

Perspective

It has been established that miRNAs play critical roles in cell differentiation, proliferation, and apoptosis, and the abnormalities of specific miRNA expression contribute to the initiation and progression of tumor. However, identification of target mRNAs negatively regulated by miRNAs remain largely to be explored. Although up to hundreds of target genes toward a single miRNA were predicted by bioinformatics approaches (4), there is no comprehensive assay to biologically validate the prediction algorithm. Therefore, establishment of a method to comprehensively and rapidly identify target mRNAs for the miRNA is necessary for understanding biological and functional mechanisms of miRNA.

References

- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116:281–297.
- Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*. 2000;408:86–89.
- Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell*. 2005;120:21–24.
- Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005;120:15–20.
- Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A*. 2004;101:2999–3004.
- Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. 2002;99:15524–15529.
- Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A*. 2005;102:13944–13949. Erratum in: *Proc Natl Acad Sci U S A*. 2006;103:2464–2565.
- He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, et al. A microRNA polycistron as a potential human oncogene. *Nature*. 2005;435:828–833.
- O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*. 2005;435:839–843.
- Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature*. 2005;435:834–838.
- Lee Y, Jeon K, Lee JT, Kim S, Kim VN. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J*. 2002;21:4663–4670.
- Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*. 2004;23:4051–4060.
- Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature*. 2003;425:415–419.
- Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, et al. The Microprocessor complex mediates the genesis of microRNAs. *Nature*. 2004;432:235–240.
- Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev*. 2003;17:3011–3016.
- Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. *Cell*. 2003;115:209–216. Erratum in: *Cell*. 2003;115:505.
- Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the assembly of the RNAi enzyme complex. *Cell*. 2003;115:199–208.
- Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell*. 2004;15:185–197.
- Hutvagner G, Zamore PD. A microRNA in a multiple-turnover RNAi enzyme complex. *Science*. 2002;297:2056–2060.
- Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell*. 2005;123:631–640.
- Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, Basyuk E, et al. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science*. 2005;309:1573–1576.
- Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol*. 2005;7:719–723.
- Andrei MA, Ingelfinger D, Heintzmann R, Achsel T, Rivera-Pomar R, Luhrmann R. A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies. *RNA*. 2005;11:717–727.
- Liu J, Rivas FV, Wohlschlegel J, Yates JR 3rd, Parker R, Hannon GJ. A role for the P-body component GW182 in microRNA function. *Nat Cell Biol*. 2005;7:1261–1266.
- Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*. 2005;433:769–773.
- Chen CZ, Li L, Lodish HF, Bartel DP. MicroRNAs modulate hematopoietic lineage differentiation. *Science*. 2004;303:83–86.
- Guimaraes-Sternberg C, Meerson A, Shaked I, Soreq H. MicroRNA modulation of megakaryoblast fate involves cholinergic signaling. *Leuk Res*. 2006;30:583–595.
- Garzon R, Pichiorri F, Palumbo T, Iuliano R, Cimmino A, Aqeilan R, et al. MicroRNA fingerprints during human megakaryocytopoiesis. *Proc Natl Acad Sci U S A*. 2006;103:5078–5083.
- Fazi F, Rosa A, Fatica A, Gelmetti V, De Marchis ML, Nervi C, et al. A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPalpha regulates human granulopoiesis. *Cell*. 2005;123:819–831.
- Esau C, Davis S, Murray SF, Yu XX, Pandey SK, Pear M, et al. miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab*. 2006;3:87–98.

GLIDA: GPCR-ligand database for chemical genomic drug discovery

Yasushi Okuno*, Jiyeon Yang, Kei Taneishi, Hiroaki Yabuuchi and Gozoh Tsujimoto

Department of Genomic Drug Discovery Science, Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida-Shimo-Adachi-cho, Sakyo-ku, Kyoto 606-8501, Japan

Received August 15, 2005; Revised and Accepted September 22, 2005

ABSTRACT

G-protein coupled receptors (GPCRs) represent one of the most important families of drug targets in pharmaceutical development. GPCR-Ligand Database (GLIDA) is a novel public GPCR-related chemical genomic database that is primarily focused on the correlation of information between GPCRs and their ligands. It provides correlation data between GPCRs and their ligands, along with chemical information on the ligands, as well as access information to the various web databases regarding GPCRs. These data are connected with each other in a relational database, allowing users in the field of GPCR-related drug discovery to easily retrieve such information from either biological or chemical starting points. GLIDA includes structure similarity search functions for the GPCRs and for their ligands. Thus, GLIDA can provide correlation maps linking the searched homologous GPCRs (or ligands) with their ligands (or GPCRs). By analyzing the correlation patterns between GPCRs and ligands, we can gain more detailed knowledge about their interactions and improve drug design efforts by focusing on inferred candidates for GPCR-specific drugs. GLIDA is publicly available at <http://gdds.pharm.kyoto-u.ac.jp:8081/glida>. We hope that it will prove very useful for chemical genomic research and GPCR-related drug discovery.

INTRODUCTION

The superfamily of G-protein coupled receptors (GPCRs) forms the largest class of cell surface receptors. These molecules regulate various cellular functions responsible for physiological responses (1). GPCRs represent one of the most important families of drug targets in pharmaceutical development (2). A large majority of human-derived GPCRs still

remain 'orphans' with no identified natural ligands or functions, and thus a key goal of GPCR research related to drug design is to identify new ligands for such orphan GPCRs.

