Biochemistry

© Copyright 1991 by the American Chemical Society Volume 30, Number 13 April 2, 1991

Perspectives in Biochemistry

Protein Folding: Local Structures, Domains, Subunits, and Assemblies[†,‡]

Rainer Jaenicke

Institut für Biophysik und Physikalische Biochemie, Universität Regensburg, D-8400 Regensburg, FRG

Received November 7, 1990; Revised Manuscript Received February 13, 1991

Old Principles versus New Prospects

Globular proteins show the intrinsic property of acquiring their spatial structure in an autonomous way, based on their amino acid sequence and their aqueous or nonaqueous environment. Folding in vivo may be assumed to occur spontaneously as a cotranslational process (Bergman & Kuehl, 1979; Wetlaufer & Ristow, 1973). To generate the native state, the polypeptide chain generally requires neither extrinsic factors nor the input of energy (Anson, 1945; Anfinsen, 1973). However, recent findings seem to reinforce Anfinsen's early idea that rate-limiting steps in protein folding might be catalyzed by "shuffling enzymes". In fact, convincing evidence has accumulated which clearly indicates that the three most important late events in the self-organization of proteins, i.e., the formation of disulfide bridges, proline isomerization, and association, are catalyzed or directed by specific helper proteins compartmentalized either in the cytosol or in organelles: the respective enzymes are protein disulfide isomerase (PDI),[1] peptidyl-prolyl cis-trans isomerase (PPI), and "molecular chaperones". Obviously, they not only promote folding and disulfide exchange but also prevent aggregation, thus helping to guide the nascent protein to its final destination within or without the cell (Pelham, 1989; Fischer & Schmid, 1990; Ellis, 1990; Schmid, 1991).

The final product of protein folding, i.e., the native molecule in its functional state, may be determined by kinetic and thermodynamic constraints: *kinetic*, because of the vectorial character of protein biosynthesis from the N- to the C-terminal end of the polypeptide chain, and *thermodynamic*, because energy minimization in terms of the subtle balance of stabilizing and destabilizing weak interactions is the driving force of structure formation. Bringing both aspects together, "hierarchical condensation" has been proposed to generate the native three-dimensional structure (Gō, 1984; Jaenicke, 1987, 1988a, 1990). This means that short-range next-neighbor and short-range through-space interactions within the growing or reconstituting chain lead to the formation of elements of secondary structure that are gradually combined and reshuffled to form subdomains, domains, and subunit assemblies. Whether in this hierarchy the final state occupies a local or the global minimum of potential energy is still controversial (Anfinsen & Scheraga, 1975; Wetlaufer, 1984; Finkelstein & Ptitsyn, 1987; Ptitsyn et al., 1989). What is settled is that (in the case of sufficiently large molecules) the nascent polypeptide chain acquires its native three-dimensional structure by the modular assembly of (micro-) domains. In defining the final state, posttranslational processing such as specific proteolysis, glycosylation, cross-linking by disulfide bridges, and attachment of cofactors and ligands may also be of crucial importance.

The mechanism of in vivo folding is unknown. One way to unravel individual steps along the folding path has been the analysis of unfolding/refolding reactions in vitro. From the practical point of view, such reconstitution experiments have two aims: first, the "nativation" of inactive deposits ("inclusion bodies") of recombinant proteins (Rudolph, 1990), and second, the elucidation of the code of protein folding in connection with reaction-specific protein design. As far as true *prediction* is concerned, this "second half of the genetic code", assumed to govern the spontaneous transition from the 1D information encoded in the primary structure of a protein to its 3D functional state, is still far from being solved. If it were known, structures could be inferred from given DNA sequences; consequently, engineered changes in protein structure could be designed by recombinant DNA techniques, and hypothetical functions for given amino acid sequences could be specified (Jaenicke, 1988a).

The change from esoteric intellectual gambling into an area

[†] Work performed in the author's laboratory was supported by grants from the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie, and the BAP program of the European Commission.

[‡] Dedicated to Professor Christian B. Anfinsen on the occasion of his 75th birthday.

[1] Abbreviations: 1D, 2D, and 3D, one-, two-, and three-dimensional; N, I, U, A, and IB, native state, intermediate state, unfolded state, aggregates, and inclusion bodies, respectively; CD, circular dichroism; H-D exchange, hydrogen-deuterium exchange; NMR, nuclear magnetic resonance; BPTI, basic pancreatic trypsin inhibitor; ER, endoplasmic reticulum; LDH, lactate dehydrogenase; PDI, protein disulfide isomerase; PPI, peptidyl-prolyl cis-trans isomerase; Rubisco, ribulose bisphosphate carboxylase; tPA, tissue plasminogen activator; RNase, ribonuclease.

0006-2960/91/0430-3147$02.50/0 © 1991 American Chemical Society

of unpredicted industrial importance has initiated a most dramatic expansion of the field (Fasman, 1989). When and whether the time is approaching when new (and even useful) proteins will be (de novo) designed, synthesized, and technologically applied is a question of enthusiasm and belief.

Thermodynamics versus Kinetics

In developing a theory of the structure of biological systems, three central issues need consideration: (i) How do the intrinsic physical and chemical properties of the building blocks and their mutual interactions combine to determine stability and flexibility, (ii) what are the elementary processes on the kinetic pathway of structure formation, and (iii) what are the energetics underlying the transition from disorder to order. The unifying concept that governs the correlation of thermodynamics and kinetics in the self-organization of proteins is the assumption that evolution has been geared at optimizing the *structure-function relationship* rather than *stability* in a given environment. As pointed out by Wetlaufer (1984), proteins in general have multiple functions, the first being *folding* and the last *turnover*, i.e., programmed degradation. This argument in itself does not solve the controversy of proteins being in their global or in a local minimum of potential energy. However, it provides a plausible explanation why globular proteins in general only show marginal free energies of stabilization.

Stability versus Flexibility. The free energy of stabilization of globular proteins in aqueous solution is found to be marginal: values commonly determined for globular proteins are of the order of 50 ± 15 kJ/mol, i.e., less than $^1/_{10}kT$ per residue, independent of the mode of denaturation (Privalov, 1979; Finney, 1982; Pace, 1990a,b; Dill, 1990). On balance, the stability is based on the equivalent of a few hydrogen bonds, ion pairs, or hydrophobic patches, although numerous stabilizing and destabilizing intermolecular interactions assist in maintaining the native 3D structure. To give an example: for RNase T_1, where the folded conformation is stabilized by two disulfide bridges, apart from 87 intramolecular hydrogen bonds and prominent hydrophobic interactions involving 85% of the nonpolar residues, ΔG_{stab} is no more than 24 kJ/mol (pH 7.0, 25 °C) (Pace, 1990). Thus, ΔG_{stab} represents a small difference of big numbers.

The intricate balance of stabilizing and destabilizing forces explains why the ability of a given amino acid sequence to form a stable 3D structure is by no means trivial: of the astronomically high number of possible sequences, only a small selection seems to be able to form a compact stable structure. Among those, obviously, nature has selected biologically active conformations that are only marginally more stable than the unfolded states. There are exceptions both in synthetic peptides and in natural proteins which prove that evolution could have generated more stable proteins if they were advantageous (De Grado, 1988; Richardson & Richardson, 1988; Sauer et al., 1990).

Estimates of the contributions of the different types of molecular interactions to the free energy of protein stabilization show that protein self-organization at physiological temperature is determined mainly by the increase in entropy caused by water release from nonpolar and polar residues, as well as van der Waals interactions (Privalov & Gill, 1989). Unfavorable free energy terms are derived from the loss of rotational and translational degrees of freedom that invariably accompanies intra- and intermolecular interactions (Sturtevant, 1977). Hydrogen bonds are considered less important since they replace similar interactions with solvent molecules in the free unfolded chain; their main role is to confer structural specificity. Similarly, Coulomb interactions (ion pairs) cannot be the dominant contributor to protein stability (Kauzmann, 1959; Dill, 1990). Most charged groups are exposed to the aqueous solvent; on the average, only one ion pair per 150 amino acid residues of a globular protein is buried within the interior core (Barlow & Thornton, 1983). Thus, only surface ion pairs are expected to be involved in stabilization, in agreement with experimental findings from site-directed mutagenesis and X-ray analysis (Fersht, 1972; Perutz & Raidt, 1975; Hollecker & Creighton, 1982; Sundaralingam et al., 1987; Pace & Grimsley, 1988; Anderson et al., 1990). Even for these groups, salt effects clearly prove that a significant fraction of the electrostatic free energy arises from the entropy of proton and water release rather than charge energetics (Stigter & Dill, 1990). As mentioned, hydrophobic interactions are governed by the same energetic principle in that water release from interacting hydrophobic surfaces is predominant. The reason is that "hydrophobic energy" is gained by the reduction of nonpolar surface in contact with water. To a first approximation, the free energy of stabilization of a folded protein is proportional to the water-accessible surface area (Kauzmann, 1959; Richards, 1977). In refining the original models of Kauzmann and Richards, Eisenberg and McLachlan (1986) calculated the solvation energy from the atomic coordinates of each single amino acid residue, weighting the polar or nonpolar character of each atom rather than whole residues. In deriving the free energy of transfer, the contribution of different classes of atoms is estimated as the product of their solvent accessibility and specific solvation parameters derived from experimental transfer data. Examination of the increments of the various side chains involved in the overall solvation energy of folding proved the dominant contribution to the stability of the folded state to come from nonpolar groups, including the carbon atoms of Lys, Arg, Asp, and Glu. A number of observations have been taken to support this view: (i) solvent denaturation using amphipathic or nonpolar denaturants (Dill, 1990), (ii) the large positive change in partial heat capacity (ΔC_p), which may be ascribed to clathrate formation around nonpolar surface areas, (iii) the close similarity of the parabolic temperature profile of both protein stability and solubility of nonpolar substances in water (Gill et al., 1976; Privalov & Gill, 1988, 1989; Franks & Hatley, 1990), (iv) site-directed mutations correlating protein stability with the oil-in-water partitioning of the amino acids involved (Yutani et al., 1987; Matsumura et al., 1988; Kellis et al., 1989). On the other hand, there are indications that contradict hydrophobic interactions being the primary stabilizing factor; to give an example, myoglobin is characterized by high hydrophobicity but low stability, especially at low temperature (Griko et al., 1988b).

Recent calorimetric studies have clearly shown that there are distinct differences between the temperature dependence of the enthalpy and entropy of protein denaturation on one hand and the respective thermodynamic data for the transfer of nonpolar substances to water on the other (Privalov & Gill, 1988). In both cases, the enthalpy and entropy functions increase with temperature up to a limiting value beyond 120 °C where "hydrophobic hydration" vanishes (Baldwin, 1986; Privalov, 1988; Jaenicke, 1991; Dill et al., 1990). The principal difference is that the transfer entropy at this temperature is zero, while the entropy of denaturation is large and positive. A model explaining this difference has been previously proposed in connection with the compressibility of proteins which clearly indicates that the protein interior resembles a crystallike solid phase rather than a nonpolar liquid (Hvidt, 1979; Jaenicke, 1981; Kundrot & Richards, 1987).

Conventionally, hydrophobic interactions have been interpreted in terms of "entropic bonds" (Kauzmann, 1959; Jencks, 1969). With increasing temperature, the ordering effect on water molecules at nonpolar surfaces is accompanied by an increasingly positive transfer enthalpy and an increase in ΔC_p so that the hydrophobic "phase separation" is favored to a still higher extent. Only at exceedingly high temperatures ΔC_p converges; the corresponding temperature range (120–140 °C) coincides with the temperature at which the transfer entropy of small nonpolar molecules vanishes, i.e., at which water becomes an ordinary solvent. It is interesting to note that both the upper temperature limit of life in the biosphere and the denaturation temperature of hyperthermophilic proteins come close to the given temperature range (Wrba et al., 1990b). At their optimum temperatures these proteins maintain the dynamical characteristics of their mesophilic counterparts at their respective physiological temperature, e.g., by filling cavities or by additional local tertiary interactions (Jaenicke & Zâvodszky, 1990).

