AD NUMBER
ADB263444
NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies only; Proprietary Info.; May 2000. Other requests shall be referred to US Army Medical Research and Materiel Comd., 504 Scott St., Fort Detrick, MD 21702-5012.

AUTHORITY

USAMRMC ltr, 23 Aug 2001

AD_____

Award Number: DAMD17-97-1-7099

TITLE: Structural Determination of a Transcribing RNA Polymerase
II Complex

PRINCIPAL INVESTIGATOR: Averell Gnatt, Ph.D.

CONTRACTING ORGANIZATION: Stanford University
Stanford, California 94305-5532

REPORT DATE: May 2000

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Distribution authorized to U.S. Government
agencies only (proprietary information, May 00). Other requests
for this document shall be referred to U.S. Army Medical Research
and Materiel Command, 504 Scott Street, Fort Detrick, Maryland
21702-5012.

The views, opinions and/or findings contained in this report are
those of the author(s) and should not be construed as an official
Department of the Army position, policy or decision unless so
designated by other documentation.
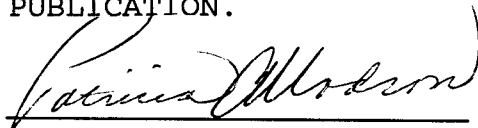
20010216 071

NOTICE

USING GOVERNMENT DRAWINGS, SPECIFICATIONS, OR OTHER DATA INCLUDED IN THIS DOCUMENT FOR ANY PURPOSE OTHER THAN GOVERNMENT PROCUREMENT DOES NOT IN ANY WAY OBLIGATE THE U.S. GOVERNMENT. THE FACT THAT THE GOVERNMENT FORMULATED OR SUPPLIED THE DRAWINGS, SPECIFICATIONS, OR OTHER DATA DOES NOT LICENSE THE HOLDER OR ANY OTHER PERSON OR CORPORATION; OR CONVEY ANY RIGHTS OR PERMISSION TO MANUFACTURE, USE, OR SELL ANY PATENTED INVENTION THAT MAY RELATE TO THEM.
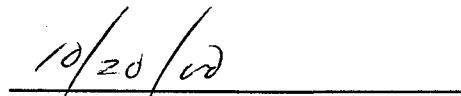
LIMITED RIGHTS LEGEND

Award Number: DAMD17-97-1-7099
Organization: Stanford University
Location of Limited Rights Data (Pages):

Those portions of the technical data contained in this report marked as limited rights data shall not, without the written permission of the above contractor, be (a) released or disclosed outside the government, (b) used by the Government for manufacture or, in the case of computer software documentation, for preparing the same or similar computer software, or (c) used by a party other than the Government, except that the Government may release or disclose technical data to persons outside the Government, or permit the use of technical data by such persons, if (i) such release, disclosure, or use is necessary for emergency repair or overhaul or (ii) is a release or disclosure of technical data (other than detailed manufacturing or process data) to, or use of such data by, a foreign government that is in the interest of the Government and is required for evaluational or informational purposes, provided in either case that such release, disclosure or use is made subject to a prohibition that the person to whom the data is released or disclosed may not further use, release or disclose such data, and the contractor or subcontractor or subcontractor asserting the restriction is notified of such release, disclosure or use. This legend, together with the indications of the portions of this data which are subject to such limitations, shall be included on any reproduction hereof which includes any part of the portions subject to such limitations.

THIS TECHNICAL REPORT HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION.

_____   10/20/00

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE May 2000	3. REPORT TYPE AND DATES COVERED Final (01 May 97 - 30 Apr 00)

4. TITLE AND SUBTITLE
Structural Determination of a Transcribing RNA Polymerase II Complex

5. FUNDING NUMBERS
DAMD17-97-1-7099

6. AUTHOR(S)
Averell Gnatt, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Stanford University
Stanford, California 94305-5532

E-Mail: gnatt@cellbio.stanford.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This report contains colored photographs

12a. DISTRIBUTION / AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT: Distribution authorized to U.S. Government agencies only (proprietary information, May 00). Other requests for this document shall be referred to U.S. Army Medical Research and Materiel Command, 504 Scott Street, Fort Detrick, Maryland 21702-5012.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)
The purpose of the proposed research was to provide a structural basis for understanding the mechanism of transcription, its regulation, and altered regulation as occurs in tumor cells. The goal of the proposed research is to determine the X-ray structure of RNA Polymerase II in the midst of transcribing RNA from a DNA template.

Despite challenges involved in this project, a major achievement is at hand. Firstly, a mainchain model of RNA Polymerase II with distinct features directly involved in transcription has been achieved. Secondly, two biochemical systems allowing for the generation of elongation complexes and thirdly the collection of X-ray diffraction data sets of two different elongation complexes has been achieved. Structural data derived from molecular replacement with the mainchain model shows that in the elongation complex, a key domain moves relative to the mainchain model, and acts as a clamp for nucleic acids in the active site cleft. A refined model of polymerase will shortly be available for molecular replacement with the elongation diffraction data, which should allow for the completion of the main goal of this project.

14. SUBJECT TERMS
Regulation, RNA Polymerase II, crystallography, elongation, ternary Complex.

15. NUMBER OF PAGES
34

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited

NSN 7540-01-280-5500

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

___ Where copyrighted material is quoted, permission has been obtained to use such material.

___ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

___ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

X In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

X In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Averell Gnatt, Ph.D.
June 1, 2000

PI - Signature   Date

TABLE OF CONTENTS

Structural Determination of a Transcribing RNA Polymerase II complex

Introduction

The goal of the proposed research is to determine the X-ray structures of RNA polymerase II in the midst of transcription complex at atomic resolution and with regulatory proteins. The purpose of the proposed research is to provide a structural basis for understanding the mechanism of transcription, regulation of the process, and altered regulation as occurs in tumor cells.

The transcription mechanism, of which RNA Polymerase II is the key player, appears to be universal. Human RNAPII subunits tested, were found able to replace their yeast counterparts in vivo (1). Therefore, studies of yeast RNAPII may be expected to reveal general principles of eukaryotic transcription and its regulation. The yeast enzyme is especially suited for 3-D structural analysis because a large amount of pure material can readily be obtained from yeast cell culture. Yeast RNA polymerase lacking subunits 4 and 7 was shown to be more homogenous than the wild type enzyme (2,3) and was therefore used for structural studies. The effort of crystallizing the polymerase in the midst of transcription has also required this project to aid in solving the polymerase structure alone (see below). After many years of effort this has just recently been achieved and a backbone structure of the ten subunit enzyme was recently determined (4).

In breast cancer research there are two major routes of study. The first is to use methods at hand or develop methods to directly intervene and eradicate tumor cells. These can be by invasive or non- invasive means. A second route of study is in advancing our knowledge of the disorder itself. The current research is of this nature. Cancer cells are different than normal cells in that they have altered regulation. The key point of regulation on the cellular level is that of transcription. Indeed mutations in tumor suppressor genes such as p53 and inherited mutations in the breast and ovarian cancer susceptibility gene, BRCA1, are directly associated with breast cancer.

The path by which mutations are capable of altering cellular traits is by affecting regulation of specific genes either at initiation or elongation of RNA polymerase II. During elongation, RNA polymerase II pauses on its transcript. Proteins such as TFIIS, can regulate the amount of the read through. Indeed, an additional elongation factor, SIII, has been shown to be a target of the VHL tumor suppressor protein and able to directly regulate its function (5). Mutations in the VHL gene predispose individuals to a variety of tumors (6). This is a clear case of point mutations in a gene, directly affecting the regulatory mechanism

In the case of breast cancer there has been growing evidence that altered regulation occurs on the level of RNA Polymerase II transcription initiation as well as elongation. Recently, p53 has been shown to regulate CAK kinase activity. CAK kinase is a component of the basal transcription factor TFIIH found to be necessary for CTD phosphorylation of RNA polymerase II in order to allow elongation of transcription (7). In addition p53 has been shown to interact directly with the general transcription factor TFIID and its TATA box-binding protein component (8). Inhibition of RNA polymerase II is a possible trigger for p53 response (9). Another example is that of BRCA1 which was found to be a component of the RNA polymerase II holoenzyme complex (10). It was further shown to, activate transcription when linked with a DNA-binding domain (11).

The most efficient means of generating RNA polymerase II elongation complexes is with the aid of tailed oligonucleotide templates. Initiation on a single strand protruding from the 3'-end of duplex DNA does not require accessory factors and allows for highly efficient generation of functional elongation complexes (12). Such a "tailed" template may be viewed as half of an unwound "bubble", which occurs at the active site of RNA polymerase molecules during transcription. Consistent with this idea, transcription starts within the single stranded region, about three bases from the junction with duplex DNA, in both tailed templates and in the unwound bubble of an elongation complex (12). At least two possible paused complexes can exist. The first is halted or paused due to the lack of a single nucleotide such as UTP (13). The second is arrested even in the presence of all nucleotides, due to the DNA structure arising from its primary sequence (14). Although structural determination of the ternary complex in a functional state until date has not been shown, use of tailed templates has allowed the applicant to develop a system with appropriate templates for the generation, purification and crystallization of this complex (15).

In previous reports from this grant proposal, the means of generating RNA polymerase II elongation complexes was described. Basically initiation on a single strand of DNA protruding from the 3'-end of duplex DNA allowed for the efficient generation of elongation complexes (15). Employing such "tailed" templates, transcription starts within the single stranded region, about three bases from the junction with duplex DNA, in both tailed templates and in the unwound bubble of an elongation complex (13). Elongation complexes were "halted" on tailed templates by transcription in the absence of UTP, so that the polymerase halted when the first T residue in the template was reached. The halted complexes generated on the tailed templates are advantageous for crystallization because of their uniformity in content of DNA and RNA sequences (15).

Despite the difficulties in determining the structure of the polymerase alone, successful crystallization and diffraction of elongation complexes were achieved under this grant. In addition, initial success at generated polymerase-TFIIS co-crystals, were also achieved. Taken together with the generation of a structural model of the polymerase alone, almost all the work required to generate the structures in question have been completed.

Body

Many substantial achievements took place during the last year of support from the breast cancer initiative.

<u>1. Collection of a complete data set to 3.2 Å of a transcribing RNA Polymease II complex in the midst of transcription.</u>

Previously, the best elongation complex crystals were plate crystals and consisted of a C2 symmetry group. They diffracted to a limit of 6.0 Å with a high mosaic spread. Furthermore when soaked with heavy metals and/or cryoprotectant, they cracked. Since diffraction of crystals in a synchrotron beam was necessary due to the size of polymerase, cryosoaking followed by freezing of the crystals became a key and crucial aspect prior to data collection. A key decision to search for new crystal forms was then implemented during the duration of the current funding, though unsuccessful.

During the search for growth of other improved crystal forms employing template 9Pause (figure 1), it was found that improved crystal diffraction could be achieved from the plate crystals by changing the cryosoaking conditions. After crystallization, the mother liquor in which the plate elongation complex crystals were grown (16%PEG 6000, 390mM Ammonium/Sodium Phosphate pH 6.0 and 5mMDTT) were gradually cryosoaked over a 10-16 hour period. The cryosoaking buffer contained 350mM Sodium Chloride, 100mM MES pH 6.2, 18% PEG 400, 15.5% PEG6000, 50 mM Dioxane and 3mM DTT. Crystals were then placed at 4°C for 4-12 days and frozen in liquid nitrogen. As a result, one in 15 crystals diffracted to better than 4Å. Since many of the plate crystals were in either twinned or cracked, a large number of crystals were screened. This indeed proved to be a successful strategy as a complete data set to 3.2Å of the plate crystals (C2 form) were collected in SSRL Beam-line 9-2 and remains a major achievement of this project (figure 2). Indeed the mosaic spread of 0.7 of this crystal is a far improvement over the original 1.5 mosaic spread of the CHESS 6Å data set (figure 2). Two diffraction data sets were taken of this crystal, the first a native data set and the second taken at the zinc anomalous peak. All data were processed within the CCP4 program suite unless otherwise indicated in the text. Data was initially processed with DENZO followed by scaling with SCALEPACK. Formatting of HKL to MTZ was by Scalepack2mtz, followed by changing of I to F by Truncate. Cad was then used to scale data and Scaleit to scale different data sets when necessary.

<u>2. Strategy for Structural Determination of the Elongation Complex and an RNA Polymerse II Mainchain Model</u>

In order to determine a protein's structure using x- ray diffraction technology, one needs good diffraction data and phase information.

Since crystal growth and diffraction of the elongation complex turned out to be a massive undertaking, it was necessary to develop an efficient strategy to achieve both good diffraction data and phase information. Although the plate crystals were fragile and difficult to grow the recent improvements allowed for the achievement of the collection of successful data as described above (section 1).

Good phase information though has proven to be an equally challenging problem. Firstly, determining of phases using heavy metals requires a fair amount of high quality crystals and very few good elongation crystals were successfully diffracted. Secondly, elongation plate crystals (C2 form) were fragile and cracked in the presence of heavy metals.

It was therefore decided that the most efficient strategy would be to perform molecular replacement with a model of the polymerase alone and use the data from the elongation complex. This project was being performed by in collaboration with co-

7

workers in Roger Kornbergs laboratory. Therefore a fair amount of effort went into helping attain a model of the polymerase. Indeed, recently a mainchain model of the polymerase has been achieved (4). This model though is not ideal for molecular replacement. It consists mostly of polyalanine. Despite this, it was recently used for molecular replacement with the elongation complex data in an attempt to extract any information from this technique. A refined model should be made available in the near future and the molecular replacement will be repeated.

In the model, the two largest subunits of RNA polymerase II form a central core and the remaining subunits are peripheral to the core. A long cleft in the molecule is ideally suited to fit nucleic acids. One side of the channel is composed of a domain consisting of subunits 1,2 and 6 and is termed the clamp domain (4).

3. Molecular replacement of the elongation complex data reveal a key movement of a clamp domain.

The mainchain model of RNA polymerase II was employed for molecular replacement of the plate C2 elongation crystal. The Amore molecular replacement programs were used initially and CNS (X-PLOR) programs to confirm results. The solutions found in both cases were nearly identical.

Structure factors were then generated using Sfall (CCP4 suite) from the model shifted into the position of the molecular replacement solution, and the scaled diffraction data from the elongation complex plate crystal (C2 form). Finally electron density maps were generated using the program FFT. Density maps were generated for the native data and for the anomalous data taken at the zinc edge. In addition an Omit electron density map was generated using the CNS program Omit.

The mainchain model of RNA polymerase II has eight zinc atoms. Using the anomalous signal from the C2 elongation complex data allowed for the localization of the zinc atoms in the elongation complex. The comparison revealed a clear shift in the positions of 4 of the eight zinc atoms. This indicates that there is protein domain shifting. Three of the zinc atoms are found in a single domain that we now call the "clamp" domain. One of the zinc atoms in the clamp indeed shifted about 16Å, which is quite substantial.

The position of the clamp domain, which has moved from its position in the model, was confirmed by two means. Firstly, the clamp domain was manually moved to align the zinc atoms of the model with the zinc atoms observed from the anomalous zinc signal. The shifted clamp domain was then observed to give a nice fit with the electron density of the elongation complex Omit map (figure 3).

As mentioned above, the clamp domain composes one side of the long DNA channel. In the RNA Polymerase II model, the clamp is found in an open conformation and in the elongation complex it is in a closed conformation. In the closed conformation, a tight binding clamp is formed on DNA placed in the channel, and it is has no room to dissociate from the polymerase. In the open conformation, double stranded DNA has more than sufficient room to either enter or exit the channel. This can be observed by placing a B-form double stranded DNA molecule in the cleft of the model structure in both the open conformation and the elongation complex closed conformation (figure 4). In addition the clamp in the elongation complex appears to be situated in a position that allows for direct interaction of the clamp domain with the nucleic acids (figure 4).

The closing of the clamp during transcription and the moveable domain help explain the ability of RNA polymerase II to be a processive enzyme. Transcription of long genes such as the blood clotting protein factor VIII, requires thousands of bases to be transcribed without polymerase disengaging. A mechanical clamp, as we observe in the elongation complex, is the mechanism of choice. On the other hand, for the DNA to rapidly move through the enzyme, a degree of freedom is required. This freedom of movement is nicely observed in the clamp domain. The clamp domain could also serve as

8

a target for other functions. Release factors would have to be involved in the "prying" open the clamp domain prior to the releasing of the DNA.

Disclosing such important features of the polymerase is a good beginning in understanding of the regulation of transcription. The next stage is in understanding where regulatory factors bind on the enzyme. A factor binding the clamp domain could have a significant effect on transcription either in causing pausing during transcription or release of polymerase from the DNA. It is likely that such regulatory factors may be present in differing amounts in breast cancer cells and in normal cells, and may be partly responsible for the altered regulation of transcription.

In previous studies, polymerase in these crystals were not only shown to contain DNA but also the to maintain and elongation competent conformation (15). No substantial electron density fitting nucleic acids were observed from the molecular replacement results. This strongly suggests that the DNA in the C2 plate crystal form may be disordered to some degree. Two possibilities for disorder exist. Firstly, initiation at three different positions, generating RNA of 3 different sizes, may be partly responsible. Secondly, the polymerase is known to "backslide" during transcription. At that point, the active site disengages from the 3' of the RNA and the polymerase slides back on the DNA. If the end position of the backsliding is not homogeneous in all or most of the elongation complexes, there may exist multiple conformations and the nucleic acid density would be difficult to observe. In this case high quality phase information may be necessary to visually the nucleic acids. For this reason, it is essential for the molecular replacement model to be as accurate as possible. Such a model should be shortly completed and the conclusions of the molecular replacement will be made known.

4. Improvement of the biochemical system allowing for more homogeneous elongation complexes

Another means of generating better diffracting crystals is to change/improve the current templates. By changing the template size and homogeneity, a different and perhaps better crystal form may be generated. It is essential to recall that on the tailed templates, initiation began at -3,-4 and -5 relative to the double stranded junction. This meant that although all the RNA was halted at a single base because of the withholding of UTP, there still existed 3 species of RNA differing in size from 1 to 3 bases in length.

