10月30日(金)

CBI 学会 2020 年大会

Chem-Bio Informatics Society (CBI) Annual Meeting 2020

10月30日（金）

CH1　チャンネル１

＜プレナリー講演＞『我が国における新型コロナウイルス感染症 COVID-19に対する計算科学の取り組み』

座長：上村 みどり（帝人ファーマ㈱）

日時：2020 年 10 月 30 日 10:00-10:40
チャンネル：1

COVID-19 の数理科学 -標準治療の早期確立を目指して

Mathematical Sciences in COVID-19

岩見 真吾
Shingo Iwami

九州大学大学院理学研究院生物科学部門
Kyushu University.

私達のグループは 100%ドライラボである利点を生かして様々な研究テーマに取り組み、生命医科学の実験・臨床研究者との共同研究を積極的に進めている。数理モデルを用いた実験データあるいは臨床データの定量的解析をはじめ、遺伝子配列データの生物情報学的解析やシミュレーションによる網羅的理論予測、データ解析のための理論自体の構築など、私が独自に築き上げた国内外の研究ネットワークを駆使して、多面的に研究を展開している。数理科学の汎用性を最大限に利用することで、私達の研究では一貫して(疾患を含む)様々な生命現象のエンジンになっている『増殖・分化・感染・変異・進化・適応する要素』が組み合わさって創発するシステムの定量的分析(生命動態定量)を行ってきた。そして、生命現象に共通して内在する問題を解決するために、ユニークで汎用性の高いアプローチを開発し、個別の生命現象に対する理解を深めてきた。

本講演では、私達のグループが構築してきた国際ネットワーク（オックスフォード大学とインディアナ大学）に基づいた、新型コロナウイルス(SARS-CoV2)による感染症(COVID-19)をテーマにした異分野融合型（生物、数学、公衆衛生学）の最新の共同研究の成果について報告したい。まず、COVID-19 の臨床データを用いたモデル駆動型アプローチにより、生体内における SARS-CoV2 感染動態が、過去に流行した同様のコロナウイルスである中東呼吸器症候群コロナウイルス(MERS-CoV)および重症急性呼吸器症候群コロナウイルス(SARS-CoV)の感染動態とどのように異なるかを定量的に分析した結果を報告する。次に、私達が開発しているモデル駆動型とデータ駆動型の融合アプローチにより、COVID-19 患者ではウイルス排出パターンが短期(発症 7 日程度)、中期(発症 14 日程度)、長期(発症 28 日程度)の 3 グループに層別化できること、および、これらの患者異質性が COVID-19 患者の治療評価において大きな交絡因子になることを説明する。そして、現在世界中の研究チームが取り組んでいる、承認薬のドラッグリポジショニングによる COIV-19 への治療効果を正確に分析するためには、何が必要で何が欠けているかをシミュレーションを駆使した分析により明らかにした結果を報告する。

K-11

日時： 2020 年 10 月 30 日 10:40-11:20
チャンネル： 1

富岳：「アプリケーションファースト」のエクサスケールスパコンと、その COVID-19 に対する戦い

Fugaku: Applications-First "Exascale" Supercomputer and its Fight Against the COVID-19

松岡　聡
Satoshi Matsuoka

理化学研究所 計算科学研究センター(R-CCS) センター長
RIKEN
東京工業大学 特任教授
Tokyo Institute of Technology

富岳は理研と富士通の共同開発により 2020 年 4 月に試験的稼働が開始され、同 6 月にスパコンランキングで Top500 を含む四冠を史上初めて達成した。だが、富岳はランキングを目標に設計された事は一度もなく、約 10 年にも及ぶ日本の広範な HPC コミュニティが参加した検討開始から設置完了に至るまで、常に「アプリケーション・ファースト」のコデザインが遂行され、その中には生命・医療・創薬などのアプリケーションが多く含まれていた。

それらのアプリを含み、富岳は Society5.0 における社会課題の幅広い新たな解決を目指しており、また、各種アプリにおける計算手法としても、従来の第一原理的なシミュレーションだけでなく、ビッグデータの分析や、AI による経験則的な未来予測、更にはそれらの統合的ソリューションなど含んでいる。富岳はこれらを駆動する全てのメトリックで世界一になっており、かつ、その性能を最大限ものづくりなどの環境で活用できるソフトウェアや運用体制の整備が行われている。その成果はいち早く新型コロナウイルス対策の複数のプロジェクトで発揮されており、今後様々な産業界の多様な利用に適用されていく事が期待されている。

特に新型コロナウィルスに対する対策プログラムは、富岳が正式に稼働する一年前に前倒しで急遽計画され、審査を経て現在 5 つの課題が走っており、随時新課題も募集中である。それらはウィルスの構造を第一原理計算で解明するミクロな生物学的シミュレーションから、人を一人一人 AI エージェントとみなし、その行動から感染伝搬をシミュレーションする社会レベルのマクロなものまでさまざまだが、いずれも非常に大規模な計算リソースを要求し、場合によっては富岳以外の日本のスパコンの計算資源を全て合算したレベルの要求があるものもある。しかしながら、その成果はめざましく、既存薬の網羅的な高精度探索による治療薬候補の早期発見や、飛沫感染に対する感染防止の政府のガイドライン作成などに科学的根拠を与えるなど、国民の安全安心に対し非常に多くの役割を果たしており、生物学的な危機における大規模スパコンの重要性を正に示しているといえる。

本講演では、富岳を紹介するとともに、コロナウィルス対策を中心とした生命科学の新たな発展への可能性も言及する。

日時： 2020 年 10 月 30 日 11:20-12:00
チャンネル： 1

医療ビッグデータの研究開発への利活用 ～次世代医療基盤法は何をもたらすか～

Utilization of Large-Scale Medical Information Data for R&D
-Next-Generation Medical Information Infrastructure Act-

小松 慶太
Keita Komatsu

内閣官房 健康・医療戦略室（併 内閣府、経済産業省）
Cabinet Office, Government of Japan

医療・バイオ分野の研究開発においてもビッグデータが活用されるようになって久しいが、これまで画像・検査値・診断名等の多種多様な情報を時系列で紐付けたデータベースを構築することは困難であり、利用者が限定されていることも多かった。そこで、2018 年、医療分野の研究開発に資するための匿名加工医療情報に関する法律（通称「次世代医療基盤法」）が施行され、時代の要請に応えたデータベースを整備するための環境が整えられた。

昨年から今年にかけて、データベース整備の主体となる法人が複数認定され、近く、利用が開始される見込みとなった。

当該データベースにより提供される情報がどのようなものになるのか、医薬品や新規治療法の開発に向けてどのように貢献できるかなど、今後の展望も含めて明らかにする。

10月30日（金）

CH1　チャンネル１

口頭発表７

＜口頭発表７＞『新型コロナ等ウィルス対策/計算化学（分子認識）』 6

座長：片倉 晋一（慶應義塾大学）、福澤 薫（星薬科大学）、
広川 貴次　（産業技術総合研究所）

01-08 Ryo Hatada (Department of Chemistry & Research Center for Smart Molecules, Faculty of Science, Rikkyo University)
"Interaction analyses between SARS-CoV-2 main protease and inhibitor N3 by using fragment molecular orbital method and molecular dynamics simulation"

01-09 Kazuki Akisawa (Department of Chemistry & Research Center for Smart Molecules, Faculty of Science,Rikkyo University)
"Interaction analyses on SARS-CoV-2 spike proteins by using fragment molecular orbital method"

01-10 Nazim Medzhidov (Elix, Inc.)
"Predicting inhibitors for SARS-CoV-2 RNA- dependent RNA polymerase using machine learning"

01-11 Mohini Yadav (Department of Engineering, Chiba Institute of Technology)
"Theoretical insights into the molecular mechanism of NA-I117V-Mediated Oseltamivir Resistance in H5N1 Avian Influenza Virus"

Interaction analyses between SARS-CoV-2 main protease and inhibitor N3 by using fragment molecular orbital method and molecular dynamics simulation

Ryo Hatada[1] **Koji Okuwaki**[1] **Kazuki Akisawa**[1] **Yuji Mochizuki**[1,2] **Yuma Handa**[3] **Kaori Fukuzawa**[2,3] **Yuto Komeiji**[4] **Yoshio Okiyama**[5] **Shigenori Tanaka**[6]

hatada@rikkyo.ac.jp 17cc044m@rikkyo.ac.jp okuwaki@rikkyo.ac.jp fullmoon@rikkyo.ac.jp d2002@hoshi.ac.jp k-fukuzawa@hoshi.ac.jp y-komeiji@aist.go.jp okiyama@nihs.go.jp tanaka2@kobe-u.ac.jp

1 Department of Chemistry & Research Center for Smart Molecules, Faculty of Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan
2 Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
3 School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa-Ku, Tokyo 142-8501, Japan
4 Health and Medical Research Institute, AIST, Tsukuba Central 6, Tsukuba, Ibaraki 305-8566, Japan
5 Division of Medicinal Safety Science, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa 201-9501, Japan
6 Graduate School of System Informatics, Department of Computational Science, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe 657-8501, Japan

Keywords: SARS-CoV2, Fragment molecular orbital (FMO) method, Molecular dynamics

Many research groups have engaged a variety of investigations against the worldwide issue of SARS-CoV-2. Here, we have performed a combination approach of fragment molecular orbital (FMO) calculation and molecular dynamics (MD) simulation to analyze the interactions between the SARS-CoV-2 Main protease (Mpro) and the N3 inhibitor in a statistical manner.

Prior to the MD-assisted investigation, a detailed analysis with inter-fragment interaction energy (IFIE) was performed for the crystal structure (PDB ID: 6LU7) [1] of complex between Mpro and N3 inhibitor, based on the FMO-MP2(PR)/6-31G* calculation by using ABINIT-MP [2,3]. Using AMBER [4], MD simulations for 100 ns were then performed from the 6LU7 initial structure, and a total of 100 structures were extracted from the trajectory. These sampled structures were subjected to a series of FMO calculations, and statistical evaluation of IFIEs was made.

