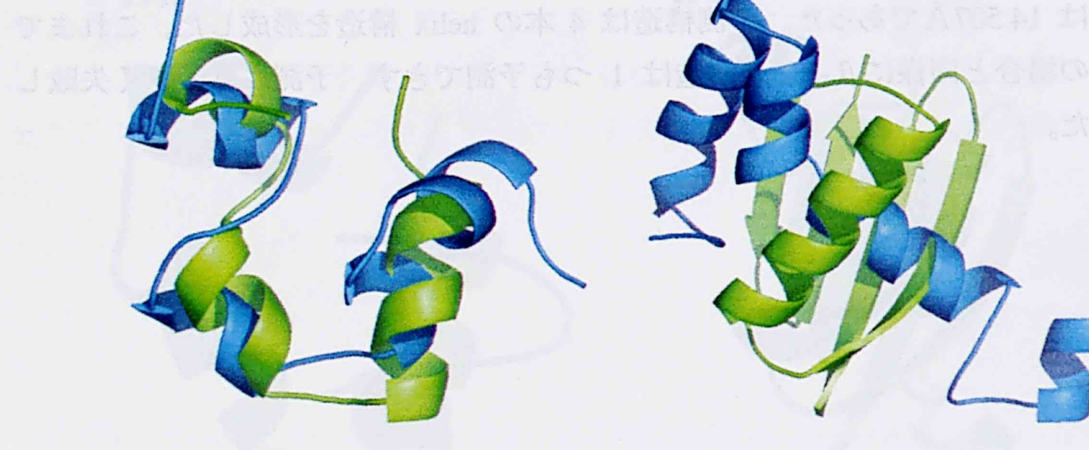


Fig. 6 : Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=5, gbsa=1 Four TRP residues in 1LE3 and TYR and TRP residues in 1L2Y were represented as a *stick* expression.

TYR and TRP residues in 1L2Y were represented as a *stick* expression.

(2) 予測構造と Native 構造のポテンシャルエネルギー比較

予測構造と Native 構造のポテンシャルエネルギー比較を行った。ポテンシャルエネルギー計算に先立ち、PDB 構造を緩和する目的で MD シミュレーションを行った。MD シミュレーション後の構造を Native 構造とし、予測構造との比較に用いた。

Table 1 に IGB=1 (gbsa=0, 1)の時の各タンパク質における 1. PDB 構造と Native 構造の主鎖 C α 原子の RMSD、2. (1) で求めた PDB 構造と予測構造の主鎖 C α 原子の RMSD、3. GB 法によるポテンシャルエネルギー差、4. PB 法によるポテンシャルエネルギー差を示す。

1LE3 における PDB-Native の RMSD(C α)は gbsa=0 で 1.722 Å、gbsa=1 で 0.828 Åであった。いずれも PDB 構造をよく保存し MD シミュレーションが実行された。1L2Y における PDB-Native の RMSD(C α)は gbsa=0 で 3.241 Å、gbsa=1 で 2.675 Åであった。両者ともに 3-10helix 構造領域において MD シミュレーション後に helix が緩くなってしまっていた。このため比較的大きな RMSD であったが、その他の領域で PDB 構造をよく保存していた。1VII における PDB-Native の RMSD(C α)は gbsa=0 で 16.466 Å、gbsa=1 で 11.432 Åであった。両者ともにたった 5-ps の MD シミュレーションで PDB 構造が崩壊した。したがって、これを PDB 構造由来の構造と考えることはできないので、ポテンシャルエネルギー比較の値は参考値とした。1PGB における PDB-Native の RMSD(C α)は gbsa=0 で 0.908 Å、gbsa=1 で 1.464 Åであった。両者ともに PDB 構造をよく保存し MD シミュレーションが実行された。IGB=1 では 1LE3、1L2Y、1PGB のタンパク質で PDB 構造をよく保存して MD シミュレーションが実行された。

1LE3 における GB-ポテンシャルエネルギー差 (ΔE_{GB} 予測構造 - ΔE_{GB} Native) は gbsa=0 の場合 6.12 kcal/mol で Native 構造の方が安定であった。gbsa=1 の場合 -4.61 kcal/mol で予測構造の方が安定であった。1L2Y における GB-ポテンシャルエネルギー差は gbsa=0 の場合 -8.84 kcal/mol で予測構造の方が安定であった。gbsa=1 の場合 -7.99 kcal/mol で予測構造の方が安定であった。1VII における GB-ポテンシャルエネルギー差は参考値として、gbsa=0 の場合 -70.88 kcal/mol で予測構造の方が安定であった。gbsa=1 の場合 -49.82 kcal/mol で予測構造の方が安定であった。1PGB における GB-ポテンシャルエネルギー差は gbsa=0 の場合 -3.08 kcal/mol で予測構造の方が安

定であった。gbsa=1 の場合 1.89 kcal/mol で Native 構造の方が安定であった。

より精度の高いエネルギー計算を実行する目的で、Native 構造と予測構造の 2 つの構造に対して、PB 法を用いてポテンシャルエネルギーの再評価を行った。1LE3 における PB-ポテンシャルエネルギー差 (ΔE_{PB} 予測構造 - ΔE_{PB} Native) は gbsa=0 の場合 0.32 kcal/mol で Native 構造の方が安定であった。GB 法では 6.12 kcal/mol あった両構造のエネルギー差が PB 法を用いた場合ほとんど無くなった。gbsa=1 の場合 1.45 kcal/mol で Native 構造の方が安定であった。GB 法では予測構造の方が安定であるが、PB 法では Native 構造が安定となった。1L2Y における PB-ポテンシャルエネルギー差は gbsa=0 の場合 -11.23 kcal/mol で予測構造の方が安定であった。GB 法よりも予測構造が 2.39 kcal/mol 安定化した。gbsa=1 の場合 -12.74 kcal/mol で予測構造の方が安定であった。GB 法よりも予測構造が 4.79 kcal/mol 安定化した。1VII における PB-ポテンシャルエネルギー差は参考値として、gbsa=0 の場合 -36.31 kcal/mol で予測構造の方が安定であった。gbsa=1 の場合 -15.46 kcal/mol で予測構造の方が安定であった。1PGB における PB-ポテンシャルエネルギー差は gbsa=0 の場合 42.63 kcal/mol で Native 構造の方が安定であった。GB 法では -3.08 kcal/mol で予測構造が安定であったが、PB 法では逆に、大きく Native 構造が安定であった。gbsa=1 の場合 52.38 kcal/mol で Native 構造の方が安定であった。GB 法では 1.89 kcal/mol しか無かった両構造のポテンシャルエネルギー差は PB 法では大きく開いた。

Table 1

Energy and RMSD($C\alpha$) comparisons between the PDB structure and the prediction structure in case of IGB=1 (gbsa=0, 1)

IGB=1
GBSA=0

	RMSD (Å) PDB-Native	RMSD (Å) PDB-Prediction	Δ EGB (kcal/mol) Δ EPrediction- Δ ENative	Δ EPB (kcal/mol) Δ EPrediction- Δ ENative
1LE3	1.722	6.374	6.12	0.32
1L2Y	3.241	1.567	-8.84	-11.23
1VLL	16.466	5.574	(-70.88)	(-36.31)
1PGB	0.908	11.520	-3.08	42.63

GBSA=1

	RMSD (Å) PDB-Native	RMSD (Å) PDB-Prediction	Δ EGB (kcal/mol) Δ EPrediction- Δ ENative	Δ EPB (kcal/mol) Δ EPrediction- Δ ENative
1LE3	0.828	5.604	-4.61	1.45
1L2Y	2.675	3.299	-7.99	-12.74
1VLL	11.432	5.085	(-49.82)	(-15.46)
1PGB	1.464	10.894	1.89	52.38

Table 2 に IGB=2 (gbsa=0, 1) の時の各タンパク質における 1. PDB 構造と Native 構造の主鎖 C α 原子の RMSD、2. (1) で求めた PDB 構造と予測構造の主鎖 C α 原子の RMSD、3. GB 法によるポテンシャルエネルギー差、4. PB 法によるポテンシャルエネルギー差を示す。

1LE3 における PDB-Native の RMSD(C α)は gbsa=0 で 0.848 Å、gbsa=1 で 0.963 Åであった。いずれも PDB 構造をよく保存し MD シミュレーションが実行された。1L2Y における PDB-Native の RMSD(C α)は gbsa=0 で 3.721 Å、gbsa=1 で 2.619 Åであった。両者ともに RMSD(C α)は小さいが、helix 構造が完全にほどけてしまった。また、3- α helix 構造も形成されていなかった。二次構造的特性を全く保存できなかったため、これを PDB 構造由来の構造として取り扱うことはできないと判断した。よって、ポテンシャルエネルギー比較の値は参考値とした。1VII における PDB-Native の RMSD(C α)は gbsa=0 で 16.350 Å、gbsa=1 で 16.948 Åであった。IGB=1 (gbsa=0, 1) のときと同様に、両者ともにたった 5-ps の MD シミュレーションで PDB 構造が崩壊した。したがって、これを PDB 構造由来の構造と考えることはできないと判断し、ポテンシャルエネルギー比較の値は参考値とした。1PGB における PDB-Native の RMSD(C α)は gbsa=0 で 2.408 Å、gbsa=1 で 0.983 Åであった。両者ともに PDB 構造をよく保存し MD シミュレーションが実行された。IGB=2 では 1LE3 と 1PGB のタンパク質で PDB 構造をよく保存して MD シミュレーションが実行された。

1LE3 における GB-ポテンシャルエネルギー差は gbsa=0 の場合 4.96 kcal/mol で Native 構造の方が安定であった。gbsa=1 の場合 11.43 kcal/mol で Native 構造の方が安定であった。1L2Y における GB-ポテンシャルエネルギー差は参考値として、gbsa=0 の場合 21.45 kcal/mol で Native 構造の方が安定であった。gbsa=1 の場合-5.63 kcal/mol で予測構造の方が安定であった。1VII における GB-ポテンシャルエネルギー差は参考値として、gbsa=0 の場合-48.56 kcal/mol で予測構造の方が安定であった。gbsa=1 の場合-36.81 kcal/mol で予測構造の方が安定であった。1PGB における GB-ポテンシャルエネルギー差は gbsa=0 の場合 11.33 kcal/mol で Native 構造の方が安定であった。gbsa=1 の場合 42.94 kcal/mol で Native 構造の方が安定であった。両者とも大きく Native 構造が安定であった。特に gbsa=1 のときに Native 構造の安定性が大きく増加した。参考値である 1L2Y と 1VII を除く全ての場

合で Native 構造の方が安定となった。

1LE3 における PB-ポテンシャルエネルギー差は gbsa=0 の場合 8.24 kcal/mol で GB 法の時と同様に Native 構造の方が安定であった。Native 構造の安定性は GB 法と比べると、3.28 kcal/mol 増加した。gbsa=1 の場合 6.52 kcal/mol で GB 法の時と同様に Native 構造の方が安定であった。Native 構造の安定性は GB 法と比べると 4.91 kcal/mol 減少した。1L2Y における PB-ポテンシャルエネルギー差は参考値として、gbsa=0 の場合 16.63 kcal/mol で Native 構造の方が安定であった。gbsa=1 の場合 -7.72 kcal/mol で予測構造の方が安定であった。1VII における PB-ポテンシャルエネルギー差は参考値として、gbsa=0 の場合 -6.40 kcal/mol で予測構造の方が安定であった。gbsa=1 の場合 -9.66 kcal/mol で予測構造の方が安定であった。1PGB における PB-ポテンシャルエネルギー差は gbsa=0 の場合 31.89 kcal/mol で Native 構造の方が安定であった。GB 法の場合と同様に、大きく Native 構造が安定であった。gbsa=1 の場合 59.09 kcal/mol で Native 構造の方が安定であった。GB 法においても Native 構造が非常に大きく安定であったが、PB 法においても同様に大きく安定した。参考値である 1L2Y、1VII を除く全ての場合で Native 構造の方が安定となった。

Table 2

Energy and RMSD($C\alpha$) comparisons between the PDB structure and the prediction structure in case of IGB=2 (gbsa=0, 1)

IGB=2
GBSA=0

	RMSD (Å) PDB-Native	RMSD (Å) PDB-Prediction	Δ EGB (kcal/mol) Δ EPrediction- Δ ENative	Δ EPB (kcal/mol) Δ EPrediction- Δ ENative
1LE3	0.848	6.823	4.96	8.24
1L2Y	3.721	6.030	(21.45)	(16.63)
1VLL	16.350	9.799	(-48.56)	(-6.40)
1PGB	2.408	10.586	11.33	31.89

GBSA=1

	RMSD (Å) PDB-Native	RMSD (Å) PDB-Prediction	Δ EGB (kcal/mol) Δ EPrediction- Δ ENative	Δ EPB (kcal/mol) Δ EPrediction- Δ ENative
1LE3	0.963	6.232	11.43	6.52
1L2Y	2.619	3.478	(-5.63)	(-7.72)
1VLL	16.948	6.701	(-36.81)	(-9.66)
1PGB	0.983	11.632	42.94	59.09

Table 3 に IGB=5 (gbsa=0, 1) の時の各タンパク質における 1. PDB 構造と Native 構造の主鎖 C α 原子の RMSD、2. (1) で求めた PDB 構造と予測構造の主鎖 C α 原子の RMSD、3. GB 法によるポテンシャルエネルギー差、4. PB 法によるポテンシャルエネルギー差を示す。

1LE3 における PDB-Native の RMSD(C α) は gbsa=0 で 0.704 Å、gbsa=1 で 0.611 Å であった。いずれも PDB 構造をよく保存し MD シミュレーションが実行された。1L2Y における PDB-Native の RMSD(C α) は gbsa=0 で 0.700 Å、gbsa=1 で 0.713 Å であった。両者ともに PDB 構造をよく保存していた。3-*iohelix* 構造についてもよく PDB 構造を保存していた。1VII における PDB-Native の RMSD(C α) は gbsa=0 で 1.584 Å、gbsa=1 で 1.600 Å であった。両者ともに PDB 構造をよく保存していた。また、7PHE、11PHE、18PHE の 3 つの側鎖環構造の作る π - π 相互作用もよく保存されていた。1PGB における PDB-Native の RMSD(C α) は gbsa=0 で 0.908 Å、gbsa=1 で 1.464 Å であった。両者ともに PDB 構造をよく保存していた。IGB=5 (gbsa=0, 1) では、全てのタンパク質について PDB 構造をよく保存して MD シミュレーションが実行された。

1LE3 における GB-ポテンシャルエネルギー差は gbsa=0 の場合 13.56 kcal/mol で Native 構造の方が安定であった。gbsa=1 の場合 -1.14 kcal/mol で予測構造の方が安定であった。1L2Y における GB-ポテンシャルエネルギー差は gbsa=0 の場合 2.25 kcal/mol で Native 構造の方が安定であった。gbsa=1 の場合 -23.53 kcal/mol で予測構造の方が安定であった。1VII における GB-ポテンシャルエネルギー差は gbsa=0 の場合 1.36 kcal/mol で Native 構造の方が安定であった。gbsa=1 の場合 -32.38 kcal/mol で予測構造の方が安定であった。1PGB における GB-ポテンシャルエネルギー差は gbsa=0 の場合 94.27 kcal/mol で Native 構造の方が安定であった。gbsa=1 の場合 56.84 kcal/mol で Native 構造の方が安定であった。両者とも大きく Native 構造が安定であった。全てのタンパク質について gbsa=0 では Native 構造の方が安定であった。gbsa=1 では最も残基数の多い 1PGB 以外で予測構造の方が安定であった。

1LE3 における PB-ポテンシャルエネルギー差は gbsa=0 の場合 -2.62 kcal/mol で予測構造の方が安定であった。GB 法では Native 構造の方が安定であったが、PB 法では逆に予測構造の方が安定であった。gbsa=1 の場

合-0.22 kcal/mol で GB 法の時と同様に予測構造の方が安定であった。1L2Y における PB-ポテンシャルエネルギー差は gbsa=0 の場合 3.94 kcal/mol で GB 法の場合と同様に Native 構造の方が安定であった。gbsa=1 の場合 -18.43 kcal/mol で GB 法の場合と同様に予測構造の方が安定であった。1VII における PB-ポテンシャルエネルギー差は gbsa=0 の場合 0.49 kcal/mol で GB 法の場合と同様に Native 構造の方が安定であった。gbsa=1 の場合 -23.62 kcal/mol で GB 法の場合と同様に予測構造の方が安定であった。1PGB における PB-ポテンシャルエネルギー差は gbsa=0 の場合 71.17 kcal/mol で GB 法の場合と同様に Native 構造の方が安定であった。gbsa=1 の場合 56.66 kcal/mol で Native 構造の方が安定であった。1LE3 を除く全てのタンパク質について gbsa=0 では Native 構造の方が安定となった。gbsa=1 では残基数の最も大きな 1PGB 以外の全てのタンパク質で予測構造の方が安定となった。また、gbsa=1 のときは GB 法と PB 法のポテンシャルエネルギー差の値がよく似ていた。

Table 3

Energy and RMSD($C\alpha$) comparisons between the PDB structure and the prediction structure in case of IGB=5 (gbsa=0, 1)

IGB=5
GBSA=0

	RMSD (Å) PDB-Native	RMSD (Å) PDB-Prediction	ΔE_{GB} (kcal/mol) $\Delta E_{Prediction} - \Delta E_{Native}$	ΔE_{PB} (kcal/mol) $\Delta E_{Prediction} - \Delta E_{Native}$
1LE3	0.704	8.019	13.56	-2.62
1L2Y	0.700	2.960	2.25	3.94
1VLL	1.584	5.903	1.36	0.49
1PGB	0.982	13.648	94.27	71.17

GBSA=1

	RMSD (Å) PDB-Native	RMSD (Å) PDB-Prediction	ΔE_{GB} (kcal/mol) $\Delta E_{Prediction} - \Delta E_{Native}$	ΔE_{PB} (kcal/mol) $\Delta E_{Prediction} - \Delta E_{Native}$
1LE3	0.611	7.812	-1.14	-0.22
1L2Y	0.713	3.133	-23.53	-18.43
1VLL	1.600	6.463	-32.38	-23.62
1PGB	0.867	14.507	56.84	56.66

(3) 既存力場パラメータ ff03 force field のアミノ酸コンフォメーション安定性と量子化学的手法によるアミノ酸コンフォメーション安定性の比較

既存力場パラメータ ff03 force field のアミノ酸コンフォメーション安定性について精査した。IGB=1 (gbsa=0) の時の全てのアミノ酸の各コンフォメーションの安定性を Table 4 に示す。ALA、ARG、GLY、MET、PHE、PRO、の 6 個のアミノ酸で β コンフォメーションが最も安定となった。ASN、ASP、CYS、HID、ILE、LEU、LYS、SER、THR、TRP、TYR の 11 個のアミノ酸で αR コンフォメーションが最も安定となった。GLN、GLU、VAL の 3 個のアミノ酸で C5 コンフォメーションが最も安定となった。C5 コンフォメーションが構造的に不利な PRO を除いた全てのアミノ酸について、C5 コンフォメーションと αR コンフォメーションだけに注目した場合、ALA、ARG、GLN、GLU、GLY、ILE、PHE、VAL の 8 個のアミノ酸で C5 コンフォメーションの方が安定であった。その他の 11 個のアミノ酸では αR コンフォメーションの方が安定であった。

IGB=1 (gbsa=1) の時の全てのアミノ酸の各コンフォメーションの安定性を Table 5 に示す。ALA、ASN、ASP、CYS、ILE、LEU、LYS、PHE、SER、THR、TRP、TYR の 12 個のアミノ酸で αR コンフォメーションが最も安定となった。ARG、GLY、MET、PRO の 4 個のアミノ酸で β コンフォメーションが最も安定となった。GLN、GLU、HID、VAL の 4 個のアミノ酸で C5 コンフォメーションが最も安定となった。PRO を除いた全てのアミノ酸について、C5 コンフォメーションと αR コンフォメーションだけに注目した場合、ALA、ASN、ASP、CYS、GLY、ILE、LEU、LYS、MET、PHE、THR、SER、TRP、TYR の 14 個のアミノ酸で αR コンフォメーションの方が安定であった。その他の 5 個のアミノ酸では C5 コンフォメーションの方が安定であった。

IGB=1 の方法を用いた場合、gbsa=0、gbsa=1 の両方法において、 αR コンフォメーションが最も安定となるアミノ酸が最も多く存在した。

IGB=2 (gbsa=0) の時の全てのアミノ酸の各コンフォメーションの安定性を Table 6 に示す。ALA、ASN、ASP、CYS、GLN、HID、ILE、LEU、LYS、MET、SER、THR、TRP、TYR の 14 個のアミノ酸で αR コンフォメーションが最も安定となった。ARG、GLY、PHE、PRO の 4 個のアミノ酸で β コンフォメーションが最も安定となった。GLU、VAL の 2 個のアミノ酸で

C5 コンフォメーションが最も安定となった。PRO を除いた全てのアミノ酸について、C5 コンフォメーションと α R コンフォメーションだけに注目した場合、ALA、ARG、ASN、ASP、CYS、GLN、HID、ILE、LEU、LYS、MET、SER、THR、TRP、TYR の 15 個のアミノ酸で α R コンフォメーションの方が安定であった。その他の 4 個のアミノ酸では C5 コンフォメーションの方が安定であった。

IGB=2 (gbsa=1) の時の全てのアミノ酸の各コンフォメーションの安定性を Table 7 に示す。ALA、ASN、ASP、CYS、GLN、HID、ILE、LEU、LYS、MET、PHE、SER、THR、TRP の 14 個のアミノ酸で α R コンフォメーションが最も安定となった。ARG、GLU、GLY、PRO、TYR の 5 個のアミノ酸で β コンフォメーションが最も安定となった。VAL だけで C5 コンフォメーションが最も安定であった。PRO を除いた全てのアミノ酸について、C5 コンフォメーションと α R コンフォメーションだけに注目した場合、ALA、ARG、ASN、ASP、CYS、GLN、GLU、HID、ILE、LEU、LYS、MET、PHE、SER、THR、TRP の 16 個のアミノ酸で α R コンフォメーションの方が安定であった。その他の 3 個のアミノ酸では C5 コンフォメーションの方が安定であった。

IGB=2 の方法を用いた場合、gbsa=0、gbsa=1 の両方法において、 α R コンフォメーションが最も安定となるアミノ酸が最も多く存在した。

IGB=5 (gbsa=0) の時の全てのアミノ酸の各コンフォメーションの安定性を Table 8 に示す。ALA、ASP、CYS、HID、ILE、LEU、LYS、MET、PHE、SER、THR、TRP、TYR の 13 個のアミノ酸で α R コンフォメーションが最も安定となった。ARG、ASN、GLU、GLY、PRO の 5 個のアミノ酸で β コンフォメーションが最も安定となった。GLN、VAL の 2 個のアミノ酸で C5 コンフォメーションが最も安定となった。PRO を除いた全てのアミノ酸について、C5 コンフォメーションと α R コンフォメーションだけに注目した場合、ALA、ARG、ASN、ASP、CYS、GLY、HID、ILE、LEU、LYS、MET、PHE、SER、THR、TRP、TYR の 16 個のアミノ酸で α R コンフォメーションの方が安定であった。その他の 3 個のアミノ酸では C5 コンフォメーションの方が安定であった。

IGB=5 (gbsa=1) の時の全てのアミノ酸の各コンフォメーションの安定性を Table 9 に示す。ALA、ASN、ASP、CYS、HID、ILE、LYS、MET、SER、

THR、TRP、TYR の 12 個のアミノ酸で α R コンフォメーションが最も安定となった。ARG、GLY、LEU、PHE、PRO の 5 個のアミノ酸で β コンフォメーションが最も安定となった。GLN、GLU、VAL の 3 個のアミノ酸で C5 コンフォメーションが最も安定となった。PRO を除いた全てのアミノ酸について、C5 コンフォメーションと α R コンフォメーションだけに注目した場合、ALA、ASN、ASP、CYS、HID、ILE、LEU、LYS、MET、SER、THR、TRP、TYR の 13 個のアミノ酸で α R コンフォメーションの方が安定であった。その他の 6 個のアミノ酸では C5 コンフォメーションの方が安定であった。