To overcome this problem a series of new templates were generated. The object was to allow for initiation at a single site and maintain a structure as close to the native elongation bubble as possible.

4a. RNA Polymerase II Elongation complexes are mobile during initiation on tailed templates.

Initiation on tailed templates occurs primarily at bases -3,-4 and -5 from the double stranded-single stranded junction. In these reactions UTP is withheld to allow for efficient pausing of the polymerase. Therefore, Cytidine bases at -4, -5, and -6 (relative to the double stranded junction) in the single stranded polyC tail of template 9Pause were changed to A residues, resulting in template 10Pause (figure 1). In transcription reactions it was quite intriguing to notice that polymerase was unable to initiate at -4, -5, of template 10Pause and in turn initiated at -2 and -1 (figure 5). This is evidence that during initiation of tailed templates are not fixed in place and the position of the active site is non-homogenous. Indeed the appearance of more than one initiation in all tailed templates to date is evidence of such mobility. It is apparent that polymerase has a kinetic preference to maintain the active site position at bases -3, -4 and -5 but is in dynamic motion, with a less preferred forward sliding as observed with template 10Pause.

9

The presence of ATP added before pausing or after pausing, was mostly inconsequential though some minor differences in the elongation pattern are observed. Compare lanes 9P, Eb and Ea in figure 5 for an example.

4b. Shortening the template size
Previously (15) it was shown that to maintain an active, yet paused elongation complex, between 16 and 22 bases upstream of the pause site in the non-coding strand is needed. Comparing transcription of templates 9Pause and 11Pause establish that 18 bases downstream of the pause site is sufficient (figure 5).

4c. Mimicking the non-template strand DNA in the transcription bubble
During transcription a 7-9 base DNA-RNA hybrid exists. In turn, some DNA on the non-template strand is displaced to form the transcription bubble and remains in a single stranded form. Mimicking this may then add stability to the complex and/or allow it to become more like the form found *in vivo*. Template 12P and 13P were designed with 5 non-homologous bases in the 5' region of the non-template strand and would therefore be unable to hybridize with the template strand, mimicking a transcription bubble (figure 1). In transcription assays (Figure 5, Gel B) the overall efficiency of transcription remains the same, however a difference in the elongation pattern is observed when comparing template 10 and 12. Some "faulty" read-through present in figure 5, with template 10P is not present compared to template 12P where the 5 base 5' overhang exists. This could be due to increased stability resulting from the binding to the 5 base overhang in the nontemplate strand.

4d. Forcing initiation at a single base and preventing leakage from the pause site
Improving the homogeneity of the RNA species, it was thought, may allow for improved crystal diffraction. As mentioned above in section 4a, polymerase can initiate at -2 and -1 relative to the double stranded junction on tailed templates when prevented from initiating at -4 and -5. This was also observed from comparing the paused transcription patterns of template 13P with those of 12P and 14P (figure 6 GelC). 15P initiates at -3 and -2 and 14P initiates at -2 and -1 because they were prevented from initiation below -4 or below -3 respectively (see sequences, figure 1). It is also observed that the paused elongation patterns using forced initiation are more homogeneous in length compared to those that are not (figure 6, GelC, compare 13P with 14P).
Finally, a transcription system could now be designed for initiation at a unique site. Template 17Pause (figure 1) is comprised of a tailed template and an RNA primer. The size of the RNA primer was chosen at 9 bases because it is also 9 bases in native elongation complexes. The tail sequence was made homologous with the RNA primer and contained 4 Adenosine bases immediately before the double stranded region. The nontemplate strand was designed with a 3 base 5' nonhomologous sequence, in keeping with the idea of mimicking a transcription bubble, with a distance of greater that 10Å, possibly allowing it to reach its binding native binding domain. In transcription reactions, in addition to A,C and GTP a deoxyUTP (dUTP) chain terminator was employed. This would have a two-fold advantage. Firstly, initiation could not start from within the A bases, immediately before the double stranded region. Secondly, contamination of reactions by UTP would lead to some read-through, resulting in a non-homogeneous mixture of RNA species, which is not ideal for crystallization. dUTP, as a chain terminator would prevent that from happening. Indeed, The combination of the various technologies in use, were successful in attaining the goal of a nearly completely homogenous RNA species. The tailed template without the RNA primer (figure 6, GelD, 17P-RNA) initiated very poorly. It is also evident that reactions in GelD figure 6, were contaminated with some UTP since synthesis went beyond the pause site in both 17P and 17P-RNA. The problem of contaminating UTP

1 0

though is non-existent when dU is employed and an almost completely homogenous band RNA species is observed (Figure 6 GelD 17P lane E dU). This template is greatly improved from the original tailed templates. Crystallization could be performed in the presence of magnesium and A,C.G, nucleotides and dU without fear of misincorporation. Indeed, these paused elongation complexes were employed for crystallization and resulted in a much better diffracting crystal form (see below).

5. Intrinsically arrested elongation complexes.

Intrinsically arrested elongation complexes contain polymerase in the midst of transcribing an RNA strand yet is unable to continue even in the presence of all four nucleic acid. It was previously reported that adding a polyT region at a specific site to tailed template allows for nearly 100% of polymerase to become arrested (15). Many of the templates generated under the current grant support contain such sites. The Poly T stretches are located immediately after the pause sites in most templates. For example, template 14P in gels C and D, figure 6, allow for nearly 100% arresting immediately after the pause site. Careful inspection however, leads us to conclude that the arresting does not occur at a single base, but is spread out over 4 or 5 bases. In figure 6, Gel D, compare template 14P lane P which has nearly a single paused complex, with lane E, where although most has arrested, it was non-uniform. Indeed, it is unclear at this point if it is possible to generate a complex, which is homogeneously arrested at a single base. In addition, the challenges involved in arriving at the paused complex structure have taught that a very homogeneous RNA species may be a key prerequisite to structural determination. Since this is not currently possible with the intrinsically paused (arrested) complex, crystallization of current arrested complexes would probably not result in useful information.

6. New elongation complex crystals with improved diffraction.

Template 17pause was then employed in growing crystals. A screen was set with the PEG6000 concentration being the only variable. Within 1 week crystals with plate-like morphology grew at 14 and 15 percent PEG6000. Those were then cryosoaked and frozen and proved to be related to the C2 elongation crystal form. From two weeks to one month, an additional crystal form in lower PEG6000 concentrations (12-13%) was observed and it appeared morphologically similar to the native enzyme crystals. After cryosoaking and freezing these crystals diffracted well.

Most crystals diffracted to 4Å or better and were isomorphous to the I222 native crystal form, with the shorter a axis (Form2, see below). A full data set was taken at a wavelength of 0.98Å and is complete to 3.1Å and a sample diffraction pattern is shown in figure 7. The wavelength 0.98Å is in the tail of the zinc anomalous signal and using phases from form1 with the difference anomalous signal of the new elongation complex allowed for immediate localization of the positions of the 8 zinc atoms in the I222 elongation crystal. This crystal form is a marked improvement over the C2 plate crystal form. Firstly, 2/3 of crystals tested diffract to 4Å or better, whereas only few C2 plate crystals diffracted to 4Å. This allows for consistent diffraction data collection. Secondly, the I222 elongation crystals were easily manipulated with little observed physical damage to them as opposed to the extreme sensitivity. Most importantly is that they are closely related to the native polymerase structures. This has allowed for direct visualization of the zinc atoms. In addition, since the native crystal form structures will shortly be refined, phases could be used directly with the data, possibly eliminating the need for molecular replacement.

The new I222 elongation crystal form required slightly lower PEG6000 and much more time to grow. It was therefore necessary to determine if indeed they remained an elongation complex. For this purpose a non-denaturing nusieve agarose gel was used for confirmation. Previously (15) it was shown that polymerase, polymerase and DNA

11

(binary complex) and Polymerase, DNA and RNA (elongation complex) migrate differently on non-denaturing nusieve agarose gels. I222 elongation crystals were then washed in mother liquor, dissolved and loaded onto a nusieve agarose gel. In figure 8, polymerase from the I222 crystals clearly migrate differently than polymerase alone or polymerase binary complex. A very small amount of the elongation complex appears dissociated which sometimes occurs, after harshly dealing with elongation complexes, such as rapid pipetation needed to dissolve the crystals.

7. Unwinding of the double stranded region is crucial in determining the site of initiation on tailed templates.

Since the site of initiation is relative to the distance from the double stranded DNA, the first two bases of the double stranded region were altered in the template strand from AA to GG. The idea behind the change is that the tighter base pairing GG/CC would lessen the ability of the DNA template to unwind. Indeed the degree of opening of the double stranded region may determine the precise initiation site on tailed templates. Template 12P and 15P differed only in that 15P had the tighter binding GG/CC at the 5' of the double stranded region (figure 6). Indeed the two base change proved to dramatically collapse the paused transcription pattern (figure 6, Gel C, compare 12P and 15P). The pattern indicates that initiation at -1 and -2 was nearly abolished by the "tighter" closed DNA.

A possible mechanism for the patterns of initiation on tailed templates can now be proposed, based on the structural information from the elongation complex. A dramatic movement of the clamp domain has been observed in the elongation complex. In addition it is highly probable that the clamp directly binds the DNA (figure 3B, notice the alpha helix in close proximity to the DNA). Once the clamp is bound to the tailed template, movement of the clamp away from the cleft could then unwind to some degree the double stranded region. When initiation begins on the tailed template, several structural states would exist. The clamp bound to the non-template strand for example, could melt the DNA to varying degrees while it moves away from the cleft. Indeed it may oscillate back and forth while bound to the nontemplate strand due to forces of the DNA rehybridizing and pulling the arm closed. Initiation at any moment in time would therefore begin at various bases. This explains the existence of multiple initiation sites as well by the constitution of the changing of the first two bases in the double stranded region of the tailed templates (figure 6 GelC, compare 12P and 15P).

12

Conclusions

1. Original Project Objectives

Several Objectives were set forth in the original proposal to be performed during the duration of this fellowship. It is clear that most of the necessary results needed to fulfill the objectives have been achieved. These will be listed and discussed below.

Objective 1. X-ray structure determination of RNA polymerase II at 6 Å resolution

This goal was achieved with the successful generation of a 5Å electron density map (16) and finally a 3.3Å mainchain model of the enzyme (4). This goal was set in order to allow for the generation of a model for molecular replacement. It was assumed *a priori* that generating functional homogeneous elongation complex crystals would be a major and difficult project alone and that the native polymerase model would greatly simplify the process of determining phases by employing molecular replacement.

Objective 2. X-ray structure determination of ternary complex at 3.5 Å resolution

Support of this research has allowed for the collection of complete data sets of elongation complexes. The first, generated using the tailed template 9Pause was complete to 3.2Å and is the C2 form, from the plate-like crystals. Although successful, many crystals were grown to obtain the current data because of the mechanically weak condition of the crystals. In addition only few crystals diffracted to better than 4Å. The second form is the I222 crystal form, which is isomorphous with the native form2 crystals. Diffraction from this crystal was superior to that of the C2 form in that most crystals diffracted to better than 4Å, and crystals were less prone to damage by the various techniques employed in crystal manipulation. A complete data set to 3.1Å was collected and some diffraction was observed even beyond 3.1Å.

It must be noted that a large effort was made not only in the field of x-ray crystallography, but also on the level of the biochemistry of generating elongation complexes. This can be observed in part from results included in this document whereby a new biochemical system was devised to allow for the generation of more homogeneous elongation complexes, with a nearly homogeneous RNA species.

Although this project has not been completed, it is indeed very close to completion. For molecular replacement to give good phase information the initial model needs to be well refined. In our case, the initial model is an unrefined mainchain model, which could only supply limited phase information. This though is not of serious consequence because a refined model with sidechains is currently being generated. Indeed all the information to complete the structure is at hand and the work is underway. This means that within a short period of time, a high quality model will be available for molecular replacement.

In the mean time, molecular replacement with the current mainchain model and the C2 diffraction data has yielded valuable information about the elongation complex. It is clear that the cleft domain described in this document is a moveable domain. The clamp domain forms one wall of a long nucleic acid channel. In the mainchian model it is found in an open conformation (figures 3 and 4).This conformation is structurally compatible with the entrance or disengaging of DNA. The clamp in the elongation complex however is in a closed conformation (figures 3 and 4). In this conformation it moves closer to the other wall of the nucleic acid cleft. When a DNA molecule is placed in the cleft, the clamp in the elongation complex comes into direct contact and would prevent it from disengaging. This is crucial since polymerase needs to be highly processive, yet retain the DNA for thousands of bases. In addition the clamp allows for some "breathing" (flexibility) which is necessary for moving rapidly along the DNA template.

The results however were inconclusive as to the exact location of the nucleic acids. It appears that there may be some movement of the polymerase on the DNA. As we have

13

observed from the biochemistry in this report, polymerase is quite dynamic. It is highly mobile on its template. If there are multiple conformations then it would be difficult to directly visualize the nucleic acids. Employing a refined model for molecular replacement will allow for the highly improved generation of phase data that could allow for nucleic acid location with the current data. This should be shortly at hand. Secondly, In last years report, efficient generation of elongation complexes, were observed employing Mercuri-CTP and/or Brominated DNA templates. Since the zinc anomalous signal was easily detected using phases from the molecular replacement with the mainchain model, it is highly likely that the location of the DNA and RNA can be found using an anomalous signal generated by the use of nucleic acids containing Hg or Br.

3. Determining the structure of DNA sequences in the ternary complex caused by intrinsic pausing, a point of cellular regulation of elongation complexes

Intrinsically paused complexes are complexes that even in the presence of all four nucleotides are arrested and unable to elongate their RNA chain. Indeed in this report we observe that employing well designed, templates nearly all the complex can arrest. The sequence involved in inducing the arresting, is the poly T tracts added immediately after the pause site (15). It does not appear though that arresting occurs at a single base in the poly T track but rather at ~3 different residues that are in sequence. An important conclusion is that a new system needs to be defined to allow crystallization of intrinsically arrested polymerase. Indeed, it may be very challenging to develop such a system.

4. Co-crystallization of ternary complex with TFIIS, one of the proteins that regulates elongation at pause sites

TFIIS is and elongation factor that causes RNA polymerase II to cleave a small portion of RNA in arrested elongation complexes in a mechanism which allows polymerase to read-through the arrest site. During the second year of research under support of this grant, co-crystals of RNA polymerase II were grown and diffracted. 60% of a complete data set to 3.6Å was collected. Since then, it has been noted that these crystals are quite anisotropic and 30-40% of the diffraction was limited to 8Å. Currently more crystals are being grown. During the duration of the Breast Cancer Initiative support, co-crystals of TFIIS-Polymerase were grown and it its structural determination seems quite promising.

14

Figures

Figure 1 Oligonucleotides employed for generating elongation complexes

```
                            Pause Site
                                *
9Pause            AAGACCAGGCATTTTTTCTTGTTGCGGAAGGGG
        CCCCCCCCCCCCCTTCTGGTCCGTAAAAAAGAACAACGCCTTC
        ----Tail---              ---Upstream Region----

10Pause           AAGACCAGGCATTTTTTCTTGTTGCGGAAGGGG
        CCCCCCAAACCCTTCTGGTCCGTAAAAAAGAACAACGCCTTC

11Pause           AAGACCAGGCATTTTTTCTTGTTGCGGAA
        CCCCCCCCCCCCCTTCTGGTCCGTAAAAAAGAACAACGCCTT

12Pause      CACAC
                  AAGACCAGGCATTTTTTCTTGTTGCGGAA
        CCCCCCAAACCCTTCTGGTCCGTAAAAAAGAACAACGCCTT

13Pause      CACAC
                  AAGACCAGGCATTTTTTCTTGTTGCGGAA
        CCCCCCCCCCCCCTTCTGGTCCGTAAAAAAGAACAACGCCTT

14Pause      CACAC
                  AAGACCAGGCATTTTTTCTTGTTGCGGAA
        CCCCCCCAAACCTTCTGGTCCGTAAAAAAGAACAACGCCTT

15Pause      CACAC
                  CCGACCAGGCATTTTTTCTTGTTGCGGAA
        CCCCCCAAACCCGGCTGGTCCGTAAAAAAGAACAACGCCTT

17Pause                Pause Site
             GGC          *
                  AAGACCATTCGGCGAAGAACAAGCAA
        CCGGTCTAAAACTTCTGGTAAGCCGCTTCTTGTTCGTT
RNA      CCAGATTTT          --Upstream Region--
```

Figure 2. Crystallographic data for yeast RNA polymerase II and its complexes

| Molecule	Unit Cell	Space Group	X-ray Source	Resolution Limit (Å)	Rsymm	<I/ sigma >	Complete-ness
Native pol II Form1	a = 131.3 b = 223.7 c = 368.9	I222	SSRL BL7-1	3.0	9.9 % for 50.0-4.0 Å	30.0	75% for 50-4.0 Å
Form2	a = 211.4 b = 222.3 c = 321.6	I222 or I212121	SSRL BL7-1	5.0	7.5% for 50.0-5.0 Å		98% for 50.0-5.0 Å
Transcription elongation complex	a = 174.9 b = 222.2 c = 196.5 b = 103.3°	C2	CHESS F-1	6.0	6.1% for 70-6.4 Å	6.2	71% for 70-6.4
Wild-type pol II	a = 218.9 b = 390.5 c = 279.8	C222	SSRL BL 1-5	6.0	6.3% for 50-6.0 Å	4.7	95% to 6.0 Å
*Wild-type elongation complex	a = 219.8 b = 389.1 c = 272.0	C222	SSRL BL 1-5	6.0	7.6% for 23.7-7.5 Å	7.2	74% to 7.5 Å
*TFIIS- pol II	a = 170.9 b = 218.9 c = 193.9 b = 99.9°	C2	SSRL BL 9-1	3.34 (Edge of plate, perhaps better)	7.6% for 16.0-3.6 Å	7.6	60% to 3.6 Å
*Transcription elongation complex C2	a = 157.3 b = 220.7 c = 191.3 b = 97.5°	C2	SSRL BL 9-2	3.2	11.7% for 40 to 3.2 Å	7.2	99.8% to 3.2 Å
*Transcription elongation complex I222	a = 122.2 b = 222.2 c = 375.5	I222	SSRL BL 9-2	3.1	7.7% for 40 to 3.1 Å	16.3	99.5% to 3.1 Å

* Performed under the current study.

