Proteomics on the rims: insights into the biology of the nuclear envelope and flagellar pocket of trypanosomes

MARK C. FIELD1*, VINCENT ADUNG’A1, SAMSON OBADO2, BRIAN T. CHAIT3 and MICHAEL P. ROUT2

1 Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, UK, 2 Laboratories of Cellular and Structural Biology and 3 Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, USA

(Received 14 September 2011; revised 31 October 2011; accepted 3 November 2011)

SUMMARY

Trypanosomatids represent the causative agents of major diseases in humans, livestock and plants, with inevitable suffering and economic hardship as a result. They are also evolutionarily highly divergent organisms, and the many unique aspects of trypanosome biology provide opportunities in terms of identification of drug targets, the challenge of exploiting these putative targets and, at the same time, significant scope for exploration of novel and divergent cell biology. We can estimate from genome sequences that the degree of divergence of trypanosomes from animals and fungi is extreme, with perhaps one third to one half of predicted trypanosome proteins having no known function based on homology or recognizable protein domains/architecture. Two highly important aspects of trypanosome biology are the flagellar pocket and the nuclear envelope, where in silico analysis clearly suggests great potential divergence in the proteome. The flagellar pocket is the sole site of endo- and exocytosis in trypanosomes and plays important roles in immune evasion via variant surface glycoprotein (VSG) trafficking and providing a location for sequestration of various invariant receptors. The trypanosome nuclear envelope has been largely unexplored but, by analogy with higher eukaryotes, roles in the regulation of chromatin and most significantly, in controlling VSG gene expression are expected. Here we discuss recent successful proteomics-based approaches towards characterization of the nuclear envelope and the endocytic apparatus, the identification of conserved and novel trypanosomatid-specific features, and the implications of these findings.

Key words: Proteomics, nuclear pore complex, flagellar pocket, systems biology, molecular evolution, protein complex, protein interactions.

INTRODUCTION

The identification and definition of function remains a highly challenging aspect of modern biology, despite the recent huge technical advances for discovering new genes and their products. One major innovation in the past decade has been the vast improvement in mass spectrometric identification of polypeptides, brought about by a combination of improved instrument sensitivity together with computational resources that allow the assignment of mass spectra based on genome sequences (discussed in Chait, 2011). Coupled with the falling costs of large-scale DNA sequencing, many organisms are now amenable to molecular level analysis for the first time as the economic ‘barriers to entry’ have begun to fade. The concept of a ‘proteome’ has emerged through these technological advances, and to a first approximation is equivalent to the protein composition of a defined biological specimen, be it a whole organism lysate, a tissue fluid, an isolated organelle or a specific protein complex. The application of proteomic methods to pathogens has the exciting potential to inform broadly on functions relevant to virulence, host range restriction and immune evasion mechanisms. Examples include the identification of developmental changes through the life cycle, discovery of complexes responsible for signaling, gene expression and intracellular transport and potentially monitoring the responses of a pathogen to a drug or other insult.

The African trypanosomes, Trypanosoma brucei spp. are the causative agents of African sleeping sickness in humans and nagana in cattle (Simarro et al. 2010). These organisms have been, and remain, major causes for concern in terms of public health and agricultural productivity. Unquantified (and likely unquantifiable), but major impacts on the flora and fauna of Africa have resulted from infection of a great many animal species by T. brucei. Related organisms, including Euglena and the parasitic Leishmania and Phytomonas, most likely also exert huge impacts, both ecologically and economically, on much of the planet (see Camargo, 1999; Duttagupta et al. 2004; Antinori et al. 2011, for some examples and discussion). Many of the diseases associated with these organisms have been historically classed as ‘neglected’, in part due to the absence of high quality chemotherapeutic agents or vaccines with which to combat infections, but also due to the absence of a financial incentive as many of those afflicted live in the poorest parts of the world.
Fig. 1. Considerations for proteomics. Schematics to illustrate some of the conceptual and technical challenges that need to be addressed or considered when analyzing the protein composition of a cell, organelle or complex. Panel A: Much difficulty can lie in the selection of appropriate gating parameters, with the significant issue that, ab initio, it is almost impossible to know where these parameters should be set. An initial estimate may be gleaned from prior art or scouring the genome sequence for clues, but this can remain a very inaccurate process and is full of assumptions. In the examples, the gating is used to either stringently eliminate >95% of the experimentally detected proteome (left), to reject >70% as a moderate cut-off (centre) or to allow into consideration essentially all of the proteome (right). Depending on the question, the nature of the sample, the quality of the data and the quality of the available predicted proteome database (genome) a gate approximating one (or none) of these possibilities may be appropriate. Obviously, more stringent gating reduces the number of proteins to consider going forward but will, in all probability, exclude important information.

