Acc. Chem. Res. **2003,** *36,* 199−206

Structural Genomics of Membrane Proteins

YOSHIMASA KYOGOKU,*
YOSHINORI FUJIYOSHI,§ ICHIO SHIMADA,†
HARUKI NAKAMURA,# TOMITAKE TSUKIHARA,‡
HIDEO AKUTSU,‡ TAKAYUKI ODAHARA,
TETSUJI OKADA, AND NOBUO NOMURA

Biological Information Research Center (BIRC), National Institute of Advanced Industrial Science and Technology (AIST), and Japan Biological Informatics Consortium (JBIC), 2-41-6 Aomi, Koto-ku, Tokyo 135-0064, Japan

Received July 3, 2002

ABSTRACT

A program on the structural genomics of membrane proteins has started at the BIRC, AIST, involving other academic institutions and industrial companies. Emphasis is being put on the development of techniques for the structural determination of membrane proteins of biological importance and ligand−receptor interactions by means of electron microscopy, X-ray diffraction, NMR, and computer simulation. Most efforts at the present stage, however, are being directed to finding suitable expression and purification systems and crystallization conditions for such proteins. The program is expected to be linked with the human full-length cDNA project and should lead to medical and industrial uses.

1. Introduction

One of the most popular areas of research in the postgenomic era is the high-throughput structural determination of proteins based on genome information. Most projects aim to determine the structures of as many soluble proteins or soluble parts of proteins as possible. About 25−30% of proteins, however, are embedded in membranes. The latter include receptors, channels, transporters, energy converters, etc., which have significant biological functions. Their structures should be determined, not only to explain biological functions but also to develop new drugs and other industrial uses, because about 60% of newly developed drugs are related to membrane proteins.[1] However, these proteins are insoluble in aqueous solutions and hard to overexpress, purify, and crystallize. Thus, only 2% of the structures of proteins deposited in the Protein Data Bank belong to the category of membrane proteins. We cannot apply high-throughput structure analysis methods to membrane proteins.

The Biological Information Research Center, AIST, opened in 2001 for postgenomic research in collaboration with the Japan Biological Informatics Consortium, which consists of 90 industrial companies. There are three research groups in the center: the structural genomics group focuses on the structural determination of membrane proteins; the functional genomics group devotes itself to the functional analysis of human full-length cDNAs; and the integrated database group has started to annotate the human genome sequence database. Here we will describe the research of the structural genomics group, emphasizing the development of techniques for the structural determination of membrane proteins.

2. Structural Determination of Membrane Proteins by Electron Microscopy

2.1. Characters of Membrane Proteins. While the number of determined structures of soluble proteins is increasing dramatically, the structures of only about 40 membrane proteins have been solved so far at high resolution by means of either X-ray or electron crystallography. This can be attributed to the following problems: (i) it is hard to find natural sources from which a membrane protein can be purified in a sufficient amount; (ii) it is difficult to find an effective overexpression system; and (iii) membrane proteins tend to be denatured on extraction and purification. The third problem is serious for 3D crystallization, but less serious for electron crystallography of 2D crystals, since membrane proteins are reconstituted to two-dimensional membrane forms even after solubilization, which keeps the sample relatively stable. The single-particle analysis technique provides a chance to reveal a three-dimensional structure without the need to make 2D crystals and has been applied to sodium and calcium channels.[2,3] The resolution with this analysis technique, however, has not been sufficient to obtain an atomic model so far.

* To whom correspondence should be addressed.
§ Also a member of the Department of Biophysics, Graduate School of Science, Kyoto University.
† Also a member of the Graduate School of Pharmaceutical Sciences, The University of Tokyo.
# Also a member of the Institute for Protein Research, Osaka University.
‡ Permanent address: Institute for Protein Research, Osaka University.

Yoshimasa Kyogoku is the Director of BIRC, AIST, the Leader of the Structural Genomics Group, and a Professor Emeritus of Osaka University. His research interests are in the fields of structural biology and molecular biophysics.

Yoshinori Fujiyoshi is the Leader of the Structure Analysis Team, BIRC, AIST, and a Professor of Kyoto University. He is engaged in electron crystallography.

Ichio Shimada is the Leader of the Molecular Recognition Analysis Team, BIRC, AIST, and a Professor of the University of Tokyo. He works in the fields of bio-NMR spectroscopy and structural biology.

Haruki Nakamura is the Leader of the Structural Bioinformatics Team, BIRC, AIST, and a Professor of Osaka University. His major interests are in protein science and bioinformatics.

Tomitake Tsukihara is a Professor of Osaka University, where he works on protein X-ray crystallography.

Hideo Akutsu is a Professor of Osaka University. He works in the fields of solid-state NMR spectroscopy and structural biology.

Takayuki Odahara is a Principal Researcher, Structure Analysis Team, BIRC, AIST. He is currently engaged in research on the crystallization of membrane proteins.

Tetsuji Okada is a Principal Researcher, Molecular Function Analysis Team, BIRC, AIST. His research interest is in the structural and functional analysis of rhodopsin.

Nobuo Nomura is the Deputy Director of BIRC, AIST, and the Leader of the Functional Genomics Group. He was a Senior Researcher at the Kazusa DNA Research Institute and works in the fields of functional genomics and molecular biology.

10.1021/ar0101279 CCC: $25.00
Published on Web 11/27/2002

2.2. Expression and Purification of Membrane Proteins. Structural analysis of membrane proteins is strongly linked to methodological innovations, especially to recombinant DNA techniques. Expression systems involving prokaryotic membrane proteins are rarely useful for mammalian ones. Good candidates for this purpose are Sf9 insect cells. We applied this system to express human endothelin B receptor (ET_BR), a G protein-coupled receptor (GPCR). Procedures have been described in detail by Doi et al.,[4] who constructed mutants of ET_BR. For large-scale preparation, the receptor proteins are expressed in insect cells grown in 3 L of medium. Significant amounts of ET_BR with and without ligands were purified with the respective elution systems together with avidin and Ni affinity column chromatography. The final yields were about 0.3 and 0.6 mg, respectively, the ligand binding activity as well as the G protein coupling activity being preserved.

