Chapter 6

Development of a β-sheet propensity scale via Backbone Thioester Exchange (BTE)

6.1 Introduction

6.1.1 Study of the protein secondary structure: the β-sheet

The research of this chapter was performed at the University of Wisconsin-Madison (USA), during a six-months leave in the group of Prof. Samuel H. Gellman.

Although it is well-documented that a protein's folded three-dimensional structure is encoded by its amino acid sequence, currently the folded structure cannot be predicted from sequence information alone,[1] so the studies of protein stability, protein secondary structure and *de novo* protein design are intimately interconnected. Stability studies provide insight for the design of proteins that will fold into predetermined structures and perform specified functions, while protein design provides an opportunity to test the understanding of the rules that underlie protein structure and stability.

Three types of regular secondary structures are found in proteins: helix, sheet and turn.[2] The frequency with which these few secondary structural patterns are found in folded proteins has led many researcher to focus on model systems in which a well-defined secondary structure occurs in the absence of a tertiary structural context, to provide a way to separate intrinsic sources of secondary structural stability from tertiary context effects. Among the regular secondary structures, sheets have received the least attention because, until recently, there were no well-behaved model systems. Within the past 20 years discrete β-sheet model systems have become available. These systems are beginning to provide insight into the factors that influence β-sheet stability and they promise to give important information in the next few years.

β-Hairpin is a common supersecondary structural element in which two adjacent antiparallel strands are linked by a short loop. Understanding β-sheet formation is the key to a host of problems and applications involving protein folding and design. For example, the formation of a **β-hairpin** has a profound effect on reducing the conformational space and defining the long-range interactions for a folding protein. Recent studies have emphasized that there are many proteins in which β-sheets play functionally important roles, particularly in protein-DNA,[3] protein-RNA,[4] protein-protein recognition[5] and aggregated protein fibrils exhibiting predominantly β-structure have been implicated in amyloid diseases.[6] The loop size can vary, usually from two to six residues,[7] where the two loop residues are the central two residues of a β-turn. In 1993 the first short peptide was reported[8] which folds to a **β-hairpin** conformation in aqueous solution. Some other initial studies showed that β-sheet could form

in aqueous solution at concentrations that do not allow aggregation. The monomeric nature of these **β-hairpins** was important because of prior demonstrations that short β-strand-forming peptides can readily self-associate.

Apparently, proper covalent linkage of strand-forming segments within **β-hairpin** peptides allows the strands to engage in intramolecular contacts, at concentration sufficient for spectroscopic analysis. These initial studies also showed that extracting **β-hairpin** sequences from folded proteins does not reliably produce free-standing **β-hairpins**, therefore tertiary **context** is often crucial for **β-hairpin** stability. Use of autonomously folding **β-hairpins** to explore β-sheet stability requires the ability to design **β-hairpins** of a specific shape.

6.1.2 Designing β-hairpin peptides and turns

The design of β-sheet proteins has proven more difficult than that of α-helical proteins.[9] The β-sheet is composed of many structurally distinct regions with different conformational requirements, all of which must be incorporated into a single design. To form hydrogen-bonded β-sheets, β-strands necessarily interact with one another. Design not only must include favourable interactions between several β-strands, but also must disfavour unplanned, competing interactions with other β-strands. Despite these diverse considerations, designs for β-sheet proteins and **β-hairpin** have met with some success. β-Sheet structure (**Fig. 6.1**) is very complex: the different hydrogen bonding patterns of antiparallel and parallel sheets produce definite structural differences within the sheet itself. β-Sheets may twist, curl, and even fold back on themselves to varying degrees. A distinct feature of a β-sheet is the pattern of hydrogen bonds formed between the amide and carbonyl groups of the protein backbone. In parallel β-sheets, where the β-strands run in the same amide-to-carbonyl direction, the backbone hydrogen bonds are evenly spaced and angle crossed to the adjacent main chain. In contrast, the hydrogen bonds formed in antiparallel sheets are approximately perpendicular to the main chain.