Panel B: Most proteins participate in a range of interactions, some of which are direct, while others are indirect. Conceptually one can consider a core of tight associations mediating the basic function of a given protein or complex (core, in red). However, this is biologically inaccurate as even tight complexes, organelles or other biological assemblies exist in association with other complexes and systems. These interactions become functionally as well as physically more tenuous and will eventually come to include proteins that are off target but which may still retain a genuine affinity for components of the target complex/organelle (light grey and blue). In some cases a given protein may be a bona fide member of several complexes; Sec13 in *S. cerevisiae* is now known to participate in at least three complexes, COPII, the NPC and the SEA complex. The point at which one considers such interactions to represent contaminants is hard to determine and, to some level, is subjective. Panel C: Considerations for the analysis of complex mixtures. Regardless of the appropriateness or not of the gating algorithm used, any initial list of proteins will contain a mixture of genuine components (‘realerons’, red), together with additional moderate or low probability members of the complex/organelle (‘mysterons’, black); probability here equates to additional evidence for membership to a complex. The biological sample will also contain contaminants, which are frequently impossible to exclude, as well as miscalls due to low signals, errors in the predictive algorithms, suppression and database quality problems. All of these criteria are variable (and
(Magez and Radwanska, 2009; Wilkinson and Kelly, 2009).

**TRYPANOSOMES, DIVERGENCE AND PROTEOMIC INSIGHTS**

The completion of a genome sequence for one strain of *T. brucei* yielded many insights into the biology of trypanosomes, and provided the vital framework for going forward with molecular level dissection of trypanosome biology (Berriman et al. 2005). This has been coupled with the emergence of RNA interference (RNAi) for suppression of gene expression in a conditional manner, RNA sequencing approaches to monitor transcription (Kolev et al. 2010) and most recently RNAi-based expression knockdown screens (RIT-seq, Alsford et al. 2011), with the result that our understanding of the cell biology and metabolism of *T. brucei* has advanced at an accelerated pace during the past five to ten years. However, many of the investigations in this period have been centered around ‘candidate’-based approaches, i.e. mining the genome for gene products with either known functions or at least functions in known processes or pathways or predictions based on similarity of either sequence or domain architectures; transcription, histone modification, intracellular trafficking and the cytoskeleton are all good examples of where this type of approach has been of great value (see Kawahara et al. 2008; Luz Ambrósio et al. 2009; Field and Carrington, 2009; Wickstead et al. 2010).

Regardless of how informative, this is nevertheless still an introspective strategy which ignores much of the potential novel biology and therapeutic opportunity within the trypanosome.

It is estimated that up to 50% of the trypanosome protein-coding content is ‘divergent’, in the sense that orthology or paralogy with higher eukaryote genes cannot be reliably established, opening up the potential for novel and trypanosome-specific functions. For example, the trypanosome kinase families seemingly appear highly divergent from higher eukaryotes, with few conserved domain architectures beyond the kinase domains themselves, making functional prediction extremely difficult (Parsons et al. 2005). It is, however, very likely that this 50% of novel gene products is an overestimate, as many orthologous relationships are simply too divergent to be detected by sequence-based algorithms alone. As we have demonstrated previously, the nuclear pore complex is apparently highly divergent based on just *in silico* analysis, but is in fact rather well conserved and simply an example of BLAST failing to identify highly diverged sequences (deGrasse et al. 2009). Thus, targeted proteomic analyses have a major role to play in this regard, as subcellular organelles, macromolecular structures and complexes can be isolated, using evolutionarily conserved handles, i.e. tagged proteins that are sufficiently well conserved for function to be confidently inferred, and potentially less well conserved proteins that are in association can be identified. This has the strategic advantage of finding genuinely trypanosome-specific proteins but which interact with known proteins. At a minimum, evidence for a likely role in a specific process or pathway is already available for the new protein, facilitating the design of strategies to test these hypotheses directly.

