

Supplementary Figures and Tables

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

Shreyas Arvindekar¹, Matthew J. Jackman², Jason K.K. Low³, Michael J. Landsberg^{2, *}, Joel P. Mackay^{3, *}, and Shruthi Viswanath^{1, *}

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

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

³School 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.)

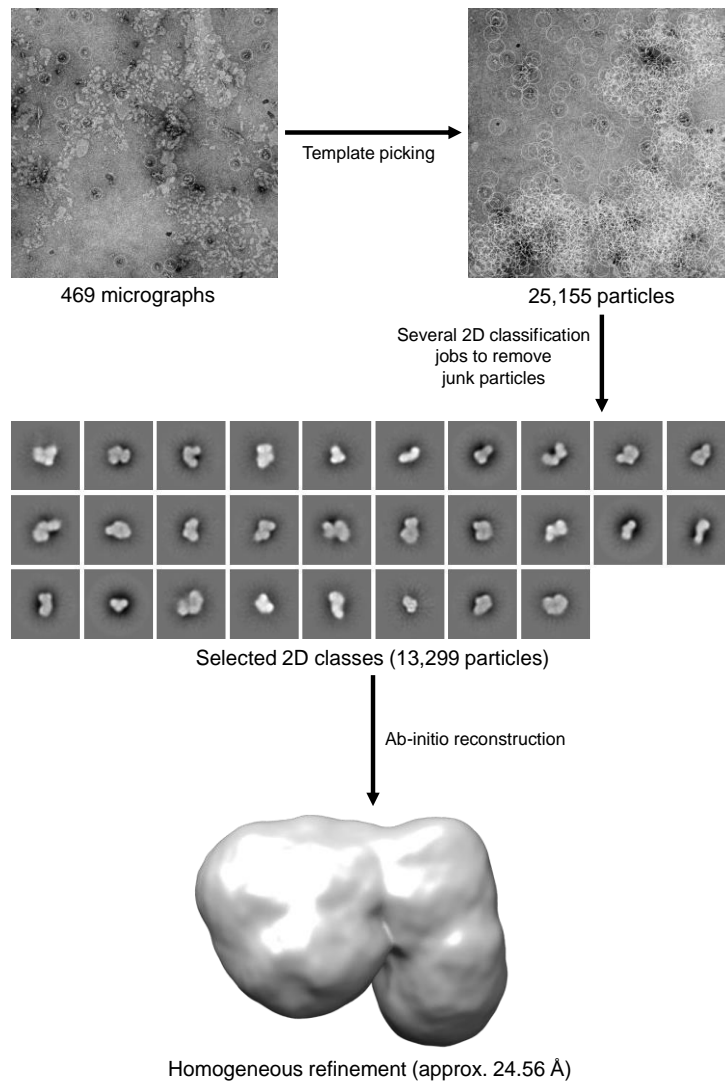


Fig. S1 Single particle analysis of the MHR complex. A flowthrough from the micrographs imported into Cryosparc (Punjani et al., 2017) to the final 3D reconstruction of the MHR complex using the ab-initio reconstruction job. 2D classes were generated using 50 classes per run, with particles from non-junk classes used as input to subsequent 2D classification jobs. Non-uniform refinement was not performed on the final structure due to its low resolution.

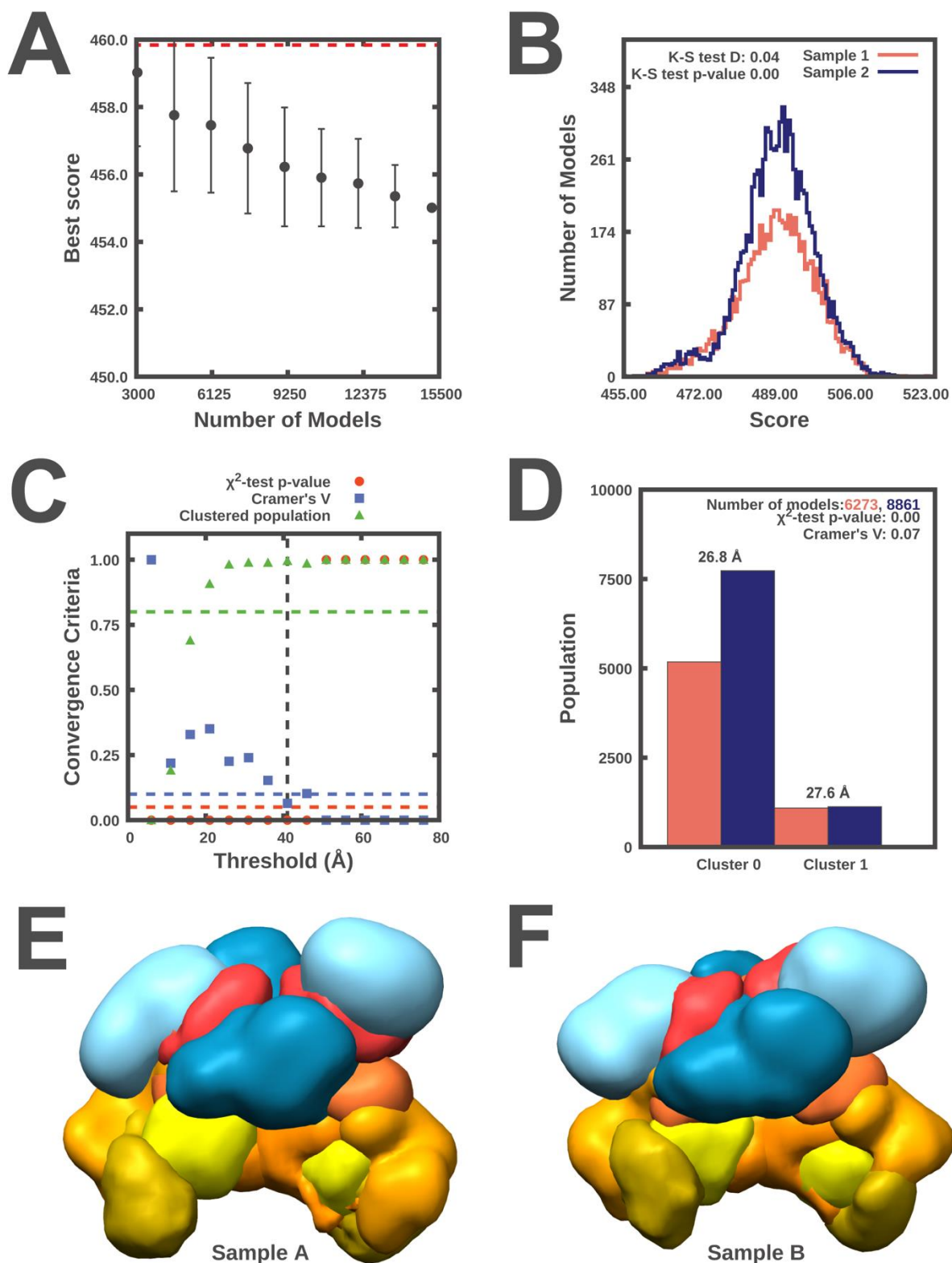


Fig. S2 Sampling exhaustiveness protocol on MHR models Results of test 1, convergence of the model score, for the 15200 good-scoring models; the scores do not continue to improve as more models are computed essentially independently. The error bar represents the standard deviations of the best scores, estimated by repeating sampling of models 10 times. The red dotted line indicates a lower bound reference on the total score. B. Results of test 2, testing similarity of model score distributions between samples 1 (red) and 2 (blue); the difference in the distribution of scores is significant (Kolmogorov-Smirnov two-sample test p-value less than 0.05) but the magnitude of the difference is small (the Kolmogorov-Smirnov two-sample test statistic D is 0.04); thus, the two score distributions are effectively equal. C. Results of test 3, three criteria for determining the sampling precision (Y-axis), evaluated as a function of the RMSD

clustering threshold (X-axis). First, the p-value is computed using the χ^2 -test for homogeneity of proportions (red dots). Second, an effect size for the χ^2 -test is quantified by the Cramer's V value (blue squares). Third, the population of models in sufficiently large clusters (containing at least 10 models from each sample) is shown as green triangles. The vertical dotted grey line indicates the RMSD clustering threshold at which three conditions are satisfied (p-value > 0.05 [dotted red line], Cramer's V < 0.10 [dotted blue line], and the population of clustered models > 0.80 [dotted green line]), thus defining the sampling precision of 41 Å. D. Populations of sample 1 and 2 models in the clusters obtained by threshold-based clustering using the RMSD threshold of 41 Å. Cluster precision is shown for each cluster. E. and F. Results of test 4: comparison of localization probability densities of models from sample A and sample B for the major cluster (84.95% population). The cross-correlation of the density maps of the two samples is greater than 0.96.

