RNA polymerase II (Pol II) is the central enzyme of eukaryotic gene expression machinery. Complete definition of the three-dimensional structure of Pol II is essential for understanding the mechanisms that regulate transcription via protein-protein interactions within the Pol II apparatus. To date a series of Pol II-related crystal structures have been reported. However, certain peptide regions, including several that are implicated to interact with regulatory factors, remain obscure. Here we describe conformations for two such regions that are close to the Pol II surface and assume seemingly flexible loop structures. One is located in the TFIIF-interacting Protrusion domain, whereas the other is in the TFIIE-interacting Clamp domain. This structural definition was aided by the application of an advanced crystallographic refinement approach that utilizes the single anomalous diffraction (SAD) from zinc ions bound intrinsically in Pol II. The SAD-based strategy allowed the 12-subunit Pol II model to be fully refined up to 3.8 Å with excellent stereochemical properties, demonstrating the effectiveness of the SAD approach for the refinement of large structures at low-to-moderate resolutions. Our results also define additional components of the free Pol II, including the functionally critical Fork Loop-1 and Fork Loop-2 elements. As such, this refined Pol II model provides the most complete structural reference for future analyses of complex structures formed between Pol II and its regulatory factors.

RNA polymerase II (Pol II) and its associated factors form an elaborate protein machinery that transcribes DNA sequences into pre-mRNAs to carry out the first step of gene expression in eukaryotic cells (1). Pol II is a dodecameric protein assembled from a tightly associated 10-subunit core and a heterodimERIC subcomplex of subunits Rpb4 and Rpb7 (2). Although the 10-subunit core harbors the central transerase activity that catalyzes RNA-chain polymerization, the Rpb4/7 heterodimer enables promoter-dependent initiation by the polymerase and supports yeast growth under stress conditions (3, 4). An x-ray crystallographic breakthrough (5) combined with biochemical characterizations of the Pol II system using materials from Baker’s yeast has led to a high resolution structure of the 10-subunit core (6) and the complete 12-subunit structure at moderate resolutions (7, 8). These structural results have shed new light on the molecular mechanism that underlies the enzymatic process of pre-mRNA synthesis (9–11) and, together with structural information on related factors, provided a three-dimensional framework for integrating and rationalizing volumes of biochemical and genetic data accumulated over the past four decades (12–14). The success of this line of structural work along with the techniques and precautions developed during the process (5, 15) has opened the door to elucidating the structures of Pol II-factor complexes and understanding the regulatory mechanisms that govern class II transcription activities in the nucleus.

Structural refinement is an important step in establishing crystallographic models for macromolecules, and its feasibility is traditionally determined by the resolution limit of diffraction datasets collected from macromolecular crystals in question. More explicitly, refinement feasibility depends on whether there are a sufficient number of observations per atom to be refined. In general, refinement for macromolecules is greatly helped by the fact that amino or nucleic acid residues in the biopolymers obey their standard stereochemical rules. However, refinement of the 12-subunit Pol II has been hampered by the fact that related crystal structures diffracted to no better than 3.4 Å (7, 8). These structural results have shed new light on the enzymatic process of pre-mRNA synthesis (9–11) and, together with structural information on related factors, provided a three-dimensional framework for integrating and rationalizing volumes of biochemical and genetic data accumulated over the past four decades (12–14). The success of this line of structural work along with the techniques and precautions developed during the process (5, 15) has opened the door to elucidating the structures of Pol II-factor complexes and understanding the regulatory mechanisms that govern class II transcription activities in the nucleus.

Here, we describe a full refinement of the 12-subunit Pol II structure at 3.8 Å using a newly developed algorithm that is capable of parameter refinement for crystals that diffract to lower resolutions (e.g. 3–4 Å). This technique utilizes the diffraction signals from anomalous scatterers in the crystal and incorporates this additional information into a multivariate likelihood target function (19, 20) to increase the number of
Fully Refined Structure of RNA Polymerase II

Full refinement of Pol II, a large multiprotein complex, is difficult due to its complexity, high resolution requirements, and limited crystal availability. Here, we describe the fully refined structure of Pol II at 3.8 Å resolution, which reveals new structural features and improves our understanding of its function.

