Cell, Vol. 108, 599–614, March 8, 2002, Copyright ©2002 by Cell Press

Structural Organization of Bacterial RNA Polymerase Holoenzyme and the RNA Polymerase-Promoter Open Complex

Vladimir Mekler,[1,2,4] Ekaterine Kortkhonjia,[1,2,4] Jayanta Mukhopadhyay,[1,2,4] Jennifer Knight,[2,4] Andrei Revyakin,[1,2] Achillefs N. Kapanidis,[1,2,5] Wei Niu,[1,2,6] Yon W. Ebright,[1,2] Ronald Levy,[2] and Richard H. Ebright[1,2,3]

[1]Howard Hughes Medical Institute
Waksman Institute
Rutgers University
Piscataway, New Jersey 08854

[2]Department of Chemistry
Rutgers University
Piscataway, New Jersey 08854

Summary

We have used systematic fluorescence resonance energy transfer and distance-constrained docking to define the three-dimensional structures of bacterial RNA polymerase holoenzyme and the bacterial RNA polymerase-promoter open complex in solution. The structures provide a framework for understanding σ^{70}-(RNA polymerase core), σ^{70}-DNA, and σ^{70}-RNA interactions. The positions of σ^{70} regions 1.2, 2, 3, and 4 are similar in holoenzyme and open complex. In contrast, the position of σ^{70} region 1.1 differs dramatically in holoenzyme and open complex. In holoenzyme, region 1.1 is located within the active-center cleft, apparently serving as a "molecular mimic" of DNA, but, in open complex, region 1.1 is located outside the active center cleft. The approach described here should be applicable to the analysis of other nanometer-scale complexes.

Introduction

Transcription initiation is a multistep process (reviewed in Record et al., 1996; de Haseth et al., 1998). RNA polymerase (RNAP), together with one or more initiation factor(s): (1) binds to promoter DNA to yield an RNAP-promoter closed complex, (2) clamps down on DNA to yield an RNAP-promoter intermediate complex, (3) melts ~14 bp of DNA (transcription bubble; positions −11 to +3 relative to transcription start) to yield an RNAP-promoter open complex, (4) enters into abortive cycles of synthesis and release of short RNA products as an RNAP-promoter-initial transcribing complex, and, ultimately, (5) escapes the promoter and enters into productive RNA synthesis as an RNAP-DNA elongation complex.

High-resolution crystallographic structural information is available for three molecular species relevant to this multistep process: bacterial RNAP core enzyme (Zhang et al., 1999; subunit composition $\beta'/\beta/\alpha^{I}/\alpha^{II}/\omega$), eukaryotic RNAP II core enzyme Δ4/7 (Cramer et al., 2001; subunit composition 1/2/3/5/6/8/9/10/11/12, where 1, 2, 3, 11, and 6 are homologs of bacterial RNAP β', β, α^{I}, α^{II}, and ω; Ebright, 2000), and a eukaryotic RNAP II tailed-template elongation complex (Gnatt et al., 2001).

Further progress in understanding structure and mechanism in transcription will require development of methods to "leverage" the available crystallographic structural information for these three molecular species, in order to obtain structural information about each molecular species on the pathway, to define the structural transitions in protein and nucleic acid at each step in the pathway, to define the kinetics of each step, and to define the impact of promoter sequence and regulators on each step.

We have been developing methods to use fluorescence resonance energy transfer (FRET), a physical phenomenon that permits measurement of distances (Förster, 1948; reviewed in Lilley and Wilson, 2000; Selvin, 2000), to define structure and mechanism in transcription (Mukhopadhyay et al., 2001). FRET occurs in a system having a fluorescent probe serving as a donor and a second fluorescent probe serving as an acceptor, where the emission wavelength of the donor overlaps the excitation wavelength of the acceptor. In such a system, upon excitation of the donor with light of its excitation wavelength, energy can be transferred from the donor to the acceptor, resulting in excitation of the acceptor and emission at the acceptor's emission wavelength. The efficiency of energy transfer, E, is a function of the Förster parameter, R_o, and of the distance between the donor and the acceptor, R:

$$E = [1 + (R/R_o)^6]^{-1} \quad (1)$$

Thus, if one quantifies E and R_o, one can determine R. With commonly used fluorescent probes, FRET permits accurate determination of distances in the range of ~20 to ~100 Å. Thus, FRET permits accurate determination of distances up to more than one-half the diameter of a transcription complex (diameter ~150 Å; Zhang et al., 1999; Cramer et al., 2001; Gnatt et al., 2001).

In our work, we use FRET to determine distances between donors and acceptors incorporated at specific sites in transcription complexes. By collecting data systematically—measuring large numbers of donor-acceptor distances (tens to hundreds of donor-acceptor distances)—we are able to define the relative spatial orientation of components of a transcription complex in solution, and thus to define the three-dimensional structure of a transcription complex in solution.

In this report, we describe work in which we have used FRET to measure large numbers of distances between fluorescent probes incorporated into bacterial RNAP core enzyme and DNA and complementary fluorescent probes incorporated into the bacterial initiation factor σ^{70}—both in the context of RNAP holoenzyme (composition $\beta'/\beta/\alpha^{I}/\alpha^{II}/\omega/\sigma^{70}$) and in the context of the RNAP-

[3]Correspondence: ebright@mbcl.rutgers.edu

[4]These authors made equal contributions to this work.

[5]Present address: Department of Chemistry, University of California, Los Angeles, California 90024

[6]Present address: Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138

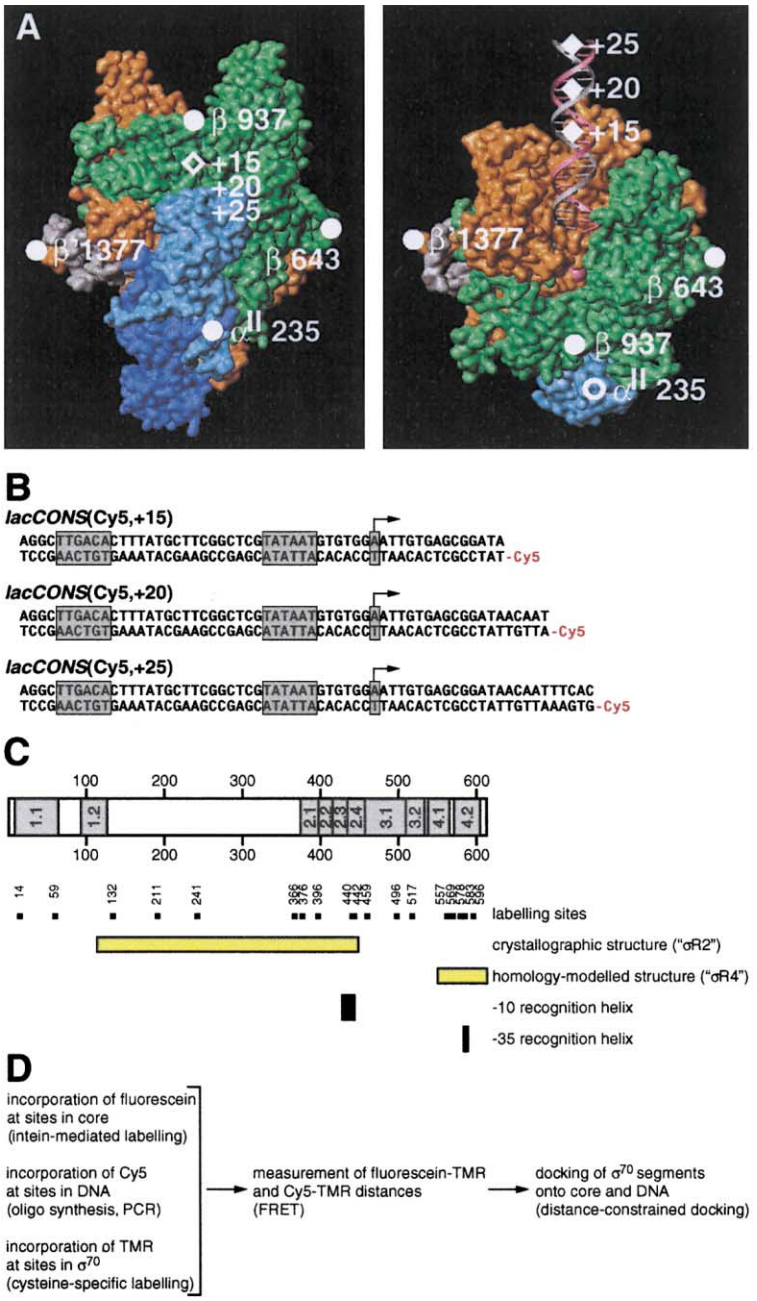

Figure 1. Strategy

(A) Probe sites within RNAP core and DNA. Two orthogonal views of the crystallographic structure of RNAP core (Zhang et al., 1999; Minakhin et al., 2001; PDB accession 1HQM), showing the location of positions +4 through +25 of DNA in RP_o ("downstream duplex"; Naryshkin et al., 2000; Ebright, 2000; see also Korzheva et al., 2000), and the locations of positions at which fluorescent probes were incorporated in this work (β' 1377, β 643, β 937, α^{II} 235, DNA +15, DNA +20, and DNA +25; filled circles and diamonds where visible; open circles and diamonds where not visible). Left, "upstream face." Right, "top face" (view directly into the active-center cleft, toward the active-center Mg^{2+} [magenta sphere at center]). β', β, α^{I}, α^{II}, and ω subunits are in orange, green, light blue, dark blue, and gray; DNA nontemplate and template strands are in pink and gray.

(B) Probe sites within DNA. Sequences of *lacCONS* (Mukhopadhyay et al., 2001) derivatives having fluorescent probe at +15, +20, or +25. Shaded boxes, transcription start site (with arrow), promoter −10 element, and promoter −35 element.

(C) Probe sites within σ^{70}. Map of σ^{70} showing conserved regions 1.1, 1.2, 2.1, 2.2, 2.3, 2.4, 3.1, 3.2, 4.1, and 4.2 (shaded boxes; Gross et al., 1998); σR2, for which a crystallographic structure is available (long yellow bar; Malhotra et al., 1996); σR4, for which a homology-modeled structure is available (short yellow bar; Baikalov et al., 1996; Lonetto et al., 1998); determinants for interaction with promoter −10 and −35 elements (solid bars; Gross et al., 1998); and sites at which fluorescent probe was incorporated in this work (small filled squares).

(D) Outline of experimental approach. TMR, tetramethylrhodamine.

promoter open complex (RP_o; composition $\beta'/\beta/\alpha^{I}/\alpha^{II}/\omega/\sigma^{70}$/DNA). The results define the structures of holoenzyme and RP_o in solution; reveal a striking difference in the position of σ^{70} region 1.1 in holoenzyme versus in RP_o, with region 1.1 being located deep within the active-center cleft in holoenzyme, but being located outside the active-center cleft in RP_o; and validate the method.

Strategy

Our analysis involved the following structural inputs: the crystallographic structure of RNAP core enzyme (Figure 1A; Zhang et al., 1999); the position, based on protein-DNA photocrosslinking, of the 25 bp DNA segment downstream of the RNAP active center in RP_o ("downstream DNA duplex"; Figure 1A, right; Naryshkin et al., 2000; Ebright, 2000; see also Korzheva et al., 2000); the crystallographic structure of a segment of σ^{70} spanning regions 1.2 through 2.4 and containing the α helix responsible for recognition of the promoter −10 element ("σ^{70} region 2," σR2; Figure 1C; Malhotra et al., 1996); and the homology-modeled structure of a segment of σ^{70} spanning regions 4.1 through 4.2 and containing the α helix responsible for recognition of the promoter −35 element ("σ^{70} region 4," σR4; Figure 1C; homology modeled based on the crystallographic structure of sequence homolog NarL; Baikalov et al., 1996; Lonetto et al., 1998).

Our analysis involved five steps (Figure 1D):

(1) incorporation of the fluorescent probe fluorescein at each of a series of sites within core;
(2) incorporation of the fluorescent probe Cy5 at each of a series of sites within DNA;
(3) incorporation of the fluorescent probe tetramethylrhodamine at each of a series of sites within σ^{70};
(4) measurement of fluorescein-tetramethylrhodamine and Cy5-tetramethylrhodamine distances (105 distinct distances for RP_o; 66 distinct distances for holoenzyme); and
(5) distance-constrained docking of structures of core, DNA, and segments of σ^{70}.

In step 1 of the analysis, we incorporated fluorescein at each of four sites within core: namely, residue 1377 of β', residue 643 of β, residue 937 of β, and residue 235 of α^{II} (Figure 1A). The four probe sites are well separated in the structure of core, are located at the periphery of core, and bracket the central portion of core and the active-center cleft (Figure 1A). The positions of the probe sites permit accurate three-dimensional triangulation of the position of a complementary probe in σ^{70} (or in any other component of a transcription complex) by measurement of distances to each of the sites and application of trigonometry, and permit particularly accurate three-dimensional triangulation of positions of complementary probes in the central portion of core and the active-center cleft. The four probe sites are in regions of core that are conformationally equivalent in crystallographic structures of bacterial RNAP core, eukaryotic RNAP II core, and the eukaryotic RNAP II elongation complex (Zhang et al., 1999; Ebright, 2000; Cramer et al., 2001; Gnatt et al., 2001), and thus are expected to be conformationally equivalent in the absence and presence of DNA. To incorporate probe at each site, we used intein-mediated C-terminal fluorescent labeling (Supplemental Figure S1 [see Supplemental Data, below]; Mukhopadhyay et al., 2001; see also Chong et al., 1996, 1997; Muir et al., 1998). To incorporate probe at sites in β' and α^{II}, we labeled subunit C termini in core derivatives assembled in vivo (Supplemental Figure S1). To incorporate probe at sites in β, we labeled subunit-fragment C termini in "split-subunit" core derivatives (Severinov et al., 1995, 1996; Naryshkin et al., 2000, 2001) reconstituted in vitro (Supplemental Figure S1). Control experiments verify that the resulting core derivatives are labeled site specifically, are labeled with high efficiency, and retain full function in transcription initiation and σ^{70} interactions.

