Survivin prevents the Polycomb Repressor Complex 2 from methylating Histone 3 lysine 27

Maja Jensen[a], María-José García Bonete[b], Atsarina Larasati Anindya[a], Karin Andersson[c], Malin C. Erlandsson[c], Nina Oparina[c], Ulrika Brath[d], Venkataragavan Chandrasekaran[c], Maria Bokarewa[c,e] and Gergely Katona[a,*]

[a]Department of Chemistry and Molecular Biology, Faculty of Science, University of Gothenburg, Sweden.

[b]Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, University of Gothenburg, Gothenburg SE-405 30, Sweden.

[c]Department of Rheumatology and Inflammation Research, Institute of Medicine, University of Gothenburg, Box 480, 40530 Gothenburg, Sweden.

[d]Department of Chemistry and Molecular Biology and the Swedish NMR Centre, University of Gothenburg, SE-412 96 Gothenburg, Sweden.

[e]Rheumatology Clinic, Sahlgrenska University Hospital, Gröna stråket 16, 41346 Gothenburg, Sweden

[*]Corresponding author Gergely Katona, Department of Chemistry and Molecular Biology, Faculty of Science, University of Gothenburg, Box 462, 40530 Gothenburg, Sweden, phone +46-709-123309, E-mail: gergely.katona@gu.se

Abstract

Survivin is a small protein that belongs to the inhibitor of apoptosis protein family and participates in cell division and apoptosis. It was actively studied in human cancers, inflammatory diseases and in autoimmune diseases. Here, we reveal that survivin takes part in epigenetic gene silencing by interaction with the polycomb repressive complex 2 (PRC2). PRC2 silences gene expression through tri-methylation of lysine 27 on histone 3 (H3K27). We detected differential expression of PRC2 core subunits in CD4+T cells with different survivin expression. ChIP-seq experiments indicated that survivin binds chromatin that overlap with the regions occupied by PRC2. ChIP-seq of H3K27 in CD4+T cells indicate that inhibition of survivin leads to a substantial increase in H3K27 tri-methylation by PRC2 in contrast to other histone modifications, which lends support to that survivin prevents PRC2 from functioning. Survivin binds peptides derived from PRC2 subunits Ezh2, EED, Suz12 and Jarid2 in a peptide microarray that cover intersubunit interfaces, catalytic residues, and present binding sites for substrates, DNA, and regulatory proteins. Amino acid composition of the peptides has substantial predictive power for survivin interaction in the peptide microarray as determined by multilayer perceptron classification analysis. NMR experiments with ^{15}N labelled survivin indicate that peptide colocalization does not entirely depend on binding mediated by short range interactions. These results indicate that survivin interacts with PRC2, preventing the methylation of H3K27 and specific gene silencing.

Introduction

For the human body to develop and maintain cell homeostasis a tight regulation of cellular processes, such as apoptosis and cell division is necessary. Survivin is one of the proteins that play an important role in these processes. With a length of 142 amino acids and a molecular weight of 16.5 kDa, survivin is the smallest member of the IAP family. In contrast to the other members of the IAP family, survivin has a single baculovirus-inhibitor of apoptosis repeat (BIR) which is connected to the α-helical C-terminus. [1, 2] The expression level and location of survivin in the cell is highly dependent on the phase of the cell cycle. During cell division, survivin takes part in the chromosomal passenger complex (CPC) together with Aurora B kinase, INCENP and borealin.[3, 4] When not part of the CPC, during interphase, survivin is in the cytoplasm and prevents apoptosis. During apoptotic signaling Smac/DIABLO is released from the mitochondria and prevents survivin from inhibiting apoptosis. A relatively unknown function of survivin is its role as a transcription factor or cofactor in gene regulation. A more systematic search of survivin binding sites in the chromatin is necessary to strengthen this emerging view.[5] In order to map potential association with chromatin we performed deep sequencing of DNA regions bound by survivin in CD4+ T-cells.

We found that the precipitated regions were strongly associated with chromatin locations of subunits associated with PRC2. PRC2 is a histone methyltransferase whose main function is to methylate lysine 27 on histone 3 (H3K27), thus regulating gene expression by epigenetic modification.[6] When H3K27 is trimethylated the genes are maintained in a repressed state, preserving the identity of the cell.[7] PRC2 is intensively studied because of its role in various types of cancer and its ability to restore pluripotency.[8-10] The structure of its core subunits is well characterized. The core consists of four subunits; enhancer of zeste homologue 2 (Ezh2), embryonic ectoderm development (EED), suppressor of zeste 12 (Suz12) and retinoblastoma binding protein 46/48 (RBAP46/RBAP48). Ezh2 contains the enzymatic component of the PRC2, this is where the trimethylation of H3K27 is taking place. The seven WD40-repeat containing protein EED is the smallest component of the complex. EED has an aromatic cage, which consists of four aromatic side chains, Phe-97, Tyr-148, Trp-364 and Tyr-365. This motif binds the already triple methylated H3K27, stimulating activity of the complex and autocatalysing the methylation of other histone subunits by Ezh2.[11] Suz12 encompasses the other subunits contributing to the stability of the complex and keeping them together.[12] Jumonji and AT-rich interaction containing domain 2 (Jarid2) and adipocyte enhancer-binding protein 2 (Aebp2) have been reported as important cofactors of PRC2. Jarid2 is mimicking a methylated H3 tail, which triggers PRC2 activity. Aebp2 on the other hand is mimicking an unmodified H3 tail.[12]

Based on the in vivo results, we were searching for potential direct interactions between survivin and short peptide segments derived from the subunits of PRC2. We expected that these may be mediated by relatively short-range bonds (hydrogen bonds, ion bridges, van der Waals forces, hydrophobic shape complementarity). The evidence presented here leads to a contradiction with this view and suggests a more intriguing model. Thus, we rationalized the terminology to make the discussion efficient. We will use the expressions “direct interaction” or just “interaction” when we assume that two partners are sufficient for an association, but we do not want to specify the physical mechanism. We will use the word “binding” for describing an association resulting from a random, initial collision between partners and establishing a pattern of short-range physical interactions or “bonds”. We reserve the word “tethering” for specific and predictable interactions resulting in colocalization of partners for which there is no evidence of “binding”, or the interaction “binding” is not plausible due to the great variety of potentially incompatible short range physical interactions. We assume that at least one of the partners must be localized and the partners do not necessarily obey the law of mass action. We will use the word “colocalization” when we simply mean that two or more components occupy a limited shared volume with or without direct interaction.

Specifically, we studied the role of survivin as a transcription factor in association with chromatin and identified its interaction with PRC2. We studied the direct interaction of PRC2 with survivin in peptide microarrays, biolayer interferometry (BLI), and NMR spectroscopy. Machine learning analysis indicates that the tethering of survivin to PRC2 derived peptides is not primarily governed by the sequence of the peptides, but their amino acid/atomic composition. NMR titration experiments revealed that this predictable tethering association does not likely involve binding on the surface of survivin. We argue that to explain the location and activity of the PRC2 complex together with survivin, tethering type of interactions are more important than binding type of interactions.

Results and Discussion

It is well known that survivin is binding nucleosomes through phosphorylated Thr-3 of histone H3. Although PRC2 also targets the N-terminal tail of H3, the two actors may not be present at the same time. We explored if H3K27 histone modifications involved in epigenetic regulation colocalize with survivin using ChiP-seq with antibodies against survivin, H3K27ac, H3K27me3 and H3K4me3.

