Review

Recognition between flexible protein molecules: induced and assisted folding[†]

Alexander P. Demchenko*

The Palladin Institute of Biochemistry of the Academy of Sciences of Ukraine, Kiev 252030, Ukraine and TUBITAK Marmara Research Center, 41470 Gebze-Kocaeli, Turkey

This review focuses on a very important but little understood type of molecular recognition — the recognition between highly flexible molecular structures. The formation of a specific complex in this case is a dynamic process that can occur through sequential steps of mutual conformational adaptation. This allows modulation of specificity and affinity of interaction in extremely broad ranges. The interacting partners can interact together to form a complex with entirely new properties and produce conformational signal transduction at substantial distance. We show that this type of recognition is frequent in formation of different protein–protein and protein–nucleic acid complexes. It is also characteristic for self-assembly of protein molecules from their unfolded fragments as well as for interaction of molecular chaperones with their substrates and it can be the origin of 'protein misfolding' diseases. Thermodynamic and kinetic features of this type of dynamic recognition and the principles underlying their modeling and analysis are discussed. Copyright © 2001 John Wiley & Sons, Ltd.

Keywords: molecular recognition; flexible molecules; protein–ligand interaction; protein folding; molecular chaperones; folding diseases

1. INTRODUCTION

Molecular recognition is a process by which biological macromolecules interact with each other or with smaller molecules to form a specific complex (Janin, 1995; Otlewski and Apostoluk, 1997). What distinguishes molecular recognition from other types of interactions exhibited by biological macromolecules is that it is able to distinguish highly specific from less specific binding. In order to make an interaction specific, the bonds between correct partners should be strong, while for other partners showing only minor differences in structure they should be weak or even repulsive. Another feature of molecular recognition is that the ligand concentration effects are usually nonlinear. The binding is usually cooperative and a high concentration of weakly interacting ligands cannot replace the effect of a small concentration of a specific ligand interacting with high affinity. Most important is the fact that the recognition is usually not a process in itself, but is an element of a more complex, functionally important mechanism such as allosteric regulation of enzyme activity, signal transduction, protein folding or the formation of multisubunit and supramolecular structures. This requires important and sometimes dramatic changes in the properties of interacting partners.

Physically, the strong binding of correct ligand and discrimination against incorrect ligands can be achieved by formation of rigid patterns of interacting groups of atoms on complementary molecular surfaces of the partners, which requires their pre-existing and stable conformations (Jones and Thornton, 1996; Lo Conte *et al.*, 1999). This case is easily conceivable, has a lot of analogies in our macroscopic world and although not without difficulties, is solvable by computational methods based on convenient rigid-body approximation. This mechanism does not require and does not provide conformational reorganization of partners, which is often necessary for achieving the functional result of binding.

Thus, in order to achieve cooperativity in interaction, allosteric effects and signal transduction through the molecule, a certain level of mobility should exist in the ligand–receptor system, which can propagate over substantial distance (see Table 2 in Lo Conte *et al.*, 1999). This requires the solution of a more difficult problem, namely to describe how the conformations of the partners change and to establish the mechanisms of these changes. Modeling of these processes is difficult, because the dynamics of interactions involve a much larger conformational space intrinsic to interacting partners, together with new interactions formed during the process of recognition.

There are even more difficult cases when one of the partners or both of them are very flexible or even completely disordered, and their interaction results in formation of an ordered structure in the complex. In this case the whole process of complex formation develops over a huge configurational space under the influence of a multitude of attractive and repulsive interactions acting on every amino acid residue, on hierarchial length and time scales, and

* Correspondence to: A. Demchenko, TUBITAK Marmara Research Center, 41470 Gebze-Kocaeli, Turkey. E-mail: dem@rigeb.gov.tr

probably not at equilibrium. So the problem of physical analysis in these systems becomes as difficult as the problem of protein folding (Demchenko, 2000b). Mechanical models in this case are inapplicable, while stochastic modeling is complicated and may not allow a clear visual representation of results. In order to analyze topological, thermodynamic and kinetic properties of molecular recognition in such systems, the researcher has to select a new methodology.

The aim of this review is to discuss the problem of molecular recognition in systems with high flexibility. Particular emphasis will be given to conformational adaptation in the formation of functionally important protein–protein, protein–nucleic acid and protein–small ligand complexes, to the coupling of intermolecular assembly with protein folding and of protein folding with complex formation involving molecular chaperones. We will try to derive general principles and approaches for the modeling and analysis of these systems.

2. FLEXIBILITY IN MOLECULAR RECOGNITION

There are numerous examples of proteins existing in disordered configurations that may exhibit self-organization or induced organization on complex formation with their targets (Gast et al., 1995; Yoo, 1995; Weinreb et al., 1996; Sohl et al., 1998; Lo Conte et al., 1999; Duggan et al., 1999; Wright and Dyson, 1999; Lydakis-Simantiris et al., 1999). Most of these proteins are involved in very important regulatory functions, and the lack of a particular structure in unbound form may be to their advantage since it provides larger flexibility in the interaction with different targets and rapid turnover in the cell.

2.1. Enzyme–substrate and enzyme–inhibitor interactions

Catalytic functions of many enzymes require conformational isomerization of both enzymes and substrates. The extent of these motions may be quite different — from reorientations of rigid domains and segments to disorder–order transitions. Thus, for glutathione S-transferase A 1-1 (Nieslanik et al., 1999; Stella et al., 1999) the C-terminal loop covering the active site is disordered in apo-form and becomes α-helical in the presence of glutathione conjugates. For a mutant of this protein, the increase in binding affinity is associated with dramatic decrease in rates for the C-terminal order–disorder transition associated with product release (Haley et al., 2000).

A very interesting result has been reported recently. It was found that aspartic proteinase A from yeast is inhibited by a small (68 residues) protein IA_3 which when alone in solution does not have any detectable secondary structure. However, upon formation of inhibitory complex it becomes ordered, and its segment between residues 2 and 32 adopts a nearly perfect α-helical structure. Thus, aspartic proteinase A acts as a folding template and 'folds its own inhibitor into a helix' (Li et al., 2000).

2.2. Antigen–antibody interactions

The location of protein antigenic determinants predominantly at the protein surface and at its conformationally flexible sites is well known (Morris et al., 1998). The phenomenon of conformational stabilization (Rizzo et al., 1992) and selection between different antigen conformers can be demonstrated by means of antibodies that act on a population of antigen molecules with different extents of conformational order. Thus, antibodies raised against conformational antigenic determinants of the native protein can react with denatured protein molecules by inducing their folding, at least locally (Leder et al., 1995; Berger et al., 1999). Thus, angiotensin II which is unfolded in solution adopts a regular structure on interaction with specific antibody (Murphy et al., 1993). Large conformational changes in antibodies can also accompany antigen binding, and these can alter dramatically the size, shape and charge distribution in the antigen-binding pocket (Stanfield and Wilson, 1994).

2.3. Proteins of signal-transduction systems

Calmodulin is a small protein involved in the regulation of a wide variety of intracellular processes. It was found that calmodulin-binding peptide, which is an unstructured random-coil in solution, attains an α-helical conformation on binding to native calmodulin (O'Neil et al., 1987). This means that the folded peptide conformation is induced during the dynamic recognition with the peptide binding site. In contrast, the peptide corresponding to calmodulin-binding domain of smooth muscle myosin light chain kinase interacts with calcium-saturated calmodulin in such a way that the initially helical peptide reorganizes in the complex to a more open state exhibiting a helix–coil transition (Ehrhardt et al., 1995). Thus, by criteria of amide hydrogen exchange kinetics, the local unfolding is demonstrated.

Structural comparison of two GTPase activating proteins p120 and p50 in complexes with Ras and Rho, respectively, allows the structural features responsible for their remarkable structural flexibility to be specified (Soucheta et al., 2000). Comparative studies of cyclophilin A in the free form and in complexed with cyclosporin revealed the transitions of polypeptide loops surrounding the ligand-binding site from locally flexible conformations in the free protein to well-defined spatial arrangements in the complex (Ottiger et al., 1997).

The cyclin-dependent kinase (Cdk) inhibitor p21Waf1/Cip1/Sdi1, important for p53-dependent cell cycle control, mediates G1/S arrest through inhibition of Cdks and possibly through inhibition of DNA replication. A striking disorder–order transition for p21 upon binding to one of its biological targets, Cdk2, was demonstrated (Kriwacki et al., 1996). It was shown that p21 and NH2-terminal fragments that are active as Cdk inhibitors lack stable secondary or tertiary structure in the free solution state. The p21 NH2 terminus adopts an ordered stable conformation when bound to Cdk2. This structural transition has profound implications in light of the ability of p21 to bind and inhibit a diverse family of cyclin-Cdk complexes, including cyclin A-Cdk2, cyclin E-Cdk2, and cyclin D-Cdk4.

Src-homology-2 (SH2) domains bind to proteins containing phosphorylated tyrosines, with additional specificity provided by interactions with residues C-terminal to the phosphotyrosine (pTyr) residue. While the C-terminal SH2 domain of phospholipase C-gamma 1 (PLCC SH2) interacts with eight residues of a pTyr-containing peptide from its high-affinity binding site on the beta-platelet-derived growth factor receptor, it can still bind tightly to a phosphopeptide containing only three residues. Certain regions of the PLCC SH2 domain contacting the residues C-terminal to the pTyr have a high degree of mobility in both the free and peptide bound states (Kay et al., 1996, 1998). In contrast, there is significant restriction of motion in the pTyr binding site. These results suggest a correlation between the dynamic behaviour of certain groups in the PLCC SH2 complex and their contribution to high affinity binding and binding specificity (Kay et al., 1996, 1998).

FlgM is the inhibitor of sigma 28, a transcription factor specific for the expression of bacterial flagella and chemotaxis genes. During flagella filament assembly it is also exported from the cytoplasm to the outside of the cell. It was reported that FlgM is mostly unfolded, but about 50% becomes structured when bound to sigma 28 (Daughdrill et al., 1997). In solution the C-terminal part of FlgM lacks sufficient intramolecular contacts to form stable secondary and tertiary structures, and on binding to sigma 28 this structure is stabilized.

2.4. Flexible proteins in macromolecular assembly

In the course of the cell cycle, microtubules exhibit dramatic reorganization, which is modulated by a number of stabilizing and destabilizing factors (Andersen, 2000). Assembly and stability of microtubules is supported by stabilizing factors, the most important of which are tau and MAP2. When studied in solution, these proteins exhibit low helical content, no temperature-dependent unfolding transitions and a high sensitivity to proteases (Hernandez et al., 1986), which suggests their disordered conformation. The ordering of these proteins occurs on interaction with tubulin filaments, and both interaction partners acquire new properties. Thus the binding to microtubules becomes highly cooperative, while the isolated microtubule-binding segments do not display any cooperativity (Coffey and Purich, 1995). Through their high conformational flexibility in the unbound state (Friedhoff et al., 2000) these proteins are involved in pathological conditions such as amyloidosis (see Section 5.2).

Among microtubule-destabilizing factors, the most important is phosphoprotein OP18/stathmin (Larsson et al., 1999; Martin et al., 2000). This protein does not have a defined three-dimensional structure, although it contains three distant regions of sequence with different helix propensities (Wallon et al., 2000). The separated segments can bind to tubulin filaments but are not able to produce the functional effect that modulates the dynamics of microtubules. The binding to tubulin occurs by multiple, physically distinct, but cooperative interaction sites (Larsson et al., 1999). The binding of stathmin to tubulin modulates the binding of GTP to tubulin as a consequence of a conformational change in the β-tubulin subunit (Moreno et al., 1999). Thus, we observe that on interaction with microtubule modulating factors the tubulin molecules and their ordered filaments acquire new properties.

Caldesmon is a major F-actin binding protein in smooth muscle, which is also involved in assembly of myosin rods (Katayama et al., 1995). It was reported that in solution it has a highly extended flexible conformation devoid of secondary structure (Lynch et al., 1987).

2.5. Proteins interacting with nucleic acids

Protein–nucleic acid complexes possess both stability and specificity: the binding constants under typical ionic conditions range from 10^9 to $10^{12}\,M^{-1}$ or higher, and the ratios of specific to nonspecific binding constants range from 10^3 to 10^7 (Spolar and Record, 1994). The phosphodiester backbone is uniformly charged, and non-specific binding can be easily achieved electrostatically. Sequence-specific recognition occurs through hydrogen bonding interaction with the bases, which modifies DNA and RNA flexibility making these molecules more rigid and bent to a strained conformation. Conformational changes observed in various proteins interacting with nucleic acids include quaternary rearrangement of domains or subunits, ordering of disordered loops or N-terminal segments and formation of α-helices and β-hairpins from the unfolded structure in the free state (Hard, 1999).

Most of the folding or ordering transitions in proteins interacting with DNA occur on complementary surfaces of double helical structure, which comprise the binding site. Often this recognition occurs by insertion of an α-helical segment (helix–coil–helix) into the major groove, and this allows the recognition of a specific sequence by direct interaction with the bases. These proteins commonly contain the characteristic 'leucine zipper' motif as well as basic domains. The yeast transcription activator GCN4 provides a well-studied example of bZIP recognition, wherein B-DNA serves essentially as a template for protein folding (Weiss et al., 1990; O'Neil et al., 1991; Ellenberger et al., 1992; Berger et al., 1996; Benevides et al., 2000). Crystallographic data demonstrate that the basic region of GCN4 protein and its fragments containing the leucine zipper and basic segments are highly disordered, and that the interaction with DNA induces their fully helical conformation (Weiss et al., 1990; O'Neil et al., 1991). This results in highly specific binding with the DNA recognition site (Konig and Richmond, 1993; Benevides et al., 2000). Non-specific binding does not produce this ordering effect.

The SKN-1 transcription factor of Caenorhabditis elegans binds DNA with a high affinity as a monomer, by means of a basic region like that of basic-leucine zipper (bZIP) proteins, which bind DNA only as dimers. A flexible arm at the Skn domain amino terminus binds in the minor groove, while a support segment adjacent to the carboxy-terminal basic region stabilizes independently the basic region-DNA binding. Without DNA the basic region and arm are unfolded and the support segment forms a loose conformation (molten globule) of four α-helices. On binding DNA, the Skn domain adopts a tertiary structure in

which the basic region helix extends directly from a support segment α-helix, which is required for binding. The remainder of the support segment anchors this uninterrupted helix on DNA, but leaves the basic region exposed in the major groove. This is similar to the way the bZIP basic region extends from the leucine zipper, indicating that positioning and cooperative stability provided by helix extension are conserved mechanisms that promote binding of basic regions to DNA (Carroll *et al.*, 1997).

A flexible segment is responsible for the interaction of the λ repressor with DNA (Clarke *et al.*, 1991), and molecular dynamics simulation of this protein suggests the existence of two dynamic sub-states (Kombo *et al.*, 2000). Monomer–dimer equilibrium in phage λ Cro repressor is interesting: there are no folded monomers in the free state and only unfolded monomers and folded dimers exist in equilibrium. However, on interacting with DNA, this protein can attain a folded monomeric conformation (Jana *et al.*, 1997).