2.3. Two-Dimensional Crystallization. An excellent review of 2D crystallization of membrane proteins is available.[5] The most important but difficult step is selection of the best detergent for crystallization. Generally, a detergent with a lower CMC value is mild and prevents the denaturation of membrane proteins. However, it solubilizes them less effectively from membrane patches and causes loss of proteins. More seriously, the remaining lipid molecules might prevent the proteins from growing into crystals of high crystallinity. Therefore, detergent selection requires many trials. Also for the crystallization of membrane proteins, we have to perform many trials to determine the optimum conditions to obtain high-resolution crystals. To avoid such tedious trials, an automatic electron microscope and a high-resolution and high-contrast optical microscope should be developed. The former would automatically search EM grids to find better 2D crystallization conditions, and the latter would enable us to search crystallization conditions with a similar manner of 3D crystallization.

Small bacteriorhodopsin (bR) crystals have been enlarged by means of a fusion technique, because larger crystals enable us to analyze structures at higher resolution. The bR crystals, which are also called "purple membrane", could be isolated from *Halobacterium salinarium* without solubilization with a detergent.[6] A small crystal of about 100 nm in diameter could be fused to larger patches, more than 5 μm in diameter, by use of a detergent mixture of *n*-octyl-β-D-glucoside (OG) and cationic dodecyltrimethylammonium chloride (DTAC). The most important task for making 2D crystals might be the development of a new crystallization technique that makes it possible to grow crystals in membranes without the use of detergents at high concentration.

2.4. Electron Crystallography. For high-resolution data collection, 2D crystals are embedded in a thin layer of amorphous ice and/or a sugar solution. A sugar, trehalose, is recommended for this purpose. The specimen is then mounted on the cryostage of an electron microscope using a cryotransfer device. The resolution of an image of a biological macromolecule is usually limited to a value much larger than 3 Å. This limitation is due not to the resolution of the instrument but to damage to the specimen by the electron beam. The irradiation damage to biological crystals was found to be reduced to about 1/10 and 1/20 of the value at room temperature when the specimen was cooled to below 20 and 8 K, respectively.[7] In 1991, we developed a superfluid helium stage that can achieve a resolution of 2.6 Å.[8] However, an instrument yielding better resolution would be highly beneficial, because biological molecules consist mainly of light atoms that exhibit small atomic scattering factors in a high-resolution range. Resolution higher than 2.5 Å might be required to identify water molecules. We then improved the instrumental resolution and overcome the operational difficulties. Eventually, a fifth-generation electron cryomicroscope equipped with an automatic cryotransfer system, a field emission gun, an omega filter, and a 4K × 4K slow-scan CCD camera was developed and installed at BIRC (Figure 1).

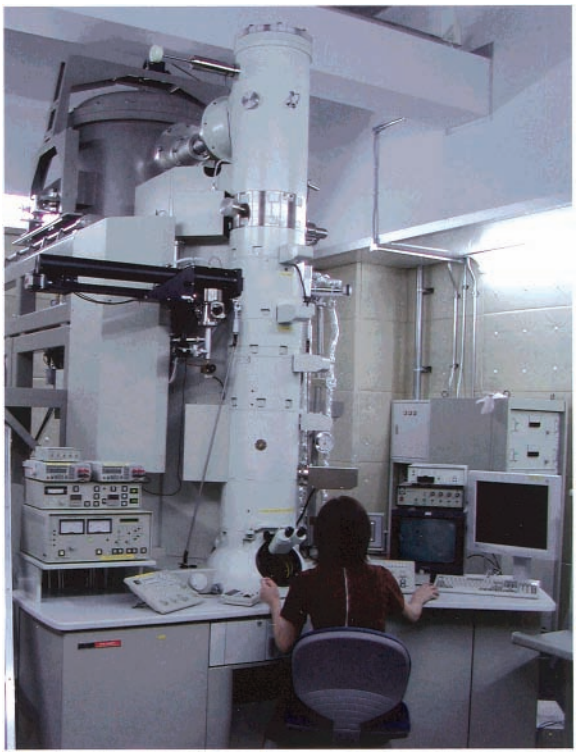

FIGURE 1. A fifth generation electron cryomicroscope equipped with an automatic cryotransfer system, a field emission gun, an omega filter, and a 4K × 4K slow-scan CCD camera was developed and installed at BIRC, AIST.

2.5. Structural Analysis of Bacteriorhodopsin. The computer programs used for the analysis of electron crystallographic data were mainly developed by Henderson and co-workers.[9] An atomic model of bR was built into the experimental 3D density map, which was calculated from 129 electron micrographs and 366 electron diffraction patterns.[10] The resolution of 3.0 Å was achieved by using an electron microscopic system equipped with a liquid helium stage.[11] The experimental map revealed eight lipid molecules related to one bR protomer (Figure 2). On crystallographic refinement, these lipid molecules were modeled as phosphatidyl glycerophosphate mono-

