Measurement Modalities and Literature Review

Reporters

This section enumerates the most likely reporters of amino acid identities, although it does not claim to be complete.

Endogenous properties

The simplest possible reporter of an amino acid residue is the acid itself. Amino acid side chains have various properties which permit them to perform the myriad functions requried by biological systems, and these properties can be measured. These properties include but are not limited to mass, rigidity, positive or negative charge, acidity or basicity, polarity, hydrophobicity, and endogenous fluorescence. These endogenous properties can, for example, provide contrast in a nanopore electrical measurement, provide a mass/charge fingerprint in a mass spectrometer, or provide an excitation signal in a fluorescence measurement. However, these signals are generally quite weak, and constraints on instrument sensitivity typically (although not always) means that they are only useful in bulk measurements.

Exogenous labels

Here, a label is defined as any agent that modifies the protein of interest in a residue- or sequence-specific manner in order to simplify downstream detection or analysis. The label could, for example, be a synthetic fluorophore to enhance brightness in a optical micrograph, a heavy metal to enhance contrast in an electron micrograph, a conductive or resistive element to enhance contrast in an electrical measurement, and so on. It is not practical to enumerate the particulars of the many possible labels in this report; rather, we will consider a label any agent that improves the sensitivity of the measurement modalities reviewed in Section **??**.

Labeling Methods

Chemical reporters on side chains

Here, chemical diversity in amino acid side chains is exploited in order to covalently bond, in a residue-specific manner, a reporter that is more amenable to downstream analysis than the properties of endogenous amino acid. An ideal system of chemical label reporters would involve one reporter per amino acid; however, many amino acids are non-reactive, and so only the reactive amino acids should be considered as targets for chemical conjugation. Moreover, many (most) chemical reactions on amino acid side chains are not orthogonal: one reaction may label multiple species. It is therefore a challenge in chemical biology to find **orthogonal, resuide specific** reactions on amino acid side chains.

The best characterized chemical label conjugation reactions are:

1. The NHS-ester-activated amide formation reaction, which specifically targets the free primary amine side chain on lysine residues in mildly basic conditions.

2. The maleimide-activated thioether formation reaction, which specifically targets the free sulfhydryl side chain on cysteine residues in mildly acidic conditions.

3. The carbodiimide-activated amide formation reaction, which specifically targets the free carboxylic acid side chain on glutamic acid residues in mildly acidic conditions.

These three reactions can be carried out sequentially and will feature prominently in subsequent discussion. However, other candidate reactions for chemical label conjugation, including reactions on reactive residues such as tyrosine, arginine, histidine, and aspartic acid, have received a great deal of attention and can also be considered should additional diversity prove a desirable objective.

Enzymatic reporters

Here, enzymatic specificity is exploited to covalently bond, in a residue-specific manner, a reporter group. This occurs frequently in biological systems in the form of post-translational modifications, where a vast range of chemical groups are appended to amino acid side chains; phosphorylation, acylation, and glycosylation are among the most frequent examples[1]. However, the use of post-translational modifications to report amino acid sequence is not commonplace, and most efforts to develop methods for single-molecule detection of post-translational modifications have focused on measurement of endogenous modifications rather than using them as a component in a protein sequencer, such as in the case of a nanopore-based detector for phosphorylation[2].

Affinity reagents

Here, non-covalent binding interactions are exploited in order to attach a reporter to a sequence of one or more amino acids. This is traditionally embodied in the form of an antibody[3], but aptamers[4], nanobodies[5], and other emerging classes of affinity reagent[6] can also be considered; moreover, single-molecule counting methods are also under development[7]. However, specificity and reproducibility are often major difficulties with this type of measurement, and the typical size of an epitope - many sequential residues - essentially demands one reporter per type of protein [8, 3]. One promising method to circumvent this requirement is to use a affinity reagent specific to the N-terminus of a particular amino acid, and combine multiple N-terminal measurements with sample degradation[9]. However, research in this area is still in its infancy.

References

[1] Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific reports* **1** (2011).

[2] Rosen, C. B., Rodriguez-Larrea, D. & Bayley, H. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nature biotechnology* **32**, 179–181 (2014).

[3] Taussig, M. J. *et al.* Proteomebinders: planning a european resource of affinity reagents for analysis of the human proteome. *Nature Methods* **4**, 13–17 (2007).

[4] Ellington, A. D. & Szostak, J. W. In vitro selection of rna molecules that bind specific ligands. *nature* **346**, 818–822 (1990).

[5] Rothbauer, U. *et al.* Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nature methods* **3**, 887–889 (2006).

[6] Skerra, A. Alternative non-antibody scaffolds for molecular recognition. *Current opinion in biotechnology* **18**, 295–304 (2007).

[7] Tessler, L. A., Reifenberger, J. G. & Mitra, R. D. Protein quantification in complex mixtures by solid phase single-molecule counting. *Analytical chemistry* **81**, 7141–7148 (2009).

[8] Baker, M. Blame it on the antibodies. *Nature* **521**, 274–6 (2015).

[9] Havranek, J. J. & Borgo, B. Molecules and methods for iterative polypeptide analysis and processing (2014). US Patent App. 14/211,448.