

科学研究費助成事業 研究成果報告書

平成 26 年 5 月 26 日現在

機関番号：13301

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23790137

研究課題名(和文)多目的最適化と対話型最適化を利用した蛋白質-リガンドドッキングプログラムの開発

研究課題名(英文)Development of the protein-ligand docking program using multi-objective optimization and interactive optimization

研究代表者

小田 彰史(Oda, Akifumi)

金沢大学・薬学系・准教授

研究者番号：50433511

交付決定額(研究期間全体)：(直接経費) 3,300,000円、(間接経費) 990,000円

研究成果の概要(和文)：本研究ではタンパク質-リガンドドッキングにおいて多目的最適化の概念を導入し、複数のスコア関数を使用してドッキングポーズを選別するための技術を開発した。またドッキングの際に生じる偽陽性を取り除くために、ドッキング以外の方法を組み合わせて選別を行う対話型最適化の概念についても導入した。従来のコンセンサスコア技術では使用するスコア関数の選別を注意深く行う必要があるが、多目的最適化は原理上スコア関数を単純に増やすほど正解を見落とすことがなくなるという利点がある。一方で増加することになる偽陽性については対話型最適化によって削減するというドッキング戦略を構築した。

研究成果の概要(英文)：In this study, multi-objective optimization was adopted for protein-ligand docking calculations. Multiple scoring functions were simultaneously used, and the Pareto optima were selected as the solutions of docking trials. In addition, the concept of the interactive optimization was also used, because the reduction of the number of the false positives is one of the most important problems in docking calculations. In interactive optimization, false positives were eliminated by using several criterion, such as experimental data and estimated pharmacophore obtained by ligand-based approach. The combination use of multi-objective optimization and interactive optimization was tested for large test set, and was applied for actual drug design trials.

研究分野：医歯薬学

科研費の分科・細目：薬学・創薬化学

キーワード：インシリコ創薬 タンパク質-リガンドドッキング 多目的最適化 対話型最適化 コンセンサスコア

1. 研究開始当初の背景

計算機を利用した創薬において、タンパク質 - リガンド複合体構造は重要な役割を果たす。実験的に複合体立体構造が得られている場合は当然それを利用することができるが、創薬初期の段階においては複合体の立体構造はほとんどの場合で得られていない。そのため、計算機的にタンパク質 - リガンド複合体構造を予測するドッキングが用いられることがほとんどである。

計算機によるドッキングでは、タンパク質 - リガンド複合体の立体構造候補を多数作成するドッキングと呼ばれるステップと、作成した候補構造を評価し、多くの候補構造の中から少数(典型的には1つ)の構造に絞り込むスコアリングと呼ばれるステップの2つの段階がある。現時点までに計算機によるドッキングのための手法は多数考案されており、構造を構築するドッキングの段階についてはある程度の精度での実行が可能な状況になっている。たとえば一切工夫することなく、ソフトウェアにあらかじめ決められた設定をそのまま使用するだけでも、70%程度の系で実験構造に近いタンパク質 - リガンド複合体構造を得ることができる。それに対してスコアリングについては現在に至るまで決定版となる方法は存在せず、試行錯誤が繰り返されている状態である。スコアリングではタンパク質 - リガンド複合体構造候補を評価するのが目的であるため、結合自由エネルギーを評価する必要がある。現実の複合体形成では、複合体を形成することによって最も自由エネルギーが低下するため、この結合自由エネルギーが候補構造の評価に用いられる。しかし一方で自由エネルギー計算は統計力学的処理が必要なため、精密に行うには極めて多大なコンピュータ資源を必要とする。一般に計算機によるドッキングは数分程度で終わらせる必要があるため、そのような自由エネルギー計算を行うのは非現実的である。そこで結合自由エネルギーを簡易的に算出するスコア関数が用いられることになる。これまでに様々なスコア関数が提案されているが、大きく分けると力場に基づくスコア関数、経験的スコア関数、知識に基づくスコア関数の3種類に分類される。しかしいずれも一長一短あり、様々な系に対して常に有効なスコア関数は見いだされていない。

このような状況から、複数のスコア関数を組み合わせて使用するコンセンサススコア概念が提唱されている。これは複数のスコア関数を組み合わせて新たな評価関数を作成し、それによって複合体候補構造を評価するというもので、互いの欠点を補う手法として注目されている。研究代表者らは以前の研究でコンセンサススコアの有用性を評価し、それを様々な系に適用してきた。しかしその有用性一方で、コンセンサススコアの能力はどのようなスコア関数を使用するかによって強く依存するため、スコア関数の選別に注意を

