

Gridテクノロジーを用いた創薬プラットフォームの構築

徳島大学大学院 ヘルスバイオサイエンス研究部 中馬 寛

Development of a platform for drug discovery using Grid technologies.

Hiroshi Chuman, The University of Tokushima

Abstract:

Our project has aimed at developing a platform for drug discovery using Grid technologies, on which various molecular simulation, computation, and analyses for structure based Quantitative Structure-Activity Relationships of drug molecules can now be executed. “Drug-ML” database was also developed as Data Grid for storing and retrieving data of the results.

1. 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および計算(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。従来、論理的創薬の方法として知られているもののほぼすべての手法は一連の薬物分子間の構造類似性(仮定:生物活性・機能の強度は薬物分子の構造類似性の程度で決定される)に基づくものであり一般に定量的構造活性相関解析として使われてきている。一方、近年の計算科学技術を活用した構造ゲノミクス(バイオインフォマティクス)と大規模分子科学計算・シミュレーション技術の進展にともない生理活性分子とその受容体タンパク質等との原子・電子レベルでの詳細な相互作用解析が標準に入りつつある。

本プロジェクトは新しい計算科学技術を活用し、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の獲得のためのプラットフォーム構築

を目的とする。すなわち、Gridテクノロジーを用いて計算機およびデータベース等の広域かつ高速なネットワーク上に分散した計算機資源を接続して一つの仮想的な巨大計算機を構成することにより、生理活性分子ならびにその受容体構造情報を基に各種の大規模分子科学計算を大量・多種の創薬分子系へ適用し、得られる大量の計算結果データを

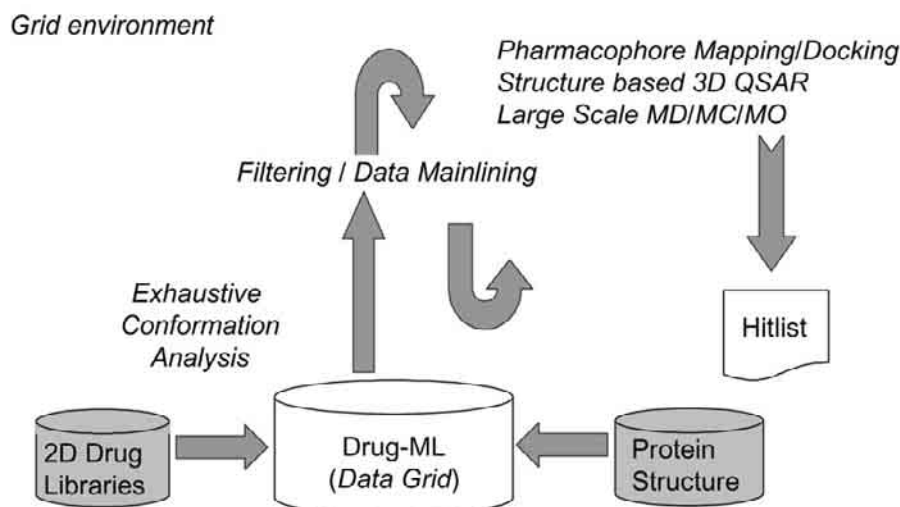


図1. 創薬プラットフォームの全体概要

共有・統合し、創薬に資することである(図1)。従来のローカルな計算資源のみを使用する手法とは異なり、ハイスループットかつ確度の高い新しい創薬のアプローチとその資源を創薬研究者に提供することになる。本プロジェクトは創薬研究者に有用なアプローチと資源を資するとともに、学術的意義として薬物分子の活性発現に関する電子・原子レベルでの新規な定量的構造活性相関解析法の基盤の一つを与える。

近年注目を浴びている構造ゲノミクス側からのアプローチであるバイオインフォマティクスと構造活性相関解析側からのアプローチであるケモインフォマティクスの統合的活用が今後の創薬の研究・開発において焦点になると予想されるが、本プロジェクトでの成果はその基盤の一つになる。

