Structure of the nucleosome-bound human BCL7A

Franck Martin, Asgar Abbas Kazrani, Julie Lafouge, Dana Mariel Diaz Jimenez, Stéphanie Siebert, Leonie Fabbro-Burtschell, Emma Maillard, Karine Lapouge, Haydyn David Thomas Mertens, Claude Sauter, et al.

Nucleic Acids Research, 2025, **53**, gkaf273
https://doi.org/10.1093/nar/gkaf273
Structural Biology

Structure of the nucleosome-bound human BCL7A

Franck Martin[1,2,3,4,†], Asgar Abbas Kazrani[1,2,3,4,†], Julie Lafouge[1,2,3,4,†], Dana Mariel Diaz-Jimenez[1,2,3,4], Stéphanie Siebert[1,2,3,4], Leonie Fabbro-Burtschell[1,2,3,4], Emma Maillard[5], Karine Lapouge[6], Haydyn David Thomas Mertens[7], Claude Sauter[8], Alexander Leitner[9], Françoise Ochsenbein[5], Alexandre Blais[10,11,12,13], Elisa Bergamin[1,2,3,4,*]

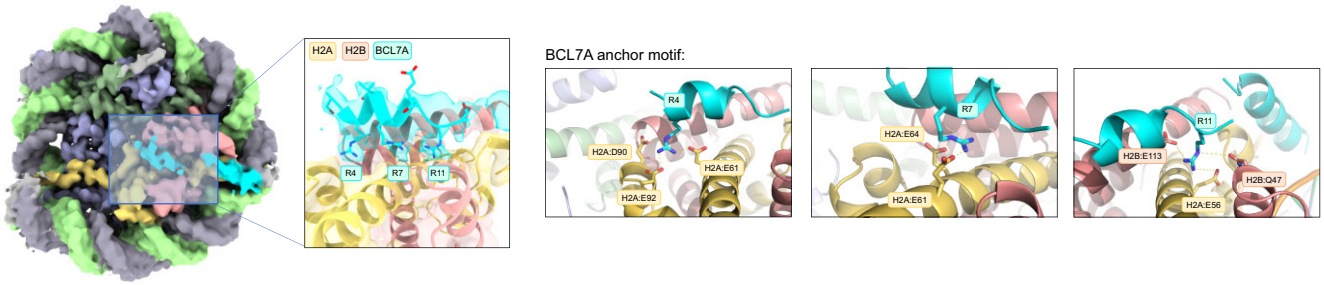
[1]Department of Functional Genomics and Cancer & Department of Integrated Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), 67400 Illkirch-Graffenstaden, France
[2]Université de Strasbourg, 67084 Strasbourg, France
[3]Centre National de la Recherche Scientifique UMR 7104, 67400 Illkirch-Graffenstaden, France
[4]Institut National de la Santé et de la Recherche Médicale U1258, 67400 Illkirch-Graffenstaden, France
[5]Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, 91190 Gif-sur-Yvette, France
[6]Protein Expression and Purification Core Facility, European Molecular Biology Laboratory, 69117 Heidelberg, Germany
[7]European Molecular Biology Laboratory (EMBL), Hamburg Outstation, c/o DESY, Notkestraße 85, 22607 Hamburg, Germany
[8]Architecture et Réactivité de l'ARN, Université de Strasbourg, CNRS, IBMC, 2 All. Konrad Roentgen, 67084 Strasbourg, France
[9]Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland
[10]Faculty of Medicine, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, 451 Smyth Road, Ottawa, Ontario K1H 8M5, Canada
[11]Ottawa Institute of Systems Biology, Ottawa, Ontario K1H 8M5, Canada
[12]University of Ottawa Centre for Inflammation, Immunity and Infection (CI3), Ottawa, Ontario K1H 8M5, Canada
[13]Éric Poulin Centre for Neuromuscular Disease, Ottawa, Ontario K1H 8M5, Canada

[*]To whom correspondence should be addressed. Email: bergamie@igbmc.fr
[†]The first three authors should be regarded as Joint First Authors.

Abstract

Proteins of the BCL7 family (BCL7A, BCL7B, and BCL7C) are among the most recently identified subunits of the mammalian SWI/SNF chromatin remodeler complex and are absent from the unicellular version of this complex. Their function in the complex is unknown, and very limited structural information is available, despite the fact that they are mutated in several cancer types, most notably blood malignancies and hence medically relevant. Here, using cryo-electron microscopy in combination with biophysical and biochemical approaches, we show that BCL7A forms a stable, high-affinity complex with the nucleosome core particle (NCP) through binding of BCL7A with the acidic patch of the nucleosome via an arginine anchor motif. This interaction is impaired by BCL7A mutations found in cancer. Further, we determined that BCL7A contributes to the remodeling activity of the mSWI/SNF complex and we examined its function at the genomic level. Our findings reveal how BCL7 proteins interact with the NCP and help rationalize the impact of cancer-associated mutations. By providing structural information on the positioning of BCL7 on the NCP, our results broaden the understanding of the mechanism by which SWI/SNF recognizes the chromatin fiber.

Graphical abstract



Received: November 2, 2023. **Revised:** March 21, 2025. **Editorial Decision:** March 21, 2025. **Accepted:** March 25, 2025

Introduction

At the most fundamental level, eukaryotic DNA is packaged into chromatin by making nearly two turns around an octamer of histone proteins, forming nucleosomes [1]. ATP-dependent chromatin remodeling complexes modulate the compaction of DNA through sliding, adding, or evicting nucleosomes from the chromatin fiber, or replacing histones with histone variants [2]. Deregulation of this process can severely impact gene expression patterns and genome integrity. Their importance is such that chromatin remodeling protein mutations are strongly associated to several diseases, including cancer [3]. Unfortunately, despite encouraging advances with certain mutations [4–6], in most cases we currently do not understand well enough the molecular details of their mode of action to be able to translate this into improved cancer treatment.

All eukaryotes have multiple ATP-dependent chromatin remodeling complexes, which tend to be highly conserved, from simple yeasts to metazoans, testifying to their importance. Their ATPases all have homology to the *Saccharomyces cerevisiae* enzyme SNF2 and are classified into sub-families based on the amino acid sequence of the ATPase: SNF2, ISWI, CHD, and INO80 [7]. The complexes accomplish their remodeling function by antagonizing the strong interaction between the histone octamer and DNA, to alter chromatin structure. In addition to their ATPase subunit, these complexes contain more associated proteins that contribute to their function. The mammalian switch/sucrose nonfermentable (mSWI/SNF) complex appears as a ∼2 MDa complex and is composed of at least 11 subunits, coming from the product of 29 genes and multiple paralogs generating wide diversity in its composition [8]. mSWI/SNF complexes exist in three forms: canonical Bromo-Associated Factor (BAF or cBAF) complexes, polybromo-associated BAF (PBAF) complexes, and noncanonical BAF (ncBAF) complexes, each distinguished by the incorporation of distinct accessory subunits [8]. Each complex comprises a set of evolutionary conserved 'core and enzymatic' subunits, but also 'auxiliary' subunits present only in animals thought to reflect increasing biological complexity. All complexes contain a conserved ATPase subunit, either BRG1 or BRM (encoded by the SMARCA4 and SMARCA2 genes), that catalyses the hydrolysis of ATP. The advent of sensitive and large-scale proteomics methods has shed new light on the composition of the human SWI/SNF. Studies focused on identifying mSWI/SNF components [9, 10], as well as others determining the composition of most nuclear multiprotein complexes [11–13], have revealed the identity of previously unsuspected mSWI/SNF subunits, namely BCL7A, BCL7B, BCL7C, BRD7, BRD9, SS18, SS18L1, BICRA, and BICRAL. Without exception, these proteins do not exist in yeast and only participate in the metazoan complexes. The molecular function of most accessory subunits and their role within mSWI/SNF remain incompletely defined [10, 14–19]. It has been reported that BCL7 proteins accumulate at sites of DNA damage, but the role they may play in DNA damage repair is unknown [20].

BCL7A is labeled as a "Cancer Gene" by the COSMIC database as it is often mutated in a variety of cancers and is an established tumor suppressor [21–23]. Despite this, the structure, the molecular details of interaction and the function of BCL7 subunits within mSWI/SNF are poorly defined at best. They participate to all three forms of mSWI/SNF. The topological information currently available is limited to the observation that BCL7 proteins make contacts with BRG1, BAF47 (encoded by the SMARCB1 gene), DPF2, actin, and histone H2B [8, 24]. Recent studies have reported the cryo-electron microscopy (cryo-EM) structure of the SWI/SNF complex. Unfortunately, they do not inform us on BCL7 proteins: some studies concerned the yeast SWI/SNF complex and BCL7 are absent in this species [25–27]; in studies of reconstituted mSWI/SNF, BCL7 proteins were either omitted [28, 29] or not detectable [24]; and studies with the endogenous human complex did not have sufficient resolution to unambiguously locate BCL7 proteins within the complex [30].

Given this important knowledge gap, we have investigated the structure and molecular function of the BCL7 proteins. We used cryo-EM, biophysical and biochemical approaches to obtain evidence that BCL7A tightly binds the acidic patch of the nucleosome core particle (NCP) through an arginine anchor motif. We find that single amino acids substitutions within the highly conserved arginine anchor of BCL7A, which are frequent in cancer patients, impair NCP binding.

Experimental procedures

Protein expression and purification

The complementary DNA (cDNA) encoding full length human BCL7A (isoform Q4VC05-2, 231 amino acids), BCL7A (1–100), full length BCL7B (isoform Q9BQE9-1, 202 amino acids), BCL7C (isoform Q8WUZ0-2, 242 amino acids), and BAF47 were subcloned into the parallel vector pGST2 that includes a Tobacco Etch Virus (TEV)-cleavable, N-terminal glutathione sulfotransferase (GST) tag. BCL7A mutants were obtained by site directed mutagenesis (QuickChange II site directed mutagenesis kit, Agilent) and the constructs were verified by automated DNA sequencing. Each vector encoding BCL7 proteins or BAF47 was transformed into *Escherichia coli* strain Rosetta (DE3), and cultures were grown in Terrific Broth media at 37°C to an OD_{600} of 0.6. Protein expression was induced by the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) 0.2 mM for 16 h at 19°C. Cells were harvested, resuspended in lysis buffer [50 mM Tris (pH 8.0), 500 mM NaCl, 0.1% (v/v) Triton X-100, 5 mM 2-mercaptoethanol], lysed by sonication and clarified by centrifugation. The supernatant was collected and applied onto glutathione Sepharose beads (Cytiva) for 2 h at 4°C and washed extensively with lysis buffer. GST-BCL7 or GST-BAF47 were TEV cleaved on beads for 16 h at 4°C. BCL7 untagged proteins were further purified by heparin affinity chromatography and all the proteins were purified by size exclusion chromatography (SEC) (Superdex 200 Increase 10/300) pre-equilibrated in 20 mM Tris (pH 8.0), 150 mM NaCl and 1 mM Dithiothreitol (DTT). To produce protein for nuclear magnetic resonance (NMR) experiments, full length human BCL7A was cloned into pSMT3 vector carrying a N-terminal ULP1 cleavable His_6-SUMO tag and it was transformed into *E. coli* strain Lemo21 (DE3). For ^{15}N labeling, cells were grown to OD_{600} of 0.6 in minimal medium containing 0.5 g/l $^{15}NH_4Cl$ as the sole nitrogen source. For $^{15}N/^{13}C$-labeled sample, 0.5 g/l $^{15}NH_4Cl$ and 2 g/l ^{13}C glucose were the sole sources of nitrogen and carbon, respectively. Protein expression was performed as indicated above. Cells were harvested and resuspended in lysis buffer [50 mM Tris (pH 8.0), 500 mM NaCl, 0.1% (v/v) Triton X-100,

5 mM 2-mercaptoethanol], lysed by sonication and clarified by centrifugation. The supernatant was collected and applied onto TALON beads (Cytiva) for 1 h at 4°C and washed extensively with lysis buffer containing 1 M NaCl and then equilibrated in the same buffer containing 300 mM NaCl. His6-SUMO BCL7A was ULP1 cleaved on beads for 16 h at 4°C. Untagged protein was further purified by SEC (Superdex 200 Increase 10/300) pre-equilibrated in 20 mM HEPES (pH 7.0), 150 mM NaCl, and 1 mM DTT.

To isolate the BCL7A-BAF47 complex, the cDNA encoding full length human BCL7A and human BAF47 were subcloned independently into the vector pSMT3, transformed into the *E. coli* strain Lemo21 (DE3), and protein expression was carried out as mentioned above. Cells were harvested, resuspended in lysis buffer [50 mM Tris (pH 8.0), 500 mM NaCl, 5 mM 2-mercaptoethanol], lysed by sonication, and clarified by centrifugation. The clarified lysates were combined, incubated with TALON beads (Cytiva) for 1 h at 4°C, and washed thoroughly with lysis buffer. Both proteins were eluted with elution buffer [50 mM Tris (pH 8.0), 200 mM NaCl, 250 mM imidazole], and then dialyzed for 16 h at 4°C against a buffer consisting of 20 mM HEPES (pH 7.0), 150 mM NaCl, 0.5 mM ethylenediaminetetraacetic acid (EDTA), 1 mM DTT in presence of ULP1 to cleave the SUMO tag off. The untagged proteins were further purified by SEC (Superdex 200 Increase 10/300) pre-equilibrated in 20 mM HEPES (pH 7.0), 150 mM NaCl, 0.5 mM EDTA and 1 mM DTT.

The BAF47 shorter constructs encompassing amino acids 182–245 (RPT1), 182–382 (RPT1-RPT2-CTD), and 100–325 (RPT1-RPT2) were also subcloned into pSMT3 vector and overexpressed and purified as mentioned above for the full length BAF47.

BRG1 was expressed in Sf9 cells using the Bac-to-Bac Baculovirus expression system (Invitrogen). pFastBac1 Flag BRG1 was a gift from Robert Kingston (Addgene plasmid #1957; http://n2t.net/addgene:1957; RRID:Addgene_1957) and purification of BRG1 was carried out as previously reported [31].

Nucleosomal DNA and large-scale nucleosome reconstitution

The 147-bp strong positioning 601 DNA sequence [32] was used. To produce sufficient quantity of NCPs for biochemical studies, a large-scale protocol was employed. The DNA was prepared in milligram quantities using a plasmid with multiple concatenated copies of the 601 sequence, followed by restriction digestion of the individual units and size selection, as reported previously [33]. *Xenopus laevis* core histones were produced by co-expression in *E. coli* and were purified in native form following an established protocol [34]. The two components were assembled into NCPs by overnight dialysis to reduce NaCl from 2 to 0.25 M. NCPs were heat-shifted at 37°C for 120 min. Proper assembly was assessed on nondenaturing polyacrylamide gels [5% acrylamide 37.5:1 acrylamide:bis-acrylamide ratio, 0.25× Tris-Glycine-EDTA (TGE) buffer, 0.03% (v/v) NP-40] pre-run for 1 h at 4°C and run for 1 h at 120 volts on Mini-Protean 3 gels (Bio-Rad).

Electrophoretic mobility shift assays

Fluorescent NCPs were assembled using the salt dilution method with recombinant histones and Cy5-labeled 601 DNA produced by polymerase chain reaction (PCR). The labeled probes were incubated at 10 nM in 30 μl binding reactions containing 20 mM Tris–HCl (pH 7.6), 50 mM NaCl, 10% (v/v) glycerol, 1 mM EDTA. Purified proteins were omitted or added to the binding reactions at six varying concentrations ranging from 75 to 1800 nM. The concentrations were chosen after preliminary tests with a broader concentration range and selected so that in general there was a large molar excess of protein over ligand (total protein is a reasonable approximation of free protein and binding reactions occurred in the "binding regime") [35]. Reactions were incubated on ice for 30 min and then loaded on nondenaturing 5% acrylamide gels (60:1 ratio of acrylamide to bis-acrylamide) containing 10% glycerol, using 0.25× TGE (40 mM Trizma, 45 mM glycine, and 1 mM EDTA) as running buffer.

After electrophoresis, gels were immediately scanned using the Typhoon Phosphorimager, at 100 μm resolution, adjusting the photomultiplier Tube (PMT) value in order to obtain the strongest signal without saturating any pixel. Image files were saved in 16-bit.gel format and imported into the ImageQuant TL 10.1 software (GE Healthcare). Free NCP signal was quantified in boxes of equal area as relative intensities, and background was subtracted using the "rubber band" method. The free NCP values in each lane were used, expressed as fraction of free NCP in the absence of protein, and fitted to various equations, including a quadratic binding equation [35] and the Hill equation [36] fitting the fraction of free probe as a function of the micromolar concentration of protein, and total ligand concentration in the assay (0.01 μM). It was found that for each protein assayed, the Hill equation gave the best agreement with the experimental points. The trace amount of probe assumption required for the Hill equation was considered valid as it is ∼5%–10% of the calculated K_D [37]. Fitting was performed using the nlsLM function from the 'minpack.lm' package in R. The initial value for K_D was set to 0.2 and Hill coefficient n (when applicable) was set to 1. Using an initial K_D value 10 times higher or 90% lower did not affect the results. Multiple replicates of each experiment were performed. For each replicate, the residuals sum of squares from measured versus fitted values was used to perform a goodness-of-fit analysis; this returned chi-squared *P*-values lower than 1E-4 in every case. The final reported values are the arithmetic mean ± standard error of replicates. Student's *t*-tests (unpaired, one-tailed, equal variance) were used to determine statistical significance when comparing wild type (WT) with mutant proteins.

Biolayer interferometry assay

His-tagged NCPs were assembled using the salt dilution method with histidine tagged recombinant histones (His_6-H2A and His_6-H4) and 601 DNA prepared as mentioned above. Biolayer interferometry (BLI) experiments were conducted using an Octet R8 System (Sartorius) at 20°C and 1000 rpm and samples (200 μl) were prepared in wells of a 96-well plate [microplate 96-well, polypropylene, F-bottom (chimney well) from Greiner Bio-One]. Ni-NTA biosensors (Sartorius octet NTA biosensors) were first incubated in reaction buffer [20 mM HEPES (pH 8.0), 100 mM KCl, and 0.02% (v/v) Tween], and then loaded in the same buffer with His-tagged mono nucleosomes until a BLI signal of 1.5 nm was reached. The binding experiments with BCL7A WT or mutants were performed in the reaction buffer, except that 2% bovine serum

albumin (BSA) was added to reduce BCL7A nonspecific binding. Protein concentrations ranged from 1 μM to 15.6 nM with association and dissociation times both set at 600 s. The nonspecific BCL7A-sensor interactions were determined by measuring binding signal of empty sensors in various concentrations of BCL7A, which was subsequently subtracted from the signal of the binding data. Overall, the nonspecific interaction of BCL7A with the sensor was of about 35%. K_D values were determined by plotting the local fit maximum response (nm) as a function of small molecule concentrations (nM) using Octet Data Analysis 13.0.1.35 Software. Global fitting was applied to all data, using a heterogeneous binding model and the titration curves were fit using the following steady-state analysis equation: Response = $R_{max} \times conc/(K_D + conc)$, where R_{max} is the local fit maximum response, conc is the concentration of the small molecule and K_D is the equilibrium dissociation constant. All the experiments were repeated at least twice, with values averaged and standard errors calculated.