Unstructured elements of polypeptide sequence participate also in ligand binding by *Escherichia coli* CytR regulator (Jorgensen *et al.*, 1998; Gavigan *et al.*, 1999). The nucleotide excission repair occurs by insertion of a flexible β-hairpin between two DNA strands (Theis *et al.*, 1999; Ikegami *et al.*, 1999).

The *trp* repressor is an important regulatory protein in *Escherichia coli*, which, when activated by its cofactor, L-tryptophan, controls the uptake and synthesis of tryptophan. The relative disorder of its DNA-binding domain and its stabilization on interaction with L-tryptophan is recorded by a variety of experimental methods (Zhao *et al.*, 1993; Zhang *et al.*, 1994; Gryk *et al.*, 1995). A point mutation in the DNA binding domain which increases the affinity towards DNA and decreases the domain mobility limits the effectiveness of *trp* repressor to interact with some operators (Gryk *et al.*, 1996).

Studies of transcription factor Ets-1 provide evidence that local protein unfolding (in contrast to more common folding) can also accompany DNA binding (Petersen *et al.*, 1995). Circular dichroism and partial proteolysis showed that the secondary structure of the Ets-1 DNA-binding domain is unchanged in the presence of DNA. In contrast, DNA allosterically induced the unfolding of an α-helix that lies within a flanking region involved in the negative regulation of DNA binding. These findings suggest a structural basis for the intramolecular inhibition of DNA binding and a mechanism for the cooperative partnerships that are common features of many eukaryotic transcription factors.

The co-factors in interaction of transcription factors with DNA can also display disorder–order transition on interaction with their partners (Wendt *et al.*, 1998). When they interact with transcription factors they can induce their ordering (Hua *et al.*, 1998).

In the case of Eco RV endonuclease, in addition to changes of quaternary structure, the folding of two short loops is coupled to DNA binding (Winkler *et al.*, 1993). One loop (68–71) which contacts with minor groove and the sugar-phosphate backbone, leads to a β-turn following both specific and non-specific binding, although the protein is poorly ordered in the free state. Another loop (182–187) interacting specifically with the base pairs in the major groove remains disordered in non-specific complexes. It is interesting that the total contact area is larger in the case of non-specific compared to specific binding to DNA. When the binding with DNA is specific, bending of the DNA helix occurs (Horton and Perona, 1998).

RNA recognition differs from DNA recognition, which is connected with the different functions of the formed complexes (Draper, 1995, 1999; Varani, 1997; Frankel and Smith, 1998). Two main themes appear in protein–RNA interactions. A 'groove binder' class of proteins places a protein structure (α-helix, 3–10-helix, β-ribbon, or irregular loop) in the groove of an RNA helix, recognizing both the specific sequence of bases and the shape or dimensions of the groove, which is sometimes distorted from the normal A-form. A second class of proteins uses β-sheet surfaces to create pockets that recognize the bases of single-stranded RNA. Some of these proteins recognize completely unstructured RNA and, in others, RNA secondary structure indirectly promotes binding by constraining bases in an appropriate orientation.

What is common in DNA and RNA binding is the involvement of dynamic interactions with the participation of flexible elements of structure. Formation of complexes with proteins can order the flexible single-stranded RNA loops (Allain *et al.*, 1996), and also highly flexible protein segments become structured on this interaction (Markus *et al.*, 1997; Nanduri *et al.*, 1998). Thus, the interaction between ribosomal protein L25 with a fragment of 5S rRNA (Stoldt *et al.*, 1999) demonstrates two types of recognition, preformed and induced. In the latter case a flexible loop converts to α-helix. Another ribosomal protein, L11, possesses an extended unstructured loop, which becomes structured on RNA binding (Markus *et al.*, 1997).

A remarkable case is U1A protein, which binds, very tightly ($K_d \approx 10^{-11}$ M), an RNA hairpin during splicing, despite the presence of a very small interface area. The binding involves a disorder–order transition in the loop on the template of β-sheeted protein structure (Allain *et al.*, 1996).

Protein kinase PKR is activated by a double-stranded RNA (Nanduri *et al.*, 1998). It was found that a highly flexible interdomain linker enables two domains to wrap around the RNA duplex for cooperative and high-affinity binding. This leads to an overall change of PKR conformation resulting in its activation.

Thus, high flexibility of at least one of the interacting partners in protein–nucleic acid recognition and often of both of them is necessary for highly selective functional events. The results of many experiments demonstrate, however, that there is no need for whole molecules to be unfolded, a high mobility being required only for its functional/recognition part. In coupling with local folding a high specificity can be realized where specific local sequences serve as templates for folding transition, while nonspecific sequences do not. Since both the driving force (binding free energy) and the driven process (binding and folding) are a function of DNA sequence, the final conformation of the complex may induce a highly selective function of this DNA sequence.

3. SELF-ASSEMBLY OF NATIVE PROTEIN BY ASSOCIATION OF ITS DISORDERED CHAINS: AN EXTREME CASE OF MOLECULAR RECOGNITION BETWEEN FLEXIBLE STRUCTURES

Protein folding is the process of attaining a unique structure in an enormous configurational space driven by the search for global free energy minimum. In order to achieve this minimum, the flexible molecule has to try a huge variety of configurations between partially ordered elements of structure. If the polypeptide chain is fragmented, can these fragments associate to form the folded structure? Can we, by proper manipulation with protein chains and their fragments, observe the coupling of protein folding with intermolecular recognition between flexible and self-organizing elements of this macromolecular structure? This section will try to answer these questions.

3.1. Folding of proteins cut into fragments

The folding of protein from its separated unstructured fragments may present a unique model for dynamic recognition. This field of research started to be explored already in the early seventies, and the reader is referred to excellent reviews covering the early research (Anfinsen, 1973; Zabin and Villarejo, 1975; Wetlaufer, 1981; Taniuchi et al., 1986). At that time the studies were focused primarily on the implications of structure and functional activity determination in general terms, and much less attention was paid to local structural and kinetic effects. Results of more recent experiments (Prat-Gay, 1996) involve detailed structural analysis, information on folding kinetics and exploration of genetically engineered protein forms.

Limited proteolysis of ribonuclease A, a molecule composed of 124 amino acids, cuts a single peptide bond, which results in the appearance of an N-terminal fragment of 20 residues, the S-peptide. The latter has no regular structure, but on its interaction with the rest of molecule (S-protein), with all four of its disulfide bonds being intact, a complex is formed in which the S-protein regular structure induces the S-peptide to adopt an α-helix conformation (Labhardt, 1982, 1984). The rate-limiting step of this process is the formation of α-helix (Goldberg and Baldwin, 1999). The S-protein, if denatured with reduced disulfide bonds, cannot fold correctly without the help of S-peptide, but the folding can be readily achieved when both unfolded S-protein and S-peptide are mixed in solution in folding conditions.

Staphylococcal nuclease consists of a single polypeptide chain of 149 residues containing no disulfide bonds. Association of two or three of its long fragments (Andria et al., 1971; Light et al., 1974), including those with overlapping sequence (Taniuchi and Anfinsen, 1971; Taniuchi et al., 1977), results in restoration of native structure and of some level of activity. A common feature of formation of all these complexes is the first-order kinetics, which suggests that the rate-limiting step is not the diffusion of components but the reorganization of their complex (Light et al., 1974).

Cytochrome c, a 104-residue heme-containing protein, was also explored as a model for the assembly of a protein from its fragments. Its molecules assembled from two or three chain fragments resemble closely the intact native protein in structure and biological activity (Hantgan and Taniuchi, 1977; Juillerat et al., 1980; Juillerat and Taniuchi, 1982), although assembled protein has a decreased temperature stability.

Thioredoxin, a small α/β domain protein, can fold together out of its fragments (Reutimann et al., 1983). Reassembly is possible not only after a cleavage inside an exposed loop but also in an α-helix (Yang et al., 1999). Although both sets of fragments produce native-like complexes, there are clear differences between them in interface geometry, stability of the folded state and mechanism of association/folding (Chaffotte et al., 1997; Ghoshal et al., 1999).

Barnase is a small (110 residues) bacterial ribonuclease free from disulfide bonds. A peptide corresponding to barnase residues 1–22 which contains in its native structure the major α-helix (residues 6–18) binds rapidly to the complementary peptide (residues 23–110) containing a β-sheet to form a catalytically active complex with near-native structure and properties (Kippen and Fersht, 1995). Both fragments, when they are separate, appear to be disordered, and become structured as a result of their association. A set of mutations in the helical region (1–22) were introduced in order to destabilize the α-helix. The mechanism of assembly of the peptides was investigated by analyzing the kinetics and equilibria of association of mutants. The reaction was found to follow second-order kinetics. Virtually the entire change in stability of the complex on mutation was reflected in changes in the association rate constant, while the dissociation rate constant was very little affected. The complexes formed by all preformed mutant peptides (1–22) with (23–110) were only 10% active. It was found that the noncovalent complex was destabilized less by mutations in the α-helix than is the intact protein.

Two fragments (20–59) and (60–83) of chymotrypsin inhibitor-2 from barley seeds associate to give a native-like structure (Ruiz-Sanz et al., 1980; Mohana-Borges et al., 1999). The kinetics and equilibria of association of mutant fragments derived from cleaving mutant proteins were analyzed. The changes in free energy of association have been measured both by isothermal studies of the binding of fragments and by thermal denaturation of the complexes. In general, there is a good correlation between the changes of free energy of association of fragments following mutation and the changes in free energy of folding of the parent protein. The second-order rate constants for the major phase of association change with mutation. The rate constants for association correlate well with the rate constants of refolding of the respective intact proteins.

These earlier results were extended recently to other fragments of chymotrypsin inhibitor-2 (Prat-Gay, 1996). For the protein assembled from fragments, both crystal and NMR structures have been solved, and they were found to be similar to that of intact protein except for the cleaved loop and its closely neighbouring groups. Kinetics of association of fragments demonstrate cooperative and simultaneous formation of secondary and tertiary structures, as in the intact protein. By application of high pressure the

protein assembled from fragments can be transferred to the denatured form, in which the fragments are still bound together (Mohana-Borges *et al.*, 1999).

Subdomain-size proteolytic fragments of *trp* repressor from *Escherichia coli* have been produced. They assemble in defined order to regenerate fully native dimers (Tasayco and Carey, 1992). By characterization of the secondary and tertiary structures of isolated and recombined fragments, the structure of assembly intermediates can be correlated with the kinetic folding pathway of the intact repressor deduced from spectroscopic measurement of folding rates. The native-like structure of these intermediates provides further evidence that protein folding pathways involve the acquisition of stability of secondary structure units and assemblies that are found in the native state.

It is known that many proteins are synthesized together with N-terminal pro-peptides which are necessary for correct folding. It was shown that the refolding of a calcium-free mutant subtilisin BPN′ is readily catalyzed by the isolated pro-peptide (Strausberg *et al.*, 1993). The pro-peptide is unstructured at the start of the complexation-folding reaction, and the rate-limiting step in this case is the formation of initial collision complex. Once formed, this complex results in a rapid isomerization to the fully folded structure.

Thus, we can conclude that, even in cases when the peptide fragment exhibits no or very limited ordered structure in aqueous solution, its latent structure determined by the protein folding code may be realized upon interaction with complementary fragments derived from the same protein molecule. This is achieved by the specificity of the intermolecular process of recognition between flexible structures.

3.2. Interaction of folding protein with its own fragments

Protein fragments, independently of their possession of regular elements of structure, are usually more flexible than their parent folded proteins. When the protein folds, its structural fragments get together, recognize each other by complementary pattern of non-covalent bonds and condense into a protein globule or domain. One can intervene in this process by adding the isolated protein fragments to the solution of intact protein undergoing the folding reaction. The fragments may interact with complementary segments of the intact chain and interfere with the normal process of intramolecular recognition in the folding pathway by formation of non-native intermolecular complexes. Experimentally we can then observe the retardation of the folding reaction and/or imperfect folding with the decrease of activity of the folded enzyme. Experiments performed on dihydrofolate reductase demonstrated clearly the effect of decreased activity of the protein folded in the presence of its fragments (Hall and Frieden, 1989). It is interesting that this effect was observed only in conditions when the folding rate was relatively slow. Among a number of tested peptides only three (belonging to one chain segment) were effective, and the higher effect was observed for the shorter peptide. Thus, in this unnatural system the externally added peptides may be recognized as intrinsic elements of the folding protein structure and become trapped.

3.3. Intertwined dimers: self-assembly precedes folding

There are dimeric proteins which cannot exist as monomers in the same conformation because of high level of integration between subunits. They cannot be assembled from already folded subunits, and the question arises regarding the steps of their folding and assembly. An example of such proteins is the *trp*-repressor, a dimer of two identical chains of 108 residues and six α-helices each, in which the two chains are intertwined. The kinetics of its folding-assembly is very complicated and has been interpreted to involve three parallel channels with multiple folding and isomerization reactions. In order to resolve it, a polypeptide corresponding to the core/dimerization domain was constructed (Gloss and Matthews, 1998). Kinetic properties of its folding showed that the second-order rate constant for the association reaction approaches that of the diffusion limit. The dimeric structure is formed via a dimeric intermediate, in other words the formation of secondary and tertiary structure is concurrent with or precedes dimerization.

Another example of coupling folding and assembly is the cell cycle regulatory protein sucl, which can exist in two thermodynamically stable forms — as a monomer and as a dimer. The dimer is a structure with intertwined chains belonging to the monomers. On folding to dimer, it forms quite specific secondary structure that cannot be formed by association of already folded monomers (Endicott *et al.*, 1995; Bourne *et al.*, 1995).

Recognition coupled with folding and recognition between folded structures are present together in the assembly of the small tetrameric protein tumor suppressor p53 (Mateu *et al.*, 1999). Unfolded monomers form intertwined dimers, which associate to produce functional tetramers.

Protein folding by association of subunits is not limited to these spectacular cases. In titrations by chemical denaturants it is often observed that the only stable species detected are unfolded monomers and folded oligomers with no folded monomeric intermediates (Bowie and Sauer, 1989; Barry and Matthews, 1999). This process can even be modeled by rather small peptides such as melittin, which exhibits a concentration-dependent equilibrium between disordered monomer and α-helical tetramer (Spolar and Record, 1994).

It can be concluded that the protein folding is a transformation of linear information encoded in amino acid sequence into a well-determined three-dimensional structure. On the folding pathway numerous recognition steps resulting in the formation of elements of regular structure should be involved. The protein folding code which is a rule for this transformation is not local but is distributed over the whole sequence (Demchenko and Chinarov, 1999). As a result the cutting of protein sequence into long segments does not produce a substantial damage and, the segments may self-assemble and fold into the native structure, as does the intact chain. However, in the case of associating fragments the process of folding involves additional processes of diffusion, encounter and complex-formation of distantly located elements of structure. Therefore, this system may serve as a powerful model for studying

molecular recognition between flexible structures, a process which is only partly understood.