2. 研究開発項目とその成果概要

2.1 Grid分子科学計算

・網羅的配座探索計算から受容体結合構造の候補絞り込み

一般に同一受容体タンパク質に結合すると想定される一連の候補薬物分子を2次元分子構造ライブラリから選択し、それらの受容体との結合時の3次元構造(活性配座)を推定することが最初のステップとなる。このため、候補分子についての網羅的配座探索、その結果得られる複数の3次元構造について情報化学的手法による活性配座の候補を絞り込むことを行う。この過程での計算はそれほど一般に計算負荷はあまり大きくないものの入出力データは分子数、配座自由度等の組み合わせによって膨大となり、Grid計算に適している(Combinatorial computing)。

2種類の網羅的配座探索法の開発を行った；Conflex(MM力場)は局所変形により、Xsi(MMFF94s)は二面体角値の変更により初期配座を系統的に発生させ、立体エネルギーの極小化を行う。それぞれのプログラムの並列化およびGrid対応化を行い、ともに効率よく低エネルギー安定配座を網羅的に求めることが可能なことを確認した。Conflexの並列化・Grid化による性能向上の結果を図2に示す(ペプチド(Ala)₄および(Ala)₁₆、並列度45~160)。なお、Conflex、Xsiについては溶媒効果の取り込み、基準振動解析、Monte Carlo法等の機能が組み込まれている。

上記の配座探索計算によって得られる薬物分子の各低エネルギー安定配座について立体構造、電荷分布、水素結合供与・受容体分布、疎水性分布を表わす座標依存性のない記述子(2次モーメント量)を定義した。複数の薬物分子-安定配座のそれらの記述子を用いたクラスタリングと類似性判定から共通する3次元構造特性を抽出し(pharmacophore mapping)、活性配座候補の絞り込みを行った(Drug-MLデータベースシステム(2.2))。

薬物-受容体相互作用予測において、両者の結合構造予測および相互作用自由エネルギーの定量的評価が最重要となる(docking)。前者には上記の配座探索法と安定配座のクラスタリング・類似性判定、後者には広範囲なサンプリング(熱平衡分布)が可能な分子動力学法およびそのポテンシャルエネルギーの精度を確認する大規模分子系に適用可能な非経験的分子軌道法が必須となると考えた。これらの方法は計算負荷が大きいため、プログラムの並列・Grid化が必要となる。この目的のためにGrid対応のレプリカ交換

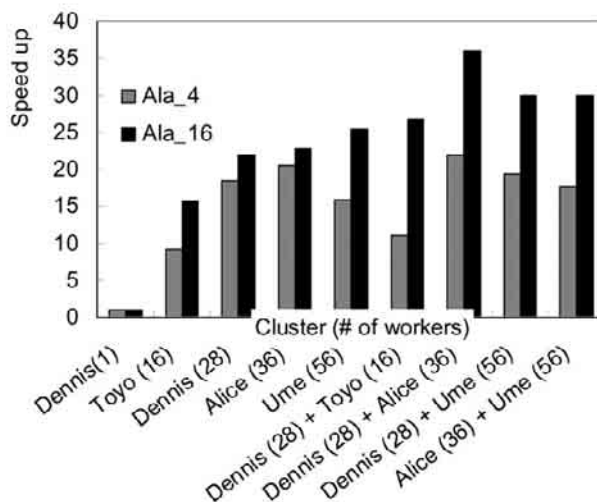


図2. Grid環境でのConflex-Gの性能向上
横軸はクラスターと(CPU数)を表わす。

分子動力学法(REMD)を開発し、現在テスト分子系についての稼動検証を終了している。一方、大規模分子系に対応し、Grid対応のフラグメント非経験的分子軌道法(FMO)を検証例(2.3)に適用し、今後の有用性と実用性を確認した。現在FMOを用いた分子動力学計算(FMO-QMD)の開発と検証を行っている(JST-CREST、研究代表者 長嶋雲兵)。

