STRUCTURAL BASIS OF PROTEIN METASTABILITY

Peter Flecker

Johannes Gutenberg University, Department of Chemistry and Pharmacy, Duesbergweg 10-14, D-55128 Mainz, FRG. E-mail: flecker@mail.uni-mainz.de

Keywords:
protein folding, protein metastability, protease inhibitors, local energy minimum

Abstract
This review article summarises a series of mutational and structural investigations with the double-headed Bowman Birk inhibitor (BBI) of trypsin and chymotrypsin. BBI represents an highly unusual ´inside-out´ structure with a polar interior and a hydrophobic exterior constrained by a rigid framework of disulfide bridges. This molecule none the less responds extremely sensitively to mutations in marked contrast to most other globular proteins. The fully active state of these mutants can be reached by assisted refolding on a template with complementary structure in contrast to the autonomous refolding competence of the parent protein. It is shown by means of equilibration experiments that - at least for the variants - the fully active state represents a local rather than a global minimum on their free energy landscape. Compensatory structural adaptations within the hydrophobic core do not seem to be possible with this kind of conformationally constrained ´inside out´ structure.

Introduction

A central proposition of protein structural biology is that all information for an acquisition of the fully active conformation of proteins is encoded in their amino acid sequence. A general feature of protein folding processes seems to be that all essential structural features of a protein are formed rather early and retained to a significant extent until the final collapse into the fully active conformation [1]. Specific kinetic pathways that guide the nascent protein conformation into the native state along a single preferred pathway or a dynamic ensemble of flexible conformers [2] can unravel the classical problem [3] how the nascent polypeptide chain can find the native, fully active conformation on a biologically meaningful time scale.

There has been a considerable debate whether protein-folding reactions are under thermodynamic or kinetic reaction control [4-5]. Thermodynamic reaction control represents a random selection of the most stable species from all conformers that are possible for a given amino acid sequence. In kinetic reaction control, the preferred species formed along the preferred reaction pathway is the most rapidly but not necessarily the most stable one. The occurrence of the active state within a local minimum that can be converted into a more stable state by equilibration experiments would provide experimental proof for kinetic reaction control. However, this would be difficult to demonstrate on an experimentally feasible time scale in case of local energy minima (metastable states or pseudo

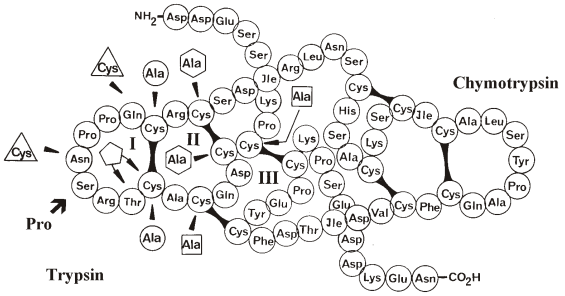

Fig. 1. Covalent structure of rBBI, the recombinant Bowman Birk inhibitor and the mutations.

equilibriua) separated from the most stable state by high barriers of activation. Indeed, the appearance of such relative energy minima in protein folding pathways has been proposed theoretically by model calculations [6-8]. In the present survey the results of our previous mutational [9-12] and structural [13-15] studies related to this problem will be discussed and related to the literature. The soybean Bowman Birk inhibitor of trypsin and chymotrypsin (termed as sBBI further below) was selected as a model protein because its double-headed structure should facilitate the detection of irregularities by amino acid replacements. As a parent protein of our mutational studies we have designed a Lys16Arg and Met27Ile double replacement in sBBI that will be termed in this article as the recombinant inhibitor rBBI.

The recombinant inhibitor was obtained by means of chemical synthesis, molecular cloning and expression of gene coding for rBBI [16]. The amino acid replacements in figure 1 were introduced into the first subdomain of BBI directed against trypsin while using the second subdomain directed against chymotrypsin as an internal standard. By means of this approach is possible to detect even structurally minor but functionally relevant irregularities and distinguish local perturbations within the first subdomain from more global ones transmitted into the second subdomain. The replacements in figure 1 induce substantial irregularities, greatly differing in their extent.

Structural Studies

In order to facilitate a clear-cut interpretation of these irregularities in became necessary to determine the crystal structure of the soybean Bowman Birk inhibitor [13] that is compared with a NMR structure [17] in figure 2. The structure of a related pea inhibitor could be solved with the crystal structure of the soybean inhibitor as a search model [14]. The structure of sBBI is built up from two antiparallel

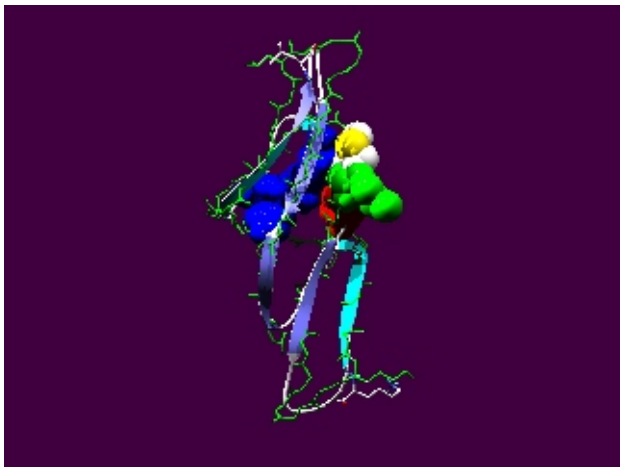

Fig. 2. Crystal and NMR structure of the soybean Bowman Birk inhibitor. The crystal structure is coloured in CPK, the NMR structure is depicted in green. The side chains of Lys16 and Leu43 at the scissile peptide bonds, and that of Met27 (in space filling) are shown in both models. The buried polar side chains of the acidic (red) Asp26 and the basic (blue) Arg28 and His33 are shown (in space filling) for the crystal structure only. The differing orientation of Met27 in the crystal and the NMR-structure is especially notable.

