P01-01

Cyclosporin A: Conquering Conformational Complexity and Chameleonicity

Satoshi Ono[1] **Matthew R. Naylor**[2] **Chad E. Townsend**[2]
Ono.satoshi@mg.mt-pharma.co.jp

Chieko Okumura[1] **Okimasa Okada**[1] **Hsiau-Wei Lee**[2]

R. Scott Lokey[2]

[1] Discovery Technology Laboratories, Innovative Research Division, Mitsubishi Tanabe Pharma Corporation, 1000, Kamoshida-cho, Aoba-ku, Yokohama, Kanagawa 227-0033, Japan
[2] Department of Chemistry and Biochemistry, University of California Santa Cruz, Santa Cruz, California 95064, United States

Keywords: Cyclic peptide, Conformation, bRo5 compound, Multicanonical MD

The chameleonic behavior of cyclosporin A (CsA) was investigated through conformational ensembles employing multicanonical molecular dynamics simulations that could sample the wide range of conformational space including cis and trans isomers of N-methylated amino acids. These assessments were conducted in various explicit solvents ranging polar to apolar using AMBER ff03, AMBER10:EHT, AMBER12:EHT, and AMBER14:EHT force fields. The conformational details were discussed employing the free-energy landscapes (FELs) at T = 300 K. It was observed that the experimentally determined structures of CsA were only a part of the conformational space. Comparing the ROESY measurements in cyclohexane-d12 and hexane-d14, the major conformations in those apolar solvents were essentially the same as that in $CDCl_3$ except for the observation of some sidechain rotamers. The effects of the metal ions on the conformations, including the cis/trans isomerization, were also investigated. Based on the analysis of FELs, it was concluded that the AMBER ff03 force field best described the experimentally derived conformations, indicating that CsA intrinsically formed membrane-permeable conformations and that the metal ions might be key to the cis/trans isomerization of N-methylated amino acid before binding a partner protein.

[1] Ono, S.; Naylor, M. R.; Townsend, C. E.; Okumura, C.; Okada, O.; Lee, H.-W.; Lokey, R. S. Cyclosporin A: Conformational Complexity and Chameleonicity. *J. Chem. Inf. Model.* **2021**, *61*, 5601−5613.

P01-02

Toward a *in silico* screening of broad-spectrum antibody that is effective against wide variety of variants

Kanji Oshima[1] **Shinichi Yoshida**[1]
Kanji.Oshima@kaneka.co.jp Shinichi.Yoshida@kaneka.co.jp

Hiroshi Kita[1]
Hiroshi.Kita@kaneka.co.jp

[1] KANEKA CORPORATION 1-8, Miyamae-cho, Takasago-cho, Takasago Hyogo 676-8688 Japan

Keywords:broad spectrum antibody, In silico screening, mutation, fluctuation

Influenza viruses and novel coronaviruses such as SARS-CoV-2 are prone to frequent mutations of the viral genome, resulting in mutations in the amino acid sequence of the surface protein of the virus that can be recognized as an antigen (epitope site of neutralizing antibodies), changing its antigenicity. As a result, antibodies that have neutralizing ability in the wild type become less effective in variants. Therefore, attention is being paid to the development of broad-spectrum neutralizing antibodies that are less susceptible to viral mutations. In the study, we have focused on viral genome mutations and investigated a method of predicting epitope sites where mutations will seldom occur in the future.

We have obtained the amino acid sequence data of human influenza A hemagglutinin (HA) from the NCBI influenza-related protein sequence database (Influenza Virus Resources) and performed sequence analysis (total 45,545 sequences, unique sequence after removing duplicated sequences ~14.173 sequences), We have found that many amino acid residues (560 out of 565 residues in total length) accumulated mutations during the period from 1918 to about 100 years. Next, we have performed MD (Molecular Dynamics) analysis using the oldest HA sequence structure in 1918 (PDBID:3gbn), and calculated the fluctuation of each amino acid residue (rmsf). As a result of examining the relationship between fluctuation and mutation accumulation rate, we have found that the mutation accumulation rate tended to decrease as the fluctuation became smaller. Furthermore, in a known broad-spectrum neutralizing antibody for HA (CR6261), we have confirmed that most of the amino acid residues of the epitope are characterized by small fluctuations, small mutation accumulation rates, and exposure to antibody-accessible surfaces.

These results suggest that if the amino acid residues on the antibody-accessible surface have small fluctuations, they will be less likely to be mutated in the future, then neutralizing antibodies with these amino acid residues as epitopes become effective therapeutic agents (broad-spectrum antibodies) against various mutant strains in the long term.

Structure prediction of cyclic peptides in solvent with integrating multiple technologies

Kentaro Takai takai.kentaro@fujitsu.com
Yoshiaki Tanida tanida.yoshiaki@fujitsu.com
Toshio Manabe tmanabe@fujitsu.com
Chieko Terashima c.terashima@fujitsu.com
Mitsuhito Wada mwada@fujitsu.com
Takashi Mitsui tmitsui@fujitsu.com
Hiroyuki Sato shryk@fujitsu.com

Fujitsu Ltd, 1-5 Omiya-cho, Saiwai-ku, Kawasaki-shi, Kanagawa, 212-0014, Japan

Keywords: structure prediction, cyclic peptides

Drug discovery in middle sized molecules such as cyclic peptides is an attracting area because they have higher selectivity to target proteins and lower toxicity compared to small molecules and higher membrane permeability and lower production cost compared to antibodies. Although these properties depend on their structures, the standard technology to analyze their large conformational space has not yet been established. Thus, we developed a method to predict three-dimensional structures of cyclic peptides in solvent.

Our proposed method consists of multiple technologies. The procedure is as follows (Figure 1). (i) We prepared a coarse-grained model for a sequence of an input peptide, where one amino acid is represented as one particle [1]. (ii) We explored the stable configuration in this model by Fujitsu Quantum-inspired Computing Digital Annealer [2]. The interaction potentials between particles are predicted by AI. (iii) We transformed the coarse-grained model to the all-atom model using the monomer structures in DB. (iv) We prepared input files for REST2 simulation [3]. Here, we used our original force field (FF-FOM) [4] and the RESP charges of monomers which are registered in DB. (v) We performed REST2 simulations on high-performance computing. (vi) We extracted representative structures by density peak clustering [5].

We predicted the stable structure of Cyclosporin A in the chloroform solvent and found that the cluster center structure in most frequent cluster was similar to the experimentally observed one (RMSD of the main chain and CB atoms is 0.87 Å). Although expert knowledge is required to use above each technology, we automated our method so that anyone can use it easily.

Our method achieved the proposal of a few stable structures of cyclic peptides in solvent. Aiming to improve accuracy and processing speed, we are energetically developing our technologies.

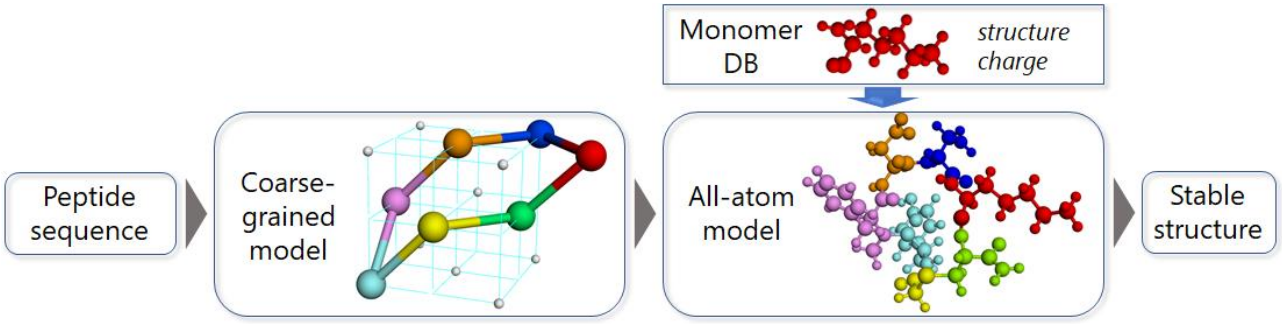


Figure 1: The procedure of our method.

[1] A. Godzik, A. Kolinski, and J. Skolnick. *J. Comput. Chem.*, 14, 10, 1194 (1993).
[2] https://www.fujitsu.com/global/services/business-services/digital-annealer/
[3] L. Wang, R. A. Friesner, and B. J. Berne, *J. Phys. Chem. B*, 115, 9431 (2011).
[4] H. Fujitani, A. Matsuura, S. Sakai, H. Sato, and Y. Tanida, *J. Chem. Theory Comput.* 5, 1155 (2009).
[5] A. Rodriguez, A. Laio, Science 344, 6191, 1492 (2014).

Molecular dynamics study of the interaction between a GST dimer and a novel peptidic covalent aptamer

Lisa Matsukura[1] liz@miyashita-lab.net
Naoyuki Miyashita[1,2] miya@waka.kindai.ac.jp
– **Masumi Taki**[2] taki@pc.uec.ac.jp
– **Shinichi Watanabe**[2] shin.watanabe@uec.ac.jp
– **Jay Yang**[3,4] jyang75@wisc.edu

[1] Department of Biological System Engineering, Graduate school of Biological Oriented Science and Technologies, KINDAI University, 930 Nishimitani, Kinokawa, Wakayama 649−6493, Japan
[2] Department of Engineering Science, Graduate School of Informatics and Engineering, University of Electro-Communications, 1-5-1, Chofugaoka, Chofu, Tokyo 182-8585, Japan
[3] Department of Anesthesiology, University of Wisconsin, School of Medicine and Public Health, Madison, Wisconsin 53706, USA
[4] Department of GI Surgery II, Hokkaido Univ Grad Sch Medicine, Sapporo, Hokkaido, Japan

Keywords: Molecular dynamics Simulation, Ligand Docking

Recently, peptidic covalent aptamer (PecA)[1] has been reported by Tabuchi et. al., as a novel peptide drug for Glutathione S-transferase (GST). It specifically binds to the dimer structure of GST. PecA consists of a peptide, which has seven amino acids (LESCAWY), and a compound, aryl-fluorosulfate (aryl- OSO2F; AFS), which connects to the cysteine in the peptide. A previous study suggested that PecA binds around the glutathione binding site of GST dimer, located in the hydrophobic regions. In this system, TYR111 in GST dimer and the fluorine of AFS in PecA are covalently bound. However, the detailed binding pose has not been clarified experimentally. In this study, we investigated the binding pose of PecA on GST and the drug-bound state using molecular docking and molecular dynamics simulations.

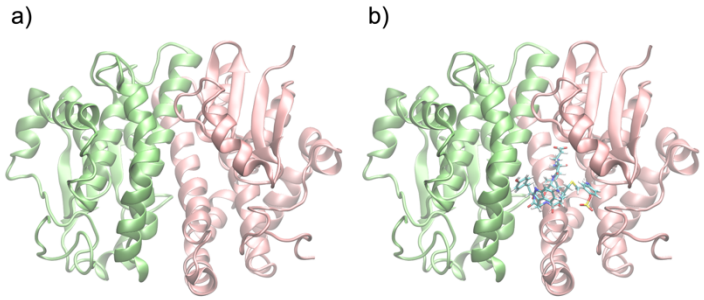


Fig. 1 The structure of a) GST dimer, b) the complex structure of GST dimer and PecA.

[1] Tabuchi, S.; Watanabe, T.; Katsuki, R.; Ito, Y.; Taki, M., Direct screening of a target-specific covalent binder: stringent regulation of warhead reactivity in a matchmaking environment, *Chemical Communications*, **2021**, *57*, 5378-5381.

QM/MM simulations of artificial ion channel in membrane-water system

Mayuko Nakagawa[1] w215421d@yokohama-cu.ac.jp
Toru Ekimoto[1] ekimoto@yokohama-cu.ac.jp
Tsutomu Yamane[2] tsutomu.yamane@riken.jp
Kohei Sato[3] koheisato@bio.titech.ac.jp
Kazushi Kinbara[3] kinbara.k.aa@m.titech.ac.jp
Mitsunori Ikeguchi[1,2] ike@yokohama-cu.ac.jp

[1] Yokohama City Univ., 1-7-29 Suehiro-chou, Tsurumi-ku, Yokohama-shi, Kanagawa, 230-0045, Japan
[2] RIKEN, 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama-shi, Kanagawa, 230-0045, Japan
[3] Tokyo Inst. of Tech., 4259 Nagatsuda-chou, Midori-ku, Tokyo, 226-8503, Japan

Keywords: Artificial ion channel, Ion selectivity, QM/MM

The development of artificial molecular machines mimicking biomolecular functions is progressing. Recently, inspired by potassium channels found in nature, a cyclic artificial ion channel **C$_{FF}$** has been developed. [1] **C$_{FF}$** consists of two rigid parts including perfluorinated aromatic units and linkers connecting them. According to experimental results, **C$_{FF}$** transports cations across the membrane and has potassium selectivity.

To assess channel structures of **C$_{FF}$** in the membrane-water system at atomic resolution, we performed quantum mechanical/molecular mechanical (QM/MM) simulation implemented in GENESIS. [2-5] Specific interactions between the potassium ion in the pore center and the perfluorinated aromatic units were observed: The potassium ion was surrounded by four fluorine-substituted aromatic units and the ion interacted with eight fluorine atoms, rather than the planes of the aromatic rings. In addition, different interaction modes were observed in the potassium and sodium ions at the channel entrance: Potassium ion was not completely hydrated and maintained interactions with the fluorine atoms. In contrast, sodium ion completely dissociated from **C$_{FF}$** and was hydrated with water molecules, except for the pore center. Comparing these different interaction modes, the mechanism of the ion selectivity could be understood by a balance of interaction strengths between the ion and fluorine atoms and between the ion and water molecules.

[1] Sato, K.; *et al*., Supramolecular Mechanosensitive Potassium Channel Formed by Fluorinated Amphiphilic Cyclophane, *J. Am. Chem. Soc.,* **2022**, *144*(26), 11802–11809; Correction: *J. Am. Chem. Soc*., **2022**, *144*(30), 13983-13984.
[2] Yagi, K.; *et al*., Anharmonic Vibrational Analysis of Biomolecules and Solvated Molecules Using Hybrid QM/MM Computations. *J. Chem. Theory Comput.*, **2019**, *15*, 1924–1938.
[3] Yagi, K.; *et al*., Exploring the Minimum-Energy Pathways and Free-Energy Profiles of Enzymatic Reactions with QM/MM Calculations. *J. Phys. Chem. B*, **2021**, *125*, 4701–4713.
[4] Jung, J.; *et al*., GENESIS: a Hybrid-Parallel and Multi-Scale Molecular Dynamics Simulator with Enhanced Sampling Algorithms for Biomolecular and Cellular Simulations. *WIREs Comput. Mol. Sci.*, **2015**, *5*, 310–323.
[5] Kobayashi, C.; *et al*., GENESIS 1.1: A Hybrid-Parallel Molecular Dynamics Simulator with Enhanced Sampling Algorithms on Multiple Computational Platforms. *J. Comput. Chem.*, **2017**, *38*, 2193–2206.

Pursuit of elementary reactions by the difference of H/D using the Nuclear-Electronic Orbital method

Kohei Motoki[1] a17.3rsh@g.chuo-u.ac.jp
Hirotoshi Mori [1] qc-forest.19d@g.chuo-u.ac.jp

[1] Department of Applied Chemistry, Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, 112-8551, Japan

Keywords: Nuclear-Electronic orbital, Deuterium kinetic isotope effect, Elementary reaction

Recent progress in molecular science has shown that the molecular properties of deuterium-containing systems would be slightly different from the normal systems. The fact means that a way to control molecular electronic state without changing its structure has been opened by introducing Deuterium (D-) instead of substituent groups, often applied in conventional drug discovery. [1,2]

A quantum chemical method considering the nuclear quantum effects, i.e., beyond Born-Oppenheimer (BO) treatment, is mandatory to describe the electronic difference between light hydrogen and deuterium. Nuclear-Electronic Orbital (NEO) method [3] is one of the beyond BO methods. Recently, we have initiated a set of deuterium science research related to pharmacy [4].

The most important in the research field is the Kinetic Isotope Effect (KIE). Previously, the KIE has been explained only by the mass effect in quantum mechanics. However, many experimental results cannot be explained without hypothesizing electronic states' change caused by D-substitution. In this study, we have theoretically examined several organic elementary reactions of which KIE have been precisely observed [5]. In the poster session, we will discuss the importance of electronic state tuning by D- substitution, including a well-known pharmaceutical example case, i.e., an optical isomerization of deuterium-substituted thalidomide [6].

[1] C. Schmidt, First deuterated drug approved, *Nature Biochemistry,* **2017**, *35*, 493-494
[2] T. Pirali, *et. al*., Applications of Deuterium in Medicinal Chemistry, *J. Med. Chem.*, **2019**, *62*, 5276−5297
[3] F. Pavosĕvic, *et. al*., Multicomponent Quantum Chemistry: Integrating Electronic and Nuclear Quantum Effects via the Nuclear−Electronic Orbital Method, *Chem. Rev.*, **2020**, *120*, 4222−4253
[4] K. Motoki, H. Mori, Nuclear-Electronic Orbital study of quantum property changes of organic molecules by H/D difference, The 16th Molecular Science Symposium Abstract, **2022**, 2P098
[5] S. Gronert, *et. al.,* Deuterium Isotope Effects in Gas-Phase Reactions of Alkyl Halides: Distinguishing E2 and S_N2 Pathways, *J. Am. Chem. Soc.*, **1991**, *113*, 4009-4010
[6] C. Tian. *et*. *al*., *Chem*. *Eur. J*., Enantiomerization Mechanism of Thalidomide and the Role of Water and Hydroxide Ions, **2012**, *18*, 14305 – 14313

P01-07

Prediction of Structural Change of siRNA by 2′-formamide, a Newly-synthesized Chemical Modification, via Density Functional Theory

Seongjin An[1]
7826759202@edu.k.u-tokyo.ac.jp

Yoshiaki Kobayashi[2]
yoshiaki-kobayashi@g.ecc.u-tokyo.ac.jp

Kohei Nomura[3]
nomura.kohei@f.mbox.nagoya-u.ac.jp

Yasuaki Kimura[3]
kimura.yasuaki.r9@f.mail.nagoya-u.ac.jp

Hiroshi Abe[3]
h-abe@chem.nagoya-u.ac.jp

Dai Akase[4]
dai-akase@hiroshima-u.ac.jp

Misako Aida[4]
maida@hiroshima-u.ac.jp

Kumiko Ui-Tei[1,2]
ktei@bs.s.u-tokyo.ac.jp

[1] The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-Shi, Chiba, 227-8561, Japan
[2] The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
[3] Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan
[4] Hiroshima University, 1-3-2 Kagamiyama, Higashi-Hiroshima City, Hiroshima, 739-8511, Japan

Keywords: siRNA, 2′-formamide, Density Functional Theory, chemical modification, off-target effect

RNA interference (RNAi) is an innate mechanism of post-transcriptional gene silencing mediated by small interfering RNA (siRNA) that regulates mRNA expression with complete sequence complementarity. On the other hand, siRNA often exhibits off-target effect that down-regulates unintended mRNAs with partial sequence complementarities with the seed region (nucleotides 2 to 8 from the 5′ end) of the guide strand. Our previous report revealed that the degree of off-target effect is positively correlated with the thermodynamic stability in base-pairing between the seed region of guide strand and unintended mRNAs [1].

In this study, we introduced newly synthesized 2′-formamide chemical modification into sugars of the guide strand nucleotides, capable of reducing base-pairing stabilities, into seed nucleotides and investigated their effects on off-target effect. 2′-formamide modification reduced off-target effects at all of the positions in the seed region. However, the degrees of the off-target effects differed depending on their positions. The 2′-formamides introduced into the former positions (nucleotides 2, 3, 4, 5) inhibited off-target effects stronger than those in the latter positions (nucleotides 6, 7, 8). Such position dependency may due to the steric hindrance induced by the sugar modification [2]. To investigate this, we used density functional theory calculation and predicted the structural changes. As the result, structural changes were observed only in the former positions, suggesting that this structural distortion may have had a great effect on the reduction of off-target effect, and the 2′-formamides in the latter positions are considered to inhibit off-target effects by reducing thermodynamic stabilities.

[1] Ui-Tei, K.; Naito, Y.; Nishi, K.; Juni, A.; Saigo, K. Thermodynamic Stability and Watson–Crick Base Pairing in the Seed Duplex Are Major Determinants of the Efficiency of the SiRNA-Based Off-Target Effect. Nucleic Acids Research **2008**, 36 (22), 7100–7109.

[2] Iribe, H.; Miyamoto, K.; Takahashi, T.; Kobayashi, Y.; Leo, J.; Aida, M.; Ui-Tei, K. Chemical Modification of the SiRNA Seed Region Suppresses Off-Target Effects by Steric Hindrance to Base-Pairing with Targets. ACS Omega **2017**, 2 (5), 2055–2064.

P01-08

Metadynamics simulation of ferritin encapsulation drugs

Kazutoshi Takahashi[1]
kazutoshi.takahashi.pu7@asv.ajinomoto.com

Ippei Inoue[1]

[1] Research Institute for Bioscience Products & Fine Chemicals, Ajinomoto Co., Inc. 1-1 Suzuki-cho, Kawasaki-ku, Kawasaki 210-8681, Japan

Keywords: Metadynamics simulation, Ferritin, Protein fluctuation

Metadynamics is one of a computational simulation method that can be used for molecular dynamics simulation to accelerate rare events. Molecular dynamics simulations are powerful methods, but it is difficult to perform enough calculations to obtain all the conformations in a typical molecular dynamics simulation. If the relevant conformations are separated by a high free energy barrier, this free energy is rarely exceeded. Therefore, in metadynamics, collective variables (CVs) are set up and a bias potential is added to the CVs to accelerate state transitions in the system and facilitate sampling. In this study, metadynamics was used to simulate the drug internalization process of ferritin protein.

Protein-based nanoparticles are one of the carriers in promising drug delivery systems because of their excellent biocompatibility, solubility, etc. In fact, the iron storage protein human ferritin is an excellent candidate for a carrier. This protein is a robust cage-like complex composed of 24 identical subunits with an external size of 12 nm. Although the conventional method of encapsulation of drug molecules into ferritin-protein complexes has used the degradation and resealing method, by incubating ferritin and drug under appropriate conditions, the drug molecules are allowed to pass through the pores in ferritin to obtain a complex between ferritin and the target drug. The process of passing through the pore was analyzed by metadynamics simulation.

[1] Laio, A.; Parrinello, M., Escaping free-energy minima. Proc Natl Acad Sci U S A 2002, 99 (20), 12562-6.

