P1-01

The juxtamembrane region of TrkA kinase is critical for inhibitor selectivity

Noritaka Furuya[1,2], **Takaki Momose**[1], **Kenji Katsuno**[1], **Nobuhiko Fushimi**[1],
noritaka_furuya@pharm.kissei.co.jp

Hideyuki Muranaka[1], **Chiaki Handa**[1], **Tomonaga Ozawa**[1], **Takayoshi Kinoshita**[2]
kinotk@b.s.osakafu-u.ac.jp

[1] Kissei Pharmaceutical, 4365-1, Kashiwabara, Hotaka, Azumino City, Nagano Pref. 399-8304, Japan
[2] Graduate School of Science, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan

Keywords: TrkA kinase, Juxtamembrane, Crystal structure, Selective allosteric inhibitor

Although numerous crystal structures for protein kinases have been reported, many include only the kinase domain but not the juxtamembrane (JM) region, a critical activity-controlling segment of receptor tyrosine kinases (RTKs). In this study, we determined the X-ray crystal structure of the tropomyosin receptor kinase (Trk) A selective inhibitor **A1** complexed with the TrkA kinase domain and the JM region. This structure revealed that the unique inhibitor-binding pocket created by a novel JM configuration yields significant potency and high selectivity against TrkB and TrkC. Moreover, we validated the importance of the JM region for the potency of **A1** using in vitro assays. The introduction of moieties that interact with the JM region will be one of the most effective strategies for producing highly selective RTK inhibitors.

[1] Furuya, N.; Momose, T.; Katsuno, K.; Fushimi, N.; Muranaka, H.; Handa, C.; Ozawa, T.; Kinoshita, T., The juxtamembrane region of TrkA kinase is critical for inhibitor selectivity. Bioorg Med Chem Lett **2017**, 27 (5), 1233-1236.
[2] Benito-Gutiérrez, E.; Garcia-Fernàndez, J.; Comella, J. X., Origin and evolution of the Trk family of neurotrophic receptors. Mol Cell Neurosci **2006**, 31 (2), 179-92.
[3] Bertrand, T.; Kothe, M.; Liu, J.; Dupuy, A.; Rak, A.; Berne, P. F.; Davis, S.; Gladysheva, T.; Valtre, C.; Crenne, J. Y.; Mathieu, M., The crystal structures of TrkA and TrkB suggest key regions for achieving selective inhibition. J Mol Biol **2012**, 423 (3), 439-53.
[4] Chang, D. S.; Hsu, E.; Hottinger, D. G.; Cohen, S. P., Anti-nerve growth factor in pain management: current evidence. J Pain Res **2016**, 9, 373-83.
[5] Wang, T.; Lamb, M. L.; Scott, D. A.; Wang, H.; Block, M. H.; Lyne, P. D.; Lee, J. W.; Davies, A. M.; Zhang, H. J.; Zhu, Y.; Gu, F.; Han, Y.; Wang, B.; Mohr, P. J.; Kaus, R. J.; Josey, J. A.; Hoffmann, E.; Thress, K.; Macintyre, T.; Omer, C. A.; Yu, D., Identification of 4-aminopyrazolylpyrimidines as potent inhibitors of Trk kinases. J Med Chem **2008**, 51 (15), 4672-84.

P1-02

Calculation of intramolecular Reaction of Hydroxymethyl Rhodamine Derivatives for Development of Fluorescent Probes Based on Computational Chemistry

Ryo Tachibana[1]
ryo-tachibana@g.ecc.u-tokyo.ac.jp

Mako Kamiya[2,4]
mkamiya@m.u-tokyo.ac.jp

Satoshi Suzuki[3]
suzuki.satoshi.8v@kyoto-u.ac.jp

Keiji Morokuma[3]
morokuma.keiji.3a@kyoto-u.ac.jp

Yasuteru Urano[1,2,5]
uranokun@m.u-tokyo.ac.jp

[1] Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[2] Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[3] Fukui Institute for Fundamental Chemistry, Kyoto University, Takano-Nishihiraki-cho 34-4, Sakyou-ku, Kyoto 606-8103, Japan
[4] JST PRESTO
[5] AMED CREST

Keywords: Free Energy Calculation, Intramolecular Reaction, Acid-base Balance, Transition State Calculation, First-Shell Hydration

HMR (Hydroxymethyl Rhodamine) derivatives can exist in two forms (open, closed) by intramolecular spirocyclization. The proportion of two forms at equilibrium state changes according to pH. Various fluorescence probes have been developed using pK_{cycl}(pH value where the concentration of open form and that of closed form are the same) as an indicator[1][2]. In this research, we analyzed this intramolecular reaction computationally and found a powerful method to predict pK_{cycl} of HMR derivatives directly, without synthesizing any reference compounds.

We focused on closed-to-open reaction at acidic condition, and performed free energy calculation of local minima of HMR derivatives, considering first-shell hydration and proton transfer. Next, we searched each transition states which corresponded to calculated local minimum, and evaluated their contribution ratio to the reaction. As a result, we found energetically favorable chemical pathway, which dominated almost all of close-to-open reaction. We made a formula for calculating pK_{cycl}, based on equilibrium model of HMR derivatives including acid-base balance of their amino group and hydroxymethyl group. By calculating free energy gap throughout above-mentioned reaction path, we succeeded in reproducing pK_{cycl} of measured HMR derivatives and predicting proportion of unknown derivatives.

Moreover, we succeeded in developing new red fluorescence peptidase probes by applying pK_{cycl} prediction, which shows the possibility of accurate and efficient molecular design using calculation with careful analysis of solvent behavior including first-shell hydration.

[1] Sakabe, M.; Asanuma, D.; Kamiya, M.; Iwatate, R. J.; Hanaoka, K.; Terai, T.; Nagano, T.; Urano, Y. *Journal of the American Chemical Society* **2013**, 135, 409.
[2] Uno, S. N.; Kamiya, M.; Yoshihara, T.; Sugawara, K.; Okabe, K.; Tarhan, M. C.; Fujita, H.; Funatsu, T.; Okada, Y.; Tobita, S.; Urano, Y. *Nature chemistry* **2014**, 6, 681.

Evaluation of log *P* of endocrine disruptors using DFT methods

Masao Fujisawa[1] fujisawa@waka.kindai.ac.jp
Mamu Tabe[1] b14035tm@waka.kindai.ac.jp
Hirohito Ikeda[2] ikeda@fukuoka-u.ac.jp
Takayoshi Kimura[3] kimura@chem.kindai.ac.jp

[1] Department of Biotechnological Science, Kindai University, 930 Nishimitani, Kinokawa-City, Wakayama, Tokyo 649-6493, Japan
[2] Department of Pharmaceutical Science, Fukuoka University, 8-19-1 Nanakuma Fukuoka 814-0180, Japan
[3] Department of Chemistry, Kindai University, Kowakae, Higashi-osaka 577-8502, Japan

Keywords: DFT-D, Vibrational frequencies, log *P*

Solvation phenomena play a significant role in chemical reactions and biomolecular recognition; however, it can be very difficult to determine the related thermodynamic quantities. For example, to calculate the solvation Gibbs energy, one must know temperature dependence of the vaporization enthalpy or the vapor pressure. Moreover, the dispersion energy is an important contributor to the solvation energy of a solute. In an attempt to predict solvation Gibbs energy, theoretical methods including explicit or continuum solvent models have been developed and applied [1-3]. In an explicit solvent model, there have been an attempt of hydration Gibbs energies computed using the molecular dynamics simulation and the energy-representation theory of solvation [4]. In this study, the solvation Gibbs energies for several endocrine disruptors were calculated using dispersion-corrected density functional theory (DFT-D) method, which were then compared with experimental thermodynamic data. To decide the lowest energy structure, conformation search was performed. The obtained lowest energy conformer was the initial structure of the DFT-D calculations in gas phase. Vibrational frequencies were calculated for these optimized geometries at the same level of theory in gas phase. The solute geometries were optimized in water using DFT-D functions. Similarly, vibrational frequencies were calculated for these optimized geometries at the same level of theory in water. Gibbs energies of solutes were determined by vibrational frequencies in gas phase and in water. Partition coefficients (log *P*) were determined using the differences in Gibbs energy between two solvents (water, *n*-octanol). These predicted values well reproduced the experimental log *P*.

[1] Gallicchio, E., Zhang, L. V., Levy R. M., The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators, *Journal of Computational Chemistry*, 23:517-529, 2002.
[2] Shivakumar, D., Williams, J., Yujie, W., Damm, W., Shelley, J., and Woody, S., Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field, *Journal of Chemical Theory and Computation*, 6:1509-1519, 2010.
[3] Takano, Y. U., and Houk K. N., Benchmarking the Conductor-like Polarizable Continuum Model (CPCM) for Aqueous Solvation Free Energies of Neutral and Ionic Organic Molecules, 1:70-77, 2005.
[4] Matubayashi, N., Nakahara, M., Theory of solutions in the energy representation. III. Treatment of the molecular flexibility, *The Journal of Chemical Physics*, 119:9686-9702, 2003.

P1-04

Conformational regularity of the condensed tannins investigated by free-energy calculation

Reina Takahashi[1]
14a3031h@shinhsu-u.ac.jp

Takuma Todoroki[1]
14a3035a@shinshu-u.ac.jp

Hiroshi Fujii[1,2]
hfujii@shinshu-u.ac.jp

Hidefumi Makabe[1]
makabeh@shinshu-u.ac.jp

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

[1] Faculty of agriculture, Shinshu University, 8304 Minami-minowa, Kami-ina, Nagano, 399-4598, Japan
[2] Department of Interdisciplinary Genome Sciences and Cell Metabolism, Institute for Biomedical Sciences, Interdisciplinary Cluster for Cutting Edge Research, Shinshu University, Minami-minowa, Kami-ina, Nagano, 399-4598, Japan

Keywords: Natural product, Flavan-3-ol, Proanthocyanidin

Three-dimensional (3D) molecular structure is responsible for biological function. Biopolymers can adopt some stable conformations. The condensed tannins are known as the polymer of polyphenol in natural products. However, most of their conformations are still elusive. Then, we have investigated the stable conformations of the condensed tannins by free-energy calculation to figure out the conformational regularity.

Free-energy calculation was done for the natural polymer, the condensed tannins. The unit of the condensed tannin is flavan-3-ol such as catechin derivatives. The two adjacent units are connected by a single covalent bond; inter-flavan bond. The conformational freedom around the inter-flavan bond designates the relative direction between the upper and the lower units. In the free-energy calculation, we conducted the umbrella-sampling molecular dynamics simulations with the reaction coordinate of the dihedral angle around the inter-flavan bond. Herein, we chose the unit of catechin (cat), epicatechin (epi), gallocatechin (gc) and epigallocatechin (egc). For dimers of tannins, all permutation of two units were calculated (e.g. cat-cat, cat-epi, epi-cat, cat-gc, gc-cat, etc; total of 16 dimers). For trimers of tannins, total of 4 trimers (cat-cat-cat, epi-epi-epi, gc-gc-gc and egc-egc-egc) were investigated.

The results showed that the free-energy landscape (FEL) of the dimer depended on the upper unit. The FELs of the dimers with the upper unit of cat and gc had one minima while those of epi and egc did two minimal basins. It may be explained by hydrogen-bond pattern between upper and lower units. The cat unit differs from the epi one mainly at the point of chirality of connection to the lower unit. The hydrogen bonds between the upper and lower units can be formed according to the upper-unit conformation, which can contribute the structural stability. For the trimers, the FELs of cat and gc showed two low regions whereas those of epi and egc did four, which might indicate that the conformational stability of polymer depends on the adjacent units. In our presentation, the detail of conformation will be displayed. Furthermore, we will show the recent results for cat- and epi-polymers up to 7-mer.

P1-05

Functional profiling of asymmetrically-organized human CCT/TRiC chaperonin

Kazutaka Araki[1]
k-araki@aist.go.jp

Atsushi Suenaga[1, 2]
atsushi.suenaga@aist.go.jp

Tohru Natsume[1, 3]
t-natsume@aist.go.jp

Kazuhiko Fukui[1]
k-fukui@aist.go.jp

[1] Molecular Profiling Research Center for Drug Discovery, National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan
[2] Department of Biosciences, College of Humanities and Sciences, Nihon University, 3-25-40 Sakurajosui Setagaya-Ku, Tokyo 156-8550, Japan
[3] Robotic Biology Institute, Inc., Tokyo 135-0064, Japan

Keywords: Chaperonin, CCT, Cysteine, Protein modeling, Electrostatic calculation

Molecular organization of the eukaryote chaperonin known as CCT/TRiC complex was recently clarified. Eight distinct subunits are uniquely organized, providing a favorable folding cavity for specific client proteins such as tubulin and actin. Because of its heterogeneous subunit composition, CCT complex has polarized inner faces, which may underlie an essential part of its chaperonin function. In this study, we structurally characterized the closed and open states of CCT complex, using molecular dynamics analyses. Our results showed that the inter-subunit interaction energies were asymmetrically distributed and were remodeled during conformational changes of CCT complex. In addition, exploration of redox related characteristics indicated changes in inner surface properties, including electrostatic potential, pKa and exposure of inner cysteine thiol groups, between the closed and open states. Cysteine activation events were experimentally verified by interaction analyses, using tubulin as a model substrate. Our data highlighted the importance of dynamics-based structural profiling of asymmetrically oriented chaperonin function.

[1] Araki, K., Suenaga, A., Kusano, H., Tanaka, R., Hatta, T., Natsume, T., and Fukui, K., Functional profiling of asymmetrically-organized human CCT/TRiC chaperonin, *Biochemical and biophysical research communications*, 481:232-238, 2016.
[2] Araki, K., Kusano, H., Sasaki, N., Tanaka, R., Hatta, T., Fukui, K., and Natsume, T., Redox Sensitivities of Global Cellular Cysteine Residues under Reductive and Oxidative Stress, *Journal of proteome research*, 15:2548-2559, 2016.
[3] Araki, K., Ushioda, R., Kusano, H., Tanaka, R., Hatta, T., Fukui, K., Nagata, K., and Natsume, T., A crosslinker-based identification of redox relay targets, *Analytical biochemistry*, 520:22-26, 2017.

P1-06

Flexible Docking using Replica-Exchange Molecular Dynamics Simulation

Suyong Re[1] suyongre@riken.jp
Hiraku Oshima[1] hiraku.oshima@riken.jp
Motoshi Kamiya[2] motoshi.kamiya@riken.jp
Yuji Sugita[1,2,3] sugita@riken.jp

[1] RIKEN Quantitative Biology Center (QBiC), Integrated Innovation Building 7F, 6-7-1 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan
[2] RIKEN Advanced Institute for Computational Science (AICS), Integrated Innovation Building 7F, 6-7-1 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan
[3] RIKEN Theoretical Molecular Science Laboratory and iTHES, 2-1 Hirosawa Wako-shi, Saitama 351-0198, Japan.

Keywords: Molecular dynamics simulations, enhanced sampling, protein-ligand binding

Accurate prediction of ligand binding structure and energetics remains a challenge. Accounting for both ligand binding and protein structure changes, such as binding site reorganization, is still very difficult. Two-dimensional replica-exchange approach[1] is promising in that it enhances the binding events to predict the binding structure with high statistical accuracy. Here, we extend the method by introducing the generalized replica-exchange with solute tempering (gREST)[2]. The protein flexibility is incorporated by scaling the temperature of a wide solute region, involving the ligand and the active site residues, using gREST. We applied the method to study the ligand binding of Src kinase. We show that the native binding structure is correctly predicted. Importantly, the consideration of protein flexibility significantly improves the docking efficiency compared to the existing methods. We also stress that the method enhances the binding events to give high statistical accuracy, while long-time simulation could sample the event once per several microseconds. This approach, followed by binding free energy calculation, allows us to accurately predict protein-ligand binding affinity.

[1] Kokubo, H. et al. “Two-dimensional replica-exchange method for predicting protein-ligand binding structures.” J. Comput. Chem. (2013) 34:2601-2614.
[2] Kamiya, M. et al. in preparation.

P1-07

FMO calculations on specific interactions between vitamin-D receptor and its ligands

Ryosuke Takeda[1], Ittetsu Kobayashi[1], Kanako Shimamura[1], Hiromi Ishimura[1], Ryushi Kadoya[1], Kentaro Kawai[2], Atsushi Kittaka[3], Midori Takimoto-Kamimura[4], Noriyuki Kurita[1]

kurita@cs.tut.ac.jp

1 Department of Computer Science and Engineering, Toyohashi University of Technology, Tempaku-cho, Toyohashi, Aichi, 441-8580, Japan
2 Drug Research Center, Kaken Pharmaceutical Co. Ltd., 14, Shinomiya, Minamigawara-cho, Yamashina-ku, Kyoto, 607-8042, Japan
3 Faculty of Pharmaceutical Sciences, Teikyo University, 2-11-1 Kaga, Itabashi, Tokyo 173-8605, Japan
4 Teijin Institute for Bio-Medical Research, Teijin Pharma Ltd., 4-3-2 Asahigaoka, Hino, Tokyo, 191-8512, Japan

Keywords: Molecular simulation, Fragment molecular orbital, Vitamin D receptor, Inhibitor

Vitamin D_3 is hydroxylated at both the 25 and the 1α sites in a liver and a kidney, resulting in the active form of vitamin D_3 metabolites 1α,25-dihydroxyvitamin D_3 (1α,25$(OH)_2D_3$), which plays important roles in the regulation of calcium and phosphorus metabolism as well as in the bone formation. The physiological actions caused by 1α,25$(OH)_2D_3$ are triggered by its specific interaction with vitamin D receptor (VDR)[1]. The previous study [2] confirmed that almost all cell tissues in a living organism have VDR and that the binding of 1α,25$(OH)_2D_3$ to the VDR is deeply related to the pathogenesis of the immunological diseases such as cancer as well as the response anomaly to hormone. For elucidating the biological effect of VDR, it is indispensable to determine the structures of VDR bound by vitamin D derivatives. Therefore, various types of 1α,25$(OH)_2D_3$ derivatives were synthesized [3] and the structures of their complexes with human VDR were determined to clarify that the derivatives form hydrogen bonds to Arg274 residue of VDR and have high binding affinity for VDR. However, it is not elucidate the reason why a slight difference in structure of the derivatives causes a large difference in binding affinity between VDR and the derivatives. This fact might be a bottleneck for proposing novel derivatives as a potent modulator or inhibitor to VDR.

In the present study, we employed two types of 1α,25$(OH)_2D_3$ derivatives, whose structures are almost the same but their effect on VDR activity is significantly different, and their specific interactions with VDR were investigated at an electronic level, using *ab initio* molecular simulations based on fragment molecular orbital (FMO) method [4]. Based on the results of FMO calculations, we elucidated which parts of the derivatives and which residues of VDR are important for the specific binding between VDR and the derivatives [5].

[1] Darwish H., DeLuca H. F., “Vitamin D-regulated gene expression,” Crit. Rev. Eukaryot. Gene Expr., 3: 89−116, 1993.
[2] Walters M. R., “Newly identified actions of the vitamin D endocrine system,” Endocr. Rev., 13: 719−764, 1992.
[3] Matsuo M., Hasegawa A., Takano M., Saito H., Kakuda S., Takagi K., Ochiai E., Horie K., Takimoto-Kamimura M., Takenouchi K., Sawada D., Kittaka A., “Design and synthesis of 2α-(tetrazolylethyl)-1α,25-dihydroxyvitamin D_3 as a high affinity ligand for vitamin D receptor,” J. of Steroid Biochem. & Mole. Biol., 144: 201−203, 2014.
[4] Kitaura K., Ikeo E., Asada T., Nakano T., Uebayashi M., “Fragment molecular orbital method: an approximate computational method for large molecules,” Chem. Phys. Lett., 313: 701−706, 1999.
[5] Takeda R., Kobayashi I., Shimamura K., Ishimura H., Kadoya R., Kawai K., Kittaka A., Takimoto-Kamimura M., Kurita N., “Specific interactions between vitamin-D receptor and its ligands: *ab initio* molecular orbital calculations in water”, J. of Steroid Biochem. & Mole. Biol., 171: 75−79, 2017.

P1-08

Repetition of binding and unbinding processes between protein and ligand by supervised molecular dynamics

Takashi Mitsui[1]
tmitsui@jp.fujitsu.com

[1] Bio-IT R&D Office, Research& Development Division, Healthcare Systems Unit, Fujitsu Limited, 1-17-25 Shin-Kamata, Ota-ku, Tokyo 226-8504, Japan

Keywords: Molecular dynamics, Ligand binding

The mechanism of the binding/unbinding processes between protein and ligand is a crucial issue in *in-silico* drug design. In recent published papers, unbiased long-time molecular dynamics (MD) simulations reproduced experimentally observed protein-ligand complex structures. However, ligand binding to its receptor in correct pose is rare event, therefore it requires microsecond time-scale MD simulation.

Supervised molecular dynamics (SuMD) [1][2] is tabu-like computational algorithm developed to follow protein-ligand approaching processes within a relatively short time-scale compared to traditional MD simulation. In this study, we extended the original protocol to an unbinding process to enable repetition of multiple binding/unbinding events.

A demonstrative simulation was carried out using the X-ray co-crystal structure, 1QY2(PDB)[3], of major urinary protein (MUP) and its ligand, 2-isopropyl-3-methoxypyrazine (IPMP). The initial structure was prepared by taking the ligand apart from the binding pocket of the protein. Multiple SuMD simulations were performed and multiple binding/unbinding events occurred in each trajectory. Resultant complex structures were clustered into several binding poses. They included a ligand binding pose observed in traditional MD simulations of the X-ray co-crystal structure.

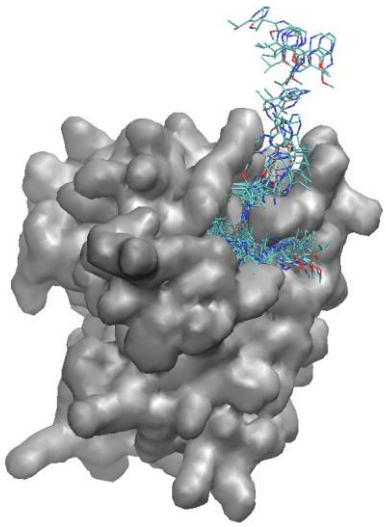


Figure 1. A typical trajectory of ligand binding process.

[1] D. Sabbadin, S. Moro, “Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale”, J. Chem. Inf. Model., **54**, 372 (2014)

[2] A. Cuzzolin, et al., “Deciphering the complexity of ligand-protein recognition pathway using supervised molecular dynamics (SuMD) simulations”, J. Chem. Inf. Model., 56, 687 (2016)

[3] R. J. Bingham, et al., “Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-Methoxy-3-isobutylpyrazine to the major urinary protein”, J. Am. Chem. Soc., 126, 1675 (2004).

P1-09

In silico protein design for functional modification of photoactivated adenylate cyclase

Mayu Tanaka[1] **Toru Ekimoto**[1] **Mio Ohki**[1] **Tsutomu Yamane**[1]
w165425b@yokohama-cu.ac.jp

Sam-Yong Park[1]
park@yokohama-cu.ac.jp

Mitsunori Ikeguchi[1]
ike@yokohama-cu.ac.jp

[1] Graduate School of Medical Life Science, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

Keywords: Photoactivated adenylate cyclase, Mutation analysis, Protein design

Photoactivated adenylate cyclase (PAC) is a photoreceptor protein that produces the second messenger cyclic-AMP (cAMP) upon light illumination. PAC consists of the blue light using flavin (BLUF) and the adenylate cyclase (AC) domains. Recently, the crystal structure of *Oscillatoria acuminate* photoactivated adenylate cyclase (OaPAC) has been solved [1].

Optogenetics is a rapidly growing field in which light is used to control biological systems. OaPAC is expected as an optgenetic tool to produce cAMP upon blue-light illumination. In addition to cAMP, the control of the cGMP level upon light illumination would be useful to investigate the effects of the cGMP on biological systems. Therefore, in this study, we attempt *in-silico* protein design to modify OaPAC to produce cGMP instead of cGMP.

In the crystal structure of OaPAC, ATP or ATP analog was not bound to the ATP binding site. Therefore, in this study, the crystal structure of CyaC, a homolog of OaPAC, was used to design the nucleotide-binding site, because ApCpp, an ATP analog, was bound to the crystal structure of CyaC. First, ApCpp was replaced to GTP in the crystal structure of CyaC. Then, amino acid in the binding site were comprehensively mutated, and the changes in GTP affinity upon mutations were calculated. Also, the changes in stability of the protein upon mutations were calculated. Mutants predicted to have high affinity of GTP forms hydrogen bonds to GTP and not to ATP. Interactions with both two GTPs of the CyaC dimer were found to be important to raise affinity to GTP.

[1] Ohki, M. *et al.,* Structural insight into photoactivation of an adenylate cyclase from a photosynthetic cyanobacterium, *PNAS*, 113:6659-6664, 2016

Validation of Epigenetic Therapeutic Target Proteins for Homogenous Assay Performance

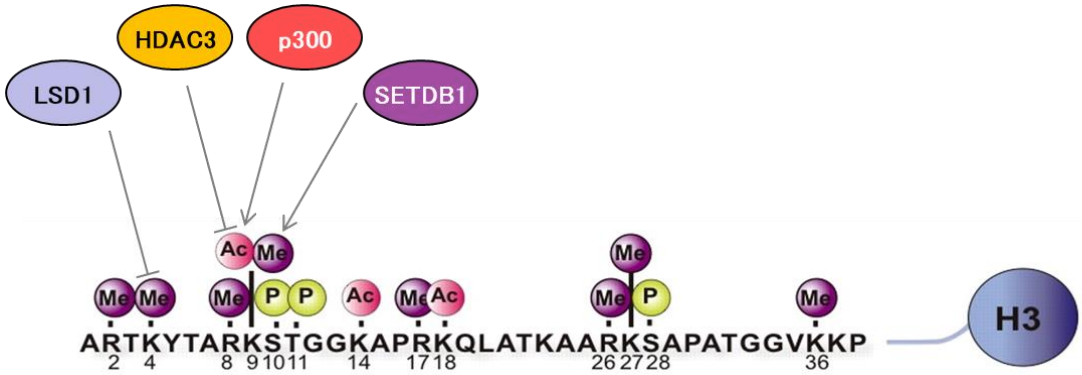
Masato Yonezawa[1], Mary Anne Jelinek[1], Melissa Ritland[1], Jake Dabrowski[1], Wei Gong[1], Lingchun Kong[2] and Fei Lan[2]

[1] Active Motif Inc., Carlsbad CA, USA
[2] Active Motif China, Shanghai, China
Email: yonezawa@activemotif.com

Keywords: Protein-small molecule interaction, Enzymatic reaction

The cancer genome atlas project (TCGA) revealed a great number of potential new therapeutic targets among epigenetic factors. To accelerate epigenetic drug discovery, we produced a protein toolbox of reagents including active enzymes, recombinant substrates and detection antibodies, and set up homogeneous screening assay platforms for histone deacetylases (HDACs), histone acetyltransferases (HATs), lysine methyltransferases (KMTs), lysine demethylases (KDMs), and bromodomain proteins (BRDs). Here, we present data obtained by AlphaLISA (amplified luminescent proximity homogeneous assay) platforms for HDAC3, LSD1, p300, SETDB1, and BRD family members. In general, IC50 values of these proteins are well within published results for the various compounds. This protein toolbox continues to expand to include “designer” oligo nucleosomes which include site directed installation of modified amino acids. Additional candidates can now be screened using these alternate nucleosome-based substrates. Thus, our assay platforms will be very useful to screen and validate candidate epigenetic drugs as well as to study properties of epigenetic factors.

[1] https://cancergenome.nih.gov/
[2] www.activemotif.com/catalog/5
[3] www.activemotif.com/catalog/744
[4] www.activemotif.com/catalog/1068



P1-11

Comprehensive database analysis of protein kinase structures

Kei Moritsugu moritugu@yokohama-cu.ac.jp
Yoshihiko Nishino nisshy155416@yokohama-cu.ac.jp
Akinori Kidera kidera@yokohama-cu.ac.jp

Graduate School of Medical Life Science, Yokohama City University, 1-7-29 Suehirocho, Tsurumi-ku, Yokohama 230-0045, Japan

Keywords: Protein-small molecule interaction, Molecular dynamics method, Protein kinase, KLIFS database, Structural bioinformatics

Protein kinases are ATP phosphotransferases targeting at Ser/Thr/Tyr residues of specific substrate proteins and then play an important role for signaling in cells. While some kinase proteins involved in diseases have been extensively studied as drug-discovery targets, few of studies have been done to understand complete structural dynamics for overall protein kinases. A protein kinase database, KLIFS [1], collects a variety of human/mouse protein kinases with various kinase families and species, kinase activities, bound (drug) molecules, site-directed mutations, and so on, which however takes interest on local atom interactions between the proteins and ligands. In the present study, by use of the KLIFS database, we have attempted comprehensive analyses of protein kinase collective motions in relation to the way of local protein-drug interactions.

Data collection for this study was carried out from 849 human tyrosine kinases in the KLIFS database by excluding structures with a number of lacking atoms or with large similarities, leading to 150 representative kinase structures. Motion Tree [2,3] calculated using 207 Cα atoms after multiple alignment of the 150 structures demonstrated a significance of the domain motion between kinase N-lobe and C-lobe, as well as the motions of both activation loop and C-helix which have often been discussed in previous studies. Structural classifications were performed by projecting of each protein kinase onto the three motions and by assigning the information on kinase activities (DFG-in/C-helix-in, etc) and kinase species, which were then analyzed in relation to the atom interactions with bound molecules such as ATP-analog and drugs, and in combination with the structural dynamics data calculated from comprehensive molecular dynamics simulations. These findings yielded a general picture of how protein kinases are inactivated due to what kinds of drugs on which binding pockets, which will be useful for designing new drug molecules for specific protein kinases.

[1] A. J. Kooistra et al., “KLIFS: a structural kinase-ligand interaction database”, *Nucleic Acids Res* 44: D365-D371 (2016).
[2] K Moritsugu, R Koike, K Yamada, H Kato, A Kidera, "Motion Tree Delineates Hierarchical Structure of Protein Dynamics Observed in Molecular Dynamics Simulation", *PLoS ONE* 10: e0131583 (2015).
[3] R Koike, M Ota, A Kidera, "Hierarchical description and extensive classification of protein structural changes by Motion Tree", *Journal of Molecular Biology* 426: 752-762 (2014).

Finite-size effect on the charging free energy in the alchemical perturbation and ``warp drive'' method

Toru Ekimoto ekimoto@yokohama-cu.ac.jp
Tsutomu Yamane tyamane@yokohama-cu.ac.jp
Mitsunori Ikeguchi ike@yokohama-cu.ac.jp

Graduate School of Medical Life Science, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

Keywords: Binding free energy, Free energy perturbation, Charging free energy, Finite-size effect

With increase of computational power, calculations of the binding free energy by exact methods, such as the alchemical free energy perturbation (alchemical-FEP), routinely become possible with all-atom molecular dynamics (MD) simulations. In the alchemical-FEP, the bound ligand is thermodynamically eliminated from the binding site and is emerged to a bulk region. The alchemical-FEP can treat buried ligands in deep pocket, however, accuracy for highly charged ligands is poor. One of the reasons is due to the finite-size effect on the charging free energy (CFE) in periodic systems. The CFE is the free energy change for tuning on the electrostatic interaction of the solute with the solvent. The finite-size effect refers to the cell-size dependence of the CFE at different cell sizes. [1] When the CFE is left uncorrected, a comparison of the CFE among different charged states is erroneous as shown in the examinations for ions and proteins [1,2]. The CFE is a key component of the binding free energy, therefore, a correction scheme is necessary to obtain the CFE at the limit of large cell size.

In this study, we examine the finite-size effect on the CFE in the alchemical-FEP with systematically varied cell sizes of the MD unit cell, assess the performance of the correction scheme formulated by Hummer et al. [1], and introduce an alternative perturbation, termed “warp drive method” [3], providing the CFE at the limit of large cell size without any corrections. The phosphotyrosine peptide (-5e) bound to the Src homology 2 domain (+1e) is employed as a test complex, and the CFE is calculated by thermodynamic integration method. We show that the finite-size effect arises at ~92 kcal/mol in the annihilation process and ~180 kcal/mol in the emerging process. The self-energy correction essentially corrects them within ~2 kcal/mol and ~3 kcal/mol, respectively, and the additional correction from the solvation effect reduces the remains to ~2 kcal/mol and ~1 kcal/mol, respectively. This shows that the corrections scheme is absolutely necessary to obtain the CFE at the limit of large cell size, however, its performance at the small cell sizes is less effective in this examination. We also examine the warp drive method in which both processes in the alchemical-FEP is simultaneously executed using a solution system consisting of a protein-ligand complex and a distantly-positioned unbound-ligand in one MD unit cell. The eliminated partial charges of the bound ligand simultaneously emerge on the other ligand in bulk, and therefore, the total charge of the system does not change at all intermediate sates. The CFE by this method does not show the finite-size effect even at small cell sizes, and the value without any corrections is in good agreement with the summation of the corrected CFEs calculated by the alchemical-FEP.

[1] Hummer, G., Pratt, L.R., Garcia, A.E., *Journal of Physical Chemistry*, 100: 1206-1215, 1996.
[2] Ekimoto, T., Matubayasi, N., and Ikeguchi, M., *Journal of Chemical Theory and Computation*, 11:215-223, 2015.
[3] Ekimoto, T., Yamane, T., and Ikeguchi, M., in preparation.

P1-13

Molecular dynamics and *ab initio* FMO calculations for amyloid-β nonamer

Shogo Tomioka[1], Haruki Sogawa[1], Hiromi Ishimura[1], Akisumi Okamoto[1], Sergiy Shulga[2], Pavel Karpov[2], Yaroslav Blume[2], Noriyuki Kurita[1]
kurita@cs.tut.ac.jp

1 Department of Computer Science and Engineering, Toyohashi University of Technology, Tempaku-cho, Toyohashi, Aichi, 441-8580, Japan
2 Institute for Food Biotechnology and Genomics, National Academy of Sciences of Ukraine, 2a. Osypovskogo Str., Kyiv-123, 04123, Ukraine

Keywords: Molecular simulation, fragment molecular orbital, molecular dynamics, Alzheimer's disease, amyloid-beta, aggregate

The accumulation of amyloid-β (Aβ) oligomers and fibrils in a brain has been recognized to be a major cause of the onset of Alzheimer's disease (AD). It is thus expected that the inhibition of Aβ aggregations can prevent the onset of AD, and many kinds of agents with strong binding affinity to Aβ have been developed. In the design of these potent inhibitors against the Aβ aggregation, it is necessary to make clear the structures of the Aβ oligomers and fibrils as well as the mechanism of the aggregation. Recently, Lu *et al.*[1] revealed structural models for the *in vivo* Aβ nonamer, based on the solid-state nuclear magnetic resonance analyses for the Aβ fibrils derived from the brains of two different AD patients. Each of these Aβ nonamer models has a single and patient-specific structure possessing three-fold symmetry with respect to the axis of fibril growth. However, it is not elucidated why such high symmetry structure of Aβ fibrils can be stabilized.

Molecular simulations such as molecular mechanics (MM) and molecular dynamics (MD) ones for Aβ fibrils are efficient for investigating their stable structures and the mechanism of aggregation at an atomic level. Kahler *et al.*[2] conducted a systematic computational study based on all-atom classical MD simulations for many types of fibrillary Aβ oligomers and concluded that the pairs of Aβ protofilaments are important as a seed for forming Aβ fibrils. However, they considered only Aβ(9-42) fragment, missing the residues of N-terminal region, which were found to be important for Aβ aggregation in the previous experiment.

In our previous studies [3,4], replica-exchange MD (RE-MD) simulations were conducted in water for monomer and dimer models of Aβ(1-42) peptides, in order to search widely for their stable conformations in water. In addition, we carried out *ab initio* fragment molecular orbital (FMO) calculations for the conformations obtained by the RE-MD simulations and determined the stable conformations of the solvated Aβ(1-42) monomer and dimer, with considering water molecules explicitly. We furthermore investigated the specific interactions between Aβ peptides in the Aβ hexamers at an electronic level, using *ab initio* FMO method [5]. Our *ab initio* simulations elucidated the importance of structural water molecules for the stabilization of Aβ fibrils.

