Chapter 10a

Recent Developments in Molecular Biology — An Overview

Reprint from:
H.-J. Rehm and G. Reed (Editors)
Biotechnology
Vol. 7b
VCH Verlagsgesellschaft, Weinheim, 1989

Wolfram Hörz

Institut für Physiologische Chemie
der Universität München
München, Federal Republic of Germany

Manfred Kröger

Institut für Mikrobiologie und Molekularbiologie
der Universität Gießen
Gießen, Federal Republic of Germany

Dietrich Kuschmitz

Max-Planck-Institut für Ernährungsphysiologie
Dortmund, Federal Republic of Germany

Andreas Plückthun

Gen-Zentrum der Universität München
Max-Planck-Institut für Biochemie
Martinsried, Federal Republic of Germany

Following are some short reviews of topics covered in Volume 7b, first published in the Annual Review of Biochemistry and Molecular Biology for 1988 in the German journal *Nachrichten aus Chemie Technik und Laboratorium 37* (2) (1989).

I. Molecular Biology of Plants

Certain domestic plants have proved difficult to be transformed as cell cultures. For many years, only dicotyledons were believed to be suitable. After some successful transformations of monocotyledons (primarily of grain plants) were published, a systematic approach to this problem is now in sight. A symbiosis between the tumor-inducing (Ti) plasmids from the soil bacteria *Agrobacterium tumefaciens* and plants is induced by certain phenolic substances (e.g., acetosyringon). It was believed that monocots could not form these substances, until SCHELL and coworkers successfully achieved this through direct DNA transformation of young floral tillers (DE LA PEÑA et al., 1987). USAMI and colleagues (1988) have proposed that monocots, e.g., in wheat and oats, possess also this type of inductors, which are growth-dependent macromolecules. This means that resting seeds do not contain these inductors and, therefore, cannot be transformed. Transformations may now be generally possible in certain tissues, when it is possible to overcome the inhibitors present (NAS, 1988). Noteworthy progress has also been made in studies on the mechanisms of Ti-plasmid virulence. In general, Ti transfer is similar to conjugation between two bacteria, in which again a single DNA strand is transferred (CITOVSKY et al., 1988; DAS, 1988). Integration itself strikingly resembles the mechanisms of insertion elements and retroviruses (TIMMERMAN et al., 1988).

As with grain, many attempts have been made to adapt soybeans to genetic manipulation. The soybean accounts for some 25% of world-wide fat requirements and is therefore of utmost economic importance. CHRISTOU and colleagues at Agrocetus have bombarded soy cells with DNA adsorbed onto gold. This technique uses an electrical field to accelerate metal spherules through the cellular wall and membrane (CHRISTOU et al., 1988; MCCABE et al., 1988; WIDHOLM, 1988). KLEIN has successfully used tungsten in similar experiments at Cornell University (GRAY, 1988). Even though yields have not yet been optimized, this technique opens up the astonishing possibility that one could introduce foreign DNA into a cell without using a vector. The recombination system of the cell assimilates the DNA into the genome without the risk that the DNA could become autonomous through viral or vectorial means. This is seen as an important step in the improvement of biological safety and has been successfully applied in animal systems (MANSOUR et al., 1988). Moreover, recent findings on the mechanism of damage caused by asbestos fibers show that Nature already "used" this method (APPEL et al., 1988).

The alteration of a plant's genetic composition serves two main applications: to make the plant more resistant to disease and pests and to increase its value as food. This former goal is often met with skepticism, as domesticated plants are bred more and more away from originally variable gene pools. For instance, LEE and coworkers (1988) have presented an especially elegant way of plant protection, which if improperly handled could have really serious consequences. These investigators have isolated the tobacco gene for the enzyme acetolactate synthetase, manipulated it, and put it back into the plant. Before, tobacco was sensitive to the herbicide sulfonyl urea. Now, the binding characteristics of the target enzyme to the herbicide were changed through mutation, and a viable, resistant plant was the result. This approach is interesting, but worrisome since the destruction of non-resistant plants could result in a

domestic plant monoculture (see also SCHLOSS et al., 1988).

In 1987, the heterologous expression of bacterial insecticides was considered to be an especially elegant method of active plant protection. Now TOONG and coworkers have shown that Nature itself has done the same. A type of Malaysian cyperus grass contains a grasshopper development hormone. Female grasshoppers that feed on this grass lay significantly underdeveloped eggs. Since this is not a peptide hormone, it would be very difficult to artificially transfer this type of resistance; one would have to clone and transfer entire pathways of genes. But at least a cheap source for the hormone has been found (TOONG et al., 1988).

Two examples of attempts to create more valuable food plants are corn and tomatoes. Work is in progress to supplement the storage protein of corn with the amino acids lysine and tryptophan. In a heterologous expression experiment in frog oozytes, a fully synthetic mRNA could be transcribed into a aggregatable protein (WALLACE et al., 1988). In general, the genetic transformation of corn no longer seems to present any problems (RHODES et al., 1988). Several groups have begun work on the improvement of the quality of quickly ripening supermarket tomatoes. Antisense RNA was merely a concept several years ago; success is now in sight. This synthetic inhibitor complexes with indigenous mRNA and prevents premature expression of polygalacturonase in the ripening fruit, which digests fruit cell walls. The tomatoes turn red and sweet, but the skin remains firm (SMITH et al., 1988; SHEEHY et al., 1988; ROBERTS, 1988). In principle a number of studies could use antisense RNA to help understand the regulation of gene expression in plants (LICHTENSTEIN, 1988; VAN DER KROL et al., 1988 a, b).

The inducing mechanism of plant hormones is another topic to which much effort is directed. Best known are auxin and ethylene, which control cell growth and ripening, respectively. The molecular mechanisms of both hormones are being studied in order to clarify binding sites and the pathways within signal chains. Auxin seems to induce a signal chain across the membrane similar to that found in animals. ETTLINGER and LEHLE (1988) were able to determine that auxin induces a sudden change in phosphatidylinositol metabolism. According to MARGOSSIAN and colleagues (1988), ethylene regulates the ripening process of tomatoes via a protease inhibitor. DEIKMAN and FISCHER (1988) investigated an ethylene-dependent tomato gene to find the binding site for the regulating factor. Since the DNA sequence is now known and the factor has also been isolated, one can expect further progress on this topic in the near future. A more classical route was taken by BLEECKER and colleagues (1988) who studied ethylene-insensitive mutants of *Arabidopsis thaliana*. This plant is especially interesting to botanists because it is the smallest of the higher plants, can easily be grown in a test tube and is mature at 6 weeks. *Arabidopsis thaliana* was the first higher plant to produce a mutant which had a defect in an amino acid synthesis pathway (LAST and FINK, 1988).

A less well-known group of hormones are the stress hormones. WALBOT and BRUENING (1988) presented the status of this research in a congress report. It will still take some time before it is understood why the stress hormone abscisic acid (ABA) reacts to a lack of water through synthesis of a protein especially rich in glycine (GOMEZ et al., 1988).

Of greatest influence on general molecular biology in 1988 have probably been the newly discovered plant RNA enzymes. These RNAs do indeed deserve the name since they are independent catalytic substances. They were discovered as covalently bound activities of viroids, which have puzzled researchers for years. After it was discovered that viroids and other RNAs could split themselves, it was discussed whether it would be possible to bring this activity to RNA from the outside. This idea was proposed by HASELOFF and GERLACH (1988) of Canberra. They could pinpoint the catalytic activity so that the mechanism can be accepted as mostly explained. Only one RNA sequence of at least 39 nucleotides

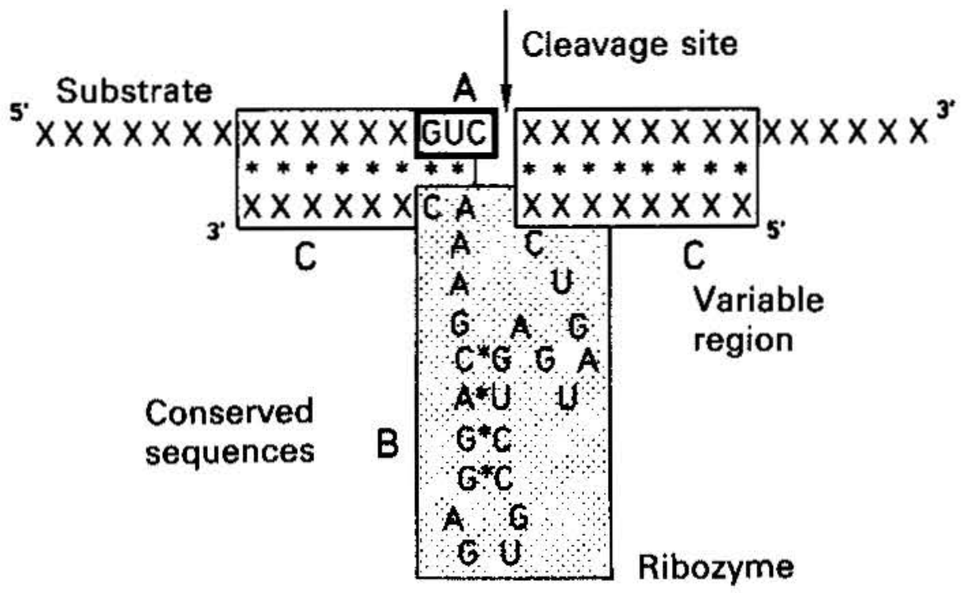

Figure 1. Model for the design of ribosomes.

with a high degree of self-structure is required. The target RNA is spliced just behind the GUC sequence (area A in Fig. 1). Twenty-two nucleotides are catalytically active (area B), while 14 nucleotides (area C) can be variable. Through base-pairing of these nucleotides to the target RNA, the specificity of the reaction is defined. The system has been tested *in vitro* on different positions of identical target RNA (HASELOFF and GERLACH, 1988). WALBOT and BRUENING (1988) comment that these RNA enzymes could replace the use of antisense RNA *in vivo* since previously synthesized RNA can be actively destroyed. This could be applicable to animal systems. Technical problems, however could always occur when the target RNA contains a high degree of secondary structure.

