

ブラウン動力学法によるタンパク質-リガンド複合体予測構造の評価. インフルエンザノイラミニダーゼと阻害剤の複合体のシミュレーション

小田 彰史,^{*,a,b} 小林 佳奈,^a 高橋 央宜,^a 山乙 教之,^c 広野 修一^c

^a 東北薬科大学薬学部, ^b 大阪大学蛋白質研究所, ^c 北里大学薬学部

Evaluation of Predicted Structures of Protein-Ligand Complex by Brownian Dynamics Simulation. Simulations of a Neuraminidase-Inhibitor Complex

Akifumi ODA,^{*,a,b} Kana KOBAYASHI,^a Ohgi TAKAHASHI,^a Noriyuki YAMAOTSU,^c
and Shuichi HIRONO^c

(Received November 20, 2009)

The prediction of protein-ligand complex structures plays an important role in structure based drug design. In this study, evaluations of predicted protein-ligand structures by using Brownian dynamics simulations were carried out. The complex structures were generated by computational docking trials, which were frequently used in actual drug design. From the obtained complex structures, "correct answer" which reproduced the experimental structure and "incorrect answers" which were different from crystal structure were selected, and simulations of them were carried out. The complex between influenza A virus neuraminidase and its inhibitor was used as test set for the test calculations. For "incorrect answers", the 3D structures were broken by the simulations, and the results seem to be caused by instabilities of the predicted structures.

Key words — protein-ligand complex; Brownian dynamics; computational docking

タンパク質の立体構造を利用した創薬 (Structure-Based Drug Design, SBDD) において, タンパク質-リガンド複合体構造を計算機的に予測することは非常に重要である.¹⁾ 予測された構造に基づいて論理的に化合物の最適化を行ったり, あるいは予測複合体構造を参考にしてコンピュータ上で化合物の絞り込みを行うヴァーチャルスクリーニングを実行したりすることになる. タンパク質-リガンド複合体構造をコンピュータを用いた計算で予測する方法を計算機によるドッキング (あるいは単にドッキング) と呼ぶが, これは複合体構造を構築するドッキング過程と, 構築した複合体構造を評価するスコアリング過程の2つの過程から構成されている. ドッキング過程ではタンパク質・リガンド双方の物理化学的・幾何学的性質を考慮し, それらの組み合わせによって複合体構造を構築する. したがってドッキング過程はしばしば組み合わせ最適化問題として扱われることがあり, 種々の数学的・情報科学的手法を用いて解かれる. 組み合わせ最適化問題では, 解候補の数が爆発的に増加する「組み合わせ爆発」と呼ばれる困難があるものの, これに対する最適解探索手法が多く開発されており, 現在ドッ

キング過程についてはある程度の成功率 (ソフトウェアによっては70%近く) での実行が可能となっている.²⁾

一方のスコアリング過程では, タンパク質-リガンド間の結合自由エネルギーを計算する. 自由エネルギーを正確に算出するためには統計力学的手法を用いる必要があるが, これにはきわめて長時間のシミュレーションが必要となるため,³⁾ 高速な計算が要求されるドッキングにおいては現実的ではない. そのため自由エネルギーを簡易的に算出するための近似関数が使用されている.¹⁾ この近似関数をスコア関数と呼ぶ. 今までにスコア関数は多数提案されているが, それらは分子力場を基にしたスコア関数, 経験的スコア関数, 知識を基にしたスコア関数の3種類に分類することができる. このように様々なアプローチでスコア関数が作成されているが, いずれも近似的に自由エネルギーを算出する関数であるため精度に限界があり, 現在のところ決定版となり得るスコア関数は存在しない.