Analytical ultracentrifugation

Sedimentation velocity (SV) experiments were performed at 20°C and 42 000 rpm in a Beckman Coulter ProteomeLab XL-I ultracentrifuge with an AnTi-50 eight-hole rotor. NCP was further purified on gel filtration column before SV experiment. Samples of 420 μl or 110 μl were loaded into analytical cell assemblies with 12- or 3-mm charcoal-filled Epon double-sector centerpieces and sapphire windows. The reference buffer was 20 mM Tris (pH 8), 100 mM NaCl, and 1 mM DTT. After temperature equilibration at 20°C for 3 h, the rotor was accelerated to 42 000 rpm and 300 scans were collected with absorbance optics at 280 nm and interference optics. Sedimentation data were analyzed with SEDFIT software (version 16.1c) using the continuous sedimentation coefficient distribution model c(s). Buffer density, buffer viscosity and protein partial specific volumes were calculated using SEDNTERP software. A partial specific volume of 0.65 ml/g was used for NCP or NCP with BCL7A, and 0.7107 ml/g for BCL7A. A solvent density of 1.03 g/ml and a solvent viscosity of 1.017 cP were used for data analysis. GUSSI was used to integrate the sedimentation peaks and to produce the graphs [38].

Size exclusion chromatography combined with multi-angle light scattering

Size exclusion chromatography combined with multi-angle light scattering (SEC-MALS) was performed with a Superose 6 (10/300 GL) column at 0.5 ml/min, 25°C, pre-equilibrated in 20 mM Tris (pH 8.0), 150 mM NaCl, 1 mM DTT, to separate the sample before performing the MALS measurements. BCL7A wild type, NCP alone or in presence of a 1:5 molar excess of BCL7A were individually injected in a Superose 6 10/300. The molar mass for each molecule was determined with the ASTRA software (Wyatt Technologies).

Cancer mutation analysis

The COSMIC database was queried on 1 October 2023 for samples with mutations to the BCL7A gene and the results were retrieved for plotting. The number of samples with mutation at each amino acid position, regardless of the change (missense, nonsense, insertion, or deletion) are reported. For the diffuse large B-cell lymphoma (DLBCL) mutation hotspot analysis, the data from the FishHook analysis by Mlynarczyk *et al.* were downloaded [39]. This analysis was performed with a cohort of 101 DLBCL cases from the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium [40]. As reported by the authors, a Q–Q plot of the log-transformed observed *P*-values versus the log-transformed quantiles from the uniform distribution was prepared.

NMR experiments

NMR experiments were carried out on Bruker DRX-950MHz spectrometer equipped with a cryo-probe. All NMR data were processed using Topspin (Bruker) and analyzed using Sparky (T.D. Goddard and D.G. Kneller, UCSF). Samples were prepared in 3 mm NMR tubes, in a HEPES (pH 7.0) buffer with 100 mM NaCl, 1 mM DTT, 0.1 mM disuccinimidyl suberate (DSS), 0.1% NaN_3, and 5% D_2O. For the titration experiments, BCL7A concentration was 6 μM. Heteronuclear Multiple Quantum Correlation (sofast-HMQC) spectra were recorded at 287°K. For the backbone resonances assignments, 2D and 3D data were collected at 287°K with a μM uniformly ^{13}C-^{15}N labeled sample, including BTROSY, CBCANH, CBCA(CO)NH, HN(CA)CO, HNCO, and HNCANH experiments from the NMRLIB package [41]. Proton chemical shifts (in ppm) were referenced relative to internal DSS and ^{15}N and ^{13}C references were set indirectly relative to DSS using frequency ratios [42]. Chemical shift index was calculated according to the sequence-specific random coil chemical shifts [43].

Cryo-EM studies of the BCL7A–NCP complex

Sample preparation and data acquisition

BCL7A and nucleosomes used in the preparation of the complex for cryo-EM studies were produced as indicated above. 147-bp nucleosomes and BCL7A were mixed at a molar ratio of 1:5 (NCP:BCL7A). The complex was incubated at 4°C for 1 h before being purified by SEC (Superose 6 Increase 3.2/300). To prepare the cryo-EM grids, the freshly purified complex was first cross-linked with 0.1% glutaraldehyde, then, immediately, 5 μl of sample were applied on Quantifoil holey carbon grids (R2/2, 300 mesh) plasma cleaned on a model 1070 NanoClean (Fischione Instruments) for 90 s at 30% power and with a mixture of 80% argon and 20% oxygen. After application the grids were blotted for 4 s at a force setting of 5 and then plunged into liquid ethane using a Vitrobot mark IV (FEI) system with a chamber humidity of 95% at 10°C. Images were collected on a 300 kV Titan Krios (FEI) at the Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC) equipped with a Falcon 4i camera in super resolution counting mode using the SerialEM software [44] with a physical pixel size of 0.729 Å and a defocus range from −0.8 to −2.2 μm. Exposures of 3 s were dose-fractionated into 40 frames with a dose rate of 7.64 electrons $pixel^{-1}$ s^{-1}, which gives a total dose of 43.13 $e^-/Å^2$ (Supplementary Table S1). A total of 15 749 movies were collected.

Cryo-EM data processing

The images in the dataset were aligned and corrected for particle motion using MotionCor2 [45]. Contrast transfer function (CTF) estimates were performed with CTFFIND4 [46]. CryoSPARC [47] was used for all data processing. Particles were selected using the blob picker and then the low-quality/false particles were filtered out using 2D classes. The best 2D classes were selected to generate templates for tem-

plate picking. Template picker was used to select particles more specifically and in greater number (2453 488 particles). *Ab initio* reconstructions were generated respectively for the NCP-BCL7A complex and for the particles of poor quality. This allowed to conduct a heterogenous refinement with decoy classes and the particles were cleaned more efficiently (1236 492 particles). Once the particles had been thoroughly cleaned, another heterogeneous refinement was performed, followed by a 3D refinement that allowed to keep only the best particles to obtain optimal resolution (563 788 particles). Once the best particles have been selected, a homogeneous refinement was performed. The final reconstruction was obtained at a resolution of 2.95 Å (FSC = 0.143) (Supplementary Table S1).

Cryo-EM model building

The initial model of the complex was built using the structure of the nucleosome (PDB ID: 3LZ1) and BCL7A 3D structure predicted by AlphaFold [48] (AF-Q4VC05-F1). BCL7A N-terminal α-helix predicted by AlphaFold [48] did not perfectly fit in the observed extra cryo-EM density on the nucleosome. Therefore, the AlphaFold [48] predicted α-helix was broken in two segments (amino acids 1–12 and 15–30, respectively) and docking into the EM map was performed in UCSF Chimera [49] using the two fragments. Models were manually fitted and reconstructed as necessary in Coot, followed by rigid body refinements and global minimization in Phenix [50, 51] (Supplementary Table S1). Due to the poor cryo-EM density in the region, amino acids 15–30 of BLC7A were modeled as a polyalanine α-helix and rigid-body docking was performed. The fitted and well-reconstructed model was subjected to iterative rounds of refinement in real-space using secondary structure constraints in Phenix [51], and the errors were inspected and adjusted in Coot [52]. Molecular graphics and analysis were performed with PyMOL (PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LLC) or with UCSF Chimera software [49].

Cryo-EM studies of the BCL7A-BAF47-NCP complex

Cryo-EM sample preparation

BCL7A-BAF47 complex and NCP were reconstituted as previously described and mixed at a molar ration of 1:3 (NCP: BCL7A-BAF47). The complex was incubated for 1 h at 4°C before being purified by SEC (Superose 6 Increase 3.2/300). Cryo-EM grids were prepared and imaged as indicated above with acquisition parameters of a physical pixel size of 0.729 Å and a defocus range from −0.8 to −1.9 µm, exposures of 3.84 s which gives a total dose of 48.92 $e^{-}/Å^{2}$ (Supplementary Table S2).

Cryo-EM data processing

The images in the dataset were analyzed as mentioned above and data processing was performed with CryoSPARC2 [47] using the strategy previously described. Data processing statistics are reported in Supplementary Table S2. The final reconstruction was obtained at a resolution of 3.05 Å (FSC = 0.143).

Cross-linking and mass spectrometry studies

Cross-linking and sample processing for mass spectrometry

Complexes of BCL7A and BAF47 in presence or absence of nucleosomes, prepared in 20 mM HEPES (pH 7.5), 100 mM NaCl, 1 mM DTT, were cross-linked without freezing/thawing. Aliquots corresponding to an estimated 50 µg of total protein were diluted to 1 µg/µl final concentration. DSS (d_0/d_{12}, Creative Molecules) was added to a final concentration of 0.25 or 1 mM from a 25 mM stock solution freshly prepared in anhydrous dimethyl formamide, and samples were incubated at 25°C in a ThermoMixer (Eppendorf) with mild shaking. After 1 h, ammonium bicarbonate (AmBic) was added to a final concentration of 50 mM (from a 1 M stock solution) to stop the cross-linking reaction, and samples were incubated for an additional 30 min. Quenched samples were evaporated to dryness in a vacuum centrifuge and redissolved in 50 µl of 8 M urea for reduction, alkylation and enzymatic digestion. Cysteines were reduced by addition of tris(2-carboxyethyl)phosphine to 5 mM, and samples were incubated for 30 min at 37°C. Free thiol groups were then capped with iodoacetamide (10 mM final concentration) by incubating for 30 min at room temperature in the dark. Samples were diluted down to 5.5 M urea with 150 mM AmBic and 0.5 µg of endoproteinase LysC (Wako) was added. After an incubation for 2.5 h at 37°C, the urea concentration was further reduced to 1 M with 50 mM AmBic and 1 µg of trypsin (Promega) was added. Samples were incubated overnight at 37°C, digestion was stopped by adding pure formic acid to a final concentration of 2% (v/v) and digests were cleaned up by solid-phase extraction (50 mg SepPak tC18 cartridges, Waters). Purified peptide mixtures were fractionated by peptide-level SEC on a Superdex 30 Increase column (GE/Cytiva) as described previously [53]. Four fractions were collected per sample, for some samples, the first two low-intensity fractions were pooled for MS analysis.

Liquid chromatography–tandem mass spectrometry

Liquid chromatography–tandem mass spectrometry was performed with an Easy-nLC 1200 chromatography system coupled to an Orbitrap Fusion Lumos mass spectrometer with a Nanospray Flex electrospray ionization source (all from Thermo Fisher Scientific). Duplicate injections for all SEC fractions were performed. Peptides were separated on an Acclaim PepMap RSLC C18 column (250 mm × 75 µm, 2 µm particle size, Thermo Fisher Scientific) at a flow rate of 300 nl/min. The mobile phase gradient was set up as 11 to 40% B in 60 min, and mobile phases were A = water/acetonitrile/formic acid (98:2:0.15, v/v/v) and B = acetonitrile/water/formic acid (80:20:0.15, v/v/v). Mass spectrometric sequencing was carried out in the data-dependent acquisition mode with top speed setting and a cycle time of 3 s. MS scans were performed with detection in the orbitrap analyzer at a resolution of 120 000. Precursors of charge states between 3+ and 7+ were selected for fragmentation in the linear ion trap at a normalized collision energy of 30% and fragments were detected in the orbitrap analyzer at a resolution of 30 000. Dynamic exclusion was activated for 30 s after one sequencing event.

Identification of cross-linked peptides

Mass spectrometry data was converted into the mzXML format using msconvert [54] and searched using xQuest v2.1.5 [55]. The database consisted of the sequences of the target proteins (including canonical H2, H3, and H4 histone sequences)

and identified contaminants/copurified proteins (human keratins, proteins from the *E. coli* expression system). Key xQuest search settings included: enzyme = trypsin, maximum number of missed cleavages = 2, MS mass tolerance = 15 ppm, MS/MS mass tolerance = 15 ppm. Reactive sites for DSS were K and the protein N terminus. Primary search results were further filtered with narrower mass tolerance windows according to the actual mass accuracy at the time of data acquisition. Filters for a minimum xQuest main score ($\geq$25), the fraction of assigned total ion current ($\geq$0.15), and the delta score ($\leq$0.9) were applied, and the remaining candidate assignments were manually evaluated. Only identifications with at least four bond cleavages in total or three consecutive cleavages were accepted (both criteria applied to each of the two peptides). The false discovery rate was controlled by searching against reversed sequences. Due to the small number of target identifications, an accurate estimation of false discovery rates is difficult. However, over all data sets only two decoy hits remained in the curated identification lists. Xlink analyzer [56] v1.1.4 was used to visualize cross-links on the BCL7A–NCP structure.

SAXS data acquisition and analysis

Size-exclusion chromatography small-angle X-ray scattering (SEC-SAXS) experiments were conducted at the SWING beamline [57] at the Synchrotron SOLEIL (Saint-Aubin, France). The incident beam had a wavelength of $\lambda = 1.033$ Å and was detected using an EigerX4M detector (Dectris) positioned 2000 mm from the sample. The direct beam was off-centered, yielding an exploitable q-range of 0.005–0.5 $Å^{-1}$, where the scattering vector is defined as $q = 4\pi \sin \theta/\lambda$, with 2θ being the scattering angle. The samples consisted of 80 µl of BCL7A or NCP, with a stock concentration of 15 and 5 µM, respectively, or the complex NCP:BCL7A at 1:3 molar ratio, and were injected onto a Superose 6 3.2/300 SEC column, which was installed upstream of the SAXS flow cell. Prior to experiments, the column was pre-equilibrated with 20 mM Tris–HCl pH 8.0, 150 mM NaCl, and 1 mM DTT. SAXS data were acquired continuously during elution, which was performed at a flow rate of 40 µl/min.

Primary data reduction was carried out using FOXTROT [58], while further analysis was performed using BioXTAS-RAW [59] and ATSAS [60]. As the NCP sample exhibited a fraction of larger molecular species, Evolving Factor Analysis and regularized alternating least squares (REGALS) deconvolution were applied in BioXTAS-RAW [61, 62] to resolve contributions from distinct molecular populations in NCP and NCP:BCL7A solutions. Structural parameters, including the radius of gyration (R_g), were estimated from the Guinier region and via the pair distance distribution function using Bayesian Indirect Fourier Transform [63]. The maximum particle dimension (D_{max}) was also determined. The degree of compaction was assessed using the normalized Kratky plot in BioXTAS-RAW [61, 62], while molecular masses were estimated using a concentration-independent method based on the volume of correlation (V_c), which is particularly suited for protein–nucleic acid complexes [64].

Circular dichroism spectrometry analysis

Circular dichroism (CD) spectra of the samples at a concentration of 20 µM in a buffer composed of 150 mM NaCl, 20 mM HEPES (pH 7.0), 1 mM DTT were recorded using a Jasco J-815 CD spectrophotometer using a 0.2 mm pathlength CD Quartz cuvette, at 20°C, between 260 and 190 nm, with 10 accumulations, standard sensitivity, a bandwidth of 1 nm and a data-pitch of 0.1 nm. A buffer spectrum was first recorded as baseline and subtracted from the sample spectrum. The spectra were analysed with the CDSSTR [65], Contin-LL [66], and Selcon 3 [67] methods with the protein reference data 7 [67] using the Dichroweb server [68].

Protein sequence bioinformatic analyses

The AlphaFold3 algorithm [69] was queried on the dedicated server, with the protein sequences indicated. The D2P2 server [70] was queried, searching the database with the protein sequences. Analysis with MobiDB-lite [71] was performed using the Python code version 4.0 using Python 3.8.10. The diagram of states was calculated using localCIDER with Python 3.8.10. Amino acid sequence patterns were searched using NARDINI [72]. The Python code was modified to allow search of the extended series of amino acid patterns used in Patil *et al.* [73] For each protein, the native sequence was tested against a null distribution calculated from 100 000 randomizations of the sequence. The z-score matrices were plotted in R v4.3.1 (https://www.r-project.org/) using the Complex Heatmap package v2.18.0 [74].

Remodeling assay

The mammalian SWI/SNF complex was purified from K-562 cells (chronic myelogenous leukemia, ATCC CCL-243), which are wild-type for all the genes coding its subunits (from the COSMIC database). Purification was made possible by the installation of an affinity purification tag at the N-terminus of BAF47, using CRISPR/Cas9 knocking in at the endogenous SMARCB1 locus. The strategy used employed a modified version of plasmid pX335-U6-Chimeric_BB-CBh-hSpCas9n(D10A) (a gift from Feng Zhang [75], Addgene plasmid #42 335; RRID:Addgene_42335) with two U6-sgRNA cassettes in tandem, expressing single-guide RNA (sgRNAs) flanking the initiation codon of SMARCB1, co-transfected with a homology repair template of ~1700 bp centred around the initiation codon followed by the VAP affinity purification tag [76, 77] (Supplementary Fig. S1A). The strategy of tagging at the N-terminus of BAF47 was previously employed successfully to purify mSWI/SNF [78]. Analysis of a selected clone showed successful homozygous editing (Supplementary Fig. S1B–D)

To purify mSWI/SNF from BAF47-tagged cells, ~500E6 cells were grown in spinner flasks and processed as in Mak *et al.* [76] for anti-Flag affinity purification using Flag M2 agarose beads (Sigma). Complexes were eluted in HEPES remodeling assay buffer [HEPES (pH 8.0), 20 mM; KCl, 100 mM; DTT, 1 mM] supplemented with 3xFlag peptide at 0.1 µg/µl. The eluted complexes were concentrated on Amicon Ultra 100 kDa cutoff filters and rinsed twice with HEPES remodeling assay buffer without peptide and used in remodeling assays. This strategy resulted in good purification of the complex (Supplementary Fig. S1E). The presence of all the subunits in the purified complex, including the BCL7 proteins, was confirmed via mass spectrometry analysis (Supplementary Fig. S2).

