

Fig. 1. Maturation of BMDCs co-cultured with fibroblasts in the presence of SCF. (A) BMDCs co-cultured for 16 days were collected (Day-0, -4, -8, -12, and -16) and stained with Alcian blue and Safranin-O. Bars = 10  $\mu\text{m}$ . (B–E) Granule protease activities of the cells were measured (B: chymotryptic activity, C: tryptic activity, D: carboxypeptidase A activity). Histamine content was also measured (E). Values are presented by the means  $\pm$  S.E.M. ( $n = 5$ ). (F) Surface expression of Fc $\epsilon$ R1 and c-kit was determined by flow cytometry.

ities (Fig. 1B and C). In addition, induction of other protease genes, such as cathepsin B (*Ctsb*) and transmembrane tryptase (*Tpsg1*), was also observed. (ii) Induction of histidine decarboxylase (*Hdc*), which catalyzes histamine synthesis, is also consistent with the increase in cellular histamine content (Fig. 1E). (iii) Heparan sulfate biosynthesis is mediated by an array of enzymes [19]. We found the increased gene expression of these enzymes, including exostosin 1 (*Ext1*), and heparan sulfate 6-O-sulfotransferase 2 (*Hs6st2*), which mediates synthesis of polysaccharide backbone and 6-O-sulfation after *N*-deacetylation and *N*-sulfation of *N*-acetyl glucosamine, respectively. Glucosaminyl *N*-deacetylase/*N*-sulfotransferase-2 (*Ndst2*) is known to be essential for heparin biosynthesis in mast cells, and peritoneal mast cells in the *NDST-2*-deficient mice exhibit aberrant granule morphology and decreased granule proteases [13,20]. However, we unexpectedly found a decrease in the *NDST-2* gene expression during maturation. Analysis of these enzymes at the protein level is surely required to address this inconsistency. (iv) As for arachidonic acid metabolism, up-regulation of hematopoietic

prostaglandin (PG) D synthase (*Ptgds2*) and down-regulation of leukotriene (LT) C<sub>4</sub> synthase (*Ltc4s*) are consistent with the previous report indicating the production of PGD<sub>2</sub> in preference to LTC<sub>4</sub> in CTMCs [17]. Given that CTMCs adhere to the extracellular matrix in the tissues, it is interesting that up-regulation of the genes related to adhesion, such as integrins (*Itga5*, *Itga9*, and *Itgb1*), CD44 (*Cd44*), and ICAM-1 (*Icam1*), was detected during the co-culture period. Altogether, these changes in the gene expression are in good accordance with phenotypic changes observed during mast cell maturation, indicating that our culture system successfully mimics the maturation process.

Immunoblot analyses confirmed that the changes in the expression of various proteins paralleled those of the corresponding gene expression in the microarray analyses (Fig. 2A and B). GATA-1 and GATA-2 were reported to be required for mast cell differentiation [21,22]. Recently, Masuda et al. demonstrated that these transcription factors are required for maintaining mast cell-specific gene expression and are required

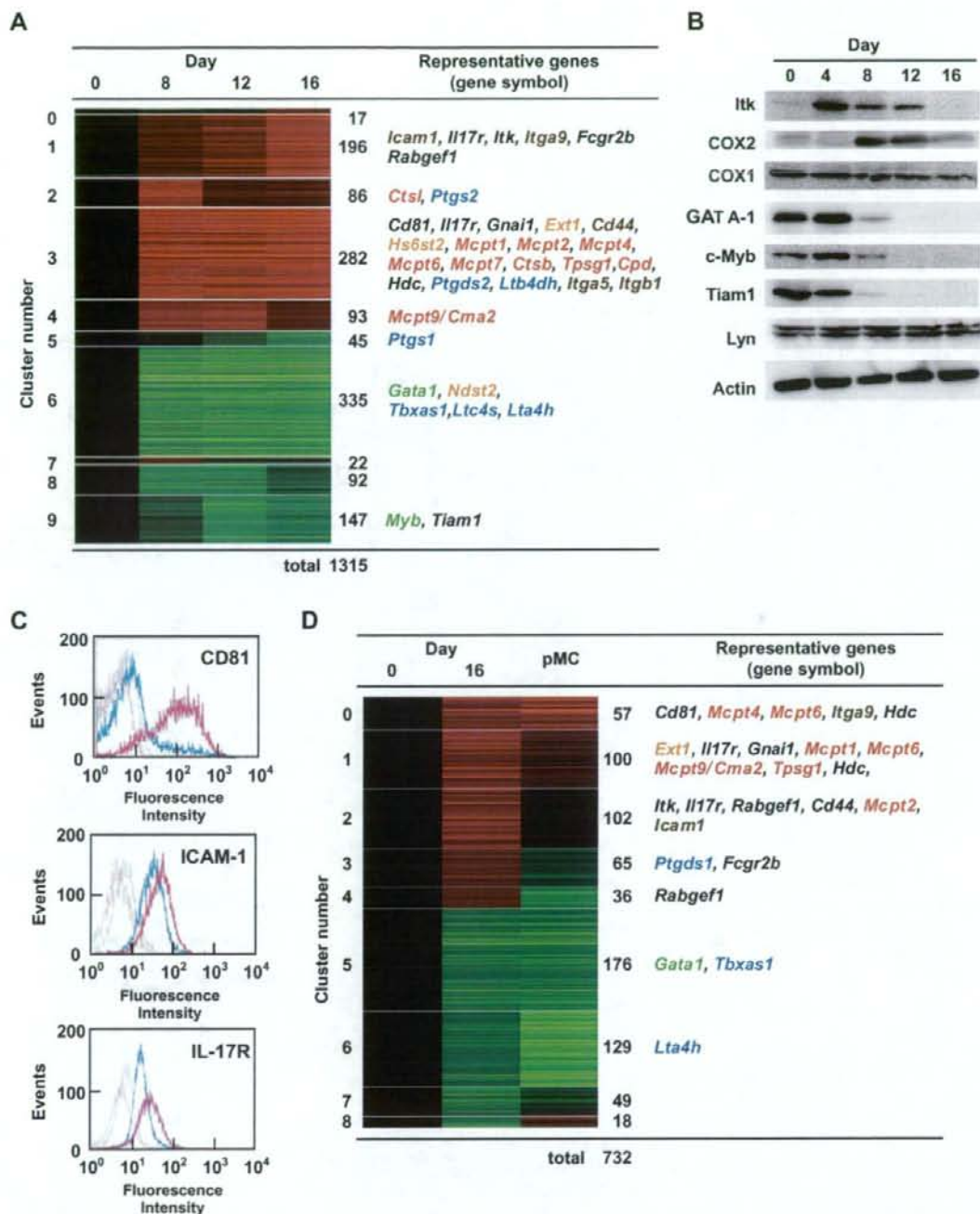


Fig. 2. Clustering analysis of the gene expression profile during maturation of cultured mast cells. (A) Co-cultured mast cells were subjected to analysis of the gene expression profiles using microarray (Affymetrix, 430A murine GeneChip, ~22000 genes). Selected 1315 genes, of which maximum minus minimum expression values were greater than 2 (2-fold change), were classified into 10 clusters on the basis of similarity in the expression profiles and presented by the columns (red: increased, green: decreased). The number of genes and the representative genes (proteases; red, enzymes related to arachidonate metabolism; blue, adhesion molecules; brown, enzymes related to heparin synthesis; orange, and transcription factors; green) in each cluster are indicated on the right. (B) The results of immunoblot analyses of the selected proteins are presented. Immunoblot analyses for Lyn and actin were performed as the loading control. (C) Surface expression of CD81, ICAM-1, and IL-17Rs was investigated by flow cytometry in BMMCs (blue lines) and co-cultured mast cells (Day-16, red lines). The isotype controls were shown by the dotted lines. (D) BMMCs (Day-0), Day-16 co-cultured mast cells (Day-16), and peritoneal c-kit<sup>+</sup> cells (pMC) were subjected to analysis of the gene expression profiles using microarray (430A murine GeneChip). Selected 732 genes, of which differences in expression values were greater than 2 between Day-0 and Day-16 co-cultured mast cells, were classified into 9 clusters on the basis of similarity in the expression profiles. The number of genes and the representative genes in each cluster are indicated.

for activation upon IgE-mediated antigen stimulation [23]. We observed, however, a drastic GATA-1 down-regulation during the later stage of co-culture period (Fig. 2B), indicating that GATA-1 may be dispensable for the final step of mast cell maturation.

We then analyzed expression of several membrane proteins, including CD81, ICAM-1, and the IL-17 receptor (IL-17R), by flow cytometry (Fig. 2C). *CD81* was one of the strikingly up-regulated genes during the co-culture period (supplemental data). Since Yamada et al. demonstrated that the higher surface expression of CD81 is one of the signatures of cutaneous mast cells [10], it is likely that the drastic up-regulation at Day-16 reflects the differentiation towards CTMC-like cells. The surface expression of ICAM-1 was also confirmed to be up-regulated in the Day-16 cultured mast cells. Consistent with induction of the IL-17R gene on the microarray analysis (Fig. 2A, clusters #1 and #3, *Il17r*), the surface expression of IL-17R, which was detectable on BMMCs, was significantly up-regulated in the Day-16 cultured mast cells. Nakae et al. demonstrated that IL-17 enhanced ionomycin-induced TNF- $\alpha$  production in BMMCs, although IL-17 alone failed to induce TNF- $\alpha$  [24]. IL-17R may determine the sensitivity to IL-17 in mature tissue mast cells. Our results indicate that these membrane proteins are good candidates for the surface marker of CTMCs.