IGB=5 の方法を用いた場合、gbsa=0、gbsa=1 の両方法において、 α R コンフォメーションが最も安定となるアミノ酸が最も多く存在した。

Table 4
Conformational stabilities of all amino acids in case of IGB=1 (gbsa=0)

ALA		ARG		ASN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.713602	C5	= -1.160085	C5	= -1.218142
α R	= -1.614826	α R	= -0.910372	α R	= -1.525324
β	= -1.816121	β	= -1.554038	β	= -1.487849
C7ax	= 2.774533	C7ax	= 3.215497	C7ax	= 3.058271
β 2	= 0.023155	β 2	= 0.435313	β 2	= 0.266944
α L	= 2.774533	α L	= 3.215497	α L	= 3.058271
ASP		CYS		GLN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.516537	C5	= -1.182871	C5	= -2.752889
α R	= -2.008663	α R	= -2.385930	α R	= -1.744823
β	= -1.463298	β	= -1.449614	β	= -1.988183
C7ax	= 2.763067	C7ax	= 2.589629	C7ax	= 2.184091
β 2	= -0.284525	β 2	= -0.544711	β 2	= -0.567287
α L	= 2.763067	α L	= 2.589629	α L	= 2.184091
GLU		GLY		HID	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.804383	C5	= -1.193631	C5	= -1.761482
α R	= -1.064322	α R	= -0.808197	α R	= -2.039865
β	= -1.707797	β	= -1.486534	β	= -1.324140
C7ax	= 2.897255	C7ax	= -0.574369	C7ax	= 2.695736
β 2	= 0.347652	β 2	= 0.478332	β 2	= -0.819946
α L	= 2.897255	α L	= -0.072498	α L	= 2.695736
ILE		LEU		LYS	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -3.358615	C5	= -0.660390	C5	= -0.489441
α R	= -2.718716	α R	= -1.584672	α R	= -2.256005
β	= -2.534770	β	= -1.534903	β	= -1.415138
C7ax	= 1.529135	C7ax	= 3.102452	C7ax	= 2.739454
β 2	= -1.052700	β 2	= 0.595757	β 2	= -0.322334
α L	= 1.529135	α L	= 3.102452	α L	= 2.739454
MET		PHE		PRO	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.407845	C5	= -1.719853	C5	= 2.479384
α R	= -1.720304	α R	= -1.554471	α R	= -1.692344
β	= -1.761776	β	= -1.913767	β	= -2.575552
C7ax	= 2.839046	C7ax	= 2.751378	C7ax	= 2.479384
β 2	= -0.240023	β 2	= 0.000000	β 2	= 2.479384
α L	= 2.839046	α L	= 2.751378	α L	= 2.479384
SER		THR		TRP	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.274581	C5	= -2.223640	C5	= -1.459736
α R	= -2.654167	α R	= -2.530461	α R	= -2.878749
β	= -1.087391	β	= -2.050582	β	= -1.771297
C7ax	= 2.399777	C7ax	= 2.199184	C7ax	= 2.119577
β 2	= -0.815720	β 2	= -0.763649	β 2	= -1.369306
α L	= 2.399777	α L	= 2.199184	α L	= 2.119577
TYR		VAL			
C7eq	= 0.000000	C7eq	= 0.000000		
C5	= -1.355849	C5	= -2.088319		
α R	= -2.754390	α R	= -1.195975		
β	= -1.718502	β	= -1.710974		
C7ax	= 2.242300	C7ax	= 2.739454		
β 2	= -1.156111	β 2	= 0.206646		
α L	= 2.242300	α L	= 2.739454		

Table 5
Conformational stabilities of all amino acids in case of IGB=1 (gbsa=1)

ALA		ARG		ASN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.606811	C5	= -1.145650	C5	= -1.137087
α R	= -1.790811	α R	= -1.092865	α R	= -1.762193
β	= -1.654952	β	= -1.560548	β	= -1.369159
C7ax	= 2.802266	C7ax	= 3.184918	C7ax	= 1.096484
β 2	= -0.122092	β 2	= 0.482736	β 2	= -0.170518
α L	= 2.802266	α L	= 3.184918	α L	= 1.205178
ASP		CYS		GLN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.536978	C5	= -0.490110	C5	= -1.718428
α R	= -1.788635	α R	= -1.797398	α R	= -1.544112
β	= -1.490710	β	= -1.309538	β	= -1.419647
C7ax	= 2.863994	C7ax	= 3.089203	C7ax	= 0.592613
β 2	= 0.009854	β 2	= 0.201263	β 2	= -0.233849
α L	= 2.863994	α L	= 3.089203	α L	= 0.968955
GLU		GLY		HID	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.888919	C5	= -0.887141	C5	= -1.783676
α R	= -1.175249	α R	= -0.888741	α R	= -1.698793
β	= -1.641203	β	= -1.102098	β	= -1.321033
C7ax	= 2.863994	C7ax	= -0.449373	C7ax	= 2.844120
β 2	= 0.366264	β 2	= 0.623844	β 2	= -0.461726
α L	= 2.863994	α L	= 0.139273	α L	= 2.844120
ILE		LEU		LYS	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.567611	C5	= -1.030407	C5	= -0.423335
α R	= -1.913509	α R	= -1.715024	α R	= -2.157097
β	= -1.663475	β	= -1.529739	β	= -1.260964
C7ax	= 2.774533	C7ax	= 2.995066	C7ax	= 2.839046
β 2	= 1.534840	β 2	= 0.226238	β 2	= -0.334351
α L	= 2.774533	α L	= 2.995066	α L	= 2.839046
MET		PHE		PRO	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.229463	C5	= -1.719645	C5	= 2.524230
α R	= -1.461692	α R	= -1.985986	α R	= -1.751267
β	= -1.553071	β	= -1.967185	β	= -2.503721
C7ax	= 3.058271	C7ax	= 2.581836	C7ax	= 2.524230
β 2	= 0.133913	β 2	= -0.680546	β 2	= 2.524230
α L	= 3.058271	α L	= 2.581836	α L	= 2.524230
SER		THR		TRP	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -0.917392	C5	= -0.762296	C5	= -0.980786
α R	= -2.839509	α R	= -2.719459	α R	= -2.235964
β	= -0.914477	β	= -1.568955	β	= -1.582917
C7ax	= 2.256005	C7ax	= 2.320171	C7ax	= 2.669236
β 2	= -1.103265	β 2	= -0.964339	β 2	= -0.715052
α L	= 2.256005	α L	= 2.320171	α L	= 2.669236
TYR		VAL			
C7eq	= 0.000000	C7eq	= 0.000000		
C5	= -1.417238	C5	= -3.618019		
α R	= -1.981019	α R	= -1.108212		
β	= -1.926440	β	= -2.393535		
C7ax	= 2.675972	C7ax	= 1.481417		
β 2	= -0.131689	β 2	= 0.241724		
α L	= 2.675972	α L	= 1.481417		

Table 6
Conformational stabilities of all amino acids in case of IGB=2 (gbsa=0)

ALA		ARG		ASN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.611821	C5	= -1.073133	C5	= -0.860755
α R	= -2.005884	α R	= -1.449528	α R	= -1.854076
β	= -1.712492	β	= -1.702502	β	= -1.291321
C7ax	= 2.034266	C7ax	= 3.025646	C7ax	= 2.995066
β 2	= -0.343831	β 2	= 0.042464	β 2	= -0.357781
α L	= 2.689221	α L	= 3.025646	α L	= 2.581836
ASP		CYS		GLN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.486584	C5	= -0.754339	C5	= -1.755727
α R	= -1.710653	α R	= -1.802945	α R	= -1.947242
β	= -1.474705	β	= -1.586211	β	= -1.617565
C7ax	= 2.919925	C7ax	= 2.975126	C7ax	= 2.675972
β 2	= 0.237292	β 2	= 0.091899	β 2	= -0.652718
α L	= 2.919925	α L	= 2.975126	α L	= 2.675972
GLU		GLY		HID	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.809288	C5	= -1.183156	C5	= -1.881575
α R	= -1.389517	α R	= -0.931596	α R	= -2.213706
β	= -1.671420	β	= -1.471878	β	= -1.397516
C7ax	= 2.849151	C7ax	= -0.480287	C7ax	= 2.557826
β 2	= 0.086083	β 2	= 0.177211	β 2	= -0.927622
α L	= 2.849151	α L	= -0.167233	α L	= 2.557826
ILE		LEU		LYS	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -2.544985	C5	= -1.169983	C5	= -0.683761
α R	= -3.714010	α R	= -1.744490	α R	= -2.158264
β	= -1.974599	β	= -1.706613	β	= -1.613660
C7ax	= 1.309911	C7ax	= 2.892615	C7ax	= 2.751378
β 2	= -2.036445	β 2	= 0.090349	β 2	= -0.341165
α L	= 1.309911	α L	= 2.892615	α L	= 2.751378
MET		PHE		PRO	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.825546	C5	= -1.934901	C5	= 2.460457
α R	= -2.062416	α R	= -1.733761	α R	= -1.789310
β	= -1.836965	β	= -2.196548	β	= -2.575348
C7ax	= 2.573940	C7ax	= 2.515527	C7ax	= 2.460457
β 2	= -0.602359	β 2	= -0.118835	β 2	= 2.460457
α L	= 2.573940	α L	= 2.515527	α L	= 2.460457
SER		THR		TRP	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.345707	C5	= -1.401571	C5	= -1.496311
α R	= -2.364671	α R	= -1.803478	α R	= -2.470165
β	= -0.965499	β	= -1.651983	β	= -1.860885
C7ax	= 1.031600	C7ax	= 2.849151	C7ax	= 2.399777
β 2	= -0.870103	β 2	= -0.240052	β 2	= -1.012429
α L	= 0.735641	α L	= 2.849151	α L	= 2.399777
TYR		VAL			
C7eq	= 0.000000	C7eq	= 0.000000		
C5	= -1.511482	C5	= -2.050663		
α R	= -2.446366	α R	= -1.082083		
β	= -1.829547	β	= -1.677710		
C7ax	= 2.420698	C7ax	= 2.780184		
β 2	= -1.008790	β 2	= 0.222358		
α L	= 2.420698	α L	= 2.780184		

Table 7
Conformational stabilities of all amino acids in case of IGB=2 (gbsa=1)

ALA		ARG		ASN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.769751	C5	= -1.744087	C5	= -0.877838
α R	= -2.149663	α R	= -1.829865	α R	= -1.522449
β	= -1.934000	β	= -1.899482	β	= -1.375103
C7ax	= 2.524230	C7ax	= 2.662422	C7ax	= 3.158676
β 2	= -0.501416	β 2	= -0.525341	β 2	= 0.021240
α L	= 2.524230	α L	= 2.662422	α L	= 3.158676
ASP		CYS		GLN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.483573	C5	= -0.661863	C5	= -1.591987
α R	= -1.640647	α R	= -1.895862	α R	= -1.746074
β	= -1.473808	β	= -1.506586	β	= -1.465338
C7ax	= 0.972595	C7ax	= 2.958679	C7ax	= 2.849151
β 2	= 0.216352	β 2	= 0.129912	β 2	= -0.593440
α L	= 1.977969	α L	= 2.958679	α L	= 2.849151
GLU		GLY		HID	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.609843	C5	= -1.013061	C5	= -2.471346
α R	= -1.712538	α R	= -1.085758	α R	= -2.587635
β	= -1.724413	β	= -1.411676	β	= -1.757164
C7ax	= 0.053962	C7ax	= -0.577044	C7ax	= 2.119577
β 2	= -0.230148	β 2	= 0.388283	β 2	= -1.357192
α L	= 0.348130	α L	= -0.236607	α L	= 2.119577
ILE		LEU		LYS	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -3.005114	C5	= -0.871962	C5	= -0.669421
α R	= -3.329251	α R	= -1.802377	α R	= -2.228868
β	= -2.321184	β	= -1.542273	β	= -1.550472
C7ax	= 1.481417	C7ax	= 2.966959	C7ax	= 2.714866
β 2	= -1.773402	β 2	= 0.025194	β 2	= -0.481354
α L	= 1.481417	α L	= 2.966959	α L	= 2.714866
MET		PHE		PRO	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.009075	C5	= -1.850849	C5	= 2.641502
α R	= -2.043642	α R	= -2.643097	α R	= -1.457277
β	= -1.457190	β	= -2.016140	β	= -2.426300
C7ax	= 2.818308	C7ax	= 2.199184	C7ax	= 2.641502
β 2	= -0.685798	β 2	= -1.216906	β 2	= 2.641502
α L	= 2.818308	α L	= 2.199184	α L	= 2.641502
SER		THR		TRP	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.888154	C5	= -1.199743	C5	= -1.903204
α R	= -2.602345	α R	= -1.849262	α R	= -2.349955
β	= -1.297866	β	= -1.848269	β	= -2.190298
C7ax	= 2.332215	C7ax	= 2.791327	C7ax	= 2.320171
β 2	= -1.027055	β 2	= -0.203739	β 2	= -0.847382
α L	= 2.332215	α L	= 2.791327	α L	= 2.320171
TYR		VAL			
C7eq	= 0.000000	C7eq	= 0.000000		
C5	= -2.264285	C5	= -1.945457		
α R	= -2.157644	α R	= -0.884014		
β	= -2.503033	β	= -1.826691		
C7ax	= 2.184091	C7ax	= 2.807662		
β 2	= -0.649838	β 2	= 0.593756		
α L	= 2.184091	α L	= 2.807662		

Table 8
Conformational stabilities of all amino acids in case of IGB=5 (gbsa=0)

ALA		ARG		ASN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.605101	C5	= -0.907287	C5	= -1.112840
α R	= -1.978209	α R	= -0.980466	α R	= -1.349976
β	= -1.763683	β	= -1.599865	β	= -1.497503
C7ax	= 1.372723	C7ax	= 3.239400	C7ax	= 3.143583
β 2	= -0.502284	β 2	= 0.689797	β 2	= 0.435025
α L	= 2.682633	α L	= 3.239400	α L	= 3.143583
ASP		CYS		GLN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.544229	C5	= -0.790190	C5	= -2.124907
α R	= -1.639506	α R	= -1.922781	α R	= -1.943021
β	= -1.524952	β	= -1.391068	β	= -1.698509
C7ax	= 2.910960	C7ax	= 2.962833	C7ax	= 2.532808
β 2	= 0.202402	β 2	= 0.021068	β 2	= -0.688084
α L	= 2.910960	α L	= 2.962833	α L	= 2.532808
GLU		GLY		HID	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.530124	C5	= -1.046220	C5	= -1.898316
α R	= -1.450612	α R	= -1.052579	α R	= -2.037653
β	= -1.648889	β	= -1.353945	β	= -1.552452
C7ax	= 2.941765	C7ax	= -0.413231	C7ax	= 2.619821
β 2	= 0.139499	β 2	= 0.176146	β 2	= -0.704370
α L	= 2.941765	α L	= -0.031503	α L	= 2.619821
ILE		LEU		LYS	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -2.512891	C5	= -1.091232	C5	= -0.580853
α R	= -2.725441	α R	= -1.685023	α R	= -2.413311
β	= -1.882941	β	= -1.590893	β	= -1.521945
C7ax	= 2.027678	C7ax	= 2.979168	C7ax	= 2.589629
β 2	= -1.109759	β 2	= 0.150401	β 2	= -0.636610
α L	= 2.027678	α L	= 2.979168	α L	= 2.589629
MET		PHE		PRO	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.640911	C5	= -1.611496	C5	= 2.488628
α R	= -1.925794	α R	= -1.974062	α R	= -1.919718
β	= -1.762013	β	= -1.967314	β	= -2.496199
C7ax	= 2.708558	C7ax	= 2.619821	C7ax	= 2.488628
β 2	= -0.197869	β 2	= -0.272794	β 2	= 2.488628
α L	= 2.708558	α L	= 2.619821	α L	= 2.488628
SER		THR		TRP	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.640145	C5	= -1.602155	C5	= -1.854276
α R	= -2.543361	α R	= -1.894647	α R	= -2.533011
β	= -1.222137	β	= -1.696533	β	= -2.223180
C7ax	= 0.676636	C7ax	= 0.686175	C7ax	= 2.228271
β 2	= -1.016313	β 2	= -0.204059	β 2	= -0.801089
α L	= 0.200593	α L	= 1.010260	α L	= 2.228271
TYR		VAL			
C7eq	= 0.000000	C7eq	= 0.000000		
C5	= -1.456044	C5	= -3.253750		
α R	= -2.824866	α R	= -1.186391		
β	= -1.993602	β	= -2.138981		
C7ax	= 2.136372	C7ax	= -1.582214		
β 2	= -1.260041	β 2	= 0.273957		
α L	= 2.136372	α L	= -0.252093		

Table 9
Conformational stabilities of all amino acids in case of IGB=5 (gbsa=1)

ALA		ARG		ASN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.578332	C5	= -1.176687	C5	= -0.924754
α R	= -1.953117	α R	= -1.140455	α R	= -1.528486
β	= -1.772009	β	= -1.748402	β	= -1.260758
C7ax	= 2.702182	C7ax	= 3.068761	C7ax	= 3.193414
β 2	= -0.352551	β 2	= 0.471439	β 2	= 0.453960
α L	= 2.702182	α L	= 3.068761	α L	= 3.193414
ASP		CYS		GLN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.567796	C5	= -0.801924	C5	= -2.649815
α R	= -1.702211	α R	= -1.793217	α R	= -2.118703
β	= -1.562574	β	= -1.491271	β	= -1.895284
C7ax	= 1.563777	C7ax	= 3.006718	C7ax	= 2.184091
β 2	= 0.239326	β 2	= 0.073593	β 2	= -0.986392
α L	= 2.460457	α L	= 3.006718	α L	= 2.184091
GLU		GLY		HID	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.699957	C5	= -1.209113	C5	= -1.655782
α R	= -1.286726	α R	= -1.176880	α R	= -2.071047
β	= -1.665851	β	= -1.541241	β	= -1.472617
C7ax	= 2.919925	C7ax	= -0.737316	C7ax	= 2.689221
β 2	= 0.285563	β 2	= 0.070218	β 2	= -0.738371
α L	= 2.919925	α L	= -0.389584	α L	= 2.689221
ILE		LEU		LYS	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.681418	C5	= -1.284538	C5	= -0.748163
α R	= -2.481863	α R	= -1.551345	α R	= -2.510379
β	= -1.492898	β	= -1.600425	β	= -1.647741
C7ax	= 2.440909	C7ax	= 2.995066	C7ax	= 2.488628
β 2	= -0.588451	β 2	= 0.312433	β 2	= -0.642599
α L	= 2.440909	α L	= 2.995066	α L	= 2.488628
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
MET		PHE		PRO	
C5	= -1.408227	C5	= -2.114231	C5	= 2.440909
α R	= -2.191171	α R	= -1.403582	α R	= -1.802127
β	= -1.512758	β	= -2.272500	β	= -2.597194
C7ax	= 2.669236	C7ax	= 2.479384	C7ax	= 2.440909
β 2	= -0.713006	β 2	= 0.028621	β 2	= 2.440909
α L	= 2.669236	α L	= 2.479384	α L	= 2.440909
SER		THR		TRP	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.571846	C5	= -0.978221	C5	= -1.942875
α R	= -2.546454	α R	= -2.192006	α R	= -2.254799
β	= -1.230630	β	= -1.553979	β	= -2.415421
C7ax	= 1.181006	C7ax	= 1.135242	C7ax	= 2.269403
β 2	= -1.050804	β 2	= -0.546261	β 2	= -0.579748
α L	= 0.847382	α L	= 2.708558	α L	= 2.269403
TYR		VAL			
C7eq	= 0.000000	C7eq	= 0.000000		
C5	= -1.427130	C5	= -2.796431		
α R	= -1.995703	α R	= -1.181983		
β	= -1.870138	β	= -1.866181		
C7ax	= 0.927259	C7ax	= 2.228271		
β 2	= -0.288423	β 2	= 1.160085		
α L	= 1.723141	α L	= 2.228271		

量子化学的手法により、各アミノ酸コンフォメーション安定性について精査した。C5 コンフォメーションと α R コンフォメーションの最適化された構造の phi 角、psi 角と、両コンフォメーションのポテンシャルエネルギー差を Table 10 に示す。全てのアミノ酸について、初期構造コンフォメーションをよく保存した最適化構造が得られた。PRO 以外のアミノ酸について、HF/6-31G**によるポテンシャルエネルギー差は GLY、MET、THR を除く全てについてマイナス値となり、C5 コンフォメーションの方が安定であった。一方、GLY、MET、THR は α R コンフォメーションが安定となった。PRO について、ポテンシャルエネルギー差はマイナス値となり β コンフォメーションが安定となった。B3LYP/cc-pVTZ によりポテンシャルエネルギーの再評価を行った。PRO 以外のアミノ酸について、MET、THR を除く全てのアミノ酸について、C5 コンフォメーションの方が安定となった。PRO は β コンフォメーションの方が安定となった。

以上の結果から、 α R コンフォメーションと C5 コンフォメーションの2つのコンフォメーションに注目して、既存力場パラメータ ff03 force field のアミノ酸コンフォメーション安定性と量子化学的手法によるアミノ酸コンフォメーション安定性の比較を行った。IGB=1 の場合の ff03 force field による結果と量子化学計算の結果との比較を Fig 7 に示す。量子化学計算結果と gbsa=0 の結果を比べた時、ASN、ASP、CYS、HID、LEU、LYS、SER、TRP、TYR の 9 個のアミノ酸でそれぞれ矛盾する結果が得られた。いずれのアミノ酸についても、量子化学計算では C5 コンフォメーションが安定なのだが、ff03 force field では α R コンフォメーションが安定となった。gbsa=1 の時、ALA、ASN、ASP、CYS、ILE、LEU、LYS、PHE、SER、TRP、TYR の 11 個のアミノ酸で、量子化学計算結果と反対の α R コンフォメーションが安定という結果が得られた。IGB=2 の場合の結果と量子化学計算の結果との比較を Fig 8 に示す。gbsa=0 の時、ALA、ARG、ASN、ASP、CYS、GLN、HID、ILE、LEU、LYS、SER、TRP、TYR の 13 個のアミノ酸で量子化学計算結果とは逆の α R コンフォメーションが安定という結果になった。gbsa=1 の時、ALA、ARG、ASN、ASP、CYS、GLN、GLU、GLY、HID、ILE、LEU、LYS、PHE、SER、TRP の 15 個のアミノ酸で量子化学計算結果とは逆の α R コンフォメーションが安定という結果になった。IGB=5 の場合の結果と量子化学計算の結果との比較を Fig 9 に示す。gbsa=0 の時、

ALA、ARG、ASN、ASP、CYS、GLY、HID、ILE、LEU、LYS、PHE、SER、TRP、TYR の 14 個のアミノ酸で量子化学計算結果とは逆の α R コンフォメーションが安定という結果になった。gbsa=1 の時、ALA、ASN、ASP、CYS、HID、ILE、LEU、LYS、SER、TRP、TYR の 11 個のアミノ酸で量子化学計算結果とは逆の α R コンフォメーションが安定という結果になった。

量子化学計算において、他とは異なり α R コンフォメーションの方が安定となった THR について、 β コンフォメーションの構造最適化計算を行った。その結果、 $(\phi, \psi)=(-83.1352, 122.9532)$ の主鎖二面角をもつ β コンフォメーション構造を得た。最適化された C5 コンフォメーション、 α R コンフォメーション、 β コンフォメーションについて、B3LYP/6-311++g(2d,2p) でポテンシャルエネルギーの再評価を行った。すると、 β コンフォメーションが最も安定なコンフォメーションであることが分かった。次いで α R コンフォメーション、C5 コンフォメーションの順に安定となった。 β コンフォメーションと α R コンフォメーションのポテンシャルエネルギー差は 0.2947 kcal/mol であった。 β コンフォメーションと C5 コンフォメーションのポテンシャルエネルギー差は 1.1511 kcal/mol であった。

