Capture Compounds for the methylome analysis

Christian Dalhoff,[a] Peter Poot,[a] Michael Hüben,[a] Hubert Köster,[b] and Elmar Weinhold,[a]

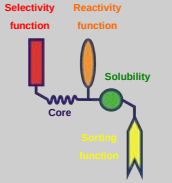
[a]*Institute of Organic Chemistry, RWTH Aachen University, D-52056 Aachen, Germany;* [b]*caprotec GmbH, Berlin, Germany*

Introduction

The isolation and identification of proteins from complex cell lysate is a challenging task in proteomics and a targeted isolation of enzyme families would be highly desirable. This requires to separate functional proteome subsets from the whole proteome in a form which allows identification by state-of-the-art bioanalytics and bioinformatics.
Separation by physical properties like molecular weight and charge can reduce the complexity of protein mixtures but lacks any selectivity concerning protein function. Affinity chromatography and chip technology achieve functional separations by immobilised selectivity functions but struggle from surface effects and steric hindrance. Additionally, the polymeric network has to be swollen in appropriate solvents, which have to be compatible with solubility properties of the proteins to be analysed. Since size and solubility properties of the targeted proteins are *a priori* unknown many proteins may remain undetected.

Capture Compounds

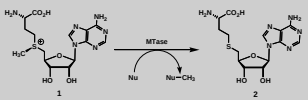
Capture Compounds (CC) are small molecules with three different functionalities. They comprise a *selectivity function* (red) for specific, reversible binding to targeted proteins, a *reactivity function* (orange) for covalent attachment to the proteins and a *sorting function* (yellow) for isolation of the captured proteins.



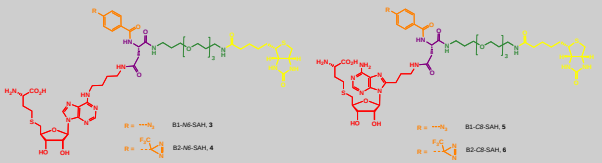
The general concept of the CC approach (A) reduces the proteome to a preprogrammed subset of proteins by initial affinity binding in solution (B) with subsequent covalent photo-crosslinking (C) followed by isolation using magnetic beads (D). Thus, the critical selective affinity interaction takes place in solution and is separated from the pull-out step involving immobilisation on a solid support. Either directly or after tryptic fragmentation the captured proteins (E) are identified and characterized by MALDI-MS (F), peptide map fingerprinting and/or MS-MS sequencing of the fragments followed by database queries.
We designed a CC scaffold bearing a position for facile attachment of the selectivity function, an aromatic azide for photoactivated crosslinking and biotin for pull-down. The modularity of the concept allows facile variation of the selectivity function to isolate different subsets of the proteome, such as drug-interacting proteins.[1]

Capturing the methylome

S-Adenosyl-L-methionine (SAM, **1**) is the major source of methyl groups in nature and hence utilized by a large number of methyltransferases (MTases) having different targets. The enzymatic reactions range from sequence-specific methylation of nucleic acids and proteins, methylation of small biomolecules, inorganic arsenite and halides to methylene transfer to unsaturated fatty acids. In all these transformations *S*-adenosyl-L-homocysteine (SAH, **2**) serves as the leaving group.

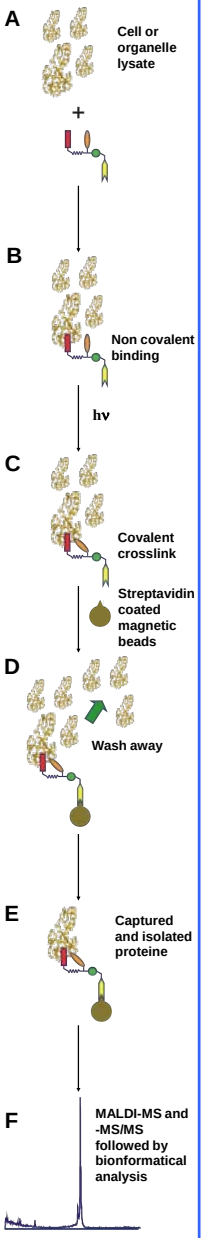


DNA methylation occurs throughout most living organisms. Mammalians use DNA methylation mainly as an epigenetic marker for repression of gene expression.[2] In addition to DNA methylation, modification of histone proteins plays also an important role in regulating gene expression by affecting the chromatin structure [3]. Methylation of RNA occurs either on the 2'-hydroxyl group of ribose or at the nucleobases albeit with more diversity than DNA methylation.[4] Methylation of small molecules is exemplified by the reactions catalyzed by catechol-*O*-MTase (COMT). Transfer of the methyl group of AdoMet to one of the phenolic hydroxyl groups of biologically active catechols, like dopamin or of its precursor l-Dopa, is the first step in their biological degradation.

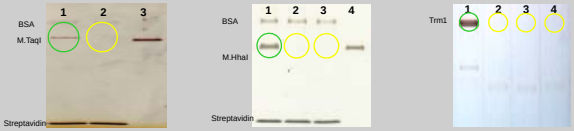


The entirety of enzymes transferring the methyl group of SAM, here named the methylome, is therefore an interesting target in proteomics. For the isolation of the methylome we used SAH derivatives as selectivity function. The sulfonium centre of SAM is relatively labile and thereby less favourable as selectivity function. The CC scaffold is attached via *N6* or *C8* of SAH as prior work showed, that modifications at these positions are tolerated by a large number of MTases.[5] The photocrosslink can be achieved by aromatic azides (B1) or aromatic diazirines (B2). We synthesized the CC s B1-*N6*-SAH (**3**), B2-*N6*-SAH (**4**) and B2-*C8*-SAH (**6**). The data presented here concentrates on B1-*N6*-SAH (**3**).

