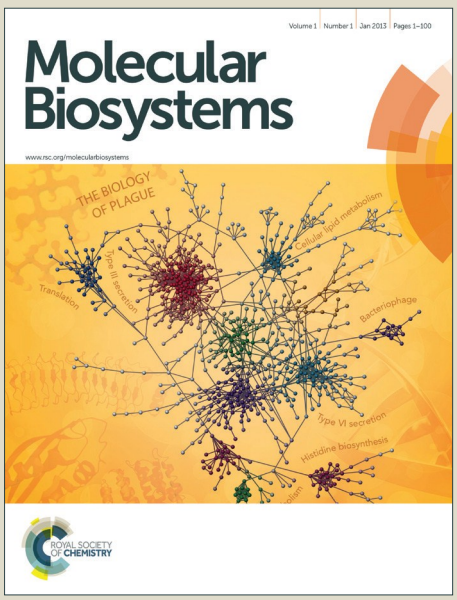
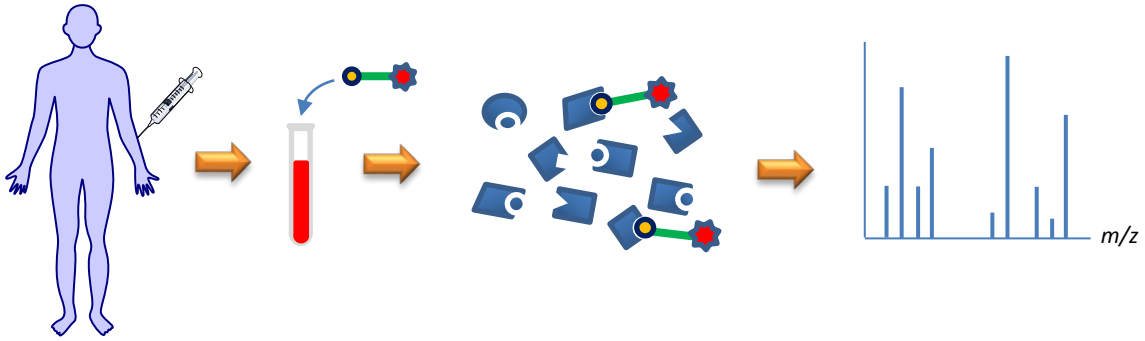
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems



Journal Name

Opinion

Understanding the Chemically-Reactive Proteome

Lyn H. Jones[a]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

The reactivity of amino acid residues in proteins is context-dependent and difficult to predict. Chemical biology can be used to understand the chemical modifications of proteins to help elucidate the nature of the reactive proteome. The resulting insights can be applied to pharmacoproteomics, target identification and molecular pathology.

Introduction

Protein labelling has found significant utility in chemical biology. For example, activity-based protein profiling (ABPP) relies on the use of a tagged chemical probe with a reactive warhead that captures mechanistically-related protein targets in a cell or tissue. The probe reacts with a key catalytic residue in the protein binding site, so enabling an assessment of protein activity, and competition of labelling with a drug of interest allows target occupancy and selectivity to be determined, often through the use of mass spectrometry (MS) proteomics.[1] ABPP nicely illustrates the power of synthetic chemical biology and better understanding of amino acid reactivity in protein binding sites would significantly impact chemoproteomic technology development.

Another important example of protein labelling is that of post-translational modification (PTM). This is often an enzymatic process, but non-enzyme-mediated chemical reactions also occur. Aberrant redox chemistry in inflammatory disease for instance results in a plethora of amino acid modifications that can be mapped using suitably functionalised chemical probes. Chemoproteomics may therefore provide a means to characterise disease at the molecular level, and create new opportunities for precision medicine and proactive therapeutic targeting.

Here I highlight some current efforts being used to understand the reactive proteome and show how this information can be applied to advance drug discovery.

