RSC Biomolecular Sciences

Edited by Henri Buc and Terence Strick

RNA Polymerases as Molecular Motors

RSC Publishing

RNA Polymerases as Molecular Motors

RSC Biomolecular Sciences

Editorial Board:

This Series is devoted to coverage of the interface between the chemical and biological sciences, especially structural biology, chemical biology, bio- and chemo-informatics, drug discovery and development, chemical enzymology and biophysical chemistry. Ideal as reference and state-of-the-art guides at the graduate and post-graduate level.

Titles in the Series:

Biophysical and Structural Aspects of Bioenergetics
Edited by Mårten Wikström
Computational and Structural Approaches to Drug Discovery: Ligand-Protein Interactions
Edited by Robert M Stroud and Janet Finer-Moore
Exploiting Chemical Diversity for Drug Discovery
Edited by Paul A. Bartlett and Michael Entzeroth
Metabolomics, Metabonomics and Metabolite Profiling
Edited by William J. Griffiths
Nucleic Acid-Metal Ion Interactions
Edited by Nicholas V. Hud
Oxidative Folding of Peptides and Proteins
Edited by Johannes Buchner and Luis Moroder
Protein-Carbohydrate Interactions in Infectious Disease
Edited by Carole A. Bewley
Protein Folding, Misfolding and Aggregation: Classical Themes and Novel Approaches
Edited by Victor Muñoz
Protein-Nucleic Acid Interactions: Structural Biology
Edited by Phoebe A. Rice and Carl C. Correll
Quadruplex Nucleic Acids
Edited by Stephen Neidle and Shankar Balasubramanian
Quantum Tunnelling in Enzyme-Catalysed Reactions
Edited by Nigel S. Scrutton and Rudolf K. Allemann
Ribozymes and RNA Catalysis
Edited by David MJ Lilley FRS and Fritz Eckstein
RNA Polymerases as Molecular Motors
Edited by Henri Buc and Terence Strick
Sequence-specific DNA Binding Agents
Edited by Michael Waring
Structural Biology of Membrane Proteins
Edited by Reinhard Grisshammer and Susan K. Buchanan
Structure-based Drug Discovery: An Overview
Edited by Roderick E. Hubbard
Therapeutic Oligonucleotides
Edited by Jens Kurreck

Visit our website on www.rsc.org/biomolecularsciences

For further information please contact:
Sales and Customer Care, Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge, CB4 0WF, UK
Telephone: +44 (0)1223 432360, Fax: +44 (0)1223 426017, Email: sales@rsc.org

RNA Polymerases as Molecular Motors

Edited by

Henri Buc
CIS, Institut Pasteur, Paris, France

Terence Strick
Centre National de la Recherche Scientifique, Institut Jacques Monod, and University of Paris Diderot-Paris 7, Paris, France

RSC Publishing

ISBN: 978-0-85404-134-3
A catalogue record for this book is available from the British Library

Published by The Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road,
Cambridge CB4 0WF, UK

Registered Charity Number 207890

For further information see our web site at www.rsc.org

Foreword

The process of making RNA obliges the DNA-dependent RNA polymerases to perform a remarkable number of separate tasks:

- They scan duplex DNA to find the sites for initiating transcription.
- At these sites, the promoters, they select the DNA strand that will serve as the transcription template and expose it for base-paired binding of their ribonucleoside triphosphate substrates.
- They polymerize RNA chains through distinctive processes of initiation, promoter escape and elongation.
- They discriminate against the incorporation of deoxyribonucleotides into RNA.
- They recognize specific DNA sites for disassembling the elongating transcription complex, releasing the nascent transcript and disengaging from the DNA template.

They also do many things that are less obviously intrinsic to a nucleic acid polymerase:

- They abortively and repetitively initiate transcription at the promoter, generating short ribo-oligonucleotides.
- They elongate RNA chains at non-uniform rates, pausing at characteristic sites.
- They detect incorrect nucleotide addition to the elongating transcript and increase the accuracy of transcription by proofreading.
- They set intrinsically unequal rates of transcription at different promoters by DNA sequence-determined variations of binding affinity, rates of promoter opening and ease of escape from the promoter.
- They exert force on transcription-obstructing DNA-bound proteins, clearing them out of their path as they elongate RNA chains.
- They exert force on their DNA templates, with diverse consequences for the architecture of the eukaryotic nucleus and the prokaryotic nucleoid/chromoid.
- They monitor their DNA templates for damage.

- They recruit accessory proteins for several of these activities, particularly transcript elongation, termination and DNA repair, and they also interact with other molecular machines in order to couple transcription with subsequent RNA processing.

This is an exciting time for understanding and admiring transcription in its machine-like and mechanistic terms. Within the past approximately 10 years, the determination of the structure of the multisubunit RNA polymerases has profoundly transformed the way in which ideas about transcription mechanisms are formulated and tests of these ideas are designed. The determination of the structure of the eukaryotic RNA polymerase II from budding yeast and, most recently, of an archaeal RNA polymerase has made the common evolutionary roots of the multisubunit enzymes vividly apparent. At the same time, comparison of structures of single-subunit nucleic acid polymerases with multisubunit RNA polymerases establishes the existence of catalytic mechanisms that are common to all nucleic acid polymerizations.

During the same period, the spectacular development of methods for examining single molecules of RNA polymerase in action has opened up entirely new possibilities for probing the mechanical and motor-like aspects of RNA polymerases. The process of directly observing RNA synthesis one molecule at a time reveals insights that are difficult to retrieve from, or are entirely obscured in, observations of ensembles in bulk solution. At the same time, technical advances have significantly increased the power of longer-established analytical approaches (*e.g.*, fast reaction kinetics).

This book presents a synthesis of these streams of endeavor. Overview chapters that focus on the mechanism–structure interface and the structure–machine interface introduce the two sections of the book, while individual chapters within each section concentrate more specifically on particular processes – kinetic analysis, single-molecule spectroscopy, and termination of transcription, for example. Seen from the perspective of (nearly) 50 years ago, the detail in which every step of transcription is currently understood is remarkable, and the ways in which that detail illuminates every aspect of gene regulation is enormously satisfying. From the current perspective, the dominant sense is of unanswered questions, of experiments addressing key aspects of mechanism that remain open to competing interpretation, of further technical development that would yield insights currently just out of reach, of molecular computations that have not yet been done – *i.e.*, of a work in progress and a field of activity urgently engaged in finding fascinating new questions to answer.