In this analysis survivin ChIP-seq peak locations (Figure 1D) overlap substantially with Ezh2, EED, Suz12 and Jarid2 ChIP-seq peak locations supporting a potential concerted action. The highest significance is obtained when colocalization was analyzed in the 1k base interval in contrast to direct overlap or wider 10k base window. The most significant association of survivin peaks is with the catalytic Ezh2 among the compared subunits. This indicates that survivin may not directly compete for the DNA region corresponding to the modified nucleosomes, but it is part of a larger complex.

These results suggest that survivin may take part in H3K27 methylation pathways through PRC2, which is important for epigenetic silencing of sensitive genes. The experiments were repeated with and without sepantronium bromide (YM155) present, which is a specific inhibitor of survivin function.[13] Figures 1A, 1B and 1C show that colocalized survivin peaks and all histone modifications are associated with high peak score and substantially overlap. The combination of histone H3 modifications is associated with the highest score of survivin peaks. It was also reasonable to assume that open regions of the chromatin are more accessible to chromatin modifying enzymes and survivin access to histone tails is also less restricted. However, we detected an important difference in histone post-translation modification in Figure 1E and 1F. The largest increase in tag change upon survivin inhibition occurs in relation to H3K27me3 modification, whereas PRC2 independent modifications such as H3K27 acetylation and H3K4 trimethylation are less affected. For H3K27ac and H3K4me3, we could not see any significant change in the presence of survivin. Contrarily, we observed a significant difference in the amount of H3K27me3 upon the inhibition of survivin. This result indicates that survivin interferes with PRC2 and high amounts of survivin prevents PRC2 from methylating H3K27.

Figure 1G shows a strong association of PRC2 subunit expression with the expression level of survivin. This indicates a functional link between the two systems and presents indirect evidence of concomitant action of survivin and PRC2. To shed light on how survivin interferes with the function of PRC2 complex, we designed a peptide microarray of overlapping peptides derived from Ezh2, EED, Suz12 and Jarid2. These subunits represent the core of PRC2 and the interaction of survivin with the PRC2 molecular scaffold provides indication about the potentially disruptive role of survivin on the function of PRC2. In Figure 2, we compare the fluorescence signal arising from labeled survivin as a function of position along the polypeptide chain of PRC2 subunits. The raw fluorescence images are displayed in Figure S1. Figure 3 illustrates the microarray fluorescence intensities mapped on the three-dimensional structure of the PRC2 core complex. The entire data set is going to be deposited in the Zenodo data base upon acceptance of the manuscript.

Strongest interaction with Ezh2 derived peptides was observed in one continuous region (172-191). The region is located in the SANT1 domain of the enzyme, which is a general binding motif for

histone tails and promotes chromatin association.[14] A strong continuous signal was also observed at peptides in the region 707-736, which included the catalytic residue Tyr-726. Tyr-726 was initially proposed to be the general base for deprotonation of the substrate lysine, but it is likely that the solvent performs this catalytic role.[15, 16] Nevertheless, this is an important part of the active site and when the same region is missing or mutated (Tyr-736, Glu-740) in humans, it results in a specific phenotype described as Weaver's syndrome, a type of overgrowth disease in the affected patients (Tyr-736, Glu-740).[17] Lower level of fluorescence was also observed in peptide 627-656 which contains Tyr-641, which is also part of the active site region that brings together the SAM cofactor methyl donor and the H3K27 substrate.[18] Y641F mutation in Ezh2 results in higher H3K27 tri-methylation activity.[16] The active site residue Cys-588 is also important for the methyltransferase activity of Ezh2 enzyme and peptides containing this residue display moderately high fluorescence. (Figure 2A and supporting online material) Several mutations are associated with overgrowth diseases in Ezh2 [19]. These include the active site residues Y726, Y736 and E740. In addition, survivin binding was observed in a continuous region in the N-terminal direction from the SANT1 region where mutations were associated with overgrowth syndrome (P132 & Y133).[20]

The SBD domain is located on the N-terminal region of the enzyme, and it is linked to DNA binding. This domain contains residue Ser-21 which is targeted by phosphorylation. Intriguingly, mutation in this position affects the methyltransferase activity of Ezh2.[21] This region is surprisingly far from the active site in the PRC2 core structure.[12] On the other hand, both the entire N-terminal region (residues 1-36) and peptides containing the residues of the active site display high, continuous fluorescence in the survivin interaction assay. The involvement of survivin may interfere or facilitate the connection between these already distant sites and at the same time it may influence the DNA binding properties of Ezh2.

In subunit EED, we see repeated (Figure 2B and supporting online material), more evenly distributed regions of high colocalization with survivin including regions 13-37, 58-77, 123-147 and 223-247. The separation of these regions loosely follows the periodicity of the WD repeats, and the regions are concentrated on one side of the doughnut shaped protein. There is a substantial overlap between survivin interacting peptides and the 70-79 region of EED. This region is highly basic and contributes to DNA binding and chromatin positioning. [22] Similarly, a mutation at the conserved site R236 is also known as EED-linked Weaver syndrome [23, 24]. Almost all residues that participate in the binding of the trimethylated lysine are present in the peptides that have high fluorescence signal in the survivin microarray. In addition, H1K26, H3K9 and H3K27 are part of an AXK motif, and the alanine residue of this motif fits into a small pocket on the surface of EED formed by the hydrophobic moieties of Trp-364, Tyr-308 and Cys-324.

In Suz12 (Figure 2C and supporting online material), the highest affinity peptides (533-607) are concentrated in the C-terminal VEFS-box domain [25], but a continuous region with low binding affinity is present between residue range 248-367. Peptides from the zinc finger region (448-471), have relatively small, but noticeable fluorescence signal. Mutation E610V in Suz12 is known for association with Weaver's syndrome.[23] This amino acid residue belongs to the VEFS-box region of the protein.

Jarid2 contains a sharply defined interacting region between residues 165-199 and several moderately fluorescent regions. (Figure 2D and supporting online material) The ARID domain is required to direct the PRC2 complex to its target location on chromatin and this domain is abundant with overlapping peptides interacting with survivin.[26]

The large number of peptides (5395) on the microarray allowed machine learning of the characteristics that prompt survivin with interaction. On the microarray approximately 40% of the peptides were associated with larger than zero fluorescence intensity and approximately 20% of the peptides had 1000 or higher fluorescence intensity (Figure S2). Seven peptides with high histidine content were excluded since these reacted strongly with the anti-His-tag antibody directly. Thus, the

proportions of interacting and non-interacting peptides are reasonably balanced in this microarray experiment.

Rather than focusing on the sequence of the peptides, we associated the peptides with the abundance of certain atom types in their amino acids (Table S1). Among many ways to categorize the chemical/positional nature of atoms in amino acids, the exact choice had limited practical influence on the outcome. Firstly, the peptides were clustered using Ward's method based on their atom abundance. (Figure 4). The clustering revealed that the similarity in atom type abundance between peptides is correlated with the fluorescence intensity originating from survivin interaction.

Secondly, the peptides were classified as interacting and non-interacting, and a neural network was trained on half of the data set to recognize these two classes based on the features of atom composition. The strength of prediction was tested on the remaining half of the data set. The confusion matrix is shown in Table 1. The prediction specificity and sensitivity appeared to be the most successful for the non-interacting peptides providing solid ground for the negative selection of the peptides, although approximately 2/3 of the interacting peptides were correctly recognized.

