OPEN ACCESS Freely available online

Fungal Genomics & Biology

ISSN: 2165-8056

Mini Review

The Growth of sPEP-Specific Techniques in Fungi

Paige E Erpf[1*], **James A Fraser**[2]

[1]*Department of Synthetic Biology Natural Sciences, Macquarie University, Sydney, Australia;* [2]*Australian Infectious Diseases Research Centre, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia*

ABSTRACT

Short open reading frame (sORF)-encoded polypeptides (sPEPs) have become an increasing focus of investigation over the last decade, both due to the development of sPEP-specific experimental techniques and the growing understanding of the importance of sORFs in biological functions. In this mini review, we highlight the advancements made in sPEP identification and characterization across several fungi species. Here we briefly describe some of the techniques that have been developed to facilitate identification and characterization of sORFs and sPEPs across a range of fungal species.

Keywords: Fungi; sPEPs; *Saccharomyces cerevisiae*; *Schizosaccharomyces pombe*; *Cryptococcus neoformans*; sORF; Proteomics; Ribosome profiling; CRISPR/Cas9

INTRODUCTION

The acceleration in research around short open reading frames (sORFs) and their encoded polypeptides has been enabled through a variety of advances in proteomics, ribosome profiling (ribo-seq), and bioinformatics [1-5]. sORF-encoded polypeptides (sPEPs) are proteins no larger than 150 amino acids in length, originating from short ORFs found in genomic contexts such as the 5' and 3' untranslated regions (UTRs) of mRNAs encoding larger translated ORFs, in alternative reading frames of mRNA producing larger translated ORFs, and in RNAs erroneously annotated as "non-coding". They have been found across the domains of life, including bacteria, insects, plants, humans, and fungi [6].

Some of the best characterized sPEPs come from fungi, particularly the ascomycete budding yeast *Saccharomyces cerevisiae*, and many of these brewer's yeast sPEPs are conserved in many other fungal species [7]. Even though these studies in *S. cerevisiae* have significantly contributed to the understanding of sORFs in eukaryote genomes and sPEPs in proteomes, the identification, validation, and characterization of sPEPs in many organisms still remains limited; despite the knowledge that they exist, the genomic locations that encode them, their low abundance, and small size continues to make them difficult to identify. To help address this many proteomic protocols now enrich for sPEPs to overcome these hurdles and utilize a combination of analytical methods optimized for their identification [8-10].

HIGH-THROUGHPUT CRISPR/CAS9 TO CHARACTERIZE sPEPS IN *S. CEREVISIAE*

As one of the first sequenced eukaryotes, *S. cerevisiae* has a comprehensive repertoire of databases available, including some dedicated to sORFs [11,12], yet these genomic features still remain relatively understudied in this important species.

A primary reason sORFs can be difficult to genetically manipulate is because they often residue in UTRs, out of frame to larger ORFs, or in overlapping genes. Disruptions of these genomic locations may lead to confounding results where phenotypes cannot always be attributed to just the loss of the sORF and its product.

Markless mutagenesis can address some of these issues by ensuring phenotypic artefacts do not arise due to marker-driven positional effects. To address the difficulty of working with sORFs, a markless CRISPR/Cas9 strategy was developed for sORF mutagenesis in *S. cerevisiae* [13]. Using a "guide+donor" strategy to investigate 315 poorly characterized sORFs, high-throughput mutagenesis enabled the generation of pools of mutants *via* Cas9-mediated deletion of the first 60 base pairs of

Correspondence to: Paige E Erpf, Department of Synthetic Biology Natural Sciences, Macquarie University, Sydney, Australia, E-mail: paige.erpf@mq.edu.au

Received: 28-Jun-2022, Manuscript No. FGB-22-17331; **Editor assigned:** 06-Jul-2022, PreQC No. FGB-22-17331 (PQ); **Reviewed:** 25-Jul-2022, QC No. FGB-22-17331; **Revised:** 29-Jul-2022, Manuscript No. FGB-22-17331 (R); **Published:** 05-Aug-2022, DOI: 10.35841/2165-8056.22.12.192

Citation: Erpf PE, Fraser JA (2022) The Growth of sPEP-Specific Techniques in Fungi. Fungal Genom Biol.12: 192.

the sORF 5' UTR, usually including the start codon; barcoding then facilitated sORF mutation verification using high-throughput sequencing [13]. The mutant library was screened for essentiality/non-essentiality and cellular fitness. None of the 315 sORFs were found to be essential, but 68 played a role in fitness. The high- throughput method proved to be a rapid technique for simultaneously screening the phenotypes of hundreds of sORFs and their encoded products.

