Awards Review

Studies of the Structure-Activity Relationships of Peptides and Proteins Involved in Growth and Development Based on Their Three-Dimensional Structures

Koji Nagata

Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences,
The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

Online Publication, March 7, 2010
[doi:10.1271/bbb.90678]

Peptides and proteins with similar amino acid sequences can have different biological functions. Knowledge of their three-dimensional molecular structures is critically important in identifying their functional determinants. In this review, I describe the results of our and other groups’ structure-based functional characterization of insect insulin-like peptides, a crustacean hyperglycemic hormone-family peptide, a mammalian epidermal growth factor-family protein, and an intracellular signaling domain that recognize proline-rich sequence.

Key words: nuclear magnetic resonance; bombyxin; molt-inhibiting hormone; epidermal growth factor-like domain of heregulin-α; WW domain

The growth and development of multicellular organisms are controlled in a coordinated manner by the activity of a number of signaling molecules and their corresponding pathways. There are many signaling proteins with similar amino acid sequences but different biological functions. My research objective is to identify the functional determinants of such signaling proteins by determining their three-dimensional structures by the use of nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography and by evaluating the importance of amino acid residues by mutational analysis. In this review article, I describe our results mainly, and also some recent progress by other groups on the function-structure relationships of insect insulin-like peptides (bombyxin-II and Samia bombyxin-related peptides), crustacean hyperglycemic hormone-family peptides (molt-inhibiting hormone), mammalian epidermal growth factor (EGF)-family peptides (the EGF-like domain of heregulin/neuregulin-α), and an intracellular signaling domain recognizing proline-rich sequences (the WW domain).

I. Insect and Crustacean Neuropeptides That Can Regulate Molting

1. Bombyxin

Bombyxin is a brain-secretory peptide of the silkworm Bombyx mori. It was the first molecule identified as an insulin-related peptide of invertebrate origin based on sequence similarity to insulin (50% and 32% identical in the A- and B-chains, respectively). When bombyxin is administered to a brain-removed dormant pupa of the saturniid moth Samia cynthia ricini, it promotes adult development of the pupa by stimulating the prothoracic glands to synthesize and release ecdysteroid, the insect molting hormone. In addition, bombyxin lowers the concentration of hemolymph trehalose, the major blood sugar of B. mori, in a dose-dependent manner, induces meiosis in the ovary, induces morphological changes in BM-N4 cells established from ovarian tissue of B. mori, and acts together with 20-hydroxyecdysone to stimulate the cell division and growth of wing imaginal disks. Thus bombyxin is similar to insulin not only in primary structure but also in biological function, but no cross-activity is observed between them; porcine insulin fails to show bombyxin-like prothoracicotropic activity even at a dose of 1 μg/pupa, about 10^6 times the minimal active dose of bombyxin, and bombyxin fails to bind to the human insulin receptor even at a concentration of 1 μM, about 3 × 10^12 times the concentration of the half-maximal inhibitory concentration (IC_{50}) of human insulin. Bombyxin binds to ovarian cells of three different species of lepidoptera, B. mori, S. cynthia ricini, and the fall armyworm Spodoptera frugiperda, as well as to their ovarian cell lines. BM-N4 and SP9 cells display 15,800 and 20,000 putative bombyxin receptors per cell. The bombyxin receptors expressed on the surface of BM-N4 and SP9 cells are similar to the mammalian insulin receptor in molecular size and subunit structure, and bind bombyxin in a specific, saturable, and reversible manner with dissociation constants of 2,360 ± 560 pm and 260 ± 90 pm, respectively.

In order to elucidate the structural bases of different functions, that is, different receptor specificities, between bombyxin and insulin and to investigate the molecular evolution of insulin-superfamily peptides, we determined the three-dimensional structure of bombyxin-II, a representative molecular species of bombyxin, based on two-dimensional proton NMR data (Fig. 1A). This was the first three-dimensional structure solved for an invertebrate insulin-superfamily
peptide. The overall main-chain fold of bombyxin-II, Protein Data Bank (PDB): 1BOM and 1BON, is similar to those of insulin (PDB: 1HIU) and relaxin-2 (PDB: 6RLX). Acidic, basic, neutral, and hydrophobic residues are colored red, blue, yellow, and green, respectively.

