On the need for a standardized set-up for validation studies of fish vitellogenin assays as an endpoint in endocrine disruptor testing and screening – a proposal

Anders Goksøyr[1,2], Janne K. Eidem[1], Sven Inge Kristiansen[1] and Bente M. Nilsen[1]
[1]*Biosense Laboratories AS, Thormøhlensgate 55, N-5008 Bergen, Norway*
[2]*Department of Molecular Biology, University of Bergen, Norway*

Summary

Over the past few years, the use of vitellogenin (Vtg) as a sensitive biomarker for endocrine disrupting effects, specifically for (anti-) estrogenic chemicals, has moved from research laboratories into contract work and international guideline development. In this process, various studies have aimed at comparing both in-house laboratory methods and a few commercially available assay kits. However, the issue of endocrine disruptor testing and screening using fish exposure and vitellogenin (among others) as endpoint, involves several aspects that need to be handled separately to be able to clearly evaluate the results and identify problems. These range from exposure protocols and sampling procedures to Vtg analysis and data handling, including quality assurance. Here we propose that the OECD initiate a process for establishing procedures for preparation of certified reference material and for performing a full validation of available assays for relevant test fish Vtg according to standards as set by the Association of Official Analytical Chemists (AOAC).

Background

Vitellogenin (Vtg) is a large phospholipoglycoprotein produced as the yolk protein precursor in the liver of oviparous vertebrates such as fish. The Vtg protein is secreted from the cell through the secretory pathway, enters circulation and is taken up by the growing oocyte. Plasma Vtg concentrations are normally an indication of the maturational status of the female fish (for reviews, see Mommsen & Walsh, 1988; Arukwe & Goksøyr, 2003). More than a decade ago, several studies demonstrated that even male fish caught in rivers and streams had high levels of plasma Vtg (e.g. Purdom et al., 1994; Jobling et al., 1998), caused by chemicals acting like estrogens present in the environment. Since then, numerous studies have shown the fish Vtg to be a highly responsive biomarker for estrogenic compounds in both in vitro hepatocyte cell cultures, in vivo aquaria studies, and field studies (for reviews, see Kime, 1995; Sumpter & Jobling, 1995; Arukwe & Goksøyr, 1998; 2003). Hence, Vtg in fish has become an accepted measure of xenoestrogenic potency of chemicals, effluents, and discharges.

Within international bodies such as the Organization for Economic Cooperation and Development (OECD), work is ongoing to develop screening and testing programmes for endocrine disrupting effects of new chemicals. In the focus of this development are small fish test species including the fathead minnow (*Pimephales promelas*), zebrafish (*Danio rerio*), and Japanese medaka (*Oryzias latipes*). These fish share several attributes that make them ideal test species for reproductive toxicity testing, including small size at maturity, relatively short generation times, asynchronous spawning, and overall ease of culture.

Against this background, there is a need for specific, sensitive, and robust assays for detecting Vtg in these fish species, assays that are readily available and give reproducible results in different laboratories. The ELISA (enzyme-linked immunosorbent assay) technique is a sensitive laboratory method widely used to detect and quantify antigens or antibodies in a variety of biological samples. It can be quantitative (with a standard curve) or qualitative (semi-quantitative - without a standard curve). The two most widely used principles for quantitative detection of proteins are the competitive ELISA and the sandwich ELISA techniques (Crowther, 2001).

In the process of evaluating endpoints for endocrine disruptor testing and screening over the recent years, various studies have been conducted, involving both in-house laboratory methods for Vtg analysis and a few commercially available assay kits. However, the issue of endocrine disruptor testing and screening using fish exposure, and Vtg as endpoint, encompasses several aspects that need to be handled separately to be able to clearly evaluate the results and identify problems. These range from exposure protocols and sampling procedures to the Vtg analysis itself, data handling, and quality assurance. Some of the studies indicate that varying results can be obtained in different laboratories with different methods, and also that the same method used in different laboratories can give different results. Adding confusion to this, some recent papers (e.g. Mylchreest et al., 2003) mis-interpret data on VTG ELISAs from other papers. An appreciation of the real challenges involved in Vtg quantification is apparently needed.