払う必要がある。スコア関数を正しく選択しないと単独で使用した場合よりも能力が低下することさえあり、コンセンサススコアの使用には注意が必要となる。またドッキング全般に付随する問題として偽陽性(実際の複合体構造とは異なっているにもかかわらず、スコアの値が良好となってしまいうような候補構造)の出現があるが、コンセンサススコアの中でも特に有用な vote を用いた手法ではこの偽陽性が主な問題である。

2. 研究の目的

本研究では、多目的最適化の技法を計算機によるドッキング、特にスコアリングの過程に適用し、スコアリングの成功率を上げることを目的とする。さらに偽陽性の問題に対して対話型最適化の方法を用いた解決を目指す。多目的最適化は、複数の目的関数を最適化するような問題を解くための方法である。計算機によるドッキングの場合はスコア関数が目的関数となるため、複数のスコア関数を使用したドッキングとなる。その点ではコンセンサススコアと同様であるが、コンセンサススコアではスコア関数を組み合わせて新たな関数を作成するため、目的関数としては1つである。それに対して多目的最適化では複数のスコア関数を複数のままで最適化を行う。また、対話型最適化ではドッキングによって得られた複合体候補構造を、スコア関数以外の方法によって選別する。通常実際の創薬の現場においては、薬物標的および医薬候補化合物について何らかの情報が得られていることが多い。そのような場合、純粋にドッキングのみで複合体構造を得ようとするのではなく、他の情報と合わせて複合体構造の構築・選別を行うほうが合理的である。特に多目的最適化では偽陽性が多くなることが想定されるため、それを削減するためにも様々な情報を考慮して対話的にタンパク質 - リガンドドッキングを行うほうが現実的であろうと考えられる。多目的最適化に加えてこの対話型最適化を用い、タンパク質 - リガンドドッキングを効率的に行うための手法を確立することを目標とする。

3. 研究の方法

まず、タンパク質 - リガンドドッキングに関する手法の整備を行う。手法の確立に際してはまず既存の *in silico* 創薬プログラムを用いて検討を行うが、それら既存プログラムの有用性の評価を行う。特にドッキングプログラムについては初期構造や計算の設定に依存して結果が変化することが予想されるものの、それらに踏み込んだ評価がこれまでになかったため、新規手法を確立する上でまずそれらを調査する必要がある。また研究代表者が共同研究を行っているグループの開発したソフトウェアに対して、それらを新たな手法に組み込むことが可能であるかどうかのベンチマーク試験も行う。

既存手法の有用性および限界を明らかにした後、多目的最適化・対話型最適化のためのスクリプトを作成する。既存の手法が使用できる部分については使用し、それらをつなぐためのスクリプトを構築する。その後ベンチマーク試験を行ったテストセットを用いて、新たな手法の有用性について評価を行う。その際には、通常のスコア関数の評価と同様に選んだ候補構造が妥当な構造（一般に結晶解析によって得られた実験構造との root mean square deviations (RMSDs) が 2.0 Å 以下の構造）であるかどうかだけでなく、その妥当な構造を得るためにどれだけの解を選別したかについても評価する。これは偽陽性の数と対応しており、多目的最適化の技術で増加することが予想される偽陽性をコントロールする上で必要となる。

手法を構築した後は、実際の系に適用する。ホスホイノシチド-3-キナーゼ (PI3K) 阻害剤の開発において必要な PI3K と阻害剤との複合体構造の予測や、シトクロム P450 (CYP, P450) と阻害剤とのドッキング、また生体内に生じた D-アミノ酸残基を修復するとされるタンパク質イソアスパラギン酸 D-アスパラギン酸メチルトランスフェラーゼ (PIMT) と D-アスパラギン酸含有ペプチドとの複合体形成の推定を行う。研究代表者は共同研究者とともに PI3K 阻害剤の開発を行っているが、その過程でヒストンデアセチラーゼ (HDAC) 阻害剤であるロミデプシンが PI3K 阻害剤としての機能を持っていることを解明した。そこで PI3K とロミデプシンとの複合体構造を推定する必要があったが、ロミデプシンは大環状化合物であり、その配座自由度のため単純なドッキングでは適切な解が得られない恐れがある。そこで新たに開発した手法を用い、実験データや類縁化合物のドッキング結果などを参考にした対話型操作によって妥当と考えられる複合体構造を得た。また P450 についても実験者と共同で遺伝多型と基質代謝との関連について研究を行っているが、CYP2B6 の遺伝多型が薬物代謝に与える影響について分子動力学 (MD) シミュレーションとドッキングによって検討を行った。これについても野生型での結果を生かした対話型最適化の方法を利用した。さらに PIMT と D-アスパラギン酸含有ペプチドとのドッキングについても行った。これらいずれに対しても MD シミュレーションによる構造精密化を行っているが、これはドッキングのみでは構造の歪みなどが残されているためである。MD シミュレーションにおいては古典的分子力場が必要となり、アミノ酸については十分用意されているが、補酵素やリガンドについては適切なパラメータが存在しない場合がある。そこで MD シミュレーションを行う上で必要な力場パラメータについても、モデル化合物に対する量子化学計算によって算出した。

