MycoSynVac

WP5 report of Deliverable

Deliverable D5.2

List of surface proteins in target species

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 634942.

Project Acronym: MycoSynVac

Project Title: Engineering of *Mycoplasma pneumoniae* as a broad-spectrum animal vaccine

Grant Agreement no.: 634942

Project Duration: 1 April 2015 - 31 March 2020

Work Package no.: WP5

Deliverable Number: D5.2

Deliverable Title: List of surface proteins in target species

Lead Beneficiary: Partners– CRG and MSD

Nature: R (Report)

Dissemination Level: CO (Confidential)

Deliv. Date (Annex I): 31 July 2016

Description: D5.2 List of surface proteins in target species

Version	Contributors	Comments
1.0	CRG	CRG first draft
2.0	CRG	Revised and improved
3.0	CRG	Final review and formatting

Table of Contents

1 Introduction

Mycoplasmas are the smallest free-living microorganisms. Their unique characteristic among the rest of bacteria is that they lack a cell wall. Therefore, they are resistant to many of the common antibiotics used to treat bacterial infections. They are parasites of a wide range of hosts from animals to humans. These obligatory parasites colonize mucosa, such as the respiratory and genital tracts and rarely penetrate in the bloodstream unless the host is weakened or immunosuppressed. Every year, infections caused by *Mycoplasmas* in poultry, cows and mainly pigs, result in important economic losses in livestock. There is an urgent need for better vaccines against Mycoplasma pathogens to decrease the infection levels more efficiently, have better safety profile and more economical, than the current commercial inactivated or live attenuated vaccines.

At this moment there are three main commercial *Mycoplasma* vaccines used in European farms: *M. hyopneumoniae* (inactivated whole cell) for pigs, *M. gallisepticum and M. synoviae* (live attenuated) for poultry (chicken and turkey). Also, there are some other mycoplasma vaccines that are used throughout the world. Historically, the first one is against contagious bovine pleuropneumonia (CBPP) with attenuated *M. mycoides* subsp. *mycoides* (strains T1/44 or T1sr, used in Africa, KH3J used in Australia or others). Vaccines that protect against *M. hyopneumoniae* are based on inactivated whole cells in a strong adjuvant. The vaccines are applied parenterally and are thought to produce (systemic) antibodies that prevent or limit adherence and colonization of the pathogen. However, as *M. hyopneumoniae* is mainly found at the mucosal surface of the trachea, bronchi, and bronchioles, and adherence to the ciliated epithelium is a prerequisite for initiation of the infection, a mucosal antibody response is considered important to prevent colonization and the excessive (cell-mediated) immune response upon infection, which is part of the pathogenesis. However, the cell-mediated immunity is important to prevent systemic spread of *M. hyopneumoniae* within the pig. Therefore, an attenuated vaccine administered locally may be best suited to provide a strong, long lasting mucosal humoral response and a proper cell-mediated immune response.

Vaccines for *M. gallisepticum* and *M. synoviae* are obtained from attenuated bacterial strains grown, mixed and sold as single and a combination vaccine. The issue here is that the vaccines strains were isolated in the 80's after continuous passages and tested to be attenuated in poultry. It has been recently reported that the temperature-sensitive phenotype of the *M. synoviae* vaccine strain is due to point mutation(s) in the *obgE* gene (Shahid, Markham et al. 2013). This could explain the possible reversion to virulence after vaccinating chicken as some customers have claimed. The nature of the attenuation of the various *M. gallisepticum* strains appears to be more complex i.e. the genome comparison of one of the vaccine strain

(F strain) with its parental strain indicated that they are several sites that have been changed during serial passage (Szczepanek, Tulman et al. 2010). Aside from this, vaccines have to be cheap and therefore increased yields in production would be a benefit for the animal industry.

Identification of proteins involved in immunomodulation and development of infection-associated inflammatory diseases is very important for vaccine design, as well as, for diagnosis. One-dimensional SDS-PAGE and two-dimensional gel electrophoresis are the most common used techniques (expensive, tedious and not high throughput). Also, membrane proteome maps have been reconstructed by immunoblotting in several Mycoplasma species using whole cell extracts, as well as Triton X-114 extraction (to enrich amphiphilic proteins including putative lipoproteins of the terminal organelle). For example, in *M.fermentans*, a total of 181proteins have been identified. Further bioinformatics analysis of these ORFs encoding proteins, with known or so far, unknown orthologues among bacteria, revealed that a total of 131 proteins are homologous to known proteins, 11 proteins are conserved hypothetical proteins, and the remaining 39 proteins are likely *M.fermentans*-specific proteins. Moreover, Triton X-114-enriched fraction was shown to activate NF-kB activity of raw 264.7 macrophages and 21 lipoproteins with predicted signal peptide were identified therefrom. It was the first proteome reference map of *M.fermentans*, as well as, several putative virulence-associated proteins were identified as diagnostic markers or vaccine candidates for further functional study of this human pathogen. The inconvenient of this technique is that only the proteins that generate immune response are identified and does not consider the total cell extract. Thus, exposed proteins that are not antigenic naturally but they could elicit a protective immune response if exposed in large amounts. In addition, since this technique involves immunoprecipitation, it is very tedious to identify enriched surface proteins.

Other techniques using fluorescent marked reagents have been used for the identification of exposed proteins by proteomics. The inconvenient of these approaches is that these molecules usually enter in the cell, labelling cytosolic proteins and could result in significant background, preventing the complete identification of exposed proteins. Tryptic digestion has been also used to study the surfomes by “Shaving” the cells. However, often it results in cell lysis and therefore in a strong background.

