Molecular architecture of nucleosome remodeling and deacetylase sub-complexes by integrative structure determination

Shreyas Arvindekar[1], Matthew J. Jackman[2], Jason K.K. Low[3], Michael J. Landsberg[2,*], Joel P. Mackay[3,*], and Shruthi Viswanath[1,*]

[1]National Center for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India

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

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

*Corresponding authors E-mail: m.landsberg@uq.edu.au (M.J.L.); joel.mackay@sydney.edu.au (J.P.M.); shruthiv@ncbs.res.in (S.V.)

Short title: Integrative models of NuRD sub-complexes

Abstract

The Nucleosome Remodeling and Deacetylase (NuRD) complex is a chromatin-modifying assembly that regulates gene expression and DNA damage repair. Despite its importance, limited structural information is available on the complex and a detailed understanding of its mechanism is lacking. We investigated the molecular architecture of three NuRD sub-complexes: MTA1-HDAC1-RBBP4 (MHR), MTA1^{N}-HDAC1-MBD3^{GATAD2CC} (MHM), and MTA1-HDAC1-RBBP4-MBD3-GATAD2 (NuDe) using Bayesian integrative structure determination with IMP (Integrative Modeling Platform), drawing on information from SEC-MALLS, DIA-MS, XLMS, negative-stain EM, X-ray crystallography, NMR spectroscopy, secondary structure and homology predictions. The structures were corroborated by independent cryo-EM maps, biochemical assays, and known cancer-associated mutations. MBD3 is a demethylase that connects the deacetylase and chromatin-remodeling modules in NuRD. Localization of the full-length MBD3 in NuRD was not previously characterized. Our models indicate two different localizations for MBD3 in NuRD, suggesting a mechanism by which MBD3 in the presence of GATAD2 asymmetrically bridges the two modules in NuRD. Further, our models indicate three previously unrecognized subunit interfaces in NuDe: HDAC1^{C}-MTA1^{BAH}, MTA1^{BAH}-MBD3, and HDAC1$^{60-100}$-MBD3. We observed that a significant number of cancer-associated mutations mapped to protein-protein interfaces in NuDe. Our approach also allows us to localize regions of unknown structure, such as HDAC1^{C} and MBD3^{IDR}, thereby resulting in the most complete structural characterization of these NuRD sub-complexes so far.

Introduction

The Nucleosome Remodeling and Deacetylase (NuRD) complex is a multi-protein chromatin-modifying assembly, expressed in most metazoan tissues, and conserved across multi-cellular animals (Basta and Rauchman, 2017; Denslow and Wade, 2007; Lejon et al., 2011; Yoshida et al., 2008). It regulates gene expression and DNA damage repair and It modulates nucleosome accessibility in enhancers and promoters for transcription factors and RNA polymerases, thereby regulating the expression of target genes (Basta and Rauchman, 2017, 2015; Bornelöv et al., 2018; Burgold et al., 2019; Denslow and Wade, 2007; Li and Kumar, 2010; Reynolds et al., 2013; Smeenk et al., 2010; Yoshida et al., 2008). Subunits of NuRD are implicated in human cancers and various congenital defects (Basta and Rauchman, 2015; Toh and Nicolson, 2009). Considerable diversity is observed in subunit isoforms and NuRD-associated factors across tissues (Burgold et al., 2019; Denslow and Wade, 2007; Hoffmann and Spengler, 2019).

NuRD comprises two catalytic modules – a histone deacetylase module and ATP-dependent chromatin-remodeling module (Burgold et al., 2019; Denslow and Wade, 2007; Low et al., 2020). The deacetylase module contains metastasis-associated proteins (MTA1/2/3) that form a dimeric scaffold for the histone deacetylases (HDAC1/2). It also contains the chaperones RBBP4/7, which mediate interactions of NuRD with histone tails and transcription factors (Basta and Rauchman, 2017, 2015; Hong et al., 2005). The chromatin-remodeling module contains methyl-CpG DNA-binding proteins (MBD2/3) that recruit NuRD to methylated and/or hemi-methylated DNA, GATA-type zinc-finger proteins (GATAD2A/B), and an ATP-dependent DNA translocase (CHD3/4/5) (Burgold et al., 2019; Low et al., 2020).

Some structural information is available for the complex. Several attempts made to determine the stoichiometry of the endogenous NuRD complex have returned variable results (Bode et al., 2016; Guo et al., 2019; Kloet et al., 2015; Sharifi Tabar et al., 2019; Smits et al., 2013; Spruijt et al., 2016; Zhang et al., 2016). A recent characterization by quantitative mass

spectrometry from (Low et al., 2020) reported a 2:2:4:1:1:1 (MTA1:HDAC1:RBBP4:MBD3:GATAD2A:CHD4) stoichiometry for the full NuRD complex. Atomic structures of parts of the NuRD complex, including the MTA1-HDAC1 dimer, RBBP4 bound to MTA1, the MBD domain of MBD3, and the coiled-coil dimer of MBD2 and GATAD2A have been determined by X-ray crystallography and NMR spectroscopy (Alqarni et al., 2014; Cramer et al., 2014; Gnanapragasam et al., 2011; Millard et al., 2016, 2013). Structures of the 2:2 MTA1-HDAC1 dimer, the 2:2:4 MTA1-HDAC1-RBBP4 complex (MHR), the 2:2:2 $MTA1^{N}$-$HDAC1\text{-}MBD3^{GATAD2CC}$ (MHM) complex, the 2:2:4:1:1 MTA1-HDAC1-RBBP4-MBD3-GATAD2 (NuDe complex), and CHD4 bound to a nucleosome substrate have also been characterized at various resolutions by negative stain and/or cryo-electron microscopy (Farnung et al., 2020; Low et al., 2020; Millard et al., 2020, 2016).

Pairwise interactions between domains and subunits within the MHR, MHM, NuDe, and the endogenous NuRD complexes have also been characterized by chemical crosslinking and mass spectrometry (XLMS) (Low et al., 2020; Millard et al., 2016). A model of the MHM complex, based on crosslinks-driven rigid-body docking of known atomic structures with a pair of MTA1-RBBP4 structures manually placed, has also been reported (Low et al., 2020). While this represents the most complete model of NuRD architecture, it only accounts for 30% of residues in the NuRD complex. In fact, only 50% of residues in NuRD have known or readily modeled atomic structures, and the structures of proteins such as MBD3, CHD4, and GATAD2 are largely uncharacterized. More recent artificial intelligence-based methods such as AlphaFold are also unable to resolve these uncharacterized regions (Jumper et al., 2021). Owing to variability in the paralogue composition as well as significant structural dynamics, the atomic structure of the entire NuRD complex is still undetermined and is likely to remain a challenge for some time to come.

The 2:2:4 MTA1-HDAC1-RBBP4 complex (MHR) forms the deacetylase core of NuRD. MBD3 binds to the N-terminal half of MTA1 to form a 2:2:2 $MTA1^{N}$-$HDAC1\text{-}MBD3^{GATAD2CC}$ (MHM) complex, which contains two copies of MBD3 (Low et al., 2020). However, the 2:2:4:1:1 MTA1-HDAC1-RBBP4-MBD3-GATAD2 (NuDe complex) and the endogenous NuRD complex are asymmetric, both containing a single copy of MBD3 (Low et al., 2020). The mechanism by which this asymmetry is introduced in NuDe/NuRD is not known. The structure of full-length MBD3 is unknown and it contains a significant intrinsically disordered region (IDR; $MBD3^{71\text{-}213}$), which has been shown to be critical for recruiting the deacetylase core in case of MBD2 (Desai et al., 2015). The localization of full-length MBD3 in NuDe/NuRD is also not known.

Here, we investigated the molecular architecture of the MHR, MHM, and NuDe complexes using an integrative approach. Integrative structure determination is a method for investigating structures of large assemblies which are refractory to a single experimental method such as X-ray crystallography or cryo-electron microscopy (Alber et al., 2007; Ganesan et al., 2020; Gutierrez et al., 2020; Kim et al., 2018; Rout and Sali, 2019; Russel et al., 2012; Viswanath et al., 2017a; Ward et al., 2013). Using Bayesian integrative modeling with the Integrative Modeling Platform (IMP), we combined data from complementary experiments, physical principles, statistical inference, and prior models. This approach allowed us to combine noisy, sparse, ambiguous, and incoherent data at various resolutions (Alber et al., 2007; Rieping et al., 2005; Rout and Sali, 2019; Schneidman-Duhovny et al., 2014). It produced an ensemble of

models consistent with the input information, allowing us to obtain precise uncertainty bounds on the structure (Saltzberg et al., 2019, 2021; Viswanath et al., 2017b; Webb et al., 2018). We used data from SEC-MALLS, DIA-MS, XLMS, negative stain EM, X-ray crystallography, NMR spectroscopy, secondary structure and homology predictions, and stereochemistry considerations (Alqarni et al., 2014; Connelly et al., 2006; Cramer et al., 2014; Gnanapragasam et al., 2011; Low et al., 2020; Millard et al., 2016, 2013). These integrative structures were corroborated by independent cryo-EM maps, biochemical assays, and known cancer-associated mutations, a significant number of which mapped to protein-protein interfaces in the structures (Desai et al., 2015; Forbes et al., 2006; Millard et al., 2020; Pflum et al., 2001; Zhang et al., 1999). By using all available information, the accuracy, precision, completeness, and efficiency of structure determination were maximized (Alber et al., 2007; Rout and Sali, 2019; Russel et al., 2012).

The integrative approach allows for regions of unknown structure to be modeled in the context of regions of known structure. This facilitated the modeling of NuRD proteins with significant regions of unknown structure, such as MBD3, at full-length, for the first time. Our models indicate that MBD3 localizes in two different sites in NuRD. This suggests a mechanism by which MBD3, in the presence of GATAD2, asymmetrically bridges the deacetylase and chromatin-remodeling modules. Finally, our models enable us to compare the structure of the MHR complex in the presence and absence of MBD3 and GATAD2. We show that, while the MHR complex alone is dynamic, the presence of MBD3 and GATAD2 makes it less dynamic.

