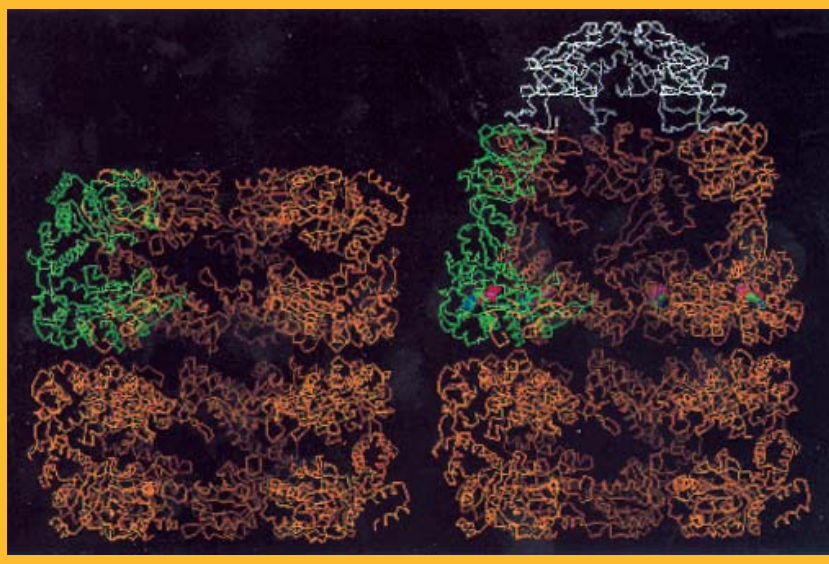
The Agouron Institute

SUPRAMOLECULAR ASSEMBLIES

CURRENT TECHNOLOGY *and* RESOURCE NEEDS



AGOURON INSTITUTE

The specific and primary purposes are to perform research in the sciences and in mathematics, to disseminate the results obtained therefrom, all to benefit mankind.

Cover photo: C-alpha drawings of GroEL (left) and GroEL-GroES-$(ADP)_7$ (right), sliced vertically along the seven-fold axis. (Xu, Z., Horwich, A.L., and Sigler, P.B. 1997)

SUPRAMOLECULAR ASSEMBLIES

CURRENT TECHNOLOGY *and* RESOURCE NEEDS

The Agouron Institute

INTRODUCTION

With the discovery of recombinant DNA technology in the mid-1970's, came a tremendous set of new opportunities in biology and chemistry. The Agouron Institute (AI), a non-profit research organization, was formed in 1978 by John Abelson, Melvin Simon, and Joseph Kraut, then all Professors at the University of California, San Diego, as a vehicle by which new research frontiers and technologies in biology and chemistry could be investigated in an expeditious manner free from what was perceived then as constraints in the University.

The first scientists working at the Agouron Institute in La Jolla came from the laboratories of the founders. The Institute commenced its activities with a small grant from the Office of Naval Research on the marine fouling problem — the beginning of a long standing program in environmental microbiology. By 1982 the research program had expanded considerably and had obtained additional funding from the NSF and the NIH. A group of molecular biologists and chemists were collaborating to exploit new technology in which synthetic oligonucleotides were used to direct specific mutations in genes. A crystallography group had been formed and they were collaborating with the molecular biologists to study the properties of the altered proteins. These were among the very first applications of the new technology to form what is now the very large field of protein engineering. In addition a substantial group of computational chemists had formed which had the goal of predicting protein structure and the specificity of protein ligand interactions.

Early successes in the protein engineering and computational groups led to the idea that this technology could be used in the rational design of ligands that interact with and inhibit the activity of proteins, i.e. rational drug design. In 1984 a commercial entity, Agouron Pharmaceuticals, was formed to exploit the potential of rational drug design. A commercial cooperation agreement was instituted between Agouron Pharmaceuticals and the Institute and in return the Institute received a major portion of the founding stock. Many of the molecular biologists, chemists and structural biologists shifted to the company and within a few years the two institutions moved in different directions

Agouron

with the Institute specializing in environmental microbiology and molecular biology.

The vision of rational drug design turned out to be a prescient one. Agouron Pharmaceuticals is today a major biotechnology company and its first rationally designed drug, Viracept™ is the leading HIV protease inhibitor. The use of protease inhibitors together with reverse transcriptase inhibitors in a multi-drug therapy regime has led to a dramatic decrease in deaths due to AIDS. Agouron Pharmaceuticals has a number of other drugs in the pipeline including a metalloproteinase inhibitor designed to function as a non-cytotoxic inhibitor of metastasis and a rhino virus protease inhibitor that may provide relief from the common cold.

With the success of Viracept™ the value of the Agouron Pharmaceutical stock has appreciated and with it the endowment of the Institute. As a result, the Board of Directors of the Institute has been able to consider new strategies to achieve their goal of making significant contributions in biology and chemistry. The Institute has decided to expand its activities beyond research at the La Jolla site and to explore the possibilities of making highly leveraged investments in basic and applied biology and chemistry. To this end the decision was made to carry out a study program in several promising areas of research. This study of prospects for the determination of the structures of supramolecular complexes is the first such effort.

In January 1998, the Agouron Institute held two meetings, one in New York and one in La Jolla, on the topic of Supramolecular Assemblies. The invited scientists David Agard (UCSF), David Eisenberg (UCLA), Richard Henderson (MRC), Wayne Hendrickson (Columbia), Sung-Hou Kim (UC Berkeley), John Kuriyan (Rockefeller), Harry Noller (UC Santa Cruz), Douglas Rees (Caltech), Paul Sigler (Yale), and Tom Steitz (Yale) were asked to speak about their own research and to participate in an open-ended discussion assessing the current technology, needs and resources of the field.

One of the recommendations of this group was that the institute include among its research activities, the public issuance of reports on the status of fields of research to bring to the attention of federal funding agencies and private foundations resource gaps and how the strategic input of funds can accelerate the progress of science. The Agouron Institute has accepted this challenge. Not only can we impartially evaluate an area of research and define specific resource needs, but we will prove our commitment to our recommendations by supporting some of them ourselves.

Institute

Current Technology

Over the last 10 or 20 years, it has become apparent that fundamental cellular processes do not generally occur as the result of protein and nucleic acid molecules colliding randomly within the cytoplasm, nucleoplasm, or cell membrane. Rather, most processes are carried out by molecular machines composed of large numbers of individual protein and RNA subunits. Because these machines are so large and complex, understanding how they function is a daunting task. We cannot hope to find out how cells work, however, unless we know the three-dimensional structures of these machines (Alberts, 1998). As Richard Feynman put it nearly forty years ago, "It is very easy to answer many ... fundamental biological questions, you just look at the thing! (Feynman, 1960)." Thus we expect that structural biology will play an increasingly important role in paving the way to understanding the molecular and chemical basis of cellular function. Just as the three dimensional structures of individual macromolecules such as DNA and proteins revolutionized the biological sciences in the 20th century, the structures of complex molecular machines will be central for achieving the next level of understanding in the 21st century.

The study of protein machines requires the application of molecular biology, biochemistry, and structural biology techniques (for general review see Alberts et al., 1995). Analysis of machine structure at atomic resolution is the primary subject of this review. In this introduction, however, we will briefly discuss how molecular biology and biochemistry can help us understand how machines work.

MOLECULAR BIOLOGY

Many molecular biology techniques can be usefully applied to the understanding of machines. First, in most systems, it is straightforward to introduce a cloned gene into cells and express its protein product in its normal context. The open reading frame encoded by a gene can then be tagged with an epitope recognized by a monoclonal antibody or a fluorescent moiety such as green fluorescent protein (GFP) (Chalfie et al., 1994; Jones et al., 1995; Wang et al., 1996; Jarvik and Telmer, 1998; Tsien, 1998). The subcellular localization of the tags can be examined using light or electron microscopy in order to determine how the protein is distributed within the cell (Niswender et al., 1995). One can also attach biochemical 'handles', such as hexahistidine sequences, to proteins. Histidine tagging allows purification of native complexes containing the protein of interest using nickel-based resins (Crowe et al., 1996).

Second, molecular biology can be used to examine pairwise interactions between any two proteins that are suspected to be part of the same complex. This can be done by expressing individual histidine, GFP or glutathione S-transferase (GST)-tagged proteins in bacteria, yeast, or animal cell systems, then mixing these proteins together and examining whether they associate (Sharrocks, 1994; Park and Raines, 1997; Tsien, 1998). One can also detect protein-protein interactions with 'two-hybrid' systems, which use transcriptional activation as an assay for whether two proteins can interact within a yeast or bacterial cell (Song et al., 1994; Phizicky and Fields, 1995; Bai and Elledge, 1997; Brent and Finley, 1997).

Third, after one has identified the machine to which an individual protein belongs, its function can be examined by mutating the gene encoding it and introducing the mutant gene into the appropriate cells (Penner et al., 1997). By examining how cellular processes are affected in cells expressing the mutated gene, one can obtain insights into machine functions. In some cases, mutagenesis experiments can also define how an individual protein contributes to the activities of the machine as a whole.

BIOCHEMISTRY

Sophisticated biochemical techniques are essential for understanding machine composition, assembly, and function. It is often necessary to develop new protocols to purify each macromolecular complex to be examined. In many cases, the precise conditions used for machine purification can dramatically change which proteins are purified as part of the complex. Introducing a histidine tag into one identified component of a machine can sometimes allow purification of an entire complex as an intact entity. If, as in yeast, one also knows the entire sequence of the genome and can isolate biochemical quantities of the complex, it is then possible to identify every protein within it by analyzing protein fragments either by direct peptide sequencing or by comparing the masses of peptides using a mass spectrometer with those in known protein sequences.

Once a complex thought to represent a machine has been purified, biochemical assays must be devised to examine which functions the complex retains. Specialized methods are required to perform kinetic analyses of the individual steps catalyzed by a machine. For example, the replication machine that copies double-stranded DNA must unwind DNA at the replication fork, separately replicate the leading and lagging strands, and remove errors in the new DNA molecules. The unwinding, replication, and editing processes are in turn composed of many individual subprocesses (Baker and Bell, 1998). In order to understand how the machine copies cellular DNA, one must devise assays to analyze the rates and mechanisms of each of the component subprocesses. This is obviously a very large task. We are still far from a complete understanding of the bacterial replisome, which has been studied intensively for forty years.

SUPRAMOLECULAR ASSEMBLIES: MOLECULAR MACHINES

In the next sections, we briefly describe our present understanding of a small subset of the molecular machines that function within eukaryotic and prokaryotic cells. This subset has been chosen to illustrate some principles of machine structure and organization. We cite only a few references, although thousands exist. Many of these cited references are reviews from a recent and very useful issue of Cell (vol. 92, #3, 1998) on molecular machines. These reviews in turn reference other, more detailed reviews in addition to many primary references.

It must be emphasized that most machines do not function in isolation, but also interact with each other. Thus, for example, transcription complexes in vivo may be assemblies that also interact with the machines that remodel and assemble chromatin and splice mRNAs. Since we know even less about the interactions among machines than we do about the structures and functions of individual machines, however, such interactions will be described only for a few cases in which such knowledge exists.