Molecular Pharmacology

Serine reactive probes were the basis of much of the initial efforts in ABPP. In particular, the enhanced reactivity of an activated serine in a catalytic dyad or triad of the serine hydrolase (SH) family has been targeted selectively using a number of fluorophosphates,[2] diphenyl phosphonates,[3] ureas and carbamates.[4] Importantly, these efforts have led to the development of selective SH inhibitors with therapeutic potential.

We recently reported a cheminformatic analysis of the cysteine residues in and around the kinase ATP-site (Figure 1).[5] Each Cys residue in the 'kinase cysteinome' represents a potential opportunity for irreversible kinase inhibitor design and highlights kinases with the same reactive Cys that should be considered for off-target reactivity and potential toxicities. Each Cys position-cluster also represents opportunities for the development of 'semi-promiscuous' chemical probes, which can be used to measure intracellular target engagement and selectivity, thus better defining molecular pharmacology.[1, 6] Additionally, similar probes can be used to canvass kinase clusters as potential new targets from phenotypic screens.[7] A related approach recently used a clickable BTK covalent inhibitor to assess broad protein reactivity in an unbiased manner beyond the obvious kinase cysteinome cluster (Hinge 6 in Figure 1b).[8]

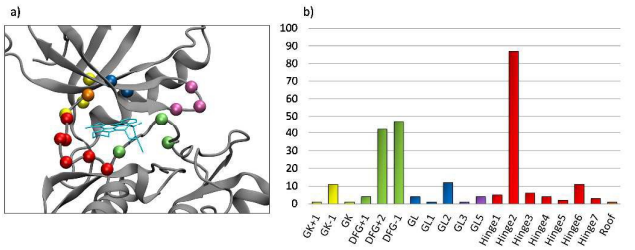


Figure 1. The kinase cysteinome. a) Crystal structure of the JAK3 kinase domain in complex with a staurosporine analogue (PDB 1YVJ)[9] used to calculate the positions of Cys residues in the ATP-site that are available to react with covalent binders; b) number of kinases corresponding to Cys positions in the kinase domain (GK – gatekeeper, GL – glycine loop).[5]

An important ABPP study in this area was reported by the Cravatt group that used a clickable iodoacetamide derivative to globally characterise and quantify reactive cysteine residues in various proteomes.[10] This work serves as an excellent

[a] Worldwide Medicinal Chemistry, Pfizer, 610 Main Street, Cambridge MA, 02139, USA.

resource to further understand and interrogate the broader reactive cysteine proteome.

The privileged nature of sulfonyl fluoride reactivity with context-specific Tyr, Lys, Ser, Thr, His and Cys residues in protein binding sites has enabled reactive proteome mapping experiments.[11] In a mass spectrometry (MS) proteomics study, the ATP-affinity reagent 5'-fluorosulfonylbenzoyl-5'-adenosine (FSBA, Figure 2) was shown to react with Tyr and Lys residues in a variety of proteins, including non-kinases.[12] By analysing the labelled protein residues and searching the Protein Data Bank for the presence of crystal structures for those proteins, it was possible to understand the microenvironment of the reactive Tyr and Lys residues (Figure 2). Most reactive tyrosines are proximal to basic amino acid residues, and reactive lysines are proximal to acidic residues, presumably facilitating deprotonation of the phenol and ammonium functionalities respectively. This knowledge can be used to help design irreversible inhibitors and chemical probes that target specific tyrosine and lysine residues.[13]

The current landscape of electrophiles known to target various nucleophilic amino acids has been explored somewhat.[14, 15] However, more work needs to be done to understand the microenvironment of each reactive residue to then efficiently incorporate these learnings into prospective molecular design of activity-based probes and selective covalent pharmacological agents. This could include the use of computational techniques to predict the pKa of reactive residues in proteins.[16]