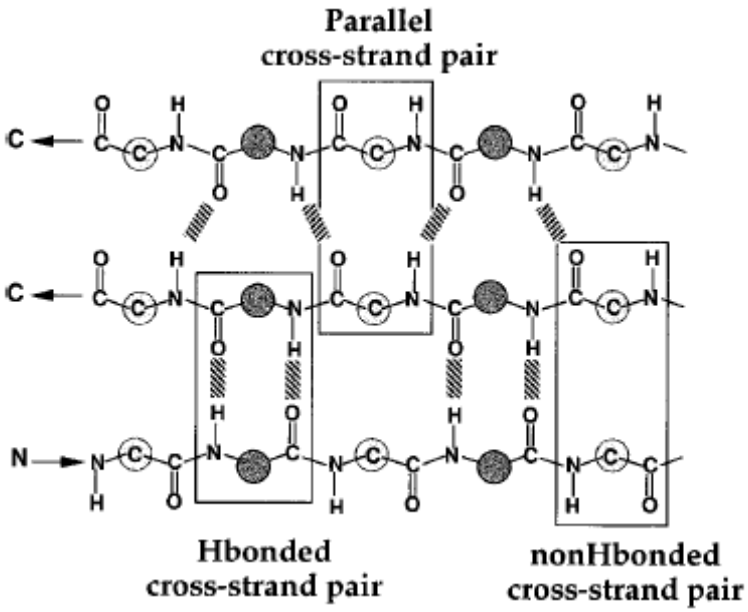


Fig. 6.1. Hydrogen-bonding pattern for parallel and antiparallel β-strands and illustration of cross-strand side-chain pairs. Hydrogen bonds are represented by hatched blocks and side chains by grey filled circles. Arrows show the amide (N) to carbonyl (C) direction of the strand.[1]

Because each β-strand necessarily interact with others to form these hydrogen bonds, the β-sheets can bring close together amino acids which are very distant in sequence. β-Sheet side chains alternate above and below the plane of the sheet along each strand. A parallel β-sheet arrangement produces an asymmetric cross-strand pair of amino-acids which do not share hydrogen bonds with each other. In antiparallel β-sheets, there are two types of symmetrical cross-strand pairs: a narrow hydrogen bond pair and a wide pair[10] also called H-bonded and non-H-bonded pairs, hydrogen bond either to solvent or to another adjacent strand.

To allow detailed thermodynamic and structural analyses of a **β-hairpin** peptide, it is critical that it exhibits a significant degree of secondary structure, a discrete oligomeric state, and high solubility in water solution. One of the first successful designs was the conversion of a homodimeric protein, λ-Cro, to a folded and functional monomeric structure.[11] The intermolecular dimer interface, formed by two antiparallel β-strands, was replaced by an intramolecular **β-hairpin** in which the β-turn sequence was optimized by selection. The protein containing the best turn (Asp-Gly) behaved as the predicted monomer in solution, and its crystal structure confirmed that the **hairpin** design was successful.[12] Recently, a certain degree of native-like **β-hairpin** structure was observed in aqueous solution for short, monomeric, linear peptides, derived from natural proteins.[13]

The relationship between β-turn conformation and right-handed twist of strands in β-sheets has turned out to be critical in developing small β-sheet models.[2] One strategy for creating **β-hairpins** with tight two-residue loops has been to use at least one D residue at loop position.

The rational design of the turn regions of anti-parallel β-sheets has greatly benefitted from the statistical analysis of protein structures in the protein data bank.[14] The early systematic classification of β-turns reveals a wide variety of geometries and sizes of loop.[7,15] For the design of **β-hairpins**, the emphasis has been on incorporation of the smallest turn sequence possible to limit the entropic destabilisation effects.

The classic definition of a tight turn[16] is in terms of a four-residue corner with a hydrogen bond between the CO of the first residue and the NH of the fourth residue. There are at least five distinct combinations of ϕ and ψ torsion angles for the two central residues that can allow formation of this H-bond. The only turn type, called type I, that has both ϕ and ψ pairs with freely permissible values is the most common type of turn, since any residues are permitted, with residue 2 α-helical and residue 3 in 3_{10} conformation. The second most common type is the "glycine" turn, called type II, with residue 2 in poly-Pro conformation and residue 3 in left-handed 3_{10} conformation. Their mirror images are called type I' and type II', respectively.[17]

