Review

Structural insights into transcription complexes

Imre Berger [a], Alexandre G. Blanco [b,c], Rolf Boelens [d], Jean Cavarelli [e,f,g,h], Miquel Coll [b,c], Gert E. Folkers [b], Yan Nie [a], Vivian Pogenberg [i], Patrick Schultz [e,f,g,h], Matthias Wilmanns [i], Dino Moras [e,f,g,h,*], Arnaud Poterszman [e,f,g,h,*]

[a] EMBL-Grenoble, BP 181, 6 Rue Jules Horowitz, 38042 Grenoble Cedex 9, France
[b] Institut de Biologia Molecular de Barcelona (CSIC), Barcelona Science Park, Baldiri Reixac 10, 08028 Barcelona, Spain
[c] Institute for Research in Biomedicine, Barcelona Science Park, Baldiri Reixac 10, 08028 Barcelona, Spain
[d] Bijvoet Center for Biomolecular Research, NMR Spectroscopy, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
[e] Institut de Génétique et de Biologie Moléculaire et Cellulaire, BP 163, 67404 Illkirch Cedex, France
[f] Institut National de Santé et de Recherche Médicale, U964 Illkirch, France
[g] Centre National de Recherche Scientifique, UMR 7104 Illkirch, France
[h] Université de Strasbourg, Illkirch, France
[i] EMBL-Hamburg, Hamburg Outstation, Building 25A, DESY, Notkestrasse 85, 22603 Hamburg, Germany

ARTICLE INFO

ABSTRACT

Control of transcription allows the regulation of cell activity in response to external stimuli and research in the field has greatly benefited from efforts in structural biology. In this review, based on specific examples from the European SPINE2-COMPLEXES initiative, we illustrate the impact of structural proteomics on our understanding of the molecular basis of gene expression. While most atomic structures were obtained by X-ray crystallography, the impact of solution NMR and cryo-electron microscopy is far from being negligible. Here, we summarize some highlights and illustrate the importance of specific technologies on the structural biology of protein–protein or protein/DNA transcription complexes: structure/function analysis of components the eukaryotic basal and activated transcription machinery with focus on the TFIID and TFIIH multi-subunit complexes as well as transcription regulators such as members of the nuclear hormone receptor families. We also discuss molecular aspects of promoter recognition and epigenetic control of gene expression.

1. Introduction

The ultimate goal of research on transcription is an understanding of transcriptional control and of the capacity of living cells to respond to environmental changes. In the human body, modulation of gene expression is a very complex process as given physiological response involves different stimuli in a time-dependent manner. Complexity stems from multiple interactions between the molecules involved in distinct pathways. The molecular mechanisms governing transcription regulation are of primordial importance and have major biomedical relevance. Although inappropriate regulation or execution of apoptosis leads to disease, such as cancer, there is now evidence for their great therapeutic potential especially if apoptosis could be targeted at defined organs, rather than acting ubiquitously like chemotherapy.

The SPINE2-COMPLEXES consortium whose aim was to develop new methods and technologies for structural analysis of multicomponent complexes was driven by the choice of 'high-value human health targets' and number of them targets are associated with transcription initiation and regulation. We have investigated components the eukaryotic basal and activated transcription machinery with focus on (i) the TFIID and TFIIH general transcription factors as well as (ii) transcription regulators including members of the nuclear hormone receptor family, and have addressed (iii) molecular aspects of promoter recognition and (iv) epigenetic control of gene expression.

This work has benefited from HTP technologies for the structural genomic implemented in the context of SPINE I and has required development of new technologies adapted for the production, characterization and structural analysis of multicomponent assemblies. It has led to methodological developments to cope with technical challenges and to the determination of more than 50 three-dimensional structures of proteins and complexes directly involved in transcription and its control. While most atomic structures were obtained by X-ray crystallography (Table 1), the

* Corresponding author at: Institut de Génétique et de Biologie Moléculaire et Cellulaire, BP 163, 67404 Illkirch Cedex, France.
E-mail addresses: Dino.MORAS@igbmc.fr (D. Moras), Arnaud.POTERSZMAN@igbmc.fr (A. Poterszman).

Table 1
List of representative structure solved.

Complex	Protein (s)	Ligand	Access number	Method	Resolution	Reference
PhoB complex	PhoB,o4, RNAP (3-fiap	DNA		X-ray	4.3 Å	Submitted
Transcription factor IID (TFIID)	15 subunits	none	EMD-5026	Cryo-EM	22 Å	Papai et al. (2009)
	15 subunits	DNA	EMD-5075, EMD-5076, EMD-5077, EMD-5078	Cryo-EM	29 Å, 24 Å, 19 Å, 31 Å	Papai et al. (2010)
	TAF3 module	H3K4me3 peptide, Zn2+	2K16, 2K17	NMR	rmsd 0.9 Å, 0.8 Å	van Ingen et al. (2008)
	TAF5 modules	none	2J4B, 2J49	X-ray	2.2, 2.3 Å	Romier et al. (2007)
Transcription factor IIH (TFIIH)	p8-TTD-A	none	2JNJ	NMR	rmsd 0.9 Å	Vitorino et al. (2007)
	Tfb2, Tfb5 (p52 and p8-TTD-A)	none	3DGP, 3DOM	X-ray	2.9, 1.9 Å	Kainov et al. (2008)
Coactivator-Associated arginine methyl transferase I (CARM1)	CARM1 modules	none	3B3F, 3B3G, 3B3 J, 2OQB	X-ray	2.2 Å, 2.4 Å, 2.5 Å, 1.7 Å	Troeffer et al. (2007b)
	CARM1 catalytic domains	SFG, ligands analogues	10 structures	X-ray	2,2–2,7A	To be published
Transcription elelongation complexes	Spt6 C-terminal domain	none	2XP1	X-ray	2,20 Å	Diebold et al. (2010a,b)
	Iws1/Spt6 complexes	none	2XPL, 2XPN, 2XPO, 2XPP	X-ray	2.2 Å, 1.9 Å, 2.1 Å, 1.7 Å	Diebold et al. (2010a,b)
SAGA complex	ATXN7L3 SCA7	Zn2+	2KKT, 2KKR	NMR	rmsd 0.5 Å, 0.6 Å	Bonnet et al. (2010)
Lac repressor/Lac DNA complexes	Lac repressor	DNA	2KEI, 2KEJ, 2KEK	NMR	rmsd 0.9 Å, 1.0 Å, 1.7 Å	Romanuka et al. (2009)
Ets-1 dimer DNA complex	Ets-1	DNA	2NNY	X-ray	2.8 Å	Lamber et al. (2008)
MafB DNA complexes	MAfB, c-Fos	DNA	2WT7, 2WTY	X-ray	2.3 Å	To be published
Nuclear Hormone receptors	RXR/RAR heterodimer (LBDs)	atRA/LG100754	3A9E	X-ray	2.7 Å	Sato et al. (2010)
	Tribolium castaneum heterodimer EcR/USP ecysone receptor	ponasterone A	2NXX	X-ray	2.7 Å	Iwema et al. (2007)
	Heliothis virescens heterodimer EcR/USP ecdysone receptor	20-hydroxyecdysone	2R40	X-ray	2.4 Å	Browning et al. (2007)
	Amphioxus RXR tetramer	none	3EYB 2HC4, 2HCD	X-ray	2.8 Å, 2.2 Å, 2.6 Å	Tocchini-Valentini et al. (2009), Ciesielski et al. (2007)
	Vitamin D receptor (LBD)	Vit D synthetic ligands	3A32, 3A40 3CS4, 3CS6	X-ray	1.7 Å, 1.4 Å, 2.0 Å, 1.8 Å	Antony et al. (2010), Rochel et al. (2011)
	ERR ligand binding domain	PGC-alpha peptide	3D24	X-ray	2.1 Å	Greschik et al. (2008)
	RXR/VDR heterodimer (DBDs + LBDs)	DNA, Vit D, retinoid	–	cryo-EM	12 Å	Submitted
	heterodimer (DBDs + LBDs)	DNA	–	cryo-EM	12 Å	To be published
	AR DBD WT and T575A mutant	none	–	NMR	rmsd 0.8 Å	To be published

impact of solution NMR and cryo-electron microscopy is far from being negligible. Here, we summarize some highlights and illustrate the importance of specific technologies on the structural biology of protein–protein or protein/DNA transcription complexes.

2. Challenges for sample preparation: New methods for protein complex production

Many important protein complexes such as multicomponent transcription factors exist in very low quantities in their natural hosts, which renders their extraction from endogenous source difficult. Purification techniques such as tandem-affinity purification (TAP) of tagged open reading frames (Rigaut et al., 1999) are now widely used to isolate native complexes for analysis of protein subunit stoichiometry, interactions, post-translational modifications, and in some cases, for structural studies by cryo-electron microscopy (see below) or exceptionally by X-ray crystallography (Kornberg, 2007). Yet, preparation of complexes in the quality and quantity required for high-resolution structural studies from endogenous source, particularly for human targets is often virtually impossible, or requires very large culture volumes. Heterogeneity of the complexes purified from endogenous source further

complicates their study. Often, transcription factor complexes are highly regulated and can exist as mixtures of isoforms differing in subunit composition and/or containing differential post-translational modifications, representing the kaleidoscope of states the specimens were in at the moment of cell disruption for purification. For example, in-depth profiling of endogenously purified human general transcription factor TFIID by high-resolution mass spectrometry revealed 118 unique phosphorylation sites and 54 unique lysine acetylation sites, distributed over the ensemble of the TFIID molecules purified, giving insights into interesting functional details (Mousson et al., 2008; Pijnappel et al., 2009).

Overproduction of recombinant proteins had a decisive impact on biological research and in particular in structural biology. Producing proteins in a heterologous host can furthermore overcome a number of the above outlined impediments. High-level production of proteins of interest can result in several milligrams of high quality sample from comparatively small culture volumes. Thus, recombinant sample production techniques, in particular using *Escherichia coli* as a prokaryotic expression host organism, have become commonplace for the production of proteins of interest in virtually every molecular biology laboratory. Many plasmid-based systems exist for overexpressing proteins in *Escherichia coli* each with their own merit, several are distributed by commercial

suppliers. Eukaryotic proteins may impose particular requirements on the expression host, such as requiring post-translational modifications for activity. Consequently, eukaryotic expression systems have begun to complement prokaryotic expression systems most commonly used are expression systems using mammalian cells or baculovirus systems that infect insect cell cultures (Jarvis, 2009; Nettleship et al., 2010).

