PROXIMITY LIGATION: A SPECIFIC AND VERSATILE TOOL FOR THE PROTEOMIC ERA

Ola Söderberg, Karl-Johan Leuchowius, Masood Kamali-Moghaddam, Malin Jarvius, Sigrun Gustafsdottir, Edith Schallmeiner, Mats Gullberg, Jonas Jarvius and Ulf Landegren

University of Uppsala
Department of Genetics and Pathology
Rudbeck laboratory
SE-751 85 Uppsala
Sweden

ABSTRACT

Knowledge about the total human genome sequence now provides opportunities to study its myriad gene products. However, the presence of alternative splicing, post-translational modifications, and innumerable protein-protein interactions among proteins occurring at widely different concentrations, all combine to place extreme demands on the specificity and sensitivity of assays. The choice of method also depends on matters such as whether proteins will be analyzed in body fluids and lysates, or localized inside single cells. In this review we discuss commonly used detection methods and compare these to the recently-developed proximity ligation technique.

INTRODUCTION

The problem can be illustrated by the counting of swans in a pond inhabited by swans (*Cygnus olor*) and ducks (*Anas platyrhynchos*). To discriminate

Genetic Engineering, Volume 28, Edited by J. K. Setlow

between the two species we set up the following criteria characterizing a swan: it is white, and it has a long neck. Either one of these criteria may suffice to determine if a bird in the pond is a swan or not. However, in a more complex environment such as a lake the analysis may result in an overestimate of the number of swans. Seagulls (*Larus canus*) are white birds and fulfill the first criterion. Similarly, the second criterion (long neck) is also applicable to herons (*Ardea cinerea*). Specificity of detection increases greatly, however, if the two criteria are combined.

Turning now to proteomics, antibodies do not recognize whole proteins, but merely epitopes on the proteins, each composed of just a few amino acids. Assays that require positive identification of target proteins by two antibodies specific for different epitopes on the same protein exhibit profoundly enhanced specificity over assays that depend on single-binding events. The strategy of double recognition has in fact been used for almost forty years in sandwich immunoassays for soluble proteins, where one immobilized antibody traps the protein that is then detected by a second, labeled, antibody (1, 2). Such assays can reach pM detection levels.

Antibody-based measurements of target protein concentrations generally involve detection of bound antibodies *via* labels such as heavy metals, radioisotopes, fluorophores (3), or using chemoluminescence or linked enzymes (4). Enzyme-linked detection reactions offer enhanced sensitivity because of the catalytic activity of the enzymes, generating detectable products. It is possible to further enhance detection by attaching strands of DNA to antibodies. Such DNA strands can be exponentially amplified by methods like polymerase chain reaction (PCR) (immuno-PCR) (5), or they can be used to prime a rolling circle amplification (RCA) reaction for localized signal amplification by generating a long concatemeric product, covalently linked to the antibody, to which labeled probes can hybridize (immuno-RCA) (6-8). All these means of detecting bound antibodies fail to distinguish among specifically and nonspecifically bound antibodies, however, limiting the increase of signals over background that can be attained, and thus the assay sensitivity.

Proteins are being analyzed in many different formats including simplex assays of proteins in solution, *in situ* analyses and using protein arrays. Protein microarrays (reviewed by Espina et al., 9) allow simultaneous detection of several different proteins. In one format all proteins in a sample to be analyzed are directly labeled, and then the ones that have bound to antibodies immobilized in an array can be detected after washes. The specificity and thereby sensitivity of such assays are inherently limited by the fact that single-binding-events per target molecule are scored. Alternatively, different immobilized bait molecules capture their cognate proteins to specific locations on an array. Subsequently, the bound proteins can be detected and quantified by binding of specific-labeled-antibodies in a sandwich format. The requirement for dual binding increases specificity, but the parallel assay format creates increased opportunities for crossreactive binding between noncognate pairs of protein-binding reagents compared to single-analyte assays, limiting-assay sensitivity and ultimately the degree of multiplexing that can be achieved. The use of arrays of photoaptamers can provide increased specificity without a concomitant increase in risks of crossreactivity with increasing