### 3.3. Comparison of gene expression profiles between cultured mast cells and peritoneal mast cells

We compared the gene expression profile of the Day-16 co-cultured mast cells with that of peritoneal mast cells, which are regarded as typical CTMCs. The characteristic changes in gene expression profile between the Day-0 and Day-16 co-cultured mast cells were also confirmed in the case of the comparison between BMMCs (Day-0) and peritoneal mast cells; approximately 70% of the extracted genes (511 genes in 732 genes; parallel increase in clusters 0 and 1, parallel decrease in clusters 5, 6, and 7), of which differences in expression values were greater than 2-fold between Day-0 and Day-16 co-cultured mast cells, exhibited similar trends in the peritoneal mast cells (Fig. 2D). This result suggests that the approach with the co-culture system can efficiently select the genes, of which expression is characteristic of CTMCs.

### 3.4. Degranulation of cultured mast cells in response to cationic secretagogues

BMMCs and the co-cultured mast cells responded to IgE-mediated antigen stimulation in a similar fashion up to Day-12 (Fig. 3A). A significant decrease in degranulation was found at Day-16, which may be attributed to up-regulation of the molecules that can suppress activation of mast cells upon IgE-mediated antigen stimulation, such as Fc $\gamma$ RIIB, Rab-GEF1, and CD81 (Fig. 2A, clusters #1 and #3) [25–27]. The constant levels of degranulation in response to a Ca<sup>2+</sup> ionophore, A23187, were observed, indicating that the process of Ca<sup>2+</sup> influx-mediated degranulation was not affected during the co-culture period (Fig. 3B).

CTMCs can be distinguished from MMCs or BMMCs in their sensitivity to cationic secretagogues, such as compound 48/80. At Day-16, the cultured mast cells released 60% and 40% of  $\beta$ -hexosaminidase activity in response to compound 48/80 and substance P, respectively, whereas no or little release

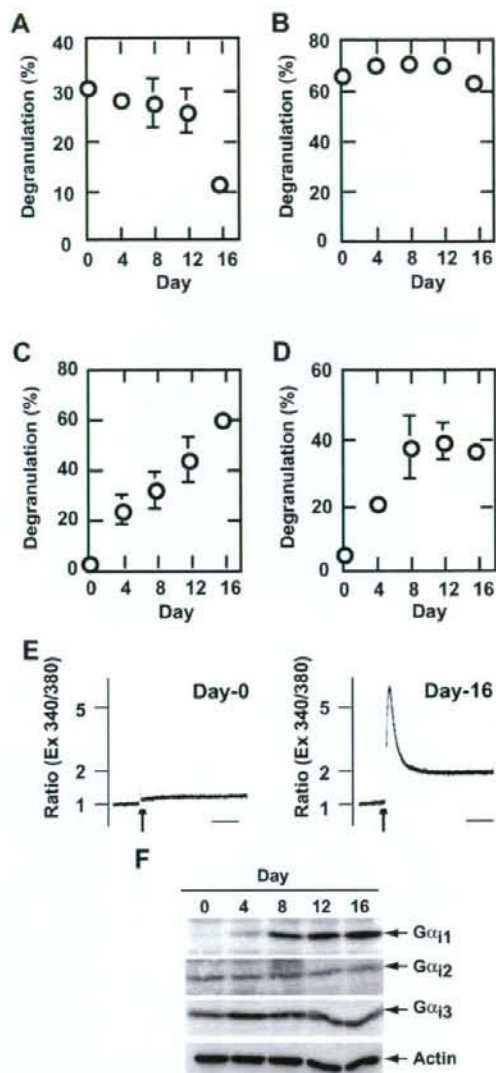


Fig. 3. Cationic secretagogue-induced degranulation and G $\alpha$ <sub>i</sub> expression. (A–D) Co-cultured mast cells were separated and stimulated with 1  $\mu$ M A23187 (B), 10  $\mu$ g/ml compound 48/80 (C), or 100  $\mu$ M substance P (D). In case of IgE-mediated antigen stimulation, the sensitized mast cells were stimulated with the antigen as described in Section 2 (A). Values are presented by the means  $\pm$  S.E.M. ( $n = 5$ ). (E) Intracellular Ca<sup>2+</sup> mobilization was monitored using Fura-2/AM. BMMCs (Day-0) and co-cultured mast cells (Day-16) were stimulated with 10  $\mu$ g/ml compound 48/80 (indicated by the arrows). These are the representatives of three similar independent experiments. Bars = 1 min. (F) Expression of three G $\alpha$ <sub>i</sub> subtypes in the co-cultured mast cells at each time point was examined by immunoblot analyses. An anti-actin antibody was used as a loading control.

from BMMCs (Day-0) was detected (Fig. 3C and D). Since a previous study demonstrated that compound 48/80 causes intracellular Ca<sup>2+</sup> mobilization in mast cells [28], we then examined cytosolic Ca<sup>2+</sup> response to the secretagogue. As expected, compound 48/80 increased the cytosolic Ca<sup>2+</sup> concentrations in

the Day-16 cultured mast cells, but not in BMDCs (Day-0) (Fig. 3E).

Degranulation induced by basic secretagogues was reported to be mediated by the heterotrimeric G protein,  $G_{\beta\gamma}$ , especially by  $G_{13}$ , in rat peritoneal mast cells [29]. Ferry et al. suggested that the  $\beta\gamma$  subunits of  $G_{12}$  and  $G_{13}$  induced degranulation through activating phospholipase C $\beta$  [30]. Furthermore, a recent study using cultured human mast cells revealed that the  $G_{\alpha_{13}}$  protein was induced upon co-culture with 3T3 fibroblasts and suggested an involvement of  $G_{13}$  in eosinophil major basic protein-induced degranulation [31]. We then examined the levels of three  $G_{\alpha_i}$  subtypes (Fig. 3F). In good agreement with the microarray results (Fig. 2A, cluster #3, *Gnai1*), the level of the  $G_{\alpha_{11}}$  protein was drastically increased during the co-culture period, whereas the level of  $G_{\alpha_{12}}$  or  $G_{\alpha_{13}}$  was not significantly changed. Furthermore, the  $G_{\alpha_{11}}$  induction is well correlated with the elevated degranulation in response to compound 48/80 (Fig. 3C). Previous studies using basic secretagogues have paid little attention to the  $G_{\alpha_{11}}$  protein, since they mostly used rat peritoneal mast cells, which lack the expression of  $G_{\alpha_{11}}$ . Our results suggest that  $G_{\alpha_{11}}$  is a potent candidate for the regulator of basic secretagogue-induced degranulation in mouse mast cells.

### 3.5. Conclusion

In summary, we succeeded in improving the culture system to develop the CTMC-like mast cells. This system allows us to investigate the gene expression profiles during maturation of mast cells, which should contribute to understanding the mechanism underlying mast cell terminal differentiation.

**Acknowledgements:** This study was supported in part by 21st Century COE Program "Knowledge Information Infrastructure for Genome Science", a grant from Uehara Memorial Foundation, Naito Foundation, Kao Foundation for Arts and Sciences, Sankyo Foundation of Life Science, and by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports and Technology of Japan and from the Ministry of Health and Labor of Japan.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2008.03.033.

### References

- [1] Metz, M., Grimaldeston, M.A., Nakae, S., Piliponsky, A.M., Tsai, M. and Galli, S.J. (2007) Mast cells in the promotion and limitation of chronic inflammation. *Immunol. Rev.* 217, 304–328.
- [2] Metz, M. and Maurer, M. (2007) Mast cells—key effector cells in immune responses. *Trends Immunol.* 28, 234–241.
- [3] Kitamura, Y. (1989) Heterogeneity of mast cells and phenotypic change between subpopulations. *Annu. Rev. Immunol.* 7, 59–76.
- [4] Stevens, R.L. and Austen, K.F. (1989) Recent advances in the cellular and molecular biology of mast cells. *Immunol. Today* 10, 381–386.
- [5] Benoist, C. and Mathis, D. (2002) Mast cells in autoimmune disease. *Nature* 420, 875–878.
- [6] Bryce, P.J., Miller, M.L., Miyajima, I., Tsai, M., Galli, S.J. and Oettgen, H.C. (2004) Immune sensitization in the skin is enhanced by antigen-independent effects of IgE. *Immunity* 20, 381–392.
- [7] Lu, L.F., Lind, E.F., Gondek, D.C., Bennett, K.A., Gleeson, M.W., Pino-Lagos, K., Scott, Z.A., Coyle, A.J., Reed, J.L., Van

