Crystal Structure of a Bacterial RNA Polymerase Holoenzyme at 2.6 Å Resolution

The DNA-dependent RNA polymerase (RNAP) is the key enzyme of the transcription process, and is a final target in many regulatory pathways that control gene expression in all living organisms. Bacterial RNAP exists in two forms: core and holoenzyme. The core enzyme (~ 400 kDa) consists of five subunits: α-dimer(α2), β, β', and ω. The transcription cycle in bacterial cells can be divided into three major stages: initiation, elongation, and termination. Although it is catalytically active, the core enzyme is incapable of initiating transcription efficiently and with specificity. For this, it must bind an initiation factor, σ, to form a holoenzyme that can recognize specific DNA sequences (promoters) [1,2]. During initiation, the holoenzyme specifically binds to two conserved hexamers in the promoter at nucleotide (nt) positions –35 and –10 relative to the transcription start site (+1), to form a closed promoter complex. Then, it unwinds the double-stranded DNA around the –10 region (between nt –12 and +2), resulting in the open promoter complex, and initiates transcription in the presence of nucleoside triphosphate substrates [3]. After the synthesis of a 9 - 12 nt-long RNA, of which 8 - 9 are base-paired with the DNA template strand (RNA-DNA hybrid), the transcription complex passes from the initiation to elongation stage [4]. This transition is characterized by the escape of the RNAP from the promoter, the dissociation of σ from the core, and the formation of a highly processive ternary elongation complex. The σ-factor plays a key role in initiation, being directly involved in promoter recognition, DNA melting, and promoter escape and clearance [2]. The family of related σ proteins shares four regions of sequence homology, designated 1 to 4, which are further divided into several subregions [5]. The regions 4.2, 2.3 - 2.4, and 2.5 were shown to recognize the –35, –10, and the so-called "extended –10" elements of the promoter, respectively.

We have determined the crystal structure of the *T. thermophilus* RNAP holoenzyme, containing the major σ-factor (σ70), at 2.6-Å resolution (Fig. 1) at beamline **BL45XU**. The σ70 subunit is located almost entirely on the core surface, except for a short segment (σ313 - 342), which is buried within the core molecule. The modeled structure of σ70

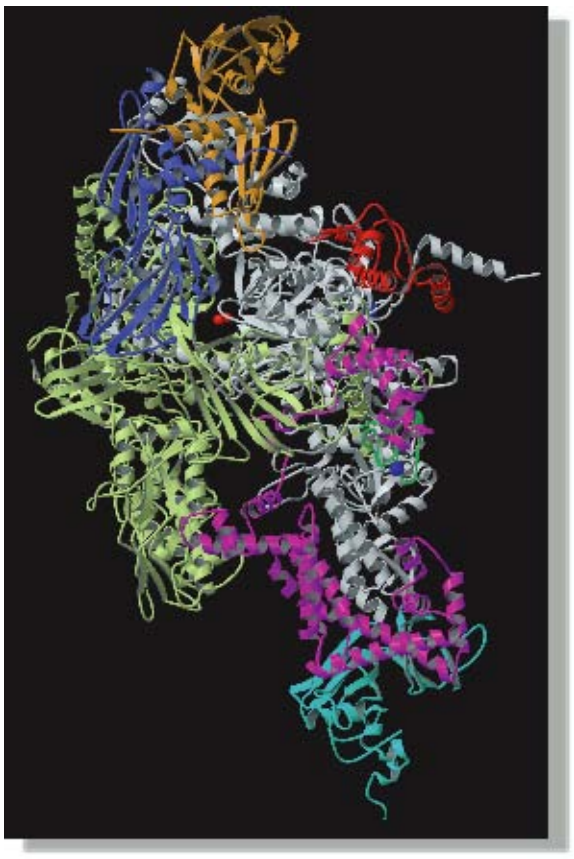

Fig. 1. Holoenzyme crystal structure. The subunits colors are: β, sage; β', white (β'163-452, cyan; β' Zn-finger, green); αI, blue; αII, light orange; σ, magenta; and ω, red. Two catalytic Mg^{2+} (red) and two Zn^{2+} ions (blue) are shown as spheres.

consists entirely of α-helices connected by either turns or loops, and it can be divided into four structural domains: N-terminal domain 1 (ND1), N-terminal domain 2 (ND2), linker domain (LD), and C-terminal domain (CD) (Fig. 2).

