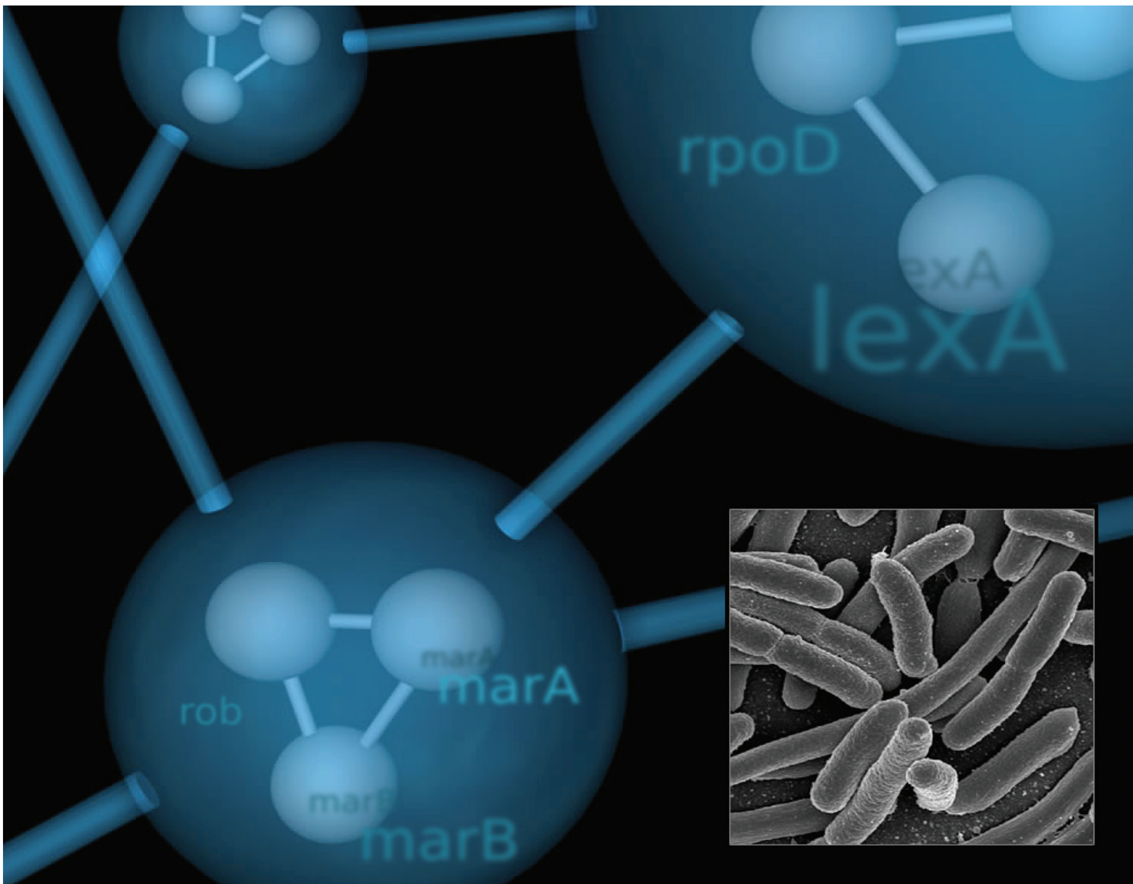
Molecular BioSystems

This article was published as part of the

Computational and Systems Biology themed issue

Please take a look at the full table of contents to access the other papers in this issue.



DNA-centered approaches to characterize regulatory protein–DNA interaction complexes†

Jovan Simicevic and Bart Deplancke*

Received 5th August 2009, Accepted 6th October 2009
First published as an Advance Article on the web 2nd December 2009
DOI: 10.1039/b916137f

Gene regulation is mediated by site-specific DNA-binding proteins or transcription factors (TFs), which form protein complexes at regulatory loci either to activate or repress the expression of a target gene. The study of the dynamic properties of these regulatory DNA-binding complexes has so far been dominated by protein-centered methodologies, aiming to characterize the DNA-binding behavior of one specific protein at a time. With the emerging evidence for a role of DNA in allosterically influencing DNA-binding protein complex formation, there is renewed interest in DNA-centered approaches to capture protein complexes on defined regulatory loci and to correlate changes in their composition with alterations in target gene expression. In this review, we present the current state-of-the-art in such DNA-centered approaches and evaluate recent technological improvements in the purification as well as in the identification of regulatory DNA-binding protein complexes within or outside their biological context. Finally, we suggest possible areas of improvement and assess the putative impact of DNA-centered methodologies on the gene regulation field for the forthcoming years.

Why DNA-centered methods?