B. A possible structural change in insulin, from the closed state to an open state, upon binding to its receptor. If the conformation of insulin is fixed to the closed state, it cannot activate the insulin receptor. The solution structure of an active mutant, [Gly\text{B24}]insulin (PDB: 1HIT), which represents an open state, can reflect the bound state of insulin to the insulin receptor. The exposed hydrophobic patch of bombyxin-II, which corresponds to the putative receptor-recognition site, shown on the right side of the molecule in Fig. 1A, is colored magenta. A similar hydrophobic patch is present also in insulin (magenta), but it is usually covered by the aromatic cluster, Ph\text{B24}-Ph\text{B25}-Tyr\text{B26}, of the B-chain C-terminal \(\beta\)-strand (yellow) in insulin to form a characteristic molecular surface both in the crystalline state and in solution. The crystal structure of IGF-II (magenta: the hydrophobic patch; yellow: the B-region C-terminal part and the C-region; cyan: the other regions on the molecular surface) bound to IGF2R (white; PDB: 2V5P) shows that the IGF2R-recognition site of IGF-II is located on the opposite side of the molecule from the hydrophobic patch. This figure was prepared with QUANTA (MSI) and PyMOL.

Fig. 1. Three-Dimensional Structures and Putative Receptor Recognition Sites of Insulin-Family Peptides.
A. Three-dimensional structures and putative receptor-recognition sites of bombyxin-II (PDB: 1BOM), human insulin (PDB: 1HIU), and human relaxin-2 (PDB: 6RLX). Acidic, basic, neutral, and hydrophobic residues are colored red, blue, yellow, and green, respectively.
B. A possible structural change in insulin, from the closed state to an open state, upon binding to its receptor. If the conformation of insulin is fixed to the closed state, it cannot activate the insulin receptor. The solution structure of an active mutant, [Gly\text{B24}]insulin (PDB: 1HIT), which represents an open state, can reflect the bound state of insulin to the insulin receptor. The exposed hydrophobic patch of bombyxin-II, which corresponds to the putative receptor-recognition site, shown on the right side of the molecule in Fig. 1A, is colored magenta. A similar hydrophobic patch is present also in insulin (magenta), but it is usually covered by the aromatic cluster, Ph\text{B24}-Ph\text{B25}-Tyr\text{B26}, of the B-chain C-terminal \(\beta\)-strand (yellow) in insulin to form a characteristic molecular surface both in the crystalline state and in solution. The crystal structure of IGF-II (magenta: the hydrophobic patch; yellow: the B-region C-terminal part and the C-region; cyan: the other regions on the molecular surface) bound to IGF2R (white; PDB: 2V5P) shows that the IGF2R-recognition site of IGF-II is located on the opposite side of the molecule from the hydrophobic patch. This figure was prepared with QUANTA (MSI) and PyMOL.
Leu, and Leu), A19 (Tyr, Tyr, and Phe), B6 (Tyr, Leu, and Leu), B11 (Leu, Leu, and Leu), B15 (Leu, Leu, and Gln) and B18 (Leu, Val, and Ile), and the A6–A11 and A20–B19 disulfide bonds. Despite the overall structural similarity, the structure of the B-chain C-terminal part of bombyxin-II is quite different from that of insulin. Bombyxin-II adopts a relaxin-2-like helix and a coiled structure instead of an insulin-like sharp turn and an extended β-strand. In insulin, the extended structure of the B-chain C-terminus is stabilized by intramolecular hydrophobic interactions between PheB24 and ValB112/LeuB15/TyrB16, and between TyrB26 and LeuB11/IleA2/VaA3. On the other hand, the extension of the B-chain helix in the corresponding part of bombyxin-II and relaxin-2 results from the lack of a Gly residue at either B20 or B23 as well as the shorter B-chain C terminal tail terminating at B25. Almost all the invertebrate insulin-related peptides so far characterized lack Gly at B20, which suggests that they take a helix extension, as do bombyxin-II and relaxin. Before we determined the solution structure of bombyxin-II, its three-dimensional structure was predicted by interactive computer graphics and energy minimization techniques by assuming homology with the X-ray structure of porcine insulin.34 In the modeled structure, the B-chain C-terminal part of bombyxin-II adopts a type-III β-turn between CysB19 and AlaB22 and an extended C-terminal segment in a way similar to insulin, but the modeled structure is not correct, at least in solution.

