OBC
www.rsc.org/obc

# Chemical approaches to studying transcription

**Ross V. Weatherman***
*Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, IN, USA*

***Received 14th July 2003, Accepted 19th August 2003***
***First published as an Advance Article on the web 27th August 2003***

Gene transcription is one of the most important and complex processes in biology but great advances are being made into understanding its molecular mechanisms. Selective modulators of nuclear receptors that can regulate transcription of specific genes allow for the comparative analysis of different states of transcription. Techniques to monitor the binding of proteins to DNA leading up to transcription have also increased our knowledge of the events involved in the initiation of transcription. While still in its infancy, the use of chemical tools to study transcription shows great promise in dissecting a complex molecular process.

## Introduction

My fascination with the chemistry of transcription began back in the mid-nineties. I would soon be getting my doctorate in organic chemistry working in the newly coined field of chemical biology with our lab and others synthesizing molecules to study all sorts of complex cellular processes. I knew that I wanted to continue bringing chemical approaches to the world of biology for my postdoctoral work. I decided to venture into one of the most complex cellular processes—gene transcription. But much like a visitor in a strange land, I quickly found myself immersed in a world with almost *too much* complexity and a different language spoken almost entirely in new acronyms and abbreviations. Nevertheless, I gradually began to understand this language and embrace the complexity as a never-ending source of interesting questions and important research. I also have grown to recognize some of the areas related to gene transcription where chemical approaches have been or could be extremely useful. It is the purpose of this article to share some of these areas with the reader, albeit not in an all-encompassing review format, by highlighting some of outstanding work performed by researchers applying chemical approaches to the general area of gene transcription. This article is also meant to complement an outstanding article in this series that has appeared previously[1] by presenting another philosophy on how to use chemical methods to begin to untangle this complex web.

In a general sense, gene transcription can be envisioned as a template-driven synthesis accomplished by a multi-component machine perfectly designed to transcribe genes into messenger RNA. The molecular mechanisms underlying the activity of RNA polymerase and the other associated proteins that make up this transcriptional machine have long been an area of intense study by chemists, biochemists and geneticists and are a major component of the foundation of modern molecular biology.[2] This article will focus on the use of chemical methods to study the processes involved in regulating the recruitment of the transcriptional machinery to specific genes because, ultimately, we would like to be able to selectively modulate a particular gene or subset of genes while leaving all other transcription unperturbed. To begin to understand how to accomplish such an approach, one needs to understand how the transcriptional machinery is recruited to a specific gene.

*Ross Weatherman received an A.B. degree in chemistry from Wabash College in 1991 before moving on to the University of Wisconsin-Madison to obtain a Ph.D. degree in organic chemistry with Professor Laura Kiessling. He then moved to University of California, San Francisco to complete postdoctoral work with Professor Tom Scanlan. In the fall of 2001, he joined the faculty at Purdue University where his research group is focused on developing chemical approaches to control specific steroid hormone signaling pathways.*

**Ross Weatherman**

## It all starts at the promoter

Upstream of every gene lies a region of DNA known as the promoter. It is here where the molecular events occur to recruit RNA polymerase and accessory proteins to bind to the DNA and begin transcribing the gene. The number and nature of proteins involved in this recruitment varies greatly depending on the DNA sequence of the promoter and the cell type, thus the promoter becomes a focal point of regulation by numerous signal transduction pathways in the cell. Despite this potentially complex situation at the promoter, chemical methods have been used in primarily two different types of approaches to begin to unravel the mechanisms of transcriptional regulation. The first consists of building the regulatory complex from the ground up—making molecules that can bind to a specific DNA sequence in the promoter region and recruit other proteins to begin the process of transcription. There has been interesting progress in this area and a previous article in this series has covered it.[1] The other approach is to analyze the promoter complex from the top down—finding molecules that alter the transcription of a particular gene and then analyzing the differences in the regulatory complex caused by that molecule. I will highlight some interesting applications of this approach and discuss potential areas where new chemical approaches are needed.

