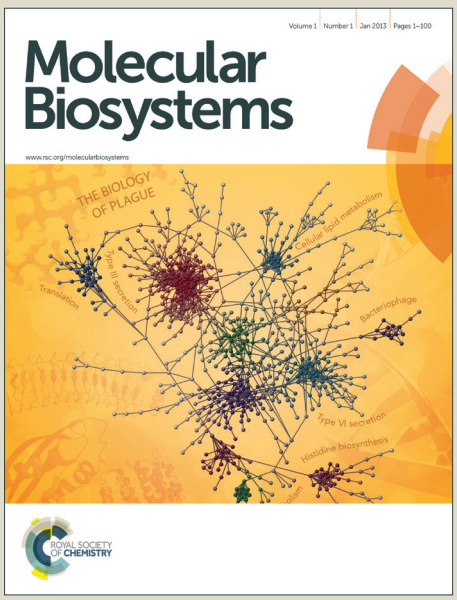
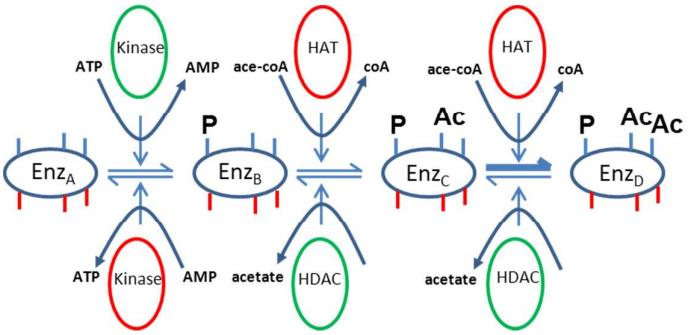
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems



Do proteins really exist? This paper explores the limitations to the definition in face of the epistemological challenge of rates of data deposition from bottom-up proteomics pertaining to post-translational modifications. We demonstrate that this information cannot *ever* be resolved and exists highly disconnected from biology. We propose that rather than dealing with impossibly complex multi-state species, the application of relational biology ideas, coupled with complexity reduction via a normative filter of top-down proteomics and analysed by mathematical and computational platforms for metabolism, will enable the description of "proteins" as metabolic pathways or networks, and place them in a conceptual continuum with metabolites.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

OPINION

Are proteins a redundant ontology? Epistemological limitations in the analysis of multistate species

Bernard M. Corfe,*[ab] and Caroline A. Evans[c]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

DOI: 10.1039/b000000x

Advances in proteomics have exponentially increased the numbers of post-translational modifications identified, the resulting volume of data is overwhelming both databases and empiricists. We review methodologies for chemical and functional PTM assignment. Using β–oxidation as a paradigm, we discuss epistemic limitations and conceptual approaches to resolving them combining relational biology, proteomics, and the erosion of "protein" and "metabolite" as distinct ontologies.

Post translational modification (PTM) of proteins, the addition of a chemical group, is a key regulatory mechanism in regulating protein function. To date over 400 PTM species have been demonstrated and catalogued [1]. In the analysis of PTM function, the key questions are: what type of PTM? how many different types are present? The functional state is modulated by PTM and itself subject to flux. As such the key questions *should also be*: do they co-occur on the same species? what is the stoichiometry and what is the functional consequence of the (combination of) PTM? Mass Spectrometry (MS) provides a useful tool for the analysis of PTMs since each PTM is associated with addition of a defined mass (e.g. lysine acetylation adds a mass of 42 Da, phosphorylation adds a mass of 80 Da). However an ever-increasing number of high-throughput depositions derive from top-down approaches and consequently there is significant loss of biologically important information. In considering PTM status and its relation to function, and the ever increasing assignments of PTM, we here review the current technical challenges and current methodologies to studying co-occurring combinations of PTMs on a single backbone and discuss a new philosophical construct to modelling the biochemical impact of the PTM status on protein pathways and networks.

Mass spectrometry based analysis of PTM

i) identification of PTM status

In general analytical strategies for PTM operate in two main modes, 1. PTM mapping of single proteins 2. Identification of PTM peptides/protein by global analysis of complex samples. There is currently no single method for PTM profiling, instead a range MS-based methods have been developed, targeted to specific PTM types e.g. phosphorylation, lysine acetylation.

Due to the plethora of methodologies, only the basic principles and concepts are outlined here with examples. PTMs are typically detected at low stoichiometry, specific enrichment steps have been developed which are required for analysis of individual proteins and for complex samples. For individual proteins, the enrichment is targeted with the protein purified by classical biochemical techniques. Such approaches include immunoprecipitation with protein-specific antibodies, subcellular fractionation followed, SDS-PAGE and band excision prior to analysis. Two dimensional gel electrophoresis can be of value as differentially modified proteins will migrate to distinct positions due to alteration of pI by the PTM.

