JOURNAL OF PROTEOMICS 94 (2013) 89–109

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Review

Mass spectrometry-based identification of proteins interacting with nucleic acids

*A. Tacheny, M. Dieu, T. Arnould, P. Renard**

Laboratory of Biochemistry and Cell Biology (URBC), NAmur Research Institute for LIfe Sciences (NARILIS), University of Namur, 61 rue de Bruxelles, 5000 Namur, Belgium

ARTICLE INFO

Article history:
Received 27 May 2013
Accepted 13 September 2013
Available online 20 September 2013

Keywords:
DNA-affinity capture
Mass spectrometry
Protein–DNA interactions
Protein–RNA interactions
RNA-affinity capture
Transcription factors

ABSTRACT

The identification of the regulatory proteins that control DNA transcription as well as RNA stability and translation represents a key step in the comprehension of gene expression regulation. Those proteins can be purified by DNA- or RNA-affinity chromatography, followed by identification by mass spectrometry. Although very simple in the concept, this represents a real technological challenge due to the low abundance of regulatory proteins compared to the highly abundant proteins binding to nucleic acids in a nonsequence-specific manner. Here we review the different strategies that have been set up to reach this purpose, discussing the key parameters that should be considered to increase the chances of success. Typically, two categories of biological questions can be distinguished: the identification of proteins that specifically interact with a precisely defined binding site, mostly addressed by quantitative mass spectrometry, and the identification in a non-comparative manner of the protein complexes recruited by a poorly characterized long regulatory region of nucleic acids. Finally, beside the numerous studies devoted to *in vitro*-assembled nucleic acid–protein complexes, the scarce data reported on proteomic analyses of *in vivo*-assembled complexes are described, with a special emphasis on the associated challenges.

Contents

* *Corresponding author*. Tel.: +32 81724128; fax: +32 81724135.
E-mail address: patsy.renard@unamur.be (P. Renard).

1874-3919/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jprot.2013.09.011

1. Introduction

Understanding gene expression represents a major key to decipher most fundamental processes in cell biology. It is now largely accepted that gene expression regulation is a complex cellular mechanism integrating several levels of control including epigenetics, transcription and translation. Both transcriptional and translational controls depend on interactions with proteins and with regulatory non-coding RNAs, this latter type of interaction being pointed out more recently. The identification of the regulatory proteins that control DNA transcription or RNA stability and translation represents thus a key step in the comprehension of those processes, but constitutes a very difficult task. For this goal, affinity chromatography strategies have been developed, based on a theoretically simple concept. In the first step, a nucleic acid (NA) sequence of interest, either RNA or DNA, is used as a bait, immobilized on a chromatographic support, to capture NA-interacting proteins contained in a protein extract. Once NA/protein complexes are formed, the proteins or protein/NA complexes are eluted and the second step consists of the identification of captured proteins by mass spectrometry.

In this review, we will discuss the different strategies that have been set up to reach this purpose, first in the case of DNA-interacting proteins, and secondly when RNA is used to capture RNA-interacting proteins. The different biological properties of DNA and RNA are important to consider in the design of DNA- or RNA-affinity chromatography features and will therefore be discussed separately for the two types of nucleic acids. However, in both types of NA-interacting protein complexes, the chromatography eluates consist of complex protein mixtures containing low abundant sequence-specific interacting proteins and highly abundant proteins that interact in a non-specific manner with those nucleic acids. As the mass spectrometry-based identification of the specific partners represents a similar challenge for DNA- and RNA-affinity purified proteins, this step will be discussed simultaneously for both nucleic acids interacting proteins.

2. DNA-affinity purification

2.1. *History of DNA-affinity: From biased to -omics methods…*

Selection and modulation of gene transcription in response to environmental changes or developmental signals depend on the coordinated influence of transcriptional regulator interactions with *cis*-regulatory sequences and chromatin modifications [1]. Among transcriptional regulators, one can distinguish at least two categories: those that directly bind to DNA *cis*-regulatory sequences, called transcription factors (TFs), and the co-regulators interacting indirectly with DNA sequences through dynamic interactions with transcription factors. While the general TFs (GTFs) are necessary for transcription to occur, and usually belong to the large pre-initiation complex, the various specific TFs, often activated in response to intra- or extracellular cues, bind to specific binding sequences that can be located in the close proximity of the transcription start site (promoter region) or at long distances (enhancers) [2,3]. The latter can then recruit the basal transcriptional machinery or play on chromatin architecture and nuclear 3D organization through diverse enzymatic activities at the origin of epigenetic modifications [4]. The main difficulty in the study of gene expression regulation comes from the fact that the presence of a binding site for a specific TF does not ensure that this factor actually binds to this site as this also depends on the genomic and protein context surrounding the DNA site, on the abundance, translocation and/or activation of the TF (by post-transcriptional modifications for instance) and on the chromatin status of the loci. Moreover, the precise spatiotemporal modulation of gene expression is due to complex and regulated networks of protein interactions taking part on several binding sites at the level of the promoter or other *cis*-regulatory sequences. The elucidation of these specific interactomes is of major importance to better understand gene expression regulation.

Various experimental methods to assay protein–DNA interactions exist, but many of these methods are based on the fact

that at least one of the partners of the interaction has been previously identified. Those methods will not be detailed here, as they have been recently reviewed [5–7]. Briefly, one can distinguish the protein-centered methods, like chromatin immunoprecipitation (ChIP)-derived methods [8,9], from the DNA sequence-centered methods [10]. In this second category, individual DNA sequences are investigated for their protein-binding capacity by DNaseI protection assays [11,12], or by electrophoretic mobility shift assay (EMSA) *in vitro* [13]. EMSA can reveal that a short synthetic double strand DNA sequence is able to bind proteins present in a nuclear extract, as visualized by a shift in the electrophoretic migration of the labeled oligonucleotide. However, neither DNaseI protection assays nor EMSA do reveal the identity of the proteins involved in the interaction, even if, for EMSA, such information can be brought by the use of good quality antibodies that are specific for a candidate transcription factor suspected or known to interact with the DNA sequence. In this condition, the addition of antibodies induces a further shift in the electrophoretic mobility ("supershift").

A major drawback of such methods is that they are biased since they are conditioned by a pre-existing knowledge of the regulatory DNA binding site and/or the transcription factor(s) involved. However, many researchers have been facing the simple question: "What are the transcriptional regulators responsible for the modifications in the expression of a gene of interest?" This question, that requires unbiased methods, has become more crucial with the arrival of -omics era, when genome-wide analyses of gene expression modifications are easily obtained. Tackling this question usually starts with an *in silico* analysis of the promoter sequence of the gene of interest [14,15]. Such analyses are based on transcription factor databases containing a collection of consensus sequences, and on diverse algorithms like Matinspector or Transfac, to look for putative DNA binding sites in the sequence of interest [16,17]. Although they may be useful and usually constitute a first step in the approach, *in silico* analyses present several drawbacks. First, they typically provide several hundreds of candidate transcription factors that largely diverge depending on the algorithm used. Second, they generate a large number of false positive and false negative [18,19]. Finally and very importantly, as those databases are fed with already validated consensus/transcription factor pairs, putative identification by such *in silico* analyses are restricted to already characterized TF [20]. Considering that the validation of candidates by EMSA, ChIP or reporter assays is handy and time-consuming, it is of the upmost importance to carefully select the candidates. For these reasons, as well as to get a chance to identify "novel" regulatory proteins that are not necessarily known as TF, DNA-affinity chromatography protocols to isolate the proteins interacting with a DNA regulatory sequence, followed by an unbiased identification by mass spectrometry, have been developed. Although the concept is very simple, achieving this goal represents a real technological challenge for various reasons that will be discussed below.

2.2. *Challenges of DNA-affinity purification*

Identifying transcription factors by mass spectrometry after DNA affinity purification is a challenge for at least 3 different reasons: the sensitivity and the specificity of the method, as well as the high dynamic range of a complex protein mixture.

2.2.1. *Sensitivity*

TFs are low abundant proteins, estimated to be 10^{-2} to 10^{-5} copies per cell, or 0.01 to 0.001% of total protein content of cells [3,21]. In addition, many TFs need to be activated (by ligand binding, post-translational modification, *etc.*) to gain the capacity to bind DNA, which means that only a proportion of a transcription factor of interest is able to bind DNA. For these reasons, most of the experiments leading to the identification of DNA-captured proteins have required dozens of mg of nuclear proteins [22–25]. In addition, TF identification after DNA-affinity capture has only been made possible thanks to the sensitivity improvement of mass spectrometry technology.

2.2.2. *Specificity*

A broad range of proteins are able to bind DNA in a non-sequence-specific manner, such as proteins involved in DNA maintenance, replication, and repair [26]. Although it is generally considered that these proteins bind DNA with a lower affinity than sequence-specific transcription factors, they are much more abundant, making their capture by the DNA bait very likely.

2.2.3. *Complex mixture with a high dynamic range*

This point is directly linked to both the sensitivity and specificity considerations. Indeed, even if the sequence-specific transcription factors are captured by the immobilized DNA bait, the presence of many other highly abundant proteins that are captured by the bait (or adsorbed on the chromatographic support) in a nonspecific manner might hamper the identification of the sequence-specific transcription factors by mass spectrometry, as illustrated by the group of Jiang [27,28]. Other authors have identified up to 650 different proteins when using a bait as short as 30 bp DNA [29], a still more critical point when studying longer DNA sequences (of over 100 bp). To overcome the problem of a complex protein/peptide mixture containing low abundant TF mixed with hundreds of other proteins of high abundance, a key point will therefore consist of an efficient protein separation, in a 1D or 2D gel-dependent or gel-independent fashion, to allow an optimal protein identification (see Section 4).

With a historical perspective, Fig. 1 illustrates how DNA affinity purification has evolved from totally biased approaches – focused on a highly purified and known DNA-binding protein for purposes of either biochemical characterization, antibody production or gene cloning [21,30,31] – to progressively unbiased approaches allowing *de novo* identification of protein partners bound to a specific DNA sequence. The message delivered by this historical overview is that the identification of proteins captured by a DNA sequence requires to succeed in several consecutive steps, detailed in the next section: the starting capture material, the protein elution, and the mass spectrometry-based identification of these proteins requiring eventually the separation of proteins or peptides. In addition, several strategies have been developed throughout this process to ensure specificity, as much as possible. Each of those steps has been improved over time, as illustrated in Fig. 1, but the relative impact of each improvement

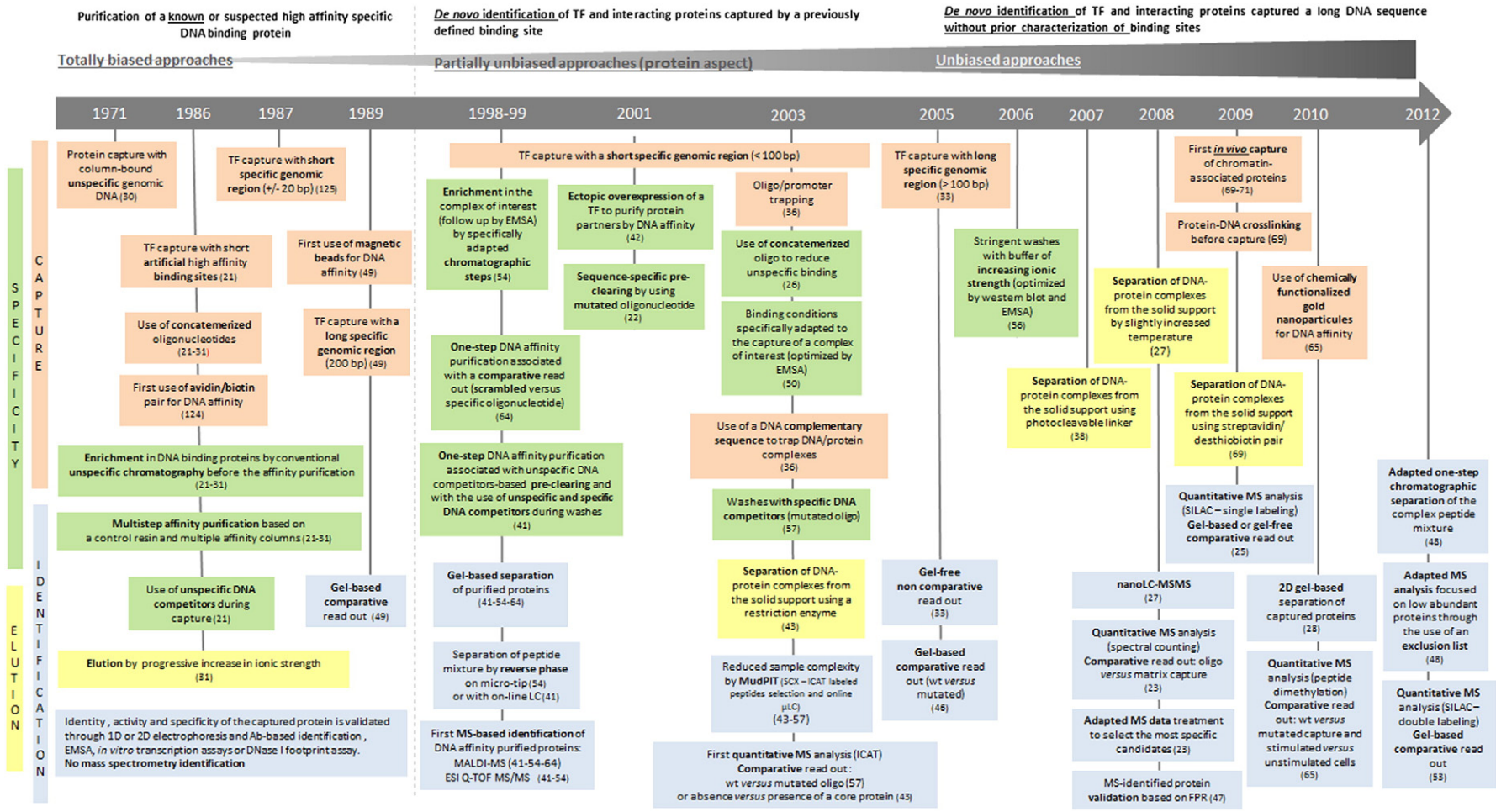

Fig. 1 – Key milestones in the historical development of DNA affinity approaches. The critical steps of the identification of proteins after DNA-affinity capture, *i.e.* the capture, elution, separation and identification of the proteins are colored in orange, yellow and blue, respectively. In addition, the tricks devoted to ensure the specificity of the methods are highlighted in green. For more clarity, only the first representative studies bringing innovative elements, according to us, are represented. Three categories of approaches are emphasized (from left to right): 1) totally biased approaches: both the binding sequence and the protein partners were previously known or suspected. 2) Partially unbiased approaches: the regulatory binding sequence was previously defined, but the identity of interacting proteins is unraveled by MS. 3) Totally unbiased approaches: a large regulatory binding sequence for which precise binding sites were not defined was used to capture a maximum of protein partners without any *a priori* about their identity. Abbreviations: TF: transcription factor; EMSA: electrophoretic mobility shift assay; LC: liquid chromatography; MALDI: matrix-assisted laser desorption/ionization; ESI Q-TOF electrospray ionization quadrupole - time of flight; MudPIT: multidimensional protein identification technology; SCX: strong cation exchange; ICAT: isotope-coded affinity tag; SILAC: stable isotope labeling by amino acids in cell culture; FPR: false positive rate.

is difficult to address as there is only one final read-out (the identity of interacting proteins) conditioned by the success of each consecutive step. For instance, the separation of the DNA–protein complex from the chromatographic support is a major improvement in such strategy (reviewed in [32]). Some researchers have invested a lot of efforts to make the specificity better for the protein capture, but much of these efforts were spilled by an unspecific elution step [33].