With the unprecedented accumulation of the genomic information, databases and bioinformatics have become essential tools to guide GPCR research. The GPCRDB (<http://www.gpcr.org/7tm/>) (2) and IUPHAR (<http://iuphar-db.org/iuphar-rd/index.html>) (3) receptor databases are representatives of widely used public databases covering GPCRs. These databases, which provide substantial data on the GPCR proteins and pharmacological information on receptor proteins containing GPCRs, are mainly focused on biological aspects of the gene products or proteins. In spite of the significance of ligand compounds as drug leads, the relationships between GPCRs and their ligands and/or chemical information on the ligands themselves are not yet fully covered.

On the other hand, there is increasing interest in collecting and applying chemical information in the post-genome era. This new trend is called 'chemical genomics', in which biological information and chemical information are integrated on the genome scale (4,5). PubChem (<http://pubchem.ncbi.nlm.nih.gov/>) (6), KEGG/LIGAND (<http://www.genome.jp/kegg/ligand.html>) (7) and ChEBI (<http://www.ebi.ac.uk/chebi/>) (8) have been developed as databases related to chemical genomics. KEGG/LIGAND and ChEBI contain primarily biochemical information on reported enzymatic reactions. Recently, NIH (the National Institutes of Health) opened PubChem, a public database providing information on the chemical structures of small molecules. However, one cannot retrieve direct information relating these chemical structures to gene or protein entries. Although chemical genomic approaches have thrown new light on relationships between receptor sequences and compounds that interact with particular receptors, the GPCR-ligand information is not well represented in these large-scale databases for chemical genomics.

There are still very few publicly available databases or tools for GPCR-specialized drug discovery from the viewpoint of chemical genomics. Herein, we have developed a novel relational database, GLIDA (GPCR-Ligand Database) (9).

*To whom correspondence should be addressed. Tel: +81 75 753 9264; Fax: +81 75 753 4544; Email: okuno@pharm.kyoto-u.ac.jp

Table 1. The current numbers of GLIDA ligands and GPCRs and their respective links

Information item	Number of entries
GPCR entries	3738
Links to Entrez Gene	3073
Links to GPCRDB	3738
Links to UniProt	3738
Links to IUPHAR	389
Links to KEGG	595
Ligand entries	649
Cas registry number	320
Molecular structure	364
Links to PubChem	242
Links to ChEBI	28
Links to KEGG	109
GPCR-ligand pair entries	1989
GPCR entries	281
Ligand entries	632

GLIDA contains biological information on GPCRs and chemical information on their ligand compounds. Furthermore, it provides various analytical data on GPCR-ligand correlations by incorporating bioinformatics and chemoinformatics methods, and thus it should prove very useful for chemical genomic research in GPCR-related drug discovery.

DATA CONTENTS

GLIDA contains three types of primary data: biological information on GPCRs, chemical information on their ligands and information on binding of specific GPCR-ligand pairs. The GPCR entries were acquired from the deposits of human, mouse and rat entries in the GPCRDB because these three species include sufficient information regarding ligands, and rats and mice are representative model animals for drug discovery. The ligand information was manually collected and curated using various public web sites and commercial DBs, such as the IUPHAR Receptor Database, PubMed, PubChem and MDL ISIS/Base 2.5. Table 1 indicates the size and scope of the GLIDA database.

GPCR and ligand data

The database lists general information on GPCR and ligand data, respectively. The general information table of GPCR contains gene names, family names, protein sequences and links to other biological databases, such as GPCRDB, UniProt, IUPHAR, Entrez Gene and KEGG. The ligand result page provides a general information table containing names, molecular structures, CAS registry numbers, formulas, molecular weights, MOLfiles and links to the other chemical databases KEGG, PubChem and ChEBI.

Information on binding of GPCR-ligand pairs

The correlation information relating GPCRs to particular ligands, a key issue for GPCR-related drug discovery, is stored in a relational database. GLIDA allows users to retrieve GPCR-ligand binding information dynamically and continuously. When users retrieve a GPCR (or ligand) entry, its result page displays all entries showing the corresponding ligands (or GPCR entries) with their binding activity types, as well as

references. The references are hyperlinked with the corresponding PubMed literature or the IUPHAR pages that were used to collect the information regarding GPCR-ligand binding. The activity types include agonist, inverse agonist, antagonist and so on. An agonist will bind to and activate the corresponding GPCRs, whereas an antagonist will bind to and block the activity of the corresponding GPCRs. An inverse agonist binds to GPCRs and reduces the fraction of them that are in an active conformation, and a partial agonist is an agonist that in a given tissue, under specified conditions, cannot elicit as large an effect as another agonist acting through the same GPCRs in the same tissue can.

WEB INTERFACE AND APPLICATION

GLIDA was constructed on the LAMP (Linux, Apache, MySQL and PHP) platform. GLIDA is available at <http://gdds.pharm.kyoto-u.ac.jp:8081/glida>. The web interface of GLIDA includes a GPCR search page (Figure 1a) and a ligand search page (Figure 1b). Each page consists of a classification table and a keyword search box. The user can search a GPCR (or ligand) manually from the guide-tree of the classification table, or automatically by using the keyword search function of MySQL. Every GPCR (or ligand) has its own result page (Figure 1c or d) containing a general information table for a GPCR (or ligand), a table of its correlated ligands (or GPCRs) and a button to carry out a similarity search and correlation analysis. Clicking the button starts the calculation, and an analytical report page (Figure 1e) then appears with a list of the top 25 entries that are most similar to the GPCR (or ligand) and a correlation map of the 25 GPCRs (or ligands) and their corresponding binding pairs. A search starting from ligand retrieval proceeds in the same way.