Since the peptides overlap in sequence, the sequence similarity between the test and training set could potentially contaminate the test set. By considering every third peptide only for the test and training set, we eliminated the effect of the overlaps. The relative proportions of elements in the confusion matrix are very similar to relative proportions of elements in the confusion matrix when the classification included every peptide. Therefore, we concluded that the strength of predictions did not originate from training set contaminating the test set. The prediction of interacting peptides was only slightly improved when the peptides with fluorescence intensity below 100 were excluded from the training and testing sets. (Table 2) Nevertheless, the number of falsely labelled true interaction partners was substantially reduced when overlapping peptides were considered. For non-overlapping peptides the precision for finding the correct interacting peptide was better, but recall was slightly worse, presumably because of the limited number of interacting peptides in the training set.

We also examined the uniqueness of peptide features as described with amino acid composition. Only eight peptides were significantly associated with at least one pair that had identical amino acid composition, again this cannot explain the strength of predictions. The near exclusive uniqueness of peptides is expected from a random model which assumes uniform probability of the 20 amino acids to occur at any position in the 15 amino acid long peptide. The probability of obtaining a specific peptide sequence is $1/20^{15} = 3.1 \times 10^{-20}$, but the abundance of amino acids is not specific to a particular sequence. Since all permutations of the sequence results in identical atom composition, the probability of obtaining a signature composition in a 15 amino acid peptide is $15!/20^{15} = 4 \times 10^{-8}$. In this microarray design, 10 amino acids are constrained between adjacent peptides. Adjacent peptides with identical amino acid composition is still not expected by chance ($5!/20^{5} = 4 \times 10^{-5}$) in an array of 5395 peptides. Nevertheless, functional protein sequences are not random. This explains presence of a minor fraction of peptides that share identical amino acid composition, and such repeated occurrences were exclusively associated with polypeptide chains belonging to the same protein.

Survivin concentration was 1 µg/ml in in the microarray experiment, which suggests that survivin was present in its monomeric form. At this low concentration, there was a clear link between the amino acid composition of the peptide and survivin location. For evaluating the strength of predictions, one should consider that the local torsions and global structure of the peptide was not even considered in this analysis. Moreover, the limit of detection was set to the smallest possible fluorescence signal. At low fluorescence intensity, counting errors could easily result in missing out weak associations. Nevertheless, the rules governing peptide tethering do not necessarily involve specificities of the peptide sequence. They are probably relatively simple since they are inferred from a relatively small training sample set which can work with similar prediction strength as larger data sets. Even the composition of 15 amino acid peptide segments provides a high degree of

uniqueness and highly scalable intensity for survivin tethering. Survivin has a remarkable three-dimensional structure that lacks deep binding pockets. Still, survivin can selectively associate with and avoid multiple peptides with signature amino acid compositions that do not require a specific level of net charge or hydrophobicity.

To describe how the interaction affects survivin, ^{15}N isotope labeled survivin was titrated while measuring 2D nuclear magnetic resonance spectra (NMR) with PRC2 derived peptides. The peaks associated with assigned protons do not shift substantially during the titration. Figure 5 shows the unclear contrast in NMR chemical shifts in survivin between no Ezh2 peptide present and at the maximum concentration of 5 mM. Although not all positions of the peaks are identical, the central part of survivin, where most of the assigned peaks are located, does not substantially change at the highest peptide concentration. The total signal intensity decreases due to the interaction with the peptide. This NMR observation argues against direct binding mediated by short range interactions and shape complementarity as this could introduce chemical shift alterations of surface residues. We cannot completely discount the possibility that the monomeric and dimeric forms of survivin have different affinity to the studied peptides. Since a high survivin concentration is required for NMR spectroscopy, we could not unambiguously connect these results to the observed tethering in the microarray experiment and the observed colocalization and functional links on the cellular level.

To better mimic the conditions of the microarray experiment and provide kinetic information about the interaction, we performed bio-layer interference (BLI) experiments. We selected three peptides ($EZH2_{171-210}$, $EZH2_{682-701}$, $EZH2_{722-741}$) and one peptide ($EED_{295-314}$) based on the sequences of EZH2 and EED, respectively. The selection of peptides was based on the approximate fluorescence intensity from peptide microarray experiments.

Figure 6 shows BLI interference shift trace profiles associated with the thickness of biolayer as a result of peptide-survivin interaction. The different peptides interact with survivin (at low 1 µg/ml concentration) with varied interaction intensity. (Figure 6a) Incubating survivin with the peptide $EZH2_{171-210}$ reaches the highest amplitude of interference shift during the association phase of the experiment (Figure 6a). The total amplitude of interference shift does not show any tendency of saturation at different survivin concentrations (10 ng/ml-100 µg/ml). The survivin concentration dependence of interference shift for peptide $EZH2_{171-210}$ is shown in Figure 6b.

An empirical kinetic model was developed to simulate the association and dissociation phase of the 1 µg/ml survivin concentration BLI experiment. This model assumes one and two independent exponential processes for the association and dissociation phases, respectively. The two phases of the experiment were modelled independently with zero wavelength shift as starting point for the association phase and a similarly fixed starting wavelength shift for the dissociation phase. The total modeled amplitudes were constrained to the sum of the independent dissociation exponential processes. While the amplitude was assumed to be unknown for the association phase, the dissociation was expected to reach zero interference shift in the model of the dissociation phase. The modeled progress curves are displayed in Figure 6c. A simple exponential association model was sufficient with an amplitude of 0.081 and a rate constant 3.87×10^{-3} s^{-1}, respectively. The more complex dissociation model revealed that the fast dissociation process reverted only approximately 7.6 % of the accumulated amplitude ($\Delta\lambda(0) = 0.071$, $A_{-1} = 0.005$ and $A_{-2} = 0.066$). The dominant fraction of survivin undergoes very slow dissociation process ($k_{-1} = 2.86 \times 10^{-3}$ s^{-1} and $k_{-2} = 2.40 \times 10^{-16}$ s^{-1}). The model predicts that to reach 10% of wavelength shift signal above the pre-association level takes 3.0×10^{8} years, which indicates that the interaction is practically irreversible.

The fast dissociation phase may also be attributed to additional buildup after the first layer of the survivin is saturated. The dissociation profiles suggest irreversible binding to survivin in all examined peptides above a certain survivin concentration that varies from peptide to peptide. This irreversible binding is scalable and peptide specific. In this light, it is not surprising that in the microarray experiment the relatively harsh washing steps did not remove survivin from the array surface and a wide range of fluorescence intensity was discernible. It is important to note that at

higher survivin concentrations the dissociation progress curves become more complex, and more than two exponentials are necessary for modelling. Although it is tempting to assign the independent exponential dissociation processes to multiple binding sites, the number of binding sites is not expected to increase with analyte concentration and the association phase does not indicate multiple parallel kinetic processes. Moreover, the total amplitude does not appear to show any sign of saturation at higher survivin concentrations, which is at odd with the irreversible interaction observed at low survivin concentration. Therefore, simple models with multiple well-defined binding sites appear to be unlikely valid because these models assume that a single layer of survivin accumulate on a limited surface area. It is more likely that multiple layers of survivin are deposited through an interaction we describe as tethering. On the microscopic level this may involve subtle induced conformational changes or other dynamic processes that trigger survivin-survivin self-assembly and phase changes on the surface.

Conclusion

Survivin and PRC2 share functional links and colocalization in the human chromatin. Specific peptides derived from PRC2 subunits attracted survivin at low concentration in a peptide microarray assay and built an irreversible complex. The interacting peptides were frequently originated from important functional regions comprising catalytic, pocket forming, and disease associated amino acid residues. We showed that associating the composition of peptides with colocalization of interaction partners is a promising approach to predict interaction and results in a complex, but reliable pattern. Multiple such peptide regions can be combined into an extremely selective protein-protein targeting system even in the absence of a well-defined 3D structure.