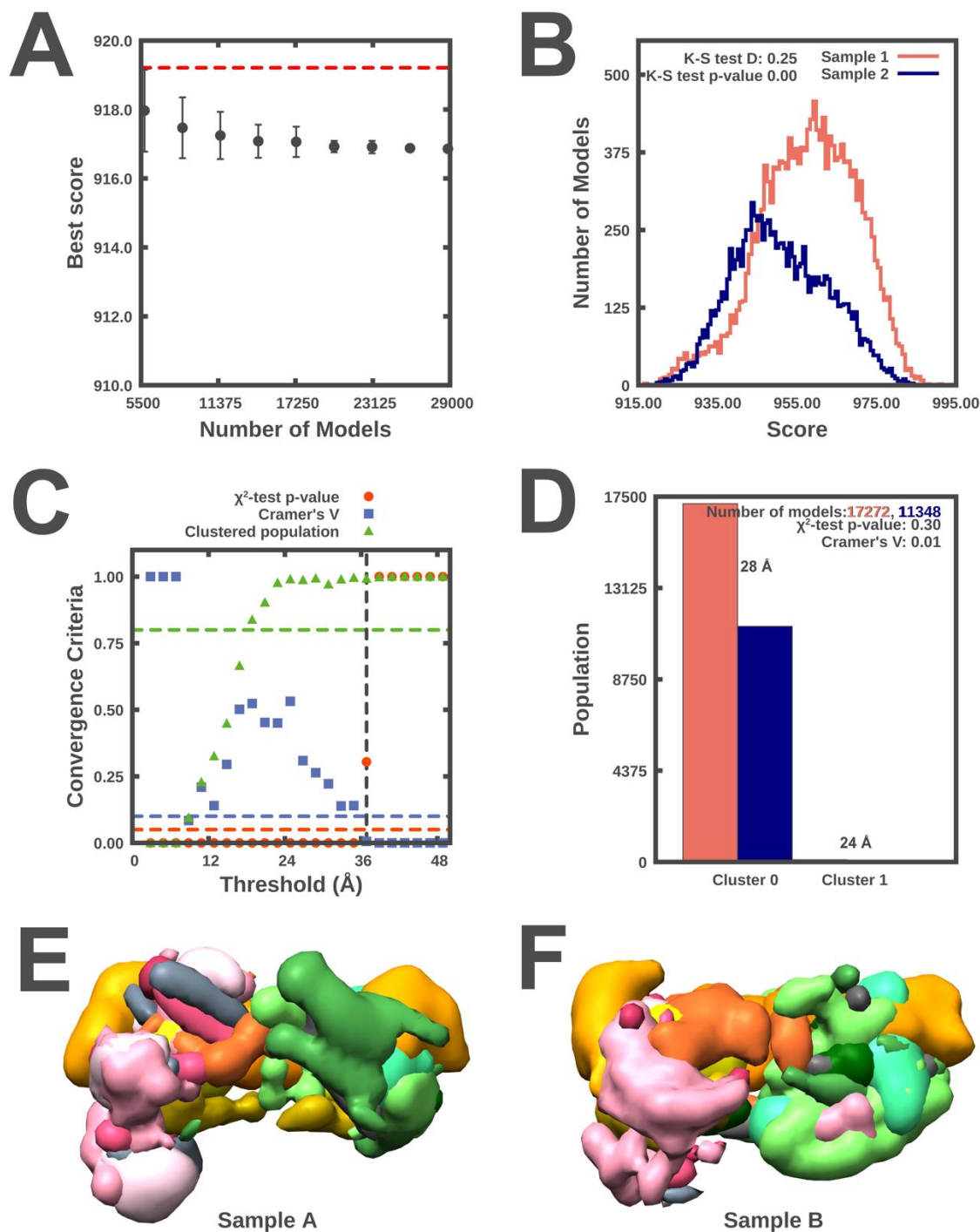


Fig. S3 Sampling exhaustiveness protocol on MTA1^N-HDAC1-MBD3^{GATAD2CC} (MHM) models Results of test 1, convergence of the model score, for the 28836 good-scoring models; the scores do not continue to improve as more models are computed essentially independently. The error bar represents the standard deviations of the best scores, estimated

by repeating sampling of models 10 times. The red dotted line indicates a lower bound reference on the total score. B. Results of test 2, testing similarity of model score distributions between samples 1 (red) and 2 (blue); the difference in the distribution of scores is significant (Kolmogorov-Smirnov two-sample test p-value less than 0.05) but the magnitude of the difference is small (the Kolmogorov-Smirnov two-sample test statistic D is 0.25); thus, the two score distributions are effectively equal. C. Results of test 3, three criteria for determining the sampling precision (Y-axis), evaluated as a function of the RMSD clustering threshold (X-axis). First, the p-value is computed using the χ^2 -test for homogeneity of proportions (red dots). Second, an effect size for the χ^2 -test is quantified by the Cramer's V value (blue squares). Third, the population of models in sufficiently large clusters (containing at least 10 models from each sample) is shown as green triangles. The vertical dotted grey line indicates the RMSD clustering threshold at which three conditions are satisfied (p-value > 0.05 [dotted red line], Cramer's V < 0.10 [dotted blue line], and the population of clustered models > 0.80 [dotted green line]), thus defining the sampling precision of 37 Å. D. Populations of sample 1 and 2 models in the clusters obtained by threshold-based clustering using the RMSD threshold of 37 Å. Cluster precision is shown for each cluster. E. and F. Results of test 4: comparison of localization probability densities of models from sample A and sample B for the major cluster (99% population). The cross-correlation of the density maps of the two samples is 0.99.

