Peptidome project: Analysis of peptides in pig and mouse brain for the Peptidome database

Naoto Minamino, Junko Isoyama-Tanaka, Takahiro Kihara,
Yasuko Matsui and Hiromiki Kuwahara
Department of Pharmacology, National Cardiovascular Center Research Institute,
Suita, Osaka 565-8565, Japan
(e-mail: minamino@ri.ncvc.go.jp)

Peptides play crucial roles in many physiological events as hormones, neurotransmitters and modulators, but the database construction for endogenous peptides has not been undertaken. This is mainly due to the facts that most of peptides in the cells and tissues are metabolically generated fragments of proteins, and that peptides are easily susceptible to proteolysis during extraction and purification. In 1999, we started the Peptidome project, which is aimed to comprehensively analyze peptides in the cell and tissue by minimizing degradation of peptides. As we usually use ion exchange (IEX) and reverse phase (RP) high performance liquid chromatography (HPLC) for the peptide separation, the degrees of hydrophobicity and net charge as well as molecular mass of the peptides are employed as major parameters to register the peptide information in addition to their names and sequences.

We set up the preparation procedures of the peptide fraction, and the peptides obtained by the gel filtration (Mr<6,000) were separated into about 5,000 fractions by 2D-HPLC (1st IEX and 2nd RP). Molecular masses and sequences of the peptides were determined by MALDI-TOF, ESI-Q-TOF and MALDI-TOF-TOF mass spectrometers. By these procedures, about 20,000 and 4,000 peptides are deduced to be detectable and identifiable, respectively, starting from about 1 g of pig and mouse brain tissue. Laboratory information management system and an automated 2D-HPLC system are also being developed in this project, which facilitates to accumulate and compare the peptide data in different targets and conditions.

We are applying this database to identification of new bioactive peptides. The elution positions of about 100 bioactive peptides have been determined on the 2D-HPLC, and screening of new bioactive peptides are on-going by using the brain peptide mixture separated under the standard conditions for the Peptidome analysis. The screening data are easily compared with the known peptide data and the Peptidome data on this common 2D-HPLC platform, which is expected to increase a chance for discovery of new bioactive peptides. Comparison of the Peptidome data between two different conditions, such as normal and diseased conditions, helps us to find diagnostic markers of the specific disease and to develop its new therapeutic strategy. Moreover, comparison of the accumulated sequence data in the Peptidome database with each other as well as those with their precursor sequences will provide us the processing and cleavage profiles for generating and degrading peptides and proteins, which will make us possible to understand the real status of the peptides in the cells, tissues, and living organisms.

Peptidome: Data acquisition, storage and presentation

Masaharu Isoyama
Protein Research Foundation, Toyonaka, Osaka 560-0082, Japan
(e-mail: isoyama@prf.or.jp)

The peptidome database is a fact database expressing whole image of peptides in a living body. In order to build the peptidome database, it is necessary to develop the database system with which various information of peptides acquired from experiments, such as the amount of existence, processing, modifications, biological activity, receptors and structures, can be contained and searched exactly. Moreover, it is necessary to also develop means to build a database efficiently from references or other databases. In consideration of using a scalable network, a WEB-Based-user interface was adopted and the database system by which researchers can register and use data was built.

The structural and functional information of peptides that are acquired from references were examined in detail, items of peptidome database were selected, and the core table was created.

Based on reference information, construction of the database about information of modifications of peptides and proteins was started (PRF/MODDB). The system for data registration for researchers to the PRF/MODDB via the Internet was developed.

In order to express the structure of peptides exactly, the extended amino acid code table was created. The sequence database with modification data (Expanded Amino Acids Sequence Database: PRF/EXSEQ) was created.

Peptidomics-based discovery of peptide tumor markers

Kazuki Sasaki
Growth Factor Division, National Cancer Center Research Institute,
Chuo-ku, Tokyo 104-0045, Japan.
(e-mail: ksasaki@ncc.go.jp)

The tumor marker is a substance whose determination in body fluids may provide information in cancer diagnosis. A secretory peptide from tumor cells should serve as a useful tumor marker so long as it is measurable in blood. Recent advances in mass spectrometry techniques allow us to analyze and identify endogenous peptides much easier than ever before. To identify potential peptide tumor markers, we have started with the mass analysis of unfractionated peptides found in serum-free culture supernatant from cultured cells, using surface-enhanced laser desorption/ionization (SELDI) mass spectrometry. We have examined culture supernatants from more than 70 cultured cells (mostly cell lines) to discover peptides that seem to occur specifically in pancreatic cancer cells but not in non-malignant pancreatic duct cells or other types of cancer cells. Antibody studies on one of the candidate peptides, a 29-residue carboxyl-terminal fragment of the putative tumor-suppressor protein DMBT1, verified the SELDI data and further revealed its presence in the pancreatic juice as well as the tumor tissues from pancreatic cancer. Another approach to identify potential markers is a peptidomics-based analysis of endogenous secretory peptides from cultured pancreatic cancer cells. This study is based on our observation that tumor cells in culture tend to secrete or shed peptides in greater amounts than non-malignant cells. A developing list of identified peptides include peptide fragments from some bioactive peptides or proteins known to be over-expressed in pancreatic cancer. Along this line, we could find a peptide elevated in plasma from patients with advanced pancreatic cancer. These findings point to the advantage of cultured cells over clinical samples when analyzing and identifying secretory peptides of interest.

Peptidomics-based discovery of novel neuropeptides

Per E. Andrén
Laboratory for Biological and Medical Mass Spectrometry, Uppsala University,
SE-75123 Uppsala, and Amersham Biosciences, Uppsala, Sweden
(e-mail: per.andren@bmms.uu.se)

Modern proteomic methodologies have significantly improved the possibilities of large-scale identification of proteins. However, these methodologies are limited by their inability to reliably detect endogenously expressed peptides. We describe an approach of combining sample preparation, comprising focused microwave irradiation and mass spectrometric peptide profiling that has enabled us to simultaneously detect more than 550 endogenous neuropeptides in 1mg of hypothalamic extracts. Automatic switching tandem mass spectrometry and amino acid sequence determination of the peptides showed that they consist of both novel and previously described neuropeptides. The methodology includes virtual visualization of the peptides as two- and three-dimensional image maps. In addition, several novel and known post-translational modifications of the neuropeptides were identified. The peptidomic approach proved to be a powerful method for investigating endogenous peptides and their post-translational modifications in complex tissues such as the brain. Further, we have employed this approach to an investigation of the endogenous neuropeptide content of experimental Parkinson's disease (PD). PD is a common neurodegenerative disorder characterized by the degeneration of dopamine producing nigrostriatal neurons. The neurotoxin MPTP induces such neuronal degradation and produces symptoms analogous to PD in mouse animal models. We have studied the neuropeptide expression differences between MPTP lesioned mice (levodopa or saline treated) and untreated saline injected controls. The relative concentrations of the peptides were compared and the differences were analyzed and identified by MS-MS. The comparisons of neuropeptides revealed several differences in relative abundance both between the different animal groups and the different brain regions. The compared peptides comprise precursor families of both known secretory pathway proteins, including granin-like neuroendocrine peptide, CART, POMC, prosomatostatin, proenkephalin, and precursors with neuropeptides identified as novel. Several novel post-translational modifications of both novel and known peptides were identified.

