Author/s:
Schmidberger, JW;Sharifi Tabar, M;Torrado, M;Silva, APG;Landsberg, MJ;Brillault, L;AlQarni, S;Zeng, YC;Parker, BL;Low, JKK;Mackay, JP

Title:
The MTA1 subunit of the nucleosome remodeling and deacetylase complex can recruit two copies of RBBP4/7

Date:
2016-08-01

Citation:
Schmidberger, J. W., Sharifi Tabar, M., Torrado, M., Silva, A. P. G., Landsberg, M. J., Brillault, L., AlQarni, S., Zeng, Y. C., Parker, B. L., Low, J. K. K. & Mackay, J. P. (2016). The MTA1 subunit of the nucleosome remodeling and deacetylase complex can recruit two copies of RBBP4/7. Protein Science, 25 (8), pp.1472-1482. https://doi.org/10.1002/pro.2943.

The MTA1 subunit of the Nucleosome Remodeling and Deacetylase complex can recruit two copies of RBBP4/7

Jason W Schmidberger*[1], Mehdi Sharifi Tabar*[1], Mario Torrado*[1], Ana PG Silva[1], Michael J Landsberg[2], Lou Brillault[3], Saad AlQarni[1], Yi Cheng Zeng[1], Benjamin L Parker[1], Jason KK Low[1], Joel P Mackay[1]

*These three authors contributed equally.

¶Correspondence should be addressed to JPM (joel.mackay@sydney.edu.au), or JWS (jason.schmidberger@sydney.edu.au)

Phone: +61-2-9351-3906; Fax: +61-2-9351-4726

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

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

[3]Institute for Molecular Bioscience, University of Queensland, Qld, Australia

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version record. Please cite this article as doi:10.1002/pro.2943.

Abstract

The Nucleosome Remodeling and Deacetylase (NuRD) complex remodels the genome in the context of both gene transcription and DNA damage repair. It is essential for normal development and is distributed across multiple tissues in organisms ranging from mammals to nematode worms. In common with other chromatin-remodeling complexes, however, its molecular mechanism of action is not well understood and only limited structural information is available to show how the complex is assembled. As a step towards understanding the structure of the NuRD complex, we have characterized the interaction between two subunits: the metastasis associated protein MTA1 and the histone-binding protein RBBP4. We show that MTA1 can bind to two molecules of RBBP4 and present negative stain electron microscopy and chemical crosslinking data that allow us to build a low-resolution model of an MTA1-$(RBBP4)_2$ subcomplex. These data build on our understanding of NuRD complex structure and move us closer towards an understanding of the biochemical basis for the activity of this complex.

Keywords

NuRD complex, RBBP4, MTA1, Transcription regulation, Chromatin, Protein structure

Introduction

Physical remodeling of the eukaryotic genome is an essential aspect of a wide range of processes, including transcription, replication and DNA repair. ATP-dependent chromatin remodeling complexes are instrumental in such remodeling events. Remodeling complexes all contain a DNA translocase enzyme that harnesses ATP-derived energy to alter the positions, occupancy and composition of nucleosomes, thereby altering the accessibility of DNA to other DNA-binding

factors. The activity of the translocase is modulated by additional subunits that harbour, for example, domains that recognize specific chromatin modifications or other regulatory proteins.

Despite their central role in regulating the genome, a detailed mechanistic description of remodeling is still lacking and no high-resolution structures of remodeling complexes are available. Four classes of remodeling complex have been defined, based on their central remodeling subunit: INO80, ISWI, SWI/SNF and CHD type complexes. Low-resolution (23–50 Å) models of several yeast remodeling complexes derived from single-particle electron microscopy data have been reported (*e.g.*[1,2]) and partial X-ray crystal structures have been reported for several components or domains or other complexes[3-5] but far less is known about CHD-family remodeling complexes, the best-described of which is the nucleosome remodeling and deacetylase (NuRD) complex.

The NuRD complex is conserved across all complex animals and is expressed in most, if not all, tissues. NuRD can repress or activate genes and its activity is required, for example, at all stages of haematopoiesis, regulating both haematopoietic stem cell (HSC) maintenance and differentiation of these cells into distinct lineages (e.g.[6]). NuRD is also emerging as a significant player in efforts to reprogram somatic cells into pluripotent stem cells.[7-9]

In mammalian cells, NuRD comprises the ATP-dependent remodeling enzyme CHD4, the histone deacetylases HDAC1 and -2, the DNA-binding proteins MBD2 and -3,[10] the metastasis-associated proteins MTA1, -2 and -3, the WD40-repeat proteins RBBP4 and -7 and the poorly understood proteins GATAD2A and –B. The various isoforms of each protein are encoded by separate genes but little is known about their functional significance. The MTA-family proteins (**Figure 1A**) contain BAH and SANT domains, which in other proteins have been implicated in nucleosome recognition, whereas RBBP4 and -7 have been shown to bind the transcriptional co-regulator FOG1 [11] as well as both histone H3 [12] and histone H4.[13]

Currently, we have little understanding of how all of these components come together to make the NuRD complex, although structures of a number of domains and several limited subcomplexes have been determined. These subcomplex structures (**Figure 1B**) include an HDAC1-MTA1 dimer of

dimers involving the ELM and SANT domains of MTA1 [14] and a complex formed between RBBP4/7 and a short motif at the C-terminal end of MTA1.[15]

As part of efforts to delineate the architecture of the NuRD complex, we have examined the interaction between RBBP- and MTA-family proteins and we show here that MTA1 and -2 carry two motifs that can each independently recruit a molecule of RBBP4. *In vitro* binding studies show that these two interactions can occur simultaneously and negative-stain single particle electron microscopy combined with covalent crosslinking and mass spectrometry (XL-MS) reveals the overall shape of an MTA:$(RBBP4)_2$ subcomplex. Taken together with the existing subcomplex structures of HDAC1:$MTA_{ELM-SANT}$ and RBBP4:MTA1, these data begin to provide an outline for the physical arrangement of subunits within the NuRD complex.

Results

The C-terminal half of MTA1 and -2 can bind directly to two RBBP subunits

We previously demonstrated that a short sequence from the C-terminal end of MTA1 is able to bind with micromolar affinity to either RBBP4 or RBBP7.[15] This sequence encompasses a helical ^{678}KRAARR motif (an RBBP-binding motif, or RBM) that forms a number of electrostatic interactions with the RBBP partner. Examination of the MTA1 sequence revealed a second RBM with a closely related sequence that is highly conserved across complex eukaryotes (**Figure 1A, 2**). To assess whether this motif can also bind RBBP4, we co-expressed FLAG-$MTA1_{440–550}$ and HA-RBBP4 and captured the HA-RBBP4 with anti-HA beads. As shown in **Figure 3A**, this fragment was able to efficiently pull down FLAG-$MTA1_{440–550}$. We also carried out a triple co-expression, expressing FLAG-$MTA1_{440–550}$, HA-RBBP4 and FLAG-RBBP4. In this case, affinity captured HA-RBBP4 still pulled down FLAG-$MTA1_{440–550}$ but FLAG-RBBP4 did not co-purify, indicating that FLAG-$MTA1_{440-550}$ can bind only a single molecule of RBBP4 (**Figure 3A**).

To assess whether the entire C-terminal half of MTA1 can bind two RBBP4 subunits simultaneously, we co-expressed combinations of FLAG-RBBP4, HA-RBBP4 and a longer (untagged) $MTA1_{449–715}$ construct in HEK293 cells and purified the FLAG-RBBP4 using anti-FLAG Sepharose beads. Western blot analysis (**Figure 3B**, *lane 4*) showed that HA-RBBP4 was robustly pulled down when $MTA1_{449–715}$ was co-expressed, indicating that this portion of MTA1 is able to bind to two RBBP4 molecules simultaneously. Mutation of the ^{678}KRAARR motif to AAAAAA, however, abrogates the ability of $MTA1_{449–715}$ to bind the second RBBP4 (**Figure 3B**, *lane 5*). For corroboration, we purified a co-expressed RBBP4-$MTA1_{449–715}$ complex from HEK293 cells and subjected it to covalent crosslinking using glutaraldehyde. **Figure 3C** shows that the crosslinked complex runs on denaturing SDS-PAGE with an apparent molecular mass of ~130 kDa, which is consistent with the expected mass of a 2:1 RBBP4-$MTA1_{449–715}$ complex (128 kDa).

Covalent crosslinking combined with mass spectrometry (XL-MS)

We next sought to map the RBBP4-$MTA1_{449–715}$ interface using XL-MS. Purified RBBP4-$MTA1_{449–715}$ complex was crosslinked with either disuccinimidyl suberate (DSS) or adipic acid dihydrazide (ADH) as previously described.[16,17] The use of two crosslinkers was to provide us with orthogonal complementary data. Briefly, DSS specifically crosslinks primary amines (i.e. lysine residues and the protein N-terminus) while ADH crosslinks carboxylic acids (i.e. glutamate, aspartate residues and the protein C-terminus). In addition, in the ADH reaction, a separate side-reaction arising from the 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) used can also lead to 'zero-length' crosslinks to form between carboxylic acids and primary amines; a direct condensation reaction between the primary amine and carboxylic acid. These crosslinks will be referred to as 'ZLXL' crosslinks from here on forth. Post-crosslinking, the samples were digested with trypsin, fractionated via size exclusion to enrich for crosslinked peptides and then analysed by LC-MS/MS to identify crosslinked peptides from RBBP4 and MTA1. We identified 58 high-confidence crosslinks (**Supplementary Table I**). Amongst these 58

crosslinks, 13 could be mapped to our previously determined crystal structure of RBBP4 bound to $MTA1_{670–695}$ (15), including $RBBP4_{K22}$-$MTA1_{K686}$, $RBBP4_{K317}$-$MTA1_{K686}$ and RBBP4-RBBP4 intra-protein crosslinks. Importantly, all 13 of these crosslinks were within the maximum distance limits for the crosslinker used, providing confidence in the XL-MS data. We note that a small number of crosslinks were also detected that involved residues from other NuRD components, including MBD3, HDAC1, GATA2DA and CHD4. These most likely indicate that a fraction of our expressed FLAG-RBBP4, which was used as the purification 'handle' in our experiments, has docked with endogenous NuRD subunits.

Negative-stain EM of an $(RBBP4)_2$-$MTA1_{449–715}$ complex

To gain structural insight into the RBBP4-MTA1 interaction, we first purified the RBBP4-$MTA1_{449–715}$ complex in two steps, using FLAG-affinity chromatography and sucrose gradient centrifugation combined with mild glutaraldehyde crosslinking (using the GraFix protocol;[18] **Figure 4A**). We then subjected the complex to negative-stain electron microscopy (EM). **Figure 4B** shows examples of individual particles and 2D class averages obtained using RELION.[19] Approximately 1,700 particles were manually picked and used to generate templates for automatic particle picking; this process yielded a total of 12,000 particles, which was reduced to 9,000 following manual inspection of the dataset and further reduced to 4,000 particles following 2D classification and 3D classification. Two cycles of unsupervised 3D classification in RELION were used to generate first ten, then four 3D classes. The best class from the first cycle was used to generate a new starting model for the next cycle, which resulted in models that had a similar overall shape (**Figure 4C**). The most heavily populated 3D class, which contained 1,700 particles, was further refined to obtain a final model with an estimated resolution of 29.8 Å according to gold standard Fourier shell correlation (0.143 cut-off). 2D projections were calculated from the 3D model and match well to independently generated 2D class averages (**Figure 4B**).

Simple correlation-based fitting of the crystallographic dimer observed for the RBBP4-MTA1 structure (PDB: 4PBY) into the EM-derived envelope (using the "fit in map" tool in UCSF Chimera) revealed a very good overall fit for the two RBBP subunits (**Figure 4D**). Placement of the $MTA1_{449–715}$ into the envelope was not, however, as straightforward. Bioinformatic analysis does not reveal a clear prediction for the structure of this region of MTA1 (or MTA2 or -3). Disorder predictions suggest that the region encompassing residues 545–650 has a high probability of being disordered, although a number of Molecular Recognition Features (MoRFs, short sequences that undergo a disorder-to-order transition upon binding their partner[20]) are also predicted within this region. Secondary structure and domain predictions predict several helical segments within $MTA1_{449–715}$ (**Figure 5A**) but no identifiable domains, suggesting that MTA1 might undergo a significant conformational change upon binding RBBP4 and form an extended interface with the two RBBP4 subunits, similar to the structure observed for the N-terminal part of the MTA1 ELM domain that binds to HDAC1.[14]

In order to model the likely overall structure and placement of $MTA1_{449–715}$ within the complex, we made use of the PORTER secondary structure prediction server[21] and the PRALINE multiple alignment server[22] to define six putative helical regions for $MTA1_{449–715}$ (**Figure 4D**). Helix 6 is the RBM taken directly from the RBBP4-MTA1 crystal structure (PDBi: 4PBY), and Helix 2 ($MTA1_{491-499}$) – corresponding to the RBM identified above (**Figure 2)** – was modeled on Helix 6 ($MTA1_{677–684}$) and placed in an equivalent binding location in the second RBBP4 monomer. Helices 1 (461–488), 3 (505–515), 4 (534–541), and 5 (663–671) were modeled based on helical predictions made by the PRALINE sequence alignments (**Supplementary Figures 1 and 2**). We also made use of distance restraints derived from 19 crosslinks detected in the XL-MS data (**Table I**). This subset comprised the crosslinks between RBBP4 and MTA1 (or between two residues in MTA1 that are greater than 10 residues apart), excluding those that involved MTA1 residues that were not represented in the model (*e.g.*, the $MTA1_{541–663}$ region that is predicted to be disordered). We used the XL-MS data to guide manual placement of the six predicted MTA1 helices into the EM-derived

molecular envelope into which had been fitted two copies of RBBP4. Care was taken to keep charged residues relatively exposed and, where possible, to juxtapose hydrophobic surfaces. The lengths of the sequences linking the six MTA1 helices were also considered in the course of the modeling.

Using this strategy, we were able to generate a model that satisfied 14 of 19 crosslinks (**Figure 5**). Given that our XL-MS dataset included crosslinks that involved endogenous NuRD components, it is possible that some or all of the unsatisfied crosslinks correspond to connections made in larger NuRD subcomplexes. In addition, a recent XL-MS study of the full NuRD complex[23] yielded three RBBP4-MTA2 crosslinks that correspond to pairs of residues found in our model (B1, B2 and B3 in **Figure 5B**). These crosslinks are entirely consistent with our model. Spectra for crosslinks that satisfied the modeling are shown in **Supplementary Figure 3**. Helix 1 was placed in a position such that it formed a coiled coil with the N-terminal helix ($RBBP4_{3–31}$) of one of the RBBP4 monomers, consistent with coiled coil predictions for this portion of MTA1 from PAIRCOIL2.[24] It is notable that this arrangement juxtaposes a cluster of arginines (R484, R484 and R487) at the C-terminal end of Helix 1 with a cluster of acidic residues on the RBBP4 monomer (E357, D358, and E360). Helix 3 was also placed to interact with $RBBP4_{3–31}$ on one side of the model, linking across to Helix 4 on the other side of the model through an 18-residue linker. This arrangement was important for satisfying crosslinks A2 and Z2 for helix 3, and D1 for Helix 4 (**Table I**). The disordered region ($MTA1_{541–663}$) is assumed to pass under the model as indicated in **Figure 5B**, connecting with Helix 5 on the opposite side of the model. Helix 5 leads in turn to the crystallographically defined Helix 6. Notably, a loop region of the RBBP4 crystal structure (4PBY – residues 86–113) had to be remodeled to accommodate MTA1 helices 1 and 2. This loop is absent from other RBBP4-MTA1 structures (PDBs 4PBZ, 4PC0), suggesting it is relatively mobile. Collectively, the model posits that a large portion of $MTA1_{449–715}$ forms a helical core that lies at the interface between two RBBP4 subunits.

Discussion

Our data demonstrate that MTA1 can coordinate the binding of two RBBP subunits through two closely related motifs that are highly conserved throughout phyla that assemble a NuRD complex, suggesting that this arrangement is a characteristic aspect of NuRD structure. The structure of HDAC1 bound to an N-terminal fragment of MTA1 [14] displays a 2:2 stoichiometry, indicating that the MTA1-RBBP4 interactions observed in our work can rationalize the presence of four RBBP4 subunits in the NuRD complex. Given the presence of a number of regions of MTA1 that have strong predictions of disorder (*e.g.*, $MTA1_{350–380}$ and $MTA1_{541–663}$), it is tempting to speculate that a certain degree of flexibility will be observed in the intact NuRD structure – both in the region bordered by the ELM-SANT region and our $MTA1_{449–715}$ ($MTA1_{350–380}$) and even perhaps between the two RBBP subunits.