- Snick, R.J., Strom, T.B., Zheng, X.X. and Noelle, R.J. (2006) Mast cells are essential intermediaries in regulatory T-cell tolerance. *Nature* 442, 997–1002.
- [8] Levi-Strauss, F., Austen, K.F., Gravalles, P.M. and Stevens, R.L. (1986) Coculture of interleukin 3-dependent mouse mast cells with fibroblasts results in a phenotypic change of the mast cells. *Proc. Natl. Acad. Sci. USA* 83, 6485–6488.
- [9] Dayton, E.T., Pharr, P., Ogawa, M., Serafini, W.E., Austen, K.F., Levi-Strauss, F. and Stevens, R.L. (1988) 3T3 fibroblasts induce cloned interleukin 3-dependent mouse mast cells to resemble connective tissue mast cells in granular constituency. *Proc. Natl. Acad. Sci. USA* 85, 569–572.
- [10] Yamada, N., Matsushima, H., Tagaya, Y., Shimada, S. and Katz, S.I. (2003) Generation of a large number of connective tissue type mast cells by culture of murine fetal skin cells. *J. Invest. Dermatol.* 121, 1425–1432.
- [11] Tanaka, S., Takasu, Y., Mikura, S., Satoh, N. and Ichikawa, A. (2002) Antigen-independent induction of histamine synthesis by immunoglobulin E in mouse bone marrow-derived mast cells. *J. Exp. Med.* 196, 229–235.
- [12] Ogasawara, T., Murakami, M., Suzuki-Nishimura, T., Uchida, M.K. and Kudo, I. (1997) Mouse bone marrow-derived mast cells undergo exocytosis, prostaglandin generation, and cytokine expression in response to G protein-activating polybasic compounds after coculture with fibroblasts in the presence of c-kit ligand. *J. Immunol.* 158, 393–404.
- [13] Forsberg, E., Pejler, G., Ringvall, M., Lunderius, C., Tomasini-Johansson, B., Kusche-Gullberg, M., Eriksson, I., Ledin, J., Hellman, L. and Kjellén, L. (1999) Abnormal mast cells in mice deficient in a heparin-synthesizing enzyme. *Nature* 400, 773–776.
- [14] Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B. and Speed, T.P. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 31, e15.
- [15] Tanaka, S., Mikura, S., Hashimoto, E., Sugimoto, Y. and Ichikawa, A. (2005)  $Ca^{2+}$  influx-mediated histamine synthesis and IL-6 release in mast cells activated by monomeric IgE. *Eur. J. Immunol.* 35, 460–468.
- [16] Dayton, E.T., Caulfield, J.P., Hein, A., Austen, K.F. and Stevens, R.L. (1989) Regulation of the growth rate of mouse fibroblasts by IL-3-activated mouse bone marrow-derived mast cells. *J. Immunol.* 142, 4307–4313.
- [17] Metcalfe, D.D., Baram, D. and Mekori, Y. (1997) Mast cells. *Physiol. Rev.* 77, 1033–1079.
- [18] Tsai, M., Wedemeyer, J., Ganiatsas, S., Tam, S.Y., Zon, L.I. and Galli, S.J. (2000) In vivo immunological function of mast cells derived from embryonic stem cells: an approach for the rapid analysis of even embryonic lethal mutations in adult mice in vivo. *Proc. Natl. Acad. Sci. USA* 97, 9186–9190.
- [19] Kusche-Gullberg, M. and Kjellén, L. (2003) Sulfotransferase in glycosaminoglycan biosynthesis. *Curr. Opin. Struct. Biol.* 13, 605–611.
- [20] Humphries, D.E., Wong, G.W., Friend, D.S., Gurish, M.F., Qiu, W.-T., Huang, C., Sharpe, A.H. and Stevens, R.L. (1999) Heparin is essential for the storage of specific granule proteases in mast cells. *Nature* 400, 769–772.
- [21] Walsh, J.C., DeKoter, R.P., Lee, H.J., Smith, E.D., Lancki, D.W., Gurish, M.F., Friend, D.S., Stevens, R.L., Anastasi, J. and Singh, H. (2002) Cooperative and antagonistic interplay between PU.1 and GATA-2 in the specification of myeloid cell fates. *Immunity* 17, 665–676.
- [22] Migliaccio, A.R., Rana, R.A., Sanchez, M., Lorenzini, R., Centurione, L., Bianchi, L., Vannucchi, A.M., Migliaccio, G. and Orkin, S.H. (2003) GATA-1 as a regulator of mast cell differentiation revealed by the phenotype of the GATA-1<sup>low</sup> mouse mutant. *J. Exp. Med.* 197, 281–296.
- [23] Masuda, A., Hashimoto, K., Yokoi, T., Doi, T., Kodama, T., Kume, H., Ohno, K. and Matsuguchi, T. (2007) Essential role of GATA transcription factors in the activation of mast cells. *J. Immunol.* 178, 360–368.
- [24] Nakae, S., Suto, H., Berry, G.J. and Galli, S.J. (2007) Mast cell-derived TNF can promote Th17 cell-dependent neutrophil recruitment in ovalbumin-challenged OTII mice. *Blood* 109, 3640–3648.
- [25] Malbec, O. and Daeron, M. (2007) The mast cell IgG receptors and their roles in tissue inflammation. *Immunol. Rev.* 217, 206–221.

- [26] Fleming, T.J., Donnadiou, E., Song, C.H., Laethem, F.V., Galli, S.J. and Kinet, J.P. (1997) Negative regulation of FcεRI-mediated degranulation by CD81. *J. Exp. Med.* 186, 1307–1314.
- [27] Tam, S.Y., Tsai, M., Snouwaert, J.N., Kalesnikoff, J., Scherrer, D., Nakae, S., Chatterjea, D., Bouley, D.M. and Galli, S.J. (2004) RabGEF1 is a negative regulator of mast cell activation and skin inflammation. *Nat. Immunol.* 5, 844–852.
- [28] Senyshyn, J., Baumgartner, R.A. and Beaven, M.A. (1998) Quercetin sensitizes RBL-2H3 cells to polybasic mast cell secretagogues through increased expression of G<sub>i</sub> GTP-binding proteins linked to a phospholipase C signaling pathway. *J. Immunol.* 160, 5136–5144.
- [29] Aridor, M., Rajmlevich, G., Beaven, M.A. and Sagi-Eisenberg, R. (1993) Activation of exocytosis by the heterotrimeric G protein G<sub>13</sub>. *Science* 262, 1569–1572.
- [30] Ferry, X., Eichwald, V., Daeffler, L. and Landry, Y. (2001) Activation of βγ subunits of G<sub>12</sub> and G<sub>13</sub> proteins by basic secretagogues induces exocytosis through phospholipase Cβ and arachidonate release through phospholipase Cγ in mast cells. *J. Immunol.* 167, 4805–4813.
- [31] Piliponsky, A.M., Gleich, G.J., Nagler, A., Bar, I. and Levi-Schaffer, F. (2003) Non-IgE-dependent activation of human lung- and cord blood-derived mast cells is induced by eosinophil major basic protein and modulated by the membrane form of stem cell factor. *Blood* 101, 1898–1904.

バイオインフォマティクスとケモインフォマティクスの融合による  
インシリコ創薬研究

奥野 恭史

*In silico* Drug Discovery Based on the Integration of Bioinformatics  
and Chemoinformatics

Yasushi OKUNO

*Department of PharmacoInformatics, Centre for Integrative Education of Pharmacy Frontier,  
Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida  
Shimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan*

(Received May 28, 2008)

With the near completion of the human genome sequencing, bioinformatics and chemoinformatics are expected as promising tools in genome-based drug discovery. The emerging field of chemical genomics is accumulating large-scale assay data on compound-protein interactions. We are now developing new mining methods for the chemical genomics data based on the integration of bioinformatics and chemoinformatics. Here we present a GPCR-ligand database (GLIDA) and a novel *in silico* screening method, which we have developed. GLIDA is a novel public GPCR-related chemical genomics database that is primarily focused on the correlation of information between GPCRs and their ligands. Our *in silico* screening method is based on statistical machine learning of the conserved patterns of molecular recognition extracted from comprehensive compound-protein interaction data. These are promising approaches to accelerating drug discovery processes.