The large difference in partial heat capacity between the folded and denatured states leads to a significant temperature dependence of both the enthalpy and entropy of unfolding. Due to enthalpy–entropy compensation, the free energy change is less dramatic. It exhibits an optimum curve indicating that unfolding may occur at high *and* low temperatures. The corresponding ΔH and ΔS values have opposite signs whereas ΔC_p is the same. Thus, exothermic "cold denaturation" may be considered a consequence of the more favorable solvation of nonpolar surfaces at low temperature. Taking staphylococcal nuclease and myoglobin as examples, Griko et al. (1988a,b) were able to show that decreasing temperature leads to a release of heat and a decrease in entropy, in accordance with the above concept. Spectral data clearly demonstrate that denaturation at high and low temperatures exhibits similar characteristics (Griko et al., 1988a,b; Chen & Schellman, 1989; Franks & Hatley, 1990).

It is obvious that there are variables other than temperature, especially cosolvents such as polyols, urea, and guanidinium chloride, that are equally important in stabilizing and destabilizing proteins. Their action may be the result of direct binding of the additive or of alterations of the physical properties of water. The effects may be quantitatively described by two ΔG increments, (i) the energy required to form a cavity in water and (ii) the interactions between the solute and water in filling the cavity (Arakawa & Timasheff, 1984, 1990a,b; Timasheff & Arakawa, 1989). Additives excluded from the hydration sphere of a protein lead to decreased solubility and increased stability of the folded state. Obviously, the effect can be overcome if the additive interacts with the protein surface. Thus, glycerol stabilizes the folded state due to preferential interactions with polar groups, while urea and guanidinium chloride are strong denaturants due to their solubilizing effect on nonpolar residues or surfaces.

The driving force determining the self-organization of proteins is the minimization of the free energy of stabilization:

$$\Delta G_{\text{stab}} = \Delta H_{\text{b}} - T\Delta S_{\text{solv}} - T\Delta S_{\text{chain}} \qquad (1)$$

In order to accomplish a stable state, the sum of all intermolecular interactions (ΔH_{b}) and the entropy of solvation (ΔS_{solv}) have to compensate the decrease in chain entropy (ΔS_{chain}). As far as folding and/or association involve hydrophobic and electrostatic interactions, water release will contribute significantly to the increase in solvent entropy, thus increasing the net free energy of stabilization. Due to the charge distribution on the protein surface, maximum stability is observed near the isoelectric point.

As indicated by the comparison of homologous proteins of organisms from extreme biotopes, very minor changes in the amino acid sequence may have dramatic effects on protein stability: in order to increase the stability to the extremes, e.g., of thermophilic conditions, a few additional "weak interactions" are sufficient. In all cases presently investigated in detail, no significant differences in the overall structure of the wild-type protein and its extremophilic counterpart were detectable; the alterations in ΔG_{stab} have their correspondence in minute structural alterations (Jaenicke, 1981; Jaenicke & Zâvodszky, 1990). In this context, it has been shown that folding requires highly specified environmental conditions, as active mesophilic or halophilic enzymes cannot be expressed in thermophilic or nonhalophilic hosts (in contrast to thermophilic proteins, which have been successfully cloned in mesophiles).

In spite of intensive work on both the structure and the energetics of proteins from extremophilic organisms, so far no clear-cut strategies of "molecular adaptation" to extremes of temperature, hydrostatic pressure, pH, and salt concentration in thermophiles, barophiles, acidophiles, alkalophiles, and halophiles have been uncovered. Similarly, efforts to generate "extremophilic" behavior by genetic engineering techniques have been of only limited success. With respect to flexibility, recent advances connected with thermophilic and halophilic proteins revealed that molecular adaptation tends to maintain "corresponding states" with respect to the structural and functional properties (Jaenicke & Zâvodszky, 1990). For example, NAD-dependent dehydrogenases from mesophilic, thermophilic, and halophilic sources show similar enzymological properties under their respective in vivo conditions (Hecht et al., 1989; Wrba et al., 1990a,b).

As mentioned, in many cases biological function requires flexibility of functional groups, chain segments, or domains. In contrast, processes such as energy or electron transfer depend on conformational rigidity to guarantee high efficiency (Huber, 1988). The significance of chain flexibility for protein folding and protein association is clearly illustrated by (i) oxygen binding to myoglobin or hemoglobin, where local oscillations and movements of chain segments of up to 10 Å are involved in the "channeling" of the ligand, (ii) binding of substrates or antigens, which may be accompanied by large domain motions or hinge bending, and (iii) the decrease in static disorder upon conjugation of proteins, e.g., trapping of RNA by tobacco mosaic virus protein (Bloomer & Butler, 1986; Ohada, 1986).

In connection with the intrinsic structural flexibility of proteins and the maintenance of corresponding states, the question whether a protein in its native state occupies a local or the global minimum of potential energy has been discussed for more than a decade (Anfinsen & Scheraga, 1975; Wetlaufer, 1984). The vectorial character of (cotranslational) protein folding suggests that the nascent polypeptide chain ends up in the "kinetically accessible minimum of free energy". This minimum is not necessarily the global one belonging to a hypothetical most stable state. If the two alternative states were (meta-) stable, their corresponding conformers should differ significantly; otherwise they should be interconvertible within a reasonable time. The fact that renaturation from different denaturants, including renaturation from the fully denatured state, as well as folding in the presence and in the absence of "helper proteins" (see below) yields the native protein shows that the vectorial character of cotranslational folding obviously does not determine the final state. The whole ensemble of structural intermediates from the various unfolded states to late nativelike conformers is rapidly equilibrating but separated from the native state by a relatively high energy barrier. This leads to just one activation energy for the various renaturation reactions. Since there is obviously one common final folding step, and since the energy of stabilization is only marginal, it seems reasonable to assume that the native protein occupies the global energy minimum. Incomplete renaturation due to wrong domain pairing, aggregation, chemical modification, etc., which was previously taken to indicate that there are energetically more favorable states, can be considered as kinetic artifacts. In most cases they can be avoided by carefully selecting appropriate folding conditions; in vivo, specific enzymes or polypeptide chain binding proteins are assumed to minimize side reactions (see below).

Pathways versus Puzzle. As shown by in vitro renaturation experiments, protein folding follows a sequential pathway where short stretches of secondary structure combine to compact local supersecondary structures, which subsequently either merge to nativelike intermediates or collapse to form the "molten globule" as a metastable set of states (Kuwajima, 1989). Characteristics of the molten globule state as a hypothetical common intermediate on the overall folding pathway may be illustrated by using α-lactalbumin as the best-known example (Dolgikh et al., 1985; Kuwajima, 1989). (1) According to its spectral properties, the molecule has regained a nativelike secondary structure while its aromatic residues show high flexibility. (2) A slight "expansion" of the molecule leads to the exposure of hydrophobic surfaces, which promote the formation of aggregates. (3) Intramolecular structural fluctuations are slowed down, however, thermal transitions show low cooperativity; correspondingly, both the change in enthalpy and heat capacity resemble the denatured protein. (4) In ^{1}H NMR, the low-field and aromatic spectral regions differ substantially from those of both the native and fully unfolded states, reflecting an intermediate level of order. The most strongly perturbed residues are among those that form the native hydrophobic core; on the other hand, certain segments that are helical in the native state are found to be highly protected from solvent exchange, as one would predict from the sequential pathway of folding (Baum et al., 1989).

Obviously, the given criteria are unspecific enough to subsume the wide variety of early transient intermediates accumulating during the first fractions of a second of in vitro renaturation under one term; it is because of possible misleading associations connected with the term molten globule that certain properties of this state have been discussed in some detail. However, it should be borne in mind that it is a whole class of intermediate states that may be trapped under a wide variety of conditions such as high temperature, intermediate denaturant concentrations, low pH ("A-state"), high pH, etc. (Wong & Tanford, 1973; Kuwajima, 1977; Ohgushi & Wada, 1983; Brems et al., 1985; Ptitsyn, 1987). Their common denominator is that, starting from the unfolded state, they are formed within 0.1–0.2 s, independent of the topologies and overall folding rates of their parent proteins (Ptitsyn et al., 1990). They have been observed as equilibrium intermediates mainly for proteins with low overall stability and remain undetected under conditions favoring the native state or in the presence of stabilizing agents. An example illustrating the involvement of the solvent in the formation of the molten globule is the "acid-induced folding" of a number of proteins and synthetic polypeptides (Rudolph & Jaenicke, 1976; Ebert & Kuroyanagi, 1982; Goto et al., 1990; Buchner et al., 1991). Using various strong acids, Goto et al. (1990a,b) were able to show that, in the case of β-lactamase, cytochrome *c*, and apomyoglobin, preferential anion binding to positively charged sites of the proteins is responsible for the structural transition; thus, it is the shielding of intramolecular repulsive forces at strongly acidic pH that favors the formation of the A-state. Other solvent conditions leading to molten globule intermediates may stabilize different states.

The acquisition of the final tertiary (and quaternary) structure occurs by a limited number of pathways characterized by well-defined structural intermediates. The rate-limiting steps are late events in the overall reaction sequence (Creighton, 1978; Segawa & Sugihara, 1984; Goldberg, 1985; Goldenberg & Creighton, 1985; Jaenicke, 1987; Creighton, 1988; Baldwin, 1990; Kim & Baldwin, 1990).

It is evident that restriction of flexibility is essential in the folding process and is responsible for the uniqueness of the native 3D structure. As proven by a variety of experimental methods, including limited proteolysis, NMR, H–D exchange, binding of antibodies, etc., in approaching the native state, free movements of the polypeptide chain are progressively frozen, and finally the molecule is trapped in its native configuration.

The complete description of the folding path of a given protein implies the characterization of the nascent (denatured) state and the final (native) state, together with all intermediates along the $U \rightarrow N$ transition. That there exists a defined pathway of structure formation in monomeric proteins such as basic pancreatic trypsin inhibitor (BPTI) and ribonuclease (RNase) has been shown in pioneering studies by Wetlaufer, Creighton, Baldwin, and co-workers (Ristow & Wetlaufer, 1975; Anderson & Wetlaufer, 1976; Creighton, 1978, 1988; Baldwin & Creighton, 1980; Kim & Baldwin, 1982, 1990). In no case, not even for the given simple model systems, has the complete description of the elementary processes on the kinetic pathway been accomplished.

2D NMR experiments promise direct insight into emerging local structures within the refolding polypeptide chain (States et al., 1987; Udgaonkar & Baldwin, 1988; Roder et al., 1988; States & Kim, 1990; Bycrofft et al., 1990; Matouschek et al., 1990). The way to get hold of intermediate states is to follow the proton exchange rate of individual deuterated peptide NH groups after pulse-labeling with hydrogen at different stages of structure formation. ^{2}H on amide groups exposed to the solvent at the time of labeling exchange for ^{1}H, which can be monitored by 2D NMR when folding is completed; as the amide groups become involved in structure formation, they are protected from exchange.

One further approach to detect emerging local structures has been the use of monoclonal antibodies. In this case, the reappearance of conformation-specific epitopes in the time course of reconstitution is monitored to characterize the detailed folding mechanism (Goldberg & Zetina, 1980; Murry-Brelier & Goldberg, 1988; Friguet et al., 1989; Blond-Elguindi & Goldberg, 1990).

The outcome of these and related investigations has been that there are well-defined folding pathways with identifiable intermediate states, including the above-mentioned molten globule state. Summarizing the results for BPTI, RNases A and T_1, barnase, cytochrome *c*, α-lactalbumin, the β_2 dimer of tryptophan synthase, and ubiquitin, there is now good evidence supporting vectorial folding models ("framework model"); random mechanisms such as the "jigsaw puzzle model" (Harrison & Durbin, 1985) can be excluded.

Due to present limitations to 2D NMR, short peptides or small proteins and protein fragments were used as models in order to determine local structural elements. From these studies it became clear that in aqueous solution short peptides can form stable elements of secondary structure. The fact that there seems to be a correlation between the turn-forming potential of such peptides, on one hand, and the turn probabilities of the same sequences in the protein data base, on the other, suggested that specific sequences may determine local secondary structural elements (Marqusee & Baldwin, 1987; Dyson et al., 1988a,b; Wright et al., 1988). As a consequence of their restricting effect on the conformational space, they may be significant in the overall folding reaction. Due to the space-filling properties of both the polypeptide backbone and the amino acid side chains, available free space in the Ramachandran diagram is a priori reduced to <15%. Further restriction provides a framework directing the folding path in even narrower limits, thus enhancing the rate of the overall folding reaction.

