

Grid 対応型分子設計支援システム MolWorks

MolWorks — Grid-Enabled Integrated Software Tool for Molecular Design —

田島 澄恵^a, 八木 徹^a

Sumie Tajima and Toru Yagi

1. はじめに

分子設計支援システム MolWorks は、(株)ベストシステムズで開発されてきた。ビヨンド・コンピューティング(株)が平成 19 年 6 月 21 日に設立され、(株)ベストシステムズより一部事業の譲渡を受け、MolWorks はビヨンド・コンピューティング(株)の製品となっている。

現在、MolWorks は無償 Java アプリケーションソフトウェア (MolWorks2.0) として提供を行っている (<http://www.molworks.com/>)。世界各国から月平均 700 ~ 800 件程度がダウンロードされており、認知度が高まってきている。そこで、GUI の改良や機能追加・強化を行い、Grid 対応型の有償版提供を開始する予定である。

本投稿では、現在提供中の MolWorks2.0 の主な機能、および、有償版として提供予定の物性推算関連機能、Grid 関連機能について述べる。

2. MolWorks2.0 機能 (無償版)

MolWorks 概要図を Figure 1 に示す。MolWorks2.0 には、以下に示す機能が組み込まれている。

Computational Chemistry (計算機化学支援) には、分子構造ファイルの取り扱い、3D 分子モデリング、量子化学計算アプリケーション (Gaussian/GAMESS/Q-Chem/MOPAC) 用プリポスト機能が含まれている。また、半経験的手法のひとつである CNDO/2 法プログラムが内蔵されており、計算実行・ダイポールモーメント表示・分子軌道表示が可能である。Property Estimation (物性推算) では、純物質について Joback 法¹をベースに 19 種類の物性値予測が可能である。また、蒸気圧については温度変化グラフの表示が可能である。Chemical Engineer (化学工学計算) では、物性推算機能から得られた純物質物性値を元に、2 成分系混合物に対する PVT 線図の予測が可能である。Data Base については、分子構造データが約 200 件登録されている。Network Computing では、ネットワークに接続されたマシン上に存在する

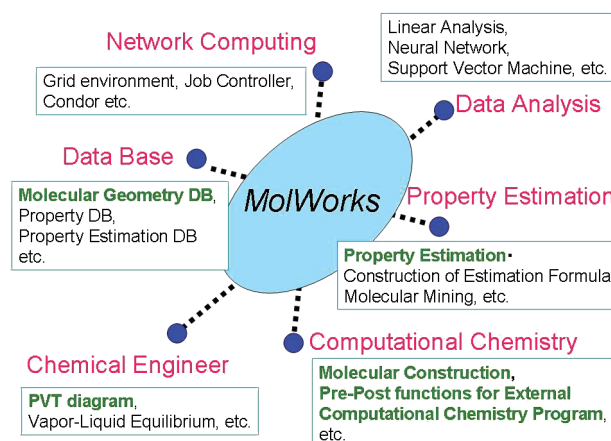


Figure 1. MolWorks conceptual diagram.

データや計算機資源を簡便に利用することを支援する。例えば、手元のマシン上で量子化学計算用プログラムの入力ファイルを作成し、ネットワークに接続された別マシン上で計算実行し、計算終了後、計算結果を手元のマシン上で解析するというプロセスが、MolWorks 画面内操作だけで行うことができるようになる。無償バージョンでは、MolWorks インストールマシン上に Gaussian がインストールされている場合には、MolWorks 画面内から Gaussian 計算を開始することが可能である。Data Analysis (データ解析) については、検証段階であり未実装である。

3. 新機能 (有償版)