The backbone conformation of type I and type II turns, defined by ϕ and ψ torsion angles, results in a local left-handed twist, which is not compatible with the right-handed twist found in protein β-sheets. It was shown[18] the importance of ϕ and ψ angle preferences for the residues in the turn sequence by comparing the stability of a number of **β-hairpin** peptides derived from ubiquitin sequence. Replacing in the turn the ^{L}Pro with ^{D}Pro, the twist is switched from left-handed (type I or II) to right-handed (type I' or II') making the latter compatible with the **hairpin** conformation. Statistical analyses from a number of groups have identified X-Gly as a favoured type I' turn. A number of studies have used the NG sequence to design **β-hairpin** motifs that showed a high population of the folded structure with the required turn conformation and strand alignment.[19] Using NMR and CD measurements it was concluded that X= Asn > Asp > Gly > Ala > Ser in promoting **hairpin** folding, in agreement with the intrinsic ϕ and ψ preferences of these residues.

6.1.3 Length, aggregation and cooperative folding

Understanding the forces that control conformational preferences within the common secondary structures[20] should contribute to interpret the conformational preferences at tertiary

and quaternary levels. α-Helices become more stable as the chain length increases,[21] because helix initiation is thermodynamically unfavourable but helix propagation is favourable, at least for some residues.[22] Analogous length-dependent stabilization is observed for double-helical nucleic acid conformations[23] and for unnatural oligomers that adopt helical secondary structures.[24] Length-dependent stabilization is a more complex issue for sheet secondary structure: two dimensions must be considered for a sheet,[25] along the strand direction and perpendicular to the strand direction, while only one dimension is important for a helix,[22,26] along the helix axis. Each dimension can be evaluated independently to design short peptides that display β-sheet conformations containing a predetermined number of strands with defined lengths. There is an intrinsic limit on strand length in antiparallel β-sheets, at least for some sequences.[20]

The earlier difficulties with peptide aggregation and low solubility appear to have been partially overcome.[14] Examining successfully designed sequences, one strategy seems to be the insertion of solubilising charged residues that lead to mutual repulsion in the aggregated state, even if these problems appear to run hand-in-hand with stability of the folded β-sheet, resulting in problems with intermolecular interactions at NMR concentrations.

While the number of **β-hairpin** model systems is expanding rapidly, the number that has been amenable so far to quantitative analysis and thermodynamic characterisation remains small.

The problems are several, related to the nature of the "folded" state, not entirely clear in terms of the dominant stabilising interactions. The fact that limiting[14] parameters for the fully folded state are subject to some uncertainty, making quantitative analysis of folded populations problematic, and finally because folding models are based on the assumption of a two-state cooperative process, which has not generally been clearly justified an may not be appropriate.

So this field is lacking the rigorous quantitative approaches to measuring β-sheet stability that are now well-established in the field of α-helical model systems. Even **β-hairpin** contains several levels of complexity requiring dissection of both the β-turn and the β-strand contributions to the stability of the folded conformation and an understanding of the nature of the stabilising interactions and the influence of intrinsic propensities to secondary structure.

The physical techniques, particularly NMR, are being used to probe folding in solution at the individual residue level, but the whole area of β-sheet peptide design would significantly benefit from a more quantitative approach to understand the nature of the stabilising weak interactions involved and the thermodynamics and kinetics of β-sheet formation.

6.1.4 Side chain interactions in the stabilisation

The origin of the specificity of **β-hairpin** folding has been discussed above and relates largely to the conformational preferences of the turn sequence. However, the origin of the stability of the folded state has been attributed to interstrand hydrogen bonding and/or hydrophobic interactions.

While hydrogen bonds in aqueous solution have long been associated with very small free energy contributions, on the basis that a peptide-peptide hydrogen bond is isoenergetic with a peptide-water hydrogen bond, hydrogen bonds in the solvent-excluded core of proteins are thought to contribute significantly to stability, with a substantial cost associated with burying an unsatisfied hydrogen bond donor or acceptor.[27] More recent examination of the effects of interstrand salt bridges between C- and N- terminal residues of a **β-hairpin** has established a small but measurable effect, despite the solvent exposed nature of the interacting residues.[28] Other groups have highlighted pH-dependent effects on **hairpin** stability.[29] β-Strands necessarily interact to form a β-sheet, and it seemed likely that side-chains interactions could make a substantial contribution to its stability. Statistical survey reveals a non-random pairwise distributions of amino acids in cross-strand positions in antiparallel β-sheets.[30] The specific pairing of amino acids depends on their positions in H-bonded and non H-bonded sites.[30a] Furthermore, theoretical analyses suggest that specific interactions between side chains play an important role in determining β-sheet stability[31] and such pairwise interactions have been included in protein structure prediction.[32] However, β-sheets are often amphipathic, with one face solvent exposed and the other contributing to the hydrophobicity. The statistical distribution is consistent with this possibility and as a result may not reflect any particular interaction between cross-strand side chains. So it was set out to measure the energetic contributions of side-chain interactions to β-sheet stability using the B1 domain,[33] measuring interactions between residues which occupied a H-bonded site in an antiparallel β-sheet environment.

