Minimum Information About a Peptide Array Experiment (MIAPepAE)

Gordon Botha, Judit Kumuthini*

Centre for Proteomic and Genomic Research (CPGR), Cape Town, South Africa

* corresponding author (jkumuthini@gmail.com)

Abstract

Peptide-array screening is currently a well-established high-throughput technique with growing numbers of applications. Peptide-array technology is used for protein recognition, quantification of peptide expression levels, and detection of protein-protein interactions. The use of protein/peptide arrays in medical life science studies is becoming increasingly widespread. Their increased use in diagnostic applications and protein function profiling calls for a standardised set of guidelines to be followed by future experimenters to enable reproducible, high-quality data and accurate findings.

We aim to provide preliminary guidelines describing the Minimum Information About a Peptide-Array Experiment (MIAPepAE). We propose a checklist of data and meta-data that should accompany a peptide-array experiment, and invite fellow researchers in the field to collaborate in this effort to create a sustainable and coherent set of guidelines for the benefit of the protein/peptide-array research community. Although this article focuses on spotting peptide arrays, MIAPepAE is intended to be a work-in-progress to be adopted for other peptide-array types, such as in situ synthesised peptide arrays.

Introduction

The use of protein/peptide arrays in medical life science studies is becoming increasingly widespread (Reimer et al., 2002; Wulfkuhle et al., 2003; Cretich *et al.*, 2006). Broadly speaking, they are used for two main purposes: diagnostic applications (bio-markers or antibody detection) and protein function profiling. Peptide arrays are powerful diagnostic tools, as they allow both multiple analyses of identical samples and single-instance analyses of differential samples. For example, they have been applied to immune-response profiling experiments by measuring antibody-antigen interactions (Davies et al., 2005; Ingvarsson *et al.*, 2008; Andresen and Grötzinger, 2009); they have also been instrumental in protein-function profiling studies (Katz et al., 2011), in part because they use very little sample material and can process many proteins in parallel (Haab, 2001), and partly also because they can quantify very low concentrations of protein (Korf *et al.*, 2008) and take into account protein/peptide tertiary structures. Overall, peptide arrays are becoming pivotal to protein studies, spurring developments in related fields.

The technology and methodology is steadily advancing, in terms of slide preparation (Kopf et al., 2005; Beyer *et al.*, 2006) and sample preparation (Ghazani *et al.*, 2006; Usui *et al.*, 2006), and, in turn, is leading bioinformaticians to develop new software tools (Li *et al.*, 2005) and Web applications (Li *et al.*, 2009). Specific statistical techniques for the analysis of peptide arrays have also been developed (Royce *et al.*, 2006). Furthermore, high-throughput sequencing methods, such as real time PCR (Heid *et al.*, 1996), have delivered

an abundance of genomic and proteomic data for many species (Love *et al.*, 1990; Blattner *et al.*, 1997; Dean *et al.*, 2002). With so much proteomic information and analysis tools available, it is inevitable that many more peptide-array experiments will be conducted in the foreseeable future.

Our aim is to provide preliminary guidelines for the Minimum Information About a Peptide-Array Experiment (MIAPepAE). We propose a checklist of data and meta-data that should accompany a peptide-array experiment, aiming to fulfill the following main objectives:

• MIAPepAE should provide authors, reviewers, editors and readers with the specifics required to critically evaluate, understand and reproduce a peptide-array experiment;

• MIAPepAE should provide sufficient information to aggregate/integrate similar experimental data, independently of the platform on which the experiment was performed;

• MIAPepAE should allow secondary data, such as clinical patient and epidemiology data, to be integrated, enabling the extraction of more meaningful information from peptide-array experiments.

We emphasise meta-data pertaining to the sample. Variation in preparation of protein/peptide samples and their assaying to the array slides can be a major contributor to experimental variation and, as such, warrants a focused effort toward the proposed guidelines.