What will be the function of these four RBBP subunits? Published work suggests that although RBBP4/7 interacts with histone H4, the interaction makes use of (and interacts similarly with) the same surface on RBBP4 that is contacted by the C-terminal RBM.[13] It therefore seems unlikely that any of the four RBBPs will make direct contact with histone H4 in the context of an assembled NuRD complex. In contrast, the surface that contacts both histone H3 and FOG1 is more likely to be available. **Figure 6** shows a simple model of a 2:2:4 HDAC1-MTA1-RBBP4 complex. The arrangements shown makes it clear that the N-terminal tail of a histone H3 tail (shown as a blue dashed line) in the context of a nucleosome can readily contact both the histone H3 binding site on an RBBP4 subunit *and* the HDAC1 active site, which might act on the tail (or on other acetylated lysines in the nucleosome). Similar interactions can take place with the other copy of histone H3 via the second (right-hand side) RBBP4-MTA1 unit. It is tempting to speculate that these RBBP4-containing 'arms' might simultaneously grasp the nucleosome from opposite sides, allowing the chromatin remodeling component of the NuRD complex, CHD4 (not shown), to operate on the nucleosomal DNA, moving it relative to the histone octamer. Alternatively, it is possible that the

substrate of NuRD is a dinucleosome in which each MTA-$RBBP4_2$ entity might contact a separate nucleosome, allowing CHD4 to act on a two-nucleosome unit.

It is worth noting that the first of two papers that quantitatively analysed the stoichiometry of the NuRD complex suggested that NurRD harbours six RBBP subunits.[25] The second paper suggested that some RBBP subunits might be less tightly associated with NuRD. If these additional subunits are present, they might well be attached to NuRD via a different interaction – which would then leave the MTA1/histone H4 binding site on these RBBP4 subunits available for interactions with nucleosome substrates. Finally, given that FOG1 and several other coregulators make use of the histone H3 binding surface of RBBP4, there is clearly still much to learn about how the different RBBP subunits in the NuRD complex coordinate their activity to recognize both histone substrates and co-regulatory binding partners such as FOG1.

Materials and Methods

Plasmids

A DNA sequence of human *MTA1* (Uniprot: Q13330) corresponding to residues 449–715, with no tag ($MTA1_{449\text{-}715}$), cloned in the expression vector pcDNA3.1, was a kind gift from Prof. Gerd A. Blobel. A version of this construct including the mutation ^{678}KRAARR to ^{678}AAAAAA ($MTA1_{449\text{-}715\ MUT}$) was obtained by site-directed mutagenesis. Additionally, genes corresponding to human *RBBP4* (full length; Uniprot: Q09028) with both FLAG and HA N-terminal tags, and residues 440–550 of human *MTA1* ($MTA1_{440–550}$) with an N-terminal FLAG tag, were cloned into pcDNA3.1.

Immunoprecipitation analysis

Combinations of constructs in pcDNA3.1 were co-transfected into suspension HEK Expi293F™ cells (Thermo Fisher) using Polyethylenimine (PEI) (Polysciences) and equimolar plasmid ratios. Prior to each co-transfection, a 1.9-ml culture was grown to a density of 2×10^6 cells/ml in Expi293™ Expression Medium (Thermo Fisher). 3.8 μg of the DNA mix was diluted in 205 μl of

PBS (Thermo Fisher) and vortexed briefly; 7.6 μg of PEI was then added and the suspension was vortexed again and incubated for 20 min at room temperature before being added to the cells. Cells were collected in 1-ml aliquots 65 h after transfection, and each of these lysed by sonication in 0.5 ml of buffer A [50 mM Tris pH 7.9, 150 mM NaCl, 1 mM PMSF, 1× Roche complete protease inhibitor, 0.2 mM DTT, 1% (v/v) Triton X-100]. Cell lysates were incubated on ice for 30 min and then cleared by centrifugation. 'Input' samples were collected at this stage. Cleared lysates were then mixed with 20 μl anti-FLAG or anti-HA Sepharose 4B beads (BioTool) overnight at 4^{o}C. Next, the beads were washed five times with 1 ml of buffer B [50 mM Tris pH 7.5, 150 mM NaCl, 0.5% (v/v) Igepal], and the bound proteins eluted in 60 μl of buffer C [10 mM HEPES pH 7.5, 150 mM NaCl, and either 150μg/ml 3×FLAG peptide or 2 mg/ml 1×HA peptide]. Samples were analysed by Western blot, using anti-FLAG, anti-HA or anti-MTA1 antibodies (Cell Signalling Technologies).

Protein production

The FLAG-RBBP4 and MTA1$_{449–715}$ constructs were co-transfected into suspension Expi293F™cells as described above, and the resulting sub-complex purified also as described above. Briefly, a 150-ml culture of cells was transfected with a mixture of 0.3 mg DNA, 16.7 ml PBS, and 0.6 mg PEI. 50 ml of buffer A and 1 ml of FLAG beads were used, and each wash was done with 100 ml of buffer B. The sub-complex was eluted in 4 ml buffer C containing 150 μg/ml 3×FLAG peptide, and DTT was added to 1 mM final concentration.

Preparation of the RBBP4-MTA1$_{449-715}$ subcomplex for crosslinking-mass spectrometry (XL-MS) analysis

The RBBP4-MTA1$_{449-715}$ subcomplex was purified as described above using FLAG-affinity chromatography. Prior to crosslinking, the excess 3× FLAG peptide used during the elution step

was removed using Zeba Spin desalting gel filtration columns (7K MWCO; ThermoFisher Scientific), and the complex was exchanged into a buffer comprising 30 mM HEPES (pH 7.5) and 225 mM NaCl. For each crosslinking experiment, ~15 µg of subcomplex at a concentration of ~0.15 mg/mL was used. The purified samples were then prepared for XL-MS essentially as described in[16,17] with some modifications. Briefly, for disuccinimidyl suberate (DSS) crosslinking, H_{12}/D_{12}-DSS (1:1 ratio, 25 mM stock solution in anhydrous dimethylformamide; Creative Molecules) was added to a final concentration of 1 mM and incubated for 30 min at 37°C with constant mixing. The excess DSS was then quenched with 100 mM NH_4HCO_3 (50 mM final concentration) and further incubated at 37°C for 20 min. For adipic acid dihydrazide (ADH) crosslinking, H_8/D_8-ADH (1:1 ratio, 100 mg/mL stock solution in 20 mM HEPES pH 7.4; Creative Molecules) and 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) (144 mg/mL stock solution in 20 mM HEPES pH 7.4; Sigma-Aldrich) were added to final concentrations of ~8.3 mg/mL and 12 mg/mL, respectively. The sample was then incubated for 1.5 h at 37°C with constant mixing. The excess DSS was then removed by using the Zeba Spin Desalting gel filtration columns (7K MWCO; ThermoFisher Scientific) into 20 mM HEPES pH 7.5, 150 mM NaCl. Both DSS and ADH crosslinked samples were then dried in a vacuum centrifuge.

Dried, crosslinked protein samples were resuspended in 50 µL of 8 M urea and the proteins were reduced (5 mM TCEP, 37°C, 30 min) and alkylated (10 mM iodoacetamide, 20 min, room temperature in the dark). The samples were then diluted to 6 M urea using 200 mM Tris-HCl pH 8 and Trypsin/Lys-C mix (Promega) was added to an enzyme:substrate ratio of 1:25 (w/w). The solution was incubated at 37°C for 4 h. Following this digestion step, the sample was further diluted to 0.75 M urea using 50 mM Tris-HCl pH 8 and additional Trypsin (Promega) was added at an enzyme:substrate ratio of 1:50 (w/w). The sample was then incubated at 37°C overnight (~16 h). Following overnight digestion, the samples were acidified by addition of formic acid to a final concentration of 2% (v/v) and centrifuged at 16,000 *g* for 10 min. The supernatant was then

desalted using 50-mg Sep-Pak tC18 cartridges (Waters), eluted in 50:50:0.1 acetonitrile:water:formic acid (v/v/v), and dried in a vacuum centrifuge.

For size exclusion chromatography fractionation (SEC), the dried desalted peptides were resuspended in 150 µL of SEC mobile phase [acetonitrile:water:trifluoroacetic acid, 70:30:0.1 (v/v/v)] and loaded onto a Superdex Peptide HR 10/30 column connected to a Biologic DuoFlow™ system (Bio-Rad) with a BioFrac™ fraction collector (Bio-Rad). A flow rate of 0.5 mL/min was used and the separation was monitored by UV absorption at 215, 225, 254 and 280 nm using a Biologic QuadTec™ UV/Vis detector (Bio-Rad). Fractions were collected as 2-min windows over 1.25 column volumes (30 mL). Based on the UV absorption traces, fractions of interest (retention volumes ~12-17 mL) for mass spectrometry were dried in a vacuum centrifuge.

Mass spectrometry

For LC-MS/MS, peptides were resuspended in 3% (v/v) acetonitrile, 0.1% (v/v) formic acid and loaded onto a 20 cm × 75 µm inner diameter column packed in-house with 1.9-µm C18AQ particles (Dr Maisch GmbH HPLC) using an Easy nLC-1000 nanoHPLC (Proxeon). Peptides were separated using a linear gradient of 5–30% Buffer B over 120 min at 200 nL/min at 55^{o}C (Buffer A = 0.1% (v/v) formic acid; Buffer B = 80% (v/v) acetonitrile, 0.1% (v/v) formic acid). Mass analyses were performed using a Q-Exactive mass spectrometer (Thermo Scientific). Following each full-scan MS1 at 70,000 resolution at 200 *m/z* (300–1750 *m/z*; 3×10^6 AGC; 100 ms injection time), up to 10 most abundant precursor ions were selected for MS/MS (17,500 resolution; $1x10^5$ AGC; 60 ms injection time; 32 normalized collision energy; 2 *m/z* isolation window; $1.7x10^5$ intensity threshold; minimum charge state of +3; dynamic exclusion of 20 s).

Peak lists were generated using the msConvert tool[26] and submitted to the database search program Mascot (Matrix Science). The data was searched with oxidation (M) and carbamidomethyl (C) as variable modifications using a precursor-ion and product-ion mass tolerance of ±15 ppm and

±0.02 Da, respectively. The enzyme specificity specified was trypsin with up to two missed cleavages and all taxonomies in the Swiss-Prot database (March 2016; 550,740 entries) were searched. A decoy database of reversed sequences was used to estimate the false discovery rates. To be considered for further analysis, identified peptides had to be top-ranking and statistically significant (p<0.05) according to the Mascot expect metric.

Analysis of XL-MS data

Analysis of the XL-MS data was performed with the pLink software[27] pLink search parameters that differ from the default settings were as follows: precursor mass tolerance ±15 ppm, product-ion mass tolerance ±20 ppm, variable modifications of oxidation (M) and carbamidomethyl (C), enzyme specificity of trypsin with up to two missed cleavages per chain. H_{12}/D_{12}-DSS crosslinker settings were: crosslinking sites were Lys and protein N-terminus, isotope shift 12.075 Da, xlink mass-shift 138.068 Da, monolink mass-shift 156.079 Da. For the H_8/D_8-ADH crosslinker settings: crosslinking sites were Asp, Glu and protein C-terminus, isotope shift 8.050 Da, xlink mass-shift 138.090 Da, monolink mass-shift 156.100 Da. For the DMTMM crosslinker (side reaction in the ADH crosslinking) settings: crosslinking sites were Asp, Glu and protein C-terminus to Lys, xlink mass-shift -18.0106 Da, monolink mass-shift 0 Da. The protein database used for searching consisted of all the NuRD components (CHD4, MTA1, MTA2, MTA3, HDAC1, HDAC2, HDAC3, GATAD2A, GATAD2B, RBBP4, RBBP7, MBD2, MBD3) plus the top 10 contaminants identified (from the Mascot search) in the samples. The default FDR of 5% was used and only peptides with scores $\leq 1x10^{-4}$ were considered for further analysis. Finally, all spectra were also manually verified; only crosslinks with at least four fragment ions on both the alpha- and beta-chain each were retained for modelling.

Sucrose density gradient ultracentrifugation

Sucrose density gradients 2–25% (w/v) in 50 mM HEPES-KOH pH 8.2, 150 mM NaCl were prepared in 12 ml ultracentrifugation tubes (Beckman Coulter) using the GraFix protocol,[18] with 0.15% (v/v) glutaraldehyde in the 25% (w/v) sucrose buffer as the cross-linking agent. Prepared gradients were left standing at 4°C for at least 1 h prior to usage. A glutaraldehyde-free cushion of 600 µl of the 2% (w/v) sucrose buffer was placed on top of the gradients. The sample (200 µl) was then layered on top of the cushion and ultracentrifuged (186,000 *g*, 4°C, 18 h). The gradients were fractionated as 200-µl aliquots collected from the top, and each of these mixed with 20 µl of 1 M Tris pH 8.0 to deactivate remaining glutaraldehyde.

EM sample preparation

Samples of the RBBP4-MTA1$_{449-715}$ subcomplex recovered from GraFix density gradient centrifugation had the sucrose diluted to an approximate final concentration of 0.1% (w/v) by buffer exchange using a Vivaspin 2 device (CTA, 20,000-Da MWCO) (Sartorius Stedim), previously blocked with 0.1% (v/v) Tween-20. The exchange buffer used was 50 mM HEPES KOH pH 8.2, 150 mM NaCl, 1 mM DTT. Finally, the sample was centrifuged at 5,000 *g* for 5 min to remove possible aggregates. Final concentration of the FLAG-RBBP4/MTA1$_{449–715}$ sub-complex was estimated to be ~7 ng/µl.

Negative-stain EM

The RBBP4-MTA1 complex (10 µL) was applied to a glow-discharged, carbon-coated 400-mesh copper grid (GSCu400CC – ProSciTech). After an incubation time of 5 min, the grid was blotted and washed with five drops of distilled water, blotted again and subsequently stained with a 2% (w/v) uranyl acetate solution for one minute. Excess stain was then blotted away and the grid allowed to dry under air at ambient conditions. Images were acquired using a Tecnai T12 TEM operated at 120 kV and equipped with a Direct Electron LC-1100 (4k x 4k) lens-coupled CCD

camera. Images were recorded at a nominal magnification of 52,000× and were binned by a factor of 2, resulting in a pixel size of 5.58 Å at the specimen level. Defocus values ranged from –1 to –2.5 μm.

EM image processing and 3D reconstruction

Data were processed using the RELION software package.[19] Four hundred micrographs were recorded, a subset of which were used for manual picking of an initial 1,730 particles. These were used to generate templates for autopicking, which yielded a dataset of 12,114 particles. Initial, manual inspection of the dataset (removal of obvious false positives) reduced this to 9,000 particles, which were used to generate 90 reference free 2D classes. The same 9000 particles were used for initial 3D classification. Fir this a starting model was built by placing two copies of the monomeric RBBP4 crystal structure (PDB 4PBY) alongside each other in a way that they were touching end on. The model was sampled at 40 Å and low pass filtered in RELION to 50 Å. Ten 3D classes were generated.

A second round of 3D classification was performed, using a smaller but higher quality particle dataset. Particles falling into the 68 best classes from the initial 2D classification were grouped and subjected to a further, more stringent round of manual inspection, with poor quality particles (*e.g.*, obvious aggregates, small particles) removed, leaving the 4,000 best particles to be taken forward for 3D classification and refinement. This time the initial reference model was generated by refitting the RBBP4 monomers into the best 3D model from the first round of 3D classification, which again was converted to a mrc formatted map at 40 Å resolution and low pass filtered to 50 Å in RELION. Four 3D classes were generated by unsupervised 3D classification. Class 1 contained the most particles (1,709 particles) and the model generated from this class was refined further to obtain a final 3D envelope that was used for subsequent modeling of the complex (**Figure 4C/D**). All single particle image processing was performed with no symmetry assumed or imposed. No absolute hand determination was performed and, for subsequent model fitting, the handedness of

the map was flipped as this appeared to give a fit more consistent with the data. The EM map will be deposited in the EMDB prior to publication.

Modeling of the RBBP4-MTA1$_{449-715}$ complex

The RBBP4-MTA1$_{449-715}$ model was generated by fitting the RBBP4 crystallographic dimer from 4PBY[15] into the output EM envelope using the 'fit in map' function of the CHIMERA software.[28] A series of bioinformatics tools were used to analyse the MTA1$_{491-499}$ sequence and guide modeling. This included: a multiple sequence alignment (PRALINE[22]) across a range of taxa, secondary structure prediction (PORTER[21]), and coiled coil analysis using PAIRCOIL2.[24] Predicted helices and short connecting loops were modeled in PYMOL and COOT.[29] Crosslink distances were assessed on the basis of adhering to the following approximate distance restraints: ADH = 22 to 26 Å, DSS = 29 to 30 Å, and ZLXL = 12 to 14 Å.

Supplementary material

Included are three supplementary figures and one supplementary table. Supplementary Figures 1 and 2 are output from a multiple sequence alignment program (PRALINE) indicating the level of conservation across MTA1 variants, as well as secondary structure predictions. Supplementary Figure 3 and Table 1 both displays the raw data associated with the crosslinking work.

Supplementary Figure 1. Multiple sequence alignment of the full MTA1$_{449-715}$ across a range of taxa as output from the PRALINE server. Residues are coloured on the basis of level of conservation. Species included are human (HsMTA1 – UniprotID: Q13330, HsMTA2 – O94776), *Mus musculus* (MmMTA1 – Q8K4B0), *Struthio camelus australis* (ScaMTA1 – A0A093HEK0), *Callorhinchus milii* (CmMTA1 – V9KH55), *Danio rerio* (DrMTA3 – E7EY65), *Drosophila melanogaster* (DmMTA1 – Q9VNF6), and *Habropoda laboriosa* (HlMTA1 – A0A0L7R1K5).