The most useful information for the design and engineering of stable and soluble protein can be gained from studying the interaction between pairs of amino acid with high intrinsic β-sheet forming propensities and pairs of complementary charge. Such pairs were substituted into a solvent-exposed “double” guest site and the thermal stabilities of these proteins were measured. The side-chains interaction energy was calculated as the difference in free energy beyond that which the simple additive sum of the propensities would predict. For example, side chains interact favourably if the stability of the double mutant is greater than that

predicted by the sum of the intrinsic propensities. Stabilizing and destabilizing interaction energies within a range of nearly 2 kcal/mol were found, which are comparable to that of the measured propensities. The pairs found to be most interactive experimentally, Phe-Phe, Phe-Tyr, Glu-Arg, and Glu-Lys are indeed the pairs found statistically together most often. Conversely, the least interactive pairs experimentally (Thr-Val and Thr-Trp) are found together with low frequency. It was interesting to observe that absolute effects on stability and side-chain interaction energy are not necessarily correlated. The intrinsic β-sheet-forming propensities and side-chain interaction energies work in conjunction to make a stable protein. Long-range interactions may dominate if local stabilisation is low.

Overall protein stability is a balance between local and long-range interactions, and for protein design it is important to optimize both β-sheet-forming propensities and cross-strand side-chain interactions. In attempts to deconvolute the factors that contribute to a particular residue in a particular context adopting a particular conformation, intrinsic backbone ϕ, ψ preferences have been examined using also other statistically based analyses of high resolution structures in the protein data bank. A novel approach[34] is to determine ϕ, ψ propensities of different residues in coil regions of protein structures, so residues not in regular secondary structure (β-sheet or α-helix). The advantage of this approach is that intrinsic conformational preferences can be identified free of the interactions associated with regular secondary structure.

In this context ϕ and ψ angles are far from randomly distributed, but most occupy regions of Ramachandran space associated with regular secondary structure.[14]

The observed ϕ, ψ distributions for the individual residues have been taken as representative of those found in denatured state of protein providing the base of “random coil” state from which residue-specific NMR parameters can be derived as a reference state for folding studies.[35] While β-**propensity** is found to vary significantly from one residue to the next, **context dependent** effects appear to play an important part.[36]

6.1.5 β-Sheet construction and forming propensities

Statistical studies of proteins of known structure reveal that β-branched (Ile, Val, Thr) and aromatic amino acids (Trp, Tyr, Phe) occur most frequently in β-sheets.[36] The implication is that the β-sheet-forming propensities of the amino acids are important for β-sheet stability. Although early studies of β-sheet formation suggested that there are qualitative differences in

the β-sheet forming propensities of selected amino acids,[37] there was a clear need for detailed thermodynamic measurements in water-soluble, unaggregated model systems to extend these initial results. Experimental measurements of intrinsic conformational propensities of the amino acids are based on a "host-guest" method first described for α-helical propensities.[22b,38] In host-guest studies, the stability of a standard protein or peptide is compared with those of mutants in which the other 19 amino acids are individually substituted into the guest site.