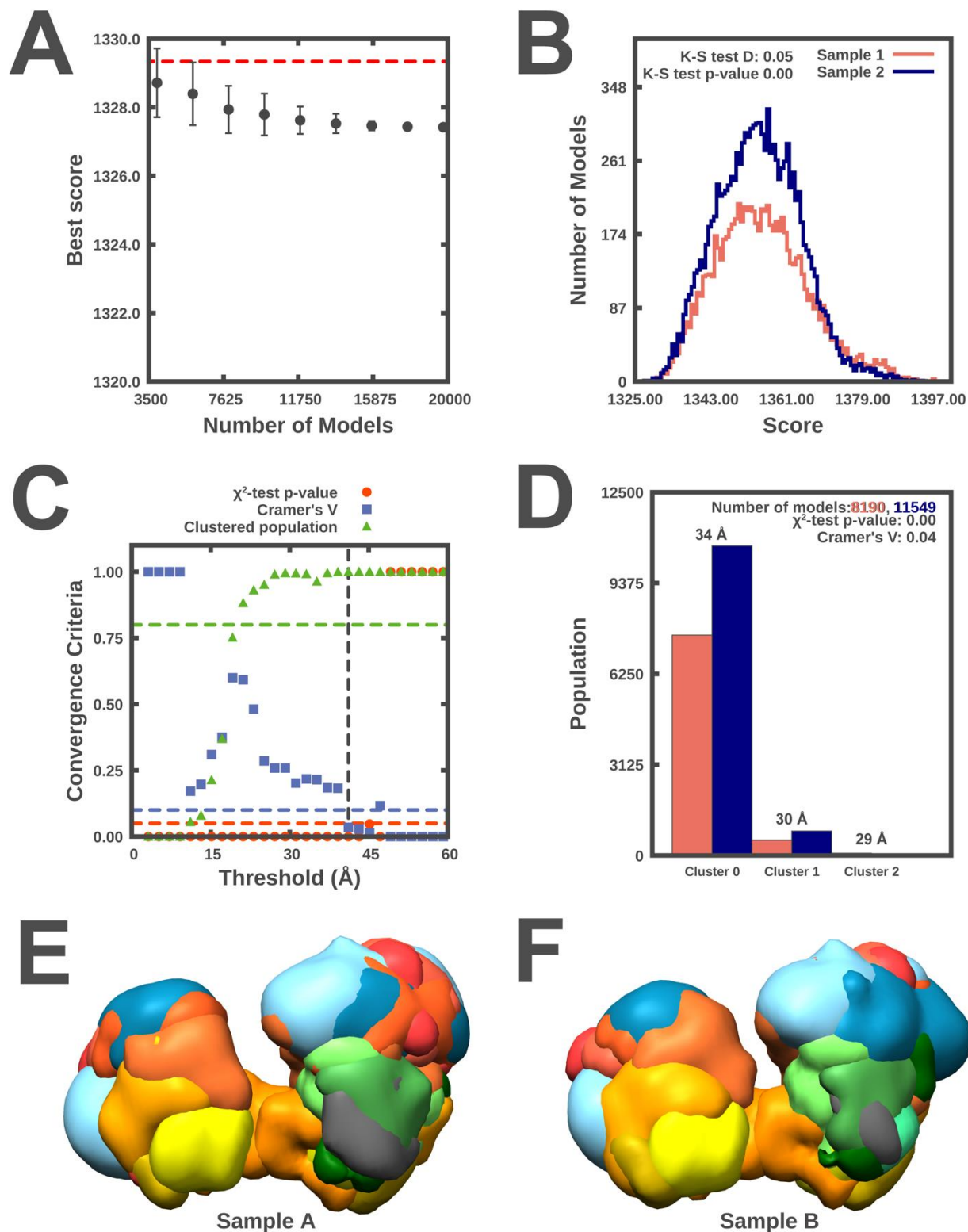


Fig. S4 Sampling exhaustiveness protocol on NuDe integrative models Results of test 1, convergence of the model score, for the 18005 good-scoring models; the scores do not continue to improve as more models are computed essentially independently. The error bar represents the standard deviations of the best scores, estimated by repeating sampling of models 10 times. The red dotted line indicates a lower bound reference on the total score. B. Results of test 2, testing similarity of model score distributions between samples 1 (red) and 2 (blue); the difference in the distribution of scores is significant (Kolmogorov-Smirnov two-sample test p-value less than 0.05) but the magnitude of the difference is small (the Kolmogorov-Smirnov two-sample test statistic D is 0.05); thus, the two score distributions are effectively equal. C. Results of test 3, three criteria for determining the sampling precision (Y-axis), evaluated as a function of the RMSD clustering threshold (X-axis). First, the p-value is computed using the χ^2 -test for homogeneity of proportions (red

dots). Second, an effect size for the χ^2 -test is quantified by the Cramer's V value (blue squares). Third, the population of models in sufficiently large clusters (containing at least 10 models from each sample) is shown as green triangles. The vertical dotted grey line indicates the RMSD clustering threshold at which three conditions are satisfied (p-value > 0.05 [dotted red line], Cramer's V < 0.10 [dotted blue line], and the population of clustered models > 0.80 [dotted green line]), thus defining the sampling precision of 41 Å. D. Populations of sample 1 and 2 models in the clusters obtained by threshold-based clustering using the RMSD threshold of 41 Å. Cluster precision is shown for each cluster. E. and F. Results of test 4: comparison of localization probability densities of models from sample A and sample B for the major cluster (71% population). The cross-correlation of the density maps of the two samples is 0.99.

Key: — <10 Å — <20 Å — <30 Å — <40 Å — >40 Å

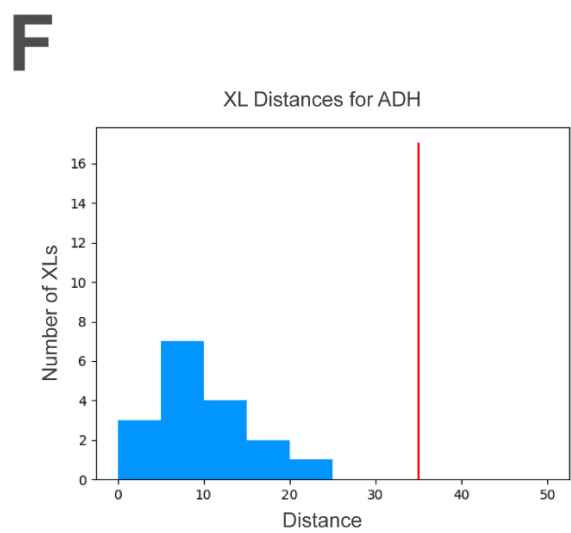
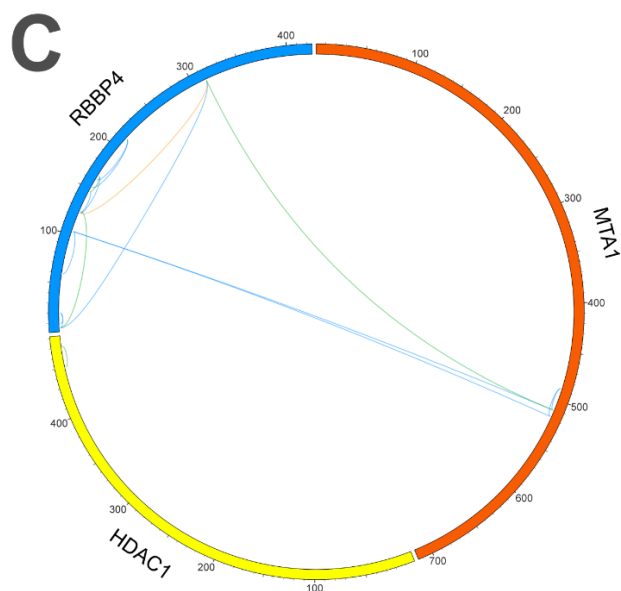
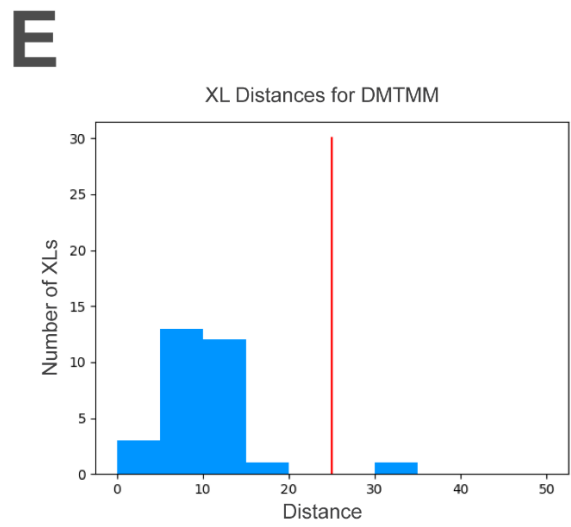
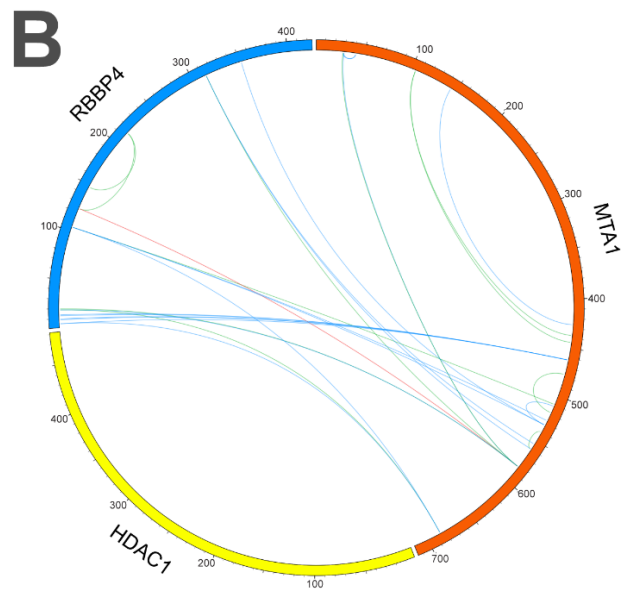
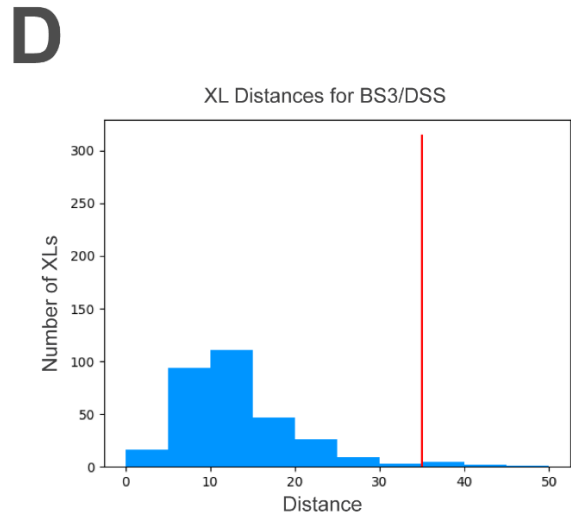
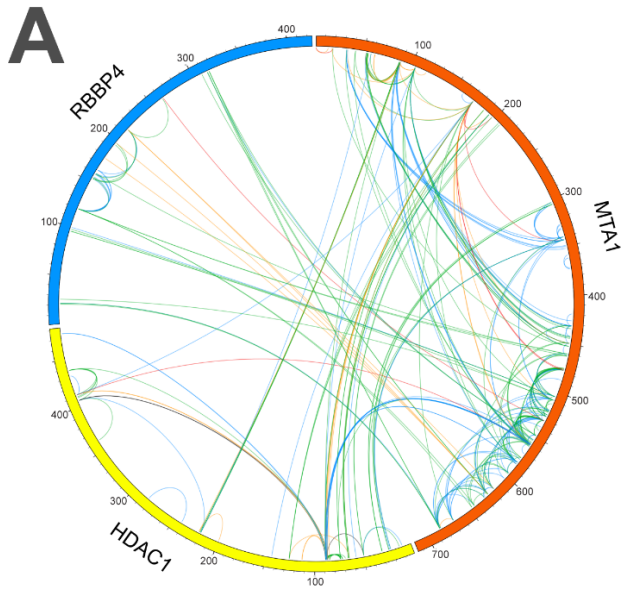


