Turncoat polypeptides: We adapt to our environment

Héctor Zamora-Carreras,^{[a]#} Beatriz Maestro,^{[b]#}
Jesús M. Sanz^{*[b]} and M. Angeles Jiménez^{*[a]}

Cover text: French statesman Talleyrand showed an astonishing ability to adapt to the circumstances in the political scenario of the pre- and post-revolutionary France. He remained as the classical example of a fruitful turncoat in politics. In a similar way, some polypeptides possess a remarkable capacity to accommodate to changes occurring in their environment or by the presence of triggering agents. These polypeptides “turn their coats”, i.e. rearrange their structural features, in order to gain stability under the new circumstances.

WILEY-VCH

Abstract: A common interpretation of Anfinsen's hypothesis states that one amino acid sequence should fold into a single native ordered state, or a highly similar set thereof, coinciding with the global minimum in the folding energy landscape, which in turn is responsible for the protein's function. However, this classical view is challenged by many proteins and peptide sequences, which can adopt exchangeable, significantly dissimilar conformations that even fulfill different biological roles. We review the similarities and differences of concepts related to these proteins, mainly chameleon sequences, metamorphic proteins and switch peptides, and denote all of them "turncoat" polypeptides. Besides adding a twist to the conventional view of protein folding, the lack of structural definition adds a clear versatility to the activity of proteins and can be used as a tool for protein design and further application in biotechnology and biomedicine.

1. Introduction

Very often the macroscopic world reflects events that occur in the microscopic level, and vice versa. Charles Maurice de Talleyrand (1754-1838) is considered as one of the most influential politicians and diplomats in France history. He became Agent-General of the Clergy and later worked as diplomat under environments as different (and contradictory) as the Louis XVI regime, the French Revolution years, the Napoleon Bonaparte's empire and the Louis XVIII and Louis-Philippe reigns. Yet, he played throughout an essential role in the foreign politics of his country using an ability to adapt barely matched in history. Does this example of political versatility have a molecular counterpart? Perhaps we might find it (*"toute proportion gardée"*) in the protein realm.

Proteins are essential building blocks of living organisms. They constitute a vast and diverse set of biomolecules, which not only play essential roles in every biological process, but are also involved in the development of numerous diseases. For decades, the study of proteins has been a major topic in science given the relevance and complexity of these systems. The cornerstone of protein science is the so-called "folding problem", a concept comprising two main questions: i) what are the thermodynamic principles behind the folding of a given amino acid sequence? and ii) what are the kinetic principles governing the folding

pathways followed by a protein? The answer is imperative to predict, in an accurate way, the native structure of a protein from its sequence. ^[1]

Héctor Zamora-Carreras earned his Ph.D. in Biochemistry in 2016 (Universidad Complutense de Madrid, Spain). His doctoral project was focused on NMR techniques applied to the study of interactions between peptides/proteins and other biomolecules. He spent two doctoral stays in Prof. Anne S. Ulrich's group at the Karlsruhe Institute for Technology (KIT, Germany), where he employed novel solid state NMR techniques to elucidate the mechanism of membrane-active peptides. Currently, he works on NMR applied to the study of polypeptide systems at Instituto de Química Física Rocasolano (IQFR-CSIC), in Madrid.

Beatriz Maestro obtained her Ph.D. in Pharmacy in 2001 (Universidad Complutense de Madrid, Spain). At the Miguel Hernández University (Elche, Spain, 2001-2009 and 2013-2018), she carried out studies on protein structure, engineering and function, and was awarded the SALVAT prize (2006) for the best medical sciences project. She also worked at the University of Alicante (Spain, 2009-2012), investigating on microbial fuel cells and enzymatic electrodes. She is now working at CSIC, in Madrid, on the development of new antibiotics that involve the use of switch peptides.

Jesús M. Sanz received his Ph.D. in Chemistry in 1992 (Universidad Complutense de Madrid, Spain). He spent a postdoctoral stay at the Centre for Protein Engineering in Cambridge (U.K.) (1992-1994), and a second postdoc in the CSIC (Madrid, Spain, 1994-1997). He was Professor at the Miguel Hernández University (Elche, Spain) until 2018, when he obtained a full-position at CSIC in Madrid. The research lines he currently leads involve the structural study of proteins and the functionalization of nanoparticles and bioplastics, with the aim of designing new antimicrobial and biotechnological tools.

- [a] Dr H. Zamora-Carreras,* Dr M. A. Jiménez*
Instituto de Química-Física Rocasolano (IQFR)
Consejo Superior de Investigaciones Científicas (CSIC)
Serrano 119, 28006-Madrid, Spain
E-mail: majimenez@iqfr.csic.es
- [b] Dr. B. Maestro,* Dr. J. M. Sanz*
Centro de Investigaciones Biológicas (CIB)
Consejo Superior de Investigaciones Científicas (CSIC)
Ramiro de Maeztu 9, 28040-Madrid, Spain
E-mail: jmsanz@cib.csic.es

* These authors contributed equally to the work

M. Angeles Jiménez got her Ph.D. in Chemistry in 1987 (Universidad Autónoma de Madrid, Spain). Her Thesis project performed at Prof. M. Rico's group (CSIC, Madrid), involved NMR studies of peptides. After a postdoc at Univ. Leicester, UK (1988-1989) studying protein/ligand interactions by NMR, she obtained a full-position at CSIC, Madrid, Spain. Since then, her research fields are peptide design, in particular β -hairpins and therapeutics, as well as protein structure and recognition, aimed to establish molecular basis of protein functionality, using solution NMR as main tool.