Supplementary Figure 2. Multiple sequence alignment of MTA1$_{449-715}$ across a range of taxa as output from the PRALINE server. Residues are coloured on the basis of secondary structure prediction (PORTER). Species included are human (HsMTA1 – UniprotID: Q13330, HsMTA2 – O94776), *Mus musculus* (MmMTA1 – Q8K4B0), *Struthio camelus australis* (ScaMTA1 – A0A093HEK0), *Callorhinchus milii* (CmMTA1 – V9KH55), *Danio rerio* (DrMTA3 – E7EY65), *Drosophila melanogaster* (DmMTA1 – Q9VNF6), and *Habropoda laboriosa* (HlMTA1 – A0A0L7R1K5).

Supplementary Figure 3. Fragmentation spectra for the crosslinks identified in this study that were used for the EM model. The figure panels are labelled according to the Id of the crosslink used in the EM model. Alpha peptide y- and b-ions are labelled in yellow and green, respectively, whereas the beta peptide y- and b-ions are labelled in russet and cyan, respectively. Cysteine residues coloured in red refers to them carrying the carbamidomethyl modification.

Supplementary Table 1. Intermolecular crosslinks found in this study

Acknowledgements

MJL and LB would like to acknowledge the Australian Microscopy and Microanalysis Research Facility (AMMRF) for access to their facilities as part of this work. This work was funded by grants from the National Health and Medical Research Foundation to JPM (101, 2161, 1058916, 1063301).

References

1. Nguyen VQ, Ranjan A, Stengel F, Wei D, Aebersold R, Wu C, Leschziner AE (2013) Molecular architecture of the ATP-dependent chromatin-remodeling complex SWR1. Cell 154:1220-1231.

2. Tosi A, Haas C, Herzog F, Gilmozzi A, Berninghausen O, Ungewickell C, Gerhold CB, Lakomek K, Aebersold R, Beckmann R, Hopfner KP (2013) Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. Cell 154:1207-1219.

3. Grune T, Brzeski J, Eberharter A, Clapier CR, Corona DF, Becker PB, Muller CW (2003) Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. Mol Cell 12:449-460.

4. Horton JR, Elgar SJ, Khan SI, Zhang X, Wade PA, Cheng X (2007) Structure of the SANT domain from the Xenopus chromatin remodeling factor ISWI. Proteins 67:1198-1202.

5. Yamada K, Frouws TD, Angst B, Fitzgerald DJ, DeLuca C, Schimmele K, Sargent DF, Richmond TJ (2011) Structure and mechanism of the chromatin remodelling factor ISW1a. Nature 472:448-453.

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

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

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

9. Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour AA, Caspi I, Krupalnik V, Zerbib M, Maza I, Mor N, Baran D, Weinberger L, Jaitin DA, Lara-Astiaso D, Blecher-Gonen R, Shipony Z, Mukamel Z, Hagai T, Gilad S, Amann-Zalcenstein D, Tanay A, Amit I, Novershtern N, Hanna JH (2013) Deterministic direct reprogramming of somatic cells to pluripotency. Nature 502:65-70.

10. Scarsdale JN, Webb HD, Ginder GD, Williams DC, Jr. (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. Nucleic Acids Res 39:6741-6752.

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

12. Schmitges FW, Prusty AB, Faty M, Stutzer A, Lingaraju GM, Aiwazian J, Sack R, Hess D, Li L, Zhou S, Bunker RD, Wirth U, Bouwmeester T, Bauer A, Ly-Hartig N, Zhao K, Chan H, Gu J, Gut H, Fischle W, Muller J, Thoma NH (2011) Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell 42:330-341.

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

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

15. Alqarni SS, Murthy A, Zhang W, Przewloka MR, Silva AP, Watson AA, Lejon S, Pei XY, Smits AH, Kloet SL, Wang H, Shepherd NE, Stokes PH, Blobel GA, Vermeulen M, Glover

DM, Mackay JP, Laue ED (2014) Insight into the architecture of the NuRD complex: structure of the RbAp48-MTA1 subcomplex. J Biol Chem 289:21844-21855.

16. Leitner A, Joachimiak LA, Unverdorben P, Walzthoeni T, Frydman J, Forster F, Aebersold R (2014) Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. Proc Natl Acad Sci U S A 111:9455-9460.
17. Leitner A, Reischl R, Walzthoeni T, Herzog F, Bohn S, Forster F, Aebersold R (2012) Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. Mol Cell Proteomics 11:M111 014126.
18. Stark H (2010) GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. Methods Enzymol 481:109-126.
19. Scheres SH (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 180:519-530.
20. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 44:12454-12470.
21. Pollastri G, McLysaght A (2005) Porter: a new, accurate server for protein secondary structure prediction. Bioinformatics 21:1719-1720.
22. Simossis VA, Heringa J (2005) PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. Nucleic Acids Res 33:W289-294.
23. Kloet SL, Baymaz HI, Makowski M, Groenewold V, Jansen PW, Berendsen M, Niazi H, Kops GJ, Vermeulen M (2015) Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. FEBS J 282:1774-1785.
24. McDonnell AV, Jiang T, Keating AE, Berger B (2006) Paircoil2: improved prediction of coiled coils from sequence. Bioinformatics 22:356-358.
25. Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res 41:e28.
26. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak MY, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, Lefebvre B, Kuhlmann F, Roark J, Rainer P, Detlev S, Hemenway T, Huhmer A, Langridge J, Connolly B, Chadick T, Holly K, Eckels J, Deutsch EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P (2012) A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol 30:918-920.
27. Yang B, Wu YJ, Zhu M, Fan SB, Lin J, Zhang K, Li S, Chi H, Li YX, Chen HF, Luo SK, Ding YH, Wang LH, Hao Z, Xiu LY, Chen S, Ye K, He SM, Dong MQ (2012) Identification of cross-linked peptides from complex samples. Nat Methods 9:904-906.
28. Yang Z, Lasker K, Schneidman-Duhovny D, Webb B, Huang CC, Pettersen EF, Goddard TD, Meng EC, Sali A, Ferrin TE (2012) UCSF Chimera, MODELLER, and IMP: an integrated modeling system. J Struct Biol 179:269-278.
29. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Cryst D66:486-501.

Table I. Crosslinks used as EM modelling restraints obtained from XL-MS analyses

Crosslinks from this study

Id	Crosslinker	Protein 1	Residue	Protein 2	Residue	Length in model (Å)
D1	DSS	RBBP7‡	263	MTA1	532	28.1
D2	DSS	RBBP4	120	MTA1	462	33.2
D3	DSS	RBBP4/7	22/21	MTA1	462	17.5
D4	DSS	RBBP4	26	MTA1	509	17.1
D5	DSS	RBBP4/7	22/21	MTA1	686	25.6
D6	DSS	RBBP4	317	MTA1	686	21.7
D7	DSS	MTA1	447	MTA1	527	25.9
D8	DSS	MTA1	509	MTA1	532	27.0
D9	DSS	MTA1	462	MTA1	509	19.7
D10	DSS	MTA1	670	MTA1	686	22.0
D11	DSS	RBBP7	119	RBBP4	156	**
D12	DSS	RBBP7	211	RBBP4	156	**
D13	DSS	RBBP7	119	RBBP4	160	**
A1	ADH	RBBP4	104	MTA1	518	22.9
A2	ADH	RBBP4	166	MTA1	511	30.7
A3	ADH	MTA1	488	MTA1	518	**
A4	ADH	MTA1	511	MTA1	686	**
Z1	ZLXL	RBBP4	104	MTA1	509	19.5*

Z2	ZLXL	MTA1	511	MTA1	686	16.4

Crosslinks from in Kloet *et al.*, 2015

Id	Crosslinker	Protein 1	Residue	Protein 2	Residue	Length in model (Å)
B1	BS3[§]	RBBP4/7	306/307	MTA2[‡]	531	17.7
B2	BS3[§]	RBBP4/7	306/307	MTA2[‡]	533	18.9
B3	BS3[§]	RBBP4/7	306/307	MTA2[‡]	539	14.3

Note: [§]BS3 refers to the use of bis-sulfosuccinimidyl suberate as the crosslinker. [‡]Where possible, RBBP7 and MTA2 residues are mapped onto corresponding RBBP4 and MTA1 residues, respectively, in the EM model. *While this linker distance exceeds the theoretical limit of ~14 Å, residue $RBBP4_{104}$ is located on a flexible loop. **These crosslinks were not compatible with the model.

Figure legends

Figure 1. Schematics of HDAC1, RBBP4 and MTA1

A. Schematics of three components of the NuRD complex. Domains (shown as labelled ovals/boxes) have known structures or the structures of related domains are known. BAH, Elm, SANT, and ZF (zinc finger) are domains of MTA1, whereas RBM refers to RBBP Binding sequence Motifs defined in this paper. Brown regions are low complexity sequences. Black underlining refers to protein sections for which crystal structures are available. Colouring is preserved in B.

B. Three-dimensional structures of NuRD subcomplexes. *Left:* HDAC1-MTA1$_{ELM-SANT}$ dimer of dimers (PDB 4BKX). *Right:* RBBP4-MTA1$_{675-686}$ (PBB 4PBY).

Figure 2. Sequence alignment of a portion of MTA1/2 from a range of eukaryotes.

The alignment highlights the high conservation of the first of the RBBP-binding motifs across a wide range of species, including human (HsMTA1 – UniprotID: Q13330, HsMTA2 – O94776), *Mus musculus* (MmMTA1 – Q8K4B0), *Struthio camelus australis* (ScaMTA1 – A0A093HEK0), *Callorhinchus milii* (CmMTA1 – V9KH55), *Danio rerio* (DrMTA3 – E7EY65), *Drosophila melanogaster* (DmMTA1 – Q9VNF6), and *Habropoda laboriosa* (HlMTA1 – A0A0L7R1K5).

Figure 3. The C-terminal half of MTA1 can bind two molecules of RBBP4. A. Immunoprecipitation analysis showing that MTA1$_{440–550}$, which encompasses a single RBBP-binding motif (RBM) at positions 493–498 (^{493}WHAARH), is able to bind a single RBBP4. Cleared cell lysates (Input) of HEK293 cells expressing the proteins indicated on top, as well as anti-HA immunoprecipitates (anti-HA pulldown), were analysed by Western Blot using anti-HA or anti-FLAG antibodies, to detect the proteins indicated on the left. The absence of FLAG-RBBP4 in the elution of the triple co-transfection (*lane 3*) suggests that only one RBBP4 can bind to MTA1$_{440–550}$.

B. Immunoprecipitation analysis showing that $MTA1_{449-715}$, which contains the two RBMs at positions 493–498 and 678–683, is able to bind two molecules of RBBP4. Cleared cell lysates (Input) of HEK293 cells expressing the proteins indicated on top, as well as anti-FLAG immunoprecipitates (anti-FLAG pulldown), were analysed by Western Blot using anti-FLAG, anti-HA or anti-MTA1 antibodies, to detect the proteins indicated on the left. The results show that FLAG-RBBP4 can efficiently pull down HA-RBBP4 in the presence of $MTA1_{449–715}$, suggesting that this fragment can bind two molecules of RBBP4. However, this effect is not observed when using an $MTA1_{449–715}$ version mutated at one of the RBMs (^{678}KRAARR to AAAAAA). **C.** Sypro-stained SDS-PAGE showing purified RBBP4-$MTA1_{449–715}$ before and after treatment with glutaraldehyde. The crosslinked complex runs at a molecular weight consistent with the formation of a 2:1 (RBBP4:$MTA1_{449–715}$) complex.

Figure 4. Negative stain EM of the RBBP4:$MTA1_{449–715}$ complex.

A. Sypro-Ruby stained SDS-PAGE showing selected fractions of a 2–25% sucrose plus 0–0.15% glutaraldehyde "GraFix" gradient after anti-FLAG affinity purification of the complex. The position of the species of interest is indicated on the right by a triangle, and fraction #15, which was selected for the EM analysis, is boxed. The lane corresponding to the non-crosslinked sample loaded onto the sucrose gradient is labelled as "Input". "R" indicates the position of RBBP4, and "M" the position of $MTA1_{449–715}$.

B. Examples of typical individual particles (*top left*) and reference-free 2D class averages observed in negative-stain EM analysis of RBBP4-$MTA1_{449–715}$ (*bottom left*). An FSC plot (*top right*) and sample 2D projections of the final reconstruction (*bottom right*) are shown.

C. 3D reconstruction workflow for RBBP4-$MTA1_{449–715}$ obtained from RELION. Numbers shown under models refer to the number of particles associated with each stage of model development.

D. Refined 3D reconstruction of negatively-stained RBBP4-MTA1$_{449–715}$. A copy of the RBBP4 crystallographic dimer (PDB 4PBY) is fitted into the envelope as a rigid body (*right*). The RBM helices from MTA1 are shown in orange and the RBBP4 subunits are shown in grey.

Figure 5. Proposed model of RBBP4-MTA1$_{449–715}$ with associated cross-linking data.

A. Schematic of MTA1$_{449–715}$ showing the predicted helices. Colour coding is used in the rest of the Figure. **B–D.** Model of the complex. The RBBP4 crystallographic dimer taken from PDB 4PBY is shown in two shades of grey. Modeled MTA1$_{449–715}$ helices are: Helix 1 (461–488), Helix 2 (also RBM1, 491–499), Helix 3 (505–515), Helix 4 (534–541), Helix 5 (663–671), and Helix 6 (also RBM2, 677–684). A large putative disordered region (541–663) is not included in the model; the position of this region is indicated by a dashed orange line in B (*right*). Crosslinking data are shown as blue (DSS — basic), red (ADH — acidic), black (ZLXL) and orange (BS3 — basic; from (23)) and are labelled accordingly.

Figure 6. Model of the HDAC-MTA1-RBBP4 portion of the NuRD complex.

The model has been made by simply juxtaposing the HDAC-MTA1 crystal structure with two copies of our model of the RBBP4-MTA1$_{449–715}$ complex and adding the crystal structure of the nucleosome (1AOI). The HDAC1 active sites are shown as yellow spheres and the histone H3 binding site on RBBP4 is circled. The N-terminal 40-residue tail of histone H3 is indicated schematically to show that it is within reach of both the histone-H3-binding site on RBBP4 and the HDAC1 active site. Similar interactions are possible for the right-hand RBBP4 subunits, which could interact with the second histone H3 tail on the back of the model as drawn.

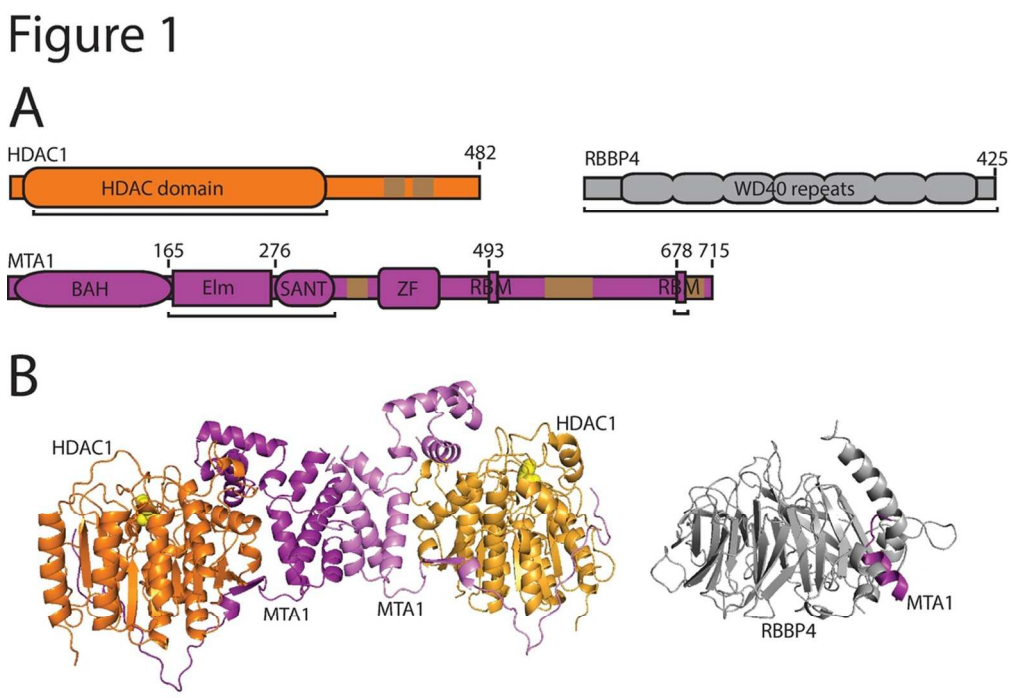


Figure 1. Schematics of HDAC1, RBBP4 and MTA1.
A. Schematics of three components of the NuRD complex. Domains (shown as labelled ovals/boxes) have known structures or the structures of related domains are known. BAH, Elm, SANT, and ZF (zinc finger) are domains of MTA1, whereas RBM refers to RBBP Binding sequence Motifs defined in this paper. Brown regions are low complexity sequences. Black underlining refers to protein sections for which crystal structures are available. Colouring is preserved in B.
B. Three-dimensional structures of NuRD subcomplexes. Left: HDAC1-MTA1$_{ELM-SANT}$ dimer of dimers (PDB 4BKX). Right: RBBP4-MTA1$_{675-686}$ (PBB 4PBY).
99x70mm (300 x 300 DPI)