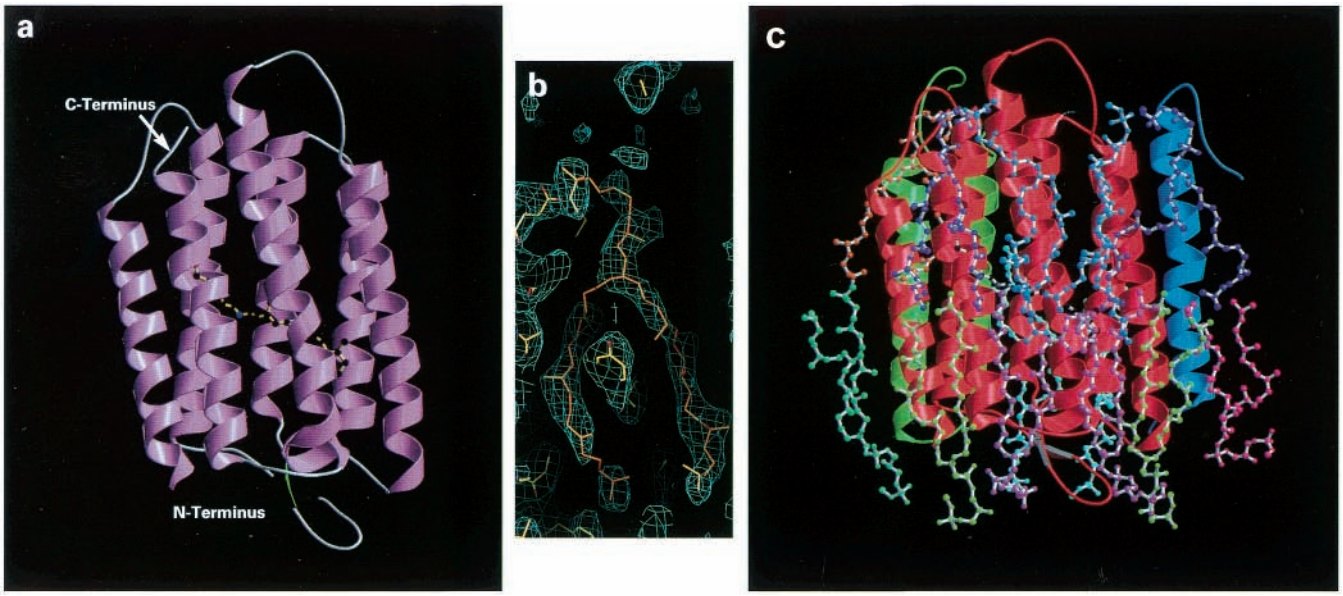

FIGURE 2. Structure of bacteriorhodpsin analyzed at 3.0 Å resolution by electron crystallography. (a,b) An overall view of a bR monomer and a typical map of lipid molecules on the cytoplasmic side. (c) Ball-and stick models of lipid molecules superposed on the bR structure including bR helices (green and blue), which directly interact with the central bR monomer (red).

methyl ester with dihydrophytol chains, which is the major phospholipid found in intact purple membranes. Very recently, the structure of bR was analyzed at a resolution of 2.5 Å, and water molecules in bR were detected by obtaining about 600 electron diffraction patterns for bR crystals at pH 10, and thus the map obtained with the previous data was improved. Another successful result is that for aquapolin 1 from human blood cell membranes, which was resolved at 3.8 Å[12] and is being refined at higher resolution.

3. Structure Determination of Membrane Proteins by X-ray Crystallography

3.1. Crystallization of Membrane Proteins. To obtain the 3D crystals of membrane proteins, we have to obtain a stable purified protein in a large quantity. Membrane proteins themselves are unstable in either an organic solvent or aqueous solution, because they have both hydrophilic and hydrophobic surfaces. Any purified membrane protein is stable in an aqueous solution when the hydrophobic surface is covered with detergent molecules. An adequate detergent for the isolation of each membrane protein is generally chosen by means of a trial-and-error method. The integrity of the conformation of a protein is much more important than the chemical purity of the protein. Slightly denatured protein molecules greatly interfere with crystallization of the protein. Trials to completely remove impurities often cause denaturation. Sometimes, purer proteins obtained by repeated chromatographic separations fail to crystallize.

The most important interaction in the crystal lattice of membrane proteins is the one between the hydrophilic surfaces of the proteins, since only this interaction is specific to the protein. Detergent molecules on the hydrophobic surfaces are unlikely to participate in any specific and strong interactions between the protein molecules in a crystal. On the basis of these considerations, a strategy for the crystallization of membrane proteins has been proposed.[13] According to it, membrane proteins should be solubilized with small-size detergents in order to promote interaction between the hydrophilic surfaces of adjacent protein molecules. However, successful crystallization conditions are not always consistent with this theory, as shown below.

Mitochondrial cytochrome *c* oxidase is isolated from the mitochondrial inner membrane by means of cholic acid. The acid should be replaced with other detergents for crystallization. The purified enzyme was stabilized with Brij-35[$CH_3(CH_2)_{11}O(OCH_2CH_2O)_{23}H$], which provides a hexagonal bipyramidal crystal at low ionic strength.[14] Neither a longer ethyleneglycol unit of 23 nor a shorter one of 6 gives a crystal.[15] This indicates that the size of detergents is critical for the crystallization of membrane proteins. Finally, crystals exhibiting diffraction of higher than 2.8 Å resolution were obtained when decyl maltoside, as a stabilizing agent, and polyethylene glycol, as a precipitant, were employed.[16] Figure 3 shows the molecular packing of an enzyme in a crystal lattice. Only a few intermolecular interactions among those stabilizing the crystal packing contribute to high-resolution diffraction. Since a large volume, up to 70%, of the unit cell is occupied by solvent and detergent molecules, even such a large detergent molecule as decyl maltoside does not prevent the enzyme molecules from forming a highly ordered lattice.

3.2. Systematic Survey of Crystallization Conditions for Membrane Proteins. Most crystals of integral membrane proteins have been obtained by means of an empirical search for crystallization conditions based on the trial-and-error method, although some rules are found in a specific case, as above. However, such rules often fail for another membrane protein. To overcome this situation

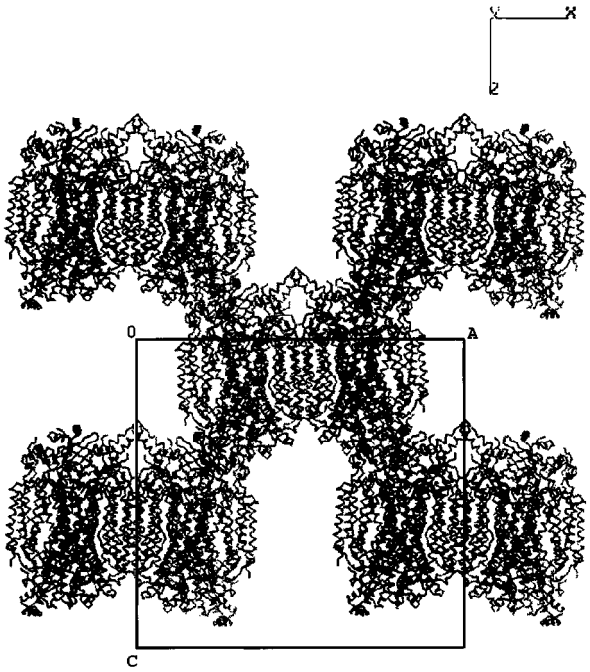

FIGURE 3. Molecular packing of mitochondrial cytochrome *c* oxidase projected on the *ac* plane. Five dimeric enzyme molecules and a unit cell are shown. The top and bottom ends of the molecule are extramembrane parts with a hydrophilic property. Transmembrane helices with a hydrophobic property form the middle part of the molecule. Only the extramembrane parts of the molecule interact with each other to stabilize the crystal structure. The transmembrane surface is not in contact with any other molecule.