Material and Methods

Gene expression analysis of CD4+ cells

Blood samples of 24 RA patients were collected at the Rheumatology Clinic, Sahlgrenska Hospital, Gothenburg. All RA patients fulfilled the EULAR/ACR classification criteria (Aletaha 2010 Arthritis Rheum 62, 2569-2581) and gave their written informed consent prior to the blood sampling. The study was approved by the Swedish Ethical Evaluation Board (659-2011) and was performed in accordance with the Declaration of Helsinki. The trial is registered at ClinicalTrials.gov with ID NCT03449589. Human PBMC were isolated using density gradient separation on Lymphoprep (Axis-Shield PoC As, Norway). CD4+ cells were isolated using positive selection (Invitrogen, 11331D), and cultured ($1.25x10^6$ cells/ml) in complete RPMI-medium in aCD3 precoated wells (0.5ug/ml, OKT3, Sigma-Aldrich), for 24 h.

RNA from the CD4+ cell cultures was prepared using the Norgen Total micro mRNA kit (Norgen, Ontario, Canada). Quality control was done by Bioanalyzer RNA6000 Pico on Agilent2100 (Agilent, St.Clara, CA, USA). Deep sequencing was done by RNAseq (Hiseq2000, Illumina) at the LifeScience Laboratory, Huddinge, Sweden. Raw sequence data were obtained in Bcl-files and converted into fastq text format using the bcl2fastq program from Illumina. Mapping of transcripts was done using Genome UCSC annotation set for hg38 human genome assembly. Analysis was performed using the core Bioconductor packages in R-studio v.3.6.3. Differentially expressed genes were identified using DESeq2 (v.1.26.0) with Benjamini-Hochberg adjustment for multiple testing.

ChIP seq experiments and analysis

CD4+ cells were isolated from 12 subjects as stated above, stimulated with concanavalin A (0.625 ug/ml, Sigma-Aldrich), and LPS (5 ug/ml, Sigma-Aldrich) for 72 h and pooled in 4 independent samples for chromatin purification. CD4+ cells were cross-linked and lysed according to EpiTect ChIP OneDay kit (Qiagen). After shearing and pre-clearing of chromatin, 1% of the sample was saved as an input. The pre-cleared chromatin was incubated with 2ug anti-Survivin (Tang, Ling, et al (2012) BBRC 421(2) 249-254) (10811, Santa Cruz Biotechnology), anti-H3K27ac (C15410196, Diagenode), anti-H3K27me3 (C15410195, Diagenode) or anti-H3K4m3 (C15411003, Diagenode).

The immune complexes were washed, the cross-links reversed, and the DNA purified according to the EpiTect ChIP OneDay kit (Qiagen). Resulting purified DNA was quality controlled using TapeStation (Agilent, Santa Clara, CA, USA). DNA libraries were prepared using ThruPLEX (Rubicon) and sequenced using the Hiseq2000 (Illumina) following the manufacturer's protocols. Bcl-files were then converted and demultiplexed to fastq using the bcl2fastq program from Illumina.

ChIP-seq analysis

The fastq sequencing files were mapped to the human reference genome (hg38) using the STAR aligner [27] with default parameters apart from setting the alignIntronMax flag to 1 for end-to-end mapping. Quality control of the sequenced material was performed by FastQC tool using MultiQC v.0.9dev0 (Babraham Institute, Cambridge, U.K.). Peak calling was performed using the default parameters. Peaks were identified for enrichment of the immunoprecipitation fraction against input (adjusted $p<10^{-5}$) and annotated by HOMER software [28] in standard mode to closest transcription start site (TSS). Peaks with overlapping localization by at least 1 nucleotide in several samples were merged and further on referred to as one peak.

ChIP seq metanalysis

To identify the transcription regulators in vicinity of survivin-ChIP peaks, we performed colocalization analysis of aggregated human ChIPseq datasets of 1034 transcriptional regulators using ReMap database (http://remap.univ-amu.fr/, accessed 15nov2020). Colocalization enrichment analysis was performed using ReMapEnrich R-script (https://github.com/remap-cisreg/ReMapEnrich, accessed 15nov2020). All comparisons were carried out using hg38 human genome assembly. Two-tailed p-values were estimated and normalized by Benjamini-Yekutielli, maximal allowed value of shuffled genomic regions for each dataset (n=15), kept on the same chromosome (shuffling genomic regions parameter byChrom=TRUE). Default fraction of minimal overlap for input and catalogue intervals was set to 10%. Bed interval files of survivin-ChIP peaks with 0, 1 and 10 kb flanks were prepared. The dataset with 0kb flanks was compared with the Universe. Statistically significant enrichment of peak overlaps (q-value < 0.05, number >100) were selected.

Production and purification of survivin

Survivin in pHIS8 [29] was expressed in E. coli BL21(DE3) star (Merck). Luria Bertani (LB) medium was used, and expression was induced with 0.5 mM IPTG and incubated for 4 h at 30C. Cells were harvested at 6000 g for 30 minutes and resuspended in lysis buffer containing 50 mM Tris pH 7.5, 300 mM NaCl, 20 mM imidazole, 0.25 mM Pefabloc (Sigma-Aldrich)), 20 mg/ml DNAse (Sigma-Aldrich) and 0.2 mg/ml lysozyme (Sigma-Aldrich). The cells were lysed by sonication with an amplitude of 30 A for 10 minutes (10 s on/30 s off) by a Q700 sonicator (Qsonica). Survivin was purified by immobilized metal ion chromatography (IMAC) with a 5 ml HisTrap FF (GE Healthcare) column followed by overnight dialysis in 50 mM Tris pH 7.5, 150 mM NaCl and 1 mM DTT (survivin SEC buffer) and gel filtration using a Superdex 75 100/300GL column (GE Healthcare).

Peptide Microarray

A peptide microarray was designed with the subunits of PRC2: EED, Ezh2, Suz12 and Jarid2. The protein sequences were divided into peptides of 15 amino acids with an overlap of 10 amino acids. Pre-staining of one of the PEPperCHIP Peptide Microarrays was done with the secondary 6X His Tag Antibody DyLight680 antibody at a dilution of 1:1000 and with monoclonal anti-HA (12CA5)-DyLight800 control antibody at a dilution of 1:1000 to investigate background interactions with the protein-derived peptides that could interfere with the main assays. Subsequent incubation of other peptide microarray copies with survivin at a concentration of 1 μg/ml in incubation buffer was followed by staining with the secondary 6X His Tag Antibody DyLight680 (Rockland Immunochemicals) antibody and the monoclonal anti-HA (12CA5)-DyLight800 control antibody

(Rockland Immunochemicals) as well as by read-out at scanning intensities of 7/7 (red/green). HA and His tag control peptides were simultaneously stained as internal quality control to confirm the assay quality and to facilitate grid alignment for data quantification. Read-out was performed with a LI-COR Odyssey Imaging System, while quantification of spot intensities and peptide annotation were done with PepSlide Analyzer. Quantification of spot intensities and peptide annotation were based on the 16-bit gray scale tiff files at scanning intensities of 7/7 that exhibit a higher dynamic range than the 24-bit colorized tiff files shown in Figure S1.

Classification by machine learning

The scikit-learn python library was used to implement the machine learning process.[30] The features of the peptides were the number of atoms that belong to specific atom type categories. For this study, the atoms in the amino acids were assigned according to Table S5. After translating each amino acid in the peptide to atom types, these were accumulated. 5388 peptides from the microarray were divided into equally large training and test sets. The training set was labelled as interacting (fluorescence intensity > 0) or non-interacting (fluorescence intensity = 0). The features were standardized before performing the training. Training was performed with the multi-layer perceptron classifier using the default parameters of scikit-learn. The confusion matrix and prediction accuracy were evaluated by the tools provided by the scikit-learn library.