numbers of analytes (10). After excess proteins have been removed from arrays of immobilized photoaptamers by washes, specifically-bound proteins can be covalently cross-linked to the aptamers by treatment with UV light. The requirement for the proteins to be correctly positioned for cross-linking introduces an additional level of specificity and allows nonspecifically bound proteins to be removed by extremely stringent washes, leaving only covalently-bound proteins, followed by detection *via* general protein stains.

Western blots achieve increased specificity of detection by distinguishing proteins both according to electrophoretic mobility and by antibody binding to the separated proteins blotted onto a membrane, resulting in nM detection limits. Localized protein detection reactions are also used in immunohistochemistry, where the distribution of proteins in tissues and cells is revealed by the binding of specific-labeled-antibodies. This information represents an invaluable resource in research and diagnostics. *In situ* assays provide increased information over solution-phase quantification assays since heterogeneity among cells can be revealed and spatial relationships among structures are visualized. All current localized detection reactions suffer from the problem that visualization depends on single binding events, however, limiting specificity and sensitivity. It is also a problem that evaluation of *in situ* staining reactions is subjective, and limited to a relatively crude estimation of the degree of staining. Simultaneous phenotyping of multiple cell types can be performed in solution by flow cytometry determining expression levels of several proteins on the surfaces of individual cells. The protein expression pattern in combination with the light scattering properties of the cells enables highly accurate identification of cell types and maturation stages, with the additional advantage that cells also can be sorted for subsequent analysis.

Despite the multitude of formats for protein detection, there clearly remains a pressing need for highly specific detection reactions in order to negotiate protein concentration ranges that may exceed 10 orders of magnitude in serum, and to allow detection of even single proteins in cells and evaluate the company they keep. The choice of method also depends on matters such as whether proteins will be analyzed in body fluids and lysates, or localized inside single cells. In this review we discuss commonly-used detection methods and compare these to the recently-developed proximity ligation technique

PROXIMITY LIGATION

Recently a new and quite general approach to protein detection—proximity ligation—was described. In this technique specific and sensitive detection can be achieved by utilizing a combination of highly-specific target recognition and powerful signal-amplification as required for detection of low abundant proteins (11, 12). The probes used in proximity ligation are composed of an antigen binding part (e.g., an antibody or an aptamer) to which short single-stranded DNA molecules have been conjugated. Upon binding of two such proximity probes to the same target molecule, a subsequently added connector oligonucleotide can hybridize to the ends of the conjugated DNA strands and guide their joining by enzymatic ligation. This creates a DNA molecule that can then be amplified by PCR (Figure 1). Recognition of target molecules by proximity