The quality of a proteome, and its ultimate utility, rests on sample preparation, and the rather difficult determination or definition of purity (Fig. 1 and legend). There are many strategies for preparing material for mass spectrometric analysis, ranging from whole cell extracts to more specific protein complexes, but in practice the issues remain similar, defining the material being analysed, being aware that there are likely contaminants and validating the proteins that are identified. Given that many complexes are known that participate in multiple interactions, the determination of specificity in itself is a very crucial step.

One approach we have found to be highly successful has been analysis of the trypanosome nuclear envelope. Using as starting material a highly enriched pore complex lamina fraction (Rout and Field, 2001), we were able to identify ~800 distinct peptides, amounting to 10% of the predicted *T. brucei* proteome (DeGrasse et al. 2009). A recent innovation, pioneered by one of us, is the isolation of protein complexes using a cryogrinding method that was originally developed to facilitate the lysis of *Saccharomyces cerevisiae* cells due to the presence of a cell wall that does not dissolve in mild detergents. We have successfully adapted this method to trypanosomes and demonstrated its usefulness using the NPC as a test-bed, as well as to elements of the nuclear skeleton/lamina and, more recently, cytoplasmic complexes. Although still at the development phase, the methodology promises to be generally applicable to trypanosomes (and maybe other
protists), with the potential to allow the elucidation of protein-protein interactions in *T. brucei* in a systematic manner (data not shown).

Here we briefly consider two major cellular systems of trypanosomes, the endocytic apparatus and the nuclear envelope (Fig. 2). Both of these systems are essential, possess numerous diverse functions and are likely contribute in unique ways to the biology of trypanosomes. We will also briefly discuss several approaches to the generation of biological samples for proteomics and some advantages and disadvantages of each.

DEFINING A CLATHRIN INTERACTOME – CLASSICAL ANTIBODY PULLOUT

Based on primary structure-based searches, membrane transport in *T. brucei* apparently lacks multiple major factors found in other species that are known to play important roles in transport (discussed in Field and Carrington, 2005). These apparent absences result in an ‘incomplete’, and paradoxically non-functional, predicted protein network with many of the trypanosome endocytic protein orthologues lacking identifiable binding partners (see Field et al. 2009). As the endocytic system clearly is functional, this observation implies either that many critical players are too divergent to be detected by sequence alone or that truly novel proteins substitute for these factors. To some extent, both of these possibilities have the same consequences and highly divergent proteins are predicted to be operating within the trafficking machinery of trypanosomes.

All endocytosis in trypanosomes proceeds *via* the flagellar pocket, a membrane invagination surrounding the base of the flagellum and uniquely excluded from the subpellicular microtubule array subtending the bulk of the plasma membrane. The parasite surface is dominated by the GPI-anchored VSG, and endocytosis operates at an extremely rapid rate in the bloodstream stage of the parasite (Pal *et al.* 2003; Engstler *et al.* 2004). Furthermore, all endocytosis appears to be clathrin-mediated, as RNAi against clathrin completely blocks endocytic activity (Allen *et al.* 2003), and many of the proteins that mediate clathrin-independent endocytosis are restricted to...
metazoan taxa and therefore absent from trypanosomatids and in fact the majority of organisms (Field et al. 2007). In addition, the AP-2 complex, which mediates recognition of cargo and their concentration in clathrin-coated pits is absent from T. brucei, and the major trans-membrane domain endocytic cargo, the invariant surface glycoproteins 65 and 75, both lack classical AP-based sorting signals and are internalized and degraded via a ubiquitylation-based mechanism (Chung et al. 2004, 2008; Leung et al. 2011).