Systematic identification of peptide signaling molecules in *Hydra*
-Recent progress in functional analysis of neuropeptides

Toshitaka Fujisawa
Department of Developmental Genetics, National Institute of Genetics,
Mishima, Shizuoka 411-8540, Japan
(e-mail: tfujisaw@lab.nig.ac.jp)

Hydra is a member of cnidarians whose ancestor first acquired nervous system in evolution. It has been generally believed that cnidarians have a simple net-like nervous system utilizing primarily peptides as transmitters and therefore exhibit rather poor behaviors. Thus, assuming that *Hydra* contains rather a limited number of signaling peptides, we started to identify systematically all the peptide molecules (MW<5,000 Da) in the organism ("Hydra Peptide Project"). Contrary to our expectation, however, it is estimated that *Hydra* contains a few hundreds of peptide signaling molecules and that a half of them are neuropeptides and the rest epitheliopeptides (peptides derived from epithelial cells). Recently we also initiated "*Hydra* EST Project". Combination of *Hydra* Peptide Project and EST Project has allowed us efficient and rapid identification of peptides as well as their genes. So far we have identified 30 neuropeptides and 10 neuropeptide genes and analyzed their localization/expression patterns along the body. The results showed that *Hydra* nervous system is composed of several distinct subpopulations that divide the hydra body into several compartments with a sharp border between them. *Hydra* also exhibits complex responses to external cues such as prey, oxygen, gravity etc. When the animal captures prey, a series of movement is initiated: feeding response, digestive movement and pumping of gastric fluid to the periphery. We have found two neuropeptides that antagonistically control tentacle movement. One is a neuro-modulator, Hym-357 that induces contraction and the other is a neurotransmitter, Hym-248 that does opposite. The effect of these peptides on feeding response generated by adding glutathione to the culture solution was analyzed. The peptides modulated the response, suggesting their involvement in feeding behavior. Since these peptides are produced by different subsets of neuron population, there must be some neural circuit that regulates the concerted secretion of these neuropeptides.

Peptidomics in insects: Peptide identification in the post-genome era

Liliane Schoofs and Geert Baggerman
Laboratory for Developmental Physiology, Genomics and Proteomics,
K.U.Leuven, B-3000 Leuven, Belgium
(e-mail: liliane.schoofs@bio.kuleuven.ac.be)

Bioactive peptides are of considerable biological, medical and industrial interest. In pre-genomic times, peptides had to be purified from tissue extracts taken their activity as a starting point. Their purification often proved heroic, requiring kg or even ton quantities of starting material and many painstaking chromatography rounds. Only after purification to homogeneity of a sufficient amount of peptide, its sequence could be determined (usually by Edman degradation based sequencing). The completion of numerous genome projects has opened new opportunities for rapid identification and functional analysis of novel peptides. Peptides can in principle be predicted from genomic sequences, but this is quite a challenge as for peptides: one not only needs to predict the precursor protein, but in addition the (often times) multiple cleavage sites as well as post-translation modifications. Peptide processing is also regulated in space and time. Different tissues have the same genome but will express different peptides.

During the last few years, the genome projects have been the driving force for drastic improvements in chromatography and mass spectrometry (MS). Modern mass spectrometers, with high resolution, sensitivity and speed permit the scanning of tissue extracts for virtually all peptide masses present (single MS) and enable sequence determination of peptides as well (tandem-MS). In addition, improvements in informatics have produced specialised software allowing MS-data processing and comparison with database sequence information.

At the 15th International MS Conference (Barcelone, Spain) in August 2000, we introduced –by analogy with proteomics- the concept of peptidomics i.e. the characterisation of the peptidome or all peptides (with their post-translational modifications) expressed in a cell, tissue, or organism. This technology shortens from years to days the time required for identifying new peptides and promises a flood of new peptide sequences. In 2001, we published the first data (1) and it appeared that the time of peptidomics had come because within months the same term was used (presumably independently) by other groups (2,3). In the mean time, 'peptidomics' has been introduced into the 'omics' hall of fame http://www.genomicglossaries.com/content/omes.asp.

As in proteomics, the core identification tool in peptidomics is mass spectrometry; mostly MALDI-TOF and Electrospray ionisation (ESI) tandem MS (4). ESI has the major advantage that the electrospray process transfers ions from solution directly into the gas phase enabling on-line coupling to LC-systems. This means that peptides are sequenced as they elute from the LC-column. Using LC-MSMS, we analyzed the peptidome of the larval *Drosophila* central nervous system (CNS) (5). We were able to provide the amino acid sequences of 28 neuropeptides using an extract of only 50 larval *Drosophila* CNS. This is an impressive hit rate considering the fact that pre-genomic peptide identification required ± 400.000 fruit flies. Many of the identified *Drosophila* peptides had already

been predicted, although several incorrectly. Only 11 peptides had been purified before and 8 others are entirely novel and are encoded in five unknown genes. This expression profiling study also opens perspectives for other eukaryotic model systems, for which genome projects are completed or in progress. We also successfully characterized the neuropeptidome of other arthropods such as *Locusta migratoria* and *Cancer borealis* (6,7), although in these cases, de novo sequencing remains a challenge due to lack of genomic sequence information.

References: (1) Clynen E. et al., (2001), Eur. J. Biochem. 268, 1929-1939; (2) Verhaert P. et al., (2001), Proteomics 1, 118-131; (3) Schulz-Knappe P. et al., (2001), Com. Chem. High Throughput Screen. 4, 207-217; (4) Clynen et al., (2003) Gen Comp Endocrinol. 132:1-9. (5) Baggerman G. et al., (2002) J. Biol. Chem. 277:40368-40374; (6) Clynen E. et al. (2002) Biochem. Biophys. Res. Commun. 300, 422-428. (7) Huybrechts J. et al., (2002) Biochem. Biophys. Res. Commun 308, 535-544.

Structure and action of the cardiovascular regulatory neuropeptides in *Aplysia*

Yasuo Furukawa, Kosei Sasaki, Fumihiro Morishita and Osamu Matsushima
Department of Biological Science, Graduate School of Science, Hiroshima University,
Kagamiyama, Higashi-Hiroshima 739-8526, Japan
(e-mail: yasfuru@hiroshima-u.ac.jp)

Neuropeptides are one of the most ubiquitous signaling molecules in the animal kingdom. Fairly large numbers of different neuropeptides are identified in both vertebrate and invertebrate nervous systems, and the number is still increasing. In the last few years, we isolated bioactive peptides from the central and peripheral nervous systems of a marine mollusc, *Aplysia*. To delineate physiological roles of identified peptides, we used immunohistochemistry and examined the distribution of neurons and neuronal processes containing the peptides in *Aplysia*. Among the several systems in *Aplysia*, the cardiovascular system was found to have rich peptidergic innervation. Here, we show the structure and action of the neuropeptides that are potentially acting as the cardiovascular regulators in *Aplysia*. In the cardiovascular system, we concentrate on three targets: the heart, the anterior aorta, and the vasoconstrictor muscle of the abdominal aorta. The anterior aorta is the major artery supplying the blood into the head region. The contraction of the vasoconstrictor muscle restricts the blood flow into the abdominal aorta, and results in the increase of the flow volume into the anterior aorta.

NdWFamide is a cardioactive tripeptide containing D-tryptophan, and potently augments the beating amplitude of the heart. NdWFamide also evokes longitudinal shortening of the anterior aorta and enhances the nerve-evoked contraction of the vasoconstrictor muscle. Enterins are nona/decapeptides, and most of them share a C-terminal sequence of HSFVamide. Enterins inhibit the contractile activity of the anterior aorta as well as the vasoconstrictor muscle. Although the contraction of the vasoconstrictor muscle is strictly brought about by the activity of the motoneurons, the extent of contraction of the muscle was found to be modulated by peptides including NdWFamide, enterins, and *Aplysia Mytilus* inhibitory peptide-related peptides (AMRPs). AMRPs are hexapeptides (XXPXFXamide), and similar peptides are quite common in molluscs. Excitatory junctional potentials of the vasoconstrictor muscle were enhanced by NdWFamide, and depressed by enterins and AMRPs, resulting in either stronger or weaker constriction of the abdominal aorta. These results suggest that the circulation in *Aplysia* is finely tuned by the action of distinct neuropeptides at several levels.