These RNA enzymes are leading to an understanding of the biosynthesis of viroids (FORSTER et al., 1988). Although intensive research on viroids is underway, not much is known about their pathogenic mechanisms. HIDDINGA and coworkers (1988) suggest that viroids can induce phosphorylation of host-proteins. These modified plant proteins are obviously structurally related to a double-stranded RNA-binding protein found in mammalian cells. The plant/viroid interaction shows some similarities to previous experience in interferon action. The importance of this discovery has yet to be seen.

II. Membrane Biochemistry

The crystallization of membrane proteins is still rarely reported by membrane biochemists. Genetic engineering, including site-directed mutagenesis, is used for the analysis of structure-function relationships. The summary of a wealth of detailed experimental data, which sometimes involves only the exchange of one amino acid for another, can lead to the development of a unified concept. The involvement of allosteric receptor oligomerization in the EGF (epidermal growth factor) receptor: signal transfer is an example (SCHLESSINGER, 1988). Different mutants show that intrinsic tyrosine kinase activity is necessary for signal transfer. Autophosphorylation results from intermolecular cross-phosphorylation within an oligomeric receptor complex.

Potassium channels probably form the most complicated groups of ion channels. A new principle has been found that explains the diversity of the potassium A channel: alternative splicing of transcripts from a single gene. While the membrane potential-dependent sodium channels as well as the calcium channels of the dihydropyridine receptor are large proteins consisting of four repeated 30 kD pseudo-subunits, the Shaker locus proteins of the *Drosophila* genome, forming potassium channels, consist of only one such pseudo-subunit. SCHWARZ et al. (1988) have shown that the Shaker locus is a complicated transcription unit that codes for many different proteins. All proteins are probably multiple components of the K^+ A channel and are formed from a family of transcripts through alternative splicing. These proteins seem to contribute to the different sub-types of the channel and offer a biological explanation for the diversity of this class of potassium channels.

III. Protein Folding

In the last few years a surge of activity in the once somewhat quieter field of protein folding was noted. While many key experiments had been carried out before this time (reviewed e.g. by ANFINSEN, 1973, and JAENICKE, 1987), the advent of gene technology and new methodology are mostly responsible for the increasing activity. Recombinant proteins are often obtained from inclusion bodies by heterologous expression in bacteria, but have to be refolded to an active conformation to be sold by industry. Gene technology is not only the trigger for this research, but it also supplies at least three key methodologies: (1) The production of proteins in sufficient quantity for physical studies, (2) the alteration of a protein to (almost) any desired amino acid sequence, and (3) the tools for the investigation of protein folding *in vivo*. Recent advances in NMR have also shown great promise for progress in this field.

The folding of a protein (more precisely: of a single domain) is generally seen as a cooperative process. That is, in a denaturant induced unfolding experiment, the protein solution consists of a mixture of molecules in either the native or the unfolded state at any concentration of denaturant and stable intermediates cannot be obtained. This makes it very challenging to elucidate the structure of transient folding-intermediates. Two research groups have attempted to solve this problem using NMR methods (UTGAONKAR and BALDWIN, 1988; RODER et al., 1988). The protein (for which all amide proton resonance assignments must be known) is denatured in D_2O, and the progression of proton capture during the renaturation in H_2O is monitored. This method was applied in an investigation of the folding of ribonuclease and cytochrome c. It essentially determines the time point during the folding at which individual regions of the protein are no longer in contact with the surrounding water.

In a different approach, OAS and KIM (1988) investigated a peptide model of bovine pancreatic trypsin inhibitor (BPTI), which may represent a disulfide-bridged folding intermediate.

It has been known for some time that in some cases the slow step of protein folding can be the isomerization of a peptide bond that is located *N*-terminal to proline. An enzyme was discovered (FISCHER et al., 1984) that catalyzes this process *in vitro*, though only very moderately. LANG et al. (1987) could show that it is not identical with the enzyme that causes the isomerization of disulfide bonds (LANG and SCHMID, 1988). The latter enzyme appears to be essential for the correct folding of at least some proteins (BULLEID and FREEDMAN, 1988). On the other hand, it is not yet known whether proline-*cis-trans*-isomerase is really an essential protein nor whether there might even be several isomerases. It is also not clear that this protein actually catalyzes this reaction *in vivo*. It is thus even more perplexing that this "enzyme" was now shown to be identical to the protein cyclophilin, which binds the immuno-suppressant peptide cyclosporin very tightly, and was also shown to be homologous to a protein involved in visual transduction in *Drosophila* (FISCHER et al., 1989; TAKAHASHI et al., 1989; SHIEH et al., 1989).

Another long-standing question has been the relationship between protein folding and protein transport through membranes. An increasing number of experiments now suggest that this transport is only possible for an unfolded form of a protein. PARK et al. (1988) and LAMINET and PLÜCKTHUN (1989) have now provided evidence that the presence of a signal sequence can influence the kinetics of folding without significantly changing the final folded state.

The exact function of the eukaryotic heat-shock proteins (hsp), the glucose regulated proteins (GRP) and their relatives is still shrouded in mystery, but evidence accumulates that these proteins may have something to do with cellular protein folding. KASSENBROCK et al. (1988) found that a protein (named BiP) that is known to be involved in the assembly of antibodies, also

binds to aberrant proteins, at least *in vitro*. When a mutated hemagglutinine is produced that cannot form all disulfide bonds, BiP and some other proteins are apparently induced *in vivo* (KOZUTSUMI et al., 1988). Two research groups have published data indicating a participation of Hsp 70-related proteins in the transport of proteins into the endoplasmic reticulum as well as into mitochondria (DESHAIES et al., 1988; CHIRICO et al., 1988). The effect of the Hsp 70 protein is ATP dependent, and the authors suggest that its function is the enzymatic "unfolding" of precursor proteins.

A chloroplast protein that is involved in the assembly of oligomeric proteins in plants was discovered by HEMMINGSEN et al. (1988). Interestingly, it is homologous to the *GroEL* gene of *Escherichia coli*. GroEL and GroES are known to be required for phage head assembly, but their mechanism of action is unknown. GroEL may have a function similar to that of the eukaryotic heat-shock proteins (BOCHKAREVA et al., 1988), since it associates with newly synthesized proteins and can only dissociate from them in the presence of ATP. GroES/GroEL promotes the assembly of foreign oligomeric proteins in *E. coli* (GOLOUBINOFF et al., 1989).

Several current ideas about the relation between protein folding and protein transport have recently been summarized by MEYER (1988).

IV. Protein Structures

A. X-Ray Crystallography

The crucial importance of the knowledge of the structure of a protein for understanding its function was emphasized through the selection of three crystallographers and biochemists as recipients of the 1988 Nobel prize in chemistry. DEISENHOFER, HUBER, and MICHEL (1988a, b) elucidated the structure of the photosynthetic reaction center of *Rhodopseudomonas viridis*. This work not only provided a most decisive contribution to the understanding of photosynthesis, but also, by the development of techniques for crystallizing membrane proteins, may open a completely new area: the structural study of membrane proteins.

The study of conformational changes is one of the current topics in protein crystallography. Phosphorylase certainly qualifies as an interesting case in point. This molecule was crystallized in two different conformations by two different research groups so that now, through the comparison of the two structures, the mechanism of the allosteric conformational changes can be elucidated (SPRANG et al., 1988). It is possible that some general principles of conformational changes through protein phosphorylation can be deduced from this model. The conformational transitions of the most thoroughly investigated of all allosteric proteins, hemoglobin, can now be described in more detail, since in addition to the deoxyT- and oxyR forms, the crystal structure of the oxyT form has now also been determined (LIDDINGTON et al., 1988). The conformational changes which take place in glyceraldehyde phosphate dehydrogenase by cofactor binding have now also been investigated (SKARZYNSKI and WONACOTT, 1988) as has a conformational change of adenylate kinase (DREUSICKE and SCHULZ, 1988).

The year 1988 was also marked by the elucidation of several structures of DNA-binding proteins complexed with DNA (MATTHEWS, 1988). By means of the structural analysis of the *cro* repressor of phage 434 complexed with DNA, WOLBERGER et al. (1988) could show that the conformation of DNA changes with the binding of the repressor. The comparison of the complex of the λ repressor with the similar repressor from phage 434 (JORDAN and PABO, 1988) will provide substantial information about the exact mechanism of operator-sequence recognition. In the *trp* repressor (OTWINOWSKI et al., 1988), it is not the bases, but

the precise position of the phosphates of DNA which is recognized. The location of the phosphate groups is in turn sequence-dependent. This information was derived from co-crystals of the *trp* repressor and its operator. From this crystal structure, the mechanism of the inducer could be elucidated as well (LAWSON and SIGLER, 1988). Enzymes that work on DNA were also crystallographically studied: among these were DNA-polymerase I (DERBYSHIRE et al., 1988) and DNAse I complexed with "nicked" DNA (SUCK et al., 1988).

The structure of DNA itself continues to be actively investigated. Unusual DNA duplex structures with unpaired or looped-out bases were solved by JOSHUA-TOR et al. (1988) and MILLER et al. (1988).