In step 2 of the analysis, we incorporated Cy5 at each of three sites within downstream-duplex DNA: namely, positions +15, +20, and +25 (Figures 1A and 1B). In RP_o, these sites are contained within a DNA segment located deep within the RNAP active-center cleft, with a location, orientation, and rotational phasing precisely defined by results of site-specific protein-DNA photocrosslinking (Figure 1A, right; Naryshkin et al., 2000; Ebright, 2000: see also Korzheva et al., 2000). To prepare DNA derivatives, we performed solid-phase synthesis of labeled primers, followed by PCR. Control experiments verify that the resulting DNA derivatives are labeled site specifically, are labeled with high efficiency, and retain full function in transcription initiation.

In step 3 of the analysis, we incorporated tetramethylrhodamine at each of eighteen sites within σ^{70} (Figure 1C). Eight probe sites are contained within the segment of σ^{70} for which a high-resolution crystallographic structure is available (σR2; Figure 1C; Malhotra et al., 1996). Five probe sites are located within the segment of σ^{70} for which a homology model is available (σR4; Figure 1C; Baikalov et al., 1996; Lonetto et al., 1998). The remaining probe sites are in segments of σ^{70} for which no three-dimensional structural information has been reported (Figure 1C): two probe sites are in σ^{70} region 1.1 (σR1.1), two probe sites are in σ^{70} region 3.1 (σR3.1), and one probe site is in σ^{70} region 3.2 (σR3.2). We prepared eighteen σ^{70} derivatives, one labeled at each site. To prepare each σ^{70} derivative, we first prepared a σ^{70} derivative containing a single Cys residue at the site of interest, and then performed Cys-specific chemical modification with tetramethylrhodamine maleimide to introduce tetramethylrhodamine at the Cys residue (methods as in Mukhopadhyay et al., 2001). Control experiments indicate that 15 of the 18 resulting σ^{70} derivatives are fully functional in transcription initiation, and that all 18 are functional in core-σ^{70} interaction.

In step 4 of the analysis, for each combination of labeling site in core or DNA and functional labeling site in σ^{70}, we used FRET to measure the probe-probe distance—both in the context of RP_o and in the context of holoenzyme (Table 1). For RP_o, this involved measurement of 105 distances (with seven probe sites in core and DNA and fifteen functional probe sites in σ^{70}; Table 1, top panel). For holoenzyme, this involved measurement of 66 distances (with 4 probe sites in core and 18 functional probe sites in σ^{70}; Table 1, bottom panel). To minimize complications due to fluorescence arising from free components, from aggregates of components, and from nonproductive complexes, we performed experiments using RP_o and holoenzyme isolated by nondenaturing PAGE and analyzed in situ in gel slices (Supplemental Figure S2; Mukhopadhyay et al., 2001). In all experiments, energy transfer efficiencies decreased to $\sim$0 after addition of SDS to gel slices and incubation for 10 min at 37°C ("random-coil" control; Experimental Procedures, FRET), justifying the assumption that observed FRET efficiencies are dependent on the proximity of donor and acceptor within an intact, native three-dimensional structure. Steady-state fluorescence anisotropy measurements carried out in gel slices, and time-resolved fluorescence anisotropy measurements carried out in solution, indicate that probes reorient on the time scale of the probe life times (Experimental Procedures, Fluorescence Anisotropy), justifying the assumption that $\kappa^2 = 2/3$ in calculation of the Förster parameter, R_o, and thus justifying the assumption that FRET efficiencies can be interpreted exclusively in terms of mean distances, R (equations *1,4,* and *5*).

In step 5 of the analysis, we used an automated, objective distance-constrained docking algorithm to dock segments of σ^{70} onto the structures of core and DNA (Experimental Procedures, Distance-Constrained Docking). The initial phase of docking employed only geometric information—namely, the positions of probe sites in core, DNA, and σ^{70}, and the measured FRET distances

Table 1. Systematic FRET Data

RNAP-promoter open complex

σ region	σ residue	β'1377 E	β'1377 Ro (Å)	β'1377 R (Å)	β643 E	β643 Ro (Å)	β643 R (Å)	β937 E	β937 Ro (Å)	β937 R (Å)	αI/235 E	αI/235 Ro (Å)	αI/235 R (Å)	DNA +15 E	DNA +15 Ro (Å)	DNA +15 R (Å)	DNA +20 E	DNA +20 Ro (Å)	DNA +20 R (Å)	DNA +25 E	DNA +25 Ro (Å)	DNA +25 R (Å)
σR1.1	14	0.085	57.9	86	0.11	57.1	81	0.20	58.4	74	0.025	58.2	107	0.49	61.4	62	0.45	61.4	64	0.39	61.4	66
	59	0.11	57.7	82	0.14	56.9	77	0.22	58.2	72	0.02	58.0	111	0.78	61.4	50	0.60	61.4	57	0.48	61.4	62
σR2	132	0.10	57.8	83	0.06	57.0	91	0.19	58.4	75	0.02	58.2	111	0.69	61.4	54	0.36	61.4	68	0.17	61.4	80
	211	0.11	57.8	82	0.07	57.1	88	0.65	58.4	53	0.01	58.2	125	0.08	61.4	92	0.07	61.4	95	0.04	61.4	104
	241	0.12	57.9	81	0.04	57.2	95	0.20	58.5	74	0.02	58.3	112	0.40	61.4	66	0.26	61.4	73	0.19	61.4	78
	366	0.25	57.7	69	0.07	57.0	89	0.35	58.3	64	0.02	58.1	111	0.45	61.4	64	0.33	61.4	69	0.13	61.4	84
	376	0.22	57.6	71	0.07	56.9	88	0.42	58.2	61	0.02	58.0	111	0.45	61.4	64	0.24	61.4	74	0.17	61.4	80
	396	0.25	57.8	69	0.34	57.0	64	0.71	58.4	50	0.07	58.2	91	0.44	61.4	64	0.31	61.4	70	0.13	61.4	84
	440	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
	442	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
σR3.1	459	0.20	57.9	73	0.18	57.2	74	>0.99	58.5	<27	0.07	58.3	90	0.21	61.4	77	0.13	61.4	84	0.12	61.4	86
	496	0.32	57.8	66	0.13	57.0	78	>0.99	58.3	<27	0.14	58.2	79	0.19	61.4	78	0.11	61.4	87	0.08	61.4	92
σR3.2	517	0.46	57.9	60	0.31	57.2	65	0.98	58.5	31	0.29	58.3	68	0.34	61.4	69	0.19	61.4	78	0.08	61.4	92
σR4	557	0.65	58.0	52	0.07	57.1	89	0.48	58.5	59	0.11	58.4	82	0.09	61.4	90	0.06	61.4	97	0.03	61.4	110
	569	0.54	57.9	56	0.09	57.1	83	0.71	58.4	50	0.12	58.2	81	0.14	61.4	83	0.07	61.4	95	0.01	61.4	>125
	578	0.50	57.9	58	0.08	57.1	85	0.47	58.4	60	0.12	58.2	81	0.16	61.4	81	0.04	61.4	104	0.02	61.4	118
	583	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
	596	0.37	57.9	63	0.05	57.2	94	0.67	58.5	52	0.29	58.3	68	0.05	61.4	100	0.03	61.4	109	<0.01	61.4	>125

RNAP holoenzyme

σ region	σ residue	β'1377 E	β'1377 Ro (Å)	β'1377 R (Å)	β643 E	β643 Ro (Å)	β643 R (Å)	β937 E	β937 Ro (Å)	β937 R (Å)	αI/235 E	αI/235 Ro (Å)	αI/235 R (Å)
σR1.1	14	0.23	57.9	71	0.41	57.1	61	0.37	58.4	64	0.07	58.2	90
	59	0.34	57.7	64	0.35	56.9	63	0.41	58.2	62	0.10	58.0	84
σR2	132	0.14	57.8	78	0.06	57.0	90	0.19	58.4	74	0.02	58.2	111
	211	0.09	57.8	85	0.11	57.1	80	0.72	58.4	50	<0.01	58.2	>125
	241	0.14	57.9	78	0.04	57.2	95	0.20	58.5	74	0.03	58.3	104
	366	0.33	57.7	65	0.08	57.0	86	0.37	58.3	64	0.03	58.1	104
	376	0.28	57.6	67	0.08	56.9	86	0.35	58.2	64	0.03	58.0	104
	396	0.30	57.8	67	0.34	57.0	64	0.74	58.4	49	0.05	58.2	95
	440	0.30	57.7	66	ND	ND	ND	ND	ND	ND	0.05	58.0	95
	442	0.37	57.9	63	ND	ND	ND	ND	ND	ND	0.06	58.3	92
σR3.1	459	0.29	57.5	67	0.14	56.7	78	>0.99	58.5	<27	0.06	57.8	91
	496	0.45	57.8	60	0.12	57.0	79	>0.99	58.3	<27	0.14	58.2	79
σR3.2	517	0.44	57.9	60	0.28	57.2	67	0.94	58.5	37	0.24	58.3	71
σR4	557	0.54	57.9	56	0.09	57.2	84	0.48	58.5	59	0.10	58.2	84
	569	0.45	57.9	60	0.10	57.1	82	0.65	58.4	53	0.12	58.2	81
	578	0.40	57.9	62	0.12	57.1	79	0.53	58.4	57	0.11	58.2	82
	583	0.28	57.9	68	ND	ND	ND	ND	ND	ND	0.06	58.2	92
	596	0.43	57.9	61	0.05	57.2	94	0.74	58.5	49	0.26	58.3	69

Values of R in holoenzyme (bottom panel) that differ by ≥10 Å from values of R in RP_o (top panel) are indicated in red. Values of R have precision of ~±2.5% for 30 Å < R < 110 Å, and ~±5-10% for 30 Å ≥ R ≥ 110 Å. In the top panel, data are not presented for RNAP derivatives defective in formation of RP_o (i.e., RNAP derivatives labeled at residues 440, 442, and 583 of σ^{70}; Experimental Procedures, RNAP Holoenzyme). In the bottom panel, data for [β′(1-1377)-F]-RNAP and [[Alα45]α^{II}(1-235)-F]-RNAP are means of data for holoenzyme prepared directly and prepared by salt-dissociation of RP_o; data for [β(1-643)-F/β(643-1342)]-RNAP and [β(1-937)-F/β(938-1342)]-RNAP are for holoenzyme prepared by salt dissociation of RP_o (Experimental Procedures, Core-σ^{70} FRET, RNAP Holoenzyme).

(or corresponding FRET efficiencies). The initial phase of docking involved a grid search (8,000-point grid) to define coarse translations and rotations of each segment of σ^{70} relative to the structures of core and DNA that best fit the FRET constraints, followed by Markov-chain Monte Carlo sampling (Metropolis et al., 1953) of 100,000 trial configurations centered on specific grid points to define fine translations and rotations of the segment of σ^{70} relative to the structures of core and DNA that best fit the FRET constraints. σR2 and σR4 were modeled using three-dimensional structures containing probe sites; σR1.1, σR3.1, and σR3.2 were modeled as spheres containing probe sites. For σR2 and σR4, the model from the initial phase of docking was refined in a second, "postprocessing" phase, involving Markov-chain Monte Carlo simulations employing both FRET constraints and steric constraints.

Results and Discussion

Structural Organization of the RNAP-Promoter Open Complex

The results of systematic FRET and distance-constrained docking define both the translational and the rotational orientations of σR2 and σR4 relative to core and downstream-duplex DNA in RP_o (42 distance constraints for σR2; 28 distance constraints for σR4; Table 1, top panel; Figure 2). In addition, the results define the translational orientations of σR1.1, σR3.1, and σR3.2 relative to core and downstream-duplex DNA in RP_o (14 distance constraints for σR1.1; 14 distance constraints for σR3.1; 7 distance constraints for σR3.2; Table 1, top panel; Figure 2). The positions of σ^{70} segments are essentially similar in models from distance-constrained-docking runs employing only FRET constraints ("FRET-only model"; Supplemental Figure S3A), employing both FRET constraints and steric constraints ("FRET+steric model"; Supplemental Figure S3B), or employing both FRET constraints and steric constraints and employing a structure of core in which the "β′ pincer" of the crab-claw-shaped core molecule is rotated 16° into the active-center-cleft channel—as observed in the crystallographic structure of a eukaryotic RNAP II tailed-template elongation complex (Gnatt et al., 2001), and as proposed for bacterial RP_o (Korzheva et al., 2000; Ebright, 2000) ("FRET+steric, closed-claw model"; Figure 2; Supplemental Figure S3C). However, the calculated σR2-core interaction energy is significantly more favorable in the

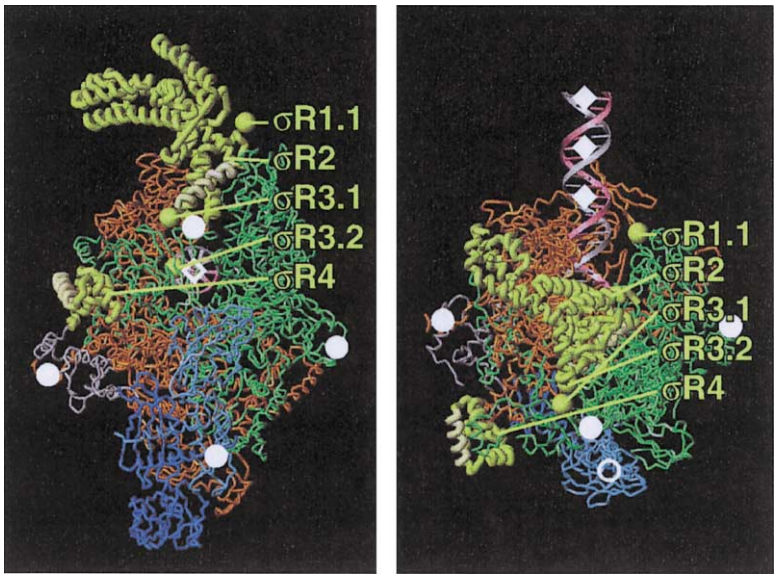

Figure 2. Structural Organization of the RNAP-Promoter Open Complex: Model

Model for the structural organization of RP_o from distance-constrained-docking run employing both FRET constraints and steric constraints, and using structure of core having the β′ pincer rotated 16° into the active-center cleft—as observed in the crystallographic structure of a eukaryotic RNAP II tailed-template elongation complex (Gnatt et al., 2001), and as proposed for bacterial RP_o (Korzheva et al., 2000; Ebright, 2000) (FRET+steric, closed-claw model). View orientations, colors of core subunits and downstream-duplex DNA, and symbols for probe sites are as in Figure 1A. σR2 and σR4 are shown as yellow ribbons, with the α helices that mediate recognition of the promoter −10 element (region 2.3/region 2.4 α helix) and −35 element (region 4.2 α helix) highlighted in light yellow. σR1.1, σR3.1, and σR3.2 are shown as yellow spheres. (Supplemental Figure S3 presents models from distance-constrained-docking runs employing only FRET constraints [FRET-only model; Supplemental Figure S3A] or employing both FRET constraints and steric constraints, and using an unmodified structure of core [FRET+steric model; Supplemental Figure S3B].)