Machines that operate on nucleic acids: ribosomes, replisomes, spliceosomes, and transcription complexes

The paradigm for molecular machines is defined by the ribosome, which exists as a stable 2.5 MDa complex within all cells and was already thought of as a machine by the early 1960s. Ribosomes were described at that time as being analogous to tape recorders that 'read' the message inscribed in the mRNA tape. It was originally thought that the three ribosomal RNAs were passive structural elements. The numerous processes executed by the ribosome, including mRNA and tRNA binding, peptide bond formation, mRNA translocation, and translational termination, were thought to be mediated primarily by ribosomal proteins. Now, however, we think of the ribosome as an 'RNA machine' that contains about 50 associated protein cofactors.

It is difficult to understand the functions of most intrinsic ribosomal components, since they function only in the context of the intact assembly. We have more structural information about the ribosome, however, than about any other similarly large asymmetric complex. Structures of many of the individual protein components have been solved to high resolution. Electron microscopic structures have been obtained in the 15-20 Å resolution range (Malhotra et al., 1998) and biochemical mapping methods have been used to identify the locations of individual protein and RNA components. Whole ribosomes and the large and small subunits of ribosomes can be crystallized,

and some crystal forms diffract x-rays to atomic resolution. Completion of a high resolution structure of a ribosome is still a goal for the future, but a three-dimensional crystal structure of the large subunit of an archaebacterium ribosome was recently determined at 9 Å resolution (Ban et al., 1998). The ribosome appears to be a surprisingly 'open' structure composed of an RNA scaffold to which proteins are attached. Unlike viruses, the only complexes of comparable size whose structures are known, the ribosome contains many water-filled holes and tunnels. Some of these presumably represent the A, P, and E sites to which tRNAs bind. The open structure may be required to allow the tRNAs and mRNA to move during the translation process (for review see Wilson and Noller, 1998).

Replisomes have been easier to understand mechanistically than ribosomes, because they can be readily separated into component parts that retain individual activities. These parts include polymerases, editing exonucleases, primases, helicases, and sliding clamps. In vivo, however, the replisome is a moving complex at the replication fork that contains all of these proteins, and may also be more loosely associated with ligases, topoisomerases, and many other replication enzymes. The replisome simultaneously replicates the leading strand continuously and the lagging strand discontinuously, as well as proofreading both new strands, and it accomplishes this at a speed of up to 1000 nucleotides per second (for review see Baker and Bell, 1998). The replisome is initially assembled onto DNA at replication origins via interactions with origin-binding protein assemblies such as ORC (in eukaryotes) and the DnaA complex (in E. coli).

The spliceosome, like the ribosome is a large complex of RNA and protein and as in the case of the ribosome, it is thought that the RNA is the fundamental catalyst in the splicing reaction (for review see Staley and Guthrie, 1998). The spliceosome carries out a complex set of transesterification reactions that remove intron sequences from mRNAs and ligate the exons together. In each splicing reaction the spliceosome is assembled de novo on the pre-mRNA substrate in a closely controlled assembly reaction in which five different small ribonucleoprotein complexes, the snRNPs, are added to the complex. Each assembly step is characterized by an RNA rearrangement. For example, in the first step of splicing, the U1 RNA base pairs with the 5' splice site. In a subsequent step this interaction is replaced by interactions between U6 RNA and the 5' splice site. Despite the fact that the two phospho-transfer reactions which constitute the splicing reaction do not require added energy, the overall process requires ATP and there is good evidence that at least some of this requirement is due to the action of a set of ATP-dependent RNA helicases which mediate these assembly steps. Progress in understanding RNA splicing will depend not only on obtaining structures of the individual snRNP particles (most of which have been purified) but also the structures of the assembly intermediates and of the functionally active spliceosome. Because of the dynamic nature of this process, this is a daunting task. It is not even clear that all of the assembly intermediates have been

characterized. Thus techniques which can give structural information from small amounts of material will probably be essential in this area.

Like replisomes, transcription complexes contain separable components that retain activities such as transcription of naked DNA into RNA. In eukaryotes, the RNA polymerase II holoenzyme contains 'core' RNA polymerase, which is an enzyme with many subunits, and a set of basal transcription factors. The holoenzyme is in turn associated with several other protein assemblies. These include the SWI/SNF complex, which is involved in chromatin remodeling, 'coactivator' complexes that have histone acetylase and deacetylase activities, and a 'mediator' complex (for review see Kadonaga, 1998). One or more of these associated complexes can copurify with RNA polymerase II activity under some conditions (see Wilson et al., 1996). The basal transcription/chromatin remodeling complex also binds to cell type-specific components that control whether a particular gene will be transcribed. These components bind to promoter/enhancer elements adjacent to genes and direct the basal transcription complex to the appropriate initiation sites.

Machines that move and sort proteins within the cell: nuclear pores, protein translocation channels, and vesicle transport machines.

Nuclear pores are huge 'grommet'-shaped complexes containing 50 to 100 different protein components. They are 30 times as large as a ribosome. Nuclear pores pierce the double nuclear membrane and regulate the traffic of proteins and mRNAs between the nucleus and cytoplasm. Many cellular proteins have intrinsic nuclear localization or nuclear export signals within their sequences, and these signals direct their association with the appropriate 'importin' or 'exportin' transport complex. These transport complexes interact with specialized structures on the cytoplasmic and nucleo-plasmic faces of the nuclear pore and are then moved in the correct direction across the nuclear membrane. Directionality can be determined by regulated hydrolysis of GTP bound to a small G protein, Ran (for review see Ohno et al., 1998).

Protein translocation channels move secreted, single-spanning, and multi-spanning proteins across the membrane of the endoplasmic reticulum. Movement can occur either co- or posttranslationally. Cotranslational movement is best understood, and involves an interaction between the signal sequence of a nascent protein as it emerges from the ribosome and the signal recognition particle (SRP). SRP binds to a receptor on the endoplasmic reticulum membrane, allowing the ribosome to interact directly with the translocation channel. The protein can then move through the channel as it is synthe-

sized. Channel rearrangements must occur in a variety of different ways in order to allow correct assembly of type I (N-terminal outside), type II (N-terminal inside), and multispanning proteins into the ER membrane. Possible mechanisms for these rearrangements are discussed in a review by Matlack (Matlack et al., 1998).

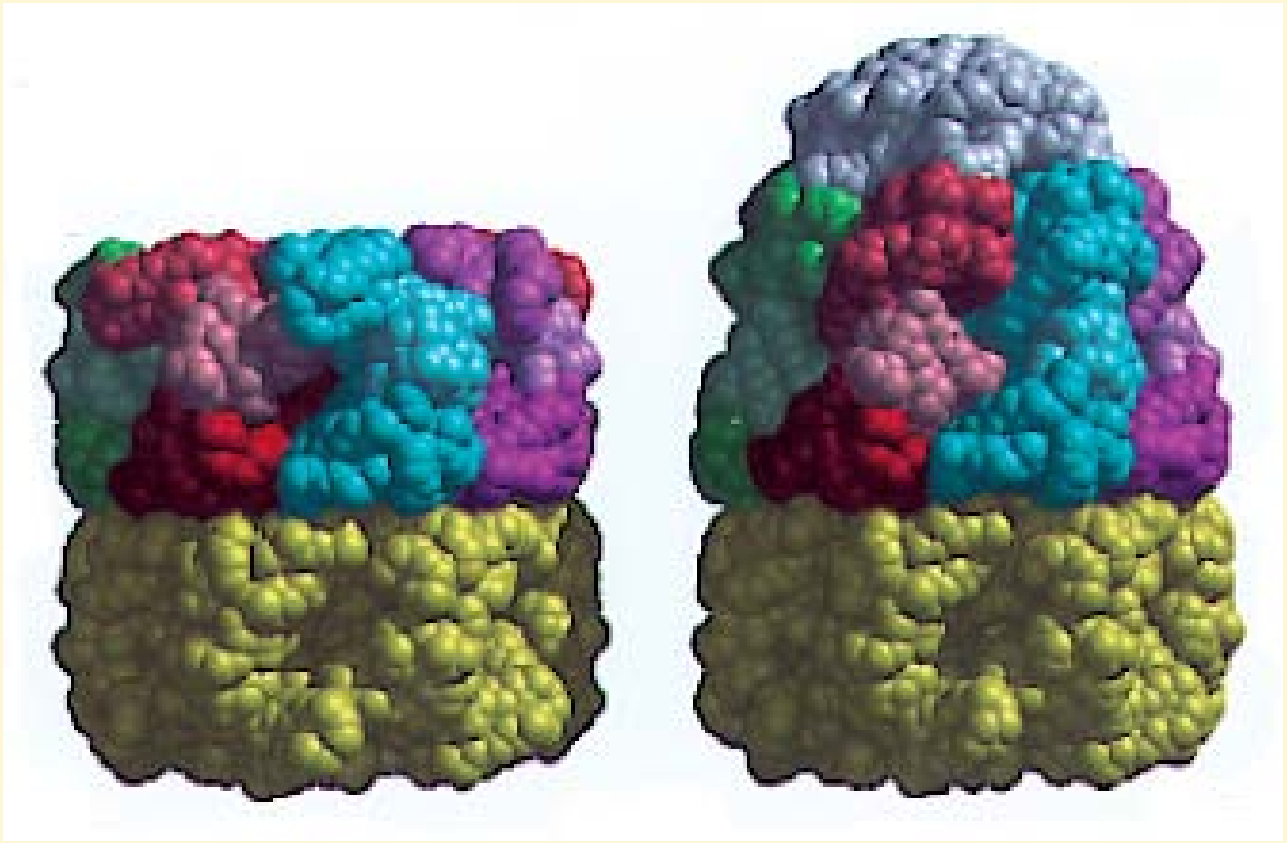
After a secreted, vesicular, or transmembrane protein has been inserted into or deposited within an endoplasmic reticulum vesicle, it is glycosylated in a specific pattern, and the vesicle containing it is delivered to the appropriate location within the cell. Many of these glycosylation and targeting steps occur within the cisternae of the Golgi complex. Vesicle transport is exceedingly complex, and involves interactions between many different protein complexes, including vesicle fusion machines, vesicle coat complexes, and signaling complexes that determine targeting. It occurs in both inward and outward directions.

One might consider the entire Golgi, and perhaps the entire network of intracellular membranes, to be a single gigantic molecular machine that directs the appropriate trafficking of proteins within the cell. Three-dimensional electron microscopic reconstructions of Golgi stacks from R. Macintosh's laboratory reveal that they have well-defined 'passages' that connect the various levels. Some connect adjacent levels, as expected, but others appear to form 'short-circuits' that connect nonadjacent levels. These may be conduits for as yet undefined protein trafficking events.

In addition to vesicle trafficking machines that function in all nucleated cells, many specialized machines exist that are only present in certain cell types. The processes carried out by these machines include synaptic vesicle release in neurons, vectorial transport in intestinal cells, and phagocytosis and cell killing by cells of the immune system.

Machines that fold and degrade proteins: chaperones and the proteasome

Chaperonins are a subset of proteins that are called "molecular chaperones", because they facilitate correct folding of other proteins within cells. They are some of the best understood machines, because three-dimensional structures of chaperonin complexes have been solved at atomic resolution. The GroEL complex in bacteria is about 1 MDa, and is composed of two back-to-back rings of protein subunits that surround a central cavity. The rings are capped at one end by a GroES complex. The walls of the central cavity undergo conformational changes driven by ATP hydrolysis, which allow the binding, releasing, and rebinding of partially folded polypeptide chains. When their folding is complete, polypeptides no longer bind to GroEL-GroES (for review see Bukau and Horwich, 1998). Comparisons of the high resolution structures of GroEL alone and the GroEL-GroES complex combined with sophisticated biochemical and molecular biological experiments have allowed a detailed model of the conformational changes that accompany and facilitate protein folding to be formulated. Structural studies of chaperonins combined with biochemical and biophysical



Van der Waals space-filling models of GroEL (left) and GroEL-GroES-$(ADP)_7$ (right). *(Xu et al., 1997)*

characterization of chaperonin-mediated protein folding illustrates how other approaches in combination with structural analysis at atomic resolution can be used to facilitate understanding of how machines function.