Since the end of the nineties, two types of issues related to DNA-affinity have been considered: 1) the identification of proteins captured by a short (≈30 bp) DNA sequence previously shown to be crucial for gene regulation, typically through point mutations-dependent experiments and/or DNase I footprinting assays, and 2) the identification of proteins captured by a long regulatory sequence (more than 100 bp) in order to resolve a promoter interactome in a more "native" context.

Quantitative proteomics is particularly well suited to answer the first kind of question: despite a large number of proteins captured and identified, the transcription factors that specifically bind to the DNA-binding site of interest are the few proteins that are differentially abundant. However, such approaches are somehow biased towards the TFs of interest, since the identification of the DNA-binding site of interest requires pre-existing knowledge based on heavy experimental work. On the contrary, when trying to identify a maximum of proteins that bind to a long DNA sequence, quantitative proteomics cannot be used, as there is no suitable "control bait" (see Section 4.2.1.) [25]. The challenge of combining sensitivity, specificity and a highly complex mixture with a high dynamic range is thus even more crucial for this second type of biological question which is more prone to generate false negative and false positive results. The candidates emerging from such analyses require, more than ever, biological validations.

2.3. *Key parameters in the purification of in vitro-assembled DNA–protein complexes*

Capturing the sequence-specific interacting proteins probably represents the major difficulty, due to the low abundance of specific TF and the high abundance of proteins that bind to the bait in a non-sequence specific manner. Several key parameters can be considered to overcome this problem (summarized in Table 1).

2.3.1. *Design of the oligonucleotide probe*

Due to the presence of a blunt DNA extremity on the bait, proteins involved in the DNA repair like the abundant DNA-dependent protein kinase (DNA-PK) are easily captured, possibly hindering the identification of low abundant specific TF [26]. Concatemerized oligonucleotides can be used to reduce the ends-to-binding site ratio, as well as the use of short unspecific DNA competitors [34,35]. Even if largely used since the early times of DNA affinity purification [21], the concatemerization of a DNA sequence presents some putative drawbacks [33,36]. Indeed, it might introduce novel DNA binding sites, thereby increasing the probability of false positive candidates during the capture step. Furthermore, this strategy is excluded with long DNA sequences. Another possibility to reduce unspecific capture linked to the blunt end of the bait is to use chemically modified oligonucleotides, like thiol group-modified oligonucleotides [37] or cyanine-modified oligonucleotides (personal unpublished data). Additionally, one should pay attention to maintain some space between the sequence of interest and the solid support, to avoid steric hindrance [38].

The length of the probe is another important element to take into consideration, as the analysis of long DNA sequences by DNA-affinity chromatography offers several advantages. First, it does not require deep promoter dissection that necessitates time-consuming experiments. Second, it takes into consideration the possibility that protein binding to DNA can be cooperative, the binding of one protein on a sequence being dependent on the presence of another protein bound nearby or at some distance [39,40]. In addition, co-activators or co-repressor recruitment might depend on the presence of several transcription factors bound to DNA, making this second layer of transcriptional co-regulators barely accessible to studies based on short DNA baits. Indeed, as discussed in [25], the purification of DNA-binding protein complexes, requires oligonucleotide of a certain length to ensure a minimal genomic context and thus a biological relevance.

2.3.2. *Promoter trapping or promoter immobilization*

One can distinguish two protein capture strategies: i) the proteins can be incubated with the DNA bait in solution before trapping of the DNA bait on the solid support, a method called "promoter trapping" and detailed below, or ii) the DNA bait is first immobilized on the solid support before incubation with the protein extracts. This latter approach, called "promoter immobilization", represents the large majority of DNA-affinity based studies. While the early studies used oligonucleotides covalently linked to sepharose columns through CNBr activation [21], most studies are now based on biotinylated oligonucleotides immobilized on paramagnetic streptavidin-coated beads, among others [9,24,25,29,34,38,41–49]. The biotin–(strept)avidin interaction is so strong (with a Kd of 10^{-13} M) that it can be considered as equivalent to a covalent link, and is much more convenient to use than the hazardous CNBr. In addition, chemical coupling might provoke modifications in some nucleotides of the sequence [50]. When compared with chromatographic columns, paramagnetic beads are easy to handle and offer the advantages of large incubation/wash volumes, and small elution volumes, thereby avoiding further lyophilisation processes before MS analyses.

Theoretically, the interaction between proteins and DNA should be favored when incubated in solution, with respect to the situation in which the DNA is immobilized. This was confirmed with a GFP-tagged TF (the C/EBP for CAAT-enhancer binding protein) that is less efficiently captured by a short (about 35 bp) oligonucleotide immobilized on a sepharose column, than when the bait is kept in solution [50]. However this has not been observed by the group of Praseult with paramagnetic beads [38]. One may hypothesize that in the case of paramagnetic beads, as the binding step takes place for hours in the presence of a mild agitation, the equilibrium can be reached more easily than when the bait is immobilized on a column.

The group of Jarrett has set up the promoter trapping method [51], characterized by a DNA–protein binding step in

Table 1 – Comparison of the major features, advantages and limitations of DNA- and RNA-affinity chromatographies.

Purification step	DNA-affinity		RNA-affinity	
	Characteristics	Advantages/limitations	Characteristics	Advantages/limitations
Bait sequence (length, concatemeriz., spacer)	• Mostly short (<40 bp) DNA sequences. Long (>100 bp) DNA sequences are unfrequent. • Short DNA sequences may be concatemerized. • A minimal spacing sequence must be inserted between the support and the binding site of interest [38].	• Cooperative DNA binding [39,40] and higher order complexes formation (containing coactivators/corepressors) [25] require long DNA sequences. • The longer the DNA sequence, the more complex is the protein mixture. • Concatemerization [33,36] might introduce artificial DNA binding sites.	• Mostly several hundreds of nucleotides (to allow the formation of RNA secondary structures required for protein binding). • Unfrequently: short RNA sequences (22 nt to 45 nt).	• The longer the RNA sequence, the more complex is the protein mixture, making the identification of low abundant proteins more difficult.
Bait synthesis	Synthesized by PCR	• Modified primers are useful to link the bait to the chromatographic support. • The unspecific capture associated with the blunt end of the bait is limited with chemically modified oligonucleotides [37].	• *In vitro* synthesized (biotinylated or not [83,87,92]). • *In vitro* transcribed. • *In vivo* transcribed.	The bait is sensitive to RNAse.
Chromatog. support	• Chromatographic columns • Paramagnetic beads	Advantages of paramagnetic beads: • easy to handle • large incubation/washes volumes and small elution volumes • the equilibrium is reached more easily [38] with beads under agitation.	Idem DNA affinity	
Bait immobilization (must be considered together with the different elution possibilities)	• Mostly through biotin/streptavidin interaction (using biotinylated primers captured by (strept)avidin-coated chromatographic columns or beads) [9,24,25,29,34,38,41–49]. • Occasionally covalent links after chemical coupling [21]. • Promoter trapping: antisense oligonucleotides bound to the chromatographic support are used to fish the DNA/protein complexes [27,50].	• The biotin–(strept)avidin interaction is robust and convenient to use. • Chemical coupling might provoke modifications in some nucleotides of the sequence [50].	• Chemical modifications of RNA [90,92,93]. • Hybrid nucleotidic probes, with biotinylated deoxyribonucleotides linked to the ribonucleotide bait [87,98]. • Tags introduced in the RNA sequence: aptamers (defined sequences able to bind to molecules with a high affinity and specificity [99]) or polyA (for recovery with oligo(dT)-beads [86]). • Antisense oligonucleotides bound to the chromatographic support are used to fish the RNA bait [102–104].	• Chemical modifications of nucleotides or incorporation of a foreign sequence in the RNA bait may modify the RNA structure and the RNP complex composition. • The recovery yield of the RNP complex might be low with aptamers (although not always defined) but could be improved with scaffolding technology [89]. • The RNA fishing with antisense oligonucleotide does not require any RNA modification (chemical or sequence) or cell transfection (for *in vivo*-assembled complexes) but is restricted to RNA with accessible regions.
Protein extract	• Nuclear extracts • Possible cross-linking with DNA (for *in vivo*-assembled complexes).	More stringent washing conditions can be used when proteins and DNA are cross-linked, thereby limiting the presence of non-sequence specific	• Cytosolic or nuclear extracts, depending on the biological question. • Possible cross-linking with RNA (for *in vivo*-assembled complexes).	More stringent washing conditions can be used when proteins and RNA are cross-linked, thereby limiting the presence of non-sequence specific

		interacting proteins, and making the need for specific elution less crucial.		interacting proteins [106], and making the need for specific elution less crucial.
Pre-clearing/ blocking	• Blocking agents (casein, BSA) to saturate the solid support. • Unspecific pre-clearing by (successive) chromatography steps and selection of the fraction retaining a DNA-binding activity. • Unspecific pre-clearing in the presence of the solid support, naive, coated with biotin or with unrelated DNA sequences. • Specific pre-clearing in the presence of mutated oligonucleotides.	• Blocking proteins increase the noise during the MS-based identification step. • Pre-clearing is useful to decrease the complexity of protein mixture. • Successive chromatographies are time-consuming, require large amounts of material and are not always compatible with the identification of large protein complexes. • Unspecific pre-clearing represents a risk to lose the specific TFs. • Specific pre-clearing is limited to the study of well-characterized binding sites. • The binding conditions should be set up individually for each binding site; this is not applicable to uncharacterized sequence. • DNA competitors decrease the non-sequence specific binding, but they might also capture the proteins of interest when used at high concentration [33,34,51].	• Blocking of the immobilized bait with high concentrations of unspecific RNA (yeast tRNA), proteins (BSA) [85], heparin, and/or salmon sperm DNA [91]. • Unspecific pre-clearing by (successive) chromatography steps [86,90,91]; selection of the fraction retaining a RNA-binding activity. • Unspecific pre-clearing in the presence of fresh beads (not linked with RNA) [83,92], egg white avidin and/or yeast RNA [89].	• Blocking agents increase the noise during the MS-based identification step. • Pre-clearing is useful to decrease the complexity of protein mixture. • Successive chromatographies are time-consuming, require large amounts of material. • Unspecific pre-clearing represents a risk to lose the specific TFs. • Adsorbed blocking molecules might be eluted together with RNP complexes.
NA–protein binding and washes conditions	Binding and washing buffer compositions could be adapted: • Ionic stringency • Detergent concentration • Unspecific competitors: polydI-dC, salmon sperm DNA • Specific competitors: scramble sequence, mutated oligonucleotide, …		Binding and washing buffer compositions could be adapted: • Ionic stringency • Detergent concentration • Unspecific competitors: RNAs [95], negatively charged molecules (heparin) [81,94]. • Sequence specific competitors	• RNA competitors decrease the non-sequence specific binding, but they might also capture the proteins of interest. Therefore, the use of dedicated RNA sequences is advised [87].
Elution	*Unspecific*: salt, temperature, detergent *Specific:* 1) restriction enzyme 2) photocleavable biotin 3) displacement of desthiobiotin with biotin. 4) annealing of an $(AC)_5$ tail to its complementary $(GT)_5$ sequence on the bait; elution by a mild temperature elevation (37 °C) in low salt buffer conditions [27,50].	*Unspecific* elution: the proteins adsorbed on the solid support contaminate the eluate (might be affordable with a strong comparative control downstream). *Specific elution:* 1) restriction enzyme. High recovery yield. Requires to modify the DNA sequence of the bait. The restriction enzyme contaminates the DNA/protein complexes and could increase the noise of the sequencing analysis. 2) photocleavable biotin. Recovery yield of 75%. The use of intense UV light might induce covalent cross-linking between some proteins and the DNA sequence [68]. 3) displacement of desthiobiotin with biotin. High recovery yield (100%). A high concentration of free biotin is collected with the protein–DNA complexes and should be eliminated by an additional incubation with naive streptavidin-coated beads % [48]. 4) annealing of an $(AC)_5$ tail to its complementary $(GT)_5$ sequence requires to modify the sequence of the bait, but the DNA–proteins complexes are eluted without inducing the release of proteins adsorbed to this support.	*No elution*: on bead digestion of RNP *Unspecific*: salt, temperature, detergent, urea. *Specific*: 1) displacement of aptamers with an excess of their ligand molecule (ex: biotin [95]) 2) insertion of a cleavage sequence in the fishing device [106]	*No elution*: 17–26% of identified proteins are proteins adsorbed on the beads [106]. *Unspecific* elution: the proteins adsorbed on the solid support contaminate the eluate (might be affordable with a strong comparative control downstream). Several authors use a step gradient of salt to elute the RNP and select for MS analysis the fraction presenting RNA-binding activity [86,90,93]. *Specific elution:* is strongly advised to limit the number of adsorbed proteins identified, but this requires to modify the bait sequence.

solution, a capture of the DNA–protein complexes based on the specific interaction between complementary strands (using an $(AC)_5$ –sepharose column and DNA probes with a $(TG)_5$ tail on both strands) and by specific conditions regarding the DNA-protein interactions. These are based on the concept that TF can bind with low affinity (μM) to any DNA sequence – which would be part of the sliding model of TF–DNA binding [52] – and with high affinity (nM–pM) with their specific binding site. Therefore, using low concentrations (nM) of oligonucleotide, typically 10× Kd, ensures that most of the proteins captured by the bait are the specific transcription factors [50]. This purification strategy, associated with a 2D gel electrophoresis and nLC–MSMS (nano liquid chromatography–tandem mass spectrometry) analysis, has been successfully applied to identify USF-2 (upstream stimulatory factor-2) as an E-box binding factor in the human telomerase reverse transcriptase gene promoter [28] and was recently adapted by [53] in a quantitative proteomic approach. Although theoretically elegant, a major drawback of such strategy is that it requires careful characterization by EMSA of the interaction between TF and the DNA binding site, in terms of DNA concentrations and buffer composition (competitors, modifiers like heparin, detergent, *etc.*). Moreover, this cannot be applied to long DNA sequences for which the key response element(s) has(ve) not been defined yet.