FRET+steric, closed-claw model (Figure 2) than in the other models (differences = 1,000 kcal/mol and 700 kcal/mol; values calculated following all-atoms energy minimization). We offer the FRET+steric, closed-claw model (Figure 2) as a working model for the structural organization of core, downstream-duplex DNA, and σ^{70} in RP_o.

The model is well-defined; thus, the top solutions from five independent distance-constrained-docking runs superimpose with mean rmsd in positions of centers of mass of σ^{70} segments (mean $rmsd_c$) of 0.9 Å, and with mean rmsd in positions of Cα atoms of σR2 and σR4 (mean $rmsd_{C\alpha}$) of 1.6 Å. The model also is robust; thus, the top solutions from five independent runs, each omitting a different randomly selected 10% of distance constraints for each σ^{70} segment, superimpose with mean $rmsd_c$ of 6.0 Å and mean $rmsd_{C\alpha}$ of 9.8 Å. The model also is relatively insensitive to values used for adjustable parameters in the distance-constrained-docking run; thus, the top solutions from parallel runs using higher or lower values for each adjustable parameter superimpose with mean $rmsd_c$ of 0.8–2.8 Å and mean $rmsd_{C\alpha}$ of 1.2–3.0 Å (Experimental Procedures, Distance-Constrained Docking).

In the model, segments of σ^{70} are arranged on core in a linear, sequential fashion—with σR1.1 near the σR2 N terminus, the σR2 C terminus near σR3.1, σR3.1 near σR3.2, and σR3.2 near σR4 (Figures 2 and 3). As predicted from results of affinity-cleaving, protein-protein-interaction, and genetic studies (Gross et al., 1998; Owens et al., 1998b; Traviglia et al., 1999; Sharp et al., 1999; Gruber et al., 2001), the interface between σ^{70} and core is extensive, involving interactions with the β′ pincer, the β′ rudder, the β pincer, and the β flap (Figures 2 and 3). σR1.1 is located outside the active-center cleft, and is positioned to interact with the tip of lobe 1 of the β pincer. σR2 is located above and within the active-center cleft, "capping" the active-center cleft, analogously to a champagne cork in a champagne bottle. The conserved part of σR2 (σ^{70} regions 1.2 and 2.1–2.4) is located within the active-center cleft, and is positioned to interact with the β′ pincer, the β′ rudder, and lobes 1 and 2 of the β pincer. The nonconserved part of σR2 (the connector between σ^{70} regions 1.2 and 2.1) is located above the active-center cleft, and is positioned to interact with the tip of the β′ pincer. σR3.1 and σR3.2 are located near the β flap, with σR3.1 located above the edge of the flap, and with σR3.2 located below the edge of the flap, wedged between the β′ pincer and the flap, directly within the RNA exit channel (see below). σR4 is located at the extreme tip of the β flap, and is positioned to interact with β′ residues 393–402 and with the flap tip.

The model accommodates—without adjustment, or with only minor adjustment—the paths proposed by Naryshkin et al. (2000) for nucleotides +3 to −5 and −12 to −40 of the nontemplate strand of promoter DNA and nucleotides +3 to −40 of the template strand of promoter DNA (Figure 4). The probable path of nucleotides −6 to −11 of the nontemplate strand can be identified as the shortest sterically permitted route connecting the positions of nucleotides −5 and −12 of the nontemplate strand (Figure 4).

The modeled presence of σR2 within the active-center cleft, and of σR3.1 at the edge of the active-center cleft, restricts access to the active-center cleft to five distinct channels (Figures 3 and 4): (1) a channel corresponding to the path of the downstream duplex (downstream-duplex channel; defined by the β′ pincer, the β′ central mass, and lobe 1 of the β pincer); (2) a channel corresponding to the path of the nontemplate strand of the transcription bubble proposed in Naryshkin et al. (2000) (nontemplate-strand channel; defined by σ^{70} regions 1.2, 2.1, and 2.3 and lobes 1 and 2 of the β pincer); (3) a

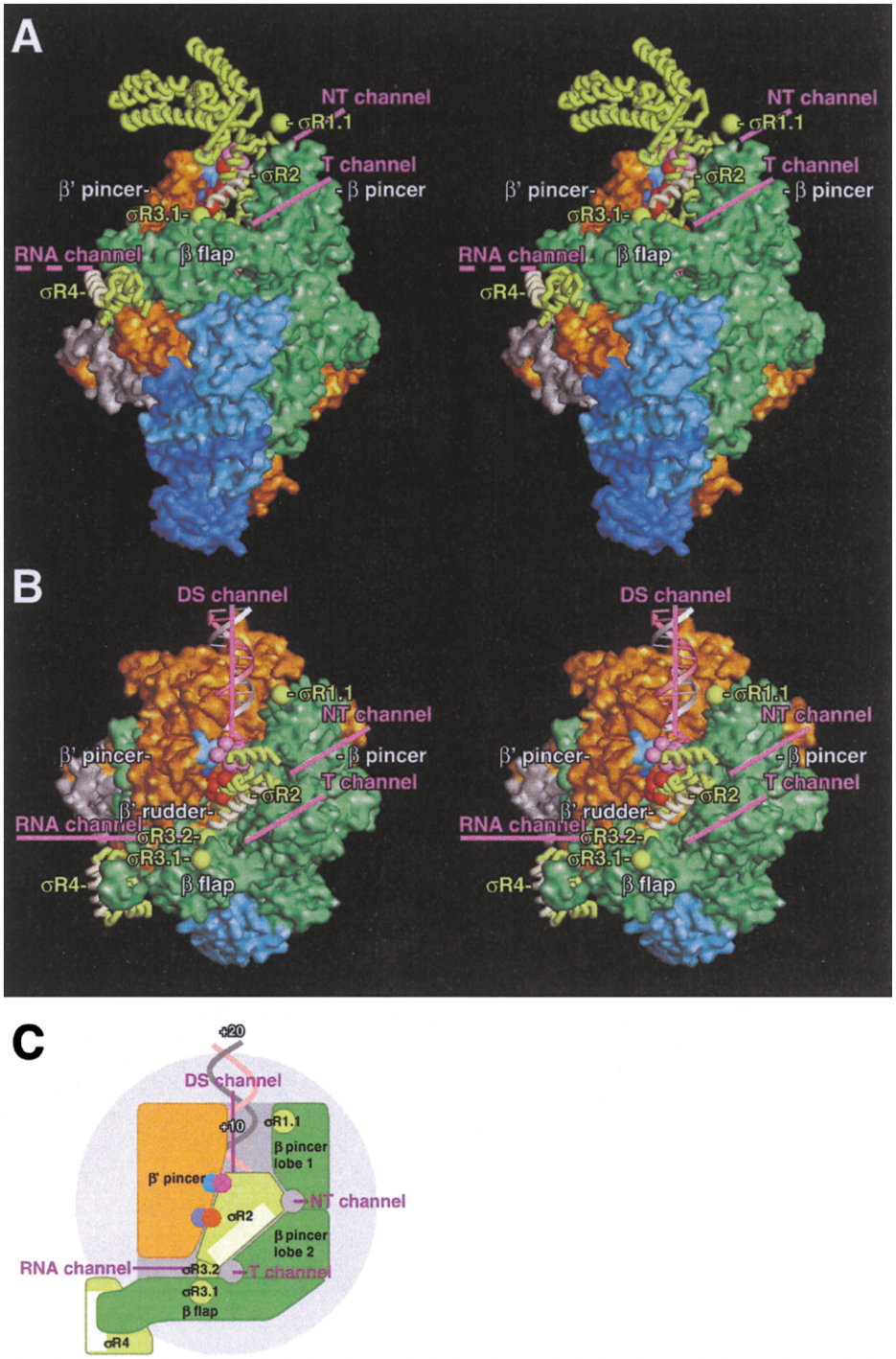

Figure 3. Structural Organization of the RNAP-Promoter Open Complex: σ^{70}-Core Interactions

Stereoviews and schematic view of model for the structural organization of RP_o (FRET+steric, closed-claw model; Figure 2). σR2 and σR4 are shown as yellow ribbons, with the α helices that mediate recognition of the promoter −10 element (region 2.3/region 2.4 α helix) and −35 element (region 4.2 α helix) highlighted in light yellow. σR1.1, σR3.1, and σR3.2 are shown as yellow spheres. Core is shown as a van der Waals surface (colors of core subunits as in Figures 1A and 2). Nontemplate and template strands of downstream-duplex DNA are shown as pink and gray ribbons. Residues 373–381 within σ^{70} region 2.1 and residues 288–294 within the β′ pincer are shown in magenta and light blue (see text). Residues 402, 403, 406, 407, 409, and 413 within σ^{70} region 2.2 and residues 275, 297, and 302 within the β′ pincer are shown in red and dark blue (see text). Residues 316–320 within the β′ rudder (nonconserved residues at the tip of the β′ rudder) are omitted for clarity. Locations of the nontemplate-strand (NT), template-strand (T), downstream-duplex (DS), and RNA-exit channels (RNA) are indicated; the secondary channel (through which NTPs enter the active-center cleft; Korzheva et al., 2000; Ebright, 2000) is not visible in these views.

(A) Stereoview of upstream face of RP_o (view as in Figure 2, left).

(B) Stereoview of active-center cleft of RP_o (view as in Figure 2, right; nonconserved part of σR2 [connector between regions 1.2 and 2.1; Figure 1C] omitted for clarity).

(C) Schematic view of active-center cleft of RP_o (view as in [B]; outline of RNAP in light gray; floor of active-center cleft in medium gray).

channel corresponding to the path of the template strand of the transcription bubble proposed in Naryshkin et al. (2000) (template-strand channel; defined by σ^{70} regions 2.2 and 2.4, σR3.1, σR3.2, lobe 2 of the β pincer, and the β flap); (4) the RNA exit channel (through which the RNA 5′ end exits; Naryshkin et al., 2000; Korzheva et al., 2000; Ebright, 2000); and (5) the secondary channel (through which NTPs enter; Korzheva et al., 2000; Ebright, 2000). The nontemplate-strand and template-strand channels are particularly noteworthy. σR2, together with β lobes 1 and 2, nearly completely encloses the proposed path of nucleotides -2 to -5 of the nontemplate strand, creating a $\sim$10 Å wide tunnel; σ^{70} regions 2.1 and 2.3 provide a strip of positively charged residues lining one wall of this tunnel (residues 385, 392, 393, and 426), potentially facilitating interaction with the negatively charged nontemplate strand. σR2, together with σR3.1, σR3.2, β lobe 2, and the β flap, nearly completely encloses the proposed path of nucleotides -8 to -10 of the template strand, creating a $\sim$12 Å wide tunnel; σ^{70} region 2.2 provides a positively charged residue on one wall of this tunnel (residue 397), potentially facilitating interaction with the template strand. The modeled threading of the nontemplate and template strands through enclosed channels in RP_o implies that, to permit entry of the nontemplate and template strands during formation of RP_o, σR2 must at least temporarily be displaced, and/or the β' pincer and σR2 must at least transiently rotate to yield a more open claw conformation (analogous to the conformation in core; Zhang et al., 1999).

The model provides a structural rationalization for biochemical and genetic results defining σ^{70}-core interactions in RP_o. Affinity-cleaving results have defined regions of core proximal to specific residues of σ^{70} (Owens et al., 1998b; Traviglia et al., 1999); the model is consistent, in detail, with these results. A peptide corresponding to σ^{70} region 2.1 has been shown to interact with core (Lesley and Burgess, 1989); in the model, residues 373–381 within region 2.1 are positioned adjacent to residues 288–294 of β' (Figure 3; residues numbered as in *E. coli* RNAP). σ^{70} region 2.2 has been shown to interact with a coiled coil within the β' pincer, and residues 402, 403, 406, 407, 409, and 413 within region 2.2 and residues 275, 295, and 302 within the β' coiled coil have been shown to be involved in this interaction (Arthur et al., 2000; Arthur and Burgess, 1998: Sharp et al., 1999; Young et al., 2001); in the model, the α helix comprising σ^{70} region 2.2 is positioned adjacent to the two α helices comprising the β' coiled coil, and σ^{70} residues 402, 403, 406, 407, 409, and 413 are positioned adjacent to β' residues 275, 295, and 302 (Figure 3). σR4 has been shown to interact with the β flap (Kuznedelov et al., 2002); in the model, the primary interactions of σR4 with core involve the flap tip (Figure 3). Residues 555, 562, and 565 within σ^{70} region 4.1 have been shown to be important for interaction with core (Sharp et al., 1999); in the model, these residues are adjacent to an α helix comprised of β' residues 393–402.