4. 研究成果

多目的最適化を計算機によるドッキングに適用した。多目的最適化では、「他のどれかの解と比較して、すべてのスコア関数で劣っている」解だけを排除し、「他のすべての解と 1 対 1 で比較したとき、どんな解に対しても最低一つのスコア関数で劣っていない」解を残す。これをパレート最適解（非劣解）と呼び、これが多目的最適化の解となる（図 1）。図 1 において解 1 は目的関数 A に関して最適、解 4 は目的関数 B に関して最適であるため、これらはいずれもパレート最適解である（他のどんな解と比較しても、解 1 は目的関数 A に関して、解 2 は目的関数 B に関して、常に優れているため）。解 2 は解 1 と比べると、目的関数 A に関しては劣っているものの、目的関数 B に関しては解 1 より優れている。また、解 2 を解 3 や 4 と比べると、目的関数 B では劣っているものの、目的関数 A に関しては優れている。すなわち、目的関数 A と B の両方で解 2 を上回るような解は存在せず、解 2 もパレート最適解となる。それに対して解 5 は解 3 と比べて目的関数 A, B のいずれも劣っており、パレート最適解ではない（解 5 と比べて完全に優れているのは解 3 のみだが、それでもパレート最適解ではなくなる）。このパレート最適解の概念をスコアリングに応用した。すなわち、複数のスコア関数を用いてパレート最適解集合を求め、それをスコアリングの結果として用いる方法を採用した。

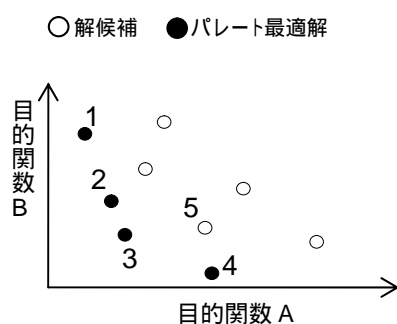


図 1 パレート最適解

多目的最適化の試験として、我々が以前リガンド結合部位探索プログラムの評価に用いたタンパク質 - リガンド複合体構造既知で Protein DataBank に記載されている 458 個の系に対してテストを行った。ドッキング手法として FRED を用い、10 種類のスコア関数の全ての組み合わせに対してテストを行った。各テストセットに対して、妥当な解が 1 つでもパレート最適解の中に入っていれば「成功」と定義した。その結果を図 2 に示す。パレート最適解の定義上スコア関数（目的関数）を増やせば増やすほど成功した系の数は増えていくが、この図に示したとおりスコア関数の数が増えるにつれて成功数の増加は鈍化しており、8 で一定値に到達する。