**Key words**—chemoinformatics; bioinformatics; *in silico* screening

## 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掛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。われわれは、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という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。なお本研究は、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. ケミカル空間とケミカルゲノミクス

2004年12月のNature誌において、Chemical Space特集号が発表された。<sup>1)</sup>そこでは、化合物の種類は $10^{60}$ 個を超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いか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

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（〒606-8501 京都市左京区吉田下阿達町46-29）

e-mail: okuno@pharm.kyoto-u.ac.jp

本総説は、平成20年度日本薬学会奨励賞の受賞を記念して記述したものである。

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、われわれは、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

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。したがって、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を持つ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、われわれ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ともに、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、Fig. 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、われわれは、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考案した。

### 4. ケミカル空間とバイオ空間の融合モデル

情報科学の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てきた。これに対し、われわれの手法は、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（Fig. 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なモデルを構築できる（Fig. 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。

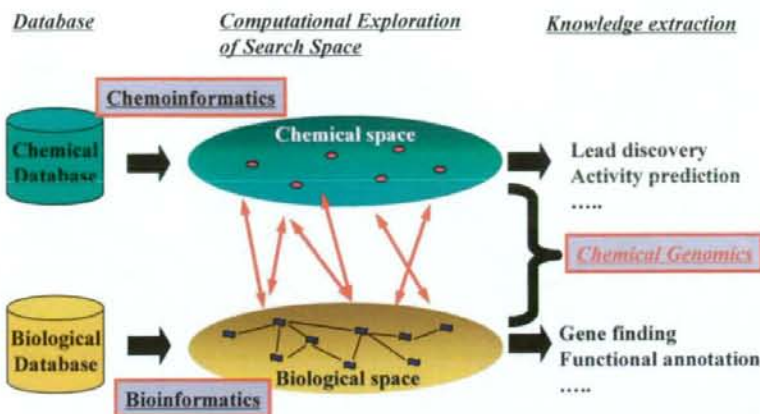
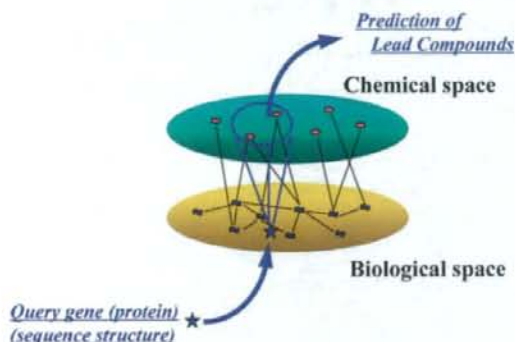


Fig. 1. Bioinformatics and Chemoinformatics

Fig. 2. *In silico* Screening for Chemical Genomics Data

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)。

われわれは、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<sup>24)</sup>としてWebサービスを行っている。GLIDAは、GPCRのバイオ情報、そのリガンドのケミカル情報、及びGPCRとリガンドの相互作用情報の3種類の情報より構成されている。GPCRのエントリーはヒト、ラット、マウスに限定し、バイオ情報はGPCRDBから取得した。また、GPCRと結合するリガンドのエントリーとそのケミカルデータ(化学名、構造式、分子量、MDL Molファイルなど)はIUPHAR Receptor Database, PubMed, Pub-

Chem及びMDL ISIS/Base 2.5などの公共又は商用のデータベースから取得した。具体的には、2008年1月現在、24077件のリガンドエントリー、及び39140件のリガンド-GPCRの相互作用エントリーの登録に至っている。

各エントリーの検索は、GPCR(またはリガンド)のキーワード検索及びクラス分類テーブルから行うことが可能である。ここで、GPCR分類は、GPCRDBに定義されている進化系統樹由来の分類に従っている。またリガンド分類は、KEGGで定義されている原子タイプの原子数/結合数に基づいた頻度プロファイルから距離行列を計算し、主成分分析(PCA)に基づいてGLIDA独自のリガンド分類を行っている。検索された各GPCR(またはリガンド)のページには、バイオ情報(またはケミカル情報)、及びそれらに結合するリガンド(またはGPCR)のリストが同時に表示される。さらに、GLIDAのGPCR(またはリガンド)のページはGPCR-リガンド相互作用の解析機能を有している。すなわち、検索されたGPCR(またはリガンド)と最も高い類似性を持つ25個のGPCR(またはリガンド)リストを表示するとともに、これら25個のエントリーと結合するリガンド(またはGPCR)との相互作用様式を2次元マップ表示する。このマップの2軸に並ぶGPCRとリガンドの順番は、各々GPCRとリガンドのクラスタリング結果を反映している。したがって、GPCR、リガンドの類似性情報と相互作用情報を同時に視覚化し、このパターンを分析してGPCRとリガンドの相互



作用予測を実現し、薬物と作用基点の相互作用に関する情報を得ることができる (Fig. 3, Fig. 4)。

### 5. ケミカルゲノミクスに基づくバーチャルスクリーニング

活性化化合物を効率よく迅速に探し出すために、計算機を用いた候補化合物の絞り込み手法「バーチャルスクリーニング (VS)」が開発されてきた。現在よく用いられている VS として、既知リガンドとの構造類似性に基づく「Ligand-based virtual screening (LBVS)」と標的タンパク質の立体構造に基づ

く「Structure-based virtual screening (SBVS)」がある。<sup>5,6)</sup> この2つの手法は、近年の情報技術の進歩と相まって、この10-20年で著しい発展を遂げ、ゆるぎない地位を確立した。しかしながら現在、VSのヒット確率は1-10%もあればよしとされており(例えば、LBVSでは、既知活性化化合物の骨格構造に強く影響される嫌いがあるし、SBVSでは、パラメータの恣意性、予測的中率の低さなどが指摘されている)、さらなる技術的な改良や革新的技術の開発が切望されていることは間違いない。

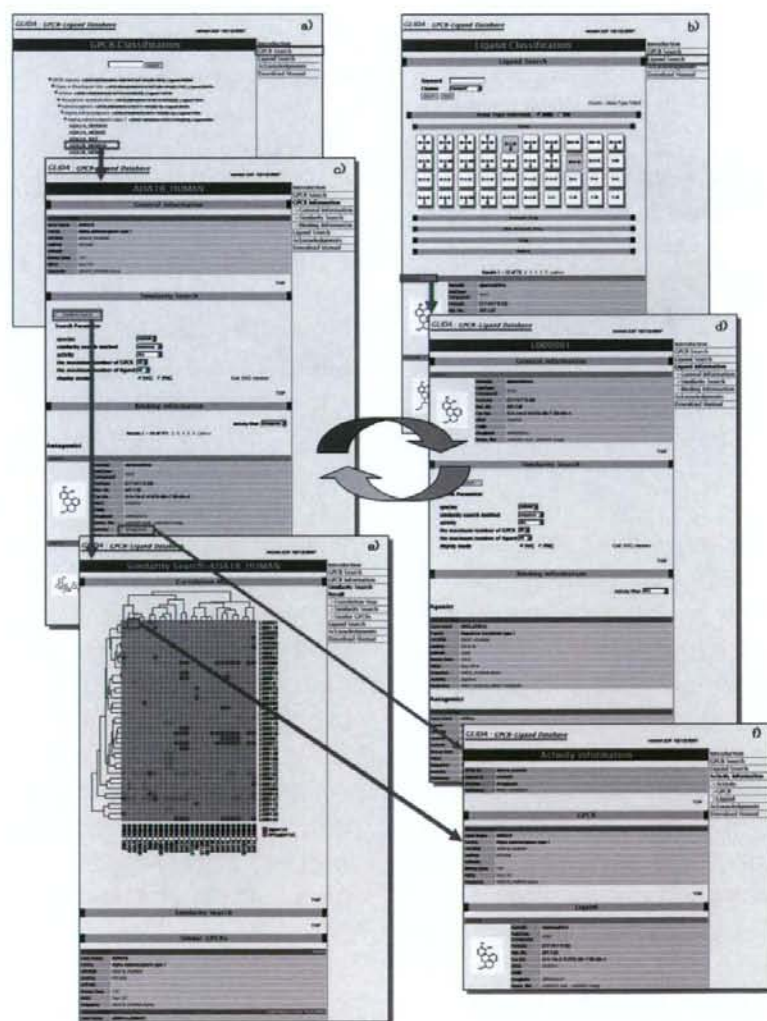


Fig. 3. A Screenshot of GLIDA Showing Linked Relations among Search Pages (a, b), Result Pages (c, d), an Analytical Report Page (e), and a Binding Information Page (f)

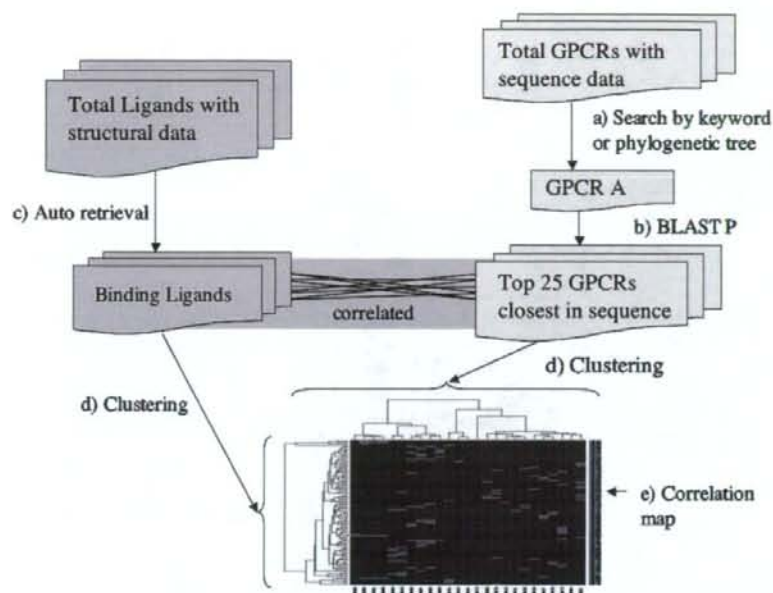


Fig. 4. A Schematic Example of the Search and Analysis Process Showing GPCR-ligand Correlations Produced from a GPCR Query Using GLIDA

