Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2022 March 10.

Published in final edited form as:

Cell Rep. 2020 December 01; 33(9): 108450. doi:10.1016/j.celrep.2020.108450.

The Nucleosome Remodeling and Deacetylase Complex Has an Asymmetric, Dynamic, and Modular Architecture

Jason K.K. Low[1,7,*], **Ana P.G. Silva**[1,7], **Mehdi Sharifi Tabar**[1,8], **Mario Torrado**[1,8], **Sarah R. Webb**[1], **Benjamin L. Parker**[1], **Maryam Sana**[1], **Callum Smits**[6], **Jason W. Schmidberger**[1], **Lou Brillault**[2], **Matthew J. Jackman**[2], **David C. Williams Jr.**[3], **Gerd A. Blobel**[4], **Sandra B. Hake**[5], **Nicholas E. Shepherd**[1], **Michael J. Landsberg**[2,*], **Joel P. Mackay**[1,9,*]

[1]School of Life and Environmental Sciences, University of Sydney, NSW, Australia

[2]School of Chemistry and Molecular Biosciences, University of Queensland, QLD, Australia

[3]Dept of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, NC, USA

[4]The Division of Hematology, Children's Hospital of Philadelphia, and the Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[5]Institute for Genetics, FB08 Biology, Justus-Liebig-University Giessen, Giessen, Germany

[6]Sydney, NSW, Australia

[7]These authors contributed equally

[8]These authors contributed equally

[9]Lead Contact

SUMMARY

The nucleosome remodeling and deacetylase (NuRD) complex is essential for metazoan development but has been refractory to biochemical analysis. We present an integrated analysis of the native mammalian NuRD complex, combining quantitative mass spectrometry, cross-linking, protein biochemistry, and electron microscopy to define the architecture of the complex. NuRD is built from a 2:2:4 (MTA, HDAC, and RBBP) deacetylase module and a 1:1:1 (MBD, GATAD2, and Chromodomain-Helicase-DNA-binding [CHD]) remodeling module, and the complex displays considerable structural dynamics. The enigmatic GATAD2 controls the asymmetry of the complex and directly recruits the CHD remodeler. The MTA-MBD interaction

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

*Correspondence: jason.low@sydney.edu.au (J.K.K.L.), m.landsberg@uq.edu.au (M.J.L.), joel.mackay@sydney.edu.au (J.P.M.).

AUTHOR CONTRIBUTIONS

A.P.G.S., M.S.T., M.T., S.R.W., and J.W.S. prepared samples for EM experiments; A.P.G.S., S.R.W., J.W.S., L.B., M.J.J., and M.J.L. contributed to EM data collection and processing; A.P.G.S., C.S., J.W.S., and J.P.M. contributed to the modeling; J.K.K.L., M.S.T., M.T., and B.L.P. contributed to DIA-MS and XLMS sample preparation and data collection; J.K.K.L. and B.L.P. performed the DIA-MS and XLMS data analysis; M.T. and M.S. performed the PWWP2A experiments; S.B.H., D.C.W., G.A.B., and N.E.S. provided intellectual input on the studies and manuscript; A.P.G.S., J.K.K.L., M.J.L., and J.P.M. designed and supervised the studies and wrote the manuscript with input from all the authors.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.108450.

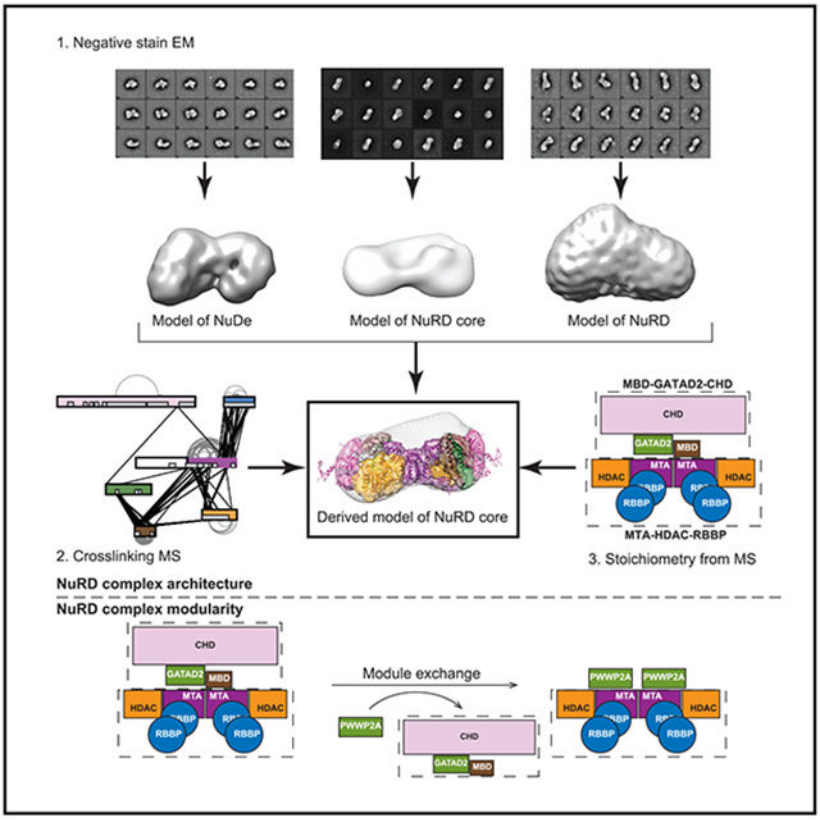
DECLARATION OF INTERESTS

The authors declare no competing interests.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

acts as a point of functional switching, with the transcriptional regulator PWWP2A competing with MBD for binding to the MTA-HDAC-RBBP subcomplex. Overall, our data address the long-running controversy over NuRD stoichiometry, provide imaging of the mammalian NuRD complex, and establish the biochemical mechanism by which PWWP2A can regulate NuRD composition.

Graphical Abstract



In Brief

Low et al. examine the architecture of the nucleosome remodeling and deacetylase complex. They define its stoichiometry, use cross-linking mass spectrometry to define subunit locations, and use electron microscopy to reveal large-scale dynamics. They also demonstrate that PWWP2A competes with MBD3 to sequester the HDAC-MTA-RBBP module from NuRD.

INTRODUCTION

The physical organization of DNA is a critical determinant of genome function. ATP-dependent chromatin remodeling enzymes use a conserved DNA translocase domain to alter the positions, occupancy, and composition of nucleosomes, thereby regulating the availability of DNA for transcription, replication, or repair. Chromatin remodelers typically exist as large multisubunit complexes *in vivo*, and despite recent high-resolution structures of nucleosomes bound to the INO80 (Ayala et al., 2018; Eustermann et al., 2018) and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

SWR1 complexes (Willhoft et al., 2018), as well as to the isolated Snf2 (Liu et al., 2017), Chromodomain-Helicase-DNA-binding 1 (CHD1) (Farnung et al., 2017), and CHD4 remodelers (Farnung et al., 2020), our understanding of how such enzymes bring about remodeling is still underdeveloped. This is particularly true for the nucleosome remodeling and deacetylase (NuRD) complex, which is not found in yeast and therefore represents a considerable challenge for structural analysis.

The NuRD complex is widely distributed among metazoans and is expressed in most, if not all, tissues. It is essential for normal development (Gómez-Del Arco et al., 2016; Yoshida et al., 2008) and is a key regulator in the reprogramming of differentiated cells into pluripotent stem cells (dos Santos et al., 2014; Luo et al., 2000; Rais et al., 2013). Age-related reductions in NuRD subunit levels are strongly associated with memory loss, metastatic potential in human cancers (Toh and Nicolson, 2009), and the accumulation of chromatin defects (Pavlopoulos et al., 2013; Pegoraro et al., 2009).

The mammalian NuRD complex comprises at least six subunits (Figure S1A), and for each subunit, there are at least two paralogs. CHD4 (and its paralogs CHD3 and CHD5) is the DNA translocase in the complex but NuRD is distinguished from many other remodelers by its second catalytic activity, which is imparted by the histone deacetylases HDAC1 and HDAC2. MBD2 and MBD3 can bind hydroxymethylated and/or methylated DNA (Cramer et al., 2014; Liu et al., 2018; Yildirim et al., 2011), and RBBP4 and RBBP7 can each bind histone tails (Murzina et al., 2008) and other transcriptional regulators (Lejon et al., 2011; Liu et al., 2015; Moody et al., 2018; Sankaran et al., 2008). The metastasis-associated proteins MTA1, MTA2, and MTA3 contain several domains associated with nucleosome recognition, whereas GATAD2A and GATAD2B bind to both MBD2 and MBD3 (Gnanapragasam et al., 2011; Torrado et al., 2017) as well as to CHD proteins (Sher et al., 2019; Torrado et al., 2017) but otherwise have no known functions.

Some structures are available for portions of the NuRD complex (Figure S1B): (1) HDAC1 forms a 2:2 complex with an N-terminal segment of MTA1 (Millard et al., 2013), (2) two copies of RBBP4 can bind the C-terminal portion of MTA1 (Alqarni et al., 2014; Millard et al., 2016; Schmidberger et al., 2016), (3) MBD2 and GATAD2A form a heterodimeric coiled-coil (Gnanapragasam et al., 2011; Walavalkar et al., 2013), and (4) a cryoelectron microscopy (cryo-EM) structure shows the catalytic domain of CHD4 bound to a nucleosome (Farnung et al., 2020). However, no structural data at any resolution exist for the intact mammalian NuRD complex, and even the subunit stoichiometry is uncertain. Recent studies using label-free mass spectrometry (Bode et al., 2016; Guo et al., 2019; Kloet et al., 2014; Smits et al., 2013; Spruijt et al., 2016; Sharifi Tabar et al., 2019) have yielded variable results that are often at odds with the stoichiometries demonstrated in the known subcomplex structures. For example, MTA:HDAC ratios ranging from 2:2 to 1:3 have been reported in the literature.

The mechanisms by which NuRD selects target sites are also poorly understood. Transcriptional regulators such as FOG1 (Hong et al., 2005; Lejon et al., 2011) and BCL11A (Moody et al., 2018; Xu et al., 2013) can bind to NuRD via the RBBP subunits, but this mechanism is likely to account for only a small proportion of NuRD-genome

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

interactions. Recently, we demonstrated that the chromatin-binding protein PWWP2A, which can selectively recognize H2A.Z-containing nucleosomes (Link et al., 2018; Pünzeler et al., 2017), interacts robustly with the MTA, HDAC, and RBBP subunits of NuRD but not with GATAD2, MBD, and CHD. The nature of this interaction is currently unknown.

Against this background, we have used structural, biophysical, and biochemical data to define the architecture of the native NuRD complex purified from murine erythroleukemia (MEL) cells. We first determined the subunit stoichiometry of the complex using the gold-standard absolute quantification (AQUA) strategy (Gerber et al., 2003) in combination with data-independent acquisition mass spectrometry (DIA-MS) (Gillet et al., 2012), providing a substantial increase in accuracy over previous measurements. We derived a 4:2:2:1:1:1 ratio of RBBP, HDAC, MTA, MBD, GATAD2, and CHD subunits, showing that the complex is a mixture of symmetric and asymmetric subcomplexes.

We identified stable subcomplexes within NuRD and showed that the full complex is composed of two parts with separable enzymatic activities (an MTA-HDAC-RBBP subcomplex for deacetylation and an MBD-GATAD2-CHD complex for remodeling). Covalent cross-linking combined with mass spectrometry (XLMS) analysis of four separate complexes provided >300 unambiguous intermolecular cross-links (XLs), a ~20-fold increase compared with that of previous studies. These XLs allowed us to infer positions for several domains in the complex that were previously uncharacterized. These models are corroborated by negative-stain EM data that provide the first images of the full mammalian NuRD complex and demonstrate that NuRD displays considerable conformational structural dynamics. We propose that these dynamics involve large-scale movements of the RBBP, GATAD2, and CHD subunits relative to the HDAC-MTA core.

Our data also indicate that, despite the underlying dimeric structure of the deacetylase module, the presence of the CHD-containing translocase module confers overall asymmetry on the complex. The stoichiometry of the connection between the modules is coordinated by the GATAD2 subunit, and unexpectedly, we observe that the interface between the MTA-HDAC-RBBP and MBD-GATAD2-CHD modules is a site of regulation by other proteins. We demonstrate that direct competition between MBD and the co-regulator PWWP2A for binding to the same surface of MTA-HDAC-RBBP is responsible for the formation of distinct complexes with very different functionality.

RESULTS

Subunit Stoichiometry Measurements Partition the NuRD Complex into Symmetric and Asymmetric Modules with Distinct Catalytic Activities

We purified native mammalian NuRD complex from MEL cells by using our established protocol (Figure 1A; Low et al., 2016). This strategy also yields a complex lacking the CHD subunit, which we have previously termed the nucleosome deacetylase (NuDe) complex. To rigorously establish the subunit composition of these complexes, we carried out absolute quantification using AQUA peptide standards together with DIA-MS (Gerber et al., 2003; Gillet et al., 2012). The use of subunit-specific $^{13}C/^{15}N$-labeled peptides as internal standards provides a significant improvement in accuracy over published label-free derived

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

NuRD stoichiometries (Bode et al., 2016; Guo et al., 2019; Kloet et al., 2014; Smits et al., 2013; Spruijt et al., 2016; Sharifi Tabar et al., 2019).

For both NuDe and NuRD, we derived ratios of ~2:2:7:1:1 for MTA:HDAC:RBBP:MBD:GATAD2 (Figure 1B; Data S2). CHD4 was consistently sub-stoichiometric, reflecting the fact that complete separation of NuDe and NuRD was not possible because of overlap in their sedimentation profiles. The numbers are calculated relative to an MTA:HDAC ratio of 2:2 (see STAR Methods), based on the stoichiometry of the crystal structure of this complex (Millard et al., 2013).

Of all subunits, RBBP displayed the highest variability, with values ranging from 5 to 9 subunits per complex (Figure 1B). However, our purification protocol involves initial affinity capture with a peptide from FOG1 fused to glutathione S-transferase (GST) (Hong et al., 2005), and this fusion remains with the complex during the purification process. Because of the propensity of GST to dimerize (Riley et al., 1996), it is possible that additional GST-FOG1:RBBP complexes co-purify with NuRD, confounding the determination of RBBP stoichiometry.

To resolve this issue and more precisely measure the RBBP content, we reconstituted a number of NuRD subcomplexes. Previous structural and biochemical work shows that it is the C-terminal half of MTA1 that is responsible for recruiting RBBPs to the NuRD complex (Millard et al., 2016; Schmidberger et al., 2016; Torrado et al., 2017). We therefore expressed and purified a subcomplex comprising the N-terminal half of MTA2 (residues 1–429; MTA^{N}), HDAC1, and $MBD3^{GATAD2CC}$ (MTA^{N}-HDAC-$MBD^{GATAD2CC}$; Figure S2A). $MBD3^{GATAD2CC}$ is full-length MBD3 stabilized by fusion to its dimerization partner—the first coiled-coil domain of murine GATAD2A (residues 133–174) (Desai et al., 2015; Figure S1A). Quantitative DIA-MS analysis showed that this MTA^{N}-HDAC-$MBD^{GATAD2CC}$ complex—the isolation of which does not rely on a GST-FOG1 affinity purification step—contains little or no RBBP protein, consistent with the idea that RBBP subunits are recruited to the complex solely by the C-terminal half of MTA (Data S2). To complement this result, we expressed an MTA^{C}-RBBP subcomplex comprising the C-terminal half of MTA1 (residues 449–715) and RBBP4 (Figure S2A); DIA-MS showed that this complex has a stoichiometry of ~1:2.5 (Figure 1B; Data S2), which is much closer to the expected 1:2 ratio. Finally, we co-expressed and purified a complex comprising full-length MTA2, HDAC1, and RBBP7 (MTA-HDAC-RBBP; Figure S2A), which yielded a subunit ratio of ~2:2:4 (Figure 1B; Data S2).