Fig. S5 Results of crosslink fitting for MHR models CX-CIRCOS (<http://cx-circos.net/>) plots are shown for A. BS3/DSS, B. DMTMM, and C. ADH crosslinks on the ensemble of MHR models from the major cluster. Each line depicts a crosslink; its color depicts the minimum distance between the corresponding crosslinked residues in the ensemble, as shown in the color key (top). Histograms showing the distribution of the minimum crosslink distance in the ensemble for D. BS3/DSS, E. DMTMM, and F. ADH crosslinks. The red line indicates the distance threshold for a crosslink type.

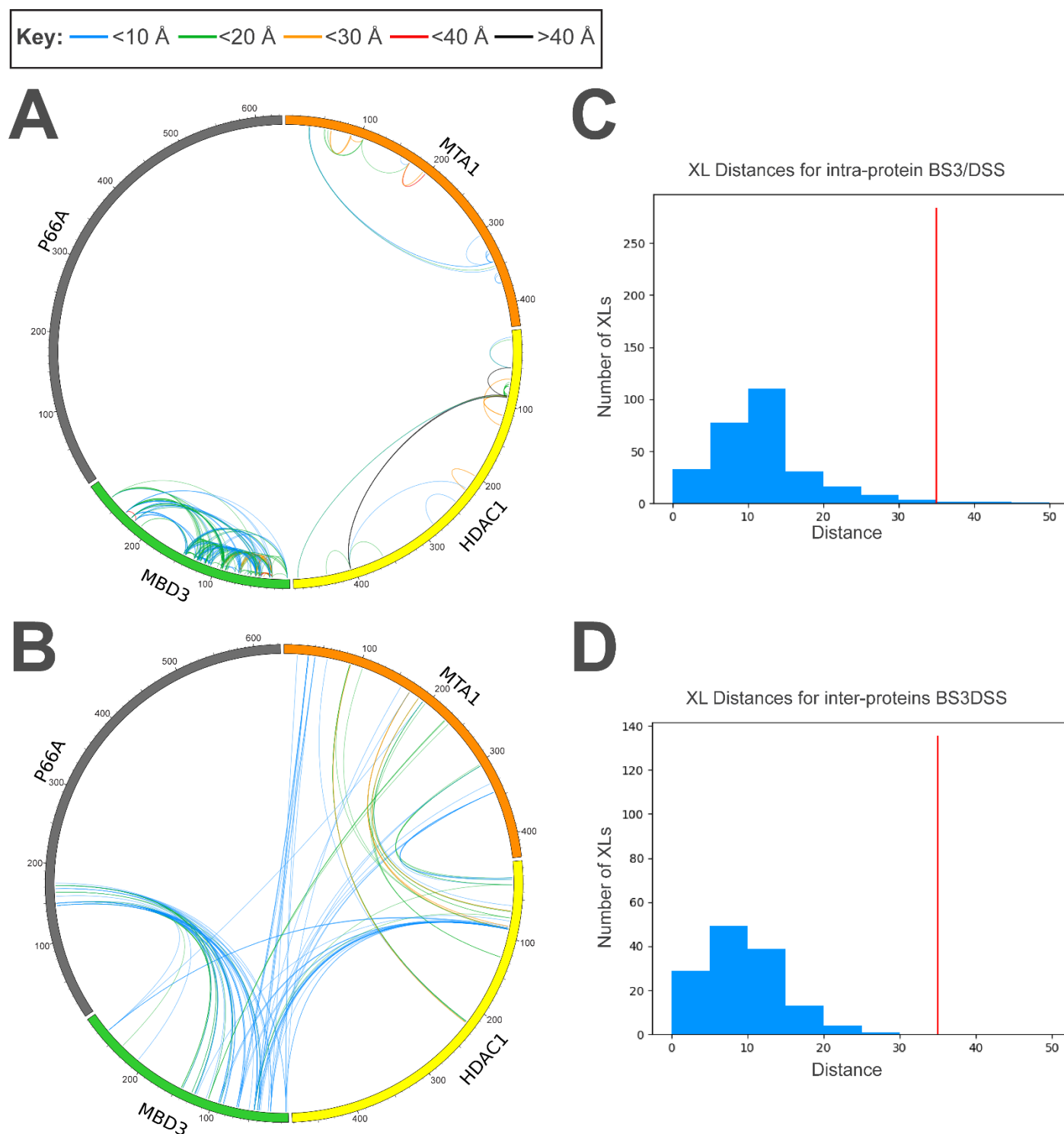


Fig. S6 Results of crosslink fitting for MTA1^N-HDAC1-MBD3^{GATAD2CC} (MHM) models CX-CIRCOS (<http://cx-circos.net/>) plots are shown for A. BS3/DSS, B. DMTMM, and C. ADH crosslinks on the ensemble of MHM models from the major cluster. Each line depicts a crosslink; its color depicts the minimum distance between the corresponding

crosslinked residues in the ensemble, as shown in the color key (top). Histograms showing the distribution of the minimum crosslink distance in the ensemble for D. BS3/DSS, E. DMTMM, and F. ADH crosslinks. The red line indicates the distance threshold for a crosslink type.