In the present study, to clarify the reason why the Aβ nonamer with three-fold symmetry obtained by Lu *et al.*[1] is stable, we investigated the change in structure of the Aβ nonamer by use of classical MD simulations in water. In addition, *ab initio* FMO calculations were carried out for some snapshots obtained by the MD simulations. The results revealed that the interactions between the Aβ peptides of stacked Aβ pairs make resultant contribution in stability of the Aβ nonamer.

[1] Lu J.X., Qiang W., Yau W.M., Schwieters C.D., Meredith S.C., Tycko R., Cell 154 (2013) 1257–1268.
[2] Kahler A., Sticht H., Horn A.H.C., Plos One 8 (2013) e70521.
[3] Yano A., Okamoto A., Nomura K., Higai S., Kurita N., Chem. Phys. Lett. 595/596 (2014) 242–249.
[4] Okamoto A, Nomura K., Yano A., Higai S., Kondo T., Kamba S., Kurita N., Chem. Phys. Lett. 577 (2013) 131–137.
[5] Ishimura H., Tomioka S., Kadoya R., Simamura K., Okamoto A., Shulga S., Kurita, N., Chem. Phys. Lett. 672 (2017) 13–20.

Specific interactions between mycobacterial FtsZ and curcumin derivatives: molecular docking and *ab initio* molecular simulations

Mitsuki Fujimori[1], Haruki Sogawa[1], Shintaro Ota[1], Pavel Karpov[2], Sergiy Shulga[2], Yaroslav Blume[2], Noriyuki Kurita[1]
kurita@cs.tut.ac.jp

1 Department of Computer Science and Engineering, Toyohashi University of Technology, 1-1 Hibarigaoka, Tenpaku-cho, Toyohashi, Aichi, 411-8580, Japan
2 Institute for Food Biotechnology and Genomics, National Academy of Sciences of Ukraine, 2a. Osypovskogo Str., Kyiv-123, 04123, Ukraine

Keywords: Curcumin; Inhibitor; FtsZ; Tuberculosis; Cell division; Fragment molecular orbital; Protein-ligand docking

Tuberculosis (TB) is one of the most widespread infectious diseases caused by the bacillus Mycobacterium tuberculosis (*Mtb*). In the treatment of TB, many kinds of drugs such as isoniazid, rifampicin, pyrazinamide and ethambutol have been administered. However, there is a considerable potential for *Mtb* to have resistance against these drugs. In particular, the *Mtb* having resistance against multiple drugs is called multidrug-resistant TB (MDR-TB), and the number of MDR-TBs is increasing rapidly. It is thus necessary to develop new anti-TB drugs targeting the most conservative proteins, which cannot be mutated easily [1]. As a candidate of anti-TB drugs, new compounds were proposed, which suppress the growth of *Mtb* by inhibiting the division of *Mtb* cell [2]. These drugs are targeted to cytoskeletal protein FtsZ (filamenting temperature-sensitive mutant Z), which plays an essential role in the cell division mechanism [3]. The inhibitor against FtsZ function is expected to suppress the *Mtb* cell growth.

As a novel inhibitor against *Mtb* FtsZ, we here considered curcumin derivatives. Curcumin is a natural product and contained in the root of *Curcumae Rhizoma*, while other curcuminoids such as demethoxycurcumin and bisdemethoxycurcumin are also included in the root [4]. These curcumin derivatives have been widely used [4] as conventional drugs for treating many diseases. It was found that curcumin suppresses bacterial cell proliferation by inhibiting the FtsZ function [5]. However, since there are some ligand-binding pockets in *Mtb* FtsZ, the binding site of curcumin on FtsZ and the specific interactions between curcumin and FtsZ are not elucidated yet.

In the present study, we investigated the specific interactions between *Mtb* FtsZ and some curcumin derivatives, using *ab initio* molecular simulations based on protein-ligand docking, classical molecular mechanics optimization and *ab initio* fragment molecular orbital (FMO) calculation. Based on the FMO results, we attempted to reveal which curcumin derivative can bind more strongly to FtsZ. In addition, we elucidated which parts of FtsZ and curcumin derivative are important for the specific interactions between them. The result will be useful for proposing novel anti-TB drugs based on curcumin derivatives.

[1] H. Tomioka, *Kekkaku*, 85: 815–822, 2010.
[2] P.A.J de Boer, *et al.*, *Cell*, 56: 641–649, 1989.
[3] M. Osawa, *et al.*, *Cell*, 56: 641–649, 1989.
[4] F. Kiuchi, *et al.*, *Chem. Pham. Bull*, 41: 1640–1643, 1993.
[5] D. Rai, *et al, Biochem. J.*, 410: 147–155, 2008.

QCLObot: an automation engine of canonical MO calculation in proteins

Toshiyuki Hirano[1]
t-hirano@iis.u-tokyo.ac.jp

Fumitoshi Sato[1]
satofumi@iis.u-tokyo.ac.jp

[1] Institute of Industrial Science, the University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

Keywords: canonical molecular orbital calculation, electronic structure calculation, proteins

The initial guess of the SCF calculation is one of the most important factors to successfully achieve canonical molecular orbital (CMO) calculation. In order to get a precise initial guess for large systems, the QCLO method [1, 2] has demonstrated some great performances [3-5]. The QCLO is a kind of localized orbital, which is not only localized in a certain region of the molecule but also canonical in that region. Gathering of the subsystem QCLOs, we can obtain the precise initial guess of the super-system. Considering the protein structure, such as secondary structure and ion pairs, we can get more precise initial guess which is almost equal to the SCF convergence [2]. However, the procedure of computing and gathering QCLOs is so complicated and difficult.

We have developed an automation program, QCLObot [6]. The QCLObot is inspired by an open-source automation provisioning program, Ansible [7]. Both of the applications are written by the Python, and the format of input file is based on the YAML human-readable serialization data format [8]. The QCLObot playbook is also idempotent, in order to prevent redundant calculations and unexpected side-effects on the QCLO method. The QCLObot has also used template engine based on Jinja2 [9], so using variables and dynamic expression are enabled in the playbook.

The fragments, which are components of the computing molecule named as the frame molecule, are encouraged to define by the hierarchic structure of the YAML format. The corresponding QCLOs of the fragments are automatically formed the basis for the YAML playbook. Therefore, not only contiguous amino-acid residues, but also spatially-designated chemical components can be treated as a QCLO fragment. Heteroatoms, such as coenzyme and prosthetic group, can be also easily computed based on the QCLO method by using the QCLObot program.

We have performed some CMO calculations of proteins by using the QCLObot program. Recently, the simple geometrical optimization method (steepest descent and conjugate gradient method) based on the QCLO method has been implemented. Computational efficiency strategy by using the combination of corresponding orbitals in the previous optimization step is under consideration. And, in order to provide an easy way to make a calculation model, an elementary modeling system for proteins has been developed. In this presentation, we report these new features and some simulation results using the QCLObot application.

[1] H. Kashiwagi, H. Iwai, K. Tokieda, M. Era, T. Sumita, T. Yoshihiro, F. Sato, *Mol. Phys.*, 101:81-86, 2003.
[2] N. Nishino-Uemura, T. Hirano, F. Sato, *J. Chem. Phys.*, 127:184106, 2007.
[3] K. Chiba, T. Hirano, F. Sato, M. Okamoto, Quant. Chem., 113: 2345-2354, 2013.
[4] Y. Yagi, Y. Naoshima, *Trans. J. Soc. Sim. Tech.*, 4:127-136, 2012.
[5] T. Inaba, N. Tsunekawa, T. Hirano, T. Yoshihiro, H. Kashiwagi, F. Sato, *Chem. Phys. Lett.*, 434:331-335, 2007.
[6] https://github.com/ProteinDF/QCLObot
[7] https://www.ansible.com
[8] http://yaml.org
[9] http://jinja.pocoo.org

FMO calculations on specific interactions between androgen receptor and non-steroid agents

Ittetsu Kobayashi[1], Ryosuke Takeda[1], Rie Suzuki[1], Kanako Shimamura[1], Hiromi Ishimura[1], Ryushi Kadoya[1], Kentaro Kawai[2], Midori Takimoto-Kamimura[3], Noriyuki Kurita[1]
kurita@cs.tut.ac.jp

[1] Department of Computer Science and Engineering, Toyohashi University of Technology, Tempaku-cho, Toyohashi, Aichi, 441-8580, Japan
[2] Drug Research Center, Kaken Pharmaceutical Co. Ltd., 14, Shinomiya, Minamigawara-cho, Yamashina-ku, Kyoto, 607-8042, Japan
[3] Teijin Institute for Bio-Medical Research, Teijin Pharma Ltd., 4-3-2 Asahigaoka, Hino, Tokyo, 191-8512, Japan

Keywords: Molecular simulation, Fragment molecular orbital, Androgen receptor, Non-steroid agent

Androgen receptor (AR) is a family of nuclear receptor proteins and a ligand-activated transcription factor [1]. The bindings of ligand as well as coactivator to AR are considered to be a trigger for the onset of the progression mechanism of prostate cancer. Accordingly, many types of antagonists against AR have been developed as promising agents for treating prostate cancers. Among these agents, bicalutamide [2] with a non-steroid skeleton compound has been widely used. However, since some mutant-type ARs with strong drug-resistance have appeared, it is essential to develop novel agents alternative to bicalutamide.

In the present study, we investigated the specific interactions between AR and several types of non-steroid agents at an electronic level, using *ab initio* molecular simulations. We first obtained the stable structures of AR+ligand complexes in water by classical molecular mechanics (MM) calculations, and the electronic properties of the stable structures were investigated by the *ab initio* fragment molecular orbital (FMO) method [3]. The interaction energies between AR and each of the ligands evaluated by FMO were confirmed to explain the trend of the observed binding affinity between AR and the ligands. It was highlighted that Asn705 and Arg752 of AR contribute mainly to the binding between AR and the ligands.

In addition, we proposed some novel agents promising as potent ligands against AR and investigated the binding properties between AR and these agents [4]. The *ab initio* molecular simulations indicate that some of our proposed agents can bind more strongly with AR than the existing ligands and be a potent ligand against AR.

[1] S. Grosdidier, L.R. Carbó, V. Buzón, G. Brooke, P. Nguyen, J.D. Baxter, C. Bevan, P. Webb, E. Estébanez-Perpiná, J. Fernández-Recio, Allosteric conversation in the androgen receptor ligand-binding domain surfaces, *Mol. Endocrinol.*, 26:1078–1090, 2012.
[2] C.E. Bohl, Z. Wu, J. Chen, M.L. Mohler, J. Yang, D.J. Hwang, S. Mustafa, D.D. Miller, C.E. Bell, J.T. Daltona, Effect of B-ring substitution pattern on binding mode of propionamide selective androgen receptor modulators, *Bioorg. Med. Chem. Lett.*, 18:5567–5570, 2008.
[3] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayashi, Fragment molecular orbital method: an approximate computational method for large molecules, *Chem. Phys. Lett.*, 313:701–706, 1999.
[4] I. Kobayashi, R. Takeda, R. Suzuki, K. Shimamura, H. Ishimura, R. Kadoya, K. Kawai, M. Takimoto-Kamimura, N. Kurita, Specific interactions between androgen receptor and its ligand: *ab initio* molecular orbital calculations in water, *J. of Mole. Graph. & Model.*, 75:383-389, 2017.

Binding properties between curcumin and malarial tubulin: molecular-docking and *ab initio* FMO calculations

Shintaro Ota[1], Mitsuki Fujimori[1], Sergiy Shulga[2], Noriyuki Kurita[1]
kurita@cs.tut.ac.jp

[1] Department of Computer Science and Engineering, Toyohashi University of Technology, Tempaku-cho, Toyohashi, Aichi, 441-8580, Japan
[2] Institute for Food Biotechnology and Genomics, National Academy of Sciences of Ukraine, 2a. Osypovskogo str., Kyiv-123, 04123, Ukraine

Keywords: Fragment molecular orbital, Protein and ligand docking, in silico drug design

Malaria is caused by the protozoa of the genus Plasmodium. Because of its prevalence, virulence, and drug resistance, it is the most serious and widespread parasitic disease. In spite of decades of intense researches, malaria remains a lethal disease in the world. Therefore it is urgently required to develop inexpensive and effective anti-malarial drugs.

Curcumin, a polyphenolic organic molecule derived from turmeric, has been used widely as a traditional medicine [1]. It was found that curcumin is effective against malaria parasites, and the possibility that curcumin can be used a novel therapeutic agent against malaria was recognized [2].

Curcumin is also known to have inhibitory effect on mammalian tubulin, a protein that is essential for cell division and cellular trafficking. It was found that curcumin can bind strongly to tubulin *in vitro* to inhibit polymer assembly and induce tubulin dimer aggregation, resulting in depolymerization of mitotic microtubules during cell division [3].

Recent immunofluorescence as well as molecular docking studies for the human cancer cell lines suggested that curcumin might bind at the interface of alpha-beta tubulin heterodimer leading to altered microtubule morphology [4]. This binding site was confirmed to be similar to that of the microtubule destabilizing drug colchicine. Therefore, curcumin is expected to have a similar effect on tubulin as the existing drug colchicine. However, it is not elucidated why curcumin binds to the colchicine site and not to the other ligand-binding sites of tubulin.

In the present study, we searched widely the binding site of curcumin and its derivatives on the alpha- and beta-tubulin monomers, using *ab initio* molecular simulations based on fragment molecular orbital (FMO) method. In addition, the same simulations were conducted for the existing drugs (colchicine and vinblastine), and the results for these different types of drugs were compared. The results obtained by our simulations will contribute to the development of new therapeutic agents for malaria.

[1] R. K. Maheshwari, A. K. Singh, J. Gaddipati, R. C. Srimal, Multiple biological activities of curcumin: A short review, *Life Sciences*, 78:2081-2087, 2006.
[2] R. C. Reddy, P. G. Vatsala, V. G. Keshamouni, G. Padmanaban, P. N. Rangarajan, Curcumin for malaria therapy, *Biochemical and Biophysical Research Communications*, 326:472-474, 2005.
[3] K. K. Gupta, S. S. Bharne, K. Rathinasamy, N. R. Naik, D. Panda, Dietary antioxidant curcumin inhibits microtubule assembly through tubulin binding, *The FEBS Journal*, 273: 5320-5332, 2006.
[4] R. Chakrabarti, P. S. Rawat, B. M. Cooke, R. L. Copeel, S. Patankar, Cellular Effects of Curcumin on Plasmodium falciparum Include Disruption of Microtubules, *PLOS ONE*, 8(3), 2013.

A soft-core potential suitable to a special purpose computer for molecular dynamics

Teruhisa S. Komatsu [1] **Makoto Taiji**[1]
teruhisa.komatsu@riken.jp **taiji@riken.jp**

[1] Research and Development Project of Special Purpose Supercomputers for Drug Design, Quantitative Biology Center, RIKEN,
2F QBiC Building B, 6-2-4 Furuedai, Suita Osaka 565-0874, Japan

Keywords: molecular dynamics simulation, soft-core potential, special purpose computer

We are developing a special purpose computer for classical molecular dynamics, MDGRAPE-4A, a improved version of MDGRAPE-4 [1], which aimes a long time molecular dynamics simulation with the speed of 10 microseconds per day. One of the main application of its ability to perform a long time simulation is expected to accelerate a drug design process. The system is designed to perform simple molecular dynamics simulation faster as possible. Although the basic software interface is developed based on GROMACS [2], most of the molecular dynamics code for our special purpose computer should be rewritten from a scratch and thus only a limited number of the specific schemes available in GROMACS will be implemented from the viewpoints of computational speed and software developmental cost.

Soft-core potential is a common scheme to treat particle creation/annihilation process often used in double-decoupling/alchemical binding free energy calculations [3]. The full calculation of these schemes require massively parallel computation for many lambda states and is not considered as the task expected/suitable for MDGRAPE-4A. But long time simulation for a lambda state may have a merit. We are planing to implement (without loss of computational speed) soft-core potential in a limited way, mainly usable for double-decoupling scheme. The soft-core potential form implemented in MDGRAPE-4A is chosen as a suitable form to special purpose computer, more concretely, the force exerted by the new soft-core potential inside the softening region ($r<rc$) is simply proportional to the particle distance r,

$$f \propto r$$

which is different from that [4] used in GROMACS and also from recently proposed one [5]. In order to confirm that our new soft-core potential form performs well, we simulate simple test systems and compare the results.

[1] I. Ohmura, G. Morimoto, Y. Ohno, A. Hasegawa, M. Taiji, MDGRAPE-4: a special-purpose computer system for molecular dynamics simulations, Phil. Trans. R. Soc. A 372 (2014) 20130387.
[2] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1 (2015) 19-25.
[3] Y. Deng, B. Roux, Calculation of standard binding free energies: aromatic molecules in the T4 Lysozyme L99A mutant. JCTC 2 (2006) 1255-1273.
H. Fujitani, Y. Tanida, A. Matsuura, Massively parallel computation of absolute binding free energy with well-equibrated states, Phys. Rev. E79 (2009) 021914.
[4] T. C. Beutler, A. E. Mark,R. C. van Schaik, P. R. Gerber, W. F. van Gunsteren, Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations, Chem. Phys. Lett. 222 (1994) 529-539.
[5] V. Gapsys, D. Seeliger, B. L. de Groot, New soft-core potential function for molecular dynamics based alchemical free energy calculations, J.C.T.C. 8 (2012) 2373-2382.

Free-energy analysis of the urea and alkylurea effects on the structure of protein

Yu Yamamori
yamamori@cheng.es.osaka-u.ac.jp

Nobuyuki Matubayasi
nobuyuki@cheng.es.osaka-u.ac.jp

Division of chemical Engineering, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan

Keywords: protein denaturation, cosolvent effect, free-energy calculation, molecular dynamics

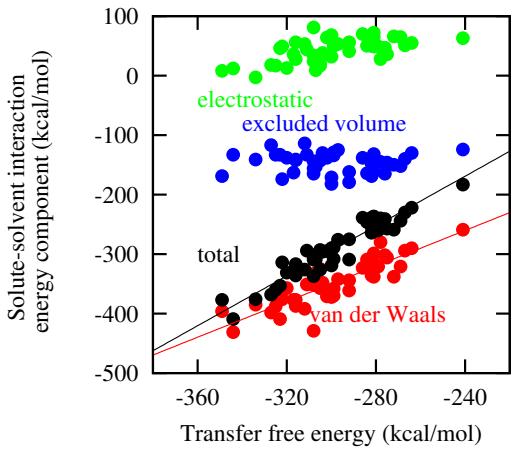


Figure Transfer free energy against the solute-solvent interaction energy (black), the electrostatic component (green), the van der Waals component (red), and the excluded-volume component (blue). 50 structures in total were subject to the free energy calculations.

The mechanism of urea-induced denaturation of a protein has been discussed for a long time. The major controversy is which interaction component is responsible for denaturation: electrostatic, van der Waals (dispersion) or excluded-volume effect (cavitation). In this study, the transfer energetics of cytochrome *c* [1] and T4-lysozyme [2] from pure-water solvent to cosolvent-water mixed solvent was analyzed. For cytochrome *c*, urea is adopted and for T4-lysozyme, urea, methylurea, 1,1-dimethylurea and isopropylurea was adopted as cosolvent. A variety of protein structures from near-native to partially unfolded ones were generated by molecular dynamics simulation in each mixed solvent. Each structure in this set was subject to the calculation of the solvation free energy in pure-water solvent and cosolvent-water mixed solvent. The energy-representation method was employed to compute the solvation free energy. The correlations of the transfer free energy of cytochrome *c* from pure-water solvent to urea-water mixed solvent against the transfer values of the average of solute-solvent interaction and its components are shown in the figure. The strong correlation is found against the van der Waals component, while the correlations are weak for the electrostatic and excluded-volume components. Thus the van der Waals interaction is the dominant component among the three. The transfer free energy is further decomposed into the urea and water contributions; the former is referred as the “direct mechanism” and the latter as the “indirect mechanism”. A similar analysis of correlation supports the “direct mechanism”. The direct mechanism is also supported in the cases of denaturation of T4-lysozyme caused by alkylurea.

[1] Yamamori, Y.; Ishizuka, R.; Karino, Y.; Sakuraba, S.; Matubayasi, N. *J. Chem. Phys.* **2016**, *144*, 085102.
[2] Yamamori, Y.; Matubayasi, N. *J. Chem. Phys.* **2017**, *146*, 225123.

Efficient calculation of electrostatic potential of biomolecule based on fragment molecular orbital method

Takeshi Ishikawa[1]
t-ishi@nagasaki-u.ac.jp

[1] Department of Molecular Microbiology and Immunology, Graduate School of Biomedical Science, Nagasaki University, 1-12-4 Sakamoto, Nagasaki, 852-8523, Japan

Keywords: electrostatic potential, fragment molecular orbital method, point-charge approximation

Accurate calculation of electrostatic potential (ESP) is one of the important issues in the computational chemistry and biology. For example, ESP at the molecular surface plays a key role in the protein-protein or protein-ligand interaction, and is also responsible for the catalytic activity of many enzymes. In this study, efficient quantum chemical calculations of the ESP were performed based on fragment molecular orbital (FMO) method[1]. In FMO method, ESP at a position $\boldsymbol{r}_m$ is defined as the following equation:

$$\phi^{FMO}(\boldsymbol{r}_m) = \sum_I \phi^I(\boldsymbol{r}_m) + \sum_{I<J} \Delta\phi^{IJ}(\boldsymbol{r}_m) + \sum_\alpha \frac{Z_\alpha}{|\boldsymbol{r}_m - \boldsymbol{R}_\alpha|}, \quad (1)$$

where I and J are indices of the fragments, and α is index of the atom. The φ^I is a direct contribution from the monomer, and $\Delta\varphi^{IJ}$ is a two-body correction from the dimer. They are calculated as

$$\phi^I(\boldsymbol{r}_m) = \sum_{\mu\nu\in I} D^I_{\mu\nu} u_{\mu\nu}(\boldsymbol{r}_m), \qquad \Delta\phi^{IJ}(\boldsymbol{r}_m) = \sum_{\mu\nu\in I,J} \Delta D^{IJ}_{\mu\nu} u_{\mu\nu}(\boldsymbol{r}_m), \quad (2)$$

where μ and ν are indices of the basis functions. The matrix $\boldsymbol{D^I}$ is the density matrix of the monomer, $\Delta\boldsymbol{D^{IJ}}$ is the difference density matrix[2], and $u_{\mu\nu}$ is a one-electron integral like a nuclear attraction integral. The numerical errors of the ESP associated with FMO scheme were examined at HF, MP2, and RI-MP2[3,4] levels of theory. As a result, the FMO errors in ESP were significantly smaller than the amplitude of the electron correlation effect, indicating that the FMO method provides sufficiently accurate electrostatic properties for chemical and biological researches.

Additionally, an attempt to reduce the computational effort was proposed by combining the FMO scheme and a point-charge approximation. When the distance between a position $\boldsymbol{r}_m$ and monomer or dimer is sufficiently large, the monomer contribution and two-body correction in equations (2) are evaluated using a point-charge approximation as the following equations:

$$\phi^I(\boldsymbol{r}_m) = \sum_{\alpha\in I} \frac{q^I_\alpha}{|\boldsymbol{r}_m - \boldsymbol{R}_\alpha|}, \qquad \Delta\phi^{IJ}(\boldsymbol{r}_m) = \sum_{\alpha\in I,J} \frac{\Delta q^{IJ}_\alpha}{|\boldsymbol{r}_m - \boldsymbol{R}_\alpha|}, \quad (3)$$

where $q_\alpha{}^I$ and $\Delta q_\alpha{}^{IJ}$ are a monomer contribution and a two-body correction to the atomic point-charge of the electron, respectively. The error of this approximation was examined using two proteins, prion protein and human immunodeficiency type 1 protease. Finally, as an illustrative example, ESP at the molecular surface of these proteins at MP2 level of theory were performed. In this study, PAICS program package[5] was used for all the calculations.

[1] K. Kitaura, et al, *Chem. Phys. Lett.*, **313** (1999) 701
[2] T. Nakano, et al, *Chem. Phys. Lett.*, **351** (2002) 475
[3] T. Ishikawa and K. Kuwata, *Chem. Phys. Lett.*, **474** (2009) 195
[4] T. Ishikawa and K. Kuwata, *J. Phys. Chem. Lett.*, **3** (2012) 375
[5] T. Ishikawa, et al, *J. Comput. Chem.*, **30** (2009) 2594

Mechanism analysis of estrogen receptor beta selective agonist by molecular dynamics and fragment molecular orbital method

Kensuke Misawa[1]
misawa.kensuke@kao.co.jp

Tadahiro Ozawa[1]
ozawa.tadahiro@kao.co.jp

Yoshiya Sugai[1]
sugai.yoshiya@kao.co.jp

Taketoshi Fujimori[1]
fujimori.taketoshi@kao.co.jp

Takatsugu Hirokawa[2,3]
t-hirokawa@aist.go.jp

[1] Biological Science Laboratories, Kao Corporation, 2606 Akabane, Ichikai-machi, Haga-gun, Tochigi 321-3497, Japan
[2] Molecular Profiling Research Center for Drug Discovery (molprof), National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan
[3] Division of Biomedical Science, Faculty of Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8575, Japan

Keywords: Estrogen receptor, Ligand docking, Molecular dynamics, Fragment molecular orbital

Estrogen receptors (ERs) are nuclear transcription factors that play an important role in the regulation of various physiological processes in humans and located in the various tissues such as breast, bone and skin[1]-[2]. There are two major subtypes of ERs: estrogen receptor alpha (ERα) and estrogen receptor beta (ERβ) [1]. Estradiol binds both the ERs equally, but this often leads to an increased risk of breast and endometrial cancers because of ERα agonism[2]-[3]. Therefore, many ERβ selective agonists have been developed so far [4].

We focused on ERβ selective agonist *meso*-2,3-bis(4-hydroxyphenyl)succinonitrile[5] and analysed the mechanism of ERβ selectivity by the combination of induced-fit docking, molecular dynamics and fragment molecular orbital. The details will be argued on the day.

[1] Paterni, I., Granchi, C., Katzenellenbogen, J. A., and Minutolo, F., Estrogen receptors alpha (ERα) and beta (ERβ): Subtype-selective ligands and clinical potential, *Steroids*, 90:13-29, 2014.
[2] Eyster, K. M., *The Estrogen Receptors: Methods and Protocols*, Humana Press, 2016.
[3] Jordan V. C., The secrets of selective estrogen receptor modulation: cell-specific coregulation, *Cancer Cell,* 1:215-217, 2002.
[4] Mohler, M. L. *et al.*, Estrogen receptor β selective nonsteroidal estrogens: seeking clinical indications, *Expert Opinion on Therapeutic Patents,* 20:507-534, 2010.
[5] Meyers, M. J. *et al.*, Estrogen Receptor-β Potency-Selective Ligands: Structure-Activity Relationship Studies of Diarylpropionitriles and Their Acetylene and Polar Analogues, *J. Med. Chem.,* 44:4230-4251, 2001.

P1-22

Study of BACE1/2 selectivity mechanism for highly selective inhibitor using metadynamics

Naoya Asada[1] naoya.asada@shionogi.co.jp
Kazunari Hattori[1] kazunari.hattori@shionogi.co.jp
Ken-ichi Kusakabe[1] ken-ichi.kusakabe@shionogi.co.jp

[1] Shionogi Pharmaceutical Research Center, SHIONOGI & CO., LTD., 3-1-1, Futaba-cho, Toyonaka-shi, Osaka 561-0825, Japan

Keywords: Selectivity, Molecular Dynamics, Metadynamics, Mutation analysis

BACE1 (beta-site APP cleaving enzyme 1) is one of the aspartic proteases and protein targets for Alzheimer's disease. Selectivity against BACE2 is desired for BACE1 inhibitor to avoid potential side effects. Highly selective inhibitor of BACE1 was reported [1], however the mechanism of selectivity has been unknown. We performed metadynamics simulation using Desmond [2] for the elucidation of the selectivity mechanism of this compound.

It was noted that a shift of flap loop region, which is a part of the binding pocket, to a more open orientation was occurred when the selective inhibitor bound [1] and comparison of X-ray crystal structures between BACE1 and BACE2 revealed that the flap orientation was different between them. Based on this information, we hypothesized that flap motion got involved in the selectivity. To validate this hypothesis, metadynamics simulation was performed with the setting of collective variables as the distance between catalytic region and the flap for the calculation of Free Energy Surface (FES). As a result, it was observed that FES was different between two proteins. In BACE1, free energy kept stable from close to open orientation of the flap, on the other hand, free energy became high when the flap was open in BACE2. This result showed that BACE2 was more unstable than BACE1 when the selective inhibitor bound with the flap opening and this could be the main factor for BACE1/2 selectivity. We also performed protein mutation analysis in virtual model for identifying amino acid residues responsible for the difference of FES. On the flap region, four residues are different between BACE1 and BACE2 (BACE1/BACE2: Y68/T84, P70/K86, K75/S91, E77/T93). We made four BACE2 models which mutated one of the four residues to that correspond to BACE1 and FES was calculated using metadynamics simulation for each model. Finally, we found that FES of two mutation models (T84Y and S91K) showed similar results with those of BACE1, which indicated that these two residues might contribute to the difference of the flap motion.

[1] Michael S. Malamas, Keith Barnes, Matthew Johnson, Yu Hui, Ping Zhou, Jim Turner, Yun Huc, Erik Wagner, Kristi Fan, Rajiv Chopra, Andrea Olland, Jonathan Bard, Menelas Pangalos, Peter Reinhart, Albert J. Robichaud, *Bioorg. Med. Chem*. **2010**, *18*, 630–639.
[2] Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2016. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2016.

Comparing two molecular dynamics simulation trajectories in terms of residue-residue interaction

Chie Motono c.motono@aist.go.jp
Takatsugu Hirokawa t-hirokawa@aist.go.jp

Molecular Profiling Research Center for Drug Discovery (molprof), National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo, 135-0064, Japan

Keywords: molecular dynamics simulation, Protein, Mutation, residue-residue interaction.

Molecular Dynamics (MD) simulations provide hi-resolution information on biomolecular dynamics. It has been used to elucidate the effects of mutations to a protein molecule, or binding/dissociation of chemicals to biomolecules on the motion and function of the molecules at a atomic level.

It is non-trivial a task to compare the two MD trajectories with different conditions (ex. with or without mutations, binding other molecules) and to detect the fundamental differences. As initial and intuitive analyses, time evolutions of root mean square deviations (RMSD), root mean square fluctuations (RMSF) of each position, and average structures of each trajectory are often compared. To observe statistical differences in molecular motions, Principal Component Analysis (PCA) is used.[1] It was proposed to use a more sophisticated technique, Partial Least Square for Discriminant Analysis (PLS-DA) for comparison of two MD trajectories.[2] These statistical methods suffer from the anisotropy of the molecular structures. Linear Discriminant Analysis with ITERative procedure (LDA-ITER) is an excellent method which maximizes the ratio of the between-ensemble fluctuation compared to the within-ensemble fluctuation to diminish the anisotropy effect.[3]

In this study, we propose a new approach to compare the fluctuations of residue-residue interactions of two MD trajectories. Identification of the positions where the dynamics of residue-residue interactions differs between two trajectories leads to the determination of the cause of different molecular motions or functions. This method indexes the residues with which a residue interacts and the durability of the interactions within a trajectory, and then compare the index with that in another trajectory to give a similarity. We applied the procedure to compare the trajectories of a wild-type protein and its mutant, and identify the most affected positions. This method is versatile and applicable to any different conditions like mutations, binding of other molecules, or unfolding, etc.

[1] Kitao, A. and Go,N., Investigating protein dynamics in collective coordinate space, *Current Opinion in Structural Biology*, 9:164-169, 1999.
[2] Peters, J. H. and de Groot, B. L., Investigating protein dynamics in collective coordinate space, *PLoS Comput. Biol.*, 8: e1002704, 2012.
[3] Sakuraba, S. and Kono, H., Spotting the difference in molecular dynamics simulations of biomolecules, *J. Chem. Phys.*, 145074:116, 2016.

P1-24

An integrated workflow for accurately predicting standard free energy of binding

Yoshiaki Tanida
tanida.yoshiaki@jp.fujitsu.com

Azuma Matsuura
matsuura.azuma@jp.fujitsu.com

Fujitsu Laboratories Ltd., Atsugi, Kanagawa 243-0197, Japan

Keywords: standard free energy, alchemical free energy calculation, metadynamics, cluster analysis, reweighting

To realize computational “de novo” drug design (e.g. w/o structural information in ligand-receptor complex), we propose a new strategy of binding affinity evaluation and binding structure search in combination.

(i) Well-tempered metadynamics is used for exploring the available binding structures of new drug-like molecule to a target in a suitable few collective variables space.
(ii) Some “pose” of the ligand providing local free energy minima is obtained by cluster analysis and reweighting.
(iii) “Local” standard free energies can be estimated by alchemical free energy calculations; the standard free energy ligand binding is obtained.

We also introduce the application of this procedure in some typical case. We believe that our workflow described here will become a promising way in rational drug discovery.

An efficient approach for finding fragment-binding conformations

Hiroyuki Sato[1]
shryk@jp.fujitsu.com

Yoshiaki Tanida[1]
tanida.yoshiaki@jp.fujitsu.com

Azuma Matsuura[1]
matsuura.azuma@jp.fujitsu.com

[1] Fujitsu Laboratories Ltd., 10-1 Morinosato-Wakamiya, Atsugi 243-0197, Japan

Keywords: Protein pocket, Fragment molecule, Binding conformation, Non-bonding potential

We present a novel and efficient approach for finding various fragment-binding conformations to a protein using molecular dynamics (MD) simulation. In general, a long-time MD simulation is required to get possible binding variations in the conventional approaches[1]. Our approach adopts randomly exchangeable protein replicas with high concentration of fragment molecules whose non-bonding potential functions are different from each other. We applied this approach to human coagulation factor Xa[2] with 4 M concentration of 6-chloro-1-benzothiophene-2-ol, and found that fragment molecules bound to protein pocket more than 60 times during a 10-ns MD simulation. This indicates that our approach provides a sufficient number of fragment-binding conformations through a short-time MD simulation.

[1] Guvench, O., *et al*., *PLoS Comput. Biol.* **2009**, 5, e1000435.
[2] Maignan, S., *et al*., *J. Med. Chem.* **2003**, *46*, 685-690.

Computational Study on Structure-Activity Relationship in FAS Inhibitors Based on Three-Dimensional Electronic Similarity

Takafumi Inoue[1]
175d8303@st.kumamoto-u.ac.jp

Toshihiro Ideo[1]
152d9201@st.kumamoto-u.ac.jp

Manabu Sugimoto[1,2,3]
sugimoto@kumamoto-u.ac.jp

[1] Graduate School of Science and Technology, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan

[2] Faculty of Advanced Science and Technology, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan

[3] Research Center for Advanced Science and Technology (RCAST), University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, Japan

Keywords: Structure-activity relationship, FAS Inhibitor, Electronic similarity

The fatty acid synthase (FAS) inhibitor is expected to work as an anti-cancer drug. Several candidates as the FAS inhibitor have experimentally suggested by Wang et al. [1] and Li et al. [2-3]. In order to investigate the structure-activity relationship of these molecules, we are developing a computational method evaluating three-dimensional topology similarity where the topology is obtained by using electronic-structure calculations. We call this similarity "three-dimensional electronic-shape similarity (3D-ESS)". Herein we focus on electro-static potential (ESP) as an example of 3D-ESS. In the presentation, we will introduce a program for scoring ESP similarity among the compounds. We will discuss correlation between the similarity score and IC_{50} for the FAS inhibitors through evaluation of 3D-ESS. The combined use of 3D-ESS with energy-based electronic- similarity (EB-ES) will also be discussed.

[1] Wang, X., Song, K. S., Guo, Q. X., and Tian, W. X., The galloyl moiety of green tea catechins is the critical structural feature to inhibit fatty acid synthase, *Biochem. Pharma.,* 66: 2039-2047, 2003.

[2] Li, B. H., and Tian, W. X., Inhibitory effects of flavonoids animal fatty acid synthase, *J. Biochem.,* 135: 85-91, 2004.

[3] Li, B. H., Ma, X. F., Wang, Y., and Tian, W. X., Structure-activity relationship of polyphenols that inhibit fatty acid synthase, *J. Biochem.,* 138: 679-685, 2005.

