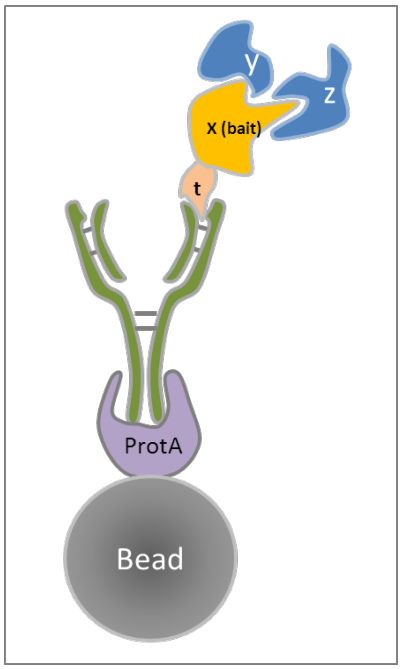
Discovery of protein-protein interactions by affinity purification and mass spectrometry (AP-MS)

General suggestions and guidelines for sample preparation

Mass spectrometry has become a very useful tool to discovery protein-protein interactions, the knowledge of which can be crucial to determine the function of a protein and its involvement in cellular pathways.



AP-MS (Affinity Purification – Mass-Spectrometry) is one of the most popular tools for this and is based on the principle of "guilty by association", i.e. on the hypothesis that proteins physically interacting are functionally related.

The goal of one such experiment is thus to identify possible unknown protein(s) **Y** and **Z** from cells or tissues which specifically co-purify with a **protein of interest** (here **X**, or **"bait"** protein) .

In the majority of cases the affinity reagent used is an immunoglobulin (antibody, Ab) which recognizes directly X or an affinity **tag (t)** which has been added to the N- or C-terminus of X.

Cells/tissues are then lysed and the Ab is added. Recovery of the Ab-t-X complex is then done mostly using a **capture resin** made of ProteinA-coated beads which bind the Fc portion of the Ab. However other systems exist which rely on other types of affinity interactions, i.e. biotinylation⇔ (strept)avidin systems or (His)6 ⇔ Ni2+ columns etc. It is important to note that, several interactions (often non covalent ones) have to be conserved all along the purification procedure for the final complex to be recovered.

Irrespective of the type of affinity reagents used, some general principles apply :

1- Interactions Ab to X and Ab to capture resin should be specific, strong and stable (low k_{off}) during capture and washing but reversible at the end of the procedure for elution from the resin.

2- Conditions (T, buffer, ionic strength, pH) should be sufficiently mild to retain the relevant interactions during capture and wash.

3- Sufficient time must be allowed for the resin to enter in contact with a maximum of the sample to adsorb as much as possible of the X-Y-Z complex.

Now, fishing in a single step a single molecule from an extremely complex and dense mixture of promiscuously interacting proteins has some practical limitations. Indeed, all AP-MS "purifications" are - at best - enrichment steps and a more or less important level of undesirable nonspecific contaminants is always present. These can be proteins that are recovered in the final sample because they are :

i) extremely abundant proteins that are simply not washed away sufficiently well during the purification

ii) proteins which interact nonspecifically (?) with the Ab

iii) proteins which interact nonspecifically with the beads

iv) proteins which sediment in aggregates together with the resin (w/o any binding) and are mechanically recovered