In the 1960s, Anfinsen issued his celebrated thermodynamic hypothesis. This states that, under native conditions, a protein uniquely folds to the three-dimensional structure with the lowest Gibbs energy of the system, which is fully encoded in its amino acid sequence. For many years, this hypothesis has been insistently oversimplified to the “one sequence, one structure, one function” canon.^{[12] [13] [14]}

During the last decades, however, biochemical, biophysical, and computational results have challenged this classical view. In fact, many polypeptides undergo massive spatial reorganizations (far beyond the limited rearrangements that are commonly termed as “conformational changes”) in response to a diversity of factors. This feature, that is, the capacity to change between different folded states or from unfolded to folded, adds an unsuspected versatility for protein function, as big structural changes qualify a protein to perform highly different activities.^{[12] [15]}

Many labels have been coined to refer to these proteins, but sometimes the concepts are unclear, being difficult to discern among them. To clarify this question, we define herein some of these terms:

Figure 1. Examples of proteins defying Anfinsen's canon. A) Metamorphic protein XCL1: $2\alpha+3\beta$ (PDB: 1J9O) and 4β structures (PDB: 2N54) of XCL1.^[2] B) Transformer protein: all- α (PDB: 2OUG) and tall- β (PDB: 2LCL) structures of the C-terminal domain of RfaH protein.^[3] C) Morpheein: a scheme of the two different oligomeric equilibria of PBGS protein.^[4] D) Chameleon sequence: The different structures of a chameleon 8-mer sequence are highlighted in red (PDB: 2Q0Y and 3S30).^{[5] [6]} E) Conformational switch: a region of the EF-Tu protein, whose conformation depends on the bound ligand (random coil with GDP/PDB:4ZV4; and helical with GTP/PDB:2C77).^[7] F) Switch peptide: A 14-mer peptide derived from autolysin LytA, which forms a native-like, stable β -hairpin in aqueous solution, and becomes helical in DPC micelles.^[8] G) Intrinsically disordered protein (IDP): the 50S ribosomal protein L11 (L11-C76) has a dynamic, disordered loop in its structure (PDB: 1FOX; loop highlighted in yellow), that becomes partially structured upon RNA binding.^[9] H) β -amyloid peptide, containing a helix and unstructured regions (PDB: 2LFM), which can misfold to a mostly- β structure that oligomerises to form pathogenic amyloid fibrils (PDB: 5OQV).^{[10] [11]}

- **Metamorphic protein** (Fig. 1A): a protein with two or more native structures that can reversibly interconvert between them. Generally, each native structure shows different functionality.^{[12] [16] [17] [18]} The term “transformer protein” has been proposed for extreme cases of metamorphism in which the structures are completely different, such as the C-terminal domain of RfaH (CtD) (Fig. 1B).^[14]

- **Morpheein** (Fig. 1C): a protein in which structural and oligomerization changes are associated. A morpheein protein exists as a dynamic equilibrium of distinct oligomeric states, and the conversion from one to another requires dissociation into subunits and a conformational change in each subunit. This structure rearrangement does not necessarily involve changes in the core protein fold, but small backbone and side chain angle adjustments.^{[4] [19]}

- **Chameleon sequence** (Fig. 1D): an amino acid sequence that can adopt different secondary structures depending on the protein in which it is found.^{[14] [20]}

- **Conformational switch** (Fig. 1E): a delimited protein region capable of acquiring different secondary structures upon changes in the environment (pH variations, ligand binding, post-translational modifications).^[21] Some conformational switches are located at regions linking protein domains, and act as hinges leading to the so-called open and closed states of proteins.^[22]

- **Switch peptide** (Fig. 1F): an isolated peptide whose structure is totally modified when triggered by a specific change in the environment (pH, redox conditions, temperature, light...). It can be regarded as an isolated conformational switch.^{[21] [23]}

- Intrinsically disordered protein (IDP; Fig. 1G): a protein possessing no well-defined three-dimensional structure, but rather adopting an ensemble of conformations in solution that are functional.^[24] They frequently contain certain regions, called MorFs (molecular recognition features), capable of undergoing disorder-to-order transitions upon binding to partners.^[25] Interestingly, these kinds of MorFs sequences may adopt different structures when interacting with different partners.^[26]
- Amyloidogenic or amyloid-forming protein (Fig. 1H): A protein, which mis-folds upon certain conditions, and the mis-folded structure experiences irreversible aggregation forming amyloid fibrils. They are mostly related to neurodegenerative diseases, but some functional amyloids have also been described.^{[27] [28] [29]}

Hereinafter, we will focus on three closely related types of sequences: chameleon sequences, metamorphic proteins, and switch peptides. We will use the generic term "turncoat" polypeptides to encompass all of them.

2. Chameleon sequences

It has been long known that amino acid composition and sequence in a protein stretch do not convey sufficient information to define (and predict) its secondary structure. The importance of non-local interactions was first brought to light in the mid-1980s, when identical amino acid sequences were found to display different secondary structures in different proteins.^[30] These kinds of sequences, denoted chameleon sequences (Fig. 1D), are defined as strings of identical amino acids arranged in the same order that can adopt different conformations in protein structures.^{[20] [6, 31]}

The chameleon sequences usually do not have strong α -helical or β -sheet propensities, and thus they can adapt to both structures depending on the protein they are part of. Since the first chameleon sequences were reported,^[30] the length of the described chameleon sequences has increased, currently varying from 6 to 10 residues, or even 11 residues in a designed sequence inserted in a bigger domain.^[20] A large database containing thousands of chameleon sequences (<http://prodata.swmed.edu/chseq>) has been built.^{[6, 31] [32] [33] [34] [35]}

According to some studies,^[32] the conformation of a chameleon sequence is defined by a combination of the flanking sequences and long-range interactions. Indeed, some authors have coined the term "conformational contagion", meaning that the folding of a chameleon sequence and, by extension, the folding of a sequence within a protein, is affected by the influence of the sequentially neighboring secondary structures.^[33]

Chameleon sequences might play important functional roles. Many chameleon sequences have been shown to be involved in biological events.^[6] Moreover, it has been suggested that

chameleon sequences may be involved in the conservation of protein structural fold and functional diversity in alternative splicing forms. For instance, if an alternative spliced isoform of a protein is shorter due a sequence deletion, the integrity of secondary structure elements can be affected. But, if the flanking sequence is a chameleon sequence, it could still adopt the adequate structure, filling the void, and maintaining the main protein fold.^[33] In this sense, it is easy to understand that the presence of chameleon sequences within the protein structure may favor evolutionary alterations in that structure. The plasticity of chameleon sequence confers higher adaptability to the protein, opening new paths to reach more evolved forms of the protein.^[36]

The presence of chameleon sequences in human proteins plays an important role in many diseases. Many studies have evinced the relationship between chameleon sequences and the induction of misfolding-based pathological processes, such as neurodegeneration induced by amyloid fibril formation. In this context, understanding the behavior of chameleon sequences is essential to gain knowledge about some fatal human diseases, such as Alzheimer's, Parkinson's, Huntington's, Creutzfeld-Jakob disease, etc. The molecular basis of these pathological conditions involves an initial α -helix to β -sheet transition upon interaction of the neuronal membranes with monomers of amyloid proteins.^{[37] [38]}

2.1. Prediction of chameleon sequences

The biotechnological potential of engineering structurally versatile regions in proteins amenable to be rationally controlled has prompted research to elucidate rules that govern the acquisition of heterogeneous conformational states by a certain sequence. The challenge resides in the difficulty to predict the structure of sequences that have alternative conformations differing little in stabilization energy and, therefore, are very sensitive to changes in their environment.