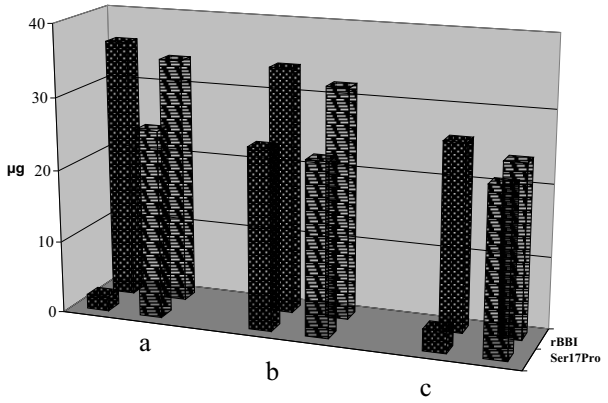

Fig. 3. Titrations of rBBI and the Ser17Pro variant with trypsin (dark grey bars) and chymotrypsin (hatched).

-hairpin motifs, each containing three antiparallel -strands arranged according to ABFCED topology [18]. The -strands A and B are connected by the trypsin-inhibition and the -strands D and E by the chymotrypsin-binding loop. The -strands are considerably twisted in a clockwise manner resulting in an almost perpendicular orientation of the two enzyme-inhibition loops in the crystal structure. One can see in figure 2, that the angle between the two loops in the NMR structure is reduced quite significantly to only 45 . There is a slight overall structural divergence of the two models. The rmsd of the C atoms of the two models is only 1.8 Å and the highest structural convergence occurs in the -sheets. More significant rmsd are observed between the side chains especially in the interdomain region. Larger differences between the crystal and the solution structure are observed with regards to the side chain conformation (see figure 4 in reference [13] for more details). This is especially true for the long side chain of Met27 projecting out to the surface of the molecule in opposite direction in the crystal and the solution structure as a result of opposite 1 and 3 angles and a slightly displaced C trace between Asp26 and Cys32. In the crystal structure the side chain of Met27 forms hydrogen bond contacts with the side chain of Gln48, thus connecting the chymotrypsin-inhibitory region with the centre of the molecule. This may account to the slight but significant effect upon the chymotrypsin-inhibitory activity induced by replacing Met27 of sBBI for an isoleucine in position 27 of rBBI. Met27 of sBBI has more recently turned out as an important determinant of molecular recognition and misrecognition of the second subdomain of sBBI[15]; discussed also in reference [19]). Close hydrophobic contacts with Asn97 of the second trypsin molecule that are possible for the longer side chain of Met27 in sBBI (but not the shorter side chain of Ile27 in rBBI) seem to provide for a hydrophobic environment. This must be regarded to be highly unfavourable for buried polar interactions across the P1-S1 subsite responsible for trypsin specificity.

Mutational studies

In analogy with the behaviour [20] of the soybean inhibitor (sBBI) the recombinant inhibitor rBBI folds spontaneously in a redox buffer system [16] into its fully active state. The stoichiometric 1:1 ratio of trypsin- and chymotrypsin-inhibitory material with rBBI (see second row in figure 3) is as expected by the covalent structure in figure 1.

The inhibition curves of the parent protein in figure 4a are indistinguishable before and after trypsin-affinity chromatography. The enzyme-inhibitory activities of the first and the second subdomain are both attained at rate constants of approximately 5 $10^{-4}s^{-1}$ and the correct ratio of the two subdomains is preserved after 24 hrs of incubation in the refolding buffer as shown in figure 4a. The Ki-values of both subdomains are in close accord with the properties of the soybean protein. The molecular mass of 7867.6 (0,5) Da, determined by electrospray mass spectrometry after affinity chromatography on trypsin-Sepharose [16] confirms the identity of the parent protein and does not support a proteolytic cleavage on the affinity matrix. For the parent protein, the notion [21] seemed to be confirmed, that all information required to reach a single and fully active minimum of free energy is contained within its amino acid sequence.

However, the variants deviated from an autonomous refolding competence as a basic tenet of protein structure and folding. The Ser17Pro replacement, constructed in order to account to the role of the hydrogen bonds at the back side of the scissile peptide bond as a highly conserved structural feature of Bowman Birk and other types of protease inhibitors [22-23] resulted in significant irregularities. As shown in the front row of figure 3, the amount of trypsin-inhibitory material was reduced to a small fraction of the chymotrypsin-inhibitory material with the Ser17Pro variant after refolding in solution. The inhibition curves of

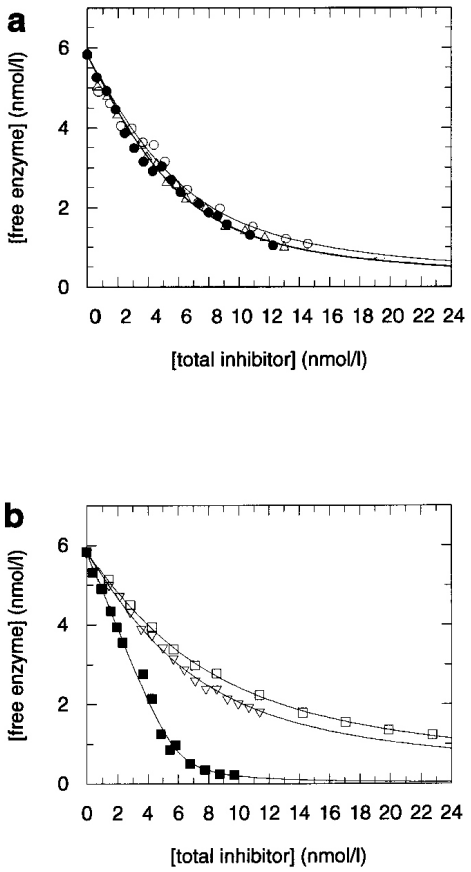

Fig. 4. Enzyme inhibition curves of the parent protein (a) and the Ser17Pro variant (b). The open symbols refer to refolding experiments in solution. The filled symbols refer to refolding experiments in the presence of trypsin-Sepharose. The triangles refer to the subsequent equilibration in solution.