Analysis of Ras/Raf-RBD Complex by Molecular Dynamics Simulation

Shota Matsunaga
matsunaga@eniac.scitec.kobe-u.ac.jp

Shigenori Tanaka
tanaka2@kobe-u.ac.jp

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

Keywords: Ras/Raf-RBD complex, cancer, molecular dynamics (MD) method, GTPase-activating proteins (GAPs)

The Ras/Raf/MEK/ERK signal transduction pathway plays an important role in controlling cell proliferation and differentiation[1]. Abnormal activation of Ras or Raf genes has frequently been reported in human cancer[2]. In this study, to elucidate the mechanism of the Ras/Raf interaction, we investigate the interconversion between two conformational states, State 1 and State 2[3], of the Ras/Raf-RBD complex using the molecular dynamics (MD) method. In particular, we analyze the two switch regions (Switch Ⅰ and Switch II) of both GTP-bound form and GDP-bound form. To accelerate the conformational change of GDP-bound form, the GDP molecule was given the thermal energy generated when the GTP molecule was hydrolyzed by GTPase-activating proteins (GAPs)[4].

As a result of calculating *dT* (the distance between the oxygen atom of Thr35 of Ras and either the γ-phosphorus atom of GTP or the β-phosphorus atom of GDP) and *dY* (the oxygen atom of Tyr64 of Ras and either the γ-phosphorus atom of GTP or the β-phosphorus atom of GDP)[5], it was found that Switch II is repeatedly opened and closed because of the allosteric property of binding in the standard MD simulation. On the other hand, we found that the thermal energy from hydrolysis of GTP is important for the Switch I to be opened sufficiently. From the results of the simulations with and without the added energy, we succeeded to obtain the comprehensive pictures of the signal transduction in the Ras/Raf-RBD complex.

References

[1] S. K. Fetics, H. Guterres, B. M. Kearney, G. Buhrman, B. Ma, R. Nussinov, C. Mattos, Allosteric Effects of the oncogenic RasQ61L mutant on Raf-RBD, *Struct. Des.*, 23: 3, 2015.

[2] L. Santarpia, S. L. Lippman, A. K. EI-Naggar, Targeting the mitogen-activated protein kinase Ras-Raf signaling pathway in cancer therapy, *NIH Pub.*, 16: 1, 2012

[3] S. Matsumoto, N. Miyano, S. Baba, J. Liao, T. Kawamura, C. Tsuda, A. Takeda, M. Yamamoto, T. Kumasaka, T. Kataoka, F. Shima, Molecular mechanism for conformational dynamics of Ras·GTP elucidated from In-Situ structural transition in crystal, *Sci. Rep.*, 6: 25931, 2016.

[4] K. Ogata, J. Shen, S. Sugawa, S. Nakamura, MD simulation study of Ras/Raf dissociation and the resonating structure of deactivated Ras, *Chem. Soc*., 85: 12, 2012.

[5] S. Matsunaga, Y. Hano, Y. Saito, K. Fujimoto, T. Kumasaka, S. Matsumoto, T. Kataoka, F. Shima, S. Tanaka, Structural transition of solvated H-Ras/GTP revealed by molecular dynamics simulation and local network entropy, *J. Mol. Graph. Model.*, 2017, in press.

P1-28

Energy correlation of protein at equilibrium fluctuation and its connection with ligand binding

Hikaru Iba[1] iba0119@cheng.es.osaka-u.ac.jp
Yuu Ymamori[1] yamamori@cheng.es.osaka-u.ac.jp
Nobuyuki Matubayasi[1] nobuyuki@cheng.es.osaka-u.ac.jp

[1] Division of Chemical Engineering, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan

Keywords: Molecular dynamics simulation, Ligand binding, Energy correlation

Protein exchanges the energy among constituent residues and the surrounding solvent. The mode of energy exchanges is then coupled to the structural fluctuation and can lead to the structural change due to an external perturbation such as ligand binding. In the present work, we investigate the correlations of intra-residue structural energy and the interaction energy with solvent water. We treat shikimate kinase and address the connection to the structural change upon binding of ATP ligand.

To sample a set of protein structures, all-atom MD simulations were conducted over 10 ns at a sampling interval of 1 ns in the *NPT* ensemble with 300 K and 1 bar. Each sampled structure was subject to the energy analyses, and the intra-residue structural energy and the interaction energy with solvent water (v) were computed at fixed structure of the protein in the ensembles of the apo and holo forms.

Figure shows the correlation coefficient of v among the residues of the apo and holo forms. The off-diagonal correlations are significant in the apo form in Figure (a) between residues 10-20 and 115-120, between residues 10-20 and 135-140, and between residues 115-120 and 135-140. The ATP binding does not affect the correlation features and strengthens the correlations for 10-30 and 105-110, for 10-20 and 145-155, and for 105-110 and 150-160. These residues correspond to the ATP binding sites. They come close at the binding, leading to the strengthened correlation of v.

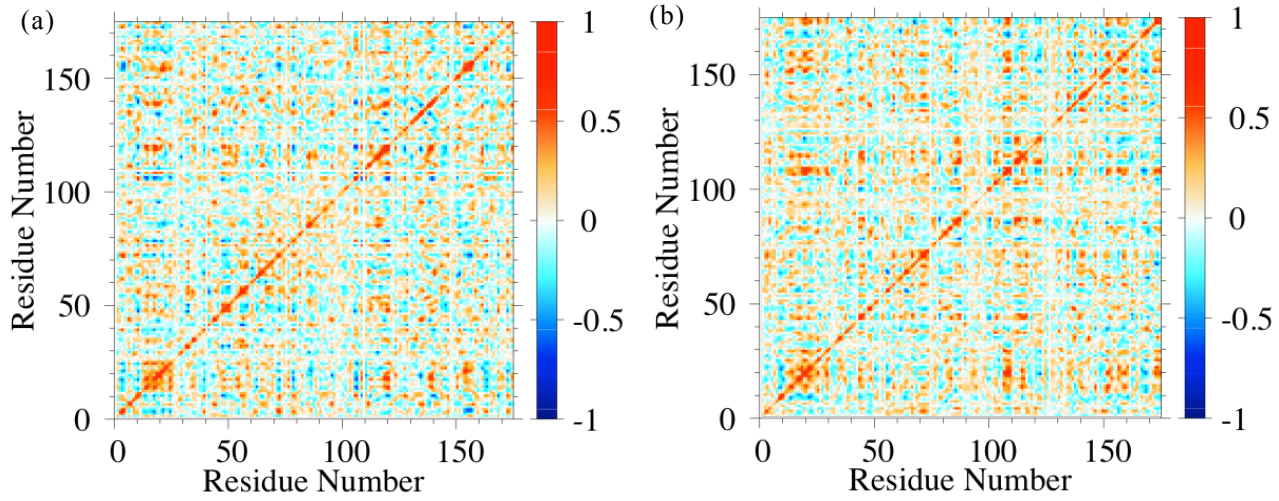


Figure. Energy correlation map of v among the residues of shimate kinase in the (a) apo and (b) holo states.

P1-29

In silico binding affinity analysis for phosphodiesterase-10A inhibitors.

Chisa Yuasa[1]
w175438e@yokohama-cu.ac.jp

Toru Ekimoto[1]
ekimoto@tsurumi.yokohama-cu.ac.jp

Tsutomu Yamane[1]
tyamane@tsurumi.yokohama-cu.ac.jp

Mitsunori Ikeguchi[1]
ike@tsurumi.yokohama-cu.ac.jp

[1] Graduate School of Medical Life Science, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama-shi, Kanagawa, Japan

Keywords: Ligand docking, Binding free energy calculation, Molecular Dynamics simulation

Phosphodiesterase (PDE) is an enzyme that hydrolyzes cyclic-AMP (c-AMP) and cyclic-GMP (c-GMP), which play important roles in signal transduction. PDE10A specifically expresses in brains, and PDE10A is considered as a drug target of schizophrenia.
In order to understand how the drug activity relates to interactions manner between PDE10A and ligand, we performed docking simulations and calculated the binding free energies between PDE10A and 24 ligands using 4 approaches: the Glide docking score, MM-GBSA, FEP+ and MP-CAFEE method. The calculated binding free energies were compared with experimental IC_{50}. Although no correlation between the Glide docking score and experimental IC_{50} was observed, the binding free energies calculated by MM-GBSA and FEP+ were correlated with IC_{50}.
In addition, we researched the docking-pose dependence of the binding free energy calculations. We examined effects of water molecules in the binding site on the docking poses. Inclusion of water molecules improved the correlation between the calculated binding free energy and IC_{50}. FEP+ showed the best performance among the methods considered here. From these examinations, we found two important points for the high correlation: precise docking poses of ligands and account of hydration water molecules in binding free energy calculations.

[1] J. Kunitomo, M. Yoshikawa, M. Fushimi, A. Kawada, J F. Quinn, H Oki, H. Kokubo, M. Kondo, K. Nakashima, N. Kamiguchi, K. Suzuki, H. Kimura, and T. Taniguchi, Discovery of 1-[2-Fluoro-4-(1H-pyrazol-1-yl)phenyl]-5-methoxy-3-(1-phenyl-1H-pyrazol-5-yl)pyridazin-4(1H)-one (TAK-063), a Highly Potent, Selective, and Orally Active Phosphodiesterase 10A (PDE10A) Inhibitor. J. Med. Chem. 2014, 57, 9627−9643
[2] L. Wang, B. J. Berne and A. Friesner On achieving high accuracy and reliability in the calculation of relative protein–ligand binding affinities. Proc. Natl. Acad. Sci. USA 2012, 109, 1937-1942
[3] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, and R. A. Friesner, OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J. Chem Theory Comput. 2016, 12, 281-296

Computational Determination of the Partial Charges of Neutral Sialic Acid

Makoto Ikejo[1]
ikejo@eniac.scitec.kobe-u.ac.jp

Hirofumi Watanabe[2]
hirofumi.watanabe@port.kobe-u.ac.jp

Shigenori Tanaka[1]
tanaka2@kobe-u.ac.jp

1 Graduate School of System Informatics, Department of Computational Science, Kobe University, 1-1, Rokkodai, Nada, Kobe 657-8501, Japan

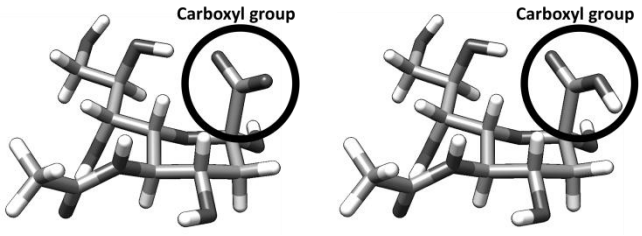
2 Education Center on Computational Science and Engineering, Kobe University, 7-1-48, Minatojimaminamimachi, Chuo-ku, Kobe 650-0047, Japan

Keywords: Electronic structure calculation, Force field, Glycan, Parameter fitting

The sugar chain is actively studied as the third "life molecule" in recent years. Molecular dynamics (MD) simulation is a useful tool to clarify the relationship between structure and function of biomolecules including proteins and sugar chains, where force field called GLYCAM [1] is introduced in MD software AMBER. However, in this force field, only carboxyl groups in deprotonated state are prepared, which are considered to be unsuitable for calculations in acidic solution.

Sialic acid is a monosaccharide having the carboxyl group and is known to play an important role in biological mechanisms such as virus infection. Here we report the determination of the atomic charges in the neutral state of this sialic acid (*N*-acetylnueraminic acid). Furthermore, we discuss the difference in the atomic charges between neutral and deprotonated states, and its influence on MD results.

Sialic acid (*N*-acetylnueraminic acid)



Deprotonated state Neutral state

[1] Kirschner, K. N., *et al.*, GLYCAM06: a generalizable biomolecular force field. Carbohydrates, *J. Comput. Chem.*, 29(4):622-55, 2008.

P1-31

Time-dependent Variance-covariance Matrix Analysis of Molecular Dynamics for Protein-ligand Interaction

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 dynamics, Protein ligand interaction, Variance-covariance matrix, Random matrix, Cluster analysis

We study the protein-ligand interaction using the characteristic scale of coupling formation [1] developed in random matrix theory. The eigensystems of the timescale-dependent variance-covariance matrix, which are obtained from the time series data of the atomic coordinates of a protein produced by the all-atom molecular dynamics of the solvent, are analyzed. [2] As an example, we present a result for a protein, N-terminal PDZ domain of ZASP in complex with myotilin C-terminal peptide, PDBID:4YDP, which has a ligand of peptide with three amino acids. We find that there are at least four different time scales involved in the coupling formation of correlated sectors of atoms and an oscillation of the eigenvalue as a function of the sampling interval of the variance-covariance matrices. These time scales and the oscillation behavior coexist simultaneously. From the study of the eigenvector components, the weights of the ligand atoms are dominant in a small time scale. We further compare the results with those without ligand and with the dataset which we artificially displace the coordinates of the ligand.

[1] For example see a review article, J. Kwapień and S. Drożdż, Phys. Rep. **515**, 115 (2012).
[2] M. Yamanaka, J. Phys. Soc. Jpn. 84, 063801 (2015).

In silico protein-protein interaction analysis of axon guidance molecule semaphorin and receptor plexin

Erena Shimoji[1]
eshimoji@tsurumi.yokoama-cu.ac.jp

Tsutomu Yamane[1]
yamane@tsurumi.yokohama-cu.ac.jp

Toru Ekimoto[1]
ekimoto@tsurumi.yokohama-cu.ac.jp

Nogi Terukazu[1]
nogi@tsurumi.yokohama-cu.ac.jp

Mitsunori Ikeguchi[1]
ike@tsurumi.yokohama-cu.ac.jp

[1] Graduate School of Medical Life Science, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Keywords: Semaphorin, Guidance molecule, Protein-protein interaction

Semaphorin has been known as an axon guidance molecule, secreted around cells. Plexin is a receptor of semaphorin, and their interactions are guidance cues for axons of neurons during development of the nervous system. Because the protein-protein interactions (PPI) of semaphorin and plexin have wide interaction surfaces, it is not trivial to understand the determinants of PPI. The surface of protein is too shallow and wide to bind small compound, so it is difficult to discover drug for PPI target. It is important to understand the mechanism to control PPI for drug discovery. Semaphorin and plexin include “Sema domain” with high sequence similarity, and they form PPI and interact through the sema domain. Therefore, we analyzed the PPI using *in silico* approaches to recognize the feature of specific interactions. More specifically, we used program Rosetta to calculate the binding free energy and to perform *in-silico* alanine scanning [1,2], and identified the “hotspots” which are significant amino acids in binding. Then, the details of the hotspot interactions were analyzed to understand the determinants of PPI.

[1] T. Kortemme and D. Baker, A simple physical model for binding energy hot spots in protein–protein complexes. Proc. Natl. Acad. Sci. 99, 14116-14121, 2002.
[2] T. Kortemme, D. E. Kim, and D. Baker, Computational Alanine Scanning of Protein-Protein Interfaces. Sci. STKE 10, 2004.
[3] B. J. C. Janssen, R. A. Robinson, F. Perez-Brangul , C. H. Bell, K. J. Mitchell, C. Siebold & E. Y. Jones, Structural basis of semaphorin–plexin signaling. Nature 467 1118-1122, 2010.

P1-33

Characterization of TCR-pMHC interaction at residue-level based on FMO and PIEDA

Hiromichi Tsurui[1] tsurui@juntendo.ac.jp
Yuko Tsuchiya[2] tsuchiya@protein.osaka-u.ac.jp
Yoshiki Namiuchi[3] Yoshiki.namiuchi@riken.jp
Hiroshi Wako[4] wako@waseda.jp

[1] Department of Pathology, Juntendo University School of Medicine, Bunkyo-ku, Tokyo, Japan
[2] Institute for Protein Research, Osaka University, Suita, Osaka, Japan
[3] RIKEN Quantitative Biology Center, Suita, Osaka, Japan
[4] School of Social Sciences, Waseda University, Shinjuku-ku, Tokyo, Japan

Keywords: TCR-pMHC, FMO, PIEDA, CH-π, hydrogen bond,

Interaction between T cell receptor (TCR) and peptide-MHC (pMHC) is essential for determining T cell fate and immunological response. However, determining actual binding profile of TCR by experiment requires vast amount of labor and by computer simulation is still a challenge. To investigate what interactions determine binding profile of TCR, we examined TCR-pMHC complexes having common pMHC and different TCRs with variety of binding profile, namely narrow (PDBID: 3c5z, 3rdt, 4p5t), modestly (3c6l), and highly cross-reactive (3c60). TCRs with specific profile formed more hydrogen bonds (about 10) between CDRs to presented peptide than that with cross-reactive profile (about 6 to 7), suggesting the contribution of hydrogen bond to the specific interaction. CH-p interaction seemed between CDRβ3 aromatic residues and MHCα helix in both complexes containing specific and cross-reactive TCRs. Both hydrogen and CH-p bonds are principally determined geometrically and actual interaction-energies are obscure. We examined those bonds with FMO calculation at MP2/6-31G* level and PIEDA.

A QM/MM study on binding affinity computation of tankyrase 2-ligand system

Yoshinori Hirano[1] hirano@riken.jp
Takao Otsuka[1] totsuka@riken.jp
Noriaki Okimoto[1] okimoto@riken.jp
Makoto Taiji[1] taiji@riken.jp

[1] QBiC, RIKEN, QBiC Building B, 6-2-4 Furuedai, Suita, Osaka 565-0874, Japan

Keywords: Binding affinity computation, QM/MM method

Accurate binding affinity computation of protein-ligand system is of significant importance for the drug discovery. There are various computational methods to predict the protein-ligand binding affinity. One of the most popular approach for protein-ligand binding affinity is the free energy calculation based on molecular dynamics (MD) simulation, such as the linear interaction energy method, the molecular mechanics/Poisson-Boltzmann and surface area (MM-PB/SA) method, the alchemical free energy calculation, and so on. Most of these methods use classical molecular mechanics (MM) force fields.

Protein-ligand interactions consist of various non-bonded interactions such as π-stacking, charge transfer, polarization effects, and dispersion interaction, as well as the usual and/or weak hydrogen bonds. Therefore, it is difficult to treat these interactions by the classical MM force fields. Quantum mechanical (QM) calculations can consider these non-bonded interactions. We have expected to improve the accuracy of protein-ligand binding affinity compared with that of the MM force fields. In our previous study, we focused on the evaluation of binding affinity through a full QM calculation and proposed the accurate and efficient computational scheme by using multilayer fragment molecular orbital (MFMO) method[1].

In this study, we explore a computational-cost effective scheme to calculate protein-ligand binding affinity by using QM/MM method. One of the key points is how the size of the QM region influences the performance of binding affinity prediction. We will report the binding affinities of tankyrase 2-ligand systems using QM/MM method[2-4] with some different QM regions.

[1] Otsuka, T.; Okimoto, N.; Taiji, M., J. Comput. Chem., 2015, 36, 2209-2218.
[2] Schrödinger Release 2017-2: QSite, Schrödinger, LLC, New York, NY, 2017.
[3] Murphy, R. B.; Philipp, D. M.; Friesner, R. A., "A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments," J. Comp. Chem., 2000, 21, 1442-1457
[4] Philipp, D. M.; Friesner, R. A., "Mixed ab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide," J. Comput. Chem., 1999, 20, 1468-1494

P1-35

The distinct dynamic mechanisms of peptide recognition by specific and cross-reactive T-cell receptors

Yuko Tsuchiya[1]
tsuchiya@protein.osaka-u.ac.jp

Yoshiki Namiuchi[2]
yoshiki.namiuchi@riken.jp

Hiroshi Wako[3]
wako@waseda.jp

Hiromichi Tsurui[4]
tsurui@juntendo.ac.jp

[1] Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
[2] RIKEN Quantitative Biology Center, 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan
[3] School of Social Sciences, Waseda University, 1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050, Japan
[4] Department of Pathology, Juntendo University School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan

Keywords: Cross-reactive TCR recognition, binding affinity-interface area relationship, Molecular dynamics (MD) simulation, Fragment molecular orbital (FMO) method

Some T-cell receptors (TCRs) recognize an antigen peptide presented by Major Histocompatibility Complex (pMHC) specifically, whereas others recognize it cross-reactively. We focused on five TCRs in complex with the same class II MHC-peptide, three and two of which recognize the peptide specifically and cross-reactively, respectively. These complexes have similar binding affinities, whereas the interface areas between TCR and pMHC differ significantly. In this study, we investigated static and dynamic structural features of the TCR-pMHC complex and of TCR in free state, and the relationship between binding affinity and interface area. Our dynamics study revealed differences in the recognition mechanisms of TCRs, i.e., differences in interaction types and balances of enthalpy gain and entropy loss upon binding.

Interaction Analysis between Beta-Secretase and its Inhibitors by Fragment Orbital Method

Norihito KAWASHITA[1] kawashita@life.kindai.ac.jp
Yuji HASHIMOTO[2] yuji71070516@gmail.com
Hirotomo MORIWAKI[2] hirotomo.moriwaki@gmail.com
Yu-Shi TIAN[2] yushi-tian@phs.osaka-u.ac.jp
Tatsuya TAKAGI[2] ttakagi@phs.osaka-u.ac.jp

[1] Faculty of Science and Engineering, Kindai University, 3-4-1 Kowakae, Higashiosaka City, Osaka 577-8502, Japan
[2] Graduate School of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565-0871, Japan

Keywords: Beta-Secretase, Fragment molecular orbital (FMO) method, Inter-fragment interaction energy (IFIE)

Beta-Secretase (BACE) 1 is an aspartyl protease, it's active site is made up of two aspartate residues: Asp32 and Asp228. The β-hairpin loop over the active site, known as the "flap", is also an important feature to BACE1. While the active site remains inactive, the flap stays in its open conformation. However, the flap is stabilized while closed over its substrate or some inhibitor [1].

Currently, aminoquinolines, aminoisoquinolines, and aminopyrimidone with cyclic-amidine structure were reported as potent fragment hits binding to BACE1 [2]. Since then, several studies on similar inhibitors with cyclic-amidine structure have been reported.

In this study, we performed FMO calculation to examine the molecular interactions between BACE1 and its ligands with cyclic-amidine structure for identified the key information for screening of more potent inhibitors.

Cocrystal structures of ligands and BACE1 were obtained from PDB. MOE [3] was used for structure preparation. MMFF94x force field was used for energy minimizations. FMO calculations were carried out in ABINIT-MP (Ver. 6.0+) [4] with MP2/6-31G*.

From the results, we confirmed the correlation between IC_{50} and IFIE-SUM (IFIE between ligand and receptor). The coefficient of determination R^2 was 0.728, suggesting a good correlation. Next, Asp28 and 228's IFIE notably contributed to the stabilization of protein-ligand interaction. And we detected tendency that ligands without aromaticity contributed to stabilization more than ligands with aromaticity. Therefore, when we design BACE1 inhibitors with cyclic-amidine structure, it is better that we choose structure without aromaticity.

In this study, we performed FMO method to examine the molecular interactions between BACE1 and its ligands. Not only good correlation between IC_{50} and IFIE-SUM but also interaction tendency in BACE1-ligand complex were obtained. Hereafter, we will execute further calculations and more detailed IFIE analyses. Eventually, we aim at contribution to design novel inhibitors.

This research was done in activities of the FMO drug design consortium (FMODD). The results were obtained using the K computer (project ID: hp150160, hp160103, and hp170183). PIEDA calculation was done by using MIZUHO/BioStation software package.

[1] Lin *et al.*, *Science*, 290: 150-153, 2008.
[2] Congreve *et al.*, *J. Med. Chem*. 50: 1124-1132, 2007.
[3] Chemical Computing Group, MOE. 2015.
[4] Nakano *et al.*, *Chem. Phys. Lett*. 351: 475-480, 2002.

In silico modeling of PAX8–PPARγ fusion protein with unknown three-dimensional structure in follicular thyroid adenoma and carcinoma

Kaori Sakaguchi[1]
sakaguchi@eniac.scitec.kobe-u.ac.jp

Yoshio Okiyama[2]
okiyama@nihs.go.jp

Shigenori Tanaka[1]
tanaka2@kobe-u.ac.jp

[1] Department of Computational Science, Graduate School of System Informatics, Kobe University, 1-1, Rokkodai, Nada, Kobe, Hyogo 657-8501, Japan
[2] National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan

Keywords: Follicular neoplasm, fusion protein, protein modeling, Molecular dynamics simulation

Follicular neoplasms including follicular thyroid adenoma (FTA) and follicular thyroid carcinoma (FTC) have no molecular-targeted agents. Specifically in these tumor cells, a t(2,3) (q13,p25) chromosomal translocation results in production of paired-box gene 8 (PAX8)-peroxisome proliferator-activated receptor γ gene (PPARγ) fusion protein (PPFP). While the three-dimensional (3D) structure of PPFP and its functions are not reported so far, a transgenic mouse model shows that pioglitazone, an effective agonist of PPARγ, makes PPFP behave as PPARγ-like transcription factor; it turns tumor cells into adipocytes and deprives their malignant characters [1]. Based on this finding, it is possible to obtain more effective anticancer agents for FTA and FTC if we can design compounds specifically bound to PPFP.

We here propose a 3D structure of PPFP for establishment of targeted therapies in FTA and FTC. We constructed amino acid sequence of PPFP that was composed of PAX8 (Met1-Ala396) and PPARγ (Met1-Tyr505) based on the currently proposed schematic diagram of PPFP [2]. In PDB, there are X-ray crystal structure of PAX6 paired domain with DNA (PDB ID: 6PAX) having high homology with PAX8 and the intact PPARγ on DNA (PDB ID: 3DZY). Using them, we assumed a fusion construct containing both PAX8 and PPARγ on DNA, and predicted the structure of PAX8 and the fusing region by I-TASSER [3] online servers. After some modification and relaxation, the final 3D structures were obtained in this study. To validate the predicted models, we analyzed their physicochemical characteristics, Ramachandran plot distributions and dynamical stabilities with molecular dynamics simulation.

[1] B. Xu *et al.*, J. Biol. Chem. 291, 19274-19286 (2016).
[2] P. Raman *et al.*, Nat. Rev. Endocrinol. 10, 616-623 (2014).
[3] J. Yang *et al.*, Nat. Methods. 12, 7-8 (2015).

Computational characterization of influenza A virus hemagglutinin recognition sites by a cross-neutralizing monoclonal antibody

Manabu Igarashi[1,2] igarashi@czc.hokudai.ac.jp
Masakazu Sekijima[3] sekijima@gsic.titech.ac.jp
Nobuaki Yasuo[3] yasuo.n.aa@m.titech.ac.jp
Takashi Abe[4] takaabe@ie.niigata-u.ac.jp
Teruaki Watabe[5] twatabe-mi@umin.ac.jp
Rashid Manzoor[1] manzoor@czc.hokudai.ac.jp
Takatsugu Hirokawa[6] t-hirokawa@aist.go.jp
Ayato Takada[1,2] atakada@czc.hokudai.ac.jp

[1] Research Center for Zoonosis Control, Hokkaido University, Sapporo, Hokkaido, Japan
[2] Global Station for Zoonosis Control, Global Institution for Collaborative Research and Education (GI-CoRE), Hokkaido University, Sapporo, Hokkaido, Japan
[3] Global Scientific Information and Computing Center, Tokyo Institute of Technology, Meguro-ku, Tokyo, Japan
[4] Department of Information Engineering, Niigata University, Niigata, Japan
[5] Center of Medical Information Science, Kochi Medical School, Kochi University, Kochi, Japan
[6] Molecular Profiling Research Center for Drug Discovery (molprof), National Institute of Advanced Industrial Science and Technology (AIST), Koto-ku, Tokyo, Japan

Keywords: Molecular dynamics, Influenza virus, Antibody, Protein modeling, Binding free energy

Background
The hemagglutinin (HA) of influenza A viruses is classified into 16 subtypes (H1-H16). It is generally known that HA-specific antibodies have little cross-neutralizing activity against multiple HA subtypes. Recently, however, several broadly neutralizing antibodies were reported and have attracted attention due to their potential application to therapeutics and vaccine design. We have previously reported a cross-reactive antibody, designated S139/1, which neutralizes H1, H2, H3, H13, and H16 subtypes and its crystal structure in complex with the HA of the A/Victoria/3/1975 (H3N2) strain. However, detailed structural basis of its cross-neutralizing activity still remain to be elucidated. In this study, we characterized the S139/1 recognition sites on different HAs using computational structural biology methods.

Materials and Methods
The HAs from eight strains (subtypes H1, H2, H3, H6, H9, and H13) were analyzed. The structure models of the HA-S139/1 complex were constructed by homology modeling. Using the structures as starting points, we performed molecular dynamics (MD) simulations, and then calculated binding free energies (ΔG) between S139/1 and each HA.

Results and Discussions
The ΔG values of the strains neutralized by S139/1 were lower than the other strains tested. We next investigated the contribution of individual residues on each HA to the interaction with S139/1 and found that amino acids at positions 98, 136, 153, 156, 158, 159, 193, 194, 196, and 226 (H3 numbering) on HA strongly contributed to S139/1 binding as for the strains neutralized by S139/1. Additionally, we found that, in the strains neutralized by S139/1, K156 formed a salt bridge with the Ig heavy chain E50. Indeed, amino acid substitutions at adjacent residues were experimentally observed in the mutant viruses escaping from neutralization by S139/1. Thus, our computational methods identified the amino acid residues critical for the cross-neutralizing activity of S139/1.

Characteristics of Biomolecule Dynamics under the Crowding Environment of Cytoplasm Discovered by Massive All-atom simulation and Big-data analysis

Isseki Yu[1,2] isseki@riken.jp
Po-hung.wang[2] po-hung.wang@riken.jp
Michael Feig[5] feig@msu.edu
Yuji Sugita[1,2,3,4] sugita@riken.jp

[1]iTHES Research Group, RIKEN, Saitama, Japan
[2]Theoretical Molecular Science Laboratory, RIKEN, Saitama, Japan
[3]Laboratory for Biomolecular Function Simulation, RIKEN Quantitative Biology Center, Kobe, Japan
[4]Computational Biophysics Research Team, RIKEN Advanced Institute for Computational Science, Kobe, Japan
[5]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, United States

Keywords: Cellular environment, Molecular crowding, Protein diffusion, Metabolites

Inside of a cell is highly crowded with a large number of macromolecules together with solvents and metabolites. How variable specific/non-specific interactions within dense cellular environments may affect the structure and dynamics, and ultimately function is one of the most fundamental questions in life science. We constructed full atomistic model of the cytoplasm of bacteria and various levels of protein crowding models. Using these model, we performed massive all-atom molecular dynamics (MD) simulation with the highly parallelized MD simulator GENESIS on K-computer. Influence of crowding and non-specific interaction between macromolecules on the translational/rotational/collective motion of proteins are analyzed. Anomalous diffusion and the localization of metabolites, such as ATP and amino acids, on the macromolecule surface are also investigated. These work is an important step towards physically realistic in silico whole-cell models that connect molecular with cellular biology.

[1] Michael Feig, Isseki Yu, Po-Hung Wang, Grzegorz Nawrocki, Yuji Sugita, Crowding in Cellular Environments at an Atomistic Level from Computer Simulations, *Journal of Physical chemistry B* (in press)
[2] Po-hung Wang, Isseki Yu, Michael Feig, Yuji Sugita, Influence of Protein Crowder Size on Hydration Structure and Dynamics in Macromolecular Crowding, *Chemical Physics Letters* 671, 63-70, (2017)
[3] Isseki Yu, Takaharu Mori, Tadashi Ando, Ryuhei Harada, Jaewoon Jung, Michael Feig, Yuji Sugita, Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm, *eLife* 2016;5:e19274
[4] Michael Feig, Ryuhei Harada, Takaharu Mori, Isseki Yu, Koichi Takahashi, Yuji Sugita, Complete atomistic model of a bacterial cytoplasm for integratingphysics, biochemistry, and systems biology *Journal of Molecular Graphics and Modelling*, 58, 1–9, (2015)

P1-40

Molecular Dynamics Simulations of HIV Tat proteins and Amyloid-β peptides

Kazumi Omata[1]
komata@hosp.ncgm.go.jp

Satoru G. Itoh[2]
itoh@ims.ac.jp

Hisashi Okumura[2,3]
hokumura@ims.ac.jp

[1] Department of Data Science, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku, Tokyo 162-0052, Japan
[2] Institute for Molecular Science, 38 Nishigonaka, Myodaiji, Okazaki, Aichi 444-8585, Japan
[3] The Graduate University for Advanced Studies, 38 Nishigonaka, Myodaiji, Okazaki, Aichi 444-8585, Japan

Keywords: HIV Tat protein, Amyloid-β, Ligand docking, Molecular dynamics, neurocognitive dysfunction

To identify binding sites of an HIV transactivator of transcription (Tat) protein and an amyloid-β (Aβ) peptide, the present study has investigated their docking, using molecular dynamics (MD) simulations.
Recently, the experiments by Hategan et al. suggested that HIV Tat binds to the exterior surfaces of Aβ fibrils, increasing β-sheet formation and lateral aggregation into thick multifibrillar structures [1]. As a result, the fibers increase their rigidity and mechanical resistance. They also showed that the aggregate of Tat-Aβ complexes synergistically induced neurotoxicity both *in vitro* and in animal models. Apart from such a study, Okumura and Itoh examined the structure of Aβ fibrils by all-atom MD simulations, and demonstrated that structural fluctuations at the fibril ends play an important role for the fibril formation [2]. Based on the knowledge about the structure of Aβ fibrils derived from their MD simulations, the present study has attempted to find which parts of Aβ peptides are tightly associated with HIV Tat proteins, using MD simulations.
The all-atom MD simulations were carried out using the Generalized-Ensemble Molecular Biophysics (GEMB) program [3]. The AMBER99SB force field was used for the HIV Tat and Aβ, and the TIP3P rigid-body model was used for the water molecules. The electrostatic potential was calculated by the particle mesh Ewald method. Temperature was controlled at 298 K with the Nosé-Hoover thermostat, and pressure was controlled at 0.1 MPa with the Andersen barostat. The symplectic quaternion scheme was employed for the rigid-body water molecules.
The results of our MD simulations could be used for the inhibition of the Tat-Aβ complex formation and the development of drugs for neurocognitive dysfunction of HIV-infected individuals.

[1] Hategan, A., et al., HIV Tat protein and amyloid-β peptide form multifibrillar structures that cause neurotoxicity. *Nature Structural & Molecular Biology* **24**, 379–386, 2017.
[2] Okumura, H. and Itoh, S. G., Structural and fluctuational difference between two ends of A*β*amyloid fibril: MD simulations predict only one end has open conformations. *Scientific Reports* 6, Article number:38422, 2016.
[3] Okumura, H., Temperature and pressure denaturation of chignolin: Folding and unfolding simulation by multibaric-multithermal molecular dynamics method. *Proteins*, 80: 2397-2416, 2012.

An index to select appropriate homologous sequences for functional region prediction of a protein

Shoichiro Kato[1] 16rmb05@ms.dendai.ac.jp
Hiroyuki Toh[2] tohhir@kwansei.ac.jp

Wataru Nemoto[1] w.nemoto@mail.dendai.ac.jp

[1] Division of Life Science and Engineering, Graduate School of Science and Engineering, Graduate School of Tokyo Denki University (TDU), JAPAN.
[2] Department of Biomedical Chemistry, School of Science and Technology, Kwansei Gakuin University, JAPAN.

Keywords: sequence selection, phylogenetic tree, conservation score,

The conserved amino acid residues among homologous amino acid sequences are often closely located each other at and near the functional regions in the tertiary structures. Based on the observations, several methods have been developed to predict functional regions, in which both homologous amino acid sequences and tertiary structures are used for the analysis. We have also proposed a method with the same idea [1]. Given a target protein for the prediction, homologous sequences are collected under the condition that the sequence identity to the target protein is less than or equal to a given threshold value. Then, a multiple sequence alignment is made for the target protein and the collected sequences, and the conservation score is calculated at each alignment site. Next, the conservation score of each alignment site is mapped on the corresponding residue of a tertiary structure of one of the homologous proteins or the target protein. Finally, clusters of conserved residues are detected as functional regions in the tertiary structure.

In this strategy, a set of homologous sequences to calculate conservation score is required to contain the sequences with their functional regions at the corresponding sites on their structures. If a set of homologous sequences includes the sequences with their functional regions at distinct sites on their structures, it would be difficult to accurately evaluate residue conservation for this strategy. Hence, we introduced a new index based on the spatial autocorrelation to evaluate the spatial clustering of the conserved residues [1]. By changing the threshold value for the sequence collection from 90 % to 10 % with an interval of 5 %, different sets of the homologous sequences are generated, to each of which the new index is calculated. We selected a set of the homologous sequences with the maximum value of the index as the most appropriate set for the prediction because the set is appropriate to evaluate the spatial clustering of the conserved residues. Our method can not only predict the functional regions, but also determine the best set of the homologous sequences for the prediction.