The proteasome is a tightly regulated molecular machine that catalyzes the degradation of proteins into small peptides. Core (20S) proteasome structures have been determined at atomic resolution (Lowe et al., 1995; Groll et al., 1997). The core proteasome contains 28 protein subunits. Remarkably, in archaebacteria it is composed of 14 copies each of two different subunits, while in eukaryotes it is built of two copies each of 14 different subunits. These structures contain features such as 'central channels' and 'side windows', which lead into the central cavity that contains the active sites. At present, little is known about how the protein substrate is delivered to these active sites, but the distance between active sites determined from the crystal structures has implications for the generation of self and foreign peptides that are presented by MHC molecules to T cells during immune surveillance.

Access to the proteasome must be tightly controlled in order to prevent unregulated protein degradation. Proteins are targeted to the proteasome via ubiquitination, a process that may also be catalyzed by a machine. Ubiquitinylated proteins are recognized by 19S 'cap complexes', which interact with the core proteasome to form an elongated structure that is about 45 nm long. Cap complexes may also unfold the ubiquitinylated proteins to prepare them for degradation by the proteasome (for review see Baumeister et al., 1998).

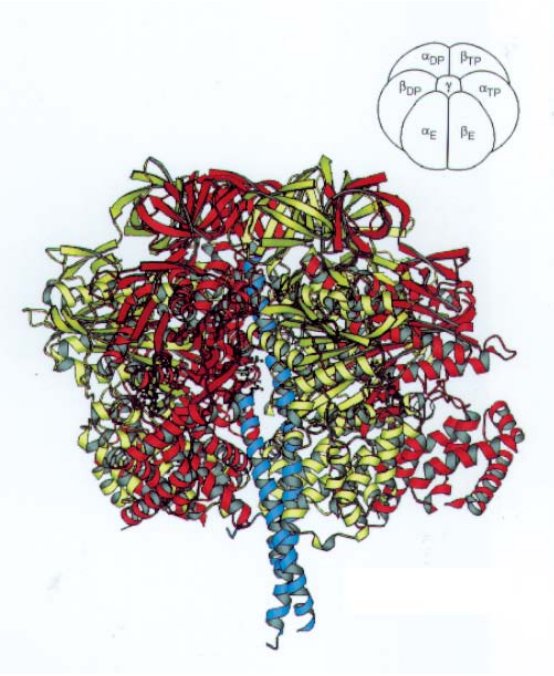
Machines that are motors: the bacterial flagellum and the F1 ATPase

Molecular motors propel free-living cells, drive shape changes in stationary cells, move organelles and vesicles within cells, and provide energy by generating ATP. The two motors that are best understood at the structural level are kinesin and myosin (Amos and Cross, 1997). These motors, however, are composed of only a few protein chains, and are more accessible to analysis by conventional methods than larger motors that contain many chains and span membrane bilayers. Here we focus on two large rotary motors that probably cannot be analyzed by standard crystallographic techniques.

The ATP synthase of mitochondria and the bacterial flagellar motor are both driven by proton gradients. Rotor movement is much better understood for ATP synthase because the structure of the portion of the enzyme outside the membrane, known as the F1 ATPase, has been solved at atomic resolution by x-ray crystallography (Abrahams et al., 1994). The structure shows that the γ subunit rotor is surrounded by a hexameric $\alpha\beta$ ring. Rotation of the γ subunit relative to the $\alpha\beta$ ring can be directly visualized in the light microscope by linking γ to a fluorescent actin filament (Noji et al., 1997). In mitochondria, γ subunit rotation is driven by proton flow through the F_0 portion of the ATP synthase, which spans the membrane bilayer. Movement of the γ rotor past the catalytic sites of the β subunits drives synthesis of ATP from ADP and inorganic phosphate. ATP hydrolysis by isolated F1 ATPase, however, can

drive rotation of γ in the opposite direction. Using the fluorescent actin filament assay, Yasuda et al. (1998) showed that the rotor moves in discrete 120˚ steps, with each step catalyzed by hydrolysis of one ATP molecule. The motor is almost 100% efficient, because the mechanical work performed to rotate the filament by 120˚ is almost equal to the free energy obtainable from hydrolysis of each ATP. Although the mechanism of the F1 ATPase has been analyzed in considerable detail, we still know very little about how proton flow through F_0 is coupled to movement of the rotor. To understand this will require methods of structural analysis in addition to x-ray crystallography, because it has not been possible to crystallize intact ATP synthase using existing techniques.

The bacterial flagellar motor drives bidirectional rotation of the flagellar filament,



F1-ATPase *(Abrahams et al., 1994)*

which is 10 μm in length and has a molecular mass of about a billion daltons (for review see DeRosier, 1998). The filament, which is analogous to the propeller on a power boat, rotates at up to 100,000 rpm and can push the bacterial cell along at velocities of hundreds of microns per second. The direction of filament rotation can be rapidly switched in response to environmental conditions, and torque generation and switching are both catalyzed by proteins of the motor complex, which is embedded in the double bacterial cell wall. The motor complex has a molecular mass of about 5 million daltons and bears an amazing resemblance to a fabricated mechanical device. The filament is connected to this motor by the hook, which is analogous to a universal joint. The hook is in turn attached to a drive shaft that passes through the L and P rings (analogous to bushings) in the outer membrane and peptidoglycan layers. The drive shaft ends in a socket attached to a rotor assembly known as the M ring, which spans the inner bacterial membrane. The M ring is also connected to the C ring, which extends into the cytoplasm. Rotation of the M ring is driven by a stator, which contains ten elements called studs that are arranged in an annular pattern around the drive shaft.

The motor is composed of multiple copies of a number of different subunits. The exact number of different proteins that are integral parts of the motor is unknown, but about 40 genes encode the complete collection of flagellar structural proteins, regulatory proteins, and proteins required for motor and filament assembly. Some of the structural proteins have been approximately placed

within the motor structure. For example, it is thought that the FliF, G, M, and N proteins are likely to be part of the rotor assembly. MotA and MotB are transmembrane proteins that are components of the studs. Torque generation is probably driven by interactions between the studs and the rotor. Proton flow through MotA, which forms a channel through the inner membrane, is coupled to torque generation, but we know nothing about the mechanisms involved in this coupling.

The bacterial flagellar motor resembles the ATP synthase in that both assemblies couple proton flow down a transmembrane gradient to rotary movement of a central element, but it is much larger and more complex, containing up to 110 copies of individual protein subunits. Understanding this motor at a structural level will be a daunting task.

Machines that transduce signals: receptor tyrosine kinase/Ras cassette complexes, G protein signaling complexes, and postsynaptic densities.

Cells respond to many different external signals via cell surface receptors. Information about receptor occupancy is transmitted into the cell interior by signaling modules that interact with activated receptors. Many of these modules are likely to exist as preformed complexes. Unlike some of the other examples discussed above, however, complexes involved in signaling events are seldom isolated as stable biochemical entities. Rather, a picture of the interactions among signaling components has been obtained by identifying pairwise interactions amongst them. Together, these pairwise interactions have allowed the description of conserved signaling pathways that can be used to transmit many kinds of information. Although signal transduction complexes are seldom associated into a defined structure that can be easily studied biochemically, their modular and flexible natures allow them to be examined and re-engineered using molecular biology techniques in ways that are not possible for complexes such as ribosomes or proteasomes. Here we will discuss three examples that illustrate different structural aspects of signaling machines.

Specific tyrosine residues within the cytoplasmic domains of receptor tyrosine kinase are phosphorylated after receptor activation. The phosphorylated tyrosines serve as docking sites for SH2-domain-containing adapter proteins. Such adapters can interact both with the tyrosine kinase and with downstream signaling modules. The best understood example of this is the recruitment of the Ras signaling cassette by activated receptor kinases. Here the GRB2 SH2-SH3 domain adapter, which is likely to exist in a stable complex with the GTP/GDP exchange factor Sos, associates with the phosphorylated kinase. This brings Sos to the plasma membrane and allows it to catalyze GTP exchange onto Ras. Ras is constitutively linked to the membrane by fatty acyl modifications. Ras-GTP then recruits the MAP kinase signaling module to the membrane and activates the

first enzyme in the MAP kinase cascade, Raf. Raf in turn phosphorylates and activates MEK, which phosphorylates and activates MAP kinase.

Although this pathway is normally depicted as the stepwise recruitment of individual components, it is more likely that segments of the pathway exist as preformed complexes. The MAPK module is a good example of this. Three kinases in this module are bound to scaffolding proteins that hold them in close proximity to each other. Ferrell has shown that the purpose of the three-enzyme cascade structure is to sharpen the transitions between the non-activated and activated states (Ferrell, 1996). Since the three enzymes are contained in a single machine, Raf and MEK probably have no opportunity to phosphorylate substrates outside of the machine. Thus, this module is a switch that can exist in only two states: the quiescent state, in which MAPK, the output kinase, is inactive, and the receptor-activated state, in which MAPK is active as a kinase and phosphorylates downstream substrates such as transcription factors.

The separable and modular nature of the complexes forming these types of pathways (receptor kinase and associated proteins, adapter complexes, downstream signaling complexes) allow tyrosine kinase pathways to be easily re-engineered for different signaling purposes. For example, although SH2 domains all interact with phosphotyrosine, different SH2 domains bind to phosphotyrosines in different sequence contexts. Thus, by switching SH2 domains or altering sequences around target tyrosines, different connections can be made among these signaling components.

G protein-coupled receptors activate GTP exchange onto heterotrimeric G proteins, which in turn activate effectors. Components of these signaling pathways often exist in stable complexes. Through the use of x-ray crystallography, we have a structural understanding of how the Gs and Gi proteins affect the enzymatic activity of their effector, adenylyl cyclase, which is an intramolecular heterodimer of C1 and C2 domains (Raw et al., 1997; Sprang, 1997; Sunahara et al., 1997). Gsα-GTP and Giα-GTP- bind to opposite faces of the cyclase dimer. Remarkably, however, they interact with the individual cyclase domains in very similar ways, even though one G protein activates the cyclase and one inhibits it. The distinction between activation and inhibition appears to be due to small differences in the conformational changes that the adenyl cyclase structure undergoes in response to binding the two different G proteins. Here, 'the devil is in the details', because these distinctions can only be understood, or even observed, by examining the structures of several complexes in great detail. It is possible that in vivo, the G protein-effector complex is also constitutively associated with receptors. For example, a receptor-G protein-effector complex was identified for the vasopressin receptor (Aiyar 1989).

Activation of some G protein-coupled receptors can produce effector activation within 50-100 msec. This seems consistent with the idea that many of the signaling components must be part of a single complex.