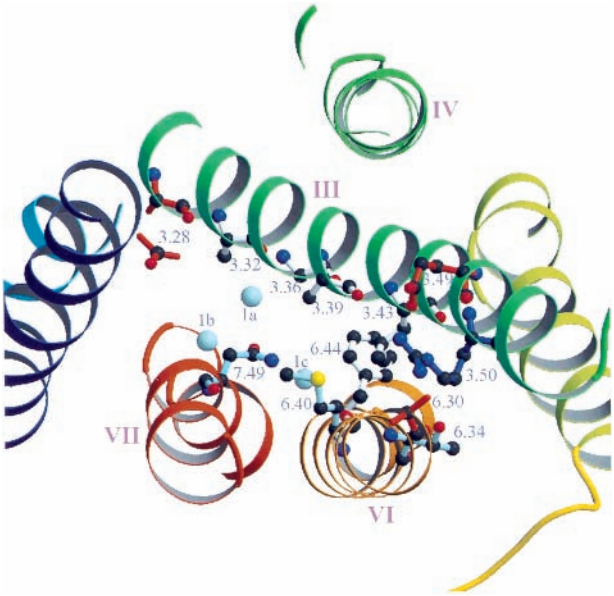

FIGURE 4. Constitutively active mutants (CAM) known for rhodopsin-like GPCRs. A projection view of the seven transmembrane helices of bovine rhodopsin from the cytoplasmic side is shown. The positions of known CAM residues in rhodopsin-like GPCRs are shown as ball-and-stick figures. The three water molecules (1a, 1b and 1c) of bovine rhodopsin at 2.6 Å are also shown as light blue spheres.

and to obtain rational guidelines for crystallization, it is necessary to derive useful common features of the crystallization conditions that were previously successful. So far, we have systematically obtained solubility diagrams for several integral membrane proteins with different detergents, precipitants, pH, and temperatures. It was shown that the prediction of the solubility of an amorphous precipitate, which provides useful information for the choice of crystallization conditions, is possible. Once solubility diagrams with multiple dimensions were determined, the obtained guidelines for the crystallization were applied to the reaction center with light-harvesting proteins, and 10 different crystallization conditions were obtained.[17] We are continuously accumulating solubility diagrams to establish rational guidelines.

3.3. Development of a Synchrotron X-ray Diffraction Beam Line for Microcrystals. Membrane proteins often form small crystals with a size of several micrometers. Assuming a cell constant of 100 Å, a crystal with a size of 1 μm can diffract X-rays since it contains 100 unit cells along the cell edge. The standard beam size of BL44XU at the SPring-8 in Harima, Japan, is $50 \times 50\ \mu m^2$, which is the smallest among the beam lines for protein crystallography at the SPring-8. When a small crystal with a size of several micrometers is exposed to the X-rays of the standard beam of BL44XU, only a small part of the X-rays hits the crystal, most of the X-rays contributing to the background on detection. A focusing technique is usually applied to obtain a small beam with a high intensity. However, focusing increases the divergence of X-rays, and too much focusing decreases the signal-to-noise ratio of the diffraction intensity. A collimator is another apparatus for reducing the beam size. Although the low divergence of X-rays is preserved after the X-rays pass through a collimator, the total intensity of the beam decreases in proportion to the beam size. Another experimental technique to be developed for microcrystal experiments is a method for locating a tiny crystal at the beam position coincident with the rotation center of the crystal.

3.4. A Current Project on Membrane Proteins—Rhodopsin as a Model of GPCRs. Rhodopsin is a photoreceptor membrane protein containing 11-*cis*-retinal, an intrinsic inverse agonist, which is buried inside the seven-transmembrane (TM) helical bundle. The crystal structure of rhodopsin recently determined[18] from bovine retinal rod cells thus represents a prototypical template for the quiescent form of GPCR.[19] The ground-state structure of rhodopsin has been refined to 2.6 Å, at which water molecules in the TM region can be well assigned (Figure 4).[20] Success in its three-dimensional crystallization and X-ray structure determination thus demonstrated the possibility of obtaining atomic models of other GPCRs. A majority of GPCRs belong to the so-called rhodopsin family that is characterized by a number of conserved residues and motifs in the TM domain. The heptahelical TM bundle links a ligand-induced conformational change on the extracellular side of a receptor to the cytoplasmic surface, where activation of heterotrimeric G protein molecules is mediated. Thus, these conserved residues are also presumed to play critical roles in structural rearrangement evoked upon activation of the receptor. Details of the molecular mechanism of the structural change will be revealed by solving the structures of photoreaction intermediates of bovine rhodopsin in the three-dimen-

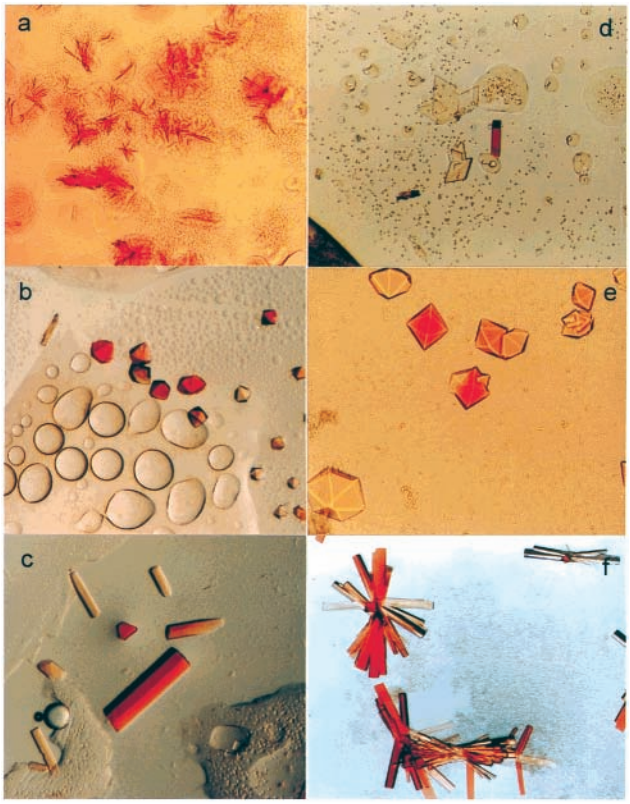

FIGURE 5. Three-dimensional crystals of bovine rhodopsin. (a) Needles and thin plates obtained using a contaminated sample. (b–d) Crystals obtained under phase separation. (e,f) Crystals obtained in the isotropic mother liquor. These crystals were obtained successively, from a to f, during the course of optimization.

sional crystal. The bimolecular complex between the activated form of rhodopsin and the G protein or its partial peptide should also be investigated.

