

令和元年6月4日現在

機関番号：32666

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00059

研究課題名(和文) 動的データに内在するレアイベント性の新規な特徴付けと生体時系列への応用

研究課題名(英文) Novel characterization of rare events in dynamic data and its application to biological time series

研究代表者

藤崎 弘士 (Fujisaki, Hiroshi)

日本医科大学・医学部・准教授

研究者番号：60573243

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：動的なデータからレアイベントと呼ばれる稀な現象を抽出し、その特徴づけを行うのは難しい問題であり、シミュレーションでデータを生成するときにも同じような問題が現れる。生体分子の時系列を分子動力学シミュレーションで生成する場合は安定な状態間を乗り移る構造変化がレアイベントに対応し、上記の問題が現れるために、ここでは重み付きアンサンブル法と呼ばれる超並列計算手法を用いることで、その問題を解決できるか調べた。具体例としてはシニョリンという小さいタンパク質に適用し、動的な構造変化を効率よく生成できることが示された。また、多様体学習の拡散マップ法を用いて、構造変化に重要な反応座標を抜き出すこともできた。

研究成果の学術的意義や社会的意義

現在は非常に大量のデータが生成・収集され、それを機械学習などの統計的な手法を用いて解析することが標準的になりつつあるが、レアイベントに関しては稀に起こる現象であるためにデータ点が少なく、その生成や収集が難しかった。しかし、本研究で、重み付きアンサンブル法という超並列計算手法を用いることで、レアイベントを生体分子に関しては効率よく生成することは可能であることが示された。また、重み付きアンサンブル法は統計学の粒子フィルター法と似た手法であり、重みのついた軌道のアンサンブルを用いているので、生体分子と関係のない時系列にも適用可能であり、本研究の成果は一般性をもつと言える。

研究成果の概要(英文)：It is a difficult problem to extract rare phenomena called rare events from dynamic data and to characterize them, and similar problems appear when generating data in simulation. When generating a time series of biomolecules by molecular dynamics simulation, structural change between stable states corresponds to a rare event, and the above problem appears. To solve this problem, we here employ a massively parallel computing method called weighted ensemble method. As a specific example, the method was applied to a small protein called chignorin, and it was shown that dynamic structural changes can be efficiently generated. In addition, we could extract reaction coordinates that are important for structural change using the diffusion map method of manifold learning.

研究分野：化学物理・生物物理

キーワード：分子動力学シミュレーション レアイベント 反応座標 多様体学習

1. 研究開始当初の背景

現在はデータが過剰に存在し、生成される時代である。その巨大なデータを記憶・解析するために、記憶媒体（ハードディスクなど）の容量も増えてきており、コンピュータも高速化している。あまりにもデータ量が多いために、その処理を人手を介さず自動的に行うという研究も盛んであるが、自動化以前にさまざまなアイデアを試す必要がある。画像解析のような「静的なデータ」の処理はかなり自動化できるようになっているが、より挑戦的で応用の幅も広いのは「動的なデータ（時系列）」を処理することである。その際に、動的なデータからどのように有益な量を取り出すか、もしくはデータを圧縮するかが問題になる。静的なデータに対しては主成分分析、クラスター分析、カーネル法といった標準的な方法が知られており、これらによってデータを分類したり、圧縮したりすることができる。しかし、こういった手法をそのまま動的なデータに用いることは難しい。というのも、データ圧縮とは「似たもの」を抽出する手法なのに対して、動的なデータはそういった「似たもの」の空間から抜け出すダイナミクスを有し、また時間の順序を保ったまま解析しなければならないからである。動的なデータの特徴付けに関しては、例えば再帰型ニューラルネットワーク(recurrent neural network)を用いる手法などが発展してきているが、上記の「抜け出す」ダイナミクスはレアイベントと総称される稀に起こる現象であり、そのサンプリングや解析に関してはまだまだ問題が多く、挑戦的な課題である。

