Quality control issues for the quantitative analysis of protein arrays: a case study

Protein arrays have emerged as a powerful tool in the post-genomic era. While DNA microarrays are currently more widespread and technologically simple, they are limited in that they only examine transcript abundance. Gene function is manifested by the activity of its translated protein, and therefore probing protein function is a more direct route to elucidating gene function. Protein arrays provide a means to simultaneously analyze the biochemical activity and interactions of entire proteomes. As with DNA microarrays, methodical and quantitative interpretation of array data has lagged behind in the field and thus many of the same issues that arise in interpreting DNA microarray data also apply to the analysis of protein arrays. In this paper I shall take one example of a protein chip and discuss quality control issues that limit its quantitative analysis.

Nanowell Protein Chip

Nanowell protein chips consist of an array of microwells in a disposable silicone elastomer into which small volumes of different analytes are placed. Proteins are covalently attached to the wells using a crosslinker molecule. The biochemical activity of these proteins or their reactivity as a substrate can thus be analyzed in a high-throughput manner. For example, Zhu *et al.* used nanowell protein chips to analyze 119 yeast kinases for their activity on 17 different substrates. Substrate was attached to each of the wells; kinases were added individually in the presence of γ^{33}P-ATP and allowed to react. The chip was rinsed and imaged using a phosphorimager. Novel tyrosine kinase activity was discovered for a large number of yeast kinases using this assay.

Nanowells present unique considerations in quantitative analysis. Saturation of signal can occur and limit the accuracy of intensity values. Background from the elastomer and from nonspecific adsorption of the radiolabeled molecule must be subtracted. Strong-signal wells may “bleed” onto neighbors and can occlude weaker spots. Rings of signal are commonly seen, as substrate will attach to the sides as well as the bottoms of the wells. If wells are punctured during manual loading of the radioactive substance, artifacts identifiable as small, dark spots can occur. Quantitation of spot

intensity raises other questions. Should spot intensity be calculated by summing over all pixel values within a given spot, or by averaging over the spot? How should results from different experiments (chips) be compared and can they even be compared? How should nanowell data be normalized?

Data acquisition procedures and analysis should fulfill several conditions. The approach should be sensitive enough to detect weak signals, yet a threshold needs to be established to eliminate false positives. Ideally, a wide range of linearity should be accessible in order to simultaneously monitor changes of both strong and weak signals. Lastly, comparison between experiments performed at different times on different chips with different protein samples should be possible. Pelizzari *et al*. have described a method based on quantitative electron film dosimetry utilized in radiotherapy physics for the quantitative analysis of DNA array autoradiographs. I discuss here how this method can be applied to nanowell experiments.

An OD (optical density) standard was included in scans of autoradiographs to calibrate pixel value with film OD. While this was done for filter-based DNA microarrays, an analogous calibration curve can be utilized for nanowells. Similarly, a standardization curve correlating known amounts of radiation to OD/pixel values was generated. For nanowells, this could be achieved by directly attaching varying concentrations of radiolabeled proteins to a column of wells in each chip. (Alternatively, for the kinase assay, a series of controlled enzymatic reactions could be performed in which the substrate attached to the well is the limiting reagent and an excess of enzyme (kinase) and γ^{33}P-ATP are added.) This would also be a means of assessing the efficiency of protein attachment to the wells. Calibration and standardization permit the correlation of signal to amount of radioactivity and, indirectly, biochemical activity. Furthermore, by ensuring linearity of signal with exposure (radioactivity), these operations enable comparison and combination of results from different chips.

The image of a DNA array was divided into cells which nominally contain two well-separated symmetrical spots of equal intensity. This permits the identification of artifacts, which will not be symmetrically positioned within the cell. Nanowell images could be similarly partitioned, with each cell containing a perfect-circle well centered in

the cell. Rings could thus be distinguished from puncture artifacts, as the former have curvature and positional symmetry while the latter are of ill-defined shape and placement.

In order to obtain a signal that is linear with the amount of radiation, it is necessary to integrate over the entire well. Image processing techniques can be used to obtain a reliable estimate of integrated spot intensity. The image intensity model employed by Pelizzari et al had three components: (i) a global background value, (ii) a local background value that varies across the chip, and (iii) one or more Gaussian peaks per cell. The use of a low energy radiation source such as ^{33}P can reduce the occurrence of signal saturation and "bleeding". Saturation can also be addressed by the three-dimensional fitting of spots using a mathematical model of expected peak shape, as used by Pelizzari *et al*. The saturated values of central pixels are replaced by the fitting function, and the analytical integral of the resulting fitted peak provides a measure of integrated peak density.

Conclusion

While the definition of protein function remains vague, one important realm of protein function is biochemical activity. Nanowell protein chips are highly suitable to biochemical assays. Microwells minimize cross-contamination and provide a liquid environment that prevents proteins from drying out. Furthermore, they require only a small amount of protein sample. However, their power as a high-throughput tool in proteomics is limited by reproducibility and accuracy. The systematic application of image analysis procedures to protein array data, such as that proposed by Pelizzari et al,, should increase the reliability of these data and thereby advance our understanding of protein function.

References

Zhu, H., *et al.* (2000) Analysis of yeast protein kinases using protein chips. *Nature Genetics* **26** (3): 283-289.

Pelizzari, C.A., Khodarev, N.N., Gupta, N., Calvin, D.P., Weichselbaum, R.R. (2000) Quantitative analysis of DNA array autoradiographs. *Nuc. Acids Res.* **28** (22): 4577-4581.