2.3.3. *Specificity of the capture and/or elution of transcriptional regulators*

2.3.3.1. Enrichment in (specific) DNA-binding proteins. The first DNA affinity purification approaches were based on multiple unspecific pre-clearing chromatographic steps to enrich the sample in DNA-binding proteins prior to the final specific DNA affinity capture. Negatively charged proteins were eliminated through ion exchange chromatography [54] while positively-charged proteins were enriched through ion exchange chromatography [54] or affinity chromatography for heparin-sepharose [55] or phosphocellulose columns [35]. Such unspecific enrichment strategies were eventually refined towards the complex of interest by monitoring the DNA-binding activity of the complex in the different eluted fractions, by EMSA for instance. Only the fractions positive for this assay were conserved to pursue the enrichment procedure [22,34,55]. Laborious and time-consuming, such pre-fractionation of the nuclear extract requires huge amount of starting material (up to several hundreds of mg of proteins), as well as large elution volumes that require additional lyophilisation treatments prior to the MS-based identification. Such material handling is prone to lose low abundant proteins thereby affecting the output [35]. Moreover, some pre-fractionation strategies are not necessarily compatible with the identification of large protein complexes, as the different partners might be eluted in different fractions.

The rationale of above pre-clearing steps is based on the negative charges of the DNA bait, and the consequent overall positive charges of interacting proteins. However, other considerations have also been taken into account such as the unspecific adsorption of the proteins on the chromatographic support. These aspects led to the use of a pre-clearing step of the nuclear protein extract in the presence of the solid support, either naive [23,56], or covered with biotin moieties [45], with $(AC)_5$ tail [51], with unrelated DNA sequences like calf thymus DNA [23] or with scramble oligonucleotide [55]. The principle of unspecific pre-clearing steps has progressively evolved towards sequence-specific pre-clearing in the presence of the immobilized DNA sequence mutated in the binding site of interest [22,57]. Pre-incubation in the presence of mutated oligonucleotides was also used in the strategy called NAPSTER (Nucleotide-Affinity Pre-incubation Specificity Test of Recognition) [58]. This protocol was used among others to identify YB-1 (Y-box binding protein-1) as a specific binder of an AP-1 (activator protein 1) binding site [46]. More recently, it has been used to confirm interactions suggested by other techniques such as ChIP [59,60]. Although useful to overcome the challenge of complex mixture analysis, these various pre-clearing steps represent a non-negligible risk to lose the specific TFs, as illustrated by the pull-down of ERa adsorbed on agarose beads [61]. They should be avoided, as much as possible, except in the case of mutated oligonucleotides that clear out the unspecific proteins and allow to focus on the MS identification of proteins that specifically interact with the DNA binding site of interest, but such strategy cannot be applied to uncharacterized sequences.

2.3.3.2. Binding and wash conditions. The binding buffer used to allow the interactions between DNA and proteins represents a key point in the success of DNA-affinity chromatography. The pH as well as the concentration in salts, detergents, EDTA and mono/bivalent metallic cations are important to consider [21] and should be as close as possible as physiologic nuclear conditions [23]. A stringent binding buffer, containing moderately high concentrations in salts or detergents will limit the binding of unspecific proteins, but might also hamper the binding of specific TFs and/or provoke the destabilization of multi-protein complexes. The redox conditions must also be taken into consideration, as the binding capacity of several TFs is conditioned by the formation of disulfide bonds [62]. Similarly, the composition of the washing buffer, particularly in terms of ionic strength, can be more or less stringent, directly affecting the number of non-sequence specific interactions [56].

When studying a precise binding site, this problem can be solved by optimizing, step by step, the composition of the binding buffer and its impact on the DNA-binding capacity assessed by EMSA. The ideal binding buffer is often slightly different from one tested TF to another, as attested by the large diversity of binding conditions used for EMSA in the literature. Once the binding conditions are optimized, they can be transferred to DNA-affinity chromatography [34,57]. However, such systematic optimization is not possible when studying the proteins that bind a DNA sequence of several hundreds of base pairs, thereby increasing the risk of false negative. Using an ELISA-like assay for DNA-binding activity [63], we have systematically assessed the effect of binding buffer composition on the DNA-binding activity of 23 different transcription factors to their consensus sequence. We came to the conclusion that a binding buffer containing 120 mM NaCl, 0.44 mM EDTA, 0.8 mM $MgCl_2$ and 6% glycerol (final concentrations) is consensual for most of the transcription factors tested (unpublished personal data).

The washing step could also be adapted to remove a maximum of weakly DNA-bound proteins, with a progressive increase in the salt concentration of the washing buffer [41,56] or with an excess of specific or unspecific DNA competitors as explained below. A combination of different kinds of washing conditions could also be used to ensure a better result [41].

2.3.3.3. *Use of DNA competitors*. In a majority of DNA affinity procedures, unspecific DNA competitors such as *Escherichia coli* double or single strand DNA, salmon sperm DNA or poly (dI:dC) are used during a pre-clearing step, during the binding reaction and/or during washes [9,22,28,33,41,48,53,55]. Through the capture of proteins that display a general affinity for nucleic acid molecules or free DNA ends, the use of DNA competitors reduces efficiently the unspecific noise, thereby improving the capacity for identifying low abundant proteins. However, in this situation, the risk encountered is that DNA competitors might also capture some of the proteins of interest, particularly when used at high concentration, and thus compete for the protein of interest [33,34,51]. Therefore, although unspecific DNA competitors are necessary, their use at high concentrations (above 0.1 mg/ml) should be considered with caution. One can also emphasize that a combination of different competitors could be interesting to capture a larger variety of unspecific DNA binding proteins [51].

To counteract this problem, several authors use sequence-specific DNA competitors that consist of fragments derived from the sequence of interest digested with a restriction enzyme [41] or, more frequently, in the same sequence than the bait mutated for the binding site of interest [22,46]. Used at high concentration, these mutated baits capture high abundant proteins characterized by a low affinity for DNA, but not the specific transcription factors [24,57].

2.3.3.4. *Elution of interacting proteins*. Classically, DNA-interacting proteins are eluted with denaturing solutions, whether these are based on salts, detergents or heat accompanied by reducing agents [22,34,41,42,44–46,53,61,64,65]. Although efficient to free the proteins from the DNA sequence, such methods present a major drawback: the proteins that are adsorbed on the solid support in an unspecific manner are also exposed to denaturation and thus contaminate the eluate. For these reasons, several authors use blocking agents such as casein or BSA (bovine serum albumin) to saturate the solid support, but these highly abundant blocking proteins also increase the noise during the MS-based identification step [42,43]. Proteins containing a biocytin group that interacts with streptavidin-coated supports are also denatured and thus contaminate the suspension of proteins eluted from the DNA bait. To limit this problem, streptavidin-coated magnetic beads functionalized with the biotinylated oligonucleotide can be incubated with free biotin to saturate the streptavidin binding sites [66]. Alternatively, solid support characterized by a low unspecific binding, such as NeutrAvidin beads [53] or gold nanoparticles [61] are used. It has to be underlined that despite all these cautions, the presence of background contaminants cannot be completely avoided, and that the identity of contaminant proteins depends on numerous parameters such as the cell type, the DNA sequence, the chromatographic support, the binding/washing buffer composition, and the elution. It is thus very difficult to determine whether a protein is a background contaminant or not, but it might be helpful for the researcher to consult the CRAPome (www.crapome.org), an online resource aggregating the proteins identified by mass spectrometry in the negative controls of protein-affinity purifications [67]. Although the contaminants might differ after protein-affinity or NA-affinity purification, at least, the influence of the chromatographic support and the type of subcellular fractionation could be estimated thanks to this database.

However, to avoid the elution of unspecifically adsorbed proteins, we strongly recommend the use of a strategy for selective separation of the DNA–protein complexes from the solid support. This topic, already reviewed [32], is briefly summarized below.

Using a DNA bait bound to the paramagnetic beads through a photocleavable linker, the group of D. Praseuth estimated that 75% of the immobilized oligonucleotide-bound tetracycline repressor protein is recovered in the supernatant after UV exposure [38]. Although this is an indirect way to estimate the recovery of the DNA–protein complexes (some proteins could directly bind to the beads), this suggests that 25% of the proteins captured by the oligonucleotide are lost for MS identification. Moreover, it has to be mentioned that the use of intense UV light to separate DNA/protein complexes from the solid support could also induce covalent cross-linking between some proteins and the DNA sequence [67], making those proteins inaccessible for trypsin digestion and further MS-based identification (personal observation).

Using DNA/protein complexes trapping through the specific annealing of an $(AC)_5$ tail to its complementary $(GT)_5$ sequence on the bait, the promoter trapping strategy developed by [27,50] uses a mild temperature elevation (37 °C) in low salt buffer conditions to elute DNA–proteins complexes from the solid support without inducing the release of proteins that are nonspecifically bound to this support.

Several groups have used restriction enzymes to release the protein–DNA complexes from the solid support [25,43,53]. In this case, the bait sequence is modified to contain a specific restriction site, separated from the moiety in interaction with the solid support by a spacer. This strategy was shown to efficiently and specifically recover the bait of interest without contaminant proteins [25]. However, several drawbacks of this approach have to be mentioned. First, one cannot exclude that the sequence modified with the inserted restriction site might capture proteins that would not bind to the unmodified sequence, thereby generating false positive candidates. Second, a non-negligible amount of recombinant restriction enzyme used to release the oligonucleotide is collected together with the DNA/protein complexes. Once digested and analyzed by mass spectrometry, the restriction enzyme-derived peptides could increase the noise of the sequencing analysis. While such consideration can be overcome in the case of comparative or quantitative proteomics strategies like SILAC (stable isotope labeling by amino acids in cell culture) (explained in Section 4.2.3.), this is of importance in the case of unbiased approaches.

More recently, the desthiobiotin–streptavidin pair has been successfully used to selectively elute the DNA–protein complexes. As desthiobiotin has a much lower affinity for streptavidin than biotin (with Kd of 10^{-5} M and 10^{-13} M,

respectively) [68], a biotin excess can displace the desthiobiotinylated oligonucleotide–protein complexes. First reported in 2002 to purify proteins [68], this strategy was adapted in 2009 by Déjardin and Kingston [69] to purify oligonucleotide-bound chromatin/proteins complexes and more recently by us to purify *in vitro*-assembled DNA–protein complexes [48]. One disadvantage of this strategy is that a high concentration of free biotin is collected with the protein–DNA complexes that might hamper the peptide separation and identification. Ideally, an additional incubation with naive streptavidin-coated beads has to be inserted in the process to capture the free biotin excess. On the other hand, this strategy offers the considerable advantage of a high recovery yield: although the binding efficiency of desthiobiotinylated oligonucleotides to streptavidin-coated beads is fairly low (50%), the recovery of the DNA–protein complexes after the displacement reaction has been estimated to be around 100% [48], indicating that all the proteins captured by the bait are likely to be involved in the next steps of analysis leading to protein identification.

2.4. *Purification of in vivo-assembled DNA–protein complexes*

All the DNA-affinity approaches described above identified protein/ADN complexes in an *in vitro* context while it is more relevant to characterize such interactions in an *in vivo* context, taking into account the chromatin environment and the exact nucleoplasm composition. However, defining the protein composition of an *in vivo*-formed DNA–protein complex represents a technological challenge in terms of sensitivity, as most DNA sequences of interest do exist in a limited number of copies per cell (usually 2 copies for human cells). Therefore, so far, the successful identifications by mass spectrometry of proteins interacting *in vivo* with a DNA sequence are restricted to DNA sequences present in multiple copies in each cell, *i.e.* on mitochondrial DNA (reviewed in [70]), on a transfected low copy plasmid [71], or on repetitive DNA regions like telomeric sequences [69].