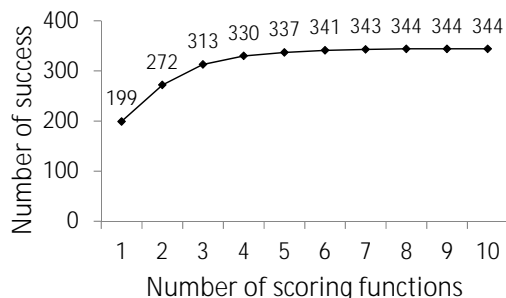


図2 ドッキングに成功した系の数

パレート最適解の数について図3に示す。ここに示したように、目的関数であるスコア関数の数を増やせば増やすほど偽陽性の数が増え、真の「正解」が見だしにくくなる。これは多目的最適化問題に一般的に起こる問題であり、これを解決するために対話型最適化など、多数のパレート解から特定の解を選択するための方法が必要となる。ドッキングの問題について言えば何らかの構造生物学的な実験データや、リガンドに基づいた創薬(LBDD)の手法によるファーマコフォアモデルなどを用いてドッキングポーズを選択することになる。その場合でも偽陽性が多いと正解の選択が困難になることが予想されるため、パレート最適解の数が極端に多くなく、かつ成功率も高い設定を選択する必要がある。スコア関数の数と成功数とのバランスについて検討し、偽陽性の数を少なく保ちつつ成功率を上げるようなスコア関数の組み合わせについて調査した結果、スコア関数の数を4つから5つに限定して用いた場合、比較的偽陽性を抑えつつ成功率が上げられることが判明した。また、スコア関数が4つか5つであればコンセンサススコアほどスコア関数の組み合わせを詳細に検討しなくてもある程度良好な結果が得られることも分かった。しかしスコア関数の組み合わせに悩まされることなく偽陽性を取り除くためには、対話型の操作が必要となる。

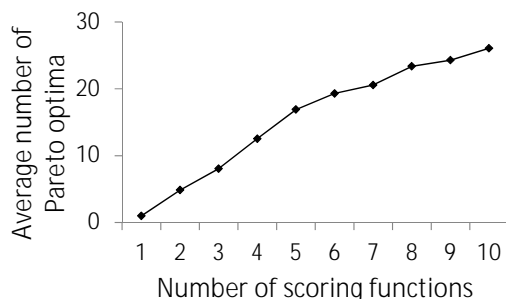


図3 パレート最適解の数の平均値

そこで対話型最適化を用いた場合のドッキングについて検討するため、実際の系に対してドッキング以外の情報も用いたドッキングを行った。PI3Kに対して、阻害作用を示したロミデプシンとFK-A5をドッキングし、

それらの構造を比較した。FK-A5はロミデプシン類縁体であり、かつどちらもPI3K阻害作用、HDAC阻害作用の両方を持つ。これらのことから、この両者は類似した結合様式でPI3Kに認識されるのではないかと推測される。そこでFK-A5の複合体候補構造とロミデプシンの複合体候補構造を網羅的に重ね合わせ、最もよく重なった構造が正解に近いのではないかと推定する。これはLBDDの思想に基づいた選別法であるが、ドッキングによって得られた多数の解候補の中から少数に絞り込む上でLBDDが使用できるかどうかを判断した。計算の結果、十分な重なりをもってFK-A5とロミデプシンが重なったのは1組のみであり、それを正解候補として選別した(図4)。この構造では阻害剤はPI3Kの触媒サブユニットのポケットの奥深くに位置し、多くの疎水性残基によって囲まれていた。またポケットには芳香族アミノ酸が存在しており、それがFK-A5の芳香環と相互作用している可能性が示唆された。PI3Kに対する阻害活性はFK-A5のほうが高いが、この認識の違いが阻害活性につながっている可能性がある。

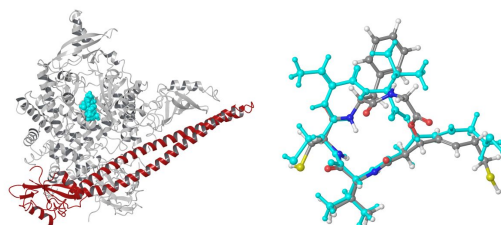


図4 得られた複合体構造

左: PI3K-ロミデプシン複合体、右: ロミデプシン(水色)とFK-A5(原子ごとに色分け)の結合構造の比較

また、CYP2B6と基質アーテメターとの複合体構造についても、多数得られた構造の中から実験データを再現しうるもののみを抽出して評価した。CYP2B6は遺伝多型によって基質の代謝活性を失うことがしばしばあるが、そのような変異体のうちいくつかでは適切なタンパク質-リガンド複合体構造をドッキングによって得ることができなかった。ドッキングによって得られた解自体は多数存在していたとしても、ドッキング外の情報を入れることでそのような結果となることもあり、対話型操作の重要性が示唆される。PIMTとD-アスパラギン酸含有ペプチドとのドッキングに関しては、基質ペプチドとして6残基ペプチドVYPDGAを用いて行った。これについてはアスパラギン酸残基としてL- α -Asp, D- α -Asp, L- β -Asp, D- β -Aspの4種類を考慮し、その全てについて複合体構造を求めた。この際にもPIMTがD- α -AspとL- β -Aspのみを基質とすることを考慮して複合体構造を選択したが、実験結果の通りL- α -Asp含有ペプチドとD- β -Asp含有ペプチドは構造が類似しているにもかかわらず複

