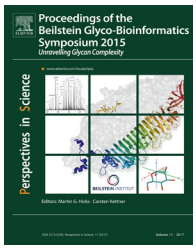
Perspectives in Science (2017) 11, 31—39

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/pisc



Bioinformatics applications to aid high-throughput glycan profiling☆

Ian Walsh, Roisin O'Flaherty, Pauline M. Rudd*

NIBRT GlycoScience Group, NIBRT—The National Institute for Bioprocessing, Research and Training, Foster's Avenue, Mount Merrion, Blackrock, Co. Dublin, Ireland

Received 28 September 2015; accepted 4 January 2016
Available online 25 November 2016

KEYWORDS
Glycomics;
Bioinformatics;
High-throughput;
Databases;
Software tools;
Biomarker

Summary High-throughput methods to identify and quantify glycans in a given sample are rare. We have optimised a robotic platform for analysing biopharmaceuticals at each stage of the manufacturing process. In addition, it can be applied to basic research. The plate format makes it convenient for large sample sets; it is relatively cheap, robust and quantitative. However, the large datasets churned out by this platform require significant time to interpret. Consequently, informatics tool are required to help with this annotation. This article briefly describes our robotic platform and concentrates on a set of software tools for the interpretation of quantitative glycoprofiling data.

Introduction

All cell surfaces are covered with glycoconjugates and more than fifty percent of human proteins are linked to glycan moieties. Increasing evidence indicates that carbohydrates have critical roles in major biological events such as immunological recognition, metastasis, cell signalling and cell differentiation to name just a few (Ohtsubo and Marth, 2006; Calarese et al., 2003). In addition, alterations in glycosylation are common in physiological and pathological processes enabling the fine tuning of biological pathways. Tailoring glycosylation during the manufacture of biotherapeutics improves the safety and efficacy of drugs such as monoclonal antibodies and erythropoietin. The analysis of glycans is challenging, evident in the fact that glycomics is significantly lagging behind genomics and proteomics. A possible reason for the lag is the rarity of high-throughput analytical methods. One successful attempt to overcome this is the development of a robotic platform to release and label glycans from glycoproteins in a 96/384 well plate format (Royle et al., 2008; Stöckmann et al., 2015b). It has been developed as a front end to glycan separations technologies including HILIC/MS/MS and capil-

* Corresponding author.
E-mail address: pauline.rudd@nibrt.ie (P.M. Rudd).