Through these studies, several amino acid residues such as His163 and Glu166 was identified to be important in interacting with N3 with hydrogen bond. Interestingly, the structural fluctuation via MD sampling could alter the relative importance among the residues. We will report on the details of these results on the presentation day.

[Acknowledgment] The present work was supported by MEXT as a social and scientific priority issue #6 to be tackled by using post-K computer (FS2020), AMED-BINDS (JP20am0101113), and Rikkyo SFR. MD simulations were performed using TSUBAME 3.0 (BINDS quota) at the Tokyo Institute of Technology. For FMO calculations, computational resources of FX-100 at Nagoya University, RIKEN R-CCS Fugaku (priority trial use against coronavirus) and Oakforest-PACS (hp200146 quota) at JCAHPC were used.

[1] Jin, Z., *et al.*, Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature*, 582 (7811): 289-293, **2020**.
[2] Tanaka, S., *et al.*, Electron-Correlated Fragment-Molecular-Orbital Calculations for Biomolecular and Nano Systems. *Phys. Chem. Chem. Phys.*,16 (22): 10310–10344, 2014.
[3] Hatada, R., *et al.*, Fragment Molecular Orbital Based Interaction Analyses on COVID-19 Main Protease − Inhibitor N3 Complex (PDB ID: 6LU7). *J. Chem. Inf. Model.*, 60 (7): 3593–3602, 2020.
[4] <https://ambermd.org/>.

Interaction analyses on SARS-CoV-2 spike proteins by using fragment molecular orbital method

Kazuki Akisawa[1] Ryo Hatada[1] Koji Okuwaki[1] Yuji Mochizuki[1,2] Kaori Fukuzawa[2,3] Yuto Komeiji[4] Shigenori Tanaka[5]

17cc044m@rikkyo.ac.jp hatada@rikkyo.ac.jp okuwaki@rikkyo.ac.jp
fullmoon@rikkyo.ac.jp k-fukuzawa@hoshi.ac.jp
y-komeiji@aist.go.jp tanaka2@kobe-u.ac.jp

[1] Department of Chemistry & Research Center for Smart Molecules, Faculty of Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan
[2] Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
[3] School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa-Ku, Tokyo 142-8501, Japan
[4] Health and Medical Research Institute, AIST, Tsukuba Central 6, Tsukuba, Ibaraki 305-8566, Japan
[5] Graduate School of System Informatics, Department of Computational Science, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe 657-8501, Japan

Keywords: SARS-CoV2, Spike protein, Fragment molecular orbital (FMO) method, MP4(SDQ)

It is well known that the spike protein of SARS-CoV-2 commits the infection to human cells. Thus, considerable efforts in worldwide have been conducted to understand the natures of this protein build with 3 chains (a total of 3.3 thousand amino acid residues). Here, we have performed highly correlated FMO (fragment molecular orbital) calculations with ABINIT-MP [1]. The calculated systems were the spike protein of both closed (PDB ID: 6VXX) and open (PDB ID: 6VYB) structures and also a couple of RBD (receptor binding domain) complexes with human ACE2 (angiotensin-converting enzyme 2) (PDB ID: 6M0J) and B38 antibody (PDB ID: 7BZ5).

The structures of 6VXX and 6VYB were prepared with a homology modeling and relaxed by an MD (molecular dynamics) annealing using AMBER [2]. The structure preparations for 6M0J and 7BZ5 were rather simply done with MOE [5]. The level of FMO calculations were MP2, MP3 [2], and MP4(SDQ) [3], and a scaling for correlation energy such as MP2.5 was adopted to improve accuracy. The basis sets were 6-31G* and cc-pVDZ. A series of IFIE (inter-fragment interaction energy) analyses were then made for these systems.

It was found that the inter-chain stabilization energies (A-B, B-C and C-A) of the open structure (6VYB) were sizably different from those of the closed structure (6VXX). In particular, the stabilization energy of RBD in chain B of 6VYB is remarkably decreased relative to that of 6VXX. The importance of charged residues in chain interactions was illuminated as well.

The calculated IFIEs of 6M0J and 7BZ5 suggested that the stabilization energy loss of RBD in chain B is partly compensated by the binding with ACE2 and B38 antibody. These results could be consistent with the fact that the open structure is responsible for the infection to human cell and the target of antibody. The details of these results will be reported on the day of the presentation.

[Acknowledgment] The present work was supported by Rikkyo SFR and AMED-BINDS (JP20am0101113). The world's largest scale higher-order correlated FMO calculations were carried out using the RIKEN R-CCS "Fugaku" (priority trial use against coronavirus) and the ITO System A (hp200147 quota) at the Kyushu University.

[1] Tanaka, S., et al., Electron-Correlated Fragment-Molecular-Orbital Calculations for Biomolecular and Nano Systems. Phys. Chem. Chem. Phys.,16 (22): 10310−10344, 2014.
[2] <https://ambermd.org>.
[3] Mochizuki, Y., et al., Large-Scale FMO-MP3 Calculations on the Surface Proteins of Influenza Virus, Hemagglutinin (HA) and Neuraminidase (NA). Chem. Phys. Lett. 2010, 493 (4), 346–352.
[4] Mochizuki, Y., et al., Higher-Order Correlated Calculations Based on Fragment Molecular Orbital Scheme. Theor. Chem. Acc. 2011, 130 (2), 515–530.
[5] < https://www.chemcomp.com/>.

O1-10

Predicting inhibitors for SARS-CoV-2 RNA-dependent RNA polymerase using machine learning

Nazim Medzhidov[1]
nazim.medzhidov@elix-inc.com

Romeo Cozac[1]
romeo.cozac@elix-inc.com

Shinya Yuki[1]
shinya.yuki@elix-inc.com

[1]Elix, Inc., Daini Togo Park Building 3F, 8-34 Yonbancho, Chiyoda-ku, Tokyo 102-0081, Japan

Keywords: SARS-CoV-2, RdRp, Machine Learning

Global coronavirus disease pandemic (COVID-19) caused by newly identified SARS-CoV-2 coronavirus continues to claim the lives of thousands of people worldwide. The unavailability of specific medications to treat COVID-19 has led to drug repositioning efforts using various approaches, including computational analyses. Such analyses mostly rely on molecular docking and require the 3D structure of the target protein to be available. In this study, we utilized a set of machine learning algorithms and trained them on a dataset of RNA-dependent RNA polymerase (RdRp) inhibitors to run inference analyses on antiviral and anti-inflammatory drugs solely based on the ligand information. We also performed virtual screening analysis of the drug candidates predicted by machine learning models and docked them against the active site of SARS- CoV-2 RdRp, a key component of the virus replication machinery. Based on the ligand information of RdRp inhibitors, the machine learning models were able to identify candidates such as remdesivir and baloxavir marboxil, molecules with documented activity against RdRp of the novel coronavirus. Among the other identified drug candidates were beclabuvir, a non-nucleoside inhibitor of the hepatitis C virus (HCV) RdRp enzyme, and HCV protease inhibitors paritaprevir and faldaprevir. Further analysis of these candidates using molecular docking against the SARS-CoV-2 RdRp revealed low binding energies against the enzyme active site. Our approach also identified anti-inflammatory drugs lupeol, lifitegrast, antrafenine, betulinic acid, and ursolic acid to have potential activity against SARS-CoV-2 RdRp. We propose that the results of this study are considered for further validation as potential therapeutic options against COVID-19.

O1-11

Theoretical insights into the molecular mechanism of NA-I117V-Mediated Oseltamivir Resistance in H5N1 Avian Influenza Virus

Mohini Yadav[1] s1879602EC@s.chibakoudai.jp
Manabu Igarashi[2] igarashi@czc.hokudai.ac.jp
Norifumi Yamamoto[1] norifumi.yamamoto@p.chibakoudai.jp

[1] Department of Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan
[2] Research Center for Zoonosis Control, Hokkaido University, 5 Chome Kita 8 Jonishi, Kita Ward, Sapporo, Hokkaido 060-0808, Japan

Keywords: Avian Influenza, Neuraminidase, H5N1, NA-I117V, Drug Resistance, Oseltamivir Resistance

Influenza is a contagious respiratory illness that infects the nose, throat, and lungs, causing mild to severe illness. Glycoproteins, hemagglutinin (HA) and neuraminidase (NA), of the influenza virus play essential roles in its replication process. HA binds to sialic acid (SA) at the end of a sugar chain present on the surface of a host cell, and then the virus is taken into the cell. NA on the virus surface plays a role of scissors to cleave the bond between SA and the sugar chain, which allows the daughter virus to liberate to begin infecting the surrounding cells.

Anti-influenza drugs that inhibit the interaction between NA and SA have been developed. Currently, four types of NA inhibitors (oseltamivir, zanamivir, peramivir, and laninamivir) have been used for the treatment of influenza in Japan. But resistance against available anti-influenza drugs is emerging rapidly. Hence, it's crucial to study the detailed mechanism of drug resistance to develop anti-influenza drugs which are less prone to resistance.

Most of the receptor mutations that lead to drug resistance occurs at its active site residues, since they are the key residues interacting with the drug. But residue 117 is not a part of active site of neuraminidase and still an Ile-to-Val substitution at position 117 (NA-I117V) causes a slight reduction in susceptibility to oseltamivir (OT) *in vitro* and dramatically *in vivo*.[1] However, the molecular mechanism of OT resistance caused by NA-I117V mutation is still not clear.

In this study, to clarify the binding affinities of SA and OT to the NA-I117V, we computed the corresponding binding free-energies using the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method. We also computed the pairwise per-residue free energy decomposition calculation to clarify the molecular mechanism of OT resistance in NA-I117V mutant.