Our structure-activity relationship studies of bombyxin show that the residues at the A-chain N and C termini (GlyA1, ValA5, and CysA25/CysB19) and the B-chain central part (TyrB6, LeuB11, AlaB12, and ThrB14 to LeuB18) are important to its activity, while the residues at the B-chain N terminus (pGluA1–2 to GlnA11) and the B-chain C terminus (AlaB22 to AspB25) are not.15–18 In the case of insulin, the residues at the A-chain N terminus (GlyA1 to GluA5), the A-chain C terminus (TyrA19 and AsnA21), the B-chain central helix (ValB112 and TyrB16), and the B-chain C-terminal β-strand (PheB23/PheB25/TyrB26) have been found to be important to receptor recognition.19–21 Thus, these two molecules use the corresponding molecular surfaces for cognate receptor recognition, although the structures and properties of their surfaces are dissimilar, probably due to their divergent molecular evolution.

The insulin receptor, the type-I insulin-like growth factor (IGF) receptor (IGF1R),23 which binds insulin-like growth factors-I and -II (IGF-I and IGF-II) strongly and insulin weakly, and the putative bombyxin receptor6,9 are αβ2 heterotetrameric tyrosine kinases. Insulin-like growth factor-II (IGF-II) and relaxin use totally different classes of receptor proteins. The type-2 IGF receptor (IGF2R), also known as the cation-independent mannose 6-phosphate receptor (CI-M6PR), acts as a clearance receptor for IGF-II that activates no intracellular signaling pathways.23 At present, the only known ligand-receptor complex structure for the insulin-family peptide and the cognate receptor protein is the complex of IGF-II and a part of the ectodomain of IGF2R (PDB: 2V5P), in which IGF-II recognizes IGF2R via the side of the molecular surface opposite to the putative receptor-recognition hydrophobic patch of insulin and bombyxin (Fig. 1B).24 The relaxin receptors, RXFP-1–4, belong to two different classes of G-protein-coupled receptors (GPCRs). RXFP-1 and -2 are leucine-rich repeat-containing GPCRs that act as the receptors of relaxin-2, the ovarian relaxin, while RXFP-3 and -4 are more conventional GPCRs without a large ectodomain that act as the receptors of relaxin-3, the neuropeptide relaxin.25 In relaxin-2, residues ArgB13, ArgB19, and IleB16 on the B-chain central helix and residue TrpB25 at the B-chain C-terminus as well as A-chain residues ThrA12, LysA13, and PheA19 are involved in the recognition of RXFP-1 and -2.13 Thus, despite the different classes of receptors, insulin/bombyxin and relaxin-2 use the corresponding molecular surfaces to recognize the cognate receptors (Fig. 1A).

There is an exposed hydrophobic patch in the proposed receptor-recognition surfaces of bombyxin-II, while the corresponding hydrophobic patch is covered by the aromatic cluster, PheB24/PheB25/TyrB26, of the B-chain C-terminal β-strand in insulin to form a characteristic molecular surface (Fig. 1B). The B-chain C-terminal β-strand of insulin is critically important to receptor recognition,26 but it inhibits receptor binding when its conformation is fixed by cross-linking.27 Hence, it has been proposed that the B-chain C-terminal β-strand of insulin detaches from the core structure when it binds to the receptor, which can result in the exposition of a similar hydrophobic patch, as in bombyxin (Fig. 1B).11 In contrast to insulin, bombyxin does not need to reorganize the molecular surface when binding to its receptor. The receptor-binding site of the ancestral molecule probably involved a common framework (the A-chain N and C termini and the B-chain central helix),21 and the involvement and associated conformational changes in the B-chain C terminus of insulin11,26 might have appeared at a later stage of evolution. Furthermore, the B-chain C-terminal part of insulin contributes to insulin hexamer formation,22 which is required for the storage of enough insulin in the storage granules of pancreatic β cells.