Key: — <10 Å — <20 Å — <30 Å — <40 Å — >40 Å

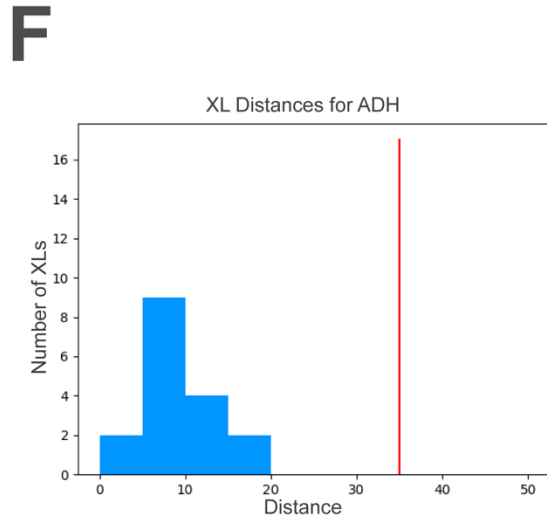
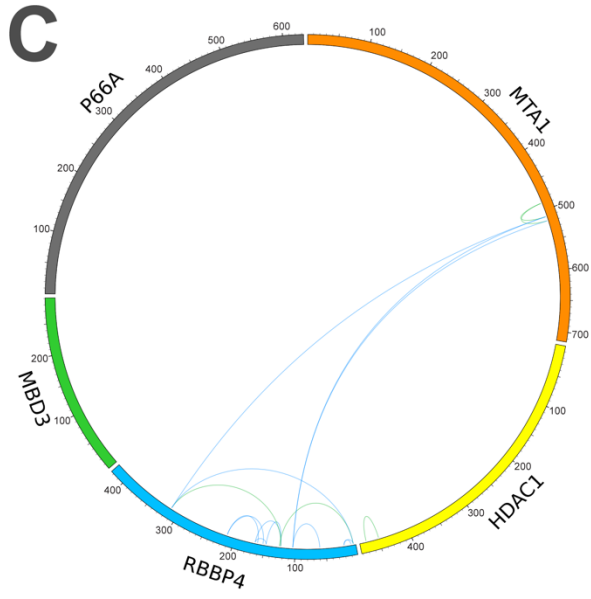
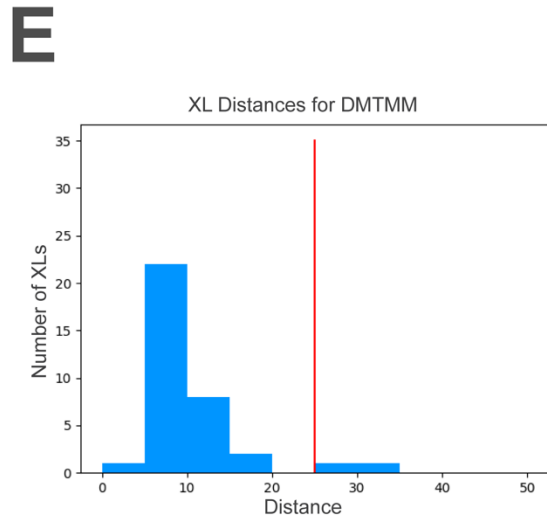
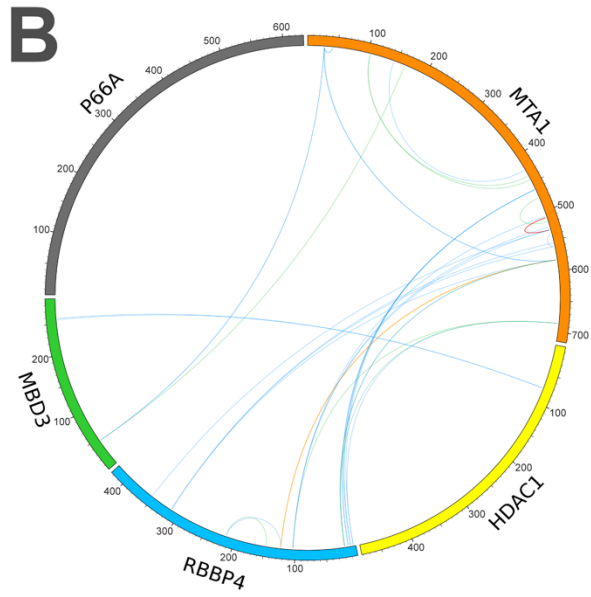
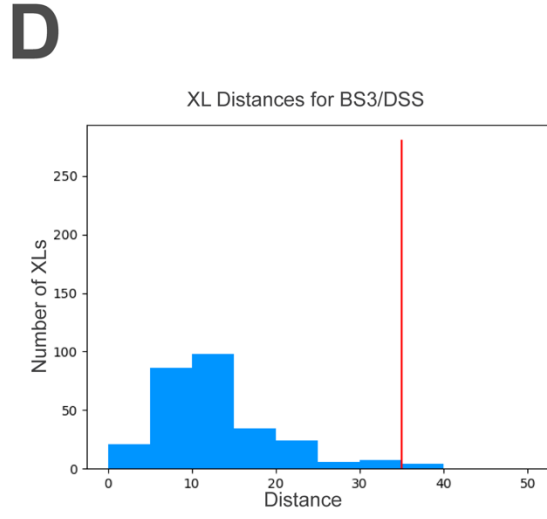
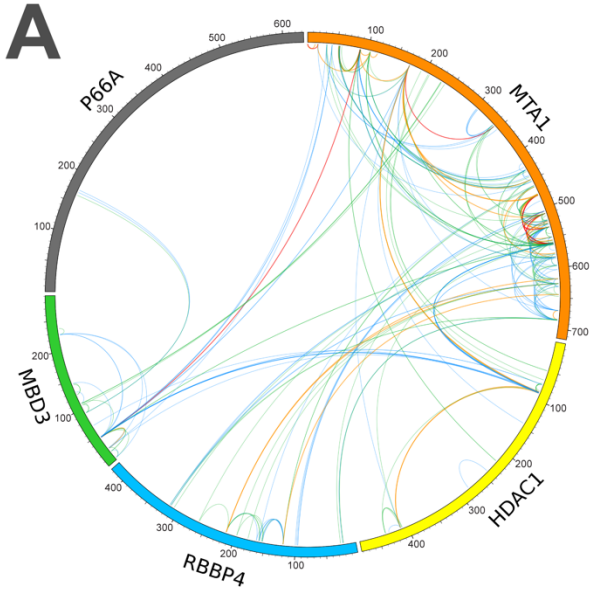


Fig. S7 Results of crosslink fitting for NuDe models CX-CIRCOS (<http://cx-circos.net/>) plots are shown for A. BS3/DSS, B. DMTMM, and C. ADH crosslinks on the ensemble of NuDe models from the major cluster. Each line depicts a crosslink; its color depicts the minimum distance between the corresponding crosslinked residues in the ensemble, as shown in the color key (top). Histograms showing the distribution of the minimum crosslink distance in the ensemble for D. BS3/DSS, E. DMTMM, and F. ADH crosslinks. The red line indicates the distance threshold for a crosslink type.