ここで溶媒効果に注目し、連続溶媒モデルではなく実際に水分子を発生させた時のコンフォメーション安定性について検討を行った。ACE-ALANME のジペプチドモデルについてのみ、水分子を実際に発生させた計算モデルを作成し、MD シミュレーションを実行した。各コンフォメーションの安定性の図を Fig. 10 に示す。C7eq コンフォメーションのポテンシャルエネルギーを 0 kcal/mol とした時、C5、 α R、 β 、C7ax、 β 2、 α L コンフォメーションはそれぞれ、-1.6235 kcal/mol、-2.0161 kcal/mol、-1.9119 kcal/mol、2.5973 kcal/mol、-0.7971 kcal/mol、2.5973 kcal/mol であった。C5 コンフォメーションと α R コンフォメーションを比べた時、量子化学計算では C5 コンフォメーションの方が安定であったが、MD シミュレーションの結果では α R コンフォメーションの方が約 0.4 kcal/mol 安定となった。さらに、その他のコンフォメーションと比較しても、 α R コンフォメーションが最も安定となっており、水分子を実際に発生させたモデルでも量子化学計算を再現することはできなかった。

Table 10

The main-chain torsion angles and the energy differences of all amino acids

	C5 conformation Φ value Ψ value	α R conformation Φ value Ψ value	$\Delta E = E(\text{C5}) - E(\alpha \text{R})$ [kcal/mol] HF/6-31G** b3lyp/cc-pVTZ
ALA	-156.4302 149.4884	-77.9649 -26.2124	-0.2549981011 -0.6849447352
ARG	-152.7190 130.6847	-73.6131 -33.6495	-0.7460347163 -1.2718510732
ASN	-160.5020 168.9922	-79.1209 -25.4354	-0.0492036866 -0.1267519999
ASP	-153.6053 142.3152	-71.8354 -35.6883	-1.122632962 * -1.9433005784
CYS	-152.3512 116.5909	-74.1422 -35.8960	-2.9288119360 -3.3552325914
GLN	-151.8713 130.5908	-73.2673 -34.0387	-0.0622590321 -0.4031801950
GLU	-149.8313 132.7859	-71.6697 -37.7150	-0.5042569958 -0.9359104571
GLY	179.4482 -179.3906	-80.5794 -20.3529	0.0610240924 -0.1464131432
HID	-146.8105 122.9691	-71.1386 -40.9116	-3.0410377069* -3.6913896984
ILE	-151.0311 137.9397	-76.5114 -24.8808	-0.9254743383 -1.3785051828
LEU	-153.9757 125.7698	-71.1250 -37.8365	-1.2845456005 -1.5961394560
LYS	-152.5715 129.9444	-75.2446 -33.1864	-0.9178049111 -1.6454906075
MET	-151.8990 130.9927	-80.7571 -24.0864	0.3114753386 0.0074780366
PHE	-155.5316 130.6818	-68.4814 -41.5152	-4.0599708747 -3.7391356793
PRO	-61.0145 147.3009	-66.4779 -30.6649	-0.7320255555 -1.1486156393
SER	-152.1266 129.3507	-73.3657 -43.4088	-1.9588245483 -2.0127929183
THR	-119.9441 165.4372	-94.4409 -6.6334	1.2044552442 0.9267826967
TRP	-153.2736 126.7709	-64.1460 -43.0071	-0.4024874240 * -0.4995795363
TYR	-154.1476 128.9232	-70.2024 -40.4023	-1.4395481006 -1.6607936939
VAL	-151.8550 139.0828	-75.3002 -24.6655	-0.8786766005 -1.5299164432

ΔE suggests the energy differences between C5 conformation and α R conformation calculated by *Gaussian03*.

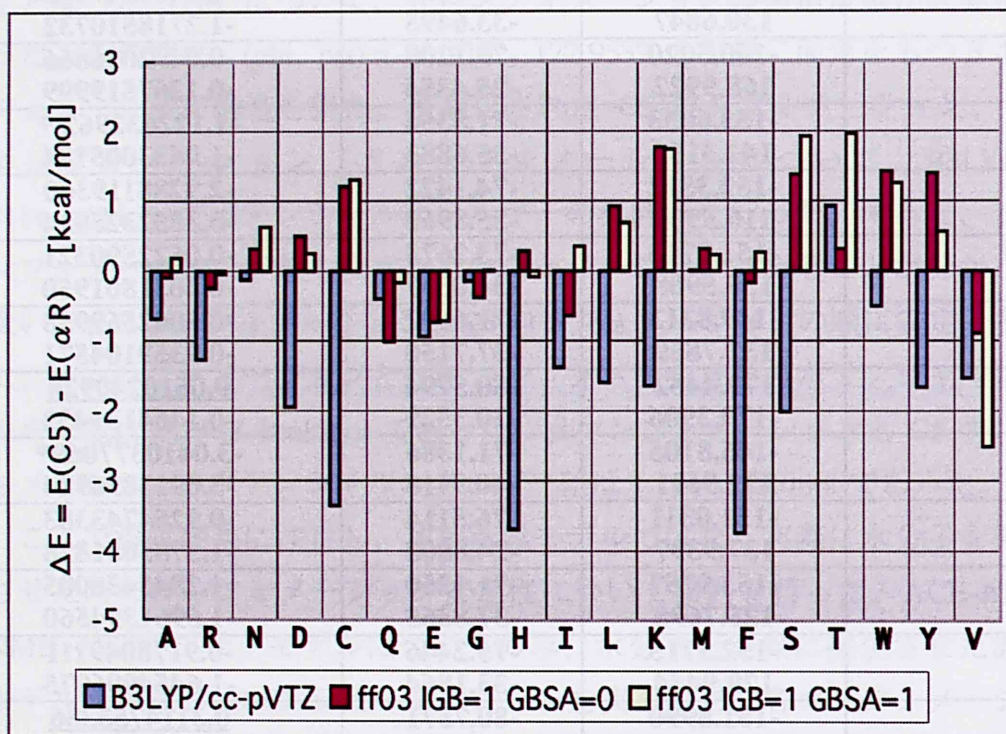


Fig. 7

Comparison between QM and MM energies of the all dipeptide models.

The blue bar suggests the QM energy differences between C5 conformation and αR conformation. The red and yellow bars suggest the MM energy differences. The case of the IGB=1(gbsa=0) is represented by red color, and IGB=1(gbsa=1) is represented by yellow color. All amino acids are expressed by the one-letter code in horizontal axis.

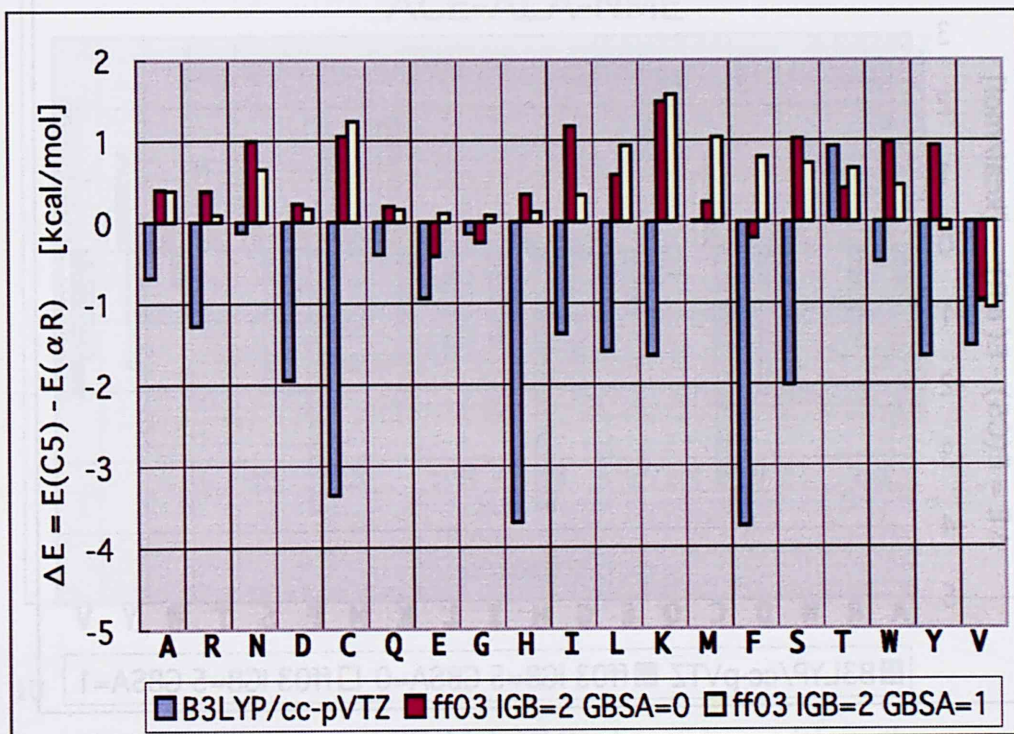


Fig. 8

Comparison between QM and MM energies of the all dipeptide models.

The blue bar suggests the QM energy differences between C5 conformation and αR conformation. The red and yellow bars suggest the MM energy differences. The case of the IGB=2(gbsa=0) is represented by red color, and IGB=2(gbsa=1) is represented by yellow color. All amino acids are expressed by the one-letter code in horizontal axis.

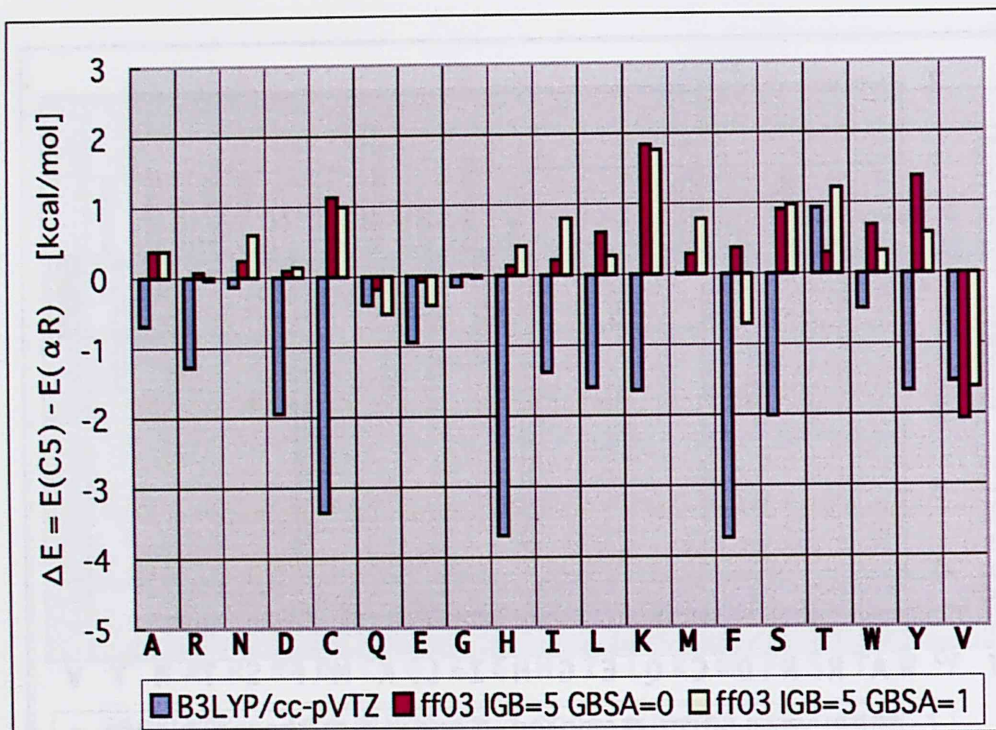


Fig. 9

Comparison between QM and MM energies of the all dipeptide models.

The blue bar suggests the QM energy differences between C5 conformation and αR conformation. The red and yellow bars suggest the MM energy differences. The case of the IGB=5(gbsa=0) is represented by red color, and IGB=5(gbsa=1) is represented by yellow color. All amino acids are expressed by the one-letter code in horizontal axis.

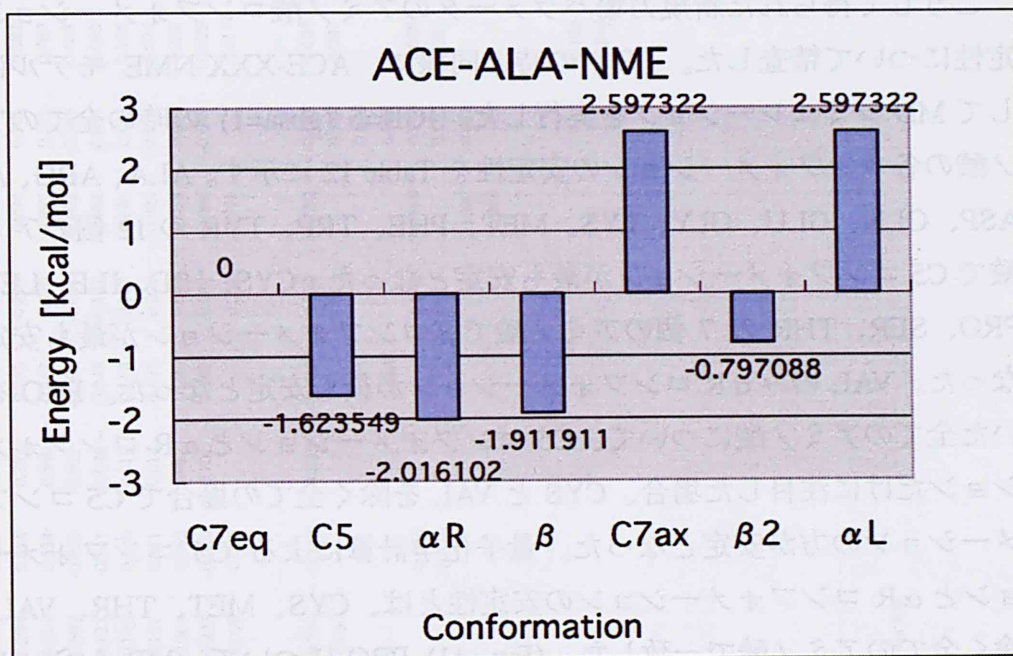


Fig. 10
 Conformational stabilities of ACE-ALA-NME dipeptide model in the presence of explicit water molecules.

(4) 高精度新規力場パラメータの開発

量子化学的手法により求めた各アミノ酸コンフォメーションの最適化構造に対して、静電ポテンシャル計算を実行した。得られた静電ポテンシャルに対して RESP 計算を実行し、水の溶媒効果を含めた各アミノ酸の電荷パラメータを新規に作成した。各アミノ酸の電荷を Table 11 に示す。

こうして得られた新規力場パラメータのアミノ酸コンフォメーション安定性について精査した。(3)の時と同様に、ACE-XXX-NME モデルに対して MD シミュレーションを実行した。IGB=5 (gbsa=1) の時の全てのアミノ酸の各コンフォメーションの安定性を Table 12 に示す。ALA、ARG、ASN、ASP、GLN、GLU、GLY、LYS、MET、PHE、TRP、TYR の 12 個のアミノ酸で C5 コンフォメーションが最も安定となった。CYS、HID、ILE、LEU、PRO、SER、THR の 7 個のアミノ酸で β コンフォメーションが最も安定となった。VAL のみ αR コンフォメーションが最も安定となった。PRO を除いた全てのアミノ酸について、C5 コンフォメーションと αR コンフォメーションだけに注目した場合、CYS と VAL を除く全ての場合で C5 コンフォメーションの方が安定となった。量子化学計算による C5 コンフォメーションと αR コンフォメーションの安定性とは、CYS、MET、THR、VAL を除く全てのアミノ酸で一致した。(Fig. 11) PRO について、MD シミュレーションの結果、 β コンフォメーションの方が αR コンフォメーションよりも安定であった。これは量子化学計算の結果を支持する結果であった。MD シミュレーションにおける THR の C5、 αR 、 β の 3 つのコンフォメーション安定性に注目した場合、量子化学計算同様、 β コンフォメーションが最も安定となった。

Table 11
Atomic partial charges of standard amino acid

	GLY	ALA	SER	CYS	VAL	THR	PRO	ILE	LEU	MET	ASP	ASN	GLU	GLN	HIS (HID)	LYS	ARG	TRP	PHE	TYR
N	-0.254	-0.339	-0.269	-0.296	-0.377	-0.208	-0.086	-0.281	-0.300	-0.325	-0.374	-0.230	-0.320	-0.320	-0.328	-0.269	-0.317	-0.380	-0.354	-0.385
H	0.303	0.298	0.303	0.275	0.310	0.264	0.436	0.286	0.294	0.314	0.297	0.245	0.293	0.294	0.301	0.282	0.305	0.309	0.305	0.321
C	0.515	0.508	0.521	0.536	0.560	0.407	-0.586	0.525	0.514	0.534	0.540	0.467	0.520	0.497	0.515	0.489	0.532	0.517	0.525	0.521
O	-0.612	-0.604	-0.604	-0.570	-0.612	-0.577	-0.586	-0.602	-0.599	-0.605	-0.611	-0.591	-0.608	-0.602	-0.585	-0.603	-0.608	-0.602	-0.602	-0.599
C _α	-0.252	-0.023	-0.160	-0.112	-0.116	-0.050	-0.070	-0.194	-0.123	-0.141	-0.046	-0.043	-0.141	-0.086	-0.057	-0.085	-0.141	-0.014	-0.039	-0.041
H _α	0.150	0.124	0.176	0.175	0.133	0.140	0.121	0.145	0.147	0.148	0.119	0.118	0.167	0.146	0.143	0.136	0.158	0.132	0.132	0.146
C _β	-0.063	0.107	0.053	-0.094	0.250	0.233	-0.023	0.157	-0.127	0.011	-0.014	-0.125	0.001	0.022	-0.118	-0.024	-0.008	-0.032	-0.012	-0.043
H _β	0.033	0.033	0.053	0.124	0.016	0.044	0.033	0.033	0.064	0.062	0.017	0.084	0.044	0.033	0.095	0.028	0.040	0.066	0.052	0.058
C _γ , O _γ , S _γ			-0.620	-0.357	-0.251	-0.681	-0.009	-0.029	0.223	-0.080	0.701	0.592	-0.058	-0.080	-0.049	0.029	0.009	-0.205	-0.005	-0.017
H _γ			0.441	0.195	0.056	0.463	0.032	0.025	0.008	0.096			0.001	0.041	0.041	0.017	0.043			
C _δ						-0.180		-0.217												
H _δ (α,β)						0.049		0.052					0.739	0.682	-0.165	-0.024	-0.022	-0.083	-0.103	-0.161
C _ε , O _ε , N _ε							-0.024	-0.095	-0.234	-0.332	-0.823	-0.610			0.318	0.037	0.078	0.169	0.112	0.142
H _ε							0.056	0.022	0.051			-0.797		0.099	0.099		0.078	0.078		
C _ζ , N _ζ												0.403		0.089						
H _ζ													-0.842	-0.649	0.140	0.028	-0.334	-0.341	-0.180	-0.224
C _η															0.123	0.081	0.317	0.364	0.138	0.153
H _η															-0.618		0.187	0.187		
C _θ																		-0.210		
H _θ																		0.151		
C _ι																			-0.265	0.284
H _ι																			0.157	0.127
C _κ , O _κ , N _κ (α,β)																				
H _κ , H _κ																				
C _λ																				
H _λ																				

^aH_δ(α,β) for Gly.
^bH_δ(α,β) for Ala and H_δ for Thr, Ile, and Val, H_δ(α,β) for all others.
^cC_γ for Glu, Asp, Lys, Pro, Met, Asn, and Gln; C_γ(α,β) for Val; O_γ for Ser, O_γ for Thr, S_γ for Cys.
^dH_γ for Thr, H_γ(α,β) for Gln, Arg, H_γ(α,β) for Ile, H_γ(α,β,γ,δ) for Val.
^eC_δ for Ile; Trp, C_δ(α,β) for Leu, Phe, Tyr; S_δ for Met; O_δ for Asn; O_δ(α,β) for Asp; C_δ for Pro, Glu, Gln, Lys, Arg; N_δ for His.
^fH_δ(α,β,γ) for Ile, H_δ(α,β) for Arg, Lys, Pro; H_δ for Trp; H_δ(α,β) for Phe, Tyr; H_δ(α,β,γ,δ) for Leu.
^gC_δ for His, Trp, N_δ for Asn.
^hH_δ(α,β) for Asn.
ⁱC_γ for Met, Lys; C_γ for His; C_γ(α,β) for Tyr, Phe, O_γ for Gln; O_γ(α,β) for Glu, N_γ for Arg; N_γ for Trp.
^jH_δ(α,β) for Lys; H_δ(α,β,γ) for Met, H_δ for Arg; H_δ for His, Trp; H_δ(α,β) for Phe, Tyr.
^kC_δ for Trp, Phe, Tyr; N_δ for Gln, His.
^lH_δ(α,β) for Gln.
^mC_δ for Trp; N_δ for Lys.
ⁿH_δ for Trp; H_δ(α,β,γ) for Lys.
^oC_δ for Trp; OH for Tyr; N_δ(α,β) for Arg.
^pH_δ for Trp; H_δ for Tyr; H_δ(α,β,γ) for Arg.

Table 12
Conformational stabilities of all amino acids in case of IGB=5 (gb sa=1) using the original developed force field.