16

Figure 3. Elongation Complex Omit Map Confirms Clamp Domain Movement
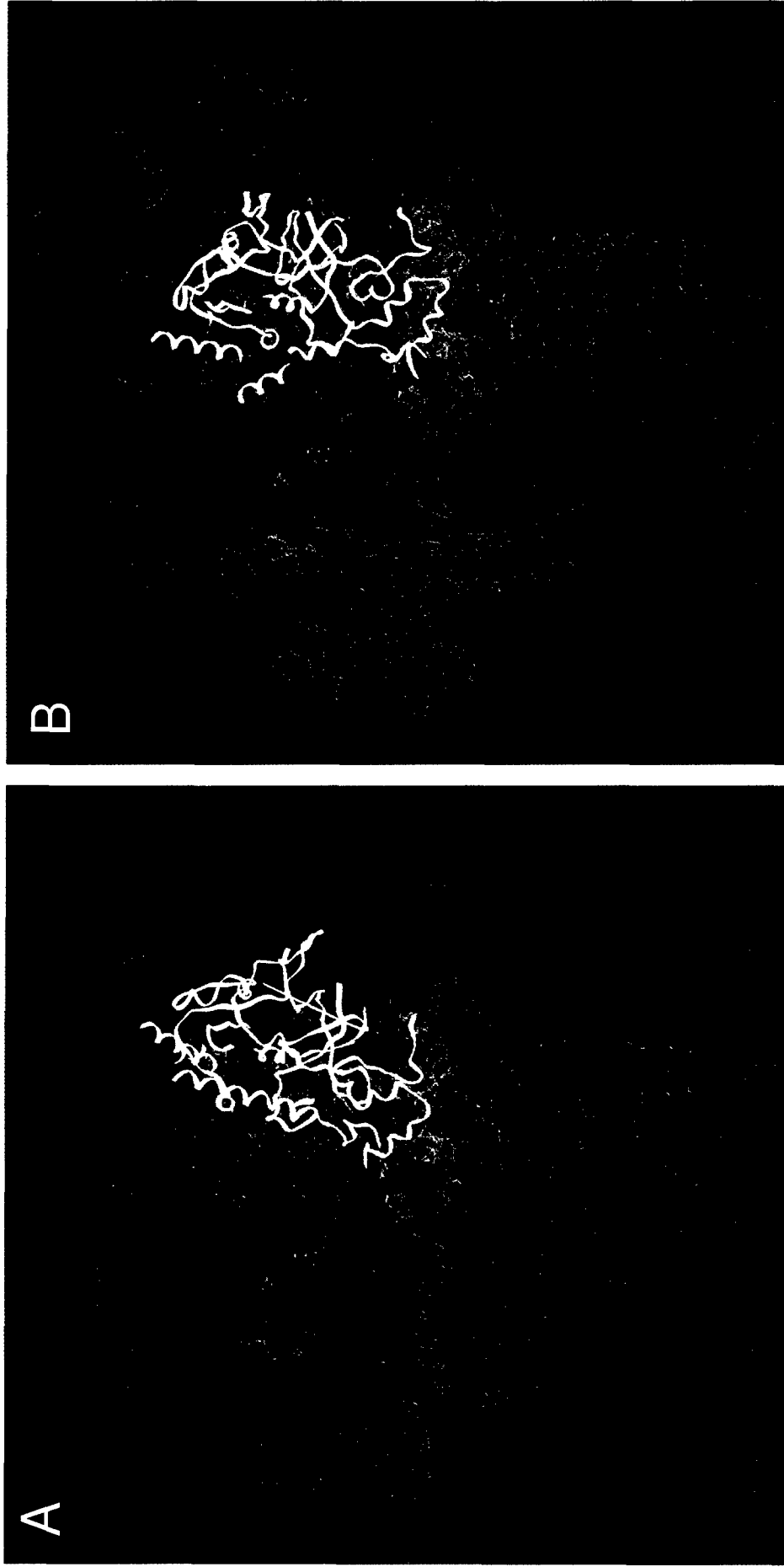

The bulk of the RNA Polymerase II mainchain model is depicted in red as a stick model and the clamp domain of the model is depicted as a yellow ribbon. The blue elctron density of the elongation complex Omit map (see text) in the region of the clamp domain is also shown. The position of the clamp domain in the model (A) does not fit the elongation complex electron density map, whereas the elongation complex zinc aligned clamp domain (B) does fit into the electron density (see text).

17

Figure 4. Small Domain is a DNA Clamp in the Elongation Complex
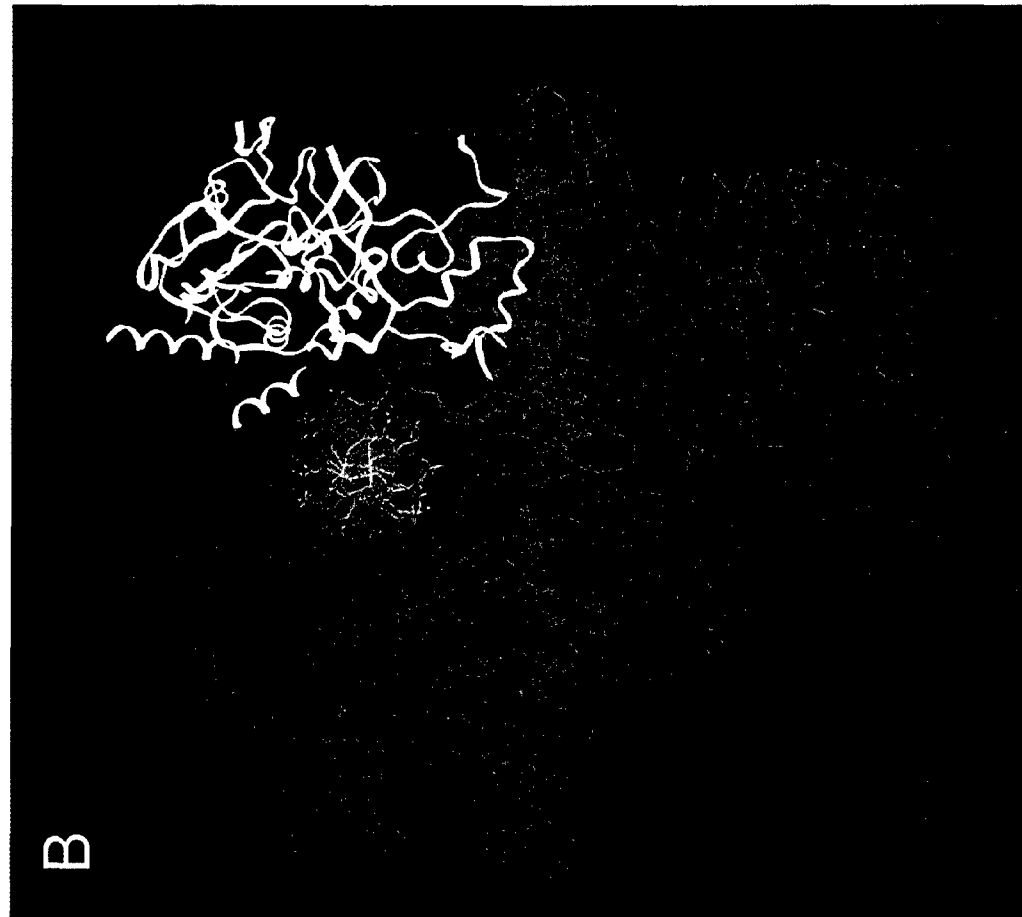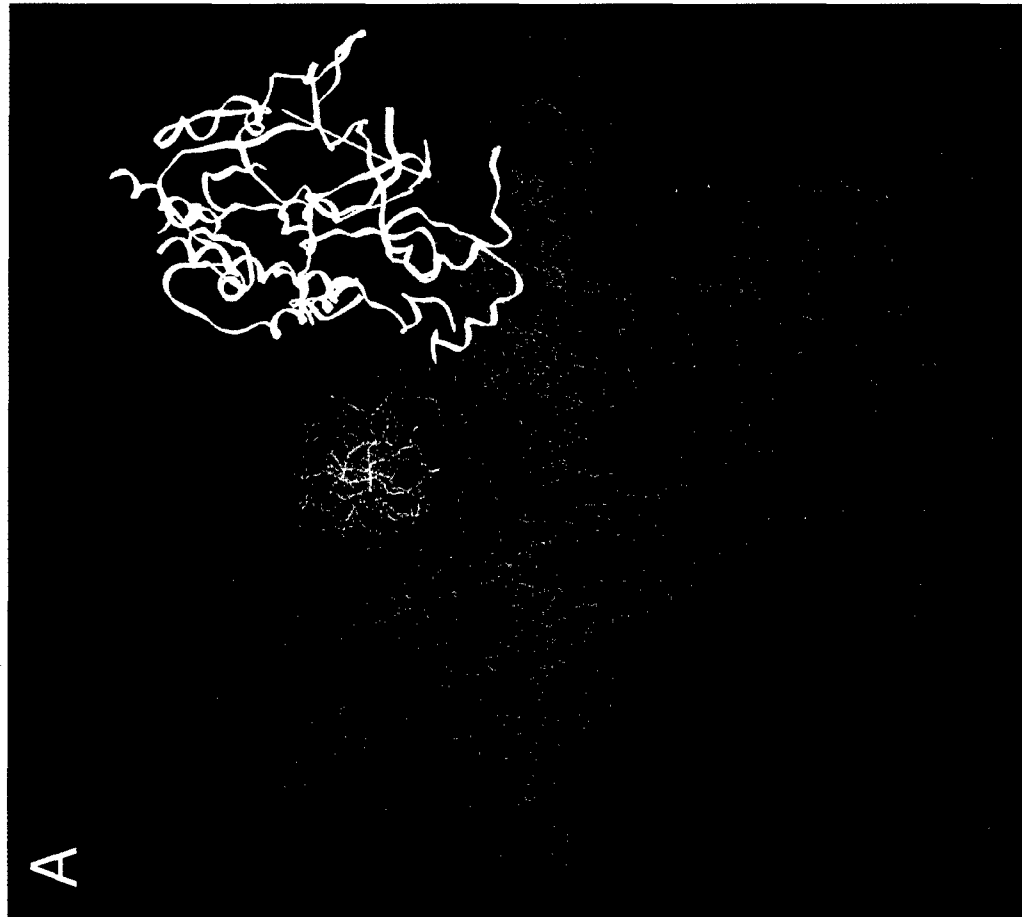
The bulk of the RNA Polymerase II mainchain model is depicted in red as a stick model and the clamp domain of the model is depicted as a yellow ribbon. A B-form double stranded DNA in green was placed in the proposed DNA biniding cleft. In **A**, the clamp domain is positioned as it is in the original mainchain model and in **B**, the position of the clamp is that of the elongation complex C2 crystal.

18

# Figure 5. RNA Polymerase II is in dynamic motion, sliding at the site of initiation on tailed templates.
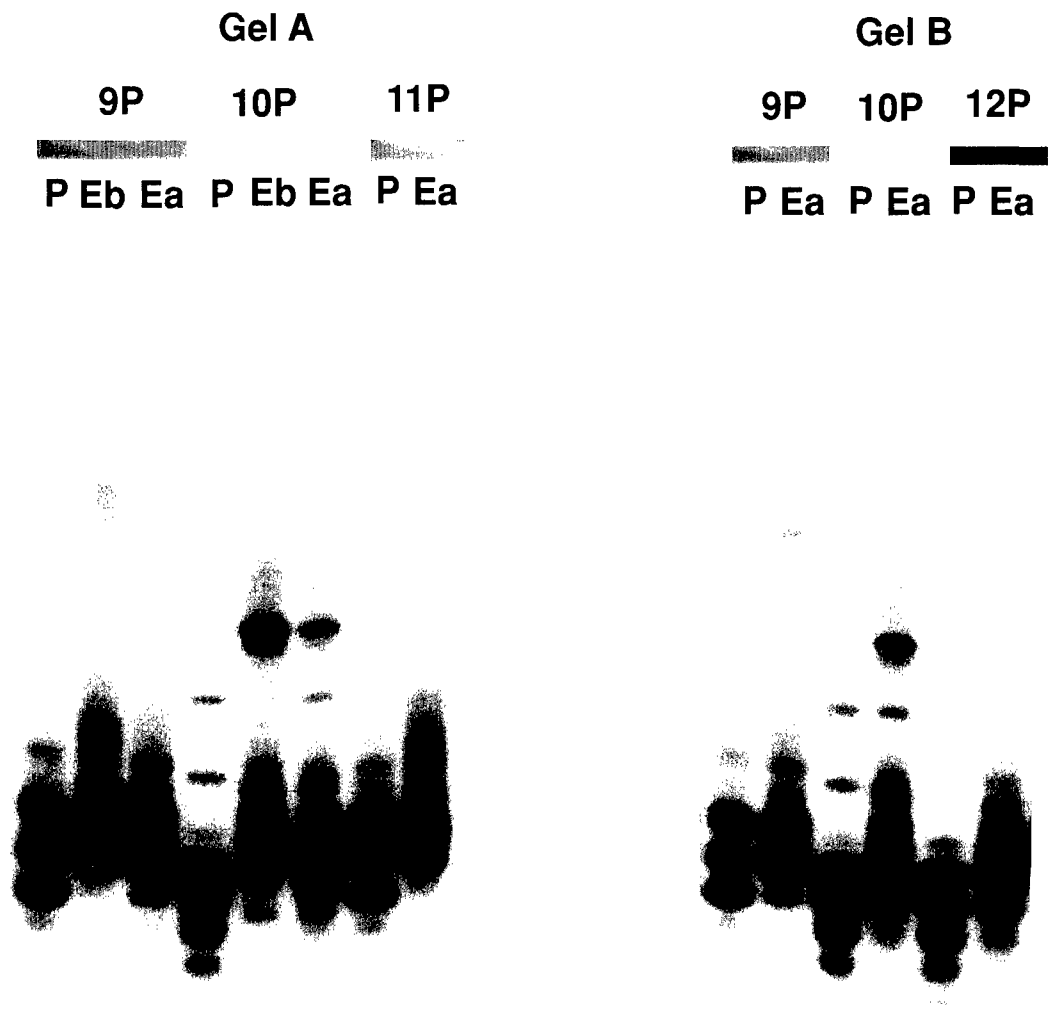
Tailed templates 9-13Pause (9P-13P) were employed for transcription reactionsas previously described (15), and paused in the presence of A,C and GTP by withholding UTP (lanes P). Transcripts were elongated by adding UTP before pausing (lanes Eb) or 15 minutes after pausing (lanes Ea). RNA in gels migrate from top to bottom. In GelA no significant difference is observed when adding UTP before or after pausing.

Figure 6. RNA Polymerase II can be made to pause at a single base on tailed templates.
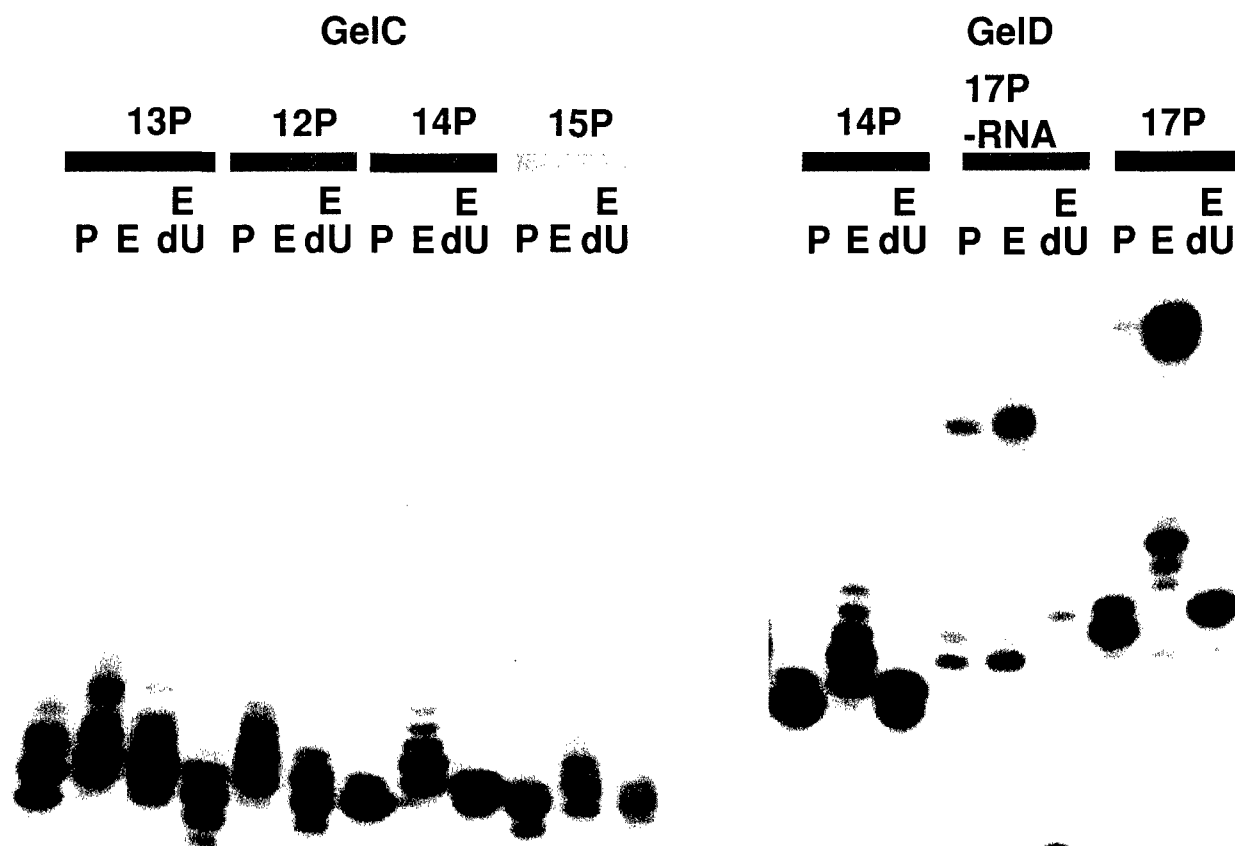
Tailed templates 12-15Pause and 17Pause (shown in figure 1) employed for transcription reactions as in figure 5, were paused in lanes P and elongated in lanes E. Template 17Pause contains a nine base RNA primer whereas template 17P -RNA does not contain the RNA primer. To prevent readthru while pausing by residual contaminating UTP, or misincorporation of nucleotides, the RNA chain terminator deoxyUTP was employed (dU).