Differential gene expression is central to most fundamental biological processes and is controlled by site-specific DNA binding protein complexes. The latter transcriptional complexes, of which transcription factors (TFs) are the core members, function by integrating extra- and intracellular cues through protein–protein or protein–ligand interactions and translating these cues into a gene regulatory output by binding to gene regulatory elements.[1] Signal integration can thereby be directly mediated by the TF itself, for example, through post-translational modification (PTM) of TF domains which modulates its activity[2] or cellular location,[3] or can be controlled indirectly through interaction with co-regulators. These higher-order interactions can result in, or can also be the result of PTMs, and can then determine whether the TF-containing complex acts as an activator or repressor of gene expression. This concept has perhaps been best characterized for nuclear receptor TFs, for which multiple PTM-dependent co-activator and co-repressor complexes have been identified (*e.g.* reviewed in ref. 4 and 5). It is currently unclear to what extent these higher-order protein interactions and resulting PTMs can affect DNA binding specificities and affinities. Given that TFs have often been observed to act both as activators and repressors,[4,6,7] it is possible that the associated DNA binding complex influences DNA binding specificity or affinity, for example to distinguish genes that need to be repressed from those that need to be activated. This could occur through modulation not only of the TF DNA binding domain,[8] but also of regions located outside the DNA binding domain that can alter DNA recognition and affinity through protein domain intercommunication.[9–11] In recent years however, there has been increasing evidence for the reverse notion of the DNA dictating complex formation rather than the DNA binding complex differentiating between gene targets.[12–15] In other words, while the implicated TF(s) is still responsible for target gene identification, the nucleotide composition of the respective TF binding site allosterically influences co-regulator recruitment and thus whether the resulting DNA binding complex will activate or repress gene expression (Fig. 1). So far, this phenomenon has been elucidated for only a couple of TFs, including Oct-1,[15] NFκB[13] and glucocorticoid receptor,[14] but it is possible that this regulatory principle extends to many, if not most, other TFs. To validate this, it will be important to perform a comprehensive analysis of transcriptional complexes while bound to DNA, ideally without using protein-specific antibodies as this would significantly limit the experimental scope since such antibodies are available for only a low number of TFs and co-regulators. This DNA-centered approach to transcriptional regulation would also allow assessment of the dynamic properties of these complexes, as the same "DNA bait" could be used in distinct biological contexts. With the rapidly growing amount of experimentally defined regulatory element data (*e.g.* the Fantom and Encode consortia[16,17]), there is a wealth of suitable DNA bait candidates which could provide instrumental insights into gene regulatory mechanisms. The use of actual regulatory elements as DNA baits rather than TF binding site-representing double-stranded oligonucleotides may thereby be more informative given the often relatively poor correlation between *in vitro*-derived TF binding sites and *in vivo*-observed binding events.[18]

Ecole Polytechnique Fédérale de Lausanne, School of Life Sciences, Institute of Bioengineering, Station 15, 1015 Lausanne, Switzerland. E-mail: bart.deplancke@epfl.ch; Tel: +41-21-6931821

† This article is part of a *Molecular BioSystems* themed issue on Computational and Systems Biology.

 This journal is © The Royal Society of Chemistry 2010

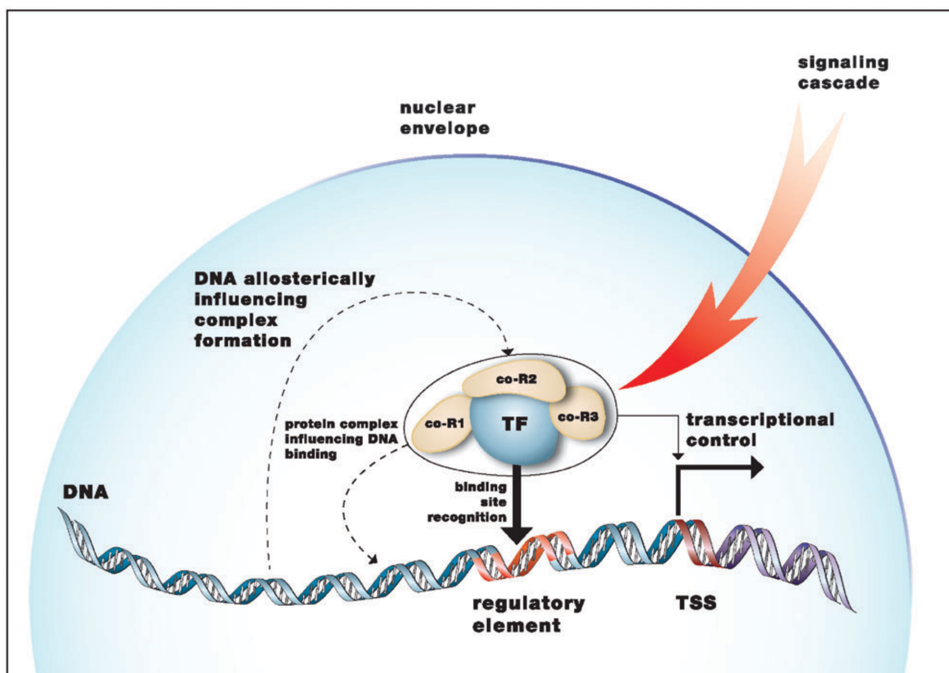


Fig. 1 Drawing illustrating the complex interplay between TF binding site recognition and DNA-binding protein complex formation at a specific regulatory locus.