[2] Inoue, I.; Chiba, M.; Ito, K.; Okamatsu, Y.; Suga, Y.; Kitahara, Y.; Nakahara, Y.; Endo, Y.; Takahashi, K.; Tagami, U.; Okamoto, N., One-step construction of ferritin encapsulation drugs for cancer chemotherapy. Nanoscale 2021, 13 (3), 1875-1883.

P01-09

Simulation study of the interaction between lipids and the complex structure of γ-secretase and APP or Notch.

Chika Minami[1]
minami.373@miyashita-lab.net

Lisa Matsukura[1]
liz@miyashita-lab.net

Naoyuki Miyashita[1*]
miya@waka.kindai.ac.jp

[1] Graduate School of Biology-Oriented Science and Technology, KINDAI University, 930 Nishimitani, Kinokawa, Wakayama 649-6493, Japan

Keywords: Coarse-grained simulation, Molecular dynamics simulation, APP, Notch

γ-secretase has been known for cleaving various membrane proteins. It consists of four membrane proteins, namely Nicastrine (NCT), Presenilin (PS), APH-1, and PEN-2. Amyloid precursor protein (APP), which is also membrane protein, is cleaved by β-secretase and γ-secretase in the early stage of Alzheimer's disease proceeds. The relation between APP and secretases is linked to the production of senile plaques. The γ-cleavage site in the APP has two possible sites, Aβ40 and Aβ42. However, the mechanism of these selections has not been unknown yet. Notch, the signal transduction mediating membrane protein, is also known as another one of the targets of the cleavage of γ-secretase. It does not know the mechanism of the binding between γ-secretase and Notch yet. The APP and the Notch usually bind to γ-secretase in the raft environment.

We investigate how lipids affect the docking process between γ-secretase and APP or Notch. We performed seven Coarse-grained molecular dynamics simulations, namely APP, Notch, γ-secretase, γ-secretase with APP or Notch, γ-secretase and APP or Notch complex with the membrane. The membrane contains three lipids, glycerophospholipids, sphingolipids, and cholesterols. The ratio of them is 5:3:2. The distributions of lipids will be shown, and we will discuss the interaction between the lipids and those membrane proteins.

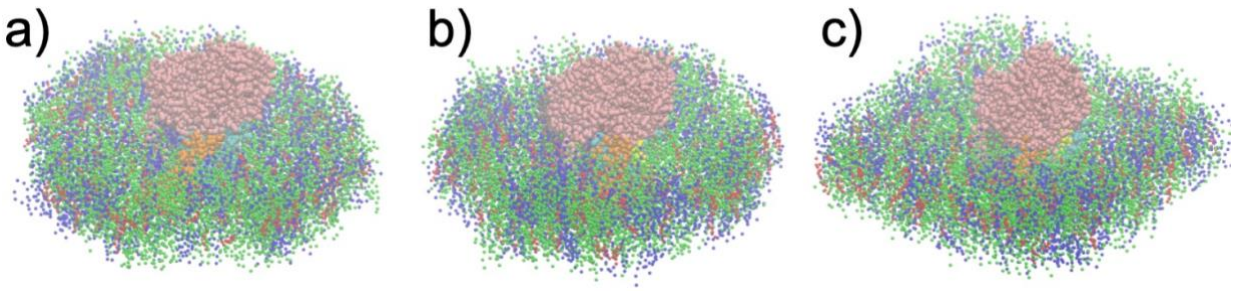


Fig 1. Coarse-grained model structures of a) γ-secretase with the membrane, b) γ-secretase and APP with the membrane and c) γ-secretase and APP complex structure with the membrane. The green, blue, and red indicate the glycerophospholipids, sphingolipids, and cholesterol, and the pink, cyan, gray, orange, and yellow show Nicastrine(NCT), Presenilin(PS), APH-1, PEN-2, and APP, respectively.

[1] Zhou, R *et.al, Science* **363** (2019)
[2]Yang, G *et. al, Nature,* **565**, 192–197 (2019)

Conformational behavior and dynamics of G7A mutant IgG-aptamer

Seiichiro Ishii[1], **Ryoji Yamazaki**[1], **Ozora Kudo**[1], **Takeshi Ishikawa**[2], **Taiichi Sakamoto**[3],

Kenji Yamagishi[1]
yamagishi.kenji@nihon-u.ac.jp

[1] College of Engineering, Nihon University, 1 Nakagawara, Tokusada, Tamura, Koriyama, Fukushima 963-8642, Japan
[2] Faculty of Engineering, Kagoshima University, 1-21-24 Koorimoto, Kagoshima, Kagoshima, 890-8580, Japan
[3] Faculty of Advanced Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan

Keywords: RNA aptamer, Nucleic Acid, Dynamics, Interaction energy, MD, FMO

RNA aptamers are short, single-stranded oligonucleotides that bind to specific target molecules such as proteins, nucleic acids, and small molecules. The single strand of RNA aptamers can fold unique tertiary structures that enable the RNA aptamer to interact with high affinity and specificity for its target molecule. An optimized 24-nucleotide aptamer was designed, which was shown to bind with high affinity to the Fc domain of human Immunoglobulin G (IgG) [1]. RNA mutagenesis experiments were performed, to investigate the structural and functional contribution of each nucleotide [1]. Destruction of the interaction between fU6 and G7 by mutation, such as fU6-to-U/mU/u and G7-to-A/fU, caused a loss of affinity. Although these data suggest the functional importance of the interaction between G7 and U6, the structural basis of the binding mechanism of the RNA aptamer with IgG is poorly understood.

In this study, in order to elucidate the conformational behavior and dynamics of G7A mutant IgG-aptamer, we performed the molecular dynamics simulation and fragment molecular orbital calculation.

Reference

[1] Y. Nomura, et al., *Nucleic Acids Research*, **38**, 7822 (2010).
[2] H. Yoshida, T. Ishikawa, T. Sakamoto, K. Yamagishi, et al., *Chem. Phys. Letts.*, **738**, 136854 (2020).

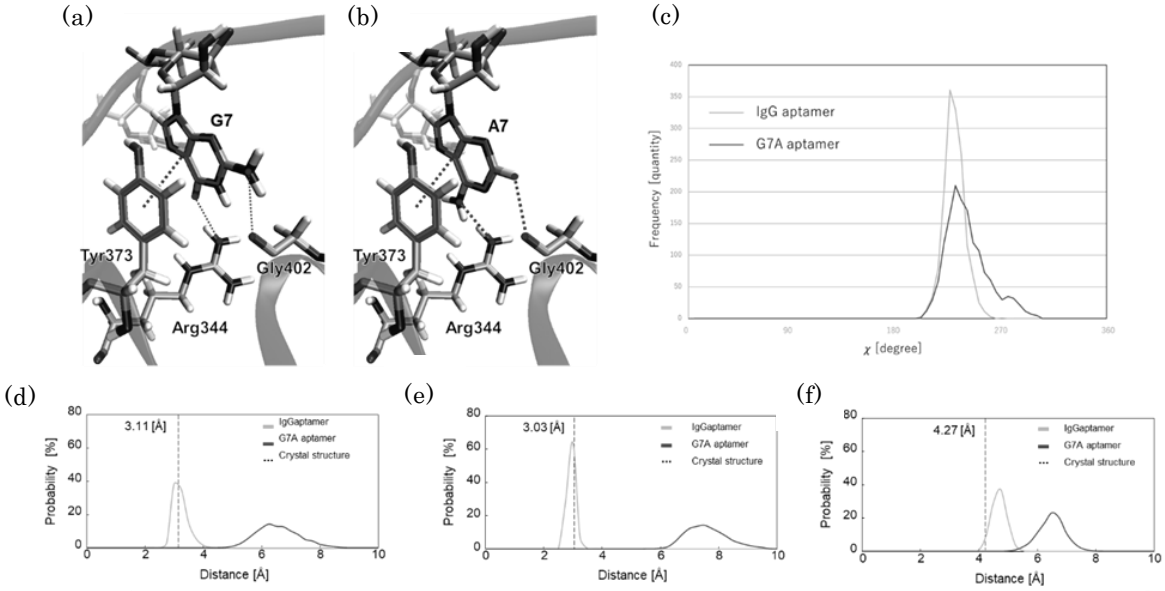


Figure. Structure of (a) original IgG aptamer (G7), (b) mutant IgG aptamer (G7A). Distribution of (c) kai angle, (d) G7/A7-Gly402 distance, (e) G7/A7-Arg344 distance, (f) G7/A7-Tyr373 distance, obtained from MD simulation.

P02-01

Energy decomposition analysis for cyclodextrins-furosemide complexes

Masao Fujisawa[1] fujisawa@waka.kindai.ac.jp
Tomonori Ohata[2] ohata@fukuoka-u.ac.jp
Hiroyuki Tsutsumi [2] h-tsutsumi@fukuoka-u.ac.jp
Hirohito Ikeda[2] ikeda@fukuoka-u.ac.jp

[1] Department of Biotechnological Science, Kindai University, 930 Nishimitani, Kinokawa-City, Wakayama, 649-6493, Japan
[2] Faculty of Pharmaceutical Science, Fukuoka University, 8-19-1 Nanakuma Fukuoka 814-0180, Japan

Keywords: Energy decomposition, DFT, Molecular dynamics, Inclusion complex

Cyclodextrin (CyD) can accommodate various kinds of guest molecules in its internal cavity, recognizing the difference in structure of guest molecules in aqueous solution. Improvement in stability and solubility of medicines can be expected when using CyD to form pharmaceutical drug conjugates. We determined systematically thermodynamic functions for the molecular inclusion of simple molecules into α- and β-CyD cavities in dilute aqueous solutions using microcalorimetry[1-2]. In this study, we have used computer simulation methods e.g., energy decomposition analysis (EDA) and molecular dynamics (MD) to investigate the thermodynamic aspects of the inclusion of pharmaceutical drugs into β-CyD cavities, specifically detailing the interaction energy for the complexation of furosemide which is a loop diuretic with hydroxypropyl-β-CyD (Hpc-β-CyD) in aqueous solutions. Furosemide contains a primary aryl sulfonamide and a furan moiety. We examined two conformations for each CyD complex, including (1a) the sulfonamide of furosemide existing in the Hpc-β-CyD cavity and the furan moiety located on the outside of the secondary hydroxyl group (SULFO_Hpc-β-CyD), and (1b) the furan ring of furosemide existing in the Hpc-β-CyD cavity and the sulfonamide located on the outside of the secondary hydroxyl group (FURAN_Hpc-β-CyD). Molecular dynamics calculations were performed for these two conformations. The final structure after 20 ns of MD simulation was analyzed using an energy decomposition method by decomposition into electrostatic (ES), exchange-repulsion (EX), charge transfer (CT), and dispersion (DI) components. In the SULFO-Hpc-β-CyD system at b3lyp/6-31g level, it was realized that dispersion forces of this complex by -19.6 kJ mol^{-1}. Furosemide (SULFO) is stabilized largely on inclusion into Hpc-β-CD cavity. The FURAN_Hpc-β-CyD system and β-CyD systems are will be discussed on the day of the annual meeting.

[1] Kimura T., Yukiyama T., Fujie S., Fujisawa M., Aki H., Kamiyama T., Enthalpy and entropy changes on molecular inclusion of pentane derivertives into α-cyclodextrin cavities in aqueous solutions,, *Journal of Inclusion Phenomena and Macrocyclic Chemistry*, 70(3-4):269-278, 2011.
[2] Ikeda H., Fukushige Y., Matsubara T., Inenaga M., M. Kawahara M., Yukawa M., Fujisawa M., Yukawa E., Aki H., Improving water solubility of nateglinide by complexation of β-cyclodextrin, *Journal of Thermal Analysis and Calorimetry*, 123(3):1847-1850, 2016.

Insight into Allosteric ERK2 Inhibitors by using Metadynamics Simulations

Hajime Sugiyama[1]
sugiyama.hajime.ma@m-chemical.co.jp

Seisuke Hasegawa[2]
se22077x@st.omu.ac.jp

Mayu Yoshida[2]
yoshida19@b.s.osakafu-u.ac.jp

Takayoshi Kinoshita[2]
kinotk@omu.ac.jp

[1] Mitsubishi Chemical Corporation, 1000 Kamoshida-cho, Aoba-ku, Yokohama, Kanagawa 227-8502, Japan
[2] Osaka Metropolitan Univ., 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan

Keywords: ERK2, Allosteric site, Metadynamics

ERK2 (Extracellular signal-regulated kinase 2) is a member of the mitogen-activated protein kinase family. STAT3 (signal transducer and activator of transcription 3), which plays an essential role in normal glucose homeostasis which plays an essential role in normal glucose homeostasis, and regulates cell proliferation, differentiation, and various other cellular responses. Based on previous studies on the ERK2/STAT3 pathway, the authors focused on ERK2 as a candidate target molecule for diabetes treatment and identified small molecule compounds with inhibitory activity by in silico screening [1].

In order to establish the basis for drug design, we investigated the static interaction between ERK2 and its compounds in the bound state from the crystal structure, aiming at structural modification of the small molecule compounds. On the other hand, the association and dissociation processes are also important in drug design. In order to efficiently generate these events in a time scale that allows molecular dynamics simulations, we attempted to apply a combined technique of Metadynamics [2] and Hamiltonian Replica Exchange [3].

As a result, we confirmed the occurrence of these events during the calculation and succeeded in calculating the free energy surfaces. The detailed results of the analysis will be presented on the same day. These results would greatly enhance to develop highly potent and selective ERK2 inhibitors.

[1] T. Kinoshita, et al., Chem. Biol. Drug Des. 78: 471- 476 (2011).
[2] A. Barducci, et al., Wiley Interdiscip. Rev.: Comput. Mol. Sci. 1: 826−843 (2011).
[3] V. Oleinikovas et al., J. Am. Chem. Soc. 138: 14257−14263 (2016).

Interaction analyses of inhibitors against threonyl-tRNA synthetase by fragment molecular orbital calculations

Yoshiharu Mori ymori@landscape.kobe-u.ac.jp
Shigenori Tanaka tanaka2@kobe-u.ac.jp

Graduate School of System Informatics, Kobe University, 1-1, Rokkodai, Nada, Kobe, Hyogo 657-8501, Japan

Keywords: Electronic structure calculation, Protein-ligand interaction, Translation inhibitor

Threonyl-tRNA synthetase (ThrRS) is an enzyme that covalently binds threonine to tRNA and is one of a group of enzymes required for translation processes. They are also required for protein synthesis and are therefore the target of inhibitors. Borrelidin is a well-known inhibitor, and it is understood from a structural analysis that this compound occupies several parts of the ligand binding region necessary for enzyme function. However, because borrelidin can also bind to human threonyl-tRNA synthetase, alternative compounds that bind to bacterial ThrRS but not to human ThrRS are desirable. Recently, such inhibitors have been discovered and synthesized. In this study, we used the fragment molecular orbital (FMO) method to analyze the interactions between such compounds and ThrRS to find similarities and differences in the binding mode between species.

The structures of human ThrRS bound with borrelidin and *E. coli* ThrRS bound with borrelidin were selected for the interaction analyses. Structural optimization using the Amber force field was performed on these structures to obtain structures for the FMO calculations. The FMO calculations were performed at the MP2/6-31G* level of theory. Interaction analyses were performed using the pair interaction energy decomposition analysis (PIEDA).

Because borrelidin forms a salt bridge with ThrRS, the contribution of electrostatic interactions was the largest in both species. On the other hand, in human and *E. coli*, there were variations in the interactions due to differences in the amino acid sequence of the ligand binding site. For comparison, FMO calculations were also performed on a structure containing an alternative compound with low binding affinity to human ThrRS. The interaction analyses provide insights into the inhibition mechanism for ThrRS.

P02-04

Amino acid preference mapping on protein-protein interaction surface using mixed-solvent molecular dynamics

Genki Kudo[1]
s2120165@s.tsukuba.ac.jp

Keisuke Yanagisawa[2]
yanagisawa@c.titech.ac.jp

Ryunosuke Yoshino[3,4]
yoshino.r.aa@md.tsukuba.ac.jp

Takatsugu Hirokawa[3,4]
t-hirokawa@md.tsukuba.ac.jp

[1] Physics Department, Graduate School of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577, Japan
[2] Department of Computer Science, School of Computing, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro 152-8550, Tokyo, Japan
[3] Faculty of Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8575, Ibaraki, Japan
[4] Transborder Medical Research Center, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577, Japan

Keywords: Protein-Protein Interaction, Mixed-Solvent Molecular Dynamics, Hotspot Detection

Over the past decade, compounds with a molecular weight range of 700-2000, such as cyclic peptides, have been developed as protein-protein interaction (PPI) inhibitors [1]. Several reports identified hotspots on PPI surface using mixed-solvent molecular dynamics (MSMD) which involves MD in the presence of explicit water molecules mixed with probe molecules or functional group fragments [2]. However, accessibility and amino acid preference on PPI surface has not been evaluated because they employed probes which are substructures of traditional drug molecules.

In this study, we propose MSMD with amino acid probes to detect hotspots on PPI surfaces. In addition, amino acid probes can also provide their preferred surfaces (Figure 1). In our presentation, we assessed our method with several proteins, and compared amino acid preference obtained from MSMD and PPI interface residues obtained from co-crystal structures.

(a) Arg probes (b) Glu probes

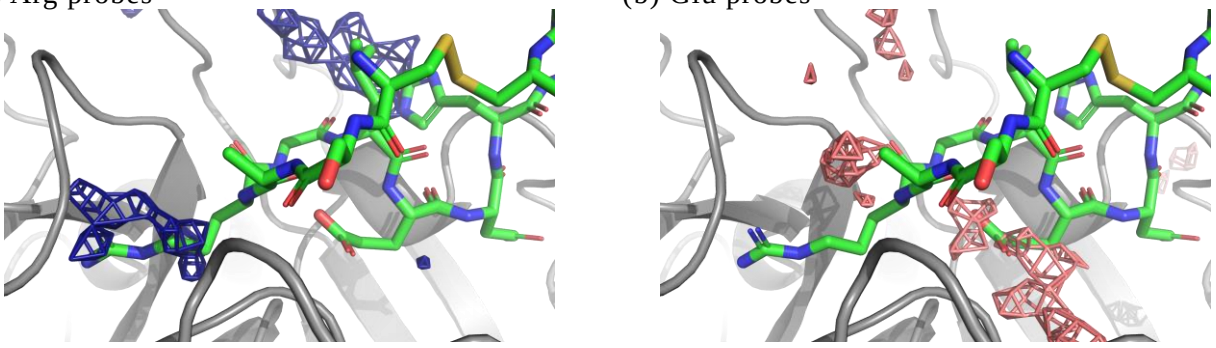


Figure 1. Example of preference maps on PPI surface obtained from MSMD

[1] Dougherty, P.G.; Qian, Z.; Pei, D. Macrocycles as protein-protein interaction inhibitors, *Biochem J.*, **2017**, 474(7), 1109-1125.
[2] Ghanakota, P.; *et al.*, Large-scale validation of mixed-solvent simulations to assess hotspots at protein–protein interaction interfaces, *J Chem Inf Model.*, **2018**, 58(4), 784-793.

P02-05

Practical visualization of interaction energies by FMO for drug design

Hirofumi Watanabe[1] **Yuji Tanakamaru**[1]
watanabe@withmetis.jp

[1] WithMetis Co.,Ltd., Wembley Building 7F 6-1-17 Isogamidori, Chuo-ku, Kobe, Hyogo, 671-0086, Japan

Keywords: Fragment molecular orbital method (FMO), Inter-fragment interaction energy (IFIE), Computer-aided drug design (CADD), Protein–ligand interaction

Inter-fragment energy (IFIE), evaluated by the fragment molecular orbital (FMO) method, is useful information for analyzing protein–ligand interactions based on quantum mechanics [1-4]. Because IFIEs between amino acid residue units or even smaller subunits can be further decomposed into components based on their physical origin (PIEDA) [5], the FMO method is a powerful tool for compound design in lead generation and lead optimization processes. To make this method more useful and more intuitive, we are developing a visualization tool for the IFIEs, Q-AIR FLAIR Viewer. The non-covalent molecular interactions are classified by composed elements and geometries, such as hydrogen bonds, salt bridges, and weaker interactions, such as CH–π, π–π interactions, where the contribution of dispersion interactions is important. This tool supports interpretation of the relationship between IFIEs and these molecular interactions based on structural information. We will discuss how we should visualize the IFIEs and PIEDAs in the research in the pharmaceutical company. Example analysis was done using data in a previous p38 study [6] and in the FMODD [7].

[1] Kitaura, K.; *et al.,* Fragment molecular orbital method: an approximate computational method for large molecules, *Chemical Physics Letters*, **1999**, *313,* 701-706.
[2] Tanaka, S.; *et al.,* Electron-correlated fragment-molecular-orbital calculations for biomolecular and nano systems, *Physical Chemistry Chemical Physics*, **2014**, *16*, 10310-103.
[3] Nakano, T.; *et al.,* Fragment molecular orbital method: use of approximate electrostatic potential, *Chemical Physics Letters*, **2002**, *351*, 475–480.
[4] Mochizuki, Y.; *et al.,* Large scale MP2 calculations with fragment molecular orbital scheme, *Chemical Physics Letters*, **2004**, *396*, 473-479**.**
[5] Fedorov, D.; *et al.,* Pair interaction energy decomposition analysis *Journal Computational Chemistry*, **2007**, *28*, 222–237.
[6] Sheng, Y.; Watanabe, H.; *et al.*, Towards good correlation between fragment molecular orbital interaction energies and experimental IC50 for ligand binding: A case study of p38 MAP kinase, *Computational and Structural Biotechnology Journal*, **2018**, *16,* 421-434.
[7] Takaya, D.; *et al.,* FMODB: The World's First Database of Quantum Mechanical Calculations for Biomacromolecules Based on the Fragment Molecular Orbital Method, *Journal Chemical Information Modeling*, **2021**, *61*, 777-794.