BLI experiments on survivin interactions with PRC2 derived peptides

Peptide hits from previous peptide microarray experiments were immobilized using amine reactive second-generation biosensor (AR2G, Fortebio). Ethanolamine was used to quench unbound AR2G surface to prevent non-specific interactions. The biosensor was incubated in MQ water for 15 minutes before being used in the Octet instrument (Fortebio) with the following settings: equilibration in water for 60 s and 1000 rpm centrifugation, activation in EDC/s-NHS for 300 s at 1000 rpm, peptide immobilization in immobilization buffer pH 6 (Fortebio) for 900 s and 1000 rpm (50 μM for each dissolved peptide), quenching with 1 M ethanolamine for 300 s and 1000 rpm, baseline measurement in survivin SEC buffer for 600 s and 1000 rpm, association with survivin (100 μM, 10 μM, 1 μM, 0.1 μM, and 0.01 μM concentrations) for 600 s and 1000 rpm, and dissociation step in survivin SEC buffer for 1200 s and 1000 rpm. The assay was performed at 30 ºC. The reference wells and reference biosensors were subtracted by the Octet Data Analysis software.

$\Delta\lambda(t)$ values were converted to dimensionless quantities by dividing the wavelength shifts with 1 nm prior the analysis. The kinetic modelling was performed with the scipy python package [31] using rate equations for the association:

$$\Delta\lambda(t) = A(1 - e^{-kt}) \qquad \text{(eq. 1)}$$

, where A and k are the amplitude and rate constant of the exponential association process, respectively. The rate equation of the dissociation phase was:

$$\Delta\lambda(t) = (\Delta\lambda(0) - A_{-2})e^{-k_{-1}t} + A_{-2}e^{-k_{-2}t} \qquad \text{(eq. 2)}$$

, where $(\Delta\lambda(0) - A_{-2})$ and A_{-2} are the amplitudes of the independent exponential dissociation processes with rate constants k_{-1} and k_{-2}, respectively. The total amplitude is represented by the interference shift at the beginning of the dissociation phase (constant $\Delta\lambda(0)$). The bounds for all parameters were 0 to positive infinity. The starting values were A=1 and k=1 1/s for the association and $A_{-2}=10^{-3}$, $k_{-1}=10^{-4}$ 1/s and $k_{-2}=10^{-9}$ 1/s for the dissociation phase, respectively. The non-linear regression was performed with the default Levenberg-Marquardt algorithm [32] of scipy. The residual plots are displayed in Figure S3.

Isotopic labeling of survivin and NMR titration experiments of survivin by PRC2 derived peptides

^{15}N survivin was produced in M9 medium containing ^{15}N labeled ammonium chloride (Sigma-Aldrich). Expression was induced with 0.5 mM IPTG and incubated for 20-22 h at 15 °C. Cell harvesting and purification were performed using identical protocol as for unlabeled survivin.
The peptides were synthetized and purified to 95% by Biomatik. The peptides were titrated into ^{15}N labelled survivin (0.3 mM) at different measure points in the concentration range described in Table 3. Standard HSQC spectra were recorded for ^{15}N survivin.

Acknowledgements
This work has been funded by grants from the Röntgen-Ångström Cluster Framework of the Swedish Research Council (GK, 2015-06099), the Swedish Research Council (MB, 521-2017-03025), the Swedish Association against Rheumatism (MB, R-566961), the King Gustaf V:s 80-year Foundation (MB), the Regional agreement on medical training and clinical research between the Western Götaland county council and the University of Gothenburg (MB, ALFGBG-717681).

References
1. Mita, A.C., et al., *Survivin: key regulator of mitosis and apoptosis and novel target for cancer therapeutics.* Clin Cancer Res, 2008. **14**(16): p. 5000-5.
2. Altieri, D.C., *Survivin - The inconvenient IAP.* Semin Cell Dev Biol, 2015. **39**: p. 91-6.
3. Jeyaprakash, A.A., et al., *Structure of a Survivin-Borealin-INCENP core complex reveals how chromosomal passengers travel together.* Cell, 2007. **131**(2): p. 271-85.
4. Ruchaud, S., M. Carmena, and W.C. Earnshaw, *The chromosomal passenger complex: One for all and all for one*. Cell, 2007. **131**(2): p. 230-231.
5. Tang, L., et al., *Transcriptional inhibition of p21(WAF1/CIP1) gene (CDKN1) expression by survivin is at least partially p53-dependent: Evidence for survivin acting as a transcription factor or co-factor.* Biochemical and Biophysical Research Communications, 2012. **421**(2): p. 249-254.
6. Kasinath, V., S. Poepsel, and E. Nogales, *Recent Structural Insights into Polycomb Repressive Complex 2 Regulation and Substrate Binding.* Biochemistry, 2019. **58**(5): p. 346-354.
7. Wiles, E.T. and E.U. Selker, *H3K27 methylation: a promiscuous repressive chromatin mark.* Curr Opin Genet Dev, 2017. **43**: p. 31-37.
8. Varambally, S., et al., *The polycomb group protein EZH2 is involved in progression of prostate cancer.* Nature, 2002. **419**(6907): p. 624-629.
9. Kleer, C.G., et al., *EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(20): p. 11606-11611.
10. Pereira, C.F., et al., *ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency*. Cell Stem Cell, 2010. **6**(6): p. 547-56.
11. Qi, W., et al., *An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED.* Nature Chemical Biology, 2017. **13**(4): p. 381-+.
12. Kasinath, V., et al., *Structures of human PRC2 with its cofactors AEBP2 and JARID2.* Science, 2018. **359**(6378): p. 940-944.
13. Nakahara, T., et al., *YM155, a novel small-molecule survivin suppressant, induces regression of established human hormone-refractory prostate tumor xenografts*. Cancer Res, 2007. **67**(17): p. 8014-21.
14. Weaver, T.M., et al., *The EZH2 SANT1 domain is a histone reader providing sensitivity to the modification state of the H4 tail.* Scientific Reports, 2019. **9**.
15. Wu, H., et al., *Structure of the catalytic domain of EZH2 reveals conformational plasticity in cofactor and substrate binding sites and explains oncogenic mutations*. PLoS One, 2013. **8**(12): p. e83737.
16. Kipp, D.R., C.M. Quinn, and P.D. Fortin, *Enzyme-dependent lysine deprotonation in EZH2 catalysis*. Biochemistry, 2013. **52**(39): p. 6866-78.