Supported by Grant-in-Aid from the Ministry of Education, Science, Sports and Culture of Japan, and the SUNBOR GRANT from the Suntory Institute for Bioorganic Research.

Identification of novel RFamide peptide genes by utilizing databases

Shuji Hinuma
Discovery Research Laboratories, Pharmaceutical Research Division, Takeda Chemical Industries Ltd., Tsukuba, Ibaraki 300-4293, Japan
(e-mail: Hinuma_Shuji@takeda.co.jp)

Recent advances in cDNA and genomic DNA analysis have revealed that hundreds of G protein-coupled receptor (GPCR) genes exist on the human genome. Those for which ligands have not yet been identified are referred to as orphan GPCRs. GPCRs play pivotal roles in cell-to-cell communication and in the regulation of cell functions. Furthermore, as GPCRs are considered to be some of the most important drug target molecules, identifying the ligands of orphan GPCRs provides opportunities for developing novel drugs. We have previously established a widely applicable method to identify ligands for orphan GPCRs through monitoring specific signal transductions in cells expressing orphan GPCRs. Using this method, we have succeeded in identifying various orphan GPCR ligands.

In our search for orphan GPCR ligands, we first identified prolactin-releasing peptide (PrRP) as a ligand for orphan GPCR hGR3 (1). As we noticed that PrRP belonged to the arginine-phenylalanine-amide (RFamide) peptide family, whose members have a common RFamide structure at their C-termini, we hypothesized that other RFamide peptides would exist in mammals and that these should also be ligands for orphan GPCRs. We therefore utilized databases to search for novel RFamide peptide genes. Through this search, we identified a gene encoding RFamide-related peptide (RFRP) (2). We subsequently found that an orphan GPCR, OT7T022, functioned as a receptor for RFRP. RFRP was closely related to neuropeptide FF (NPFF) in structure. Similar to NPFF, RFRP showed activity to modulate pain in mice. Furthermore, we recently identified another gene encoding pyroglutamylated RFamide peptide (QRFP), and subsequently demonstrated that an orphan GPCR, AQ27, functioned as a receptor for QRFP (3). The administration of QRFP into rats induced aldosteron secretion. Our results indicate that a variety of RFamide peptides exist and function widely not only in invertebrates, but also in mammals.

References:
(1) Hinuma, S., *et al.* (1998) *Nature* **393**, 272-276
(2) Hinuma, S., *et al.* (2000) *Nat. Cell Biol.***10**, 703-708
(3) Fukusumi, S., *et al.* (2003) *J. Biol. Chem.* **278**, 46387-46395

Receptor-Galpha fusion proteins as a peptide-screening system

Tatsuya Haga[1], Shigeki Takeda[2], Hinako Suga[3] and Naoto Minamino[4]
[1]Institute for Biomolecular Science, Gakushuin University, Toshima-ku, Tokyo 171-8588,
[2]Faculty of Engineering, Gunma University, Kiryu, Gunma 376-8515, [3] Institute Molecular Cellular Regulation, Gunma University, Maebashi, Gunma 371-8512, [4]National Cardiovascular Center Research Institute, Suita, Osaka 565-8565, Japan
(e-mail: tatsuya.haga@gakushuin.ac.jp)

We have prepared fusion proteins of G protein-coupled receptors (GPCRs) and G protein alpha subunits and examined if the fusion proteins could be used as a ligand-screening system. Fusion proteins of muscarinic acetylcholine receptor M_2 subtype and $G\alpha_{i2}$ and of $beta_2$ adrenergic receptor and $G\alpha_s$ were expressed in insect cells (Sf9) using baculovirus. The membrane preparations expressing these fusion proteins were shown to have low, intermediate and high affinity for GDP in the presence of full-, partial-agonists and antagonists, respectively, and then these ligands could be easily discriminated by $[^{35}S]GTP\gamma S$ binding assay in the presence of GDP. On the other hand, agonists and antagonisits could not be clearly discriminated for Sf9 membrane preparations expressing fusion proteins of muscarinic M_1, M_3 and M_5 receptors with $G\alpha_{11}$, and fusion proteins of $beta_2$ adrenergic, M_2 and M_1 muscarinic receptors with promiscuous $G\alpha_{16}$. Fusion proteins with $G\alpha_{16}$ of CX_3C chemokine, urotensin II, M_1, and M_2 muscarinic receptors, however, could provide efficient screening systems when they were expressed in cultured CHO cells and the ligand-induced Ca^{2+} increase or prostaglandin E_2 production were used as monitors. These results indicate that fusion proteins with $G\alpha_i$, $G\alpha_s$ or $G\alpha_{16}$ of orphan GPCRs could be used to screen endogenous ligands including peptides. Using nociceptin receptor-$G\alpha_{i2}$ fusion proteins as a model system, we could show that the stimulating activity of $[^{35}S]GTP\gamma S$ binding could be detected in a fraction of reverse-phase chromatography of the extract derived from 2 g of porcine brain, and that authentic nociceptin is eluted at the same fraction.

We have identified GPCRs from the human genome sequence by taking advantage of the fact that all known GPCRs have seven trans-membrane segments and that most GPCR genes do not have an intron in their open reading frames. We have detected 50 novel GPCR genes, and prepared fusion proteins with $G\alpha_{i1}$, $G\alpha_s$, $G\alpha_{16}$ of 27, 10 and 11 GPCRs among 50 novel GPCRs, respectively. Sf9 membrane preparations expressing these GPCR-$G\alpha_{i1}$ and -$G\alpha_s$ fusion proteins and CHO cells expressing GPCR-$G\alpha_{16}$ fusion proteins were used to screen their ligands from the ligand libraries including a peptide library and peptide fractions of the porcine brain extracts. So far we have detected a number of surrogate agonists, an endogenous ligand that is 5-oxoeicosatetraenoic acid, but no peptide ligand yet.

Peptidomics from amphibian peptides to mammalian peptides

Hubert Vaudry[1], Cynthia Dujardin[1], Hervé Tostivint[1], Jérôme Leprince[1], Marie-Christine Tonon[1], J. Michael Conlon[2], Youssef Anouar[1], Isabelle Lihrmann[1] and Nicolas Chartrel[1]
[1]European Institute for Peptide Research (IFRMP 23), Laboratory of Cellular and Molecular Neuroendocrinology, INSERM U 413, UA CNRS, University of Rouen, 76821 Mont-Saint-Aignan Cedex, France,
[2]Department of Biochemistry, Faculty of Medicine and Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates
(e-mail: hubert.vaudry@univ-rouen.fr)

Studies conducted in frogs have shown that the brain of these animals can be a valuable source for the discovery of novel neuropeptides: *(i)* the concentration of neuropeptides is much higher in the brain of amphibians than in that of mammals, and *(ii)* the sequence of biologically active peptides has been strongly preserved during evolution from amphibians to mammals. Novel peptides isolated from the amphibian brain can thus be readily used to characterize their counterparts (orthologues) in mammals. Here, we describe the characterization of three biologically active peptides from the frog brain that have subsequently led to the discovery of three novel neuropeptides in human, and to the identification of their cognate receptors.