The current understanding of the mechanistic and machine-like aspects of transcription has been formed primarily through work with the bacterial RNA polymerases. While the common evolutionary and mechanistic basis of all transcription can now be appreciated at the structural level, research on the eukaryotes has been dominated by the challenge of enumerating and understanding elaborations of the core transcription machinery with extrinsic initiation, elongation and termination factors and complexes (the core

transcription initiation factors of budding yeast alone comprise nearly 50 polypeptide chains with an aggregate mass of more than 2.5×10^6 MDa), the dominant role of chromatin structure and modification in regulation of transcription, direct coupling of transcription with post-transcriptional RNA processing and, recently, the role of small RNAs in these processes. Many questions about the RNA polymerase machine that are specific to eukaryotic transcription, especially transcription of nucleosomal chromatin and the important role of elongation factors, are open to lines of experimentation and analysis that are described here for the bacterial enzymes. For these lines of inquiry, the work on bacterial transcription that is presented here points to, and lights up, the path.

A Foreword is a good place to relate how and where this very large endeavor started. The activity of DNA-dependent RNA polymerase was discovered at the University of Chicago's Argonne Cancer Research Hospital in 1959 and announced in a brief note in the *Journal of the American Chemical Society* in August of that year by S.B. Weiss and his assistant L. Gladstone. They showed that incorporation of ^{32}P-labeled CTP into acid-insoluble (*i.e.*, polymeric) form in a preparation of rat liver nuclei required all four ribo NTPs, ATP, GTP, CTP and UTP. The product of their synthesis was degraded by pancreatic ribonuclease but not by deoxyribonuclease I. Moreover, degradation with alkali of the radioactive product made with CTP ^{32}P-labeled in the α position distributed radioactivity to all four ribonucleoside 2′ and 3′ monophosphates. This implied the synthesis of RNA polymers of complex sequence, as opposed to the mere addition of CMP to the ends of nucleic acid chains, either adventitiously to DNA or as the matured CCA adduct of tRNAs. RNA synthesis was strongly inhibited by pyrophosphate but indifferent to orthophosphate, distinguishing the enzyme from polynucleotide phosphorylase. The role of DNA in RNA synthesis was, however, not resolved.

A year later, A. Stevens, then a postdoctoral fellow at NIH, and J. Hurwitz, A. Bresler and R. Diringer at the NYU School of Medicine separately announced the existence of a comparable activity in extracts of *Escherichia coli*. RNA synthesis by the abundantly active bacterial extracts was readily shown to be profoundly dependent on DNA. In contrast, Weiss could not separate the mammalian RNA polymerase activity from DNA and also turned to work with a bacterial enzyme (from *Micrococcus luteus*). With the bacterial preparations, J.J. Furth, Hurwitz and M. Goldman and also A. Stevens, followed by Weiss and T. Nakamoto soon showed the correspondence of the relative incorporation of (AMP + UMP) to (GMP + CMP) into synthesized RNA with the guanine-cytosine content of added DNA. Weiss and Nakamoto extended the analysis of the DNA template–RNA product relationship to the level of nearest neighbor nucleotide pairs by essentially copying an elegant analytical strategy devised for DNA polymerase (by J. Josse, A.D. Kaiser and A. Kornberg) that had been published just months before.

Using CsCl density centrifugation, B.D. Hall and S. Spiegelman had just provided the definitive proof that RNA made in phage T2-infected *E. coli* was specific to the infecting virus by showing that it was able to form DNA–RNA

hybrid duplexes with phage DNA. Weiss, Nakamoto and I adopted this approach to show that the RNA synthesized *in vitro* by the bacterial RNA polymerase generated a polymeric product that was fully complementary to the eliciting double-stranded T2 phage DNA and corresponded, in that sense, to RNA made in the phage-infected cell. This (simple) experiment showed that DNA, which was still widely referred to as the "primer" of RNA synthesis, in fact was its template. The same series of experiments yielded the additional information that the newly synthesized RNA was released from its template and that transcription did not separate template DNA strands. However, both strands of T2 DNA were transcribed, yielding RNA that was self-complementary. In similar experiments with the RNA polymerase activity from *E. coli* and phage ϕX174 DNA, M.N. Hayashi and S. Spiegelman at the University of Illinois, as well as M. Chamberlin and P. Berg at Stanford, also found both strands of their phage DNA templates transcribed. In contrast, the RNA isolated from cells infected with diverse phages was soon found to be DNA-strand-selected, or "asymmetric," which is consistent with the requirements of messenger RNA function in instructing protein synthesis. Was something missing from the *in vitro* RNA synthesis system, or was it conceivable that mRNA might have to be selected after transcription by an additional process, with unusable transcripts rapidly disposed of? The first alternative implied an ability to select specific sites on DNA for starting transcription; the alternative hypothesis clearly failed on two counts: it was too elaborate and inelegant to be plausible, and it implied that transcription yields its functional products at the cost of an energy-consuming futile cycle. Thus, the hunt for DNA strand-selective "asymmetric" transcription was on. Of course, it can now be appreciated that the dichotomy was, to some extent, false. The human genome is pervasively transcribed, with both complementary DNA strands frequently serving as transcription templates. Moreover, cellular processes for very fast disposal of nonfunctional/unusable transcripts do exist.

In any case, the question of template strand selection in transcription was soon answered by J. Marmur's group at Brandeis, Hayashi and Spiegelman, as well as G.P. Tocchini-Valentini and co-workers at Chicago and the International Laboratory of Genetics and Biophysics in Naples, with bacterial RNA polymerase preparations that yielded strand-selective transcription, implying the ability of the enzyme to select specific DNA sites for production of RNA *in vitro*. In hindsight, the something missing from, or inactivated in, the polymerase preparations used for the initial experiments must have been σ, the initiation-specific subunit of bacterial RNA polymerases, discovered several years later by R.R. Burgess and A.A. Travers at Harvard, and E.K.F. Bautz and J.J. Dunn at Rutgers. Within the next year, partially purified bacterial RNA polymerase had been prepared in several laboratories, including those already referred to, W. Zillig's group and others. Those first experiments also provided what we now understand to have been a straightforward and simple demonstration of the existence of genes under positive transcriptional control: "asymmetric" transcription of phage T4 DNA, and the DNA of large-tailed

phages infecting *Bacillus subtilis*, selectively yielded transcripts that correspond to RNA produced at the outset of phage infection, the so-called early RNA. This is, in a sense, the historical baseline for the "modern synthesis" represented by this book.

