

創薬のための仮想スクリーニング統合環境 Xsi の開発

稲垣 祐一郎

みずほ情報総研(株), 〒101-8443 東京都千代田区神田錦町 2 - 3

e-mail: yuichiro.inagaki@gene.mizuho-ir.co.jp

(Received: July 12, 2005; Accepted for publication: November 9, 2005; Published on Web: December 15, 2005)

分子力場法をベースとした薬物候補化合物の仮想スクリーニングのための統合環境 Xsi を開発した。創薬の手法として、ターゲット蛋白質の構造情報を用いない Ligand Based Drug Design(LBDD) と、ターゲット蛋白質の構造情報を用いる Structure Based Drug Design(SBDD) の二通りがある。Xsi はこの両手法をひとつのプラットフォーム上に一体化させ、専用のスクリプト言語 Xsi-Script によりユーザーが機能を自由に組み合わせることにより様々な解析フローを実現できる統合環境となっている。

LBDD の中心的機能の低分子の配座生成部分について、SH4 版 EHPC システムを用いて、階層型並列化を行い、CPU21 個で並列化効率 96.6% が得られた。

更に、SBDD 機能を用いて、c-ABL とリガンドの再ドッキング試験を行い、Root Mean Square Deviation 0.13 が得られている。

キーワード: 分子力場法, 専用計算機, Ligand based drug design, Structure based drug design

1 はじめに

High Throughput Screening(HTS)、コンビナトリアルケミストリーなどの技術の進展に伴い、製薬会社ではターゲット蛋白質ひとつに対して数百万化合物のスクリーニングが行われるようになってきている。これに対応して、X 線結晶解析等により得られる蛋白質立体構造情報の増大、計算機の計算速度の向上を背景として高精度かつ高速度の仮想スクリーニングの技術も発展して来ている。実験によるスクリーニングにおいても偽陰性、偽陽性は避けられず、コストもかかるため、実験と計算を適切に組み合わせ、ヒット化合物からリード化合物、さらにはリード最適化を行うことが求められている。

創薬における計算手法にはターゲット蛋白質の構造情報を用いない Ligand Based Drug Design(LBDD) と、ターゲット蛋白質の構造情報を用いる Structure Based Drug Design(SBDD) の二通りがある。従来この二つの手法は蛋白質立体構造情報の有無の条件に応じてそれぞれ独立に行われることが多かった。しかし、蛋白質

立体構造の情報とある程度の活性を持つ化合物の情報との双方が得られる場合も多い。このような場合に入手できる情報をフルに活用し、計算の精度を上げるため、LBDD の機能と SBDD の機能を一体化した創薬のための統合環境 Xsi を開発して来た [1–3]。

さらに、LBDD の中心的機能のひとつである低分子の配座生成を高速に実行するため、EHPC-SH4 システム(汎用 CPU である SH4 を用いた EHPC のプロトタイプシステム)を用いて並列化を行い、高い並列化効率を得られている。

本稿では、Xsi の概要と EHPC-SH4 システムによる並列化、および SBDD の計算例として c-ABL のドッキング構造の再現試験を行ったので報告する。

2 仮想スクリーニング統合環境 Xsi

Xsi は、その機能のすべてが専用のスクリプト言語 Xsi-Script により制御されるようになっている。Xsi-Script にはオブジェクト指向的なクラス、メソッド、イ

インスタンスといった概念があり、ユーザーは例えば分子をあらわす Molecule クラス、ドッキング計算を行う Docking クラスのインスタンスを生成し、インスタンスに対してメソッドを実行させることで所望の機能を実現する。任意のクラスの任意次元の配列、フロー制御 (if 文、while 文)、四則演算、論理演算などの機能も備えている (但しユーザーが自分でクラスを定義することは出来ない)。

Xsi の主なクラスと機能を Table 1 に示す。このうち Clique クラスについて説明を補足する。クリーク (Clique) とは、グラフにおける頂点の集合であって、