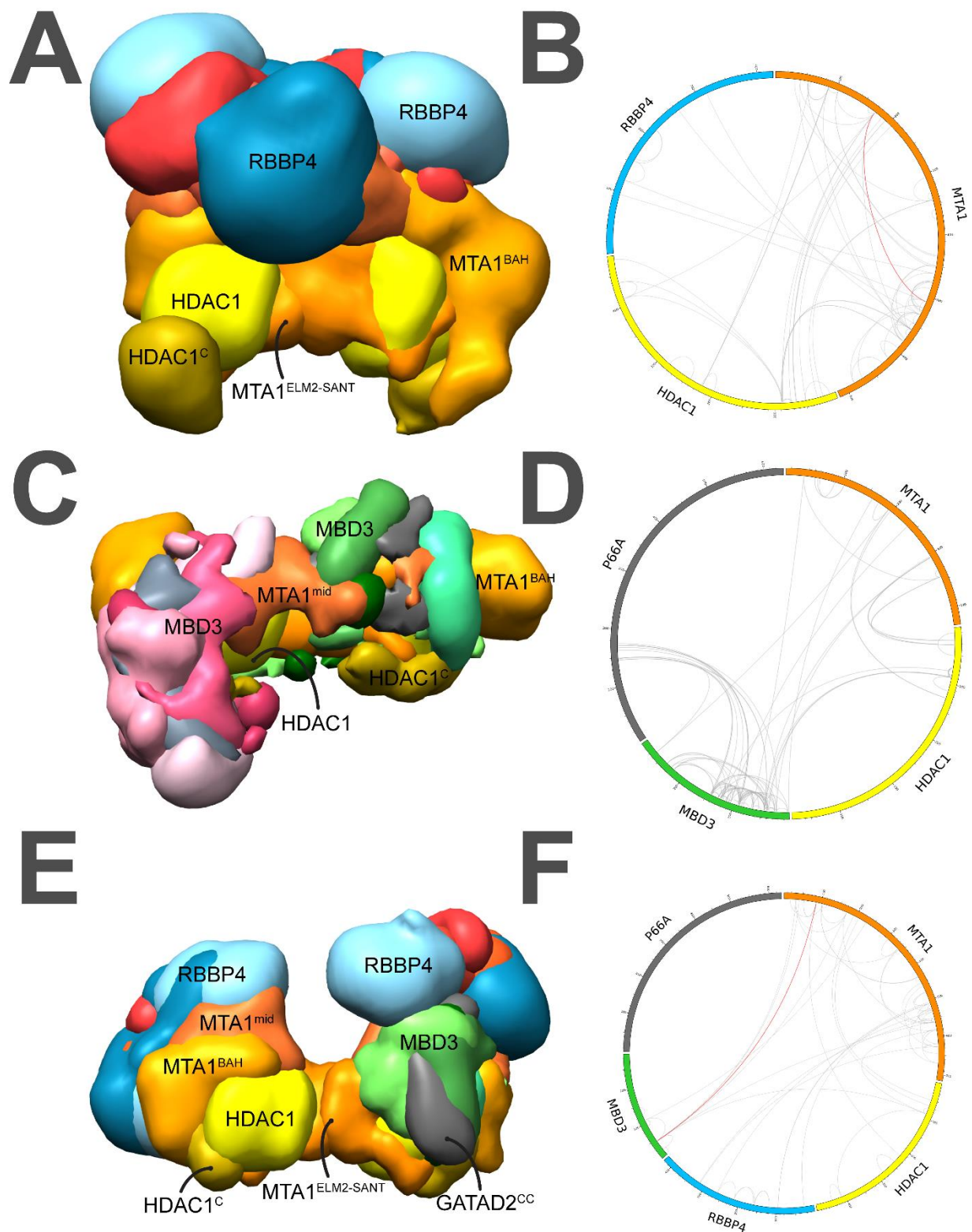


Fig. S8 Results of jack-knifing tests MHR, MHM, and NuDe models were recomputed by omitting a random subset of the input BS3/DSS crosslinks. The recomputed models were validated by their fit to the omitted crosslinks. Localization probability densities for the ensemble of recomputed models from the top cluster for A. MHR, C. MHM, and E. NuDe sub-complexes. Map colors and contour levels are similar to the main runs (Fig. 3- Fig. 5). CX-CIRCOS (<http://cx-circos.net/>) plots showing fit to the omitted crosslinks for the recomputed B. MHR, D. MHM, and F. NuDe models. Grey (red) lines indicate satisfied (violated) crosslinks.

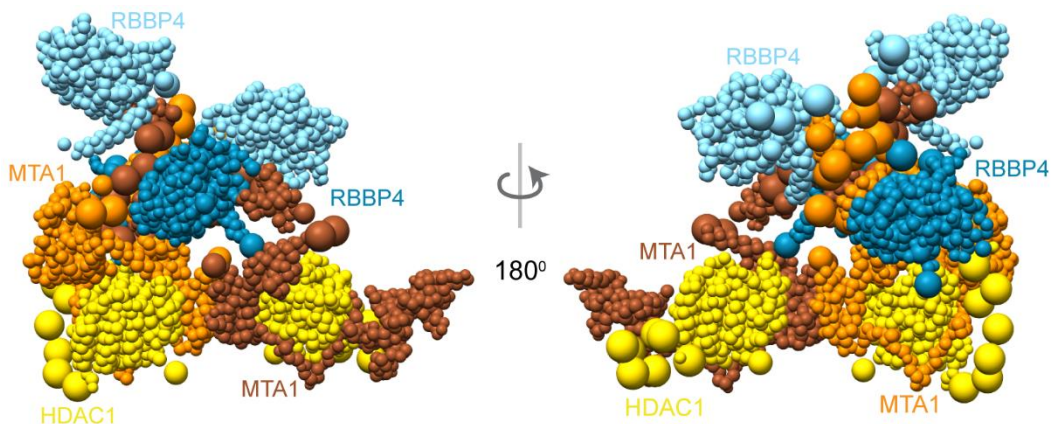


Fig. S9 Integrative model of the MHR complex based on crosslinks alone Representative bead model from the most populated cluster of analyzed integrative models for the MHR complex, colored by subunit. The two copies of MTA1 are shown in different colors (orange and brown) to illustrate the crossover.

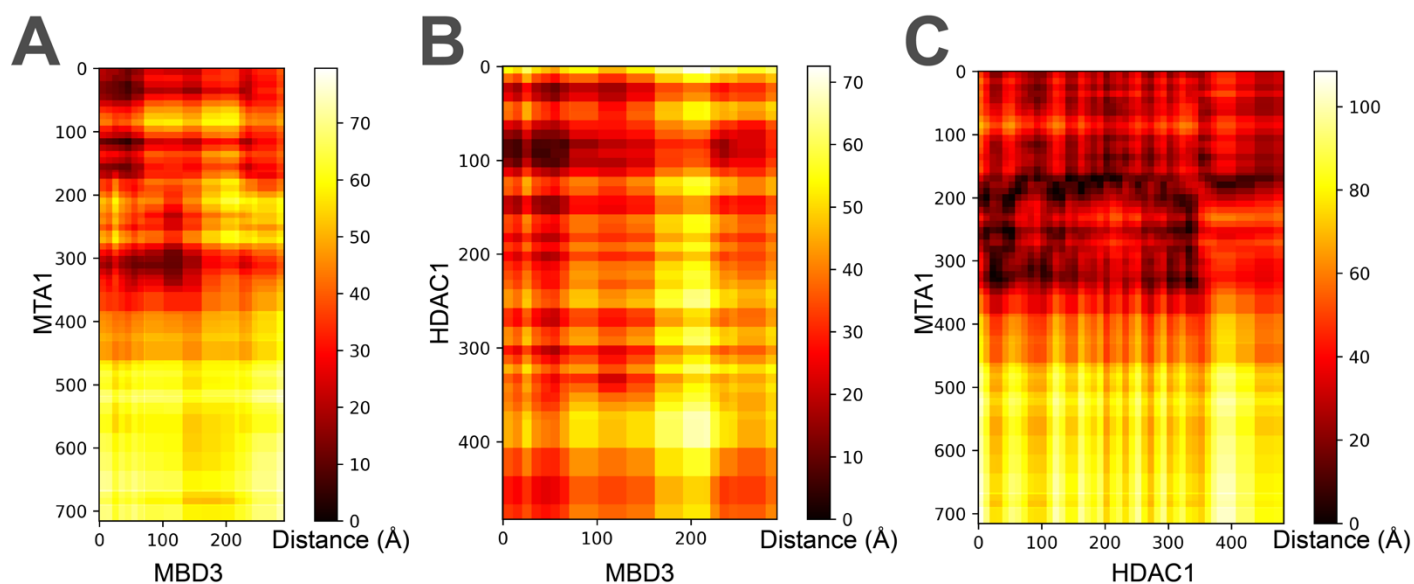


Fig. S10 Distance maps for protein interactions in NuDe. The distance maps show the average pairwise residue distances in the ensemble of NuDe models for the A. MBD3-MTA1, B. MBD3-HDAC1, and C. HDAC1-MTA1 protein pairs. For a pair of residues, the map indicates the distance between the surfaces of the corresponding beads averaged over the ensemble.