In a PTM profiling analysis, a PTM is targeted rather than specific protein and enrichment protocols are directed to the distinct chemical features of the PTM type. Affinity capture methods such as use of metal affinity enrichment of phosphopeptides/phosphoprotein, the use PTM specific antibodies e.g. anti acetyl lysine, anti phosphotyrosine antibodies have proved of value. Glycosylated peptides can be enriched through lectin binding [2]. Complementary approaches include chromatographic enrichment of phosphorylated and glycosylated peptides by electrostatic repulsion hydrophilic interaction chromatography (ERLIC) [3].

ii) Site localization of PTM

The presence of a PTM and its site of location in a peptide can be determined by from information of the precursor mass (measured as *m/z*) and its product ion fragments. The commonly used Collision Induced Dissociation (CID) mode of fragmentation, results in fragmentation across the peptide backbone to yield N- and C- terminal containing fragments from which peptide sequence and PTM sites can, theoretically, be inferred from the MS/MS spectrum generated. For collisionally stable modifications eg Lysine acetylation, fragments containing the PTM site which have an additional mass due to presence of the PTM, enable site assignments. Labile modifications such as serine/threonine phosphorylation can be inferred by phosphorylation-specific neutral losses resulting from the β-elimination of phosphoric acid. Analysis of PTM is aided by specific fragmentation patterns generated by MS to yield diagnostic fragment ions e.g. 126.1 Da for Lysine acetylation, 216.1 Da for tyrosine phosphorylation in positive mode, -79 Da

for phosphopeptides in negative mode. The diagnostic fragments are utilized in targeted analysis, by methods designed to monitor for the presence of these ions. Examples include precursor ion scanning, multiple reaction monitoring, these methods are outlined in detail for phosphopeptides [4]

Some PTM are relatively simple to assign, including lysine acetylation, methylation since they generate unique diagnostic ions, whilst others such as glycosylation[5], sumoylation and ubiquitination are more complex. Peptide modifiers, including Ubiquitin and related modifications (Ubl) pose challenges since the modifiers themselves produce various mass produce fragments upon proteolytic digestion. This is problematic since proteolysis is a common step in proteomic workflows. Novel MS and database searching tools such as the approach basis consecutive addition to lysine (characteristic of Ub and Sumo) have been developed [6].

iii) Identifying PTM combinations in single species

PTM have typically been studied at the peptide level, in a so called 'bottom up' or 'shotgun' approach, where proteins are proteolytically cleaved to yield peptide fragments that are amenable to separation and analysis by LC-MS/MS. Identification of co-occurrence of PTM in the same protein is limited to individual peptides in these approaches. A key challenge in PTM research is to identifying the number of modified forms (PTM isoforms) of a *protein* and the full extent of its modification. The 'top down' approach has considerable strengths: the PTM is inferred from the experimentally determined mass value by adding masses of the modified residues to the non-modified mass of the protein. PTM modifications are inferred from accurate measurement of intact mass using high resolution mass spectrometry [7]. While technical advances have been achieved for the 'top down' in practice, it is typically applied for PTM analysis proteins <100 kDa. Application of top down has been demonstrated for higher masses [8].
Top down proteomic methodologies have been comprehensively reviewed recently [9, 10]. The power of mapping PTM (and ability to discriminate isoforms, which is completely lost in 'bottom up' approaches) on intact proteins has been shown but has required development of protein fractionation workflows which were previously inadequate relative to peptide separation capabilities. A novel 4 dimensional protein separation ahead of MS analysis facilitated a high throughput approach, not previously possible [11]. This represents a major step forward in the field with the study identifying an impressively high number of 3,000 protein species corresponding to 1,043 gene products from the human HeLa cancer cell line [11]

A hybrid approach, 'middle down' . (also termed Extended Range Proteomic Analysis)combines the strengths of both approaches whilst addressing their limitations. The use of outer membrane protease T (OmpT) is a major step forward, to generate fragments of average mass 6.3 kDa by cleavage between Lys-Lys, Lys-Arg, Arg-Lys and Arg-Arg) [12]. Previous work using restricted proteolysis, alternative proteases (Glu C, Asp N)or chemical cleavage by cyanogens bromide has met with some success[13] but can be limited due to generating fragment sizes similar to trypsin (800-2500 kDa) rather than larger fragments [10]. The method has been successfully applied to 20-100 kDa proteins from HeLa cell lysate, identifying 3,697 unique peptides from 1,038 unique proteins. Softwares such as PTMSearchPlus [14] and ProSight2.0 allow integration of top down and bottom up data to aid PTM isoform assignment.

Catalogues, repositories and their limitations

It is possible to routinely identify and catalogue thousands of -lysine acetylations [15, 16] and phosphorylation sites [17]. Data repositories of PTM sites are readily available online which are curated and updated. Examples include PhosphoSite www.phosphosite.org [18], PHOSIDA www.phosida.com [19], Phospho.ELM www.phospho.elm.eu.org [20]. Protein databases such as UniProt list published PTM sites and are also updated to provide a valuable resource (www.uniprot.org) [21].

It should be noted that there are many published large scale PTM-'discovery' focused MS studies, where the emphasis has been on generating high amounts of information from high complexity samples. It is imperative that this is coupled to quality control of PTM and site assignment, particularly in cases where follow up work is planned (e.g. site-directed mutagenesis, generation of specific antibodies) to ensure cost effectiveness by avoiding focus on false-positive or unconfident PTM assignments. In general application of false discovery rate (FDR) and score filters and in some cases, manual verification of spectra are potential strategies. The 'good, the bad and the ugly' aspects of PTM assignment and strategies for quality control mechanisms are outlined by Beck et al. [22].