ここでは、LBVS や SBVS とは概念の異なる第 3 の VS として、われわれが開発している「ケミカルゲノミクスに基づく VS 手法 (Chemical Genomics-based virtual screening; CGBVS)」を紹介する。「ケミカルゲノミクス」は、興味を持つ化合物が生物に与える影響についてゲノム規模で研究する学問であり、<sup>7,8)</sup> マイクロアレイやハイスループットスクリーニングなどの同時大量解析技術の革新に後押しされ、近年、注目を浴び始めている。それに伴い、化合物と遺伝子の関連性について、膨大な実験データが蓄積されている。そこでわれわれは、情報科学技術の 1 つであるパターン認識技術を用いて、タンパク質と化合物との結合情報 (ケミカルゲノミクス情報) から抽出したタンパク質のリガンド認識パターンに基づいて活性化化合物を効率的に発見する新たな VS、「CGBVS」を開発している。

タンパク質とリガンドとの相互作用パターンの認識とそのリガンド予測を開発するために、われわれは学習アルゴリズムの一種であるサポートベクターマシン (SVM)<sup>9)</sup> を用いた。SVM は、2 クラス分類器の一種であり、与えられた 2 つのグループに属する特徴ベクトルを最大マージンで分離するような超平面を構築する。ここで、最大マージンとは、分離

した超平面から各サンプル間までの最短距離を指す。

われわれは、この SVM を用いて、化合物-タンパク質相互作用の有/無を判別する手法を開発した。その手法の流れを Fig. 5 に示す。まず、収集した相互作用をベクトルとして表現するために、各化合物の化学構造、各タンパク質のアミノ酸配列について、様々な属性 (記述子と呼ぶ) を計算する。次に、正例 (相互作用する化合物-タンパク質ペア) 及び負例 (相互作用しない化合物-タンパク質ペア) に対応する記述子をそれぞれ組み合わせ特徴ベクトルを構成し、SVM を用いて学習モデルを構築する。このモデルが得られると、(未知の化合物-タンパク質ペアに相当する) 新しいベクトルが相互作用有/無のどちらのクラスに属するかを予測することができる。

既存の VS 手法との比較検討を行うため、今回開発した CGBVS と LBVS との予測性能を比較した。収集した化合物-GPCR 相互作用を用いて、負例を交換しながら 5 分割交差検定法 (5 fold cross-validation) を試行した。交差検定の結果、最近傍法を用いた LBVS では  $84.4 \pm 0.3\%$ 、CGBVS では  $91.6 \pm 0.2\%$  の相互作用を正しく予測した。また、ROC 曲線からも、CGBVS の予測性能の高さが確

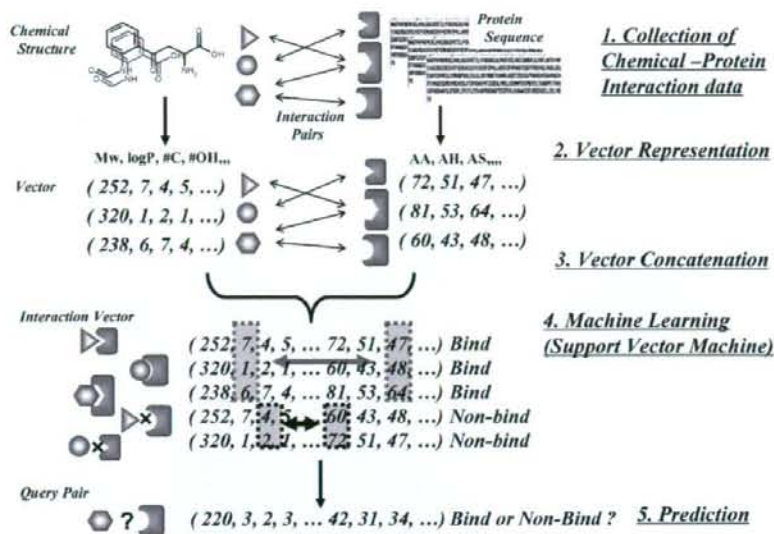


Fig. 5. Overview of Chemical Genomics-based Virtual Screening

認された (Fig. 6)。したがって、ケミカルゲノミクス情報の活用がリガンド予測性能の向上につながったといえる。

さらに、ヒト  $\beta 2$  アドレナリン受容体 ( $\beta 2AR$ ) を標的 GPCR とし、構築した学習モデルを用いてリガンド予測を行い、*in vitro* 実験による検証を行った。リガンド予測の対象化合物は、 $\beta 2AR$  以外の GPCR のリガンドとして知られている 826 化合物とした。ここで、CGBVS により予測された  $\beta 2AR$  リガンド候補上位 50 の化合物のうち、文献・特許調査により 14 種の化合物について  $\beta 2AR$  との相互作用に関する報告を確認した。さらに、残りの相互作用不明な化合物のうち、入手可能な 21 種類について *in vitro* 結合阻害実験を行ったところ、17 種類の化合物が相互作用 ( $10^{-5} M < IC_{50} < 10^{-3} M$ ) を示した。結合阻害実験のヒット率は 81% (17/21) に上り、ここにおいても高い予測的中率が示された。

また今回新たに発見した化合物の化学構造を精査したところ、典型的な  $\beta 2AR$  作動薬の構造 (カテコラミン骨格、イソプレナリン誘導体) 及び  $\beta 2AR$  拮抗薬の構造 (アリルアルキルアミン誘導体) とは異なる多様な骨格を含んでおり、化合物の構造類似性に基づく従来の方法では発見が困難なリガンド群が含まれることが明らかになった。したがって、ケミカルゲノミクス情報が、リガンド予測精度の向上

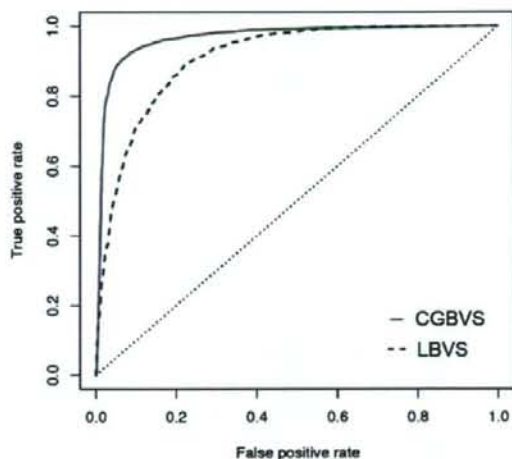


Fig. 6. Comparison of Prediction Performance between Chemical Genomics-based and Ligand-based Virtual Screenings

のみならず、新規骨格を持つリガンドの検出にも有用であることが示唆された。

## 6. おわりに

従来、生物活性を示す化合物の探索には、化合物の構造類似性とその指標とされてきた。しかし、化合物の構造類似性だけでタンパク質との多種多様な相互作用を見出すには限界があり、新しい探索手法の開発が望まれていた。われわれは、「ケミカル

ゲノミクス情報の中から相互作用パターンを抽出する」という新しいアプローチを試み、1) 化合物-GPCR 質間相互作用データベース GLIDA を開発し、2) 学習アルゴリズムを用いた予測モデルにより、GPCR リガンドを精度よく予測できることを実証した。

なお、 $\beta 2AR$  の立体構造が 2007 年決定された<sup>10)</sup> ことを受けて、われわれは SBVS との予測性能比較も行っており、その結果、CGBVS が SBVS よりも高精度で  $\beta 2AR$  リガンドを予測できることを確認している。

われわれが開発している CGBVS は、標的タンパク質のリガンド情報がないケースにも適用可能であり、実際に現在、オーファン GPCR を標的としたリガンド予測を行っている。タンパク質の立体構造情報を必要としない点は、詳細な 3D 構造が不明な GPCR にとって、魅力的なりガンド予測手法といえる。将来的には、結合阻害定数 (Ki 値) など定量的な活性データから回帰分析を行うことにより、相互作用の強さを反映させたりガンド予測手法への拡張を行う予定である。ケミカルゲノミクス情報を利用するという新規なアプローチは、GPCR に限らずすべてのタンパク質に適用可能であり、化合物-タンパク質相互作用の理解、ひいては創薬プロセスの効率化につながる事が期待される。

**謝辞** 本研究の一部は、文部科学省、経済産業省 (NEDO 若手グラント)、厚生労働省の助成金支援によって行われており、深く感謝申し上げます。本総説は、平成 20 年度日本薬学会奨励賞の受賞を記念して記述したものであり、日本薬学会役員、審査

員の先生方をはじめご関係の皆様にご心より感謝申し上げます。また、これまでご指導、ご鞭撻賜りました学生時代の恩師・杉浦幸雄先生 (現同志社女子大学教授)、研究員時代の恩師・金久 實先生 (京都大学化学研究所教授)、助手時代の恩師・辻本豪三先生 (京都大学薬学研究科教授)、並びに京都大学薬学研究科研究科長・藤井信孝先生に謹んで感謝の意を表します。

## REFERENCES

- 1) *Nature*, **432** (7019) (Insight), 823-865 (2004).
- 2) Okuno Y., Yang J., Taneishi K., Yabuuchi H., Tsujimoto G., *Nucleic Acids Res.*, **34**, D673-677 (2006).
- 3) Okuno Y., Tamon A., Yabuuchi H., Nijijima S., Minowa Y., Tonomura K., Kunimoto R., Feng C., *Nucleic Acids Res.*, **36**, D907-D912 (2008).
- 4) (<http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/>)
- 5) Muegge I., Oloff S., *Drug Discov. Today Technol.*, **3**, 405-411 (2006).
- 6) Oprea T. I., Matter H., *Curr. Opin. Chem. Biol.*, **8**, 49-58 (2004).
- 7) MacBeath G., *Genome Biol.*, **2**, COMMENT2005 (2001).
- 8) Salemme F. R., *Pharmacogenomics.*, **3**, 257-267 (2003).
- 9) Vapnik V. N., "The Nature of Statistical Learning Theory," Springer, New York, 1995.
- 10) Cerezov V., *Science.*, **318**, 1258-1265 (2007).