As an orthogonal approach to assessing stoichiometry, we directly measured the molecular mass of the NuDe and NuRD complexes by using size exclusion chromatography coupled to multi-angle laser light scattering (SEC-MALLS). In both cases, we observed masses that were within 7% of the expected mass for a 2:2:4:1:1:1 complex (MTA:HDAC:RBBP:MBD:GATAD2:CHD4; Figure S3A). Taken together, our data argue that the intact NuRD complex derived from M EL cells has a stoichiometry of 2:2:4:1:1:1 (MTA:HDAC:RBBP:MBD:GATAD2:CHD4). The DIA-MS data also allowed us to quantify paralog abundance in the native NuRD and NuDe complexes isolated from MEL cells. As shown in Figure S1C, a distinct preference for HDAC1 over HDAC2 is observed, whereas

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

MTA3 and MBD2 are nearly absent from this complex. The two GATAD2 paralogs are similarly abundant. Little or no CHD3 or CHD5 were detected in our samples.

XLMS Data Establish the Core Architecture of the NuRD Complex

By integrating the findings of our stoichiometry analysis with published biophysical and structural work on *Drosophila* NuRD and several subcomplexes (Alqarni et al., 2014; Gnanapragasam et al., 2011; Millard et al., 2013, 2016; Schmidberger et al., 2016; Zhang et al., 2016), we can conclude that NuRD is built from a compositionally symmetric 2:2:4 MTA-HDAC-RBBP deacetylase module and an asymmetric 1:1:1 MBD-GATAD2-CHD remodeling module. Building on this conclusion, we next asked how the deacetylase and remodeling modules of NuRD are physically connected. Our prior work has established that there are no direct interactions between the deacetylase module (MTA-HDAC-RBBP) and either GATAD2 or CHD and thus that the only interaction linking these two modules is between MTA and MBD (Torrado et al., 2017). Our ability to co-express and purify a stable MTA^{N}-HDAC-$MBD^{GATAD2CC}$ subcomplex (see above) confirms that MBD directly binds the deacetylase module, thereby bridging the two halves of the complex. This result also shows that neither the C-terminal half of MTA nor the RBBPs are required to couple the two modules. A structure of the MBD-GATAD2 complex (Gnanapragasam et al., 2011) demonstrates that these two proteins directly interact, meaning that GATAD2 is recruited to the complex by MBD.

Consistent with this idea, co-expression of MTA2, HDAC1, MBD3, and GATAD2A yields a stable complex containing all four proteins (MTA-HDAC-MBD-GATAD2; Figure S2A). Notably, this complex recruits significant amounts of endogenous CHD, whereas the MTA^{N}-HDAC-$MBD^{GATAD2CC}$ complex does not (Figures S2A and S2B). We therefore conclude that the architecture of NuRD consists of a dimeric deacetylase core of MTA, HDAC, and RBBP subunits that directly interacts with MBD; in turn, MBD binds GATAD2 and it is GATAD2 alone that dictates the recruitment of CHD and ultimately confers DNA translocase activity on the full complex (Figure 2A).

To examine the placement of subunits and, in particular, how the MBD subunit couples the two halves of the complex, we carried out XLMS on NuRD and several subcomplexes (Figure S2B). These experiments yielded a total of 752 unique XLs that were highly conserved across MTA^{N}-HDAC-$MBD^{GATAD2CC}$, MTA-HDAC-MBD-GATAD2, MTA-HDAC-RBBP, NuDe, and NuRD complexes (Figure S4A; Data S1). We observed a high density of XLs between (1) RBBP and the C-terminal half of MTA, (2) HDAC and the N-terminal half of MTA, (3) MBD and the N-terminal half of MTA, and (4) the first coiled-coil region of GATAD2 (residues 136–178) and MBD3. In total, 174 of the XLs connect pairs of residues in known structures of subunits and subcomplexes. Ninety-four percent of these XLs are consistent (within the accepted crosslinker cutoff) with the extant structural data (Figure S4B; Data S1), providing a high level of confidence that these subcomplex structures recapitulate the architecture of the full NuRD complex in solution.

In addition to the XLs above, which establish the veracity of the approach, 120 XLs connect pairs of domains with known or readily modeled structures, but for which relative positions in the NuRD complex are unknown. We have defined these domains as shown in Figure S1:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

MTA1^{BAH} (the BAH domain of MTA), the dimer of heterodimers HDAC1-MTA1^{ES} (the ELM-SANT domains of MTA), MTA^{H} (a 20-residue predicted α-helix to which 19 XLs were observed), MTA^{ZF} (a GATA-type zinc-finger domain), MBD^{MBD} (the methyl-DNA binding domain), and MBD^{CC}-GATAD2^{CC1} (the heterodimeric coiled-coil formed by the MBD3 coiled-coil domain and the first coiled-coil domain of GATAD2A). XL-driven rigid-body docking in HADDOCK (van Zundert et al., 2016) finds a solution for the domain architecture that is supported by 93% of these 120 XLs (Figures 2B and S5). The model places the BAH domain at the distal ends of the MTA^{ES}-HDAC dimer of dimers where it is juxtaposed with the MBD^{MBD} domain. In turn, MBD^{MBD} packs against the MBD^{CC}-GATAD2^{CC1} coiled-coil, despite the two domains of MBD being separated by a 145-residue intrinsically disordered region (MBD^{IDR}; Desai et al., 2015). Only one copy of MBD3 is placed in the model because of the complex stoichiometry dictated by our DIA-MS data.

In our model, the MTA^{H} helix lies directly adjacent to the HDAC active site. Figure 2C highlights the well-described phenylalanine at the entry to the active site (Phe150, blue), and we have indicated the reduced degree of access to the active site by modeling in a known HDAC inhibitor, hydroxamic acid (based on a structure of HDAC8; PDB: 5FCW) (Tabackman et al., 2016). Our model thus suggests that the MTA subunits in NuRD could potentially modulate HDAC activity.

Although the MBD^{IDR} is disordered in isolation (Desai et al., 2015), it is predicted to be ordered by programs such as PONDR (Obradovic et al., 2003). We observed 162 XLs to residues in MBD^{IDR}, exclusively in the N-terminal half of the sequence (Figure S4C). Ninety-two of these XLs were to structured portions of HDAC, MTA^{N}, GATAD2^{CC1}, MBD^{CC}, and MBD^{MBD} (Figure 2D), suggesting that MBD^{IDR} forms an irregular but ordered structure by packing against the MTA-HDAC core. In line with this conclusion, the N-terminal part of this region can immunoprecipitate HDAC2, MTA2, and RBBP4 (Desai et al., 2015).

One hundred and seventy-seven XLs were observed within and between RBBP proteins and the C-terminal half of MTA (which encompasses the two RBBP binding sites R1 and R2 as well as a ~100-residue region that is predicted to be disordered; Figure S1A). Of these XLs, 63 out of 69 can be mapped within the crosslinker cutoff in crystal structures of the MTA1^{R1}-RBBP4 and MTA1^{R2}-RBBP4 subcomplexes (Alqarni et al., 2014; Millard et al., 2016; Figure S4B). An additional 7 XLs were between structured regions but not in published crystal structures, and we used HADDOCK, together with published structures, to generate a model of MTA^{C}-RBBP$_2$. However, although this model fitted within our published low-resolution EM map of the complex (Schmidberger et al., 2016), only 4 out of 7 XLs were satisfactorily mapped in the model (Figure S4D). This low percentage might indicate the presence of substantial dynamics in this region of the complex, which is in line with the prediction of substantial intrinsic disorder in this region of MTA.

Finally, a further 60 XLs were observed between the C-terminal half of MTA and our HADDOCK-modeled core complex. Again, the XLs highlight a surface (Figure 2E, yellow) that indicates the likely location of the MTA^{C}-RBBP$_2$ subcomplex and allows us to represent the general position taken up by the two MTA^{C}-RBBP$_2$ units in a full MBD-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$HDAC_2$-MTA_2-$RBBP_4$ complex (Figure 2F). The remaining portion of the complex, namely CHD and the C-terminal half of GATAD2, displayed very few inter-subunit XLs, meaning that their locations cannot be confidently modeled. This lack of XLs suggests that the CHD-GATAD2 subcomplex might also display considerable dynamics.

GATAD2 Controls the Asymmetry of the NuRD Complex

Our model indicates that there are two equivalent sites on the MTA-HDAC subcomplex that could accommodate an MBD subunit, raising the question as to why only one MBD is observed in the full NuRD complex. To investigate further, we performed DIA-MS on the MTA^N-HDAC-$MBD^{GATAD2CC}$ subcomplex. Unexpectedly, we derived a 2:2:2 stoichiometry (Figure 3A). This stoichiometry was corroborated by SEC-MALLS (Figure 3B), which yielded a molecular weight (MW) measurement of ~290 kDa, in close agreement with the 288-kDa mass predicted for a 2:2:2 complex and significantly higher than the mass of 248 kDa expected for a 2:2:1 complex. FLAG-tagged MBD3 can also immunoprecipitate untagged MBD3 when both are co-expressed with HDAC and MTA, confirming that there are two accessible MBD binding sites in the MTA^N-HDAC complex (Figure S3D).

Using negative-stain EM, we recorded 325 micrographs of the purified MTA^N-HDAC-$MBD^{GATAD2CC}$ complex (Figures 3C, S2C, and S6A) and generated 2D class averages that represent different orientations of the complex. The dataset was relatively homogeneous and, using 3D classification and refinement routines in RELION without imposing any symmetry, we obtained a low-resolution structural envelope (29 Å according to the FSC 0.143 criterion). The envelope has a bi-lobed structure with an approximate 2-fold symmetry (Figures 3D and 3E), which is in line with our stoichiometry data. The dimensions and shape of the model independently derived from our XLMS data (Figure 2) match the EM density envelope very well (Figures 3D and 3E). Only the MTA^{ZF} domain significantly protrudes from the map, which might be a function of the low resolution of our map or of domain dynamics.

Segmentation of the EM envelope according to occupancy by domains in our XLMS-derived model reveals regions of the map for which no model density is assigned. These are likely accommodated by regions of the sequence for which no structures are available, namely the ~100 C-terminal residues of HDAC and the 145-residue MBD^{IDR}. Encouragingly, the largest region of unassigned density lies in the position predicted for MBD^{IDR} by our XLMS data (Figures 2D and 3D).

Given the unexpected 2:2:2 stoichiometry of the MTA^N-HDAC-$MBD^{GATAD2CC}$ complex, we hypothesized that the GATAD2 subunit must prevent binding of a second MBD subunit in the larger NuDe and NuRD complexes. Indeed, when we co-expressed GATAD2A with MTA2, HDAC1, and MBD3 (the MTA-HDAC-MBD-GATAD2 complex; Figure S2A), we observed the expected 2:2:1:1 stoichiometry by DIA-MS (Figure 3A; Data S2). We therefore conclude that the GATAD2 subunit is responsible not only for recruiting CHD remodeling activity to the NuRD complex but also for controlling the architecture and asymmetry of the complex by permitting the recruitment of only a single MBD subunit.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RBBP Subunits Introduce Substantial Conformational Dynamics

To corroborate our emerging picture of how the structural modules of NuRD are progressively assembled into a functional complex, we next analyzed three other complexes by negative-stain EM. Datasets were collected for an MTA-HDAC-RBBP complex assembled from recombinantly expressed components, as well as for the NuDe and NuRD complexes isolated from MEL cells (Figures S2C and S6-S8). The 2D classes for MTA-HDAC-RBBP exhibited significant shape heterogeneity, even for particles of a similar apparent size (Figure 4A). Given the relative homogeneity of the MTA^{N}-HDAC-$MBD^{GATAD2CC}$ complex in comparison, we concluded that the MTA^{C}-$RBBP_2$ units exhibit considerable dynamics relative to the MTA^{N}-HDAC-$MBD^{GATAD2CC}$ complex. Cartoons in Figure 4A indicate how these 2D classes can be readily interpreted as comprising a central MTA^{N}-HDAC "bridge" (magenta and orange, Figure 3D) and two MTA^{C}-$RBBP_2$ "arms" (blue, Figure S4D) that can rotate relative to the MTA^{N}-HDAC core. The only difference between the cartoon models shown is the relative rotation of the MTA^{C}-$RBBP_2$ arms.

For the NuDe complex, 2D classes derived from 23,401 particles again showed significant shape heterogeneity (Figure S7), suggestive of conformational flexibility. We attempted to cautiously deconvolute this apparent variability by performing *ab initio* 3D reconstruction in CryoSPARC (Punjani et al., 2017), specifying multiple output models followed by several rounds of multi-model and non-uniform refinement (Figure S7). Using this approach, we were able to produce two maps (corresponding to 22% and 21 % of the particles) with nominal resolutions of ~20 Å. These maps are qualitatively similar to the MTA-HDAC-RBBP structure described above but appear to be distinguished by a twisting motion in one of the two arms of the complex (Figure 4B). This observation is consistent with the MTA-HDAC-RBBP analysis and again suggests that the MTA^{C}-$RBBP_2$ arms exhibit substantial dynamics relative to the deacetylase core complex.

A much larger dataset (202,986 particles from 1323 images) was collected for the NuRD complex. Although some 2D classes were consistent with the MTA-HDAC-RBBP and NuDe data, a significant proportion exhibited clear additional density in the middle of the complex (Figure S8). Following repeated rounds of *ab initio* multi-model reconstruction and non-uniform refinement in CryoSPARC, the majority of particles (55%) remained in a class that, when compared to the maps obtained for NuDe (Figures 4C and S8), features additional density adjacent to the MTA^{N}-HDAC bridge. This density likely corresponds to parts of the MBD-GATAD2-CHD4 module. Although the resolved density cannot accommodate the entire module, this was not surprising; in addition to evidence from our XLMS data pointing to this region having substantial conformational variability, a recent structure of a CHD4-nucleosome complex resolved only one-third of CHD4 (Farnung et al., 2020), pointing to significant flexibility even in the presence of substrate. Similarly, the majority of GATAD2 is predicted to be disordered. We therefore conclude that our data resolve the attachment point of the remodeling module but are unable to resolve a complete envelope for this region of the complex.

Overall, these data outline the architecture of the NuRD complex and suggest that the complex undergoes substantial conformational dynamics.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

PWWP2A Competes Directly with MBD-GATAD2-CHD for Binding to the MTA-HDAC-RBBP Module of NuRD

Our data demonstrate that the MTA-MBD interface is the nexus between the symmetry-mismatched deacetylase and remodeling modules of NuRD. This finding raises the question as to whether either module has a physiological role outside the NuRD complex. We and others recently demonstrated that the transcriptional regulator PWWP2A selectively binds MTA-HDAC-RBBP but not intact NuRD (Link et al., 2018; Zhang et al., 2018). To understand the basis for this selectivity, we first expressed FLAG-PWWP2A with combinations of HDAC1, MTA1, MBD3, and GATAD2B and carried out pull-downs. As expected, FLAG-PWWP2A pulled down MTA1 and HDAC1 (Figure 5A, lane 5). Strikingly, however, the addition of either MBD3 or a combination of MBD3 and GATAD2B (lanes 4 and 6, respectively) made no difference—only HDAC1 and MTA1 were retained by PWWP2A. These data indicate that PWWP2A competes directly with MBD for exclusive binding to the deacetylase module of NuRD.

We next collected XLMS data on a PWWP2A-MTA-HDAC-RBBP complex (Data S1). A total of 146 unique XLs were observed between PWWP2A and either HDAC or MTA^N. Figure 5B shows that the set of residues in MTA1 and HDAC1 that cross-link to PWWP2A closely matches those that cross-link to MBD in the MTA^N-HDAC-$MBD^{GATAD2CC}$ complex. These data together support the conclusion that PWWP2A directly competes with MBD for binding to a common surface on the MTA-HDAC core complex. In an *in vivo* context, this provides a mechanism by which PWWP2A can recruit the deacetylase activity of NuRD to specific genomic loci without co-recruiting the remodeling activity of the CHD subunit.

