

計算化学を利用した分子構造のスクリーニング Automated screening of molecular structures with computational chemistry

一般財団法人高度情報科学技術研究機構
太田 幸宏

半経験的分子軌道法計算パッケージMOPACに基づき、化学反応経路自動探索ツールGRRMの利用を可能とするインターフェイス計算コードを開発する。開発コードにより、半経験的分子軌道法による高効率な反応経路探索が可能となる。その応用として、two-step structural searchによる安定構造の網羅的探索を提案する。すなわち、半経験的分子軌道法による安定分子構造の網羅的なスクリーニング後、分子軌道法や密度汎関数法など高精度計算による再最適化により、安定構造分布を系統的に得ることができる。開発コードとtwo-step structural searchの適用事例として、セルロースにおけるカチオン吸着構造探索を実施する。本インターフェイス計算コードは、半経験的分子軌道法が正当な範囲で、大規模分子系における様々な化学反応経路探索に利用できると期待される。

1. はじめに

系統的な化学反応経路探索は、原理的な関心に加え、多彩な応用とつながるため、計算化学における重要な課題の一つである [1]。その実現は計算機シミュレーションによる様々な材料開発、例えば、触媒反応の設計 [2]、構成物質のみによる結晶構造の生成 [3]、に大きく寄与する。そのためには、ポテンシャルエネルギー面上において最適化問題を解くための高効率アルゴリズム、膨大な計算結果に対する効率よいデータ処理法の整備が要請される。こうした要請を満たす手法について精力的な研究が実施されている [4, 5, 6, 7]。GRRMは化学反応経路を自動的に探索する目的で開発された計算パッケージである [7, 8]。GRRMの利用により、最安定構造の周りに存在し得る局所安定構造の全探索、安定構造間をつなぐ反応経路と付随する遷移状態の探索、任意の反応物から生成物に至る反応経路の非経験的探索が可能となる。その適用事例は、シリコン表面における酸素原子の反応経路探索 [9]、ハロゲ

ン化有機物のホウ素化反応経路探索 [10]、未知の炭素構造体の予測 [11] など多岐に渡る。

GRRMにおいて、ポテンシャルエネルギー面の評価が計算実行時間を支配する。特に、タンパク質や生体分子など、中・大規模分子系において、計算コストの低減化は必須となる。GRRMの探索アルゴリズムでは、ポテンシャルエネルギー面の勾配が要求され、それらは外部計算パッケージ、例えばGAUSSIAN [12] やGAMESS [13] などを通し、分子軌道法あるいは密度汎関数法により計算される。よって、現実的な全自動探索の実行には、利用可能な計算機資源に応じて、エネルギー勾配の計算に必要なコストを削減することが望まれる。

本稿では、中・大規模分子系における反応経路自動探索の効率化を目指し、半経験的分子軌道法計算パッケージMOPAC [14, 15] に基づいたGRRMの利用を可能とするインターフェイス計算コード [16] の開発について報告する。加えて、開発されたコードの応用として、two-step structural searchによる

安定構造の網羅的探索法を提案する[17]。この手法では、以下の手順で安定構造を探索する。すなわち、(1) インターフェイス計算コードを使い、半経験的分子軌道法により、低コストかつ低精度で安定構造候補を網羅的に見つける、(2) 各候補に対し、より高コストかつ高精度の計算手法を用いて再構造最適化を実行する。言い換えれば、計算機シミュレーションにより、安定な分子構造の網羅的なスクリーニングが実行される。この提案の適用事例としてセルロースにおけるカチオン吸着サイト探索の結果を報告する。

2. 化学反応経路自動探索ツール：GRRM

GRRMは化学反応経路を自動的に探索することを目指し開発されている計算パッケージである[1, 7, 8]。本節では、文献[1]に従い、その特徴をまとめる。

効率的な反応経路探索ツールを開発する動機を理解するため、反応経路全探索を概観する。化学反応を理解するアプローチとして、Born-Oppenheimer近似の下、ポテンシャルエネルギー面を計算することは主要な位置を占める。それは多変数非線形関数の最適化問題として定式化される[16]。ポテンシャルエネルギー面を記述する変数は、例えば結合距離や結合角など、分子に含まれる内部自由度に対応する。N個の原子からなる（直線分子でない）分子であれば、大域的な並進と回転を除き、 $3N - 6$ 個となる。ポテンシャルエネルギー面の極小値（山の谷底）は局所安定構造に対応する。異なる極小値の間に存在する鞍点（山の谷底と尾根に挟まれたような峠）は遷移状態に対応する。ある安定構造から遷移状態を経て別の安定構造（あるいは別種化合物への分離に対応する解離チャンネル）に最小のエネルギーコストで到達する経路が反応経路に対応すると考えられる。安定構造の探索のため、最急降下法やBroyden-Fletcher-Goldfarb-Shanno法などが利用さ