構造活性相関部会・ニュースレター <1 April, 2008>

# SAR News No.14

## 「目次」

### //// Perspective/Retrospective ////

ケミカルゲノミクス情報を用いた新規リガンド探索手法

藪内 弘昭、奥野 恭史 … 2

### //// Cutting Edge ////

分子シミュレーションによる高活性リガンド探索の試み

温品 由美 … 7

セフェム系抗生物質のラット脳脊髄液移行性に関する QSAR 解析

吉田 麻衣、坂和 園子、阿部 覚、木原 勝、内藤 真策、山内 あい子 … 12

### //// Activities ////

#### <報告>

第 35 回構造活性相関シンポジウム報告

赤松 美紀 … 16

5th International symposium Causalities explored by indirect observation 参加報告

錦織 理華 … 17

#### <会告>

構造活性フォーラム 2008 「標的蛋白質志向のケミカルバイオロジーと構造活性相関」

… 20

第 36 回構造活性相関シンポジウム

… 21

第 8 回薬物の分子設計と開発に関する日中合同シンポジウム

… 22

2008 年度構造活性相関部会 SAR Promotion Award 受賞候補者募集

… 23

## //// Perspective/Retrospective ////

## ケミカルゲノミクス情報を用いた新規リガンド探索手法

京都大学薬学研究科 藪内 弘昭、奥野 恭史

## 1. はじめに

生体内には様々な分子が存在しているが、タンパク質は特定の分子を認識してその機能を発揮する。基礎生命科学から医薬品開発にわたる多くの分野では、この分子認識機構を利用して目的のタンパク質に結合して機能を制御するリガンド化合物の開発が長年進められてきた。しかし、化合物の化学構造は非常に多様性に富んでいる ( $10^{60}$ 種類以上と推定されている<sup>1)</sup>) ため、片っ端から合成、生物実験を行うことにも限界がある。

活性化化合物を効率よく迅速に探し出すために、計算機を用いた候補化合物の絞り込み手法「ヴァーチャル・スクリーニング (VS)」が開発されてきた。現在よく用いられている VS として、既知リガンドとの構造類似性に基づく「Ligand-based virtual screening (LBVS)」と標的タンパク質の立体構造に基づく「Structure-based virtual screening (SBVS)」がある<sup>2,3)</sup>。この二つの手法は、近年の情報技術の進歩と相まって、この 10-20 年で著しい発展を遂げ、ゆるぎない地位を確立した。しかしながら現在、VS のヒット確率は 1% - 10% もあれば良しとされており (例えば、LBVS では、既知活性化化合物の骨格構造に強く影響される嫌いがあるし、SBVS では、パラメータの恣意性、予測的中率の低さなどが指摘されている)、さらなる技術的な改良や革新的技術の開発が切望されていることは間違いない。

ここでは、LBVS や SBVS とは概念の異なる第三の VS として、著者らが開発している「ケミカルゲノミクスに基づく VS 手法 (Chemical Genomics-based virtual screening: CGBVS)」を紹介したい。

## 2. ケミカルゲノミクス

「ケミカルゲノミクス (Chemical Genomics)」とは、興味を持つ化合物が生物に与える影響についてゲノム規模で研究する学問である<sup>4,5)</sup>。この研究分野は、マイクロアレイやハイスループットスクリーニングなどの同時大量解析技術の革新に後押しされ、近年、注目を浴び始めている。それに伴い、化合物と遺伝子の関連性について、膨大な実験データが蓄積されている。そこで我々は、情報科学技術の一つであるパターン認識技術を用いて、タンパク質と化合物との結合情報 (ケミカルゲノミクス情報) から抽出したタンパク質のリガンド認識パターンに基づいて活性化化合物を効率的に発見する新たな VS、「CGBVS」を開発している。

## 3. ケミカルゲノミクス情報の収集・整備

CGBVS を実行するには、多くの「化合物-タンパク質間相互作用データ」が必要である。しかし残念ながら、それらのデータは基本的に製薬企業や各研究室の極秘資料となっている。そこで、我々はまず、文献などから利用可能な化合物-タンパク質間相互作用データを収集、整備することからスタートした。データ収集の対象とするタンパク質としては、創薬標的として最も注目されている遺伝子ファミリーのひとつ、G タンパク質共役型受容体 (GPCR) に的を絞った。

収集したデータは、公共データベース GLIDA (GPCR-Ligand Database)<sup>6,7)</sup> として現在 WEB で公開している (URL: <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/>)。GLIDA の面白い特徴は、相互作用データを可視化するために、化合物、タンパク質それぞれの類似性を反映させた 2 次元カラーマップを用いている点である。マッピングされた相互作用において近傍に位置するタンパク質や化合物は、化学者にとっても生物学者にとっても、新たな研究対象として興味深いところである。

なお、現在利用可能なケミカルゲノミクス情報の公共データベースとして、ほかに DrugBank<sup>8</sup> がある。このデータベースは、米国食品医薬品局 (FDA) で承認された薬物とその標的遺伝子に関する情報を提供している。

#### 4. CGBVS のためのアルゴリズム： サポートベクターマシンを用いた相互作用学習モデル

相互作用パターンの抽出、そしてリガンド予測の手段として、学習アルゴリズムであるサポートベクターマシン (SVM)<sup>9</sup> が利用可能である。SVM は、2クラス分類器の一種であり、与えられた2つのグループに属する特徴ベクトルを最大マージンで分離するような超平面を構築する。ここで、最大マージンとは、分離した超平面から各サンプル間までの最短距離を指す。

我々は、この SVM を用いて、化合物-タンパク質相互作用の有/無を判別する手法を開発した。その手法の流れを図1に示す。まず、収集した相互作用をベクトルとして表現するために、各化合物の化学構造、各タンパク質のアミノ酸配列について、様々な属性 (記述子と呼ぶ) を計算する。次に、正例 (相互作用する化合物-タンパク質ペア) および負例 (相互作用しない化合物-タンパク質ペア) に対応する記述子をそれぞれ組み合わせる特徴ベクトルを構成し、SVM を用いて学習モデルを構築する。このモデルが得られると、(未知の化合物-タンパク質ペアに相当する) 新しいベクトルが相互作用 有/無 のどちらのクラスに属するか、を予測することができる。

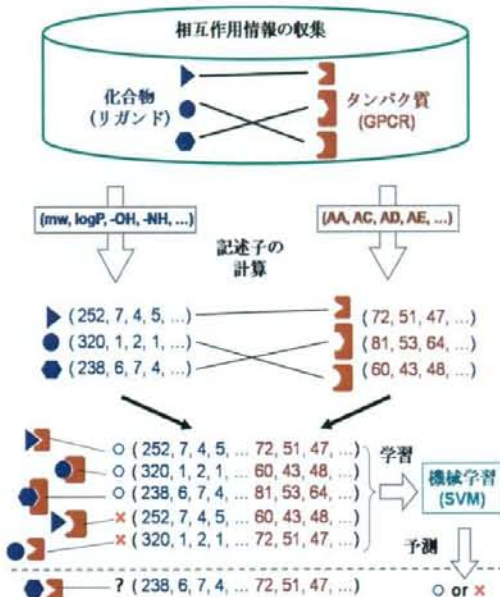


図1 CGBVS の流れ

#### 5. 交差検定法によるモデルの評価

既存の VS 手法との比較検討を行うため、今回開発した CGBVS と LBVS との予測性能を比較した。収集した化合物-GPCR 相互作用を用いて、負例を交換しながら 5 分割交差検定法 (fivefold cross-validation) を試行した。ただし、負例については、データが入手できなかったため、ここでは化合物-タンパク質ペアをランダムに組み合わせる生成した。交差検定の結果、最近傍法を用いた LBVS では  $84.4 \pm 0.3\%$ 、CGBVS では  $91.6 \pm 0.2\%$  の相互作用を正しく予測した。また、ROC 曲線からも、CGBVS の予測性能の高さが確認された (図2)。したがって、ケミカルゲノミクス情報の活用がリガンド予測性能の向上につながったといえる。なお、予測に要する計算的負荷が一般的な LBVS によるものほとんど変わらないことから、CGBVS は大規模スクリーニングに十分適用可能と考えられる。