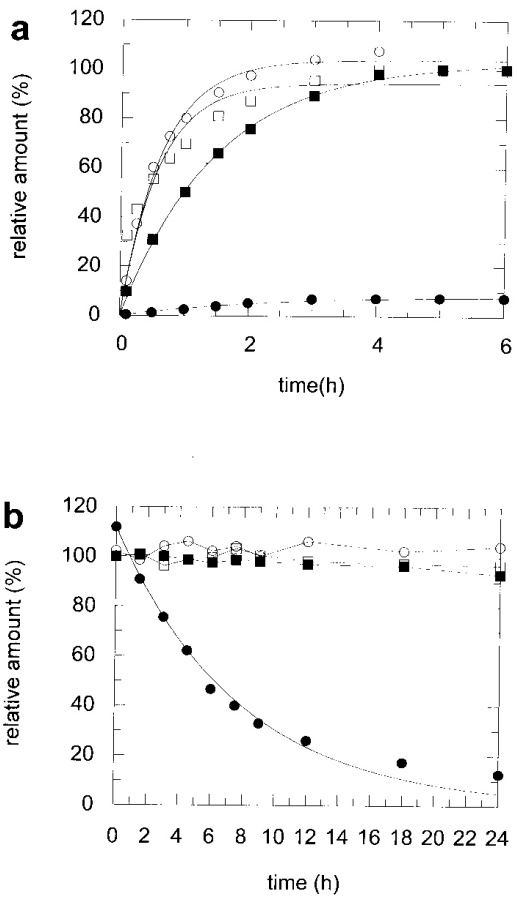

Fig. 5. Refolding and equilibration experiments with the parent protein and the variant in solution. Squares refer to trypsininhibitory activity. Circles refer to chymotrypsin-inhibitory activity. Open symbols refer to the parent protein and filled symbols to the Ser17Pro variant.

the variant in figure 4b differed before and after template-directed refolding in marked contrast to those of the parent protein in figure 4a.

The equilibrium dissociation constant of the minor amount of trypsin-reactive material obtained from the data in figure 4b is 2.3 nM and that of the apparently inactive major portion of trypsin-reactive material is estimated to be near 2μM (results not shown) from experiments at higher concentration of enzyme and inhibitor. This indicates that the trypsin-reactive region of Ser17Pro is apparently distributed into a heterogeneous population of partially active and apparently non-active conformers, greatly differing in their activity. As shown in figure 5a, the appearance of the minor amount of trypsin-inhibitory material with Ser17Pro laggs behind that of the chymotrypsin-inhibitory region. The refolding rate of the second subdomain is also significantly, but not as dramatically affected as a result of the mutation. This points to long-range effects transmitted from the first into the second subdomain and argues against an entirely independent refolding mode of the two subdomains.

As shown under procedure *b* in figure 3, a single and fully active state of Ser17Pro characterised by the correct ratio (1:1) of trypsin- and chymotrypsin reactive material could be obtained only upon refolding in the presence of trypsin-Sepharose as a template with complementary structure followed by affinity chromatography [9]. The Ki-values of trypsin-inhibitory material differs substantially before and after contact with the template as shown by the differing inhibition curves in figure 4b. Although the Ki-value of the second subdomain directed against chymotrypsin is not significantly affected (results not shown) on the template, it is slightly but none the less significantly increased (about twofold) with respect to the parent protein. The variant migrates a single band at approximately 8 kDa in SDS gels under reducing conditions and it is eluted as a single active peak in reversed phase HPLC after affinity chromatography. The chemical identity of the variant is confirmed by electrospray mass spectrometry as a single polypeptide chain of 7879.7 1.3 Da with all disulfide bridges and peptide bonds intact (see figure 4 in reference [11]). This finding is in agreement with the proteolytic stability of Bowman Birk inhibitors upon affinity chromatography on trypsin-Sepharose [24]. The equilibrium dissociation constant of the variant against bovine trypsin is decreased to a Ki of 0.072 0.0023 nM after template directed refolding. As this is a further improvement of activity we may conclude, that the trypsin-reactive region of the variant exists in at least three different forms with Ki-values differing as much as four orders of

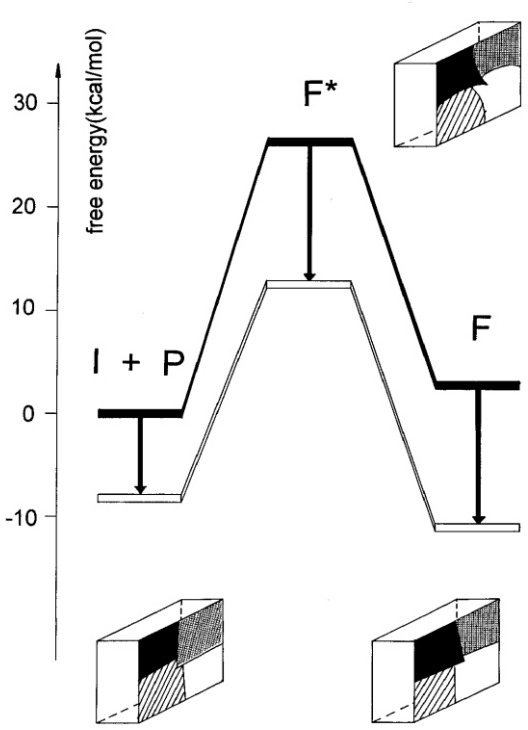

Fig. 6. Reconstructed free energy profiles for the refolding of the Ser17Pro variant in solution (filled line) and on the template (open line). The minimal thermodynamic stabilisation of the partially active states I and P and the fully active state F and the minimal kinetic stabilisation of the transition state F* connecting these states is indicated by arrows. The close structural similarity of these states is indicated by means of the cardbox model of protein folding.