一方、動的なデータの特徴付けということに関しては、いわゆる力学系の分野や、分子動力学(Molecular Dynamics, MD)法を用いた研究分野においてさまざまな手法が発達してきている。代表者や分担者は本研究開始までの10年以上に渡りこれらの分野で研究をしてきており、生体分子における重要な経路の探索、結合カオス力学系における同期現象の発見といった成果を上げている。しかし、それらの手法には経験的な要素も多く、大自由度の動的なデータにレアイベントが含まれる場合には一般的に使えるものではない。そこで、ここでは分子動力学(化学物理・生物物理)を専門とする代表者と、力学系や機械学習を専門とする分担者が共同で、動的なデータの特徴付けの新技术を考案し、それをそれぞれの分野の具体例(生体分子の時系列や、ニューラルネットワークモデルの時系列)へ適用することを試みた。

2. 研究の目的

まず動的なデータ(時系列)の例として、MD計算によってコンピュータから得られる生体分子のデータを考える。現在はタンパク質やDNAなどの生体分子のシミュレーションは比較的容易に行えるようになっているが、その結果得られる時系列の動的なデータとしての側面はあまり研究されていない(長時間平均の振る舞いなどを考えることが多い)。ただし、その動的なデータの特徴付けは、生体分子の機能、ひいては動的な生命現象を考えるうえでも重要であり、本研究の手法を用いる対象として適している。

まずシニョリン(chignolin)と呼ばれる10残基のペプチド(ユニークな安定状態をもつことからミニタンパク質とも呼ばれる)の構造変化の時系列を研究対象とする。この系に関しては既にさまざまな実験や計算の先行研究があり、例えば、森次圭氏(横浜市大)らはこの系の自由エネルギー面の性質を独自のサンプリング手法を用いて調べ、光武亜代理氏(明治大学)らは1マイクロ秒ほどのMD計算から緩和モード解析(relaxation mode analysis)を行って構造変化に関連する反応座標を抜き出した。ただし、まだこの系の構造変化のダイナミクスもしくはキネティクス(速度論)に関しては十分に調べられていない。そこで、本研究では重み付きアンサンブル法を用いて、この系のキネティクスについて調べる。また上記の先行研究にある緩和モード解析は線形の解析であるので、非線形の効果も含む拡散マップ法を用いた解析を行う。その結果として、この系における反応座標の問題、もしくは次元削減の問題に関しても、新たなアプローチを行うことができると考えた。

3. 研究の方法

(1) 重み付きアンサンブル法

本研究では構造変化する生体分子の動的なデータを生成するために、重み付きアンサンブル法(weighted ensemble method)を用いる。これは動的な(非平衡な)軌道のアンサンブルを得るための手法であり、途中で自由エネルギー障壁のようなボトルネックがある場合でも有効に使える方法である。図1にあるように、あるオーダーパラメータで指定される状態空間を用意し、それを複数のセル(ピン)に区切る。そして、通常のMD計算をするが、異なるセルに移ったときに、その軌道を分割(もしくは消滅)させる。ただし、それらの軌道には「重み」がついており、各セル

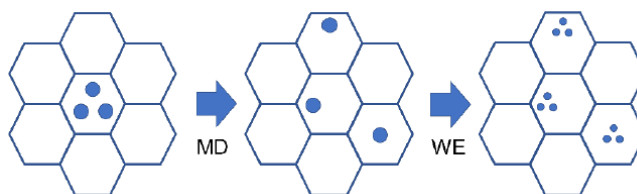


図1 重み付きアンサンブル法のアルゴリズム

のトータル確率（ポピュレーション）が同じになるように変化させる。これによって、状態空間の確率の発展は正しくシミュレートされ、なおかつセルの間にバリアーがある場合でも、重みを小さく変えることで軌道は発展していくので、構造変化のようなレアイベントの軌道を効率よく生成できる。この手法は G.A. Huber, S. Kim (Biophys. J. **70** (1996) 97-110) によって提案されたが、最近では Daniel M. Zuckerman のグループが精力的にこの手法を研究を行っている (Annu. Rev. Biophys. **46** (2017) 43-57)。

(2) 拡散マップ法

生体分子の反応座標を求めるために、ここでは近年提案された多様体学習 (manifold learning) 法の一つである、拡散マップ法 (diffusion map method) を用いる。多様体学習とは高次元の空間に散らばったデータ点において連続的な構造（多様体）を抜き出すための手法であり、これらの手法によってデータの次元削減（次元圧縮）が可能となる。その結果、データの可視化が容易になり、また現象の本質的な自由度を抜き出すことができる。具体的には図 2 にあのように、構造のデータ x_i から、ガウシアン・カーネル $k(x, y)$ を定義し、それを適当に規格化して $M(i, j)$ とした後に対角化する。その結果として抜き出される座標 $u_1(x_i), u_2(x_i), u_3(x_i), \dots$ が拡散マップの座標と呼ばれ、それらを用いてデータを低次元に射影することができる。本研究ではタンパク質の重い原子のみを構造データとして用いて（溶媒分子は無視して）、拡散マップ法を適用した。

