**DBP: A systems approach towards the therapeutic modulation of the acetylome**

|  |  |
| --- | --- |
| Collaborating PI(s): | Anne-Claude Gingras, Ph.D. |
| Institution: | Samuel Lunenfeld Research Institute, Toronto, Canada |
| Funding: |
| CIHR/IRSC | 275594, Sep 2012 to Aug 2016, PI Gingras |

## 1 Significance

The cellular machinery responsible for the acetylation of lysines is critically implicated in cancer biology. In particular the only protein domain that recognizes acetylated lysines, the bromodomain, is present in multiple proteins implicated in chromatin biology and regulation of gene expression, and many of these proteins are deregulated in human tumors. Structurally [], the recognition of acetylated peptides by the bromodomain makes it an attractive druggable target. Through the work of one of the co-PIs on this CIHR grant (Filippakopoulos at the Structural Genomic Consortium, SGC, Oxford, UK), the first anti-cancer bromodomain inhibitor (JQ1; specific for the BET family of bromodomains) was developed and shows promise against a number of tumor types. However, this work also highlighted the need for more information regarding the cellular context in which bromodomain containing proteins and acetylation enzymes act, and what are the cellular targets for bromodomain recognition of acetylated peptides. Here, we propose to systematically define the specificity and function of human bromodomains through identification of their binding partners and acetylated targets. This project synergizes with other funded projects by co-applicant Filippakopoulos that employ structural biology to investigate protein-protein interactions involving epigenetic readers.

**The acetylation machinery (acetylome).** Lysine acetylation (KAc) is the process by which an acetyl group is transferred (from Acetyl Coenzyme A) to the epsilon amine of a lysine residue, a modification that is catalyzed by a family of lysine acetyltransferases (KATs; 18 genes in human). Similar to protein phosphorylation in which kinases and phosphatases oppose each other for the addition (“writing”) and removal (“erasing”) of the phosphate group, KAc added by KATs can be removed by lysine deacetylases (KDACs, histone-specific KDACs are known as HDACs, histone deacetylases; 18 genes). KAc marks are recognized by a single dedicated protein domain, the bromodomain, which acts as a “reader” of the acetylation mark and provides context-specific recognition (44 genes in human; some proteins contain more than one bromodomain). Lysine acetylation first came to the forefront as a histone modification. Histone tails are modified by a number of post-translational modifications (PTMs) that present different cues to the cellular machinery (the so-called histone code). In general, acetylation of histone tails acts to open the chromatin structure, activating transcription, although certain acetylation marks are associated instead with chromatin compaction and with other processes as well, such as metabolism and DNA repair. However, while first studied intensively in the context of the histone code, lysine acetylation is now recognized as a widespread PTM occurring on a large proportion of the proteome. So far, 18330 sites on 6870 proteins have been reported in total in human cells in a PTM repository (http://www.phosphosite.org), which is likely still an underestimation (we consistently detect new sites by mass spectrometry). The prevalence of KAc on thousands of cellular proteins underlies the vital importance of this modification and also highlights the lack of information regarding which KATs and KDACs are responsible for these acetylation events, and, just as importantly, which sites on which proteins are “read” by the different bromodomains.

**Targeting the acetylome in disease**. Tight control of the acetylome is critical to cellular homeostasis, and many of the acetylome components have been associated with disease, particularly cancer. development and(HDAC) , and two of them (Voronistat, a broad specificity inhibitor, and Romidepsin, a KDAC1/KDAC2 inhibitor) have been approved for However, whileclear Given the more specific nature of the acetylation recognition (see below), bromodomains are attractive drug targets. Bromodomain alterations in disease include expression changes, but also a number of genetic alterations, most notably deletions, point mutations and translocations. For example, translocations of the BET (Bromodomain and Extra Terminal domain) family members BRD4 and BRD3 to the NUT protein cause a rare and aggressive cancer known as midline carcinoma. Disrupting the KAc binding with the small molecule inhibitor JQ1 reverses the tumor phenotype of midline carcinoma cells, offering an excellent proof-of-concept for targeting protein-protein interactions for epigenetic “readers”. Importantly, the anti-proliferative effects of JQ1 and other recently developed BET-family inhibitors extents beyond this rare malignancy, as BRD4 is critical for the growth of c-Myc driven cancers. As several other bromodomain-containing proteins are implicated in disease, there is a large ongoing effort for the development of additional specific and pan-specific inhibitors of the KAc-bromodomain interaction, including in the unique industry/academic partnership led by the SGC (early access to newly developed compounds is provided through our collaborations; http://www.thesgc.org/scientists/epigenetics).