The identification of proteins interacting with mitochondrial DNA (mtDNA) is not representative of the difficulties inherent in the identification of proteins interacting with a defined nuclear genomic locus. Indeed, mtDNA is a relatively small size (16 kb in humans) circular genome, packed in a nucleoid form, located in an organelle that can be relatively easily isolated and purified by centrifugation, and, more importantly, present in thousands of copies per cell. MtDNA nucleoids can therefore be biochemically purified, either under their native form, or after exposure to formaldehyde to induce DNA–protein crosslinks (reviewed in [70]). The second approach authorizes the use of stringent conditions, in terms of salt and detergent concentrations, to remove possible contaminating proteins and low-affinity interacting proteins. Altogether, biochemical purification of native or formaldehyde cross-linked nucleoids followed by mass spectrometry analysis has lead to a layered model of mtDNA-interacting proteins (reviewed in [72]).

Regarding the identification of proteins interacting with a defined nuclear genomic locus, the challenge lies in the purification of low abundant protein/DNA complexes. Therefore, Lee's group developed a strategy called "DNA sampling" to isolate the different proteins bound to a specific plasmidic DNA region (the promoter of the colicin K gene) directly from *E. coli* [71]. For this purpose, they designed a low expression DNA plasmid containing the sequence of interest (594-bp long) adjacent to binding sites for the LacI repressor and circumscribed by sensitive sites for a specific nuclease. When the nuclease expression is induced, the DNA fragment of interest is cut from the plasmid DNA, and purified by FLAG-affinity capture targeting the co-expressed FLAG-tagged LacI repressor. After unspecific elution, captured proteins were resolved by SDS-PAGE. Although the authors successfully applied this method to several different DNA regions, they underlined some shortcomings: i) such strategy requires to transfect 3 constructs, and thus cannot be used on any cell type, and the ratio between the low expression plasmid and the FLAG-tagged LacI repressor must be tightly controlled; ii) it probably generates false negative in the category of ligand-dependent activated DNA-binding proteins and iii) false positive candidates are also suspected due to the systematic presence of protein contaminants captured by the solid support [71]. Despite these shortcomings, we think that this promising method would deserve further improvement, particularly by including a specific elution step to avoid contamination by the proteins adsorbed in an unspecific manner.

Déjardin and Kingston developed an original purification scheme allowing elucidation of locus-specific composition in an endogenous context, without the need of genetic engineering, by identifying proteins associated to a specific genomic locus in eukaryotic cell chromatin through formaldehyde DNA–protein cross-linking. This method, called PICh (Proteomics of Isolated Chromatin segments) uses desthiobiotinylated LNA (locked nucleic acid) probes to specifically isolate the chromatin fragment of interest with a reduced nonspecific protein contamination. Free biotin excess is then used to displace desthiobiotin (see point 2.3.3, elution of interacting proteins) and to elute the DNA–protein complexes specifically. After reversal of the cross-link, proteins are resolved on SDS-PAGE and submitted to MS analysis for identification. Even if successfully applied to human telomere-associated protein complex identification, so far, the developed method has not yet demonstrated sufficient sensitivity to identify *in vivo* chromatin-interacting proteins when the locus of interest is unique. Déjardin and Kingston have estimated that several hundred liters of cell culture would be necessary to provide a sufficient amount of material to sequence the proteins captured by a unique genomic locus [69]. Clearly, this strategy is interesting, and is ready to be declined in a multiplexing version called GENECAPP (Global ExoNuclease-based Enrichment of Chromatin-Associated Proteins for Proteomics) [73], but it still needs further refinements to be applied to unique genomic loci.

3. RNA-affinity purification

Although traditionally less emphasized than DNA–protein interactions, RNA–protein interactions are of great biological importance and complexity. Indeed, RNA-binding proteins (RBPs) control the processing, localization, decay and translation level of mRNAs [74]. In addition, RNA molecules can adopt a

large variety of conformations, that can be modified upon interaction with proteins in an "induced fit" manner, thereby allowing the recruitment of additional protein partners [75]. The protein–RNA interactions are thus highly dynamic and difficult to predict due to the versatility of the RNA molecule structure. RBPs are estimated to number over 600 genes in yeasts, and 2500 in mammals [76], and recent data suggest that these numbers are underestimated, as proteins that were not predicted or annotated as RBPs were shown to associate with specific RNAs [77]. While some RBPs are thought to bind RNA with low (or no) sequence-specificity, most of these RBPs specifically bind to particular subpopulations of mRNAs [77].

3.1. *Challenges of RNA affinity purification*

Most of the assays available to investigate protein–DNA interactions, when the partners are already identified or at least suspected, have been adapted to study protein–RNA interactions, such as EMSA, UV-crosslinking, RNA-immunoprecipitation, PAR-CLIP (PhotoActivated-Ribonucleoside-enhanced CrossLinking and ImmunoPrecipitation), a method to identify the binding sites of a defined RBP in a transcriptome-wide manner [78]. Readers interested by method reviews will consult [79,80]. Several methods also exist to map interactions between RNA and proteins [75,81]. Although such methods can provide useful information regarding the location of the RNA sequence interacting with proteins, or the size of the interacting proteins, they do not provide the identity of proteins of interest. For these reasons, RNA-affinity purification procedures of proteins have been developed. Although several different strategies have been developed, the general scheme drawn for DNA-affinity purification of proteins can be applied to RBP analysis.

Briefly, in most publications, a RNA bait is immobilized on a solid support, and the ribonucleoprotein (RNP) complexes are *in vitro* assembled, in the presence of cell/tissue lysates [82–94]. The formed RNP can be either eluted in native conditions or after the formation of cross-links consecutive to either UV exposure or formaldehyde treatment. Another strategy consists of purifying *in vivo* assembled RNP complexes, usually thanks to the addition of special tags to the RNA sequence of interest. Although closer to the physiological situation, these latter approaches also present some drawbacks, like the requirement for cell transfection. It has to be underlined that cross-linking can also be applied to *in vivo*-assembled RNP [95]. One of the clear advantages of RNP cross-linking, either used on *in vivo* or *in vitro*-assembled complexes, is to freeze the composition of RNP complexes, allowing the use of stringent washing conditions, thereby strongly reducing the capture of non-specifically binding proteins.

3.2. *Key parameters in the purification of* ***in vitro*** *assembled RNA–protein complexes*

3.2.1. *Design and immobilization of the RNA bait*

On the contrary to DNA-affinity purification, most of the publications deal with long RNA baits (up to several hundreds of nt) to purify interacting proteins, a feature linked with the requirement for RNA to adopt secondary structures to capture proteins. Those long RNA baits are produced by *in vitro* translation of a linearized plasmid. On the contrary, few studies focused on short RNA sequences (22 nt to 45 nt) or used synthetic RNA, biotinylated or not [83,87,92].

Different protocols to immobilize the RNA bait on a solid support have been developed, each one bearing limitations. As this has been reviewed by [96], we will only summarize the basic principles here below. The immobilization can be based on chemical modifications of RNA or on tags introduced in the RNA sequence.

First, the RNA bait can be covalently linked to the solid support, as is the case for oxidized RNA linked to adipic acid dihydrazide agarose beads [91,92] or for cyanogen-activated sepharose beads [89]. However, these chemical modifications of nucleotides might affect the secondary structure of RNA and consequently the subsequent captured proteins. Second, the RNA bait can be immobilized through non-covalent – although robust – chemical interactions with the solid support. RNA can be chemically tagged during *in vitro* transcription, thanks to the incorporation of rNTPs modified with different moieties like biotin or digoxygenin (see for instance [84]). Other strategies consist of using hybrid nucleotidic probes, with biotinylated deoxyribonucleotides linked to the ribonucleotide bait [86,97], or to biotinylate the 3′end of *in vitro* transcribed RNA bait. Alternatively, the RNA bait can be polyadenylated *in vitro*, for subsequent recovery with oligo(dT)-sepharose beads [85].

Inspired by the multiplicity of useful protein tags, different tags for nucleic acids have been generated. A selected RNA motif, called an aptamer, can be incorporated during transcription, whether this occurs *in vitro* or *in vivo*. An aptamer is defined as a RNA (or DNA) sequence able to bind to small (or large) molecules, with a high affinity and specificity (reviewed in [98]). The inserted motif is chosen according to its high affinity for a defined protein (called hereafter a "fishing" protein), used to immobilize the RNA bait on a solid support. Affinity tag aptamers may have a natural origin, like the stem loop RNA motif U1hpII presenting a high affinity for a protein (U1A) [99] or the RNA motif binding the bacteriophage RNA binding protein MS2 [100]. Synthetic RNA aptamers have also been defined after *in vitro* selection procedure (SELEX for Systematic Evolution of Ligands by EXponential enrichment) for their high affinity towards tobramycin [82], sephadex or streptavidin [96]. However, because of the high affinity between the fishing protein and the RNA motif, the insertion of a cleavage site in the RNA sequence may be required to release the native ribonucleoprotein complexes from the fishing protein coupled to the solid support, after the TAP-tag purification principle [100].

However, using this type of aptamers does present some limitations. First, in the case of *in vivo*-assembled complexes (see point 3.3), it requires cell transfection to express the RNA of interest genetically modified with the aptamer. Second, the incorporation of a foreign sequence in the RNA bait may modify the RNA structure, and consequently the RNP complex formation and composition. Finally, the sensitivity of the assay may be hampered by a relatively low recovery of the RNP complex formed. Indeed, Srisawat and co-workers showed that only 20–25% of the RNA, genetically modified with the streptavidin-affinity aptamer, is recovered after elution with an excess of biotin [101].

Nevertheless, the fact that the RNP complexes are eluted in native conditions, thus allowing further activity assays as shown for RNAse P [101], may represent a considerable advantage. In addition, the recent developments in scaffolding technology will probably improve the efficiency of RNA aptamer-based purification strategies, as suggested by Iioka and co-workers. These authors used a streptavidin-affinity aptamer, scaffolded to a tRNA to improve its stability, to pull down the proteins captured by a RNA sequence of interest. As a proof of concept, these authors showed, by immunoblotting, that this strategy purified the Fragile X Mental Retardation Protein interacting with a specific RNA motif more efficiently (10×) than the nonscaffolded streptavidin-affinity aptamer, by a standard biotinylated-RNA pull down assay. This improvement was attributed to the scaffold-dependent stabilization of the secondary structure of the RNA [88].

Similarly to the trapping of DNA–protein complexes through the annealing of $(AC)_5$ bait tail to the immobilized (GT)5 complementary sequence, developed by the group of Jarrett [50,51] (see Section 2.3.2.), several authors have also taken advantage of the base pair complementarity of nucleic acids to immobilize RNA sequences, as demonstrated for oligo(dT)-sepharose beads to capture poly(A)-tailed RNA [85]. More generally, antisense oligonucleotides can be easily bound to a chromatographic support, for instance through a streptavidin–biotin interaction, to fish the RNA bait of interest, whether synthesized *in vitro* or *in vivo*. The RNP complexes can then be eluted in their native form, by using an excess of the fishing oligonucleotide, or simply by denaturation. This method has the advantage that it suits to cells that are difficult to manipulate genetically, and that the RNA of interest is not modified, neither chemically nor in the sequence itself. However, this approach is restricted to RNA species bearing accessible regions that can be defined by prior RNAse H digestion assays. This method was first developed to selectively isolate U4/U6 small nuclear ribonucleoprotein particles (snRNPs) from HeLa nuclear extracts [102], then adapted to isolate telomerase [103], and more recently, was used to identify the proteins composing the snoRNP complex MBII-52 [104].

3.2.2. *Specificity of the capture and/or elution of RNPs*

The major difficulty in the identification of proteins composing RNP complexes probably lies in the specificity of the capture, the challenge regarding sensitivity being relatively less crucial than for DNA-affinity, as RNA binding proteins are usually considered to be relatively abundant proteins [74]. The specific capture and identification of RNP complexes-contained proteins involve several steps, detailed below.

3.2.2.1. *Blocking/pre-clearing*. The chromatographic support functionalized with the RNA bait can be blocked prior to the incubation with protein lysates with high concentrations of unspecific RNA (typically yeast tRNA), proteins (typically BSA) [84], heparin, and/or salmon sperm DNA [90]. Alternatively, other authors clear the protein extracts by adding fresh beads (not linked with RNA) [82,91], or egg white avidin and yeast RNA, to block endogenous biotinylated proteins and non-specific RNPs [88]. However, if the adsorbed BSA or avidin is eluted together with RNP complexes, this represents an important problem for further protein identification by mass spectrometry.

2.2.2.2. *Enrichment in proteins of interest*. The immobilized RNA bait is usually incubated with nuclear extracts or with cytosolic extracts, depending on the biological question related to the RNA of interest. Alternatively, whole cell extracts can be fractionated, either by gel filtration [90], by affinity chromatography for a heparin-sepharose column [85] or by progressive ammonium salt precipitation [89], and the fraction used for RNA-affinity purification is selected according to its ability to retain a RNA-binding activity, visualized by RNA EMSA or by UV-cross-linking experiments.

3.2.2.3. *RNA competitors*. As in the case for DNA-affinity, unspecific RNA competitors in the binding buffer are used to limit the capture of proteins with an overall affinity towards RNA and negatively charged molecules. For instance, large quantities of unspecific competitor RNAs (twice the amount of the RNA bait) like yeast tRNA [94] and negatively charged molecules, like heparin, can be used [80,93]. However, one cannot exclude that the proteins of interest also bind to RNA competitors, in addition to the RNA bait, thus decreasing the sensitivity of the assay. For these reasons, the use of dedicated RNA sequences is advised, as illustrated by the work of Harris and collaborators who have used a 10-fold excess (at least) of different RNA competitors during the incubation of the protein lysate with the immobilized RNA bait (a 308 nt 3′UTR (untranslated region) of the Hepatitis C Virus (HCV)) [86]. Some of the competitor RNAs were totally unspecific (poly(U) or poly(A)), unrelated (MS2 mRNA) or a bit closer to the sequence of interest (collagen 3′UTR). Silver stained SDS-PAGE analysis of the collected proteins after RNA-affinity purification clearly highlighted the efficiency of the collagen 3′UTR to increase the purification of true 3′NTR HCV binders, as additional bands were seen in the presence of this competitor.