General concept



Purified methyltransferases

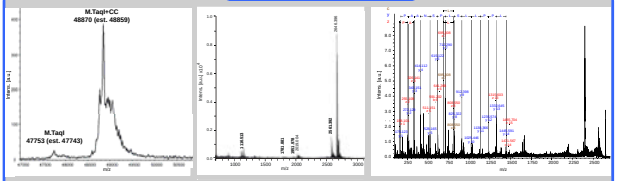


In a first step purified MTases were treated with CC and the capture event monitored by isolation using streptavidin coated magnetic beads followed by SDS-PAGE and silver staining.

Left: The adenosine N6-DNA-MTase M.TaqI was treated with CC in the absence (1) and presence (2) of 10 mM SAH as competitor. Lane 3 shows unmodified M.TaqI for comparison. M.TaqI is captured under standard conditions (green circle) but no capturing takes place if the CC is displaced by SAH before covalent crosslinking occurs. Showing that the covalent crosslink is guided by the non-covalent binding of M.TaqI to the SAH-moiety of the CC. Additional bands for streptavidin and BSA derived from the magnetic beads are detected. Middle: The cytosin C5-DNA-MTase M.HhaI was treated with CC in the absence (1) and presence (2) of 10 mM SAH as competitor yielding identical results as for M.TaqI. Without photocrosslinking no capturing is observed too (lane 3). Untreated M.HhaI is shown in lane 4 for comparison. Right: The RNA MTase Trm1 can be isolated with CC (lane 1). Control reactions without light activation (lane 2), in the presence of SAH (lane3) or using CC-scaffold without selectivity function (lane 4).

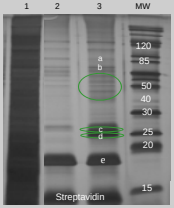
These results proof that the initial, non-covalent binding of the selectivity function is essential for the follow up photocrosslinking as displacement by a competitor (SAH) and absence of the selectivity function do not lead to crosslinking reaction. The isolation on magnetic beads allows for effectiv washing procedures without the risk of loosing isolated target protein. SDS-page analysis proofed practical for a first analysis of capture results.

MALDI identification



Captured proteins are identified by MALDI-mass spectrometry. Full length proteins or tryptic fragments can be eluted from magnetic beads using a 1:1 mixture of TFA and CH_3CN. As the CC is attached covalently to the enzyme, the mass value of the CC is added to the mass of the captured protein (left, sample protein M.TaqI). Tryptic fragmentation of captured proteins yields a fingerprint of the protein which can be used for direct identification (data not shown) or only fragments with directly attached CC are isolated yielding information about the position of the SAM-binding site (middle). Afterwards MALDI-MS/MS sequencing of tryptic fragments can be used to verify identification by fingerprinting (right). Combining the fingerprint of tryptic fragments and the sequence of several fragments like the N-terminal part of M.TaqI in this case, opens the door to identify new proteins which interact with SAM. The combination of these results with protein or DNA sequence databases can be achieved by bioinformatic tools like MASCOT.

E.coli lysates



First capture experiments in *E.coli* cell lysate demonstrate the ability of the CC approach to reduce the complexity of protein mixtures. Several proteins can be isolated from lysate using CC (lane 3). A control experiment with scaffold without selectivity function revealed a background consisting especially of highly expressed proteins. Lane 1 shows the composition of the starting material (0.5% of the lysate used in this experiment). Proteins isolated only in the presence of the selectivity function are marked green. MALDI analysis after tryptic digest identified following proteins: a)Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, b) Chaperone protein dnaK, c) Trans-aconitate 2-methyltransferase, d) 5'-methylthioadenosine nucleosidase
e) acetyl-CoA-carboxylase. The proteins c) and d) are SAM and SAH converting enzymes respectively, while a) and e) contain a lipoyl- or biotin-binding moiety and are thus isolated from *E.coli* lysate even in the absence of CC. Protein b) is known to bind ATP and other nucleotides.

Summary and Outlook

The capture compound approach showed its ability to isolate proteins with comparable function, in this case SAM dependent MTases. The crosslinking step is guided by the non-covalent interaction between selectivity function and target proteins in solution. We demonstrated the facile isolation and follow up MALDI characterization of proteins and a first decomplexation of cell lysate.
Ongoing tasks are the reduction of background reactions in cell lysate, the comparison of attachment positions and reactivity functions and the expansion to other selectivity functions, e.g. ATP, carbohydrates or other small biomolecules.

References

[1] Köster, H., Little, D. P., Luan, P., Müller, R., Siddiqi, S. M., Marappan, S. and Yip. P. *ASSAY and Drug Development Technologies*. **2007**, 5(3): 381-390. [2] C. Brenner, F. Fuks, *Curr. Top. Microbiol. Immunol.* **2006**, *301*, 45 - 66. [3] A. C. D'Alessio, Szyf, M., *Biochem. Cell Biol.* **2006**, *84*, 463 - 476. [4] G. R. Björk, J. U. Ericson, C. E. D. Gustafsson, T. G. Hagervall, Y. H. Jönsson, P. M. Wikström, *Annu. Rev. Biochem*. **1987**, *56*, 263 - 287. [5] Pljevaljcic, G., Schmidt, F. and Weinhold, E., *ChemBioChem* **2004**, *5(3)*, 265-269.