The results of the experimentally measured α-helical propensities show correlations with each other and with statistical preferences.[22b] The first comprehensive experimental measurement of the β-sheet-forming propensities employed a guest site located in a solvent-exposed, non-H-bonded position on the antiparallel β-strand of a consensus zinc-finger peptide (**Fig. 6.2**).[39]

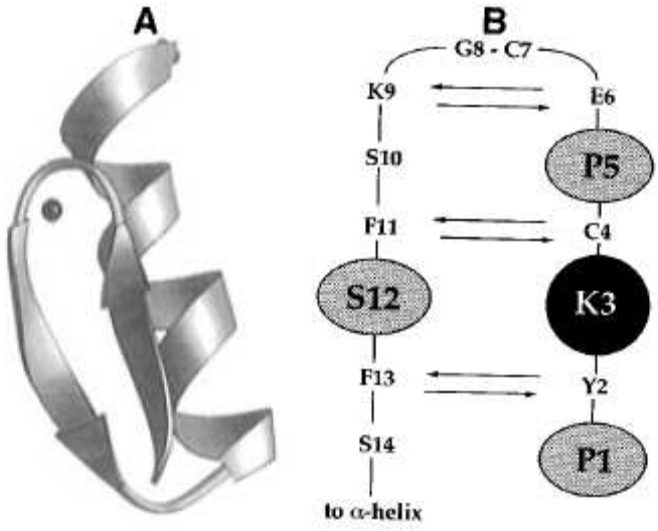


Fig. 6.2. (A) Ribbon diagram[40] of the consensus zinc-finger peptide. Bound zinc is illustrated by the black sphere. (B) Hydrogen-bonding diagram of the β-sheet region. Hydrogen bonds are indicated by arrows. The guest site is highlighted in black. The nearest neighbours to the guest residue on the same face of the β-sheet are highlighted in grey.[1]

Two subsequent host-guest studies used different variants of the B1 domain of streptococcal protein G (the B1 domain) as the model system.[41] The guest sites in both B1 domain studies were positioned along a central strand in a solvent-exposed, H-bonded, antiparallel β-sheet.

The three experimental studies showed, all together, that there are measurable differences between the β-sheet forming propensities of different amino acids. β-Branched and aromatic amino acids tend to be the best β-sheet forming residues, while glycine and proline tend to be the poorest one. There was not an exact correspondence between the experimental studies, in which host specific interactions can modulate the specific ranking, and the statistical data, which average over different environments, and the various types of β-strands. However, the

results of these studies show a strong overall correlation with statistical and theoretical analyses[27b,36,42] and indicate that the amino acids have different intrinsic propensities to adopt β-sheet conformation.

The energetic range between the best and poorest β-sheet-forming amino acids measured in the zinc-finger study (0.5 kcal/mol) is less than that of the B1 domain system (2.8 kcal/mol). This difference may results from the different locations of the two guest sites, in a non-H-bonded edge and a H-bonded central strand, respectively. Both, the edge location, in which there are fewer interactions with the β-sheet than in a internal strand, and the lack of β-sheet hydrogen bonds in a non-H-bonded site can increase the conformational freedom of the guest site. These factors could attenuate the magnitude of the differences in the β-sheet-forming propensities at this edge-strand position.

Edge-strand propensities have also been measured in the B1 domain on a solvent-exposed H-bonded position on an antiparallel β-sheet.[43] The β-branched and aromatic residues were still among the better β-sheet-forming amino acids, and glycine and proline were among the poorest. Greater deviations from the Chou-Fasman statistical preferences[36] were observed in the edge-strand B1 domain study, than in the edge-strand zinc-finger study. Perhaps related to the H-bonded position of the guest site, the range in the B1 domain edge-strand study was not significantly attenuated relative to the central-strand study. These results demonstrate that positioning the guest site in an edge *vs* central strand can modulate the rankings of the measured β-sheet propensities.