計算機によるドッキングにおいて, スコア関数は3つの役割を果たしている.⁴⁾ 1つはドッキング過程の最中に構築中の複合体構造を評価する役割であ

り, これは最適化問題における目的関数として働いている. この場合, ドッキング過程はスコア値が最も低くなるような構造を探索する最適化問題となる. この役割においてはスコア関数はドッキング過程と密接に関連しているため, スコアリング過程ではなくドッキング過程の一部として述べられることも多い. 2つ目の役割は, ドッキング過程によって構築された複合体構造を評価することである. ドッキング過程においては一般に複数の複合体予測構造を出力する. これらのうちのいずれが現実の複合体構造と近いかを推測する作業がこの段階になる. 自然界においては, タンパク質とリガンドは最安定(あるいは準安定)の構造をとっていると考えられるため, 自由エネルギーを計算して最も安定な構造が現実に近い構造であると推定する. 3つ目の役割は, 複数の化合物に対してドッキングを行ったとき, どの化合物が最も医薬品となり得るかを評価することである. 薬物標的とリガンドのみに注目した場合, 標的に対して強く結合する化合物が医薬品となる可能性が高いと考えることができる. そこで薬物標的とリガンドとの結合自由エネルギーを計算し, それを化合物間で比較することにより, リガンドとして有望な化合物を抽出する. この作業はヴァーチャルスクリーニングと呼ばれる.

スコア関数の3つの役割はいずれも重要であり, それぞれに応じてスコア関数の評価および改良が行われている. それぞれの役割について, 様々な実験データを用いながら評価・改良が行われているものの, 2つ目の役割に関しては「現実の複合体構造に近い構造(正解構造)」と「遠い構造(不正解構造)」の両方の結合自由エネルギーを計算する必要があり, 前者については実験値が存在するものの後者については実験値がないという問題がある. これは, 後者がいわば「ドッキングに失敗した構造」であるため, 必然的に不安定な構造となり, 自然界には存在し得ないためである. その結果, 2つ目の役割についてはスコア関数の評価・開発を行うに際して限界があることになる.

そこで本研究では, 分子シミュレーション手法を用いて「正解構造」と「不正解構造」の両方について評価を行った. その結果を比較し, 両者に構造的な差異が現れているかどうかを判断した. 正解構造と不正解構造との間でシミュレーション結果に差が現れた場合, 分子シミュレーションが「2つ目の役割」に関してスコア関数の代替として使用できるこ

とを意味する. 分子シミュレーション手法としてはブラウン動力学(Brownian Dynamics, BD)法⁵⁾を用いた. 分子シミュレーションにおいては溶媒となる水分子の扱いに困難が生じることが多いが, BDでは溶媒効果を粘性とランダム力として取り込むことで溶媒分子を明示的に取り扱うことを避け, 計算コストの低減につなげている. BDはこれまでにタンパク質のフォールディングシミュレーション⁶⁾や電子伝達系の機構の解析⁷⁾などにおいて有意義な成果を上げている. BDのためのプログラムとして, UHBD⁸⁾や安藤らによるプログラム⁹⁾などいくつか開発されているが, 本研究では我々のグループが開発したBDプログラム *brownian*¹⁰⁻¹²⁾を使用した. *brownian*は分子動力学法(Molecular Dynamics, MD)プログラム AMBER¹³⁾と操作性を統一することでより容易にBD計算を導入することを可能としたプログラムである. トポロジーや座標ファイルの作成に AMBER の LEaP モジュールが使用できるのはもちろんのこと, 医薬品のシミュレーションで困難を生じることの多い非ペプチド性低分子化合物のためのファイルの作成に *antechamber*¹⁴⁾をそのまま使用できるといった利点を持っている. 本研究ではインフルエンザノイラミニダーゼとその阻害剤との複合体に対して *brownian* を用いて, 複合体構造評価に *brownian* が有効であるかどうかを判断した. ノイラミニダーゼはインフルエンザウイルス表面に存在するスパイクタンパク質の一種で, シアル酸を基質とする酵素である. 細胞表面に存在する糖鎖をシアル酸部分で切断することによってインフルエンザウイルスに感染した細胞からウイルスを遊離させ, ウイルスはほかの細胞へと感染していく. そのためノイラミニダーゼは抗インフルエンザウイルス薬の標的として注目されており, SBDDの対象としても重要視されている. 本研究ではこのインフルエンザノイラミニダーゼと阻害剤の複合体について *brownian* の能力を評価することにより, SBDDにおける *brownian* の有用性について検討した.