The model also provides a structural rationalization for biochemical and genetic results defining σ^{70}-DNA interactions in RP_o. Affinity-cleaving, crosslinking, and genetic results have defined residues of promoter DNA proximal to specific segments and residues of σ^{70} (Gross et al., 1998; Owens et al., 1998a; Bown et al., 1999; Naryshkin et al., 2000); the model is consistent, in detail, with these results. σ^{70} regions 2.3 and 2.4, which comprise an aromatic-rich loop followed by an α helix, have been shown to mediate sequence-specific recognition of the promoter -10 element (Gross et al., 1998), with residues 421–434 within region 2.3 proposed to interact with nucleotides -7 to -11 of the nontemplate strand (as single-stranded DNA; Juang and Helmann, 1994; Aiyar et al., 1994; Marr and Roberts, 1997; Huang et al., 1997; Callaci and Heyduk, 1998; Fenton et al., 2000), and with residues 434–440 within regions 2.3 and 2.4 proposed to interact with base pairs -12 and -13 (as double-stranded DNA; Daniels et al., 1990; Waldburger et al., 1990; Fenton et al., 2000). In the model, regions 2.3 and 2.4 are prominently exposed on the molecular surface (Figures 3 and 4). The region 2.3 loop forms part of the mouth of the nontemplate-strand channel, with residues 421–426 located adjacent to the proposed positions of nucleotides -4 to -9 of the nontemplate strand, and the region 2.3/region 2.4 α helix extends away from the mouth of the nontemplate-strand channel, with residues 437 and 440 located adjacent to the proposed positions of base pairs -12 and -13 (Figure 4). σR3.1 has been shown to mediate sequence-specific recognition of the promoter extended -10 element, a discrete promoter element located immediately upstream of the -10 element (base pairs -14 to -17; Gross et al., 1998). In the model, σR3.1 is located adjacent to the proposed position of the major groove of base pairs -14 to -19 (Figure 4). σR4 has been shown to mediate sequence-specific interaction of the promoter -35 element (base pairs -30 to -35), with the region 4.2 α helix proposed to interact with the -35 element major groove, and with residues immediately C-terminal to the region 4.2 α helix proposed to interact with transcriptional activators (Gross et al., 1998). In the model, the region 4.2 α helix and residues immediately C-terminal are prominently exposed on the molecular surface, and the region 4.2 α helix is located adjacent to the proposed position of the major groove of base pairs -30 to -35 (Figure 4). The modeled distance between the centers of the region 2.3/region 2.4 α helix (responsible for recognition of the $-12/-13$ region) and the region 4.2 α helix (responsible for recognition of the -35 region) is 72 Å—a distance that can be spanned by 20 bp of canonical B form DNA (-12 to -32; 70 Å) positioned essentially as proposed in Naryshkin et al. (2000) (Figure 4).

The model also provides a structural rationalization for the observation that the affinity of σ^{70} for the remainder of the transcription complex decreases upon synthesis of $\geq$9–11 nt of RNA and transition from initiation to elongation (Gross et al., 1998). In the model, σR3.2 is located squarely within the RNA exit channel (Figures 3B, 3C, and 5), with σR3.2 apparently serving as a molecular mimic or molecular placeholder for RNA. (Like RNA, σR3.2 is highly negatively charged [net charge -5].) The modeled position of σR3.2 would result in steric clash between σR3.2 and a $\geq$9–11 nt RNA product. We suggest that, upon synthesis of a $\geq$9–11 nt RNA product, RNA ejects σR3.2 from the RNA exit channel, permitting threading of RNA into and through the exit channel, and we suggest that the loss of σR3.2-core interactions upon

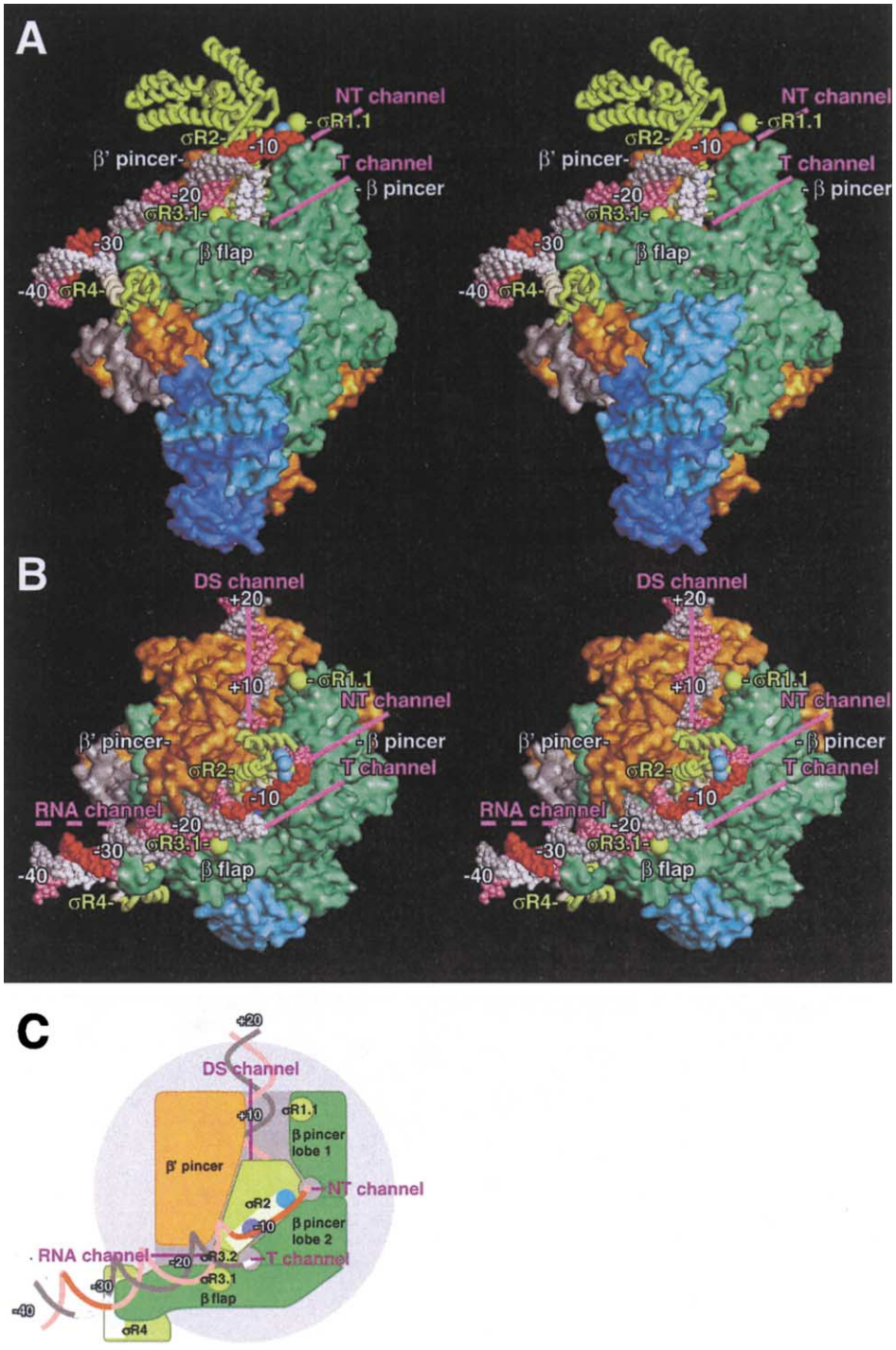

Figure 4. Structural Organization of the RNAP-Promoter Open Complex: σ^{70}-DNA Interactions

Stereoviews and schematic view of model for the structural organization of RP_o, showing the inferred path of promoter DNA. View orientations, representations of σ^{70} segments and core subunits, and colors of σ^{70} segments and core subunits are as in Figure 3. The paths of nucleotides +3 to −5 and −12 to −40 of the nontemplate strand of promoter DNA and nucleotides +3 to −40 of the template strand of promoter DNA are modeled essentially as in Naryshkin et al. (2000); the path of nucleotides −6 to −11 of the nontemplate strand is modeled along the shortest sterically permitted route connecting nucleotides −5 and −12. The nontemplate strand is shown in pink, with the promoter −10 and −35 elements highlighted in red. The template strand is shown in gray, with the promoter −10 and −35 elements highlighted in white. Residues 421–426 within σ^{70} region 2.3 and residues 437 and 440 within σ^{70} region 2.4 are shown in, respectively, light blue and dark blue (see text).

(A) Stereoview of upstream face of RP_o.

(B) Stereoview of active-center cleft of RP_o.

(C) Schematic view of active-center cleft of RP_o (outline of RNAP in light gray; floor of active-center cleft in medium gray).

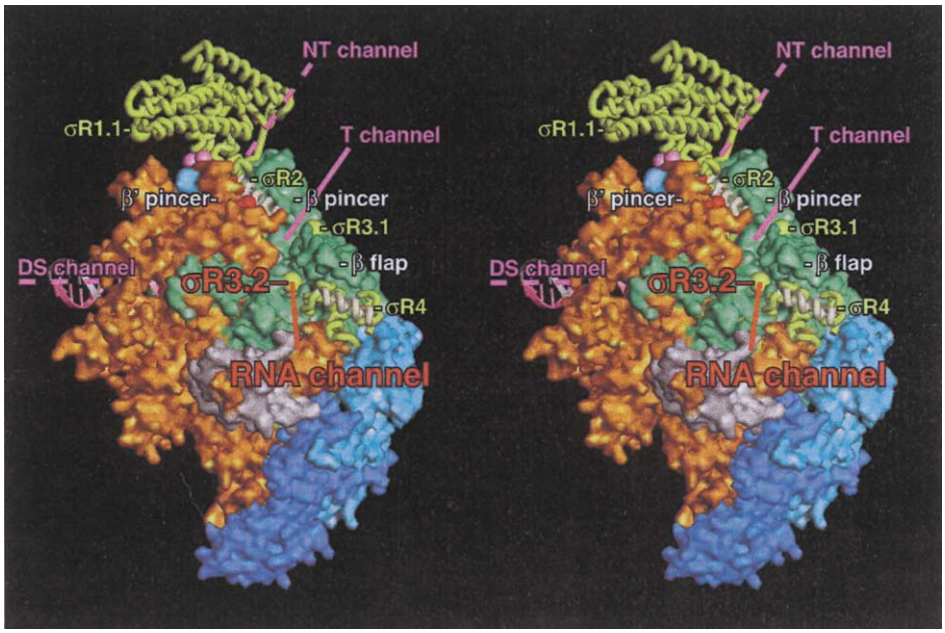

Figure 5. Structural Organization of RNAP-Promoter Open Complex: σ^{70}-(RNA Exit Channel) Interactions

Stereoview of model for the structural organization of RP_o, showing presence of σR3.2 within the RNA exit channel (red annotation). View rotated 90° on *y* relative to Figure 3A.

the ejection of σR3.2 accounts, at least in part, for the decrease in affinity of σ^{70} for the remainder of the transcription complex. This interpretation receives strong support from results indicating that σ^{70} and an RNA oligomer of length ≥9–11 nt compete for interaction with the remainder of the transcription complex (Daube and von Hippel, 1999) and, especially, from results indicating that the position of σR3.2 relative to the remainder of the transcription complex (Marr et al., 2001), but not the positions of other σ^{70} segments relative to the remainder of the transcription complex (Marr et al., 2001; Mukhopadhyay et al., 2001), differs in RP_o and in transcription complexes containing ≥9–11 nt of RNA.

Structural Organization of RNAP Holoenzyme

We have measured 66 σ^{70}-core distances within holoenzyme (Table 1, bottom). All distances for σR2, σR3.1, σR3.2, and σR4 in holoenzyme are identical within 8 Å—generally within 3 Å—to corresponding distances in RP_o (black in Table 1, bottom panel). In contrast, all distances for σR1.1 in holoenzyme are different by at least 10 Å—generally by at least 15 Å—from corresponding distances in RP_o (red in Table 1, bottom panel). We conclude that the positions of σR2, σR3.1, σR3.2, and σR4 in holoenzyme are similar to those in RP_o, but that the position of σR1.1 in holoenzyme is different from that in RP_o (Figure 6A).

The modeled positions of σR1.1 in holoenzyme and in RP_o differ by a full 51 Å, which corresponds to one-third to one-half the diameter of core (red sphere and yellow open circle in Figure 6A). In holoenzyme, σR1.1 is located deep within the active-center cleft, just above the floor of the downstream-duplex channel, and is positioned to interact with the floor and walls of the downstream-duplex channel (β′ pincer, β′ central mass, and lobe 1 of β pincer; red sphere in Figure 6A). In contrast, as described in the preceding section, in RP_o, σR1.1 is located outside the active-center cleft, and is positioned to interact with the tip of lobe 1 of the β pincer (yellow open circle in Figure 6A).

σR1.1 is extremely highly negatively charged (net charge −13). The location of σR1.1 in holoenzyme (the downstream-duplex channel) is highly positively charged. The location of σR1.1 in RP_o (lobe 1 of the β pincer) also is highly positively charged. We infer that σR1.1 makes alternative electrostatic interactions: with the downstream-duplex channel in holoenzyme, or with lobe 1 of the β pincer in RP_o.