How is it possible for certain fish to live in water colder than 0 °C? Although these fish do not have an elevated body temperature, their body fluids do not freeze. A polypeptide, whose structure has now been elucidated, prevents this (YANG et al., 1988). The polypeptide is a single α-helix that apparently attaches to the nuclei of ice crystals and prevents their growth. The detailed understanding of the interactions between proteins and water or ice is obviously of great general interest.

Obtaining appropriate single crystals of macromolecules is a difficult art and new methods of crystallization are continuously being sought. A new example is the growth nucleation of protein crystals by means of mineral surfaces (MCPHERSON and SHLICHTA, 1988).

Recently, the Laue X-ray diffraction method (which employs a stationary crystal and polychromatic X-rays) has been revived. At least in principle, a complete set of data can be obtained in milliseconds (FARBER et al., 1988) – an exciting prospect for the investigation of conformational changes.

B. NMR Spectroscopy

Recently R. ERNST's group (GRIESINGER et al., 1987) pioneered methods by which NMR resonances can be split into three dimensions. This extension from the currently used 2D-NMR gives yet another increase in resolution and reliability of the assignments. In 1988, two groups showed independently that this technique can be applied to proteins and offers distinct advantages (OSCHKINAT et al., 1988; VUISTER et al., 1988).

There is a limit to the size of a protein that can be investigated by NMR. As the size of the protein increases, the spin-spin relaxation time decreases and the resulting increase in resonance linewidth makes the resolution of individual peaks impossible. Recently, LEMASTER et al. (1988) found a way to circumvent this problem. The solution consists of partially deuterating the protein (to about 75%) which then causes proton-proton-relaxation to be much less dominant, leading to narrow lines. Also, some decoupling will be observed, again causing multiplets to narrow. The production of the large amounts of deuterated protein required will normally only be possible in microorganisms.

The computational methods, with which protein structures can be obtained from NMR data, are also continuously being improved (NILGES et al., 1988; HOLAK et al., 1988).

How do protein structures derived from crystallography compare with those derived from NMR? One of the pioneers of the latter method has recently critically reviewed this question (WÜTHRICH, 1989a, b). Clearly, the overall folding of the proteins for which structures are available from both methodologies were usually found to be very similar. Yet, there are differences in the conformation of surface residues that appear to be real and beyond experimental uncertainty. Small proteins will be more distorted by crystal contacts than larger ones and if several conformations are available, only one might crystallize. These factors might contribute to the fact that metal-

lothionein has different architecture and metal coordination in the crystal than in solution (ARSENIEV et al., 1988; SCHULTZE et al., 1988).

V. Protein Engineering

Trying to understand electrostatics is one of the current themes in protein engineering. LINSE et al. (1988) altered some surface charges in the protein calbindin and investigated the effect on ion binding. The altered amino acids were not in direct contact with the ion, but merely changed the electric field. The influence of the protein structure on the binding of ions was also analyzed using free energy calculations. HWANG and WARSHEL (1988) came to the conclusion that, as a rule, ion pairs cannot be exchanged. BRANGE et al. (1988) described a practical application of the engineering of ionic interactions. They developed monomeric insulins that have improved pharmacokinetic characteristics over those of the natural molecule.

Another aspect of protein engineering where increasing activity can be noted is the investigation of protein stability through the crystallographic and biochemical analysis of a large number of different mutants (MATSUMURA et al., 1988; ALBER et al., 1988; NICHOLSON et al., 1988). It could be shown in these studies on T4 lysozyme that the hydrophobic stabilization of a protein is proportional to the surface area that, after folding, is no longer accessible to the solvent. Interestingly, the replacement of a proline residue with various other amino acids right in the middle of an α-helix did not lead to a stabilization of this protein. While the helix itself might be more stable after substitution it might interact less will with the rest of the protein. Stabilization of the protein was, however, achieved through mutations of residues that interact with helix dipoles. This last phenomenon was also observed in studies of a bacterial ribonuclease ("barnase") (SALI et al., 1988). The importance of the precise packing of hydrophobic amino acids was demonstrated by KELLIS et al. (1988) through the use of several barnase mutants. All these studies show the enormous complexity of the protein stabilization problem. The net effect on stability is dependent upon the sum of all the individual interactions within the protein.

The influence of newly constructed disulfide bridges in T4-lysozyme was also investigated in detail (WETZEL et al., 1988). According to WETZEL's results, disulfides may prevent the denatured protein from aggregating or reaching conformations from which refolding is kinetically impossible. Disulfides may thus stabilize a protein against irreversible denaturation, but not necessarily against denaturation under reversible conditions.

The study of enzyme mechanisms through the use of carefully selected mutants together with both kinetic and structural investigations is now being carried out for an ever increasing number of systems. Pioneering work in this area was undertaken in the last few years by the groups of KNOWLES (using triosephosphate isomerase) and FERSHT (using tyrosyl-tRNA-synthetase). In this context, the discussion of the validity of linear free energy relationships between mutant enzymes between ESTELL and FERSHT (1987) is worthwhile reading.

VI. Antibodies

In 1988 the expression of functional, antigen-binding antibody fragments in *Escherichia coli* was first reported. It was shown (SKERRA and PLÜCKTHUN, 1988) in these experiments that the so-called F_V fragment, which consists of only the variable domains

and was obtained from *E. coli* as a correctly associated heterodimer, had the same binding constant to the antigen as the entire mouse antibody. This model system is suitable for protein engineering studies since the crystal structure of the F_{ab} fragment is known. With a similar system, a chimeric F_{ab} fragment could also be expressed in *E. coli* (BETTER et al., 1988). With a comparable strategy, yeast was employed as the host for the secretion of an entire chimeric antibody or an F_{ab} fragment (HORWITZ et al., 1988), although with lower yields than was possible with the *E. coli* system. The microbial expression is versatile (and easier to use than the expression in cell lines) and antibody research will probably benefit from this development (WETZEL, 1988).

With this new technology, the rapidly developing area of catalytic antibodies may also receive a new impetus. Up until now, monoclonal antibodies have been used as "black box" catalysts in several experiments, but protein engineering, that is the rational redesign of the catalytic protein itself, has not been reported. The idea of catalytic antibodies has been worked on for two decades, but has led to success only in the last two years, mostly as a consequence of the availability of monoclonal antibodies. All can be traced back to Linus PAULING's 1946 recognition that an enzyme stabilizes transition states better than the starting material or product of a reaction. The key strategy was therefore to raise antibodies to stable analogs of the transition state and investigate the antibody-catalyzed turnover of substrate (ground state) molecules. This concept has been used to catalyze several types of reactions, albeit still with quite moderate rate enhancements (reviews: KRAUT, 1988; SCHULTZ, 1988).

Two research groups have attempted to produce a single chain molecule from the F_V fragment of an antibody. These so-called "single chain antibodies" were obtained from inclusion bodies of *E. coli* and refolded *in vitro* to molecules that had antigen-binding activity, albeit with lower affinity than the corresponding original antibody (HUSTON et al., 1988; JACOBSON et al., 1988).

The binding sites of the complement component C1q on IgG could be characterized through the expression of appropriately mutated whole antibodies expressed in myeloma cells (DUNCAN and WINTER, 1988). By the grafting of whole hypervariable loops onto a human framework, the binding specificity of several antibodies could be changed ("humanization of antibodies") (RIECHMANN et al., 1988; VERHOEYEN et al., 1988).

VII. Peptides with Designed 3D-Structure

Some years ago, DEGRADO's work on the design and synthesis of peptides with a high propensity for amphipathic helix-structure lead to peptides that have a high tendency to aggregate as helix bundles. In recent work, a 4-helix bundle was covalently connected to form a contiguous sequence. The long peptide was not synthesized but obtained from *E. coli* by using a synthetic gene (REGAN and DEGRADO, 1988). The folded peptide was surprisingly stable. Such a bundle of (unconnected) helices was also used in the design of an ion-specific membrane channel (LEAR et al., 1988).

The design of water-soluble β-pleated sheet structures is apparently much more difficult but has now also met with success (J. RICHARDSON, quoted by KIM, 1988; ERICKSON et al., Proceedings of the 1988 Miami Biotechnology Winter Symposium).

DYSON et al. (1988) have investigated the intriguing problem of which peptides could function as the nuclei in the folding process through a systematic NMR investigation of the structure of a series of very similar short peptides in solution. In this work, the intrinsic tendency of a molecule to nucleate a reverse turn as well as an α-helix was studied.

VIII. Theory of Protein Structure and Enzyme Function

Quantum mechanics and molecular dynamics have now been combined by several research groups (ZHENG et al., 1988; WARSHEL et al., 1988). This combination of methods was employed to theoretically analyze free energy changes between different protein-ligand complexes. Improved computer power now also allows molecular dynamics simulations that directly complement structural experiments – an approach essential for the improvement of the methodology. One recent example is an insulin derivative which was simulated in the crystalline state with water and ions present (YUN-YU et al., 1988). Another example is a comparison of experimental inelastic neutron scattering spectra of bovine pancreatic trypsin inhibitor to spectra calculated from molecular dynamics (CUSACK et al., 1988). The quality of the force-field (MAPLE et al., 1988; HOWARD et al., 1988) is also currently being investigated, as is the quality of different water models (HERMANS et al., 1988).

Several recent publications deal with comparative studies of protein structures. ROOMAN and WODAK (1988) have very systematically analyzed the Brookhaven data base to determine whether short peptide sequences with strong propensities for certain secondary structures exist. While some sequence patterns have been found, only a very small part of a sequence can be predicted with acceptable accuracy, and it remains to be seen whether the situation will be much remedied by a larger structural database. Two research groups report that they have found preferences for certain amino acids at the beginning and at the end of α-helices (PRESTA and ROSE, 1988; RICHARDSON and RICHARDSON, 1988). This may be one more of the many facets that all together determine the final tertiary structure of a protein. NOVOTNY et al. (1988) describe criteria which may be useful for better distinguishing between correct and incorrect model structures.