Figure 2

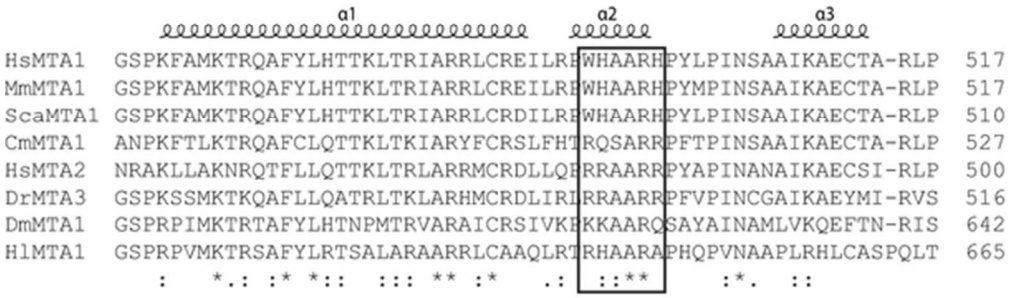


Figure 2. Sequence alignment of a portion of MTA1/2 from a range of eukaryotes. The alignment highlights the high conservation of the first of the RBBP-binding motifs across a wide range of species, including human (HsMTA1 – UniprotID: Q13330, HsMTA2 – O94776), Mus musculus (MmMTA1 – Q8K4B0), Struthio camelus australis (ScaMTA1 – A0A093HEK0), Callorhinchus milii (CmMTA1 – V9KH55), Danio rerio (DrMTA3 – E7EY65), Drosophila melanogaster (DmMTA1 – Q9VNF6), and Habropoda laboriosa (HlMTA1 – A0A0L7R1K5).
51x18mm (300 x 300 DPI)

Figure 3

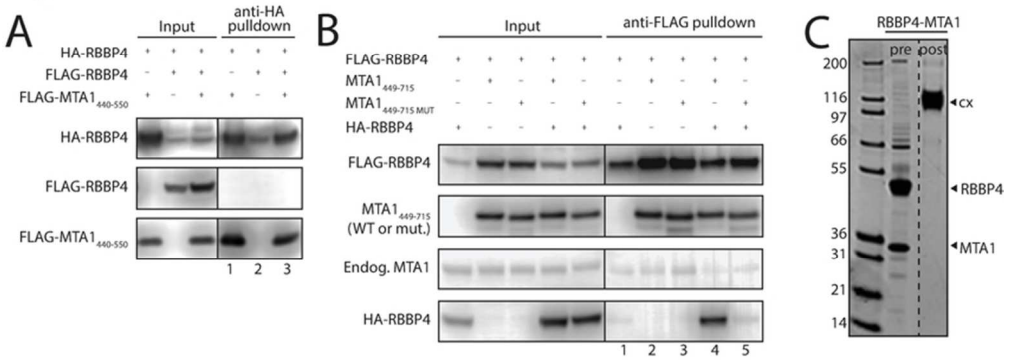


Figure 3. The C-terminal half of MTA1 can bind two molecules of RBBP4. A. Immunoprecipitation analysis showing that $MTA1_{440–550}$, which encompasses a single RBBP-binding motif (RBM) at positions 493–498 (^{493}WHAARH), is able to bind a single RBBP4. Cleared cell lysates (Input) of HEK293 cells expressing the proteins indicated on top, as well as anti-HA immunoprecipitates (anti-HA pulldown), were analysed by Western Blot using anti-HA or anti-FLAG antibodies, to detect the proteins indicated on the left. The absence of FLAG-RBBP in the elution of the triple co-transfection (lane 3) suggests that only one RBBP4 can bind to $MTA1_{440–550}$. B. Immunoprecipitation analysis showing that $MTA1_{449-715}$, which contains the two RBMs at positions 493–498 and 678–683, is able to bind two molecules of RBBP4. Cleared cell lysates (Input) of HEK293 cells expressing the proteins indicated on top, as well as anti-FLAG immunoprecipitates (anti-FLAG pulldown), were analysed by Western Blot using anti-FLAG, anti-HA or anti-MTA1 antibodies, to detect the proteins indicated on the left. The results show that FLAG-RBBP4 can efficiently pull down HA-RBBP4 in the presence of $MTA1_{449–715}$, suggesting that this fragment can bind two molecules of RBBP4. However, this effect is not observed when using an $MTA1_{449–715}$ version mutated at one of the RBMs (^{678}KRAARR to AAAAAA). C. Sypro-stained SDS-PAGE showing purified RBBP4-$MTA1_{449–715}$ before and after treatment with DSS. The crosslinked complex runs at a molecular weight consistent with the formation of a 2:1 (RBBP4:$MTA1_{449–715}$) complex.
71x29mm (300 x 300 DPI)

Figure 4

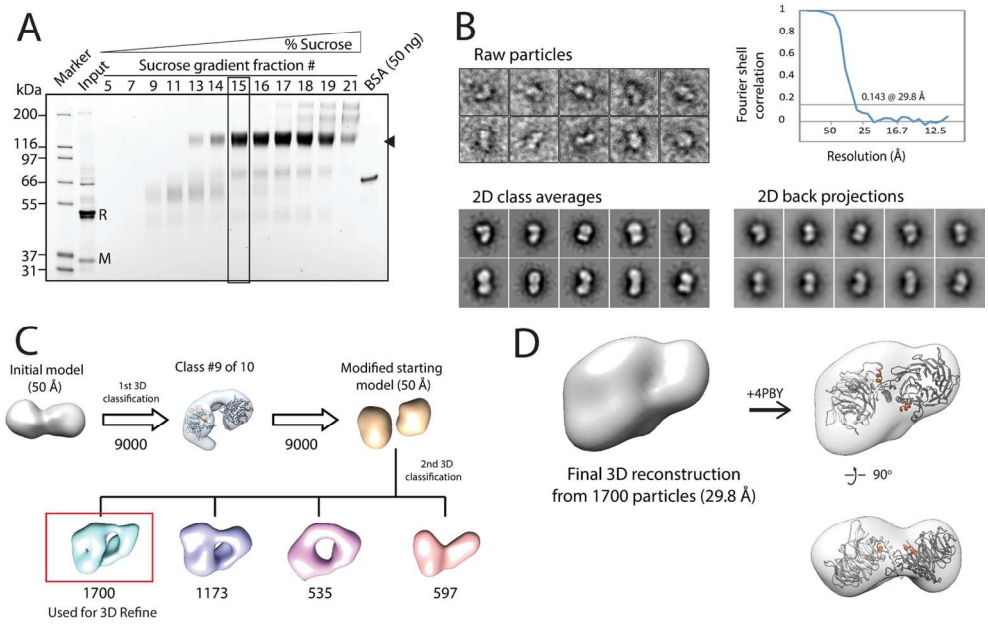


Figure 4. Negative stain EM of the RBBP4:MTA1$_{449–715}$ complex.\nA. Sypro-Ruby stained SDS-PAGE showing selected fractions of a 2–25% sucrose plus 0–0.15% glutaraldehyde "GraFix" gradient after anti-FLAG affinity purification of the complex. The position of the species of interest is indicated on the right by a triangle, and fraction #15, which was selected for the EM analysis, is boxed. The lane corresponding to the non-crosslinked sample loaded onto the sucrose gradient is labelled as "Input". "R" indicates the position of RBBP4, and "M" the position of MTA1$_{449–715}$.\nB. Examples of typical individual particles (top left) and reference-free 2D class averages observed in negative-stain EM analysis of RBBP4-MTA1$_{449–715}$ (bottom left). An FSC plot (top right) and sample 2D projections of the final reconstruction (bottom right) are shown.\r\nC. 3D reconstruction workflow for RBBP4-MTA1$_{449–715}$ obtained from RELION. Numbers shown under models refer to the number of particles associated with each stage of model development. \nD. Refined 3D reconstruction of negatively-stained RBBP4-MTA1$_{449–715}$. A copy of the RBBP4 crystallographic dimer (PDB 4PBY) is fitted into the envelope as a rigid body (right). The RBM helices from MTA1 are shown in orange and the RBBP4 subunits are shown in grey.
136x93mm (300 x 300 DPI)

Figure 5

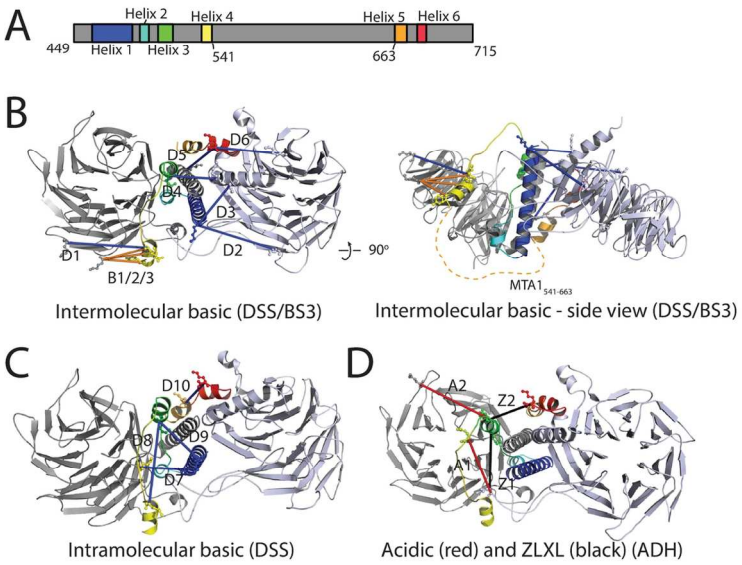


Figure 5. Proposed model of RBBP4-MTA1$_{449–715}$ with associated cross-linking data. A. Schematic of MTA1$_{449–715}$ showing the predicted helices. Colour coding is used in the rest of the Figure. B–D. Model of the complex. The RBBP4 crystallographic dimer taken from PDB 4PBY is shown in two shades of grey. Modeled MTA1$_{449–715}$ helices are: Helix 1 (461–488), Helix 2 (also RBM1, 491–499), Helix 3 (505–515), Helix 4 (534–541), Helix 5 (663–671), and Helix 6 (also RBM2, 677–684). A large putative disordered region (541–663) is not included in the model; the position of this region is indicated by a dashed orange line in B (right). Crosslinking data are shown as blue (DSS — basic), red (ADH — acidic), black (ZLXL) and orange (BS3 — basic; from (8)) and are labelled accordingly.
138x101mm (300 x 300 DPI)

Figure 6

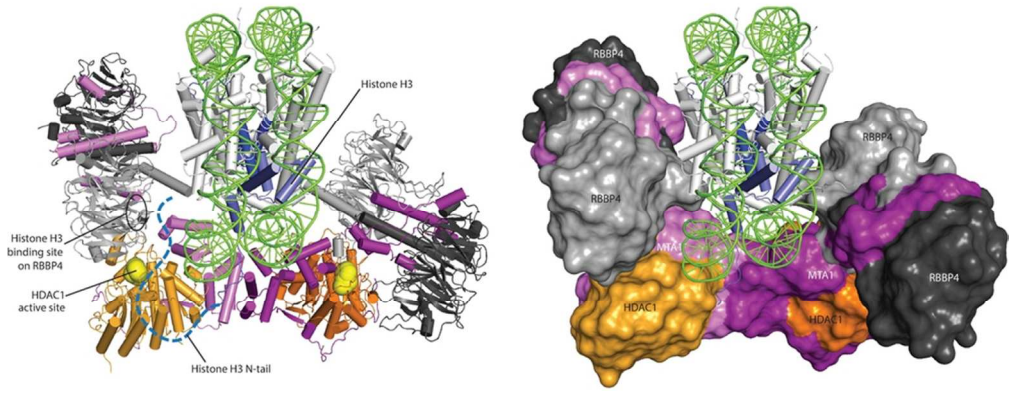


Figure 6. Model of the HDAC-MTA1-RBBP4 portion of the NuRD complex. The model has been made by simply juxtaposing the HDAC-MTA1 crystal structure with two copies of our model of the RBBP4-MTA1$_{449–715}$ complex and adding the crystal structure of the nucleosome (1AOI). The HDAC1 active sites are shown as yellow spheres and the histone H3 binding site on RBBP4 is circled. The N-terminal 40-residue tail of histone H3 is indicated schematically to show that it is within reach of both the histone-H3-binding site on RBBP4 and the HDAC1 active site. Similar interactions are possible for the right-hand RBBP4 subunits, which could interact with the second histone H3 tail on the back of the model as drawn.
84x37mm (300 x 300 DPI)

Supplementary Table 1. Intermolecular crosslinks found in this study.

Crosslinker	Crosslink type	Protein 1	Residue	Protein 2	Residue	Length in model (Å)	Used as restraints in EM model (Id)
DSS	Inter-protein	RBBP4/7	22/21	MTA1	462	17.5	D3
DSS	Inter-protein	RBBP4/7	22/21	MTA1	686	25.6	D5
DSS	Inter-protein	RBBP4	26	MTA1	509	17.1	D4
DSS	Inter-protein	RBBP4	120	MTA1	462	33.2	D2
DSS	Inter-protein	RBBP4	178	MTA1	626	Not in model	-
DSS	Inter-protein	RBBP4	215	MTA1	626	Not in model	-
DSS	Inter-protein	RBBP7[‡]	263	MTA1	532	28.1	D1
DSS	Inter-protein	RBBP4	317	MTA1	554	Not in model	-
DSS	Inter-protein	RBBP4	317	MTA1	686	21.7	D6
DSS	Inter-protein	RBBP4	156	RBBP7[‡]	119	84.1 (inconsistent with model)	-
DSS	Inter-protein	RBBP4	156	RBBP7[‡]	211	102.1 (inconsistent with model)	-
DSS	Inter-protein	RBBP4	160	RBBP7[‡]	119	78.7 (inconsistent with model)	-
DSS	Inter-protein	MBD2/3	190/46	HDAC1/2	89/90	Not in model	-
DSS	Intra-protein	RBBP7[‡]	119	RBBP7[‡]	159	6.2	-
DSS	Intra-protein	RBBP4	120	RBBP4	160	6.2	-
DSS	Intra-protein	RBBP4	156	RBBP4	215	14.5	-
DSS	Intra-protein	RBBP4	160	RBBP4	215	23.1	-
DSS	Intra-protein	MTA1	447	MTA1	527	25.9	D7
DSS	Intra-protein	MTA1	462	MTA1	509	19.7	D9
DSS	Intra-protein	MTA1	509	MTA1	532	27.0	D8

DSS	Intra-protein	MTA1	527	MTA1	532	13.8	-
DSS	Intra-protein	MTA1	532	MTA1	549	Not in model	-
DSS	Intra-protein	MTA1	532	MTA1	554	Not in model	-
DSS	Intra-protein	MTA1	532	MTA1	582	Not in model	-
DSS	Intra-protein	MTA1	532	MTA1	686	50 (inconsistent with model)	-
DSS	Intra-protein	MTA1	532	MTA1	626	Not in model	-
DSS	Intra-protein	MTA1	582	MTA1	670	Not in model	-
DSS	Intra-protein	MTA1	596	MTA1	626	Not in model	-
DSS	Intra-protein	MTA1	626	MTA1	686	Not in model	-
DSS	Intra-protein	MTA1	626	MTA1	670	Not in model	-
DSS	Intra-protein	MTA1	670	MTA1	686	22.0	D10
DSS	Intra-protein	MTA2/3[‡]	285/288	MTA2[‡]	323	Not in model	-
DSS	Intra-protein	MTA2[‡]	531	MTA2[‡]	539	Not in model	-
DSS	Intra-protein	MTA2[‡]	611	MTA2[‡]	618	Not in model	-
DSS	Intra-protein	CHD4	959	CHD4	969	Not in model	-
DSS	Intra-protein	HDAC1	457	HDAC1	466	Not in model	-
DSS	Intra-protein	HDAC2	462	HDAC2	468	Not in model	-
DSS	Intra-protein	MBD3	41	MBD2/3	46/190	Not in model	-
DSS	Intra-protein	MBD3	114	MBD3	124	Not in model	-
DSS	Intra-protein	GATA2DA	27	GATA2DA	30	Not in model	-
DSS	Intra-protein	GATA2DA	451	GATA2DA	459	Not in model	-
ADH	Inter-protein	RBBP4	104	MTA1	518	22.9	A1
ADH	Inter-protein	RBBP4	166	MTA1	511	30.7	A2