Gaussian kernel:
$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\epsilon}\right)$$

“Transition” Matrix:
$$M(i, j) = \frac{k(\mathbf{x}_i, \mathbf{x}_j)}{\sum_j k(\mathbf{x}_i, \mathbf{x}_j)}$$

Solve the Eigenvalue Problem:
$$Mu = \lambda u \begin{cases} u_1(\mathbf{x}_i), \lambda_1 = 1 \\ u_2(\mathbf{x}_i), \lambda_2 < \lambda_1 \\ u_3(\mathbf{x}_i), \lambda_3 < \lambda_2 \\ \dots \end{cases}$$

図 2 拡散マップ法のアルゴリズム

(3) マイルストーン法 (milestoning) とノンマルコフな解析法 (non-Markov analysis)

これらは状態空間をセルに区切ったときに、その間の遷移確率を計算して、そのデータに基づき大域的な遷移率やセルにおける平衡分布などを計算する手法である。現在は生体分子のキネティクスを扱うためにはマルコフ状態モデル (Markov State Model, MSM) が非常によく使われている（例えば、以下を参照: G.R. Bowman, V.S. Pande, F. Noe, (editors), An Introduction to Markov State Models and Their Applications to Long Timescale Molecular Simulation, Springer 2014）が、それに対する別のアプローチとみなせる。これらの手法も重み付きアンサンブル法と同じく状態空間におけるセルを用いるので、大域的には非常に時間のかかるゆっくりとした過程であっても、効率よく時定数を計算できるという利点をもつ。また数学的には非常に単純な原理に基づいているので計算も簡単である。

4. 研究成果

(1) まず先行研究との対応を考えるために、シニョリンに関する amber 力場で計算した 1 マイクロ秒の時系列を光武亜代理氏から提供していただいた。それを用いて、上で述べたマイルストーン法とノンマルコフな解析法を行い、1 本の軌道から得られるキネティクスの情報を調べた。この場合のキネティクスとは、安定な状態からミスフォールドした状態への遷移レート（もしくはその逆数である平均遷移時間）を意味する。その結果は状態をどのように決めるかというパラメータ依存性があるが、420K という高温（転移温度に相当）において、10 ナノ秒ほどの値が得られた（図 3 の左）。

(2) また同じ軌道を用いて拡散マップ法の計算も行った。その結果を自由エネルギー面として表示すると図 3 の右ようになった。これから拡散マップによって得られる低次元の座標も反応座標の候補として使えることが分かった。また、面白いことに、拡散マップの座標と光武氏らにより計算された緩和モード解析から得られる座標を比較して

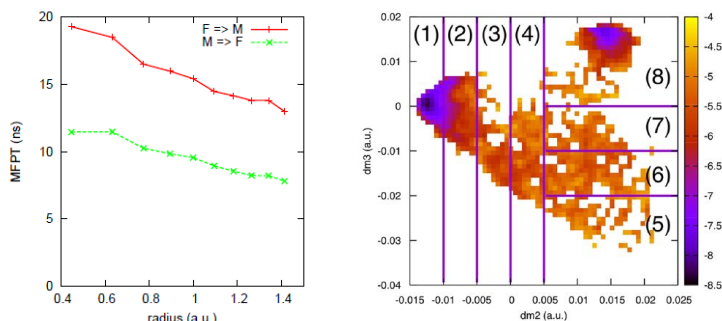


図 3 シニョリンの F 状態から M 状態への構造変化に要する平均時間のパラメータ依存性（左）。拡散マップ法によって得られた座標を軸としたときの自由エネルギー面（右）。

みると、この系に対してはほぼ同じ振る舞いをすることが分かった。これらの計算手法はともに分布の Fokker-Planck 的な振る舞いに基づいているので、似ていることは合理的だが、計算手法としては全く異なるので、面白い結果である。