The importance of accurate site assignment is highlighted by the demonstration that several phosphorylation sites within a relatively short peptide sequence, may represent different kinase substrates and or differential modulation [23]. Ambiguity in site localization is a common problem where more than one modifiable residue exists in the peptide. Alternative activation modes for fragmentation to ion trap collision-induced dissociation (CID) are of value in this scenario since they provide complementary information; these include Higher energy collision dissociation (HCD), Electron Transfer Dissociation (ETD) [24]. In terms of complexity, positional isomers can occur, which can be indistinguishable based on retention time analysis [25]. In principle this may relate to other PTM. Emerging novel techniques such as ion mobility separation add another dimension of separation for separation of peptides, which occurs in the gas phase of MS prior to fragmentation and is of particular value in separating resolution of co-eluting isobaric species such as phosphoisomers [26].

In order to determine the range of PTM that are known to exist biologically, there has been emphasis on retrospective mining of tandem MS data to search for PTM. Several bioinformatics based software solutions have been developed including Modi [27] Semop [28] PTMClust [28] amongst others which deal with 2 key issues: mass inaccuracy of the modification mass and site assignment uncertainty.

Functional consequences of PTM regulation

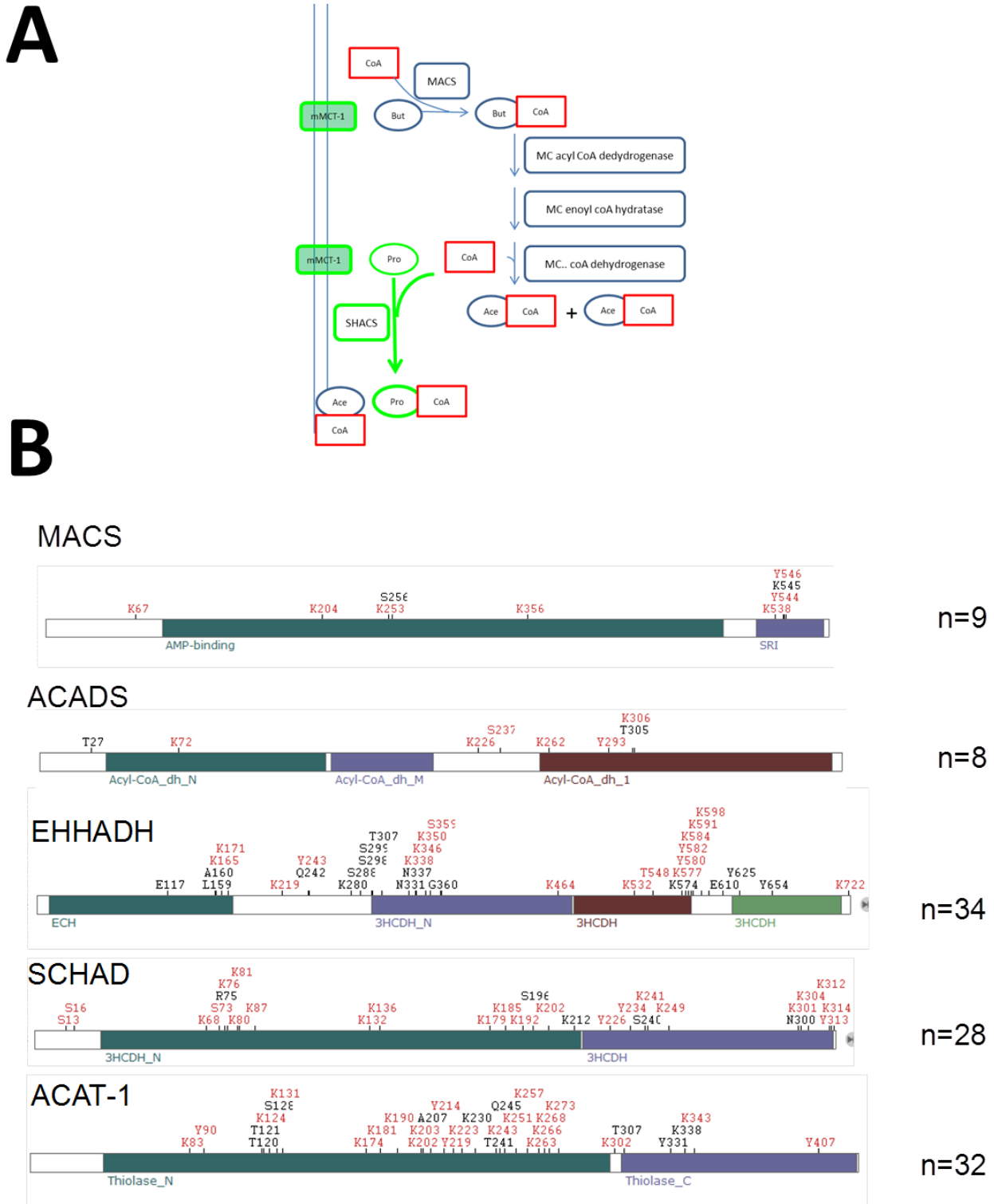


Fig. 1 Outline scheme of the β oxidation pathway. A. The five enzymes converting butyrate to Acetyl CoA B. Post-translation modifications, as downloaded from www.phosphosite.org, March 2012.