In the interest of coherent and coordinated development of such guidelines, the project is registered on the [MIBBI porta](http://mibbi.org/index.php/MIBBI_portal)[1]. The MIBBI project is a collaboration between leaders in the biological and biomedical fields, acting as a meeting point for the coordination of minimum information guidelines and checklists (Taylor *et al.*, 2008)

We have also based our checklist format on the guidelines for peptide-array experiments provided in the Minimum Information About a Proteomics Experiment (MIAPE) article (Taylor *et al.*, 2007), and the Minimum Information About a Microarray Experiment (MIAME) article (Brazma *et al.*, 2001b). The original MIAME checklist for microarray experiments has been revised (see (Abeygunawardena, 2007)), and we have based

our checklist for peptide-arrays on the revised checklist. Hence, we have drafted our checklist with the following main subjects: Raw Data, Final Processed Data for Set of Hybridisations, Sample Annotation and Experimental Factors, Experimental Design, Sufficient Annotation of Array Design, Essential Experimental and Data-Processing Protocols.

We endeavour to adhere to two criteria introduced by the MIAPE article: those of *Sufficiency* and *Practicability*. *Sufficiency* states that the minimum information requirements are constructed in such a way that the reviewer is able to "understand and critically evaluate the interpretation and conclusions". The reader must also be able to support the findings. *Practicability* states that the incorporation of a minimum information requirement for a proteomic experiment need not be so taxing on the experimenters that its adoption is impaired.

The checklist is still under development and will undoubtedly undergo revision as more peptide array experiments are performed and more comments and suggestions from colleagues in the field are incorporated.

Key Concepts

Our approach towards the formulation of guidelines for a peptide-array experiment takes several key concepts into account. These need to be defined clearly before proceeding, as this is necessary to interpret our guidelines.

Microarray Nomenclature

We have compiled a nomenclature from previous definitions (Brazma et al., 2001a; Royce *et al.*, 2006). The molecules bonded to the slide at the time of manufacture are termed *probes*. Any subsequent binding molecules are termed *targets*. A *spot* or *feature* is defined as a group of probes with identical sequences, concentrated at a known position on the microarray. A group of *targets* from the same biological entity is defined as a *sample*. One instance of the introduction of one or more samples to the array is known as *probing*. Finally, a series of probing to investigate a hypothesis is known as an *experiment*.

Unique Peptide

A unique peptide, as used in a peptide-array experiment, should conform to the following

_____
1　http://mibbi.org/index.php/MIBBI_portal

properties. It should have a unique identification number (ID) such as a National Centre for Biotechnology Information (NCBI) or a Protein Data Bank (PDB) number. If the peptide is synthetic, the full amino-acid sequence must be made available. A list of the protein(s) in which the peptide can be found should be given, including the starting position in the protein. The peptide length should be specified, and the overlap used when

aligning the peptide to a protein. Finally, any unidentified/ambiguous amino acids within the peptide sequence must be noted.

Table 1. MIAPepAE checklist for authors, revie-wers and editors. All essential information (E) must be submitted with the manuscript. Desirable information (D) should be submitted if available.

EXPERIMENTER INFO		
Author (submitter), laboratory, contact information (e-mail, postal address), links (URL), citation		

RAW DATA		
Typically, these are the data-files produced by microarray image-analysis software		

	IMPORTANCE	CHECKLIST
Raw data-files provided		
Native format	E	
Type: *e.g.*, image, binary data	D	
The file matches the respective array design	D	
Scanned image files for each slide	D	
Data location	E	

FINAL PROCESSED DATA FOR SET OF HYBRIDISATIONS (EXPERIMENT)		
Normalised/Summarised data on which conclusions are based		

	IMPORTANCE	CHECKLIST
Processed (normalised) data-files	E	
Normalisation application: *e.g.*, pin-to-pin, array-to-array, slide-to-slide, background correction	E	
Normalisation method	E	
The identifiers match the array annotation/location	D	
Control(s) on which normalisation was based	E	

SAMPLE ANNOTATION & EXPERIMENTAL FACTORS		
Describes the key experimental variables in the experiment. Additional information regarding sample, such as storage conditions, preparation methods, etc., are of great importance.		