The overarching objective of this work is to engineer *M. pneumoniae* as a chassis to be used to display antigens corresponding to other Mycoplasma species and/or viruses to immunize poultry and pigs, as well as other farm animals and thus it could be used as universal vaccine. Herein, we have developed the technique MemBioID, to define the membrane proteins in different Mycoplasma species. This technique uses the biotin ligase protein (BirA) that activates biotin added to the medium to form biotinyl 5'-adenylate that reacts with nearby solvent exposed Lysine residues in proteins. Biotinylated proteins are recovered by

affinity purification using streptavidin resin. Biotinylation is done in intact cells, which should only label solvent exposed proteins, and total protein extracts which should label all proteins. As controls, we purified membrane proteins and prepared total protein extracts without biotinylation. All samples are the trypsin digested (cuts at Lys or Arg residues) and biotinylated peptides should not be cut. Comparison of biotinylated and non-biotinylated samples and membrane purified proteins and total extracts, allow identification of membrane proteins as well as peptides accessible to biotinylation in membrane proteins. Thus, we could in principle identify solvent exposed aa sequences of membrane proteins. Additionally, this technique can be combined with immunoblot and immunoprecipitation to identify surface exposed proteins that are recognized by immune system.

This Deliverable was scheduled to be submitted on July 2016 but it have suffered some delay due to some technical issues: To identify the proteins exposed in the Surface of Mycoplasma pneumoniae (surfome), CRG had the idea of using the biotin ligase protein (BirA) to biotinylate the proteins at the surface and purify them by affinity. As explained in the kick off, CRG realized that the BirA protein was internalized in *M. pneumoniae* cells and all the proteins were labelled without observing a clear enrichment in the surface proteins. Dr. Bijlsma from MSD proposed to bind the BirA protein to beads (to increase the size of the molecule and avoiding the entrance of the protein). Then, MSD was in charge to continue developing the protocol. CRG sent them BirA protein and they stablished the protocol to identify the surface proteins of *M. bovis*. They did the affinity purification of surface proteins and observed a clear difference between the total fraction and the surface proteins. MSD sent to CRG the samples to do Mass Spectrometry to identify the exposed proteins of *M. bovis*. The results and analysis of data (including required replicates) required 6 more months. In parallel CRG was developing the system to expose proteins and peptides in the surface of the chassis. A preliminary list of 4 antigens (P97, PCV2, GP40, flaggelin), were described as effective antigens to generate immune response by MSD, which has been used to generate the first versions of the vaccines. The efficiency of *M. pneumoniae* to expose these proteins will be tested in a near future. Thanks to the fact we have already identified some surface proteins (4 antigens), this delay does not involve the alteration of other WP, or affect the work in other aspect of the project. On the other hand, it will give us the chance to prepare a more detailed list of the surface proteins to be included in the chassis.

2 Results

2.1 Description of samples used

Samples were obtained by MSD using a protocol discussed with CRG. The BirA protein and idea of the use of the labelling to mark with biotin exposed proteins was provided by CRG. The idea of improving the protocol by using BirA coated with beads, as well as, including a step of membrane extraction in the protocol was suggested by MSD. Samples were obtained by MSD and sent to CRG to be analysed by MS. The MS raw data (Table S1 and S2) was analysed by CRG and CRG has stablished the statistical test and analysis to classify membrane proteins and to identify possible exposed peptides.

Samples send to MS at the CRG Proteomics Facility (Request:20161130143413).

#	Strain	Sample	Concentration (mg/ml)	Biotinylated proteins	Negative control to:	Buffer
1	*M.bovis* JF4278	Biotinylated proteins	0,8	Yes		0.01M PBS + 0.1% SDS
2		Proteins	8,18	No	Reaction 1	6M urea + 8% triton
3		Biotinylated membrane isolation - unbound proteins	7,75	Yes		2.3% triton X114 in 0.01M PBS
4		Biotinylated membrane isolation - elution	1	Yes		0.01M PBS + 0.1% SDS
5		Membrane isolation - membrane proteins	9	No	Reaction 4	2.3% triton X114 in 0.01M PBS
6		Lysate biotinylated proteins	0,87	Yes		0.01M PBS + 0.1% SDS
7		Lysate proteins	39,31	No	Reaction 6	Reaction buffer (CRG)
8		Biotinylated proteins with BirA-linked to beads	0,46	Yes		0.01M PBS + 0.1% SDS
9	*M.bovis* PG45	Biotinylated proteins	0,86	Yes		0.01M PBS + 0.1% SDS
10		Proteins	20,98	No	Reaction 9	6M urea + 8% triton
11		Biotinylated membrane isolation - unbound proteins	76,46	Yes		2.3% triton X114 in 0.01M PBS
12		Biotinylated membrane isolation - elution	0,89	Yes		0.01M PBS + 0.1% SDS
13		Membrane isolation - membrane proteins	104,08	No	Reaction 12	2.3% triton X114 in 0.01M PBS
14		Lysate biotinylated proteins	0,76	Yes		0.01M PBS + 0.1% SDS
15		Lysate proteins	19,97	No	Reaction 14	Reaction buffer (CRG)
16		Biotinylated proteins with BirA-linked to beads	0,57	Yes		0.01M PBS + 0.1% SDS
17	*M.hyopneumoniae*	Biotinylated proteins	0,17	Yes		0.01M PBS + 0.1% SDS
18		Proteins	23	No	Reaction 17	6M urea + 8% triton

#	Strain	Sample	Concentration (mg/ml)	Biotinylated proteins	Negative control to:	Buffer
19		Biotinylated membrane isolation - unbound proteins	6,66	Yes		2.3% triton X114 in 0.01M PBS
20		Biotinylated membrane isolation - elution	1,69	Yes		0.01M PBS + 0.1% SDS
21		Membrane isolation - membrane proteins	99,95	No	Reaction 20	2.3% triton X114 in 0.01M PBS
22		Lysate biotinylated proteins	18,97	Yes		0.01M PBS + 0.1% SDS
23		Lysate proteins	23,58	No	Reaction 22	Reaction buffer (CRG)

Table 1. Description of the samples provided by MSD

Samples names in MS facility are: 161130 S TFLS 01 to 23

Description of samples in Table 1:

Biotinylated membrane isolation-elution: 161130_S_TFLS_12_03_5pto
Cells (intact) were biotinylated with BirA, then, membrane proteins were isolated and the biotinylated proteins purified with streptavidin. We collected the non-bound and bound fractions.