ALA		ARG		ASN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -2.057688	C5	= -1.838395	C5	= -1.843627
α R	= -1.536362	α R	= -1.359440	α R	= -0.438506
β	= -1.901574	β	= -1.645616	β	= -1.802846
C7ax	= 2.627136	C7ax	= 2.844120	C7ax	= 1.458396
β 2	= -0.118310	β 2	= -0.122839	β 2	= 1.314623
α L	= 2.627136	α L	= 2.844120	α L	= 2.887939
ASP		CYS		GLN	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -2.182278	C5	= -1.408158	C5	= -1.934958
α R	= -0.368447	α R	= -1.608253	α R	= -1.070914
β	= -2.129138	β	= -1.716428	β	= -1.797061
C7ax	= 2.581836	C7ax	= 2.897255	C7ax	= 0.927547
β 2	= 1.271925	β 2	= 0.004639	β 2	= 0.247219
α L	= 2.581836	α L	= 2.897255	α L	= 2.796822
GLU		GLY		HID	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.935347	C5	= -1.554905	C5	= -1.805347
α R	= -1.444667	α R	= -0.211456	α R	= -0.893466
β	= -1.922344	β	= -1.552727	β	= -1.811827
C7ax	= 2.695736	C7ax	= -0.444842	C7ax	= 2.873688
β 2	= 0.113901	β 2	= 0.764831	β 2	= 0.553517
α L	= 2.695736	α L	= 0.685224	α L	= 2.873688
ILE		LEU		LYS	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.766330	C5	= -1.396580	C5	= -2.283782
α R	= -1.212094	α R	= -1.073454	α R	= -1.055456
β	= -1.978175	β	= -1.689895	β	= -2.020639
C7ax	= 2.774533	C7ax	= 3.058271	C7ax	= 2.541265
β 2	= -0.016795	β 2	= 0.905565	β 2	= 0.755311
α L	= 2.774533	α L	= 3.058271	α L	= 2.541265
MET		PHE		PRO	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -2.132463	C5	= -2.111109	C5	= 2.242300
α R	= -1.079762	α R	= -1.235907	α R	= -1.888843
β	= -1.938849	β	= -2.041601	β	= -2.825139
C7ax	= 2.648557	C7ax	= 2.604916	C7ax	= 2.242300
β 2	= 0.406258	β 2	= 0.362616	β 2	= 2.242300
α L	= 2.648557	α L	= 2.604916	α L	= 2.242300
SER		THR		TRP	
C7eq	= 0.000000	C7eq	= 0.000000	C7eq	= 0.000000
C5	= -1.889002	C5	= -1.560939	C5	= -2.234765
α R	= -1.724323	α R	= -1.150678	α R	= -2.096658
β	= -1.894056	β	= -1.724274	β	= -2.165774
C7ax	= 2.641502	C7ax	= 2.971057	C7ax	= 2.332215
β 2	= 0.021681	β 2	= 0.520294	β 2	= -0.654955
α L	= 2.641502	α L	= 2.971057	α L	= 2.332215
TYR		VAL			
C7eq	= 0.000000	C7eq	= 0.000000		
C5	= -2.378513	C5	= -1.013276		
α R	= -1.777457	α R	= -1.754008		
β	= -2.115708	β	= -1.301207		
C7ax	= 2.366953	C7ax	= 3.029360		
β 2	= -0.487187	β 2	= -0.440375		
α L	= 2.366953	α L	= 3.029360		

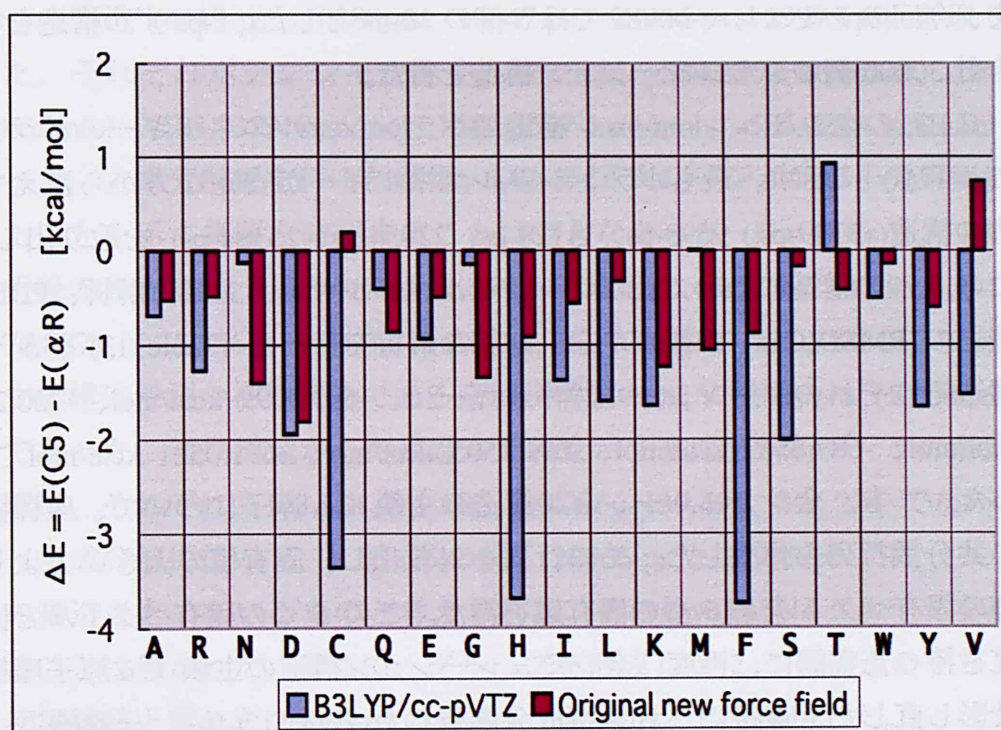


Fig. 11 Comparison between the QM and MM energies of all dipeptide models using the original developed force field.

The blue bar suggests the QM energy differences between C5 conformation and αR conformation. The red bar suggests the MM energy differences of original developed force field in the case of the IGB=5(gbsa=1). All amino acids are expressed by the one-letter code in horizontal axis.

(5) 高精度新規力場パラメータを用いたタンパク質立体構造予測

IGB=5 (gbsa=1) の溶媒効果を用いて、1LE3、1L2Y、1VII、1PGB の4つのタンパク質について立体構造予測を行った。加速的にシミュレーションを行う目的で 375K という比較的高い温度を保ったまま MD シミュレーションを実行した。また、(1) の場合 6-ns であったシミュレーション時間を、今回のシミュレーションでは5倍の 30-ns とした。1-ps ずつ構造を得、合計 30000 個のシミュレーション構造を得た。

1LE3 の全シミュレーション構造のポテンシャルエネルギーのグラフを Fig. 12(a) に示す。最もポテンシャルエネルギーが安定な方から数えて 5 つの構造 (3149-ps、9698-ps、11128-ps、22954-ps、25384-ps) を取り出した。この 5 つの構造について、ポテンシャルエネルギー、主成分解析、PDB 構造との RMSD(C α) の観点から、詳細な解析を行った (Table 13) 。5 つの構造それぞれのポテンシャルエネルギーは、-690.1099 kcal/mol、-700.2021 kcal/mol、-687.6569 kcal/mol、-691.1505 kcal/mol、-691.0521 kcal/mol であった。一方、全シミュレーション構造の主鎖 C α 原子について、座標変位の共分散行列を作成した。この行列を対角化し、固有値の大きい方より 5 つの固有ベクトルと 5 つの固有値を得た。この 5 つの固有ベクトルと固有値を持った空間に、再度、全シミュレーション構造の主鎖 C α 原子座標を投影し直した。最後にこれを主成分解析した (Fig. 12(c)) 。一般的に MD シミュレーションにより構造が平衡に達した状態とは、この主成分解析において、ある領域で集団を形成し、その状態からあまり動かなくなる状態を言う。しかし、そのような状態を確認することはできなかった。主成分解析後の X、Y 軸成分として、5 つの構造に対してそれぞれ、(X , Y) = (-160.344 , -1.763)、(-57.297 , -21.278)、(6.431 , 29.334)、(142.092 , -102.974)、(49.603 , -103.619) を得た。立体構造予測の初期構造である直鎖状構造の X、Y 軸成分は (X , Y) = (-89.704 , 23.147) であった。この初期構造からの各成分の距離はそれぞれ、74.903、54.989、96.334、263.886、188.351 であった。

主成分解析後の X、Y 軸成分から、5 つの構造は同グループにいないことが示された。そこで、ポテンシャルエネルギーが最も低い方から数えて 5 つの構造のうち、最も初期構造と離れた 22954-ps の構造を予測構造として採用した。

ここで、PDB 構造と全シミュレーション構造の、主鎖 C α 原子における RMSD(C α)を Fig. 12(b) に示す。5つの構造に対する RMSD(C α) はそれぞれ、6.260 Å、4.852 Å、8.190 Å、3.869 Å、7.271 Å であった。

1L2Y の全シミュレーション構造のポテンシャルエネルギーのグラフを Fig. 13(a) に示す。最もポテンシャルエネルギーが安定な方から数えて 5つの構造 (4119-ps、10288-ps、10647-ps、28966-ps、29995-ps) を取り出した。それぞれポテンシャルエネルギーは、-584.2869 kcal/mol、-589.4253 kcal/mol、-583.7379 kcal/mol、-582.6607 kcal/mol、-585.0238 kcal/mol であった。主成分解析を行った結果、MD シミュレーションが十分平衡に達したことを示す集団を確認することはできなかった (Fig. 13(c))。主成分解析後の X、Y 軸成分は 5つの構造に対しそれぞれ、(-152.997, 17.993)、(-41.197, 76.887)、(-75.456, -19.949)、(140.106, 25.170)、(127.465, 29.301) であった。立体構造予測の初期構造である直鎖状構造の X、Y 軸成分は (5.104, 60.290) であった。初期構造からの構造のずれを示す主成分解析データの各成分距離は、それぞれ、163.661、49.186、90.096、139.495、126.224 であった。

主成分解析後の X、Y 軸成分から、28966-ps と 29995-ps が同一グループの構造をとっているものと考えられた。その他の構造はグループを形成することは無かった。そこで、グループを形成した 2つの構造のうち、最も初期構造と離れた 28966-ps の構造を予測構造として採用した。

ここで、1L2Y の RMSD(C α) を Fig. 13(b) に示す。5つの構造に対する RMSD(C α) はそれぞれ、3.689 Å、4.464 Å、3.143 Å、2.646 Å、2.319 Å であった。

1VII の全シミュレーション構造のポテンシャルエネルギーのグラフを Fig. 14(a) に示す。最もポテンシャルエネルギーが安定な方から数えて 5つの構造 (24067-ps、24341-ps、25207-ps、25388-ps、26541-ps) を取り出した。それぞれポテンシャルエネルギーは、-1193.6375 kcal/mol、-1195.6899 kcal/mol、-1222.8089 kcal/mol、1191.1037 kcal/mol、-1201.0559 kcal/mol であった。主成分解析を行った結果、MD シミュレーションが十分平衡に達したことを示す集団を確認することはできなかった (Fig. 14(c))。主成分解析後の X、Y 軸成分は 5つの構造に対しそれぞれ、(102.521, -25.685)、(88.729, -3.683)、(87.398, 71.529)、(91.530, 89.550)、(87.997, 87.025) であった。立体構造予測の初期構造である直鎖状構造の X、Y 軸成分は (-347.833,

22.757) であった。初期構造からの構造のずれを示す主成分解析データの各成分距離は、それぞれ、452.952、437.362、437.955、444.411、440.543 であった。

主成分解析後の X、Y 軸成分から、25207-ps、25388-ps、26541-ps の 3 つがグループを形成していることが考えられた。他の 2 つの構造はグループを形成することは無かった。そこで、グループを形成した 3 つの構造のうち、最も初期構造と離れた 25388-ps の構造を予測構造として採用した。

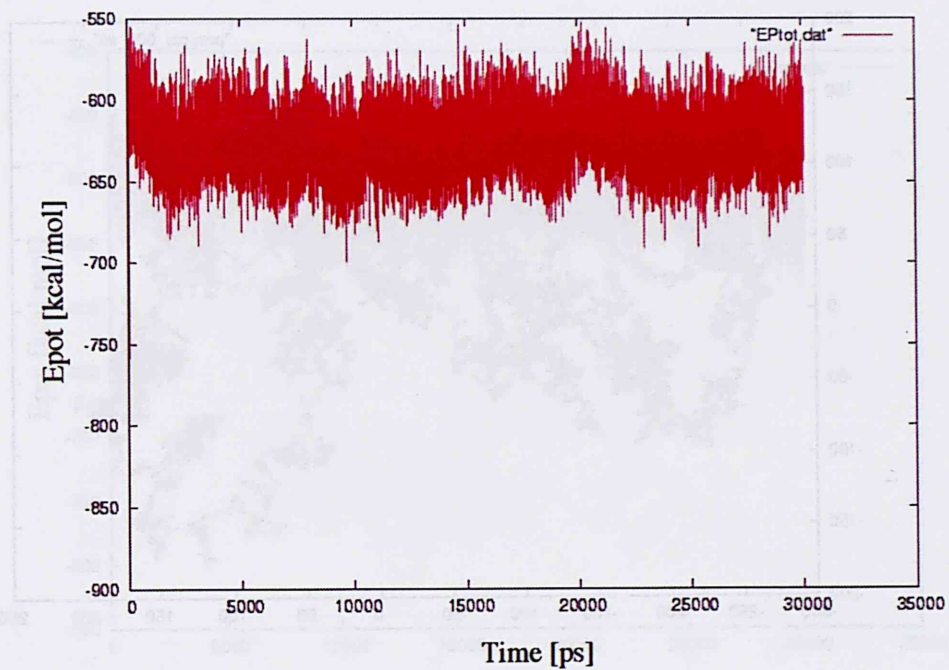
ここで、1VII の RMSD(C α) を Fig. 14(b) に示す。5 つの構造に対する RMSD(C α) はそれぞれ、7.183 Å、6.958 Å、6.537 Å、5.809 Å、7.674 Å であった。

1PGB の全シミュレーション構造のポテンシャルエネルギーのグラフを Fig. 15(a) に示す。最もポテンシャルエネルギーが安定な方から数えて 5 つの構造 (15273-ps、19054-ps、29564-ps、29855-ps、30006-ps) を取り出した。それぞれポテンシャルエネルギーは、-2021.5143 kcal/mol、-2012.2551 kcal/mol、-2015.4301 kcal/mol、-2022.9451 kcal/mol、2016.9702 kcal/mol であった。主成分解析を行った結果、MD シミュレーションが十分平衡に達したことを示す集団を確認することはできなかった (Fig. 15(c))。主成分解析後の X、Y 軸成分は 5 つの構造に対しそれぞれ、(52.492 , 226.854)、(115.482 , 108.432)、(274.627 , -183.678)、(273.976 , -205.036)、(319.212 , -209.225) であった。立体構造予測の初期構造である直鎖状構造の X、Y 軸成分は (-509.793 , -52.367) であった。初期構造からの構造のずれを示す主成分解析データの各成分距離は、それぞれ、627.797、645.620、795.335、798.500、843.714 であった。

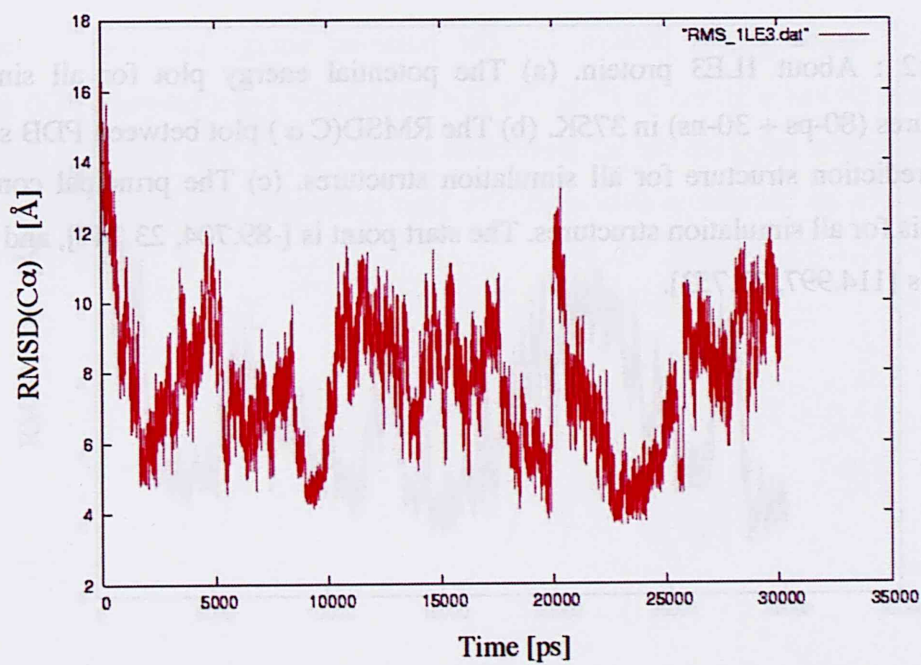
主成分解析後の X、Y 軸成分から、29564-ps と 29855-ps の 2 つがグループを形成していることが考えられた。そこで、グループを形成した 2 つの構造のうち、最も初期構造と離れた 29855-ps の構造を予測構造として採用した。

ここで、1PGB の RMSD(C α) を Fig. 15(b) に示す。5 つの構造に対する RMSD(C α) はそれぞれ、17.787 Å、14.702 Å、13.054 Å、15.976 Å、16.308 Å であった。

(a)



(b)



(c)

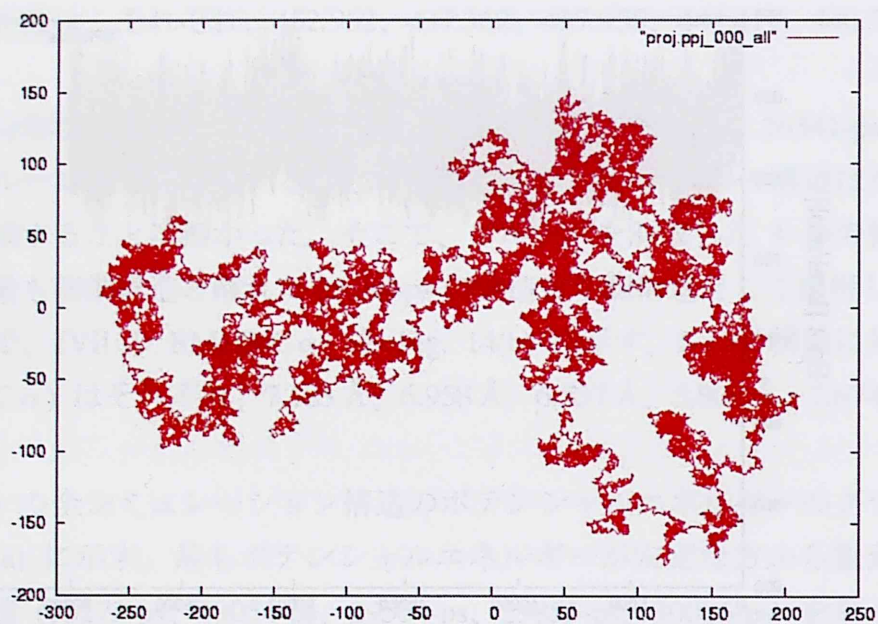
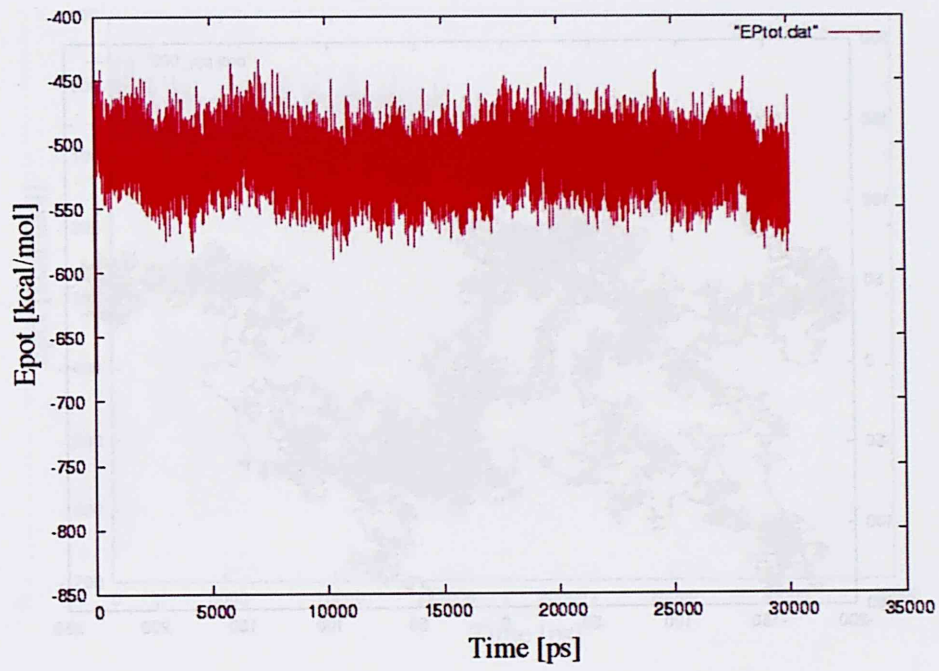
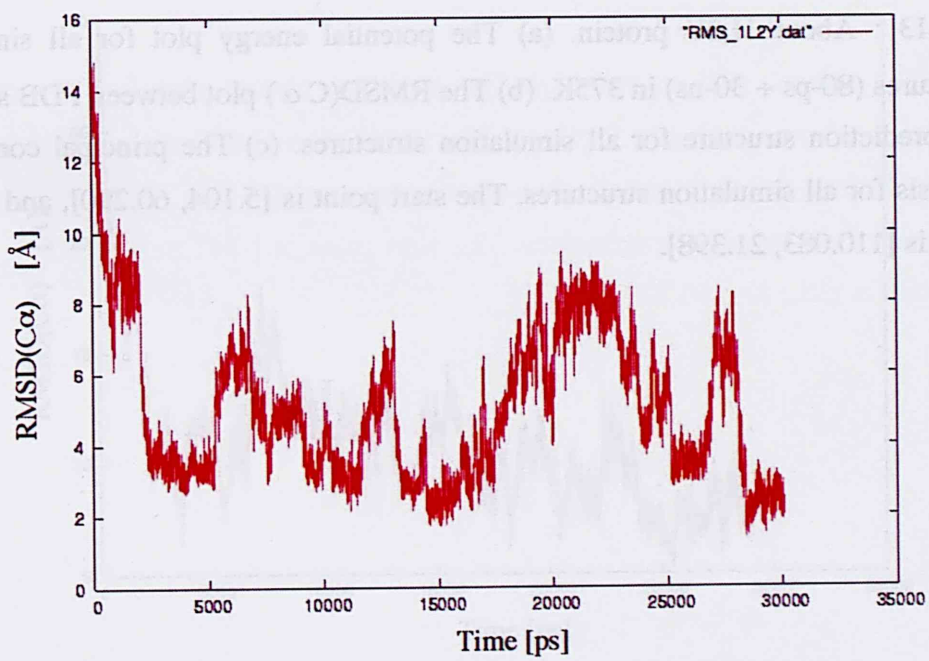


Fig. 12 : About 1LE3 protein. (a) The potential energy plot for all simulation structures (80-ps + 30-ns) in 375K. (b) The RMSD($C\alpha$) plot between PDB structure and prediction structure for all simulation structures. (c) The principal component analysis for all simulation structures. The start point is [-89.704, 23.147], and the end point is [114.997, 27.753].

(a)



(b)



(c)

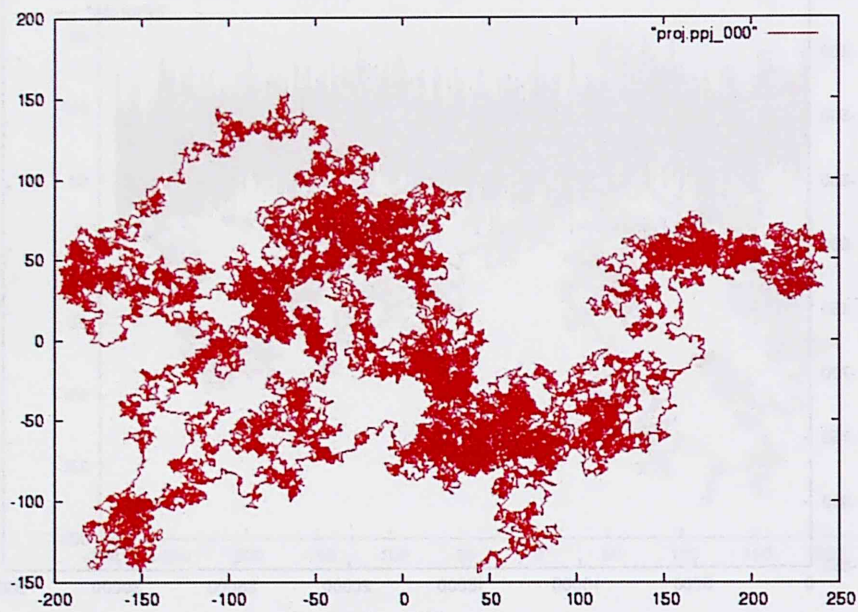
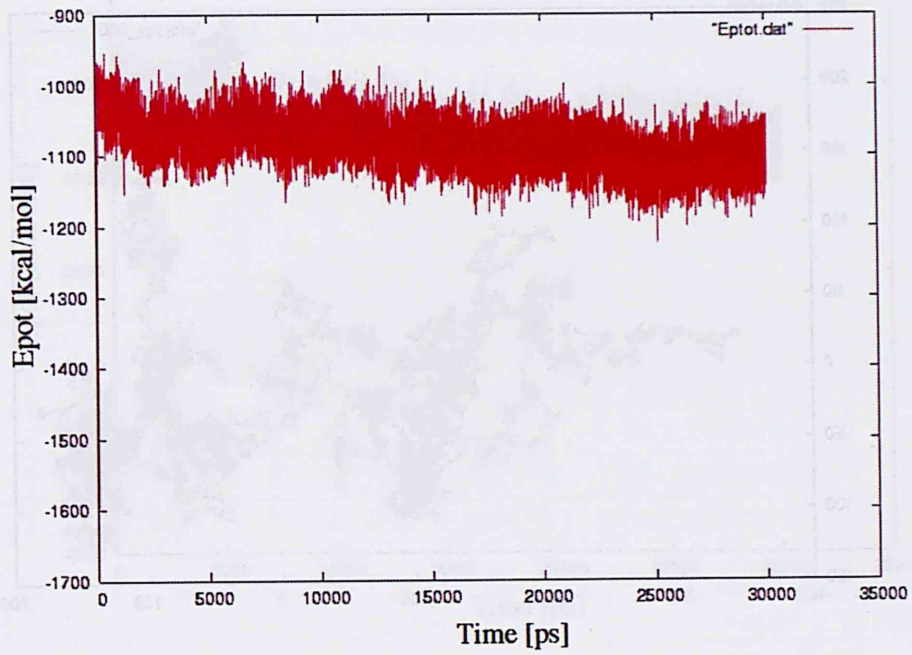
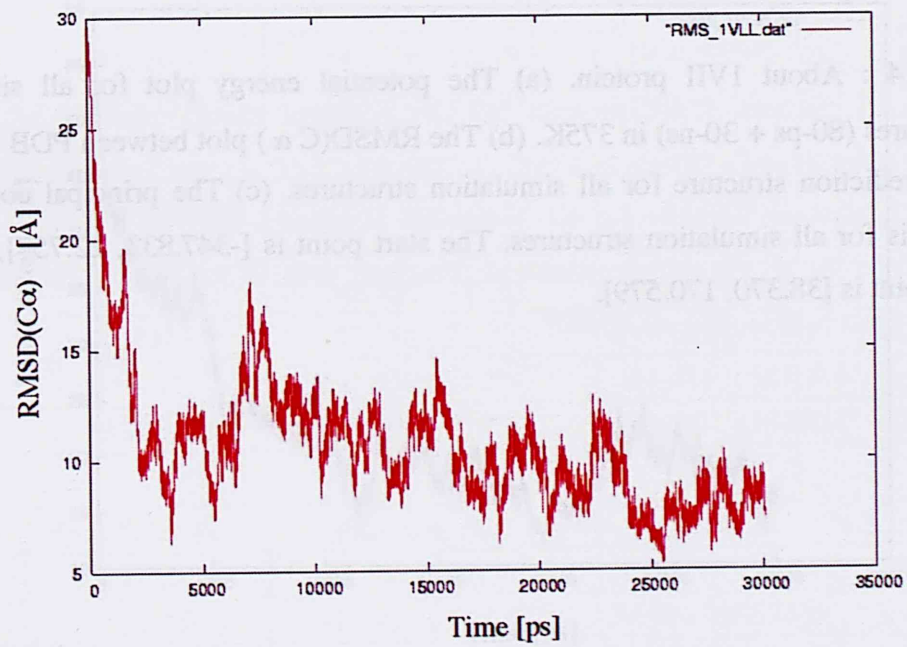


Fig. 13 : About 1L2Y protein. (a) The potential energy plot for all simulation structures (80-ps + 30-ns) in 375K. (b) The RMSD($C\alpha$) plot between PDB structure and prediction structure for all simulation structures. (c) The principal component analysis for all simulation structures. The start point is [5.104, 60.290], and the end point is [110.083, 21.398].

