1. Introduction
There are many enzymes and receptors in a living body which play several roles. GPCR (G protein-coupled receptor) is one of such protein super families. All GPCRs have a seven transmembrane (7-TM) region (Fig 1). GPCR is considered as drug targets of about half commercially-supplied drugs. But there are many GPCRs whose ligands have not yet been discovered. These GPCR are called as orphan GPCR. For example, Rhodopsin family, one of GPCR families, has 95 orphan GPCRs [1]. Discovering ligands of orphan GPCR is one of the most important tasks in medicinal field because it would lead to reveal unknown physiological phenomena and discovering new drug targets. A general method for discovering orphan GPCRs is to check activity of compounds in chemical libraries experimentally. But these methods require high cost and success rate of them is not necessarily high. Therefore, more efficient method for discovering ligands of orphan GPCRs is needed.

2. Objective
In this study, we aim to develop an in silico method for discovering ligands of orphan GPCRs. In silico prediction of ligand enable us to reduce amount of experiment and save high cost. Furthermore, it also would lead to reveal unknown physiological phenomena and discover new drug targets.

3. Method
A brief overview of proposed methods is shown below. First, by using a structure generator, we prepare many virtual chemical structures in a computer. Next, in order to predict activities of each chemical structure, we apply prediction models to them. Finally, we experimentally check activities of only compounds which show high prediction values.

Whether we can propose ligands of orphan GPCRs accurately by this method or not depends on accuracy of prediction models. In this study we use quantitative structure activity relationships (QSAR) methods in order to build models. QSAR is assumed to have some merits that it needs short calculation times and does not necessarily require mechanism of phenomena of the objects and so

KO-15 Development of a method for discovering orphan GPCR ligands by using chemoinformatics

Michio Koyama*, Masamoto Arakawa, Kimito Funatsu

The University of Tokyo

E-mail: funatsu@chemsys.t.u-tokyo.ac.jp
on. QSAR is the method to build models to predict activities by using only chemical compounds, not proteins. Therefore, it is impossible to build models to predict activities of proteins with no activity data. There is no activity data of orphan GPCR. Thus, no model can be built by using previous QSAR methods. But new QSAR methods which use both compounds and protein information have been researched.

Fig 2 shows the abstract of this method. First, we translate both compounds and proteins information into numerical data. Next, we apply machine learning methods to the data set. Machine learning methods are preferable to be ones that represent interactions of compounds and proteins. The method that uses a following kernel as a kernel of kernel learning machines is developed \cite{2}.

\[ K((c_1,p_1),(c_2,p_2)) = K(c_1,c_2) \times K(p_1,p_2) \]

\( c,p \) means descriptor vectors of compounds or proteins, respectively. This kernel is called as Target-Ligand Kernel (TLK). TLK is considered to represent interactions because TLK has product term in itself. When we use Gaussian kernel as kernel of TLK, TLK can be represented like below.

\[ K((c_1,p_1),(c_2,p_2)) = e^{-\gamma \|c_1-c_2\|^2} \times e^{-\gamma \|p_1-p_2\|^2} \]

When we apply TLK to support vector regression (SVR), predicted value of an SVR model can be represented like below.

\[ y_{pred} = \sum_{i \in SV} (\alpha_i - \alpha_i^*) \times (K(c_i,c_i) \times K(p_i,p_i)) \]

Although there are several kernel learning methods such as kernel PLS, Gaussian Process and so on, we first try to use SVR-TLK and inspect its prediction power.

4. Result and Discussion

We used ligands of Dopamine receptors of type 2 and type 3 as data set. Total number of compounds is 369, whose activity values (pKi) toward both proteins are measured.

We used Dragon\cite{3} to calculate descriptors of compounds. By using Dragon, we can calculate several types of descriptors such as topological and electrical information of molecules. We also used frequency of pairs of amino acids as descriptors of proteins.

The criterion of model evaluation is \( Q^2 \) values of ten-fold cross validation. Table 1 shows the result of analysis. In table 1, SVR-LK shows the result of SVR-Ligand Kernel, which is the SVR model use only information of compounds. From the result, \( Q^2 \) value of SVR-TLK is better than those of SVR-LK. Therefore, it is proven that by using SVR-TLK, activities of multiple proteins can be predicted precisely and simultaneously, which show the usefulness of this method. We plan to continue to examine usefulness of SVR-TLK focusing on whether activities of proteins with no ligand data can be predicted precisely by using SVR-TLK.

5. Conclusion

In this study, we have proposed the method for discovering ligands of orphan GPCR by using chemoinformatics. We investigated the usefulness of SVR-TLK. The result show better prediction power than past method. After more advanced investigation of SVR-TLK, we aim to propose ligands of orphan GPCRs by using this method.

References

3. http://www.talete.mi.it/