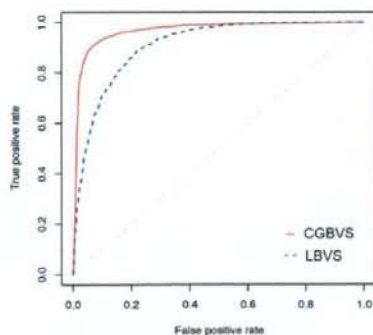


図2 ROC 曲線による相互作用予測性能の比較。赤い実線が CGBVS、青い破線が最近傍法を用いた LBVS を示す。

## 6. ヒト $\beta_2$ アドレナリン受容体リガンド予測への適用

我々が収集した化合物-GPCR 相互作用情報は、今までの研究により「強く結合する」と知られているもののみであり、その他の大部分の化合物-GPCR ペア間の相互作用の有無は不明である。我々の疑問は、CGBVS が期待に反して相互作用すると予測したペアは本当に相互作用しないのか、ということである。そこで、*in vitro* での実験により、予測された相互作用スコアと実際の相互作用の有無との関連性を確認した。

まず、ヒト $\beta_2$ アドレナリン受容体 ( $\beta_2$ AR) を標的 GPCR とし、構築した学習モデルを用いてリガンド予測を行った。リガンド予測の対象化合物は、他 GPCR のリガンドとして知られている 826 化合物とした。そして、CGBVS により予測された $\beta_2$ AR リガンド候補上位 50 の化合物のうち、文献・特許調査により 14 種の化合物について $\beta_2$ AR との相互作用に関する報告を確認した(図 3A 左)。さらに、残りの相互作用不明な化合物のうち、入手可能な 21 種類について *in vitro* 結合阻害実験を行ったところ、17 種類の化合物が相互作用 ( $10^{-5}\text{M} < \text{IC}_{50} < 10^{-3}\text{M}$ ) を示した(図 3A 右)。結合阻害実験のヒット率は 81% (17/21) にのぼり、ここにおいても高い予測的中率が示された。

一方、下位 50 化合物については、 $\beta_2$ AR リガンドとして報告されているものは文献および特許調査では確認されなかった(図 3B 左)。さらに、入手可能な 9 化合物についても結合阻害実験を行ったが、2 個の化合物が同程度の強さの相互作用を示したものの、残りの 7 化合物は相互作用を示さなかった(図 3B 右)。

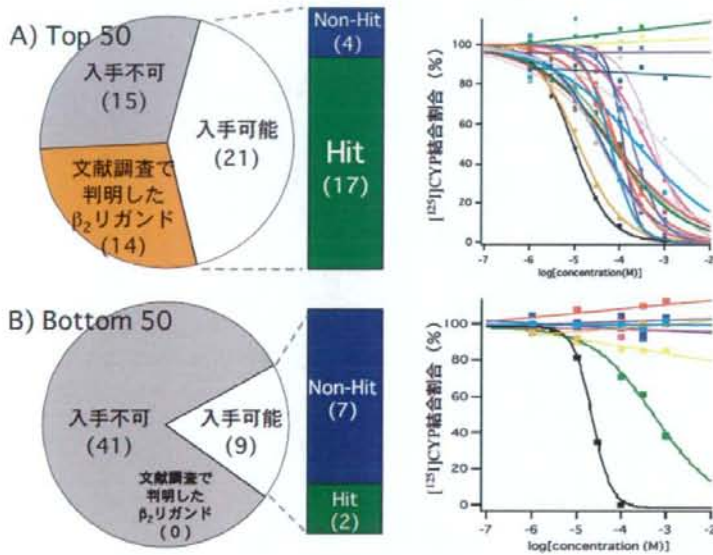


図3  $\beta_2$ AR リガンド予測結果の検証。予測スコア上位 50 (A) と下位 50 (B) の化合物に対する調査・実験結果を示す。

左: 文献調査および実験検証で判明した化合物の内訳。

右:  $[^{125}\text{I}]$ -シアノピンドロールに対する結合阻害曲線。縦軸が阻害の割合、横軸が各化合物の濃度を示す。

これらの予測結果を LBVS によるものと比較したのが図 4 である。結合阻害実験で相互作用を確認した化合物の半数近くは、LBVS ではスコアが低かった。実際に、これらの化合物は、典型的な $\beta_2$ AR 作動薬の構造(カテコラミン骨格、イソプレナリン誘導体)および $\beta_2$ AR 拮抗薬の構造(アリルアルキルアミン誘導体)とは異なる多様な骨格を持っており、化合物の構造類似性に基づく従来の方法では発見が困難なリガンド群である。また、これらの化合物の中には、ニューロペプチド受容体アンタゴニストなど、従来はペプチド受容体に作用する化合物として知られていたものも含まれており、遠縁にあたる $\beta_2$ AR とも相互作用することが今回の結合阻害実験により確認



された。以上から、ケミカルゲノミクス情報が、リガンド予測精度の向上のみならず、新規骨格を持つリガンドの検出にも有用であることが示唆された。

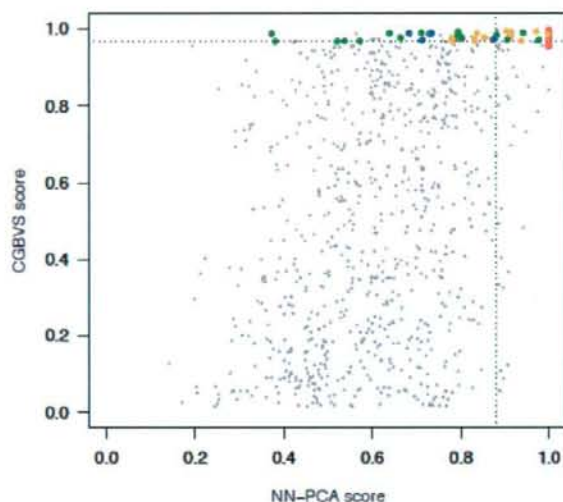


図4 CGBVS と LBVS による $\beta_2$ AR リガンド予測結果の比較

各点が化合物であり、縦軸は CGBVS、横軸は LBVS (主成分空間における最近傍法) による相互作用予測スコアを示している。文献調査および結合阻害実験の結果を以下の色分けにて表示した。

- 既知  $\beta_2$ AR リガンド
- 文献調査  $\beta_2$ AR リガンド
- 相互作用あり (結合阻害実験)
- 相互作用なし (結合阻害実験)

なお、図中の点線は、各手法において上位 50 位のスコアを示す。

## 7. おわりに

従来、生物活性を示す化合物の探索には、化合物の構造類似性がその指標とされてきた。しかし、化合物の構造類似性だけでタンパク質との多種多様な相互作用を見いだすには限界があり、新しい探索手法の開発が望まれていた。

我々は、「ケミカルゲノミクス情報の中から相互作用パターンを抽出する」という新しいアプローチを試み、(1) 化合物-GPCR 質間相互作用データベース GLIDA を開発し、(2) 学習アルゴリズムを用いた予測モデルにより、GPCR リガンドを精度よく予測できることを実証した。なお、 $\beta_2$ AR の立体構造が昨年決定された<sup>10</sup>ことを受けて、我々は SBVS との予測性能比較も行っており、その結果、CGBVS が SBVS よりも高精度で $\beta_2$ AR リガンドを予測することを確認している。

我々が開発している CGBVS は、標的タンパク質のリガンド情報が無いケースにも適用可能であり、実際に現在、オーファン GPCR を標的としたリガンド予測を行っている。タンパク質の立体構造情報を必要としない点は、詳細な 3D 構造が不明な GPCR にとって、魅力的なリガンド予測手法といえる。将来的には、結合阻害定数 ( $K_i$  値) など定量的な活性データから回帰分析を行うことにより、相互作用の強さを反映させたリガンド予測手法への拡張を行う予定である。

最後に、ケミカルゲノミクス情報を利用するという新規なアプローチは、GPCR に限らずすべてのタンパク質に適用可能であり、化合物-タンパク質相互作用の理解、ひいては創薬プロセスの効率化に繋がるのが期待される。

## 8. 参考文献

1. Dobson, C. M. Chemical space and biology. *Nature* **432**, 824–828 (2004).
2. Muegge, I. & Oloff, S. Advances in virtual screening. *Drug Discov. Today Technol.* **3**, 405–411 (2006).
3. Oprea, T. I. & Matter, H. Integrating virtual screening in lead discovery. *Curr. Opin. Chem. Biol.* **8**, 49–58 (2004).

4. MacBeath, G. Chemical genomics: what will it take and who gets to play? *Genome Biol.* **2**, COMMENT2005 (2001).
5. Salemme, F. R. Chemical genomics as an emerging paradigm for postgenomic drug discovery. *Pharmacogenomics* **3**, 257–267 (2003).
6. Okuno, Y., Yang, J., Taneishi, K., Yabuuchi, H. & Tsujimoto, G. GLIDA: GPCR-ligand database for chemical genomic drug discovery. *Nucleic Acids Res.* **34**, D673–D677 (2006).
7. Okuno, Y. *et al.* GLIDA: GPCR–ligand database for chemical genomics drug discovery–database and tools update. *Nucleic Acids Res.* **36**, D907–D912 (2008).
8. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).
9. Vapnik, V. N. *The Nature of Statistical Learning Theory* (Springer, New York, 1995).
10. Cerezov, V. *et al.* High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265 (2007).