Biotinylated membrane isolation-unbound proteins==corresponds to the proteins not bound by streptavidin (161130_S_TFLS_11_03_5pto)

Membrane isolation-membrane proteins: this fraction is obtained from a separate reaction without biotinylation after membrane purification (negative control 161130_S_TFLS_13_03_5pto)

Biotinylated proteins: total cells (not lysed) were incubated with BirA, and the biotinylated proteins purified with streptavidin after breaking the cells. This sample is comparable with sample 161130_S_TFLS_13_03_5pto, only this reaction/experiment was done without membrane isolation with triton. (161130_S_TFLS_09_04_5pto)

Proteins=negative control to sample of biotinylated proteins (161130_S_TFLS_10_03_5pto)

Biotinylated proteins with BirA-linked to beads: different experiment than 161130_S_TFLS_13_03_5pto) because the BirA enzyme was linked to beads.

Lysate biotinylated proteins: The cells that were used in this reaction were lysed, so cytoplasmic proteins and membrane proteins were all biotinylated (161130_S_TFLS_14_03_5pto)

Lysate proteins: Cells were lysed in this reaction and no purification was done.

Note: Samples Lysate and Lysate Biotinylated proteins:
These samples have been compared to address if biotinylation is affecting the detection and quantification of some proteins. For this experiment, we incubated a total cell extract with the BirA enzyme and the two samples treated and not treated derived from the same extract were analyzed by MS. The number of proteins detected in the fraction without BirA is lower than in the labelled sample. The sum of the areas of detected proteins revealed that the non-label sample has 16 times less protein: This explains the smaller number of detected proteins. Therefore, the impact of biotin labelling in the number of detected proteins cannot be assessed.
Samples prepared with BirA bound to beads or without beads: we observe a difference in the total protein amount of each sample. We did not observe significant differences regarding the number of proteins detected. Differences are probably due to different amounts of total protein injected.

2.2 Pipeline for data analysis

First, we want to find the ratio and the threshold that allow us to define if a protein detected by MS is enriched in the samples that correspond to purified membrane proteins. Data of *M. bovis* PG45 is used to stablish and define the best protocol to analyze the MS results.
Of *M. bovis* PG45 experiment, two **different experiments** are considered for this study:

1. Biotinylated membrane isolation: elution sample (161130_S_TFLS_12_03_5pto) and as reference total lysate biotinylated (161130_S_TFLS_14_03_5pto)
2. Membrane isolation (no biotinylated): Membrane sample (161130_S_TFLS_13_03_5pto) and as reference, protein extract passed through a streptadivin column with no biotin labeling(161130_S_TFLS_10_03_5pto)

Note: Areas values have been corrected. Detection: We know that peptides with areas below 10^5 are not detected by mass spectroscopy. Thus when we don't see a protein the area could be between 0 and 10^5. Thus, to be conservative, we assigned a 10^5 value to a protein when it has been found in one of the two samples being compared and not in the other. If a protein has not been found in both samples then it is discarded for the analysis.

Different ratios are calculated:

-R1=Rm-Rt

Rm=ratio membrane proteins= Area of protein in membrane sample/Sum area of all detected proteins; Rt=Area of protein in lysated sample/Sum area of all detected proteins in the sample.

-R2=Am/At

Am=area of protein in membrane fraction; At=area of protein in lysated sample

-R3=Ratio of elution or membrane sample

ROC curves to establish the Threshold (Thr) to define if one protein is enriched in the membrane fraction

Two sets (true negatives and positives) have been stablished:

1. Sets stablished by functional annotation at NCBI (MLS)

   False positives: this set comprises the ribosomal proteins detected in the reference sample. The non-detected in the reference sample are discarded because it could be that they are not detected due to their small size and abundance of Lys, Arg residues. Total number of proteins is 48.

   True negatives: this set comprises 55 membrane proteins detected in the reference sample or in the elution or membrane samples.

   Sets (LS) stablished by looking at the characteristic signature of lipoproteins at the N-terminus of proteins and in some cases the prediction of transmembrane segments for prokaryotes using http://www.sbc.su.se/~miklos/DAS/

Finally, we used the predicted LS set to evaluate which is the percentage of proteins that are properly classified as membrane or cytosolic (% of agreement with LS in Tables 2 and Table3).

ROC	AUC	Thr	Max of substraction of ratios true and false positive	Nr Membrane Proteins	Nr Cytosol Proteins	% of Agreement with LS
R1_MLS	0.9767892	-0,00108095	0.85725338	472	112	49,4863
R1_LS	0.8158467	0,000040171	0.5383638623	241	343	75,85616
R2_MLS	0.9396518	0,666666667	0.83636364	280	304	72,94521
R2_LS	0.819933	0,6315789	0.536094451	289	295	72,77397
R3_MLS	0.5062126	0,001519555	0.1633686148	216	368	54,96575
R3_LS	0.6659723	0,000709126	0.291160189	292	292	54,28082

Table 2. Summary of the results of the different analysis by ROC for the Biotinylated membrane sample. AUC, is the area under the curve; Thr is the threshold defined by the ROC analysis defined in column1 (ROC).