E.P. Geiduschek
Division of Biological Sciences, UCSD, La Jolla

Preface

It is now fashionable to view the cell as an ensemble of molecular machines or motors working in concert. What exactly do we mean by employing these terms normally reserved for industry and engineering? Up to what point is the analogy drawn between man-made motors and Nature's molecular motors justified? When does this analogy become counterproductive? If one considers RNA polymerase, the analogy is evidently useful for helping ask the right questions about its function. How does a given RNA polymerase molecule cope with the various biochemical and biomechanical tasks it has to accomplish – physically translocating along DNA, acting as a faithful chemical replicator, and properly targeting biological "start" and "stop" signals along a given DNA sequence? These are some of the naïve questions posed and addressed within the ten chapters of this book.

Tackling these challenges and concisely reporting the present achievements and bottlenecks encountered in this venture requires both clarity and humility. We are very grateful to all the authors for their illuminating contributions, and for their compliance with our numerous requests.

We are also extremely grateful to David Lilley for having offered us the opportunity to publish this book in the prestigious collection of the RSC Biomolecular Science series. It has been a pleasure working with Janet Freshwater and Annie Jacob of the Royal Society of Chemistry and we thank them for their support, patience and humor.

In Dante's Inferno there must exist a tiny, poorly-lit room where pale fellows frantically scratch at thick piles of papers while still receiving "final" versions, totally ruining their previous efforts. We have left Tantalus and Sisyphus there, thanks to our saviors, Marie-Hélène and Cary.

Henri Buc and Terence Strick

RSC Biomolecular Sciences
RNA Polymerases as Molecular Motors
Edited by Henri Buc and Terence Strick

Published by the Royal Society of Chemistry, www.rsc.org

Contents

RSC Biomolecular Sciences
RNA Polymerases as Molecular Motors
Edited by Henri Buc and Terence Strick

Published by the Royal Society of Chemistry, www.rsc.org

There and Back Again: A Structural Atlas of RNAP

SETH DARST

Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

RSC Biomolecular Sciences
RNA Polymerases as Molecular Motors
Edited by Henri Buc and Terence Strick

Published by the Royal Society of Chemistry, www.rsc.org

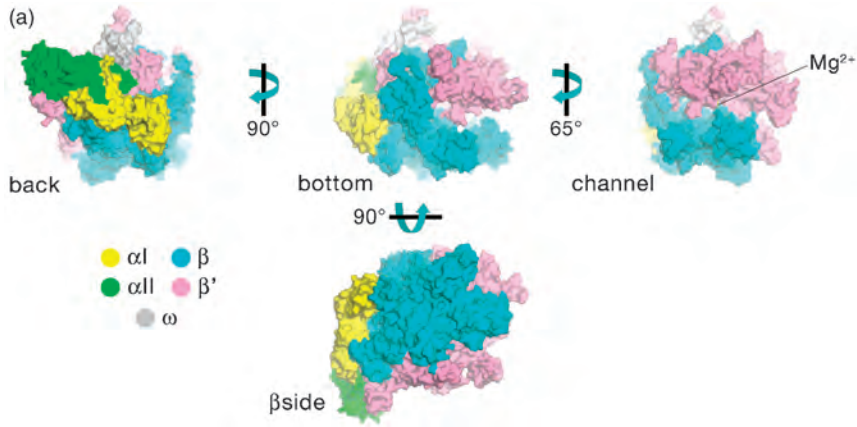

Figure A.1 Structural overview of bacterial core RNAP, σ, and holoenzyme. (a) Reference views of *Thermus aquaticus* core RNAP,[1] shown as a molecular surface with subunits colored as indicated. A Mg^{2+}-ion chelated at the RNAP active center is shown as a yellow sphere (barely visible in the channel view). A 295-residue non-conserved insert (β′ residues 158–452) is not shown (see ref. 2). For parts (b)–(e), each of the reference views is expanded into four components: *Upper left*: Core RNAP [molecular surface colored as part (a)]. *Lower left*: σ^A domains 2–4 [(σ_2, σ_3, σ_{3-4} loop, σ_4; see ref. 3); molecular surface colored orange] as they are seen in the RNAP holoenzyme structure,[4,5] along with the RNAP active center Mg^{2+} (purple sphere). *Upper right*: RNAP holoenzyme [refs 4. and 5; molecular surface colored as in part (a) and with σ^A orange]. *Lower right*: RNAP holoenzyme (molecular surface with the core RNAP colored white and rendered transparent). Structural components, when visible, are labeled: **β1**: Upstream lobe of the β subunit.[1] β2: Downstream lobe of the β subunit.[1] **Clamp**: Structural element consisting of parts of both the β and β′ subunits.[6] Undergoes large conformational change, "clamping" down on the RNA/DNA hybrid in the RNAP active site to stabilize the elongation complex. **Flap**: Structural element of the β subunit.[1] Occludes the RNA exit channel[1,7,8] and interacts with σ_4.[4,9] **Jaw**: Structural element of the β′ subunit. **Lid**: Structural element of the β′ subunit.[10] Interacts with the σ_{3-4} loop[4] and is involved in nucleic acid interactions at the upstream edge of the RNA/DNA hybrid in the elongation complex. Believed to be involved in "peeling off" the RNA from its hybrid with the template DNA strand.[11] **Rudder**: Structural element of the β′ subunit[1] involved in nucleic acid interactions at the upstream edge of the RNA/DNA hybrid in the elongation complex (Westover et al. 2004). Elongation complexes prepared with rudderless RNAP mutant displays reduced stability.[12] **Zbd**: Zinc-binding domain – Structural element of the β′ subunit.[5] Involved in nucleic acid interactions at the upstream edge of the RNA/DNA hybrid in the elongation complex. Zbd mutants form stable TECs but are defective in termination and antitermination.[13] **Zipper**: Structural element of the β′ subunit[10] involved in interactions at the upstream edge of the RNA/DNA hybrid in the elongation complex.