DISCUSSION

NuRD Assembly Is Regulated at an Interface between Separable Histone Deacetylase and DNA Translocase Units

Our data define the subunit stoichiometry of the mammalian NuRD complex, elucidate key features of NuRD architecture, and demonstrate that the NuRD complex is built from two separable structural entities that have distinct catalytic activities. One of these entities, namely MTA-HDAC-RBBP, has two-fold symmetry, can be readily purified, and harbors the histone deacetylase activity. In contrast, the other entity, namely MBD-GATAD2-CHD, has a 1:1:1 stoichiometry, is inherently asymmetric, and contributes the ATP-dependent remodeling activity.

A central and unresolved question regarding NuRD architecture has been how the symmetric 2:2:4 MTA-HDAC-RBBP module and the asymmetric 1:1:1 MBD-GATAD2-CHD unit are connected. Our data show that MBD alone is insufficient to “break the symmetry” of MTA-HDAC-RBBP and it is the recruitment of GATAD2 that leads to asymmetry in the full complex. GATAD2 also recruits a CHD subunit to the complex, via the GATA-type zinc finger of GATAD2 (Sher et al., 2019; Torrado et al., 2017). We do not observe GATAD2 binding to MTA, HDAC, or RBBP (Torrado et al., 2017), and thus, the means by which GATAD2 prevents a second copy of MBD from binding to MTA-HDAC-RBBP remains

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

unclear. Binding of GATAD2 to one MBD might sterically block access to a second MTA subunit without involving a direct GATAD2-MTA interaction.

We note in passing that although CDK2AP1 (DOC1) has been consistently identified in NuRD preparations from HeLa cells (Kloet et al., 2014; Smits et al., 2013; Spruijt et al., 2010, 2016), we only identified it in one-third of our NuRD preparations (n = 12; data not shown). We made a similar observation in our purification of NuRD from three cancer cell lines (Sharifi Tabar et al., 2019), and it is possible that this subunit is either relatively labile or plays a tissue-specific role. As a result, we did not include CDK2AP1 in our analyses. Nonetheless, CDK2AP1 abundance does track closely with CHDs and moderately with GATAD2 (Sharifi Tabar et al., 2019), suggesting that CDK2AP1 might directly interact with CHDs in the asymmetric remodeling module of NuRD.

The modular nature of the NuRD complex is functionally relevant. We have shown recently that the coregulator PWWP2A can selectively bind the MTA1-HDAC-RBBP module and likely deacetylates H3K27 and H2A.Z (Link et al., 2018). Here, we demonstrate that the PWWP2A-MTA1-HDAC-RBBP interaction effectively rejects the MBD-GATAD2-CHD module by competing directly with MBD for binding to a common surface on MTA. Recent work on PWWP2A/B suggests that two copies of PWWP2A/B interact with MTA-HDAC-RBBP (Zhang et al., 2018), mirroring the 2:2:2 stoichiometry we see for MTA-HDAC-MBD. Thus, the MTA-HDAC-RBBP module appears to have a cellular function independent of the intact NuRD complex, with the MTA subunit acting as a regulatory control point for competitive interactions that can lead to NuRD-independent activities for NuRD subcomplexes.

Our stoichiometry and XLMS data, combined with published findings, allow us to present a structural model for the region of the NuRD complex comprising MTA^{N}, HDAC, RBBP, MBD, and the MBD^{CC}-$GATAD2^{CC1}$ coiled-coil domain. The model, which is independently consistent with our low-resolution EM data, provides a picture of mammalian NuRD architecture and delineates the point of connection between the deacetylase and remodeling modules, as well as other interacting partners.

The NuRD Complex Is Dynamic on Multiple Levels

Our EM data indicate that the MTA-HDAC-RBBP, NuDe, and NuRD complexes exhibit substantial conformational heterogeneity, indicating that the RBBP subunits are a major source of structural dynamics. Negative-stain EM data do carry a degree of ambiguity because of difficulties disentangling stain heterogeneity and dehydration artifacts from genuine dynamics and biological heterogeneity. However, our multiple EM datasets are not only internally consistent (i.e., maps of successively built-up subcomplexes demonstrate a degree of consistency) but also agree with the architectural models of NuRD that we derive independently from XLMS, DIA-MS, and existing high-resolution structures of NuRD subunits and subcomplexes. The EM data thus provide orthogonal and complementary evidence for the proposed architecture, while also supporting the conclusions of substantial dynamics suggested by other data.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The model shown in Figure 6A indicates the likely extent of motion exhibited by the MTA^{C}-$RBBP_2$ units. This mobility could have a role in docking the NuRD complex (or the competing PWWP2A-MTA-HDAC-RBBP complex; Figure 6B) onto nucleosomes by binding to histone H3 (Millard et al., 2016), in preparation for deacetylation and/or remodeling. RBBPs also work as adaptors for interactions with transcriptional regulators (e.g., FOG1, Lejon et al., 2011; ZNF827, Yang et al., 2018; and PHF6, Liu et al., 2014), which could help guide the NuRD complex to specific genomic sites.

Considering that CHD and GATAD2 subunits together make up 25% of the mass of the NuRD complex, the lack of inter-subunit XLs to CHD and the C-terminal half of GATAD2 suggests that these subunits are even more dynamic than the RBBPs relative to the HDAC-MTA core. This conclusion is in accord with our EM data for the intact NuRD complex, which showed a smaller than expected increase in particle size compared to the subcomplexes. Furthermore, the recent structure of CHD4 bound to a nucleosome (Farnung et al., 2020) displayed density for only 35% of the CHD4 protein. Thus, our data demonstrate that interactions of CHD4 with other NuRD subunits do not cause significant dampening of these dynamics.

The NuRD complex also displays considerable compositional dynamics, and there are several examples of functional consequences of such dynamics. MBD2- and MBD3-specific NuRD complexes have distinct biochemical properties (Le Guezennec et al., 2006; Yu et al., 2019), and CHD3, CHD4, and CHD5 NuRD complexes operate at different stages of mammalian brain development (Nitarska et al., 2016). Furthermore, PWWP2A exhibits a preference for an MTA-HDAC-RBBP complex built on MTA1, rather than MTA2 (Link et al., 2018). Our DIA-MS data allowed us to interrogate the paralog composition of the NuRD complex, which has not been examined in detail to date. The NuRD complex that we isolate from MEL cells shows an approximate equal representation of the two GATAD2 paralogs and a distinct preference for HDAC1 (80%) and MTA2 (60%), and both MTA3 and MBD2 (≤10% each) are essentially absent from our complex (Figure S1C). Unlike MTA1 and MTA2, MTA3 can bind only one molecule of RBBP, suggesting that NuRD complexes with different stoichiometries might exist in different cellular or organismal contexts. In all cases except for GATAD2, these percentages do not reflect the transcript abundance in MEL cells, suggesting that some selective incorporation of specific paralogs most likely takes place. Considering that shifts in the composition of subunit paralogs can translate to changes in activity, DIA-MS measurements made in different cell types in the future will be informative in this regard. Finally, we note that our observation of both NuDe and NuRD complexes (both here and previously; Low et al., 2016) emphasizes the peripheral placement of CHD4 in the NuRD complex. This placement supports the idea that CHD4 can act in several different capacities in the cell: as part of the NuRD complex or as part of the recently described ChAHP complex that can regulate chromatin structure (Kaaij et al., 2019; Ostapcuk et al., 2018). There have also been reports suggesting that CHD4 can act in isolation (e.g., O'Shaughnessy and Hendrich, 2013 and references within), although they are less definitive.

In summary, we present an integrated structural and biochemical analysis of the mammalian NuRD complex, which reveals that the complex is a highly dynamic assembly made from

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

symmetric and asymmetric units that carry separable enzymatic activities. The interface between these units serves as a point of regulation in complex assembly, showing that direct competition for binding surfaces can give rise to variant complexes with distinct functions.

STAR★METHODS

RESOURCE AVAILABILITY

Lead Contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Joel P Mackay (joel.mackay@sydney.edu.au).

Materials Availability—Plasmids generated in this study will be made available on reasonable request.

Data and Code Availability—All mass spectrometry data that have been generated and analyzed during the current study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository. Accession numbers are PRIDE: PXD010110 for DIA-MS and PRIDE: PXD010111 for XLMS data. Electron microscopy maps presented here has been deposited in the EMDB repository with accession numbers MTA^{N}-HDAC-$MBD^{GATAD2CC}$ (EMD-21382), NuDe (EMD-22904, EMD-22905 and EMD-22906) and NuRD (EMD-22895). Raw EM datasets have been deposited into EMPIAR with accession numbers EMPIAR-10537 (NuRD), EMPIAR-10538 (NuDe) and EMPIAR-10539 (MTA^{N}-HDAC-$MBD^{GATAD2CC}$). These data are also available from the corresponding authors upon reasonable request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

HEK293 cell culture and transient protein expression—For both cell maintenance and transient protein expression, suspension-adapted HEK Expi293F™ cells (Thermo Fisher Scientific, Waltham, MA, USA) were grown in Expi293™ Expression Medium (Thermo Fisher Scientific) at 37°C with 5% CO_2 and horizontal orbital shaking at 130 rpm. For seed cultures, 10 mL volumes were maintained and cell densities were always kept below 4×10^6 cells/mL. For transient protein expression, cells were grown to a density of 2×10^6 cells/mL and plasmids were transfected into cells using linear 25-kDa polyethylenimine (PEI) (Polysciences, Warrington, PA, USA) at a 1:2 (wt/wt) ratio. Proteins were allowed to express for 65–72 h. Cells were harvested by centrifugation at 300 $g \times 5$ min, washed twice in PBS, frozen in liquid nitrogen and stored at −80°C until use.

Murine erythroleukemia (MEL) cell culture—Murine erythroleukemia (MEL) seed cultures were maintained at 10–20 mL in DMEM/F-12 nutrient mix (Thermo Fisher) and 5% (v/v) FBS at 37°C with 5% CO_2. For large-scale grow ups, cells were grown in 250-mL batches inoculated with 2×10^6 cells. The cells were harvested at a density of $\leq 1 \times 10^6$ cells/mL by centrifugation at 300 $g \times 5$ min. Cells were washed twice in PBS, frozen in liquid nitrogen and stored at −80°C until use.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

METHOD DETAILS

Expression and purification of NuRD subcomplexes—The transfection and transient overexpression of proteins in suspension HEK Expi293F™ cells are as described above. Plasmid combinations used were FLAG-MTA2, tagless HDAC1 and HA-RBBP7 to express the MTA-HDAC-RBBP subcomplex, and FLAG-MBD3^{GATAD2CC}, HA-MTA2^{N} and tagless HDAC1 for the MTA^{N}-HDAC-MBD^{GATAD2CC} subcomplex, and FLAG-MTA2, HA-MBD3 and tagless HDAC1 and HA-GATAD2A for the MTA-HDAC-MBD-GATAD2 subcomplex, FLAG-MTA1^{C} and HA-RBBP4 for the MTA^{C}-RBBP$_2$ subcomplex, and FLAG-PWWP2A, HA-MTA1, tagless HDAC1 and HA-RBBP4 for the PWWP2A-MTA-HDAC-RBBP complex. Plasmid combinations were the same for all experiments except for the MTA-HDAC-MBD-GATAD2 subcomplex: HA-MTA2^{N}, FLAG-MBD3 and tagless HDAC1 and HA-GATAD2B were used for cross-linking mass spectrometry.

For protein purification by FLAG-affinity pulldown, cells were resuspended in lysis buffer (50 mM Tris-HCl, 150–500 mM NaCl, 3 mM ATP, 3 mM $MgCl_2$, 1% (v/v) Triton X-100, 1 × cOmplete® EDTA-free protease inhibitor (Roche, Basel, Switzerland), 0.2 mM DTT, pH 7.9), sonicated, incubated on ice for 30 min to precipitate chromatin, and then clarified via centrifugation (≥16,000 *g*, 20 min, 4°C). Anti-FLAG Sepharose 4B beads (BiMake; pre-equilibrated with lysis buffer) was added to the cleared HEK cell lysate and the mixtures were incubated overnight at 4°C with orbital rotation. Post-incubation, the beads were washed with ‘wash’ buffer (50 mM Tris-HCl, 150–500 mM NaCl, 3 mM ATP, 3 mM $MgCl_2$, 0.5% (v/v) IGEPAL® CA630, 0.2 mM DTT, pH 7.5). Bound proteins were eluted (× 3) with ‘elution’ buffer (10 mM HEPES, 150 mM NaCl, 300 μg/mL 3 × FLAG peptide (MDYKDHDGDYKDHDIDYKDDDDK; APExBIO), pH 7.5) at 4°C. Selected elutions were pooled and concentrated using Amicon Ultra-2, 100 kDa MWCO devices (Merck Millipore), pre-blocked with 5% (v/v) Tween-20. Typical yields were 1–2 mg per liter of culture.

NuDe and NuRD purification—Samples of NuDe and NuRD were isolated by FOG1 affinity pulldown from cultured mouse erythroleukemia (MEL) cells as described in Low et al. (2016). Briefly, GST-FOG1(1–45) was expressed in *Escherichia coli* BL21(DE3) using standard IPTG induction. Cells were lysed via sonication in GST binding buffer (50 mM Tris, 500 mM NaCl, 0.1% (v/v) β-mercaptoethanol, 0.5 mM PMSF, 0.1 mg/mL lysozyme, 10 μg/mL DNase I, pH 7.5), clarified via centrifugation (≥ 16,000 *g*, 20 min, 4°C), and incubated with pre-equilibrated glutathione Sepharose 4B beads (GE Healthcare) for 1 h at 4°C. The beads were then washed in GST wash buffer (20 CV, 50 mM Tris, 500 mM NaCl, 1% (v/v) Triton X-100, 1 mM DTT, pH 7.5), and then in NuRD binding buffer (10 CV). The beads were then used for pulldown experiments below.

MEL cell nuclear extracts were prepared by incubating the thawed cell pellets with hypotonic lysis buffer (5 mL/g of cells; 10 mM HEPES-KOH, 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT, cOmplete® protease inhibitor, pH 7.9) for 20 min at 4°C. IGEPAL® CA630 was then added (final concentration of 0.6% (v/v)), incubated for 10 min, and the cell suspension was vortexed for 10 s. Nuclei were recovered via centrifugation at 3,300 *g* × 5 min. The nuclear pellet was gently washed once with lysis buffer (+0.6% (v/v) IGEPAL®

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

CA-630), resuspended in NuRD binding buffer (50 mM HEPES-KOH, 150 mM NaCl, 1% (v/v) Triton X-100,1 mM DTT, 1 × cOmplete® protease inhibitor (Roche), pH 7.4), sonicated, incubated on ice for 30 min to precipitate chromatin, clarified via centrifugation (≥ 16,000 *g*, 20 min, 4°C), and incubated with the FOG1-affinity resin (prepared above) overnight at 4°C. Post-incubation, the beads were washed with 20 CV of NuRD wash buffer 1 (50 mM HEPES-KOH, 500 mM NaCl, 1% (v/v) Triton X-100, 1 mM DTT, pH 7.4) and then 10 CV of NuRD wash buffer 2 (50 mM HEPES-KOH, 150 mM NaCl, 0.1% (v/v) Triton X-100, 1 mM DTT, pH 7.4). Captured proteins were then eluted (× 3; 50 mM reduced glutathione, 50 mM HEPES-KOH, 150 mM NaCl, 0.1% Triton X-100, 1 mM DTT, pH 8.0) for 30 min at 4°C.

