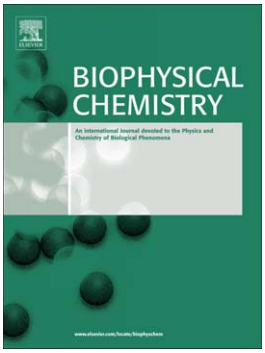
folding efficiencies for mutants of the P22 tailspike -helix protein correlate with predicted stability changes

Lothar Reich, Marion Becker, Robert Seckler, Thomas R. Weikl



In vivo folding efficiencies for mutants of the P22 tailspike β-helix protein correlate with predicted stability changes

Lothar Reich, Marion Becker, Robert Seckler, Thomas R. Weikl

PII: S0301-4622(09)00029-5
DOI: doi: 10.1016/j.bpc.2009.01.015
Reference: BIOCHE 5227

To appear in: *Biophysical Chemistry*

Received date: 29 November 2008
Revised date: 29 January 2009
Accepted date: 29 January 2009

Please cite this article as: Lothar Reich, Marion Becker, Robert Seckler, Thomas R. Weikl, *In vivo* folding efficiencies for mutants of the P22 tailspike β-helix protein correlate with predicted stability changes, *Biophysical Chemistry* (2009), doi: 10.1016/j.bpc.2009.01.015

In vivo folding efficiencies for mutants of the P22 tailspike β-helix protein correlate with predicted stability changes

Lothar Reich[1], Marion Becker[1,2], Robert Seckler[2], and Thomas R. Weikl[1]

[1] *Max Planck Institute of Colloids and Interfaces, Department of Theory and Bio-Systems, Science Park Golm, 14424 Potsdam, Germany*

[2] *Department of Physical Biochemistry, University of Potsdam, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam-Golm, Germany.*

Abstract

Parallel β-helices are among the simplest repetitive structural elements in proteins. The folding behavior of β-helix proteins has been studied intensively, also to gain insight on the formation of amyloid fibrils, which share the parallel β-helix as central structural motif. An important system for investigating β-helix folding is the tailspike protein from the *Salmonella* bacteriophage P22. The central domain of this protein is a right-handed parallel β-helix with 13 windings. Extensive mutational analyses of the P22 tailspike protein have revealed two main phenotypes: temperature-sensitive-folding (tsf) mutations that reduce the folding efficiency at elevated temperatures, and global suppressor (su) mutations that increase the tailspike folding efficiency. A central question is whether these phenotypes can be understood from changes in the protein stability induced by the mutations. Experimental determination of the protein stability is complicated by the nearly irreversible trimerization of the folded tailspike protein. Here, we present calculations of stability changes with the program FoldX, focusing on a recently published extensive data set of 145 singe-residue alanine mutants. We find that the calculated stability changes are correlated with the experimentally measured *in vivo* folding efficiencies. In addition, we determine the free-energy landscape of the P22 tailspike protein in a

nucleation-propagation model to explore the folding mechanism of this protein, and obtain a processive folding route on which the protein nucleates in the N-terminal region of the helix.

Key words: repeat proteins; β-helix; protein stability; mutational analysis; energy landscape

Introduction

A considerable fraction of proteins have repetitive elements in their structure. These structural repeats often arise from internal sequence repeats [1, 2], but can also be elements without any sequence similarity [3]. The repeats usually lead to an overall elongated shape of the protein, for example to the superhelical shape of solenoid proteins [4]. In the simplest cases, each structural repeat consists of just a few secondary elements, such as two or three α-helices or β-strands, or a combination of strands and helices [4]. A characteristic of repeat protein structures is that the residue contacts are contacts within a structural unit or between adjacent units. A central question is how this characteristic structural feature affects the folding behavior of these proteins [5–9].

A particularly interesting class of solenoid proteins are proteins that contain β-helices. In a parallel β-helix, the polypeptide backbone winds up to form a right- or left-handed helix [10, 11]. Each helical winding consists of three short β-strands, which are connected by turns and loop regions of different length. The strands form three elongated β-sheets, which constitute the walls of a prism-like motif. Parallel β-helix structures are also assumed for amyloid fibrils [12–14], one-dimensional protein aggregates that can cause severe diseases such as Alzheimer's, Parkinson's, or the variant Creutzfeldt-Jakob disease [15]. Understanding the folding and stability of β-helix proteins therefore may provide important insights on these aggregates.

We focus here on a protein with a right-handed β-helix, the tailspike protein

of *Salmonella* bacteriophage P22. Right-handed parallel β-helix motifs have been observed in different organisms, e.g. in other bacteriophage tailspikes like Sf6 [16, 17], HK620 [18] and Det7 [19] and in many bacterial, plant and fungal enzymes involved in degradation and modification of polysaccharides [20]. With computational methods, right-handed β-helices have been predicted for a large number of proteins, mostly from bacteria and fungi [21–24]. Unlike other solenoid proteins, such as left-handed β-helix proteins, leucine-rich-repeat proteins or ankyrins, right-handed β-helix proteins do not show internal sequence repeats [2, 4, 25]. Nevertheless, repetition as a structural feature seems to play an important role in parallel β-helix stability and folding. In particular, linear side chain stacking along the helix axis is presumably one of the major factors contributing to their stability [26, 27]. In such stacks, side chains of adjacent coils are aligned and oriented in a similar direction. Three different classes of stacks can be observed in right-handed β-helix proteins: polar stacks of Asn, Cys and Ser residues, aliphatic stacks of Ala, Val, Leu and Ile and aromatic stacks of Phe and Tyr [25].

The central domain of the P22 tailspike protein is a right-handed parallel β-helix with 13 windings, flanked by a N-terminal capsid-binding domain and an C-terminal trimerization domain [28, 29]. The stacks in the helix lumen contain predominantly hydrophobic residues such as Val, Ile, Leu, Phe, and Met. Three P22 tailspike protein monomers with a length of 666 residues form the trimeric tailspike, which mediates binding and hydrolysis of the *Salmonella* O-antigen polysaccharide [3,30,31]. The native tailspike trimer is highly resistant to detergent denaturation by sodium dodecyl sulfate (SDS). Via SDS-gel electrophoresis, the correctly folded, native tailspike trimer therefore can be distinguished by its reduced mobility from non-native oligomers and aggregates that are solubilised by SDS. Gel electrophoresis thus allows to follow and characterize the folding efficiency of P22 tailspike proteins, both in vivo and in vitro [32–35].

The folding of the P22 tailspike protein has been investigated via extensive

mutational analyses [26, 35–39]. Two main phenotypes have been observed: Temperature-sensitive-folding (tsf) mutations reduce the folding efficiency and lead to aggregation and inclusion body formation at elevated temperatures. In contrast, global suppressor (su) mutations increase the tailspike folding efficiency at high temperatures. A link between these phenotypes and protein stability has been found in mutational studies of an isolated tailspike fragment (see Fig. 1). This fragment comprises only the central β-helix domain, does not trimerize, and exhibits reversible folding transitions. The reverse folding allows to measure the protein stability, the free energy difference between the native, folded state and the denatured, unfolded state of the protein. Stabilities have been measured for the wildtype, two tsf mutants, and four su mutants [26]. The two tsf mutations decrease the protein stability, whereas the su mutations all increase the stability.