Results

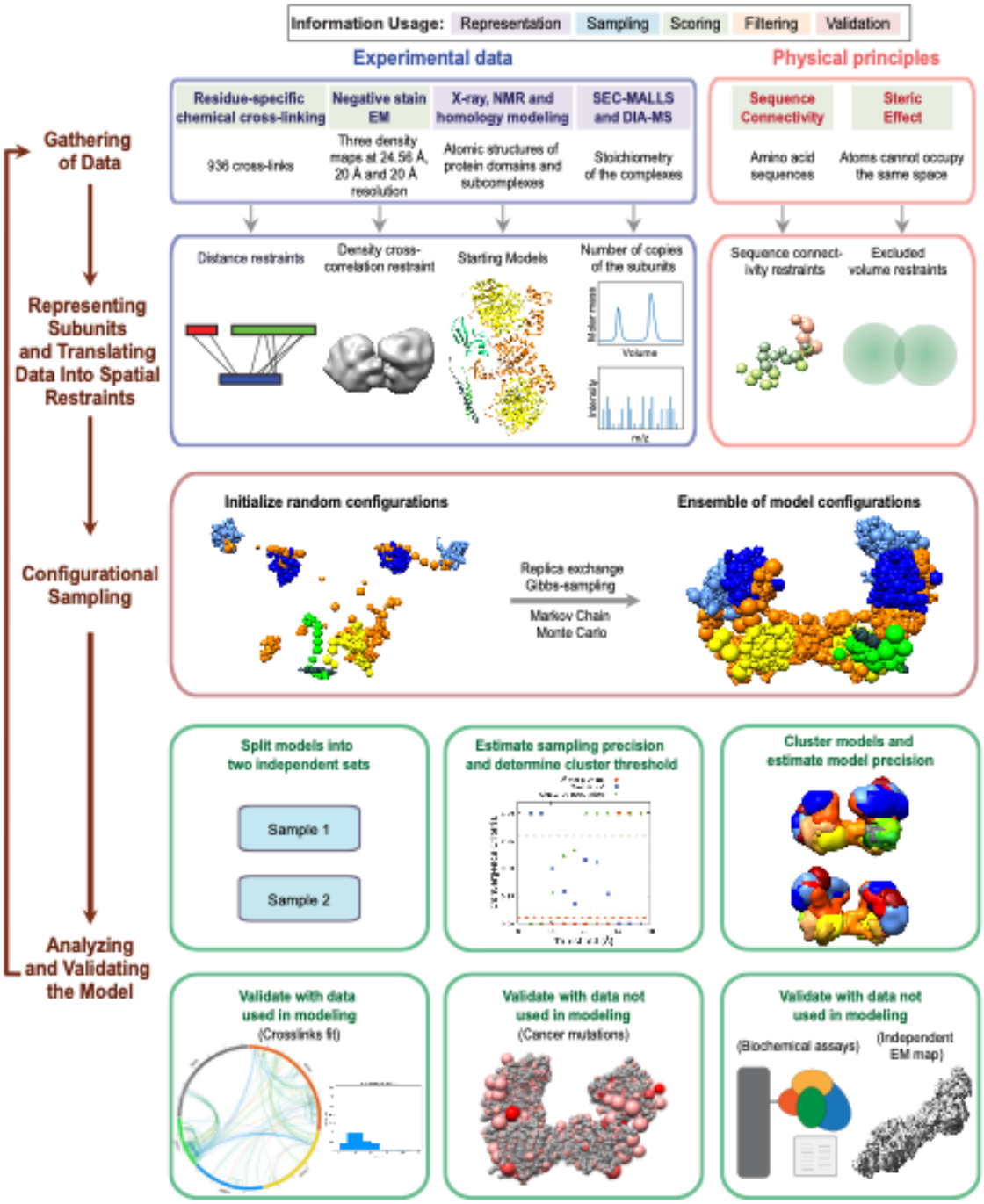

Fig. 1 Integrative structure determination of NuRD sub-complexes Schematic describing the workflow for integrative structure determination of NuRD sub-complexes. The first row describes the input information. The second-row details how data is used to encode spatial restraints. The third row mentions the sampling method, and the last two rows illustrate the analysis and validation protocol. The background colors of the input information show the stage of modeling in which the information is used, as shown in the legend at the top.

Integrative modeling workflow

The integrative modeling of the MHR, MHM, and NuDe complexes proceeded in four stages (Fig. 1, Material and Methods) (Alber et al., 2007; Rout and Sali, 2019; Russel et al., 2012). The modeled NuRD proteins (subunits), their domains, their representations, and the number of copies in the modeled complexes are shown (Fig. S1A-S1C). The stoichiometry of the modeled proteins was informed by DIA-MS and SEC-MALLS experiments (Fig. S1C) (Low et

al., 2020). Regarding paralogs, a single representative for each protein was chosen for this work for simplicity, namely MTA1, HDAC1, RBBP4, MBD3 and GATAD2A (Fig. S1A).

We first represented each protein as a series of beads of sizes that depend on the degree of knowledge of the structure (and can vary throughout the sequence). Protein domains with known atomic structures (such as the MTA1-HDAC1 dimer) were represented at 1 and 10 residues per bead and modeled as rigid bodies, whereas domains without known structure (such as the $MBD3^{IDR}$) were coarse-grained at 30 residues per bead and modeled as flexible strings of beads (Fig. S1A-S1B). Data from chemical crosslinking combined with mass spectrometry (XL-MS) were used to restrain the distance between cross-linked residues. Negative-stain EM maps were used to restrain the shape of the complexes (Low et al., 2020).

The simulations started with randomized configurations for the rigid bodies and flexible beads. Over 40 million models per complex were sampled using a Monte Carlo approach (Replica Exchange Gibbs Sampling MCMC; Materials and Methods). The models were scored based on agreement with XL-MS and EM data, together with additional stereochemistry restraints such as connectivity and excluded volume. For each complex, about 20,000 models that sufficiently satisfied the input information were selected for further analysis (Saltzberg et al., 2021).

These models were clustered based on structural similarity and the precision of the clusters was estimated (Fig. S3-S5) (Saltzberg et al., 2019, 2021; Viswanath et al., 2017b). The quality of the models was assessed by fit to input data (Fig. S6-S8), as well as data not used in modeling, such as independent cryo-EM maps (Millard et al., 2020), published biochemical data (Desai et al., 2015; Millard et al., 2020; Pflum et al., 2001; Zhang et al., 1999) and human cancer-associated mutations (COSMIC) (Table S1) (Forbes et al., 2006). The resulting integrative models were visualized in two ways - a representative bead model and a localization probability density map, and represented in UCSF Chimera and ChimeraX (Pettersen et al., 2021, 2004). The bead model represents the centroid of the major cluster, whereas the localization probability density map represents all models in the major cluster, by specifying the probability of a voxel (3D volume unit) being occupied by a bead in the set of superposed cluster models.

MHR

First, to support the integrative modeling of the MHR complex, an *ab initio* 3D EM map for the MHR complex was produced by further analysis of the MHR 2D class averages reported in a previous study (Low et al., 2020). Integrative modeling of the 2:2:4 MHR complex produced effectively a single cluster of models (85% of a total of 15200 models) with a model precision of 27 Å; model precision is the average RMSD between the cluster centroid and models in the cluster (Fig. S3). The models fit very well to the input data as measured by the EM and crosslink scores. 98% of the input crosslinks were satisfied within their uncertainty (Fig. S6). An adipic acid dihydrazide (ADH) / bis(sulfosuccinimidyl)suberate - disuccinimidyl suberate (BS3DSS) / dimethoxy triazinyl methyl-morpholinium chloride (DMTMM) crosslink is violated if the corresponding cross-linked beads are greater than 35 / 35 / 25 Å apart in all models in the cluster. The cross-correlation between the localization probability density map for the models in the major cluster and the input EM map was 0.74, indicating the fit to EM is reasonable but not too high. This could partly be due to unoccupied density in the lobes of the experimental EM map.

Surprisingly, the representative bead model from the dominant cluster (cluster centroid model) shows the C-terminal half of the two MTA subunits ($MTA1^{432-715}$) crossing over (brown and orange MTAs, Fig. 2A, Movie M1). Integrative models of the MHR complex created in the absence of the EM map also showed the MTAs crossing over (Fig. S9).

The $MTA1^{BAH}$ domain ($MTA1^{1-164}$) is positioned distal to the MTA1 dimerization interface ($MTA1^{200-290}$, $MTA1^{dimer}$), consistent with its position in an independent EM map (Fig. 2B, Fig. 2C) (Millard et al., 2020). It is proximal to the HDAC1 active site and might regulate HDAC1 activity (Fig. 2A). This conclusion is consistent with histone deacetylation assays in which MTA1 was shown to modulate HDAC1 deacetylase activity in NuRD (Zhang et al., 1999). Further, for one of the MTAs, the $MTA1^{BAH}$ is located near an RBBP4 (Fig. 2A, Fig. 2B); $MTA1^{BAH}$ proximity to RBBP4 was also indicated in an independent cryo-EM map (Millard et al., 2020). Finally, $MTA1^{BAH}$ is also proximal to the $MTA1^{mid}$ region ($MTA1^{334-431}$) containing the predicted helix (H) and zinc finger regions (ZF) (Fig. 2B, Fig. 2C).

The $MTA1^{mid}$ region is juxtaposed between $MTA1^{dimer}$ and the $MTA1^{BAH}$ domain (Fig. 2B). In contrast, in a previous crosslink-based MHR model (Low et al., 2020), $MTA1^{mid}$ was proximal to the $MTA1^{BAH}$ domain and distal from the $MTA1^{dimer}$. The MTA1 C-terminus ($MTA1^{C}$; *i.e.*, $MTA1^{692-715}$) shows considerable conformational heterogeneity and is co-located with $MTA1^{USR}$ ($MTA1^{547-669}$), the MTA1 disordered region between the R1 and R2 RBBP4 binding regions (Fig. 2B, Movie M1). Overall, many MTA1 domains in the MHR model, such as $MTA1^{BAH}$ domain, $MTA1^{mid}$, and $MTA1^{C}$, are exposed and could interact with nucleosomal DNA and/or other proteins.

The HDAC1 C-terminus ($HDAC1^{C}$; *i.e.*, $HDAC1^{377-482}$) interacts with the $MTA1^{BAH}$ domain (Fig. 2B). Although it has been shown that the MTA1-HDAC1 dimer can form in the absence of $MTA1^{BAH}$ (Millard et al., 2013), this additional interaction between MTA1 and HDAC1 could be functionally important. Consistent with this possibility, mutations in $HDAC1^{C}$ (Δ391-482, S421A, S423A, E426A) have been known to disrupt binding to NuRD subunits (Pflum et al., 2001). There are also post-translational modifications in the HDAC1 tail that might modulate its interaction with MTA1 (Pflum et al., 2001; Rathert et al., 2008).

Both the $MTA1^{R1}$-RBBP4 units are located between the two lobes in the EM map, with one complex in the front and the other at the back (dark blue beads and densities, Fig. 2A-2C). On the other hand, the $MTA1^{R2}$-RBBP4 complexes are located in separate lobes (light blue beads and densities, Fig. 2A-2C). The densities of RBBP4 are spread out, indicating its localization in MHR is imprecise (Fig. 2B, Movie M1). This is consistent with the structural heterogeneity observed in 2D class averages of the MHR EM data (Low et al., 2020). This flexibility could facilitate RBBP4 interactions with transcription factors and histones.