It has emerged that many PTMs are dependent on a precursor series of events, presence of other PTMs, and exist within the context of an event sequence, just like any biochemical pathway and demonstrated for histone H3 [29]. PTMs are subject to feedback with 'cross talk' with other PTM. Lysine acetylation and phosphorylation can co-occur within the same protein to regulate function as clearly demonstrated by use of a genome reduced

bacterium *Mycoplasma pneumoniae*, to facilitate demonstration of systematic perturbation of acetylation following deletion of (the only) two protein kinases and its unique protein phosphatase [30]. The study identified that cross talk between lysine acetylation and other PTM including phosphorylation, ubiquitination is emerging a key area in PTM research and there is recognition that determination of the type and extent of PTM provides valuable mechanistic information. For example. a bioinformatic study used *in silico* mutation, to demonstrate that lysine acetylation has potential to impact on phosphorylation, methylation and ubiquitination status [31]. Furthermore, a recent study by Minguez et al analysed 115,149 non-redundant PTMs of 13 distinct types, from 8 eukaryote types in the first large scale survey of conservation of "the global landscape of post-translational regulation" [1]. Key findings include the co-evolution of PTM pairs, supporting the presence of specific PTM clusters in provision 'regulatory centres', regions of protein sequence that contain multiple sites for PTM [32]. Phosphorylation, acetylation, ubiquitination and O-linked glycosylation were identified as central to controlling temporal events and processes that govern protein localization [1].The combinatorial nature of PTM in regulating distinct aspects of protein function has been demonstrated, with specific PTM combinations forming "histone code" and the analagous "tubulin codes" where different PTM combinations are associated with specific protein activity [33, 34].

The technical and epistemic challenge of the PTM data explosion (using β-oxidation pathway as a paradigm)

Lysine acetylation has emerged as a key post-translational modification involved in cell regulatory mechanisms, particularly of metabolism. For example, multiple enzymes in the key metabolic pathways of glycolysis, TCA, β-oxidation are acetyl proteins, as elegantly demonstrated by two key publications, published back-to-back in Science in 2010 [35, 36]. The first examined a selected enzyme for one step in each pathway, demonstrated that the target enzyme was acetylated, and further showed that acetylation had a regulatory effect, increasing or decreasing activity (K_{cat} value). In the case of one of the β-oxidation pathway proteins, enoyl–coenzyme A hydratase/3-hydroxyacyl–coenzyme A dehydrogenase (EHHADH), this was associated with an increased K_{cat} for the forward reaction of this bidirectional enzyme. Similar effects were found for other enzymes and the acetylations were often substrate driven, with glucose, long chain fatty acids (LCFA), and mixed amino acids affecting the acetylation status of metabolic enzymes to modulate metabolism in human cells by regulation of enzyme activity[1] . These behaviours underpin the hypothesis that cells use linear motifs or codes, to and from which PTMs can be written, read, erased to modulate protein activity [37] by providing so called logic gates for progression from one activity state to another [38]). These concepts are discussed and reviewed by Creixell and Linding [32] in light of the global PTM survey of Minguez et al.[1].

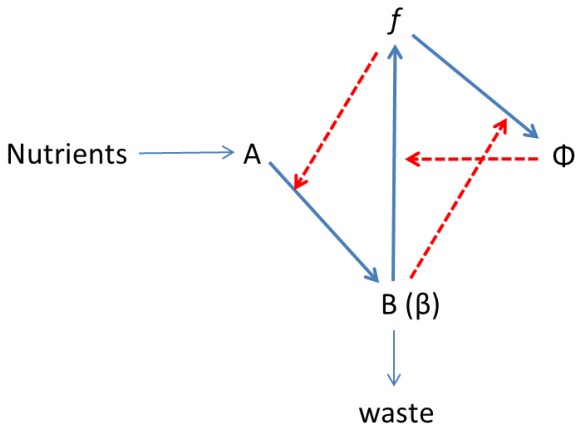


Fig. 2 Summary of the Rosen M,R system. A way of conceptualising and simplifying all biological processes . There are two fundamental processes – material causation and efficient conversion of a metabolite to a second metabolite ie conversion of a metabolite to another metabolite and catalysis respectively. Thus the set of metabolites A that enter the system are converted to metabolites B, a process catalysed by f, also metabolites that are produced materially from set B. In turn the production of set f from B is catalysed by set Φ, metabolic products of f. Finally the formation of Φ from f is catalysed by β. β is not B, but is a property of B. An engaging feature of the (M,R) is that every biochemical, including the macromolecules, is a metabolite - a product of metabolism and converted from input masses. This version of the M,R is described after Letelier et al.[39].

In terms of systems modelling, the β-oxidation pathway of butyrate (Fig. 1A), appears conceptually to be a simple system: 5 enzymes, 1 substrate. The reality is more complex since all 5 enzymes are subject to substantial additional post-translational modifications (Fig 1B) with a total of 111 PTM reported from empirical data (www.phosphosite.org, accessed March 2012). MACS, ACAD, SCHAD-1 and ACAT-1 enzymes carry 9, 8, 34, 28 and 32 PTM sites respectively, of which 52 are acetylated lysine [2]. If one makes a basic (and probably unsafe) assumption that each PTM is a binary possibility and is independent of other PTMs, this means that the number of states that the pathway can exist in is 2^{111}. The simplest way to compute this would be with a binary star topology of 111 nodes. However, the assumption that each PTM is a binary possibility and is independent of other PTMs is oversimplistic since there is no accounting for weighting of effects; some PTMs may have non-equal effects. Furthermore, single amino acid residues have the potential to modified by different PTM, for example lysine can be acetylated, propionylated, butyrylated, methylated, dimethylated, trimethylated, ubiquitinated, sumoylated – nine states at a minimum).

