Crystallization of protein-DNA complexes

Macromolecular interaction is essential, necessary and unavoidable in a living organism. Specific interactions among macromolecules are required for molecular machinery assembly and for progression and regulation of metabolic reactions. To fully understand a biological process, it is essential to determine the atomic structures of and interactions among components of a macromolecular complex and to decipher how these structures and interactions change during a reaction or signaling cycle. Some macromolecular complexes are naturally stable, for example tetrameric hemoglobin, nucleosome, and ribosome. But most macromolecular complexes are formed only transiently, e.g. an enzyme and substrate complex, a growth factor and its receptor interaction, or transcription factors assembled on a promoter. To determine structures of macromolecular complexes, whether stable or transient, has become a common practice of structural biologists in the 21st century.

Protein-DNA complexes were among the first macromolecular complexes characterized by X-ray crystallography. Since the isolation of the *lac* repressor with the lac operator by Gilbert and Müller-Hill, understanding the nature of specific protein-DNA interactions has captivated and occupied many scientists. How are specific interactions formed between protein and DNA? How do such interactions alter the involved macromolecules so that they progress to or initiate the next reaction step? The crystallographic studies of repressor-DNA and CAP-DNA complexes in the 1980s and early 1990s (Anderson et al., 1987; Jordan and Pabo, 1988; Otwinowski et al., 1988; Schultz et al., 1991) paved the road to pursuit of more complex and intricate protein-DNA assemblies in the last 10 years. In the last decade, the crystal structures of ribosome (Ban et al., 2000; Harms et al., 2001; Schuwirth et al., 2005; Selmer et al., 2006; Wimberly et al., 2000), nucleosome (Luger et al., 1997; Richmond and Davey, 2003; Schalch et al., 2005) and yeast RNA polymerase II holoenzyme complexed with RNA-DNA substrate (Westover et al., 2004) are the poster child of the modern structural biology. In the following sections, I will describe strategies and current approaches employed to obtain cocrystals of protein-DNA complexes.

1. <u>Sample preparation</u>
   a. protein

      Many proteins are composed of multiple functional domains with internal or terminal flexible regions. As commonly believed and practiced, one usually attempts to remove all the flexible and non-functional parts and retains only the smallest functional domains to facilitate crystal growth. In principle, this approach increases the probability of getting crystals because any flexible parts might inhibit the orderly packing of macromolecules in a crystalline array. However, the definition of "smallest functional domain" varies in practice. For example, lambda repressor contains both a compact helix-turn-helix domain that specifically recognizes DNA in the major groove (Jordan and Pabo, 1988) and a flexible arm that wraps DNA in the adjacent minor groove (Clarke et al., 1991). The crystal structure of the helix-turn-helix domain, although displaying the sequence specific recognition, misses the remaining protein-DNA interactions,

which considerably stabilize the repressor-DNA association. Personally, I would emphasize function over size and believe that we should aim for a trimmed down protein that has the complete function.

Many DNA-interacting proteins are insoluble at low salt concentrations, for example, γδ resolvase and MarA repressor (Reed and Grindley, 1981; Rhee et al., 1998). One reason that these proteins need high salt to be soluble is that they contain highly charged surface regions, which promote self-association to form an irregular and possibly branched polymers. High ionic strength may disrupt charge-charge interactions hence keeping protein molecules apart. The property of low solubility is actually useful in protein purification. By reducing salt concentrations, one can fractionate a protein mixture and precipitate the protein of interest. After centrifugation, the precipitated DNA-binding protein can be re-dissolved in high salt solution. One caution is that when centrifuging protein precipitates, whether under high or low salt conditions, the rotation speed or centrifugation force should be kept moderate (<4000g). High centrifugal forces often denature protein in the precipitated state and render the protein permanently insoluble.