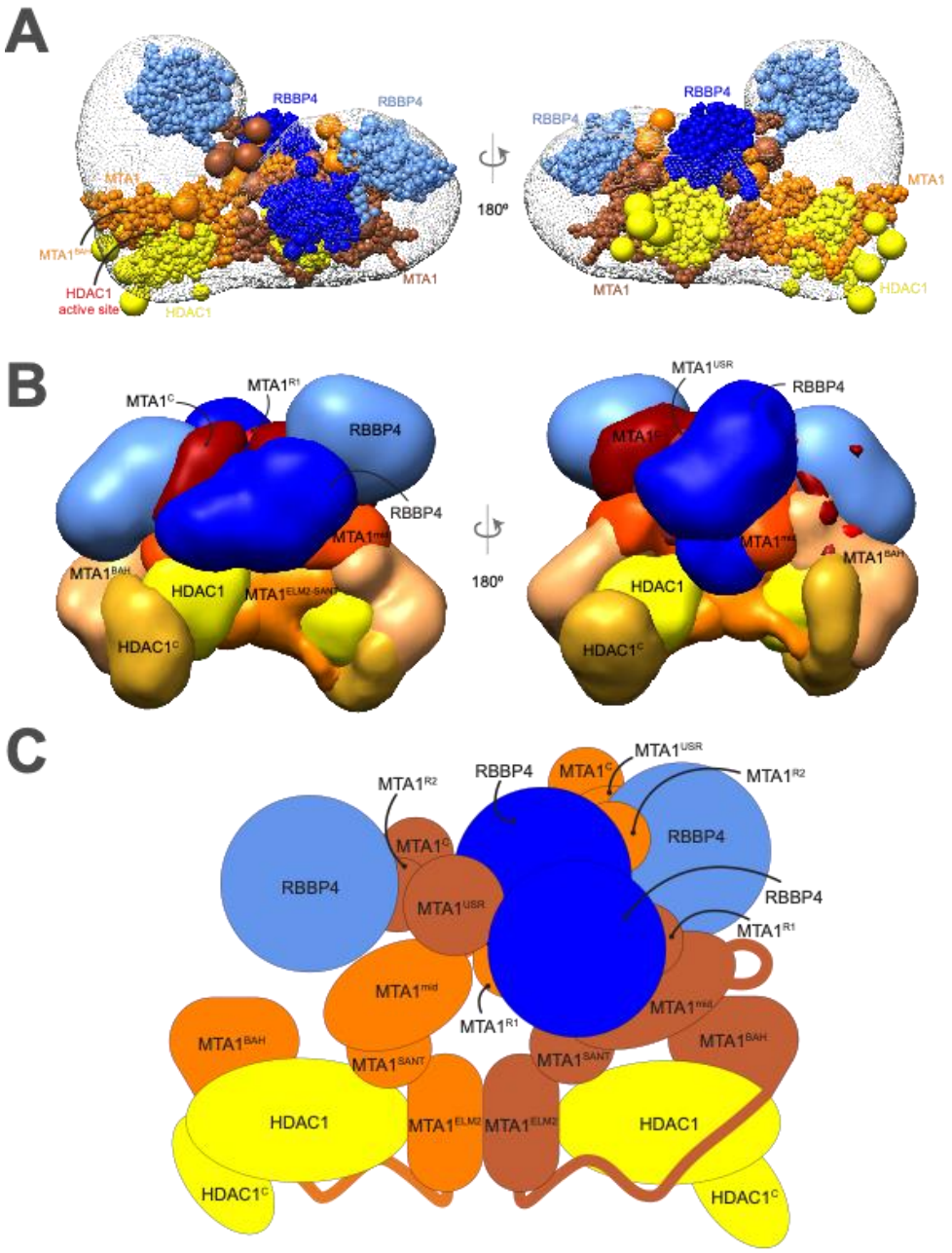

Fig. 2 Integrative model of the MTA1-HDAC1-RBBP4 (MHR) complex A. Representative bead model from the most populated cluster of integrative models for the MHR complex, shown with the MHR EM map. The model is colored by subunit. For MTA1, the two copies are shown in different colors (brown and orange) in panels A and C, to illustrate the crossover. The HDAC1 active site is shown in red. B. Localization probability density maps showing the position of different domains/subunits in the cluster. The map specifies the probability of any volume element being occupied by a domain in the ensemble of superposed models from the cluster. The domain densities are colored according to Fig. S1. These maps are contoured at ~10% of their respective maximum voxel values. C. Schematic representation of the integrative model of the MHR complex. See also Figs. S1, S3, S6, and Movie M1.

MHM

Integrative modeling of the 2:2:2 MHM complex resulted in a major cluster containing 60% of 21960 models. The model precision was 24 Å and 99% of the input crosslinks were satisfied (Fig. S4, Fig. S7). The cross-correlation between the localization probability density map for the models in the major cluster and the input EM map was 0.90.

First, in a control run, where MHM was modeled as a 2:2:1 complex with a single copy of MBD3, we observed two symmetric MBD3 binding sites (Fig. S10). However, our 2:2:2 MHM model shows that the two copies of MBD3 bind asymmetrically on the MTA1-HDAC1 dimer (Fig. 3A-3C, Movie M2). One MBD3 localizes exclusively to one end of the MTA1-HDAC1 dimer, making contacts predominantly with a single MTA1 and HDAC1 (pink MBD3, Fig. 3C), whereas the other MBD3 is more spread out and interacts with both copies of MTA1 and HDAC1 (green MBD3, Fig. 3C).

In our models, both the copies of $MBD3^{MBD}$ localize close to the $MTA1^{BAH}$ domain, which is similar to the location observed for $MBD2^{MBD}$ in an independent cryo-EM map of a 2:2:1 MTA1:HDAC1:MBD2 complex (Fig. 3A-3C) (Millard et al., 2020). Although there are two MBD3s in our models, only a single $MBD3^{IDR}$ localizes to the MTA1 dimerization interface, $MTA1^{dimer}$ (green MBD3, Fig. 3D-3E). This localization of $MBD3^{IDR}$ is consistent with its previously predicted localization from the crosslinks-based model (Low et al., 2020) and the localization of $MBD2^{IDR}$ based on cryo-electron microscopy (Millard et al., 2020). It is also supported by two separate mutagenesis and co-immunoprecipitation studies, one of which showed that $MBD2^{IDR}$ was essential for binding to the MTA1-HDAC1 dimer (Desai et al., 2015), while the other showed that $MTA1^{dimer}$ was essential for its interaction with MBD2 (Millard et al., 2020). It is known that MBD3 binding in NuDe/NuRD is asymmetric; although there are two symmetric MBD3 binding sites, there is a single MBD3 in these complexes (Low et al., 2020). It is possible that a single MBD3 is present in NuDe/NuRD since a single $MBD3^{IDR}$ can bind to the $MTA1^{dimer}$ (see also Discussion).

For both MBDs, the $MBD3^{MBD}$ domain is proximal to the $MTA1^{BAH}$ domain, and the $MBD3^{IDR}$ is close to the $MTA1^{mid}$ region (Fig. 3B, Fig. 3E). The $MBD3^{CC}$-$GATAD2^{CC}$ coiled-coil domain is exposed. Finally, in a small minority of models, both MBDs are localized together on the same side of the $MTA1^{dimer}$; however, we believe this latter configuration is possibly due to numerous intra-MBD3 crosslinks (Movie M2).

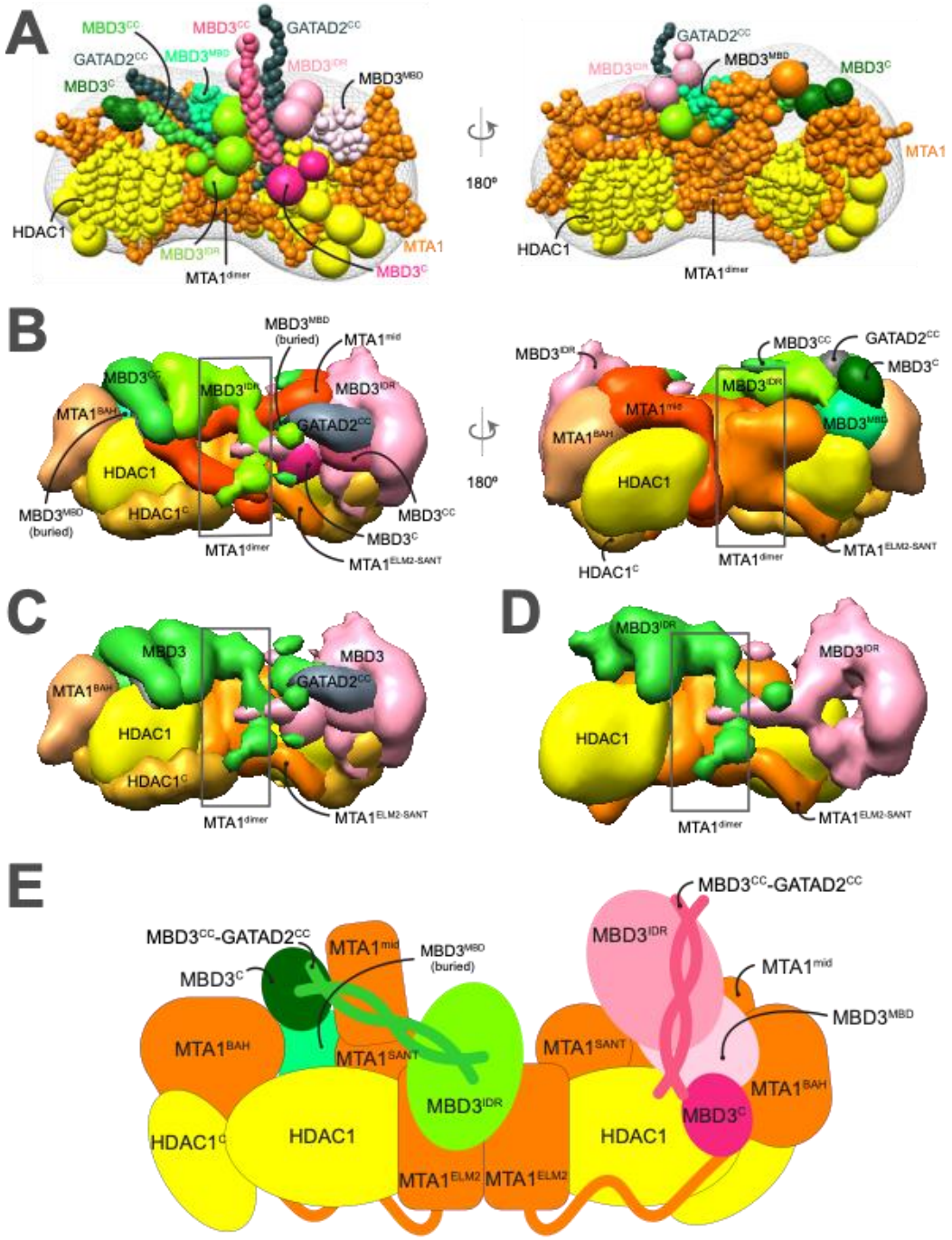

Fig. 3 Integrative model of the MTA1^{N}-HDAC1-MBD3^{GATAD2CC} (MHM) complex A. Representative bead model from the major cluster of analyzed integrative models for the MHM complex, with the corresponding EM map (EMD-21382) (Low et al., 2020), colored by subunit. The domains of the two MBD3s are shown in shades of pink and green respectively. B. Localization probability density maps showing the position of different domains in the ensemble of models from the cluster. The domain densities are colored according to Fig S1. C. The same density maps as B (front view), showing the two MBDs in pink and green respectively, illustrating that they localize differently on the MTA1-HDAC1 dimer. The density maps of MTA1^{mid} and GATAD2^{cc} were omitted for clarity. D. The density maps of the two MBD3^{IDR} domains on the MTA1-HDAC1 dimer. Almost all the maps are contoured at 20% of their respective maximum voxel values (GATAD2 maps are contoured at 27% whereas MBD3^{IDR} maps are contoured at 10%). E. Schematic representation of the integrative model of the MHM complex. Note that MTA1^{mid} in this model corresponds to MTA1$^{334-431}$. See also Figs. S1, S4, S7, and Movie M2.