Taken together, the secondary structure of proteins is formed at an early stage in the folding process, providing a scaffolding on which the remaining parts of the polypeptide chain can arrange themselves to end up in the compromise of high packing density and low solvent-accessible surface area (Richards, 1977; Wodak et al., 1987; Jaenicke, 1987; Janin et al., 1988; Kim & Baldwin, 1990). In the overall hierarchical condensation reaction, docking of preformed elements seems to be the rate-limiting step. As in the case of quaternary assembly of subunits, where water release has been clearly proven to be the driving force of structure formation (Lauffer, 1975; Jaenicke, 1987), the formation of the hydrophobic core seems to be crucial in the docking of secondary and supersecondary structural elements. Certain characteristics of the present hierarchical condensation or framework model seem to contradict the above-mentioned properties of the molten globule intermediate, especially its noncooperativity and the exposure of hydrophobic side chains. Presently there is no solution to these discrepancies (Baldwin, 1990).

Folding versus Reconstitution

The structural integrity of proteins in solution depends on the solvent parameters. Accordingly, one would predict that protein folding is strongly influenced by the environment. However, a variety of experimental findings have proven that the effects of solvent conditions upon translation and reconstitution are less critical than expected: in vitro folding and assembly may be accomplished in dilute buffer solution in the absence of components involved in cellular folding events; biologically active thermophilic proteins may be expressed in mesophilic hosts; cotranslational and posttranslational modifications such as glycosylation or processing do not necessarily interfere with the intrinsic capacity of the polypeptide chain to acquire its native 3D structure. Except for the influence of viscosity and specific ligands (coenzymes, substrates, ions), hardly any attempts have been made to mimic the solvent conditions in the cytoplasm in folding experiments. The fact that the refolding chain requires neither extrinsic factors nor the input of energy in order to generate the native structure has been considered sufficient evidence to postulate that the genetic code governs both translation and folding. Whether there is a unique folding code remains to be shown (Fasman, 1989). That it cannot be collinear (as the genetic code of translation) is trivial for a number of reasons: both local next-neighbor ("short-range") *and* nonlocal through-space ("long-range") interactions are involved in the minimization of energy, so as a consequence, identical stretches of polypeptide chains may assume different 3D structures; widely differing sequences from "homologous proteins" code for identical topologies; subdomains and domains as cooperative entities are separated by connecting peptides exhibiting anomalous configurations; in certain cases, extrinsic effects or effectors (not inherent in the amino acid sequence) have been shown to play a significant role in folding. The latter argument has been shown to be essential in cases where ligands such as cofactors serve to stabilize intermediates of folding or assembly (Gerschitz et al., 1978). Other cell biological implications that may interfere with a general 1D → 3D algorithm of protein folding are cellular compartmentalization, co- or posttranslational modification, protein splicing, genome organization, transcription control, codon usage, amino acid pools, discontinuity in the rate of translation, kinetic competition of folding and association, etc. (Jaenicke, 1987, 1988a). Whether, or to what extent, these phenomena affect the 3D structure of a given protein is still under dispute. A variety of observations suggest, however, that as a rule the 3D structure of a protein is fully determined by its amino acid sequence and the solvent environment; it occupies the state of minimum potential energy, determined by physicochemical rather than cell biological criteria.

A few examples will suffice to prove the point: Human tissue plasminogen activator (t-PA), overexpressed in *Escherichia coli*, forms inclusion bodies consisting of carbohydrate-free inactive protein. Complete reduction of disulfide bonds and subsequent formation of the 17 cystine bridges yields >70% of the biologically active protein, which—according to all available criteria—is indistinguishable from the native molecule (Opitz, 1988; Rudolph et al., 1987, 1990b; Jaenicke, 1988a). This implies that, in vitro as well as in vivo, just one out of the 2.2 × 10^{20} theoretically possible combinations of SH groups has been selected, independent of the presence (human t-PA) or absence (recombinant t-PA from *E. coli*) of the carbohydrate moiety. In addition we may conclude that disulfide bridges *stabilize* the native state rather than *determining* the spatial arrangement of the polypeptide backbone. Finally, considering the complex kringle structure of t-PA, it is obvious that not only the local tertiary structure but also the *domain structure* of proteins must be encoded in the amino acid sequence. Neither directionality due to cotranslational "folding by parts" nor discontinuity or "punctuation" due to specific codon usage can play a significant role. A variety of approaches may be considered to prove these conclusions: (i) So far, no host-specific anomalies of recombinant proteins have been found in widely differing guest–host pairs. (ii) Reverting the directionality by using Merrifield synthesis (from the C- to the N-terminus) instead of translation leads to authentic protein. (iii) "Topology-directed modifications" such as cyclization, polymerization, or generation of new terminal ends of polypeptide chains do not necessarily affect the 3D structure.

As indicated in connection with t-PA, glycosylation does not seem to have a significant effect on protein folding. Careful studies using ribonuclease (RNase) and invertase clearly confirm this result: neither the overall structure nor the kinetic mechanism of folding exbibits significant differences when the glycosylated and nonglycosylated proteins are compared. What may be affected is the stability and the tendency to aggregate: the carbohydrate moiety increases the solubility so that upon secretion glycosylation keeps the nascent polypeptide chain from association or aggregation (Krebs et al., 1983; Schmid & Jaenicke, 1987; Schülke & Schmid, 1989; Kern et al., 1991).

In Vivo versus in Vitro Folding. In comparing the in vivo and in vitro self-organization of proteins, in vivo structure formation is generally assumed to yield 100% native protein, while reconstitution in vitro, after optimization, may range from 0 to 100%. As indicated by a number of observations, this assumption may not be true. (1) It is increasingly clear from a variety of systems that a protein's 3D structure and oligomeric state controls not only its functional properties but also its intracellular transport, as well as its ultimate localization in the cell and its overall life span. During secretion, misfolded, misassembled, and unassembled polypeptides are retained in the ER and specifically degraded. This property of the ER provides an inherent quality control leading to the apparent yield of 100% (Hurtley & Helenius, 1989; Pelham 1989). (2) There is clear evidence that under unbalanced physiological conditions in vivo folding may lead to inactive "wrong" conformers, which interfere with unperturbed structure formation. In the case of the phage P22 tailspike protein, even under optimal growth conditions of the bacterium, the yield of in vivo folding is lower than 50% (Haase-Pettingell & King, 1988). (3) The fact that proteins, due to their low free energy of stabilization, are basically close to their denaturation transition implies that "the native state" in vivo may represent a whole set of conformers where "misfits" are continuously removed by proteolysis. (4) Chaperones may be involved in the "spontaneous and autonomous folding" within the cell so that in vivo self-organization of proteins may be catalyzed and/or regulated by protein–protein interactions, including ATP-dependent processes (Pelham, 1986; Rothman, 1989).

Folding in vivo generally takes place within the time range of seconds to minutes. In the case of small monomeric single-domain proteins, in vitro rates of the same order or even faster are observed. In certain cases, extremely fast folding reactions have also been recorded for large multidomain proteins, and even for oligomers (aldolase, triosephosphate isomerase, tryptophan synthase). However, in most cases reconstitution in vitro is exceedingly slow. For example, the reoxidation of RNase requires 20 min for 50% reactivation (Anfinsen et al., 1961); the respective half-times of the reassembly of the pyruvate dehydrogenase complex from *Bacillus stearothermophilus* and the reshuffling of the Fab fragment of immunoglobulin are ≈8 h and ≈15 h. In the case of multichain systems, one of the reasons is the difference in concentration: in vivo folding of polypeptide chains on the polysome yields high local concentrations of "structured monomers", which would cause aggregation in the corresponding in vitro experiment (Zettlmeissl et al., 1979; Light, 1985; Jaenicke, 1987; Rudolph, 1990; Rudolph et al., 1991). Thus, cotranslational folding and folding of the complete chain may show significant kinetic differences, not because of the vectorial "directionality" of the folding process but because of rate-determining association steps often involved in the reconstitution of oligomeric proteins. Other mechanisms that may accelerate in vivo folding include "nucleating effects" of ligands (Gerschitz et al., 1978), catalysis by specific isomerases such as protein disulfide isomerase and peptidyl–prolyl cis–trans isomerase, and regulation by heat-shock proteins and related chaperones [for reviews see Fischer and Schmid (1990), Ellis (1990), and Morimoto et al. (1990)]. Mimicking intracellular redox conditions in the case of proteins stabilized by cystine bridges does not enhance in vitro shuffling to the rate observed in vivo (Rudolph & Fuchs, 1983). Systematic investigations of the effects of cellular components or physicochemical parameters characteristic for the cytoplasm on the rate and mechanism of protein folding are still lacking.

Catalyzed versus Noncatalyzed Folding. The idea that folding, rather than being "simply a function of the order of the amino acids" (Crick, 1958), could be influenced by intermolecular interactions with other proteins goes back to Anfinsen's early redox experiments (Epstein et al., 1963). In the case of the exchange of disulfide groups, the corresponding shuffling enzyme, protein disulfide isomerase (PDI), is localized in the ER, as one would expect for a catalyst involved in protein secretion (Freedman, 1984, 1989). Its biological significance has been clearly proven by depletion and folding experiments (Bulleid & Freedman, 1988). The enzyme catalyzes the formation and breakage of correct disulfide bonds without affecting the folding mechanism (Creighton et al., 1980). The fact that PDI forms a heterodimer with the microsomal triglyceride transfer protein (Wetterau et al., 1990) indicates that various components involved in protein folding and processing in the ER may be organized as a multienzyme complex.

Another well-established class of folding enzymes are peptidyl–prolyl cis–trans isomerases (PPI's), which catalyze the rate-determining step in the folding of proteins containing "accessible" and "essential" *cis*-proline residues (Brandts et al., 1975; Levitt, 1981; Lin et al., 1988; Baldwin, 1989; Fischer & Schmid, 1990). Since translation yields the all-trans configuration of the polypeptide chain and since *cis*-proline bonds are occasionally found in globular proteins, catalysis of proline cis–trans isomerization is expected to be essential for cellular folding.

For long, Brandts' "proline hypothesis" has been debated. However, in recent years, physicochemical, enzymological, and chemical evidence (including catalysis by PPI and site-directed mutagenesis) has gained so much weight that the mechanism can now be considered well-established (Garel & Baldwin, 1973; Schmid & Baldwin, 1978; Kim & Baldwin, 1982, 1990; Fischer & Bang, 1985; Lang et al., 1986, 1987; Lang & Schmid, 1988, 1990; Fischer et al., 1989; Kiefhaber et al., 1990a,b). The sequence specificity of the enzyme catalyzing the reaction has not been unraveled yet. The observation that PPI's are ubiquitous from bacteria to mammals suggests that the enzyme plays a biologically important role. Whether its function is related to catalysis of *folding* by kinetic competition with unproductive side ways on the folding path needs still to be proven. It is worth mentioning that some members of the PPI family are inhibited by immunosuppressives such as cyclosporin (Fischer & Schmid, 1990).

Mechanisms different from conventional enzyme catalysis seem to be involved in the protein interactions connected with the inhibition of premature folding. Keeping the folding polypeptide chain in a more or less unfolded form different from the compact native state is a requirement (i) for protein translocation across membranes and (ii) for correct folding (and assembly) of large polypeptide chains (Neupert & Schatz, 1981; Pelham, 1988; Ellis, 1990; Schlesinger, 1990). In targeting proteins to specific compartments, the leader sequence of ATP-dependent heat-shock proteins may arrest a polypeptide chain in its transport-competent form (Park et al., 1988; Wiech et al., 1990). In the case of large polypeptide chains, molecular chaperones (Laskey et al., 1974; Ellis, 1990) or polypeptide chain binding proteins (PCB's) (Rothman, 1989) keep partially folded or misfolded molecules from aggregation (Jaenicke, 1974; London et al., 1974; Zettlmeissl et al., 1979). They mediate correct folding rather than converting incorrect structures or aggregates back to the native state. They are not components of the final functional state; instead they seem to function by binding specifically and noncovalently to protein surfaces that are transiently exposed during structure formation. This binding is probably reversed under conditions favoring correct tertiary interactions within the nascent protein. ATP hydrolysis is involved in the release of the folded molecule.