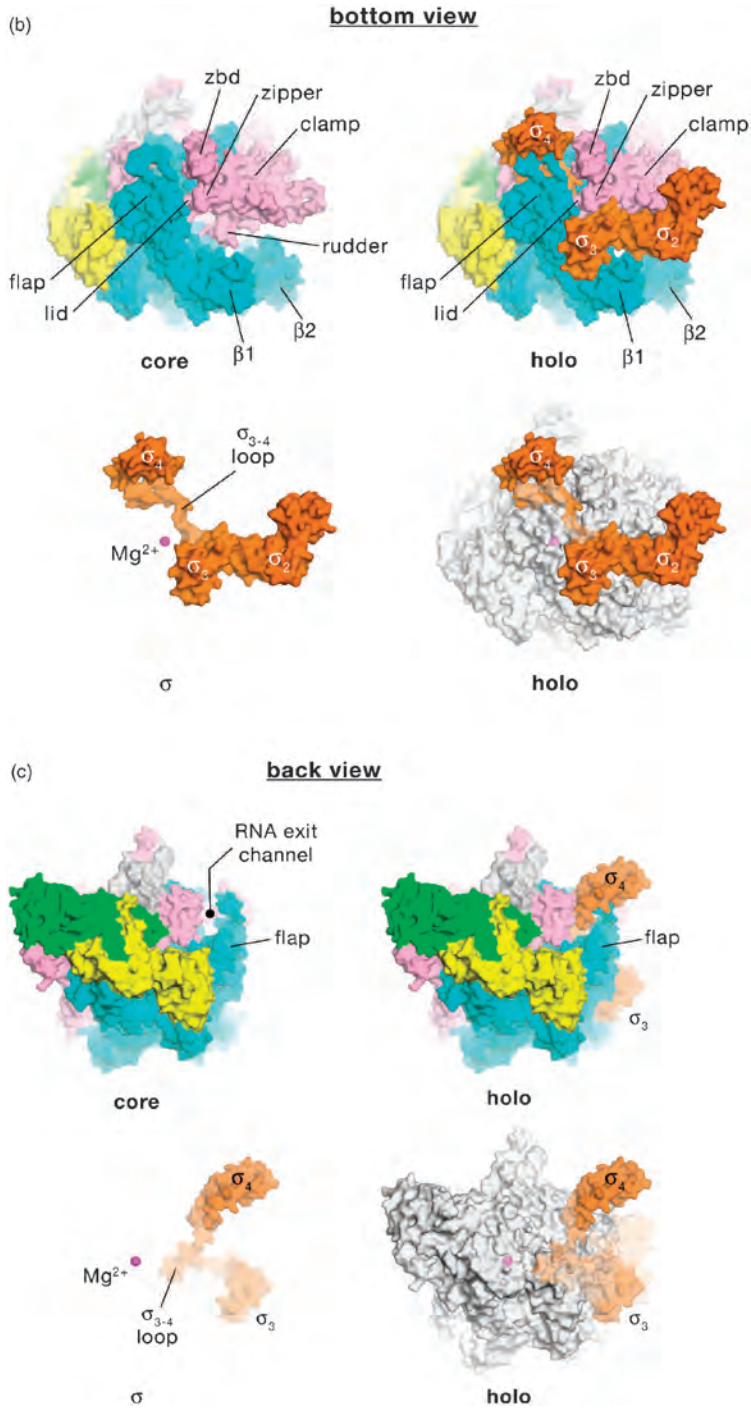

Figure A.1 Continued

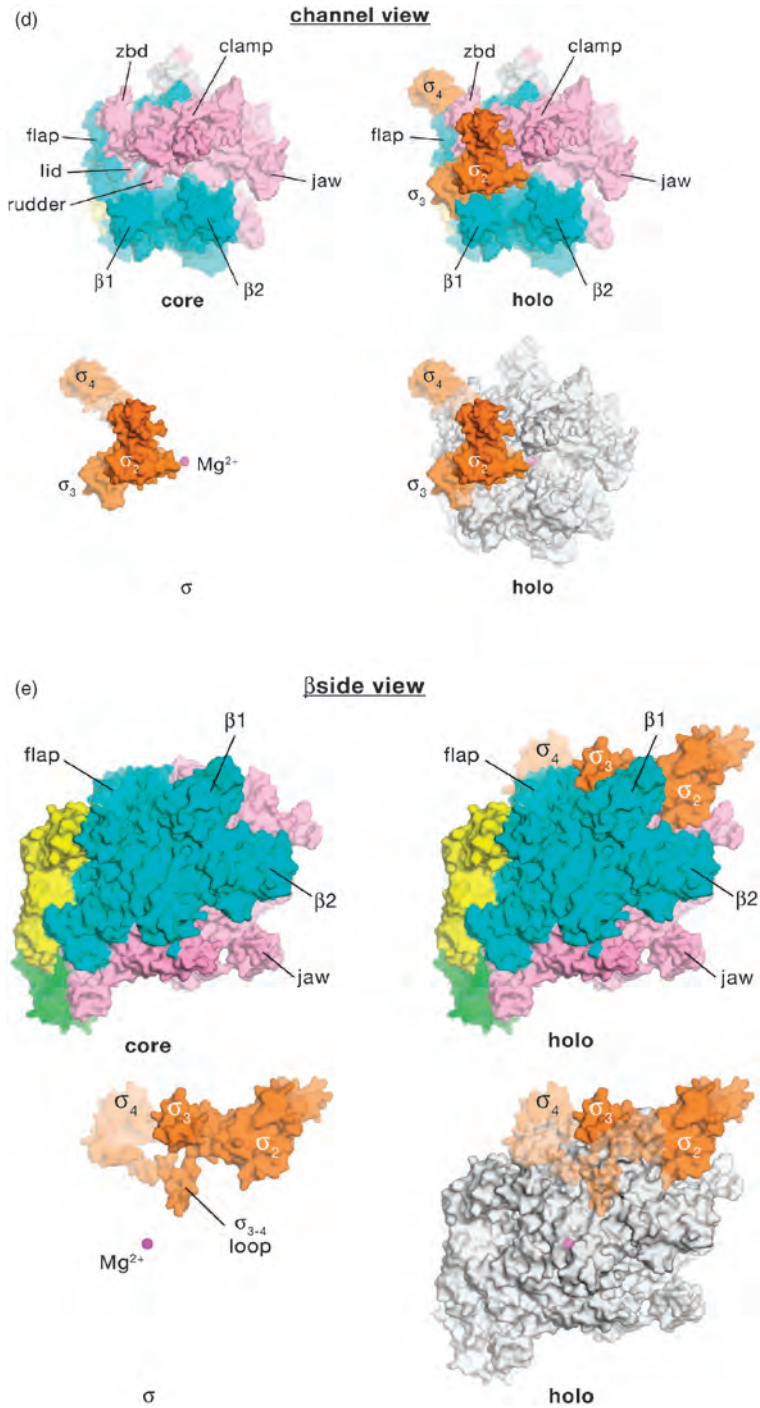

Figure A.1 Continued

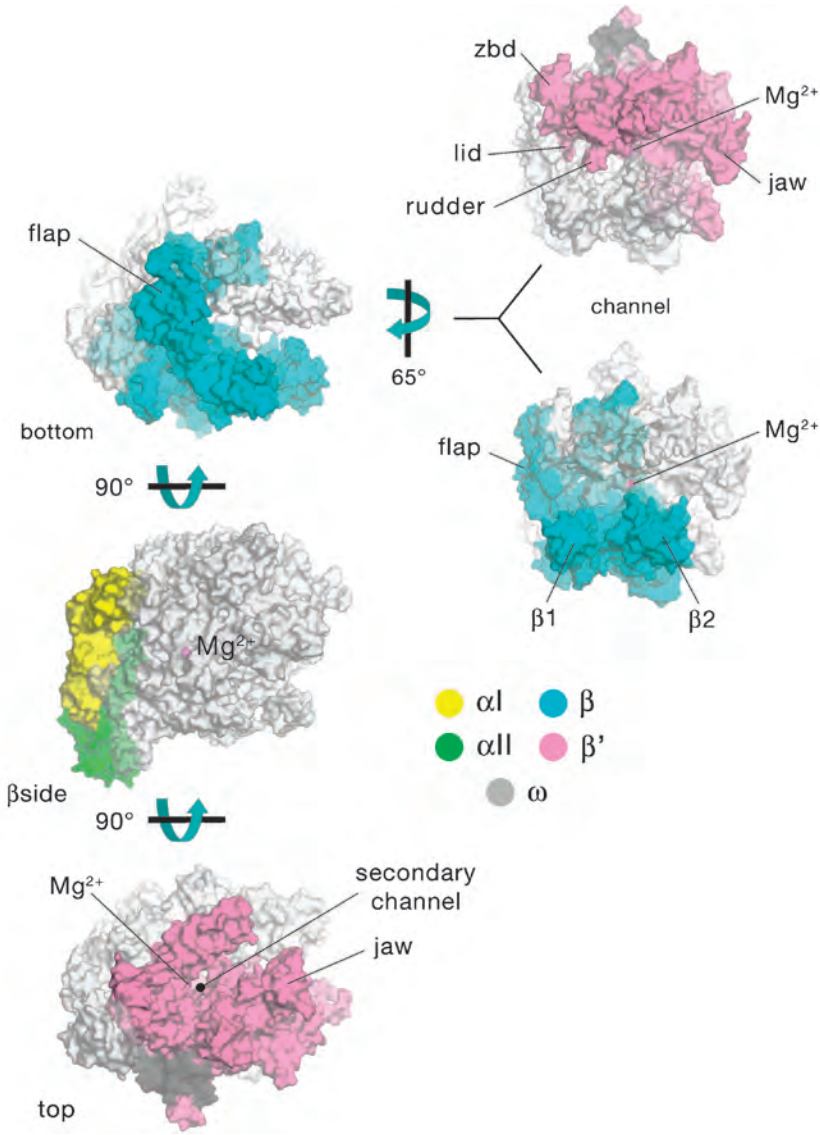

Figure A.2 Core RNAP subunits. Various views of core RNAP.[1] In each view, the RNAP is shown as a white, transparent molecular surface – except one subunit is shown non-transparent and colored (to show how the subunit fills up the space of the core RNAP). The active center Mg^{2+} is shown as a purple sphere. The bottom view highlights the β subunit. The β side view highlights the α subunit dimer. The top view highlights the β′ subunit. There are two channel views, one highlighting β′ (above), one highlighting β (below).