NuDe

Although the NuDe complex contains full-length GATAD2, due to the lack of information on full-length GATAD2, we modeled only the GATAD2 region that forms a coiled-coil with MBD3. Integrative modeling of the NuDe complex resulted in effectively a single cluster (99% of 21632 models). The model precision was 35 Å and 99% of the input crosslinks were satisfied (Fig. S5, Fig. S8). The cross-correlation between the localization probability density map for the models in the major cluster and the input EM map was 0.88.

In contrast to our MHM model, where one MBD3 was proximal to the MTA1^{SANT} domain, MBD3 in NuDe localizes precisely to one end of the MTA1-HDAC1 dimer, away from the MTA1^{SANT} domain. It is juxtaposed next to the MTA1^{BAH} and MTA1^{mid} domains, as well as HDAC1 (Fig. 4A-4C, Movie M3). An independent cryo-EM map of MTA1$^{1-546}$-HDAC1-MBD2-RBBP4 also showed that MBD3 was proximal to MTA1^{BAH} and MTA1^{dimer} (Millard et al., 2020). It is possible that the presence of GATAD2 sterically precludes MBD3 from occupying the MTA1 dimerization interface (see also Discussion).

From protein-protein distance maps of the cluster, HDAC1$^{60-100}$ and MTA1^{BAH} are most proximal to MBD3 (Fig. S11A, S11B). MBD3^{CC}-GATAD2^{CC} is exposed. The MBD3^{MBD} domain is buried, consistent with the failure of MBD3 to bind DNA in NuRD noted in immuno-precipitation experiments (Fig. 4A-4C) (Zhang et al., 1999). Interestingly, several nucleosome-interacting domains such as MTA1^{BAH} and MTA1^{ZF} are co-localized in the NuDe model (Fig. 4A-4C).

Similar to the MHR models, the HDAC1^{C} domain is proximal to MTA1^{BAH} (Fig. S11C). Mutagenesis and co-immunoprecipitation studies have shown that HDAC1^{C} makes critical interactions with NuRD subunits (Pflum et al., 2001). In contrast to the MHR models which showed crossover of MTAs, the two MTAs are well-separated in NuDe (Fig. 4A-4C). The localization of RBBPs is also more precise in NuDe than in MHR (Fig. S12).

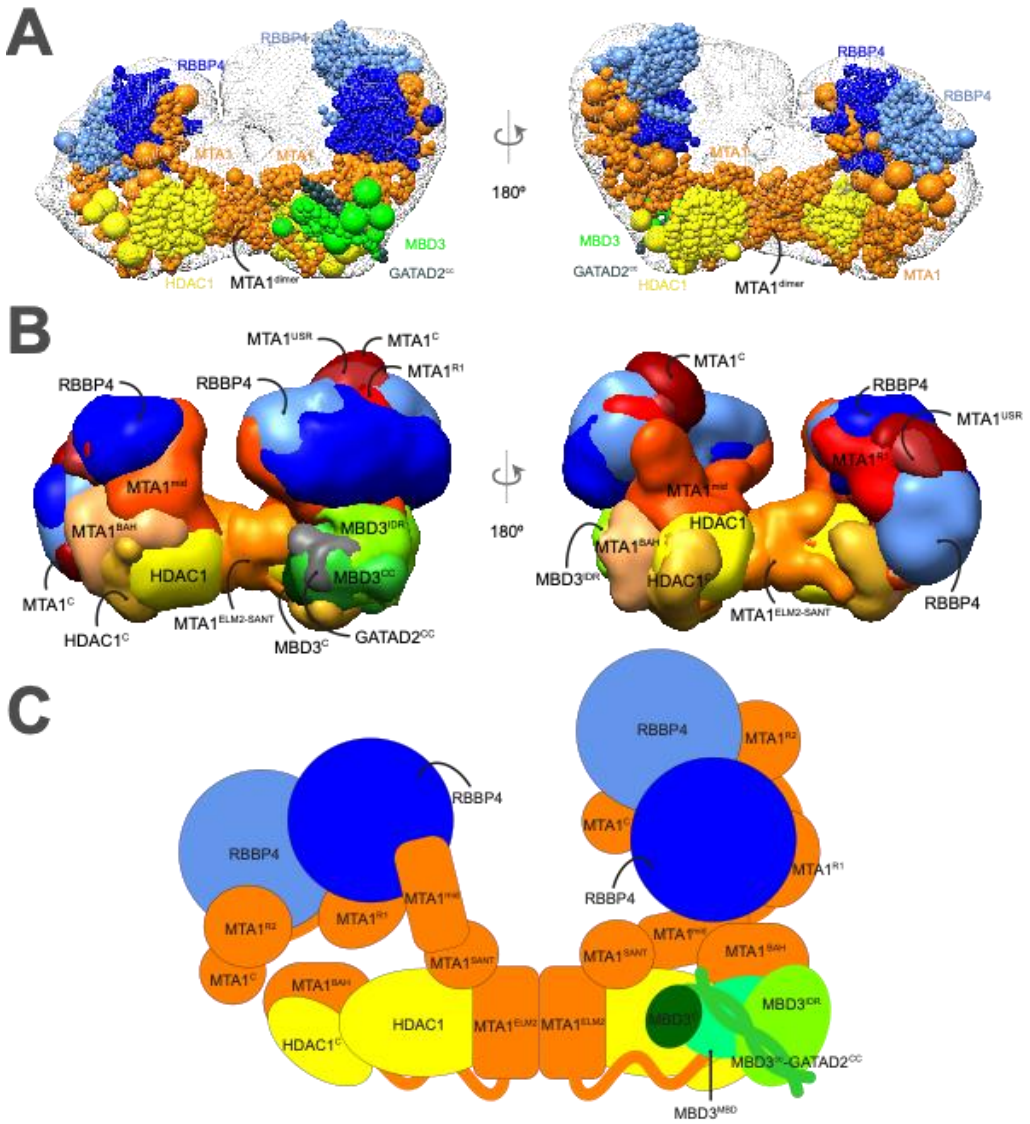

Fig. 4 Integrative model of the nucleosome deacetylase (NuDe) complex A. Representative bead model from the dominant cluster of integrative models for the NuDe complex, with the corresponding EM map (EMD-22904) (Low et al., 2020), colored by subunit. B. Localization probability density maps showing the position of different domains in the ensemble of models from the cluster. The domain densities are colored according to Fig S1. Maps are contoured at ~10% of their respective maximum voxel values. C. Schematic representation of the integrative model of the NuDe complex. See also Figs. S1, S5, S8, S11, and Movie M3.

Mapping COSMIC mutations

We next consulted the COSMIC (Catalogue of Somatic Mutations in Cancer) database for somatic, confirmed pathogenic, point mutations of the NuRD subunits, MTA1, HDAC1, RBBP4, and MBD3 (Forbes et al., 2006). In total, 356 point mutations were identified and mapped onto the cluster of NuDe integrative models (Methods, 4.5 COSMIC data analysis). Analysis of these mutations revealed that 74% of mutations mapped to protein-protein interfaces within NuDe; for this analysis, a mutation was considered to be at an interface if the average distance of the corresponding residue to a residue in an interacting protein is less than 10 Å. 29% of the mapped mutations were located in previously uncharacterized binding interfaces predicted by our model, such as MBD3-MTA1^{BAH}, MBD3-HDAC1$^{60\text{-}100}$, and HDAC1^{C}-MTA1^{BAH}, consistent with the idea that these interfaces are important for NuRD function (Fig. 5, Fig. S13, Table S1).

Mutations at protein-protein interfaces could alter protein-protein interactions, reducing the stability of the complex, thereby leading to pathogenicity. Moreover, 19% of all mutations mapped to exposed regions that are known to bind to nucleosomes and transcription factors, such as the HDAC1 active site and RBBP4 H3 interaction site (Fig. 5, Fig. S13, Table S1). These mutations could impair NuRD binding to partners such as nucleosomes or transcription factors, contributing to the pathogenesis of disease. Of the 19% that map to exposed regions, more than half (57%) of the mutations map to regions of unknown structure (regions for which no experimental structure or reliable model is available), such as MTA1^{USR} and MBD3^{IDR} (Fig. 5, Table S1). The functional significance of these mutations is therefore difficult to predict, but could indicate that these regions of unknown structure also have important roles in protein stability, regulating interactions between NuRD subunits, or interactions with binding partners of NuRD. An important consideration for all these disease-causing mutations is that many of the NuRD subunits function in cellular contexts independent of other NuRD subunits, and so in some cases, these mutations may be rationalised in the context of other functional roles.