IX. Protein Chemistry

Part of today's standard procedure for cloning a gene includes the determination of the amino acid sequence of some peptides of the protein to be cloned, usually using a very small amount of the purified protein. Appropriate oligonucleotides are then synthesized to identify the clone in a DNA-library by hybridization. The correct determination of as long a peptide sequence as possible from extremely small amounts of protein (usually a few picomoles) is therefore a key step in molecular cloning. Several research groups have developed techniques by which an *N*-terminal protein sequence may be obtained directly from a protein that is blotted onto a special membrane (reviewed by T. E. KENNEDY, 1988). A new membrane has recently been developed that, in addition to improved amino acid sequences, also allows the more delicate amino acid analysis directly from protein blots (ECKERSKORN et al., 1988). Amino acid analyses are now possible with still smaller quantities of material with the relatively new technique of capillary zone electrophoresis. Using laser-induced fluorescence, fluorescein derivatives of amino acids can be separated and determined at the sub-attomol level (only several thousand molecules) (CHENG and DOVICHI, 1988; GORDON et al., 1988). However, this technique is not ready yet for practical applications, since normally it will be very difficult to obtain samples of the extreme purity required.

X. Protein Biosynthesis

The decisive factor which is ultimately responsible for the nature of the genetic code is the specificity of the enzymes that couple the amino acids to their cognate tRNAs. HOU and SCHIMMEL (1988) found that, at least in the cases they investigated, the identity of the tRNA is recognized by a single base pair in the acceptor helix. Glutamine and serine specific tRNAs (ROGERS and SÖLL, 1988) are also distinguished in the stem region while the recognition of $tRNA^{Phe}$ (MCCLAIN and FOSS, 1988) is more complicated. With methionine- and valine-tRNA, an exchange of anticodons is sufficient to change the identity of the tRNA (SCHULMAN and PELKA, 1988). Further clarification of these issues can be expected since two research groups have now reported successful co-crystallizations of tRNAs with their cognate tRNA-synthetases (PERONA et al., 1988; RUFF et al., 1988).

Proteins do not only consist of the usual 20 amino acids. Many different types of post-translational modification reactions can provide certain residues in proteins with a wide variety of chemical moieties. It has been a long-standing puzzle of how selenocysteine might enter a protein (cf. MAELICKE, 1988). In the last few years, two research groups have independently determined the existence of opal stop-codons at the corresponding position in a gene where a selenocysteine appears in the protein sequence. LEINFELDER et al. (1988) recently showed that in *E. coli*, selenocysteine is co-translationally incorporated with the help of a special tRNA that recognizes this opal codon.

One might eventually envisage the development of an analogous strategy for incorporating other unusual amino acids into "engineered" proteins. A first small step in this direction was achieved by a Japanese research group (KOIDE et al., 1988). They reported the specific incorporation of aminohexanoic acid instead of methionine into a protein in *E. coli in vivo* and the purification of the unnatural protein.

SCHÖN et al. (1988) have made the surprising discovery that even a normal amino acid (glutamine) is incorporated into the proteins of chloroplasts in an unusual way: through the conversion of $Glu\text{-}tRNA^{Gln}$ by the action of an amidotransferase.

XI. Expression of Heterologous Proteins

How important is codon usage for obtaining good expression levels of foreign proteins in *Escherichia coli*? This controversial topic has now been taken up again by WILLIAMS et al. (1988), who have provided new data that are consistent with a stronger expression of a synthetic gene using only codons of highly expressed *E. coli* genes. BONEKAMP and JENSEN (1988) have also reported data relevant to this question, and SHARP et al. (1988) present useful tables of codon usage. Unfortunately, it is difficult to prove whether the observed effects on expression are really the result of codon usage or rather an effect of the primary sequence itself on, e.g., translation initiation or RNA secondary structure caused by the RNA sequence change. The issue of bacterial translation initiation was in turn investigated with a novel approach by DREYFUS (1988), while an interesting practical solution to this problem was proposed by OLINS et al. (1988).

A difficult problem in the industrial fermentation of *E. coli* can sometimes be to provide enough oxygen for supporting aerobic growth to high cell densities. KHOSLA and BAILEY (1988) presented a highly original solution to this problem. They cloned and functionally expressed a bacterial hemoglobin in *E. coli*. The bacte-

ria actually grew to higher cell densities with the heterologous oxygen-binding protein.

XII. DNA Sequencing

Several new developments have occurred in DNA sequencing and were recently summarized by HEINRICH and DOMDEY (1988). Through an ingenious simplification of sequencing methods, the genome sequence of *E. coli* was brought closer to realization. CHURCH and KEIFFER-HIGGINS (1988) developed this method, known as multiplex sequencing. It consists of simultaneously establishing several genomic libraries in different vectors, combining clones (each from a different vector), using a *common* Maxam-Gilbert reaction for the combined plasmids, and finally detecting all the *individual* sequences in the mixture through the repeated blotting with different vector-specific oligonucleotides. Another new method, genomic sequencing through the polymerase chain reaction (PCR), is described in Sect. XXVI of this chapter and in Chapter 10b (see also NAKAMAJE et al., 1988).

XIII. Pattern Formation

A large step towards the understanding of the molecular mechanisms that control the development of *Drosophila* was made last year. One of the first steps in the differentiation of the egg is the generation of anterior/posterior as well as dorsal/ventral polarities, for which the so-called maternal effect genes are responsible. The *bicoid* gene, which has been cloned and sequenced, is responsible for the establishment of the anterior/posterior axis (FRIGERIO et al., 1986; BERLETH et al., 1988).

It could directly be shown that the *bicoid* protein exists in the egg and early embryo as a morphogenic gradient, within the maximum concentration at the anterior end, while the protein is practically non-existent in the posterior third (DRIEVER and NÜSSLEIN-VOLLHARD, 1988a). The corresponding mRNA is synthesized in the maternal cells and transported to the anterior pole of the egg. The protein gradient results from localized synthesis at the anterior end, followed by diffusion of the product, and is maintained by continuous proteolytic degradation. If the protein gradient is altered through genetic manipulation, the development of the embryo is exactly correlated to the local concentration of the *bicoid* protein which has all the properties of a morphogen (DRIEVER and NÜSSLEIN-VOLLHARD, 1988b). The *bicoid* protein functions as a regulator of the segmentation genes, for example the *hunchback* gene (TAUTZ, 1988).

XIV. Gene Regulation

One of the most notable discoveries of last year was that the mechanism of gene regulation is much more strongly conserved across the species than had previously been thought. Numerous studies have shown that mammalian factors are functional in yeast, as are yeast factors in higher eukaryotic cells. It was determined in 1987 that the regulatory yeast protein GCN4 is identical to the mammalian transcription factor AP-1.

More detailed investigations have now shown that a DNA element from the SV40

enhancer, to which the AP-1 protein binds, is also active in yeast where it is the target not only of GCN4, but also of another, previously unknown transcriptional factor yAP-1 (y for yeast) (HARSHMAN et al., 1988). It appears that in yeast as well as in higher eukaryotes, there is a family of AP-1 proteins, which GCN4 is a member of.

The paradigm of yeast transcriptional factors, the GAL4 protein, can activate an artificial promoter in mammalian cells, when the GAL4 binding site is present upstream (KAKIDANI and PTASHNE, 1988; WEBSTER et al., 1988). Corresponding experiments were also successful with *Drosophila* (FISCHER et al., 1988) and with the promoter of a plant virus in tobacco cells (MA et al., 1988).

Not only enhancers and upstream activating sequence (UAS) binding proteins are conserved, but factors that bind to the TATA element are also exchangeable. A yeast factor can functionally replace the TFIID protein in an *in vitro* mammalian cell system (BURATOWSKI et al., 1988; CAVALLINI et al., 1988).

Progress has been made in characterizing the CAAT box factor which has previously been reported to be identical to the nuclear factor I. The reason for certain discrepancies between different research groups is now clear. There is a family of related proteins that are apparently encoded by one gene but with differentially spliced transcripts (SANTORO et al., 1988). In addition, there is evidence that proteins encoded by other genes have an overlapping binding specificity and biological activity (CHODOSH et al., 1988a). There is again a yeast connection. The yeast proteins HAP2 and HAP3, which are necessary for the activation of the CYC1 (iso-1-cytochrome c) gene, have the same DNA binding characteristics as a human CAAT box factor, and are also functionally interchangeable (CHODOSH et al., 1988b). In light of these results, it is perhaps not surprising that the human estrogen receptor is functional in yeast in a hormone dependent fashion and induces transcription when the appropriate DNA binding site is introduced, exactly as in mammalian cells (METZGER et al., 1988). It is thus not without reason that yeast has acquired the status of "honorary mammal".

The binding of a number of regulatory proteins to DNA is a property of so-called zinc fingers, structural elements that were originally found for the transcriptional factor TFIIIA and, as is now known, are apparently quite widespread (review EVANS and HOLLENBERG, 1988). A chain of 30 amino acids characteristically contains four cysteine moieties or two cysteine and two histidine residues at precisely defined positions in the protein sequence and coordinately bind a zinc ion. However, FRANKEL and PABO (1988) rightly caution that not everything that fits into this broadly defined scheme is a zinc finger.

In one example it could be convincingly shown that the zinc finger domain of a protein alone is capable of binding DNA, and that zinc is necessary for this (NAGAI et al., 1988). An exchange of cysteine into serine in the *Drosophila* segmentation gene "cripple", which contains a zinc finger, destroys the biological activity of this protein (REDEMANN et al., 1988). This is in agreement with the central role of cysteine in binding zinc, the basis of the zinc finger concept.