れる[18, 19]。遷移状態探索は安定構造に比べより難しいが、様々な手法で対応できる。例えば、反応物と生成物が既知の場合、その間を多点補間することで、遷移状態の探索と反応経路を見出すNudged Elastic Band法[20] やストリング法[21, 22] は有効である。

個別の安定構造や遷移状態の探索において、上記の計算手法は効果的である。しかし、次に示すような大域的な全探索への適用は困難である。すなわち、ある安定構造の周りの異なる安定構造を網羅的に探索し、かつ出発点となった安定構造との間に存在する遷移状態を系統的に見つけ出す課題である。また、任意の反応物から出発して、非経験的に可能性のある全ての生成物とそれに至る反応経路を予測する課題も含まれる。これらは挑戦的であるが、計算機シミュレーションによる未知の反応予測を実現するため、解決すべき課題である。そのため、ポテンシャルエネルギー面の重要な構造を詳細に探索する手法が望まれる。特に、ポテンシャルエネルギー面は多変数（N原子系であれば $3N - 6$ 個）の関数であることから、多次元空間における高効率探索アルゴリズムが課題解決に必要となる（素朴な格子点サンプリング法における計算コストの見積もりは興味深い。例えば、ギ酸のような5原子分子でサンプル点を100点とすると、1点あたりの計算時間を1秒として、探索終了時間は「宇宙の年齢」を超えてしまう[1]）。

GRRMでは、上記の大域的探索を実現すべく考案された2種のアルゴリズムを含む計算パッケージである[1, 7, 8]。第1に、非調和下方歪追跡法（Anharmonic Downward Distortion Following法、略してADDF法）は、与えられた安定構造から出発し、その周辺に存在し得る安定構造（すなわち、配座異性体）、遷移状態、そして別種化合物へ分離する解離チャンネルを全探索するアルゴリズムである。ADDF法の基本的アイディアは、分

子系のポテンシャルエネルギー面に典型的な構造に着目したものである。すなわち、遷移状態を経て異なる安定構造をつなぐポテンシャルは、各安定構造周りでは、調和近似されたそれに比べ、非調和性のため小さくなる(非調和下方歪み)。この非調和下方歪みを追跡することで、安定構造候補、およびその中間ある遷移状態候補が系統的に探索される。その探索はスケールされた超球面上(Scaled Hypersphere、略してSHS)で実行される。SHSは、調和近似の範囲で等エネルギー面になるよう定められる。よって、SHS上で分子のエネルギーを計算することで、非調和下方歪みが効率よく探索され得る。第2に、人工力誘起反応法(Artificial Force Induced Reaction法、略してAFIR法)は、与えられた反応物から出発し、可能な生成物候補と付随する反応経路候補を全探索するアルゴリズムである。AFIR法の基本的アイデアは、人工力を付与することで生成物候補を探索するというものである。一般に、化学反応には共有結合の切断や再結合が含まれるため、反応のポテンシャル障壁は非常に高いと予測される(例えば、OH結合の結合解離エネルギーは101.76kcal/molであり[23]、これを温度に換算すると約50000Kに相当する)。人工力の付与は、こうしたポテンシャル障壁を下げ、効率のよい生成物探索に導く。生成物候補が一旦検出されれば、その付随する反応経路から人工力の寄与を引くことで、反応物から生成物へ至る反応経路候補(また、遷移状態候補)も見出される。人工力は原子間距離について線形なポテンシャルとして与えられ、人工力強度に対応する単一のパラメータ、model collision parameterを調整することにより、生成物候補数が系統的に制御される。