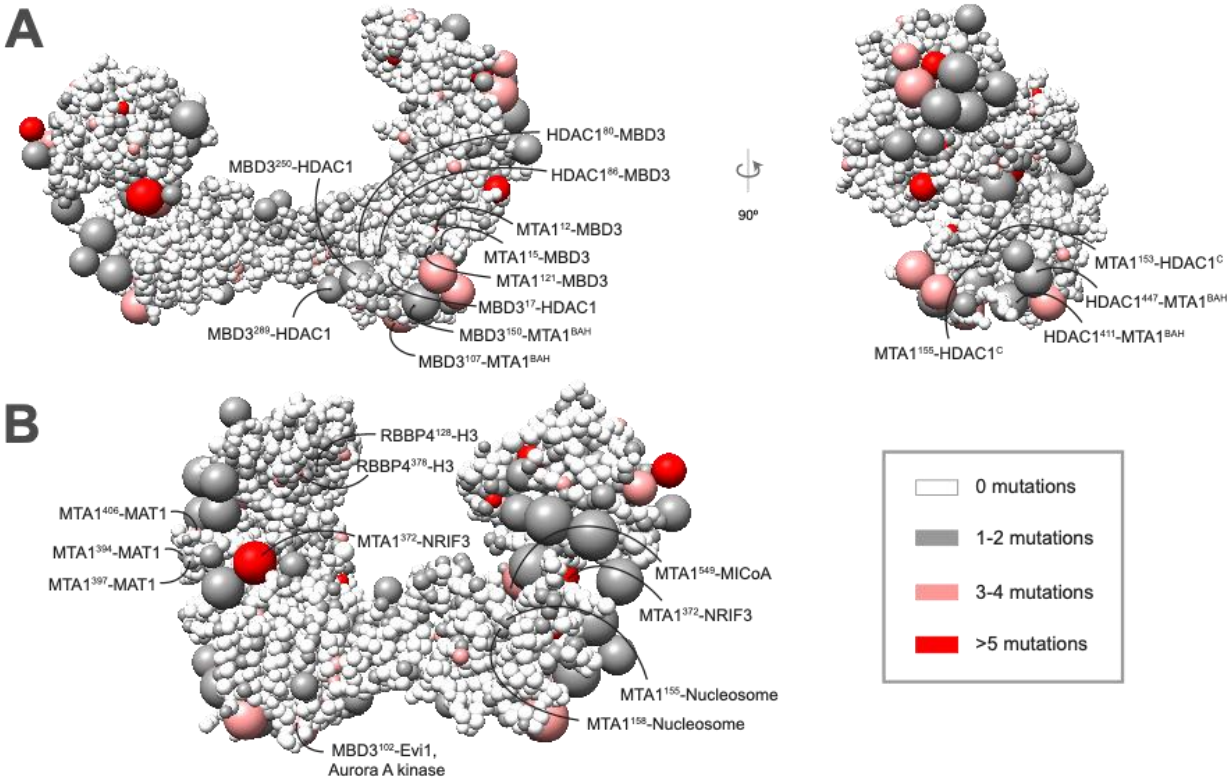

Fig. 5 COSMIC mutations mapped onto the NuDe integrative model. Somatic pathogenic point mutations from the COSMIC database (Forbes et al., 2006) mapped onto the representative bead model of the NuDe complex (Fig. 4A). A. Mutations of residues that map to previously undescribed protein-protein interfaces within our model. Residues from two proteins are at an interface if the average distance between their corresponding bead surfaces is less than 10 Å in the cluster of NuDe integrative models. B. Mutations on residues that map to exposed binding sites between modeled proteins and known binding partners. A bead is colored according to the maximum number of mutations on any residue in the bead, according to the legend. Representative mutations are labeled in both A. and B. See also Table S1 and Fig. S13.

Docking the nucleosome

We next attempted to dock the CHD4-nucleosome structure (Farnung et al., 2020) into the cleft in the NuDe structure between the MTA1 C-terminal arms (Fig. 6). Although there are limitations to this docking (for example, it is a coarse placement, lacks histone tail densities, and binding sites for the nucleosome outside the cleft might also exist), this positioning of the nucleosome indicates its size complementarity to the integrative model, further corroborating the latter. This

placement allows for the histone H3 and H4 tails to be located towards the HDAC1 active site. It also accommodates the known interactions between the RBBPs and the histone H3 (Fig. 6). The partial CHD4 structure is exposed. MTA1^{mid}, which contains the zinc finger, can also potentially interact with the nucleosome in this position. Finally, MBD3 does not interact with the nucleosome, since MBD3^{MBD} is buried in NuDe (Fig. 4A-4C), consistent with MBD3 in NuRD failing to bind DNA in immuno-precipitation experiments (Zhang et al., 1999).

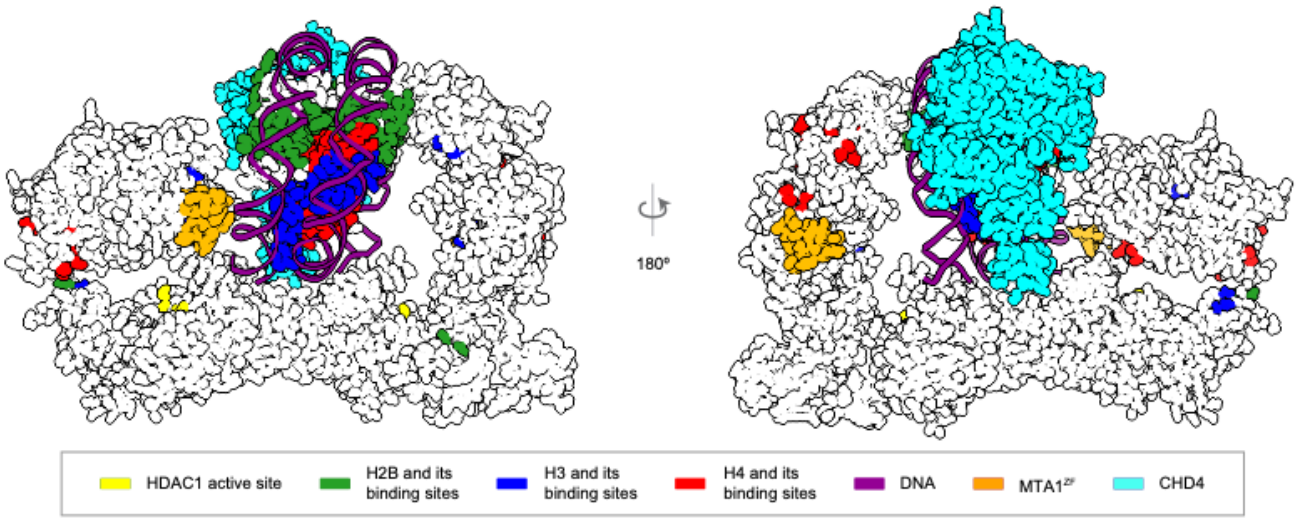

Fig. 6 Integrative model of NuDe complex with the nucleosome The CHD4-nucleosome structure (Farnung et al., 2020) is placed in the cleft of the NuDe integrative model. The regions with known atomic structure are shown in the NuDe integrative model from Fig. 4A. Nucleosome proteins/DNA and the corresponding NuDe subunit residues they are known to bind to, are depicted in the same color, as given by the legend.

Discussion

Here, we obtained structural models of the MTA1-HDAC1-RBBP4 (MHR), MTA1^{N}-HDAC1-MBD3 (MHM), and MTA1-HDAC1-RBBP4-MBD3-GATAD2 (NuDe) complexes using Bayesian integrative modeling. The approach allowed us to combine all available structural information, including data from SEC-MALLS, DIA-MS, chemical crosslinking mass spectrometry (XLMS), negative stain EM, X-ray crystallography, NMR spectroscopy, secondary structure, and homology predictions (Methods, Fig. 1) (Alqarni et al., 2014; Connelly et al., 2006; Cramer et al., 2014; Gnanapragasam et al., 2011; Low et al., 2020; Millard et al., 2016, 2013; Söding et al., 2005; Tjandra et al., 1997). The models were corroborated by independent cryo-EM maps, enzyme assays, mutagenesis, co-immunoprecipitation studies, and the mapping of cancer mutations (Fig. 5, Fig. S13, Table S1) (Desai et al., 2015; Forbes et al., 2006; Millard et al., 2020; Pflum et al., 2001; Zhang et al., 1999). Importantly, our approach allowed us to localize regions of unknown structure, for *e.g.*, HDAC1^{C} and MBD3^{IDR}, in context of regions of known structure, for *e.g.*, the MTA1-HDAC1 dimer, resulting in the most complete structural characterization of these NuRD sub-complexes so far.

NuDe complex is more ordered than MHR

A comparison of MTA1 and RBBP4 in the MHR and NuDe models suggests that these subunits are more conformationally heterogenous in MHR, as shown by the broader localization probability densities for the C-terminal half of MTA1 and RBBPs in MHR (volume enclosed by the corresponding maps = 1120 nm^3) compared to NuDe (volume enclosed = 883.6 nm^3) (Fig. S12). Also, the cross-correlation of the MHR localization probability density to the

corresponding EM map is lower than that of NuDe, indicating higher heterogeneity for the former. This indicates that the MHR is initially dynamic and the presence of MBD3-GATAD2 possibly makes it less dynamic.

MBD3^{IDR} – MTA1^{dimer} interaction

In our MHM models, one MBD3^{IDR} is near the MTA1^{dimer}, consistent with the previously predicted localization of MBD3^{IDR} based on chemical crosslinks (Low et al., 2020) and MBD2^{IDR} based on a cryo-electron density map (Fig. 3C-3E) (Millard et al., 2020). Two separate mutagenesis and co-immunoprecipitation studies have shown that the MBD^{IDR} and the MTA1 dimerization interface are each essential for MBD2 interaction with the MTA1-HDAC1 dimer (Desai et al., 2015; Millard et al., 2020). Despite the corresponding region of MBD2 being disordered in solution (Desai et al., 2015), MBD3$^{125-175}$ is predicted to be ordered based on PONDR® analysis (Fig. S14) (http://www.pondr.com) (Romero et al., 2001, 1997). Because this region is well conserved across species (Cramer et al., 2017), it is likely that it becomes ordered upon binding, similar to the region of MTA1 that winds irregularly across the surface of HDAC1 (MTA1$^{165-226}$). Further, the crosslinks between MBD3^{IDR} and MTA1 involve a loop (MTA1$^{229-236}$) of the MTA1^{dimer} that is not visible in the MTA1-HDAC1 crystal structure. It is possible that this region of MTA1 may also become ordered upon binding MBD3.

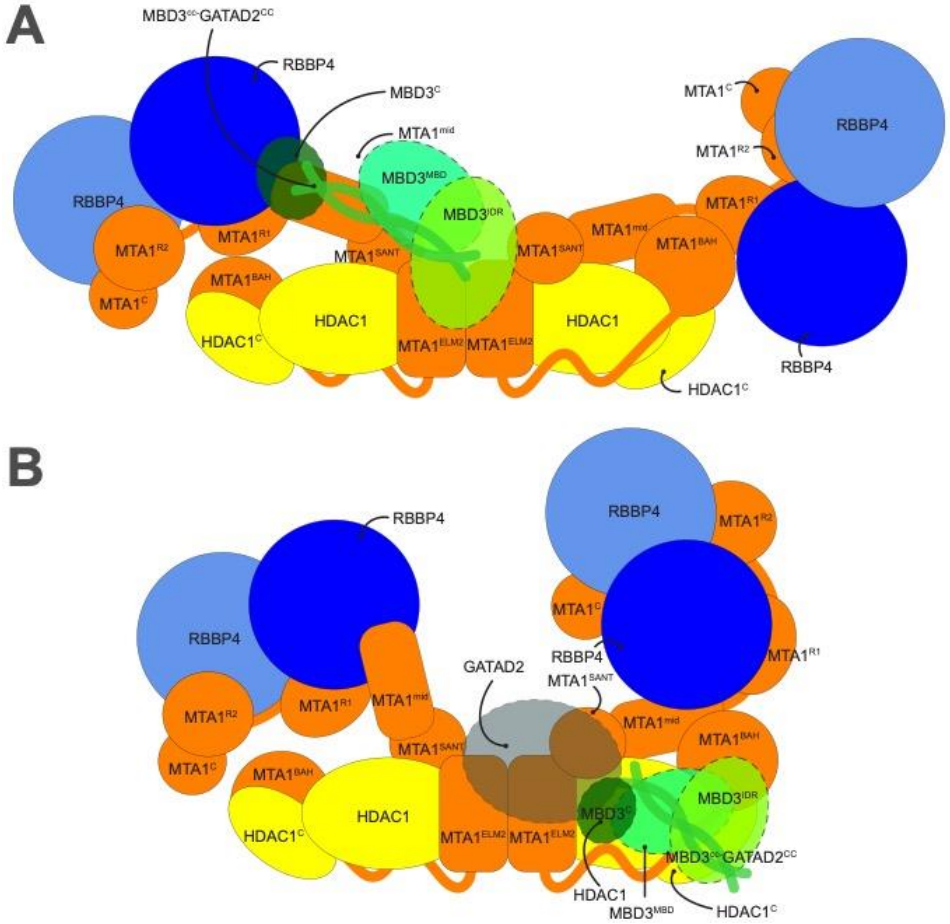

Fig. 7 Bind-and-shift model of MBD3 binding to MHR The figure shows two stages of MBD3 assembly in NuRD. A. In the first stage, the MTA1 dimerization interface is accessible for MBD3^{IDR} to bind. B. In the second stage, upon binding, MBD3 recruits GATAD2 and the chromatin remodeling module and shifts to one end of the MTA1-HDAC1 dimer. GATAD2 localizes near MTA1dimer, precluding a second MBD3 from binding to it.