3.2.2.4. *Elution*. As for DNA-affinity, the prior separation of the RNP complexes from the solid support before protein digestion and analysis represents a key step to limit the identification of proteins adsorbed on the beads, even if the protein sample has been pre-cleared [91]. From this point of view, strategies developed with a cleavage sequence in the fishing device [105] or some aptamers used to immobilize the RNA bait on the solid support [94] are of great interest. For instance, an excess of free biotin can efficiently displace the RNP complexes immobilized on the paramagnetic beads by an aptamer tag against streptavidin [94]. In a more sophisticated way, Piekna-Przybylska and co-workers have even used a two-step enrichment procedure. The snoRNA complexes were first purified following a TAP-tag method, thanks to the expression of a tagged core protein common to all snoRNPs. After cleavage of the RNP with the TEV protease, the second purification step is based on a RNA tag, as the snoRNA contains the U1hpII domain, which is specifically recognized by the U1A protein. The myc-tagged U1A protein, co-expressed in the same cells, can therefore be used to purify the snoRNA complexes [99].

However, most publications describe a standard elution procedure, based on salt [90,92], detergent [80,84], urea [87], or

heat [83]. In these conditions, as unspecific proteins adsorbed on the chromatographic support inevitably contaminate the proteins of the RNP complexes, the requirement for a strong comparative control downstream in the process is crucial. This is still reinforced in the case of experiments conducted with on bead digestion of the RNP, for which the contribution of RNA-independent binding of proteins to magnetic beads has been estimated at 17–26% [105]. To limit the contamination of unspecific proteins during the elution step, several authors use a step gradient of salt to elute the RNP and select for MS analysis the fraction presenting RNA-binding activity, as shown by UV cross-linking or RNA EMSA [85,89,92].

3.3. *Purification of in vivo-assembled RNA–protein complexes*

Although the vast majority of RNA-affinity based purification literature is dedicated to *in vitro*-assembled RNP complexes, the scarce data generated from *in vivo*-assembled RNA–protein complexes look more relevant. Indeed, *in vitro*-assembled RNA–protein complexes are based on the association between RNA immobilized on a solid support, and proteins in a cell extract. Such complexes may not reflect authentic RNA–protein complexes as immobilized RNA may not fold properly, especially if chemically or sequence-modified, and unspecific RNA–protein interactions can form during purification processes. Three recent studies have described the analysis of *in vivo*-assembled RNP complexes, one related to strictly native complexes (extracted from mouse brain) [104], the second one to an overexpressed genetically modified RNA sequence of interest [105] and the third one to endogenous RNP complexes captured by a peptide nucleic acid (PNA) coupled to a cell penetrating peptide (CPP) [95].

Soeno and co-workers used an immobilized oligonucleotide complementary to a stem loop of the MBII-52 snoRNA to purify the MBII-52 snoRNP from mouse brains [104]. These authors estimate that about 7.3% of the MBII-52 snoRNP was purified by this RNA antisense purification approach. By comparing the captured proteins with those purified with another snoRNA (U20), based on silver stained SDS-PAGE, the authors could highlight several bands corresponding to 17 sequenced proteins specifically interacting with MBII-52 snoRNA. This approach is interesting because the RNA of interest is not modified, neither chemically nor in its sequence, and the RNPs were not assembled *in vitro* but were directly purified from the biologically relevant tissue. However, the extraction procedure might modify the composition of RNPs, although these authors controlled by sedimentation that the approximate molecular weight of the complex was maintained throughout the extraction procedure. In addition, one may suspect that subpopulations of the RNA of interest have been purified, since only a low percentage of MBII-52 snoRNP was captured (7.3%). The success of this experiment was probably partly due to the fact that MBII-52 snoRNP is extremely abundant in mouse brains. One may anticipate that such a strategy could not be applied to any RNPs [104].

Tsai and co-workers have developed an integrated strategy called MS_2 *in vivo* biotin tagged RNA affinity purification (MS_2-BioTRAP) to capture any *in vivo*-assembled RNP complexes. The principle relies on the co-expression of 2 vectors, one coding for the RNA of interest tagged with an aptamer (the stem loop sequence characterized by a high affinity for the bacteriophage MS2 protein), and the second one coding for the MS2 protein fused to a hexahistidine tag, a TEV cleavage site and a signal sequence for *in vivo* biotinylation. RNP complexes assembled with the tagged RNA bait include the biotinylated MS protein that can be purified from cell lysates with streptavidin-coated beads or Ni-coated solid support [105]. The authors estimate that about 1% of tagged-RNA from whole cell extracts was captured by streptavidin-coated beads. At this step, at least two possibilities are offered to collect and analyze the proteins: a sequence-specific cleavage with the TEV protease, keeping the RNP complexes native, or a denaturing elution. The authors also evaluated the opportunity to use UV crosslinking before elution. Indeed, if the objective is to keep the RNP complexes under their native form, mild washing conditions must be used, resulting in the identification of 535 proteins after on bead digestion. On the contrary, if UV cross-linking is carried out prior to cell lysis, the integrity of the RNP complexes is maintained throughout the purification process, and more stringent washing conditions can be used, resulting in 326 proteins identified. The MS_2-BioTRAP, combined with a SILAC-based quantitative mass spectrometry analysis (see Section 4.2.3) was used to identify the proteins captured by Internal Ribosome Entry Site (IRES)-specific RNA, *versus* canonical Cap-dependent translated RNAs. Under stringent denaturing conditions, 36 IRES-enriched RNA binding proteins were identified with a fairly low (1.84) threshold ratio, including several expected proteins [105]. The MS_2-BioTRAP strategy will probably be used in the future to analyze other RNP complexes, although it presents some shortcomings. First, it is only applicable in easy-to-transfect cells, such as 293 HEK cells. Second, importantly, while the *in vivo* assembled-complexes are undoubtedly more relevant than *in vitro* assembled complexes, the fact that two of the interacting molecules are overexpressed might modify the composition of RNP complexes.

Another methodology to capture endogenously formed RNP complexes, called PAIR (PNA-Assisted Identification of RBPs), has been developed by the group of Eberwine [95]. Briefly, a peptide nucleic acid (PNA) coupled with a compound that can be photo-activated is delivered into living cells thanks to a cell penetrating peptide. PNAs are nucleic acid analogs that bind RNA with high sequence specificity, forming highly stable RNA-PNA hybrids. Linked to the PNA is a *p*-benzoylphenylalanine (Bpa) photoactivatable amino acid adduct, that, after UV irradiation, releases a free phenylalanine radical that able to crosslink the nearest molecules (located at a distance $\leq$ 4.5 Å). The PNA–RNP complexes are then collected with a biotinylated oligonucleotide complementary to the PNA, and coupled to paramagnetic streptavidin-coated beads. As there are covalent links between RNA and RBPs, stringent washes can be used to eliminate any contaminant bound material. The proteins are then eluted by heat and resolved by SDS-PAGE before MS analysis.

Compared with the method developed by Tsai and co-workers, this strategy has the advantage to analyze the authentic endogenously formed RNP complexes, without possible modifications caused by the relative expression levels of transfected plasmids, and to allow a precise specific dissection of the sequence analyzed, as RBPs can be isolated

in an exon-specific manner [106]. Although this method can theoretically be applied to any mRNA expressed in cell cultures, one can suspect that, due to RNA secondary structure, some mRNA sequences may not be accessible to PNA. However, this point nor the recovery yields of the mRNA of interest were discussed by the authors.

4. MS-based identification of NA-affinity purified proteins

Once the TFs or RBPs have been captured by the DNA or RNA baits, respectively, and (selectively) eluted from the chromatographic support, the next challenge is to identify the low abundant proteins of interest in a complex protein solution containing multiple highly abundant proteins. As a reminder, the unspecific proteins may have been captured by the bait in a non-sequence specific way (*e.g.* DNA/RNA maintenance proteins), adsorbed on the chromatographic support (in the case of denaturing elution processes), or introduced during the NA-affinity purification process (*e.g.* proteins blocking the chromatographic support, restriction enzymes used to selectively elute the protein–DNA complexes, TEV protease used to free the RBP complex from the fishing proteins). Therefore, reducing the sample complexity is a key step towards the successful identification of low abundant sequence-specific interacting proteins.

4.1. *Reduction of sample complexity*

The large majority of NA affinity-based procedures include a step of SDS-PAGE between the elution of the RNPs or the TFs and the MS analysis. The bands of interest or the entire gel lanes are cut and processed for trypsin digestion and MS analysis. This has several advantages: i) it gives information on the molecular weight of the interacting proteins, ii) it simplifies the protein mixture to analyze and iii) to some extent, it allows to compare the proteins captured by the sequence of interest and the proteins captured by a control bait (see below). However, including a SDS-PAGE separation step of proteins before MS analysis limits the overall sensitivity of the assay, whether the authors use a poorly sensitive protein staining such as Coomassie blue [84,85,89], or a more sensitive staining like silver [83,88] or Sypro Ruby [86,93]. This aspect has been clearly highlighted by Soeno and co-workers who studied the proteins interacting *in vivo* with the snoMBII-52 RNP, by comparison with proteins captured by a U20 snoRNA sequence of the same length [104]. After analysis of the differential bands detected on silver stained SDS-PAGE, these authors could identify 17 proteins. Several expected proteins were not identified, although they were present in the RNP complexes, as shown by Western blot analysis.

To overcome the limitation of gel-based protein separation, gel-independent proteomics has emerged. The first liquid chromatographies used to decrease the sample complexity were performed off line, often with microtips [54], and were progressively improved by the development of nanoLC, of long elution gradients [56] and of multidimensional peptide separation [24,43–45,57,61]. In most recent papers, these more sensitive gel-free approaches can be used to simplify complex protein mixtures prior to MS analysis, eventually coupled with chemical or metabolic labeling allowing quantitative analyses to focus on specific partners [107] as discussed in Section 4.2. In this approach, the digested proteins are separated under one (typically nano-RPLC (reverse-phase liquid chromatography)–MS/MS) or two dimensions (2D nano-LC–MS/MS) prior to MS analysis. In 2D-nano-LC-MS/MS methodology, peptides could be separated by either two successive reverse phase chromatography steps [61] or a method called multidimensional protein identification technology (MudPIT) that usually combines a strong cation exchange chromatography (SCX) with a reverse phase chromatography [87,105]. In line of what has been illustrated above, the number of interacting proteins identified is generally higher, for example, 500 proteins as in the gel-free study of Talukdar and co-workers [92] while a typical gel-dependent study leads to the identification ranging between 20 [27,43,53] and 250 different proteins [25]. The question of the sequence-specificity of the identified proteins is then more important than ever, and can be answered by using comparative proteomic analyses.

4.2. *Comparative proteomic analyses*

4.2.1. *Defining a "control" condition*

To discriminate between the proteins of interest – captured by the bait in a sequence-specific manner - and the unspecific proteins, a control condition must be defined. This control condition corresponds to an experimental set up in which the proteins of interest should not be contained in the eluate. This can be achieved by modifying either the NA partner or the protein partners. Most authors compare the RNA bait-purified proteins with proteins captured by a control RNA sequence, ranging from totally unrelated sequences [83,87,91] to antisense sequences [90], coding sequence when UTR is studied [82,88,94], or closely related sequences like those mutated in a few nucleotides [92–94]. Similarly, regarding the DNA partner, we can cite the use of i) a blank (*e.g.* the proteins adsorbed on the chromatographic support not linked to the bait) [23], ii) an immobilized scramble DNA [42,45,65], iii) a sequence-specific negative control, like a bait mutated in the binding site of interest [25,43,44,56], or a (de)methylated DNA sequence when the proteins captured by methylated DNA are the proteins of interest [25]. The relevance of the data generated is clearly conditioned by a judicious choice of the control condition, itself determined by the biological question considered. We would like to underline that the closer the control sequence is, the more relevant the data in terms of sequence-specific interacting proteins will be. However, such strategy can only be applied to precisely defined binding sites, and not to uncharacterized long sequences.

Considering the protein partner, some studies are based on the comparison of proteins extracted from cells stimulated or not [29,61,95]. Alternatively, the group of Ranish used an experimental condition in which the protein of interest is not expressed as a control in order to unravel the partners of this transcription factor [43,108].

4.2.2. *Qualitative comparison*

Whatever the conditions considered, the sample comparison can be achieved in a qualitative or in a quantitative manner

(Fig. 2 - Upper part). Gel-based separations are used for qualitative comparisons, both after RNA-affinity [83–86,88,89,95,104] or DNA-affinity, either in one dimension [22,27,42,46,65] or in two dimensions [28,38].

Although the electrophoretic protein separation is easy to set up, its main drawback is, as already mentioned above, a poor sensitivity possibly associated with a limited resolution. Indeed, the low abundant proteins might be undetected after PAGE or masked by highly abundant contaminant proteins of similar molecular weight [27,43,53] as also illustrated by Mittler and co-workers who compared the proteins captured by a 40-bp long oligonucleotide containing a wild type or mutated site for AP-1 [25]. The protein–DNA complexes, specifically eluted from the paramagnetic beads after digestion with a restriction enzyme, were denatured and the proteins resolved by silver-stained SDS-PAGE. No differences between wild-type (wt) *versus* mutated bait could be observed after gel examination, although the AP-2 transcription factor was specifically detected by Western blot in the wt condition but not in the mutated one [25]. This illustrates the necessity for quantitative proteomics to analyze such proteins.

4.2.3. *Quantitative comparison*

Quantitative proteomics can be achieved through different strategies, dependent on chemical or metabolic labeling (reviewed in [107,109–112]), or based on label-free approaches (reviewed in [113]). As the respective pros and cons of each strategy have been reviewed [112,114], this will not be discussed here.