DOI: 10.1039/b307987b

This journal is © The Royal Society of Chemistry 2003

## Nuclear hormone receptors: an ideal system

The nuclear hormone receptor superfamily[3] is an excellent system for studying the mechanisms of transcriptional regulation. The ligands that bind to the members of this superfamily are well known to any organic chemist—compounds such as estradiol, testosterone, and retinoic acid have long been used as proving grounds for the development of synthetic methodology and as the basis of a large number of analogs with potential therapeutic applications. But in addition to their therapeutic importance, the nuclear hormone receptor superfamily is essentially a group of small molecule-dependent initiators of transcription and thus provides an excellent tool to dissect the molecular mechanisms by which gene transcription is regulated. Nuclear receptors can either activate or repress transcription upon ligand binding depending on the structure of the ligand, the promoter region and the type of cell. This has been referred to as a "tripartite pharmacology" and allows for comparison of the different transcriptional states to help elucidate the molecular mechanisms underpinning the start of transcription. While there is still much to be learned, the general picture that has emerged is that the nuclear receptor acts as a foundation for the building of a large multi-protein coregulatory complex that ultimately alters the structure of the genomic DNA to either activate or block transcription of the downstream gene.[4] Elucidating the molecular mechanisms of how these complexes arise has now become a major area of new research.

Nuclear receptors generally are organized into three domains—a ligand binding domain, a DNA binding domain and a poorly defined N-terminal domain. As the names suggest, the DNA binding domain binds to DNA sequences in the promoter region and the ligand binding domain binds to the small molecule.[5] After ligand and DNA binding, other proteins in the complex then begin to bind to the nuclear receptor at the promoter. Depending on whether the transcription is activated or repressed, different combinations of coregulatory proteins are recruited. Some combinations lead to activation, others lead to repression. Ultimately, it is the structure of the small molecule at the core of the ligand binding domain that determines which coregulatory proteins are invited to join the complex.

Structural studies combined with biochemical studies have greatly increased our knowledge of the first step in the recruitment of the coregulatory protein complex. In the case of the estrogen receptor, the endogenous hormone estradiol activates the transcription of a certain subset of genes while the anticancer agent tamoxifen blocks transcription.[6] A comparison of structures determined by X-ray crystallography of the ligand binding domain of estrogen receptor bound to either estradiol or tamoxifen shows a major perturbation in one alpha helix of the domain (Fig. 1). It was then shown that the cleft created by that helix upon estrogen binding is recognized by a coactivating protein that allows further buildup of a complex that activates transcription at a particular promoter. This cleft is obstructed when tamoxifen binds to the ligand domain which leads to an interaction with a corepressor protein that promotes the buildup of a complex that represses transcription.

This mode of coactivator and corepressor recruitment appears to be very general across the nuclear receptor superfamily and has led to a new paradigm in the design of antagonists of nuclear receptor action by synthesizing peptidomimetic agents to block the activation cleft.[7] There are many questions regarding the interaction between certain coactivators and nuclear receptors. The current models are complicated by the ability of some compounds like tamoxifen to activate transcription in some tissues while repressing transcription in others. Chemical approaches to study this critical issue have revealed that nuclear receptors are involved in cross-talk with other transcription factors and cell signaling pathways.[8] The discovery of novel chemical entities with novel regulatory properties will be vital for dissecting this important problem.