Discrepancies in secondary structure predictions of a protein may unveil subsequences whose intrinsic propensities differ from the structures found in the entire polypeptide.^[39] Such discrepancies could be thought, in principle, to be responsible,^[40] at least in part, for the level of accuracy achieved by current prediction methods, which is 85 % at most for a 3-state assignment (i.e. helix, extended and coil conformations).^[12] Secondary structure predictors are usually classified in three classes or "generations", according to their theoretical basis and procedures (see ^[41] and ^[42] for extensive reviews). First generation methods analyze individual propensities of amino acids on an almost purely statistical basis, whereas second generation procedures take into account the neighboring residues and carry out the prediction in sliding segments (windows) of amino acids. Finally, third generation advanced algorithms benefit from the increase in elucidated three-dimensional structures, make use of evolutionary information through conservation analyses of multiple sequence alignments and incorporate long-range interactions in their calculations.

Therefore, it is hardly surprising that the structure of chameleon sequences in their native proteins can be mostly accurately predicted using these profile-based methods, even with similar accuracy to the rest of sequences.^{[6, 33, 43] [44] [45]} However, for this same reason, they are of limited use to discover fold-switching regions. On this basis, Porter and Looger^[40] developed a searching procedure that employed "incorrect" secondary structure predictions (carried out with SPIDER 2,^[46] a method which does not employ long-range sequence information), together with the estimation of independent folding cooperativity. In this way, a relatively high number of proteins (ranging from 0.5 to 4 % of all members of the Protein Data Bank database), were identified as prone to display alternative folds.

Regarding amino acid composition, hydrophobic residues such as leucine, isoleucine, valine, and alanine are found with high frequency in chameleon sequences.^[35, 45] Moreover, these structures contain a mix of amino acids that strongly favor both α -helix and β -strand conformations,^[33, 45, 47] while secondary structure breakers such as proline are in turn underrepresented.^[45]

3. Metamorphic proteins

Many proteins able to adopt multiple folded conformations under native conditions have been reported.^{[17, 48] [49]} In some cases the protein possesses an energy landscape with two or more local minima separated by relatively low energy barriers, and thus it can switch reversibly among different states that are native topologies by themselves. These proteins have been recently termed metamorphic proteins (Fig. 1A).^[16]

It is not clear yet whether metamorphism is an exclusive characteristic of some proteins, or it is a more widespread feature. The different native topologies of metamorphic proteins show distinct functionalities, and this fact points to a possible biological role. It can be interpreted as a way of post-translational regulation of the protein activity by a mechanism of mutual exclusion. That is, when folded in one of the native states, the protein can only perform the activity associated to that topology. The switch in the protein structure, and hence in protein function, usually takes place as a result of changes in the local physiological conditions, although spontaneous interconversion under native conditions has been reported as well.^[17]

The origin of metamorphic proteins remains unclear, but some authors suggest that they can represent evolutionary intermediates. That is, they are proteins immersed in the process of evolving from an old structure to a new one.^[50] Other authors see these proteins as evolved proteins that have acquired the ability of metamorphosing in order to gain the plasticity required to accomplish multiple functions or allow new regulatory mechanisms.^[12]

Metamorphic proteins are usually folded proteins with low thermodynamic stability (less than 2–3 kcal·mol⁻¹), and sometimes they show disordered regions that play an important role in the structural transition. These transitions tend to be susceptible to environmental changes, and imply large-scale conformational changes with substantial modifications in the backbone ϕ/ψ angles of structural motifs or even entire domains. In addition, metamorphic proteins usually have latent binding sites that become more accessible when conformational switching takes place.^[51]

The specific structural changes described in the metamorphic proteins reported so far are quite diverse. For instance, the IscU protein has two native structures: one with 4 α -helices and 3 β -strands, which is able to interact with the co-chaperone HscB, and the other one, a dynamically disordered structure that binds to the IscS cysteine desulfurase.^[52] In the case of CLIC1, a chloride ion channel protein, an all- α N-terminal domain acts as a channel protein, but it can also show an stable $\alpha 3+\beta 4$ arrangement of still unknown activity.^[53] Another example is the RfaH protein (defined as a transformer protein; Fig. 1B), whose C-terminal domain possesses an all- β open structure, acquiring the capability to recruit the RNA polymerase and the ribosome, and an all- α closed structure that auto-inhibits the recruitment to the selected targets.^[3b, 14] In the case of XCL1, a chemokine showing the structural arrangement typical of this class of proteins, it can experience a structural transition to a β -sheet form to gain glycosamino-glycan binding capacity.^[48] Other reported metamorphic proteins include Mad2,^[54] HIV-1 reverse transcriptase,^[55] and KaiB.^[56]

4. Switch peptides

Switch peptides are short amino acid sequences that undergo a structural transition upon changes in the environmental conditions (Fig. 2). The transition can occur between an unfolded and a folded state or between two different folded conformations. In both cases, the folded states can be either monomeric or oligomeric. Some examples of folded/unfolded transitions may involve α -helices,^[57] β -hairpins,^[58] and self-assembled hydrogel β -sheets.^[59] Nearly all instances of transitions between two folded states involve at least one oligomeric state, e.g. a soluble monomeric α -helix to self-associated oligomeric β -sheet,^[60] two coiled-coil helices with different registers^[61] or a dimer of coiled-coil helices to a helical-hairpin.^[62] Nevertheless, a monomeric peptide, which exhibits a β -hairpin/ α -helix transition (Fig. 1G), has been recently reported.^[8] In all the cases, the initial state is perturbed upon variations in the surrounding conditions leading to the adoption of an alternative stable conformational state. This alternative topology is usually driven by weak propensities present in the sequence, and dictated by the peptide nature. A very important feature of switch peptides is the reversibility of the structural transition. The original topology is recovered when the conditions revert to the initial state.^[63]