The way chaperones work, i.e., how intermolecular interactions promote intramolecular interactions, is difficult to rationalize. There are examples such as scaffolding proteins in phage morphopoiesis (Kellenberger, 1984) and nucleoplasmins (Laskey et al., 1978; Dingwall & Laskey, 1990) that have been in the literature for many years. However, during the last 2 years, more than a dozen proteins have been detected in bacteria, plants, and higher eukaryotes which seem to indicate that helper proteins are a general concept in the transport, folding, and assembly of proteins (Ellis, 1990; Schlesinger, 1990). Faced with the wide variety of proteins, at present no general mechanism can be envisaged for these diverse functions. Since physicochemical investigations are only beginning, four examples may serve to illustrate certain principles.

(i) The structural basis underlying the recognition of a wide range of unrelated signal peptides by the signal recognition particle is a repeating pattern of methionine residues that are postulated to form a "bristle" accommodating hydrophobic sequences (Bernstein et al., 1989).

(ii) The only high-resolution crystal structure of a chaperone presently available is the one of PapD, a protein that mediates the assembly of pili in *E. coli* without being part of the final complex. The protein reveals a CH2 immunoglobulin fold (Holmgren & Brändên, 1989). The best candidate for the "intermediary binding site" for the pili subunits is a wide hydrophobic crevice between the two domains that shows features not commonly observed at protein surfaces.

(iii) GroEL from *E. coli*, a weak potassium-dependent ATPase, interacts with GroES to form a complex with 7-fold symmetry. In the presence of MgATP, this complex facilitates the in vitro reconstitution of active Rubisco from the unfolded state (Goloubinoff et al., 1989; Viitanen et al., 1990). The mitochondrial GroEL homologues (Hsp60 and -10) have been reported to be essential for the correct association of a number of oligomeric enzymes (including the suppression of temperature-sensitive mutations of the tailspike protein of *Salmonella* phage P22) as well as for the correct folding of monomeric imported proteins (Cheng et al., 1988; Ostermann et al., 1989; van Dyk et al., 1989). Recently, the GroE system has been used to promote the in vitro refolding of citrate synthase. With the use of light scattering to monitor the kinetic competition of reactivation and aggregation, it is evident that GroE inhibits aggregation with high efficiency, thus facilitating refolding and reactivation in a specific, ATP-dependent fashion (Buchner et al., 1991).

(iv) Chaperones are assumed to show sequence specificity in polypeptide chain binding and release. Flynn et al. (1989) were able to show that BIP (Hsp70) and cytoplasmic Hsc70 are able to bind short peptides in an ATP-dependent manner, albeit with high K_m values. Whether there is a unique consensus sequence or a structural motif or any colligative property of the peptide involved in the interaction is still unresolved. In the case of the SecB protein from *E. coli*, the chaperone arrests folding by binding unfolded precursor proteins with high affinity, in this way keeping them in their translocation-competent form. There are no specific interactions with the signal sequence involved; its function is merely the retardation of folding in order to optimize the complexation with the chaperone (Liu et al., 1989; Randall et al., 1990).

The "repair" function of polypeptide chain binding proteins in terms of the capacity to dissolve aggregates is still hypothetical. In both the GroE–citrate synthase and the GroE–Rubisco system, solubilization of aggregates in vitro has been shown to be unsuccessful (Goloubinoff et al., 1989; Buchner et al., 1991). On the other hand, Hsc70 does uncoat newly budded clathrin-coated vesicles to enable their fusion; ATP drives the dissociation–association cycle (Rothman & Schmid, 1986). There is some evidence that the structure of the target protein dictates the release from the chaperone: obviously, proteins incapable of correct folding are trapped and finally subjected to degradation; in this context, Hsp70 is assumed to be part of a larger complex connected with general "editing function" (Sheffield et al., 1990; Schmid, 1991).

Chaperones in the overall scheme of protein self-organization ultimately have two functions: first, they stabilize some (late) folding intermediate, arresting the nascent polypeptide chain in a translocation-competent form, and second, they protect this intermediate from premature association and turnover as the most important side reactions off the folding pathway. How they perform (or catalyze) these reactions is an open question that has been addressed so far mainly by teleological speculation, sometimes approaching vital force concepts. The hard facts have been reviewed in a most lucid way by Fischer and Schmid (1990).

Structure Prediction versus Postdiction. The in vivo → in vitro issue is mainly a kinetic problem. It does not affect Anfinsen's orginal assumption that the folding process is thermodynamically determined and that no genetic information other than that present in the amino acid sequence of the protein is required. Thus, in spite of the cell biological implications, it should be possible to translate by calculations the 1D amino acid sequence into its corresponding 3D configuration.

There have been numerous attempts to forecast the 3D structure of proteins or their mode of folding: search programs for sequence homologies have been successfully applied to correlate given primary structures to a limited number of protein "families" (Doolittle & Richardson, 1981). Statistical analyses of preferences for α-helices, β-strands, turns, loops, or random structures have been used to predict the secondary structure with reliabilities of the order of 70%. Topological considerations and docking procedures have been developed to optimize both minimum hydrophobic surface area and maximum packing (Wodak et al., 1987). Energy minimization and molecular dynamics calculations, as well as semi-quantum-mechanical and statistical mechanical methods proved useful in reducing the number of possible conformations from an astronomically high value to only a few (Brooks et al., 1988). They have been most valuable in characterizing conformational changes with high precision (McCammon & Harvey, 1988).

A combination of all available methods in terms of knowledge-based computer-aided structure predictions has been conceived by Blundell et al. (1987). The result of $\simeq$90% correct prediction with an rms deviation < 3 Å is highly satisfactory from the theoretical point of view; however, in order to predict the structure *and function*, i.e., to define the active site of an enzyme or a binding site of a given protein with some degree of confidence, the local configuration has to be determined to significantly higher precision. Thus, predictions have still to be taken with a grain of salt. A number of reasons are responsible for the limited success. As indicated, the predictive methods on the whole are based on empirical data obtained from crystallized (i.e., crystallizable) globular proteins. Therefore, they are fundamentally *post*dictive and may be biased. As pointed out by Rooman and Wodak (1988), even for the resticted set of crystallizable soluble globular proteins, the small size of the data-base puts a limit to any reasonable structure prediction. Modifications at the transcriptional level, as well as co- or posttranslational processing of the polypeptide chain, may affect the 3D structure so that the "translation" of the amino acid sequence from the DNA sequence via the genetic code may lead to erroneous results. The present state of the art has been summarized in a number of recent reviews (Jaenicke, 1987, 1988a; Fasman, 1989). Promising new concepts explaining the prevalence of secondary structural elements in regenerating compact conformations and the occurrence of well-defined protein families come from polymer statistics and pattern-recognition techniques (Dill, 1990; Kaden et al., 1990; M. van Heel, unpublished results).

Local versus Global Structure

The self-organization of proteins reflects the hierarchy of protein structure, gradually proceeding from local secondary and supersecondary structure via subdomains and domains to the complete tertiary structure (Rossmann & Argos, 1981). Correct docking of domains to form the native tertiary structure can only occur if distinct local conformations in the interface are present. Similarly, in the case of structures requiring subunit assembly, subunit recognition requires at least the domains involved in quaternary interactions to be folded. It is obvious that in order to avoid misfolding and misassembly, the various structural levels must be reached one after the other. This implies that the overall mechanism must involve a sequence of (first-order) folding steps, generating domains and/or association-competent "structured monomers" and subsequent (second-order) subunit association. Whether the corresponding kinetics can be quantitatively described by a consecutive uni-bimolecular or a more complex model, and which of the single reactions is rate-limiting, depends on both the protein and the experimental conditions (Jaenicke, 1987).

Local Structures. The analysis of the folding mechanism at the subdomain level poses two questions: (i) Do "local nonrandom conformations" in the nascent or reconstituting polypeptide chain initiate folding, and (ii) do short polypeptides or fragments of a given protein form a nativelike 3D structure, and what is the minimum size of stable subdomains.

(i) As mentioned, local nonrandom conformations in short peptide fragments of proteins as short as 4 or 5 residues have been found to be detectable under conditions where native proteins fold (Wright et al., 1988). Due to the stabilizing effect of charged side chains interacting with the helix dipole, short α-helical peptides have been found to be more stable in solution than would be predicted from theory (Shoemaker et al., 1987). By using oligopeptides with 16 or 17 residues, the role of ion pairs on the secondary structure of synthetic peptides has been studied in detail. Spacing the charged residues at different positions clearly established the stabilizing effect predicted for an α-helical array (Marqusee & Baldwin, 1987).

β-turns were observed, e.g., in an immunogenic peptide fragment of the influenza virus hemagglutinin HA 1 chain; truncation of the nonapeptide YPYDVPDYA to the N-terminal tetrapeptide still allowed retention of the reverse-turn conformation (Dyson et al., 1988a,b).

In summarizing available data, it is obvious that local transient structures (in rapid equilibrium with the fully randomized chain) should be present in the nascent polypeptide chain. For small proteins such as RNAse, the "unfolded state" seems still to retain residual structure (Chavez & Scheraga, 1980; Haas et al., 1988). It is tempting to assume that the formation or retention of local structures in certain regions of the polypeptide chain could be initiating steps in protein folding. Although such structures would be only marginally stable, they would efficiently reduce the conformational space, thus directing (and this way enhancing) the folding reaction. Clearly, the role of such "initiation sites" must be restricted to early folding steps, prior to the formation of well-populated intermediates and much before the rate-limiting step. A number of experimental findings seem to support the idea that α-helices, β-turns, and hydrophobic clusters are "seeds" of protein self-organization (Wright et al., 1988; Yu & King, 1988). However, there is some indication that the previously mentioned oligopeptides have no substantial tendency to adopt the same conformation in unrelated protein structures; also, some reverse turns observed in the small peptides seem to be absent in the known 3D structures of proteins with these sequences.

In the case of BPTI and RNAse, Creighton (1988) has shown that nonrandom conformations in the unfolded or nascent proteins are insignificant for the kinetics of folding. The intrinsically stable α-helix at the N-terminus of RNase A does not serve as an initiation site but is incorporated into the folded conformation only as the last detectable step (Brems & Baldwin, 1984). Thus, folding models involving initiation by local structural elements and subsequent modular assembly, although seemingly plausible, are still hypothetical.

(ii) As indicated, short helices and β-bends may show high intrinsic stability; however, their structure as separate entities is not necessarily identical with the one observed within their intact parent protein. In studying protein fragments obtained by limited proteolysis or semisynthesis the same problem arises, apart from the question of whether a given primary structure altogether yields a stable and well-defined 3D structure (Jaenicke, 1987). With thermolysin as a model, the folding/unfolding of the CNBr fragments 1–120, 121–205, and 206–316 has shown that the N-terminal portion of the enzyme stabilizes the all-helical C-terminal domain. The latter may be shortened drastically without significantly affecting its structural features (Vita et al., 1989). Only the C-terminal helix (residues 297–316) is too short to maintain its native conformation in aqueous solution (Rashin, 1984; Fontana, 1990). From this we may conclude that subdomains down to the size of a two-helix bundle not only show intrinsic stability but also fold as independent entities ("folding by parts"; Wetlaufer, 1981). In this connection it is of interest to find out whether the N- or C-terminal ends of the polypeptide chain are indispensable for proper folding. Taking RNase and LDH as examples, it has been shown that the N-terminal end of both proteins can be cleaved without altering the overall topology. However, the stability is found to be drastically reduced: the S-protein of RNase S cannot be reoxidized in the absence of the S-peptide, and in the case of LDH, the "proteolytic dimer" is inactive unless stabilizing salt is added (Richards & Vithayathil, 1959; Girg et al., 1983; Opitz et al., 1987). Cleaving off the C-terminus in the case of RNase is sufficient to block the regain of activity after reduction/reoxidation, while LDH can still be reconstituted (Opitz et al., 1987; Teschner & Rudolph, 1987). Which part of the backbone is indispensable for the unaltered topology of a given protein cannot be predicted. In the case of BPTI, RNase, and larger peptide sequences up to the range of β-galactosidase, circular permutation of parts of the sequence, fragmentation, chain extension, derivatization with oligopeptide branches [poly(DL-Ala)–RNase], joining of subunits, and hybrid formation by peptide interchange have clearly shown that tertiary structure formation and even subunit assembly may tolerate an unexpectedly wide range of sequence variations, as well as variations in chain connectivities (Epstein et al., 1963; Wetlaufer, 1981; Goldenberg & Creighton, 1984; Kuchinke & Müller-Hill, 1985; Opitz et al., 1987; Jaenicke, 1987; Luger et al., 1989; Buchner & Rudolph, 1991). Obviously, certain core regions determine the overall topology, while "peripheral parts" of the polypeptide chain may be altered or even lacking.