**Recognition of acetylated lysines by bromodomains.** Despite variation in primary amino acid sequence, all bromodomains (a domain of 120 aa) share a common fold, and bind KAc in a similar manner. The bromodomain consists of a left-handed bundle of four alpha helices that are linked by more variable loop insert regions. KAc is recognized in a central deep and largely hydrophobic cavity, as determined by co-crystal structures of bromodomains complexed with KAc-containing peptides. Anchoring of the acetylated lysine (or lysines, see below) is provided by hydrogen bonding to a conserved asparagine residue. How bromodomains specifically recognize distinct acetylated peptides is beginning to be understood, in large part due to our group’s efforts to systematically determine the structures of multiple bromodomains []. For example, while the structural fold of all bromodomains is similar, the electrostatic potential of the surfaces surrounding the KAc binding site is diverse, suggesting that they recognize different sequences (e.g. positively charged histones may not be favored by bromodomains that have a positive charge, such as that of the first bromodomain of PBRM1). Note that differences in surface charge have been observed even within the same protein, including the first and second domains of BRD4, suggesting that different bromodomains within a protein may target different sites, an hypothesis that we addressed directly for BRD4, as described below. Additional selectivity elements are provided by the diverse loop regions that distinguish the different bromodomains.

Figure 1: **Influence of Neighboring PTMs on BRD Interaction with Histone H3.** Shown are interactions shaded by spot intensity as indicated in the legend at the upper-left corner of the figure. The influence of lysine trimethylation (Kme3), acetylation, and phosphorylation (pT, pS) has been studied on binding to a central acetylated lysine epitope. The combination of the different modifications is indicated in the right panel. Nonmodified peptides have been included as controls to identify interactions independent of lysine acetylation. (from Filippakopoulos et al. [])

**How is the acetylome organized?**  There is a pressing need to characterize systematically and in an unbiased manner the function of each of the components of the acetylation machinery and the relationships between the writer-reader-eraser modules. As mentioned above, structurally, bromodomains offer attractive targets for therapies and are the topic of an intense drug discovery program, yet very little is known regarding the cellular context in which many of the bromodomain containing proteins reside or their specificity for their targets in vivo. We have used synthetic libraries of histone-derived peptides to provide information regarding the specificity of isolated bromodomains in vitro (see Figure ). This expanded the number of bromodomain substrates, but also led to the realization that flanking PTMs (especially phosphorylation and acetylation) have a strong influence on the recognition of marks, indicating that bromodomains often recognize combinations of marks rather than isolated KAc. This observation was previously made for BRDT which requires the presence of several acetylation sites for high-affinity binding to histone tails. Structural determination of BRD4 with different diacetylated peptides derived from histone H4 showed that both KAc groups bind within the same pocket, and that the different sequences exhibit the same mode of binding. However, our studies have also outlined the need to use alternative approaches to the synthetic libraries to identify specific sequences to which each of the bromodomains are associated in a cellular context. This is important especially in light of the therapeutic design of inhibitors: the inhibitors should be able to displace relevant (and presumably higher affinity) KAc peptides from the bromodomain.