4. RECOGNITION BETWEEN UNFOLDED PROTEINS AND MOLECULAR CHAPERONES

Molecular chaperones is the common name of a diverse family of proteins which help other proteins to fold, refold after transport through biological membrane or maintain the folded state under conditions of stress. Different chaperones only partially substitute each other; they interact with different families of substrate proteins and on different steps of their folding. They differ in a number of properties, and their common feature is that they interact only with unfolded or partially folded but not with native proteins, while the formed complex can dissociate into chaperone and folded protein molecule. The chaperones are not enzymes in a strict sense, they do not accelerate the folding, but substantially suppress the appearance of misfolded and aggregated protein states (Demchenko, 1999). Many chaperones are ATPases, and the suggested general role of ATP hydrolysis is the control on the dissociation of productive chaperone-substrate complexes. At present there is no clear understanding how chaperones help proteins to fold. In the author's 'fettered folding model' (Demchenko, 1999, 2000a) the temporal occlusion of particular hydrophobic sites in polypeptide sequence reduces greatly the conformational space for folding. This allows the self-assembly of structure to proceed with the formation of folding determinants at sites that are free from chaperone protection. After the chain is released from the complex, a collapse of structure with formation of intramolecular hydrophobic bonding occurs.

The diversity of chaperones, their broad but well-determined substrate specificity and the fact that a great variety of newly synthesized and functioning proteins with very dissimilar structures interact with chaperones suggest the absence of highly specific protein recognition features. This is in contrast to other examples of 'flexible' recognition.

4.1. Chaperonins (Hsp60) bind unfolded protein substrates in a caged structure

A typical representative of this family is the GroEL/GroES protein complex. It is a 14-meric particle of 840 kDa forming a porous cylinder with an internal cavity of about 6 nm in diameter, which can accommodate proteins of a size of about 60 kDa. Each of the GroEL 60 kDa subunits forming seven-membered rings consists of three domains. The large equatorial domains associate to produce the central core of the structure possessing maximum contacts within the ring and between the two rings. They contain the nucleotide binding (ATPase) site adjacent to the junction with the small intermediate domain. The binding of both non-native substrates and of co-chaperone GroES occurs to the apical domain, which is able to move as a rigid body adopting different orientations and probably possessing

some intrinsic flexibility (Braig et al., 1995). In contrast to other chaperones, GroEL is able to undergo muti-point binding of substrate proteins by seven subunits simultaneously (Farr et al., 2000). It has been suggested that its ability to re-fold the misfolded proteins lies in an ATP-dependent (Von Germar et al., 1999) stretching of the complex (Hammarstrom et al., 2000) with the generation of force (Shtilerman et al., 1999).

Hsp60s can assist the folding of an extended variety of proteins (Houry et al., 1999). Their binding ability probably excludes only initial and final states in folding kinetics, i.e. totally unfolded polypeptide chains and already folded native proteins. In some cases, when the binding with native proteins is recorded, the bound forms are probably the pre-existing non-native conformations, presumably late folding intermediates, rather than the native states. The forms of bound protein studied at equilibrium demonstrate intermediate properties between unfolded and native states. They may possess elements of secondary structure such as α-helices (Landry and Gierasch, 1991; Chatellier et al., 1999; Preuss and Miller 1999) and β-sheets (Chatellier et al., 1999). The critical factors for binding appear to be the features of early folding intermediates, such as clustering and exposure of hydrophobic residues (Preuss and Miller, 1999; Wang et al., 1999a,b) rather than a defined secondary structure motif (Hendrick and Hartl, 1993). In addition to low-selective hydrophobic interactions (Lin et al., 1995), electrostatic interactions probably also play a role in the recognition. Being acidic, GroEL interacts more strongly with basic peptides and proteins (Itzhaki et al., 1995; Lau and Churchich, 1999). Long-range ionic interactions may participate in the initial steps of recognition before hydrophobic binding occurs, and this can allow fast association and slow dissociation of unfolded peptide (Perrett et al., 1997).

An elegant model to study substrate specificity of GroEL has been introduced recently (Chatellier et al., 1999). It uses the isolated apical domain of GroEL monomer (minichaperone) immobilized on a solid support. Using this model it was found that the side-chains of the recognized peptide do not have to be totally hydrophobic, and that polar and positively charged chains can also be accommodated. Further, the spatial distribution of the side-chains is compatible with that in an α-helix. This implies that GroEL can bind a wide range of structures, from extended β-strands and α-helices to folded states with exposed hydrophobic side-chains (Wang et al., 1999a). The binding site can accommodate substrates of approximately 18 residues in a helical or seven residues in an extended conformation. These data support the existence of two GroEL functions: the ability to retain sticky intermediates by binding many motifs and an unfolding activity by binding an extended sequential conformation of the substrate.

The question whether in addition to these very general regularities observed for all proteins which are substrates of GroEL, there exist specific recognition sites (chaperonin recognition determinants), which are transiently exposed and then hidden into protein interior in the native state, was discussed (Preuss and Miller, 1999). In favor of this possibility are the results on chaperonin binding to substrate protein mutants. Thus, a mutant of maltose binding protein is completely arrested by interaction with GroEL, while

wild-type protein can fold successfully (Sparrer *et al.*, 1996). Against this suggestion is the fact that GroEL can bind artificial proteins with random sequences (Aoki *et al.*, 2000). It may be possible that the chaperonin recognizes somehow the physical state of the protein molecule known as 'highly flexible intermediate' or 'molted globule' (Buckle *et al.*, 1997). However, the latter mechanism by now has very little experimental support.

How can a compact but flexible substrate protein structure interact dynamically with GroEL recognition sites allowing optimal induced fit (Chen and Sigler, 1999)? Experiments demonstrate that during this interaction the already existing secondary structure is partially destabilized (Zahn *et al.*, 1994; Preuss and Miller, 1999). ATP binding and hydrolysis induce conformational changes in GroEL (Galan *et al.*, 1999), which depend on binding of protein or polypeptide substrate (Mendoza and Campo, 1996). This demonstrates the existence of communication between two structurally remote functional sites.

Thus, chaperonin GroEL binds a variety of polypeptides that share no obvious sequence similarity. The precise structural, chemical and dynamic features that are recognized remain largely unknown. The GroEL structure combines both the rigidity necessary for formation of oligomeric structure and performing control on ATP binding and hydrolysis and the flexibility important for binding different unfolded proteins.

4.2. Hsp70 chaperones — monomeric molecules binding single polypeptide strands

The functional role of Hsp70 chaperones is associated with the binding and release of segments of polypeptide chains belonging to unorganized random-coil proteins and their subunits, which is necessary to prevent nonspecific aggregation already during initial steps of protein folding. Hsp70 proteins are ATPases activated by the binding of peptide substrate. The most extensively studied representative of Hsp70 group is bacterial DnaK. For this protein it was established that two functional entities, the ATPase domain and the peptide-binding domain, are structurally separated. The N-terminal, ~45 kDa, ATPase domain is followed by a C-terminal ~18 kDa domain, which contains the peptide-binding site (Demchenko, 1999).

The structure of peptide-binding domain of DnaK in the complex with bound synthetic heptapeptide has been resolved by X-ray crystallography (Zhu *et al.*, 1996) and by NMR in solution (Wang *et al.*, 1998; Morshauser *et al.*, 1999; Pellecchia *et al.*, 2000). This structure consists of two sub-domains, one represented by β-sandwich and the other (C-terminal part) by a sequence of α-helices. The substrate peptide in stretched conformation is bound in the site formed by the loops of β-structure and interacts with it via seven hydrogen bonds of the main chain and numerous other main-chain and side-chain interactions. Such a configuration suggests that the recognition of the binding site on the chain is local, and that chaperone binding occludes all interchain interactions with other peptide segments. The chaperone α-helical sub-domain may occupy two positions. One of them is on top of the structure of the substrate-biding domain, covering the binding site but without contact with the bound peptide. The other is an extended conformation, which allows the peptide-binding site to open. This suggests that conformational changes should be associated with protein functioning, at least during the steps of peptide binding and release. These changes were recorded by a variety of physical methods (Slepenkov and Witt, 1998) and many attempts were made to couple the cycle of substrate binding-release with the ATPase cycle (McCarty *et al.*, 1995; Pierpaoli *et al.*, 1997; Farr *et al.*, 1998; Gisler *et al.*, 1998; Demchenko, 1999).

The key difference in substrate specificity between the Hsp70 and Hsp60 systems is that, while Hsp60 is an oligomeric complex which allows multipoint interaction with the substrate, Hsp70 is usually monomeric and can interact with unfolded chains only at one site and can occlude only a short segment of the sequence. Unlike Hsp60 chaperonins, Hsp70s cannot recognize the globular structures and therefore cannot perform 'proofreading' and correction of misfolded forms on a globular level. It should operate on a smaller scale i.e. in the binding of short sequences.

The experimental data are in accord with this specific feature of Hsp70. The frequency of appearance of DnaK binding sites in different proteins is on average every 36 residues. The affinities towards these sites are very high (K_d as low as 100 nM), and 'DnaK is thus capable of distinguishing, though not exclusively, secondary structure elements by recognition of primary structure' (Rudiger *et al.*, 1997a). The elements of structure that are strongly favored for binding are the hydrophobic β-strands, whereas those which are strongly avoided are the amphiphilic α-helices. This preference is completely different from that observed for GroEL (Landry *et al.*, 1992). Thus, the structures which frequently appear on the surface of the folded protein are avoided from Hsp70 binding, while those which in the native structure have to form a hydrophobic nucleus are favored. In line with this suggestion are the results of comparative studies of DnaK binding to peptides of different sizes and sequences (Fourie *et al.*, 1994; Gragerov *et al.*, 1994; Rudiger *et al.*, 1997a, b). The sequences containing short (four to five residues in length) segments of hydrophobic residues surrounded by basic residues one preferentially bound (Rudiger *et al.*, 1997a). A multistep mechanism of interaction of unfolded peptide with Hsp70 was suggested (Mayer *et al.*, 2000). It involves the concept of mechanistic motion of rigid Hsp70 segments together with a conformational adaptation of the flexible protein substrate.

4.3 Hsp90 chaperones with a specificity determined by functional rather than structural properties of substrate

Hsp90 is one of the most abundant proteins in the cytosol of both prokaryotic and eukaryotic cells. In addition to a general chaperone effect, it has a specific role in the folding of a range of signal transduction molecules, including steroid hormone receptors and protein kinases (Csermely *et al.*, 1998; Caplan, 1999; Buchner, 1999). Hsp70 can bind and hydrolyze ATP, but its ATP-hydrolyzing activity is not obligatory (Prodromou *et al.*, 1997). Binding affinity to

different peptide substrates has not been studied in detail, and existing data (Csermely *et al.*, 1997) correlate this affinity with both hydrophobicity and positive charges of peptide substrates. These discriminative features are similar to those of other chaperones. However, many Hsp90 substrates retain high flexibility after folding.

One of the interaction partners of Hsp90 is co-chaperone p23, which binds to Hsp90 in its ATP-bound state and, on its own, interacts specifically with non-native proteins performing the chaperone function. It contains an unstructured region that maps to the C-terminal part of the protein sequence (Weikl *et al.*, 1999). This unstructured region of p23 is not necessary for interaction with Hsp90, but important for the ability to bind non-native proteins and to prevent their non-specific aggregation. The isolated C-terminal region itself is unable to act as a chaperone or to complement truncated p23 by addition of this peptide. These results imply that different sites with different flexibility of interaction may coexist within one protein molecule, and that for efficient interaction of p23 with non-native proteins, both the folded domain and the C-terminal unstructured region are equally important.

4.4 SecB binds the unfolded substrates with diffusion-limited rates

SecB is a bacterial chaperone with the apparent function to bind unfolded proteins, which are to be transported across the membranes, and to keep them in a transport-competent conformation (Randall, 1992; Randall and Hardy, 1995). SecB is a tetramer made up of 17 kDa subunits. It binds to and promotes the export of periplasmic and outer membrane proteins by stabilizing the unfolded or loosely folded precursor conformation and thus preventing premature folding or aggregation. Since the proteins designed for export cannot translocate across the membrane in the native form, binding chaperones and keeping them in an unfolded transport-competent form is a necessary step in their production in the cell. A distinguishing feature of SecB interaction with unfolded protein substrates is an extremely high association rate, which approaches the limit set by diffusion of interacting partners, $\sim 10^9\,\mathrm{M}^{-1}\mathrm{s}^{-1}$.

SecB is highly acidic with a p$K \approx 4$, which allows the electrostatic recognition of positively charged sites in the polypeptide sequence. Specific interactions of SecB with different signal sequences of precursor protein forms have been shown (Watanabe and Blobell, 1995), and this may be an important factor which determines recognition of the target protein. It is essential also that, being a tetramer, SecB can interact simultaneously with four small-size peptides, but only with one unfolded protein chain. This suggests a cooperation between subunits.

The mechanism by which SecB recognizes secretory proteins and eliminates cytosolic proteins is poorly understood. To identify its binding motif, 2688 peptides covering the sequences of 23 proteins were recently screened for SecB binding (Knoblauch *et al.*, 1999). The motif was found to be approximately nine residues long and was enriched in aromatic and basic residues, whereas acidic residues were disfavored. Its identification required the search for binding regions among different proteins. Surprisingly, it was found

that SecB lacks specificity toward signal sequences. In unfolded proteins, SecB-binding regions occur statistically every 20–30 residues. The occurrence and affinity of binding regions are similar in SecB-dependent and -independent secretory proteins and in cytosolic proteins. According to these data, SecB cannot differentiate between secretory and non-secretory proteins via its binding specificity.

The analysis of equilibrium and kinetic regularities of substrate binding suggested a multistep binding mechanism (Randall, 1992), according to which the primary interaction triggers a conformational change in SecB, which leads to exposure of new hydrophobic binding sites resulting in much stronger binding. There are two types of binding sites on SecB that interact with different regions of a polypeptide ligand (Randall *et al.*, 1998). One type in the flexible regions can interact electrostatically and is involved in the initial recognition, while the other interacting with hydrophobic areas becomes exposed at a latter stage as a result of subsequent conformational change. Due to multiple binding, concerted functioning of several sites interacting with one ligand and large contact area, a high-affinity binding is selective for non-native ligands without high specificity at any ligand binding site (Randall and Hardy, 1995).

It was reported that flexible contact area of SecB interacting with the flexible segments of substrate protein induces the latter to adopt a β-sheet structure (Fasman *et al.*, 1995). Recording the kinetics of CD spectra shows that the rate of initial binding approaches the encounter limit. This allows a kinetic partitioning of polypeptides, which exhibit-slow spontaneous folding into complexes with higher affinity, while the forms exhibiting fast spontaneous folding are kinetically eliminated. This mechanism, which involves the transition from low-affinity to high-affinity binding, requires a conformational change in the chaperone, and the latter was observed directly (Randall, 1992).