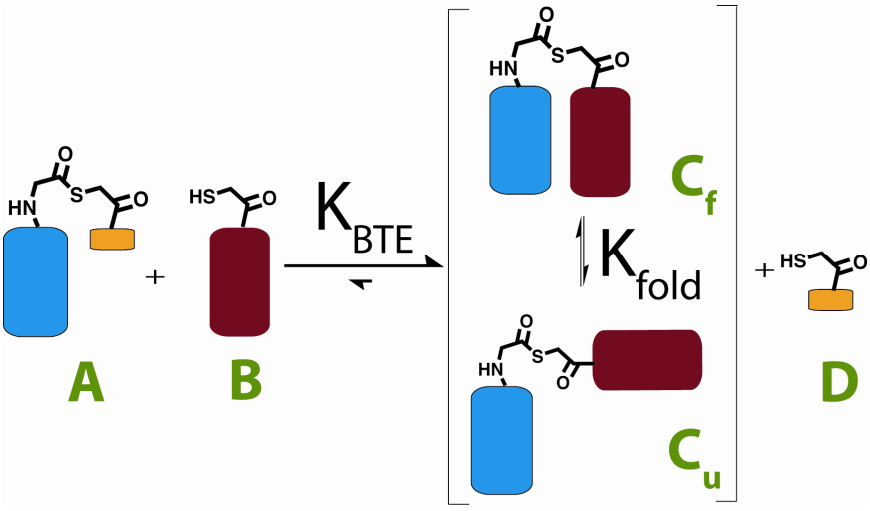
Statistical and theoretical studies also reveal differences between the ranking of the residues preferred in a β-sheet when edge and central strands are considered separately.[30a,44] In conclusion all these previous studies resulted **context dependent**.

6.1.6 Backbone Thioester Exchange (BTE) method

The contribution of a specific residue to the folded stability of a protein is usually studied via mutagenesis, whereby the side chain of interest is mutated to another one that is expected to diminish the folded stability. By comparing the folded stability of the mutant protein to the wild type protein (i.e. $\Delta\Delta G_{fold} = \Delta G_{fold/\ Wild\ Type} - \Delta G_{fold}/_{Mutant}$, where ΔG_{fold} = free energy of folding), one can gain an estimate for the contribution of the wild type side chain to the stability of the protein. Measurement of ΔG_{fold} is possible via a variety of techniques, each with its own advantages and disadvantages. The methods of protein denaturation,[45] hydrogen-deuterium exchange[46] and dynamic combinatorial chemistry (DCC)[47] are the most common.

Backbone thioester exchange (BTE) is a dynamic combinatorial chemistry method developed in Samuel H. Gellman's group for the determination of conformational stability in small folded peptides and proteins.[48] This approach is expected to provide information complementary to those available from thermal and chemical denaturation studies because all measurements are made under native conditions. This strategy should permit the separation of the contributions of autonomously formed secondary structure and those of higher order structure to net stability. BTE is implemented by replacing a backbone amide with a thioester linkage bond in a protein of interest. Secondary amide and thioester groups have comparable conformational properties,[49] but they differ in their hydrogen bonding abilities. This modification allows the equilibration of the thioester with thiols via a transthioesterification reaction, which achieve equilibrium rapidly in aqueous solution at neutral pH and at ambient temperature (native conditions for the folded protein). Transthioesterification occurs via attack on the carbonyl by a thiolate ion (depending on its pKa) to form a tetrahedral intermediate, followed by expulsion of a thiolate or a thiol to reform the thioester bond.[47b]

Knowledge of K_{fold}, the equilibrium constant for the exchange between these two states, allows the determination of the folded stability of the peptide, through the relationship $\Delta G_{fold} = -RTlnK_{fold}$. K_{fold} cannot be determined directly, but can be indirectly assessed via BTE. Transthioesterification (**Scheme 6.1**) is initiated by the addition of thioester (A) to thiol (B) to form a full length thioester (a thiodepsipeptide, analogue to the native amide peptide of interest) (C) and a small thiol molecule (D), methylthioglycolate for this project. After a proper time the equilibrium is reached and it has a constant K_{BTE}.



Scheme 6.1. Cartoon depicting a general backbone thioester exchange reaction (from Erik B. Hadley's thesis).

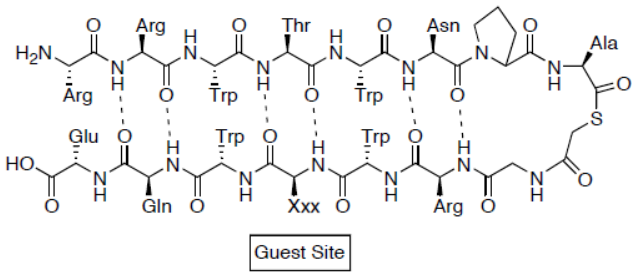
K_{BTE} can be measured via the relative HPLC quantification of A, B, C equilibrating species, which have cromophores. Concentration of C is assumed to be equal to concentration of D, starting from the same concentration of A and B, quantified by HPLC. K_{BTE} could be calculated from a simple equation from which it results that $K_{fold} = K_{BTE} - 1$, because conservative assumptions can be made (see Supporting Information for algebraic expressions in ref. 46).