17. Tatton-Brown, K., et al., *Germline mutations in the oncogene EZH2 cause Weaver syndrome and increased human height.* Oncotarget, 2011. **2**(12): p. 1127-33.
18. Dillon, S.C., et al., *The SET-domain protein superfamily: protein lysine methyltransferases.* Genome Biology, 2005. **6**(8).
19. Gibson, W.T., et al., *Mutations in EZH2 Cause Weaver Syndrome.* American Journal of Human Genetics, 2012. **90**(1): p. 110-118.
20. Cohen, A.S.A., et al., *Weaver Syndrome-Associated EZH2 Protein Variants Show Impaired Histone Methyltransferase Function In Vitro.* Human Mutation, 2016. **37**(3): p. 301-307.
21. Cha, T.L., et al., *Akt-mediated phosphorylation of EZH2 suppresses methylation of lysine 27 in histone H3.* Science, 2005. **310**(5746): p. 306-10.
22. Poepsel, S., V. Kasinath, and E. Nogales, *Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes.* Nat Struct Mol Biol, 2018. **25**(2): p. 154-162.
23. Imagawa, E., et al., *Mutations in genes encoding polycomb repressive complex 2 subunits cause Weaver syndrome.* Human Mutation, 2017. **38**(6): p. 637-648.
24. Cohen, A.S.A., et al., *A novel mutation in EED associated with overgrowth.* Journal of Human Genetics, 2015. **60**(6): p. 339-342.
25. Birve, A., et al., *Su(z)12, a novel Drosophila Polycomb group gene that is conserved in vertebrates and plants.* Development, 2001. **128**(17): p. 3371-3379.
26. Patsialou, A., D. Wilsker, and E. Moran, *DNA-binding properties of ARID family proteins.* Nucleic Acids Res, 2005. **33**(1): p. 66-80.
27. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
28. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.
29. Verdecia, M.A., et al., *Structure of the human anti-apoptotic protein survivin reveals a dimeric arrangement.* Nat Struct Biol, 2000. **7**(7): p. 602-8.
30. Pedregosa, F., et al., *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 2011. **12**: p. 2825-2830.
31. Pauli Virtanen, R.G., Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E.A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.* Nature Methods, 2020. **17**(3): p. 261-272.
32. Levenberg, K., *A Method for the Solution of Certain Non-Linear Problems in Least Squares.* Quarterly of Applied Mathematics, 1944. **2**: p. 164-168.

Table 1 Confusion matrices of predictions

		Predicted label	All peptides		Predicted label	Non-overlapping
		Non-interacting	Interacting		Non-binder	Interacting
Correct label	Non-interacting	1404	297	Non-interacting	474	99
	Interacting	352	645	Interacting	108	218
Precision/recall		0.80/0.83	0.68/0.65		0.81/0.83	0.69/0.67

Table 2 Confusion matrices of predictions with weak interaction partners (0<FI<100) excluded from the analysis.

		Predicted label	All peptides FI >100 or = 0		Predicted label	Non-overlapping FI >100 or = 0
		Non-interacting	Interacting		Non-interacting	Interacting
Correct label	Non-interacting	1409	304	Non-interacting	492	77
	Interacting	282	644	Interacting	118	193
Precision/recall		0.81/0.83	0.69/0.67		0.81/0.86	0.71/0.62

Table 3 Summary of NMR titration experiments

Subunit	Residues	Peptide sequence	Conc. range (mM)	Obs.
EZH2	11-30	PVCWRKRVKSEYMRLRQLKR	6.8×10^{-2} – 1.2	Precipitation started at 1.4 µM
EZH2	171-210	VELVNALGQYNDDDDDDDGDDPEEREEKQKDLEDHRDDKE	4.5×10^{-4} – 2.9	No binding
EZH2	571-590	PCYLAVRECDPDLCLTCGAA	8.4×10^{-2} – 1.5	No binding
EZH2	637-656	FISEYCGEIISQDEADRRGK	7.7×10^{-2} – 1.4	No binding
EZH2	682-701	NKIRFANHSVNPNCYAKVMM	7.6×10^{-2} – 1.4	No binding
EED	58-82	NTPNAPGRKSWGKGKWKSKKCKYSF	4.5×10^{-4} – 2.6	Precipitation started at 3.1 µM
EED	128-152	SQGEIRLLQSYVDADADENFYTCAW	4.5×10^{-4} – 3.8	No binding
EED	231-250	VEGHRDEVLSADYDLLGEKI	7.9×10^{-2} – 1.4	No binding
EED	295-314	FPDFSTRDIHRNYVDCVRWL	3.5×10^{-2} – 1.0	No binding
SUZ12	441-470	TEARDDLHCPWCTLNCRKLYSLLKHLKLCH	2.5×10^{-2} – 0.7	No binding
SUZ12	572-596	PLRPQEMEVDSEDEKDPEWLREKTI	4.5×10^{-4} – 4.4	No binding
JARID2	169-193	PNSMVYFGSSQDEEEVEEEDDETED	4.5×10^{-4} – 4.7	No binding