[1] Takano, R., Kiso, M., Igarashi, M., Le Q.M., Sekijima M., Ito, K., Takada, A., Kawaoka, Y., Molecular Mechanisms Underlying Oseltamivir Resistance Mediated by an I117V Substitution in the Neuraminidase of Subtype H5N1 Avian Influenza A viruses, *The Journal of Infectious Diseases,* 207:89-97, 2013.

10月30日（金）

CH1　チャンネル１

口頭発表10

座長：石川 岳志　（鹿児島大学）

O1-12 Kentaro Takai (Fujitsu Ltd.)
"Advanced methods to predict properties of cyclic peptides: conformation analysis of peptides cyclized by non-amide bonds"

O1-13 Kei Moritsugu (Yokohama City University)
"Enhanced conformational sampling of cyclic peptide Cyclorasin by coupled Nosé-Hoover equation"

O1-14 Yuji Mochizuki (Department of Chemistry & Research Center for Smart Molecules, Faculty of Science, Rikkyo University & Institute of Industrial Science, University of Tokyo)
"Development status of ABINIT-MP program in 2020"

O1-15 Koji Okuwaki (Faculty of Science, Rikkyo University)
"Development of multiscale FMO-DPD simulations for molecular design"

Advanced methods to predict properties of cyclic peptides: conformation analysis of peptides cyclized by non-amide bonds

Kentaro Takai[1] takai.kentaro@fujitsu.com **Sakiko Mori**[1] mori.sakiko@fujitsu.com **Atsushi Tomonaga**

Azuma Matsuura **Mutsuyo Wada**[1] **Akihiko Ueda**[1] **Seiji Fukui**[1] **Nozomu Kamiya**[1]

[1] Fujitsu Ltd., 1-2-70 Konan, Minato-ku, Tokyo 108-0075, Japan

Keywords: Conformation analysis, Cyclic peptides

Drug discovery in middle sized molecules such as cyclic peptides is an attracting area as they have higher selectivity compared to small molecules and higher membrane permeability compared to antibodies. Since the properties depend on their conformation[1], we developed a "Ring Expansion Method"[2] for the generation of cyclic glycine conformers and an addition method of side chain conformers[3] to analyze the stable conformation exhaustively. In CBI annual meeting 2019, we reported these methods and calculation results for peptides cyclized by only the amide bond[4,5]. Recently, we expanded the scope of application to analyze conformations of peptides cyclized by non-amide bonds.

Our new conformation analysis methods are as follows.

(i) Generation of cyclic glycine conformers by "Ring Expansion Method", and addition of heavy atoms between N and CA atom in the main chain to generate non-amide bonds.

(ii) Substitution of a cyclic residue like the proline for a glycine residue and addition of a heavy atom whose position is uniquely determined (CB atom of the alanine or C atom of the N-methylated amide bond). These conformers are named "stem ring conformers".

(iii) Addition of side chain conformers to the stem ring conformers.

We analyzed the conformation for the peptide (YDYPGDYCYLY) cyclized by non-amide bonds (PDB ID: 5KGN). In this study, we analyzed residues of the ring part (YDYPGDYC).

The analysis time was less than 2 hours using 400 cores (Xeon E5-2660 (2.2 GHz)). The conformations of several stable conformers obtained were found to be similar to that of the experimentally observed one. The lowest RMSD between the PDB and obtained conformations is 1.79 Å, here these conformers are superposed on all heavy atoms (Figure 1). When these are superposed on backbone atoms, the lowest RMSD is 1.21 Å.

Our advanced methods achieved the conformation analysis of peptides cyclized by not only the amide bond but also non-amide bonds.

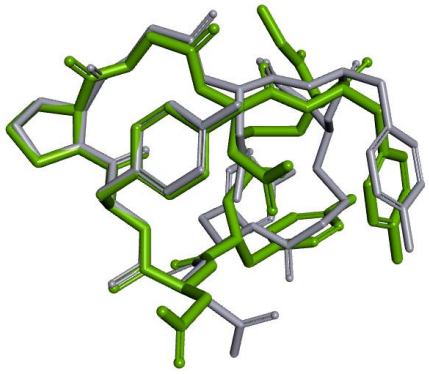

Figure 1: The closest conformer (green line) to the conformer reported in PDB (gray line). These conformers are superposed on all heavy atoms.

[1] A. Whitty, M. Zhong, L. Viarengo, D. Beglov, D. R. Hall and S. Vajda, *Drug Discovery Today*, 21, 712 (2016).
[2] A. Tomonaga, H. Sugiyama and A. Ueda, WO2018/154789A1.
[3] A. Tomonaga, H. Sugiyama and A. Ueda, JP2019/179304A.
[4] S. Mori, *et al.*, CBI annual meeting 2019, P1-27.
[5] K. Takai, *et al.*, CBI annual meeting 2019, P1-28.

Enhanced conformational sampling of cyclic peptide Cyclorasin by coupled Nosé-Hoover equation

Kei Moritsugu[1]
moritugu@yokohama-cu.ac.jp

Koh Takeuchi[2]
koh-takeuchi@aist.go.jp

Yoshifumi Fukunishi[2]
y-fukunishi@aist.go.jp

Ikuo Fukuda[3]
ifukuda@sim.u-hyogo.ac.jp

[1] Graduate School of Medical Life Science, Yokohama City University, Yokohama 230-0045, Japan
[2] National Institute of Advanced Industrial Science and Technology, Cellular and Molecular Biotechnology Research Institute, Tokyo 135-0064, Japan
[3] Graduate School of Simulation Studies, University of Hyogo, Kobe 650-0047, Japan

Keywords: cyclic peptide, Cyclorasin, membrane permeability, enhanced conformational sampling, coupled Nosé-Hoover equation

The importance has been increasing in recent drug discovery to design cyclic peptides that have high affinity to target proteins for inhibition and high membrane permeability for drug delivery. The affinity is estimated by simulating the fluctuation around the protein-peptide complexes and by calculating the associated binding free energies, whereas the membrane permeability should rely on intrinsic flexibility of the cyclic peptides. To reveal the permeability in terms of atomistic viewpoints, we have attempted to fully obtain the structural ensembles of our target cyclic peptides by use of the coupled Nosé-Hoover equation (cNH) of our recent development [1,2].

The original Nosé-Hoover equation generates a canonical distribution at a predetermined static temperature. In cNH, in contrast, a physical system is simulated at nonequilibrium temperature that is fluctuating in a dynamic manner. This corresponds to a temperature replica exchange simulation with continuous temperatures by adopting continuous time development using a single replica. The dynamics are realized by coupling the physical system with a temperature system of small degrees of freedom, both of which are described by the Nosé-Hoover method. The desired distribution of the fluctuating temperature is achieved by the established formalism, allowing an enhanced conformational sampling and a reconstruction of the canonical distribution for the physical system at an arbitrary temperature.

Here, we chose Cyclorasin 9A5 [3] and a derivative as target cyclic peptide systems. Cyclorasin inhibits Ras signaling by blocking Ras-effector protein interactions, which will have therapeutic benefits in cancer patients. A derivative, 9A54, achieved higher affinity than 9A5 after small modifications, which, however, resulted in lower membrane permeability [3]. To explain this controversial result from fully atomistic descriptions, we applied the cNH to sample the structural ensembles of both 9A5 and 9A54 in both explicit water and DMSO environments. The derived structures indicated minor but significant changes in the structural dynamics via the modification from 9A5 to 9A54, which is reasonably in agreement with the NMR experiment [4].

[1] I. Fukuda and K. Moritsugu, Phys. Rev. E (2016) 93: 033306.
[2] I. Fukuda and K. Moritsugu, J. Phys. A: Math. Theoret., in press.
[3] P. Upadhyaya, et al., Angew. Chem. Int. Ed. (2015) 54:7602.
[4] K. Takeuchi, et al., submitted.

O1-14

Development status of ABINIT-MP program in 2020

Yuji Mochizuki[1,2] **Kota Sakakura**[3] **Hiromasa Watanabe**[4] **Koji Okuwaki**[1]
fullmoon@rikkyo.ac.jp sakakura@j-focus.or.jp hi-watanabe@hpc.co.jp okuwaju@rikkyo.ac.jp
Kouichiro Kato[5] **Naoki Watanabe**[6] **Yoshio Okiyama**[7]
kato.koichiro.957@m.kyushu-u.ac.jp naoki.watanabe@mizuho-ir.co.jp okiyama@nihs.go.jp
Kaori Fukuzawa[2,8] **Tatsuya Nakano**[7]
k-fukuzawa@hoshi.ac.jp nakano@nihs.go.jp

[1] Department of Chemistry & Research Center for Smart Molecules, Faculty of Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan
[2] Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
[3] Foundation for Computational Science, 7-1-28 Minatojima Minamicho, Chuo-ku, Kobe, Hyogo 650-0047, Japan
[4] HPC SYSTEMS Inc., Kaigan 3-9-15, Minato-ku, Tokyo 108-0022, Japan
[5] Department of Applied Chemistry, Graduate School of Engineering & Center for Molecular Systems, Kyushu University, Nishi-ku, Fukuoka 819-0395, Japan
[6] Mizuho Information & Research Institute Inc., 2-3 Kanda-nishiki-cho, Chiyoda-ku, Tokyo 101-8443, Japan
[7] Division of Medicinal Safety Science, National Institute of Health, 3-25-26 Tonomachi Kawasaki-ku, Kawasaki, Kanagawa 210-9501, Japan
[8] School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa-Ku, Tokyo 142-8501, Japan

Keywords: Fragment molecular orbital, FMO, ABINIT-MP, Interaction energy analysis

We have been developing the ABINIT-MP program [1] for about 20 years, and this system has been widely used in the fields of pharmaceutical chemistry and biophysics in Japan. The latest distribution package is the Open Version 1 Revision 22 (hereafter denoted as Rev.22) as of June 2020. In Rev. 22, the calculation of PIEDA in ABINIT-MP [2] is greatly accelerated by improving the loop structures; several bottlenecks of processing have been removed. Only the specified region can be calculated at the MP2 or MP3 levels [3] in a multilayer FMO (MFMO) framework [4]. Additionally, the frozen-domain (FD) [5] is supported for partial geometry optimizations.