6.2 Results and Discussion

6.2.1 Development of a β-sheet propensity scale via BTE

The propensities reported to date, based on "host-guest" studies in tertiary folding units that contain a surface-exposed β-sheet;[2,39a,41,43,50] appear to be all **context dependent** and could therefore not yield to a general scale.[43] Because of their intrinsic simplicity, **β-hairpins** can potentially be used to remove the **contextual effects** present in more complicated systems.

Previously, in Samuel H. Gellman's laboratory it was shown that the stability of a **β-hairpin** can be studied via BTE.[51] Using this methodology, the goal of the project described in this chapter is the design of an antiparallel **β-hairpin** in an effort to determine an experimental β-sheet **propensity** scale for α-amino acid residues, based on the stability of a **β-hairpin** conformation in aqueous solution. Before my arrival in Madison this model system was designed, inspired by a **hairpin** with a tryptophan zipper motif, a structural pattern which greatly stabilizes this secondary structure (**Fig. 6.3**).[52] This work was carried out by Erik B. Hadley and Dr. Felix Freire-Iribarne.



I

Fig. 6.3. Atomic detail of backbone structure of **I**. Hydrogen bonds are indicated by dotted lines. The guest site is labelled as Xxx (from Erik B. Hadley's thesis).

Initially hydrogen bonded position 13 was chosen as "guest site", that would be changed to different residues to determine their effect on **hairpin** stability. For each mutant a great variation in the measured ΔG_{BTE} was observed when the experiments were repeated. The cause of this variability remains unknown, but it was concluded that a site between two Trp residues in a tryptophan zipper motif may have specific conformational requirements that render it unsuitable as a guest site.

To circumvent this problem of irreproducibility, peptide **I** was modified (**Fig. 6.4**) to include an inserted segment derived from a **β-hairpin** previously studied in Samuel Gellman's group (Xxx = Ile, the red residue shown in **Fig. 6.4**[53]). Initial efforts involved the placement of the insert either between the loop and the TrpZip motif (**II**) or between the TrpZip motif and the strand termini (**III**) as indicated by the green boxes. In most cases, reproducible thermodynamic stabilities measurements could be made via BTE. However, repeated measurement of the Xxx= Ala mutant indicated that this particular substitution resulted in larger errors in the BTE stability measurements.

Fig. 6.4. Sequence of extended **hairpin** peptides **II-III**. Inserted sequence is surrounded by the green box and the amino acid designated for substitution is shown in red. Hydrogen bonded positions are indicated by a dotted line. Hydrogen bonded positions of the insert vary between the two structures (from Erik B. Hadley's thesis).

Unfortunately, the peptides were observed to precipitate out upon preparation of 50 μM samples for CD analysis in 50 mM phosphate buffer pH 7; the resultant CD spectra were therefore uninterpretable.

Modifications to either the parent peptide sequence or the conditions of analysis were needed, to improve the solubility. Two approaches were taken, changing the buffer or, alternatively, chemically modifying the peptide to discourage aggregation.

The modification of the buffer involved the identification of appropriate components that would allow thiol-thioester exchange and were available in deuterated form for NMR analysis. Buffers were chosen from Good's list.[54] HEPES (50 mM, pH 7) and MES (50 mM,

pH 7) were found to solubilize peptide **III** and were therefore tested for their ability to allow thioester exchange. In each case, the BTE experiment yielded reproducible results, however, the determined ΔG_{fold} varied based on the chosen buffer indicating that the folded stability is solvent dependent. To discourage aggregation the chemical modification of the **β-hairpin** model system was undertaken. Peptides **IV-VIII** (**Fig. 6.5**) were designed to vary the sequence of the insert and/or the position of the inserted region. Peptides **IX-XIV** (**Fig. 6.5**) were designed to incorporate charged residues at the turn region, which is primarily hydrophobic and could potentially promote aggregation. Unfortunately, each of the all amide peptides was found to precipitate from solution when dissolved in 50 mM phosphate buffer pH 7. This result suggests that the aggregation issue in phosphate buffer is related to the extended length of the peptide and not to the identity of the insert sequence or to the hydrophobicity of the turn region. Subsequent solubility studies at 70 μM in HEPES and MES indicate that most of these designed peptides are soluble under these conditions.

