Exploitation of proteomics strategies in protein structure–function studies*

Gennaro Marino[1,2,‡], Piero Pucci[1,2], Leila Birolo[1], and Margherita Ruoppolo[2,3]

[1]Dipartimento di Chimica Organica e Biochimica, Facoltà di Scienze Biotecnologiche, Università di Napoli "Federico II", Napoli, Italy; [2]CEINGE, Biotecnologie Avanzate, scrl, Napoli, Italy; [3]Dipartimento di Biochimica e Biotecnologie Mediche, Università di Napoli "Federico II", Napoli, Italy

Abstract: Mass spectrometry plays a central role in structural proteomics, particularly in highly intensive structural genomics projects. This review paper reports some examples taken from recent work from the authors' laboratory and is aimed at showing that modern proteomics strategies are instrumental in the integration of structural genomic projects in fields such as: (i) protein–protein interactions, (ii) protein–DNA interactions, (iii) protein–ligand interactions, and (iv) protein-folding intermediates.

INTRODUCTION

Mass spectrometry (MS) is nowadays the core methodology in structural proteomics [1]. It plays a noteworthy role in highly intensive projects aimed at giving a "form" to the monodimensional inputs of genome sequencing (structural genomics). A number of strategies that combine protein chemistry procedures such as limited proteolysis, selective chemical modifications, and/or hydrogen/deuterium (H/D) exchange with the unsurpassed sensitivity and dynamic range of analysis of MS, may actually provide direct answers to different structural problems such as: (i) probing protein–protein interactions, (ii) probing protein–DNA interactions, (iii) probing protein–ligand interactions, and (iv) probing protein-folding intermediates.

Some examples aimed at showing that modern proteomics strategies are instrumental in the integration of structural genomic projects are outlined and discussed in this review paper.

PROBING PROTEIN–PROTEIN INTERACTIONS

A critical step toward the understanding of protein–protein interactions consists in the structural description of the contact regions within the complex. Focal points of the protein complex "form" may be zoomed-out by the integration of protein chemistry and MS advanced tools. The mono-dimensional data then can eventually be locked on a 3D template by the judicious use of molecular modeling and dynamics. Structural proteomics actually combine the use of traditional protein chemistry tools (i.e., limited proteolysis and chemical reagents) duly revisited, with a modern MS armory. The revisitation essentially consists in spotting only single chemical events occurring on the protein substrate. In order to finely tune the specified reaction to this end, an array of experiments can be monitored by electro-

spray (ES)-MS [2]. In contrast to the previous classical approach of protein chemistry, one can now make use of both specific and aspecific reagents, and proteases can be used without taking into consideration the total extent of the reaction [3].

As a result of "complementary" proteolysis experiments, the exact molecular mass of the two complementary (poly)peptides, as determined by ES-MS often in trace amounts, immediately defines the most exposed/flexible peptide bond in the selected conditions. By using the extensive armory of available proteases, it is then possible to get a sufficiently detailed picture of the exposed protein backbone. Differential protein conformers can be highlighted when preferential proteolytic patterns of unbound protein structures and their complex(es) with molecular effector(s), either of low and high molecular weight, are compared using identical proteolytic conditions. As a corollary, cross-linking reagents, in conditions to give a single molecular derivative, may be of paramount importance in providing structural constraint information, including orientation of secondary structure elements [4].

The topology of the Ca^{2+}-calmodulin (CaM)-melittin ternary complex has been investigated by the integrated strategy described above. Calcium-bound calmodulin is rapidly digested by proteases with the most exposed regions being located within the second half of the long central helix connecting the two domains in the native molecule. However, upon binding melittin, most of the cleavage sites of calmodulin disappeared. Moreover, the central region of melittin was fully protected in the complex, whereas the isolated peptide had been completely digested under the same conditions. The analysis of the CaM-melittin complex using these integrated approaches showed that during the interaction, calmodulin undergoes a dramatic conformational change in which the N- and C-terminal domains are brought in close proximity by the disruption of the central helix, giving rise to a globular complex that engulfs most of the melittin structure. Surprisingly, the ligand peptide was shown to interact with calmodulin by adopting an opposite orientation within the complex (Fig. 1) as compared to all the peptide substrate examined so far [4]. This finding adds a further dimension to the already remarkable capability of calmodulin in binding different protein substrates, thus suggesting that this protein might be involved in regulating an even larger number of enzymatic activities.
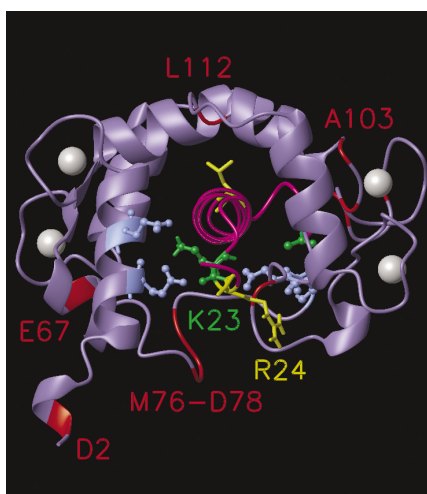