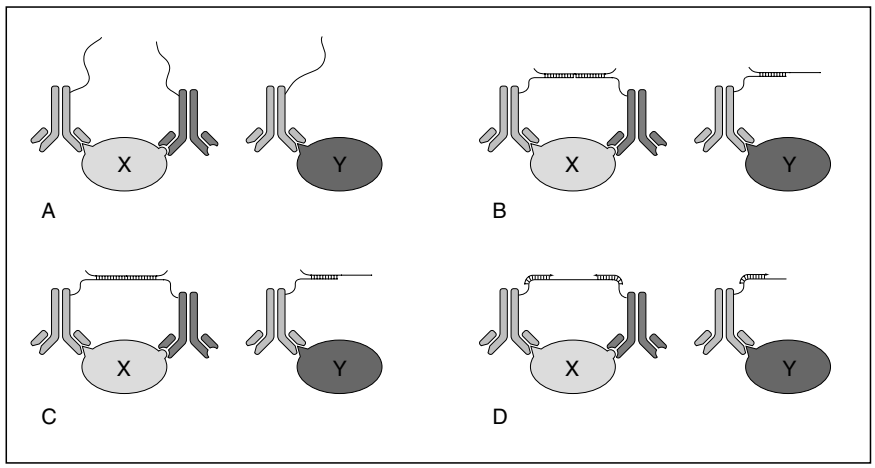

Figure 1. (A) The proximity ligation procedure. Two proximity probes bind protein X, while one probe also crossreacts by binding to protein Y. (B) A complementary connector oligonucleotide is added that hybridizes to the oligonucleotides attached to pairs of adjacent proximity probes, allowing the free oligonucleotide ends to be joined by ligation. (C) Only reagents brought in proximity by binding pairwise to protein X will be ligated together. (D) Addition of PCR primers allows sensitive detection by exponential amplification of ligated proximity probes having bound protein X, but not of unreacted proximity probes.

ligation thus strictly depends on dual recognition in order to generate an amplifiable DNA strand that serves as a surrogate marker for the detected protein. Signal amplification by real-time PCR allows sub-pM levels of proteins to be detected in a homogenous assay that is performed without any washes, just the addition of a ligation/amplification cocktail, followed by amplification and detection (Figure 2). Alternatively, a sandwich format can be used, where the target proteins are first trapped on a solid support *via* specific binding followed by addition of pairs of proximity probes that are joined by ligation. The removal of excess reagents by washes lowers the background from chance proximity by unbound proximity probes and also reduces the concentration of substances that may inhibit ligation, amplification or detection. The assay involves three recognition events of any target molecule, further increasing the ability to discriminate among closely similar protein molecules.

By virtue of the presence of many copies of the same proteins on their surfaces, even single viral particles or bacteria have been successfully detected using the proximity ligation mechanism (13). It is also possible to design homogenous assays that require three recognition events and two ligations. As a consequence of the reduced chance for proximity of three rather than two reagents, and the increased biological specificity of the three binding events, detection levels of just a few hundred molecules have been achieved (Schallmeiner et al., submitted).

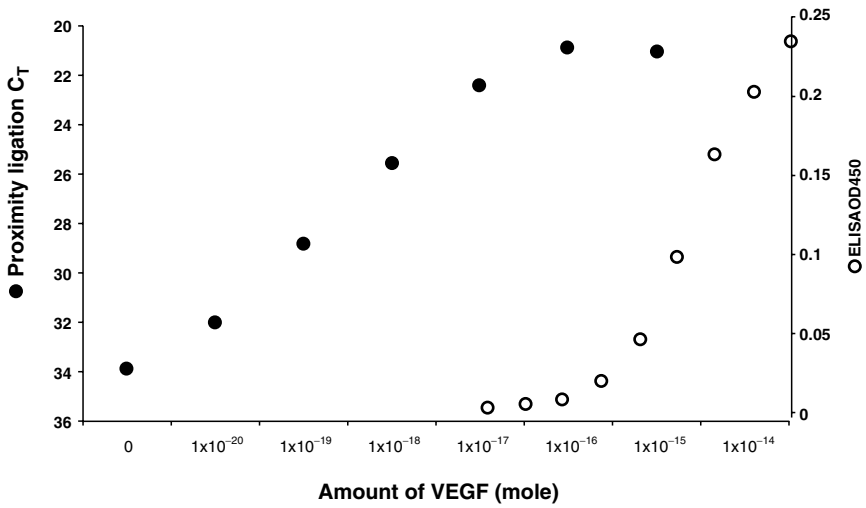

Figure 2. Comparison of detection of VEGF by proximity ligation (filled circles) and by ELISA (open circles). The molar amount of target protein present in 1 µl samples for proximity ligation and 100 µl samples for ELISA is plotted against the cycle threshold values from real-time PCR assays or absorbance at 450 nm for the ELISA

PROXIMITY-LIGATION *IN SITU* ASSAY (P-LISA)