Bind-and-shift model of MBD3 binding to MHR

The stoichiometry of MBD3 in NuRD is intriguing. The MHM complex has two copies of MBD3, while a single MBD3 is seen in the NuDe and NuRD complexes (Low et al., 2020). Based on our integrative models, we propose a two-stage mechanism to explain the asymmetric binding of MBD3 in NuRD (Fig. 7).

In the first stage (Fig. 7A), the C-terminal arms of MTA1 in MHR are heterogenous and adopt a range of configurations including an extended, open state (Millard et al., 2020) and crossed-over state (Fig. 2, MHR models). In the open state, the MTA1 dimerization interface is accessible for $MBD3^{IDR}$ to bind. This interaction of $MBD3^{IDR}$ with the $MTA1^{dimer}$ is also observed in a cryo-EM study on MTA1:HDAC1:MBD2 and is known to be critical for MBD recruitment to the deacetylase module (Fig. 3, MHM models) (Desai et al., 2015; Millard et al., 2020).

In the second stage (Fig. 7B), upon binding to $MTA1^{dimer}$, MBD3 recruits GATAD2 and the chromatin remodeling module and shifts to one end of the MTA1-HDAC1 dimer (Fig. 4, NuDe models). In this state, GATAD2 localizes near $MTA1^{dimer}$, precluding a second MBD3 from binding to it. Although we did not model full GATAD2 in NuDe due to unavailability of structures and crosslinks involving the protein, the proximity of CHD4, and hence GATAD2, to the $MTA1^{dimer}$ in our coarse nucleosome docking supports this idea (Fig. 6). This possibly explains how GATAD2 introduces asymmetry of MBD3 binding in NuRD. Moreover, upon binding the chromatin remodeling module, the C-terminal arms of MTA1 with the RBBPs are less heterogenous and adopt a closed configuration (Fig. 4, Fig. S12).

The novel NuRD protein interfaces predicted by our model need to be confirmed by future experiments. High-resolution structures of regions such as $MBD3^{IDR}$ will delineate their roles in NuRD. Ultimately, a complete atomic characterization of the NuRD complex will aid in understanding NuRD-mediated regulation of gene expression.

Materials and Methods

Integrative modeling

The integrative structure determination of the NuRD sub-complexes proceeded through four stages (Fig. 1) (Alber et al., 2007; Rout and Sali, 2019; Russel et al., 2012). The modeling protocol (i.e., stages 2, 3, and 4) was scripted using the Python Modeling Interface (PMI) package, a library for modeling macromolecular complexes based on open-source Integrative Modeling Platform (IMP) package, version 2.13.0 (https://integrativemodeling.org) (Russel et al., 2012). The current procedure is an updated version of previously described protocols (Ganesan et al., 2020; Gutierrez et al., 2020; Kim et al., 2018; Saltzberg et al., 2019, 2021; Viswanath et al., 2017a; Webb et al., 2018). Files containing the input data, scripts, and output results are publicly available at https://github.com/isblab/nurd .

Stage 1: Gathering data

The stoichiometry and isoforms of subunits were based on DIA-MS and SEC-MALLS experiments (Fig. S1) (Low et al., 2020). Known atomic structures were used for the MTA1-HDAC1 dimer, $MTA1^{R1}$ and $MTA1^{R2}$ domains in complex with RBBP4, and MBD domain of MBD3 (Fig. S1) (Alqarni et al., 2014; Cramer et al., 2014; Millard et al., 2016, 2013). The $MTA1^{BAH}$ domain, $MTA1^{H}$, $MTA1^{ZF}$, and $MBD3^{CC}$-$GATAD2A^{CC}$ structures were homology-modeled based on the structures of related templates (Fig. S1A) (Connelly et al., 2006; Gnanapragasam et al., 2011; Tjandra et al., 1997).

The shapes of the complexes were based on 3D negative-stain EM maps; MHR: to be deposited (24.56 Å), MHM: EMD-21382 (20 Å), and NuDe: EMD-22904 (20 Å) (Low et al., 2020). The negative-stained EM map for the MHR complex was produced by further analysis of data reported in a previous study (Fig. S2) (Low et al., 2020). 25,155 particle images were subjected to multiple rounds of 2D classification in CryoSparc (Punjani et al., 2017), following which an *ab initio* 3D reconstruction was obtained and refined by homogenous 3D refinement. The final map was produced from 13,299 particles and had an estimated resolution of ~25 Å according to the FSC0.143 criterion.

Chemical crosslinks informed the relative localization of the NuRD subunits. A total of 936 crosslinks, including 877 BS3DSS (bis(sulfosuccinimidyl)suberate - disuccinimidyl suberate), 40 DMTMM (dimethoxy triazinyl methyl-morpholinium chloride), and 19 ADH (adipic acid dihydrazide) crosslinks were used (Low et al., 2020). The models were validated by independent EM maps (Millard et al., 2020), biochemical assays (Desai et al., 2015; Pflum et al., 2001; Zhang et al., 1999), and human cancer-associated mutations on NuRD proteins (Forbes et al., 2006).

Stage 2: Representing the system and translating data into spatial restraints

The stoichiometry and representation of subunits is shown (Fig. S1). The domains with known atomic structures were represented in a multi-scale manner with 1 and 10 residues per bead to maximize computational efficiency. These domains were modeled as rigid bodies where the relative distances between beads is constrained during sampling. In contrast, domains without known structure were coarse-grained at 30 residues per bead and modeled as flexible strings of beads.

We next encoded the spatial restraints into a scoring function based on the information gathered in Stage 1, as follows:

(1) *Cross-link restraints:* The Bayesian cross-links restraint (Rieping et al., 2005) was used to restrain the distances spanned by the cross-linked residues (Shi et al., 2014).

(2) *EM restraints:* The Bayesian EM density restraint was used to restrain the shape of the modeled complexes and was based on the cross-correlation between the Gaussian Mixture Model (GMM) representations of the NuRD subunits and the GMM representation of the corresponding negative-stain EM density maps (Bonomi et al., 2019).

(3) *Excluded volume restraints*: The excluded volume restraints were applied to each bead, using the statistical relationship between the volume and the number of residues that it covered (Alber et al., 2007).

(4) *Sequence connectivity restraints*: We applied the sequence connectivity restraints, using a harmonic upper distance bound on the distance between consecutive beads in a subunit, with a threshold distance equal to twice the sum of the radii of the two connected beads. The bead radius was calculated from the excluded volume of the corresponding bead, assuming standard protein density (Shi et al., 2014).

Stage 3: Structural sampling to produce an ensemble of structures that satisfies the restraints

We aimed to maximize the precision at which the sampling of good-scoring solutions was exhaustive (Stage 4). The sampling runs relied on Gibbs sampling, based on the Replica Exchange Monte Carlo algorithm (Saltzberg et al., 2019, 2021). The positions of the rigid bodies (domains with known structure) and flexible beads (domains with unknown structure) were sampled.

The initial positions of the flexible beads and rigid bodies in all complexes were randomized, with one exception. For MHR, we were able to unambiguously dock the structure of the MTA1-HDAC1 core in the EM map, with the help of the previous EM map (EMD-3399) (Millard et al., 2016). Hence, the position of the corresponding rigid body was fixed throughout.

The Monte Carlo moves included random translations of individual beads in the flexible segments and rigid bodies (around 3.7 Å and 1.3 Å respectively). A model was saved every 10 Gibbs sampling steps, each consisting of a cycle of Monte Carlo steps that moved every bead and rigid body once.

The sampling produced a total of 40 million MHR, 48 million MHM, and 80 million NuDe integrative models.

Stage 4: Analysing and validating the ensemble of structures and data

The sampled models were analysed to assess sampling exhaustiveness and estimate the precision of the structure, its consistency with input data and consistency with data not used in modeling. The structure was further validated by experiments based on the predictions from the models. We used the analysis and validation protocol published earlier (Rout and Sali, 2019; Saltzberg et al., 2019, 2021; Viswanath et al., 2017b). Assessment began with a test of the thoroughness of structural sampling, including structural clustering of the models, estimating model precision, and visualizing the variability in the ensemble of structures using localization probability density maps (Viswanath et al., 2017b). The precision of a domain refers to its positional variation in an ensemble of superposed models. It can also be visualized by the localization probability density map for the domain. A localization probability density map specifies the probability of a voxel (3D volume unit) being occupied by a bead in a set of superposed models. The models and densities were visualized with UCSF Chimera and ChimeraX (Pettersen et al., 2021, 2004).

(1) *Determining good-scoring models*

Starting from the millions of sampled models, first, we selected models obtained after score equilibration and clustered them based on the restraint scores (Saltzberg et al., 2021). For further analysis, we considered 15200 MHR, 21960 MHM, and 21632 NuDe good-scoring models that satisfy the data restraints sufficiently well.

(2) *Clustering and structure precision*

We next assessed the sampling exhaustiveness and performed structural clustering (Saltzberg et al., 2019, 2021; Viswanath et al., 2017b). Integrative structure determination resulted in effectively a single cluster for all complexes, at a precision of 26.8 Å (MHR), 23.8 Å (MHM), and 34.6 Å (NuDe). The cluster precision is the bead RMSD from the cluster centroid model averaged over all models in the cluster (Viswanath et al., 2017b).

(3) *Fit to input information*

The dominant clusters from each modeled NuRD sub-complex satisfied over 95% of all the BS3DSS, ADH, and DMTMM crosslinks used; a crosslink is satisfied by a cluster of models if the corresponding Cα-Cα distance in any model in the cluster is less than 35Å, 35Å, 25Å for BS3DSS, ADH, and DMTMM crosslinks respectively. The agreement between the models and the corresponding EM maps was computed by calculating the cross-correlation of the combined localization probability densities of all subunits for the major cluster with the experimental EM map using the *fitmap* tool in UCSF Chimera (Fig. 2-4) (Pettersen et al., 2004). The remainder of the restraints are harmonic, with a specified standard deviation. The cluster generally satisfied the excluded volume and sequence connectivity restraints. A restraint is satisfied by a cluster of models if the restrained distance in any model in the cluster (considering restraint ambiguity) is violated by less than 3 standard deviations, specified for the restraint. Most of the violations are small, and can be rationalized by local structural fluctuations, coarse-grained representation of the model, and/or finite structural sampling.