P03-01

Investigation of the applicability domain for prediction model using in-house data

Toshiaki Watanabe[1] **Kosuke Takeuchi**[1] **Atsushi Kasuya**[1] **Daisuke Sugiyama**[1]
watanabe.toshiaki.h5@daiichisankyo.co.jp

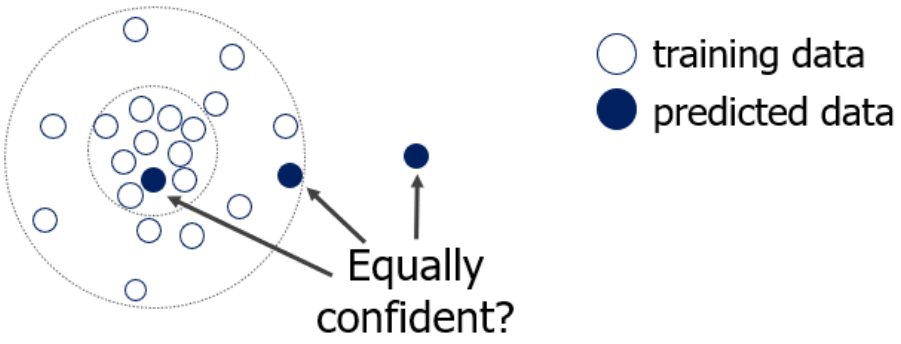
Ryo Kunimoto[1] **Kentaro Rikimaru**[2] **Shin Saitou**[2]

[1] R&D Division, Daiichi Sankyo Co., Ltd., 1-2-58 Hiromachi, Shinagawa-ku, Tokyo 140-8710, Japan
[2] ExaWizards Inc., Tokyo 105-0021, Japan

Keywords: machine learning, applicability domain

The drug discovery process would be expected to accelerate by using machine learning methods. On the other hand, if not used appropriately case, there is a risk of misleading rather than improving efficiency. We thought that the understanding of AD (Applicability Domain)[1,2,3] is an essential element to properly interpreting the predicted results. Therefore, we conducted an analysis concerning AD using in-house data.

Investigation of AD using in-house metabolic stability data, we found that the distance between training and test affected the accuracy. By designing a validation set based on this, we succeeded in improving the accuracy. In a case study of in-house project data, it became possible to determine the applicability of the global model using the AD concept. Furthermore, we found that switching to a local model is effective once data has been accumulated as the project progresses.



[1] Baskin, I. Kireeva, N. and Varnek, A. The One-Class Classification Approach to Data Description and to Models Applicability Domain. *Mol. Inf.* **2010**, *29*, 581-587.
[2] Sushko, I. et al. Applicability Domains for Classification Problems: Benchmarking of Distance to Models for Ames Mutagenicity Set. *J. Chem. Inf. Model*. **2010**, *50, 12*, 2094–2111.
[3] Kaneko, H. and Funatsu K. Applicability Domain Based on Ensemble Learning in Classification and Regression Analyses *J. Chem. Inf. Model*. **2014**, *54, 9*, 2469–2482.

P03-02

Deep Learning for Cellular Morphological Change Detection with High Throughput Images

Yuko Nirei[1]
nirei.yuuko@ma.mt-pharma.co.jp

Kentaro Matsumoto[2]
matsumoto.kentarou@ma.mt-pharma.co.jp

Ken Ohno[2]
oono.ken@mr.mt-pharma.co.jp

Ryuta Saito[1]
saitou.ryuuta@mc.mt-pharma.co.jp

[1] Discovery Technology Laboratories, Sohyaku. Innovative Research Division, Mitsubishi Tanabe Pharma Corporation, 1000, Kamoshida, Aoba, Yokohama, Kanagawa, 227-0033, Japan
[2] Discovery Technology Laboratories, Sohyaku. Innovative Research Division, Mitsubishi Tanabe Pharma Corporation, 2-26-1, Muraoka-Higashi, Fujisawa, Kanagawa, 251-8555, Japan

Keywords: image processing, high content screening, CellProfiler, deep learning, transfer learning

Cellular morphological changes observed in microscopic images of high-content screening (HCS) are detected by various image processing tools including software solutions designed for HCS. Software for HCS offers feature sets such as area, shape, and texture of cellular structures multi-stained with fluorescence and measured in different color channels [1]. These features are designed for detecting various morphological changes observed in cells, which are comprehensible, however, can lead to limitation in detecting those unexpected in the design. In contrast, deep learning algorithms can represent incomprehensible, but complex and unnoticeable features observed in morphological changes. Especially, features of deep neural networks pre-trained with various image objects are applicable to different tasks and can achieve high accuracy and speedy processing. Here, we clarify advantages of applying deep learning to HCS, and compare between an image processing software for HCS, CellProfiler (version 1) [2], and one of the deep neural networks, AlexNet, in hit-judgement of morphological changes induced by various chemical compounds and the processing speed. We prepared images of U2OS cells processed with 1680 compounds or DMSO. The cell structures were multi-stained with fluorescent, which were observed in four channels for each cell. Morphological changes induced by each compound were measured with 1085 features in CellProfiler. Significant morphological changes were identified by five-fold SD from DMSO, resulting in 228 hit and 1452 non-hit compounds. For the same image dataset, 4097 features of pre-trained AlexNet model were extracted from individual cellular images. Based on these AlexNet features, cells processed with each compound and DMSO were binary classified with k-NN, where accuracy of a classification is expected to increase when a compound induced larger morphological changes from non-treated cells. The average accuracy of classifications with AlexNet based method for the 228 CellProfiler hits was 67.7 ± 7.8%, and for the 1452 non-hit compounds was 58.2 ± 2.5%. Among the non-hit compounds, 290 compounds showed more than 60% accuracy with AlexNet based classifications between each compound and DMSO, indicating that AlexNet has capability to detect morphological changes that are not detected with predefined features of CellProfiler. In addition, processing speed on our local computational environment was about three months with CellProfiler, and about two and a half days (33 seconds for a compound well) with the AlexNet based method. These results show that deep learning can offer higher capability in detecting morphological changes and faster processing compared to software solutions designed for HCS.

[1] Lin, S.; Schorpp, K; Rothenaigner, I; Hadian, K. Image-based high content screening in drug discovery, *Drug Discovery Today*, **2020**, *25*, 1348-1361.
[2] https://cellprofiler.org/

P03-03

Investigation of the relationship between the performance of Encoder-Decoder model and the characteristic of molecular representation obtained

Shumpei Nemoto[1]
nemo88@g.ecc.u-tokyo.ac.jp

Tadahaya Mizuno[1, †]
tadahaya@mol.f.u-tokyo.ac.jp

Hiroyuki Kusuhara[1, †]
kusuhara@mol.f.u-tokyo.ac.jp

[1] Laboratory of Molecular Pharmacokinetics, Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo, Japan
[†] Author to whom correspondence should be addressed

Keywords: Encoder-Decoder model, chemoinformatics, molecular representation, descriptor

In chemoinformatics tasks such as virtual screening and toxicity prediction, it is essential to describe chemicals as numerical information (molecular representation) beforehand. Since the accuracy of downstream chemoinformatics tasks depends on the nature of the molecular representation various methods have been developed. After the first proposal by Gomez-Bómbarelli et al. in 2016, methods using Encoder-Decoder (ED) models, inspired by natural language processing and takes SMILES as input, have attached much attention for generating molecular representation[1]. Molecular representation derived from the ED model-based methods has two characteristics: continuous and structure-returnable natures. Many methods have been developed, however, most of them evaluate the performance of their methods based on the accuracy of downstream tasks using molecular representations as inputs. This approach can evaluate the ability of molecular representations to describe chemical structures, but it cannot evaluate the ability to return to structure. In addition, most of the approaches are comparisons between models, and the relationship between model accuracy itself and the performance of the molecular representation is unclear.

The purpose of this study is to clarify the relationship between the performance of ED model as an NLP model (translation accuracy) and a generator of molecular representation. In particular, since the structure representation ability of the molecular representation is related to the encoder performance and the structure restoration ability is related to the decoder performance of ED model, we tried to evaluate them separately.

Using ZINC dataset as input, we prepared a set of ED models with various translation accuracies by controlling the progress of training. To evaluate the structure representation ability, predictive models for ToxCast HTS assay data were built using molecular representation derived from the ED model set as input. Interestingly, accuracy for these downstream tasks was similar, independent of accuracy across model sets, except for the model with the poorest translation accuracy, suggesting that the structure representation ability of molecular representation is acquired early in the training. To evaluate the structure restoration ability, we firstly collected chemicals successfully translated (output structure is the same as input) in all accuracy models. Random numbers with specific correlation with molecular representations of these chemicals were generated for each model and the characteristics of the structures restored from the random numbers were evaluated with some metrics of chemical sets such as diversity. The results showed that the higher the accuracy of the ED model, the more it converged to the original structure, suggesting that enough training is necessary for unique structure restoration.

This study suggests that, in structure numerization using ED model, a high-performance model is unnecessary when the objective is only to capture structural features, while a high-performance model is necessary when the objective includes structure restoration.

[1] Rafael Gomez- Bómbarelli; *et al*, Automatic Chemical Design Using a Data-Driven Continuous Repreesntation of Molecules, *ACS Central Science*, **2018**, 4, 268-276

Integrated browsing of chemical reactions registered in Electronic Lab Notebook and external data

Yu Endo[1]
yu.endo@po.nippon-shinyaku.co.jp

Akihiro Oda[1]
a.oda@po.nippon-shinyaku.co.jp

[1] Nippon Shinyaku Co., Ltd., 14, Nishinosho-monguchi-cho, Kisshoin, Minami-ku, Kyoto, Kyoto, 601-8550, Japan

Keywords: Electronic Lab Notebook, ELN, data integration, Reaxys

Nippon Shinyaku has integrated intrinsic reaction information within the Elsevier Reaxys product. The incorporation and discoverability of proprietary and public information significantly improves researcher productivity. With this system, researchers can search Reaxys for internal Electronic Lab Notebook (ELN) data and external data published in journals and patents at the same time, unifying and organizing search results that are directly relevant to the researcher's workflow.

To achieve this integrated workflow, we developed a BIOVIA Pipeline Pilot protocol to automate a series of steps: extraction of reaction data from ELN, conversion to UDM[1] data format, and uploading of the converted UDM files to Amazon S3 buckets.

Key points of the ELN integration and data modeling will be discussed.

[1] *Pistoia Alliance UDM Project page* https://www.pistoiaalliance.org/projects/current-projects/unified-data-model/ (accessed 2022-08-08)

P03-05

Construction D4 environment for Patent informatics

Takayuki Serizawa[1]
serizawa.takayuki.a3@daiichisankyo.co.jp

Ryo Kunimoto[1]
kunimoto.ryo.si@daiichisankyo.co.jp

1 R&D Division, Daiichi Sankyo Co., Ltd., 1-2-58 Hiromachi, Shinagawa-ku, Tokyo 140-8710, Japan

Keywords: Database Development, Open Data Collection

A patent is one of the useful data sources in drug discovery projects. It has not only structural data but also biological data.
Chemoinformatician often spends lots of time to extract data from patents because the information is not provided as machine readable format, like PDF from OCR.
To overcome the issue, we build in-house patent database with public data source, SureChEMBL[1].
Here we would like to present our internal effort to construct patent analysis infrastructure on AWS cloud.
(D4: Data Driven Drug Discovery)

[1] George Papadatos et al., *Nucleic Acids Research*, Volume 44, Issue D1, 4 January 2016, Pages D1220–D1228
[2] https://github.com/chembl/surechembl-data-client

Estimation of amino-acid interaction potentials with structure similarity calculated by Digital Annealer

Chieko Terashima[1]
c.terashima@jp.fujitsu.com

Yusuke Yoshida[1]
yoshida_yusuke@fujitsu.com

Tetsuya Onuki[1]
onuki.tetsuya@fujitsu.com

Hiroyuki Satou
shryk@fujitsu.com

Go Fujisawa
fujisawa.go@fujitsu.com

[1] Fujitsu Limited. 1-5 Omiya-cho, Saiwai-ku, Kawasaki-shi, Kanagawa, 212-0014, Japan

Keywords: amino-acid, interaction potential, machine learning, structure similarity

Cyclic peptide is a kind of the promising molecules for middle drug discovery, and their structure is important. For structure investigation, a coarse-grained model with amino-acid pair interactions was treated [1]. Amino-acid pair interactions developed by umbrella sampling [2], require a long time and enormous computational resources. To use the coarse-grained model, amino-acid pair interactions were estimated by machine learning with structure similarity calculated by Fujitsu Quantum-inspired Computing Digital Annealer [3, 4] (Figure 1). The estimated model from this descriptor shows higher accuracy than other structural descriptors.

For searching the estimation model, we tried two types of data segmentation methods. First, the data of amino-acid pair interactions were randomly divided. Second, amino-acid was divided into a learning data group or validation data group. As the result, it was easy to estimate the amino-acid pair interaction with the known amino-acid, but it was difficult to estimate the interaction with the unknown amino-acid. On the day, we'll show the solution way.

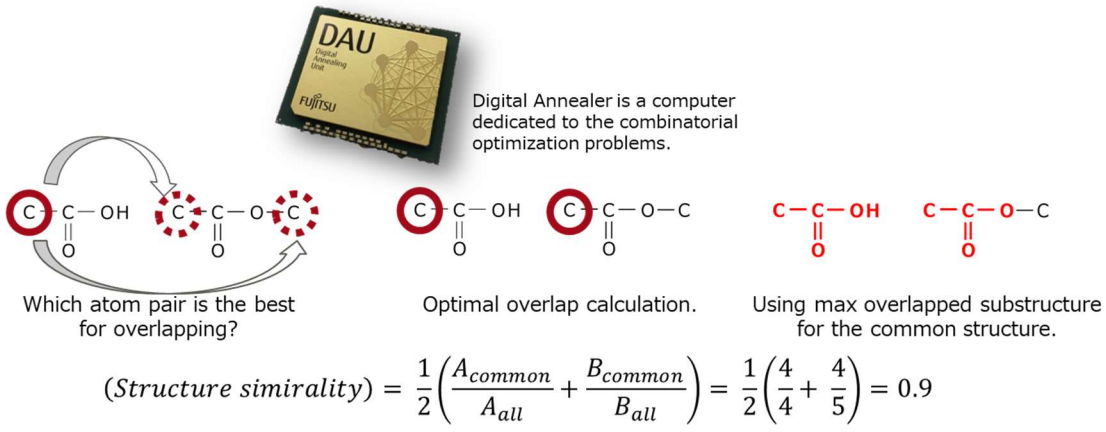


Figure1: Structure similarity calculated by Digital Annealer

[1] Droi T., Ron E PROTEINS: Structure, Function, and Genetics 41:40-46 (2000).
[2] GROMACS Tutorials http://www.mdtutorials.com/gmx/index.html.
[3] https://www.fujitsu.com/global/services/business-services/digital-annealer/
[4] M. Aramon, G. Rosenberg, E. Valiante, T. Miyazawa, H. Tamura, H. G. Katzgraber, Frontiers in Physics 2019, 7.

P03-07

Recent developments of FMODB: enhancement of the IFIE/PIEDA interface for analyzing the related FMO calculation data of bio-macromolecules

Kikuko Kamisaka[1] kikuko.kamisaka@riken.jp
Chiduru Watanabe[1] chiduru.watanabe@riken.jp
Daisuke Takaya[2] takaya-d@phs.osaka-u.ac.jp
Teruki Honma[1] honma.teruki@riken.jp

[1] RIKEN Center for Biosystems Dynamics Research, 1-7-22 Suehiro-Cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
[2] Graduate School of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565-0871, Japan

Keywords: FMO, Database, MD Snapshots, interaction energy diagram

The fragment molecular orbital (FMO) method, one of the quantum chemical calculations, can quantitatively and accurately estimate the inter-fragment interaction energy (IFIE) in ligand-protein, protein-protein, and nucleic acids systems. In addition, the pair interaction energy decomposition analysis (PIEDA) enables a more detailed interaction analysis. Therefore, FMO is expected to be helpful in drug discovery research, such as selecting lead compounds and compound design.

Our group has accumulated FMO data calculated by skilled researchers from the FMO drug design (FMODD) consortium and made it public as an FMO database (FMODB [1]). The database provides simple analysis functionality such as a graph of IFIE/PIEDA values via a web interface. Recently, in FMODB, there has been an increase in registration of FMO calculation data for a series of data such as snapshot time-series structure of molecular dynamics (MD) simulation and docking pose of a complex. However, the IFIE/PIEDA analysis functionality provided via the web interface in the database was only for single calculation data. We have developed new functionality to collectively analyze ligand-protein interactions for multiple structures registered in FMODB, such as a series of datasets of MD snapshots. In this study, we introduce the ability to collectively analyze the interactions between protein-ligand, using MD snapshots [2] for complexes of cyclin-dependent kinase-2 (CDK2) with six inhibitors (PDB IDs: 4FKL, 4FKI, 4FKQ, 4FKR, 4FKS, and 4FKW). We also derive the correlation between the dynamically averaged IFIE/PIEDA and the activity value considering structural fluctuations.

In addition, we developed a function to automatically generate an interaction energy diagram that considers the bond network and its strength between fragments using IFIE/PIEDA. The diagram will be helpful for analyzing and understanding intricate bound networks and their interaction energy. Thus, we introduce the function and usage using the complex between Remdesivir, RNA, and RNA-dependent RNA polymerases [3].

Acknowledgment

The authors thank Dr. Kazumi Tsuda, Mr. Daisuke Murayama, Dr. Shu Koyama, and Dr. Tomoharu Isobe of Science & Technology Systems, Inc., Japan, for technical support. This research was done in activities of the FMO Drug Design Consortium, https://fmodd.jp/top-en/.

[1] Takaya, D.; *et al., J. Chem. Inf. Model.* **2021**, 61, 2, 777–794.
[2] Takaba, K.; *et al., J. Comput. Chem.*, **2022**, 43(20), 1362-1371.
[3] Kato, K.; *et al.*, *J. Mol. Graph. Model.,* **2020**, 100, 107695.

Reservoir Computing for Efficient Prediction of Optical Response of Digital Metamaterial

Tomoya Meguro[1]
phf67gbv@s.okayama-u.ac.jp

Masaaki Misawa[1]
misawa@okayama-u.ac.jp

Kei Tokita[2]
tokita@i.nagoya-u.ac.jp

Kenji Tsuruta[1]
tsuruta@okayama-u.ac.jp

[1] Graduate School of Natural Science and Technology, Okayama University,
3-1-1 Tsushimanaka, Kita-ku, Okayama-shi, Okayama 700-8530, Japan
[2] Graduate School of Informatics, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi 464-8601, Japan

Keywords: Machine Learning, Reservoir Computing, Digital Metamaterials

Application of artificial intelligence technology has been rapidly expanding in a variety of industries, yet it requires big data to acquire sufficient accuracy in the application. In this study, we focus on an artificial material called “digital metamaterials.” Metamaterials are artificial materials that exhibit exotic responses to incident waves, such as light and sound, in ways not found in nature. By periodically arranging metallic structures smaller than the wavelength of light, they have unique optical properties depending on their shape. The principle of operation is that the metallic nanostructure acts as an antenna in which the optical response is triggered by resonance at a frequency corresponding to the length of the structure. As a specific example, it is expected to be applied to a sensor, which can be used to detect nanostructures simply by exposing them to infrared light. Conventional approach to the design of metamaterials relies strongly on the human experience and wisdom along with try-and-error processes. Digital metamaterial [1] is a two-dimensional nanostructure that is represented by a pixel pattern with the data structure similar to that in the sequences of genes. Application of machine learning technologies to optimize metamaterials properties with respect to such digitized data structure is expected to lead to a high-throughput development of next-generation devices.

Here we adopt a machine learning model called reservoir computing, which has shown high utility for temporal series of data, to predict the optical response of metamaterials. Reservoir computing is a type of artificial neural network, but unlike other models, it achieves a reduction of computational cost by partially fixing neural network (reservoir) and by optimizing only a readout layer. We have previously reported that the precise prediction of the digital metamaterial becomes more difficult by using either an artificial neural network or a metaheuristic algorithm as the structure be more complex [2]. Based on this, we here apply reservoir computing to the digital metamaterial, expecting to improve the prediction accuracy of data with complex structures while reducing the computational cost. Also, the Fast Fourier Transform (FFT) for the structural data is applied as a preprocessing method for an effective feature extraction from the input data [3]. In this presentation, we will report the results of reservoir computing, FFT processing, and their effectiveness on the accuracy specifically in terms of MAE (Mean Absolute Error).

[1] C. D. Giovampaola and N. Engheta, Digital metamaterials, *Nat. Mater.* **13**, pp.1115-1121 (2014).
[2] Y. Sugino, A. Ishikawa, Y. Hayashi, and K. Tsuruta, Automated design of infrared digital metamaterials by genetic algorithm, *Proc. SPIE* **10343**, 103432O (2017).
[3] N. Md Isa, A. Amir, M. Ilyas and M. Razalli, Motor imagery classification in Brain computer interface (BCI) based on EEG signal by using machine learning technique, *Bull. Electr. Eng. Inform.* **8**, pp. 269-275, 2019.

P03-09

Assessment of iterative screening for drug discovery using an unbiased dataset

Akira Kimura[1] akira.yamazaki@riken.jp
Tomohiro sato[1] tomohiro.sato@riken.jp
Teruki Honma[1] honma.teruki@riken.jp

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

Keywords: drug screening, high throughput screening, machine learning, AI

In drug discovery, screening of compound libraries for hit identification is time-consuming, labor-intensive, and expensive process. Iterative screening is a technology that improves the efficiency by utilizing machine learning to iteratively select a batch of compounds in order of their expected activities based on the assay results of previous batches, and there are several reports about its effect. However, there are concerns that the data sets conventionally used to assess iterative screening may contain issues about data biases such as false positive compounds or too high ratio of hit compounds, which could distort the assessment of screening efficiency.
In this study, we examined the effect of iterative screening using a dataset named LIT-PCBA, which is created form PubChem BioAssay and processed to remove biases. Specifically, we performed iterative screening using several machine learning and deep learning models (DNN, LGBM, RF, SVM, XGB). For three targets in LIT-PCBA dataset, approximately 70% of the active compounds were obtained when 35% of the library was screened, which was roughly the same retrieval as previously reported. On the day of the conference, we plan to discuss the results including active learning, hyperparameter tuning, and so on.

