

○浜田 道昭、稲垣 祐一郎（みずほ情報総研）、中馬 寛（徳島大学）

はじめに

創薬の研究開発は化学・生物学を中心とし、多種の関連分野の知識を必要とする。近年の構造ゲノミクス及び計算（機）科学より得られる大量の情報の整理・解析は創薬の研究開発をより効率的にすると期待されている。我々のプロジェクト「g-drug discovery」は、Grid テクノロジーや XML データベースなどの新しい IT 技術を活用し、従来の定量的構造活性相関解析と大規模分子科学計算・シミュレーション・構造ゲノミクスからの情報を統合し、創薬に関する情報の解析と知識獲得のためのプラットフォームの構築を目指している（図 1）。このプラットフォーム上では、様々な計算科学ソフトウェアが、3次元分子構造データベースを介してシームレスに接続することが可能となる。今回の発表においては、この創薬プラットフォームの概要及びその有効性を実験により確かめた結果について発表をおこないたい。

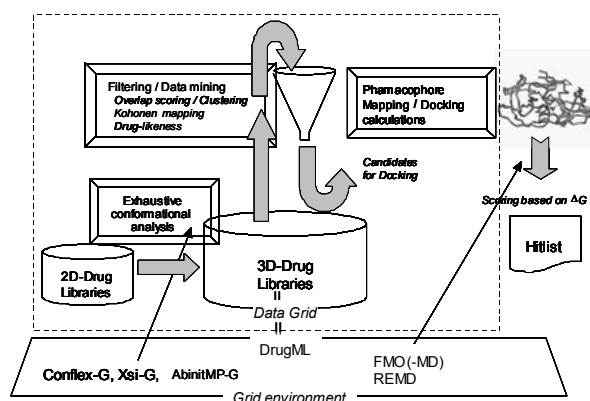


図 1. 創薬プラットフォームの全体像 Grid 環境は OmniRPC（後述）にもとづき構築されている。また、データベースにはネイティブ XML データベースを用いている。

DrugML (Drug Markup Language)

創薬プラットフォーム上で様々なアプリケーションが連携を行うためには、アプリケーションの間で共通のデータ構造が必要となる。近年 XML (eXtensible Markup Language) の隆盛により、様々な分野で、このようなアプリケーション間の共通のデータ構造としてマークアップ言語が用いられてきている。そして、化学分野においては化合物の構造からスペクトル等の実験情報までを間便に記述可能な CML (Chemical Markup Language)¹⁾ が広く利用されている。本研究においては、この CML を出来る限り利用し、創薬のためのマークアップ言語 DrugML を開発した。DrugML の主要な構造を図 2 に示す。DrugML では CML には存在しない、以下の概念（タグ）を新たに追加した。

- **universe** : 複数の分子のスナップショットを表す。これを用いることにより、例えば結合しているレセプターとリガンドの状態を表現することが可能となる。
- **conformation** : 分子の「配座」を表す。CML においては、分子を構成する各原子の座標情報を表すことはできるが、「配座」という概念は明確になっていない。このタグを用いることにより、1 つの分子に対して計算で求めた多数の安定配座を効率的に記述することが可能となっている。
- **descriptor** : 分子の様々な記述子を表す。記述子は、分子の 1 次元・2 次元・3 次元構造に依存するものを、それぞれ別々に記述する。記述子の数値データは、CML の scalar・array・matrix タグを用いて表す。



創業プラットフォームの中心に位置し、アプリケーション間のシームレスな接続を可能とするデータベースシステムを開発



The diagram illustrates the architecture of the Xsi system, organized into several layers and components:

- Application Layer:** Contains the **Xsi** application and the **Browser**.
- Schema Layer:** Contains **DrugML or CML**.
- General Interface Layer:** Contains **ID Operation** and **General DB Operations**.
- DB Connection Layer:** Contains two **XML-RPC Interface** components. A bidirectional arrow labeled **HTTP (XML-RPC)** connects these two interfaces.
- Database Layer:** Contains **ID Server**, **Xindice**, **Servlet**, and **XML:DB**.

Communication flows are indicated by arrows:

- A vertical arrow labeled **HTTP** connects the **Browser** to the **Servlet**.
- A bidirectional arrow labeled **HTTP (XML-RPC)** connects the two **XML-RPC Interface** components.