Such complexity of PTM leads to the question of how this diversity of PTM be addressed by systems modelling approaches? Gatherer [40] suggests that such problems rapidly exceed normal computational limits, and that as Bremmerman's limit is

[1] A key area of interest is to investigate the effects of substrates on pathways other than for their own utilisation: for example, does exposure of cells to LCFA reduce the activity of glycolytic and TCA enzymes by acetylation? Follow-up papers are sure to address this

[2] During production of this article we noted that the sequence of proteins converting butyrate to acetyl coA have 3, 4, 12, 16 and 17 acetylations respectively – as the series gets closer to the product it becomes more likely to become a co-substrate for KATs with its own product.

approached, there may be a limit to what is ever computationally tractable. The intractability of problems at this scale has been classified as part of the epistemiological anti-reductionist school [40]. The modelling community addressing kinetic data use what is effectively net parameter data for multistate species. PTM can be included as variable but this is not trivial to do. The Systems Biology Markup Language approach to PTMs/multistates is to consider all PTMs on a backbone. The concept of multistate protein species[3] seems pre-emptively reasonable, but doesn't address the epistemic limitation

Alternative approaches to the complexity of enzyme regulation by PTM

Consideration of the inherent difficulties of current approaches as discussed above, led us to question the utility of the current use of 'protein' as a definition or ontology? The PTM status of an enzyme is one aspect of a more complex situation where multiple effects of competing substrates should also be considered. As such, the term 'protein' encompasses not just the catalytic activityof the protein, but additional properties: as a substrate for enzymes mediating addition and removal of PTM which modulate activity and being multi-state when existing in unmodified and PTM modified forms. The property of the protein here links closely to top-level gene ontology classifications (e.g. cell component, biochemical process, molecular function) reflecting the location, interaction and enzymatic properties rendered by a particular combination of PTMs.

An alternative, radical view, which may be useful, is to consider 'protein' states: PTM modified and (potentially multi-PTM, and the unmodified state) as 'metabolites' instead. Consider this analogy: AMP, ADP and ATP each differ by a single phosphate residue, in terms of metabolism, ADP and ATP are considered distinct metabolites rather than phosphorylation products of adenosine. By the same rationale an enzyme, for example, MACS, and PTM modified forms; phospho-MACS, and diphospho-acetyl MACS should, perhaps, not be considered as states of the same protein, but as points along a biochemical pathway of MACS metabolism, just as progressive modifications to a mass feature. A scheme for considering the protein-as-metabolome, based on that of Rosen's Metabolism Repair systems (M, R) [41] to represent cell metabolism, in a relational biology context (Figure 2). In this approach the model is to be seen as not representative of any one reaction, but as a way of conceptualising and simplifying all biological processes. There are two fundamental processes – material causation and efficient conversion of a metabolite to a second metabolite. These processes represent conversion of a metabolite to another metabolite and catalysis respectively.

The scheme of Rosen introduces a second relational biology concept, that of *property*, whereby the separate forms of the protein (enzyme)-metabolome would have distinct properties, for example enzyme kinetics, affinity for binding partners, affinity for processing enzymes (Fig 3). This 'protein-as-metabolite' approach has particular value in moving away from the naïve

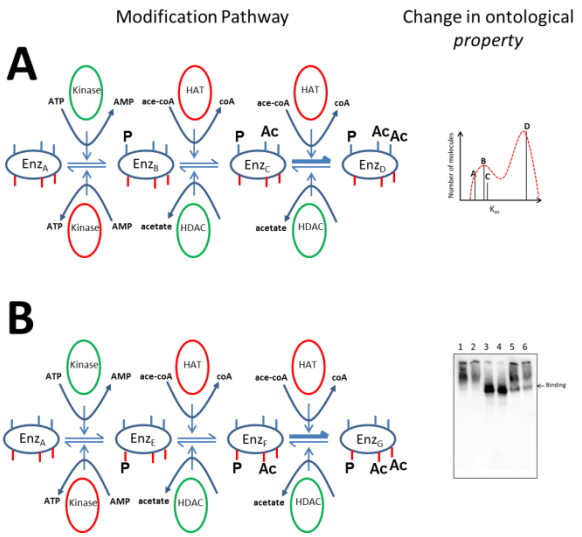


Fig 3 Conceptualization of the protein-as-metabolome A. a peptide backbone with enzymatic potential is the substrate firstly for a kinase, which increases affinity for a HAT, in turn leading to greater acylation at a second site. All molecules are metabolites in so far as they are in turn substrates for other activities. The consequence of successive modifications is arrival at the final Km for the Biochemical Function (right panel) although the observed Km will be a product of the cumulative activities of species and their relative abundance. B. In B a parallel series of modifications governs the transition to the molecular function (For example DNA binding).

At any one point the individual A series and B series will overlap in observed combinations in species within a cell. The linking of gene ontology (GO) terms to specific modification series and the separation of sub-combinations might be achieved through Markov modeling.