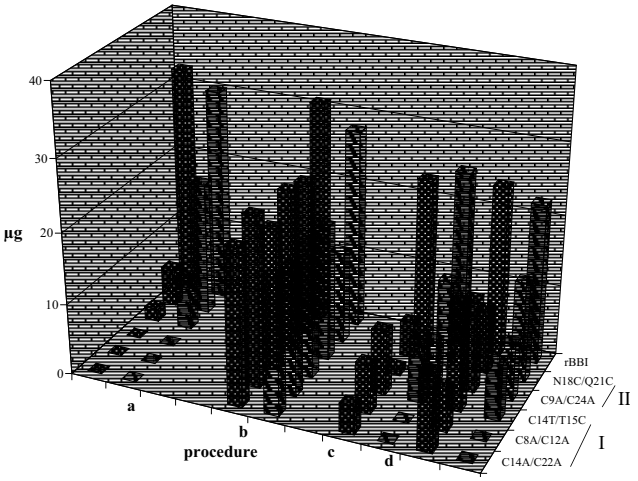

Fig. 7. Refolding and equilibration experiments with disulfide mutants of rBBI. The amounts of trypsin (darker bars) and chymotrypsin-reactive material (lighter bars) refer to 32 mg of crude CNBr cleavage material. The procedures were a: refolding in solution; b: refolding on trypsin-Sepharose; c: equilibration experiments with b in refolding buffer under redox-conditions; d: equilibration under non-redox conditions in the presence of 1,5I-AEDANS used as a SH-trapping reagent.

magnitude. These three forms are designated as the apparently inactive form I (Ki in the µM range), the partially active form P (Ki in the nanomolar range) and F (fully active state). Thus, the properties of the variant differ before and after template directed refolding and subsequent affinity chromatography which is in marked contrast to the strictly Anfinsonian behaviour observed with the parent protein. The increased activity of the Ser17Pro variant is unexpected because a proline in the corresponding position occurs only in nearly inactive forms of the Bowman Birk and other families of protease inhibitors. It appears that this kind of variant has been disfavoured in evolution mainly due to its impaired refolding competence in solution.

The fully active state of Ser17Pro obtained by template-directed refolding represents a local rather than a global energy minimum that unfolds slowly in the refolding buffer even in the absence of denaturant. The non-stoichiometric ratio of the two subdomains (see figure 5b) and the initial Ki-value (see triangles in figure 4b) of the trypsin-inhibitory region are slowly restored after a 24 h incubation of the fully active form in the refolding buffer at 37 C. The partially inactive state obtained after equilibration could be converted back into the fully active state on the template in a fully reversible manner. The molecular mass of 7876.6 (2) Da after equilibration excludes chemical side reactions during the equilibration experiments such as side chain modifications. The kinetic analysis of the data in figure 5b reveals a rate constant of k_u = 3.6 (0.2) $10^{-5}s^{-1}$ for the backward return to the initial mixture of partially active states. The rate constant of the forward reaction in the absence and the presence of the template could only be estimated by means of the logarithmic form of the kinetic equation for a first order conversion of the partially active state (P) into the fully active state (F):

$$\ln(1 - A_t/A_{oo}) = -k_f t$$

From the lack of observing an increase of trypsin-reactive material within a prolonged incubation period of three days (t 2.6 10^5s; A_{oo} = 100; A_t = 5, corresponding to a standard error of 5%) the rate constant for the forward reaction in solution is estimated to be $k_f < 2$ $10^{-7}s^{-1}$. Because on the other hand, the formation of the fully active state is completed within less than 15min (the time needed for packing the column) on the template, a $k_f > 3.33$ $10^{-3}s^{-1}$ may be estimated for the rate constant of template directed refolding (t < 900s; A = 100, A_t = 95). Similar estimates are also obtained from simple thermodynamic considerations in reference [11]. This would correspond to a rate acceleration of at least four orders of magnitude on the template with respect to that in solution. The resulting free energy profiles for the refolding of the variant in solution and on the template are compared in figure 6.

A global minimum of free energy supports a kinetic reaction control of protein folding (at least for the variant) although it cannot exclude thermodynamic control. In fact, the template facilitates folding kinetically, by reducing the

high barrier of activation of the variant in solution and also thermodynamically by stabilising the fully active state by means of protein-protein-interactions as shown in figure 6. The significant activity of the other, partially active forms are in favour of minor, but functionally significant structural differences between the three species that seem to be formed with Ser17Pro. A close structural resemblance is also supported by the proteolytic stability of the variant upon trypsin-Sepharose assisted refolding and affinity chromatography. Minor conformational distortions are also supported by the absence of free thiol groups after refolding in solution.