2. Samia bombyxin-related peptides (SBRPs)

In addition to bombyxin in the silkmoth B. mori, the presence of insulin-like peptides in insects has been discovered, e.g., Locusta insulin-related peptide (LIRP) in the grasshopper Locusta migratoria,28 Samia bombyxin-related peptide (SBRP) in the saturniid moth S. cynthia ricini,29 and Agris bombyxin-related peptide (ABRP) in the sweet potato hawkmoth Agris convolvuli.30 These peptides are present in neurosecretory cells of the central nervous system, and are considered to play important roles in metabolism, growth, and reproduction, as do vertebrate insulin-superfamily members, but little is known about their physiological functions. Bombyxin stimulates the prothoracic glands of S. cynthia ricini to synthesize and release the molting hormone ecdysteroid.3 Curiously, bombyxin fails to promote the pupal-adult development of B. mori, from which it originates.31,32 Although the very low prothoracicotropic activity of bombyxin has been confirmed by in vitro assay, this activity was judged to be unphysiological.32 The brain of S. cynthia ricini contains two groups of molecules that show prothoracicotropic activity, big prothoracicotropic hormone (big PTTH, ca. 35 kDa) and small prothoracicotropic...
tropic hormone (big PTTH, ca. 4 kDa), as assessed by S. cynthia ricini in vivo assay.\textsuperscript{33} The genes for the bombyxin homologs of S. cynthia ricini (SBRPs) were cloned and found to be expressed in the brain\textsuperscript{29} in a manner similar to bombyxin.\textsuperscript{34,35} Because of the similar molecular sizes of SBRPs (5 kDa) and S. cynthia ricini small PTTH (4 kDa), SBRPs are considered to be the most probable candidates for the small PTTH of S. cynthia ricini.

We chemically synthesized two molecular species, SBRP-A1 and -B1, and examined their prothoracicotropic activity by in vivo S. cynthia ricini pupal assay.\textsuperscript{36} SBRP-A1 and -B1 are active at minimal effective doses (ED\textsubscript{50}) of 50 and 10 ng/pupa, i.e., minimal effective concentrations (EC\textsubscript{50}) of 5 and 1 nM, respectively, given that the hemolymph volume in a pupa is about 2 ml, but their prothoracicotropic activity is weaker than that of bombyxin-II by 20–100 fold on a molar basis. Since Southern hybridization has indicated that the S. cynthia ricini genome contains many other SBRP genes,\textsuperscript{29} other SBRP molecular species with higher activity might exist. We also synthesized a chimeric molecule consisting of the SBRP-B1 A-chain and the bombyxin-II B-chain. The chimeric molecule showed 5-fold higher prothoracicotropic activity than SBRP-B1.\textsuperscript{36} This indicates that the B-chain central part, particularly His\textsuperscript{B10}, Arg\textsuperscript{B13}, and/or Asp\textsuperscript{B17} in bombyxin-II, is important to prothoracicotropic activity. The brain of S. cynthia ricini contains two different classes of molecules with prothoracicotropic activity, big PTTH (about 35 kDa) and small PTTH (about 4 kDa),\textsuperscript{33} as in the case of the tobacco hornworm Manduca sexta.\textsuperscript{37} Bioassay results indicated that the SBRP molecules are S. cynthia ricini small PTTH. On the other hand, the putative S. cynthia ricini big PTTH, which is homologous to B. mori PTTH, shows 200 to 1,000-fold higher prothoracicotropic activity than SBRP-A1 and -B1 on a molar basis.\textsuperscript{36} Since putative S. cynthia ricini big PTTH and SBRPs share no sequence similarity, they should express prothoracicotropic activity in different ways, perhaps via different receptors. They might act stage-specifically in a manner similar to those of the tobacco hornworm Manduca sexta, in which big PTTH (about 25–30 kDa) stimulates equally larval and pupal prothoracic glands, while small PTTH (about 7 kDa) activates larval prothoracic glands much more than pupal prothoracic glands.\textsuperscript{37}

3. Molt-inhibiting hormone (MIH)

Molting is one of the most significant processes occurring during the arthropod life cycle. It is triggered by the molting hormone ecdisyordinates. In crustaceans, it is assumed that the synthesis and the secretion of ecdisyordinates by the Y-organ are suppressed by molt-inhibiting hormone (MIH).\textsuperscript{38} MIH is produced by the X-organ and is released from the sinus gland, located in the eyestalk. Crustacean hyperglycemic hormone (CHH, which increases hemolymph glucose levels), gonad-inhibiting hormone (GIH, which suppresses vitellogenesis in the ovary), and mandibular organ-inhibiting hormone (MOIH, which suppresses the synthesis of methyl farnesoate in the mandibular organ) are also synthesized in and released from the X-organ/sinus gland complex. These peptide hormones, 72–78-amino acids in length, exhibit similar amino acid sequences, and thus form the CHH family of peptide hormones.\textsuperscript{39} CHH family peptides commonly have six cysteine residues, which form three intramolecular disulfide bonds.\textsuperscript{40} In general, each CHH family peptide shows only one type of biological activity, although a few peptides show multiple hormonal activities.