(binary) assumption, that all modifications are equally likely and that they act independently. Letelier et al. [42] demonstrate the utility of combining the concept of self-generation/autopoiesis in combination with Rosen's M,R in dealing with the 'circularity of metabolism, and the new epistemologies that they imply'. These approaches potentially offer relevant new modelling strategies.

We propose a novel approach for conceptualisation of the protein-as-metabolome as outlined in Fig. 3. The model proposed suggests that an enzyme backbone ENZ, can be subject to 6 PTMs, but that these may be divided into an A series and a B series. The A series have a hierarchy towards yielding and regulating a biochemical function (e.g. K_m); the B series have an independent hierarchy regulating molecular function (e.g. DNA binding). A tertiary series could likewise be imagined influencing cellular component. As such we suggest that the PTM profile of a given backbone is a combination of several series of linear patterns superimposed on each other. Any one species will be subject to a combination of modifications driving these independent pathways. In reality many proteins (for example the β-oxidative proteins shown in Fig 1) will have far more than six modifications. The top-down proteomics approaches described above are essential to progress biological understanding of which combinations of PTMs, or alternatively, which states along the metabolic pathway actually occur (bottom-up approaches cannot yield such insights on combinations). MS data from top-down can

[3] Oellerich et al 2010: http://sbml.org/images/8/8d/Multi_2010_November_29.pdf

therefore rapidly and effectively reduce the computational requirement, by excluding the majority of states which do not occur in reality. Nonetheless the observed states will be fewer that the actual states as some species may be refractory to MS analysis and others may be extremely transient (in the hypothetical construct in Fig 3, the last state in each path is produced at an accelerated rate, and by implication the preceding state may be transient).

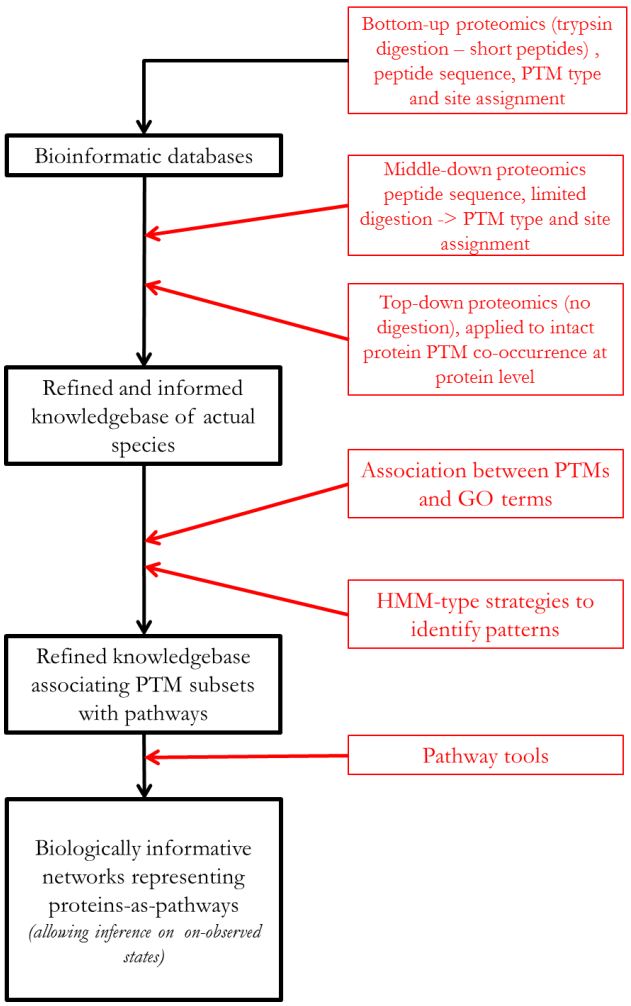


Figure 4 Schematic workflow for resolution of bottom-up and top-down proteomic data coupled with modelling and pathway tools to yield insights into the protein-as-pathway concept.

Having deployed top-down proteomics to delimit the likely complexity of numbers of states, two complementary approaches could be applied to identify pathways. We are hypothesizing that the actual combinations of PTMs on a backbone are the product of a series of superimposed pathways (as suggested in Fig 3b), and that there are therefore a series of patterns identifiable in the cumulative dataset. Identification and separation of such patterns may be tractable as a Hidden Markov (HMM) problem, which would provide the framework of flexibility required where data are incomplete and sequence uncertain. HMM approaches have been successfully applied elsewhere in molecular biology for pattern analysis [43] and in clinical analysis of pathways in disease progression[44]. For reasons described data will be incomplete, but partial information taking the form of an outline metabolic framework can be applied inferentially. Pfau et al have summarised a suite of approaches to modelling including integrative approaches to use metabolic and proteomic data to infer metabolic networks.[45] Approaches generally rely on existing metabolic schemes, methods described therein, including Minimal Cut Sets and Christian's Method for Network Expansion[46] could bridge the gap from data-driven incomplete networks, to predictive and computable networks. A hypothetical workflow is presented in Fig 4.

Perspective

The purpose of this article is to describe a combined philosophical and technical approach to address in a pragmatic and useful way the effect of combinations of PTMs on an amino-acid backbone. Rather than addressing and investigating the combinatorial explosion, we propose treating proteins much in the manner of studying a metabolic pathway: looking at a series of modifications that yields a biochemically important species (enzyme) with properties (subcellular location, catalytic activity, substrate affinity). Some PTM may have relatively low importance in function of the backbone (for example those present in the cellular pool, but only involved in directing the backbone to the right compartment).There is a key requirement for functional assignment of PTM significance, particularly given the recent demonstration that not all sites of PTM have a significant biological role [47]. The description of individual PTM isoforms at the protein level is a key factor in assigning mechanistic significance of PTM status.