A plausible working hypothesis that could account to the close structural reminiscence between the partially active states and the fully active state is offered by the cardbox-model of protein folding [25]. The key feature of this model is a close structural reminiscence between the folded and partially folded states. The cardbox model also seems to imply, that the transition state (F* in figure 6) corresponding to the final step of refolding represents a distorted structural analogue of the final state (F in figure 6). In order to increase the rate of folding the template must be distorted from structural complementarity to the final state into perfect structural complementarity to the transition state folding. Conformational differences between free and immobilised trypsin are known [26] and are obviously responsible for the fact, that only the immobilised enzyme is capable to facilitate the refolding of the variants. Minor structural distortions of conformation late in folding may explain related (so far only poorly understood) observations in the literature. For example, small proteinase inhibitors direct the folding of immobilised trypsin [27] in contrast to the Kunitz inhibitor, being without a similar effect. Apparently, the transition state of this refolding reaction is not stabilised by the Kunitz inhibitor because of its more excessive structural complementarity to the ground state of the enzyme [28]. Significant conformational distortions are also known to occur in prosequences, known to facilitate protein-folding reactions [29].

The more radical replacements in the framework of disulfide bonds in figure 1 result in dramatic irregularities greatly differing in their extent [12]. The results of titrations with trypsin and chymotrypsin after refolding in solution, on the template and subsequent equilibration experiments (under redox- and non-redox conditions) are depicted in figure 7.

Minor amounts after refolding in solution, a stoichiometric ratio after template-directed refolding followed by a slow and fully reversible return for the initial state is observed with the Cys9Ala/Cys24Ala and the Asn18Cys/Gln21Cys double replacement. The absolute amounts of the chymotrypsin-inhibitory subdomain and its activity obtained after refolding in solution and on the template and after the equilibration experiments are in close agreement. The behaviour of these variants, designated as class II in figure 7 was in close accord with that of Ser17Pro presented in figure 3.

However, more dramatic perturbations are induced by the Cys14Ala/Cys22Ala, the Cys8Ala/Cys12Ala and the Cys14Thr/Thr15Cys replacements belonging to class I in figure 7. Not only the activity of the trypsin- but also that of the chymotrypsin-inhibitory is abolished in case of these variants after refolding in solution. This indicates that these mutation do not only affect the refolding of the trypsin-inhibitory but also that of the chymotrypsin-inhibitory region. Refolding on trypsin-Sepharose as a template with complementary structure gave the expected stoichiometric 1:1 ratio of both subdomains with the class I and II variants. The substantial amounts of trypsin-inhibitory material retained after overnight incubation of the fully active state of the variants belonging to class I seem to be in line with a non-reversible behaviour. It is notable that a larger amount of trypsin-reactive material is retained under non-reducing condition in the presence of 1,5I-AEDANS used as a sulfhydryl-trapping reagent.

These results indicate, that the two subdomains of BBI cannot be regarded as independent refolding units even though only intradomain disulfide bonds are present in the molecule. The dramatic perturbations induced by the disulfide replacements in figure 1 differ substantially from the conformational stability observed with many globular proteins [30-32], which may be explained by compensatory structural readaptations in their hydrophobic cores [33]. The dramatic effects induced by the disulfide replacements in BBI seems to be more reminiscent of the behaviour of insulin-like growth factor IGF-I [34] and lysozyme [35] that are also devoid of substantial hydrophobic cores. The destabilisation induced by introducing an additional disulfide bond in the Asn18Cys/Gln21Cys double replacement resembles other examples where additional disulfide bonds were engineered into a protein [36-37].

The pronounced clustering of disulfide bonds may explain only local perturbations within the trypsin-inhibitory subdomain. A clustering of disulfide bonds belonging to the two different subdomains is discernible only for Cys8, Cys24 and Cys51. This may perhaps account to the breakdown of chymotrypsin-inhibitory activity in Cys8Ala/Cys12Ala. However, the dramatic effects observed with Cys14Ala/Cys22Ala and Cys14Thr/Thr15Cys would not be expected from the clustering of the cysteine residues in position 8, 24 and 51. The local perturbations induced by the Cys9Ala/Cys24Ala replacement would also not be expected on the basis of disulfide clustering because this variant would be predicted to induce more global perturbations also into the second subdomain directed against chymotrypsin. Therefore, we have to conclude that other, more indirect effects are responsible for the overall perturbations induced by the class I mutations. An inspection of the crystal structure of sBBI in figure 2 reveals, that all disulfide bonds affected by the mutations that belong to class I protrude out from the -strand A (Asp10- Thr15) of the structure. In contrast, the disulfide bonds affected by the mutations that belong to class II protruding out from -strand B (Gln21- Cys24). A primary role of -strand A for structure and folding seems to be also reflected by the protective effect of 1,5I-AEDANS against the relaxation of the trypsin-inhibitory subdomain with the class I but not the class II variants in figure 7d.

What kind of interactions could be responsible for the significant effects transmitted into the chymotrypsin-inhibitory subdomain induced by the replacements within -strand A? The hydrogen bond contacts between the NH

atoms of Asp10 and Gln11 donated to the carboxylate side chain of Asp26, thus connecting the N termini of the two subdomains appear as a less likely possibility, which cannot account to the dramatic effects induced by Cys14Thr/Thr15Cys and also the Cys14Ala/Cys22Ala replacement near the C-terminus of strand A. The five tightly bound internal water molecules in the structure are also rather unlikely because they are involved in hydrogen bond contacts with strands B, C, E and F with His33, but not with the critical strand A. The present data cannot rigorously exclude a possible relevance of loosely bound water molecules that are not observed in the crystal structure. The close electrostatic contacts between the carboxylate side chain of Asp10 at the N-terminus of strand A with Arg28 and His33 in the chymotrypsin-inhibitory region stabilising the relative orientation of the two subdomains [13] appears as a more likely possibility. A structural significance of the polar interactions across the interdomain boundary as shown in figure 2, seems to be also supported by sequence similarities in the Bowman Birk family [38]. Asp10, Arg28 and His33 are strictly conserved in the major groups I and II of the Bowman Birk family whereas they are covariant in the groups III and IV comprising only few members. A structural significance of long-range interactions between the two subdomains such as the proposed ion pairs was also supported by the restored native properties of a reconstituted mixture of two separated subdomain fragments [39]. Therefore, we favour ion pairs as the predominant force, because electrostatic interactions are effective over longer distances than hydrogen bonds. We feel, that long-range interactions in the nascent polypeptide chain would restrict the number of possible conformers more effectively than short distance interactions. Indeed, there is a growing awareness in the literature for a significance of electrostatic interactions [40-45] in protein folding in addition to the hydrophobic effect.