(4) *Fit to data not used in modeling*

The MHR integrative models were supported by histone deacetylation assays, mutagenesis, and co-immunoprecipitation, showing that MTA1 and the HDAC1^{C} regulate HDAC1 deacetylase activity and NuRD assembly (Pflum et al., 2001; Zhang et al., 1999). The localization of domains such as MTA1^{BAH} and RBBP4 were validated by their consistency with independently determined cryo-EM maps (Millard et al., 2020).

The MHM integrative models were supported by independent cryo-EM maps of the complex showing similar localizations for MBD2^{MBD} and MTA1^{BAH} (Millard et al., 2020). The MBD3^{IDR}-MTA1^{dimer} interaction was also supported by two separate mutagenesis and co-immunoprecipitation studies (Desai et al., 2015; Millard et al., 2020).

The NuDe integrative models were corroborated by immunoprecipitation experiments showing that the MBD domain of MBD3 is buried in NuRD (Zhang et al., 1999). They were also supported by independent cryo-EM maps showing that MBD3 is proximal to MTA1^{BAH}, and biochemical assays showing the importance of HDAC1^{C} interactions in NuRD (Millard et al., 2020; Pflum et al., 2001). The mapping of cancer mutations to protein-protein interfaces in the NuDe model also supported them (Fig. 5, Fig. S13, Table S1) (Forbes et al., 2006).

4.1 Mapping COSMIC mutations

We obtained a total of 356 somatic, confirmed pathogenic, point mutations for the modeled NuRD subunits (MTA1, HDAC1, RBBP4, MBD3) from the COSMIC (Catalogue of Somatic Mutations in Cancer) database (Forbes et al., 2006). For each subunit, point mutations were selected from search results based on the presence of census genes and correct documentation of current structures. To ensure the mutations studied significantly affect the function, folding, and protein-protein interaction of the protein, the “confirmed pathogenic” and “somatic” filters were applied in all cases.

Acknowledgements

We thank Vinothkumar Kutti Ragunath and lab members Satwik Pasani, Praveen Roy DS, and Varun Ullanat for useful comments on the manuscript.

Molecular graphics images were produced using the UCSF Chimera and UCSF ChimeraX packages from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081).

Funding

This work has been supported by the following grants: Department of Atomic Energy (DAE) TIFR grant RTI 4006 and Department of Science and Technology (DST) SERB grant SPG/2020/000475 from the Government of India to S.V, National Health and Medical Research Council of Australia project grants: APP1012161, APP1063301, APP1126357 to M.J.L. and J.P.M. and a fellowship (APP1058916) from the same organization to J.P.M.

Conflict of Interest

None declared.

References

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., Rout, M.P., Sali, A., 2007. Determining the architectures of macromolecular assemblies. Nature 450, 683–694. https://doi.org/10.1038/nature06404

Alqarni, S.S.M., Murthy, A., Zhang, W., Przewloka, M.R., Silva, A.P.G., Watson, A.A., Lejon, S., Pei, X.Y., Smits, A.H., Kloet, S.L., Wang, H., Shepherd, N.E., Stokes, P.H., Blobel, G.A., Vermeulen, M., Glover, D.M., Mackay, J.P., Laue, E.D., 2014. Insight into the Architecture of the NuRD Complex. J. Biol. Chem. 289, 21844–21855. https://doi.org/10.1074/jbc.M114.558940

Basta, J., Rauchman, M., 2017. Chapter 3 - The Nucleosome Remodeling and Deacetylase Complex in Development and Disease, in: Laurence, J., Beusekom, M.V. (Eds.), Translating Epigenetics to the Clinic. Academic Press, Boston, pp. 37–72. https://doi.org/10.1016/B978-0-12-800802-7.00003-4

Basta, J., Rauchman, M., 2015. The nucleosome remodeling and deacetylase complex in development and disease. Transl. Res. J. Lab. Clin. Med. 165, 36–47. https://doi.org/10.1016/j.trsl.2014.05.003

Bode, D., Yu, L., Tate, P., Pardo, M., Choudhary, J., 2016. Characterization of Two Distinct Nucleosome Remodeling and Deacetylase (NuRD) Complex Assemblies in Embryonic Stem Cells*. Mol. Cell. Proteomics 15, 878–891. https://doi.org/10.1074/mcp.M115.053207

Bonomi, M., Hanot, S., Greenberg, C.H., Sali, A., Nilges, M., Vendruscolo, M., Pellarin, R., 2019. Bayesian Weighing of Electron Cryo-Microscopy Data for Integrative Structural Modeling. Structure 27, 175-188.e6. https://doi.org/10.1016/j.str.2018.09.011

Bornelöv, S., Reynolds, N., Xenophontos, M., Gharbi, S., Johnstone, E., Floyd, R., Ralser, M., Signolet, J., Loos, R., Dietmann, S., Bertone, P., Hendrich, B., 2018. The Nucleosome Remodeling and Deacetylation Complex Modulates Chromatin Structure at Sites of Active Transcription to Fine-Tune Gene Expression. Mol. Cell 71, 56-72.e4. https://doi.org/10.1016/j.molcel.2018.06.003

Burgold, T., Barber, M., Kloet, S., Cramard, J., Gharbi, S., Floyd, R., Kinoshita, M., Ralser, M., Vermeulen, M., Reynolds, N., Dietmann, S., Hendrich, B., 2019. The Nucleosome Remodelling

667 and Deacetylation complex suppresses transcriptional noise during lineage commitment. EMBO
668 J. 38, e100788. https://doi.org/10.15252/embj.2018100788
669 Connelly, J.J., Yuan, P., Hsu, H.-C., Li, Z., Xu, R.-M., Sternglanz, R., 2006. Structure and Function of
670 the Saccharomyces cerevisiae Sir3 BAH Domain. Mol. Cell. Biol. 26, 3256–3265.
671 https://doi.org/10.1128/MCB.26.8.3256-3265.2006
672 Cramer, J.M., Pohlmann, D., Gomez, F., Mark, L., Kornegay, B., Hall, C., Siraliev-Perez, E., Walavalkar,
673 N.M., Sperlazza, M.J., Bilinovich, S., Prokop, J.W., Hill, A.L., Williams, D.C., 2017.
674 Methylation specific targeting of a chromatin remodeling complex from sponges to humans. Sci.
675 Rep. 7, 40674. https://doi.org/10.1038/srep40674
676 Cramer, J.M., Scarsdale, J.N., Walavalkar, N.M., Buchwald, W.A., Ginder, G.D., Williams, D.C., 2014.
677 Probing the Dynamic Distribution of Bound States for Methylcytosine-binding Domains on
678 DNA* *This work was supported, in whole or in part, by National Institutes of Health Grants
679 R01 GM098624 (to D. C. W.) and R01 DK029902 (to G. D. G.). NMR data were acquired using
680 NMR instrumentation in the Virginia Commonwealth University Massey Cancer Center
681 Structural Biology Core, and surface plasmon resonance data were acquired in the Virginia
682 Commonwealth University Massey Cancer Center Flow Cytometry Core, both of which are
683 supported, in part, by NCI Cancer Center Core Support Grant P30 CA016059. J. Biol. Chem.
684 289, 1294–1302. https://doi.org/10.1074/jbc.M113.512236
685 Denslow, S.A., Wade, P.A., 2007. The human Mi-2/NuRD complex and gene regulation. Oncogene 26,
686 5433–5438. https://doi.org/10.1038/sj.onc.1210611
687 Desai, M.A., Webb, H.D., Sinanan, L.M., Scarsdale, J.N., Walavalkar, N.M., Ginder, G.D., Williams,
688 D.C., 2015. An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2)
689 recruits the histone deacetylase core of the NuRD complex. Nucleic Acids Res. 43, 3100–3113.
690 https://doi.org/10.1093/nar/gkv168
691 Farnung, L., Ochmann, M., Cramer, P., 2020. Nucleosome-CHD4 chromatin remodeler structure maps
692 human disease mutations. eLife 9, e56178. https://doi.org/10.7554/eLife.56178
693 Forbes, S., Clements, J., Dawson, E., Bamford, S., Webb, T., Dogan, A., Flanagan, A., Teague, J.,
694 Wooster, R., Futreal, P.A., Stratton, M.R., 2006. Cosmic 2005. Br. J. Cancer 94, 318–322.
695 https://doi.org/10.1038/sj.bjc.6602928
696 Ganesan, S.J., Feyder, M.J., Chemmama, I.E., Fang, F., Rout, M.P., Chait, B.T., Shi, Y., Munson, M.,
697 Sali, A., 2020. Integrative structure and function of the yeast exocyst complex. Protein Sci. Publ.
698 Protein Soc. 29, 1486–1501. https://doi.org/10.1002/pro.3863
699 Gnanapragasam, M.N., Scarsdale, J.N., Amaya, M.L., Webb, H.D., Desai, M.A., Walavalkar, N.M.,
700 Wang, S.Z., Zu Zhu, S., Ginder, G.D., Williams, D.C., 2011. p66α–MBD2 coiled-coil interaction
701 and recruitment of Mi-2 are critical for globin gene silencing by the MBD2–NuRD complex.
702 Proc. Natl. Acad. Sci. 108, 7487–7492.
703 Guo, T., Luna, A., Rajapakse, V.N., Koh, C.C., Wu, Z., Liu, W., Sun, Y., Gao, H., Menden, M.P., Xu,
704 C., Calzone, L., Martignetti, L., Auwerx, C., Buljan, M., Banaei-Esfahani, A., Ori, A., Iskar, M.,
705 Gillet, L., Bi, R., Zhang, J., Zhang, H., Yu, C., Zhong, Q., Varma, S., Schmitt, U., Qiu, P., Zhang,
706 Q., Zhu, Y., Wild, P.J., Garnett, M.J., Bork, P., Beck, M., Liu, K., Saez-Rodriguez, J., Elloumi,
707 F., Reinhold, W.C., Sander, C., Pommier, Y., Aebersold, R., 2019. Quantitative Proteome
708 Landscape of the NCI-60 Cancer Cell Lines. iScience 21, 664–680.
709 https://doi.org/10.1016/j.isci.2019.10.059
710 Gutierrez, C., Chemmama, I.E., Mao, H., Yu, C., Echeverria, I., Block, S.A., Rychnovsky, S.D., Zheng,
711 N., Sali, A., Huang, L., 2020. Structural dynamics of the human COP9 signalosome revealed by
712 cross-linking mass spectrometry and integrative modeling. Proc. Natl. Acad. Sci. 117, 4088–
713 4098. https://doi.org/10.1073/pnas.1915542117
714 Hoffmann, A., Spengler, D., 2019. Chromatin Remodeling Complex NuRD in Neurodevelopment and
715 Neurodevelopmental Disorders. Front. Genet. 0. https://doi.org/10.3389/fgene.2019.00682

Hong, W., Nakazawa, M., Chen, Y.-Y., Kori, R., Vakoc, C.R., Rakowski, C., Blobel, G.A., 2005. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. EMBO J. 24, 2367–2378. https://doi.org/10.1038/sj.emboj.7600703