3. インターフェイス計算コードの開発

GRRMの主要アルゴリズム、ADDF法と

AFIR法のどちらも、分子系におけるエネルギー勾配のデータを要求する。ここでは、半経験的分子軌道法計算パッケージMOPACによりエネルギー勾配を計算することでGRRMの利用を可能にするインターフェイス計算コードの開発について説明する。

まず、半経験的分子軌道法[18]について簡潔にまとめる。半経験的分子軌道法は、典型的な分子系の実験データを再現するよう定められた経験的なパラメータを利用することで、分子軌道法において計算コストがかかる手続きを簡略化した電子状態の計算手法である。具体的には、内殻の電子の寄与は考えず、価電子に関係する基底関数のみを利用する(周期表の二列目であれば、s軌道とp軌道のみ考慮することになる)。さらに、ハートリー・フォック法で最も計算コストが高いフォック行列の評価に関する2電子積分について簡略化を行う。すなわち、異なる原子間で基底関数の重なりをゼロと近似することで2電子積分の多くを無視し、残りの2電子積分については実験データに基づく経験的パラメータで評価をする。こうした近似手法の特徴から、遷移金属を含む化合物や共有結合の切断・再結合を含む化学反応の適用には注意を要する。一方、有機低分子や生体分子の安定構造を調べる目的には非常に適している。MOPAC[14, 15]は、このような半経験的分子軌道法を利用した電子状態計算パッケージである。なお、遷移金属を含むような場合について半経験的な手法により電子状態を評価する計算パッケージとしては、DFTB[24]が知られている。DFTBとGRRMを組み合わせた計算事例は文献[25]で報告されている。

では、開発されたインターフェイス計算コードについて説明する。コード開発はGRRM14[8]とMOPAC2012[14, 15]の利用を念頭に実施された。開発言語はFortran 90が使用された。GRRMにおける

ADDF法あるいはAFIR法で要求されるデータはエネルギーおよびエネルギー勾配である。開発コードが呼び出されると、GRRMの要求に従い、MOPACによりこれらのデータは計算される。そして、MOPACの標準的な出力データはGRRM側で利用可能なデータ形式に整形され、GRRM側に引き渡される仕組みである(図1)。なお、本インターフェイスの使用においては、エネルギー勾配の差分からHessianを評価することを指定するオプションをGRRM14側の入力ファイルで指定する必要がある。GRRM14の実行開発されたコードは、GitHub上にGNU GPL Version 3.0のライセンス形態で公開されている[16]。

今回開発されたコードは、MOPACとGRRMを直接結び付けるインターフェイス、という特徴がある。半経験的分子軌道法は、GAUSSIANやGAMESSを通して利用が可能

である。これは、内部的にMOPACを呼び出すことに相当している。この呼び出しのオーバーヘッドのため、今回開発されたインターフェイスの利用の方が効果的な探索を可能にする[17]。例えば、セルロース単量体におけるK⁺吸着サイト探索では、MOPACの直接利用におけるエネルギー勾配計算1回あたりに対する(平均)経過時間は、GAMESS[13]を経由した場合の約1/10になることが観測された[17]。よって、インターフェイスを通すことで、GAMESSの利用と比べ、約10倍程度早く探索は終了し得る。すでにGAUSSIANを通した半経験的分子軌道法の計算によりGRRMを利用する研究報告はあった[26]。しかし、この分析から明らかのように、MOPACを直接利用するほうが、より効果的な探索が可能となる。