Conclusions

The equilibration experiments with the variants support the notion of the fully active state of a protein being located within local rather than global minimum on its structural free energy landscape. The fully active state of Ser17Pro (see results in figures 3 and 5) and the class II replacements (see results in figure 7) obtained by means of template-assisted renaturation is converted into the initial mixture of conformers as obtained after refolding in solution within a subsequent 24 h incubation period. It is especially notable that the reattainment of the initial state represents a slow process (see results in figure 3b) due to a high barrier of activation. This has been also observed also with the active form of the human plasminogen-activator inhibitor PAI-I that relaxes slowly into its latent form [46]. The significant amounts of trypsin-reactive material observed with the disulfide replacements belonging to class I (especially under non-redox conditions) are also more in line with a slow return to the initial situation that is not completed after overnight incubation. Thus, the template-directed folding observed with the variants arises not only from thermodynamic (stabilisation of the final state on the template) but also from kinetic (stabilisation of the transition state between the partially and the fully active state) factors. This

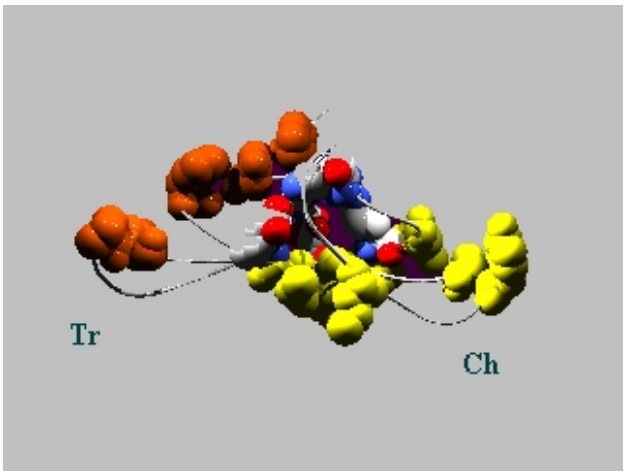

Fig. 8. Exposed hydrophobic patches and buried polar interior of sBBI (see text).

seems to be in line with a parallelism of kinetic and thermodynamic reaction control in protein folding and the notion of foldability as an inherent feature of those protein-sequence motifs that are preferred by Nature [47].

The fully active state of the variants is not literally metastable because their relaxation to the equilibrium state is a slow but still measurable process. None the less, a solely thermodynamic reaction control such as with insulin-like growth factor I (which returns into the initial state rapidly after removal of its binding protein [48]) must be clearly excluded.

The occurrence of local minima has been reported for several other proteins since our original publication [11]. The active state of chymopapain resides within a local minimum that is convertible into the equilibrium state at elevated temperatures [49]. A similar behaviour has also been established with the active state of processed alpha Lytic Protease (aLP) attained in the presence of its prodomain. This fully active form of aLP has been shown to relax into an inactive, more stable state at experimentally feasible rates with increasing concentrations of denaturant [50]. By extrapolating back to the native buffer system the authors arrive at an energy profile being closely reminiscent of the energy profile in figure 6 which has been published by us previously [11]. The active state O6-Methylguanin-Transferase also resides within a relative minimum that is stabilised by the binding of DNA [51]. A medically important example for metastable structures in biology is the conversion of infectious particles into the fusion competent state in the course of viral infections. The native conformation of the HA2 chain of influenza hemagglutinin (HA) resides within a metastable state at pH 8 that is converted into the thermodynamically more stable and fusion competent state at pH 5-6 during endocytosis [52]. This conversion into the more stable non-native state can also be mimicked with denaturants at physiological pH [53].

Does the refolding behaviour of the soybean and the recombinant Bowman Birk inhibitor prove a solely thermodynamic reaction control in protein folding? In the authors opinion more information is needed about other (and perhaps more stable) conformers that may exist on the folding

landscape of sBBI before a definitive answer can be given. The reason for this reservation is that the structure of sBBI differs from the majority of other more globular proteins. The structure of globular proteins exhibit a hydrophobic core and a hydrophobic surface exposed to the solvent. This has also been documented for several other types of protease inhibitors (reviewed in references [22-23]) including the Bowman Birk inhibitor A-II of the peanut forming a small hydrophobic core in the interdomain region in the centre of the molecule [54]. However, a hydrophobic core of this kind is not observed with our crystal ray structure of the Bowman Birk inhibitor of the soybean (sBBI). To the contrary, the structure of sBBI displays several peculiarities that are in contrast to the situation typically observed with most other globular proteins that are highlighted in figure 8.

On the one hand, the exposed hydrophobic patches are highly unusual for a folded protein in its most stable form. The side chains of Met27, Leu29, Ile40, Tyr45, Phe50 and Val52 together with the CH2 groups of Pro46 form a hydrophobic patch on the right-lower side of the molecule highlighted in yellow in figure 8 and the other exposed hydrophobic patch on the upper-left side of the molecule highlighted in brown is formed by the proline residues in positions 7, 19 and 20 and by the two aromatic side chains of Phe57 and Tyr59. The disulfide bridges are also exposed to the solvent, except the more buried disulfide linkages Cys9-Cys24 and Cys36-Cys51.