Figure1. Recovery ratio of active compounds during screening

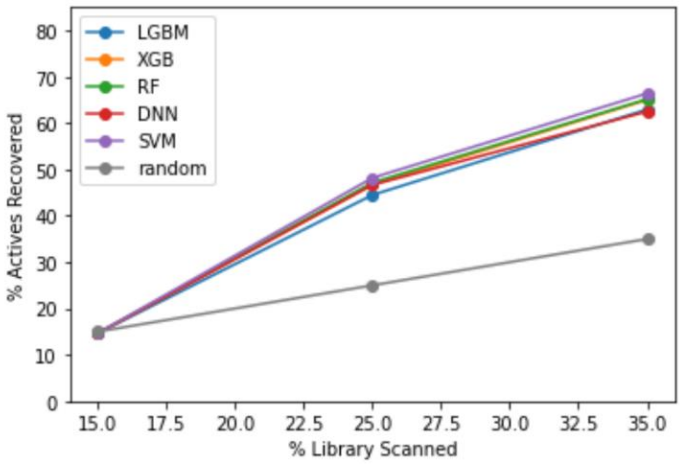
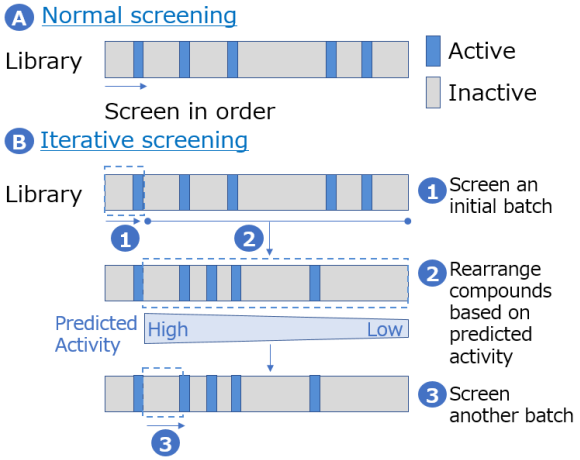


Figure2. Difference between normal screening and iterative screening



[1] Macarron R.; *et al.*, Impact of High-Throughput Screening in Biomedical Research, *Nat. Rev. Drug Discov*, **2011**, 10, 188-195.
[2] Viet-Khoa Tran-Nguyen; *et al.*, LIT-PCBA: An Unbiased Data Set for Machine Learning and Virtual Screening, *Journal of Chemical Information and Modeling*, **2020**, 60(9), 4263-4273.
[3] Gabriel H.S. Dreiman; *et al.*, Changing the HTS Paradigm: AI-Driven Iterative Screening for Hit Finding, *SLAS Discovery*, **2021**, 26(2), 257-262.
[4] Isidro Cortés-Ciriano; *et al.*, Discovering Highly Potent Molecules from an Initial Set of Inactives Using Iterative Screening, *Journal of Chemical Information and Modeling*, **2018**, 58(9), 2000-2014.

Constructing a systematic knowledge graph using real-world data

Takeshi Takagi[1]
takagi.takeshi489@eps.co.jp

Hideki Inoue[1]
inoue.hideki296@eps.co.jp

Atsushi Kikkawa[1]
kikkawa.atsushi109@eps.co.jp

Yukio Kitajima[1]
kitajima.yukio181@eps.co.jp

Nobuyuki Okajima[1]
okajima.nobuyuki505@eps.co.jp

[1] EPS Corporation, Tsuruya Building, 2-23 Shimomiyaibicho, Shinjuku-ku, Tokyo, 162-0822, Japan

Keywords: Real-world data, Knowledge graph, tf-idf

Objective: A knowledge graph is a knowledge base that systematically connects various types of information and represents them in a graph structure. By representing both structured and unstructured data as a knowledge graph, it becomes easier to analyze the relationship of data.
In this study, we investigated a method for systematically creating knowledge graphs from tabular data, performed transformations using the investigated method on actual real-world data, and examined methods for evaluating the validity of the created knowledge graphs.

Methods: Using the medical information database (RWD-DB) operated by Real World Data, Co., Ltd., we created a knowledge graph based on the method of Linfeng et al[1].
For each patient, only the receipt disease names that matched the electronic medical record disease names were extracted, and patient information, diseases, drugs, and laboratory values were obtained for each month. Data represented by continuous numerical values, such as laboratory values, were categorized as above/below the reference value and represented as nodes in the graph structure.
Next, tf-idf used in natural language processing was applied as a method to evaluate the validity of the knowledge graph. The tf-idf used in natural language processing is an index that expresses “how important each word is within each document.” In this study, a disease was considered as a document and a laboratory test value as a word, and the tf-idf value was calculated with the intention to determine “how important each laboratory test value is for the disease.”

Results: In disease E11 (type 2 diabetes mellitus), laboratory values indicating high HbA1c and decreased renal function had higher tf-idf values. In N18 (chronic kidney disease), laboratory values indicating decreased renal function also had higher tf-idf values.

Discussion: The above results suggest that the knowledge graph created in this study captures the characteristic laboratory values of each disease. Knowledge graphs using medical data are expected to be applied to support diagnosis and drug prescriptions in the medical field in the future of medicine[2].

[1] Linfeng, L.; Peng, W.; Jun, Y.; *et al.*, Real-world data medical knowledge graph: construction and applications, *Artificial Intelligence In Medicine*, **2020**, *103*, 101817.
[2] Huang, X.; *et al*., A knowledge graph based question answering method for medical domain. *PeerJ Comput Sci*, **2021**, *7*, e667.

Proposal and Outline of the “Autonomous Chemistry” Research

Kohtaro Yuta

k-yuta@insilicodata.com
In Silico Data, Ltd., 5-19-5, Akitsu, Narasino, Chiba, 275-0025, Japan

Keywords: Autonomous Chemistry, Data Science, Autonomous drug design

The advocacy of "autonomous chemistry" is the first presentation in Japan (1). In terms of technology, “autonomous” is the evolution of technology that comes after automation, and is a form of technology in the near future (Figure1). Currently, the word "autonomous chemistry" is extremely rare in the world. In this presentation, we will explain the outline of “autonomous chemistry” research and discuss "autonomous drug design", which is one of the application cases.

Once the definition and basis of "autonomous chemistry" are determined, it will be applicable to all research fields related to chemistry (drug discovery, toxicity, physical properties, functional compounds, food, etc.). The difference between “automation” and “autonomy” lies in the high level of human judgment. There are no judgment items in “automation”, but “autonomous” is characterized by a large number of advanced and complex judgment items.

Autonomous is strongly deployed in the automotive field (Figure2). Next-generation automated driving is “autonomous driving,” and research examples in this field can be used as references for applications in the field of chemistry. Autonomous driving has 6 levels, from level 0 to 5, until completion. (2) Technology is accumulated sequentially from the starting level 0 (no automation), and it will be completed at level 5 (complete driving automation).

“Autonomous chemistry” in the field of chemistry also sets a level so that the degree of completion can be judged by the value of this level. For example, the implementation of drug design requires processing and judgment of various information, and there are extremely complicated processes such as evaluation of toxicity and pharmacological activity, and reorganization of projects.

It is difficult to solve drug design research involving such complicated decisions by simple “automation”, and “autonomous” research can handle it. (Figure3).

In real drug design process, complex technologies are continuously applied and sophisticated judgments are made according to changing circumstances. This is the category of autonomous drug discovery. Autonomous drug design is carried out while the system judges the collaboration and continuity of a wide variety of technologies required for drug design. In other words, Dr. Horii proposes in his lecture (3) that "an approach to system construction from the point of contact between artificial intelligence and human intelligence is desired as a future prospect". Autonomous drug discovery is exactly what makes this possible.

[1] http://www.insilicodata.com/index-others.html
[2] https://www.macnica.co.jp/business/maas/columns/135343/
[3] https://cbi-society.org/taikai/taikai21/FS/FS-07.pdf

Figure 1. Technological advancements

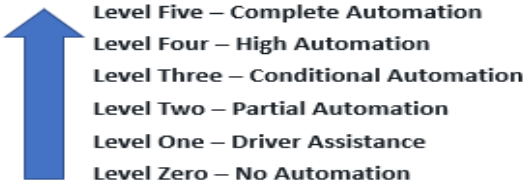


Figure 2. 5 Autonomous Driving Levels

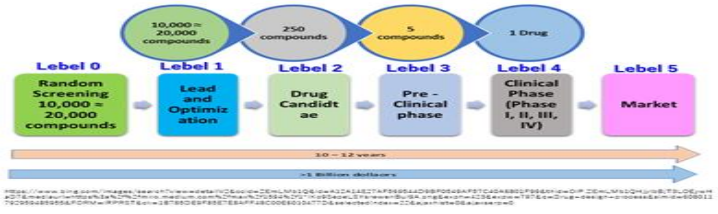


Figure 3. Level division of the drag design process according to the autonomous concept

P03-12

Protein-ligand complex structure generation with diffusion-based generative models

Shuya Nakata[1] 218x219x@stu.kobe-u.ac.jp
Yoshiharu Mori[1] ymori@landscape.kobe-u.ac.jp
Shigenori Tanaka[1] tanaka2@kobe-u.ac.jp

[1] Graduate School of System Informatics, Kobe University, Japan

Keywords: Ligand docking, Machine learning, Generative model

The complex structures of proteins and ligand molecules provide valuable information about their interactions and are important for structure-based drug design. Docking simulations are widely used to computationally predict complex structures, but the number of degrees of freedom that can be addressed is limited, making it difficult to obtain accurate results using uncomplexed structures.

In this study, we developed a machine learning model to generate protein-ligand complex structures taking into account both receptor and ligand flexibility, which is difficult to achieve with conventional methods. To generate complex structures allowing for conformational changes upon binding, we extended and applied the diffusion-based generative models [1,2] to the protein-ligand complex structures in the PDBbind database and directly modeled the 3D coordinates of protein Cα atoms and ligand non-hydrogen atoms. Using a low-resolution distance matrix between protein residues and a 2D structure of the ligand as input, trained models were able to generate a variety of complex structures, including ones close to the experimental structure (L-RMSD<3Å).

The details of the method and the results of the evaluation of the generated structures will be discussed in the presentation.

[1] Kingma, D., Salimans, T., Poole, B. and Ho, J. Variational Diffusion Models, *NeurIPS*, **2021**, 34, 21696-21707

[2] Hoogeboom, E., Satorras, V., Vignac, C. and Welling, M. Equivariant Diffusion for Molecule Generation in 3D, *arXiv:2203.17003*, **2022**

P03-13

Development of a method for building a long-term disease progression model using neural networks

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

Hideki Yoshioka[1, 2]
ysokhdk@gmail.com

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

[1] Clinical Pharmacology and Pharmacometrics, Graduate School of Pharmaceutical Sciences, Chiba University, 1-8-1, Inohana, Chuo-ku, Chiba-shi, Chiba, 260-8675, Japan.
[2] Pharmaceuticals and Medical Devices Agency, Tokyo, Japan.

Keywords: Disease progression model, Neural network, Biomarker, TensorFlow

In this study, a new method was developed using machine learning to estimate long-term chronic disease progression over decades from short-term fragmented data typically found in big data or real-world data. The new method breaks through the computational limitations (several thousand subjects and less than ten biomarkers) of our previous method, the Statistical Restoration of Fragmented Time course (SReFT) [1-3], and would be applicable for precision medicine in the future.

Our new method consists of two networks. The first network relocates the short-term observations over a long-term time axis. The second network estimates the transition of each biomarker. In this presentation, we tested the validity of the analysis and the ability to reject meaningless biomarkers as noise appropriately. Simulation results using synthetic data with varying numbers of cases (500 to 5000) and biomarkers (5 to 30) will be discussed. Individual observations were synthesized assuming a sigmoid function with error and covariate effect. Noise biomarkers were given random numbers between 0 and 1 as observations. For a neural network library, we used TensorFlow (version 2.8.0). As a result, the present method predicted long-term biomarker transition from short-term synthesized observations and appropriately judged that the noise biomarkers did not have meaning throughout the entire period.

[1] Ishida, T.; *et al.*, A Novel Method to Estimate Long-Term Chronological Changes From Fragmented Observations in Disease Progression., *Clin. Pharmacol. Ther.*, **2019**, 105 (2), 436–447.
[2] Kawamatsu, S.; *et al.*, A. 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.* **2020**, 9 (8), 2676.
[3] Jin, R.; *et al.*, Statistically estimating life-long progression of chronic disease from the information of short clinical trials. *CBI Annual Meeting 2020.* O2-14.

Computational search for compounds exhibiting the same activity from database by molecular fingerprints and 3D pseudo-atom methods

Mizuki Matsushita[1]
21bs120b@shinshu-u.ac.jp

Yuka Yamashita[1]
20bs117j@shinshu-u.ac.jp

Koji Umezawa[1,2]
koume@shinshu-u.ac.jp

1 Grad. Sch. of Sci. & Tech., Shinshu University, Minami-minowa 8304, Kami-ina, Nagano, 399-4598, Japan.
2 Institute for Biomedical Sciences, Shinshu University, Minami-minowa 8304, Kami-ina, Nagano, 399-4598, Japan.

Keywords: Structure-activity relationship, Ligand-based virtual screening, Berberine

The relationship between chemical substructure and biological activity has attracted attention in drug discovery. The chemical substructures within a compound can be described as a vector of molecular fingerprint. When two compounds exhibit a similar molecular fingerprint, their activities should be similar. Comparing molecular fingerprints have been widely used to retrieve the same active compounds as a reference compound from chemical database. However, the molecular fingerprint search contains difficulty in selection of an appropriate method to collect the targeted compounds accurately. It needs to be resolved that the result of fingerprint search depends on a reference compound and its activity.

In this study, a general substructure search has been developed for different biological activities. Substructures were newly defined based on heteroatom connection and aromatic ring. We used 3D information of substructures for similarity search. The mean positions of substructures were calculated and were named as pseudo-atomic sites. The similarity was judged by satisfaction of the pseudo-atomic distances with those of the reference compound. In order to estimate performance of our method, we used the chemical database, LigandBox (T. Kawabata et al., 2013), and true collection of the same active compounds were taken from drug groups in KEGG DRUG (M. Kanehisa et al., 2021). The conventional fingerprint methods of MACCS keys, topological (RDKit), Morgan, and Avalon were used, and the similarity was judged by Tanimoto coefficient, which are implemented in RDKit open-source cheminformatics.

The 3D pseudo-atom method and the conventional fingerprint methods were applied for collection of compounds in ACE inhibitor, HMG-CoA inhibitor, and β-lactam antibiotic groups. The accuracy and completeness of collection was evaluated by the receiver operating characteristic (ROC) curve. The results showed that Morgan fingerprint was most suited in the case of ACE inhibitor and HMG-CoA inhibitor groups. However, for β-lactam antibiotic group, Morgan fingerprint showed worst performance. The 3D pseudo-atom method enabled better search among them.

The 3D pseudo-atom method can be straightforwardly applied for natural compounds. Natural compound generally exhibits various biological activities. Berberine-like compounds were searched by the reference of berberine. The details will be discussed in the presentation.

P04-02

Analysis of protein-protein dynamical interaction by molecular dynamics and principal component analysis

Keima Kawada[1] cske21006@g.nihon-u.ac.jp
Masanori Yamanaka[1] yamanaka.masanori@nihon-u.ac.jp

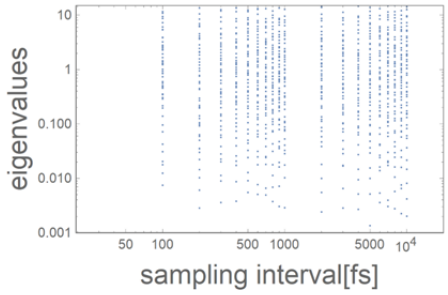
[1] Department of Physics, College of Science and Technology, Nihon University, 1-8 Kannda-Surugadai, Chiyoda-ku, Tokyo, 101-8308, Japan

Keywords: Protein interaction, Molecular dynamics, Principal component analysis

The study of protein-protein and protein-ligand interactions is of great importance in the field of drug discovery. Static analysis and statistical averaging analysis of snapshots of molecular dynamics trajectories have been carried out in literature. However, there is no analysis of interactions that preserves the causal relationships of the time series. Generally, the analysis of dynamic properties is known to be difficult, not only for proteins.

In this study, we propose an analysis method that preserves the causal relationships of time series of protein-protein or protein-ligand interactions. Note that this study is not a simple statistical average over a large number of snapshots cut from a time series. The test target protein is PDBID: 6DKY, which consists of two small ring proteins. The intermolecular interactions between the protein rings were investigate using molecular dynamics [1] and principal component analysis [2]. The trajectories of the atomic coordinates were determined numerically using molecular dynamics in water media. The total duration of the molecular dynamics is 100ns. The variance-covariance matrices between atoms were calculated from the obtained trajectory. When creating the variance-covariance matrices, the sampling interval of the atomic coordinates was systematically varied from 100fs to 10ns. These matrices were diagonalized and subjected to principal component analysis.

The results of eigenvalues as a function of the sampling time interval are shown in Fig.1. The amino acid residues and atoms with interaction estimated from the second principal component vector among the results of principal component analysis are listed in Table 1. The interacting atoms are found to depend on the sampling interval. Thus, this result leads to the following important conclusions: when sampling and taking statistical averages in molecular dynamics, it is meaningless unless the sampling interval is set to be optimal for the dynamics of the residue or ligand to be examined.



interval	PC#	Chain A/B	Residue	Atom
100fs	2	A	ILE6	C
100fs	2	B	GLY5	O
100fs	2	A	ILE6	N
100fs	2	A	ILE3	C
100fs	2	B	LEU4	C
100fs	2	B	LEU4	N
1ps	2	A	ILE3	C
1ps	2	A	ILE2	C
1ps	2	A	LEU6	O
1ps	2	B	ILE6	C
1ps	2	A	LEU7	C
10ps	2	B	ILE3	C
10ps	2	B	ILE3	N
10ps	2	B	ILE6	C
10ps	2	B	LEU7	C
10ps	2	B	LEU6	N
10ps	2	B	LEU7	N

Fig.1 (left) Eigenvalues of the variance-covariance matrices. Horizontal axis is the sampling interval of the snapshot of the molecular dynamics. Table 1 (right) Interacting residues and atoms obtained from second principal component vector.

[1] https://www.gromacs.org
[2] K. Peason, Philosophical Magazine 2 (11), 559 (1901).

P04-03

Constructing QSAR models for c-Met inhibitors

Yu-Shi Tian[1], **Miho Sasaoka**[1]
yushi-tian@osaka-u.ac.jp miho914.ikgk@gmail.com

Daisuke Takaya[1], **Tatsuya Takagi**[1], **Kaori Fukuzawa**[1]
takaya-d@phs.osaka-u.ac.jp ttakagi@osaka-u.ac.jp fukuzawa-k@phs.osaka-u.ac.jp

[1] Graduate School of Pharmaceutical Science, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565-0871, Japan

Keywords: Quantitative Structure-Property Relationships, c-Met inhibitors, anti-tumor, classification and regression models

c-Met is a receptor-type tyrosine kinase (RTK), plays an important role in cell proliferation, migration and invasion by binding hepatocyte growth factor (HGF), and is involved in normal embryonic development, wound healing, and tissue regeneration. However, abnormal HGF/Met signaling pathway activation has been identified in various tumors, including liver, breast, pancreas, lung, kidney, bladder, ovary, brain, and prostate, and c-Met amplified cell lines show high sensitivity to c-Met inhibitors. Furthermore, the HGF/c-Met pathway is involved in the acquired resistance to EGFR tyrosine kinase inhibitors (EGFR-TKI), and it has been reported that c-Met-mediated acquired resistance can be overcome by concomitant use of c-Met inhibitors in addition to EGFR-TKI. Based on these findings, c-Met is expected to be a promising molecular target in cancer therapy, and inhibitors are being developed worldwide.

We constructed and compared classic QSAR models of inhibitory activities of published c-Met inhibitors for future novel inhibitor search. Structure information and inhibitory activities of 424 inhibitors were obtained from published articles. First, we constructed classification models using 2D descriptors calculated by Mordred and general machine learning methods (i.e., kernel Support Vector Classification (kernel SVC), Random Forest Classification (RFC), Logistic Regression (LR)). The objective variable was binarized as high and low classes per pIC_{50} threshold (pIC_{50} = 9, 8, and 7). The best accuracy for test sets was obtained as about 84% using kernel SVC, RFC, and pIC_{50} = 8. We used one-class SVM (OCSVM) to improve the model further by removing outliers. Consequently, the RFC can give as high accuracy as 83%, with both recall and specificity higher than 80%. Therefore, we believe that this model can be used to filter possible candidate compounds.

Furthermore, regression models were conducted using both 2D and 3D descriptors and regression models (i.e., kernel SVR, linear SVR, RFR, and LASSO). 3D descriptors were calculated from the docking poses of all inhibitors conducted by MOE Dock. Before docking, the structure of the c-Met protein was obtained from PDB (ID: 3ZXZ), and the ligand PF-04217903 was re-docked to the protein to check the docking procedure and parameters. An RMSD of 0.6 was obtained from the re-docking. Therefore, all the compounds were docked to c-Met to achieve the docking poses that were further used to calculate 3D descriptors. A two-step explanatory variable selection was conducted using the generic algorithm SVM (GASVM) and permutation importance (PIMP). Finally, the kernel SVR showed the best prediction accuracy (R_{test}^2 = 0.603).

In summary, this study focuses on the anti-cancer target c-Met. We collected the previously reported c-Met inhibitors and constructed classic QSAR models of both classification and regression models. The accuracy of the models is high, and these models can be used in future initial screening of novel c-Met inhibitors.

Acknowledgments:
The major part of this study was conducted as the master's thesis of Ms. Miho Sasaoka, who is now graduated. Hirose Foundation financially supported this work (Research Grant to Yu-Shi Tian).

P05-01

Molecular Evolution of Peptides and Energy Level Statistics of Dipeptides and Tripeptides

Masanori Yamanaka[1]
yamanaka.masanori@nihon-u.ac.jp

[1] Department of Physics, College of Science and Technology, Nihon University, 1-8 Kanda-Surugadai, Chiyoda-ku, Tokyo, 101-8308, Japan

Keywords: Molecular evolution, Density functional theory, Energy level statistics

Based on the energy level statistics of amino acids [1, 2], dipeptides, and tripeptides, the molecular evolution of peptides and proteins was discussed. The energy levels of these peptides were calculated numerically using ab initio methods, and the statistical distributions of the energy level statistics were calculated according to the energy level analysis method in random matrix theory. Specifically, molecular orbitals were calculated by ab initio methods and Kohn-Sham orbitals were calculated by density functional theory. To improve the statistical accuracy, we performed a multipoint calculation on 10000 molecular structures produced via the molecular dynamics simulation. For the valence orbitals, the energy-level statistics exhibit repulsion, but the universality in the random matrix cannot be determined. For the unoccupied orbitals, the energy-level statistics agrees with the critical level statistics which is an intermediate distribution between the Gaussian orthogonal ensemble and the semi-Poisson statistics for all 20 kinds of amino acids and for all 400 dipeptides. The energy level statistics for most molecules were found to be critical statistics.

During the hundreds of millions of years of molecular evolution of proteins, peptides in a state of critical statistics are thought to have differentiated into Poisson statistics and Gaussian orthogonal ensemble [1]. The present study [3] suggests this prediction.

[1] G. Vattay, D. Salahub, I. Csabai, A. Nassimi, and S. A. Kaufmann, 2015 J. Phys.: Conf. Ser. 626, 012023.
[2] M. Yamanaka, J. Phys. Soc. Jpn, 86 (2017), 023801.
[3] This work is supported by JSPS KAKENHI, Grant Number 19K21857.

