

Title	水和水が媒介するタンパク質リガンド相互作用をデータマイニングで解明する
Author(s)	水上, 卓
Citation	科学研究費助成事業研究成果報告書: 1-5
Issue Date	2017-06-15
Type	Research Paper
Text version	publisher
URL	<a href="http://hdl.handle.net/10119/14318">http://hdl.handle.net/10119/14318</a>
Rights	
Description	挑戦的萌芽研究, 研究期間: 2013 ~ 2016, 課題番号: 25650050, 研究者番号: 50270955, 研究分野: 生物物理学

## 科学研究費助成事業 研究成果報告書

平成 29 年 6 月 15 日現在

機関番号：13302

研究種目：挑戦的萌芽研究

研究期間：2013～2016

課題番号：25650050

研究課題名(和文) 水和水が媒介するタンパク質リガンド相互作用をデータマイニングで解明する

研究課題名(英文) Datamining approach for protein-ligand interaction mediated by hydration water

研究代表者

水上 卓 (Mizukami, Taku)

北陸先端科学技術大学院大学・先端科学技術研究科・助教

研究者番号：50270955

交付決定額(研究期間全体)：(直接経費) 1,600,000円

研究成果の概要(和文)：本研究の目標は、タンパク質・リガンド結合系を対象に、データマイニング法により定義された水和水の可視化(現象論)と、溶媒和自由エネルギー計算による水和水効果の見積り(定量化)である。

MDシミュレーションの出力データから混合分布モデル(GMM)によって水の振る舞いをデータマイニングする系を確立した。これによりリガンド結合タンパク質水和水自由エネルギーと水の振る舞いの強い相関を検知し、ある程度の予測を可能とした。次に原子の多体相互作用を考慮した記述子を設計し、一般線型モデルとクラスタリングにより機械学習・予測を行った。小さなサイズの分子系に適用したところ、より詳細なエネルギーの予測が可能となった。

研究成果の概要(英文)：The motivation of the research is 1) visualization of hydration water defined by data-mining method, and 2) estimation of hydration effect by means of solvation free energy.

We founded a system for mining water behavior by means of the mixture model from MD simulation data. It detected a strong correlation between solvation free energy and hydration water behavior that enables a prediction of free energy.

The machine learning and prediction system was build with the combination of three methods, linear model, clustering and a new descriptor that was designed by accounting the multi-body interaction between atoms. Application of the machine for the small-scale solution / molecular systems resulted in a more detailed prediction for the energies.

研究分野：生物物理学

キーワード：タンパク質 水和水 分子動力学計算 データマイニング 特徴空間 機械学習

### 1. 研究開始当初の背景

タンパク質表面の水和水が、ダイナミクスや機能発現におおきな影響を与えることは、古くから指摘されてきた。しかしながら水分子の非対称な構造、比較的大きなエネルギーをもつ水素結合、長距離におよぶ静電相互作用、などの理由により複雑な振る舞いや反応を示し、その直接的な解明に困難があった。またタンパク質のリガンド結合において、分子間に存在する水和分子の影響が指摘されてきたが、同様な理由によってその解明に困難があった。

一方で近年、機械学習・データマイニングの技術が発展し、複雑な問題の取り扱いが確立されてきた。このアプローチをこれら溶液系や生体高分子の複雑な問題に適用することが望まれていた。

### 2. 研究の目的

タンパク質・水の分子動力学計算から生じた大量データから、データマイニングによって水分子の振る舞いを特徴化し、特徴空間上でのクラスタリング等の解析による現象のクラス分けを行う。また逆に、それら特徴空間上のクラスを実空間に投影して物理化学的データに換算し、各クラス分けされた水分子の振る舞いの特徴と、物理化学的な意味との関係をあきらかにする。

さらに機械学習の技術を用い、得られた特徴空間を利用して、物理化学量の予測を可能とするシステムを構築する。

目標とする対象は、水分子による溶媒和現象であり、タンパク質のリガンド結合過程における水和水の役割である。

### 3. 研究の方法

(1)水分子の振る舞いのデータマイニング、Mixture Model によるクラスタリング、およびそれらのディスクリプション(記述);

水分子の振る舞いを特徴化し分類するために、固体やタンパク質表面の存在やそこか

らの距離を前提としない、溶質-溶媒系の空間上のあらゆる位置に存在し得る仮想的な“水和サイト”を想定する。分子動力学(Molecular Dynamics; MD)計算によるトラジェクトリデータから、水分子の酸素原子あるいは水素原子の、MD空間上の位置を経時的に記録し、Mixture Model をもちいて解析した。確率密度関数として Gaussian を採用しその線形結合として記述する。得られた個々の3次元ガウス分布を仮想的な水和サイトと想定し、そのサイズ、エリプソイドの方向ベクトル、含まれる水分子のMD経時点数などの特徴量とその線形結合からなるベクトルデータを生成し、特徴空間を構築した。特徴空間上で、変数選択や変数の統合、PCA(Principal Component Analysis)などを行い、溶液中においてタンパク質の有無によるデータ分布の差が検知されるように部分特徴空間を探索した。最後に特徴空間上でデータのクラスタリングを行い、水分子の“振る舞い”を複数の種類に分解した。