Domains. Proceeding from structural elements and subdomains to higher levels in the hierarchy of protein structure, the sensitive distance dependence of the stabilizing and destabilizing weak interactions causes a high degree of specificity. The minimization of accessible surface area, accomplished by docking and association, allows domains and subunits to "recognize" their respective counterparts (Wodak et al., 1987). This holds even if the linker peptide connecting the domains is missing: in the case of LDH it has been shown that nicked subunits can be reconstituted due to the specificity of the interdomain contacts (Opitz et al., 1987). On the other hand, wrong pairing reactions may occur, giving rise to incomplete reconstitution. The fact that the yield of the reaction of nicked LDH is only ≈20% may be explained this way. Octopine dehydrogenase, a monomeric homologue of LDH is another example: upon denaturation/renaturation, the two-domain enzyme shows only 70% reactivation. After separating inactive material, a second denaturation/renaturation cycle again yields 70% (of 70%). The remaining 30% in each cycle represents inactive monomers with nativelike CD but nonnative fluorescence properties (Zettlmeissl et al., 1984; Jaenicke, 1988a). As suggested by the decelerating effect of increased solvent viscosity on reactivation (Teschner et al., 1987), the rate-limiting reaction in this case is domain pairing rather than folding. PPI does not affect the kinetics of reactivation. Thus, the reduced yields may be attributed to wrong domain interactions. In vivo, the directionality of the folding reaction is assumed to minimize side reactions.

Domains, defined by visual inspection of crystal structures, surface area calculations, deconvolution of equilibrium transitions, limited proteolysis, sequence homology at the protein level, or intron–exon organization at the DNA level, are suggested to fold cotranslationally in discrete modules (Bergman & Kuehl, 1979). Making use of modular assembly or folding by parts, the self-organization of large proteins is speeded up by many orders of magnitude (Wetlaufer, 1973); at the same time, wrong intramolecular nonlocal interactions are minimized. Summarizing presently available data [for review, cf. Jaenicke (1987)], the overall folding reaction is characterized by multistep transitions with independent folding/unfolding kinetics involving consecutive folding and merging of the individual "lobes". In the simplest case, then, folding of a two-domain protein may be described by superimposing on one another the individual folding reactions of the constituent parts of the complex with a subsequent pairing reaction. Depending on mutual stabilization effects, domain pairing may or may not be significant. To give an example, the equilibrium transition and the unfolding/folding kinetics of γ-II-crystallin from calf eye lens, a prototype two-domain protein, has been shown to be quantitatively described by a three-state model;

$$\mathrm{N} \rightleftharpoons \mathrm{I} \rightleftharpoons \mathrm{U} \qquad (2)$$

where N indicates native γ-II-crystallin, I the intermediate with the N-terminal domain intact and the C-terminal domain in its random state, and U the denatured state, respectively (Rudolph et al., 1990a). The model has been nicely corroborated by fragment studies (Sharma et al., 1990). The merging of the two domains in the overall reaction remains undetectable: increasing viscosity of the solvent has no effect on the N → U folding kinetics, indicating that the rate-limiting process must be the folding rather than the pairing of the two domains (Siebendritt, 1989). As shown by the incomplete reactivation of octopine dehydrogenase (see above), this observation cannot be generalized (Zettlmeissl et al., 1984).

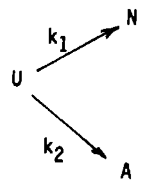
Association versus Aggregation. The early stages during the (re-) folding of *oligomeric proteins* are expected to be identical with those involved in the self-organization of single-chain proteins. Thus, subunit polypeptide chains may be assumed to fold first into subdomains and/or domains that will subsequently collapse to form structured monomers with nativelike structure. They will finally undergo association and shuffling to yield the native quaternary structure so that the overall process may be written as a sequence of (unimolecular) folding reactions and (bimolecular) association steps according to

$$n\mathrm{M} \rightarrow n\mathrm{M}' \rightarrow (n/2)\mathrm{D} \rightarrow (n/2)\mathrm{D}' \rightarrow (n/4)\mathrm{T} \rightarrow (n/4)\mathrm{T}' \rightarrow \ldots \rightarrow \mathrm{P}_n \qquad (3)$$

where n is the number of subunits and M, M′, D, D′, T, T′, and P_n are monomers, dimers, and tetramers in different conformational states and the n-mer, respectively (Jaenicke & Rudolph, 1986). The number of intermediate steps may vary so that the single arrows clearly oversimplify the actual mechanism. This holds for both the folding of monomers and the various stages of assembly (Murry-Brelier & Goldberg, 1988; Blond-Elguindi & Goldberg, 1990). In the case of multisubunit assemblies such as the pyruvate dehydrogenase multienzyme complex, reactivation may be determined by unimolecular shuffling at the multimer level (Jaenicke & Perham, 1982). Presently, available methods are insufficient to analyze the detailed folding/assembly mechanism. What has been accomplished in a number of cases is the elucidation of the detailed association mechanism, including "assembly maps" of complex organelles such as the ribosome, UDP-glucose dehydrogenase, or ferritin (Nomura & Held, 1974; Nierhaus, 1982; Jaenicke et al., 1986; Gerl et al., 1988). For a detailed discussion of the correlation of folding and association for a wide variety of proteins, see Jaenicke (1987).

There is no qualitative difference between interdomain and intersubunit interactions. Therefore, it is obvious that in the given hypothetical mechanism kinetic competition between folding and association may occur. This holds because association (in contrast to folding) is of higher than first order. In vitro, the corresponding side reaction leads to "wrong aggregation" (Jaenicke, 1974; London et al., 1974; Zettlmeissl et al., 1979). Its influence can be minimized by choosing low protein concentration (slowing down bimolecular processes) and solvent conditions that stabilize the native state. This is the reason why renativation experiments with oligomeric proteins are commonly performed under essentially irreversible conditions, i.e., far from equilibrium. Accordingly, eq 3 was written as a unidirectional reaction (Jaenicke & Rudolph, 1989).

In overexpressing strains of bacteria, the formation of inclusion bodies (IB's) may be assumed to result from precisely the same mechanism responsible for wrong aggregates in vitro. In fact, deposition of recombinant proteins at high expression rates can be quantitatively described by the kinetic competition of first-order folding and diffusion-controlled higher-order aggregation according to



(4)

where U, N, and A are partially folded intermediates and native and aggregated states, respectively, and k_1 and k_2 are the rate constants for the rate-determining first-order folding (k_1) and the second-order aggregation steps (k_2). On the basis of this simple kinetic model, reactivation yields can be computed as a function of the initial concentration of the denatured protein (Rudolph, 1990; Rudolph et al., 1991).

The physical nature of IB's has become a central issue because of their technological importance in the recovery of recombinant proteins (Light, 1985; Marston, 1986; Hoess et al., 1988; Mitraki & King, 1989; Rudolph, 1990). From the foregoing paragraphs it is obvious that aggregates and IB's equally derive from partially folded intermediates rather than unfolded or mature native protein. In most cases the foreign recombinant gene product is the main component. In addition, IB's may contain proteins of the host, such as RNA polymerase and ribosomal proteins, as well as nucleic acids. As has been discussed in connection with the renaturation of plasminogen activator (tPA), general strategies for the downstream processing of recombinant proteins have been worked out with unprecedented success (Rudolph et al., 1987, 1990b; Rudolph, 1990; Buchner & Rudolph, 1991).

Approaches restricting aggregation in vitro may help in our understanding of how the cell under normal conditions avoids IB formation. (i) The fact that at low protein concentration aggregation is minimal (corresponding to low expression in vivo) has already been mentioned. (ii) In the case of protein immobilization, the aggregation is hindered sterically, comparable to the folding of the nascent chain at the ribosome (Light, 1985). (iii) Chemical modification has been shown to increase the solubility of the unfolded protein and, thus, decrease the potential for aggregation (Light, 1985). Closely related are attempts to increase the yield of renaturation by adding low concentrations of moderate denaturants (e.g., arginine) to the renaturation buffer (Rudolph, 1990). Considering glycoproteins (e.g., invertase from yeast), the carbohydrate moiety shows a similar solubilizing effect, leading to a decrease in aggregation in the case of the highly glycosylated extracellular form compared to the core-glycosylated secretion intermediate or the carbohydrate-free internal form of the enzyme (Kern et al., 1991). (iv) Temperature, viscosity, and specific ligands have been shown to be important in preventing wrong aggregation. In a number of cases, reconstitution could only be accomplished at low temperature (Rubisco, rhodanese, tailspike protein of *Salmonella* phage P22); in the case of thermophilic enzymes, high temperature seems to be required (Jaenicke, 1987). In connection with the effect of ligands, hemoglobin and metalloproteins may serve as examples (Schneider et al., 1969; Gerschitz et al., 1978). In both cases, thermal stability and solubility have been shown to depend on the respective ligands, heme or metal ions. In the case of alcohol dehydrogenase, folding in the absence of Zn^{2+} leads to 100% aggregates. After addition of Zn^{2+}, the side reaction is quenched because the ligand stabilizes a folding intermediate, thus guiding the protein on the correct folding path. (v) Chaperones have been previously discussed in detail concerning their capacity to arrest partially folded or nonassembled polypeptide chains and prevent them from premature aggregation. Referring again to the renaturation of citrate synthase as an example, it is obvious that low yields of in vitro reconstitution can be improved dramatically in the presence of GroE. The chaperone is specific: it suppresses aggregation but it does not redissolve aggregates present in advance (Goloubinoff et al., 1989; Buchner et al., 1991). Taking the previous findings together, it is evident that in the cytoplasm and the ER conditions prevail that allow proper folding and association of the nascent polypeptide chain. However, misfolding and misassembly do occur, leading to wrong conformers or aggregates. They may undergo turnover, or they may be deposited as inclusion bodies or retained in the ER (Pelham, 1989; Mitraki & King, 1989; Hurtley & Helenius, 1989; Rudolph, 1990).

How proper assembly actually takes place at the cellular level is still unknown. The simplest model involves self-assembly via random collisions; the way chaperones enter the game is presently one of the most challenging questions in the field. The only well-established example that allows some insight into the correlation of folding, association, and misassembly in vivo is the tailspike protein of *Salmonella* phage P22. The tailspike in its native state is a trimer that differs from the nascent polypeptide chain and the partially folded "protrimer" by its exceptional resistance to SDS, proteases, and heat (King et al., 1987). The fraction of protein capable of maturing to the native form decreases with increasing temperature due to wrong aggregation. Because intermediates and the final product are energetically and kinetically well separated and because a large set of temperature-sensitive folding (*tsf*) mutants have been analyzed, it was possible to correlate alterations of the folding pattern with specific changes in the amino acid sequence. Data fully agree with the previously described aggregation mechanism (eq 4), with temperature as the critical parameter for the partitioning (Haase-Pettingell & King, 1988; Sturtevant et al., 1989). Chains synthesized at high temperature either aggregate or reenter the productive pathway if shifted to low temperature early enough. The data are consistent with a scheme in which a productive intermediate (I^{pf}) is thermolabile and can partially melt to a species prone to aggregation (I^{pr*}); [I^{pf}] is a soluble partially folded early intermediate, and [I^{pT}] is the intermediate competent to associate into the protrimer, [pT]:

$$\begin{tikzcd}
\text{Translation} \arrow[r] & {[I^{pf}]} \arrow[r, leftrightarrow] \arrow[d] & {[I^{pT}]} \arrow[r, leftrightarrow] & {[pT]} \arrow[r] & \text{native} \\
 & {[I^{pf*}]} \arrow[d] & & & \\
 & {[\text{Aggregate}]} & & &
\end{tikzcd} \tag{5}$$

In the case of *tsf* mutants, the early intermediate is destabilized, speeding up the off pathway, thus inhibiting phage formation.