属するメンバー同士が全て辺により直接に連結しているものを指す。また、準クリークとは、同じく頂点の集合であって、属するメンバー同士が全て辺により直接に連結している割合が一定の値より高いものを指す。Clique クラスは、与えられたグラフ構造に対して、準クリークを確率的に探索する機能を実現している。この機能を用いることによって、大量の分子の配座に対して、何らかの類似性指標が任意の閾値以上のものを辺で結んで出来るグラフ構造の中で準クリークを見つけることによって、複数分子の重ね合わせ構造を推定することが出来る。

Table 1. Functions implemented in Xsi ver.2.1

クラス名	機能
Molecule	分子を表現
Universe	複合体を表現
Constraints	力場計算時の分子に対する拘束を表現
UniverseConstraints	力場計算時の複合体に対する拘束を表現
MoleculeFileMol	MDL Mol ファイル [4, 5] の入出力
MoleculeSetFilePDB	PDB ファイル [6] の入出力
MoleculeSetFileSD	MDL SD ファイル [4, 5] の入出力
MolecularMechanics	MMFF 力場 [7-13] による分子の配座生成、構造最適化。距離依存誘電率、GB/SA モデルによる溶媒効果を考慮可能
MonteCarlo	モンテカルロ法による分子の配座生成
Docking	力場による複合体の構造最適化 (蛋白質側を任意の範囲でフレキシブルに設定可能)、リガンドのシミュレーテッドアニーリング
RMSMinimizer	二つの分子の Root Mean Square 計算
ClusterAnalyzer	Ward 法による任意のベクトルセットのクラスタリング
Descriptor	分子の構造記述子計算。WHIM[14]、極性接触表面積など
Similarity	二つのベクトルの Euclid 距離、Cosine 係数、Tanimoto 係数の計算
Clique	グラフ構造中の quasi clique 同定。
Statistics	線形重回帰計算、線形判別分析
SVM	サポートベクターマシンによる機械学習
CSVFile	CSV ファイルの入出力
RandomNumberGenerator	乱数生成
LigandAlignment	格子上にマップされた物理化学的性質 (形状、水素結合、疎水性など) に対する重ね合わせ探索
MolecularOrbital	MOPAC[15] とのインターフェース
Field	任意の場を表現し保存
PotentialField	ポテンシャルの場を表現し保存