An additional level of regulation has been demonstrated by the finding that certain transcriptional activators must be phosphorylated before they bind to DNA and/or exert their effects. For example, the CREB protein, which activates many eukaryotic genes in a cAMP dependent way, dimerizes after phosphorylation and only then acquires DNA-binding properties and functions as a transcriptional activator (YAMAMOTO et al., 1988). For the factor that mediates the response to heat shock, the situation is different depending on the organism from which it is derived. The yeast factor binds without being phosphorylated but does not become active until after phosphorylation (SORGER and PELHAM, 1988). The protein from mammalian cells binds and activates only after phosphorylation (ZIMARINO and WU, 1987; KINGSTON et al., 1987; SORGER et al., 1987).

XV. Mechanism of Transcriptional Activation

It has been shown for GAL4 that the activating domain consists of a run of acidic amino acid residues, the crucial factor being the generation of a negatively charged, local surface. The actual amino acid sequence plays only a secondary role (GINIGER and PTASHNE, 1987). A similar situation holds for other yeast activators such as GCN4 (HOPE et al., 1988). With the discovery of the exchangeability of such factors between yeast and higher eukaryotes, the importance of these results has further increased. Such domains have also been found for the glucocorticoid receptor (EVANS, 1988). A protein from herpes simplex virus, VP16, which contains a highly acidic region, is one of the strongest activators known and can assume the function of GAL4 in yeast (SADOWSKI et al., 1988) when it is equipped with the corresponding GAL4-DNA binding domain. Physiologically it functions through non-covalent interaction with a cellular DNA binding protein (TRIEZENBURG et al., 1988).

The mechanism of activation by the acidic domain is still unclear. Conceivable explanations are interactions with RNA-polymerase itself or a TATA box binding protein resulting in the formation of a stable transcription pre-initiation complex. According to this theory, favored primarily by PTASHNE (1988), the DNA between the enhancer element and the TATA box is looped out.

XVI. Chromatin Structure and Gene Regulation

Another possibility for transcriptional activation by means of negatively charged protein domains is that the proteins cause a local change in the chromatin structure, perhaps an "open window", for the transcriptional machinery. HAN et al. (1988) have shown that interfering with histone biosynthesis leads to the activation of certain genes in yeast. This would be in agreement with the concept that the expression of these genes is normally repressed through their organization in chromatin.

A more detailed picture emerges from studies in which the *N*-terminal 30 amino acids of histone H4 in yeast were removed through manipulation of the corresponding gene (KAYNE et al., 1988). The cells divided somewhat slower but behaved normally as fas as the regulation of a number of different genes is concerned. One function was impaired, however, the repression of a certain locus by a so-called silencer. A silencer is the opposite of an enhancer, that is, it represses a gene in its vicinity through interaction with regulatory proteins. The new results show that the *N*-terminal section of the histone H4 is necessary for this repression.

Evidence that the chromatin structure is involved in gene regulation also stems from the analysis of certain mutants which showed impaired regulation of a number of different genes. Analysis of these mutants revailed that the defect was in the copy number, and therefore the gene dosage, of certain histone genes (CLARK-ADAMS et al., 1988). This phenomenon could be directly confirmed through the targeted deletion or duplication of individual histone gene copies (NORRIS et al., 1988).

XVII. Oncogenes

The *ras* genes were the first for which the principle of proto-oncogene to oncogene activation was recognized. The corresponding proteins are associated with the cell membrane, bind GTP and GDP, and take part in the signal transfer to the cell interior. Mutations that lead to amino acid changes, particularly at positions 12 and 61, turn the proto-oncogene into an oncogene. Even in small amounts this oncogene can transform mammalian cells, and it is found in about 30% of all human tumors. These amino acid changes stabilize the GTP complex of the protein, which is active in signal transduction, while the GDP complex is inactive in this respect. It is still unclear which system receives the signal from the ras protein.

A protein has now been discovered that interacts with the ras protein in mammalian cells and stimulates its GTPase activity (GAP for GTPase activating protein) (TRAHEY and MCCORMICK, 1987). The binding occurs via the effector region of the ras protein (CALES et al., 1988), so GAP might be the long-sought next link in signal transduction.

The ras product is the first oncogene protein whose crystal structure has been determined (DE VOS et al., 1988). In fact, it has been shown that the ominous position 12 contacts the β-phosphate group of the GDP molecule, and interacts with a glutamine residue at position 61. The region 10–14 is quite probably the catalytic center of GTPase activity.

Of all the oncogenes and proto-oncogenes, *c-fos* has probably received the most attention during 1988. It is among those genes that are most rapidly induced in response to a growth stimulus and which in turn activate genes that are responsible for cell division. It has been found that *Fos*, the protein product of *c-fos*, occurs in the cell in a complex with other proteins, especially with p39. This protein has been identified as *Jun*, that is the product of the proto-oncogene *c-jun* (RAUSCHER et al., 1988; CHIU et al., 1988; SASSONE-CORSI et al., 1988a). This discovery is very interesting because *Jun* is identical with the transcription factor AP-1. The complex of *Fos* and *Jun* binds more tightly to regulatory DNA elements than either of the two proteins alone (KOUZARIDES and ZIFF, 1988). This is in accordance with the role of the complex in gene regulation (CHIU et al., 1988). It could also be shown that the *Fos-Jun* complex binds to the *c-fos* promotor and represses *c-fos* transcription in a negative auto-regulation (SASSONE-CORSI et al., 1988b). This could be the reason that after stimulation by a mitogen a *c-fos* transcription peak appears only for a short time before quickly leveling off again.

LANDSCHULZ et al. (1988) believe that a series of leucine residues at intervals of seven amino acids are responsible for the interaction between *Fos* and *Jun*. This structure has been termed a "leucine zipper" because in α-helical regions these leucine residues lie next to one another and could intract in a zipper-like pattern. In fact, *in vitro* mutagenesis studies with *Fos* in which leucine residues have been replaced support this model, which is also invoked for the interactions of other oncogene products (KOUZARIDES and ZIFF, 1988).

XVIII. Anti-Oncogenes

It has been known for some time that tumors can result not only through the activation of certain genes, but also through inactivation. This mechanism is becoming increasingly important; already about 20 such tumors are known, among which the one most intensively studied is retinoblastoma. A direct relationship between oncogenes and anti-oncogenes has been established.

In the cell, the product of an adenovirus oncogene, the E1A protein, forms a tight

complex with a cellular protein. It could be shown that this protein is the product of the retinoblastoma gene, which has been cloned (WHYTE et al., 1988). Furthermore, it turned out that the *SV40-T* antigen also forms such a complex (DECAPRIO et al., 1988). It, therefore, seems that these oncogenes do not exert their transforming effect directly, but rather through titration of the retinoblastoma gene product.

XIX. Targeted Gene Manipulation in Mammals

The substitution of a chromosomal copy of a gene by a copy modified *in vitro* is a standard technique in yeast, but is still very difficult in mammalian cells. This is in part due to the fact that mammalian genomes are more than one-hundred times larger than the yeast genome. Secondly, illegitimate, that is non-homologous, recombinations is more frequent in higher eukaryotes than in yeast. The success of gene transfer experiments in mammalian cells is, therefore, dependent on efficient selection methods. MANSOUR et al. (1988) describe a well thought out strategy that is generally applicable. Embryonal stem cells are transformed with linear DNA fragments which contain a segment that is homologous to the target region containing the gene *X*, but that is interrupted in the center through the insertion of a neomycin-resistance gene (*neo*^{r}). Located at the very end of the fragment is the thymidine kinase gene (*tk*) of the herpes-simplex virus. Two homologous crossovers occurring in the homologous region to the left and right of *neo*^{r} lead to the incorporation of *neo*^{r} into the target region and simultaneous inactivation of the gene *X* (through the insertion of *neo*^{r}). This process is also known as "gene disruption". As a result the *tk* gene at the end should be lost. On the other hand, non-homologous illegitimate integration of the DNA into other regions of the genome would lead to incorporation of the *tk* gene since these reactions almost always start from the DNA ends. The legitimate recombinants should be X^-, *neo*^{r}, tk^-, while non-homologous integration should lead to the X^+, *neo*^{r}, tk^+ phenotype. Selection for *neo*^{r} can be achieved through the addition of the neomycin related antibiotic G418. Gancyclovir, a virostatic nucleoside analog, that can only be phosphorylated through the viral thymidine kinase, selects tk^- recombinants. MANSOUR et al. (1988) have found that, by selection for tk^-, homologous integration events are scored at a 2000-fold higher frequency and they show for two genes (*hprt* and *int-2*) that their strategy works. In the case of *hprt*, integration into the *hprt* locus occurs in almost all the cells that survive the selection. The frequency of homologous recombination in the *int-2* locus was about 20 times less frequent, i.e., 4 of 81 colonies contained the desired insertion.

XX. Transcription in Mitochondria of Trypanosomes

A very unusual observation was reported concerning transcription of the sub-unit III of cytochrome oxidase in *Trypanosoma brucei* (SHAW et al., 1988; FEAGIN et al., 1988; FEAGIN and STUART, 1988). In all eukaryotes this polypeptide is coded for by the mitochondrial DNA, and this is not different in the two protozoa *Leishmania tarentolae* and *Crithidia fasciculata*, which are closely related to *Trypanosoma brucei*. Practically identical mRNAs were found in all

three protozoa, but in *T. brucei* the appropriate gene is missing. "Missing" is perhaps not quite the correct term: there is a gene, in which thymidine is missing at almost all the positions where uridine is found in mRNA, but thymidine occurs occasionally where no uridine is present in the mRNA. The sequence of C, G, and A in the gene is precisely the same as in the mRNA. The authors were left with no conclusion other than that after transcription, uridine moieties are inserted and occasionally deleted – a process they call "RNA editing". That would mean that almost 50% of the sequence information is inserted after transcription, an event that is without precedent in molecular biology.