ND1 consists of eight α-helices (σ74-254) comprising four helix-turn-helix motifs (HtH). This domain encompasses region 1.2 up to the N-terminal half of region 2.4, including the non-conserved segment between regions 1 and 2. ND1 has a U-shaped structure, and is connected to ND2 (σ261 - 312) by a short linker loop (σ255 - 260), That lies at the C-terminus of region 2.4. ND2, corresponding to conserved regions 2.4 - 2.5 and 3.1, consists of three α-helices that fold into an α-helical bundle. The C-terminal helix of ND1 and the N-terminal helix of ND2 (σ234 - 281) form a V-shaped structure near the opening of the upstream DNA binding channel, which is likely to be a binding site for the –10 element of promoter. The 30 residue-long "linker" domain, LD (σ313 - 339), intervenes between the globular N- and C-terminal portions of σ and has a mostly extended, unfolded conformation. Roughly at its midpoint, LD forms a hairpin loop (σ318 - 329) that protrudes into the active site cleft. The C-terminal domain, CD (σ340

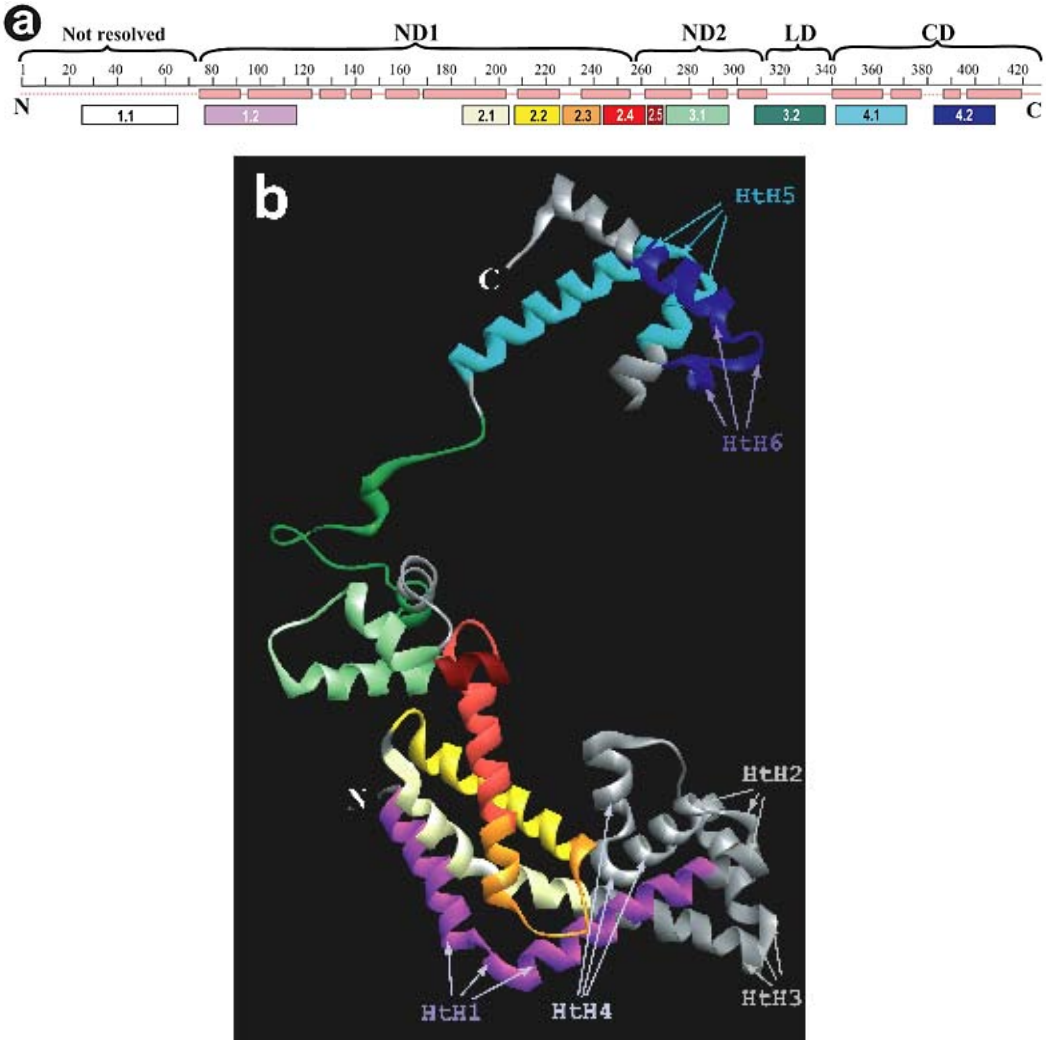

*Fig. 2. σ domain organization and structure. (**a**) Scheme of structural domains and conserved regions. (**b**) Ribbon diagram of σ. The color coding is the same as in (**a**) except for the non-conserved regions (grey).*

- 423), which includes conserved regions 4.1 and 4.2, contains four α-helices, which are arranged as a pair of HtH motifs. This site is about 57 Å from the N-terminal half of the protein, containing regions 2.3 - 2.5, that allows unambiguous modeling of promoter bound to the holoenzyme (Fig. 3).