From April 2020, the trial usage of Fugaku has been provided for ABINIT-MP, and a series of FMO calculations have been performed for several proteins related with COVID-19. With 3072 nodes of Fugaku, an FMO-MP3/cc-pVDZ job for a spike protein (consisting of 3.3 thousand amino acid residues) could be completed in only 3.4 h.

In the presentation, we will show the features of Rev. 22 and some demonstrative results.

[Acknowledgment] This work was supported by the following projects; a social and scientific priority issue #6 to be tackled by using post-K computer (Fugaku) operated by MEXT, Kaken-hi (16H04635), AMED/BINDS (JP19an010101113) and Rikkyo SFR. The trial usage of Fugaku was conducted by RIKEN R-CCS and MEXT under the special project against COVID-19 [6].

[1] Tanaka, S., et al., Electron-Correlated Fragment-Molecular-Orbital Calculations for Biomolecular and Nano Systems, *Phys. Chem. Chem. Phys.*, 16:10310-10344, 2014.
[2] Tsukamoto, T., et al., Implementation of Pair Interaction Energy Decomposition Analysis and Its Applications to Protein-Ligand Systems, *J. Comp. Chem. Jpn.*, 14:1-9, 2015.
[3] Mochizuki, Y., et al., Large-scale FMO-MP3 calculations on the surface proteins of influenza virus, hemagglutinin (HA) and neuraminidase (NA), *Chem. Phys. Lett.*, 493:346-352, 2010.
[4] Fedorov, D.G., et al., Multiconfiguration self-consistent-field theory based upon the fragment molecular orbital method, *J. Chem. Phys.*, 122:054108, 20005.
[5] Fedorov, D.G., et al., Geometry Optimization of the Active Site of a Large System with the Fragment Molecular Orbital Method, *J. Phys. Chem. Lett.*, 2:282-288, 2011.
[6] <https://www.r-ccs.riken.jp/library/topics/fugaku-coronavirus.html>.

Development of multiscale FMO-DPD simulations for molecular design

Koji okuwaki[1] okuwaki@rikkyo.ac.jp
Eiji Shinsho[2] m1813@hoshi.ac.jp
Ruka Nishida[2] s162515@hoshi.ac.jp
Kaori Fukuzawa[2,3] k-fukuzawa@hoshi.ac.jp
Yuji Mochizuki[1,4] fullmoon@rikkyo.ac.jp

[1] Department of Chemistry & Research Center for Smart Molecules, Faculty of Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan
[2] School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa-Ku, Tokyo 142-8501, Japan
[3] Graduate School of Engineering, Tohoku University, 6-6-04, Aramaki Aza Aoba Aoba-ku, Sendai, Miyagi 980-8579, Japan
[4] Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku,Tokyo, 153-8505, Japan

Keywords: FMO-DPD, Multiscale simulation, Dissipative particle dynamics, Reverse map

In the fields of drug design or materials science, control of molecular structures at an atomistic level is highly desired for the advancement of products, and pre-screening or understanding of molecular mechanism using computational chemistry has attracted much attention. In recent years, the prediction of meso-scale structures formed by molecular aggregates is important issue because meso-scale structure has a significant effect on various chemical and biological mechanisms. In order to predict meso-scale structures, simulations using coarse-grained models such as dissipative molecular dynamics (DPD) [1] are commonly used for handling long time-scale and large systems. The parameters that describe the interactions among particles are crucial for these simulations, and they are generally closely related to the so-called χ parameters that correlated with the affinities between components by the Flory-Huggins theory. However, it is well known that the evaluation of a reliable set of χ parameters is a difficult problem, and thus values based on experimental data or empirical values are generally used in DPD simulation. The parameter-dependent applications should have several limitations, and accurate evaluation of χ parameters is still a critical concern.

Our group has developed a new framework for calculating effective interaction parameters [2-4] between particles for meso-scale DPD simulations, based on nano-scale fragment molecular orbital (FMO) calculations [5] by using ABINIT-MP program [6]. In addition, the workflow of detailed FMO analyses of fully atomistic structure generated from meso-scale structure using the reverse map technology has been established recently. In other words, a series of the multiscale FMO-DPD simulation systems have been completed. In this presentation, we outline the frameworks of FMO-DPD and also the demonstrative applications [7-9].

[Acknowledgment] The present work was supported by MEXT as a social and scientific priority issue #6 to be tackled by using post-K computer (FS2020), AMED-BINDS (JP20am0101113), and Rikkyo SFR.

[1] R. Groot et al., J. Chem. Phys. 107 (1997) 4423. [2] C.F. Fan et al., Macromolecules 25 (1992) 3667. [3] K. Okuwaki et al., J. Phys. Chem. B, 122 (2018) 338. [4] K. Okuwaki, et al., J. Comput. Chem. Japan 17 (2018) 102. [5] D.G. Fedorov et al., CRC Press, Boca Raton, 2009. [6] S. Tanaka et al., Phys. Chem. Chem. Phys. 16 (2014) 10310. [7] H. Doi et al., Chem. Phys. Lett. 684 (2017) 427. [8] E. Shinsho et al., J. Comput. Chem. Japan 17 (2018) 172. [9] K. Okuwaki et al., Appl. Phys. Express 13 (2020) 017002.

10月30日（金）

CH2　チャンネル２

口頭発表８

座長：水間 俊（帝京平成大学）、小長谷 明彦（恵泉女学園大学）

02-08 Yusuke Hoshino (Chiba University)
"The prediction of food effect on drug oral absorption by machine learning"

02-09 Mizuki Nakamori (Graduate School of Pharmaceutical Sciences, Nagoya City University)
"Prediction of the inhibitory activity of rat drug-metabolizing enzyme by in silico method"

02-10 Daitaro Misawa (SyntheticGestalt)
"Introduction to the Issues of Dataset Construction in Pharmacokinetic Models"

02-11 Tetsuo Kitamura (Nonclinical Research Center, LSI Medience Corporation)
"Analysis of monkey pose estimation using deep learning."

O2-08

The prediction of food effect on drug oral absorption by machine learning

Yusuke Hoshino[1] **Hideki Yoshioka**[1]

Akihiro Hisaka[1]

[1] Clinical Pharmacology and Pharmacometrics, Graduate School of Pharmaceutical Sciences, Chiba University, 1-8-1, Inohana, Chuo-ku, Chiba 260-8675, Japan

Keywords: Pharmacokinetics, oral absorption, food effects, machine learning, random forest

The prediction of food effect on drug absorption has been previously reported by several methods including BCS, GUT framework and PBPK model, which were investigated based on observed availability changes of approximately 100 drugs. The purpose of present study was to develop a reliable model for prediction of food effect on drug absorption from the chemical structure based on observations of 470 drugs by using machine learning. Changes in oral availability by food were collected from previous reports[1-4] and the product labels. The food effects were classified into negative, no, and positive for drugs whose availability ratios (fed/fasted), mainly AUC ratios, are less than 0.8(14% of the drugs), 0.8 to 1.25 (57% of the drugs), and greater than 1.25 (29% of the drugs), respectively. Approximately 360 features including solubilities in various conditions were calculated from chemical structures with ADMET Predictor (Simulations Plus, Inc.). Features of monotonous or highly correlated were excluded, and then Boruta method was used to select informative features. The data was divided into a training set and a test set, and a training set was further divided into a real training set and a validation set for hyperparameter optimization (nested cross validation). The decision tree algorithm of random forest was used for training of food effect from the selected features. F1-score was used for evaluation of the model performance. F1-scores of the final model in the test set (average of five trainings of different division ways) for negative, no, and positive food effects were 0.304, 0.651 and 0.669, respectively. The F1-score of negative food effect was the smallest due to the lowest frequency of this class (14%). Using information collected in this study, the prediction performance was evaluated also with the BCS method in which drugs were classified by dose, predicted permeability and solubility. F1-scores of negative (class 3), no (class 1), and positive (class 2 and 4) effects were 0.350, 0.379, and 0.580, respectively. The ratio of erroneous negative prediction for positive drugs and that of erroneous positive prediction for negative drugs were 0.058 and 0.166, respectively, in the machine learning model, whereas those in the BCS method were 0.146 and 0.328, respectively. Therefore, the machine learning model predicted changes in availability of drugs more accurately for no or positive food effects and with noticeably less risk of opposite predictions, compared with the BCS method.

[1] Singh, B.N., A quantitative approach to probe the dependence and correlation of food-effect with aqueous solubility, dose/solubility ratio, and partition coefficient (Log P) for orally active drugs administered as immediate-release formulations, Drug Dev Res, 65: 55-75, 2005.

[2] Gu, C.H. *et al.*, Predicting effect of food on extent of drug absorption based on physicochemical properties, *Pharm Res*, 24: 1118-1130, 2007.

[3] O'Shea, J.P. *et al.*, Food for thought: formulating away the food effect - a PEARRL review, *J Pharm Pharmacol*, 71:510-535, 2019.

[4] Deng, J. *et al.*, A Review of Food-Drug Interactions on Oral Drug Absorption, Drugs, 77:1833-1855, 2017.