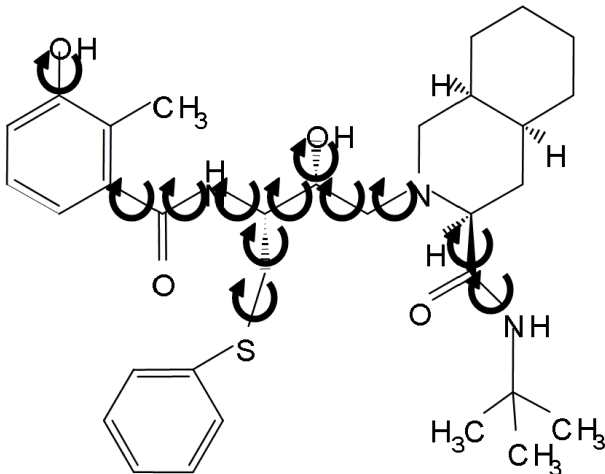


Figure 1. Viracept

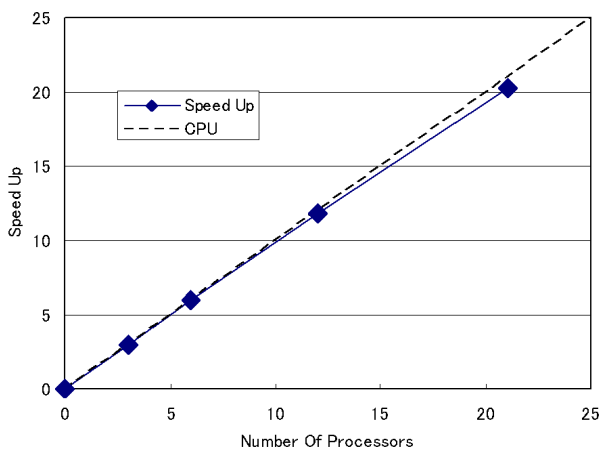


Figure 2. Speed up of conformation search on EHPC-SH4 system

3 EHPC-SH4システムでの並列化

Xsi 機能のうち、MolecularMechanics クラスによる網羅的配座探索機能を EHPC-SH4 システムに移植し、HIV プロテアーゼ阻害薬である Viracept[16] (Figure 1) の網羅的配座探索を行いパフォーマンスを測定した結果を Figure 2 に示す (計算の詳細については既報 [3])。Figure 1 に示す Rotatable な結合に関して、各々 2 ~ 3 個の二面角を設定して生成される配座の中からランダムに初期配座 2048 を選択した。それら初期構造に関して、衝突解析 (原子の衝突が激しいものを除外する) を行い、1000 配座を得た。この配座を各 CPU に割り振り、構造最適化を行った。

Figure 2 が示すように、CPU 数と Speed up は 21CPU

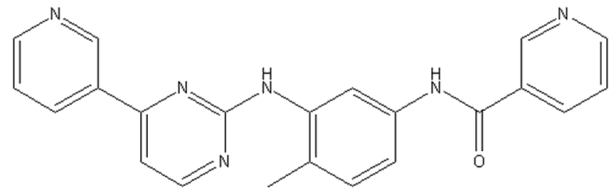


Figure 3. 1FPU のリガンド

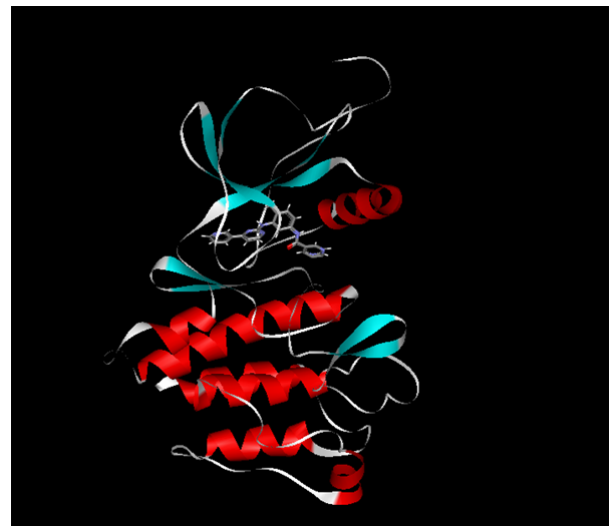


Figure 4. PDB (1 FPU) 複合体のリガンド構造 (stick 表示) と計算結果 (line 表示)

まで良好な線形性を示している (並列化効率 96.6%)

4 c-ABL との再ドッキング試験

SBDD に関する Xsi の基本的性能評価を行うため、リガンド (Figure 3) とそのターゲット蛋白である c-ABL との複合体構造 (PDB コード 1FPU、Figure 4) の再現試験を行った。