(2)マイニングによって得られた水分子振る舞いクラスの評価;

特徴空間上でクラスタリングの結果から分類された水分子の振る舞いを元にして、物理化学的な量を算出し、それを指標にして、クラスの評価を行った。これは特徴空間上のクラスを分子動力学計算で表現される実空間へと投影する操作に対応する。

見積もる物理化学量は、温度、溶質表面からの動経分布関数、中間散乱関数、時空相関関数、動的構造因子、分子系全体のポテンシャルエネルギー、およびエネルギー表示法(松林ら)による溶媒和自由エネルギーをもちいた。

(3)原子間の多体相互作用を配慮した記述子の構築と、それを用いた機械学習と予測;

水分子の動的な「振る舞い」と共に、分子間・原子間の相互作用の情報を内包する物理量を学習・予測するために、あらたな記述子

を設計・採用した。原子の2体相互作用を想定した原子間の距離を変数とする分布、および3体相互作用を想定した角度分布を基礎において、MDの1スナップショット構造あたり約500次元の特徴空間を構築した。

この特徴空間上で一般化線形モデルを用い、変数選択やPCAなどを援用して、ターゲットとする物理量の学習と予測を行った。(4)対象とする分子系；

大きさの異なる複数の可溶性タンパク質、リガンド結合性のタンパク質、およびNa<sup>+</sup>Cl<sup>-</sup>イオンやジアラニンなどの小分子を溶質とし水分子を溶媒とする分子系を構築し、AmberおよびGROMACSを用いて分子動力学計算を行い、トラジェクトリデータの取得を行った。

#### 4. 研究成果

(1)Gaussian Mixture Modelによる水分子の振る舞いの特徴空間の構築と、特徴空間上で分離された水分子の振る舞いの物理化学的な解釈；

水分子の振る舞いを分類するために、Gaussian Mixture Modelを採用して特徴空間を構築した。それを用いて、大きさの異なる3種類のタンパク質(PDBID=1psv,4pti,1hel)の、水溶液中の溶媒分子の振る舞いを調査した。特徴空間上で、タンパク質水溶液と純水(バルク水)の2つのケースでデータ点の分布を比較し、バルク水にタンパク質を挿入したときに特徴空間に現れる新しい分布を単離した。これにより300Kにおいて、タンパク質の水和に関するクラスを2つ同定した。

各クラスに属する水分子の振る舞いを分子動力学計算のデータ空間に投射し、各物理化学量を算出した。その結果、平均の拡散係数がバルク水と同等、バルク水より遅い、

バルク水より速い、3つのクラスが存在することがわかった。それぞれのクラスでタンパク質表面からの動経分布関数を見積もる

と、バルク水、第1水和水、第1水和水+第2水和水+バルク水、の領域と一致した。以上の結果からタンパク質表面と水分子のあいだの距離情報を用いることなく、水分子の振る舞いのみをマイニングすることで、第1水和水と一致するクラスを同定することが出来、本研究のアプローチの有効性を示した。

タンパク質周辺の第1・第2水和層内において、拡散の速いクラスの水分子は、タンパク質の疎水・親水面でそれぞれ異なった振る舞いをしめした。疎水面では第1水和・第2水和に特徴的な2つの鋭いピークによる動経分布を示し、親水面ではほぼなだらかな分布を示す。これは、親水面と疎水面での水分子の振る舞いに根本的な違いがあることを示唆している。また、これら2つの速い拡散

・遅い拡散を示すクラスは、時空相関関数に構造をもち、そこから水分子における手続きの・文脈的な相互作用が存在することが示唆された。今までの調査の範囲において、これはバルク水の領域ではみられず、タンパク質の水和領域にのみ見られる現象である。(2)タンパク質のドッキングにおける水の振る舞い；

サブユニット間でドッキングをおこなうタンパク質(PDBID=1brs)を用い、ドッキングの際の溶媒の振る舞いを調べた。1brsの2つのユニット間の距離 $r$ を変化させ、Mixture Modelによる水分子の振る舞いの特徴空間上の変化を調べた。それによると、距離 $r$ が10Å以下で遅い拡散を示す水和水のサイトが増加していくことがわかった。これは10Å程度の比較的長距離においても水分子の振る舞いに相関があることを示している。また、距離 $r$ が小さくなるにつれて第2水和層から第1水和層へ水分子の密度分布の移動を観測した。同じ条件での溶媒和自由エネルギーを見積もると、ユニット間距離 $r$ が小さくなるにつれ、水和自由エネルギーは安

定化し強い相関を示した。

(3)原子間の多体相互作用に着目した記述子を用いた、溶液系における物理化学量の機械学習と予測；

バルク水および小分子による水溶液を用い、溶液の各スナップショット構造における全ポテンシャルエネルギーをターゲットとして原子間の局所的な配置構造と多体相互作用に着目した記述子をもちいた一般化線形モデルを構築した。特徴空間上で変数選択と分子の局所的な配置構造によるクラスタリングを行い、各構造クラスからなる混合分布モデルを構築した。次に各クラスで独立な一般化線形モデルを導入して機械学習を行った。形成した学習器を用いて予測を行うと、溶液系の局所構造の各クラスにおいて、ポテンシャルエネルギーの寄与を見積もることに成功した。