To mutate the three BCL7 genes, CRISPR/Cas9 editing was used to create indels downstream of the initiation codon of each gene. Three U6-sgRNA expression cassettes were

inserted into plasmid pX330-U6-Chimeric_BB-CBh-hSpCas9 (a gift from Feng Zhang [75], Addgene plasmid # 42 230; RRID:Addgene_42230), with one guide targeting each BCL7 gene. The BCL7A guide plus PAM sequence underlined (GGTCGGTTCGAGCCGAGACGAGG) has the Cas9 cutting site after nucleotide 30 of the coding sequence. The BCL7B guide plus PAM (CCGGTCGGTCCGGGCGGAGACCC) has the cut site after nucleotide 11 of the coding sequence. In the case of BCL7C, a microRNA gene (miR762) is present on the opposite strand overlapping the initiation codon, so targeting was performed at exon 2 after nucleotide 128 (CCTTCGTATCTTCAAGTGGGTGC). The plasmid was electroporated into the VAP-tagged K-562 cell line and cells were grown as single-cell clones, which were amplified and screened for the presence of indels using gene-specific PCR and the T7E1 assay. A clone containing indels at the three genes was selected for further use. Presence of indels was confirmed by cloning products of PCR amplification of the three loci on genomic DNA from the selected clone followed by sequencing at least 20 colonies for each locus (Supplementary Fig. S3A). Western blots on purified complexes were performed (Supplementary Fig. S3B) using the following antibodies: rabbit polyclonal against BCL7A (Atlas antibodies HPA019762, raised against residues 136–208 of Q4VC05.1), rabbit polyclonal BCL7B (Proteintech #11740-1-AP, raised against the entire protein), and mouse monoclonal anti-BCL7C (US Biologicals #243754, raised against the entire protein, unmapped epitope). This characterization showed that BCL7A has three alleles, with insertion of 1 base pair or deletion of 2 or 28 bp, all causing frameshifts. A shorter BCL7A protein was detected in western blot, likely arising from initiation at Met22, past the N-terminal alpha helical region. BCL7B showed three alleles, with deletion of 2, 12, or 13 base pairs. They cause frameshifts affecting codons 6 and onward or a deletion of 4 codons (from Val6 to Glu9). In western blots, the abundance of BCL7B was not visibly affected but the protein ran at a slightly lower molecular weight, suggestive of initiation at Met22 (Supplementary Fig. S3B). BCL7C has also three alleles, with deletions of 1, 5, or 13 base pairs causing frameshifts after codon Thr41, Arg44, or Ile45. The BCL7C protein was undetectable. Considering that all alleles of the BCL7 genes carry some defect, coding for proteins that are either very short, or truncated in their N-terminal helix, or entirely missing that helix. This genotype was further validated by inspecting high-throughput sequencing results from ATAC-seq experiments (see below). For simplicity, the cells are referred to as triple knock out or TKO.

The western blot membranes also contained equal amounts of each recombinant protein (100 ng, or 4.0, 4.5, and 3.8 pmoles for BCL7A, BCL7B, and BCL7C, respectively), which revealed that the anti-BCL7B antibody is not entirely specific and crosses slightly with BCL7A and BCL7C. Further, while comparing the western blot signal from the purified complex with that from the recombinant proteins, it appears that the amount of BCL7A in the loaded aliquot of WT complex is around 100 ng, that BCL7C appears above 100 ng, and that BCL7B is the least abundant family member, at levels markedly lower than 100 ng. Finally, the size of the BCL7A protein in the WT complex is lower than that of the recombinant protein, suggesting that the predominant isoform expressed in K-562 cells is the shorter one (Q4VC05-1, 210 residues). These results match those obtained by mass spectrometry on the purified mSWI/SNF complex (Supplementary Fig. S2). Protein content of the parental and TKO eluates was estimated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis of various eluate amounts by western blot against BRG1 and BAF47 (Supplementary Fig. S4A). Mass spectrometry on purified complexes confirmed the reduction in BCL7 paralog abundance in the TKO samples (Supplementary Fig. S4B and C).

The template DNA for remodeling was generated by PCR using the Widom 601 positioning sequence cloned in pBluescript and amplified by PCR with primers placed such that the 147 nucleotides of positioning sequence are flanked by 51 and 65 base pairs of non-nucleosomal DNA and bearing Cy5 dyes at the 5′ termini. Due to the position of the extensions relative to the MfeI site, only octamer eviction or sliding in one direction will fully expose the MfeI restriction site. The remodeling reactions were carried out using 1 μl of eluted complexes, corresponding to ∼10 ng of BRG1, in a final volume of 20 μl, implying a complex concentration of ∼2.5 nM. Reactions contained 2 mM HEPES (pH 8.0), 10 mM KCl, 2% (v/v) glycerol, 0.2 mM EDTA, 0.1 mM DTT, 5 mM $MgCl_2$, 1 mM ATP, 0.1 mg/ml BSA, 40 U MfeI (NEB), and 9 nM Cy5-labeled NCP, supplemented with protease inhibitor cocktail. Reactions were carried out at 30°C for various durations and stopped by addition of 4% (w/v) sodium dodecyl sulfate and 10% (v/v) glycerol. Samples were digested with 20 μg of proteinase K and incubated 1 h at 37°C, precipitated with 2 volumes of ethanol. DNA pellets were dried, resuspended in 20 μl of Tris-EDTA (TE) buffer and 4 μl of 50% (v/v) glycerol. Samples were separated on 2% agarose gels and scanned on a GE Healthcare Typhoon scanner, using the highest photomultiplier tube voltage that avoided pixel saturation. Raw fluorescence data were acquired in Fiji (ImageJ v1.54), using the Analyze Gels function. Signal across gel lanes for the intact DNA fragment as well as the two restriction products was acquired. The signal of each of the three bands was normalized by dividing the raw signal by the sum of raw signals in the corresponding lane. The normalized signal of the two cut fragments was summed, since each has half the Cy5 dye contained in the intact fragment. Cut fragment signal was plotted as a function of time and genotype. The experiment was performed with two different preparations of the parental and TKO complex, each time in three replicate reactions. Repeated measures two-way Analysis of Variance (ANOVA) was used to test significance.

ATAC-sequencing

ATAC-seq was performed on the same CRISPR/Cas9-edited cells used for purifying the mSWI/SNF complex for *in vitro* remodeling assays. To confirm the edited cells genotype, trimmed fastq files from the ATAC-seq experiment were aligned to GRCh38 using bwa-mem2 [79] using default settings. BAM files from the same cell line were merged, variants were called using FreeBayes (Preprint at https://doi.org/10.48550/ARXIV.1207.3907) and results were inspected on the IGV browser [80].

To help identify genomic loci where DNA accessibility depends on mSWI/SNF activity, cells were left untreated or treated for 6 h with BRM014 (MedChemExpress) at 10 μM in Dimethyl Sulfoxide (DMSO) to block BRG1 and BRM en-

zymatic activity [81, 82]. Each condition was represented by two biological replicates (experiments performed on different passages of the cells, on different days).

Fifty thousand cells were used per ATAC-seq library. A kit from Active Motif was used (cat. #53150) and manufacturer instructions were followed. Sequencing was performed in paired-end mode (2 × 50 bp) on an Illumina NextSeq 2000 to an average of 41 million pairs per sample. Raw data were processed as described in [83], aligning on GRCh38 version of the human genome. At least 32 million read pairs were uniquely mapped per sample. Differential DA analysis was performed using a sliding window approach (50 bp wide, shifted by 50 bp) quantifying the transposase insertion sites to be published elsewhere (https://doi.org/10.1101/2022.03.16.484118). All analyses were performed in R/Bioconductor (https://www.r-project.org/). To focus the analysis on mSWI/SNF binding sites where ATPase activity is needed for DNA accessibility, binding sites of BRG1 in K-562 cells identified as part of the ENCODE project were used (accession ID ENCFF267OGF [84]). BRG1 binding sites overlapping sliding windows where BRM014 treatment led to a decrease in transposase insertions of at least 50% with a Benjamini–Hochberg-corrected *P*-value no larger than 0.05 were retained. Over these BRG1-bound and BRG1-dependent regions, the presence of sliding windows with significantly different DA in BCL7 TKO compared to parental (corrected *P*-value < .05) were recorded along with their associated $\log_2$ fold-change.

The heatmap of differential ATAC-seq signal was plotted using *ComplexHeatmap* (v2.18) [85], where each row represents a region where BRM014 elicits a significant decrease in DNA accessibility and overlaps a BRG1 binding site. The values plotted were the sum of Trimmed Mean of M-values (TMM) normalized counts over the windows making up the DA region, scaled (null mean and unity standard deviation) to obtain Z-scores. Rows were clustered based on whether the effect of TKO compared to parental was a decrease in accessibility, an increase, or no statistically significant change. Boxplots represent the $\log_2$FC values used to cluster the heatmap and were plotted using *ggplot2* (v3.4.4). Example plots of ATAC-seq read coverage were generated using Gviz [86], employing the counts-per-million normalized read counts in bins of 10 base pairs. Assessment of distance to genes was performed using *ChIPseeker* (v1.38.0) [87] defining the promoter region as the 2 kb area centered around gene TSS. Overlap with previously published ChIP-seq datasets was calculated at the ChIP Atlas website [88], using regions that were stretched to 2 kb in length (fixing the center) and using a ChIP-seq peak score cutoff of 50; only peaks from datasets in the K-562 cell line were retained. Regions with lower Donor Allele (DA) in TKO were compared with regions with higher DNA accessibility. Enriched DNA sequence motifs were identified using *universalmotif* (https://doi.org/10.18129/B9.BIOC.UNIVERSALMOTIF) within the 300 bp around the center of the DA regions, and motifs from the JASPAR 2022 database accessed through Bioconductor [84]. The background set of sequences used in the motif enrichment calculations were the sequences in the entire set of BRG1-bound and BRG1-dependent regions. ATAC-seq and ChIP-seq read coverage plots at various regions were generated using deepTools (v3.5.1) with the functions bamCoverage, computeMatrix, and plotHeatmap [89].

Perturb-seq analysis

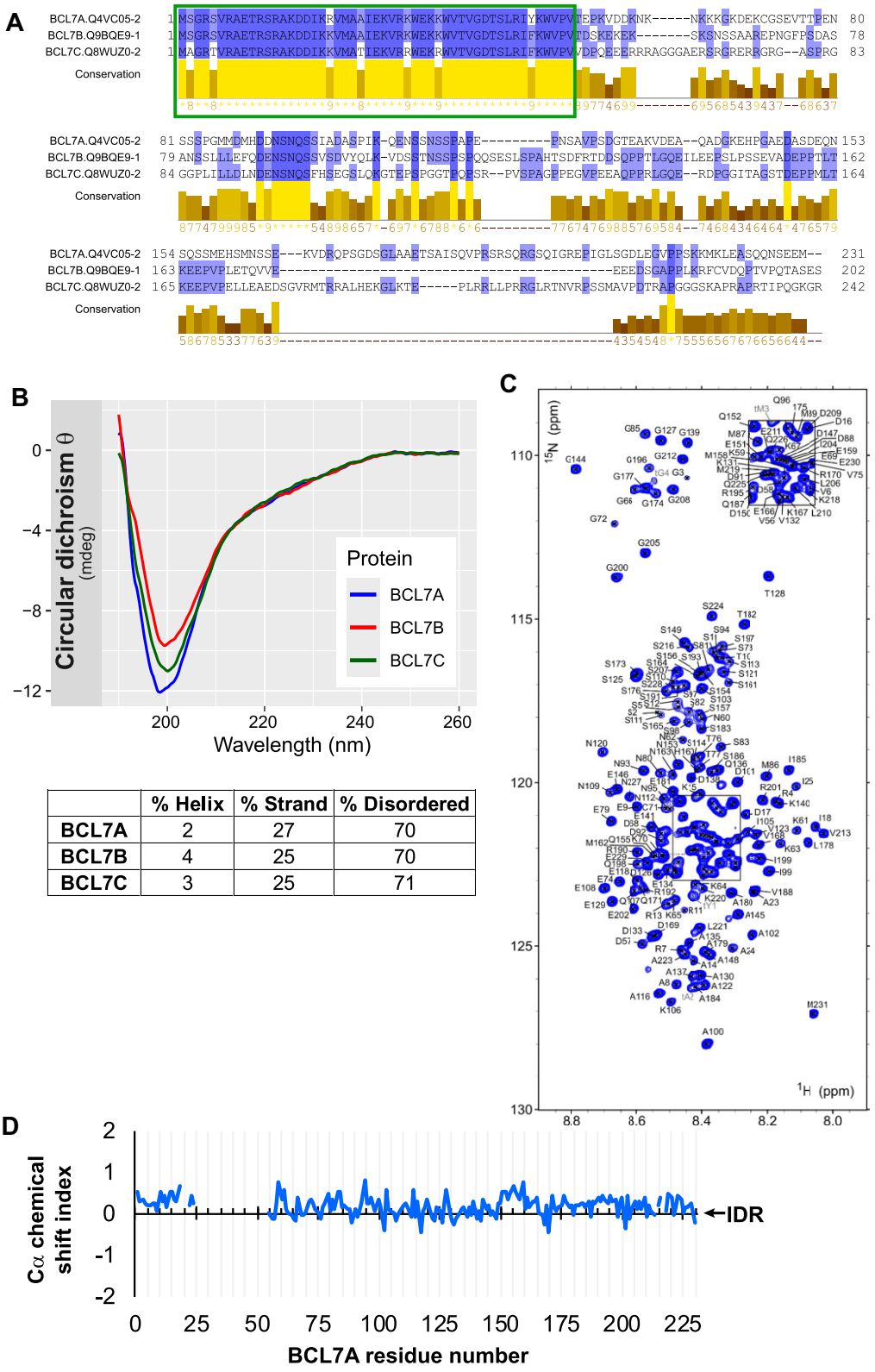
Perturb-seq sequencing data from Otto *et al.* were downloaded from the NCBI GEO under accession GSE200201 [90]. Similar data processing steps as described in Otto *et al.* were employed, using the Seurat package [91] in R/Bioconductor. The data processing scheme involved retaining cells with at least 200 detected messenger RNA (mRNA) and retaining genes expressed in at least three cells, then filtering out cells with <1000 reads or with >20% mitochondrial genome reads. The counts were normalized so that each cell would have a sum of counts of 10 000, then expression values were added to 1 and $\log_2$-transformed. The 23 863 assayed genes were further filtered to retain only those that are reliably detected, with mean normalized expression across cells between 0.0125 and 4. This resulted in retaining 11 212 genes. The normalized expression matrix was then scaled so that each gene would have a mean expression of zero and a standard deviation of 1. For differential expression calculation, sgRNAs were collapsed to the mSWI/SNF subunit gene they target and all cells with perturbation of a given subunit were retained. Only cells containing a single perturbation sgRNA were retained, except for the analysis of the double perturbation of BCL7 genes. Cells with nontargeting sgRNAs were randomly split into two groups, one to serve as invariant reference set for differential expression calculation and the other to serve as a control, null-effect set. Differential expression was calculated by taking the mean of scaled normalized expression values of the cells with a subunit perturbation and subtracting the mean of the reference set of cells. The boxplot was generated with the dedicated base R function and defaults parameters, except that notches were drawn with defaults and outliers beyond 1.5 times the interquartile range were removed for clarity.

Results

Sequence comparison and structural predictions of BCL7 proteins

The BCL7 family is composed of three proteins encoded by three different genes: BCL7A, B, and C. They lack any domain of known molecular function or shared with other proteins. The longest mRNA isoform of human BCL7A comprises six exons and codes for a protein of 231 amino acids. A second mRNA isoform exists that shortens the end of exon 5 by 21 codons. According to multiple sequence alignment results of the three human BCL7 family members, the first 51 amino acids of each protein are highly homologous, while homology is more limited beyond this position (Fig. 1A). The sequence of each family member is conserved throughout evolution across the entire length of the proteins (Supplementary Fig. S5). The three family members likely exist because of ancient duplication events of a single primitive BCL7 gene and are considered paralogs.

Sequence-based analyses using AlphaFold3 [69] and a panel of intrinsic disorder prediction algorithms [70, 71] predict that the majority of BCL7A length is disordered except for an alpha helix from residues R4–R29, and a beta hairpin from residues W31–P50 (Supplementary Fig. S6A–C). Comparison with the sequences of BCL7B and BCL7C yielded qualitatively similar results. Diagram of states analysis using localCIDER [92] of the BCL7 protein sequences, or just their predicted disordered segments, places them in the boundary zone where a majority of intrinsically disordered proteins (IDPs)



	% Helix	% Strand	% Disordered
BCL7A	2	27	70
BCL7B	4	25	70
BCL7C	3	25	71

Figure 1. Sequence and disorder of BCL7 proteins. (**A**) Multiple sequence alignment of the three human BCL7 proteins performed using Muscle in JalView. The first 51 amino acids, highly homologous between the three proteins, are highlighted in a green box. They correspond to the BCL_N domain reported in the Pfam database of protein domains (PF04714). (**B**) CD analysis of the three BCL7 proteins. (**C**) Analysis of free BCL7A by NMR. ^{1}H-^{15}N SOFAST-HMQC spectrum of the full length uniformly ^{15}N labeled BCL7A. The assigned signals are labeled (labels in the central region are shown in the top-right corner inset). (**D**) Cα chemical shift index for all assigned residues.

localize [93] (Supplementary Fig. S6D). We also searched the disordered segments' amino acid sequences for short patterns that would be enriched, as this was shown to be implicated in the phase separation of ARID1 proteins [73]. This analysis performed with NARDINI [72] revealed that each protein has a distinct set of such patterns (Supplementary Fig. S6E): BCL7A most notably has blocks of polar residues alone or combined with positively charged ones; the most salient pattern in BCL7B are blocks of serine residues; BCL7C is characterized by blocks of glycines with hydrophobic or negatively charged residues. These are markedly different from the patterns most characteristic of ARID1A IDR1 (blocks containing glutamine, alanine, proline or glycine), or of ARID1A IDR2 (blocks of glutamine with serine or glycine) [73].

CD spectra analysis revealed that the three BCL7 paralogs have a very similar secondary structure (Fig. 1B). To obtain structural information on BCL7A, NMR analysis of a ^{15}N-labeled BCL7A solution was performed. As predicted, the protein spectrum exhibits all the characteristics of an IDP, with limited proton amide spectral dispersion. All visible signals were assigned and span from residues 1–15 and 57–231 (Fig. 1C). The calculated chemical shift index shows near to zero values for most residues except slightly positive values for >4 consecutive residues in two stretches (1–20 and 150–160), indicating that the two assigned segments are mostly fully disordered with a very low helical propensity for the first 20 residues and for the segment 150–160 (Fig. 1D). In contrast, the signal from residues in the 26–56 segment was absent from the spectrum most probably due to some conformational exchange in the microsecond-millisecond timescale, as can often be seen for secondary structures in the absence of tertiary structure. Interestingly, this particular segment correlates with the hairpin secondary structure predicted by AlphaFold3 (Supplementary Fig. S6A).

BCL7 proteins form a stable complex with the nucleosome

Because mSWI/SNF acts on nucleosomes, electrophoretic mobility shift assays (EMSA) experiments were performed to examine binding of BCL7A to the NCP. NCPs were assembled using recombinant *X. laevis* histone octamers and fluorescently labeled 601 positioning DNA sequence [32]. EMSA on these nucleosomal probes revealed that BCL7A can stably bind this substrate (Fig. 2A). Curve fitting to explain probe electrophoretic retardation worked best using the Hill equation and indicated that the binding affinity of BCL7A on the NCP probe has an apparent dissociation constant (K_D) of 290 nM (Fig. 2B and D). The Hill coefficient n of 3.7 suggested binding cooperativity of BCL7A to the probe. As shown in previous studies [94] the Hill coefficient only describes the transition cooperativity, rather than exactly estimate the number of protein molecules bound to the nucleosome.