KEK GoToCloud project: An Application of Cloud Computing for Structural Biology Research in KEK

Yusuke Yamada[1]
yusuke.yamada@kek.jp

Toshio Moriya[1]
toshio.moriya@kek.jp

Toshiya Senda[1]
toshiya.senda@kek.jp

[1] Structural biology research Center, Institute of Materials Structure Science, High Energy Accelerator Research Organization, 1-1 Oho, Tsukuba, Ibaraki, 305-0801, Japan

Keywords: Cloud computing, Cryo-electron microscope, Macromolecular crystallography

Structural Biology Research Center in High Energy Accelerator Research Organization (KEK) develops and operates structural biology research facilities such as macromolecular crystallography beamlines, solution scattering beamlines at synchrotron radiation rings and a cryo-electron microscope. We develop measurement and data analysis techniques with these facilities and provide them to the academic and industrial users. The recent advances of methods, especially the automation of measurements, has made it possible to acquire large amounts of data in a short period of time, requiring a large amount of computing resources for analysis, which is an urgent issue to be addressed. We have been focusing on the abundant computing resources and convenience of the public cloud and has been promoting its utilization for the analysis of measurement data (KEK GoToCloud project).

Amazon Web Service (AWS) is one of the providers of cloud computing, and they serve a high-performance computing service, AWS ParallelCluster, which enables us to generate a virtual computer cluster with many types of computer instances and high-performance shared storage. Recently we successfully deploy AWS ParallelCluster for data analysis of macromolecular crystallography and single particle analysis of cryo-electron microscope.

In this presentation, we will introduce our efforts to use cloud computing and show some examples on macromolecular crystallography and single particle analysis of cryo-electron microscope.

Prediction of drug-induced liver injury *in silico* using large-scale adverse event database

Sarara Doi[1]
c172035@ed.nagoya-cu.ac.jp

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

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

[1] Graduate School of Pharmaceutical Sciences, Nagoya City University, 3-1 Tanabe-dori, Mizuho-ku, Nagoya-shi, Aichi, 467-8603, Japan

Keywords: Applicability Domain, Drug-Induced Liver Injury, JAPIC AERS, Random Forest, Synthetic Minority Over-sampling Technique

The risk of drug-induced liver injury (DILI) is difficult to be predicted from data of non-clinical and clinical trials because some of them are idiosyncratic adverse drug reactions. JAPIC AERS [1] built by cleaning the FDA Adverse Event Reporting System (FAERS) is useful as large-scale clinical adverse events information database. In this study, we aimed to build a machine-learning model to predict positive or negative of DILI in drugs from chemical structural information using JAPIC AERS.

We used JAPIC AERS data from the fourth quarter of 1997 to the second quarter of 2020 to extract DILI-positive/negative drugs, which was defined by the Proportional Reporting Ratios (PRR) method and the number of reports. We calculated the 2D molecular descriptors of drugs' chemical structure by alvaDesc [2]. After the feature selection, we constructed a DILI classification model using Random Forest (RF). To improve the imbalanced data, we applied the Synthetic Minority Over-sampling Technique (SMOTE). Furthermore, external validation was done using data of DILIrank which lists the DILI-positive drugs [4] to evaluate the performance of our model.

A dataset of 424 positive drugs and 210 negative drugs was created from JAPIC AERS. We established the machine-learning clarification model to predict the positive or negative of DILI with a ROC-AUC of 0.8 as the internal validation. With the improvement in data imbalance by SMOTE, specificity was improved, resulting in better overall performance. To create a dataset for external validation, 76 positive drugs and 139 negative drugs were obtained from the DILIrank. The ROC-AUC is over 0.7 for external validation after setting Applicability Domain (AD). These results suggested that we successfully constructed a model with high generalization performance.

[1] https://www.japic.or.jp
[2] Alvascience, 2020, https://www.alvascience.com
[3] Chawla NV.; *et al.*, SMOTE: synthetic minority over-sampling technique, *J Artif Intell Res*, **2002**, *16*, 341-378
[4] Chen M.; *et al.*, FDA-approved drug labeling for the study of drug-induced liver injury, *Drug Discov Today*, **2011**, *16(15–16)*, 697–703.

P06-02

Characterizing Predictive Effects of Autophagy Agents on the Early Stages of Embryo Development

Kengo Matsugo 204031_pf@yok.hamayaku.ac.jp
Yuumi Namiki 194018_md@yok.hamayaku.ac.jp
Kohsuke Hayamizu k.hayamizu@hamayaku.ac.jp
Hideko Sone hideko.sone@yok.hamayaku.ac.jp

Yokohama Univ. Pharm., 601 Matano Totsuka Yokohama 245-0066, Japan

Keywords: autophagy; neuronal development; toxicity prediction; teratogens

Autophagy is the major cellular pathway to degrade dysfunctional organelles and protein aggregates during the early stage development. Especially, neurons depend largely on autophagy to maintain cell homeostasis to get rid of damaged. In this way, as the understanding of autophagy progresses, the modulators of the autophagy pathway are being organized. However, under chemical-induced stress conditions, changes in survival mechanisms of cell functions are unknown. Therefore, whether autophagy modulators induces dysfunction of this pathway should be investigated since the dysfunction of this process contributes to the pathology of many human diseases. In this study, we characterized predictive effects of autophagy modulators with known toxic agents and teratogens based on the current knowledge of the different pathways, molecular mechanisms of mammalian autophagy. [1-6].

[1] Klionsky, D.J. Guidelines for the use and interpretation of assays for monitoring autophagy (4th edition)1. Autophagy. 2021 17:1-382.
[2] Yamane J, et al.,StemPanTox: A fast and wide-target drug assessment system for tailor-made safety evaluations using personalized iPS cells. iScience. 2022 Jun 6;25(7):104538.
[3] Yamane J, et al., Prediction of developmental chemical toxicity based on gene networks of human embryonic stem cells. *Nucleic Acids Res*. **2016,** 44, 5515-5528.
[4] Filippone, A. et al., The contribution of altered neuronal autophagy to neurodegeneration. *Pharmacol Ther*. **2022**, 238,108178.
[5] Tang, F et al., The Role of Vti1a in Biological Functions and Its Possible Role in Nervous System Disorders. *Front Mol Neurosci*. **2022**, 15, 918664.
[6] Andrea, K. H. et al., Autophagy in Neurons *Annu Rev Cell Dev Biol*, **2019**, *35*, 477-500.

Application of Deep Learning to the Evaluation of Seizure-like Behavior Using Zebrafish

Mizuho Wakasa[1] **Riki Kajiwara**[2] **Kaori Abe**[1]
Wakasa.Mizuho@otsuka.jp

[1] Department of Investigative Toxicology, Nonclinical Research Center, Tokushima Research Institute, Otsuka Pharmaceutical Co., Ltd., 463-10 Kagasuno, Kawauchi-cho Tokushima 771-0192, Japan
[2] Department of Research Management, Otsuka Pharmaceutical Co., Ltd., 463-10 Kagasuno, Kawauchi-cho Tokushima 771-0192, Japan

Keywords: Toxicity prediction, Deep Learning, Multi-animal tracking, Central Nervous System

Drug-induced convulsive seizures are a serious adverse event and drug evaluation is usually carried out using behavioural indicators in laboratory animals such as rodents. However, these evaluation methods use large numbers of animals and require large amounts of test substance, resulting in low throughput. While tests using laboratory animals provide a lot of useful information, they are also considered to be an area for improvement in terms of animal welfare, speed of research and development and cost. To address this issue, a test method using zebrafish, which can be evaluated with a small amount of test substance (-µg) and can contribute to the promotion of the 3Rs, has been reported using the amount of behaviour immediately after drug addition as an indicator. However, in the previously reported evaluation system, the presence or absence of convulsions is finally determined by visual observation by researchers, which causes problems such as inconsistency in evaluation, reduced throughput and difficulty in multi-tracking. Therefore, in this study, we aimed to solve the above issues by introducing AI into the evaluation of seizure-like behaviour using zebrafish and automating some of the tasks.

In this presentation, an overview of automated tracking for each designated site using DeepLabCut[1-3], extraction of behavioural parameters that change in relation to compound administration, and evaluation of a seizure-like behaviour classification model using the behavioural parameter dataset will be presented.

[1] Lauer, J.; *et al*., Multi-animal pose estimation and tracking with DeepLabCut, *BioRxiv*, **2021.**
[2] https://github.com/DeepLabCut/DeepLabCut
[3] http://www.mackenziemathislab.org/deeplabcut

P07-01

Artificial Intelligence to Accelerate New Drug Discovery: Target identification and drug discovery by data-driven approach and experimental validation in Idiopathic Pulmonary Fibrosis (IPF)

Mari N Itoh[1] mari@nibiohn.go.jp
Yayoi Natsume-Kitatani[1] natsume@nibiohn.go.jp
Masataka Kuroda[1,2] m-kuroda@nibiohn.go.jp
Jun Adachi[3] jun_adachi@nibiohn.go.jp
Yi-An Chen[1] chenyian@nibiohn.go.jp
Kenji Mizuguchi[1] kenji@nibiohn.go.jp
Yoshito Takeda[4] yoshito@imed3.med.osaka-u.ac.jp
Atsushi Kumanogoh[4] kumanogo@imed3.med.osaka-u.ac.jp
Naonori Ueda[5] naonori.ueda@riken.jp

[1] AI Center for Health & Biomedical research, Bioinformatics Project, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki, Osaka, 567-0085, Japan
[2]Discovery Technology Laboratories, Mitsubishi Tanabe Pharma Corporation, 1000, Kamoshida-cho, Aoba-ku, Yokohama, Kanagawa 227-0033, Japan
[3]Laboratory of Proteome Research, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki, Osaka, 567-0085, Japan
[4]Osaka University Graduate School of Medicine, 2-2 Yamada-oka, Suita, Osaka, 565-087, Japan
[5]RIKEN Center for Advanced Intelligence Project, RIKEN, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, 103-0027, Japan

Keywords: Pathway analysis, Exosome Proteomics, Target validation

With the support of PRISM (Public/Private R&D Investment Strategic Expansion PrograM) project, we aim to develop the novel strategy such as data-driven target identification and drug discovery based on patient data.
We have created a dataset linking proteome data in serum exosomes obtained by state-of-the-art data independent acquisition (DIA) and structured electronic medical record (EMR) data for each patient. Furthermore, a novel algorithm named subset binding (SB) made it possible to extract as many features linked to each other as possible from the two different pieces of information.
In this conference, we will report the results of using this algorithm to search for novel drug targets for idiopathic pulmonary fibrosis (IPF), an intractable disease with poor prognosis, as well as the results of pathway analysis and experimental validation as follows,

1) 20 proteins that are strongly connected with medical records, including novel molecules not known to be associated with IPF, were detected. They were up-regulated in the fibrotic areas of patient lungs.
2) Interpreted and analyzed these 20 proteins using bioinformatics methods, utilizing Ingenuity Pathway Analysis (IPA) and TargetMine, to investigate PPI networks and upstream master regulators.
3) Ponatinib, one of the candidates which inhibits upstream master regulators of the detected IPF-related proteins, was experimentally validated in an epithelial-mesenchymal transition of human airway epithelial cells.

Conclusion: The involvement of the molecules identified by SB in the pathogenesis of IPF and their plausibility as drug targets were confirmed from multiple perspectives.

P07-02

Five cluster groups found in the health state space of 96,093 population

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

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

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

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

[1] Dept. of Informatics for Genomic Medicine, Tohoku Medical Megabank Organization, Tohoku University, Tohoku Medical Megabank Organization Bldg 7F room #736, Seiryo 2-1, Aoba-ku, Sendai, Miyagi, 980-8573

Keywords: health state space, dimensional reduction, real-world data

What is health? Can the degree of health be quantified? While it is important to look at clinical values individually, by looking at those multidimensional values as a whole, it is possible to understand the extent to which the subject deviates from a "healthy" population, which may contribute to improving the subject's health status. Therefore, the purpose of this study is to provide a quantitative understanding of quantified health status as a "health state space" that represents the health status of the subjects.

In this study, we analyzed 4 years (2014-2017) health checkup data which consists of 42 variables without missing values (17 clinical variables, 25 lifestyle variables) in 96,093 populations (64,128 males and 31,965 females) from the JMDC database. At first, dimensional reduction was conducted by using PCA and UMAP to define variables that span the health state space. In the health state space, we then detected clusters by conducting DBSCAN. Finally, we categorized detected clusters into cluster-groups by analyzing the adjacency matrix of cluster networks.

As a result, we detected twelve clusters in the health state space, and we found that the health state space of populations consists of five cluster-groups: one healthy cluster-group A and four cluster-groups B-E with health problems. Cluster-group B had features related to diabetes metabolism with higher HbA1c and glucose levels as one moved away from the center of the health state space. Cluster-group C had features related to obesity, lipid abnormalities, liver function abnormalities, and daily alcohol consumption. Cluster-group D was characterized by red blood cell-related abnormalities such as very low red blood cell count. Cluster-group E was characterized by hypertension, abnormal liver function, and a history of cerebrovascular and cardiovascular disease. Regarding transitions between cluster groups, transitions from healthy cluster-group A to each of non-healthy cluster-groups B-E were observed, but not among cluster group B-E. To further quantify the degree of health, we are planning to stratify health cluster-group A and analyze it to quantitatively assess health status within the health state space.

P08-01

Proposal of a simple pose descriptor for prediction of correct docking poses by machine learning approaches

Kentaro Kawai[1] **Daiki Fujioka**[1] **Daiki Minami**[1] **Yukiko Karuo**[1]
kentaro.kawai@pharm.setsunan.ac.jp

Atsushi Tarui[1] **Kazuyuki Sato**[1] **Shinya Nakamura**[2] **Masaaki Omote**[1]

1 Faculty of Pharmaceutical Sciences, Setsunan University, 45-1 Nagaotoge-cho, Hirakata, Osaka 573-0101, Japan
2 Faculty of Pharmaceutical Sciences, Kindai University, 3-4-1 Kowakae, Higashi-osaka, Osaka 577-8502, Japan

Keywords: Ligand docking, Pose prediction, Machine learning

To facilitate SBDD, it is necessary to understand the correct binding pose of a compound to the target protein, but it is difficult to predict the exact binding pose based on docking score alone. Therefore, we tried to predict the correct binding pose using various machine learning methods, i.e., SVM, random forest, and lightGBM [1]. a dataset was generated using binding poses generated by AutoDock Vina [2] and rDock [3], then the RMSD with the crystal structure ligand values less than 2.5 Å were considered as correct binding poses. Here, we compared two methods for numerical expression of binding poses, PLIP [4] and LCP [5]; LCP was implemented by us and is a simple binding pose representation method, which is calculated based on the atom type and coordinates of the ligand atoms.
As a test case, xanthine oxidase was used in this study, and 21 different complex structures were obtained from the PDB, each of which was self-docked to obtain a maximum of 120 different binding poses per complex. The binding poses were represented numerically using PLIP and LCP expressions. The results showed that random forests gave better results for PLIP, and lightGBM gave better results for LCP, respectively. In both cases, the prediction of binding poses for nucleic acid derivatives was found to be poor. We considered this was attributed to their small-sized molecules with many functional groups capable of forming hydrogen bonds.

References:
[1] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T. Y. Liu, NIPS'17: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, **2017**, 3149–3157.
[2] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. **2010**, *31*, 455-461.
[3] Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, Barril X, Hubbard RE, Morley SD. rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput Biol*. **2014**, *10*, e1003571.
[4] Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res*. **2015**, *43*, W443-7.
[5] Kawai K, Asanuma Y, Kato T, Karuo Y, Tarui A, Sato K, Omote M. LCP: Simple Representation of Docking Poses for Machine Learning: A Case Study on Xanthine Oxidase Inhibitors. *Mol Inform*. **2022**, *41*, e2100245.

P08-02

Global Assessment of Substituents on the Basis of Analogue Series

Kosuke Takeuchi[1]
takeuchi.kosuke.pm@daiichisankyo.co.jp

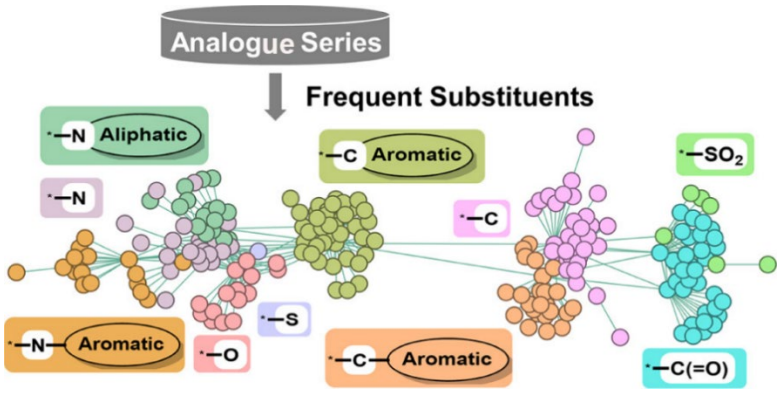
Ryo Kunimoto[1]
kunimoto.ryo.si@daiichisankyo.co.jp

Jürgen Bajorath[2]
bajorath@bit.uni-bonn.de

[1] R&D Division, Daiichi Sankyo Co., Ltd., 1-2-58 Hiromachi, Shinagawa-ku, Tokyo 140-8710, Japan
[2] Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology & Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität Bonn, Friedrich-Hirzebruch-Allee 6, Bonn, D-53113, Germany

Keywords: Substituent Analysis, R-group, Functional group, Fragmentation, Analog series

While bioisosteric replacements have been extensively investigated, comprehensive analyses of R-/functional groups have thus far been rare in medicinal chemistry. We introduce a new analysis concept for the exploration of chemical substituent space that is based upon bioactive analogue series as a source. From ~24,000 analogue series, more than 19,000 substituents were isolated that were differently distributed. A subset of ~400 substituent fragments occurred most frequently in different structural contexts. These substituents contained well-known R-groups as well as novel structures. Substitution site-specific replacement and network analysis revealed that chemically similar substituents preferentially occurred at given sites and identified intuitive substitution pathways that can be explored for compound design. Taken together, the results of our analysis provide new insights into substituent space and identify preferred substituents on the basis of analogue series.



[1] Takeuchi, K.; Kunimoto R.; Bajorath J. Global Assessment of Substituents on the Basis of Analogue Series, *J. Med. Chem*. **2020**, *63*, 15013–15020.
[2] Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. RECAP Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511−522.
[3] Naveja, J. J.; Vogt, M.; Stumpfe, D.; Medina-Franco, J. L.; Bajorath, J. Systematic Extraction of Analogue Series from Large Compound Collections Using a New Computational Compound-Core Relationship Method. *ACS Omega* **2019**, *4*, 1027- 1032.

P08-03

Validation of low-resolution protein crystal structures using deep learning

Takaaki Kuribayashi[1]
kuribayashi-takaaki@mki.co.jp

Ikuko Miyaguchi[2]
miyaguchi.ikuko@mv.mt-pharma.co.jp

Hiroaki Hata[1]
hata-hiroaki@mki.co.jp

Akiko Kashima[2]
kashima.akiko@mb.mt-pharma.co.jp

Shigeyuki Matsumoto[3]
sigeyuki.matsumoto@riken.jp

Masateru Ohta[4]
masateru.ota@riken.jp

Mitsunori Ikeguchi[5]
ike@tsurumi.yokohama-cu.ac.jp

[1] Mitsui Knowledge Industry Co., LTD. 2-5-1 Atago, Minato-ku, Tokyo 105-6215, Japan
[2] Mitsubishi Tanabe Pharma Co., LTD. 1000 Kamoshida, Aoba-ku, Yokohama, Kanagawa 227-0033, Japan
[3] Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan
[4] RIKEN Center for Computational Science, 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan
[5] Graduate School of Medical Life Science, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

Keywords: Electron density, Protein structure, 3D-CNN, X-ray crystallography

Many protein structures have been determined by X-ray crystallography. High-resolution data (electron density map) is required to determine atomic coordinates with high accuracy. For determining side chain orientations, the resolution of 2 Å or higher is often necessary. However, more than half of registered entities in the Protein Data Bank (PDB) have less than 2 Å resolution [1], so that it brings difficulties in drug development or molecular simulation studies.

We therefore started to develop a method named Quality Assessment based on an Electron density map (QAEmap) [2], which is applied 3D-convolutional neural network (3D-CNN) to evaluate local protein structures determined from such a low-resolution data.

Our method estimates how well the structure fits the putative high-resolution structure by calculating a correlation between the coordinate structure to be evaluated and the electron density map of the correct structure in amino acid units. We expect that QAEmap enables protein structures from low-resolution data to be utilized in many studies.

[Acknowledgement] This research was carried out as activities of Life Intelligence Consortium (LINC).

[1] Burley, S.K. et al. Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Res*. 47, D520-D528 (2019).
[2] Miyaguchi, I. et al. Machine learning to estimate the local quality of protein crystal structures. *Sci. Rep*. 11, 1–13 (2021).

Improvement of Compound-Protein Interactions Prediction with Semi-Supervised Learning

Takuto Koyama[1] koyama.takuto.82j@st.kyoto-u.ac.jp
Shigeyuki Matsumoto[1] matsumoto.shigeyuki.4z@kyoto-u.ac.jp
Hiroaki Iwata[1] iwata.hiroaki.3r@kyoto-u.ac.jp
Ryosuke Kojima[1] kojima.ryosuke.8e@kyoto-u.ac.jp
Yasushi Okuno[1] okuno.yasushi.4c@kyoto-u.ac.jp

1 Graduate School of Medicine, Kyoto Univ., 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto, 606-8507, Japan

Keywords: Compound-protein interactions, Deep learning, Data augmentation, Semi-supervised learning

Accurate prediction of Compound-Protein Interactions (CPIs) can accelerate the process of drug discovery. Along with the recent development and expansion of databases associated with CPIs, supervised machine-learning frameworks, especially using deep learning techniques, have become one of the powerful approaches to CPIs prediction [1, 2]. However, the differences in the sample sizes and the affinity distributions cause non-negligible data imbalance for each target protein in the database, significantly impairing the performance of the model trained with the dataset.

To solve the problem, we here propose a semi-supervised learning framework based on a data augmentation approach. We first constructed a deep learning model to predict the CPI activities using a training dataset derived from the ChEMBL database. The model has a multimodal architecture of both chemical structures and protein sequences with the corresponding activity labels as “positive” or “negative” according to a threshold of 1uM. In the present semi-supervised framework, we adopted a bootstrap-based approach that iteratively generates pseudo-labels for compound-protein pairs without a priori interaction information and refines the model parameters. The data imbalance in the generated datasets was shown to be progressively alleviated, and further, the finally constructed model outperformed the initial model which was trained only with the known CPIs data. The present results indicate that semi-supervised learning is a useful approach for augmenting data in the extremely vast CPI space, providing a clue to realize the practical *in-silico* drug discovery using a deep learning model.