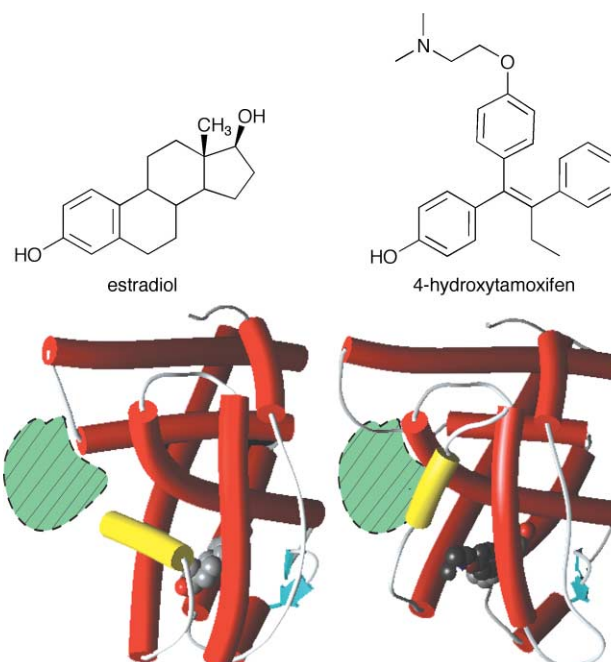


**Fig. 1** A schematic representation of the structure of the estrogen receptor alpha ligand binding domain bound to either estradiol or 4-hydroxytamoxifen. Alpha helices are represented as cylinders and the key helix 12 is highlighted in yellow. The green hatched area indicates the coactivator-binding cleft formed upon estradiol binding which is blocked by helix 12 upon 4-hydroxytamoxifen binding.

## Following complex formation

After the binding of the ligand-bound nuclear receptor to the promoter, a number of other proteins then associate with the nuclear receptor. Traditional antibody-based methods have detected sometimes as many as 20–30 proteins, many uncharacterized, in this complex that recruits the polymerase to a particular gene. While these techniques have provided a wealth of information as to the identity of the proteins, they do not provide much information into specific interactions between the different proteins nor does it provide much information on the interactions at a specific promoter. Two approaches, one based on chemical crosslinking, the other based on fluorescence, are beginning to address these issues.

One technique that has been rapidly advancing our knowledge of specific molecular interactions at promoters is chromatin immunoprecipitation, also known as chIP (Fig. 2).[9] In this approach, cells are treated with formaldehyde to form chemical crosslinks between amine groups of interacting proteins or between proteins and the exocyclic amines of adenosine and cytosine in the promoter DNA sequence. The chromosomal DNA is then sheared by sonication and purified along with any covalently attached protein complexes—the DNA complexes are isolated using antibodies against proteins of interest. The crosslinks are then reversed and the proteins and DNA from the complex can be analyzed using standard methods.

This method has been used to determine such things as the identity of transcription factors at a given promoter or the genome-wide distribution of binding sites of a particular transcription factor.[10] One particularly interesting application is the study of the dynamics of transcriptional complex formation. chIP analysis of an estrogen regulated promoter revealed the order and timing of various proteins assembling on the promoter region.[11]

Another approach to study interactions in the transcription complex uses fluorescence to look at specific proteins. The

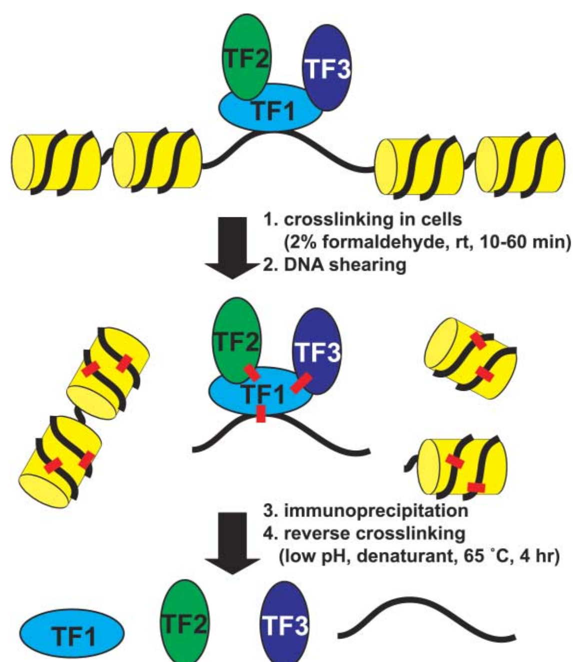


**Fig. 2** A schematic representation of the chromatin immunoprecipitation (chIP) assay. Chromosomal DNA that is bound to both histones (in yellow) and transcription factors (TF1-3) are chemically crosslinked using formaldehyde. The DNA is then sheared by sonication and DNA fragments covalently attached to proteins are isolated (crosslinks denoted in red). Antibodies (against TF1 in this case) are used to immunopurify TF1 containing fragments and then the crosslinks are reversed and the molecules associated with TF1 can be analyzed.