	IMPORTANCE	CHECKLIST
Basic experimental factors (dose, time, disease state, treatment) provided for all samples	E	
Additional sample information		
Sample type	D	
Sample storage condition	D	
Sample dilution buffer	D	

	Sample name/annotation	E	
	Sample dilution used in the assay	E	
	Blocking agent	D	
	Detection antibody	E	
	Concentration of detection antibody	E	
	Hybridisation and washing conditions	D	
	Type of dye	D	
	Source organism (NCBI ta-xo-nomy)	D	
	Laboratory protocol for sample treatment (name, version, availability)	D	
Any post-printing processing, including cross-linking			
	Protein from which peptide was extracted (incl. ID) (NCBI/ UniProtKB/ SwissProt)	E	
	Peptide position in protein	E	
	Peptide overlap in protein alignment	E	
	Peptide conservancy	E	
	Peptide/protein sequence ID (NCBI/ UniProtKB /SwissProt)	E	

EXPERIMENTAL DESIGN

Describes the basic way in which the experiment was set up. Associations between samples and raw data generated from using these samples are critical. Note that the representation of an experimental design is best done via a graphical representation. The MAGE-TAB spreadsheet template (see text) provides a simple format for encoding such graphs.

		IMPORTANCE	CHECKLIST
Experimental design description			
	Table showing (sample) - (raw-data file) associations	E	
	Essential relationships between sample and array biomaterial noted	E	
	Experiment variables: *e.g.,* treated vs untreated	E	
Replicates			
	Identify which, if any, of the arrays are replicates	E	
	Identify whether replicates are technical/biological	E	

SUFFICIENT ANNOTATION OF ARRAY DESIGN

Essential information regarding array design, such as layout, probe information, slide surface preparation, etc.

		IMPORTANCE	CHECKLIST
Probe sequence information			
	Probe sequence database ID or complete peptide sequence, if synthetic for every probe**	D	
	**: Disclosure of the probe sequence is highly desirable and strongly encouraged. However, as not all commercial pre-designed assay vendors provide this information, it cannot be an essential requirement. Use of such assays is advised against.		
Controls			
	Positive controls, incl. sequence	E	
	Negative controls, incl. sequence	E	
	Synthetic/organic	E	
	Other buffer or empty spots?	E	
Array Design			
	GenePix Array List GAL file (or similar) with complete grid and labelling for all probes on array (incl. replicates, controls, sequence, and annotation if possible)	E	
	Surface type	D	
	Number of pins per array	E	
Slide Preparation			
	Number of array per slide	E	
	Preparation info (blocked, *etc.*)	D	

ESSENTIAL EXPERIMENTAL AND DATA PROCESSING PROTOCOLS

Essential experimental and data-processing protocols are typically described in the methodology/method. If protocols that allow for variable/user-defined variables are used, these must be adequately described. As for novel analysis methods, the protocol should be sufficiently documented to allow a reviewer to fully understand the process involved. Most software packages are able to output these parameter settings into files such as ArrayPro 'Spot Descriptor' or ArrayPro 'Grid Overlay' files (APG).

		IMPORTANCE	CHECKLIST
Spot intensities			
	Method (cell boundary definition/edge detection, *etc.*)	D	

	Pixels per spot	D	
	Spot dimensions (approximate diameter)	D	
	Spot local background dimensions (approximate diameter)		
	Net intensity calculation (raw minus mean background, raw minus spot background, *etc*.)	E	
Grid-finding methodology			
	Grid layout file (incl. spacing between sub-grids, grid rotation, spot shape and size, *etc*.)	E	
Background intensities			
	Method (local ring, local corners, global from image, global from background cells, *etc*.)	E	
Normalisation			
	Method (Loess, quantile, scaling, *etc*.)	E	
	Normalisation parameter (mean, median, *etc*.)	E	
	Spots used for normalisation (controls, all, subset, *etc*.)	E	
Instruments used			
	Scanner name	D	
	Model	D	
	Proprietary software name, version	D	
Data-extraction software used			
	Name	E	
	Version	D	
	Gain setting	E	
	Minimum threshold	E	
	Macro or script used for data extraction	D	
	Settings file	E	
Data Filtration method			
	Negative controls	D	
	Signal qualities (…from PROCAT)	D	
	Flagged spots criteria	E	
	Criteria 1	E	
	Criteria 2, *etc*.	E	

Reproducibility

Our guidelines aim to maximise the reproducibility of an experiment. They will also ease the interpretation of findings by peers, as a clear idea of experimental procedures will more effectively orient a reviewer.