2.2 Gridデータベース

近年、XML (eXtensible Markup Language) の隆盛により様々な分野でマークアップ言語が用いられている。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簡便に定義可能なCML (Chemical Markup Language) がすでに広く利用されてきている。これらのCMLのタグを可能な限り利用した創薬における研究・開発を対象としたDrug-MLの開発を行った。Drug-MLでは一連の薬物分子の集合や一つの化合物に対して複数の配座情報や構造記述子を表現するためのタグを追加している。さらに各サイトで得られたそれぞれの分子構造情報を分散・集中管理できるよう設計している(Data Grid)。2.1で述べた本データベースに格納する情報は、階層的分子構造情報として格納され、それらの情報を用いて活性発現に必要な共通構造(立体構造、電子構造、水素結合性、疎水性)を抽出し、構造のクラスタリングおよび構造間の類似性判定を行う(Drug-ML データベースシステム)。なお、Drug-ML データベースシステムの開発および今後の利用についてはACT-JSTプロジェクト「DNAナノ領域ダイナミクスの第一原理解析」(研究代表者 田中成典)と共同で検討を行っている。

2.3 実証

前節の一連の処理を用いたHIV-プロテアーゼの阻害剤についての解析結果の概略を述べる。

HIV-プロテアーゼの同一部位に結合し、かつ骨格が異なる(構造多様性の大きい)幾つかのHIV-プロテアーゼの阻害剤分子を選択し、Conflex および Xsiによる網羅的安定配座解析を行い、各分子約1000個の低エネルギー安定配座構造を算出した。立体・電子・疎水性・水素結合受容・供与体分布の比較法のため、座標系に依存しない2次のモーメント展開法および Self Organization Map (Kohonen Map) 法を開発・導入し、上記の配座ファミリー間の定量的類似性評価を行った結果、各阻害剤分子の予測候補構造の中にX線結晶解析によるHIV-プロテアーゼとの結合構造と $rmsd$ 値 2.5 以下の類似配座が含まれていることを確認した(図3 (b))。この結果は、「薬物候補分子の2次元構造ライブラリー 低エネルギー安定配座探索情報化学的方法による受容体との結合構造候補の絞込み」の一連の過程の検証に対応する。

次に上記の結果に基づき、阻害剤-タンパク質複合体の全原子(約3200原子)を対象とし、分子力場法(Xsi)、半経験的分子軌道法(LocalSCF; AM1)およびフラグメント非経験的分子軌道法(HF/STO-3G)を適用した結果、妥当と思われる結合エネルギー(エンタルピー) E を得た。レプリカ交換分子動力学法およびフラグメント非経験的分子軌道法の本格的使用については今後の計画としている。

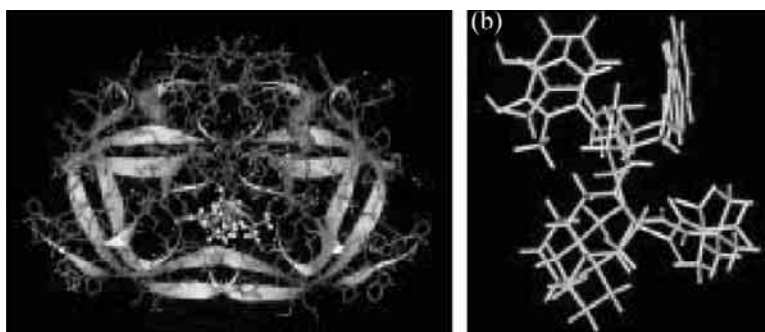


図3. ViraceptとHIVプロテアーゼの相互作用様式
(a)複合体構造, (b)X線結晶解析と予測結果($rmsd=1.78$)

以上の基本的検証に加えて、実際の創薬の研究・開発での使用を想定し、計算機での薬物分子のスクリ

ーニング(virtual screening, *in silico* screening)を行った。5個の既知HIV-プロテアーゼ分子に加えて、薬物の2次元分子構造ライブラリーからランダムに883個の薬物を選択し、検証スクリーニングセットとした。分子力場法(Xsi)によるHIV-プロテアーゼとの相互作用エネルギーおよび各記述子の類似度から定義されるスコアを定義し、合計888個の分子についてのスコアリングを行っ