Rhodopsin and related photoreceptor proteins are excellent sources of targets for obtaining valuable information on structure–function relationships in TM receptors by X-ray crystallography, because procedures for overexpression, purification, and crystallization have become available (Figure 5).[21] Various interesting mutants of rhodopsin should also be subjected to structure determination. The major focus of attention should be directed to the fact that it has been known to exhibit so-called constitutive activity without light stimulation. All these methodological advances made on the basis of the study on photoreceptor proteins are about to be applied to the members of the GPCR family as diffusible ligands that are pharmacologically very important.

4. Membrane Proteins Studied by NMR

4.1. A Method for Surveying the Interaction Surface—Transferred Cross-Saturation. High-resolution NMR is often used for structure determination of proteins. Recent advances in TROSY-based measurements have made it possible to determine the structure of membrane proteins.[22] However, the sample should be soluble, and its molecular weight is limited up to about 30K for detailed NMR analyses. These limitations are quite severe for us

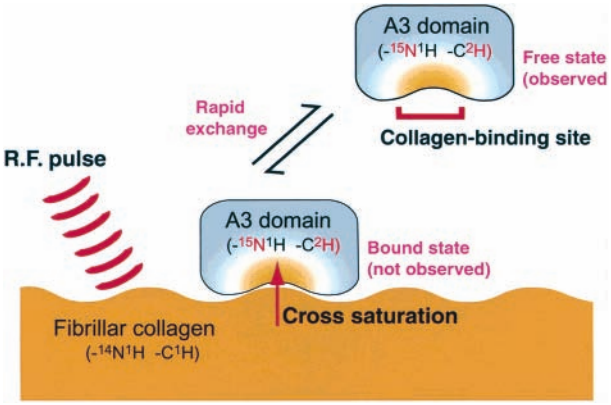

FIGURE 6. Schematic illustration of transferred cross-saturation experiments for the fibrillar collagen and the A3 domain of von Willebrand factor.

to use high-resolution NMR for structure determination of membrane proteins. Thus, we focus on using NMR to investigate the ligand-recognition mechanism of membrane proteins at an atomic resolution.

Recently, we proposed a novel NMR measurement, termed "cross-saturation", which enables the identification of the contact residues in a huge protein complex in a rigorous manner.[23] The ligand protein for which we want to identify the interface residues is uniformly labeled with ^{2}H and ^{15}N, and then is allowed to form a complex with a nonlabeled receptor protein. Although the protein uniformly labeled with ^{2}H and ^{15}N is not directly affected by the radio frequency field, it is expected that the saturation can be transferred from the receptor protein to the doubly labeled ligand protein through the interface of the complex by cross-relaxation. Therefore, we can identify the residues at the interface by observing the reduction of the peak intensities in the $^{1}H-^{15}N$ HSQC spectra measured through TROSY coherence transfer.

However the cross-saturation method is difficult to apply to protein complexes with molecular weights over 150K. To overcome this limitation, we have developed another version of the cross-saturation measurement, termed "transferred cross-saturation" (TCS).[24] Under the conditions of an excess amount of the ligand protein relative to the target protein, and a fast exchange rate between the free and bound states of the ligand protein, the cross-saturation effect that occurs in the bound state of the ligand protein could be efficiently observed. Our experimental evaluation for the efficiency of TCS phenomena revealed that TCS experiments are applicable for systems in which the dissociation rate is greater than 0.1 s^{-1}. As shown in Figure 6, we applied TCS to a blood coagulation system, the A3 domain of von Willebrand factor (vWF) and the fibrillar collagen. On the basis of the data from TCS, it is indicated that the binding site of the A3 domain was located at the hydrophobic "front" surface and totally different from that of the I domain from the α2 subunit of integrin (α2-I domain), which was previously reported, although the A3 domain and the α2-I domain fold similarly and possess the identical collagen binding function (Nishida, N., et al. *Nat. Struct. Biol.*, in press).

Since the resonances originating from the ligand protein in the free state are used in the TCS experiments to identify the contact residues in the complex, the method would be easy to apply to protein−protein interactions, such as those between ligand and receptor proteins in bilayer lipids, and also those between ligand proteins and cells.

4.2. Solid-State NMR for Structure Analysis of Membrane Proteins. Solid-state NMR has no limitation as to molecular weight measurements. Furthermore, magic angle sample spinning (MASS) has enabled us to obtain high-resolution spectra of solid samples. However, it is necessary to recouple a particular interaction under MASS conditions to obtain structural information, as proposed by Griffin et al.[25] and Gullion and Schaefer.[26] The rotational resonance (RR) method is widely used to determine a distance in a solid sample specifically labeled at two positions. We have modified the constant time RR method[27] for application to membrane systems. Mastoparan X, which is a wasp venom and is known to activate a G-protein, was used as a target sample. Selectively labeled mastoparan X, as shown below, was bound to DPPC−

Ile-Asn-Trp-Lys-Gly(^{13}CO)-Ile-Ala-Ala($^{13}CH_3$)-
Met-Ala-Lys-Lys-Leu-Leu-NH_2

DPPG bilayer membranes under hydrated conditions. Using the constant time RR pulse sequence, the magnetization transfer from Gly5(^{13}CO) to Ala8($^{13}CH_3$) through the dipole−dipole interaction was measured. The internuclear distance was determined by analyzing simulation of the decay curve.

To determine a tertiary structure, it is convenient for us to obtain multiple structural parameters from a single sample. The combination of multidimensional NMR under MASS conditions and uniformly isotope-labeled samples is useful for this purpose.[28,29] The strategy for the structural determination should be similar to that of solution NMR. Uniformly ^{13}C- and ^{15}N-labeled mastoparan X was expressed in *Escherichia coli* as a fusion protein with ubiquitin. The intraresidue correlations of the backbone and side-chain signals were determined in 2- and 3D NMR experiments with dipolar mixing periods for directly bonded spins. The magnetization transfer takes place through peptide bonds via ^{15}N. These experiments made sequence-specific assignments of most signals possible. The chemical shifts of the assigned CO, C_α, and C_β signals indicate that mastoparan X forms an α-helix. The correlation between anisotropic interactions of different spins specifies dihedral angles.[29] Since all elemental techniques have been established, we are now trying to determine the tertiary structure of mastoparan X bound to phospholipid bilayers.