As it is shown shadowed in yellow in Table 2, R1_LS is the analysis that gives the higher percentage of agreement with the expected classification of membrane and cytosolic proteins. The corresponding roc curve graph is shown in Figure 1.

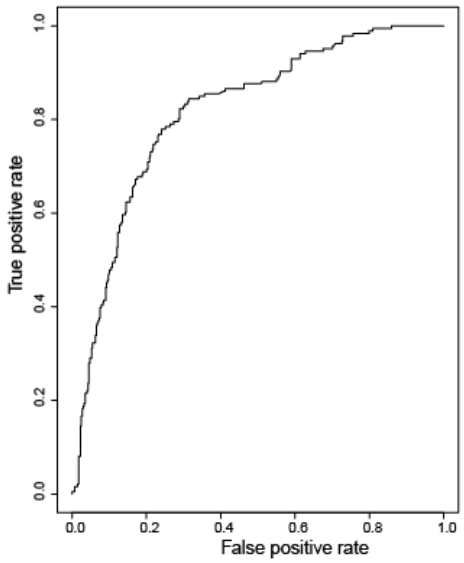

Figure 1. ROC R1_LS for biotinylated membrane sample

The same analysis was performed for the Membrane proteins experiment and the summary of the results of ROC curve are shown in Table 3.

	AUC	Thr	Max of substraction of ratios true and false positive	Nr Membrane Proteins	Nr Cystosol Proteins	% of Agreement with LS
R1_MLS	0.9860884	-0,001995998	0.87234043	538	86	46
R1_LS	0.845693	0,000030053	0.6446214000	173	451	87
R2_MLS	0,9815876	0,2207792	0.88052373	339	285	73
R2_LS	0.8678416	0,6705882	0.6380411978	179	445	87
R3_MLS	0.5605993	0,000842905	0.248211091	216	408	66
R3_LS	0.7323774	0,000365259	0.3702214162	317	307	62

Table 3. Summary of the results of the different analysis by ROC for the Membrane sample. AUC, is the area under the curve; Thr is the threshold defined by the ROC analysis defined in first column (ROC).

Also, in this case the analysis R1_LS is the one that gives higher percentage of agreement with the list of predicted membrane proteins. The graph of corresponding ROC curve is shown in Figure 2.

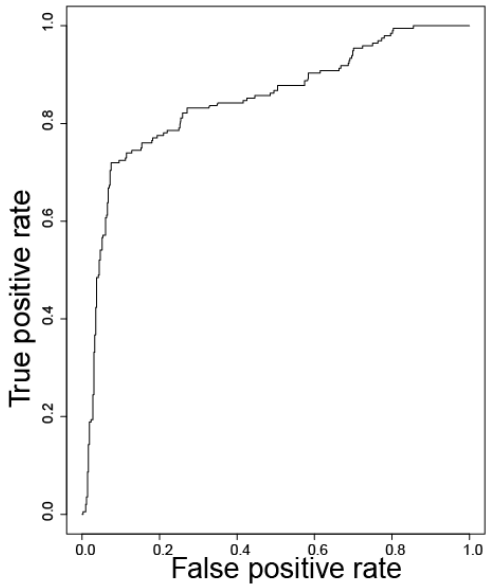

Figure 2. ROC R1_LS for membrane sample

Proteins without a clear cytosol membrane assignment based on their function were run on the membrane prediction server http://www.sbc.su.se/~miklos/DAS/ in those cases in which the server predicted a clear trans-membrane segment we assigned the corresponding proteins to membrane, if not to cytosol. Once we have done the ROC analysis and decide on a threshold for classification (R1_LS), we run on the membrane prediction server on all those proteins that were incorrectly assigned by the cutoff value from the ROC curve

analysis and in some cases the cytosol/membrane assignment was changed. The updated assignment for the prediction is (Table S3). We have repeated the ROC curves for the R1_LS2 analysis by using this classification (LS2) for the true positive and negative set. The results of this analysis are shown in Table 4 (Supplementary excel A1 to A6).

Analysis name	AUC (confidence ratio)	Thereshold	Max of substraction of ratios true and false positive	Nr Membrane Proteins	Nr Cystosol Proteins	% LS2 Class
R1_LS_Biotin	0.8158467	0,000040171	0.5383638623	241	343	75,85616
R1_LS2_Biotin	0.8614103	4,01714E-05	0.618268575	240	344	79,40141
R1_LS_M	0.845693	0,000030053	0.6446214000	173	451	87
R1_LS2_M	0.933066	3,00534E-05	0,806556604	173	451	91,77632

Table 4. Comparison of ROC curves obtained in the analysis R1_LS and R1_LS2 for both experiments: with birA (Biotin) and membrane extraction (M)

By using this new set of true positives and negatives we have improved the AUC in the ROC curves of both experiments and also the % of classification is better adjusted. In table R1_LS2_Analysis are collected all the information and data of the analysis.

We repeat the pipeline for *M. hyopneumoniae*. For this analysis the two **different samples** used for this study are:

1. Biotinylated membrane isolation: elution sample (161130_S_TFLS_20_04_5pto) and total biotinylated proteins (161130_S_TFLS_17_04)
2. Membrane isolation (no biotinylated): Membrane sample (161130_S_TFLS_21_04_5pto_21_21B) and protein extract (161130_S_TFLS_18_04_5pto)

Note: Areas values have been corrected by detection: We know that peptides with areas below 10^5 are not detected by mass spectroscopy. Thus when we don't see a protein the area could be between 0 and 10^5. Thus, to be conservative, we assigned a 100000 value to a protein when it has been found in the sample with which it is compared. If a protein has not been found in both samples then it is discarded for these analysis.