Similar K_D values were obtained for the three BCL7 proteins. To identify which part of the BCL7A protein might be implicated in NCP binding, EMSA experiments were repeated with the BCL7A (1–100) truncation mutant, where most of the predicted disordered C-terminal region is removed. It was found that the first 100 amino acids of BCL7A are sufficient for NCP binding, with an affinity not significantly different from that of BCL7A WT (K_D of 320 nM).

Further, we examined BCL7-nucleosome interactions using BLI experiments. To perform BLI measurements, poly-histidine-tagged mononucleosomes were immobilized on Ni-NTA biosensors and incubated with different amounts of BCL7 proteins; wavelength shifts of the reflected light from the biosensor surface were measured and plotted against the different BCL7 concentrations. Analysis of the binding profiles resulted in calculated apparent K_D values of 220 nM for BCL7A, 170 nM for BCL7B, and 220 nM for BCL7C (Fig. 2C and D), supporting the measurements obtained via EMSA assays.

Characterization of the BCL7A–NCP complex

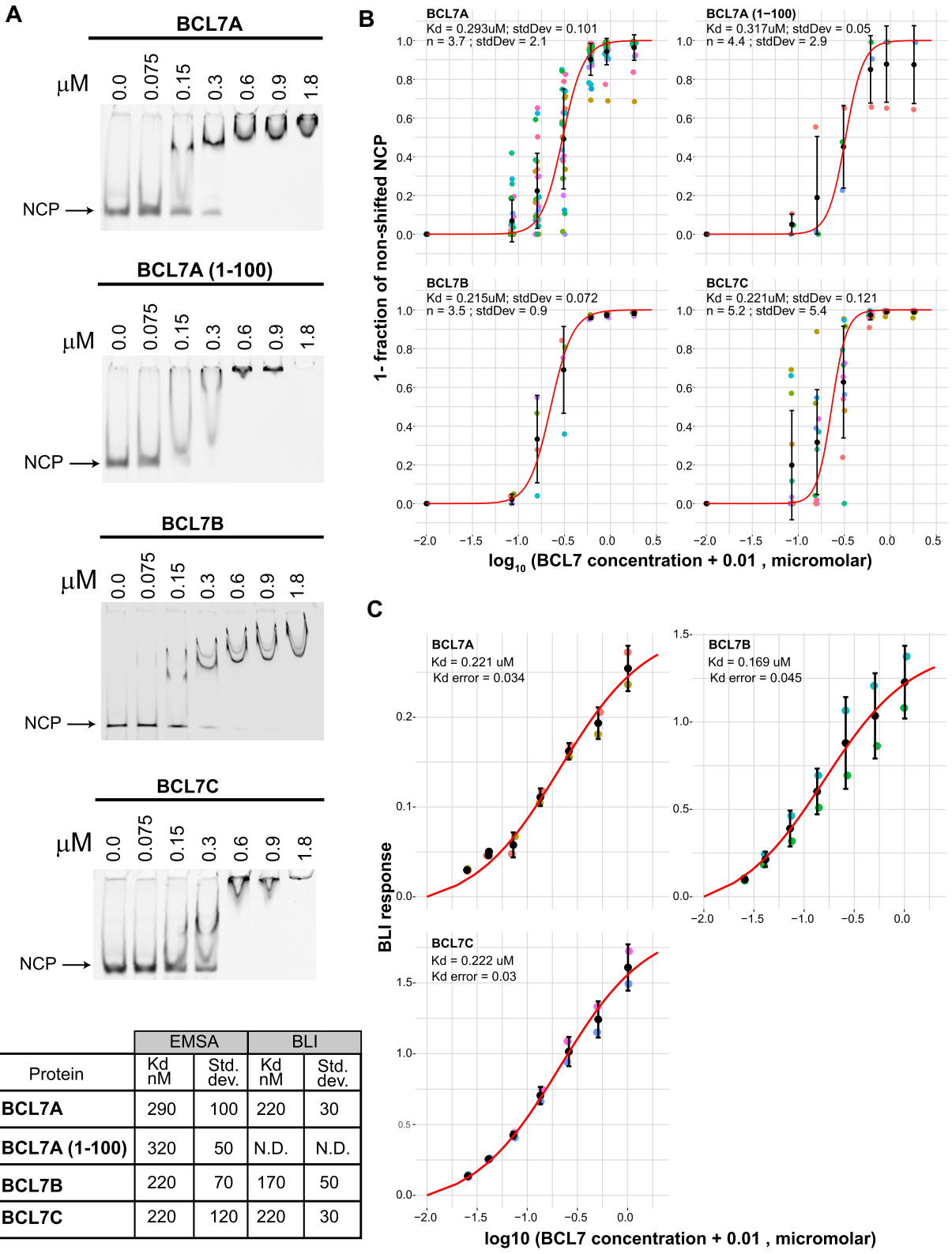
The complex formed by BCL7A and the NCP was analyzed by SEC. NCPs were incubated with a 1:5 molar excess of full length BCL7A and the sample was purified on a Superose-6 Increase 3.2/300 column. The chromatogram displayed a single symmetric peak, and the corresponding fractions contained BCL7A and the histones, indicating the formation of a stable complex of BCL7A with the NCP (Supplementary Fig. S7A).

Complex stoichiometry was examined by SEC followed by MALS, comparing NCP alone and the NCP-BCL7A complex with an excess of BCL7A (Supplementary Fig. S7B). Both samples appeared mono-disperse. The NCP showed a mass of 214.67 kDa ± 1.68% (198.8 kDa calculated), the excess free BCL7A showed a molecular mass of 29.53 ± 10.32% kDa (25.0 kDa calculated) and the NCP-BCL7A complex yielded a molecular mass of 233.48 ± 1.79% KDa (223.8 kDa calculated). The results suggest that BCL7A is a monomeric species and that the NCP-BCL7A complex has a 1:1 stoichiometry at this concentration.

To better understand the BCL7A–NCP complex identified by EMSA experiments, we employed SV analytical ultracentrifugation (SV-AUC). Full length BCL7A is a monomer in solution with a sedimentation coefficient of 1.87 S, which corresponds to the MW of about 26 KDa, and has a high frictional coefficient due to the protein's disordered nature (Supplementary Fig. S7C and D). The NCP alone sediments as a homogeneous 11.35 S species, consistent with earlier studies of isolated nucleosome particles [95, 96] and the BCL7A–NCP complex sediments as 10.52 S species, suggesting the formation of a 1:1 molar ratio complex. The fact that the BCL7A–NCP complex sediments more slowly than the NCP alone could be imputable to the increased frictional coefficient of the complex compared to NCP alone due to the binding of the NCP to an IDP as previously described [96].

Cryo-EM structure of the BCL7A–NCP complex

To gain further structural information on how BCL7A interacts with the NCP, we used cryo-EM and obtained a structure of the BCL7A–NCP complex (Fig. 3) at an overall resolution of 2.95 Å (Supplementary Figs S8 and S9). The structure shows that the first 30 amino acids of BCL7A engage the acidic patch of the NCP (Fig. 3A–C, Figs 4 and 5, and Supplementary Fig. S9). BCL7A N-terminal region folds in two distinct α-helices: a short α1 helix (aa 3–9) and a longer α2 helix (aa 15–30) connected by a loop of which two residues (R13 and A14) are not visible, probably due to their flexibility (Fig. 3B and C). The N-terminal residues of BCL7A (aa 1–12) host an arginine anchor structural motif that contacts the cavity of the NCP acidic patch defined by helices α2 and α3 of histone H2A (Figs 3A–C and 5). BCL7A α2 helix bends on histone H2B C-terminal α-helix and continues toward the nucleosomal DNA to the super helical location 4.5 (SHL + 4.5;



D

Protein	EMSA		BLI	
	Kd nM	Std. dev.	Kd nM	Std. dev.
BCL7A	290	100	220	30
BCL7A (1-100)	320	50	N.D.	N.D.
BCL7B	220	70	170	50
BCL7C	220	120	220	30

Figure 2. Binding of BCL7 proteins to the nucleosome. (**A**) EMSA gels using Cy5-labeled NCP and the indicated BCL7 proteins. Protein amounts were varied as indicated above each lane. (**B**) Quantification of the calculated amount of bound NCP and curve fitting using the Hill equation (red curve). Dissociation constants and Hill coefficients 'n' are indicated, along with their standard deviation from replicates of the experiment. The colors of the dots represent the various replicate titrations, with $n \geq 3$. Black dots and brackets indicate the mean and standard error. (**C**) BLI response curves for the indicated proteins, with fitting of a heterogenous binding model curve in red. Dots of different color indicate the two biological replicates of the experiment. Black dots and brackets indicate the mean and standard deviation of the replicates. (**D**) Table of the calculated dissociation constants. Numbers rounded to the closest ten. N.D.: not determined.

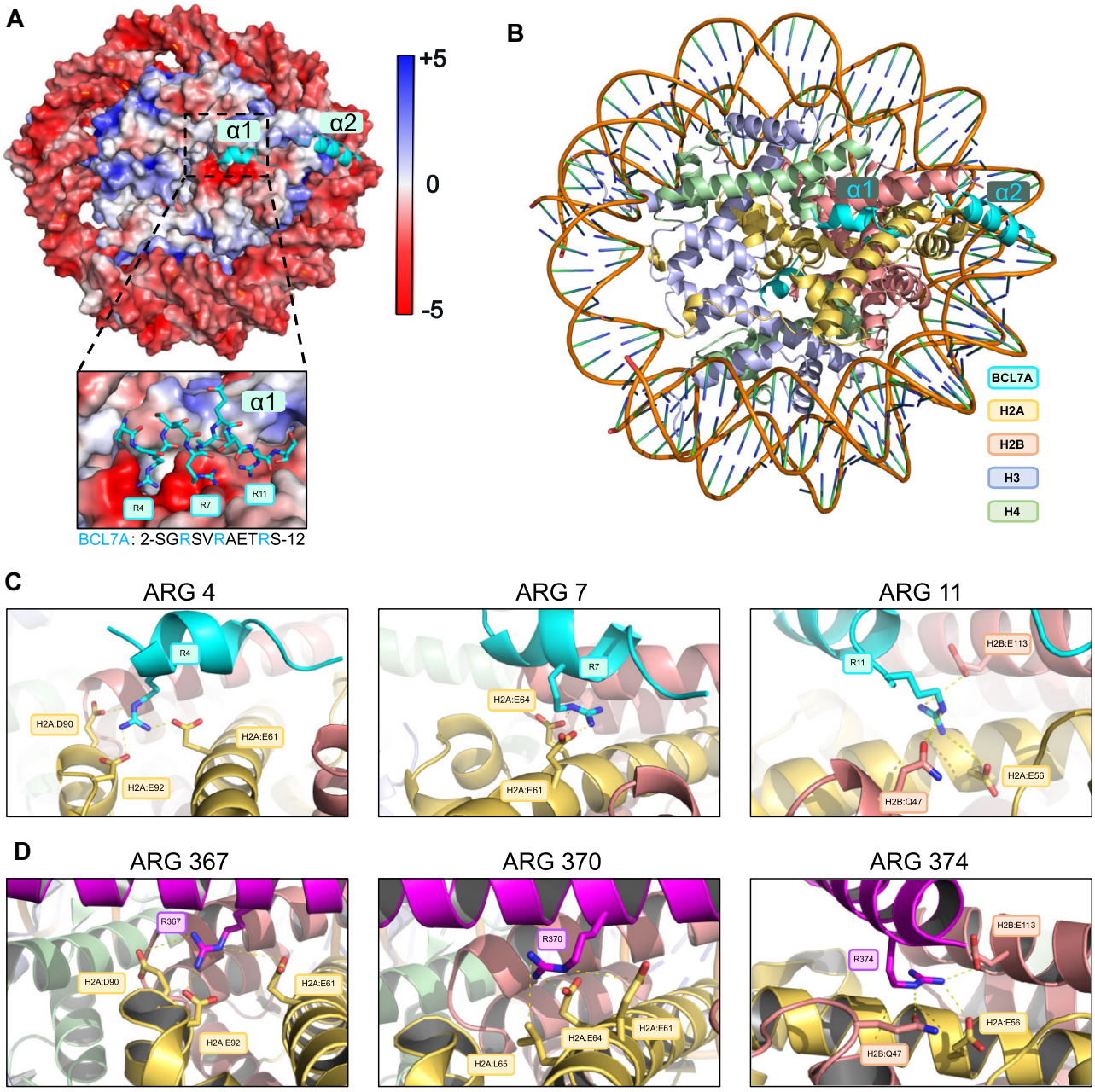


Figure 3. Cryo-EM structure of BCL7A in a nucleosome bound state. (**A**) Molecular surface representation of the nucleosome core particle colored according to the electrostatic potential (blue, positive [+5 kT]; white, neutral; red, negative [−5 kT]) in complex with BCL7A; BCL7A N-terminal region (aa 2–30) is shown in ribbon diagram representation and colored cyan. The panel below shows a zoomed view of BCL7A interaction with the nucleosome: amino acids 2–12 of the N-terminal α helix BCL7A are shown in stick representation where carbon atoms are colored cyan, oxygen atoms are red, nitrogen atoms are blue. The exact amino acid sequence of BCL7A N-terminus (aa 2–12) is indicated and the three Arg within the arginine anchor motif are colored blue. (**B**) Ribbon diagram representation of BCL7A in complex with the acidic patch of the nucleosome. BCL7A is colored cyan, DNA is colored orange and the histones H2A, H2B, H3, and H4 are colored yellow, red, light blue, and green, respectively. (**C**) Zoomed-in views of the interactions of BCL7A R4 (left panel), R7 (center panel), and R11 (right panel) with H2A and H2B. Residues of interest are shown in stick representation. BLC7A and histones are color coded as above. (**D**) Zoomed view of BAF47 C-terminal helix (PDB ID: 6LTJ) interaction with the acidic patch. BAF47 is colored purple and histones are colored as mentioned above. Residues directly involved in the interactions are in stick representation. Panels (A), (B), and (C) were rendered with PyMOL.

Supplementary Fig. S10). BCL7A arginine anchor motif is characterized by three arginines, R4, R7, and R11, that establish a secure network of hydrogen bonds and ionic interactions with the conserved acidic residues of histones H2A and H2B. R4 projects deep in the canonical arginine-anchor binding pocket forming hydrogen bonds with H2A E61, D90 and E92. BCL7A R7 establishes hydrogen bonds with H2A E61 and a salt bridge with E64; R11 contacts H2A E56, and H2B Q47 and E113 establishing hydrogen bonds (Fig. 3C; Fig. 4). Following the nomenclature of McGinty and Tan [97], R4 constitutes a canonical arginine anchor, R7 is a type 2 arginine anchor binding the cavity right next to the canonical arginine binding cleft, and R11 is a type 1 variant arginine. BCL7A α2 helix appears as a tube of density, clearly in contact with hi-

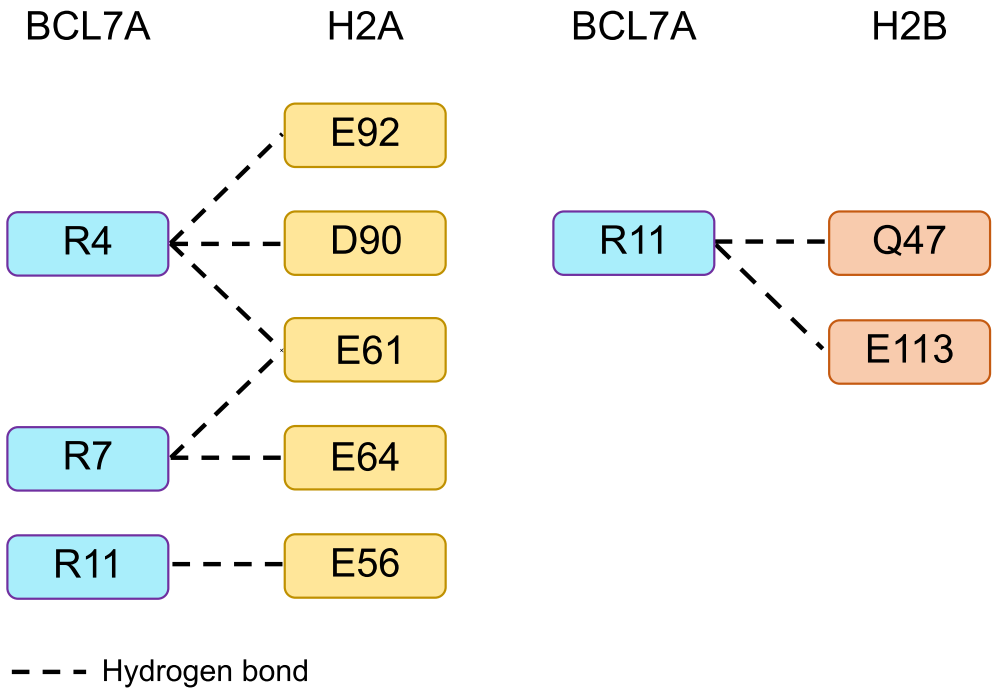


Figure 4. Contact points between BCL7A and histones H2A and H2B. Residue numbers are indicated in boxes colored light blue for BCL7A, yellow for H2A, and red for H2B. The interactions are marked by dotted lines.

stone H2B C-terminal α helix, H2A α1 helix, the beginning of H2A tail and it ends in close proximity with the phosphate backbone of the nucleosomal DNA, on the minor grove of the DNA at SHL + 4.5. However, the interpretation of the electron-density map in this region is challenged by the lower resolution and the fact that BCL7A α2 helix was built as a poly-alanine chain. There is no observed density for the disordered C-terminal tail of BCL7A. A second molecule of BCL7A is visible on the acidic patch on the opposite side of the nucleosome, with identical nucleosome binding mode.

Characterization of the BCL7A–NCP complex by NMR

To map the interaction region between BCL7A with NCP, ^{15}N-^{1}H 2D NMR spectra of uniformly labeled ^{15}N-BCL7A were recorded at different ratio BCL7A:NCP (1:0, 1:0.2, 1:0.33, and 1:1) (Fig. 6A and B). Due to the large NCP size, the intensity of signals corresponding to interacting residues are expected to vanish. The loss of intensity for isolated signals was plotted for the different BCL7A:NCP ratio (Fig. 6C). It was observed that signals for the residues located in the N-terminal segment of BCL7A (amino acids 1–15) almost vanish even after addition of only 0.2 molar equivalent of NCP. An important signal intensity decrease was also observed in two C-terminal segments (amino acids 192–202 and 216–224, respectively; Fig. 6C). Finally, a 50% intensity loss was observed in the segment corresponding to amino acids 57–80, but only at 1:1 ratio. From this experiment, we conclude that the regions whose signal intensity vanishes upon titration with the NCP are likely to be contacting the nucleosome.