One very rapid response is the transduction current generated in fly photoreceptors after a light flash. The transient current response is generated via activation of a phospholipase effector and subsequent opening of the Trp/Trpl heteromeric ion channel, and it is rapidly shut off by the action of a protein kinase that phosphorylates the channel. All of the downstream components of this pathway are parts of a single machine. This machine is assembled by interactions between the C termini of the transduction proteins and a scaffolding protein called InaD. InaD contains multiple PDZ domains (Ranganathan and Ross, 1997) which interact with specific C-terminal sequences. Trp/Trpl, the kinase, and the phospholipase each bind to a different PDZ domain of InaD. One can thus remove individual components from the machine by mutating specific PDZ domains of inaD and reintroducing the mutated inaD gene into flies (C. Zuker, personal communication). Different components could presumably be added to the machine by installing new PDZ or other protein interaction domains into InaD that would recruit these proteins. Like the MAPK module, the phototransduction machine has only one output, current through the channel, and only one input, activated G protein. Remarkably, however, it is always able to rapidly generate a current response of a fixed duration in response to binding of G protein, because it contains built-in on (phospholipase) and off (kinase) switches.

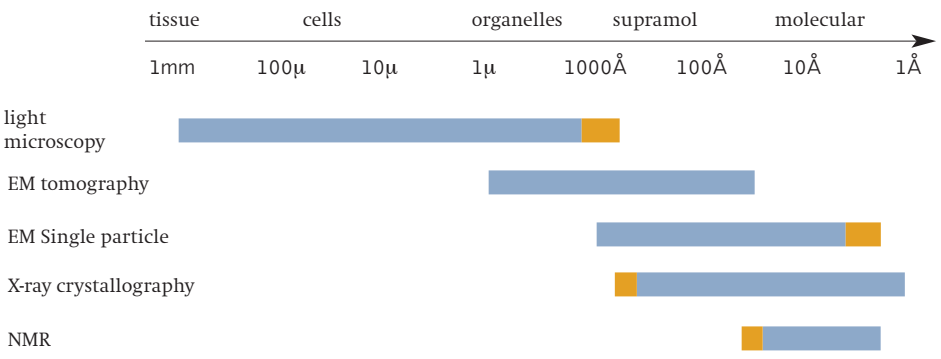
The postsynaptic density (PSD) of mammalian brain neurons is a very large structure that is visible in the electron microscope and can be isolated as a biochemical entity. It contains cytoskeletal proteins, cell adhesion molecules, ion channels, and transduction components such as kinases and phosphatases. The PSD is more like an organelle such as the Golgi apparatus than like a single machine, and probably should be considered as a 'supermachine' composed of many different machines that execute subsets of its functions. Like the fly phototransduction machine, each of these individual machines may be assembled onto a scaffold defined by one or a few proteins. The machines could then be linked together into the PSD by other PDZ domain interactions. The outputs of the PSD are presumably activated or inhibited enzymes, while its primary input is likely to be postsynaptic Ca^{2+}. All postsynaptic densities probably contain some common ingredients, but postsynaptic densities in different neurons are likely also to have divergent components that allow them to generate the signals appropriate to those neurons.

"JUST LOOKING" AT MACROMOLECULAR MACHINES

Once the components of a machine have been identified, and some aspects of their function and regulation have been elucidated, a three dimensional structure of the machine can be crucial to understanding its function. Techniques for determining 3D structures of biological macromolecules and assemblages include multidimensional NMR (Wüthrich, 1995), x-ray crystallography, and electron microscopy. At the present time, solution NMR spectroscopy is problematical for molecules above ~30 kD, so the techniques that have been and will continue to be used for structural studies of molecular machines are x-ray diffraction and electron microscopic methods. The methods are complementary, in that atomic resolution structures are most easily obtained by x-ray crystallographic techniques for relatively small macromolecules (i.e., ≤100 kD) though there is no intrinsic limit to the size of an assembly whose crystal structure could be solved, whereas the power of electron cryomicroscopy increases as the size of the object increases. The theoretical and practical aspects of using each technology for solving three-dimensional structures will be discussed in turn, followed by a summary of prospects for combining the two technologies for further structural insights into the mechanisms of cellular functions.

Hierarchical Approach to Cellular Structure

(courtesy of D. Agard)

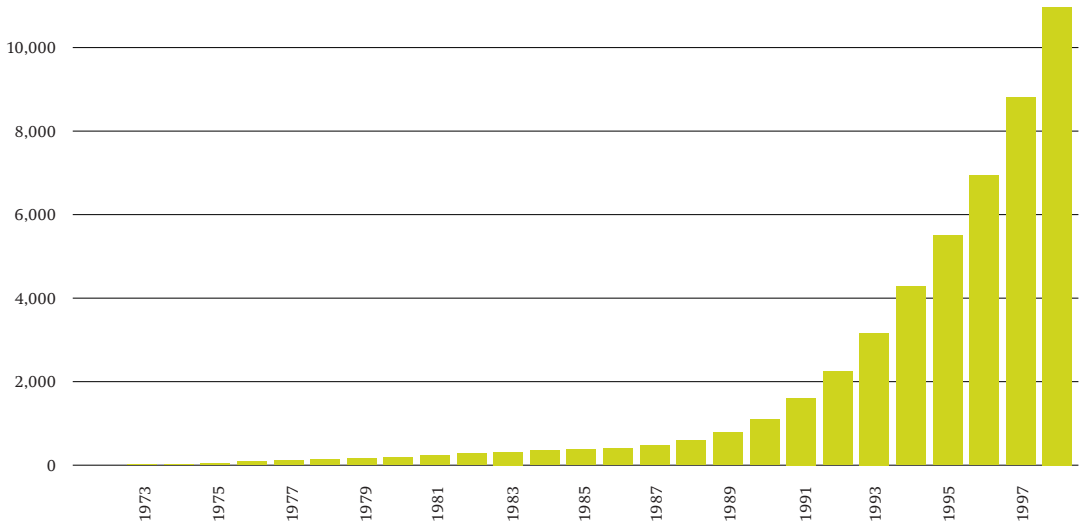


Crystallizing Machines: Solving Structures by X-ray Crystallography

X-ray crystallographic structure determinations of biological macromolecules have had a major impact on the advancement of the biological sciences and upon issues related to human health. X-ray crystallography is currently the most commonly used technique for obtaining 3D structures at atomic resolution, which for the purposes of this review, will be defined as resolutions ≤ 3.5 Å. Advances in protein production and crystallographic software and hardware have led to an explosion in the number of structure determinations, as witnessed by the number of new structures deposited in the Brookhaven National Laboratory Protein Data Bank (PDB) each year: from fewer than 50 in 1986 to more than 2000 in 1998 with the expectation of continued exponential growth. X-rays are used for biological macromolecular structure determinations because their wavelength (0.1 to 100 Å) is in the appropriate range to resolve the details of interest to the structural biologist (e.g., C,N,O,S,P atoms in biological macromolecules). Although x-rays have a number of advantages, their major drawbacks for structure determinations are that they cannot be focused by a lens into an image, and that they scatter relatively weakly from biological samples. To solve the problem of the unavailability of a physical lens, crystallographers use a computer to act as a computational lens by calculating a Fourier transform using the amplitudes and phases of the diffracted x-rays. The weak scattering by x-rays necessitates using a large number of macromolecules (>10^{10}) arranged into a crystalline array.

The first step in solving a structure by x-ray crystallography, obtaining suitable crystals, is often the most difficult. Crystallizing

Total Number of Submissions to Protein Data Bank (PDB) by Year

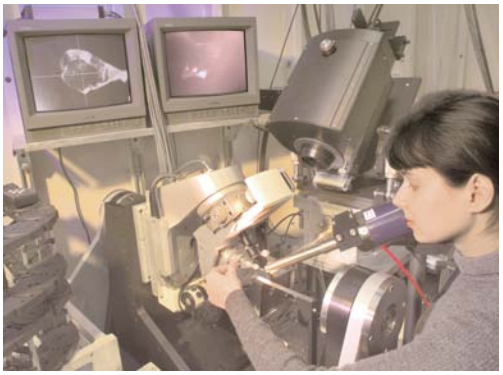


a protein, nucleic acid, or macromolecular complex remains an art rather than a science. The most time consuming step is producing large quantities (10-100 mg) of the multiple components of a macromolecular assembly and producing suitable variations (species or length) of each. Crystals are usually obtained by trial and error methods, although the introduction of sparse matrix factorial techniques for screening conditions has improved the success rate for crystallization. The next step in the structure solution is to collect x-ray diffraction data, using either a laboratory rotating anode x-ray generator or the intense x-rays produced at a synchrotron facility. The intensities of the diffraction maxima of the scattered x-rays are used to obtain amplitudes, but all phase information is lost in the recorded diffraction patterns.

Macromolecular crystallographers can obtain phase information by an indirect procedure called multiple isomorphous replacement (MIR) that involves using amplitude differences between crystals derivatized with heavy atom compounds and underivatized crystals (for review see Rhodes, 1993). In this procedure, the word "isomorphous" describes an important requirement for success, in that the introduction of the heavy atom must not change the properties of the crystal, a requirement that can be difficult to achieve in practice. In those cases in which some portion of the protein or complex that was crystallized is homologous to a macromolecule of known structure, initial phases can be obtained by molecular replacement (MR), once the known structure has been located in the unit cell of the crystals under investigation. Upon obtaining phases, the next step is to do a Fourier transformation using the measured amplitudes and initial phases to produce an electron density map. The map is then interpreted by the crystallographer (these days with help from automatic density interpretation programs) to build a molecular model.

Synchrotron Radiation

The use of synchrotron radiation has had a major impact for macromolecular crystallography (Biosync, 1997). Synchrotron radiation is produced when charged particles are accelerated in a curved path. The radiation is emitted as a continuum range of energies, and tunability of the synchrotron source allows wavelengths useful for macromolecular x-ray diffraction experiments (0.5 to 3.5 Å) to be selected. Synchrotron-produced x-rays are highly collimated and more intense than x-rays produced by laboratory generators. The low beam divergence and high degree of collimation that can be achieved at a synchrotron source enables good spatial resolution for crystals with large unit cells (in which diffraction maxima are close together) and



permits collection of usable data from very small, weakly diffracting crystals. With the introduction of cryogenic techniques whereby x-ray diffraction data are collected at ~−170˚C to reduce or virtually eliminate radiation damage, intense synchrotron-produced x-rays can be used to particular advantage to obtain high resolution data from single crystals.

In addition to the already mentioned advantages, the tunability of synchrotron radiation can be exploited to extract phase information in recently emerging technology called MAD (multiwavelength anomalous diffraction) (Hendrickson et al., 1990). In this method, phase information is obtained from datasets collected at different wavelengths from one crystal into which an appropriate chemical element has been incorporated. The MAD technique therefore eliminates problems with non-isomorphism that occur when comparing data from derivatized and underivatized crystals. MAD phasing requires introduction of a compound that produces anomalous dispersion into the crystals, which is often done by producing a protein with selenomethione substituted for its naturally occurring methionines or by using brominated uracil in nucleic acids.

A recent survey of the literature indicated that synchrotron radiation was used in nearly half of newly reported crystal structures (BioSync, 1997). The increasing use of MAD phasing, which must be done at a tunable synchrotron source, is likely to result in an even higher proportion of structures being solved with the help of data collection at a synchrotron facility in the future.

Crystal structure determinations of large macromolecules and macromolecular complexes.