In addition to the general issues of antibody specificity and sensitivity, there are some specific challenges related to the development of quantitative immunoassays for Vtg. Vtg is relatively easily purified from plasma of estrogenized fish (where it can reach concentrations of 50-150 mg/ml). It is however an inherently unstable protein. The instability of Vtg is due to its role as a precursor for shorter peptide fragments, and it is very sensitive to proteolytic breakdown into these fragments. This instability leads to some problems with immunization, since breakdown products may be more immunogenic than Vtg itself. In addition, it is a challenge for the use of Vtg as a standard in quantitative assays (see discussion in Nilsen et al., 2003).

The dynamic range of Vtg concentrations found in fish plasma creates another challenge. Plasma Vtg can vary maybe 100 million-fold, from a few ng/ml in unexposed male fish, to the 50-150 mg/ml or above found in estrogenized salmonids (e.g. Tyler et al., 2002). To be able to quantify this enormous range in samples analyzed blind, the working range of the assay should preferentially be as wide as possible.

To summarize, a validation of Vtg assays must include solving issues pertaining to:

- the standard
- the assay itself
- sample preparation

In order to have a validated method where you can rely on your results, the sampling procedure must also be established and validated.

Here we propose the following strategy:

1. Development of a procedure for the production of certified reference material (CRM) for fathead minnow (or carp) Vtg, zebrafish Vtg, and medaka Vtg
2. Specification of performance criteria for the Vtg assays

3. Establishment of a validation study protocol (collaborative study protocol) according to (e.g.) AOAC guidelines
4. Establishment of guidelines for sample preparation

Production of Certified Reference Material (CRM)

The CRM should consist of intact Vtg purified to apparent homogeneity from each test species (fathead minnow FHM or carp C, zebrafish ZF, and medaka M), quantified according to accepted methods for protein quantification. The CRM needs to be stable during shipping and storage. The issues here pertain to the choice of source for purification, the purification procedure, the quantification method, and the stabilization procedure. Here we discuss these issues, and propose that the OECD initiates a process that leads to the establishment of detailed standard procedures for CRM production. The need for a negative ("matrix blank") CRM should also be considered.

Source for CRM production

Various sources of Vtg can be envisaged, such as whole body homogenates (WBH), plasma samples, liver, ascites fluid, and egg yolk, all of which are being employed by different laboratories. In each of these sources, the state and stability of the Vtg protein will be different. E.g. the liver cell will contain immature Vtg that has not undergone full post-translational modifications, as well as mature Vtg ready for secretion, whereas the egg will contain the lipovitellin form processed after uptake. WBH will contain a mixture of all of these (unless ovaries and /or liver have been removed), in addition to a high level of proteolytic activity that may act to degrade the protein during preparation. The plasma appear to be the best source of Vtg, since it contains the Vtg in its circulatory form at high concentrations. The challenge of taking blood samples from small fish needs to be addressed in this regard (see also Establishment of guidelines for sample preparation below).

A CRM should preferentially contain the protein to be detected in a form similar to the one found in samples to be tested. Both WBH, liver, and plasma samples have been proposed for sampling protocols, and the matrix used for sampling should also be chosen as a source for the CRM. The matrix for CRM production should be obtained from estrogenized fish exposed to a standard level of 17β-estradiol (or another given reference estrogen) in the water for a given period of time. A suitable protease inhibitor (e.g. aprotinin or protease inhibitor cocktail) should be added to the matrix during sampling.