Stability measurements for the full P22 tailspike protein are complicated by the nearly irreversible trimerization of the protein. In this article, we provide further evidence for a close link between the folding efficiency and stability of the tailspike protein by calculating the mutation-induced stability changes for a large number of tailspike mutants. We observe significant correlations between the calculated stability changes and experimentally measured folding efficiencies in *E. coli* at the temperatures 18°C, 30°C and 37°C. Our focus here is on folding efficiency data from a recent extensive mutational analysis by Simkovsky and King [35], in which 145 residues throughout the isolated β-helix domain of the tailspike protein have been individually mutated to alanine. The calculations are performed with the program FoldX [40, 41], which has been specifically designed for determining stability changes, and calibrated against a large set of mutational data on small single-domain proteins.

In addition, we calculate the energy landscape of the P22 tailspike protein to explore the . Helices are typically assumed to form via a nucleation-propagation mechanism [42, 43]. After nucleation, which may occur in central or terminal helix regions, the helix grows during propagation via successive addition of

adjacent helical turns. In such a nucleation-propagation scenario, partially folded states of the helix thus can be characterized by a contiguous stretch of native-like structured residues. Each of these states can be described by two variables, for example the center and the length of the native stretch of residues. We have calculated the stability of each partially folded state with the program FoldX, which results in a three-dimensional energy landscape. This energy landscape reveals a dominant folding route on which the protein nucleates in the N-terminal region of the helix, with subsequent processive formation of the central β-helix region and C-terminal region. The energy landscape sections that correspond to the formation of the central β-helix region are rather flat, which may help to understand why mutations in this region have a particularly strong effect on the folding efficiency.

Results

Stability changes and folding efficiencies

Methods to predict mutation-induced stability changes have been investigated intensively in the past years [40, 41, 44–51]. We calculate here the stability changes for mutants of the P22 tailspike protein with the program FoldX. The force field of the FoldX program includes terms for backbone and sidechain entropies, which have been weighted against other terms using experimental data from mutational stability analyses of a variety of proteins [40]. FoldX has been tested on a set of 1088 point mutants, and reproduces the stability changes of 1030 of these mutants with a Pearson correlation coefficient of 0.83 and a standard deviation of 0.81 kcal/mol [40]. Stability calculations with the FoldX program have been used in protein design [52], evolutionary fitness models [53–55], and folding models [56–60].

As a first test, we compare here calculated stability changes with experimentally measured stability changes for mutants of the isolated β-helix domain

of the P22 tailspike protein shown in Fig. 1. In contrast to the full length protein, the isolated β-helix domain exhibits reversible folding transitions required for stability measurements. Two of the six experimentally considered single-residue mutations decrease the stability, while four mutations increase the stability [26]. As shown in Table 1, FoldX correctly predicts the sign of the stability change for five of the six mutants and only errs on the sign for the mutant V331A with the smallest experimentally measured stability change.

The four stabilizing mutations of Table 1 increase the folding efficiency of the P22 tailspike protein, while the two destabilizing mutations decrease the folding efficiency, which has led to the suggestion that the folding efficiency and stability of this protein are closely linked [26]. We investigate this link here by comparing calculated stability changes and *in vivo* folding efficiencies of a large set of P22 tailspike mutants. In the mutational analysis of Simkovsky and King [35], 145 nonalanine sites in the P22 tailspike protein were individually mutated to alanine. These sites include 103 non-alanine residues in the hydrophobic lumen of the β-helix, and all nonalanine residues in the N-terminal and C-terminal capping regions of the isolated β-helix domain. The mutants were expressed in *E. coli* at the temperatures 18°C, 30°C and 37°C, and the folding efficiency of the SDS-resistant trimer was measured *via* gel electrophoresis.

We have calculated the stability change for 144 of these 145 mutations. The stability change for one of the mutations, the mutation G511A, cannot be calculated since the mutated residue is part of a short disordered loop that is not resolved in the crystal structure of the protein [28, 30], see Fig. 1. Plots of the experimentally measured folding efficiencies of the P22 mutants versus the calculated stability changes $\Delta\Delta G$ are shown in Fig. 2. Despite substantial scattering of the data, a clear correlation between these two quantities can be observed. The folding efficiencies tend to decrease with the stability changes. At the temperature T=37°C, relatively small changes of the stability already have a rather drastic effect on the folding efficiency. At the temper-

ature T=18°C, the folding efficiency is more robust with respect to stability changes.

The correlations between stability change and folding efficiency can be quantified by the Spearman correlation coefficient r_S and the Pearson correlation coefficient r_P. In principle, the Spearman coefficient r_S here is preferable to the Pearson coefficient r_P since it tests for monotous relations, not necessarily linear relations as the Pearson coefficient. A Spearman coefficient of -1 indicates a monotonously decreasing relation between two quantities, and a Spearman coefficient of 1 a monotonously increasing relation. Nonetheless, also the Spearman coefficients obtained here have to be interpreted with care. At the temperature T=18°C, for example, the folding efficiency appears to be constant for mutations with small stability changes, within error bounds, rather than nonlinearly decreasing. This results in a, in absolute values, smaller Spearman coefficient of $r_S = -0.42$ at this temperature, compared to the coefficients $r_S = -0.68$ and -0.66 at the temperatures T=37°C and T=30°C, respectively. However, all three correlations are highly significant, with p-values of $3.6 \cdot 10^{-21}$ for $r_S = -0.68$, $1.2 \cdot 10^{-19}$ for $r_S = -0.66$, and $8.0 \cdot 10^{-8}$ for $r_S = -0.42$. The p-value is the probability for obtaining the observed correlation, or a stronger one, by chance.

Do stability changes have different effects on the folding efficiency in different regions of the protein? The β-helix of the tailspike protein has two main loops, a large structured loop from residues 198 to 256 ('dorsal fin', see Fig. 1), and a smaller partially unstructured loop from residues 397 to 426. Simkovsky and King [35] have not considered mutations in these loop regions. The performed mutations therefore fall into three regions, divided by these two loops. Plots of the folding efficiency versus the stability changes for mutations in these regions are shown in Fig. 3. We obtain high Spearman coefficients of $r_S = -0.76$ and $r_S = -67$ in the N- and C-terminal regions. In the central region, the majority of mutations has a rather drastic effect on the folding efficiency, even for small stability changes, which results in a smaller yet still highly significant

Spearman coefficient of $r_S = -0.60$. This is in agreement with the clustering of the majority of the sites of tsf mutations generated by random mutagenesis in the central part of the β-helix domain [61].

The stabilities calculated here should not be confused with the thermostabilities at 80°C and 90°C measured by Simkovsky and King [35]. The thermostability, defined as the percentage of native chains after heating, is a non-equilibrium quantity that reflects momomerization and unfolding barriers of the protein. The stability, in contrast, is the equilibrium free energy difference between the native and the denatured state. The thermostabilities at 80°C and 90°C measured by Simkovsky and King are not correlated with the folding efficiencies of the P22 tailspike protein mutants at 37°C.