SAXS analysis of the BCL7A–NCP complex

SAXS was employed to provide structural insights into the BCL7A–NCP complex, including the disordered segment of BCL7A (Fig. 7). SAXS analysis revealed that BCL7A eluted as a single peak, with scattering data characteristic of an elongated protein, exhibiting large radius of gyration (R_g) and maximum dimension (D_{max}) values for a 231 residues protein (Fig. 7A-B). The dimensionless normalized Kratky representation (Fig. 7C) confirmed that BCL7A is predominantly unstructured, consistent with predictions from AlphaFold [69, 98], suggesting the presence of secondary structure elements restricted to the N-terminal region. Molecular mass estimation based on volume correlation [64] yielded a value of 39 kDa (Fig. 7D), which is slightly higher than but compatible with the theoretical mass of the monomer (25 KDa). The Kratky plot (Fig. 7C) confirmed that the complex was composed of compact particles. Guinier analysis of the main peak revealed an increase in R_g from ∼43 to ∼45 Å upon complex formation, accompanied by an increase in estimated molecular mass from 217 to 243 kDa (Fig. 7D). These findings support the binding of BCL7A to the NCP. Notably, BCL7A appears to adopt a more compact conformation upon NCP binding, as indicated by a D_{max} increase relative to NCP alone, yet remaining significantly lower than that of free BCL7A in solution. This observation aligns with NMR data, which suggest additional interactions involving the C-terminal region of BCL7A upon complex formation with NCP.

BCL7A- BAF47 interaction

Our observation of an interaction of BCL7A at the NCP acidic patch brought up the question of how this can accommodate BAF47, given that this protein has been reported to contact the acidic patch as well [24, 28–30, 99]. Previous CX-MS data reported an interaction between BCL7A and BAF47 [8]. We have identified that BAF47 and BCL7A are direct binding partners and isolated the BCL7A-BAF47 complex by SEC;

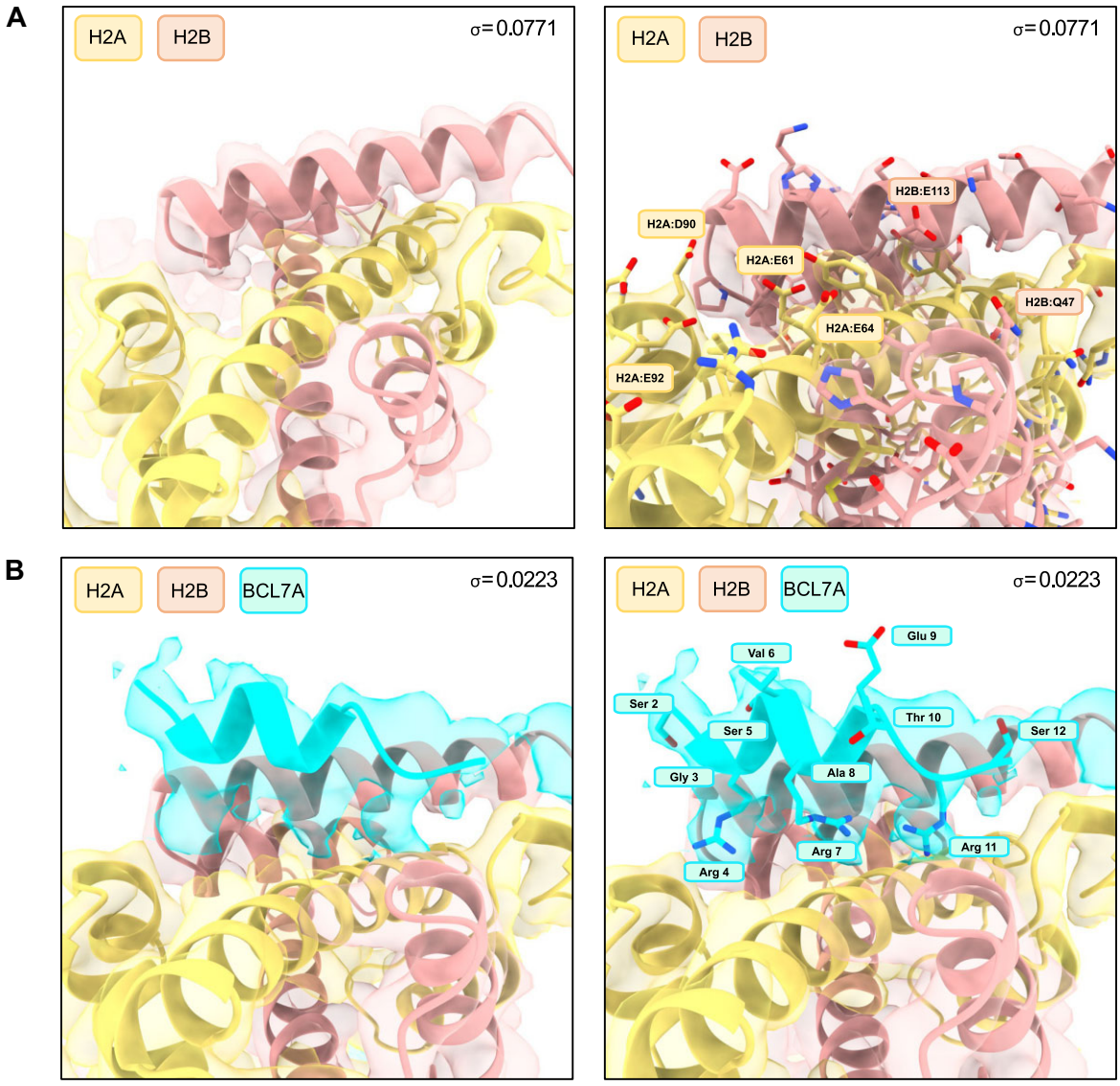


Figure 5. Detailed view of the BCL7A–NCP cryo-EM density map at the acidic patch filled with BCL7A coordinates displayed at two different map thresholds. (**A**) Atomic models derived from the density maps show histones H2A (yellow) and H2B (red) in ribbon representation (left panel) and in stick representation (right panel) at a threshold of σ = 0 077. (**B**) Same as in A but the map threshold is set at σ = 0,0223. BCL7A density (displayed in cyan) is visible and positioned on the top of H2A and H2B and showed in ribbon representation (left panel) and stick representation (right panel). The figure was prepared using ChimeraX software in order to adjust the threshold and color the different histone and BCL7A densities as needed.

point of contacts between the two molecules were further verified by cross-linking coupled with mass spectrometry (CX-MS) (Supplementary Table S3 and Supplementary Fig. S11) and by NMR experiments performed with BCL7A and shorter fragments of BAF47 (Supplementary Fig. S12). While the N-terminal region of BCL7A seems to be responsible for direct binding to BAF47, BAF47 appears to contact BCL7A at different points throughout its entire sequence, encompassing residues within the WH-DBD, RPT1, RPT2, CTD domains (Supplementary Fig. S12A). Further we reconstituted the BCL7A-BAF47 complex in presence of nucleosome, purified a stable ternary complex by SEC (Supplementary Fig. S13A and B) and performed cryo-EM studies (Supplementary Figs S13C–F and S14). Upon cryo-EM data processing we obtained a low-resolution reconstruction of the BAF47–BCL7A–NCP complex: an arch shaped density branches out from the acidic patch and reaches the DNA on the opposite side of the nucleosome at the dyad axis (Supplementary Fig. S13D). The map features are elongated and cannot be interpreted with an atomic model. This is probably due to BAF47 flexibility when not in contact with its binding partners within the SWI/SNF complex. To resolve the regions that are less well defined in the cryo-EM density, CX-MS provided valuable information on the proximity between amino-acid residues of the molecules (Supplementary Table S3). CX-MS confirmed that in a nucleosome bound state several BCL7A–BABF47 points of contact are retained and that BCL7A still contacts the nucleosome at least at positions H2BK109 and H3K28 (Supplementary Fig. S15 and Supplementary Table S3). To better understand the location of BCL7A and BAF47 on the acidic patch, the map was refined to 3.05 Å in proximity

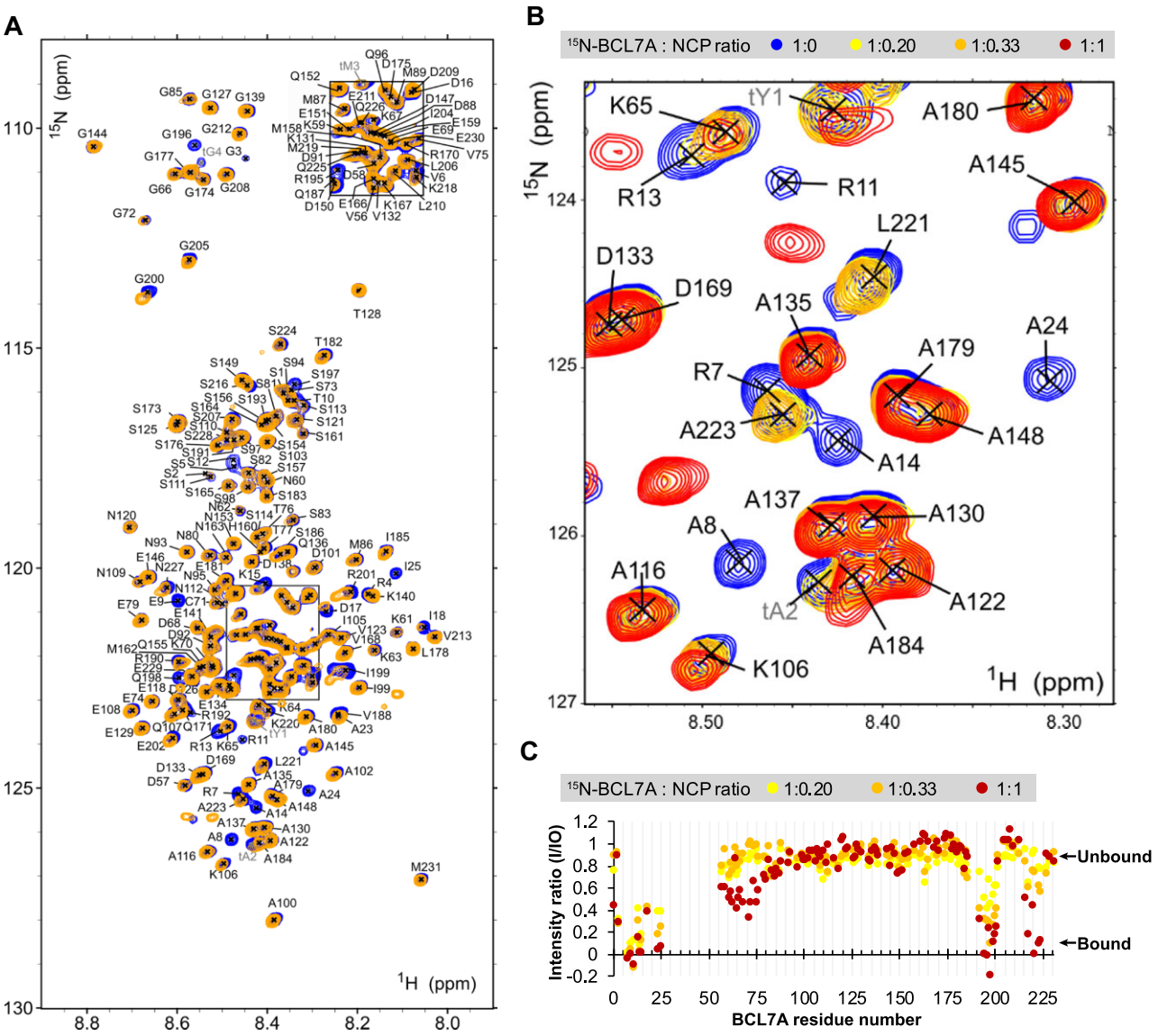
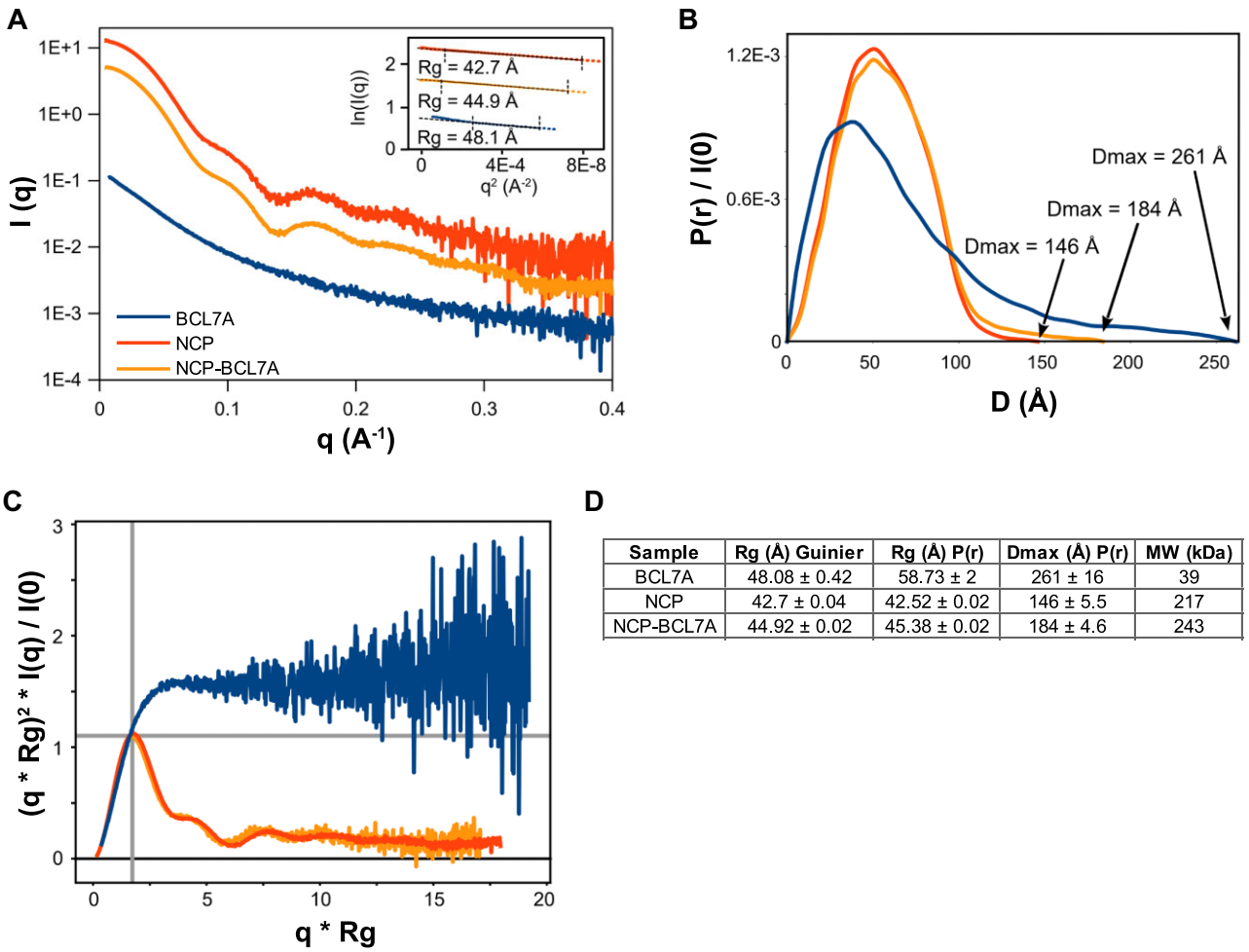


Figure 6. Analysis of BCL7A–NCP complex by NMR. (**A**) Overlay of the ^{1}H-^{15}N spectra of uniformly ^{15}N-labeled BCL7A alone (in blue, as in Fig. 1E) and after addition of 0.33 equivalent of unlabeled NCP (in orange). (**B**) Zoom of the ^{1}H-^{15}N spectra of uniformly ^{15}N-labeled BCL7A alone (in blue) and after addition of 0.2, 0.33, and 1 molar equivalent of unlabeled NCP (in yellow, orange, and red, respectively). (**C**) Mapping BCL7A–NCP interaction region with the ratio of intensities (I/I0), where I and I0 are the intensities of isolated signals of the ^{1}H-^{15}N spectra before and after addition of NCP at different ratio (with the same color code as in B).

to the nucleosome. However, BCL7A-BAF47 density has a resolution of ∼4–5 Å making its interpretation difficult. Via 3D classification, we focused on a subset of particles where the nucleosome has one molecule of BCL7A on one side (Supplementary Fig. S13E, left panel) and both BCL7A and BAF47 molecules on the other side (Supplementary Fig. S13E, right panel). By fitting the BAF47–BCL7A–NCP density into the cryo-EM structure of the BAF complex (PDB 6LTJ), we were able to position the C-terminal helix of BAF47 on the acidic patch. By using our BCL7A–NCP structure (PDB 9QAJ), we positioned BCL7A α2 helix, but the rest of BCL7 density cannot be surely attributed although it seems to run along BAF47. As mentioned above CX-MS experiments (Supplementary Table S3 and Supplementary Fig. S15) support the idea that the two molecules interact and can both be retained on the nucleosome surface in proximity to the acidic patch.

BCL7A cancer-derived mutations located in the arginine anchor motif impair NCP binding

BCL7A is frequently mutated in cancer, especially in DLBCL [17, 28, 31, 22]. The mutations tend to accumulate at the N-terminal region (Supplementary Fig. S16A) and some of them represent cancer drivers (Supplementary Fig. S16B). Although R11 is the most frequently mutated residue in cancer reported to date, we noted that all three arginine anchor residues have been reported as cancer mutations. We generated recombinant BCL7A proteins comprising these mutations (R4A, R7A and R11S or R11A, each single mutant and the triple mutant) and tested their ability to bind the NCP by EMSA assays. It was observed that R11S mutation significantly reduced the affinity of BCL7A for the NCP (K_D 519 nM compared to 293 nM, Fig. 8); R4A and R7A similarly reduced the affinity for the NCP (R4A K_D is 517 nM and R7A K_D is 428 nM). Finally, upon mutation of all three arginines (R4A, R7A, R11A) the affinity



D

Sample	Rg (Å) Guinier	Rg (Å) P(r)	Dmax (Å) P(r)	MW (kDa)
BCL7A	48.08 ± 0.42	58.73 ± 2	261 ± 16	39
NCP	42.7 ± 0.04	42.52 ± 0.02	146 ± 5.5	217
NCP-BCL7A	44.92 ± 0.02	45.38 ± 0.02	184 ± 4.6	243

Figure 7. SAXS analysis of the BCL7A–NCP complex. (**A**) Resulting SAXS curves with Guinier plots and their respective gyration radii (Rg) in inset. (**B**) Pair distance distribution functions with respective maximum dimensions Dmax. (**C**) Normalized Kratky plots: BCL7A shows the characteristic plot of a flexible and mainly unstructured protein, while NCP and NCP:BCL7A complex show the plot of particles with compact core with a peak at $\sqrt{3}$. (**D**) Summary of biophysical parameters derived from the SAXS analysis.

for the NCP dropped dramatically and K_D could not be assessed reliably. All K_D reductions were significant with *P*-value lower than 0.01. Instead, none of the mutations affected the Hill coefficient in a manner that was statistically significant. We further confirmed these results by performing BLI binding experiments with nucleosome and BCL7A carrying single or triple arginine mutants as described above. The assays showed a noticeably reduced dissociation constant for the R4A mutant (Supplementary Fig. S17). Results with the other arginine anchor mutants were plagued by weak binding response and/or high noise yielding poor fit and preventing the calculation of a reliable K_D value (Supplementary Fig. S17). Collectively, the data show that the arginine-anchor is the driving force of the binding to the nucleosome. These results validate the NCP binding mechanism that we elucidated above using cryo-EM, NMR, and SAXS (Figs 3, 6, 7) by highlighting the importance of the BCL7A arginine anchor. Additionally, they indicate that BCL7A is involved in direct NCP recognition.

