The Construction and Analysis of the Multidrug Resistance Transporter Knowledge-Base

Yoshinobu Igarashi igarashi@kuicr.kyoto-u.ac.jp
Yasushi Okuno okuno@kuicr.kyoto-u.ac.jp
Susumu Goto goto@kuicr.kyoto-u.ac.jp
Minoru Kanehisa kanehisa@kuicr.kyoto-u.ac.jp
Bioinformatics Center, ICR, Kyoto University, Uji, Kyoto 611-0011, Japan

Keywords: multidrug resistance, antibiotics, chemical compound, principal component analysis

1 Introduction

Multidrug resistance (MDR) proteins serve as transporters for chemical compounds, small molecules like antibiotics in almost all species from bacteria to higher eukaryote. They provide the resistance against antibiotics and agrichemicals for the cells like cancer, fungi and parasite, etc. Therefore, the research on MDR protein is essential and important in medical, agricultural and scientific fields.

Five transporter protein families are known to contain the MDR members as follows. ABC (ATP-Binding Cassette superfamily), MFS (Major-Facilitator Superfamily), SMR (Small Multidrug Resistance family), RND (Resistance-Nodulation-cell Division family) and MATE (Multidrug and Toxic compound Extrusion family).

The ABC and MFS consist of some subfamilies. Only the ABC transporter is driven by ATP hydrolysis energy, and other four families are driven by proton motive force. It was reported that more than 90% of MDR genes are composed of ABC and MFS genes in completely sequenced bacteria genomes [1].

In general, the relationship of the transporter and its substrate is highly specific, but MDRs are known to be transporters which have broad spectrums of drugs and chemical compounds. Therefore, the MDRs are not specific but poly-specific rather than non-specific transporters. Many substrates have been found and used in experiments for elucidating the mechanism and property of MDRs. The spectrum of each MDR family is known to be somewhat distinct, but the relationship to the structural and physicochemical properties of substrates has not been well established.

In this study, we collected the substrates from the literature and tried to retrieve structural and physicochemical properties that are representative of each substrate of MDRs and to define relationships between MDR families and their substrates.

2 Method

2.1 Constructing MDR Knowledge-Base

We collected published literature which described the relations of MDRs and their substrates, by searching PubMed with specific keywords (multidrug, ABC, MFS, etc.). We identified 71 MDR protein sequences and 212 chemical compounds in 24 species from 100 literatures, which consistuted our knowledge-base.

2.2 Converting Chemical Compound Information

The chemical structure of each compound was registered to the KEGG/LIGAND database [2]. We used a vector representation of chemical structures according to the 67 kinds of atom types with environmental information, reflecting their physicochemical properties.

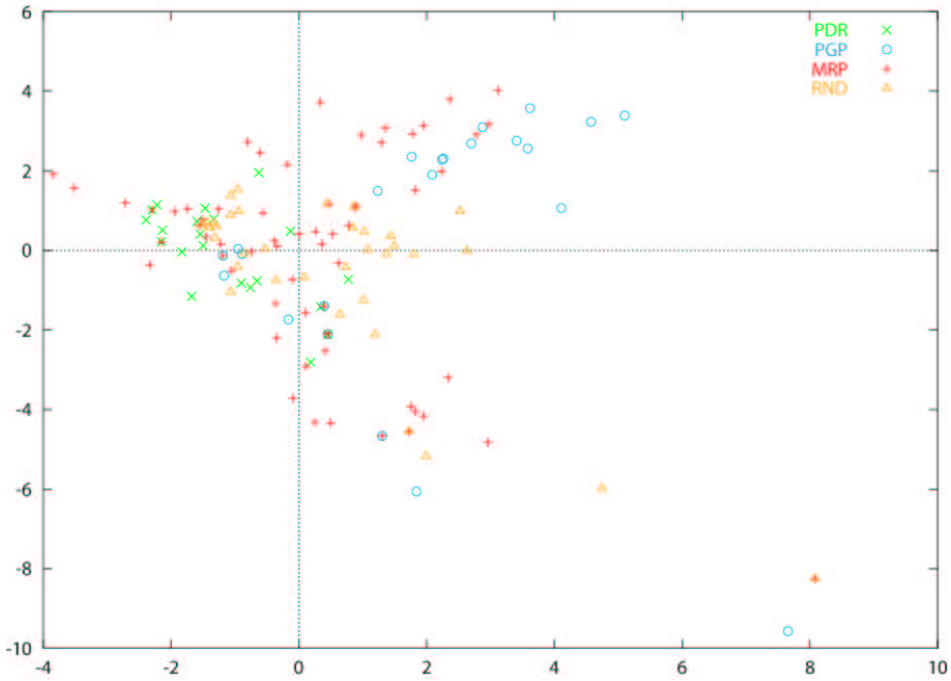

Figure 1: Plot of the first and second principal components score.

2.3 Classification of MDR Proteins

The classification of MDR proteins were performed by manual operation using the phylogenetic tree which was constructed by multiple alignment ClustalW ver.1.81 with default parameters and NJplot. This classification result was almost consist with the result of the complete linkage hierarchical cluster analysis.

2.4 Classification of MDR Substrates

The principal component analysis (PCA) was applied to the vector representation matrix of 212 chemical compounds by using GnuR statistical analysis package.

3 Results and Discussion

In the result of PCA, we observe several high factor loading coefficients in the first and second principal components. The five high coefficients in the first principal component consist of the atoms of the peptides. It is caused by MRP family which can transport glutathione conjugate compounds and pgp family of ABC which can transport peptide and peptide-like compounds, while other MDRs can transport hydrophobic compounds principally. The plot of the first and second principal component scores for the substrates of ABC (PGP, MRP, PDR) and RND is shown in Figure 1. We can observe distinct spectrum of substrates of each transport family. Further analysis and results will be reported.

Acknowledgments

We thank Sanae Asanuma, Tomoko Komeno, Rumiko Yamamoto and Koichiro Tonomura for improvement and maintenance of LIGAND database. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] Palusen, I.T., Sliwinski, M.K., and Saier, M.H.Jr., Microbial genome analyses: Global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities, *J. Mol. Biol.*, 277:573–592, 1998.

[2] Yamamoto, R., Goto, S., Okuno, Y., Hattori, M., Nishioka, T., and Kanehisa M., Improvement of the LIGAND chemical database, *Genome Informatics*, 13:492–493, 2002.