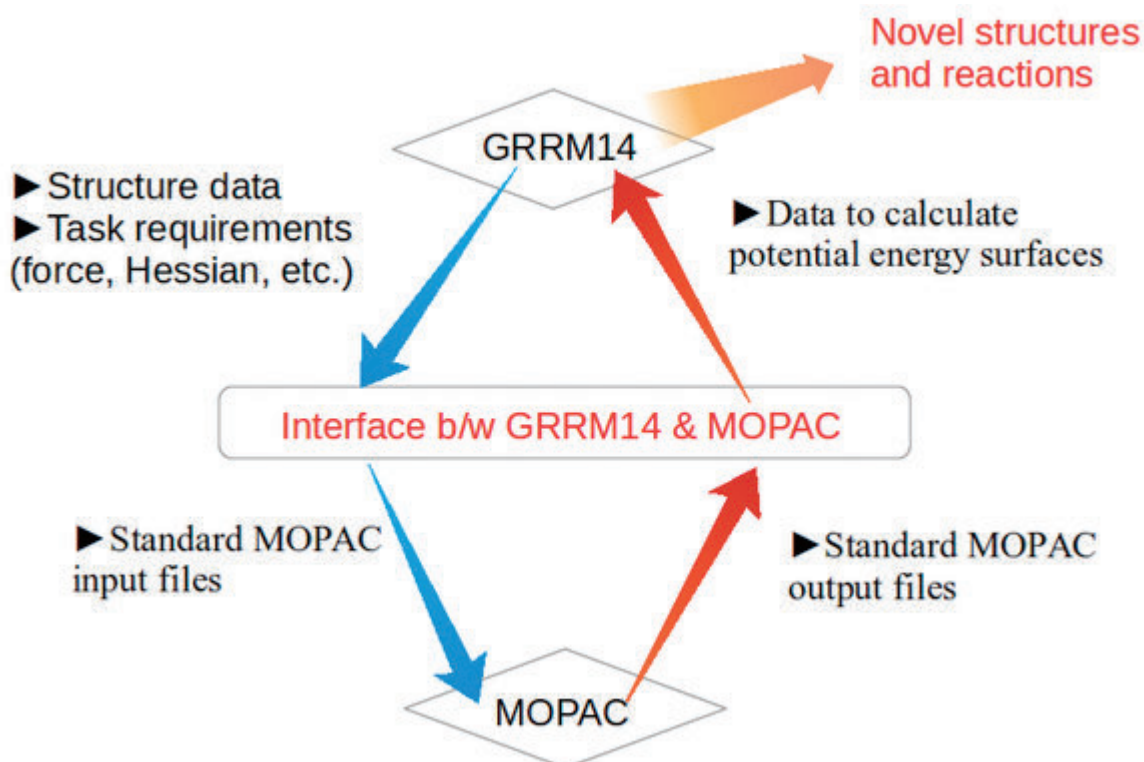


図1 GRRMとMOPACのインターフェイス計算コードの概念図

4. Two-step structural searchによる

セルロースへのカチオン吸着サイト探索

開発コードの適用事例として、two-step structural searchによるセルロースへのイオン吸着サイト探索の結果を示す [17]。セルロース ($C_6H_{10}O_5$)_nの単量体 (n=1) と二量体 (n=2) におけるカチオン (Li^+ , K^+ , Cs^+) の吸着サイト探索を実施した。まず、半経験的分子軌道法により、網羅的なスクリーニングを実行する。エネルギー及びエネルギー勾配はMOPAC2012 [14, 15] におけるPM6で計算された。カチオン吸着サイト探索の問題をGRRMにおける反応経路探索の問題へ対応させるべく、AFIR法の発展形であるmulti-component AFIR (MC-AFIR) 法 [27, 28] が採用された。すなわち、分子およびカチオンを反応物とみなし、それらの間に反応誘起のための人工力が付与され、生成物として分子-カチオン複合体が生成される。人工力はセルロースの酸素原子とカチオンの間に加えられ、model collision parameterは100kJ/molに設定された。探索は、50回連続して同一反応経路が見つけれ出されるまで繰り返された。PM6に基づくMC-AFIR法によるスクリーニング後、GAUSSIAN09 [12] を利用し、得られた構造を初期構造として密度汎関数法による再構造最適化が行われた。計算レベルはB3LYP/LanL2DZとし、ノード内8プロセス並列で実行された。さらに、水の溶媒効果を取り入れるため、polarizable continuum model (PCM) [29, 30] の範囲で密度汎関数法による構造最適化も実行した。

計算データおよびその解析方法についてまとめる。two-step structural searchの結果、安定吸着構造とそのエネルギーが、PM6 (低レベル計算でのスクリーニング) およびB3LYP/LanL2DZ (高レベル計算での精密化) でそれぞれ求まる。エネルギーの増加に伴い昇順で、得られた構造に識別番号を割り当てる。吸着構造の相対的な安定性は、エネルギー