Fig. 1 3D structure of the CAM-mellittin complex showing the reverse orientation of the peptide (in magenta).

PROBING PROTEIN–DNA INTERACTIONS

The approach described above can also be applied to the investigation of protein–DNA binding. In protein–DNA complexes, in fact, the (poly)peptide regions responsible for DNA recognition, usually accessible in the isolated molecules, become protected following the interaction.

Thyroid transcription factor 1 (TTF-1) is a 378-residue protein responsible for transcriptional activation of genes expressed in follicular thyroid cells and lung epithelial cells. TTF-1 binds DNA by a homeodomain (HD) of 61 amino acid residues, which specifically recognizes oligonucleotide sequences containing the 5′-CAAG-3′ core motif. The 3D structure of recombinant TTF-1HD has been solved by NMR analysis, whereas no structural data are available for its complex with DNA. The topology of the TTF-1HD/DNA complex was studied by integrating traditional protein chemistry tools with MS methodologies [5]. When limited proteolysis digestions were carried out on the protein in the absence or presence of its target oligonucleotide, differential peptide maps were obtained from which the amino acid residues involved in the interaction were inferred. Similarly, selective acetylation of lysine residues in both the isolated and the complexed HD led to the identification of those amino acids masked following DNA interaction.

Surface topology analysis of isolated TTF-1HD performed at neutral pH was in good agreement with the 3D structure of the molecule as determined by NMR studies in acidic conditions. Minor differences were detected in the C-terminal region of the protein circled in Fig. 2, which, contrary to NMR data, showed no accessibility to proteases. Analysis of the TTF-1HD/DNA complex provided an experimental validation of the model proposed for the interaction on the basis of the HD structures described so far. An increased accessibility of the C-terminal region was observed following the interaction, indicating that the C-terminal portion of HD is displaced from the protein surface by the oligonucleotide molecule and adopts a disordered and flexible conformation. These findings suggest a direct effect of DNA binding on the molecular conformation of the entire TTF-1 protein; the displacement of the HD C-terminal tail might affect the C-terminal domain of the TTF-1 protein, causing this region to assume a biologically relevant conformation. Comparative experiments with DNA fragments differing for sequence and binding capability highlighted structural differences among the complexes, mainly located in the N-terminal region of the HD, thus accounting for their different dissociation constants.
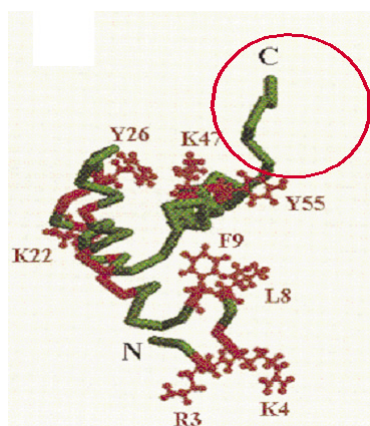
Fig. 2 Structure of TTF-1HD. The circle indicates the C-terminal region of the protein.

PROBING PROTEIN–LIGAND INTERACTIONS

Two examples concerning either covalently and noncovalently bound effectors are described below, showing how the proteomics strategies can integrate molecular modeling in the study of interactions of ligands to proteins, leading eventually to the inactivation (phospholipase) or to the gain of function (lactoalbumin).