There is a vast number of reported switch peptides, which can be classified in different families depending on the cause that triggers the structural transition:

- **Temperature switches:** it is known that thermal activation can control and direct the conformation of systems with inherent structural ambiguity. In this regard, temperature has been exploited to generate α -to- β switches. Some temperature-dependent switches based on α -helical coiled-coil motifs has been reported,^[64] as well as other polypeptides that undergo conversions from helix-turn-helix conformations to β -sheet rich amyloids.^[65]

- **Polarity switches:** Coulombic interactions are a major contributor to protein folding and stability, and thus, conformational switching can sometimes be triggered by the modulation of charges present in the amino acid side chains, or by changes in the polarity, ionic strength or pH of the solvent.^[66]

- **Photo-switches:** some designed peptides are able to experience conformational switching thanks to the presence of chromophoric moieties that change their configuration or constitution upon absorption of light of a certain wavelength. Photo-switchable peptides with transitions from an unstructured or a β state to a helical conformation upon irradiation have been reported.^{[67] [68]}

- **Redox switches:** a few amino acids have redox properties (Cys, Met), and changes in their oxidation state can produce conformational transitions. The most usual way to induce a structural switching by redox perturbations is to modify the hydrophobic/hydrophilic amino acid balance by oxidizing the sulfur moiety of Met or Cys side chains, generally to sulfoxides or sulfinic acid, respectively. In both cases, the final product has a higher polarity and is more hydrophilic.^[68-69]

- **Metal-ion switches:** various peptides can bind to metal ions, and a specific secondary structure is stabilized upon binding. For instance, formation of α -helix is typically favored if metal-binding groups are located in positions i and $i+4$ or $i+5$, whereas β structures or turns can be induced when the separation between binding sites is more than five residues long.^[70] The most frequent metal-binding groups are N-heterocycles, such as imidazole, pyridine and bipyridine, and for this reason, His residues are common in metallopeptides. Sometimes other coordinating groups are introduced in the peptide structure, such as crown ethers or phosphano serine.^[71] The effect of metal coordination in peptides can be affected by pH variations or changes in the redox conditions.^[60, 68]

- **Other switches:** other strategies can be explored to produce new peptide switches. Switch peptides based on X-N acyl group migration has been described.^[58] Furthermore, a peptide derived from the choline-binding domain of pneumococcal autolysin LytA has been reported to undergo a reversible transition from a β -hairpin structure in aqueous

solution to a well-defined, stable α -helix in the presence of zwitterionic and anionic detergent micelles.^[8]

The study of switch peptides is a useful and simple tool to unveil the relationship between the primary structure and the conformational and physicochemical properties of peptides and proteins, and ultimately, to understand the impact of these features on biological processes.^[23] They are also attracting much interest by their potential applications in Biotechnology and Biomedicine (see section 6).

Figure 2. Switch peptides: Examples of transitions and triggering agents.

5. Design of turncoat polypeptides

Early efforts on *de novo* design of switch peptides contemplated simple systems that could respond to external conditions such as pH, temperature and solvent composition (Fig. 2). Mutter *et al.*^[72] designed a 16-aa amphiphilic β -sheet peptide, whose sequence alternated hydrophobic and hydrophilic residues, which were then subjected to substitutions. In this way, they achieved the stabilization of additional amphiphilic α -helical conformations in equilibrium with the original β -sheet. Shifting between those two conformations could be modulated by pH or cosolvents, such as trifluoroethanol (TFE). On the other hand, computational approaches are necessary for most specific outlines. As an example, Kuhlman and Ambroggio^[21] developed an algorithm employing a Monte Carlo search with simulated annealing and RosettaDesign to simultaneously favor two conformations of a peptide (a trimeric coiled-coil and a 2Cys-2His zinc finger-like fold), whose inter-conversion could be modulate by pH or by the presence of transition metals.

Nevertheless, most remarkable results have been obtained by creating alternate folds in a polypeptide through the identification of a chameleon (or chameleon-prone) sequence and the engineering of a switch in its conformation by site-directed mutagenesis. As early as in 1994, the so-called Paracelsus challenge was issued to completely change the fold of a

polypeptide by replacing less than 50 % of its residues.^[73] This was soon fulfilled^[74] and even surpassed, as it has been shown afterwards that only as little as one mutation could suffice to successfully convert an IgG-binding, 4 β + α fold into an albumin-binding, 3- α structure.^[75] On the other hand, as stated by Guo *et al.*,^[33] the chameleon-flanking regions are key in determining its final secondary structure within a particular protein. This phenomenon has been used to rationally design a change of α -helical sequences to β -strands by incorporating helix breakers such as proline at the ends.^[76]

Protein engineering may also be useful in more complex systems through the merging of consensus sequences belonging to different motifs. For instance, Hori and Sugiura^[77] introduced as little as four mutations to selected positions in the mostly α -helical *Antennapedia* homeodomain to transform it into a zinc-finger (mixed α/β) structure, so that the interchange between the two conformations could be reversibly achieved upon addition of Zn^{2+} . A similar approach was followed by Woolfson,^[78] using an α -helical coiled-coil trimer as the starting point. It is noteworthy that helical coiled-coils are present in an important number of reports, as their structure and energetics have been profoundly studied and provide an additional stabilization to the designed α -helix by self-oligomerization. In this line, Kokschi^[79] worked on a 26-aa α -helical coiled-coil dimer and substituted the solvent-exposed b, c and f positions in the heptad repeat (the least involved in helix dimerization; Fig. 3) by more hydrophobic, β -favoring residues. Furthermore, some histidine residues were introduced in selected positions to create new metal-binding sites in the intended β conformation. As a consequence, the peptide was able to switch between the α and β conformations in response to the presence of divalent metals, such as Zn^{2+} or Cu^{2+} .

Figure 3. Helical wheel representation for a heptad repeat in a helical coiled-coil dimer. The nomenclature for a-f positions is shown. Hydrophobic residues are shown in pink; red and blue indicate anionic and cathionic residues, respectively. White residues are solvent-exposed. Arrows symbolize interactions. Figure adapted from Pagel & Kokschi.^[80]

Finally, other systems have been described that respond to changes in temperature or the redox environment. Woolfson^[64a] described a peptide system in which transition from an α -helical coiled-coil to a β -hairpin motif is induced by temperature, through the incorporation of β -sheet preferring residues in the f position of the heptad repeat (Figure 3). Moreover, Gellman^[81] reported the change from a β -sheet to an aggregating α -helix in a methionine-rich peptide caused by changes in the oxidation state of this amino acid.