Figure 7. Consistent Better Quality Diffraction
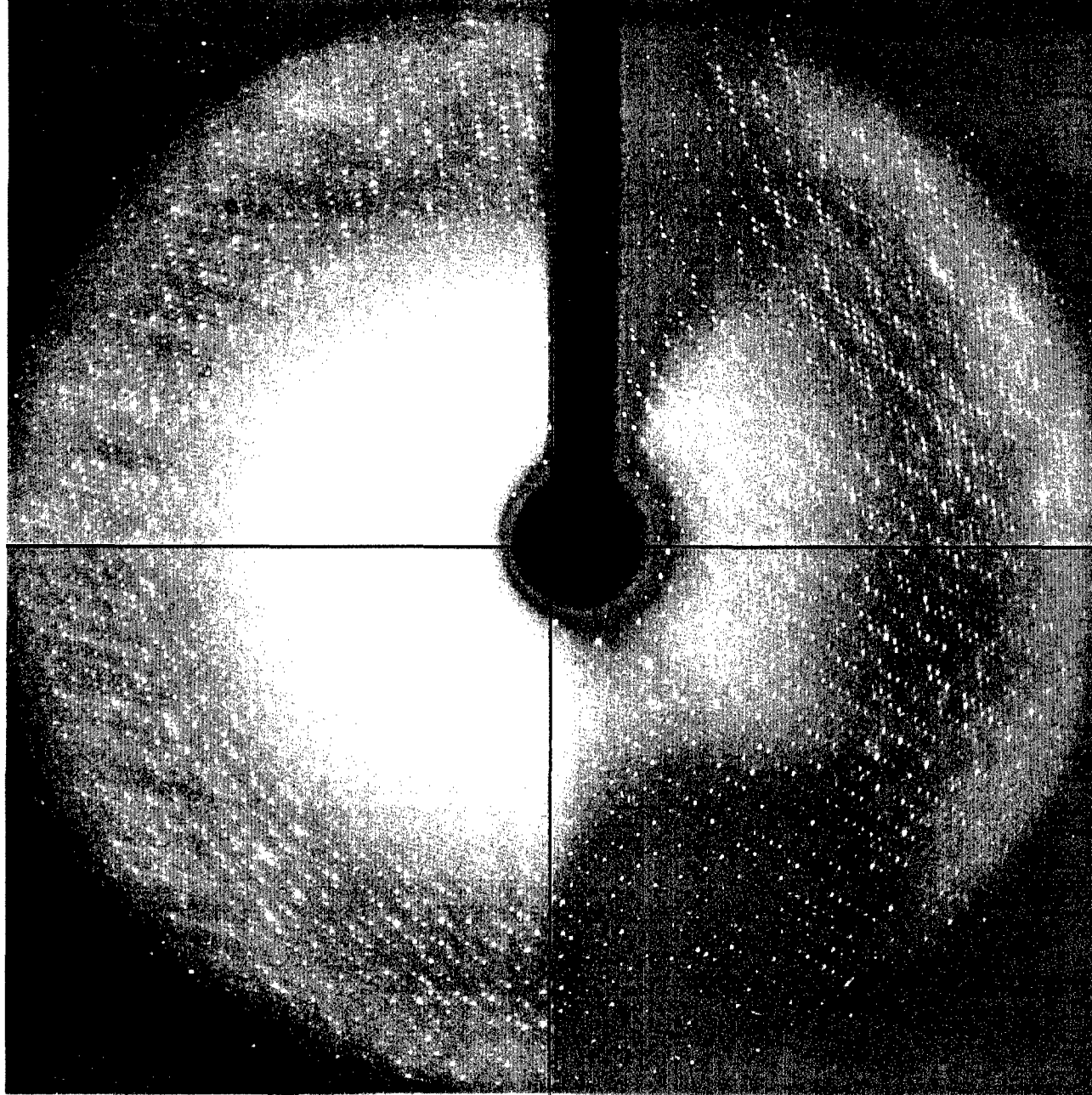
S.G.	Unit Cell
I222	121.9 Å
	221.1 Å
	374.7 Å

Resolution
3.1 Å

Figure 8. I222 Elongation Complex crystals contain polymerse that migrates as an intact elongation complex

P   T   E

Polymerase (P), Binary complex consisting of polymerase and template 17Pause (B), and polymerase from elongation complex I222 crystals generated with template17Pause (E) were were electrophoretically separated in a non-denaturing nusieve-agarose gel as previously described (15). The nearly all the elongation complex polymerase migrates faster than Polymerase alone or Binary complex. This confirms that the polymerse in the I222 crystals maintains an elongation state prior to the freezing of crystals.

22

References

1. Mckune, K. Moore, P.A. Hull, M.W. & Woychik, N. A. Molecular Cell Biology 15:6895-6900 (1995).

2. Edwards, A.E., Kane C.M., Young, R.A. & Kornberg, R.D. J. Biol. Chem. 266:71-75 (1991).

3. Edwards, A.M., Darst, S.A., Feaver, W.J., Thompson, N.E., Burges, R.R., & Kornberg, R.D., . Proc. Natl. Acad. Sci. USA 87:2122-2126 (1990).

4. Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L. Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M. David, P.R., and Kornberg, R.D. Science 288: 640-649 (2000).

5. Aso, T., Lane, W.S., Weliky, Conaway, J.W., & Conaway, R.C. Science 269:1439-1443 (1995).

6. Duan, R.D., Pause, A., Burgess, W.H., Aso, T., Chen, D.Y.T., Garret, K. P., Conaway, R.C., Conaway J.W., Linehan W.M., & Klausner, R.D. Science 269:1402-1406 (1995).

7. Schneider E., Montenarh M., andWagner P. Oncogene 17:2733-41 (1998).

8. Truant, R., Xiao, H., Ingles, C.J., and Greenblatt, J., J Biol Chem 268:2284-7 (1993).

9. Ljungman, M., Zhang, F., Chen, F., Rainbow, A.J., McKay, B.C., Oncogene 18:583-92 (1999).

10. Scully R, Anderson SF, Chao DM, Wei W, Ye L, Young RA, Livingston DM, Parvin JD Proc Natl Acad Sci U S A 94:5605-10 (1997).

11. Anderson S.F., Schlegel B.P., Nakajima T., Wolpin E.S., Parvin J.D., Nat Genet 19:254-6 (1998).

12. Kadesch, T.R. & Chamberlin, M.J. . J. Biol. Chem. 257:5286-5295 (1982).

13. Dedrick, R.L., Chamberlin, M. J. Studies on transcription of 3'-extended templates by mammalian RNA polymerase II. Parameters that affect the initiation and elongation reactions. Biochemistry 24:2245-2253 (1985).

14. Rice, G.A., Chamberlin, M. J., & Kane, C.M. Nucleic Acids. Res. 21:113-118 (1993).

15. Gnatt, A., Fu, J., & Kornberg, R.D. J. Biol. Chem. 272:30799-30805 (1997).

16. Fu, J., Gnatt, A., Bushnell, D., Jensen, G.J., Thompson, N.E., Burgess, R.T., David, P., and Kornberg, R.D. Cell 98:799-810 (1999).

Architecture of RNA Polymerase II and Implications for the Transcription Mechanism

Patrick Cramer,[1] David A. Bushnell,[1] Jianhua Fu,[1]
Averell L. Gnatt,[1] Barbara Maier-Davis,[1] Nancy E. Thompson,[2]
Richard R. Burgess,[2] Aled M. Edwards,[3] Peter R. David,[1]
Roger D. Kornberg[1]*

A backbone model of a 10-subunit yeast RNA polymerase II has been derived from x-ray diffraction data extending to 3 angstroms resolution. All 10 subunits exhibit a high degree of identity with the corresponding human proteins, and 9 of the 10 subunits are conserved among the three eukaryotic RNA polymerases I, II, and III. Notable features of the model include a pair of jaws, formed by subunits Rpb1, Rpb5, and Rpb9, that appear to grip DNA downstream of the active center. A clamp on the DNA nearer the active center, formed by Rpb1, Rpb2, and Rpb6, may be locked in the closed position by RNA, accounting for the great stability of transcribing complexes. A pore in the protein complex beneath the active center may allow entry of substrates for polymerization and exit of the transcript during proofreading and passage through pause sites in the DNA.

RNA polymerase II (pol II), the central enzyme of gene expression, synthesizes all messenger RNA in eukaryotes. The intricate regulation of pol II transcription underlies cell growth and differentiation. The size and complexity of pol II befit this important role. The best characterized form of the enzyme, that from the yeast *Saccharomyces cerevisiae*, comprises 12 different polypeptides, with a total mass of about 0.5 megadaltons (MD) (Table 1). The human enzyme must be virtually identical, as the human genes for all subunits show a high degree of sequence conservation (Table 1), and at least 10 mammalian pol II genes can be substituted for their counterparts in yeast (1).

Pol II is the core of the transcription machinery. On its own, it can unwind the DNA double helix, polymerize RNA, and proofread the nascent transcript. In the presence of additional proteins, it assembles even larger initiation and elongation complexes, capable of promoter recognition and response to regulatory signals. A regulated initiation complex comprises pol II, five general transcription factors, and a multiprotein Mediator (2–4). It contains some 60 proteins, with a total

mass of 3.5 MD. In transcription elongation complexes, Mediator and some of the general transcription factors are replaced by SII (TFIIS), Elongator, other elongation factors, and RNA processing proteins (3, 5, 6).

Determination of molecular models for the pol II transcription machinery has so far been limited to a half dozen of the smallest proteins and protein fragments (7–17). Detailed structural studies of the larger proteins and multiprotein complexes, essential for understanding the mechanism and regulation of transcription, pose a more formidable challenge. We report here the x-ray analysis of a 10-subunit yeast pol II. As nine of the subunits are conserved among RNA polymerases I, II, and III (18), our findings provide a basis for understanding the entire eukaryotic transcription machinery. They suggest roles for each of the many subunits and give insight into the remarkable features of the transcription mechanism.

Our investigation stemmed originally from the development of a yeast cell extract capable of accurately initiated pol II transcription (19) and the development of a general method of forming single-layer [two-dimensional (2D)] protein crystals (20). An active extract opened the way to the isolation of functional pol II (21), whereas the 2D crystallographic approach extended the reach of structure determination to such scarce, large, fragile multiprotein complexes. The first 2D crystallization trials gave crystals too small and too poorly ordered for structure determination (21). However, the ease and small amount of material required for 2D

crystallization allowed its use as a structural assay to guide the preparation of pol II that would form better crystals. It soon emerged that heterogeneity of pol II, owing to substoichiometric levels of two small subunits, Rpb4 and Rpb7, was an impediment to crystallization. The problem was solved by the isolation of pol II from an *RPB4* deletion strain of yeast, yielding a "deletion" enzyme lacking both Rpb4 and Rpb7, which together account for only 8% of the mass of the wild-type protein. The deletion enzyme, unimpaired in transcription elongation and also fully active in transcription initiation when supplemented with the missing subunits (22), formed exceptionally large, well-ordered 2D crystals (23). Structures of pol II alone, and complexed with general transcription factors and nucleic acids, were determined by 3D reconstruction from electron micrographs of 2D crystals to about 15 Å resolution (24–27). In the course of this work, it became apparent that even at the low protein concentration used for 2D crystallization, typically about 0.1 mg/ml, there was a tendency of the crystals to grow epitaxially, adding additional layers in register with the first (23). This tendency was exploited by the use of 2D crystals as seeds for growing 3D crystals (28), which are now readily obtained by conventional methods as well.

X-ray diffraction from 3D crystals of pol II was initially undetectable. The problem proved to be oxidation. Maintenance of an inert atmosphere during the final stages of protein purification and throughout crystal growth, as well as improvements in crystallization conditions, enabled the collection of diffraction data to 3.5 Å resolution (29). Because of the great size of the protein and unit cell, only large heavy atom clusters, such as an 18–tungsten-atom cluster, could be used for initial phase determination. The validity of the initial phases was shown by a close fit of the electron density map computed at 6 Å resolution to the pol II map from 2D crystallography (29). There was only one deviation between the two maps, which was attributed to movement of a protein domain, suggested to clamp nucleic acid in a transcribing complex (29).

With a 6 Å phase set, it should have been possible to locate individual heavy atoms in isomorphous derivatives and to extend structure determination to higher resolution. There were, however, three major obstacles. First, diffraction to 3.5 Å resolution could not be obtained reproducibly. Second, the crystals were nonisomorphous, varying by as much as 10 Å in one dimension of the unit cell. Very few crystals could be derivatized and matched with an isomorphous native crystal. Because of the low abundance of pol II, approximately 10,000 liters of cell culture had to be processed to obtain the 6 Å electron density map, and far more would have been required for extension to high resolution. The final obstacle was that heavy atom compounds commonly used for protein

[1]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305–5126, USA. [2]McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Madison, WI 53706, USA. [3]Banting and Best Department of Medical Research, University of Toronto, Toronto, M5G 1L6, Canada.

*To whom correspondence should be addressed. E-mail: kornberg@stanford.edu

phase determination destroyed diffraction from the crystals.

A crystallographic backbone model for RNA polymerase II. These difficulties were overcome in the present work by a soaking procedure that shrank the crystals to an apparent minimum of the variable unit cell dimension (*30*). The resulting crystals were isomorphous and diffracted isotropically to 3.0 Å resolution (*31*). Because the improved crystals were non-isomorphous with the original crystals, initial phases were redetermined by multiple anomalous dispersion (MAD) with a six–tantalum-atom cluster derivative, which showed a single peak in difference Pattersons (Fig. 1) (*32*). These phases sufficed to reveal individual heavy atoms in other crystals by means of cross-difference Fouriers (Fig. 1) (*33*). An extensive search identified nonstandard mononuclear heavy atom compounds that gave useful derivatives (Table 2) (*34*). Phases were determined by multiple isomorphous replacement with anomalous scattering (MIRAS) from 10 data sets, ranging from 4.0 to 3.1 Å resolution (Table 2) (*35*). The resulting molecular envelope was in good agreement with that previously obtained at 6 Å resolution (*29*). After solvent flattening, an electron density map was obtained that revealed the course of the polypeptide chain and many amino acid side chains (Fig. 2) (*36*).

Available structures of pol II subunits and subunit fragments, comprising 14% of all pol II amino acid residues, were manually fit into the electron density (*37*). The complete structures of yeast Rpb5 and Rpb8 were used, whereas structures of *Escherichia coli* and archaebacterial homologs of yeast Rpb3, 6, 9, 10, and 11 were truncated to the conserved regions (Table 1). In all cases, a unique fit of the subunit fold to regions of the electron density map was observed. Subunit placement was facilitated by the location of eight zinc ions, revealed by a zinc anomalous difference Fourier (Fig. 1 and Table 1). Most parts of the yeast subunits missing from the homologous proteins could be modeled as polyalanine into adjacent regions of electron density. The remaining density, about 70% of the total volume, was attributed to the two large subunits, Rpb1 and Rpb2, with a minor contribution from the smallest subunit, Rpb12. It was modeled as polyalanine fragments, with the use of standard secondary structure elements wherever possible. Combination of phases from MIRAS and an initial polyalanine model resulted in an improved map, which allowed adjustment and extension of the model (*38*). The polyalanine fragments were assigned to Rpb1 or Rpb2 on the basis of (i) the location of the active-site metal bound by Rpb1 (see below); (ii) two zinc-binding motifs in the NH₂-
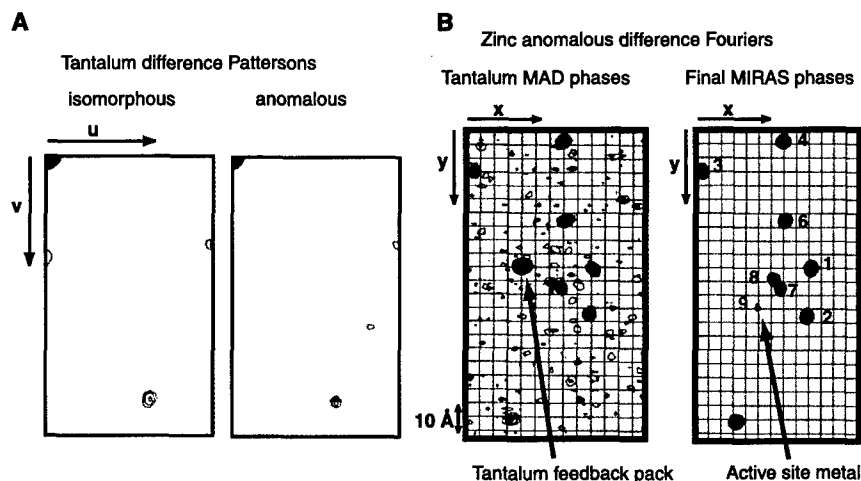

Fig. 1. Localization of heavy atoms. **(A)** Harker sections of isomorphous and anomalous difference Patterson maps of the tantalum cluster derivative (Table 2). A single peak at the same position in the two maps is observed. Heights of the Harker peaks in the isomorphous and anomalous difference Pattersons were 6 σ and 5 σ, respectively. The resolution range of the data used is 40 to 5.5 Å. The contour levels are 3 σ (background) and 1 σ (steps). **(B)** Anomalous difference Fourier calculated with native data collected at the zinc anomalous peak energy using initial tantalum MAD phases (left) and final MIRAS phases (right). The projection of one asymmetric unit along the z axis is shown for tantalum and MIRAS phases at a contour level of 3 σ and 7 σ, respectively, with 1 σ steps. The eight strong peaks correspond to structural zinc atoms (Table 1). The ninth peak corresponds to the active site metal and likely arises from partial replacement of magnesium by zinc.