Recombinant production of protein complexes with many subunits, in particular for high-resolution structural studies, has its own challenges and intricacies. Many protein subunits in a complex require cloning and combination of many genes for co-expression. In structural biology, especially for crystallization, proteins often need to be modified by truncation, mutation or deletion of low complexity regions to achieve a sample which can form a well-ordered three-dimensional crystal lattice that diffracts the incident X-ray radiation to high resolution. This necessitates a flexible system of gene assembly into multigene expression vectors, which allows for replacement and manipulation of genes encoding for individual subunits in a rapid and uncomplicated fashion.

Within the SPINE2-COMPLEXES consortium a wide panel of cloning strategies and vector sets have been developed to streamline construct design for expression/co-expression screening in Escherichia coli (Busso et al., 2005; de Jong et al., 2006; Berrow et al., 2007; Scheich et al., 2007; Fogg and Wilkinson, 2008; Bieniossek et al., 2009; Unger et al., 2010; Diebold et al., 2011) (Luna-Vargas et al., 2011) as well as in insect and mammalian cells (Aricescu et al., 2006; Berrow et al., 2007; Abdulrahman et al., 2009; Pradeau-Aubreton et al., 2010; Trowitzsch et al., 2010). A variety of new technologies for DNA manipulation including ligation independent or restriction free procedures, in-fusion or gateway approaches are now being used in addition to classical restriction-based strategies (see Busso et al., 2011 for examples and test cases). Partner Grenoble has developed a system for combinatorial gene assembly into multigene expression vectors called ACEMBL. This system (Bieniossek et al., 2009; Nie et al., 2009) uses a single multigene plasmid which is rapidly built from custom-designed, tiny progenitor DNA molecules by a method termed "tandem recombineering" (TR) (Nie et al., 2009). Tandem recombineering exploits the exonuclease activity of T4 DNA polymerase in the absence of nucleotides to create long (20–30 bases) overhangs on double stranded DNA molecules such as PCR fragments or linearized plasmids. By properly designing these long stick ends, genes, regulatory elements or entire expression cassettes can be concatenated and inserted into small plasmids by sequence and ligation independent cloning methods (SLIC). An array of plasmids, called donor and acceptor plasmids, can be conveniently charged with recombinant DNA cargo in this way (Fig. 1). The main specificity of the ACEMBL approach lies in the use of donor and acceptor plasmid molecules that can easily be assembled into multigene constructs containing all desired genes encoding for subunits of a protein complex of choice. The assembly is catalyzed by Cre recombinase, which creates acceptor–donor fusions by joining the plasmids via a short DNA sequence, LoxP, present on each plasmid. The Cre-LoxP reaction is an equilibrium reaction, therefore, all combinations of donor and acceptor plasmid molecules with their selection of genes co-exist in the reaction vessel in which the Cre-fusion is carried out. The combinations can then be selected by challenging with combinations of antibiotic, as the acceptor and donor plasmids each encode for a different resistance marker.

ACEMBL has been originally designed for multigene expression in Escherichia coli and a series of protein complexes, including factors involved in transcription and gene regulation, has been produced by this method (Bieniossek et al., 2009). Automation is a vital prerequisite in contemporary protein complex research. A fully automated pipeline for producing multiprotein complexes in Escherichia coli has been achieved using the TR approach, made possible by the implementation of robust and simple protocols for PCR, gene insertion by SLIC and the reliance of the method on only two enzymes (T4 DNA polymerase and Cre recombinase) for multigene assembly using the TR approach.. The ACEMBL pipeline is described in a separate contribution in this SPINE2-COMPLEXES special issue (Vijayachandran et al., 2011). More recently, the ACEMBL TR pipeline has been extended successfully to include also multigene assembly for complex expression in eukaryotic systems (Kriz et al. 2010; Vijayachandran et al., 2011).

## 3. Insights into the eukaryotic basal transcription machinery

In eukaryotes, the core promoter serves as a platform for the assembly of the transcription preinitiation complex (PIC) that includes transcription factors IIA, IIB, IID, IIE, IIF, IIH, and RNA polymerase II, which function collectively to specify the transcription start site. While RNA polymerase II as well as general transcription factors IIA, IIB and TBP are now characterized at the atomic level (Liu et al., 2010), the structure and architecture of the multisubunit complexes IID (TFIID) and IIH (TFIIH) are still under investigation.

These complexes composed of 14 and 10 subunits, respectively, are difficult to purify to homogeneity and their crystallization is still out of reach. On the way towards an atomic description of theses macromolecular assemblies and to provide insight into functional aspects, Strasbourg follows a multi-scale approach that combines electron microscopy to obtain a global view of the architecture as well as X-ray crystallography and NMR for atomic scale details.

3.1. X-ray and solution structures of the p8/TTD-A TFIIH subunit: structural basis for trichothiodystrophy

The multi-protein transcription factor TFIIH is involved in the transcription of classes I and II genes as well as in DNA repair (Egly, 2001; Mydlikova et al., 2010). Mutations in its XPB, XPD helicase subunit as well as in its p8/Tfb5 subunit (Giglia-Mari et al., 2004; Coin et al., 2006) have been incriminated in trichothiodystrophy (TTD), a rare autosomal recessive multisystem disorder characterized by sulfur-deficient brittle hair, mental and physical retardation, ichthyosis and, in many cases, cutaneous photosensitivity but no predisposition to cancer. To gain insights into the molecular basis of this disease, the Strasbourg team has determined the solution and X-ray structures of the p8/Tfb5 TFIIH subunit isolated (Vitorino et al., 2007) as well as in complex with the p52/Tfb2 (Kainov et al., 2008), another TFIIH component. The minimal complex between Tfb5, the yeast ortholog of p8, and the carboxy-terminal domain of Tfb2, the yeast p52 subunit of TFIIH revealed that these two polypeptides adopt the same fold, forming a compact pseudo-symmetric heterodimer via a β-strand addition and coiled coils interactions between terminal α-helices. Furthermore, Tfb5 protects a hydrophobic surface in Tfb2 from solvent, providing a rationale for the influence of p8 in the stabilization of p52 (Fig. 2A) and explaining why mutations that weaken p8–p52 interactions lead to a reduced intracellular TFIIH concentration and a defect in nucleotide-excision repair, a common feature of TTD cells.

Key to the successful structure determination of a minimal Tfb2:Tfb5 complex was the use of limited proteolysis combined with mass spectrometry to map the Tfb2 domain required for interaction with Tfb5. A bottleneck in the structure determination was the limited quality of the initial crystals which diffracted to 2.6 Å but were difficult to handle. Despite extensive efforts to control cryoprotection, only a minor proportion of crystals exhibited reasonable diffraction and mosaicity, which hampered the possibility to solve the structure using heavy atom derivatives. From 250 crystals tested, only three yielded usable datasets. Of major
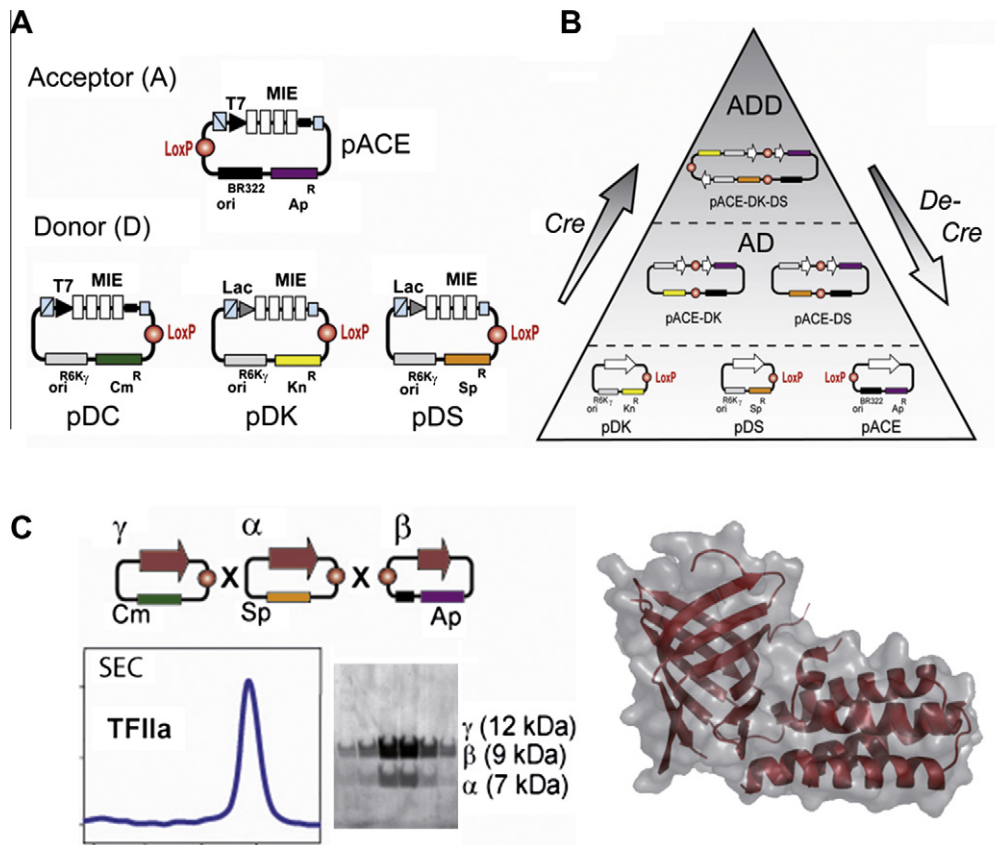
Fig. 1. Protein complex expression by ACEMBL. (A) Genes encoding for subunits of a protein complex are introduced into the multiple integration element (MIE) of small (∼2 kb) plasmid DNA molecules called acceptor and donor. Donors contain a conditional replicon derived from R6 Kγ phage. The acceptor has a regular ColE1 replicon. Promoters (T7, Lac) are indicated. Resistance markers are Ap (ampicillin), Cm (Chloramphenicol), Kn (kanamycin), Sp (spectinomycin). All plasmids contain a LoxP sequence (marked in red). (B) Cre recombinase generates multigene constructs by Cre-LoxP fusion. The multigene constructs are characterized by unique combinations of resistance markers and can be selected for by challenge with the corresponding antibiotics. (C) ACEMBLing transcription factor TFIIA from three subunits (α,β,γ) from a multigene fusion constructed by recombineering. A size exclusion profile (SEC) of the purified complex is shown, with a corresponding SDS–PAGE gel section of the three polypeptides (left). The three-dimensional structure of TFIIA (based on PDB submission 1NH2) is illustrated on the right (panels adapted from Ref. 10, Bieniossek et al., with kind permission of the publisher).

importance was the shortening of the Tfb2 construct, which affects crystal packing along the c axis, leading to a new crystal form. These crystals diffracted to 1.7 Å on a synchrotron beamline, which facilitated heavy atom screening and structure determination (Kainov et al., 2010).