To determine the structural basis of the functional specificity of CHH family peptides, we determined the solution structure of MIH from the kuruma prawn Marsupenaeus japonicus based on three-dimensional heteronuclear NMR data (Fig. 2).\textsuperscript{41} This was the first three-dimensional structure solved for a crustacean neuropeptide (PDB: 1J0T). The MIH molecule consists of an N-terminal region, five α-helices, four loops between the α-helices, and a C-terminal tail region. The conformation of MIH is stabilized by a hydrophobic cluster and three disulfide bonds. These hydrophobic and cysteine residues are conserved in most CHH family peptides, suggesting that CHH family peptides adopt similar folds. A search for peptides and proteins structurally similar to MIH using the DALI server\textsuperscript{42} indicated that CHH family peptides form a novel class of folds. The tertiary structure of a M. japonicus CHH was modeled based on 32% sequence identity to M. japonicus MIH\textsuperscript{33} using the SWISS model server (Fig. 2).\textsuperscript{44} The modeled structure of CHH was very similar to the solution structure of MIH. MIH has a characteristic insertion of a glycine residue at position 12 located at α1. On the other hand, the C-terminal sequence of CHH is shorter than that of MIH by several amino acids. The C-terminal region of MIH is located close to α1 in the tertiary structure (Fig. 3A), where the surface properties, the electrostatic potential and hydrophobicity, of MIH and CHH are different. The functional site of CHH involves the C-terminal region based on the fact that the C-terminal amide moiety of CHH is required for hyperglycemic activity\textsuperscript{43,46} and the fact that a truncated form of CHH that lacks six C-terminal amino acid residues shows no activity.\textsuperscript{46} In Fig. 2, the highly conserved, characteristic residues to MIH and CHH are colored. The colored residues are located mainly on one side of the molecular surface (the front side in Fig. 2). Hence we proposed that the colored molecular surfaces on MIH and CHH are important to molt-inhibiting and hyperglycemic activities, respectively.\textsuperscript{41} This was later confirmed by a mutational study showing the importance of Asn\textsuperscript{13}, Ser\textsuperscript{71}, and Ile\textsuperscript{72} to the molt-inhibiting activity of MIH.\textsuperscript{47}

II. Mammalian Growth Factor

The epidermal growth factor (EGF)-like domain of heregulin (HRG)-α

ErbB-2 (HER-2 or Neu\textsuperscript{48,49}) and ErbB-4 (HER-4)\textsuperscript{50} are receptor-type tyrosine kinases that are similar to ErbB-1 (the epidermal growth factor receptor, EGFR, or HER-1)\textsuperscript{51} and ErbB-3 (HER-3).\textsuperscript{52} Simultaneous overexpression of these ErbB family tyrosine kinases is often observed in human cancer cells.\textsuperscript{53} In particular, overexpression of ErbB-2 is known to correlate with a poor prognosis in breast, ovarian, and endometrial cancers and non-small cell lung adenocarcinoma.\textsuperscript{54} Heregulin (HRG)-α, a 45 kDa glycoprotein isolated from a con-
ditioned medium of MDA-MB-231 human breast carcinoma cells, and related molecules (HRG-β1, β2, and β3) were purified and cloned as specific activators for ErbB-2. Each HRG molecule is a mosaic glycoprotein containing an immunoglobulin unit and an EGF-like domain. The EGF-like domain has been found to be sufficient to activate ErbB-2. It was found later that HRGs bind not to ErbB-2 but to homologous receptors ErbB-3 and -4, and activates ErbB-2 indirectly through receptor heterodimerization and transphosphorylation. The EGF-like domain, which is defined by six cysteine residues characteristically spaced over a sequence of 35–40 amino acid residues, is shared by many functionally diverse proteins, including growth factors, e.g., EGF and transforming growth factor (TGF)-α, cell adhesion molecules, e.g., laminin, and plasma proteins, e.g., protein C. The EGF-like domain of HRG-α shares sequence identity with EGF family growth factors EGF (27%) and TGF-α (32%). In spite of the sequence similarity, the EGF-like domains of HRG-α and EGF/TGF-α are functionally different. The EGF-like domains of HRG-α bind specifically to ErbB-3 and -4 but not to ErbB-1, whereas EGF and TGF-α bind to ErbB-1 but not to ErbB-3 or -4.