6. Perspectives in biotechnology and biomedicine for turncoat polypeptides

Polypeptides showing a global fold-switching behavior (including unfolding-folding events) in response to a controllable signal are of great interest for the development of novel biosensors as well as “smart” drugs that interact with their surroundings.^[12, 82]

In the field of drug discovery, the dynamic equilibrium between two native protein conformations opens up the possibility to control it at convenience in accordance with the external environment.^[12] For instance, a metamorphic or morpheic protein that selectively presents a therapeutic activity in only one of its states could be administered in the alternative non-active fold to reduce any side effects so that the therapeutic conformation is acquired only at the target as triggered by specific conditions, such as pH or temperature. Besides, these proteins can also be considered as vehicles in drug delivery, as the enclosed drug could be disassembled *in situ* thanks to the conformational change. Other therapeutic applications include the development of new drugs against one specific conformation responsible of the pathology^[83] or the possibility to shift the equilibrium to the non-pathological folded state.^[12]

Regarding biosensor development, metamorphic proteins might provide at once both the recognition and transduction functions, reducing the need for complex and expensive detection systems that are necessary in current devices. This has fostered the investigations to rationally design switchable proteins that recognize a ligand and transduce a signal.^[84]

The knowledge and affordable complexity of Zn^{2+} binding sites has favored the appearance of one of the first biosensors in this field. The zinc-finger motifs show binding to several cations, mostly divalent transition metal ions.^[85] In the absence of metal, zinc fingers are disordered, whereas if appropriate metal cations are present, the coordination residues lead to a defined tertiary structure, including a cluster of conserved hydrophobic residues. Walkup and Imperiali^[86] replaced one of these residues by an environment-sensitive fluorophore to monitor the metal-binding event. A similar strategy was followed independently by Godwin and Berg,^[87] but using in this case two fluorophores and the resonant energy transfer as signal transduction mechanism. Both devices are capable of quantifying Zn^{2+} concentrations at nanomolar level.

The challenge of developing a switchable protein biosensor is that the sensing probe must undergo an intense change upon binding. Conceptually, the most extensive transformation in a protein is that from the unfolded to folded conformation. Therefore, a method can be envisaged to mutate a protein in such a way that its folded state would be more stable than the denatured state only if the ligand provides the necessary bonds, so it should experience folding strictly upon ligand binding. Based in these basic protein folding principles, Kohn and Plaxco rationally modified, by successive amino acid deletion (remote from its target binding site), the conventional folded FynSH3 small protein until it became unfolded, in such a way that it undergoes a ligand-induced folding upon recognition of the p85 α -2 peptide part of the regulatory subunit of the phosphatidylinositol 3-kinase, and described an optical sensor for detecting this specific binding even in complex contaminant-riddled samples.^[88]

Other authors have taken advantage of natural intrinsic disordered proteins to develop quantitative biosensor systems. The specific interaction between the intrinsically disordered polypeptide regions of the BRCA1 and p53 proteins induces the folding of both. This structural change can be used to detect p53 by quantitative quenching using both intrinsic tryptophan residues and extrinsic fluorescent probes, the latter being strategically inserted into environmentally-sensitive amino acid positions and detected upon binding on the visible region.^[89]

Figure 4. Schematic representation showing the N and N' states of the "Alternate Frame Folding".^[90] (A) The sequence and native structure of a protein are schematically represented, and a yellow triangle indicates a ligand-binding site in the C-terminal domain. (B) The C-terminal segment containing the binding site (in blue), which is duplicated, and inserted at the N-terminus (in orange), with a spacer loop. The original binding site in the C-terminal domain is mutated to inactivate it. Ligand binding forces the alternative N'-form folding, and this rearrangement can be reported by inserted fluorescent groups (indicated by stars) whose relative proximities change as a result of this transition.

A complete unfolding of the protein is not always necessary to have a powerful biosensor based on turncoat proteins. Stratton *et al.* developed an elegant, general method, designated "Alternate Frame Folding (AFF)", that may be applied to numerous proteins and by which protein conformational changes can be engineered in response to a specific signal and coupled to a report output.^[90] In essence, the general approach consists in the duplication of a segment located in the C-terminus, which is appended to the N-terminus, or vice versa (segment chosen depends on the location of key functional residues). This segment must contain at least one critical ligand-binding amino acid, and should finish in an exposed loop. Since the duplicated segments compete to interact with the rest of the protein, this generates two alternative, mutually exclusive conformations: the wild-type (N) (with the appended duplicated sequence disordered) and an alternative state (N'), designed to be more unstable than N in the absence of the ligand and in which the appended sequence substitutes the wild-type one and occupies its place. The two duplicated sequences differ in that an inactivating mutation is inserted in the critical position for ligand binding, but only in the original protein, so only N' is active. In the presence of ligand, only N' will be able to bind it, inducing the refolding of the appended sequence and the unfolding of the original stretch. These massive loop movements can be easily followed by fluorescence spectroscopy by putting probes in selected positions of both duplicated loops. The authors postulate the system as amenable for many proteins, since different permutation sites can be chosen, especially with computational methods that nowadays can predict viable permutation sites. With this procedure the protein does not undergo a complete unfolding transition (which may lead to technical problems, such as aggregation), but only a limited and localized region. This procedure has been followed to convert calbindin D9k into an optical calcium sensor^[90-91] and, with some modifications, for turning a fibronectin type III domain (FN3), a monobody binding scaffold, to an optical biosensor to specifically recognize the Src homology 2 (SH2) domain of cAbl kinase.^[92]

7. Summary and Outlook

The existence of many types of polypeptidic entities defying the "one sequence, one structure, one function" canon is an example of how complex is the relationship between an amino acid sequence and the folding process leading to a three-dimensional structure. In this work, three closely related concepts have been discussed: chameleon sequences, metamorphic proteins, and switch peptides, which we term as turncoat polypeptides. Nature has taken advantage of the plasticity of these turncoat polypeptides for regulatory purposes and for gaining functionality. This versatility may be also very useful for biotechnological and clinical purposes, and many research lines are focused on understanding and controlling these polypeptides to develop powerful tools.