EXPERIMENTAL PROCEDURES

Purification of Pol II from yeast (Saccharomyces cerevisiae) and its crystallization, x-ray diffraction data collection, and experimental phasing have been described previously (15). To refine the model, the anomalous scatterers (the eight intrinsic sites of zinc) were first refined with the BP3 program (24) against the zinc SAD data. Next, the previous unrefined 12-subunit model (15) was subjected to geometry regularization against the zinc SAD data. Then, the positional parameters of the polymerase model were refined with the TLS groups (described below) in addition to the anomalous scatterers. The refinement process involved several cycles of refinement, with the final refinement producing a structure with a lower R-factor than previous models.

RESULTS AND DISCUSSION

Refined Structure of Complete Pol II—We found that the 12-subunit Pol II structure could be fully refined at 3.8 Å resolution with the multivariate-likelihood strategy of Pannu and co-workers (19, 20). We describe the crystallographic performance in the supplemental information. As a result of applying this technique and using the intrinsic zinc SAD data as additional observations, the quality of 12-subunit Pol II structure is significantly improved in comparison with the models published previously. As a demonstration of such improvement, Fig. 1 shows three structural regions (A, B, and C) where our model showed increased mobility not accounted for by the existing groups. For example, a modification to the TLS groups was made when (i) atoms with negative displacements were indicated in the log file and (ii) negative Fo-Fc density was observed following a region of the model that showed otherwise good density. B-factor sharpening, as described in DeLaBarre and Brunger (32), was tested but did not improve map quality.

The root-mean-square deviations from standard amino acid stereochemistry in supplemental Table S1B were calculated using REFMAC5 (33) and ADIT. β-Carbon (Cβ) deviations and Ramachandran statistics in supplemental Table S1B were calculated using MOLPROBITY. Omit maps were calculated using the program OMIT in the CCP4 suite (34). Figures were generated using PyMol (DeLano Scientific).

Throughout the refinement process, severe stereochemical distortions were reduced by the weighting determined automatically by the program, and the B-factors were adjusted similarly. In the final stages of refinement, problematic regions were identified using ADIT (work station version) (26), MOLPROBITY (27), and COOT (25).
Loop-1 (Fig. 2, A–C, respectively) and Fork Loop-2 (supplemental Fig. S2A), following the nomenclature for the 10-subunit Pol II. The former three regions were covered by density in both the omit (Fig. 2, right panels) and experimentally phased zinc multiwavelength anomalous diffraction (Fig. 2, left panels) maps, whereas Fork Loop-2 and 12 other residues were defined only in the model-based omit map (supplemental Fig. S2; Table 1). Altogether, 56 residues have been built throughout the refinement/rebuilding process. We discuss the functional relevance of these newly modeled regions below.

Conformation of the TFIIF-Interacting Rpb2 Protrusion—The earlier structures of Pol II had lacked a model for 10 amino acid residues in the protrusion domain of subunit Rpb2 (residues 437–446). A polyalanine chain was placed for this loop in an unrefined 12-subunit structure (15). We have now been able to build a starting model for this region based on the experimental zinc multiwavelength anomalous diffraction density (Fig. 2A, left) and adjust the model during the cycle of refinement with SAD information. The Protrusion adopts a loop-like conformation (Fig. 3A, yellow) consistent with its electron densities in both the experimental and omitted 2Fo-Fc maps (Fig. 2A). The Protrusion Loop might be considered as partially disordered under the solution condition of the 10-subunit crystal because it was not visible in high resolution maps of the 10-subunit Pol II (6, 35). The Protrusion Loop connects the gap from Glu-437 to Leu-446 at one end of the two-helix coil in Rpb2.