In order to elucidate the structural basis of receptor specificity, we determined the solution structure of the EGF-like domain of HRG-α based on two-dimensional proton NMR data, and compared it with the reported

![Fig. 2. Three-Dimensional Structures of MIH (the solution structure; PDB: 1J0T) and CHH (a homology-based model).](image-url)

The highly conserved, characteristic residues to MIH and CHH are colored in the same color code as in Fig. 1A, while the other residues are shown in white. The colored residues are important to the respective activities of MIH and CHH. In fact, the labeled residues on MIH have been shown to be important to MIH activity. This figure was prepared with PyMOL.

![Fig. 3. Three-Dimensional Structures of the EGF-Like Domain of HRG-α (PDB: 1HRE), EGF (PDB: 1IVO), and TGF-α (PDB: 1MOX).](image-url)

The EGF and TGF-α molecules are bound to their receptor ErbB-1 (presented by a line), and this is a receptor’s view to the ligands. The colored residues of EGF and TGF-α are all involved in receptor binding. The EGF-like domain of HRG-α binds to its receptor, ErbB-4, in a similar manner via a molecular surface equivalent to the receptor-binding surfaces of EGF and TGF-α. The putative ErbB-4 binding residues of the EGF-like domain of HRG-α are colored and labeled. The color code is the same as in Fig. 1A. This figure was prepared with PyMOL.

![Fig. 4. Ligand-Recognition Surfaces of WW Domains.](image-url)

Group I, II/III, and IV WW domains (PDB: 1EGF, 2HO2, and 1FSA) recognize the PY motif, the polyproline/PL motif/PR motif, and the pS/pT-P motif, respectively. Ligand specificity is determined by the characteristic ligand binding motif for each group: group I, Tyr-pocket (red); group II/III, XP2 groove (orange); and group IV, p patch (cyan). On the other hand, the XP groove (magenta) is common to all WW domains. This figure was prepared with PyMOL.
structures of EGF and TGF-α (Fig. 3). The tertiary structure of the EGF-like domain of HRG-α (PDB: 1HRE and 1HRF) is characterized by two subdomains: the N-terminal subdomain and the C-terminal subdomain. The N-terminal subdomain consists of a triple-stranded antiparallel β-sheet (βI). An irregular helix is attached to βI by two disulfide bonds. The C-terminal subdomain forms a small antiparallel double-hairpin structure. In spite of the low sequence similarity, the overall main-chain fold of the EGF-like domain of HRG-α is similar to those of EGF/TGF-α, previously reported. When compared the structure of the EGF-like domain of HRG-α with those of EGF/TGF-α, the RMS deviations are 1.74 and 1.44 Å, respectively, for the main-chain atoms, excluding the divergent and less well-defined regions. The linkages of the three disulfide bonds in the EGF-like domain of HRG-α are the same as those in EGF/TGF-α, and the conserved glycine residues in both HRGs and EGF/TGF-α take positive φ values, which indicates that the conserved cysteine and glycine residues are essential to construct the characteristic backbone fold of the EGF-like structural unit. The conserved Arg220 is located in the subdomain interface with its side chain lying in a hydrophobic pocket formed by Phe178 and Val181, as is the case for EGF and TGF-α. In EGF, Arg220 is essential for ErbB-1 recognition. Similarly, Arg220 in HRG-α is probably required functionally to confer on the molecule an affinity for ErbB-3/4. Some structural features are uniquely observed in HRG-α. Firstly, the N-terminal β-strand, β1, in HRG-α is well-defined due to tight hydrogenic interactions with β2, instead of being disordered or less well-defined, as in EGF/TGF-α. Secondly, the orientation of the irregular helix relative to the N-terminal β-sheet I is different between HRG-α and EGF/TGF-α. The helix in HRG-α runs perpendicularly across the bottom of the antiparallel β-strands, β2 and β3, while the helix in EGF/TGF-α runs obliquely across the β3-strands.