差 $E^{(s)} - E^{(s=1)}$ で評価される。ここで、 $E^{(s)}$ は識別番号sを有する安定構造のエネルギーである。このエネルギー差は気相における安定性評価を与える。水溶液中における統計的性質を記述するため、PCMでの計算データを利用し、温度Tにおける分配関数を次式で定める。

$$Z_{PCM} = \sum_s' \exp(-E_{PCM}^{(s)}/k_B T)$$

ここで、 k_B はボルツマン定数である。和におけるダッシュ記号は、幾何学的に区別される構造についてのみ和をとることを意味する。構造の識別は原子間距離を計算することで実行された。2個の安定吸着構造について、すべての原子間距離が0.1Å以下のとき、それらは等価な幾何的構造とみなす。分配関数 Z_{PCM} で定まるカノニカル分布に従い、温度Tにおいて、最安定吸着構造が見出される確率は次式で与えられる。

$$p_1(T) = \frac{1}{Z_{PCM}} \exp(-E_{PCM}^{(1)}/k_B T)$$

分配関数を利用することで、自由エネルギーや安定構造の統計分布に関するエントロピーも評価することができる [17]。

図2に、PM6によるMC-AFIR法により発見された安定吸着構造の相対エネルギー差、およびB3LYP/LanL2DZレベルの密度汎関数法により再構造最適化された安定吸着構造のそれを示す。横軸に構造の識別番号、縦軸に相対エネルギー差を示す。図中の差し込み図には、低エネルギー側から数えて3番目までのデータを示す。図2から、幅広いエネルギー領域に渡り、カチオンの安定吸着構造が得られたことがわかる。再最適化データ (青三角プロット) に着目する。セルロース二量体について、カチオンの原子番号が増加するに伴い (すなわち、イオン半径が増加するに伴い)、複数の最安定吸着構造が見出される

ことがわかる。K⁺については、0.6kcal/mol (温度換算で300Kに相当) 内に2個の異なる配座が存在した。同様に、Cs⁺については、3個の異なる配座が存在した。それ以外のデータについては、エネルギー的に最も安定な吸着構造がただ一つ発見された。

図3にセルロースにおけるカチオン吸着構造を示す。分子構造の可視化および幾何学的解析には、VMD [31] を利用した。図3中の各パネルの下部には、カノニカル分布から定められた最安定構造の検出確率 $p_1(T)$ を示した。まず、図3(a)に着目する。左および中央パネルに、two-step structural searchにおける典型的な系列、すなわちスクリーニング(PM6)から再最適化(B3LYP/LanL2DZ)の系列を示す。各パネルの下部に示された番号は図2の識別番号に対応する。右パネルに、PCM下での最安定構造を示す。すなわち、PCMにおける最安定構造は、スクリーニングでは11番目の安定構造から来ることがわかる。こうして、密度汎関数法による再最適化は、半経験的分子軌道法で予測された安定性の順序を変更する可能性がある。ここで重要なことは、MOPACと組み合わせたGRRM

の使用は、分子構造の精密な予測を与えるものではなく、高精度計算のための良い初期構造をバイアスなく生成することを目的とする、ということである。

次に、図3(a)の右パネルと図3(b)に着目し、水溶液中における最安定構造の出現確率について考察する。温度は $T=300\text{K}$ に固定して解析をした。PCMで記述される水溶液は気相中の相対構造安定性を変更する。特に、Cs⁺吸着において劇的な変化を予測した。セルロース単量体について、最安定構造を見出す確率は0.6以下であった。すなわち、溶媒効果により吸着構造分布にブロードニングが起きたことを意味する。これは、気相中で単一の安定構造が存在することと対照的である[図2(a)]。他方、セルロース二量体について、溶媒効果によって最安定構造を見出す確率は非常に高くなることが確認された($p_1=0.9$)。これも、気相中では0.6kcal/mol内に複数の最安定構造が存在することと対照的である[図2(b)]。このように、two-step structural searchは、スクリーニングを通して多数の安定構造候補を生成できることから、吸着構造分布を調べることに適している。