Acknowledgements

This work was supported by Spanish Ministry of Science, Innovation and Universities (MCIU Grants BIO2016-79323-R and CTQ2017-84271P), Autonomous Community of Madrid (CM Grant S2017/BMD-3817) and Spanish National Research Council (CSIC Grant 2018201132).

Keywords: chameleon sequences • metamorphic proteins • protein folding • protein structure • switch peptides

- [1] a) K. A. Dill, *Protein Sci.* **1999**, *8*, 1166-1180; b) K. A. Dill, S. B. Ozkan, M. S. Shell, T. R. Weikl, *Annu. Rev. Biophys.* **2008**, *37*, 289-316.
- [2] a) E. S. Kuloglu, D. R. McCaslin, M. Kitabwalla, C. D. Pauza, J. L. Markley, B. F. Volkman, *Biochemistry* **2001**, *40*, 12486-12496; b) J. C. Fox, R. C. Tyler, C. Guzzo, R. L. Tuinstra, F. C. Peterson, P. Lusso, B. F. Volkman, *ACS Chem Biol* **2015**, *10*, 2580-2588.
- [3] a) G. A. Belogurov, M. N. Vassilyeva, V. Svetlov, S. Klyuyev, N. V. Grishin, D. G. Vassilyev, I. Artsimovitch, *Mol Cell* **2007**, *26*, 117-129; b) B. M. Burmann, S. H. Knauer, A. Sevostyanova, K. Schweimer, R. A. Mooney, R. Landick, I. Artsimovitch, P. Rösch, *Cell* **2012**, *150*, 291-303; c) D. Shi, D. Svetlov, R. Abagyan, I. Artsimovitch, *Nucleic Acids Res* **2017**, *45*, 8835-8843.
- [4] E. K. Jaffe, *Trends Biochem Sci* **2005**, *30*, 490-497.
- [5] A. Parmeggiani, I. M. Krab, S. Okamura, R. C. Nielsen, J. Nyborg, P. Nissen, *Biochemistry* **2006**, *45*, 6846-6857.
- [6] W. Li, L. N. Kinch, P. A. Karplus, N. V. Grishin, *Protein Sci.* **2015**, *24*, 1075-1086.
- [7] J. C. Whitney, D. Quentin, S. Sawai, M. LeRoux, B. N. Harding, H. E. Ledvina, B. Q. Tran, H. Robinson, Y. A. Goo, D. R. Goodlett, S. Raunser, J. D. Mougous, *Cell* **2015**, *163*, 607-619.
- [8] a) H. Zamora-Carreras, B. Maestro, E. Strandberg, A. S. Ulrich, J. M. Sanz, M. A. Jimenez, *Chemistry* **2015**, *21*, 8076-8089; b) H. Zamora-Carreras, B. Maestro, E. Strandberg, A. S. Ulrich, J. M. Sanz, M. A. Jimenez, *Chemistry* **2018**, *24*, 5825-5839.
- [9] M. A. Markus, A. P. Hinck, S. Huang, D. E. Draper, D. A. Torchia, *Nat Struct Biol* **1997**, *4*, 70-77.
- [10] S. Vivekanandan, J. R. Brender, S. Y. Lee, A. Ramamoorthy, *Biochem Biophys Res Commun* **2011**, *411*, 312-316.
- [11] L. Gremer, D. Scholzel, C. Schenk, E. Reinartz, J. Labahn, R. B. G. Ravelli, M. Tusche, C. Lopez-Iglesias, W. Hoyer, H. Heise, D. Willbold, G. F. Schroder, *Science* **2017**, *358*, 116-119.
- [12] A. F. Dishman, B. F. Volkman, *ACS Chem Biol* **2018**, *13*, 1438-1446.
- [13] H. Wasserman, E. O. Saphire, *Ann Rev Virol* **2016**, *3*, 373-386.
- [14] S. H. Knauer, P. Rösch, I. Artsimovitch, *RNA biology* **2012**, *9*, 1418-1423.
- [15] a) V. N. Uversky, *Cell Mol Life Sci* **2017**, *74*, 3065-3067; b) V. N. Uversky, C. J. Oldfield, A. K. Dunker, *J Mol Recognit* **2005**, *18*, 343-384.
- [16] A. G. Murzin, *Science* **2008**, *320*, 1725-1726.
- [17] S. C. Goodchild, C. N. Angstmann, S. N. Breit, P. M. G. Curmi, L. J. Brown, *Biochemistry* **2011**, *50*, 10887-10897.
- [18] M. Lella, R. Mahalakshmi, *Biochemistry* **2017**, *56*, 2971-2984.
- [19] T. Selwood, E. K. Jaffe, *Arch Biochem Biophys* **2012**, *519*, 131-143.
- [20] D. L. Minor, P. S. Kim, *Nature* **1996**, *380*, 730-734.
- [21] X. I. Ambroggio, B. Kuhlman, *Curr Opin Struct Biol* **2006**, *16*, 525-530.
- [22] K. S. Keating, S. C. Flores, M. B. Gerstein, L. A. Kuhn, *Protein Sci.* **2009**, *18*, 359-371.
- [23] L. Saucède, S. D. Santos, A. Chandravarkar, B. Mandal, R. Mimna, K. Murat, M.-S. Camus, J. Bérard, E. Grouzmann, M. Adrian, *CHIMIA Int J Chem* **2006**, *60*, 199-202.
- [24] J. Habchi, P. Tompa, S. Longhi, V. N. Uversky, *Chem Rev* **2014**, *114*, 6561-6588.
- [25] a) A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker, V. N. Uversky, *J Mol Biol* **2006**, *362*, 1043-1059; b) W. L. Hsu, C. J. Oldfield, B. Xue, J. Meng, F. Huang, P. Romero, V. N. Uversky, A. K. Dunker, *Protein Sci* **2013**, *22*, 258-273.
- [26] C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Yang, V. N. Uversky, A. K. Dunker, *BMC Genomics* **2008**, *9 Suppl 1*, S1.
- [27] A. Relini, N. Marano, A. Gliozzi, *Biomolecules* **2013**, *4*, 20-55.
- [28] D. Otzen, R. Riek, *Cold Spring Harb Perspect Biol* **2019**.
- [29] A. Bleem, G. Christiansen, D. J. Madsen, H. Maric, K. Stromgaard, J. D. Bryers, V. Daggett, R. L. Meyer, D. E. Otzen, *J Mol Biol* **2018**, *430*, 3751-3763.
- [30] W. Kabsch, C. Sander, *Proc Natl Acad Sci U S A* **1984**, *81*, 1075-1078.
