

Protein Crystallization on Lipid Layers and Structure Determination of the RNA Polymerase II Transcription Initiation Complex*

Francisco J. Asturias and Roger D. Kornberg‡

From the Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305-5126

Structure determination of proteins on the order of 10^4 Da is routinely performed at atomic resolution by x-ray crystallography and NMR. Proteins on the order of 10^5 – 10^6 Da (especially symmetrical protein assemblies) have been solved by x-ray methods, but these larger molecules and molecular arrays are generally studied by electron microscopy. Several small (10^4 Da) proteins have been determined in near atomic detail by electron methods, but for larger structures, the resolution is typically limited to 10–30 Å.

A single organic molecule is destroyed by a dose of x-rays or electrons far less than that required to determine its structure (1). This limitation is overcome by irradiation at a low level and averaging over many molecules for detection. The images or diffraction patterns to be averaged must come from molecules that are identically oriented, as any deviations will limit the resolution of the result. In x-ray analysis, this requirement is met by the use of three-dimensional crystals. In electron microscopy, there are two alternatives: the use of crystals, which must be very thin, preferably a single layer thick (or “two-dimensional”) to allow transmission of the electron beam and facilitate interpretation of the data; or correlation analysis, whereby images of single particles are grouped on the basis of similar features and thus, presumably, similar orientations (2). The single particle approach has the advantage that crystals are not required, but the effectiveness of correlation analysis for determining orientations is limited for smaller molecules (less than about 10^6 Da). As a result of these considerations, electron crystallography is often the method of choice for complexes larger than about 10^5 Da.

Improvements in all aspects of electron crystallography, from sample preparation to data collection and image processing, have extended the range of application and resolution of the approach. Here we review some recent developments, with particular regard to structure determination of two-dimensional crystals formed on lipid layers. The utility of the approach and potential for solution of biological problems are illustrated by results obtained for the $\sim 10^6$ -Da RNA polymerase II transcription initiation complex.

Lipid Layer Crystallization

Electron crystallography originated from studies of naturally occurring, ordered arrays of proteins, such as regular viruses and two-dimensional crystals of membrane proteins (3). A

three-dimensional structure can be derived from a single electron microscope picture of an array if its natural symmetry affords many different views, related in a known manner. For example, structures of viral capsid proteins have been determined from single images of helically or icosahedrally symmetric particles. In the case of two-dimensional crystals, a three-dimensional structure may be reconstructed from multiple images of the specimen tilted at various angles to the incident electron beam. Multiple images are readily obtained from specimens embedded in negative stain, allowing rapid structure determination to the resolution of the molecular envelope in stain (typically about 15 Å). Unstained specimens, preserved in amorphous ice or in a very concentrated solution of a small organic molecule such as glucose or tannin, may be solved to near atomic resolution (4, 5).

As natural ordering of proteins is rare, a general method of forming two-dimensional arrays was developed to expand the range of the electron crystallographic approach (6, 7). The method entails binding the protein of interest to a lipid layer, either by the use of an appropriate lipid-ligand or through electrostatic interaction with a charged lipid layer. Bound proteins are constrained in two dimensions without loss of mobility because of the rapid lateral diffusion of lipids. A high concentration of bound protein drives the crystallization process, with ordering in two dimensions determined by protein-protein interaction. Dozens of proteins have been crystallized in this way (8), and the following advantages of the procedure have emerged. First, only a small amount of material is required; under favorable circumstances, 10–1000 μ g of protein will suffice for three-dimensional structure determination. Second, two-dimensional crystals form under a wide range of aqueous solution conditions, including physiologic pH and ionic strength. Third, a two-dimensional crystal exposes a face for interacting molecules, facilitating the analysis of multiprotein complexes. Fourth, structure determination by electron crystallography is comparatively rapid, with no limitation on the size of the protein studied up to 10^6 Da and beyond. To summarize, electron microscopy of two-dimensional crystals on lipid layers is of particular benefit for proteins whose size, scarcity, or fragile nature precludes structure determination by other methods.

Generality of Lipid Layer Crystallization

The first requirement of the lipid layer crystallization approach is for a protein of sufficient quality to form ordered arrays. Homogeneity and structural regularity are crucial, although deviations may be tolerated if they are confined to the crystal faces and do not affect intermolecular contacts in the crystal plane. For example, monoclonal antibodies were crystallized in two dimensions (9) but failed to do so in three dimensions, perhaps because of carbohydrate chains projecting from the crystal faces whose heterogeneity or poor capacity for ordering interfered with three-dimensional crystal growth. In general, it appears that any protein capable of ordering in three dimensions will do so in two dimensions, but not the reverse.