ADH	Intra-protein	RBBP4	5	RBBP4/7	20/19	21.1	-
ADH	Intra-protein	RBBP4	60	RBBP4	104	22.1	-
ADH	Intra-protein	RBBP4/7	124/123	RBBP4	166	7.7	-
ADH	Intra-protein	RBBP4/7	148/147	RBBP4	166	15.4	-
ADH	Intra-protein	RBBP4/7	153/152	RBBP4	162	16.1	-
ADH	Intra-protein	MTA1	488	MTA1	518	41.3 (inconsistent with model)	-
ADH	Intra-protein	MTA1	511	MTA1	518	15.1	-
ZLXL	Inter-protein	RBBP4/7	20/19	MTA1	582	Not in model	-
ZLXL	Inter-protein	RBBP4	104	MTA1	509	19.5*	Z1
ZLXL	Inter-protein	RBBP4/7	124/123	MTA1	582	Not in model	-
ZLXL	Intra-protein	RBBP4/7	148/147	RBBP4	215	12.2	-
ZLXL	Intra-protein	MTA1	488	MTA1	582	Not in model	-
ZLXL	Intra-protein	MTA1	511	MTA1	686	16.4	Z2
ZLXL	Intra-protein	MTA2	523	MTA2	539	Not in model	-
ZLXL	Intra-protein	CHD4	1060	CHD4	1064	Not in model	-

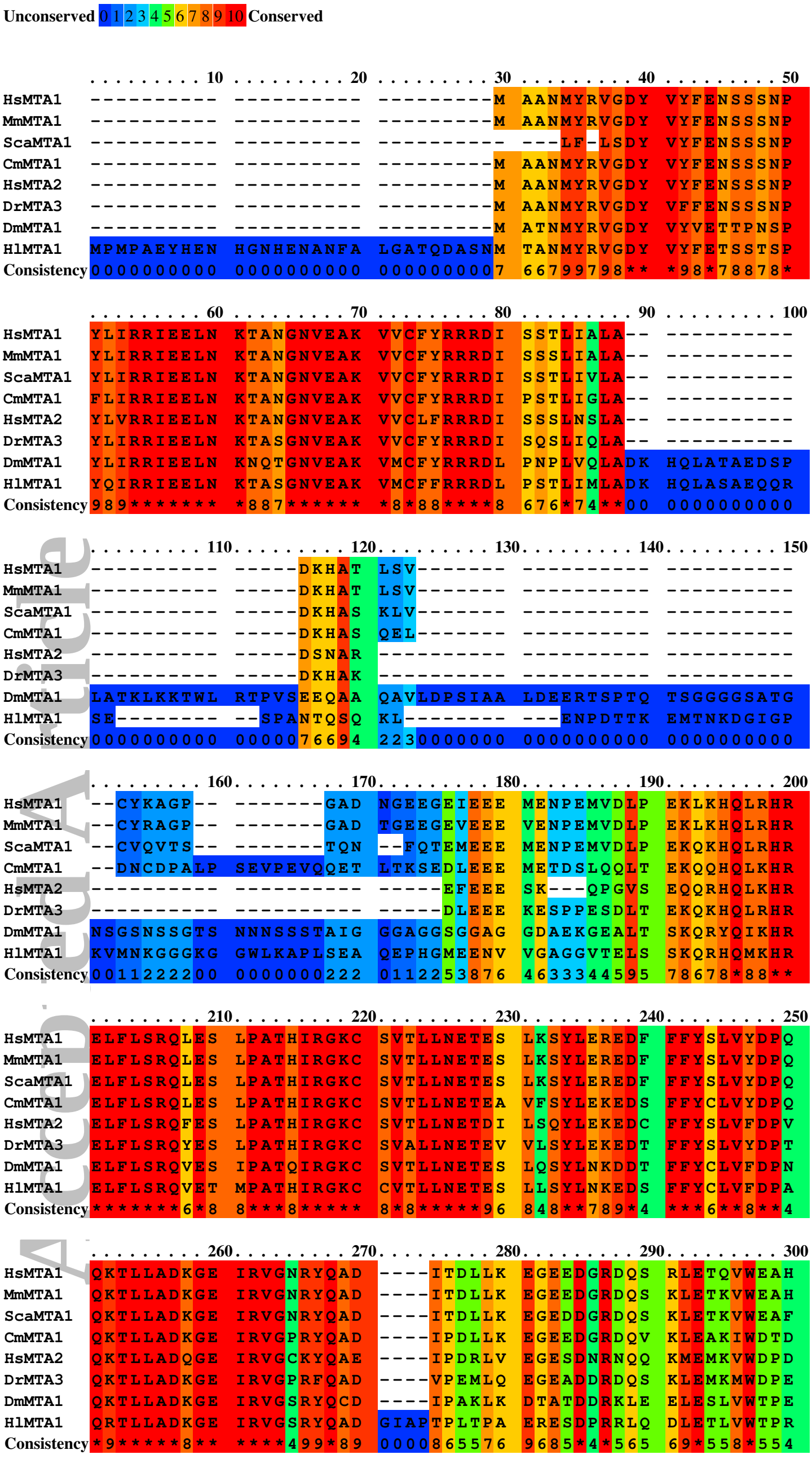
Notes: [†]Where possible, RBBP7 and MTA2 residues are mapped onto corresponding RBBP4 and MTA1 residues, respectively, in the EM model. *While this linker distance exceeds the theoretical limit of ~14 Å, residue $RBBP4_{104}$ is located on a flexible loop.

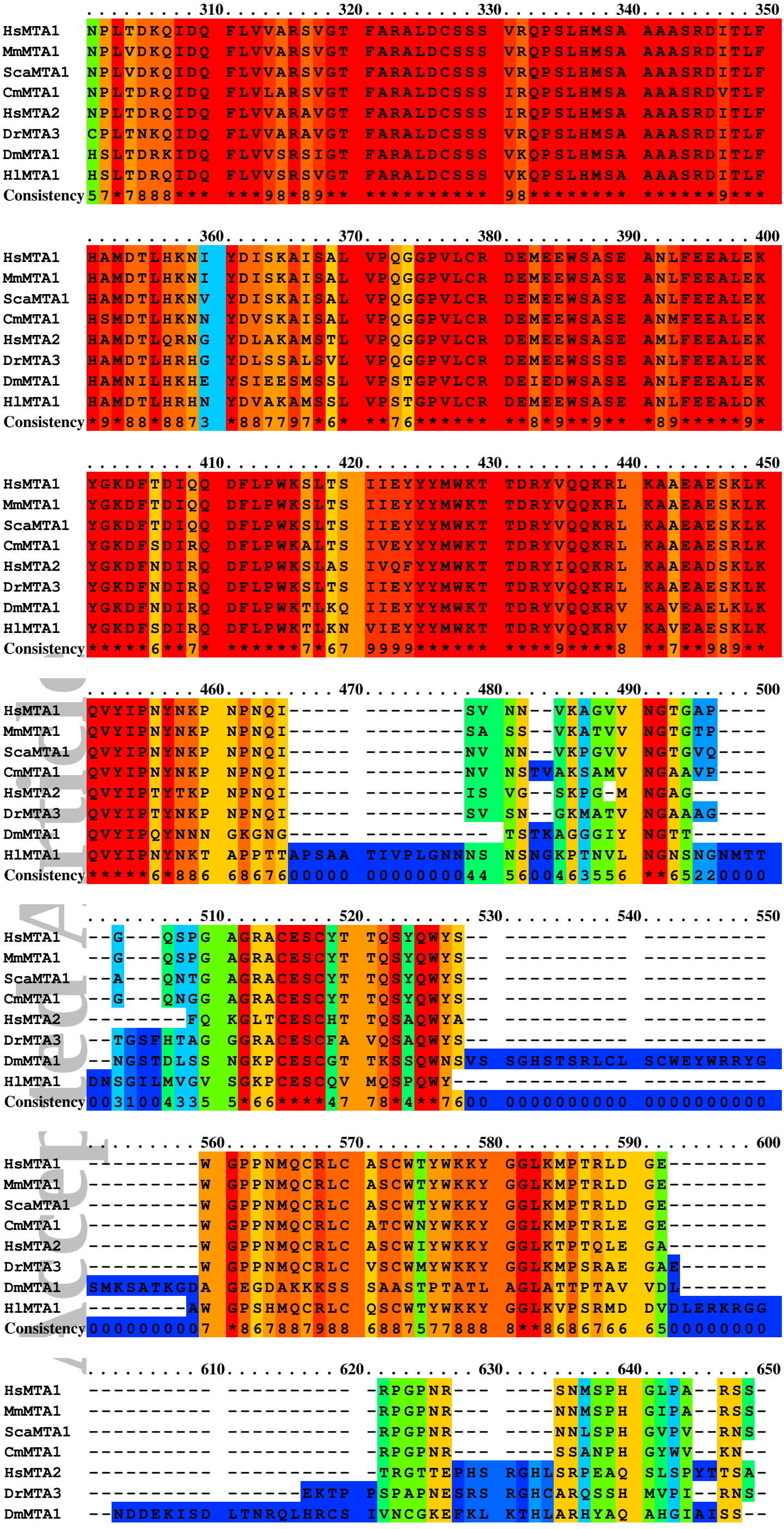
04/17/2016 09:03:39 AM

Results colour-coded for amino acid conservation

The current colourscheme of the alignment is for **amino acid conservation**.

The conservation scoring is performed by PRALINE. The scoring scheme works from 0 for the least conserved alignment position, up to 10 for the most conserved alignment position. The colour assignments are:





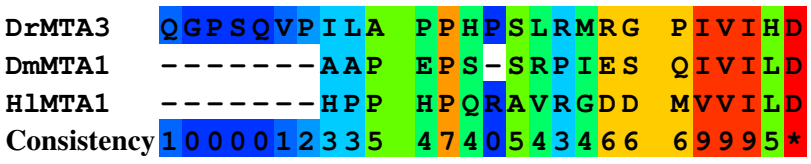
Results colour-coded for amino acid conservation

04/17/2016 09:03:39 AM

```
HlMTA1       TGSDEESKGI  GGAHRPHRCS  IPSCGKEFKL  KAHLSRHYAS  AHGVDLRGSG
Consistency  0000000000  0000000000  0455566011  1021663556  5435006640

                      660         670         680         690         700
HsMTA1       -------GSP  KFAMKTRQAF  YLHTTKLTRI  ARRLCREILR  PWHAARHPYL
MmMTA1       -------GSP  KFAMKTRQAF  YLHTTKLTRI  ARRLCREILR  PWHAARHPYM
ScaMTA1      -------GSP  KFAMKTRQAF  YLHTTKLTRI  ARRLCRDILR  PWHAARHPYL
CmMTA1       -------ANP  KFTLKTRQAF  CLQTTKLTKI  ARYFCRSLFH  TRQSARRPFT
HsMTA2       -------NRA  KLLAKNRQTF  LLQTTKLTRL  ARRMCRDLLQ  PRRAARRPYA
DrMTA3       -------GSP  KSSMKTKQAF  LLQATRLTKL  ARHMCRDLIR  LRRAARRPFV
DmMTA1       -------GSP  RPIMKTRTAF  YLHTNPMTRV  ARAICRSIVK  PKKAARQSAY
HlMTA1       ASGGGGGGSP  RPVMKTRSAF  YLRTSALARA  ARRLCAAQLR  TRHAARAPHQ
Consistency  0000000778  8357*8978*  5*6876988 7  **57*85676  5459**5854

                      710         720         730         740         750
HsMTA1       PINSAAIKAE  CTA-RLPEAS  Q-SPLVLKQA  V-----RKPL  EAVLRYLETH
MmMTA1       PINSAAIKAE  CTA-RLPEAS  Q-SPLVLKQV  V-----RKPL  EAVLRYLETH
ScaMTA1      PINSAAIKAE  CTA-RLPEAS  E-NPLLLKQV  V-----RKPL  EAVLRYLESH
CmMTA1       PINSAAIKAE  CTA-RLPEIG  E-KPIKLKPS  V-----RKTL  ESIAKYLELH
HsMTA2       PINANAIKAE  CSI-RLPKAA  K-TPLKIHPL  V-----RLPL  ATIVKDLVAQ
DrMTA3       PINCGAIKAE  YMI-RVSEGM  TGRPLKPKSS  P-----RSTL  TSVLQYLESR
DmMTA1       AINAMLVKQE  FTN-RISGKS  Q-AEIKKLLL  LK-PKDRGSV  TKIANRLGAP
HlMTA1       PVNAAPLRHL  CASPQLTSKS  S-VELRILAR  AVRPRPRPRV  TDIATRLGDH
Consistency  89*6569967  6650886646  5047855544  600000*448  5596654*545

                      760         770         780         790         800
HsMTA1       P---------  ----------  -------RPP  KPDPVKSV--  -------SSV
MmMTA1       P---------  ----------  -------RPP  KPDPVKSS--  -------SSV
ScaMTA1      P---------  ----------  -------CPP  KPDPAKSL--  -------SSS
CmMTA1       P---------  ----------  --------SV  KPICSLA---  --------VH
HsMTA2       A---------  ----------  --------PL  KPKTPRGT--  -------K--
DrMTA3       P---------  ----------  --------AT  HVQRPHRT--  -------PGL
DmMTA1       G----SGPHE  WLVLTPKDKM  PLPAVVSFPK  PPKAPDGSLV  YDRVPNKSPD
HlMTA1       PAPRQPGDWD  WLALTAAAQR  KQPDRVSFPR  PPKAPDGSLL  YERVPNKSEV
Consistency  6000000000  0000000000  0000000163  5843435400  0000000532

                      810         820         830         840         850
HsMTA1       LSSLTPAKVA  PVINNGSPTI  LGKR--SYEQ  HNGVDGNMKK  RLLMPSRGLA
MmMTA1       LSSLTPAKSA  PVINNGSPTI  LGKR--SYEQ  HNGVDGNMKK  RLLMPSRGLA
ScaMTA1      LNNLTPAKFT  PVINNGSPTI  LGKR--SYEQ  HNGMDGNMKK  RLLMPSRGLP
CmMTA1       SGSVTPAKIL  PVLNHGSPTI  LGKR--GYEQ  HNGIDGTMKK  RLLMPFRGKT
HsMTA2       -TPINRNQLS  Q--NRGLGGI  MVKR--AYET  MAG-------  ----------
DrMTA3       QVQPPRRLLS  SLPSHGPLGM  LGKR--SYHH  HSRVESAERR  AGATG-----
DmMTA1       VVAVPADKEL  TIIPTQATST  IRKRAHEDQQ  LNGTEVTIVP  SGPPAKRPNK
HlMTA1       DRLTVTPPQP  QPSMQAQQTI  LKRTRPPFDE  INGSDGIALS  AGLPGGPPAK
Consistency  2345444534  5435465367  8498005766  5784643444  4343323322

                      860         870         880         890         900
HsMTA1       ----------  --NHGQARHM  GP-SRNLLLN  GKSYPTKVRL  IRGGSLPPVK
MmMTA1       ----------  --NHGQTRHM  GP-SRNLLLN  GKSYPTKVRL  IRGGSLPPVK
ScaMTA1      ----------  --NHGQTRQM  GP-SRNLLLN  GKSYPTKVRL  IRGGSMPPVK
CmMTA1       LAQSCSRLHS  HPNHGQVRQM  AAVNRGYMIN  GKPYSRSSKI  LKPGSLPPMK
HsMTA2       ----------  ----------  A--GVPFSAN  GRPLA---SG  IRSSSQPAAK
DrMTA3       ----------  ---QENPAHI  V--GPILQHN  GSSTG---GS  SLRGSGLMLR
DmMTA1       DPMPSHCPSP  EQFAAMMAAS  GQPLSRHHLN  GKQ-----KI  AQMARGGNGR
HlMTA1       RAHHSQQLHP  KHTLEHTAPT  VLPLAPPLNG  RAAHPHTLPH  GPPLSRSNAR
Consistency  0000000000  0023343434  5204434448  8653312243  4535834458

                      910         920         930         940         950
HsMTA1       RRRMNWIDAP  DDVFYMATEE  TRKIRKLLSS  SETKRAARRP  YKPIALRQSQ
MmMTA1       RRRMNWIDAP  DDVFYMATEE  TRKIRKLLSS  SETKRAARRP  YKPIALRQSQ
ScaMTA1      RRRMNWIDAP  DDVFYMATEE  TRKIRKLLSS  SEAKRAARRP  YKPIILRP--
CmMTA1       GRRMNWIDAP  DDVLFIATEE  TRKIRKLLGP  VDMKRAARQP  YKQIFIRQAR
HsMTA2       RQKLNPADAP  NPVVFVATKD  TRALRKALTH  LEMRRAARRP  NLPLKVKPTL
DrMTA3       KRRPNWIDAP  DDSFFLVSRE  TRKARRMLSR  SQLRRACRQP  CEQISLRRVP
DmMTA1       KQVISWMDAP  DDVYFRANDT  HKKTRKILSA  VDLRRAARKP  WRTLPIKP--
HlMTA1       KQVISWMDAP  DDVYFRASDQ  TKRLRKTLSS  VELRRAARKP  WRRLPAPL--
Consistency  6756876***  8886858766  7876*96*75  5758**8*7*  4648376411

                      960         970
HsMTA1       A-----LPPRP  PPP-APVNDE  PIVIED
MmMTA1       A-----LPLRP  PPP-APVNDE  PIVIED
ScaMTA1      ------MQAV   QLR-QPMNDE  PIIIED
CmMTA1       V-----SCAVS  QPSRQPVNEE  PIIIED
HsMTA2       IAVRPPVPLP  APSHPASTNE  PIVLED
```