(a)



(b)



(c)

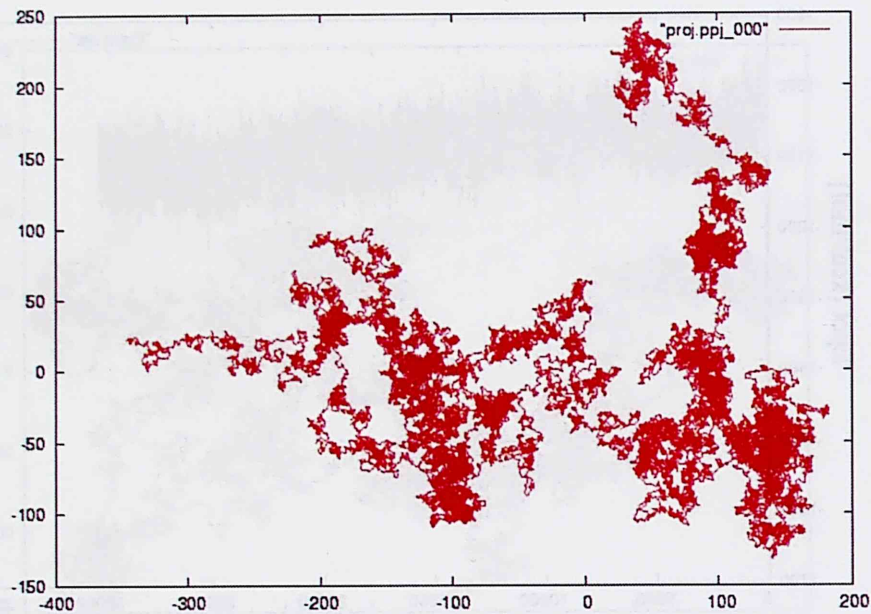
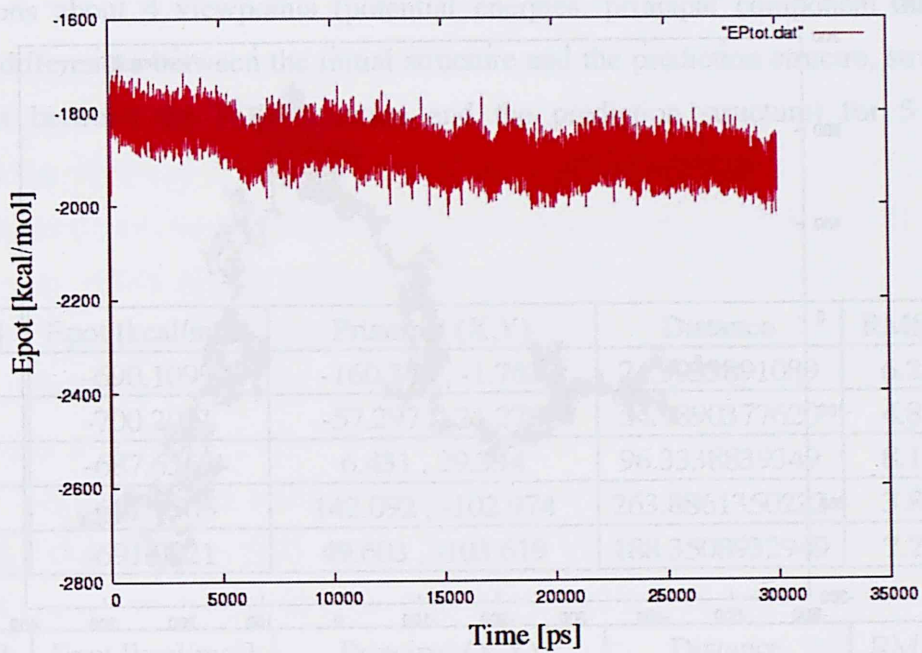
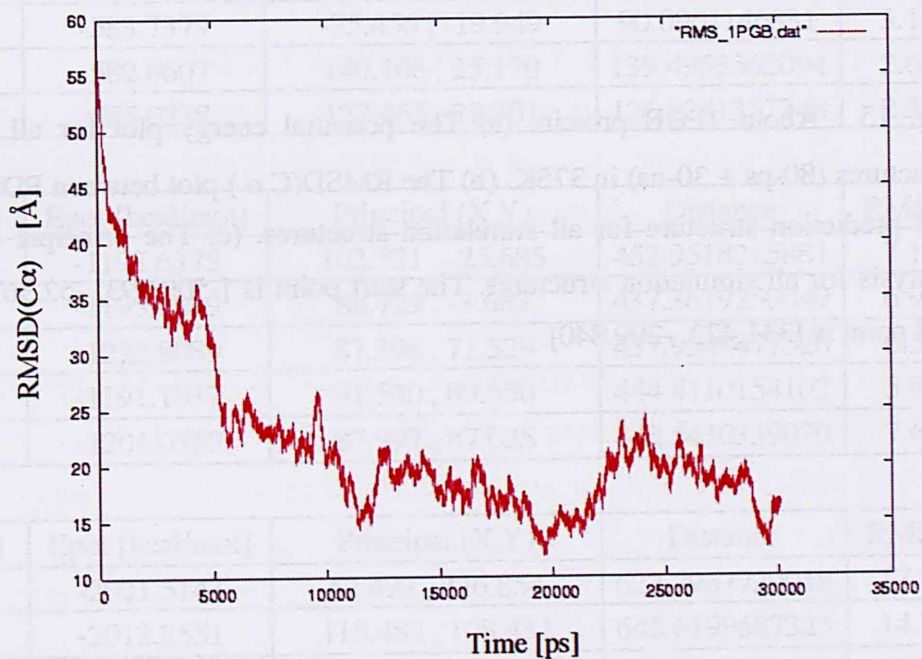


Fig. 14 : About 1VII protein. (a) The potential energy plot for all simulation structures (80-ps + 30-ns) in 375K. (b) The RMSD($C\alpha$) plot between PDB structure and prediction structure for all simulation structures. (c) The principal component analysis for all simulation structures. The start point is $[-347.833, 22.757]$, and the end point is $[38.370, 170.579]$.

(a)



(b)



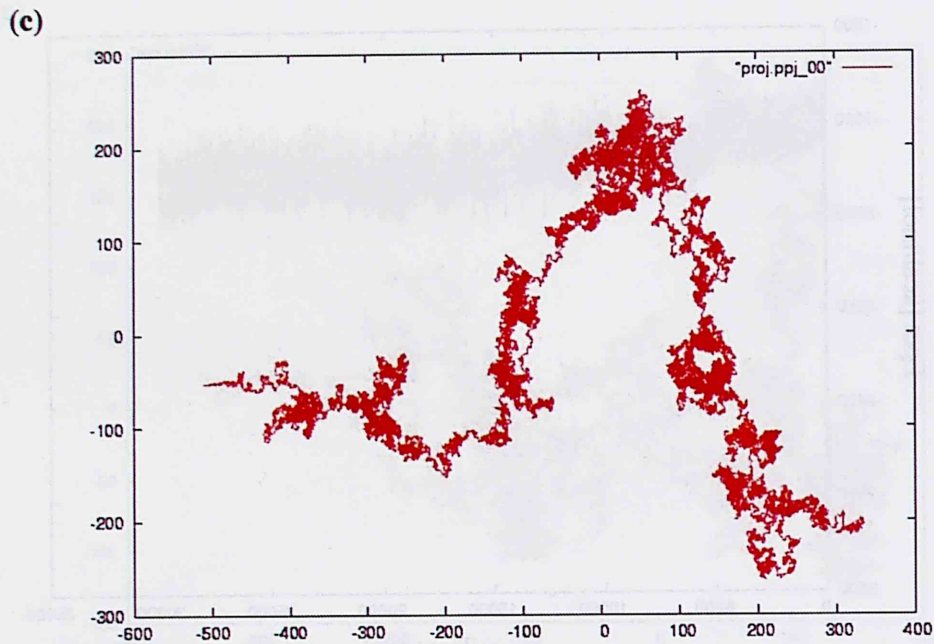


Fig. 15 : About 1PGB protein. (a) The potential energy plot for all simulation structures (80-ps + 30-ns) in 375K. (b) The RMSD($C\alpha$) plot between PDB structure and prediction structure for all simulation structures. (c) The principal component analysis for all simulation structures. The start point is $[-509.793, -52.367]$, and the end point is $[331.422, -209.840]$.

Table 13

Comparisons about 4 viewpoints (potential energies, principal component analysis, structural differences between the initial structure and the prediction structure, structural differences between the PDB structure and the prediction structure) for 5 stable structures.

1LE3

Time [ps]	Epot [kcal/mol]	Principal (X,Y)	Distance	RMSD [Å]
3149	-690.1099	-160.344 , -1.763	74.9033891089	6.25961
9698	-700.2021	-57.297 , -21.278	54.9890377620	4.85249
11128	-687.6569	6.431 , 29.334	96.3338839349	8.19030
22954	-691.1505	142.092 , -102.974	263.8861350222	3.86907
25384	-691.0521	49.603 , -103.619	188.3508932949	7.27131

1L2Y

Time [ps]	Epot [kcal/mol]	Principal (X,Y)	Distance	RMSD [Å]
4119	-584.2869	-152.997 , 17.993	163.6611206426	3.68851
10288	-589.4253	-41.197 , 76.887	49.1858008982	4.46437
10647	-583.7379	-75.456 , -19.949	90.0961146831	3.14319
28966	-582.6607	140.106 , 25.170	139.4953562094	2.64643
29995	-585.0238	127.465 , 29.301	126.2241357348	2.31854

1VII

Time [ps]	Epot [kcal/mol]	Principal (X,Y)	Distance	RMSD [Å]
24067	-1193.6375	102.521 , -25.685	452.9518215881	7.18271
24341	-1195.6899	88.729 , -3.683	437.3619250049	6.95795
25207	-1222.8089	87.398 , 71.529	437.9549477526	6.53695
25388	-1191.1037	91.530 , 89.550	444.4110154102	5.80869
26541	-1201.0559	87.997 , 87.025	440.5430339070	7.67407

1PGB

Time [ps]	Epot [kcal/mol]	Principal (X,Y)	Distance	RMSD [Å]
15273	-2021.5143	52.492 , 226.854	627.7967729018	17.78742
19054	-2012.2551	115.482 , 108.432	645.6199687323	14.70230
29564	-2015.4301	274.627 , -183.678	795.3347189209	13.05362
29855	-2022.9451	273.976 , -205.036	798.4996361489	15.97558
30006	-2016.9702	319.212 , -209.225	843.7142420209	16.30779

予測構造と PDB 構造を、主鎖 C α 原子について最も距離変位が小さくなるように重ねあわせた図を Fig. 16 に示す。

1LE3 について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 3.87 Å であった。完全な β -sheet 構造を予測することはできなかったが、非常に近い構造の予測に成功した。また予測構造について、4 つの TRP の側鎖の芳香族環は PDB のように整然と並ぶことは無かったが、(1) の ff03 force field を用いた場合と比べると、4 つの TRP はよく集まっていた (Fig. 17 (a))。turn 構造領域において、予測構造は PDB 構造と同じ 7ASP、8ALA、9THR、10LYS で turn 構造を形成し、正しい位置での turn 構造予測に成功した。PDB 構造と予測構造のポテンシャルエネルギー差について、 Δ EGB は 3.41 kcal/mol であり PDB 構造の方が、 Δ EPB は -16.64 kcal/mol であり予測構造の方が安定であった。1L2Y について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 2.32 Å であった。2 つの helix 構造と 1 つの strand 構造を予測し、全体的に PDB 構造とよく似た構造の予測に成功した。PDB 構造の TRP 側鎖は PRO12 と疎水的相互作用を形成するが、予測構造の TRP 側鎖は PRO17 と疎水的相互作用を形成した (Fig. 17 (b))。このことにより、予測構造の TRP 側鎖の向きは PDB 構造の向きとは異なった。PDB 構造と予測構造のポテンシャルエネルギー差について、 Δ EGB は -39.73 kcal/mol、 Δ EPB は -36.18 kcal/mol で、いずれも予測構造の方が安定であった。1VII について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 5.81 Å であった。予測構造は PDB 構造と同様の領域に 3 本の helix 構造を形成し、二次構造的には高精度で構造予測に成功した。N 末端側から数えて 1 本目の helix 構造と 2 本目の helix 構造の間の π - π 相互作用について、PDB 構造では PHE7-PHE18、PHE11-PHE18 の 2 つの π - π 相互作用が存在するが、予測構造では PHE11-PHE18 の 1 つの π - π 相互作用のみであった (Fig. 17 (c))。予測構造の PHE7-PHE11 の π - π 相互作用は 1 本目の helix 構造どうしの π - π 相互作用であった。この π - π 相互作用の違いにより、1 本目の helix 構造の向きが PDB 構造とは異なったが、全体としては良く似た構造を予測できた。PDB 構造と予測構造のポテンシャルエネルギー差について、 Δ EGB は -59.58 kcal/mol、 Δ EPB は -37.28 kcal/mol で、いずれも予測構造の方が安定であった。1PGB について、PDB 構造と予測構造の主鎖 C α 原子の RMSD は 15.98 Å であった。ALA23~ASN37 の helix 領域における、PDB 構造と予

測構造の主鎖 C α 原子の RMSD は 1.891 Å であった (Fig 17 (d))。helix 構造領域における構造予測には高精度で予測に成功した。 β -sheet 構造領域について、構造予測には完全に失敗した。また、構造全体としても立体構造折り畳みが不十分であり、二次構造-二次構造間の距離が比較的大きかった。PDB 構造と予測構造のポテンシャルエネルギー差について、 Δ EGB は 127.54 kcal/mol、 Δ EPB は 72.64 kcal/mol で、いずれも予測構造の方が安定であった。PDB 構造のポテンシャルエネルギー計算において、全てのタンパク質で Native 構造は PDB 構造を良く保存し MD シミュレーションが実行された。

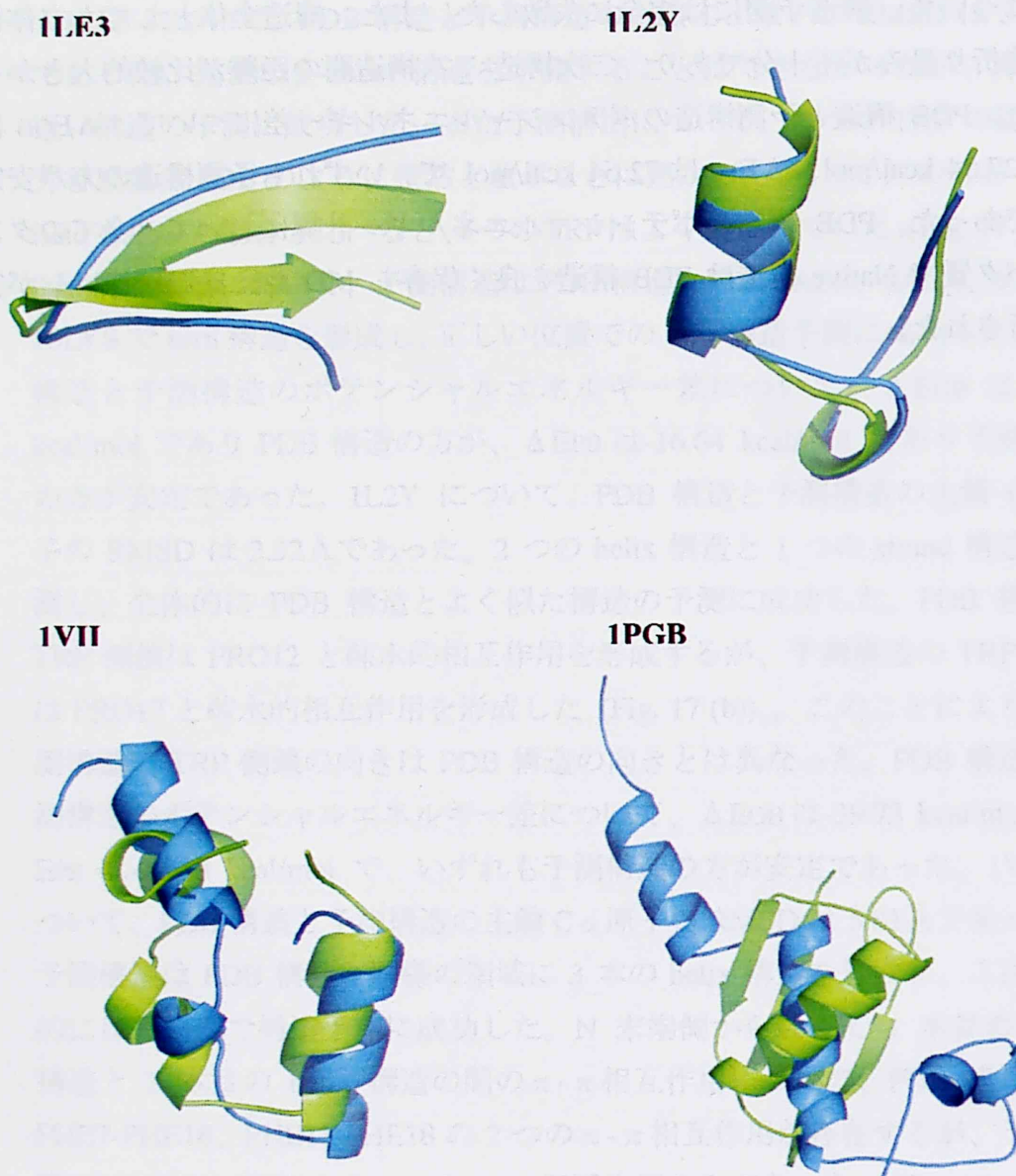


Fig. 16: Prediction structures (cyan) and PDB structures (green) were represented as a *cartoon* expression in case of the IGB=5, gbsa=1 by using the *original* force field.

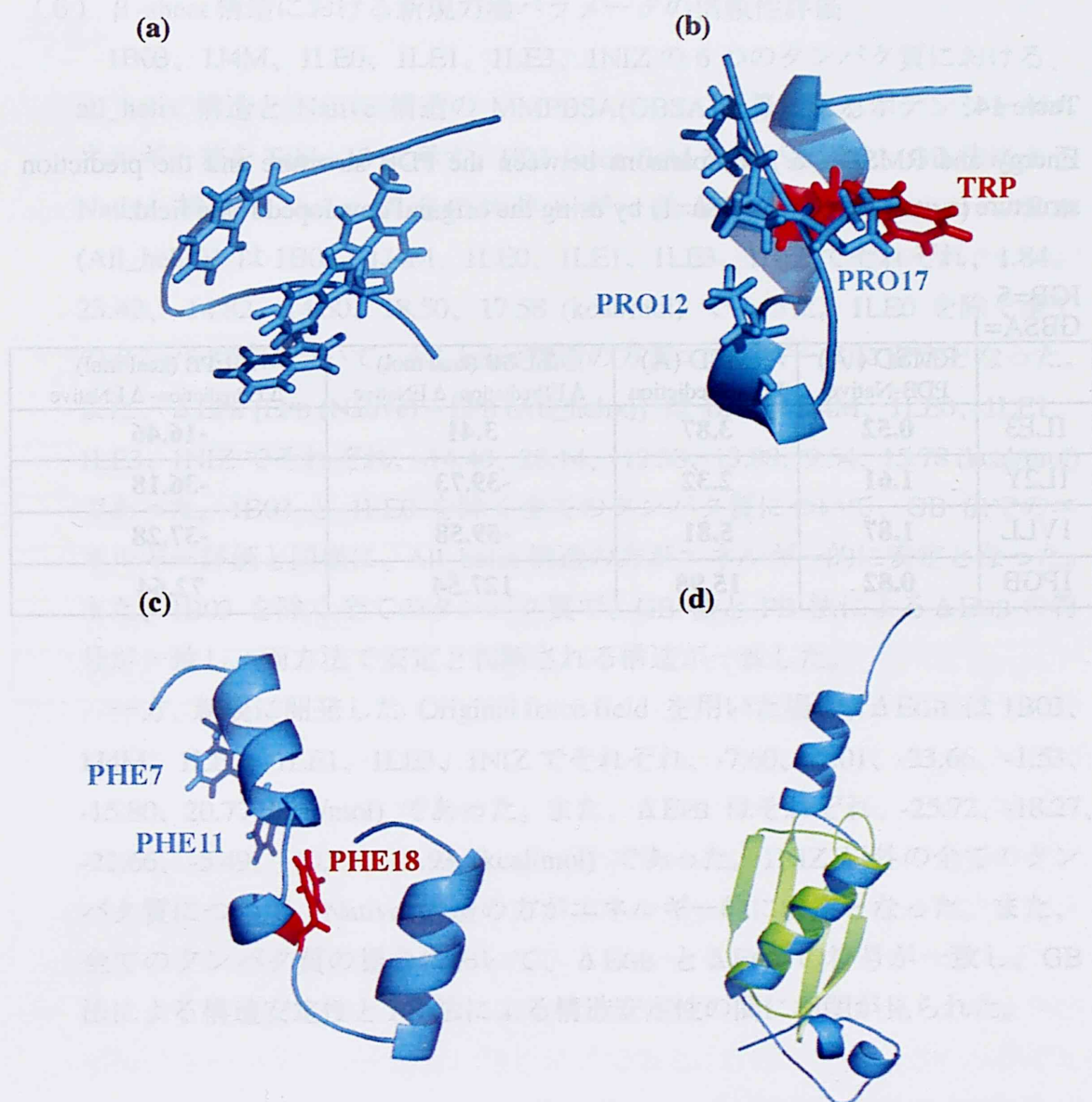


Fig. 17: (a) 4 TRP residues can not be formed π - π interactions. (b) TRP forms the hydrophobic interaction with PRO17 instead of PRO12. (c) PHE18 makes π - π interaction with PHE11, but cannot interact with PHE7. (d) The helix region of PDB structure and prediction structure is conformed to be highly accurate.