As it was previously shown for the report of *M. bovis*, R1 ratio and LS sets for ROC curves are the parameters that allow better classification of membrane proteins by using ROC (Table 5).

	AUC	Thr	Max of substraction of ratios true and false positive	Nr Membrane Proteins	Nr Cystosol Proteins	% of Agreement with LS
R1_LS_M	0.803	-1,16E-05	0.508	111	401	82,6
R1_LS_BioM	0.806	3,12E-06	0.521	112	400	83,4

Table 5. Results of ROC curves analysis for the experiments of *M. hyopneumoniae* membrane and biotinylated membrane, using ration R1 and set LS for the membrane proteins classification.

2.3 Identification of *M. bovis* PG45 membrane proteins

➢ Membrane proteins predicted by analysis R1_LS2 of biotin and membrane samples

We have identified 240 and 173 putative membrane proteins in the samples of biotinylated membrane sample and membrane experiments, respectively. The number of cytosolic proteins are 344 and 451 for both experiments, respectively. When this classification is compared with the LS2 prediction we find that 79% and 91% of the proteins are well classified for the biotin and membrane experiments, respectively. Interestingly, 60% of the identified membrane proteins in the experiment with biotin are also classified as membrane proteins in LS2. In contrast, in the experiment of membrane this percentage is 87%. This result suggests that the membrane experiment without biotin allows a better identification of membrane proteins than the experiment with biotin.

The number of proteins common to both experiment is 139 (Table S4). A total of 90,1% of those proteins were classified as membrane proteins by LS2 study. Interestingly, 6 out 13 proteins assigned as cytosolic in LS but classified as membrane in both experiments are hypothetical proteins, suggesting a new role for those proteins as membrane components.

We have found 102 proteins classified as membrane proteins in the experiment of biotinylated membrane that do not appear in the experiment of membrane (Table S4). Only 19% of those proteins were assigned as membrane proteins in LS2 assignment. The reason is that several membrane proteins are not enriched in the biotinylated sample and this results in a lower cutoff value in the ROC curve that as a result allows cytosolic proteins to pass the cutoff. In total there are 34 proteins that are present in the experiment of the membrane that are not present in the biotinylation experiment (Table S4). In this case 76% of the proteins are assigned as membrane proteins in LS2 study.

As a summary, it looks like biotinylation results in less membrane proteins detected probably because they are not accessible to the BirA enzyme and therefore they are not good targets for vaccination.

➢ Identification of putative antigens and exposed protein regions in *M. bovis* PG45

We have studied the unique peptides detected for the 139 proteins identified in both experiments. In the biotinylated membrane sample since biotinylation was done with the cell membrane intact we expect that only regions of the proteins outside the cell membrane will be labeled, while transmembrane segments and cytoplasmic regions will be not. However, since previous experiments suggested that biotin labeling was happening in intracellular proteins, because BirA could be entering in the cells, we did first an analysis of the labelling specificity of extracellular proteins by BirA. For this propose, we have compared the protein extract treated with BirA (and enriched by affinity) with the biotinylated cells (not lysated) without doing the membrane extraction. We would expect to find an enrichment in extracellular proteins in the sample derived from biotinylated cells. However, we found cytosolic proteins (studying levels of ribosomal proteins) in similar levels than in the sample obtained from the extract. Thus, it looks that when BirA was used for labeling it was entering into the cells as found in previous experiments. Then, we compared the sample of cells treated with BirA-beads with the sample cells with BirA. The idea was that by coating the BirA we would prevent the entrance in the cells. However, again we were finding similar enrichment or separation for extracellular and intracellular proteins; BirA or radicals from the reaction were entering in the cell and labelling intracellular proteins.

However, biotinylation could give us information of the regions of the protein that are exposed to the solvent. Lys biotinylated residues prevent trypsin digestion and we did not include that modification in the Mascot software search. As a result, we expect that the biotinylated peptides won't be detected in the streptavidin elution fraction. Thus, all the peptides of these 139 identified proteins that are detected in the membrane sample but not in the biotinylated sample could be possible regions of the membrane proteins exposed to the outside and therefore targets for vaccination.

We have identified 815 unique peptides corresponding to 793 regions of 123 out of the 139 proteins that are not detected in the biotinylated sample and are present in the membrane fraction. The fact that a peptide is not detected in the biotinylated sample could be also due to low abundance of the corresponding proteins which results in random detection. Thus, we could assign a confidence to these peptides based on the area of the peptide detected in the non-biotinylated sample. The most reliable ones are those properly detected in all samples except in the biotin experiment. Furthermore, peptides should be filtered by only considering those

that finish or start with an K because the ones that have an R (not followed by any K) should be also detected in the biotinylated sample. After applying this filter of K and R we obtain 777 peptides, 168 with an area >10^8, corresponding to 63 out the 139 identified membrane proteins (Table S5). Peptides with areas of >10^7 (401; corresponding to 100 proteins) should also be considered as good candidates, while areas lower than 10^7, should be treated with caution (Table S5).

➢ Targets for vaccination against *M. bovis* PG45

The most interesting targets for vaccination would be those membrane proteins accessible to biotinylation that are essential for *M. bovis* growth. Doing an orthologue search and using the essentiality data of *M. pneumoniae* (Lluch-Senar M., MolSysBio 2015), we assigned the essentiality to 26 of the 139 studied membrane proteins. This number could be refined by improving the orthologue search (work ongoing at the CRG) and looking at the essentiality in other mycoplasma species. Interestingly, 20 out of this 26 conserved proteins are essential for *M. pneumoniae* growth. An interesting candidate could be FtsH, different ABC transporters, F0F1 ATP synthase as well as proteases (Table 6; Table S6).