Results colour-coded for amino acid conservation

04/17/2016 09:03:39 AM



Results colour-coded for amino acid conservation

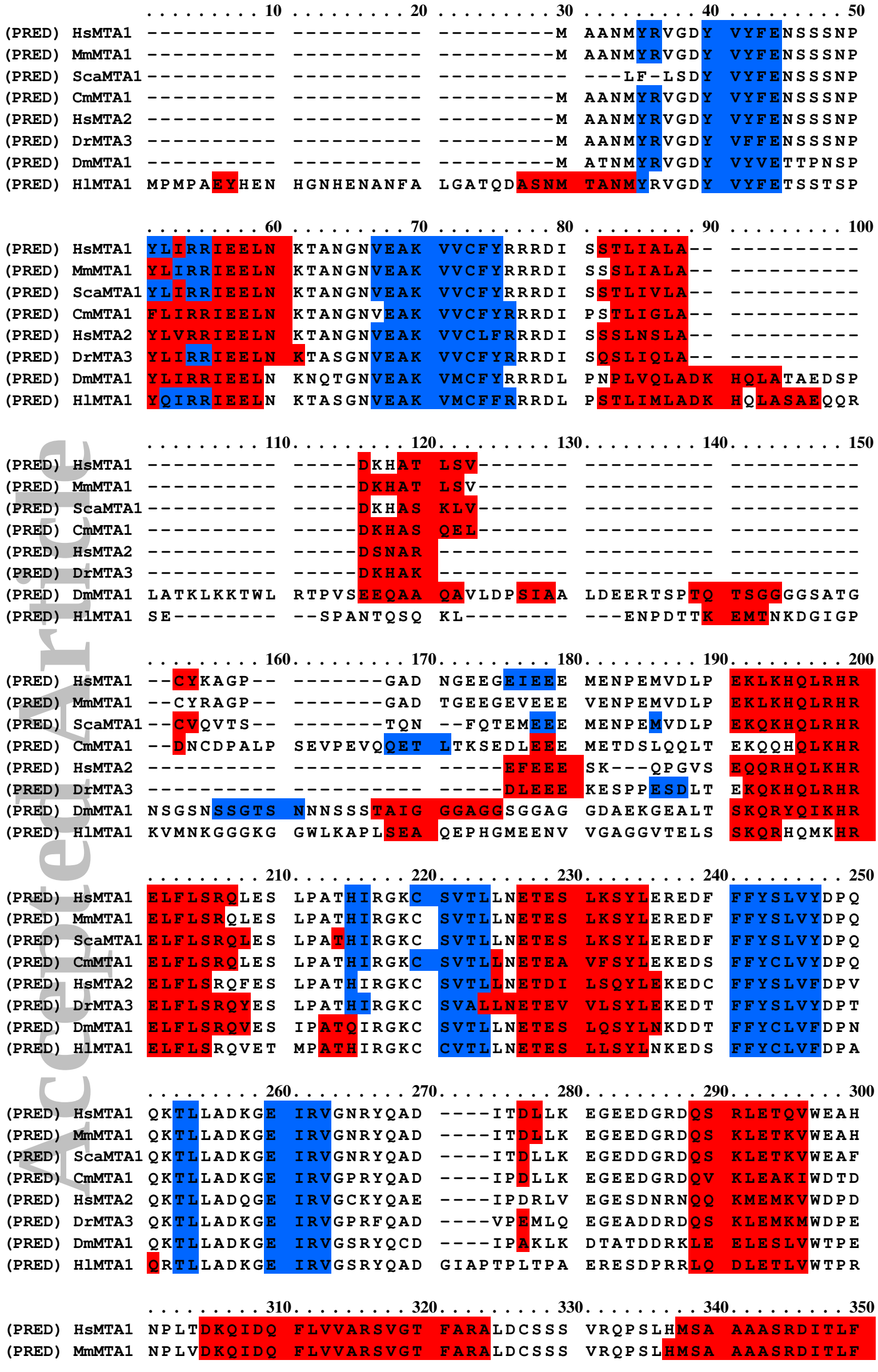
04/17/2016 09:03:38 AM

Results colour-coded for secondary structure

The current colourscheme of the alignment is for **secondary structure type**.

The 3-state (H, E C) secondary structure for each sequence is represented by a colour. If a sequence in the alignment has no colours assigned, this means that either there is no DSSP information available (if this was requested), or that no prediction was possible for that sequence (if this was requested).
The colour assignments are:

HELIX (H) **STRAND (E)** You have selected to perform secondary structure prediction using **DSSP** (Kabsch and Sander, 1983) and **Porter** (Pollastri and McLysaght, 2005).



```
                    .........360.........370.........380.........390.........400
(PRED) HsMTA1       HAMDTLHKNI  YDISKAISAL  VPQGGPVLCR  DEMEEWSASE  ANLFEEALEK
(PRED) MmMTA1       HAMDTLHKNI  YDISKAISAL  VPQGGPVLCR  DEMEEWSASE  ANLFEEALEK
(PRED) ScaMTA1      HAMDTLHKNV  YDISKAISAL  VPQGGPVLCR  DEMEEWSASE  ANLFEEALEK
(PRED) CmMTA1       HSMDTLHKNN  YDVSKAISAL  VPQGGPVLCR  DEMEEWSASE  ANMFEEALEK
(PRED) HsMTA2       HAMDTLQRNG  YDLAKAMSTL  VPQGGPVLCR  DEMEEWSASE  AMLFEEALEK
(PRED) DrMTA3       HAMDTLHRHG  YDLSSALSVL  VPQGGPVLCR  DEMEEWSSSE  ANLFEEALEK
(PRED) DmMTA1       HAMNILHKHE  YSIEESMSSL  VPSTGPVLCR  DEIEDWSASE  ANLFEEALEK
(PRED) HlMTA1       HAMDTLHRHN  YDVAKAMSSL  VPSTGPVLCR  DEMEEWSASE  ANLFEEALDK

                    .........410.........420.........430.........440.........450
(PRED) HsMTA1       YGKDFTDIQQ  DFLPWKSLTS  IIEYYYMWKT  TDRYVQQKRL  KAAEAESKLK
(PRED) MmMTA1       YGKDFTDIQQ  DFLPWKSLTS  IIEYYYMWKT  TDRYVQQKRL  KAAEAESKLK
(PRED) ScaMTA1      YGKDFTDIQQ  DFLPWKSLTS  IIEYYYMWKT  TDRYVQQKRL  KAAEAESKLK
(PRED) CmMTA1       YGKDFSDIRQ  DFLPWKALTS  IVEYYYMWKT  TDRYVQQKRL  KAAEAESRLK
(PRED) HsMTA2       YGKDFNDIRQ  DFLPWKSLAS  IVQFYYMWKT  TDRYIQQKRL  KAAEADSKLK
(PRED) DrMTA3       YGKDFNDIRQ  DFLPWKSLTS  IIEYYYMWKT  TDRYVQQKRL  KAAEAESKLK
(PRED) DmMTA1       YGKDFNDIRQ  DFLPWKTLKQ  IIEYYYMWKT  TDRYVQQKRV  KAVEAELKLK
(PRED) HlMTA1       YGKDFSDIRQ  DFLPWKTLKN  VIEYYYMWKT  TDRYVQQKRV  KAVEAESKLK

                    .........460.........470.........480.........490.........500
(PRED) HsMTA1       QVYIPNYNKP  NPNQI-----  --------SV  NN--VKAGVV  NGTGAP----
(PRED) MmMTA1       QVYIPNYNKP  NPNQI-----  --------SA  SS--VKATVV  NGTGTP----
(PRED) ScaMTA1      QVYIPNYNKP  NPNQI-----  --------NV  NN--VKPGVV  NGTGVQ----
(PRED) CmMTA1       QVYIPNYNKP  NPNQI-----  --------NV  NSTVAKSAMV  NGAAVP----
(PRED) HsMTA2       QVYIPTYTKP  NPNQI-----  --------IS  VG--SKPG-M  NGAG------
(PRED) DrMTA3       QVYIPTYNKP  NPNQI-----  --------SV  SN--GKMATV  NGAAAG----
(PRED) DmMTA1       QVYIPQYNNN  GKGNG-----  ----------  TSTKAGGGIY  NGTT------
(PRED) HlMTA1       QVYIPNYNKT  APPTTAPSAA  TIVPLGNNNS  NSNGKPTNVL  NGNSNGNMTT

                    .........510.........520.........530.........540.........550
(PRED) HsMTA1       --G---QSPG  AGRACESCYT  TQSYQWYS--  ----------  ----------
(PRED) MmMTA1       --G---QSPG  AGRACESCYT  TQSYQWYS--  ----------  ----------
(PRED) ScaMTA1      --A---QNTG  AGRACESCYT  TQSYQWYS--  ----------  ----------
(PRED) CmMTA1       --G---QNGG  AGRACESCYT  TQSYQWYS--  ----------  ----------
(PRED) HsMTA2       --------FQ  KGLTCESCHT  TQSAQWYA--  ----------  ----------
(PRED) DrMTA3       --TGSFHTAG  GGRACESCFA  VQSAQWYS--  ----------  ----------
(PRED) DmMTA1       --NGSTDLSS  NGKPCESCGT  TKSSQWNSVS  SGHSTSRLCL  SCWEYWRRYG
(PRED) HlMTA1       DNSGILMVGV  SGKPCESCQV  MQSPQWY---  ----------  ----------

                    .........560.........570.........580.........590.........600
(PRED) HsMTA1       ---------W  GPPNMQCRLC  ASCWTYWKKY  GGLKMPTRLD  GE--------
(PRED) MmMTA1       ---------W  GPPNMQCRLC  ASCWTYWKKY  GGLKMPTRLD  GE--------
(PRED) ScaMTA1      ---------W  GPPNMQCRLC  ASCWTYWKKY  GGLKMPTRLD  GE--------
(PRED) CmMTA1       ---------W  GPPNMQCRLC  ATCWNYWKKY  GGLKMPTRLE  GE--------
(PRED) HsMTA2       ---------W  GPPNMQCRLC  ASCWIYWKKY  GGLKTPTQLE  GA--------
(PRED) DrMTA3       ---------W  GPPNMQCRLC  VSCWMYWKKY  GGLKMPSRAE  GAE-------
(PRED) DmMTA1       SMKSATKGDA  GEGDAKKKSS  SAASTPTATL  AGLATTPTAV  VDL-------
(PRED) HlMTA1       --------AW  GPSHMQCRLC  QSCWTYWKKY  GGLKVPSRMD  DVDLERKRGG

                    .........610.........620.........630.........640.........650
(PRED) HsMTA1       ----------  ----------  -RPGPNR---  ----SNMSPH  GLPA--RSS-
(PRED) MmMTA1       ----------  ----------  -RPGPNR---  ----NNMSPH  GIPA--RSS-
(PRED) ScaMTA1      ----------  ----------  -RPGPNR---  ----NNLSPH  GVPV--RNS-
(PRED) CmMTA1       ----------  ----------  -RPGPNR---  ----SSANPH  GYWV--KN--
(PRED) HsMTA2       ----------  ----------  -TRGTTEPHS  RGHLSRPEAQ  SLSPYTTSA-
(PRED) DrMTA3       ----------  ------EKTP  PSPAPNESRS  RGHCARQSSH  MVPI--RNS-
(PRED) DmMTA1       --NDDEKISD  LTNRQLHRCS  IVNCGKEFKL  KTHLARHYAQ  AHGIAISS--
(PRED) HlMTA1       TGSDEESKGI  GGAHRPHRCS  IPSCGKEFKL  KAHLSRHYAS  AHGVDLRGSG

                    .........660.........670.........680.........690.........700
(PRED) HsMTA1       -------GSP  KFAMKTRQAF  YLHTTKLTRI  ARRLCREILR  PWHAARHPYL
(PRED) MmMTA1       -------GSP  KFAMKTRQAF  YLHTTKLTRI  ARRLCREILR  PWHAARHPYM
(PRED) ScaMTA1      -------GSP  KFAMKTRQAF  YLHTTKLTRI  ARRLCRDILR  PWHAARHPYL
(PRED) CmMTA1       -------ANP  KFTLKTRQAF  CLQTTKLTKI  ARYFCRSLFH  TRQSARRPFT
(PRED) HsMTA2       -------NRA  KLLAKNRQTF  LLQTTKLTRL  ARRMCRDLLQ  PRRAARRPYA
(PRED) DrMTA3       -------GSP  KSSMKTKQAF  LLQATRLTKL  ARHMCRDLIR  LRRAARRPFV
(PRED) DmMTA1       -------GSP  RPIMKTRTAF  YLHTNPMTRV  ARAICRSIVK  PKKAARQSAY
```

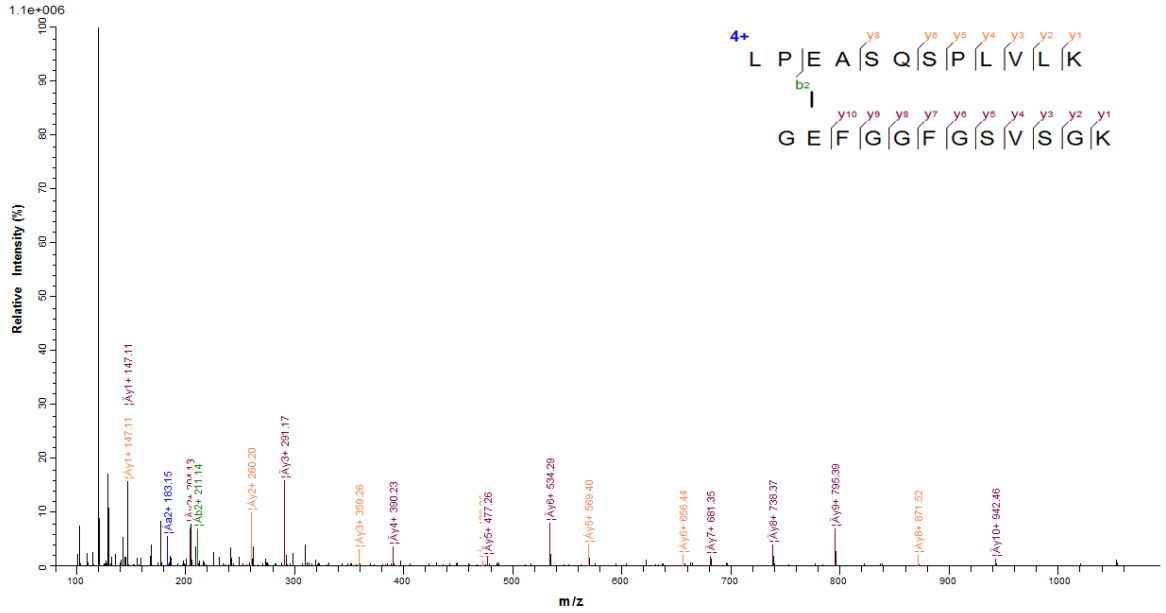
Results colour-coded for secondary structure