XXI. Cell Cycle

The product of the cell cycle gene *CDC2* from *Saccharomyces pombe* is necessary to bring the cell past two points in the cell cycle. Depending on the nutrient content of the medium, it can be phosphorylated and thereby acquires protein kinase activity. It has now been discovered that this protein is identical with a subunit of the maturation promoting factor (MPF) of *Xenopus*, which is responsible for the onset of mitosis. This factor consists of a 45000 (45K) molecular weight protein and a 34K subunit. GAUTIER et al. (1988) showed that this small subunit reacts with antibodies against the yeast CDC2 protein, and has protein kinase activity. They conclude that it is identical with the CDC2 protein. DUNPHY et al. (1988) came to the same conclusion using different experiments. They showed that a protein that binds to the CDC2 product in yeast also binds to the 34K protein in *Xenopus*, thereby preventing the induction of mitosis in interphase extracts from *Xenopus* eggs. MPF is identical with a histone H1-kinase (ARION et al., 1988) that is found in many eukaryotic cells. This histone could be a physiological substrate of MPF in these organisms.

Another cell cycle gene from *Saccharomyces pombe*, *CDC13*, has recently been sequenced. It has been deduced from genetic investigations that the product of this gene interacts with the *CDC2* product. When the amino acid sequence of the CDC13 protein was deduced from the DNA sequence, it turned out that this protein was highly homologous to the so-called cyclins (SOLOMON et al., 1988). Cyclins are a group of proteins that accumulate during interphase, are destroyed after mitosis is complete, and have a regulatory function in the cell cycle.

XXII. T-Lymphocyte Antigen Receptor

Although the existence of a second T-cell receptor, consisting of a γ- and a δ-chain, was proven last year, its function and cellular distribution remain unclear. GOODMAN and LEFRANÇOIS (1988) reported that this receptor is found in mouse T-cells obtained from the intestinal mucosa. BONNEVILLE et al. (1988) also came to the same conclusion. Most surface epithelia contain lymphocytes that have a local immuno-defensive function, and these cells appear to express the second T-cell receptor (JANEWAY et al., 1988).

XXIII. AIDS

The suspicion that the third HIV virus, called HTLV4 according to the old nomenclature, was the result of a laboratory contamination by a monkey virus, could be verified. KESTLER et al. (1988) investigated the DNA of the monkey virus SIV_{Mac} and found that isolates from different macaques were quite different, but that one of these isolates was identical with HTLV4. This isolate had been made available to KANKI whose laboratory characterized HTLV4 for epidemiological studies. In a reply ESSEX and KANKI (1988) themselves took the view that HTLV4 must be considered a laboratory contamination. As a consequence, the theory that the human AIDS virus was recently evolved from a virus of the African Green Monkey (AGM) no longer holds. A complete sequencing of SIV_{AGM} DNA (FUKASAWA et al., 1988) shows as many differences from HIV1 and HIV2 as the sequences of the two human viruses do from each other. Despite enormous world-wide research activity, the AIDS virus still holds many puzzles (LEVY, 1988).

XXIV. Scid Mice

Scid (severe combined immunodeficiency) is an autosomal recessive mutation that appeared spontaneously several years ago in a mouse strain (BOSMA et al., 1983) and that has attracted much attention ever since. Scid mice allow new insight into the function of the immune system, and they are also of increasing importance for basic and applied research.

Mice which are homozygous for the scid mutation have no functional B- and T-cells and therefore no immune system of their own. It had been known before that scid mice are unable to correctly rearrange their immunoglobulin gene segments and those of the T-lymphocyte antigen receptors. The underlying defect has now been investigated in more detail by characterizing the rearrangement products of the immunoglobulin genes in cell lines from scid mice (MALYNN et al., 1988), or by transfecting plasmids that contain the recombination signals from T-cell receptor gene segments (LIEBER et al., 1988). Recombination events did take place; they were aberrant, however, *V*-, *D*-, and *J*-gene segments were never fused together. This defect in recombinase activity can explain the scid phenotype, since development of the B- and T-lymphocytes appears to be perfectly normal otherwise (SCHULER et al., 1988).

Stem cells from the human immune system have been successfully transplanted into scid mice to give so-called scid-hu-mice (hu for human) (MCCUNE et al., 1988; MOSIER et al., 1988). Mature human T- and B-lymphocytes have indeed been obtained from hematopoietic fetal liver cells, fetal thymus, and fetal lymph node tissue. Due to their defect, scid mice cannot reject foreign cells, as would normally be the case. It is obvious that such a system is invaluable for investigating the immune system and also as a model system for the study of the AIDS virus (NAMIKAWA et al., 1988), which infects human T-lymphocytes. Even the production of human antibodies should at least in principle be possible.

XXV. Genetic Diseases

In 1987, the gene defective in Duchenne muscular dystrophy was elucidated. The gene product, a protein of molecular weight 400000 called "dystrophin", has since been studied intensively. A first report claiming that dystrophin is located at the border be-

tween the sarcoplasmic reticulum and the actin filaments could not be confirmed. Several research groups now reported that dystrophin is localized on the outer membrane (the sarcolemma) of heart and skeletal muscle fibers (ZUBRZYCKA-GAARN et al., 1988; ARAHATA et al., 1988; WATKINS et al., 1988; BONILLA et al., 1988). The secondary structure of the protein is similar to that of spectrin, a cytoskeletal protein from erythrocytes (KOENIG et al., 1988). It was, therefore, suggested that dystrophin plays a role in muscle similar to that of spectrin in erythrocytes – the link between the plasma membrane and the cytoskeleton of the myofibrils.

Molecular analyses have now also provided a reason why the clinical manifestations of Becker muscular dystrophy are less severe. This disease is also caused by deletions in the coding region of dystrophin. However, they do not change the reading frame and, therefore, a shortened protein is synthesized. In the Duchenne form, by contrast, there is a frame shift, so that a functional protein is completely missing (MONACO et al., 1988; FORREST et al., 1988; KOENIG et al., 1988).

Alzheimer's disease, a form of senility, is diagnosed from the appearance of characteristic intracellular neurofilaments and extracellular amyloid plaques in certain regions of the brain. A cDNA for protein A4, consisting of 42 amino acids, could be derived from amyloid plaques. This cDNA codes for a much larger precursor of the A4 protein (695 amino acids). Three independent research groups have now reported independently that a further cDNA, which codes for a protein 57 amino acids longer, has been isolated (PONTE et al., 1988; TANZI et al., 1988; KITAGUCHI et al., 1988). The corresponding mRNA is derived from the same gene and is formed by alternative splicing of a common precursor. Interestingly, the additional peptide belongs to the serine protease inhibitor family. This finding is quite remarkable because the pathological changes that are characteristically found in brains from Alzheimer's patients might be due to a defect in the proteolytic processing of the precursor protein (review: GLENNER, 1988). Since not all proteases are likely to react to the newly discovered inhibitor in the same way degradation of the precursor protein might well depend on the presence or absence of this inhibitor domain. Moreover, CARRELL (1988) pointed out that surface-bound serine protease inhibitors have a regulatory function in the differentiation of nerve cells (GLOOR et al., 1988).

The other pathological change in Alzheimer's patients are intracellular aggregates of neurofilaments. It has now been shown that they consist of a protein fragment called tau which is associated with microtubules (GOEDERT et al., 1988; WISCHIK et al., 1988 a, b). These structures could represent the final stage in the disintegration of the cytoskeleton.

DNA diagnostics has been successfully applied to a complex mental illness, manic depression. Similar results have been reported for schizophrenia. By correlating genetic markers with the appearance of schizophrenia in some Icelandic and British families, there were indications that a predisposition to this disease is caused by one gene on chromosome 5 (SHERRINGTON et al., 1988). Not unexpectedly, another research group found no correlation with this region for a large Swedish family (J. L. KENNEDY et al., 1988). This is in agreement with the fact that schizophrenia is a complex disease with many different factors contributing to its etiology. A similar situation has also been reported for manic-depression.

XXVI. Polymerase Chain Reaction

The so-called polymerase chain reaction (PCR) constitutes one of the most important new developments in molecular biology (MULLIS and FALOONA, 1987). This technique was developed in the laboratory of H. A. EHRLICH, at the Cetus Corp. in California. By this method, any desired DNA fragment can be amplified. First, two short DNA pieces (primers) are synthesized, one complementary to one strand on the left, the other to the other strand on the right side of the region to be amplified. Then the starting DNA is denatured and renatured in the presence of the primers, which bind to their complementary sequences. Starting with these primers, two new strands, complementary to the original ones, are synthesized using DNA polymerase. After denaturation and renaturation the same reaction is repeated, and the primers can bind to both the original as well as the newly-synthesized DNA.

With each cycle, the DNA region between the primers is duplicated. For example, after twenty cycles this region is present 2^{20} times, a million times the original amount. Recently, the *Escherichia coli* polymerase, which is inactivated with each cycle and, therefore, must be added again, was replaced by a polymerase from the bacterium *Thermus aquaticus* (SAIKI et al., 1988). These bacteria normally live in hot springs and have a stable DNA polymerase that survives the DNA denaturation step.

Using this method and starting with the total genomic DNA, one can investigate any gene directly. For example, one can sequence it without having to clone it first. Although the first publication was rather recent (SAIKI et al., 1985), there is already a flood of applications including the study of genetic lesions, diagnosis of infectious pathogens like the AIDS virus, oncogene activations, or the analysis of DNA polymorphisms in DNA diagnosis.