3.1 物性推算関連機能

未知の物性値を効率的に、かつ精度良く予測する物性推算機能は、広範囲の分野で必要とされている。我々は、線形的解析手法に加えて、非線形解析手法のひとつであるニューラルネットワーク法 (NN 法) を利用した物性推算式構築を試みている。NN 法は、入力情報と出力情報との隠れた因果関係を抽出することが得意であり、非線形的要素を含む分子構造情報と物性値の関係を表現するのに適した手法である。これまでに、NN 法を利用して、沸点・logP (n-オクタノール/水分分配係数)・溶解度・pKa・Tg (ガラス転移温度) などの推算式を構築してきた。その際、入力情

^a ビヨンド・コンピューティング株式会社
連絡先 〒 305-0032 つくば市竹園 1-6-1 三井ビルディング 8F
電子メール tajima@beyond-computing.com

報には原子団をベースとした分子構造情報を利用した。その結果、全ての物性値において、高精度な推算式が得られた。NN 法物性推算式を利用することにより、構造最適化や量子化学計算などの計算時間を要する作業を必要とせず、高速に精度の良い推算結果を得ることが可能である。

Table 1 に、logP 値 (n-オクタノール/水分係数) の推算結果を示す。Existing Formula [1]² は量子化学計算 (B3LYP/6-31g*) 結果を入力とした線形的手法による推算結果、Existing Formula [2]³ は量子化学計算 (AM1 および PM3) 結果を入力とした NN 法による推算結果、Neural Network は我々の行った NN 法による logP 推算結果を示している。Existing Formula [1] では分子群を、non-hydrogen bonders (group1), hydrogen acceptors (group2), hydrogen donors (group3) に分類している。我々の NN 法での入力パラメータには、原子団情報に加えて立体構造情報を用いた。近似直線および R² 値を比較すると、我々の行った NN 法は、量子化学計算の結果を利用していないにも関わらず良い結果となっていることが分かる。つまり、時間のかかる量子化学計算の結果を利用せずとも、高精度な推算を行うことが可能であることが分かった。

Table 1. Results of logP estimation.

Method	Data	Collinear Approximation	R ²
Existing Formula[1]	group1 (74 molecules)	$y = 0.9825x + 0.0481$	0.982
	group2 (41 molecules)	$y = 0.9639x + 0.0448$	0.964
	group3 (51 molecules)	$y = 0.9135x + 0.0729$	0.908
	All (166 molecules)	$y = 0.9747x + 0.0427$	0.972
Existing Formula[2]: AM1	Teach (985 molecules)		0.931
	Test (105 molecules)	$y = 0.9319x + 0.0906$	0.903
Existing Formula[2]: PM3	Teach (985 molecules)		0.883
	Test (105 molecules)	$y = 0.8557x + 0.2601$	0.845
Neural Network	Teach (425 molecules)	$y = 0.9885x + 0.0424$	0.990
	Test (202 molecules)	$y = 0.9594x + 0.1334$	0.911
	All (627 molecules)	$y = 0.9885x + 0.0158$	0.988

通常の物性推算は、分子構造情報を入力し、各種物性を予測するシステムである。我々は、全く逆のアプローチを行う Molecular Mining 法を提案している。Molecular Mining 法では、物性情報を入力し、分子構造情報を出力する。

パラキシレンの物性値に基づき、沸点: 403.4~419.4 [K], 臨界温度: 607~627 [K], 臨界圧力: 32 ~ 38 [bar], 臨界体積: 368~388 [cm³/mol] という物性条件のもと、Joback 法を利用して Molecular Mining を行った。先ず初めに、20 種類のフラグメントを、主鎖が 2 フラグメントから 5 フラグメントまでという条件で機械的に組み合わせて、258907 の鎖状分子を自動発生させた。構造を自動発生させることは、候補分子の漏れを防ぐ効果がある。発生させた分子について物性推算を実行し、物性条件に対応した分子を抽出

すると、13 分子にまで絞り込むことが出来た。これらの操作を、CPU: Pentium III 1 GHz, Memory: 512 MB の計算機上で行うと、258907 分子の生成に約 4 分、258907 分子の物性推算に約 20 分、合計約 25 分を要した。