Our method successfully improved the performance of functional region predictions [1]. As suggested in our manuscript, however, the collection of the homologous sequences based on the sequence identity to the target protein is too rough for such an approach, and the phylogenetic relationship among the collected sequences should be taken into account to generate a set of the homologous sequences for the prediction. In this work, hence, a set of homologous sequences was constructed as a cluster corresponding to a subtree in the phylogenetic tree of the target and its homologs. The performance of the prediction was evaluated by shifting the subtree from the closely related node to the distantly related node. The experimentally identified catalytic sites obtained in Catalytic Site Atlas was used for the performance evaluation. We will explain details of the procedure of our method and discuss advantages and disadvantages of the new approach.

[1] Nemoto, W. & Toh, H. *BMC Struct. Biol.* 12**,** 11, 2012

Influence of Molecular Force Field on Performance of Protein Ligand Docking

Daisuke Takaya[1] daisuke.takaya@riken.jp
Kikuko Kamisaka[1] kikuko.kamisaka@riken.jp
Shunpei Nagase[1] shunpei.nagase@riken.jp
Teruki Honma[1] honma.teruki@riken.jp

[1] RIKEN Center for Life Science Technologies, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan.

Keywords: SBDD, Protein-Ligand docking, Benchmark

Structure Based Drug Design (SBDD) is one of the most powerful approach when a target protein is known and its 3D-coordinates are experimentally determined by X-ray crystal structure analysis or can be constructed based on its template by using homology modeling. Since protein-ligand complex structures are not always available in drug design, protein-ligand docking programs (*e.g.* AutoDock Vina[1], FRED[2], Glide[3], and so on) are used for ligand binding mode prediction. Moreover, the docking programs are also useful for predicting its affinity as a docking score to the target protein, searching putative inhibitors from compound database and discriminating between active and inactive compounds. Docking programs are widely used for *in silico* screening by not only computational chemists but also biological researchers.[4,5]

Generally, docking programs provide some parameters and conditions which influence the output such as predicted binding modes and docking score of ligands. Therefore, it affects the prediction performance of docking programs. For example, preprocess of the receptor structure such as addition of hydrogen atoms, assignment of atom types of molecular force field, coefficients of interaction terms of the score function, exploring methods for molecular conformation and position. Additionally, their combination should be considered.

There are several approaches to handle the docking conditions. Firstly, the docking condition is performed under default setting. Secondly, if there is data of known inhibitors for the target protein, cross docking and test docking are performed for optimizing the conditions for the target. Finally, the conditions are determined by visual inspection of the prediction results and experience of computational researchers. It was thought that it would be useful for selecting docking setting if the influence of the parameters and conditions were known in advance. In this study, we examined the extent to which docking conditions affect the prediction performance by using benchmark sets. The results will give guidance on whether the condition should be changed or not from the default setting.

[1] Trott, O. *et al.*, AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading *J. Comput. Chem.* 31:455-461, 2010
[2] McGann, M. FRED Pose Prediction and Virtual Screening Accuracy *J. Chem. Inf. Model.*, 51(3):578-596, 2011
[3] Friesner, R. A. *et al.*, Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy *J. Med. Chem.* 2004, 47, 1739-1749
[4] Takaya, D., *et al.*, A new method for induced fit docking (GENIUS) and its application to virtual screening of novel HCV NS3-4A protease inhibitors, *Bioorg. Med. Chem.* 19:6892–6905, 2011
[5] Tajiri, H., *et al.*, Targeting Ras-Driven Cancer Cell Survival and Invasion through Selective Inhibition of DOCK1,*Cell Reports* 19(5):969-980, 2017

Prediction of cancer-associated hotspot mutations that affect GPCR oligomerization

Wataru Nemoto[1,2] Vachiranee Limviphuvadh[3] Sebastian Maurer-Stroh[3]
w.nemoto@mail.dendai.ac.jp vachiraneel@bii.a-star.edu.sg sebastianms@bii.a-star.edu.sg

Yoshihiro Yamanishi[4] Motoaki Yamanoi[2] Hiroyuki Toh[5]
yamanishi@bioreg.kyushu-u.ac.jp 17rmb24@ms.dendai.ac.jp tohhir@kwansei.ac.jp

[1] Division of Life Science and Engineering, School of Science and Engineering, Tokyo Denki University, ishizaka, Hatoyama-machi, Saitama 350-0311, Japan
[2] Division of Life Science and Engineering, Graduate School of Science and Engineering, Graduate School of Tokyo Denki University, ishizaka, Hatoyama-machi, Saitama 350-0311, Japan
[3] Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, #07-01 Matrix, 138671, Singapore
[4] Medical Institute of Bioregulation, Kyushu University, Fukuoka, 812-8581, Japan
[5] Department of Biomedical Chemistry, School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda-shi, Hyogo, 669-1337, Japan.

Keywords: membrane proteins, protein–protein interaction, interaction partner prediction, support vector machine, drug design;

At present, approximately 30% of marketed pharmaceutical medicines target G Protein-Coupled Receptors (GPCRs). Numerous studies have reported that GPCRs function not only as monomers but also as homo- or hetero- dimers or higher-order molecular complexes (oligomers). Many GPCRs exert a wide variety of functions by forming specific combinations of GPCR subtypes. Several hetero GPCR hetero dimers are reportedly associated with diseases. Therefore, GPCR oligomerization is recognized as an important event in various biological phenomena, and many researchers are investigating this subject. Hence, we developed a method to predict interacting pairs for GPCR oligomerization by integrating the structure and sequence information (GGIP). The area under the Receiver Operating Characteristic curve was 0.938. GGIP is the only prediction method for interacting pairs among GPCRs. Our prediction server is available at http://protein.b.dendai.ac.jp/GGIP/.

A recent study reported that somatic mutations in GPCR-encoding genes are frequently found in various types of cancer. Hotspot mutations are defined as recurrent amino acid changes occurring in coding sequences among hotspot mutations. It has been revealed that many GPCRs have hotspot mutations in the same types of cancer tissues although most of the hotspot mutations have not been characterized yet. In addition, their effects on the cancer pathways remain unknown. Some of the hotspot mutations may be related to cancers through modifying GPCR oligomerization since they are considered to be present on the surface of transmembrane helices. In this work, we examined the predicted interacting pairs including the GPCRs with hotspot mutations. We will discuss the characteristics of these mutations and introduce several examples.

[1] Nemoto, W. *et al*., *Proteins*. 12**,** 11, 2012

Structure and Dynamics of RNA Aptamer to Human Immunoglobulin G

Hisae Yoshida[1,2] cehs17001@g.nihon-u.ac.jp
Masahiro Sekiguchi[1] cema16020@g.nihon-u.ac.jp
Keisuke Masukawa[1] ceke14138@g.nihon-u.ac.jp
Sho Yamazaki[1] cesi14155@g.nihon-u.ac.jp
Emire Inomata[3] e.inomata@ribomic.com
Kazumasa Akita[3] k.akita@ribomic.com
Takeshi Ishikawa[4] t-ishi@nagasaki-u.ac.jp
Taiichi Sakamoto[5] taiichi.sakamoto@p.chibakoudai.jp
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] JSPS Research Fellow, Kojimachi Business Center Building, 5-3-1 Kojimachi, Chiyoda-ku, Tokyo 102-0083, Japan
[3] Ribomic Inc., 3-16-13 Shirokanedai, Minato-ku, Tokyo 108-0071, Japan
[4] Graduate School of Biomedical Sciences, Nagasaki University, 1-12-4 Sakamoto, Nagasaki, Nagasaki 852-8523, Japan
[5] Faculty of Advanced Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan

Keywords: RNA, Aptamer, Tertiary structure, Dynamics, Interaction energy, MD, FMO

RNA aptamers are short single-stranded nucleic acids with high affinity and specificity for their target molecules, which can be nucleic acids, proteins, or small organic compounds. RNA aptamers may be of value as therapeutic agents for the treatment of a wide variety of diseases. In order to be used in biological applications, the 2' position of the sugar moiety of one or more nucleotide units has to be suitably modified to make it resistant to nuclease digestion. However, it is not well understood how dose chemical modification for nucleotide units affect the uniquely tertiary structure of RNA aptamer.

The aim of this study is to understand the structure and dynamics of RNA aptamer. Our target aptamer is an optimized 23-nucleotide aptamer, which was shown to bind with high affinity to the Fc domain of human IgG (hFc1) [1]. The crystal structure of the hFc1 complexed with the RNA aptamer has been determined [2].

In order to elucidate the conformational behaviors and dynamical features of IgG-RNA aptamer, we performed molecular dynamics (MD) calculations for some modified IgG-RNA aptamers. We evaluated the interaction energies of base-pair and base-stacking using the *ab initio* Fragment Molecular Orbital (FMO) calculation of IgG-RNA aptamers.

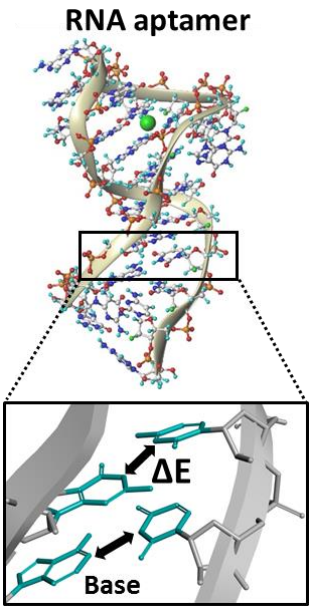


Figure. Interaction energies of the base-pair in the aptamer

Reference
[1] S. Miyakawa, et al., *RNA*, 14:1154-1163, 2008.
[2] Y. Nomura, et al., *Nucleic Acids Research*, 38:7822-7829, 2010.

P2-01

WhichP450: Predicting Which Cytochrome P450 Isoforms are involved in the Metabolism of a Xenobiotic

Peter Hunt[1] peter@optibrium.com
Jonathan Tyzack[1] jon@optibrium.com
Sumie Tajima[2] tajima@hulinks.co.jp
Marina Takahashi[2] m_takahashi@hulinks.co.jp
Matthew Segall[1] matt@optibrium.com

[1] Optibrium Ltd., 7221 Cambridge Research Park, Beach Drive, Cambridge, CB25 9TL, UK
[2] Hulinks Inc., 5-14 Nihonbashi Hakozaki-cho, Chuo-ku, Tokyo 103-0015, Japan

Keywords: Cytochrome P450 metabolism, isoform specificity, QSAR models, metabolite profile

Predicting which cytochrome P450 isoforms are involved in the metabolism of a molecule is important in assessing its metabolic fate in vivo. The various P450 isoforms have active sites of different shapes, sizes and characters [1], favouring different binding pharmacophores which can lead to metabolism at different sites within the molecule. Therefore, predicting the P450 isoforms likely to be involved in metabolism is a useful precursor to predicting the metabolites that are likely to be formed and considering which isoform-specific predictive models should be applied. 'Which P450' models also have application in assessing the risk of drug-drug interactions and the impact of genetic polymorphisms. A molecule reliant on a single isoform for metabolic clearance is at an increased risk of exhibiting drug-drug interactions or a greater impact of genetic polymorphisms on its pharmacokinetics.

Here we present a QSAR model to predict which P450 isoforms, out of a list of the 7 leading CYP450 isoforms, are likely to be involved in the metabolism of novel molecules. This model was generated using a multi-class categorical Random Forest method using the Auto-Modeller module in StarDrop [2] and has a 75% success rate with a Top-1 criterion and a 90% success rate for a Top-2 criterion and showed significantly improved performance over random estimates.

We furthermore illustrate how 'which P450' models can be combined with isoform-specific models for prediction of site of metabolism to predict the metabolite profile due to P450 metabolism of a xenobiotic.

[1] M.B.G. Cruciani et al., Molecular Interaction Fields, *Wiley-VCH*, 12:273-290, 2006.
[2] StarDrop. www.optibrium.com/stardrop

P2-02

Escape from ‘availability bias’ in compound design

Andras Stracz[1] **Akos Tarcsay**[1]

Gabor Imre[1] **Ivan Solt**[1] **Daisuke Miyazaki**[2]
miyazaki@patcore.com

[1] ChemAxon Ltd. 1031 Budapest. Zahony u. 7, Hungary
[2] Patcore,Inc. Sapia Tower 26F, 1-7-12 Marunouchi, Chiyoda-ku, Tokyo 100-0005, Japan

Keywords: Compound Design Support, Fast Structure Search

Small molecule design is an information demanding activity, since all relevant knowledge is to be accessible within a single space and requires synchronized application of computational models to assist decision making on synthesis candidates. Our study aims to evaluate a software platform coping with this complexity (Marvin Live). The tool provides central management of innovative ideas and a framework that helps triage these. Decision support is based on predicted properties that span phys-chem descriptors, combined metrics like CNS MPO score, 3D overlay and modelling results conducted with KNIME on the one hand, and cross-checking with available knowledge collected from a variety of sources to do rapid freedom to operate analysis and SAR by catalog: by ultra-fast searching of patent and compound catalog databases, respectively on the other. This hypothetic study shows the potential evolution of a compound idea from the reference compound to a synthesis candidate through an example discovery project of 5-hydroxytryptamine receptor 6 (HTR6) ligands.

[1] Wager, Travis T, et al. ACS Chem. Neurosci., 2010, 1 (6), pp 435-449
[2] Papadatos, George, et al. Nucleic Acids Res. 2016 Jan 4; 44(Database issue): 01220-01228. Database downloaded 01 May 2017
[3] https://nodejs.org/ Accessed 04 June 2017
[4] https://www.chemaxon.com/products/madfast/ Accessed 04 June 2017
[5] Shivanyuk, A. N., et al. Chemistry Today 2007,25,58-59. Database downloaded 01 May 2017
[6] Tolmachev, Andrey, et al. ACS Combinatorial Science 2016, 18.10,616-624.
[7]https://docs.chemaxon.com/display/docs/Chemical+Hashed+Fingerprint Accessed 04 June 2017
[8]https://www.chemaxon.com/products/calculator-plugins/molecular-modelling/ Accessed 04 June 2017

P2-03

Virtual Screening of New Inhibitors from Indonesian Herbal Compounds against Mycobacterium tuberculosis CYP121

Vivitri Dewi Prasasty[1]
vivitri.dewi@atmajaya.ac.id

Sandra Cindana[2]
sandra.cendana@surya.ac.id

Fransiskus Xaverius Ivan[2]
fx.ivan@surya.ac.id

[1] Faculty of Biotechnology, Atma Jaya Catholic University of Indonesia, Jakarta, 12930, Indonesia
[2] Faculty of Life Sciences, Surya University, Tangerang, Banten, 15810, Indonesia

Keywords: *M. tuberculosis* CYP121, molecular docking, molecular dynamics, kaempferol

Tuberculosis (TB) is a serious threat with the emergence of resistant bacterial variants, namely multidrug-resistant tuberculosis (MDR-TB) and extremely drug-resistant tuberculosis (XDR-TB). This study aims to find new inhibitors for *M. tuberculosis* CYP121 from Indonesian medicinal plants, especially *Rhoeo spathacea* and *Pluchea indica*. A virtual structure-based screening was performed on 16 molecules through molecular docking using AutoDock Vina and subsequent molecular dynamics simulations were performed using YASARA. The two best compounds, i.e. KAE (kaempferol) and KAE3 (kaempferol derivative) were selected based on their lowest binding energies (-9.1 kcal/mol and -10.5 kcal/mol, respectively) and their adherence to the Lipinski rule. Results from 50 ns molecular dynamics simulation suggest that KAE and KAE3 have an inhibitory mechanism againsts *M. tuberculosis* CYP121 that is different from GGJ (control inhibitor). Based on the RMSF scores comparison, GGJ alters the overall dynamics of the receptor characterized by the changes in the flexibility of amino acid residues far away from the receptor binding side. In contrast, the binding of KAE and KAE3 changes the flexibility of amino acids around the binding site of the receptor.

[1] Collins, D., Hafidz, F. & Suraratdecha, C., *The Economic Burden of Tuberculosis in Indonesia*, Cambridge: TB CARE I - Management Sciences for Health, 2013.
[2] Radji, M., Kurniati, M. and Kiranasari, A., Comparative antimycobacterial activity of some Indonesian medicinal plants against multi-drug resistant *Mycobacterium tuberculosis*. Journal of Applied Pharmaceutical Science, 05(01): 019-022, 2015.
[3] McLean, K. J. and Munro, A. W., Structural biology and biochemistry of cytochrome P450 systems in *Mycobacterium tuberculosis*. Drug Metabolism Review, 40: 427-446, 2008.
[4] Sliwoski, G., Kothiwale, S., Meiler, J. and Lowe, E. W., Computational methods in drug discovery. Pharmacological Reviews, 66(1): 334-395, 2014.

P2-04

A new rotation-translation invariant molecular descriptor for Ligand-Based Virtual Screening and beyond

Francois Berenger[1] **Yoshihiro Yamanishi**[1,2]
berenger@bioreg.kyushu-u.ac.jp yamanishi@bioreg.kyushu-u.ac.jp

[1] System Cohort Division, Medical Institute of Bioregulation, Kyushu University
3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan
[2] PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan

Keywords: Ligand-Based Virtual Screening, molecular similarity, chemical fingerprints, autocorrelation, 3D.

Various methods exist to compare ligands: fingerprints like MACCS, ECFP4 or PubChem and specialized software like ROCS[1], EON, ElectroShape[2], ShaEP[3] and Blaze[4]. There are also various methods to compare protein binding sites like sup-CK[5], SuMo[6], Shaper[7], SiteAlign[7] and FuzCav[7].

While fingerprints are only working in 2D, binding site comparison methods are complex and incorporate several parameters. Interestingly, both kind of methods perform the same task of comparing two sets of atoms.

In this study, we propose a novel method, scarce in parameters and working in 3D, that can do both ligand and binding site similarity search, modulo re-parametrization of the method towards the intended use case (to maximize performance). The proposed method will be able to directly find ligands given a binding site; without docking or prior knowledge of ligands for this binding site. Similarly, the opposite task of finding a binding site given a ligand is theoretically possible. We will show the usefulness of our proposed method by comparing its performance with popular 2D chemical fingerprints. Our method considers ligands in 3D and is rotation-translation invariant. Ligands do not need to be optimally superposed prior to being compared. Another innovative point of our method is that it can work in any or a combination of "chemical spaces" like shape, electrostatic, hydrophobic and hydrogen bonding.

[1] Rush, T. S., Grant, J. A., Mosyak, L., & Nicholls, A. (2005). A shape-based 3-D scaffold hopping method and its application to a bacterial protein− protein interaction. *JMedChem*, *48*(5), 1489-1495.
[2] Armstrong, M. S., Morris, G. M., Finn, P. W., Sharma, R., Moretti, L., Cooper, R. I., & Richards, W. G. (2010). ElectroShape: fast molecular similarity calculations incorporating shape, chirality and electrostatics. *JCAMD*, *24*(9), 789-801.
[3] Vainio, M. J., Puranen, J. S., & Johnson, M. S. (2009). ShaEP: molecular overlay based on shape and electrostatic potential.
[4] Cheeseright, T., Mackey, M., Rose, S., & Vinter, A. (2006). Molecular field extrema as descriptors of biological activity: definition and validation. *JCIM*, *46*(2), 665-676.
[5] Hoffmann, B., Zaslavskiy, M., Vert, J. P., & Stoven, V. (2010). A new protein binding pocket similarity measure based on comparison of clouds of atoms in 3D: application to ligand prediction. *BMC bioinformatics*, *11*(1), 99.
[6] Jambon, M., Andrieu, O., Combet, C., Deléage, G., Delfaud, F., & Geourjon, C. (2005). The SuMo server: 3D search for protein functional sites. *Bioinformatics*, *21*(20), 3929-3930.
[7] Kellenberger, E., Schalon, C., & Rognan, D. (2008). How to measure the similarity between protein ligand-binding sites? *CCADD*, *4*(3), 209.

P2-05

Improving 3D-QSAR in AutoGPA using Gaussian-distributed physicochemical descriptors

Yoshirou Kimura[1], **Akitoshi Okada**[1], **Ryoichi Kataoka**[1], **Atsushi Kanou**[1]
kimura.yoshirou@molsis.co.jp

[1] Life Science Department, MOLSIS Inc., 1-28-38 Shinkawa, Chuo-ku, Tokyo 104-0033, Japan

Keywords: 3D-QSAR, AutoGPA, CoMFA, CoMSIA,

AutoGPA [1] is fully automated 3D-QSAR application developed on MOE [2]. It can build 3D-QSAR models based on CoMFA [3] from the collection of ligand structural formulae and activity values without a need of active conformers. AutoGPA can predict active conformers by aligning ligands based on common pharmacophore features by making use of Pharmacophore Elucidation function implemented in MOE.

In this study, we employed CoMSIA [4] method in AutoGPA with more and unique 3D descriptors compared to the original CoMSIA. Weights and center positions of the new 3D descriptors are assigned by pharmacophore features and atomic contribution of molecular properties. The available pharmacophore features include cation, anion, aromatic ring, hydrogen-bond (H-bond) donor/acceptor. Furthermore, in case of H-bond donor/acceptor, H-bond donor/acceptor strength are derived from modified Hückel theory [5] and used as a weight of their Gaussian distribution. The available molecular properties are volume, charge, logP, Molecular Refractivity.

We applied the molecular alignment of AutoGPA and the new 3D-QSAR method to phosphoinositide-dependent kinase-1 (PDK1) inhibitor dataset. The obtained 3D-QSAR model was compared to one of the crystal structure of PDK1 and inhibitor complex (PDB ID: 2PE2). We found the positions and the properties of 3D descriptors which contribute to activity are similar to those of the ligand binding site (Fig. 1). The prediction accuracy of the 3D-QSAR model ($q^2 = 0.834$, $r_{pred}^2 = 0.846$) was better than that of the CoMFA model reported by previous AutoGPA ($q^2 = 0.760$, $r_{pred}^2 = 0.811$) [1]. We believe our new method enhances AutoGPA in its applicability and flexibility by the use of additional 3D descriptors.

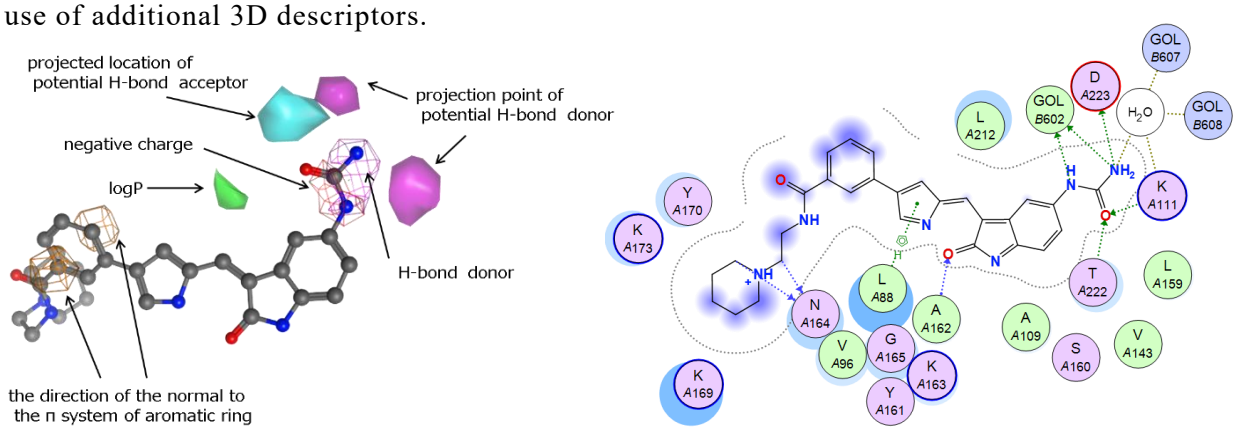


Figure 1. The isosurfaces of the 3D-QSAR model (left) and the 2D depiction around ligand binding site of PDK1.

[1] Asakawa, N., Kobayashi, S., Goto, J., Hirayama, N.; *Int. J. Med. Chem.* **2012**, *2012*, 498931.
[2] *Molecular Operating Environment (MOE)*, 2016.08; Chemical Computing Group Inc., 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2016.
[3] Cramer, R.D., Patterson, D.E., Bunce, J.D.; *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.
[4] Klebe, G., Abraham, U., Mietzner, T.; *J. Med. Chem.* **1994**, *37*, 4130-4146.
[5] Gerber, P.R.; *J. Comput.-Aid. Mol. Des.* **1998**, *12*, 37–51.

P2-06

In silico search for proanthocyanidins binding proteins

Takuma Todoroki[1]
14a3035a@shinshu-u.ac.jp

Reina Takahashi[1]
14a3031h@shinhsu-u.ac.jp

Hiroshi Fujii[1,2]
hfujii@shinshu-u.ac.jp

Hidefumi Makabe[1]
makabeh@shinshu-u.ac.jp

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

[1] Faculty of agriculture, Shinshu University, 8304 Minami-minowa, Kami-ina, Nagano, 399-4598, Japan
[2] Department of Interdisciplinary Genome Sciences and Cell Metabolism, Institute for Biomedical Sciences, Interdisciplinary Cluster for Cutting Edge Research, Shinshu University, Minami-minowa, Kami-ina, Nagano, 399-4598, Japan

Keywords: Natural products, Ligand docking, Binding-pocket search

There are a lot of natural products good for health via a specific biological function while, in most cases, the receptors for the molecules derived from the natural products are still unknown. Proanthocyanidins belong to the functional natural products of polyphenols involved in grape, apple and so on, which exhibit various functions such as suppression of cell proliferation. The receptors for proanthocyanidins have not been identified widely.

The proanthocyanidin is the polymer conjugating the units of catechin (cat), epicatechin (epi) and their derivatives of gallocatechin (gc) and epigallocatechin (egc). Therefore, the chemical structure of proanthocyanidin has a lot of variety depending on combination of the species of units and the number of polymerization. For instance, the dimer of egc (egc-egc) is called prodelphinidin B2 (PDB2). Interestingly, the PDB2 exhibits a biological activity although the dimer of gc (gc-gc) does not. Thus, the chemical variety is crucial for the biological activity. It should be related to a receptor for proanthocyanidins. However, the receptor proteins that proanthocyanidins bind are not found comprehensively. Then, we have explored the proanthocyanidin-binding proteins from Protein Data Bank.

For the purpose of searching the proanthocyanidin-binding proteins from the database, we took the two steps: local-similarity search and ligand-docking simulation. First, in the local-similarity search, we chose the known cat-binding and epi-binding protein structures to construct binding-pocket templates for cat and epi. The binding-pocket template is composed of the coordinates of protein atoms in the vicinity of cat or epi. We compared the binding-pocket templates with the all structures deposited in Protein Data Bank. We collected the proteins that have the similar sites with the binding-pocket templates. In the next step, we conducted the ligand-docking simulation for the collected proteins with proanthocyanidins. The proanthocyanidin-binding candidates were ranked in order of the docking scores. Then, we could search proteins related to biological functions including cell proliferation. These proteins might be future drug targets as lead compounds of proanthocyanidins.

P2-07

Construction of FMO IFIE-database

Chiduru Watanabe[1]
chiduru.watanabe@riken.jp

Daisuke Takaya[1]
daisuke.takaya@riken.jp

Shunpei Nagase[1]
shunpei.nagase@riken.jp

Kikuko Kamisaka[1]
kikuko.kamisaka@riken.jp

Yoshio Okiyama[1,2]
yoshio.okiyama@riken.jp

Kaori Fukuzawa[3]
k-fukuzawa@hoshi.ac.jp

Teruki Honma[1]
honma.teruki@riken.jp

1 RIKEN Center for Life Science Technologies, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
2 Division of Medicinal Safety Science, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan
3 School of Pharmacy and Pharmaceutical Sciences, Hoshi University 2-4-41 Ebara, Shinagawa-ku, Tokyo 142-8501, Japan

Keywords: interaction energy database, FMO, ligand binding

The fragment molecular orbital (FMO) method proposed by Kitaura *et al.* enables us to efficiently perform *ab initio* quantum mechanical calculations for large biomolecules by many-body expansion technique with fragmentation. Inter-fragment interaction energy (IFIE) analysis based on FMO is able to reveal important interactions in molecular recognition such as protein-ligand binding in units of amino acid residues and a ligand. Moreover, the information of important interactions estimating from decomposed electronic energy terms (e.g. electrostatic and dispersion interactions) is quite useful for drug design. On the other hand, it is not easy to perform FMO calculations for not experienced researchers in the matter of structure preparation and FMO setting. To publicly disclose the IFIE data, we have started to comprehensively collect IFIE data and construct IFIE-database (Figure 1). However, there are limits to prepare huge amount of various protein structure data (e.g. PDB having ca. 40,000 entries of protein-ligand complex) by manual labor. Thus, we developed “semi-automated FMO calculation protocol” to efficiently carry out structure preparation and subsequent FMO calculations for multiple X-ray crystal structures. To validate the FMO protocol, we manually prepared protein structure and carry out FMO calculations using the same crystal structures. As a result, IFIE values of the FMO protocol data showed good agreement with those of manual labored data. Therefore, we confirmed that this protocol can be used for structure preparation and FMO calculations of massive protein structures. There are at present the IFIE data of 300 complexes (e.g. ERα, ERβ, p38-MAP kinase, Aurora, CHK1) obtained from the FMO protocol. They were registered in newly developed IFIE-database with GUI. This research was done in activities of the FMO drug design (FMODD) consortium [http://eniac.scitec.kobe-u.ac.jp/fmodd/index.html]. The results of FMO calculations were obtained using the K computer (project ID: hp170183).

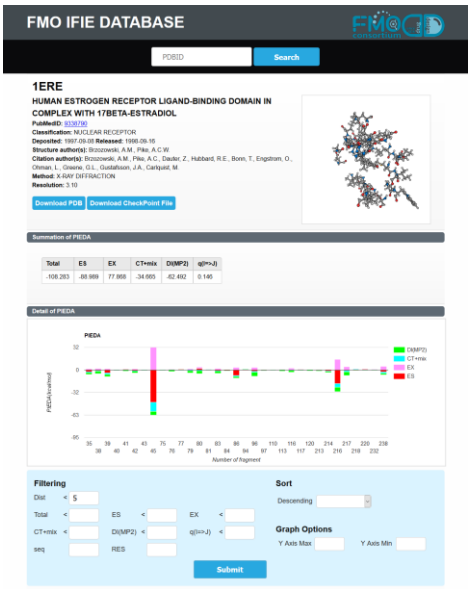


Figure 1 IFIE-database

The Role of Water Molecules in Protein-Ligand Binding: Fragment Molecular Orbital Calculations on the Complexes of Renin with its Inhibitors

Yoichiro Yagi[1]
yagi@ee.ous.ac.jp

Takatomo Kimura[2]
kimura@konankako.co.jp

Chiduru Watanabe[3]
chiduru.watanabe@riken.jp

Yoshio Okiyama[4]
okiyama@nihs.go.jp

Shigenori Tanaka[5]
tanaka2@kobe-u.ac.jp

Teruki Honma[3]
honma.teruki@riken.jp

Kaori Fukuzawa[6]
k-fukuzawa@hoshi.ac.jp

[1] Department of Electrical and Electronic Engineering, Okayama University of Science, 1-1 Ridai-cho, Kita-ku, Okayama 700-0005, Japan
[2] Konan Chemical Industry Co., Ltd., 5-12 Nakagawa-cho, Takatsuki, Osaka 569-0066, Japan
[3] RIKEN Center for Life Science Technologies, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
[4] National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan
[5] Department of Computational Science, Graduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe, Hyogo 657-8501, Japan
[6] School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa-Ku, Tokyo 142-8501, Japan

Keywords: Fragment Molecular Orbital (FMO) Calculation, Renin, Inhibitor, Water molecules, Interaction energy, Binding energy, Pair interaction energy decomposition analysis (PIEDA)

We have carried out the fragment molecular orbital (FMO) calculations on the 22 different renin-inhibitor complexes at FMO2-MP2/6-31G* level to clarify a relationship between the calculated binding energy for the renin-inhibitor complexes and the activity value on 50% inhibitory concentration (IC_{50}) of the inhibitors. The PDB structures of complexes with TIP3P model waters around 8Å of complex were optimized by an energy minimization using AMBER. After energy minimization, TIP3P model waters were removed and FMO calculations using MIZUHO/BioStation program were performed on the K computer [1]. FMO computations indicated that the binding energy can be correlated to IC_{50}. In the FMO calculation that contains crystal waters within 5Å of ligand, the binding energies showed a good correlation with IC_{50} for each inhibitor [2]. The aim of this study is to elucidate a role of water molecules in binding of renin and inhibitor. By analyzing of the interactions between ligand and water molecules and between water molecules and amino acid residues in protein, we found that ligand interacts with protein through the water molecules. This result suggests that the water molecules may support the binding of ligand in active site of protein. In addition, we performed pair interaction energy decomposition analysis (PIEDA) calculation to examine the energy components.

This research was done in activities of the FMO drug design consortium (FMODD). The results were obtained using the K computer (project ID: hp170183).

[1] Tanaka, S., Mochizuki, Y., Komeji, Y., Okiyama, Y., and Fukuzawa, K., Electron-correlated fragment-molecular-orbital calculations for biomolecular and nano systems, *Physical Chemistry Chemical Physics*, 16:10310-10344, 2014.
[2] Yagi, Y., Kimura, T., Watanabe, C., Okiyama, Y., Tanaka, S., Honma, T., and Fukuzawa, K., Fragment Molecular Orbital Calculations on the Interaction and Binding Energies of Human Protease Renin and its Inhibitors, *Chem-Bio Informatics Society (CBI) Annual Meeting 2016*, P1-08, 2016.

P2-09

Using multiple molecular fingerprints for improvement of drug activity predictions

Yusuke Matsuyama[1,3] matsuyama@cb.cs.titech.ac.jp
Takashi Ishida[2,3] ishida@c.titech.ac.jp

1 Department of Computer Science, Tokyo institute of technology, Meguro, Tokyo, Japan
2 School of Computing Department of Computer Science, Tokyo institute of technology, Meguro, Tokyo, Japan
3 Education Academy of Computational Life Sciences, Tokyo Institute of Technology, Meguro, Tokyo, Japan

Keywords: Virtual screening; Machine learning; Molecular fingerprint; Stacking; Ensemble learning

Drug activity prediction based on machine learning method is useful for ligand-based virtual screening. And it is drawing attention as a key to keep down drug discovery costs which are said to be 300 billion[1] and still increasing now.

When using machine learning technique, the information of molecule have to be converted into a vector with fixed length.

For many years, huge number of molecular fingerprints are proposed, however, it is known that the best fingerprint is different for the target[2], and it is difficult to select the most suitable fingerprint.

To solve this problem, we propose a new technique to use multiple fingerprint by emsemble learning for drug activity prediction.

As a result of performance evaluation, the proposed method greatly increased the predicted performance compared to the prediction model using single molecular fingerprint.

In addition, we compared the feature and prediction performance of major molecular fingerprints, and analyzed the reason why our ensemble method works well.

[1] Mullard, A.: New drugs cost US$2.6 billion to develop. Nature reviews. Drug discovery 13(12), 877 (2014). doi:10.1038/nrd4507

[2] Riniker, S., Landrum, G.A.: Open-source platform to benchmark fingerprints for ligand-based virtual screening. Journal of Cheminformatics 5(5) (2013). doi:10.1186/1758-2946-5-26

K^4: An *in silico* Drug Design Suite combined with HPCs

Biao Ma[1,2] biao-ma@fbri.org
Yuta Isaka[1,2] isaka@fbri.org
Mitsugu Araki[3,4] araki.mitsugu.6w@kyoto-u.ac.jp
Hiroaki Iwata[3] iwata.hiroaki.3r@kyoto-u.ac.jp
Yasushi Okuno[2,3,4] okuno.yasushi.4c@kyoto-u.ac.jp

[1] Research and Development Group for In Silico Drug Discovery, Pro-Cluster Kobe, Foundation for Biomedical Research and Innovation (FBRI), 1-6-5, Minatojima-Minamimachi Chuo-ku, Kobe 650-0047, Japan
[2] Simulation-Driven Drug Discovery Group, IBRI laboratory, Foundation for Biomedical Research and Innovation (FBRI), 1-6-5, Minatojima-Minamimachi Chuo-ku, Kobe 650-0047, Japan
[3] Graduate School of Medicine, Kyoto University, 53 Kawahara-cho, Shogoin Sakyo-ku, Kyoto 606-8507, Japan
[4] RIKEN Advanced Institute for Computational Science, 7-1-26 Minatojima-Minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

Keywords: SBDD, Molecular dynamics simulation, Docking simulation, Virtual screening, Binding free energy calculation, MP-CAFEE, HPCs, K-computer, Graphical user interface, GUI

In drug discovery process, computational methods to efficiently explore and optimize drug candidates are strongly required. Since virtual screening from huge amount of chemical compounds ($> 10^{60}$) and accurate prediction of the protein-drug binding affinity need a large computational cost, we attempt to overcome this problem by utilizing high-performance computers (HPCs) such as K-computer, Osaka Univ VCC, and so on. However, in silico drug discovery on HPCs requires complicated operation. To solve this problem, we have developed an in silico drug design suite, K^4(K-Force), which has a friendly graphical user interface (GUI). The medicinal chemists can easily utilize computational power of K-computer and other HPCs.