計算手順を以下に示す。

(1) 蛋白質ドッキング初期構造の設定

蛋白質の PDB 構造に対し水素付加を行い、水素の位置のみをフレキシブルに設定して構造最適化を行う。

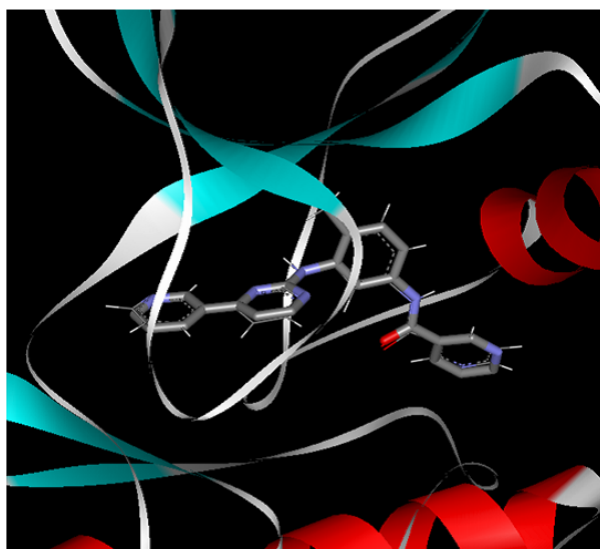


Figure 5. PDB (1 FPU) 複合体のリガンド構造 (stick 表示) と計算結果 (line 表示) (拡大図)

(2) リガンド初期構造の設定

リガンド構造が複合体構造に依存しないように、別途 ChemDraw によりリガンド構造を生成し、構造最適化を行う。

(3) リガンド配座生成

(2) で得られたリガンドの配座をモンテカルロ法により 100 配座程度発生する。

(4) リガンドアラインメント

蛋白質とリガンドに関して、形状の相補性指標を最適化を行う。ここで用いる相補性指標には、結合サイト周辺に間隔 0.5 で生成した格子点上に、ガウス分布 (原子中心位置を中心とし、van der Waals 半径を分散値とする) の値を設定した格子データと、リガンド側から同様に生成される格子データとをそれぞれ多次元ベクトル化したもののユークリッド距離を用いた。

この指標を用いて、(3) で求めた配座各々に関して重心の周りにランダムに 10 通り回転したものをそれぞれ分子全体の並進と回転の自由度に関して最適化する。

(5) RMSD によるクラスタリング

以降のステップの計算量を減らすため、(4) の結果に関してリガンド構造間の RMSD (Root Mean Square Deviation) によりクラスタリングを行い、各クラスターから代表 100 配座を得る。代表配座は、各クラスターの重心 (全ての原子座標を並べた多次元ベクトルの分布の重心) に最も近いものを選択した。

(6) ドッキング

(5) の結果それぞれに対してドッキング計算を行い、全系のエネルギーに関してリガンド配座のソーティングを行う。

以上の計算により、ドッキングの結果 100 配座のうち、全系のエネルギーが最低のものが、PDB の complex 構造におけるリガンド構造に最も近く、PDB 構造におけるリガンド構造と計算結果の間の RMSD が 0.13 であった (Figure 4、Figure 5)。

5 まとめ

コンピュータを用いた薬物設計がコンピュータ化学の重要な応用分野のひとつであることは言を待たない。使用できる蛋白質立体構造が増えて今後の計算精度の向上、PC クラスタやグリッド計算、本研究で用いたような専用計算機による計算速度向上も期待でき、近い将来には創薬に必要な不可欠なツールのひとつとなると信じる。

MMFF 力場はこれまでドッキング計算に使用されることは少なかったが、c-ABL の計算例により、ドッキングに使用しても何ら精度的に問題がない可能性が示された。

今後さらに Xsi の性能向上のため、リガンド依存の水分子挿入、Induced fit への対応、スコア関数の高精度化等を行っていく予定である。

なお、本研究の一部は科学技術振興調整費 総合研究「科学技術計算専用ロジック組込み型プラットフォーム・アーキテクチャに関する研究 (代表 村上和彰 九州大学教授) によるものである。また、Xsi の開発に多大な支援をいただいている徳島大学薬学部中馬寛教授、住友製薬山崎一人氏に深く感謝する。

参考文献

- [1] <http://www.mizuho-ir.co.jp/science/xsi>
- [2] 山崎一人、金岡昌治、住友化学, 2005-I, 52-62 (2005).
- [3] 浜田道昭、馮誠、稲垣祐一郎、長嶋雲兵、村上和彰、中馬寛, 日本応用数理学会論文誌, 14, 267-288 (2004).