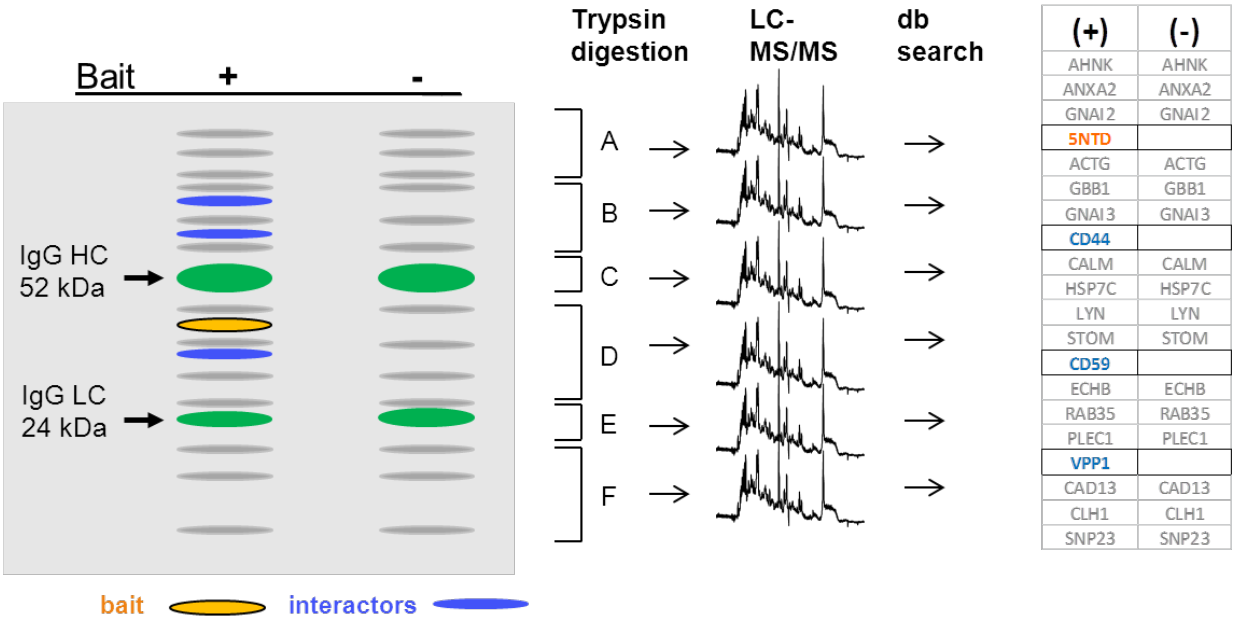
v) components of the protein synthesis machinery which bind to immature forms of X being synthesized, folded, exported, etc. Although these can be interesting depending on the goal of the study, such interactors can become quite abundant especially when X is overexpressed and the synthesis machinery struggles to produce and fold such amounts.

The optimization of the conditions to obtain a fruitful purification for AP-MS requires the optimization of the amount of X recovered and at the same time the maximal reduction of the background(noise) of nonspecific contaminants. Since the latter can never be completely eliminated, a suitable **negative control AP experiment** used to define the background **is mandatory**. Such a negative control should reproduce a maximum of the elements that generate the "noise" in the AP sample but without the "active" component, that is the interacting portions of X.

In practice, some general principles apply : you need to have enough bait protein for its detection and for detecting interacting proteins. The scale, yield and efficiency of purification are therefore important parameters to optimize. At the same time, one should try to maximize specificity and "cleanliness" of the final sample, which in turn may lead to the choice of more elaborate strategies. Such alternatives are OK as long as these do not compromise too much the yield. In other words, it is better to have a somewhat dirtier sample but with a good amount of material than a super-clean sample with barely enough protein to detect. More practical suggestions can be found below in the form of Q&A.

Our analytical workflow :

After affinity purification, proteins are dissociated from the resin (beads) and separated by limited SDS-PAGE electrophoresis (a short gel migration). Whole lanes are cut into 5-7 slices and proteins are in-gel digested with trypsin. MS is then used to identify a maximum of the peptides present in the resulting mixtures by liquid chromatography-tandem mass spectrometry ("LC-MS/MS"). The analysis generates datasets that contain sequence-related information. Comparison of these data with the sequence database of the organism studied allows to match sequences and identify confidently up to hundreds of proteins in one analysis.



Data from all slices of one sample are pooled to give a single list of proteins. The lists for the two samples (neg. control and positive AP) are then aligned to determine common proteins as well as proteins uniquely found in one sample, as the bait and specific interactors should be.

Practical considerations : FAQ

The most frequently asked questions are listed here below. These cover most of the aspects related to experiment planning and the choice of strategy and are therefore very important.

Choice of source material: AP from cell lines, primary cells, tissues ?

Obviously this choice is largely dictated by the biological question, the goal of the project, the characteristics and the expression pattern of the bait protein. Based on our experience, however, a cell line transfected with a tagged version of the bait protein almost always constitutes a useful initial test bench, which can provide quickly useful hints for further experiments and often some reasonable candidate interactors. The advantage is that such experiments can be performed quickly and scale is usually not a limiting factor. We have seen groups to spend months or years trying (and failing) to get a good IP from primary cells or tissues, only to later move to cell lines and obtain quickly some promising results. We should consider that, even if a given protein of interest is not expressed in HeLa's or HEK293 cells, some of the interactors probably are, due to the modularity and ubiquity of cellular pathways. With transfectable cells it is also easier to validate the interactions found by reverse IP, and only later move to primary cells or tissues.