Table 1. Yeast RNA polymerase II subunits.

Subunit	Mass (kD)	Residues in sequence	Identity to human (%)[1]	Structure used in modeling — Organism	Protein	Method	PDB code	Reference	Conserved residues[4]	Residues in model (%)	Zinc site (σ)[2]	Surface cysteines[3]
Rpb1	191.6	1733 (1449)[5]	52							1213 (84)[5]	Zn6 (23.2), Zn8 (19.3), "Zn9" (9.8)[6]	
Rpb2	138.8	1224	61							949 (78)	Zn7 (23.1)	
Rpb3	35.3	318	46	E. coli	α	X-ray	1bdf	(69)	8–69, 162–180, 233–251	264 (83)	Zn2 (30.2)	Cys207
Rpb4	25.4	221	30									
Rpb5	25.1	215	45	S. cerevisiae	Rpb5	X-ray	1dzf	(47)		211 (98)		Cys83
Rpb6	17.9	155	59	Human	RPABC14.4	NMR	1qkl	(12)	79–154	140 (96)		
Rpb7	19.1	171	61									
Rpb8	16.5	146	43	S. cerevisiae	Rpb8	NMR	1a1d	(13)		114 (78)		Cys24, Cys36
Rpb9	14.3	122	37	Thermococcus celer	Rpb9 COOH-term. domain	NMR	1qyp	(14)	67–112	117 (96)	Zn3 (27.5), Zn4 (26.4)	
Rpb10	8.3	70	73	Methanobacterium thermoautotrophicum	Rpb10 homolog	NMR	1ef4	(81)	3–55	65 (93)	Zn1 (31.9)	
Rpb11	13.6	120	50	E. coli	α	X-ray	1bdf	(69)	19–101	110 (92)		
Rpb12	7.7	70	43							36 (51)	Zn5 (24.7)	
Total	513.6	4565	53							3219 (83)[7]	8+1	

[1]Percentage of identical amino acid residues, for Rpb1 excluding the COOH-terminal domain. [2]Peaks in the zinc anomalous difference Fourier shown in Fig. 1. Peaks are numbered according to their height which is given in parentheses in multiples of the standard deviation. Zn6 and Zn8 are located in the NH₂-terminal region of Rpb1. Zn7 is located in the COOH-terminal region of Rpb2. Zn3 is located in the NH₂-terminal and Zn4 in the COOH-terminal domain of Rpb9. [3]These exposed cysteine residues coincide with mercury sites in two independent derivatives [mercury, Table 2, and ethylmercuryphosphate (89)], confirming the modeling at several places. [4]Conserved residues (yeast protein numbering) to which the model structure was truncated before placement in the electron density. [5]The numbers in parentheses correspond to Rpb1 without the unstructured COOH-terminal domain (CTD). [6]The ninth peak in the zinc anomalous difference Fourier corresponds to the active site metal and likely arises from partial replacement of the active site metal by zinc. [7]The number in parentheses corresponds to the pol II mutant used in structure determination, which lacks subunits Rpb4 and Rpb7.
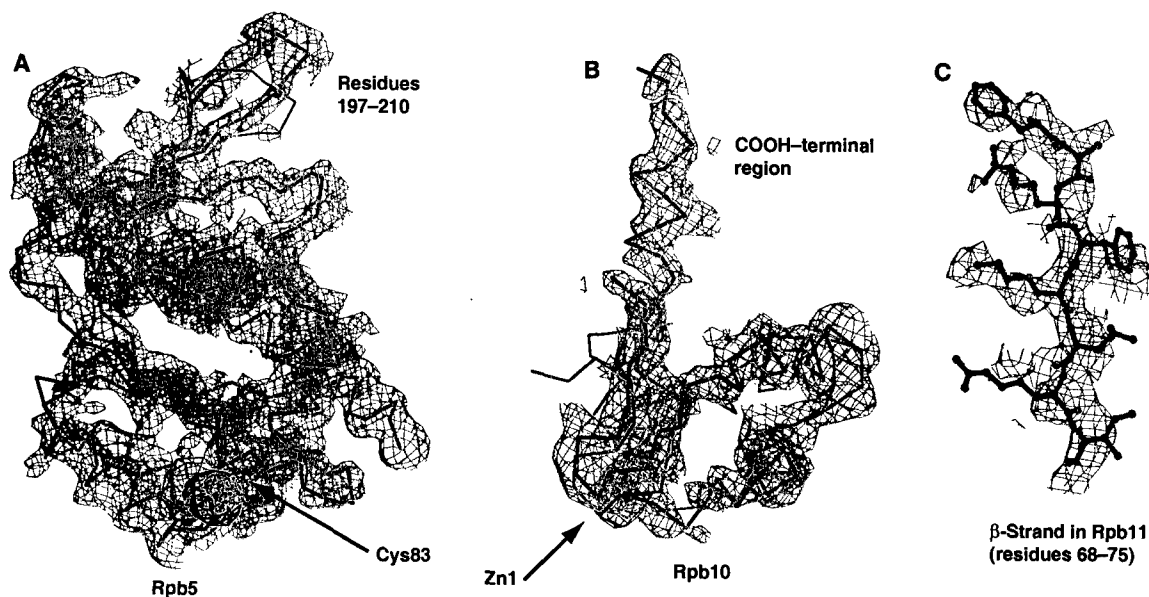
terminal region of Rpb1, connected by a linker of appropriate length; (iii) one zinc site in the COOH-terminal region of Rpb2; and (iv) cross-linking of Rpb5 to the COOH-terminal region of Rpb1 and of Rpb3 to residues 901 to 992 of Rpb2 (*39*).

The current backbone model comprises 8 polyalanine fragments for Rpb1, 10 fragments for Rpb2, and major portions of all small sub-units (Table 1). It accounts for the entire molecular volume observed in the crystals and contains 3219 residues, about 83% of the total,

assuming all residues are ordered except the COOH-terminal domain of Rpb1. Building of an atomic model is well advanced.

General architecture and DNA binding. The two largest subunits, Rpb1 and Rpb2, form distinct masses with a deep cleft between them

Fig. 2. Subunit structures determined previously or rebuilt here fitted to the experimental pol II electron density. The solvent-flattened MIRAS electron density map (blue) is contoured at 1.0 σ. Experimental phases in the resolution range 40 to 3.1 Å were used to calculate the map. In (A) and (B), the map was filtered with program MAPMAN to reduce noise (*84*). This map facilitated fold recognition but appears to be at lower resolution, and side chain density is largely removed. In (C), the original map is shown, which is noisier but reveals many details. (**A**) Cα model of Rpb5 [black (*47*)] fitted to the density (blue). A loop that is involved in packing against Rpb1 is in a different conformation in pol II than in the structure of free Rpb5 (orange). Peaks of anomalous difference Fourier transforms of two mercury derivatives (pink, yellow, both contoured at 5 σ) coincide with the position of Cys83. (**B**) Cα traces of the NMR structure of the Rpb10 homolog from *Methanobacterium thermoautotrophicum* [orange (*81*)] fitted to the density (blue) and the rebuilt backbone model for yeast Rpb10 (black). The location of the zinc ion in the NMR structure coincides with a strong peak in the zinc anomalous Fourier (pink, contoured at 7 σ). (**C**) One of the β strands in Rpb11 (black, residues 68 to 75) fitted to the density (blue). Distinct electron density is present for several side chains. The model was obtained by placing the conserved core of *E. coli* α (*69*) and replacing the side chains with those in yeast Rpb11 using the most common rotamer. This figure was prepared with BOBSCRIPT (*85*) and MOLSCRIPT (*86*).

Table 2. Data collection and MIRAS phasing.

Dataset[1]	X-ray source[2]	Wavelength (Å)	Resolution (Å)	Unique reflections[3]	Completeness (%)[3]	Redundancy	Mosaicity (°)[4]	R_{sym} (%)[1,5]	Soaking time and concentration		Sites	R_{iso} (%)[6]	R_{Cullis} iso/ano[7]	Phasing power iso/ano[7]
Data collection and MIRAS phasing. I222, 131 x 225 x 370 Å														
Native (zinc)	SSRL	1.283	40–3.1	98,315 (9,073)	99.2 (92.7)	4.7	0.44	8.4 (29.8)			8		–/0.97	–/0.79
Tantalum,[8] peak	SSRL	1.2551	40–3.5	64,756 (5,724)	92.9 (82.9)	3.2	0.95	8.2 (25.7)	3 h	2 mM	6	11.3	0.93/0.88	0.97/1.06
Tantalum, inflection	SSRL	1.2553	40–3.5	61,506 (5,682)	88.3 (82.3)	3.1		7.3 (27.3)						
Tantalum, remote	SSRL	1.3776	40–3.5	64,624 (5,808)	92.8 (84.2)	3.2		6.6 (31.3)						
Iridium-1a	ALS	1.105	40–3.2	89,734 (8,869)	99.5 (99.6)	4.0	0.50	6.3 (30.3)	22 h	1 mM	6	16.6	0.76/0.91	1.69/0.94
Iridium-1b	SSRL	1.106	40–3.3	80,297 (6,397)	96.6 (79.0)	3.4	0.58	5.5 (21.4)	5 h	5 mM	7	16.4	0.76/0.77	1.76/1.13
Iridium-2	SSRL	1.107	25–4.0	46,373 (4,540)	99.6 (99.5)	4.3	0.68	7.5 (24.7)	22 h	1 mM	4	18.7	0.96/0.98	0.59/0.50
Mercury-a	SSRL	1.009	40–3.2	90,934 (9,064)	98.4 (99.0)	4.3	0.62	6.9 (33.8)	4 h	1 mM	7	8.9	0.92/0.98	0.76/0.71
Mercury-b	SSRL	1.009	40–3.8	55,003 (5,143)	99.4 (94.3)	5.2	0.75	8.9 (25.6)	6 h	1 mM	7	15.7	0.83/0.94	0.99/0.38
Rhenium-1	SSRL	1.181	40–4.0	45,791 (4,421)	98.4 (96.9)	3.8	0.72	9.6 (24.2)	5 h	20 mM	30	33.4	0.99/0.90	0.70/0.99
Rhenium-2a	ALS	1.176	40–3.2	89,814 (8,818)	99.6 (99.0)	4.5	0.45	8.0 (34.3)	5 h	10 mM	14	16.0	0.90/0.92	0.87/1.03
Rhenium-2b	ALS	1.176	40–3.3	79,025 (6,820)	96.0 (83.6)	3.2	0.90	6.4 (33.0)	5 h	5 mM	11	15.1	0.88/0.82	0.90/0.88
MIRAS Figure of merit (FOM) with resolution														
Resolution (Å)	39.24–8.58	8.58–6.14	6.14–5.03	5.03–4.37	4.37–3.91	3.91–3.57	3.57–3.31	3.31–3.10	Overall, 40–3.1					
FOM (centrics)	0.616	0.690	0.689	0.609	0.562	0.557	0.524	0.255	0.585					
FOM (acentrics)	0.801	0.810	0.743	0.621	0.555	0.514	0.440	0.192	0.529					

[1]Numbers following the element names indicate different heavy atom compounds. Lowercase letters indicate different soaking concentrations or soaking times, leading to differences in the numbers and occupancies of heavy atom sites. Although data sets from derivative pairs obtained in this way were correlated, additional phase information could be extracted that proved crucial for obtaining an interpretable electron density map. The heavy atom compounds used were as follows: tantalum, $Ta_6Br_{12}^{2+}$; iridium-1, chloro-pentamethylcy-clopentadienyl-1,2-bis(diphenylphosphino)ethane-iridium chloride; iridium-2, pentamethylcyclopentadienyl-iridiumchloride dimer; mercury, $Hg_3N_3C_{18}O_{12}H_{24}$, a 1, 3, 5-triazine-based compound. Although the same compound, methyltrioxorhenium, was used for rhenium-1 and rhenium-2, the observed binding sites differ, leading to independent derivatives. We believe that the compound was altered with time in solution leading to a different chemical specificity. Tantalum and iridium-2 derivatives were found previously (*29*), and gave diffraction to higher resolution in this study. [2]SSRL, beamline 9-2 at the Stanford Synchrotron Radiation Facility; ALS, beamline 5.0.2 at the Advanced Light Source at Berkeley. [3]Statistics for the highest resolution shell are given in parentheses. [4]Mosaicity was refined with SCALEPACK (*82*). [5]$R_{sym} = \Sigma_{i,h}|I(i,h) - \langle I(h) \rangle|/\Sigma_{i,h}|I(i,h)|$, where $\langle I(h) \rangle$ is the mean of the *I* observations of reflection *h*. R_{sym} was calculated with anomalous pairs merged; no sigma cut-off was applied. [6]R_{iso} = isomorphous difference = $\Sigma|F_{PH} - F_P|\Sigma F_{PH}$, where F_{PH} and F_P are the derivative and native structure factor amplitudes, respectively. [7]R_{Cullis}, mean lack of closure divided by the mean isomorphous/anomalous difference. Phasing power, mean value of heavy atom structure factor amplitudes divided by the lack of closure. The numbers given are for acentric reflections. These statistics were calculated with SHARP (*83*). [8]Owing to random orientation of the cluster, it was treated as a point scatterer and data were used to only 4.5 Å resolution. The MAD data were used for initial phasing but only the peak wavelength data were used in the final MIRAS phasing.

(Fig. 3). Each of the small subunits occurs in a single copy, arrayed around the periphery. The structure is cross-strutted by elements of Rpb1 and Rpb2 that traverse the cleft: A helix of Rpb1 bridges the cleft, and the COOH-terminal region of Rpb2 extends to the opposite side. The Rpb1-Rpb2 complex is anchored at one end by a subassembly of Rpb3, Rpb10, Rpb11, and Rpb12.

The active site was located crystallographically by replacement of the catalytic Mg^{2+} ion with Zn^{2+}, Mn^{2+}, or Pb^{2+} (40). A native zinc anomalous Fourier showed a 10-σ peak that likely results from partial replacement of the active site Mg^{2+} by Zn^{2+} during protein purification (Fig. 1), and difference Fouriers obtained from crystals soaked with either Mn^{2+} or Pb^{2+} showed a single peak at the same location (41). The metal ion site occurs within a prominent loop of Rpb1 (Fig. 3), which, on the basis of preliminary sequence assignment, harbors the conserved aspartate residue motif (42). Only one catalytic metal ion was found, and only one was reported for a bacterial RNA polymerase (43), although a two-metal ion mechanism, as described for single-subunit polymerases (44), is not ruled out.

The location of duplex DNA downstream of the active site (ahead of the transcribing polymerase) was previously determined by difference 2D crystallography of an actively transcribing complex (27). Canonical B-form DNA placed in this location lies in the Rpb1-Rpb2 cleft, and can follow a straight path to the active site (Fig. 3). About 20 base pairs are readily accommodated between the edge of the polymerase and the active site, consistent with nuclease digestion studies showing the protection of about this length of downstream DNA (45). This proposal for the pol II–DNA complex is also consistent with results of protein-DNA cross-linking experiments: Rpb1 and Rpb5
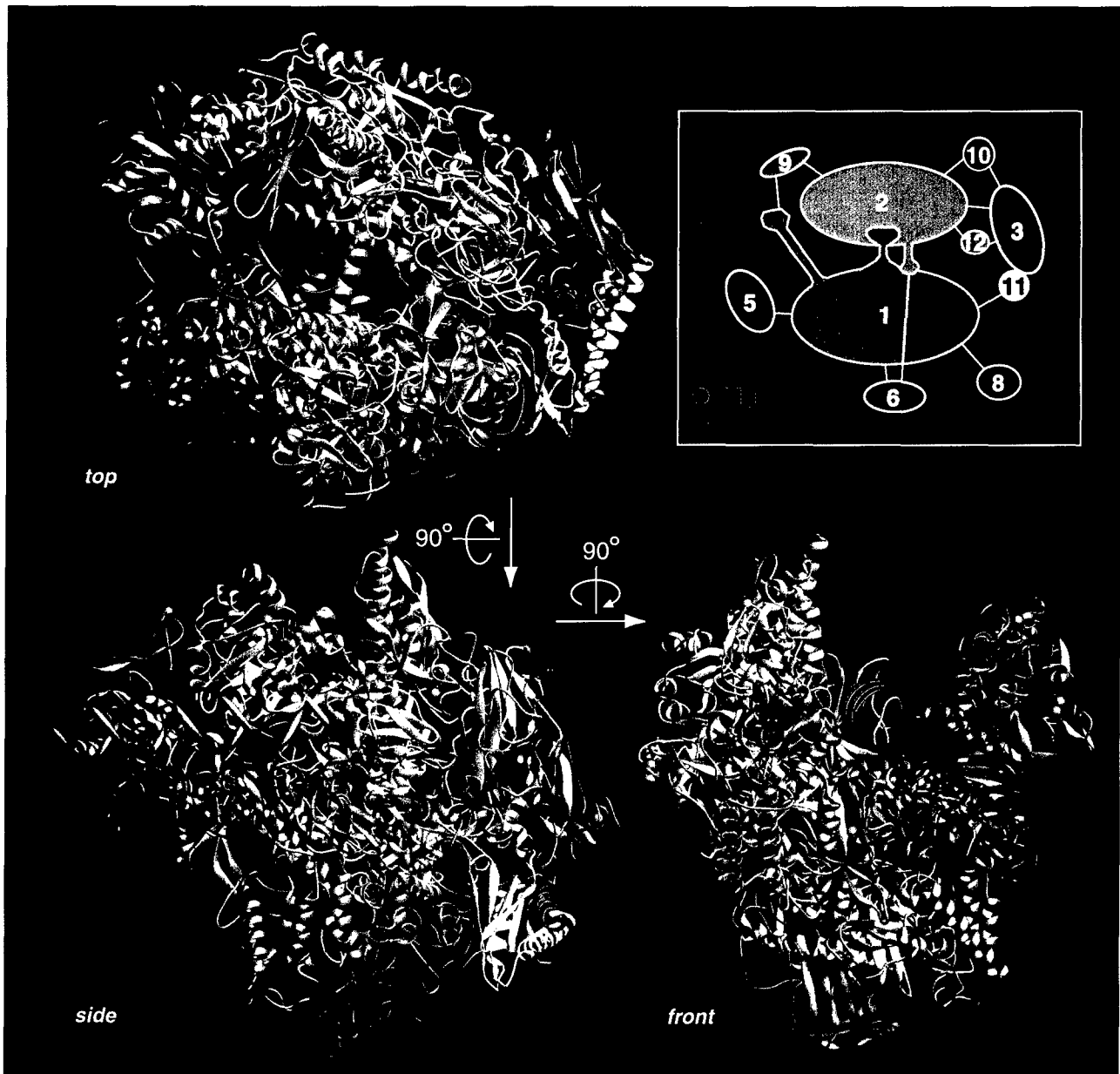

Fig. 3. Architecture of yeast RNA polymerase II. Backbone models for the 10 subunits are shown as ribbon diagrams. Secondary structure has been assigned by inspection. The three views are related by 90° rotations as indicated. Downstream DNA, though not present in the crystal, is placed onto the ribbon models as 20 base pairs of canonical B-DNA (blue) in the location previously indicated by electron crystallographic studies (27). Eight zinc atoms (blue spheres) and the active site magnesium (pink sphere) are shown (Table 1). The box (upper right) contains a key to the subunit color code and an interaction diagram. The same views and color coding are used throughout the article. This and other figures have been prepared with RIBBONS (87).

cross-link to one side of the DNA and Rpb2 to the other; and in the case of Rpb5, the cross-links are located about 5 to 15 base pairs downstream of the active site (46).