Energy landscape

Statistical-thermodynamical models of helix formation, pioneered by Zimm and Bragg [42, 43], usually assume that each residue can be in either of two states: helically structured (H) or coil-like unstructured (C). Partially folded states of a helix then can be described by sequences of H's and C's, where each letter in the sequence indicates whether the corresponding residue is in state H or C. In the past decade, this modeling framework has also been widely used for protein folding, with the central modeling assumption that each residue of a protein is either native-like structured or unstructured [56, 62–70]. To reduce the number of possible partially folded states in such models, the native-like structured residues are often taken to occur in one, two, or three contiguous stretches. [1]

We present here a model for the β-helix domain of the P22 tailspike protein in which we assume that the native-like structured residues in a partially folded state occur in a single contiguous stretch. This assumption is reasonable

[1] For the models of Munoz and Eaton [63] and Wako and Saito [70], however, exact solutions of equilibrium properties in the full state space have been found [71, 72].

for helices, which are typically thought to form via a processive nucleation-propagation mechanism [42,73]. Each partially folded state of the protein then can be described by two variables, e.g. the numbers of the first and last residue of the native-like structured stretch, or the center and the length of this stretch of residues. We use here the program FoldX to obtain stability estimates $\Delta G_{\text{I-D}}$ for each partially folded state I in this model. We first consider the whole β-helix domain, which comprises the residues 113 to 544 of the P22 tailspike protein (see Fig. 1). After iterative energy minimization with FoldX, we obtain the calculated stability $\Delta G_{\text{N-D}} = -5.0$ kcal/mol at T=10°C, which is close to the experimental stability of -7.6 kcal/mol measured at this temperature [26]. To estimate the stability of a partially folded state I of the β-helix domain, we then simply 'cut out' the residues structured in this state from the energy-minimized protein structure, and calculate the stability of this 'substructure' with FoldX. The unstructured ends of the protein in the partially folded state can be neglected since their free energy contribution in this state is close to the free energy contribution in the denatured state.

The obtained energy landscape for the isolated β-helix domain is shown in Fig. 4. All mutations in the full-length protein considered by Simkovsky and King lie in this domain. In the landscape, the native state of the protein corresponds to the right corner of the triangle, with low free energies shown in blue. In this state, the length of the native-like structured stretch is equal to the length of the whole β-helix domain. The denatured state corresponds to model states along the left line of the triangle. The denatured state is separated from the native state by a free-energy barrier shown in red. The folding routes on this landscape are routes from the left line to the right corner of the triangle. Folding nucleation steps correspond to reaching saddlepoints of the red free-energy barrier, and propagation steps to diffusing down the barrier towards the native state at the right triangle corner. On the main folding routes, the high-energy sections of the landscape are avoided, and the protein nucleates either at the N- or the C-terminus.

The dominant folding route on this landscape with the saddlepoint of lowest energy is the route on which the protein nucleates in the N-terminal region. The saddle point of this route is located around the partially folded state with length 100 and center 170 of the native stretch of residues. The central region of the β-helix domain is a propagation region on this route. Interestingly, mutations with small stability changes in the central region have more drastic effects on the folding efficiency, compared to mutations with similar stability changes in the N- or C-terminal regions (see Fig. 3). This sensitivity in the central region of the protein may be related to the fact that the energy landscape on the N-terminal folding route beyond the saddle point is rather flat up to states with a native stretch of 300 residues. On this rather flat landscape section, green colors dominate, while blue colors for negative free energies occur only close to the native state. A relatively slow helix propagation on rather flat landscape sections may be vulnerable to competing non-productive folding or aggregation processes and, therefore, highly sensitive to small mutational stability changes. Overall, the relatively small experimentally measured stability of -7.6 kcal/mol at 10°C for the β-helix domain [26] makes it plausible that destabilizing mutations throughout this domain affect the folding efficiency of the full-length protein.

A processive nucleation-propagation mechanism for the tailspike protein has also been proposed by Simkovsky and King [35]. Based on the more drastic effects of mutations in the central helix region on folding efficiency, Simkovsky and King have proposed that nucleation occurs in this region. On the landscape of Fig. 4, in contrast, nucleation occurs in the N-terminal region of the helix, while the central helix region is formed during propagation.

Conclusions

We have found highly significant correlations between calculated stability changes and *in vivo* folding efficiencies of the P22 tailspike protein at 18°C,

30°C and 37°C. The folding efficiency of the protein decreases with increasing temperature. Mutations with small stability changes thus have a more drastic effect on the folding efficiency at higher temperatures (see Fig. 2). At 37°C, the central region of the helix is particularly sensitive to mutations with small stability changes (see Fig. 3). On the energy landscape of Fig. 4, the folding process nucleates in the N-terminal region, and propagates in the central region. The high sensitivity of the folding efficiency with respect to mutations in the central region may be related to rather flat landscape sections for helix propagation.

References

[1] E. M. Marcotte, M. Pellegrini, T. O. Yeates, D. Eisenberg, A census of protein repeats., J. Mol. Biol. 293 (1998) 151–160.

[2] M. A. Andrade, C. Perez-Iratxeta, C. P. Ponting, Protein repeats: structures, functions, and evolution, J. Struct. Biol. 134 (2001) 117–131.

[3] R. Seckler, Folding and function of repetitive structure in the homotrimeric phage P22 tailspike protein, J. Struct. Biol. 122 (1998) 216–222.

[4] B. Kobe, A. V. Kajava, When protein folding is simplified to protein coiling: the continuum of solenoid protein structures, Trends. Biochem. Sci. 25 (2000) 509–515.

[5] E. Kloss, N. Courtemanche, D. Barrick, Repeat-protein folding: new insights into origins of cooperativity, stability, and topology, Arch. Biochem. Biophys. 469 (2008) 83–99.

[6] E. R. G. Main, A. R. Lowe, S. G. J. Mochrie, S. E. Jackson, L. Regan, A recurring theme in protein engineering: the design, stability and folding of repeat proteins, Curr. Opin. Struct. Biol. 15 (2005) 464–471.

[7] C. C. Mello, D. Barrick, An experimentally determined protein folding energy landscape, Proc. Natl. Acad. Sci. USA 101 (2004) 14102–14107.

[8] A. R. Lowe, L. S. Itzhaki, Rational redesign of the folding pathway of a modular protein, Proc. Natl. Acad. Sci. USA 104 (2007) 2679–2684.

[9] N. Courtemanche, D. Barrick, The leucine-rich repeat domain of Internalin B folds along a polarized N-terminal pathway, Structure 16 (2008) 705–714.

[10] M. D. Yoder, F. Jurnak, Protein motifs. 3. The parallel β helix and other coiled folds, FASEB J. 9 (1995) 335–342.

[11] J. Jenkins, O. Mayans, R. Pickersgill, Structure and evolution of parallel β-helix proteins, J. Struct. Biol. 122 (1998) 236–246.

[12] R. Wetzel, Ideas of order for amyloid fibril structure, Structure 10 (2002) 1031–1036.