Purification of CRM

Due to the instability of Vtg, the purification procedure should be as rapid as possible while maintaining the integrity of the protein and yielding a pure product. Various methods involve ion exchange chromatography and gel permeation clean-up (e.g. Brion et al., 2002). Although efficient for some species, others seem vulnerable to degradation during or after purification. Selective precipitation of Vtg from plasma using $MgCl_2$ is a very rapid method that has been successfully used with some species (e.g. trout and carp: Norberg & Haux, 1985, and Nilsen et al., 2003), but appears to be less useful for some other species and especially when Vtg levels in plasma is below 20 mg/ml (unpublished results). When specific antibodies are available a one-step, selective, rapid purification procedure can be obtained by using immunoaffinity-based procedures (e.g. chromatography or magnetic beads). In this strategy, a specific

antibody is immobilized on chromatographic beads (e.g. Sepharose), or magnetic beads (e.g. Dynabeads). In chromatography, the beads are packed in a column and used to capture the Vtg protein from the sample passed over the column. In magnetic bead separation, the capture is performed in solution. The captured Vtg is then released from the antibody by passing an acidic buffer over the beads, and the solution is subsequently neutralized. Within a short period of time (10-30 min.), mg quantitites of pure Vtg can be obtained. The process should be performed with cold buffers and in a cold environment to prevent degradation of the protein.

Different purification protocols may lead to different compositions of the Vtg holo-protein (i.e. different parts of the phospholipoglyco-modifications may be retained), which also may lead to different affinities to the antibodies in the ELISA assays. The production of CRM needs to take these issues into account, to ensure a reliable and stable source of homogeneous material.

The purity and homogeneity of the CRM need to be established using e.g. both native and SDS-PAGE gel electrophoresis. For each species the relative molecular mass (m_r) of the major Vtg band(s), and the relative purity of the band(s) in the sample (e.g. obtained by densitometric analysis of the gel) need to be established and specified to meet certain criteria.

Quantification of CRM

Various methods have been used to quantify purified Vtg. Methods such as Lowry et al. (1951) or Bradford (1976), rely on the use of a protein standard, normally bovine serum albumin (BSA), and assume a similar staining response of the protein in question to this standard protein. In many cases, this will not be true; specifically for a protein like Vtg, containing various non-peptide groups (sugars, phosphates, lipids), the staining reaction may be influenced by these factors. Ideally, therefore, the standard used in the protein determination should be the protein itself, and we are then placed in a chicken-and-egg situation (see discussions in Arukwe & Goksøyr, 2003; Nilsen et al., 2003). Also, our own results demonstrate that species differences in Vtg behaviour is a great concern in the quantification of Vtg.

A very simple method often used is the measurement of absorbance at 280□m. In order to convert this value into a protein concentration, this method relies on the use of an extinction coefficient that will be different for each protein, due to the intramolecular environment affecting the exposure of UV absorbing aromatic amino acid residues. A theoretical extinction coefficient can be calculated when the full amino acid sequence of a protein is known (e.g. using the ProtParam tool at Expasy: http://us.expasy.org/tools/protparam.html). However, this will only be correct if a specific denaturing buffer is used, and it is not known how posttranslational modifications will affect this value. A280 absorbance is also sensitive to other components of the system being analyzed. One way to obtain a specific extinction coefficient is to prepare a sufficient amount of material to be able to dry and weigh it in a precise manner (e.g. Norberg & Haux, 1988).

The best solution to the quantification problem may be to use a more precise method of protein quantification such as *quantitative amino acid analysis*. This analysis can be performed by independent analytical laboratories using standard instrumentation. A limitation of this method is that we only quantify the peptide portion of the Vtg molecule. The lipid and phosphate parts have been reported for some species to represent 15–20% and 0.6–0.8%, respectively (e.g. Silversand & Haux, 1995), whereas the carbohydrate portion is not well studied.

To demonstrate the homogeneity of the Vtg CRM, all portions of the phospholipoglycoprotein should ideally be characterized, including lipid, phosphate, and sugar moieties, using appropriate methods.