On the other, the side chains of Asp26, Arg28 and His33 (shown in CPK in figure 8) form an electrically charged cluster of buried amino acids at the interdomain boundary together with the partially buried aspartates in positions 10 and 53 in contrast to the side chains of Arg23, Lys37 and Asp56 projecting out towards the solvent shell. The charged protein interior at the interdomain boundary is shielded from the solvent shell by the aromatic plane of Tyr59. The presence of three buried non-functional charged amino acids in a 71 amino-acid protein is peculiar because only one ion pair per 150 residues is usually placed in the interior of globular proteins [55]. A structural role seems to be played by the carboxylate side chains of Asp26 and Asp53 involved in the hydrogen bond contacts with NH atoms of residues belonging to the other subdomain. The charged interior of BBI resembles the buried polar cluster observed also with fungal lipases [56]. The Coulomb interactions between the buried residues at the interdomain boundary seem to be responsible for the nearly perpendicular orientation of the two subdomains of sBBI. A strong electric field between the two subdomains seems to be also supported by the presence of five tightly bound water molecules embedded into the polar interdomain boundary. The polar hydroxyl function of Tyr59 points into the interdomain cleft that contains five internally bound water molecules (not shown) rather than being exposed into the solvent. Since the side chains of Asp26 and Asp53 are placed favourably for interactions with the internal water molecules 102 and 108 we may assume that the tightly bound water molecules embedded into the polar interdomain boundary provide a solvent shell around the buried and thermodynamically unfavourable [57] luster of charged amino acid residues in the protein interior.

It is especially notable that the buried cluster of oppositely charged residues at the interdomain-boundary is sandwiched between two exposed hydrophobic patches on the opposite sides of the molecule as shown in figure 8. Therefore, it is possible that a hydrophilic protein interior surrounded by a hydrophobic protein exterior represents a local rather than a global energy minimum. The polar interior and the exposed hydrophobic patches are structural peculiarities that are much more reminiscent of the kind of structural features that have been postulated to occur in partially folded [55] rather than in globular proteins. These exposed hydrophobic patches of sBBI are responsible for the reactivity of the second subdomain with both trypsin and chymotrypsin [15]. Such broad specificity may be functionally important for an effective plant defence against digestive proteases of insects. A hydrophobic collapse of the exposed hydrophobic patches into a more regular, stable hydrophobic core seems to be prevented by the rigid array of seven disulfide bridges of Bowman Birk inhibitors. This hypothesis seems to be supported by the observation of considerable changes in the CD spectra of BBI under reducing conditions in the absence of denaturant [58]. More recent unfolding studies with a Bowman Birk inhibitor from the Indian bean Doliochus biflorus [59] have shown that its disulfide bonds are hyperreactive in spite of an extraordinary (meta?)-stability under extreme pH, high temperature and the presence of denaturants. The trypsin and the chymotrypsin-inhibitory activity are lost at an early stage of reduction without accumulation of intermediates as expected for a protein with a polar protein interior devoid of a hydrophobic core.

References

1. K. Kuwata, R. Shastry, H. Cheng, M. Hoshino, C.A. Batt, Y. Goto & H. Roder, *Nat. Struct. Biol.* **8** (2001) 151-155.
2. V.S. Pande, A.Y. Grosberg, T. Tanaka & D.S. Rokhsar (1998), *Curr. Opin. Struct. Biol.* **8** (1998) 68-79.
3. C. Levinthal, *J. Chim. Phys*. **65** (1968) 44-45.
4. T. E. Creighton, *Nature* **356** (1992) 194-195.
5. D. Baker & D.A. Agard, *Biochemistry* **33** (1994) 223-230.
6. J. D. Honeycutt & D. Thirumalai, *Proc.Natl. Acad. Sci. (USA)*, **87** (1990) 3526-3529.
7. M. Ebeling & W. Nadler, *Proc. Natl. Acad. Sci (USA)*, **92** (1995) 8798-8802.
8. A. R. Dinner, & M. Karplus, *Nat. Struct. Biol*., **5** (1998) 236-241.
9. P. Flecker, *FEBS Lett.*, **252** (1989) 153-157.
10. P. Flecker in Modern Methods in Protein- and Nucleic Acid Research pp.173-185. 1990.Springer.
11. P. Flecker, *Eur. J. Biochem.*, **232** (1995) 528-535.
12. S. Philipp, Y.M. Kim, I. Dürr, G. Wenzl, M. Vogt & P. Flecker, *Eur. J. Biochem.*, **251** (1998) 854-862.
13. R.H. Voss, U. Ermler, L.O. Essen, G. Wenzl, Y.M. Kim & P. Flecker, *Eur. J. Biochem.*, **242** (1996) 122-131.
14. I.L. De la Sierra, , L. Quillien, P. Flecker, J. Gueguen, & S. Brunie, *J. Mol. Biol*., **285** (1999) 1195-1207.