In the holoenzyme, certain σ structural elements greatly reduce the available space in the functionally important protein cavities and channels that accommodate promoter DNA, RNA-DNA hybrid, and RNA product. Thus, the holoenzyme structure additionally restrains possible orientations and conformations of the nucleic acids in the transcription intermediates, allowing better understanding of various steps of transcription initiation.

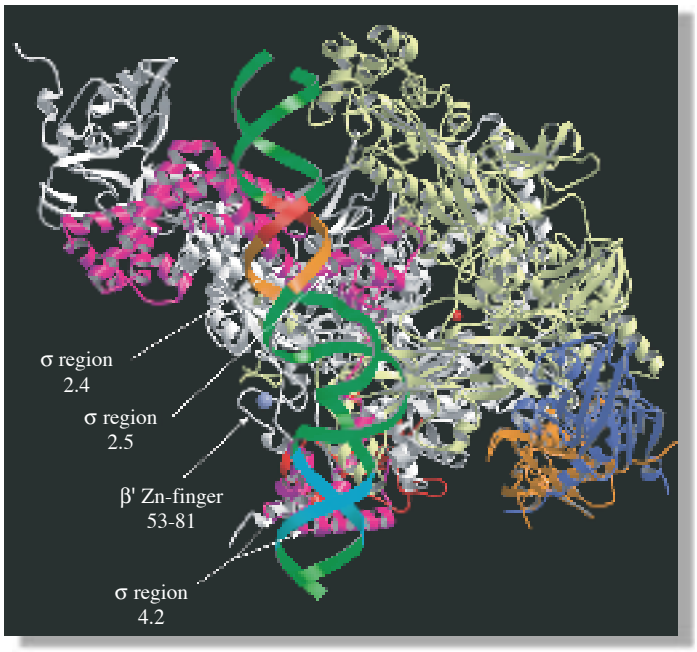

*Fig. 3. Modelling of the holoenzyme/closed promoter complex. The protein colour coding is the same as in Fig. 1(**c**), except all of β' is shown in white. The dsDNA (green) contains the –35 (cyan), –10 (red), extended –10 (orange) promoter regions. The Mg^{2+} (red) and Zn^{2+} (blue) ions are shown as spheres.*

Dmitry G. Vassylyev[a], Sergei Borukhov[b] and Shigeyuki Yokoyama[a,c,d]

(a) SPring-8 / RIKEN
(b) SUNY Health Science Center, USA
(c) RIKEN Genomic Sciences Center
(d) The University of Tokyo

E-mail: dmitry@yumiyoshi.harima.riken.go.jp

References

[1] R.R. Burgess *et al.*, Nature **221** (1969) 43.
[2] C. Gross *et al.*, Transcriptional Regulation (eds S.R. McKnight & K.R. Yamamoto) - Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1992) 129.
[3] M.T.J. Record *et al.*, *Escherichia coli* and *Salmonella* (ed F.C. Neidhart) - ASM Press, Washington, D.C. (1996) 792.
[4] P. von Hippel, Science **281** (1998) 660.
[5] M. Lonetto *et al.*, J. Bacteriol. **174** (1992) 3843.