The protrusion domain has recently been shown to interact with the general transcription factor TFIIF in the context of transcription pre-initiation complex (21). The newly modeled loop sits at the apex of the helical coil (Fig. 3, A and B, yellow) and is bracketed in space by two TFIIF-cross-linking sites, one marked by Tyr-57 and Thr-68 of Rpb2 and the other by Gin-469 of Rpb2, the blue sites near Protrusion (Fig. 3B). The shortest

**TABLE 1**

| Subunit | Residues | Functional comments | Shown in |
|---------|----------|---------------------|----------|
| Rpb2   | 71, 437–446 | Protrusion, TFIIF interaction | Fig. 2A |
| Rpb1   | 187–195 | Clamp top, TFIIE interaction | Fig. 2B |
| Rpb2   | 467–477 | Fork Loop-1, DNA/RNA hybrid interactions | Fig. 2C |
| Rpb2   | 502–510 | Fork Loop-2, DNA unwinding | Fig. S2A |
| Rpb11  | 115 | | Fig. S2B |
| Rpb3   | 269 | | Fig. S2C |
| Rpb8   | 1, 64, 75 | | Fig. S2D |
| Rpb4   | 117 | | Fig. S2E |
| Rpb2   | 336 | | Fig. S2F |
| Rpb1   | 1091 | | Fig. S2G |
| Rpb2   | 720–722, 714–716 | | Fig. S2H |
| Rpb2   | 669 | | Fig. S2I |
| Rpb12  | 24 | | |
The distances from residues in Protrusion Loop to Tyr-57, Thr-68, and Gln-469 are 28, 18, and 25 Å, respectively.

We observed density for yet another un-modeled region in the protrusion domain, Leu-71 through Ile-90 (Fig. 3C, red cages). The density shows that this region runs side-by-side with the protrusion loop. The strength of this density in our current map was not sufficient to allow model building, suggesting that this region is more mobile than Protrusion Loop. Based on the peptide chain directions before and after the break, we would expect this region to be hairpin-like if the structure should be stabilized. This region includes Leu-74 that is cross-linked to TFIIF in the pre-initiation complex (21). Although Leu-74 has not been modeled, it is likely located within the bounds of the density. Because the Protrusion Loop is situated close to both Leu-74 and the above-mentioned TFIIF-interacting residues, it may be envisaged to interact with TFIIF upon approaching by the proteins. It is reasonable to expect this loop structure to undergo conformational changes when interacting with TFIIF, and such flexibility would be consistent with the aforementioned suggestion that it is partially disordered in the crystal lattice of 10-subunit Pol II. The peptide segment between Lys-134 and Lys-164 in the same region (Fig. 3C) remains disordered, as no density could be seen for it.

Conformation of the TFIIE-Interacting Clamp Top—The newly modeled Rpb1 region from Lys-187 to Asp-195 completes a gap at the top of Rpb1 Clamp domain (Fig. 3D, red). The Clamp interacts extensively with both DNA and RNA in the transcription elongation complex (17, 36) and cross-links with TFIIE in assembled pre-initiation complex (21). This loop structure is delineated in space by two TFIIE-cross-linking sites, Asn-169 and Leu-199 of Rpb1, the blue sites next to the Clamp Top (Fig. 3B). The shortest distances from residues in Clamp Top to Asn-169 and Leu-199 are 11 and 13 Å, respectively. The loop folds over by 90°. Similar to the other protrusion loops, the Clamp Top structure is expected to interact with TFIIE and possibly undergo conformational alterations when doing so.

Conformational Movement and Dynamics of Fork Loop-1—We have observed electron density for Fork Loop-1 (Rpb2 461–480) in the 12-subunit Pol II uncomplexed with nucleic acids and placed a main-chain model for the loop (15). We are now able to refine a complete model of Fork Loop-1 in free Pol II. After refinement with SAD data, the 2Fo-Fc omit map showed well defined density for the loop (Fig. 2C, right), whereas the previous 2Fo-Fc map based on unrefined model failed to do so (15). This is an indication of improvement of the current model as a result of the refinement performed in this work.