The location of σR1.1 in holoenzyme corresponds to the location of downstream-duplex DNA in RP_o—specifically, the location of base pair +9 of downstream-duplex DNA (Figure 6B). We propose that the extremely highly negatively charged σR1.1 serves as a molecular mimic or molecular placeholder for downstream-duplex DNA. We propose that σR1.1 occupies the downstream-duplex channel in holoenzyme, and must be displaced, out of the active-center cleft, to lobe 1 of the β pincer, to permit formation of RP_o. The proposal that σR1.1 must be displaced to permit formation of RP_o provides a structural rationalization for the observation that deletion of σR1.1 affects kinetics of formation, but not stability, of RP_o (Wilson and Dombroski, 1997; Vuthoori et al., 2001). The proposals that σR1.1 is a molecular mimic of double-stranded DNA, and that σR1.1 must be displaced to permit formation of RP_o, are reminiscent of the observations that the $dTAF_{II}230(1–81)$ component of eukaryotic transcription factor IID is a molecular mimic of double-stranded DNA, and that $dTAF_{II}230(1–81)$ must be displaced to permit formation of the IID-DNA complex (Liu et al., 1998; Burley and Roeder, 1998). As has been

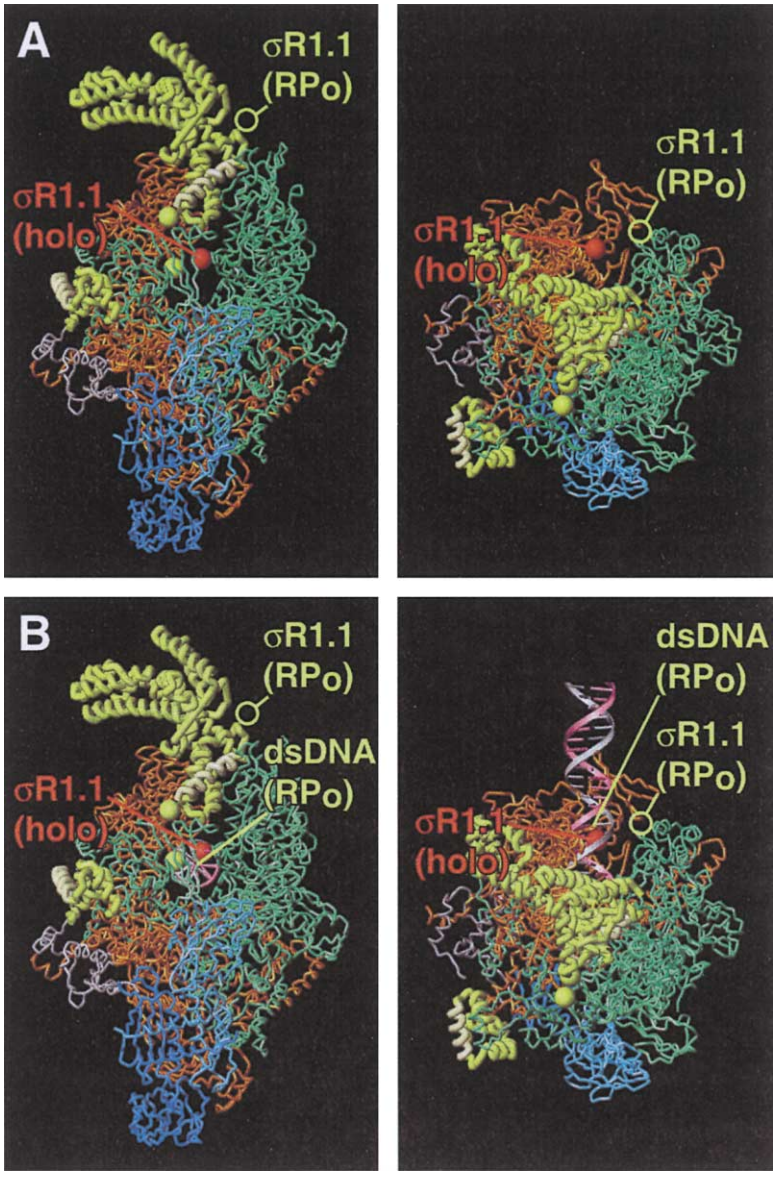

Figure 6. Structural Organization of RNAP Holoenzyme

(A) Model for the structural organization of holoenzyme. σR2, σR3.1, σR3.2, and σR4 (yellow ribbons and spheres) are positioned as in the model for RP_o (Figure 2). σR1.1 (red sphere) is positioned within the downstream-duplex channel, based on distance-constrained docking. The location of σR1.1 in RP_o (yellow open circle), 51 Å away, at the tip of lobe 1 of the β pincer, is shown for reference. (The line relating modeled positions of σR1.1 in holoenzyme and in RP_o does not intersect any probe site in core [cf. Figure 2]. Therefore, the difference in modeled positions of σR1.1 in holoenzyme and in RP_o [51 Å] exceeds the difference in each measured σR1.1-core distance [10–27 Å; Table 1]).

(B) As (A), but with downstream-duplex DNA from RP_o superimposed (pink and gray ribbons; cf. Figure 2) to document overlap between positions of σR1.1 in holoenzyme and downstream-duplex DNA in RP_o.

noted in the case of $dTAF_{II}230(1–81)$ (Liu et al., 1998; Burley and Roeder, 1998), a requirement that a molecular mimic of double-stranded DNA be displaced to permit complex formation provides an obvious target for regulation of complex formation by activators and repressors.

In free σ^{70}, σR1.1 interacts with DNA binding determinants of σR2 and σR4 and inhibits DNA binding (possibly also via molecular mimicry of double-stranded DNA; Gross et al., 1998). During formation of holoenzyme, σR1.1 transiently interacts with the β flap tip (Gruber et al., 2001). Our finding that σR1.1 interacts with the downstream-duplex channel in holoenzyme and with lobe 1 of the β pincer in RP_o (sites located >30 Å from DNA binding determinants of σR2, >75 Å from DNA binding determinants of σR4, >70 Å from the β flap tip, and 51 Å from each other) underscores the remarkable structural dynamism of σR1.1, with distinct protein-protein interactions by σR1.1 occurring in, and possibly coordinating, distinct steps in transcription.

Prospect

In this work, we have developed a method that enables us to leverage available crystallographic information in order to define three-dimensional structures of transcription complexes in solution, and we have applied the method to define the three-dimensional structures of RP_o and holoenzyme in solution. In further work, we have shown that the method can be modified to permit monitoring of structural changes in protein and nucleic acid during transcription (Mukhopadhyay et al., 2001), to permit measurement of kinetics of structural changes (by used of stopped-flow spectroscopy; J.M. and R.H.E., unpublished data), and to permit single-molecule measurement of the kinetics of structural changes (by use of confocal optical microscopy; S. Weiss, A.N.K., and R.H.E., unpublished data). With the method now established, and with labeled core and σ^{70} derivatives now in hand, detailed structural and mechanistic analysis of each step in transcription initiation, elongation, editing, pausing, and termination should be possible.

We note that the labeling, FRET, and docking procedures described in this report should be generalizable to permit analysis of any nanometer-scale multiprotein or nucleoprotein complex for which there exists (1) structural information for components and (2) procedures for reconstitution of the complex from components.

Experimental Procedures

5-(Amino-acetamido)-fluorescein HCl

Into 5-iodoacetamido fluorescein (Molecular Probes, Inc.; 200 mg; 0.388 mmol) in 10 ml DMF was added 15 ml concentrated NH_4OH. The solution was stirred for 1 hr at 25°C in the dark, and then evaporated under vacuum. The resulting semisolid was suspended in 10 ml water, the pH was adjusted to 1.5–2 by addition of 1% HCl, the suspension was cooled, and the product was collected by centrifugation and dried over high vacuum. Yield: 162 mg, 95%. (See Supplemental Figure S1A.)

5-(N^{α}-trityl-S-trityl-L-cysteinylamido-acetamido)-fluorescein

5-(Amino-acetamido)-fluorescein HCl (100 mg; 0.23 mmol), N^{α}-trityl-S-trityl-L-cysteine N-hydroxysuccinimide ester (Novabiochem, Inc.; 240 mg, 0.34 mmol), and triethylamine (0.1 ml) were dissolved in 10 ml anhydrous dimethylformamide and stirred 6 hr at 25°C in the dark. The reaction was terminated by addition of 1 ml water and stirring for 20 min at 25°C in the dark. The reaction mixture was evaporated under vacuum, and the product purified by flash chromatography (silica gel, 200-400 mesh; 5% [v/v] NH_4OH in EtOH; R_f ~0.8). Yield: 227 mg, 100%. (See Supplemental Figure S1A.)

5-(L-cysteinylamido-acetamido)-fluorescein (Cys-F)

Into 5-(N^{α}-trityl-S-trityl-L-cysteinylamido-acetamido)-fluorescein (227 mg; 0.23 mmol) in 3 ml freshly prepared 10% (v/v) triethylsilane in dichloromethane was added 3 ml trifluoroacetic acid. The solution was stirred 30 min at 25°C in the dark, and then evaporated under vacuum. The resulting solid was dissolved in 5 ml methanol, and the solution was evaporated under vacuum. The product was triturated with 30 ml ether and purified by flash chromatography (4:1:1:1 [v:v] ethyl acetate/ethanol/acetic acid/water; R_f ~0.6). Yield: 62 mg, 53%. (See Supplemental Figure S1A.)

Plasmids

Plasmid pGEMD(-Cys), encoding a σ^{70} derivative with no Cys residues, and plasmids encoding σ^{70} derivatives with single Cys residues at positions 59, 132, 366, 396, 459, 496, 517, 569, 578, 583, and 596 were described in Owens et al. (1998b), Callaci et al. (1998), Bown et al. (1999), and Mukhopadhyay et al. (2001). Plasmids encoding σ^{70} derivatives with single Cys residues at positions 14, 211, 241, and 557 were constructed from plasmid pGEMD(-Cys) by use of site-directed mutagenesis. Plasmid pVMβ'(1–1377)-IC, encoding β'(1–1377)-intein^{VMA}-CBD (wherein intein^{VMA} is a modified *Saccharomyces cerevisiae* VMA-I intein preceded by Pro-Gly, and CBD is the *Bacillus circulans* chitin binding domain), was described in Mukhopadhyay et al. (2001). Plasmids pARβ(1–643)-IC and pARβ(1–937)-IC, encoding β(1–643)-intein^{Mxe}-CBD and β(1–937)-intein^{Mxe}-CBD (wherein intein^{Mxe} is a modified *Mycobacterium xenopi GyrA* intein preceded by Met-Arg-Met), were constructed by replacement of the NdeI-SapI segment of pTXB1 (New England Biolabs, Inc.) by NdeI-SapI DNA fragments carrying coding sequences for β(1–643) and β(1–937) (prepared by PCR of pRL706 [Severinov et al., 1997]). Plasmid pREIIα-45A(1–235)-IC, encoding [Ala45]α(1–235)-intein^{VMA}-CBD, was constructed by replacement of the XbaI-BamHI segment of pREIIα (Tang et al., 1994) by the XbaI-BamHI segment of a pHTF1α-Bam (Niu, 1999) derivative carrying the coding sequence for [Ala45]α(1–235) and XhoI and SmaI sites (prepared by use of site-directed mutagenesis), followed by replacement of the XhoI-SmaI segment by a XhoI-SmaI DNA fragment carrying the coding sequence for intein^{VMA}-CBD (prepared by PCR of pCYB2 [Chong et al., 1997]). Plasmids pT7β', p$\beta_{643-1342}$, and pHTT7f1-NHα, encoding, respectively, β', β(643–1342), and hexahistidine-tagged α, were described in Tang et al. (1995) and Naryshkin et al. (2000). Plasmid p$\beta_{938-1342}$, encoding β(938–1342), was constructed by replacement of the NcoI-XhoI segment of pET28a (Novagen, Inc.) by an NcoI-XhoI DNA fragment prepared by PCR of pRL706 (Severinov et al., 1997). Plasmid pT7ω, encoding ω, was constructed by replacement of the NdeI-EcoRI segment of pET21a by an NdeI-EcoRI DNA fragment prepared by PCR of *E. coli* genomic DNA.

DNA Fragments

Oligodeoxyribonucleotides were prepared by automated β-cyanoethyl-phosphoramidite synthesis; Cy5 was incorporated using Cy5-CE Phosphoramidite (Glen Research, Inc.). DNA fragments were prepared by PCR with synthetic primers and templates and were purified by urea-PAGE.

σ^{70}

Labeled σ^{70} derivatives were prepared using Cys-specific chemical modification (Figure 1C; methods as in Mukhopadhyay et al., 2001). Reaction mixtures for labeling of σ^{70} contained (1 ml): 20 μM single-Cys σ^{70} derivative (subjected to solid-phase reduction on Reduce-Imm [Pierce, Inc.] per manufacturer's instructions immediately before use), 200 μM tetramethylrhodamine-5-maleimide (Molecular Probes, Inc.), 100 mM sodium phosphate (pH 8.0), and 1 mM EDTA. Following 1 hr on ice, products were purified by gel-filtration chromatography on Bio-Gel P6DG (Bio-Rad, Inc.) and stored in 20 mM Tris-HCl (pH 7.9), 100 mM NaCl, 0.1 mM EDTA, 0.1 mM DTT, and 50% glycerol at −20°C. Efficiencies of labeling, determined by measurement of UV absorbance, were ~90%; site-specificities of labeling, determined by comparison to products of control reactions with Cys-free σ^{70} derivative, were ~90%.