We have isolated two forms of somatostatin from the brain of the frog *Rana ridibunda*, *i.e.* somatostatin 14 (S14) that exhibits the same sequence as human S14 and a somatostatin variant that exhibits two amino acid substitutions as compared to S14, [Pro^2,Met^{13}]S14. Molecular cloning of the cDNAs has shown that the two peptides derive from two distinct genes. In frog, the S14 gene is widely expressed in the CNS and the gastrointestinal tract while the expression of the gene encoding [Pro^2,Met^{13}]S14 seems to be restricted to the brain and the pars intermedia of the pituitary. The mammalian orthologue of this somatostatin variant, that also exhibits the $Gly^2 \rightarrow Pro^2$ substitution, has been characterized and called cortistatin (CST). Like the amphibian [Pro^2,Met^{13}]S14 variant, CST is exclusively expressed in the brain. It has been recently reported that the orphan receptor MrgX2 is selectively activated by CST.

Urotensin II (UII) is a cyclic peptide initially isolated from the fish urophysis. UII has been subsequently identified in a frog brain extract and the frog UII cDNA has been characterized. In amphibians, the UII peptide is located in motoneurons of the brainstem and spinal cord. The UII cDNA has been cloned in human and it has been found that, in mammals as in amphibians, UII is primarily expressed in motoneurons. Finally, it has been shown that UII is the endogenous ligand of the orphan receptor GPR14.

A novel neuropeptide of 26 amino acids, that exhibits the Arg-Phe-NH_2 motif at its C-terminus, has been recently isolated from the frog brain and termed 26RFa. The cDNA encoding the 26RFa precursor has been cloned in human and rat. In the rat brain, the 26RFa

gene is exclusively expressed in the ventromedial hypothalamic nucleus and in the lateral hypothalamic area, two diencephalic regions that play a crucial role in the control of food intake. Intracerebroventricular injection of 26RFa in mice produced a dose-dependent stimulation of food consumption, indicating that 26RFa is actually involved in the control of feeding behavior. It has also been shown that 26RFa and its N-terminally elongated form 43RFa are the endogenous ligands of the orphan receptor GPR103.

In conclusion, we believe that the identification of CST, UII and 26RFa, as well as the characterization of their receptors, validate the use of the frog brain for the discovery of novel biologically active peptides.

Supported by grants from INSERM (U 413) and the Conseil Régional de Haute-Normandie.

Identification and functional characterization of bioactive peptides from amphibian and reptile

Tai Kubo[1], Suzuko Kobayashi[1], Gandhi Radis Baptista[1,2], Xianyu Yang[1],
Hidetoshi Inagaki, Yuka Nakajima[1], Mami Takeda[1], Hikaru Urano[1],
Tadashi, Kimura[1], Alvaro Prieto-da-Silva[2] and Tetsuo Yamane[2]
[1]Neuroscience Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8566, Japan,
[2]Center for Applied Toxinology, Institute Butantan, Sao Paulo, Brazil
(e-mail: tai.kubo@aist.go.jp)

Bioactive material such as biogenic amines, steroids and peptides are found in amphibian skin and its secretion. Snake venoms are also unique mixture with reference to their biochemical and pharmacological properties. Bioactive peptides so far identified from frogs have counterparts in the gastrointestinal tract and/or the nervous system of higher vertebrates. Toward intensive peptidome analysis of amphibian skin and snake venom glands, we took advantage of molecular genetic approach by which a limited amount of bioresources is converted to a cDNA library. We constructed cDNA libraries which may represent nonredundant and full-length cDNAs. The library could be amplified, if necessary, and be subjected to the following structure- or activity-based screenings.

We first prepared a cDNA library from dorsal skin of Japanese toads. Among the cDNAs, four clones coding for a novel and characteristic cysteine framework, called the four disulfides core (4DC) or WAP motif, were analyzed. They have common structural features; a signal sequence followed by one to five 4DC motif(s), in which eight cysteine residues are strictly conserved at the corresponding positions while the interval sequences are diverged. The recombinant peptides showed specific inhibition to pancreatic kallikrein among the serine proteinases.

A bite by South American coral snake *Micrurus corallinus*, one of the Elapidae species, often causes serious neurogenic symptoms, such as convulsions and paralysis. To elucidate the molecular and physiological basis of the mode of toxin actions, we prepared a cDNA library from the venom glands and identified eleven new neurotoxin-like peptides so far, MicTx1 to MicTx11. Amino acid sequences of the peptides are diverged, while the cysteine frameworks similar to that of the α-bungarotoxin (BTx) are strictly conserved; suggesting that the peptides are the family of three-finger type neurotoxins. Physiological and biochemical functions of the MicTxs were examined by using recombinant peptides produced in *E. coli*, yeast and/or *Xenopus* oocytes. Interaction between the MicTx and acetylcholine binding protein (AChBP) was monitored on a biosensor chip. All the peptides showed differential kinetics of interaction with the AChBP. Furthermore, the MicTx peptides blocked ACh-induced currents of the nicotinic AChRs α7 and/or $\alpha_2\beta\delta\varepsilon$. We also observed that the MicTxs modulate the activities of serine proteases such as trypsin and elastase. By

calculating divergence degrees in the nucleotide sequences, it is suggested that accelerated evolution of ancestral gene(s) for the toxin may facilitate to produce divergent sequences and thus resulting in functional variations in the three-fingered toxin family.

Calcitonin receptor-stimulating peptides: New members of the CGRP family with unique biological activity

Takeshi Katafuchi and Naoto Minamino
National Cardiovascular Center Research Institute, Suita, Osaka 565-8565, Japan
(e-mail: katafuch@ri.ncvc.go.jp)

The first calcitonin receptor-stimulating peptide (CRSP-1) was purified from acid extracts of porcine brain by monitoring cAMP production in LLC-PK_1 cells. The brain extracts were successively separated by gel filtration, ion exchange chromatography and reverse phase HPLC, and the biological activity of each fraction was evaluated at every purification step. After separation by cation-exchange and reverse phase chromatography, we found the biological activity at the elution fraction distinct from those of known peptides with the cAMP-producing activity on the 2-dimensional chromatogram similar to that of the Peptidome analysis. We isolated CRSP-1 from this biologically active fraction, and subsequent molecular cloning made us possible to identify two other CRSPs in pig brain.

Amino acid and cDNA sequence analysis demonstrated that CRSP-1 is a new member of the CGRP family having high sequence identity with calcitonin gene-related peptides (CGRPs), but this peptide did not stimulate an endogenous or a recombinant CGRP receptor. Although CRSP-1 showed low structural similarity with calcitonin (CT), this peptide stimulated CT receptor at a potency more than 100-fold stronger than that of CT in the COS-7 cell expression system. Northern blot analysis and radioimmunoassay showed that CRSP-1 is expressed and produced mainly in the central nervous system (CNS) and thyroid gland. Although CT has not been detected in the brain, the CT receptor is expressed in the CNS, and intracerebroventricular injection of CT is known to reduce food intake and induce analgesia. These results suggest that CRSP-1 is an endogenous ligand for the central CT receptor.

To identify peptides that have structural similarity with CRSP-1, we screened a porcine hypothalamus cDNA library with a CRSP-1 cDNA probe under the low-stringent conditions and identified two cDNAs each encoding a CRSP-1-like peptide, CRSP-2 or CRSP-3. Although CRSP-2 and CRSP-3 show high sequence identity with that of CGRP and belong to the CGRP family, they did not stimulate CGRP, AM or CT receptor. RT-PCR analysis demonstrated that mRNAs of these peptides were expressed mainly in the CNS and thyroid gland, and tissue expression profiles were similar to those of CRSP-1 and CGRP. This evidence suggests that CRSP-2 and CRSP-3 participate in the physiological regulation of the CNS as well as the peripheral organs, and that their receptor-signal transduction systems are distinct from that of CRSP-1, CT, CGRP, or AM.