Stabilization of CRM

Purified Vtg is normally sensitive to thawing and freezing, and needs to be stored at –80°C and shipped on dry ice. For a CRM to be useful to a wider audience, it needs to be aliquoted into smaller vials containing the normal amounts used in a single experiment. It also needs to be stabilized for storage and shipping, to avoid possible degradation that may occur if shipments thaw in transport. In the Biosense laboratory, we have developed a lyophilization procedure that gives excellent stability to Vtg preparation, with some species variation. This procedure (proprietary information) can provide 3-10 month stability to Vtg CRM at 4°C, depending on the species.

Specification of performance criteria for Vtg assays

Important performance criteria that need to be considered in Vtg assays are:

- Selectivity
- Calibration
- Accuracy
- Repeatability
- Reproducibility
- Limit of Detection (LOD) and Limit of Quantification (LOQ)

Definitions

Selectivity: "The ability to measure accurately the analyte (Vtg) in the presence of components that may be expected to be present in the matrix (plasma). Selectivity is demonstrated by using "matrix blanks".

"Matrix blank" is a representative sample that does not contain detectable levels of analyte (Vtg).

Calibration, often thought of as linearity, is the empirical determination of the relationship between the parameter measured (e.g. ELISA absorbance) and the analyte concentration. The range of concentrations of analyte where such relationship is established is often referred to as "calibration range" or the "standard curve working range"

Accuracy is usually expressed as recovery and is here defined as "fraction or percent of the analyte (Vtg) that is recovered when a sample is tested using the method of analysis under validation. Determination of accuracy requires standard reference material. Accuracy is an expression of bias or systematic error whereas

Repeatability (often referred to as precision) is an expression of random error. Repeatability is the degree of agreement of results when conditions are maintained as constant as possible. Repeatability can be divided into within-day precision and between-day precision and are usually expressed as relative standard deviation, RSD_r.

Reproducibility or precision under reproducible conditions, when test results are obtained by an analyst in another laboratory using the same analytical method to analyze the same test sample at the same or different time.

LOD - the lowest content that can be measured with reasonable statistical certainty. Often expressed as reagent blank + 3x standard deviation of reagent blank.

LOQ - The content equal to or greater than the lowest concentration point in the calibration curve (AOAC). Also defined as reagent blank + 10x standard deviation of reagent blank (IUPAC "orange" Book).

General performance criteria

Based on experience so far, there is a need to realize that Vtg assays have a higher variability than many other biological assays and immunoassays. The criteria listed below should be regarded as the minimal acceptable performance as defined from a user standpoint on the purpose of performing Vtg analysis. Specific performance criteria need to be established for each specific kit or assay to be used in the validation study based on in-house (within laboratory) performance.

Selectivity	Matrix blank < LOD (with the necessary dilution factor to avoid matrix effects)
Calibration	Standard curve working range > 10-fold, preferably 50-100 fold to be practical with the dynamic range found in Vtg levels
Accuracy	Ideally 50-200%
Repeatability	<20%
Reproducibility	<50%
Limit of Detection (LOD)	<10 ng/ml
Limit of Quantification (LOQ)	<10 ng/ml
Sample LOQ (=LOQ x necessary matrix dilution)	200 – 500 ng/ml

The characterization of the “matrix effect” is an important challenge in this regard. It may be difficult to ensure that a “matrix blank” sample is really devoid of any Vtg. One method to obtain a “matrix blank” could be by removal of Vtg using a specific antibody. Criteria for establishing the matrix effect need to be developed. A negative (“matrix blank”) CRM should also be considered.

Establishment of a validation study protocol (collaborative study protocol) according to AOAC guidelines

Validation is the process of demonstrating or confirming the performance characteristics of a method of analysis. The primary purpose of validating a method of analysis is to show that the method is fit for its intended purpose. The study should conform to accepted international requirements as much as possible. The overall objective of this study in particular is to establish documentation on the performance of (commercially) available ELISA kits or assays for quantification of fish vitellogenin (Vtg).