In K^4, we constructed a single platform that integrates the following main functions: 1) protein structure sampling using temperature replica-exchange molecular dynamics (T-REMD), 2) docking simulation (DS) or virtual screening (VS) based on rDock[1], 3) Ab initio calculation of compound partial charges based on GAMESS, 4) the protein-compound free energy calculation based on molecular dynamics (MD) simulation using MP-CAFEE[2,3] or MM-PB/GBSA method.

We will provide more methods for rational design of drug molecules in next version, such as protein-ligand inter-fragment interaction energy calculation(IFIE) by use of fragment molecular orbital (FMO) method and ensemble docking, which enables to dock a chemical library against multiple receptor conformations. If you are interested into our applications suite, please visit KOBE-CITY and FBRI booth.

[1] Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., Garmendia-Doval, A. B., Juhos, S., Schmidtke, P., Barril, X., Hubbard, R. E., and Morley, S. D. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. PLoS Comput. Biol. (Prlic, A., Ed.) 10, e1003571, 2014.
[2] Fujitani, H., Tanida, Y., and Matsuura, A. Massively parallel computation of absolute binding free energy with well-equilibrated states. Phys. Rev. E 79, 21914, 2009.
[3] Araki, M., Kamiya, N., Sato, M., Nakatsui, M., Hirokawa, T., and Okuno, Y. The Effect of Conformational Flexibility on Binding Free Energy Estimation between Kinases and Their Inhibitors. J. Chem. Inf. Model. 56, 2445–2456, 2016.

P2-11

ReCGen: Development of a novel approach to generate virtual compounds

Chisato Kanai[1] kanai@k-ct.jp

Edmund Taylo[1] taylo@k-ct.jp

[1] Kyoto Constella Technologies Co., Ltd., fourth floor Kyozome Kaikan, 481 Tourouyama-cho, Nakagyo-ku, Kyoto City, 604-8225, Japan

Keywords: *de novo* design, ECFP, RECAP, synthetic accessibility

It is widely accepted that one of the important elemental technologies of *de novo* design [1] is the creation of novel structures. To date, various methods have been proposed which are roughly classified into the following three types: atom-based method, fragment-based method, and chemical reaction (synthetic route)-based method. The most common type is fragment-based structure generation method which is represented by those using RECAP [2], BRICS [3] and MMPA (Matched molecular pair analysis) [4]. RECAP, in particular, is a type of retrosynthetic analysis method focusing on chemical reactions frequently used in drug discovery. The structures generated by this method are relatively easy to synthesize, so the method is considered good for *de novo* design.

We have recently developed a new fragment-based method that utilizes the concept of ECFP [5], a type of fingerprint, in the generation of new fragments that are subsequently used as building blocks in the assembly of new structures. In our method, the size of ECFP fragments can be set arbitrarily by specifying the diameter of the graph, so a wide variety of fragments can be obtained as compared with the RECAP method. Therefore, compared to RECAP, which is subject to rough structural transformation, relatively fine structure transformation is possible when using our method. Furthermore, synthetic accessibility (SA) [6] of structures generated by our method is as favorable as that generated using the RECAP method. Because of those features, we refer to it as “Refined Compound Generator (ReCGen)”.

In this poster we present a comparison between the principle of our method and the features of RECAP method.

[1] Schneider G. and Fechner U., Nature Reviews Drug Discovery 4:649–663, 2005.
[2] Lewell, X.Q., Judd, D.B., Watson, S.P. and Hann, M.M., J. Chem. Inf. Comput. Sci. 38:511-522, 1998.
[3] Degen J, Wegscheid-Gerlach C, Zaliani A and Rarey M., ChemMedChem, 3:1503-1507, 2008.
[4] Kenny P. W., Sadowski J., “Structure Modification in Chemical Databases”. In Chemoinformatics in Drug Discovery. Oprea, T. I., Ed. Wiley-VCH: Weinheim, Germany. pp. 271−285, 2004.
[5] David R. and Mathew H., J. Chem. Inf. Model. 50:742-754, 2010.
[6] Fukunishi Y., Kurosawa T., Mikami Y. and Nakamura H., J. Chem. Inf. Model. 54:3259-3267, 2014.

Quick and precise homology modeling method of GPCRs

Mika Nabeno[1] **Takatsugu Hirokawa**[2]
nabeno.mika@mv.mt-pharma.co.jp t-hirokawa@aist.go.jp

[1] Discovery Technology Laboratories, Sohyaku. Innovative Research Division, Mitsubishi Tanabe Pharma Corporation, 1000 Kamoshida-cho, Aoba-ku, Yokohama, Kanagawa, 227-0033, Japan
[2] Molecular Profiling Research Center for Drug Discovery, National Institute of Advanced Industrial Science and Technology, 2-4-7 Aomi, Koto-ku, Tokyo, 135-0064, Japan
Division of Biomedical Science, Faculty of Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8575, Japan

Keywords: GPCR, Homology modeling, Extracellular loop II, Ligand docking, Molecular dynamics simulation

G protein-coupled receptors (GPCR) are major drug discovery targets, and the reports on the X-ray crystal structures of GPCRs, which are rapidly increasing in recent years, are very useful for GPCR-targeted drug design. However, because X-ray crystallography of GPCRs is still challenging, we cannot acquire the structures complexed with compounds at the time when they can be utilized in drug design effectively. Therefore, the needs of homology modeling are high at the drug discovery field. A number of modeling methods of GPCRs have been studied to date[1]. Particularly in recent years, there is a tendency to perform a large amount of calculation to build a highly accurate model[2]. However, in the drug discovery research, we often have multiple compounds to be modeled, and the predicted models need to be revised according to SAR of compounds. In addition to that, when a new X-ray crystal structure more suitable for a template is revealed, the homology models have to be built from scratch. For the reasons shown above, it is desirable that the modeling method be not only highly accurate but also simple.

In this study, we propose the quick and precise homology modeling method in the case of the compound binding to the orthosteric site of GPCR. The site is covered with the extracellular loop II (ECL2), which affects binding of the compound seriously. However, because of its diversity among GPCRs, its modeling is generally difficult. Here, a multi-template method was adopted which selected different templates for transmembrane(TM) region and ECL2. We found out that the accuracy of modeling is improved by selecting one template for the TM region with high sequence homology and another template for the ECL2 with almost the same number of amino acid residues. In drug design, although the prediction of the docking pose of the compound is quite vital, it is the most difficult step in GPCR modeling. Then, we built an easy method to narrow down the possible docking poses after searching them in a wide range. It was also confirmed that the accuracy of the models was remarkably improved by performing molecular dynamics simulation for 50 ns on finishing of the obtained models. In our validation, the RMSDs of the bound ligands predicted by the above method were less than 2 Å from those of the X-ray crystal structures. Although further verification is necessary, this method can be promising for drug design.

[1] Koehler Leman, J., Ulmschneider, M. B., Miller, and W., Gray, J.J., Computational modeling of membrane proteins, *Proteins*, 83(1):1-24, 2015.
[2] Rodríguez, D., Ranganathan, A., Miller, W., and Carlsson, J., Strategies for Improved Modeling of GPCR-Drug Complexes: Blind Predictions of Serotonin Receptors Bound to Ergotamine, *Journal of Chemical Information and Modeling*, 54:2004-2021, 2014.

P2-13

KampoDB: An integrated platform for mode-of-action analysis and repositioning of natural medicines

Ryusuke Sawada[1]
rsawada@bioreg.kyushu-u.ac.jp

Michio Iwata[1]
m-iwata@bioreg.kyushu-u.ac.jp

Masahiro Umezaki[2]
masume@inm.u-toyama.ac.jp

Yoshihiko Usui[2]

Takaki Kubono[3]
m1661217@ems.u-toyama.ac.jp

Makoto Kadowaki[3]
makotok@inm.u-toyama.ac.jp

Yoshihiro Yamanishi[1,4]
yamanishi@bioreg.kyushu-u.ac.jp

[1] Division of System Cohort, Multi-scale Research Center for Medical Science, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka, Fukuoka 812-8582, Japan.
[2] Division of Chemo-Bioinformatics, Institute of Natural Medicine, University of Toyama, Toyama 930-0194, Japan.
[3] Division of Gastrointestinal Pathophysiology, Institute of Natural Medicine, University of Toyama, Toyama 930-0194, Japan.
[4] PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan.

Keywords: Kampo medicines, functional analysis, target prediction, Kampo repositioning

Natural medicines (i.e., herbal medicines, traditional formulas) are useful for treatment of multifactorial and chronic diseases. Here, we present KampoDB (http://wakanmoview.inm.u-toyama.ac.jp/kampo/), a novel database of natural medicines, which provides various useful scientific resources on Japanese traditional formulas Kampo medicines, constituent herbal drugs, constituent compounds, and target proteins of these constituent compounds. Potential target proteins of these constituent compounds were predicted by docking simulations and machine learning methods based on large-scale omics data (e.g., genome, proteome, metabolome, interactome). The current version of KampoDB contains 42 Kampo medicines, 54 crude drugs, 1230 constituent compounds, 460 known target proteins, and 1369 potential target proteins, and has functional annotations for biological pathways and molecular functions. KampoDB is useful for mode-of-action analysis of natural medicines and prediction of new indications for a wide range of diseases.

Integrated multiomics analysis for dengue hermorrhogic fever

Takayuki Amemiya[1]
takayuki-amemiya@aist.go.jp

M. Michael Gromiha[1,2]
gromiha@iitm.ac.in

Katsuhisa Horimoto[1]
k_horimoto@aist.go.jp

Kazuhiko Fukui[1]
k-fukui@aist.go.jp

[1] Molecular Profiling Research Center for Drug Discovery (molprof), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo 135-0064, Japan
[2] Department of Biotechnology, Indian Institute of Technology Madras, Tamilnadu 600 036, India

Keywords: Multi-omics analysis, Drug repositioning, Dengue hermorrhagic fever

Due to recent advances in the development of various kinds of high throughput technology in life sciences, it is possible to collect and analyze multiple "omics" data. Omics analysis as integrative system biology has been studied for drug discovery, biomarker and assessment of their toxicity and efficacy. Here, a multi-omics approach is applied for dengue hermorrhagic fever (DHF) characterized by vasculopathy, which results in sudden plasma leakage that reduces the blood volume and may result in hypovolemic shock, known as dengue shock syndrome [1]. However, there are no approved drug, DHF is classified as a neglected tropical disease by WHO.

In our analysis, three layers of experimental data such as transcriptomic, proteomic and interactomic data are considered. In each layer, we obtained the signature genes and proteins to find the drug candidates and the significant biological pathway [2]-[4]. The effective drug candidates are identified by using a drug repositioning method [5] to inverse drug-disease relationship in patients and healthy control gene expression profile and using interactomic relationships between signature proteins and chemical compounds. We will present the integrated multi-omics analysis for DHF, which leads to identify and narrow down drug candidates based on expression profiles and molecular interaction.

[1] D. J. Gubler, *Dengue and dengue hemorrhagic fever.*, Clin. Microbiol. Rev., 11:480–96, 1998.
[2] J. Lamb, *The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease*,,Science, 313:1929–1935, 2006.
[3] A. Subramanian et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.*, Proc. Natl. Acad. Sci. U. S. A., 102:15545–15550, 2005.
[4] D. Szklarczyk et al., *STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data*, Nucleic Acids Res., 44:D380–D384, 2016.
[5] Lamb, J., et al., The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and diseases, *Science*, 313: 1929-1935, 2006.

P2-15

Development of Structure-Free Design and Analysis for Compound Libraries Based on Chemical and Bioactive Diversity

Yugo Shimizu[1]
shimizu-yg@pha.keio.ac.jp

Kazuyoshi Ikeda[1]
ikeda-kz@pha.keio.ac.jp

Masanori Osawa[1]
osawa-ms@pha.keio.ac.jp

[1] Keio University, Faculty of Pharmacy, Division of Physics for Life Functions, 1-5-30 Shibakoen, Minato-ku, Tokyo 105-8512, Japan

Keywords: Library design, Structure-blinded analysis, Bioactivity space

High through screening is a starting point for drug discovery in many pharmaceutical companies. To expand collaborative drug discovery between industry and academia, the Drug-discovery Innovation and Screening Consortium (DISC) of over 20 pharmaceutical companies in Japan has been established [1]. DISC generated a library that consists of compounds collected from the attending companies, but diversity of compound structures in the library is unknown due to the intellectual property, and therefore, it is necessary to analyze the compound library without compound structure. For this purpose, we started a project, "Structure-Free Compound-Library Diversity Analysis", supported by Japan Agency for Medical Research and Development to develop a computational system for diversity analysis of large-scale structure-blinded compound library.

In consideration of security, we adopted a secure cloud-based web system for collecting physicochemical descriptors and fingerprints of the compounds from companies, and Pipeline Pilot was used as a platform for manipulating and calculating large-scale compound data. An automated protocol for converting 2D/3D molecular descriptors of physicochemical property including principal moment of inertia and lead/drug-likeness score was developed for subsequent analysis of compound library diversity. The developed protocol can compare chemical spaces of small-molecule libraries from approved drugs, candidates, commercial compounds, and patent compounds from ChEMBL database [2]. We also developed a method for analyzing coverage of bioactive space. We extracted compounds in the ChEMBL database with bioactivity information for over 1,400 targets. Models for each target were created by several machine learning (ML) methods such as naïve Bayes, random forest, support vector machine, and deep neural networks using the descriptors and fingerprints. The prediction performance of each ML method was tested by cross-validation and the best one was used for predicting the bioactivity of compounds from large-scale libraries to evaluate the distribution of diversity in bioactive space.

[1] http://www.amed.go.jp/program/list/06/03/05.html
[2] Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., Krüger, F. A., Light, Y., Mak, L., McGlinchey, S., Nowotka, M., Papadatos, G., Santos, R., and Overington, J. P., The ChEMBL bioactivity database: an update, *Nucleic Acids Research*, 42(Database issue):D1083-1090, 2014.

This work is conducted as a part of "Construction of Drug Discovery Informatics System" supported by Japan Agency for Medical Research and Development (AMED).

Developing novel inhibitors of Heme-copper oxygen reductases

Yuya Nishida[1]
nishida@medbio.med.osaka-u.ac.jp

Yasunori Shintani[1]
yshintani@medbio.med.osaka-u.ac.jp

Hitomi Yuki[2]
hitomi.yuki@riken.jp

Teruki Honma[2, 3]
honma.teruki@riken.jp

Seiji Takashima[1]
takasima@cardiology.med.osaka-u.ac.jp

[1] Medical Biochemistry, Osaka University Graduate School of Medicine, 2-2 Yamadaoka, Suita-shi, Osaka, 565-0871, Japan
[2] RIKEN Center for Life Science Technologies, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
[3] Tokyo Institute of Technology, School of Computing, Department of Computer Science, 4259 Nagatsuta-cho, Midori-ku, Yokohama, 226-8503, Japan

Keywords: Heme-copper oxygen reductase, cytochrome c oxidase, allosteric inhibitor, ligand-based virtual screening

Heme-copper oxygen reductase (HCO) is a family of transmembrane proteins that work as a member of the respiratory electron transport chain. HCO receives electrons from cytochrome c or ubiquinol and then reduces oxygen to water, which couples with proton transfer across membrane. The generated transmembrane difference in proton concentration is a major part of proton-motive force, which is a driving force of ATP synthase. These reactions are crucial in most organisms, therefore the composing proteins are considered as drug targets. Here we describe inhibitors of two HCOs: mitochondrial cytochrome c oxidase (mtCcO) and bacterial ubiquinol oxidase (UqO). Recently, we identified novel allosteric inhibitors of mtCcO, and successfully determined crystal structures of mtCcO in complex with two inhibitors. The structures showed that both inhibitors bound to the same position of mtCcO, which is far from the ligand-binding site. This suggests that mtCcO activity is inhibited by a novel allosteric mechanism. In addition, as comparing mtCcO with bacterial UqO, the inhibitor binding site in mtCcO has some similarity to the homologous position in bacterial UqO, though mtCcO is much larger complex than bacterial UqO. These led us to hypothesize that the novel allosteric site has a potential to be a novel antibiotic target. To elucidate the inhibitory mechanism by approach of the structure-activity relationship and to develop a novel inhibitor targeting bacterial UqO, we performed ligand developments with the inhibitors we found. By using of the ligand-based virtual screening system LAILAPS in commercially available approximately 70,000,000 compounds, 270 compounds were selected and subsequently, we assessed these inhibitory activities against mtCcO and bacterial UqO. As results with mtCcO, 17 compounds had IC50 less than approximately 10 uM, and the most improved compound had IC50 of 0.1 uM, which is one-order better than that of the query inhibitor. Moreover, the result of the assay with bacterial UqO from *Escherichia coli* showed that 13 compounds had IC50 less than approximately 100 uM, and above all, the most effective compound had IC50 less than 10 uM. In conclusion, we succeeded to obtain various novel inhibitors of both mtCcO and bacterial UqO, which can be subsequently applied to the structure-activity relationship analysis and developing a novel antibiotic.

P2-17

The Contribution of Entropy in Drug-Protein Binding

Takeshi Tanaka tanaka@interprotein.com
Ken Ikeda ikeda@interprotein.com
Hirotsugu Komatsu komatsu@interprotein.com
Takao Matsuzaki tak-matsuzaki-45@cb3.so-net.ne.jp

Interprotein Corporation, SIC-2-205, 5-4-30 Nishihashimoto, Midori-ku, Sagamihara, Kanagawa, 252-0131, Japan

Keywords: in Silico Screening, Protein-Protein Interaction, Small Molecule Inhibitor, Entropy, Drug Design

The entropy change of drug-protein binding can now be observed by Isothermal Titration Calorimetry (ITC). The reported values of the entropy change of drug-protein binding range up to 10 kcal/mol, which consists significant portion of free energy change. The entropy effect for drug-protein binding can be positive or negative depending on the nature of binding reaction. In either case, drug-protein binding can not be predicted correctly without considering entropy change. This seems to be the main reason why conventional docking methods fail in most cases. We have developed a system which consider entropy contribution and concentrates compounds with high total energy into higher rank. It has been successfully applied for 20 PPI targets and 10 enzyme targets. The details about Runx1-CBFβ are shown.

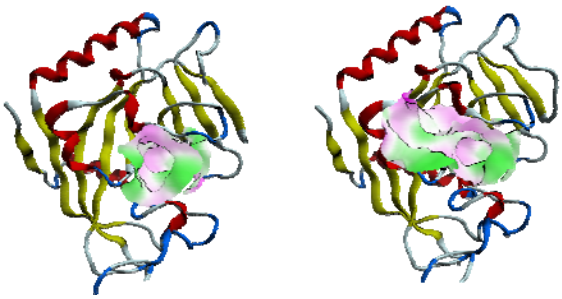
Evaluation of incorporation of protein flexibility for computational structure-based drug design

Noriaki Okimoto[1] okimoto@riken.jp
Yoshinori Hirano[1] hirano@riken.jp
Shigeo Fujita[1] shigeo.fujita@riken.jp
Makoto Taiji[1] taiji@riken.jp

1 QBiC, RIKEN, QBiC Building B, 6-2-4 Furuedai, Suita, Osaka 565-0874, Japan, Japan
Keywords: molecular dynamics simulation, molecular docking, binding pocket, ligand binding

A large number of available therapeutic drugs target protein molecules. Proteins fluctuate dynamically under physiologic conditions, and this flexibility is essential for ligand binding, a key protein function. Apparent conformational changes upon ligand binding have been observed in several pharmacologically relevant proteins. Although the importance of the effects on protein dynamics and solvent water molecules is widely recognized, many computational methods for structure-based drug design do not fully consider related effects. Thus far, there have been considerable efforts to incorporate these effects on protein flexibility and solvent water molecules into molecular docking. Here, we focus on the effect on the incorporation of protein flexibility for molecular docking. The protein flexibility was represented by the use of the multiple receptor structures obtained from the computational approaches. The computational approaches including molecular dynamics (MD) simulations are used to explore the conformations of the receptor protein before molecular docking.

In this study, tankyrase-ligand systems[1,2] were evaluated for the effect of incorporation of protein flexibility. First, we utilized several MD-based sampling techniques to obtain multiple protein structures. Furthermore, we try to select characteristic druggable binding pocket structures. Finally, we evaluate performance of molecular docking using these multiple conformations. We will report the detailed results.



Typical crystal structures of tankyrase protein:
The molecular surface representations indicate the binding pockets.

References

1. Karlberg, T.; Markova, N.; Johansson, I.; Hammarstrom, M.; Schutz, P.; Weigelt, J.; Schuler, H. J Med Chem 2010, 53(14), 5352-5355.

2. Narwal, M.; Venkannagari, H.; Lehtio, L. J Med Chem 2012, 55(3), 1360-1367.

Computer simulation-based prediction of drug-induced arrhythmia by evaluating repolarization reserve

Shingo Murakami shingo.murakami@med.toho-u.ac.jp
Taichiro Tomida tomida@med.toho-u.ac.jp
Yoshinori Mikami ymikami@med.toho-u.ac.jp
Masanori Ito itomasanori@med.toho-u.ac.jp
Satomi Adachi-Akahane satomiaa@med.toho-u.ac.jp

Department of Physiology, School of Medicine, Toho University, 5-21-16 Omori-Nishi, Ota-ku, Tokyo 143-8540, Japan

Keywords: Drug induced arrhythmia. Action potential simulation, Repolarization reserve

Some of drugs under development and in clinical use turned out to induce arrhythmia and were withdrawn from the market. The drug-induced arrhythmia is mainly accounted for by prolongation of action potential duration (APD) due to a block of one type of repolarizing K^+ currents in cardiac myocyte called I_{Kr}. Therefore, time-consuming and costly *in vitro* hERG screenings are conducted for all candidate compounds. However, it turned out that *in vitro* hERG screening also excludes potentially useful drug candidates that may not induce arrhythmia. The risk of drug-induced arrhythmia differed among various I_{Kr} blockers because normal repolarization of membrane potential in cardiac myocytes is determined by multiple and redundant mechanisms. The compensation of the repolarizing component by other currents is called repolarization reserve, which has been used only for a qualitative explanation to account for different risks among drugs.

To assess the risk of drug-induced arrhythmia, we redefined the repolarization reserve and examined if the quantified repolarization reserve can be used to predict the APD prolongation caused by various drugs. In the present study, we developed a method for quantification of repolarization reserve. We redefined repolarization reserve as a source of repolarizing currents that are activated during prolonged depolarization. By using human ventricular myocyte simulation models, we searched a novel quantification method based on this definition. We calculated a new quantitative index of repolarization reserve by simulating action potential (AP) clamp with prolonged AP waveforms and the human ventricular myocyte models.

We verified the effectiveness of this method by using the O'Hara and Rudy model of human ventricular myocytes. The calculated index quantitatively accounted for prolongation of APD under various conditions in cardiac myocytes. We also confirmed that the quantified repolarization reserve could be used to predict the prolongation of APD by different drugs. The proposed index of repolarization reserve is expected to contribute to further understanding of APD prolongation and predicting the risk of drug-induced arrhythmia.

P2-20

Ligand binding site analysis with protein flexibility for fragment to lead optimization

Masa-aki Ito [1]
31343u@ube-ind.co.jp

Takatsugu Hirokawa [2, 3]
t-hirokawa@aist.go.jp

[1] Pharmaceuticals Research Laboratory, UBE INDUSTRIES,LTD.,
1978-5,kogushi,Ube,Yamaguchi 755-8633,Japan
[2] Molecular Profiling Research Center for Drug Discovery, National Institute of Advanced Industrial Science and Technology,
2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan
[3] Division of Biomedical Science, Faculty of Medicine, University of Tsukuba,
1-1-1 Tennodai, Tsukuba-shi, Ibaraki 305-8575, Japan

Keywords: Molecular dynamics, Drug design, Fragment to lead optimization

With the development of increasing number of protein structure determination, pharmacological and drug discovery research based on the structural biology data have been accelerated. However, some of experimental structural data may be inconvenient to use for drug discovery due to the constraints on crystallization conditions or undesirable biological state. In silico analysis including molecular dynamics (MD) simulation can help to bridge the gap between original state of experimental structural data and drug discovery oriented structural models. In case of ligand biding analysis, a variety of pocket detection algorithms are available for the analysis of static protein structures. However, since proteins are dynamic entities. Static structures represent a serious limitation to account for the intrinsic plasticity of the binding pocket including pre-existing equilibrium, induced fit model, cryptic and allosteric sites. Characterization of cavities taking into account protein conformational ensembles should be valuable.

In this work, we developed workflow of ligand binding site analysis using MD simulation and chemoinformatics: (1) generating conformational ensemble of the target protein using MD simulation with 100ns time step based on the input structure of apo-form and fragment-bound form, (2) alpha spheres are detected on different pre-aligned conformation of protein and (3) analysis of the density of surrounding alpha spheres and frequency for fragment to lead optimization. For example, we applied this workflow to evaluate the fragment to lead expansion by fragment growing on DPP4 protein. Two type of MD simulation were performed, apo-form and fragment-bound form. As the result of MD simulation based on input structure of apo-form, fragment and lead were overlapped with frequency map at about 40 and 20%, respectively. We found the additional two grid spaces around fragment at frequency map at 20% and lead functional group by fragment growing occupied one of them. These results tell us if ligand design by fragment free approach, three types of lead shape will be acceptable. These findings should prove useful in lead optimization from fragment by our approach.

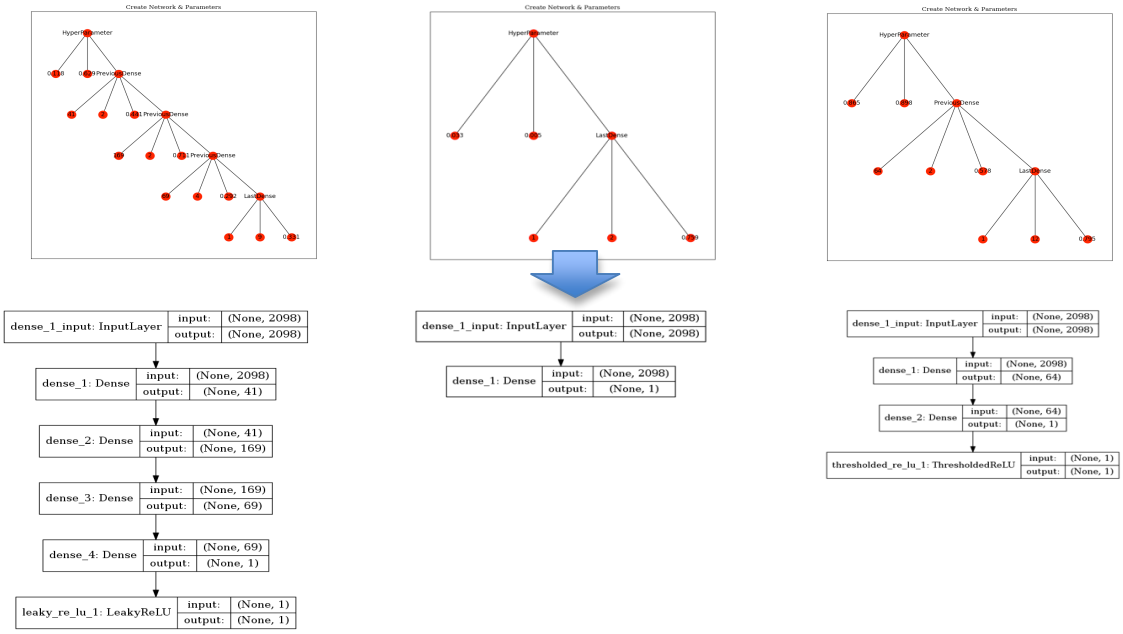
A Genetic Approach to Deep Leaning in Prediction of Molecular Properties

Yoshinori Wakabayashi[1] wakabayashi@by-hex.com
Tadaaki Mashimo[2,3] mashimo-tadaaki@aist.go.jp
Mitsuhito Wada[2] m-wada@aist.go.jp

[1] BY-HEX LLP, 1-19-14 Shimizu Simizu, Suginami-ku, Tokyo 167-0033, Japan
[2] Technology Research Association for Next generation natural products chemistry, 2-3-26 Aomi, Koto-ku, Tokyo 135-0064, Japan
[3] IMSBIO Co., Ltd., 4-21-1 Higashi-ikebukuro, Toshima-ku, Tokyo 170-0013 Japan

Keywords: myPresto, RDKit, Descriptor, Regression, Deep Learning, Deep Neural Network, Genetic Programing, Auto-Building, Auto-Turning, Cloud

Deep learning, a branch of machine learning utilizing multiple processing layers, is rapidly advancing many areas including image, text and voice recognition, and now being applied to drug discovery and cheminfomatics. However, it is difficult to design deep neural networks (DNN) and to find optimal hyper-parameters. In this work, we propose an efficient method to automatically set the number of layers and the initial value of the parameters of neural networks on the basis of genetic programming. To demonstrate how the automated method works well on predicting chemical properties, DNN models were trained for estimating the octanol/water partition coefficient. We used experimental LogP data retrieved from public database and linked those to molecular fingerprints, topological properties and physicochemical properties. Cross-validation experiments showed that the DNN-based models achieved sufficient accuracy in predicting LogP values.



P2-22

Seed Generation by Integrating Computer-aided Drug Design with Complementary Experimental Methods

Taiji Oashi[1,2]
taiji.oashi@kyowa-kirin.co.jp

Michihiko Suzuki[2]
michihiko.suzuki@kyowa-kirin.co.jp

Osamu Saku[3]
osamu.saku@kyowa-kirin.co.jp

Yuki Takayama[2]
yuki.takayama@kyowa-kirin.co.jp

Yuichi Takahashi[2]
yuichi.takahashi@kyowa-kirin.co.jp

Hikaru Miyagi[1]
hikaru.miyagi@kyowa-kirin.co.jp

Kuniyuki Kishikawa[2]
kuniyuki.kishikawa@kyowa-kirin.co.jp

Fumikazu Shinohara[4]
fumikazu.shinohara@kyowa-kirin.co.jp

Jun-ichi Saito[5,6]
jun.saito@kyowa-kirin.co.jp

[1] Research Core Function Laboratories, Research Functions Unit, R&D Division, Kyowa Hakko Kirin Co., Ltd 3-6-6 Asahi-machi, Machida-shi, Tokyo, 194-8533, Japan
[2] Chemical Research Laboratories, Research Functions Unit, R&D Division, Kyowa Hakko Kirin Co., Ltd 1188 Shimotogari, Nagaizumi-cho, Sunto-gun, Shizuoka, 411-8731, Japan
[3] Environment and Safety Section, Ube Plant, Production Division, Kyowa Hakko Kirin Co., Ltd 2547-3 Fujimagari, Ube-shi, Yamaguchi, 755-8501, Japan
[4] Innovative Technology Laboratories, Research Functions Unit, R&D Division, Kyowa Hakko Kirin Co., Ltd 3-6-6 Asahi-machi, Machida-shi, Tokyo, 194-8533, Japan
[5] R&D Planning Department, R&D Division, Kyowa Hakko Kirin Co., Ltd 1-9-2 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan
[6] Open Innovation Department, R&D Division, Kyowa Hakko Kirin Co., Ltd 1-9-2 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan

Keywords: Structure-based drug design, 3D-database screening, Druggability evaluation, Oral bioavailability prediction,

Structure-based drug design (SBDD) has been used to obtain the compounds that could fit the target site of interest, modulating the disease-associated biological process. Here we will present how a combination of computational and experimental methods was utilized in the early seed generation process by two examples; (1) identification of novel small-molecule inhibitor for a specific target potentially involved in cancer-related disease via 3D-database screening and *in vitro* enzyme inhibition assay, (2) discovery of potency-enhancing non-natural nucleotide for siRNA by using 3D-database screening, NMR-based binding assay, X-ray crystallography, and *in vitro* siRNA knockdown experiment. For further improving the seed generation process, we will also discuss key elements in the computational approach, including identification and druggability evaluation of putative ligand-binding site [1] and *in silico* screening as well as prediction of potential bioavailability and druglikeness for compounds selected from database screen prior to the final selection for experimental assay [2].

[1] Zhong, S., and MacKerell, A.D. Binding response: a descriptor for selecting ligand binding site on protein surfaces. *Journal of Chemical Information and Modeling*, 47:2303-2315, 2007.
[2] Oashi, T., Ringer, A.L., Raman, E.P., and Mackerell, A.D. Automated selection of compounds with physicochemical properties to maximize bioavailability and druglikeness. *Journal of Chemical Information and Modeling*, 51:148-158, 2011.

P3-01

A Graph Genome Browser

Toshiyuki Yokoyama tfdvrt@gmail.com
Yoshitaka Sakamoto yafdc1115@outlook.jp
Masahide Seki mseki@edu.k.u-tokyo.ac.jp
Yutaka Suzuki ysuzuki@edu.k.u-tokyo.ac.jp
Masahiro Kasahara mkasa@edu.k.u-tokyo.ac.jp

Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa City, Chiba 277-8583, Japan

Keywords: NGS, Graph Genome, Structural Variations, Visualization, Genome Browser

Background
Continuous reduction in sequencing cost and continuous improvement in sequencing throughput realized population-scale sequence data, which underscores the importance of the development of graph genome analysis for handling a large number of genomes with variations of different scales. Since most part of the genomes is shared between individuals, we need to focus on differences in order to minimize the computation time. Single-nucleotide variants or small insertions and deletions (indels) called from those data can be handled more accurately by a graph genome. On top of that, long-read sequencing enables us to more accurately identify large-scale structural variants (SVs) and fusion genes, which include duplications or indels as loops or branching and merging against a reference genome. Graph genomes have been considered promising especially in human variation analysis.

Although graph genome analysis is inextricably linked to the visualization of graph genomes, means to visualize graph genomes are yet to be developed. Existing genome browsers can visualize linear genomes but not graph genomes. Existing tools for handling graph genomes at the human genome scale cannot visualize genomes interactively and intuitively.

Results
We propose a graph genome browser for visualizing genome constructed by combining mathematical graphs with biological annotations. We define a sequence graph as a bidirected graph composed of a set of multiple DNA sequences as nodes and the corresponding end-to-end connections as edges. With these representations, we aim to visualize cancer genomes including SVs or fusion genes through comparison with the reference genome using sequencing data including long reads.

Conclusions
We implement a novel graph genome browser in order to provide an infrastructure for graph genome analysis. Our graph genome browser enables genome researchers to interactively explore biologically meaningful features observed as structures on a graph.

P3-02

Fragmentation prediction using neutral loss-fragment pairs

Hiroyuki Yamamoto[1] h.yama2396@gmail.com
Hiroshi Tsugawa[2] hiroshi.tsugawa@riken.jp

[1] Human Metabolome Technologies, Inc., 246-2 Mizukami Kakuganji, Tsuruoka, Yamagata 997-0052, Japan
[2] RIKEN Center for Sustainable Resource Science, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Keywords: MS/MS Fragmentation prediction, Machine learning, Metabolomics

In metabolomics, unknown peaks that chemical structures have not been determined are sometimes selected as metabolomic biomarker candidates in disease associated biomarker discovery. Recently, some chemoinformatics approaches using MS/MS spectra with the objective with structural elucidation of unknown peaks have been reported [1, 2]. In this study, we aimed to predict fragmentation by using neutral loss-fragment pairs.