Although three CRSPs show high sequence identity with each other and with CGRP, CRSP-1 elicits its activity through CT receptor, while CRSP-2 and CRSP-3 are found to have unique properties different from those of CRSP-1, CGRP and the other members of the CGRP superfamily.

Ghrelin, a new regulator in the endocrine, central nervous and cardiovascular systems

Kenji Kangawa[1,2], Hiroshi Hosoda[1,2] and Masayasu Kojima[3]
[1]Department of Biochemistry, National Cardiovascular Center Research Institute, Suita, Osaka 565-8565,
[2]Translational Research Center, Kyoto University Hospital, Kyoto 606-8507,
[3]Institute of Life Science, Kurume University, Kurume, Fukuoka 839-0861, Japan
(e-mail: kangawa@ri.ncvc.go.jp)

Small synthetic molecules termed growth hormone secretagogues (GHSs) act through an orphan G-protein coupled receptor, GHS-R, to stimulate growth hormone (GH) release from pituitary. To search for the natural ligand for GHS-R, we constructed a stable CHO cell line expressing rat GHS-R for monitoring intracellular calcium concentration $[Ca^{2+}]_i$ change induced by rat tissue extracts. Using this cell line, we purified a peptide that potently activated GHS-R in 1999.

The purified peptide was consisted of 28 amino acids, in which the Ser-3 residue was n-octanoylated. Surprisingly, this n-octanoyl modification at Ser-3 was essential to the activity, since des-acyl-peptide could not induce $[Ca^{2+}]_i$ change. We named the GH-releasing peptide "ghrelin" ("ghre" is the Proto-Indo-European root of the word "grow"). We also purified peptides and cloned cDNAs of mammalian and non- mammalian ghrelins. N-terminal 10 amino acids are highly homologous, which supports the importance of octanoyl modification for ghrelin's activity.

Ghrelin induced GH release in a dose-dependent manner both in vitro and in vivo. Ghrelin immunoreactive cells in the stomach were found in X/A-like cells, an endocrine cell in stomach. In rat brain, ghrelin immunoreactive neurons were found to be localized in the hypothalamic arcuate nucleus, a region where regulates GH release and food intake. In fact, ICV injection of ghrelin induced potent appetite-stimulating effects in rat. In human studies, IV injection of ghrelin elicited a potent GH release and had beneficial hemodynamic effects via decreasing mean arterial pressure and increasing cardiac output without an increase in heart rate.

Thus, the occurrence of ghrelin in both stomach and hypothalamus will give a new dimension to the regulation of GH release and feeding. Further, GHS-R is widely expressed in peripheral tissues. Ghrelin also thus have multifaceted roles in the cardiovascular system and metabolism.

Feeding regulation by recently discovered peptides

Masamitsu Nakazato[1], Yukari Date[1], Naoto Minamino[2] and Kenji Kangawa[3]
[1]Third Department of Medicine, Miyazaki Medical College,
University of Miyazaki, Kiyotake, Miyazaki 889-1692,
[2]Department of Pharmacology and [3]Department of Biochemistry, National Cardiovascular Center Research Institute, Suita, Osaka 565-8565, Japan
(e-mail: nakazato@med.miyazaki-u.ac.jp)

Energy balance is controlled by the complicated and minute interactions of substances to stimulate or suppress food intake and energy expenditure. The molecular mechanisms of energy balance are coming to light by the recent robust progresses in the molecular biology and neuroscience. Hypothalamus, the center of energy homeostasis, receives information related to satiety and fast from the body and other brain regions, integrate them, and mediate interactions with efferent pathways. We have identified peptide ligands for orphan G-protein coupled receptors and studied their physiological and clinical implications in the regulation of energy homeostasis. I will focus on recently discovered peptides, ghrelin, neuropepitde W and calcitonin receptor-stimulating peptides. These findings will provide a clue to our better understanding of the molecular etiology of body weight control and the pathogenesis of obesity and anorexia in humans.

Integration of peptide database to multiple databases, and touch and feel system for peptide representation

Hiroshi Mizushima[1], Hiroshi Nagata[2] and Hiroshi Tanaka[3]
[1]National Cancer Center Research Institute, Center for Medical Genomics, Chuo-ku, Tokyo 104-0045, [2]KDDI Research Institute, Kamifukuoka, Saitama 356-8502, [3]Tokyo Medical and Dental University, Department of Bioinformatics, Bunkyo-ku, Tokyo 113-8510, Japan
(e-mail: hmizushi@ncc.go.jp)

There are hundreds of Molecular Biological Databases on the Internet, and they are managed and administrated independently. Major databases such as Genbank, EMBL, SwissProt, which are commonly used as a standard database, also are updated very frequently, and sometimes changes its structure. We made a flat file database linking system using SRS, which was developed by EBI in Europe and maintained by LION bioscience. This system was good to create and maintain the links between flat file databases, which is the common type of database structure in this field.

To use this flat file database relational system with the new “peptide database”, we first make the flat file database or XML format database, and use it as a source. After indexing the information automatically, the links of each field will be automatically done. However, this system can only make links from peptide database to other databases. It is required to ask the database owner to make the reverse direction link, which requires human negotiation and making peptide database aware to the world. Until that, we are making cross-reference database to this.

Peptide is a small molecule which has many biological and physical characters and functions. To understand its character, we made a virtual reality system with force feedback device. Usually 3D structure of molecules is represented in the computer screen as a turning around object or stereo-diagram. In this way, people imagine the 3D structure in their head. This new virtual reality system can “Touch and Feel” the molecules. It calculates the physical and electrostatic potential real time. Until this system, you couldn’t feel the electrostatic potential even if you made an actual model.

In this system, we made TouchMol and DockMol system. TouchMol enables to touch the molecule with a probe with electro potential. You can find the binding spot on the protein surface and mark it for drug design. DockMol can move two molecules and make a docking simulation. You can feel a drug going into active center of the protein, or even transcription factor binding to the DNA. We are planning to make an easy demonstration of TouchMol and DockMol at the symposium.

Studies on receptor-bound peptide conformation

Masaji Ishiguro, Kenji Sugase and Takahiro Hirano
Suntory Institute for Bioorganic Research, Shimamoto, Osaka 618-8503, Japan
(e-mail: ishiguro@sunbor.or.jp)

Biologically active peptides exert its activity by binding their receptors with specific conformations at the cellular surface. Due to flexible conformations in solution, it is very difficult to elucidate the receptor-bound conformation of peptides without any structural information on the receptors. Among various peptide receptors, we focused on G protein-coupled receptors (GPCR), integral membrane proteins with seven membrane-spanning helices, since GPCRs constitute a majority of biologically active peptide receptors and the crystal structure of rhodopsin, a prototype of GPCR, affords a three-dimensional image of other GPCRs for diffusible ligands such as biogenic amines and peptides. However, the crystal structure of rhodopsin is a inactive form of GPCR and thus modeling of structures of the active form of GPCR is prerequisite for understanding the receptor-peptide interactions since endogenous peptides are agonists, which bind active forms of their receptors.

In the first step of the modeling of peptide receptors of GPCR family, the two photointermediates, lumirhodopsin and metarhodopsin I, and the activated form of rhodopsin, metarhodopsin II, were constructed from the crystal structure of rhodopsin. Major structural changes of the photointermediates involve an outward swing of the C-terminal portion of transmembrane segment (TM) 3, and the rigid-body motion of TM6. These motions led to structural models of the two photointermediates and metarhodpsin II. The arrangement of the entire transmembrane segments of the photointermediate models closely agreed with the biophysical spectral data and provided a structural basis for the mechanism of ligand recognition.