Closely related species are of interest as they may have slightly shifted properties, altered directionality. It is, however, only by questioning the value of the concept of "a protein" that we can arrive at the "protein-as-metabolome" or "protein-as-pathway" concept that will enable us to progress the study of, ironically, proteins. The removal of one hierarchy (metabolites versus proteins) and replacement with a sequential hierarchy (the pathway) offers the chance reduce complexity around the explosion of data on PTMs being yielded at ever-accelerating rates by high-throughput approaches.

At the moment such approaches are, *in lieu* of good biochemistry can be questionable without functional validation, a bottle neck as the rate of identification far outstrips analysis of PTM variants *in vitro*. Beyond this what is needed is a connection between proteomic data (list of PTM sites, splice variants that represent the recently termed Proteoform[48] for further understanding mechanistic biology.

Notes and references

[a] *Molecular Gastroenterology Research Group, Academic Unit of Surgical Oncology, Department of Oncology, University of Sheffield. Tel: +44 (0) 114 271 3004; E-mail: b.m.corfe@sheffield.ac.uk*

[b] *InSigneo Institute for Insilico Medicine, University of Sheffield,*

[c] *Chemical Engineering at the Life Science Interface (ChELSI) Institute, University of Sheffeld*

[c] *Biological and Systems Engineering Group, ChELSI Institute, Department of Chemical and Biological Engineering, University of Sheffield S1 3JD; E-mail: caroline.evans@sheffield.ac.uk*