Comparability and Re-usability

Another key concept that we want to capture in the guidelines is that of platform-independent comparability. Findings between studies can be compared effectively if standardised data formats are in place. Furthermore, if data are extracted in a concise and correct manner, they can be used in subsequent experiments. We feel that quality of data supersedes quantity of data, and using a concise method of data extraction from peptide arrays can greatly increase the experimenter's ability to sort biological meaning from experimental error.

Specificity

The specificity of an experiment, or an experiment in a more general case, measures the ability to correctly classify positive events. In a peptide-array context, this could measure the probability that a peptide/protein-binding event is in fact a specific binding, and hence biologically significant, and not due to a non-specific binding event or experimental error.

Quantification

The crux of the guidelines is to enable correct quantification of spot intensity within an array. It is paramount that spot intensities are biologically significant readings and not the result of experimental variation. The correct quantification of experimental parameters lends itself to effective verification of findings. We aim to achieve this with the proposed guidelines.

Conclusion

We have provided a checklist for capturing essential information when conducting peptide-array experiments. By conforming to this checklist, experimenters will:

provide authors, reviewers, editors and readers with the specifics required to critically evaluate, understand and reproduce a peptide-array experiment. This will lead to more accurate conclusions and higher quality data;

be able to compare and combine experiments across different platforms, greatly enhancing the re-usability of the data;

be able to extract and combine meta-data from experiments that might bring to light interesting observations. In so doing, experimental data can be utilised fully to discover biologically relevant observations.

The MIAPepAE form/format contains most of the required fields and sections in one document type, and allows for continuous updating as procedural standards become apparent from discussions within the community. Certainly, as technological advancements are made, the guideline will be appropriately adjusted. The document is version controlled and is available on the MIBBI portal.

In the interests of speeding up adoption of the MIAPepAE checklist in peptide-array experiments, we urge experimenters to provide at least the essential fields in an electronic format with published data and articles. Only with other researchers' input can the ease of conforming to the standards, and accuracy of field prioritisation within the checklist, be assessed. We do, however, note that, for the full benefits of the MIAPepAE guidelines to be reached, project conformity will have to be enforced at a higher level. Like other minimum information protocols, compliance can be required for: i) the publication of research articles (at journal level); ii) data submission to proprietary and public data repositories (at project and framework level); iii) funding and grant proposals (from funders); and possibly iv) encouragement from open-source project repositories.

We hope that the MIAPepAE guidelines are useful to data generators, data consumers and end users. This will, however, depend entirely on the willingness of the scientific community to adopt the guidelines and, more importantly, the willingness of fellow peptide-array experimenters to contribute to (and criticise) the development of the guidelines. In the end, the success of this project depends entirely on the community that it serves.

References

1. Andresen H, and Grötzinger C (2009). Deciphering the antibodyome - peptide arrays for serum antibody biomarker diagnostics. *Current Proteomics* **6 (1)**, 1-12.

2. Beyer M, Felgenhauer T, Ralf Bischoff F, Breitling F, and Stadler V (2006). A novel glass slide-based peptide array support with high functionality resisting non-specific protein adsorption. *Biomaterials* **27**, 3505–3514.

3. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland *et al.* (1997). The Complete Genome Sequence of Escherichia coli K-12. *Science* **277**, 1453-1462.

4. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman et al. (2001a). Minimum information about a microarray experiment (MIAME) toward standards for microarray data. *Nat Genet* **29**, 365-371.

5. Cretich M, Damin F, Pirri G and Chiari M (2006). Protein and peptide arrays: Recent trends and new directions. *Biomolecular Engineering* **23**, 77-88.