(3) 以上の結果を踏まえて、シニョリンに対する重み付きアンサンブル法の適用を行った。その際にはまずオーダーパラメータとして、先行研究で用いられた 2 つの水素結合の距離（これを定めるには化学的な直観を要する）と、上で計算した拡散マップの 2 次元座標を用いた。また、状態空間を 8 つのセルに分割し、それぞれのセルに 4 つの粒子を用意した。10 ピコ秒おきに粒子を観測し、それを 100 回繰り返した。すなわちそれぞれの粒子は 1 ナノ秒だけ時間発展した。初期状態としては安定な状態（天然状態）F と準安定な状態（ミスフォールドした状態）M をとり、これらの状態間の遷移を計算した。そして、F 状態、M 状態に対応するセルの確率を計算し、それらの時間発展の振る舞いから、遷移時間を概算した。その結果、両方の遷移に対して、10 ナノ秒ほどの遷移時間が得られ、上のマイルストーン法などによる計算とコンシステントであることが分かった（図 4）。また、どちらのオーダーパラメータを用いても同じ値が得られたので、これは化学的な直観を要しない、多様体学習の座標でも計算が可能であることを意味する。また、拡散マップで得られた、もっとも有効的な 1 次元の座標は、グリシンの 2 面角と同様の振る舞いをすることも分かったので、これから構造変化の化学的な理解に関しても、拡散マップ（多様体学習）が有効であることが結論づけられた。

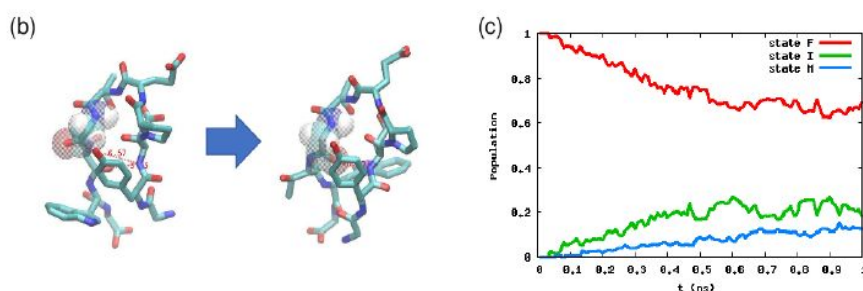


図 4 シニョリンの F 状態から M 状態への構造変化（左）。重み付きアンサンブル法で得られたポピュレーションのダイナミクス（右）。

(4) 上で述べた結果は化学物理

の雑誌 Journal of Chemical Physics の招待論文として掲載済みである。当初の計画では重み付きアンサンブル法と軌道のリアプノフ不安定性を用いたサンプリングを組み合わせる予定であったが、モデル系で試してもうまくいかなかったため、ここでは重み付きアンサンブル法と拡散マップ法を組み合わせる手法に変更した。リアプノフ不安定性を用いたサンプリングに関しては、最近、東大の明石らによる試み(R. Akaishi, Y.S. Nagornov, J. Phys. Soc. Jpn. **87** (2018) 063801)があるので、それを今後は使うことを考えている。

現在は重み付きアンサンブル法を PIN1 酵素やアデニル酸キナーゼといった、より大きな系に適用しており、ある程度成功しているが、単純に重み付きアンサンブル法を適用しても自由エネルギー障壁が高すぎる場合はうまくいかないことが分かったので、その点をどのように克服するかに取り組んでいる。

重み付きアンサンブル法は、統計学の粒子フィルター(particle filter)法と密接に関連のある手法である。どちらも並列でシミュレーションを行い、粒子（軌道）に重みをつけて、それらの取捨選択をしていく。ただし、粒子フィルター法ではポピュレーションの多いところで粒子を増やすのに対し、重み付きアンサンブル法ではレアイベントをサンプルするために、ポピュレーションの少ないところでむしろ粒子を増やす（その分、重みは小さくなる）。これらはリサンプリング(resampling)の手法の違いとみなせるが、そのような手法の variation についても今後考えたい。また、粒子フィルター法は現在、細胞ダイナミクスのモデルにおけるパラメータ推定に使うことも考えている。

また当初の計画では重み付きアンサンブル法を神経回路網ダイナミクスに用いる予定であったが、時間の関係上遂行できなかった。神経回路網ダイナミクスも、多数の準安定状態を持ち、その間を遷移するレアイベントの問題とみなすことができるので、本研究の対象としては非常に適している。これに関しては今後の課題としたい。