Hierarchical classification

The GPCR classification table on the search page was adapted from the phylogenetic tree of the GPCRDB information system (<http://www.gpcr.org/7tm/phylo/phylo.html>). As for the ligand classification table, GLIDA offers an original one (Figure 1b) that is based on a cluster analysis of the ligand structures as follows. We converted the structural images of the ligands into computational MDL Mol files using ISIS/Draw software. Next, we calculated distance metrics among all of the ligands using the frequency profiles of the atoms and the bonds of the KEGG atom types (10), and carried out complete-linkage clustering. We manually defined sub-clusters based on their common structural skeletons. Both the GPCR and ligand classification tables display the entries of the corresponding GPCRs or ligands at the end of the tree, and these are hyperlinked with their respective result pages.

Similarity search and GPCR-LIGAND correlation maps

GLIDA has a structure similarity search function on its result pages. Alignment scores of protein sequences generated by the BLAST algorithm provide similarity measures for GPCRs. Ligand similarity is defined by the dissimilarity (distance) of frequency profile patterns generated from the constitutive atoms and bonds of the chemical structure, using the KEGG atom types (10,11). From this similarity search, the 25 most

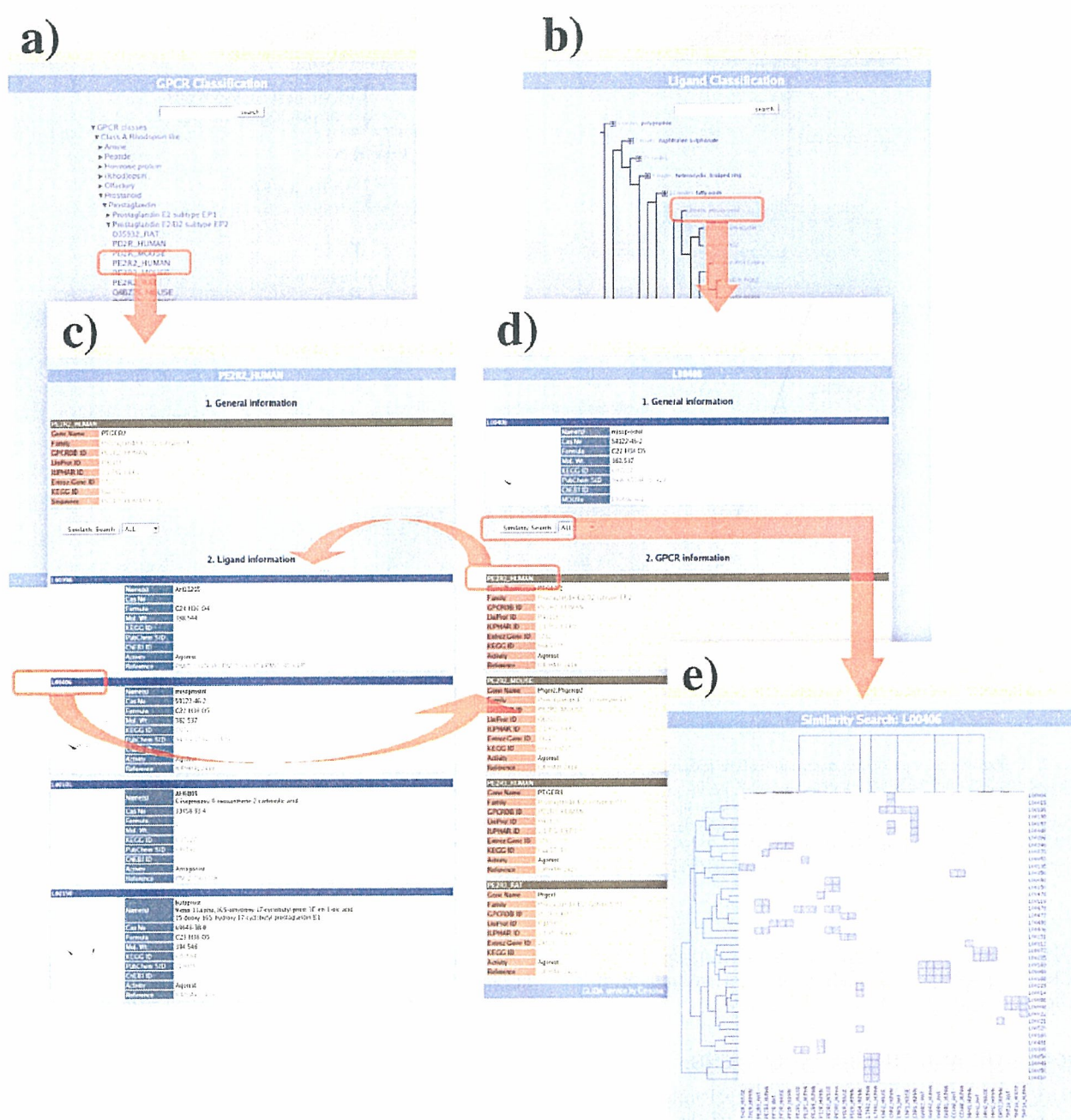


Figure 1. A screenshot of GLIDA showing its linked relations among search pages (a, b), result pages (c, d) and an analytical report page (e).

similar GPCRs (or 40 ligands) are retrieved and listed with their similarity scores on an analytical report page.