15. J. Koepke, U. Ermler, E. Warkentin, G. Wenzl, & P. Flecker, *J. Mol. Biol.*, **298** (2000) 477-491.
16. P. Flecker, *Eur. J. Biochem.*, **166** (1987) 151-156.
17. M. H. Werner, & D.E. Wemmer, *Biochemistry* **31** (1992) 999-1010.
18. D. R. Flower, *FEBS Lett.* **344** (1994) 247-250.
19. P. Flecker, *Materials Structure* **9** (2002) 36-37.
20. J. M. Hogle & I.E. Liener, *Can. J. Biochem.* **51** (1973) 1014-1020.
21. C. B. Anfinsen, *Science* **181** (1973) 223-230.
22. W. Bode & R. Huber, *Eur. J. Biochem.*, **204** (1992) 433-451.
23. R. J. Read & N.M.G. James, in Proteinase inhibitors, pp. 301-336. 1986. Elsevier.
24. T. Kurokawa, S. Hara, S. Norioka, T. Teshima & T. Ikenaka, *J. Biochem.* (Tokyo) **101** (1987) 723-728.
25. D.P. Goldenberg & T.E. Creighton, *Biopolymers* **24** (1985) 167-182.
26. D. Gabel, I.Z. Steinberg & E. Katchalski, *Biochemistry* **25** (1971) 4661-4669.
27. V. V. Mozhaev & K. Martinek, *Eur. J. Biochem.* **115** (1981) 143-147.
28. R. M. Sweet, H. T. Wright, J. Janin, C. H. Chothia & D. M. Blow, *Biochemistry* **13** (1974) 4212-4228.
29. P. Sorensen, J.R. Winther, N.C. Kaarlsholm & F.M. Poulsen, *Biochemistry* **32** (1993) 12160-12166.
30. H.X. Zhu, C.M. Dupureur, X.Y. Zhang & M.D. Tsai, *Biochemistry* **34** (1995) 15307-15317.
31. L.C. Chuang, P.Y. Chen, C.P. Chen, T.H. Huang, K.T. Wang, S.H. Chiou, & S.H. Wu, *Biochem. Biophys. Res. Commun*. **220** (1996) 246-254.
32. Y. Dai & J. G. Tang, *Biochim. Biophys. Acta* **1296** (1996) 63-68.
33. J. Pons, A. Planas & E. Querol, *Protein Eng.* **8** (1995) 939-945.
34. Q. Hua, L. Narhi, W. H. Jia, T. Arakawa, R. Rosenfeld, N. Hawkins, J.A. Miller & M.A. Weiss, *J. Mol. Biol.* **259** (1996) 297-313.
35. M. E. Goldenberg & Y. Guilou, *Protein Sci.* **3** (1994) 883-887.
36. J. Clarke, K. Henrick, & A. R. Fersht, *J. Mol. Biol.* **253** (1995) 493-504.
37. A. P. Hinck, D. M. Truckses & J.L. Markley, *Biochemistry* **35** (1996) 10328-10338.
38. T. Ikenaka & S. Norioka: Proteinase inhibitors pp. 361-374. 1986. Elsevier.
39. G. S. Townshend, D. P. Botes, & L. Visser, *Biochim. Biophys. Acta* **701** (1982) 346-356.
40. Y. Yu, O.D. Monera, R. S. Hodges & P. L. Privalov. *J. Mol. Biol.* **255** (1996) 367-372.
41. M. H. J. Cordes, A. R. Davidson, & R. T. Sauer, *Curr. Opin. Struct. Biol.* **6** (1996) 3-10.
42. M. Oliveberg & A.R. Fersht, *Biochemistry* **35** (1996) 6795-6805.
43. A. C. Tissot, S. Vuilleumier & A. R. Fersht, *Biochemistry* **35** (1996) 6786-6794.
44. H. Inoue, T. Hayashi, X. P. Huang, J. F. Lu, S. B. P. Athauda, K.H. Kong, H. Yamagata, S. Udaka & K. Takahashi, *Eur. J. Biochem.* **237** (1996) 719-725.
45. S. Tiwari-Woodruff, C. T. Schulheis, A. F. Mock & D. M. Papazian, *Biophys. J.* **72** (1997) 1489-1500.
46. A. E. Franke, D. E. Danley, F. S. Kaczmarek, S. J. Hawrylik, R. D. Gerard, S. E. Lee & K. F. Geoghegan, *Biochim. Biophys. Acta* **1037** (1990) 16-23.
47. E. I. Shakhnovich, *Phys. Rev. Lett.* **72** (1994) 3907-3910.
48. S. Hober, A. Hansson, M. Uhlén & B. Nilsson, *Biochemistry* **33** (1994) 6758-6761.
49. S. Solis-Mendiola, L.H. Gueriérrez-Gonzáles, A. Arryo-Reyna, J. Padila-Zuniga, A. Rojo- Domínguez & H. Hernandez-Arana, Biochim. *Biophys. Acta* **1388** (1998) 363-372.
50. I. L. Sohl, S. S. Jaswal, D. A. Agard, *Nature* **395** (1998) 462-464.
51. D. Bhattacharya, T. K. Hazra, D. Behnke, P. L. G. Chong, A. Kurosky, J. C. Lee, & S. Mitra, *Biochemistry* **37** (1998) 37, 1722-1730.
52. J. Chen, S. A. Wharton, W. Weissendorn, L. J. Calder, F. M. Hughson, J. J. Skebel., & D. C. Wiley, *Proc. Natl. Acad. Sci. (USA)* **92** (1995) 12205-12209.
53. C. M. Carr, C. Chaudhry & P. S. Kim, *Proc. Natl. Acad. Sci (USA)* **94** (1997) 14306-14313.
54. A. Suzuki, T. Yamane,T. Ashida, S. Norioka, S. Hara & T. Ikenazka, *J. Mol. Biol.* **234** (1993) 722-734.
55. K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, & H. S. Chan, *Protein Science* **4** (1995), 561-602.
56 U. Derewenda, L. Swenson, R. Green, Y. Wei, G. G. Dodson, S. Yamauchi, M. J. Haas, & Z. S. Derewenda, *Nat. Struct. Biol*. **1** (1994) 36-47.
57. C. D. Waldburger, J. F. Schiedbach, & R.T. Sauer, *Nat. Struct. Biol.* **2** (1995) 122-128.
58. Y. V. Wu & D. J. Sessa, *J. Agric. Food Chem*. **42** (1994) 2136-2138.
59. R. R. Singh & AGA Rao. *Biochim. Biophys. Acta* **1597** (2002) 280-291.