The practical realization of such DNA-centered analyses has so far been difficult, as there are inherent difficulties associated with studying TF function such as their low expression and involvement in many transient and context-dependent interactions. Nevertheless, in recent years, important experimental progress has been made, which promises to significantly improve our ability to study the dynamic properties of transcriptional complexes in a DNA-centered fashion. Here, we provide a critical overview of these advances by highlighting their technical improvements over previously available DNA-centered methods and by pinpointing the remaining limitations. In addition, we briefly compare their output against protein-centered DNA binding complex detection methods, and highlight the advantages and disadvantages of each strategy.

DNA affinity chromatography

General concept

The most familiar DNA-centered method is DNA affinity chromatography. This approach to study TF-containing complexes is based on DNA bait-mediated protein purification, which is achieved by exploiting the inherent capacity of TFs to bind to DNA. DNA is thereby either absorbed or linked covalently to a chromatographic support before being used for DNA affinity chromatography. Originally, heterogeneous non-specific DNA such as salmon or herring DNA was linked to a cellulose or sepharose chromatographic support (reviewed in ref. 19). This approach is not optimal for the purification of specific DNA binding complexes because of the abundant prevalence of contaminant proteins which non-specifically bind either to the support material or to the DNA. Rather, this approach is now commonly used as one of the many steps involved in TF purification as it efficiently removes contaminating proteins from complex protein mixtures. To subsequently isolate selected TFs or DNA binding complexes, specific DNA sequences are preferred. These are typically double-stranded oligonucleotides either in single-copy or concatemerized format, which represent TF consensus binding sites[20] or very small DNA regions with known DNA binding function identified, for example, through DNAse I footprinting.[21] Since TFs have an affinity several orders of magnitude greater for their consensus binding site sequence compared to non-specific DNA, the use of TF-specific double-stranded oligonucleotides allows a relatively straightforward purification of the respective TF and associated proteins from complex protein mixtures. However, this approach has also important limitations. First, while binding site concatemerization has been the preferred format for TF purification, it also introduces novel DNA sites, increasing the probability that other proteins will bind to the DNA bait and thus reducing purity.[19] Second, DNA binding and complex assembly occurs *in vitro* and since also stringent washing is required, only proteins that bind with high affinity to the respective TF will be retained, making this approach not ideal to study the dynamic properties of DNA binding complexes within their endogenous context. Third, it requires prior knowledge of specific TF binding sites. For a large number of TFs such corresponding binding sites are still unavailable, limiting the scope of this method.[1] Consequently, the approach is TF-centered, and will therefore not provide a comprehensive view of the factors controlling the transcription of your gene of interest. Fourth, such DNA bait typically represents just one of the possible binding site possibilities. This is important given the *in vivo* observation in both prokaryotic[22] and

eukaryotic systems[1] that TFs also bind to sub-optimal binding sites. Thus, while the use of high affinity binding sites will provide a significant insight into DNA-binding protein complexes involving the respective TF, it will be by no means comprehensive. Moreover, it is becoming increasingly clear that multiple binding sites with varying TF affinities can cooperate, for example through DNA looping, to stabilize the TF-containing complex.[23] Thus, individual sites would fail to capture this complex. Finally, double-stranded oligonucleotides or other short DNA fragments usually fail to preserve the same DNA topology as that of the endogenously occurring TF binding site, which has also been shown to affect DNA binding. For example, p53 binding to DNA was enhanced with increasing negative superhelix density.[24]

Recent advances

In recent years, there have been attempts to overcome many limitations by considering the use of single regulatory elements (or at least short fragments thereof) such as enhancers or promoters as DNA baits.[25–28] This approach does not require *a priori* knowledge of TF binding site properties, and since elements are typically linked to a specific gene, it also provides information of immediate relevance to how the respective gene may be transcriptionally controlled. Recent examples of the DNA affinity chromatography approach include the isolation of a *Drosophila* TF, DEAF-1, binding to the enhancer of an immunity gene,[26] as well as several proteins binding to promoter fragments of, respectively, the human *ESRRA* and *MTA*2 genes.[27] In both studies, DNA was immobilized onto a solid phase by biotin-labelling the DNA and coupling it to either streptavidin-coated columns or magnetic beads. This is in contrast with the technique of DNA trapping used by Jiang and co-workers[29] in which a 250 bp region of the human *c-jun* promoter with a single stranded $(GT)_5$ tail was annealed to single-stranded $(AC)_5$-Sepharose. Several pre-initiation components such as RNA polymerase II, TBP, and the TFIF subunit RAP 74 were retrieved as well as the TF SP1. DNA trapping typically supports better purification as streptavidin-coated supports are known to bind to contaminating proteins (Table 1). In addition, DNA trapping allows a non-denaturating elution of bound proteins,[29] which, because of the strength of the interaction, is not possible with DNA immobilization where high temperatures and denaturating agents such as SDS are required to elute proteins. On the other hand, DNA immobilization is less time consuming and amenable to automatization due to its relatively simple workflow, and is therefore often the method of choice. As indicated above, a critical aspect of both techniques is the use of competitor DNA such as salmon sperm DNA, poly(dI:dC) or scrambled bait DNA to eliminate proteins that have low affinity for the DNA bait but would otherwise be retained because of the high concentration of bait DNA.