3.2. Cryo-EM structures of TFIID and transcription activation

The general transcription factor TFIID is composed of the TATA binding protein (TBP) and thirteen TBP associated factors (Tafs) which recognize gene promoters in a activator dependant way. The Strasbourg node has determined an improved structural model of the TFIID complex at 23 Å (Papai et al., 2009) and determined the 3-D organization of different TFIID-containing complexes from cryo electron microscopy (CryoEM) images (Papai et al., 2010) to better understand the activator-dependant promoter recruitment of *S. cerevisiae* TFIID. The purification of endogenous TFIID from affinity Tagged yeast strains was instrumental in the production of highly homogeneous complexes. In this respect several yeast strains were prepared in order to introduce different type of tags and to place the tag on different Tafs subunits and to screen the constructs were the integrity of the complex is least affected. For example, when the 140 kDa Taf1 subunit was Tap tagged on its carboxy-terminus, a sub stoichiometric amount of Taf2 was found in the purified TFIID suggesting that this large 37 kDa Tag fragilizes the interaction of Taf2 with the TFIID core. In contrast when the

same Taf1 subunit was HA tagged on its amino-terminus the Taf2 composition was not affected. The interaction of TFIID with DNA was studied in the presence of TFIIA and the CryoEM images revealed that TFIIA interacts close to TBP as predicted by the TBP–TFIIA–DNA crystal structure (Fig. 2B). To exert its coactivator function TFIID was shown in several systems to directly contact transactivators. The Rap1 transactivator was shown to directly bind the TFIID complex through a network of interactions with Taf4, -5, and -12. CryoEM revealed that Rap1 binds to lobe B away from TBP, which is located at the junction of A and C lobes.

In order to obtain deeper insights into the activation mechanism the structure of a committed activation complex formed between Rap1, TFIID and TFIIA, all assembled on a ribosomal enhancer–promoter DNA fragment was determined. A major difficulty in this analysis came from the heterogeneity of the dataset and the use of new methods of particle separation according to defined functional states was instrumental in deciphering this complex mixture of functional states. The results revealed an unexpected interaction between TFIIA and Rap1 which form a protein bridge between TBP and the lobe B-bound Rap1 thus resulting in a large conformational change in the position of TFIIA. We speculate that these rearrangements could (i) stimulate an activator-dependant binding of TBP to the promoter; (ii) stabilize the TFIID–promoter interaction since the protein bridge topologically traps the DNA; or (iii) facilitate subsequent recruitment of TFIIB, Pol II and/or the additional components involved in PIC formation.
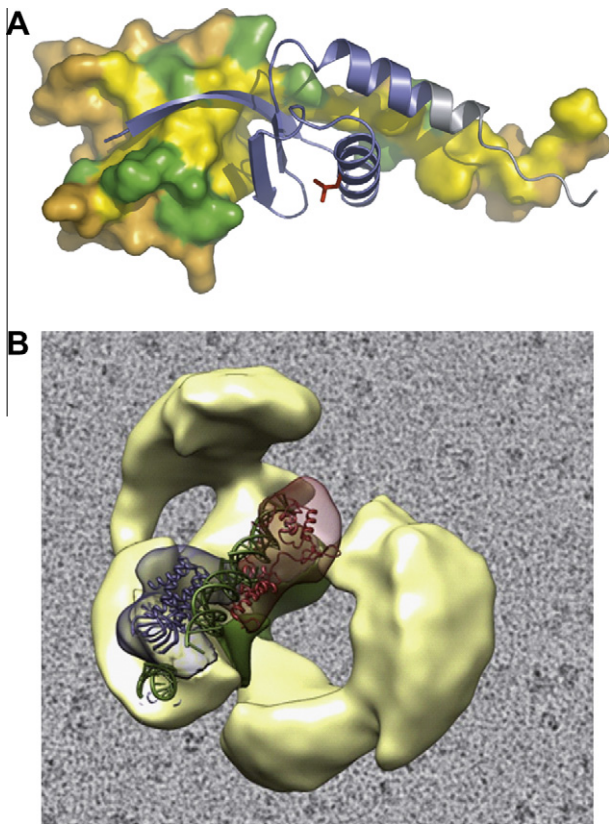
Fig. 2. Insights into the basal Transcription machinery. (A) X-ray structure of the p8/Tfb5 TFIIH subunit in complex with the carboxy-terminal domain of p52/Tfb2, another TFIIH component. Tfb5 is shown as a ribbon lying on the surface of Tfb2C. Hydrophobic side chains in the binding region are in yellow, and others are in green (Kainov et al., 2008). (B) Cryo-EM structure of TFIID in complex with TFIIA and the transactivator Rap1 which cooperate to commit TFIID for transcription initiation (Papai et al., 2010). The analysis of different functional intermediates revealed the mode of binding of Rap1 and TFIIA to TFIID, as well as a Rap1-induced reorganization of TFIIA.

4. Promoter recognition

The transcription factors assemble to DNA to either activate or inhibit transcription of their target genes. These regulatory events are governed by cooperative protein–protein or protein–DNA interactions in a dynamic network of multi-component complexes.

4.1. Structure of the RNAP σ_4-β-flap chimera/PhoB^E/pho box DNA transcription activation sub-complex

One of the strategies that have proven to be successful for the structural characterization of biological complexes is the formation of sub-complexes comprising only the most relevant regions or domains of each of the protein components within the whole complex. The obvious advantage of this approach is that the often difficult purification of full-length proteins can be avoided, but then other drawbacks may arise. The absence of regions that play critical roles in the stabilization of one of the components can be a major problem that can be circumvented by the design and construction of a chimeric protein. This strategy was followed by Barcelona to determine the crystal structure of a ternary transcriptional initiation sub-complex.

PhoB, a two-component response regulator, activates transcription by interacting with the σ^{70} subunit of the *Escherichia coli* RNA polymerase in promoters in which the *pho box* replaces the –35 σ^{70}-recognition sequence. Mutations and carboxy-terminal dele-

tions of σ^{70} had shown the implication of its σ_4 subdomain in the transcriptional activation mediated by PhoB (Makino et al., 1993). Barcelona already had solved the crystal structure showing the tandem DNA recognition by the PhoB effector domain (PhoB^E) (Blanco et al., 2002), but initial efforts to get the structure of the PhoB^E-DNA-σ_4 ternary complex were fruitless because all the σ_4 domain constructs were very poorly expressed or the protein precipitated during the purification process. An analysis of genetic studies (Kuznedelov et al., 2002) and available RNAPH crystal structures (Darst et al., 2001; Murakami et al., 2002; Vassylyev et al., 2002) indicated that σ_4 has a hydrophobic surface that interacts with a region of the RNAP β-flap. This finding inspired the Barcelona group to design a chimera by fusing σ_4 with the β-flap tip helix through an artificial flexible linker. The resulting construct provided a soluble and stable globular domain that could be easily overexpressed in *Escherichia coli*.

Once purified, the σ_4-β-flap chimeric construct was incubated with the PhoB^E-*pho box* DNA complex and the resulting ternary complex was isolated by using size exclusion chromatography. The stability of the complex was assessed by SDS–PAGE and the final sample was subsequently used in crystallization trials. After testing many crystal forms that systematically turned out to be formed by PhoB-DNA binary complexes, a crystal form that enabled the determination of the ternary complex structure was obtained.

The data revealed that σ_4 recognizes the upstream *pho box* repeat (Fig. 3A). As with the –35 element, σ_4 achieves this recognition capacity through the amino-terminal portion of its DNA recognition helix, although in this case the helix is less extended onto the DNA groove. As a consequence, σ_4 establishes less direct contacts with the DNA *pho box* than with the canonical –35 promoter sequence. However, the lost direct contacts of σ_4 with the DNA are compensated by new contacts with the PhoB^E activator which is bound to the *pho-box* as well. This observation suggests a simple recruitment mechanism of the polymerase to the Pho promoters which occurs only in the presence of already-bound transcriptional activator dimer.

4.2. DNA recognition and allosteric regulation by the Lac repressor

The expression of genes involved in the lactose metabolism of *Escherichia coli* is effectively controlled by the Lac repressor (Wilson et al., 2007). The presence of multiple Lac repressor operator binding sites within the *lac* operon is responsible for the effective down regulation of these genes. The main operator O1 overlaps with the lac promoter and is essential for the function of the lac operon. In addition there exist two auxiliary operators O2 and O3, located 401 base pairs (bp) downstream of O1 and 92 bp upstream of O1, respectively, which contribute significantly to the transcriptional repression. Mutation or deletion of O1 leads to an almost complete loss of repression even in the presence of both auxiliary operators, and thus O1 appears indispensable (Betz et al., 1986; Oehler et al., 1990). Inactivation of either O2 or O3 results in a slight decrease of repression, apparently compensating each other, while the combined loss of both O2 and O3 leads to a significant (~70-fold) decrease of repression (Oehler et al., 1990). This cooperativity can be well explained, since the tetrameric Lac repressor functions as a dimer of dimers and binds simultaneously to the O1 operator and to either of the auxiliary O2 and O3 operators creating one of two alternative DNA loops (Kramer et al., 1987).