O2-09

Prediction of the inhibitory activity of rat drug-metabolizing enzyme by *in silico* method

Mizuki Nakamori[1]
c172528@ed.nagoya-cu.ac.jp

Riku Tohno[1]
c162036@ed.nagoya-cu.ac.jp

Kaori Ambe[1]
ambek@phar.nagoya-cu.ac.jp

Masahiro Tohkin[1]
tohkin@phar.nagoya-cu.ac.jp

Takamitsu Sasaki[2]
t-sasaki@u-shizuoka-ken.ac.jp

Kouichi Yoshinari[2]
yoshinari@u-shizuoka-ken.ac.jp

[1] Graduate School of Pharmaceutical Sciences, Nagoya City University, 3-1 Tanabe-dori, Mizuho-ku, Nagoya-shi, Aichi, 467-8603, Japan
[2] School of Pharmaceutical Sciences, University of Shizuoka, 52-1, Yada, Suruga-ku, Shizuoka-shi, Shizuoka, 422-8526, Japan

Keywords: Applicability domain, Drug-metabolizing enzymes, Machine learning method, Synthetic Minority Over-sampling TEchnique

Some chemical substances exhibit their toxicity, especially hepatotoxicity, through their reactive metabolites formed by drug-metabolizing enzymes. In this study, we sought to develop in silico models for predicting the inhibitory activity of chemical substances against rat drug-metabolizing enzymes based on chemical structure information to gain an insight into the chemical-induced toxicity.

We selected 218 commercially available substances from the rat repeated-dose toxicity test database of Hazard Evaluation Support System Integrated Platform (HESS) [1] and used them as test substances for inhibition assays against rat drug-metabolizing enzymes. We used the data of 6 rat cytochrome P450s (CYP1A1, CYP1A2, CYP2B1, CYP2C6, CYP2D1, CYP3A2) and rat UDP-glucuronosyltransferases (UGT) as targets. Substance that showed inhibition by 15% or more were considered as “reactive” and the others as “non-reactive”. The performance of generated prediction models was evaluated in terms of sensitivity and area under the curve (ROC-AUC).

Using the results of the in vitro inhibition assays, we constructed classification models for each target discriminating “reactive” and “non-reactive” using a random forest (RF) algorism with molecular descriptors calculated by Mordred [2]. We randomly divided the 218 substances into the training set and test set at a 3 to 1 ratio, maintaining the reactive/non-reactive ratio. To improve the prediction of imbalanced data, we applied the Synthetic Minority Over-sampling TEchnique (SMOTE). We also set the applicability domains (ADs), which are the chemical spaces that can provide highly reliable prediction results in the test set. The established models showed the following prediction performance: Sensitivity; 0.79 or more, and ROC-AUC; 0.83 or more.

In this study, we have demonstrated that the incorporation of the SMOTE and ADs is useful to develop high-performance and reliable classification models that predict the inhibitory activity of chemical substances against drug-metabolizing enzymes.

This study was supported in part by the Artificial Intelligence-based Substance Hazard Integrated Prediction System project of the Ministry of Economy, Trade and Industry of Japan.

[1] http://www.nite.go.jp/en/chem/qsar/hess-e.html
[2] http://mordred-descriptor.github.io/documentation/master/index.html

Introduction to the Issues of Dataset Construction in Pharmacokinetic Models

Daitaro Misawa[1] d.misawa@syntheticgestalt.com
Yoshihiro Hori[1] y.hori@syntheticgestalt.com
Kotaro Kamiya[1] k.kamiya@syntheticgestalt.com
Reiko Watanabe[2] reiko-watanabe@nibiohn.go.jp
Hitoshi Kawashima[2] hkawashi@nibiohn.go.jp
Masataka Kuroda[2, 3] m-kuroda@nibiohn.go.jp
Kenji Mizuguchi[2] kenji@nibiohn.go.jp

[1] SyntheticGestalt, 1-6 Naitomachi Shinjuku-ku, Tokyo 160-0014, Japan
[2] National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi Ibaraki City, Osaka 567-0085, Japan
[3] Mitsubishi Tanabe Pharma Corporation, 1000 Kamoshida-cho Aoba-ku Yokohama-shi, Kanagawa 227-0033, Japan

Keywords: Pharmacokinetics, Dataset

Developing a new drug normally takes over ten years and costs hundreds of billions of yen. In addition, Significant loss of time and money occurs when development is discontinued due to pharmacokinetic problems. The prediction of pharmacokinetic parameters using machine learning is expected to allow the development of new drugs at a lower cost over a shorter period of time. However, the available data are limited, and it is rare that the framework of machine learning can be applied immediately. In this presentation, we will discuss the problems faced when combining the Well-Stirred Model and machine learning, using the estimation of hepatic clearance (CL_h), one of a pharmacokinetic index, as an example. Machine learning is used for the regression prediction of pharmacokinetic parameters (intrinsic clearance (CL_{int}), fraction unbound in plasma ($f_{u,p,}$), blood/plasma ratio (R_b)). In particular, we point out two problems that were revealed through the construction of evaluation data for CL_h and the model construction of R_b. The first is "lack of data." The number of compounds with a set of ((CL_{int}, $f_{u,p,}$ R_b), CL_h) necessary for the evaluation of CL_h prediction was only 5 out of 5,201 (0.096%). The second is "quality of data." With regard to the distribution of R_b, approximately 20% of all data are unevenly distributed at R_b=1. It can be inferred that 20% of the values for Rb have been set to 1 for some reason. The identification flag is necessary to eliminate the mixture of data, but no metadata for quality assurance was included to flag which values are set and which are actual measurements. Finally, we would like to summarize the insights gained from this activity and propose the data curation policy for machine-learning on pharmacokinetic parameters and its predictions.

[1] Brown RP, Delp MD, Lindstedt SL, Rhomberg LR, Beliles RP. Physiological parameter values for physiologically based pharmacokinetic models. *Toxicol Ind Health*. 1997;13(4):407-484. doi:10.1177/074823379701300401
[2] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ. 2016;47:20-33. doi:10.1016/j.jhealeco.2016.01.012
[3] Levy V. A new future for R&D? Measuring the return from pharmaceutical innovation 2017, Deloitte Centre for Health Solutions 2017(https://www2.deloitte.com/ch/en/pages/ life-sciences-and-healthcare/articles/new-future-for-r-and- d.html)

O2-11

Analysis of monkey pose estimation using deep learning.

Tetsuo Kitamura[1] kitamura.tetsuo@mh.medience.co.jp
Modong Tan[2] modong.tan@phchd.com
Yousuke Ochiai[1] ochiai.yousuke@mn.medience.co.jp
Munehiro Nakagawa[1] nakagawa.munehiro@md.medience.co.jp
Yusuke Suzuki[2] yusuke.suzuki@phchd.com
Takumi Sakimura[2] takumi.sakimura@phchd.com

[1] Nonclinical Research Center, LSI Medience Corporation, 14-1, Sunayama, Kamisu-shi, Ibaraki, 314-0255, Japan
[2] Advanced Technology Development Center, PHC Corporation, Shonan Health Innovation Park, 2-26-1, Muraoka-Higashi, Fujisawa-shi, Kanagawa, 251-8555, Japan

Keywords: Nonclinical study, behavior, deep learning

Recent technological innovations have had a great impact on the development of new drugs, one of which is deep learning. This technology allows us to utilize a much larger amount of data for drug development than humans alone could. This study was conducted with the expectation of discovering a new method of utilizing deep learning in video monitoring of animals, which would allow monitoring for a longer period than previously performed by humans alone. The ultimate goal of this study is to develop a deep learning model that can classify toxic features such as changes in activity level, abnormal posture, ataxia, and vomiting/nausea from videos recorded for a long period of time in normal cages with a shield. As a preliminary step, we herein report on detection of monkey skeleton.

A skeleton detection model was used. In order to avoid bias in the learning data, still images cut out from RGB + depth videos were clustered by image pattern, and still images were selected evenly from each cluster. The coordinates of the head, neck and extremities were input to the selected still images and used as learning data. The skeletal detection was learned by CNN from the still images, and inferences were made based on videos of the same individuals over different time zones.

In the same individuals, the accuracy of the detected parts in the selected still images was 94%.

As a result, identification of the positions of the monkey's head, neck and extremities using a skeleton detection model was considered effective. We will further investigate whether toxic features can be detected from these positions in monkeys showing toxic features.

10月30日（金）

CH2　チャンネル２

口頭発表11

座長：植沢 芳広（明治薬科大学）、
湯田 浩太郎（株式会社インシリコデータ）

O2-12 Haris Hasic (Department of Computer Science, School of Computing, Tokyo Institute of Technology)
"Single-step Retrosynthesis Prediction based on the Identification of Potential Disconnection Sites"

O2-13 Tomohiro Sato (Center for Biosystems Dynamics Research, RIKEN)
"Deep neural network models to predict ion channel inhibitors with pre-learning of activity information of analogous proteins"

O2-14 Ryota Jin (Laboratory of Clinical Pharmacology and Pharmacometrics, Graduate School of Pharmaceutical Sciences, Chiba University)
"Statistically estimating life-long progression of chronic disease from the information of short clinical trials"

O2-15 Yasunari Matsuzaka (Department of Medical Molecular Informatics, Meiji Pharmaceutical University)
"Prediction models of peroxisome proliferator-activated receptor-gamma through a deep learning-based QSAR analysis, DeepSnap"

O2-12

Single-step Retrosynthesis Prediction based on the Identification of Potential Disconnection Sites

Haris Hasic[1] hasic@cb.cs.titech.ac.jp
Takashi Ishida[1] ishida@c.titech.ac.jp

[1] Department of Computer Science, School of Computing, Tokyo Institute of Technology, W8-85, 2-12-1 Ookayama, Meguro, 152-8550 Tokyo

Keywords: Machine Learning – AI Method Development, Descriptor (Parameter) Development

Retrosynthetic analysis or retrosynthesis is a technique used by chemists during the planning of chemical synthesis routes [1]. The manual application of retrosynthesis requires a lot of knowledge and experience to identify possible transformations for a given target compound. Furthermore, since the complexity scales with the complexity of the targets, efficient application on compounds with complex structures becomes an almost impossible task for human chemists. The idea of using computers in such situations has existed for a long time, but the accuracy was insufficient for practical applications. However, with the steady improvement of artificial intelligence in recent years, computer-assisted retrosynthesis has been gaining research attention again. Because of the overall lack of available chemical reaction data, the main challenge of recent computer-assisted retrosynthesis approaches is low exploration ability during the analysis of target and intermediate compounds [2]. This analysis is referred to as single-step retrosynthesis.