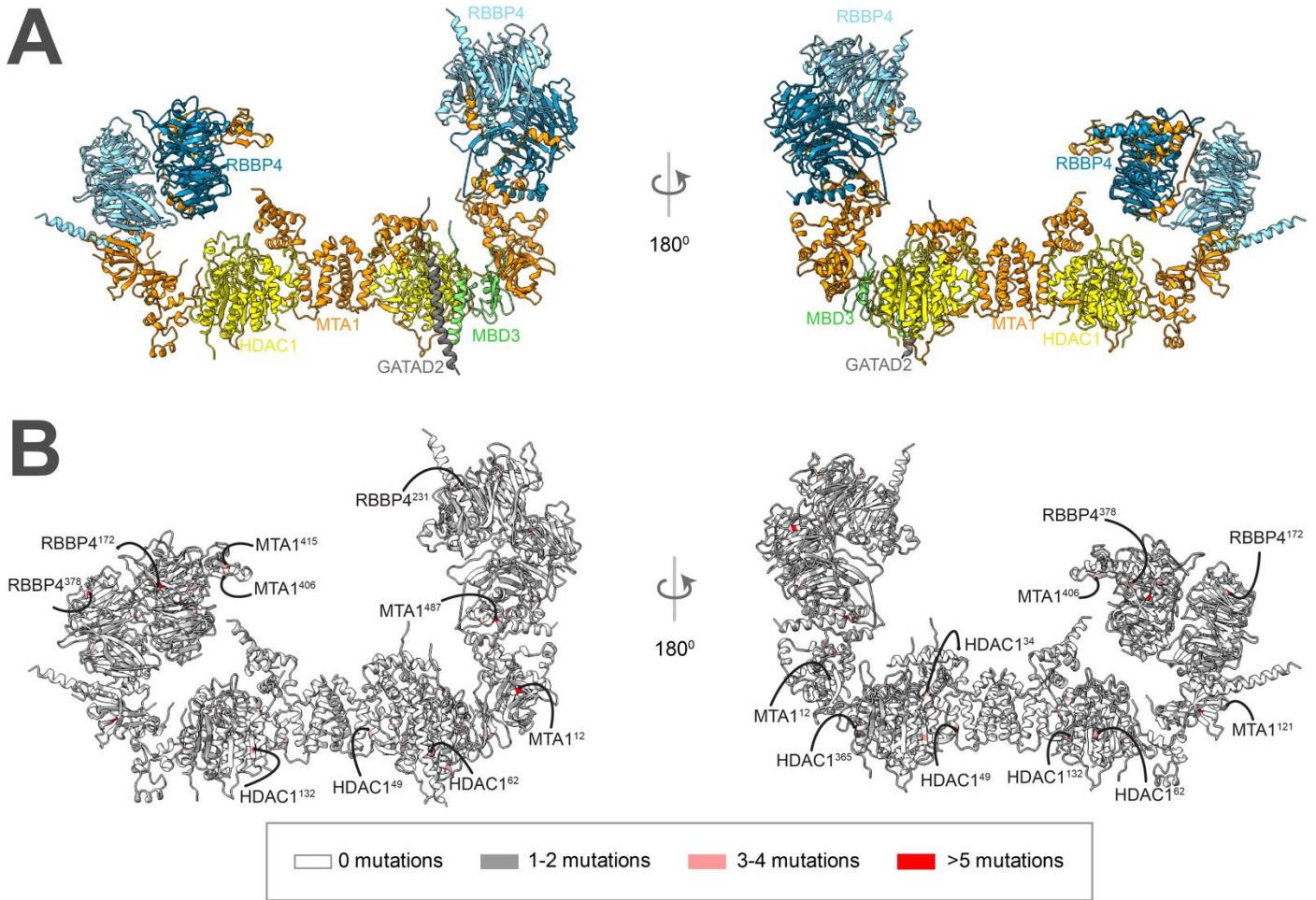


Fig. S11 COSMIC mutations on the structured regions of NuDe complex Somatic, confirmed pathogenic, point mutations from the COSMIC database (Forbes et al., 2006) mapped onto regions with known structure in the NuDe integrative model shown in Fig. 4A. A. Regions with known atomic structure in the integrative model colored by subunit. B. Mutations on residues in regions of known structure, colored according to the legend.

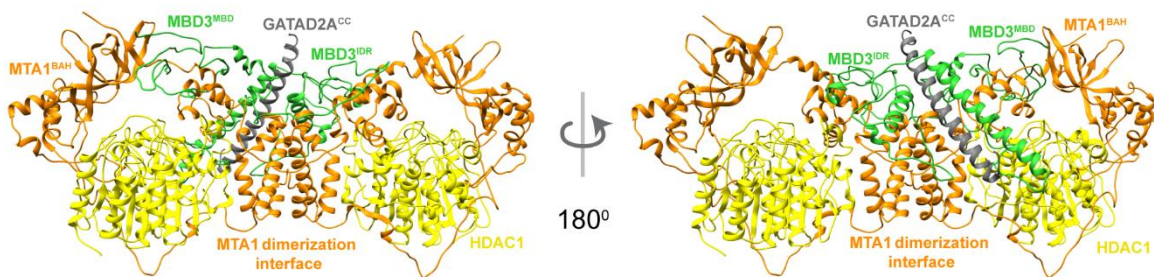


Fig. S12 AlphaFold multimer (Evans et al., 2022) predictions for 2:2:1:1 MTA1:HDAC1:MBD3:GATAD2^{CC} complex Front and back views of the top ranked model predicted by AlphaFold 2 multimer showing the proximity of MBD3^{MBD} to MTA1^{BAH} and the MBD3^{IDR}-MTA1 dimerization interface.

Intrinsic disorder profile

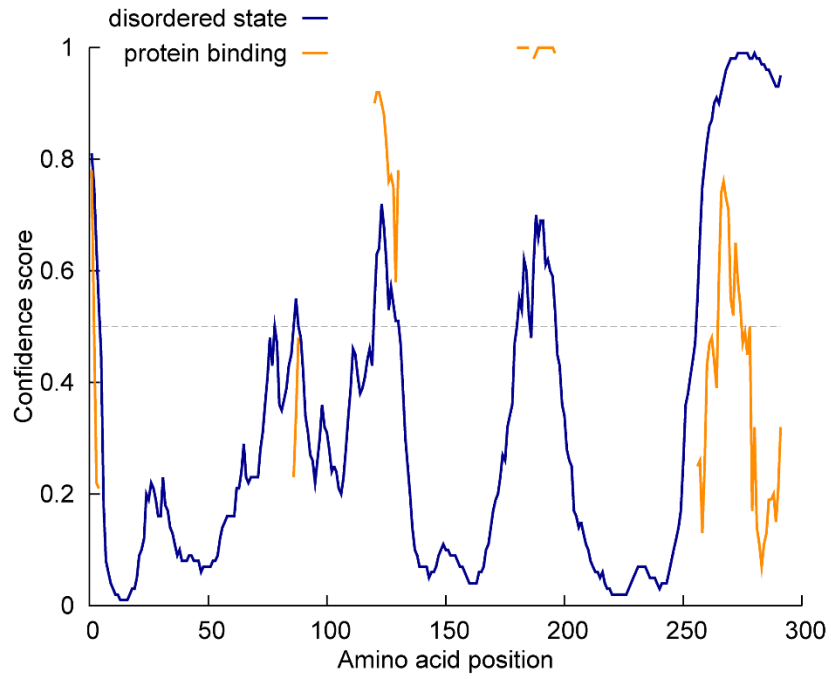


Fig. S13 Disorder prediction for MBD3 PONDR® (<http://www.pondr.com>) (Romero et al., 1997, 2001) disorder prediction for MBD3.

Table S1. COSMIC mutations on the NuDe integrative model Somatic, confirmed pathogenic point mutations from the COSMIC database [(Forbes et al., 2006)] mapped on the NuDe integrative model. A. Mutations in previously undescribed protein-protein interfaces in the model. Residues from two proteins are at an interface if the average distance between their corresponding bead surfaces is less than 5 Å in the cluster of models. B. Mutations in exposed binding sites between modeled proteins and known binding partners, based on the representative NuDe model.