Sucrose density gradient ultracentrifugation—Sucrose density gradients for both GraFix (Stark, 2010) and non-GraFix samples were performed essentially as described in Schmidberger et al. (2016). Either a 5.5 mL SW55 Ti or a 12 mL SW41 Ti Beckman Coulter rotor was used. Briefly, sucrose density gradients (2%–25% (w/v) for MTA^{N}-HDAC-$MBD^{GATAD2CC}$, MTA-HDAC-RBBP, MTA^{C}-$RBBP_2$, 5%–35% (w/v) for NuDe, NuRD and MTA-HDAC-MBD-GATAD2) in buffer 50 mM HEPES-KOH pH 8.2, 150 mM NaCl, 1 mM DTT were prepared as described in Stone, 1974; gradients were allowed to form by layering the low sucrose buffer on top of the high sucrose buffer, slowly tilted to a horizontal position for 4 h at room temperature, and stored vertically for at least 1 h at 4°C before use. For EM samples, 0.15% (v/v) glutaraldehyde was added in the high sucrose buffer as the cross-linking agent as per the GraFix protocol (Stark, 2010). Prior to ultracentrifugation, 400 or 600 μL of the low sucrose buffer was removed from the top of the gradients and a protein sample of up to 200 μL was then layered on top of the cushion and ultracentrifuged using a SW55 Ti or a SW41 Ti Beckman Coulter rotor (28,000 rpm or 33,000 rpm, 4°C, 18 h). The gradients were fractionated as 100 or 200 μL aliquots, collected from the top, prior to further downstream analyses. To each fraction from GraFix gradients, 1 M Tris pH 8.0 was added to a final concentration of 100 mM to deactivate the remaining free glutaraldehyde. Fractions were checked by SDS-PAGE to confirm successful fractionation. For XLMS and DIA-MS, glutaraldehyde was omitted.

Crosslinking-mass spectrometry (XLMS)—Post-sucrose gradient separation, fractions containing NuRD and NuDe complexes were pooled, buffer exchanged into sucrose-free buffers and concentrated using Vivaspin tricellulose acetate centrifugal filters (20 K MWCO; Sartorius). The centrifugal filters were pre-blocked with 0.1% (v/v) Tween-20. For each crosslinking experiment, ~10–15 μg of the complex at a concentration of ~0.1–0.15 mg/mL was used. For MTA-HDAC-RBBP, it is essentially the same as for NuRD and NuDe except that sucrose gradient separation was omitted.

For the MTA^{N}-HDAC-$MBD^{GATAD2CC}$, the MTA-HDAC-MBD-GATAD2 subcomplexes, and the PWWP2A-MTA-HDAC-RBBP complex, eluates were buffer exchanged into a buffer comprising 50 mM HEPES pH 7.5 and 150–250 mM NaCl using an Amicon Ultra 0.5 mL or 4 mL centrifugal filter (50 K MWCO; Merck Millipore). For each crosslinking experiment, ~10–20 μg of subcomplex at a concentration of 0.2–1 mg/mL was used.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Sample preparation for XLMS using disuccinimidyl suberate (DSS) and adipic acid dihydrazide (ADH) crosslinkers were essentially as described previously (Schmidberger et al., 2016). Crosslinking reactions using bissulfosuccinimidyl suberate (BS3) were performed essentially as described for DSS crosslinking with the exception that the stock was made in Milli-Q water instead of dimethylformamide. Briefly, DSS (H_{12}/D_{12}-DSS; 1:1 ratio, 25 mM stock solution in anhydrous dimethylformamide; Creative Molecules) or BS3 (50 mM stock solution in Milli-Q water; Thermo Fisher) were added to a final concentration of 0.5–1 mM and incubated for 30 min at 37°C with constant mixing. The excess DSS/BS3 were then quenched with either 100 mM NH_4HCO_3 (50 mM final concentration) and further incubated at 37°C for 20 min. For adipic acid dihydrazide (ADH) crosslinking, H_8/D_8-ADH (1:1 ratio, 100 mg/mL stock solution in 20 mM HEPES pH 7.4; Creative Molecules) and 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) (144 mg/mL stock solution in 20 mM HEPES pH 7.4; Sigma-Aldrich) were added to final concentration of ~8.3 mg/mL and 12 mg/mL, respectively. The sample was then incubated for 1.5 h at 37°C with constant mixing. The excess ADH/DMTMM were then removed by using the Zeba Spin Desalting gel filtration columns (7K MWCO; ThermoFisher Scientific) into 20 mM HEPES pH 7.5, 150 mM NaCl. All crosslinked samples were then dried in a vacuum centrifuge.

Dried, crosslinked samples were resuspended in 50 μL of 8 M urea, reduced (5 mM TCEP, 37°C, 30 min) and alkylated (10 mM iodoacetamide, 20 min, room temperature in the dark). The samples were then diluted to 6 M urea using 200 mM Tris-HCl pH 8, Trypsin/Lys-C mix (Promega) was added to an enzyme:substrate ratio of 1:25 (w/w), and the solution was incubated at 37°C for 4 h. Following this, the urea was diluted to 0.75 M using 50 mM Tris-HCl pH 8, additional Trypsin (Promega) was added at an enzyme:-substrate ratio of 1:50 (w/w), and the sample was then incubated at 37°C overnight (~16 h). Following overnight digestion, the samples were acidified with formic acid to 2% (v/v) and centrifuged at 16,000 *g* for 10 min. The supernatant was then desalted using 50-mg Sep-Pak tC18 cartridges (Waters), and eluted in 50:50:0.1 acetonitrile:water:formic acid (v/v/v), and dried in a vacuum centrifuge.

For size exclusion chromatography fractionation (SEC), the dried desalted peptides were resuspended in 150 μL of SEC mobile phase [acetonitrile:water:trifluoroacetic acid, 70:30:0.1 (v/v/v)] and separated on a Superdex Peptide HR 10/30 column. A flow rate of 0.5 mL/min was used and the separation was monitored by UV absorption at 215, 254 and 280 nm. Fractions were collected as 1-mL fractions. Based on the UV absorption traces, fractions of interest (retention volumes ~12-17 mL) for mass spectrometry were dried in a vacuum centrifuge.

For LC-MS/MS, peptides were resuspended in 3% (v/v) acetonitrile, 0.1% (v/v) formic acid and loaded onto a 20 cm × 75 μm inner diameter column packed in-house with 1.9 μm C18AQ particles (Dr Maisch GmbH HPLC) using an Easy nLC-1000 UHPLC (Proxeon) or a Dionex UltiMate 3000 UHPLC (ThermoFisher Scientific). Peptides were separated using a linear gradient of 5%–30% buffer B over 120–150 min at 200 nL min-1 at 55°C (buffer A consisted of 0.1% (v/v) formic acid; while buffer B was 80% (v/v) acetonitrile and 0.1% (v/v) formic acid). Mass analyses were performed using either a LTQ Orbitrap Velos

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Pro, Q-Exactive or Q-Exactive Plus mass spectrometer (ThermoFisher Scientific). Following each full-scan MS1 at 30,000 (Velos Pro) or 70,000 resolution at 200 *m/z* (300 – 1750 *m/z*; 1–3 × 10^6 AGC; 100 ms injection time), up to 10 most abundant precursor ions were selected for MS/MS (HCD; 17,500 resolution; 1 × 10^5 AGC; 60 ms injection time; 32 NCE; 2 *m/z* isolation window; 1.7 × 10^5 intensity threshold; minimum charge state of +3; dynamic exclusion of 20 s).

Peak lists were generated using the MSConvert tool (Chambers et al., 2012) and submitted to the database search program Mascot (Matrix Science). The data was searched with oxidation (M) and carbamidomethyl (C) as variable modifications using a precursor-ion and product-ion mass tolerance of ± 15 ppm and ± 0.02 Da, respectively. The enzyme specificity was set to trypsin with up to two missed cleavages allowed and all taxonomies in the UniProt database (Nov 2013 – Jul 2016; 541,762–551,705 entries) were searched. A decoy database of reversed sequences was used to estimate the false discovery rates. To be considered for further analysis, identified peptides had to be top-ranking and statistically significant ($p < 0.05$) according to the Mascot expect metric.

XLMS data for the MTA^C-RBBP complex were recorded previously (Schmidberger et al., 2016).

Analysis of XLMS data—Analysis of the XLMS data was performed with pLINK v2.3.5 (Yang et al., 2012). pLINK search parameters that differ from the default settings were as follows: Peptide mass between 600–10,000 Da and peptide length between 6–100 were considered, precursor mass tolerance ± 15 ppm, product-ion mass tolerance ± 20 ppm, variable modifications of oxidation (M) and carbamidomethyl (C), enzyme specificity of trypsin with up to three missed cleavages (excluding the site of crosslinking) per chain. BS3 and H_{12}/D_{12}-DSS crosslinker settings: crosslinking sites were Lys, Ser, Tyr, Thr and protein N terminus, isotope shift (H_{12}/D_{12}-DSS) 12.075 Da, xlink mass-shift 138.068 Da, monolink mass-shift 156.079 Da. For the H_8/D_8-ADH crosslinker settings: crosslinking sites were Asp, Glu and protein C terminus, isotope shift 8.050 Da, xlink mass-shift 138.090 Da, monolink mass-shift 156.100 Da. For the DMTMM crosslinker (side reaction in the ADH crosslinking) settings: crosslinking sites were Asp, Glu and protein C terminus to Lys and protein N terminus, xlink mass-shift – 18.0106 Da, monolink mass-shift 0 Da. The protein database used for searching consisted of all the NuRD components (CHD4, MTA1, MTA2, MTA3, HDAC1, HDAC2, HDAC3, GATAD2A, GATAD2B, RBBP4, RBBP7, MBD2, MBD3) plus the top 10 contaminants identified (from the Mascot search) in the samples. The default FDR of 5% was used. Only peptides with a precursor mass error of ≤ ± 10 ppm, E-value scores of ≤ $1x10^{-3}$, and with at least four fragment ions on both the alpha-and beta-chain each were retained for further analysis. XLs that were identified as possibly either inter-protein or intra-protein (due to isoform sequence similarities) were considered as intra-protein as the likelihood of intra-protein XLs is higher than inter-protein XLs.

We included and re-processed our data from Schmidberger et al. (2016) for this work using the above settings.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

For crosslinking schematics as shown in Figure S4A, the xVis webserver was used (Grimm et al., 2015).

For modeling, the list of XLs was further filtered to remove redundant hits and crosslinked intra-protein residues of ≤ 10 residues apart were discarded. This final non-redundant list and unfiltered pLINK search outputs can be found in Data S1. All mass spectrometry data and XLMS search results have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Vizcaino et al., 2016) with the dataset identifier PRIDE: PXD010111.

Peptide selection for MS-based quantification—For quantification with internal standard peptides, the selection of reliable and representative proteotypic peptides (PTPs) for the protein targets of interest is an important step. Hence, in the first instance, we performed standard LC-MS/MS shotgun proteomics experiments to characterize our purified NuRD samples. From this proteomics dataset, we selected PTPs based on a pre-determined set of criteria: (1) Reproducible and reliable detection of the peptide across different MS runs and biological replicates; (2) Absence of missed cleavages; (3) Absence of known post-translational modifications (PTMs) (by querying protein databases including PhosphoELM (Dinkel et al., 2011) and UniProt (The UniProt Consortium, 2017) and large scale proteomic datasets that had not been incorporated into UniProt (at the time of experimental design) (Chen et al., 2012; Choudhary et al., 2009; Lundby et al., 2012) or potential predicted PTM sites (GPS 2.1.1, (Xue et al., 2011)); (4) Hydrophobicity scores in the range 10–46 (extremely hydrophobic peptides could present solubility issues and aberrant chromatographic behavior); (5) Related to point (4), avoidance of series of hydrophobic residues (C, F, I, L, V, W, Y). A series of these residues can be difficult to synthesize; (6) Avoidance of extremely long or short peptides (9–10 residues is ideal); (7) Where possible, to avoid the following residues: methionine – unpredictable oxidation; cysteines – unpredictable oxidation but can be mitigated by reduction and alkylation of the samples; prolines – unusual elution profile due to isomerisation and predominant CID fragmentation pattern; N-terminal glutamines and glutamates – can undergo spontaneous modification to pyroglutamate; (8) Of less importance and where possible, to select for peptides with precursor ions < 1000 *m/z* (this is due to limitations of the instrument used for data acquisition). Using these criteria, we selected 2–3 PTPs per target protein and where paralogs are very similar (e.g., HDAC1/2 and RBBP4/7), an additional 1–3 ‘pseudo PTPs’ shared between the paralogs were also chosen. Selected peptides were synthesized with stable isotopes (carboxy-terminal Arg ($^{13}C_6$; $^{15}N_4$) or Lys ($^{13}C_6$; $^{15}N_2$)) as AQUA™ peptides. A total of 31 peptides were chosen (exact numbers are indicated in parentheses): CHD4 (3), GATAD2A (3), GATAD2B (3) MTA1 (3), MTA2 (3), MTA3 (2), MBD2 (2), MBD3 (3), HDAC1 (2), HDAC2 (2), RBBP4 (3) RBBP7 (3). For HDAC1/2, one AQUA peptide was used for each HDAC1 and HDAC2, plus one additional ‘shared’ peptide for both HDAC1/2. For RBBP4/7, three ‘shared’ AQUA peptides were used. The list of peptides used is available in Data S2. The linear intensity range of our selected peptides were determined as described below.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

DIA-MS and sample preparation—Selected PTPs were synthesized with stable isotopes (carboxy-terminal Arg ($^{13}C_6$; $^{15}N_4$) or Lys ($^{13}C_6$; $^{15}N_2$)) as Absolutely Quantified SpikeTidesTM TQL with the cleavable Qtag by JPT Peptide Technologies. Peptides were supplied as 5 × 1 nmol aliquots and each 1 nmol aliquot per PTP were resuspended with 20% (v/v) acetonitrile in 80% (v/v) 0.1 M NH_4HCO_3, pooled, aliquoted, freeze-dried and stored at −20°C. For each MS experiment, each aliquot was resuspended in 20% (v/v) acetonitrile in 80% (v/v) 0.1 M NH_4HCO_3, the required amount taken and the remaining excess peptide was then discarded.

Post-affinity chromatography and sucrose gradient fractionation, heavy-labeled synthetic peptides were added to the highly-purified NuRD or NuRD subcomplexes, and the samples were then freeze-dried. Once dried, the samples were then re-solubilised with 50 μL of 8 M urea dissolved in 50 mM Tris.Cl pH 8. The samples were then reduced (5 mM TCEP, 37°C, 30 min) and alkylated (10 mM iodoacetamide, room temperature in the dark, 20 min). The samples were then diluted to 6 M urea with 50 mM Tris.Cl pH 8 and Trypsin/Lys-C mix (Promega) was added to a final enzyme:substrate ratio of 1:20 (w/w). The sample was then incubated for 37°C for 4 h. Following this initial digestion step, the sample was further diluted to 0.75 M urea with 50 mM Tris.Cl pH 8 and additional Trypsin/Lys-C mix was added to a final enzyme:substrate ratio of 1:10 (w/w). The sample was then incubated at 37°C overnight for ~16 h. Following the overnight digestion, the samples were acidified with formic acid to a final concentration of 2% (v/v) and centrifuged (20,000 *g*, 10 min). The supernatant was then desalted using in-house made C18 stagetips.

All samples for stoichiometry analysis were performed using the same Q-Exactive mass spectrometer. In every MS experiment/injection, 100 fmol of each reference heavy synthetic peptide was injected, unless otherwise stated. For every biological replicate that was analyzed, 1–2 replicates of DDA runs and 3–5 replicates of DIA runs were performed.