Table 14

Energy and RMSD($C\alpha$) comparisons between the PDB structure and the prediction structure in case of IGB=5 (gbsa=1) by using the original developed force field.

IGB=5
GBSA=1

	RMSD (Å) PDB-Native	RMSD (Å) PDB-Prediction	Δ EGB (kcal/mol) Δ EPrediction- Δ ENative	Δ EPB (kcal/mol) Δ EPrediction- Δ ENative
1LE3	0.52	3.87	3.41	-16.46
1L2Y	1.61	2.32	-39.73	-36.18
1VLL	1.87	5.81	-59.58	-37.28
1PGB	0.82	15.98	127.54	72.64

(6) β -sheet 構造における新規力場パラメータの信頼性評価

1B03、1J4M、1LE0、1LE1、1LE3、1NIZ の 6 つのタンパク質における、all_helix 構造と Native 構造の MMPBSA(GBSA)計算によるポテンシャルエネルギー差を Table 13 に示す。ff03 force field を用いた場合、GB 法による Native 構造と helix 構造のエネルギー差 ΔEGB [EGB (Native) - EGB (All_helix)] は 1B03、1J4M、1LE0、1LE1、1LE3、1NIZ でそれぞれ、1.84、23.42、-14.82、14.50、18.50、17.58 (kcal/mol) であった。1LE0 を除く全てのタンパク質について、All_helix 構造の方がエネルギー的に安定となった。また、 ΔEPB [EPB (Native) - EPB (All_helix)] は 1B03、1J4M、1LE0、1LE1、1LE3、1NIZ でそれぞれ、-14.40、20.14、-12.38、13.88、9.54、15.78 (kcal/mol) であった。1B03 と 1LE0 を除く全てのタンパク質について、GB 法でのエネルギー評価と同様に、All_helix 構造の方がエネルギー的に安定となった。また、1B03 を除く全てのタンパク質で、GB 法と PB 法による ΔEGB の符号が一致し、両方法で安定と判断される構造が一致した。

一方、新規に開発した Original force field を用いた場合、 ΔEGB は 1B03、1J4M、1LE0、1LE1、1LE3、1NIZ でそれぞれ、-7.60、-2.01、-23.66、-1.53、-15.80、20.77 (kcal/mol) であった。また、 ΔEPB はそれぞれ、-25.72、-18.27、-22.66、-5.49、-10.71、15.93 (kcal/mol) であった。1NIZ 以外の全てのタンパク質について、Native 構造の方がエネルギー的に安定となった。また、全てのタンパク質の場合において、 ΔEGB と ΔEPB の符号が一致し、GB 法による構造安定性と PB 法による構造安定性の間に相関が見られた。

Table 15

Comparisons of energy differences between the native structure and the all_helix structure. Two different force fields, ff03 force field and the original developed force field, are compared.

	ΔE_{GB} (kcal/mol) ff03	ΔE_{PB} (kcal/mol) ff03	ΔE_{GB} (kcal/mol) Original FF	ΔE_{PB} (kcal/mol) Original FF
1B03	1.84	-14.40	-7.60	-25.72
1J4M	23.42	20.14	-2.01	-18.27
1LE0	-14.82	-12.38	-23.66	-22.66
1LE1	14.50	13.88	-1.53	-5.49
1LE3	18.50	9.54	-15.80	-10.71
1NIZ	17.58	15.78	20.77	15.93

考察

既存力場の helix 構造過剰評価とタンパク質立体構造予測の力場依存性

既存の力場パラメータを用いてタンパク質立体構造予測を行った場合、一つも β -sheet 構造の予測に成功しなかった。 β -sheet 構造となるべき領域で helix 構造が形成されてしまった。1LE3 の PDB 構造は GLU2、TRP3、THR4、THR11、THR13、THR15 の 6 残基が C5 コンフォメーションをとり、TRP5、TRP12、TRP14 の 3 残基が β コンフォメーションをとっている。その結果、GLU、TRP、THR の 3 つのアミノ酸が β -sheet 構造を形成している。ここで、 β -sheet 構造を形成している 9 残基中 8 残基を占める THR と TRP のコンフォメーション安定性について注目すると、MD シミュレーションにおける IGB=1 (gbsa=0, 1)、IGB=2 (gbsa=0, 1)、IGB=5 (gbsa=0) の場合で α R コンフォメーションが最も安定であった。また、IGB=5 (gbsa=1) においては β コンフォメーションが最も安定であった。これらの安定性は量子化学計算による安定性とは異なる結果である。通常 helix 構造は 4 残基ごとの水素結合によって形成される。しかし、この水素結合系が存在しない ACE-XXX-NME ジペプチドモデルにおいて α R コンフォメーションが最も安定になるということは、helix 構造を過剰に安定化してしまう力場パラメータであると考えられる。IGB=5 (gbsa=1) の場合は β コンフォメーションが最も安定であったものの、 β -sheet 構造を形成する C5 コンフォメーションは α R コンフォメーションよりも不安定であった。この安定性順位は量子化学的安定性とは逆の結果である。IGB=5 (gbsa=1) の場合でも α R コンフォメーションを過剰に安定化し、helix 構造を不当に安定評価する力場であると考えられる。よって、THR、TRP が連続して存在する 1LE3 では、 α R コンフォメーションが過剰に安定化する力場パラメータに依存して β -sheet 構造が予測できなかったと考えられる。また、1PGB の β -sheet 構造領域について、PDB 構造は TYR3、LYS4、LEU5、ILE6、LEU7、LEU12、LYS13、GLU15、THR16、THR17、THR18、GLU19、ALA20、GLY41、TRP43、THR44、TYR45、ASP46、THR51、PHE52、THR53、VAL54、THR55 が C5 コンフォメーションをとっている。その結果、ALA、ASP、GLU、GLY、ILE、LEU、LYS、PHE、THR、TRP、TYR、VAL の 12 種類のアミノ酸によって 1PGB の β -sheet 構造は形成されている。これらのアミノ酸のコン

フォメーション安定性に注目すると、GLU、GLY、PHE、VAL の 4 種類以外のアミノ酸は MD シミュレーションのほぼ全ての方法で α R コンフォメーションが最も安定であった。この結果は量子化学計算の結果と異なる結果である。1LE3 の場合と同様に 1PGB の場合も、 β -sheet 構造が予測できなかった原因は helix 構造を過剰に安定化してしまう力場パラメータにあると考えることができる。

一方、1L2Y は IGB=2 (gbsa=0) の場合を除くすべての方法で比較的精度良く予測に成功している。1L2Y の PDB 構造は 3 つの主要な二次構造から形成されている。LEU2、TYR3、ILE4、GLN5、TRP6、LEU7、LYS8、ASP9 の 8 残基が α R コンフォメーションをとり helix 構造を形成し、PRO12、SER13 の 2 残基が α R コンフォメーションをとり 3-10 helix 構造を形成し、ARG16 が C5 コンフォメーションを、PRO17、PRO18、PRO19 が β コンフォメーションをとり、strand 構造を形成している。また、各構造と構造の間のターン様構造領域には、主鎖二面角の回転自由度がアミノ酸 20 種の中で最も大きい GLY が存在している。まず helix 構造を形成している、ASP、GLN、ILE、LEU、TRP、TYR のコンフォメーション安定性について注目すると、GLN 以外のアミノ酸について、MD シミュレーションのほぼ全ての方法で量子化学計算の結果とは異なり α R コンフォメーションが最も安定となった。helix 構造をとるべき領域で都合良く、helix 構造を安定化する力場パラメータが存在したため、helix 構造の予測に成功したものと考えられる。次に 3-10helix 構造を形成している PRO、SER のコンフォメーション安定性について注目すると、SER では量子化学計算の結果と異なり α R コンフォメーションが、PRO では量子化学計算と同じく β コンフォメーションがそれぞれ安定であった。この領域の予測に成功した方法は IGB=1 (gbsa=0, 1) の 2 つの方法だけである。その他の方法では helix の巻が緩くなってしまい、予測には失敗した。SER のコンフォメーション安定性について、MD シミュレーションのどの方法においても α R コンフォメーションが最も安定であり、次に安定なコンフォメーションは C5 コンフォメーションである。両コンフォメーションの差は IGB=1 (gbsa=0) : 1.38 kcal/mol であり、IGB=1 (gbsa=1) : 1.92 kcal/mol であった。その他の方法では、0.71 kcal/mol ~ 1.02 kcal/mol であった。IGB=1 (gbsa=0, 1) が最も α R の安定性が高い。そのため IGB=1 (gbsa=0, 1) のみ、この領域の helix 構造の予測に成功

したものと考えられる。最後に strand 構造を形成している ARG、PRO に注目した。コンフォメーション安定性について、MD シミュレーションの全ての方法で β コンフォメーションが最も安定であった。ARG のみ量子化学計算とは異なる結果であった。この場合も helix 構造領域の時と同様に、strand 構造領域で都合良く strand 構造をとりやすい β コンフォメーションを安定化する力場パラメータを持つアミノ酸が多く存在したため、strand 構造の予測に成功したものと考えられる。1L2Y は helix 構造を過剰に安定化する力場を用いれば、GLY が turn 様領域に存在することと、PRO がその構造特性上 strand 用構造をとりやすいことから、比較的容易に予測可能なモデルであると言える。これまで報告されている MD シミュレーションによるフォールディングの成功例が 1L2Y だけであることも、以上の理由から理解することができる。

1VII は IGB=1 と IGB=5 の場合、ある程度予測精度は良好であった。二次構造的特性は比較的良く予測に成功している。1VII の PDB 構造は、ASP4、GLU5、ASP6、PHE7、LYS8、ALA9、ARG15、SER16、ALA17、PHE18、ALA19、LEU23、TRP24、LYS25、GLN26、GLN27、ASN28、LEU29、LYS30、LYS31、GLU32 が α R コンフォメーションをとり helix 構造を形成している。helix 構造と helix 構造の間の 2ヶ所の無定形部位には、GLY または PRO が存在している。まず helix 構造を形成している、ALA、ARG、ASP、GLN、GLU、LEU、LYS、PHE、SER、TRP のコンフォメーション安定性について注目すると、ARG、GLN、GLU、PHE 以外のアミノ酸について、MD シミュレーションのほぼ全ての方法で量子化学計算の結果とは異なり α R コンフォメーションが最も安定となった。この安定性により、helix 領域は比較的よく予測に成功していたと考えることができる。ARG、GLN、GLU、PHE は他のアミノ酸に比べ、 α R コンフォメーションの安定性が低い。予測構造でもこのアミノ酸の存在する部分で helix 構造の巻きが緩くなる傾向があった。

1PGB の helix 領域について、PDB 構造では ALA23、ALA24、THR25、ALA26、GLU27、LYS28、VAL29、PHE30、LYS31、GLN32、TYR33、ALA34、ASN35、ASP36、ASN37 が α R コンフォメーションをとり、helix 構造を形成している。ALA、ASN、ASP、GLN、GLU、LYS、PHE、THR、TYR、VAL のコンフォメーション安定性は MD シミュレーションでは GLN、GLU、PHE、

VAL を除くすべてのアミノ酸で、ほぼすべての方法で αR コンフォメーションが安定となった。この結果は量子化学的安定性とは異なる結果である。1PGB は β -sheet 領域では予測が全く外れてしまったが、helix 領域では比較的予測が成功していた。これも、helix 構造をとる領域で都合よく helix 構造を過剰に安定評価する力場があったためと考えられる。

これまでに述べてきた過剰な helix 構造の安定性は、(6) の β -sheet 構造における力場パラメータの信頼性評価でも確認された。ff03 force field を使い、IGB=5 (gbsa=1) を使用した場合、6 個中 4 個で PDB 構造とは全く異なる all_helix 構造を過剰に安定化した。このことから、ff03 force field は過剰に helix 構造を安定化する力場パラメータであり、MD シミュレーションにより安定化される構造は、力場パラメータの影響を強く受けることが示された。

量子化学計算の妥当性と水の溶媒効果の重要性

ACE-XXX-NME のジペプチドモデルを使用し、各アミノ酸の C5 コンフォメーションと αR コンフォメーションの安定性を量子化学的に評価した。C5 コンフォメーションは β -sheet 構造を形成する重要なコンフォメーションとして、 αR コンフォメーションは helix 構造を形成する重要なコンフォメーションとして注目した。このうち ACE-ALA-NME モデルについては 2004 年に Yong Duan 等によって、真空中、Ether 溶媒中 ($\epsilon = 4.335$)、Water 溶媒中 ($\epsilon = 78.39$) の 3 つの異なる条件におけるコンフォメーション安定性の詳細な研究が行われた。^[42] その研究によれば、真空中と Ether 溶媒中では共に C7eq コンフォメーションが最も安定となり、helix 構造や β -sheet 構造などの主要な二次構造を形成する αR コンフォメーション、C5 コンフォメーション、 β コンフォメーションは C7eq コンフォメーションよりも不安定となっていた。一方、Water 溶媒中では αR コンフォメーション、C5 コンフォメーション、 β コンフォメーションが C7eq コンフォメーションよりも安定となっていた。さらに、この 3 つのコンフォメーションの中で C5 コンフォメーションが最も安定となっていた。 αR コンフォメーションよりも C5 コンフォメーションの方が安定であるという結果は我々の量子化学計算ともよく一致した。また、彼等の求めた αR コンフォメーションの最適化構造と C5 コンフォメーションの最適化構造の主鎖 Φ 、 Ψ 角はそ

れぞれ、 $(\Phi, \Psi) = (-70.5, -32.1)$ 、 $(\Phi, \Psi) = (-156.4, 143.8)$ であり、我々の求めた最適化構造の $(\Phi, \Psi) = (-77.9, -26.2)$ 、 $(\Phi, \Psi) = (-156.4, 149.5)$ はよく一致した。ここで、Yong Duan 等の量子化学計算では構造最適化計算に MP2/6-31G**を用い、各コンフォメーションポテンシャルエネルギー計算には MP2/cc-pVTZ を用いており (MP2/6-31G**//MP2/cc-pVTZ)、我々の使用した HF/6-31G**//B3LYP/cc-pVTZ に比べると精度が高い方法を使用している。しかし、我々が行った量子化学計算は、彼等が行った量子化学計算とコンフォメーション安定性的にも、構造的にもよく一致した結果を得ており、妥当な計算であると言える。

Yong Duan 等の真空中と Ether 条件下の量子化学計算の結果は、誘電率が小さな系では、タンパク質構造が正しく折り畳まれない可能性があることを強く示している。特に、この条件下で最も安定となった C7eq コンフォメーションは、分子内で ACE の C=O と NME の N-H が水素結合を取ることによって形成される構造であることから、近傍の電氣的相互作用を過剰に見積った構造を安定化してしまう恐れがある。また、タンパク質の二次構造的知見から考えると、分子内水素結合は β -sheet 構造よりも helix 構造の方が多いため、誘電率が小さいと helix 構造を過剰に安定化させてしまう可能性がある。実際にいくつかの実験において、誘電率が低ければタンパク質は helix 構造を形成しやすくなることが確認されている。^[43-45] 一方、水の溶媒効果を加味した場合、helix 構造や β -sheet 構造などの主要な二次構造を形成する α R コンフォメーション、C5 コンフォメーション、 β コンフォメーションが他のコンフォメーションに比べ良く安定化されていた。さらに、真空中と Ether 条件下では構造が不安定すぎるため、構造最適化計算の後、C5 コンフォメーションか C7eq コンフォメーションの構造になってしまっていた β コンフォメーションが、Water 条件下ではよく安定化されている。 β コンフォメーションは strand 構造や一部の β -sheet 構造を形成する上で重要なコンフォメーションである。このことから、タンパク質の正常な折り畳みには水の溶媒効果が大変重要な役割を示すことが考えられる。これまでの我々の研究においても、タンパク質構造の正常な折り畳みには水分子が重要であることが示されている。^[25-26]

既存力場を用いた MD シミュレーションによるコンフォメーション安定性と溶媒効果

既存力場パラメータである ff03 force field を使い、ACE-XXX-NME ジペプチドモデルに対して MD シミュレーションを実行した。IGB オプションが 1、2、5 の 3 つの全ての連続溶媒モデルを試行し、その各々の場合について SA 計算を導入する場合 (gbsa=0) と導入しない場合 (gbsa=1) の 2 つの方法を試した。しかし、いずれの IGB オプションを使用しても、また、SA 計算を導入してもしなくても、多くのアミノ酸で α R コンフォメーションの過剰な安定性が確認された。特に、アミノ酸主鎖の Φ 角が -60° 付近で大きく安定化しているため、 α R や β コンフォメーションが過剰に安定化してしまい、 Φ 角が -180° 付近の C5 コンフォメーションが不安定化している。一般的にタンパク構造の振る舞いは MD シミュレーションの力場パラメータと周囲の環境に大きく依存する。周囲の環境を IGB と gbsa オプションにより 6 つの方法でそれぞれ試したが、どの方法でも α R コンフォメーションの過剰安定性を改善することができなかった。よって、このような α R コンフォメーション過剰安定性は力場パラメータに強く依存した結果誤って導出された安定性である可能性が高い。

一方、連続溶媒モデルではなく、実際に水分子を ACE-ALA-NME ジペプチドモデルの周囲に発生させた系について MD シミュレーションを行った。しかし、この場合も連続溶媒モデルを使用した場合と同様に、 Φ 角の -60° 付近の大きな安定化が見られ、それにともない α R、 β コンフォメーションが過剰に安定化した。水分子を実際に発生させた系であっても、連続溶媒モデルにより溶媒効果を取り入れた系とコンフォメーション安定性には差はないことが考えられる。

タンパク質構造は種々のアミノ酸構造の集まりからなる。このアミノ酸構造の正しいコンフォメーションを評価できないパラメータを使っている限り、周囲の環境やモデル計算系をどんなに精度良く表現したとしても、正しい MD シミュレーションは実行されない可能性が高い。このことを示すように、 α R コンフォメーション安定性が高い ff03 force field を使ったタンパク質立体構造予測は、 β -sheet 構造領域で全く成功していない。

既存力場パラメータによるタンパク質立体構造予測に対する、MMPBSA(GBSA)計算を使った計算条件検討

既存パラメータ ff03 force field を用いたタンパク質立体構造予測の後、Native 構造と予測構造のポテンシャルエネルギー計算を MMPBSA(GBSA)法を用いて評価した。IGB=1 (gbsa=0, 1) では 1LE3 と 1PGB で ΔE_{GB} と ΔE_{PB} の差が大きく、特に、gbsa=0 の 1PGB、gbsa=1 の 1LE3 と 1PGB ではエネルギーの符号が逆転し、GB 法と PB 法で安定と判断される構造が逆になってしまった。GB 法と PB 法では同じ構造のエネルギー差を評価しているにも関わらず、両エネルギー差に相関が見られないことから、IGB=1 (gbsa=0, 1) は正しく溶媒効果を表現できていないことが考えられる。また、1LE3 と 1PGB はともに β -sheet 構造をとることから、IGB=1 (gbsa=0, 1) は特に、 β -sheet 構造を正しく評価できない方法であるとも言える。これまでの研究で β -sheet 構造には水の溶媒効果が重要であることが分かっているが、^[25] 溶媒効果が不十分なために β -sheet 構造を持つ 1LE3 と 1PGB のみエネルギー評価が正しく行われなかったというこの結果は、この研究結果とも矛盾しない。1VII では Native 構造が MD シミュレーションにより完全に崩壊してしまっただ。1VII は 3 つの helix 構造が 3 つの PHE 側鎖の π - π 相互作用と helix どうしの疎水的相互作用によって折り畳まれている構造である。よって、IGB=1 (gbsa=0, 1) は疎水的相互作用や π - π 相互作用を正しく表現できていないことが考えられる。

IGB=2 (gbsa=0, 1) ではすべてのタンパク質で ΔE_{GB} と ΔE_{PB} のエネルギーの符号が一致し、GB 法と PB 法で安定と判断される構造が同じものとなった。しかし、1L2Y と 1VII で Native 構造が MD シミュレーションにより即座に崩壊してしまっただ。1L2Y では TYR、TRP の π - π 相互作用と、TRP、PRO の疎水的相互作用が構造形成に重要な役割を示している。1VII も π - π 相互作用と疎水的相互作用によって構造が形成されている。このことから、IGB=2 (gbsa=0, 1) は疎水的相互作用や π - π 相互作用を正しく表現できていないことが考えられる。

IGB=5 (gbsa=0) では 1LE3 を除くすべてのタンパク質について ΔE_{GB} と ΔE_{PB} のエネルギーの符号が一致し、GB 法と PB 法で安定と判断される構造が同じものとなった。また、IGB=1 (gbsa=0, 1)、IGB=2 (gbsa=0, 1) に比べ、 ΔE_{GB} と ΔE_{PB} のエネルギー値そのものが近い値をとっている。また、

IGB=1 (gbsa=0, 1)、IGB=2 (gbsa=0, 1) ではこれまで 1L2Y や 1VII について MD シミュレーションにより構造が即座に壊れてしまっていたが、IGB=5 (gbsa=0) では PDB 構造をよく保ったまま MD シミュレーションが実行された。IGB=5 (gbsa=0) は疎水的相互作用や π - π 相互作用は比較的良く表現できるものの、一部の β -sheet 構造に対して溶媒効果を正しく表現できない方法であると言える。一方 IGB=5 (gbsa=1) では、すべてのタンパク質について ΔE_{GB} と ΔE_{PB} のエネルギーの符号が一致し、GB 法と PB 法で安定と判断される構造が同じものとなった。また、これまでの方法の中で最も ΔE_{GB} と ΔE_{PB} のエネルギー値が近い値となった。これは IGB=5 (gbsa=1) が全ての GB 法の中で最も精度が高いことを示している。また、IGB=1 (gbsa=0, 1)、IGB=2 (gbsa=0, 1) で見られた MD シミュレーション中の構造崩壊は見られなかった。よって、IGB=5 (gbsa=1) は最もよく溶媒効果を表現した連続溶媒和モデルであると判断できる。

IGB=5 (gbsa=1) について、1PGB を除く全てのタンパク質で、Native 構造よりも予測構造の方が安定であった。これは、1LE3、1L2Y、1VII ではエネルギー的に十分安定な構造を予測できており、エネルギーの底を見つけるという目的に対して、この計算条件は妥当であることを示している。しかし 1LE3 と 1VII に対して、得られた予測構造は PDB 構造とは大きく異なった。特に、1LE3 では β -sheet 構造を形成すべき領域で helix 構造を形成した。予測構造は PDB 構造と全く異なるにも関わらず、Native 構造よりも予測構造の方がエネルギー的に安定であるということは、ff03 force field を用いた MD シミュレーションによる立体構造予測では PDB 構造を予測することは不可能であると判断できる。1PGB では Native 構造の方がエネルギー的に安定である。これは、まだ十分にエネルギーの底を見つけることができていないことを示しており、より長い MD シミュレーション時間を必要としている可能性がある。1PGB は残基数 56 で、1LE3 (残基数 : 17)、1L2Y (残基数 : 20)、1VII (残基数 : 36) と比べても大きいことが分かる。通常、タンパク質フォールディングはミリ秒~秒オーダーでおこる言われている。^[46] しかし、今回行われたタンパク質立体構造予測シミュレーションは秒よりも非常に短いナノ秒オーダーのシミュレーションであり、 10^9 秒スケールの非常に短い現象を追ったに過ぎない。1PGB のような helix 構造と β -sheet 構造の両方を持つ複雑なタンパク質については、秒