Examples of recent structure determinations of machines or large macromolecular complexes include the 9 Å crystal structure of the large subunit of the ribosome (Ban et al., 1998), and atomic resolution structures of the proteasome (Lowe et al., 1995), GroEL/GroES chaperonins (Xu et al, 1997), the nucleosome (Luger et. al., 1997), T cell receptor/MHC complexes (Garboczi et al. 1996; Garcia et al., 1996), muscle proteins (Rayment et al, 1993a,b), and a CD4/gp120/Fab complex (Kwong et al., 1998). What are the future prospects for crystallizing machines and solving their structures to atomic resolution? With the use of synchrotron radiation, x-ray crystallography can be used for structure determinations of large macromolecular complexes that present very challenging crystallization problems. It is often possible to produce small microcrystals of complexes or difficult to crystallize proteins such as integral membrane proteins, but refining conditions to produce large crystals can be a major stumbling block. The high brilliance of new synchrotron sources can permit data collection from very small crystals — reportedly as small as 10 μm (BioSync, 1997). In addition, the highly collimated beams at the newer synchrotron sources can be used for data collection from crystals with very large unit cells; a recent example being the successful data collection from crystals of the core particle of bluetongue virus, a spherical particle of >700 Å in diameter that crystallizes with unit cell edges of ~800 Å (Grimes et al., 1998a). The

crystal structure was recently solved to 3.5 Å resolution, and represents the largest molecular structure determined to this degree of detail (Grimes et al., 1998b).

Continued improvements in cryopreservation techniques have made it routine to collect data at low temperatures to avoid most effects of radiation damage, so that diffraction from weakly diffracting crystals of large complexes can be recorded. The introduction of MAD techniques have made the "phase problem" in macromolecular crystallography less of a problem, and MAD will see increasing use in the future as more synchrotron beamlines are equipped to do MAD experiments. Continued theoretical and computational development in the area of "direct methods", whereby phase information is obtained directly from the diffraction amplitudes, may eliminate the phase problem altogether.

"Who needs crystals anyway?": Atomic resolution structures using electron microscopy

Recent advances in the field of electron cryomicroscopy offer the hope of solving atomic resolution structures of large assemblies of macromolecules that do not crystallize or require larger quantities of homogeneous material than it is practical to produce. The problem, of course, with crystals is that they can be difficult or impossible to obtain, especially crystals of supramolecular assemblages with dimensions >1000 Å, which often do not have defined structures. Over the past several decades, electron microscopy has been increasingly used to look directly at the structures of biological macromolecules, rather than just for visualizing their imprints in negative stain. The principles behind reconstructing three-dimensional images from electron micrographs were developed over 30 years ago by DeRosier and Klug, and recent innovations in electron microscopy, including improvements in sample preparation, instrumentation, and computation, have acted synergistically to use these principles to visualize specimens at increasingly higher resolutions (DeRosier, 1993; Chiu and Schmid, 1997; DeRosier, 1997, DeRosier and Harrison 1997; Koster et al., 1997; Walz et al., 1997).

Solving 3D structures by electron microscopy

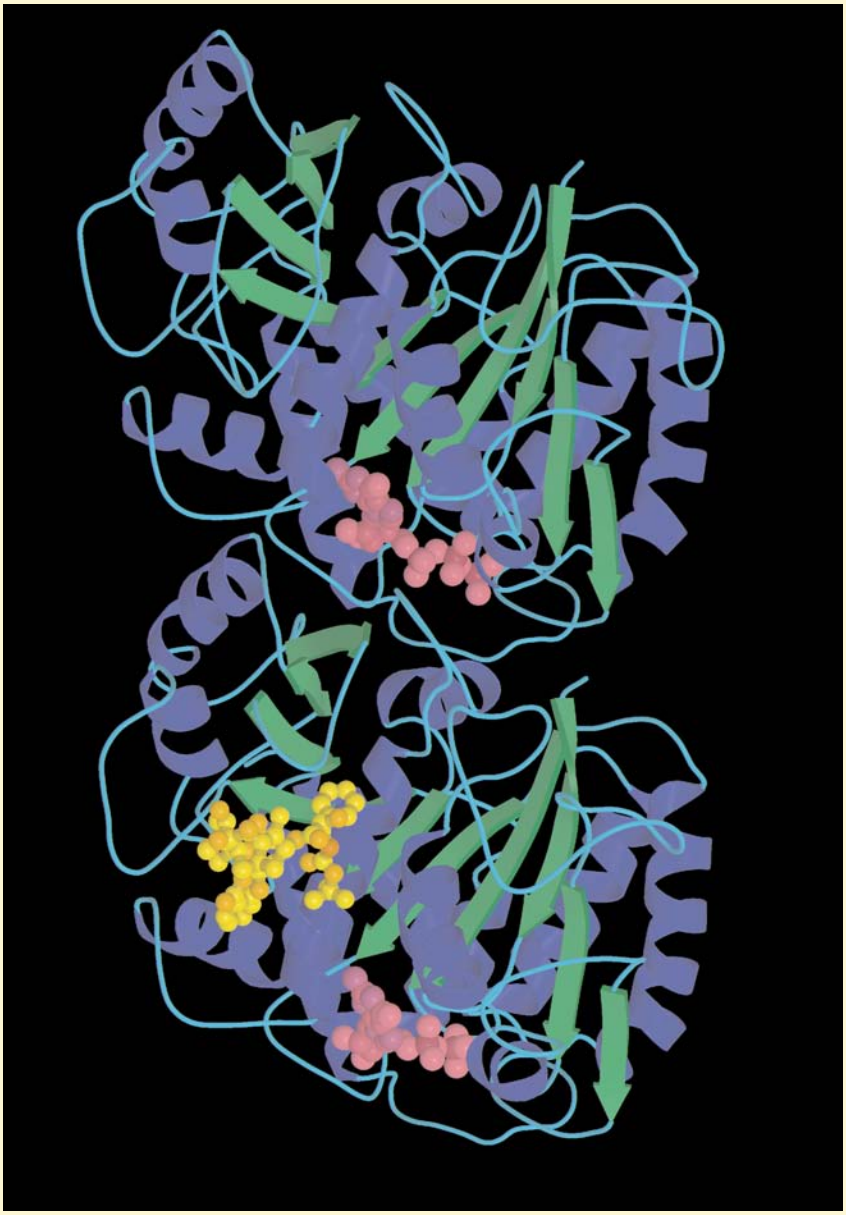
As previously discussed, the big problem for x-ray crystallographic structure determinations is that x-rays cannot be focused by a lens, so one must resort to indirect phasing methods to obtain the structure. Electrons, however, can be focused by lenses, thus electron microscopes can provide direct images of samples. Why then doesn't a typical EM picture of a macromolecule provide its 3D structure to atomic resolution? The answer to this question is multi-part: (i) An EM image corresponds to a projection of a sample and therefore contains only two-dimensional information. (ii) Electrons are scattered by air. Thus samples must be kept under a vacuum, conditions that are far from ideal for biological samples. (iii) Radiation from the electron beam damages biological specimens and can induce structural changes.

Low dose imaging reduces the effects of radiation damage, but the resulting micrographs are noisy and fine details are not visually apparent. (iv) The atoms in typical biological macromolecules do not scatter electrons differently enough to produce much contrast. Contrast can be obtained by staining the specimen to achieve either a "positive" or "negative" image, but the stain itself obscures high resolution details and limits the possible resolution to ~20 Å. Some contrast of unstained specimens can be achieved by defocusing, but the images themselves are altered under the defocus conditions.

Fortunately there are ways around, or in some cases, almost around, these problems so that 3D images of interesting biological specimens can be obtained using electron microscopy. For example, 3D information is obtained from the two-dimensional projections by combining images of the structure taken at different angles. This can be achieved either by tilting the stage on the microscope to record images of a single particle at different tilt angles, or by combining images of many individual particles that are oriented randomly on a grid. The effects of radiation damage and the vacuum cannot be completely overcome, but can be greatly reduced by preserving biological specimens in a frozen, hydrated state using cryopreservation techniques and low doses of radiation. The noisy images obtained from low-dose methods are then computationally averaged to enhance the signal to noise ratio and computational methods are used to remove artifacts introduced by various electron optical factors.

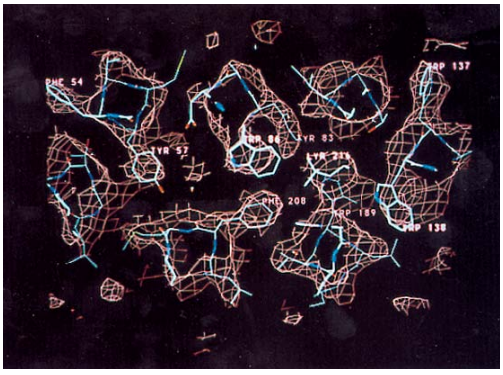
Electron Crystallography

There are several possible ways that 3D structures of biological macromolecules or assemblages can be obtained using EM techniques. The first, electron crystallography, is currently the best-established method for obtaining atomic resolution structures. This method, which was pioneered by Henderson and Unwin in their groundbreaking studies of the purple membrane protein from *Halobacterium halobium* (Henderson and Unwin, 1975), involves using a two-dimensional crystalline array of the specimen. 3D structures of the protein in the two dimensional crystal are obtained in much the same way structures are obtained from 3D crystals using x-rays, except that the images of the ordered array already contain the phase information. The EM images can be "back-transformed" by Fourier transformation into calculated amplitudes and phases. The calculated amplitudes can then be substituted by the amplitudes in electron diffraction patterns, which are collected in the back focal plane of the microscope. The experimental amplitudes are more accurate than the amplitudes calculated from the image, so when they are combined with the phases obtained from the image, the resulting calculated image is a more true representation of the projection image of the molecule than the original image. By collecting images and diffraction patterns of the 2D crystal at different tilt angles and correcting for the various instrumental effects, a three-dimensional image of the protein in the array can be obtained. This technique is particularly well-suited for solving the structures of membrane



Tubulin dimer *(Nogales, E., Wolf, S.G., and Downing, K.H., 1998)*

proteins, which can be difficult to coax into forming ordered three-dimensional crystals suitable for x-ray diffraction studies, but which often form ordered two-dimensional arrays in bilayers. To date, three atomic resolution structures at 3-4 Å resolution have been obtained entirely using electron crystallography: bacteriorhodopsin (purple membrane protein, the light driven proton pump from Halobacteria) (Grigorieff et al., 1996), plant light harvesting complex (Kühlbrandt et al., 1994), and the microtubule protein tubulin (Nogales et al., 1998). Other electron microscopic analyses that have reached near atomic resolution with the promise of achieving higher resolution in the near future include the nicotinic acetylcholine receptor (Unwin, 1995; Unwin, 1998), which forms an ordered array of tubes with helical symmetry, the plasma membrane Ca^{2+} ATPase (Zhang et al., 1998) and rhodopsin (Unger and Schertler, 1995; Baldwin et al., 1997). The systems being studied include many large membrane proteins that are generally difficult to crystallize for x-ray diffraction structure determinations, but form two-dimensional arrays suitable for electron diffraction.