For LC-MS/MS, peptides were resuspended in 3% (v/v) acetonitrile, 0.1% (v/v) formic acid and loaded onto a 40–50 cm × 75 μm inner diameter column packed in-house with 1.9 μm C18AQ particles (Dr Maisch GmbH HPLC) using an Easy nLC-1000 UHPLC (Proxeon). Peptides were separated using a linear gradient of 5%–30% buffer B over 120 min at 200 nL/min at 55°C (buffer A consisted of 0.1% (v/v) formic acid; while buffer B was 80% (v/v) acetonitrile and 0.1% (v/v) formic acid). For DDA: after each full-scan MS1 (R = 70,000 at 200 *m/z*, 300–1750 *m/z*; 3 × 10^6 AGC; 100 ms injection time), up to 20 most abundant precursor ions were selected for MS/MS (R = 35,000 at 200 *m/z*; 1 × 10^6 AGC; 120 ms injection time; 30 normalized collision energy; 2 *m/z* isolation window; 8.3 × 10^5 intensity threshold; minimum charge state of +2; dynamic exclusion of 60 s). For DIA: after each full-scan MS1 (R = 140,000 at 200 *m/z* (300–1600 *m/z*; 3 × 10^6 AGC; 120 ms injection time), 16 × 25 *m/z* isolations windows in the 425–825 *m/z* range were sequentially isolated and subjected to MS/MS (R = 17,500 at 200 *m/z*, 3 × 10^6 AGC; 60 ms injection time; 30 normalized collision energy). After the first eight 25 *m/z* isolations and MS/MS and before the last eight 25 *m/z* isolations, a full-scan MS1 was performed. 25 *m/z* isolation window placements were optimized in Skyline (MacLean et al., 2010) to result in an inclusion list starting at 437.9490 *m/z* with increments of 25.0114 *m/z*.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Assessing the linearity in MS of selected peptides—We constructed a standard curve to assess the linearity of our heavy-labeled synthetic peptides in our MS quantification assay. We prepared and analyzed a dilution series of heavy-labeled synthetic peptides with a fixed quantity of highly purified NuRD proteins as the sample matrix. A range of 10–300 fmol was used a starting point, which was increased to 1000 fmol for the RBBP4/7 peptides. The MS data was processed in Skyline (MacLean et al., 2010) as described below. To determine the degree of linearity, we utilized simple linear regression analysis and response factor plots (Kuzyk et al., 2009). From these experiments, 90 MS1 ions and 103 MS2 ions from 27 peptides (these numbers are doubled when considering heavy peptides as well) were found to have usable linear regions (Data S2), with linear regression R^2 values of ≥ 0.972 and a response factor of < ± 20% of the average.

DIA data processing—All DDA data were processed using Proteome Discoverer v1.4 and searched with Sequest HT against all taxonomies in the UniProt database (Nov 2013 – Apr 2015; 541,762 – 548,208 entries). The data were searched with oxidation (M) and carbamidomethyl (C) as variable modifications using a precursor-ion and product-ion mass tolerance of ± 20 ppm and ± 0.02 Da. All results were filtered using Percolator to a false discovery rate of 1%.

All DIA data were processed using Skyline v3.5.0.9319 (MacLean et al., 2010). Reference spectral libraries were built in Skyline with .msf files using the BiblioSpec algorithm (Frewen and MacCoss, 2007). Precursor and production extracted ion chromatograms were generated using extraction windows that were two-fold the full-width at half maximum for both MS1 and MS2 filtering. Ion-match tolerance was set to 0.055 *m/z*. For MS1 filtering, the first three isotopic peaks with charges +2 to +4 were included while for MS2, *b*- and *y*-type fragments ions with charges +1 to +3 were considered. To ensure correct peak identification and assignment, the following criteria had to be met: (i) co-elution of light and heavy peptides; (ii) averaged dot product values of ≥ 0.95 between peptide precursor ion isotope distribution intensities and theoretical intensities (idotp) for both the heavy and light peptides; (iii) averaged dot product values of ≥ 0.85 between library spectrum intensities and current observed intensities (dotp); (iv) averaged relative dot product values of ≥ 0.95 between the observed heavy and light ion intensities (rdotp); (v) coefficient of variation (CV) values of ≤ 20% for all dot product values; (vi) matching peak shape for precursor and product ions of both the heavy and light peptides. Individual peptide ions were quantified using the light-to-heavy area-under-the-curve ratio of the M, M+1 and M+2 peaks for MS1 and a manually curated set of *y*-ions for MS2. Curation was based on: (i) linearity as determined from the standard curve (Data S2); and (ii) consistent observation between technical and biological replicates. To ensure only reliable ions were used for peptide quantification, further filtering was performed: (i) only light-to-heavy area-under-the-curve ratios with CV values of ≤ 20% between technical replicates were accepted; and (ii) the calculated value for each ion type had to fall within its pre-determined linear range. Finally, peptides were quantified by using the weighted average ratios between the M, M+1 and M+2 ions for quantification at the MS1 level and the selected *y*-ions for quantification at the MS2 level. This calculation was done in Skyline. Excellent correlation was observed for quantification results between the MS1 and MS2 levels (R^2 = 0.95, Pearson r = 0.97, n =

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

221; Figure S1D); hence, we used both MS1- and MS2-derived data to inform us about the NuRD stoichiometry. All reported stoichiometry values in the text were median values from both MS1-and MS2-derived data. All processed data related to the DIA-MS experiments can be found in Data S2. All raw mass spectrometer and Skyline datafiles have also been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Vizcaino et al., 2016) with the dataset identifier PRIDE: PXD010110.

As we had multiple peptides to quantify each of our NuRD proteins, we were able to see that these peptides seldom agree with each other in terms of reporting the target protein's quantity. There are a number of possible reasons why a particular peptide might under-report the true peptide quantity (e.g., a protease-resistant endogenous peptide could lead to missed cleavages, or unknown PTMs on the endogenous peptide could alter the apparent peptide quantity). On the other hand, it was very unlikely that there would be instances of over-reporting. Thus, the commonly accepted solution of averaging multiple peptides for a target protein will result in under-reporting of the true quantity of the target protein. Hence, we opted to pick the highest quantified value for each protein instead – which should lead to a value that is as close to the true peptide quantity as possible.

Analysis of NuRD stoichiometry—To calculate NuRD stoichiometry, we first normalized all our quantified values from the DIA-MS analyses to the average value between MTAs and HDACs ((*MTA*1 + *MTA*2 + *MTA*3) + (*HDAC*1 + *HDAC*2) /2) then multiplied them by two. This yielded a value of ~2 for both MTAs and HDACs; it has been previously shown through X-ray crystallography that HDAC and MTA form a 2:2 complex (Millard et al., 2013) and subunit exchange data that show that the MTA and HDAC subunits are very stable in the NuRD complex (Kloet et al., 2014). Based on these reasons, it was deemed reasonable to make both MTA and HDAC the stoichiometric reference subunits.

SEC-MALLS experiments and data analysis—Due to the instability and polydispersity of the NuDe and NuRD complexes upon concentration, size exclusion chromatography coupled to multiangle laser light scattering (SEC-MALLS) runs were carried out under denaturing conditions. NuDe and NuRD were first purified on GraFix sucrose gradients. The fractions containing the purified complexes were pooled, buffer exchanged (50 mM HEPES, 150 mM NaCl, 100 mM Tris, 6 M guanidine HCl, pH 8) and concentrated using Amicon® Ultra, MWCO 100-kDa concentrators (Merck Millipore), pre-blocked with 5% (v/v) Tween-20. Chromatography runs were performed on a Superose 6 Increase 3.2/300 column (GE Healthcare) at 50 μL/min, in a buffer comprising 50 mM HEPES, 150 mM NaCl, 100 mM Tris, 2 M guanidine-HCl, pH 8. Similarly, both bovine thyroglobulin and bovine serum albumin (BSA) were purified by GraFix, and then subjected to SEC-MALLS under denaturing conditions. For untreated thyroglobulin and BSA samples, guanidine-HCl was omitted in all steps.

SEC-MALLS of the MTA^{N}-HDAC-$MBD^{GATAD2CC}$ complex was performed under non-denaturing conditions. The MTA^{N}-HDAC-$MBD^{GATAD2CC}$ sample used was eluted from FLAG-beads, after pulling down the complex from HEK cell lysate. Chromatography run was performed on a Superose® 12 10/300 GL (GE Healthcare) at 0.5 mL/min, in 10 mM HEPES, 75 mM NaCl, 0.2 mM DTT, pH 7.5.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

All data analyses were performed using the ASTRA® software (v6.1.1.17; Wyatt Technology). Both light scattering and UV data were used to calculate the molecular mass. Extinction coefficients (mL/(mg cm), at 280 nm) used were 1.148 for NuDe, 1.118 for NuRD, 1.044 for MTA^{N}-HDAC-$MBD^{GATAD2CC}$, 1.069 for bovine thyroglobulin and 0.614 for BSA. The BSA runs (untreated, 100 μL, 2 mg/mL) were used to align the UV and light scattering signals.

Glutaraldehyde, the fixative agent used in the GraFix protocol (Stark, 2010), forms homopolymers with a range of sizes and the expected mass gain as a result of glutaraldehyde crosslinking is unknown. To this end, we first calibrated the expected mass gain from crosslinking by using two separate protein standards (BSA and thyroglobulin; Figures S3B and S3C). We applied the following formula to the BSA and thyroglobulin standards: *Mass gain per lysine residue* = ((*grafixed mass*) −(*untreated mass*) /*number of lysine residues*). As there are 59 and 148 lysine residues in BSA and thyroglobulin, respectively, we arrived at the conclusion that the GraFix protocol added an average of ~540 Da per lysine residue (lysines are the major target of glutaraldehyde).

EM sample preparation—Post-sucrose density gradient separation, crosslinked fractions containing the complex of interest (based on the corresponding fractions collected from a non-crosslinked sucrose gradient run in parallel; Figure S2C) had the sucrose diluted to an approximate final concentration of 0.02% (w/v) by buffer exchange using an Amicon® Ultra, 100-kDa MWCO device (Merck Millipore), pre-blocked with 5% (v/v) Tween-20. The exchange buffer used was 10 mM HEPES-KOH pH 7.5, 75 mM NaCl, 0.3 mM DTT. The concentrated sample was centrifuged at 5,000 *g* for 5 min to remove possible aggregates and aliquots of the supernatant were frozen in liquid nitrogen and stored at −80°C. Final concentrations were in the range 3–30 ng/μL.

Negative stain single particle EM—Sample (5 μL) was applied to a glow-discharged, carbon-coated 400-mesh copper grid (GSCu400CC—ProSciTech). After 2 min incubation time, the grid was blotted and washed with ten drops of distilled water, blotting on a filter paper in between washes. The grid was then washed in a drop of 1% (w/v) uranyl acetate, blotted and subsequently incubated in another drop of 1% (w/v) uranyl acetate solution for 30 s, blotted and allowed to air dry at room temperature. Images were acquired using a Tecnai T12 TEM operated at 120 kV (MTA^{N}-HDAC-$MBD^{GATAD2CC}$, MTA-HDAC-RBBP and NuRD) or a Tecnai F30 TEM operated at 300 kV (NuDe), both equipped with a Direct Electron LC-1100 (4k × 4k) lens-coupled CCD camera. Images were recorded at a nominal magnification of 52,000 × , 59,000 × , or 67,000 × corresponding to an unbinned pixel size of 2.79, 2.4 or 2.17 Å at the specimen level, respectively. Defocus values ranged from −1 to −2.5 μm.

EM image processing and 3D reconstruction—All data processing was carried out using the SPHIRE (Moriya et al., 2017), RELION 2.0 (Scheres, 2012) and CryoSPARC (Punjani et al., 2017) software packages. A total of 325, 469, 400 and 1332 micrographs were recorded for MTA^{N}-HDAC-$MBD^{GATAD2CC}$, MTA-HDAC-RBBP, NuDe and NuRD samples, respectively. For each dataset, a subset of images was used for manual picking of around 1000 particles, subsequently used to generate templates for autopicking the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

complete datasets, with the exception of NuRD which was fully-picked manually. The total number of extracted particles was 49,216, 25,155, 34,346 and 506,064 for MTA^{N}-HDAC-$MBD^{GATAD2CC}$, MTA-HDAC-RBBP, NuDe and NuRD, respectively. Various unsupervised 2D classifications were performed in order to select out poor quality particles (e.g., obvious aggregates, small particles). 2D classification was primarily performed in SPHIRE as this provides the option to specify a target number of particles per class and avoids the pitfall of having too many particles aggregating in few classes that can in turn lead to over-averaging. Iterative Stable Alignment and Clustering (ISAC) was used during 2D classification.

3D classification and refinement of MTA^{N}-HDAC-$MBD^{GATAD2CC}$ was performed in RELION using a low pass filtered map derived from the crystal structure of the MTA1-HDAC1 dimer (PDB: 4BKX) with a low pass cut-off 60 Å. A single 3D class was produced containing 8,233 particles, and the final map was 3D refined and post-processed. The resolution of the final map was estimated at the post-processing stage using gold-standard Fourier shell correlation. All single particle image processing was performed with no symmetry assumed or imposed.

Multi-model *ab initio* 3D reconstruction and non-uniform refinement of NuDe and NuRD datasets was performed in CryoSPARC. For NuDe, two rounds of 2D classification into 50 classes were first performed, with poor quality classes removed after each iteration (26 and 17 bad classes respectively), with *ab initio* 3D reconstruction with three classes subsequently carried out. The best class, containing 14,019 particles was subjected to a second round of *ab initio* reconstruction with three classes, from which two classes were identified and processed further using the non-uniform refinement algorithm. A similar workflow was employed for analysis of the much larger NuRD dataset, with three rounds of 2D classification performed. The data were initially processed as two subsets, which were combined after the final 2D classification step, following which a final 2D classification was performed and 40 good classes retained. *Ab initio* reconstruction was trialled with 2, 3 and 5 classes, with the three-class reconstruction producing a single good class containing 126,149 particles, which were subjected to a second round of *ab initio* reconstruction with three classes. The vast majority of particles (111,503) remained in a single class with a much smaller percentage of particles (9,929) separating into a second class that yielded a map resembling NuDe. Extensive attempts to fractionate the major class further yielded maps with few if any discernible differences.

Interaction studies using pulldown assays—Pulldown assays and western blot analyses used for the MTA^{N}-HDAC-$MBD^{GATAD2CC}$ subcomplex (in Figure S3D) and PWWP2A interaction studies (in Figure 5) were performed essentially as described in Torrado et al. (2017). Briefly, after FLAG-affinity purification (as described above), protein samples were separated via SDS-PAGE using Bolt™ 4%–12% Bis-Tris Plus gels (Thermo Fisher Scientific) run in MES buffer at 165 V for 45 min. The gel-separated proteins were blotted onto PVDF membranes and probed with antibodies at the following concentrations: from Cell Signaling Technology: anti-HA-HRP (2999S, 1:40,000) and anti-HDAC1-HRP (59581S, 1:80,000); from Sigma-Aldrich: anti-FLAG-HRP (A8592, 1:80,000); from Abcam: anti-MBD3 (ab157464, 1:2,500); from Leinco Technologies: anti-rabbit-HRP (R115, 1:10,000). Blots were imaged using enhanced chemiluminescence and X-ray films.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Structural modeling—To model the MTA^{N}-HDAC-$MBD^{GATAD2CC}$ subcomplex, we combined the 3D structures available for NuRD subunits with our XLMS data. The structures used in our modeling were $MTA1^{ES}$-HDAC1 (PDB: 4BKX) and the MBD domain of MBD3 (PDB: 2MB7). We also constructed homology models using the SWISS-MODEL server (Waterhouse et al., 2018): MTA1 BAH domain was based on the structure of the Sir3 BAH domain (PDB: 2FVU; ~19% identity); MBD3-GATAD2A coiled-coil was based on the MBD2-GATAD2A structure (PDB: 2L2L; ~85% identity between MBD coiled-coils); and the MTA1 ZF domain was based on the C-terminal ZF of GATA1 (PDB: 2GAT; ~32% identity). Finally, a single α-helix was built in PyMOL to represent the predicted helical region for MTA1(334-354).