For 438 metabolites, we cut of all combination of chemical bonds in two step computations. The disconnected structure pairs were compared with neutral loss and fragment pairs registered in MassBank database [3], and matched pairs were assigned as positive samples and unmatched pairs as negative samples. For 424 metabolites, PubChem fingerprint for the disconnected structures was computed. A fragment feature vector was generated by the similar way as generation of the feature vector from the compound pair in the prediction of the enzymatic reaction by Kotera et al. [4]. The fragment feature vector was set as explanatory variable and matched or unmatched information as a binary response variable to make a prediction model of fragmentation. Regularized logistic regression approaches of ridge logistic regression, elastic net and lasso were performed. A regularized parameter was optimized by cross-validation, and the prediction accuracy was computed for independent test set. A maximum value of area under the curve of the receiver observation curve for the test set was 0.7839 by using logistic regression with lasso. We plan to examine to improve prediction accuracy by using machine learning approaches such as deep learning in the future.

[1] Tsugawa, H. et al., Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software, Analytical Chemistry, 88(16):7946-58 (2016).
[2] Duhrkop, K. et al., Searching molecular structure databases with tandem mass spectra using CSI:FingerID, PNAS, 112(41):12580-5 (2015).
[3] Horai, H. et al., MassBank: a public repository for sharing mass spectral data for life sciences, Journal of Mass Spectrometry, 45(7):703-14 (2010).
[4] Kotera, M. et al., Supervised de novo reconstruction of metabolic pathways from metabolome-scale compound sets, Bioinformatics, 29(13):i135-44 (2013).

P3-03

Target Genes and Pathways in Drug Development for Rare Diseases in Clinical Trials

Ryuichi Sakate[1] rsakate@nibiohn.go.jp
Akiko Fukagawa[1] a-fukagawa@nibiohn.go.jp
Masashi Suzuki[2] suzuki.masashi@me.mt-pharma.co.jp
Masami Morita[2] morita-opir@jpma.or.jp
Akifumi Matsuyama[1] akifumi-matsuyama@nibiohn.go.jp

[1] National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki-shi, Osaka 567-0085, Japan
[2] Office of Pharmaceutical Industry Research, Japan Pharmaceutical Manufacturers Association, Nihonbashi Life Science Bldg., 2-3-11 Nihonbashi-honcho, Chuo-ku, Tokyo 103-0023, Japan

Keywords: Rare disease, Drug development, Clinical trial

Unmet medical needs in drug development for rare diseases are growing worldwide. In order to contribute to the drug development and related studies, trends in clinical trials for 306 rare diseases (intractable diseases) designated by the Ministry of Health, Labour and Welfare Japan had been analyzed [1-4]. In this study, further analyses were conducted to clarify the relation between tested drugs and their target genes and pathways. As a result, it was found that, for 122 rare diseases, 287 genes in 109 pathways (KEGG) were targeted by 338 drugs (DrugBank). Information of how drugs, genes and targets were shared by diseases was obtained. This information is expected to facilitate drug repositioning and studies of pathogenic mechanism.

[1] Suzuki, M., Sakate, R., and Fukagawa, A., [Status of implementation of clinical trials on intractable diseases designated in Japan], *The Office of Pharmaceutical Industry Research News*, 48:23-27. 2016
[2] Suzuki, M., Sakate, R., and Fukagawa, A., [Implementers of clinical trials for the designated intractable diseases], *The Office of Pharmaceutical Industry Research News*, 49:29-34. 2016
[3] Suzuki, M., Sakate, R., and Fukagawa, A., [Connecting information of the designated intractable diseases to databases], *The Office of Pharmaceutical Industry Research News*, 49:35-46. 2016
[4] Suzuki, M., Sakate, R., and Fukagawa, A., [Launched drugs used for clinical trials of the designated intractable diseases], *The Office of Pharmaceutical Industry Research News*, 50:39-43. 2017

P3-04

Prediction of natural biosynthetic pathway by reverse synthetic analysis

Tsubasa Matsumoto
matsumoto.t.ap@m.titech.ac.jp

Kohei Amano
amano.k.ac@bio.titech.ac.jp

Miyabi Hishinuma
hishinuma.m.aa@m.titech.ac.jp

Masaaki Kotera
maskot@bio.titech.ac.jp

School of Life Science and Technology Department of Life Science and Technology, Tokyo Institute of Technology, 2-12-1, Ookayama, Meguro-ku, Tokyo, 152-8550, Japan

Keywords: Reverse synthetic analysis, Machine learning

The relationship between natural products and enzymes that synthesize them and their genes has been extensively studied in the fields of synthetic biology, metabolic engineering, phytochemistry and the like. However, the number of natural products whose metabolic pathways are unknown, and the number of enzyme genes whose function is unknown are both increasing. So it is desired to develop a method for exhaustively connecting these two types of information.

Most of the informatics methods on the metabolic pathway are gene expression level and metabolite quantity for known metabolic pathways such as glycolysis system and Krebs cycle. On the other hand, Kotera and colleagues have established the methods such as predicting an enzyme gene (or enzyme protein) that catalyzes a specified reaction[1][2] or selecting a combination that can be a substrate / product pair of an enzyme reaction from tens of thousands of compounds[3][4].

In this study, we develop a chemical informatics method for predicting the biosynthetic pathway of various natural products using large scale computer system and extracting biosynthetic genes. By integrating reverse synthetic analysis and machine learning, we propose a framework to predict chemical transformations that make up the unknown metabolic pathway and support the identification of enzyme genes and enzyme proteins involved in it.

[1] Tabei Y, Yamanishi Y, Kotera M. Simultaneous prediction of enzyme orthologs from chemical transformation patterns for de novo metabolic pathway reconstruction. Bioinformatics, 32, i278-i287, 2016.
[2] Moriya Y, Yamada T, Okuda S, Nakagawa Z, Kotera M, Tokimatsu T, Kanehisa M, Goto S. Identification of Enzyme Genes Using Chemical Structure Alignments of Substrate-Product Pairs. J Chem Inf Model. 56(3):510-516, 2016.
[3] Yamanishi Y, Tabei Y, Kotera M. Metabolome-scale de novo pathway reconstruction using regioisomer-sensitive graph alignments. Bioinformatics. 31(12):i161-i170, 2015.
[4] Kotera M, Tabei Y, Yamanishi Y, Tokimatsu T, Goto S. Supervised de novo reconstruction of metabolic pathways from metabolome-scale compound sets. Bioinformatics, 29(13), i135-i144, 2013.

P3-05

Construction of a Transcriptome-based Toxicity Prediction Model of Adjuvants

Natsuko Kishishita[1,2,4]
nkishishita@nibiohn.go.jp

Yoshinobu Igarashi[3]
y-igarashi@nibiohn.go.jp

Hiroshi Yamada[3]
h-yamada@nibiohn.go.jp

Ken Ishii[1,2,4]
kenishii@biken.osaka-u.ac.jp

[1-3] [1]Laboratory of Mockup Vaccine, [2]Laboratory of Adjuvant Innovation, and [3]Laboratory of Toxicogenomics Informatics, National Institutes of Biomedical Innovation, Health and Nutrition, Ibaraki, Osaka, 567-0085, Japan
[4] Laboratory of Vaccine Science, Immunology Frontier Research Center, World Premier International Research Center, Osaka University, Suita, Osaka 565-0871, Japan

Keywords: Adjuvant, Toxicity, Transcriptome, Biomarker

Research and development of adjuvants is very active today to improve the efficacy of vaccines against various diseases such as cancers and chronic disorders. Many adjuvants are ligands for pattern-recognition receptors that are important for innate immunity, and they skew immune responses in a certain way to induce Th1- or Th2-type response, for instance. Many adjuvants are developed over the world by pharmaceutical and venture companies as well as academia, and some of them are in clinical trials. And yet, it is not easy to evaluate the safety and efficacy of adjuvants accurately in preclinical studies or clinical trials to avoid adverse reactions in humans.

We have constructed the Adjuvant Database (http://adjuvantdb.nibiohn.go.jp/) to evaluate safety and efficacy of adjuvants with the grant from the Ministry of Health Labour and Welfare and AMED. Toxicity tests with rats were performed to obtain gene expression profiles of major tissues such as liver and kidney, and dataset were analyzed collectively with a large toxicogenomics database developed by our institute (http://toxico.nibiohn.go.jp/). We set drugs and adjuvants with known toxicity as “positive controls” to construct a toxicity prediction model based on gene expression data of each tissue. In addition, we analyzed several blood factors including conventional tissue-damage biomarkers (e.g., AST and ALT) and microRNAs to search for new biomarkers. As a result, a toxicity prediction model determined reactogenic adjuvants as “toxic” and those adjuvants that are clinically-approved or under clinical trials as “safe”. In fact, one of the “toxic” adjuvant was known for liver toxicity, greatly affected a set of genes involved in cell death and increased multiple tissue-damage biomarkers in the periphery blood. Meanwhile, “safe” adjuvants did not show toxicity-related features as expected. We further plan to optimize a toxicity prediction model to generalize molecular signatures of toxicity and aim to identify such toxicity biomarkers that are applicable to humans.

P3-06

Prediction of Biosynthetic Basic Parts of Compounds with Complicated Structures

Kohei Amano[1] amano.k.ac@m.titech.ac.jp
Tsubasa Matsumoto[1] matsumoto.t.ap@m.titech.ac.jp
Miyabi Hishinuma[1] hishinuma.m.aa@m.titech.ac.jp
Masaaki Kotera[1] maskot@bio.titech.ac.jp

[1] School of Life Science and Technology, Tokyo Institute of Technology, M6-2, 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan

Keywords: Biosynthetic pathway analysis, Genetic Algorithm

Many pharmaceutical compounds such as antibiotics are derived from natural products. Identification of their metabolic pathways is required not only for metabolic engineering but also for synthetic biology. Natural product chemists estimate reaction pathways or identify enzyme proteins on the basis of their own knowledge and experiences.

Most of the informatics methods on the metabolic pathway deal with known metabolic pathways such as glycolysis system and Krebs cycle, and do not focus on identifying previously unknown metabolism. On the other hand, Kotera and colleagues published the series of methods such as predicting an enzyme gene or an enzyme protein that catalyzes a specified reaction[1][2] or selecting a combination that can be a substrate / product pair of an enzyme reaction from tens of thousands of compounds[3][4].

We focused on the fact that living things make a lot of natural products from a limited number of basic parts. This study provides the method of predicting the biosynthetic basic parts from natural products that contain complicated structures. We propose a new approach that we clarify the basic metabolite in its metabolic pathway in advance to avoid combinatorial explosion. The database of compounds which can be candidates of the biosynthetic basic parts is prepared. These compounds are supposed to be detectable by metabolome analysis. We virtually cut any chemical bonds contained in the interesting compound, and decompose it into partial structures. We make several evaluation functions with some molecular fingerprints and molecular similarities, and decide which bonds are to be cut by comparing with the database. It is tested by genetic algorithm to adopt compounds which metabolic pathways are known. We believe our method can support the elucidation of biosynthetic pathway.

[1] Tabei Y, Yamanishi Y, Kotera M. Simultaneous prediction of enzyme orthologs from chemical transformation patterns for de novo metabolic pathway reconstruction. Bioinformatics, 32, i278-i287, 2016.
[2] Moriya Y, Yamada T, Okuda S, Nakagawa Z, Kotera M, Tokimatsu T, Kanehisa M, Goto S. Identification of Enzyme Genes Using Chemical Structure Alignments of Substrate-Product Pairs. J Chem Inf Model. 56(3):510-516, 2016.
[3] Yamanishi Y, Tabei Y, Kotera M. Metabolome-scale de novo pathway reconstruction using regioisomer-sensitive graph alignments. Bioinformatics. 31(12):i161-i170, 2015.
[4] Kotera M, Tabei Y, Yamanishi Y, Tokimatsu T, Goto S. Supervised de novo reconstruction of metabolic pathways from metabolome-scale compound sets. Bioinformatics, 29(13), i135-i144, 2013.

P3-07

A system level investigation to unravel mechanisms and potential direct targets of the osteogenic small molecule TH

Mochammad Ichsan[1]
ichsan@hamfaro.or.id

Kosuke Kanke[2]
kankekosuke@gmail.com

Hironori Hojo[1,3]
hojo@tetrapod.t.u-tokyo.ac.jp

Ung-il Chung[1,3]
tei@bioeng.t.u-tokyo.ac.jp

Shinsuke Ohba[1,3]
ohba@m.u-tokyo.ac.jp

[1] Department of Bioengineering, University of Tokyo, Tokyo 113-8656, Japan
[2] Department of Sensory and Motor System Medicine, University of Tokyo, Tokyo 113-0033, Japan
[3] Center for Disease Biology and Integrative Medicine, University of Tokyo, Tokyo 113-8655, Japan

Keywords: Kinase inhibition, Microarray, Molecular docking, Osteoblast differentiation, Reverse docking

Direct reprogramming of human somatic cells such as dermal fibroblasts (DFs) into several cell types without introduction of transcription factors or genetic manipulation has drawn much attention. Although TH, an osteogenic helioxanthin-derivative, has been shown to act as a robust small molecule inducing differentiation of several cell types into osteoblasts [1,2,3,4], detailed mechanisms and direct targets of TH remain unclear. In the present study, we aimed to unravel them in order to develop novel strategies to chemically reprogram DFs into osteoblasts. Microarray analysis revealed 1,166 differentially expressed genes during an osteoblast induction phase in TH-mediated stepwise differentiation of mouse embryonic stem cells (mESCs) into osteoblasts. We also confirmed the increased expression of osteoblast differentiation-related genes, such as *Col1a1*, *Ibsp*, *Runx2* and *Sp7*, in the dataset. Furthermore, we identified focal adhesion, ECM-receptor interaction, and PI3K-Akt signaling pathways as the most enriched pathways modulated by TH. Bioactivity prediction showed that TH potentially acted as a kinase inhibitor with the highest bioactivity score compared to other 22 kinase inhibitors obtained from Osteogenics DB, an osteoblast-related molecule database which we newly developed. We then identified 145 potential targets of TH by pharmacophore- and structure-based reverse docking and 62 kinases, which were potentially inhibited during the osteoblast induction phase, by a kinase enrichment analysis. By intersecting all identified potential targets of TH with 344 kinases which are targeted by osteogenic kinase inhibitors in Osteogenics DB, we identified 42 proteins as the most potential direct targets of TH that fulfilled all criteria used in this study. Lastly, we performed molecular docking analyses to predict binding affinity and binding specificity of TH against the most potential direct targets. Taken together, these findings suggest that TH achieves its osteogenic activity at least partially via modulating the focal adhesion, ECM-receptor interaction, and PI3K-Akt signaling pathways and potentially acts as a kinase inhibitor against several kinases.

[1] Ohba S et al. Biochem Biophys Res Commun. 357(4):854-60, 2007
[2] Nakajima K et al. Biochem Biophys Res Commun. 395(4):502-8, 2010
[3] Kanke K et al. Stem Cell Reports. 2(6):751–60, 2014
[4] Zujur D et al. Sci Adv. 3(5):e1602875, 2017

P3-08

Large-scale diseasome analysis of gene expression signatures toward understanding disease–disease associations and drug discovery

Michio Iwata[1]
m-iwata@bioreg.kyushu-u.ac.jp

Ryusuke Sawada[1]
rsawada@bioreg.kyushu-u.ac.jp

Yasuo Tabei[2]
yasuo.tabei@gmail.com

Yoshihiro Yamanishi[1,3]
yamanishi@bioreg.kyushu-u.ac.jp

[1] Division of System Cohort, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka, Fukuoka 812-8582, Japan
[2] RIKEN Center for Advanced Intelligence Project, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
[3] PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan

Keywords: diseasome, patients, gene expression, drug discovery, machine learning

Diseases are caused by dysfunction of the human biological system consisting of genes, proteins, and pathways. The pathogenesis is basically considered disease-specific, but characteristic molecular features are often similar among patients of different diseases, suggesting the commonality of the underlying molecular mechanisms. Most drugs are small compounds that modulate the activity of biological systems for the treatment of diseases. Thus, the knowledge on molecular correlations among diseases can be useful for drug discovery. Diseases are characterized by impaired expression of genes, and drugs modulate the activity of gene expression system. Thus, the use of transcriptome data for diseases and those for drug candidate compounds is a rational approach toward drug discovery. However, the integration of disease-specific and drug-induced transcriptome data has been an open question.

In this study we performed a large-scale diseasome analysis of gene expression signatures for 79 diseases (e.g., adrenoleukodystrophy, leukemia, Alzheimer's disease, asthma, atopic dermatitis, breast cancer, cystic fibrosis, dengue and inflammatory bowel disease). We extracted characteristic gene expression patterns that explain the specificity of each disease and commonality between different diseases. We also proposed a novel machine learning method for drug discovery from the integration of disease-specific and drug-induced transcriptome data using multi-layered diseasome similarities computed from various molecular features (e.g., gene expressions, diagnostic markers, disease-causing genes, disordered pathways, and environmental factors), and demonstrated the usefulness of the proposed method in terms of accuracy and applicability. The diseasome analysis is expected to be useful for understanding disease-disease associations and for drug discovery.

P3-09

Prediction of novel mitochondrial proteins having disease-associated mutations toward discovery of drug target

Kenichiro Imai[1]
kenichiro.imai@aist.go.jp

Yoshinori Fukasawa[2]
y-fukasawa@aist.go.jp

Kentaro Tomii[2]
k-tomii@aist.go.jp

Paul Horton[2]
horton-p@aist.go.jp

[1] Molecular Profiling Research Center for Drug Discovery, National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan
[2] Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

Keywords: Mitochondria, Disease-associated mutations, Proteome, Transcriptome

Mitochondria provide numerous essential functions for cells and their dysfunction leads to a variety of diseases. Thus, obtaining a complete mitochondrial proteome should be a crucial step toward understanding the roles of mitochondria in the diseases. Many mitochondrial proteins have been identified experimentally but a complete list is not yet available. To fill this gap, methods to computationally predict mitochondrial proteins from amino acid sequence have been developed and are widely used, but unfortunately, their accuracy is far from perfect. We developed MitoFates [1], an improved prediction method for cleavable N-terminal mitochondrial targeting signals (presequences) and their cleavage sites. MitoFates is a Support Vector Machine-based predictor and considers positively charged amphiphilicity, presequence motifs, position weight matrices modeling the presequence cleavage sites and classical features such as amino acid composition and physico-chemical properties. On independent test data, MitoFates attains better performance than existing predictors in both detection of presequences and in predicting their cleavage sites. In this study, we used MitoFates to look for novel mitochondrial proteins involved in human diseases. We obtained 83583 curated disease-associated variants from 6388 human diseases, which are collected form UniProt, ClinVar [2], and the GWAS Catalog [3] using DisGeNET [4]. Then we analyzed transcripts possessing the disease-associated variants using MitoFates and found more than 100 transcripts cording novel mitochondrial protein candidates involved in human diseases. We will report the functional annotation of the candidates. Also we will discuss interpretation of relevance between the candidates, the mutations and the diseases. The candidate list would be helpful in elucidating the role of mitochondria in health and disease.

[1] Fukasawa et al., MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites, *Mol Cell Proteomics*, 14(4):1113-26, 2015
[2] Landrum M.J. *et al.*, ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.*, 44:D862–D868, 2016.
[3] Welter D. *et al.*, I The NHGRI GWAS Catalog, a curated resource of SNP-trait associations, *Nucleic Acids Res.*, 42:D1001–D1006, 2014
[4] Piñero J., *et al.*, DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants, *Nucleic Acids Res.*, 45(D1):D833-D839, 2017

Image-based phenotypic profiling associated with EGF receptor trafficking and signal transductions reveals novel mode of action of drugs

Kenji Tanabe
tanabe.kenji@twmu.ac.jp

Medical Research Institute, Tokyo Women's Medical University, 8-1 Kawada-cho, Shinjuku-ku, Tokyo 162-8666, Japan

Keywords: High contents screening, Mode of action of drugs, EGF receptor, small-molecule inhibitors

Small-molecule compounds, including kinase inhibitors, are widely used as anti-cancer drugs. However, as almost all small-molecule compounds have multiple targets, it may cause an unexpected side effect and contribute to high attrition rate in clinical trials. Therefore, target deconvolution of drug candidates is a critical theme for drug development. Although various approaches are used for the target deconvolution, revealing a functionally unrelated target molecule, such as non-kinase targets for kinase inhibitors, is difficult. Recently, we established an image-based compound profiling, and found an unexpected non-kinase target (microtubules) for a most uni-specific kinase inhibitor (Tanabe, Sci. Rep. 2016). This study applied the image-based compound profiling for the target deconvolution of an inhibitor library, which consists of ~400 clinical drugs and various inhibitors. A549 cells, which express GFP-EGF receptor, were exposed to inhibitors, stimulated by EGF and visualized several major signaling molecules. Cells were photographed using automated fluorescent microscopy, and analyzed to extract ~500 image features to quantitatively evaluate their phenotypes. These extracted image features were used to classify inhibitors using unsupervised machine learning. As result, many clusters with unexpected combination of inhibitor were observed. For example, sodium channel inhibitor or transferase inhibitor were classified with some PI3K-Akt inhibitors. Interestingly, biochemical assay showed that they share target molecules associated with PI3K-Akt pathway. Similar observations were obtained in other clusters including cytoskeleton inhibitors. This result indicates that this image-based compound profiling is a powerful tool to reveal mode of action of drugs.

[1] Tanabe, K., Image-based compound profiling reveals a dual inhibitor of tyrosine kinase and microtubule polymerization, *Scientific Reports*, 6:25095, 2016.

P3-11

Analysis of “uncallable” regions in the human reference genome associated with gene annotation and segmental duplication

Shu Saikawa[1]
5944563671@edu.k.u-tokyo.ac.jp

Masahiro Kasahara[2]
mkasa@edu.k.u-tokyo.ac.jp

[1] Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa 277-8568, Japan.

Keywords: Next generation sequencing, Disease omics analysis

Genome wide association study (GWAS) is one of the most important applications of sequencing technology and contributed to the detection of gene regions associated with diseases in many researches[1][2]. However, most of those projects did not result in the detection of highly correlated SNV and the effectiveness of GWAS study is questioned by researchers [3].

We expect that GWAS study so far does not achieve adequate results because there are some problems in bioinformatics analysis processes and software. Particularly, we focused on the problems in mapping procedures. Read mapping software often used in genome sequencing research such as BWA [4] and Bowtie [5] discard reads or select a certain position randomly from multiple positions (in this case, map quality becomes zero) when they deal with multi-mapping reads. In both cases multi-mapping reads are not used in subsequent analysis.

The probability that certain number of reads are uniquely mapped in each genome position can be modeled by poisson distribution. Therefore, in this research we calculated the probability that enough number of reads for SV detection are mapped in each genome position using poisson distribution. Then we focused on the regions where this probability is low (which we call “uncallable region”) and associated with gene annotation and segmental duplication.

[1] Okada, Y., *et al.,* Genetic of rheumatoid arthritis contributes to biology and drug discovery, *Nature* 506:376-381, 2013.

[2] Imamura, M., *et al.,* Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes, *Nature Communications.* 7:10531, 2016.

[3] Gallaway, E., New concerns raised over value of genome-wide disease studies, *Nature*, 546:463, 2017.

[4] Li, H. & Durbin, R., Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics*, 25:1754-1760, 2009.

[5] Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biology*, 10:R25, 2009.

Toward the next step in G protein-coupled receptor research: a knowledge-driven analysis for the next targets in drug discovery

Koji Nagata[1] aknagata@mail.ecc.u-tokyo.ac.jp
Yukie Katayama[1] ykatayam@ims.u-tokyo.ac.jp
Tomomi Sato[2] tomomi@post.kek.jp
Yeondae Kwon[1] ayekwon@mail.ecc.u-tokyo.ac.jp
Takeshi Kawabata[3] kawabata@protein.osaka-u.ac.jp

[1] Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
[2] Structural Biology Research Center, Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan
[3] Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

Keywords: DDSS, OMIM, Orphan GPCR, VaProS

More than 800 G protein-coupled receptor (GPCR) genes have been discovered in the human genome. Towards the next step in GPCR research, we performed a knowledge-driven analysis of orphan class-A GPCRs that may serve as novel targets in drug discovery [1]. We examined the relationship between 61 orphan class-A GPCR genes and diseases using the OMIM database [2] and the DDSS tool [3]. The OMIM database contains data on disease-related variants of the genes. Particularly, the variants of GPR101, GPR161, and GPR88 are related to the respective diseases: growth hormone-secreting pituitary adenoma 2; pituitary stalk interruption syndrome (interpretations conflicting); and childhood-onset chorea with psychomotor retardation. On the other hand, the DDSS tool suggests that 49 out of the 61 orphan receptor genes are related to diseases, judging from their co-occurrences in abstracts of biomedical literature. Notably, GPR50 and GPR3 are related to as many as 25 and 24 disease-associated keywords, respectively. The aforementioned five orphan GPCRs were characterized genetically, structurally and functionally using the structural life science data cloud VaProS [4], so as to evaluate their potential as next targets in drug discovery.

[1] Nagata, K., Katayama, Y., Sato, T., Kwon, Y., and Kawabata, T., Toward the next step in G protein-coupled receptor research: a knowledge-driven analysis for the next potential targets in drug discovery, *J. Struct. Funct. Genomics*, 17, 111-133, 2017.
[2] Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., and Hamosh, A., OMIM.org: Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic disorders, *Nucleic Acids Res.*, 43, D789-798, 2015.
[3] Kwon, Y., Shimizu, S., Sugawara, H., and Miyazaki, S., A novel evaluation measure for identifying drug targets from the biomedical literature, *IPSJ Transactions on Bioinformatics*, 7, 16-23, 2014.
[4] Gojobori, T., Ikeo, K., Katayama, Y., Kawabata, T., Kinjo, A. R., Kinoshita, K., Kwon, Y., Migita, O., Mizutani, H., Muraoka, M., Nagata, K., Omori, S., Sugawara, H., Yamada, D., and Yura, K., VaProS: a database-integration approach for protein/genome information retrieval, *J. Struct. Funct. Genomics*, 17, 69-81, 2017.

P3-13

Impact of read trimming on Illumina paired-end-sequencing samples in the microbiome analysis using Qiime

Attayeb Mohsen attayeb@nibiohn.go.jp
Jonguk Park jonguk@nibiohn.go.jp
Yi-An Chen chenyian@nibiohn.go.jp
Hitoshi Kawashima hkawashima@nibiohn.go.jp
Kenji Mizuguchi kenji@nibiohn.go.jp

[1] National institutes of Biomedical Innovation, Health, and Nutrition, 7-6-8, Saito-Asagi, Ibaraki City, Osaka 567-0085, Japan

Keywords: Quality trimming, Next generation sequencing, Microbiome data analysis.

Paired-end sequences are generated by sequencing the target DNA from its both sides producing partially overlapped two pair-end reads for the target DNA, then both reads are merged to produce a longer final read. The quality of the resulting final read is affected by the base-specific quality for the original reads. The Illumina paired-end sequencing technology yielded a good quality product, and is the most commonly used technology nowadays. However, base-specific quality scores reduce substantially with machine cycles, leading to low-quality bases as sequencing approaches to the end of the read, which worsens the merging process and the subsequent analysis steps in Qiime[1].

Trimming[2] is used to remove the parts with low quality, and helps improve the outcome of the merging step and overall analysis. In this poster, we examined the importance of the trimming quality threshold using different datasets, and showed the effect of this threshold on the number of reads preserved after trimming, the number of picked operational taxonomic units (OTUs) and the diversity measures.

[1] Caporaso J. G., *et al.*, QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7(5): 335-336, 2010.
[2] Bushnell B., BBTools available at: (http://sourceforge.net/projects/bbtools/).

Has Proximity in Human Genome been Conserved through 2-Round Whole Genome Duplication ?

Hyogo Kajiyama[1] dbo64283@kwansei.ac.jp

Hiroyuki Toh[1] tohhir@kwansei.ac.jp

[1] Department of Biomedical Chemistry, School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda, Hyogo, 669-1337, Japan

Keywords: Two-Round Whole Genome Duplication, Hi-C

Recent development of the NGS technology has enabled us to analyze various biological phenomena. One of the applications of the NGS is “Hi-C”, which can detect spatially adjacent chromosomal regions in human genome. Then, a model for the structural conformation of the human genome is built based on the Hi-C data [1, 2]. On the other hand, it is suggested that the vertebrates have undergone the two-round whole genome duplication (2RWGD). Associated with the events, four copies of paralogs were generated in the vertebrate genomes, although the numbers may have changed due to the lineage specific deletions and/or duplications after the events. In this research, we examined whether the adjacency of chromosomal regions has been conserved after the 2RWGD or not. At first, we obtained a data from NCBI GEO [3], which includes proximity scores between chromosomal regions generated from a Hi-C experiment [1]. In the data, each chromosome is divided into non-overlapped regions or bins of 1Mbp in length. We also obtained a whole set of pairs, gene name and protein ID, included in the human genome (hg19) from UCSC Genome Browser [4]. In addition, the information about the chromosomal locations of human genes was obtained through Galaxy [5]. Then, these data were integrated to generate a table, in which each row corresponds to a gene and consists of three elements, gene name, chromosomal location and protein ID of the gene. Next, we obtained the amino acid sequence data encoded by human genes from Uniprot [6]. The sequences thus obtained were sorted so that the proteins encoded in the same bin constitute one set, to make the sequence comparison described below easier. Thridly, we obtained a list of the paralogous genes associated with 2RWGD from OHNOLOGS [7,8]. Then, the bins including the paralogs were identified. The gene products encoded by the bins spatially adjacent to the different bins containing the paralogs were compared by making the alignments for all possible pairs between the adjacent bins. If paralogs are observed among the adjacent bins, it suggests the conservation of chromosomal adjacency after the 2RWGD. The conservation of the chromosomal contingency will be discussed based on the analysis.

[1] Lieberman-Aiden, E. *et al*. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950), 289-293.
[2] Li, C. *et al*. (2014). The 3DGD: a database of genome 3D structure. *Bioinformatics*, 30.11: 1640-1642.
[3] http://www.ncbi.nlm.nih.gov/geo/
[4] https://genome.ucsc.edu/
[5] https://usegalaxy.org/
[6] http://www.uniprot.org/
[7] Singh, P. P., Arora, J., & Isambert, H. (2015). Identification of ohnolog genes originating from whole genome duplication in early vertebrates, based on synteny comparison across multiple genomes. *PLoS Comput Biol*, *11*(7), e1004394.
[8] http://ohnologs.curie.fr/

Development of a pharmacokinetics prediction system using multiscale integrated modeling: 1. A curated database combining public and newly acquired experimental data

Hitoshi Kawashima[1] hkawashi@nibiohn.go.jp
Reiko Watanabe[1] reiko-watanabe@nibiohn.go.jp
Tsuyoshi Esaki[1] tsuyoshi-esaki@nibiohn.go.jp
Rikiya Ohashi[1,2] rikiya-ohashi@nibiohn.go.jp
Daisuke Satoh[3] daisuke.satoh@level-five.jp
Kazuyoshi Ikeda[3] ikeda@level-five.jp
Chioko Nagao[1] chio@nibiohn.go.jp
Yayoi Natsume-Kitatani[1] natsume@nibiohn.go.jp
Kenji Mizuguchi[1] kenji@nibiohn.go.jp

[1] Laboratory of Bioinformatics, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8, Saito Asagi, Ibaraki-shi, Osaka 567-0085, Japan
[2] Discovery Technology Laboratories, Mitsubishi Tanabe Pharma Corporation, 2-2-50, Kawagishi, Toda-shi, Saitama 335-8505, Japan
[3] Software Division, Level Five Co., Ltd., 2-25, Kanda Suda-cho, Chiyoda-ku, Tokyo 101-0041, Japan

Keywords: Pharmacokinetics, Database, Prediction models

To address pharmacokinetic and toxicological issues in drug development, once the main source of late attrition of drug candidates, many pharmaceutical companies have now implemented early ADME or early toxicological studies. However, such approaches are difficult to emulate in the academic drug discovery environment. Therefore, we began an initiative “Development of a Drug Discovery Informatics System”, with support from the Japan Agency for Medical Research and Development (AMED) in collaboration with several other research groups. The main aim of this initiative is to develop accurate prediction systems for ADME and tox properties primarily targeting academic scientists. Our group’s focus is to develop a pharmacokinetics database and prediction models.

Any good prediction system depends on high-volume, high-quality training datasets. We collected pharmacokinetic and physicochemical parameters from the public bioactivity database, ChEMBL [1]. However, since ChEMBL compiles data obtained in different experimental conditions, we developed a curation workflow to select the data measured in compatible conditions and to reformat the results as appropriate for our prediction system.

In addition to the public data, we have been acquiring both *in vitro* and *in vivo* experimental data using consistent protocols. The *in vitro* experiments include physicochemical parameters such as solubility and distribution coefficient, and pharmacokinetic parameters such as metabolic stability, protein binding in plasma, protein binding in brain homogenate, and blood-to-plasma concentration ratio. The *in vivo* data include the drug concentrations in plasma and tissues after oral or intravenous administration of the drug and pharmacokinetic parameters calculated therefrom.

We stored all the data in a PostgreSQL database and developed a web application to view and analyze the database content. We plan to integrate this system with the cardiotoxicity and hepatotoxicity databases developed within the AMED initiative described above.

[1] https://www.ebi.ac.uk/chembl/

This work was conducted as part of “Development of a Drug Discovery Informatics System” supported by Japan Agency for Medical Research and Development (AMED).

Development of a pharmacokinetics prediction system using multiscale integrated modeling: 2. Classification of aqueous solubility measured using the dried-DMSO method

Tsuyoshi Esaki[1]
tsuyoshi-esaki@nibiohn.go.jp

Reiko Watanabe[1]
reiko-watanabe@nibiohn.go.jp

Hitoshi Kawashima[1]
hkawashima@nibiohn.go.jp

Yayoi Natsume-Kitatani[1]
natsume@nibiohn.go.jp

Chioko Nagao[1]
chio@nibiohn.go.jp

Rikiya Ohashi[1,2]
rikiya-ohashi@nibiohn.go.jp

Kenji Mizuguchi[1]
kenji@nibiohn.go.jp

[1] Laboratory of Bioinformatics, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8, Saito Asagi, Ibaraki-shi, Osaka 567-0085, Japan
[2] Discovery Technology Laboratories, Mitsubishi Tanabe Pharma Corporation, 2-2-50, Kawagishi, Toda-shi, Saitama 335-8505, Japan

Keywords: Aqueous solubility, Dried-DMSO, Machine learning

It has been demonstrated that high-throughput *in vitro* measurements for the evaluation of physicochemical, pharmacokinetic, and safety parameters can save time and cost in drug development and ultimately improve the success rate. Solubility is an essential physicochemical property having great influences on drug absorption. The oral administration of compounds with low solubility can not only have low effects, but also be affected by the individual variation. Therefore, *in silico* methods to estimate the solubility of compounds are needed and both free and proprietary software programs have been developed. However, solubility is one of the properties that have been reported to be difficult to predict with high accuracy [1]. In the academic drug discovery environment, it is unfeasible to collect a large volume of new experimental data and although a considerable amount of data are available from public databases, they came from a variety of different experimental conditions. Thus, in this study, we compiled publicly available solubility data that were measured using a single procedure, the dried-DMSO method, and constructed computational models to predict solubility and to classify target compounds from chemical structures alone. The dried-DMSO method is a high-throughput experimental method widely used in pharmaceutical companies. All the experimental data were collected from the public bioactivity database, ChEMBL (ver.23) [2], often by checking the original literature and excluding the data measured using other (or unclear) methods. The collected compounds data were separated into two groups, training and test sets. Different combinations of chemical descriptors and machine learning algorithms were examined on the training set and the constructed models were evaluated on the test set, resulting in an accurate system for predicting compound solubility measured using the dried-DMSO method.

[1] Hughes L.D., *et al.*, Why are some properties more difficult to predict than others? A study of QSAR models of solubility, melting point, and LogP, *J. Chem. Inf. Model.*, 48:220-232, 2008.
[2] https://www.ebi.ac.uk/chembl/

This work was conducted as part of “Development of a Drug Discovery Informatics System” supported by Japan Agency for Medical Research and Development (AMED).