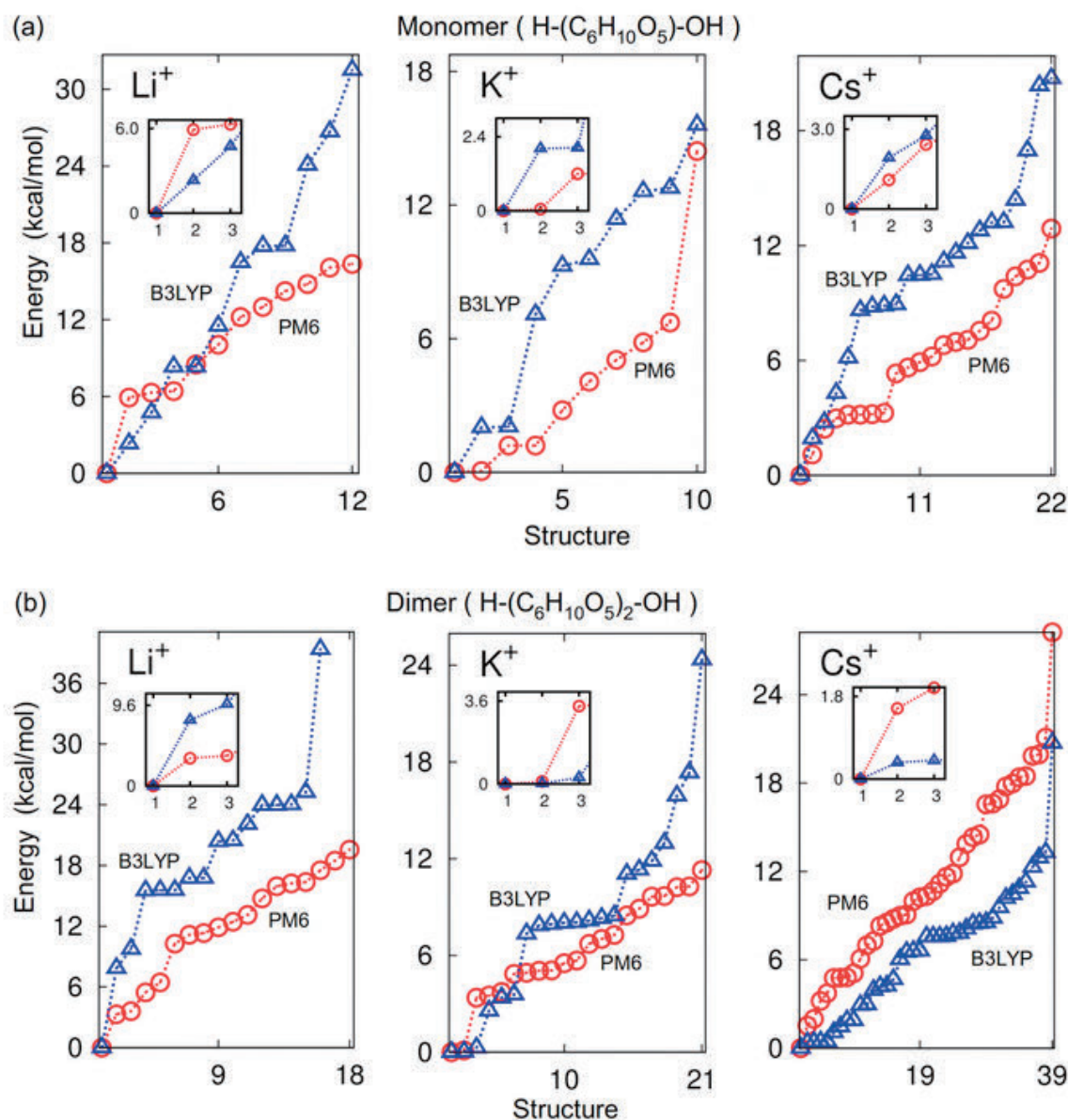


図2 安定吸着構造におけるエネルギー差

赤丸プロットがPM6を用いたmulticomponent Artificial Force Induced Reaction法による結果（スクリーニング）、青三角プロットが密度汎関数法（B3LYP/LanL2DZ）による結果（再最適化）に対応する。(a) セルロース単量体 (monomer)。 (b) セルロース二量体 (dimer)。 Reprinted from Chem. Phys. Lett., 648, Y. Ota, S. Ruiz-Barragan, M. Machida, and M. Shiga, A screened automated structural search with semiempirical methods, 119-123, Copyright (2016), with permission from Elsevier.