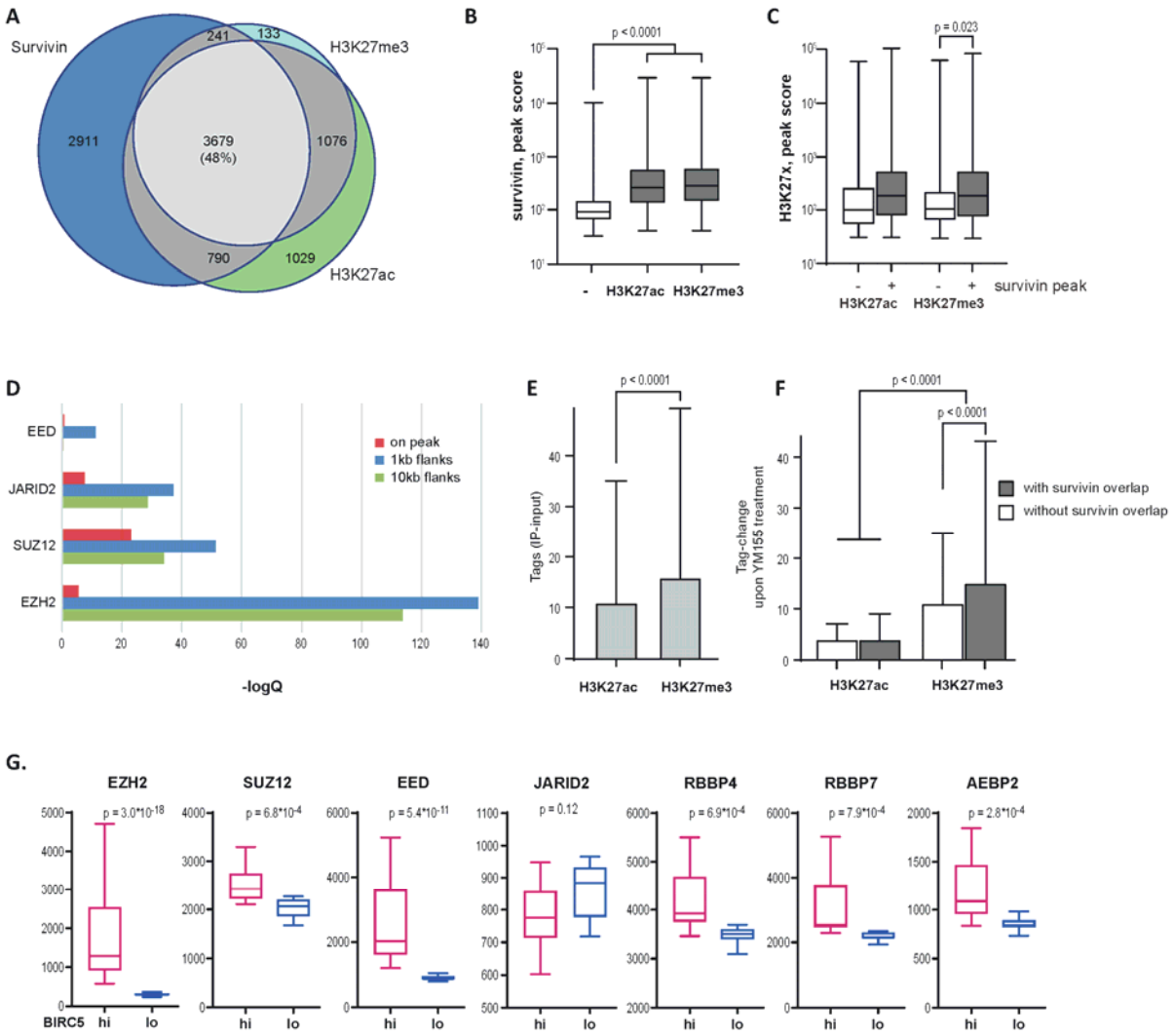

Figure 1 (**A)** Co-localization of the survivin ChIP-peaks with histone H3K27 epigenetic marks. Genome localization of the survivin ChIP-peaks was compared to those of ChIPseq obtained with H3K27ac and H3K27me3 antibodies, on CD4 cells. The comparison was done by annotation of the peaks to the nearest TSS. Venn diagram shows that 48% of the survivin ChIP-peaks are annotated to the same gene as H3K27ac and H3K27me3. (**B)** Peaks annotated to protein coding genes in survivin-ChIP which overlap with peaks in H3K27me3 and H3K4me3-ChIP have higher peak score than those without overlap (overlap = less than 1kb between peak-startbase). Median, IQR and range indicated. Kruskal Wallis with Dunn's multiple comparisons test. (**C)** Peaks, annotated to protein coding genes, in H3K27ac and H3K27me3 -ChIP which overlap with peaks in survivin-ChIP have higher peak score than those without overlap (overlap = less than 1kb between peak-startbase). Median, IQR, range, and unpaired t-test result indicated. (**D)** Co-localisation of ChIPseq peaks for survivin and PRC2 proteins was done by ReMapErichR-script for survivin ChIP-peaks with 0kb flanks (red), 1kb flank (blue), 10kb flanks (green). (**E)** Tags from H3K27ac and H3K27me3-ChIP on CD4 cells. Median and IQR, and unpaired t test result indicated. (**F**) Survivin inhibition using YM155 leads to an increase in H3K27me3 but not H3K27ac, ChIPseq analysis of H3K27ac and H3K27me3 in CD4 cells. H3K27me3-peaks overlapping with survivin-peaks showed significantly more change than those without overlap. Median and IQR, and unpaired t-test result indicated. (**G)** Expression of PRC2 genes in survivin (BIRC5) high and low expressing RA patients. CD4+ T cells from 24 female RA patients were stimulated with anti-CD3 for 48h and sent for RNAseq analysis. RNAseq data was analyzed using DESeq2 split on median of the level of BIRC5 expression. Shown are median with IQR and range, p-values from the paired DESeq2 analysis.

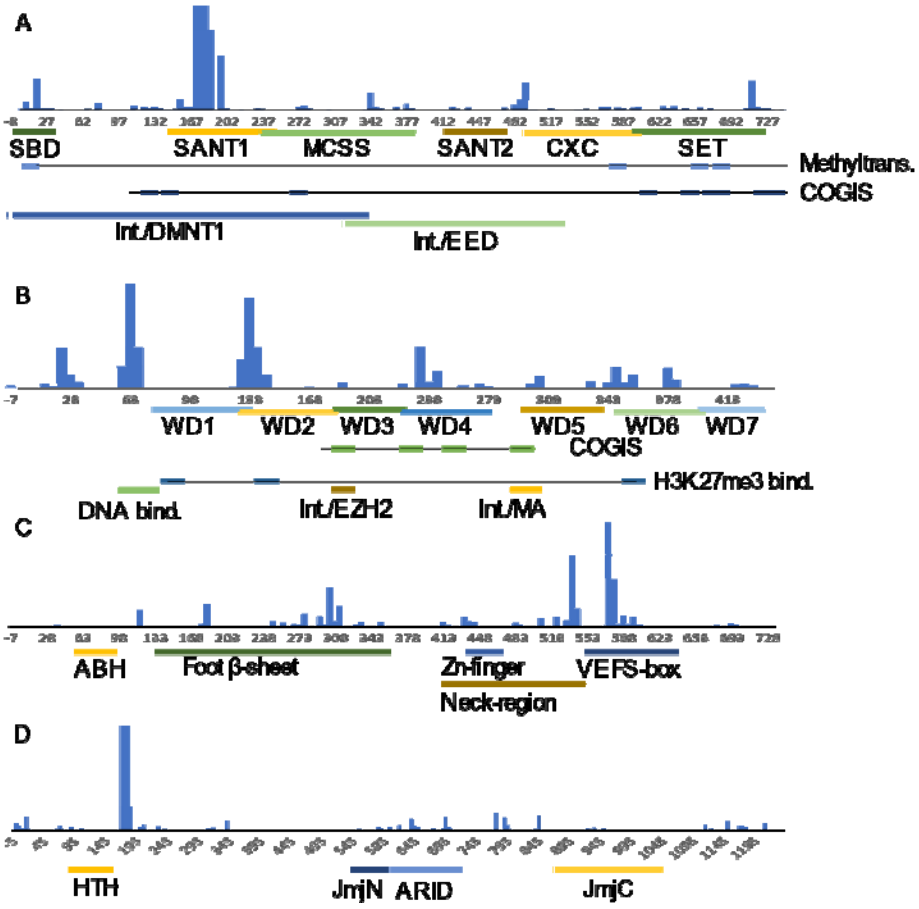

Figure 2 Fluorescence intensity in survivin-peptide microarray of peptides derived from (**A**) Ezh2, (**B**) EED, (**C**) Suz12 and (**D**) Jarid2. The residue numbers refer to the first amino acid of the peptides. Negative numbers mean filling amino acid residues that are not part of the natural sequence of the subunit. The annotation bars are inclusive i.e. it is enough to have one amino acid residue from the region on a peptide to be marked as part of the region. Since typically three peptides contain the same amino acid position, the annotation bars overlap even if the described regions do not.

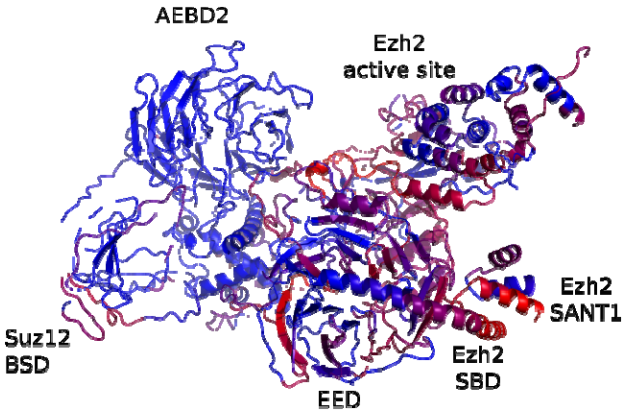

Figure 3 Fluorescence intensity mapped on the 3D structure of the PRC2 core complex (PDB id 6c23).[12] Blue indicates no interaction or no information about interaction (Aebd2). Progressively brighter red indicates logarithmic magnitude of fluorescence intensity in the peptide microarray experiment. The coloring is inclusive i.e., the residue is associated with the average fluorescence intensity of all the peptides the amino acid residue is present on.