Endogenous protein AP or transfection of a tagged protein ?

It is known that a tag may interfere with a protein's function and stability and sometimes intracellular localization. However using tags also has clear advantages for the purification and offers an easy way to have a negative control, i.e. the tag expressed alone. Purifying the endogenous proteins X with direct antibodies is in principle a more "physiological" approach. However three problems exist: i) the anti-X antibody may have unexpected cross-reactivities and recognize other proteins under native conditions and ii) the choice of a negative control is more arbitrary (often generic IgG's or an Ab against an "unrelated"(?) protein are chosen) ; iii) the anti-X Ab may interfere with the binding of the protein of interest to its interactors. In our experience, problem i) may be quite a serious one, because false positives due to Ab cross-reactivity are not easy to identify and dismiss.

I choose to tag : which tag should I use ?

Generally, "classical" small tags (Flag, HA, Myc) work well and their interference with protein function is often minimal. Still, the choice of tag location (C-term, N-term) is important. Larger tags (e.g. GFP) have also given good results in our experience. Small tags can be used in tandem, e.g. twice the same tag to increase purification efficiency or two different tags for different purposes (one for purification, one for detection by Western). Two-step purifications ("TAP tags") have been made popular by some studies due to their increased specificity and "cleanliness" but are not always easy to implement. The critical point is that the elution from the first affinity column must be made under very mild conditions to preserve non-covalent interactions and allow binding to a second affinity resin. These steps sometimes lead to important losses of material. Still, using two tags from the start is good because it gives choices and flexibility to evaluate the best possible strategy (using tag1 or tag2 or both sequentially).

Is it worth cross-linking ?

In principle, cross-linking may seem a logical step to freeze complexes in their physiological state and prevent losses occurring during purification. Two technical limitations however exist. First, treatment with the cross-linker should not interfere with the affinity purification step (this can happen by modification of the Ab epitope). Second, cross-linking can form large complexes and in extreme cases this may interfere with gel separation, protein digestion and MS. Reversible cross-linkers should be used when possible. A titration of the cross-linker concentration with monitoring of purification efficiency may be necessary to optimize conditions.

Which beads should I use ? Which ones give the least background ?

A variety of affinity beads exist, from standard agarose to magnetic beads. Agarose beads display non-negligible levels of unspecific binding, although the background is determined by many other factors. The consensus from many studies is unfortunately that the level of nonspecific binding is lab- and protocol-specific and lists of contaminants are difficult to transfer. Several experiments done in standardized conditions are essential if one wants to build know-how on unspecific contaminants. Nevertheless it's worth visiting the CRAPome website (http://www.crapome.org) which contains such a list of proteins identified in a large number of negative control experiments. We have also our own list of "common contaminants" which are frequently found in this type of experiments. Please do not hesitate to ask for it.

How to do lysis / IP ?

Standard lysis buffers containing 1% Triton or 1% NP-40 are a good starting choice. 1% CHAPS can be tested as an alternative detergent. Of course mechanical lysis and no detergents is also possible. Based on our experience, **lysates should not be prepared at too high protein concentration** (max 1.5-2.0 mg/ml), because this can result in the formation of **aggregates**, which stick to the beads or sediment with them, greatly increasing the background. In the same way, aggregates may form with time during incubation and shaking. **We recommend clearing lysates at 15'000 xg, adding the Ab, incubate 1-3h then centrifuge the samples again and discarding the pellet before adding the beads**.

Amount of beads :

Beads are indeed one of the major sources of background. The amount of beads used should be minimized. Since beads have often a very high capacity (10 ul of settled ProteinA-SepharoseCL-4B can bind up to 200 ug of Ab !!), not much is needed. Typically we recommend a maximum of 50 ul of packed beads, with 10-20 ul being the optimal amount. A good solution to recover the beads after in-batch incubation is to pass the lysate-beads suspension through an empty spin filter cartridge, which can also be used for washing.

How to elute ?

Elution is dependent on the affinity step and should be as efficient and as specific as possible. Generic elution methods include acidic pH, basic pH, denaturing conditions (SDS, Urea, + reducing agents), competition with free tag (Flag, HA peptides,…),… Importantly, the recovered sample should be compatible with further steps. As a first choice in general we recommend simple direct elution in SDS gel sample buffer, which is highly effective, though not specific. More specific elutions may result in partial recovery and samples that are for example very diluted or contain high salt, peptides or detergents

which interfere with further separation and MS analysis. Nevertheless, there is always the possibility to optimize procedures and we have obtained sometimes good results with acidic elution (5% acetic acid followed by lyophilization).