5. Structural Bioinformatics of Membrane Proteins and Ligand Binding

For effective application of structural information on membrane proteins, a breakthrough is strongly required in the technology for in silico screening and structure-based drug design. One of the target technologies is the development of high-accuracy and high-speed stereo-structure modeling for membrane proteins and other proteins including flexible loop regions through the preparation of basic software. Another target is the development of quantitative interaction simulation aiming at practical use of in silico screening, considering the structural changes of both proteins and ligands.

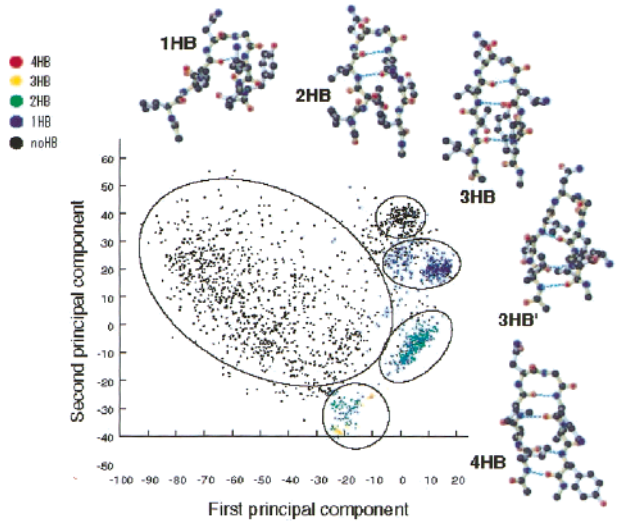

FIGURE 7. Conformational distribution of the Boltzmann−Gibbs ensemble at 300 K for a peptide [Ace-Ile-Thr-Val-Asn-Gly-Lys-Thr-Tyr-Nme] designed for a β-hairpin in explicit water, plotted against principal axes *a1* and *a2* obtained by principal component analysis.[32] Here, Ace and Nme are the N-terminal acetyl and C-terminal *N*-methyl groups, respectively. The peptide conformations are classified as to the numbers of typical intramolecular hydrogen bonds, *n*, which are indicated by different colors: $n = 4$, red; $n = 3$, yellow; $n = 2$, green; $n = 1$, blue; and $n = 0$, black. Typical conformations corresponding to the individual clusters are shown.

For both targets, we developed basic algorithms for enhanced conformational sampling, with which we can compute the free energy values of an isolated protein and the complex of the protein with a ligand. This approach has been successfully applied for the conformational prediction of short peptides,[30−32] as shown in Figure 7, and for the preliminary simulation of the flexible docking between the SH3 protein and a Pro-rich peptide.[33] Thus, structural models and ligand screening should be assessed by means of simulated free energy values. So far, we have developed a new algorithm for Tsallis dynamics and a method for constructing a Boltzmann−Gibbs ensemble from the Tsallis ensemble.[34] Another new technology has also been applied to multicanonical ensembles by means of the temperature scaling method.

Using these basic algorithms, we have developed new programs, prestoX (eXtended version of PRESTO[31]) and TPLgene (ToPoLogy Generating EnGinE). The former is for the molecular simulation of membrane proteins (Figure 8) with several enhanced sampling methods (multicanonical molecular dynamics, replica exchange method, and Tsallis dynamics), and the latter is for the construction of the complete molecular topology and for the assignment of the force field parameters to individual

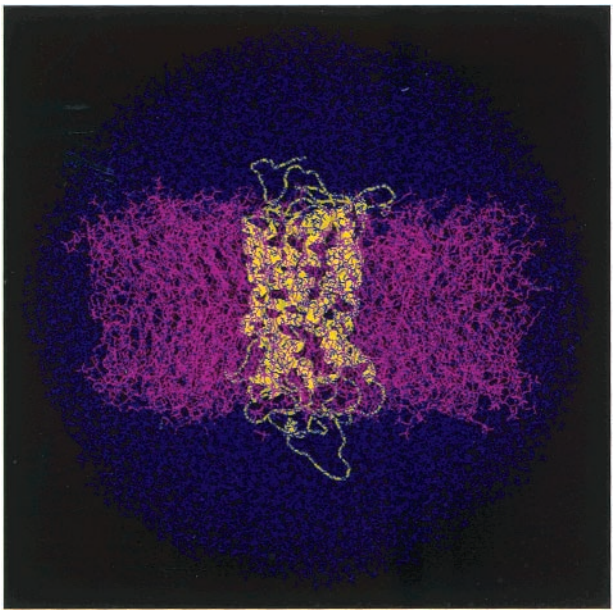

FIGURE 8. Molecular model of bovine rhodopsin for dynamic simulations with the phospholipid and water molecules.

potential energy terms. As an actual application, we made structural models of the cytoplasmic loops of bovine rhodopsin, the structures of which were not previously determined because of their flexible nature. Since these loops interact with a G-protein, the structural models, as an ensemble of several putative conformations, should provide a new insight for drug design.

6. Human Full-Length cDNA Clones and Their Use as Resources for Proteomics of Membrane Proteins

6.1. Human Full-Length cDNA Projects. The generation of a set of full-length cDNAs, in terms of both sequences and physical clones, which contains the entire and uninterrupted protein coding regions of all human genes, is one of the key issues for the systematic and comprehensive analysis of protein structure and function. The Japanese FLJ cDNA project (Sugano, S., et al., in preparation), Kazusa DNA Research Institute,[35] German cDNA project,[36] and NIH Mammalian Gene Collection team[37] are major groups that have been making enormous efforts toward the systematic collection and complete sequencing of cDNA clones.

The FLJ cDNA project, which is a METI-funded research program for large-scale archiving and sequencing of putative human full-length cDNA clones, was initiated in 1998 and finished in the spring of 2002. In this project, cDNA libraries were constructed by means of a cap-targeted selection procedure for full-length cDNAs.[38,39] With the use of the cDNA libraries, the efficiency of isolating full-length cDNA clones greatly increased. cDNA clones thus far selected have been subjected to complete sequencing. So far, the DNA sequences of 21 243 clones have been deposited in GenBank/EMBL/DDBJ. The results of various in silico analyses of these cDNA clones are available from the web sites of the Helix Institute (http://www.hri.co.jp), the Kazusa DNA Research Institute (http://www.kazusa.or.jp), and the Institute of Medical Science, the University of Tokyo (http://cdna.ims.u-tokyo.ac.jp).