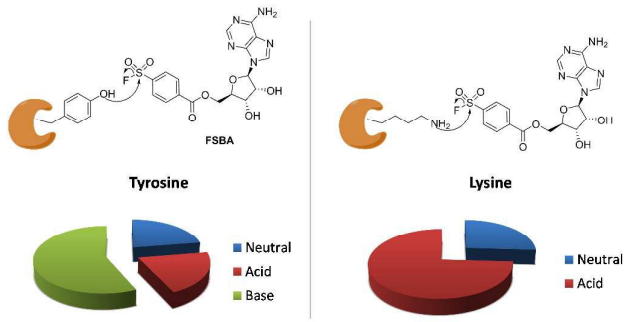


Figure 2. Proportion of acidic, basic or neutral amino acid residues that are nearest to FSBA-reactive tyrosine or lysine based on proteomic data from Hanoulle et al[12] and correlated to structures in the PDB. Most reactive tyrosine residues are proximal to a basic amino acid side chain (Lys, Arg, His),[13] whilst most lysines are proximal to acidic residues (Asp, Glu).

Molecular Pathology

Chemical protein modifications also play a wide role in the causation and modulation of disease-relevant processes. Disease classification according to symptoms and gross pathology is becoming an outdated concept, and by understanding its molecular aetiology, a more sophisticated definition of disease could be realised, that in turn could lead to more successful therapeutic approaches. A systematic screen for covalent modification events that are linked to disease conditions could therefore be highly rewarding in terms of developing molecular diagnostics and finding novel points of therapeutic intervention by modulating protein activity.

4-Hydroxy-2-nonenal (4-HNE, Figure 3a) is an unsaturated aldehyde produced during lipid peroxidation in cells, and is found in higher amounts during oxidative stress. It is believed to play an important role in the molecular pathogenesis of disease due to direct protein labelling, although broader profiling of 4-HNE adducts would further understanding of aberrant signalling events and adaptive responses. To this aim, a quantitative chemoproteomic approach was employed to determine protein alkylation in cells that used an alkyne-tagged derivative (aHNE, Figure 3a) to enable subsequent biotinylation using click chemistry, protein isolation and identification using high resolution LC-MS/MS.[17] 386 cysteine and 12 histidine sites were modified by aHNE in live cells, which could serve as a useful resource for further analysis in disease (oxidative stress) states. Interestingly, a characteristic sequence motif (CxxK) was identified for S-alkylation by aHNE likely due to the positively charged lysine lowering the pKa of the proximal Cys thus enhancing reactivity with electrophiles.

Endogenous electrophiles may also provide protection from disease. Nitro-oleic acid (10-nitrooctadec-9-enoic acid, NO_2-OA, Figure 3a) is a highly reactive, endogenous lipid electrophile known to target several cysteine sensors, including Keap1 and TrpA1. A recent study showed that protection from hypertension in mice is mediated by NO_2-OA inhibition of soluble epoxide hydrolase (through reaction with Cys521 proximal to its catalytic centre) suggesting a molecular mechanism behind the antihypertensive benefit of the Mediterranean diet.[18] A clickable version of NO_2-OA (such as NO_2-OA-yne, Figure 3a) and other electrophilic lipids could be used to canvass the broader reactive proteome in an unbiased manner (similar to aHNE) to assess other potential benefits or deleterious effects of protein labelling. Methods targeted to analysis of the membrane proteome in particular would be highly significant since the concentration of certain reactive lipids will be highest at the outer membrane, resulting in considerably higher rates of reaction (due to the 'molecular lens effect').[19] Lipid complexity created by the microbiota, and the subsequent electrophilic chemistries of these lipids could also be investigated using related chemical biology techniques.

Protein tyrosine nitration (PTN) and 3-nitrotyrosine (3-NT, Figure 3b) itself are hallmarks of many inflammation-associated diseases that form under conditions of nitrative stress.[20] PTN often leads to loss-of-function, such as the ablation of MnSOD function following Y34 nitration.[21] Gain-of-function has also been reported, as for the increase in peroxidase activity of peroxiredoxin 2 following Y193 nitration.[22] PTN can also lead to the generation of neoepitopes that play a role in autoimmune diseases. Chemoproteomic methods have been developed to assess the PTNs, including the use of chemical reduction of the nitro group followed by tagging of the resulting aniline with enrichment handles. These approaches provide opportunities to create PTN signatures that could differentiate related inflammatory diseases at a molecular level.[23] Tyrosine in particular can undergo a variety

of chemical PTMs in disease states, including chlorination, bromination and diverse oxidations (e.g. hydroxylations and crosslinking with other amino acid residues).[23] New global profiling methods are required to assess these transformations, particularly in disease. Cancer-associated neoepitopes resulting from these chemical transformations provide opportunities for biomarker generation and even personalised vaccines.