As the similarity search calculation is proceeding, GLIDA illustrates the correlation map (Figure 2c) showing the homologous GPCRs (or ligands) and their ligands (or GPCRs) that are retrieved. This map shows spots that match the GPCRs and their ligands in a two-dimensional matrix. The ordering along the x-axis and the y-axis are calculated respectively by

two-way clustering of the GPCRs and the ligands based on their similarities. In particular, the ordering along the x- and y-axis allows users to evaluate information regarding similarities and correlations between GPCRs and ligands simultaneously. By analyzing the correlation patterns between GPCRs and ligands that are illustrated by these maps, we can gain detailed knowledge about their interactions and utilize this information to infer possible candidates for development

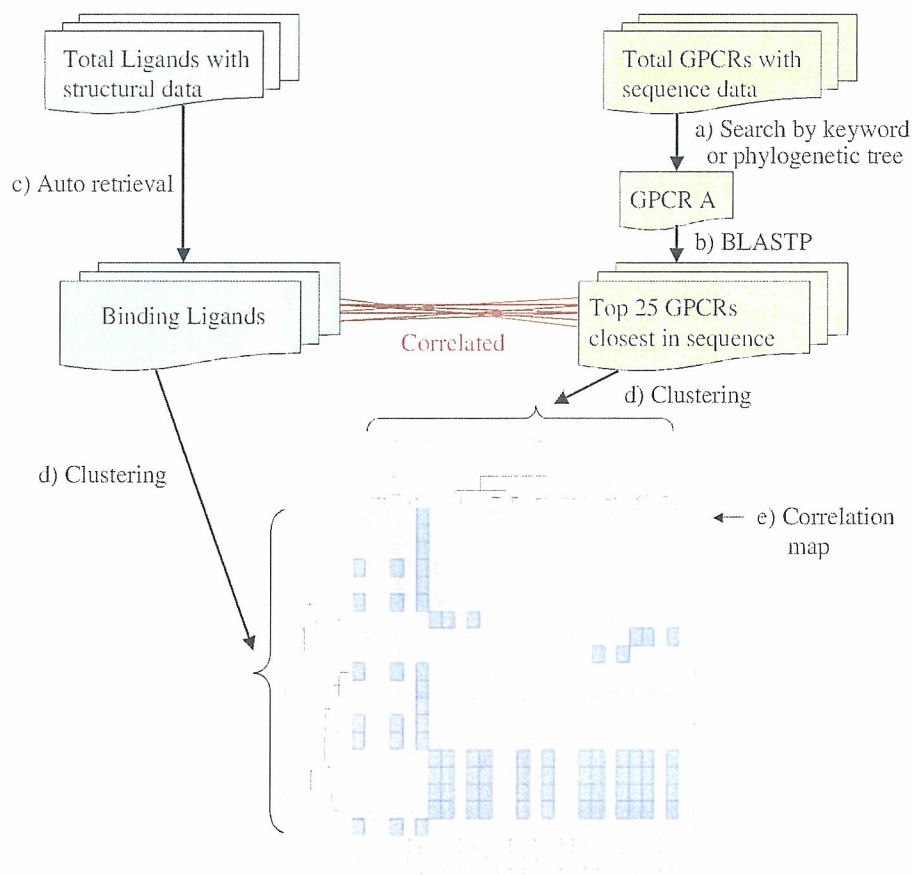


Figure 2. A schematic example of the search and analysis process showing GPCR-ligand correlations produced from a GPCR query using GLIDA. (a) If GPCR A is selected using a keyword search or a guide-tree search on the GPCR search page, its retrieved data will be displayed in its result page. (b) By clicking an analysis button on the result page, a list of the top 25 GPCRs that are most similar in sequence, including GPCR A, are obtained by the BLASTP calculation. (c) The server retrieves a list of corresponding ligands, which are respectively correlated with the 25 GPCRs. (d) Finally, a map is displayed to help visualize the matching spots linking GPCRs with particular ligands. The *x*-axis and *y*-axis respectively indicate the clustering results for GPCRs and ligands, calculated using sequence alignment scores among the GPCRs and structural profile distances among the ligands.

of GPCR-specific drugs. Figure 2 shows an example of the GPCR-ligand search and analysis process starting from a GPCR query using GLIDA.

DISCUSSION AND FUTURE DIRECTIONS

GLIDA provides a unique database for GPCR-related chemical genomic research and drug discovery. GLIDA is distinct from other public chemical genomic databases because it contains original, GPCR-specific chemical entries, although the total scale of its contents is not yet large (Table 1). GLIDA provides several advantages over other databases, in that a search can be started either from a GPCR or from a ligand. Thus, searches may be carried out in a dynamic and user-friendly way. GLIDA's coverage of chemical and biological information simultaneously also provides an advantage to users by saving them the time and labor required to search multiple databases. The ligand search page is another distinct characteristic of GLIDA in that it displays the structural distribution of ligands, and thereby facilitates research on

GPCR-related drugs by incorporating structural aspects of the ligand compounds. The analytical report pages resulting from the calculated structural similarities of GPCRs and ligands can give the user deep insights into the GPCR-ligand relationships. The lists of neighboring ligands (or GPCRs) and the correlation maps are useful visualizing tools for analyzing correlations among their structural features and their GPCR-ligand binding properties. Because the GLIDA algorithms can be applied to proteins other than the GPCR family, it may also be considered as a promising database for chemical genomics research.