DNA-binding protein identification

Antibodies

The most straightforward and therefore widespread strategy to identify captured TFs or TF-containing protein complex members is based on the use of antibodies. Their application is thereby not restricted to DNA affinity chromatography,[28] as other protein–DNA interaction detection approaches also benefit greatly from antibody availability. An excellent example is the supershift assay in which the identity of a DNA-binding protein is confirmed only when a protein-specific antibody reduces the electrophoretic mobility of a protein–DNA interaction complex.[30] A significant advantage of such a gel shift procedure over other protein–DNA interaction detection methods is its ability to distinguish single from multimeric forms of bound protein and to immediately relate this information to the respective DNA bait. For example, using supershift assays, Tantin and colleagues[31] determined that DNA baits were more likely to induce di- or multimerization of the TF Oct-4 when they contained at least three Oct-4 half sites. Thus, and as discussed already above, binding site cooperativity can influence the formation of distinct TF complex configurations, with each possibly having a differential impact on how the respective target gene is transcriptionally controlled. Nevertheless, despite their utility, antibodies restrict the scope of the assay as only a limited number of highly specific DNA-binding protein antibodies are currently available. Moreover, antibody implementation requires an *a priori* assumption about the identity of interacting proteins, making this approach protein-centered. Thus, while several protein-centered methods have already contributed in significant fashion to our understanding of the molecular mechanisms underlying protein–DNA interactions *in vitro* and *in vivo* (reviewed extensively in ref. 1), we will not discuss them here given this review's focus on DNA-centered protein–DNA interaction approaches. Instead, we will briefly discuss new efforts to eliminate the protein-centered bias of current DNA bait-based techniques such as gel shift and DNA affinity chromatography by linking them to *de novo* protein detection and identification methods such as two-dimensional gel electrophoresis (2-DE)[32] and mass spectrometry.[33] Since it is desirable still to confirm the identity of 2-DE-detected proteins using mass spectrometry, we will briefly focus on the latter technology.

Mass spectrometry

Traditionally, the detection and identification of DNA-binding proteins or complexes by mass spectrometry has always been difficult owing to the low cellular abundance of the majority of these types of proteins.[34] In recent years, mass spectrometry has become increasingly sensitive, driven by fast-paced technological advances in instrumentation. This significant increase in mass accuracy and resolving power now allows for the first time a more detailed functional analysis of such lowly expressed proteins as their spectral peaks become increasingly distinguishable from background noise in complex mixtures. Based on overall sensitivity, two mass analyzers stand out. The first is the Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR).[35] The second is the Orbitrap, which also uses an FT-based strategy.[36] For example, Mann and co-workers[27] have used an FT-ICR to identify sequence-specific DNA-binding proteins in HeLa S3 cells purified by DNA affinity chromatography. Interestingly,

This journal is © The Royal Society of Chemistry 2010

Table 1 Summary of the strengths (+) and weaknesses (–) of the discussed DNA-centered methods to characterize regulatory protein–DNA interaction complexes

			Time-consuming	Unambiguous protein purification/identification	Unambiguous complex purification/identification	Prone to artificial DNA binding	Unbiased (truly DNA-centered) interaction screen	Protein complex characterization in its natural context
Homogeneous DNA				–	–	–		–
Heterogeneous DNA	Short DNA sequences as DNA bait	TF-specific double-stranded oligonucleotide trapping	–	+	–			–
		TF-specific double-stranded oligonucleotide immobilization		+	–			–
		Concatemers		+	–	–		–
	Long DNA sequences as DNA bait	DNA trapping	–	+	+		+	–
		DNA immobilization		+	+		+	–
PICh			–	+	+		+	+
Supershift (Electrophoretic Mobility Super Shift Assay)			–	+	–	–	–	–
High-throughput yeast one-hybrid				+	–	–		–

when comparing the eluted protein SDS-PAGE profiles from the wild-type and negative control DNA bait, there was virtually no difference and thus no clear bands were revealed corresponding to true specific DNA-binding proteins. Nonetheless, because of the sensitivity of FT-ICR, candidate DNA-binding proteins that were more abundant in the wild-type *versus* negative control samples were in the end identified (see also below). Protein-centered approaches aiming to characterize TF-specific protein interaction partners or complexes are also benefiting greatly from the recent sensitivity increase as evidenced by the fact that many of the detected TF interactors were themselves TFs.[37,38] Thus, we are entering an exciting era in which proteins such as TFs that have traditionally been for the most part off-limit become increasingly accessible and thus characterizable.