The proximity ligation mechanism can also be used to achieve dual-recognition *in situ* immuno-staining, by modifying the method to provide localized detection signals. In order to obtain highly specific detection *in situ*, the creation of a circular amplifiable DNA template was made dependent on the proximal binding of two proximity probes, in analogy to padlock probe-based detection of single target DNA sequences *in situ* (14). In both cases—using padlock probes for DNA detection and proximity ligation to detect proteins—circular DNA strands form upon highly-specific target detection, and next give rise to single-stranded amplification products composed of hundreds of complements of the circular DNA strands, anchored at the site of probe binding. The RCA products bundle up in random coils less than a micrometer in diameter to which fluorophore-labeled oligonucleotide probes are hybridized. Even single molecules can thus be detected and enumerated easily in a standard fluorescence microscope either by the investigator or using dedicated software, increasing throughput and objectivity (15). Compared with previous methods the requirement for dual recognition significantly increases the specificity of detection.

CURRENT METHODS TO DETECT PROTEIN INTERACTIONS

Measurement of expression levels of a protein often is not sufficient to determine its activity state. Interaction with partners in the formation of protein complexes, and post-translational modifications such as phosphorylation, are

often crucial for the functionality of a protein. Post-translational modifications can be studied with specific antibodies binding the modified residues, although as usual crossreactivity remains a problem in single-recognition strategies. An even more difficult task is studies of interactions between proteins, as microscopic co-localization of signals offers too poor resolution to determine if two or more proteins are interacting, due to the limiting resolution and sensitivity of light microscopy. Recent improvements in confocal microscopy such as 4Pi and STED have enhanced the resolution down to 28 nm (see review by Hell (16)), but detection reactions still face problems of crossreactivity and poor detectability of single fluorophores. If information about sub-cellular localization or inter-cellular variation is not required, then gel electrophoresis-based methods such as co-immunoprecipitation are applicable for studies of protein interaction.

In recent years several methods have been developed for detecting protein interactions based upon split-enzymes, where one part of an enzyme is fused with one protein and the other part of the enzyme with a possible interaction partner (17-21). In yeast two-hybrid assays the DNA binding domain of a transcription factor is fused to one protein and the transcription activating domain fused to another protein, restoring the function of the transcription factor only if the two fused proteins interact (17, 18). By using yeast two-hybrid or split ubiquitin (21) whole protein interaction networks can be determined.

Techniques utilizing either split fluorescent/bioluminescent proteins (22, 23) or resonance energy transfer (24-27) have been developed during the last few years for visualization of protein interactions in living cells. In bimolecular fluorescence complementation (BiFC) analysis, the gene encoding yellow fluorescent protein (YFP) is split in two parts and fused with genes for proteins whose interactions are to be monitored (22). Upon interaction between the two fusion proteins, the two halves of YFP are brought together, resulting in fluorescence. An analogous method utilizing light emission by luciferase was recently published (23). Both fluorescence resonance energy transfer (FRET) (24-26) and bioluminescence resonance energy transfer (BRET) (27) are based on donor molecules exciting acceptor molecules that then emit light. Only when the donor and acceptor are in close proximity, within a few nm and in a favorable orientation, will resonance energy transfer occur.

Although methods such as FRET, BRET and BiFC are very efficient and widely used for interaction studies in living cells, the non-physiological levels of expression of the transgenes, along with the risk that properties of the fusion-proteins may differ from those of the native proteins, may seriously influence the results. However, until now, no methods have been available for the detection of endogenous protein interactions *in situ*.

DETECTION OF PROTEIN INTERACTIONS AND MODIFICATIONS BY PROXIMITY LIGATION

The properties of proximity ligation make it ideal also to detect and measure protein interactions and modifications. By using two or more antibodies directed against interacting partner proteins, the interacting molecules can be detected in a homogenous assay allowing detection of low abundant molecules or

rare interactions (Gustafsdottir et al., in progress). For *in situ* detection, the P-LISA method results in highly-specific and strongly-amplified signals, allowing detection of individual endogenous protein interactions (15) or detection of post-translationally modified proteins, such as phosphorylated receptors with little or no background (Jarvius et al., unpublished). An additional benefit is that the technique enables multiplexed detection, as different proximity probes can give rise to distinct RCA products detectable using specific fluorescence labeled oligonucleotide probes.