To our knowledge, only two papers reported spectral counting to compare the proteins captured by a wt RNA [92] or DNA [23] bait to those captured respectively by the mutated counterpart or by a blank composed of unconjugated cellulose. Briefly, spectral count takes into account the number of peptides sequenced, the number of spectra acquired as well as the sequence coverage to evaluate the relative abundance of a protein in one sample *versus* the other one. However, an important number of replicates are required to get statistically relevant data with this approach, especially when the authors consider relatively low ratio between the two samples.

Mostly used strategies applied to NA-affinity purified proteins include ICAT (isotope coded affinity tag labeled peptides), stable-isotope dimethyl labeling of peptides and SILAC.

In the case of ICAT, proteins from each sample are chemically labeled with cysteine reactive ICAT reagents before being combined together and digested. The peptide sample complexity is then reduced through strong cation exchange chromatography. The cysteine-labeled peptides are further purified by avidin affinity chromatography and finally analyzed by μLC-MS/MS. Due to the chemical labels, slightly different in their m/z ratio, the relative abundance of each peptide in the different conditions can be deduced. This strategy has been used by the group of Ranish in several promoter or enhancer contexts [44,45,57,115]. Those experiments confirmed the large number of proteins captured by DNA-affinity in a nonspecific manner: for instance, in the case of the identification of Six4 transcription factor as a regulator of the muscle creatine kinase gene [57], only 3 proteins presented a higher abundance in the wt *versus* mutated sample, among 900 identified co-purified proteins characterized by a one-to-one ratio.

Stable isotope dimethyl labeling has been used in order to highlight proteins specifically captured by an estrogen response element (ERE) in cells stimulated or not with estradiol [61]. In this case, primary amine of peptides from both conditions is differentially labeled by formaldehyde in combination with cyanoborohydride before to be pooled [116]. The sample complexity is then reduced by two successive reverse phase high performance liquid chromatography (RP-HPLC) separations before MS analysis. This MS-based quantitative comparative approach led to the identification of 43 proteins, including ERα, presenting a significant binding to the probe as well as a significant change under estradiol stimulation. These enriched proteins were then sorted according to their functions and their possible interactions, which led the authors to propose an ERE-bound complex model [61].

Interestingly, the power of such quantitative approach might exempt from heavy TF [25,117] or RBP [82,94,105] purification or from thorough sample complexity reduction. This is illustrated by the studies conducted by the group of M. Mann in SILAC-based analyses of transcriptional regulators purified by DNA-affinity [25,117] or RBP purified by RNA-affinity [94]. In SILAC-based proteomic analyses, the proteins are metabolically labeled during cell culture in the presence of different amino acid isotopes (for instance heavy or light lysine) to allow discrimination based on differences in peptide mass. The proteins metabolically labeled with one isotope or another were incubated with the immobilized mutated or wt bait. After washing the beads, the two conditions were combined before a restriction enzyme-based specific elution of the protein–DNA complexes. Proteins were resolved by 1D SDS-PAGE, and the gel cut into pieces. After LC–MS/MS analysis of the different pieces, 250 proteins were identified, among which only 4 were differentially abundant, corresponding to TFs specific for the binding site of interest [25]. A similar strategy was adopted by Kern's research group, with a double labeling of both arginine and lysine, which improves the number of quantified peptides [53].

Such remarkable achievement can raise the following question: in the case of quantitative proteomics such as SILAC-based strategies, is it really necessary to reduce the sample complexity to allow the identification of the specific TF? To answer this question, the same authors compared the results obtained with an in-gel digestion, as described above, to those obtained with the in-solution digestion of the whole eluate analyzed by LC–MS/MS with a LTQ–FTMS (Linear ion Trap Quadrupole–Fourier Transform Mass Spectrometer). With the second strategy, a lower number of different proteins were identified (197 instead of 703), and among them only one specific TF, instead of 2 with the gel-dependent strategy. Even if the heavy/light ratio of this specific protein was similar to the gel-dependent experiment, this was obtained from a lower number of quantified peptides (5 instead of 14). This suggests that even with quantitative proteomics, the identification of the partners interacting with a short oligonucleotide (in this case, 26 bp) requires a step of reduction of the sample complexity. However, with quantitative proteomics, a simple 1D gel separation of the proteins might be sufficient to reach this goal, which is easy to handle and less time and material consuming than multiple chromatographic separation including off line SCX [43,57]. In addition, it has to be underlined that the interesting results obtained in quantitative proteomics

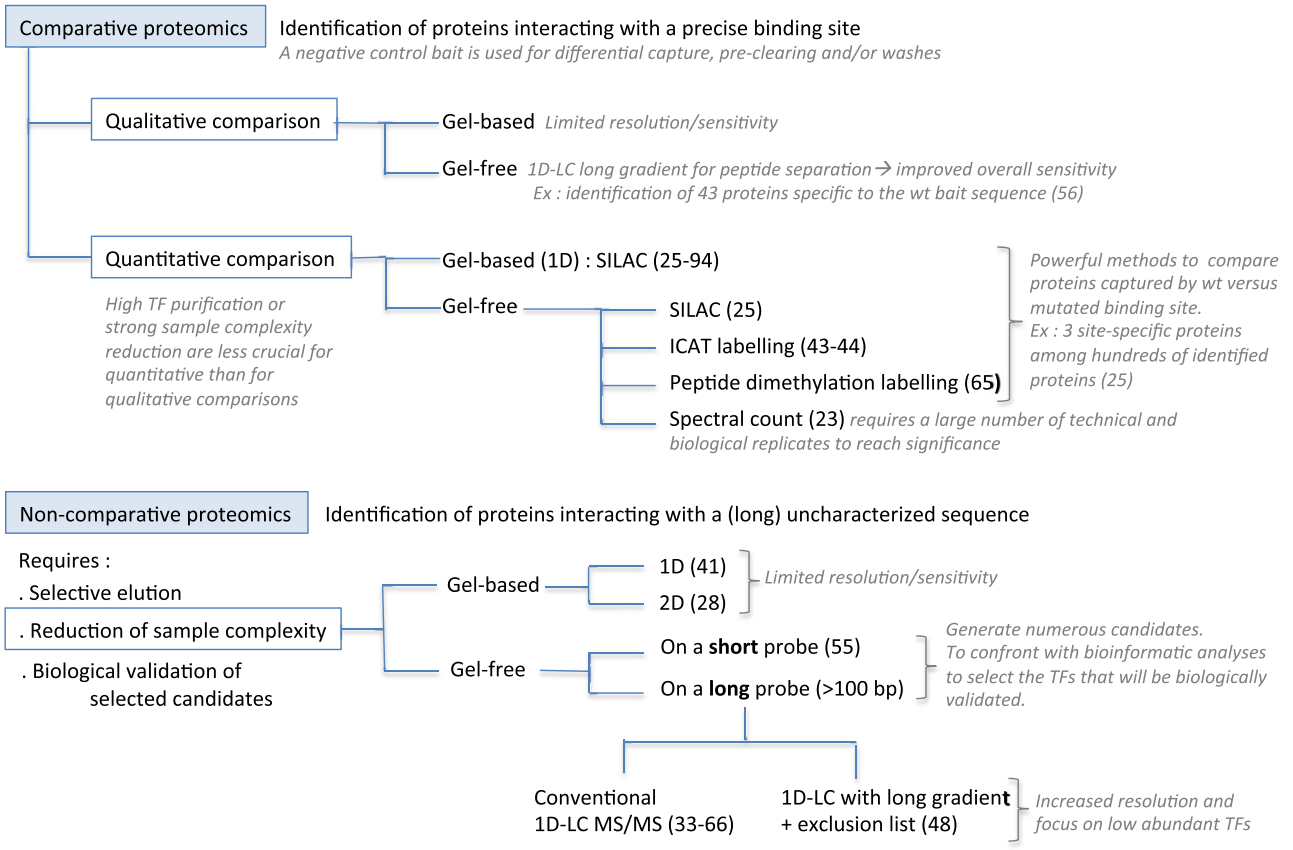

Fig. 2 – Most DNA-affinity strategies used to identify proteins *in vitro*-captured by DNA baits. Abbreviations: 1D-LC: one dimensional-liquid chromatography; MudPIT: multidimensional protein identification technology; ICAT: isotope-coded affinity tag; SILAC: stable isotope labeling by amino acids in cell culture.

applied to DNA-affinity purified proteins have been made possible by the never ending progresses made in MS instrumentations, allowing high speed of analysis, sensitivity, resolution and mass accuracy required for the analysis of highly complex mixtures.

4.3. *Noncomparative proteomics*

Although such quantitative proteomics strategies are efficient to allow the identification of transcription factors specific for a precise DNA binding site, they depend on the choice of an appropriate negative control. Consequently, they can only be applied to well-characterized binding sites, for which the mutated version has been shown previously to be non-functional. The biological question is totally different when the key binding sites have not been defined and when the authors want to identify a maximum of proteins interacting with this DNA sequence [23,27,28,33,41,47,48,55,66]. Such question corresponds to what we have called "totally unbiased analysis of DNA-interacting proteins" (the right part of Fig. 1). However, the unbiased aspect is somehow limited if the authors concentrate on short regulatory sequences defined inside a long region following a heavy experimental work like directed mutagenesis [47] or EMSA tiling [55]. DNA-affinity purification performed on short sequences lacks the above-cited advantages of long sequences, including the cooperation between transcription factors and the reconstitution of "enhanceosomes".

There are few reports analyzing the proteins captured by relatively long DNA sequences (>100 bp) that could correspond to enhancer or promoter fragments. Let's mention that Kim's group applied a DNA affinity capture assay to identify unknown transcriptional regulators involved in antibiotic biosynthesis through interaction with 350 bp-long promoter regions of *actI-ORF4* and *redD* in *Streptomyces coelicolor* A3(2) [66]. MS analysis following this kind of large-scale unbiased DNA affinity assay on a relatively long sequence generates long lists of putative candidates (>100 [23,48]). However, a major difficulty is to discriminate between the sequence-specific and unspecific interacting proteins as, in the absence of appropriate control, quantitative proteomics cannot be used. To be relevant, such analyses should be characterized by the following features: a selective elution to avoid the identification of proteins adsorbed on the chromatographic support, an efficient process to make the sample less complex, the absolute need for biological independent replicates and further biological validation of selected candidates (Fig. 2 - Bottom).

Indeed, due to the long bait, a very large number of different proteins are captured, and the low abundant sequence-specific TFs may be undetectable in a high dynamic range mixture. Therefore, gel-free approaches inspired from complex proteome

analyses and based on improved chromatographic separation of the complex peptide mixture prior to the MS analysis [118] should be recommended. Among different possibilities, MudPIT-based strategy, combining multidimensional LC separation prior to ESI (electrospray ionization)–MS/MS detection, could be considered even if, to our knowledge, DNA affinity purification followed by MudPIT separation has systematically been associated with quantitative proteomics. Such analyses are not only laborious and time-consuming, they also represent a non-negligible risk of low abundant peptides loss in the case of offline successive chromatographic separations. Alternatively, increasing the column length and the run time can been used to improve peptide separation before MS sequencing. This approach was shown to be useful in DNA affinity capture experiments performed on short [29,56] and long DNA sequence [48]. Beside the improved resolution of peptide separation, a peptide exclusion list can also be used to specifically focus on the sequencing of low abundant DNA binding proteins. In the analysis of proteins interacting with a long (226 bp) fragment of the HIV-1 LTR5′, we have adapted the MS/MS analysis to proceed in two successive runs for each sample. Raw data from the first run, largely composed of peptides from abundant proteins, were used to generate a schedule precursor list (SPL) which contains peptide masses of the most abundant proteins detected in the first run. Those peptide masses were excluded from sequencing during the second run to preferentially focus on the sequencing of low abundant peptides. Peptides sequenced from both runs were then merged to proceed to protein identification [48]. A careful examination of the data obtained after the first and second runs indicates that the use of a peptide exclusion list led to an increase in TF identification and a global improvement in the quality of sequencing of transcriptional regulators.

With such strategy, the number of identified proteins is inevitably high, even after subtraction of the proteins identified in a "blank" experiment, in which the beads are not functionalized with the DNA [23,48].

To be considered as relevant candidates, the proteins should be identified out of several independent biological replicates and then functionally validated. Indeed, due to the natural tendency of numerous proteins to bind RNA [84,86] or DNA [119] with high affinity, the candidate interacting proteins absolutely require further biological validations. For instance, Blackwell and co-workers compared the proteins captured by the 45-nt piALU RNA sequence to those interacting with a control hexamer-repeat sequence of the same length. Silver stained SDS-PAGE electrophoretic profiles of the eluted proteins were highly similar. In 5 bands that were more pronounced in piALU-purified proteins, the authors identified by MSMS analysis 64 proteins involved in various biological processes like chromatin remodeling, DNA repair, cell cycle control, centromeres/pericentromeres function. These data probably contain false negative and false positive candidates, but since they lack functional assessment, it is not possible to evaluate their significance [83]. Considering this example, one can wonder if it is worth to undergo an unbiased analysis of the proteins interacting with a NA sequence of interest, since it is difficult, without an appropriate control, to assess the biological significance of the data.

However, in the case of DNA-affinity, it is easier to select the candidates that will undergo further biological validations than for RNA-affinity. Indeed, the databases of transcription factor binding sites can be used to predict the transcription factors that may bind the DNA sequence of interest. Although, as already mentioned, such *in silico* analyses are well known to generate a large number of false positive and false negative [120], such list of candidates can be crossed with the experimentally generated list of interacting proteins obtained after DNA-affinity chromatography. We applied this strategy to perform a totally unbiased analysis of a 226 bp fragment of the HIV-1 5′LTR [48]. A list of 24 putative transcription factors and 38 putative co-regulators was highlighted, among which 12 expected transcription factors. After *in silico* analysis, putative DNA-binding sites could be located on the DNA sequence of interest for 11 of the 12 unexpected TF. One selected TF was finally functionally validated.