RIBOSOME PROFILING REVEALS SPEPS DURING *SCHIZOSACCHAROMYCES POMBE MEIOSIS*

The development of robust sORF databases and repositories has been accelerated through advancements in ribo-seq and proteomics. An excellent example is how the combination of ribo-seq, genomics and proteomics (proteogenomics) has enabled not only the identification of sORFs but also their encoded product, key data required to validate sORF existence and potential function in the proteome.

A combination of the classic technique of ribosome profiling with high throughput sequencing methodologies, ribo-seq provides a global snapshot of translation *in vivo* [14]. The underlying mechanism behind ribo-seq involves stalling ribosomes with protein translation inhibitors such as cycloheximide, harringtonine, and lactimidomycin.

Upon treatment of isolated polysomes with RNAse, stalled ribosomes create ribosome protected fragments (RPF) along the transcript being translated that are then purified and subsequently amplified, sequenced, and mapped. As a result, features such as upstream translated ORFs, sORFs, start sites, novel coding regions, folded nascent peptides, and specific protein products have been identified and quantified [14].

An excellent example of the power and utility of ribo-seq in sPEP research has been shown in the ascomycete fission yeast *Schizosaccharomyces pombe*, an unparalleled model for analysis of the cell cycle. Meiosis can be induced through the inactivation of a conditional mutant of the protein kinase Pat1, a negative regulator of the RNA-binding protein Mei2, an initiator of meiotic division. Because induction of meiotic division can be done with ease, it was one of the first organisms to undergo ribo-seq to study the translational landscape of diploid cells undergoing sexual differentiation [15,16]. Within this study, 373 sORFs were captured as being translated during meiosis by detecting polysome binding of their mRNAs. However, ribo-seq only provides a snapshot of translation occurring, not if the protein produced has a function [16].

Huraiova and colleagues used stable isotope labelling by amino acids cell culture (SILAC) quantitative proteomics in an attempt to identify peptides corresponding to the 373 sORFs [8]. SILAC-based proteomics compares the relative levels of proteins present during various stages of meiosis. As a consequence, any protein being expressed, including sPEPs, has the potential of being detected through Mass Spectrometry (MS) at any given time of the 373 originally identified sORFs, only nine were validated using this approach. The researchers noted that the low detection of small encoded proteins could be accredited to the absence of a sPEP enrichment step, a process that could have been included during MS preparation. Enrichment is important as it removes larger proteins dominating the proteome which mask the presence of bona fide sPEPs. Of the nine verified sPEPs, five were selected for gene deletion and phenotypical analysis.

Unlike Guo and colleagues' study where a markless strategy was employed for sORF mutagenesis, here mutants were created using a drug resistance marker and homologous recombination. Vegetative growth, cell morphology, sensitivity to DNA damaging agents, chromosome segregation and spore viability were all tested alongside a wild-type control. For every test condition, mutants had a similar phenotype to the wild-type, and no major meiotic defects were observed. However, this does not preclude the identified sPEPs playing a role in the cell, as they may function under different growth conditions that were not tested in this study, or have a subtle effect during meiosis which may require more sensitive analysis such as competition assays across multiple rounds of meiotic division.

sPEP-OPTIMIZED PROTEOGENOMICS UNMASKS sPEPS IN THE HUMAN FUNGAL *CRYPTOCOCCUS NEOFORMANS*

As the literature expands around sPEP identification, it has become increasingly clear that dedicated enrichment steps are essential to appropriately detect the presence of these small translated products, particularly when an organism has features that occlude the existence of sPEPs even further than usual. A good example of this is the pathogenic basidiomycete budding yeast *Cryptococcus neoformans* where a complex polysaccharide capsule combined with a fungal cell wall makes sPEPs extremely difficult to isolate [10].