同様のプロセスを溶媒和自由エネルギーに対して進め、溶液構造のクラス分けとそれによる寄与の算出が限定した条件において可能となった。

#### 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 2 件)

(1)Hiroaki Saito, Masashi Iwata, Kazutomo Kawaguchi, Taku Mizukami, Takeshi Miyakawa, Masako Takazu, and Hidemi Nagao, “Molecular Dynamics Study of Gramicidin A in Lipid Bilayer: Electrostatic Map and Ion Conduction”, JPS Conf. Proc., 1, (2014), 012053-1,4

DOI: <http://dx.doi.org/10.7566/JPSCP.1.012053> (査読有り)

(2)Hiroaki Saito, Masashi Iwayama, Taku Mizukami, Jiyoung Kangb, Masaru Tatenob, Hidemi Nagao, “Molecular dynamics study on binding free energy of

Azurin-Cytochrome c551 complex”, Chemical Physics Letters, 556, (2013), 297-302

DOI: <https://doi.org/10.1016/j.cplett.2012.12.016> (査読有り)

〔学会発表〕(計 10 件)

(1) Machine-learning approach for behavior of hydration water; A pathway to discovery and prediction for biophysical and physiological properties of protein, Taku Mizukami, Nguyen Viet Cuong, Pham Tien Lam, Dam Heiu Chi, International Symposium on Biophysics of Rhodopsins, 国際学会, 2017年5月11日～2017年5月12日, 京都大学 (京都府京都市)

(2)機械学習によるタンパク質親水/疎水表面における水分子の動的振る舞いの解析  
水上 卓, Viet Cuong Nguyen, Ho Tu Bao, and Dam Hieu Chi, 日本生物物理学会 第54回年会 2016年11月25日～2016年11月27日, つくば国際会議場 (茨城県つくば市)

(3) Simulation-based data-mining approach for the protein hydration water behavior, Taku Mizukami, Nguyen Viet Cuong, Ho Tu Bao, and Dam Hieu Chi, Pacificchem2015 国際学会, 2015年12月14日～2015年12月22日, (Honolulu, USA)

(4) シミュレーション・データマイニングによるリガンド結合系における水分子の握る舞い, 水上 卓, Nguyen Viet Cuong, Ho Tu Bao, and Dam Hieu Chi, 日本生物物理学会 第53回年会, 2015年11月25日～2015年11月27日, 金沢大学 (石川県金沢市)

(5)Hydration Water Behavior Characterization by Simulation-Based Data-Mining Approach, Taku Mizukami, Nguyen Viet Cuong, Ho Tu Bao, and Dam Hieu Chi, ACCMS-8 (Conference of Asian Consortium on Computational Materials

Science) 国際学会, 2015年06月16日~2015年06月18日, NTUST (Taipei, Taiwan)

(6) 混合分布モデルにより分離されたタンパク質水和水の振る舞い: シミュレーション・データマイニングによるアプローチ, 水上卓, Viet Cuong Nguen, Ho Tu Bao, and Dam Hieu Chi, 日本生物物理学会 第52回年会, 2014年09月13日~2014年09月15日, 札幌コンベンションセンター (北海道札幌市)

(7) シミュレーション・データマイニングアプローチによる蛋白質ドッキング過程における水和水ダイナミクス, 水上卓, 杉山 歩, Ho Tu Bao, and Dam Hieu Chi 生物物理学会 第51回年会 2013年10月28日~2013年10月30日京都国際会議場 (京都府京都市)

(8) L1型正則化線形回帰モデルを用いたシミュレーションデータからの二元合金の物性予測, 鈴木 大輔, 川崎 隆史, 杉山 歩, 水上卓, Dam Hieu Chi, Ho Tu Bao 第7回分子科学討論会 2013年9月24日~2013年9月27日 京都テルサ (京都府京都市)

(9) スパースモデルによる水溶液構造の解析 水上卓, 杉山 歩, 川崎 隆史, 高木 啓行, Ho Tu Bao, Dam Hieu Chi, 第7回分子科学討論会 2013年9月24日~2013年9月27日 京都テルサ (京都府京都市)

(10) A novel category of protein hydration water classified by simulated data mining approach, Taku Mizukami, Ayumu Sugiyama, Dam Hieu Chi, Ho Tu Bao, ACCMS-7(Conference of Asian Consortium on Computational Materials Science) 国際学会, 2013年7月25日, (Korat, Thailand)

〔図書〕(計 1件)  
水上卓, 朝倉書店, 光と生命の事典, 2016, 第2章 光のエネルギー利用, 25 生物による光エネルギーの利用, p36-37

〔産業財産権〕

出願状況(計 0件)

取得状況(計 0件)

〔その他〕  
無し

6. 研究組織  
(1) 研究代表者  
水上卓 (TAKU MIZUKAMI)  
北陸先端科学技術大学院大学・先端科学技術研究科・助教

研究者番号: 50270955