The photointermediate models suggested that G protein-coupled receptors bind four functionally distinct ligands; inverse agonists, antagonists, partial agonists, and full agonists. Hence, putative structural models for biogenic amine receptors corresponding to the ligand function (inverse agonist-, antagonist-, partial agonist-, and full agonist-bound receptor models) were built using the photointermediate models in the rhodopsin photocascade. The ligand-receptor recognition of each was examined by modeling receptor-ligand complexes with functional ligands. The complex models suggested that each functional ligand binds the corresponding receptor structure.

The fully activated receptor models of endogenous peptides, which are classified in group A of GPCRs, could be constructed using the fully activated model of rhodopsin as the template. Examination of docking of biologically active peptides, such as angiotensin II, enkephalins, tachykinins and so on, into the binding cleft of the receptor suggested that the agonist peptides bind in a similar site enclosed by the seven transmembrane helices with

roughly similar conformations. Moreover, peptide analogues and non-peptide compounds bound other functionally distinct receptor structures. Details of the receptor bound conformation obtained from the docking study and the distinct recognition of the receptor structure by the synthetic peptides and non peptides will be discussed.

Free-energy landscape of peptides and conformational classification

Junichi Higo
School of Life Science, Tokyo University of Pharmacy and Life Science, Hachioji,
Tokyo 192-0392, Japan, and BIRD, JST (Japan Science and Technology Corporation)
(e-mail: higo@ls.toyaku.ac.jp)

The potential energy surface of a peptide/protein is generally high dimensional and involves a number of energy local minima. Thus, the efficiency of the sampling is essentially important. We have developed an enhanced conformational sampling method, multicanonical molecular dynamics simulation, and applied it to peptides in explicit water. The free-energy landscape is obtained from the sampling, and thermodynamically stable structures are extracted from the landscape. We have shown that the stable structures of peptides at room temperature in water correlate with experimentally obtained conformations. The advantage of the multicanonical sampling method is an ability of wide conformational sampling with overcoming energy barriers. Then, we can get not only the most stable but also semi-stable conformations, which may not be detected in the experimental measurement. In fact, the free-energy landscape consisted of some conformational clusters thermodynamically stable. With applying a principal component analysis on the simulation trajectories, the sampled conformations are classified into the clusters and cluster-cluster transition pathways are visible on the free-energy landscape.

Mass spectrometry for peptidomic and proteomic analysis

Toshifumi Takao[1], Yoshinori Satomi[1], Jorge Fernandez-de-Cossio[1, 2] and Masami Matsubae[1]
[1]Research Center for Structural and Functional Proteomics, Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan,
[2]Center for Genetic Engineering and Biotechnology, Havana, Cuba
(e-mail: tak@protein.osaka-u.ac.jp)

Mass spectrometry represents a well-accepted and reliable method for characterization of proteins. The method has great advantages in terms of high throughput, high accuracy, and high sensitivity in measurements, which is well suited for the identification of a wide variety of proteins and peptides, such those separated by 2D-PAGE and LC, and for the analysis of post-transnational modifications, which play important roles in various biological events. Taking advantages of accumulating protein/DNA sequence databases, proteomic or peptidomic analysis of tissues and body fluid such as urine and blood has become one of the most important tasks that might lead to discovery of biomarkers. We here wish to describe the current status of MS and its capability for overall analysis of peptides existing in human urine.

Peptides could be promptly isolated from urine by several batch-wise separation steps followed by LC, subjected to MALDI-MS/MS, and identified by database searching. The peptide profiles of urine and their sample preparations will be discussed. In addition, a database search engine targeted to the analysis of peptides has been developed, which allows the setting of modes of fragmentations according to the propensities of fragment ions to be observed in MS/MS. Especially, high-energy CID spectra of peptides obtained by MALDI-MS/MS could be efficiently analyzed with a combination of the developed search engine and PRF/SEQDB sequence database.

Hypothesis-driven Proteomics

Markus Kalkum[1] and Brian T. Chait[2]
[1] Beckman Research Institute of the City of Hope, Duarte, CA 91010,
[2] The Rockefeller University, New York, NY 10021, USA
(e-mail: MKalkum@coh.org)

Antibodies that combine high affinity with exclusive specificity are excellent tools to test for the presence of proteins of interest. However, high-quality immuno-reagents remain elusive or time-consuming to generate for a variety of targets. One alternative to antibodies involves mass spectrometric (MS) detection protein-specific proteolytic fragments. Unfortunately, MS observation of components from complex biological matrixes is often impaired by the limited dynamic range of the currently available methods. Thus, for example, signal suppression effects and the so-called "chemical noise" can greatly reduce sensitivity and can often hamper reliable application of MS methods. We describe here a novel MS approach, termed *hypothesis-driven multistage mass spectrometry* (HMS-MS), that allows us to overcome these limitations. The detection principle is based on a search for the absence or presence of a predicted peptide in a manner analogous to the use of a antigen-specific antibody. Characteristic fragmentation patterns differentiate the signal from the background noise and provide reliable evidence for the presence of the predicted peptide. We present a broad range of applications of our approach. For example, we demonstrate a rapid (< 1 min) screen for determining mating types of yeast cells (as used in our lab for yeast genetic projects). Another example describes the detection of autoinducing peptides from supernatants of pathogenic Staphylococci that are involved in *quorum sensing* and *bacterial interference* and reveals several novel facts about their structures. The broad versatility of HMS-MS is extended to neutral loss screens that allow a comprehensive description of site-specific phosphorylation in proteins. We also describe the use of "signature peptides" to determine the presence or absence of proteins in complexes, transforming HMS-MS into a powerful tool for studying protein-protein interactions.

New tools for proteome analysis

Osamu Nishimura
Life Science Laboratory, Shimadzu Corporation
Nakagyo-ku, Kyoto 604-8511, Japan
(e-mail: n_osamu@shimadzu.co.jp)

The sequence of human genome has been decoded and a post-genome era is now beginning. It is a start of protein era. Working at a fundamental research division of a pharmaceutical company and being engaged in studies on the research and development of new drug based upon genomic information, it was clear to me that genomic information was useful, significant and indispensable in driving forward new drug discovery. However, it is also true that we need more than that. Actually, I had been facing multiple barriers in pursuing for genomic information alone. There are numerous dead-end stories in genome-based new drug discovery and in many cases, those were the hurdles hard to get over. What approach can be a breakthrough of this bottleneck? I believe that one of the answers is a research on proteins. The whole of proteins is called proteome. The proteome analysis research field promises a bright future in discovery of new drug target, diagnosis and therapeutics. What is a key instrument for this research? It is obviously Mass Spectrometer (MS). The contribution of mass spectrometry was recently recognized by the award of the Nobel Prize Chemistry 2002 for 'the discovery of Soft Laser Desorption Ionization' to Mr. Koichi Tanaka. It is no doubt that MS plays an essential role to accelerate proteome analysis.

However, the present proteome analysis technology has not reached to the sufficient scientific level yet. Further improvements of both its hardware and software are necessary. For the hardware, improvement of sensitivity, accuracy and high-throughput, moreover equipment of ultra-micro analysis applied to the analysis of the proteome should be addressed. From the standpoint of the software, we need to develop new chemistries for proteome analyses and propose original, user-friendly proteome analysis methods.