In the ternary complex with DNA/RNA, Fork Loop-1 (FL-1) interacts with both the RNA and the loop structure named “Rudder” (Rpb1 310–324), which is a part of the Rpb1 clamp domain (9). In the refined structure of free Pol II, FL-1 also contacts Rudder. Solvent-accessible surfaces of the two loop

Protrusion is marked. There is no density for the break between Lys-134 and Lys-164 of Rpb2. D, the Clamp Top conforms to a bent loop (red) and connects the Rpb1 gap (Lys-187—Asp-195) exiting in previous models. The rest of the Clamp (tan) is also shown.
structures merge at their closest point of interaction (Fig. 4A, surfaces). A distance of 3.9 Å is found between the amide of Lys-470 in Fork Loop-1 and the carbonyl of Gly-319 in Rudder (Fig. 4A, pink dashed line), suggesting that the main-chain groups may potentially serve as a pair of hydrogen-bond donor and acceptor or candidates for water-mediated hydrogen-bond interaction. The two loop structures approach one another from opposite sides of the nucleic acid binding cleft, forming a gate-like barrier that separates the cleft into two compartments, the DNA/RNA hybrid binding tunnel behind the gate of Fork Loop-1/Rudder and the downstream DNA duplex binding channel that lies perpendicularly down the hybrid binding tunnel as described previously (15).

Although Rudder maintains the same conformation in all of the reported structures (Fig. 4B, gray), Fork Loop-1 adopts one of the two main conformations observed in each of the crystals of Pol II-DNA/RNA ternary complexes, differing primarily in their distance to Rudder. Compared with its conformations in representative structures that contained DNA/RNA (Fig. 4B, yellow and red models), Fork Loop-1 in unengaged Pol II (cyan) shifts toward the floor of the downstream-DNA cleft by ~7 Å at its tip. Identities of the facing residues from each loop vary accordingly (Table 2). Together, these structures define the range at which the Fork Loop-1 conformation may fluctuate when engaging the single-stranded DNA or DNA/RNA hybrid during transcription initiation or elongation, respectively. Transient opening of the gate by Fork Loop-1 changing its conformation was first pointed out in our earlier work as a necessary step involved in the formation of open promoter complex and the process of transcription initiation (15). Based on the structural information from both the free and DNA/RNA bound Pol IIs, we suggest that the basic residues Lys-470 and Lys-471 of Rpb2, which are located at the edge of Fork Loop-1 facing Rudder, play important roles in engaging template DNA during initiation and in stabilizing DNA/RNA hybrid during elongation. In particular, we suggest that the gate of Fork Loop-1/Rudder may “lock” into a more stable interaction after the formation of the nascent RNA (longer than eight bases); the non-coding DNA strand may, thus, be kept separate from the coding strand to prevent collapse of the transcription bubble (9). This mechanism is consistent with the observation that deletion of the two residues in the Fork Loop-1 edge of mammalian Pol II that face Rudder (equivalent to Lys-471—Ala-472 in yeast Pol II) results in a loss of transcription.

**TABLE 2**

| PDB code | Dist. (Å) | FL-1 residue, atom | Rudder residue atom | Potential H-bond | Model in Fig. 4B | Reference |
|----------|-----------|---------------------|---------------------|------------------|------------------|-----------|
| Unengaged, this work | 3.5 | Lys-470, NZ | Leu-315, CD1 | | Cyan | |
| Complexed with DNA/RNA | | | | | | |
| 2E2J | 2.9 | Gln-469, O | Arg-320, NH₂ | Yes | | 22 |
| 2NVZ | 2.9 | Lys-470, O | Arg-320, NE | Yes | Yellow | 22 |
| 1Y77 | 3.0 | Lys-470, O | Ser-318, O | | | 17 |
| 2E2I | 3.1 | Lys-470, O | Arg-320, NH₁ | Yes | | 22 |
| 2NVT | 3.1 | Lys-471, NZ | Arg-320, O | Yes | Red | 17 |
| 1Y1W | 3.2 | Lys-471, O | Leu-315, CD1 | | | |
| 2E2H | 3.2 | Lys-470, O | Arg-320, CD | | | |
| 2I5 | 3.5 | Lys-471, CA | Gly-319, CA | | | 44 |
| 2I7 | 3.5 | Lys-471, O | Leu-315, CD1 | | | 44 |
| 2I8 | 3.5 | Lys-471, O | Leu-315, CD1 | | | 44 |
| 2NVQ | 3.5 | Lys-471, CE | Arg-320, O | | | |
| 2NVX | 3.5 | Lys-471, NZ | Pro-321, CD | | | 22 |
| 2VUM | 3.6 | Lys-471, CG | Gly-319, CA | | | 11 |
| 2I6 | 3.7 | Lys-471, CB | Leu-315, CD1 | | | 44 |
| 2YU9 | 4.1 | Gln-469, O | Arg-320, CD | | | |
| Complexed with duplex RNA | | | | | | |
| 2R92 | 3.3 | Lys-471, O | Leu-315, CD1 | | | 45 |
| 2R93 | 3.6 | Lys-471, O | Leu-315, CD1 | | | 45 |