1. P. Minguez, L. Parca, F. Diella, D. R. Mende, R. Kumar, M. Helmer-Citterich, A.-C. Gavin, V. van Noort and P. Bork, *Mol. Syst. Biol.*, 2012, 8.
2. K. Kubota, Y. Sato, Y. Suzuki, N. Goto-Inoue, T. Toda, M. Suzuki, S.-i. Hisanaga, A. Suzuki and T. Endo, *Analytical Chemistry*, 2008, 80, 3693-3698.
3. P. Hao, T. Guo and S. K. Sze, *Plos One*, 2011, 6.
4. M. R. Larsen, M. B. Trelle, T. E. Thingholm and O. N. Jensen, *Biotechniques*, 2006, 40, 790-798.
5. S. M. Patrie, M. J. Roth and J. J. Kohler, *Methods in molecular biology (Clifton, N.J.)*, 2013, 951, 1-17.
6. G. J. Chicooree N, Connolly J, Tan C, Malliri A, Eyers CE and Smith DL. , *Rapid Communications in Mass Spectrometry*, 2012, in press.
7. Y. Ge, B. G. Lawhorn, M. ElNaggar, E. Strauss, J. H. Park, T. P. Begley and F. W. McLafferty, *Journal of the American Chemical Society*, 2002, 124, 672-678.
8. N. M. Karabacak, L. Li, A. Tiwari, L. J. Hayward, P. Hong, M. L. Easterling and J. N. Agar, *Molecular & Cellular Proteomics*, 2009, 8, 846-856.
9. F. Lanucara and C. E. Eyers, *Mass Spectrometry Reviews*, 2013, 32, 27-42.
10. H. Zhou, Z. Ning, A. E. Starr, M. Abu-Farha and D. Figeys, *Analytical Chemistry*, 2012, 84, 720-734.
11. J. C. Tran, L. Zamdborg, D. R. Ahlf, J. E. Lee, A. D. Catherman, K. R. Durbin, J. D. Tipton, A. Vellaichamy, J. F. Kellie, M. Li, C. Wu, S. M. M. Sweet, B. P. Early, N. Siuti, R. D. LeDuc, P. D. Compton, P. M. Thomas and N. L. Kelleher, *Nature*, 2011, 480, 254-U141.
12. C. Wu, J. C. Tran, L. Zamdborg, K. R. Durbin, M. Li, D. R. Ahlf, B. P. Early, P. M. Thomas, J. V. Sweedler and N. L. Kelleher, *Nature Methods*, 2012, 9, 822-+.
13. A. Moradian, A. Kalli, M. J. Sweredoski and S. Hess, *Proteomics*, 2013.
14. V. Kertesz, H. M. Connelly, B. K. Erickson and R. L. Hettich, *Analytical Chemistry*, 2009, 81, 8387-8395.
15. C. Choudhary, C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen and M. Mann, *Science*, 2009, 325, 834-840.
16. N. Mischerikow and A. J. R. Heck, *Proteomics*, 2011, 11, 571-589.
17. K. E, *Seminars in Cell & Developmental Biology*, 2012, in press.
18. P. V. Hornbeck, I. Chabra, J. M. Kornhauser, E. Skrzypek and B. Zhang, *Proteomics*, 2004, 4, 1551-1561.
19. F. Gnad, J. Gunawardena and M. Mann, *Nucleic Acids Research*, 2011, 39, D253-D260.
20. H. Dinkel, C. Chica, A. Via, C. M. Gould, L. J. Jensen, T. J. Gibson and F. Diella, *Nucleic Acids Research*, 2011, 39, D261-D267.
21. G. A. Khoury, R. C. Baliban and C. A. Floudas, *Scientific Reports*, 2011, 1.
22. F. Beck, U. Lewandrowski, M. Wiltfang, I. Feldmann, J. Geiger, A. Sickmann and R. P. Zahedi, *Proteomics*, 2011, 11, 1099-1109.
23. M. Courcelles, G. Bridon, S. Lemieux and P. Thibault, *Journal of Proteome Research*, 2012, 11, 3753-3765.
24. C. K. Frese, A. F. M. Altelaar, M. L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A. J. R. Heck and S. Mohammed, *Journal of Proteome Research*, 2011, 10, 2377-2388.
25. H. Marx, S. Lemeer, J. E. Schliep, L. Matheron, S. Mohammed, J. Cox, M. Mann, A. J. R. Heck and B. Kuster, *Nature Biotechnology*, 2013, 31, 557-+.
26. G. Bridon, E. Bonneil, T. Muratore-Schroeder, O. Caron-Lizotte and P. Thibault, *Journal of Proteome Research*, 2012, 11, 927-940.
27. S. Kim, S. Na, J. W. Sim, H. Park, J. Jeong, H. Kim, Y. Seo, J. Seo, K.-J. Lee and E. Paek, *Nucleic Acids Research*, 2006, 34, W258-W263.
28. C. Baumgartner, T. Rejtar, M. Kullolli, L. M. Akella and B. L. Karger, *Journal of Proteome Research*, 2008, 7, 4199-4208.
29. S. Liokatis, A. Stuetzer, S. J. Elsaesser, F.-X. Theillet, R. Klingberg, B. van Rossum, D. Schwarzer, C. D. Allis, W. Fischle and P. Selenko, *Nature Structural & Molecular Biology*, 2012, 19, 819-823.
30. V. van Noort, J. Seebacher, S. Bader, S. Mohammed, I. Vonkova, M. J. Betts, S. Kuhner, R. Kumar, T. Maier, M. O'Flaherty, V. Rybin, A. Schmeisky, E. Yus, J. Stulke, L. Serrano, R. B. Russell, A. J. R. Heck, P. Bork and A. C. Gavin, *Mol. Syst. Biol.*, 2012, 8.
31. Z. Lu, Z. Cheng, Y. Zhao and S. L. Volchenboum, *Plos One*, 2011, 6.
32. P. Creixell and R. Linding, *Mol. Syst. Biol.*, 2012, 8.
33. M. S. Cosgrove and C. Wolberger, *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire*, 2005, 83, 468-476.
34. J. Kilner, B. M. Corfe and S. J. Wilkinson, *Molecular Biosystems*, 2011, 7, 975-983.
35. S. M. Zhao, W. Xu, W. Q. Jiang, W. Yu, Y. Lin, T. F. Zhang, J. Yao, L. Zhou, Y. X. Zeng, H. Li, Y. X. Li, J. Shi, W. L. An, S. M. Hancock, F. C. He, L. X. Qin, J. Chin, P. Y. Yang, X. Chen, Q. Y. Lei, Y. Xiong and K. L. Guan, *Science*, 2010, 327, 1000-1004.
36. Q. J. Wang, Y. K. Zhang, C. Yang, H. Xiong, Y. Lin, J. Yao, H. Li, L. Xie, W. Zhao, Y. F. Yao, Z. B. Ning, R. Zeng, Y. Xiong, K. L. Guan, S. M. Zhao and G. P. Zhao, *Science*, 2010, 327, 1004-1007.
37. W. A. Lim and T. Pawson, *Cell*, 2010, 142, 661-667.
38. W. A. Lim, *Current Opinion in Structural Biology*, 2002, 12, 61-68.
39. M. L. Cardenas, J. C. Letelier, C. Gutierrez, A. Cornish-Bowden and J. Soto-Andrade, *Journal of Theoretical Biology*, 2010, 263, 79-92.
40. D. Gatherer, *Bmc Systems Biology*, 2010, 4.
41. R. Rosen, *SOME RELATIONAL CELL MODELS THE METABOLISM REPAIR SYSTEMS*, 1972.
42. J. C. Letelier, G. Marin and J. Mpodozis, *Journal of Theoretical Biology*, 2003, 222, 261-272.
43. J. Wu and J. Xie, in *Statistical Methods in Molecular Biology*, eds. H. Bang, X. K. Zhou, M. Mazumdar and H. L. VanEpps, 2010, vol. 620, pp. 405-416.
44. A. C. Titman and L. D. Sharples, *Statistical Methods in Medical Research*, 2010, 19, 621-651.

45. T. Pfau, N. Christian and O. Ebenhoeh, *Briefings in Functional Genomics*, 2011, 10, 266-279.
46. N. Christian, P. May, S. Kempa, T. Handorf and O. Ebenhoeh, *Molecular Biosystems*, 2009, 5, 1889-1903.
47. P. Beltrao, V. Albanese, L. R. Kenner, D. L. Swaney, A. Burlingame, J. Villen, W. A. Lim, J. S. Fraser, J. Frydman and N. J. Krogan, *Cell*, 2012, 150, 413-425.
48. L. M. Smith, N. L. Kelleher and P. Consortium Top Down, *Nature Methods*, 2013, 10, 186-187.