Conclusions

The mechanism of protein biosynthesis is well-established. Although highly complex, each piece in the puzzle of DNA, RNA, and the various factors has been put into its proper place. However, the real importance of a gene remains obscure as long as we do not understand how the one-dimensional information at the sequence level is translated into the corresponding unique three-dimensional arrangement of a given polypeptide chain. To the present day, the code of protein folding is unknown, even for short peptides or small single-chain proteins. In the case of long polypeptide chains and assembly systems, domains and subunit interactions complicate the situation due to their involvement within and without the respective structural entities. However, even highly complex systems have been reconstituted in vitro, thus proving that, apart from short-range interactions responsible for the secondary structure, also cystine cross-links, domains, and intersubunit interactions must be programmed by the amino acid sequence. There are only very few exceptions where all attempts to renature a protein failed. Either people did not try long enough or the nascent and denatured polypeptide chain differ significantly due to posttranslational modification or other cell biological implications (Jaenicke, 1988a).

Such differences would interfere not only with folding but equally well with structure prediction. In this context, summarizing all available prediction methods, presently semiempirical combinatorial approaches are the only way to arrive at a satisfactory result. The highest accuracy (rms deviations < 2 Å) has been achieved by proceeding in a hierarchical manner from "joint predictions" of structural motifs via "template search" to simplified and finally refined energy calculations, including molecular dynamics (Blundell et al., 1987; Fasman, 1989). Interactive 3D computer graphics methods have been most useful tools at all stages of the given procedures. Their data-base has been systematically collected from X-ray crystallography and sequence homology or some general analogy with protein families, making use of the basic motifs of protein structure (Blundell & Sternberg, 1985; Crawford et al., 1987; Richardson & Richardson, 1989a,b). It remains to be seen whether new concepts such as lattice simulation, pattern recognition, or neural network methods keep their promise.

Energy minimization has to take into account the compensation of repulsive forces (including the decrease in chain entropy) on one hand and the sum of the attractive intermolecular interactions (plus the entropy of solvation) on the other. Considering the magnitude of the various contributions, the free energy of stabilization turns out to be a small difference of big numbers. It is essentially this observation that renders energy calculations ambiguous, especially because the energy increments of hydration are most difficult to assess (Levitt, 1988).

The "energy hyperspace" yielding the native configuration as the final state of folding is highly complex. Taking the reshuffling of BPTI as an example, the various intermediates differ in their stability by 3–5 kJ/mol; the respective activation energies are of the order of 10–20 kJ/mol (Creighton, 1978). From these and similar data for other proteins, we may conclude that under physiological conditions native proteins are near the borderline of denaturation and that the native state is expected to occupy the global minimum of potential energy.

The time scale of folding in vitro and in vivo has been a controversial issue. For small monomeric proteins the in vitro folding rate may exceed the rate of protein synthesis by 2 orders of magnitude; on the other hand, proper protein folding and assembly within the cell typically require more time than chain synthesis (Rothman, 1989). For large multimeric proteins, in vitro reconstitution may take hours. Tsou (1988) argued that cotranslational folding and posttranslational adjustments, together with "effects of the local environment in the cell", may speed up the process. It is the environment of the machinery of protein synthesis, i.e., chaperones (PCB proteins), which—due to their "antifolding" and "antiaggregation" activity—regulate the kinetics of structure formation within a relatively narrow time range. Further studies are required to understand the physical chemistry as well as the cell biology behind chaperone action.

The detailed folding mechanism in terms of the description of the initial and final states and significantly populated intermediates on the pathway of folding and association has not been accomplished for any protein. Certain rate-limiting steps have been clarified with the result that there are catalysts available in the cell for the three most important slow steps: disulfide exchange, proline cis–trans isomerization, and association. The respective enzymes, PDI, PPI, and chaperones, are ubiquitous in both prokaryotes and eukaryotes and are expressed to high levels. In the case of PDI and chaperones the biological significance of these helper proteins in the context of protein folding has been proven unambiguously; for PPI it is most suggestive.

In the case of the traditional model proteins, BPTI, RNase, and tryptophan synthase, studies on folding intermediates have reached a level that promises to come close to a full description of the folding path. Also a number of multisubunit systems, such as ACTase, ferritin, pyruvate dehydrogenase, and the ribosome, have been worked out so that assembly pathways and assembly maps have begun to become clear. The most important gap in our present knowledge, and the most unjustified restriction in this review, refers to membrane proteins. Very few data have been reported in this area. The most detailed analysis refers to the reconstitution of bacteriorhodopsin (Popot et al., 1987). Problems connected with the oligomerization of nascent proteins in the ER have been addressed by Hurtley and Helenius (1989). Joint efforts of physicists, protein chemists, and cell biologists are required to further investigate this issue, which will give new and highly relevant insights into the correlations between the structure, function, and energetics of proteins at the cellular level.

Acknowledgments

The kind help of Agathe Jaenicke in preparing this manuscript is gratefully acknowledged. I thank the Fogarty International Center for Advanced Study, National Institutes of Health, Bethesda, MD, for generous support and hospitality. Fruitful discussions with G. Böhm, J. Buchner, M. E. Goldberg, P. L. Privalov, R. Rudolph, F. X. Schmid, and R. Seckler have been most important in clarifying many aspects presented in this review.

References

Anderson, D. E., Becktel, W. J., & Dahlquist, F. W. (1990) *Biochemistry 29*, 2403–2408.

Anderson, W. L., & Wetlaufer, D. B. (1976) *J. Biol. Chem. 251*, 3147–3153.

Anfinsen, C. B. (1973) *Science 181*, 223–230.

Anfinsen, C. B., & Scheraga, H. A. (1975) *Adv. Protein Chem. 29*, 205–300.

Anfinsen, C. B., Haber, E., Sela, M., & White, F. H., Jr. (1961) *Proc. Natl. Acad. Sci. U.S.A. 47*, 1309–1314.

Anson, M. L. (1945) *Adv. Protein. Chem. 2*, 361–384.

Baldwin, R. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8069–8072.

Baldwin, R. L. (1989) *Trends Biochem. Sci. 14*, 291–294.

Baldwin, R. L. (1990) *Nature 346*, 409–410.

Baldwin, R. L., & Creighton, T. E. (1980) in *Protein Folding* (Jaenicke, R., Ed.) pp 217–260, Elsevier-North Holland, Amsterdam, New York.

Barlow, D. J., & Thornton, J. M. (1983) *J. Mol. Biol. 168*, 867–885.

Baum, J., Dobson, C. M., Evans, P. A., & Hanley, C. (1989) *Biochemistry 28*, 7–13.

Bergman, L. W., & Kuehl, W. M. (1979) *J. Supramol. Struct. 11*, 9–24.

Bernstein, H. D., Poritz, H. A., Strub, K., Hoben, P. J., Brenner, S., & Walter, P. (1989) *Nature 340*, 482–486.

Blond-Elguindi, S., & Goldberg, M. E. (1990) *Biochemistry 29*, 2409–2417.

Bloomer, A. C., & Butler, P. J. G. (1986) in *The Plant Viruses* (van Regenmortel, M. H. V., & Fraenkel-Conradt, H., Eds.) Vol. 2, pp 19–57, Plenum, New York.

Blundell, T. L., & Sternberg, M. J. E. (1985) *Trends Biochem. Sci. 3*, 228–235.

Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E., & Thornton, J. M. (1987) *Nature 326*, 347–352.

Brandts, J. F., Halverson, H. R., & Brennan, M. (1975) *Biochemistry 14*, 4953–4963.

Brems, D. N., & Baldwin, R. L. (1984) *J. Mol. Biol. 180*, 1141–1156.

Brems, D. N., Plaisted, S. M., Havel, H. A., Kauffman, E. W., Stodola, J. D., Eaton, L. C., & White, R. D. (1985) *Biochemistry 24*, 7662–7668.

Brooks, C. L., III, Karplus, M., & Montgomery Pettitt, B. (1988) *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, p 259, J. Wiley, New York.

Buchner, J., & Rudolph, R. (1989) *Dechema Biotechnol. Conf. 3B*, 1035–1040.

Buchner, J., & Rudolph, R. (1991) *Bio/Technology 9*, 157–162.

Buchner, J., Schmidt, M., Fuchs, M., Jaenicke, R., Rudolph, R., Schmid, F. X., & Kiefhaber, T. (1991) *Biochemistry 30*, 1586–1591.

Bulleid, N. J., & Freedman, R. B. (1988) *Nature 335*, 649–651.

Bycrofft, M., Matouschek, A., Kellis, J. T., Jr., Serrano, L., & Fersht, A. R. (1990) *Nature 346*, 488–490.

Chavez, L. G., & Scheraga, H. A. (1980) *Biochemistry 19*, 1005–1012.

Chen, B., & Schellman, J. A. (1989) *Biochemistry 28*, 685–699.

Cheng, M. J., Hartl, F.-U., Martin, J., Pollock, R. A., Kalousek, F., Neupert, W., & Horwich, A. L. (1989) *Nature 337*, 620–625.

Crawford, I. P., Niermann, T., & Kirschner, K. (1987) *Proteins: Struct., Funct., Genet. 2*, 118–129.

Creighton, T. E. (1978) *Prog. Biophys. Mol. Biol. 33*, 231–297.

Creighton, T. E. (1988) *Biophys. Chem. 31*, 155–162.

Creighton, T. E. (1990) *Biochem. J. 270*, 1–16.

Creighton, T. E. (1991) *Curr. Opinion Struct. Biol.* (in press).

Creighton, T. E., Hillson, D. A., & Freedman, R. B. (1980) *J. Mol. Biol. 142*, 43–62.

Crick, F. H. C. (1958) *Symp. Soc. Exp. Biol. 13*, 138–163.

De Grado, W. F. (1988) *Adv. Protein. Chem. 39*, 51–124.

Dill, K. A. (1985) *Biochemistry 24*, 1501–1509.

Dill, K. A. (1990) *Biochemistry 29*, 7133–7155.

Dill, K. A., & Alonso, D. O. V. (1988) in *Protein Structure and Protein Engineering* (Huber, R., & Winnacker, E.-L., Eds.) Colloquium Mosbach, Vol. 39, pp 51–58, Springer Verlag, Berlin, Heidelberg, New York.

Dill, K. A., Privalov, P. L., Gill, S. J., & Murphy, K. P. (1990) *Science 244*, 297–298.

Dingwall, C., & Laskey, R. A. (1990) *Semin. Cell Biol. 1*, 11–17.

Dolgikh, D. A., Abaturov, L. V., Bolotina, I. A., Brazhnikov, E. V., Bushuev, V. N., Bychkova, V. E., Gilmanshin, R. I., Lebedev, Y. O., Semisotnov, G. V., Tiktopulo, E. I., & Ptitsyn, O. B. (1985) *Eur. Biophys. J. 13*, 109–121.

Dyson, H.-J., Rance, M., Houghton, R. A., Lerner, R. A., & Wright, P. E. (1988a) *J. Mol. Biol. 201*, 161–200.

Dyson, H.-J., Rance, M., Houghton, R. A., Wright, P. E., & Lerner, R. A. (1988b) *J. Mol. Biol. 201*, 201–217.

Ebert, G., & Kuroyanagi, Y. (1982) *Polymer 23*, 1147–1158.

Eisenberg, D., & McLachlan, A. D. (1986) *Nature 319*, 199–203.

Ellis, R. J. (1990) *Semin. Cell Biol. 1*, 1–9.

Epstein, C. J., Goldberger, R. F., & Anfinsen, C. B. (1963) *Cold Spring Harbor Symp. Quant. Biol. 28*, 439–449.