P4-03

Development of a pharmacokinetics prediction system using multiscale integrated modeling: 3. Prediction of human plasma protein binding from chemical structures

Reiko Watanabe[1]
reiko-watanabe@nibiohn.go.jp

Tsuyoshi Esaki[1]
tsuyoshi-esaki@nibiohn.go.jp

Hitoshi Kawashima[1]
hkawashi@nibiohn.go.jp

Yayoi Natsume-Kitatani[1]
natsume@nibiohn.go.jp

Chioko Nagao[1]
chio@nibiohn.go.jp

Rikiya Ohashi[1,2]
rikiya-ohashi@nibiohn.go.jp

Kenji Mizuguchi[1]
kenji@nibiohn.go.jp

[1] Laboratory of Bioinformatics, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8, Saito Asagi, Ibaraki-shi, Osaka 567-0085, Japan
[2] Discovery Technology Laboratories, Mitsubishi Tanabe Pharma Corporation, 2-2-50, Kawagishi, Toda-shi, Saitama 335-8505, Japan

Keywords: Plasma protein binding, Protein unbound, Machine learning

The fraction of chemicals unbound to plasma proteins is an important determinant of the efficacy of a drug in pharmacokinetic and pharmacodynamic studies, since the free (unbound) form is available for pharmacological interactions. The unbound fraction affects the membrane permeability, metabolism and glomerular filtration rate, and consequentially it influences the volume of distribution in different tissues and the total clearance. In addition, unbound drugs exert pharmacological and toxicological effects. In the early stages of drug discovery, determining the fraction of drug unbound to plasma proteins in humans and in animal models provides an understanding of the ADME properties to assist candidate selection. Furthermore, it is important to predict the fraction of unbound drug in the brain for assessing brain drug delivery. Accordingly, the *in silico* prediction of unbound drug in plasma is perceived as an effective tool to mitigate the risk of late-stage attrition during drug development and to optimize further screening [1]. Human plasma contains several binding proteins such as human serum albumin (HSA), α-1 acid glycoprotein (AGP), lipoproteins and globulins. Consequently, only a few studies have reported *in silico* modeling to predict the binding ratio in intact plasma [2], although previous studies have reported the prediction of protein binding to simple substances such as HSA or AGP.

The aim of this study is to build high precision models to predict the fraction unbound in the intact human plasma. We employed three machine learning techniques, using topological descriptors to represent molecular structures, to generate models of human plasma protein binding based on the large data set of experimental values; these models were subsequently used to estimate the unbound fraction. For this purpose, multiple prediction models were created; a regression model to predict the protein binding in human plasma, a classification model to predict the fraction unbound in human plasma and different regression models based on the compound polarity such as acidic, basic and neutral.

[1] Lambrinidis G., *et al.*, In vitro, in silico and integrated strategies for the estimation of plasma protein binding. A review, *Adv. Drug Deliv. Rev.*, 86:27-45, 2015.
[2] Votano J.R., *et al.*, QSAR modeling of human serum protein binding with several modeling techniques utilizing structure-information representation, *J. Med. Chem.*, 49:7169-7181, 2006.

This work was conducted as part of “Development of a Drug Discovery Informatics System” supported by Japan Agency for Medical Research and Development (AMED).

P4-04

Development of a pharmacokinetics prediction system using multiscale integrated modeling: 4. Development of a program for predicting the sites of metabolism by CYP1A2 using a two dimensional template

Junya Ohori[1]
oohori.junya@jp.fujitsu.com

Hiromi Koga[1]
koga.hiromi-01@jp.fujitsu.com

Mayumi Matsushita[1]
matusita.mayumi@jp.fujitsu.com

Kenji Mizuguchi[2]
kenji@nibiohn.go.jp

[1] Life Innovation Department, Healthcare Solutions Unit, Fujitsu Kyushu Systems Limited, 5-13 Higashihie 1-chome, Hakata-ku, Fukuoka 812-0007, Japan
[2] Laboratory of Bioinformatics, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito Asagi, Ibaraki-shi, Osaka 567-0085, Japan

Keywords: In Silico, Drug Metabolism, Regioselectivity, Cytochrome P450

In drug discovery research it is important to predict whether metabolic enzymes metabolize drugs and what metabolites would they produce. In silico methods for predicting various drug metabolisms, mainly by cytochrome P450s, have been studied, but their very limited accuracy is the main reason why they still remain only for supplementary use at the early stages of drug discovery. In order to predict the safety of drug candidates as early as possible and to accelerate the development of drugs, a prediction system that can predict metabolism with high accuracy is desired.

In this study, we developed a program to predict the metabolic site of CYP1A2 based on the method [1] of superimposing compounds to be predicted on a template representing the relevant binding and metabolic sites of the enzyme. The template was prepared by overlaying several known substrates of the metabolic enzyme and assigning additional physicochemical constraints to it. The program developed to predict regioselectivity of the enzyme only requires the structural formula of the query compound and does not require any complicated computational chemistry methods. Here we report the prediction results of the program which were validated and compared with the actual experiment results collected from literature.

[1] Yamazoe, Y. *et al.*, *Prediction of regioselectivity and preferred order of metabolisms on CYP1A2-mediated reactions. Part 1. Focusing on polycyclic arenes and the related chemicals.* Drug Metabolism and Pharmacokinetics. 31(5):363-384, 2016.

This work was conducted as part of “Development of a Drug Discovery Informatics System” supported by Japan Agency for Medical Research and Development (AMED).

Development of a pharmacokinetics prediction system using multiscale integrated modeling: 5. Prediction of sites of drug metabolism by cytochrome P450 by molecular simulation

Hiroaki Saito[1], hiroaki.saito@riken.jp,
Taku Mizukami[2], mizukami@jaist.ac.jp,
Yoshinori Hirano[1], hirano@riken.jp,
Takao Otsuka[1], totsuka.riken.jp,
Noriaki Okimoto[1], okimoto@riken.jp,
Makoto Taiji[1] taiji@riken.jp

[1] Laboratory of computational molecular design computational biology, computational biology research core, RIKEN Quantitative Biology Center (QBiC)
[2] Department of Materials Sciences, Japan Institute of Science and Technology (JAIST)

Keywords: Molecular dynamics, Molecular docking, Site of metabolism, Cytochrome P450 (CYP)

Cytochrome P450s (CYPs), a superfamily of haem-containing enzymes, are the major enzymes involved in drug metabolism. In humans, it has been estimated that CYPs metabolize approximately 75% of all marketed drugs, 95% of which are metabolized by CYP3A4, 2D6, 2C9 and 1A2 [1]. In the drug metabolism by the CYPs, identification of sites of metabolism (SOMs) on molecules and the structure of their metabolites can be decisive for the design of molecules with favorable metabolic properties. However, experimental techniques to determine SOMs and structures of metabolites are still highly resource-demanding and challenging [2]. Thus, developing fast and accurate computational methods to predict the SOMs/products of compounds metabolized by the CYPs is one of the important tasks for the optimization of ADME and toxicity properties.

In this study, we present a new computational method (score function) to predict SOMs of compounds metabolized by CYP1A2. The new score function is composed of accessibility and reactivity scores. The accessibility scores for the sites (atoms) of compounds are estimated by ensemble docking simulation, while the activation energies of atoms stored in the SMARTCyp program[3] are adopted as the reactivity scores in the method. We carried out a molecular dynamics (MD) simulation of apo type of CYP1A2 for 10 micro second to sample the receptor (pocket) structures for the ensemble docking simulation. In the molecular docking, DOCK program with Grid score, which are optimized for CYP system [4], was used. We prepared 42 test set compounds metabolized by the CYP1A2 and 100,000 receptor snapshots (which correspond to the 10 micro second MD simulation) for the ensemble docking simulation. The top three ranked atoms by the new score function are defined as possible SOMs for each compound in this study. We found the success rate of the predicted SOMs was 94 %, showing better result than those by the ensemble docking (80%) and SMARTCyp program (88%) alone.

[1] Wienkers LC, heath TG., Nature Rev. Drug. Discov. 4 (2005), 825-833.
[2] Johannes Kirchmair et al., J. Chem. Inf. Model. 52 (2012), 617−648.
[3] P. Rydberg et al., ACS Med. Chem. Lett., 1 (2010), 96-100
[4] Kiumars Shahrokh, Anita Orendt, Garold S. Yost, Thomas E. Cheatham, III, J. Comp. Chem., 33 (2012), 119-33.

This work is conducted as a part of “Construction of Drug Discovery Informatics System” supported by Japan Agency for Medical Research and Development (AMED).

Development of a pharmacokinetics prediction system using multiscale integrated modeling: 6. Assessment of quantum chemical reaction indices using Fukui function

Takao Otsuka[1]
totsuka@riken.jp

Hiroaki Saito[1]
hiroaki.saito@riken.jp

Noriaki Okimoto[1]
okimoto@riken.jp

Makoto Taiji[1]
taiji@riken.jp

[1] RIKEN Quantitative Biology Center (QBiC), QBiC Building B, 6-2-4 Furuedai, Suita, Osaka 565-0874, Japan

Keywords: Electronic structure calculation, Fukui function, chemical reaction indices, frontier molecular orbital

One of the most popular methodology to predict the potential site of chemical reaction of a molecule by electronic structure calculations is Fukui function or Fukui indices method. The Fukui indices are described as the response between the chemical potential and the external field of a molecule in the framework of density functional theory exactly. By using some approximations, the Fukui indices can be evaluated through the electron density or the molecular orbitals, especially the frontier orbitals, reflecting the electronic states such as electron adding or removing of whole molecular systems[1]. In practical quantum chemical calculations, the Fukui indices can be obtained by so-called population analysis method[2,3].

In this study we report a study on the numerical assessment of the Fukui indices derived by some types of population analysis method such as Mulliken, Löwdin, Hirschfeld and so on. We also apply the Fukui indices method to the metabolic molecular systems, and analyze and evaluate the reactivity site of metabolic reaction through the Fukui indices.

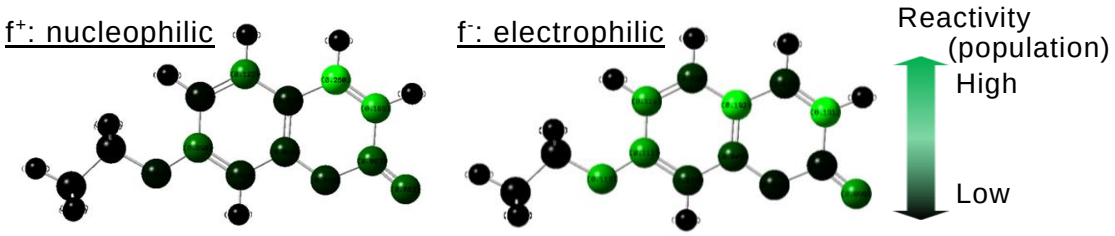


Figure. Quantum chemical reaction indices of 7-Ethoxycoumarin obtained by Fukui function.

[1] Parr, R. G., Wang, W., *Density-functional theory of atoms and molecules*, Oxford University Press, New York, Oxford, 1989.
[2] Yang, W., Mortier, W. J., *J. Am. Chem. Soc.*, 108, 5708-5711, 1986.
[3] For example, Contreras, R. R., Fuentealba, P., Galván, M., Pérez, P., *Chem. Phys. Lett.*, 304, 405-413, (1999).

This work is conducted as a part of “Construction of Drug Discovery Informatics System” supported by Japan Agency for Medical Research and Development (AMED).

P4-07

Development of a pharmacokinetics prediction system using multiscale integrated modeling: 7. Enhancement of the custom SoC LSI functionalities to improve the system performance

Gentaro Morimoto gentaro.morimoto@riken.jp
Yousuke Ohno ohno@riken.jp
Hao Zhang hao.zhang@riken.jp
Itta Ohmura ohmura@riken.jp
Keigo Nishida keigo.nishida.jg@riken.jp
Teruhisa Komatsu teruhisa.komatsu@riken.jp
Yohei M. Koyama ym.koyama@riken.jp
Aki Hasegawa ahasegawa@riken.jp
Hiroshi Koyama hkoyama@riken.jp
Makoto Taiji taiji@riken.jp

Research and Development Project of Special Purpose Supercomputers for Drug Design, RIKEN Quantitative Biology Center, 6-2-4 Furuedai, Suita, Osaka 565-0874, Japan

Keywords: Molecular dynamics simulations, Ligand binding, Protein modeling, Drug design

We have been developing a series of special purpose computers for molecular dynamics simulations. The current system is the MDGRAPE-4[1]. The whole system consists of 512 custom System on Chip(SoC) LSI chips in 3D torus network. The target performance is to make it possible to simulate typical protein-ligand complex surrounded by water for 10 microseconds per day, which is two orders of magnitude faster than commercially available systems. One of the applications of long-term MD simulations of proteins is to predict the site of metabolism.

The MDGRAPE-4 hardware development completed in 2014, and the stabilization of operation and software development process succeeded for two years. However, the final performance does not reach the expected one. In the process of the development, we realized that the classical MD simulation of the typical system size has already reached the strong scaling limit of parallel computation and not only accelerating the computationally demanding part but also more specialized hardware for the type of computation is required to improve the performance.

To achieve the target performance, we have started the development of the improved system that is named MDGRAPE-4A. The system architecture is basically common with the MDGRAPE-4. The enhancement of the functionalities are as follows:

1) support of management and migration of atoms by the memory hardware
2) more sophisticated treatment of exclusion and reduced non-bonded interaction pairs in the special purpose pipelines
3) implementation of SIMD functions and special instructions required in the MD simulations in the general-purpose cores
4) reduction of the latency and more flexible routing patterns in the network interface
5) support of a long-range electrostatic interaction calculation method that is suitable for the parallel systems with torus network topology

Currently the development process is under the design and the verification of the revised LSI. The new system hardware will be completed in 2018. In this presentation, we would like to show the ideas to improve the performance, progress and simulation results of the new LSI design.

[1] Ohmura, I., *et al.*, MDGRAPE-4: a special-purpose computer system for molecular dynamics simulations, *Phil. Trans. R. Soc. A*, 372:20130387, 2014.

Some of this work is conducted as a part of “Construction of Drug Discovery Informatics System” supported by Japan Agency for Medical Research and Development (AMED).

P4-08

Library Design With ChEMBL and Virtual Fragments Based On Electrostatic and Shape Bioisosterism Approach

Kazuyoshi Ikeda [1] ikeda-kz@pha.keio.ac.jp
Tomoki Yonezawa [2] yonezawa@level-five.jp

1 Keio University, Faculty of Pharmacy, Division of Physics for Life Functions, 1-5-30 Shibakoen, Minato-ku, Tokyo 105-8512, Japan
2 Division of Drug Discovery, Software Division, Level Five Co., Ltd., 2-25, Kanda Suda-cho, Chiyoda-ku, Tokyo 101-0041, Japan

Keywords: Field-based ligand design, Scaffold hopping, Tyrosine-protein kinase inhibitor, WaterSwap

Computational approaches play an important role for efficient ligand design. Our group is analyzing a compound library of a diverse set of compounds as well as developing a core/focused library design method by utilizing information from drug discovery data-resources such as drugs/compounds/bioactivities and the targets.

In this study, we have developed a protocol to generate a focused library using a molecular-field ligand-based method for a target. The field-based method, which is an improved method of the conventional pharmacophore model, can directly express molecule interactions of ligand to a protein target including electrostatic and hydrophobic interactions. The method can be also applied for scaffold/fragment hopping of a given reference-molecule to make the bioisosteric compound dataset [1]. Using the Pipeline Pilot [2], an automatic protocol has been developed to extract frequent rings and fragments from the ChEMBL database [3] and split based on the frequency of occurrence of the fragments in the databases. We also added the fragment data from historical/recent approved-drugs in the DrugBank [4] and a novel ring information from the VEHICLe (virtual exploratory heterocyclic library) to expand the chemical space. This protocol can filter drug-like/lead-like compounds with pharmaceutically desirable physicochemical properties.

Tyrosine-protein kinase ABL/BTK inhibitor was selected as a test-set in this study and expanded their known drugs to the diverse set of the bioisosteric compounds with the similar electrostatic patterns. Additionally, the protein structure information was used for analyzing the residue-level ligand-protein interaction energetics in the active-site and identifying the substitution point of ligand for scaffold hopping using the WaterSwap method which is recently developed by Woods et al. [implemented in the software 1.]. In this presentation, we will explain and discuss the automatic protocol of generating a forced library and the applications to the known drug targets.

[1] http://www.cresset-group.com/products
[2] http://accelrys.co.jp/products/collaborative-science/biovia-pipeline-pilot
[3] Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., Krüger, F. A., Light, Y., Mak, L., McGlinchey, S., Nowotka, M., Papadatos, G., Santos, R., and Overington, J. P., The ChEMBL bioactivity database: an update, *Nucleic Acids Research*, 42(Database issue):D1083-1090, 2014.
[4] Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A, Gabriel G, Ly C, Adamjee S, Dame ZT, Han B, Zhou Y, Wishart DS. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014 Jan 1;42(1):D1091-7.
[5] Rapid decomposition and visualisation of protein-ligand binding free energies by residue and by water.
[6] Woods CJ, Malaisree M, Michel J, Long B, McIntosh-Smith S, Mulholland AJ. Faraday Discuss. 2014;169:477-99.

P4-09

Development of an informatics system for predicting cardiotoxicity: 1. Exploring binding modes of hERG inhibitors based on docking simulation and the validation with site-directed mutagenesis data

Hitomi Yuki
hitomi.yuki@riken.jp

Tomohiro Sato
tomohiro.sato@riken.jp

Keiji Ogura
keiji.ogura@riken.jp

Teruki Honma
honma.teruki@riken.jp

RIKEN Center for Life Science Technologies, 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Keywords: hERG, Ligand docking, Pharmacophore, Site-directed mutagenesis

ADMET (absorption, distribution, metabolism, excretion and toxicity) studies are indispensable to lead efficacious and safe drug candidates. Since the 2000s, toxicity has become a major part of the cause of development failure,[1] among which the cardiotoxicity accounts for about 40% of the total toxicity.[2] hERG (human ether-à-go-go-related potassium channel) is a member of potassium ion channels and involved in cardiac repolarization. Drug-induced hERG blocking causes long QT syndrome and it is a major cause of cardiotoxicity. Therefore, drug design for avoiding the hERG inhibition is quite important through the all stages of drug discovery. Although 3D models of hERG would provide worthy compound design ideas for reducing hERG inhibition, there was no experimentally validated 3D model for several decades. This year, using emerging cryo-electron microscopy technologies, the molecular structure of hERG has been determined.[3]

We performed MD simulation of the hERG cryo-electron microscopy structure and subsequent docking simulations of known representative hERG inhibitors such as dofetilide to identify the essential pharmacophores for the hERG inhibition. hERG is a channel and shows high mobility like cytochrome P450. It makes difficult to determine the appropriate binding mode from diverse docking poses by only docking scores. Thus, we used the site-directed mutagenesis (SDM) results for the hERG pore to select and validate the binding mode. SDM reveals which residues are associated with the inhibition of hERG for each compound. We would like to discuss the analysis of these docking results and the pharmacophores of the hERG inhibition.

This work was done as a part of “Construction of Drug Discovery Informatics System” supported by Japan Agency for Medical Research and Development (AMED).

[1] Kola, I., and Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3:711–716, 2004.

[2] Laverty, H., Benson, C., Cartwright, E., Cross, M., Garland, C., Hammond, T., Holloway, C., McMahon, N., Milligan, J., Park, B., Pirmohamed, M., Pollard, C., Radford, J., Roome, N., Sager, P., Singh, S., Suter, T., Suter, W., Trafford, A., Volders, P., Wallis, R., Weaver, R., York, M., and Valentin, J. How can we improve our understanding of cardiovascular safety liabilities to develop safer medicines? *Br. J. Pharmacol.* 163:675–693, 2011

[3] Wang, W., and MacKinnon, R. Cryo-EM Structure of the Open Human Ether-à-go-go-Related K+ Channel hERG. *Cell* 169:422–430.e10, 2017

P4-10

Development of an informatics system for predicting cardiotoxicity: 2. Update of integrated database for hERG blocking small molecules considering deviation of assay results.

Tomohiro Sato tomohiro.sato@riken.jp
Hitomi Yuki hitomi.yuki@riken.jp
Keiji Ogura keiji.ogura@riken.jp
Teruki Honma honma.teruki@riken.jp

Center for Life Science Technologies, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan

Keywords: hERG, human ether-a-go-go, QT prolongation, database

The inhibition of hERG potassium channel is closely related to the prolonged QT interval [1], and to assess the risk could greatly contribute to the development of safer therapeutic compounds. To utilizing the recent increase of information about hERG inhibitors in public databases, integrated database for hERG blocking compounds was built and reported in CBI annual meeting 2016. Here, we updated the database using latest versions of ChEMBL, PubChem, GOSTAR, and hERG Central [2], and refined the merging procedure to accommodate to compounds for which contradictory assay results were found across multiple databases. Assessment of structural diversity using Murcko frameworks revealed that the integrated database contains more than 2 times as many scaffold for hERG inhibitors as any of the existing databases, and covering 18.2% of all chemical space occupied by whole compounds in ChEMBL (438,551 frameworks). (Table 1) The database could provide most comprehensive information about hERG inhibitors and be useful to predict the inhibitory activity and design safer compounds for drug discovery.

This work was done as a part of “Construction of Drug Discovery Informatics System” supported by Japan Agency for Medical Research and Development (AMED).

Table 1. Number of hERG inhibitors and inactive compounds in each database.

Database	Class	Compounds	Murcko frameworks
ChEMBL (v22)	Inhibitors	4,793	2,474
	Inactives	5,275	3,012
GOSTAR	Inhibitors	3,260	1,727
	Inactives	3,509	1,692
NCGC	Inhibitors	232	173
	Inactives	1,234	504
hERGCentral	Inhibitors	4,321	2,708
	Inactives	274,536	73,419
Integrated DB	Inhibitors	9,890	5,516
	Inactives	281,329	76,420

[1] S. Picard, S. Goineau, P. Guillaume, J. Henry, J.-L. Hanouz, R. Rouet, Supplemental studies for cardiovascular risk assessment in safety pharmacology: a critical overview, *Cardiovasc. Toxicol.*11, 285–307, 2011
[2] Du F., Yu H., Zou B., Babcock J., Long S., Li M. hERGCentral: a large database to store, retrieve, and analyze compound-human Ether-à-go-go related gene channel interactions to facilitate cardiotoxicity assessment in drug development, *Assay Drug Dev Technol.* 9(6), 580-588, 2011

P4-11

Development of an informatics system for predicting cardiotoxicity: 3. hERG prediction model based on integrated database combined with in silico molecular design for generation of safer compounds

Keiji Ogura keiji.ogura@riken.jp
Tomohiro Sato tomohiro.sato@riken.jp
Hitomi Yuki hitomi.yuki@riken.jp
Teruki Honma honma.teruki@riken.jp

RIKEN Center for Life Science Technologies, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan

Keywords: hERG, machine learning, in silico molecular design

Inhibition of the human ether a-go-go related gene product (hERG) has been identified as a major contributor to QT interval prolongation. It can cause critical cardiovascular side effects and was accounted for the withdrawal of several drugs from the market. Therefore, blockade of hERG has been extensively investigated since the early 2000s, and in silico models for hERG prediction have been continuously reported in recent years [1,2]. Despite the increasing number of hERG-related information in publicly available database [3], most of the previous researches were performed based on insufficient number or imbalanced dataset, resulting in limited the predictability and/or scope of models. In addition to the prediction of hERG blocking activity, suggestion of structural conversion for safer compounds to hERG is a quite useful tool for drug discovery process.

In this study, in silico model to predict hERG blockade was built using integrated dataset consisting of hERG information from multiple databases, and combined with fragment replacement algorithm to suggest structural conversion for safer compounds. This system enables us to not only predict hERG inhibition but also suggest synthetic candidates with drug-likeness to release the hERG issue. The validation using the test set (87,361 compounds) revealed the superior prediction performance of the constructed hERG prediction model (accuracy: 0.984, kappa: 0.732, ROC-AUC: 0.962) compared to commercially available prediction models (accuracy: 0.573-0.905, kappa: 0.0622-0.291, ROC-AUC: 0.827-0.890). Then, to demonstrate the efficiency of the combined system to suggest structural conversion, the system was applied to our in house HCK inhibitor dataset to assess whether the system can generate and predict hERG inactive compounds from a compound which inhibit hERG.

This work was done as a part of “Construction of Drug Discovery Informatics System” supported by Japan Agency for Medical Research and Development (AMED).

[1] Braga, R. C., Alves, V. M., Silva, M. F.B., Muratov, E., Fourches, D., Lião, L. M., Tropsha, A., and Andrade, C. H., Pred-hERG: A Novel web-Accessible Computational Tool for Predicting Cardiac Toxicity, *Mol. Inf.* 10:698-701, 2015.
[2] Villoutreix, B, O., Taboureau, O., Computational investigations of hERG channel blockers: New insights and current predictive models, Adv Drug Deliv Rev. 86:72-82, 2015.
[3] Wang, S., Li, Y., Xu, L., Li, D, and Hou, Y., Recent Developments in Computational Prediction of hERG Blockage, Curr. Top. Med. Chem. 13:1317-1326, 2013.

Development of an informatics system for predicting drug-induced liver injury: 1. An ontological approach for structuring knowledge of drug-induced liver toxicity

Yuki Yamagata y-yamagata@nibiohn.go.jp
Yoshinobu Igarashi y-igarashi@nibiohn.go.jp
Noriyuki Nakatsu nnakatsu@nibiohn.go.jp
Hiroshi Yamada h-yamada@nibiohn.go.jp

Toxicogenomics, Informatics Project, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Asagi Saito Ibaraki-City Osaka 567-0085, Japan

Keywords: Ontology, Drug induced liver injury, Hepatotoxicity, Toxic course, Toxic process

One of the major issues in drug development is the safety assessment of toxicity. The liver toxicity has been recognized as a primary cause of withdrawal of drug development.

In the poster, we discuss the course of the drug-induced liver toxicity. First, we introduce our visualization model of hepatotoxic course map, which shows causal relationships of cellular processes from early cellular responses to toxicological end points. Next, we propose our ontological model to structuralize the knowledge of liver toxicity. Based on upper ontology such as YAMATO [1] and BFO [2], we classified concepts involved in hepatotoxic courses as follows:

- Occurrent: toxic process (e.g., unfolded protein refolding), toxic course (e.g., ER stress)
- Continuant: molecular entity, structure (e.g., organ, cell, organelle), organism
- Dependent entity: role (e.g., transcription factor), function

In the course of realization of toxicity, toxic processes are diverse with multiple granularities. Therefore, in our model, a process can be decomposed into sub-processes represented by “hasPart” relations. For example, Unfolded protein response (UPR) hasPart three sub-processes, i.e., translation attenuation, unfolded protein refolding and unfolded protein degradation. Furthermore, sub-processes can also be decomposed for instance, “unfolded protein refolding” hasPart “Perk-Bip dissociation” at the molecular level. Based on role theory [3], we also define the toxic process by specifying the roles assigned to participants, focusing on feature genes, e.g., UPR process hasParticipant Bip/Grp78, and Bip/Grp78 hasRole “chaperone.”

This work provides high-quality data with manual curation from articles/textbooks and covers granularities of toxic processes from molecule to organ level. Hence, it will contribute to understanding knowledge about liver toxicity including bridging a gap between in vivo and in vitro experiments.

Acknowledgment
This research is supported by the Drug Discovery Support Promotion Project from Japan Agency for Medical Research and Development, AMED.

[1] Mizoguchi, R., YAMATO: Yet Another More Advanced Top-level Ontology, *Proc. of the Sixth Australasian Ontology Workshop*, 1-16, 2010.
[2] Smith, B., and Spear, A. D. *Building Ontologies with Basic Formal Ontology*, MIT Press, 2015.
[3] Mizoguchi, R., *et al.*, *Applied Ontology* 2 (2): 159-179, 2007.

P4-13

Development of an informatics system for predicting drug-induced liver injury: 2. Drug-induced liver toxicity prediction system using transcriptomics data of human primary hepatocyte

Yoshinobu Igarashi
y-igarashi@nibiohn.go.jp

Yuki Yamagata
y-yamagata@nibiohn.go.jp

Noriyuki Nakatsu
nnakatsu@nibiohn.go.jp

Hiroshi Yamada
h-yamada@nibiohn.go.jp

Toxicogenomics Informatics Project, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Asagi Saito Ibaraki-City Osaka 567-0085, Japan

Keywords: DILI, Toxicogenomics, Transcriptomics, Human primary hepatocyte

Drug-induced liver injury (DILI) is a primary safety issue in drug development. The DILI often causes failures in clinical trials as well as withdrawals from the market. To decrease the failures, it is important to characterize and predict potential mechanism of liver toxicity of a drug candidate compound in advance.

We are developing the *in silico* integrated analysis platform for DILI. This platform consists of two components. One component is the database of newly obtained toxicogenomics data of DILI compounds which are defined in FDA/Liver Toxicity Knowledge Base [1]. We also assembled other public transcriptomics data like NIH/LINCS [2] to complement the database. Another component is the DILI prediction system using multiple approaches. One of the approaches is the DILI prediction using transcriptomics data of human primary hepatocyte. We utilize transcriptomics data of human primary hepatocyte because the animal testing on rodents has limitations to predict unexpected DILI in human due to the species gap. Each DILI prediction result is shown up with a statistical probability and with current predicted location in each hepatotoxicity course map which is constructed based on the ontological model.

This DILI prediction system will assist to prioritize compounds to select further toxicology testing.

Acknowledgment
This research is supported by the Drug Discovery Support Promotion Project from Japan Agency for Medical Research and Development, AMED.

[1] Chen M, Vijay V, Shi Q, Liu Z, Fang H, Tong W. FDA-approved drug labeling for the study of drug-induced liver injury. Drug Discov Today. 2011 16(15-16):697-703

[2] http://www.biorxiv.org/content/early/2017/05/10/136168

P4-14

Development of an informatics system for predicting drug-induced liver injury: 3. Bridging for drug-induced hepatotoxicity using in vivo and in vitro gene expression data

Yang Zhao[1] y-zhao@aist.go.jp
Kazuhiko Fukui[1] k-fukui@aist.go.jp
Katsuhisa Horimoto[1] k.horimoto@aist.go.jp

[1] Molecular Profiling Research Center for Drug Discovery, National Institute for Advanced Industrial Science and Technology, 2-4-7Aomi, Koto-ku, Tokyo 135-0064 Japan

Keywords: statistical analysis, gene expression, Toxicogenomics, hepatotoxicity

The Japanese Toxicogenomics Project (TGP) [1] measured multiple toxic compounds in liver and have provided a large amount of gene expression data for rat in vivo/vitro and human in vitro. Transcriptomic understanding of in vivo/in vitro toxic mechanisms and phenotypic changes when drugs are exposed is important for risk management in drug discovery. In particular, bridging in vivo and in vitro data is a key topic to interpret potential hepatotoxicity of drug candidates [2].

In this work, we developed *toxBridge* – a statistical interpretation of building relationship between in vivo and in vitro of TGP data. Based on the rat in vivo and human in vitro gene expression data, fold change and data normalization were taken into account to analyze the significance (p-value) of gene signatures and pathways for about 160 medical compounds at four dose levels and time points, where the compounds are selected from toxicological model compounds known to possess hepatotoxicity/nephrotoxicity. These results are compiled the *toxBridge*, which provides 1) the significant signatures and pathways depending on dose and time series for each compound and 2) bridging rat in vivo and rat in vitro by using basic statistical analyses at pathway level. Indeed, at the pathway level, target compounds are likely to take feasible bridging capabilities between rat in vivo/vitro and human in vitro whose pathway the Jaccard index is up to 0.4. The results of combined probability and rank correlation analyses demonstrate that particular compounds with high (combined probability or rank correlation) values are capable to bridge rat in vivo/vitro and human in vitro.

We further proposed *toxRank* – rank analyses of drugs using rat in vivo/vitro and human in vitro gene expression data. Given a signature of query drug, *toxRank* computes the similarity with reference to expression profiles which are our own developed rank matrix data, and predicts the hepatotoxicity of the query drug.

The interpretation we give of the *toxBridge* and *toxRank* make an interesting link with TGP gene expression data that measure how many medical compounds are translatable from rat in vivo/vitro to human in vitro. Both of these two studies are expected to be useful for prediction of drug-induced hepatotoxicity.

Acknowledgment

This research is supported by the Drug Discovery Support Promotion Project from Japan Agency for Medical Research and Development, AMED.

[1] Igarashi, Y., Nakatsu, N., Yamashita, T., Ono, A., Ohno, Y., Urushidani, T. and Yamada, H., Open TG-GATEs: a large-scale toxicogenomics database, *Nucleic Acids Research*, 43:D921-D927, 2015.

[2] Hardt, C., Beber, M.E., Rasche, A., Kamburov, A. and Hebels, D.G., ToxDB: pathway-level interpretation of drug-treatment data, *The journal of biological databases and curation*, 2016.

P4-15

Development of an informatics system for predicting drug-induced liver injury: 4. QSAR prediction system for drug-induced liver toxicities using adverse drug event database

Yoshihiro Uesawa
uesawa@my-pharm.ac.jp

Department of Clinical Pharmaceutics, Meiji Pharmaceutical University
2-522-1 Noshio, Kiyose, Tokyo 204-8588, JAPAN

Keywords: DILI, nuclear receptor, stress response pathway, JADER, FAERS

Adverse drug events and toxicities are one of the most important factors in the dropout of new drugs during their development period. Currently, in the early stages of drug development, such as lead optimization step for a large number of chemicals, the in silico approach is expected to offer powerful tools for discriminating among compounds that are toxic and have adverse effects. In this presentation, the development of the QSAR prediction model for a wide range of compounds that possibly result in drug-induced liver toxicities, one of the most frequent toxic events of drugs, was mentioned. In our strategy for developing the QSAR prediction model, activities related to the adverse outcome pathway (AOP) [1], such as nuclear receptors and stress response pathways stored in Tox21-AOP database, were first targeted. Then, the predicted activities and molecular descriptors were applied to develop the prediction model for hepatic disorders in adverse drug event databases such as PMDA-JADER [2] and FDA-FAERS [3].

Acknowledgment
This research is supported by the Drug Discovery Support Promotion Project from Japan Agency for Medical Research and Development, AMED.

[1] Ankley GT., Bennett RS., Erickson RJ., Hoff DJ., Hornung MW., Johnson RD., Mount DR., Nichols JW., Russom CL., Schmieder PK., Serrrano JA., Tietge JE., Villeneuve DL., Environ. Toxicol. Chem., 29, 730-741 (2010).

[2] Pharmaceuticals and Medical Devices Agency
(http://www.pmda.go.jp/safety/info-services/drugs/adr-info/suspected-adr/0005.html)

[3] Questions and Answers on FDA's Adverse Event Reporting System (FAERS)
(https://www.fda.gov/drugs/guidancecomplianceregulatoryinformation/surveillance/adversedrugeffects/)

Enantiomeric Molecular Recognition of Barbitals in Propofol Binding Site of $GABA_A$ Receptor

Tomoyoshi Seto[1], **Ken Koyano**[2]
tseto@belle.shiga-med.ac.jp

[1]Department of Anesthesiology, Shiga University of Medical Science Hospital, [2]Faculty of Medicine, Shiga University of Medical Science, Seta-Tsukinowa-cho, Otsu 520-2192, Japan

Keywords: Barbital, Enantiomer, Propofol, Molecular Recognition, $GABA_A$ Receptor

Stereoisomers of isoflurane, barbitals, etomidate and neuro steroid have been reported to have difference in anesthetic potency, the minimum alveolar concentration (MAC) [1].

$GABA_A$ receptor ($GABA_AR$) is ligand-gated Cl^- channel, which exits post-synaptic membrane of inhibitory neurons of central nerve system, has important role in loss of consciousness in anesthesia [2]. In 2014, 3D structure of $GABA_AR$ has resolved [3]. Propofol binding site of $GABA_AR$ has recently specified [4].

We address the questions as to how enantiomeric anesthetics bind to anesthetic target site, and why there are differences between the enantiomers in their mode of binding. To clarify the mechanisms of enantiomeric molecular recognition and binding to $GABA_AR$, we investigated the binding mode and energy of amobarbital (amobar), pentobarbital (pentobar) and isobarbital (isobar) using docking simulation.

Method: 4COF (Protein Data Bank) was used as $GABA_AR$ structure [3]. Amobar, *(R)*-, *(S)*-pentobarbital (pentobar) and *(R)*-, *(S)*-isobarbital (isobar) bindings were studied. ASEDock (2016.01.07) (Ryoka system, Japan) [5] were used for the dockings and energy calculations on MOE2016.0802 [6]. Ligand's chirality was kept in flexible ligand docking. Amber10:ETH force field was used. MOE-ligand Interactions was used to find interactions.