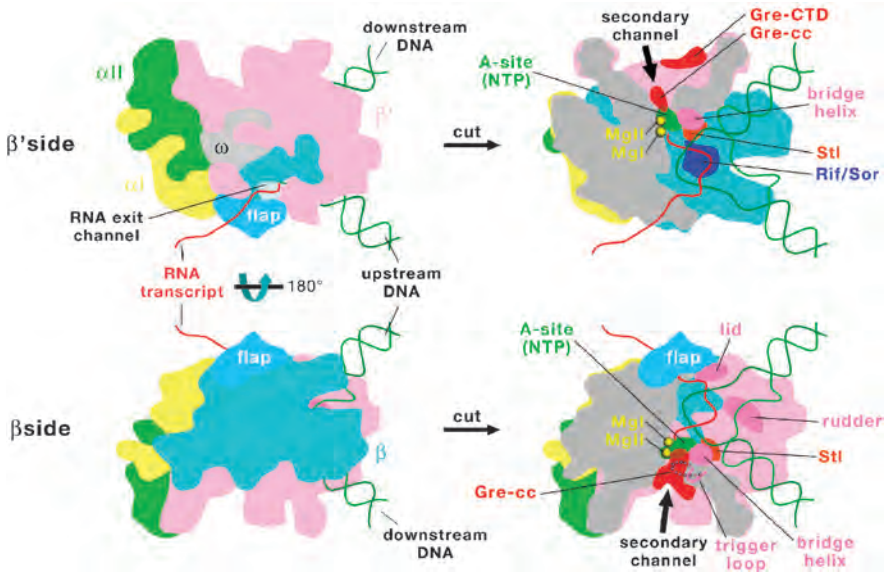

Figure A.3 Cartoon schematics of the RNAP elongation complex.[6–8,11,14–16] The top row shows the β′ side view, the bottom row gives the β side view. The left-hand column shows the entire elongation complex structure, the right-hand column depicts cross-sections in which obscuring protein elements have been cut away to reveal the inside of the active site channel. The active center Mg^{2+} is shown as a yellow sphere labeled "MgI". Important structural elements of the RNAP are shown in magenta (bridge helix, lid, rudder, trigger loop). The trigger loop is shown in two positions, one pointing away from the active center, and one seen in structures of elongating complexes bound to the incoming NTP substrate (magenta dots).[17] (Vassylyev et al. 2007). In the cross-sectional views, binding sites for structurally characterized RNAP ligands or effectors are highlighted as follows: **A-site (NTP)**, green: Binding site for the incoming NTP substrate (Westover et al. 2004; Vassylyev et al. 2007), also sometimes called the insertion site (IS), or also the $i+1$ site. The incoming NTP substrate arrives in the A-site via the secondary channel,[1] (Westover et al. 2004) and chelated to a Mg^{2+} ion, which becomes an essential component of the active site (MgII, also sometimes called metal B; see ref. 18). **Gre**, red: Gre-factors (ref. 19) bind via their C-terminal domain (Gre-CTD) outside the secondary channel, and insert their coiled-coil finger (Gre-cc) into the secondary channel.[20] The tip of the Gre-cc contains absolutely conserved acidic residues that serve to stabilize the binding of MgII, which is required for the endonucleolytic cleavage of backtracked RNA.[21,22] **Rif/Sor**, blue: Rifamycins,[23] a key component of tuberculosis therapy,[24] bind in a pocket of the β subunit in the active site channel[25,26] and inhibit RNAP by blocking the path of the elongating RNA transcript.[25] Sorangicin A, a chemically unrelated inhibitor,[27] binds in the identical site and likely inhibits RNAP by a similar mechanism.[28] **StI**, orange: The bacterial RNAP inhibitor streptolydigin[29] binds to a site along the bridge helix.[30,31]

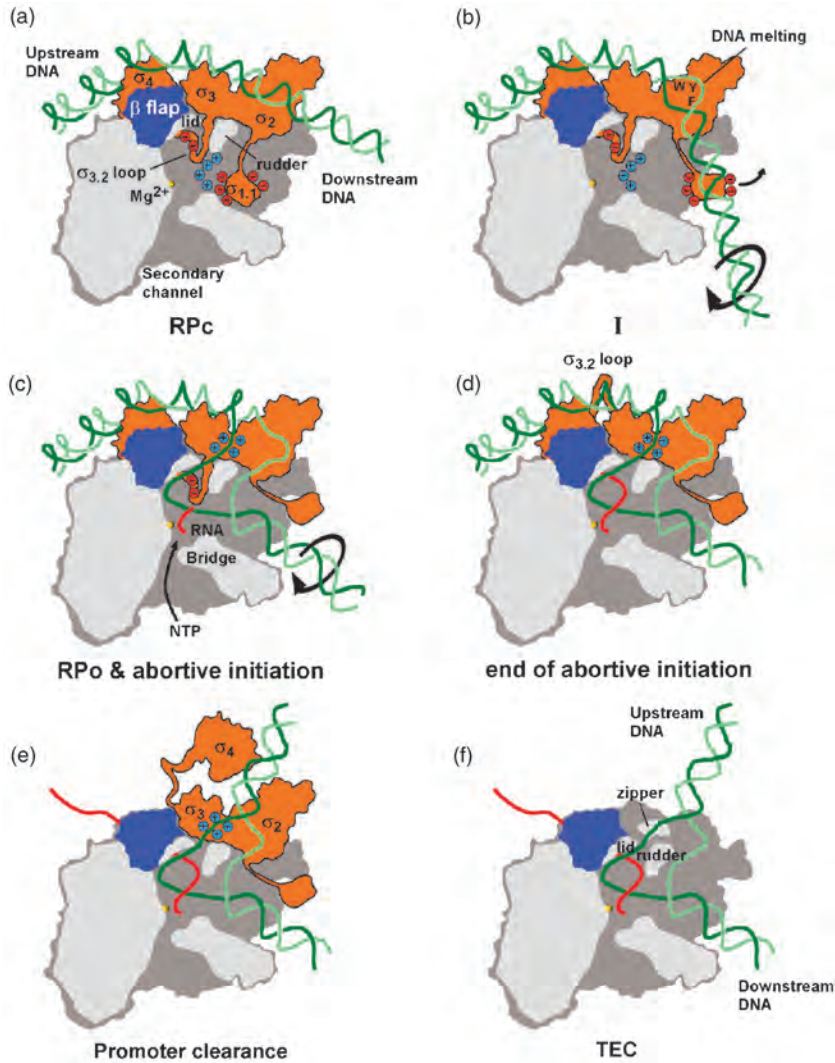

Figure A.4 Structural transitions during the steps of transcription initiation. Shown are cross-sectional views of the RNAP holoenzyme (β flap, blue; σ, orange; rest of RNAP, gray; catalytic Mg^{2+}, yellow sphere), promoter DNA (template strand, dark green; nontemplate strand, light green), and the RNA transcript (red) at (a) the RPc, (b) intermediate (I), (c) RPo and abortive initiation, (d) end of abortive initiation, (e) promoter clearance and (f) TEC stages of transcription initiation. The view is looking down on top of the β subunit, but with most of β removed, revealing the inside of the RNAP active site channel.

References

1. G. Zhang, E. A. Campbell, L. Minakhin, C. Richter, K. Severinov and S. A. Darst, Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution, *Cell*, 1999, **98**, 811–824.
2. M. Chlenov, S. Masuda, K. S. Murakami, V. Nikiforov, S. A. Darst and A. Mustaev, Structure and function of lineage-specific sequence insertions in the bacterial RNA polymerase b′ subunit, *J. Mol. Biol.*, 2005, **353**, 138–154.
3. E. A. Campbell, O. Muzzin, M. Chlenov, J. L. Sun, C. A. Olson, O. Weinman, M. L. Trester-Zedlitz and S. A. Darst, Structure of the bacterial RNA polymerase promoter specificity sigma factor, *Mol. Cell*, 2002, **9**, 527–539.
4. K. Murakami, S. Masuda and S. A. Darst, Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution, *Science*, 2002, **296**, 1280–1284.
5. D. G. Vassylyev, S. Sekine, O. Laptenko, J. Lee, M. N. Vassylyeva, S. Borukhov and S. Yokoyama, Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution, *Nature*, 2002, **417**, 712–719.
6. A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell and R. D. Kornberg, Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution, *Science*, 2001, **292**, 1876–1882.
7. N. Korzheva, A. Mustaev, M. Kozlov, A. Malhotra, V. Nikiforov, A. Goldfarb and S. A. Darst, A structural model of transcription elongation, *Science*, 2000, **289**, 619–625.
8. D. G. Vassylyev, M. N. Vassylyeva, A. Perederina, T. H. Tahirov and I. Artsimovitch, Structural basis for transcription elongation by bacterial RNA polymerase, *Nature*, 2007, **448**, 157–162.
9. K. Kuznedelov, L. Minakhin, A. Niedziela-Majka, S. L. Dove, D. Rogulja, B. E. Nickels, A. Hochschild, T. Heyduk and K. Severinov, A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition, *Science*, 2002, **295**, 855–857.
10. P. Cramer, D. A. Bushnell and R. D. Kornberg, Structural basis of transcription: RNA polymerase II at 2.8 Å resolution, *Science*, 2001, **292**, 1863–1876.
11. K. D. Westover, D. A. Bushnell and R. D. Kornberg, Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center, *Cell*, 2004, **119**(4), 481–9.
12. K. Kuznedelov, N. Korzheva, A. Mustaev and K. Severinov, Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA-DNA hybrid length, *EMBO J.*, 2002, **21**, 1369–1378.
13. R. A. King, D. Markov, R. Sen, K. Severinov and R. A. Weisberg, A conserved zinc binding domain in the largest subunit of DNA-dependent RNA polymerase modulates intrinsic transcription termination and