Protein (accesion) name	Product name (function)	Experimental classification	Predicted LS2	MPN	Essentiality in M. pneumoniae
MBOVPG45_RS00670	polyamine ABC transporter permease	Membrane	Membrane	MPN056	E
MBOVPG45_RS00675	spermidine/putrescine ABC transporter permease	Membrane	Membrane	MPN057	E
MBOVPG45_RS01540	ABC transporter permease	Membrane	Membrane	MPN135	E
MBOVPG45_RS01535	ABC transporter permease	Membrane	Membrane	MPN136	E
MBOVPG45_RS01430	preprotein translocase subunit SecY	Membrane	Membrane	MPN184	E
MBOVPG45_RS01780	cation transporter	Membrane	Membrane	MPN209	E
MBOVPG45_RS00555	peptide ABC transporter permease	Membrane	Membrane	MPN215	E
MBOVPG45_RS00495	prolipoprotein diacylglyceryl transferase	Membrane	Membrane	MPN224	E
MBOVPG45_RS01970	hypothetical protein	Membrane	Membrane	MPN244	E
MBOVPG45_RS00100	sugar ABC transporter substrate-binding protein	Membrane	Membrane	MPN260	E
MBOVPG45_RS02085	PTS sugar transporter subunit IIA	Membrane	Membrane	MPN268	E
MBOVPG45_RS02990	peptidase A8	Membrane	Membrane	MPN293	E
MBOVPG45_RS00775	pyruvate kinase	Membrane	Cytosol	MPN303	E
MBOVPG45_RS02945	membrane protein	Membrane	Membrane	MPN396	E
MBOVPG45_RS02715	ABC transporter ATP-binding protein	Membrane	Membrane	MPN416	E
MBOVPG45_RS03480	membrane protein	Membrane	Membrane	MPN482	E
MBOVPG45_RS01640	arginyl-tRNA synthetase	Membrane	Cytosol	MPN556	E
MBOVPG45_RS02195	F0F1 ATP synthase subunit A	Membrane	Membrane	MPN604	E
MBOVPG45_RS03935	phosphatidate cytidylyltransferase	Membrane	Membrane	MPN637	E
MBOVPG45_RS04320	cell division protein FtsH	Membrane	Membrane	MPN671	E

Table 6. Putative targets for vaccination against *M. bovis*.

For example, in the case of FtsH we find several putative antigenic peptides shown in Table 7. Study of the location of detected biotinylated peptides by transmembrane predictor program indicates that the Cter region of this protein is exposed to biotinylation (Figure 3).

Name	Biotin	Reference Biotin	Membrane	Reference Membrane	Peptide sequence	Position in Protein
MBOVPG45_RS04320	N.D.	1,90E+07	5,90E+08	5,60E+08	[K].TFLFLPDK.[F]	MBOVPG45_RS04320 [100-107]
MBOVPG45_RS04320	N.D.	2,30E+05	6,50E+08	5,50E+08	[K].NAPAIIFIDELDAIGR.[T]	MBOVPG45_RS04320 [291-306]
MBOVPG45_RS04320	N.D.	N.D.	2,90E+08	6,40E+07	[K].SAQFSAQVAHEIDIEVR.[N]	MBOVPG45_RS04320 [581-597]
MBOVPG45_RS04320	N.D.	N.D.	1,50E+08	1,50E+08	[K].MVTEWGMSELGPIQYEQDEGSPFLGR.[D]	MBOVPG45_RS04320 [551-576]
MBOVPG45_RS04320	N.D.	N.D.	1,50E+08	3,10E+07	[K].ELIAMITSFMGGR.[A]	MBOVPG45_RS04320 [509-521]
MBOVPG45_RS04320	N.D.	N.D.	2,60E+08	4,40E+08	[K].SELTMVAYHEAGHAVVGIK.[M]	MBOVPG45_RS04320 [454-472]
MBOVPG45_RS04320	N.D.	N.D.	3,90E+08	6,20E+08	[K].RTPGYSGAQLENVINEAGLLAVR.[R]	MBOVPG45_RS04320 [400-422]
MBOVPG45_RS04320	N.D.	N.D.	2,90E+08		[K].EIVDYLK.[N]	MBOVPG45_RS04320 [214-220]

Table 7. Putative FtsH peptides that could be good candidates for vaccination.

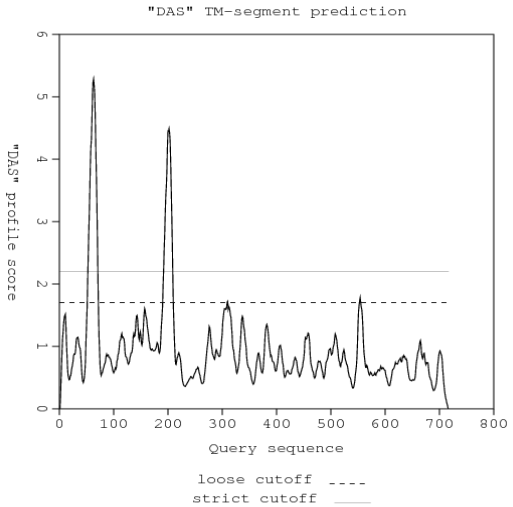

Figure 3. Prediction of transmembrane regions of FtsH protein.