5 . 主な発表論文等

〔雑誌論文〕(計 9 件)

(1) [Hiroshi Fujisaki](#), Kei Moritsugu, Ayori Mitsutake, [Hiromichi Suetani](#), Conformational change of a biomolecule studied by the weighted ensemble method: Use of the diffusion map method to extract reaction coordinates, J. Chem. Phys. **149** (2018) 134112. <https://doi.org/10.1063/1.5049420> 査読あり

(2) [Hiroshi Fujisaki](#), Kei Moritsugu, Y. Matsunaga, Exploring Configuration Space and Path Space of Biomolecules Using Enhanced Sampling Techniques --- Searching for

Mechanisms and Kinetics of Biomolecular Functions, *Int. J. Mol. Sci.* **19** (2018) 3177.

DOI: [10.3390/ijms19103177](https://doi.org/10.3390/ijms19103177) 査読あり

(3) Y. Matsunaga, T. Yamane, T. Terada, K. Moritsugu, H. Fujisaki, S. Murakami, M. Ikeguchi, A. Kidera, *Energetics and Conformational Pathways of Functional Rotation in the Multidrug Transporter AcrB*, *eLife* 10.7554/eLife.31715 (2018).

DOI: 10.7554/eLife.31715 査読あり

(4) Hirofumi Suetani, Yoko Mizuno, Keiichi Kitajo, A manifold learning approach to chart human brain dynamics using resting EEG signals. *Unifying Themes in Complex Systems IX*. 359-367, 2018.

https://doi.org/10.1007/978-3-319-96661-8_37 査読あり

(5) 藤崎弘士, モンテカルロ法を使った席順決めと経路の再重み付けについて, *日本医科大学基礎科学紀要* **47** (2018) 1-13

https://www.nms.ac.jp/var/rev0/0019/3470/47thebulletin_1.pdf 査読なし

(6) 藤崎弘士, 生体分子におけるレアイベントサンプリング: Onsager-Machlup 作用を用いた計算手法, *京都大学数理解析研究所講義録* **2028** (2017) 38-48

www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/pdf/2028-06.pdf 査読無し

(7) H. Fujisaki, K. Yagi, H. Kikuchi, T. Takami, and G. Stock, Vibrational energy transport in acetylbenzotrile described by an ab initio-based quantum tier model, *Chem. Phys.* **482** (2017) 86-92

<https://doi.org/10.1016/j.chemphys.2016.09.010> 査読あり

(8) 藤崎弘士, 生体分子におけるレアイベントの探求, *生物物理* **57** (2017) 40-41.

<https://doi.org/10.2142/biophys.57.040> 査読あり (依頼原稿)

(9) 藤崎弘士, 統計力学に基づく細胞ダイナミクスの計算手法について, *日本医科大学基礎科学紀要* **45** (2016) 29-50.

https://www.nms.ac.jp/var/rev0/0005/3210/45thebulletin_hiroshi_fujisaki_3.pdf 査読無し

[学会発表](計 19 件)

(1) Hiroshi Fujisaki, Weighted ensemble simulations of biomolecules: Applications to peptides and proteins, APS March Meeting 2019, Boston Convention Center, Mar. 4-8, 2019.

(2) H. Kikuchi, H. Fujisaki, Applications of manifold learning techniques to protein dynamics, 56th Annual Meeting of the Biophysical Society of Japan, Okayama Univ. (Okayama), 9/15-17, 2018.

(3) 藤崎弘士, 森次圭, 米澤康滋, 楯真一, プロリン異性化酵素の重み付きアンサンブル法による計算とその加速, *日本物理学会 2018 年秋季大会* 2018. 9. 9-12.

(4) 小田切健太, 藤崎弘士, 機械的刺激の効果を考慮した創傷治癒の数理モデル II, *日本物理学会 2018 年秋季大会* 2018. 9. 9-12.

(5) 菊地浩人, 藤崎弘士, 溶液中のアセチルベンゾニトリルにおけるエネルギー移動: 非平衡 MD による計算, *日本物理学会 2018 年秋季大会* 2018. 9. 9-12.

(6) H. Fujisaki, K. Moritsugu, A. Mitsutake, Y. Matsunaga, H. Suetani, Weighted ensemble simulations of biomolecules, 11th European Conference on Mathematical and Theoretical Biology, Lisbon, July 23 - 27, 2018.