GLIDA will be updated continuously. In particular, we are planning to computationally extract GPCR-ligand information from the literature and from patents using a text-mining tool, and to increase the number of ligand entries immediately. Further information on ligands from various computable chemical descriptors is currently being incorporated, and GLIDA will be combined with a system for predicting novel ligands of orphan GPCRs in the future. Furthermore, we also plan to carry out XML publication of GLIDA.

ACKNOWLEDGEMENTS

This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan, from the Ministry of Health, Labor and Welfare of Japan and from the 21st Century COE program 'Knowledge information infrastructure for Genome Science'. Funding to pay the Open Access publication charges for this article was provided by the Ministry of Health, Labor and Welfare of Japan.

Conflict of interest statement. None declared.

REFERENCES

1. George,S.R., O'Dowd,B.F. and Lee,S.P. (2002) G-protein-coupled receptor oligomerization and its potential for drug discovery. *Nature Rev. Drug Discov.*, **1**, 808–820.
2. Horn,F., Bettler,E., Oliveira,L., Campagne,F., Cohen,F.E. and Vriend,G. (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.*, **31**, 294–297.
3. Fredholm,B.B., Fleming,W.W., Vanhoutte,P.M. and Godfraind,T. (2002) The role of pharmacology in drug discovery. *Nature Rev. Drug Discov.*, **1**, 237–248.
4. Lipinski,C. and Hopkins,A. (2004) Navigating chemical space for biology and medicine. *Nature*, **432**, 855–861.
5. Dobson,C.M. (2004) Chemical space and biology. *Nature*, **432**, 824–828.
6. Zerhouni,E. (2003) Medicine: the NIH roadmap. *Science*, **302**, 63–72.
7. Goto,S., Okuno,Y., Hattori,M., Nishioka,T. and Kanehisa,M. (2002) LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.*, **30**, 402–404.
8. Brooksbank,C., Cameron,G. and Thornton,J. (2005) The European Bioinformatics Institute's data resources: towards systems biology. *Nucleic Acids Res.*, **33**, D46–D53.
9. Yang,J., Okuno,Y. and Tsujimoto,G. (2004) GLIDA: GPCR and Ligand Database. *Genome Informatics*, **15**, P057.
10. Hattori,M., Okuno,Y., Goto,S. and Kanehisa,M. (2003) Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.*, **125**, 11853–11865.
11. Kotera,M., Okuno,Y., Hattori,M., Goto,S. and Kanehisa,M. (2004) Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions. *J. Am. Chem. Soc.*, **126**, 16487–16498.

第4回(平成18年度)薬学研究ビジョン部会部会賞受賞者(1)

ケミカル・バイオ情報に基づく創薬インフォマティクス研究

奥野 恭史 京都大学大学院薬学研究科 統合薬学フロンティア教育センター

1、はじめに

ヒトゲノムが解読された今日、莫大なゲノム情報から創薬への手がかりを発見すること、すなわち「ゲノム創薬」に大きな期待が寄せられている。ゲノム創薬は、ゲノム情報を出発点とし創薬の標的遺伝子探索からリード化合物探索を経て臨床段階に至る広範で高度に専門化した複合領域であり、その実践にはこれらの複合領域の橋渡しを実現する統合的なインフォマティクス基盤(創薬インフォマティクス)が必須となる。我々は、創薬インフォマティクスという新たな研究分野の創成に向け、バイオ情報を扱うバイオインフォマティクスとケミカル情報を扱うケモインフォマティクスの独立に発展してきた2つの情報科学分野の統合を図り、バイオ情報とケミカル情報の両者を同時に統合的にマイニングする新しい情報技術の開発に着手している。さらに本研究は、現在、国内外で注目されているケミカルゲノミクス・ケミカルバイオロジーのための有力な情報基盤ともなり得るものと考えられる。

2、化合物の宇宙探索(ケミカル空間の探索)

2004年12月のNature誌において、Chemical Space特集号が発表された(1)。そこでは、化合物の種類は1060個を越える天文学的なバリエーションを有しており、化合物空間を探索することは宇宙探索と同様に壮大な課題であることが提示されている。このことは医薬品の候補化合物となり得る新規な活性化化合物を見つけ出すことが如何に困難でセレンディップなことであるかを示唆するものである。

これらケミカル空間の探索の基礎研究としてケミカルゲノミクス・ケミカルバイオロジー研究が近年注目されている。ケミカルゲノミクスでは、その命題として「莫大な数の化合物と生体系(タンパク質や細胞など)との相互作用を包括的に明らかにすること」が挙げられている。実際、米国では、ケミカルゲノミクスプロジェクトを掲げ、数百万もの膨大な化合物に関する情報を収集し、

有用化合物の探索に国策として取り組んでいる。

しかしながら、広大な化合物空間から生物活性を有する化合物を探し当てる化合物探索には、天文学的な数量に対応できる新たなインフォマティクス技術とハイスループット技術の研究開発が必須である。そこで、我々は、莫大な化合物群とタンパク質群との相互作用様式をゲノムスケールで解析することを目的とした情報学的技術、すなわちケミカルゲノミクスのためのインフォマティクス技術の研究開発を行っている。