We have also described a putative clathrin-associated sorting protein (CLASP), TbEpsinR, 2011). We have also described a putative clathrin-et al. 2007). In addition, the AP-2 complex, which mediates recognition of cargo and their concentration in clathrin-coated pits is absent from trypanosomatids and in fact the majority of organisms (Field and Carrington,2009). All three antigens demonstrate significant co-localization with clathrin (Fig. 3 and data not shown). These data suggest that, together with TbHsc70, TbCAP100 and TbCAP161 are potential members of a trypanosome clathrin interactome network. However, as the clathrin interactions likely span endocytosis, endocytic recycling and also post-Golgi complex sorting and transport, the precise functions of TbCAP100 and TbCAP161 cannot be inferred from location alone and more evidence is needed to clarify the roles of these proteins (Morgan et al. 2001; Grunfelder et al. 2003). For further confirmation of clathrin interaction, the HA-tagged proteins were immunoprecipitated from whole cell lysates and these precipitates demonstrated to also contain clathrin by Western analysis (Fig. 3). In summary, bi-directional co-immunoprecipitation and co-localization between TbHsc70, TbCAP100 and TbCAP161 with clathrin provides very strong evidence that TbHsc70, TbCAP100 and TbCAP161 are bona fide clathrin-binding proteins.

The highly significant finding here is that while TbHsc70 and TbCAP100 are clearly well conserved between trypanosomes and higher eukaryotes, TbCAP161 is lineage-specific and cannot be detected by homology searches beyond the Kinetoplastida (data not shown). A similar level of conserved and lineage-restricted interaction partners was described recently by us for trypanosome Rab11 (Gabernet-Castello et al. 2011). The major advantage of the direct antibody pull-out is that it is minimally invasive with wild-type parasites being the target. The method is also highly accessible – any good polyclonal is a potential starting point, and little in the way of specialised equipment is required. Of course, the approach may fail to capture...
many associated proteins due to dissociation and the absence of a second affinity step can increase the incidence of off-target hits. In the present example, there was a requirement for lysis to provide access to the intracellular antigen using partial detergent solubilisation which is potentially disruptive. Finally, there is the significant issue that the antibody resource is finite, and it is (too) easy to consume significant amounts of antibody in such experiments.

THE NUCLEAR ENVELOPE – SUBCELLULAR FRACTIONATION

The nuclear envelope is a defining feature of eukaryotic cells, and serves to separate nuclear activities such as transcription and mRNA processing from translation. Trypanosome nuclei are conventional in that they possess electron-dense material at the nuclear periphery which is most probably heterochromatin, a recognisable nucleolus and a double bilayer nuclear envelope which is contiguous with the endoplasmic reticulum. The limited evidence to hand has suggested that trypanosomes have comparatively conventional systems for RNA and protein import and export, as well as at least one direct study that demonstrated a probable conventional protein import pathway (Marchetti et al. 2000). However, novel aspects are likely present, for example in mRNA transport, where RNA-recognition motif (RRM) domains appear sufficient to specify nuclear accumulation, likely by binding to a specific RNA (Cassola and Frasch, 2009). Further, ultrastructural analysis of trypanosomes has indicated the presence of nuclear pore complexes (NPCs) that morphologically resemble those in opisthokont taxa (Rout and Field, 2001). The obvious importance of the NPC to mRNA export pathways is enhanced by the assumed absence of promoter control from expression of many protein-coding genes due to polycistronic transcription and the presence of a spliced leader at the 5′ end of mature coding mRNAs.

A problem similar to the identification of the endocytic protein network was initially presented for the analysis of the trypanosome nuclear envelope and specifically identification of NPC proteins (nucleoporins). Based on BLAST analyses very few nucleoporins could be identified or clear orthologous relationships derived which raised the possibility that the trypanosome NPC was genuinely highly divergent from metazoan and yeast NPCs (Mans et al. 2004). The opposing interpretation was that the NPC in fact conserved, but that the nucleoporins were too divergent to be identified using sequence-based searches. In fact the success in identifying potential
NPC orthologues was significantly poorer than for the endocytic network.