*Note that PDB entries 1R9T and 1SFO (DNA/RNA complexes) are not included in this survey due to their violations of amino acid stereochemistry in the region of interest.*
During transcript elongation, the details of which remain elusive. In terms of modeling based on electron density, Resi-
dues 503–508 were initially seen as disordered in the original ternary complex containing DNA/RNA hybrid (36), and the loop segment from Arg-504 through Lys-510 could not be built for free Pol II due to the lack of reliable density (15, 18). Upon full refinement of the free polymerase using the zinc SAD information, however, we observed continuous electron density in the 2Fo-Fc difference map. The density allowed the loop structure to be completed, filling the gap between residues 503 and 511 in Rpb2 (supplemental Fig. S2A). When compared with the Fork Loop-2 models from Pol II transcription bubble complexes (Fig. 5, red and white loops), our final refined conformation (green loop) is obviously an intermediate between the two conformation extremes that have been observed for the ternary complex. This observation is consistent with the notion that Fork Loop-2 can sample several free energy minima, and each of the energetic states may be associated with a particular step during transcript elongation, the details of which remain to be seen from defined transcribing complex structures further along the RNA polymerization pathway.

Implication for Structural Refinement of Large Macromolecular Complexes—Supramolecular machineries carry out vital cellular functions in processes such as gene transcription and translation, DNA replication, chromosome organization and segregation, biogenesis of energy, cross-membrane translocation, and certain anabolic synthesis. Initial crystallographic advances had allowed structural analyses for two such systems, those of mRNA and polypeptide syntheses (5, 38) and, therefore, have established a methodological framework (39) for future investigations of large biological apparatus. However, the final step in such crystallographic studies, structural refinement of large complexes, could be cumbersome due primarily to limited resolutions (e.g. lower than 3 Å) of diffraction data from these crystals. This technical problem is not restricted to large assemblies, although it occurs more often to large structures because of their greater surface variations and, hence, their non-unique lattice packing choices.

Although the problem is usually rooted in fundamental properties of crystal lattice packing and of protein mobility intrinsic to the complex in question and may not be resolved within the predictable future, efforts can now be made to lower the resolution requirement for refinement and, thus, maximize the biological insights that may be learned from results at moderate-to-low resolutions (e.g. 3.5–5 Å) (40–43). The success that has been made in this regard relies on the increase in the number of observations made possible by the availability of experimental phase information from either multiwavelength or single anomalous diffraction data and incorporating this information into the target function for maximum-likelihood refinement (19, 20, 32). Our result demonstrates that this approach is so powerful that it effectively refines Pol II, a 513-kDa complex, to a complete structure with good stereochemical geometry (supplemental Table S1B) and sound statistics (supplemental Table S1, C and D). It is worth noting that our SAD phase information was derived solely from a total of eight zinc sites, at a very low scatterer-to-protein ratio (1 zinc per 570-residues) (15). As such, we conclude that the algorithm developed by Skubak et al. (20) is effective as a general technique for refining structures at moderate resolution limits. As the future of structural biology will see increasing efforts to crystallize multicomponent macromolecular complexes, this refinement technique will find growing applications in projects where only moderate diffraction power can be attained.

Acknowledgments—Use of the Research Collaboratory for Structural Bioinformatics, Brookhaven National Laboratory, was supported by the United States Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-AC02-98CH10886. This work was also based upon research supported by the National Science Foundation under Award DMR 0225180 and conducted at the Cornell High Energy Synchrotron Source (CHESS) using the Macromolecular Diffraction at CHESS (MacCHESS), facility supported by award RR-01646 from the National Institutes of Health through its National Center for Research Resources.