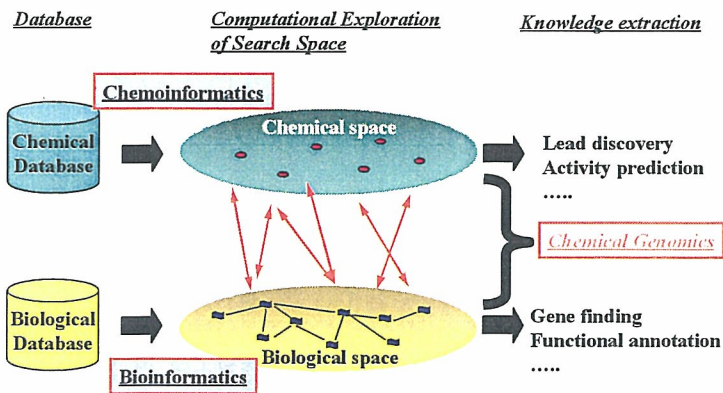
3、ケモインフォマティクスとバイオインフォマティクス(図1)

ケミカルゲノミクス・ケミカルバイオロジーでは、化合物のケミカル情報と生体系のバイオ情報の2種の異なる情報が対象となる。従って、ケミカルゲノミクスのための情報処理技術には、ケミカル情報を処理するケモインフォマティクスとバイオ情報を処理するバイオインフォマティクスを融合する新たなインフォマティクス技術の開発が必須となる。しかしながら、化学と生物学という異なる分野を背景にもつ2つのインフォマティクスは、独立して発展してきており現状では互いに相容れない。そこで、我々はケモインフォマティクスとバイオインフォマティクスにおける方法論的なアナロジーに着目しその融合を図った。すなわち、ケモインフォマティクスもバイオインフォマティクスも共に、個体(化合物やタンパク質)の特徴量を数値やベクトルで表現することにより、各個体の相対的な特性の違いを探索空間上の個体間の距離として定量的に算出する方法論を基本としている。例えば、ケモインフォマティクスでは、データベースに集積された膨大な化合物エントリーは化学構造や特性を定量的に表すベクトルとして表現され、その相対的な違いを距離の尺度として持つ座標空間(探索空間)をコンピュータ内部に構築する。データベース検索はこの探索空間において距離が近接する化合物を類似化合物として選出してくることに

なる。また、バイオインフォマティクスでも同様の考え方であり、遺伝子・タンパク質エントリーは配列や構造として表現され、それぞれの相同性（類似度）を尺度として持つ探索空間（バイオデータの場合、探索空間は系統樹やネットワーク構造になっている場合もある。）が構築され、データベース検索にはこの探索空間に基づき、類似（類縁）遺伝子・タンパク質が選出される。

一方、ケミカルゲノミクスとは、ケミカル空間の個体（化合物）とバイオ空間の個体（遺伝子・タンパク質）との相互作用関係を網羅的に明らかにする研究であり、図1の赤線に示す対応関係を付加したモデルであると考えられる。ここで、我々は、ケミカル情報とバイオ情報を統合的に処理するために、ケミカル空間（緑色）とバイオ空間（黄色）を独立して扱うのではなく、2つの空間を融合したモデルをケミカルゲノミクスのためのインフォマティクスモデルとして初めて考案した。

図1. ケモインフォマティクスとバイオインフォマティクス



4. ケミカル空間-バイオ空間の融合モデル

情報科学的アプローチによる化合物探索は、これまで化合物のケミカル情報のみを用いたケモインフォマティクス手法が用いられてきた。これに対し、我々の手法は、このケミカル情報のみの従来手法にバイオインフォマティクス技術を融合させ、バイオ情報を考慮に入れた化合物探索を実現する新しいインフォマティクス手法と言える。（図1）

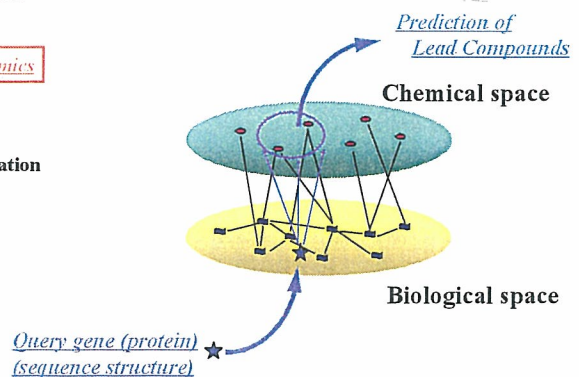
例えば、化合物について構造や特性の類似性を相対的な位置関係として表現したものをケミカル空間（赤が化合物、緑領域がケミカル空間）と

して定義するとともに、タンパク質についても類似関係（配列や構造の相同性）を相対的な位置関係として表現したものをバイオ空間（青がタンパク質、黄色領域がバイオ空間）として定義する。さらに個々の化合物とタンパク質の結合をリンク（黒線）することによって、これらケミカル空間とバイオ空間を融合した単純モデルをコンピュータ内部に構築できる。（図2）

ここで、標的タンパク質に作用する化合物候補を探索する *In silico* スクリーニングにこの融合モデルを適用する場合を考えると、

- 1) 標的タンパク質（青星）の配列構造から、そのタンパク質がバイオ空間座標にマッピングされる。
- 2) バイオ空間にマッピングされた標的タンパク質の近隣タンパク質からのケミカル空間へのリンク情報をたどること（青矢印）により、その標的タンパク質が関係するケミカル空間のエリア（青円内）を指定することができる。
- 3) 上記エリア内の化合物群が、標的タンパク質に相互作用する可能性のある化合物群と推定される。（ここでは、類似のタンパク質は、類似の化合物を結合するという前提を基にしている。）