4.5. Proteasomes and foldases with chaperone activity

Two types of enzymes are actively involved in the process of protein folding — protein disulfide isomerase and peptidyl prolyl *cis–trans* isomerase (PPI). Their common name 'foldases' indicates that they catalyze the formation of correct disulfide bonds and proline isomers when these reactions limit the protein folding process. The substrate specificity of these enzymes is very broad and they interact with folding intermediates of variable conformation (Wang and Tsou, 1998). In addition they exhibit the function of molecular chaperones and suppress the aggregation of targeted proteins by accelerating their correct folding. In a sense, these functions are independent, but they are applied to the same substrates, probably in sequential manner, so that they may share the same mechanism of recognition.

The system of intracellular protein degradation also requires the recognition of unfolded, misfolded and aggregated proteins. The proteasome is a multicatalytic protease, which is known to degrade unfolded polypeptides with low specificity in substrate selection and cleavage pattern. Molecular modeling based on crystallographic data (Loidl *et al.*, 1999) suggests the presence of six active sites in the inner chamber of the complex, which allows

multivalency in substrate binding. It is interesting that the proteasome can also act as a molecular chaperone (Braun *et al.*, 1999). During the binding of substrate proteins it first tries to fold them and then starts to hydrolyze them if they are unable to fold. The same phenomenon observed with classical chaperones are found with proteasomes, i.e.: promiscuous selective binding of unfolded proteins with precise selecting out and release after the steps of interaction of normally folded and functioning molecules.

4.6. Small heat shock proteins and α-crystallins — the barrier against aggregation of structurally damaged macromolecules

The small heat shock proteins (sHsp) are abundant and ubiquitous intracellular proteins which possess a number of functions, one of which is the prevention of aggregation of proteins denatured by heat and other stress factors (Ehrnsperger *et al.*, 1997; Lee *et al.*, 1997). They range in size from 12 to 42 kDa and can be found as loosely structured complexes of 200–800 kDa. They share structural homology with eye lens α-crystallins and can form mixed complexes with them (Merck *et al.*, 1993). Moreover, α-crystallins are found in all living cells, and they are thought to perform similar functions. Both protein groups can form stable complexes with folding intermediates of their substrates. The exact mechanism of their function is not known, and the function itself is formulated as an efficient trapping of a large number of unfolded proteins in a folding-competent state. This may create a reservoir of non-native proteins for an extended period of time, allowing refolding after restoration of physiological conditions in cooperation with other chaperones (Ehrnsperger *et al.*, 1997).

In studies of sHsp 16.3 the interesting observation was made that very mild treatments by heating, urea or guanidine hydrochloride, which enhance the dynamics of the polypeptide chain, increase dramatically the chaperone activity (Yang *et al.*, 1999). This increase is followed by exposure of hydrophobic surfaces as revealed by fluorescence probes. This suggests an important role of flexibility in performing the chaperone function.

The three-dimensional structure of one sHsp has been determined recently (Kim *et al.*, 1998). The monomeric folding unit consists of a β-sandwich in which one of the strands comes from the neighboring molecule. A 24-meric structure forms a hollow spherical complex of octahedral symmetry, with eight tragonal and six square 'windows'. The hole is 6.5 nm in diameter. The amino-terminal sequence of 32 residues is highly disordered, but from residue 33 onwards, including the entire α-crystallin domain (the segment 46–135 homologous to α-crystallin) and the carboxy-terminal extension, it is well-ordered. In an oligomeric complex the sequence of disordered residues is located inside the sphere. By analogy, one may think that in α-crystallins also, a part of the sequence is not ordered and participates in substrate binding. This possibility is supported by the fact that mutations in this region of αB-crystallin abolish its chaperone activity *in vitro* without influencing the size of the oligomeric complex (Plater *et al.*, 1996).

Because of crystallization problems, the three-dimensional structure of α-crystallin has not been resolved. The model based on comparison with highly homologous β/γ-crystallins suggests a high flexibility of the C-terminal domain (Singh *et al.*, 1996; Farnsworth *et al.*, 1998), which may also be a candidate for substrate binding (Lindner *et al.*, 1998). The other possible reason for the presence of an unstructured C-terminal domain or its highly flexible extension (Carver *et al.*, 1992) is its role in providing high protein solubility, which is necessary in view of its large hydrophobic surface. A decreased mobility of this site can be achieved by insertion of the hydrophobic Trp residue (Smulders *et al.*, 1996), which results in dramatic reduction of chaperone activity.

The substrate specificity of α-crystallin is not very clear. There is evidence that it can recognize and bind proteins with perturbed and damaged conformation, which are very close to the native state (Das and Surewicz, 1995; Das *et al.*, 1996). At the same time it can bind proteins that are heat-denatured (Das and Surewicz, 1995; Carver *et al.*, 1995; Das *et al.*, 1999) or denatured by reduction of S–S bonds (Farahbakhsh *et al.*, 1995) and prevent their aggregation. However, it does not react with stable, hydrophobic proteins (e.g. reduced and carboxymethylated α-lactalbumin and α-casein) (Carver *et al.*, 1995).

Controversial data exist regarding the overall α-crystallin structure as a function of temperature and complex-formation with substrate protein. One set of data based on NMR and protein fluorescence indicates only slight changes in hydrophobicity of the N-terminal region and mobility of the C-terminal region (Lindner *et al.*, 1998), while other results based on binding fluorescence probes indicates the transition to a 'molten globule' state (Raman *et al.*, 1995). We decided to verify if indeed the α-crystallin molecule displays a dramatic change of conformation as a function of substrate binding and temperature (Vadzuk, Ercelen and Demchenko, to be published). We observed that some modulation of protein structure does indeed occur, and is easily detected by tryptophan fluorescence, but the observation of a significant level of red-edge effect (Demchenko, 1986) together with relatively high anisotropy of emission indicates the preservation of a rigid core of N-terminal tryptophan-containing domain, which is in line with a recently published model (Farnsworth *et al.*, 1998). Thus, plasticity and rigidity coexist in the same molecule.

In conclusion, in the case of interaction of molecular chaperones with unfolded or partially folded substrate proteins, we observe that 'group-specific' recognition can be achieved between flexible macromolecular structures, which results in ordering and dissociation of one of them (substrate protein).

5. FALSE RECOGNITION: 'INCLUSION BODIES' AND PROTEIN FOLDING DISEASES

An important question arises regarding possible mistakes in the recognition between flexible structures and the cost of these mistakes. Many proteins can fold spontaneously

following their normal folding pathway, determined by their structure and the folding code imprinted in it (Anfinsen, 1973), but there are also cases when the normal folding is disrupted with the accumulation of aggregates of denatured protein. This causes serious complications such as an accumulation of insoluble 'inclusion bodies' when animal protein genes are expressed in bacteria. The protein aggregates, which appear in animal and human cells, cause a number of pathological conditions. In both cases the polypeptide chains are synthesized normally, and their inability to fold is due to interactions with other factors that cause non-specific aggregation.

5.1. Inclusion bodies — the aggregates of recombinant proteins in bacterial cells

Very often recombinant proteins expressed in bacteria do not fold normally, but form inert and inactive aggregates (inclusion bodies) that accumulate in host cells. These aggregates can be isolated, dissolved in chemical denaturants and refolded again (Rudolph and Lille, 1996). The fact that many proteins recovered from inclusion bodies can fold spontaneously *in vitro* suggests that the problem is not caused by a missing factor, but rather by the presence of some additional factor, which causes the aggregation. This factor is the aggregated protein itself. It forms a template for binding the newly synthesized proteins leading to further aggregation. Protein folding intermediates are thermally unstable (King *et al.*, 1996), and the aggregate provides a surface with high affinity for their binding.

5.2. Protein misfolding diseases

There are several diseases which are characterized by abnormal accumulation of aggregated proteins. They are commonly known as amyloidosis, a condition in which certain proteins or protein fragments precipitate in various tissues as amyloid, a fibrillar aggregate in a pleated-sheet conformation. Thus in the case of Alzheimer's disease, insoluble fibrilar deposits known as amyloid plaque are formed in neural cells (Kosik, 1992). Its major component is β amyloid protein Aβ. The formation of amyloid plaque is related to the flexible conformation of its N-terminal domain, which allows an easy transition between native α-helical and non-native β-stranded structures. The fact that the transition from monomer to neurotoxic amyloid is mediated by the template was confirmed directly in studies of deposition of soluble Aβ onto amyloid fibrils (Esler *et al.*, 2000).

How can this process start in the absence of template? A minor 'non-Aβ' component in the plaque has recently been identified, which forms amyloid fibrils *in vitro* by seeding a fibril formation by major Aβ component β 1–40 (Han *et al.*, 1995). It may therefore be a candidate for a template of *in vivo* plaque formation. Conformational analysis of this protein (Weinreb *et al.*, 1996) demonstrates that it exists as a mixture of rapidly equilibrating extended conformers and is representative of a class of 'natively unfolded' proteins, many of which are known to initiate protein–protein interactions.

Tau protein which is involved in the stabilization of microtubules and is naturally disordered (Schweers *et al.*, 1994) can, by aggregation, form fibrillar bodies characteristic of Alzheimer's disease (Kosik, 1992). These bodies consist almost exclusively of tau protein that is induced to adopt the pleated-sheet structure on transition to the aggregated form.

Abnormal aggregates of non-native states are formed in other disease conditions, in particular in systemic amyloidosis consisting of aggregated transthyretin and its peptide fragments (Jarvis *et al.*, 1993). As in previous cases, the fibrillar state is composed mainly of β-sheets (Terry *et al.*, 1993). The soluble protein form in this case is also composed of β-sheets, but in contrast to previous cases, the β-sheets are also characteristic of native protein structure. Since native protein cannot form aggregates, a profound reorganization of β-structure must occur in the pathological state.

Thus, in pathologies associated with amyloid deposits, the formation or reorganization of β-structure is concomitant with agregation. The aggregate growth occurs on interaction of unfolded or pre-folded protein with template formed by misfolded agregated proteins.

5.3. Prion protein aggregation and 'mad cow disease'

Small infectious particles called prions cause certain neurodegenerative diseases, including scrapie in sheep and goats, bovine spongiform encephalopathy ('mad cow disease') in cattle and Creutzfeldt–Jacob disease in humans. The fact that prions consist exclusively of protein drew a lot of attention to the mechanism of their replication and propagation (Cohen and Prusiner, 1998). The prion protein PrP^{Sc} has a normal cellular isoform PrP^C with a predominantly α-helical conformation (Zahn *et al.*, 2000) and an apparent absence of ability to associate. In contrast, the scrapie isoform is a protein-resistant aggregate composed of the same protein but in dense β-sheeted structure (Inouye and Kirschner, 1997). The protein is probably synthesized as a normal isoform, but on interaction with the template of aggregated protein, it adopts a conformation of a pathological protein, which allows the template to grow. This template may play a role of infectious particle. The absence of a simple and direct route for conformational isomerization from α-helical to β-sheeted conformation suggests that either the misfolding occurs after the chain synthesis before the acquisition of the folded structure or the transformation is produced by interaction with the surface of the prion particle (Horiuchi and Caughey, 1999; Cohen 1999).

Thus, inclusion bodies and protein misfolding diseases can be considered as resulting from high protein flexibility. The interaction potential between folded compact globular proteins is usually strongly repulsive, while unfolded proteins and folding intermediates with low structural order are unstable and prone to formation of aggregates. In fact, misfolded proteins, if they do not aggregate, do not produce any danger. They can be easily eliminated by intracellular proteases (Ritter and Helenius, 2000; Matouschek, 2000), while the highly dense β-sheeted fibrils are strongly protease-resistant. Therefore 'misfolding diseases' may be called 'diseases of false recognition'.

6. KINETICS AND THERMODYNAMICS OF RECOGNITION

6.1. Kinetic steps and their analysis

Protein–protein recognition kinetics are relatively easily analyzed in a rigid-body approximation, i.e. when the precise assembly of two sterically and chemically complementary surfaces occurs leaving no extra degrees of freedom to the component molecules (Janin, 1995). In this case the two partners come together by translational diffusion. If they are in the right position and orientation, an interface develops as part of the molecular surface and internal degrees of freedom relax to optimize short-range interactions. If the orientation is not correct, then the partners dissociate and collide again. The collision rate K_{coll} is determined by the Einstein–Smoluchowski equation and is of the order of $10^9 \, \text{M}^{-1}\text{s}^{-1}$. This rate is a theoretical limit for protein–protein associations. It should dramatically decrease if not every collision results in binding and if conformational adaptation of the partners is required. Purely geometric estimates suggest that the observed rates might be six orders of magnitude slower than the Smoluchowski rate limit (Northrup and Erickson, 1992), while the experimentally observed rates are only three orders of magnitude slower. The high rates were explained by nonspecific and not orientation-restricted primary binding followed by sliding in a two-dimensional diffusion along the surface (Berg, 1985). This concept is not applicable when the surface is flexible or even structureless. Substantial rate acceleration is suggested by 'electrostatic steering', when the properly designed electrostatic potential guides the binding (Wade et al., 1998), but it is not clear how this effect is realized if the charges fluctuate together with the structures. Partial desolvation of interacting partners may play a role in molecular recognition especially in cases where the long-range electrostatics is weak (Camancho et al., 1999, 2000), although protein segmental dynamics can modulate these effects dramatically.

A virtual intermediate, which is formed at the end-point of diffusional association before the isomerization of the complex, is the encounter pair. Its characteristics are important for determining association rates, but even in the case of rigid-body complexation its structure cannot be determined experimentally (Gabdoulline and Wade, 1998; 1999). When molecules are flexible, the encounter pair should be formed primarily between interacting mobile segments, which then should be reorganized into a stable complex. The characteristic time scale of elementary events in this process is the nanosecond which corresponds to the time for translational motion of a small flexible segment by a few angstrom and also to its elementary step rotational diffusion motion. How many of these elementary steps are needed and which of them determines the rate of the whole process cannot be easily studied.

A commonly applied approach to separate molecular recognition into elementary steps and evaluate rate-limiting events uses the effect of solvent viscosity on rates (Von Hippel and Berg, 1989). In the case of flexible recognition partners, this is not applicable, since the viscosity influences not only the diffusion but also the folding kinetics. In addition to concentration dependencies of rates, a variety of time-resolved techniques are currently used to follow the formation and reorganization of structure. In some cases they show that the complex-formation approaches the diffusion limit. This behavior was observed for recognition-folding of fragments of barnase (Kippen and Fersht, 1995), chymotrypsin inhibitor 2 (Mohana-Borges et al., 1999), folding-assembly of dimeric trp repressor (Gloss and Matthews, 1998), binding of unfolded substrates to molecular chaperone SecB (Randall, 1992) and the transcriptional factors binding to DNA (Berger et al., 1998; Wendt et al., 1997). In contrast, the folding from fragments of staphylococcal nuclease (Light et al., 1974), flexible antigen–antibody interactions (Lindner et al., 1999) and the protein binding to a variety of molecular chaperones is kinetically limited to isomerization steps in the complex. Kinetic measurements under pseudo-first-order conditions often reveal the complex character of these reactions which involve several reaction steps. (Chaffotte et al., 1997).