6. Davies DH, Liang X, Hernandez JE, Randall A, Hirst S et al. (2005). Profiling the humoral immune response to infection by using proteome microarrays: High-throughput vaccine and diagnostic antigen discovery. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 547-552.

7. Dean FB, Hosono S, Fang L, Wu X, Faruqi *et al.* (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 5261-5266.

8. Ghazani AA, Lee JA, Klostranec J, Xiang Q, Dacosta *et al.* (2006). High Throughput Quantification of Protein Expression of Cancer Antigens in Tissue Microarray Using Quantum Dot Nanocrystals. *Nano Letters* **6**, 2881-2886.

9. Haab BB (2001). Advances in protein microarray technology for protein expression and interaction profiling. *Curr Opin Drug Discov Devel* **4**, 116-123.

10. Heid CA, Stevens J, Livak KJ and Williams PM (1996). Real time quantitative PCR. *Genome Research* **6**, 986-994.

11. Ingvarsson J, Wingren C, Carlsson A, Ellmark P, Wahren *et al.* (2008). Detection of pancreatic cancer using antibody microarray-based serum protein profiling. *Proteomics* **8**, 2211–2219.

12. Katz C, Levy-Beladev L, Rotem-Bamberger S, Rito T, Rüdiger SGD *et al.* (2011). Studying protein–protein interactions using peptide arrays. *Chem. Soc. Rev.* **40**, 2131-2145

13. Kopf E, Shnitzer D and Zharhary D (2005). Panorama Ab Microarray Cell Signaling kit: a unique tool for protein expression analysis. *Proteomics* **5**, 2412–2416.

14. Korf U, Derdak S, Tresch A, Henjes F, Schumacher et al. (2008). Quantitative protein microarrays for time-resolved measurements of protein phosphorylation. *Proteomics* **8**, 4603–4612.

15. Li T, Zuo Z, Zhu Q. Hong A, Zhou X and Gao X (2009). Web-based design of peptide microarrays using microPepArray Pro. *Methods Mol. Biol* **570**, 391–401.

16. Li X, Yi EC, Kemp CJ, Zhang H and Aebersold R (2005). A Software Suite for the Generation and Comparison of Peptide Arrays from Sets of Data Collected by Liquid Chromatography-Mass Spectrometry. *Molecular & Cellular Proteomics* **4**, 1328 –1340.

17. Love JM, Knight AM, McAleer MA and Todd JA (1990). Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites. *Nucleic Acids Research* **18**, 4123-4130.

18. Niran Abeygunawardena (2007). MIAME 2.0 - MIAME – FGED http://www.mged.org/Workgroups/MIAME/miame.html.

19. Reimer U, Reineke U and Schneider-Mergener J (2002). Peptide arrays: from macro to micro. *Current Opinion in Biotechnology* **13**, 315–320.

20. Royce TE, Rozowsky JS, Luscombe NM, Emanuelsson O, Yu *et al.* (2006). Extrapolating traditional DNA microarray statistics to tiling and protein microarray technologies. *Meth. Enzymol* **411**, 282-311.

21. Sakanyan V (2005). High-throughput and multiplexed protein array technology: protein-DNA and protein-protein interactions. *Journal of Chromatography B* **815**, **77–95**.

22. Stears RL, Martinsky T and Schena M (2003). Trends in microarray analysis. *Nat Med* **9**, 140–145.

23. Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R *et al.* (2008). Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat Biotech* **26**, 889–896.

24. Taylor CF, Paton NW, Lilley KS, Binz PA, Julian *et al.* (2007). The minimum information about a proteomics experiment (MIAPE). *Nat Biotech* **25**, 887–893.

25. Usui K, Tomizaki K, Ohyama T, Nokihara K and Mihara H (2006). A novel peptide microarray for protein detection and analysis utilizing a dry peptide array system. *Mol. BioSyst.* **2**, 113.

26. Wulfkuhle JD, Aquino JA, Calvert VS, Fishman DA, Coukos G *et al.* (2003). Signal pathway profiling of ovarian cancer from human tissue specimens using reverse-phase protein microarrays. *Proteomics* **3**, 2085-2090.