Three regions have been proposed for the receptor (ErbB-l)-binding surface in EGF based on mutagenic studies: (i) the face of the N-terminal β-sheet βI on the opposite side of the irregular helix (Ile209,70,71) (ii) the inter-subdomain cleft (Arg220),68,69 and (iii) the C-terminus (Leu226),72,73 which was later confirmed and refined by the crystal structures of the EGF–ErbB-1 (PDB: 1IVO)74 and TGF-α–ErbB-1 (PDB: 1MOX)75 complexes (Fig. 3). Because of the structural similarity between the EGF-like domain of HRG-α and EGF/TGF-α and the sequence similarity between ErbB-3/4 and ErbB-1, HRG-α is considered to have a receptor (ErbB-3/4)-recognition surface similar to those of EGF/TGF-α. Hence we proposed that the residues in the above three regions that were strictly conserved in HRGs but not conserved in EGF/TGF-α are involved in ErbB-3/4 recognition. Based on the crystal structures of the EGF–ErbB-1 and TGF-α–ErbB-1 complexes,74,75 we have revised the putative ErbB-3/4-recognition residues in the EGF-like domain of HRG-α, which include His178, Lys187, Phe189, Val191, Asn192, Phe197, Val199, Asp201, Leu202, Arg207, Leu209, Cys210, Lys211, Cys212, Thr217, Gly218, Ala219, Arg220, Thr222, Glu223, Asn224, Pro226, Met227, Lys228, and Gln230 (Fig. 3). The distribution of potential receptor-recognition sites around the whole EGF-like domain of HRG-α implies that the ligand interacts with the ErbB-3/4 receptor at multiple sites. The receptor specificity between the EGF-like domain of HRG-α and EGF/TGF-α must be an effect of different exposed patches on the receptor binding surface (putative in the case of the EGF-like domain of HRG-α), as shown in Fig. 3.

### III. Intracellular Signaling Module

**The WW domain—Redefining the classification**

The WW domain is a small protein module composed of approximately 40 amino acids with two highly conserved tryptophans, known as an intracellular signaling module that binds proline-rich peptide motifs. It consists of a stable, triple stranded β-sheet, and it recognizes proline-containing ligands. WW domains are found in many different signaling and structural proteins, often localized in the cytoplasm as well as in the cell nucleus. They were originally classified into four groups according to ligand specificity. Group I recognizes -Pro-Pro-Xaa-Tyr- (the PY motif); group II recognizes -Pro-Pro-Leu-Pro- (the PL motif); group III recognizes proline-rich segments with Arg residues (the PR motif); and group IV recognizes -Ser(P)/Thr(P)-Pro- (the PS/PT-P motif, where Ser(P)/pS and Thr(P)/pT mean phosphoserine and phosphothreonine, respectively). Based on ligand binding assay data and several known molecular structures of WW domains, we characterized the ligand recognition mechanisms of all the classes of WW domains, and have proposed a better classification system, in which the old groups II and III are joined together as one group, group II/III (Table I and Fig. 4), because they have almost indistinguishable ligand preferences and kinetic properties by the surface plasmon resonance technique. Unlike the group I and IV WW domains, the group II/III WW domains can bind simple polyproline as well as the proline-leucine and proline-arginine motifs, and they possess two Xaa-Pro (where Xaa is any amino acid) binding grooves, XP and XP2 binding grooves, similar to the SH3 domains. (Table I) by the use of our classification system, shown in Table 1. The ligand specificity of functionally uncharacterized WW domains can be predicted. We have classified the 200 WW domains listed on the EMBL Web site, and have

<table>
<thead>
<tr>
<th>Group</th>
<th>Ligand</th>
<th>Ligand-binding motif</th>
<th>Residues forming the motif</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>PY motif</td>
<td>XP groove</td>
<td>Trp⁴, Tyr²²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tyr-binding pocket</td>
<td>Ile²⁹/Leu/Val²⁵, His²⁷, Lys/Arg/Gln³⁰</td>
</tr>
<tr>
<td>II/III</td>
<td>polyproline</td>
<td>XP groove</td>
<td>Trp³⁰, Tyr²³, Tyr²³, Trp/Tyr²⁵</td>
</tr>
<tr>
<td>PR motif</td>
<td>XP groove</td>
<td>Trp³⁰, Tyr²³, Tyr²³, Trp/Tyr²⁵</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>pS/PT-P motif</td>
<td>XP groove</td>
<td>Trp³⁰, Tyr²³, Tyr²³, Trp/Tyr²⁵</td>
</tr>
<tr>
<td></td>
<td>p patch</td>
<td>Arg/Lys/Asn/Gln³⁰</td>
<td>Arg/Lys/Asn/Gln³⁰, Tyr²³</td>
</tr>
</tbody>
</table>

Table 1. Classification of WW Domains According to Their Ligand Preferences
successfully predicted the ligand specificity for 174 (87%) WW domains; 112 (56%), 54 (27%), and 8 (4%) of WW domains satisfy all the criteria to be classified as groups I, II/III, and IV, respectively, whereas the other 26 (13%) WW domains remain unclassified. 