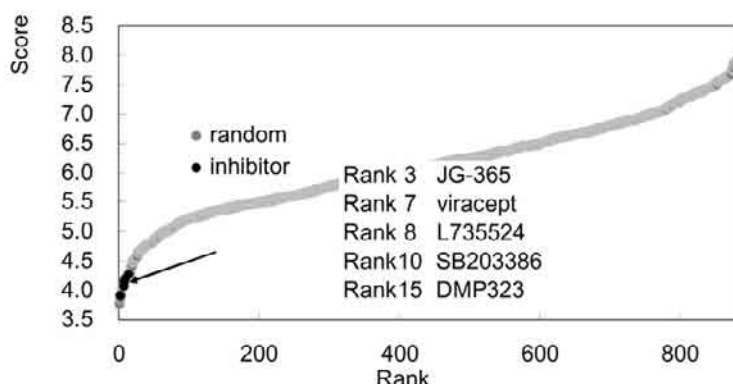


図4. Virtual screeningの結果

た結果 (36 cpu / Grid、339 min)を図4に示す。既知のHIV-プロテアーゼ5分子のスコアが低ランク(すなわち結合活性が高いと予測)に集中していることを確認した。

以上の結果は未整備の部分を残しながらも本プロジェクトの目的である創薬プラットフォームの構築と検証までほぼ到達したと考えている。

3. ネットワークの活用について

これまで、Gridコンピューティングにおける分散並列計算のためのミドルウェアであるOmniRPC用い配座探索プログラムConflex、Xsi等をGrid対応の並列化したアプリケーションを開発し、各サイトにOmniRPCをインストール、アプリケーションを配置し、Grid計算の実行環境を整備した(図5)。この結果、大規模かつ大量の薬物分子の構造探索とスクリーニングに本ネットワークシステムが有効であることを実証した。

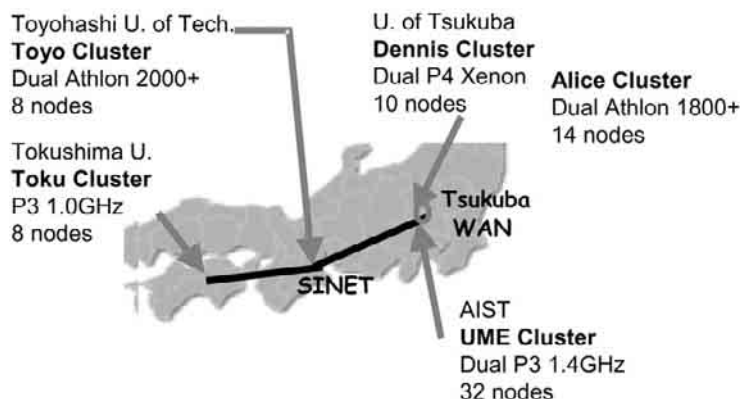


図5. 創薬Gridネットワーク

4. まとめ

本プロジェクトは創薬を直接の対象とし、ケモインフォマティクス・バイオインフォマティクス、定量的構造活性相関・大規模分子科学計算、各種データベースの統合的活用を目指す国内外には例のないものである。本プロジェクトの優位性と独自性を確保するために、継続して研究・開発する項目について今後も鋭意努力を続けていく。

5. 研究開発実施体制

代表研究者 徳島大学大学院ヘルスバイオサイエンス研究部 中馬 寛

研究分担

研究開発項目：3次元DBとその解析システムとそのインターフェースの開発(プラットフォームシステムの有用性検証のための定量的構造活性相関解析)()は平成16年度

徳島大学大学院 ヘルスバイオサイエンス研究部

生体情報科学部門 中馬 寛

研究開発項目：Grid対応大規模分子系計算プログラム開発（Grid対応並列化フラグメント分子軌道法プログラムへの分子動力学計算機能の導入と計算機シミュレーションの実行）

産業技術総合研究所

グリッド研究センター 長嶋 雲兵

研究開発項目：創薬プラットフォームの配座探索計算法（Conflex）の開発（配座探索計算法の稼動検証）

豊橋技術科学大学 工学部

知識情報工学系 後藤 仁志

研究開発項目：創薬プラットフォームのためのGrid環境の開発

筑波大学 電子・情報工学系

計算物理学研究センター 佐藤 三久、朴 泰祐、高橋 大介

研究開発項目：統合プラットフォームの開発研究（3次元DBとその情報処理システムの構築および稼動検証・配座探索計算法の稼動検証）

富士総合研究所 R&S部門

サイエンスソリューショングループ 稲垣 祐一郎