antitermination but does not stabilize the elongation complex, *J. Mol. Biol.*, 2004, **342**, 1143–1154.
14. H. Kettenberger, K. -J. Armache and P. Cramer, Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS, *Mol. Cell*, 2004, **16**, 955–965.
15. K. D. Westover, D. A. Bushnell and R. D. Kornberg, Structural basis of transcription: Separation of RNA from DNA by RNA polymerase II, *Science*, 2004, **303**, 1014–1016.
16. D. G. Vassylyev, M. N. Vassylyeva, J. Zhang, M. Palangat, I. Artsimovitch and R. Landick, Structural basis for substrate loading in bacterial RNA polymerase, *Nature*, 2007, **448**, 163–168.
17. D. Wang, D. A. Bushnell, K. D. Westover, C. D. Kaplan and R. D. Kornberg, Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis, *Cell*, 2006, **127**(5), 941–54.
18. T. A. Steitz, A mechanism for all polymerases, *Nature*, 1998, **391**, 231–232.
19. S. Borukhov, V. Sagitov and A. Goldfarb, Transcript cleavage factors from *E. coli*, *Cell*, 1993, **72**, 459–466.
20. N. Opalka, M. Chlenov, P. Chacon, W. J. Rice, W. Wriggers and S. A. Darst, Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase, *Cell*, 2003, **114**, 335–345.
21. O. Laptenko, J. Lee, I. Lomakin and S. Borukhov, Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase, *EMBO J.*, 2003, **23**, 6322–6334.
22. E. Sosunova, V. Sosunov, M. Kozlov, V. Nikiforov, A. Goldfarb and A. Mustaev, Donation of catalytic residues to RNA polymerase active center by transcription factor Gre. *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 15469–15474.
23. P. Sensi, History of the development of rifampin, *Rev. Infect. Dis.*, 1983, **5**(Supp. 3), 402–406.
24. H. G. Floss and T. W. Yu, Rifamycin-mode of action, resistance, and biosynthesis, *Chem. Rev.*, 2005, **105**, 621–632.
25. E. A. Campbell, N. Korzheva, A. Mustaev, K. Murakami, S. Nair, A. Goldfarb and S. A. Darst, Structural mechanism for rifampicin inhibition of bacterial RNA polymerase, *Cell*, 2001, **104**, 901–912.
26. I. Artsimovitch, M. N. Vassylyeva, D. Svetlov, V. Svetlov, A. Perederina, N. Igarashi, N. Matsugaki, S. Wakatsuki, T. H. Tahirov and D. G. Vassylyev, Allosteric modulation of the RNA polymerase catalytic reaction is an essential component of transcription control by rifamycins, *Cell*, 2005, **122**, 351–363.
27. H. Irschik, R. Jansen, K. Gerth, G. Hofle and H. Reichenbach, The sorangicins, novel and powerful inhibitors of eubacterial RNA polymerase isolated from myxobacteria, *J. Antibiotics*, 1985, **40**, 7–13.
28. E. A. Campbell, O. Pavlova, N. Zenkin, F. Leon, H. Irschik, R. Jansen, K. Severinov and S. A. Darst, Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase, *EMBO J.*, 2005, **24**, 674–682.

29. G. Cassani, R. R. Burgess, H. M. Goodman and L. Gold, Inhibition of RNA polymerase by streptolydigin, *Nature New Biol*., 1971, **230**, 197–200.
30. D. Temiakov, N. Zenkin, M. N. Vassylyeva, A. Perederina, T. H. Tahirov, E. Kashkina, M. Savkina, S. Zorov, V. Nikiforov, N. Igarashi, N. Matsugaki, S. Wakatsuki, K. Severinov and D. G. Vassylyev, Structural basis of transcription inhibition by antibiotic streptolydigin, *Mol. Cell*, 2005, **19**, 655–666.
31. S. Tuske, S. G. Sarafianos, X. Wang, B. Hudson, E. Sineva, J. Mukhopadhyay, J. J. Birktoft, O. Leroy, S. Ismail, A. D. J. Clark, C. Dharia, A. Napoli, O. Laptenko, J. Lee, S. Borukhov, R. H. Ebright and E. Arnold, Inhibition of bacterial RNA polymerase by streptolydigin: Stabilization of a straight-bridge-helix active-center conformation, *Cell*, 2005, **122**, 541–552.

Part I From Promoter Recognition to Promoter Escape

CHAPTER 1

Where it all Begins: An Overview of Promoter Recognition and Open Complex Formation

STEPHEN BUSBY,[a] ANNIE KOLB[b] AND HENRI BUC[c]

[a] School of Biosciences, University of Birmingham, Birmingham B15 2TT, United Kingdom; [b] Institut Pasteur, Molecular Genetics Unit and CNRS URA 2172, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France; [c] CIS Institut Pasteur, 75724 Paris Cedex 15, France.