In this symposium, I will discuss a new concept for proteome analysis based on mass spectrometry and new methodologies following our previous publications (1,2). This 'next generation' platform may be a solution for the development of an integrated system that will contribute to our understanding of clinical information/bioinformatics, therapeutics, diagnosis and even drug discovery.

1) An Approach to Quantitative Proteome Analysis by Labeling Tryptophan Residues
H. Kuyama, M. Watanabe, C. Toda, E. Ando, K. Tanaka & O. Nishimura,
Rapid Communications in Mass Spectrometry, 17, 1642 (2003)
2) An Efficient Chemical Method for Dephosphorylation of Phosphopeptides
H. Kuyama, C. Toda, M. Watanabe, K. Tanaka & O. Nishimura,
Rapid Communications in Mass Spectrometry, 17, 1493 (2003)

Design and performance of a new hybrid LC-QIT-TOF mass spectrometer

Kozo Miseki, Jun-ichi Taniguchi, Eizo Kawatoh, Shin-ichi Yamaguchi, Hiroto Itoi, Jun-ichi Koide, Takami Yamada, Junko Iida and Shaun Bilsborough
Shimadzu Corporation, Nakagyo-ku, Kyoto 604-8511, Japan
(e-mail: miseki@shimadzu.co.jp)

Quadrupole ion trap (QIT) offer many advantages over other instruments, not least the ability to obtain MSn fragmentation as an aid to structural elucidation. On the other hand TOF mass spectrometer deliver high mass resolution and high mass accuracy. Requirement for increased mass accuracy in both MS and MS^n mode has led to the idea of combination of QIT and TOF. A hybrid QIT-TOF configuration with MALDI source was reported by Tanaka et al.[1)] Based on their configuration a new hybrid LC-QIT-TOF has been developed. In this paper we describe the design and performance of a LC-QIT-TOF, capable of providing the high mass accuracy and resolution that is required for confident identification.

Ions generated at atmospheric pressure using electrospray ionization were focused towards the 3D quadrupole ion trap using newly developed ion optics that maximise ion transmission. Ions were accumulated prior to trapping using an octopole part-coated with a resistive material, which produced an axial electric potential in which the ions were accumulated near the exit side. Highly efficient filling of the trap was then achieved using a novel entrance electrode. Following isolation of the ion of interest and CID, high extraction efficiency was attained through rapidly stopping the RF voltage while simultaneously switching the voltage applied to the endcap electrodes. Precursor and products ions were ejected from the ion trap into a TOF tube.

Preliminary data of the instrument performance achieved at present will be shown. Single isotope of Angiotensin II was isolated in the Ion Trap, showing the isolation resolution greater than 1000. Additionally the resolution of the instrument has been shown to be in excess of 10,000 in MS and MS^3 mode. Mass accuracy data with small molecules in MS and MS^n mode will also be shown.

Reference:

1) Koichi Tanaka, et. al., Proceedings 47th ASMS conferences in Mass Spectrometry and Allied Topics (1999, June 13-17, Dallas, Texas), p.1823-1824

Atmospheric pressure MALDI–orthogonal acceleration TOFMS (APMALDI–oaTOFMS): A new tool for proteomics and peptidomics research

Jun Tamura[1], Tetsuichiro Morita[1], Kenji Nagatomo[1] and Phillip Tan[2]
[1] Analytical Instruments Division, JEOL Ltd., Akishima, Tokyo 196-8558, Japan,
[2] MassTech, Inc, Columbia, MD 21046, U.S.A.
(e-mail: tamura@jeol.co.jp)

An atmospheric pressure matrix assisted laser desorption/ionization (AP/MALDI or APMALDI)[1) 2)] ion source is interfaced with an orthogonal acceleration time-of-flight mass spectrometer (oaTOFMS)[3)]. Combination of these two technologies is expected to have unique advantages over conventional vacuum MALDI–TOFMS or APMALDI–QITMS (quadlupole ion trap mass spectrometer), which has been popular after the introduction of the commercial APMALDI ion source for QITMS[4)]:

1. Detection of the molecular-related ions of very fragile species such as sulfopeptides and phosphopeptides
2. Very good mass accuracy (~ 5 ppm) regardless of sample morphology
3. Possibility of very high throughput analyses using a high repetition rate (> 100 Hz) laser and robotics

These characteristics are quite suited for high throughput identification of proteins and peptides, especially for those with post-translational modifications.

All the data have been acquired with a model 611 AP/MALDI ion source (MassTech, Inc., MD, U.S.A.) and a JMS-T100LC AccuTOF API-oaTOFMS (JEOL Ltd., Tokyo, Japan). A prototype atmospheric pressure ionization interface employing a heated metal capillary has been developed for optimizing the transportation of MALDI generated ions from the atmospheric pressure into the high vacuum.

Basic performance characteristics of the instrument, such as sensitivity, mass resolution and mass accuracy will be demonstrated with peptide standards. Some applications to post-translationally modified proteins and peptides will also be discussed.

References:

1) Laiko, V. V.; Baldwin, M. A.; Burlingame, A. L. *Anal. Chem.* **2000**, *72*, 652 – 657
2) Moyer, S. C.; Cotter, R. J. *Anal. Chem.* **2002**, *74*, 469A – 476A
3) Guilhaus, M., Selby, D., Mlynski, V. *Mass Spectrom. Rev.*, **2000**, *19*, 65 – 107
4) MassTech, Inc, Columbia, MD 21046, U.S.A.

Data mining and database systems for large-scale proteome analyses

Takeshi Sakamoto,
Life Science Group, Hitachi, Ltd., Kawagoe, Saitama 350-1165, Japan
(e-mail: t-sakamoto@ls.hitachi.co.jp)

One of the goals of proteomics is a large-scale, comprehensive characterization of disease-related proteins. In a large proteomics facility, high-throughput proteome analyses produce more than thousands of mass spectra a day using a variety of state-of-the-art mass spectrometers such as MALDI-TOF-TOF and LC-ESI-MS/MS. Such a high sample throughput generates an overload of data. Therefore, a current bottleneck for high-throughput proteomics is the lack of a fully automated, sophisticated data analysis system to rapidly identify and quantify proteins from multiple protein mixtures. To overcome this limitation, we have developed an integrated bioinformatics platform for MS data analysis, which can generate confident results (protein identification, quantification, and characterization) with minimal human intervention.

Our platform is based on a relational database system running on Oracle 9i. Data acquisition and mining methods are mainly triggered from the database system as a web based client-server application. The database stores LIMS information, experimental procedures, sample clinical information, mass spec raw data, protein quantification data, a variety of annotations, and quality control data as well as identified peptides and proteins.

The platform incorporates the following essential automated steps of MS data processing:

(1) Data acquisition and archival from a variety type of mass spectrometers
(2) Peak list generation, merging, and submission to the protein search engine (Mascot)
(3) Parsing and uploading the search results into the database
(4) Several primary filters at the level of spectrum, peptide, and protein

The platform also includes sophisticated data mining tools such as:

(1) Regularly updates of an NCBInr-based annotated protein reference database
(2) Peptide/protein quantification algorithms supporting commercial methods like ICAT
(3) Novel peptide-level filtering/re-scoring algorithm utilizing an MO (molecular orbital) calculation
(4) Results integration from multiple mass spectrometers

The developed system can be applied on any large-scale protein identification and quantification approach of Proteomics projects.