Fasman, G. D., Ed. (1989) *Prediction of Protein Structure and the Principles of Protein Conformation*, p 798, Plenum, New York, London.

Fersht, A. R. (1972) *J. Mol. Biol. 64*, 497–509.

Finkelstein, A. V., & Ptitsyn, O. B. (1987) *Prog. Biophys. Mol. Biol. 50*, 171–190.

Finney, J. L. (1982) in *Biophysics of Water* (Franks, F., & Mathias, S., Eds.) pp. 55–58, 73–95, Wiley, Chichester, England.

Fischer, G., & Bang, H. (1985) *Biochim. Biophys. Acta 828*, 39–42.

Fischer, G., & Schmid, F. X. (1990) *Biochemistry 29*, 2205–2212.

Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T., & Schmid, F. X. (1989) *Nature 337*, 268–270.

Flynn, G. C., Chappell, T. G., & Rothman, J. E. (1989) *Science 245*, 385–390.

Fontana, A. (1990) in *Peptides: Chemistry, Structure and Biology* (Rivier, J. E., & Marshall, G. R., Eds.) pp 557–565, Esom, Leiden, The Netherlands.

Franks, F., & Hatley, R. H. M. (1990) *Adv. Low Temp. Biol. 1* (in press).

Freedman, R. B. (1984) *Trends Biochem. Sci. 9*, 438–441.

Freedman, R. B. (1989) *Cell 57*, 1069–1072.

Friguet, B., Djavadi-Ohaniance, L., & Goldberg, M. E. (1989) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 287–310, IRL Press, Oxford, New York, Tokyo.

Garel, J.-R., & Baldwin, R. L. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 3347–3351.

Gerl, M., Jaenicke, R., Smith, J. M. A., & Harrison, P. M. (1988) *Biochemistry 27*, 4089–4096.

Gerschitz, J., Rudolph, R., & Jaenicke, R. (1978) *Eur. J. Biochem. 87*, 591–599.

Girg, R., Jaenicke, R., & Rudolph, R. (1983) *Biochem. Int. 7*, 433–441.

Gō, N. (1983) *Annu. Rev. Biophys. Bioeng. 12*, 183–210.

Gō, N. (1984) *Adv. Biophys. 18*, 149–164.

Goldberg, M. E. (1985) *Trends Biochem. Sci. 10*, 388–391.

Goldenberg, D. P., & Creighton, T. E. (1984) *J. Mol. Biol. 179*, 497–526.

Goldenberg, D. P., & Creighton, T. E. (1985) *Biopolymers 24*, 167–182.

Goloubinoff, P., Gatenby, A. A., & Lorimer, G. H. (1989) *Nature 337*, 44–47.

Goto, Y., Calciano, L. J., & Fink, A. L. (1990a) *Proc. Natl. Acad. Sci. U.S.A. 87*, 573–577.

Goto, Y., Takahashi, N., & Fink, A. L. (1990b) *Biochemistry 29*, 3480–3488.

Griko, Y. V., Privalov, P. L., Sturtevant, J. M., & Venyaminov, S. Y. (1988a) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3343–3347.

Griko, Y. V., Privalov, P. L., Veynaminov, S. Y., & Kutyshenko, V. P. (1988b) *J. Mol. Biol. 202*, 127–138.

Haas, E., McWherter, C. A., & Scheraga, H. A. (1988) *Biopolymers 27*, 1–21.

Haas, I. G. (1990) *Curr. Top. Microbiol. Immunol. 167*, 71–82.

Haase-Pettingell, C., & King, J. (1988) *J. Biol. Chem. 263*, 4977–4983.

Harper, J. N., Auld, D. S., Riordan, J. F., & Vallee, B. L. (1988) *Biochemistry 27*, 219–226.

Harrison, S. C., & Durbin, R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4028–4030.

Hecht, K., Wrba, A., & Jaenicke, R. (1989) *Eur. J. Biochem. 183*, 69–74.

Hoess, A., Arthur, A. K., Wanner, G., & Fanning, E. (1988) *Bio/Technology 6*, 1214–1217.

Hollecker, M., & Creighton, T. E. (1983) *J. Mol. Biol. 168*, 409–437.

Holmgren, A., & Brändên, C. I. (1989) *Nature 342*, 248–251.

Huber, R. (1988) *Angew. Chem., Int. Ed. Engl. 27*, 79–88.

Hurtley, S. M., & Helenius, A. (1989) *Annu. Rev. Cell. Biol. 5*, 277–307.

Hvidt, A. (1983) *Annu. Rev. Biophys. Bioeng. 12*, 1–20.

Jaenicke, R. (1974) *Eur. J. Biochem. 46*, 149–155.

Jaenicke, R. (1981) *Annu. Rev. Biophys. Bioeng. 10*, 1–67.

Jaenicke, R. (1987) *Prog. Biophys. Mol. Biol. 49*, 117–237.

Jaenicke, R. (1988a) in *Protein Structure and Protein Engineering* (Huber, R., & Winnacker, E. L., Eds.) Colloquium Mosbach, Vol. 39, pp 16–36, Springer Verlag, Berlin, Heidelberg, New York.

Jaenicke, R. (1988b) *Naturwissenschaften 75*, 604–610.

Jaenicke, R. (1991) in *Advances in Life Sciences* (Jörnvall, H., Höög, J. O., & Gustavsson, A.-M., Eds.) Birkhäuser, Basel, Switzerland (in press).

Jaenicke, R., & Lauffer, M. A. (1969) *Biochemistry 8*, 3077–3092.

Jaenicke, R., & Perham, R. N. (1982) *Biochemistry 21*, 3378–3385.

Jaenicke, R., & Rudolph, R. (1986) *Methods Enzymol. 131*, 218–250.

Jaenicke, R., & Rudolph, R. (1989) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 191–223, IRL Press, Oxford, New York, Tokyo.

Jaenicke, R., & Závodszky, P. (1990) *FEBS Lett. 268*, 344–349.

Jaenicke, R., Rudolph, R., & Feingold, D. S. (1986) *Biochemistry 25*, 7283–7287.

Janin, J., Miller, S., & Chothia, C. (1988) *J. Mol. Biol. 204*, 155–164.

Jencks, W. P. (1969) *Catalysis in Chemistry and Enzymology*, p 644, McGraw-Hill, New York.

Kaden, F., Koch, I., & Selbig, J. (1990) *J. Theor. Biol. 147*, 85–100.

Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1–63.

Kellenberger, E. (1984) *Helv. Phys. Acta 57*, 188–201.

Kellis, J. T., Nyberg, K., & Fersht, A. R. (1989) *Biochemistry 28*, 4914–4922.

Kern, G., Schülke, N., Schmid, F. X., & Jaenicke, R. (1991) *J. Biol. Chem.* (submitted for publication).

Kiefhaber, T., Quaas, R., Hahn, U., & Schmid, F. X. (1990a) *Biochemistry 29*, 3053–3070.

Kiefhaber, T., Grunert, H.-P., Hahn, U., & Schmid, F. X. (1990b) *Biochemistry 29*, 6475–6480.

Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459–489.

Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631–660.

King, J., Haase, C., & Yu, M. H. (1987) in *Protein Engineering* (Oxender, D. L., & Fox, C. F., Eds.) pp 109–121, A. R. Liss Inc., New York.

Krebs, H., Schmid, F. X., & Jaenicke, R. (1983) *J. Mol. Biol. 169*, 619–635.

Kuchinke, E., & Müller-Hill, B. (1985) *EMBO J. 4*, 1067–1073.

Kundrot, C. E., & Richards, F. M. (1987) *J. Mol. Biol. 193*, 157–170.

Kundrot, C. E., & Richards, F. M. (1988) *J. Mol. Biol. 200*, 401–410.

Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87–103.

Lang, K., & Schmid, F. X. (1988) *Nature 331*, 453–455.

Lang, K., & Schmid, F. X. (1990) *J. Mol. Biol. 212*, 185–196.

Lang, K., Wrba, A., Krebs, H., Schmid, F. X., & Beintema, J. J. (1986) *FEBS Lett. 204*, 135–139.

Lang, K., Schmid, F. X., & Fischer, G. (1987) *Nature 329*, 268–270.

Laskey, R. A., Honda, B. M., Mills, A. D., & Finch, J. T. (1978) *Nature 275*, 416–420.

Lauffer, M. A. (1975) *Entropy Driven Processes in Biology*, p 264, Springer Verlag, Berlin, Heidelberg, New York.

Levitt, M. (1981) *J. Mol. Biol. 145*, 251–263.

Levitt, M. (1988) in *Protein Structure and Protein Engineering* (Huber, R., & Winnacker, E.-L., Eds.) Colloquium Mosbach, Vol. 39, pp 45–50, Springer, Berlin, Heidelberg, New York.

Levitt, M., & Sharon, R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7557–7561.

Light, A. (1985) *Biotechniques 3*, 298–306.

Lin, L.-N., Hasmui, H., & Brandts, J. F. (1988) *Biochim. Biophys. Acta 956*, 256–266.

Liu, G., Topping, T. B., & Randall, L. L. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9213–9217.

London, J., Skrzynia, C., & Goldberg, M. E. (1974) *Eur. J. Biochem. 47*, 409–415.

Luger, K., Hommel, U., Herold, M., Hofsteenge, J., & Kirschner, K. (1989) *Science 243*, 206–210.

Marqusee, S., & Baldwin, R. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8898–8902.

Marston, F. A. O. (1986) *Biochem. J. 240*, 1–12.

Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycrofft, M., & Fersht, A. R. (1990) *Nature 346*, 440–445.

Matsumura, M., Becktel, W. J., & Matthews, B. W. (1988) *Nature 334*, 406–410.

McCammon, J. A., & Harvey, S. C. (1988) *Dynamics of Proteins and Nucleic Acids*, p 234, Cambridge University Press, Cambridge, England.

Mitraki, A., & King, J. (1989) *Bio/Technology 7*, 690–697.

Morimoto, R. I., Tissiêres, A., & Georgopoulos, C., Eds. (1990) *Stress Proteins in Biology and Medicine*, Cold Spring Harbor Monograph Series 19, p 450, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Murry-Brelier, A., & Goldberg, M. E. (1988) *Biochemistry 27*, 7633–7640.

Neupert, W., & Schatz, G. (1981) *Trends Biochem. Sci. 6*, 1–4.

Nierhaus, K. (1982) *Curr. Top. Microbiol. Immunol. 97*, 82–155.

Nomura, M., & Held, W. A. (1974) in *Ribosomes* (Nomura, M., Tissiêres, A., & Lengyel, P., Eds.) pp 193–223, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Ohgushi, M., & Wada, A. (1983) *FEBS Lett. 164*, 21–24.

Okada, Y. (1986) *Adv. Biophys. 22*, 95–145.

Opitz, U. (1988) Dissertation, University of Regensburg, FRG.

Opitz, U., Rudolph, R., Jaenicke, R., Ericsson, L., & Neurath, H. (1987) *Biochemistry 26*, 1399–1406.

Ostermann, J., Horwich, A. L., Neupert, W., & Hartl, F. U. (1989) *Nature 341*, 125–130.

Pace, C. N. (1990a) *Trends Biochem. Sci. 15*, 14–17.

Pace, C. N. (1990b) *Trends Biotechnol. 8*, 93–98.

Pace, C. N., & Grimsley, G. R. (1988) *Biochemistry 27*, 3242–3246.

Park, S., Lin, G., Topping, T. B., Cover, W. H., & Randall, L. L. (1988) *Science 239*, 1033–1035.

Pelham, H. R. B. (1988) *Nature 332*, 776–777.

Pelham, H. R. B. (1989) *Annu. Rev. Cell Biol. 5*, 1–23.

Perutz, M. F., & Raidt, H. (1975) *Nature 255*, 256–259.

Popot, J.-L., Gerchman, S.-E., & Engelman, D. (1987) *J. Mol. Biol. 198*, 655–676.

Prevelige, P., Thomas, D., & King, J. (1988) *J. Mol. Biol. 202*, 743–757.

Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167–241.