RNAP Core: [β'(1–1377)-F]-RNAP

[β'(1–1377)-F]-RNAP was prepared by intein-mediated C-terminal labeling (Supplemental Figure S1; Mukhopadhyay et al., 2001; see also Chong et al., 1996, 1997; Muir et al., 1998). The procedure involved: (1) coexpression of genes encoding β'(1–1377)-intein^{VMA}-CBD, β'^{ts397c} (temperature-sensitive, assembly-defective β' derivative [Christie et al., 1996; Minakhin et al., 2001]), β, α, ω, and σ^{70}; (2) affinity-capture of holoenzyme containing β'(1–1377)-intein^{VMA}-CBD on chitin; (3) cleavage, elution, and concurrent labeling with Cys-F; and (4) removal of σ^{70}. A culture of strain 397c (*rpoC397 argG thi lac* ($\lambda cI_{857}h_{80}St_{68}dlac+$) [Christie et al., 1996]) transformed with plasmid pVMβ'1377-IC was shaken in 1 liter 4xLB containing 170 mM NaCl, 3 mM IPTG, and 200 μg/ml ampicillin at 37°C until OD_{600} = 1.5, and was harvested by centrifugation (5000 × g; 15 min at 4°C). Cell lysis and Polymin P fractionation were performed as in Burgess and Jendrisak (1975), except that volumes were one-twelfth those in Burgess and Jendrisak, 5 ml protease inhibitor mixture P8465 (Sigma, Inc.) was included in the lysis buffer, and 1 ml 0.8 mM σ^{70} (prepared as in Mukhopadhyay et al., 2001) was added immediately after lysis. The Polymin P eluate was applied to a 10 ml column of chitin (New England Biolabs, Inc.) preequilibrated in buffer A (20 mM Tris-HCl [pH 7.9], 0.1 mM EDTA, and 5% glycerol) containing 1 M NaCl. The column was washed with 50 ml buffer A containing 1 M NaCl, washed with 25 ml buffer B (20 mM sodium phosphate [pH 7.3], 200 mM NaCl, and 0.5 mM tris(2-carboxyethyl)phosphine [TCEP; Pierce, Inc.]), and equilibrated with 10 ml buffer B containing 150 μM Cys-F, 0.5% (v/v; saturating) thiophenol and 0.1 mM phenylmethysulfonyl fluoride. Following 8 hr at 4°C, 10–20 ml buffer A containing 200 mM NaCl and 0.1 mM DTT was applied to the column, and 1 ml fractions were collected. Fractions containing labeled holoenzyme were pooled, centrifuged to remove fine-particulate chitin (15000 × g; 15 min at 4°C), filtered through 50 ml Sephadex LH-20 (Amersham-Pharmacia Biotech, Inc.) in buffer A containing 200 mM NaCl and 0.1 mM DTT to remove excess Cys-fluorescein, and chromatographed on HiTrap Heparin (Amersham-Pharmacia Biotech, Inc.), Mono-Q (Amersham-Pharmacia Biotech, Inc.), and Bio-Rex70 (Bio-Rad, Inc.), essentially as in Kashlev et al. (1996), to remove σ^{70}. Labeled core was dialyzed into 40 mM Tris-HCl (pH 7.9), 200 mM NaCl, 0.1 mM EDTA, 0.5 mM TCEP, and 50% glycerol, and was stored at −20°C. The yield was ~0.1 mg; the efficiency of labeling, determined by measurement of UV absorbance, was ≥95%; the site-specificity of labeling, determined by comparison to products

of control reactions with RNAP derivatives not containing intein^{VMA}-CBD and with RNAP derivatives containing intein^{VMA}-CBD fused to subunits other than β′, was ≥95%.

RNAP Core: [β(1-n)-F/β([n+1]-1342)]-RNAP

β(1–643)-F and β(1–937)-F were prepared by intein-mediated C-terminal labeling under denaturing conditions (using the *M. xenopi GyrA* intein, which, unlike the *S. cerevisiae* VMA-I intein [Ayers et al., 1999; V.M. and R.H.E., unpublished data], retains significant function under denaturing conditions [M.Q. Xu, personal communication; V.M. and R.H.E., unpublished data]), and were used to reconstitute [β(1–643)-F/β(643–1342)]-RNAP and [β(1–937)/(938–1342)]-RNAP core in vitro (Supplemental Figure S1). Washed inclusion bodies containing β(1–643)-intein^{Mxe}-CBD or β(1–937)-intein^{Mxe}-CBD were prepared as described for β(1–643) or β(1–937) in Naryshkin et al. (2000, 2001), but using plasmid pARβ(1–643)-IC or pARβ(1–937)-IC, and using 10 mM TCEP in place of 1 mM DTT. Samples (10 mg) were resuspended in 1 ml buffer C (50 mM Tris-HCl [pH 8.0], 8 M urea) at 23°C, 3 ml 50 mM Tris-HCl (pH 8.0), 200 mM 2-mercaptoethanesulfonic acid (Sigma, Inc.) was added, and 20 μl 0.2 M Cys-F in dimethylformamide was added. Following 40 hr at 4°C, excess Cys-F was removed by desalting on Bio-Gel P6DG (Bio-Rad, Inc.; 30 ml column; preequilibrated and eluted with 50 mM Tris-HCl [pH 8.0], 5 M urea; 23°C), uncleaved β(1-n)-intein^{Mxe}-CBD and intein^{Mxe}-CBD were removed by affinity capture on chitin (New England Biolabs, Inc.; 3 ml column; preequilibrated and eluted with 50 mM Tris-HCl [pH 8.0], 5 M urea; 23°C), and products were stored in aliquots at −80°C. [β(1–643)-F/β(643–1342)]-RNAP and [β(1–937)-F/(938–1342)]-RNAP core were reconstituted as described for [β(1–643)/β(643–1342)]-RNAP and [β(1–937)/(951–1342)]-RNAP core in Naryshkin et al. (2000, 2001), using 500 μg β(1–643)-F and 700 μg β(643–1342), or 600 μg β(1–937)-F and 200 μg β(938–1342), including 70 μg ω (prepared as described for β′ in Naryshkin et al., 2001, but using pT7ω), and including a final incubation for 45 min at 30°C. Products were purified using metal-ion-affinity chromatography on Ni:NTA-agarose and optionally concentrated (methods essentially as in Naryshkin et al., 2000, 2001), dialyzed into 20 mM Tris-HCl (pH 7.9), 100 mM NaCl, 0.1 mM EDTA, and 0.1 mM DTT, and stored at −20°C. Yields were ~0.1 mg; efficiencies of labeling were ≥95% for [β(1–643)/β(643–1342)]-RNAP and 70%–80% for [β(1–937)/(938–1342)]-RNAP; site-specificities of labeling were ≥95%.

RNAP Core: [[Ala45]α^{II}(1–235)-F]-RNAP

[[Ala45]α^{II}(1–235)-F]-RNAP was prepared as described for [β′(1–1377)-F]-RNAP (Mukhopadhyay et al., 2001), except that the strain used was XL1-Blue (Stratagene), the plasmid used was pRElα45A(1–235)-IC (which encodes [Ala45]α(1–235)-intein^{VMA}-CBD, wherein the Ala45 substitution prevents occupancy of the α^{I} site within RNAP and thereby restricts occupancy to the α^{II} site within RNAP [Murakami et al., 1997; Niu, 1999; Estrem et al., 1999]), and chromatography on HiTrap Heparin was omitted (Supplemental Figure S1). Yield: ~0.1 mg; efficiency of labeling, ≥95%; site-specificity of labeling, ≥95%.

RNAP Holoenzyme

Holoenzyme derivatives were prepared by incubation of 4 pmol unlabeled core (prepared as in Mukhopadhyay et al., 2001) or labeled core derivative with 8 pmol unlabeled σ^{70} (prepared as in Mukhopadhyay et al., 2001) or labeled σ^{70} derivative in 20 μl transcription buffer (TB: 50 mM Tris-HCl [pH 8.0], 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 10 μg/ml bovine serum albumin, and 5% glycerol) for 20 min at 25°C. Transcriptional activities of labeled holoenzyme derivatives (determined as in Suh et al. [1992], but using the *lac(ICAP)UV5* promoter) were indistinguishable from those of unlabeled holoenzyme ($K_B = 3 \times 10^8$ M^{-1}; $k_f = 0.01$ s^{-1}), except in the cases of holoenzyme derivatives labeled at position 440 of σ^{70} (K_B not determined; $k_f = <0.001$ s^{-1}), position 442 of σ^{70} (K_B not determined; $k_f = <0.001$ s^{-1}), or position 583 of σ^{70} ($K_B = 0.6 \times 10^8$ M^{-1}; $k_f = 0.01$ s^{-1}) of σ^{70}.

FRET: Core-σ^{70} FRET, RP$_o$

For each FRET measurement, three RP$_o$ derivatives were prepared and analyzed in parallel: (1) an RP$_o$ derivative containing labeled core and labeled σ^{70} (donor-acceptor experiment, DA), (2) an RP$_o$ derivative containing labeled core and unlabeled σ^{70} (donor-only control, D), and (3) an RP$_o$ derivative containing unlabeled core and labeled σ^{70} (acceptor-only control, A). Reaction mixtures for preparation of RP$_o$ derivatives contained (20 μl): 200 nM holoenzyme derivative and 50 nM DNA fragment *lac(ICAP)CONS*(−147/+53) (positions −147 to +53 of *lac(ICAP)UV5* [Naryshkin et al., 2000] derivative having consensus −35 element, consensus −10 element, and consensus −35/−10 spacer [partial sequence in Figure 1B]) in TB. Following 15 min at 37°C, 0.5 μl 1 mg/ml heparin was added (to disrupt nonspecific complexes [Cech and McClure, 1980]), and, following a further 5 min at 37°C, reaction mixtures were applied to 5% polyacrylamide slab gels (30:1 acrylamide/bisacrylamide; 6 × 9 x 0.1 cm) and electrophoresed in 90 mM Tris-borate (pH 8.0) and 0.2 mM EDTA (20 V/cm; 1 hr at 37°C). Gel regions containing RP$_o$ were identified using an x/y fluorescence scanner (FluorImager 595; Molecular Dynamics, Inc.; Supplemental Figure S2A), excised, and mounted in sub-micro fluorometer cuvettes (Starna, Inc.; catalog number 26.40f-Q-10) containing 100 μl TB. For each gel slice, excitation spectra (emission wavelengths = 530 nm and 585 nm; excitation and emission slit widths = 5 nm; QuantaMaster QM1 spectrofluorometer; PTI, Inc.) were measured after 5 min at 37°C ("RP$_o$"), and again measured after addition of 1 μl 10% SDS and incubation for 10 min at 37°C ("random coil"; control for possible non-Förster quenching and/or transfer processes [none detected in this work]). Fluorescence intensities (F) were corrected for background by subtraction of fluorescence intensities for equivalent-size gel slices not containing RP$_o$ and were corrected for wavelength-dependence of lamp output. The fluorescence intensity attributable to FRET ($F^{FRET}_{585,492}$), the efficiency of FRET (E), the Förster parameter (R_o), and the donor-acceptor distance (R) were calculated as follows (Clegg, 1992; Mukhopadhyay et al., 2001):

$$F^{FRET}_{585,492} = F^{DA}_{585,492} - \frac{F^{DA}_{530,492} \cdot F^{D}_{585,492}}{F^{D}_{530,492}} - \frac{F^{DA}_{585,552} \cdot F^{A}_{585,492}}{F^{A}_{585,552}} \quad (2)$$

$$E = \frac{I^{FRET}_{585,492} \cdot \epsilon^{A}_{552}}{I^{A}_{585,552} \cdot \epsilon^{D}_{492} \cdot d} \quad (3)$$

$$R_o = 9780(n^{-4}\kappa^2 Q_D J)^{1/6}\ \text{Å} \quad (4)$$

$$R = R_o[(1/E) - 1]^{1/6} \quad (5)$$

where d is the efficiency of labeling of the core derivative (0.72–1.0; Experimental Procedures, Core), ϵ^{D}_{492} is the extinction coefficient of the donor at 492 nm (70,000 M^{-1} cm^{-1}), ϵ^{A}_{552} is the extinction coefficient of the acceptor at 552 nm (80,000 M^{-1} cm^{-1}), n is the refractive index of the medium (1.4 [Clegg, 1992]), κ^2 is the orientation factor relating the donor emission dipole and acceptor excitation dipole (approximated as 2/3—justified by fluorescence anisotropy measurements indicating donor and acceptor reorient on the time scale of the donor excited-state life time [Clegg, 1992; Experimental Procedures, Fluorescence Anisotropy], and, in most cases, also by the fact that E < 0.5 [Wu and Brand, 1992; Table 1]), Q_D is the quantum yield of the donor in the absence of the acceptor (0.70, 0.66, 0.73, and 0.75 for, respectively, [β′(1–1377)-F]-RNAP, [β(1–643)-F/β(643–1342)]-RNAP, [β(1–937)-F/β(938–1342)]-RNAP, and [[Ala45]α^{II}(1–235)-F]-RNAP; determined using disodium fluorescein in 0.1 M NaOH as reference [Lakowicz, 1999]), and J is the spectral overlap integral of the donor emission spectrum and the acceptor excitation spectrum (3.2–3.6 × 10^{-13} cm^3 M^{-1}; determined using corrected spectra for donor-only and acceptor-only controls [Clegg, 1992]).

FRET: Core-σ^{70} FRET, Holoenzyme

Core-σ^{70} FRET experiments with holoenzyme were performed analogously to core-σ^{70} FRET experiments with RP$_o$. Gel slices containing holoenzyme were prepared in two ways. In experiments with [β′(1–1377)-F]-RNAP and [[Ala45]α^{II}(1–235)-F]-RNAP, gel slices containing holoenzyme were prepared by application of holoenzyme (20 μl) to 5% polyacrylamide slab gels (30:1 acrylamide/bisacrylamide; 6 × 9 × 0.1 cm), followed by electrophoresis in 90 mM Tris-borate (pH 8.0) and 0.2 mM EDTA (20 V/cm; 1 hr at 25°C), followed by identification of gel regions containing holoenzyme using an x/y fluorescence scanner (FluorImager 595; Molecular Dynamics, Inc.; Supplemental Figure S2B), followed by excision of the gel region, mounting in

submicro fluorometer cuvettes, and equilibration in 100 μl TB for 10 min at 25°C (direct method). In a second set of experiments with [β′(1–1377)-F]-RNAP and [[Ala45]α^{II}(1–235)-F]-RNAP, and in all experiments with [β(1–643)-F/β(643–1342)]-RNAP and [β(1–937)-F/β(938–1342)]-RNAP, gel slices containing holoenzyme were prepared by mounting of gel slices containing RP_o in submicro fluorometer cuvettes, followed by equilibration in 200 μl TB containing 300 mM NaCl for 10 min at 37°C (salt-dissociation method). Control experiments involving measurement of DNA-σ^{70} FRET (Experimental Procedures, FRET: DNA-σ^{70} FRET) verify that the salt-dissociation method results in complete loss of DNA-σ^{70} FRET (complete dissociation of RP_o into DNA and holoenzyme). Excellent agreement between results of experiments with [β′(1–1377)-F]-RNAP and [[Ala45]α^{II}(1–235)-F]-RNAP using the direct method and the salt-dissociation method further validates the salt-dissociation method.