DNA-binding dynamics

To achieve a comprehensive, mechanistic understanding of gene regulation, it is essential to not only determine the identity of regulatory DNA-binding complex members, but to also chart compositional complex changes in relation to alterations in target gene expression. This need to monitor protein complex assembly dynamics either with other proteins or with DNA has prompted the development of quantitative proteomics approaches (*e.g.* reviewed in ref. 39). The latter involve the labelling of proteins with isotopically distinguishable tags enabling a protein abundance comparison between two or more biological samples. Brand *et al.*[37] used isotope-coded affinity tagging (ICAT) to monitor the compositional changes of the protein complex involving the TF NF-E2p18/MafK during erythroid differentiation. Results uncovered more than 100 potential protein interactors and indicated that MafK acts as a dual-function TF, exchanging dimerization partners upon induction of differentiation, leading to the replacement of interacting co-repressors with co-activators and up-regulation of the expression of its target gene β-globin. To answer questions related to the molecular mechanisms underlying this protein partner exchange, the next step would be to monitor the compositional changes of only those MafK-containing complexes that are bound to DNA. Although this is in principle feasible by monitoring the DNA occupation of individual complex members at distinct time points using chromatin immunoprecipitation,[37,40] the unavailability of antibodies for the majority of proteins limits, as indicated previously, the scope of such assays and thus prevents a functional analysis of the majority of detected protein interaction partners. Moreover, similar to other recent mass spectrometry-based TF-protein interaction detection techniques such as the streptavidin-mediated isolation of biotinylated TF complexes,[38] the approach used by Brand and colleagues[37] is again strictly TF or protein-centered and may therefore miss crucial factors that may influence MafK complex assembly on the DNA and thus β-globin gene regulation in general without physically interacting with MafK, but for example by altering DNA accessibility.[41] With this experimental mindset, Mittler *et al.*[27] combined the "stable isotope labelling with amino acids in cell culture" (SILAC) technique with DNA affinity chromatography to detect protein–DNA complex assembly differences on wild-type *versus* mutated TF binding sites or short regulatory element fragments. For both types of DNA baits, a significant number of putative binding proteins were found. Since most of these were captured in approximately equal amounts by the wild-type and negative control bait, they could however be eliminated. The identity of the remaining proteins was in line with predictions and proved the value of their method. In addition, these researchers were able to identify proteins, many of which were previously not described, that preferentially bind to methylated *versus* non-methylated CpG sites on the *MTA2* gene promoter. While the latter method clearly increases our ability to determine TF binding profiles, it still suffers from the previously mentioned important limitation that DNA binding and putative complex assembly occurs *in vitro* and thus information regarding

This journal is © The Royal Society of Chemistry 2010

the complex composition at the corresponding endogenous locus is lost.

Alternative DNA-centered approaches

PICh

To enable the *in vivo* assessment of regulatory DNA-binding complexes at specific gene loci, Déjardin and Kingston[42] have now developed the PICh (Proteomics of Isolated Chromatin segments) method. This method is a drastic departure from previous methods typically based on DNA affinity chromatography as it is better described as a reverse chromatin immunoprecipitation, since it uses cross-linking to fix protein complexes on DNA, but rather than using a protein-specific antibody to identify bound DNA regions, it employs a DNA element-specific probe to pull down the associated protein complex (Fig. 2). The probe is an oligonucleotide containing locked nucleic acid (LNA) residues. These have an altered backbone that favours base stacking, thereby significantly increasing the stability of probe–DNA interactions. After the probe has hybridized to chromatin cross-linked to protein complexes, it is captured on streptavidin magnetic beads through a desthiobiotin molecule covalently linked to the probe. Desthiobiotin is a biotin analog with weaker affinity for avidin which therefore permits a more gentle competitive elution using biotin, limiting the co-elution of non-specific factors. Thus, by maintaining the DNA-bound protein complex in its natural state and because probes can be designed against any locus, PICh provides in principle the possibility to correlate protein complex composition changes with alterations in the expression of any target gene. However, PICh has so far only been used to detect proteins associated with human telomere sequences, which, with around 100 copies per cell, are rather abundant in the genome and therefore compensate for the relatively low protein detection