- [31] A. Ghodlane, A. Joseph, A. Bornot, A. De Brevem, *Bioinformatics* **2009**, *3*, 367-369.
- [32] K. Takano, Y. Katagiri, A. Mukaiyama, H. Chon, H. Matsumura, Y. Koga, S. Kanaya, *Proteins* **2007**, *68*, 617-625.
- [33] J.-T. Guo, J. W. Jaromczyk, Y. Xu, *Proteins* **2007**, *67*, 548-558.
- [34] M. Mezei, *Prot eng* **1998**, *11*, 411-414.
- [35] M. Mezei, *Algorithms* **2018**, *11*, 114.
- [36] W. J. Anderson, L. O. Van Dorn, W. M. Ingram, M. H. J. Cordes, *Protein Eng Des Sel.* **2011**, *24*, 765-771.
- [37] S. Yoon, H. Jung, *The Protein Journal* **2006**, *25*, 361-368.
- [38] G. Bahramali, B. Goliaei, Z. Minuchehr, A. Salari, *Biochem Biophys Res Commun* **2016**, *472*, 209-216.
- [39] J. M. Sanz, M. A. Jiménez, G. Giménez-Gallego, *Biochemistry* **2002**, *41*, 1923-1933.
- [40] L. L. Porter, L. L. Looger, *Proc Natl Acad Sci U S A* **2018**, *115*, 5968-5973.
- [41] Y. Yang, J. Gao, J. Wang, R. Heffernan, J. Hanson, K. Paliwal, Y. Zhou, *Brief. Bioinform.* **2016**, bbw129.
- [42] C. J. Oldfield, K. Chen, L. Kurgan, *Methods Mol Biol* **2019**, 1958, 73-100.
- [43] I. Jacoboni, P. L. Martelli, P. Fariselli, V. De Pinto, R. Casadio, *Protein Sci.* **2001**, *10*, 779-787.
- [44] B. Rost, *J Struct Biol* **2001**, *134*, 204-218.
- [45] K. M. Saravanan, S. Selvaraj, *Peptide Sci* **2013**, *100*, 148-153.
- [46] Y. Yang, R. Heffernan, K. Paliwal, J. Lyons, A. Dehzangi, A. Sharma, J. Wang, A. Sattar, Y. Zhou, in *Prediction of protein secondary structure*, Springer, **2017**, pp. 55-63.
- [47] X. Zhou, F. Alber, G. Folkers, G. H. Gonnet, G. Chelvanayagam, *Proteins* **2000**, *41*, 248-256.
- [48] R. L. Tuinstra, F. C. Peterson, S. Kutlesa, E. S. Elgin, M. A. Kron, B. F. Volkman, *Proc Natl Acad Sci U S A* **2008**, *105*, 5057-5062.
- [49] X. Luo, H. Yu, *Structure* **2008**, *16*, 1616-1625.
- [50] T. Newlove, J. H. Konieczka, M. H. J. Cordes, *Structure* **2004**, *12*, 569-581.
- [51] P. Kulkarni, T. L. Solomon, Y. He, Y. Chen, P. N. Bryan, J. Orban, *Protein Sci.* **2018**, *27*, 1557-1567.
- [52] J. L. Markley, J. H. Kim, Z. Dai, J. R. Bothe, K. Cai, R. O. Frederick, M. Tonelli, *FEBS letters* **2013**, *587*, 1172-1179.
- [53] D. R. Littler, S. J. Harrop, W. D. Fairlie, L. J. Brown, G. J. Pankhurst, S. Pankhurst, M. Z. DeMaere, T. J. Campbell, A. R. Bauskin, R. Tonini, *J Biol Chem.* **2004**, *279*, 9298-9305.
- [54] Y. Zhao, L. Li, C. Wu, X. Jiang, B. Ge, H. Ren, F. Huang, *Protein Eng Des Sel.* **2016**, *29*, 23-29.
- [55] R. London, *Viruses* **2016**, *8*, 260.
- [56] E. M. Adler, *J Gen Physiol* **2015**, *146*, 193-194.
- [57] a) R. S. Signarvic, W. F. DeGrado, *J Am Chem Soc.* **2009**, *131*, 3377-3384; b) S. J. Smith, K. Du, R. J. Radford, F. A. Tezcan, *Chem Sci* **2013**, *4*, 3740-3747.
- [58] M. Mutter, A. Chandravarkar, C. Boyat, J. Lopez, S. Dos Santos, B. Mandal, R. Mimna, K. Murat, L. Patiny, L. Saucedo, *Angew Chem* **2004**, *43*, 4172-4178.
- [59] J. P. Schneider, D. J. Pochan, B. Ozbaz, K. Rajagopal, L. Pakstis, J. Kretsinger, *J Am Chem Soc.* **2002**, *124*, 15030-15037.
- [60] X. Wang, I. Bergenfeld, P. S. Arora, J. W. Canary, *Angew Chem* **2012**, *124*, 12265-12267.
- [61] P. Anzini, C. Xu, S. Hughes, E. Magnotti, T. Jiang, L. Hemmingsen, B. Demeler, V. P. Conticello, *J Am Chem Soc.* **2013**, *135*, 10278-10281.
- [62] M. J. Pandya, E. Cerasoli, A. Joseph, R. G. Stoneman, E. Waite, D. N. Woolfson, *J Am Chem Soc.* **2004**, *126*, 17016-17024.
- [63] P. N. Bryan, J. Orban, *Curr Opin Struct Biol.* **2010**, *20*, 482-488.
- [64] a) B. Ciani, E. G. Hutchinson, R. B. Sessions, D. N. Woolfson, *J Biol Chem.* **2002**, *277*, 10150-10155; b) R. A. Kammerer, M. O. Steinmetz, *J Struct Biol* **2006**, *155*, 146-153.
- [65] a) Y. Fezoui, D. M. Hartley, D. M. Walsh, D. J. Selkoe, J. J. Osterhout, D. B. Teplow, *Nat Struct Mol Biol* **2000**, *7*, 1095-1099; b) R. A. Kammerer, D. Kostrewa, J. Zurdo, A. Detken, C. Garcia-Echeverria, J. D. Green, S. A. Müller, B. H. Meier, F. K. Winkler, C. M. Dobson, *Proc Natl Acad Sci U S A* **2004**, *101*, 4435-4440.
- [66] W. Li, F. Nicol, F. C. Szoka Jr, *Adv Drug Delivery Rev* **2004**, *56*, 967-985.
- [67] a) R. Cerpa, F. E. Cohen, I. D. Kuntz, *Folding and Design* **1996**, *1*, 91-101; b) J. Bredenbeck, J. Helbing, J. R. Kumita, G. A. Woolley, P. Hamm, *Proc Natl Acad Sci U S A* **2005**, *102*, 2379-2384; c) Oana I. Lungu, Ryan A. Hallett, Eun J. Choi, Mary J. Aiken,