Cancer proteomics

Young-Ki Paik,
Department of Biochemistry, Yonsei Proteome Research Center, Yonsei University,
Seoul, 120-749 Korea
(e-mail: paikyk@yonsei.ac.kr)

My talk will be composed of three parts: First part of my presentation will be focused on our studies of the functional roles of proteins in hepatocelluar carcinoma (HCC) at the protein level. To this end, more than 150 paired liver tumor tissues were analyzed by three different proteomic tools including two-dimensional electrophoresis with different pI ranges, ICAT and SELDI-TOF. Some highlights will be presented from the proteomic analyses of more than 150 proteins in which emphasis will be placed on the differentially expressed proteins that are closely related to the key metabolic pathways of glucose, amino acid and energy production. Second part of my talk deals with proteome informatics. During the course of study of HCC, we have constructed a hepatocellular carcinoma reference database (HCC-RDB), a web-based database, which can execute integration, data-mining and query function using various data obtained from two dimensional electrophoresis (2DE), non-2DE data (*e.g*., isotope-coded affinity tag, surface-enhanced laser desorption ionization) and genomic data (e.g., loss of heterozeigosity). HCC-RDB enables link various protein data to ExPASy, NCBI, PDB, KEGG, and Gene card through which one can search for the names of genes, their relative expression levels, and chromosomal loci. Finally, I will discuss recent progress on the plasma protein analysis of HCC patients using non-2D platform technologies. This part of talk will be followed by brief introduction of some progress made in the KOREAN HUMAN PLASMA PROTEOME PROJECT that have been recently initiated by the Ministry of Health and Welfare through the Biomedical Proteome Research Center Program at YPRC.

References:

1. Paik, Y.K (2003) A New Horizon of Proteomics in Korea, Editorials, Proteomics 3
2. Choi, B-K et al. (2003) *Molecular & Cellular Proteomics, 2, 1086-1095*
3. Choi B-K et al. (2003) *Proteomics, 3, 1955-1961*
4. Bae, S-H et al. (2003), *Proteomics,* 3, 569-79
5. Choi Y-R. et al. (2003) *Cancer Res.,* 63, 2188-2193
6. Cho SY et al. (2002) *Proteomics, 2,* 1104-1113
7. Park K-S et al. (2002), *Intl. J. Cancer,* 97, *261-265*
8. Park K-S. et al. (2002), *Hepatology,* 35, *1459-66*

A new strategy of large-scale protein profiling in clinical proteomics

Toshihide Nishimura
Clinical Proteome Center, Tokyo Medical University,
Shinjuku-ku, Tokyo 163-0203, Japan
(e-mail: nisimura@tokyo-med.ac.jp)

The word "proteome" first appeared in 1995, and means simply the entire proteins expressed by cells in complement of genome. Proteomics is one of promising approaches to understand gene function in the post-genomic era. A goal of proteome research is to achieve high throughput qualitative and quantitative analysis of proteins expressed. Especially, the spatiotemporal study on a protein to reveal its functional aspects becomes more important because proteomics is the science studying a dynamics of protein expression and interaction in a biological system.

In the classical proteome analysis, 2-dimensional gel electrophoresis (2-DE) was utilized to grasp a change in proteins expressed, for instance, between normal and perturbed cells by such as disease, drug exposure etc. However, 2-DE method is time-consuming and has various issues upon reproducibility, sensitivity and its dynamic range. 2-DE method is feasible for soluble proteins but has a great difficulty in displaying hydrophobic membrane proteins, which are the mostly attractive targets for drug discovery research. Then, 1-D SDS polyacrilamide gel electrophoresis (PAGE) method, which can capture membrane proteins, has been combined with mass spectrometry. This approach has shown its usefulness in profiling proteins relevant to a tissue or disease. However, gel-based methodologies restrict their detection limit of not less than 50 fmol. To capture proteins expressed in the level of 1,000 –10,000 copies in a cell, a combination of a complete solubilization of all proteins with gel-free on-line proteome analytical approach is required.

Huge numbers of human sample will be subject to their clinical proteome analysis. In this presentation, we will briefly introduce our strategy of clinical proteomics based upon multi-dimensional quantitative protein profiling integrated with tandem mass spectrometry. This multi-dimensional profiling system has been optimized by our proprietary informatics tools in order to mine proteins associated significantly with a group characterized by a specific clinical /disease parameter. Here, we will utilize the results of collaborative research on metastasis of early lung carcinoma, as an example.

The strategy in proteomics and proteome informatics in TMIG

Tosifusa Toda and Hiraku Morisawa
Tokyo Metropolitan Institute of Gerontology, Itabashi-ku,
Tokyo 173-0015, Japan
(e-mail: ttoda@tmig.or.jp)

The potential applications of proteome analysis are numerous, ranging from the confirmation of open-reading frames, assigning function to novel gene products, to the analysis of changes in protein expression during development, aging, in disease, and in response to extra-cellular stimuli such as hormones, cytokines, drugs and oxidative stresses. Some of these applications have been initiated in Tokyo Metropolitan Institute of Gerontology (TMIG) right after the basic strategy in proteomics was reported by Williams and his co-workers. The original method of proteomics had been developed on two-dimensional gel electrophoresis (2-D PAGE) for fine resolution of protein. Although two-dimensional liquid chromatography was also introduced for the separation of peptides generated by tryptic digestion in “shot-gun proteomics”, the proteomic research project in TMIG adopted the 2-D PAGE-based method consistently because the original method had the advantage of differential display in the research on aging and geriatric diseased over the 2-D HPLC based method. Even in the method of 2-D PAGE-based proteomics, new technology for the differential display has been developed on pre-labeling of protein with dual-color fluorescent dyes (2-D DIGE). However, the method of 2-D DIGE could not be applicable to the longitudinal study of alterations in protein expression during aging process in our gerontology research.

The biggest problem in 2-D PAGE-based proteomics is the difficulty in management of a lot of data generated by the 2-D gel image analysis and mass spectrometry. The primary data are generally stored in local computers for image analysis and mass spectrometry. However, a server-client computer network for Laboratory Information Management System (LIMS) had been getting necessary for data sharing in members of research project in TMIG. In the above situation, the WorksBase proteome data management system was introduced into our institute in corroboration with Nippon Bio-Rad Laboratories for testing the performance of the data sharing system. We will report our proteome research on mechanisms of aging and on data management system in our institute.

Building a cell map:
Experimental approaches to pathway expansion

Walter Blackstock
Protein Partners, Hertfordshire SG12 7AR, United Kingdom, and
Tokyo Medical University, Tokyo 163-0203, Japan

Information flows within, and between, cells, triggered by proteins recognising other molecules, and selecting from them just a few of many possible choices. In most cases recognition events lead to protein complex formation and the transmission of information through protein pathways. Pathways may be less linear than genetic experiments imply, and many approaches will be needed to understand them. Rapid progress in genome sequencing, together with technical advances in protein mass spectrometry, prompted us, in 1998, to suggest a “Cell Map” as an achievable (if expensive) goal (1). To develop the concept and explore the likely cost we founded the Cell Map Unit in GlaxoWellcome. We developed a factory-like approach to producing tagged protein complexes under conditions that disturbed normal cellular events as little as possible. I will describe the tandem affinity purification (TAP) technique, which is now finding increasing use, both for large-scale and smaller focused projects. The need for biological validation and tools for visualising the data will be emphasised. Finally, I will suggest that we need to fund a large-scale functional proteomic centres on the scale of genome sequencing centres if we are to achieve a useful Cell Map, and put this in the context of the recent NIH roadmap (http://nihroadmap.nih.gov) and the DoE 20-year facilities plan (http://www.sc.doe.gov).

Reference:

1. Proteomics: quantitative and physical mapping of cellular proteins.
Blackstock WP, Weir MP, *Trends Biotechnol*. 1999, 17(3), 121-7