次に、異なる分子領域でパラキシレンの代替物質を探索した結果を示す。先の例と同様に、Joback 法を利用して Molecular Mining を行った。今回は、ベンゼン環やシクロヘキサン構造を含む 16 種類のフラグメントを、主鎖が 2 フラグメントから 7 フラグメントまでという条件で機械的に組み合わせて、1317012 分子を自動発生させた。今回は 13 種類の物性推算も同時に行った。また、計算機環境は、CPU: Pentium4 2.8 GHz, Memory: 512 MB と先の計算機よりも高精度なものを利用した。その結果、1317012 分子の生成および推算に約 11 分を要した。物性条件に従って絞り込みを行うと、パラキシレン分子以外に、14 通りのフラグメントの組み合わせが抽出された。そのうち、5 通りは先の Molecular Mining の結果と同じ組み合わせであった。つまり、9 通りのフラグメントの組み合わせからなる分子群が、新たなパラキシレンの代替物質として得られた。

Molecular Mining の高速化と汎用化を目指し、未知分子を含む分子物性推算データを集約した分子物性推算データベース: CPDB⁴ の開発を行った。現在、産業技術総合研究所 RIODB (<http://www.aist.go.jp/RIODB/cpdb/>) で一部を公開している。このように、必要に迫られる前に、様々なフラグメントの組み合わせによって得られる大規模な分子物性推算データベースを用意しておくことは、新規材料開発の大幅な時間短縮に貢献することが期待される。

3.2 Grid 関連機能

ジョブスケジューラが管理する複数のクラスターシステムや、Grid ミドルウェアが管理する Grid 環境内に計算資源が存在する場合、スケジューラや Grid ミドルウェアには様々な種類やバージョンが存在しており、これらを使い分けることは、管理者・利用者に対して非常に大きな負担となる。そこで我々は、新たな環境構築を必要とせず、既存システムをそのままに、それらを統括して利用できるインタフェース作成を試みている。

各種スケジューラや Grid ミドルウェアの種類に依存せず、統括的にジョブのスケジューリングを行うスケジューラを、メタスケジューラと呼ぶ。メタスケジューラは、最も効率よく実行できる計算資源に対して与えられたジョブをサブミットし、その実行を管理する。このようなメタスケジューラは、既存システムをそのままにし、ジョブの実行に対して同じインタフェースを提供できることから、計算資源の管理者と利用者の双方に対して非常に有益なツールとなる。現在、メタスケジューラとしては Condor⁵ が有望である。Condor は、High-Throughput Computing を目指して開発が進められており、Grid 環境との協調がはじめから意

識されている。我々は、MolWorks が Condor を利用して、ジョブをネットワーク上の計算機にサブミットできるようにする予定である。これにより、ユーザは常に同じインタフェースからジョブを実行できるようになり、スケジューラの種類に依存した利用方法の違いを意識せずに、計算実行と結果解析という本来の研究に専念することが出来るようになる。

現在、MolWorks3.0 発表に向けて、GUI 強化、Condor との連携、その他機能向上を行っている。MolWorks3.0 では、無償版（機能を限定）と有償版に分けて提供する予定である。発表時期は、2008 年 3 月を予定している。

引用文献

- (1) Joback, K. G.; Reid, R. C. *Chem. Eng. Comm.*, **1987**, 57, 233.
- (2) Chuman, H.; Mori, A.; Tanaka, H. *Analytical Sciences*, **2002**, 18, 1015–1020.
- (3) Breindl, A.; Beck, B.; Clark, T.; Glen, R. C. *J. Mol. Model.*, **1997**, 3, 142–155.
- (4) Tajima, S.; Yagi, T.; Fukuda, T.; Nagashima, U. *J. Comput. Chem. Jpn.*, **2006**, 5, 23–28.
- (5) <http://www.cs.wisc.edu/condor/>

(受理日 2007 年 10 月 1 日)