A. COSMIC mutations in previously undescribed protein-protein interfaces in NuDe			
NuRD protein domain	Residues with mutations	Possible interacting partners in NuRD	Number of COSMIC mutations
MBD3 ^{MBD}	2, 7, 12, 14, 19, 21, 29, 31, 39, 41, 43, 45, 55, 65, 68	MTA1 ^{BAH} , HDAC1 ⁶⁰⁻¹⁰⁰	1-2
	17, 60	MTA1 ^{BAH} , HDAC1 ⁶⁰⁻¹⁰⁰	3-4
MTA1 ^{BAH}	10, 15, 24, 31, 46, 53, 60, 63, 81, 88, 90, 104, 107, 112, 121, 128, 133, 135, 143, 146, 153, 155, 158	MBD3 ^{MBD} , HDAC1 ^C	1-2
	12	MBD3 ^{MBD} , HDAC1 ^C	5+
HDAC1 ⁶⁰⁻¹⁰⁰	61, 64, 69, 73, 80, 86, 94, 96	MBD3 ^{MBD}	1-2
	62	MBD3 ^{MBD}	3-4
HDAC1 ^C	376, 379, 388, 391, 396, 404, 411, 420, 423, 425, 431, 435, 447, 462, 474, 476	MTA1 ^{BAH}	1-2
	398	MTA1 ^{BAH}	3-4

B. COSMIC mutations in exposed binding sites to known interactors in NuDe			
NuRD protein	Residues with mutations	Associated partner outside NuRD if known	Number of COSMIC mutations
RBBP4	39, 46, 71, 75, 129, 321, 376, 396, 398	H3, FOG1/2, other ZF containing TFs [(Lejon et al., 2011; Liu et al., 2015; Moody et al., 2018; Schmidberger et al., 2016)]	1-2
RBBP4	40, 378	H3, FOG1/2, other ZF containing TFs [(Lejon et al., 2011; Liu et al., 2015; Moody et al., 2018; Schmidberger et al., 2016)]	3-4
MTA1 ^{BAH}	46, 53, 153, 155, 158	Nucleosome (based on Sir3-BAH PDB 3TU4 [6]), MAT1 [(Millard et al., 2014)]	1-2
MTA1 ^{mid}	390, 394, 397, 401, 410, 416, 420, 424	MAT1 [(Mazumdar et al., 2001)]	1-2
MTA1 ^{mid}	406, 415	MAT1 [(Mazumdar et al., 2001)]	3-4
MTA1 ^{mid}	372	NRIF3 [(Talukder et al., 2004)]	5+
MTA1 ^{USR}	547, 550, 562, 564, 578, 581, 601, 603, 606, 612, 617, 619, 633, 647, 652, 658, 666	MICoA [(Mishra et al., 2003)]	1-2
MTA1 ^{USR}	549, 566, 571	MICoA [(Mishra et al., 2003)]	3-4
MTA1 ^C	697, 699, 705, 709, 715	MICoA [(Mishra et al., 2003)]	1-2
HDAC1 active site	140, 150, 303	Histone tails, TFs [(Scafuri et al., 2020)]	1-2
MBD3 ^{IDR}	77, 78, 80, 85, 87, 98, 103, 107	Evi1 [(Spensberger & Delwel, 2008)], Aurora A [(Sakai et al., 2002)]	1-2

References:

- Armache, K.-J., Garlick, J. D., Canzio, D., Narlikar, G. J., & Kingston, R. E. (2011). Structural Basis of Silencing: Sir3 BAH Domain in Complex with a Nucleosome at 3.0 Å Resolution. *Science*, 334(6058), 977–982. <https://doi.org/10.1126/science.1210915>
- Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A. W., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., ... Hassabis, D. (2022). *Protein complex prediction with AlphaFold-Multimer* (p. 2021.10.04.463034). bioRxiv. <https://doi.org/10.1101/2021.10.04.463034>
- Forbes, S., Clements, J., Dawson, E., Bamford, S., Webb, T., Dogan, A., Flanagan, A., Teague, J., Wooster, R., Futreal, P. A., & Stratton, M. R. (2006). Cosmic 2005. *British Journal of Cancer*, 94(2), 318–322. <https://doi.org/10.1038/sj.bjc.6602928>
- 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 *. *Journal of Biological Chemistry*, 286(2), 1196–1203. <https://doi.org/10.1074/jbc.M110.195842>
- Liu, Z., Li, F., Zhang, B., Li, S., Wu, J., & 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. *Journal of Biological Chemistry*, 290(10), 6630–6638. <https://doi.org/10.1074/jbc.M114.610196>
- Mazumdar, A., Wang, R.-A., Mishra, S. K., Adam, L., Bagheri-Yarmand, R., Mandal, M., Vadlamudi, R. K., & Kumar, R. (2001). Transcriptional repression of oestrogen receptor by metastasis-associated protein 1 corepressor. *Nature Cell Biology*, 3(1), 30–37. <https://doi.org/10.1038/35050532>
- Millard, C. J., Fairall, L., & Schwabe, J. W. R. (2014). Towards an understanding of the structure and function of MTA1. *Cancer and Metastasis Reviews*, 33(4), 857–867. <https://doi.org/10.1007/s10555-014-9513-5>

- Mishra, S. K., Mazumdar, A., Vadlamudi, R. K., Li, F., Wang, R.-A., Yu, W., Jordan, V. C., Santen, R. J., & Kumar, R. (2003). MICoA, a Novel Metastasis-associated Protein 1 (MTA1) Interacting Protein Coactivator, Regulates Estrogen Receptor- α Transactivation Functions*. *Journal of Biological Chemistry*, 278(21), 19209–19219. <https://doi.org/10.1074/jbc.M301968200>
- Moody, R. R., Lo, M.-C., Meagher, J. L., Lin, C.-C., Stevers, N. O., Tinsley, S. L., Jung, I., Matvekas, A., Stuckey, J. A., & Sun, D. (2018). Probing the interaction between the histone methyltransferase/deacetylase subunit RBBP4/7 and the transcription factor BCL11A in epigenetic complexes. *The Journal of Biological Chemistry*, 293(6), 2125–2136. <https://doi.org/10.1074/jbc.M117.811463>
- Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3), 290–296. <https://doi.org/10.1038/nmeth.4169>
- Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J. E., & Dunker, A. K. (1997). Identifying disordered regions in proteins from amino acid sequence. *Proceedings of International Conference on Neural Networks (ICNN'97)*, 1, 90–95 vol.1. <https://doi.org/10.1109/ICNN.1997.611643>
- Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins*, 42(1), 38–48. [https://doi.org/10.1002/1097-0134\(20010101\)42:1<38::aid-prot50>3.0.co;2-3](https://doi.org/10.1002/1097-0134(20010101)42:1<38::aid-prot50>3.0.co;2-3)
- Sakai, H., Urano, T., Ookata, K., Kim, M.-H., Hirai, Y., Saito, M., Nojima, Y., & Ishikawa, F. (2002). MBD3 and HDAC1, Two Components of the NuRD Complex, Are Localized at Aurora-A-positive Centrosomes in M Phase *. *Journal of Biological Chemistry*, 277(50), 48714–48723. <https://doi.org/10.1074/jbc.M208461200>
- Scafuri, B., Bontempo, P., Altucci, L., De Masi, L., & Facchiano, A. (2020). Molecular Docking Simulations on Histone Deacetylases (HDAC)-1 and -2 to Investigate the Flavone Binding. *Biomedicines*, 8(12). <https://doi.org/10.3390/biomedicines8120568>

Schmidberger, J. W., Tabar, M. S., Torrado, M., Silva, A. P. G., Landsberg, M. J., Brillault, L., AlQarni, S.,

Zeng, Y. C., Parker, B. L., Low, J. K. K., & Mackay, J. P. (2016). The MTA1 subunit of the nucleosome remodeling and deacetylase complex can recruit two copies of RBBP4/7. *Protein Science*, 25(8), 1472–1482. <https://doi.org/10.1002/pro.2943>

Spensberger, D., & Delwel, R. (2008). A novel interaction between the proto-oncogene Evi1 and histone methyltransferases, SUV39H1 and G9a. *FEBS Letters*, 582(18), 2761–2767.

<https://doi.org/10.1016/j.febslet.2008.06.056>

Talukder, A. H., Gururaj, A., Mishra, S. K., Vadlamudi, R. K., & Kumar, R. (2004). Metastasis-Associated Protein 1 Interacts with NRIF3, an Estrogen-Inducible Nuclear Receptor Coregulator. *Molecular and Cellular Biology*, 24(15), 6581–6591. <https://doi.org/10.1128/MCB.24.15.6581-6591.2004>