Solubility of DNA-binding proteins often changes after forming complexes with a cognate DNA. For example, both γδ resolvase (Yang and Steitz, 1995) and MarA (Rhee et al., 1998), which require 2M NaCl for apoproteins to stay in solution, can be concentrated in 0.1M NaCl when complexed with their cognate DNAs. In addition, change of solubility is an indication for the formation of a specific protein-DNA complex. As mentioned previously, it is best to crystallize the smallest yet fully functional protein domains. Boundaries of such protein domain(s) can be defined by limited protease digestion of a full-length protein ni the presence of a cognate DNA. In parallel, one may be able to carry out limited protease digestion of protein alone. Different digestion patterns are equivalent to "protein footprinting". If involved in DNA binding, a flexible region that is subject to protease digestion in the apoprotein form is likely to become resistant to digestion when complexed with DNA. Such regions are essential to be included in the smallest functional domain(s) for crystallization trials.

b. DNA

The length and sequence of DNA used in crystallization is also based on the "smallest functional" principle. To summarize many trial-and-errors and successful results, DNAs that have been co-crystallized with protein molecules often contain 1 to 5 additional base pairs flanking a central essential region, which may be determined by DNA footprinting or based on functional assays. It is observed that DNA length is the key factor that dominates whether a protein-DNA complex crystallizes (Jordan et al., 1985). In practice, the minimal number of base pairs used to grow protein-DNA complex crystals is about 10 bps. Since crystals are often grown at room temperature, one needs at least 6-7 bps to keep a DNA duplex stable at 20°C. A single helical turn, 10-12 bps, is often a favorite starting point for short DNA recognition sites. We often screen DNA lengths

starting from the minimal length with increments of 1-2 bps up to 10-20 bps longer than the minimal sequence. In our experience, the most extra base pairs needed to get diffracting crystals is 5 bps on either end.

Another variable is the sequence of the "extra" DNA that flanks the central essential region, particularly bases at the ends of DNA. It is often observed that DNA ends in crystals are packed against other DNA ends or protein molecules. Therefore, sticky ends, which contain single or double unpaired bases, are often tested. Two types of sticky ends are found in protein-DNA co-crystal structures. In the first type the two ends of a single DNA are complementary. For example, one end has a 5´ protruding "G" and the other end has a 3´ overhanging "C". Such DNA can polymerize in a head–to-tail fashion to form a repetitive linear array that potentially facilitates crystal growth (Rice et al., 1996). The second type of sticky end is similar to the first but less regular. It is known that G:C and G or A:T and A can form triplexes, therefore a DNA end with a protruding G or A adjacent to C:G or T:A base pair, respectively, may stack with one another by forming a two-tier G-C-G or A-T-A triplex at the junction, which is more stable than blunt end stacking. Interestingly, such sticky ends in the CAP-DNA complex crystal was observed to form the DNA stacking by triplex as designed (Schultz et al., 1991), but the overhanging G in the γδ resolvase-DNA complex (Yang and Steitz, 1995) and MutS-DNA complex (Obmolova et al., 2000) were found to interact with protein side chains.

When a DNA substrate is too long for chemical synthesis with good yield, it is possible to divide DNA to two or three segments with adhesive ends without compromising protein-DNA association. For example, one may combine two half sites to generate a full site using appropriate stick ends (Schultz et al., 1990). This mix-and-match approach can increase the combinatorial sampling of many DNA molecules without having to synthesize each sequence in its entirety.

We recently have successes in two cases by using DNA as a crosslinking reagent to link two protein-DNA complexes together, thereby enhancing the complex stability and crystallization. In practice, the DNA oligos contain two instead of one minimal binding sites, which are arranged in tandem or as an inverted repeat. The two-site DNA gave rise to crystals, while the single-site DNA of the exact sequence didn't (Lee et al., 2005 Lee, 2006 #49).

For proteins that bind nucleic acid without sequence specificity, crystallization of protein-nucleic acid complex is not hopeless. Crystal lattice packing may help to select a specific binding register. Recent examples include the structure of BAF-DNA complex and RNase H-RNA/DNA hybrid complexes (Bradley et al., 2005; Nowotny et al., 2005; Nowotny and Yang, 2006). In these cases, more than protein molecules cover the entire length of oligo (7-12 bps).