Enzymatic and chemical PTM crosstalk can also play a role in signalling processes, although more work is required to understand this complexity and pathophysiological relevance. A recent study reported 183 site-specific tyrosine nitration-phosphorylation crosstalks,[24] and specific examples have been detailed, such as the co-occurrence of both PTMs on tyrosine residues of α-synuclein.[25]

An area that has received considerable attention in this space is that of cysteine oxidation, particularly to the corresponding sulfenyl (SOH), nitrosyl (SNO) or glutathionyl (SSG) derivatives (Figure 3b). A chemoproteomic approach recently identified the following PTMs in mouse liver simultaneously: 686 SNO sites in 438 proteins; 883 SSG sites in 580 proteins; 442 SOH sites in 392 proteins.[26] This impressive analysis provides a useful reference proteome for comparison to disease states characterised by high oxidative and nitrative stress.

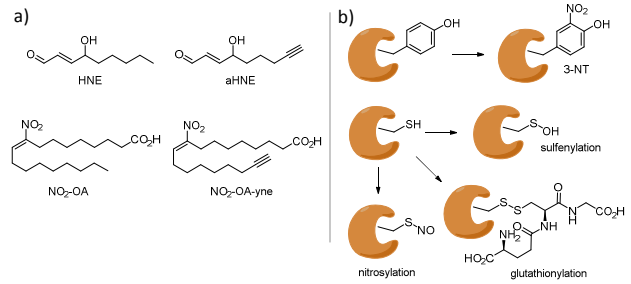


Figure 3. a) Electrophilic lipids 4-hydroxy-2-nonenal (HNE), nitro-oleic acid (NO_2-OA) and their alkynated derivatives. b) Some chemical post-translational modifications of tyrosine and cysteine residues amenable to chemoproteomic profiling.

Conclusions

A deeper knowledge of molecular pharmacology will hopefully translate to greater success in the development of innovative and transformational medicines.[27] Fundamental research into protein labelling chemistry (synthetic and natural) has the potential to advance drug discovery in a number of ways, including the elaboration of the molecular modes of action, and selectivity, of small molecule drugs and the elucidation of the molecular drivers of pathology. The development of chemoproteomic technologies such as ABPP in the clinical setting over the coming years holds great promise which will be facilitated by collaborations between clinical chemists and chemical biologists.