Implication of BCL7 proteins in chromatin remodeling

To examine if BCL7 proteins contribute to the chromatin remodeling activity of mSWI/SNF, we performed *in vitro* chromatin remodeling assays with purified intact or BCL7-mutated mSWI/SNF complex. We purified the complex from K-562 cells, which are wild-type for all the genes coding its subunits (from the COSMIC database). Purification was made possible by the installation of an affinity purification tag at the N-terminus of BAF47, using CRISPR/Cas9 knocking in at the endogenous locus. Mutations were introduced in the three BCL7 genes from these cells using CRISPR/Cas9-mediated creation of indels, which led to a cell clone, termed "TKO" where BCL7C expression is abolished, and BCL7A and BCL7B proteins are expressed as shorter proteins missing residues around the arginine anchor motif (c.f. the 'Material and methods' section and Supplementary Figs S3 and S4 for details). Remodeling activity of the WT and TKO purified complexes was assessed by restriction enzyme accessibility assay (Fig. 9A). It was observed that the mutant complex has lower activity compared to the parental cells (Fig. 9B and C, *P*-value < .01), indicating that BCL7 proteins contribute to the activity of mSWI/SNF.

Finally, to explore the implication of BCL7 proteins in mSWI/SNF-mediated chromatin remodeling in cells, we used the above-mentioned parental and TKO K-562 cell lines and performed ATAC-seq to measure DNA accessibility in a genome-wide fashion. The control, parental cell line car-

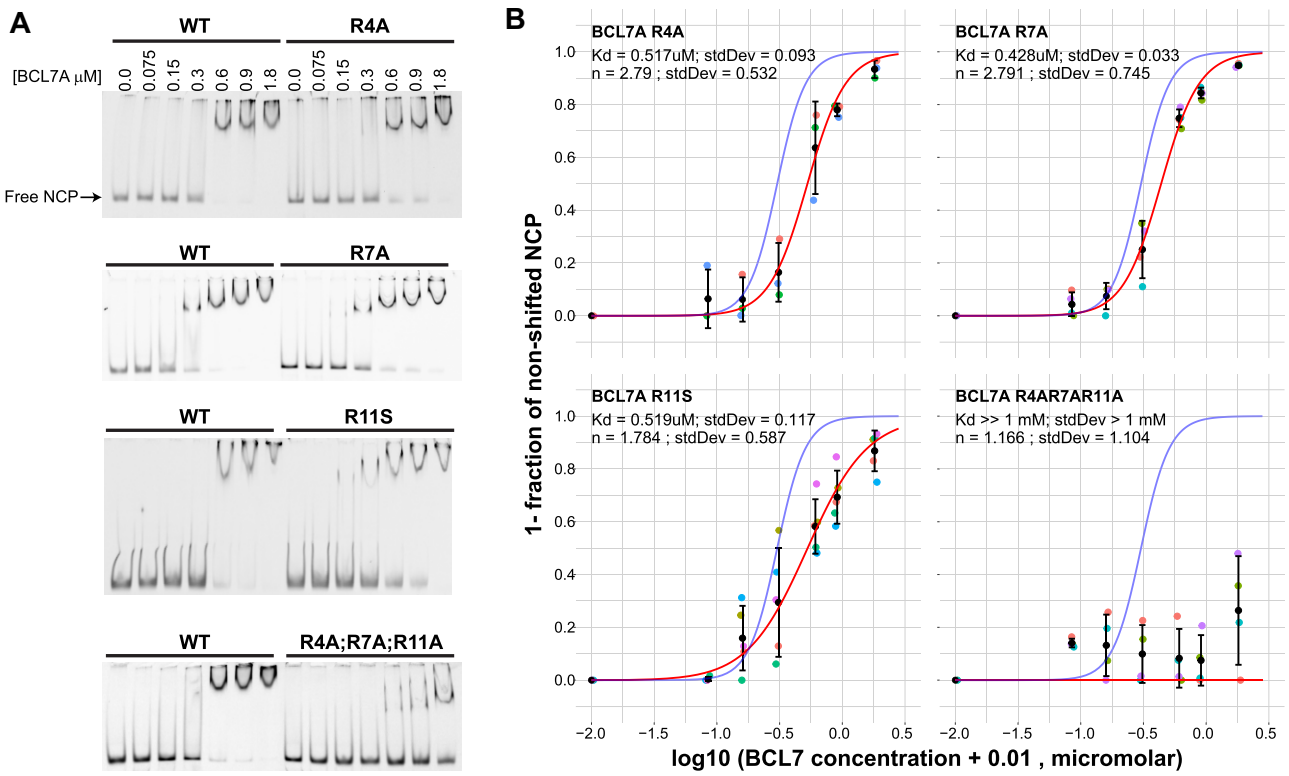


Figure 8. Nucleosome binding by BCL7A is impaired by mutations at the arginine anchor. (**A**) NCP binding experiments performed by EMSA for WT and various mutants of BCL7A. Representative replicates are shown. N ≥ 4. (**B**) Fitting of the experimental data to the Hill equation (red curves). The blue curves indicate the curve fit for the WT proteins (same as shown in Fig. 2). Dots of different color represent biological replicates of the experiment.

ried the N-terminal purification tag on the BAF47 protein but had the three BCL7 genes intact. ATAC-seq was performed on parental or TKO cells, treated with the BRM/BRG1 inhibitor BRM014 for 6 h or left untreated. The data were analyzed to identify regions of differential accessibility (DA) of DNA at BRG1 binding sites (previously identified by ChIP-seq [100, 101]) where DA is reduced by BRM014 treatment. We identified a sub-group of 3785 genomic loci (20%) where accessibility is lower in the TKO cell line and another group of 3365 loci (18%) where accessibility increased instead (Fig. 10A and B and Supplementary Fig. S18). Examples of loci with impaired DNA accessibility in the TKO, and loci unaffected, are shown in Fig. 10C and D. Examination of these loci showed that those depending on BCL7 proteins for maximum accessibility tend to have a distribution of positions with regards to genes that is comparable to the situation happening at unaffected sites, while the loci where impairing the BCL7 proteins leads to higher accessibility were comparatively more enriched in gene distal regions and correspondingly depleted in proximal promoter regions (Fig. 10E). This is also reflected by the type of transcription factor that co-occur at these sites (from ChIP-seq experiments done in the same cell line [88]). When the two differential DA groups are compared with one another, factors that typically tend to bind proximal promoter regions (e.g.TBP, NELF, DSIF subunit SUPT5H) are enriched in BCL7-dependent regions while AP-1, a factor often found at distal enhancers, is enriched in the regions where accessibility increases in the TKO (Fig. 10F and G). Together, these results indicate that intact BCL7 proteins are collectively needed for optimal mSWI/SNF activity at a subset of binding sites, with a small preference for sites that are promoter proximal.

Discussion

BCL7 proteins are IDPs for which only limited structural and functional information existed. We have provided evidence that BCL7A binds the NCP with high affinity and specificity, and we identified the N-terminal part of BCL7A as sufficient for this activity.

Molecular function of BCL7A

Although mSWI/SNF is an (NCP)-displacing multiprotein complex, until recently only the interactions of BAF47 and the ATPase BRG1 [28–30, 99] with the NCP have been described at the structural level. The interaction of PBAF with the NCP has been further clarified with the identification of additional subunits that make contact with the NCP for instance with the histone tails [29], but much remains to be discovered about the interactions between mSWI/SNF and its substrate. We report K_D values for BCL7 proteins in the range of 200–300 nM for their interaction with the NCP. This binding constant is comparable to what has been reported for the bromodomain of BRG1 and BRM (around 400–600 nM, by NMR) and stronger than the bromodomain-AT-hook regions of these two proteins (16–40 μM, by fluorescence anisotropy) [102]. Another study by isothermal titration calorimetry, for the AT-hook domain of BRG1 determined a K_D of 2–7 μM. BAF47 binds to DNA through its N-terminal winged-helix domain with a dissociation constant estimated by NMR to be in the high micromolar range [103]. It also binds the NCP acidic patch through its C-terminal alpha-helix, though a dissociation constant was not reported [99]. We measured by EMSA experiments BAF47 affinity of full length BAF47 for the NCP

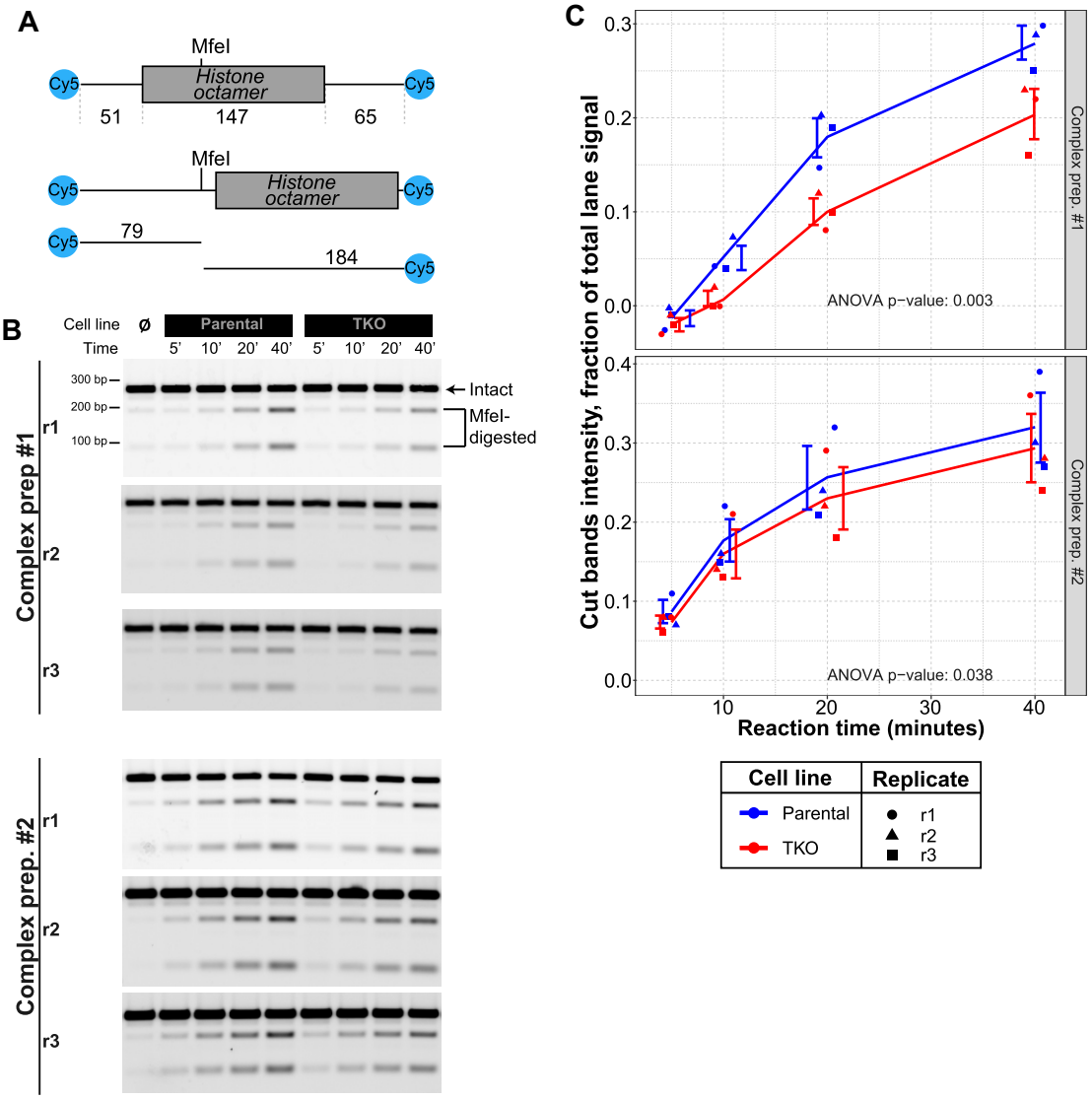


Figure 9. BCL7 proteins contribute *in vitro* to remodeling by mSWI/SNF. (**A**) Diagram of the NCP remodeling substrate. A 601 DNA sequence is flanked by the indicated lengths of exposed DNA and labeled at both extremities with Cy5 dye. Prior to remodeling, an MfeI restriction site is occluded by the histone octamer. Sliding or eviction of the octamer leads to exposure of the restriction site and allows cleavage by the enzyme. (**B**) Results of restriction enzyme accessibility assay performed with two replicate preparations of the substrate NCP and mSWI/SNF complexes (parental cell line or TKO for the BCL7 genes), each in technical triplicates. Identical amounts of complex are incubated for time periods ranging from 5 to 40 min. The two bands produced by MfeI digestion are indicated. (**C**) Graphs of the quantified data. A slight horizontal jitter was added to the individual points to prevent overlap. Error brackets indicate the standard error and are centered at the mean value. *P*-values are by repeated measures two-way ANOVA, for the factor "genotype".

and we calculated a K_D of 270 nM (Supplementary Fig. S19). Therefore, BCL7A and BAF47 seem to have similar affinities for their substrate.

Most of the biochemical work presented here was conducted with the BCL7A protein. We also presented EMSA results showing that BCL7B and BCL7C bind the NCP with high affinity, indicating a similar molecular function (Fig. 2D). We conclude from this that NCP binding is a property shared by all three members of the BCL7 family and we postulate that this involves the conserved N-terminal arginine anchors of the respective proteins. It has been proposed that several NCP- and DNA-binding activities are involved in the recruitment of the mSWI/SNF complex to chromatin [104, 105], and our work shows that BCL7 proteins are likely contributors to this process.

Unlike several of the other mSWI/SNF subunits mentioned above, the BCL7 proteins do not possess a discernible chromatin binding domain. However, all three family members share a highly related N-terminal sequence of about 50 amino acids that is followed by a less conserved, disordered tail. An EMSA performed with a truncated BCL7A (residues 1–100) revealed that this segment of the protein is sufficient for NCP binding. However, it does not rule out the possibility that parts of the divergent and mostly disordered C-terminal region might also contribute to this activity.

The NMR data reported in this study are consistent with binding of BCL7A N-terminal region to the acidic patch of the nucleosome but also provided insight into two additional points of contact toward the C-terminal end of the molecule and a weaker interaction involving BCL7A aa 57–80. This

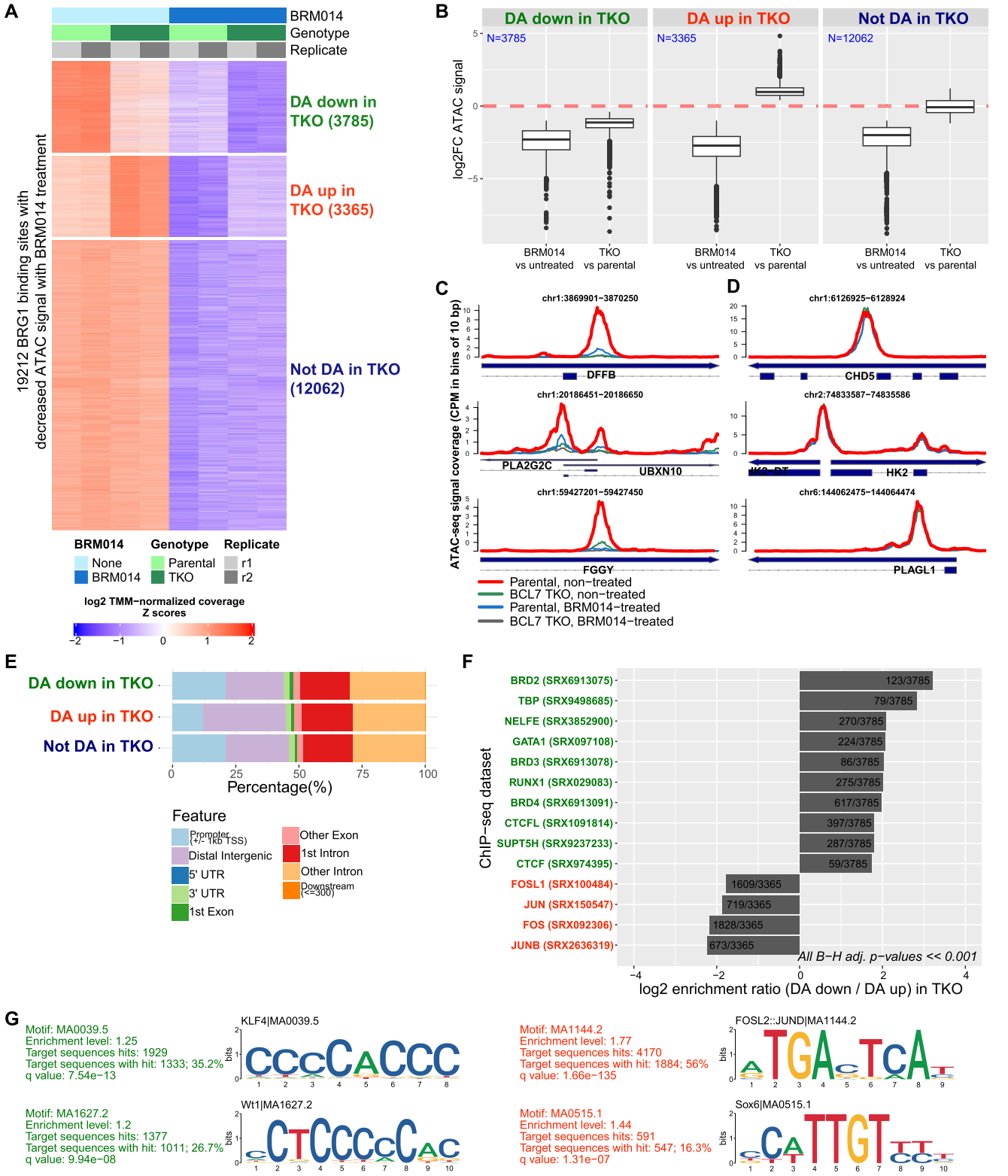


Figure 10. Identification of mSWI/SNF target loci dependent on BCL7 proteins for DNA accessibility. (**A**) Heatmap of ATAC-seq signal at regions bound by BRG1 and where BRM014 treatment leads to a significant decrease in DNA accessibility. Rows are clustered based on whether ATAC-seq signal also significantly changes in the BCL7 TKO cell line. (**B**) Boxplot showing $\log_2$ fold change in ATAC signal for the indicated contrasts, separated by clusters. (**C**) Example of ATAC-seq signal at loci from cluster "DA down in TKO". (**D**) Example of other loci not affected in any of the conditions. (**E**) Distribution of position and distances between DA regions in each of the three clusters and genes. (**F**) A selection of the ChIP-seq datasets from K-562 cells showing a significant enrichment in one of the two differential DA clusters when compared to the other. Rows labeled in green are for datasets that showed a positive $\log_2$ enrichment ratio, indicating that they are enriched in the regions where accessibility decreased in BCL7 TKO cells when compared to the regions that increased in accessibility in BCL7 TKO cells. Rows labeled in red show enrichment in the opposite direction. The fractions in each bar indicate the proportion of the regions in the cluster that overlap with the indicated ChIP-seq dataset peaks. All TFs shown have adjusted *P*-values lower than .001. (**G**) A selection of TF binding motifs enriched in the DA down or DA up clusters.

appears to be a transient interaction, and it is observed only at higher ratio (1:1, BCL7A:NCP; Fig. 6C). Thus, this flexible molecule may contact the nucleosome at multiple points. Elucidation of the contribution of the additional points of contact will require further studies.