[13] R. Tycko, Progress towards a molecular-level structural understanding of amyloid fibrils, Curr. Opin. Struct. Biol. 14 (2004) 96–103.

[14] R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe, D. Eisenberg, Structure of the cross-β spine of amyloid-like fibrils, Nature 435 (2005) 773–778.

[15] C. M. Dobson, Protein folding and misfolding, Nature 426 (2003) 884–890.

[16] A. Freiberg, R. Morona, L. Van den Bosch, C. Jung, J. Behlke, N. Carlin, R. Seckler, U. Baxa, The tailspike protein of Shigella phage Sf6. A structural homolog of Salmonella phage P22 tailspike protein without sequence similarity in the β-helix domain, J. Biol. Chem. 278 (2003) 1542–1548.

[17] J. Müller, S. Barbirz, K. Heinle, A. Freiberg, R. Seckler, U. Heinemann, An inter-subunit active site between supercoiled parallel β-helices in the trimeric tailspike endorhamnosidase of Shigella flexneri phage Sf6, Structure 16 (2008) 766–775.

[18] S. Barbirz, J. J. Müller, C. Utrecht, A. J. Clark, U. Heinemann, R. Seckler, Crystal structure of tailspike of E. coli phage HK620 reflects evolutionary relationship of three podoviridae phages, Molec. Microbiol., in press.

[19] M. Walter, C. Fiedler, R. Grassl, M. Biebl, R. Rachel, X. L. Hermo-Parrado, A. L. Llamas-Saiz, R. Seckler, S. Miller, M. J. van Raaij, Structure of the receptor-binding protein of bacteriophage Det7: A podoviral tailspike in a myovirus, J. Virol. 82 (2008) 2265–2273.

[20] A. V. Kajava, A. C. Steven, β-rolls, β-helices, and other β-solenoid proteins, Adv. Protein Chem. 73 (2006) 55–96.

[21] P. Bradley, L. Cowen, M. Menke, J. King, B. Berger, Betawrap: successful prediction of parallel β-helices from primary sequence reveals an association with many microbial pathogens, Proc. Natl. Acad. Sci. USA 98 (2001) 14819–14824.

[22] F. D. Ciccarelli, R. R. Copley, T. Doerks, R. B. Russell, P. Bork, CASH – a β-helix domain widespread among carbohydrate-binding proteins, Trends. Biochem. Sci. 27 (2002) 59–62.

[23] A. V. McDonnell, M. Menke, N. Palmer, J. King, L. Cowen, B. Berger, Fold recognition and accurate sequence-structure alignment of sequences directing β-sheet proteins, Proteins 63 (2006) 976–985.

[24] M. Junker, C. C. Schuster, A. V. McDonnell, K. A. Sorg, M. C. Finn, B. Berger, P. L. Clark, Pertactin β-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins, Proc. Natl. Acad. Sci. USA 103 (2006) 4918–4923.

[25] J. Jenkins, R. Pickersgill, The architecture of parallel β-helices and related folds, Prog. Biophys. Mol. Biol. 77 (2001) 111–175.

[26] B. Schuler, R. Seckler, P22 tailspike folding mutants revisited: effects on the thermodynamic stability of the isolated β-helix domain, J. Mol. Biol. 281 (1998) 227–234.

[27] D. E. Kamen, Y. Griko, R. W. Woody, The stability, structural organization, and denaturation of pectate lyase C, a parallel β-helix protein, Biochemistry 39 (51) (2000) 15932–15943.

[28] S. Steinbacher, R. Seckler, S. Miller, B. Steipe, R. Huber, P. Reinemer, Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer, Science 265 (1994) 383–386.

[29] S. Steinbacher, S. Miller, U. Baxa, N. Budisa, A. Weintraub, R. Seckler, R. Huber, Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 A, fully refined structure of the endorhamnosidase at 1.56 A resolution, and the molecular basis of O-antigen recognition and cleavage, J. Mol. Biol. 267 (1997) 865–880.

[30] S. Steinbacher, U. Baxa, S. Miller, A. Weintraub, R. Seckler, R. Huber, Crystal structure of phage P22 tailspike protein complexed with Salmonella sp. O-antigen receptors, Proc. Natl. Acad. Sci. USA 93 (1996) 10584–10588.

[31] U. Baxa, S. Steinbacher, S. Miller, A. Weintraub, R. Huber, R. Seckler, Interactions of phage P22 tails with their cellular receptor, Salmonella O-antigen polysaccharide, Biophys. J. 71 (1996) 2040–2048.

[32] D. P. Goldenberg, P. B. Berget, J. King, Maturation of the tail spike endorhamnosidase of Salmonella phage P22, J. Biol. Chem. 257 (1982) 7864–7871.

[33] R. Seckler, A. Fuchs, J. King, R. Jaenicke, Reconstitution of the thermostable trimeric phage P22 tailspike protein from denatured chains in vitro, J. Biol. Chem. 264 (1989) 11750–11753.

[34] A. Fuchs, C. Seiderer, R. Seckler, In vitro folding pathway of phage P22 tailspike protein, Biochemistry 30 (1991) 6598–6604.

[35] R. Simkovsky, J. King, An elongated spine of buried core residues necessary for in vivo folding of the parallel β-helix of P22 tailspike adhesin., Proc. Natl. Acad. Sci. USA 103 (2006) 3575–3580.

[36] D. P. Goldenberg, J. King, Temperature-sensitive mutants blocked in the folding or subunit of the bacteriophage P22 tail spike protein. II. Active mutant proteins matured at 30 degrees C, J. Mol. Biol. 145 (1981) 633–651.

[37] M. Beissinger, S. C. Lee, S. Steinbacher, P. Reinemer, R. Huber, M. H. Yu, R. Seckler, Mutations that stabilize folding intermediates of phage P22 tailspike protein: folding in vivo and in vitro, stability, and structural context, J. Mol. Biol. 249 (1995) 185–194.

[38] B. Schuler, F. Furst, F. Osterroth, S. Steinbacher, R. Huber, R. Seckler, Plasticity and steric strain in a parallel β-helix: rational mutations in the P22 tailspike protein, Proteins 39 (2000) 89–101.

[39] S. Betts, C. Haase-Pettingell, K. Cook, J. King, Buried hydrophobic side-chains essential for the folding of the parallel β-helix domains of the P22 tailspike, Protein Sci. 13 (2004) 2291–2303.

[40] R. Guerois, J. E. Nielsen, L. Serrano, Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations, J. Mol. Biol. 320 (2002) 369–387.

[41] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, The FoldX web server: an online force field, Nucleic Acids Res. 33 (2005) W382–W388.

[42] B. H. Zimm, J. K. Bragg, Theory of the phase transition between helix and random coil, J. Chem. Phys. 31 (1959) 526–535.

[43] K. A. Dill, S. Bromberg, Molecular driving forces: Statistical thermodynamics in chemistry and biology, Garland, 2002.

[44] D. Gilis, M. Rooman, PoPMuSiC, an algorithm for predicting protein mutant stability changes: application to prion proteins, Protein Eng. 13 (2000) 849–856.

[45] H. Zhou, Y. Zhou, Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction, Protein. Sci. 11 (2002) 2714–2726.