Notes and references

1. M. E. Bunnage, E. L. Chekler and L. H. Jones, *Nat. Chem. Biol.*, 2013, **9**, 195-199.
2. G. M. Simon and B. F. Cravatt, *J. Biol. Chem.*, 2010, **285**, 11051-11055.
3. S. Mahrus and C. S. Craik, *Chem. Biol.*, 2005, **12**, 567-577.
4. J. W. Chang, A. B. Cognetta III, M. J. Niphakis and B. F. Cravatt, *ACS Chem. Biol.*, 2013, **8**, 1590-1599.
5. Q. Liu, Y. Sabnis, Z. Zhao, T. Zhang, S. J. Buhrlage, L. H. Jones and N. S. Gray, *Chem. Biol.*, 2013, **20**, 146-159.
6. L. H. Jones, *Future Med. Chem.*, 2015, **7**, 2131-2141.
7. M. Nishino, J. W. Choy, N. N. Gushwa, J. A. Oses-Prieto, K. Koupparis, A. L. Burlingame, A. R. Renslo, J. H. McKerrow and J. Taunton, *eLife*, 2013, **2**, e00712.
8. B. R. Lanning, L. R. Whitby, M. M. Dix, J. Douhan, A. M. Gilbert, E. C. Hett, T. O. Johnson, C. Joslyn, J. C. Kath, S. Niessen, L. R. Roberts, M. E. Schnute, C. Wang, J. J. Hulce, B. Wei, L. O. Whiteley, M. M. Hayward and B. F. Cravatt, *Nat. Chem. Biol.*, 2014, **10**, 760-767.
9. T. J. Boggon, Y. Li, P. W. Manley and M. J. Eck, *Blood*, 2005, **106**, 996-1002.
10. E. Weerapana, C. Wang, G. M. Simon, F. Richter, S. Khare, M. B. Dillon, D. A. Bachovchin, K. Mowen, D. Baker and B. F. Cravatt, *Nature*, 2010, **468**, 790-795.
11. A. Narayanan and L. H. Jones, *Chem. Sci.*, 2015, **6**, 2650-2659.
12. X. Hanoulle, J. van Damme, A. Staes, L. Martens, M. Goethals, J. Vandekerckhove and K. Gevaert, *J. Proteome Res.*, 2006, **5**, 3438-3445.
13. E. C. Hett, H. Xu, K. F. Geoghegan, A. Gopalsamy, R. E. Kyne Jr., C. A. Menard, A. Narayanan, M. D. Parikh, S. Liu, L. Roberts, R. P. Robinson, M. A. Tones and L. H. Jones, *ACS Chem. Biol.*, 2015, **10**, 1094-1098.
14. E. Weerapana, G. M. Simon and B. F. Cravatt, *Nat. Chem. Biol.*, 2008, **4**, 405-407.
15. D. A. Shannon and E. Weerapana, *Curr. Opin. Chem. Biol.*, 2015, **24**, 18-26.
16. I. Uqur, A. Marion, S. Parant, J. H. Jensen and G. Monard, *J. Chem. Inf. Model*, 2014, **54**, 2200-2213.
17. J. Yang, K. A. Tallman, N. A. Porter and D. C. Liebler, *Anal. Chem.*, 2015, **87**, 2535-2541.
18. R. L. Charles, O. Rudyk, O. Prysyazhna, A. Kamynina, J. Yang, C. Morisseau, B. D. Hammock, B. A. Freeman and P. Eaton, *Proc. Natl. Acad. Sci. USA*, 2014, **111**, 8167-8172.
19. B. A. Freeman, P. R. S. Baker, F. J. Schopfer, S. R. Woodcock, A. Napolitano and M. d'Ischia, *J. Biol. Chem.*, 2008, **283**, 15515-15519.
20. L. H. Jones, *Chem. Biol.*, 2012, **19**, 1086-1092.
21. H. Neumann, J. L. Hazen, J. Weinstein, R. A. Mehl and J. W. Chin, *J. Am. Chem. Soc.*, 2008, **130**, 4028-4033.
22. L. M. Randall, B. Manta, M. Hugo, M. Gil, C. Batthyany, M. Trujillo, L. B. Poole and A. Denicola, *J. Biol. Chem.*, 2014, **289**, 15536-15543.
23. L. H. Jones, A. Narayanan and E. C. Hett, *Mol. BioSyst.*, 2014, **10**, 952-969.
24. Z. Pan, Z. Liu, H. Cheng, Y. Wang, T. Gao, S. Ullah, J. Ren and Y. Xue, *Sci. Rep.*, 2014, **4**, 7331.
25. L. Stefanis, *Cold Spring Harb Perspect Med*, 2012, **4**, a009399.

26. N. S. Gould, P. Evans, P. Martinez-Acedo, S. M. Marino, V. N. Gladyshev, K. S. Carroll and H. Ischiropoulos, *Chem. Biol.*, 2015, **22**, 965-975.
27. M. E. Bunnage, A. M. Gilbert, L. H. Jones and E. C. Hett, *Nat. Chem. Biol.*, 2015, **11**, 368-372.