6.2. Thermodynamics of folding-associated recognition

The formation of protein complexes with proteins and nucleic acids is usually associated with large negative heat capacity changes, $\Delta C^{\circ}_{assoc} < 0$. The strongest contribution to this effect may be provided by the removal of large amounts of non-polar surface on complex formation (Spolar and Record, 1994). Among entropic ($T\Delta S^{\circ}_{assoc}$) and enthalpic ($\Delta H^{\circ}_{assoc}$) contributions to ΔC°_{assoc}, the entropic factor is estimated to be most important. In this case, any ordering of structure which can result in decrease in protein surface areas and squeezing of water molecules from inter-residue contacts should dramatically increase this effect. In a flexible system, this effect should become counter-balanced not only by the decrease of translational entropy, but also by the decrease of configuration entropy of polypeptide chains. This picture is in line with simulations of DNA complex-formation with restriction nucleases (Duan et al., 1996). They demonstrate the presence of higher levels of structure ordering in specific compared to non-specific complexes, and also a significant entropy gain on specific binding. This gain is the result of opposing contributions i.e. solvent release, which increases entropy, vs configurational terms and collective terms from tight coupling between the motions of the protein and DNA.

A favorable entropic contribution to the binding was observed by isothermal titration calorimetry for angiotensin II binding antibody (Murphy et al., 1993), the binding of captopril to angiotensin-converting enzyme (Ortiz-Salmeron et al., 1998), complex-formation of ovomucoid third domain to elastase (Baker and Murphy, 1997) and folding of dimerization domain of trp repressor (Gloss and Matthews, 1998). However, this picture cannot be generalized to every case of recognition between flexible structures. Thus, the binding of calmodulin to smMLCKp proceeds with negative changes in enthalpy, heat capacity and entropy indicating that it is an enthalpy-driven and entropically unfavorable process (Wintrode and Privalov, 1997). This may be due to the fact that smMLCKp in the complex becomes more flexible (Ehrhardt et al., 1995). The increase of protein

backbone dynamics on complex formation has been reported in a number of cases (Zidek *et al.*, 1999; Forman-Kay, 1999). In general, the thermodynamic analysis strongly suggests that the ordering of structure and desolvation (or disordering and solvation) are strongly coupled, and this coupling may allow different mechanisms of sequential and selective enhancement of interactions.

Thus, in protein folding and binding two dominant and opposing contributions to entropy are manifested, one from the hydrophobic effect, or the release of water on burial of nonpolar surface, and the other from the reduction in conformational entropy. If molecular recognition is coupled with local or global folding, the hydrophobic effect may become much more significant than the entropy loss due to the loss of conformational mobility. This suggests an essentially nonlinear phenomenon: when the binding is stronger and more specific with higher positive interaction enthalpy contribution to the binding free energy, the entropy contribution can become large and positive. Thus, the binding, which immobilizes the interaction partners, may become selectively stronger, and this may be the key factor to provide highly selective recognition.

At present, because of the presence of different opposing factors, many details of the kinetics and thermodynamics of recognition between flexible structures are still obscure. I will outline only two important problems, for which a solution is needed.

1. In simple kinetics, if the strong binding is associated with fast association rates, then the dissociation rates should always be slow (the affinity constant is the ratio of association and dissociation rates). However, in many cases, specific and high affinity binding is provided with very high association rates. Many functionally important recognition processes such as chaperone-assisted folding, intracellular signaling or protein–DNA interaction require a very rapid dissociation (Felder *et al.*, 1993). How this is mechanistically realized remains unresolved. One of the suggestions to explain this fact is that the unbinding process is adiabatic, and that entropic changes occur after unbinding (Moy *et al.*, 1994).

2. Specific recognition in many cases occurs in conditions of high concentration of ligand analogs, which are able to bind non-specifically. There are also cases when the receptor has a great number of non-specific interaction sites, which may hamper specific binding. Thus, endonucleases and transcriptional activators can bind to any DNA site non-specifically, which should result in a strong effect of competitive inhibition (Jen-Jacobson, 1997), This is not observed. The fact that flexibility is different for the partners participating in specific and non-specific complexes may contain a clue to this problem.

7. MODELING OF RECOGNITION BETWEEN FLEXIBLE STRUCTURES

Computational tools for predicting ligand–receptor binding complementarity, affinity and association rates develop rapidly since they are needed for rational drug design. In this respect the application of rigid-body approximation is the simplest approach, since it does not require the searching of extensive conformational space (Walls and Sternberg, 1992; Náray-Szabó, 1989, 1994; Sobolev *et al.*, 1996; Jackson *et al.*, 1998). The fact that flexible molecules change their conformations in intermolecular interactions raises new combinatorial as well as energetic problems. Analysis of the conformational energies of flexible molecules showed that, for most of those compounds, both the crystal and protein-bound conformations were energetically well above the global minimum. In many cases there is not even any local energy minimum (Nicklaus *et al.*, 1995). This makes the energy refinement problem very difficult.

Some new computational strategies for flexible docking and design have been extended from rigid-body to flexible docking: Monte Carlo/molecular dynamics docking, in-site combinatorial search, ligand build-up, site mapping and fragment assembly (Rosenfeld *et al.*, 1995; Jackson *et al.*, 1998; Najmanovich *et al.*, 2000). In order to produce the global optimization of a multivariable function, which is the energy, and to discriminate between high-specific and low-specific binding, Monte-Carlo minimization algorithms are applied (Caflisch *et al.*, 1992; Friedman *et al.*, 1994). An algorithm based on molecular dynamics simulations (Di Nola *et al.*, 1994; Mangoni *et al.*, 1999) allows one to manipulate to some extent the ligand, receptor and solvent mobilities by assigning different temperatures to these subsystems.

A general approach to address the problem of finding an appropriate protein docking solution is to separate it into two steps: (1) a search over the N dimensional binding and conformational spaces in order to select candidate geometries of the complex; (2) application of a suitable scoring function to distinguish near native modes of binding from the other false solutions generated during the initial search (Palma *et al.*, 2000). In the case of flexible interacting partners, this approach is very difficult to apply in view of the huge number of alternative geometries (Gehlhaar *et al.*, 1995; Zhang *et al.*, 1999). Thus flexibility in modeling the molecular recognition is introduced sequentially by allowing flexibility in a small-sized ligand (Friedman *et al.*, 1994; Desmet *et al.*, 1997; Lorber and Schoichet, 1998; Wang *et al.*, 1999b), flexibility in receptor protein as movement of loops or large domains (Sandak *et al.*, 1998; Verbitsky *et al.*, 1999) and the small-scale flexibility of side chains (Desmet *et al.*, 1997; Najmanovich *et al.*, 2000). With the inclusion of one type of motion, the other types are usually ignored. It has been shown that the assumption of a rigid binding site can lead to errors in identification of the correct binding mode and assessment of binding affinity, even for proteins which show a relatively small shift in atomic positions from one ligand to the next (Lorber and Shoichet, 1998).

In order to understand the mechanism of protein folding, a new methodology has been proposed recently (Bryngelson *et al.*, 1995). It considers the process of folding as motion along the energy landscape resembling a funnel with a broad upper part representing a magnitude of conformations of the unfolded state and a narrow well representing the folded conformation. The folding proceeds downhill along the free energy gradient. Along this pathway, local minima can

exist, which will result in folding intermediates. This concept has been recently extended to model folding-associated recognition (Ma *et al.*, 1999; Kumar *et al.*, 2000). The complex-formation may change substantially the energy landscapes with different shapes and distributions of conformers. This approach allows introduction of the concept of statistical ensemble of conformers and tries to obtain the equilibrium state of the complex as a distribution of conformers in the vicinity of the free energy well.

The goal in these studies is to understand the mechanisms of formation and functioning of supramolecular complexes and of computational structure-based drug design. It is also to predict the structures of complexes that have not yet been experimentally determined by X-ray crystallography or NMR. In addition, we have to predict their affinities and sequential steps of their formation. For this task, docking of rigid ligands is inadequate because it assumes knowledge of the conformation of the bound ligand and of the bound state of the receptor. Docking of flexible ligands to flexible receptors would be desirable, but with present methodology this means searching an enormous conformational space, which is not feasible. Ideas based on protein folding funnels are attractive, but they have not yet become efficient research tools.

8. CONCLUSIONS

In view of the very frequent occurrence of dynamic recognition between flexible structures and in different structurally and functionally dissimilar systems, can we expect that there should be some general principles for this type of recognition? If yes, what are these principles? I will try to formulate them below:

1. The reaction of complex formation should consist of at least two steps — a bimolecular reaction of *diffusional formation of encounter pair* and a unimolecular *isomerization of encounter pair into a stable complex*. These steps are kinetically coupled, and in the applied range of concentrations, depending on the particular system, either of them can be rate-limiting. The formation of an encounter pair can be accelerated by long-range electrostatic interactions (Antosiewicz *et al.*, 1996; Wade *et al.*, 1998). Their preformed pattern can help to achieve proper selection and orientation of interacting partners. The encounter rate can also be accelerated by reduced diffusion in space, e.g. by sliding of a protein along double-helical DNA molecules (Von Hippel and Berg, 1989; Theis *et al.*, 1999) or moving while adsorbed to a biomembrane surface.

2. The encounter pair is a loose associate which isomerizes into a stable complex in a number of steps, which involve *sequential selection*. There is no other way to achieve the specificity of a complex being formed other than by probing it by many stochastic bond-making and bond-breaking events. Sequential selection implies that the effective ligand association occurs not in one act, but involves consecutive elementary steps during which more and more sub-sites of one partner make contacts with complementary sub-sites of the other (Neumann, 1981). The effective rate constant of ligand binding is therefore a net quantity composed of the rate constants of all the steps involved. Only if the next in sequence contact occurs within the lifetime of a given ligand–receptor configuration does the ligand become gradually bound more tightly; it is then selected as opposed to ligands with weaker sub-site binding. Each sequential step is distinguished from the previous one by stronger binding energy and longer duration.

The mechanisms of sequential selection should probably involve *kinetic proofreading* (Hopfield, 1974) or *kinetic amplification* (Ninio, 1975, 1977). These effects are best understood in the case of two successive steps, where the first one is fast, reversible and of low selectivity, whereas the second one is slow, highly selective and irreversible. A wrong ligand (with attractive but weak interaction potential) undergoes the first step, but thereafter it will be eliminated by the back reaction before it becomes involved on the second step. In contrast, the correct substrate will have a much higher probability to pass successfully both steps. Thus, on the first step there may be a rather flat energy profile, while the second one must exhibit a much higher and more discriminative energy barrier. Apparent irreversibility of the second step for correct ligand occurs because of depopulation of ligand conformers after this step due to a new sequence of events occurring in the complex. This dramatically increases the selectivity of the whole reaction, and this occurs in a sequential manner.

3. Sequential steps in recognition are not equivalent. They exhibit *hierarchy of time and length scales*, which mechanistically is similar to protein folding (Demchenko, 2000b). When a certain number of primary contacts are formed, the reaction of ordering and re-ordering of structure proceeds in the complex as an integrated unit. This is an isomerization process with sequential increase of integration. The similarity with protein folding (Miller and Dill, 1997; Tsai *et al.*, 1999) is due to involvement of the same well-known types of weak non-covalent interactions between groups of atoms and the same thermodynamic force which drives the system towards the free energy minimum global for the whole complex. Current achievements in solving the protein folding problem can be successfully applied in the analysis of dynamic complex formation.

4. Interaction of slow and fast dynamic modes may result in *nonlinear effects* and generation of *new stationary states*, which do not exist in the free, separate partners (Kharkianen and Demchenko, 2000, to be published). As a result of this process a conformational signal propagating beyond an interface and triggering an effectory function may appear. Conformational changes may not be needed for molecular recognition, but molecular recognition is absolutely needed for many biological functions in which the induction or amplification of effectory signal via conformational changes in the regions outside the contact areas is required. The resultant integrated structural unit can retain (or achieve) a highly dynamic structure, which allows combination of high specificity of interaction with easy formation and dissociation of the unit. This

makes it possible to avoid the conformational entropy cost of rigid complex formation and allows for a broad variation of contact areas. Thus the contact area and the strength of binding may not correlate (Varani, 1997).

In a dynamical system composed of interacting molecules with flexible conformations it is hard to incorporate the term 'complementarity'. What actually occurs is a stochastic process of ordering in a loosely coupled complex influenced by local energy field gradients. Thus, complementarity is created together with the acquisition of order. Many trials of binding-release must occur until the proper ligand is selected and its interactions with the receptor are reinforced with the optimal amount of short-range non-covalent bonds. Although no rigorous analysis has been performed, it seems highly probable that recognition between flexible structures with many isomerization steps is far more efficient in terms of rate and specificity than the 'lock-and-key' binding with its large number of unsuccessful fittings.

Thus, it is highly probable that the most specific recognition is achieved in nature by means of highly flexible structures. The reason may be that this type of recognition allows more possibilities to optimize the fit between the interacting partners using a broader range of structural and energetic parameters. Flexibility extends dramatically the space dimensionality for conformational search, and this factor will result in dramatic increase in selectivity. Using the 'lock-and-key' analogy we may conclude that an electronic key with millions of variants is more selective for opening the door than a mechanical key with only a few hundreds variants.

REFERENCES

Allain FH, Gubser CC, Howe PW, Nagai K, Neuhaus D, Varani G. 1996. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature* **380**: 646–650.

Andersen SS. 2000. Spindle assembly and the art of regulating microtubule dynamics by MAPs and Stathmin/Op18. *Trends Cell Biol.* **10**: 261–267.

Andria G, Taniuchi H, Cone JL. 1971. The specific binding of three fragments of staphylococcal nuclease. *J. Biol. Chem.* **246**: 7421–7428.

Anfinsen CB. 1973. Principles that govern the folding of protein chains. *Science* **181**: 223–230.

Antosiewicz J, Wlodek ST, McCammon JA. 1996. Acetylcholinesterase: Role of the enzyme's charge distribution in steering charged ligands toward the active site. *Biopolymers* **39**: 85–94.

Aoki K, Motojima F, Taguchi H, Yomo T, Yoshida M. 2000. GroEL binds artificial proteins with random sequences. *J. Biol. Chem.* **275**: 13755–13758.

Baker BM, Murphy KP. 1997. Dissecting the energetics of a protein–protein interaction: the binding of ovomucoid third domain to elastase. *J. Mol. Biol.* **268**: 557–569.

Barry JK, Matthews KS. 1999. Thermodynamic analysis of unfolding and dissociation in lactose repressor protein. *Biochemistry* **38**: 6520–6528.

Benevides JM, Li T, Lu XJ, Srinivasan AR, Olson WK, Weiss MA, Thomas GJ Jr. 2000. Protein-directed DNA structure II. Raman spectroscopy of a leucine zipper bZIP complex. *Biochemistry* **39**: 548–556.