The next requirement is for a method of binding a protein of interest to a lipid layer. The initial development of lipid layer crystallization employed specific lipid ligands, such as lipid-haptens and lipid-nucleotides. Although effective, this approach suffers both from limitation to available protein-ligand pairs and from the demand for lipid-ligand synthesis. More

* This minireview will be reprinted in the 1999 Minireview Compendium, which will be available in December, 1999.

‡ To whom correspondence should be addressed. Tel.: 650-723-6988; Fax: 650-723-8464; E-mail: kornberg@stanford.edu.

convenient and general approaches include the use of lipid-adaptors, such as metal-lipids for binding oligohistidine-tagged proteins (10) and the use of uniformly charged lipid layers, for binding almost any protein through electrostatic interactions (11). Whereas binding to lipid-ligands and lipid-adaptors is not very dependent on solution conditions, binding through electrostatic interactions can only occur within a narrow range of ionic strength and would be expected to be pH-dependent as well. At lower ionic strengths, protein binding but no crystallization may be observed, perhaps because interaction with the lipid layers is so strong as to impede rotational reorientation needed for crystalline ordering. At higher ionic strengths diminished electrostatic interaction would result in little protein binding. Because of the ease and generality of binding to charged layers, this approach will be the first to be tried for many new proteins, and variation of monovalent salt concentration in the range of approximately 100–300 mM will be necessary in each case.

The formation of two-dimensional protein crystals on lipid layers is frequently accompanied by the accretion of additional crystalline protein layers (epitaxial growth). This interferes with structural analysis by electron microscopy and can be minimized by the control of time, protein concentration, and other crystallization conditions. The formation of multiple layers can be exploited, however, for the generation of three-dimensional crystals suitable for x-ray analysis (12, 13). Such “seeding” of three-dimensional crystal growth has been performed in a number of instances and may prove general.

Structure of RNA Polymerase II

The benefits of protein crystallization on lipid layers are illustrated by application of the approach to the RNA polymerase II transcription initiation complex. When the first micrograms of a well defined RNA polymerase II, active in the initiation of transcription, became available by isolation from yeast, two-dimensional crystallization trials were performed, with the use of positively charged lipids for polymerase binding. These trials were immediately successful, but the two-dimensional crystals obtained were small and poorly ordered, diffracting to only about 30-Å resolution (14). Though unsuitable for structure determination, these crystals provided a starting point for further investigation. The ease, rapidity, and small amount of material required for two-dimensional crystallization were exploited to screen protein fractions for the formation of larger, better ordered arrays. It soon emerged that polymerase preparations varied in their content of two subunits, Rpb4 and Rpb7, and in quality of crystallization. Rpb4 and Rpb7 were generally found in substoichiometric amounts relative to the other polymerase subunits, and the resulting heterogeneity of polymerase preparations was identified as a possible impediment to crystallization. The problem was circumvented by the use of a yeast strain in which the gene for Rpb4 was deleted, and polymerase purified from this strain lacked Rpb7 as well. The deletion enzyme, termed $\Delta 4/7$ polymerase II, was homogeneous and formed two-dimensional crystals of good size and quality, diffracting to the limit of resolution in negative stain (15).

The three-dimensional structure of $\Delta 4/7$ polymerase II was derived from images of stained specimens to a nominal resolution of 16 Å (16) (Fig. 1). A notable feature of the structure was a cleft about 25 Å in diameter, appropriate in size for binding duplex DNA. A similar cleft in x-ray structures of much smaller single subunit polymerases harbors the active center (17), and the polymerase II cleft has been thought to do so as well. Support for this idea comes from the recent electron crystallographic analysis of an RNA polymerase II transcription elongation complex containing the enzyme in the act of transcrip-