Modeling was performed using HADDOCK 2.4 (van Zundert et al., 2016) using the XLMS data as unambiguous restraints. Expected maximum Cα-Cα crosslinker distances applied were 26 Å for ADH, 30 Å for BS3/DSS and 14 Å for DMTMM 'zero-length' crosslinkers. Default run parameters were used with the addition of a distance restraint of 5 Å between the Cα atoms of $MTA1^{BAH}$-K164 and $MTA1^{ES}$-G165, as well as between $MTA1^{ES}$-T333 and $MTA1^{H}$-D344, in order to maintain an appropriate distance between these contiguous parts of MTA1.

The resulting model was analyzed in Chimera (Pettersen et al., 2004), using the XLinkAnalyzertool (Kosinski et al., 2015), and in PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). After building one copy of MTA^{N}-HDAC-$MBD^{GATAD2CC}$, we used two-fold symmetry to generate a second copy of $MBD^{GATAD2CC}$ and build a model for the 2:2:2 MTA^{N}-HDAC-$MBD^{GATAD2CC}$ complex.

HADDOCK modeling of the MTA^{C}-$RBBP_2$ unit was carried out using previously published XLs (Schmidberger et al., 2016) together with XLs from this work. Based on existing structures of MTA^{R1}-RBBP and MTA^{R2}-RBBP complexes (Figure S1B), we extended the length of the MTA1 fragment in the latter structure *in silico* by drawing on sequence similarity to the longer $MTA1^{R1}$. From the 71 XLs observed for the MTA^{C}-$RBBP_2$ unit, forty involved the region of MTA1 between R1 and R2 (residues 560–650), which is predicted to be disordered. Of the remaining XLs, three are unambiguous inter-subunit XLs between RBBP4 and RBBP7, demonstrating that 'mixed' complexes can form that contain more than one paralog of a subunit. Although we had a model for the MTA^{C}-$RBBP_2$ unit from our previously published work (Schmidberger et al., 2016), we performed new modeling here, taking into account the new $MTA1^{R1}$-RBBP4 structure (Millard et al., 2016). No atomic model of MTA^{C}-$RBBP_2$ was able to fulfil all the XLs.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of NuRD DIA-MS data—All quantification with regards to the DIA-MS data were performed with Skyline v3.5.0.9319 (MacLean et al., 2010). The main statistic used in Skyline that had a major impact on data analysis were the dot product between the observed peptide precursor ion isotope distribution intensities and theoretical intensities (idotp), and the dot product between the observed fragment ion intensity distributions and the library intensities (dotp). As described above, we applied a filter of ≥ 0.95 for idotp values and ≥ 0.85 for dotp values. Quantified MS signals were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

exported as .csv files. As an additional level of quality control, we calculated coefficient of variation (CV) values for technical replicates for all dot product values. CV is calculated using the following formula:). This was done within Microsoft Excel and we applied a filter that all CV values had to be ≤ 20%. We also calculated Pearson correlation coefficient values between our observed MS1 and MS2 DIA-MS data. This was performed in Graphpad Prism and presented in Figure S1D. The value of *n* in Figure S1D refers to the number of MS1/MS2 datapoints.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The work was funded by the following grants: National Health and Medical Research Council of Australia project grants APP1012161, APP1063301, and APP1126357 and a fellowship from the same organization to J.P.M. (APP1058916). NIH R01-GM098264 was awarded to D.C.W. TRR81/Project A15 was awarded to S.B.H. This research was facilitated by access to Sydney Mass Spectrometry, a core research facility at the University of Sydney.

REFERENCES

Alqarni SS, Murthy A, Zhang W, Przewloka MR, Silva AP, Watson AA, Lejon S, Pei XY, Smits AH, Kloet SL, et al. (2014). Insight into the architecture of the NuRD complex: structure of the RbAp48-MTA1 subcomplex. J. Biol. Chem 289, 21844–21855. [PubMed: 24920672]

Ayala R, Willhoft O, Aramayo RJ, Wilkinson M, McCormack EA, Ocloo L, Wigley DB, and Zhang X (2018). Structure and regulation of the human INO80-nucleosome complex. Nature 556, 391–395. [PubMed: 29643506]

Bode D, Yu L, Tate P, Pardo M, and Choudhary J (2016). Characterization of Two Distinct Nucleosome Remodeling and Deacetylase (NuRD) Complex Assemblies in Embryonic Stem Cells. Mol. Cell. Proteomics 15, 878–891. [PubMed: 26714524]

Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, et al. (2012). A crossplatform toolkit for mass spectrometry and proteomics. Nat. Biotechnol 30, 918–920. [PubMed: 23051804]

Chen Y, Zhao W, Yang JS, Cheng Z, Luo H, Lu Z, Tan M, Gu W, and Zhao Y (2012). Quantitative acetylome analysis reveals the roles of SIRT1 in regulating diverse substrates and cellular pathways. Mol. Cell. Proteomics 11, 1048–1062. [PubMed: 22826441]

Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, and Mann M (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science 325, 834–840. [PubMed: 19608861]

Cramer JM, Scarsdale JN, Walavalkar NM, Buchwald WA, Ginder GD, and Williams DC Jr. (2014). Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA. J. Biol. Chem 289, 1294–1302. [PubMed: 24307175]

Desai MA, Webb HD, Sinanan LM, Scarsdale JN, Walavalkar NM, Ginder GD, and Williams DC Jr. (2015). An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex. Nucleic Acids Res. 43, 3100–3113. [PubMed: 25753662]

Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ, and Diella F (2011). Phospho.ELM: a database of phosphorylation sites—update 2011. Nucleic Acids Res. 39, D261–D267. [PubMed: 21062810]

dos Santos RL, Tosti L, Radzisheuskaya A, Caballero IM, Kaji K, Hendrich B, and Silva JC (2014). MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. Cell Stem Cell 15, 102–110. [PubMed: 24835571]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Eustermann S, Schall K, Kostrewa D, Lakomek K, Strauss M, Moldt M, and Hopfner KP (2018). Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. Nature 556, 386–390. [PubMed: 29643509]

Farnung L, Vos SM, Wigge C, and Cramer P (2017). Nucleosome-Chd1 structure and implications for chromatin remodelling. Nature 550, 539–542. [PubMed: 29019976]

Farnung L, Ochmann M, and Cramer P (2020). Nucleosome-CHD4 chromatin remodeler structure maps human disease mutations. eLife 9, e56178. [PubMed: 32543371]

Frewen B, and MacCoss MJ (2007). Using BiblioSpec for creating and searching tandem MS peptide libraries. Curr. Protoc. Bioinformatics 20, 13.7.1–13.7.12.

Gerber SA, Rush J, Stemman O, Kirschner MW, and Gygi SP (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc. Natl. Acad. Sci. USA 700, 6940–6945.

Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, Bonner R, and Aebersold R (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell Proteomics 11, O111.016717.

Gnanapragasam MN, Scarsdale JN, Amaya ML, Webb HD, Desai MA, Walavalkar NM, Wang SZ, Zu Zhu S, Ginder GD, and Williams DC Jr. (2011). p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. Proc. Natl. Acad. Sci. USA 108, 7487–7492. [PubMed: 21490301]

Gómez-Del Arco P, Perdiguero E, Yunes-Leites PS, Acín-Pérez R, Zeini M, Garcia-Gomez A, Sreenivasan K, Jiménez-Alcázar M, Segalés J, López-Maderuelo D, et al. (2016). The Chromatin Remodeling Complex Chd4/NuRD Controls Striated Muscle Identity and Metabolic Homeostasis. Cell Metab. 23, 881–892. [PubMed: 27166947]

Grimm M, Zimniak T, Kahraman A, and Herzog F (2015). xVis: a web server for the schematic visualization and interpretation of crosslink-derived spatial restraints. Nucleic Acids Res. 43, W362–W369. [PubMed: 25956653]

Guo T, Luna A, Rajapakse VN, Koh CC, Wu Z, Liu W, Sun Y, Gao H, Menden MP, Xu C, et al. (2019). Quantitative Proteome Landscape of the NCI-60 Cancer Cell Lines. iScience 21, 664–680. [PubMed: 31733513]

Hong W, Nakazawa M, Chen YY, Kori R, Vakoc CR, Rakowski C, and Blobel GA (2005). FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. EMBO J. 24, 2367–2378. [PubMed: 15920470]

Kaaij LJT, Mohn F, van der Weide RH, de Wit E, and Buhler M (2019). The ChAHP Complex Counteracts Chromatin Looping at CTCF Sites that Emerged from SINE Expansions in Mouse. Cell 178, 1437–1451.e1414. [PubMed: 31491387]

Kloet SL, Baymaz HI, Makowski M, Groenewold V, Jansen PW, Berendsen M, Niazi H, Kops GJ, and Vermeulen M (2014). Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. FEBS J. 282, 1774–1785. [PubMed: 25123934]

Kosinski J, von Appen A, Ori A, Karius K, Müller CW, and Beck M (2015). Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. J. Struct. Biol 189, 177–183. [PubMed: 25661704]

Kuzyk MA, Smith D, Yang J, Cross TJ, Jackson AM, Hardie DB, Anderson NL, and Borchers CH (2009). Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. Mol. Cell. Proteomics 8, 1860–1877. [PubMed: 19411661]

Le Guezennec X, Vermeulen M, Brinkman AB, Hoeijmakers WA, Cohen A, Lasonder E, and Stunnenberg HG (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol. Cell. Biol 26, 843–851. [PubMed: 16428440]

Lejon S, Thong SY, Murthy A, AlQarni S, Murzina NV, Blobel GA, Laue ED, and Mackay JP (2011). Insights into association of the NuRD complex with FOG-1 from the crystal structure of an RbAp48· FOG-1 complex. J. Biol. Chem 286, 1196–1203. [PubMed: 21047798]

Link S, Spitzer RMM, Sana M, Torrado M, Völker-Albert MC, Keilhauer EC, Burgold T, Pünzeler S, Low JKK, Lindström I, et al. (2018). PWWP2A binds distinct chromatin moieties and interacts with an MTA1-specific core NuRD complex. Nat. Commun 9, 4300. [PubMed: 30327463]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Liu Z, Li F, Ruan K, Zhang J, Mei Y, Wu J, and Shi Y (2014). Structural and functional insights into the human Börjeson-Forssman-Lehmann syndrome-associated protein PHF6. J. Biol. Chem 289, 10069–10083. [PubMed: 24554700]

Liu Z, Li F, Zhang B, Li S, Wu J, and Shi Y (2015). Structural basis of plant homeodomain finger 6 (PHF6) recognition by the retinoblastoma binding protein 4 (RBBP4) component of the nucleosome remodeling and deacetylase (NuRD) complex. J. Biol. Chem 290, 6630–6638. [PubMed: 25601084]

Liu X, Li M, Xia X, Li X, and Chen Z (2017). Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. Nature 544, 440–445. [PubMed: 28424519]

Liu K, Xu C, Lei M, Yang A, Loppnau P, Hughes TR, and Min J (2018). Structural basis for the ability of MBD domains to bind methyl-CG and TG sites in DNA. J. Biol. Chem 293, 7344–7354. [PubMed: 29567833]

Low JK, Webb SR, Silva AP, Saathoff H, Ryan DP, Torrado M, Brofelth M, Parker BL, Shepherd NE, and Mackay JP (2016). CHD4 Is a Peripheral Component of the Nucleosome Remodeling and Deacetylase Complex. J. Biol. Chem 291, 15853–15866. [PubMed: 27235397]

Lundby A, Lage K, Weinert BT, Bekker-Jensen DB, Secher A, Skovgaard T, Kelstrup CD, Dmytriyev A, Choudhary C, Lundby C, and Olsen JV (2012). Proteomic analysis of lysine acetylation sites in rat tissues reveals organ specificity and subcellular patterns. Cell Rep. 2, 419–431. [PubMed: 22902405]

Luo J, Su F, Chen D, Shiloh A, and Gu W (2000). Deacetylation of p53 modulates its effect on cell growth and apoptosis. Nature 408, 377–381. [PubMed: 11099047]

MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, and MacCoss MJ (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26, 966–968. [PubMed: 20147306]

Millard CJ, Watson PJ, Celardo I, Gordiyenko Y, Cowley SM, Robinson CV, Fairall L, and Schwabe JW (2013). Class I HDACs share a common mechanism of regulation by inositol phosphates. Mol. Cell 51, 57–67. [PubMed: 23791785]

Millard CJ, Varma N, Saleh A, Morris K, Watson PJ, Bottrill AR, Fairall L, Smith CJ, and Schwabe JW (2016). The structure of the core NuRD repression complex provides insights into its interaction with chromatin. eLife 5, e13941. [PubMed: 27098840]

Moody RR, Lo MC, Meagher JL, Lin CC, Stevers NO, Tinsley SL, Jung I, Matvekas A, Stuckey JA, and Sun D (2018). Probing the interaction between the histone methyltransferase/deacetylase subunit RBBP4/7 and the transcription factor BCL11A in epigenetic complexes. J. Biol. Chem 293, 2125–2136. [PubMed: 29263092]

Moriya T, Saur M, Stabrin M, Merino F, Voicu H, Huang Z, Penczek PA, Raunser S, and Gatsogiannis C (2017). High-resolution Single Particle Analysis from Electron Cryo-microscopy Images Using SPHIRE. J. Vis Exp, 55448. 10.3791/55448.

Murzina NV, Pei XY, Zhang W, Sparkes M, Vicente-Garcia J, Pratap JV, McLaughlin SH, Ben-Shahar TR, Verreault A, Luisi BF, and Laue ED (2008). Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. Structure 16, 1077–1085. [PubMed: 18571423]

Nitarska J, Smith JG, Sherlock WT, Hillege MM, Nott A, Barshop WD, Vashisht AA, Wohlschlegel JA, Mitter R, and Riccio A (2016). A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. Cell Rep. 17, 1683–1698. [PubMed: 27806305]

O'Shaughnessy A, and Hendrich B (2013). CHD4 in the DNA-damage response and cell cycle progression: not so NuRDy now. Biochem. Soc. Trans 41, 777–782. [PubMed: 23697937]

Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, and Dunker AK (2003). Predicting intrinsic disorder from amino acid sequence. Proteins 53, 566–572. [PubMed: 14579347]

Ostapcuk V, Mohn F, Carl SH, Basters A, Hess D, Iesmantavicius V, Lampersberger L, Flemr M, Pandey A, Thomä NH, et al. (2018). Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes. Nature 557, 739–743. [PubMed: 29795351]

Pavlopoulos E, Jones S, Kosmidis S, Close M, Kim C, Kovalerchik O, Small SA, and Kandel ER (2013). Molecular mechanism for age-related memory loss: the histone-binding protein RbAp48. Sci. Transl. Med 5, 200ra115.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Pegoraro G, Kubben N, Wickert U, Göhler H, Hoffmann K, and Misteli T (2009). Ageing-related chromatin defects through loss of the NURD complex. Nat. Cell Biol 11, 1261–1267. [PubMed: 19734887]

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem 25, 1605–1612. [PubMed: 15264254]

Punjani A, Rubinstein JL, Fleet DJ, and Brubaker MA (2017). cryo-SPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296. [PubMed: 28165473]

Pünzeler S, Link S, Wagner G, Keilhauer EC, Kronbeck N, Spitzer RM, Leidescher S, Markaki Y, Mentele E, Regnard C, et al. (2017). Multivalent binding of PWWP2A to H2A.Z regulates mitosis and neural crest differentiation. EMBO J. 36, 2263–2279. [PubMed: 28645917]

Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour AA, Caspi I, Krupalnik V, Zerbib M, et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. Nature 502, 65–70. [PubMed: 24048479]

Riley LG, Ralston GB, and Weiss AS (1996). Multimer formation as a consequence of separate homodimerization domains: the human c-Jun leucine zipper is a transplantable dimerization module. Protein Eng. 9, 223–230. [PubMed: 9005444]

Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, and Orkin SH (2008). Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science 322, 1839–1842. [PubMed: 19056937]

Scheres SH (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol 180, 519–530. [PubMed: 23000701]

Schmidberger JW, Sharifi Tabar M, Torrado M, Silva AP, Landsberg MJ, Brillault L, AlQarni S, Zeng YC, Parker BL, Low JK, and Mackay JP (2016).The MTA1 subunit of the nucleosome remodeling and deacetylase complex can recruit two copies of RBBP4/7. Protein Sci. 25, 1472–1482. [PubMed: 27144666]

Sharifi Tabar M, Mackay JP, and Low JKK (2019). The stoichiometry and interactome of the Nucleosome Remodeling and Deacetylase (NuRD) complex are conserved across multiple cell lines. FEBS J. 286, 2043–2061. [PubMed: 30828972]

Sher F, Hossain M, Seruggia D, Schoonenberg VAC, Yao Q, Cifani P, Dassama LMK, Cole MA, Ren C, Vinjamur DS, et al. (2019). Rational targeting of a NuRD subcomplex guided by comprehensive in situ mutagenesis. Nat. Genet 51, 1149–1159. [PubMed: 31253978]

Smits AH, Jansen PW, Poser I, Hyman AA, and Vermeulen M (2013). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res. 41, e28. [PubMed: 23066101]

Spruijt CG, Bartels SJJ, Brinkman AB, Tjeertes JV, Poser I, Stunnenberg HG, and Vermeulen M (2010). CDK2AP1/DOC-1 is a bona fide subunit of the Mi-2/NuRD complex. Mol. Biosyst 6, 1700–1706. [PubMed: 20523938]

Spruijt CG, Luijsterburg MS, Menafra R, Lindeboom RGH, Jansen PWTC, Edupuganti RR, Baltissen MP, Wiegant WW, Voelker-Albert MC, Matarese F, et al. (2016). ZMYND8 Co-localizes with NuRD on Target Genes and Regulates Poly(ADP-Ribose)-Dependent Recruitment of GATAD2A/NuRD to Sites of DNA Damage. Cell Rep. 17, 783–798. [PubMed: 27732854]

Stark H (2010). GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. Methods Enzymol. 481, 109–126. [PubMed: 20887855]

Stone AB (1974). A simplified method for preparing sucrose gradients. The Biochemical journal 137, 117–118. [PubMed: 4595281]

Tabackman AA, Frankson R, Marsan ES, Perry K, and Cole KE (2016). Structure of 'linkerless' hydroxamic acid inhibitor-HDAC8 complex confirms the formation of an isoform-specific subpocket. J. Struct. Biol 195, 373–378. [PubMed: 27374062]

The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169. [PubMed: 27899622]

Toh Y, and Nicolson GL (2009). The role of the MTA family and their encoded proteins in human cancers: molecular functions and clinical implications. Clin. Exp. Metastasis 26, 215–227. [PubMed: 19116762]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Torrado M, Low JKK, Silva APG, Schmidberger JW, Sana M, Sharifi Tabar M, Isilak ME, Winning CS, Kwong C, Bedward MJ, et al. (2017). Refinement of the subunit interaction network within the nucleosome remodelling and deacetylase (NuRD) complex. FEBS J. 284, 4216–4232. [PubMed: 29063705]

van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, van Dijk M, de Vries SJ, and Bonvin AMJJ (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J. Mol. Biol 428, 720–725. [PubMed: 26410586]

Vizcaino JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, et al. (2016). 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. 44, D447–D456. [PubMed: 26527722]

Walavalkar NM, Gordon N, and Williams DC Jr. (2013). Unique features of the anti-parallel, heterodimeric coiled-coil interaction between methyl-cytosine binding domain 2 (MBD2) homologues and GATA zinc finger domain containing 2A (GATAD2A/p66α). J. Biol. Chem 288, 3419–3427. [PubMed: 23239876]

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303. [PubMed: 29788355]

Willhoft O, Ghoneim M, Lin CL, Chua EYD, Wilkinson M, Chaban Y, Ayala R, McCormack EA, Ocloo L, Rueda DS, and Wigley DB (2018). Structure and dynamics of the yeast SWR1-nucleosome complex. Science 362, 199.

Xu J, Bauer DE, Kerenyi MA, Vo TD, Hou S, Hsu YJ, Yao H, Trowbridge JJ, Mandel G, and Orkin SH (2013). Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. Proc. Natl. Acad. Sci. USA 110, 6518–6523. [PubMed: 23576758]

Xue Y, Liu Z, Cao J, Ma Q, Gao X, Wang Q, Jin C, Zhou Y, Wen L, and Ren J (2011). GPS 2.1: enhanced prediction of kinase-specific phosphorylation sites with an algorithm of motif length selection. Protein Eng. Des. Sel 24, 255–260. [PubMed: 21062758]

Yang B, Wu YJ, Zhu M, Fan SB, Lin J, Zhang K, Li S, Chi H, Li YX, Chen HF, et al. (2012). Identification of cross-linked peptides from complex samples. Nat. Methods 9, 904–906. [PubMed: 22772728]

Yang SF, Sun AA, Shi Y, Li F, and Pickett HA (2018). Structural and functional characterization of the RBBP4-ZNF827 interaction and its role in NuRD recruitment to telomeres. Biochem. J 475, 2667–2679. [PubMed: 30045876]

Yildirim O, Li R, Hung J-H, Chen PB, Dong X, Ee LS, Weng Z, Rando OJ, and Fazzio TG (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell 147, 1498–1510. [PubMed: 22196727]

Yoshida T, Hazan I, Zhang J, Ng SY, Naito T, Snippert HJ, Heller EJ, Qi X, Lawton LN, Williams CJ, and Georgopoulos K (2008). The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation. Genes Dev. 22, 1174–1189. [PubMed: 18451107]

Yu X, Azzo A, Bilinovich SM, Li X, Dozmorov M, Kurita R, Nakamura Y, Williams DC Jr., and Ginder GD (2019). Disruption of the MBD2-NuRD complex but not MBD3-NuRD induces high level HbF expression in human adult erythroid cells. Haematologica 104, 2361–2371. [PubMed: 31004025]

Zhang W, Aubert A, Gomez de Segura JM, Karuppasamy M, Basu S, Murthy AS, Diamante A, Drury TA, Balmer J, Cramard J, et al. (2016). The Nucleosome Remodeling and Deacetylase Complex NuRD Is Built from Preformed Catalytically Active Sub-modules. J. Mol. Biol 428, 2931–2942. [PubMed: 27117189]

Zhang T, Wei G, Millard CJ, Fischer R, Konietzny R, Kessler BM, Schwabe JWR, and Brockdorff N (2018). A variant NuRD complex containing PWWP2A/B excludes MBD2/3 to regulate transcription at active genes. Nat. Commun 9, 3798. [PubMed: 30228260]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

- The NuRD complex has a 4:2:2:1:1:1 (RBBP:HDAC:MTA:MBD:GATAD2:CHD) stoichiometry
- The complex is built from HDAC:MTA:RBBP and MBD:GATAD2:CHD modules
- The NuRD complex demonstrates significant large-scale dynamics
- PWWP2A competes with MBD3 to bind HDAC:MTA:RBBP, forming a PWWP2A:HDAC:MTA:RBBP complex

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

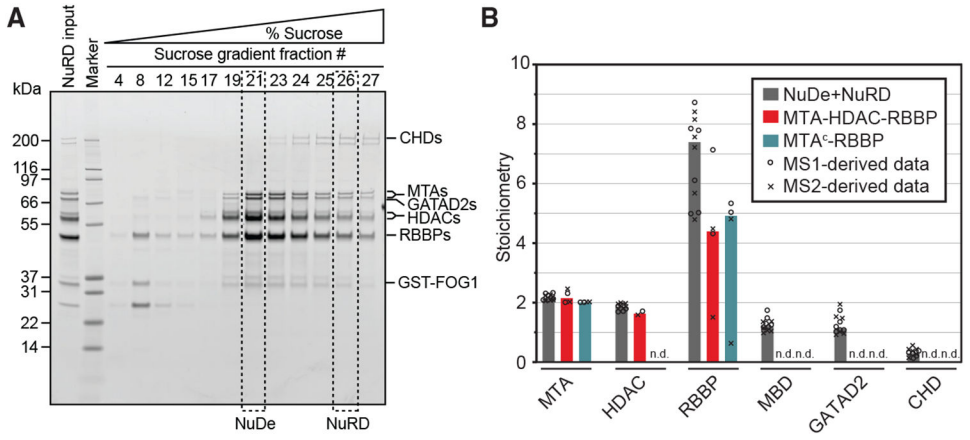


Figure 1. Absolute Quantification of NuRD Subunit Stoichiometry by Mass Spectrometry

(A) SDS-PAGE showing the NuRD complex after FOG1 pull-down (*NuRD input*) and showing fractions enriched for NuDe and NuRD following subsequent sucrose-gradient centrifugation. MBD proteins do not stain as well as other subunits.

(B) NuDe + NuRD, MTA-HDAC-RBBP, and MTA^{C}-RBBP subunit stoichiometry derived from DIA-MS data. Numbers are shown relative to the MTA/HDAC average, which is set to 2 of each based on the $MTA1^{ES}$-HDAC1 crystal structure (Millard et al., 2013). Open circles and crosses indicate the individual data points from each biological replicate, derived from MS1 and MS2, respectively. The bars show the median value, n.d., null value. See also Figures S1 and S2 and Data S2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

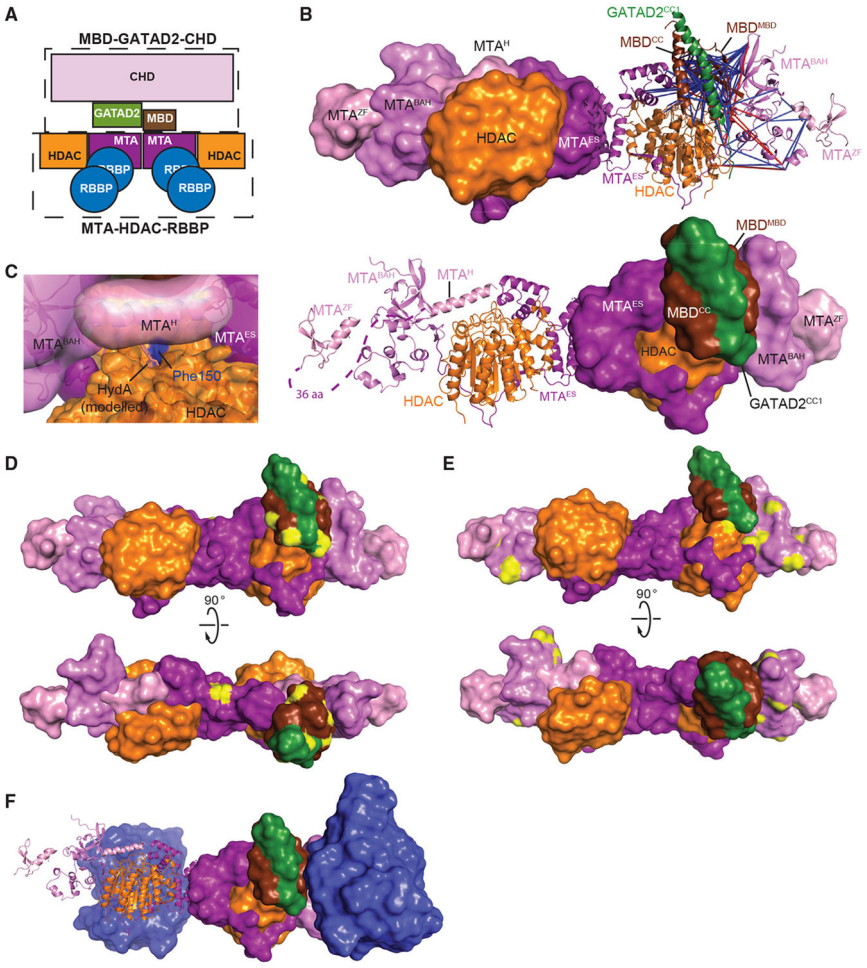


Figure 2. XLMS Data for NuRD and NuRD Subcomplexes

(A) Schematic representation of NuRD topology deduced from this work and previous studies. The two enzymatic modules (MTA-HDAC-RBBP and MBD-GATAD2-CHD) are indicated by dashed boxes. The MBD subunit bridges the two modules of the NuRD complex.

(B) Two representations of a molecular model derived from XL-driven rigid-body docking in HADDOCK. The top representation shows the XLs used to drive the docking process. Domains used in the model include the MTA1^{ES}-HDAC dimer of heterodimers, the MBD domain of MBD3 (MBD^{MBD}), a model of the MTA1 BAH domain (MTA1^{BAH}), a model of the MTA1 GATA-type zinc-finger domain (MTA^{ZF}), a model of the predicted MTA1 helix (MTA^{H}), and the heterodimeric coiled-coil formed by MBD3 (MBD^{CC}) and GATAD2A (GATAD2^{CC1}). XLs satisfied (blue) and not satisfied (red) in this model are shown (top).

(C) Relative positions of Phe150 (blue) in the active site of HDAC1, the MTA^{H}, and a modeled hydroxamic acid (HydA) HDAC inhibitor. The tight space formed between the MTA^{H} and HDAC1 suggests that MTA1 may restrict and modulate access to the HDAC active site.

(D) Residues forming XLs between the MBD3 intrinsically disordered region (MBD^{IDR}) and the MTA^{N}-HDAC-MBD-GATAD2 core. The positions of the cross-linked residues on

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the MTA^{N}-HDAC-MBD-GATAD2 core complex (yellow) point to the likely position of MBD^{IDR}.

(E) Residues forming XLs between the MTA^{C}-$RBBP_2$ and the MTA^{N}-HDAC-MBD-GATAD2 core. The positions of the cross-linked residues on the core complex (yellow) provide clues to the approximate position of the MTA^{C}-$RBBP_2$ unit.

(F) An XL-based model showing the likely position of the two MTA^{C}-$RBBP_2$ units relative to the MTA^{N}-HDAC-MBD-GATAD2 core. The model was created by manually positioning the two MTA^{C}-$RBBP_2$ units. See also Figures S4 and S5 and Data S1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

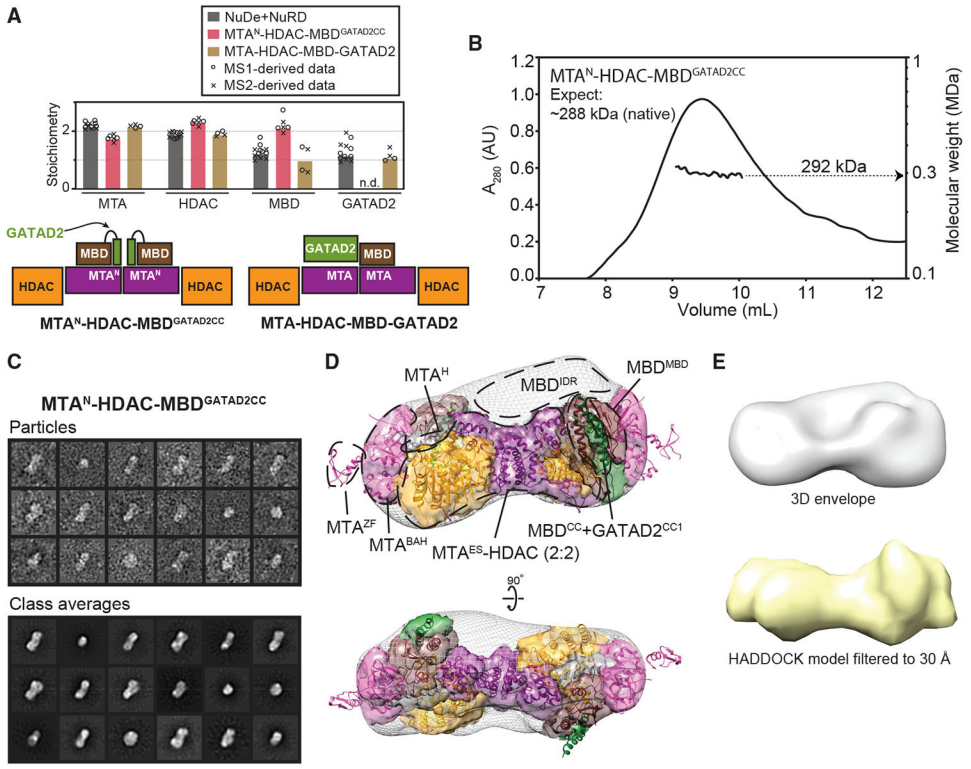


Figure 3. Structure of the MTA^{N}-HDAC-MBD^{GATAD2CC} Complex

(A) Top: subunit stoichiometry derived from DIA-MS data for MTA^{N}-HDAC-MBD^{GATAD2CC} and MTA-HDAC-MBD-GATAD2 complexes. Data for NuDe+NuRD from Figure 1B are shown for comparison. Numbers are shown relative to the MTA/HDAC average, which is set to 2 of each based on the MTA1^{ES}-HDAC1 crystal structure (Millard et al., 2013). Open circles and crosses indicate the individual data points, from each biological replicate, derived from MS1 and MS2, respectively. The bars show the median value. Bottom: schematics of MTA^{N}-HDAC-MBD^{GATAD2CC} and MTA-HDAC-MBD-GATAD2.