How many cells are needed ? How much protein is needed ?

It is impossible to predict the number of cells needed, as it is heavily dependent on the level of expression of the bait protein and the efficiency of its purification. It is necessary to run test purifications and analyse the eluate by gel with protein staining (Coomassie or Silver staining). The critical point is to be able to observe a "reasonably strong" band corresponding to the bait protein by silver or, even better, Coomassie blue. Although MS detection is potentially more sensitive than most protein stains, interactors are usually present in the sample in (much) lower amounts than the bait protein, so a relatively high amount of bait is needed to maximize the chances of detecting interactors. Really, this is a case of "the more, the better".

How do I know if my samples are good enough ? They look good by Western…

Again, you should run gels and protein stain (Coomassie, Silver) them. Western blot evaluation is a good start but is not enough. You need to evaluate the total content of the sample, the real amount of bait protein and the level of background proteins.

Can I do preliminary checks ?

YES, we really encourage this. Especially if you are not sure that the band you see on your Coomassie/silver gel is the bait protein, you can cut the region that should contain it, freeze it and bring it to us for analysis. This is very useful because it can give us a sense of how much bait protein you have and how well it is detected by MS.

Analytical workflow : why analysing the whole gel lane ?

Classically, after AP (IP), proteins purified would be separated by SDS-PAGE and stained with silver stain. Only detectable bands which appear different between the (-) and (+) bands would be analysed. This is very virtuous and stringent way of proceeding but tends to limit greatly the number of actual interactors identified. Reasons : i) MS detection now surpasses the sensitivity of most protein gel stains, ii) bands of interesting proteins may not be visible because covered by bands of abundant contaminants, iii) proteins may change their migration from sample to sample due to modifications, generating false positive bands.

Analytical workflow : why run a gel at all ? It should be possible to digest and do LC-MS/MS directly

Yes, this is true. Many groups have done this. However we found that a gel separation has major advantages, for example i) it allows a visual inspection of the sample and evaluation of concentrations, purity and complexity, ii) by isolating strong bands containing e.g. Ig heavy and light chains, whose signal could cover that of interesting proteins, we increase depth of analysis and possibility to identify low-abundance interactors. More direct approaches, not including gel steps, are often implemented for larger scale projects including dozens (hundreds) of AP-MS experiments for which the analysis time becomes the limiting step. Removing the gel separation step requires further technological implementation, for example using special tags and detergents that do not interfere with MS and/or optimizing elution conditions without using SDS.

What kind of result files will I receive ?

You will get the results in two formats. An excel .xls file will give you a quick summary of the data, reporting usually the number of peptide spectra matched ("spectral counts") to every protein identified in any of the samples. Spectral counts can have a semi-quantitative meaning but have to be taken with quite some caution, especially for values below 5. In addition you will receive a file in Scaffold Format (.sf3). Scaffold Reader is a free downloadable software for distribution of MS data (www.proteomesoftware.com). The .sf3 file contains essentially the whole information generated by MS after database search. To know more, please read our other short guide "Shotgun_data_guide_v5.doc ", which can be downloaded from our web page (http://www.unil.ch/paf/page19604_en.html).

I have my list of proteins and some candidates. How do I validate results ?

Experimentally, a reverse – IP with the candidate and identification of the bait protein as interactor is the best validation. Otherwise the nature of the molecules detected and their natural abundance can give important clues. Finding a low abundance protein as interactors implies a high degree of enrichment by the AP step and is therefore a good indication of specificity and strength of interaction.

Summary of main steps

1. Design expression system and tagged constructs or test antibodies for endogenous protein IP
2. Choose carefully a negative control

3. Test IP in cell lines if possible ; do protein staining (Coomassie or silver). Note : we recommend colloidal Coomassie staining for better results. You can send us an image of the gel and we'll give you some feedback (wwwpaf@unil.ch).
4. Cut out and freeze band corresponding to "bait" protein (prot. of interest). Bring it to the PAF for analysis. In case of uncertainty, cut broadly the region where the protein should be. Alternatively, bring us the washed, frozen beads and we will run the gel for you.
5. Scale up and do preparative experiment. At the end, wash beads 1x with PBS, remove all buffer and freeze beads at -20C. Bring the beads to the PAF.