6.2. Protein Factory: Infrastructure for the Supply of Expressed Proteins. To obtain gene products derived from the FL cDNAs described above, a new multipurpose cloning system, the Gateway cloning system (Invitrogen, Co.), has been introduced. It does not require the traditional use of restriction endonucleases and ligases for cloning of the DNA. Alternatively, it utilizes the att signal DNA sequences and unique enzymes cloned from a lambda phage for site-specific in vitro recombination. Starting from the entry clone, a large number of genes (cDNA) can be transferred to many different expression vectors in a short period, maintaining reading frames and clone orientations. Therefore, the Gateway technology is very suitable for high-throughput production of expression plasmids for large-scale analysis. So far, 15 000 entry clones have been produced. A destination vector, which is the supplier of a promoter, the N-terminal or C-terminal tag to the expression clone, can be designed for many protein-expression systems and fusion proteins.

We have initiated a program named the "Protein Factory" in which, based on Gateway entry clones, protein expression is executed in various expression systems including *E. coli* (both in vivo and in vitro), *Bacillus brevis* (in vivo), wheat germ (in vitro), insect (both in vivo and in vitro), and mammalian cell lines. About 10 Japanese companies are participating in this program to express proteins for functional and structural analyses of the proteins derived from cDNAs.

7. Conclusion

Here we described in part the experience and planning of our program. The most serious bottlenecks are the steps of overexpression, purification, and crystallization of membrane proteins. It is hard to define the structural genomics of membrane proteins. If structural genomics implies high-throughput and systematic structure determination, then that of membrane proteins has not yet arrived at this stage. Pinpoint attack is inevitable now. Thus, we have to select functionally interesting targets. However, at this time, when genome sequence information is available, we have to correlate the structural and interaction information on proteins with the sequence information. For this approach, we have to pay much more attention to the high-throughput activities of the functional genomics group and correlate the structural results with theirs. Then our activities can be regarded as structural genomics.

This work was supported by grants from the Ministry of Economics, Trade and Industry (METI) and the New Energy and Industrial Technology Development Organization (NEDO), Japan.

References

(1) Drews, J. Drug Discovery: A Historical Perspective. *Science* **2000**, *287*, 1960−1964.