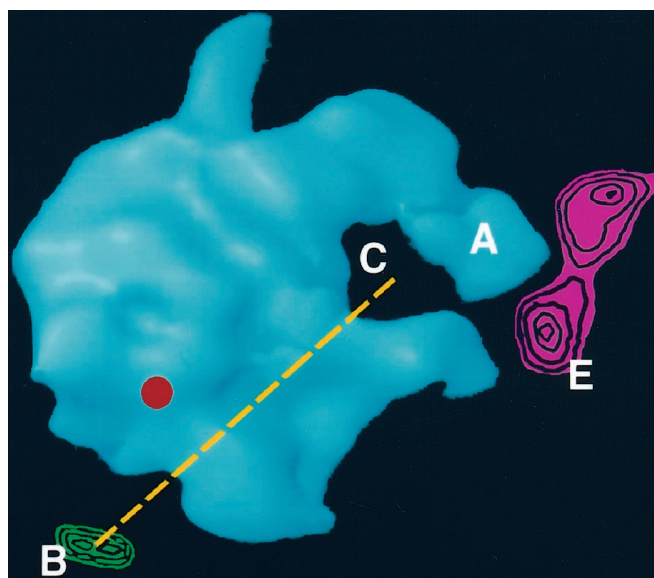


FIG. 1. Structure of RNA polymerase II-transcription factor complexes revealed by two-dimensional crystallography. In blue is shown the three-dimensional structure of yeast RNA polymerase II at 16-Å resolution. C denotes the active center cleft. A indicates the domain of protein density controlling entry and exit from the cleft. In green and magenta, denoted B and E, are the projected differences between factor B- and factor E-RNA polymerase II complexes, respectively, and the polymerase alone. The yellow dashed line marks out the 110-Å distance from B to the active center cleft and represents the presumed path of DNA from the TATA box to the transcription start site. The red dot indicates the point of attachment of the C-terminal domain of the largest subunit to the enzyme.

tion, with associated DNA and RNA. Difference structure determination between two-dimensional crystals of the elongation complex and polymerase alone revealed additional density in the complex, attributed to nucleic acid, in the 25-Å cleft.¹ There was no other difference, showing that the structure of the polymerase alone was in the elongation conformation, as previously proposed on the basis of electron crystallography of *Escherichia coli* RNA polymerase (18).

Alternative conformations of RNA polymerase II for transcription initiation and RNA chain elongation were anticipated from the nearly continuous ring of protein density surrounding the 25 Å cleft, impeding entry of DNA, and from previous studies of *E. coli* RNA polymerase and the smaller single subunit enzymes, pointing to a role of protein flexibility in opening and closing the cleft. Evidence for such flexibility of RNA polymerase II came from a second two-dimensional crystal form of the enzyme in which a domain of protein density (A in Fig. 1) at the mouth of the cleft was apparently absent because of motion or disorder (19, 20). This open conformation is presumably important not only for entry of DNA in transcription initiation but also for exit during termination.

Additional features of RNA polymerase II structure have been revealed by studies of two-dimensional crystals in negative stain. The problem of polymerase heterogeneity because of substoichiometric Rpb4 and Rpb7 could be circumvented not only by the use of $\Delta 4/7$ polymerase II but also by the isolation of wild type polymerase containing a full complement of these subunits from yeast grown under appropriate conditions. This wild type polymerase formed well ordered two-dimensional crystals, and difference structure determination revealed the locations of Rpb4 and Rpb7 (21). Difference structure determination also identified the point on the surface of the enzyme

¹ C. L. Poglitsch, G. D. Meredith, A. Gnatt, W. H. Chang, and R. D. Kornberg, manuscript in preparation.

from which the C-terminal domain of the largest subunit emanates into solution (22).

Structure of the RNA Polymerase II Transcription Initiation Complex

RNA polymerase II alone cannot recognize a promoter or initiate transcription. A set of additional proteins, termed general transcription factors, is required. Five such factors, denoted (TFII) B, E, F, H, and TBP, enable initiation at a core promoter, which comprises a TATA box and transcription start site (23). TBP recognizes the TATA box and creates a context for binding factor B which, in turn, interacts with RNA polymerase II. Biochemical studies have shown that factor B-polymerase II interaction determines the distance from the TATA box to the transcription start site (24), about 30 base pairs in most polymerase II promoters in most eukaryote organisms. Molecular measurement of this distance, the hallmark of a polymerase II promoter, underlies transcription initiation.

The mechanism of distance measurement has been inferred from electron crystallography of RNA polymerase II-factor B complexes. Difference structure determination between these complexes and polymerase alone revealed the location of factor B, about 110 Å from the active center cleft (25). The coincidence of this spacing, corresponding to about 30 base pairs of duplex DNA, with the conserved distance from the TATA box to the transcription start site, suggests a simple stereochemical basis for start site determination; promoter DNA follows a straight path on the surface of the polymerase from the B-binding site to the DNA-binding cleft, juxtaposing the start site in the DNA with the enzyme active center, where initiation occurs. (This proposal can accommodate the extraordinary situation in the yeast *Saccharomyces cerevisiae*, where the TATA box-transcription start site spacing is 40–120 base pairs, because a straight path only represents the minimum distance from the B-binding site to the DNA-binding cleft, and longer distances are not excluded.)