合体構造が異なっていた。この結果は得られた構造が妥当であることを示唆している。

本研究では多目的最適化・対話型最適化を計算機によるドッキングに組み込んだが、上述のベンチマークテストおよび実際の系に対する適用結果から、そのストラテジーの有用性が示された。従って、本手法を用いることで偽陽性の問題を回避しつつ妥当なドッキング結果を得ることができるようになるのではないかと期待できる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計8件)

A. Oda, N. Yamaotsu, S. Hirono, Y. Watanabe, S. Fukuyoshi, O. Takahashi, Effects of initial settings on computational protein-ligand docking accuracies for several docking programs, *Mol. Simul.*, 査読有, in press.

K. Kobayashi, O. Takahashi, M. Hiratsuka, N. Yamaotsu, S. Hirono, Y. Watanabe, A. Oda, Evaluation of influence of single nucleotide polymorphisms in cytochrome P450 2B6 on substrate recognition using computational docking and molecular dynamics simulation, *PLOS ONE*, 査読有, 9(5), e96789 (2014).

DOI: 10.1371/journal.pone.0096789

A. Oda, S. Fukuyoshi, R. Nakagaki, O. Takahashi, Multi-objective optimizations for selections of protein-ligand docking poses using Pareto optima as a consensus score, *JPS Conf. Proc.*, 査読有, 1, 016005 (2014).

DOI: 10.7566/JSPSC.1.016005

小田彰史, 結合型 D-アミノ酸についての計算機的研究, 査読無, D-アミノ酸学会誌, 2(1), 1-6 (2014).

A. Oda, S. Fukuyoshi, R. Nakagaki, Structural prediction of [GADV]-proteins using threading and ab initio modeling for investigations of the origin of life, *J. Comput. Aided Chem.*, 査読有, 14, 23-35 (2013).

DOI: 10.2751/jcac.14.23

A. Oda, S. Fukuyoshi, R. Nakagaki, O. Takahashi, Determination of AMBER force field parameters for thioester by quantum chemical calculations, *Chem. Lett.*, 査読有, 42(10), 1206-1208 (2013).

DOI: 10.1246/cl.130517

A. Oda, N. Yamaotsu, S. Hirono, Y. Takano, S. Fukuyoshi, R. Nakagaki, O. Takahashi, Evaluations of the conformational search accuracy of CAMDAS using experimental three-dimensional structures of protein-ligand complexes, *J. Phys.: Conf. Ser.*, 査読有, 454, 012028 (2013).

DOI: 10.1088/1742-6596/454/1/012028

A. Oda, I. Noji, K. Kobayashi, S. Fukuyoshi, R. Nakagaki, O. Takahashi, Computational

studies on the repair mechanism of peptides including isomerized aspartic acid residues by protein L-isoaspartyl (D-aspartyl) O-methyltransferase, *Peptide Sci.*, 査読有, 253-256 (2012).

[学会発表](計18件)

小田彰史, 西條憲, 石岡千加史, 加藤正, 福吉修一, 中垣良一, 高橋央宜, PI3K が HDAC 阻害剤ロミデプシンを認識する機構の分子動力学シミュレーションによる解析, 日本薬学会第 134 年会, 熊本大学 (熊本) 2014.3.29.

A. Oda, N. Yamaotsu, S. Hirono, Y. Watanabe, S. Fukuyoshi, R. Nakagaki, O. Takahashi, Effects of Initial Settings on Computational Protein-Ligand Docking Accuracies for several Docking Programs, 3rd International Conference on Molecular Simulation, Kobe International Conference Center (Kobe) 2013.11.18.

小田彰史, タンパク質中 D-アミノ酸残基の in silico 解析, ERATO 浅野プロジェクト講演会・酵素化学工学セミナー, 富山県立大学 (射水) 2013.11.1.

小田彰史, 福吉修一, 中垣良一, 高橋央宜, チオエステル周辺の AMBER 力場の開発および評価, 第 51 回日本生物物理学会年会, 京都国際会議場 (京都) 2013.10.29.

小田彰史, タンパク質中に残基として含まれる D-アミノ酸をコンピュータで取り扱う, 第 9 回 D-アミノ酸研究会学術講演会, 関西大学 (吹田) 2013.9.6.