5. Conclusion and future directions

On the long road to decipher translation and transcription regulation, the identification of proteins interacting with regulatory nucleic sequences is an obligate step, that can now progressively be overcome, mostly due to great progresses in mass spectrometry and protein/peptide separation. However, upstream the MS-based protein sequencing, the capture and purification of sequence-specific interacting proteins represent a real challenge. The choice of the NA affinity capture strategy to adopt should definitely be dictated by the biological question: either a comparative proteomic analysis focused on proteins captured by a precisely defined binding site contained in a short regulatory sequence or an unbiased analysis of a long and poorly characterized regulatory sequence.

In case of a comparative approach, the choice of the control bait is a key element: the closer the control sequence is, the more relevant will be the results towards the binding site/region of interest. Quantitative proteomics is particularly well suited to answer this type of biological question, the relevant specific interacting proteins being pointed out at the MS-analysis step. These approaches are so powerful in terms of specificity that they are moderately dependent on all tricks that may be used to guarantee the specificity of the protein capture.

On the opposite, unbiased proteomic analyses to study NA-interacting proteins generate a long list of candidates that cannot be compared with a control bait to distinguish the sequence specific interacting proteins from the "noise", composed of contaminants and non-sequence specific interacting proteins. Each step of the NA-affinity purification must therefore be carefully controlled to ensure, as much as possible, the capture specificity, elution and/or identification of these interacting proteins or interest. Indeed, beside the specific TFs or regulatory RBPs, a lot of abundant proteins are able to bind in a non-sequence specific manner, DNA or RNA, respectively. Therefore, the results generated by unbiased proteomics must necessarily be validated through several independent replicates and with classic assays. As an alternative, an unbiased proteomic analysis might eventually be pursued by a comparative proteomics, for instance when the first one has highlighted TF candidates for which a

binding site can be located on the analyzed sequence [48]. Of note, in the case of unbiased proteomics conducted for RNA-affinity purified proteins, it is more difficult to select the most promising candidates than for DNA-affinity, as, to our knowledge, no binding site database and dedicated algorithms exist to predict the binding of proteins to defined RNA sequence.

Improvement in NA-affinity purification and identification of proteins also represents a promising avenue to study the composition, stoichiometry and dynamic of NA-interacting complexes. Indeed, defining the exact composition of protein complexes, as well as understanding how complexes are assembled and regulated, is of great interest [121]. It has been shown that the purification of such protein complexes by nucleic acid affinity is more relevant than by immunoprecipitation [122]. Although most of the papers cited in this review focus on the identification of proteins that directly bind to nucleic acids, a few of them have succeeded in the identification of proteins that indirectly bind to NA [29,43,48,56], indicating that the experimental settings suit to *in vitro*-assembled NA–protein complexes reconstruction. However, those findings need to be validated *in vivo*.

In this context, the few studies performed on *in vivo*-assembled NA–protein complexes are promising as they theoretically overcome the limitations of *in vitro*-assembled complexes, in terms of NA/protein stoichiometry, buffer composition, presence of chromatin for DNA-affinity purification and presence of secondary structures for RNA-affinity purification. Although this probably represents the future directions to take in this field, those strategies currently suffer from limitations in the sensitivity of the MS-based identification step. This explains why, regarding endogenous DNA sequences, only sequences present in multiple copies (mtDNA or telomeres) have been studied so far. This sensitivity limitation is less crucial regarding *in vivo*-assembled RNP complexes as RNAs are more abundant than unique genomic DNA. As reported in this review, at least one example of endogenous RNP complex has been elucidated, and interesting RNA fishing strategies such as the recently developed (PNA)-Assisted Identification of RBPs (PAIR) technology could be applied to any RNA sequence. Interestingly, several authors use crosslinking agents to covalently freeze the *in vitro*- or *in vivo*-assembled RNP complexes, making the purification process more robust and authorizing the use of stringent washes to increase the purification specificity. Curiously, the use of reversible formaldehyde crosslinks is not widely used in DNA-affinity purification yet, although it could constitute a real benefit for the specificity and the sensitivity of the method.

List of abbreviations

AP-1 activator protein 1
AP-2 activator protein 2
C/EBP CAAT-enhancer binding protein
ChIP chromatin immunoprecipitation
CPP cell penetrating peptide
DNA-PK DNA-dependent protein kinase
ERE estrogen response element
GENECAPP Global ExoNuclease-based Enrichment of Chromatin-Associated Proteins for Proteomics
GFP green fluorescent protein
GTF general transcription factor
HCV Hepatitis C virus
HIV-1 LTR5′ human immunodeficiency virus long terminal repeat 5′
IRES internal ribosome entry site
LNA locked nucleic acid
LTQ–FTMS Linear ion Trap Quadrupole–Fourier Transform Mass Spectrometer
mtDNA mitochondrial DNA
MudPIT multidimensional protein identification technology
NA nucleic acid
NAPSTER Nucleotide-Affinity Pre-incubation Specificity Test of Recognition
NTR nontranslated region
PAIR PNA-Assisted Identification of RBPs
PAR-CLIP PhotoActivated-Ribonucleoside-enhanced CrossLinking and ImmunoPrecipitation
PICh Proteomics of Isolated Chromatin segments
PNA peptide nucleic acid
RBP RNA-binding proteins
RNP ribonucleoprotein
SCX strong cation exchange
SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis
SELEX Systematic Evolution of Ligands by EXponential enrichment
SILAC stable isotope labeling by amino acids in cell culture
snoRNP small nucleolar ribonucleoprotein
snRNP small nuclear ribonucleoprotein
SPL schedule precursor list
TAP tandem affinity purification
TEV Tobacco Etch virus
TF transcription factor
USF-2 upstream stimulatory factor-2
UTR untranslated region
wt wild type
YB-1 Y-box binding protein 1
FPR false positive rate

REFERENCES

[1] Farnham P. Insights from genomic profiling of transcription factors. Nat Rev Genet 2009;10:605–16.
[2] Thomas MC, Chiang C-M. The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol 2006;41:105–78.
[3] Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet 2009;10:252–63.
[4] Kato S, Yokoyama A, Fujiki R. Nuclear receptor coregulators merge transcriptional coregulation with epigenetic regulation. Trends Biochem Sci 2011;36:272–81.

[5] Dey B, Thukral S, Krishnan S, Chakrobarty M, Gupta S, Manghani C, et al. DNA–protein interactions: methods for detection and analysis. Mol Cell Biochem 2012;365(1-2):279–99.
[6] Gupta RK, Rosen ED, Spiegelman BM. Identifying novel transcriptional components controlling energy metabolism. Cell Metab 2011;14:739–45.
[7] Odom DT. Identification of transcription factor–DNA interactions in vivo. Subcell Biochem 2011;52:175–91.
[8] Liu ET, Pott S, Huss M. Q&A: ChIP-seq technologies and the study of gene regulation. BMC Biol 2010;8:56.
[9] Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet 2009;10:669–80.
[10] Jolma A, Taipale J. Methods for analysis of transcription factor DNA-binding specificity in vitro. In: Hughes TR, editor. A handbook of transcription factors, vol. 52. Dordrecht: Springer Netherlands; 2011. p. 155–73.
[11] Hampshire AJ, Rusling DA, Broughton-Head VJ, Fox KR. Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. Methods 2007;42:128–40.
[12] Connaghan-Jones KD, Moody AD, Bain DL. Quantitative DNase footprint titration: a tool for analyzing the energetics of protein–DNA interactions. Nat Protoc 2008;3:900–14.
[13] Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. Nat Protoc 2007;2:1849–61.
[14] Benos PV. *In silico* discovery of DNA regulatory sites and modules. In: Krawetz S, editor. Bioinformatics for Systems Biology. Humana Press; 2009. p. 353–66.
[15] Werner T. The role of transcription factor binding sites in promoters and their in silico detection. In: Krawetz S, editor. Bioinformatics for systems biology. Totowa, NJ: Humana Press; 2009. p. 339–52.
[16] Matys V. TRANSFAC(R) and its module TRANSCompel(R): transcriptional gene regulation in eukaryotes. Nucleic Acids Res 2006;34:D108–10.
[17] Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E, et al. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res 2010(Database issue):D105–10.
[18] Kahana S, Pnueli L, Kainth P, Sassi HE, Andrews B, Kassir Y. Functional dissection of IME1 transcription using quantitative promoter-reporter screening. Genetics 2010;186:829–41.
[19] Antonaki A, Demetriades C, Polyzos A, Banos A, Vatsellas G, Lavigne MD, et al. Genomic analysis reveals a novel nuclear factor-κB (NF-κB)-binding site in Alu-repetitive elements. J Biol Chem 2011;286:38768–82.
[20] Hu S, Xie Z, Onishi A, Yu X, Jiang L, Lin J, et al. Profiling the human protein–DNA interactome reveals MAPK1 as a transcriptional repressor of interferon signalling. Cell 2009;139:610–22.
[21] Kadonaga JT, Tjian R. Affinity purification of sequence-specific DNA binding proteins. Proc Natl Acad Sci U S A 1986;83:5889–93.
[22] Nirodi C, NagDas S, Gygi SP, Olson G, Aebersold R, Richmond A. A role for poly(ADP-ribose) polymerase in the transcriptional regulation of the melanoma growth stimulatory activity (CXCL1) gene expression. J Biol Chem 2001;276:9366–74.
[23] Dobretsova A, Johnson JW, Jones RC, Edmondson RD, Wight PA. Proteomic analysis of nuclear factors binding to an intronic enhancer in the myelin proteolipid protein gene. J Neurochem 2008;105:1979–95.
[24] Rubio ED, Reiss DJ, Welcsh PL, Disteche CM, Filippova GN, Baliga NS, et al. CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A 2008;105:8309–14.
[25] Mittler G, Butter F, Mann M. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. Genome Res 2009;19:284–93.
[26] Yaneva M, Kowalewski T, Lieber MR. Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies. EMBO J 1997;16:5098–112.
[27] Jiang D, Zhou Y, Moxley RA, Jarrett HW. Purification and identification of positive regulators binding to a novel element in the c-Jun promoter. Biochemistry 2008;47:9318–34.
[28] Jiang S, Galindo MR, Jarrett HW. Purification and identification of a transcription factor, USF-2, binding to E-box element in the promoter of human telomerase reverse transcriptase (hTERT). Proteomics 2010;10:203–11.
[29] Lu X, Beck Jr GR, Gilbert LC, Camalier CE, Bateman NW, Hood BL, et al. Identification of the homeobox protein Prx1 (MHox, Prrx-1) as a regulator of osterix expression and mediator of tumor necrosis factor α action in osteoblast differentiation. J Bone Miner Res 2011;26:209–19.
[30] Alberts B, Herrick G. [11] DNA-cellulose chromatography. In: Lawrence Grossman KM, editor. Methods in Enzymology, vol. 21. Academic Press; 1971. p. 198–217.
[31] Rosenfeld PJ, Kelly TJ. Purification of nuclear factor I by DNA recognition site affinity chromatography. J Biol Chem 1986;261:1398–408.
[32] Hégarat N, François J-C, Praseuth D. Modern tools for identification of nucleic acid-binding proteins. Biochimie 2008;90:1265–72.
[33] Park S-S, Ko BJ, Kim B-G. Mass spectrometric screening of transcriptional regulators using DNA affinity capture assay. Anal Biochem 2005;344:152–4.
[34] Yaneva M, Tempst P. Affinity capture of specific DNA-binding proteins for mass spectrometric identification. Anal Chem 2003;75:6437–48.
[35] Yaneva M, Tempst P. Isolation and mass spectrometry of specific DNA binding proteins. Methods Mol Biol 2006;338:291–303.
[36] Gadgil H, Oak SA, Jarrett HW. Affinity purification of DNA-binding proteins. J Biochem Biophys Methods 2001;49:607–24.
[37] Panda M, Jiang D, Jarrett HW. Trapping of transcription factors with symmetrical DNA using thiol-disulfide exchange chemistry. J Chromatogr A 2008;1202:75–82.
[38] Hégarat N, Cardoso GM, Rusconi F, Francois J-C, Praseuth D. Analytical biochemistry of DNA–protein assemblies from crude cell extracts. Nucleic Acids Res 2007;35:e92.
[39] Mrinal N, Tomar A, Nagaraju J. Role of sequence encoded κB DNA geometry in gene regulation by Dorsal. Nucleic Acids Res 2011;39:9574–91.
[40] Babayeva ND, Baranovskaya OI, Tahirov TH. Structural basis of Ets1 cooperative binding to widely separated sites on promoter DNA. PLoS One 2012;7:e33698.
[41] Masternak K, Barras E, Zufferey M, Conrad B, Corthals G, Aebersold R, et al. A gene encoding a novel RFX-associated transactivator is mutated in the majority of MHC class II deficiency patients. Nat Genet 1998;20:273–7.
[42] Drewett V, Molina H, Millar A, Muller S, von Hesler F, Shaw PE. DNA-bound transcription factor complexes analysed by mass-spectrometry: binding of novel proteins to the human c-fos SRE and related sequences. Nucleic Acids Res 2001;29:479–87.
[43] Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J, et al. The study of macromolecular complexes by quantitative proteomics. Nat Genet 2003;33:349–55.
[44] Himeda CL, Ranish JA, Hauschka SD. Quantitative proteomic identification of MAZ as a transcriptional regulator of muscle-specific genes in skeletal and cardiac myocytes. Mol Cell Biol 2008;28:6521–35.
[45] Qi Y, Ranish JA, Zhu X, Krones A, Zhang J, Aebersold R, et al. Atbf1 is required for the Pit1 gene early activation. Proc Natl Acad Sci U S A 2008;105:2481–6.