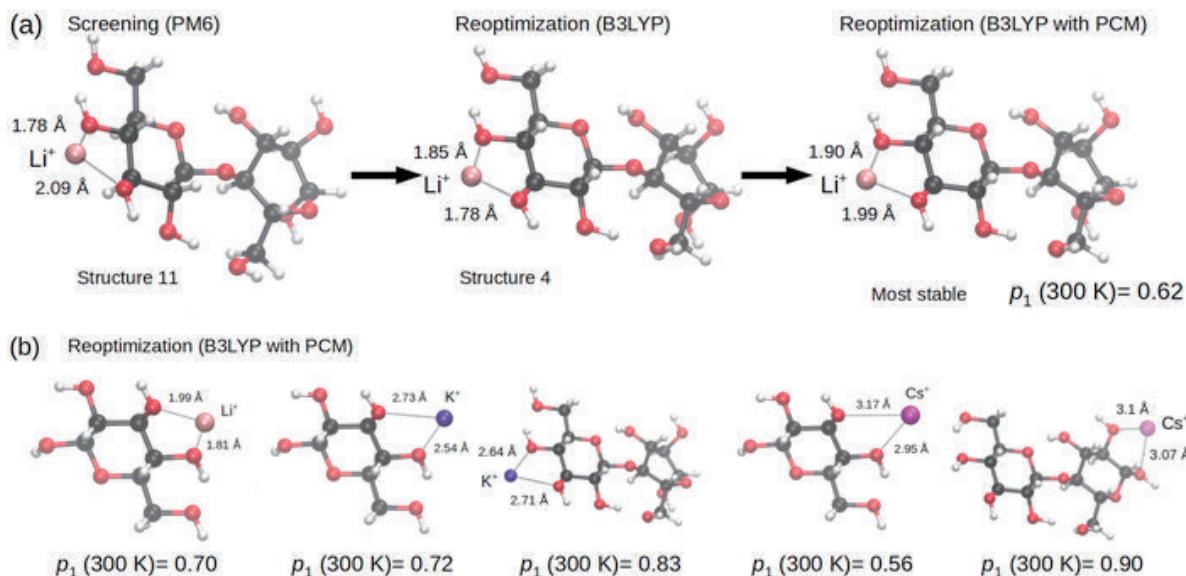


図3 セルロースにおけるカチオン吸着構造

(a) セルロース二量体におけるLi⁺吸着に対する、スクリーニングから最安定構造に至るまでの系列。
 (b) 各種カチオンに対する最安定構造。Reprinted from Chem. Phys. Lett., 648, Y. Ota, S. Ruiz-Barragan, M. Machida, and M. Shiga, A screened automated structural search with semiempirical methods, 119-123, Copyright (2016), with permission from Elsevier.

5. まとめ

中・大規模分子系における反応経路自動探索の効率化を目指し、MOPACに基づいたGRRMの利用を可能とするインターフェイス計算コードを開発した。開発された計算コードにより、GAMESSを経由した半経験的分子軌道法の利用に比べ、より高い効率で半経験的分子軌道計算による反応経路探索が可能となる。開発コードの応用として、two-step structural searchによる安定構造の網羅的探索法を提案した [17]。すなわち、半経験的分子軌道法による安定分子構造の網羅的なスクリーニング後、その結果を初期構造とした高精度計算による再最適化により、安定構造分布を系統的に得ることができる。その手法をセルロースにおけるカチオン吸着サイ

ト探索に適用し、PCMの範囲で溶媒効果を取り入れたカチオンの吸着構造分布を予測する指針を見出した。開発されたインターフェイス計算コードは、半経験的分子軌道法が正当な範囲で、大規模分子系における様々な化学反応経路探索に利用できると期待される。

謝辞

本稿で報告した研究成果は、Sergi Ruiz-Baragan博士 (RUB)、町田昌彦博士 (JAEA)、志賀基之博士 (JAEA) との共同研究 [17] に基づきます。本研究を進めるにあたり、前田理准教授 (北大) より受けた助言に感謝します。本研究における数値計算の一部は日本原子力研究開発機構におけるFujitsu PRIMERGY BX900により実施されました。