- Klaus M. Hahn, B. Kuhlman, *Chemistry & Biology* **2012**, *19*, 507-517.
- [68] M. B. Hansen, L. Ruizendaal, D. W. P. M. Löwik, J. C. M. van Hest, *Drug Discovery Today: Technologies* **2009**, *6*, e33-e39.
- [69] G. P. Dado, S. H. Gellman, *J Am Chem Soc.* **1993**, *115*, 12609-12610.
- [70] a) M. Albrecht, P. Stortz, *Chem Soc Rev.* **2005**, *34*, 496-506; b) M. R. Ghadiri, A. K. Fernholz, *J Am Chem Soc.* **1990**, *112*, 9633-9635.
- [71] a) S. R. Gilbertson, G. Chen, M. McLoughlin, *J Am Chem Soc.* **1994**, *116*, 4481-4482; b) P. Rossi, P. Tecilla, L. Baltzer, P. Scrimin, *Chemistry* **2004**, *10*, 4163-4170.
- [72] M. Mutter, R. Gassmann, U. Buttke, K. H. Altmann, *Angew Chem* **1991**, *30*, 1514-1516.
- [73] G. D. Rose, T. P. Creamer, *Proteins* **1994**, *19*, 1-3.
- [74] S. Dalal, S. Balasubramanian, L. Regan, *Nat Struct Biol* **1997**, *4*, 548.
- [75] P. A. Alexander, Y. He, Y. Chen, J. Orban, P. N. Bryan, *Proc Natl Acad Sci U S A* **2009**, *106*, 21149-21154.
- [76] a) W. Z. Yang, T. P. Ko, H. S. Yuan, L. Corselli, R. C. Johnson, *Protein Sci.* **1998**, *7*, 1875-1883; b) H. Tidow, T. Lauber, K. Vitzthum, C. P. Sommerhoff, P. Rösch, U. C. Marx, *Biochemistry* **2004**, *43*, 11238-11247.
- [77] Y. Hori, Y. Sugiura, *J Am Chem Soc.* **2002**, *124*, 9362-9363.
- [78] E. Cerasoli, B. K. Sharpe, D. N. Woolfson, *J Am Chem Soc.* **2005**, *127*, 15008-15009.
- [79] K. Pagel, T. Vagt, B. Kokschi, *Org Biomol Chem* **2005**, *3*, 3843-3850.
- [80] K. Pagel, B. Kokschi, *Curr Opin Chem Biol* **2008**, *12*, 730-739.
- [81] H. L. Schenck, G. P. Dado, S. H. Gellman, *J Am Chem Soc.* **1996**, *118*, 12487-12494.
- [82] M. M. Stratton, S. N. Loh, *Protein Sci.* **2011**, *20*, 19-29.
- [83] a) Z. A. Bornholdt, T. Noda, D. M. Abelson, P. Halfmann, M. R. Wood, Y. Kawakita, E. O. Saphire, *Cell* **2013**, *154*, 763-774; b) S. H. Lawrence, U. D. Ramirez, L. Tang, F. Fazliye, L. Kundrat, G. D. Markham, E. K. Jaffe, *Chemistry & biology* **2008**, *15*, 586-596.
- [84] a) K. W. Plaxco, H. T. Soh, *Trends Biotech* **2011**, *29*, 1-5; b) A. Vallée-Bélisle, K. W. Plaxco, *Curr Opin Struct Biol.* **2010**, *20*, 518-526.
- [85] K. Kluska, J. Adamczyk, A. Krężel, *Metallomics* **2018**, *10*, 248-263.
- [86] a) G. K. Walkup, B. Imperiali, *J Am Chem Soc.* **1997**, *119*, 3443-3450; b) G. K. Walkup, B. Imperiali, *J Am Chem Soc.* **1996**, *118*, 3053-3054.
- [87] H. A. Godwin, J. M. Berg, *J Am Chem Soc.* **1996**, *118*, 6514-6515.
- [88] J. E. Kohn, K. W. Plaxco, *Proc Natl Acad Sci U S A* **2005**, *102*, 10841-10845.
- [89] K. A. Cissell, S. Shrestha, J. Purdie, D. Kroodsma, S. K. Deo, *Anal Bioanal Chem* **2008**, *391*, 1721-1729.
- [90] M. M. Stratton, S. N. Loh, *Proteins* **2010**, *78*, 3260-3269.
- [91] a) M. M. Stratton, S. McClendon, D. Eliezer, S. N. Loh, *Biochemistry* **2011**, *50*, 5583-5589; b) M. M. Stratton, D. M. Mitrea, S. N. Loh, *ACS Chem Biol* **2008**, *3*, 723-732; c) A. J. DeGrave, J.-H. Ha, S. N. Loh, L. T. Chong, *Nat Commun* **2018**, *9*, 1013.
- [92] H. Zheng, J. Bi, M. Krendel, S. N. Loh, *Biochemistry* **2014**, *53*, 5505-5514.

Entry for the Table of Contents

REVIEW

Turncoat polypeptides are sequences that can switch between different structural states upon variations in the surrounding conditions or under the effect of trigger agents

Héctor Zamora-Carreras[#] Beatriz Maestro,[#] Jesús M. Sanz,^{*} and M. Angeles Jiménez^{*}

Page No. – Page No.

Turncoat polypeptides: We adapt to our environment