A solution has been to pursue a classical subcellular fractionation route, to produce a highly enriched pore complex lamina fraction, essentially the nuclear envelope following mild detergent extraction and again based on methods originally developed for *S. cerevisiae* (Rout and Field, 2001). Despite extensive validation of the pore complex lamina fraction by both immunocytochemistry and electron microscopy which suggested a comparatively simple fraction containing NPCs and some interconnecting fibres, over 800 distinct polypeptides could be identified by a combination of MS/MS and MALDI-MS approaches with varying degrees of confidence (deGrasse et al. 2008, 2009). This illustrates rather well an important and sobering point; despite careful monitoring of the purification and good evidence that the resulting material was comparatively homogenous, a surprisingly large proteome was uncovered, making the list of proteins identified difficult to discriminate from the cohort that might be identified from a highly impure or heterogeneous sample. In part, this is because of the high sensitivity of current mass spectrometric techniques but, in the face of such complexity, some confidence in the starting material (e.g. a validation of purity) is highly valuable before committing to further analysis with such a dataset. Informatics analysis, based on the observation that secondary structure is better conserved than primary structure, allowed the identification of candidate nucleoporins and finally validation by localization following *in situ* genomic tagging (Oberholzer et al. 2006; deGrasse et al. 2009; deGrasse and Devos, 2010). The important biological outcome of this study was that the trypanosome NPC is substantially more conserved with metazoan and yeast NPCs than sequence searches had indicated with major implications for the timing of evolution of the nucleus and sub-nuclear structures. Furthermore, the present data suggest that the NPC is more conserved than the clathrin interactome, despite the poorer return from *in silico* studies.

The advantages of subcellular fractionation are clear—the target is defined to some degree and selected to be highly biologically relevant, but the study does require a more significant investment in the design and validation of the isolation method than a simple antibody-based immunosolation. Some disadvantages are perhaps less obvious; surveying a subcellular fraction can be extremely challenging as the composition of the isolated material is likely to be highly complex. Studies of the plasma membrane, flagellum, pore complex lamina fraction and mitochondrial proteomes from trypanosomes have all revealed highly complex proteomes with the likelihood that much of this complexity is contaminants as well as putative novel components, making discrimination between these two categories vital (Broadhead et al. 2006; Bridges et al. 2008; Panigrahi et al. 2008; DeGrasse et al. 2009; Oberholzer et al. 2011). These problems are, of course, all layered upon standard concerns with purity that derive from any subcellular fractionation study.

**THE NUCLEAR PORE AND NUCLEOSKELETON COMPLEX – CRYOGRINDING**

A fuller understanding of protein function requires identification and characterisation of interaction partners as all proteins operate in the context of complexes and higher order structures, at least for part of their time. These interactions may be truly stable structures, at least on the seconds to minutes timescale, where examples include cytoskeletal elements and the NPC. More transient complexes are the clathrin triskelion cage and coatamer vesicle coats which assemble and disassemble on the seconds timescale. Examples of rapidly associating and dissociating relationships include small GTPase interactions with their constellation of effector molecules, which exhibit a range of binding affinities that vary from nanomolar to millimolar depending on both the interaction partner and the state of the complex (Will and Gallwitz, 2001; Wu et al. 2007). Affinities for protein-protein interactions can vary over several orders of magnitude, and $k_{on}$ rates likewise have a broad dynamic range, both of which can confound efforts to isolate intact complexes or reliable portions thereof. Immensely powerful though they are (see for example Panigrahi et al. 2006), affinity methods including the TAP-tag approach which use a dual affinity tag to attempt to minimise non-specific or off-target associations, suffer greatly from the prolonged period required for the pullout protocol. While facilitating the isolation of stable complexes, this can lead to loss of genuine low affinity interactions, as well as increase non-specific binding. Criticisms of other approaches to map interactions, such as yeast two hybrids, have been well rehearsed in the literature. *In vitro* or *in vivo* methods that rely on co-expression of two tagged proteins, either in a homologous system or, for example, in reticulocyte lysates, can suffer from overexpression artefacts as well as frequently non-physiological conditions and, of course, are not directed at the identification of novel factors.

One approach that avoids many of these issues is affinity isolation subsequent to cryogrinding by virtue of its rapidity, elimination of chaotropes, detergents or other factors from the generation of the initial lysate. This method preserves complexes or dynamic systems at the time of harvest, preventing mixing of components and it also reduces problems due to proteases, phosphatases or other potential factors that could degrade the sample as it is captured in liquid nitrogen. Further, inhibitors can be added to resuspension buffers when thawing the cryogrind and gentler buffer systems can be used that
potentialy preserve complexes that are disrupted under conditions where one biochemically lyses cells with stronger or high concentration detergents or chaotropes.