04/17/2016 09:03:38 AM

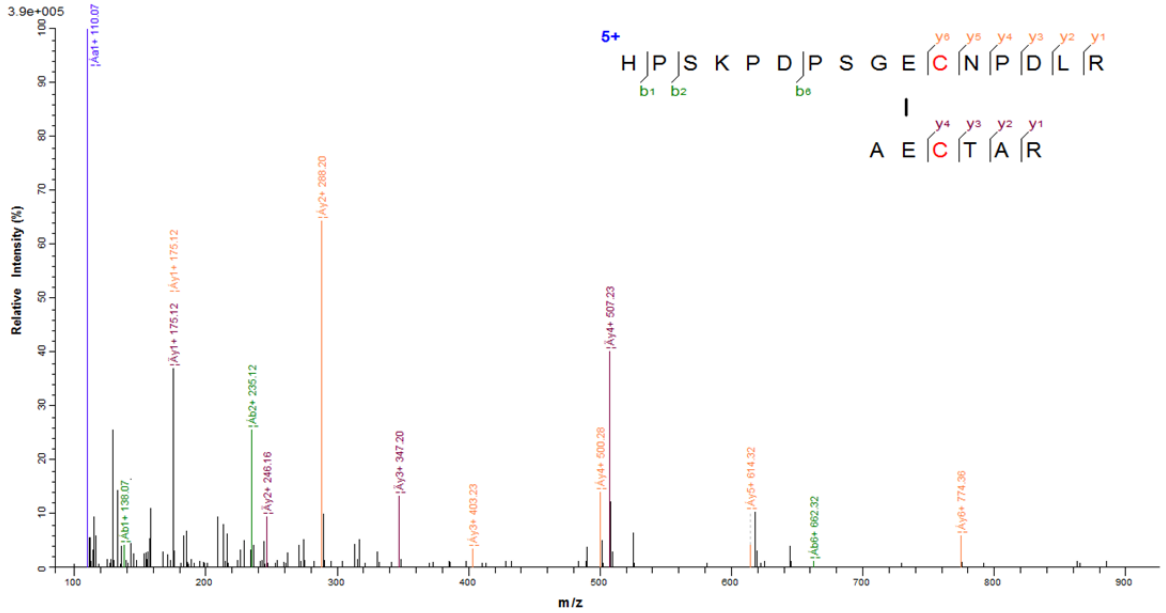
```
(PRED) HlMTA1 ASGGGGGGSP RPVMKTRSAF YLRTSALARA ARRLCAAQLR TRHAARAPHQ

                .........710.........720.........730.........740.........750
(PRED) HsMTA1  PINSAAIKAE CTA-RLPEAS Q-SPLVLKQA V-----RKPL EAVLRYLETH
(PRED) MmMTA1  PINSAAIKAE CTA-RLPEAS Q-SPLVLKQV V-----RKPL EAVLRYLETH
(PRED) ScaMTA1 PINSAAIKAE CTA-RLPEAS E-NPLLLKQV V-----RKPL EAVLRYLESH
(PRED) CmMTA1  PINSAAIKAE CTA-RLPEIG E-KPIKLKPS V-----RKTL ESIAKYLELH
(PRED) HsMTA2  PINANAIKAE CSI-RLPKAA K-TPLKIHPL V-----RLPL ATIVKDLVAQ
(PRED) DrMTA3  PINCGAIKAE YMI-RVSEGM TGRPLKPKSS P-----RSTL TSVLQYLESR
(PRED) DmMTA1  AINAMLVKQE FTN-RISGKS Q-AEIKKLLL LK-PKDRGSV TKIANRLGAP
(PRED) HlMTA1  PVNAAPLRHL CASPQLTSKS S-VELRILAR AVRPRPRPRV TDIATRLGDH

                .........760.........770.........780.........790.........800
(PRED) HsMTA1  P--------- ---------- -------RPP KPDPVKSV-- -------SSV
(PRED) MmMTA1  P--------- ---------- -------RPP KPDPVKSS-- -------SSV
(PRED) ScaMTA1 P--------- ---------- -------CPP KPDPAKSL-- -------SSS
(PRED) CmMTA1  P--------- ---------- --------SV KPICSLA--- --------VH
(PRED) HsMTA2  A--------- ---------- --------PL KPKTPRGT-- -------K--
(PRED) DrMTA3  P--------- ---------- --------AT HVQRPHRT-- -------PGL
(PRED) DmMTA1  G----SGPHE WLVLTPKDKM PLPAVVSFPK PPKAPDGSLV YDRVPNKSPD
(PRED) HlMTA1  PAPRQPGDWD WLALTAAAQR KQPDRVSFPR PPKAPDGSLL YERVPNKSEV

                .........810.........820.........830.........840.........850
(PRED) HsMTA1  LSSLTPAKVA PVINNGSPTI LGKR--SYEQ HNGVDGNMKK RLLMPSRGLA
(PRED) MmMTA1  LSSLTPAKSA PVINNGSPTI LGKR--SYEQ HNGVDGNMKK RLLMPSRGLA
(PRED) ScaMTA1 LNNLTPAKFT PVINNGSPTI LGKR--SYEQ HNGMDGNMKK RLLMPSRGLP
(PRED) CmMTA1  SGSVTPAKIL PVLNHGSPTI LGKR--GYEQ HNGIDGTMKK RLLMPFRGKT
(PRED) HsMTA2  -TPINRNQLS Q--NRGLGGI MVKR--AYET MAG------- ----------
(PRED) DrMTA3  QVQPPRRLLS SLPSHGPLGM LGKR--SYHH HSRVESAERR AGATG-----
(PRED) DmMTA1  VVAVPADKEL TIIPTQATST IRKRAHEDQQ LNGTEVTIVP SGPPAKRPNK
(PRED) HlMTA1  DRLTVTPPQP QPSMQAQQTI LKRTRPPFDE INGSDGIALS AGLPGGPPAK

                .........860.........870.........880.........890.........900
(PRED) HsMTA1  ---------- --NHGQARHM GP-SRNLLLN GKSYPTKVRL IRGGSLPPVK
(PRED) MmMTA1  ---------- --NHGQTRHM GP-SRNLLLN GKSYPTKVRL IRGGSLPPVK
(PRED) ScaMTA1 ---------- --NHGQTRQM GP-SRNLLLN GKSYPTKVRL IRGGSMPPVK
(PRED) CmMTA1  LAQSCSRLHS HPNHGQVRQM AAVNRGYMIN GKPYSRSSKI LKPGSLPPMK
(PRED) HsMTA2  ---------- ---------- A--GVPFSAN GRPLA---SG IRSSSQPAAK
(PRED) DrMTA3  ---------- ---QENPAHI V--GPILQHN GSSTG---GS SLRGSGLMLR
(PRED) DmMTA1  DPMPSHCPSP EQFAAMMAAS GQPLSRHHLN GKQ-----KI AQMARGGNGR
(PRED) HlMTA1  RAHHSQQLHP KHTLEHTAPT VLPLAPPLNG RAAHPHTLPH GPPLSRSNAR

                .........910.........920.........930.........940.........950
(PRED) HsMTA1  RRRMNWIDAP DDVFYMATEE TRKIRKLLSS SETKRAARRP YKPIALRQSQ
(PRED) MmMTA1  RRRMNWIDAP DDVFYMATEE TRKIRKLLSS SETKRAARRP YKPIALRQSQ
(PRED) ScaMTA1 RRRMNWIDAP DDVFYMATEE TRKIRKLLSS SEAKRAARRP YKPIILRP--
(PRED) CmMTA1  GRRMNWIDAP DDVLFIATEE TRKIRKLLGP VDMKRAARQP YKQIFIRQAR
(PRED) HsMTA2  RQKLNPADAP NPVVFVATKD TRALRKALTH LEMRRAARRP NLPLKVKPTL
(PRED) DrMTA3  KRRPNWIDAP DDSFFLVSRE TRKARRMLSR SQLRRACRQP CEQISLRRVP
(PRED) DmMTA1  KQVISWMDAP DDVYFRANDT HKKTRKILSA VDLRRAARKP WRTLPIKP--
(PRED) HlMTA1  KQVISWMDAP DDVYFRASDQ TKRLRKTLSS VELRRAARKP WRRLPAPL--

                .........960.........970.....
(PRED) HsMTA1  A----LPPRP PPP-APVNDE PIVIED
(PRED) MmMTA1  A----LPLRP PPP-APVNDE PIVIED
(PRED) ScaMTA1 ------MQAV QLR-QPMNDE PIIIED
(PRED) CmMTA1  V----SCAVS QPSRQPVNEE PIIIED
(PRED) HsMTA2  IAVRPPVPLP APSHPASTNE PIVLED
(PRED) DrMTA3  QGPSQVPILA PPHPSLRMRG PIVIHD
(PRED) DmMTA1  -------AAP EPS-SRPIES QIVILD
(PRED) HlMTA1  -------HPP HPQRAVRGDD MVVILD
```