[46] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, Nucleic Acids Res. 33 (2005) W306–W310.

[47] J. Cheng, A. Randall, P. Baldi, Prediction of protein stability changes for single-site mutations using support vector machines, Proteins 62 (2006) 1125–1132.

[48] L.-T. Huang, K. Saraboji, S.-Y. Ho, S.-F. Hwang, M. N. Ponnuswamy, M. M. Gromiha, Prediction of protein mutant stability using classification and regression tool, Biophys. Chem. 125 (2007) 462–470.

[49] V. Parthiban, M. M. Gromiha, C. Hoppe, D. Schomburg, Structural analysis and prediction of protein mutant stability using distance and torsion potentials: role of secondary structure and solvent accessibility, Proteins 66 (2007) 41–52.

[50] M. Bueno, C. Camacho, J. Sancho, SIMPLE estimate of the free energy change due to aliphatic mutations: Superior predictions based on first principles, Proteins 68 (2007) 850–862.

[51] S. Yin, F. Ding, N. V. Dokholyan, Modeling backbone flexibility improves protein stability estimation, Structure 15 (2007) 1567–1576.

[52] A. M. van der Sloot, V. Tur, E. Szegezdi, M. M. Mullally, R. H. Cool, A. Samali, L. Serrano, W. J. Quax, Designed tumor necrosis factor-related apoptosis-inducing ligand variants initiating apoptosis exclusively via the DR5 receptor, Proc. Natl. Acad. Sci. USA 103 (2006) 8634–8639.

[53] J. D. Bloom, J. J. Silberg, C. O. Wilke, D. A. Drummond, C. Adami, F. H. Arnold, Thermodynamic prediction of protein neutrality, Proc. Natl. Acad. Sci. USA 102 (3) (2005) 606–611.

[54] S. Bershtein, M. Segal, R. Bekerman, N. Tokuriki, D. Tawfik, Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein, Nature 444 (2006) 929–932.

[55] N. Tokuriki, F. Stricher, J. Schymkowitz, L. Serrano, D. S. Tawfik, The stability effects of protein mutations appear to be universally distributed, J. Mol. Biol. 369 (2007) 1318–1332.

[56] R. Guerois, L. Serrano, The SH3-fold family: Experimental evidence and prediction of variations in the folding pathways, J. Mol. Biol. 304 (2000) 967–982.

[57] M. A. Micheelsen, C. Rischel, J. Ferkinghoff-Borg, R. Guerois, L. Serrano, Mean first-passage time analysis reveals rate-limiting steps, parallel pathways and dead ends in a simple model of protein folding, Europhys. Lett. 61 (2003) 561–566.

[58] K. Lindorff-Larsen, E. Paci, L. Serrano, C. M. Dobson, M. Vendruscolo, Calculation of mutational free energy changes in transition states for protein folding, Biophys. J. 85 (2003) 1207–1214.

[59] T. R. Weikl, K. A. Dill, Transition states in protein folding kinetics: The structural interpretation of Φ-values, J. Mol. Biol. 365 (2007) 1578–1586.

[60] T. R. Weikl, Transition states in protein folding kinetics: Modeling ϕ-values of small β-sheet proteins, Biophys. J. 94 (2008) 929–937.

[61] C. Haase-Pettingell, J. King, Prevalence of temperature sensitive folding mutations in the parallel β coil domain of the phage P22 tailspike endorhamnosidase, J. Mol. Biol. 267 (1997) 88–102.

[62] E. Alm, D. Baker, Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures, Proc. Natl. Acad. Sci. USA 96 (1999) 11305–11310.

[63] V. Muñoz, W. A. Eaton, A simple model for calculating the kinetics of protein folding from three-dimensional structures, Proc. Natl. Acad. Sci. USA 96 (1999) 11311–11316.

[64] O. V. Galzitskaya, A. V. Finkelstein, A theoretical search for folding/unfolding nuclei in three-dimensional protein structures, Proc. Natl. Acad. Sci. USA 96 (1999) 11299–11304.

[65] E. Alm, A. V. Morozov, T. Kortemme, D. Baker, Simple physical models connect theory and experiment in protein folding kinetics, J. Mol. Biol. 322 (2002) 463–476.

[66] J. Karanicolas, C. L. Brooks III, The importance of explicit chain representation in protein folding models: An examination of Ising-like models, Proteins 53 (2003) 740–747.

[67] E. D. Nelson, N. V. Grishin, Alternate pathways for folding in the flavodoxin fold family revealed by a nucleation-growth model, J. Mol. Biol. 358 (2006) 646–653.

[68] V. J. Hilser, B. Garcia-Moreno, T. G. Oas, G. Kapp, S. T. Whitten, A statistical thermodynamic model of the protein ensemble, Chem. Rev. 106 (2006) 1545–1558.

[69] R. Zwanzig, Simple model of protein folding kinetics, Proc. Natl. Acad. Sci. 92 (1995) 9801–9804.

[70] H. Wako, N. Saito, Statistical mechanical theory of protein conformation. 1. General considerations and application to homopolymers, J. Phys. Soc. Jpn. 44 (1978) 1931–1938.

[71] P. Bruscolini, A. Pelizzola, Exact solution of the Muñoz-Eaton model for protein folding, Phys. Rev. Lett. 88 (2002) 258101.

[72] P. Bruscolini, A. Pelizzola, M. Zamparo, Rate determining factors in protein model structures, Phys. Rev. Lett. 99 (2007) 038103.

[73] A. Chakrabartty, R. L. Baldwin, Stability of α-helices, Adv. Protein. Chem. 46 (1995) 141–176.

[74] J. H. Zar, Significance testing of the Spearman rank correlation coefficient, J. Am. Stat. Assoc. 67 (1972) 578–580.

Table 1: Comparison of experimental and calculated stability changes for mutants of the isolated β-helix domain

mutation	$\Delta\Delta G_{\mathrm{exp}}$	$\Delta\Delta G_{\mathrm{FoldX}}$
D238S	1.0	2.2
G244R	2.4	2.2
V331A	−0.7	0.6
V331G	−1.4	−0.8
A334I	−1.2	−2.8
A334V	−1.7	−3.7

Experimental stability changes $\Delta\Delta G_{\mathrm{exp}}$ are from Ref. [26]. The stability changes $\Delta\Delta G_{\mathrm{FoldX}}$ have been calculated at the temperature T=10°C of the experiments with FoldX 3.0. The calculated stability changes are averages over 10 independent runs with the FoldX mutation command 'BuildModel', after minimization of the FoldX free energy for the wildtype structure (protein data bank file 1TYV) with the command 'RepairPDB'. Both FoldX commands imply adjustments of side chain rotamers for fixed main chain atoms. The stability changes are given in units of kcal/mol.