方 法

ブラウン運動を行う粒子では, ランジュバン方程式

$$m \frac{dv}{dt} = F - m\zeta v + F^B$$

に従って運動が決定されている. ここで m は粒子の質量, v は速度, t は時間, F は相互作用力, ζ は摩擦係数, F^B はランダム力である. ζ はストークスの法則から

$$\zeta = 6\pi\eta a / m$$

である. ここで η は粘度, a は粒子半径である. MD 同様 BD でも時間経過を離散化する必要があるが, *brownian* では粒子の位置および速度を Ermak および Buckholz の式¹⁵⁾

$$r(t+h) = r(t) + \frac{1}{\zeta} v(t) (1 - e^{-\zeta h}) + \frac{1}{m\zeta} F(t) \left\{ h - \frac{1}{\zeta} (1 - e^{-\zeta h}) \right\} + r^B(t+h)$$

$$v(t+h) = v(t) e^{-\zeta h} + \frac{1}{m\zeta} F(t) (1 - e^{-\zeta h}) + \frac{1}{m\zeta} (F(t+h) - F(t)) \left\{ h - \frac{1}{\zeta} (1 - e^{-\zeta h}) \right\} + v^B(t+h)$$

で決定している. ここで r は位置, h はタイムステップ, r^B はランダム変位, v^B はランダム変速である.

本研究では, 立体構造既知のタンパク質-リガンド複合体を用いて BD シミュレーションを行った. 使用した複合体はインフルエンザ A N9 ノイラミニダーゼと 1-デオキシ-2, 3-デヒドロ-N-アセチルノイラミン酸との複合体である. 立体構造については RCSB protein databank¹⁶⁾ より入手した (PDB ID: 1NNB). この複合体について, タンパク質とリガンドを分離し, それぞれに水素を付加した. リガンドについては配座探索プログラム CAMDAS¹⁷⁾ によって配座探索を行い, AM1 法¹⁸⁾ で構造最適化した. これらの配座集団のうちで AM1 法によって計算されたエネルギーが最も低い配座をリガンドの初期構造とした. この初期構造と, ノイラミニダーゼの構造とを用いて, ドッキングプログラム GOLD¹⁹⁾ でドッキングを行った. GOLD の設定はすべてデフォルトである. ドッキングによって得られた結果については結晶構造との root mean square deviations (RMSD) を求めて評価したが, 本研究では結晶構造との RMSD が 2.0 Å 以下だった構造 1 つ (pose 5, RMSD = 1.998 Å) と RMSD が 2.0 Å 以上だった構造 2 つ (pose 7, RMSD = 4.572 Å および pose 10,

RMSD = 3.121 Å) を抽出した. この pose 5 を「正解構造」, pose 7 および 10 を「不正解構造」と定義する. なお, pose 5 はドッキングによって得られた 10 個のポーズ中最も結晶構造に近かったポーズであり, pose 7 は最も結晶構造に遠かったポーズである. これらの構造を用いて, ノイラミニダーゼと阻害剤との複合体の分子シミュレーションを行った.

計算に際しては, まず MD によって予備計算を行い, 得られた構造に対して BD シミュレーションによる本計算を行った. その本計算によって, 正解構造と不正解構造がシミュレーションを通じてどのように変化するかを観察し, 正解構造および不正解構造の安定性を評価した. すなわち, *brownian* が正解構造を壊さずにシミュレーションできるか, また不正解構造では何が起こるかを観察した.

まず予備計算では, 0 K から 300 K までの昇温シミュレーションを行った. シミュレーションは Hawkins らによる一般化 Born (Generalized Born, GB) 法²⁰⁾ による連続体溶媒存在下で行った. カットオフについては使用せず, 0.5 fs のタイムステップで計 30 ps かけて昇温を行った. 本計算の BD では, *brownian* の機能により自動で誘電率および粘度を計算し, その値を用いてシミュレーションを実行した. また, カットオフは 18.0 Å を使用した. 本計算では 1 fs のタイムステップで計 5 ns のシミュレーションを行った. また, 系の温度は 300 K である. すべての計算において, タンパク質部分については AMBER の ff99 力場²¹⁾ を, リガンド部分については GAFF 力場²²⁾ を使用した. また BD によるトラジェクトリの解析については, ノイラミニダーゼの主鎖を重ね合わせた場合の主鎖のみの RMSD と, 主鎖を重ね合わせた状態でのリガンドの RMSD とを算出した. 前者はシミュレーションを通じてノイラミニダーゼの配座がどの程度変化したかを示し, 後者はノイラミニダーゼの内部でリガンドの配座および配置がどの程度変化したかを示している. RMSD 計算の参照構造としては昇温 MD を行う前の構造を用いた.