図2. ケミカル空間-バイオ空間融合モデルを用いた *In silico* スクリーニング



我々は、このケミカル空間とバイオ空間の融合モデルを用いた探索を、GPCR ファミリーとそのリガンド化合物の探索に適用し、GLIDA データベース (2,3) として Web サービスを行っている。

5. ケミカル空間-バイオ空間の相関モデル

我々が考案する融合モデルで最も重要なことは、モデルの構築方法である。我々の方法は、こ

のモデル構築方法においても、従来法より大きな優位性を有している。

ケミカル空間の構築に用いられる従来法の代表的なものに主成分分析 (PCA) がある。これは、ケミカル情報のみを用い、化合物の化学特性ができる限り多様になるように、ケミカル空間座標を定義するものであり、ここでの大きな問題点は、化合物の多様性と生物活性との直接の因果関係は無いということである。

これらの問題点を克服することを目的とし、我々は正準相関分析 (CCA) を用いて相関モデル構築を試みた。本手法は、ケミカル情報とタンパク配列情報の両情報を用いて、ケミカル空間とタンパク空間の両空間の相関が高くなるように互いの空間座標を定義するものである。これはバイオ空間の分布を考慮して、ケミカル空間座標を定義する方が、生物活性にとって都合の良い空間座標を構築できるという大きな特徴を有する。

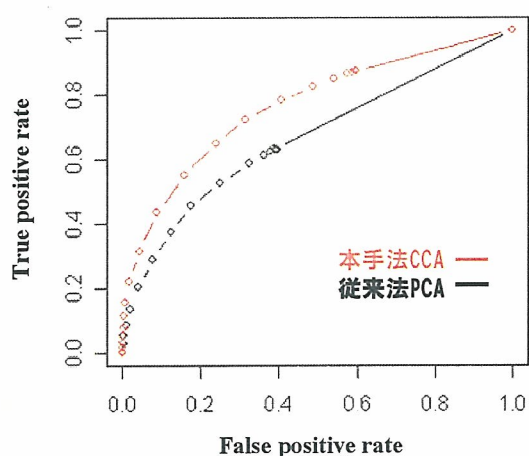


図 3. 本手法と従来法との性能比較

実際に、従来法 (PCA) と本手法 (CCA) の性能を比較するために、既知データ (DrugBank データベース (4)) を用いた 5-fold cross validation テストを実施し、化合物とタンパク質の相互作用予測の予測性能評価を行った。その結果を図 3 に示す。図は、予測性能を評価する有名な方法の一つである ROC 曲線 (横軸は化合物-タンパク質相互作用を誤って予測した割合、縦軸は正しく予測した割合を示す。) であり、このグラフは曲線が上に位置するほど、予測性能が良いことを表す。本手法の赤曲線が従来法の黒曲線より上方に位置するこ

とから、本手法の方が、従来法よりも高い予測性能であることがわかる。

6、Biologically relevant chemical space

ごく最近、宇宙観測において、暗黒物質 (ダークマター: 目には見えない物質) の宇宙空間での分布が初めて観測され宇宙の起源解明に期待が寄せられている。化合物の宇宙空間では、我々は何を探索しなければならないのであろうか? Chemical Space 特集号では、その答えとして「生物にとって意味のある化合物群 (Biologically relevant chemical space)」を探索することの重要性を提示している。しかしながら、Biologically relevant chemical space を定義する具体的方法に関する報告は未だなされていない。

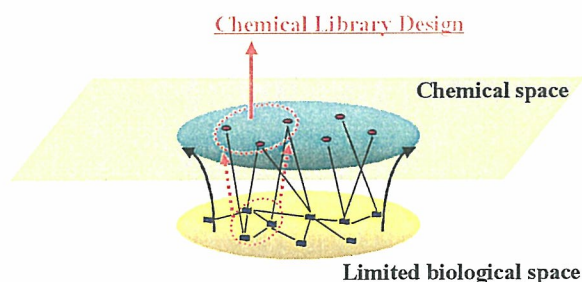


図 4. Biologically relevant chemical space

我々の開発した融合モデルはこの難問に一つの解答を与える事ができ、タンパク質群と化合物群の相互作用様式の統計モデルを構築することにより、広大なケミカル空間のうち黄色のバイオ空間に対応するケミカル空間 (biologically relevant chemical space) を限定することを可能にする。(図 4) また、バイオ空間のタンパク群を、例えば GPCR ファミリー (赤点線エリア) などに限定することにより GPCR 用のフォーカスライブラリー的设计も可能になり、生物活性を有する化合物ライブラリーの合理的設計が実現できる。

例えば、図 5 は、カナダでサービスされている DrugBank データベース (4) における 3476 個の既知薬物ターゲットタンパク質からなるバイオ空間とそれに対応する 3079 個の既知薬物からなるケミカル空間の実際の相関モデルである。このように、Biologically relevant chemical space の定義が可能となる。