Results and Discussions: (1) amobar bound to (a) M2-M2' site (propofol high affinity binding site [4]), (b) agonist binding site, (c) extracellular domain inter-subunit site and (d) M2-M3-M1' site (propofol low affinity binding site [4]) of $GABA_AR$. (2) NH of barbital ring of amobar hydrogen-bonded to Gln224 of M2-M2' site in the same as propofol binding. (3) *(R)*-, *(S)*-pentobar and isobar bound to M2-M2' site. N-H of barbital ring hydrogen-bonded to Glu270 of M2-M2' site. (4) Of pentobar's enantiomer methyl group adjacent to chiral carbon directed to opposite way each other. Steric hindrance of this methyl group determines enantiomer's fitting. (5) Barbital ring binding is presumed to be the major binding interaction of either enantiomer. The principal binding force from the ring is modulated by methyl group hindrance. The steric hindrance produced -2.9 kcal mol^{-1} molecular discrimination of binding energy.

Conclusions: Amobar, pentobar and isobar bound to propofol high affinity site [4]. The molecular discrimination of enantiomer turns out to be from modulation of major binding force by methyl-group hindrance of enantiomer.

[1] Burke, D. and Henderson, D.J., Chirality: a blueprint for the future, *Brit J Anaesth* 88:563-576, 2002.

[2] Garcia, P.S., Kolesky, S.E. and Jenkins, A., General Anesthetic Actions on $GABA_A$ Receptors, *Curr Neuropharm*, 8:2-9, 2010.

[3] Miller, P.S. and Radu Aricescu, A., Crystal Structure of a human $GABA_A$ receptor, *Nature*, 512:270-275, 2014.

[4] Franks, N.P., Structural Comparison of Ligand-gated Ion Channels in Open, Closed, and Desensitized States Identify a Novel Propofol-binding Site on Mammalian γ-Aminobutyric Acid Type A Receptors, *Anesthesiology*, 122:787-794, 2015.

[5] Goto, A., Kataoka R., *et al.*, ASEDock-docking based on alpha spheres and excluded volumes, *J Chem Inf Model*, 48:583-590, 2008.

[6] Molecular Operating Environment (MOE) 2016.08.02, Chemical Computing Group ULC, 2016.

P4-17

Genome-wide analysis of DNA methylation and gene expression patterns in human iPS cell-derived hepatocytes

Seiichi Ishida[1] **Su-Ryang Kim**[1] **Shinichiro Horiuchi**[1] **Yukie Kuroda**[1]
ishida@nihs.go.jp

Shoko Uchida[1] **Jane Synnergren**[2] **Yasunari Kanda**[1]

[1] Division of Pharmacology, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku Tokyo 1588501, Japan
[2] Systems Biology Research Center, School of Bioscience, University of Skövde Högskolevägen, Box 408, 541 28 Skövde, Sweden

Keywords: iPSC, Hepatocyte, differentiation, Gene expression, Genomic DNA Methylation

[Purpose] Human hepatocytes are frequently used for in vitro drug metabolism and toxicity tests for drug discovery and development. However, since supply of human hepatocytes is limited, construction of an alternative test system is awaited. Human induced pluripotent stem cells (hiPSCs) are useful because of its proliferative properties and differentiation capabilities into various cell types, so they are expected as a source of hepatocytes for in vitro testing. Comprehensive gene expression analysis and genomic DNA methylation analysis were performed on commercially available hiPSC-derived hepatocytes, and the methylation status of genomic DNA of hiPSC-derived hepatocytes from three vendors was compared with that of human primary hepatocyte cells.
[Materials and Methods] Three commercially available hiPSC-derived hepatocytes were seeded and cultured according to the recommended protocol. For one of hiPSC-derived hepatocytes (Vendor A), the changes of gene expression and genomic DNA methylation status during differentiation from hiPSCs to hepatocytes were analyzed. Human primary hepatocytes derived from three donors were used as controls. DNA and RNA were prepared from each cell culture. Genome-wide gene expressions were measured by GeneChip Human Gene 1.0 ST Array (Affymetrix, Santa Clara, CA, USA), and Genome DNA methylations were measured by HumanMethylation450 BeadChip (Illumina, San Diego, CA, USA), according to the manufacturer's protocol. GeneSpring GX 12.0 (Agilent) was used for data analysis.
[Results and Discussion] Genes whose expression increased and methylation rate decreased during differentiation of hiPSCs to hepatocytes were selected and 3,280 genes were obtained. Hepatocytes-specific genes such as metabolic enzymes, transporters and nuclear receptors were contained abundantly among them, while genes which show specific expression in other tissues were also found. Comparison of the methylation status of hiPSC-derived hepatocytes from three vendors and primary hepatocytes showed similar pattern, whereas the genes whose methylation status was different were also observed.
[Conclusion] The results suggest that the analysis of methylation status of genomic DNA is useful as an indicator of functional evaluation for hiPSC-derived hepatocytes.

Integrated analysis for drug toxicities in human using multi-label classification.

Koki Matsuoka[1] c401631029@tokushima-u.ac.jp
A. AMMAR GHAIBEH[2] ammargh@acm.org
Shiro Omura[1] c401203002@tokushima-u.ac.jp
Youichi Sato[1] youichi.sato@tokushima-u.ac.jp
Hiroki Moriguchi[2] h_moriguchi@ap6.mopera.ne.jp
Aiko Yamauchi[1] yamauchi.aiko@tokushima-u.ac.jp

[1] Department of Pharmaceutical Information Science, Graduate School of Biomedical Sciences, Tokushima University 1-78-1, Sho-machi, Tokushima 770-8505, Japan
[2] Department of Medical Informatics, Graduate School of Biomedical Sciences, Tokushima University 3-18-15, Kuramoto-cho, Tokushima 770-8503, Japan

Keywords: *In silico* prediction, multi-label classification, Decision tree, Random forest, SVM, carcinogenicity, teratogenicity, nephrotoxicity, hepatotoxicity

Introduction: In recent years, many studies on drug toxicity prediction using *in silico* methods have been reported. Many of them focus on a single toxicity such as carcinogenicity. However, the developed models may not be practical because drugs often have multiple toxicities. Therefore, we developed models that can predict carcinogenicity, teratogenicity, nephrotoxicity and hepatotoxicity of drugs in humans at once, using Support Vector Machine (SVM), decision tree (gini) and random forest.

Methods: We collected 139 drug data that included information regarding carcinogenicity, teratogenicity, nephrotoxicity and hepatotoxicity in human from International Agency for Research on Cancer (IARC), Drugs in Pregnancy and Lactation 8th and Japanese prescription drug information database. The *in silico* models were built on the dataset of drugs with 43 descriptors. We evaluated the model performance with Hamming score, Hamming loss and Accuracy using holdout validation (train : test = 9 : 1) in Python3.

Results and discussion: We used 'One vs All' approach to deal with the multi-label classification problem. In addition, the methods with classification algorisms corresponding to multi-labels were used in the present study. The best performances of SVM, decision tree and random forest were shown in Table.1.

Table.1 The performance based on hamming score, hamming loss and accuracy.

Classification	hamming score	hamming loss	accuracy
SVM (rbf)	0.81	0.11	0.64
Decision tree (gini)	0.69	0.16	0.43
Random forest	0.67	0.20	0.29

Conclusion: The model developed was shown to be useful for not only predicting the toxicity of new chemical entities in advance but also enabling safer pharmacotherapy.

In silico models for predicting repeated dose toxicity using machine learning

Tatsuya Ochibe c142513@ed.nagoya-cu.ac.jp
Kaori Ambe ambek@phar.nagoya-cu.ac.jp
Masahiro Tohkin tohkin@phar.nagoya-cu.ac.jp

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

Keywords: Applicability domain, Hepatocellular hypertrophy, Machine learning method

It is expected to develop the alternative method for animal experiments to reduce the number of experimental animals and improve efficiency of the risk assessment of environmental chemical substances. Machine learning *in silico* evaluation method for the repeated dose toxicity (RDT) is considered to be more useful because RDT has complex and unknown mechanisms. In this study, using machine learning methods, we tried building *in silico* prediction models of chemical-induced hepatocellular hypertrophy which is often observed in RDT studies based on molecular descriptors [1]. Our original data sets which used to build the models were created from two *in vivo* toxicological databases. The first was the risk assessment reports of pesticides, food additives, and veterinary medicine products that were published by Food Safety Commission of Japan [2], and the second was the Hazard Evaluation Support System Integrated Platform (HESS) [3] by National Institute of Technology and Evaluation. Then, we developed prediction models by support vector machine and random forest using descriptors calculated by dragon 6 [4]. Because the prediction results of chemicals inside applicability domain (AD) are high reliability [5], we defined AD by the data distance [6]. For both our models, prediction accuracy of the test set inside the AD was above 74% and AUC was about 0.8. These results suggested that our *in silico* models for hepatocellular hypertrophy are reliable and promising method for RDT prediction.

[1] A. P. Hall., *et al.*, Liver Hypertrophy: A Review of Adaptive (Adverse and Non-adverse) Changes, *Toxicologic Pathology*, 40: 971-994, 2012.
[2] http://www.fsc.go.jp/english/index.html
[3] http://www.nite.go.jp/en/chem/qsar/hess-e.html
[4] http://talete.mi.it/index.htm
[5] OECD, ENV/JM/MONO, 2, 2007.
[6] I. V. Tetko., *et al.*, Critical Assessment of QSAR Models of Environmental Toxicity against Tetrahymena pyriformis: Focusing on Applicability Domain and Overfitting by Variable Selection, *J. Chem. Inf. Model.* 48 (9): 1733-1746, 2008.

A prediction model for photolytic half-lives of chemicals

Yumi MATSUYAMA[1] u121719h@gmail.com
Norihito KAWASHITA[1, 2] nkawashita@life.kindai.ac.jp
Yu-Shi TIAN[1] yushi-tian@phs.osaka-u.ac.jp
Kousuke OKAMOTO[3] kokamoto@gen-info.osaka-u.ac.jp
Tatsuya TAKAGI[1] ttakagi@phs.osaka-u.ac.jp

[1] Graduate School of Pharmaceutical Science, Osaka University, 1-6 Yamadaoka, Suita, Osaka 562-0875, Japan
[2] Faculty of Sciences and Engineering, Kindai University, 3-4-1 Kowakae, Higashiosaka, Osaka 577-8502, Japan
[3] Faculty of Pharmaceutical Science, Hokuriku University, 1-1 Taiyogaoka. Kanazawa, 920-1154, Japan

Keywords: Statistical analysis, Quantitative Structure-Property Relationships, Evaluation of environmental pollution, Photolysis

It is an important issue to understand the physical properties and dynamics of chemicals in the environment. However, experimental surveys on the environmental fates of all compounds which have been produced is almost impossible since it takes huge amounts of cost and time. Hence, the QSPR (Quantitative Structure- Property Relationship) predictions of environmental fates are expected as an alternative approach. The QSPR models of photolysis have not been sufficiently developed yet and most models were developed only for some analogous compounds [1]. The purpose of current study is to construct a prediction model for the photolytic half-lives and to estimate some physicochemical properties which are significantly effective for the prediction.

We used Random Forests [2], a commonly used machine learning algorithm for constructing the model. Descriptors were calculated by MOE [3]. To remove the effects caused by the random split of training and test sets, models were constructed multiple times. As a result, energy differences between the HOMO and LUMO, charge-related descriptors, and some other descriptors were selected to be significant. The average accuracies of the model were approximately 73.1% and 72.9% for training and test sets, respectively. Some compounds were always correctly classified, while the others were sometimes not. This result indicated that this model was almost perfect for predicting the photolytic half-lives of the former compounds, whereas this was not suitable for the latter ones.

Therefore, we further classified the compounds into two groups (correctly predicted compounds and the others) and constructed distinct models by the two groups. In the result, although some compounds were classified almost incorrectly (the average accuracy was approximately 55.1% for test sets), accuracy improved for other compounds (the average accuracy was approximately 81.2% for test sets). Some compounds which we could not classified previous model became predictable. This result indicates that dividing dataset is effective to predict photolytic half-lives using the Random Forests method.

[1] Nolte, T.M. and Ragas, A.M. A review of quantitative structure-property relationships for the fate of ionizable organic chemicals in water matrices and identification of knowledge gaps. *Environmental science. Processes & impacts.* 19:221–246, 2017
[2] Breiman, L. Random forests. *Machine learning.* 45:5–32, 2001
[3] *Molecular Operating Environment (MOE)*, 2013.08; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2017.

Computational toxicology of cardiomyocyte gel image

Kei Kuramoto[1] kei.kuramoto@daikin.co.jp
Hiroshi Matsumoto[1] hiroshi4.matsumoto@daikin.co.jp
Tatsuya Takakuwa[1] tatsuya.takakuwa@daikin.co.jp
Shigehito Sagisaka[1] shigehito.sagisaka@daikin.co.jp
Satoshi Tokuno[1] satoshi.tokuno@daikin.co.jp
Kaoru Uesugi[2] uesugi@live.mech.eng.osaka-u.ac.jp
Keisuke Morishima[2] morishima@mech.eng.osaka-u.ac.jp

[1] Technology and Innovation Center, Daikin Industries, Ltd.
1-1, Nishi-Hitotsuya, Settsu, Osaka 566-8585, Japan
[2] Department of Mechanical engineering and Graduate School of Engineering, Osaka University
2-1, Yamada-oka, Suita, Osaka 565-0871, Japan

Keywords: computational toxicology, cardiomyocyte gel, image processing

Alternatives to animal experiments have been official policies all over the world and international standards on the welfare of animals used experiments. In this study, we performed image processing and machine learning with cuda GPGPU [1] and phase-based image processing [2] to analyze the motion and dynamics of the correlation between toxic material and cardiomyocyte gel [3]. The kinetics of myocardial cell gels has periodicity and as each moves synchronously or asynchronously, we found that region division image processing was necessary. And we also found that kinetic analysis could be performed with high accuracy by using image matching by phase-only correlation method.

[1] http://www.nvidia.co.jp/object/cuda-parallel-computing-platform-jp.html
[2] Kuglin, C. D. and Hines, D.C., The phase correlation image alignment method, *Proceedings International Conference on Cybernetics and Society*, pp. 163-165, 1975.
[3] Hoshino, T., Imagawa, and K., Morishima, K., Cardiomyocyte-Driven Wet Gel Robotics - Chemical Modulation of Cardiac Network Pattern Generator-, *The 2010 3rd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob 2010)*, IEEE Robotics and Automation Society & IEEE Engineering in Medicine and Biology Society, 547-552, 2010.

A Pneumatic Pressure-Driven Multi-Throughput Organs-On-A-Chip

Taku Satoh[1] taku.satou@aist.go.jp
Shinji Sugiura[1] shinji.sugiura@aist.go.jp
Kazumi Shin[1] shin-kazumi@aist.go.jp
Toshiyuki Kanamori[1] t.kanamori@aist.go.jp

[1] Biotechnology Research Institute for Drug Discovery, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, 305-8565, Japan

Keywords: Microfluidic device, Organs-on-a-chip, Drug discovery, Cell-based assay

Recently, multi-organs- or body-on-a-chip have been attracting attention as next-generation research tools in drug discovery. In these system, multiple types of cells were cultured in discrete chambers connected to each other via microchannels for medium circulation [1, 2]. In this study, we demonstrate a organs-on-a-chip system with parallelized medium circulation by using the pressure-driven medium circulation system [3].

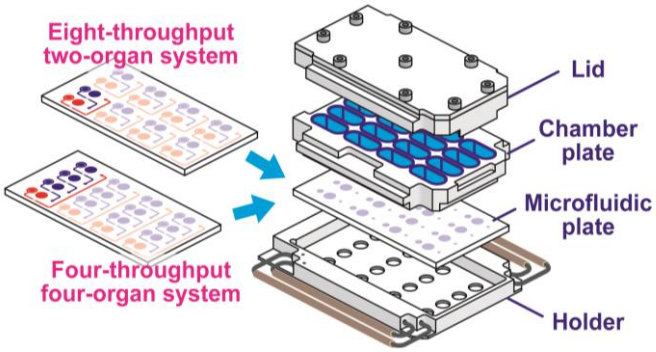


Figure 1: View of and components of the pneumatic pressure-driven multi-throughput multi-organs-on-a-chip.

The culture device in the present system is composed of a chamber plate, microfluidic plate, lid, and holder (Figure 1). The culture unit is composed of two culture chambers, which are connected each other via microchannels in the microfluidic plate. The sequentially applying pressure to the chambers induces unidirectional circulation of medium by a similar manner as described previously [3].

Medium circulation was successfully accomplished by sequential applied pressure. The eight set of the two-organ system in the single device can be simultaneously driven by the pneumatic pressure supplied from the external pressure control system via four pneumatic pressure lines in the lid. We have also successfully applied the eight-throughput two-organ system to the liver-cancer connection system to evaluate anticancer prodrug.

[1] I. Maschmeyer et al., “A four-organ-chip for interconnected long-term co-culture of human intestine, liver, skin and kidney equivalents,” Lab Chip, 15, 2688-2699, 2015.
[2] C. Oleaga et al., “Multi-Organ toxicity demonstration in a functional human in vitro system composed of four organs,” Sci. Rep. 6, 20030, 2016.
[3] T. Satoh et al., “Pneumatic pressure-driven circulation cell culture microfluidic device for investigation of vascular endothelial cells under shear stress,” Lab Chip, 16, 2339-2348, 2016.

P4-23

On/off targets based toxicity prediction analysis and Verification

Kosuke Negishi[1] k.negishi@elsevier.com

Yasunori Okazaki[1] y.okazaki@elsevier.com

1 Life Science,Elsevier Japan K.K,1-9-15 Higashi-Azabu Minato-ku, Tokyo 106-0044 Japan

Keywords: Toxicology analysis, Medicinal chemistry, On/off target discovery, data integration

SAR(Structure-activity relationship) is well known as one of major evaluation method for drug discovery and toxicity/biological activity prediction. On the other hand, protein"s" (on/off target"s") based analysis is not major yet.
In this report, we tried to evaluate the feasibility of the protein"s" based analysis. At first we focused on drugs which have few side effects and chose phenylalanine derivative of oral hypoglycemic agent. This type of drug do not have clinical adverse event besides low blood sugar. We created on/off target"s" list from 2 kind of databases (affinity/bioassay type and Text mining type). By Using the proteins listed in these databases, we calculated candidates of toxicity phenotype. Moreover, we searched FDA/EMA documents (database) to verify the consistency of these predicted toxicities.

[1] Hansch C., Smith N., Engle R., Wood H. *Quantitative structure activity relationships of antineoplastic drugs: Nitrosoureas and triazenoimidazoles* Cancer chemotherapy reports 1972 56:4 (443-456)

[2] Miller EC, Miller JA. *Searches for ultimate chemical carcinogens and their reactions with cellular macromolecules.* Cancer. 1981 May 15;47(10):2327-45.

[3]Hanefeld M, Bouter KP, Dickinson S, Guitard C. *Rapid and short-acting mealtime insulin secretion with nateglinide controls both prandial and mean glycemia.* Diabetes Care. 2000 Feb;23(2):202-7.

[4] Kawamori R, Kaku K,Hanafusa T , Kashiwabara D, Kageyama S, and Hotta N *Efficacy and safety of repaglinide vs nateglinide for treatment of Japanese patients with type 2 diabetes mellitus* J Diabetes Investig. 2012 Jun 6; 3(3): 302–308.

P5-01

Thermostability Prediction Model for Protein Single Point Mutations Using Machine Learning

Chie Imamura
chiem@mosk.tytlabs.co.jp

Keisuke Kawano
kawano@mosk.tytlabs.co.jp

Satoshi Koide
koide@mosk.tytlabs.co.jp

Yukihiro Tadokoro
tadokoro@mosk.tytlabs.co.jp

Toyota Central R&D Labs. Inc.,
41-1 Yokomichi, Ngakute, Aichi 480-1192, Japan

Keywords: Prediction of thermostability, Machine learning, SVM, Mutant protein, Protein structure

The thermostability of proteins has been improved by high-throughput experimental screenings from randomly mutated proteins so far. However, it is inefficient because it takes a lot of work and time to select a target. Therefore it is necessary to develop high-performance computational methods for predicting thermostability upon protein mutations.

In this study, we take an alternative approach to improve protein thermostability using machine learning. At first, we constructed 2,000 mutants which have single point mutation at various sites of two proteins. Then we obtained thermostability data by experimental evaluation. Also, we proposed some features based on relevant information about protein local structure around the mutation site.

Using the experimental data and those features as input data, we constructed a thermostability prediction model using support vector machine. Prediction accuracy of the model was evaluated by cross-validation. In the case which training and test data were selected from the same protein, the accuracy of predictions and the area under the curve (AUC) showed high values. On the other hand, when training and test data from different proteins were used, AUC values were decreased. Our results suggest that not only the features from local structure around the mutation site, but features from whole structure are important to improve prediction accuracy.

P5-02

Complex Network Approach for Characterization of Protein Secondary Structure

Shohei Konno[1] skon@sci.hokudai.ac.jp
Takao Namiki[2] nami@math.sci.hokudai.ac.jp
Koichiro Ishimori[1, 2] koichiro@sci.hokudai.ac.jp

[1] Graduate School of Chemical Sciences and Engineering, Hokkaido University, North-13, West-8, Kita-ku, Sapporo, 060-8628, Japan
[2] Faculty of Science, Hokkaido University, North-10, West-8, Kita-ku, Sapporo, 060-0810, Japan

Keywords: Protein secondary structure, Complex networks, Amino acid network

One difficulty in protein structure prediction is to construct appropriate coarse-grained models. An emerging approach is amino acid network (AAN), where the "vertices" represent amino acid residues in the protein and the "links" correspond to interactions between the residues. In major AAN studies, the links have been defined by distance between α carbons to provide topological information with low computational complexity. However, these networks often lose information about side chains, which remain unsuccessful to categorize protein secondary structures [1]. We successfully distinguished between all-α and all-β proteins by using a new definition of AAN, whose links are determined by different cutoff values depending on the chemical interactions [2][3].

In order to characterize the protein secondary structures by our AAN approach, we collected high-resolution all-α and all-β protein structures from Protein Data Bank [4]. To capture the types of interactions, we have constructed a network with different cutoff values for hydrophobic interactions, hydrogen bonds, ion bonds, disulfide bonds and covalent bonds as defined in [3]. The vertices are linked if any two atoms from two different residues are located within the cutoff distance. We compared our method with conventional ones in term of the ability of the secondary structure discrimination and calculated three network parameters: the average degree (k), average clustering coefficient (C) and average distance (L). k and C reflect the local structures, while the global structure affects L. As depicted in Figure A, all-α proteins are well separated from all-β proteins in the scatter plots of k versus C. The figure clearly shows that our method can distinguish between the all-α and all-β protein structures, which was unsuccessful in the conventional method (Figure B). On the other hand, it was difficult to separate the secondary structures by using L. The parameters describing the local structure (k and C) are more sensitive to the difference of the secondary structures than that describing the global structure (L). Such successful discrimination suggests that AAN we defined here can be an effective coarse-grained model for characterizing the protein structures.

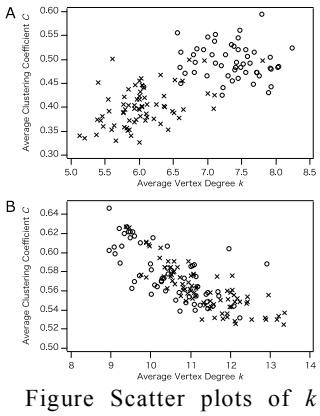


Figure Scatter plots of k versus C for all-α (○) and all-β (×) proteins by our introduced network (A) and previously used α carbon network (B).

[1] Alves, N. A. and Martinez, A. S., *Physica A*, 375:336-344, 2007
[2] Tina, K. G., Bhadra, R., Srinivasan, N., *Nucleic Acids Research*, 35: W473-W476, 2007
[3] http://pic.mbu.iisc.ernet.in
[4] http://www.rcsb.org

P5-03

Reconstruction of self-assembly structure of microtubules

Ryuzo Azuma
azuma.r.ac@m.titech.ac.jp

Gregory Spence Gutmann
ggutmann13@jcu.edu

Arif Pramudwiatmoko
pramudwiatmoko.a.aa@m.titech.ac.jp

† **Akihiko Konagaya**
kona@c.titech.ac.jp

Tokyo Institute of Technology, J3-25, 4259, Nagatsuta, Midori-ku, Yokohama, Kanagawa 226-8503, Japan
† Corresponding author

Keywords: Molecular robotics, artificial intelligence related technology

We have been developing a coarse-grained simulation method for analyses of super molecular structure formation processes. In this poster presentation, we demonstrate self-organized structures in terms of studies on mechanisms of microtubule structure formation. The halmark of our simulation is that it explicitly takes account of water molecules albeit it is based on a mathematical model employing purely stochastic rules for semi-microscopic molecular dynamics in a 3 dimensional space. Water molecules turned out to have versatile roles in governing electrostatic and hydrophobic interactions during the microtuble organization process.

In our microtuble simulation model, each tubulin monomer (α and β tubulins) is represented by a chain of coarse-grained subunits. Strength of associative interactions between those subunits are roughly double that of thermal fluctuations. For inter-monomer coupling, we assume ~0.2eV strength by taking account of electrostatics calculation in a previous study [1].

Using these simulation method and model, we have been successfully observing self-organized structure that associates well with microtuble triplet structure. Transient dynamics in those simulations reaches a plateau after ~20 milliseconds and thus the structure is stabilized. Diameter of each one in the triplet is close to that of 13 protofilament microtubles (23-4 nm.) These results support our view that coexistence of hydrophobic and electrostatic interaction between tubulin dimer drives self-organization of microtubules in presence of water molecules.

We have also been focusing on atomic-level tubulin α-β dimer structure responsible for interactions during microtubule assembly formation. In particular, surface charge and molecular dynamics are studied for scenarios where one of/both c-termini are detached from the surface, as proposed previously [2]. These features are going to be utilized to refine our coarse-grained simulation model.

This presentation is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO.)

[1] Tuszynsky, J.A., Brown, A., Crawford, E.J., Carpenter, E.J., *Molecular Dynamics Simulation of Tubulin Structure and Calculations of Electrostatic Properties of Microtubules*, Mathematical and Computer Modeling, 41:1055-1070, 2005.

[2] Priel, A., Tuszynski, J.A., Woolf, N.J., Transitions in microtubule C-termini confirmations as a possible dendritic signaling phenomenon, *Eur Biophys J*, 35:40-52, 2005.

P5-04

On Being Scientific

Tomokazu Konishi[1]
konishi@akita-pu.ac.jp

[1] Graduate School of Bioresource Sciences, Akita Prefectural University, 010-0195, Japan

Keywords: falsifiability, falsification

Our researches often make a crowded lobby; for instance, I wonder how many procedures have been introduced to normalize transcriptomics data. Most of them are totally out of use, because they lack falsifiability, which is an idea introduced by Karl Popper as a touchstone for science (1). As science has no rigid definition, distinguishing true science and others is controversial; however, this idea has been widely accepted among scientists. Science hates a situation that observation depends on analytical methodologies, as it is formed by integrating tons of observations reported by many researchers. So, most scientists do not take ideas like “anything goes” (2, 3). As far as I know, there has been found only one method that can analyze transcriptomics data fulfilling this issue (4, 5). There would not be another solution, as it is compatible to thermodynamic property of a living cell (6).

Another enthusiasm occurred to solve multiplicities of statistical tests (7). This claim encompassed scientific community a mess, and many procedures were introduced. However, the casted problem may not really exist; it could be understood as a stray bullet from war between schools of statisticians (8).

Rushing into a wrong paradigm is waste of time, not only for analysts, but also for scientists. How can we avoid such researchers’ bad habit? I believe we can learn from science. Presently, we introduce a new method by comparing it to existing ones by focusing on properties defined by ourselves. I rarely saw articles that defined problems clearly, and estimated the damages caused by the problem. Also, I rarely saw articles that declare what was assumed, certainty of the assumptions, and how they could be verified. Those should be clarified. I know it is easy to follow someone’s idea and put some new decorations; however, this is not very scientific attitude. Additionally, changing data to fulfil benefits of the analyst is prohibited in science, because it could conflict to ethical rule; it would be recognized as falsification of data.

(1) Popper K. The logic of scientific discovery, Routledge; 1959.
(2) Feyerabend P. Against method: outline of an anarchist theory of knowledge, New Left Books; 1975.
(3) Sober E. Philosophy of Biology, Oxford University Press, 1993.
(4) Konishi T, Konishi F, Takasaki S, Inoue K, Nakayama K, Konagaya A. Coincidence between transcriptome analyses on different microarray platforms using a parametric framework. *PLoS ONE*. 2008;3(10):e3555.
(5) Konishi T. Parametric analysis of RNA-seq expression data. Genes to Cells. 2016;21(6):639-47.
(6) Konishi T. A thermodynamic model of transcriptome formation. *Nucleic Acids Res.* 2005;33(20):6587-92.
(7) Dudoit S, Yang Y, Callow M, Speed T. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Stat Sin.* 2002;12:111–39.
(8) Konishi T. Microarray test results should not be compensated for multiplicity of gene contents. *BMC Systems Biology.* 2011;5(Suppl 2):S6-S.

P5-05

Using convolutional neural networks plus recurrent neural network, automatic classification of human iPS cell-derived cardiomyocytes extracellular potential waveforms.

Tetsuo Kitamura, Mayumi Obo,
kitamura.tetsuo@mh.medience.co.jp obo.mayumi@ms.medience.co.jp

Tomoharu Osada, Hiroaki Inoue,
osada.tomoharu@mg.medience.co.jp inoue.hiroaki@mb.medience.co.jp

Yamamoto Yoshinori, Yasuyuki Oonishi,
yamamoto.yoshinori@mg.medience.co.jp oonishi.yasuyuki@mp.medience.co.jp

Hideaki Hiratsuka
hiratsuka.hideaki@mg.medience.co.jp

Nonclinical Research Center, Drug Development Service Segment, LSI Medience Corp. 14-1 Sunayama, Kamisu, Ibaraki, 314-0255, Japan

Keywords: convolutional neural networks, recurrent neural network, human iPS cell-derived cardiomyocytes

[Introduction] Cardiac safety evaluation method which uses an extracellular potential of human iPS stem cell-derived cardiomyocytes is proposed. The evaluated system is expected to predict pro-arrhythmic potentials of compounds with high accuracy by detection of early after depolarizations waveforms. This detection is done by human, so there is a problem in the labor and precision. Automatic classification using convolutional neural networks plus recurrent neural networks (CNN+RNN) are reported for movie classification technique. By the machine learning with training data, CNN+RNN are possible to highly accurate classification. The purpose of this study was to evaluate the accuracy of automatic classification with CNN+RNN for the waveforms obtained in this evaluation system.
[Methods] Classification targets were waveforms of sequential beats from multi channels. Framework of neural networks was TensorFlow. Supervised learning was used. Training data and test data were selected from same experiments. These data were labeled with classification by human.
[Results] With the test data, concordance rate of classification result by human and CNN+RNN was 99.97%.
[Conclusion] CNN+RNN showed high accuracy in cardiomyocytes extracellular potential waveforms classification.

P5-06

Rational design of orthogonal gene transcription nano device on DNA origami

Takeya Masubuchi [1], Masayuki Endo [2], Ryo Iizuka [3], Ayaka Iguchi [4], Dong Hyun Yoon [4], Tetsushi Sekiguchi [5], Hao Qi [1,6], Ryosuke Iinuma [1], Yuya Miyazono [1], Shuichi Shoji [4], Takashi Funatsu [3], Hiroshi Sugiyama [2,7], Yoshie Harada [2,8], Takuya Ueda [1] and **Hisashi Tadakuma** [1,2,8,]
tadakuma@protein.osaka-u.ac.jp

[1] Graduate School of Frontier Science, The University of Tokyo, [2] Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, [3] Graduate School of Pharmaceutical Sciences, The University of Tokyo, [4] Department of Nanoscience and Nanoengineering (ASE Graduate School), Waseda University, [5] Research Organization for Nano & Life Innovation, Waseda University, [6] Department of Chemical Engineering and Technology, Tianjin University, [7] Department of Chemistry, Kyoto University, [8] Institute for Protein Research, Osaka University

Keywords: Gene expression, Synthetic Biology, Molecular Robots

In synthetic biology, the design of gene expression requires devices that alter the output depending on the situation. Inspired by RNA viruses that contain both enzymes and substrate genes, and express specific genes through architectural modalities, we integrated an enzyme, T7 RNA polymerase, and multiple target gene substrates onto a DNA origami-based nano-chip. This gene nano-chip orthogonally transcribes its own genes, and using this system, we succeeded in a rational design of gene expression by controlling the inter-molecular distances between the enzyme and the substrate genes. Our approach of component integration on a nano-chips may provide the basis for an integrated gene expression circuit.

[1] Tadakuma H, Masubuchi T, Ueda T. *RNA Study Using DNA Nanotechnology. (review)*, Prog Mol Biol Transl Sci., 139:121-163 (2016).

P5-07

Graph Neural Networks for Density Functional Theory

Masashi Tsubaki[1]
tsubaki.masashi@go.aist.jp

[1] National Institute of Advanced Industrial Science and Technology (AIST)
Artificial Intelligence Research Center, Machine Learning Research Team
Tokyo Waterfront BIO-IT Research Building, 2-4-7 Aomi, Koto-ku, Tokyo, 135-0064, Japan

Keywords: Graph Neural Networks, End-to-End Learning, Density Functional Theory

In this poster, we propose the use of machine learning technique, graph neural networks (GNNs), for predicting various electronic ground-state properties molecules based on density functional (DFT) theory. GNN is a neural network, which takes a graph-structured data (i.e., molecular graph) as an input and outputs a low-dimensional real-valued vector representation (i.e., molecular embedding). In addition, our GNN uses an end-to-end learning technique to obtain the molecular embeddings, which does not require any molecular features (e.g., descriptors and fingerprints) and knowledge of chemistry and physics. The end-to-end GNN allows us to automatically learn all parameters including input features in the model. We experiment with thirteen electronic ground-state properties of 117k distinct organic molecules and show that our end-to-end GNN can be more accurate than hybrid DFT.

ChemiSpadon: an SNS-based platform for human-assisted chemical space exploration

Kazuki Z Yamamoto[1] kayamamoto-tky@umin.net
Ryuichi Kubo[2] ryuichi.kubo@dena.com
Hiroyuki Yamasaki[3] yamasakih@pharm.kitasato-u.ac.jp
Masahito Ohue[4] ohue@c.titech.ac.jp
Yusuke Yamada[5] yamayu0127@gmail.com

[1] Isotope Science Center, University of Tokyo, 2-11-16, Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan
[2] DeNA Life Science, Inc., Shibuya Hikarie, 2-21-1 Shibuya, Shibuya-ku, Tokyo 150-8510 Japan
[3] Kitasato University School of Pharmacy, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641 Japan
[4] School of Computing, Tokyo Institute of Technology, 2-12-1, Ookayama, Meguro-ku, Tokyo 152-8550, Japan
[5] Mishima.syk, Mishima-PostOffice-Kyokudome, Minami-Tamachi, Mishima-shi, Shizuoka 411-8799, Japan

Keywords: Chemical space, Drug design, Human computation, SNS, Mastodon

Exploration of undeveloped chemical space is essential for discovery of novel drugs. Conventionally, 'druglikeness' and 'druggablity' have been mainly considered for small molecules whose molecular weights are less than 500-600[1, 2]. However, to cope with novel drug targets previously considered 'undruggable' due to absence of druggable cavities, medium-sized molecules like natural products, macrocycles, peptidomimetics or others are needed[3]. Although medium-sized molecules constitute attractive chemical space, the size of such space is too large to explore comprehensively[4]. To facilitate exploration of such vast undeveloped chemical space, here we present "ChemiSpadon", an SNS-based platform for human-assisted chemical space exploration. ChemiSpadon is based on the open-source SNS system called Mastodon[5] and equipped with chemical structure posting system, structure evaluation system, and structure generating bots. Taking advantage of human knowledge and instinct, we expect efficient exploration of undeveloped but feasible chemical space for drug discovery. Participants are very welcome. Please visit: www.chemical.space

[1] Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 23:3-25, 1997.
[2] Cheng, A. C., Coleman, R. G., Smyth, K. T., Cao, Q., Soulard, P., Caffrey, D. R., ... & Huang, E. S., Structure-based maximal affinity model predicts small-molecule druggability. *Nature biotechnology*, 25(1):71, 2007.
[3] Lipinski, C. A., Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions, *Advanced drug delivery reviews*, 101:34-41, 2016.
[4] Ruddigkeit, L., Van Deursen, R., Blum, L. C., & Reymond, J. L., Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *Journal of chemical information and modeling*, 52(11):2864-2875, 2012.
[5] Mastodon https://joinmastodon.org/