We usually purify our synthesized DNA oligos by hydrophobic interaction columns, such as oligo R3 (PerSeptive BioSystems), on an HPLC system (Yang

and Steitz, 1995). We also use gel electrophoresis to purify oligos under the denaturing conditions. The easiest way for a beginner probably is to purchase purified oligos. Our favorite company, which provides full service, is Oligos Etc. (www.oligosetc.com). If shorter than 15 bps, oligos directly off synthesis resins are already of high purity, therefore purification is not always required.

c. Trapping a reaction intermediate

In some cases a protein of interest does not recognize a specific DNA sequence. Mixing of such a protein with DNA usually results in a mixture of protein molecules distributed along the length of DNA and ultimately no crystal. For example, DNA or RNA polymerase binds DNA duplex but exhibits little sequence preference. To capture a DNA polymerase and substrate DNA binary complex, one only needs to mix the protein with a DNA that contains both a double-stranded region and a 5´ overhanging single stranded region. The junction between the double- and single-strand DNA defines a specific site for polymerase to act (Li et al., 1998). Such structure- instead of sequence-specific protein-DNA interactions have been beautifully exploited in the Ku70-Ku80 studies in addition to polymerase-nucleic acid complexes. A Ku70-Ku80 heterodimer recognizes blunt ends of DNA. To prevent the protein from binding both ends or sliding along a duplex DNA, Goldberg and colleagues used a DNA with a single blunt end and sealed the other end with a double hairpin (Walker et al., 2001). The resulting DNA resembles a nail, the tip of the nail is the blunt end bound by Ku70-Ku80 and the nail head is made of the double hairpin that prevents Ku from binding to it or sliding off the DNA.

To capture an enzyme-substrate complex, for example a DNA polymerase, template-primer DNA and incoming nucleotide ternary complex, multiple approaches to stall chemical reactions have been applied. The first approach is to use a 3´ dideoxy primer strand, which does not contain the nucleophile OH (hydroxyl group) to form a covalent bond with an incoming nucleotide (Doublie et al., 1998). The second approach is to use dideoxynucleotide triphosphate as an incoming nucleotide. A number of polymerases discriminate against dideoxynucleotide triphosphate, and these nucleotide analogs reduce the chemical reaction rate so that a enzyme-substrate complex can be captured (Li and Waksman, 2001; Ling et al., 2003). The third approach is to replace Mg^{2+}, which is often essential for phosphoryl transfer reaction, by Ca^{2+}, which often enables the enzyme-substrate complex but is ineffective in facilitating catalysis (Ling et al., 2004; Yang et al., 2006). The fourth is to mutate protein active site residues, for example, Glu to Gln or Asp to Asn, to stop the chemical reaction. The application of the fourth approach has been very successfully used in our structural studies of RNase H, a sequence-non-specific endonuclease (Nowotny et al., 2005; Nowotny and Yang, 2006). The above approaches of altering the substrate or enzyme active site can be applied to all enzyme-substrate complexes in order to capture reaction intermediates.

Another approach to capture a reaction intermediate of a protein-DNA complex is to devise a suicide substrate, which has been elegantly demonstrated in crystallographic studies of site-specific recombinase-DNA complexes and DNA glycosylase studies. A DNA recombinase rearranges a DNA segment by a cut-and-paste mechanism. Recombinase Cre (Guo et al., 1997) and lambda integrase (Aihara et al., 2003; Biswas et al., 2005) both undergo a phosphotyrosyl (DNA-protein) covalent intermediate after initial cleavage prior to final re-ligation. Protein-DNA intermediate complexes are trapped using a suicide DNA substrate, which lacks an appropriate OH group to replace the phosphotyrosyl bond with a phospho-diester bond and fails to release the Tyr and re-ligate DNA. Other chemical means have also been applied to trap reaction intermediates in crystallographic studies of DNA glycosylase (Fromme and Verdine, 2003; Scharer et al., 1997; Zharkov et al., 2002).

d. Stabilizing protein-DNA complexes by covalent cross-linking

Reverse transcriptase (RT), which makes DNA duplexes from a single-stranded viral RNA, not only has the property of non-sequence-specific interactions with nucleic acid, but such interactions are also weak. For many years, an atomic resolution structure of RT complexed with either an RNA-DNA hybrid or DNA duplex had eluded crystallographers. G. Verdine, S. Harrison and their colleagues designed a disulfide-based covalent linkage between a modified DNA base and a Cys side-chain of RT and succeeded in getting a crystal structure of RT-DNA complex (Huang et al., 1998). The basis for the design is the hypothesis that a protein α-helix tracks the DNA minor groove adjacent to the active site. By systematically replacing residues along that α-helix with Cys and modifying the DNA minor groove with a thiol group (Huang et al., 2000), they found a specific pair of modified protein and DNA that (1) produce disulfide cross-linked complexes with high efficiency and (2) the cross-linked complex retains the native conformation and allows several rounds of polymerization to occur. The rest is history. The Verdine group used a similar approach to capture several crystal structures of DNA glycosylase and DNA complexes, which demonstrates the generality of such cross-linking approaches (He and Verdine, 2002).