(B) SEC-MALLS data for MTA^{N}-HDAC-MBD^{GATAD2CC} under non-denaturing conditions.

(C) Selected particles and reference-free 2D class averages for MTA^{N}-HDAC-MBD^{GATAD2CC}.

(D) Final 3D envelope for MTA^{N}-HDAC-MBD^{GATAD2CC} refined to 29-Å resolution. Also shown is the molecular model of the complex adapted from Figure 2B to include a second molecule of MBD^{GATAD2CC} with 2-fold symmetry. Subunit colors are as in Figure 2. The wire mesh indicates regions of the map for which no model density is currently assigned; these regions could accommodate the C-terminal ~100 residues of HDAC and ~145 residues of MBD3^{IDR}, for which no structure is available.

(E) Overall shape comparison between the final 3D envelope for MTA^{N}-HDAC-MBD^{GATAD2CC} refined to 29-Å resolution (gray) and the generated map for the XLMS-derived HADDOCK model without the MTA^{ZF} domain, filtered to 30-Å resolution (yellow).

See also Figures 2, S3, S4, and S6 and Data 2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

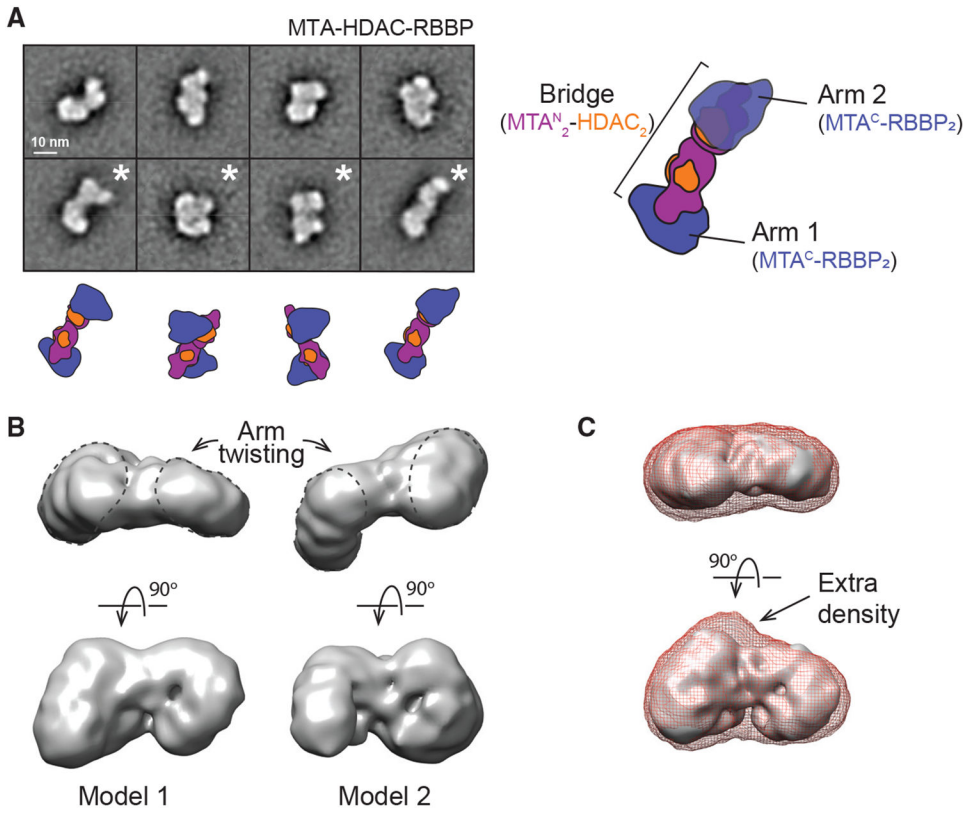


Figure 4. RBBP-Containing Complexes Display Significant Shape Heterogeneity

(A) Top left: 2D classes for the MTA-HDAC-RBBP complex highlight the inherent conformational variability that results from adding the RBBP subunits to the complex. Bottom left and right: cartoons corresponding to the indicated 2D classes, which indicate that the observed conformational variability is readily interpreted as being a consequence of motion in the MTA^C-RBBP$_2$ arms, relative to the central (MTA^N-HDAC)$_2$ bridge. Scale bar representing 10 nm has been included.

(B) The two dominant 3D models of NuDe derived from multi-model *ab initio* reconstruction in CryoSPARC emphasize this flexibility further, highlighting clear twisting motions in the arm regions that distinguish the two models. Likely locations of the MTA^C-RBBP$_2$ arms are indicated (gray dashed lines).

(C) Comparison of NuDe model 1 from (B) with a 3D NuRD model derived from the same reconstruction approach. Additional density near the (MTA^N-HDAC)$_2$ bridge is indicated. See also Figures S6, S7, and S8.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

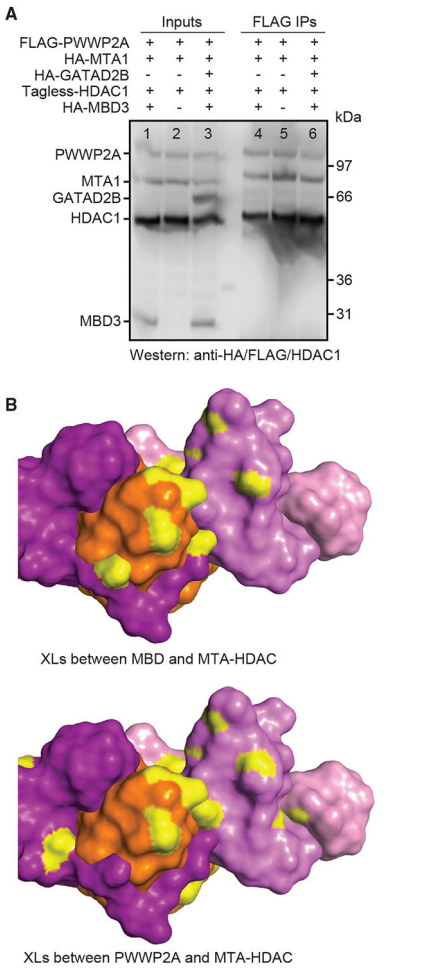


Figure 5. PWWP2A Competes with MBD3/GATAD2B for Binding to the MTA-HDAC-RBBP Complex

(A) Pull-downs showing that FLAG-PWWP2A purified on anti-FLAG beads pulls co-expressed HDAC1 and MTA1 out of cell lysate (lanes 2 + 5) but that neither co-expressed MBD3 (lanes 1 + 4) nor a combination of co-expressed MBD3 and GATAD2B (lanes 3 + 6) are pulled down. All NuRD components were hemagglutinin (HA) tagged except for HDAC1, which was untagged.

(B) Comparison of residues in the MTA-HDAC complex that exhibit XLs to either MBD (top) or PWWP2A (bottom). Colors: magenta (MTA), orange (HDAC), and yellow (cross-linked residues). See also Data S1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

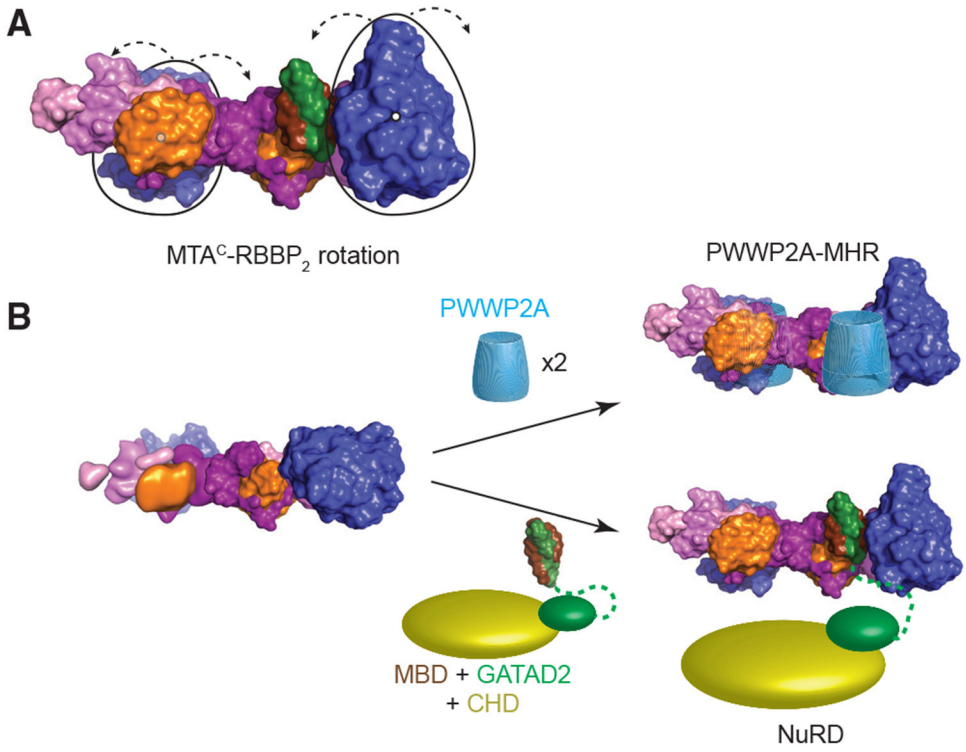


Figure 6. RBBP Subunits Introduce Conformational Dynamics and PWWP2A Competes Directly with the MBD-GATAD2-CHD Module

(A) A model based on XLMS and EM data depicting some of the possible range of movements by the MTA^{C}-$RBBP_2$ unit.

(B) PWWP2A competes directly with the MBD-GATAD2-CHD module for binding to the MTA-HDAC-RBBP module of NuRD, forming a PWWP2A-MTA-HDAC-RBBP complex. All NuRD subunit colors are as shown in Figure 2. See also Figures 2 and 5.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-HA-HRP	Cell Signaling Technology	2999S; RRID:AB_1264166
anti-HDAC1-HRP	Cell Signaling Technology	59581S; RRID:AB_2799568
anti-FLAG-HRP	Sigma-Aldrich	A8592; RRID:AB_439702
anti-MBD3	Abcam	ab157464
anti-rabbit-HRP	Leinco Technologies	R115; RRID:AB_2810875
Bacterial and Virus Strains		
Escherichia coli BL21(DE3)	Novagen	69450
Chemicals, Peptides, and Recombinant Proteins		
disuccinimidyl suberate	Creative Molecules Inc.	001S
adipic acid dihydrazide	Creative Molecules Inc.	019H
bissulfosuccinimidyl suberate	ThermoFisher Scientific	A39266
Absolutely Quantified SpikeTidesTM TQL	JPT Peptide Technologies	N/A
4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM)	Sigma-Aldrich	74104
Trypsin/Lys-C mix	Promega	V5073
Expi293 Expression Medium	ThermoFisher Scientific	A1435101
linear 25-kDa polyethylenimine	Polysciences	23966-1
DMEM/F-12	ThermoFisher Scientific	42400028
3 × FLAG peptide	Apexbio	A6001
anti-FLAG Sepharose 4B beads	BiMake	B26102
Glutathione Sepharose 4B beads	GE Healthcare	17075604
Deposited Data		
Deposited DIA-MS dataset	ProteomeXchange Consortium via the PRIDE partner repository	PXD010110
Deposited XLMS dataset	ProteomeXchange Consortium via the PRIDE partner	PXD010111
Deposited MTAN-HDAC-MBDGATAD2CC 3D map	EMDB repository	EMD-21382
Deposited NuDe raw 3D map	EMDB repository	EMD-22904; EMD-22905; EMD-22906
Deposited NuRD raw 3D map	EMDB repository	EMD-22895
Deposited MTAN-HDAC-MBDGATAD2CC raw EM micrographs	EMPIAR	EMPIAR-10539
Deposited NuDe raw EM micrographs	EMPIAR	EMPIAR-10538
Deposited NuRD raw EM micrographs	EMPIAR	EMPIAR-10537
Experimental Models: Cell Lines		
HEK Expi293F	ThermoFisher Scientific	A14527
mouse erythroleukemia (MEL) cells	N/A	N/A
Recombinant DNA		
pcDNA3-HA-GAT AD2A	This Paper	This Paper
pcDNA3-FLAG-PWWP2A	This Paper	This Paper

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pcDNA3-FLAG-MTA2	PMID: 29063705	PMID: 29063705
pcDNA3-HDAC1-tagless	PMID: 29063705	PMID: 29063705
pcDNA3-HA-RBBP7	PMID: 29063705	PMID: 29063705
pcDNA3-FLAG-MBD3GATAD2CC	PMID: 29063705	PMID: 29063705
pcDNA3-HA-MTA2N	PMID: 29063705	PMID: 29063705
pcDNA3-HA-MBD3	PMID: 29063705	PMID: 29063705
pcDNA3-FLAG-MTA1C	PMID: 29063705	PMID: 29063705
pcDNA3-HA-RBBP4	PMID: 29063705	PMID: 29063705
pcDNA3-HA-MTA1	PMID: 29063705	PMID: 29063705
pcDNA3-FLAG-MBD3	PMID: 29063705	PMID: 29063705
pcDNA3-HA-GATAD2B	PMID: 29063705	PMID: 29063705
pcDNA3-MBD3-tagless	This Paper	This Paper
pGEX6P-FOG1(1-45)	PMID: 27235397	PMID: 27235397
Software and Algorithms		
pLINK v2.3.5	PMID: 22772728	PMID: 22772728
xVis webserver	PMID: 25956653	PMID: 25956653
Proteome Discoverer v1.4	ThermoFisher Scientific	ThermoFisher Scientific
Skyline v3.5.0.9319	PMID: 20147306	PMID: 20147306
ASTRA software v6.1.1.17	Wyatt Technology	Wyatt Technology
SPHIRE	PMID: 28570515	PMID: 28570515
RELION 2.0	PMID: 23000701	PMID: 23000701
CryoSPARC	PMID: 28165473	PMID: 28165473
PyMOL	Schrödinger	Schrödinger
HADDOCK 2.4	PMID: 26410586	PMID: 26410586
Chimera	PMID: 15264254	PMID: 15264254
XLinkAnalyzer	PMID: 25661704	PMID: 25661704
GPS 2.1.1	PMID: 21062758	PMID: 21062758
MSConvert	PMID: 23051804	PMID: 23051804
Mascot search engine	Matrix Science	Matrix Science
Graphpad prism	Graphpad	Graphpad
SWISS-MODEL	PMID: 29788355	PMID: 29788355