結果および考察

Fig. 1 に, シミュレーションを通じての主鎖の RMSD を示す. ここに示したように, 5 ns のシミュレーションが終了した時点で pose 5 および pose 7 はおおよそ構造が収束していると考えられる. また

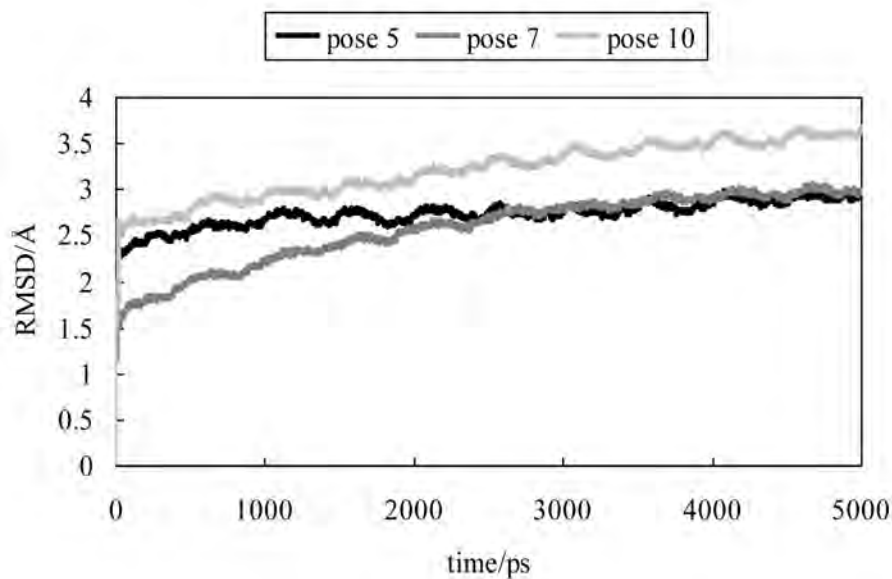


Fig. 1. RMSDs of main chain

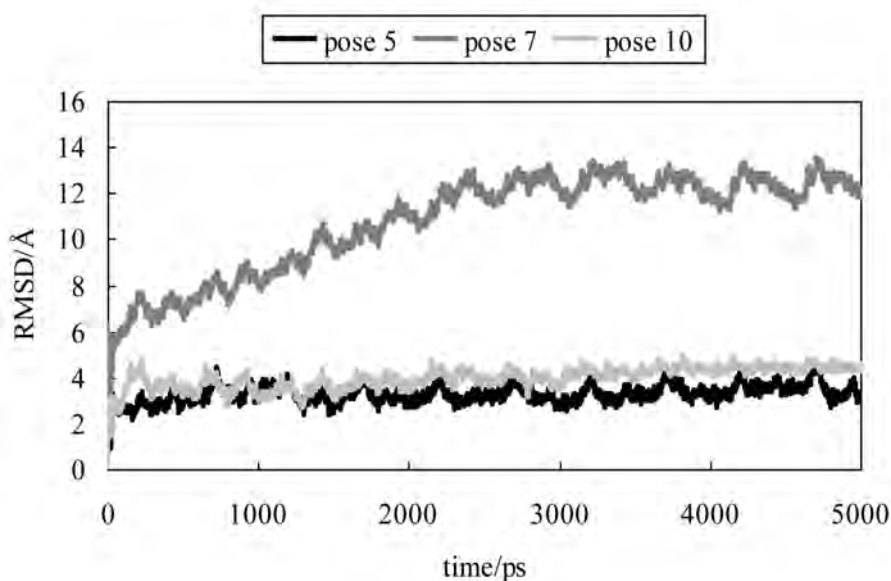


Fig. 2. RMSDs of the ligand

RMSD についても 3.0 Å 以下と、比較的小さな値に収まっている。一方で pose 10 については、5 ns のシミュレーションの終了時点でも RMSD が増加し続けており、値も 3.5 Å 以上と構造が大きく変化し続けていることが示唆されている。この pose 10 は不正解構造であることから、不適切な位置にドッキングしたりリガンドのためにタンパク質の構造が影響を受け、構造が崩れているのではないかと考えられる。また、構造が収束していると考えられる pose 5 と pose 7 についても詳しく見てみると、pose 5 では 1 ns 程度で RMSD の変化が収まっているのに対