Petrosaspongiolide M (PM) is an anti-inflammatory marine metabolite that displays a potent inhibitory activity toward groups II and III secretory phospholipase A2 (PLA2) enzymes. The details of the mechanism, which leads to a covalent adduct between PLA2 and hydroxybutenolide-containing molecules such as PM, are still a matter of debate. The authors investigated the covalent binding of PM to bee venom PLA2 by MS and molecular modeling [6]. The mass increment observed for the PM-PLA2 adduct is consistent with the formation of a Schiff base by reaction of a PLA2 amino group with the hemiacetal function at the C25 atom of the PM-hydroxybutenolide ring. Extensive proteolysis of modified PLA2 by the endoprotease LysC followed by high-performance liquid chromatography MS analysis allowed the authors to establish that the PLA2-amino terminal group residue was the single covalent binding site for PM. The stoichiometry of the reaction between PM and PLA2 was also monitored, and results showed that even in the presence of excess inhibitor, the prevalent product is a 1:1 (inhibitor:enzyme) adduct, although a 2:1 adduct is present as a minor component. The 2:1 adduct was also characterized, which showed that the second reaction site is located at the amino group of the Lys-85 residue. Similar results in terms of the reaction profile, mass increments, and location of the PLA2 binding site were obtained for manoalide, a paradigm for irreversible PLA2 inhibitors, which suggests that the present results may be considered of general interest within the field of anti-inflammatory sesterterpenes that contain the hydroxybutenolide pharmacophore. Finally, a 3D model, constrained by the above experimental results, was obtained by docking the inhibitor molecule into the PLA2 binding site through AFFINITY calculations. The model provides an interesting insight into the PM-PLA2 inhibition process and may prove useful in the design of new anti-inflammatory agents that target PLA2 secretory enzymes.

A combination of H/D exchange and limited proteolysis experiments coupled to MS analysis was used to depict the conformation of a folding variant of α-lactalbumin, endowed with the peculiar biological function of inducing apoptosis in tumor and immature cells. This conformer named HAMLET (human α-lactalbumin made lethal to tumor cells) and stabilized by the interaction with a specific fatty acid, C18:1 [7]. Human α-lactalbumin (holo) is a 14 kDa protein whose structure is made of two domains containing a single tightly bound Ca^{2+} and four disulfide bonds, and at least two different conformations of α-lactalbumin have already been described under different physicochemical conditions, i.e., at low pH (A state) and in the absence of Ca^{2+} (apo). Nevertheless, structural details of HAMLET accounting for its molecular diversity from other forms of α-lactalbumin are lacking, so far. Near and far UV CD, fluorescence, and 8-anilino-1-naphthalenesulfonate (ANS) binding were not able to discriminate between HAMLET and apo-α-lactalbumin, because they exhibit very similar spectroscopic characteristics [8,9].

H/D exchange experiments clearly indicated that HAMLET and apo are indeed two different conformers of α-LA, both differing from holo-α-lactalbumin. By extending the above concepts, D_2O can be viewed as an aspecific chemical reagent in establishing, either globally or locally, the exposition of peptide backbones. HAMLET incorporated a greater number of deuterium atoms compared to the apo and holo forms. Moreover, the apo α-lactalbumin appeared to be a mixture of at least two populations of conformers with different exchange properties.

"Complementary" proteolysis experiments showed that HAMLET and apo are both accessible to proteases in the β-domain with cleavages occurring at the same sites as illustrated on the three-dimensional structure of holo-α-lactalbumin in Fig. 3. Nevertheless, substantial differences in the kinetics of enzymatic digestion at specific sites were observed. In particular, Asp37, Tyr50, and Phe53 were found to be by far more exposed in HAMLET than in apo. These results were supported by the analysis of the
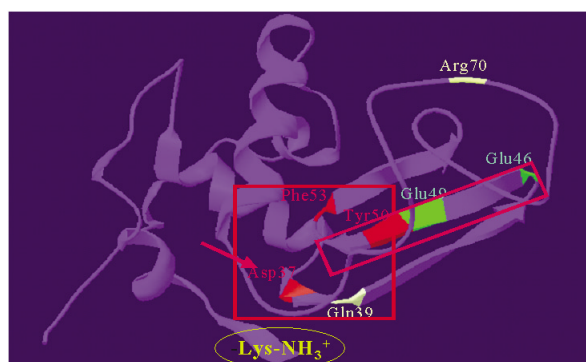
Fig. 3 Preferential cleavage sites observed in HAMLET and apo α-LA mapped on the 3D structure of holo α-LA.