(7) K. Odagiri, H. Fujisaki, Mathematical model for wound healing caused by exogeneous mechanical forces, 11th European Conference on Mathematical and Theoretical Biology, Lisbon, July 23 - 27, 2018..

(8) H. Fujisaki, K. Moritsugu, Y. Matsunaga, H. Suetani, Combining weighted ensemble method with Lyapunov weighted dynamics: Application to proteins, *Biophysical Society Meeting 62nd Annual Meeting (San Francisco)*, 2018. 2. 17-21.

(9) 藤崎弘士, 森次圭, 米澤康滋, 楯真一, PIN1 酵素における異性化反応への重み付きアンサンブル法の適用, *分子シミュレーション討論会 (金沢商工会議所)* 2017. 11.29-12.

(10) 藤崎弘士, 重み付きアンサンブル法によるタンパク質ダイナミクスの計算, 第 4 回非線形現象の捉え方 (福岡工業大学セミナーハウス, 大分県) 2017. 11. 3-5.

(11) 藤崎弘士, 森次圭, 末谷大道, 重み付きアンサンブル法を用いたタンパク質のパスサンプリング, *日本物理学会秋季大会 (岩手大学)* 2017. 9. 21-24.

(12) 小田切健太, 藤崎弘士, 機械的刺激の効果を考慮した創傷治癒の数理モデル, *日本物理学会秋季大会 (岩手大学)* 2017. 9. 21-24.

(13) 菊地浩人, 藤崎弘士, 古田忠臣, 岡本研, 西野武士, XOR の結合ポケットの中における基質の運動について --- 分子動力学による研究 --- 第 55 回日本生物物理学会大会 (熊本大学, 黒髪北地区) 2017. 9. 19-21.

(14) 藤崎弘士, 末谷大道, 光武亜代理, 生体分子の分子シミュレーションから反応座標をどう抜き出すか? *日本物理学会第 72 回年次大会* 2017. 3. 17-20.

(15) Hiroshi Fujisaki, Molecular dynamics simulations and dimensional reduction techniques, *Simulations Encounter with Data Science: Data Assimilation, Emulators, Rare*

Events and Design, (Tachikawa, Tokyo) 2017. 3. 9-11 .

(16) H. Fujisaki, Y. Yonezawa, M. Shiga, L. Maragliano, S. Tate, Exploring reaction pathways for peptidylprolyl-isomerase, Biophysical Society Meeting 61st Annual Meeting (New Orleans), 2017. 2. 11-15.

(17) 藤崎弘士, プロリン異性化酵素の反応経路シミュレーション, 2016. 12.17, 第3回日本医科大学・東京理科大学合同シンポジウム(日本医科大学 橘桜会館)

(18) 藤崎弘士, 米澤康滋, 志賀基之, 楯真一, プロリン異性化酵素の反応経路シミュレーション, 2016. 9. 13-16, 日本物理学会秋季大会(金沢大学角間キャンパス)

(19) 菊地浩人, 藤崎弘士, 古田忠臣, 岡本研, 西野武士, ヒトには効くがバクテリアには効かない痛風薬: 酵素と阻害剤 BOF との相互作用ダイナミクス, 2016. 9. 13-16, 日本物理学会秋季大会(金沢大学角間キャンパス)

〔図書〕(計0件)

〔産業財産権〕

○出願状況(計0件)

名称:

発明者:

権利者:

種類:

番号:

出願年:

国内外の別:

○取得状況(計0件)

名称:

発明者:

権利者:

種類:

番号:

取得年:

国内外の別:

〔その他〕

ホームページ等

<https://researchmap.jp/read0211704/>

6. 研究組織

(1)研究分担者

研究分担者氏名: 末谷大道

ローマ字氏名: Hiromichi Suetani

所属研究機関名: 大分大学

部局名: 理工学部

職名: 教授

研究者番号(8桁): 40507167

(2)研究協力者

研究協力者氏名: 光武亜代理

ローマ字氏名: Ayori Mitsutake

研究協力者氏名: 森次圭

ローマ字氏名: Kei Moritsugu

研究協力者氏名: 松永康佑

ローマ字氏名: Yasuhiro Matsunaga

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。