して pose 7 では 4 ns 付近まで構造変化が続いていた。これは不正解構造であった pose 7 ではその構造が不適切であるために 4 ns 付近まで構造が変化し、反対に pose 5 はドッキングの段階である程度妥当な構造が得られていたためにそれほど長時間のシミュレーションを必要としなかったのではないかと考えられる。このように、ドッキング結果が実験構造に近いか遠いかによって主鎖の RMSD に変化が生じており、BD シミュレーションによってドッキング結果の妥当性を評価できるのではないかと考えられる。

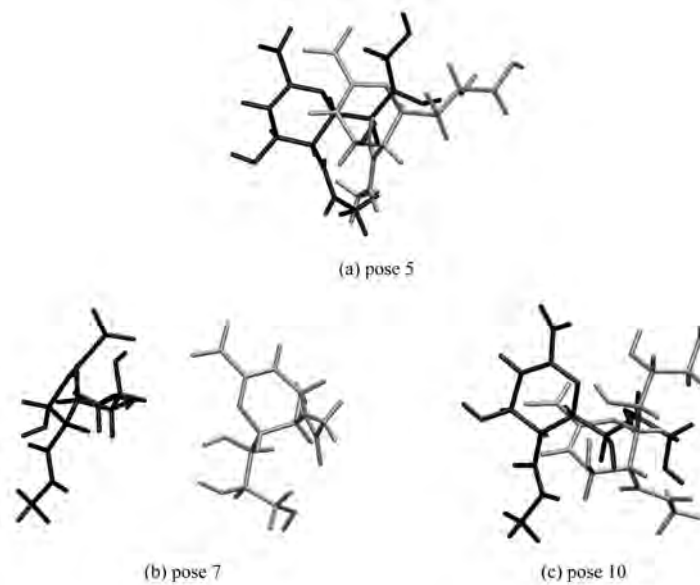


Fig. 3. Ligand structures

Structures in 5 ns are illustrated in black, and initial structures are displayed in gray.

次に, Fig. 2 にシミュレーションを通じてのリガンドの RMSD を示す. ここに示したように, 正解構造である pose 5 では RMSD が 3.0 \AA 付近で収束していた. また構造変化自体 1 ns に到達する以前に収束していた. これは pose 5 が正解構造であること, すなわち自然界における (準) 安定状態と対応すると考えられることと無矛盾であり, やはり BD シミュレーションが正解構造を正しく「正解」として扱うことができることを示している. 一方で不正解構造のうち pose 10 は, pose 5 よりも多少大きな値ではあるものの RMSD が収束していた. この系は主鎖の RMSD が大きく変化していたため, ドッキングによって得られた不安定な構造の影響がリガンド自身にはそれほど及ばず, 主に主鎖に影響を与えていることがわかる. それに対してもう一方の不正解構造である pose 7 ではリガンドの RMSD が 10 \AA 以上と大きくなっており, ノイラミニダーゼ中でリガンドが大きく動いていることがわかる. この pose 7 は pose 10 とは対照的に主鎖の RMSD が正解構造に近かったが, Fig. 2 に示した通りリガンドの RMSD が非常に大きくなっており, やはり不正解構造では BD シミュレーションによって構造が壊れることがわかる. この結果もまた, ドッキングによって得られた複合体構造が妥当であるか否か (すなわち安定構造であるか否か) が BD シミュレーションに影響を与えることを示唆している. このように, 正解構造においてはタンパク質主鎖およ

びリガンドの両方がシミュレーションを通じて安定に構造を保っているのに対して, 不正解構造ではタンパク質あるいはリガンドのどちらかの構造が大きく変動している. そのため, 主鎖・リガンド双方の RMSD を見ることでドッキング結果の評価ができるのではないかと考えられる.

最後に, Fig. 3 に 5 ns 時点での pose 5, pose 7, pose 10 の構造を示す. 参照のため昇温 MD の初期構造についても同時に示している. 5 ns 時点の構造を黒, 初期構造を灰色で表示している. ここに見られるように, pose 7 では配座はおろか配置が大きく動いており, ドッキングで得られた初期構造が不安定であることが示唆されている. また pose 10 についても pose 7 ほどは大きく動いていないものの, 環の向きが上下逆になっているなど, 配座・配置ともに変化している. それに対して正解構造の pose 5 では配座・配置ともに初期構造からそれほどは変化しておらず, 図の左方向に多少平行移動する程度に留まっている. これはドッキングの結果得られた構造が, pose 5 についてはある程度安定であったことを示している.