図4 システム構成図

本プラットフォームの Grid 環境は OmniRPC に基づいて構築されている。OmniRPC はプロジェクト内の共同研究者である筑波大学の佐藤三久教授らにより開発された GridRPC システムであり、以下の特徴を有する^{2,3,4)}。

- プログラミングモデルとして、master-slave 型の並列プログラミングをサポートする。
- 利用する計算資源として、単一の PC やワークステーションはもちろんのこと、クラスタを利用することが可能である。クラスタを構成するネットワークとしてプライベートなアドレスを用いて構成されたクラスタについてもサポートしている。
- API として、基本的に Ninf version.1 の API を踏襲している。また、リモート側の状態を保持する **persistence** をサポートしているため、これを利用した効率的なプログラミングが可能である。現在のバージョンでは C 言語の API を提供している。
- 並列プログラミングのための API としては、非同期呼び出しを用いることができる。
- パラメータ検索など並列アプリケーションを効率的にサポートするために、自動初期化実行モジュール機能を提供している。これは、初期化のための大量のデータの転送や計算が必要な場合、これを再利用することにより、効率化する機能である。
- 認証を行う Grid 環境として、Globus の他、ssh の利用も可能である。ファイヤウォールのある遠隔の計算機についても、omrpc-agent による proxy 機能を用いることにより、ssh で login できるシステムならば、計算資源として利用できる。

【概要】OmniRPC による、Grid 化の性能を測定するために、Grid 化した薬師 (Xsi)⁸⁾ (これを Xsi-G と呼んでいる) を用いて、HIV プロテアーゼとその阻害剤 (今回は **viracept** を用いた) の複合体の構造を再現した。ここで阻害剤の 3 次元構造は未知と仮定した。

【手法】計算フローを図5に示す。WHIM 記述子^{6,7)}を用いて、バインディングサイトとリガンドの向きを合わせて、Grid により Docking 計算を行っている。

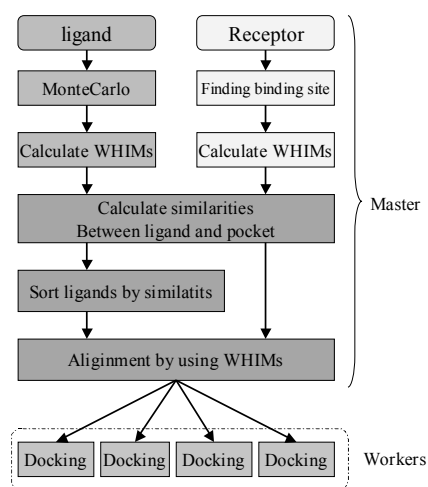


図5 計算フロー

【結果】計算した結果のリガンドの構造の中で、実験の構造に最も近かったものを図7に示す。このときの RMSD 値は 1.77 Å であった。また、Grid 計算を行うことにより、72CPU で約 67 倍の高速化が得られた (図6)。

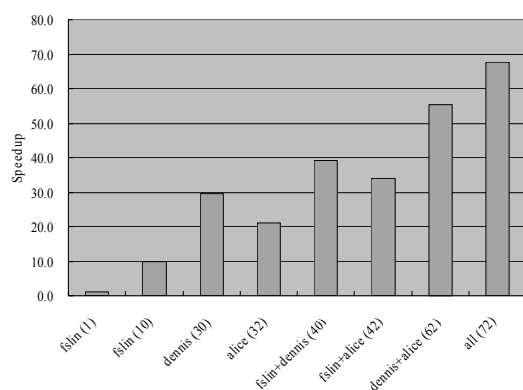


図6. Grid 計算によるスピードアップ 括弧の中の数値は CPU 数を表す

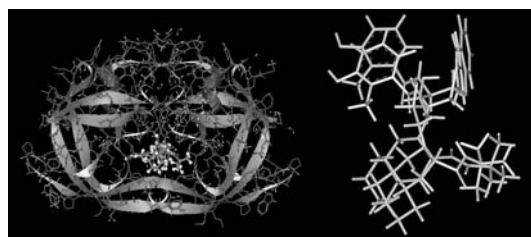


図7. 計算で得られた viracept と HIV プロテアーゼの複合体 (左)、実験で得られた構造と計算で得られた

構造の比較 (RMSD 1.7 Å) (右)

応用 2 (リガンドアラインメントを用いたバーチャルスクリーニング)

【概要】創薬プラットフォームの全体的な有効性を確かめるために、上述のデータベースに格納された 888 個のリガンドから、5 つの HIV プロテアーゼ阻害剤 (図8) のスクリーニングを行った。

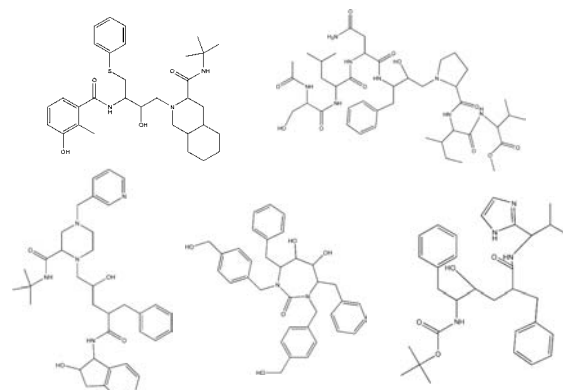


図8. HIV プロテアーゼ阻害剤

【手法】上述のプラットフォーム上で、Xsi に実装されている、リガンドアラインメントモジュールを用いて計算を行った (図9)。リガンドアラインメントとは、以下のような機能を有するものである：