The method will be published in full elsewhere but, in brief, exploits a geological ball mill for the generation of a grindate under essentially native conditions, followed by pullout of the protein handle, normally a genomically-tagged C-terminal GFP fusion protein, using anti-GFP antibodies coupled to magnetic beads, and resolution by 1D SDS-PAGE (Cristea et al. 2005; Oberholzer et al. 2006; Oeffinger et al. 2007; Fig. 4). This method requires substantial quantities of starting material but can allow the isolation of very high quality complexes. The use of large quantities of starting material also allows exploration of many buffer conditions in parallel. We have successfully used this technique to extensively explore protein-protein interactions within the NPC, as well as to demonstrate interactions between the very large coiled-coil nucleoskeletal protein NUP-1 and the NPC as well as additional coiled-coil proteins that are present within the nucleus (MPR, BTC, MCF and SO, unpublished data). We have also begun to successfully apply the method to several complexes involved in membrane trafficking pathways, including clathrin coats and the exocyst complex.

While requiring an in situ-tagged gene as an affinity handle, this strategy has the advantage of permitting assembly of the handle into the native complex in a minimally invasive manner as the protein is expressed to normal levels and only one of the two copies in the diploid genome is modified. Remarkably, most proteins appear to tolerate addition of GFP to the C-terminus without major impact on cell viability or function, at least as assessed by proliferation and correct targeting of the tagged protein. Coupled to the near native conditions that can be used for lysis, efficient cryogenic entrapment of complexes and the speed of isolation which aids in recovery of low affinity interactions and minimises non-specific binding, we consider the method to be highly useful and with great potential. The major downside is the quantity of material currently required which limits the number of candidate proteins that can be considered. However, the insights that are emerging from this approach auger well for the future exploration of protein complexes in trypanosomes, and the potential identification of novel trypanosome-specific complex members and interactions.

CONCLUSIONS

There are many approaches to the generation of a proteome. The discrimination between contaminant and real component is a fundamental one and is perhaps even more critical in such a divergent system as Trypanosoma brucei where many of the proteins that are identified may be novel; often this is precisely what the investigator is hoping to find. Frequently there is no a priori basis for a judgment in this context, hence ‘mysteron’ (Fig. 1). Validation is therefore highly critical in this regard, although the
lengths to which investigators must go here varies to some extent on the context. For example, two flagellar proteomes have now been published (Broadhead et al. 2006; Oberholzer et al. 2011), and the quality of the data here attests to the fact that validation is robust. It is comparatively straightforward to ascertain location as a flagellar protein for example, while this is further assisted by the comparative ease of isolation of a flagellar fraction from these parasites. Internal organelles are substantially less easy to purify, and consequently validation is even more critical as the proteomes described have been extensive and clearly include contaminants as well as valuable novel components. Demonstrating the frequency of these latter is important for determining the ultimate usefulness of a proteome (Bridges et al. 2008; DeGrasse et al. 2009; Panigrahi et al. 2009).

New and improved methodologies for the purification of complexes and subcellular organelles to high purity are essential for the exploration of protein-protein interactions. The coupling of genomic-tagging methods, which target a single gene copy in a diploid genome, with cryogrinding has clear advantages over simple antibody-mediated pullouts. Cryogrinding can, of course, also be applied to a conventional antibody pullout experiment, although the use of a standardised affinity tag does have significant advantages in terms of defining optimal conditions for isolation. There are, however, some obvious issues, not least of which is the amount of material required for an essentially bulk preparation and the skills required to perform the preparations themselves. It is anticipated that the latter can be surmounted relatively quickly with formalization of several of these methods, and that this will offer the community a further tool for unravelling trypanosome biology. With about 50% of the trypanosome proteome assigned as hypothetical, together with unique cellular processes and organelles present in *T. brucei*, proteomic techniques are key towards realisation of the benefit of the genome sequences. Importantly, and together with other emerging techniques such as RNA interference target sequencing (Alsford et al. 2011), these approaches will greatly enhance our knowledge of this divergent organism and hasten drug target discovery and validation.

ACKNOWLEDGEMENTS

Funding for work described in this article has been provided by the Cambridge Commonwealth Scholarships program, the National Institutes of Health (RR00862 and RR022220) and the Wellcome Trust, which is gratefully acknowledged.

REFERENCES


Mark C. Field and others