A validation study is a study of method, not of the laboratory. The chosen method must be followed as closely as practicable and any deviations must be reported to the study director.

Preliminary work (within one laboratory)

Before commencing to a full collaborative study it is necessary to conduct a single laboratory validation (SLV) of the method according to AOAC guidelines to

determine the precision characteristics of the method. Accuracy, repeatability, selectivity and limit of detection found in the SLV are the basis for the design of the collaborative study. One should also conduct as much experimentation within a single laboratory as possible with respect to optimization, ruggedness, and interferences. There is no purpose in conducting a collaborative study with a method that has not been optimized.

Inter-laboratory study design

The aim of a collaborative study is to establish the precision (RSD_r and RSD_R) for a method under realistic inter-laboratory conditions. The study design proposed below follow the standard design recommended in the AOAC guidelines for a full collaborative study to yield statistically valid data for a between laboratory validation (AOAC, 1995). The minimum requirement is for 8 laboratories to report data for 5 materials with 2 replicates. These data will be sufficient to generate values for within-laboratory repeatability RSD_r- and between-laboratory reproducibility RSD_R. If the samples are prepared from Certified Reference Material the study will also yield data on method accuracy. The comparison of calibration curves produced by the participating laboratories will also yield valuable information on this aspect of the method and could assist in setting some acceptance parameters for the kit in routine use.

Materials should be a representative matrix (plasma, whole body homogenate) with customary and extreme values for the analyte. Vtg materials should typically include low, medium and high concentrations. Including material close to limit of quantification and materials with Vtg concentrations at mg/ml levels (depending on the fish species of interest).

Data reporting and analysis

A standard system for data reporting and statistical analysis is essential for a reliable validation study. Experience from recent comparison studies (Battelle-US EPA, USA; INERIS, France) should be considered when final reports from these studies become available.

Data reporting:

Collaborators are obliged to follow the method exactly and any deviation must be recorded. This of course also includes data reporting. The collaborators are to conduct exactly the number of determinations stated in the instructions. Any other number complicates statistical analysis. Report individual values, including blanks. In making statistical calculations from the data, the full power of the calculator or computer is to be used with no rounding until the final reported mean or standard deviations are achieved. The initiating laboratory should indicate the number of significant figures to be reported, usually based on the output of the measuring instrument. Supply raw data, graphs, recorder tracings, photographs, or other documentation as requested in the instructions.

Statistical analysis:

Only valid data should be included in the statistical analysis. The study must maintain valid data from a *minimum of 8 laboratories*. Invalid data results may occur when:

- The method is not followed
- The calibration curve is nonlinear when a linear curve is expected

- Specifications in the protocol are not met
- Unexpected reactions occur
- Other atypical phenomena materialize

Rejection of more than 2/9 of the data from each material in a study, without an explanation, is considered excessive. The probability that the apparent aberrant value(s) is part of the main group of values considered as a normal population is determined by applying an appropriate test for outliers (e.g. Cochran test or Grubbs test(s)).

Establishment of guidelines for sample preparation

Guidelines for sample preparation should be established, taking into account the need to avoid degradation of the Vtg in the sample by appropriate use of protease inhibitors, by keeping samples on ice, and by reducing preparation time. The sampling technique should also be specified, since e.g. blood collection may be contaminated by other body fluids leading to differences in composition between samples. Finally, storage conditions need to be specified, and repeated freeze-thawing should be avoided for samples to be quantitatively analyzed.

Conclusions and recommendation

- Certified Reference Material (CRM) for Vtg from the relevant species should be produced and quantified by appropriate methods for use in validation studies.
- Specific performance criteria for Vtg assays should be established.
- A validation study protocol should be established, incl. guidelines for data handling and statistical analysis.
- Guidelines for sample preparation should be established.