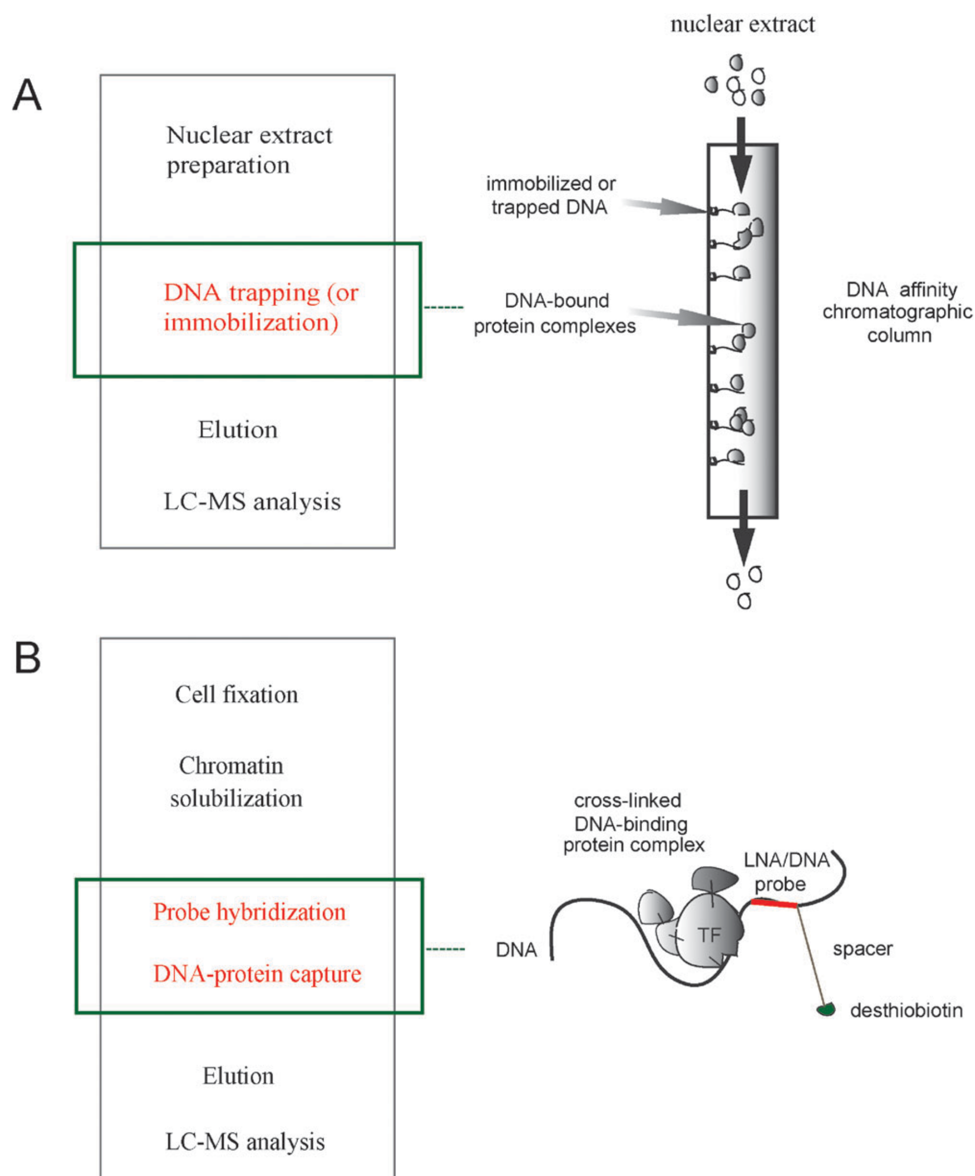


Fig. 2 Drawing illustrating the protein complex purification workflow using DNA affinity chromatography (A) and PICh (B).

 This journal is © The Royal Society of Chemistry 2010

sensitivity of the method. Evolving the method to allow screening of less abundant chromatin loci or even unique regulatory elements is now the next challenge and could involve a reconsideration of probe design, the use of even more sensitive mass spectrometers, or the integration of quantitative proteomics techniques.

High-throughput yeast one-hybrid assay

While PICh has the promise to revolutionize the gene regulation field, it remains to be seen how much the method will live up to expectations. Consequently, DNA affinity chromatography as well as other alternative methods will remain useful to study the dynamic properties of regulatory DNA-binding complexes in DNA-centered fashion. One other alternative method is the high-throughput yeast one-hybrid system, which allows the screening of regulatory elements of interest for interacting TFs or TF dimers.[43–45] Although the latter technique does not allow the detection of DNA-binding complexes and is performed in yeast and thus outside the endogenous context, it provides the unique possibility to scan the whole regulatory protein repertoire for binding to a DNA bait of choice depending on the completeness of the screened TF library.[46,47]