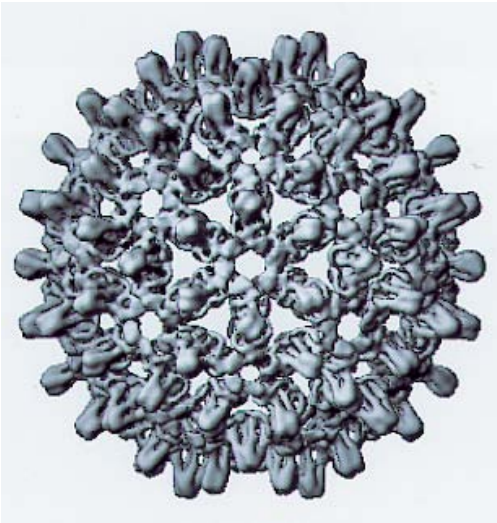
Bacteriorhodopsin *(Grigorieff et al., 1996)*

A promising approach to generate two dimensional crystals of soluble proteins suitable for electron crystallography has been developed by Roger Kornberg's laboratory (Edwards et al., 1991). In this method, proteins or macromolecular complexes interact with phospholipid head groups on a lipid bilayer. The protein or complex of interest either binds directly to the lipid head group or to a specific ligand attached to the phospholipid. So far, the tagging approach has yielded two-dimensional crystals that diffract to higher resolution; for example, 2D crystals of streptavidin grown on biotinylated phospholipids diffract to 2.0 Å, and a 3.0 Å projection map of the crystals reveals the expected β-sheet structure of the protein (Avila-Sakar and Chiu, 1996). Use of a nickel-derivatized phospholipid in a monolayer allows generalization of the method to any protein that can be tagged with a poly-histidine sequence, as demonstrated by the formation of 2D crystals of HIV reverse transcriptase (Kubalek et al., 1994).

Electron Tomography

The same reconstruction principles used in electron crystallography, whereby a series of two-dimensional images of a specimen are converted into a three-dimensional image, can be used to compute 3D structures of single particles or randomly oriented particles in a technique known as electron tomography. In its strictest definition, electron tomography refers to a 3D reconstruction derived from recording images corresponding to different two-dimensional projections of a single particle by physically tilting the stage

upon which the particle rests. In a broader definition, it includes 3D reconstructions obtained from a single electron micrograph of many particles randomly or semi-randomly oriented on a grid, in which it is assumed that all of the individual images can be considered as independent projection views of identical particles. In this case, the experimental images need to be aligned and averaged with respect to each other to refer them to a common 3D coordinate system. There are a number of computational algorithms to sort out the orientations of individual particles and to merge a large number of them coherently and statistically (Crowther et al., 1996; Baker and Cheng, 1996; van Heel et al., 1996; Frank et al., 1996). In theory, one needs only a few particles to reconstruct the three-dimensional image of an object. For instance, an object of 1000 Å in diameter needs ~1000 particles of unique and evenly sampled orientations for a 3 Å resolution reconstruction. In practice, several orders of magnitude more particles are needed because the high resolution details are blurred in the micrographs due to specimen movement under the beam, microscope imperfections, and particles of heterogeneous conformations. In the modern microscope, the instrument resolution is better than 2.7 Å, which can be set as the current theoretical limit of the technique.



hep B core protein

(Böttcher, B., Wynne, S.A., and Crowther, R.A., 1997)

Spherical viruses are tractable targets for a quasi-tomographic approach, since viruses have very similar structures even to high resolution, and individual images can be averaged using the icosahedral (60-fold) symmetry of the virus, a powerful method to improve the quality of the resulting structure. Recent examples of icosahedral virus structures determined by electron cryomicroscopy include the 7.4 Å hepatitis B virus core particle structure (which revealed a new fold for viruses — a four helix bundle) (Böttcher et al., 1997) and a 9 Å papillomavirus structure (Trus et al., 1997). It is anticipated that spherical virus structures will soon be solved to ~3 Å using electron cryomicroscopy (Chiu and Schmid, 1997). Averaging of individual images can also be applied in EM reconstructions of ion channels, chaperonins and proteasomes, large enzyme complexes that are highly symmetric (e.g., 7- to 10-fold), which have been studied to ~25 Å resolution using electron tomography.

While quasi-tomographic approaches offer the advantage that the specimen is exposed only once to the damaging radiation of the electron beam, they suffer the

disadvantage that the particles being averaged may not truly be identical to high resolution. This is especially a problem for large macromolecular assemblages that are inherently flexible. The highest resolution obtained so far from such techniques applied to a macromolecule with no internal symmetry is ~15 Å, found in structures of ribosomes derived from using over 29,000 projections of a ribosome complex in single-particle form (Stark et al., 1997; Zhu et al., 1997 Malhotra et al., 1998). Provided, however that these problems can be overcome, EM tomography of single non-uniform particles provides the most exciting potential for the solution of transient molecular machines. One can imagine that the future cell biologist will be able to take crude cell fractions, freeze them on an EM grid, and directly image their favorite machine (e.g., spliceosome, flagellar motor, intracellular signaling complex, cell cycle regulatory complex, or large transcription initiation complex) to near atomic resolution. The components of the complex will have been identified by some combination of genetics, biochemistry and genomics. Particular proteins in the complex can be tagged with an epitope tag, and the location of each tagged protein derived using gold-labeled antibodies against the epitope. Thus the molecular machine would not necessarily have to be purified in order to be studied. Successful development of an EM counterpart to green fluorescent protein, i.e. a relatively small tag that could be added to a protein to make it electron dense, would allow large assemblages to be visualized at higher resolution.

Cryoelectron microscope

Instrumentation for Electron Microscopy

*H*igh resolution microscopy is currently being done using an intermediate high voltage or a field emission gun (FEG) cryomicroscope. A FEG microscope can yield a highly coherent beam that generates low resolution contrast, which is necessary in the initial estimation of orientation, while preserving high resolution features. The use of intermediate high voltages (e.g., 300-400 kV instruments) allows higher resolution by reducing the chromatic effects of the images. Inelastically scattered electrons can also be filtered off using an imaging energy filter. To date, energy filtering microscopes have not been used extensively for collecting high resolution structural data from ice-embedded

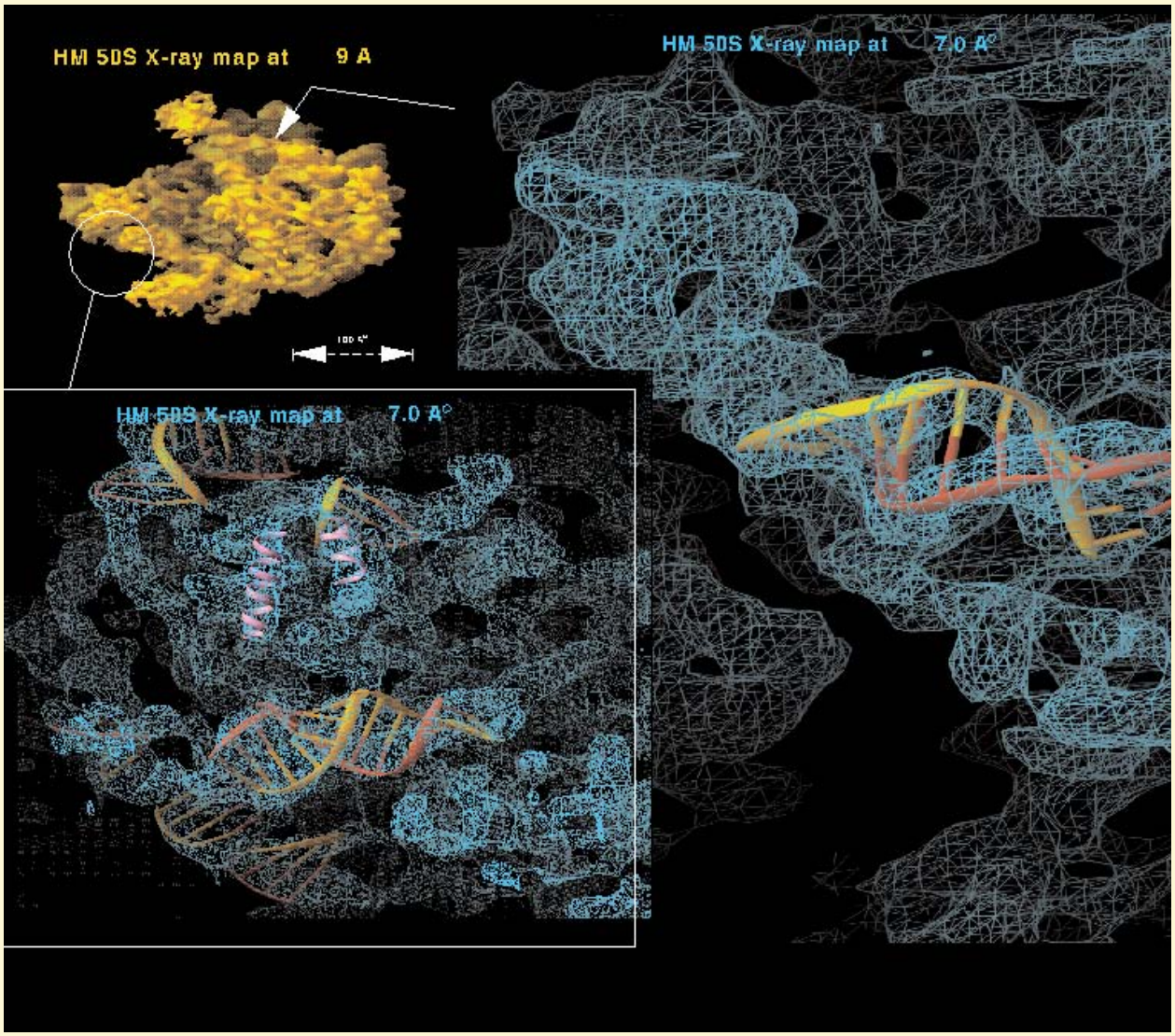
macromolecules, but a number of groups, such as those led by Wolfgang Baumeister and Jochaim Frank, are beginning to explore its usefulness for high resolution structural work. Energy filter microscopes are expected to offer advantages for examining thick specimens. An additional improvement in instrumentation involves the availability of cold stages that can operate at liquid helium or liquid nitrogen temperatures. The newer liquid helium cooled stages offer the advantage over liquid nitrogen cooled stages of a reduction in specimen temperature from about 90 K to less than 10 K, which is associated with 2-3 fold reduced sensitivity to radiation damage. The reduction in radiation damage allows images to be recorded with at least two-fold increased exposure, implying a 40% improvement in the signal-to-noise ratio of the images. The ability to increase exposure without an increase in radiation sensitivity is critical in applications involving true electron tomography, in which all of the data come from a single sample that cannot be averaged, and should allow a ~two-fold improvement in resolution.

Future prospects for single particle microscopy

Given the dramatic improvements in software, hardware and sample preparation that have led to recent breakthroughs in structure determination using electron tomography, it is likely that continued efforts will allow this resolution limit to be extended to near-atomic or atomic resolution within the next decade. Indeed, theoretical calculations (Henderson, 1995) suggest that atomic resolution structure determinations from single particles is possible if two requirements are met: (i) The particle must be large enough to carry enough information for position and orientation alignment to atomic resolution. The theoretical lower limit is 50 kD, although in practice, particles must be in the range of several hundred thousand daltons. (ii) At least 10,000 independent images of molecules in all orientations must be taken. In practice, up to a million images might be required due to losses of the signal-to-noise ratio from a variety of sources. One possible technique to extend single particle microscopy to structure determinations of smaller protein molecules using existing methods involves coupling the protein of interest to a three-dimensional geometrical aggregate that is ordered to atomic resolution. T=1 viruses, for example, have defined point group symmetry that could be used for this purpose. The protein of interest could be coupled to the T=1 virus using an antibody engineered to have two different binding specificities (a diabody). In principle, this method would allow the structure of any protein, regardless of its molecular weight, to be determined to atomic resolution without formation of two- or three-dimensional crystals.