SUMMARY

The purpose of this review has been a discussion of different methods for detection of proteins and protein-protein interactions from the point of view of specificity of analysis, focusing on proximity ligation. As all methods have their pros and cons, the choice of method must depend on the question that needs to be addressed, taking in account the time required for the analysis, cost of reagents and availability of instruments, etc. Table 1 summarizes the different methods discussed in this review.

The methods can be divided into three groups according to whether the proteins are analyzed using antibodies (e.g., ELISA, Western blot, immunohistochemistry, flow cytometry, co-immunoprecipitation and proximity ligation), by constructing ectopically expressed fluorescent proteins (FRET/BRET and BiFC), or using mass spectrometry. Future development of methods will depend on further biotechnological progress, but also upon availability of antibodies against all proteins expressed in humans (28) (http://www.proteinatlas.org). The recent development of mass spectrometry and imaging mass spectrometry (reviewed by Chaurand et al. (29) provide us with a tool to investigate protein expression within tissues in a hypothesis-free manner, allowing detection also of proteins to which there are no antibodies. However, improvements in resolution, currently in the 50 μm range, and increased sensitivity will be necessary to allow detection of low abundant proteins.

The completion of the human genome project and the emergence of new molecular tools to study biomolecules and their interactions will fundamentally impact our understanding of life and pathology. With the human genome

Table 1. Utility of different techniques for protein detection.

	Protein identification	**Localized detection**	**Endogenous proteins**	**Protein interactions**	**Analyzing living cells**
ELISA/protein array	+	−	+	+	−
Western blot	+	−	+	−	−
Mass spectrometry	+	+/−	+	−	−
Immunohistochemistry	+/−	+	+	−	−
co-immunoprecipitation	+	−	+	+	−
FRET/BRET/BiFC	−	+	−	+	+
Proximity ligation	+	+	+	+	−

mapped at maximal resolution, the task for the future is now to understand the exceedingly complex interplay between all gene products. Returning to the metaphor of the birds, a cell corresponds to a very richly populated lake indeed, inhabited by many different species busily interacting in different combinations. We can now anticipate having the binoculars to spot a swan in the midst of flocks of all other birds, and to be in a position to observe its day-to-day interactions with its partners and any passers-by.

ACKNOWLEDGMENTS

The work in our laboratory is supported by grants from the Wallenberg Foundation, by the EU Integrated Project "MolTools" and the Research Councils of Sweden for natural science and medicine.