Jaws position downstream DNA. Rpb5, and regions of Rpb1 and Rpb9 on the opposite side of the Rpb1-Rpb2 cleft, form "jaws" that appear to grip the DNA (Fig. 4). Both the upper and lower jaw may be mobile, opening and closing on the DNA. Mobility within Rpb5 is suggested by comparison with the x-ray crystal structure of the subunit alone (47). There was a nearly perfect fit of the subunit structure to the corresponding region of the pol II electron density map (Fig. 2A), except for a change in relative orientation of the NH$_2$- and COOH-terminal domains, and a conformational change of a loop in the COOH-terminal domain (Fig. 4B). The solvent-exposed, NH$_2$-terminal domain (residues 1 to 142) has apparently moved by as much as 5 Å in the direction of DNA in the pol II cleft, relative to the position in Rpb5 alone, with the COOH-terminal domain (residues 143 to 215) held fixed against the body of Rpb1 (Fig. 4B). The observed position of the NH$_2$-terminal domain in pol II is defined by crystal contacts.

Residues in the Rpb5 loops facing the DNA are conserved (Fig. 4C). Two prolines that are strictly conserved present their side chains to the DNA with a spacing and relative orientation appropriate for contacting the DNA backbone. Proline residues have been seen to interact with backbone ribose moieties of DNA in other crystal structures (48, 49). Such nonspecific van der Waals interactions might favor a particular rotational setting of the DNA, without greatly impeding the helical screw rotation required to propel the DNA toward the active site and to unwind it for transcription.

Other conserved residues of Rpb5 are located in the linker between the NH$_2$- and COOH-terminal domains and in the NH$_2$-terminal helix (Fig. 4C). Since the linker is not involved in subunit-subunit interactions, conserved residues might ensure a directed movement of the NH$_2$-terminal domain. Conserved residues in the NH$_2$-terminal helix form a positive charge cluster that is too far from DNA to contact it directly, but might attract it through long-range interactions.

Rpb5 is likely to play a role in transcriptional activation (50). The NH$_2$-terminal do-

main of Rpb5 binds to the transactivation domain of the hepatitis B virus X protein (51). Another Rpb5-interacting protein interferes with transactivation (52). Some activators might function by enhancing jaw-DNA interaction, thereby stabilizing transcription initiation or elongation complexes.

The upper jaw, formed by regions of Rpb1 and Rpb9, corresponds with a domain previously shown to be mobile by 2D crystallography (53). Rpb9 is composed of two zinc-binding domains separated by a 15-residue linker. A stretch of the linker adds a β strand to a sheet in the Rpb1 region of the jaw. Rpb9 therefore buttresses Rpb1, possibly constraining mobility of the jaw and strengthening its grip on DNA. Mutations in Rpb9 alter the locations of transcription start sites (54–56), which might be explained by a diminished grip on the DNA, or alternatively, by direct Rpb9-DNA interaction before entry of the DNA into the Rpb1-Rpb2 cleft.

A clamp retains DNA. A second mobile element of pol II, previously revealed by low-resolution structures and referred to as a "hinged" domain, was suggested to clamp nucleic acids in the cleft (29). This element, here
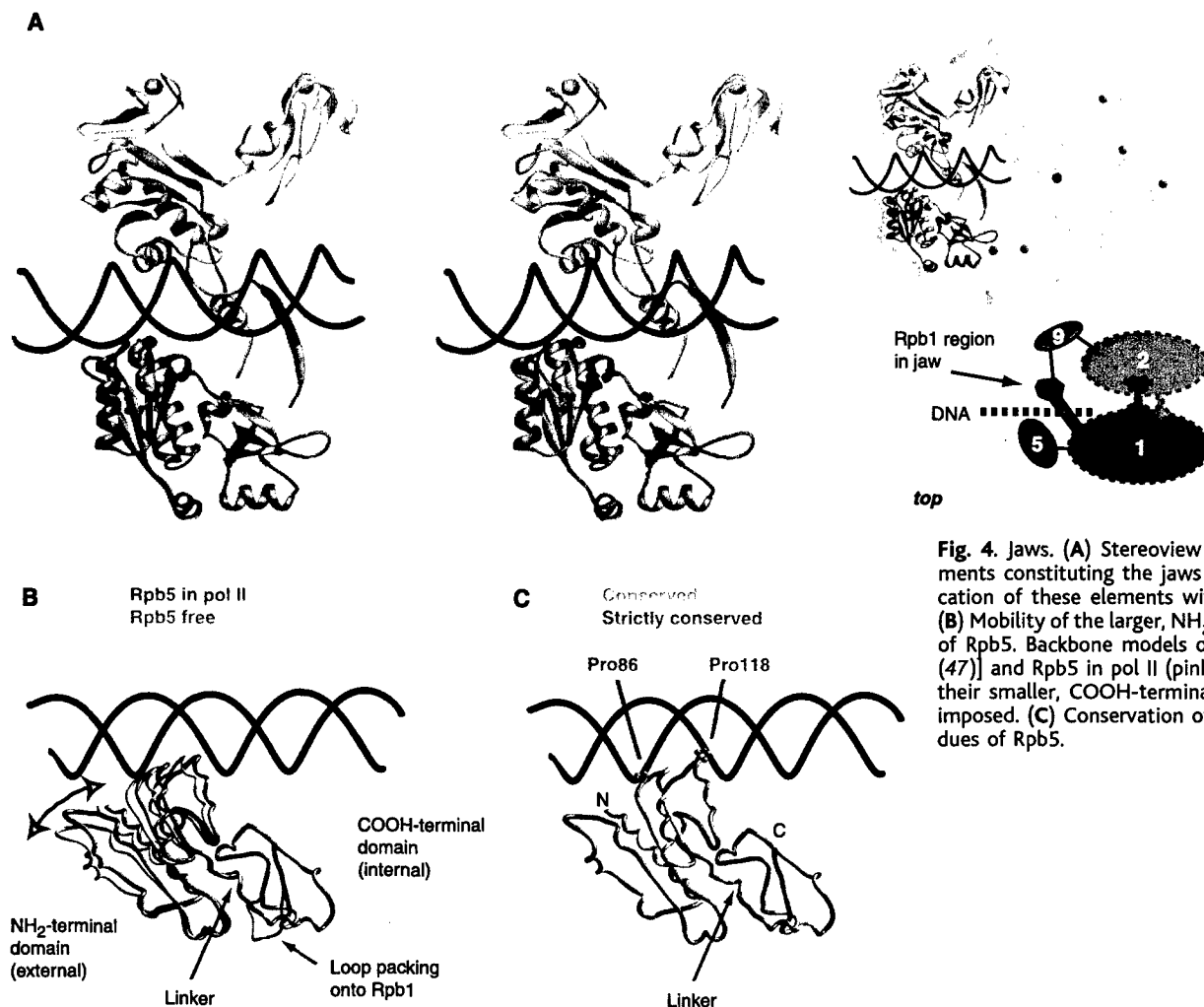

A

B
Rpb5 in pol II
Rpb5 free

NH$_2$-terminal domain (external)

Linker

COOH-terminal domain (internal)

Loop packing onto Rpb1

C
Conserved
Strictly conserved

Pro86   Pro118

N

C

Linker

Fig. 4. Jaws. (**A**) Stereoview of structural elements constituting the jaws (left) and the location of these elements within pol II (right). (**B**) Mobility of the larger, NH$_2$-terminal domain of Rpb5. Backbone models of free Rpb5 [gray (47)] and Rpb5 in pol II (pink) are shown with their smaller, COOH-terminal domains superimposed. (**C**) Conservation of amino acid residues of Rpb5.

Rpb1 region in jaw →

DNA

top

termed the "clamp," comprises NH$_2$-terminal regions of Rpb1 and Rpb6 and the COOH-terminal region of Rpb2 (Fig. 5). All three polypeptides enter at the base of the clamp near the active site, allowing a degree of conformational freedom but not unrestricted movement of the clamp. Within the Rpb6 region, 17 out of 42 residues are negatively charged, forming a cluster near the bottom of the clamp. This region of Rpb6 is also phosphorylated by casein kinase II, suggesting a regulatory role (*57*).

The clamp forms one side of the Rpb1-Rpb2 cleft, where it may interact with the DNA (and the DNA-RNA hybrid, see below) from the active site to about 15 residues downstream. This DNA region corresponds with a double-stranded DNA binding site, 3 to 12 residues downstream of the active site, defined by biochemical analysis of *E. coli* RNA polymerase (*58–60*). This binding site was referred to as a "sliding clamp" because of its importance for the great stability of a transcribing complex and processivity of transcription (*60*). Closure of the clamp over the DNA could account for this stability. Such a movement of the NH$_2$-terminal region of the largest subunit was inferred from cross-linking studies of the *E. coli* enzyme (*58*). Although the clamp is seen here in an open conformation, it is involved in crystal contacts and the observed position is likely determined by the crystal lattice. The electron density in this region is of lower quality than elsewhere in the map, and the three zinc peaks associated with the region have the lowest heights (Zn6-8, Table 1), also consistent with mobility of the clamp.

DNA-RNA hybrid binding site, RNA binding site. Transcribing polymerases have been shown to harbor an unwound region of DNA, or "bubble," within which is centered a DNA-RNA hybrid of 8 or 9 base pairs, with the 3' or growing end of the RNA at the active site (Fig. 6A) (*60*). Linear extension of duplex DNA placed in our crystallographic model, to accommodate the DNA-RNA hybrid, is impossible because of an element from Rpb2 blocking the path (Figs. 3, 4, and 6). This blocking element corresponds with a "wall" of density previously noted in the structure of bacterial RNA polymerase (*43*). Because of the wall, and because the active site lies well beneath the level of the downstream DNA, the DNA-RNA hybrid must be tilted relative to the axis of the downstream DNA (dashed line in Fig. 6C). The exact orientation of the hybrid remains to be determined.

At the upstream end of the DNA-RNA hybrid (5' end of the RNA, remote from the active site), the strands must separate. Biochemical studies show that the RNA strand enters a binding site on the protein, extending from about 10 to 20 nucleotides upstream of the active site (*61*). There are two prominent grooves in the pol II structure exiting the hybrid binding site, each of which could accommodate

one, but not two, nucleic acid strands. One groove winds around the base of the clamp (Fig. 7, groove 1). The other is between the lower part of the wall and Rpb1, and continues downward between Rpb1 and Rpb11 (Fig. 7, groove 2). We favor groove 1 as the RNA

Fig. 5. Clamp. Structural elements constituting the clamp and their location in pol II are shown. The COOH-terminal region of Rpb2 and the NH$_2$-terminal region of Rpb1 bind one and two zinc ions, respectively (blue spheres). The NH$_2$-terminal tail region of Rpb6 extends from its main body (at the bottom in the front view) into the clamp. The direction of movement of the clamp revealed by comparison with electron crystal structures (*29*) is indicated (double-headed red arrow).
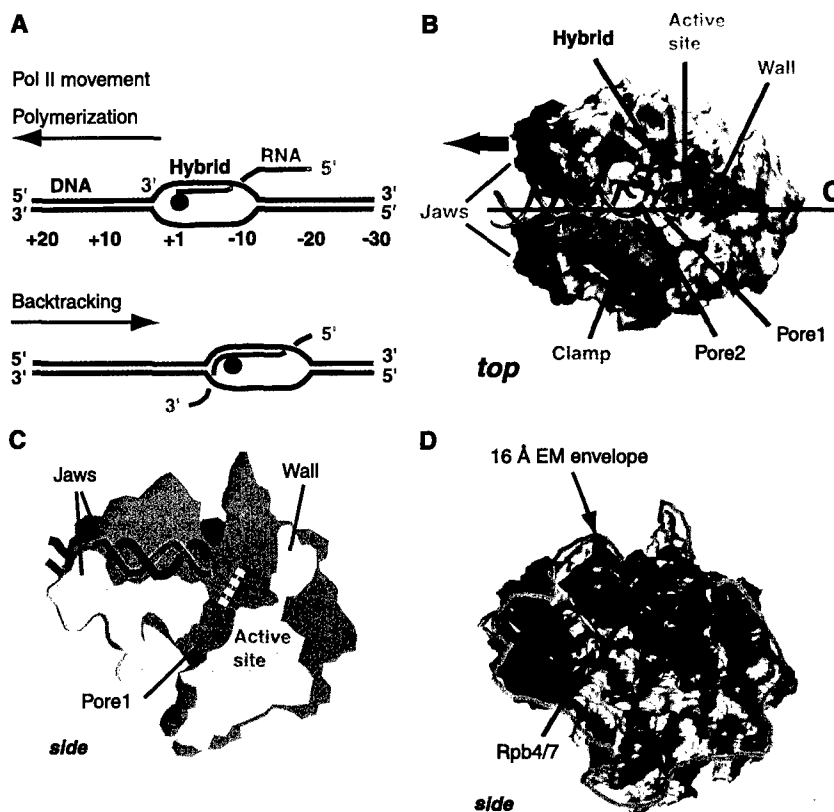

Fig. 6. Topology of the polymerizing complex, and location of Rpb4 and Rpb7. (**A**) Nucleic acid configuration in polymerizing (top) and backtracking (bottom) complexes. (**B**) Structural features of functional significance and their location with respect to the nucleic acids. A surface representation of pol II is shown as viewed from the top in Fig. 3. To the surface representation has been added the DNA-RNA hybrid, modeled as nine base pairs of canonical A-DNA (DNA template strand, blue; RNA, red), positioned such that the growing (3') end of the RNA is adjacent to the active site metal and clashes with the protein are avoided. The exact orientation of the hybrid remains to be determined. The nontemplate strand of the DNA within the transcription bubble, single-stranded RNA and the upstream DNA duplex are not shown. (**C**) Cutaway view with schematic of DNA (blue) and with the helical axis of the DNA-RNA hybrid indicated (dashed white line). An opening in the floor of the cleft that binds nucleic acid exposes the DNA-RNA hybrid (pore 1) to the inverted funnel-shaped cavity below. The plane of section is indicated by a line in (B), and the direction of view perpendicular to this plane (side) is as in Fig. 3. (**D**) Surface representation as in (B), with direction of view as in (C). The molecular envelope of pol II determined by electron microscopy of 2D crystals at 16 Å resolution is indicated (yellow line), as is the location of subunits Rpb4 and Rpb7 (arrow, Rpb4/7), determined by difference 2D crystallography (*25*).

binding site for three reasons. First, the length and location of the groove are appropriate for binding a region of RNA 10 to 20 nucleotides from the active site, in agreement with biochemical studies. Second, the RNA path would lead back toward the downstream DNA, ending in close proximity to the NH_2-terminal region of Rpb1 (defined by a zinc site). This path would accord with the reported cross-linking of RNA about 20 nucleotides upstream of the active site to the NH_2-terminal region of the largest subunit of *E. coli* RNA polymerase (*58–60*). Finally, RNA in the groove at the base of the clamp could explain the great stability of transcribing complexes. The affinity of the polymerase for the DNA template is coupled to the presence of an RNA transcript (*60*). We speculate that closure of the clamp over DNA, assuring its retention in a transcribing complex, would enlarge the groove at the base of the clamp, and subsequent binding of RNA in the groove would prevent the clamp from reopening. RNA would act as a lock on the closed conformation of the clamp.