A key step to enable sPEP enrichment in *C. neoformans* was the development of a lysis method that would overcome both the polysaccharide capsule and cell wall. Bead beating in conjugation with a lysis buffer containing the chaotropic salt guanidine hydrochloride proved to be the superior option of four different lysis agents tested for capsule and cell wall removal. For enrichment, C18 reversed-phase cartridges provided a highly hydrophobic stationary phase which retained larger proteins more securely than sPEPs that were easily eluted with an organic solvent, acetonitrile. Samples prepared in this way exhibited twice the number of unique sPEPs compared to unenriched samples when analyzed *via* LC-MS/MS. In addition to validating 215 previously predicted sORFs as encoding protein products, a total of 60 novel, previously unannotated sPEP-encoding sORFs were also identified using the enriched samples, and 41 of these were further validated through comparison with ribo-seq data [17].

As a proof-of-principle, a novel *C. neoformans* 38 amino acid sPEP originating from a "non-coding" RNA was selected for further characterization. As is typical for this species, the transcript encoding the sPEP was in a difficult to manipulate genomic location, with its 3' UTR overlapping the 3' UTR of the adjacent

convergently transcribed gene. A markless mutagenesis strategy was therefore employed using the newly developed acetamidase amdS2 Blaster dominant counterselectable marker [18]. This marker can be selected on media with acetamide as the sole nitrogen source, and counter-selected on media containing the toxic acetamide analogue fluoroacetamide. For sPEP characterization, direct repeats flanking either side of the amdS2 marker were designed to target integration to the sORF locus, introducing a substitution mutation along with the marker. Following selection on acetamide media and integration of the construct into the desired genomic location, counterselection on media containing fluoroacetamide identified cells where spontaneous mitotic recombination between direct repeats flanking the marker left behind only the single nucleotide point mutation.

The sPEP mutant exhibited a proteomic profile with changes in abundances to proteins involved in the electron transport chain, glycolysis, oxidoreductase activity, and oxidative stress [10]. Another study also found a similar proteomic profile to mutants of the bacterial SsrA-binding protein (SmpB), a protein part of a ribonucleoprotein complex which mediates the trans-translational surveillance system of stalled ribosomes on incomplete or damaged mRNAs [19-21]. To further support the potential function of the *C. neoformans* sPEP, in silico analyses revealed a similar protein secondary structure to a domain within SmpB. Overall, the data suggests there is an analogous function to trans-translation in fungi which involves this novel sPEP (dubbed Nano-Protein B1, Npb1), although additional research on this small protein and its interacting partners is needed to further support this theory.

DISCUSSION

Until recently, the identification and characterization of sPEPs has been limited due to the lack of techniques available to analyze these small products of translation. The literature cited herein has highlighted the growing nature of the field and the development of techniques focused on sPEP identification and characterization, specifically markless mutagenesis for accurate phenotyping and methods of enrichment of sPEPs to facilitate MS detection [8,10,13].

Genetically, the development of different forms of markless mutagenesis has played a crucial role in determining the function of sPEPs as their sORFs often reside in areas of the genome challenging to manipulate. The clean introduction of SNPs and/or indels reduces the possibility of observing false phenotypes through interference with adjacent genomic features. At the proteomic level, enrichment has proven to be essential for effective sPEP detection within MS samples; poor identification can be attributed to larger proteins masking sPEP presence within the proteome, and even the slightest proteolysis can produce proteolytic fragments that resemble real sPEPs.

CONCLUSION

Finally, combining genomic and proteomic techniques through ribo-seq advances detection even further, perform the critical role of assisting in validation of sORFs and their sPEPs. Given how quickly the tools to identify and characterize sPEPs have advanced over the last decade across numerous species, it is expected that this emerging field of study will continue to grow to uncover the stories of even more sPEPs in fungi and beyond.