オーダーの長期シミュレーションの実行が理想的である。しかし、現在の計算機能力ではこの長期シミュレーションは難しい。今後は短時間のシミュレーションで長期シミュレーションを再現できるような、理想的な加速試験法の考案が必要不可欠である。

新規力場パラメータの妥当性

(a) コンフォメーション安定性における新規力場パラメータの妥当性

新規に開発した力場パラメータを使い、ACE-XXX-NME ジペプチドモデルに対して MD シミュレーションを実行した。既存の ff03 force field を用いた場合で、最も精度の高い溶媒効果であった IGB=5 (gbsa=1) の連続溶媒モデルを使った。C5 コンフォメーションと α R コンフォメーションのエネルギー安定性について、CYS、MET、THR、VAL を除く全てのアミノ酸で量子化学計算と矛盾の無い結果が得られた。このことは力場パラメータの精度が大きく改善したことを示している。この計算モデルは helix 構造形成のための水素結合系の無いジペプチドモデルである。その計算モデルにおいて、 α R コンフォメーションよりも C5 コンフォメーションの方がエネルギー的に安定となったことは、タンパク質立体構造シミュレーション中に helix 構造を過剰に安定化していた問題は改善できたことを示している。ここで量子化学計算と矛盾のあった THR の最安定コンフォメーションについて注目した場合、量子化学計算においても MD シミュレーションにおいても β コンフォメーションが最も安定となった。最安定コンフォメーションが矛盾の無いコンフォメーションであったことから、THR については量子化学計算の結果を大きく無視した力場パラメータではないと判断できる。コンフォメーションレベルにおいて、量子化学計算と矛盾の無い高精度の力場パラメータ開発に成功したものと見える。

(b) タンパク質構造予測における新規力場パラメータの妥当性

既存の力場パラメータを用いてタンパク質立体構造予測を行った場合、一つも β -sheet 構造の予測に成功しなかったが、新規に開発した力場パラメータを用いてタンパク質立体構造予測を行った場合、1LE3 の β -sheet 構造となるべき領域で β -sheet 様構造を予測することに成功した。しかし、1PGB の β -sheet 構造領域では予測に失敗した。1PGB では、予測構造と PDB

構造のポテンシャルエネルギー差 (ΔE_{GB}) は 127.54 kcal/mol であり予測構造はまだ十分に折り畳みができていないものと考えられる。したがって、1PGB の予測失敗の原因は力場パラメータではなく、シミュレーション時間である可能性が高い。また、1PGB の β -sheet 構造領域で一部 helix 構造を形成してしまっているが、これもシミュレーション時間が十分でないことが原因であると考えることができる。1PGB の β -sheet 構造領域は 4 本の β -sheet 構造からなるため、この 4 本の構造が十分近づき、相互作用系を形成した時、はじめて β -sheet 構造が安定になると考えられる、実際に β -lactoglobulin は 4 本の helix 構造が集まった後、 β -sheet 構造を形成している。^[23] 一方 1L2Y、1VII において、helix 構造をとるべき領域で正しく helix 構造が予測された。このことは、新規力場パラメータが正確に、helix 構造を形成する際の水素結合による安定化によって、正しく helix 構造を安定化していることを示している。

ここで、1LE3、1L2Y、1VII、1PGB の MD シミュレーション中の全構造に対するポテンシャルエネルギーグラフと RMSD($C\alpha$) グラフに注目すると (Fig. 12(a)-(b)、Fig. 13(a)-(b)、Fig. 14(a)-(b)、Fig. 15(a)-(b))、両データに相関関係があると考えることができる。1LE3 において、0-ps ~ 2500-ps ではポテンシャルエネルギーが徐々に減少する傾向にあるが、このとき、RMSD($C\alpha$) も徐々に減少傾向を示すようになっている。次いで 2500-ps ~ 5000-ps で再びポテンシャルエネルギーが上昇するが、この時も RMSD($C\alpha$) が徐々に大きな値をとっている。次に特徴的なのが、20000-ps 前後での振る舞いである。20000-ps 手前ではポテンシャルエネルギー、RMSD($C\alpha$) とともに減少傾向を示し、20000-ps を超えたあたりから再び上昇傾向を示している。最後に 25000-ps 手前で再び両データは減少傾向を示した。1L2Y についても同様に、ポテンシャルエネルギーと RMSD($C\alpha$) が同方向への振る舞いを見せた。5000-ps ~ 10000-ps の間でこれら 2 つのデータは 1 回上昇傾向を示し、20000-ps ~ 25000-ps の間では平均的に高い値をとっていた。1VII、1PGB についても同様に、2 つのデータはシミュレーションが進むにつれ、徐々に小さな値をとる傾向を示した。これらの事実から、新規に開発した力場パラメータでは、PDB 構造に近ければ近いほどエネルギー的に安定であると考えることができる。すなわち、この新規力場パラメータはポテンシャルエネルギーで構造安定性が議論できるほど、高精度

なパラメータであると考えられる。一方、既存 ff03 force field では過剰な helix 構造への安定傾向があったため、ポテンシャルエネルギーとタンパク質構造安定性との間にこのような明確な傾向を見ることはできなかった。

(c) β -sheet 構造に対する新規力場パラメータの妥当性

1NIZ 以外のすべてのタンパク質について、Native 構造の方が安定となった。1NIZ 構造だけが新規力場パラメータを用いても all_helix 構造が安定であった。1NIZ 構造は HIV-1 gb120 の V3 loop 構造の一部であり、この領域の本来の配列は 301~335 で”CTRPNYNKRK RIHIGPGRAF YTTKNIIGTI RQAHC”である。また、301 の CYS と 335 の CYS のジスルフィド結合が形成されていることが知られている。さらに、PDB には単独の構造として登録されているが、NMR の構造決定の際には”447-52D A HUMAN HIV-1 NEUTRALIZING ANTIBODY”と複合体を形成しているため、真に独立した構造ではない。1NIZ 以外の構造はすべて Peptide synthesizer によって合成され、単独で構造が決定されたものである。これらの事実から、”ACE-<KRIHIGPGRAFYTT>-NME”単独の系では真に NMR 測定環境を再現しておらず、この系単独ならば helix 構造の方が安定である可能性がある。

ここで 1CE4 は 1NIZ の”KRIHIGPGRAFYTT”とよく似た配列を持ち、且つ、Peptide synthesizer によって合成され、単独で構造が決定されたものである。^[47] 単独で構造が決定されたものの中では最も 1NIZ と配列が良く似ている。また、1NIZ の V3 領域のように CYS によるジスルフィド結合を形成している。また、1NIZ では β -sheet 構造をとっていたが、1CE4 では β -sheet 構造を全くとらず、bend 構造、turn 構造、3-10helix 構造、helix 構造によって構造が形成されている。さらに、構造の C 末端側半分は主に helix 構造を形成している。(Fig. 18)

1CE4	CTRPNNNTRK S IHIGPGRAFYTTGEIIGDIRQAHC
2D structure	---SGGG--S---S--TTGGGGTHHHHHHHHHHT
1NIZ (model)	-----XKR I HIGPGRAFYTTX-----
2D structure	-----EEEE----EEEE-----

Fig. 18

Sequences (black) and secondary structures (blue) of 1CE4 and 1NIZ are expressed by the one-letter code. These two structures have only one different amino acid (represented by black *box* emphasis).

1CE4 について、ジスルフィド結合を持つが helix 構造を形成していることより、ジスルフィド結合は β -sheet 構造の安定化に直接関与していない可能性がある。また、1NIZ のようにジスルフィド結合が無ければ、N 末端側は伸びず、all_helix のようなすべて巻いた構造が安定になる可能性が大きい。また、turn 構造領域に注目すると、この領域に PRO や GLY がよく見られるが、これは他のタンパク質においてもよく見られる傾向である。さらに、1NIZ の all_helix 構造の MD 計算後の構造はやや歪んだ形をとり、turn 構造まではいかないものの、helix を 2 本に分けるくびれを作っている。したがって、1NIZ の MMPBSA(GBSA)計算において all_helix 構造の方が安定であったことは、妥当な結果である可能性がある。以上の事実から、新規力場パラメータは β -sheet 構造を高精度で評価できるパラメータであると考えられる。さらに、タンパク質立体構造予測において helix 構造領域においても高精度での予測に成功していること、1NIZ を all_helix 構造の方が安定であると評価したことから、helix 構造においても高精度で評価可能なパラメータであると考えられることができる。

(d) 新規力場パラメータの有用性

今回新たに開発した力場パラメータとこれまでの力場パラメータの決定

的な違いは、電荷パラメータを決める際の溶媒効果の取り扱い方である。ff03 force field 以前の力場パラメータは真空条件下で、ff03 force field は $\epsilon = 4.335$ 程度の ether 溶媒条件下で導出していた。^[22] 一方、今回新規に開発した力場パラメータは $\epsilon = 78.39$ 程度の Water 溶媒条件下で導出したものである。ff03 force field はそれ以前の力場パラメータと比べると格段に良くタンパク質構造を表現できるようになっている^[22]。これは、電荷パラメータ決定の際に溶媒効果を導入することが有効であることを示している。一方で、 $\epsilon = 4.335$ 程度の溶媒環境というものは、巨大タンパク質の内部や、細胞膜内部の環境に近く、溶媒にあらわになりやすい比較的小さなタンパク質や、細胞膜外に露出した細胞膜結合タンパク質のループ部分などは、この環境では誘電率が低すぎるものと考えられる。また、低い誘電率は helix 構造形成を誘導する実験結果も知られている。^[43-45] さらに、これまでの研究でタンパク質構造の安定性には水分子が必要不可欠であることが分かっている。^[25-26] よって、今回の立体構造予測対象であった比較的小さなタンパク質や β -sheet 構造を持つタンパク質、 β -sheet 構造安定性評価に用いた 30 残基以下のタンパク質には ff03 force field は適応できなかった可能性が高い。また、タンパク質機能に重要な役割を果たす ligand 結合部位に注目した場合、ligand が入るだけの入り口を持つことから、容易に水分子にさらされるものと考えられる。実際に多くの X 線結晶構造や MD シミュレーションにおいて、タンパク質活性部位に水分子が確認され、この水分子がタンパク質機能発現に大きな役割を示している場合が多い。^[48-50] このような観点からすると、電荷パラメータに水の溶媒効果を導入することは、単にタンパク質二次構造安定性だけではなく、タンパク質機能解析についても非常に重要であるといえる。

タンパク質立体構造予測法と MD シミュレーション時間

新規力場パラメータを使ったタンパク質立体構造予測において、375K の温度で 30-ns の MD シミュレーションを実行した。MD シミュレーション中の全構造に対する主成分解析の結果 (Fig. 12(c)、Fig. 13(c)、Fig. 14(c)、Fig. 15(c)) に注目すると、どのタンパク質の場合においても、構造が十分平衡に達しているほどの集合は確認できなかった。よって、30-ns の MD シミュレーションであっても、まだ不十分であることが考えられる。とく

に、1VII、1PGB の第一相成分と第二相成分のスケールは 1LE3、1L2Y と比べると非常に大きく、30-ns 付近でも構造が落ち着かないことが良く確認できる。通常タンパク質立体構造の折り畳みはミリ秒~秒オーダーのスケールで起ると考えられている。^[46] よって、MD シミュレーション時間の延長、効果的な加速的試験法の開発が求められる。

一方、ここで再びポテンシャルエネルギーと RMSD(C α) の関係に注目すると (Fig. 12(a)-(b)、Fig. 13(a)-(b)、Fig. 14(a)-(b)、Fig. 15(a)-(b))、比較的高い相関があるものの、完全な構造予測ができる程に高い相関関係があるわけではないことが分かる。すなわち、ポテンシャルエネルギーの低い方から数えて 5 つの構造は、パラメータの高精度化に共ない 30000 個全体の構造の中では比較的小さな RMSD(C α) をとることができた。しかし、最もポテンシャルエネルギーが低い構造が、最も小さな RMSD(C α) をとるわけではなかった。そこで、主成分解析を行い、5 つの構造の初期構造からの離れ具合、構造の集まり具合を考えることで、1LE3、1VII では 5 つの構造の中で最も RMSD(C α) が小さいものを取り出すことに成功した。また、1L2Y でも 5 つの構造の中で 2 番目に RMSD(C α) が小さい構造を取り出すことができた。このことは、ポテンシャルエネルギーを主成分解析による構造クラスタリングと組み合わせることで、効果的にタンパク質立体構造予測を行うことができることを示している。また、1PGB については 5 つの構造の中で 3 番目に RMSD(C α) が小さな構造を導いたが、ポテンシャルエネルギー、主成分解析、いずれからもタンパク質構造が平衡状態でないことが判断されることから、シミュレーション時間が足りないために最適な構造の導出に失敗したものと考えられる。

結論

MD シミュレーションは力場パラメータに強く依存し、ff03 force field は過剰な helix 構造安定化傾向を示す。

MD シミュレーションの高精度な実行には溶媒効果が必要不可欠であり、IGB=5 (gbsa=1) が最も精度良い溶媒効果を示す。

MD シミュレーションを用いたタンパク質立体構造予測には非常に大きなシミュレーション時間が必要である。

新規に開発した力場パラメータは、コンフォメーション的、タンパク質二次構造的に大変優れた高精度なパラメータである。

謝辞

本研究は、“JSPS Research Fellowships for Young Scientists”、“Grant-in-Aid for Center Of Excellence (COE) research from the Ministry of Education, Science, Sport, and Culture, Japan”、“Japan Science and Technology Agency”の支援を受けて遂行されました。また、本研究室の計算設備の他に、“Computational Astrophysics Laboratory, RIKEN”、“Research Center for Computational Science in Okazaki”、“Institute of Media and Information Technology, Chiba University”、“DRIA system at the Graduate School of Pharmaceutical Science, Chiba University”の計算機施設の協力を得て、研究を遂行することができました。心より御礼申し上げます。

本研究の遂行にあたり、終止御指導御鞭撻を賜りました、千葉大学大学院薬学研究院薬品物理化学研究室の根矢三郎教授、星野忠次助教授、畑晶之助手に謹んで御礼申し上げます。

本論文の審査をしていただき、かつ多くのご助言を賜りました、千葉大学大学院薬学研究院病態生化学研究室の五十嵐一衛教授、同研究院生化学研究室の小林弘教授、同研究院生体分析化学研究室の戸井田敏彦教授に厚く御礼申し上げます。

長きに渡りお世話になりました、千葉大学大学院薬品物理化学研究室の皆様感謝いたします。

参考文献

1. Tama F, Miyashita O, Kitao A, Go N **Molecular dynamics simulation shows large volume fluctuations of proteins** Eur Biophys J 2000 29(7),472-80
2. Fadel A .Samatey, Matsunami H, Imada K, Nagashima S, Tanvir R. Shaikh, Dennis R. Thomas, James Z. Chen, David J. De Rosier, Kitao A and Namba K **Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism** NATURE, 2004 431(28 October), 1062-1068
3. Weinzinger P, Hannongbua S, Wolschann P **Molecular mechanics PBSA ligand binding energy and interaction of Efavirenz derivatives with HIV-1 reverse transcriptase** J Enzyme Inhib Med Chem. 2005 Apr;20(2):129-34
4. Rodriguez-Barrios F, Balzarini J, Gago F **The molecular basis of resilience to the effect of the Lys103Asn mutation in non-nucleoside HIV-1 reverse transcriptase inhibitors studied by targeted molecular dynamics simulations** J Am Chem Soc. 2005 May 25;127(20):7570-8
5. Rohs R, Bloch I, Sklenar H, Shakked Z **Molecular flexibility in ab initio drug docking to DNA: binding-site and binding-mode transitions in all-atom Monte Carlo simulations** Nucleic Acids Res. 2005 Dec 13;33(22):7048-57. Print 2005
6. Settanni G, Cattaneo A, Carloni P **Molecular dynamics simulations of the NGF-TrkA domain 5 complex and comparison with biological data** Biophys J. 2003 Apr;84(4):2282-92
7. Hillisch A, von Langen J, Menzenbach B, Droescher P, Kaufmann G, Schneider B, Elger W **The significance of the 20-carbonyl group of progesterone in steroid receptor binding: a molecular dynamics and structure-based ligand design study** Steroids. 2003 Nov;68(10-13):869-78

8. Dayam R, Neamati N **Active site binding modes of the beta-diketoacids: a multi-active site approach in HIV-1 integrase inhibitor design** Bioorg Med Chem. 2004 Dec 15;12(24):6371-81
9. Nakamura Y **Significance of SNP analysis on the future medicine** Nippon Rinsho. 1999 Nov;57(11):2615-22
10. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL **GenBank** Nucleic Acids Res. 2005 Jan 1;33(Database issue):D34-8
11. Andrei Kouranov, Lei Xie, Joanna de la Cruz, Li Chen, John Westbrook, Philip E. Bourne and Helen M. Berman **The RCSB PDB information portal for structural genomics** Nucleic Acids Research, 2006, Vol. 34, Database issue D302-D305
12. Krieger E, Nabuurs SB, Vriend G **Homology modeling** Methods Biochem Anal. 2003;44:509-23
13. Chevrier B, Schalk C, D'Orchymont H, Rondeau JM, Moras D, Tarnus C **Crystal structure of Aeromonas proteolytica aminopeptidase: a prototypical member of the co-catalytic zinc enzyme family** Structure. 1994 Apr 15;2(4):283-91
14. Stehle T, Schulz GE **Refined structure of the complex between guanylate kinase and its substrate GMP at 2.0 Å resolution** J Mol Biol. 1992 Apr 20;224(4):1127-41
15. Chowdhury S, Lee MC, Xiong G, Duan Y **Ab initio folding simulation of the Trp-cage mini-protein approaches NMR resolution** J Mol Biol. 2003 Mar 28;327(3):711-7
16. Hsieh M.J. and Luo R, **Physical Scoring Function Based on AMBER Force Field and Poisson-Boltzmann Implicit Solvent for Protein Structure Prediction**, Proteins, 2004 56:475-486

17. Fan, Z. Z., Hwang J K, and Warshel A, **Using simplified protein representation as a reference potential for all-atom calculations of folding free energy.** *Theor. Chem. Acc.* 1999 103:77-80
18. Vorobjev, Y. N. and Hermans J, **ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model.** *Biophys. Chem.*, 1999 78:195-205
19. Feig M, Brooks CL 3rd, **Evaluating CASP4 predictions with physical energy functions.** *Proteins.* 2002 Nov 1;49(2):232-45.
20. Dominy BN, Brooks CL. **Identifying native-like protein structures using physics-based potentials.** *J Comput Chem.* 2002 Jan 15;23(1):147-60.
21. Felts AK, Gallicchio E, Wallqvist A, Levy RM. **Distinguishing native conformations of proteins from decoys with an effective free energy estimator based on the OPLS all-atom force field and the Surface Generalized Born solvent model.** *Proteins.* 2002 Aug 1;48(2):404-22.
22. Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo & T. Lee. **A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins** *J. Comput. Chem.* **24**, 1999-2012 (2003)
23. Kuwata K, Hoshino M, Era S, Batt CA, Goto Y. **alpha-->beta transition of beta-lactoglobulin as evidenced by heteronuclear NMR.** *J Mol Biol.* 1998 Nov 6;283(4):731-9.
24. Sheinerman FB, Brooks CL 3rd. **Molecular picture of folding of a small alpha/beta protein.** *Proc Natl Acad Sci U S A.* 1998 Feb 17;95(4):1562-7.
25. Daisuke Katagiri, Takahiro Tsuchiya, Minoru Tsuda, Masayuki Hata, and Tyuji Hoshino **Computational Analysis of Stability of the $\hat{\alpha}$ -Sheet Structure** *J.*

26. Takahiro Tsuchiya, Daisuke Katagiri, Masayuki Hata, Tyuji Hoshino, Minoru Tsuda **Theoretical analysis of the stability of helices** *Journal of Molecular Structure (Theochem)* **589–590** (2002) 413–422
27. G.D. Hawkins, C.J. Cramer & D.G. Truhlar. **Pairwise solute descreening of solute charges from a dielectric medium** *Chem. Phys. Lett.* **246**, 122-129 (1995).
28. G.D. Hawkins, C.J. Cramer & D.G. Truhlar. **Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium** *J. Phys. Chem.* **100**, 19824-19839 (1996).
29. K.A. Sharp & B. Honig **Electrostatic interactions in macromolecules: Theory and experiment** *Annu. Rev. Biophys. Biophys. Chem.* **19**, 301-332 (1990).
30. M.E. Davis & J.A. McCammon **Electrostatics in biomolecular structure and dynamics** *Chem. Rev.* **90**, 509-521 (1990).
31. **AMBER8** D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, B. Wang, D.A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J.W. Caldwell, W.S. Ross, and P.A. Kollman (2004), AMBER 8, University of California, San Francisco.
32. Cochran AG, Skelton NJ, Starovasnik MA **Tryptophan zippers: stable, monomeric beta –hairpins** *Proc Natl Acad Sci U S A* 2002 Jun 25;99(13):9081
33. Neidigh JW, Fesinmeyer RM, Andersen NH **Designing a 20-residue protein** *Nat Struct Biol.* 2002 Jun;9(6):408-10
34. McKnight CJ, Matsudaira PT, Kim PS **NMR structure of the 35-residue villin**

headpiece subdomain Nat Struct Biol. 1997 Mar;4(3):180-4

35. Gallagher T, Alexander P, Bryan P, Gilliland GL **Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR** Biochemistry. 1994 Apr 19;33(15):4721-9
36. P. Cieplak, W.D. Cornell, C. Bayly & P.A. Kollman **Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA and proteins** *J. Comput. Chem.* **16**, 1357-1377 (1995)
37. C.I. Bayly, P. Cieplak, W.D. Cornell & P.A. Kollman **A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints For Determining Atom-Centered Charges: The RESP Model** *J. Phys. Chem.* **97**, 10269 (1993).
38. Tugarinov V, Zvi A, Levy R, Anglister J **A cis proline turn linking two beta-hairpin strands in the solution structure of an antibody-bound HIV-1IIIB V3 peptide** Nat Struct Biol. 1999 Apr;6(4):331-5
39. Pastor MT, Lopez de la Paz M, Lacroix E, Serrano L, Perez-Paya E **Combinatorial approaches: a new tool to search for highly structured beta-hairpin peptides** Proc Natl Acad Sci U S A. 2002 Jan 22;99(2):614-9. Epub 2002 Jan 8
40. Cochran AG, Skelton NJ, Starovasnik MA **Tryptophan zippers: stable, monomeric beta -hairpins** Proc Natl Acad Sci U S A. 2001 May 8;98(10):5578-83. Epub 2001 May 1 *Erratum in: Proc Natl Acad Sci U S A 2002 Jun 25;99(13):9081*
41. Sharon M, Kessler N, Levy R, Zolla-Pazner S, Gorkach M, Anglister J **Alternative conformations of HIV-1 V3 loops mimic beta hairpins in chemokines, suggesting a mechanism for coreceptor selectivity** Structure.