Berg OG. 1985. Orientation constraints in diffusion-limited macromolecular association. The role of surface diffusion as a rate-enhancing mechanism. *Biophys. J.* **47**: 1–14.

Berger C, Jelesarov I, Bosshard HR. 1996. Coupled folding and site-specific binding of the GCN4-bZIP transcription factor to the AP-1 and ATF/CREB DNA sites studied by microcalorimetry. *Biochemistry* **35**: 14984–14989.

Berger C, Piubelli L, Haditsch U, Bosshard HR. 1998. Diffusion-controlled DNA recognition by an unfolded, monomeric bZIP transcription factor. *FEBS Lett.* **425**: 14–18.

Berger C, Weber-Bornhauser S, Eggenberger J, Hanes J, Pluckthun A, Bosshard HR. 1999. Antigen recognition by conformational selection. *FEBS Lett.* **450**: 149–153.

Bourne Y, Arva AS, Bernstein SL, Watson MH, Reed SI, Endicott JE, Noble ME, Johnson LN, Tainer JA. 1995. Crystal structure of the cell cycle-regulatory protein suc1 reveals a *β*-hinge conformational switch. *Proc Natl Acad. Sci. USA* **92**: 10232–10236.

Bowie JU, Sauer RT. 1989. Equilibrium dissociation and unfolding of the Arc repressor dimer. *Biochemistry* **28**: 7139–7143.

Braig K, Adams PD, Brunger AT. 1995. Conformational variability in the refined structure of the chaperonin GroEL at 2.8 A resolution. *Nature Struct. Biol.* **2**: 1083–1094.

Braun BC, Glickman M, Kraft R, Dahlmann B, Kloetzel PM, Finley D, Schmidt M. 1999. The base of the proteasome regulatory particle exhibits chaperone-like activity. *Nature Cell Biol.* **1**: 221–226.

Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**: 167–195.

Buchner J. 1999. Hsp90 & Co. — a holding for folding. *Trends Biochem Sci* **24**: 136–141.

Buckle AM, Zahn R, Fersht AR. 1997. A structural model for GroEL-polypeptide recognition. *Proc. Natl Acad. Sci. USA* **94**: 3571–3575.

Caflisch A, Niederer P, Anliker M. 1992. Monte Carlo docking of oligopeptides to proteins. *Proteins* **13**: 223–230.

Camacho CJ, Weng Z, Vajda S, DeLisi C. 1999. Free energy landscapes of encounter complexes in protein-protein association. *Biophys J.* **76**: 1166–1178.

Camacho CJ, Kimura SR, DeLisi C, Vajda S. 2000. Kinetics of desolvation-mediated protein-protein binding. *Biophys. J.* **78**: 1094–1105.

Caplan AJ. 1999. Hsp90's secrets unfold: new insights from structural and functional studies. *Trends Cell. Biol.* **9**: 262–268.

Carroll AS, Gilbert DE, Liu X, Cheung JW, Michnowicz JE, Wagner G, Ellenberger TE, Blackwell TK. 1997. SKN-1 domain folding and basic region monomer stabilization upon DNA binding. *Genes Dev.* **11**: 2227–2238.

Carver JA, Aquilina JA, Truscott RJ, Ralston GB. 1992. Identification by 1H NMR spectroscopy of flexible C-terminal extensions in bovine lens alpha-crystallin. *FEBS Lett.* **311**: 143–149.

Carver JA, Guerreiro N, Nicholls KA, Truscott RJ. 1995. On the interaction of alpha-crystallin with unfolded proteins. *Biochim. Biophys. Acta* **1252**: 251–260.

Chaffotte AF, Li JH, Georgescu RE, Goldberg ME, Tasayco ML. 1997. Recognition between disordered states: kinetics of the self-assembly of thioredoxin fragments. *Biochemistry* **36**: 16040–16048.

Chatellier J, Buckle AM, Fersht AR. 1999. GroEL recognizes sequential and non-sequential linear structural motifs compatible with extended beta-strands and alpha-helices. *J. Mol. Biol.* **292**: 163–172.

Chen L, Sigler PB. 1999. The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. *Cell* **99**: 757–768.

Clarke ND, Beamer LJ, Goldberg HR, Berkower C, Pabo CO. 1991. The DNA binding arm of lambda repressor: critical contacts from a flexible region. *Science* **254**: 267–270.

Coffey RL, Purich DL. 1995. Non-cooperative binding of the MAP-

2 microtubule-binding region to microtubules. *J. Biol. Chem.* **270**: 1035–1040.

Cohen FE 1999. Protein misfolding and prion diseases. *J. Mol. Biol.* **293**: 313–320.

Cohen FE, Prusiner SB 1998. Pathologic conformations of prion proteins. *A. Rev. Biochem.* **67**: 793–819.

Csermely P, Miyata Y, Soti C, Yahara I. 1997. Binding affinity of proteins to hsp90 correlates with both hydrophobicity and positive charges. A surface plasmon resonance study. *Life Sci.* **61**: 411–418.

Csermely P, Schnaider T, Soti C, Prohaszka Z, Nardai G 1998 The 90-kDa molecular chaperone family: structure, function, and clinical applications. A comprehensive review. *Pharmac. Ther.* **79**: 129–168.

Das KP, Surewicz WK. 1995. On the substrate specificity of alpha-crystallin as a molecular chaperone. *Biochem. J.* **311**: 367–370.

Das KP, Petrash JM, Surewicz WK. 1996. Conformational properties of substrate proteins bound to a molecular chaperone alpha-crystallin. *J. Biol. Chem.* **271**: 10449–10452.

Das KP, Choo-Smith LP, Petrash JM, Surewicz WK. 1999. Insight into the secondary structure of non-native proteins bound to a molecular chaperone alpha-crystallin. An isotope-edited infrared spectroscopic study. *J. Biol. Chem.* **274**: 33209–33212.

Daughdrill GW, Chadsey MS, Karlinsey JE, Hughes KT, Dahlquist FW. 1997. The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28. *Nature Struct. Biol.* **4**: 285–291.

Demchenko AP. 1986. *Ultraviolet Spectroscopy of Proteins*. Springer: Berlin.

Demchenko AP. 1999. Protein folding with molecular chaperones. *Comments Mol. Cell. Biophys.* **9**: 219–260.

Demchenko AP. 2000a. Protein folding with molecular chaperones: stochastic process under control. *Biofizika (Moscow)* **45**: 414–420.

Demchenko AP. 2000b. Concepts and misconcepts in the science of protein folding. *Current Protein Peptide Sci.* (submitted).

Demchenko AP, Chinarov VA. 1999. Tolerance of protein structures to the changes of amino acid sequences and their interactions. The nature of the folding code. *Protein Peptide Lett.* **6**: 115–129.

Desmet J, Wilson IA, Joniau M, De Maeyer M, Lasters I. 1997. Computation of the binding of fully flexible peptides to proteins with flexible side chains. *FASEB J.* **11**: 164–172.

Di Nola A, Roccatano D, Berendsen HJ. 1994. Molecular dynamics simulation of the docking of substrates to proteins. *Proteins* **19**: 174–182.

Draper DE. 1995. Protein–RNA recognition. *A. Rev. Biochem.* **64**: 593–620.

Draper DE. 1999. Themes in RNA–protein recognition. *J. Mol. Biol.* **293**: 255–270.

Duan Y, Wilkosz P, Rosenberg JM. 1996. Dynamic contributions to the DNA binding entropy of the EcoRI and EcoRV restriction nucleases. *J. Mol. Biol.* **264**: 546–555.

Duggan BM, Dyson HJ, Wright PE. 1999. Inherent flexibility in a potent inhibitor of blood coagulation, recombinant nematode anticoagulant protein c2. *Eur. J. Biochem.* **265**: 539–548.

Ehrhardt MR, Urbauer JL, Wand AJ. 1995. The energetics and dynamics of molecular recognition by calmodulin. *Biochemistry* **34**: 2731–2738.

Ehrnsperger M, Graber S, Gaestel M, Buchner J. 1997. Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. *EMBO J.* **16**: 221–229.

Ellenberger TE, Brandl CJ, Struhl K, Harrison SC. 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell* **71**: 1223–1237.

Endicott JA, Noble ME, Garman EF, Brown N, Rasmussen B, Nurse P, Johnson LN. 1995. The crystal structure of p13-suc1, a p34-cdc2-interacting cell cycle control protein. *Eur. Mol. Biol. Org. J.* **14**: 1004–1014.

Esler WP, Stimson ER, Jennings JM, Vinters HV, Ghilardi JR, Lee JP, Mantyh PW, Maggio JE. 2000. Alzheimer's disease amyloid propagation by template-dependent dock-lock mechanism. *Biochemistry* **39**: 6288–6295.

Farahbakhsh ZT, Huang QL, Ding LL, Altenbach C, Steinhoff HJ, Horwitz J, Hubbell WL. 1995. Interaction of alpha-crystallin with spin-labeled peptides. *Biochemistry* **34**: 509–516.

Farnsworth PN, Frauwirth H, Groth-Vasselli B, Singh K. 1998. Refinement of 3D structure of bovine lens alpha A-crystallin. *Int. J. Biol. Macromol.* **22**: 175–185.

Farr CD, Slepenkov SV, Witt SN. 1998. Visualization of a slow, ATP-induced structural transition in the bacterial molecular chaperone DnaK. *J. Biol. Chem.* **273**: 9744–9748.

Farr GW, Furtak K, Rowland MB, Ranson NA, Saibil HR, Kirchhausen T, Horwich AL. 2000. Multivalent binding of nonnative substrate proteins by the chaperonin GroEL. *Cell* **100**: 561–573.

Fasman GD, Park K, Randall LL. 1995. Chaperone SecB: conformational changes demonstrated by circular dichroism. *J. Protein Chem.* **14**: 995–600.

Felder S, Zhou M, Hu P, Urena J, Ullrich A, Chaudhuri M, White M, Shoelson SE, Schlessinger J. 1993. SH2 domains exhibit high-affinity binding to tyrosine-phosphorylated peptides yet also exhibit rapid dissociation and exchange. *Mol. Cell. Biol.* **13**: 1449–1455.

Foote J, Milstein C. 1994. Conformational isomerism and the diversity of antibodies. *Proc. Natl Acad. Sci. USA* **91**: 10370–10374.

Forman-Kay JD. 1999. The 'dynamics' in the thermodynamics of binding. *Nature Struct. Biol.* **6**: 1086–1087.

Fourie AM, Sambrook JF, Gething MJ. 1994. Common and divergent peptide binding specificities of Hsp70 molecular chaperones. *J. Biol. Chem.* **269**: 30470–30478.

Frankel AD, Smith CA. 1998. Induced folding in RNA–protein recognition: more than a simple molecular handshake. *Cell* **92**: 149–151.

Friedhoff P, von Bergena M, Mandelkowa E. 2000. Structure of tau protein and assembly into paired helical filaments. *Biochim. Biophys. Acta* **1502**: 122–132.

Friedman AR, Roberts VA, Tainer JA. 1994. Predicting molecular interactions and inducible complementarity: fragment docking of Fab–peptide complexes. *Proteins* **20**: 15–24.

Gabdoulline RR, Wade RC. 1998. Brownian dynamics simulation of protein-protein diffusional encounter. *Methods* **14**: 329–341.

Gabdoulline RR, Wade RC. 1999. On the protein-protein diffusional encounter complex. *J. Mol. Recogn.* **12**: 226–234.

Galan A, Llorca O, Valpuesta JM, Perez-Perez J, Carrascosa JL, Menendez M, Banuelos S, Muga A. 1999. ATP hydrolysis induces an intermediate conformational state in GroEL. *Eur. J. Biochem.* **259**: 347–355.

Gast K, Damaschun H, Eckert K, Schulze-Forster K, Maurer HR, Muller-Frohne M, Zirwer D, Czarnecki J, Damaschun G. 1995. Prothymosin alpha: a biologically active protein with random coil conformation. *Biochemistry* **34**: 13211–13218.

Gavigan SA, Nguyen T, Nguyen N, Senear DF. 1999. Role of multiple CytR binding sites on cooperativity, competition and induction at the Escherichia coli udp promoter. *J. Biol. Chem.* **274**: 16010–16019.

Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST. 1995. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* **2**: 317–324.

Ghoshal AK, Swaminathan CP, Thomas CJ, Surolia A, Varadarajan R. 1999. Thermodynamic and kinetic analysis of the *Escherichia coli* thioredoxin-C' fragment complementation system. *Biochem J.* **339**: 721–727.

Gisler SM, Pierpaoli EV, Christen P. 1998. Catapult mechanism renders the chaperone action of Hsp70 unidirectional. *J. Mol. Biol.* **279**: 833–840.

Gloss LM, Matthews CR. 1998. Mechanism of folding of the dimeric core domain of *Escherichia coli* trp repressor: a nearly diffusion-limited reaction leads to the formation of an on-pathway dimeric intermediate. *Biochemistry* **37**: 15990–15999.

Goldberg JM, Baldwin RL. 1999. A specific transition state for S-

peptide combining with folded S-protein and then refolding. *Proc. Natl Acad. Sci. USA* **96**: 2019–2024.

Gragerov A, Zeng L, Zhao X, Burkholder W, Gottesman ME. 1994. Specificity of DnaK-peptide binding. *J. Mol. Biol.* **235**: 848–854.

Gryk MR, Finucane MD, Zheng Z, Jardetzky O. 1995. Solution dynamics of the trp repressor: a study of amide proton exchange by T1 relaxation. *J. Mol. Biol.* **246**: 618–627.

Gryk MR, Jardetzky O, Klig LS, Yanofsky C. 1996. Flexibility of DNA binding domain of trp repressor required for recognition of different operator sequences. *Protein Sci.* **5**: 1195–1197.

Haley DA, Bova MP, Huang QL, Mchaourab HS, Stewart PL. 2000. Small heat-shock protein structures reveal a continuum from symmetric to variable assemblies. *J. Mol. Biol.* **298**: 261–272.

Hall JG, Frieden C. 1989. Protein fragments as probes in the study of protein folding mechanisms: Differential effects of dihydrofolate reductase fragments on the refolding of the intact protein. *Proc. Natl Acad. Sci. USA* **86**: 3060–3064.

Hammarstrom P, Persson M, Owenius R, Lindgren M, Carlsson U. 2000. Protein substrate binding induces conformational changes in the chaperonin GroEL. A suggested mechanism for unfoldase activity. *J. Biol. Chem.* **275**: 22832–22838.

Han H, Weinreb PH, Landsbury PT. 1995. The core Alzheimer's peptide NAC forms amyloid fibrils which seed and are seeded by beta-amyloid: is NAC a common trigger or target in neurodegenerative disease? *Chem. Biol.* **2**: 163–169.

Hantgan RR, Taniuchi H. 1977. Formation of a biologically active, ordered complex from two overlapping fragments of cytochrome c. *J. Biol. Chem.* **252**: 1367–1374.