Combining x-ray crystallography and electron microscopy

Single particle microscopy will be tomorrow's method for solving atomic resolution structures of large supramolecular assemblages. But what if you want to see the structure of your favorite uncrystallizable, large,



Haloarcula marismortui 50S subunit of the ribosome *(Ban et al., 1998)*

asymmetric, macromolecular complex today rather than tomorrow? A variety of methods have been developed to generate pseudo atomic resolution models of macromolecular complexes by combining high resolution x-ray structures of individual components with a lower resolution EM structure of the entire complex (reviewed in Holmes, 1994; Baker and Johnson, 1996). The procedure can be thought of as a three-dimensional jigsaw puzzle in which the pieces (high resolution 3D structures of the components of the complex) are assembled into a border defined by the EM-derived structure. Examples in which high and low resolution information have been combined to gain insight into the structure of a large complex include informative studies of the actin-myosin complex (Rayment et al., 1993b), virus-antibody complexes (Chiu and Smith, 1994), and complexes between microtubules and motor proteins (Sosa et al., 1997; Kozielski et al., 1998).

EM and x-ray technologies can be combined in a complementary manner such as in the recent 9 Å crystal structure of the large ribosomal subunit (Ban et al., 1998). EM-derived phases were used for the initial part of the solution of the x-ray structure. For a large asymmetric structure such as the ribosome, solving the phase problem by isomorphous replacement is especially problematic due to difficulties locating all of the bound heavy atoms by traditional difference Patterson techniques. For the ribosome work, a 20 Å EM structure was used to generate molecular replacement phases, which were used in difference Fourier calculations to locate the positions of the heavy atoms. Multiple isomorphous replacement phases were then calculated using the x-ray data to 9 Å. At 20 Å, the electron density map derived from x-ray data alone was very similar to the EM-derived image, validating both structures, and at 9 Å, the x-ray structure revealed new features corresponding to double helical RNA. Since these ribosome crystals diffract to 3 Å, an atomic resolution structure of the large ribosomal subunit is a real possibility for the near future.

Resource Needs

It is clear from this study that progress in the determination of high resolution structures of supramolecular complexes will come from x-ray crystallography, high voltage electron microscopy, and from a combination of these approaches. Progress in this field will be facilitated by wider access to synchrotron facilities and by the establishment of more high voltage electron microscope centers. We also have identified the need for more trained personnel at synchrotron facilities and the need for long term career development grants for junior level scientists.

SYNCHROTRON FACILITIES

The large unit cell size and generally smaller crystal size of supramolecular complexes dictates that virtually all of the data collected on these crystals will be at synchrotron beamlines. There have been a number of studies in recent years on the role of synchrotron beam lines in structural biology. In particular we cite the findings of the Structural Biology Subcommittee of the DOE Biological and Environmental Research Advisory Committee chaired by Jonathan Greer. Our consultants endorsed many of the recommendations of that committee.

The field will of course be aided by the availability of new beamlines for structural biology either by the construction of new facilities or more frequently by the redeployment of existing beamlines built for a different purpose. However, the Greer committee identified a number of improvements that could be made to existing structural biology beamlines that would greatly increase their usefulness. These include hardware upgrades particularly in detectors that would speed data collection time and thus increase throughput. Most importantly however, all recent studies of this subject including our own, have pointed to a need for increased staffing at each beamline that will permit efficient around the clock usage of the facility. As the field evolves and expands it is increasingly the situation that users of synchrotron facilities are not trained in the use of the facility and are ill-equipped to solve problems as they arise. The Greer committee recommends an increase of at least one FTE per beamline to a total of 3.5 to 4.0 FTE. This kind of on going salary support is often difficult to find but it is crucial if the facilities are to operate at maximum efficiency.

RECOMMENDATION

Thus an important recommendation of this report is that both public and private funding agencies interested in this field provide funds for increased staffing at each beamline in order to facilitate the efficient 24 hour per day usage of the facility.

HIGH VOLTAGE ELECTRON MICROSCOPY

Though it has become clear that high resolution electron microscopy will be an important, in some cases essential, technique in the characterization of supramolecular complexes, it is to be noted that there are relatively few groups in the United States which employ this tool and thus there are a small number of people being trained. This number could be increased by a combination of fellowship support which would attract new people to the field and the creation of a specific program to equip and/or modernize new and existing laboratories with state of the art 300 kV (or higher) field emission electron microscopes equipped with liquid nitrogen or helium cold stages. The cost of these instruments, about $2M, is an expensive item for Universities or Institutes but the installation of a relatively small number of these instruments in the United States would represent a large change in the size of the current effort. This program would provide highly leveraged funding.

RECOMMENDATION

Thus we recommend that specific postdoctoral fellowship programs be established to attract promising candidates to high resolution electron microscopy, that equipment funds be made available for the creation of new centers in this field and that the equipment in existing facilities be updated.

A CAREER DEVELOPMENT PROGRAM IN STRUCTURAL BIOLOGY

The problem of solving the structure of a supramolecular complex is a multifaceted one involving innovations in molecular biology, biochemistry and structural biology and thus requires a long term effort. In so far as these problems are currently being attacked, it is most likely to be by postdoctoral fellows in large laboratories with stable funding. The term of a postdoctoral fellowship, usually three years, is incompatible with the length of these projects. A postdoctoral fellow immersed in such a project may not be able to attract the level of support required to finish it in his or her own newly established laboratory and yet the experience he or she has gained may well be crucial to the completion of the project. Thus both science and the careers of young structural biologists could be advanced by a career development program which provided salary funding beyond the initial three year postdoctoral fellowship. This funding should be at a level commensurate with their experience and comparable to that an Assistant Professor would normally receive on assuming a position three years after the doctorate.

RECOMMENDATION

Thus we recommend that a public or private agency or a coalition of agencies establish a Career Development program that will allow creative young structural biologists to take on long term and difficult problems.

REFERENCES

Abrahams, J.P., Leslie, A.G.W., Lutter, R., and Walker, J.E. (1994). Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* 370:621–628.

Alberts, B., Bray, D., Lewis, J., Raff, M., and Roberts, K. *Molecular Biology of the Cell.* Garland Publishers, 3rd edition, 1995.

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92:291–294.

Aiyar, N. (1989). Solubilization of rat liver vasopressin receptors as a complex with a guanine-nucleotide-binding protein and phosphoinositide-specific phospholipase C. *Biochemical Journal* 261:63–70.

Amos, L.A. and Cross, R.A. (1997). Structure and dynamics of molecular motors. *Current Opinion in Structural Biology* 7:239–246.

Avila-Sakar, A.J. and Chiu, W. (1996). Visualization of beta-sheets and side-chain clusters in two-dimensional periodic arrays of streptavidin on phospholipid monolayers by electron crystallography. *Biophysics Journal* 70:57–68.

Bai, C. and Elledge, S.J. (1997). Gene identification using the yeast two-hybrid system. *Methods in Enzymology* 283:141–156.

Baker, T.A. and Bell, S.P. (1998). Polymerases and the replisome: machines within machines. *Cell* 92:295–305.

Baker, T.S. and Cheng, R.H. (1996). A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *Journal of Structural Biology* 116:120–130.

Baker, T.S. and Johnson, J.E. (1996). Low resolution meets high: towards a resolution continuum from cells to atoms. *Current Opinion in Structural Biology* 6:585–594.

Baldwin, J.M., Schertler, G.F., and Unger, V.W. (1997). An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *Journal of Molecular Biology* 272:144–164.

Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R.A., Sweet, R., Frank, J., Moore, P.B., and Steitz, T.A. (1998). A 9 Å resolution X-ray crystallographic map of the large ribosomal subunit. *Cell* 93:1105–1115.

Biosync. (1997). Structural Biology and Synchrotron Radiation: Evaluation of Resources and Needs. Report of Biosync – The Structural Biology Synchrotron Users Organization. The 1997 Biosync committee members and authors of the report were: Ian Wilson, David Davies, Ed Lattman, Jim Penner-Hahn, George Phillips, Jill Trewhella and Janet Smith.

Böttcher, B., Wynne, S.A., and Crowther, R.A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* 386:88–91.

Brent, R. and Finley, R.L. Jr. (1997). Understanding gene and allele function with two-hybrid methods. *Annual Reviews of Genetics* 31:663–704.

Bukau, B. and Horwich, A.L. (1998). The Hsp70 and Hsp60 chaperone machines. *Cell* 92:351–366.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. *Science* 263:802–805.

Chiu, W. and Smith, T.J. (1994). Structural studies of virus-antibody complexes by electron cryomicroscopy and x-ray crystallography. *Current Biology* 4:219–224.

Chiu, W. and Schmid, M.F. (1997). Pushing back the limits of electron cryomicroscopy. *Nature Structural Biology* 4:331–333.

Crowe, J., Masone, B.S., and Ribbe, J. (1996). One-step purification of recombinant proteins with 6xHis tag and Ni-NTA resin. *Methods in Molecular Biology* 58:491–510.

Crowther, R.A., Henderson, R., and Smith, J.M. (1996). MRC image processing programs. *Journal of Structural Biology* 116:9-16.

DeRosier, D.J. (1993). "Turn of the Century Electron Microscopy." *Current Biology* 3:690–692.

DeRosier, D.J. (1997). Who needs crystals anyway? *Nature* 386:26–27.

DeRosier, D.J. and Harrison, S.C. (1997). Macromolecular assemblages. Sizing things up. *Current Opinion in Structural Biology* 7:237–238.

DeRosier, D.J. (1998). The turn of the screw: the bacterial flagellar motor. *Cell* 93:17–20.

Edwards, A.M., Darst, S.A., Feaver, W.J., Thompson, N.E., Burgess, R.R., and Kornberg, R.D. (1991). Purification and lipid layer crystallization of yeast RNA polymerase II. Proc. Natl. Acad. Sci. USA 87:2122–2126.

Ferrell, J.E. (1996). Tripping the switch fantastic: How a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends in Biochemical Sciences* 21:460–466.

Feynman, R. (1960). "There is *plenty* of room at the bottom." Transcript of talk given on December 29, 1959 at the annual meeting of the American Physical Society at the California Institute of Technology. Published by Caltech in the February 1960 issue of *Engineering and Science.* Available from the web at http://nano.xerox.com/nanotech/feynman.html.

Frank J., Radermacher M., Penczek P., Zhu J., Li Y., Ladjadj M., and Leith A. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *Journal of Structural Biology* 116:190–199.

Garboczi, D.N., Ghosh, P., Utz, U., Fan, Q.R., Biddison, W.E., and Wiley, D.C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* 384:134–141.

Garcia, K.C., Degano, M., Stanfield, R.L., Brunmark, A., Jackson, M.R., Peterson, P.A., Teyton, L., and Wilson, I.A. (1996). An αβ T cell receptor structure at 2.5Å and its orientation in the TCR-MHC complex. *Science* 274:209–219.

Grigorieff, N., Ceska, T.A., Downing, K.H., Baldwin, J.M., and Henderson, R. (1996). Electron-crystallographic refinement of the structure of bacteriorhodopsin. *Journal of Molecular Biology* 259:393–421.

Grimes, J.M. (1998a). Large unit cells and cellular mechanics. *Nature Structural Biol.* 5:630–634.