When all these are in place, full validation studies of available Vtg assays should be performed to allow the use in the forthcoming endocrine disruptor testing and screening programmes.

Acknowledgements

We want to thank Charles R. Tyler (University of Exeter, UK) and Jean-Marc Porcher (INERIS, France) for having read and commented on earlier versions of this manuscript:. The responsibility for the opinions expressed in this paper lies solely with the authors.

References

AOAC INTERNATIONAL (1995) *J AOAC Int* 78: 143A-160A.

Arukwe A, Goksøyr A (1998). Xenobiotics, xenoestrogens and reproduction disturbances in fish. *Sarsia* 83:225-241.

Arukwe A, Goksøyr A (2003). Eggshell and egg yolk proteins in fish: hepatic proteins for the next generation. Oogenetic, population, and evolutionary implications of endocrine disruption. *Comp Hepatol* 2: 4.

Bradford MM (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye-binding. *Anal Biochem* 72: 248-254.

Brion F, Nilsen BM, Eidem JK, Goksøyr A, Porcher JM (2002). Development and validation of an enzyme-linked immunosorbent assay to measure vitellogenin in the zebrafish (*Danio rerio*). *Environ Toxicol Chem* 28: 1699-1708.

Crowther JR (2001). The ELISA Guidebook. In: *Methods Mol Biol* 149. Humana Press, Totowa, NJ

Jobling S, Nolan M, Tyler CR, Brighty G, Sumpter JP (1998). Widespread sexual disruption in wild fish. *Environ Sci Technol* 32: 2498–2506.

Kime DE (1995). The effects of pollution on reproduction in fish. *Rev Fish Biol Fisheries* 5:52-96

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951). Protein measurement with the folin phenol reagent. *J Biol Chem* 193:265–275.

Mommsen TP, Walsh PJ (1988). Vitellogenesis and oocyte assembly. In: Hoar WS, Randall VJ (eds) Fish Physiology XIA. Academic Press, New York, pp 347-406.

Mylchreest E, Snajdr S, Korte JJ, Ankley GT (2003). Comparison of ELISAs for detecting vitellogenin in the fathead minnow (*Pimephales promelas*). *Comp Biochem Physiol* 134C: 251-257.

Nilsen BM, Berg K, Eidem JK, Kristiansen SI, Brion F, Porcher J-M, Goksøyr A (2003). Development of quantitative vitellogenin-ELISAs for fish test species used in endocrine disruptor screening. *Anal Bioanal Chem* (in press, online DOI: 10.1007/s00216-003-2241-2)

Norberg B, Haux C (1985). Induction, isolation and a characterization of the lipid content of plasma vitellogenin from two *Salmo* species: rainbow trout (*Salmo gairdneri*) and sea trout (*Salmo trutta*). *Comp Biochem Physiol* 81B:869–876.

Norberg B, Haux C (1988). An homologous radioimmunoassay for brown trout (*Salmo trutta*) vitellogenin. Fish Physiol Biochem 5: 59-68.

Purdom CE, Hardiman PA, Bye VJ, Eno NC, Tyler CR, Sumpter JP (1994). Estrogenic effects of effluents from sewage treatment works. *Chem Ecol* 8: 275-285.

Silversand C, Haux C (1995). Fatty acid composition of vitellogenin from four teleost species. *J Comp Physiol* 164B: 593-599.

Sumpter JP, Jobling S (1995). Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. *Environ Health Perspect* 103 (Suppl 7): 173-178.

Tyler CR, van Aerle R, Nilsen MV, Blackwell R, Maddix S, Nilsen BM, Berg K, Hutchinson TH, Goksøyr A (2002). Monoclonal antibody enzyme-linked immunosorbent assay to quantify vitellogenin for studies on environmental estrogens in the rainbow trout (*Oncorhynchus mykiss*). *Environ Toxicol Chem* 21:47-54.