e. Preparation of protein-DNA complexes

One aspect that often concerns people is whether purification of a protein-DNA complex prior to setting up crystallization is necessary. In my own experience, I have not purified any complexes. Purity of protein and DNA is important and we religiously purify them separately. Protein and DNA are usually mixed at 1:1.2 to 1:1.5 molar ratios with DNA in excess. If a protein dimer binds to a single DNA molecule, the 1:1.2 molar ratio means 1 molar of protein dimers with 1.2 molar of DNA. The main reasons are that (1) DNA annealing is likely not 100% perfect, (2) estimation of either protein or DNA concentrations is approximate, and (3) DNA longer than 12 bps rarely crystallizes by itself. If a protein-DNA complex is very stable and the Kd is less than nanomolar, occasionally such complexes are purified by gel filtration before crystallization. Some material will be lost during purification, but the resulting complexes may be superior to unpurified materials.

Depending on the solubility of the protein sample, one can mix it with DNA either immediately at high concentrates or at low concentrations and exchange buffer and concentrate by dialysis and/or by centricon (a concentration device) afterwards. Precipitation is often observed after mixing DNA (usually in TE buffer) and protein in a moderate buffer (~50mM Hepes or Tris, pH < 8.0). Addition of one or two drops of 1M Tris base often clears the solution.

2. Crystallization

Crystallizing protein-DNA complexes is similar to crystallization of any macro-molecule. It depends on precipitant, ionic environment, pH and additives. Interestingly, it was observed early on that the optimal pH for crystallizing protein-DNA complexes is neutral to slightly acidic. This is probably due to polar interactions between negatively charged DNA backbones and positively charged protein side chains. Because of the polar interactions (Nadassy et al., 1999), salt as precipitant has not generated as many protein-DNA crystals as PEG or MPD. Our choices of crystal screen kits are Hampton screens I and II, Hampton PEG-ion screen and nucleic acid screen, and any variation of these.

A review on crystallization of protein and DNA by C. Wolberger and colleagues is available online: iprotocol.mit.edu/protocol/181.htm

3. References

Aihara, H., Kwon, H. J., Nunes-Duby, S. E., Landy, A., and Ellenberger, T. (2003). A conformational switch controls the DNA cleavage activity of lambda integrase. Mol Cell *12*, 187-198.
Anderson, J. E., Ptashne, M., and Harrison, S. C. (1987). Structure of the repressor-operator complex of bacteriophage 434. Nature *326*, 846-852.
Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science *289*, 905-920.
Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A., and Ellenberger, T. (2005). A structural basis for allosteric control of DNA recombination by lambda integrase. Nature *435*, 1059-1066.
Bradley, C. M., Ronning, D. R., Ghirlando, R., Craigie, R., and Dyda, F. (2005). Structural basis for DNA bridging by barrier-to-autointegration factor. Nat Struct Mol Biol *12*, 935-936.
Clarke, N. D., Beamer, L. J., Goldberg, H. R., Berkower, C., and Pabo, C. O. (1991). The DNA binding arm of lambda repressor: critical contacts from a flexible region. Science *254*, 267-270.
Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. Nature *391*, 251-258.
Fromme, J. C., and Verdine, G. L. (2003). Structure of a trapped endonuclease III-DNA covalent intermediate. Embo J *22*, 3461-3471.
Guo, F., Gopaul, D. N., and van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature *389*, 40-46.

Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001). High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. Cell *107*, 679-688.
He, C., and Verdine, G. L. (2002). Trapping distinct structural states of a protein/DNA interaction through disulfide crosslinking. Chem Biol *9*, 1297-1303.
Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science *282*, 1669-1675.
Huang, H., Harrison, S. C., and Verdine, G. L. (2000). Trapping of a catalytic HIV reverse transcriptase*template:primer complex through a disulfide bond. Chem Biol *7*, 355-364.
Jordan, S. R., and Pabo, C. O. (1988). Structure of the lambda complex at 2.5 A resolution: details of the repressor-operator interactions. Science *242*, 893-899.
Jordan, S. R., Whitcombe, T. V., Berg, J. M., and Pabo, C. O. (1985). Systematic variation in DNA length yields highly ordered repressor-operator cocrystals. Science *230*, 1383-1385.
Lee, J. Y., Chang, J., Joseph, N., Ghirlando, R., Rao, D. N., and Yang, W. (2005). MutH complexed with hemi- and unmethylated DNAs: coupling base recognition and DNA cleavage. Mol Cell *20*, 155-166.
Li, Y., Korolev, S., and Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. Embo J *17*, 7514-7525.
Li, Y., and Waksman, G. (2001). Crystal structures of a ddATP-, ddTTP-, ddCTP, and ddGTP- trapped ternary complex of Klentaq1: insights into nucleotide incorporation and selectivity. Protein Sci *10*, 1225-1233.
Ling, H., Boudsocq, F., Plosky, B. S., Woodgate, R., and Yang, W. (2003). Replication of a cis-syn thymine dimer at atomic resolution. Nature *424*, 1083-1087.
Ling, H., Boudsoçq, F., Woodgate, R., and Yang, W. (2004). Snapshots of replication through an abasic lesion:
structural basis for base substitution and frameshift. submitted.
Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.
Nadassy, K., Wodak, S. J., and Janin, J. (1999). Structural features of protein-nucleic acid recognition sites. Biochemistry *38*, 1999-2017.
Nowotny, M., Gaidamakov, S. A., Crouch, R. J., and Yang, W. (2005). Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell *121*, 1005-1016.
Nowotny, M., and Yang, W. (2006). Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. Embo J *25*, 1924-1933.
Obmolova, G., Ban, C., Hsieh, P., and Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. Nature *407*, 703-710.
Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988). Crystal structure of trp repressor/operator complex at atomic resolution. Nature *335*, 321-329.

Reed, R. R., and Grindley, N. D. (1981). Transposon-mediated site-specific recombination in vitro: DNA cleavage and protein-DNA linkage at the recombination site. Cell *25*, 721-728.
Rhee, S., Martin, R. G., Rosner, J. L., and Davies, D. R. (1998). A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. Proc Natl Acad Sci U S A *95*, 10413-10418.
Rice, P. A., Yang, S., Mizuuchi, K., and Nash, H. A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell *87*, 1295-1306.
Richmond, T. J., and Davey, C. A. (2003). The structure of DNA in the nucleosome core. Nature *423*, 145-150.
Schalch, T., Duda, S., Sargent, D. F., and Richmond, T. J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138-141.
Scharer, O. D., Deng, L., and Verdine, G. L. (1997). Chemical approaches toward understanding base excision DNA repair. Curr Opin Chem Biol *1*, 526-531.
Schultz, S. C., Shields, G. C., and Steitz, T. A. (1990). Crystallization of Escherichia coli catabolite gene activator protein with its DNA binding site. The use of modular DNA. J Mol Biol *213*, 159-166.
Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science *253*, 1001-1007.
Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. (2005). Structures of the bacterial ribosome at 3.5 A resolution. Science *310*, 827-834.
Selmer, M., Dunham, C. M., Murphy, F. V. t., Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R., and Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. Science *313*, 1935-1942.
Walker, J. R., Corpina, R. A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature *412*, 607-614.
Westover, K. D., Bushnell, D. A., and Kornberg, R. D. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell *119*, 481-489.
Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. Nature *407*, 327-339.
Yang, W., Lee, J. Y., and Nowotny, M. (2006). Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. Mol Cell *22*, 5-13.
Yang, W., and Steitz, T. A. (1995). Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. Cell *82*, 193-207.
Zharkov, D. O., Golan, G., Gilboa, R., Fernandes, A. S., Gerchman, S. E., Kycia, J. H., Rieger, R. A., Grollman, A. P., and Shoham, G. (2002). Structural analysis of an Escherichia coli endonuclease VIII covalent reaction intermediate. Embo J *21*, 789-800.