2003 Feb;11(2):225-36 *Comment in: Structure. 2003 Feb;11(2):129-30*

42. Wang ZX, Duan Y **Solvation effects on alanine dipeptide: A MP2/cc-pVTZ//MP2/6-31G** study of (Phi, Psi) energy maps and conformers in the gas phase, ether, and water** J Comput Chem. 2004 Nov 15;25(14):1699-716
43. Hirota, N., Mizuno, K., Goto, Y **Group additive contributions to the alcohol-induced alpha-helix formation of melittin: Implication for the mechanism of the alcohol effects on proteins** (1998). J. Mol. Biol. 275, 365-378
44. Luo, P., Baldwin, R. L **Mechanism of helix induction by trifluoroethanol: A framework fo extrapolating the helix-forming propensities of peptides from trifluoroethanol/water mixtures back to water** Biochemistry (1997) 36, 8413-8421.
45. Kurpin, S. Graslund, A., Ehrenberg, A., Koch, M.H.J **Nonideality of water-hexafluoropropanol mixtures as studied by X-ray small angle scattering** (1995) Biochem. Biophys. Res. Commun. 217, 1151-1156
46. Jin Wang **The Complex Kinetics of Protein Folding in Wide Temperature Ranges** Biophysical Journal 87:2164-2171 (2004)
47. Vranken WF, Budesinsky M, Fant F, Boulez K, Borremans FA **The complete Consensus V3 loop peptide of the envelope protein gp120 of HIV-1 shows pronounced helical character in solution** FEBS Lett. 1995 Oct 23;374(1):117-21
48. Zhao B, Guengerich FP, Voehler M, Waterman MR **Role of active site water molecules and substrate hydroxyl groups in oxygen activation by cytochrome P450 158A2: a new mechanism of proton transfer** J Biol Chem. 2005 Dec 23;280(51):42188-97. Epub 2005 Oct 20
49. Okimoto N, Nakamura T, Suenaga A, Futatsugi N, Hirano Y, Yamaguchi I,

Ebisuzaki T **Cooperative motions of protein and hydration water molecules: molecular dynamics study of scytalone dehydratase** *J Am Chem Soc.* 2004 Oct 13;126(40):13132-9

50. Daisuke Katagiri, Masayuki Hata, Takao Itoh, Sabro Neya, and Tyuji Hoshino **Atomic-Scale Mechanism of the GTP → GDP Hydrolysis Reaction by the G α 1 Protein** *J. Phys. Chem. B* 2003, 107, 3278-3283

MD シミュレーション変更使用オプション説明一覧

*1 DT

The time step (psec). Recommended MAXIMUM is .002 if SHAKE is used, or .001 if it isn't. Note that for temperatures above 300K, the step size should be reduced since greater temperatures mean increased velocities and longer distance traveled between each force evaluation, which can lead to anomalously high energies and system blowup. Default 0.001.

*2, *3, *4 About GB method

The generalized Born solvation model can be used instead of explicit water for non-polarizable force fields such as ff94 or ff99. There are several "flavors" of GB available, depending upon the value of *igb*. The version that has been most extensively tested corresponds to *igb*=1; the "OBC" models (*igb*=2 and 5) are newer, but appear to give significant improvements and are recommended for most projects (certainly for peptides or proteins). Users should understand that all (current) GB models have limitations and should proceed with caution. Generalized Born simulations can only be run for non-periodic systems, *i.e.* where *ntb*=0. The nonbonded cutoff for GB calculations should be greater than that for PME calculations, perhaps *cut*=16. The slowly-varying forces generally do not have to be evaluated at every step for GB, either *nrespa*=2 or 4.

*2 IGB=1

The Hawkins, Cramer, Truhlar [①,②] pairwise generalized Born model (*GBHCT*) is used, with parameters described by Tsui and Case [③]. This model uses the default radii set up by LEaP. It is slightly different from the GB model that was included in Amber6. If you want to compare to Amber 6, or need to continue an ongoing simulation, you should use the command "set default PBradii amber6" in LEaP, and set *igb*=1 in *sander*. For reference, the Amber6 values are those used by an earlier Tsui and Case paper [④]

*3 IGB=2

Use a modified GB model developed by A. Onufriev, D. Bashford and D.A. Case (*GBOBC*); the main idea was published earlier [⑤], but the actual implementation here is an elaboration of this initial idea [⑥]. The parameters α , β , and γ were determined by empirical fits, and have the values 0.8, 0.0, and 2.909125. This corresponds to model I in Ref [⑥]. With this option, you should use the LEaP command "set default PBradii mbondi2" or "set default PBradii bondi" to prepare the *prmtop* file.

*4 **IGB=5**

Same as *igb=2*, except that now α , β , γ are 1.0, 0.8, and 4.85. This corresponds to model II in Ref [⑥]. With this option, you should use the command "set default PBradii mbondi2" in setting up the *prmtop* file, although "set default PBradii bondi" is also OK. When tested in MD simulations of several proteins [⑥], both of the above parameterizations of the "OBC" model showed equal performance, although further tests [⑦] on an extensive set of protein structures revealed that the *igb=5* variant agrees better with the Poisson-Boltzmann treatment in calculating the electrostatic part of the solvation free energy.

*5, *6 **GBSA=0, 1**

Option to carry out GB/SA (generalized Born/surface area) simulations. For the default value of 0, surface area will not be computed and included in the solvation term. If *gbsa = 1*, surface area will be computed using the LCPO model. [⑧] If *gbsa = 2*, surface area will be computed by recursively approximating a sphere around an atom, starting from an icosahedra. Note that no forces are generated in this case, hence, *gbsa = 2* only works for a single point energy calculation and is mainly intended for energy decomposition in the realm of MM_GBSA.

*7 **CUT**

This is used to specify the nonbonded cutoff, in Angstroms. For PME, the cutoff is used to limit direct space sum, and the default value of 8.0 is usually a good value. When *igb>0*, the cutoff is used to truncate nonbonded pairs (on an atom-by-atom basis); here a larger value than the default is generally required. A separate parameter (**RGBMAX**) controls the maximum distance between atom pairs that will be

considered in carrying out the pairwise summation involved in calculating the effective Born radii, see the generalized Born section below.

*8 **RGBMAX**

This parameter controls the maximum distance between atom pairs that will be considered in carrying out the pairwise summation involved in calculating the effective Born radii. Atoms whose associated spheres are farther away than *rgbmax* from given atom will not contribute to that atom's effective Born radius. This is implemented in a "smooth" fashion (thanks mainly to W.A. Svrcek-Seiler), so that when part of an atom's atomic sphere lies inside *rgbmax* cutoff, that part contributes to the low-dielectric region that determines the effective Born radius. The default is 25 Å, which is usually plenty for single-domain proteins of a few hundred residues. Even smaller values (of 10-15 Å) are reasonable, changing the functional form of the generalized Born theory a little bit, in exchange for a considerable speed-up in efficiency, and without introducing the usual cut-off artifacts such as drifts in the total energy. The *rgbmax* parameter affects only the effective Born radii (and the derivatives of these values with respect to atomic coordinates). The *cut* parameter, on the other hand, determines the maximum distance for the electrostatic, van der Waals and "off-diagonal" terms of the generalized Born interaction. The value of *rgbmax* might be either greater or smaller than that of *cut*: these two parameters are independent of each other. However, values of *cut* that are too small are more likely to lead to artifacts than are small values of *rgbmax*; therefore one typically sets *rgbmax* <= *cut*.

*9 **NMROPT**

= 0

no nmr-type analysis will be done; default (Note: this variable replaces nmrmax from previous versions, and has a slightly different meaning.)

> 0

NMR restraints/weight changes will be read

= 2

NOESY volume restraints or chemical shift restraints will be read as well

*10 NRESPA

This variable allows the user to evaluate slowly-varying terms in the force field less frequently. For PME, "slowly-varying" (now) means the reciprocal sum. For generalized Born runs, the "slowly-varying" forces are those involving derivatives with respect to the effective radii, and pair interactions whose distances are greater than the "inner" cutoff, currently hard-wired at 8 Å.

If $NRESPA > 1$ these slowly-varying forces are evaluated every $nrespa$ steps. The forces are adjusted appropriately, leading to an impulse at that step. If $nrespa * dt$ is less than or equal to 4 fs the energy conservation is not seriously compromised. However if $nrespa * dt > 4$ fs the simulation becomes less stable. Note that energies and related quantities are only accessible every $nrespa$ steps, since the values at other times are meaningless.

*11, *12 About NTT option

Switch for temperature scaling. Note that setting $ntt=0$ corresponds to the microcanonical (NVE) ensemble (which should approach the canonical one for large numbers of degrees of freedom). Some aspects of the "weak-coupling ensemble" ($ntt=1$) have been examined, and roughly interpolate between the microcanonical and canonical ensembles [9]. The $ntt=2$ and 3 options correspond to the canonical (constant T) ensemble. The $ntt=4$ option is included for historical reasons, but does not correspond to any of the traditional ensembles.

*11 NTT=1

Constant temperature, using the weak-coupling algorithm [10]. A single scaling factor is used for all atoms. Note that this algorithm just ensures that the total kinetic energy is appropriate for the desired temperature; it does nothing to ensure that the temperature is even over all parts of the molecule. Atomic collisions should serve to ensure an even temperature distribution, but this is not guaranteed, and can be a particular problem for generalized Born simulations, where there are no collisions with solvent. Other temperature coupling options (especially $ntt=3$) should probably be used for generalized Born simulations.

***12 NTT=3**

Use Langevin dynamics with the collision frequency $_$ given by *gamma_ln*, discussed below. Note that when $_$ has its default value of zero, this is the same as setting *ntt* = 0.

***13 GAMMA_LN**

The collision frequency $_$, in ps⁻¹, when *ntt* = 3. A simple Leapfrog integrator is used to propagate the dynamics, with the kinetic energy adjusted to be correct for the harmonic oscillator case [(11),(12)]. Note that it is not necessary that $_$ approximate the physical collision frequency. In fact, it is often advantageous, in terms of sampling or stability of integration, to use much smaller values. Default is 0 [(12),(13)].

***14 NTB=1, 2**

Periodic boundary. If NTB .EQ. 0 then a boundary is NOT applied regardless of any boundary condition information in the topology file. The value of NTB specifies whether constant volume or constant pressure dynamics will be used. Options for constant pressure are described in a separate section below.

- = 0 no periodicity is applied and PME is off
- = 1 constant volume (default)
- = 2 constant pressure

If NTB .NE. 0, there must be a periodic boundary in the topology file. Constant pressure is not used in minimization (IMIN=1, above). For a periodic system, constant pressure is the only way to equilibrate density if the starting state is not correct. For example, the solvent packing scheme used in LEaP can result in a net void when solvent molecules are subtracted which can aggregate into "vacuum bubbles" in a constant volume run. Another potential problem are small gaps at the edges of the box. The upshot is that almost every system needs to be equilibrated at constant pressure (*ntb=2, ntp>0*) to get to a proper density. But be sure to equilibrate first (at constant volume) to something close to the final temperature, before turning on constant pressure.

***15 NTP=0, 1**

Flag for constant pressure dynamics. This option should be set to 1 or 2 when

Constant Pressure periodic boundary conditions are used (NTB = 2).

= 0 Used with NTB not = 2 (default); no pressure scaling

= 1 md with isotropic position scaling

= 2 md with anisotropic (x-,y-,z-) pressure scaling: this should only be used with orthogonal boxes (i.e. with all angles set to 90o).

Anisotropic scaling is primarily intended for non-isotropic systems, such as membrane simulations, where the surface tensions are different in different directions; it is generally not appropriate for solutes dissolved in water.

*16 **SALTCON**

Sets the concentration (M) of 1-1 mobile counterions in solution, using a modified generalized Born theory based on the Debye-Hückel limiting law for ion screening of interactions [14]. Default is 0.0 M (i.e. no Debye-Hückel screening.) Setting *saltcon* to a non-zero value does result in some increase in computation time.

MD シミュレーション変更使用オプション中の引用文献

- ①. G.D. Hawkins, C.J. Cramer & D.G. Truhlar. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **246**, 122-129 (1995).
- ②. G.D. Hawkins, C.J. Cramer & D.G. Truhlar. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **100**, 19824-19839 (1996).
- ③. V. Tsui & D.A. Case. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers (Nucl. Acid. Sci.)* **56**, 275-291 (2001).
- ④. V. Tsui & D.A. Case. Molecular dynamics simulations of nucleic acids using a generalized Born solvation model. *J. Am. Chem. Soc.* **122**, 2489-2498 (2000)
- ⑤. Onufriev, D. Bashford & D.A. Case. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **104**, 3712-3720 (2000)
- ⑥. Onufriev, D. Bashford & D.A. Case. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins* (2004). in press
- ⑦. M. Feig, A. Onufriev, M. Lee, W. Im, D. A. Case & C. L. Brooks, III. Performance Comparison of the Generalized Born and Poisson Methods in the Calculation of the Electrostatic Solvation Energies for Protein Structures. *J. Comput. Chem.* **25**, 265-284 (2004).
- ⑧. J. Weiser, P.S. Shenkin & W.C. Still. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.* **20**, 217-230 (1999).
- ⑨. T. Morishita. Fluctuation formulas in molecular-dynamics simulations with the weak

- coupling heat bath. *J. Chem. Phys.* **113**, 2976 (2000).
- ⑩. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola & J.R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690 (1984).
- ⑪. R.W. Pastor, B.R. Brooks & A. Szabo. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* **65**, 1409-1419 (1988).
- ⑫. R.J. Loncharich, B.R. Brooks & R.W. Pastor. Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-actylananyl-N'-methylamide. *Biopolymers* **32**, 523-535 (1992).
- ⑬. J.A. Izaguirre, D.P. Catarello, J.M. Wozniak & R.D. Skeel. Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **114**, 2090-2098 (2001).
- ⑭. J. Srinivasan, M.W. Trevathan, P. Beroza & D.A. Case. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **101**, 426-434 (1999).

主論文目録

本学位論文内容は下記の発表論文による。

1. Daisuke Katagiri, Takahiro Tsuchiya, Minoru Tsuda, Masayuki Hata, and Tyuji Hoshino

Computational Analysis of Stability of the β -Sheet Structure

Journal of Physical Chemistry B, 106, 9151-9158, 2002

2. Takahiro Tsuchiya, Daisuke Katagiri, Masayuki Hata, Tyuji Hoshino, Minoru Tsuda

Theoretical analysis of the stability of helices

Journal of Molecular Structure : Theochem, 589-590, 413-422, 2002

3. Daisuke Katagiri, Masayuki Hata, Takao Itoh, Sabro Neya, and Tyuji Hoshino

Atomic-Scale Mechanism of the GTP \rightarrow GDP Hydrolysis Reaction by the G α 1 Protein

Journal of Physical Chemistry B, 107, 3278-3283, 2003

本学位論文の審査は千葉大学大学院薬学研究院で指名された下記の審査委員により行われた。

主査 千葉大学教授（薬学研究院）薬学博士 五十嵐一衛

副査 千葉大学教授（薬学研究院）薬学博士 小林弘

副査 千葉大学教授（薬学研究院）薬学博士 戸井田敏彦

生体分子シミュレーションに用いる新規力場パラメータの開発と タンパク質立体構造予測への応用

ゲノム機能学講座 (薬品物理化学研究室) 片桐 大輔

[はじめに] 近年、コンピュータ技術の著しい発展により、これまで計算機シミュレーションが困難であった巨大タンパク質や複雑な生命現象について、簡便に且つ高速でシミュレーションを行うことが可能になってきた。これにともない、生命科学分野における計算科学の役割は重要なものとなってきている。特に創薬分野において、薬の直接のターゲットとなる受容体の立体構造を生化学的実験により決定することは容易では無いために、計算化学的立体構造予測に期待が寄せられている。特に Molecular Dynamics (MD) シミュレーションを用いたタンパク質立体構造予測は、配列相同性に依存することなく構造予測を行うことが可能であり大変期待されている方法である。さらに、配列相同性を用いる立体構造予測に対しても、MD シミュレーションは予測構造の最終的な構造最適化計算に使用することが可能なため、配列相同性を用いた立体構造予測を行っている研究者からも注目されている。しかし、コンピュータの計算速度が飛躍的に進歩する一方で、MD シミュレーション精度向上のための技術は、未だ成熟していないのが現状である。それゆえ、タンパク質立体構造予測を含めた多くの生体分子シミュレーションにおいて、実験的結果との相関が必ずしも高いとは言えず、深刻な問題となっている。

そこで本研究では、タンパク質立体構造予測に関する様々なアプローチを通して、タンパク質を対象とした MD シミュレーションの精度向上に取り組んだ。さらに、精度が向上した MD シミュレーションを再度タンパク質立体構造予測へ応用し、シミュレーション精度を再評価した。具体的には、これまでの「量子化学計算を用いたタンパク質局所的二次構造の安定性解析」の研究^{1, 2)}で得られた知見をもとに、(1) 既存力場パラメータを用いた MD シミュレーションによるタンパク質立体構造予測と、既存力場パラメータの問題点の提示 (2) 量子化学計算を反映した新規高精度力場パラメータの開発、(3) 新規高精度力場パラメータを用いた MD シミュレーションによるタンパク質立体構造予測、を遂行した。

[既存力場パラメータ"ff03 force field"を用いた MD シミュレーションによるタンパク質立体構造予測] 既存力場パラメータを用いた MD シミュレーションによるタンパク質立体構造予測を行った。ソフトウェアは AMBER8 を使用し、力場パラメータには最新の力場パラメータとして提供されている ff03 force field を用いた。予測対象タンパク質には、PDB code : 1LE3、1L2Y、1VII、1PGB で示される 4 つのタンパク質を用いた。初期構造は直鎖構造として MD シミュレーションを実行した。その結果 1L2Y、1VII のように主に helix

構造から成るタンパク質については、比較的精度良く構造予測が実行されたが、1LE3 と 1PGB の β -sheet 構造をとるべき領域で、helix 構造を予測してしまい構造予測は失敗した。

[既存力場パラメータの問題点と新規高精度力場パラメータの作成] 過剰に helix 構造を予測する理由を精査するため、ACE-XXX-NME のジペプチドモデルを用いて (XXX は 20 種各々のアミノ酸) MD シミュレーションを行い、ペプチドコンフォメーションの安定性解析を行った。helix 構造を形成する α R コンフォメーション (α R) と β -sheet 構造を形成する C5 コンフォメーション (C5) の双方の存在比から energy 比較を行った。その結果、

多くのアミノ酸で α R の方が安定となった (図 1-斜線)。

helix 構造は一般的に 4 残基ごとの主鎖が水素結合を形成することで安定化する。しかし、水素結合系の無いジペプチドモデルで α R が安定化してしまうことは、ff03 force field が過剰に helix 構造を安定化するパラメータであることを示している。Gaussian03 プログラムによる量子化学計算を α R、

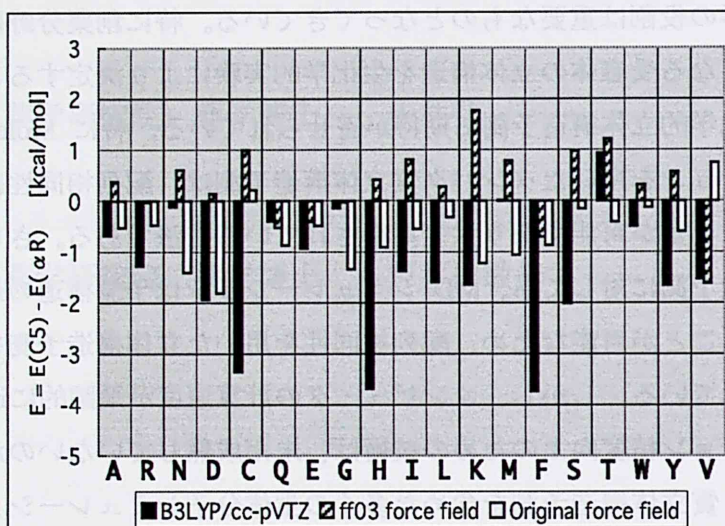


図 1 各アミノ酸コンフォメーションの安定性

C5 の双方について実行し、potential energy を比較した結果を図 1-黒に示す。量子化学計算では THR を除く全てのアミノ酸で C5 の方が安定となった。これにより、ff03 force field によるジペプチドモデルの MD シミュレーションは量子化学計算結果と相関が無いことが証明された。

そこで (3) において、力場パラメータの重要な要素の一つである電荷について、水の溶媒効果を加味したアミノ酸電荷を計算し、新たに力場パラメータを得た。その結果、CYS、THR、VAL 以外で量子化学計算結果と同様に C5 が安定化し、力場パラメータの精度は大きく改善した (図 1-白)。

[新規高精度力場パラメータを用いた MD シミュレーションによるタンパク質立体構造予測] 残基数の小さなタンパク質の立体構造を生体分子シミュレーションにより予測した。真に精度の高い力場パラメータであれば、エネルギー的に優位な立体構造を再現できるはずである。PDB code : 1LE3、1L2Y で示される 2 つのタンパク質について、配列情報のみから MD シミュレーションを使って立体構造予測を行った。その結果、1LE3 について、PDB 構造 (図 2a) と予測構造 (図 3a) の主鎖 C α 原子の RMSD は 3.87 Å であった。完全な sheet

構造を予測することはできなかつたが非常に近い予測構造を得た。既存力場パラメータの問題点であった helix 構造の過剰な安定化は改善されたと考えられる。今回新たに開発した力場パラメータは CYS、THR、VAL の 3 つのアミノ酸でコンフォメーション安定性の問題点が残っている。1LE3 には THR が 17 残基中 5 残基も存在しており、THR コンフォメーション安定性の修正が 1LE3 の構造予測の精度向上に必要であると考えられる。一方 1L2Y について、PDB 構造 (図 2b) と予測構造 (図 3b) の主鎖 C α 原子の RMSD は 2.32 Å であった。2 つの helix 領域と 1 つの strand 領域が高精度で予測に成功した。よって本研究

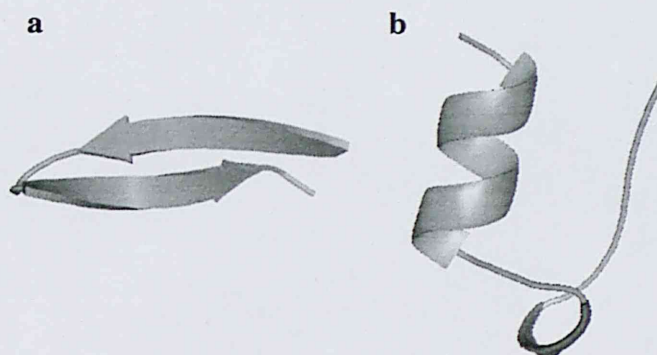


図 2 1LE3 の PDB 構造 (a) と 1L2Y の PDB 構造 (b)

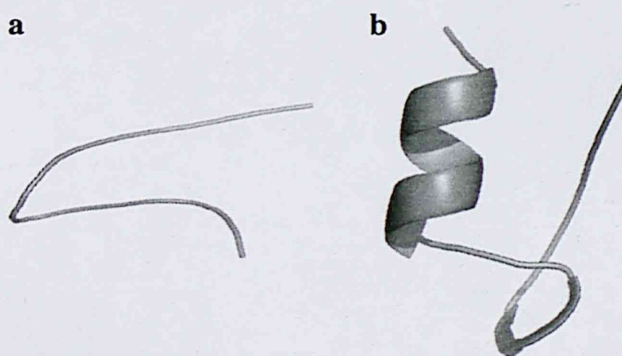


図 3 1LE3 の予測構造 (a) と 1L2Y の予測構造 (b)

で新規に開発した力場パラメータは β -sheet 構造だけでなく、helix 構造においても正当な評価ができる大変精度の高い力場パラメータであることが示された。

本研究では MD シミュレーションの根本の問題である力場パラメータの高精度化に成功した。よってタンパク質立体構造予測だけでなく、MD シミュレーションを用いる全ての生命科学の高精度化が達成されるものと強く確信している。

[参考文献] (1) Daisuke Katagiri, Takahiro Tsuchiya, Minoru Tsuda, Masayuki Hata, and Tyuji Hoshino : **Computational Analysis of Stability of the β -Sheet Structure** : *Journal of Physical Chemistry B*, 106, 9151-9158, 2002 (2) Takahiro Tsuchiya, Daisuke Katagiri, Masayuki Hata, Tyuji Hoshino, Minoru Tsuda : **Theoretical analysis of the stability of helices** : *Journal of Molecular Structure : Theochem*, 589-590, 413-422, 2002 (3) Daisuke Katagiri, Masayuki Hata, Takao Itoh, Sabro Neya, and Tyuji Hoshino : **Atomic-Scale Mechanism of the GTP \rightarrow GDP Hydrolysis Reaction by the Gi α 1 Protein** : *Journal of Physical Chemistry B*, 107, 3278-3283, 2003