今回対象としたノイラミニダーゼと阻害剤の系では, 正解構造は BD シミュレーションによって壊れることがなく, 不正解構造はシミュレーションによって大きく変化した. このような場合は構造を確認するだけで正解・不正解を判別できる可能性があるが, 場合によっては不正解構造であってもシ

ミュレーションで構造が壊れないことが考えられる。そのような場合にはシミュレーションの結果を生かしてタンパク質-リガンド間の結合自由エネルギーを計算し、それによって正解・不正解を判別する必要があるかもしれない。今後分子シミュレーション, 特にBDシミュレーションがそのような用途に使用することが可能であるかどうかを検討する予定である。

謝辞

本研究は大阪大学蛋白質研究所共同研究員制度を利用して行われた。

REFERENCES

- 1) Kitchen D. B., Decornez H., Furr J. R., Bajorath J., *Nat. Rev. Drug Discov.*, **3**, 935–949 (2004).
- 2) Oda A., Okayasu M., Kamiyama Y., Yoshida T., Takahashi O., Matsuzaki H., *Bull. Chem. Soc. Jpn.*, **80**, 1920–1925 (2007).
- 3) 肥後順一, 中村春木, “タンパク質計算科学,” 第6章, 神谷成敏, 肥後順一, 福西快文, 中村春木著, 共立出版, 東京, 2009, pp. 156–209.
- 4) Oda A., Tsuchida K., Takakura T., Yamaotsu N., Hirono S., *J. Chem. Inf. Model.*, **46**, 380–391 (2006).
- 5) 神山新一, 佐藤 明, “流体ミクロ・シミュレーション,” 朝倉書店, 東京, 1997.
- 6) Rojnuckarin A., Kim S., Subramaniam S., *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 4288–4292 (1998).
- 7) Lin J., Beratan D. N., *J. Phys. Chem. B*, **109**, 7529–7534 (2005).
- 8) Davis M. E., Madura J. D., Luty B. A., McCammon J. A., *Comp. Phys. Comm.*, **62**, 187–197 (1991).
- 9) Ando T., Meguro T., Yamato I., *J. Comput. Chem.*, *Jpn.*, **1**, 227–235 (2002).
- 10) Yamaotsu N., Hirono S., *CBI2003 Proceedings*, 184–185 (2003).
- 11) Yamaotsu N., Hirono S., *SAR News*, **6**, 18–20 (2004).
- 12) Oda A., Yamaotsu N., Hirono S., Takahashi O., *Biol. Pharm. Bull.*, **31**, 2182–2186 (2008).
- 13) Case D. A., Darden T. A., Cheatham T. E., Simmerling C. L., Wang J., Duke R. E., Luo R., Merz K. M., Pearlman D. A., Crowley M., Walker R. C., Zhang W., Wang B., Hayik S., Roitberg A., Seabra G., Wong K. F., Paesani F., Wu X., Brozell S., Tsui V., Gohlke H., Yang L., Tan C., Mongan J., Hornak V., Cui G., Beroza P., Mathews D. H., Schafmeister C., Ross W. S., Kollman P. A., *AMBER9*, University of California, San Francisco, 2006.
- 14) Wang J., Wang W., Kollman P. A., Case D. A., *J. Mol. Graphics Model.*, **25**, 247–260 (2006).
- 15) Ermak D. L., Buckholz H., *J. Comput. Phys.*, **35**, 169–182 (1980).
- 16) Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne P. E., *Nucleic Acids Research*, **28**, 235–242 (2000).
- 17) Tsujishita H., Hirono S., *J. Comput. Aided Mol. Des.*, **11**, 305–315 (1997).
- 18) Dewar M. J. S., Zoebisch E. G., Healy E. F., Stewart J. J. P., *J. Am. Chem. Soc.*, **107**, 3902–3909 (1985).
- 19) Jones G., Willett P., Glen R. C., *J. Mol. Biol.*, **245**, 43–53 (1995).
- 20) Hawkins G. D., Cramer C. J., Truhlar D. G., *Chem. Phys. Lett.*, **246**, 122–129 (1995).
- 21) Wang J., Cieplak P., Kollman P. A., *J. Comput. Chem.*, **21**, 1049–1074 (2000).
- 22) Wang J., Wolf R. M., Caldwell J. W., Kollman P. A., Case D. A., *J. Comput. Chem.*, **25**, 1157–1174 (2004).