## 膜タンパク質の結晶化技術の新展開及び創薬バリューチェーンの紹介

井上 豪,<sup>a,b</sup> 安達宏昭,<sup>a</sup> 村上 聡,<sup>c</sup> 高野和文,<sup>a</sup> 松村浩由,<sup>a</sup>  
 森 勇介,<sup>a</sup> 福西快文,<sup>d</sup> 中村春木,<sup>b,e</sup> 木下誉富,<sup>f</sup> 仲西 功,<sup>g</sup>  
 奥野恭史,<sup>g</sup> 南方聖司,<sup>a</sup> 下条真司,<sup>b,h</sup> 坂田恒昭<sup>b,h</sup>

## New Progressing Crystallization Technology of Membrane Protein and Introduction of Pharmaceutical Innovation Value Chain

Tsuyoshi INOUE,<sup>a,b</sup> Hiroaki ADACHI,<sup>a</sup> Satoshi MURAKAMI,<sup>c</sup> Kazufumi TAKANO,<sup>a</sup>  
 Hiroyoshi MATSUMURA,<sup>a</sup> Yusuke MORI,<sup>a</sup> Yoshifumi FUKUNISHI,<sup>d</sup> Haruki NAKAMURA,<sup>b,e</sup>  
 Takayoshi KINOSHITA,<sup>f</sup> Isao NAKANISHI,<sup>g</sup> Yasushi OKUNO,<sup>g</sup> Satoshi MINAKATA,<sup>a</sup>  
 Shinji SHIMOJO,<sup>b,h</sup> and Tsuneaki SAKATA<sup>b,h</sup>

<sup>a</sup>Graduate School of Engineering Osaka University, 2-1 Yamada-Oka, Suita City 565-0871, Japan, <sup>b</sup>NPO BioGrid Center Kansai, 1-4-2 Shinsenri-higashi-Machi, Toyonaka City 560-0082, Japan, <sup>c</sup>Department of Cell Membrane Biology, Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki City 567-0047, Japan, <sup>d</sup>National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6, Aomi, Koto-ku, Tokyo 135-0064, Japan, <sup>e</sup>Research Center of Structural and Functional Proteomics, Institute for Protein Research, Osaka University, 3-2 Yamada-Oka, Suita City 565-0871, Japan, <sup>f</sup>Department of Applied Biochemistry, Graduate School of Agriculture and Life Sciences, Osaka Prefecture University, 1-1 Gakuen-cho, Sakai City 599-8531, Japan, <sup>g</sup>Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshidashimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan, and <sup>h</sup>Cybermedia Center, Osaka University, 5-1 Mihogaoka, Ibaraki City 567-0047, Japan

(Received November 15, 2007)

We have recently established a Pharmaceutical Innovation Value Chain in collaboration with the SOSHO project (<http://www.so-sho.jp>) and BioGrid Project (<http://www.biogrid.jp/>) to accelerate new drug development. The SOSHO project provides novel crystallization technology with laser-irradiation and stirring growth methods, and the BioGrid Project is developing the software necessary for the *in silico* screening of promising drugs and the simulation of biological responses to proteins. In this paper, we report the recent research work on the crystallization of membrane proteins and the development of a method for *in silico* drug discovery.

**Key words**—*in silico* drug discovery; membrane protein; novel crystallization method; pharmaceutical innovation value chain

<sup>a</sup>大阪大学大学院工学研究科 (〒565-0871 吹田市山田丘 2-1), <sup>b</sup>NPO 法人バイオグリッドセンター関西 (〒560-0082 豊中市新千里東町 1-4-2), <sup>c</sup>大阪大学産業科学研究所 (〒567-0047 茨木市美穂ヶ丘 8-1), <sup>d</sup>産総研・生物情報解析センター (〒135-0064 東京都江東区青海 2-41-6), <sup>e</sup>大阪大学蛋白質研究所 (〒565-0871 吹田市山田丘 3-2), <sup>f</sup>大阪府立大学理学系研究科 (〒599-8531 堺市中区学園町 1-1), <sup>g</sup>京都大学大学院薬学研究科 (〒606-8501 京都市左京区吉田下阿達町 46-29), <sup>h</sup>大阪大学サイバーメディアセンター (〒567-0047 茨木市美穂ヶ丘 5-1)

\*e-mail: inouet@chem.eng.osaka-u.ac.jp

本総説は、日本薬学会第 127 年会シンポジウム S30 で発表したものを中心に記述したものである。

## 1. はじめに

20 世紀における科学の 3 大発見の 1 つにペニシリンの発見がある。本格的な医薬品の開発はこれを機に始まったと言われ、今では分子設計の手法も確立され、これを加速するためのコンビナトリアルケミストリー (CC) や、ロボティックスシステムを駆使したハイスループットスクリーニング (HTS) といった革新的技術も開発されている。一方、IT 技術を駆使して、バイオインフォマティクスによる新規ターゲットタンパク質の特定、リード化合物の創製や最適化を行い、研究開発のための費用の圧

縮と期間短縮が試みられている。また、わが国でもたんばく3000プロジェクトなどにより、ヒトやマウスも含めた疾患関連及び病因となるタンパク質の立体構造が次々と明らかとなっており、標的病因タンパク質の3次元立体構造に基づいた、より高効率な薬物設計手法である Structure-Based Drug Design (SBDD) の手法開発が重要性を増している。しかし、製薬会社における薬品開発では、標的タンパク質のうち約半分は GPCR などの膜タンパク質であり、その結晶化は容易ではなく、SBDD を実施することができない場合が多いことが知られている。

われわれは、立体構造を基にした医薬品の開発を加速するため、大阪大学工学研究科を中心に「創晶プロジェクト」(<http://www.so-sho.jp/>) を立ち上げ、レーザー照射や溶液かく押の技術を導入した革命的な結晶化手法の開発を行い、水溶性タンパク質を中心に、その結晶化にフェムト秒レーザーの照射が結晶化の確率向上と時間短縮に効果があり、溶液のかく押が高品質化に効果があることを示した。これが、常にサンプルの安定性が問題となっている膜タンパク質の結晶化に特に有効であることが最近判明し、いくつかの成功例が挙がってきた。一方、NPO 法人関西バイオグリッドを中心として、量子力学計算を取り入れた、オリジナリティーの高い、新しい計算方法で、酵素の活性部位に結合し得る有機低分子化合物をスクリーニングする方法が開発されている (<http://www.biogrid.jp/npo/>)。最近これら2つのグループを中心に異分野で連携して創薬研究を推進する「創薬バリューチェーン」を構築した。これには、化合物空間という、新たな阻害剤のインシリコ探索法を開発している研究者、製薬企業で豊富に経験を積んだ大学機関の研究者、さらには、長年複素環化合物の新規合成法の開拓を行ってきた研究グループも参加し、市販品を使った探索から、薬効評価、複合体の X 線構造解析、化合物の最適化、インシリコ ADME の計算を駆使した毒性や代謝評価など、薬剤候補化合物の設計提案が可能で一連の研究グループも参加している。ここでは、創薬バリューチェーンの中で結晶化及び構造解析を担当している創晶プロジェクトの技術を中心に、実際に結晶構造を解析し、インシリコでの阻害剤候補化合物の探索を行った実際例を紹介する。

## 2. 創晶プロジェクト

インシリコ創薬における阻害剤候補化合物の探索だけでなく、サブ  $\mu\text{M}$  オーダーで効果のある化合物からの SBDD を展開する際に、酵素の Native 及び阻害剤との複合体の構造情報があれば、圧倒的に薬剤開発が速く進む。したがって、NMR や X 線構造解析などの技術はインシリコ創薬にとって必須である。標的タンパク質の構造を解析する際に、NMR 法では標的タンパク質の分子量に限界があるが、X 線構造解析法にはなく、現在のところ最も強力な実験手法であるが、標的タンパク質の結晶を得る必要がある。

一般に、タンパク質を結晶化するには、その溶液に、沈殿剤と呼ばれる、塩、高分子試薬、有機溶媒を混合し、徐々にその濃度を上げることにより溶解度を上げて結晶核を発生させ、静置条件下で結晶成長させて目的タンパク質の結晶を得ている。タンパク質によって結晶化に最適の沈殿剤の種類やその濃度は大きく異なり、過去のタンパク質の結晶化条件をデータベース化されたものを参照するなど (<http://www.bmcd.nist.gov:8080/bmcd/bmcd.html>)、類縁のタンパク質の結晶化条件を応用したり、市販のスクリーニングキット [Crystal Screen and Crystal Screen 2 (Hampton Research), Wizard Screens I and II (Emerald BioStructures), Personal Structure Screens 1 and 2 (Molecular Dimensions) など] を用いて結晶化を試みるのが一般的である。様々な結晶化の方法が開発されているが、ハンギングドロップ蒸気拡散法、シッティングドロップ蒸気拡散法、サンドイッチ蒸気拡散法、静置パッチ法 (マイクロパッチ法)、自由界面拡散法及びゲルチューブ法 (GCB 法)、透析法、種結晶化法 (マクロシーディング法とマイクロシーディング法)、不溶性の液体を用いた浮遊法などが知られている。最近では、少量化やハイスループット化が進められている。



井上 豪

大阪大学大学院工学研究科応用化学専攻・准教授、工学博士、1989年大阪大学工学部卒業、1994年博士課程修了、1994年助手、1999年講師、2002年より現職。2001-2004 科学技術振興機構 さきがけ研究員兼務、大阪バイオサイエンス研究所客員研究員、創晶技術顧問など。[専門] 構造物理化学、構造生物学