A recent report showed that in neural progenitor cells, BRG1 binding to chromatin is reduced in the absence of BCL7A [15]. Our results showing lower NCP binding with arginine anchor mutants are consistent with that. Furthermore, we found that collectively, the BCL7 proteins are needed for remodeling by mSWI/SNF. Further experimentation will be required to determine if these proteins also contribute to targeting of the complex to specific genomic sites, help recruit partner proteins or carry some other function. We note that although the three human proteins diverge in their C-terminal sequences, each respective protein has a C-terminus that is better conserved throughout evolution. For example, although the C-termini of human BCL7A and BCL7C have only 18% amino acid identity past position 51, the C-termini of human and avian (*Gallus gallus*) BCL7A are 70% identical (Supplementary Fig. S5). Phylogenetic conservation of these protein regions is likely due to selective pressure, implying functional constraints that may guide the elucidation of the molecular function of these protein regions. The high conformational flexibility and structural dynamics predicted for the BCL7 C-terminus possibly allow molecular mechanisms that are unlikely for ordered proteins.

NCP binding by BCL7A through an arginine anchor

The cryo-EM structure of BCL7A in complex with the nucleosome reveals that BCL7A docks on the acidic patch of the nucleosome using an arginine anchor motif located in the N-terminal part of the molecule. BCL7A then makes a kink and continues with an α helix that contacts H2A C-terminal helix.

At least 50% of NCP-binding proteins bind it through the acidic patch via the arginine-anchor motif [97, 106–108]. BCL7A fits this scenario and constitutes an additional instance. The binding of chromatin enzymes to H2B C-terminal helix, which occupies an exposed position on the nucleosome surface, has been observed in several other proteins [97], therefore BCL7A is not unique into its mode of contacting H2B C-terminal helix.

The three BCL7A arginine side chains that compose the arginine-anchor motif match the clustered conformation of arginine anchors observed in other chromatin complexes [97]. As BCL7A is part of the mSWI/SNF complex, we have compared our cryo-EM structure of BCL7A arginine-anchor motif (aa 1–12) with the C-terminal α helix of BAF47 (PDB ID: 6LTJ) that has been shown to interact with the nucleosome acidic patch [24, 28–30, 99]. BCL7A and BAF47 arginine-anchor motifs show a very similar nucleosome binding mode; BCL7A arginine side chains responsible for contacting directly the acidic patch (R4, R7, and R11) overlay well with three equally spaced C-terminal arginines of BAF47 (R367, R370, R374) (Fig. 3D). The major difference is that while BCL7A nucleosome binding motif is located on a α helix and a loop and comprises three arginines, BAF47 arginine-anchor motif is located on a longer α helix and encompasses an additional atypical arginine (R366) and therefore accounts for a total of four arginines involved in the binding.

Implications for the architecture of mSWI/SNF

The disordered nature of BCL7A makes it challenging to clearly pinpoint its location within the mSWI/SNF complex. Previous studies have faced difficulties in accurately locating BCL7A in relation to the nucleosome, primarily due to low structure resolution [30]. Alternatively, some studies have opted to exclude BCL7A from their complex to enhance the resolution of the overall complex [28, 29]. Even in cases where BCL7A has been included, it remains challenging to visualize it in the final reconstruction of the complex [24]. Our study focused exclusively on the BCL7A subunit, rather than the entire complex, allowing us to precisely determine its coordinates. Specifically, we characterized the N-terminal arginine anchor of BCL7A and gathered information on additional nucleosome contact points within the unstructured C-terminal part. Comparison of the structures of our BCL7A-nucleosome complex with the nucleosome only [109] (PDB ID: 3LZ1) shows that BCL7A binding does not lead to a significant structural change in the nucleosome (Root Mean Square Deviation (RMSD) = 0.682 Å). This suggests that the role of this protein within mSWI/SNF has possibly more to do with recruitment/retention of the complex onto chromatin than with the remodeling activity *per se* (NCP displacement/disruption).

It is interesting to observe that both BAF47 [24, 99] and BCL7A (two components of the BAF and PBAF variant complexes) interact at the identical location on the nucleosome and according to our studies with similar affinities. Furthermore, BRG1 binds the opposite side of the nucleosome with respect to BAF47 but engages the nucleosome acidic patch in a very similar way using an arginine anchor motif composed of three arginines located in the Snf2 ATP coupling (SnAc) domain and conserved in the Snf-2 family proteins [28, 29, 110]. Thus, the BCL7A arginine anchor is similar to the arginine anchors of BAF47 and BRG1 and these components of the mSWI/SNF complex seem to share a common mechanism for nucleosome recognition. Our cryo-EM and CX-MS studies showed that BAF47 and BCL7 can converge on the same nucleosomal surface. Further studies await to be conducted to decipher the interplay of these subunits and to determine if a preferential or sequential binding occurs. A plausible hypothesis is that since nucleosome docking through the acidic patch is a critical recognition mechanism for chromatin proteins, it employs two subunits (BAF47 and BCL7A) to reach higher specificity and to prevent nonspecific interactions and enhance nucleosome binding by the complex. The loop connecting the two N-terminal helices of BCL7A seems to be very flexible and mobile and could be modulating the movement of the α1 helix as part of an ON/OFF molecular switch that helps repositioning the small N-terminal helix on the acidic patch. Interestingly, R11 is positioned in this loop and could be a key residue of a switch mechanism, perhaps regulated by PTMs or other cellular events.

The role of BCL7A within the ncBAF complex becomes especially important since this variant of the mSWI/SNF complex is missing BAF47. An analogous situation can also be imagined with BAF and PBAF complexes found in rhabdoid cancer cases characterized by biallelic deletion of the BAF47-encoding gene [111].

To further compare the roles of BCL7A and BAF47 on the function of mSWI/SNF, we reanalyzed the data of Otto *et al.*, where Perturb-seq was performed using several sgRNAs against each mSWI/SNF subunit [90]. In their analy-

sis, the authors reported that perturbations in BCL7 genes and in several other subunits clustered separately from perturbations in "cBAF functional core subunits", including BAF47 and BRG1, when mean transcriptomes were compared. Cells with BCL7 and many other subunit perturbations showed at best a modest correlation between their transcriptomes and those of BAF47 perturbations and instead correlated better with the negative control perturbation cells. We reanalyzed the data to examine the effects of BCL7 gene perturbations and to compare with those of BAF47. Single-cell RNA-seq expression profile of cells which received sgRNA against each subunit were compared to those of cells that received the nontargeting control sgRNAs and expression difference values were calculated. Mutations introduced in BAF47, BRG1, and the three BCL7 genes all yielded a general trend towards downregulation of genes, though the magnitude of the effect of BCL7 perturbations was markedly smaller (Supplementary Fig. S20A). This observation could be explained by the fact that while there is no paralog of BAF47 within mSWI/SNF, the perturbation of any BCL7 gene could in principle be compensated by the presence of the two remaining BCL7 paralogs. This hypothesis is supported by the more pronounced decrease in expression in the few cells that had a double BCL7 perturbation (Supplementary Fig. S20). The number of cells containing sgRNAs against all three BCL7 members in that dataset is too low to allow any meaningful exploration of this possibility. At the individual gene level, the effects of BAF47 and BCL7A perturbations were in general agreement as shown by the positive slope trendline on scatter plots, and the effect appeared more pronounced in cells with a double perturbation of BCL7A and BCL7B (Supplementary Fig. S20). The agreement between the effects of perturbations in BCL7 genes and BAF47 are nevertheless less pronounced than between BAF47 and BRG1. The results of this analysis suggest that across the genome, the gene sets regulated by BAF47 and by BCL7 genes overlap substantially and the two proteins probably work together at several genomic loci.

Implications for the oncogenic mechanism underlying BCL7A cancer-associated mutations

BCL7A is the target of driver mutations in DLBCL [22] and is considered to function as a tumor suppressor [17]. Truncations caused by splice site mutations lead to its exclusion from the complex [17]. The number of reported patients affected by DLBCL carrying a BCL7A mutation at R11 increased significantly in recent years and it is now considered a mutational hotspot [112–114]. However, the impact of these mutations on the molecular function of the protein, and ultimately on mSWI/SNF function and on the cancerous phenotype, was poorly characterized. We defined the impact of missense mutations at the arginine anchor of BCL7A. The reduction of nucleosome binding affinity when these key residues are mutated is consistent with BCL7A interaction with the acidic patch observed in our NMR data and cryo-EM structure. Overall, our biochemical characterization provides a molecular basis to start understanding the deleterious effects of some missense mutations in BCL7 proteins on binding to chromatin.

Role of BCL7 proteins in chromatin remodeling

We characterized that BCL7 proteins have a role in chromatin remodeling and are needed for an optimal mSWI/SNF activity by *in vitro* and genomic experiments. These data seem to be in agreement with previously reported data (https://doi.org/10.1101/2021.10.26.465931). It remains to be determined why only a fraction of BRG1-bound and BRG1-dependent sites also depend on the presence of intact BCL7 proteins. One possibility is that BCL7 proteins accomplish a function relevant to DA that is not required at all loci. Clarification of this will await further studies mapping the binding sites of BCL7 proteins and various forms mSWI/SNF in the same cell type.

Concluding statement

In conclusion, our integrated structural biology approach to characterize BCL7A–NCP interactions revealed for the first time that its N-terminal region comprised an arginine anchor motif that binds the acidic patch of the nucleosome. NMR studies provided complementary information on the dynamics of the binding. Biochemical assays provided information on the affinity of the BCL7A–NCP interaction and showed how cancer-associated mutations impacting the N-terminal arginine residues led to reduced interaction with the NCP. This was corroborated by the observation that BCL7 proteins are needed for efficient chromatin remodeling by mSWI/SNF. Both our results and those reported previously with BRG1 and BAF47 all converge to show interactions with the acidic patch. As all three proteins cannot interact simultaneously with the acidic patch, this raises the question of the contribution of each protein to NCP binding in the context of the complete BAF and PBAF complexes, but also in the ncBAF complex or in rhabdoid tumors where BAF47 is absent. Further studies will be required to determine whether there are alternative conformations of BAF and PBAF where either BAF47 or BCL7 proteins bind the acidic patch in addition to BRG1, and whether this has functional consequences concerning chromatin remodeling.

Acknowledgements

The authors wish to thank Irwin Davidson for support and helpful discussions, the staff at the Centre de Biologie Intégrative of the IGBMC structural biology platform for technical assistance, Catherine Birck for assistance with AUC experiments, Albert Weixlbaumer, Nils Marechal and Alexandre Durand for cryo-EM data processing advice, and Denis Fumagalli from the Mediaprep platform. We thank the EMBL Heidelberg Proteomics Core Facility for their support, especially Mandy Rettel and Frank Stein. A.L. would like to thank Paola Picotti at the ETH in Zürich for access to instrumentation and laboratory infrastructure for the cross-linking MS experiments. A.B. acknowledges computing support from the Digital Research Alliance of Canada (alliancecan.ca).

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This research was funded in part by the Fondation pour la Recherche Médicale (grant number AJE20171039021), an ATIP-Avenir award (contract # 188435), an award from La Ligue Contre le Cancer (JMG/SP - n° 01B.2021), startup funds from the IGBMC, ANR-23-CE11-0002-03, ANR-24-CE12-1416, ANR-the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INBS-0005, EquipEx + France cryo-EM (ANR-21-ESRE-0046), and the IR INFRANALYTICS FR2054, salary support by the Centre National de la Recherche Scientifique to E.B and C.S. and by the Commissariat à l'Energie Atomique to F.O. D.D. was supported by a 4th year fellowship from the Fondation pour la Recherche Médicale (FRM). F.M is supported by a 4th year fellowship from Fondation ARC, J.L. is supported by a scholarship from the École Doctorale de l'Université de Strasbourg and a 4th year fellowship from Fondation ARC. A.B. is supported by an operating grant from the Canadian Institutes of Health Research (PJT-183839).

Data availability

Cryo-EM coordinates of the BCL7-NCP complex have been deposited in the Protein Data Bank (PDB) with accession code 9QAJ, the map was deposited in the Electron Microscopy Data Bank (EMDB) with accession code EMD-52975 and raw data were deposited in the Electron Microscopy Public Image Archive (EMPIAR) with code EMPIAR-11579. The global map of the BCL7-BAF47-NCP complex was deposited in EMDB with accession code EMD-51375 and a local map was deposited with accession code EMD-51376. The cross-linking mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE [115] partner repository with the dataset identifier PXD054757. The ATAC-seq data were deposited to the NCBI Gene Expression Omnibus under accession number GSE285045.

The AlphaFold model of BCL7A was downloaded from https://alphafold.ebi.ac.uk/files/AF-Q4VC05-F1-model_v4.pdb.

AlphaFold3 was queried on the dedicated server: https://alphafoldserver.com/

The D2P2 server was queried at: https://d2p2.pro/

The Python code for MobiDB-lite was obtained at https://github.com/BioComputingUP/MobiDB-lite

localCIDER was accessed at https://github.com/Pappulab/localCIDER

The code for NARDINI was obtained at https://github.com/mshinn23/nardini

The PyMol software was obtained from https://pymol.org/2/

xQuest v2.1.5 was obtained from https://gitlab.ethz.ch/leitner_lab

UCSF Chimera software was obtained from https://www.cgl.ucsf.edu/chimera/download.html

Xlink Analyzer v1.1.4 was obtained at https://www.embl-hamburg.de/XlinkAnalyzer/XlinkAnalyzer.html

CryoSPARC was available at https://cryosparc.com/download

ATSAS software was available at https://www.embl-hamburg.de/biosaxs/software.html

The DichroWeb server was accessed at http://dichroweb.cryst.bbk.ac.uk/html/process.shtml

The ChIP-Atlas website was accessed at https://chip-atlas.org/enrichment_analysis

References

1. Luger K, Mäder AW, Richmond RK *et al.* Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 1997; **389**:251–60. https://doi.org/10.1038/38444
2. Narlikar GJ, Sundaramoorthy R, Owen-Hughes T Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* 2013; **154**:490–503. https://doi.org/10.1016/j.cell.2013.07.011
3. Kadoch C, Hargreaves DC, Hodges C *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat Genet* 2013; **45**:592–601.
4. Mcbride MJ, Pulice JL, Beird HC *et al.* The SS18-SSX fusion oncoprotein hijacks BAF complex targeting and function to drive synovial sarcoma. *Cancer Cell* 2018; **33**:1128–41.
5. Msaouel P, Malouf GG, Su X *et al.* Comprehensive molecular characterization identifies distinct genomic and immune hallmarks of renal medullary carcinoma. *Cancer Cell* 2020; **37**:720–34.
6. Centore RC, Sandoval GJ, Soares LMM *et al.* Mammalian SWI/SNF chromatin remodeling complexes: emerging mechanisms and therapeutic strategies. *Trends Genet* 2020; **36**:936–50.
7. Clapier CR, Iwasa J, Cairns BR *et al.* Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* 2017; **18**:407–22. https://doi.org/10.1038/nrm.2017.26
8. Mashtalir N, D'avino AR, Michel BC *et al.* Modular organization and assembly of SWI/SNF family chromatin remodeling complexes. *Cell* 2018; **175**:1272–88.
9. Middeljans E, Wan X, Jansen PW *et al.* SS18 Together with animal-specific factors defines human BAF-type SWI/SNF complexes. *PLoS One* 2012; **7**:e33834. https://doi.org/10.1371/journal.pone.0033834
10. Kadoch C, Hargreaves DC, Hodges C *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat Genet* 2013; **45**:592–601. https://doi.org/10.1038/ng.2628
11. Marcon E, Ni Z, Pu S *et al.* Human-chromatin-related protein interactions identify a demethylase complex required for chromosome segregation. *Cell Rep* 2014; **8**:297–310. https://doi.org/10.1016/j.celrep.2014.05.050
12. Huttlin EL, Ting L, Bruckner RJ *et al.* The BioPlex Network: a systematic exploration of the human interactome. *Cell* 2015; **162**:425–40. https://doi.org/10.1016/j.cell.2015.06.043
13. Wan C, Borgeson B, Phanse S *et al.* Panorama of ancient metazoan macromolecular complexes. *Nature* 2015; **525**:339–44. https://doi.org/10.1038/nature14877
14. Siguero-Álvarez M, Salguero-Jiménez A, Grego-Bessa J *et al.* A human hereditary cardiomyopathy shares a genetic substrate with bicuspid aortic valve. *Circulation* 2023; **147**:47–65. https://doi.org/10.1161/CIRCULATIONAHA.121.058767
15. Wischhof L, Lee H-M, Tutas J *et al.* BCL7A-containing SWI/SNF/BAF complexes modulate mitochondrial bioenergetics during neural progenitor differentiation-containing SWI/SNF/BAF complexes modulate mitochondrial bioenergetics during neural progenitor differentiation. *EMBO J* 2022;**41**:e110595. https://doi.org/10.15252/embj.2022110595
16. Huang C, Hao Q, Shi G *et al.* BCL7C suppresses ovarian cancer growth by inactivating mutant p53. *J Mol Cell Biol* 2021;**13**:141–50. https://doi.org/10.1093/jmcb/mjaa065
17. Baliñas-Gavira C, Rodríguez MI, Andrades A *et al.* Frequent mutations in the amino-terminal domain of BCL7A impair its tumor suppressor role in DLBCL. *Leukemia* 2020;**34**:2722–35. https://doi.org/10.1038/s41375-020-0919-5