http://dx.doi.org/10.1016/j.pisc.2016.01.013

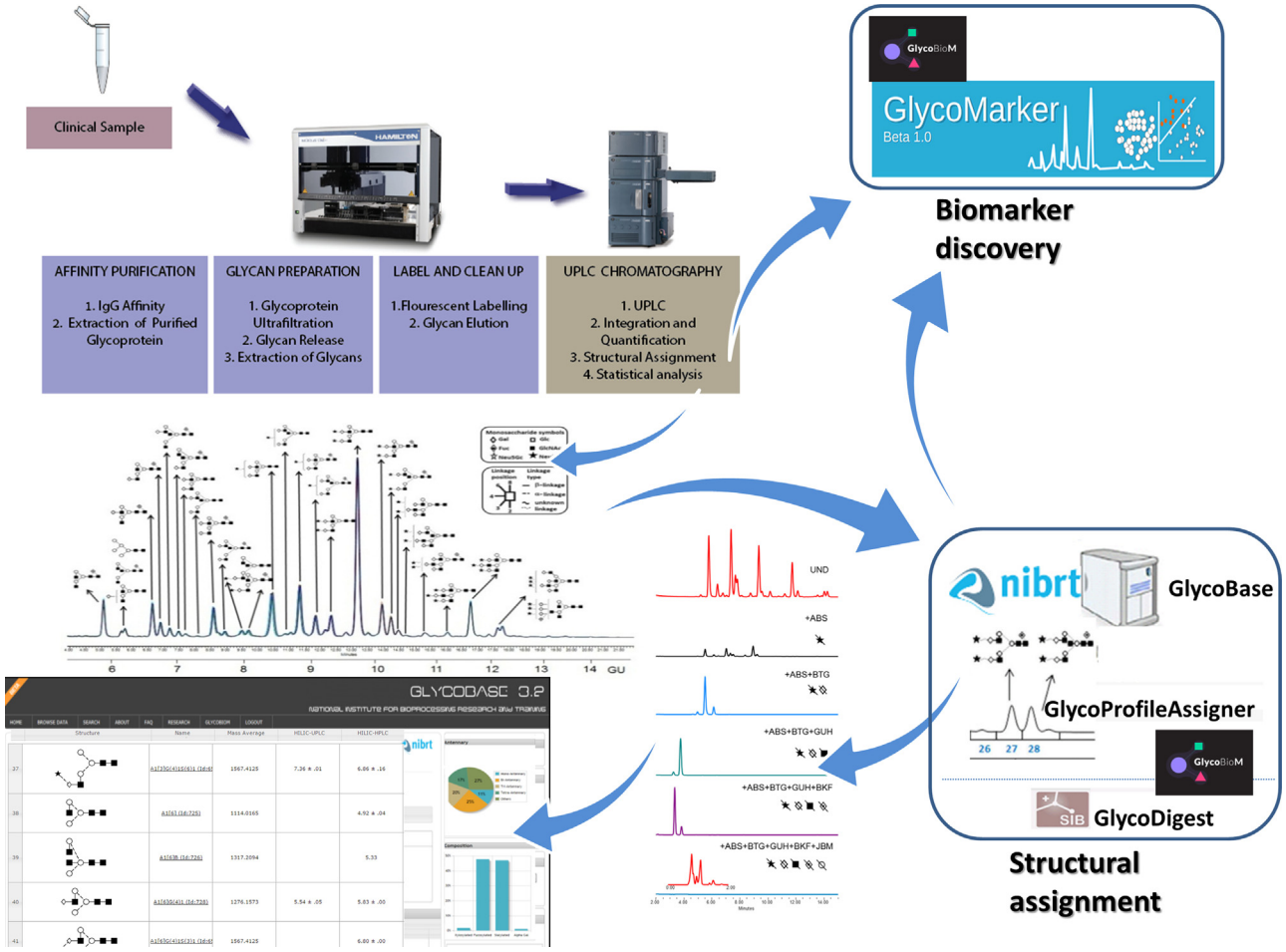


Figure 1 Glycomics workflow. Sera/plasma, individual glycoproteins, mixtures or cell culture supernatant samples are processed on a robotic workstation, resulting in fluorescently labelled glycans, which are subsequently separated into pools and quantified by ultra-high pressure liquid chromatography (UPLC). Bottom right structures are predicted using GlycoBase 3.2 and confirmed by enzyme digestion using bioinformatics software (Duffy and Rudd, 2015). Top right the high quality data produced by the platform can be analysed using new GlycoMarker bioinformatics software to find glycan biomarkers.

lary electrophoresis. The platform was initially optimised for analysing biopharmaceuticals at each stage of the manufacturing process but is also applicable to basic research; in particular the linking of extensive sets of disparate data for systems biology. The plate format makes it convenient for large sample sets; it is relatively cheap, robust and quantitative.

Calibration of the robotic platform produces massive amounts of raw data which need to be annotated and analysed. Annotation involves the integration of profile peaks and the assignment of glycan structures to each peak but this is far from a simple process. To give an idea of the complexity, in mammalian systems it has been estimated that 10^{12} structures are possible (Laine, 1994) although it has been estimated $\geq$7000 structures exist in nature (with approximately 700 proteins required to generate this diversity) (Cummings, 2009). These quantities alone present significant analytical challenges for determining detailed, quantitative glycan structural data in complex organisms. Bioinformatics applications are therefore essential to speed-up structural annotation or in the ideal case completely automate it. At the analysis stage, the large scale production of high quality experimental data can be inspected with the goal of detecting important molecular characteristics. For example, as glycans undergo rapid structural changes in response to biological stimuli they provide a unique opportunity to identify and exploit glycans as clinical markers that can be indicative of specific disease states, disease progression, and/or therapy response. Moreover, data can be linked to other parameters such as metabolomics and genomics data pointing towards the inclusion of glycan data in 'big data' sets for a better understanding of disease. However, manually analysing large collections of samples is a cumbersome process and therefore a need for bioinformatic tools is a vital component at the data analysis stage.

In this article we briefly describe our high-throughput platform and pay particular attention to the bioinformatics tools used to annotate and analyse the large quantities of data produced. Bioinformatics is a core component of our pipeline and greatly speeds up the storage, annotation and analysis of the data. The bioinformatics programmes enable the less experienced researcher to handle the data and the possibility of integrating the tools with other—omics data is now on the horizon.