Conclusion

In recent years, there has been renewed interest in obtaining a complete understanding of the complex mechanisms underlying gene regulation, driven by recent discoveries illustrating the complex interplay between all components involved (DNA, TFs, co-regulators *etc.*) in guiding the formation of functional regulatory complexes, which either activate or repress gene expression. Consequently, there are revived efforts to improve current technologies to enable an increasingly more accurate and comprehensive study of gene regulatory complex formation. Specifically, there is a significant need to monitor the formation of such complexes at defined regulatory loci in distinct biological contexts, hence the renewed interest in DNA-centered protein–DNA interaction detection technologies. Although progress in this area has been made as discussed in this review, we are still far from the complete and functional characterization of DNA-binding protein complexes and from the ability to relate changes in their composition to expression changes of their respective target genes. In this regard, we are eagerly looking forward to novel developments in the *in vivo* quantitative proteomics field, to improved TF and protein complex purification methods, and to further increases in the sensitivity of mass spectrometers, which are quickly becoming the ''gold standard'' in the analysis of DNA-binding functional protein complexes.

Abbreviations

TF	Transcription Factor
PTM	Post-Translational Modification
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
2-DE	Two-dimensional Gel Electrophoresis
FT	Fourier Transform
FT-ICR	Fourier Transform Ion Cyclotron Resonance
ICAT	Isotope-Coded Affinity Tags
SILAC	Stable Isotope Labelling with Amino Acids in Cell Culture
PICh	Proteomics of Isolated Chromatin Segments
LNA	Locked Nucleic Acid
Co-R	Co-regulatory element

Acknowledgements

We thank the members of the Deplancke lab for reading of the manuscript, and Michele Zinnai for his significant help with the preparation of the figures. Funding is provided by the Swiss National Science Foundation, the SystemsX Initiative, and the European Union (IRG, FP7); institutional support from the Ecole Polytechnique Fédérale de Lausanne.

References

1 B. Deplancke, *Briefings Funct. Genomics Proteomics*, 2009, **8**, 12–27.
2 S. H. Yang, E. Jaffray, R. T. Hay and A. D. Sharrocks, *Mol. Cell*, 2003, **12**, 63–74.
3 N. C. Reich and L. Liu, *Nat. Rev. Immunol.*, 2006, **6**, 602–612.
4 M. G. Rosenfeld, V. V. Lunyak and C. K. Glass, *Genes Dev.*, 2006, **20**, 1405–1428.
5 B. W. O'Malley, J. Qin and R. B. Lanz, *Curr. Opin. Cell Biol.*, 2008, **20**, 310–315.
6 N. Soontorngun, M. Larochelle, S. Drouin, F. Robert and B. Turcotte, *Mol. Cell. Biol.*, 2007, **27**, 7895–7905.
7 J. S. Reece-Hoyes, B. Deplancke, M. I. Barrasa, J. Hatzold, R. B. Smit, H. E. Arda, P. A. Pope, J. Gaudet, B. Conradt and A. J. M. Walhout, *Nucleic Acids Res.*, 2009, **37**, 3689–3698.
8 C. W. Garvie, J. Hagman and C. Wolberger, *Mol. Cell*, 2001, **8**, 1267–1276.
9 V. Chandra, P. Huang, Y. Hamuro, S. Raghuram, Y. Wang, T. P. Burris and F. Rastinejad, *Nature*, 2008, 350–356.
10 H. Gronemeyer and W. Bourguet, *Sci. Signal.*, 2009, **2**, pe34-.
11 M. A. Pufall, G. M. Lee, M. L. Nelson, H.-S. Kang, A. Velyvis, L. E. Kay, L. P. McIntosh and B. J. Graves, *Science*, 2005, **309**, 142–145.
12 N. Ha, K. Hellauer and B. Turcotte, *Nucleic Acids Res.*, 1996, **24**, 1453–1459.
13 T. H. Leung, A. Hoffmann and D. Baltimore, *Cell*, 2004, **118**, 453–464.
14 S. H. Meijsing, M. A. Pufall, A. Y. So, D. L. Bates, L. Chen and K. R. Yamamoto, *Science*, 2009, **324**, 407–410.
15 A. Tomilin, A. Reményi, K. Lins, H. Bak, S. Leidel, G. Vriend, M. Wilmanns and H. R. Schöler, *Cell*, 2000, **103**, 853–864.
16 The FANTOM Consortium, *Science*, 2005, **309**, 1559–1563.
17 The ENCODE Project Consortium, E. Birney, J. Stamatoyannopoulos, A. Dutta, R. Guigo, T. Gingeras, E. Margulies, Z. Weng, M. Snyder, E. Dermitzakis, R. Thurman, M. Kuehn, C. Taylor, S. Neph, C. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J. Greenbaum, R. Andrews, P. Flicek, P. Boyle, H. Cao, N. Carter, G. Clelland, S. Davis, N. Day, P. Dhami, S. Dillon and M. Dorschner, *Nature*, 2007, **447**, 799–816.
18 C. E. Massie and I. G. Mills, *EMBO Rep.*, 2008, **9**, 337–343.
19 H. Gadgil, L. A. Jurado and H. W. Jarrett, *Anal. Biochem.*, 2001, **290**, 147–178.
20 M. Yaneva and P. Tempst, *Anal. Chem.*, 2003, **75**, 6437–6448.
21 E. Ruteshouser and B. de Crombrugghe, *J. Biol. Chem.*, 1992, **267**, 14398–14404.
22 D. C. Grainger and S. J. W. Busby, *Biochem. Soc. Trans.*, 2008, **036**, 754–757.
23 L. Saiz and J. M. G. Vilar, *Curr. Opin. Struct. Biol.*, 2006, **16**, 344–350.
24 E. B. Jagelská, V. c. Brazda, P. Pecinka, E. Palecek and M. Fojta, *Biochem. J.*, 2008, **412**, 57–63.