Mutational studies of the various operators revealed that variation of the sequences leads different affinities for the Lac repressor and results in a distinct repression efficiency (Oehler et al., 1994). O1 and O2 operators have similar base pair composition while the O3 sequence differs significantly. Structural studies of DNA complexes, including those of the Lac repressor, make often use of
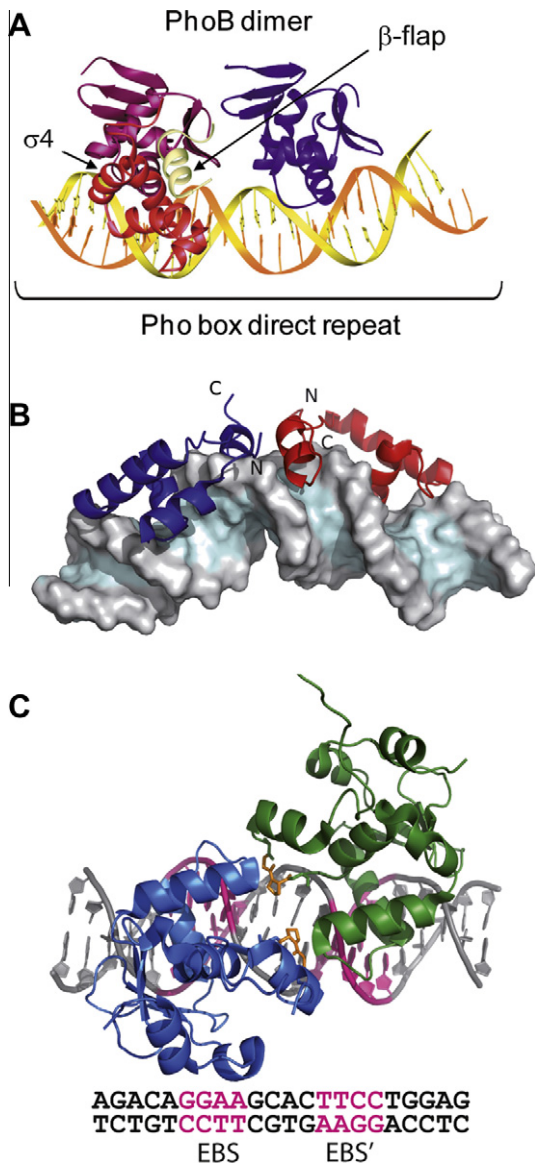
Fig. 3. Promoter recognition. (A) Structure of the RNAP σ₄-β-flap chimera/PhoBᴱ/ *pho box* DNA transcription activation sub-Complex. Ribbon representation of the structure of the quaternary complex showing the upstream (magenta) and downstream (purple) PhoBᴱ protomers and the chimera (σ₄ in red and the β-flap in beige) bound to the *pho box* DNA (gold). (B) Structure of the DNA binding domains (Headpiece HP62V52C) of the Lac repressor in complex with the O2 operator. The left and right Lac HP subunits are coloured dark blue and dark orange, respectively. (C) The structure of Ets-1 homo-dimer bound to the stromelysin-1 promoter element (S-EBS). Ribbon representation of the two components of the Ets-1 homo-dimer, Ets-1 and Ets-1′, colored in blue and green, respectively. The residues of the Glycine-Proline motif are depicted in orange. The two palindromic EBS elements (EBS and EBS′) are shown in magenta on the 22-base pairs DNA duplex corresponding to a fragment of the stromelysin-1 promoter.

symmetrical operators containing two identical half-sites. However the natural lac operators are pseudo-palindromic sequences, where the symmetry is broken by variations in the sequence between the two half-sites and by insertion of the central G:C base pair. When considered separately, the two half-sites can differ significantly in their affinity for the Lac repressor (Sasmor and Betz, 1990).

In an ongoing effort to understand specificity and recognition of various operator sequences by the Lac repressor the Utrecht team determined the NMR structures of the complexes of the dimeric Lac headpiece with its auxiliary operators O2 and O3. The structure

with O2 (Fig. 3A) shows strong similarity with that of the previously determined structure of HP62 with a symmetric SymL operator (Spronk et al., 1999) and that of HP62V52C in complex with O1 (Kalodimos et al., 2002). The Lac HP bound to a non-operator DNA (NOD) fragment is different: a major difference is that the hinge helices, which play an important role in the strong cooperative operator binding of the Headpieces are not formed (Kalodimos et al., 2002) and that of the Lac HP bound to a non-operator DNA (NOD) fragment (Kalodimos et al., 2004). The analysis of these complexes helps to understand how the Lac repressor recognizes its operators and can explain the significant differences in operator affinity (Romanuka et al., 2009).

The structure of the complex of HP62V52C with its auxiliary operator O3 presents a surprise. The left monomer of the Lac repressor in the Lac-O3 complex retains most of these specific contacts, as found in the other operator complexes. However in the right half-site of the O3 operator there is a significant loss of protein–DNA contacts, explaining the low affinity of the Lac repressor for the O3 operator. In fact the binding mode in the right half-site resembles that of the non-specific complex. In contrast to the Lac-non-operator DNA complex however where no hinge helices are formed, the stability of the hinge helices in the weak Lac-O3 complex is the same as in the Lac-O1 and Lac-O2 complexes as judged from the results of the hydrogen–deuterium experiments.

4.3. Oligomeric state and promoter recognition of the Ets-1 transcription factor

The members of the Ets family of transcription factors, which share a common DNA binding domain called ETS domain, play important roles in the development of metazoans and are sometimes involved in oncogenesis (Sharrocks, 2001). During the past fifteen years, the data published on ETS domains highlight how structural biology can provide very powerful tools to understand the mechanisms of recognition of the DNA (Kodandapani et al., 1996), the assembly of activator complexes and regulatory processes like cooperative binding (Garvie et al., 2001), auto-inhibition (Garvie et al., 2002) or post-translational modification (Pufall et al., 2005). However, the previously established mechanism for auto-inhibition of monomeric Ets-1 on DNA response elements with a single ETS-binding site (EBS: 5′-GGA(A/T)-3′) had not been observed for the stromelysin-1 promoter or the P53 promoter containing both two palindromic EBS separated by four base pairs (Venanzoni et al., 1996; Baillat et al., 2002; Baillat et al., 2009).

The Hamburg group has determined the X-ray structure of Ets-1 DNA binding domain on the stromelysin-1 promoter element (S-EBS), revealing a ternary complex in which protein homo-dimerization is mediated by the specific arrangement of the two ETS-binding sites (Fig. 3C). In this complex, both Ets-1 protomers recognize the two EBS *via* conserved residues of the DNA-recognition helix (Arg391, Arg394 and Tyr395) similarly to the way how the monomeric form of Ets-1 interacts with the single EBS. Additional data demonstrated that Ets-1 does not dimerize in solution in the absence of DNA and protein–protein interactions occur when Ets-1 binds to the S-EBS element (Lamber et al., 2008).

Several mutations of the Glycine–Proline motif (Gly333–Pro334), situated on one of the two identified protein–protein interfaces, impaired the recognition of the S-EBS by an Ets-1 dimer and decreased the ability of Ets-1 to transactivate the Stromelysin-1 promoter. The Glycine–Proline motif is not conserved in the whole Ets-1 family and therefore these data suggest that S-EBS-like promoters are specifically regulated by the Ets transcription factors sharing this particular motif (Ets-1 and Ets-2).

Altogether, this work unravels the molecular basis for relief of auto-inhibition and the ability of Ets-1 to function as a facultative dimeric transcription factor on this site. Indeed, in the structure

presented, the amino-terminal ETS-flanking region, which is known to be involved in inhibition of Ets-1 function, is observed to be unfolded when the Ets-1 dimer is bound to S-EBS similarly to what was observed in the context of monomeric Ets-1 bound to EBS. Findings from the Hamburg group may also explain previous data of Ets-1 function in the context of heterologous transcription factors, thus providing a molecular model that could also be valid for Ets-1 regulation by hetero-oligomeric assembly. In this model, the protein–protein interactions within the transcriptional regulator complexes are mediated by DNA binding and directly associated with the release of auto-inhibition.

## 5. Transcription regulation by nuclear hormone receptors

The superfamily of nuclear receptors (NRs) present in vertebrates, arthropods and nematodes plays crucial roles in the regulation of transcription, and is involved in various stages of development, maintaining the control of homeostasis and causing or preventing cellular proliferation, differentiation and death (McEwan, 2009). Some 48 members have been found in the human genome, and a smaller group in arthropoda, housing around 21 in Drosophila melanogaster. Nuclear receptors are ligand-activated transcription factors. Many members of the superfamily thus bind major hormones, such as steroids, thyroid hormones, or retinoids. These occupy a special position in gene regulation by providing a direct link between the ligand, which they bind, and the target gene, whose expression they regulate. Orphan nuclear receptors for which no known ligand has yet been found represent around half of the total number of NRs. These may have empty ligand binding pockets as in the case of estrogen-related receptor-alpha (ERRα). Others have structural ligands that constitutively bind to the LBD, such as the Drosophila USP, but for which no biological function has been established yet.

Nuclear receptors are composed of several functional domains. The amino-terminal A/B domain is highly variable in length and sequence, and contains a constitutively active transactivation function AF-1. C and E correspond to the DNA-binding domain (DBD) and the ligand-binding domain (LBD), respectively. The LBD contains the ligand-dependent transactivation function AF-2. The DBD and LBD are connected via a flexible hinge (domain D). NRs act in vivo and in vitro as ligand-dependant transcriptional regulators through binding, most often as dimers, to DNA response elements present in promoters of target genes. Activation of gene transcription occurs after binding of ligand, leading to release of corepressor and binding of coactivator to the LBD. To date, the crystal structures of more than 30 different NR LBDs have been solved but only one of full length receptors, the heterodimer PPAR/RXR (Chandra et al., 2008).