Mobility of the clamp may also be modulated by interactions with other pol II subunits and transcription factors, for example, Rpb4 and Rpb7. Although these two small subunits were absent from the form of pol II analyzed here, their approximate location is known from electron microscopy of 2D crystals (*25*). A surface representation of the crystallographic backbone model corresponds closely with the molecular envelope from 2D crystals (Fig. 6D). On this basis, Rpb4 and Rpb7 occupy a crevice in the surface between the lower jaw and the clamp (Fig. 6D). Interaction with either of these mobile elements or with downstream DNA could underlie the requirement for Rpb4 and Rpb7 for the initiation of transcription (*22*).

A funnel for substrate entry, backtracking, and elongation factor access. The floor of the Rpb1-Rpb2 cleft, which supports duplex DNA and the DNA-RNA hybrid, is very thin and perforated, exposing the nucleic acids to the space below. The perforation is bisected by the helix that forms a bridge between Rpb1 and Rpb2, creating two pores, one of which lies beneath the active site (pore 1) and the other, beneath the downstream DNA (pore 2). Both pores are about 12 Å in diameter and lie at the apex of an inverted funnel-shaped cavity, which increases to about 30 Å in diameter at the opposite side of pol II (Fig. 7, bottom). As the Rpb1-Rpb2 cleft is occupied by duplex DNA and the DNA-RNA hybrid during transcription, nucleotides may be unable to enter above the active site and may instead gain access from below, through the funnel and pore 1, as previously suggested for both pol II and bacterial RNA polymerase (*29, 43*).

The funnel and pore 1 may play similar roles in other aspects of transcription. Bacterial and eukaryotic RNA polymerases oscillate between forward (polymerization) and backward (backtracking) movement during transcription (Fig. 6A) (*60*). Backtracking is important for proofreading and for traversing obstacles such as DNA damage, bound proteins, or natural pause sites in the DNA. During backtracking, the polymerase and associated transcription bubble move backward along both the DNA and the RNA. The region engaged in the DNA-RNA hybrid retreats like a zipper, releasing the 3′ end of the RNA in single-stranded form, and incorporating single-stranded RNA on the 5′ side of the transcription bubble into the hybrid (Fig. 6A). As mentioned above for access of nucleotides to the active site during polymer-

ization, duplex DNA and hybrid in the Rpb1-Rpb2 cleft may block release of the 3′ end of the RNA into the cleft during backtracking. Rather, as suggested for entry of nucleotides, the 3′ end of the RNA may exit through the funnel and pore 1.

Backtracking beyond a certain point can result in an arrested complex, unable to reverse direction, to restore the 3′ end of the RNA to the active site, and to resume transcription (*60*). We speculate that when a certain length of RNA has been extruded by backtracking, it may interact with a site in the funnel and be trapped, preventing reversal and recovery. For recovery from arrest, cleavage of the RNA is required to generate a new 3′ end at the active site (*60*). This cleavage is achieved with the help of transcript cleavage factors (*62, 63*). The funnel and pore 1 may provide access for such factors, for example, TFIIS. A small zinc-binding domain of TFIIS has an extended β hairpin at one end with two conserved residues that come near the active site of pol II and that are critical for RNA cleavage (*15, 16, 64–66*). Also included are tryptophan and arginine side chains involved in nucleic acid binding (*67, 68*). Modeling shows that this domain, only 20 Å in diameter, can be accommodated in pore 1 with the two conserved β hairpin residues reaching the active site, while still leaving room for an extruded strand of RNA.

Comparison with bacterial RNA polymerase. Most information about core bacterial RNA polymerase structure comes from x-ray diffraction studies of the α_2 homodimer from *E. coli* (*69*) and the $\alpha_2\beta\beta'$ polymerase from *Thermus aquaticus* (*43*). Regions of sequence similarity have been noted between α, Rpb3, and Rpb11 (*69*), between β and
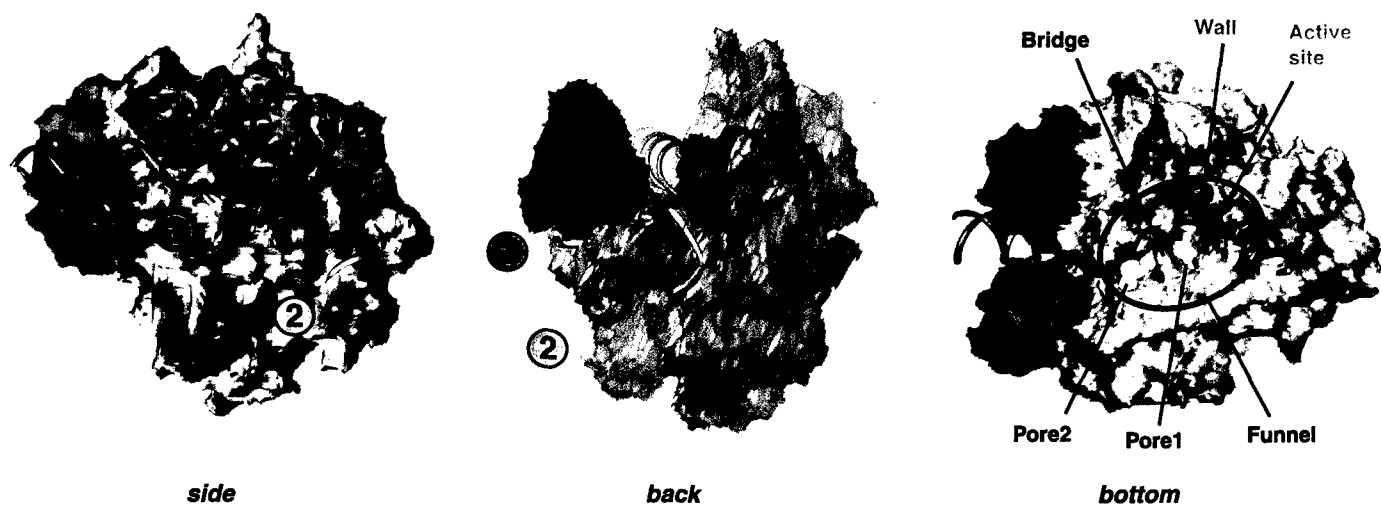

Fig. 7. Possible RNA exit grooves and funnel beneath the active site. The model of Fig. 6B is shown in two perpendicular directions of view (side, back), and also viewed from the opposite side (bottom). To the side and back views have been added dashed lines corresponding to about 10 nucleotides of RNA, lying in well-defined grooves leading away from the hybrid-binding region (groove 1, red; groove 2, orange). The nontemplate strand of the DNA within the transcription bubble and the upstream DNA duplex are not shown. To the bottom view has been added a solid line indicating the rim of the funnel-shaped cavity.

Rpb2 (70), and between β' and Rpb1 (71). The crystallographic pol II model contains a conserved core of secondary structural elements similar to those in the bacterial enzyme, surrounded by divergent elements and eukaryote-specific subunits. Conserved elements are located in the vicinity of the DNA-RNA hybrid binding site, the adjacent downstream DNA binding site, and the sides of the funnel. Consistent with the conservation of these structural elements, similar modes of interaction with nucleic acids in the vicinity of the active site have been proposed for the eukaryotic and bacterial enzymes (72). The pore beneath the active site is conserved, and the bacterial enzyme may contain a clamp as well (73). On the other hand, the jaws, which include eukaryote-specific subunits and a domain of Rpb1, are found only in pol II, possibly reflecting their interaction with the eukaryote-specific transcription initiation factor TFIIE, as revealed by 2D crystallography (26). The occurrence of jaws in pol II, but not in the bacterial enzyme, presumably accounts for the nuclease protection of about 20 base pairs of downstream DNA by pol II, compared with only about 13 base pairs by the bacterial enzyme (45, 60).

A more detailed comparison is possible, at present, for the α₂ dimer and its counterpart in pol II, the Rpb3-Rpb11 heterodimer. The α₂ dimer nucleates assembly of bacterial polymerase, binding β to form a subcomplex, which then binds β' to form a complete core enzyme (74). Similarly, the Rpb3-Rpb11 heterodimer binds Rpb2 to form a subcomplex (75). The location of the heterodimer in pol II is similar to that of α₂ in the bacterial enzyme, and the domain conserved between Rpb3, Rpb11, and α exhibits an identical fold (motif of α helices and β sheets forming the lower half of the subcomplex in Fig. 8). The conserved domain represents almost the entirety of Rpb11 and is responsible for Rpb3-Rpb11 interaction (or dimerization in the case of α). The nonconserved domain of Rpb3 (upper half of the subcomplex in Fig. 8) interacts with the eukaryote-specific subunits Rpb10 and Rpb12. Contact of Rpb10 with Rpb3 is consistent with biochemical evidence for a stable Rpb3-Rpb11-Rpb10 subcomplex (76). Rpb12 binds through a tail, which adds a β strand to a sheet in the nonconserved region of Rpb3. Rpb12 also interacts

with Rpb2 through its zinc-binding module. Consistent with this, Rpb12 has been shown to contact the second largest subunit in RNA polymerase I, and this interaction requires an intact zinc-binding motif (77). Moreover, a mutation in the COOH-terminal region of Rpb12 impairs assembly of RNA polymerase III (77). Thus, Rpb12 appears to play an essential role in the assembly or maintenance of all eukaryotic RNA polymerases by bridging between the Rpb3-Rpb11-Rpb10 subcomplex (or its homologs in polymerases I and III) and the second largest subunit.

Transcription pathway. The crystallographic model of pol II also gives insight into the transcription pathway and the still larger multiprotein complexes involved. The pathway begins with the formation of a TFIIB–TFIID–promoter DNA complex and its interaction with pol II, followed by entry of TFIIE, and finally TFIIH, whose helicase activities melt DNA around the start site of transcription. The initial interaction of pol II with the promoter must be with essentially straight, duplex DNA. The pol II model, however, requires a considerable distortion for binding at the active site, which can only occur upon melting. The transition from an initial complex to a transcribing complex will therefore be accompanied by structural changes and movement of the DNA. Transcription begins with the repeated synthesis and release of short RNAs ("abortive cycling"), until a barrier at about 10 nucleotides is traversed, and chain elongation ensues. On reaching a transcript size of about 20 nucleotides, the full stability of a transcribing complex is attained. The barrier at 10 nucleotides corresponds to the point at which the 5' end of the growing transcript must disengage from the template DNA and enter the proposed groove for RNA in the model. The transcript size needed for full stability corresponds with the length of RNA needed to fill the groove.

The interpretation along these lines may be extended and evaluated by the solution of pol II cocrystal structures, with the use of the pol II model for molecular replacement. Cocrystals with TFIIB and TFIIE (78) should reveal the trajectory of DNA in the initial pol II–promoter complex. Cocrystals containing pol II in the act of transcription (79) will show the locations of nucleic acids in an elongation complex. Cocrystals with TFIIS (80) may indicate the proposed exit pathway for RNA through a pore beneath the active site during backtracking. Other cocrystals may be sought to investigate the mechanism of transcriptional regulation by the multiprotein Mediator complex and associated activator and repressor proteins (4).
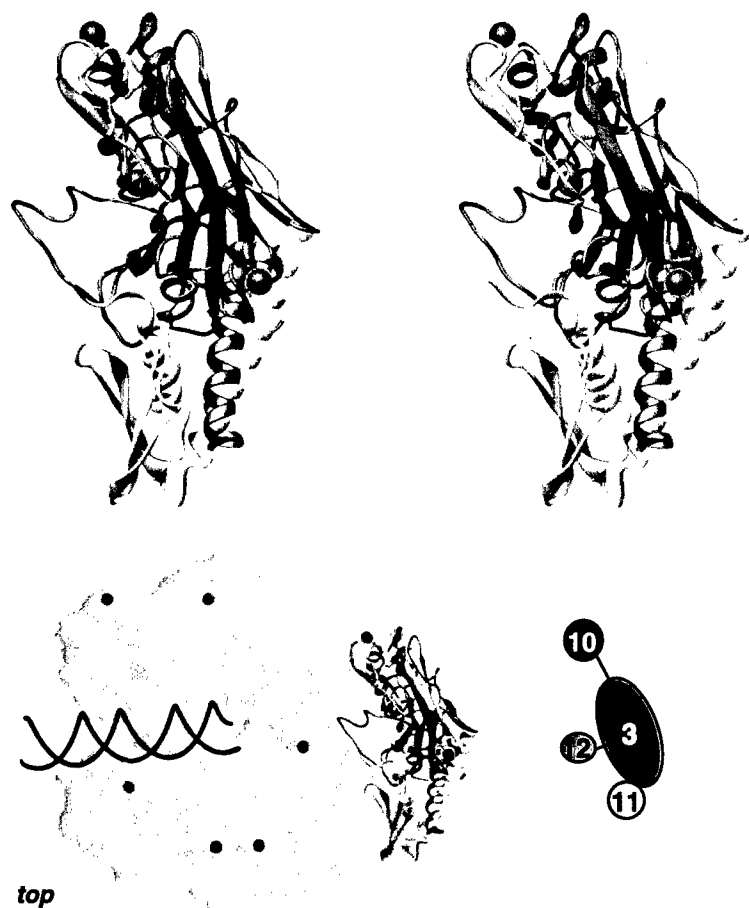

Fig. 8. The Rpb3-Rpb11-Rpb10 subcomplex and Rpb12. A stereoview of the arrangement of the four subunits is shown in the upper part, and the location of this subcomplex within pol II is shown in the lower part.

References and Notes
1. N. A. Woychik, Cold Spring Harbor Symp. Quant. Biol. **63**, 311 (1998).
2. R. G. Roeder, Trends Biochem. Sci. **21**, 327 (1996).