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kim, S.J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W., Raveh, B., Herricks, T., Slaughter, B.D., Hogan, J.A., Upla, P., Chemmama, I.E., Pellarin, R., Echeverria, I., Shivaraju, M., Chaudhury, A.S., Wang, J., Williams, R., Unruh, J.R., Greenberg, C.H., Jacobs, E.Y., Yu, Z., de la Cruz, M.J., Mironska, R., Stokes, D.L., Aitchison, J.D., Jarrold, M.F., Gerton, J.L., Ludtke, S.J., Akey, C.W., Chait, B.T., Sali, A., Rout, M.P., 2018. Integrative structure and functional anatomy of a nuclear pore complex. Nature 555, 475–482. https://doi.org/10.1038/nature26003

Kloet, S.L., Baymaz, H.I., Makowski, M., Groenewold, V., Jansen, P.W.T.C., Berendsen, M., Niazi, H., Kops, G.J., Vermeulen, M., 2015. Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. FEBS J. 282, 1774–1785. https://doi.org/10.1111/febs.12972

Lejon, S., Thong, S.Y., Murthy, A., AlQarni, S., Murzina, N.V., Blobel, G.A., Laue, E.D., Mackay, J.P., 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. https://doi.org/10.1074/jbc.M110.195842

Li, D.-Q., Kumar, R., 2010. Mi-2/NuRD complex making inroads into DNA-damage response pathway. Cell Cycle Georget. Tex 9, 2071–2079. https://doi.org/10.4161/cc.9.11.11735

Low, J.K.K., Silva, A.P.G., Sharifi Tabar, M., Torrado, M., Webb, S.R., Parker, B.L., Sana, M., Smits, C., Schmidberger, J.W., Brillault, L., Jackman, M.J., Williams, D.C., Blobel, G.A., Hake, S.B., Shepherd, N.E., Landsberg, M.J., Mackay, J.P., 2020. The Nucleosome Remodeling and Deacetylase Complex Has an Asymmetric, Dynamic, and Modular Architecture. Cell Rep. 33, 108450. https://doi.org/10.1016/j.celrep.2020.108450

Millard, C.J., Fairall, L., Ragan, T.J., Savva, C.G., Schwabe, J.W.R., 2020. The topology of chromatin-binding domains in the NuRD deacetylase complex. Nucleic Acids Res. 48, 12972–12982. https://doi.org/10.1093/nar/gkaa1121

Millard, C.J., Varma, N., Saleh, A., Morris, K., Watson, P., Bottrill, A., Fairall, L., Smith, C., Schwabe, J., 2016. The structure of the core NuRD repression complex provides insights into its interaction with chromatin. ELife 2016. https://doi.org/10.7554/eLife.13941

Millard, C.J., Watson, P.J., Celardo, I., Gordiyenko, Y., Cowley, S.M., Robinson, C.V., Fairall, L., Schwabe, J.W.R., 2013. Class I HDACs Share a Common Mechanism of Regulation by Inositol Phosphates. Mol. Cell 51, 57–67. https://doi.org/10.1016/j.molcel.2013.05.020

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., Ferrin, T.E., 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82. https://doi.org/10.1002/pro.3943

Pflum, M.K.H., Tong, J.K., Lane, W.S., Schreiber, S.L., 2001. Histone Deacetylase 1 Phosphorylation Promotes Enzymatic Activity and Complex Formation*. J. Biol. Chem. 276, 47733–47741. https://doi.org/10.1074/jbc.M105590200

765 Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A., 2017. cryoSPARC: algorithms for rapid
766 unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296.
767 https://doi.org/10.1038/nmeth.4169
768 Rathert, P., Dhayalan, A., Murakami, M., Zhang, X., Tamas, R., Jurkowska, R., Komatsu, Y., Shinkai,
769 Y., Cheng, X., Jeltsch, A., 2008. Protein lysine methyltransferase G9a acts on non-histone targets.
770 Nat. Chem. Biol. 4, 344–346. https://doi.org/10.1038/nchembio.88
771 Reynolds, N., O'Shaughnessy, A., Hendrich, B., 2013. Transcriptional repressors: multifaceted
772 regulators of gene expression. Development 140, 505–512. https://doi.org/10.1242/dev.083105
773 Rieping, W., Habeck, M., Nilges, M., 2005. Inferential Structure Determination. Science 309, 303–306.
774 https://doi.org/10.1126/science.1110428
775 Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J.E., Dunker, A.K., 1997. Identifying disordered
776 regions in proteins from amino acid sequence, in: Proceedings of International Conference on
777 Neural Networks (ICNN'97). Presented at the Proceedings of International Conference on Neural
778 Networks (ICNN'97), pp. 90–95 vol.1. https://doi.org/10.1109/ICNN.1997.611643
779 Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., Dunker, A.K., 2001. Sequence complexity
780 of disordered protein. Proteins 42, 38–48. https://doi.org/10.1002/1097-
781 0134(20010101)42:1<38::aid-prot50>3.0.co;2-3
782 Rout, M.P., Sali, A., 2019. Principles for Integrative Structural Biology Studies. Cell 177, 1384–1403.
783 https://doi.org/10.1016/j.cell.2019.05.016
784 Russel, D., Lasker, K., Webb, B., Velázquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson,
785 B., Sali, A., 2012. Putting the Pieces Together: Integrative Modeling Platform Software for
786 Structure Determination of Macromolecular Assemblies. PLoS Biol. 10, e1001244.
787 https://doi.org/10.1371/journal.pbio.1001244
788 Saltzberg, D., Greenberg, C.H., Viswanath, S., Chemmama, I., Webb, B., Pellarin, R., Echeverria, I.,
789 Sali, A., 2019. Modeling Biological Complexes Using Integrative Modeling Platform, in:
790 Bonomi, M., Camilloni, C. (Eds.), Biomolecular Simulations: Methods and Protocols, Methods
791 in Molecular Biology. Springer, New York, NY, pp. 353–377. https://doi.org/10.1007/978-1-
792 4939-9608-7_15
793 Saltzberg, D.J., Viswanath, S., Echeverria, I., Chemmama, I.E., Webb, B., Sali, A., 2021. Using
794 Integrative Modeling Platform to compute, validate, and archive a model of a protein complex
795 structure. Protein Sci. 30, 250–261. https://doi.org/10.1002/pro.3995
796 Schneidman-Duhovny, D., Pellarin, R., Sali, A., 2014. Uncertainty in integrative structural modeling.
797 Curr. Opin. Struct. Biol. 28, 96–104. https://doi.org/10.1016/j.sbi.2014.08.001
798 Sharifi Tabar, M., Mackay, J.P., Low, J.K.K., 2019. The stoichiometry and interactome of the
799 Nucleosome Remodeling and Deacetylase (NuRD) complex are conserved across multiple cell
800 lines. FEBS J. 286, 2043–2061. https://doi.org/10.1111/febs.14800
801 Shi, Y., Fernandez-Martinez, J., Tjioe, E., Pellarin, R., Kim, S.J., Williams, R., Schneidman-Duhovny,
802 D., Sali, A., Rout, M.P., Chait, B.T., 2014. Structural characterization by cross-linking reveals
803 the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex.
804 Mol. Cell. Proteomics MCP 13, 2927–2943. https://doi.org/10.1074/mcp.M114.041673
805 Smeenk, G., Wiegant, W.W., Vrolijk, H., Solari, A.P., Pastink, A., van Attikum, H., 2010. The NuRD
806 chromatin–remodeling complex regulates signaling and repair of DNA damage. J. Cell Biol. 190,
807 741–749. https://doi.org/10.1083/jcb.201001048
808 Smits, A.H., Jansen, P.W.T.C., Poser, I., Hyman, A.A., Vermeulen, M., 2013. Stoichiometry of
809 chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-
810 based proteomics. Nucleic Acids Res. 41, e28. https://doi.org/10.1093/nar/gks941
811 Söding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection
812 and structure prediction. Nucleic Acids Res. 33, W244-248. https://doi.org/10.1093/nar/gki408

Spruijt, C.G., Luijsterburg, M.S., Menafra, R., Lindeboom, R.G.H., Jansen, P.W.T.C., Edupuganti, R.R., Baltissen, M.P., Wiegant, W.W., Voelker-Albert, M.C., Matarese, F., Mensinga, A., Poser, I., Vos, H.R., Stunnenberg, H.G., van Attikum, H., Vermeulen, M., 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. https://doi.org/10.1016/j.celrep.2016.09.037

Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M., Bax, A., 1997. Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat. Struct. Biol. 4, 732–738. https://doi.org/10.1038/nsb0997-732

Toh, Y., Nicolson, G.L., 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. https://doi.org/10.1007/s10585-008-9233-8

Viswanath, S., Bonomi, M., Kim, S.J., Klenchin, V.A., Taylor, K.C., Yabut, K.C., Umbreit, N.T., Van Epps, H.A., Meehl, J., Jones, M.H., Russel, D., Velazquez-Muriel, J.A., Winey, M., Rayment, I., Davis, T.N., Sali, A., Muller, E.G., 2017a. The molecular architecture of the yeast spindle pole body core determined by Bayesian integrative modeling. Mol. Biol. Cell 28, 3298–3314. https://doi.org/10.1091/mbc.E17-06-0397

Viswanath, S., Chemmama, I.E., Cimermancic, P., Sali, A., 2017b. Assessing Exhaustiveness of Stochastic Sampling for Integrative Modeling of Macromolecular Structures. Biophys. J. 113, 2344–2353. https://doi.org/10.1016/j.bpj.2017.10.005

Ward, A.B., Sali, A., Wilson, I.A., 2013. Integrative Structural Biology. Science 339, 913–915. https://doi.org/10.1126/science.1228565

Webb, B., Viswanath, S., Bonomi, M., Pellarin, R., Greenberg, C.H., Saltzberg, D., Sali, A., 2018. Integrative structure modeling with the Integrative Modeling Platform. Protein Sci. 27, 245–258. https://doi.org/10.1002/pro.3311

Yoshida, T., Hazan, I., Zhang, J., Ng, S.Y., Naito, T., Snippert, H.J., Heller, E.J., Qi, X., Lawton, L.N., Williams, C.J., Georgopoulos, K., 2008. The role of the chromatin remodeler Mi-2β in hematopoietic stem cell self-renewal and multilineage differentiation. Genes Dev. 22, 1174–1189. https://doi.org/10.1101/gad.1642808

Zhang, W., Aubert, A., Gomez de Segura, J.M., Karuppasamy, M., Basu, S., Murthy, A.S., Diamante, A., Drury, T.A., Balmer, J., Cramard, J., Watson, A.A., Lando, D., Lee, S.F., Palayret, M., Kloet, S.L., Smits, A.H., Deery, M.J., Vermeulen, M., Hendrich, B., Klenerman, D., Schaffitzel, C., Berger, I., Laue, E.D., 2016. The Nucleosome Remodeling and Deacetylase Complex NuRD Is Built from Preformed Catalytically Active Sub-modules. J. Mol. Biol. 428, 2931–2942. https://doi.org/10.1016/j.jmb.2016.04.025

Zhang, Y., Ng, H.-H., Erdjument-Bromage, H., Tempst, P., Bird, A., Reinberg, D., 1999. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. 13, 1924–1935.