The Strasbourg node has determined and analyzed the structures of three orphan receptors, the homodimer ERR, RXR and USP associated to heterodimeric partners. The case of RXR (USP in arthropods) is especially interesting since this receptor plays a pivotal role inside the NR superfamily being required as a heterodimer partner for numerous NRs such as RARs, PPARs and VDR in human or EcR, the ecdysone receptor in insects. The molecular evolution of RXR has been investigated through LBD structures of nuclear receptors from two arthropods (Iwema et al., 2007; Iwema et al., 2009) and from that of a cephalochordate amphioxus (Branchiostoma floridae), an invertebrate chordate (Tocchini-Valentini et al., 2009). The crystal structure of this latter revealed an apotetramer (Fig. 4A) with a peculiar conformation of helix H11 filling the binding pocket. In contrast to the arthropods RXR/USPs, which cannot be activated by any RXR ligands, functional data showed that this receptor like the vertebrates/mollusk RXRs, is able to bind and be activated by RXR ligands although less efficiently than vertebrate RXRs. This suggests that amphioxus RXR is an intermediate between arthropods RXR/USPs and vertebrate RXRs.

Strasbourg has also studied the crystal and solution structures of several complexes (USP/EcR, RXR/RAR, RXR/VDR and RXR/PPAR) in different functional states. The crystal structures of LBDs, homo or heterodimers, bound to ligands and coactivator peptides provide high resolution pictures of ligand induced conformational changes. In addition these structures unravel the structural basis for understanding coactivator binding. Although structural studies on the ligand-binding domain (LBD) have established the general mode of nuclear receptor (NR)/coactivator interaction, determinants of binding specificity are only partially understood. A new crystal structure of the ERRα LBD in complex with a PGC-1α box3 peptide (Fig. 4B), explained why the LBD of estrogen receptor-alpha (ERα), interacts only with a region of the (PGC)-1α coactivator, which
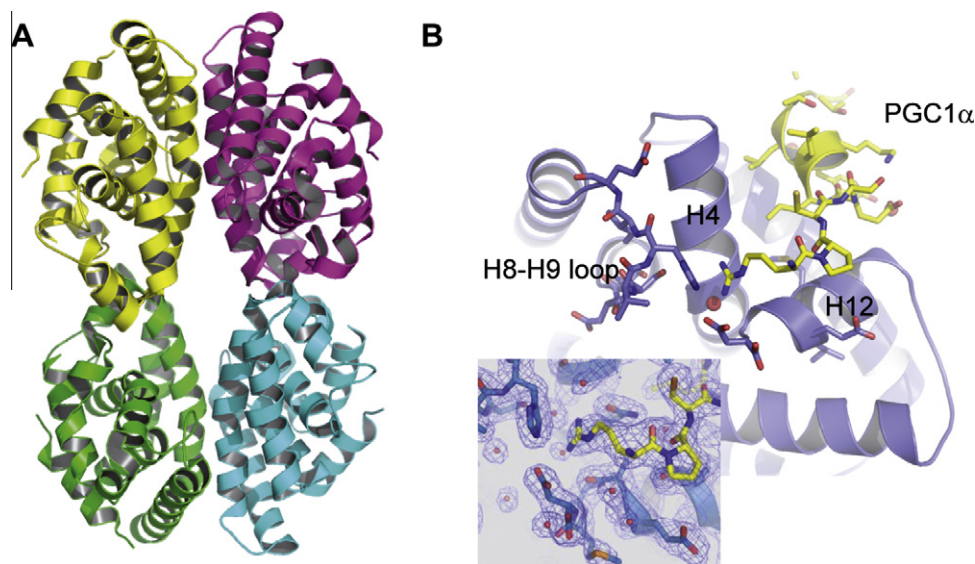
Fig. 4. Nuclear hormone receptor complexes. (A) Quaternary structure of the RXR LBD from an invertebrate chordate. (B) Structure of ERRα LBD in complex with a PGC-1α box3 peptide. Residues amino-terminal of the PGC-1α LXXYL motif contact helix 4 (H4), the loop connecting helices 8 and 9 (H8-H9), and the C terminus of the ERRα LBD. Interaction studies using wild-type and mutant PGC-1α and ERRα showed that these contacts are functionally relevant and are required for efficient ERRα/PGC-1α interaction.

contains the canonical LXXLL motif (NR box2), whereas the LBD of ERRα also binds efficiently an untypical, LXXYL-containing region (NR box3) (Greschik et al., 2008).

To address the communication between nuclear receptors, DNA and components of the basal transcription machinery, data on full length nuclear receptors are required. Strasbourg has worked in this direction and the solution structures of full length receptors in complexes with DNA direct repeat elements and the interacting regions of coactivators such a Med1 or SRC-1 were studied using SAXS, SANS and FRET methods (Rochel et al., 2011). The structures revealed an extended asymmetric shape that is markedly different from that seen in the crystal structure of PPAR/RXR, which is neither new nor extraordinary. These results pointed to the role played by the hinge domains in establishing and maintaining the integrity of the structure and showed two additional important features: the conserved position of the ligand-binding domains at the 5′ ends of the target DNAs and the binding of only one coactivator molecule per heterodimer, to RXR's partner.
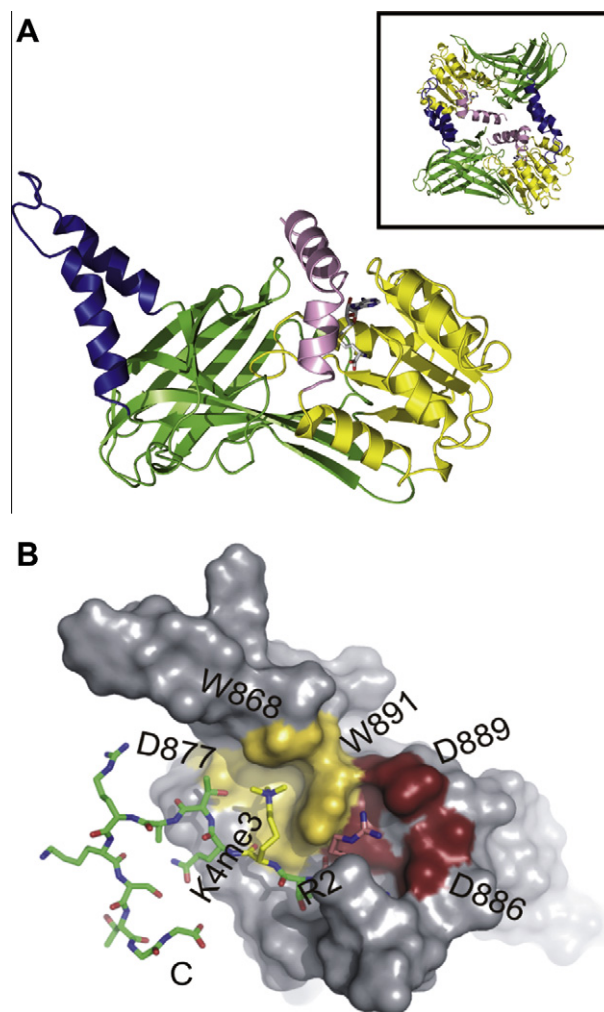
6. Epigenetics

It is well established that next to the presence of transcription factors that control promoter activity, gene expression is critically controlled by the accessibility of the gene. The higher order chromatins structure plays a key regulatory role in this process. Specific modifications in the termini of the histone tails that either lead to more or less compact chromatin structures, in turn modulates the accessibility of transcription factors to promoter and enhancer sequences. These posttranslational modifications are believed to play a key role in epigenetic gene regulation. Both the type of modification and the position within the gene determine the transcriptional outcome of the various modifications, generally referred to as the "histone code" These marks form specific interaction sites for so called reader proteins that in turn through interactions with other proteins promote or inhibit transcription (Kouzarides, 2007) by opening or compacting the chromatin structure of the gene.

6.1. CARM1

Post-translational methylation of arginine is a widespread epigenetic modification found in eukaryotes that is catalyzed by the protein arginine methyltransferases (PRMTs) (Bedford and Clarke, 2009). At least nine members of PRMTs have been identified and classified into two main classes. CARM1 (also known as PRMT4 (Spannhoff et al., 2009) is a crucial protein involved in many biological processes including the regulation of chromatin structure and transcription via methylation of histones and many transcriptional cofactors. As such, understanding the detailed mechanism of action of this protein at the structural level is important and has implications ranging from pure structural information to potential way of regulating gene expression via inhibitor design (Spannhoff et al., 2009). CARM1 contains 608 amino acids in mouse (and human) is built around a catalytic core domain (residues 150–470 in mouse CARM1) that is well conserved in sequence among all PRMTs members. CARM1 possesses two unique additional domains attached, respectively, at the amino-terminal and at the carboxy-terminal end of the PRMT active site. Both additional domains have been shown to be required for the coactivator function of human CARM1. As a first step of a process aimed at understanding at the atomic level the cooperative mechanism by which CARM1 plays its biological functions, we have reported the structure determination and the structural analysis of several crystal structures corresponding to three isolated modules of mouse CARM1: CARM1$_{28–140,}$ CARM1$_{140–480}$ and CARM1$_{28–507}$ (Troffer-Charlier et al., 2007a,b).

The 1.7 Å crystal structure of the amino-terminal domain of CARM1 (CARM1$_{28–140}$) reveals an unexpected PH domain, a scaffold frequently found to regulate protein–protein interactions in a large variety of biological processes. The structure of CARM1$_{140–480}$ has been determined in two different biological states: an apo form and a SAH-CARM1$_{140–480}$ form (both at 2.2 Å resolution) with the SAH molecule bound in the catalytic active site (Fig. 5A). The crystal structures of the CARM1 isolated modules reveal large structural modifications including disorder to order transition, helix to strand transition and active site modifications. The amino-terminal and the carboxy-terminal end of CARM1 catalytic module contain molecular switches that may inspire how CARM1 regulates its biological activities by protein–protein interactions.