[46] Samuel S, Twizere J-C, Bernstein LR. YB-1 represses AP1-dependent gene transactivation and interacts with an AP-1 DNA sequence. Biochem J 2005;388:921–8.
[47] Reed DE, Huang XM, Wohlschlegel JA, Levine MS, Senger K. DEAF-1 regulates immunity gene expression in *Drosophila*. Proc Natl Acad Sci U S A 2008;105:8351–6.
[48] Tacheny A, Michel S, Dieu M, Payen L, Arnould T, Renard P. Unbiased proteomic analysis of proteins interacting with the HIV-1 5′LTR sequence: role of the transcription factor Meis. Nucleic Acids Res 2012;40(21):e168. http://dx.doi.org/10.1093/nar/gks733 (Epub 2012 Aug 16).
[49] Gabrielsen OS, Hornes E, Korsnes L, Ruet A, Oyen TB. Magnetic DNA affinity purification of yeast transcription factor tau—a new purification principle for the ultrarapid isolation of near homogeneous factor. Nucleic Acids Res 1989;17:6253–67.
[50] Gadgil H, Jarrett HW. Oligonucleotide trapping method for purification of transcription factors. J Chromatogr A 2002;966:99–110.
[51] Jiang D, Moxley RA, Jarrett HW. Promoter trapping of c-jun promoter-binding transcription factors. J Chromatogr A 2006;1133:83–94.
[52] Berg OG, Winter RB, von Hippel PH. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry 1981;20:6929–48.
[53] Ren YR, Chaerkady R, Hu S, Wan J, Qian J, Zhu H, et al. Unbiased discovery of interactions at a control locus driving expression of the cancer-specific therapeutic and diagnostic target, mesothelin. J Proteome Res 2012;11(11):5301–10.
[54] O'Neill D, Yang J, Erdjument-Bromage H, Bornschlegel K, Tempst P, Bank A. Tissue-specific and developmental stage-specific DNA binding by a mammalian SWI/SNF complex associated with human fetal-to-adult globin gene switching. Proc Natl Acad Sci U S A 1999;96:349–54.
[55] Karmakar S, Mahajan MC, Schulz V, Boyapaty G, Weissman SM. A multiprotein complex necessary for both transcription and DNA replication at the β-globin locus. EMBO J 2010;29:3260–71.
[56] Meng Z, Camalier CE, Lucas DA, Veenstra TD, Beck Jr GR, Conrads TP. Probing early growth response 1 interacting proteins at the active promoter in osteoblast cells using oligoprecipitation and mass spectrometry. J Proteome Res 2006;5:1931–9.
[57] Himeda CL, Ranish JA, Angello JC, Maire P, Aebersold R, Hauschka SD. Quantitative proteomic identification of six4 as the trex-binding factor in the muscle creatine kinase enhancer. Mol Cell Biol 2004;24:2132–43.
[58] Kumar NV, Bernstein LR. A new analytical scale DNA affinity binding assay for analyses of specific protein–DNA interactions. Anal Biochem 2001;299:203–10.
[59] Samuel S, Beifuss KK, Bernstein LR. YB-1 binds to the MMP-13 promoter sequence and represses MMP-13 transactivation via the AP-1 site. Biochim Biophys Acta 2007;1769:525–31.
[60] Lee C, Gyorgy A, Maric D, Sadri N, Schneider RJ, Barker JL, et al. Members of the NuRD chromatin remodeling complex interact with AUF1 in developing cortical neurons. Cereb Cortex 2008;18:2909–19.
[61] Cheng P-C, Chang H-K, Chen S-H. Quantitative nanoproteomics for protein complexes (QNanoPX) related to estrogen transcriptional action. Mol Cell Proteomics 2010;9:209–24.
[62] Pastore A, Piemonte F. S-Glutathionylation signaling in cell biology: progress and prospects. Eur J Pharm Sci 2012;46:279–92.
[63] Renard P, Ernest I, Houbion A, Art M, Le Calvez H, Raes M, et al. Development of a sensitive multi-well colorimetric assay for active NFkappaB. Nucleic Acids Res 2001;29:E21.
[64] Arora T, Liu B, He H, Kim J, Murphy TL, Murphy KM, et al. PIASx is a transcriptional co-repressor of signal transducer and activator of transcription 4. J Biol Chem 2003;278:21327–30.
[65] Nordhoff E, Krogsdam AM, Jorgensen HF, Kallipolitis BH, Clark BF, Roepstorff P, et al. Rapid identification of DNA-binding proteins by mass spectrometry. Nat Biotechnol 1999;17:884–8.
[66] Park S-S, Yang Y-H, Song E, Kim E-J, Kim WS, Sohng JK, et al. Mass spectrometric screening of transcriptional regulators involved in antibiotic biosynthesis in streptomyces coelicolor A3(2). J Ind Microbiol Biotechnol 2009;36:1073–83.
[67] Angelov D, Charra M, Müller CW, Cadet J, Dimitrov S. Solution study of the NF-κB p50–DNA complex by UV laser protein–DNA cross-linking¶. Photochem Photobiol 2003;77:592–6.
[68] Hirsch JD, Eslamizar L, Filanoski BJ, Malekzadeh N, Haugland RP, Beechem JM, et al. Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. Anal Biochem 2002;308:343–57.
[69] Déjardin J, Kingston RE. Purification of proteins associated with specific genomic Loci. Cell 2009;136:175–86.
[70] Bogenhagen DF. Biochemical isolation of mtDNA nucleoids from animal cells. Methods Mol Biol 2009;554:3–14.
[71] Butala M, Busby SJW, Lee DJ. DNA sampling: a method for probing protein binding at specific loci on bacterial chromosomes. Nucleic Acids Res 2009;37:e37.
[72] Bogenhagen DF. Mitochondrial DNA, nucleoid structure. Biochim Biophys Acta 1819;2012:914–20.
[73] Wu C-H, Chen S, Shortreed MR, Kreitinger GM, Yuan Y, Frey BL, et al. Sequence-specific capture of protein–DNA complexes for mass spectrometric protein identification. PLoS One 2011;6:e26217.
[74] Mittal N, Roy N, Babu MM, Janga SC. Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks. Proc Natl Acad Sci U S A 2009;106:20300–5.
[75] Gopinath SCB. Mapping of RNA–protein interactions. Anal Chim Acta 2009;636:117–28.
[76] Keene JD. Minireview: global regulation and dynamics of ribonucleic acid. Endocrinology 2010;151:1391–7.
[77] Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biol 2008;6:e255.
[78] Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 2010;141:129–41.
[79] Thomson AM, Rogers JT, Walker CE, Staton JM, Leedman PJ. Optimized RNA gel-shift and UV cross-linking assays for characterization of cytoplasmic RNA–protein interactions. Biotechniques 1999;27:1032–9 [1042].
[80] Rattenbacher B, Bohjanen PR. Evaluating posttranscriptional regulation of cytokine genes. Methods Mol Biol 2012;820:71–89.
[81] Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, et al. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol 2011;12:R79.
[82] Ward AM, Bidet K, Yinglin A, Ler SG, Hogue K, Blackstock W, et al. Quantitative mass spectrometry of DENV-2 RNA-interacting proteins reveals that the DEAD-box RNA helicase DDX6 binds the DB1 and DB2 3′ UTR structures. RNA Biol 2011;8:1173–86.
[83] Blackwell BJ, Lopez MF, Wang J, Krastins B, Sarracino D, Tollervey JR, et al. Protein interactions with piALU RNA indicates putative participation of retroRNA in the cell cycle, DNA repair and chromatin assembly. Mob Genet Elem 2012;2:26–35.

[84] Goraczniak R, Gunderson SI. The regulatory element in the 3′-untranslated region of human papillomavirus 16 inhibits expression by binding CUG-binding protein 1. J Biol Chem 2008;283:2286–96.
[85] Bose SK, Sengupta TK, Bandyopadhyay S, Spicer EK. Identification of Ebp1 as a component of cytoplasmic bcl-2 mRNP (messenger ribonucleoprotein particle) complexes. Biochem J 2006;396:99–107.
[86] Harris D, Zhang Z, Chaubey B, Pandey VN. Identification of cellular factors associated with the 3′-nontranslated region of the hepatitis C virus genome. Mol Cell Proteomics 2006;5:1006–18.
[87] Kar A, Fushimi K, Zhou X, Ray P, Shi C, Chen X, et al. RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5′ splice site. Mol Cell Biol 2011;31:1812–21.
[88] Iioka H, Loiselle D, Haystead TA, Macara IG. Efficient detection of RNA–protein interactions using tethered RNAs. Nucleic Acids Res 2011;39:e53.
[89] Miniard AC, Middleton LM, Budiman ME, Gerber CA, Driscoll DM. Nucleolin binds to a subset of selenoprotein mRNAs and regulates their expression. Nucleic Acids Res 2010;38:4807–20.
[90] Mehta A, Driscoll DM. A sequence-specific RNA-binding protein complements apobec-1 To edit apolipoprotein B mRNA. Mol Cell Biol 1998;18:4426–32.
[91] Patel GP, Ma S, Bag J. The autoregulatory translational control element of poly(A)-binding protein mRNA forms a heteromeric ribonucleoprotein complex. Nucleic Acids Res 2005;33:7074–89.
[92] Talukdar I, Sen S, Urbano R, Thompson J, Yates III JR, Webster NJG. hnRNP A1 and hnRNP F modulate the alternative splicing of exon 11 of the insulin receptor gene. PLoS One 2011;6:e27869.
[93] Peddigari S, Li PW-L, Rabe JL, Martin SL. hnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition. Nucleic Acids Res 2013;41:575–85.
[94] Butter F, Scheibe M, Mörl M, Mann M. Unbiased RNA–protein interaction screen by quantitative proteomics. Proc Natl Acad Sci U S A 2009;106:10626–31.
[95] Zielinski J, Kilk K, Peritz T, Kannanayakal T, Miyashiro KY, Eiríksdóttir E, et al. In vivo identification of ribonucleoprotein–RNA interactions. PNAS 2006;103:1557–62.
[96] Walker SC, Scott FH, Srisawat C, Engelke DR. RNA affinity tags for the rapid purification and investigation of RNAs and RNA–protein complexes. Methods Mol Biol 2008;488:23–40.
[97] Ross AF, Oleynikov Y, Kislauskis EH, Taneja KL, Singer RH. Characterization of a beta-actin mRNA zipcode-binding protein. Mol Cell Biol 1997;17:2158–65.
[98] Wilson DS, Szostak JW. In vitro selection of functional nucleic acids. Annu Rev Biochem 1999;68:611–47.
[99] Piekna-Przybylska D, Liu B, Fournier MJ. The U1 snRNA hairpin II as a RNA affinity tag for selecting snoRNP complexes. Methods Enzymol 2007;425:317–53.
[100] Keryer-Bibens C, Barreau C, Osborne HB. Tethering of proteins to RNAs by bacteriophage proteins. Biol Cell 2008;100:125–38.
[101] Srisawat C, Engelke DR. Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. RNA 2001;7:632–41.
[102] Blencowe BJ, Sproat BS, Ryder U, Barabino S, Lamond AI. Antisense probing of the human U4/U6 snRNP with biotinylated 2′-OMe RNA oligonucleotides. Cell 1989;59:531–9.
[103] Lingner J, Cech TR. Purification of telomerase from Euplotes aediculatus: requirement of a primer 3′ overhang. Proc Natl Acad Sci U S A 1996;93:10712–7.
[104] Soeno Y, Taya Y, Stasyk T, Huber LA, Aoba T, Hüttenhofer A. Identification of novel ribonucleo–protein complexes from the brain-specific snoRNA MBII-52. RNA 2010;16:1293–300.
[105] Tsai BP, Wang X, Huang L, Waterman ML. Quantitative profiling of in vivo-assembled RNA–protein complexes using a novel integrated proteomic approach. Mol Cell Proteomics 2011;10 [M110.007385].
[106] Bell TJ, Eiríksdóttir E, Langel U, Eberwine J. PAIR technology: exon-specific RNA-binding protein isolation in live cells. Methods Mol Biol 2011;683:473–86.
[107] Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng 2009;11:49–79.
[108] Ranish JA, Yudkovsky N, Hahn S. Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. Genes Dev 1999;13:49–63.
[109] Ong S-E. The expanding field of SILAC. Anal Bioanal Chem 2012;404:967–76.
[110] Mann M. Functional and quantitative proteomics using SILAC. Nat Rev Mol Cell Biol 2006;7:952–8.
[111] Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem 2007;389:1017–31.
[112] Bantscheff M, Lemeer S, Savitski MM, Kuster B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem 2012;404:939–65.
[113] Nahnsen S, Bielow C, Reinert K, Kohlbacher O. Tools for label-free peptide quantification. Mol Cell Proteomics 2013;12:549–56.
[114] Pan S, Aebersold R. Quantitative proteomics by stable isotope labeling and mass spectrometry. In: Matthiesen R, editor. Mass spectrometry data analysis in proteomics. Humana Press; 2007. p. 209–18.
[115] Himeda CL, Ranish JA, Pearson RCM, Crossley M, Hauschka SD. KLF3 regulates muscle-specific gene expression and synergizes with serum response factor on KLF binding sites. Mol Cell Biol 2010;30:3430–43.
[116] Hsu J-L, Huang S-Y, Chow N-H, Chen S-H. Stable-isotope dimethyl labeling for quantitative proteomics. Anal Chem 2003;75:6843–52.
[117] Butter F, Kappei D, Buchholz F, Vermeulen M, Mann M. A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth. EMBO Rep 2010;11:305–11.
[118] Ly L, Wasinger VC. Protein and peptide fractionation, enrichment and depletion: tools for the complex proteome. Proteomics 2011;11:513–34.
[119] Stros M. HMGB proteins: interactions with DNA and chromatin. Biochim Biophys Acta 2010;1799:101–13.
[120] Bulyk ML. Computational prediction of transcription-factor binding site locations. Genome Biol 2003;5:201.
[121] Simicevic J, Deplancke B. DNA-centered approaches to characterize regulatory protein–DNA interaction complexes. Mol Biosyst 2010;6:462–8.
[122] Gingras A-C, Gstaiger M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol 2007;8:645–54.