How to compare ELISA kit performance between different vendors

Introduction

Selecting the right ELISA kit at the start of a long-term research project can help save time and other resources over the course of the study. This choice is not always easy because different vendors' kits often have subtle differences in components and protocols that can significantly impact results. Antibodies, protein standards, buffer composition, and detection methods all contribute to the performance of an assay kit. A direct side-by-side comparison at the lab bench is usually necessary to find the optimal kit for a particular target and sample type. Here we share some tips on how to best compare ELISA kits from different vendors.

What you should do

Each ELISA kit should be run as indicated in the respective product information sheet. Each plate should have the same layout of standards, samples, and controls (Figure 1). Natural samples like serum, plasma, or cell culture supernatant should be split under the same conditions onto each plate (e.g., same freeze–thaw conditions and sample handling). Samples with normal and elevated protein levels should be examined, such as healthy and diseased serum samples. In addition, samples with elevated protein levels should be diluted a few times over the range of the assay. Linearity-of-dilution and spike-and-recovery experiments are recommended to assess the reliability of results over a range of protein concentrations.

What you should avoid

Protein standards should not be shared between kits. Recombinant proteins are often used as standards in ELISA kits. If the standard is not a full-length protein, the capture or detection antibody may not bind to the standard. Similarly, do not use reagents from one kit (e.g., wash solution) in another kit or skip any steps in the protocol. The same protocol should not be used with different kits. Each assay is designed and tested by the vendor using the kit-specific protocol for best performance.

Std 1:1	Std 1:1	HS 1 1:1	HS 1 1:1	DS 1 1:2	DS 1 1:2	DS 1 1:8	DS 1 1:8	SR 1 High	SR 1 High	SR 1 Low	SR 1 Low
Std 1:2	Std 1:2	HS 2 1:1	HS 2 1:1	DS 2 1:2	DS 2 1:2	DS 2 1:8	DS 2 1:8	SR 2 High	SR 2 High	SR 2 Low	SR 2 Low
Std 1:4	Std 1:4	HS 3 1:1	HS 3 1:1	DS 3 1:2	DS 3 1:2	DS 3 1:8	DS 3 1:8	SR 3 High	SR 3 High	SR 3 Low	SR 3 Low
Std 1:8	Std 1:8	HS 4 1:1	HS 4 1:1	DS 4 1:2	DS 4 1:2	DS 4 1:8	DS 4 1:8	SR 4 High	SR 4 High	SR 4 Low	SR 4 Low
Std 1:16	Std 1:16	DS 1 1:1	DS 1 1:1	DS 1 1:4	DS 1 1:4	LD 1 Neat	LD 1 Neat	SR 1 Medium	SR 1 Medium	Unspiked serum 1	Unspiked serum 1
Std 1:32	Std 1:32	DS 2 1:1	DS 2 1:1	DS 2 1:4	DS 2 1:4	LD 1 1:2	LD 1 1:2	SR 2 Medium	SR 2 Medium	Unspiked serum 2	Unspiked serum 2
Std 1:64	Std 1:64	DS 3 1:1	DS 3 1:1	DS 3 1:4	DS 3 1:4	LD 1 1:4	LD 1 1:4	SR 3 Medium	SR 3 Medium	Unspiked serum 3	Unspiked serum 3
Dilution buffer	Dilution buffer	DS 4 1:1	DS 4 1:1	DS 4 1:4	DS 4 1:4	LD 1 1:8	LD 1 1:8	SR 4 Medium	SR 4 Medium	Unspiked serum 4	Unspiked serum 4

Figure 1. Example plate layout for each vendor. Protein standard and different samples (1–4) are run in duplicate. Std: protein standard at different dilutions, which may vary by manufacturer. HS: healthy serum samples with low levels of target protein. DS: diseased serum samples with elevated levels of target protein. LD: linearity-of-dilution testing to evaluate if the dose response of a sample is linear in the desired assay range. SR: spike-and-recovery testing at high, medium, and low levels of target protein to determine if detection is affected by the sample matrix; wells with unspiked serum are included for background correction.

How to analyze results

Use curve-fitting software to generate the standard curve for each kit. Ensure that raw data from both kits were run using the recommended 4-parameter logistic (4PL) curve-fitting algorithm. The standard curve from each kit should look similar to the result provided in the product documentation (Figure 2). This assessment includes examining the goodness of fit, or how well the assay fits the algorithm model. If any standard is outside of 70–130% of the expected value, consider omitting that data point.

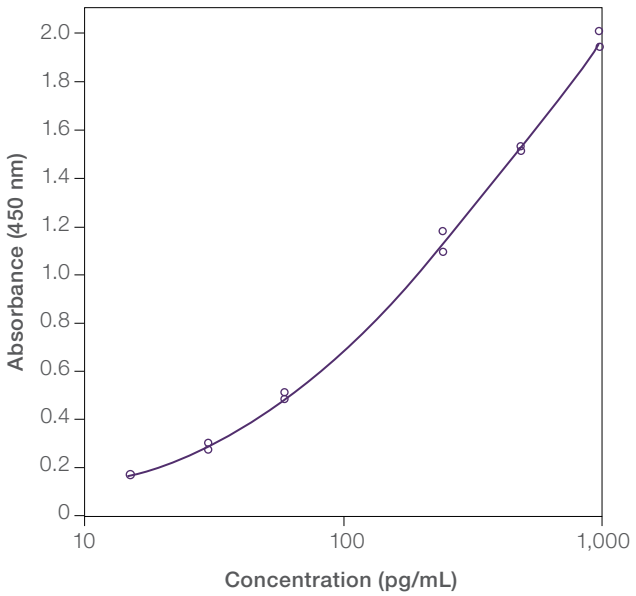

Figure 2. Standard curve profile. An example standard curve is shown from the user guide of the Invitrogen™ IL-8 Human ELISA Kit (Cat. No. BMS204-3).