The proposed mechanism of distance measurement may be tested by two-dimensional crystallography of additional polymerase-general transcription factor and polymerase-DNA complexes. The location of factor E in such complexes, adjacent to the active center cleft, has already been determined (25). Similar analyses should reveal the locations of factors F and H and promoter DNA. The combined information from two-dimensional crystallography will yield a picture of an entire transcription initiation complex, illuminating the roles of all the general transcription factors in the initiation mechanism, as well as the mode of polymerase-promoter DNA interaction.

Future Directions: Helical Protein Crystals on Lipid Nanotubes

A promising extension of the lipid layer crystallization approach entails the use of lipid nanotubes rather than planar lipid layers. This development originated from the spontaneous formation of helical crystals of *E. coli* RNA polymerase adsorbed on planar lipid layers. The helical crystallization process evidently provided the driving force for extrusion of cylindrical lipid vesicles from planar layers or spherical vesicles (18). The same result could be achieved deliberately by the use of lipid nanotubes, which represent the preferred hydrated state of some lipid mixtures. Inclusion of metal-lipids or charged lipids in the mixtures enables the binding of oligohistidine-tagged proteins or binding by electrostatic lipid-protein interaction, as described above (26).

The main advantage of lipid nanotubes over planar lipid layers is the greater ease of three-dimensional structure determination. A single view of a helical crystal usually provides sufficient information for three-dimensional reconstruction to

an acceptable resolution (about 15 Å). A planar crystal, on the other hand, must be imaged at many angles of tilt with respect to the electron beam, which poses a number of problems, some of them possibly insuperable; there is a loss of resolution upon tilting, due in part to specimen charging, which is especially severe in the case of frozen, hydrated specimens; tilting beyond about 60° is impractical, resulting in a “missing cone” of information in any three-dimensional data set from a planar two-dimensional crystal.

Another benefit of lipid nanotubes is that protein binding is more readily measured. The nanotubes can be harvested by centrifugation and the amount of bound protein simply quantitated. In studies with planar layers, binding has thus far been monitored qualitatively by the observation of single protein particles in the electron microscope.

The efficient transfer of crystals on planar layers to electron microscope specimen grids without distortion or other damage proved a major impediment in studies of RNA polymerase II and other proteins in the past and has only recently been alleviated by the “loop transfer” method (27). It remains to be seen how lipid nanotubes compare with planar layers in this regard. Lipid nanotubes are mechanically more stable than planar layers, and they are maintained in suspension, so their transfer is, in principle, straightforward.

In some instances, helical crystals arise spontaneously from proteins associated with lipid layers, driving the formation of lipid tubes of a particular diameter, presumably because a pattern of intermolecular contacts that maximizes the degree of helical order occurs for that diameter (curvature). If preformed lipid nanotubes are used, their diameters may not be appropriate, and well ordered helical crystals may be rare, or in some cases, unobtainable.

Future Directions: Two-dimensional Crystallography and Single Particle Analysis

There are problems intrinsic to the two-dimensional crystallography of large molecular complexes that will diminish the resolution of the analysis. The number of intermolecular contacts possible in a two-dimensional array is limited by the reduced dimensionality of the system. As the size of the complex increases, the density of intermolecular contacts is further reduced by the smaller surface-to-volume ratio. Larger complexes are also more prone to structural variation because of flexibility between protein domains.

The best approach for such a specimen may be to combine crystallographic and single particle techniques. Crystallographic analysis will extract the structural information available from whatever ordering is obtained and provide an initial model. Single particle analysis will then refine this model, correcting insofar as possible variation, distortion, and lack of long range order (28, 29). The application of this combined approach to RNA polymerase II in ice revealed conformational variation of the enzyme that may account for a limitation on the degree of order of the crystals and on the resolution of structure determination (19).

Summary and Conclusions

The formation of two-dimensional crystals on lipid layers is widely applicable and, in combination with electron microscopy and image processing, affords a facile route to the structure determination of even very large multiprotein complexes at moderate (~15 Å) resolution. The value of structural information at this resolution is illustrated by studies of RNA polymerase II (~600 kDa). Electron crystal structures of the polymerase alone, of cocrystals with the initiation factors B and E, and of cocrystals with DNA and RNA have revealed the locations of the various components in transcription initiation and

elongation complexes. The location of factor B led to a simple hypothesis for the basis of transcription start site selection and for the conserved distance between the TATA box and start site in almost all RNA polymerase II promoters.

Two new approaches hold promise for extension of structure determination by electron microscopy to higher resolution. Combined crystallographic and single particle analysis may be used to overcome limitations on resolution because of conformational variation, local or long range disorder, and deviations from flatness and other specimen defects. Crystallization on lipid nanotubes rather than planar layers enables three-dimensional structure determination without the requirement for recording multiple images at a range of angles of tilt to the electron beam. Lipid layer crystallization can also serve as a starting point for x-ray analysis. It can lead to three-dimensional crystallization by providing a structural "assay" for deriving protein preparations capable of forming well ordered crystalline arrays and by providing two-dimensional "seeds" for three-dimensional crystal growth.