Hard T. 1999. NMR studies of protein-nucleic acid complexes: structures, solvation, dynamics and coupled protein folding. *O. Rev. Biophys.* **32**: 57–98.

Hendrick JP, Hartl F-U. 1993. Molecular chaperone functions of heat-shock proteins. *A. Rev. Biochem.* **62**: 4349–4384.

Hernandez MA, Avila J, Andreu JM. 1986. Physicochemical characterization of the heat-stable microtubule-associated protein MAP2. *Eur. J. Biochem.* **154**: 41–48.

Hopfield JJ. 1974. Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl Acad. Sci. USA* **71**: 4135–4139.

Horiuchi M, Caughey B. 1999. Specific binding of normal prion protein to the scarpie form via a localized domain initiates its conversion to a protease-resistant state. *EMBO J.* **18**: 3193–3203.

Horton NC, Perona JJ. 1998. Role of protein-induced bending in the specificity of DNA recognition: crystal structure of EcoRV endonuclease complexed with d(AAAGAT) + d(ATCTT). *J. Mol. Biol.* **277**: 779–787.

Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU. 1999. Identification of in vivo substrates of the chaperonin GroEL. *Nature* **402**: 147–154.

Hua QX, Jia WH; Bullock BP, Habener JF, Weiss MA. 1998. Transcriptional activator–coactivator recognition: nascent folding of a kinase-inducible transactivation domain predicts its structure on coactivator binding. *Biochemistry* **37**: 5858–5866.

Huber R, Bennett WS. 1983. Functional significance of flexibility in proteins. *Biopolymers* **22**: 261–279.

Ikegami T, Kuraoka I, Saijo M, Kodo N, Kyogoku Y, Morikawa K, Tanaka K, Shirakawa M. 1999. Resonance assignments, solution structure and backbone dynamics of the DNA- and RPA-binding domain of human repair factor XPA. *J. Biochem. (Tokyo)* **125**: 495–506.

Inouye H, Kirschner DA. 1997. X-ray diffraction analysis of scarpie prion: intermediate and folded structures in a peptide containing two putative α-helices. *J. Mol. Biol.* **268**: 375–389.

Itzhaki LS, Otzen DE, Fersht AR. 1995. Nature and consequences of GroEL–protein interactions. *Biochemistry* **34**: 14581–14587.

Jackson RM, Gabb HA, Sternberg MJ. 1998. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J. Mol. Biol.* **276**: 265–285.

Jana R, Hazbun TR, Mollah AK, Mossing MC. 1997. A folded monomeric intermediate in the formation of lambda Cro dimer-DNA complexes. *J. Mol. Biol.* **273**: 402–416.

Janin J. 1995. Protein-protein recognition. *Prog. Biophys. Mol. Biol.* **64**: 145–166.

Jarvis JA, Craik DJ, Wilce MCJ. 1993. X-ray diffraction studies of fibrils formed from peptide fragments of transthyretin. *Biochem. Biophys. Res. Communs.* **192**: 991–998.

Jen-Jacobson L. 1997. Protein-DNA recognition complexes: conservation of structure and binding energy in the transition state. *Biopolymers* **44**: 153–180.

Jones S, Thornton JM. 1996. Principles of protein-protein interactions. *Proc. Natl Acad. Sci. USA* **93**: 13–20.

Jorgensen CI, Kallipolitis BH, Valentin-Hansen P. 1998. DNA-binding characteristics of the *Escherichia coli* CytR regulator: a relaxed spacing requirement between operator half-sites is provided by a flexible, unstructured interdomain linker. *Mol. Microbiol.* **27**: 41–50.

Juillerat M, Taniuchi H. 1982. Conformational dynamics of a biologically active three-fragment complex of horse cytochrome c. *Proc. Natl Acad. Sci. USA* **79**: 1825–1829.

Juillerat M, Parr GR, Taniuchi H. 1980. A biologically active, three-fragment complex of horse heart cytochrome c. *J. Biol. Chem.* **255**: 845–853.

Katayama E, Scott-Woo G, Ikebe M. 1995. Effect of caldesmon on the assembly of smooth muscle myosin. *J. Biol. Chem.* **270**: 3919–3925.

Kay LE, Muhandiram DR, Farrow NA, Aubin Y, Forman-Kay JD. 1996. Correlation between dynamics and high affinity binding in an SH2 domain interaction. *Biochemistry* **35**: 361–368.

Kay LE, Muhandiram DR, Wolf G, Shoelson SE, Forman-Kay JD. 1998. Correlation between binding and dynamics at SH2 domain interfaces. *Nature Struct. Biol.* **5**: 156–163.

Kim KK, Kim R, Kim SH. 1998. Crystal structure of a small heat-shock protein. *Nature* **394**: 595–599.

King J, Haase-Pettingell C, Robinson AS, Speed M, Mitraki A. 1996. Thermolabile folding intermediates: inclusion body precursors and chaperonin substrates. *FASEB J.* **10**: 57–66.

Kippen AD, Fersht AR. 1995. Analysis of the mechanism of assembly of cleaved barnase from two peptide fragments and its relevance to the folding pathway of uncleaved barnase. *Biochemistry* **34**: 1464–1468.

Knoblauch NT, Rudiger S, Schonfeld HJ, Driessen AJ, Schneider-Mergener J, Bukau B. 1999. Substrate specificity of the SecB chaperone. *J. Biol. Chem.* **274**: 34219–34225.

Kombo DC, Young MA, Beveridge DL. 2000. One nanosecond molecular dynamics simulation of the N-terminal domain of the lambda repressor protein. *Biopolymers* **53**: 596–605.

Konig P, Richmond TJ. 1993. The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.* **233**: 139–154.

Kophengnavong T, Carroll AS, Blackwell TK. 1999. The SKN-1 amino-terminal arm is a DNA specificity segment. *Mol. Cell. Biol.* **19**: 3039–3050.

Koshland, DE Jr and Neet, KE. 1968. The catalytic and regulatory properties of enzymes. *A. Rev. Biochem.* **37**: 359–410.

Kosik KS. 1992. Alzheimer's disease: a cell biological perspective *Science* **256**: 780–783.

Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. 1996. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl Acad. Sci. USA* **93**: 11504–11509.

Kumar S, Ma B, Tsai C-J, Sinha N, Nussinov R. 2000. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.* **9**: 10–19.

Labhardt AM. 1982. Secondary structure in ribonuclease. I. Equilibrium folding transitions seen by amide circular dichroism. *J. Mol. Biol.* **157**: 331–355.

Labhardt AM. 1984. Kinetic circular dichroism shows that the S-peptide alpha-helix of ribonuclease S unfolds fast and refolds slowly. *Proc. Natl Acad. Sci. USA* **81**: 7674–7678.

Landry SJ, Gierasch LM. 1991. The chaperonin GroEL binds a polypeptide in an alpha-helical conformation. *Biochemistry* **30**: 7359–7362.

Landry SJ, Jordan R, McMacken R, Gierasch LM. 1992. Different conformations for the same polypeptide bound to chaper-

ones DnaK and GroEL. *Nature* **355**: 455–457.

Larsson N, Segerman B, Howell B, Fridell K, Cassimeris L, Gullberg M. 1999. Op18/stathmin mediates multiple region-specific tubulin and microtubule-regulating activities. *J. Cell. Biol.* **146**: 1289–1302.

Lau CK, Churchich JE. 1999. Binding of polylysine to GroEL. Inhibition of the refolding of mMDH. *Biochim. Biophys. Acta* **1431**: 282–289.

Leder L, Berger C, Bornhauser S, Wendt H, Ackermann F, Jelesarov I, Bosshard HR. 1995. Spectroscopic, calorimetric, and kinetic demonstration of conformational adaptation in peptide-antibody recognition. *Biochemistry* **34**: 16509–16518.

Lee GJ, Roseman AM, Saibil HR, Vierling E. 1997. A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state. *EMBO J.* **16**: 659–671.

Li M, Phylip LH, Lees WE, Winther JR, Dunn BM, Wlodawer A, Kay J, Gustchina A. 2000. The aspartic proteinase from *Saccharomyces cerevisiae* folds its own inhibitor into a helix. *Nature Struct. Biol.* **7**: 113–117.

Light A, Taniuchi H, Chen RF. 1974. A kinetic study of the complementation of fragments of staphylococcal nuclease. *J. Biol. Chem.* **249**: 2285–2293.

Lin Z, Schwartz FP, Eisenstein E. 1995. The hydrophobic nature of GroEL-substrate binding. *J. Biol. Chem.* **270**: 1011–1014.

Lindner RA, Kapur A, Mariani M, Titmuss SJ, Carver JA. 1998. Structural alterations of alpha-crystallin during its chaperone action. *Eur. J. Biochem.* **258**: 170–183.

Lindner AB, Eshhar Z, Tawfik DS. 1999. Conformational changes affect binding and catalysis by ester-hydrolysing antibodies. *J. Mol. Biol.* **285**: 421–430.

Lo Conte L, Chothia C, Janin J. 1999. The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285**: 2177–2198.

Loidl G, Groll M, Musiol HJ, Huber R, Moroder L. 1999. Bivalency as a principle for proteasome inhibition. *Proc. Natl Acad. Sci. USA* **96**: 5418–5422.

Lorber DM, Shoichet BK. 1998. Flexible ligand docking using conformational ensembles. *Protein Sci.* **7**: 938–950.

Lydakis-Simantiris N, Hutchison RS, Betts SD, Barry BA, Yocum CF. 1999. Manganese stabilizing protein of photosystem II is a thermostable, natively unfolded polypeptide. *Biochemistry* **38**: 404–414.

Lynch WP, Riseman VM, Bretscher A. 1987. Smooth muscle caldesmon is an extended flexible monomeric protein in solution that can readily undergo reversible intra- and intermolecular sulfhydryl cross-linking. A mechanism for caldesmon's F-actin bundling activity. *J. Biol. Chem.* **26**: 7429–7437.

Ma R, Kumar S, Tsai C-J, Nussinov R. 1999. Folding funnels and binding mechanisms. *Protein Engng.* **12**: 713–720.

Mangoni M, Roccatano D, Di Nola A. 1999. Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation. *Proteins* **35**: 153–162.

Markus MA, Hinck AP, Huang S, Draper DE, Torchia DA. 1997. High resolution solution structure of ribosomal protein L11-C76, a helical protein with a flexible loop that becomes structured upon binding to RNA. *Nature Struct. Biol.* **4**: 70–77.

Martin L, Fanarraga ML, Aloria K, Zabala JC. 2000. Tubulin folding cofactor D is a microtubule destabilizing protein. *FEBS Lett.* **470**: 93–95.

Mateu MG, Sanchez del Pino, MM, Fersht AR. 1999. Mechanism of folding and assembly of a small tetrameric protein domain from tumor suppressor p53. *Nature Struct. Biol.* **6**: 190–198.

Matouschek A. 2000. Recognizing misfolded proteins in the endoplasmic reticulum. *Nature Struct. Biol.* **7**: 265–266.

Mayer MP, Schroder H, Rudiger S, Paal K, Laufen T, Bukau B. 2000. Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nat. Struct. Biol.* **7**: 586–593.

McCarty JS, Buchberger A, Reinstein J, Bukau B. 1995. The role of ATP in the functional cycle of the DnaK chaperone system. *J. Mol. Biol.* **249**: 126–137.

Mendoza JA, Campo GD. 1996. Ligand-induced conformational changes of GroEL are dependent on the bound substrate polypeptide. *J. Biol. Chem.* **271**: 16344–16349.

Merck KB, Groenen PJ, Voorter CE, de Haard-Hoekman WA, Horwitz J, Bloemendal H, de Jong WW. 1993. Structural and functional similarities of bovine alpha-crystallin and mouse small heat-shock protein. A family of chaperones. *J. Biol. Chem.* **268**: 1046–1052.

Miller DW, Dill KA. 1997. Ligand binding to proteins: the binding landscape model. *Protein Sci.* **6**: 2166–2179.

Mohana-Borges R, Lima Silva J, de Prat-Gay G. 1999. Protein folding in the absence of chemical denaturants. Reversible pressure denaturation of the noncovalent complex formed by the association of two protein fragments. *J. Biol. Chem.* **274**: 7732–7740.

Moreno FJ, Bagnat M, Lim F, Avila J. 1999. OP18/stathmin binds near the C-terminus of tubulin and facilitates GTP binding. *Eur. J. Biochem.* **262**: 557–562.

Morris GE, Sedgwick SG, Ellis JM, Pereboev A, Chamberlain JS, Nguyen TM. 1998. An epitope structure for the C-terminal domain of dystrophin and utrophin. *Biochemistry* **37**: 11117–11127.

Morshauser RC, Hu W, Wang H, Pang Y, Flynn GC, Zuiderweg ER. 1999. High-resolution solution structure of the 18 kDa substrate-binding domain of the mammalian chaperone protein Hsc70. *J. Mol. Biol.* **289**: 1387–1403.

Moy VT, Florin EL, Gaub HE. 1994. Intermolecular forces and energies between ligands and receptors. *Science* **266**: 257–259.

Murphy KP, Xie D, Garcia KC, Amzel LM, Freire E. 1993. Structural energetics of peptide recognition: angiotensin II/antibody binding. *Proteins* **15**: 113–120.

Najmanovich R, Kuttner J, Sobolev V, Edelman M. 2000. Side-chain flexibility in proteins upon ligand binding. *Proteins* **39**: 261–268.

Nanduri S, Carpick BW, Yang Y, Williams BR, Qin J. 1998. Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation. *EMBO J.* **17**: 5458–5465.

Náray-Szabó G, 1989. Electrostatic complementarity in molecular associations. *J. Mol. Graphics* **7**: 76–81.

Náray-Szabó G. 1994. Analysis of molecular recognition: steric, electrostatic and hydrophobic complementarity. *J. Mol. Recogn.* **6**: 205–210.

Neumann E. 1981. Dynamics of molecular recognition in enzyme-catalyzed reactions. In: *Structural and Functional Aspects of Enzyme Catalysis*, Eggerer H, Huber R (eds). Springer: Berlin; 45–58.

Nicklaus MC, Wang S, Driscoll JS, Milne GW. 1995. Conformational changes of small molecules binding to proteins. *Bioorg. Med. Chem.* **3**: 411–428.

Nieslanik BS, Dabrowski MJ, Lyon RP, Atkins WM. 1999. Stopped-flow kinetic analysis of the ligand-induced coil-helix transition in glutathione S-transferase A1-1: evidence for a persistent denatured state. *Biochemistry* **38**: 6971–6980.

Ninio J. 1975. Kinetic amplification of enzyme discrimination. *Biochimie* **57**: 587–595.

Ninio J. 1977. Are further kinetic amplification schemes possible? *Biochimie* **59**: 759–760.

Northrup SH, Erickson HP. 1992. Kinetics of protein–protein association explained by Brownian dynamics computer simulation. *Proc. Natl Acad. Sci. USA* **89**: 3338–3342.