3. R. C. Conaway and J. W. Conaway, *Prog. Nucleic Acid Res. Mol. Biol.* **56**, 327 (1997).
4. Y. J. Kim, S. Bjorklund, Y. Li, M. H. Sayre, R. D. Kornberg, *Cell* **77**, 599 (1994).
5. G. Otero *et al.*, *Mol. Cell* **3**, 109 (1999).
6. J. L. Corden and M. Patturajan, *Trends Biochem. Sci.* **22**, 413 (1997).
7. J. L. Kim, D. B. Nikolov, S. K. Burley, *Nature* **365**, 520 (1993).
8. D. B. Nikolov *et al.*, *Nature* **377**, 119 (1995).
9. Y. Kim, J. H. Geiger, S. Hahn, P. B. Sigler, *Nature* **365**, 512 (1993).
10. P. Kosa, G. Ghosh, B. S. DeDecker, P. B. Sigler, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6042 (1997).
11. W. Zhu *et al.*, *Nature Struct. Biol.* **3**, 122 (1996).
12. F. del Rio-Portilla, A. Gaskell, D. Gilbert, J. A. Ladias, G. Wagner, *Nature Struct. Biol.* **6**, 1039 (1999).
13. S. Krapp, G. Kelly, J. Reischl, R. O. J. Weinzierl, S. Matthews, *Nature Struct. Biol.* **5**, 110 (1998).
14. B. Wang, D. N. Jones, B. P. Kaine, M. A. Weiss, *Structure* **6**, 555 (1998).
15. X. Qian *et al.*, *Biochemistry* **32**, 9944 (1993).
16. V. K. Olmsted *et al.*, *J. Biol. Chem.* **273**, 22589 (1998).
17. P. E. Morin, D. E. Awrey, A. M. Edwards, C. H. Arrowsmith, *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 10604 (1996).
18. A. Sentenac *et al.*, "Yeast RNA polymerase subunits and genes," in *Transcriptional Regulation*, S. L. McKnight and K. R. Yamamoto, Eds. (Monogr. 22, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992).
19. N. F. Lue and R. D. Kornberg, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8839 (1987).
20. E. E. Uzgiris and R. D. Kornberg, *Nature* **301**, 125 (1983).
21. A. M. Edwards *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2122 (1990). The key step in pol II purification was immunoaffinity chromatography with the use of a monoclonal antibody from which pol II could be released by an elevated concentration of glycerol [N. E. Thompson, D. B. Aronson, R. R. Burgess, *J. Biol. Chem.* **265**, 7069 (1990)].
22. A. M. Edwards, C. M. Kane, R. A. Young, R. D. Kornberg, *J. Biol. Chem.* **266**, 71 (1991).
23. S. A. Darst, E. W. Kubalek, A. M. Edwards, R. D. Kornberg, *J. Mol. Biol.* **221**, 347 (1991).
24. S. A. Darst, A. M. Edwards, E. W. Kubalek, R. D. Kornberg, *Cell* **66**, 121 (1991).
25. G. J. Jensen, G. Meredith, D. A. Bushnell, R. D. Kornberg, *EMBO J.* **17**, 2353 (1998).
26. K. K. Leuther, D. A. Bushnell, R. D. Kornberg, *Cell* **85**, 773 (1996).
27. C. L. Poglitsch *et al.*, *Cell* **98**, 791 (1999).
28. A. M. Edwards, S. A. Darst, S. A. Hemming, Y. Li, R. D. Kornberg, *Nature Struct. Biol.* **1**, 195 (1994).
29. J. Fu *et al.*, *Cell* **98**, 799 (1999).
30. Purification of yeast pol II and crystallization in the orthorhombic space group I222 were as described [L. Myers *et al.*, *Methods Companion Methods Enzymol.* **12**, 212 (1997); (*29*)]. The crystals, containing a single pol II in the asymmetric unit, were transferred under argon/hydrogen in seven steps from harvest buffer [390 mM $(NH_4)_2HPO_4$/NaH_2PO_4, pH 6.0, 16% PEG 6000, 50 mM dioxane, and 3 mM dithiothreitol (DTT)] to stabilization buffer (100 mM MES, pH 6.3, 16% PEG 6000, 350 mM NaCl, 17% PEG 400, 50 mM dioxane, and 3 mM DTT). Crystals were cooled to 4°C overnight and maintained at that temperature for 5 days before flash-cooling. This treatment caused shrinkage along the crystallographic a axis to 131 Å, extended the diffraction limit, and led to high isomorphism of the crystals. Best results were obtained with crystals of about 200 × 150 × 50 μm. Crystals were mounted at 4°C in nylon loops 200 μm in diameter, plunged into liquid nitrogen and stored for data collection.
31. Data collection was carried out at 100 K using ADSC Quantum 4-CCD detectors at the tunable high brilliance undulator beamlines 9-2 at the Stanford Synchrotron Radiation Laboratory (SSRL) and 5.0.2 at the Advance Light Source (ALS) at Lawrence Berkeley National Laboratory (Table 2). All derivative data sets were collected at the anomalous peak energy of the heavy metal to maximize the anomalous signal. Care was taken to ensure completeness and reasonable redundancy in all resolution shells. Crystals were aligned with the long c axis approximately along the spindle axis to minimize overlap of reflections and to record anomalous pairs within the shortest period of time. Further, overlap of reflections at a resolution higher than 4 Å was avoided by using crystals with low mosaicity and 0.5° oscillation steps. Diffraction data were processed with DENZO and SCALEPACK (*82*)].
32. For heavy atom derivatization, crystals were soaked at 4°C in stabilization buffer containing heavy atom compounds as indicated (Table 2). MAD phases were calculated with MLPHARE (*88*) using three tantalum data sets collected at the anomalous peak, inflection, and a remote wavelength (Table 2).
33. Single heavy atom sites were located by isomorphous and anomalous cross-difference Fourier analysis. New derivatives could best be identified in anomalous difference Fourier maps, most likely because the low signal arising from the addition of only a few heavy atoms to a large unit cell is best detected in the absence of noise resulting from nonisomorphism.
34. Standard heavy atom compounds were too reactive and destroyed diffraction beyond 6 Å resolution. About 600 crystals were treated with heavy atom compounds, and a total of 60 data sets were collected to obtain those used in phasing.
35. Initially, heavy atom parameters were refined and MIRAS phases were calculated with MLPHARE (*88*). The molecular boundary was determined by the automatic procedure implemented in DM [K. Cowtan, *Joint CCP4 and ESF-EACBM Newsl. Protein Crystallogr.* **31**, 34 (1994)], and the phases were further improved by solvent flattening and histogram matching in DM. These phases revealed lower occupancy sites that were subsequently included. Parameters were further refined and final MIRAS phases were calculated with SHARP (*83*) (Table 2). The resulting phases were improved by density modification with DM.
36. Electron density maps were viewed and model building carried out with the program O [T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991)].
37. The conserved core of the *E. coli* α homodimer comprises two pairs of helices flanked by two β sheets and is responsible for dimerization (*69*). This fold motif was one of the first interpretable features of early pol II maps and was ascribed to the Rpb3-Rpb11 heterodimer. Rpb3 contains, in addition, a second domain (residues 43 to 162), whose tracing was greatly facilitated by the presence of a zinc-binding motif (residues 86 to 95). The location of residue Cys207 in a long loop agrees with the position of a mercury binding site (Table 1). The overall fold of the second domain is similar to that in the nonconserved domain of *E. coli* α although only small parts show limited sequence similarity. The core of the structure obtained by nuclear magnetic resonance (NMR) of the Rpb10 homolog from *Methanobacterium thermoautotrophicum* (residues 1 to 48 of the yeast protein) (C. D. Mackereth, C. H. Arrowsmith, A. M. Edwards, L. P. McIntosh, unpublished data) could be fitted into the pol II density in close proximity to the Rpb3-Rpb11 heterodimer with a zinc atom in the model coincident with an experimental zinc peak (Fig. 2B). The COOH-terminal region of Rpb10 could be modeled as polyalanine (Fig. 2B). Two domains in the crystal structure of yeast Rpb5 (residues 1 to 142 and 143 to 215) (F. Todone, R. O. J. Weinzierl, P. Brick, S. Onesti, unpublished data) were fitted to the electron density as rigid bodies (Fig. 2A). Except for residues 197 to 210, no rebuilding was carried out. Residue Cys83 coincided with a mercury binding site (Table 1). A backbone model of residues 51 to 59 and 78 to 155 (yeast numbering) from the NMR structure of human Rpb6 (*12*) was placed below the clamp. The model was corrected for a slight change in the angle between the two helices and partially rebuilt. Following continuous density, the model was extended at the NH_2-terminus by residues that are flexible in free Rpb6(*12*), but which become ordered in pol II. The β barrel fold of yeast Rpb8 determined by NMR (*13*) was easily detected at an exposed position in pol II that is not involved in crystal packing. Rpb8 contains two cysteine residues (Cys24 and Cys36), which lie within one sheet. Both are derivatized by mercury compounds (Table 1), confirming the location of Rpb8 and defining the orientation of the pseudosymmetric subunit. The NMR structure of the COOH-terminal domain of Rpb9 (*14*) could be placed with the aid of an experimental zinc peak. An extended density connecting to another domain that also contains an experimental zinc peak was identified as the interdomain linker. The linker length corresponds to the distance required to place the cysteines of the NH_2-terminal domain at the zinc position.
38. A first backbone model included regions in all subunits and contained a total of 2753 residues. Phases calculated from this preliminary model were recombined with the experimental MIRAS phases using SIGMAA (*88*). The electron density map obtained by application of these combined phases and density modification showed novel features such as continuous density where chain breaks were previously encountered. This map was used to adjust and complete the backbone model. Another phase combination step resulted in a further improved map in which no errors in the backbone tracing could be detected. A total of 18 polyalanine fragments were built for Rpb1 and Rpb2, ranging in length from 13 to 666 residues. No structural information was available for the smallest subunit Rpb12, which binds zinc. After seven out of the eight zinc ions in pol II were assigned to other subunits, a distinct density could be assigned for Rpb12 located around a zinc position on the outer surface of pol II, facing away from the cleft.
39. T. Miyao, A. Honda, Z. Qu, A. Ishihama, *Mol. Gen. Genet.* **259**, 123 (1998).
40. The active site metal ion in RNA polymerases can be replaced by Mn^{2+} [Y. Huang, F. Eckstein, R. Padilla, R. Sousa, *Biochemistry* **36**, 8231 (1997)], as well as Pb^{2+} (*43*). Further, Mn^{2+} and Zn^{2+} have been used to specifically replace catalytic magnesium ions in DNA polymerases [Y. Kim *et al.*, *Nature* **376**, 612 (1995); S. Doublie, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *Nature* **391**, 251 (1998); L. S. Beese and T. A. Steitz, *EMBO J.* **10**, 25 (1991); V. Derbyshire *et al.*, *Science* **240**, 199 (1988)].
41. A crystal was soaked overnight with 10 mM $MnCl_2$. The anomalous peak of manganese at 6539 eV is not accessible experimentally. However, the theoretical signal at the accessible x-ray energy of 9600 eV, in the tail region of the peak, amounts to about two electrons. Diffraction data were collected at this energy, below the absorption edge of zinc where the only measurable anomalous signal derives from manganese (P. Cramer *et al.*, data not shown). The experimental phases were of sufficient quality to reveal a single peak of height 6.2 σ in anomalous difference Fouriers. Another crystal was soaked in 1 mM $Pb(OAc)_2$ for 2 hours, and diffraction data were collected at the anomalous peak wavelength for lead. The location of the 15.6 σ peak derived from these data precisely matches the manganese site.
42. E. Zaychikov *et al.*, *Science* **273**, 107 (1996).
43. G. Zhang *et al.*, *Cell* **98**, 811 (1999).
44. T. A. Steitz, *Nature* **391**, 231 (1998).
45. G. A. Rice, J. Chamberlin, C. M. Kane, *Nucleic Acids Res.* **21**, 113 (1993).
46. T.-K. Kim *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12268 (1997).
47. F. Todone, R. O. J. Weinzierl, P. Brick, S. Onesti, unpublished data.
48. C. W. Müller, F. A. Fey, M. Sodeoka, G. L. Verdine, S. C. Harrison, *Nature* **373**, 311 (1995).
49. P. Cramer, C. J. Larson, G. L. Verdine, C. W. Müller, *EMBO J.* **16**, 7078 (1997).
50. T. Miyao and N. A. Woychik, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15281 (1998).
51. J. H. Cheong, M. Yi, Y. Lin, S. Murakami, *EMBO J.* **14**, 143 (1995).
52. D. Dorjsuren *et al.*, *Mol. Cell. Biol.* **18**, 7546 (1998).
53. F. Asturias, G. Meredith, C. Poglitsch, R. Kornberg, *J. Mol. Biol.* **272**, 536 (1997).
54. E. M. Furter-Graves, R. Furter, B. D. Hall, *Mol. Cell. Biol.* **11**, 4121 (1991).
55. E. M. Furter-Graves, B. D. Hall, R. Furter, *Nucleic Acids Res.* **22**, 4932 (1994).

56. W. W. Hull, K. McKune, N. A. Woychik, *Genes Dev.* **9**, 481 (1995).
57. K. Kayukawa, Y. Makino, S. Yogosawa, T. Tamura, *Gene* **234**, 139 (1999).
58. E. Nudler, E. Avetissova, V. Markovtsov, A. Goldfarb, *Science* **273**, 211 (1996).
59. E. Nudler, I. Gusarov, E. Avetissova, M. Kozlov, A. Goldfarb, *Science* **281**, 424 (1998).
60. E. Nudler, *J. Mol. Biol.* **288**, 1 (1999).
61. G. A. Rice, C. M. Kane, M. J. Chamberlin, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4245 (1991).
62. M. G. Izban and D. S. Luse, *Genes Dev.* **6**, 1342 (1992).
63. T. L. Johnson and M. J. Chamberlin, *Cell* **77**, 217 (1994).
64. W. Powell, B. Bartholomew, D. Reines, *J. Biol. Chem.* **271**, 22301 (1996).
65. D. E. Awrey et al., *J. Biol. Chem.* **273**, 22595 (1998).
66. C. Jeon, H. Yoon, K. Agarwal, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9106 (1994).
67. K. Agarwal et al., *Biochemistry* **30**, 7842 (1991).
68. H. Yoon, A. S. Sitikov, C. Jeon, K. Agarwal, *Biochemistry* **37**, 12104 (1998).
69. G. Zhang and S. A. Darst, *Science* **281**, 262 (1998).
70. D. Sweetser, M. Nonet, R. A. Young, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1192 (1987).
71. R. S. Jokerst, J. R. Weeks, W. A. Zehring, A. L. Greenleaf, *Mol. Gen. Genet.* **215**, 266 (1989).
72. The proposed location of downstream DNA and the DNA-RNA hybrid in the crystallographic model of pol II corresponds well with that previously put forward and recently refined on the basis of protein-nucleic acid cross-linking data for the bacterial enzyme [(*43*) and N. Korzheva et al., manuscript submitted]. The proposed location of downstream DNA is also the same as that previously shown in 15 Å electron and 6 Å x-ray maps of pol II (*27*, *29*), but the location of the DNA-RNA hybrid proposed here differs from that suggested on the basis of the lower resolution data. A conjecture about the location of the active site in the previous maps placed the hybrid at the downstream edge of the clamp. The active site and thus the hybrid

are now definitively placed near the upstream edge of the clamp.
73. The bacterial enzyme appears to contain a counterpart of the pol II clamp, on the basis of following three observations. First, the zinc-binding NH_2-terminal region of β′ and the COOH-terminal region of β form a distinct structural element, just as do the corresponding regions of Rpb1 and Rpb2, respectively, in the pol II clamp. Second, these regions include motifs conserved between the bacterial and eukaryotic enzymes. Finally, the relative position of this structural element with respect to the active site and central cleft is the same in both enzymes.
74. A. Ishihama, *Adv. Biophys.* **14**, 1 (1981).
75. M. Kimura, A. Ishiguro, A. Ishihama, *J. Biol. Chem.* **272**, 25851 (1997).
76. D. Lalo, C. Carles, A. Sentenac, P. Thuriaux, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5524 (1993).
77. L. Rubbi, S. Labarre-Mariotte, S. Chedin, P. Thuriaux, *J. Biol. Chem.* **274**, 31485 (1999).
78. D. A. Bushnell, K. Leuther, R. D. Kornberg, unpublished data.
79. A. L. Gnatt, J. Fu, R. D. Kornberg, *J. Biol. Chem.* **272**, 30799 (1997).
80. A. L. Gnatt and R. D. Kornberg, unpublished data.
81. A. M. Edwards and C. Mackereth, unpublished data.
82. Z. Otwinowski and W. Minor, *Methods Enzymol.* **276**, 307 (1996).
83. E. d. la Fortelle and G. Bricogne, *Methods Enzymol.* **B**, 472 (1997).
84. G. J. Kleywegt and T. A. Jones, *Acta Crystallogr.* **D52**, 826 (1996).
85. R. M. Esnouf, *J. Mol. Graphics* **15**, 132 (1997).
86. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
87. M. Carson, *Methods Enzymol.* **277**, 493 (1997).
88. CCP4, *Acta Crystallogr.* **D50**, 760 (1994).
89. P. Cramer et al., data not shown.
90. For commercially unavailable heavy atom compounds, we thank P.J. Alaimo and R. Bergman (University of California, Berkeley); G. Huttner, P. Schöckers, and P. Hofmann (Universität Heidelberg); W.

Windows Through the Dusty Disks Surrounding the Youngest Low-Mass Protostellar Objects

J. Cernicharo,[1]* A. Noriega-Crespo,[2] D. Cesarsky,[3] B. Lefloch,[1,4] E. González-Alfonso,[1] F. Najarro,[1] E. Dartois,[5] S. Cabrit[6]

The formation and evolution of young low-mass stars are characterized by important processes of mass loss and accretion occurring in the innermost regions of their placentary circumstellar disks. Because of the large obscuration of these disks at optical and infrared wavelengths in the early protostellar stages (class 0 sources), they were previously detected only at radio wavelengths using interferometric techniques. We have detected with the Infrared Space Observatory the mid-infrared (mid-IR) emission associated with the class 0 protostar VLA1 in the HH1-HH2 region located in the Orion nebula. The emission arises in three wavelength windows (at 5.3, 6.6, and 7.5 micrometers) where the absorption due to ices and silicates has a local minimum that exposes the central part of the young protostellar system to mid-IR investigations. The mid-IR emission arises from a central source with a diameter of 4 astronomical units at an averaged temperature of ~700 K, deeply embedded in a dense region with a visual extinction of 80 to 100 magnitudes.

Our lack of knowledge of star formation processes led to an empirical classification of the evolutionary phases of low-mass protostars into four classes: 0, I, II, and III. These describe the amount of material available for accretion versus the mass of the central object, providing the evolutionary status of the system (*1–3*). Class 0 objects are the young-

est protostars; they are surrounded by large and dusty envelopes that feed the central objects and their protoplanetary disks. These sources undergo violent ejection of matter related to accretion processes. The shockwaves created when the protostellar ejecta collides with the surrounding gas produce the Herbig-Haro (HH) jets observed at optical wavelengths. These jets seem to drive the bipolar molecular outflows (*4–6*) detected around protostars and represent a second mass loss–driven phenomenon taking place during the earliest evolutionary stages of the

[1]Consejo Superior de Investigaciones Científicas, Instituto de Estructura de la Materia, Departamento Física Molecular, Serrano 121, 28006 Madrid, Spain. [2]Space Infrared Telescope Facility (SIRTF) Science Center, California Institute of Technology, Pasadena, CA 91125, USA. [3]Institut d'Astrophysique Spatiale, Bât. 121, Université de Paris XI, 94500 Orsay Cedex, France. [4]Observatoire de Grenoble, Domaine Universitaire de Grenoble, 414 rue de la Piscine, 38406 St. Martin d'Hères, France. [5]Institute de Radioastronomie Millimétrique, Domaine Universitaire de Grenoble, 300 rue de la Piscine, 38406 St. Martin d'Hères, France. [6]Départament d'études de la Matière en Infrarouge et Millimétrique, UMR 8540 du CNRS, Observatoire de Paris, 61 Av. de l'Observatoire, F-75014 Paris, France.

*To whom correspondence should be addressed. E-mail: cerni@astro.iem.csic.es

REPLY TO
ATTENTION OF:

MCMR-RMI-S   (70-1y)   23 Aug 01

MEMORANDUM FOR Administrator, Defense Technical Information
Center (DTIC-OCA), 8725 John J. Kingman Road, Fort Belvoir,
VA 22060-6218

SUBJECT: Request Change in Distribution Statement

1. The U.S. Army Medical Research and Materiel Command has
reexamined the need for the limitation assigned to the technical
reports listed at enclosure. Request the limited distribution
statement for these reports be changed to "Approved for public
release; distribution unlimited." These reports should be
released to the National Technical Information Service.

2. Point of contact for this request is Ms. Judy Pawlus at
DSN 343-7322 or by e-mail at judy.pawlus@det.amedd.army.mil.

FOR THE COMMANDER:

Encl   PHYLIS M. RINEHART
                                     Deputy Chief of Staff for
                                        Information Management

Reports to be Downgraded to Unlimited Distribution

ADB241560	ADB253628	ADB249654	ADB263448
ADB251657	ADB257757	ADB264967	ADB245021
ADB263525	ADB264736	ADB247697	ADB264544
ADB222448	ADB255427	ADB263453	ADB254454
ADB234468	ADB264757	ADB243646	
ADB249596	ADB232924	ADB263428	
ADB263270	ADB232927	ADB240500	
ADB231841	ADB245382	ADB253090	
ADB239007	ADB258158	ADB265236	
ADB263737	ADB264506	ADB264610	
ADB239263	ADB243027	ADB251613	
ADB251995	ADB233334	ADB237451	
ADB233106	ADB242926	ADB249671	
ADB262619	ADB262637	ADB262475	
ADB233111	ADB251649	ADB264579	
ADB240497	ADB264549	ADB244768	
ADB257618	ADB248354	ADB258553	
ADB240496	ADB258768	ADB244278	
ADB233747	ADB247842	ADB257305	
ADB240160	ADB264611	ADB245442	
ADB258646	ADB244931	ADB256780	
ADB264626	ADB263444	ADB264797	