Grimes, J.M., Burroughs, J.N., Gouet, P., Diprose, J.M., Malby, R., Ziéntara, S., Mertens, P.P.C., and Stuart, D.I. (1998b). The atomic structure of the bluetongue virus core. *Nature* 395: 470–478.

Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bertunik, H.D., and Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386:463–471.

Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly Reviews of Biophysics* 28:171–193.

Henderson, R. and Unwin, P.N. (1975). Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* 257:28–32.

Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO Journal* 9: 1665–1672.

Holmes, K.C. (1994). Solving the structures of macromolecular complexes. *Structure* 2:589–593.

Jarvik, J. and Telmer, C. (1998). Epitope Tagging. *Annual Reviews of Genetics* 32:601–618.

Jones, C., Patel, A., Griffin, S., Martin, J., Young, P., O'Donnell, K., Silverman, C., Porter, T., and Chaiken, I. (1995). Current trends in molecular recognition and bioseparation. *Journal of Chromatography* 707:3–22.

Kadonaga, J.T. (1998). Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* 92:307–313.

Koster, A.J., Grimm, R., Typke, D., Hegerl, R., Stoschek, A., Walz, J., and Baumeister, W. (1997). Perspectives of molecular and cellular electron tomography. *Journal of Structural Biology* 120:276–308.

Kozielski, F., Arnal, I., and Wade, R.H. (1998). A model of the microtubule-kinesin complex based on electron cryomicroscopy and x-ray crystallography. *Current Biology* 8:191–198.

Kubalek, E.W., LeGrice, S.F., and Brown, P.O. (1994). Two-dimensional crystallization of histidine-tagged, HIV-1 reverse transcriptase promoted by a novel nickel-chelating lipid. *Journal of Structural Biology* 113:117–123.

Kuhlbrandt, W., Wang, D.N., and Fujiyoshi, Y. (1994). Atomic model of plant light-harvesting complex by electron crystalllography. *Nature* 367:614–621.

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., and Hendrickson, W.A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393:648–659.

Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4Å resolution. *Science* 268:533–539.

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 angstrom resolution. *Nature* 389:251–260.

Malhotra, A., Penczek, P., Agrawal, R.W., Gabashvili, I.S., Grassucci, R.A., Junemann, R., Burkhardt, N., Nierhaus, K.H., and Frank, J. (1998). Escherichia coli 70S ribosome at 15 Å resolution by cryo-electron microscopy: localization of fMet-tRNAfMet and fitting of L1 protein. *Journal of Molecular Biology* 280:103–116.

Matlack, K.E., Mothes, W., and Rapoport, T.A. (1998). Protein translocation: tunnel vision. *Cell* 92:381–390.

Niswender, K.D., Blackman, S.M., Rohde, L., Magnuson, M.A., and Piston, D.W. (1995). Quantitative imaging of green fluorescent protein in cultured cells: comparison of microscopic techniques, use in fusion proteins and detection limits. *Journal of Microscopy* 180:109–16.

Nogales, E., Wolf, S.G., and Downing, K.H. (1998). Structure of the αβ tubulin dimer by electron crystallography. *Nature* 391:199–203.

Noji, H., Tasuda, R., Yoshida, M., and Kinosita, K. (1997) Direct observation of the rotation of F1-ATPase. *Nature* 386:299–302.

Ohno, M., Fornerod, M., and Mattai, I.W. (1998). Nucleocytoplasmic transport — The last 200 nanometers. *Cell* 92:327–336.

Park, S.H. and Raines, R.T. (1997). Green fluorescent protein as a signal for protein-protein interactions. *Protein Science* 6:2344–2349.

Penner, C.G., Wang, Z., and Litchfield, D.W. (1997). Expression and localization of epitope-tagged protein kinase CK2. *Journal of Cellular Biochemistry* 64:525–537.

Phizicky, E.M. and Fields, S. (1995). Protein-protein interactions: methods for detection and analysis. *Microbiological Reviews* 59:94–123.

Ranganathan, R. and Ross, E.M. (1997). PDZ domain proteins: Scaffolds for signaling complexes. *Current Biology* 7:R770–R773.

Raw, A.S., Coleman, D.E., Gilman, A.G., and Sprang, S.R. (1997). Structural and biochemical characterization of the GTPγ-S-, GDP.Pi-, and GDP-bound forms of a GTPase-deficient Gly42→ Val mutant of Giα1. *Biochemistry* 36:15660–15669.

Rayment, I., Rypniewski, W.R., Schmidt-Base, K., Smith, R., Tomchick, D.R., Benning, M.M., Winkelmann, D.A., Wesenberg, G., and Holden, H.M. (1993a). Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science* 261:50–58.

Rayment, I., Holden, H.M., Whittaker, M., Yohn, C.B., Lorenz, M., Holmes, K.C., and Milligan, R.A. (1993b). Structure of the actin-myosin complex and its implications for muscle contraction. *Science* 261: 58–65.

Rhodes, G. (1993). Crystallography made crystal clear: A guide for users of macromolecular models. Academic Press, Inc., San Diego, CA.

Sharrocks, A.D. (1994). A T7 expression vector for producing N- and C-terminal fusion proteins with glutathione S-transferase. *Gene* 138:105–108.

Sprang, S.R. (1997). G Proteins, effectors and GAPs: structure and mechanism. *Current Opinion in Structural Biology* 7:849–856.

Song, H.Y., Dunbar, J.D., and Donner, D.B. (1994). Aggregation of the intracellular domain of the type 1 tumor necrosis factor receptor defined by the two-hybrid system. *Journal of Biological Chemistry* 269:22492–22495.

Sosa, H., Dias, D.P., Hoenger, A., Whittaker, M., Wilson-Kubalek, E., Sablin, E., Fletterick, R.J., Vale, R.D., and Milligan, R.A. (1997). A model for the microtubule-Ncd motor protein complex obtained by cryo-electron microscopy and image analysis. *Cell* 90: 217–224.

Staley, J.P. and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92:315–326.

Stark, H., Orlova, E.V., Rinke-Appel, J., Junke, N., Mueller, F., Rodnina, M., Wintermeyer, W., Brimacombe, R., and van Heel, M. (1997). Arrangement of tRNAs in pre- and posttranslational ribosomes revealed by electron cryomicroscopy. *Cell* 88:19–28.

Sunahara, R.K., Tesmer, J.J., Gilman, A.G., and Sprang, S.R. (1997). Crystal structure of the adenylyl cyclase activator Gsα. *Science* 278:1943–1947.

Trus, B.L., Roden, R.B., Greenstone, H.L., Vrhel, M., Schiller, J.T., and Booy, F.P. (1997). Novel structural features of bovine papillomavirus capsid revealed by a three-dimensional reconstruction to 9 Å resolution. *Nature Structural Biology* 4:413–420.

Tsien, R.Y. (1998). The green fluorescent protein. *Annual Reviews of Biochemistry* 67:509–544.

Unger, V.M. and Schertler, G.F. (1995). Low-resolution structure of bovine rhodopsin determined by electron cryomicroscopy. *Biophysics Journal* 68:1776–1786.

Unwin, N. (1995). Acetycholine-receptor channel imaged in the open state. *Nature* 373:37–43.

Unwin, N. (1998). The nicotinic acetylcholine-receptor of the Torpedo electric ray. *Journal of Structural Biology* 121:181–190.

van Heel, M., Harauz G., and Orlova, E.V. (1996). A new generation of the IMAGIC image processing system. *Journal of Structural Biology* 116:17–24.

Walz, J., Typke, D., Nitsch, M., Koster, A.J., Hegerl, R., and Baumeister, W. (1997). Electron tomography of single-ice-embedded macromolecules: Three dimensional alignment and classification. *Journal of Structural Biology* 120:387–395.

Walz, J., Erdmann, A., Kania, M., Typke, D., Koster, A.J., and Baumeister, W. (1998). 26S proteasome structure revealed by three-dimensional electron microscopy. *Journal of Structural Biology* 121:19–29.

Wang, L.F., Yu, M., White, J.R., and Eaton, B.T. (1996). Btag: a novel six-residue epitope tag for surveillance and purification of recombinant proteins. *Gene* 169:53–58.

Wilson, C.J., Chao, D.M., Imbalzano, A.N., Schnitzler, G.R., Kingston, R.E., and Young, R.A. (1996). RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell* 84:235-244.

Wilson, K.S. and Noller, H.F. (1998) Molecular movement inside the translational engine. Cell 92:337–349.

Wüthrich, K. (1995). NMR: This other method for protein and nucleic acid structure determination. *Acta Crystallography* D51:249–270.

Xu, Z., Horwich, A.L., and Sigler, P.B. (1997). The crystal structure of the asymmetric GroEL-GroES-$(ADP)_7$ chaperonin complex. *Nature* 388:741–750.

Yasuda, R., Noji, H., Kinosita, K., and Yoshida, M. (1998). F1-ATPase is a highly efficient molecular motor that rotates with discrete 120 degrees steps. *Cell* 93:1117–1124.

Zhang, P.J., Toyoshima, C., Yonekura, K., Green, N.M., and Stokes, D.L. (1998). Structure of the calcium-pump from sarcoplasmic reticulum at 8 Å resolution. *Nature* 392:835–839.

Zhu, J., Penczek, P.A., Schroder, R., and Frank, J. (1997). Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S Escherichia coli ribosome. *Journal of Structural Biology* 118:197–219.

THE AGOURON INSTITUTE

Melvin Simon, PhD	Chairman of the Board
John Abelson, PhD	President
Willis Wood, PhD	Vice President
Philip Abelson, PhD	Board member
Gustaf Arrhenius, PhD	Board member
Theodore Friedmann, MD	Board member
Gordon Gill, MD	Board member
David Hirsh, PhD	Board member
Deborah Spector, PhD	Board member
Joan Kobori, PhD	Program Director

SUPRAMOLECULAR ASSEMBLIES ADVISORY COMMITTEE

David Agard, PhD	University of California, San Francisco
David Eisenberg, PhD	University of California, Los Angeles
Richard Henderson, PhD	Medical Research Council, Cambridge, UK
Wayne Hendrickson, PhD	Columbia University
Sung-Hou Kim, PhD	University of California, Berkeley
John Kuriyan, PhD	Rockefeller University
Harry Noller, PhD	University of California, Santa Cruz
Douglas Rees, PhD	California Institute of Technology
Paul Sigler, PhD	Yale University
Thomas Steitz, PhD	Yale University

THANKS TO:

Pamela Bjorkman, PhD (California Institute of Technology) who served as our consultant and authored much of the report.

The advisory committee and Wah Chiu, PhD (Baylor University) for advice and comments on the report. Kai Zinn for assistance with the current technology section.

Jamie Williamson of The Agouron Institute, La Jolla for assistance to The Agouron Institute in Pasadena.

Writers/editors: Pamela Bjorkman, John Abelson, Joan Kobori,
Contributing writers: Willis Wood, Kai Zinn
Design: Denton Design Associates

Photo credits:
Cover: Zhaohui Xu and Paul Sigler; page12: Zhaohui Xu and Paul Sigler; page14: John Walker; page20: Tom Earnest; page24: Kenneth Downing; page25: Richard Henderson; page26: Tony Crowther; page27: FEI Company; and page29: Nenad Ban and Thomas Steitz

The Agouron Institute
201 South Lake Avenue, Suite 410
Pasadena, CA 91101
tel: (626) 577-0500
fax: (626) 577-0800