- [4] Arthur Dalby et al., *J. Chem. Inf. Comput. Sci*, **32**, 244-255 (1992).
- [5] http://www.mdl.com/solutions/white_papers/ctfile_formats.jsp
- [6] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank, *Nucleic Acids Research*, **28**, 235-242 (2000).
- [7] Halgren, T. A., Merck Molecular Force Field. I. Bases, Form, Scope, Parameterization, and Performance of MMFF94, *J. Comp. Chem.*, **17**, 490-519 (1996).
- [8] Halgren, T. A., Merck Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions, *J. Comp. Chem.*, **17**, 520-552 (1996).
- [9] Halgren, T. A., Merck Molecular Force Field. III. Molecular Geometries and Vibrational Frequencies for MMFF94, *J. Comp. Chem.*, **17**, 553-586 (1996).
- [10] Halgren, T. A., Merck Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, and Empirical Rules, *J. Comp. Chem.*, **17**, 616-641 (1996).
- [11] Halgren, T. A., Merck Molecular Force Field. VI. MMFF94s Option for Energy Minimization Studies, *J. Comp. Chem.*, **20**, 720-729 (1996).
- [12] Halgren, T. A., Merck Molecular Force Field. VII. Characterization of MMFF94, MMFF94s, and Other Widely Available Force Fields for Conformational Energies and for Intermolecular-Interaction Energies and Geometries, *J. Comp. Chem.*, **20**, 730-748 (1996).
- [13] Halgren, T. A. and Nachbar, R. B., Merck Molecular Force Field. IV. Conformational Energies and Geometries for MMFF94, *J. Comp. Chem.*, **17**, 587-615 (1996).
- [14] Todeschin, R. and Gramatica, P., New 3D Molecular Descriptors: The WHIM theory and QSAR Applications, *In 3D QSAR in Drug Design Volume 2*, ed. by Kubinyi, H., Folkers, G and Martin, Y. C., KLUWER/ESCOM, Dordrecht (1998), pp. 355-380.
- [15] Deware, M. J. S., Zoebitch, E. G., Healy, E. F., and Stewart, J. J. P., *J. Am. Chem. Soc.*, **107**, 3902 (1985).
- [16] Stephen, W. K., Vincent, J. K., Jay, F. D., Bhasker, V. S., James, E. F., Krzysztof, A., Jeffrey, A. B., Kristina, M. C., Nickolay, Y. C., David, K. C., Bruce, A. D., Steven, D. H., Deborah, A. K., Maha, B. K., Penny, P. L., Mark, A. M., Amy, K. P., Siegfried, H. R., Kenneth, S. S. and Jonm, H. T., Viracept (Nelfinavir Mesylate, AG1343): A Potent, Orally Bioavailable Inhibitor of HIV-1 Protease, *J. Med. Chem.*, **40**, 3979-3985 (1997).

Development of Xsi, An Integrated Environment for Virtual Screening and Drug Design

Yuichiro INAGAKI

Mizuho Research and Information, Inc.
2-3 Kanda-Nishikicho, Chiyodaku, Tokyo 101-8443, Japan
e-mail: yuichiro.inagaki@gene.mizuho-ir.co.jp

We developed an integrated environment for virtual screening, Xsi, based on the force field method. Two methodologies, Ligand Based Drug Design(LBDD) and Structure Based Drug Design(SBDD) are both implemented on the same platform and users can realize various analysis flows by specially designed Xsi-Script.

Calculation for conformation search on small molecules is parallelized on EHPC-SH4, with an efficiency of 96.6% by 21 CPUs.

As a test calculation for SBDD, re-docking was performed for a complex of c-ABL, and sufficiently small RMSD(0.13Å) was obtained.

Keywords: Force field method, Special purpose computer, Ligand based drug design, Structure based drug design