Table 2: Calculated stability changes for P22 tailspike mutations

mutation	$\Delta\Delta G$	mutation	$\Delta\Delta G$	mutation	$\Delta\Delta G$
K122A	0.05	F160A	5.61	M292A	4.43
Y123A	1.04	K163A	1.03	V293A	2.43
S124A	-0.14	L165A	3.67	P298A	3.43
V125A	1.92	T166A	0.86	G300A	2.38
K126A	1.33	I167A	5.24	G301A	7.13
L127A	3.54	C169A	0.93	I305A	2.66
S128A	-0.58	K172A	1.29	I306A	4.92
D129A	0.42	F173A	5.61	F308A	5.52
Y130A	3.67	G175A	10.14	N310A	1.89
P131A	1.06	L179A	4.56	N319A	2.85
T132A	1.65	F181A	5.15	V321A	3.51
L133A	3.82	L184A	4.28	G324A	9.83
Q134A	1.54	S188A	-0.05	T326A	0.76
D135A	0.38	I190A	4.87	G329A	2.95
S138A	-0.09	V193A	2.47	S333A	-0.76
V141A	0.45	M195A	4.61	F336A	5.37
D142A	0.02	S260A	-0.22	N339A	2.06
G143A	-0.44	L262A	4.50	G347A	2.62
L144A	4.31	I264A	4.53	V349A	3.94
L145A	2.21	C267A	0.91	F352A	5.04
I146A	3.99	V270A	3.2	S354A	-0.25
R148A	1.41	V272A	3.82	G361A	4.58
Y150A	4.36	G277A	7.88	V362A	3.78
F152A	5.32	F282A	5.69	T364A	1.07
Y153A	1.30	F284A	5.24	N376A	0.53
E156A	1.11	C287A	0.56	L379A	4.54
V158A	3.67	C290A	0.61	F381A	5.61

mutation	$\Delta\Delta G$	mutation	$\Delta\Delta G$	mutation	$\Delta\Delta G$
S384A	-0.52	G462A	2.32	V523A	3.85
V386A	3.43	L465A	4.51	S524A	0.51
P389A	2.63	T467A	0.96	G525A	3.05
G393A	3.93	S470A	-1.04	I526A	3.51
F394A	5.88	F472A	5.37	T527A	0.37
L396A	4.98	I475A	4.06	G528A	5.36
L424A	2.26	I477A	4.28	M529A	0.12
H426A	1.48	T480A	0.66	V530A	1.27
I428A	4.97	Q489A	0.70	D531A	1.98
L431A	4.44	I490A	4.67	P532A	1.86
V433A	3.87	I492A	4.31	S533A	-1.22
G440A	1.47	S493A	-0.51	R534A	1.56
F441A	6.29	C496A	1.14	I535A	3.87
M443A	4.53	V498A	3.93	N536A	1.83
G445A	3.7	N499A	0.59	V537A	3.12
M448A	2.65	L501A	5.36	N539A	0.08
V450A	3.77	L503A	5.11	L540A	2.80
I453A	4.76	T515A	1.03	E542A	0.48
V455A	3.81	I516A	4.80	E543A	0.04
C458A	0.86	S521A	-0.60	G544A	0.33
S461A	-0.40	T522A	0.60	L545A	0.58

$\Delta\Delta G$ values for tailspike mutants in units of kcal/mol calculated with FoldX 3.0 at the temperature T=37°C and the ionic strength 154 mM, which is an estimate for the ionic strength in the interior of Solmonella. The values are averages over 10 independent runs with the FoldX mutation command 'BuildModel', after minimization of the FoldX free energy for the wildtype structure (protein data bank file 1TYV) with the command 'RepairPDB'. Both FoldX commands imply adjustments of side chain rotamers for fixed main

chain atoms. The calculated stability changes at the temperatures T=30°C and T=18°C are identical with the values at T=37°C within the statistical errors estimated from the 10 independent runs.

Figure captions

Figure 1: Crystal structure of the isolated β-helix domain of the P22 tailspike protein (residues 113 to 544) [28, 30]. The three elongated β-sheets of the helix or indicated in different colors. The N-terminus of the helix is protected by an α-helical capping region. Two of the loops that connect the strands of the β-sheets are partially unstructured in the crystal (residues 402 to 408 and residues 510 to 514, see open ends in the structural representation). The 'dorsal fin' of the protein is a long structured loop from residues 198 to 256.

Figure 2: Calculated stability changes $\Delta\Delta G_{\text{N-D}}$ versus *in vivo* folding efficiencies of P22 tailspike mutants at the temperatures 37°C, 30°C, and 18°C. The experimentally measured folding efficiencies [35] are given in % of the folding efficiency for the wildtype protein. Mutations from glycene to the larger alanine are indicated by squares. The large stability changes for some of these mutations result from Van der Waals clashes, with may be overestimated since FoldX only relaxes side chain atoms but not backbone atoms. However, the values of the Spearman correlation coefficient r_S and Pearson correlation coefficient r_P only change by 1 or 2 % if mutations from glycenes to alanines are excluded. The p-values for the given Spearman correlation coefficients are $3.6 \cdot 10^{-21}$, $1.2 \cdot 10^{-19}$, and $8.0 \cdot 10^{-8}$. Following a standard procedure, we have calculated the p-values from a correspondence between the r_S-distribution and Student's t-distribution for large data sets [74].

Figure 3: Calculated stability changes $\Delta\Delta G$ versus *in vivo* folding efficiencies for P22 tailspike mutations in different regions at the temperature 37°C. The experimentally measured folding efficiencies for the mutants [35] are given in % of the wildtype folding efficiency. In the N-terminal region, the mutated residue positions range from 122 to 195 (see also Table 2). The residue positions for the alanine substitutions range from 260 to 396 in the central region, and from 442 to 545 in the C-terminal region. The three regions are separated by two loops in which no mutations have been performed by Simkovsky and King [35]. Mutations from glycene to the larger alanine are indicated by squares. The

correlations here are quantified both by the Spearman coefficient r_S and the Pearson coefficient r_P. The p-values for the Spearman correlation coefficients are $1.2 \cdot 10^{-9}$, $6.0 \cdot 10^{-9}$, and $8.4 \cdot 10^{-6}$, respectively.

Figure 4: Three-dimensional model energy landscapes of the isolated β-helix domain (see Fig. 1) for the temperature T=10°C at which the stability has been measured experimentally [26], see text. In the model, partially folded states are characterized by the length and the center position of the native-like structured stretch of residues. In the state with length 41 and center 200 of the native-like stretch, for example, the residues 180 to 220 are native-like structured, while all other residues are still unstructured. The folded state corresponds to the right corner of the triangle where all 432 residues of the β-helix domain are structured. The colors represent FoldX free-energy differences between the partially folded states and the unfolded state. Contour lines are drawn every 5 kcal/mol. In calculating the energy landscape, we have assumed that the two unstructured loop fragments (see Fig. 1) do not contribute to the free-energy differences between partially folded states and the unfolded state.