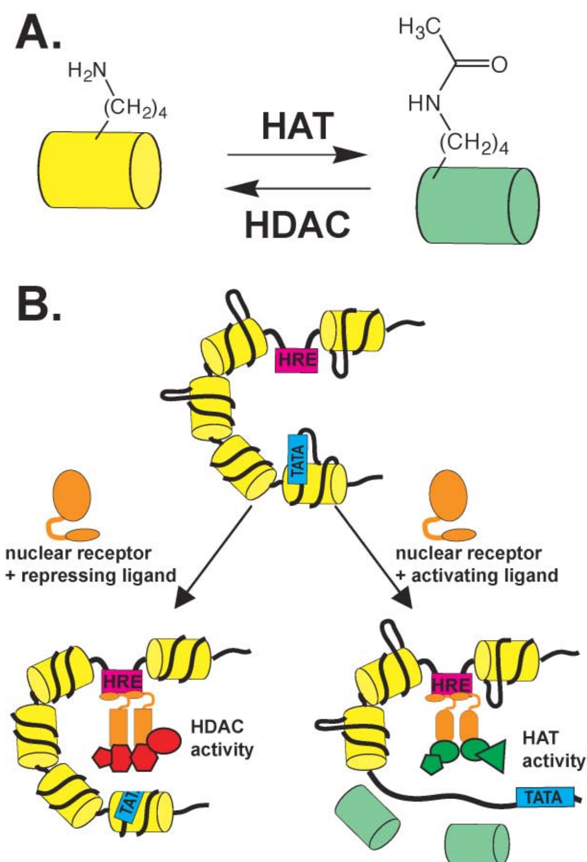


**Fig. 3** **A.** Effect of histone acteyltransferases (HAT) and histone deacetylases (HDAC) on the lysine side chains of histones. Histone lysine and arginine side chains can also be methylated by methyltransferase enzymes. **B.** Effect of nuclear receptors on chromatin structure. Initially, DNA is loosely associated around histones (shown in yellow). Binding of a nuclear receptor to its hormone response element (HRE) in the promoter along with a ligand that represses transcription causes the recruitment of corepressor proteins (shown in red) that possess histone deacetylase (HDAC) domains that cause tighter histone–DNA interactions which decrease access to the initiation site for transcription (shown as a blue TATA box). Binding of an activating ligand causes the nuclear receptor to recruit coactivator proteins (shown in green) with histone acetyltransferase (HAT) activity that relax DNA–histone interactions, freeing the TATA box for binding by the RNA polymerase complex.

advantage of fluorescence-based approaches is that the measurements can be made in real-time in living cells. Typically, the protein of interest is labeled with a fluorophore either through chemical labeling or genetically using a fluorescent protein fusion. Two fluorophores can be used to demonstrate an interaction in the transcriptional complex using fluorescence resonance energy transfer (FRET) as has been used to demonstrate an interaction inside living cells between the nuclear receptor PPAR$\gamma$ and the coactivator protein SRC-1.[12] Fluorescence can also be used to measure the time course of transcription. The exchange rate at the promoter of fluorophore-labeled transcription factors can be measured using photobleaching techniques. One such case revealed that glucocorticoid receptor exchanged rapidly with its promoter element while RNA polymerase II exchanged much more slowly.[13] This suggests that the initial interactions leading to transcription are rapidly changing while the transcriptional complex itself is much more difficult to disrupt.

While the chIP assay and fluorescence techniques have been quite successful, there is much room for improvement. The efficiency of crosslinking by formaldehyde in the chIP assays is low and purification of the complexes requires antibodies. Fluorescent techniques still rely mainly on the attachment of large and potentially interfering fluorescent proteins. More efficient crosslinking methods and more specific affinity and fluorophore labeling techniques would greatly improve these approaches.

## Studying the role of chromatin

Ultimately, all of these complexes that are formed at the promoter have to lead to the binding of RNA polymerase and the beginning of transcription. The mechanism by which this occurs appears to involve chemical modifications of the chromatin by the regulatory complex.[14] In chromatin, the DNA is wrapped up tightly around proteins known as histones by electrostatic interactions between the negatively charged DNA and the positively charged lysines of the histones (Fig. 3). Many of the transcription factors involved in activating complexes possess histone acetyltransferase domains (HAT) or histone methyltransferases (HMT) that acetylate or alkylate the side chain amines of lysine and arginine and eliminate the electrostatic interaction, leaving the DNA in a more accessible state. This allows the initiation complex of RNA polymerase to bind to the promoter and start transcription. Alternatively, transcription factors that repress transcription possess histone deacetylase (HDAC) domains that hydrolyze acetylated histone side chains and make the DNA less accessible.