[1] Hamanaka, M., Taneishi, K., Iwata, H., Ye, J., Pei, J., Hou, J., & Okuno, Y. (2017). CGBVS‐DNN: Prediction of Compound‐protein Interactions Based on Deep Learning. *Molecular informatics*, 36(1-2), 1600045.

[2] Kojima, R., Ishida, S., Ohta, M., Iwata, H., Honma, T., & Okuno, Y. (2020). kGCN: a graph-based deep learning framework for chemical structures. *Journal of Cheminformatics*, 12(1), 1-10.

P08-05

Search for allosteric chaperones for lysosomal acid α-glucosidase

Manami Yamazaki [1] ml21132@st.kitasato-u.ac.jp
Izumi Nakagome[1] nakagomei@pharm.kitasato-u.ac.jp
Nobutada Tanaka [1] tanakan@pharm.kitasato-u.ac.jp
Yuna Shimadate[2] d2162304@ems.u-toyama.ac.jp
Atsushi Kato [2] kato@med.u-toyama.ac.jp

[1] Graduate School of pharmaceutical science, Kitasato University, 5-9-1 Shirokane, Minato, Tokyo 108－8641, Japan
[2] Department of Hospital Pharmacy, University of Toyama, 2630 Sugitani, Toyama 930-0194, Japan

Keywords: Pompe disease, *N*-acetylcysteine, *in silico* fragment mapping

Pompe disease is a congenital metabolic disorder caused by genetic mutation of acid α-glucosidase (GAA), which breakdowns glycogen in lysosomes. GAA deficiency in Pompe disease results in accumulation of glycogen and myopathy. The approved treatment is enzyme replacement therapy (ERT) with recombinant human GAA, but there have been reports of cases in which treatment is limited by factors such as the emergence of antibodies. Therefore, as an alternative treatment to ERT, pharmacological chaperone therapy, in which a low-molecular-weight compound is bound to an enzyme that has become unstable due to gene mutation, to suppress the degradation of the enzyme, has attracted attention [1]. *N*-acetylcysteine (NAC) has been reported as an allosteric chaperone for GAA, but there is no allosteric chaperone currently used as a therapeutic drug for Pompe disease [2]. Therefore, in this study, with the aim of identifying novel allosteric chaperones for GAA, virtual screening was performed by combining the *in silico* fragment mapping method developed by our laboratory with the docking method. The chaperone activity of the resulting representative compounds against GAA was measured.

[1] Roig-Zamboni, V., Cobucci-Ponzano, B., Iacono, R. *et al.* Structure of human lysosomal acid α-glucosidase–a guide for the treatment of Pompe disease. *Nat Commun* **8**, 1111 (2017).
[2] Caterina Porto *et al.*, *Molecular Therapy* **20**, 2201-2211 (2012).

Prediction of pharmacological activity by deep learning using skeletal formula images

Satoshi Ito[1]
g191tg2001@dhw.ac.jp

Takafumi Nakanishi[1,2]
takafumi@dhw.ac.jp

Masatsugu Hashimoto[1]
mshashimoto@dhw.ac.jp

[1] Digital Hollywood University Graduate School, Ochanomizu Solacity Academia 3F/4F, 4-6 Surugadai, Kanda, Chiyoda-ku, Tokyo,101-8219, Japan

[2] Musashino University, 3-3-3, Ariake, Koto-ku, Tokyo, 135-8181, Japan

Keywords: AI Drug Discovery, Activity/Drug Efficacy (Potency) Prediction, in silico Screening

With machine learning to predict the pharmacological activity of compounds, molecular fingerprint which express the presence of specific molecular structure is used[1]. Since molecular fingerprint focuses on the substructure of a compound, it is presumed to contain less information than the entire structure of the compound. On the other hand, molecular graph which the molecular structure is regarded as a graph and molecular image can be used to represent the entire molecular structure[2]. In attempts to use molecular images, there is a report of case which skeletal formula images are used to predict compound rate constants[3]. And the case which prediction pharmacological activity by deep learning using skeletal formula images is not reported. Therefore, in addition to molecular fingerprint, which focus on partial structures, we confirmed the efficacy of predicting pharmacological activity using skeletal formula images that represent the entire molecule.

Assay data were extracted from ChEMBL and data were split at a ratio of 8:2 for training and testing. SMILES strings were converted into extended-connectivity fingerprint of diameter 4(ECFP4), Molecular Access System(MACCS) keys and skeletal formula images with RDKit. The pharmacological activity is discriminated from pIC_{50}($-\log IC_{50}$). For predictions in ECFP4 and MACCS keys, we used Support Vector Machine(SVM), Random Forest(RF), k-Nearest Neighbors algorithm(k-NN), and Multi layer Perceptron(MLP) as machine learning model. And we used Visual Geometry Group 16(VGG16), Dense Convolutional Network 121(DenseNet121), both are convolutional neural network(CNN) based architectures, and the model provided by AutoKeras[4], which is Automated Machine Learning(AutoML) library, for prediction with skeletal formula images.

We got the results that were 92.5% accuracy, 95.5% recall, and ROC-AUC 0.965 when classified by SVM with ECFP4 using assay data with Dipeptidyl Peptidase 4 (DPP4) as the target protein. The results for the skeletal formula images were 79.5% accuracy, 93.4% recall, and ROC-AUC 0.836 for the model provided by AutoKeras.

We would like to continue to evaluate aspects other than classification performance.

[1] Harinder, S.; *et al*., QSAR based model for discriminating EGFR inhibitors and non-inhibitors using Random forest, *Biology Direct*, **2015**, *10*, 10.
[2] Dejun J.; *et al*., Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models, *Journal of Cheminformatics*, **2021**, *13*, 12.
[3] Shifa, Z.; *et al*., Molecular image-convolutional neural network (CNN) assisted QSAR models for predicting contaminant reactivity toward OH radicals: Transfer learning, data augmentation and model interpretation, *Chemical Engineering Journal*, **2021**, *408*, 127998.
[4] AutoKeras: An AutoML system based on Keras. https://autokeras.com/. (2022.08.08)

P08-07

AI-AAM. Downsizing and scaffold hopping from a peptide to small-molecule inhibitors

Shino Ohira[1]
shino.ohira@fujifilm.com

Jun Nakabayashi[1]
jun.nakabayashi@fujifilm.com

Kyosuke Tsumura[1]
kyosuke.tsumura@fujifilm.com

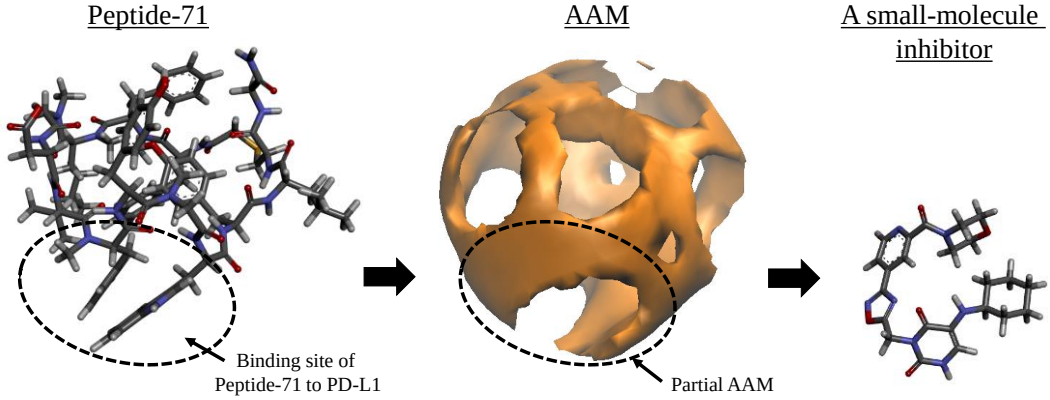
Issei Doi[1]
issei.doi@fujifilm.com

[1] Analysis Technology Center, FUJIFILM Corporation, 210 Nakanuma, Minamiashigara-shi, Kanagawa 250-0193, Japan

Keywords: AI Drug Discovery, LBDD, de novo design, PPIs, AI-AAM

The investigation of unexplored protein-protein interactions (PPIs) for diseases with an unmet medical need is highly demanding [1]. An interest to approach small molecule PPI inhibitor (MW < 500 Da) with high cell permeability instead of peptides and antibodies is increasing, but systematic methodology for finding such compounds has not yet been established.

We developed a peptide downsize and scaffold hopping methodology using AI-Amino Acid Mapping (AI-AAM) that could find and design compounds with desired binding affinities to a target starting from a known active compound with modifying its original core structure based on interactions between the ligand and the set of amino acids [2]. Using the methodology, we designed a novel compound starting from a cyclic peptide with 14 amino acids, Peptide-71 (MW = 1.8 kDa) binding to PD-L1 [3]. Whereas the starting compound Peptide-71 had high *in vitro* binding affinity with low MDCK-II permeability, we experimentally confirmed that the computed compound exhibited binding efficiency index (BEI= *p*Ki/Mw×1000) twice as high as the original compound and high permeability which was suitable for oral absorption (Fig. 1).



	Peptide-71	A small-molecule inhibitor
Ki(μM)	0.004	39
BEI(pKi/Mw×1000)	4.7	9.2
Permeability(10^{-6} cm/s)	0.08±0.02	11.1±0.8

Fig. 1 Downsizing and scaffold hopping from Peptide-71 to a novel small-molecule inhibitor via AAM

[1] Cunningham, A. D.; *et al*., Peptides and peptidomimetics as regulators of protein-protein interactions, *Curr. Opin. Struct. Biol.* 2017, 44, 59.

[2] K. Tsumura.; *et al*., A newly developed method based on AI-oriented amino acid interaction mapping (AI-AAM) for efficient virtual scaffold hopping, *CBI Annual Meeting 2018*, P2-02.

[3] K Magiera-Mularz.; *et al.*, Bioactive Macrocyclic Inhibitors of the PD-1/PD-L1 Immune Checkpoint, *Angew. Chem. Int. Ed.* 2017, 56, 1.

Graph based property and affinity prediction using open-source tool and cloud platform

Akitoshi Okada[1]
akitoshi.okada@jt.com

Hiroyuki Ogawa[1]
hiroyuki.ogawa@jt.com

Yohei Nakayama[2]
nak@amazon.com

Tatsuya Arai[2]
araitats@amazon.com

[1] Chemical Research Laboratories, Central Pharmaceutical Research Institute, Japan Tobacco Inc., 1-1, Murasaki-cho, Takatsuki, Osaka 569-1125, Japan
[2] Amazon Machine Learning Solutions Lab, Amazon Web Services Inc., 2795 Augustine Dr. Santa Clara, CA 95054

Keywords: Graph neural networks, Machine Learning, Molecular Property prediction

Molecular property prediction using machine learning model is widely used in the pharmaceutical industry [1]. Graph neural network (GNN) is one of the methods to construct such molecular property prediction model. GNN is becoming a popular method since the model prediction accuracy can outperform the other prediction methods [2], although the descriptor-based models can still be the first choice [3] depending on the dataset available to construct the model. It is therefore becoming more common to add GNN approach to other existing methods and verify the most suitable model for the target property prediction.

DGL-LifeSci [4], is an open-source python based GNN package and several GNN approach is available within the package. It is based on Deep Graph Library [5] and can be used to efficiently construct graph based molecular prediction models. With the increasing number of dataset available, computational cost can become a bottleneck in model construction and comparison. Therefore, cloud platform such as Amazon Web Service (AWS) can also be useful.

We applied DGL-Lifesci to our internal dataset to construct several properties and affinity prediction models. Computation was performed on AWS to maximize the calculation speed. There were several promising targets where GNN outperformed the other methods, and we will discuss the efficiency of both the computation and the model performance of this approach.

[1] R.S.K. Vijayan, Jan Kihlberg, Jason B. Cross, Vasanthanathan Poongavanam. *Enhancing preclinical drug discovery with artificial intelligence*, Drug Discovery Today, **2022**, 27, 4, 967-984

[2] Oliver Wieder, Stefan Kohlbacher, Mélaine Kuenemann, Arthur Garon, Pierre Ducrot, Thomas Seidel, Thierry Langer, *A compact review of molecular property prediction with graph neural networks*, Drug Discovery Today, **2020**, 37, 1-12

[3] Dejun Jiang, Zhenxing Wu, Chang-Yu Hsieh, Guangyong Chen, Ben Liao, Zhe Wang, Chao Shen, Dongsheng Cao, Jian Wu, Tingjun Hou, *Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models*, Journal of Cheminformatics **2021**, 13, 12

[4] Mufei Li, Jinjing Zhou, Jiajing Hu, Wenxuan Fan, Yangkang Zhang, Yaxin Gu, George Karypis. *DGL-LifeSci: An Open-Source Toolkit for Deep Learning on Graphs in Life Science,* ACS Omega **2021**, 6, 27233−27238

[5] Minjie Wang, Da Zheng, Zihao Ye, Quan Gan, Mufei Li, Xiang Song, Jinjing Zhou, Chao Ma, Lingfan Yu, Yu Gai, Tianjun Xiao, Tong He, George Karypis, Jinyang Li, Zheng Zhang. *A Graph-Centric, Highly-Performant Package for Graph Neural Networks*, **2019** arXiv:1909.01315

Development of a screening system for RNA-binding small molecules based on large-scale information of interactions between RNAs and a fluorescence indicator

Ryosuke Nagasawa[1,2], **Kazumitsu Onizuka**[1,2], **Kaoru R. Komatsu**[4],
nagasawa.ryosuke.p4@dc.tohoku.ac.jp, kazumitsu.onizuka.d7@tohoku.ac.jp, krk@xforestx.com,
Emi Miyashita[3,4], **Mamiko Ozawa**[1], **Hirohide Saito**[3],
emi.miyashita@xforestx.com, mamiko.ozawa.c7@tohoku.ac.jp, hirohide.saito@cira.kyoto-u.ac.jp
Fumi Nagatsugi[1,2]
fumi.nagatsugi.b8@tohoku.ac.jp

[1] IMRAM Tohoku Univ., Katahira 2-1-1, Aoba-ku, Sendai, Miyagi, 980-8577, Japan,
[2] Department of Chemistry, Tohoku Univ., Aoba-ku, Sendai, Miyagi 980-8578, Japan,
[3] CiRA Kyoto Univ., 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan,
[4] xFOREST therapeutics, 448-5 Kajiicho, Kamigyo-Ku, Kyoto, 602-0841, Japan

Keywords: RNA-binding small molecules, High-throughput screening, FID assay

RNA-binding molecules are attractive targets for their potential as therapeutic agents for gene-related diseases and infections. Fluorescent indicator displacement (FID) assay is a method to find novel RNA binding molecules. Light-up properties of RNA-binding small molecules enable us to conduct a high-throughput screening by FID assay. For example, Nishizawa's group found TO-PRO-3 as a powerful fluorescent indicator, and showed an example of FID assay for the bacterial A-site RNA.[1] However, RNA targets that have been reported in FID assay are still limited, it is essential to investigate the binding affinity and selectivity of such chemical probes toward various RNAs in order to further expand the target range.

A new method, FOREST (folded RNA element profiling with structure library), was previously developed for the large-scale analysis of protein-RNA interactions.[2] We have developed large-scale analysis to investigate the interactions of light-up probes with the RNA structure library (3000 structures form pre-miRNA loop, virus RNA and repetitive RNA) by applying FOREST to small molecules so far.

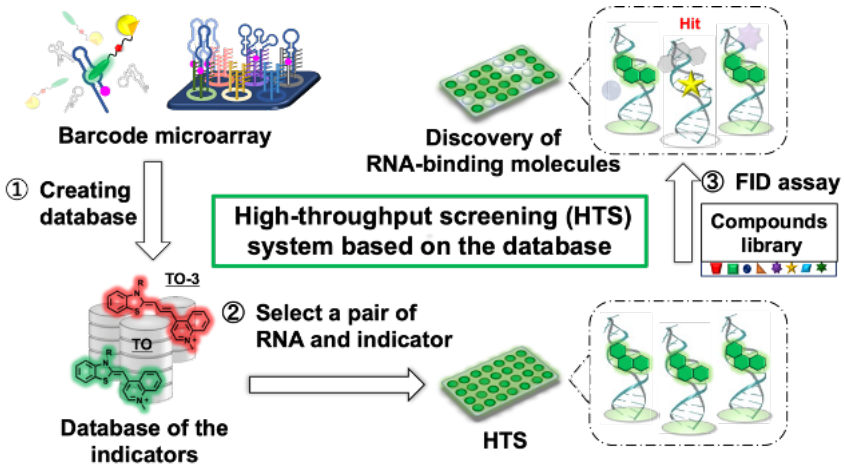


Fig. High-throughput screening system based on the database

In this study, we utilized FOREST to obtain the RNA-binding information of fluorescence indicators, thiazole orange (TO) type probes, that are known as light-up probes (Fig). From the obtained information, we selected pairs of a target disease-related RNA and a fluorescence indicator which strongly binds to the target RNA for FID assay. As a result of the FID assay, we identified several RNA-binding small molecules. In this presentation, we will report our method, the analytical results and FID assay based on the obtained information in detail.

[1] S. Nishizawa, et al., *Chem.Commun.* **2019**, *55*, 3183.
[2] H. Saito, et al., *Nat. Commun.* **2020**, *11*, 6275.

Reducing the Data Bias in the PDBbind Database

Nozomu Yamazaki[1]
yamazaki@cb.cs.titech.ac.jp

Takashi Ishida[1]
ishida@cs.titech.ac.jp

[1] Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan

Keywords: Protein-Ligand Interaction, Binding Affinity Prediction, Drug Discovery

Prediction of protein-ligand binding affinities using structural information plays an essential role in drug discovery. Deep learning models based on 3D convolutional neural networks (3DCNN), or graph neural networks (GCN) have been proposed in recent years [1-2], and many of them use the PDBbind database for the dataset. However, in the PDBbind database, similar ligands tend to have similar binding affinities. Therefore, when a simple random split is used, the binding affinity can be predicted from the ligand structure alone. The same is true when using protein structures. Therefore, existing research using the PDBbind database may be overestimating its prediction accuracy. Here, we constructed a data split that reduces the effect of such bias by performing similarity clustering based on protein sequences and ligand molecular fingerprints.

To investigate whether the bias is reduced, we created a baseline model using 3DCNN and GCN to compare the proposed data split with a random split. We also constructed datasets containing ligand structure alone or protein structure alone to confirm whether the model is learning from protein-ligand interactions.

[1] Jiménez J.; et al., K_{DEEP}: Protein–Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks, *J. Chem. Inf. Model,* **2018**, 58, 2, 287–296
[2] Jones D.; et al., Improved Protein–Ligand Binding Affinity Prediction with Structure-Based Deep Fusion Inference, *J. Chem. Inf. Model,* **2021**, 61, 4, 1583–1592
[3] Liu Z.; et al., Forging the Basis for Developing Protein–Ligand Interaction Scoring Functions, *Acc. Chem. Res.,* **2017**, 50, 2, 302–309
[4] Yang, J.; et al., Predicting or Pretending: Artificial Intelligence for Protein-Ligand Interactions Lack of Sufficiently Large and Unbiased Datasets, *Front. Pharmacol.*, **2020**, 11, 69

Estimation of Interaction Mechanism in Compound-Protein Interaction Prediction Using Interpretable Deep Learning

Sota Inoue[1]
inoue.sota.65r@st.kyoto-u.ac.jp

Shoichi Ishida[2]
ishida.sho.nm@yokohama-cu.ac.jp

Hiroaki Iwata[1]
iwata.hiroaki.3r@kyoto-u.ac.jp

Shigeyuki Matsumoto[1]
matsumoto.shigeyuki.4z@kyoto-u.ac.jp

Ryosuke Kojima[1]
kojima.ryosuke.8e@kyoto-u.ac.jp

Yasushi Okuno[1]
okuno.yasushi.4c@kyoto-u.ac.jp

[1] Graduate School of Medicine, Kyoto Univ., 53 Kawahara-cho, Syogoin, Sakyo-ku, Kyoto, 606-8507, Japan
[2] Graduate School of Medical Life Science, Yokohama City Univ., 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045, Japan

Keywords: in silico Screening, Deep Learning, Compound-Protein Interaction

Understanding compound-protein interactions (CPI) is important for identifying hit compounds in the early stages of drug discovery. Recent advances in the organization of bio-activity databases and the development of deep learning (DL) methods have enabled rapid and accurate predictions of CPIs, but DL methods still have limitations in explaining why the predictions were made [1,2]. In this study, we propose a CPI prediction model that allows physical and biological interpretation for the prediction results. In the first step, we constructed a multimodal model using Graph Convolutional Networks (GCN) for compounds and Convolutional Neural Networks (CNN) for proteins to predict CPI. In the second step, we used Integrated Gradients (IG) [3] to visualize the contribution of the atoms of compounds and amino acid residues of proteins to the predicted results, and discussed them. As a result of visualization, it is suggested that our model focuses, in the process of prediction, on the characteristics of the interaction sites in the CPI. The visualization methods in this study were shown to be useful for evaluating the validity of the prediction results in the DL-based CPI prediction model.

[1] Hamanaka, M.; Taneishi, K.; Iwata, H.; Ye, J.; Pei, J.; Hou, J.; Okuno, Y.; CGBVS-DNN: Prediction of Compound-protein Interactions Based on Deep Learning, *Mol. Inf.*, **2017**, 36, 1600045.
[2] Verma, N.; Qu, X.; Trozzi, F.; Elsaied, M.; Karki, N.; Tao, Y.; Zoltowski, B.; Larson, E.C.; Kraka, E. SSnet: A Deep Learning Approach for Protein-Ligand Interaction Prediction, *Int. J. Mol. Sci.*, **2021**, 22, 1392.
[3] Sundararajan, M., Taly, A., & Yan, Q. Axiomatic attribution for deep networks, *International conference on machine learning. PMLR*, 2017, p. 3319-3328.

Lead optimization through active learning using free energy perturbation

Kairi Furui[1]
furui@li.c.titech.ac.jp

Masahito Ohue[1]
ohue@c.titech.ac.jp

1 Department of Computer Science, School of Computing, Tokyo Institute of Technology, 4259-G3-56 Nagatsuta-cho, Midori-ku, Yokohama City, Kanagawa 226-8503, Japan

Keywords: Lead optimization, Free energy perturbation, Active learning

Exploring lead compounds with potent inhibition from hit compounds is one of the challenging tasks in drug discovery. Free energy perturbation (FEP) calculations have attracted attention as a lead optimization method that can predict binding affinities with extremely high accuracy [1]. However, FEP calculation requires huge computational time compared to other quantitative structure-activity relationships (QSAR) and molecular docking. Therefore, it is essential to explore derivatives in the possible chemical space more efficiently with fewer trials to take advantage of FEP for lead optimization.