(2) Sato, C.; Ueno, Y.; Asai, K.; Takahashi, K.; Sato, M.; Engel, A.; Fujiyoshi, Y. The Voltage-sensitive Sodium Channel is a Bell-shaped Molecule with Several Cavities. *Nature* **2001**, *409*, 1047−1051.
(3) Murata, K.; Odahara, N.; Kuniyasu, A.; Sato, Y.; Nakayama, H.; Nagayama, K. Asymmetric Arrangement of Auxiliary Subunits of Skeletal Muscle Voltage-gated L-type Ca^{2+} Channel. *Biochem. Biophys. Res. Commun.* **2001**, *282*, 284−291.
(4) Doi, T.; Hiroaki, Y.; Arimono, I.; Fujiyoshi, Y.; Okamoto, T.; Satoh, M.; Furuichi, Y. Characterization of Human Endothelin B Receptor and Mutant Receptors Expressed in Insect Cells. *Eur. J. Biochem.* **1997**, *248*, 139−148.
(5) Kuehlbrandt, W. Two-dimensional Crystallization of Membrane Proteins. *Q. Rev. Biophys.* **1992**, *25*, 1−49
(6) Oesterhelt D.; Stoeckenius, W. Isolation of the Cell Membrane of *Halobacterium halobium* and Its Fractionation into Red and Purple Membrane. *Methods Enzymol.* **1974**, *31*, 667−678.
(7) Fujiyoshi, Y. The Structural Study of Membrane Proteins by Electron Crystallography. *Adv. Biophys.* **1998**, *35*, 25−80.
(8) Fujiyoshi, Y.; Mizusaki, T.; Morikawa, K.; Yamagishi, H.; Aoki, Y.; Kihara, H.; Harada, Y. Development of a Superfluid Helium Stage for High-Resolution Electron Microscopy. *Ultramicroscopy* **1991**, *38*, 241−251.
(9) Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E.; Downing, K. H. Model for the Structure of Bacteriorhodopsin Based on High-resolution Electron Cryo-microscopy. *J. Mol. Biol.* **1990**, *213*, 899−929.
(10) Kimura, Y.; Vassylyev, D. G.; Miyazawa, A.; Kidera, A.; Matsushima, M.; Mitsuoka, K.; Murata, K.; Hirai, T.; Fujiyoshi, Y. Surface of Bacteriorhodopsin Revealed by High-resolution Electron Crystallography. *Nature* **1997**, *389*, 206−211.
(11) Mitsuoka, K.; Hirai, T.; Murata, K.; Miyazawa, A.; Kidera, A.; Kimura, Y.; Fujiyoshi, Y. The Structure of Bacteriorhodopsin at 3.0 Å Resolution Based on Electron Crystallography: Implication of the Charge Distribution. *J. Mol. Biol.* **1999**, *286*, 861−882.
(12) Murata, K.; Mitsuoka, K.; Hirai, T.; Walz, T.; Agre, P.; Heymann, J. B.; Engel, A.; Fujiyoshi, Y. Structure Determinants of Water Permeation through Aquaporin-1. *Nature* **2000**, *407*, 599−605.
(13) Michel, H. Crystallization of Membrane Proteins. *Trends Biochem. Sci.* **1983**, *8*, 56−59.
(14) Yoshikawa, S.; Tera, T.; Takahashi, Y.; Tsukihara, T.; Caughey, W. S. Crystalline Cytochrome c Oxidase of Bovine Heart Mitochondrial Membrane: Composition and X-ray Diffraction Studies. *Proc. Natl. Acad. Sci. U.S.A.* **1988**. *85*, 1354−1358.
(15) Shinzawa-Itoh, K.; Ueda, H.; Yoshikawa, S.; Aoyama, H.; Yamashita, E.; Tsukihara, T. Effects of Ethyleneglycol Chain Length of Dodecyl Polyethylene glycol Monoether on Crystallization of Bovine Heart Cytochrome c Oxidase. *J. Mol. Biol.* **1995**, *246*, 572−575.
(16) Tsukihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itoh, K.; Nakashima, R.; Yaono, R.; Yoshikawa, S. Structure of Metal Sites of Oxidized Bovine Heart Cytochrome c Oxidase at 2.8 A Resolution. *Science* **1995**, *269*, 1069−1074.
(17) Odahara, T.; Ataka, M.; Katsura, T. Phase Diagram Determination to Elucidate the Crystal Growth of the Photoreaction Center from *Rhodobacter sphaeroides*. *Acta Crystallogr. D* **1994**, *50*, 639−542.
(18) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. Crystal Structure of Rhodopsin: a G Protein-coupled Receptor. *Science* **2000**, *289*, 739−745.
(19) Okada, T.; Ernst, O.; Palczewski, K.; Hofmann, K. P. Activation of Rhodopsin: New Insights from Structural and Biochemical Studies. *Trends Biochem. Sci.* **2001**, *26*, 318−324.
(20) Okada, T.; Fujiyoshi, Y.; Silow, M.; Navarro, J.; Landau, E. M.; Shichida, Y. Functional Role of Internal Water Molecules in Rhodopsin Revealed by X-ray Crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5982−5987.
(21) Okada, T.; Le Trong, I.; Fox, B. A.; Behnke, C. A.; Stenkamp, R. E.; Palczewski, K. X-ray Diffraction Analysis of Three-dimensional Crystals of Bovine Rhodopsin Obtained from Mixed Micelles. *J. Struct. Biol.* **2000**, *130*, 73−80.
(22) Arora, A.; Abildgaard, F.; Bushweller, J. H.; Tamm, L. K.; Structure of Outer Membrane Protein A Transmembrane Domain by NMR Spectroscopy. *Nat. Struct. Biol.* **2001**, *8*, 334−338.
(23) Takahashi, H.; Nakanishi, T.; Kami, K.; Arata Y.; Shimada, I. A Novel NMR Method for the Determination of the Interface of Large Protein−protein Complexes. *Nat. Struct. Biol.* **2000**, *7*, 220−223.
(24) Nakanishi, T.; Miyazawa, M.; Sakakura, M.; Terasawa, H.; Takahashi, H.; Shimada, I. Determination of the Interface of a Large Protein Complex by Transferred Cross-saturation Measurements. *J. Mol. Biol.* **2002**, *318*, 245−249.
(25) Raleigh, D. P.; Levitt, M. H.; Griffin, R. G. Rotational Resonance in Solid State NMR. *Chem. Phys. Lett.* **1988**, *146*, 71−76.
(26) Gullion, T.; Schaefer, J. Rotational-echo Double-resonance NMR. *J. Magn. Reson.* **1989**, *81*, 196−200.
(27) Balazs, Y. S.; Thompson, L. K. Practical Methods for Solid-state NMR Distance Measurements on Large Biomolecules: Constant Time Rotational Resonance. *J. Magn. Reson.* **1999**, *139*, 371−376.
(28) Fujiwara, T.; Sugase, K.; Kainosho, M.; Ono, A.; Ono, A(M).; Akutsu, H. $^{13}C-^{13}C$ and $^{13}C-^{15}N$ Dipolar Correlation NMR of Uniformly Labeled Organic Solids for the Complete Assignment of Their $^{13}C-^{15}N$ Signals: An Application to Adenosine. *J. Am. Chem. Soc.* **1995**, *117*, 11351−11352.
(29) Fujiwara, T.; Shimomura, T.; Ohigashi, Y.; Akutsu, H. Multidimensional Solid-state NMR for Determining Dihedral Angle from the Correlation of $^{13}C-^{1}H$ and $^{13}C-^{13}C$ Dipolar Interactions under Magic-angle Spinning Conditions. *J. Chem. Phys.* **1998**, *109*, 2380−2393.
(30) Shirai, H.; Nakajima, N.; Higo, J.; Kidera, A.; Nakamura, H. Conformational Sampling of CDR−H3 in Antibodies by Multicanonical Molecular Dynamics Simulation. *J. Mol. Biol.* **1998**, *278*, 481−496.
(31) Nakajima, N.; Higo, J.; Kidera, A.; Nakamura, H. Free Energy Landscapes of Peptides by Enhanced Conformational Sampling. *J. Mol. Biol.* **2000**, *296*, 197−216.
(32) Kamiya, N.; Higo, J.; Nakamura, H. Conformational Transition States of a β-hairpin Peptide between the Ordered and Disordered Conformations in Explicit Water. *Protein Sci.* **2002**, *11*, 12297-2307.
(33) Nakajima, N.; Higo, J.; Kidera, A.; Nakamura, H. Flexible Docking of a Ligand Peptide to a Receptor Protein by Multicanonical Molecular Dynamics Simulation. *Chem. Phys. Lett.* **1997**, *278*, 297−301.
(34) Fukuda, I.; Nakamura, H. Tsallis Dynamics Using the Nose-Hoover Approach. *Phys. Rev. E* **2002**, *65*, 26105.
(35) Nomura, N.; Miyajima, N.; Suzaka, T.; Tanaka A.; Kawarabayasi Y.; Sato S.; Nagase T.; Seki N.; Ishikawa K.; Tabata S. Prediction of the Coding Sequences of Unidentified Human Genes. I. The Coding Sequences of 40 New Genes (KIAA0001−KIAA0040) Deduced by Analysis of Randomly Sampled cDNA Clones from Human Immature Myeloid Cell Line KG-1. *DNA Res.* **1994**, *1*, 27−35.
(36) Wiemann, S.; Weil, B.; Wellenreuther, R.; Gassenhuber, J.; Glassl, S.; Ansorge, W.; et al. Toward a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs. *Genome Res.* **2001**, *11*, 422−435.
(37) Strausberg, R. L.; Feingold, E. A.; Klausner, R. D.; Collins, F. S. The Mammalian Gene Collection. *Science* **1999**, *286*, 455−457.
(38) Suzuki, Y.; Yoshitomo, K.; Maruyama, K.; Suyama, A.; Sugano, S. Construction and Characterization of a Full Length-enriched and a 5′-end-enriched cDNA Library. *Gene* **1997**, *200*, 149−156.
(39) Nishikawa, T.; Ota, T.; Kawai, Y.; Ishii, S.; Saito, K.; Yamamoto, J.-i.; Wakamatsu, A.; Ozawa, M.; Suzuki, Y.; Sugano, S.; Isogai, T. Comparison of Sequences of cDNA Clones Obtained from Oligo-capping cDNA Libraries with Those from Unigene. *DNA Res.* **2001**, *8*, 255−262.

AR0101279