peptides generated from the peptic digest of deuterium-exchanged α-lactalbumin conformers, and suggested that binding of oleic acid might displace the central strand of the β-sheet, exposing Asp37, Tyr50, and Phe53 (Fig. 3).
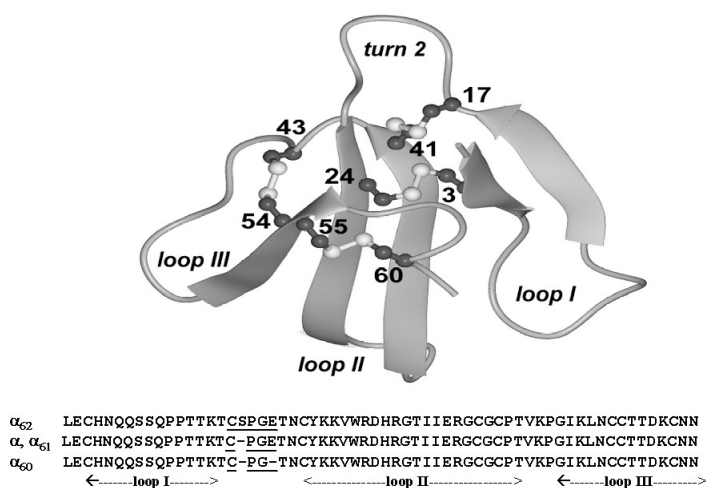
Low-resolution approaches such us limited proteolysis and H/D exchange experiments in conjunction with MS analysis were shown, therefore, to be instrumental in providing subtle structural details on highly flexible conformations such as HAMLET or apo. Modern views in structural proteomics have already recognized that the flexibility of proteins is instrumental not only to exploit catalytic activity but also to alter their biological functions, conferring peculiar properties to different conformations of the molecule. Thereof, proteomic studies which require information of the structural changes associated with different biological functions of the same (poly)peptide chain, will more and more rely on the kind of approaches described in this section.

PROBING PROTEIN-FOLDING INTERMEDIATES

In protein-folding studies, MS can be used to characterize the population of species formed on the folding pathway, and the disulfide bonds present at different times of the process. Aliquots of the reaction mixture withdrawn at different times during the oxidative folding process are trapped by alkylation of the free thiol and analyzed by ES-MS. The alkylation reaction prevents reoxidation and increases the molecular mass of the intermediates by a fixed amount for each reacted free SH group. Intermediates containing different numbers of disulfide bonds are separated by mass, and their relative abundance in the sample is determined. The nature + quantitative distribution of the disulfide-bonded species present at a given time can be used to establish the refolding pathway of the protein and, in some cases, to develop a kinetic analysis of the process.

Once the refolding intermediates have been characterized by ES-MS in terms of their content of disulfide bonds, a further step would consist in the structural assignment of the various disulfide bonds formed at different times in the entire process by peptide mapping using matrix-assisted laser desorption ionization MS (MALDI-MS). The refolding intermediates are cleaved at points between the potentially bonded cysteine residues under conditions known to minimize disulfide reduction and reshuffling in attempts to isolate any cysteine residue within an individual peptide. The MALDI-MS mapping approach is then used to search for any S–S bonded peptides, which are characterized by their unique masses. The interpretation is then confirmed by performing reduction or Edman degradation steps followed by rerunning the MS spectrum [10].

Snake neurotoxins are short all-β proteins that display a complex organization of the disulfide bonds: two bonds connect consecutive cysteine residues (C43–C54, C55–C60), and two bonds intersect when bridging (C3–C24, C17–C41) to form a particular structure classified as "disulfide ß-cross" [11]

<pre>
α₆₂   LECHNQQSSQPPTTKTCSPGETNCYKKVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDKCNN
α, α₆₁ LECHNQQSSQPPTTKTC-PGETNCYKKVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDKCNN
α₆₀   LECHNQQSSQPPTTKTC-PG-TNCYKKVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDKCNN
        ←-------loop I-------->   <---------------loop II----------------->   ←-------loop III---------->
</pre>

Fig. 4 Ribbon structure of snake neurotoxins. Sequence alignment of variants used in this study. The sequence of turn 2 is underlined.

shown in Fig. 4. The general organization of the (poly)peptide chain in snake toxins generates a trefoil structure termed "three-fingers fold". The authors showed [12] that a single mutation located in turn 2 substantially alters the rate of the sequential folding process of the protein. The authors used three variants for which the length of this turn was increased from three (variant α60) to four (toxin α) and then to five (variant α62) residues. The increase in the length of the turn slowed down the folding process so that after 2 h of incubation, about 90 % of the variant α60 was folded, whereas this value decreased to about 70 % for the 61 residues toxin α and 20 % for the variant α62. Interestingly, three disulfide-containing intermediates were the predominant species for all variants folding, but this population was markedly more abundant and persistent for the slowest-folding neurotoxin.