Recently, active learning-based lead optimization based on relative binding free energy calculations was recently proposed [2]. Still, the effectiveness of this approach in practice has not been comprehensively investigated. This study aims to establish a workflow for FEP calculations based on active learning to explore new compounds with desirable binding free energies in fewer trials.

[1] Wang, L.; *et al.*, Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field, *Journal of the American Chemical Society*, **2015**, *137*(7), 2695-2703.

[2] Guesev, F.; *et al.*, Active learning guided drug design lead optimization based on relative binding free energy modeling, *ChemRxiv* (preprint), **2022**, 10.26434/chemrxiv-2022-krs1t.

Prediction of Class A GPCR-Compound interactions by deep learning focusing on ligand binding site protein sequences

Haruki Yamane[1] yamane@cb.cs.titech.ac.jp
Takashi Ishida[1] ishida@c.titech.ac.jp

[1] Tokyo Institute of Technology

Keywords: Machine Learning, Activity Prediction

G protein-coupled receptors (GPCRs) are classified into six classes, A-F, based on physiological and structural characteristics [1]. Class A GPCRs, which constitute the largest GPCR class, have a ligand-binding site at the helix. The identification of the ligands for these GPCRs will lead to the discovery of new drugs discovery and functional roles. In recent years, deep learning has been widely used to predict protein-compound interactions, and many studies have used the full length of the protein sequence as input for various protein families [2-4]. However, if the helix region is found to be important for compound binding, as in the case of Class A GPCRs, the terminal region and intracellular and extracellular loop regions may be encoded as noise. In this study,
we created a compound interaction dataset focusing on Class A GPCRs and proposed a model for predicting protein-compound interactions that focuses on the amino acid sequence of the helix portion of Class A GPCRs.

[1] Secker A. Halling-Brown M. et al. Davies, M.N. GPCRtree: online hierarchical classification of gpcr func- tion. BMC Res Notes, Vol. 1, p. 67, 2008.
[2] Hakime O¨ztu¨rk, Arzucan O¨zgu¨r, and Elif Ozkirimli. DeepDTA: deep drug–target binding affinity prediction. Bioinformatics, Vol. 34, No. 17, pp. i821–i829, 09 2018.
[3] Masashi Tsubaki, Kentaro Tomii, and Jun Sese. Com- pound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. Bioinformatics, Vol. 35, No. 2, pp. 309–318, 07 2018.
[4] Lifan Chen, Xiaoqin Tan, Dingyan Wang, Feisheng Zhong, Xiaohong Liu, Tianbiao Yang, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, and Mingyue Zheng. TransformerCPI: improving compound–protein interac- Vol.2022-BIO-69 No.3 2022/3/10 tion prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. Bioinformatics, Vol. 36, No. 16, pp. 4406–4414, 05 2020.

AI-AAM. Lead-to-lead scaffold hopping for drug repositioning

Jun Nakabayashi[1] jun.nakabayashi@fujifilm.com
Kyosuke Tsumura[1] kyosuke.tsumura@fujifilm.com
Shino Ohira[1] shino.ohira@fujifilm.com
Kaoru Iwato[1] kaoru.iwato@fujifilm.com
Mao Tanabe[2] mtanabe@nibiohn.go.jp
Yuko Usami[2] y-usami@nibiohn.go.jp
Ryuichi Sakate[2] rsakate@nibiohn.go.jp
Tomonori Kimura[2] t-kimura@nibiohn.go.jp

[1] Analysis Technology Center, FUJIFILM Corporation, 210 Nakanuma, Minamiashigara-shi, Kanagawa 250-0193, Japan
[2] National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8, Saito-Asagi, Ibaraki City, Osaka 567-0085, Japan

Keywords: AI Drug Discovery, LBDD, Rare Disease, DDrare, AI-AAM

Drug development for rare and intractable diseases is challenging and drug repositioning (DR) has attracted attention to reduce the time and cost of the process [1]. While DR based on biological networks became common recent years, we invented a new way for DR by focusing on the efficient lead-to-lead scaffold hopping using intermolecular interactions.

We previously reported AI-Amino Acid Mapping (AI-AAM), the *in-silico* hit-to-hit scaffold hopping method in which the same order of the binding affinity as the original ligand was maintained when a new compound had the high AAM similarity to the active ligand [2]. Now, we applied the method for lead-to-lead hopping for DR. Using DDrare [3], we selected SYK, a drug target for IgA nephropathy, etc. Starting from a known SYK inhibitor candidate BIIB-057 [4], we aimed to discover other lead ligands from 44,503 compounds composed of biologically active compounds including approved drugs and compounds in clinical phase. We found a known anticancer drug candidate XC608 [5] with the highest AAM similarity score to BIIB-057 and confirmed almost the same inhibition to SYK as the original ligand (Fig.). This implies that XC608 is a repositionable drug for IgA nephropathy, etc.

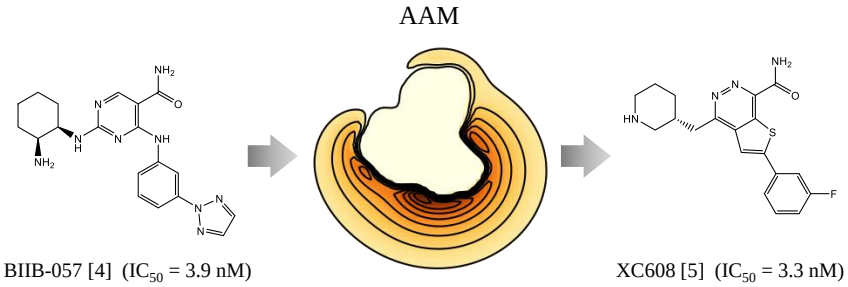


Fig. Scaffold hopping from BIIB-057 (left) to XC608 (right) via AAM (center)

[1] R. Sakate, T. Kimura. Drug repositioning trends in rare and intractable diseases. *Drug discov. today* 2022, 7, 1789.
[2] K. Tsumura.; *et al*., A newly developed method based on AI-oriented amino acid interaction mapping (AI-AAM) for efficient virtual scaffold hopping, *CBI Annual Meeting 2018*, P2-02.
[3] DDrare: Database of Drug Development for Rare Diseases, https://ddrare.nibiohn.go.jp/
[4] WO 2009/136995.
[5] WO 2011/035077.

P08-15

Understanding Molecular Mechanism of Drug Off-Target Effect in Tyrosine-Protein Kinase LCK Using Molecular Dynamics Simulation

Hidetoshi Tage[1] tage.hidetosi.86v@kyoto-u.ac.jp
Shigeyuki Matsumoto[1] matsumoto.shigeyuki.4z@kyoto-u.ac.jp
Hiroaki Iwata[1] iwata.hiroaki.3r@kyoto-u.ac.jp
Yuta Isaka[2] isaka@riken.jp
Biao Ma[2] biao-ma@riken.jp
Mitsugu Araki[1] araki.mitsugu.6w@kyoto-u.ac.jp
Atsushi Tokuhisa[2] tokuhisa@riken.jp
Ryo Kanada[2] ryo.kanada@riken.jp
Yasushi Okuno[1, 2] okuno.yasushi.4c@kyoto-u.ac.jp

[1] Graduate School of Medicine and Faculty of Medicine, Kyoto University, 53 Kawahara-cho, Shogoin Sakyo-ku, Kyoto, 606-8507, Japan
[2] Riken Center for Computational Science, 7-1-26 Minatojima-minamimachi, Chuo-ku, Kobe, 650-0047, Japan

Keywords: Molecular dynamics, Ligand docking, Adverse drug reaction, supercomputer “Fugaku”, Off-target

An unexpected interaction of a drug with a protein other than its target, i.e., off-target interaction, causes adverse drug reactions (ADRs), significantly hampering effective clinical treatment. Accordingly, the off-target interactions with the ADRs-associated proteins are investigated in the early stage of drug development. Nevertheless, the molecular mechanisms, including the binding sites and the binding manners on the off-target protein, have not been intensively investigated, limiting the establishment of the strategy to avoid the ADRs.

Here, using molecular dynamics (MD) simulation on the state-of-the-art supercomputer “Fugaku” equipped with abundant nodes, we investigated the molecular mechanism of the off-target effect attributed to lymphocyte-specific protein tyrosine kinase (LCK), a well-known off-target protein causing immunodeficiency. We selected 20 drugs, which are known to cause immunodeficiency or the other ADRs, and individually explored the binding events occurring on the overall molecular surface of LCK using an MD simulation technique, ColDock [1]. The comprehensive analysis of the obtained MD data identified several binding regions characteristic of the immunodeficiency-associated drugs. In this presentation, we will discuss the molecular mechanism underlying the off-target effect by investigating the relationship between their bound states and the functional deficiency found in LCK [2] from the point of view of dynamic behaviors.

[1] Takemura K.; Sato C.; Kitao A., ColDock: Concentrated Ligand Docking with All-Atom Molecular Dynamics Simulation, *J Phys Chem B*, **2018**, *122*, 7191–7200.
[2] Hauck F.; Randriamampita C.; Martin E.; Gerart S.; Lambert N.; Lim A.; et al., Primary T-cell immunodeficiency with immunodysregulation caused by autosomal recessive LCK deficiency, *J Allergy Clin Immunol*, **2012**, *130*, 1144–1152.e11.

P08-16

Molecular Generation for Protein-Protein Interaction Inhibitor Design focusing on Physicochemical Properties

Yuki Kojima[1]
kojima@li.c.titech.ac.jp

Takatsugu Kosugi[1]
kosugi@li.c.titech.ac.jp

Masahito Ohue[1]
ohue@c.titech.ac.jp

[1] Tokyo Institute of Technology, G3 building 10F, 4259-G3-56 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8503, Japan

Keywords: protein-protein interaction (PPI), molecular generation model, machine learning

Protein-protein interactions (PPIs) are essential targets in drug discovery because of their association with various diseases. PPI-targeting modulators have very different physicochemical properties from conventional small molecule oral drugs, such as the "Rule-of-Five" (RO5). Therefore, it has been difficult to efficiently generate and design PPI inhibitors using conventional methods, including molecular generation models. In this study, we propose a molecular generation model based on deep reinforcement learning, specialized for generating PPI inhibitor candidates. By improving the scoring function of the existing molecular generation model for small molecules, we have made it possible to generate compounds that are likely to inhibit PPIs. For future use in a biochemical assay, we also try to build a virtual library consisting of generated compounds by the proposed method. The compounds in this library are considered more suitable for recent PPI inhibitor design than those in the existing PPI-oriented library.

Multi-objective Bayesian optimization for antimicrobial peptides design using non-natural amino acids

Yuki Murakami[1]
W225431d@yokohama-cu.ac.jp

Shoichi Ishida[1]
lshida.sho.nm@yokohama-cu.ac.jp

Yosuke Demizu[1,2]
demizu@nihs.go.jp

Kei Terayama[1]
terayama@yokohama-cu.ac.jp

[1] Yokohama City Univ., 1-7-29 Suehirocho, Tsurumi-ku, Yokohama-shi, Kanagawa, 230-0045, Japan
[2] National Institute of Health Science, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki-shi, Kanagawa, 210-9501, Japan

Keywords: Bayesian optimization, Gaussian regression, Machine learning, Antimicrobial peptides, non-natural amino acids

Multidrug-resistant bacteria that have acquired resistance to many antibiotics have become a severe problem. Antimicrobial peptides (AMPs) have attracted attention as a strategy against multidrug-resistant bacteria because they act directly on cell membranes. On the other hand, AMPs screening and identification by experimental techniques are laborious and time-consuming tasks since AMPs design is often based on researchers' intuition and experience. To efficiently explore AMPs, several studies have developed machine learning models to predict antimicrobial activities of AMPs from single amino acid letter sequence information [1-3]. Recently, non-natural amino acids have been used in AMPs design for metabolic and structural stability [4]. However, most of these prediction models do not handle non-natural amino acids and efficient strategies for a design of experiments based on machine learning models have not yet been developed to design AMPs with non-amino acids.

Here, we proposed a method using multi-objective Bayesian optimization [5] to search for the optimal structure with a small number of experiments for AMP design by considering non-natural amino acids. As the input feature of surrogate models for Bayesian optimization, we used Morgan and MACCS fingerprints. The objective variables were toxicity to red blood cells and antibacterial activities against six bacteria including multi-drug resistant bacteria. To solve the multi-objective optimization problem, we calculated the probability of improvement (PI) for each objective and used the product of all the PIs scores as an acquisition function. We generated candidate AMPs with at least two mutations based on the base sequence, calculated their scores exhaustively, and selected AMPs with these better scores. Seven AMPs selected by this calculation were synthesized by the solid-phase synthesis method and evaluated by the minimum inhibitory concentration method for further optimization. At this conference, we report both the proposed method and the experimental results.

[1] Alice Capecchi; *et al.*, *Machine learning designs non-hemolytic antimicrobial peptides*, *Chem. Sci.,* **2021**, *12*, 9221-9232
[2] Supatcha Lertampaiporn; *et al.*, *Ensemble-amppred: Robust amp prediction and recognition using the ensemble learning method with a new hybrid feature for differentiating amps*, *Genes*, **2021**, *12*, 1-17
[3] Andrejs Tucs; *et al.*, *Generating ampicillin-level antimicrobial peptides with activity-aware generative adversarial networks, ACS Omega,* **2020**, *5*, 22847-22851.
[4] Motoharu Hirano; *et al.*, *Development of Antimicrobial Stapled Peptides Based on Magainin 2 Sequence*, *Molecules,* **2021**, *26,* 444
[5] Yuichi Motoyama; *et al., Bayesian optimization package: PHYSBO, Comput. Phys. Commun.*, **2022**, *278*, 108405

Automated design of novel drugs using patent literature

András Strácz [1,2]
astracz@chemaxon.com

Jenny Viklund [1]
jviklund@chemaxon.com

[1] ChemAxon Ltd., Váci út 133., Budapest, 1138, Hungary
[2] Patcore. Inc, Sapia Tower 26F, 1-7-12 Marunouchi, Chiyoda-ku, Tokyo, 100-0005, Japan

Keywords: Patent literature, MMP method, Designing novel drugs

Predictive models and data-driven generation of new compounds are key capabilities when designing and optimizing small molecule drugs. This presentation will show how knowledge from previous drug discovery projects can be utilized in rational drug design via a generic Matched Molecular Pairs (MMP) model [1] based on exemplified structures in SureChEMBL [2] (Figure 1). To put the model into context, using Design Hub [3], an example scientific hypothesis is created, and new compounds will be automatically generated using the MMP model and prioritized using a predicted parameter. We will also show how to securely share selected compounds within your project with an external CRO for synthesis.

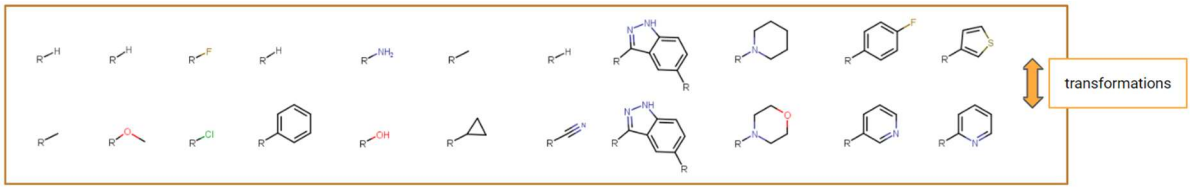


Figure 1: Examples of the most common small molecule drug transformations in SureCHEMBL

- Poster presentation
- Patcore. Inc, represents Chemaxon. This poster presentation will be read by the member of Patcore, Inc. at the CBI.

[1] Hussain and Rea. Journal of Chemical Information and Modeling, 2010, 50 (3), 339-348
[2] Papadatos, George, et al. Nucleic acids research 44.D1 (2016): D1220-D1228. DOI: 10.1093/nar/gkv1253
[3] https://chemaxon.com/products/design-hub

Automated hetero shuffling combined with FMO calculations

Sundaram Arulmozhiraja[1]
arulmozhiraja.sb@om.asahi-kasei.co.jp

Kazufumi Ohkawa[1]
okawa.kf@om.asahi-kasei.co.jp

[1] Asahi Kasei Pharma, 632-1 Mifuku, Izunokuni, Shizuoka 410-2321, Japan

Keywords: Hetero-shuffling, Fragment molecular orbital calculation, Protein-ligand interactions

Fragment molecular orbital (FMO), an accurate quantum mechanics-based theory [1-3], calculations play the leading role in estimating protein-ligand interactions, and the results derived can be utilized to support structure-based drug design [4]. In fact, several studies concluded that interaction energies obtained using FMO calculations correlate with the efficacy of drugs [5]. Hetero shuffling is an approach manipulating atoms of the scaffold to generate potential new molecules and ultimately used them in drug designing to improve physiochemical and biological activities of hit compounds. To assist and speed up drug discovery projects, we herein developed a code, to utilize both hetero shuffling and FMO calculations, to automatically generate hetero-shuffled new molecules bound to the parent protein structure. This code then prepares FMO input files for all the newly generated complex structures.

Overall, we develop a methodology for practical hetero-shuffle-based molecule generation combined with FMO calculations.

[1] Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayashi, M. Fragment molecular orbital method: an approximate computational method for large molecules, *Chem. Phys. Lett.,* **1999**, *313,* 701–706.

[2] Fedorov, DG.; Kitaura, K. Extending the power of quantum chemistry to large systems with the fragment orbital method, *J. Phys. Chem. A*, **2007**, *111*, 6904–6914.

[3] Fedorov, DG.; Nagata, T.; Kitaura, K. Exploring chemistry with fragment molecular orbital method, *Phys. Chem. Chem. Phys.*, **2012**, *14*, 7562–7577.

[4] For example, Monteleone, S.; *et al.*, Hotspot identification and drug design of protein-protein interaction modulators using fragment molecular orbital method, *J. Chem. Inf. Model*, **2022**, *62*, 3784–3799.

[5] For example, Arulmozhiraja, S.; *et al.*, Comparative binding analysis of dipeptidyl peptidase IV (DPP-4) with antidiabetic drugs – An ab initio fragment molecular orbital study, *PLoS ONE*, **2016**, *11*, e0166275.

P08-20

In Silico Identification of Natural Product Compounds as a Selective Endoplasmic Reticulum α-Glucosidase II Inhibitor using Pharmacophore-based Virtual Screening and Molecular Docking Studies

Mochammad Arfin Fardiansyah Nasution[1,2]
mochammad.arfin@protein.osaka-u.ac.jp

Chioko Nagao[1,2]
c_nagao@protein.osaka-u.ac.jp

Kenji Mizuguchi[1,2*]
kenji@protein.osaka-u.ac.jp

[1] Laboratory for Computational Biology, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
[2] National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki, Osaka, 567-0085, Japan

Endoplasmic reticulum (ER) α-glucosidase II (GluII) is a host-cell enzyme that is required by various enveloped viruses in the replication and morphogenesis processes of their life cycle[1]. The selective inhibition of GluII enzyme, without affecting other glucosidases such as ER α-glucosidase I (ER GluI) and lysosomal α-glucosidase (GAA), could lead to an effective alternative approach in treating various viral infections in the human body[2]. Natural products (NP) have been long considered as an invaluable source of drugs with diverse bioactivities, including antiviral activities[3]. Herein, 1,073,049 NP compounds from ten databases were subjected into pharmacophore modelling and molecular docking studies to determine the most potent NP compound that can selectively inhibit GluII enzyme. Pharmacophore models were created from the existing GluII crystal structure[4,5], followed by induced-fit based-docking studies. Additionally, a series of ADME-Tox screening tests were performed to determine the pharmacological and pharmacokinetic properties of the selected NP compounds. From these studies, two NP ligands, namely SID461843168 and UNPD170308 showed the highest binding affinities with the GluII enzyme, with a $\Delta G_{binding}$ score of -11.17 kcal/mol and -11.03 kcal/mol, respectively. Furthermore, SID461843168 ligand formed an H-bond interaction with both catalytic site residues Asp564 and Asp640, while UNPD170308 only formed an H-bond with Asp640. Additionally, these compounds demonstrate better affinities with GluII than GAA and GluI, according to the induced-fit docking results. Further computational analysis will be conducted on these compounds to examine the binding stability using molecular dynamics simulations.

Keywords: Endoplasmic reticulum α-glucosidase II, antiviral, natural product compounds, pharmacophore-based virtual screening, induced-fit docking.

Reference:

[1] Williams, S. J.; Goddard-Borger, E. D. α-Glucosidase Inhibitors as Host-Directed Antiviral Agents with Potential for the Treatment of COVID-19. *Biochem. Soc. Trans*. **2020**, 48 (3),
[2] Kiappes, J. L.; *et al.*, ToP-DNJ, a Selective Inhibitor of Endoplasmic Reticulum α-Glucosidase II Exhibiting Antiflaviviral Activity. *ACS Chem. Biol*. **2018**, 13 (1), 60–65.
[3] Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod*. **2020**, 83 (3), 770–803.
[4] Caputo, A. T.; *et al.*, Structures of Mammalian ER α-Glucosidase II Capture the Binding Modes of Broad-Spectrum Iminosugar Antivirals. *Proc. Natl. Acad. Sci. U. S. A*. **2016**, 113 (32), E4630–E4638.
[5] Karade, S. S.; *et al.*, N-Substituted Valiolamine Derivatives as Potent Inhibitors of Endoplasmic Reticulum α-Glucosidases I and II with Antiviral Activity. *J. Med. Chem*. **2021**.