Privalov, P. L. (1988) in *Protein Structure and Protein Engineering* (Huber, R., & Winnacker, E.-L., Eds.) Colloquium Mosbach, Vol. 39, p 6–15, Springer, Berlin, Heidelberg, New York.

Privalov, P. L., & Gill, S. J. (1988) *Adv. Protein Chem. 39*, 193–231.

Privalov, P. L., & Gill, S. J. (1989) *Pure Appl. Chem. 61*, 1097–1104.

Ptitsyn, O. B. (1987) *J. Protein Chem. 6*, 272–293.

Ptitsyn, O. B., Reva, B. A., & Finkelstein, A. V. (1989) *Highlights Mod. Biol. 1*, 11–17.

Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., & Razgulyaev, O. I. (1990) *FEBS Lett. 262*, 20–24.

Randall, L. L., Topping, T. B., & Hardy, S. J. S. (1990) *Science 248*, 860–863.

Rashin, A. A. (1984) *Biochemistry 23*, 5518–5519.

Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151–176.

Richards, F. M., & Vithayathil, P. J. (1959) *J. Biol. Chem. 234*, 1459–1464.

Richardson, J. S., & Richardson, D. C. (1989a) *Trends Biochem. Sci. 14*, 304–309.

Richardson, J. S., & Richardson, D. C. (1989b) in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., Ed.) pp 1–98, Plenum, New York, London.

Ristow, S. S., & Wetlaufer, D. B. (1973) *Biochem. Biophys. Res. Commun. 50*, 544–550.

Roder, H., Elöve, G. A., & Englander, S. W. (1988) *Nature 335*, 700–704.

Rooman, M. J., & Wodak, S. J. (1988) *Nature 335*, 45–49.

Rossmann, M. G., & Argos, P. (1981) *Annu. Rev. Biochem. 50*, 497–532.

Rothman, J. E. (1989) *Cell 59*, 591–601.

Rothman, J. E., & Schmid, S. I. (1986) *Cell 46*, 5–9.

Rudolph, R. (1990) in *Modern Methods in Protein and Nucleic Acid Research* (Tschesche, H., Ed.) pp 149–171, de Gruyter, Berlin.

Rudolph, R., & Jaenicke, R. (1976) *Eur. J. Biochem. 63*, 409–417.

Rudolph, R., & Fuchs, I. (1983) *Hoppe-Seyler's Z. Physiol. Chem. 364*, 813–820.

Rudolph, R., Fischer, S., & Mattes, R. (1987) Eur. Pat. Appl. EP-A 219 874.

Rudolph, R., Nesslauer, G., Siebendritt, R., Sharma, A. K., & Jaenicke, R. (1990a) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4625–4629.

Rudolph, R., Opitz, U., Kohnert, U., & Fischer, S. (1990b) Eur. Pat. Appl. EP-A 361475.

Rudolph, R., Kohler, H.-H., Kiefhaber, T., & Buchner, J. (1991) *Proc. Natl. Acad. Sci. U.S.A.* (submitted for publication).

Sauer, R. T., Jordan, S. R., & Pabo, C. O. (1990) *Adv. Protein Chem. 40*, 2–61.

Schlesinger, M. J. (1990) *J. Biol. Chem. 265*, 12111–12114.

Schmid, F. X. (1991) *Curr. Opinion Struct. Biol.* (in press).

Schmid, F. X., & Baldwin, R. L. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4764–4768.

Schmid, F. X., & Jaenicke, R. (1987) *Proc. ISSSSI R. Soc. Chem., Faraday Div. 8*, 115–120.

Schneider, R. G., Ueda, S., Alperin, J. B., Brimhall, B., & Jones, R. T. (1969) *New Engl. J. Med. 280*, 739–745.

Schülke, N., & Schmid, F. X. (1989) *J. Biol. Chem. 263*, 8827–8837.

Schulz, G. E. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 1–21.

Segawa, S.-I., & Sugihara, M. (1984) *Biopolymers 23*, 2473–2488.

Sharma, A. K., Minke-Gogl, V., Gohl, P. Siebendritt, R., Jaenicke, R., & Rudolph, R. (1990) *Eur. J. Biochem. 194*, 603–609.

Sheffield, W. P., Shore, G. C., & Randall, S. K. (1990) *J. Biol. Chem. 265*, 11069–11076.

Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., & Baldwin, R. L. (1987) *Nature 326*, 563–567.

Siebendritt, R. (1988) Dissertation, University of Regensburg, FRG.

Staley, J. P., & Kim, P. S. (1990) *Nature 344*, 685–688.

States, D. J., Dobson, C. M., Karplus, M., & Creighton, T. E. (1984) *J. Mol. Biol. 174*, 411–418.

States, D. J., Creighton, T. E., Dobson, C. M., & Karplus, M. (1987) *J. Mol. Biol. 195*, 731–739.

Stigter, D., & Dill, K. A. (1990) *Biochemistry 29*, 1262–1271.

Sturtevant, J. M. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2236–2240.

Sturtevant, J. M., Yu, M.-H., Haase-Pettingell, C., & King, J. (1989) *J. Biol. Chem. 264*, 10693–10698.

Sundaralingam, M., Sekharudu, Y. C., Yathindra, N., & Ravichandran, V. (1987) *Proteins: Struct., Funct., Genet. 2*, 64–71.

Teschner, W., & Rudolph, R. (1989) *Biochem. J. 260*, 583–587.

Tsou, C.-L. (1988) *Biochemistry 27*, 1809–1812.

Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694–699.

van Dyk, T. K., Gatenby, A. A., & La Rossa, R. A. (1989) *Nature 342*, 451–453.

Viitanen, P. V., Lubben, T. H., Reed, J., Goloubinoff, P., O'Keefe, D., & Lorimer, G. H. (1990) *Biochemistry 29*, 5665–5671.

Vita, C., Jaenicke, R., & Fontana, A. (1989) *Eur. J. Biochem. 183*, 513–518.

Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 697–701.

Wetlaufer, D. B. (1981) *Adv. Protein Chem. 34*, 61–92.

Wetlaufer, D. B. (1984) in *The Protein Folding Problem* (Wetlaufer, D. B., Ed.) pp 29–46, Westview, Boulder, CO.

Wetlaufer, D. B., & Ristow, S. (1973) *Annu. Rev. Biochem. 42*, 135–158.

Wetterau, J. R., Combs, K. A., Spinner, S. N., & Joiner, B. J. (1990) *J. Biol. Chem. 265*, 9800–9807.

Wiech, H., Stuart, R., & Zimmermann, R. (1990) *Semin. Cell Biol. 1*, 55–63.

Wodak, S. J., de Crombrugghe, M., & Janin, J. (1987) *Prog. Biophys. Mol. Biol. 49*, 29–63.

Wong, K.-P., & Tanford, C. (1973) *J. Biol. Chem. 248*, 8518–8523.

Wrba, A., Jaenicke, R., Huber, R., & Stetter, K. O. (1990a) *Eur. J. Biochem. 188*, 195–201.

Wrba, A., Schweiger, A., Schultes, V., Jaenicke, R., & Zấvodszky, P. (1990b) *Biochemistry 29*, 7584–7592.

Wright, P. E., Dyson, H.-J., & Lerner, R. A. (1988) *Biochemistry 27*, 7167–7175.

Yu, M. H., & King, J. (1988) *J. Biol. Chem. 263*, 1424–1431.

Yutani, K. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4441–4444.

Zettlmeissl, G., Rudolph, R., & Jaenicke, R. (1979) *Biochemistry 18*, 5567–5571.

Zettlmeissl, G., Teschner, W., Rudolph, R., Jaenicke, R., & Gäde, G. (1984) *Eur. J. Biochem. 143*, 401–407.

Articles

Designed Coiled-Coil Proteins: Synthesis and Spectroscopy of Two 78-Residue α-Helical Dimers[†]

Marisa Engel,[‡] Robert W. Williams,[§] and Bruce W. Erickson*[,‡]

Department of Chemistry, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3290, and Department of Biochemistry, Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814

Received September 6, 1990; Revised Manuscript Received December 5, 1990

ABSTRACT: Receptor-adhesive modular proteins are nongenetic proteins designed to contain ligand, spacer, coil, and linker modules and to interact strongly with integrins or other types of cell-surface receptors. We have designed, chemically synthesized, and characterized a 39-residue peptide chain having a 6-residue ligand module (Gly-Arg-Gly-Asp-Ser-Pro-) for adherence to Arg-Gly-Asp-binding integrin receptors, a 3-residue spacer module (-Gly-Tyr-Gly-) for flexibility, and a 30-residue coil module [-(Arg-Ile-Glu-Ala-Ile-Glu-Ala)$_4$-Arg-Cys-NH$_2$] containing four 7-residue repeats for dimerization. This chain was designed to form a 78-residue noncovalent dimer (P39) by folding the coils of two chains into an α-helical coiled coil through hydrophobic interaction of eight pairs of Ile residues. Air oxidation of P39 gave P78, a 78-residue covalent dimer having a disulfide bridge linking its C termini. Raman spectroscopy indicated that both synthetic proteins have high α-helical content. Ultraviolet circular dichroic spectroscopy indicated that both dimers contain stable α-helical coiled coils. Its C-terminal disulfide bridge renders P78 significantly more stable than P39 to thermal denaturation or denaturation by urea. The coiled coil of P39 was 30% unfolded near 55 °C and half-unfolded in 8 M urea, while that of P78 was 30% unfolded only near 85 °C. These studies have demonstrated the feasibility of using these ligand, spacer, and coil modules to construct the designed coiled-coil proteins P39 and P78, a stage in the nanometric engineering of receptor-adhesive modular proteins.

Integrins are cell-surface receptors present on cells such as fibroblasts (Ruoslahti et al., 1985; Ruoslahti & Pierschbacher, 1986, 1987) or platelets (Hynes, 1987; Phillips et al., 1988; Giltay & van Mourik, 1988) that interact with fibronectin, vitronectin, fibrinogen, or other extracellular matrix glycoproteins. Several integrins function through binding to regions containing the tripeptide segment -Arg-Gly-Asp-. These protein/cell interactions are important for the growth, differentiation, proliferation, and functional regulation of cells (Juliano, 1987). Development of synthetic molecules that could specifically and strongly inhibit an extracellular protein/cell-surface receptor interaction or promote cell/cell adhesion would allow diagnostic agents or drugs to be targeted to specific types of cells.

Receptor-Adhesive Modular Proteins. A common and fruitful approach to cell-surface targeting uses an antibody (a soluble dimeric receptor) to bind bivalently to two copies of an epitope (a cell-surface ligand). We are exploring a complementary approach that uses a *soluble dimeric ligand* to bind bivalently to two copies of a *cell-surface receptor*. A receptor-adhesive modular protein (RAMP)[1] is a nongenetic protein designed to function as a soluble dimeric ligand through the presence of specific peptide segments that serve as ligand, spacer, coil, and linker modules. For example, the 82-residue structure of the monomeric RAMP P82 (Figure 1) has two

[†] This work was supported by a graduate fellowship (to M.E.) during 1987–1989 from La Caxia (Spain), by Grant CO71A1 (to R.W.W.) from the Uniformed Services University, and by U.S. Public Health Service Grant GM 42031 (to B.W.E.) from the National Institute of General Medical Sciences.

[‡] University of North Carolina.

[§] Uniformed Services University.

[1] Abbreviations: Acm, acetamidomethyl; Boc, *tert*-butoxycarbonyl; BAW, 4:1:5 (v/v) 1-butanol/acetic acid/water; BSA, bovine serum albumin; CHO, chinese hamster ovary; CD, circular dichroism; DCM, dichloromethane; Dmb, dimethylbenzoyl; DIEA, *N*,*N*-diisopropylethylamine; DTT, dithiothreitol; *E*, volume percentage of TFE; EDTA, ethylenediaminetetraacetic acid; HPLC, high-pressure liquid chromatography; M_r, mass ratio; NMP, *N*-methyl-2-pyrrolidinone; RAMP, receptor-adhesive modular protein; RGD, Arg-Gly-Asp; *T*, temperature; TFA, trifluoroacetic acid; TFE, 2,2,2-trifluoroethanol; t_r, retention time; *U*, molar concentration of urea.

0006-2960/91/0430-3161$02.50/0