Keys to the successful structure determination was to benefit from HTP technologies and as a first step the ability to screen a large numbers of constructions using insect cells infected by recombinant baculovirus (Troffer-Charlier et al., 2007a,b). CARM1 is a bad candidate for structural studies as full length protein behaves in solution as large polydisperse oligomers. From sequences analysis, the first 25 amino acids and the last 120 amino acids are

Fig. 5. Epigenetics. (A) The structure of SAH-CARM1_140–480. Overview of one monomer with the SAH/SAM binding domain in yellow, the amino-terminal helices in pink, the β-barrel in green, the dimerization arm in blue. The bound SAH molecule is shown in a stick model. Ribbon representations of SAH-CARM1_140–480 dimer formed by interactions between the dimerization arm of monomer 1 with the outer surface of the Rossmann fold moiety of monomer 2 (insert). (B) Solution structure of the PHD domain of TAF3 in surface representation, the bound Histone H3 peptide is shown in stick representation, the carboxy-terminus is indicated. The binding pockets of TAF3 for Histone H3 R2 and 3methylated K4 are presented in red and yellow, respectively, with the key residues indicated.

predicted to be highly disordered. Despite extensive efforts, it has not been possible to over-express, obtain in a soluble state, and purify in quantities or concentrations compatible with structural studies any constructs encompassing those disordered regions. Moreover, constructs containing the carboxy-terminal domain of mCARM1 are prone to proteolysis. All those data prompted us to hypothesize that the carboxy-terminal domain of mCARM1 is mainly unfolded in a free state and that a disorder to order transition will take place upon binding to one or several adapted partners. CARM1 is another example of partly natively disordered protein build around a wobbly PH domain linked to a PRMT catalytic platform.

6.2. The PHD domain of TAF3

While dimethylated H3R2 correlates with inactive genes, trimethylation of lysine K4 of histone H3 within the promoter region is generally accompanied with RNA polymerase II transcription. The latter modification is recognized by Chromo, Tudor or PHD domains. The observation that the TFIID factor TAF3 contains a PHD domain argues that TAF3 is contributing to the recruitment of TFIID to promoters, thereby promoting transcription initiation. This is underscored by the observation that selective loss of H3K4 trimethylation leads to loss of binding of TFIID to the promoter region and that the TAF3 PhD domain selectively binds to trimethylated but not to non or mono methylated H3K4 peptides (Vermeulen et al., 2007).

Utrecht has determined the solution structure of the PHD domain of TAF3 in the absence or presence trimethylated H3K4 peptides (van Ingen et al., 2008). A quantitative biochemical characterization of potential Histone H3 peptides that could bind to PHD domain combined with sample condition optimization per-

mitted structural analysis of this complex by NMR. The binding pocket for trimethylated K4 clearly explains the preference for methylated histone tails (Fig. 5B). These results further provide a structural explanation for the observation that H3R2me2 prevents binding of H3K4 trimethylated peptides (Vermeulen et al., 2007). These data underscore the importance of the ability to read the modification signal and through this recognition control gene expression. The presence or absence of these modifications at position R2 and K4 act as a regulatory methyl-methyl switch that can be specifically read by the PHD domain of TAF3.

6.3. Plus3 domain of RTF1

While the structural details on the recognition of post-translationally modified histone proteins is significant, the molecular mechanism underlying the addition or removal of certain modifications is poorly understood. The Set1 protein present in the COMPASS complex is needed for methylation of H3K4. The underlying regulatory mechanism is largely unknown but the PAF complex composed of Paf1, Cdc73, Ctr9, Leo1, and Rtf1, plays an essential role. This complex is thought to interact with elongating RNA polymerase II and is required for cotranscriptional ubiquitination of H2B. Depletion of RTF1 results in loss of H3K4 methylation and transcriptional defects. The Plus3 domain, one of the conserved regions of RTF1 was, using chromatin immuno precipitation, shown to be essential for binding to open reading frames and influencing most of the other RTF1 functions, including transcription (Warner et al., 2007).

The availability of a procedure for effective optimization of expression, solubility and biophysical behavior (Folkers et al., 2004) permitted the optimization of domain boundaries showing that the domain identified by bioinformatics lacked essential
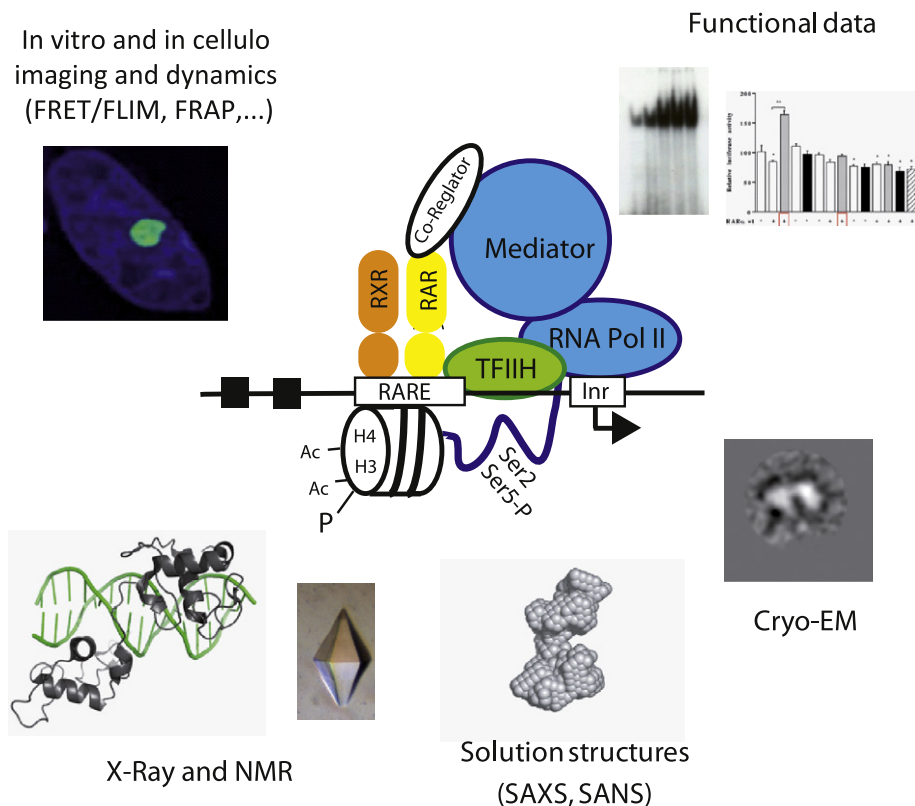

Fig. 6. Towards integrative structural biology studies of transcription complexes.

part of the structured domain. ^{15}N HSQC screening clearly established that these terminal residues were crucial for folding. Utrecht determined the solution structure of the Plus3 domain of RTF1 revealing a novel fold with a beta stranded subdomain structurally resembling Tudor domains and the Dicer/Argonaute PAZ domains (de Jong et al., 2008). Biochemical analysis revealed no evidence for specific interaction with H3 tails either methylated or non-methylated arguing that this domain is not a reader of chromatin modifications. The structural homology with a siRNA domain suggested a potential role for mRNA binding, which would agree with the proposed role for RTF1 in mRNA processing but no RNA interaction could be observed. Importantly using EMSA, NMR binding studies and site directed mutagenesis we identified an ssDNA binding surface on the RTF Plus3 domain. The ability to bind preferentially to ssDNA containing sequences suggests a role for RTF1 in binding to the transcription elongation bubble.

7. Conclusion and future directions: Imaging of transcription, integrated structural biology

Transcription and its regulation depends on the structures of the protein complexes that are its building blocks, and correct cellular function requires the dynamic association of protein complexes with regulatory elements and a myriad of macro- and small molecules. Transcription factors and their complexes can be relatively stable and a wealth of structural data at atomic resolution has been accumulated on a few well characterized complexes, such as *Escherichia coli* or yeast RNA polymerase transcription complexes. We are however still in the early stages of understanding how both general and gene-specific transcription is regulated in eukaryotes, particularly in Human. One reason for this is that the eukaryotic transcription machinery is extremely complex and that many components are multi-subunit assemblies, often poorly characterized. As discussed above, the identification of targets suitable for structural analysis is often challenging and sample preparation often constitutes a major bottleneck. Another difficulty lies in the nature of the complexes, in part because regulation often involves the formation of transient complexes with poor binding constants and in part because their composition is not fixed and can change depending upon the promoter context.

Recent years have seen intensive activities world-wide in functional genomics based around the exploitation of the ever increasing databases of sequence information from genome sequencing projects and the result of structural proteomics initiatives that pioneered high-throughput (HTP) technologies to streamline X-ray and NMR structure determination (Terwilliger et al., 2009). The SPINE2-COMPLEXES program has targeted the development and application of methodologies to address structural studies of multi-protein, protein-nucleic acid and protein–ligand complexes (see the Methods section of this issue). Data summarized above have widely benefited from these technological innovations, resulting in new and/or improved HTP procedures at all stages, from expression screening, large scale production and purification through biophysical and biochemical characterization of individual proteins and complexes, to crystallization, data collection, and solution of structures, as well as solution of smaller macromolecular structures by NMR. Data produced in the frame of the SPINE2-COMPLEXES program not only provided detailed structural information but also paved the ways towards the description of higher order structures using an integrative multi-scale approach that relies on a set of complementary technologies, both in vitro and in situ (Fig. 6). The understanding of transcription regulation that such an endeavor will produce for native and pathogenic systems, is not only an end in itself, but is also a prerequisite for the effective design of new drugs and vaccines impacting the health and quality of life.

References

Abdulrahman, W., Uhring, M., Kolb-Cheynel, I., Garnier, J.M., Moras, D., Rochel, N., Busso, D., Poterszman, A., 2009. A set of baculovirus transfer vectors for screening of affinity tags and parallel expression strategies. Anal. Biochem. 385 (2), 383–385.

Antony, P., Sigueiro, R., Huet, T., Sato, Y., Ramalanjaona, N., Rodrigues, L.C., Mourino, A., Moras, D., Rochel, N., 2010. Structure-function relationships and crystal structures of the vitamin D receptor bound 2 alpha-methyl-(20S, 23S)- and 2 alpha-methyl-(20S, 23R)-epoxymethano-1 alpha, 25-dihydroxyvitamin D3. J. Med. Chem. 53 (3), 1159–1171.