In this research, we developed a novel, template-free approach for the single-step retrosynthesis task for any given molecule. To ensure general applicability of the approach, we used only individual molecular substructures of the target compound to determine potential disconnection sites, without relying on additional information. The model for the identification of potential disconnection sites was constructed to use a novel molecular substructure fingerprint representation based on the extended connectivity fingerprints [3]. Then, a simple structural similarity-based reactant retrieval and scoring implementation was applied for each of the identified disconnection sites. As a result of the evaluation experiments, the proposed method outperforms the state-of-the-art method [4] for the single-step retrosynthesis task by achieving 48.9% Top-1 accuracy on the pre-cleaned U.S. Patent Office dataset [5]. If the reaction class predicted by the model was used to further narrow down reactant candidate search space, the performance is additionally improved to 61.4% Top-1 accuracy [6].

.

[1] Clayden, J., Greeves, N., and Warren, S., *Organic Chemistry*, Oxford University Press, 2012.
[2] Coley, C.W., Rogers, L., Green, W.H., and Jensen, K.F., Computer-Assisted Retrosynthesis based on Molecular Similarity, *ACS Central Science*, 3(12):1237-1245, 2017.
[3] Rogers, D., and Hahn, M., Extended-connectivity Fingerprints, *Journal of Chemical Information and Modelling*, 50(5):742-754, 2010.
[4] Zheng, S., Rao, J., Zhang, Z., Xu, J., and Yang, Y., Predicting retrosynthetic reactions using self-corrected transformer neural networks. *Journal of Chemical Information and Modelling*, 60:47-55, 2020.
[5] Schneider, N., Lowe, D.M., Sayle, R.A., Tarselli, M.A., and Landrum, G.A., Big Data from Pharmaceutical Patents: A Computational Analysis of Medicinal Chemists' Bread and Butter. *ACS Journal of Medicinal Chemistry*, 59(9):4385-4402, 2016.
[6] Hasic H., and Ishida T., Single-step Retrosynthesis Prediction based on the Identification of Potential Disconnection Sites using Molecular Substructure Fingerprints. Submitted to the *BMC Journal of Chemoinformatics*, 2020.

O2-13

Deep neural network models to predict ion channel inhibitors with pre-learning of activity information of analogous proteins

Tomohiro Sato **Teruki Honma**

tomohiro.sato@riken.jp honma.teruki@riken.jp

Center for Biosystems Dynamics Research, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan

Keywords: deep neural network, pre-learning, machine learning, ion channel, cardiotoxicity

The inhibitions of ion channels are closely related to various drug side effects such as torsade de pointes caused by the inhibition of hERG potassium channel. Recent studies revealed that the inhibition of various cardiac ion channels could result in diverse cardiotoxicities[1]. Comprehensive prediction of inhibitory activities for cardiac ion channels could greatly reduce the risk of cardiotoxicities in drug discovery. Though we reported the construction of a highly accurate hERG prediction model by integrating information dispersing in various databases[2,3], the limited number of inhibitory activity information could become a stabling block to construct accurate statistical models for most of the other cardiac ion channels. In this study, we demonstrated that the pre-learning about activity information of analogous proteins could effectively improve the prediction performances of deep neural network (DNN) models for ion channels with less activity information.

In the case of discrimination analysis of Nav1.5, Kv1.5, and Cav1.2 inhibitors, the pre-learning of other sodium channels, potassium channels, and calcium channels improved the kappa statistics from 0.68 by normal DNN model to 0.73 by pre-learned DNN model for Nav1.5, from 0.496 to 0.545 for Kv1.5, and from 0.387 to 0.548 for Cav1.2 (Figure 1), suggesting the usefulness of pre-learning technique when the activity information of the target protein is insufficient to construct an accurate statistical model. The prediction models were available at https://drugdesgin.riken.jp/hERG.

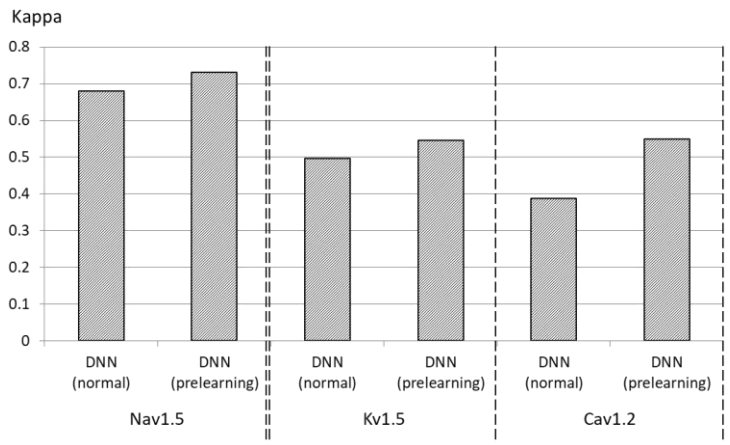

Figure 1. Kappa statistics of normal and pre-learned DNN models for Nav1.5, Kv1.5, and Cav1.2.

References

[1] Bartos DC, Grandi E, Ripplinger CM. Ion Channels in the Heart. *Compr Physiol*. **2015,** 5(3), 1423-1464

[2] Sato, T., Yuki, H., Ogura, K., Honma, T. Construction of an integrated database for hERG blocking small molecules. *PLOS ONE* **2018**, 13(7)

[3] Ogura, k., Sato, T., Yuki, H., Honma, T. Support Vector Machine model for hERG inhibitory activities based on the integrated hERG database using descriptor selection by NSGA-II. *Sci. Rep*. **2019**, 9, 12220

O2-14

Statistically estimating life-long progression of chronic disease from the information of short clinical trials

Ryota Jin[1]
ryota-j@chiba-u.jp

Shinya Kawamatsu[1, 2]
s.kawamatsu.gifu@gmail.com

Shogo Araki[1]
shogo.1202.tandf@gmail.com

Hiromi Sato[1]
hiromi-s@chiba-u.jp

Yasunori Sato[3]
Yasunori.sato@keio.jp

Akihiro Hisaka[1]
hisaka@chiba-u.jp

[1] Laboratory of Clinical Pharmacology and Pharmacometrics, Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, 260-8675, Japan.
[2] Clinical Operations, Japan Development, GlaxoSmithKline, Tokyo, Japan.
[3] Department of Preventive Medicine and Public Health, Keio University School of Medicine, Tokyo, Japan

Keywords: Nonlinear mixed-effect model, Population pharmacokinetics analysis, Long-term pathological model, Chronic disease

Progressive central nervous system diseases including Alzheimer's disease and Parkinson's disease, and various chronic diseases including diabetes, cardiovascular disease, COPD and rheumatism, persistently progress throughout life. The change of biomarkers associated with the progression and factors that affect the progression are essential information as the basis of remedy, but they are usually difficult to elucidate due to the need for decades of cohort studies. However, if a large number of patients with different degrees of disease progression are enrolled, even with short-term studies, that is, in case of that the ergodic hypothesis (anzats) in mathematics is approximately satisfied for patients, it is possible to estimate the long-term progression. Based on this concept, we developed SReFT (Statistical Restoration of Fragmented Time course) and analyzed the progression of Alzheimer's disease [1]. In this study, the applications of SReFT to Parkinson's disease and COPD [2] will be described. In each case, we succeeded in analyzing the change of biomarkers for more than 30 years and factors influencing progressions such as gene mutation and smoking from individual information of several biomarkers of hundreds to thousands of patients. The advantages of SReFT have become apparent, such as detecting long-term changes that could not be according to conventional analyses and distinguishing early and late changes in the disease. In this presentation, we will outline the algorithm of SReFT, which can be said to be a new type of machine learning, along with the currently available analysis environments, and report the results of clinical data analysis.

[1] Ishida, T., *et al.*, A Novel Method to Estimate Long-Term Chronological Changes From Fragmented Observations in Disease Progression, *Clin Pharmacol Ther*, 105:436-447, 2019.
[2] Kawamatsu, S., et al., Scores of health-related quality of life questionnaire worsen consistently in patients of COPD: estimating disease progression over 30 years by SReFT with individual data collected in SUMMIT trial, *J Clin Med*, 9:2676, 2020

O2-15

Prediction models of peroxisome proliferator-activated receptor-gamma through a deep learning-based QSAR analysis, DeepSnap

Yasunari Matsuzaka **Yoshihiro Uesawa**

matsuzys@my-pharm.ac.jp uesawa@my-pharm.ac.jp

Department of Medical Molecular Informatics, Meiji Pharmaceutical University, Tokyo,

Keywords: DeepSnap, Deep learning, peroxisome proliferator-activated receptor-gamma, QSAR, Tox21,

We have previously reported a novel technique for capturing images of chemical compounds for deep learning (DL)-based quantitative structure-activity relationship (QSAR) analysis, namely, DeepSnap-DL [1,2]. This approach can automatically perform feature extraction using snapshot images derived from 3D chemical structures and create input image using data augmentation by angle-designated rotation, zoom specification, and RGB conversion, that expect to increase in the amount of information on chemical structure, not just data padding and to prevent over-fitting that occur when learning is performed only with to match training data [1-5]. Thus, it has been shown that the prediction performances of the DeepSnap-DL approach were higher than those of conventional machine learning, including random forest, XGBoost, LightGBM, and CatBoost [3-5]. In general, the perturbation of nuclear receptor (NR) signaling pathway due to unwanted interactions of chemical compounds with some NRs is associated with various adverse effects on human health. To develop new chemical risk assessment approaches, we established cell-based *in vitro* high-throughput screening assays for agonist/antagonist activities of some NRs using the reporter assay in Toxicology in the 21st Century (Tox21) program [6]. Among these NRs, peroxisome proliferator-activated receptor-gamma (PPARg) plays a paramount role as an antioxidant response in tissue injury. Hence, we constructed prediction models of PPARg agonist activity by the DeepSnap-DL approach using the Tox21 chemical library.