FRET: DNA-σ^{70} FRET

DNA-σ^{70}-FRET experiments were performed analogously to core-σ^{70} FRET experiments, except that fluorescence intensities were determined from corrected emission spectra (excitation wavelengths = 530 nm and 620 nm; excitation and emission slit widths = 5 nm; methods as in Mukhopadhyay et al., 2001). Fluorescence intensities attributable to FRET, efficiencies of FRET, and distances were calculated using equations *2*–*5* (substituting 492, 552, 530, and 585 with, respectively, 530, 620, 580, and 665), and using d = 0.85–1.0 (Experimental Procedures, σ^{70}), ϵ^{D}_{530} = 42,600 M^{-1} cm^{-1}, ϵ^{A}_{620} = 110,000 M^{-1} cm^{-1}, and R_o = 61.4 Å.

Fluorescence Anisotropy

Steady-state fluorescence anisotropies were measured in gel slices containing RP_o and holoenzyme derivatives labeled with donor only or acceptor only (Experimental Procedures, FRET). Anisotropies were measured using a QuantaMaster QM1 spectrofluorometer equipped with T-format Glan-Thompson polarizers (PTI, Inc.). Excitation and emission wavelengths were 480 nm and 520 nm for the donor, and 550 nm and 570 nm for the acceptor, and were isolated by use of interference filters (20 nm bandpass; Omega, Inc.) in addition to monochromators in order to minimize light scattering. Anisotropy (A) was expressed (Chen and Bowman, 1965) as:

$$A = (I_{VV} - GI_{VH})/(I_{VV} + 2GI_{VH}) \quad (6)$$

where I_{VV} and I_{VH} are fluorescence intensities with the excitation polarizer at the vertical position and the emission polarizer at, respectively, the vertical position and the horizontal position, and G is the grating correction factor. Measured anisotropies for fluorescein in [β′(1–1377)-F]-RNAP, [β(1–643)-F/(643–1342)]-RNAP, [β(1–937)-F/(938–1342)]-RNAP, and [[Ala45]α^{II}(1–235)-F]-RNAP were, respectively, 0.16, 0.23, 0.25, and 0.13. Measured anisotropies for tetramethylrhodamine in σ^{70} ranged from 0.26 to 0.31. In all cases, measured anisotropies of probes were low relative to fundamental anisotropies of probes (0.13–0.25 versus 0.38; 0.26–0.31 versus 0.36 [Chen and Bowman, 1965]), and low relative to calculated anisotropies of probes linked to molecules of ~0.5 MDa and having restricted local motion (0.13–0.25 versus 0.38; 0.26–0.31 versus 0.36 [Cantor and Schimmel, 1980]), indicating that probes reorient on the time scale of the probe life times, and validating use of $\kappa^2 = 2/3$ in equation *4*. Time-resolved fluorescence anisotropy measurements carried out with [β′(1–1377)-F]-RNAP and [[Ala45]α^{II}(1–235)-F]-RNAP (methods essentially as in Kapanidis et al., 2001) further indicate that probes reorient on the time scale of the probe life times and validate use of $\kappa^2 = 2/3$ in equation *4*.

Distance-Constrained Docking: Generation of Starting Models

The structure of σR2 was from Malhotra et al. (1996) (PDB accession 1SIG). The structure of σR4 was homology-modeled based on the structure of residues 155–213 of NarL (PDB accession 1RNL; Baikalov et al., 1996; Lonetto et al., 1998). The structure of RNAP core was from Minakhin et al. (2001) (PDB accession 1HQM). The structure of RNAP core with a closed clamp conformation was modeled by rotating the β′ pincer (β′ residues 3–31, 69–155, 452–523, 536–621, and 1442–1456, and β residues 1080–1115; numbered as in PDB accession 1HQM) 16° about an axis defined by Cα atoms of β′ residues 621 and 1398 (reflecting the position of the corresponding segment of RNAP II in the structure of a RNAP II tailed-template elongation complex [PDB accession 1I6H; Gnatt et al., 2001]). The downstream DNA duplex in RP_o was modeled as in Naryshkin et al. (2000). Probes and linkers were modeled into structures of σR2, σR4, core, and DNA within IMPACT (Schrödinger, Inc.), sterically-allowed conformations of probes and linkers were identified (sampling all linker torsion angles in 30° [σR2 and σR4] or 120° [core and DNA] increments, accepting configurations with van der Waals energy <5000 kcal/mol in the OPLS all-atom force field [Jorgenson et al., 1996], and randomly selecting 20 [σR2 and σR4] or 50 [core and DNA] accepted conformations), and, for each, a probe pseudoatom, corresponding to the center of the probe chromophore, was defined. σR1.1 and σR3.1 (no structural information; two probe sites each) were modeled as spheres of radius 5 Å with probe pseudoatoms at the poles. σR3.2 (no structural information; one probe site) was modeled as a sphere of radius 2.5 Å with probe pseudoatom at the center.

Distance-Constrained Docking: Grid Search

A 250 Å × 250 Å × 250 Å cube was centered on RNAP core. For each σ^{70} segment, translational and rotational space was systematically sampled by translating the σ^{70} segment in 12.5 Å increments along *x*, *y*, and *z*, and rotating in 90° increments about *x*, *y*, and *z*. For each configuration Y, for each donor-acceptor pair i, the distances between each donor-pseudoatom d_m and each acceptor pseudoatom a_n (see Experimental Procedures, Distance-Constrained Docking: Generation of Starting Models) were assessed, and were used to calculate a configuration-specific, donor-acceptor-pair-specific FRET efficiency ($E^{calc}_{Y,i}$):

$$E^{calc}_{Y,i} = \left[\sum_{m=1}^{M}\sum_{n=1}^{N}\left[1 + \left(\frac{R_{d_m a_n}}{R_o}\right)^6\right]^{-1}\right]\left(\frac{1}{MN}\right) \quad (7)$$

where M is the total number of donor pseudoatoms d_m, and N is the total number of acceptor pseudoatoms a_n. $\pi_{FRET}(Y,i)$, the configuration-specific, donor-acceptor-pair-specific penalty function for deviation between the calculated FRET efficiency ($E^{calc}_{Y,i}$) and the experimental FRET efficiency (E^{obs}_i), and $\pi_{FRET}(Y)$, the corresponding configuration-specific, global penalty function, were calculated as:

$$\pi_{FRET}(Y,i) = (2\pi\sigma_i^2)^{-1/2}\exp\left[\frac{-(E^{calc}_{Y,i} - E^{obs}_i)^2}{2\sigma_i^2}\right] \quad (8)$$

$$\pi_{FRET}(Y) = \prod_{i=1}^{I}\pi_{FRET}(Y,i) \quad (9)$$

where σ_i is the error associated with E^{obs}_i, and I is the total number of donor-acceptor pairs i (105 for RP_o; 66 for holoenzyme). Values of E^{obs}_i >0.03 and <0.97 were assigned errors corresponding to 15% of the experimental FRET distances (converted to FRET efficiencies as in equation *1*; minimum error set at σ_i = 0.01). Values of E^{obs}_i ≤0.03 or ≥0.97 were treated as bounds and were assigned errors corresponding to 30% of the experimental FRET distances (converted to FRET efficiencies as in equation *1*). For σR1.1, σR3.1, and σR3.2 (the σ^{70} segments modeled as spheres; Experimental Procedures, Distance-Constrained Docking: Generation of Starting Models), a correction for the absence of an explicitly modeled linker between the probe attachment site in the σ^{70} segment and the center of the probe chromophore was made by increasing $E^{calc}_{Y,i}$ by a factor corresponding to the mean linker length in σR2 and σR4 (11 Å; converted to a FRET efficiency as in equation *1*). The configuration with lowest $\pi_{FRET}(Y)$ from each of eight octants of the cube was selected for further study.

Distance-Constrained Docking: Markov-Chain Monte-Carlo Simulation, FRET Constraints

For each σ^{70} segment, starting from configurations defined as described in the preceding section, Markov-chain Monte Carlo simulations (Metropolis et al., 1953) employing only FRET constraints were performed (Supplemental Figure 3A). Trial configurations (Y) were generated from previously accepted configurations (X) by translations and rotations of σ^{70} segments (translations and angles of rotations selected from Gaussian distributions [σ_{trans} = 0.75 Å; $\sigma_{rotation}$ = 4.2°] about values in X; axes of rotations selected from Fisher distri-

butions [k = 30; Fisher, 1953] about values in X) and were accepted with probability $\alpha(X,Y)$, as follows:

$$\alpha(X,Y) = \min\left[1, \frac{\pi_{FRET}(Y)}{\pi_{FRET}(X)}\right] \quad (10)$$

where $\pi_{FRET}(X)$ and $\pi_{FRET}(Y)$ are configuration-specific penalty terms for FRET violations. In each simulation, 100,000 trial configurations were generated, and 40,000–75,000 were accepted. The configuration with lowest $\pi_{FRET}(Y)$ was selected for further analysis. Results were relatively insensitive to parameter selection. Similar results were obtained using E_i^{obs} bounds of $\leq$0.02 and $\geq$0.98, or $\leq$0.04 and $\geq$0.96 (mean $rmsd_c$ = 1.6 Å or 1.7 Å), or using error terms (σ_i) corresponding to 10% and 20% of R_{obs}, or 20% and 40% of R_{obs} (mean $rmsd_c$ = 1.2 Å or 1.6 Å).

Distance-Constrained Docking: Markov-Chain Monte-Carlo Simulation, FRET Constraints and Steric Constraints

For σR2 and σR4, starting from configurations defined as described in the preceding section, Markov-chain Monte Carlo simulations (Metropolis et al., 1953) employing both FRET constraints and steric constraints were performed (Supplemental Figures 3B and 3C). Trial configurations were generated and accepted as in the preceding section, but using a configuration-specific penalty term encompassing both FRET violations and steric violations [$\pi_{FRET+steric}(Y)$], as follows:

$$\pi_{FRET+steric}(Y) = \pi_{FRET}(Y) \cdot \exp[-w[E_{steric}(Y)]] \quad (11)$$

where w is a weighting factor (0.005 in this work), and where $E_{steric}(Y)$ is a configuration-specific reduced Lennard-Jones potential (Levitt, 1976) approximating repulsive and attractive interactions between σ^{70} residues (j) and core and DNA residues (k), as follows:

$$E_{steric}(Y) = \sum_{j=1}^{J}\sum_{k=1}^{K} \epsilon\left[3\left(\frac{r_o}{r_{jk}}\right)^8 - 4\left(\frac{r_o}{r_{jk}}\right)^6\right] \quad (12)$$

where r_{jk} is the configuration-specific distance between σ^{70}-segment residue j (represented by Cα atom) and core or DNA residue k (represented by core-residue Cα atom or DNA-residue P atom; excluding β' residues 587–600 [rudder; residues numbered as in PDB accession 1HQM]), J is the total number of σ^{70}-segment residues, K is the total number of core and DNA residues, r_o is the equilibrium inter-residue distance (10 Å), and ϵ is the well depth (0.15 kcal/mol). Results were relatively insensitive to parameter selection. Similar results were obtained using E_i^{obs} bounds of $\leq$0.02 and $\geq$0.98, or $\leq$0.04 and $\geq$0.96 (mean $rmsd_c$ = 0.9 Å or 1.2 Å), using error terms (σ_i) corresponding to 10% and 20% of R_{obs}, or 20% and 40% of R_{obs} (mean $rmsd_c$ = 0.6 Å or 1.7 Å), using w = 0.0025 or 0.01 (mean $rmsd_c$ = 1.0 Å or 0.8 Å), or using r_o = 8 Å or 12 Å (mean $rmsd_c$ = 2.4 Å or 2.8 Å).

Distance-Constrained Docking: Minimization

For σR2, energy minimization using all atoms of σR2 and all atoms of the β' and β pincers (all atoms of β' residues 3–31, 69–155, 452–523, 536–586, 601–621, and 1442–1456, and β residues 20–430 and 1080–1115; numbered as in PDB accession 1HQM) was performed using the OPSL-AA force field (Jorgenson et al., 1996) with a surface-generalized Born continuum solvation model (Ghosh et al., 1998).

Supplemental Data

Supplemental data are available at http://www.cell.com/cgi/content/full/108/5/599/DC1.

Acknowledgments

We thank Drs. M. Andrec and E. Gallicchio for assistance with development of the distance-constrained-docking algorithm, M.Q. Xu for discussion, and Drs. S. Busby, T. Heyduk, C. Meares, N. Naryshkin, and K. Severinov for plasmids. This work was supported by NIH grant GM41376 and a Howard Hughes Medical Institute Investigatorship to R.H.E., and by NIH grant GM64375 and a Rutgers University SROA award to R.L. and R.H.E.

Received: January 29, 2002
Revised: February 13, 2002

References

Aiyar, S., Juang, Y.-L., Helmann, J., and deHaseth, P. (1994). Mutations in sigma factor that affect the temperature dependence of transcription from a promoter, but not from a mismatch bubble in double-stranded DNA. Biochemistry *33*, 11501–11506.

Arthur, T., and Burgess, R. (1998). Localization of a σ^{70} binding site on the N terminus of the *Escherichia coli* RNA polymerase β' subunit. J. Biol. Chem. *273*, 31381–31387.

Arthur, T., Anthony, L., and Burgess, R. (2000). Mutational analysis of β'260–309, a σ^{70} binding site located on Escherichia coli core RNA polymerase. J. Biol. Chem. *275*, 23113–23119.