## 2 Innovation

Identification of changes in protein-protein interactions associated with treatment with inhibitors, variations in patterns of post-translational modifications (PTMs) and sequence variants such as mutations is critical to understanding the consequences of such variation. Technologically speaking, this requires robust, sensitive and accurate methods for both identification and quantification from mass spectrometry data. We have successfully coupled affinity purification (AP) to data-independent mass spectrometric acquisition (SWATH) to analyze the interactome changes imparted by mutations or pharmacological inhibition of protein-protein interactions; structural information can be gained by combining this proven pipeline with chemical crosslinking. In collaboration with CCMS, we will develop an integrated computational and experimental pipeline that will be scalable to the flood of data arising from our analysis of the impact of PTM variations in bromodomain interactions. This collaboration with CCMS is synergistic and timely, based on our preliminary data, the availability of the appropriate samples, a strong expertise in structural analysis of protein-protein interactions, and the collaboration we have established with AB SCIEX. The AP-SWATH pipeline we are proposing to co-develop with CCMS will greatly accelerate biological knowledge regarding the consequences of proteome variations in protein-protein interactions.

Figure 2: **Differences between DDA and DIA mass spectrometry methods.** (top) In classical Data Dependent Acquisition (DDA), co-eluting peptides are first analyzed prior to fragmentation in a survey scan (MS1). The *n* most abundant species are then sequentially isolated in the mass spectrometer, fragmented in the CID chamber and the fragment masses are measured in the TOF chamber. A typical MS/MS spectrum is shown on the right: these spectra are used for peptide identification. (bottom) In Data Independent Acquisition (DIA, here SWATH), there is no decision to select ions for fragmentation based on a precursor scan. Instead, the Q1 filters ions by relatively large mass windows: co-filtered peptides are co-fragmented in the CID chamber and co-analyzed in the TOF chamber. This results in a mixed spectrum, shown on the right, from which identification and quantification information can be retrieved using bioinformatics tools. The process is rapid such that the entire mass range can be analyzed in 3 seconds.

## 3 Approach

**Determining specificity of bromodomains for acetyl-lysine residues.** We have already established a static interactome surrounding 77 of the acetylome components, each epitope tagged (FLAG) and expressed at near endogenous levels in human cells. Following affinity purification and mass spectrometry using a approched called mChIP which we developed for capturing interactions involving chromatin-associated proteins, ~3000 high-confidence interactions were detected that cover ~900 proteins. This constitutes our baseline interactome, which we are using to monitor the changes in interaction profiles associated with mutations (point mutations, translocations) and following cell perturbation, for example using the bromodomain-specific inhibitors (this will be done by SWATH; see below). Our interactome map is critical to understand the processes in which bromodomain-containing proteins and other acetylome components are acting; however, how the recognition of the acetylated lysine by the bromodomain is related to the interactome is unclear at this point. To begin to characterize interactome specificity, we are proposing a multi-pronged approach to identify acetylation-dependent interactions, both in the context of recombinantly-expressed isolated bromodomains (tested both on synthetic peptides, on purified histones and on cell extracts), but also with full length proteins expressed in human cells. A key element of our strategy is to be able to identify acetylation sites directly from immunoprecipitates of bromodomain-containing proteins. To start, we have modified our mChIP strategy to add, after the tryptic digest, an additional affinity step consisting of an anti-KAc pulldown at the peptide level and analysis on a high mass resolution mass spectrometer. We first tested this approach on the bromodomain-containing protein BRPF3, a component of the MOZ/MORF histone acetyltransferase complex. In addition to the known BRPF3 interaction partners, our interaction proteomics approach identified new interactors for this protein. After the KAc enrichment approach, we also found acetylation sites on BRPF3 itself, and on many of its interactors. While several of these sites were present in a public repository (PhosphoSitePlus.org), 15 are new (and have been manually validated), highlighting the sensitivity of our approach. We also optimized the mass spectrometric identification of acetylated sites by analyzing fractions of the same purification on different mass spectrometers (and in the case of the Orbitrap Velos by employing different modes of fragmentation). Importantly, we show that – in comparison to a proteome-wide study of acetylation – our targeted study of the acetylome facilitates the identification of polyacetylated peptides. This is an important result, especially in light of our recent study where we demonstrated that at least some of the bromodomains exhibit a preference for polyacetylated sites, both in vitro and in vivo, and that this preference may contribute to the recognition of histone code marks. It is also noteworthy that many of the KATs (especially those of the MYST family) require auto-acetylation for activity. As such, being able to identify and study these critical residues is important. We have also developed approaches to determine the specificity of recombinant isolated bromodomains against synthetic peptides and histone preparations.