Aricescu, A.R., Assenberg, R., Bill, R.M., Busso, D., Chang, V.T., Davis, S.J., Dubrovsky, A., Gustafsson, L., Hedfalk, K., Heinemann, U., Jones, I.M., Ksiazek, D., Lang, C., Maskos, K., Messerschmidt, A., Macieira, S., Peleg, Y., Perrakis, A., Poterszman, A., Schneider, G., Sixma, T.K., Sussman, J.L., Sutton, G., Tarboureich, N., Zeev-Ben-Mordehai, T., Jones, E.Y., 2006. Eukaryotic expression: developments for structural proteomics. Acta Crystallogr. D. Biol. Crystallogr. 62 (Pt 10), 1114–1124.

Baillat, D., Begue, A., Stehelin, D., Aumercier, M., 2002. ETS-1 transcription factor binds cooperatively to the palindromic head to head ETS-binding sites of the stromelysin-1 promoter by counteracting autoinhibition. J. Biol. Chem. 277 (33), 29386–29398.

Baillat, D., Laitem, C., Leprivier, G., Margerin, C., Aumercier, M., 2009. Ets-1 binds cooperatively to the palindromic Ets-binding sites in the p53 promoter. Biochem. Biophys. Res. Commun. 378 (2), 213–217.

Bedford, M.T., Clarke, S.G., 2009. Protein arginine methylation in mammals: who, what, and why. Mol. Cell. 33 (1), 1–13.

Berrow, N.S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I., Owens, R.J., 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucl. Acids Res. 35 (6), e45.

Betz, J.L., Sasmor, H.M., Buck, F., Insley, M.Y., Caruthers, M.H., 1986. Base substitution mutants of the lac operator: in vivo and in vitro affinities for lac repressor. Gene 50 (1–3), 123–132.

Bieniossek, C., Nie, Y., Frey, D., Olieric, N., Schaffitzel, C., Collinson, I., Romier, C., Berger, P., Richmond, T.J., Steinmetz, M.O., Berger, I., 2009. Automated unrestricted multigene recombineering for multiprotein complex production. Nat. Methods 6 (6), 447–450.

Blanco, A.G., Sola, M., Gomis-Ruth, F.X., Coll, M., 2002. Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. Structure 10 (5), 701–713.

Bonnet, J., Wang, Y.H., Spedale, G., Atkinson, R.A., Romier, C., Hamiche, A., Pijnappel, W.W., Timmers, H.T., Tora, L., Devys, D., Kieffer, B., 2010. The structural plasticity of SCA7 domains defines their differential nucleosome-binding properties. EMBO Rep. 11 (8), 612–618.

Browning, C., Martin, E., Loch, C., Wurtz, J.M., Moras, D., Stote, R.H., Dejaegere, A.P., Billas, I.M., 2007. Critical role of desolvation in the binding of 20-hydroxyecdysone to the ecdysone receptor. J. Biol. Chem. 282 (45), 32924–32934.

Busso, D., Delagoutte-Busso, B., Moras, D., 2005. Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*. Anal. Biochem. 343 (2), 313–321.

Busso, D., Peleg, Y., Heidebrecht, T., Romier, C., Jacobovitch, Y., Dantes, A., Salim, L., Troesch, E., Schuetz, A., Heinemann, U., Folkers, G.E., Geerlof, A., Wilmanns, M., Polewacz, A., Quedenau, C., Bussow, K., Adamson, R., Blagova, E., Walton, J., Cartwright, J.L., Bird, L.E., Owens, R.J., Berrow, N.S., Wilson, K.S., Sussman, J.L., Perrakis, A., Celie, P.H., 2011. Expression of protein complexes using multiple *Escherichia coli* protein co-expression systems: a benchmarking study. J. Struct. Biol.

Chandra, V., Huang, P., Hamuro, Y., Raghuram, S., Wang, Y., Burris, T.P., Rastinejad, F., 2008. Structure of the intact PPAR-gamma-RXR- nuclear receptor complex on DNA. Nature 456 (7220), 350–356.

Ciesielski, F., Rochel, N., Moras, D., 2007. Adaptability of the Vitamin D nuclear receptor to the synthetic ligand Gemini: remodelling the LBP with one side chain rotation. J. Steroid. Biochem. Mol. Biol. 103 (3–5), 235–242.

Coin, F., Proietti De Santis, L., Nardo, T., Zlobinskaya, O., Stefanini, M., Egly, J.M., 2006. P8/TTD-A as a repair-specific TFIIH subunit. Mol. Cell. 21 (2), 215–226.

Darst, S.A., Campbell, E.A., Murakami, K., Korzheva, N., Mustaev, A., Goldfarb, A., 2001. Structural studies of prokaryotic RNA polymerases. FASEB J. 15 (5), A1082.

de Jong, R.N., Daniels, M.A., Kaptein, R., Folkers, G.E., 2006. Enzyme free cloning for high throughput gene cloning and expression. J. Struct. Funct. Genomics 7 (3–4), 109–118.

de Jong, R.N., Truffault, V., Diercks, T., Ab, E., Daniels, M.A., Kaptein, R., Folkers, G.E., 2008. Structure and DNA binding of the human Rtf1 Plus3 domain. Structure 16 (1), 149–159.

Diebold, M.L., Fribourg, S., Koch, M., Metzger, T., Romier, C., 2011. Deciphering correct strategies for multiprotein complex assembly by co-expression: Application to complexes as large as the histone octamer. J. Struct. Biol.

Diebold, M.L., Koch, M., Loeliger, E., Cura, V., Winston, F., Cavarelli, J., Romier, C., 2010a. The structure of an Iws1/Spt6 complex reveals an interaction domain conserved in TFIIS, Elongin A and Med26. EMBO J. 29 (23), 3979–3991.

Diebold, M.L., Loeliger, E., Koch, M., Winston, F., Cavarelli, J., Romier, C., 2010b. Noncanonical Tandem SH2 Enables Interaction of Elongation Factor Spt6 with RNA Polymerase II. J. Biol. Chem. 285 (49), 38389–38398.

Egly, J.M., 2001. The 14th Datta Lecture. TFIIH: from transcription to clinic. FEBS Lett. 498 (2–3), 124–128.

Fogg, M.J., Wilkinson, A.J., 2008. Higher-throughtput approaches to crystallization and crystal structure determination. Biochem. Soc. Trans. 36, 771–775.

Folkers, G.E., van Buuren, B.N., Kaptein, R., 2004. Expression screening, protein purification and NMR analysis of human protein domains for structural genomics. J. Struct. Funct. Genomics 5 (1–2), 119–131.

Garvie, C.W., Hagman, J., Wolberger, C., 2001. Structural studies of Ets-1/Pax5 complex formation on DNA. Mol. Cell. 8 (6), 1267–1276.

Garvie, C.W., Pufall, M.A., Graves, B.J., Wolberger, C., 2002. Structural analysis of the autoinhibition of Ets-1 and its role in protein partnerships. J. Biol. Chem. 277 (47), 45529–45536.

Giglia-Mari, G., Coin, F., Ranish, J.A., Hoogstraten, D., Theil, A., Wijgers, N., Jaspers, N.G., Raams, A., Argentini, M., van der Spek, P.J., Botta, E., Stefanini, M., Egly, J.M., Aebersold, R., Hoeijmakers, J.H., Vermeulen, W., 2004. A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. Nat. Genet. 36 (7), 714–719.

Greschik, H., Althage, M., Flaig, R., Sato, Y., Chavant, V., Peluso-Iltis, C., Choulier, L., Cronet, P., Rochel, N., Schule, R., Stromstedt, P.E., Moras, D., 2008. Communication between the ERRalpha homodimer interface and the PGC-1alpha binding surface via the helix 8–9 loop. J. Biol. Chem. 283 (29), 20220–20230.

Iwema, T., Billas, I.M., Beck, Y., Bonneton, F., Nierengarten, H., Chaumot, A., Richards, G., Laudet, V., Moras, D., 2007. Structural and functional characterization of a novel type of ligand-independent RXR–USP receptor. EMBO J. 26 (16), 3770–3782.

Iwema, T., Chaumot, A., Studer, R.A., Robinson-Rechavi, M., Billas, I.M., Moras, D., Laudet, V., Bonneton, F., 2009. Structural and evolutionary innovation of the heterodimerization interface between USP and the ecdysone receptor ECR in insects. Mol. Biol. Evol. 26 (4), 753–768.

Jarvis, D.L., 2009. Baculovirus-insect cell expression systems. Methods Enzymol. 463, 191–222.

Kainov, D.E., Cura, V., Vitorino, M., Nierengarten, H., Poussin, P., Kieffer, B., Cavarelli, J., Poterszman, A., 2010. Structure determination of the minimal complex between Tfb5 and Tfb2, two subunits of the yeast transcription/DNA-repair factor TFIIH: a retrospective study. Acta Crystallogr. D. Biol. Crystallogr. 66 (Pt. 7), 745–755.

Kainov, D.E., Vitorino, M., Cavarelli, J., Poterszman, A., Egly, J.M., 2008. Structural basis for group A trichothiodystrophy. Nat. Struct. Mol. Biol. 15 (9), 980–984.

Kalodimos, C.G., Biris, N., Bonvin, A.M., Levandoski, M.M., Guennuegues, M., Boelens, R., Kaptein, R., 2004. Structure and flexibility adaptation in nonspecific and specific protein–DNA complexes. Science 305 (5682), 386–389.

Kalodimos, C.G., Bonvin, A.M., Salinas, R.K., Wechselberger, R., Boelens, R., Kaptein, R., 2002. Plasticity in protein–DNA recognition: lac repressor interacts with its natural operator 01 through alternative conformations of its DNA-binding domain. EMBO J. 21 (12), 2866–2876.

Kodandapani, R., Pio, F., Ni, C.Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R.A., Ely, K.R., 1996. A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. Nature 380 (6573), 456–460.

Kornberg, R.D., 2007. The molecular basis of eukaryotic transcription. Proc. Natl. Acad. Sci. USA 104 (32), 12955–12961.