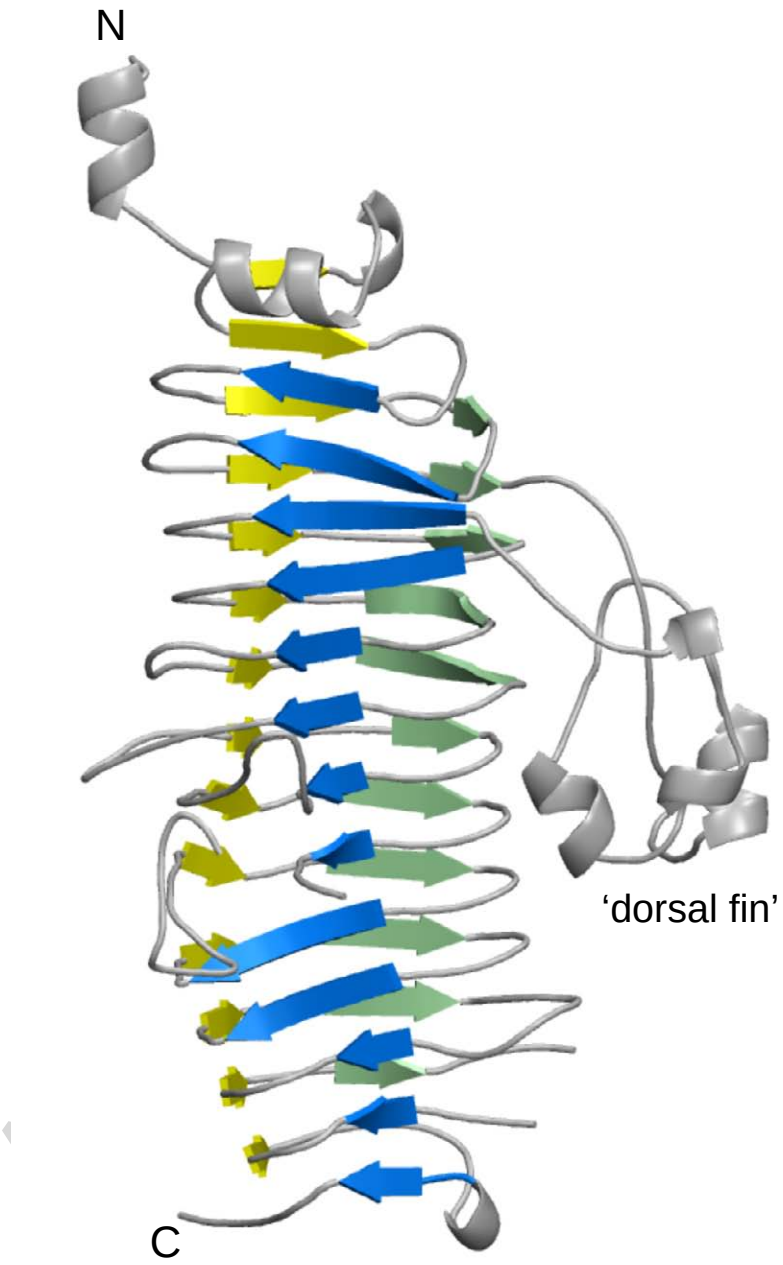


Fig. 1.

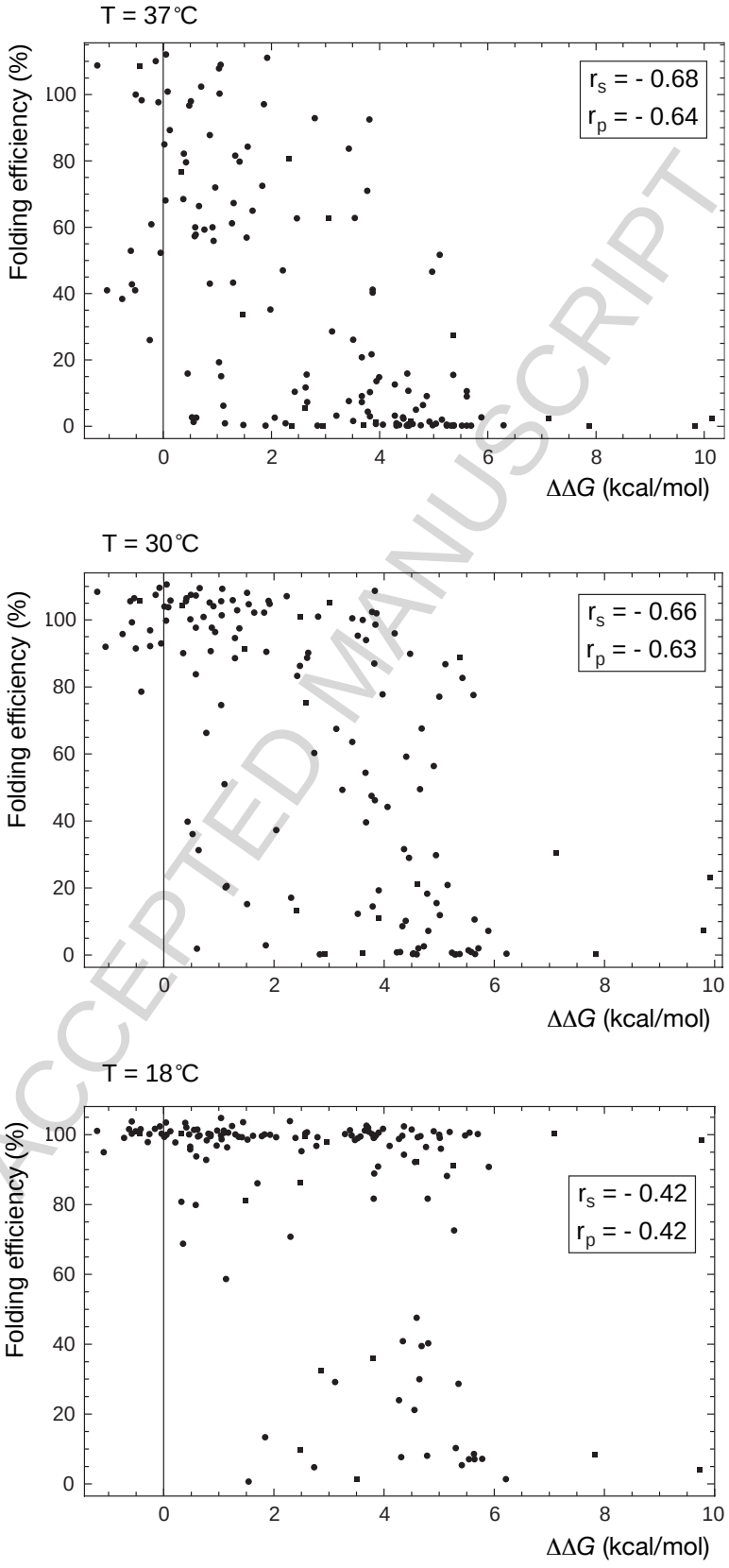


Fig. 2.