Two intermediates containing three disulfides were found to accumulate during early and late stages of neurotoxin α62 folding [13]. Both intermediates consisted of chemically homogenous species containing three of the four native disulfide bonds and lacking the C43–C54 and the C17–C41 coupling, respectively. A model of the folding pathway of neurotoxin α62 was proposed by the authors and is illustrated in Fig. 5. The des-[43–54] intermediate appeared to be the immediate precursor of the native species. Conversely, the des-[17–41] species was unable to form the fourth disulfide bond and had to rearrange into intermediates that can directly reach the native state. These isomerization reactions pro-
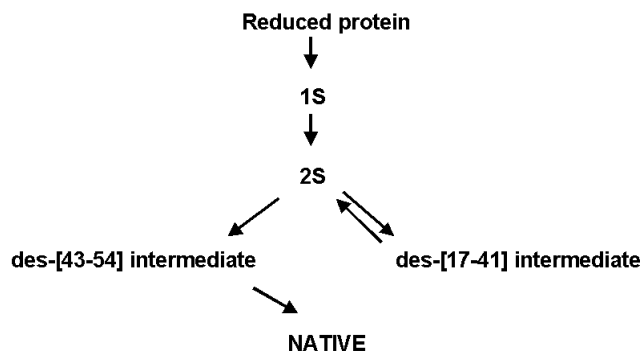
Fig. 5 Model of the folding pathway of neurotoxin α62.

vide an explanation for the accumulation of three disulfide intermediates along the pathway, which caused the slow oxidative folding, observed for neurotoxin α62.

   Integration of spectroscopic investigations and MS procedures might be instrumental in providing subtle structural details on transient conformations such as those present along a folding pathway even in the absence of disulfide bonds, as demonstrated by the characterization of a partly folded intermediate of aspartate aminotransferase from *Escherichia coli* [14]. In conjunction with spectroscopic techniques, H/D exchange and limited proteolysis experiments coupled to MS analysis were used to depict the topology of the monomeric M* partly folded intermediate in wild-type (WT) as well as in a mutant form of aspartate aminotransferase from *E. coli* (*Ec*AspAT) in which the highly conserved *cis*-proline at position 138 was replaced by a *trans*-alanine (P138A) [15]. Fluorescence analysis indicated that while M* is an off-pathway intermediate in the folding of WT *Ec*AspAT, it seems to coincide with an on-pathway folding intermediate for the P138A mutant. Gel filtration and CD analysis did not show major differences between the two M* species. On the contrary, the increased deuterium incorporation level and the higher accessibility to proteases shown by WT-M* as compared to P138A-M*, demonstrated the occurrence of conformational differences between the M* intermediates. These results, together with fluorescence data, indicate that P138A-M* is conceivably more compact than WT-M*. Limited proteolysis data, summarized in Fig. 6, suggested that the conformational differences might be related to a different relative orientation of the small and large domains of the protein induced by the presence of the *cis*-proline residue at position 138 in the WT protein. The relative rotation of the two domains might also affect the conformation of the N-terminal tail, thus explaining the accessibility of the segment 20–25, located in the small domain close to the N-terminus of the protein. The differences
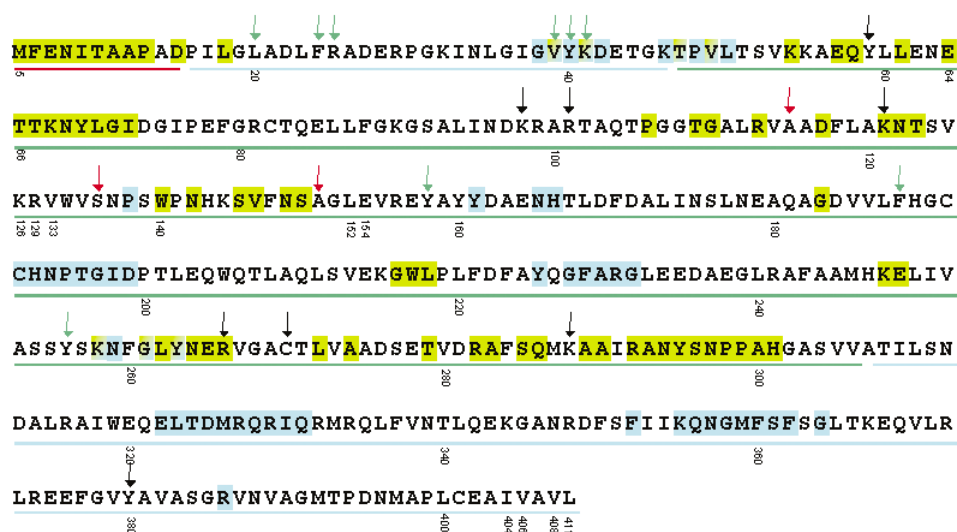
Fig. 6 Preferential proteolytic sites mapped onto the amino acid sequence of *Ec*AspAT. Cleavage sites only occurring in WT-M* are indicated by green arrows, those only present in P138A-M* by red arrows and those observed in both proteins by black arrows. Regions underlined in red, pale blue, and green indicate the N-terminal tail, the large domain and the small domain, respectively. Residues occurring at the domain interface are highlighted in pale blue, and those at the subunit interface are highlighted in yellow.