MILLS (1988) described a type of "PCR *in vivo*". Through the simultaneous expression of the replicase of the Qβ phage and a suitable mRNA substrate, he was able to show that the mRNA is amplified *in vivo*.

XXVII. DNA Diagnostics

The current status of DNA diagnostics has recently been summarized (LANDEGREN et al., 1988). Only two examples will be mentioned here. With the help of PCR, it is possible to type the DNA from a single hair (that is, from the cells that adhere to the shaft region), and to identify the previous owner of the hair (HIGUCHI et al., 1988).

Even the amount of DNA from a single cell, either a diploid body cell or a haploid sperm cell, is sufficient for characterization through PCR (LI et al., 1988). This technology will certainly open a new era in forensic medicine. DNA characterization using the "minisatellite DNA" technique developed by JEFFREYS (1988), which relies on "classical" hybridization (i.e., without amplification), has been used in England to identify and convict a rapist in court.

XXVIII. References

ALBER, T. et al. (1988). Science *239*, 631.
AGGARWAL, A. K. (1988). Science *242*, 899.
APPEL, J. D. et al. (1988). Proc. Natl. Acad. Sci. USA *85*, 7670.
ANFINSEN, C. B. (1973). Science *181*, 223.
ARAHATA, K. et al. (1988). Nature *333*, 861.
ARION, D. et al. (1988). Cell *55*, 371.
BERLETH, T. et al. (1988). EMBO J. *7*, 1749.

BETTER, M. et al. (1988). Science *240*, 1041.

BLEECKER, A. B. et al. (1988). Science *241*, 1086.

BOCHAREVA, E. S. et al. (1988). Nature *336*, 254.

BONEKAMP, F. and JENSEN, K. F. (1988). Nucleic Acids Res. *16*, 3013.

BONILLA, E. et al. (1988). Cell *54*, 447.

BONNEVILLE, M. et al. (1988). Nature *336*, 479.

BOSMA, G. C. et al. (1983). Nature *301*, 527.

BRANGE, J. et al. (1988). Nature *333*, 679.

BULLEID, N. J. and FREEDMAN, R. B. (1988). Nature *335*, 649.

BURATOWSKI, S. et al. (1988). Nature *334*, 37.

CALES, C. et al. (1988). Nature *332*, 548.

CARRELL R. W. (1988). Nature *331*, 478.

CAVALLINI, B. et al. (1988). Nature *334*, 77.

CHENG, Y.-F. and DOVICHI, N. J. (1988). Science *242*, 562.

CHIRICO, W. J. et al. (1988). Nature *332*, 805.

CHIU, R. et al. (1988). Cell *54*, 541.

CHODOSH, L. A. et al. (1988a). Cell *53*, 11.

CHODOSH, L. A. et al. (1988b). Cell *53*, 25.

CHRISTOU, P. et al. (1988). Plant Physiol. *87*, 671.

CHURCH, G. M. and KEIFFER-HIGGINS, S. (1988). Science *240*, 185.

CITOVSKY, V. et al. (1988). Science *240*, 501.

CLARK-ADAMS, C. D. et al. (1988). Genes Dev. *2*, 150.

CUSACK, S. et al. (1988). J. Mol. Biol. *202*, 903.

DAS, A. (1988). Proc. Natl. Acad. Sci. USA *85*, 2909.

DECAPRIO, J. A. et al. (1988). Cell *54*, 275.

DEIKMAN, J. and FISCHER, R. L. (1988). EMBO J. *7*, 3315.

DEISENHOFER, H. et al. (1988a). Nachr. Chem. Tech. Lab. *36*, 1189.

DEISENHOFER, H. et al. (1988b). Angew. Chem. *100*, 79.

DE LA PEÑA, A. et al. (1987). Nature *325*, 274.

DERBYSHIRE, V. et al. (1988). Science *240*, 199.

DESHAIES, R. J. et al. (1988). Nature *332*, 800.

DE VOS, A. M. et al. (1988). Science *239*, 888.

DREUSICKE, D. and SCHULZ, G. E. (1988). J. Mol. Biol. *203*, 1021.

DREYFUS, M. (1988). J. Mol. Biol. *204*, 79.

DRIEVER, W. and NÜSSLEIN-VOLHARD, C. (1988a). Cell *54*, 83.

DRIEVER, W. and NÜSSLEIN-VOLHARD, C. (1988b). Cell *54*, 95.

DUNCAN, A. R. and WINTER, G. (1988). Nature *332*, 738.

DUNPHY, W. G. et al. (1988). Cell *54*, 423.

DYSON, H. J. et al. (1988). J. Mol. Biol. *201*, 161, 201.

ECKERSKORN, C. et al. (1988). Eur. J. Biochem. *176*, 509.

ESSEX, M. and KANKI, P. (1988). Nature *331*, 621.

ESTELL, D. A. and FERSHT, A. (1987). Protein Eng. *1*, 441.

ETTLINGER, C. and LEHLE, L. (1988). Nature *331*, 176.

EVANS, R. M. (1988). Science *240*, 889.

EVANS, R. M. and HOLLENBERG, S. M. (1988). Cell *52*, 1.

FARBER, G. K. et al. (1988). Proc. Natl. Acad. Sci. USA *85*, 112.

FEAGIN, J. E. and STUART, K. (1988). Mol. Cell. Biol. *8*, 1259.

FEAGIN, J. E. et al. (1988). Cell *53*, 413.

FISCHER, G., BANG, H., and MECH, C. (1984). Biomed. Biochim. Acta *43*, 1101.

FISCHER, J. A. et al. (1988). Nature *332*, 853.

FISCHER, G., WITTMANN-LIEBOLD, B., LANG, K., KIEFHABER, T., and Schmid, F. X. (1989). Nature *337*, 476.

FORREST, S. M. et al. (1988). Genomics *2*, 109.

FORSTER, A. C. et al. (1988). Nature *334*, 265.

FRANKEL, A. D. and Pabo, C. O. (1988). Cell *53*, 675.

FRIGERIO, G. et al. (1986). Cell *47*, 735.

FUKASAWA, M. et al. (1988). Nature *333*, 457.

GAUTIER, J. et al. (1988). Cell *54*, 433.

GINIGER, E. and PTASHNE, M. (1987). Nature *330*, 670.

GLENNER, G. G. (1988). Cell *52*, 307.

GLOOR, S. et al. (1986). Cell *47*, 687.

GOEDERT, M. et al. (1988). Proc. Natl. Acad. Sci. USA *85*, 4051.

GOLOUBINOFF, P., GATENBY, A. A., and LORIMER, G. H. (1989). Nature *337*, 44.

GÓMEZ, J. et al. (1988). Nature *334*, 262.

GOODMAN, T. and LEFRANÇOIS, L. (1988). Nature *333*, 855.

GORDON, M. J. et al. (1988). Science *242*, 224.

GRAY, D. J. (1988). Trends Biotechnol. *6*, 233.

HAN, B. M. et al. (1988). EMBO J. *7*, 2221.

HARSHMAN, K. D. et al. (1988). Cell *53*, 321.

HASELOFF, J. and GERLACH, W. L. (1988). Nature *334*, 585.

HEINRICH, P. and DOMDEY, H. (1988). Nachr. Chem. Tech. Lab. *36*, 1196.

HEMMINGSEN, S. M. et al. (1988). Nature *333*, 330.

HERMANS, J. et al. (1988). J. Am. Chem. Soc. *110*, 5982.

HIDDINGA, H. J. et al. (1988). Science *241*, 451.

HIGUCHI, R. et al. (1988). Nature *332*, 543.

HOPE, I. A. et al. (1988). Nature *333*, 635.

HORWITZ, A. H. et al. (1988). Proc. Natl. Acad. Sci. USA *85*, 8678.

HOU, Y.-M. and SCHIMMEL, P. (1988). Nature *333*, 140.

HOWARD, A. E. et al. (1988). J. Am. Chem. Soc. *110*, 6984.

HUSTON, J. S. et al. (1988). Proc. Natl. Acad. Sci. USA *85*, 5879.

HWANG, J. K. and WARSHEL, A. (1988). Nature *334*, 270.

JACOBSON, J. W. et al. (1988). Science *242*, 423.

JAENICKE, R. (1987). Prog. Biophys. Mol. Biol. *49*, 117.

JANEWAY, C. A. et al. (1988). Immunol. Today *6*, 50.

JEFFREYS, A. (1988). Nachr. Chem. Tech. Lab. *36*, 552.

JORDAN, S. R. and PABO, C. O. (1988). Science *242*, 893.

JOSHUA-TOR, L. et al. (1988). Nature *334*, 82.

KAKIDANI, H. and PTASHNE, M. (1988). Cell *52*, 161.

KASSENBROCK, C. K. et al. (1988). Nature *333*, 90.

KAYNE, P. S. et al. (1988). Cell *55*, 27.

KELLIS, J. T. et al. (1988). Nature *333*, 784.

KENNEDY, J. L. et al. (1988). Nature *336*, 167.

KENNEDY, T. E. et al. (1988). Nature *336*, 499.

KESTLER III, H. W. et al. (1988). Nature *331*, 619.

KHOSLA, C. and BAILEY, J. E. (1988). Nature *331*, 633.

KIM, P. S. (1988). Protein Eng. *2*, 249.

KINGSTON, R. E. et al. (1987). Mol. Cell. Biol. *7*, 1530.

KITAGUCHI, N. et al. (1988). Nature *331*, 530.

KOENIG, M. et al. (1988). Cell *53*, 219.