Ayers, B., Blaschke, U., Camarero, J., Cotton, G., Holford, M., and Muir, T. (1999). Introduction of unnatural amino acids into proteins using expressed protein ligation. Biopolymers *51*, 343–354.

Baikalov, I., Schroder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R., and Dickerson, R. (1996). Structure of the *Escherichia coli* response regulator NarL. Biochemistry *35*, 11053–11061.

Bown, J., Owens, J., Meares, C., Fujita, N., Ishihama, A., Busby, S., and Minchin, S. (1999). Organization of open complexes at *Escherichia coli* promoters. J. Biol. Chem. *274*, 2263–2270.

Burgess, R., and Jendrisak, J. (1975). A procedure for the rapid, large-scale purification of *Escherichia coli* DNA-dependent RNA polymerase involving Polymin P precipitation and DNA-cellulose chromatography. Biochemistry *14*, 4634–4638.

Burley, S., and Roeder, R. (1998). TATA box mimicry by TFIID: autoinhibition of Pol II transcription. Cell *94*, 551–553.

Callaci, S., and Heyduk, T. (1998). Conformation and DNA binding properties of a single-stranded DNA binding region of σ^{70} subunit from *Escherichia coli* RNA polymerase are modulated by an interaction with the core enzyme. Biochemistry *37*, 3312–3320.

Callaci, S., Heyduk, E., and Heyduk, T. (1998). Conformational changes of *Escherichia coli* RNA polymerase σ^{70} factor induced by binding to the core enzyme. J. Biol. Chem. *273*, 32995–33001.

Cantor, C., and Schimmel, P. (1980). Biophysical Chemistry, Vol. II (San Francisco: W.H. Freeman).

Cech, C., and McClure, W. (1980). Characterization of ribonucleic acid polymerase-T7 promoter binary complexes. Biochemistry *19*, 2440–2447.

Chen, R., and Bowman, R. (1965). Fluorescence polarization: measurement with ultraviolet-polarizing filters in a spectrophotofluorometer. Science *147*, 729–732.

Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F., and Xu, M.Q. (1996). Protein splicing involving the *Saccharomyces cerevisiae* VMA intein. J. Biol. Chem. *271*, 22159–22168.

Chong, S., Mersha, F., Comb, D., Scott, M., Landry, D., Vence, L., Perler, F., Benner, J., Kucera, R., Hirvonen, C., et al. (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. Gene *192*, 271–281.

Christie, G., Cale, S., Iraksson, L., Jin, D., Xu, M., Sauer, B., and Calendar, R. (1996). *Escherichia coli rpoC397* encodes a temperature-sensitive C-terminal frameshift in the β' subunit of RNA polymerase that blocks growth of bacteriophage P2. J. Bacteriol. *178*, 6991–6993.

Clegg, R. (1992). Fluorescence resonance energy transfer and nucleic acids. Methods Enzymol. *211*, 353–388.

Cramer, P., Bushnell, D., and Kornberg, R. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. Science *29*, 1863–1876.

Daniels, D., Zuber, P., and Losick, D. (1990). Two amino acids in an RNA polymerase σ factor involved in the recognition of adjacent base pairs in the -10 region of a cognate promoter. Proc. Natl. Acad. Sci. USA *87*, 8075–8079.

Daube, S., and von Hippel, P. (1999). Interactions of *Escherichia coli*

σ^{70} within the transcription elongation complex. Proc. Natl. Acad. Sci. USA *96*, 8390–8395.

deHaseth, P., Zupancic, M., and Record, M.T.J. (1998). RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. J. Bacteriol. *180*, 3019–3025.

Ebright, R. (2000). RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. J. Mol. Biol. *304*, 687–689.

Estrem, S., Ross, W., Gaal, T., Chen, Z., Niu, W., Ebright, R., and Gourse, R. (1999). Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase α subunit. Genes Dev. *13*, 2134–2147.

Fenton, M., Lee, S., and Gralla, J.D. (2000). *Escherichia coli* promoter opening and -10 recognition: mutational analysis of σ^{70}. EMBO J. *19*, 1130–1137.

Fisher, R. (1953). Dispersion on a sphere. Proc. Roy. Soc. Lond. *A217*, 295–305.

Förster, T. (1948). Zwischenmolekulare Energiewanderung und Fluoreszenz. Ann. Phys. *2*, 55–75.

Ghosh, A., Rapp, C., and Friesner, R. (1998). A generalized Born model based on a surface integral formulation. J. Phys. Chem. B *102*, 10983–10990.

Gnatt, A., Cramer, P., Fu, J., Bushnell, D., and Kornberg, R. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. Science *292*, 1876–1882.

Gross, C., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998). The functional and regulatory roles of σ factors in transcription. Cold Spring Harb. Symp. Quant. Biol. *63*, 141–155.

Gruber, T., Markov, D., Sharp, M., Young, B., Lu, C., Zhong, H., Artsimovitch, I., Geszvain, K., Arthur, T., Burgess, R., et al. (2001). Binding of the initiation factor σ^{70} to core RNA polymerase is a multistep process. Mol. Cell *8*, 21–31.

Huang, X., Lopez de Saro, F., and Helmann, J. (1997). σ factor mutations affecting the sequence-selective interaction of RNA polymerase with -10 region single-stranded DNA. Nucleic Acids Res. *25*, 2603–2609.

Jorgenson, W., Maxwell, D., and Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. *118*, 11225–11236.

Juang, Y., and Helmann, J. (1994). The δ subunit of *Bacillus subtilis* RNA polymerase: an allosteric effector of the initiation and core-recycling phases of transcription. J. Mol. Biol. *239*, 1–14.

Kapanidis, A., Ebright, Y., Ludescher, R., Chan, S., and Ebright, R. (2001). Mean DNA bend angle and distribution of DNA bend angles in the CAP-DNA complex in solution. J. Mol. Biol. *312*, 453–468.

Kashlev, M., Nudler, E., Severinov, K., Borukhov, S., Komissarova, N., and Goldfarb, A. (1996). Histidine-tagged RNA polymerase of *Escherichia coli* and transcription in solid phase. Methods Enzymol. *274*, 326–334.

Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. (2000). A structural model of transcription elongation. Science *289*, 619–625.

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

Lakowicz, J. (1999). Principles of Fluorescence Spectroscopy, Vol. 27 (New York: Kluwer).

Lesley, S., and Burgess, R. (1989). Characterization of the *Escherichia coli* transcription factor σ^{70}: localization of a region involved in the interaction with core RNA polymerase. Biochemistry *28*, 7728–7734.

Levitt, M. (1976). A simplified representation of protein conformations for rapid simulation of protein folding. J. Mol. Biol. *104*, 59–107.

Lilley, D., and Wilson, T. (2000). Fluorescence resonance energy transfer as a structural tool for nucleic acids. Curr. Opin. Chem. Biol. *4*, 507–517.

Liu, D., Ishima, R., Tong, K., Bagby, S., Kokubo, T., Muhandiram, D.R., Kay, L., Nakatani, Y., and Ikura, M. (1998). Solution structure of a TBP-TAF_{II} 230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. Cell *94*, 573–583.

Lonetto, M., Rhodius, V., Lamberg, K., Kiley, P., Busby, S., and Gross, C. (1998). Identification of a contact site for different transcription activators in region 4 of the *Escherichia coli* RNA polymerase σ^{70} subunit. J. Mol. Biol. *284*, 1353–1365.

Malhotra, A., Severinova, E., and Darst, S. (1996). Crystal structure of a σ^{70} subunit fragment from *E. coli* RNA polymerase. Cell *87*, 127–136.

Marr, M., and Roberts, J. (1997). Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. Science *276*, 1258–1260.

Marr, M., Datwyler, S., Meares, C., and Roberts, J. (2001). Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins. Proc. Natl. Acad. Sci. USA *98*, 8972–8978.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations for fast computing machines. J. Chem. Phys. *21*, 1087–1091.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E., Darst, S., Ebright, R., and Severinov, K. (2001). Bacterial RNA polymerase subunit ω and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. Proc. Natl. Acad. Sci. USA *98*, 892–897.

Muir, T., Sondhi, D., and Cole, P. (1998). Expressed protein ligation: a general method for protein engineering. Proc. Natl. Acad. Sci. USA *95*, 6705–6710.

Mukhopadhyay, J., Kapanidis, A., Mekler, V., Kortkhonjia, E., Ebright, Y., and Ebright, R. (2001). Translocation of σ^{70} with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. Cell *106*, 453–463.

Murakami, K., Owens, J., Belyaeva, T., Meares, C., Busby, S., and Ishihama, A. (1997). Positioning of two α subunit carboxy-terminal domains of RNA polymerase at promoters by two transcription factors. Proc. Natl. Acad. Sci. USA *94*, 11274–11278.

Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V., and Ebright, R. (2000). Structural organization of the RNA polymerase-promoter open complex. Cell *101*, 601–611.

Naryshkin, N., Kim, Y., Dong, Q., and Ebright, R. (2001). Site-specific protein-DNA photocrosslinking: analysis of bacterial transcription initiation complexes. Methods Mol. Biol. *148*, 337–361.

Niu, W. (1999). Identification and characterization of interactions between a transcription activator and the transcription machinery. PhD Dissertation, Rutgers University, New Brunswick, New Jersey.

Owens, J., Chmura, A., Murakami, K., Fujita, N., and Ishihama, A. (1998a). Mapping the promoter DNA sites proximal to conserved regions of σ^{70} in an *Escherichia coli* RNA polymerase-*lac*UV5 open promoter complex. Biochemistry *37*, 7670–7675.

Owens, J., Miyake, R., Murakami, K., Chmura, A., Fujita, N., Ishihama, A., and Meares, C. (1998b). Mapping the σ^{70} subunit contact sites on *Escherichia coli* RNA polymerase with a σ^{70} -conjugated chemical protease. Proc. Natl. Acad. Sci. USA *95*, 6021–6026.

Record, M.T.J., Reznikoff, W., Craig, M., McQuade, K., and Schlax, P. (1996). *Escherichia coli* RNA polymerase ($E\sigma^{70}$), promoters, and the kinetics of the steps of transcription initiation. In *Escherichia coli* and *Salmonella*, F.C. Neidhart, ed. (Washington, D.C.: ASM Press), pp. 792–820.

Selvin, P. (2000). The renaissance of fluorescence resonance energy transfer. Nat. Struct. Biol. *7*, 730–734.

Severinov, K., Mustaev, A., Severinova, E., Bass, I., Kashlev, M., Landick, R., Nikiforov, V., Goldfarb, A., and Darst, S. (1995). Assembly of functional *Escherichia coli* RNA polymerase containing β subunit fragments. Proc. Natl. Acad. Sci. USA *92*, 4591–4595.

Severinov, K., Mustaev, A., Kukarin, A., Muzzin, O., Bass, I., Darst, S., and Goldfarb, A. (1996). Structural modules of the large subunits of RNA polymerase. Introducing archaebacterial and chloroplast

split sites in the β and β′ subunits of *Escherichia coli* RNA polymerase. J. Biol. Chem. *271*, 27969–27974.

Severinov, K., Mooney, R., Darst, S.A., and Landick, R. (1997). Tethering of the large subunits of *Escherichia coli* RNA polymerase. J. Biol. Chem. *272*, 24137–24140.

Sharp, M., Chan, C., Lu, C., Marr, M., Nechaev, S., Merritt, E., Severinov, K., Roberts, J., and Gross, C. (1999). The interface of σ with core RNA polymerase is extensive, conserved, and functionally specialized. Genes Dev. *13*, 3015–3026.

Suh, W., Leirmo, S., and Record, M. (1992). Roles of Mg^{2+} in the mechanism of formation and dissociation of open complexes between *Escherichia coli* RNA polymerase and the λ PR promoter: kinetic evidence for a second open complex requiring Mg^{2+}. Biochemistry *31*, 7815–7825.

Tang, H., Severinov, K., Goldfarb, A., Fenyo, D., Chait, B., and Ebright, R. (1994). Location, structure, and function of the target of a transcription activator protein. Genes Dev. *8*, 3058–3067.

Tang, H., Severinov, K., Goldfarb, A., and Ebright, R. (1995). Rapid RNA polymerase genetics: one-day, no-column preparation of reconstituted recombinant *Escherichia coli* RNA polymerase. Proc. Natl. Acad. Sci. USA *92*, 4902–4906.

Traviglia, S., Datwyler, S., and Meares, C. (1999). Mapping protein-protein interactions with a library of tethered cutting reagents: the binding site of σ^{70} on *Escherichia coli* RNA polymerase. Biochemistry *38*, 4259–4265.

Vuthoori, S., Bowers, C., McCracken, A., Dombroski, A., and Hinton, D. (2001). Domain 1.1 of the σ^{70} subunit of *Escherichia coli* RNA polymerase modulates the formation of stable polymerase/promoter complexes. J. Mol. Biol. *309*, 571–582.

Waldburger, C., Gardella, T., Wong, R., and Susskind, M. (1990). Changes in conserved region 2 of *Escherichia coli* σ^{70} affecting promoter recognition. J. Mol. Biol. *215*, 267–276.

Wilson, C., and Dombroski, A.J. (1997). Region 1 of σ^{70} is required for efficient isomerization and initiation of transcription by *Escherichia coli* RNA polymerase. J. Mol. Biol. *267*, 60–74.

Wu, P., and Brand, L. (1992). Orientation factor in steady-state and time-resolved resonance energy transfer measurements. Biochemistry *31*, 7939–7947.

Young, B., Anthony, L., Gruber, T., Arthur, T., Heyduk, E., Lu, C., Sharp, M., Heyduk, T., Burgess, R., and Gross, C. (2001). A coiled-coil from the RNA polymerase β′ subunit allosterically induces selective nontemplate strand binding by σ^{70}. Cell *105*, 935–944.

Zhang, G., Campbell, E., Minakhin, L., Richter, C., Severinov, K., and Darst, S. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. Cell *98*, 811–824.