2.4 Identification of *M. hyopneumoniae* membrane proteins

➢ Membrane proteins predicted by analysis R1_LS of biotin and membrane samples

We have identified 112 and 111 putative membrane proteins in the samples of biotinylated sample and membrane experiments, respectively. The number of cytosolic proteins are 400 and 401 for both experiments, respectively. When this classification is compared with the LS prediction we find that 83,4% and 82,6% of the proteins well classified for the biotin and membrane experiments, respectively. Interestingly, 75% and 76% of the identified membrane proteins in the membrane and biotinylation experiments are also classified as membrane proteins in LS, respectively. This percentage in the experiment of *M. bovis* PG45 was 83%. This result suggests that maybe the assignment in LS set of *M. hyopneumoniae* should be reviewed or that the experiment in *M. hyopneumoniae* did not work, as well as, in *M. bovis* PG45. We reviewed again after the prediction the sets of true positive and true negative for the ROC curves and we could change the classification of membrane and cytosolic for 6 proteins (Table S6). Thus, we do not expect that the difference versus *M. bovis* experiment is due to the sets of the ROC curves. One explanation could be the differences in the total areas of the different samples (reference/membrane) for each experiment.
In fact, the total amount of protein for the samples considered in the study are:

Total protein amount	Sample Name	Description
2,5E+10	161130_S_TFLS_21_04_5pto_21_21B	Membrane isolation-membrane proteins
1,4E+11	161130_S_TFLS_18_04_5pto	Proteins
2,9E+09	161130_S_TFLS_20_04_5pto	Biotinylated membrane isolation-elution
2,4E+09	161130_S_TFLS_17_04_5pto	Biotinylated proteins

Table 8. Total protein quantification derived from sum of areas of all *M. hyopneumoniae* proteins detected in the sample

There is one order of magnitude difference between both samples in the membrane experiment. Thus, it could be that detection of proteins is affected decreasing the number of membrane proteins detected. Currently, we are repeating the injections with 10 times more of sample of membrane.
The number of proteins common to both experiment is 86 (Table S3). The 85% of those proteins were classified as membrane proteins by LS study. Interestingly, 9 out 13 proteins assigned as cytosolic in LS but classified as membrane in both experiments are hypothetical proteins, suggesting a new role for those proteins

as membrane components. After reviewing these 13 proteins three were reassigned as membrane category increasing this percentage to 88.2% (Table S3).

We have found 26 proteins classified as membrane proteins in the experiment of biotinylated membrane that do not appear in the experiment of membrane (Table S3). 52% of those proteins were assigned as membrane proteins in LS assignment. Revision of assignments revealed that 2 proteins were miss-annotated and the percentage after correction increased to 60%. For the membrane experiment, there are 25 proteins that are present in the experiment of the membrane that are not present in the biotinylation experiment (Table S3). In this case 44% of these proteins are assigned as membrane proteins in LS study and any miss-classification was found.

➢ Identification of putative antigens and exposed protein regions (in *M. hyopneumoniae*)

We have studied the unique peptides for the 86 proteins identified in both experiments.

We have identified 7328 unique peptides in the experiment. 1440 unique peptides (1365 different aa sequences) are associated to 86 membrane proteins identified and they were found in any of the 7 samples of *M. hyopneumoniae* experiments (Table S4). 718 out of 1440 peptides are not detected in the biotinylated sample and are present in any other sample. After applying the filter of K and R (see above for *M. bovis*) we obtained 696 out 718 peptides, 99 with an area >10^8, corresponding to 17 out the 86 identified membrane proteins (Table S5). Peptides with areas of >10^7 (376; corresponding to 52 proteins) should also be considered as good candidates, while areas lower than 10^7, should be treated with caution (Table S5).

➢ Target for vaccination against *M. hyopneumoniae*

The most interesting targets for vaccination would be those membrane proteins accessible to biotinylation that are essential for *M. hyopneumoniae* growth. Doing an orthologue search and using the essentiality data of *M. pneumoniae* (Lluch-Senar M., MolSysBio 2015). I assigned the essentiality to 21 of the 86 studied membrane proteins. This number could be refined by improving the orthologue search (work ongoing at the CRG) and looking at the essentiality in other mycoplasma species. Interestingly, 12 out these 21 conserved proteins are essential; 3 Fitness and 6 non-essential for *M. pneumoniae* growth. Interesting candidate could be peptide ABC transporter permease, PTS sugar transporter and the hemolysin (Table 9 and Table S7).

Protein id M hyo	MPN Orthologue	Essentiality	Funtion
MHP_RS02120	MPN023	E	methionyl-tRNA synthetase
MHP_RS02845	MPN057	E	spermidine/putrescine ABC transporter permease
MHP_RS03000	MPN207	E	PTS sugar transporter subunit IIA
MHP_RS03195	MPN207	E	hypothetical protein
MHP_RS02575	MPN215	E	peptide ABC transporter permease
MHP_RS02570	MPN216	E	peptide ABC transporter permease
MHP_RS01430	MPN224	E	diacylglyceryl transferase
MHP_RS02500	MPN244	E	hypothetical protein
MHP_RS03175	MPN260	E	sugar ABC transporter permease
MHP_RS01645	MPN421	E	hypothetical protein
MHP_RS03110	MPN482	E	membrane protein
MHP_RS02770	MPN527	E	hypothetical protein
MHP_RS01975	MPN133	F	hypothetical protein
MHP_RS03355	MPN159	F	hemolysin
MHP_RS01935	MPN415	F	DNA repair protein HhH-GPD

Table 9. Putative targets for vaccination against *M. hyopneumoniae*.