[1] Uesawa, Y. Quantitative structure-activity relationship analysis using deep learning based on a novel molecular image input technique, Bioorg Med Chem Lett, 28(20), 3400-3403, 2018.
[2] Matsuzaka, Y., Uesawa, Y. Optimization of a Deep-Learning Method Based on the Classification of Images Generated by Parameterized Deep Snap a Novel Molecular-Image-Input Technique for Quantitative Structure-Activity Relationship (QSAR) Analysis, Front Bioeng Biotechnol, 7:65, 2019.
[3] Matsuzaka, Y., Uesawa, Y. Prediction Model with High-Performance Constitutive Androstane Receptor (CAR) Using DeepSnap-Deep Learning Approach from the Tox21 10K Compound Library, Int J Mol Sci, 20(19):4855, 2019.
[4] Matsuzaka, Y., Uesawa, Y. DeepSnap-Deep Learning Approach Predicts Progesterone Receptor Antagonist Activity With High Performance, Front Bioeng Biotechnol, 7:485, 2020.
[5] Matsuzaka Y, Hosaka T, Ogaito A, Yoshinari K, Uesawa Y. Prediction Model of Aryl Hydrocarbon Receptor Activation by a Novel QSAR Approach, DeepSnap-Deep Learning, Molecules, 25(6):1317, 2020.
[6] Matsuzaka, Y., Uesawa, Y. Molecular Image-Based Prediction Models of Nuclear Receptor Agonists and Antagonists Using the DeepSnap-Deep Learning Approach with the Tox21 10K Library, Molecules, 25(12):2764, 2020.

10月30日（金）

CH3　チャンネル３

口頭発表９

座長：荻島 創一（東北大学 東北メディカル・メガバンク機構）、
茂櫛 薫（富士通株式会社）

03-08 Shigeyuki Magi (Toho University)
"Systems biological approaches to understand the mechanisms of cancer drug resistance & cardiotoxicity"

03-09 Satoshi Nagaie (Department of Informatics for Genomic Medicine, Tohoku Medical Megabank Organization, Tohoku University)
"Development of phenotyping algorithm of type 2 diabetes mellitus: A retrospective claims database analysis"

03-10 Naoko Kasahara (Department of Informatics for Genomic Medicine, Tohoku Medical Megabank Organization, Tohoku University)
"Integrated Database “dbTMM” with family linkage for stratification of cohort participants toward drug development"

03-11 Tsuyoshi Esaki (Shiga University)
"Estimation of relationships between chemical substructures and gene expression antibiotic-resistance of bacteria: Adapting canonical correlation analysis for small sample data by gathered features using consensus clustering"

O3-08

Systems biological approaches to understand the mechanisms of cancer drug resistance & cardiotoxicity

Shigeyuki Magi[1] **Atsuhiko T Naito**[1]
Shigeyuki.magi@med.toho-u.ac.jp atsuhiko.naito@ med.toho-u.ac.jp

[1] Toho University, 5-21-16, Omori-Nishi, Ota-ku, Tokyo,143-8540, Japan

Keywords: drug resistance, cardiomyocyte dysfunction, mathematical modeling

Two major challenges in cancer drug therapy are the resistance of cancer cells and the toxicity on normal cells. Here, we introduce our approaches to understand the mechanisms of drug resistance against estrogen receptor antagonist and the cardiotoxicity against anti-HER2 monoclonal antibody.

The process of drug resistance in cancer cells can be understood as a series of changes in the molecular network in which cancer cells go through an unstable state where cells acquire heterogeneity due to the external stress of anticancer drugs and subsequent transition of surviving cells to another stable survival state. We tried to elucidate the molecular mechanisms contributing to the process of acquiring drug resistance by analyzing the changes in gene expression of breast cancer cells after the treatment with tamoxifen (TAM: estrogen receptor antagonist). We performed RNA-sequencing every week after the TAM treatment and identified that genes regulated by TGFβ, NOTCH, HIF-1 and cell adhesion molecules were increased after the growth rate of the cells was recovered, suggesting that these genes may characterize and contribute to the drug resistance against TAM. We also identified that genes related to FOXO signaling and autophagy were increased at the border period just before the growth rate of the cells was recovered, suggesting that these genes play role in “acquisition” of drug resistance and may become the biomarkers for pre-resistant state. To further analyze the transition trajectories to resistant cell types, we performed a single-cell RNA-sequencing after TAM treatment and revealed that resistant cells consisting two major subpopulations with different characteristics emerge from the pre-resistant state. We are currently constructing a mathematical model capable of reproducing the changes in subpopulation during the process of resistance acquisition.

The anti-HER2 antibody trastuzumab is known to have side effects that cause cardiac dysfunction and heart failure, but its effect on physiological HER2 signaling to human cardiomyocytes remains elusive. We report our preliminary results regarding the effects of trastuzumab on the HER2 signaling in human iPS cell-derived cardiomyocytes and human breast cancer cell lines and introduce our plan for the future research.

O3-09

Development of phenotyping algorithm of type 2 diabetes mellitus: A retrospective claims database analysis

Satoshi Nagaie[1]
nagaie@megabank.tohoku.ac.jp

Satoshi Mizuno[1]
samizuno@med.tohoku.ac.jp

Ryosuke Ishiwata[1]
rishiwata@megabank.tohoku.ac.jp

Naoko Kasahara[1]
nagaie@megabank.tohoku.ac.jp

Soichi Ogishima[1]
ogishima@megabank.tohoku.ac.jp

[1] Department of Informatics for Genomic Medicine,
Tohoku Medical Megabank Organization, Tohoku University,
Seiryo 2-1, Aoba-ku, Sendai Miyagi 980-8573 Japan

Keywords: Electronic Health Record, Phenotyping, Claims Database

As the research and development of genomic medicine based on individual genome information has progressed, highly accurate and deep genome information on genotype is being studied. On the other hand, for phenotypic information, more in-depth information on disease types as well as disease names is required. In such a situation, research is being conducted to obtain deeper information on disease types by secondary use of real-world data, such as medical information. The classification of disease types based on medical information is called “phenotyping”. In the U.S., the eMERGE project is constructing a database that aggregates 68 different pathological classification algorithms called PheKB, which is the result of extracting medical information from EHRs and classifying them into disease types.

In this study, we develop a phenotyping algorithm for type 2 diabetes mellitus (T2DM) using 3,014,656 subjects' receipt data and 3,156,110 subjects' specific health checkup data, especially medical practices (ICD-10 codes = E11)), laboratory test values (random glucose ≥ 200 mg/dL, HbA1c ≥ 6.5%) and medications (All drugs for the treatment of T2DM that have been reviewed for approval by the PMDA), registered from January 2005 to May 2019, among the data provided by JMDC, Inc. In order to evaluate the algorithm, we calculate sensitivity and specificity, which will need to be labeled with the correct answer. For this purpose, it is usually necessary for physicians to review medical records again, which is time-consuming and costly. In this study, we evaluate the performance of the algorithm using a prediction model of disease type classification as a probabilistic gold standard, without a physician's review of medical records.

O3-10

Integrated Database “dbTMM” with family linkage for stratification of cohort participants toward drug development

Naoko Kasahara[1]
n.kasahara@megabank.tohoku.ac.jp

Satoshi Mizuno[1]
samizuno@med.tohoku.ac.jp

Ryosuke Ishiwata[1]
rishiwata@megabank.tohoku.ac.jp

Satoshi Nagaie[1]
nagaie@megabank.tohoku.ac.jp

Yukie Konno[1]
y.konno@megabank.tohoku.ac.jp

Hiroshi Tanaka[2]
tanaka@megabank.tohoku.ac.jp

Soichi Ogishima[1]
ogishima@megabank.tohoku.ac.jp

[1] Department of Informatics for Genomic Medicine, Tohoku Medical Megabank Organization, Tohoku University, Seiryo 2-1, Aoba-ku, Sendai Miyagi 980-8573 Japan
[2] Department of Bioclinical Informatics, Tohoku Medical Megabank Organization, Tohoku University, Seiryo 2-1, Aoba-ku, Sendai Miyagi 980-8573 Japan

Keywords: Biobank, Cohort, Stratification, Drug Discovery

Tohoku Medical Megabank project was started to reconstruct from the Great East Japan Earthquake and Tsunami. We started cohort studies in 2013, and successfully recruited 84,000 residents for Community-Based cohort and 73,000 people for Birth and Three-Generation cohort, totally more than 150,000 participants for baseline assessment by 2017. Though these cohorts, biospecimens including serum, plasma, mononuclear cells, urine were collected more than 3.8 million sample storage (July 2020) and the answers to the questionnaires of their lifestyles including food, psychological condition, experiences of the disaster were also collected. A part of participants took more than 10 physiological examinations and also MRI examination in assessment centers and satellites.

We have developed the integrated database of Tohoku Medical Megabank project (dbTMM). In this July, we released collected data with family linkage of 68,000 subjects of Birth and Three-Generation cohort. Currently, the dbTMM stores large-scale data of 145,000 cohort participants including whole genome (n=3,500) and SNP genotyping data (n=80,000) (genetic factors), questionnaire data (environmental factors), family linkage data (n = 20,000), MRI data (n=4,300) and metabolome and proteome data (n=6,000) (molecular phenotypic factors), and clinical data (phenotypic factors). To utilize family linkage data, we developed a search function with family linkage to enable researchers to stratify cohort participants having a certain phenotype by genetic and environmental factors considering family linkage. As for clinical data, we collected medical charts and stored them into dbTMM. By using structured clinical data, we have conducted deep phenotyping of obstetric disease including hypertensive disorders of pregnancy. We believe our integrated database dbTMM will contribute to stratification of cohort participants toward drug development for common disease which is caused by complex interplay between genetic and environmental factors.