KOIDE, H. et al. (1988). Proc. Natl. Acad. Sci. USA *85*, 6237.

KOUZARIDES, T. and ZIFF, E. (1988). Nature *336*, 646.

KOZUTSUMI, Y. et al. (1988). Nature *332*, 462.

KRAUT, J. (1988). Science *242*, 533.

LAMINET, A. and PLÜCKTHUN, A. (1989). EMBO J. *8*.

LANDEGREN, U. et al. (1988). Science *242*, 229.

LANDSCHULZ, W. H. et al. (1988). Science *240*, 1759.

LANG, K. and SCHMID, F. X. (1988). Nature *331*, 453.

LANG, K. et al. (1987). Nature *329*, 268.

LAST, R. L. and FINK, G. R. (1988). Science *240*, 305.

LAWSON, C. L. and SIGLER, P. B. (1988). Nature *333*, 869.

LEAR, J. D. et al. (1988). Science *240*, 1177.

LEE, K. Y. et al. (1988). EMBO J. *7*, 1241.

LEINFELDER, W. et al. (1988). Nature *331*, 723.

LEVY. J. A. (1988). Nature *333*, 519.

LI, H. et al. (1988). Nature *335*, 414.

LICHTENSTEIN, C. (1988). Nature *333*, 801.

LIDDINGTON, R. et al. (1988). Nature *331*, 725.

LIEBER, M. R. et al. (1988). Cell *55*, 7.

LINSE, S. et al. (1988). Nature *335*, 651.

MA, J. et al. (1988). Nature *334*, 631.

MALYNN, B. A. et al. (1988). Cell *54*, 453.

MANSOUR, S. L. et al. (1988). Nature *336*, 348.

MAPLE, J. R. et al. (1988). Proc. Natl. Acad. Sci. USA *85*, 5350.

MARGOSSIAN, L. J. et al. (1988). Proc. Natl. Acad. Sci. USA *85*, 8012.

MATSUMURA, M. et al. (1988). Nature *334*, 406 (1988).

MATTHEWS, B. W. (1988). Nature *335*, 294.

MCCABE, D. E. et al. (1988). Bio/Technology *6*, 923.

MCCLAIN, W. H. and FOSS, K. (1988). J. Mol. Biol. *202*, 697.

MCCUNE, J. M. et al. (1988). Science *241*, 1632.

MCPHERSON, A. and SHLICHTA, P. (1988). Science *239*, 385.

METZGER, D. et al. (1988). Nature *334*, 31.

MEYER, D. I. (1988). Trends Biochem. Sci. *13*, 471.

MILLER, M. et al. (1988). Nature *334*, 85.

MILLS (1988). J. Mol. Biol. *200*, 489.

MONACO, A. P. et al. (1988). Genomics *2*, 90.

MOSIER, D. E. et al. (1988). Nature *335*, 256.

MULLIS, K. B. and FALOONA, F. A. (1987). Methods Enzymol. *155*, 335.

NAGAI, K. et al. (1988). Nature *332*, 284.

NAKAMAJE et al. (1988). Nucleic Acids Res. *16*, 9947.

NAMIKAWA, R. et al. (1988). Science *242*, 1684.

NICHOLSON, H. et al. (1988). Nature *336*, 651.

NORRIS, D. et al. (1988). Science *242*, 759.

NOVOTNY, J. et al. (1988). Proteins *4*, 19.

OAS, T. G. and KIM, P. S. (1988). Nature *336*, 42.

OLINS, P. O., DEVINE, C. S., RUNGWALA, S. H., and KAVKA, K. S. (1988). Gene *73*, 227.

OTWINOWSKI, Z. et al. (1988). Nature *335*, 321.

PARK, S. et al. (1988). Science *239*, 1033.

PERONA, J. J. et al. (1988). J. Mol. Biol. *202*, 121.

PRESTA, L. G. and ROSE, G. D. (1988). Science *240*, 1632.

PONTE, P. et al. (1988). Nature *331*, 525.

PTASHNE, M. (1988). Nature *335*, 683.

RAUSCHER, F. J. et al. (1988). Science *240*, 1010.

REDEMANN, N. et al. (1988). Nature *332*, 90.

REGAN, L. and DEGRADO, F. (1988). Science *241*, 976.

RHODES, C. A. et al. (1988). Science *240*, 145, 204.

RICHARDSON, J. S. and RICHARDSON, D. C. (1988). Science *240*, 1648.

RIECHMANN, L. et al. (1988). Nature *332*, 323.

ROBERTS, L. (1988). Science *241*, 1290.

RODER, H. et al. (1988). Nature *335*, 700.

ROGERS, M. J. and SÖLL, D. (1988). Proc. Natl. Acad. Sci. USA *85*, 6627.

ROOMAN, M. J. and WODAK, S. J. (1988). Nature *335*.

RUFF, M. et al. (1988). J. Mol. Biol. *201*, 235.

SADOWSKI, I. et al. (1988). Nature *335*, 563.

SAIKI, R. K. et al. (1985). Science *230*, 1350.

SAIKI, R. K. et al. (1988). Science *239*, 487.

SALI, D. et al. (1988). Nature *335*, 740.

SANTORO, C. et al. (1988). Nature *334*, 218.

SASSONE-CORSI, P. et al. (1988a). Cell *54*, 553.

SASSONE-CORSI, P. et al. (1988b). Nature *334*, 314.

SCHLESSINGER, J. (1988). Trends Biochem. Sci. *13*, 443.

SCHLOSS, J. V. et al. (1988). Nature *331*, 360.

SCHÖN, A. et al. (1988). Nature *331*, 187.

SCHULER, W. et al. (1988). EMBO J. *7*, 2019.

SCHULMAN, L. H. and PELKA, H. (1988). Science *242*, 765.

SCHULTZ, P. G. (1988). Science *240*, 426.

SCHWARZ, TH. et al. (1988). Nature *331*, 137.

SHARP, P. M. et al. (1988). Nucleic Acids Res. *16*, 8207.

SHAW, J. M. et al. (1988). Cell *53*, 401.

SHEEHY, R. E. et al. (1988). Proc. Natl. Acad. Sci. USA *85*, 8805.

SHERRINGTON, R. et al. (1988). Nature *336*, 164.

SHIEK, B. H., STAMNES, M. A., SEAVELLO, S., HARRIS, G. L., and ZUKER, C. S. (1989). Nature *337*.

SKARZYNSKI, T. and WONACOTT, A. J. (1988). J. Mol. Biol. *203*, 1097.

SKERRA, A. and PLÜCKTHUN, A. (1988). Science *240*, 1038.

SMITH, C. J. S. et al. (1988). Nature *334*, 724.

SOLOMON, M. et al. (1988). Cell *54*, 738.

SORGER, P. K. and PELHAM, H. R. B. (1988). Cell *54*, 855.

SORGER, P. K. et al. (1987). Nature *329*, 81.

SPRANG, S. R. et al. (1988). Nature *336*, 215.

SUCK, D. et al. (1988). Nature *332*, 464.

TAKAHASHI, N., HAYANO, T., and SUZUKI, M. (1989). Nature *337*, 473.

TANZI, R. E. et al. (1988). Nature *331*, 528.

TAUTZ, D. (1988). Nature *332*, 281.

TIMMERMAN, B. et al. (1988). J. Mol. Biol. *203*, 373.

TOONG, Y. C. et al. (1988). Nature *333*, 170.

TRAHEY, M. and MCCORMICK, F. (1987). Science *238*, 542.

TRIEZENBURG, S. J. et al. (1988). Genes Dev. *2*, 718.

USAMI, S. et al. (1988). Proc. Natl. Acad. Sci. USA *85*, 3748.

UTGAONKAR, J. B. and BALDWIN, R. L. (1988). Nature *335*, 694.

VAN DER KROL, A. R. et al. (1988a). Nature *333*, 866.

VAN DER KROL, A. R. et al. (1988b). Gene *72*, 45.

VERHOEYEN, M. et al. (1988). Science *239*, 1534.

WALBOT, V. and BRUENING, G. (1986). Nature *334*, 196.

WALLACE, J. C. et al. (1988). Science *240*, 662.

WARSHEL, A. et al. (1988). J. Mol. Biol. *201*, 139.

WATKINS, S. C. et al. (1988). Nature *333*, 863.

WEBSTER, N. et al. (1988). Cell *52*, 169.

WETZEL, R. (1988). Protein Eng. *2*, 169.

WETZEL, R. et al. (1988). Proc. Natl. Acad. Sci. USA *85*, 401.

WHYTE, P. et al. (1988). Nature *334*, 124.

WIDHOLM, J. M. (1988). Trends Biotechnol. *6*, 265.

WILLIAMS, D. P. et al. (1988). Nucleic Acids Res. *16*, 10453.

WISCHIK, C. M. et al. (1988a). Proc. Natl. Acad. Sci. USA *85*, 4506.

WISCHIK, C. M. et al. (1988b). Proc. Natl. Acad. Sci. USA *85*, 4884.

WOLBERGER, C. et al. (1988). Nature *335*, 789.

WÜTHRICH, K. (1989a). Science *243*, 45.

WÜTHRICH, K. (1989b). Acc. Chem. Res. *22*, 36.

YAMAMOTO, K. K. et al. (1988). Nature *334*, 494.

YANG, D. S. C. et al. (1988). Nature *333*, 232.

YUN-YU, S. et al. (1988). J. Mol. Biol. *200*, 571.

ZHENG, C. et al. (1988). Nature *334*, 726.

ZIMARINO, V. and WU, C. (1987). Nature *327*, 727.

ZUBRZYCKA-GAARN, E. E. et al. (1988). Nature *333*, 466.