A. Oda, I. Noji, K. Kobayashi, S. Fukuyoshi, R. Nakagaki, O. Takahashi, Computational studies on substrate recognition mechanism of protein L-isoaspartyl (D-aspartyl) O-methyltransferase, International Conference on Structural Genomics 2013 -Structural Life Science-, Keio Plaza Hotel Sapporo (Sapporo) 2013.7.31.

A. Oda, S. Fukuyoshi, R. Nakagaki, O. Takahashi, Novel consensus scoring strategy for evaluations of protein-ligand docking poses using multiobjective optimization, The 12th Asia Pacific Physics Conference, Makuhari Messe (Chiba) 2013.7.16.

小田彰史, 西條憲, 石岡千加史, 加藤正, 福吉修一, 中垣良一, 高橋央宜, PI3K が HDAC 阻害剤ロミデプシンを認識する機構についてリガンドベース手法と構造ベース手法を組み合わせて計算機的に解明する, 日本薬学会第 133 年会, パシフィコ横浜 (横浜) 2013.3.29.

A. Oda, N. Yamaotsu, S. Hirono, Y. Takano, S. Fukuyoshi, R. Nakagaki, O. Takahashi, Evaluations of conformational search accuracy of CAMDAS using experimental three-dimensional structures of protein-ligand complexes, Conference on Computational

Physics 2012, Nichii Gakkan conference center (Kobe) 2012.10.17.

A. Oda, K. Saijo, C. Ishioka, T. Katoh, S. Fukuyoshi, R. Nakagaki, O. Takahashi, Prediction of the complex structure between PI3K and romidepsin (FK-228) for drug design of PI3K/HDAC dual inhibitors, 生命医薬情報学連合大会, タワーホール船堀 (東京) 2012.10.15-17.

小田彰史, 高橋央宜, 福吉修一, 中垣良一, タンパク質 - リガンドドッキングにおける多目的最適化の有用性の評価, 日本コンピュータ化学会 2012 秋季年会, 山形大学 (山形) 2012.10.14

小田彰史, 高橋央宜, 残基として D-アミノ酸を含む系の分子力学計算のための力場パラメータの評価, 第 12 回日本蛋白質科学会年会, 名古屋国際会議場 (名古屋) 2012.6.22.

小田彰史, 小林佳奈, 高橋央宜, ヒト PIMT が L-β-Asp と D-α-Asp の両方を認識する機構についての計算機的研究, 日本薬学会第 132 年会, 北海道大学 (札幌) 2012.3.30.

小田彰史, 小林佳奈, 高橋央宜, 異性化アスパラギン酸を含むペプチドに対する分子力学計算の妥当性の評価, 第 39 回構造活性相関シンポジウム, 東京理科大学 (野田) 2011.11.28.

A. Oda, O. Takahashi, Pareto optimal consensus score for evaluations of protein-ligand complex structures, CBI/JSBi 2011, Kobe international conference center (Kobe) 2011.11.8.

小田彰史, アミノ酸残基の立体反転についての計算機的研究, 2011 年度 C A C フォーラム一泊研修会, リゾーピア熱海 (熱海) 2011.10.7.

小田彰史, 小林佳奈, 高橋央宜, 非金属元素に対する高次項を含んだ電荷平衡法, 平成 23 年度化学系学協会東北大会, 東北大学 (仙台) 2011.9.17.

A. Oda, Y. Takano, O. Takahashi, Validation of quantum chemical methods for geometrical optimizations of sulfonamide derivatives, XVIth International Workshop on Quantum Systems in Chemistry and Physics, Ishikawa Prefecture Museum of Art (Kanazawa) 2011.9.12-13.

[図書] (計 2 件)

A. Oda, et al., Nova Science Publishers, Inc., Molecular Dynamics: Theory, Kinetics and Implementation (D. E. Garcia, et al., Eds.), 2012, chapter 2, 25-39.

A. Oda, et al., Quantum Systems in Chemistry and Physics: Progress in Methods and Applications (K. Nishikawa, et al., Eds.), 2012, chapter 17, 331-344.

[その他]

ホームページ等

<http://www.p.kanazawa-u.ac.jp/~bukka/>

6 . 研究組織

(1)研究代表者

小田 彰史 (ODA, Akifumi)

金沢大学・医薬保健研究域薬学系・准教授

研究者番号：50433511