18. Wischhof L, Maida S, Piazzesi A *et al.* The SWI/SNF subunit Bcl7a contributes to motor coordination and Purkinje cell function. *Sci Rep* 2017; **7**:17055. https://doi.org/10.1038/s41598-017-17284-3
19. Uehara T, Kage-Nakadai E, Yoshina S *et al.* The tumor suppressor BCL7B functions in the wnt signaling pathway. *PLoS Genet* 2015; **11**:e1004921. https://doi.org/10.1371/journal.pgen.1004921
20. Izhar L, Adamson B, Ciccia A *et al.* A systematic analysis of factors localized to damaged chromatin reveals PARP-dependent recruitment of transcription factors. *Cell Rep* 2015; **11**:1486–500. https://doi.org/10.1016/j.celrep.2015.04.053
21. Forbes SA, Beare D, Gunasekaran P *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 2015; **43**:D805–11. https://doi.org/10.1093/nar/gku1075
22. Reddy A, Zhang J, Davis NS *et al.* Genetic and functional drivers of diffuse large B cell lymphoma. *Cell* 2017; **171**:481–94. https://doi.org/10.1016/j.cell.2017.09.027
23. Patiño-Mercau JR, Baliñas-Gavira C, Andrades A *et al.* BCL7A is silenced by hypermethylation to promote acute myeloid leukemia. *Biomark Res* 2023; **11**:32. https://doi.org/10.1186/s40364-023-00472-x
24. Wang L, Yu J, Yu Z *et al.* Structure of nucleosome-bound human PBAF complex. *Nat Commun* 2022; **13**:7644. https://doi.org/10.1038/s41467-022-34859-5
25. Ye Y, Wu H, Chen K *et al.* Structure of the RSC complex bound to the nucleosome. *Science* 2019; **366**:838–43. https://doi.org/10.1126/science.aay0033
26. Wang C, Guo Z, Zhan X *et al.* Structure of the yeast Swi/Snf complex in a nucleosome free state. *Nat Commun* 2020; **11**:3398. https://doi.org/10.1038/s41467-020-17229-x
27. Wagner FR, Dienemann C, Wang H *et al.* Structure of SWI/SNF chromatin remodeller RSC bound to a nucleosome. *Nature* 2020; **579**:448–51. https://doi.org/10.1038/s41586-020-2088-0
28. He S, Wu Z, Tian Y *et al.* Structure of nucleosome-bound human BAF complex. *Science* 2020; **367**:875–81. https://doi.org/10.1126/science.aaz9761
29. Yuan J, Chen K, Zhang W *et al.* Structure of human chromatin-remodelling PBAF complex bound to a nucleosome. *Nature* 2022; **605**:166–71. https://doi.org/10.1038/s41586-022-04658-5
30. Mashtalir N, Suzuki H, Farrell DP *et al.* A structural model of the endogenous human BAF complex informs disease mechanisms. *Cell* 2020; **183**:802–17. https://doi.org/10.1016/j.cell.2020.09.051
31. Phelan ML, Sif S, Narlikar GJ *et al.* Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. *Mol Cell* 1999; **3**:247–53. https://doi.org/10.1016/S1097-2765(00)80315-9
32. Lowary PT, Widom J New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* 1998; **276**:19–42. https://doi.org/10.1006/jmbi.1997.1494
33. Dyer PN, Edayathumangalam RS, White CL *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Meth Enzymol*2003;**375**:23–44, https://doi.org/10.1016/S0076-6879(03)75002-2
34. Shim Y, Duan M-R, Chen X *et al.* Polycistronic coexpression and nondenaturing purification of histone octamers. *Anal Biochem* 2012; **427**:190–2. https://doi.org/10.1016/j.ab.2012.05.006
35. Jarmoskaite I, AlSadhan I, Vaidyanathan PP *et al.* How to measure and evaluate binding affinities. *eLife* 2020; **9**:e57264. https://doi.org/10.7554/eLife.57264
36. Hill AV The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. *J Physiol* 1910;**40**:iv–vii.
37. Pagano JM, Clingman CC, Ryder SP Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes. *RNA* 2011; **17**:14–20. https://doi.org/10.1261/rna.2428111
38. Brautigam CA Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Methods Enzymol* 2015; **562**:109–33. https://doi.org/10.1016/bs.mie.2015.05.001
39. Mlynarczyk C, Teater M, Pae J *et al.* BTG1 mutation yields supercompetitive B cells primed for malignant transformation. *Science* 2023; **379**:eabj7412. https://doi.org/10.1126/science.abj7412
40. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* 2020; **578**:82–93. https://doi.org/10.1038/s41586-020-1969-6
41. Favier A, Brutscher B NMRlib: user-friendly pulse sequence tools for Bruker NMR spectrometers. *J Biomol NMR* 2019; **73**:199–211. https://doi.org/10.1007/s10858-019-00249-1
42. Wishart DS *et al.* 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 1995; **6**:135–40. https://doi.org/10.1007/BF00211777
43. Tamiola K, Acar B, Mulder FAA Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc* 2010; **132**:18000–3. https://doi.org/10.1021/ja105656t
44. Mastronarde DN Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* 2005; **152**:36–51. https://doi.org/10.1016/j.jsb.2005.07.007
45. Zheng SQ, Palovcak E, Armache J-P *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* 2017; **14**:331–2. https://doi.org/10.1038/nmeth.4193
46. Rohou A, Grigorieff N CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J Struct Biol* 2015; **192**:216–21. https://doi.org/10.1016/j.jsb.2015.08.008
47. Punjani A, Rubinstein JL, Fleet DJ *et al.* cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 2017; **14**:290–6. https://doi.org/10.1038/nmeth.4169
48. Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021; **596**:583–9. https://doi.org/10.1038/s41586-021-03819-2
49. Pettersen EF, Goddard TD, Huang CC *et al.* UCSF chimera? A visualization system for exploratory research and analysis. *J Comput Chem* 2004; **25**:1605–12. https://doi.org/10.1002/jcc.20084
50. Liebschner D, Afonine PV, Baker ML *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr D Struct Biol* 2019; **75**:861–77. https://doi.org/10.1107/S2059798319011471
51. Adams PD, Afonine PV, Bunkóczi G *et al. PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010; **66**:213–21. https://doi.org/10.1107/S0907444909052925
52. Emsley P, Lohkamp B, Scott WG *et al.* Features and development of *Coot*. *Acta Crystallogr D Biol Crystallogr* 2010; **66**:486–501. https://doi.org/10.1107/S0907444910007493
53. Leitner A *et al.* Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol Cell Proteomics* 2012; **11**:M111.014126.
54. Chambers MC, Maclean B, Burke R *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* 2012; **30**:918–20. https://doi.org/10.1038/nbt.2377
55. Walzthoeni T, Claassen M, Leitner A *et al.* False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat Methods* 2012; **9**:901–3. https://doi.org/10.1038/nmeth.2103
56. Kosinski J, Von Appen A, Ori A *et al.* Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J Struct Biol* 2015; **189**:177–83. https://doi.org/10.1016/j.jsb.2015.01.014
57. Thureau A, Roblin P, Pérez J BioSAXS on the SWING beamline at synchrotron SOLEIL. *J Appl Crystallogr* 2021; **54**:1698–710. https://doi.org/10.1107/S1600576721008736

58. David G, Pérez J Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *J Appl Crystallogr* 2009; **42**:892–900. https://doi.org/10.1107/S0021889809029288
59. Hopkins JB, Gillilan RE, Skou S *BioXTAS RAW*: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J Appl Crystallogr* 2017; **50**:1545–53. https://doi.org/10.1107/S1600576717011438
60. Manalastas-Cantos K, Konarev PV, Hajizadeh NR *et al. ATSAS 3.0*: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Crystallogr* 2021; **54**:343–55. https://doi.org/10.1107/S1600576720013412
61. Meisburger SP, Taylor AB, Khan CA *et al.* Domain movements upon activation of phenylalanine hydroxylase characterized by crystallography and chromatography-coupled small-angle X-ray scattering. *J Am Chem Soc* 2016; **138**:6506–16. https://doi.org/10.1021/jacs.6b01563
62. Meisburger SP, Xu D, Ando N *REGALS*: a general method to deconvolve X-ray scattering data from evolving mixtures. *IUCrJ* 2021; **8**:225–37. https://doi.org/10.1107/S2052252521000555
63. Hansen S Bayesian estimation of hyperparameters for indirect fourier transformation in small-angle scattering. *J Appl Crystallogr* 2000; **33**:1415–21. https://doi.org/10.1107/S0021889800012930
64. Rambo RP, Tainer JA Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 2013; **496**:477–81. https://doi.org/10.1038/nature12070
65. Compton LA, Johnson WC Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal Biochem* 1986; **155**:155–67. https://doi.org/10.1016/0003-2697(86)90241-1
66. Provencher SW, Gloeckner J Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 1981; **20**:33–7. https://doi.org/10.1021/bi00504a006
67. Sreerama N, Woody RW Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* 2000; **287**:252–60. https://doi.org/10.1006/abio.2000.4880
68. Miles AJ, Ramalli SG, Wallace BA DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data, a website for calculating protein secondary structure from circular dichroism spectroscopic data. *Protein Sci* 2022; **31**:37–46. https://doi.org/10.1002/pro.4153
69. Abramson J, Adler J, Dunger J *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024; **630**:493–500. https://doi.org/10.1038/s41586-024-07487-w
70. Oates ME, Romero P, Ishida T *et al.* D^2P^2: database of disordered protein predictions. *Nucleic Acids Res* 2013; **41**:D508–16. https://doi.org/10.1093/nar/gks1226
71. Necci M, Piovesan D, Dosztányi Z *et al.* MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics* 2017; **33**:1402–4. https://doi.org/10.1093/bioinformatics/btx015
72. Cohan MC, Shinn MK, Lalmansingh JM *et al.* Uncovering non-random binary patterns within sequences of intrinsically disordered proteins. *J Mol Biol* 2022; **434**:167373. https://doi.org/10.1016/j.jmb.2021.167373
73. Patil A, Strom AR, Paulo JA *et al.* A disordered region controls cBAF activity via condensation and partner recruitment. *Cell* 2023; **186**:4936–55. https://doi.org/10.1016/j.cell.2023.08.032
74. Gu Z Complex heatmap visualization. *Imeta* 2022; **1**:e43. https://doi.org/10.1002/imt2.43
75. Cong L, Ran FA, Cox D *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013; **339**:819–23. https://doi.org/10.1126/science.1231143
76. Mak AB, Ni Z, Hewel JA *et al.* A lentiviral functional proteomics approach identifies chromatin remodeling complexes important for the induction of pluripotency. *Mol Cell Proteomics* 2010; **9**:811–23.
77. Dalvai M, Loehr J, Jacquet K *et al.* A scalable genome-editing-based approach for mapping multiprotein complexes in human cells. *Cell Rep* 2015; **13**:621–33. https://doi.org/10.1016/j.celrep.2015.09.009
78. Sif S, Stukenberg PT, Kirschner MW *et al.* Mitotic inactivation of a human SWI/SNF chromatin remodeling complex. *Genes Dev* 1998; **12**:2842–51.
79. Vasimuddin M, Misra S, Li H *et al.* Efficient architecture-aware acceleration of BWA-MEM for Multicore systems. In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. Rio de Janeiro, Brazil: IEEE, 2019; 314–24. https://doi.org/10.1109/IPDPS.2019.00041
80. Robinson JT, Thorvaldsdóttir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011; **29**:24–6. https://doi.org/10.1038/nbt.1754
81. Papillon JPN, Nakajima K, Adair CD *et al.* Discovery of orally active inhibitors of brahma homolog (BRM)/SMARCA2 ATPase activity for the treatment of brahma related gene 1 (BRG1)/SMARCA4-mutant cancers. *J Med Chem* 2018; **61**:10155–72. https://doi.org/10.1021/acs.jmedchem.8b01318
82. Iurlaro M, Stadler MB, Masoni F *et al.* Mammalian SWI/SNF continuously restores local accessibility to chromatin. *Nat Genet* 2021; **53**:279–87. https://doi.org/10.1038/s41588-020-00768-w
83. Madden K, El Hamra R, Berton S *et al.* Mycobacterium tuberculosis infection triggers epigenetic changes that are enriched in a type I IFN signature. *Microlife* 2023; **4**:uqad006. https://doi.org/10.1093/femsml/uqad006
84. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2022; **50**:D165–73. https://doi.org/10.1093/nar/gkab1113
85. Gu Z, Eils R, Schlesner M Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016; **32**:2847–9. https://doi.org/10.1093/bioinformatics/btw313
86. Hahne F, Ivanek R Visualizing genomic data using Gviz and bioconductor. *Methods Mol Biol* 2016; **1418**:335–51. https://doi.org/10.1007/978-1-4939-3578-9_16
87. Yu G, Wang L-G, He Q-Y ChIPseeker: an R/bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 2015; **31**:2382–3. https://doi.org/10.1093/bioinformatics/btv145
88. Zou Z, Ohta T, Miura F *et al.* ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and bisulfite-seq data. *Nucleic Acids Res* 2022; **50**:W175–82. https://doi.org/10.1093/nar/gkac199
89. Ramírez F, Dündar F, Diehl S *et al.* deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014; **42**:W187–91. https://doi.org/10.1093/nar/gku365
90. Otto JE, Ursu O, Wu AP *et al.* Structural and functional properties of mSWI/SNF chromatin remodeling complexes revealed through single-cell perturbation screens. *Mol Cell* 2023; **83**:1350–67. https://doi.org/10.1016/j.molcel.2023.03.013
91. Hao Y, Stuart T, Kowalski MH *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* 2024; **42**:293–304. https://doi.org/10.1038/s41587-023-01767-y
92. Holehouse AS, Das RK, Ahad JN *et al.* CIDER: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. *Biophys J* 2017; **112**:16–21. https://doi.org/10.1016/j.bpj.2016.11.3200
93. Das RK, Pappu RV Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc Natl Acad Sci USA* 2013; **110**:13392–7. https://doi.org/10.1073/pnas.1304749110

94. Behrouzi R, Lu C, Currie MA *et al.* Heterochromatin assembly by interrupted Sir3 bridges across neighboring nucleosomes. *eLife* 2016; **5**:e17556. https://doi.org/10.7554/eLife.17556
95. Van Holde KE. Chromatin. Series in molecular biology. Springer-Verlag, New York. 1988. 530 pp. $98.00. *J Mol Recognit* 1989;**2**:i–i. https://doi.org/10.1002/jmr.300020308
96. Yang C, van der Woerd MJ, Muthurajan UM *et al.* Biophysical analysis and small-angle X-ray scattering-derived structures of MeCP2-nucleosome complexes. *Nucleic Acids Res* 2011; **39**:4122–35. https://doi.org/10.1093/nar/gkr005
97. McGinty RK, Tan S Principles of nucleosome recognition by chromatin factors and enzymes. *Curr Opin Struct Biol* 2021; **71**:16–26. https://doi.org/10.1016/j.sbi.2021.05.006
98. Receveur-Brechot V, Durand D How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr Protein Pept Sci* 2012; **13**:55–75.
99. Valencia AM, Collings CK, Dao HT *et al.* Recurrent SMARCB1 mutations reveal a nucleosome acidic patch interaction site that potentiates mSWI/SNF complex chromatin remodeling. *Cell* 2019; **179**:1342–56. https://doi.org/10.1016/j.cell.2019.10.044
100. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**:57–74. https://doi.org/10.1038/nature11247
101. Luo Y, Hitz BC, Gabdank I *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res* 2020; **48**:D882–9. https://doi.org/10.1093/nar/gkz1062
102. Morrison EA, Sanchez JC, Ronan JL *et al.* DNA binding drives the association of BRG1/hBRM bromodomains with nucleosomes. *Nat Commun* 2017; **8**:16080. https://doi.org/10.1038/ncomms16080
103. Allen MD, Freund SMV, Zinzalla G *et al.* The SWI/SNF subunit INI1 contains an N-terminal winged helix DNA binding domain that is a target for mutations in schwannomatosis. *Structure* 2015; **23**:1344–9. https://doi.org/10.1016/j.str.2015.04.021
104. Wu JI, Lessard J, Crabtree GR Understanding the words of chromatin regulation. *Cell* 2009; **136**:200–6. https://doi.org/10.1016/j.cell.2009.01.009
105. Ho PJ, Lloyd SM, Bao X Unwinding chromatin at the right places: how BAF is targeted to specific genomic locations during development. *Development* 2019; **146**:dev178780. https://doi.org/10.1242/dev.178780
106. McGinty RK, Henrici RC, Tan S Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature* 2014; **514**:591–6. https://doi.org/10.1038/nature13890
107. McGinty RK, Tan S Recognition of the nucleosome by chromatin factors and enzymes. *Curr Opin Struct Biol* 2016; **37**:54–61. https://doi.org/10.1016/j.sbi.2015.11.014
108. Lagadec F, Parissi V, Lesbats P Targeting the nucleosome acidic patch by viral proteins: two birds with one stone? *mBio* 2022; **13**:e01733–21. https://doi.org/10.1128/mbio.01733-21
109. Vasudevan D, Chua EYD, Davey CA. Crystal structures of nucleosome core particles containing the '601' Strong positioning sequence. *J Mol Biol* 2010; **403**:1–10. https://doi.org/10.1016/j.jmb.2010.08.039
110. Sen P, Ghosh S, Pugh BF *et al.* A new, highly conserved domain in Swi2/Snf2 is required for SWI/SNF remodeling. *Nucleic Acids Res* 2011; **39**:9155–66. https://doi.org/10.1093/nar/gkr622
111. Sévenet N, Sheridan E, Amram D *et al.* Constitutional mutations of the hSNF5/INI1 gene predispose to a variety of cancers. *Am Hum Genet* 1999; **65**:1342–8. https://doi.org/10.1086/302639
112. Lacy SE, Barrans SL, Beer PA *et al.* Targeted sequencing in DLBCL, molecular subtypes, and outcomes: a Haematological Malignancy Research Network report. *Blood* 2020; **135**:1759–71. https://doi.org/10.1182/blood.2019003535
113. Gunawardana J, Chan FC, Telenius A *et al.* Recurrent somatic mutations of PTPN1 in primary mediastinal B cell lymphoma and Hodgkin lymphoma. *Nat Genet* 2014; **46**:329–35. https://doi.org/10.1038/ng.2900
114. Pillonel V, Juskevicius D, Ng CKY *et al.* High-throughput sequencing of nodal marginal zone lymphomas identifies recurrent BRAF mutations. *Leukemia* 2018; **32**:2412–26. https://doi.org/10.1038/s41375-018-0082-4
115. Perez-Riverol Y, Bai J, Bandla C *et al.* The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 2022; **50**:D543–52. https://doi.org/10.1093/nar/gkab1038

Received: November 2, 2023. **Revised:** March 21, 2025. **Editorial Decision:** March 21, 2025. **Accepted:** March 25, 2025