1.1 Gene Expression as a Driver of Life

The importance of transcription, the process by which information encoded in DNA is copied into RNA, cannot be overstated. As soon as the dogma that DNA makes RNA makes protein was established, the hunt was on for the machinery that orchestrates transcription. Thus, in the late 1950s and early 1960s, classical methods of protein fractionation were used to identify DNA-dependent RNA polymerase activity. Remarkably, in parallel, primarily using *Escherichia coli* genetics, Jacob, Monod and their colleagues were discovering gene regulatory proteins and establishing the paradigm that gene transcription was the key point at which gene regulation is effected.[1] Thus, right from the start, *Escherichia coli* K-12 was established as the model system to use and, with the benefit of hindsight, it is easy to see now how 40 years of amazing progress

RSC Biomolecular Sciences
RNA Polymerases as Molecular Motors
Edited by Henri Buc and Terence Strick

Published by the Royal Society of Chemistry, www.rsc.org

was sparked by the fusion of two very different worlds, one populated by the biochemists and the other by the bacterial geneticists. Put very simply, the stories in this book expand on how the biochemistry explains the genetics and how the genetics gives reason to the biochemistry. The crucial discoveries that set the scene for these stories were made in the late 1960s: the characterization of the single multi-subunit RNA polymerase in *E. coli*, the discovery of promoters and terminators, and the realization that different genes are transcribed at widely differing frequencies. The pace accelerated with the arrival of cloning and DNA sequencing in the 1970s and in-depth studies of how different promoters are regulated exploiting increasingly sophisticated methodologies. The arrival of whole genome sequences in the late 1990s led to the complete catalogue of the different players and attempts to integrate our knowledge with systems biology approaches. And finally, the structural biologists have provided us with models of many of the major players, including the multi-subunit RNA polymerases, the principal topic of this book.

1.2 *Escherichia coli* RNA Polymerase

The view of bacterial RNA polymerase as a 500 kDa enzyme with subunit structure $\alpha_2\beta\beta'\omega\sigma$ had a long and slow birth, emerging from heroic biochemistry in both the USA and in Germany. It is easy to overlook the difficulties encountered by the pioneers in this field of proving the integrity and function of such a large multi-subunit complex. DNA cloning technologies had not yet arrived and early efforts to demonstrate specific DNA-directed transcription mostly had to exploit viral templates, notably bacteriophages. Perhaps the most influential single observation was the chance discovery by Dick Burgess and colleagues in 1969 that passage of the preparation of *E. coli* RNA polymerase through phosphocellulose led to loss of its ability to initiate specific transcripts and that this loss was due to the loss of the σ factor.[2] This led to the definition of two forms of RNA polymerase, the holo-enzyme with composition $\alpha_2\beta\beta'\omega\sigma$, and the core enzyme, $\alpha_2\beta\beta'\omega$, devoid of σ, and the notion of σ as the factor controlling transcript initiation. Another influential early finding came from Mike Chamberlin and colleagues, who showed that the transcriptionally competent complexes formed between the holoenzyme and DNA were resistant to heparin[3] (see also ref. 4). In these complexes, which could form in the absence of any nucleotides, the heparin resistance arises from the template DNA strands being locally unwound around the transcription start site. These observations gave birth to the idea of a pathway to transcription initiation, with the kinetically competent or open complex being preceded by a heparin-sensitive closed complex in which the DNA strands are not open.[5,6] Amazingly, the nature of closed complexes, the mechanics of the closed to open transition and the number of intermediates remain hot topics for study and debate today.[7,8]

One of the early proofs that *E. coli* contained a single core RNA polymerase was that RNA synthesis could be completely inhibited by the drug, rifampicin,

but a single point mutation can confer complete resistance and normal RNA synthesis.[9] The location of these *rif*^{R} mutations led to the identification of the co-transcribed *rpoB* and *rpoC* genes, which encode the RNA polymerase large β and β′ subunits (1342 and 1407 amino acids respectively).[10] Subsequently, the genes encoding the other RNA polymerase subunits were identified at different locations on the *E. coli* chromosome, and the pathway of subunit assembly was established. The first step is the formation of a dimer of two 329 amino acid α subunits, which acts as a scaffold for the addition of first β and then β′ω to give core enzyme. The holoenzyme is then formed by the addition of the σ subunit. This pathway was established by Akira Ishihama, who later showed that the C-terminal 100 amino acids of each α subunit are dispensable for RNA polymerase assembly.[11,12] The reason for this is that the RNA polymerase α subunit consists of two domains, with the 230 amino acid N-terminal domain being essential for enzyme assembly, whilst the C-terminal contains a separate independently folding domain that plays a key role at certain promoters. We now have detailed structures for both the core and holo enzymes, due largely to the efforts of Seth Darst, Dmitry Vassylyev and their coworkers using RNA polymerases from thermophilic bacteria.[13–15] The structures show the large β and β′ subunits assembled on the two α subunit N-terminal domains, with the β and β′ subunits forming a "crab claw" to accommodate DNA, with the catalytic centre of the enzyme right at the heart of the claw (Figure 1.1; full details are in Chapter 2). This organization is echoed in the structures of yeast RNA polymerase II that emerged from Roger Kornberg's laboratory at the same time, underlining its importance at all levels of life.[16]

Another major landmark in the study of bacterial RNA polymerases has been the realization that most bacteria contain multiple σ factors. This idea first emerged from studies by Rich Losick and others of the genes needed for spore formation by *Bacillus subtilis*. The products of some of these genes showed striking sequence similarities to already discovered σ factors. Since it was clear that σ factors were needed for both promoter specificity and open complex formation, Losick's proposal that the sporulation pathway was driven by the synthesis of new σ factors, which switch on new sets of genes, was soon accepted.[17] In fact, most bacteria contain one dominant σ factor and between 0 and 64 alternatives. The dominant σ, known as the "housekeeping σ" is an essential protein that is responsible for most transcription initiation. In *E. coli*, the predominant σ is σ^{70} with a molecular size of 70 kDa (613 amino acids) and it is this σ factor, encoded by the *rpo*D gene, that is found in most RNA polymerase preparations. The *E. coli* genome encodes six alternative σ factors (encoded by the *rpo*S, *rpo*H, *rpo*E, *rpo*F, *rpo*N and *fec*I genes), which are concerned with the management of different stresses.[18] An increase in the intracellular level of an alternative σ factor (*e.g.*, in response to a specific stress) results in the formation of a subpopulation of RNA polymerase holoenzyme molecules dedicated to initiate transcription at a particular subset of promoters.[19] As more bacterial genomes have been studied, the belief that alternative σ factors have evolved to drive programs of microbial adaptation and differentiation has been reinforced.