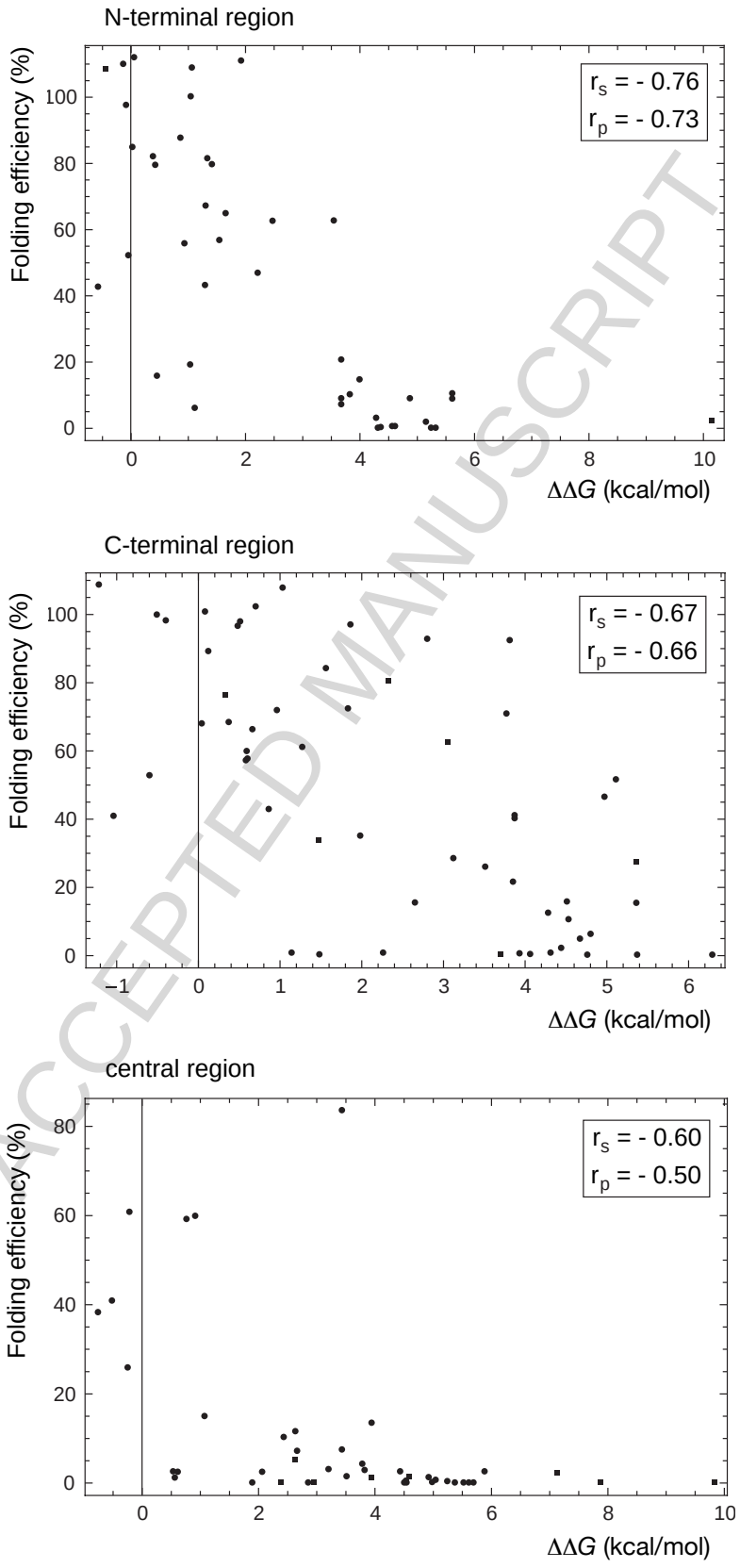


Fig. 3.

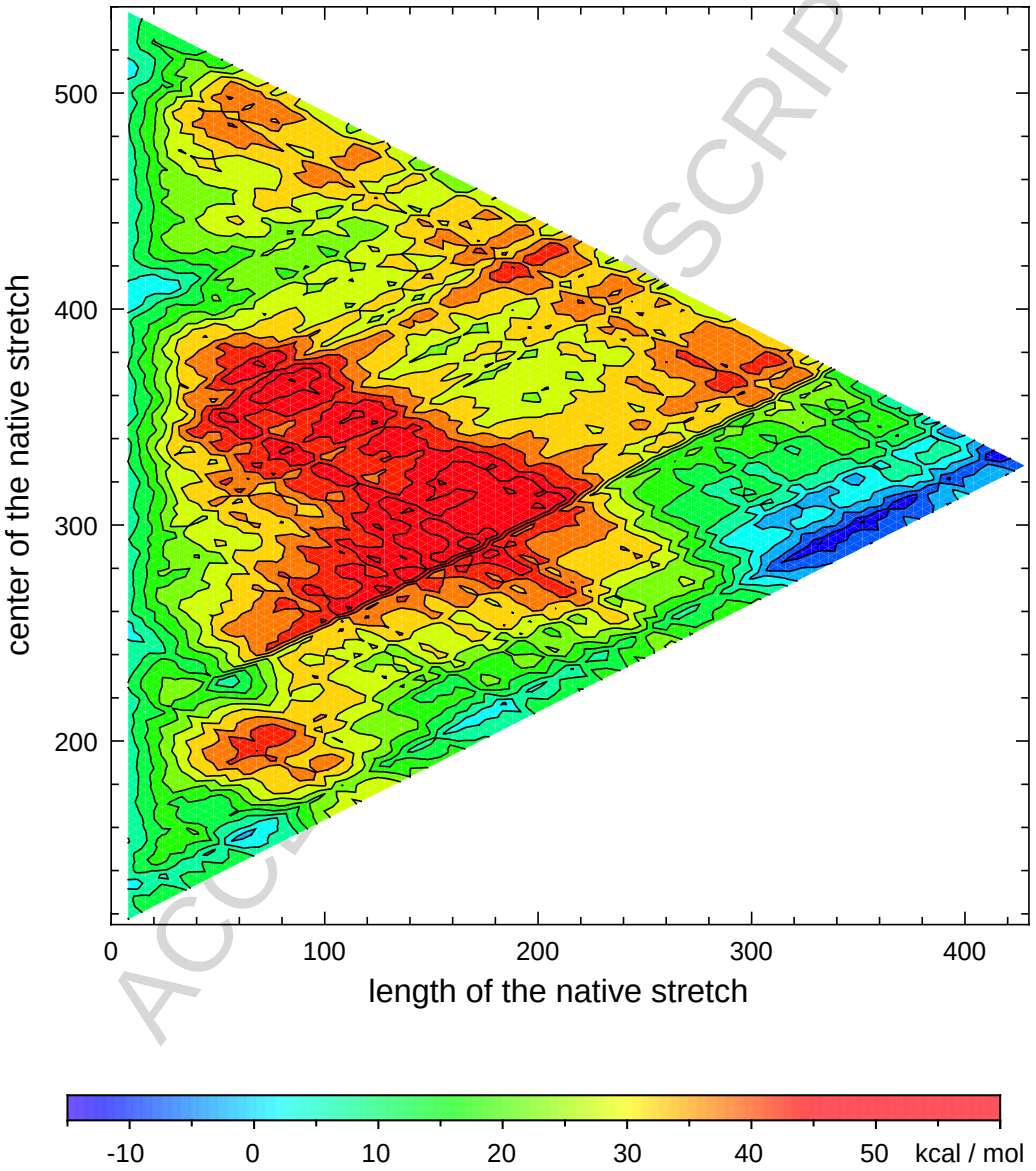


Fig. 4.