Results colour-coded for secondary structure

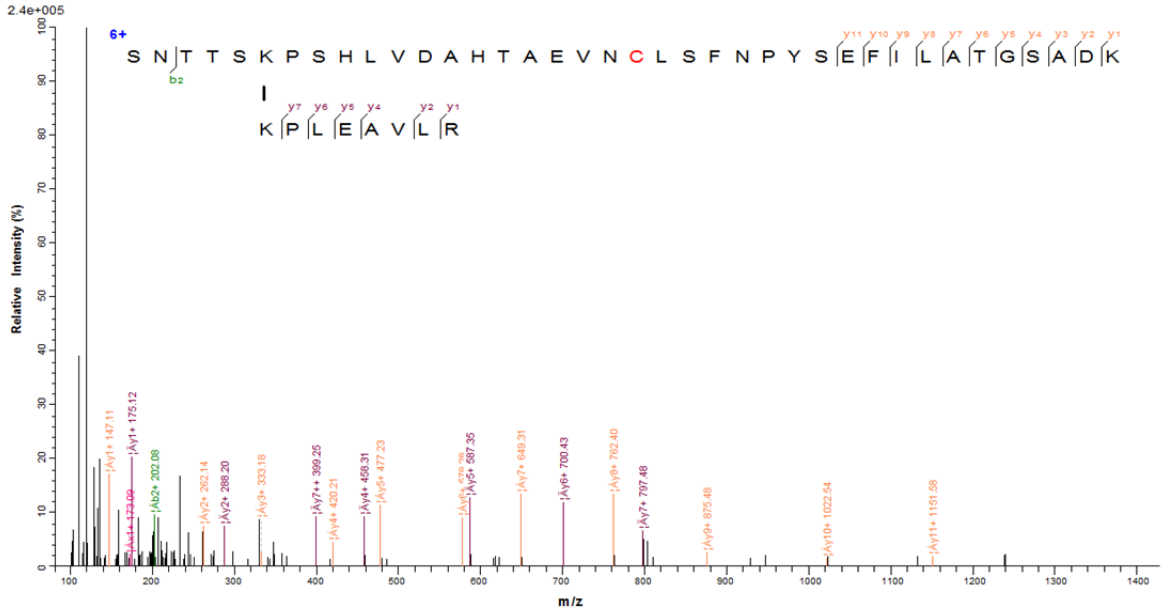
A1



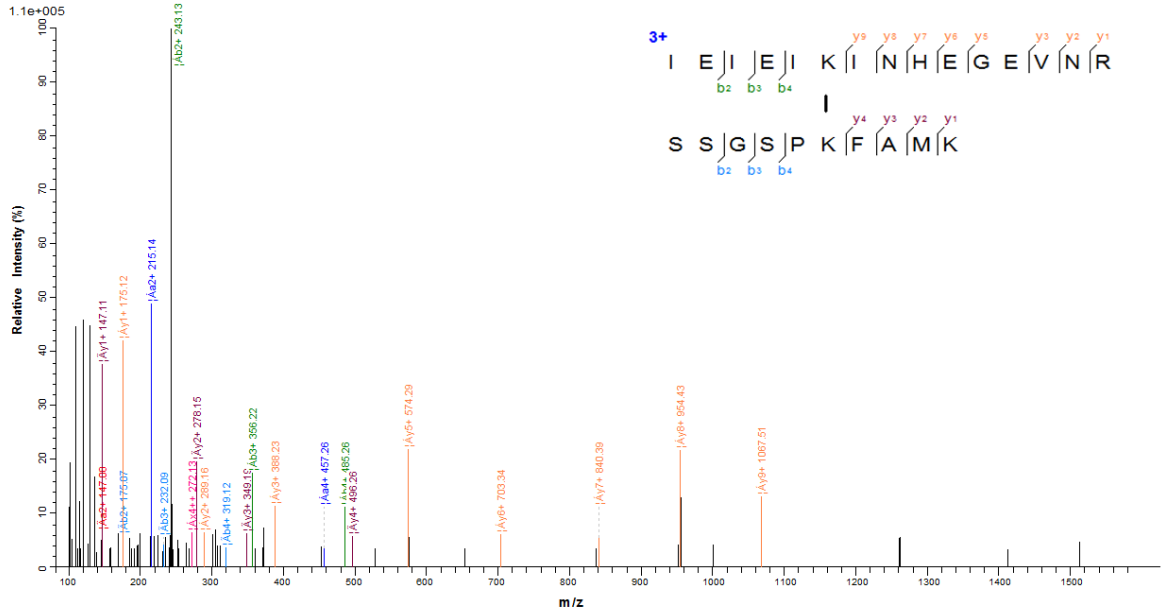
A2



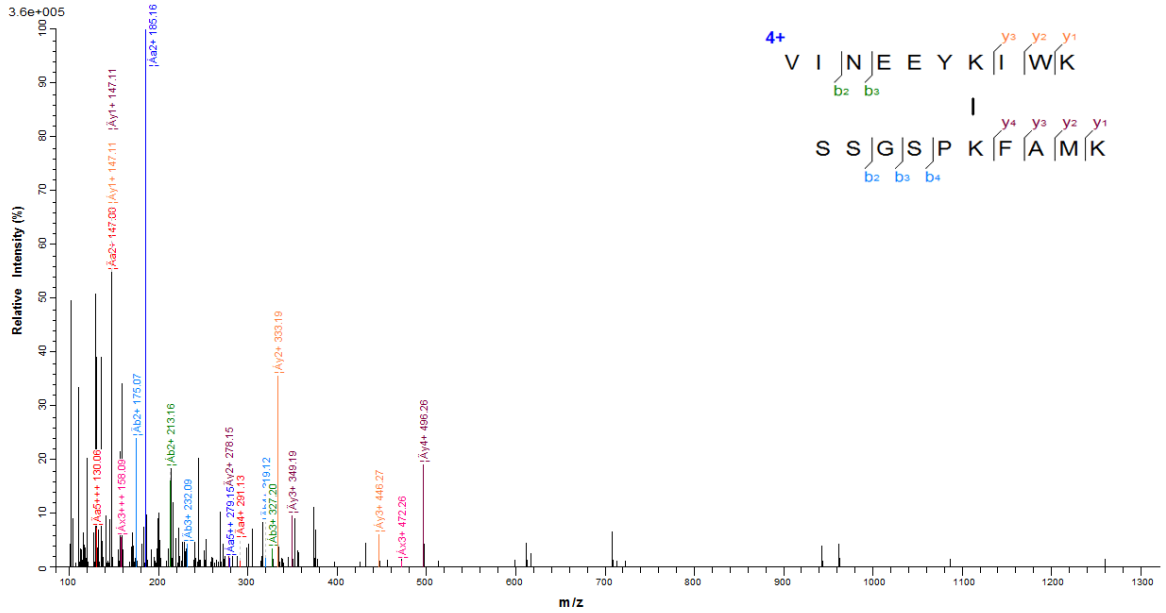
D1



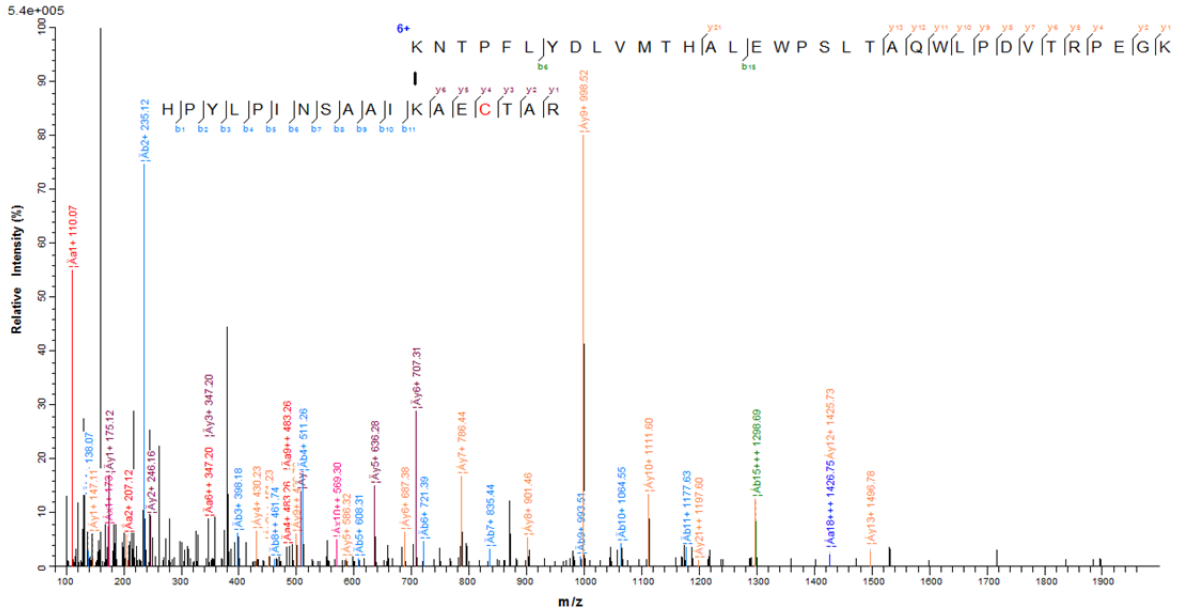
D2



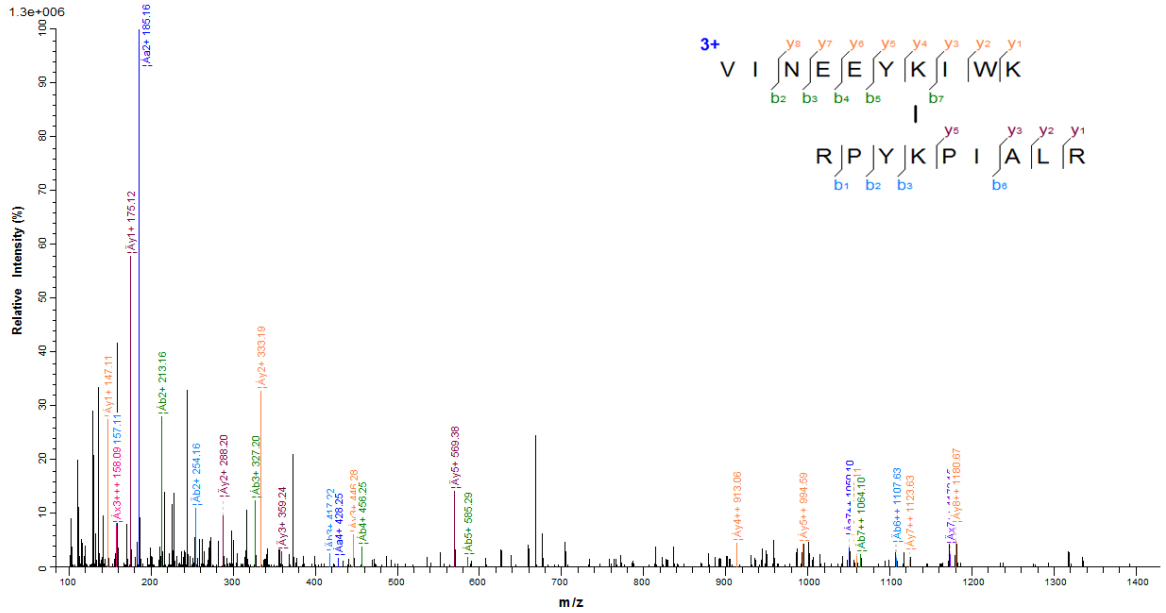
D3



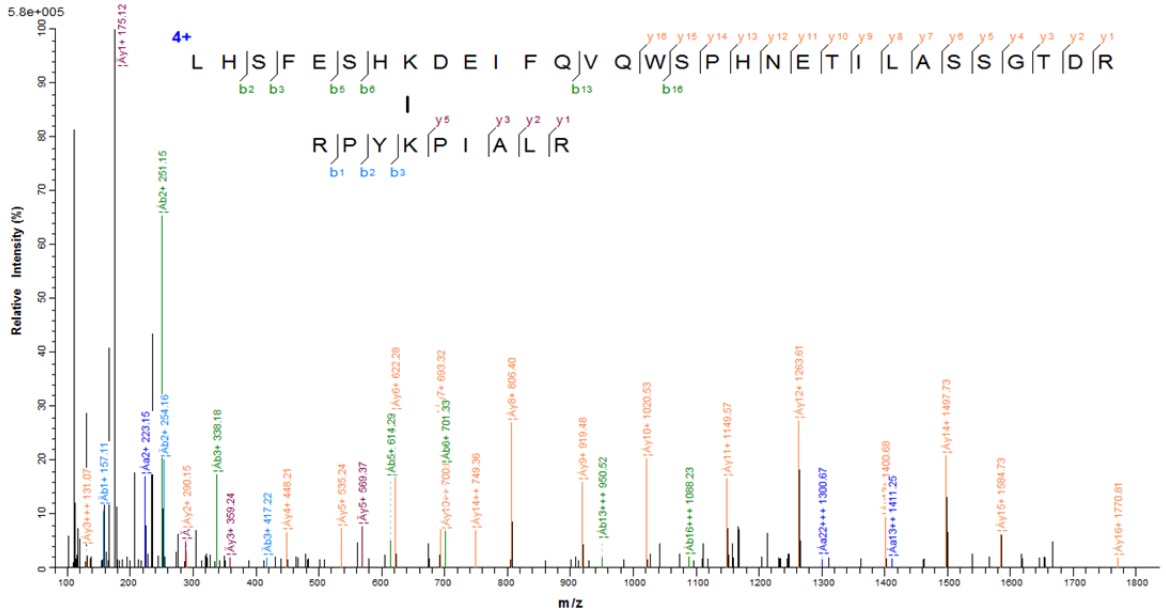
D4



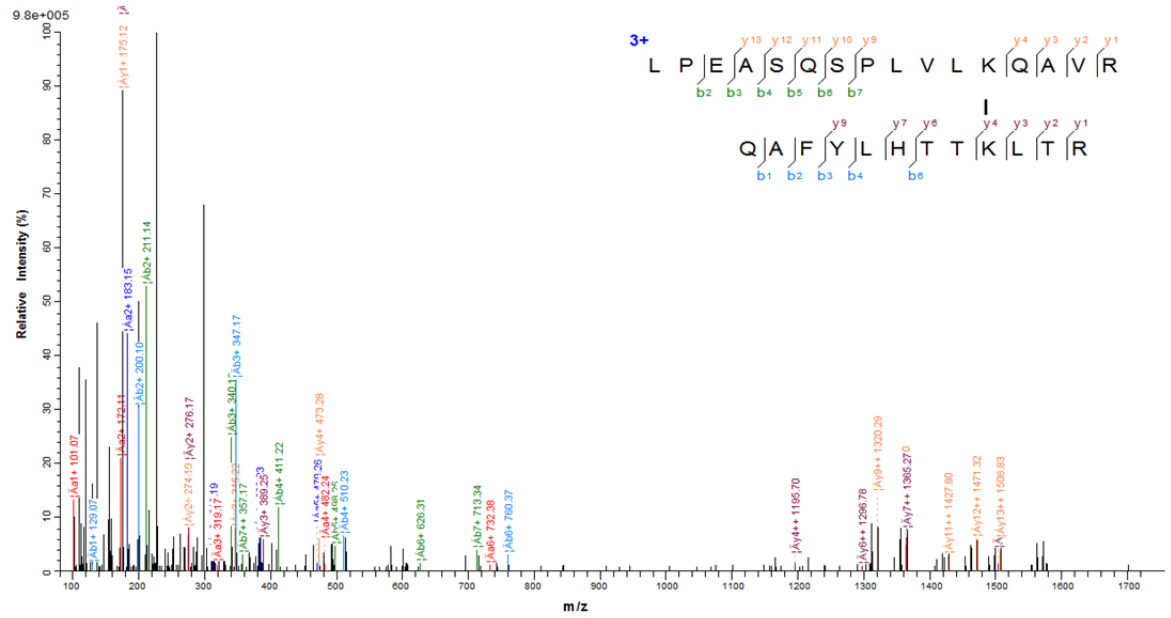
D5



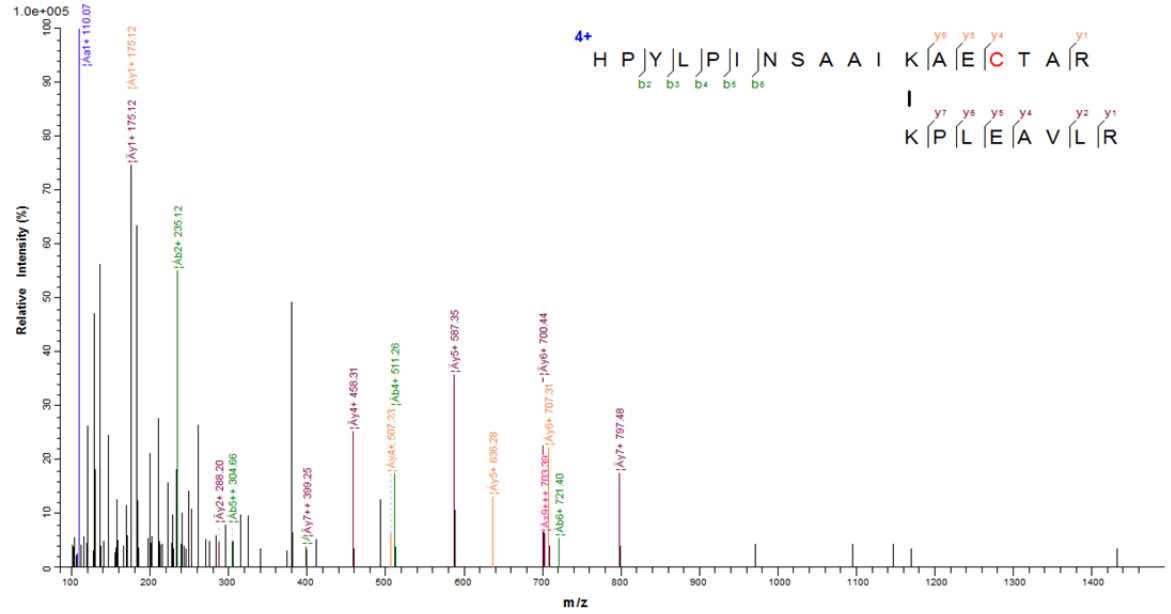
D6



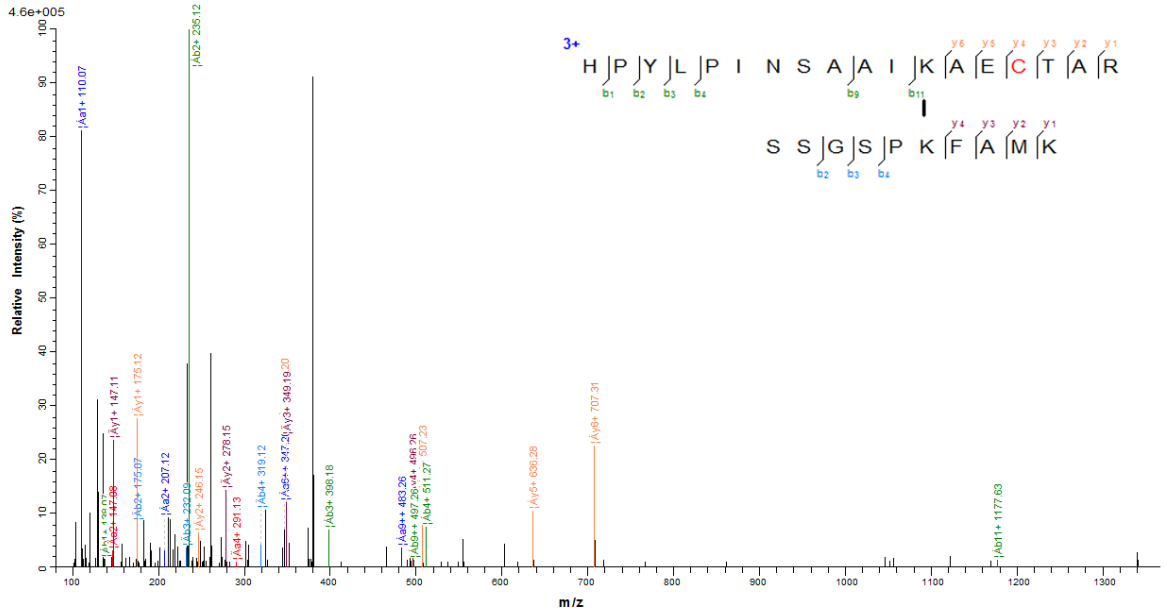
D7



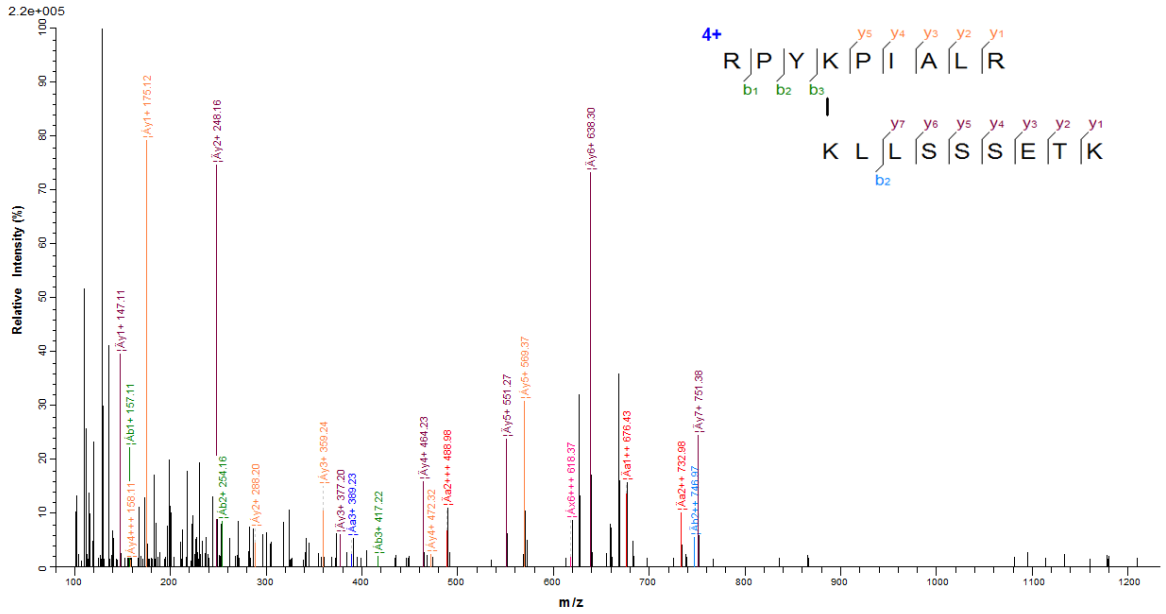
D8



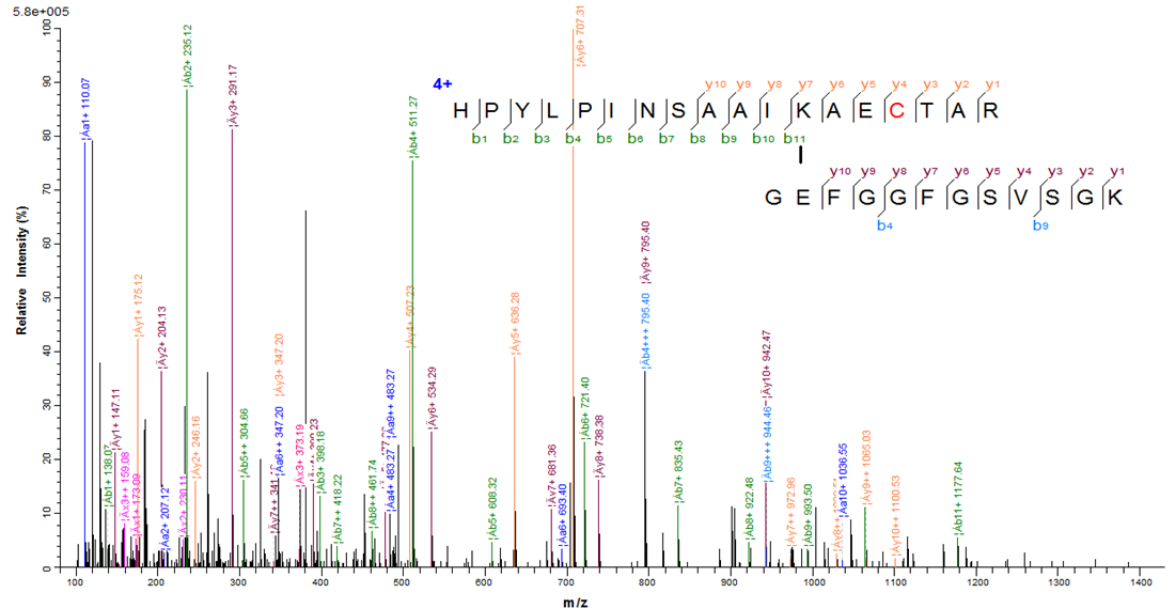
D9



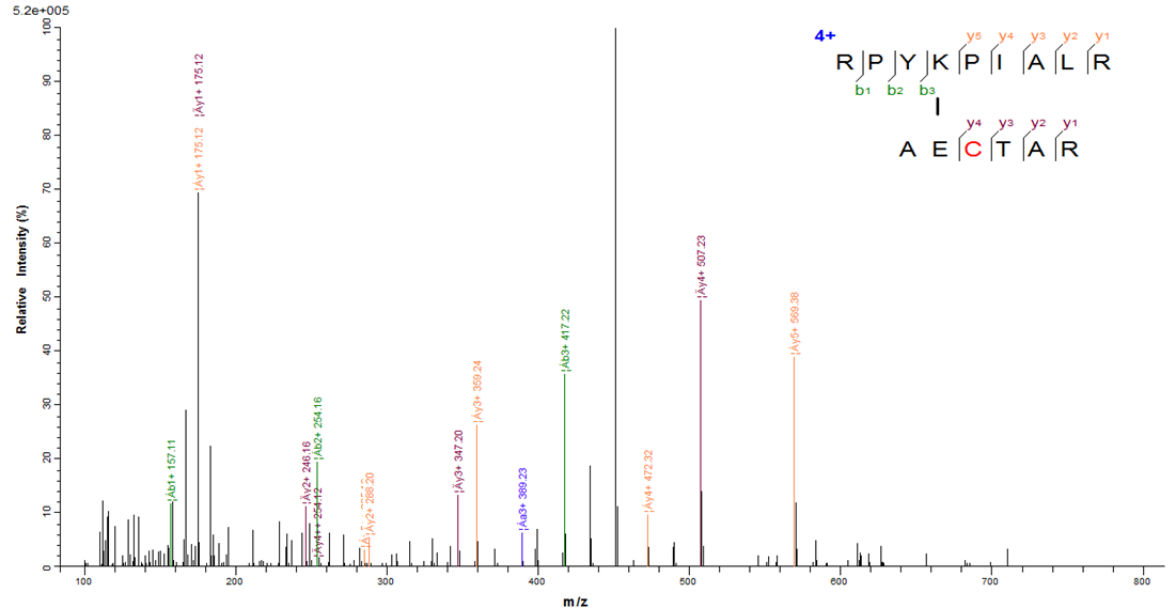
D10



Z1



Z2



Supplementary Figure 3. Fragmentation spectra for the crosslinks identified in this study that were used for the EM-model. The figure panels are labelled according to the 'Id' of the crosslink used in the EM model. Alpha peptide *y*- and *b*-ions are labelled in yellow and green, respectively, while the beta peptide *y*- and *b*-ions are labelled in russet and cyan, respectively. Cysteine residues coloured in red refers to them carrying the carbamidomethyl modification.