Much of the information that has been learned about histone modification and biological processes has come from the use of small molecule inhibitors of the modification enzymes. The natural products trichostatin and trapoxin are inhibitors of HDAC enzymes and a trapoxin affinity column was used to isolate the first mammalian HDAC.[15] With a number of different subtypes of HDAC and HAT enzymes as well as a number of other enzymes capable of histone modification such as histone kinases, phosphatases, and methyltransferases the challenge is now to find new selective inhibitors of these enzymes to deduce their specific role in chromatin remodeling. Some screening strategies have been developed to try to discover selective compounds.[16]

## Conclusions

After years of residing firmly in the world of the molecular biologists and geneticists, the complex processes of transcription are now beginning a transition into an area of research that is ripe for discovery and analysis by chemists. Whether it's the dissection of complex processes with new compounds or the discovery of new analytical techniques to observe transcription, the field needs people with a molecular mindset and the willingness to venture into this complex world and gain insight into its inner workings using new exciting approaches.

## Acknowledgements

I am grateful to the American Cancer Society (IRG-58-006-41A), the Purdue Cancer Center and Purdue University for financial support.

## References

1 A. K. Mapp, *Org. Biomol. Chem.*, 2003, **1**, 2217–2220.
2 P. Cramer, *Curr. Opin. Struct. Biol.*, 2002, **12**, 89–97.
3 D. J. Mangelsdorf, C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon and R. Evans, *Cell*, 1995, **83**, 835–839.
4 N. J. McKenna and B. W. O'Malley, *Cell*, 2002, **108**, 465–474.
5 R. V. Weatherman, R. J. Fletterick and T. S. Scanlan, *Annu. Rev. Biochem.*, 1999, **68**, 559–581.
6 A. K. Shiau, D. Barstad, P. M. Loria, L. Cheng, P. J. Kushner, D. A. Agard and G. L. Greene, *Cell*, 1998, **95**, 927–937.
7 T. R. Geistlinger and R. K. Guy, *J. Am. Chem. Soc.*, 2003, **125**, 6852–6853.
8 R. V. Weatherman, D. C. Carroll and T. S. Scanlan, *Bioorg. Med. Chem. Lett.*, 2001, **11**, 3129–3131; R. V. Weatherman, N. J. Clegg and T. S. Scanlan, *Chem. Biol.*, 2001, **8**, 427–436.
9 V. Orlando, *Trends Biochem. Sci.*, 2000, **25**, 99–104.
10 B. Ren, F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C. J. Wilson, S. P. Bell and R. A. Young, *Science*, 2000, **290**, 2306–2309; J. D. Lieb, X. Liu, D. Botstein and P. O. Brown, *Nat. Genet.*, 2001, **28**, 327–334.
11 Y. Shang, X. Hu, J. DiRenzo, M. A. Lazar and M. Brown, *Cell*, 2000, **103**, 843–852.
12 J. Llopis, S. Westin, M. Ricote, Z. Wang, C. Y. Cho, R. Kurokawa, T. M. Mullen, D. W. Rose, M. G. Rosenfeld, R. Y. Tsien, C. K. Glass and J. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 4363–4368.
13 M. Becker, C. Baumann, S. John, D. A. Walker, M. Vigneron, J. G. McNally and G. L. Hager, *EMBO Rep.*, 2002, **3**, 1188–1194; J. G. McNally, W. G. Muller, D. Walker, R. Wolford and G. L. Hager, *Science*, 2000, **287**, 1262–1265.
14 B. E. Bernstein and S. L. Schreiber, *Chem. Biol.*, 2002, **9**, 1167–1173.
15 J. Taunton, C. A. Hassig and S. L. Schreiber, *Science*, 1996, **272**, 408–411.
16 S. J. Haggarty, K. M. Koeller, J. C. Wong, R. A. Butcher and S. L. Schreiber, *Chem. Biol.*, 2003, **10**, 383–396.