- **リガンド間相同性指標の作成および適合度計算**：グリッドの各格子点に割り当てられた単独あるいは複数分子の特性情報から、LBDD 用モジュールで用いられる各種情報解析機能を用いて、リガンド間相同性指標を作成する。また、任意のリガンドについて座標データが与えられた時、リガンド間相同性指標に対する適合度を計算する。
- **最適解探索**：上記グリッド内において、リガンドの位置、向きおよび配座を任意に変化させ、“標的蛋白質との結合自由エネルギー”と“リガンド間相同性指標に対する適合度”のいずれかあるいは双方について、最適解を探索する。適合度関数には以下のものを用いた。

$$A = \varepsilon \sum_{t=\text{type}} c_t S_t + c_0 E_{\text{interaction}} + c_1 (E_{\text{internal}} - E_0)$$

ここで、 S_t はファーマコファタイプの similarity (Euclid, Cosine, Tanimoto)、 $E_{\text{interaction}}$ は分子間相

相互作用のエネルギー、 E_{internal} は分子の内部エネルギー、 E_0 は内部エネルギーの原点（真空中での最小エネルギー）、 $c_i \cdot c_0 \cdot c_i$ は重み定数をそれぞれ表す。

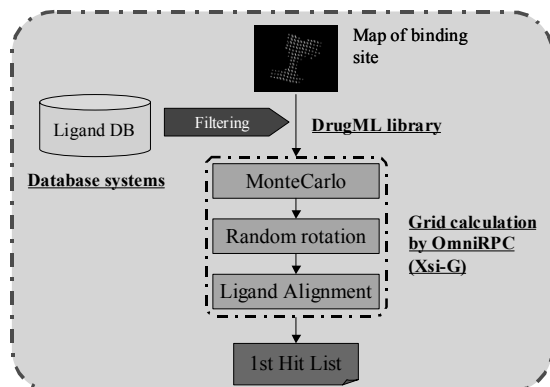


図 9. 計算フロー

【結果】5つの HIV プロテアーゼ阻害剤が低ランク（すなわち結合活性が高いと予測）に集中している（図 10）。計算時間は 40CPU（Grid）で約 320 分であった。図 11 は阻害剤の一つとマップと間の最適なアラインメントを示している。マップからはみ出している部分は HIV プロテアーゼの外側となっている。

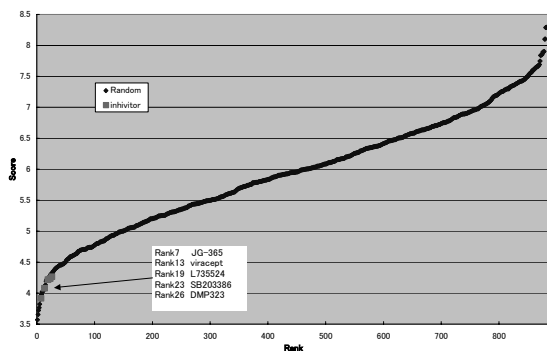


図 10. バーチャルスクリーニングの結果

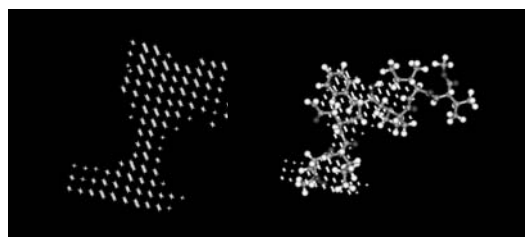


図 11. Binding サイトと阻害剤の最適アラインメント

まとめ

本研究では、Grid 技術と最新の XML 技術を用いて、創薬の研究開発を効率化する創薬プラットフォームを構築した。また、その有用性を、実際に計算を行うことによって確かめた。

謝辞

本研究は科学技術振興事業団（JST）計算科学活用型特定研究開発推進事業により行われた。ここに記して謝意を表する。

References

- [1] <http://www.xml-cml.org/>
- [2] Mitsuhsa Sato, Taisuke Boku, Daisuke Takahashi, OmniRPC: a Grid RPC system for Parallel Programming in Cluster and Grid Environment 3rd International Symposium on Cluster Computing and the Grid (CCGrid2003), May 12 - 15, 2003, Tokyo, Japan.
- [3] Mitsuhsa Sato, Motonari Hirano, Yoshio Tanaka, Satoshi Sekiguchi, OmniRPC: A Grid RPC Facility for Cluster and Global Computing in OpenMP WOMPAT 2001, 130-136.
- [4] <http://www.omni.hpcc.jp/OmniRPC/index.html.en>
- [5] <http://csb.stanford.edu/koehl/ProShape/>
- [6] Todeschin, R. and Gramatica, P., New 3D Molecular Descriptors: The WHIM theory and QSAR Applications, In 3D QSAR in Drug Design Volume 2, Eds., Kubinyi, H., Folkers, G and Martin, Y.C., 355-380, KLUWER/ESCOM, Dordrecht, 1998.
- [7] R. Todeschini, M. Lasagni, and E. Marengo, "New Molecular Descriptors for 2D and 3D Structures. Theory", J. Chemometrics, 8, pp.263-272, (1994).
- [8] <http://www.fuji-ric.co.jp/st/xsi/index.html>