P10-01

Gentamicin-induced hearing loss: a retrospective study using the Food and Drug Administration Adverse Event Reporting System and a drug−gene network analysis using the DIseAse MOdule Detection algorithm

Mizuki Tanaka[1]
165050@gifu-pu.ac.jp

Kiyoka Matsumoto[1]
165079@gifu-pu.ac.jp

Riko Satake[1]
175058@gifu-pu.ac.jp

Yu Yoshida[1]
175127@gifu-pu.ac.jp

Misaki Inoue[1]
175016@gifu-pu.ac.jp

Shiori Hasegawa[1]
135056@gifu-pu.ac.jp

Takaaki Suzuki[1,2]
t07tor5489@hi3.enjoy.ne.jp

Mari Iwata[1,3]
m-iwata@kifune.info

Kazuhiro Iguchi[4]
iguchi@gifu-pu.ac.jp

Mitsuhiro Nakamura[1]
mnakamura@gifu-pu.ac.jp

[1] Laboratory of Drug Informatics, Gifu Pharmaceutical University, 1-25-4, Daigaku-Nishi, Gifu, 501-1196, Japan
[2] Gifu Prefectural Government, 2-1-1 Yabutaminami, Gifu, 500-8570, Japan
[3] Kifune Pharmacy, 2-23-2 Hasuike, Yanaizu-cho, Gifu, 501-6103, Japan
[4] Laboratory of Community Pharmacy, Gifu Pharmaceutical University, 1-25-4, Daigaku-Nishi, Gifu, 501-1196, Japan

Keywords: hearing loss, gentamicin, drug−gene interactions, disease module detection algorithm, Food and Drug Administration adverse event reporting system

Hearing loss has a significant effect on patients' quality of life. The objectives of the study were to evaluate the relationship between gentamicin (GEN) and hearing loss using the Food and Drug Administration Adverse Event Reporting system (FAERS) database and elucidate the potential toxicological mechanism of GEN-induced hearing loss through a drug–gene network analysis. We extracted GEN-associated genes (seed genes) and analyzed drug−gene interactions using the DIseAse MOdule Detection (DIAMOnD) algorithm. The FAERS database had 10 745 188 reports from January 2004 to June 2019. The number of reports for drug induced hearing loss was 40 073. The top three adverse events reported were hypoacusis (16 557 cases), deafness (13 140 cases), and deafness unilateral (3177 cases). The lower limits of the 95% confidence intervals of the reporting odd ratios for streptomycin, tobramycin, GEN, kanamycin, neomycin, amikacin, and netilmicin were greater than one. The DIAMOnD algorithm presents a systematic analysis of the connectivity patterns of disease proteins and can be used to determine the predictive topological property using a Python script. For DIAMOnD analysis, we retrieved the human protein−protein interaction dataset from the BioGRID database. To identify the GEN-associated module constructed by this algorithm, we used the human protein−protein interaction dataset retrieved from the BioGRID database and the seed genes identified using PharmGKB and DGIdb. We identified 60 GEN-associated genes using the DIAMOnD algorithm. Several GEN-associated genes in the DIAMOnD algorithm were highly enriched in "Ras signaling pathway," "focal adhesion," "MAPK signaling pathway," "regulation of actin cytoskeleton," "oxidative phosphorylation," and "ECM-receptor interaction." Our analysis demonstrated an association between several AGs and hearing loss using the FAERS database. Drug−gene network analysis demonstrated that GEN may be associated with oxidative phosphorylation-associated genes and integrin genes, which may be associated with hearing loss.

[1] Tanaka M. et al. *HELIYON,* 2021.7, e07429, doi: 10.1016/j.heliyon.2021.e07429

P10-02

Comprehensive analysis of the drugs that may induce severe cutaneous adverse reactions using the adverse drug event report database

Risa Tonegawa[1]
m226235@std.my-pharm.ac.jp

Mizuho Asada[1]
mizuho@my-pharm.ac.jp

Yoshihiro Uesawa[1]
uesawa@my-pharm.ac.jp

[1] Department of Medical Molecular Informatics, Meiji Pharmaceutical University, 2-522-1 Noshio, Kiyose, Tokyo, 204-8588, Japan

Keywords: JADER, adverse effect, Severe cutaneous adverse reactions, Stevens–Johnson syndrome, toxic epidermal necrolysis

Purpose: Severe cutaneous adverse reactions (SCAR) like Stevens–Johnson syndrome and toxic epidermal necrolysis are grave conditions that often have prognoses. Such conditions are characterized by necrotic and epidermal lesions, and are frequently accompanied by high fever, mucous membrane rash, blistering, and epidermal exfoliation. Moreover, few studies of medications known to induce SCAR exist. We analyzed medications associated with SCAR induction using the Japanese Adverse Drug Event Report (JADER) published by the Pharmaceuticals and Medical Device Agency.

Method: We utilized 14 preferred terms meaning SCAR based on standardized MedDRA queries. We examined the tables that included JADER drug information, adverse events, and patient demographics collected from April 2004 to March 2022. Furthermore, reported instances of SCAR onset were investigated for all listed drugs. Finally, the reporting odds ratios (ROR) and Fisher's exact test P-values were determined for all the identified drugs.

Results: We identified 542 potential SCAR-inducing drugs (P-value <0.05 and ROR >1), 45 of which demonstrated high statistical significance (P-value <10−50 and ROR >2). These included medications for acid-related disorders in addition to various systemic anti-microbial, anti-epileptic, anti-inflammatory, and anti-rheumatic medications.

Discussion: We developed a comprehensive list, including previously identified medications such as carbamazepine and loxoprofen, which appear to be associated with SCAR development. These findings may be helpful for the early detection and treatment of SCAR.

P10-03

Development of Prediction Model for Adverse Events by Using Spontaneous Report Database and Chemical Structures

Keiko Ogawa[1] ogawak23@fc.ritsumei.ac.jp
Hiroaki Iwata[2] iwata.hiroaki.3r@kyoto-u.ac.jp
Ryosuke Kojima[2] kojima.ryosuke.8e@kyoto-u.ac.jp
Yasushi Okuno[2] okuno.yasushi.4c@kyoto-u.ac.jp
Rumiko Hosoki[1] r3hosoki@fc.ritsumei.ac.jp

[1] Ritsumeikan Univ., 1-1-1 Noji-Higashi, Kusatsu, Shiga, 525-8577, Japan
[2] Graduate School of Medicine, Kyoto Univ., 53 Kawahara-cho, Syogoin, Sakyo-ku, Kyoto, 606-8507, Japan

Keywords: spontaneous report database, FAERS, adverse event, chemical structure, graph convolutional networks (GCN)

The occurrence of adverse events caused by drugs can interrupt drug therapy. There is a need to the prevention of adverse events by expecting their occurrence at an early stage of drug development. To date, various prediction models for specific adverse events have been developed from diverse information such as chemical structures, aiming to predict the occurrence and the probable factors of adverse reactions [1-2]. On the other hand, it is still unclear what kind of chemical structure of a drug affects the occurrence of any adverse events.

Recently, an approach using graph convolutional networks (GCN) has been successfully applied to chemical information. GCN can represent the entire chemical structures of compounds by recognizing compounds as graphs and convolving the information of not only neighboring atoms but also distant atoms. In this study, we tried to develop a multi-task model for predicting adverse events from the chemical structures of drugs using kGCN [3], a graph-based deep learning framework.

To construct the training dataset, we used the Czeek V dataset (obtained on March 2019) curated from FDA Adverse Event Reporting System (FAERS), a spontaneous reporting database maintained by the U.S. Food and Drug Administration (FDA). First, we performed data cleaning, which were removing combination drugs and polymeric drugs, and eliminating inaccurate reports, then defined those with a PRR $\geq$ 2.0 as having a risk of adverse event occurrence. As a result of data construction, we extracted 9,621,930 reports containing 2932 drugs and 14027 adverse events. From the extracted data, a multi-task model was constructed by kGCN for 177 adverse events which are connected with 300 or more drugs. In this presentation, the model construction, and the performance of constructed GCN model will be explained together with which adverse events were predictable and which were difficult to predict from the structural information. The influence of the chemical structures on the occurrence of adverse events will be also discussed.

[1] Ambe, K.; *et al.*, In Silico Approach to Predict Severe Cutaneous Adverse Reactions Using the Japanese Adverse Drug Event Report Database, *Clinical and Translational Science*, 2021, 756-763.
[2] Gadaleta, D., *et al.*; Prediction of the Neurotoxic Potential of Chemicals Based on Modelling of Molecular Initiating Events Upstream of the Adverse Outcome Pathways of (Developmental) Neurotoxicity, *International Journal of Molecular Science*s, 2022, 23, 3053.
[3] Kojima, K.; *et al.*, kGCN: a graph-based deep learning framework for chemical structures, *Journal of Cheminformatics*, 2020, 12, 32.

Analysis of Ion Transport Properties in Artificial DNA Channels with Molecular Dynamics

Jun Takahashi[1,2]
jun.takahashi.r2@dc.tohoku.ac.jp

Akihiro Nakamura[1,2]
nakamura.akihiro.p1@dc.tohoku.ac.jp

Ibuki Kawamata[1]
ibuki.kawamata@tohoku.ac.jp

Yusuke Sato[3]
ysato@ics.kyutech.ac.jp

Takashi Tokumasu[2]
tokumasu@ifs.tohoku.ac.jp

Takuya Mabuchi[2,4]
mabushi@tohoku.ac.jp

[1] Tohoku Univ., 2-1-1, Katahira, Aoba-ku, Sendai 980-8577, Japan
[2] Institute of Fluid Science, Tohoku Univ., 2-1-1, Katahira, Aoba-ku, Sendai 980-8577, Japan
[3] Kyushu Institute of Technology, 680-4 Kawazu, lizuka, 820-8502, Japan
[4] Frontier Research Institute of Interdisciplinary Sciences, Tohoku Univ., 2-1-1, Katahira, Aoba-ku, Sendai 980-8577, Japan

Keywords: Molecular dynamics, DNA nanotechnology, Artificial molecular sensor

The transport of ions across lipid bilayers is done selectively by membrane protein channels, which control various reactions in our body. The development of artificial ion channels by biomimicking the characteristics of natural membrane protein channels leads to their application as artificial nano-biological sensors. The DNA nanotechnology enables the creation of arbitrary three-dimensional structures and multimolecular modification and facilitates the fabrication and modification of structures by the precise sequence designs at the single nucleotide level. Various artificial DNA channels have been developed to achieve the selective transport of molecules with different particle sizes[1]. However, the artificial DNA channels capable of more selective transport, such as transport of specific ions in specific proportions, have not been realized. Moreover, experimental methods have found it difficult to discuss the details of ion transport phenomena at the nanoscale. Therefore, the objective of this study is to elucidate the effects of the pore diameter and the modification of internal pores on ion transport properties using molecular dynamics simulations. For the models with different pore sizes, three different DNA channel models (3-helix DNA channel[2], 4-helix DNA channel[3], 6-helix DNA channel[4]) corresponding to 3 nm, 8 nm, and 10 nm, are used. For the modification models, the ethyl group modification, which makes partially hydrophobic inside the pores, is employed in the present study.

[1] Thomsen, R. P.; Malle, M. G.; Okholm, A. H.; Krishnan, S.; Bohr, S. S. R.; Sørensen, R. S.; Ries, O.; Vogel, S.; Simmel, F. C.; Hatzakis, N. S.; Kjems, J. A Large Size-Selective DNA Nanopore with Sensing Applications. *Nat. Commun*., **2019**, 10 (1).
[2] Park, S. H.; Barish, R.; Li, H.; Reif, J. H.; Finkelstein, G.; Yan, H.; LaBean, T. H. Three-Helix Bundle DNA Tiles Self-Assemble into 2D Lattice or 1D Templates for Silver Nanowires. *Nano Lett*., **2005**, 5 (4), 693–696.
[3] Rangnekar, A.; Gothelf, K. V.; Labean, T. H. Design and Synthesis of DNA Four-Helix Bundles. *Nanotechnology,* **2011**, 22 (23).
[4] Yoo, J.; Aksimentiev, A. Molecular Dynamics of Membrane-Spanning DNA Channels: Conductance Mechanism, Electro-Osmotic Transport, and Mechanical Gating. *J. Phys. Chem. Lett*., **2015**, 6 (23), 4680–4687.

P11-02

Luminescent control of DNA-scaffolded BRET system using strand displacement reaction

Fumiaki Takano
k179263@kansai-u.ac.jp

Yuki Minamide
k119802@kansai-u.ac.jp

Tomoya Niki
k349342@kansai-u.ac.jp

Shiori Tabana
k053995@kansai-u.ac.jp

Akinori Kuzuya
kuzuya@kansai-u.ac.jp

Department of Chemistry and Material Engineering Kansai University, 3-2-35 Yamate-cho, Suita-Shi, Osaka, Japan

Keywords: DNA, Bioluminescence resonance energy transfer (BRET), Fluorescence resonance E energy transfer (FRET)

Fluorescent dyes are used as functional sensors to visualize the physiological state of cells. In microscopic observation methods that label with fluorescent dyes, there are problems such as fading of the fluorescent dyes and cytotoxicity due to irradiation with strong excitation light. These problems can be avoided by using luciferase as an excitation source to convert chemical energy into light energy without the need for excitation light.

We therefore focused on luciferase called NanoLuc [1]. This is a modified version of luciferase from deep-sea shrimp, and NanoLuc, in which NanoLuc is split into two parts, is already available. It is composed of a small fragment of 11 amino acids and a large fragment. Although they have no luminescence activity by themselves, they are reconstituted with high affinity to become NanoLuc and regain luminescent activity. However, the wavelength of luciferase is fixed depending on the substrate. It usually a fusion protein of luciferase and fluorescent proteins through an appropriate linker is formed to modulate the wavelength [2]. While this bioluminescence resonance energy transfer (BRET) system allows efficient energy transfer, it requires optimization of the linker sequence and length through labor-intensive genetic modification experiments.

In this study, we developed a BRET system that uses DNA as a scaffold between luciferase and fluorescent dyes. White light emission was also achieved by mixing RGB emission. This system allows the user to select multi-colored luminescence by simply adjusting the concentrations of the three DNA complexes. And the color tone changes were observed in real time by using hybridization and strand displacement reactions.

[1] A. S. Dixon, *et al., Anal. Chem.* **2016**, *11*, 400-408
[2] S. Sihuai, *et al., Int. J. Mol. Sci*, **2016**, *17*, 1704.

P11-03

Long-range and multi-step intramolecular energy transfer by BRET/FRET system

Yuki Minamide k119802@kansai-u.ac.jp
Fumiaki Takano k179263@kansai-u.ac.jp
Koichi Tanimoto k769965@keio.ac.jp
Akinori Kuzuya kuzuya@kansai-u.ac.jp

Department of Chemistry and Material Engineering Kansai University, 3–2–35 Yamate–cho, Suita–shi, Osaka, Japan

Keywords: DNA, Bioluminescence resonance energy transfer (BRET)

Plants have the complex between the photosynthetic reaction center and the light-harvesting antenna core proteins to use scarce solar energy efficiently. This complex has hundreds of chlorophyll. The received light energy is passed from molecule to molecule to the reaction center by this complex. This energy transfer is difficult to imitate. Because all fluorescent dye are excited at the same time with bulk excitation source, we cannot observe energy transfer directly.

In this study, we focused on Bioluminescence Resonance Energy Transfer (BRET). BRET is a phenomenon in which the energy used by a luminescent substrate to emit light is transferred to a fluorescent substance in the vicinity. It is necessary to link the luminescent and fluorescent proteins with an appropriate linker to induce BRET [1]. This BRET system requires optimization of the linker sequence and length through labor-intensive genetic modification experiments while it allows efficient energy transfer. We have developed a DNA-linked BRET system by modifying one of the two complementary strands of DNA with NanoLuc luciferase [2], a luminescent protein, as a donor, and a fluorescent dye as an acceptor on the other strand of DNA. In this system, color can be freely changed by changing the type of fluorescent dye.

We attempted to directly observe energy transfer by modifying multiple fluorescent dyes within the DNA molecule that serves as the linker. First, an azide was modified on the C-terminal side of HiBiT, a split form of NanoLuc. The azide-modified HiBiT was conjugated with dibenzocyclooctyne (DBCO)-modified DNA via Cu-free click reaction. After hybridization with DNA inserted with fluorescent dye, NanoLuc was reconstituted and the emission spectrum was observed. As a result, it was observed that the greater the number of inserted fluorescent dyes, the more energy was transferred to the opposite side of the NanoLuc-modified end.

[1] S. Sihuai, *et al*., *Int. J. Mol. Sci*, **2016**, *17*, 1704.
[2] A. S. Dixon, *et al*., Anal. Chem. **2016**, *11*, 400-408.

P12-01

Predicting Excipient Modulated Viscosity of Monoclonal Antibody Formulations

Yoshirou Kimura[1]
kimura.yoshirou@molsis.co.jp

[1] MOLSIS Inc., 3-19-9, Hatchobori, Chuo-ku, Tokyo 104-0032, Japan

Keywords: Antibody medicine, Biologics, Molecular Property Prediction, Formulation, Excipient

When an antibody drug is administered by injection, a high-concentration solution of 100 mg/mL or more is required because of the limited amount administered at one time. But at high concentrations these formulations can cause an increase in viscosity. Subcutaneous injection typically requires viscosities below 15‐20 cP. Various excipient molecules, such as saccharides and amino acids, are used to adjust the viscosity. This experimental process is expensive and time-consuming.

Therefore, we attempted to develop an in silico method to predict the tendency of viscosity in the combination of antibodies and excipients in MOE [1]. First, we predicted the aggregation region of the antibodies by various methods similar to SAP [2]. Next, docking simulation of the excipient was performed over the entire regions of the antibody, and the region the the excipients bind to was identified from the result. The degree of agreement between the two regions was used as an index of viscosity. In this presentation, we will introduce the details of these methods and the results of comparison with experiments.

[1] *Molecular Operating Environment (MOE)*, 2020.09; Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, **2020**.
[2] Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L. Design of therapeutic proteins with enhanced stability. *Proceedings of the National Academy of Sciences*, **2009**, *106*, 11937-11942.

Specific knockdown of *KRAS* mutant gene using SNPD-siRNA repressed cell proliferation of pancreatic cancer cells in vitro and in vivo

Yoshiaki Kobayashi[1]
yoshiaki-kobayashi@g.ecc.u-tokyo.ac.jp

Yoshimasa Asano[1]
yoshimasa.asano@bs.s.u-tokyo.ac.jp

Atushi Sato[1]
atsushisato0728@gmail.com

Hiroaki Taniguchi[2]
h-tani@keio.jp

Kumiko Ui-Tei[1]
ktei@bs.s.u-tokyo.ac.jp

[1] Department of Biological Sciences, Graduate School of Science, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[2] Keio Cancer Center, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan

Keywords: SNPD-siRNA, *KRAS* mutation, pancreatic cancer

In RNA interference (RNAi), small interfering RNA (siRNA) functions to repress the expression of its target mRNA with perfect sequence complementarity. Thus, siRNAs are a powerful tool not only for functional genomics but also for therapeutic applications. To date, five siRNAs have been approved as medicines, but all these siRNAs repress their target genes regardless of the mutations. However, the nucleotide mutations in many oncogenes cause abnormal activation of cell growth. Specific types of mutations observed in *Kirsten rat sarcoma virus* (*KRAS*) gene are known to associate with tumor behavior such as accelerated tumor growth and metastasis in pancreatic or colorectal cancer patients. Furthermore, KRAS-deficient mice are embryonic lethal. Therefore, it is necessary to knock down the mutated *KRAS* gene alone without affecting the expression of its wild-type gene for therapeutic application of siRNA.

In this study, we developed a single nucleotide polymorphism-distinguishable siRNA (SNPD-siRNA) which specifically represses the mutated *KRAS* gene. Furthermore, we confirmed that it repressed cell proliferation of pancreatic cancer-derived cell lines with the mutant *KRAS* gene in vitro and in vivo. This is the first report of mismatch distinguishable siRNA and is expected to be applicable for the treatment of undruggable diseases, including hereditary diseases.

Development of Precise Genome Editing Technology by Cell Cycle Depending Activation of CRISPR-Cas9

Daisuke Matsumoto[1], dmatsumo@hiroshima-u.ac.jp
Kanae Kishi[1], d213023@hiroshima-u.ac.jp
Erina Matsugi[1], d223872@hiroshima-u.ac.jp
Yuto Inoue[1], b193576@hiroshima-u.ac.jp
Kiyomi Nigorikawa[1], knigo@hiroshima-u.ac.jp
Wataru Nomura[1] wnomura@hiroshima-u.ac.jp

[1] Hiroshima Univ.,1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan

Keywords: Genome Editing, CRISPR-Cas9, Homology directed repair, Cell cycle

The CRISPR-Cas9 system was originally discovered as part of the bacterial immune system against external DNA from organisms such as bacteriophages and plasmids [1]. It has become the predilected simplified genome editing tool, because it is easier and less expensive to construct various target libraries compared to other editing technologies such as ZFN and TALEN. The system has shown its efficient editing when used in human cells and model organisms. Although CRISPR technology is the most useful method for genome editing, off-target effects that cause unexpected mutations at pseudo-target DNA sequences could occur, similarly to those seen using as ZFN and TALEN. Thus, in addition to endeavor to increase the efficiency of precise editing at on-targets, off-target effects should be carefully addressed when genome editing tools are used, especially for clinical applications. Homology-directed Repair (HDR) is a DNA repair pathway based on template DNA having homologous arm sequences adjacent to the cleavage site. In HDR events, repair of target sequences introduces precise sequence change. Therefore, increasing the ratio of DNA repair through HDR over non-homologous end joining (NHEJ) is important for precise genome editing.

Recently, anti-CRISPR (Acr) protein inhibitors of the CRISPR-Cas9 system have been found [2]. The inhibitors, including AcrIIA4, were derived from bacteriophages targeting pathogenic bacterial strains. AcrIIA4 from Listeria monocytogenes prophage binds strongly to SpyCas9-sgRNA complexes, but the binding affinity to ApoSpyCas9 is lower. It has been reported that AcrIIA4 efficiently inhibits SpyCas9 activity in mammalian cells. Furthermore, the inhibition of SpyCas9 activity by AcrIIA4 reduces off-target editing [3]. Thus, when anti-CRISPR expression can be controlled by cell cycle, the activity of Cas9 endonuclease could also be controlled in the cells.

In this research, we fused the anti-CRISPR AcrIIA4 with the N-terminal region of human chromatin licensing and DNA replication factor 1 (hCdt1) for activation in the S/G2 phases and inactivation in the G1 phase. hCdt1 is degraded by ubiquitin-mediated proteolysis through the SCF^{Skp2} complex in the S/G2 phases. The cell cycle dependent Cas9 activation system was validated using SpyCas9 endonuclease and AcrIIA4 in the cells. As expected, the system displayed autonomous Cas9 activity switch dependent on the cell cycle and increased the preciseness of target gene editing [4].

[1] Barrangou, R.; *et al*., CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes, *Science*, **2007**, *315*, 1709-1712.
[2] Rauch, B. J.; *et al*., Inhibition of CRISPR-Cas9 with bacteriophage proteins, *Cell*, **2017**, *168*, 150–158.
[3] Shin, J.; *et al*., Disabling Cas9 by an anti-CRISPR DNA mimic, *Sci. Adv.*, **2017**, *3*, e1701620.
[4] Matsumoto D. ; *et al*., A cell cycle-dependent CRISPR-Cas9 activation system based on an anti-CRISPR protein shows improved genome editing accuracy, *Commun. Biol.*, **2020**, *3*, 601.