Kouzarides, T., 2007. SnapShot: Histone-modifying enzymes. Cell 128 (4), 802.

Kramer, H., Niemoller, M., Amouyal, M., Revet, B., von Wilcken-Bergmann, B., Muller-Hill, B., 1987. Lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. EMBO J. 6 (5), 1481–1491.

Kriz, A., Schmid, K., Baumgartner, N., Ziegler, U., Berger, I., Ballmer-Hofer, K., Berger, P., 2010. A plasmid-based multigene expression system for mammalian cells. Nat. Commun. 1 (8), 120.

Kuznedelov, K., Minakhin, L., Niedziela-Majka, A., Dove, S.L., Rogulja, D., Nickels, B.E., Hochschild, A., Heyduk, T., Severinov, K., 2002. A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition. Science 295 (5556), 855–857.

Lamber, E.P., Vanhille, L., Textor, L.C., Kachalova, G.S., Sieweke, M.H., Wilmanns, M., 2008. Regulation of the transcription factor Ets-1 by DNA-mediated homo-dimerization. EMBO J. 27 (14), 2006–2017.

Liu, X., Bushnell, D.A., Wang, D., Calero, G., Kornberg, R.D., 2010. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. Science 327 (5962), 206–209.

Luna-Vargas, M.P., Christodoulou, E., Alfieri, A., van Dijk, W.J., Stadnik, M., Hibbert, R.G., Sahtoe, D.D., Clerici, M., Marco, V.D., Littler, D., Celie, P.H., Sixma, T.K., Perrakis, A., 2011. Enabling high-throughput ligation-independent cloning and protein expression for the family of ubiquitin specific proteases. J Struct Biol.

Makino, K., Amemura, M., Kim, S.K., Nakata, A., Shinagawa, H., 1993. Role of the sigma 70 subunit of RNA polymerase in transcriptional activation by activator protein PhoB in Escherichia coli. Genes Dev. 7 (1), 149–160.

McEwan, I., 2009. Nuclear receptors: one big family. Methods Mol. Biol. 505, 3–18.

Mousson, F., Kolkman, A., Pijnappel, W.W., Timmers, H.T., Heck, A.J., 2008. Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes. Mol. Cell Proteomics 7 (5), 845–852.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., Darst, S.A., 2002. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science 296 (5571), 1285–1290.

Mydlikova, Z., Gursky, J., Pirsel, M., 2010. Transcription factor IIH – the protein complex with multiple functions. Neoplasma 57 (4), 287–290.

Nettleship, J.E., Assenberg, R., Diprose, J.M., Rahman-Huq, N., Owens, R.J., 2010. Recent advances in the production of proteins in insect and mammalian cells for structural biology. J. Struct. Biol. 172 (1), 55–65.

Nie, Y., Bieniossek, C., Frey, D., Olieric, N., Schaffitzel, C., Steinmetz, M.O., Berger, I., 2009. ACEMBLing multigene expression vectors by recombineering. Nature Protocols.

Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B., Muller-Hill, B., 1994. Quality and position of the three lac operators of E. coli define efficiency of repression. EMBO J. 13 (14), 3348–3355.

Oehler, S., Eismann, E.R., Kramer, H., Muller-Hill, B., 1990. The three operators of the lac operon cooperate in repression. EMBO J. 9 (4), 973–979.

Papai, G., Tripathi, M.K., Ruhlmann, C., Layer, J.H., Weil, P.A., Schultz, P., 2010. TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. Nature 465 (7300), 956–960.

Papai, G., Tripathi, M.K., Ruhlmann, C., Werten, S., Crucifix, C., Weil, P.A., Schultz, P., 2009. Mapping the initiator binding Taf2 subunit in the structure of hydrated yeast TFIID. Structure 17 (3), 363–373.

Pijnappel, W.P., Kolkman, A., Baltissen, M.P., Heck, A.J., Timmers, H.M., 2009. Quantitative mass spectrometry of TATA binding protein-containing complexes and subunit phosphorylations during the cell cycle. Proteome Sci. 7, 46.

Pradeau-Aubreton, K., Ruff, M., Garnier, J.M., Schultz, P., Drillien, R., 2010. Vectors for recombinational cloning and gene expression in mammalian cells using modified vaccinia virus Ankara. Anal. Biochem. 404 (1), 103–105.

Pufall, M.A., Lee, G.M., Nelson, M.L., Kang, H.S., Velyvis, A., Kay, L.E., McIntosh, L.P., Graves, B.J., 2005. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. Science 309 (5731), 142–145.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., Séraphin, B., 1999. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol. 17 (10), 1030–1032.

Rochel, N., Ciesielski, F., Godet, J., Moman, E., Roessle, M., Peluso-Iltis, C., Moulin, M., Haertlein, M., Callow, P., Mély, Y., Svergun, D., Moras, D., 2011. Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. Nat. Struct. Mol. Biol. 18 (5), 564–570.

Romanuka, J., Folkers, G.E., Biris, N., Tishchenko, E., Wienk, H., Bonvin, A.M., Kaptein, R., Boelens, R., 2009. Specificity and affinity of Lac repressor for the auxiliary operators O2 and O3 are explained by the structures of their protein–DNA complexes. J. Mol. Biol. 390 (3), 478–489.

Romier, C., James, N., Birck, C., Cavarelli, J., Vivarès, C., Collart, M.A., Moras, D., 2007. Crystal structure, biochemical and genetic characterization of yeast and E. cuniculi TAF(II)5 N-terminal domain: implications for TFIID assembly. J Mol Biol. 368 (5), 1292–1306.

Sasmor, H.M., Betz, J.L., 1990. Symmetric lac operator derivatives: effects of half-operator sequence and spacing on repressor affinity. Gene 89 (1), 1–6.

Sato, Y., Ramalanjaona, N., Huet, T., Potier, N., Osz, J., Antony, P., Peluso-Iltis, C., Poussin-Courmontagne, P., Ennifar, E., Mély, Y., Dejaegere, A., Moras, D., Rochel, N., 2010. The "Phantom Effect" of the Rexinoid LG100754: structural and functional insights. PLoS One 5 (11), e15119.

Scheich, C., Kummel, D., Soumailakakis, D., Heinemann, U., Bussow, K., 2007. Vectors for co-expression of an unrestricted number of proteins. Nucl. Acids Res. 35 (6), e43.

Sharrocks, A.D., 2001. The ETS-domain transcription factor family. Nat. Rev. Mol. Cell Biol. 2 (11), 827–837.

Spannhoff, A., Hauser, A.T., Heinke, R., Sippl, W., Jung, M., 2009. The emerging therapeutic potential of histone methyltransferase and demethylase inhibitors. ChemMedChem 4 (10), 1568–1582.

Spronk, C.A., Bonvin, A.M., Radha, P.K., Melacini, G., Boelens, R., Kaptein, R., 1999. The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator. Structure 7 (12), 1483–1492.

Terwilliger, T.C., Stuart, D., Yokoyama, S., 2009. Lessons from structural genomics. Annu. Rev. Biophys. 38, 371–383.

Tocchini-Valentini, G.D., Rochel, N., Escriva, H., Germain, P., Peluso-Iltis, C., Paris, M., Sanglier-Cianferani, S., Van Dorsselaer, A., Moras, D., Laudet, V., 2009. Structural and functional insights into the ligand-binding domain of a nonduplicated retinoid X nuclear receptor from the invertebrate chordate amphioxus. J. Biol. Chem. 284 (3), 1938–1948.

Troffer-Charlier, N., Cura, V., Hassenboehler, P., Moras, D., Cavarelli, J., 2007a. Expression, purification, crystallization and preliminary crystallographic study of isolated modules of the mouse coactivator-associated arginine methyltransferase 1. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 63 (Pt. 4), 330–333.

Troffer-Charlier, N., Cura, V., Hassenboehler, P., Moras, D., Cavarelli, J., 2007b. Functional insights from structures of coactivator-associated arginine methyltransferase 1 domains. EMBO J. 26 (20), 4391–4401.

Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F., Berger, I., 2010. New baculovirus expression tools for recombinant protein complex production. J. Struct. Biol. 172 (1), 45–54.

Unger, T., Jacobovitch, Y., Dantes, A., Bernheim, R., Peleg, Y., 2010. Applications of the restriction free (RF) cloning procedure for molecular manipulations and protein expression. J. Struct. Biol. 172 (1), 34–44.

van Ingen, H., van Schaik, F.M., Wienk, H., Ballering, J., Rehmann, H., Dechesne, A.C., Kruijzer, J.A., Liskamp, R.M., Timmers, H.T., Boelens, R., 2008. Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. Structure 16 (8), 1245–1256.

Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S., Yokoyama, S., 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. Nature 417 (6890), 712–719.

Venanzoni, M.C., Robinson, L.R., Hodge, D.R., Kola, I., Seth, A., 1996. ETS1 and ETS2 in p53 regulation: spatial separation of ETS binding sites (EBS) modulate protein: DNA interaction. Oncogene 12 (6), 1199–1204.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., Timmers, H.T., 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell 131 (1), 58–69.

Vijayachandran, L.S., Viola, C., Garzoni, F., Trowitzsch, S., Bieniossek, C., Chaillet, M., Schaffitzel, C., Busso, D., Romier, C., Poterszman, A., Richmond, T.J., Berger, I., 2011. Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells. J. Struct. Biol.

Vitorino, M., Coin, F., Zlobinskaya, O., Atkinson, R.A., Moras, D., Egly, J.M., Poterszman, A., Kieffer, B., 2007. Solution structure and self-association properties of the p8 TFIIH subunit responsible for trichothiodystrophy. J. Mol. Biol. 368 (2), 473–480.

Warner, M.H., Roinick, K.L., Arndt, K.M., 2007. Rtf1 is a multifunctional component of the Paf1 complex that regulates gene expression by directing cotranscriptional histone modification. Mol. Cell Biol. 27 (17), 6103–6115.

Wilson, C.J., Zhan, H., Swint-Kruse, L., Matthews, K.S., 2007. The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. Cell Mol. Life Sci. 64 (1), 3–16.