2.5 List of putative target proteins

We provide the two lists of putative target proteins to use as antigens for vaccination against *M. bovis* and *M. hyopneumoniae*:

For *M. bovis*:

Protein (Accession) name	Product name (function)	Experimental classification	Predicted LS2	MPN	Essentiality in *M. pneumoniae*
MBOVPG45_RS00670	polyamine ABC transporter permease	Membrane	Membrane	MPN056	E
MBOVPG45_RS00675	spermidine/putrescine ABC transporter permease	Membrane	Membrane	MPN057	E
MBOVPG45_RS01540	ABC transporter permease	Membrane	Membrane	MPN135	E
MBOVPG45_RS01535	ABC transporter permease	Membrane	Membrane	MPN136	E
MBOVPG45_RS01430	preprotein translocase subunit SecY	Membrane	Membrane	MPN184	E
MBOVPG45_RS01780	cation transporter	Membrane	Membrane	MPN209	E
MBOVPG45_RS00555	peptide ABC transporter permease	Membrane	Membrane	MPN215	E
MBOVPG45_RS00495	prolipoprotein diacylglyceryl transferase	Membrane	Membrane	MPN224	E
MBOVPG45_RS01970	hypothetical protein	Membrane	Membrane	MPN244	E
MBOVPG45_RS00100	sugar ABC transporter substrate-binding protein	Membrane	Membrane	MPN260	E
MBOVPG45_RS02085	PTS sugar transporter subunit IIA	Membrane	Membrane	MPN268	E
MBOVPG45_RS02990	peptidase A8	Membrane	Membrane	MPN293	E
MBOVPG45_RS00775	pyruvate kinase	Membrane	Cytosol	MPN303	E
MBOVPG45_RS02945	membrane protein	Membrane	Membrane	MPN396	E

Protein (Accession) name	Product name (function)	Experimental classification	Predicted LS2	MPN	Essentiality in *M. pneumoniae*
MBOVPG45_RS02715	ABC transporter ATP-binding protein	Membrane	Membrane	MPN416	E
MBOVPG45_RS03480	membrane protein	Membrane	Membrane	MPN482	E
MBOVPG45_RS01640	arginyl-tRNA synthetase	Membrane	Cytosol	MPN556	E
MBOVPG45_RS02195	F0F1 ATP synthase subunit A	Membrane	Membrane	MPN604	E
MBOVPG45_RS03935	phosphatidate cytidylyltransferase	Membrane	Membrane	MPN637	E
MBOVPG45_RS04320	cell division protein FtsH	Membrane	Membrane	MPN671	E

Table 10. Putative targets for vaccination against *M. bovis*.

For *M. hyopneumoniae*:

Protein id M hyo	MPN Orthologue	Essentiality	Funtion
MHP_RS02120	MPN023	E	methionyl-tRNA synthetase
MHP_RS02845	MPN057	E	spermidine/putrescine ABC transporter permease
MHP_RS03000	MPN207	E	PTS sugar transporter subunit IIA
MHP_RS03195	MPN207	E	hypothetical protein
MHP_RS02575	MPN215	E	peptide ABC transporter permease
MHP_RS02570	MPN216	E	peptide ABC transporter permease
MHP_RS01430	MPN224	E	diacylglyceryl transferase
MHP_RS02500	MPN244	E	hypothetical protein
MHP_RS03175	MPN260	E	sugar ABC transporter permease
MHP_RS01645	MPN421	E	hypothetical protein
MHP_RS03110	MPN482	E	membrane protein
MHP_RS02770	MPN527	E	hypothetical protein
MHP_RS01975	MPN133	F	hypothetical protein
MHP_RS03355	MPN159	F	hemolysin
MHP_RS01935	MPN415	F	DNA repair protein HhH-GPD

Table 11. Putative targets for vaccination against *M. hyopneumoniae*.

3 Future plans

Identified membrane proteins and exposed peptides could be used as antigens for vaccination; as well as for exposition of them in *M. pneumoniae* chassis surface. However, in order to do so, additional experiments are required. By expressing identified proteins recombinant, and test the serum of infected animals, we could evaluate the ones that are recognized by the immune system. Also, vaccination with the recombinant antigens and challenging the animals could be other option to evaluate the effectiveness of the identified candidates. However, a lot of animals would be required and it would be very expensive and time consuming. The MycoSynVac Consortium plan is that ATG will analyze the identified list of candidate proteins by using their pipeline to define and reduce the list to the best antigens and peptides that could promote a good protection. Once we have the results, we will be able to test the immune response by suing serum of infected animals and later one injecting in pigs and testing the immune response. It will reduce the number of animals required and also the cost and timing of these experiments.

4 Conclusions

The MemBioID protocol described in this deliverable allowed the identification of membrane proteins in *M. bovis* and *M. hyopneumoniae*. The final list includes 20 proteins for *M. bovis* and 15 for *M. hyopneumoniae.* This technique can be applied to more species to find possible antigens to expose on *M. pneumoniae* chassis surface.

Identification of proteins exposed in the surface of Mycoplasmas is interesting because it could:

1) Identify **signals to expose antigens** in the surface to engineer Mycoplasma as vaccine.
2) Identify **"specific" Mycoplasma antigens.** By doing surfome and combining with Western blot of infected patients (human for *M. pneumoniae* and animals for other Mycoplasma species), we can identify the potential proteins that generate an immune response. Applications: **Kits of diagnosis and vaccination.**
3) Identify **"universal" antigens** for Mycoplasma infections. Combining surfome of different bacteria and sera of different infected animals we can study common antigenic proteins and identify **common exposure signals** and common antigenic peptides that can be relevant if other mycoplasma species want to be used as chassis or if we want to obtain **a universal vaccine**.

Technology Transfer Department at CRG and Business Development Unit at MSD are evaluating these possible exploitation lines. MycoSynVac Innovation Board is aware of these results and monitoring the progress. The best alternatives, considering protection and exploitation, are being analyzed to see which is the best way to proceed.