between the two M* species suggest that in WT-M* Pro138 is in the *cis* conformation at this stage of the folding process.

CONCLUSIONS

Large structural genomic projects face several different bottlenecks owing to the innate methodological requirements by X-ray diffraction and NMR studies (for a recent review, see [16]). Modern proteomic strategies, as outlined above, may be instrumental in high-throughput structure determination studies both at the top and the bottom of the conveyor belt of structural genomics in providing: (a) protein quality control; (b) folding optimization; (c) models validation; (d) structural constraints; and in probing: (i) functioning structures; (ii) ligand interactions; and (iii) metastable states.

REFERENCES

1. J. Godovac-Zimmermann and L. R. Brown. *Mass Spectrom. Rev.* **20** (1), 1 (2001).
2. F. Zappacosta, A. Pessi, E. Bianchi, S. Venturini, M. Sollazzo, A. Tramontano, G. Marino, P. Pucci. *Protein Sci.* **5**, 802 (1996).
3. F. Zappacosta, P. Ingallinella, A. Scaloni, A. Pessi, E. Bianchi, M. Sollazzo, A. Tramontano, G. Marino, P. Pucci. *Protein Sci.* **6**, 1901 (1997).
4. A. Scaloni, N. Miraglia, S. Orru, P. Amodeo, A. Motta, G. Marino, P. Pucci. *J. Mol. Biol.* **277**, 945 (1998).
5. A. Scaloni, M. Monti, R. Acquaviva, G. Tell, G. Damante, S. Formisano, P. Pucci. *Biochemistry* **38**, 64 (1999).
6. F. Dal Piaz, A. Casapullo, A. Randazzo, R. Riccio, P. Pucci, G. Marino, L. Gomez-Paloma. *ChemBioChem* **3**, 664 (2002).
7. A. Casbarra, L. Birolo, G. Infusini, F. Dal Piaz, M. Svensson, P. Pucci, C. Svanborg, G. Marino. *Proc. Natl. Aca. Sci. USA* (2003). Submitted for publication.
8. M. Svensson, H. Sabharwal, A. Håkansson, A. K. Mossberg, P. Lipniunas, H. Leffler, C. Svanborg, S. Linse. *J. Biol. Chem.* **274**, 6388–6396 (1999).
9. M. Svensson, A. Håkansson, A. K. Mossberg, S. Linse, C. Svanborg. *Proc. Natl. Acad. Sci. USA* **97**, 4221–4226 (2000).
10. M. Ruoppolo, C. Torella, F. Kanda, G. Marino, M. Panico, P. Pucci, H. R. Morris. *Folding Design* **1**, 381 (1996).
11. P. M. Harrison and M. J. E. Sternberg. *J. Mol. Biol.* **264**, 603 (1996).
12. M. Ruoppolo, M. Moutiez, M. F. Mazzeo, P. Pucci, A. Menez, G. Marino, E. Quéméneur. *Biochemistry* **37**, 16060 (1998).
13. M. Ruoppolo, F. Talamo, P. Pucci, M. Moutiez, E. Quèmèneur, A. Mènez, G. Marino. *Biochemistry* **40**, 15257 (2001).
14. L. Birolo, F. Dal Piaz, P. Pucci, G. Marino. *J. Biol. Chem.* **277**, 17428–37 (2002).
15. Y. A. Ovchinnikov, C. A. Egorov, N. A. Aldanova, M. Y. Feigina, V. M. Lipkin, N. G. Abdulaev, E. V. Grishin, A. P. Kiselev, N. N. Modyanov, A. E. Braunstein, O. L. Polyanovsky, V. V. Nosikov. *FEBS Lett.* **29**, 31–34 (1973).
16. M. B. Schmid. *Curr. Opin. Chem. Biol.* **2**, 529 (1998).