Duplicate runs should have a coefficient of variation (CV) of <10%. Samples with low and high protein levels should fall within the dynamic range of the assay, minimizing the need for sample manipulations. Linearity-of-dilution experiments should show measured concentrations at 80–120% of the expected value when diluting samples in the linear portion of the standard curve (Figure 3). Spike-and-recovery experiments should result in mean recoveries close to 100% of the expected value (Figure 4).

Dilution	Measured (pg/mL)	Expected (pg/mL)	Recovery of expected (%)
Neat	708	–	–
1:2	328	354	93
1:4	178	177	101
1:8	91	89	102

Figure 3. Linearity of dilution. In this example, human serum was spiked with IFN-γ and serially diluted in standard diluent buffer. Measured values were in the expected range. Performance data were obtained using the Invitrogen™ Human IFN-γ ELISA Kit (Cat. No. KHC4021).

Sample matrix	Spike high (%)		Spike medium (%)		Spike low (%)	
	Mean	Range	Mean	Range	Mean	Range
Serum	94	80–111	102	95–114	105	91–118

Figure 4. Spike and recovery. The mean recovery was determined after spiking 3 levels of human IL-2 into serum samples. Performance data were obtained using the Invitrogen™ IL-2 ELISA Kit (Cat. No. BMS221-2).

What to expect

One kit may show an advantage over the other kit for a particular sample type and target. For example, one pair of capture and detection antibodies may offer greater sensitivity and specificity than another pair when using a complex sample matrix. Underestimates or overestimates of protein concentrations should become apparent when examining spike-and-recovery data. If the spiked-in protein has values outside of a 70–130% expected range, the assay may not be able to accurately quantitate that protein.

It is common to see only minor differences in quantitation results, which can arise from differences in the calibration of protein standards between kits. In this scenario, you should see similar trends in protein quantitation in which measured concentrations are high, medium, or low for both kits. If the kits perform similarly, other factors such as ease of use and workflow time may impact your final choice. For example, Invitrogen™ Instant ELISA™ kits include plates precoated with capture antibody, detection antibody, streptavidin–HRP, and sample diluent. All of the preparation work has been done for you, which reduces workflow time and helps minimize variation.

How to interpret performance information

Vendors typically provide a set of performance data for ELISA kits. This section is a summary of the most important performance parameters to assess before making a purchasing decision.

- **Analytical sensitivity:** It is an industry standard for ELISA manufacturers to report the analytical sensitivity, more commonly known as the limit of detection (LOD). Analytical sensitivity is the lowest analyte concentration that can be detected above the background. The background concentration is calculated by adding 2 standard deviations to the mean optical density of the dilution medium. Analytical sensitivity should not be confused with functional sensitivity. Functional sensitivity, or the limit of lowest quantitation (LLOQ), is the lowest concentration at which the analyte can be reliably detected. While functional sensitivity might be more useful to researchers, analytical sensitivity is what is reported by most vendors.
- **Intra-assay precision:** Intra-assay precision refers to sample replicates on the same plate. Replicate wells should give comparable results for each sample. The CV across the replicates should be less than 10%.
- **Inter-assay precision:** Inter-assay precision refers to sample replicates assayed on different days and different plates. This test tends to be more variable than the intra-assay precision, so CV values may be higher. Inter-assay precision helps ensure that results obtained will be consistent over time and between kits.
- **Spike and recovery:** This test shows if the true concentration of a sample is obtained without interference. Some target proteins may be in complex matrices such as plasma or serum, which can cause issues with antibody capture and binding. To determine if all of a target protein is captured during the assay, a known concentration of the protein is spiked into the sample matrix and measured using the assay. Ideally, recovery should be close to 100%.
- **Parallelism:** This test refers to similar accuracy of the assay with respect to the recombinant protein standard of the kit and the protein in the natural sample. The results for the natural sample protein and the recombinant standard should be parallel.
- **Linearity of dilution:** A sample with low protein levels is spiked with a known amount of analyte and evaluated at different dilutions. Recovery should be close to 100% for spike concentrations in the linear portion of the standard curve.

invitrogen

Conclusion

Before investing in a long-term or critical study, it is important to understand the performance of an ELISA kit with a particular research sample. Due to the wide variety of ELISA kits available from different vendors, the choice of kit is not always straightforward. The experimental approach and performance parameters outlined here are designed to help researchers choose the kit that best fits the needs of their project.

Find out more at **thermofisher.com/elisa**

ThermoFisher SCIENTIFIC