While the static interactome work described above was employing semi-quantitative measures to discriminate between true and false positive interactions (all work is currently handled in our interaction proteomics LIMS ProHits), this approach proved insufficiently accurate to capture mutation/treatment-induced changes in association. Furthermore, the approach is not appropriate for quantification at the peptide level, which is what is required for the identification of the acetylated peptide specificity. Co-applicants on this grant (Pawson and Gingras) have coupled affinity purification (AP) to targeted mass spectrometry (SRM) in the past, to successfully quantify interactome changes (ref, and Zheng et al., under review at *Nature*), but needed to develop a more rapid method to rapidly quantify multiple interactome components across varied conditions. In collaboration with AB SCIEX, we therefore implemented an AP-SWATH approach, which was first validated on known regulated interactions (Lambert et al., *submitted*). We next applied this pipeline to the investigation of the changes in BET family interaction profiles imparted by treatment with the JQ1 inhibitor (an inactive isomer was used as a control). As expected based on the mode of action of the pharmacological compound, association of BRD3 with histones was markedly decreased, while interaction with several other proteins was maintained, indicating that they are not mediated by the KAc-bromodomain interaction. Surprisingly, new interaction partners were found to be enriched following JQ1 treatment, suggesting new links to DNA damage and transcription. Similar data were obtained for other BET family members, with JQ1 and other inhibitors. In these experiments, quantification of SWATH data was accomplished by a targeted data extraction strategy using a spectral library which we built from the same samples by standard DDA. In collaboration with CCMS, we aim to extend this approach to *i*) aggregate all DDA runs into a reusable spectral library that can be used for analysis of any SWATH run (interacts with TRD3), *ii*) develop methods for peptide identification (instead of just quantification) by spectral library and database searching of SWATH spectra (interacts with TRD3 and TRD7) and *iii*) develop methods for identification of modified peptides for which only the unmodified or less-modified version of the peptide is present in either a database or the spectral library (interacts with TRD3 and TRD7). As a logical and direct extension of this work, we will also combine the use of chemical crosslinkers to explore the structural organization within protein complexes: proof of principle publications indicate that the approach is directly compatible with our affinity purification strategy (e.g. Ruedi science paper), and we have successfully crosslinked several complexes using BS3 and DSP. The SWATH approach should be more sensitive than the previously employed DDA acquisition, enabling us to achieve a more extensive characterization of the intra- and intermolecular contacts in our protein complexes. In collaboration with CCMS, we have already begun investigating fragmentation patterns for linked peptides using SWATH. Combining crosslinking with SWATH will enable us to more rapidly identify direct interaction partners in our experiments, thereby facilitating the selection of proteins to follow-up in co-crystal studies at the SGC. At the same time, knowledge of the interacting regions provided by AP-crosslinker-SWATH will permit the selection of the appropriate regions to analyze structurally, which is critical in this project in which many of the proteins are extremely large (upwards of 200kDa) and not possible to express recombinantly in amounts sufficient for crystallography.

The data already planned for acquisition in this funded project will be ideal for these TRDs because it contains validated identifications of modified peptides (e.g., the subset of quantifiable peptides from Figure ) but also provides opportunities for potentially discovering new modified peptides or new interaction partners as we become able to directly search SWATH spectra instead of limiting their use to peptide quantification only.

References

[1] P. Filippakopoulos, S. Picaud, M. Mangos, T. Keates, J. P. Lambert, D. Barsyte-Lovejoy, I. Felletar, R. Volkmer, S. Muller, T. Pawson, A. C. Gingras, C. H. Arrowsmith, and S. Knapp. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell*, 149:214–231, 2012.