25 N. Ahmad and J. B. Lingrel, *Biochemistry*, 2005, **44**, 6276–6285.
26 D. E. Reed, X. M. Huang, J. A. Wohlschlegel, M. S. Levine and K. Senger, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 8351–8356.
27 G. Mittler, F. Butter and M. Mann, *Genome Res.*, 2009, **19**, 284–293.
28 A. Krehan, H. Ansuini, O. Bocher, S. Grein, U. Wirkner and W. Pyerin, *J. Biol. Chem.*, 2000, **275**, 18327–18336.
29 D. Jiang, R. A. Moxley and H. W. Jarrett, *J. Chromatogr., A*, 2006, **1133**, 83–94.
30 L. M. Hellman and M. G. Fried, *Nat. Protoc.*, 2007, **2**, 1849–1861.
31 D. Tantin, M. Gemberling, C. Callister and W. Fairbrother, *Genome Res.*, 2008, **18**, 631–639.
32 J. A. Stead, J. N. Keen and K. J. McDowall, *Mol. Cell. Proteomics*, 2006, **5**, 1697–1702.
33 E. Nordhoff, A.-M. Krogsdam, H. F. Jorgensen, B. H. Kallipolitis, B. F. C. Clark, P. Roepstorff and K. Kristiansen, *Nat. Biotechnol.*, 1999, **17**, 884–888.
34 J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann and N. M. Luscombe, *Nat. Rev. Genet.*, 2009, **10**, 252–263.
35 M. B. Comisarow and A. G. Marshall, *J. Mass Spectrom.*, 1996, **31**, 581–585.
36 A. Makarov, *Anal. Chem.*, 2000, **72**, 1156–1162.
37 M. Brand, J. A. Ranish, N. T. Kummer, J. Hamilton, K. Igarashi, C. Francastel, T. H. Chi, G. R. Crabtree, R. Aebersold and M. Groudine, *Nat. Struct. Mol. Biol.*, 2004, **11**, 73–80.
38 J. Wang, S. Rao, J. Chu, X. Shen, D. N. Levasseur, T. W. Theunissen and S. H. Orkin, *Nature*, 2006, **444**, 364–368.
39 A.-C. Gingras, M. Gstaiger, B. Raught and R. Aebersold, *Nat. Rev. Mol. Cell Biol.*, 2007, **8**, 645–654.
40 J. Sun, M. Brand, Y. Zenke, S. Tashiro, M. Groudine and K. Igarashi, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 1461–1466.
41 E. Segal and J. Widom, *Nat. Rev. Genet.*, 2009, **10**, 443–456.
42 J. Déjardin and R. E. Kingston, *Cell*, 2009, **136**, 175–186.
43 B. Deplancke, A. Mukhopadhyay, W. Ao, A. M. Elewa, C. A. Grove, N. J. Martinez, R. Sequerra, L. Doucette-Stamm, J. S. Reece-Hoyes, I. A. Hope, H. A. Tissenbaum, S. E. Mango and A. J. Walhout, *Cell*, 2006, **125**, 1193–1205.
44 B. Deplancke, D. Dupuy, M. Vidal and A. J. Walhout, *Genome Res.*, 2004, **14**, 2093–2101.
45 G. Chen and J. A. Shin, *Anal. Biochem.*, 2008, **382**, 101–106.
46 J. S. Reece-Hoyes, B. Deplancke, J. Shingles, C. A. Grove, I. A. Hope and A. J. Walhout, *Genome Biol.*, 2005, **6**, R110.
47 V. Vermeirssen, B. Deplancke, M. I. Barrasa, J. S. Reece-Hoyes, H. E. Arda, C. A. Grove, N. J. Martinez, R. Sequerra, L. Doucette-Stamm, M. R. Brent and A. J. Walhout, *Nat. Methods*, 2007, **4**, 659–664.

This journal is © The Royal Society of Chemistry 2010