The balance between two-dimensional crystallography and single particle analysis techniques may change with time. The resolution of structures obtained by single particle techniques is continually improving. As the quality of images, the efficiency of particle alignment, and the power of data analysis algorithms increase, single particle analysis may become the technique of choice in the future.

REFERENCES

- Henderson, R. (1995) *Q. Rev. Biophys.* **28**, 171–193
- Penczek, P., Radermacher, M., and Frank, J. (1992) *Ultramicroscopy* **40**, 33–53
- Crowther, R. A., and Klug, A. (1975) *Annu. Rev. Biochem.* **44**, 161–182
- Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckman, E., and Downing, K. H. (1990) *J. Mol. Biol.* **213**, 899–929
- Kühlbrandt, W., and Wang, D. N. (1991) *Nature* **350**, 130–134
- Uzgiris, E. E., and Kornberg, R. D. (1983) *Nature* **301**, 125–129
- Kornberg, R. D., and Darst, S. A. (1991) *Curr. Opin. Struct. Biol.* **1**, 642–646
- Brisson, A., Olofsson, A., Ringler, P., Schmutz, M., and Stoylova, S. (1994) *Biol. Cell* **80**, 221–228
- Reidler, J., Uzgiris, E. E., and Kornberg, R. D. (1986) in *Handbook of Experimental Immunology* (Weir, D. M., and Herzenberg, L. A., eds) 4th Ed., Vol. 1, pp. 17.1–17.5, Blackwell Scientific Publications Ltd., Oxford
- Kubalek, E. W., Le Grice, S. F. J., and Brown, P. O. (1994) *J. Struct. Biol.* **113**, 117–123
- Darst, S. A., Ribi, H. O., Pierce, D. W., and Kornberg, R. D. (1988) *J. Mol. Biol.* **203**, 269–273
- Edwards, A. M., Darst, S. A., Hemming, S. A., Li, Y., and Kornberg, R. D. (1994) *Nat. Struct. Biol.* **1**, 195–197
- Hemming, S. A., Bochkarev, A., Darst, S. A., Kornberg, R. D., Ala, P., Yang, D. S. C., and Edwards, A. M. (1995) *J. Mol. Biol.* **246**, 308–316
- Edwards, A. M., Darst, S. A., Feaver, W. J., Thompson, N. E., Burgess, R. R., and Kornberg, R. D. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2122–2126
- Darst, S. A., Kubalek, E. W., Edwards, A. M., and Kornberg, R. D. (1991) *J. Mol. Biol.* **221**, 347–357
- Darst, S. A., Edwards, A. M., Kubalek, E. W., and Kornberg, R. D. (1991) *Cell* **66**, 121–128
- Brautigam, C. A., and Steitz, T. A. (1998) *Curr. Opin. Struct. Biol.* **8**, 54–63
- Polyakov, A., Severina, E., and Darst, S. A. (1995) *Cell* **83**, 365–373
- Asturias, F. J., Chang, W., Li, Y., and Kornberg, R. D. (1998) *Ultramicroscopy* **70**, 133–144
- Asturias, F. J., Meredith, G. D., Poglitsch, C. L., and Kornberg, R. D. (1997) *J. Mol. Biol.* **272**, 536–540
- Jensen, G. J., Meredith, G., Bushnell, D. A., and Kornberg, R. D. (1998) *EMBO J.* **17**, 2353–2358
- Meredith, G. D., Chang, W.-H., Li, Y., Bushnell, D. A., Darst, S. A., and Kornberg, R. D. (1996) *J. Mol. Biol.* **258**, 413–419
- Conaway, R. C., and Conaway, J. W. (1993) *Annu. Rev. Biochem.* **62**, 161–190
- Li, Y., Flanagan, P. M., Tschochner, H., and Kornberg, R. D. (1994) *Science* **263**, 805–807
- Leuther, K. K., Bushnell, D. A., and Kornberg, R. D. (1996) *Cell* **85**, 773–779
- Wilson-Kubalek, E., Brown, R., Celia, H., and Milligan, R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8040–8045
- Asturias, F. J., and Kornberg, R. D. (1995) *J. Struct. Biol.* **114**, 60–66
- Morgan, D. G., Grant, R. A., Chiu, W., and Frank, J. (1992) *J. Struct. Biol.* **108**, 245–256
- Frank, J., Chiu, W., and Henderson, R. (1993) *Ultramicroscopy* **49**, 387–396