REFERENCES

1. Stringer A, Smith C, Mangano K, Wade JT. Identification of Novel Translated Small Open Reading Frames in Escherichia coli Using Complementary Ribosome Profiling Approaches. J Bacteriol. 2021 ; 204(1):e00352-21.
2. Olexiouk V, Van Criekinge W, Menschaert G. An update on sORFs. org: a repository of small ORFs identified by ribosome profiling. Nucleic Acids Res. 2018;46(D1):D497-D502.
3. Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, et al. Peptidomic discovery of short open reading frame–encoded peptides in human cells. Nat Chem Biol. 2013;9(1):59-64.
4. Ma J, Diedrich JK, Jungreis I, Donaldson C, Vaughan J, Kellis M, et al. Improved identification and analysis of small open reading frame encoded polypeptides. Anal. Chem. 2016;88(7):3967-3975.
5. Verbruggen S, Ndah E, Van Criekinge W, Gessulat S, Kuster B, Wilhelm M, et al. Proteoformer 2.0: further developments in the ribosome profiling-assisted proteogenomic hunt for new proteoforms. Mol Cell Proteom. 2019;18(8):S126-S140.
6. Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet. 2014;15(3):193-204.
7. Erpf PE, Fraser JA. The long history of the diverse roles of short ORFs: sPEPs in Fungi. Proteomics. 2018;18(10):1700219.
8. Huraiova B, Kanovits J, Polakova SB, Cipak L, Benko Z, Sevcovicova A, et al. Proteomic analysis of meiosis and characterization of novel short open reading frames in the fission yeast *Schizosaccharomyces pombe*. Cell Cycle. 2020;19(14):1777-1785.
9. Wang N, Yin Z, Duan W, Zhang X, Pi L, Zhang Y, et al. sORF-encoded polypeptide SEP1 is a novel virulence factor of Phytophthora pathogens. Mol. Plant Microbe Interact. 2021;34(2):157-167.
10. Erpf PE, Chua SM, Phung TK, Kerr ED, Rothnagel JA, Schulz BL, et al. Identification and characterisation of sPEPs in *Cryptococcus neoformans*. Fungal Genet Biol. 2022;160:103688.
11. Kastenmayer JP, Ni L, Chu A, Kitchen LE, Au WC, Yang H, et al. Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. Genome Res. 2006;16(3):365-373.
12. Hao Y, Zhang L, Niu Y, Cai T, Luo J, He S, et al. SmProt: a database of small proteins encoded by annotated coding and non-coding RNA loci. Brief. Bioinform. 2018;19(4):636-643.
13. Guo X, Chavez A, Tung A, Chan Y, Kaas C, Yin Y, et al. High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR–Cas9 in yeast. Nat Biotechnol. 2018;36(6):540-546.
14. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. science. 2009;324(5924):218-23.
15. Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. High-resolution view of the yeast meiotic program revealed by ribosome profiling. science. 2012;335(6068):552-557.
16. Duncan CD, Mata J. The translational landscape of fission-yeast meiosis and sporulation. Nat Struct Biol. 2014;21(7):641-647.
17. Wallace EW, Maufrais C, Sales-Lee J, Tuck LR, de Oliveira L, Feuerbach F, et al. Quantitative global studies reveal differential translational control by start codon context across the fungal kingdom. Nucleic Acids Res. 2020;48(5):2312-2331.

18. Erpf PE, Stephenson CJ, Fraser JA. amdS as a dominant recyclable marker in *Cryptococcus neoformans*. Fungal Genet Biol. 2019;131:103241.
19. Liu Z, Hu K, Tang Y, Li H, Tang H, Hu X, e al. SmpB down-regulates proton-motive force for the persister tolerance to aminoglycosides in Aeromonas veronii. Biochem Biophys Res Commun. 2018;507(1-4):407-413.
20. Himeno H, Kurita D, Muto A. tmRNA-mediated trans-translation as the major ribosome rescue system in a bacterial cell. Front genet. 2014;5:66.
21. Neubauer C, Gillet R, Kelley AC, Ramakrishnan V. Decoding in the absence of a codon by tmRNA and SmpB in the ribosome. Science. 2012;335(6074):1366-1369.