[1] Kuriyama, S. *et al.* The Tohoku Medical Megabank Project: design and mission. *J. Epidemiol.* 26, 493–511 (2016).

Estimation of relationships between chemical substructures and gene expression antibiotic-resistance of bacteria: Adapting canonical correlation analysis for small sample data by gathered features using consensus clustering

Tsuyoshi Esaki[1]
tsuyoshi-esaki@biwako.shiga-u.ac.jp

Takaaki Horinouchi[2]
takaaki_horinouchi@rike.jp

Yayoi Natsume-Kitatani[3,4]
natsume@nibiohn.go.jp

Yosui Nojima[3]
yosui@nibiohn.go.jp

Iwao Sakane[5]
i-sakane@itoen.co.jp

Hidetoshi Matsui[1,6]
hmatsui@biwako.shiga-u.ac.jp

[1] The Center for Data Science Education and Research, Shiga University, 1-1-1 Banba, Hikone, Shiga 522-8522, Japan
[2] Center for Biosystems Dynamics Research, RIKEN, 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan
[3] Artificial Intelligence Center for Health and Biomedical Research, National Institutes of Biomedical Innovation, Health and Nutrition, 7-8-6 Saito Asagi, Ibaraki, Osaska 567-0085, Japan
[4] Center of Drug Design Research, National Institutes of Biomedical Innovation, Health and Nutrition, 7-8-6 Saito Asagi, Ibaraki, Osaka 567-0085, Japan
[5] Central Research Institute, ITO EN Ltd., 21 Megami, Makinohara, Shizuoka 421-0516, Japan
[6] Faculty of Data Science, Shiga University, 1-1-1 Banba, Hikone, Shiga 522-8522, Japan

Keywords: canonical correlation analysis, clustering, antibiotic resistance

The emergence of antibiotic-resistant bacteria is a serious public health concern worldwide. To understand the drug resistance mechanisms employed by bacteria, quantitative phenotypic changes associated with drug resistance acquisition in *Escherichia coli* have been investigated. However, relationships between specific antibiotic structures and the acquisition of resistance by bacteria have not been clarified. Understanding such relationships will both provide new insights into the control of bacterial drug resistance and support the development of novel antibiotic drugs.

To investigate relationships between chemical structures and the acquisition of drug resistance, high dimensional datasets can be treated as two different datasets. To analyze such datasets, we applied canonical correlation analysis (CCA) and analyze the relationships between chemical substructures and phenotypic changes related to the acquisition of drug resistance, the mRNA transcript expression levels of 4444 genes from drug-resistant *E. coli* evolved *in vitro* for drug resistance to 10 antibiotic compounds. Structure data files (SDF) of 10 compounds were collected from ChEMBL and Morgan fingerprint with 2 radius was used to generate 2048 substructures from SDF in rdkit (ver. 2019.09.3) within Python.

To gather lots of features, consensus clustering was employed and proper cluster number were determined to be 5 for substructure and transcriptome. Average values in the same clusters were calculated to perform CCA.

From a two-dimensional scatter plot, we calculated two orthogonal canonical coefficient vectors under the correlation contributions of 0.901 and 0.822 for canonical coefficients 1 and 2, respectively. Overall, we observed clusters with large coefficient values in each axis. These findings suggest that these clusters represent different mechanisms of resistance for each class of antibiotics.

This work was supported in part by research grant for priority areas from Shiga University.

10月30日（金）

CH3　チャンネル３

口頭発表12

座長：清水 祐吾（慶應義塾大学）

03-12 Satoshi Mizuno (Tohoku Medical Megabank Organization, Tohoku University)
"Development of machine learning based prediction model for hypertensive disorders of pregnancy (HDP)"

03-13 Kota Fujisawa (Graduate school of Engineering and Science, University of the Ryukyus)
"Selecting the genes related to COVID-19 with PCA-based unsupervised feature extraction"

O3-12

Development of machine learning based prediction model for hypertensive disorders of pregnancy (HDP)

Satoshi Mizuno[1] samizuno@med.tohoku.ac.jp
Satoshi Nagaie[1] satoshi.nagaie@gmail.com
Gen Tamiya[2] gtamiya@genetix-h.com
Shinichi Kuriyama[3] kuriyama@med.tohoku.ac.jp
Hiroshi Tanaka[4] tanaka@cim.tmd.ac.jp
Nobuo Yaegashi[5] nobuo.yaegashi@gmail.com
Junichi Sugawara[6] jsugawara@med.tohoku.ac.jp
Soichi Ogishima[1] ogishima@megabank.tohoku.ac.jp

[1] Department of Informatics for Genomic Medicine, Group of Integrated Database Systems, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan
[2] Department of Statistical Genetics and Genomics, Group of Disease Risk Prediction, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan
[3] Department of Molecular Epidemiology, Group of the Birth and Three-Generation Cohort Study, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan
[4] Department of Bioclinical Informatics, Group of Integrated Database Systems, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan
[5] Department of Gynecology and Obstetrics, Tohoku University Graduate School of Medicine, Tohoku University, Miyagi, Japan
[6] Department of Feto-Maternal Medical Science, Group of Community Medical Supports, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan

Keywords: Machine learning, hypertensive disorders of pregnancy, Phenotyping

In the resent years, machine learning (ML) is widely applied to clinical tasks including detection of tumors from medical images [1] and predict clinical events from electronic health records（EHRs）[2]. ML is also expected to be applied to early prediction of common diseases such as diabetes and heart failure from various data including genetic factors and exposures. As major limitations underlying current effort to early prediction of common diseases from big data, static and informatic issues including multi-modality, high dimension and variety of data remain to be addressed.

In this study, we developed prediction model of hypertensive disorders of pregnancy (HDP) with 22,256 pregnancy women in the BirThree cohort study [3]. Time-series exposures, laboratory tests and medical records were used as input data. To address static and informatic issue, we compared predictive power of several types of ML models and preprocessing methods in combination. Both training and test labels were identified by previously developed precise phenotyping algorism (PPV = 0.94). Evaluation of predictive powers were performed based on ten-fold cross validation.

The predictive power of developed ML model was up-to 0.95 of F1 score. Among the developed models, interpretable models show high importance for blood pressure around the mean of onset, eating habit and lifestyle.

Our developed ML models enable us not only to conduct risk prediction but also knowledge acquisition for drug development of HDP.

[1] S Kudo *et al, Artificial Intelligence-assisted System Improves Endoscopic Identification of Colorectal Neoplasms*, Clin Gastroenterol Hepatol, 18(8):1874-1881.e2, 2020
[2] A Rajkomar *et al, Scalable and accurate deep learning with electronic health Records*, NPJ Digit Med, 8;1:18, 2018.
[3] S Kuriyama *et al*, *The Tohoku Medical Megabank Project: Design and Misson*, J Epidemiol, 26(9):493-511, 2016.

O3-13

Selecting the genes related to COVID-19 with PCA-based unsupervised feature extraction

Kota Fujisawa[1] k198422@eve.u-ryukyu.ac.jp
Mamoru Shimo[2] e175238@eve.u-ryukyu.ac.jp
Yoshihiro Taguchi[3] tag@granular.com
Shinya Ikematu[4] ikematsu@okinawa-ct.ac.jp
Ryota Miyata[2] miyata26@tec.u-ryukyu.ac.jp

[1] Graduate school of engineering and science, University of the Ryukyus, 1 Senbaru, Nishihara, Nakagami, Okinawa 903-0213, Japan
[2] Faculty of Engineering, University of the Ryukyus, 1 Senbaru, Nishihara, Nakagami, Okinawa 903-0213, Japan
[3] Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo, Tokyo 112-8551,Japan
[4] Department of Bioresources Engineering, National Institute of Technology, Okinawa College, 905 Henoko, Nago, Okinawa 905-2192, Japan

Keywords: COVID-19, SARS-CoV-2, machine learning, gene selection

Coronavirus disease 2019 (COVID-19) is raging all over the world. This potentially fatal infectious disease is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). However, the mechanism of COVID-19 is not well understood. Therefore, we analyze the gene expression profiles of COVID-19-infected patients to identify the disease-related genes using an innovative machine learning method, which allows us to perform gene selection from a dataset with small samples and many candidates based on a data-driven strategy.

First, we applied Principal-components-analysis-based unsupervised feature extraction (PCAUFE [1]) to the mRNA expression profiles of 17 patients and 17 healthy controls (GSE152418 [4]), identifying 123 genes as critical for COVID-19 progression from 60,683 candidate genes. Second, we also applied PCAUFE to GSE1739 [2], a dataset of SARS, which was the other bat-origin coronavirus disease and was caused by SARS-CoV. An integrated analysis of the two datasets revealed 83 genes uniquely selected from the COVID-19 dataset, such as B2M, EIF4G2, and HLA-DPA1. Moreover, we also found 40 genes commonly selected from both the datasets such as CD74, HLA-DRA, and HLA-DRB1.

Finally, to investigate the biological reliability of these selected genes, we uploaded them an enrichment analysis sever called GeneSetDB [3]. Both the 83 genes unique to the COVID-19 dataset and the 40 genes common to these zoonotic-coronavirus datasets mainly included immune-related genes. These results suggest that PCAUFE could successfully identify a biologically feasible set of COVID-19-related genes.

[1] Taguchi, Y., *Unsupervised Feature Extraction Applied to Bioinformatics: A PCA Based and TD Based Approach*, Springer, 2019.
[2] Reghunathan, R., *et al.*, Expression profile of immune response genes in patients with Severe Acute Respiratory Syndrome, *BMC Immunol*, 6:2, 2005.
[3] https://genesetdb.auckland.ac.nz/haeremai.html
[4] https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152418