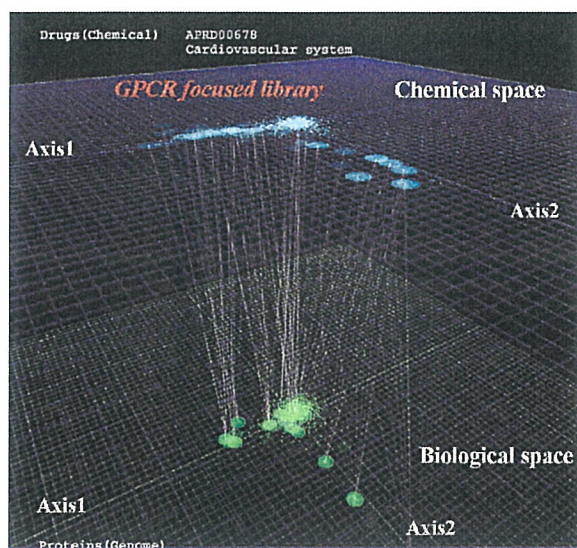


図5. DrugBank データを用いた実際のケミカル空間とバイオ空間

7. おわりに

医薬品の開発プロセスにおいて、現在用いられている In silico 技術には、我々が研究開発を行う情報科学的アプローチの他に、立体構造モデルを用いたドッキングシミュレーションのような計算化学的アプローチが有名である。情報科学的アプローチと計算化学的アプローチには、それぞれ一長一短があるが、欠点を互いに補完し合い In silico 創薬の確度向上を図ることが今後の課題であろう。特に、ケミカルゲノミクスが盛んな今日、日々増加し続ける莫大なデータを処理することは必須であり、情報科学的アプローチである創薬インフォマティクスのさらなる研究開発が必要である。

謝辞

本研究の一部は、文部科学省、厚生労働省の支援によって行われている。また、検証実験等の共同研究を行って頂いている京都大学薬学研究科ゲノム創薬科学分野の辻本豪三教授に深く感謝申し上げます。

参考文献

- 1) Nature, 432 (7019) (Insight), 823-865
- 2) Okuno, Y., Yang, J., Taneishi, K., Yabuuchi, H., Tsujimoto, G., Nucleic Acids Research, 34, D673-677
- 3) <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/>
- 4) <http://redpoll.pharmacy.ualberta.ca/drugbank/>

◆略歴◆ 奥野 恭史 (Yasushi OKUNO) : 1995年京大薬・修士課程修了、京大薬・博士後期課程進学、1996年京大薬・博士後期課程中途退学、京都大学化学研究所教務職員、2000年京大薬博士号取得、2001年京都大学化学研究所博士研究員、2002年京都大学化学研究所特任助手、2003年京都大学大学院薬学研究科特任助手、2005年産業技術総合研究所外来研究員(併任)、2006年京都大学大学院薬学研究科助教授

バイオ関連企業の動向

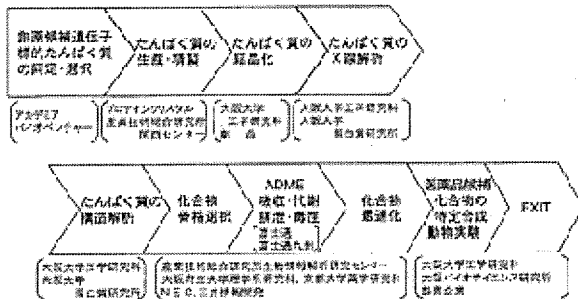
バイオグリップドセン ター関西

バイオグリップドセンター
関西、大阪府豊中市、
下柳副理事長、06・
6873・2156)は
「創薬バリューチェーン
」プロジェクト(代表
「坂田福則理事)を通過
している。同プロの提案
「坂田福則理事)を通過
化合物の発見(探索)もま

「インシリコでの創薬手
法の確立とその実証研
究」は、科省の大蔵省
(彩都)地域知的クラス
ター形成事業に採択され
た。コンビューター技術
を用いて、医薬品候補化

コンビューターによる

創薬バリューチェーンの概念



創薬バリューチェーン プロジェクトを進行

医薬品候補化合物探索の道探る

候補化合物の探索の取り
組としては、京都大学
大分薬学研究所の野村
基功助教が推進したコ
ンビューターを用いた機
能解析のたんぱく質に対
する薬物候補化合物の探
索手法「相対マシニング
」が有力である。シミュレシ
ョンにより結合モードを
探索するドッキングシミュ
レーションはたんぱく質
の構造解析が必要にな
る。構造の動的な変化を
正確にシミュレーション
するには難しい。予期
結果の信頼性、再現性た
らば問題がある。特にたん
ぱく質は新薬開発の
標的となるが、構造解析が
難しく構造はほとんど分か
っていない。このため
構造が分からなくても、
薬を作用させるも求められ
ている。

相対マシニングマシ
ニング法は化合物空間とた
んぱく質空間を、それぞれ
れぞれの次元に基いて定義
する。既知の多対多の相
互作用パターンを統計的
にリアル化し、相互作用
レベルに基いて、化合物
を選別、予測する。同法
では既知のたんぱく質の立
体構造は不要。換えては
く質の結合、一次配列の
みでの高精度予測できる
こと(特許取得済)。相互
作用空間を定義し、化学
構造の向上を評価する
ため、新薬候補の発見の
可能性が高い。計算時間
が短縮することで、大規
模である。

薬物は候補化合物の
製法たんぱく質を用いた。
既法を用い、論文などの既
知化合物の情報を利用し、
計算機で解析し約1000
の化合物を抽出。その化
合物を購入して生物活性
を測定する。すでに既に
論文で発表されている化
合物活性より1000倍
の作用と同等の作用を示
す化合物を抽出した。1
000倍の活性を示す化
合物は、すでに薬として
使えるナンセルオーダー
の活性を有し、新薬開発
に近づけることになる。ま
た低阻化合物と同等の活
性を示す20個の化合物を
探索することを考えてい
る。

国際級クラスターの 形成でわが国をけん引