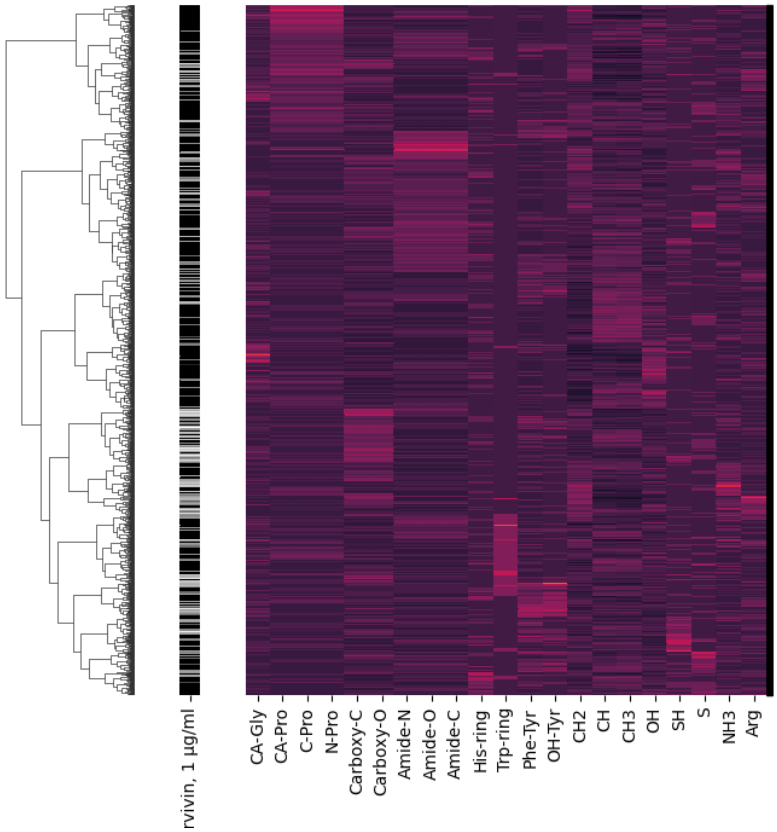

Figure 4 Clustering of peptides based on their atom type abundance (light/dark high/low abundance of atom types, respectively). The grayscale bar indicates the logarithm of fluorescence intensity of the peptide in the survivin peptide microarray experiment (*black* zero intensity, *white* highest level of intensity).

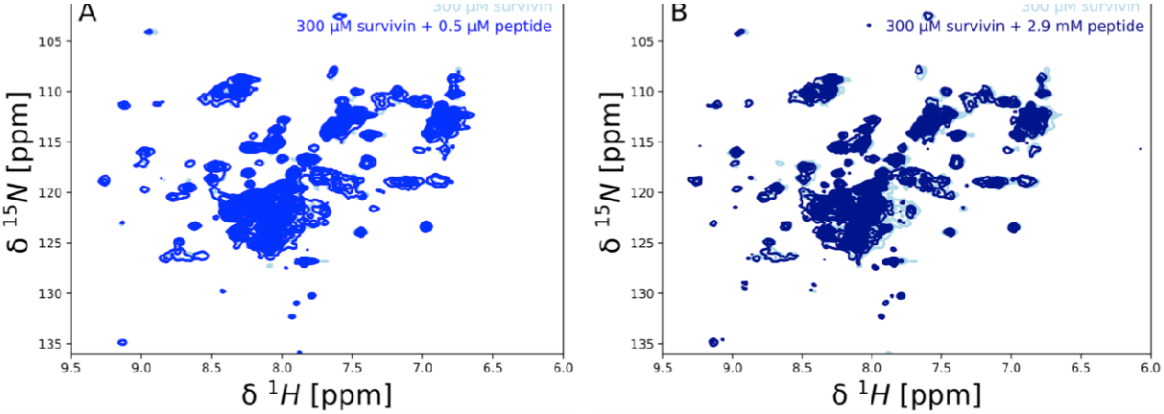

Figure 5 (**A**) Overlay of 2D [^{15}N,^{1}H] NMR spectra of 300 µM survivin in the absence (light blue) and presence (blue) of 0.5 µM peptide VELVNALGQYNDDDDDDDGDDPEEREEKQKDLEDHRDDKE which corresponds to residues 172-211 from EZH2. (**B**) Overlay of 2D [^{15}N,^{1}H] NMR spectra of 300 µM survivin in the absence (light blue) and presence (dark blue) of 2.9 mM of the same peptide.

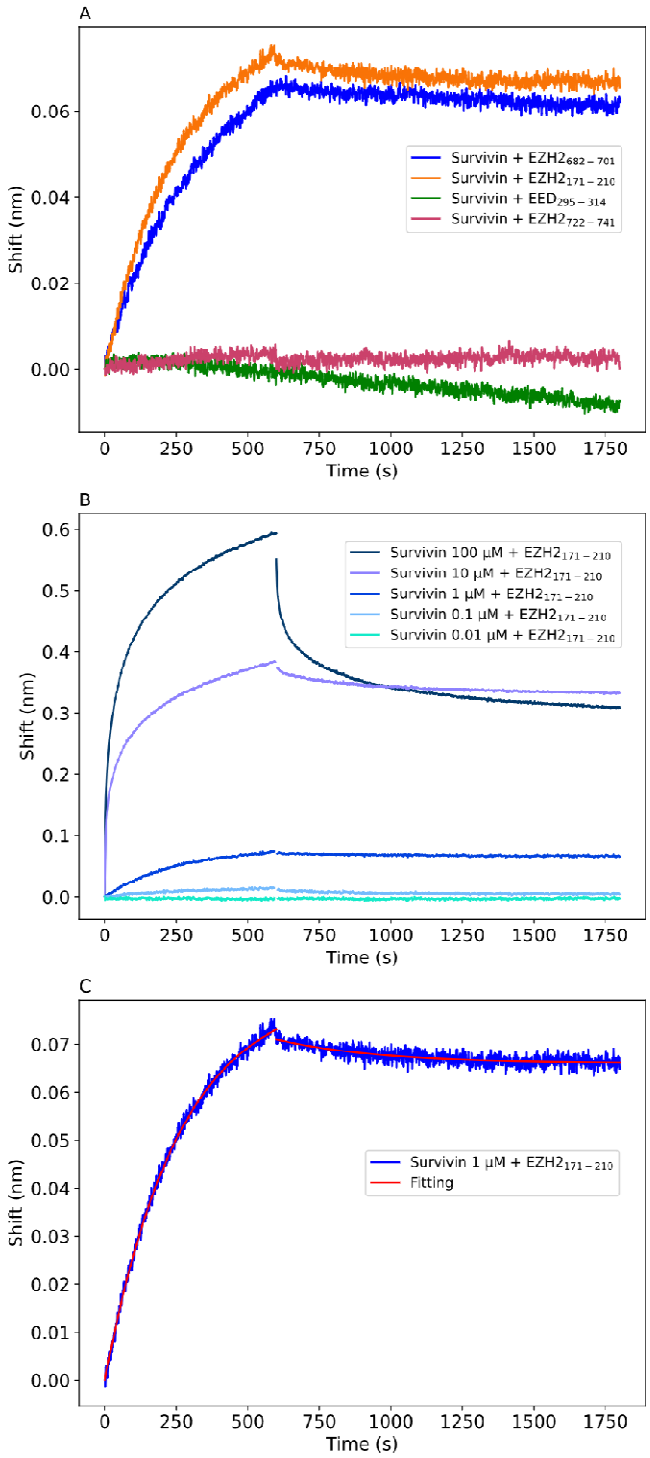

Figure 6 (**A**) Wavelength shift profile from the interaction of 1 μM survivin with different peptides (before loading 50 μM each). (**B**) Wavelength shift profile from the interaction of survivin at different concentrations (100 μM, 10 μM, 1 μM, 0.1 μM, 0.01 μM) with peptide $EZH2_{171\text{-}210}$ (50 μM). (**C**) Curve fitting (in red) of wavelength shift profile (in blue) from the interaction of 1 μM survivin with 50 μM (before loading) peptide $EZH2_{171\text{-}210}$. The wavelength shift profiles were monitored over association (0-600 s) and dissociation (601-1750 s) periods.