O'Neil KT, Wolfe HR, Eriksson-Viitanen S, DeGrado KT. 1987. Fluorescence properties of calmodulin-binding peptides reflect alpha-helical periodicity *Science* **236**: 1454–1456.

O'Neil KT, Shuman JD, Ampe C, DeGrado WF. 1991. DNA-induced increase in the alpha-helical content of C/EBP and GCN4. *Biochemistry* **30**: 9030–9034.

Ortiz-Salmeron E, Baron C, Garcia-Fuentes L. 1998. Enthalpy of captopril-angiotensin I — converting enzyme binding. *FEBS Lett.* **435**: 219–224.

Otlewski J, Apostoluk W. 1997. Structural and energetic aspects of protein-protein recognition. *Acta Biochim. Pol.* **44**: 367–387.

Ottiger M, Zerbe O, Guntert P, Wutrich K. 1997. The NMR solution conformation of unligated human cyclophilin A. *J. Mol. Biol.* **272**: 64–81.

Palma PN, Krippahl L, Wampler JE, Moura JJ, 2000. BiGGER: a

A. P. DEMCHENKO

new (soft) docking algorithm for predicting protein interactions. *Proteins* **39**: 372–384.

Pellecchia M, Montgomery DL, Stevens SY, Vander Kooi CW, Feng HP, Gierasch LM, Zuiderweg ER. 2000. Structural insights into substrate binding by the molecular chaperone DnaK. *Nature Struct. Biol.* **7**: 298–303.

Perrett S, Zahn R, Steinberg G, Fersht AR. 1997. Importance of electrostatic interactions in the rapid binding of polypeptides to GroEL. *J. Mol. Biol.* **269**: 892–901.

Petersen JM, Skalicky JJ, Donaldson LW, McIntosh LP, Alber T, Graves BJ. 1995. Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an alpha helix. *Science* **269**: 1866–1869.

Pierpaoli EV, Sandmeier E, Baici A, Schonfeld HJ, Gisler S, Christen P. 1997. The power stroke of the DnaK/DnaJ/GrpE molecular chaperone system. *J. Mol. Biol.* **269**: 757–768.

Plater ML, Goode D, Crabbe MJC. 1996. Effects of site-directed mutations on the chaperone-like activity of αB-crystallin. *J. Biol. Chem.* **271**: 28558–28566.

Prat-Gay G. 1996. Association of complementary fragments and the elucidation of protein folding pathways. *Protein Engng.* **9**: 843–847.

Preuss M, Miller AD. 1999. Interaction with GroEL destabilises non-amphiphilic secondary structure in a peptide. *FEBS Lett.* **461**: 131–135.

Prodromou C, Roe SM, O'Brien R, Ladbury JE, Piper PW, Pearl LH. 1997. Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. *Cell* **90**: 65–75.

Qian M, Spinelli S, Driguez H, Payan F. 1997. Structure of a pancreatic alpha-amylase bound to a substrate analogue at 2.03 A resolution. *Protein Sci.* **6**: 2285–2296.

Raman B, Ramakrishna T, Rao CM. 1995. Temperature dependent chaperone-like activity of alpha-crystallin. *FEBS Lett.* **365**: 133–136.

Randall LL. 1992. Peptide binding by chaperone SecB: implications for recognition of nonnative structure. *Science* **257**: 241–245.

Randall LR, Hardy SJS. 1995. High selectivity with low specificity: how SecB has solved the paradox of chaperone binding. *Trends Biochem. Sci.* **20**: 65–69.

Randall LL, Hardy SJ, Topping TB, Smith VF, Bruce JE, Smith RD. 1998. The interaction between the chaperone SecB and its ligands: evidence for multiple subsites for binding. *Protein Sci.* **7**: 2384–2390.

Reutimann H, Luisi PL, Holmgren A. 1983. Folding of protein fragments: conformational and biological studies on thioredoxin and its fragments. *Biopolymers* **22**: 107–111.

Ritter C, Helenius A. 2000. Recognition of local glycoprotein misfolding by the ER folding sensor UDP-Glucose:glycoprotein glucosyltransferase. *Nature Struct. Biol.* **7**: 278–282.

Rizzo P, Tinello C, Punturieri A, Taniuchi H. 1992. A study of hydrogen exchange of monoclonal antibodies: specificity of the antigen-binding induced conformational stabilization. *Biochim. Biophys. Acta* **1159**: 169–178.

Rosenfeld R, Vajda S, DeLisi C. 1995. Flexible docking and design. *A. Rev. Biophys. Biomol. Struct.* **24**: 677–700.

Rudiger S, Germeroth L, Schneider-Mergener J, Bukau B. 1997a. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J.* **16**: 1501–1507.

Rudiger S, Buchberger A, Bukau B. 1997b. Interaction of Hsp70 chaperones with substrates. *Nature Struct. Biol.* **4**: 342–349.

Rudolph R, Lille H. 1996. In vivo folding of inclusion bodies proteins. *FASEB J.* **10**: 49–56.

Ruiz-Sanz J, de Prat Gay G, Otzen DE, Fersht AR. 1980. Protein fragments as models for events in protein folding pathways: protein engineering analysis of the association of two complementary fragments of the barley chymotrypsin inhibitor 2 (CI-2). *J. Biol. Chem.* **255**: 845–853.

Sandak B, Nussinov R, Wolfson HJ. 1998. A method for biomolecular structural recognition and docking allowing conformational flexibility. *J. Comput. Biol.* **5**: 631–654.

Schweers O, Schonbrunn-Hanebeck E, Marx A, Mandelkow E. 1994. Structural studies of tau protein a Alzheimer paired

helical filaments show no evidence for beta-structure. *J. Biol. Chem.* **269**: 24290–24297.

Shtilerman M, Lorimer GH, Englander SW. 1999. Chaperonin function: folding by forced unfolding. *Science* **284**: 822–825.

Singh K, Zewge D, Groth-Vasselli B, Farnsworth PN. 1996. A comparison of structural relationships among alpha-crystallin, human Hsp27, gamma-crystallins and beta B2-crystallin. *Int. J. Biol. Macromol.* **19**: 227–233.

Slepenkov SV, Witt SN. 1998. Peptide-induced conformational changes in the molecular chaperone DnaK. *Biochemistry* **37**: 16749–16756.

Smulders RHPH, Carver JA, Lindner RA, van Boekel MAM, Bloemendal H, de Jong WW. 1996. Immobilization of the C-terminal extension of bovine alpha A-crystallin reduces chaperone-like activity. *J. Biol. Chem.* **271**: 29060–29066.

Sobolev V, Wade RC, Vriend G, Edelman M. 1996. Molecular docking using surface complementarity. *Proteins* **25**: 120–129.

Sohl JL, Jaswal SS, Agard DA. 1998. Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature* **395**: 817–819.

Soucheta M, Pouponb A, Callebautc I, Legera I, Mornonc J, Brila A, Calmelsal TP. 2000. Functional specificity conferred by the unique plasticity of fully alpha-helical Ras and Rho GAPs. *FEBS Lett.* **477**: 99–105.

Sparrer H, Lilie H, Buchner J. 1996. Dynamics of the GroEL-protein complex: effects of nucleotides and folding mutants. *J. Mol. Biol.* **258**: 74–87.

Spolar RS, Record T. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**: 777–784.

Stanfield RL, Wilson IA. 1994. Antigen-induced conformational changes in antibodies: a problem for structural prediction and design. *Trends Biotechnol.* **12**: 275–279.

Stella L, Di Iorio EE, Nicotra M, Ricci G. 1999. Molecular dynamics simulations of human glutathione transferase P1-1: conformational fluctuations of the apo-structure. *Proteins* **37**: 10–19.

Stoldt M, Wohnert J, Ohlenschlager O, Gorlach M, Brown LR. 1999. The NMR structure of the 5S rRNA E-domain-protein L25 complex shows preformed and induced recognition. *EMBO J.* **18**: 6508–6521.

Strausberg S, Alexander P, Wang L, Schwarz F, Bryan P. 1993. Catalysis of a protein folding reaction: thermodynamic and kinetic analysis of subtilisin BPN' interactions with its pro-peptide fragment. *Biochemistry* **32**: 8112–8119.

Taniuchi H, Anfinsen CB. 1971. Simultaneous formation of two alternative enzymology active structures by complementation of two overlapping fragments of staphylococcal nuclease. *J. Biol. Chem.* **246**: 2291–2301.

Taniuchi H, Parker DS, Bohnert JL. 1977. Study of equilibration of the system involving two alternative, enzymatically active complementing structures simultaneously formed from two overlapping fragments of staphylococcal nuclease. *J. Biol. Chem.* **252**: 125–140.

Taniuchi H, Parr GR, Juillerat MA. 1986. Complementation in folding and fragment exchange. *Meth. Enzymol.* **131**: 185–217.

Tasayco ML, Carey J. 1992. Ordered self-assembly of polypeptide fragments to form nativelike dimeric *trp* repressor. *Science* **255**: 594–597.

Terry CJ, Damas AM, Oliveira P, Saraiva MJ, Alves IL, Costa PP, Matias PM, Sakaki Y, Blake CC. 1993. Structure of Met30 variant of transthyretin and its amyloidogenic implications. *EMBO J.* **12**: 735–741.

Theis K, Chen PJ, Skorvaga M, VanHouten B, Kisker C. 1999. Crystal structure of UvrB, a DNA helicase adapter for nucleotide excision repair. *EMBO J.* **18**: 6899–6907.

Tsai C-J, Ma B, Nussinov R. 1999. Folding and binding cascades: shifts in energy landscapes. *Proc. Natl Acad. Sci. USA* **96**: 9970–9972.

Varani G. 1997. RNA-protein intermolecular recognition. *Acc. Chem. Res.* **30**: 189–195.

Verbitsky G, Nusinov R, Wolfson H. 1999. Flexible structural comparison allowing hinge-bending, swiveling motions. *Proteins* **34**: 232–254.

Von Germar F, Galan A, Llorca O, Carrascosa JL, Valpuesta JM,

Mantele W, Muga A. 1999. Conformational changes generated in GroEL during ATP hydrolysis as seen by time-resolved infrared spectroscopy. *J. Biol. Chem.* **274**: 5508–5513.

Von Hippel PH, Berg OG. 1989. Facilitated target location in biological systems. *J. Biol. Chem.* **264**: 675–678.

Wade RC, Gabdoulline RR, Luty BA. 1998. Species dependence of enzyme-substrate encounter rates for triose phosphate isomerases. *Proteins* **31**: 406–416.

Wallon G, Rappsilber J, Mann M, Serrano L. 2000. Model for stathmin/OP18 binding to tubulin. *EMBO J.* **19**: 213–222.

Walls PH, Sternberg MJ. 1992. New algorithm to model protein-protein recognition based on surface complementarity. Applications to antibody-antigen docking. *J. Mol. Biol.* **228**: 277–297.

Wang C-C, Tsou C-L. 1998. Enzymes as chaperones and chaperones as enzymes. *FEBS Lett.* **425**: 382–384.

Wang H, Kurochkin AV, Pang Y, Hu W, Flynn GC, Zuiderweg ER. 1998. NMR solution structure of the 21 kDa chaperone protein DnaK substrate binding domain: a preview of chaperone-protein interaction. *Biochemistry* **37**: 7929–7940.

Wang Z, Feng Hp, Landry SJ, Maxwell J, Gierasch LM. 1999a. Basis of substrate binding by the chaperonin GroEL. *Biochemistry* **38**: 12537–12546.

Wang J, Kollman PA, Kuntz ID. 1999b. Flexible ligand docking: a multistep strategy approach. *Proteins* **36**: 1–19.

Watanabe M, Blobell G. 1995. *Proc. Natl Acad. Sci. USA* **92**: 10133–10136.

Weikl T, Abelmann K, Buchner J. 1999. An unstructured C-terminal region of the Hsp90 co-chaperone p23 is important for its chaperone function. *J. Mol. Biol.* **293**: 685–691.

Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT Jr. 1996. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* **35**: 13709–13715.

Weiss MA, Ellenberger T, Wobbe CR, Lee JP, Harrison SC, Struhl K. 1990. Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **347**: 575–578.

Wendt H, Leder L, Harma H, Jelesarov I, Baici A, Bosshard HR. 1997. Very rapid, ionic strength-dependent association and folding of a heterodimeric leucine zipper. *Biochemistry* **36**: 204–213.

Wendt H, Thomas RM, Ellenberger T. 1998. DNA-mediated folding and assembly of MyoD-E47 heterodimers. *J. Biol. Chem.* **273**: 5735–5743.

Wetlaufer DB. 1981. Folding of protein fragments. *Adv. Protein Chem.* **34**: 61–92.

Winkler FK, Banner DW, Oefner C, Tsernoglou D, Brown RS, Heathman SP, Bryan RK, Martin PD, Petratos K, Wilson KS. 1993. The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12**: 1781–1795.

Wintrode PL, Privalov PL. 1997. Energetics of target peptide recognition by calmodulin: a calorimetric study. *J. Mol. Biol.* **266**: 1050–1062.

Wright PE, Dyson HJ. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**: 321–331.

Yang H, Huang S, Dai H, Gong Y, Zheng C, Chang Z. 1999. The *Mycobacterium tuberculosis* small heat shock protein Hsp 16.3 exposes hydrophobic surfaces at mild conditions: conformational flexibility and molecular chaperone activity. *Protein Sci.* **8**: 174–179.

Yoo SH. 1995. pH- and Ca(2 +)-induced conformational change and aggregation of chromogranin B. Comparison with chromogranin A and implication in secretory vesicle biogenesis. *J. Biol. Chem.* **270**: 12578–12583.

Zahn R, Spitzfaden C, Ottiger M, Wuthrich K, Pluckthun A. 1994. Destabilization of the complete protein secondary structure on binding to the chaperone GroEL. *Nature* **368**: 261–265.

Zahn R, Liu A, Luhrs T, Riek R, von Schroetter C, Garcia FL, Billeter M, Calsolai L, Wider G, Wuthrich K. 2000. NMR solution structure of the human prion protein. *Proc. Natl Acad. Sci. USA* **97**: 145–150.

Zhang H, Zhao D, Revington M, Lee W, Jia X, Arrowsmith C, Jardetzky O. 1994. The solution structures of the *trp* repressor-operator DNA complex. *J. Mol. Biol.* **238**: 592–614.

Zhang C, Chen J, DeLisi C. 1999. Protein-protein recognition: exploring the energy funnels near the binding sites. *Proteins* **34**: 255–267.

Zhao D, Arrowsmith CH, Jia X, Jardetzky O. 1993. Refined solution structures of the Escherichia coli *trp* holo- and aporepressor. *J. Mol. Biol.* **229**: 735–746.

Zhu X, Zhao X, Burkholder F, Gragerov A, Ogata CM, Gottesman ME, Hendrickson WA. 1996. Structural analysis of substrate binding by molecular chaperone DnaK. *Science* **272**: 1606–1614.

Zidek L, Novotny MV, Stone MJ. 1999. Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nature Struct. Biol.* **6**: 1118–1121.