REFERENCES

1 Wide, L., Bennich, H. and Johansson, S.G. (1967) Lancet 2(7526), 1105-1107.
2 Engvall, E. and Perlman, P. (1971) Immunochemistry 8(9), 871-874.
3 Coons. A.H., Creech, H.J., Jones, R.N. and Berliner, E. (1942) J. Immunol. 45, 159-170.
4 Nakane, P.K. and Pierce, G.B., Jr. (1967) J. Cell Biol. 33(2), 307-318.
5 Sano, T.S.C. and Cantor, C.R. (1992) Science 258(5079), 120-122.
6 Schweitzer, B., Wiltshire, S., Lambert, J., O'Malley, S., Kukanskis, K., Zhu, Z., Kingsmore, S.F., Lizardi, P.M. and Ward, D.C. (2000) Proc. Nat. Acad. Sci. U.S.A. 97(18), 10113-10119.
7 Wiltshire, S., O'Malley, S., Lambert, J., Kukanskis, K., Edgar, D., Kingsmore, S.F. and Schweitzer, B. (2000) Clin. Chem. 46(12), 1990-1993.
8 Gusev, Y., Sparkowski, J., Raghunathan, A., Ferguson, H., Jr., Montano, J., Bogdan, N., Schweitzer, B., Wiltshire, S., Kingsmore, S.F., Maltzman, W. and Wheeler, V. (2001) Amer. J. Pathol. 159(1), 63-69.
9 Espina, V., Woodhouse, E.C., Wulfkuhle, J., Asmussen, H.D., Petricoin, E.F., 3rd and Liotta, L.A. (2004) J. Immunol. Methods 290(1-2), 121-133.
10 Brody, E.N., Willis, M.C., Smith, J.D., Jayasena, S., Zichi, D. and Gold, L. (1999) Mol. Diagn. 4(4), 381-388.
11 Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gustafsdottir, S.M., Östman, A. and Landegren, U. (2002) Nat. Biotechnol. 20(5), 473-477.
12 Gullberg, M., Gustafsdottir, S.M., Schallmeiner, E., Jarvius, J., Bjarnegard, M., Betsholtz, C., Landegren, U. and Fredriksson, S. (2004) Proc. Nat. Acad. Sci. U.S.A. 101(22), 8420-8424.
13 Gustafsdottir, S.M., Nordengrahn, A., Fredriksson, S., Wallgren, P., Rivera, E., Schallmeiner, E., Merza, M. and Landegren, U. (2006) Clin. Chem. 52(6):1152-1160.
14 Larsson, C., Koch, J.E., Nygren, A., Janssen, G., Raap, A.K., Landegren, U. and Nilsson, M. (2004) Nat. Methods 1(3), 227-232.

15 Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K-J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L-G. and Landegren, U. Nature Methods, *in press.*
16 Hell, S.W. (2003) Nat. Biotechnol. 21(11), 1347-1355.
17 Fields, S. and Song, O. (1989) Nature 340(6230), 245-246.
18 Chien, C.T., Bartel, P.L., Sternglanz, R. and Fields, S. (1991) Proc. Nat. Acad. Sci. U.S.A. 88(21), 9578-9582.
19 Rossi, F., Charlton, C.A. and Blau, H.M. (1997) Proc. Nat. Acad. Sci. U.S.A. 94(16), 8405-8410.
20 Pelletier, J.N., Campbell-Valois, F.X. and Michnick, S.W. (1998) Proc. Nat. Acad. Sci. U.S.A. 95(21), 12141-12146.
21 Johnsson, N. and Varshavsky, A. (1994) Proc. Nat. Acad. Sci. U.S.A. 91(22), 10340-10344.
22 Hu, C.D., Chinenov, Y. and Kerppola, T.K. (2002) Mol. Cell 9(4), 789-798.
23 Paulmurugan, R. and Gambhir, S.S. (2003) Anal. Chem. 75(7), 1584-1589.
24 Mitra, R.D., Silva, C.M. and Youvan, D.C. (1996) Gene 173(1 Spec No), 13-17.
25 Mahajan, N.P., Linder, K., Berry, G., Gordon, G.W., Heim, R. and Herman, B. (1998) Nat. Biotechnol. 16(6), 547-552.
26 Galperin, E., Verkhusha, V.V. and Sorkin, A. (2004) Nat. Methods 1(3), 209-217 Epub 2004 Nov 2018.
27 Xu, Y., Piston, D.W. and Johnson, C.H. (1999) Proc. Nat. Acad. Sci. U.S.A. 96(1), 151-156.
28 Uhlén, M., Björling, E., Agaton, C., Szigyarto, C.A., Amini, B., Andersen, E., Andersson, A.C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergström, K., Brumer, H., Cerjan, D., Ekström, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Björklund, M.G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G. and Kampf, C. (2005) Mol. Cell Proteomics Aug 27; [Epub ahead of print].
29 Chaurand, P., Sanders, M.E., Jensen, R.A. and Caprioli, R.M. (2004) Amer. J. Pathol. 165(4), 1057-1068.