参考文献

- [1] 大野公一, 前田理, *Mol. Sci.* 5, A0042 (2011).
- [2] K. H. Houk and P. H.-Y. Cheong, *Nature* 455, 309 (2008)
- [3] S. M. Woodley and R. Catlow, *Nat. Matter.* 7, 937 (2008).
- [4] D. J. Wales, *Energy Landscapes* 2nd ed. (Cambridge University Press, Cambridge, 2003); OPTIM: <http://www-wales.ch.cam.ac.uk/>
- [5] A. O. Lyakhov, A. R. Oganov, H. Stokes, and Q. Zhu, *Comput. Phys. Commun.* 184, 1172 (2013); USPEX: <http://uspex-team.org>
- [6] S. Q. Wu, M. Ju, C. Z. Wang, M. C. Nguyen, X. Zhao, K. Umemoto, R. M. Wentzcovitch, and K. M. Ho, *J. Phys. : Condens. Matt.* 26, 035402 (2014)
- [7] S. Maeda, K. Ohno, and K. Morokuma, *Phys. Chem. Chem. Phys.* 15, 3683 (2013).
- [8] S. Maeda, Y. Harabuchi, Y. Osada, T. Taketsugu, K. Morokuma, and K. Ohno, see <http://grrm.chem.tohoku.ac.jp/GRRM> (accepted 11.04.14).
- [9] S. Ohno, K. Shudo, M. Tanaka, S. Maeda, and K. Ohno, *J. Phys. Chem. C* 114, 15671 (2010).
- [10] R. Uematsu, E. Yamamoto, S. Maeda, H. Ito, and T. Taketsugu, *J. Am. Chem. Soc.* 137, 4090 (2015)
- [11] K. Ohno, H. Satoh, T. Iwamoto, H. Tokoyama, And H. Yamakado, *Chem. Phys. Lett.* 639, 178 (2015).
- [12] Gaussian 09, Revision C.01, M. J. Frisch et al., Gaussian, Inc., Wallingford CT, 2010.
- [13] General Atomic and Molecular Electronic Structure System, M. W. Schmidt et al., *J. Comput. Chem.* 14, 1347 (1993).
- [14] MOPAC 2012, J. J. P. Stewart, Stewart Computational Chemistry, Version 15.180L. <http://www.openmopac.net/index.html>
- [15] J. D. C. Mala, G. A. U. Carvalho, C. P. Mangueira Jr., S. R. Santana, L. A. F. Cabral, and G. B. Rocha, *J. Chem. Theory Comput.* 8, 3072 (2012).
- [16] gr 2 moc: <http://github.com/otayuki/gr2moc>
- [17] Y. Ota, S. Ruiz-Barragan, M. Machida, and M. Shiga, *Chem. Phys. Lett.* 648, 119 (2016).
- [18] F. Jensen, *Introduction to Computational Chemistry* 2nd ed. (John & Wiley, West Sussex, 2007).
- [19] J. E. Dennis, Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (SIAM, Philadelphia, 1996).
- [20] G. Henkelman and H. Jónsson, *J. Chem. Phys.* 113, 9978 (2000).
- [21] W. E. W. Ren and E. Vanden-Eijnden, *Phys. Rev. B* 66, 52301 (2002).
- [22] 松永康佑, *統計数理* 62, 285 (2014).
- [23] S. J. Blanksby and G. B. Ellison, *Acc. Chem. Res.* 36, 255 (2003).
- [24] M. Elstner and G. Seifert, *Phil. Trans. R. Soc. A.* 372, 20120483 (2016).
- [25] H. Tokoyama, H. Yamakado, S. Maeda, and K. Ohno, *Bull. Chem. Soc. Jpn.* 88, 1284 (2015).
- [26] T. Takayanagi, K. Takahashi, A.

- Kakizaki, M. Shiga, and M. Tachikawa, *Chem. Phys.* 358, 196 (2009).
- [27] S. Maeda and K. Morokuma, *J. Chem. Phys.* 132, 241102 (2010).
- [28] S. Maeda and K. Morokuma, *J. Chem. Theory Comput.* 7, 2335 (2011).
- [29] J. Tomasi and M. Persico, *Chem. Rev.* 94, 2027 (1994).
- [30] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* 105, 2999 (2005).
- [31] W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graphics* 14, 33 (1996).