Before the complex structure of the group II/III WW domain and its ligand peptide was solved, we predicted the ligand recognition mode of the group II/III WW domain that it should bind a ligand forming the polyproline-II (PPII) helix. In this model, if an Xaa-Pro segment at positions 1' and 2' in the PPII helix stacks onto the XP groove, the second Xaa-Pro segment at positions 4' and 5' should come onto the XP2 groove. Our predicted binding model was later proven to be true by the complex structure of the Fe65 WW domain and a polyproline peptide (PDB: 2HO2) (Fig. 4). The two XP grooves between the three ridges formed by the aromatic rings of Trp, Tyr, and Trp/Tyr fit well the ridges and grooves of the PPII helix, as predicted. Ligand recognition by the group II/III WW domains is more closely similar to that by the SH3 domains than to that by the group I and IV WW domains. The group II/III WW domain has two XP binding grooves, an arrangement very similar to the SH3 domain, whereas the group I and IV WW domains have only one XP binding groove. Polyproline, one of the most frequently found motifs in the genomes of fruit fly, worm, and yeast, should act as one of the ligands for the group II/III WW domains and the SH3 domains, but not for the group I and IV WW domains. A large number of polyproline-containing proteins play crucial roles in development, cell movement, cytoskeleton reconstruction, and many other cellular and physiological processes. Therefore, the group II/III WW domain and the SH3 domain should also play critical roles in regulating growth and development.

IV. Concluding Remarks and Perspectives

Our results in the structure-based functional characterization of bombyxin, SBRP, MHH, and the EGF-like domain of HRG-α allow us to identify visually the functional determinants of these molecules. Furthermore, my collaborators and I performed structure-function studies on insulin-like growth factor (IGF-II), acidic fibroblast growth factor (aFGF), midkine, LECT2, the SH2 and SH3 domains, oryzacystatin-I, troponin-C, and S100A13, and some enzymes. We have so far focused on the structure-function relationships of soluble signaling proteins and peptides (ligands) so that we can identify the functionally important sites on the ligand molecules. However, by this strategy, we cannot obtain any information about the receptor molecules. To address this issue, we have started crystallographic studies of ligand-receptor complexes to determine the detailed molecular mechanisms of ligand-receptor interaction and receptor activation.

Acknowledgments

I started structure-based functional studies of peptides and proteins under the supervision of Professor Akinori Suzuki (now Professor Emeritus) at the University of Tokyo. I would like to express my sincere thanks to my bosses, Professors Akinori Suzuki at the University of Tokyo, Fuyuhiko Inagaki at Tokyo Metropolitan Institute of Medical Science, Masaru Tanokura at the University of Tokyo, So Iwata at Uppsala University and Imperial College London, and Hisakazu Yamane at the University of Tokyo. I am also grateful to all my collaborators and laboratory members, especially, Professors Hironori Ishizaki, Akira Isogai, Hiromichi Nagasawa, Shogo Matsumoto, Hiroshi Kataoka, Daisuke Kohda, Shin-ichi Tate, Keiko Abe, Kazuo Suzuki, Makoto Nishiyama, Osato Miyawaki, Minoru Nakano, Drs. Shohei Sakuda, Jiro Nakayama, Kazunori Maruyama, Minoru Tanaka, Hidekazu Katayama, Hideki Hatanaka, Kenji Ogura, Hiroaki Terasawa, Shigeru Tsujiya, Masataka Horiiuchi, Motohiko Nishida, Wakanana Iwasaki, Hiroshi Sasaki, Rintaro Suzuki, Mie Ito, Yusuke Kato, Yoriko Sawano, Norio Kudo, Fumiaki Yamoto, Fabiana Lica Imai, Momi Iwata, Ken Yokoyama, Bernadette Byrne, Mika Jormakk, and Tatsuro Shimamura. This research was partly supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, by the National Project on Protein Structural and Functional Analyses of the MEXT, Japan, and by the program JSPS Postdoctoral Fellowship for Research Abroad. Finally, I thank the committee members of The Japan Bioscience, Biotechnology, and Agrochemistry Society for their nomination of me for the Young Scientists Award in 2007.

References