H_2N-RRRYTVWTWNPAGGRWTWITVSQE-CO_2H	**IV**
H_2N-RRYVEVWTWNPAGGRWTWRRILQE-CO_2H	**V**
H_2N-RRRLQYWTWNPAGGRWTWVNITQE-CO_2H	**VI**
H_2N-RRYVERWTWNPAGGRWTWQRILQE-CO_2H	**VII**
H_2N-RRRYTRWTWNPAGGRWTWQTVSIE-CO_2H	**VIII**
H_2N-RRYVERWTWRPDGGRWTWQRILQE-CO_2H	**IX**
H_2N-RRYVERWTWDPRGGRWTWQRILQE-CO_2H	**X**
H_2N-RRYVERWTWRPRGGRWTWQRILQE-CO_2H	**XI**
H_2N-RRYVERWTWRPAGGRWTWQRILQE-CO_2H	**XII**
H_2N-RRYVERWTWDPAGGRWTWQRILQE-CO_2H	**XIII**
H_2N-RRYVERWTWNPRGGRWTWQRILQE-CO_2H	**XIV**

Fig. 6.5. Sequences of extended **hairpin** designs with modifications intended to increase solubility in phosphate buffer. Insert sequences are highlighted in green and other modified residues are highlighted in red. The inserted sequence for peptides **IV** and **VIII**, and **IX-XIV** were adapted from a **hairpin** designed by Stanger,[53] the inserted sequence for peptides **V** and **VII** were adapted from a **hairpin** developed by Searle,[55] and the inserted sequence for peptide **VI** was de novo designed using data reported for the frequency of amino acids in β-sheet proteins in the PDB.[56]

6.2.2 Synthesis of the BTE system components

From the preliminary studies performed before my arrival at the University of Wisconsin-Madison and described in the previous section the last peptide of the series **XIV**, from now on **83**, had the most promising properties for the developing a β-sheet **propensity** scale based on BTE.

Here are the sequences (**Fig. 6.6**) of the full-amide peptide **83**, its analogue full length thioester **84**, of the thioester **85** and the thiols **86-100**. They were synthesized, purified, analyzed and studied via BTE.

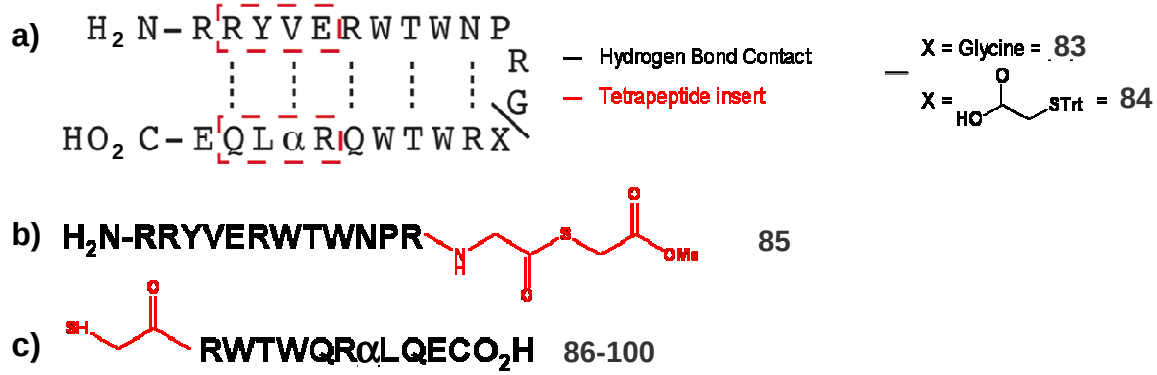


Fig. 6.6. Primary sequences of a) the full-amide peptide **83** and its full length thioester (thiodepsipeptide) analogue **84**, b) the thioester **85**, c) the thiols **86-100** (α is the substitution position), studied in the experiments.

Full-amide peptide **83** (**Fig. 6.7**) was synthesized manually on a Wang resin, using standard Fmoc solid phase peptide synthesis, and used for the NMR and Sedimentation Equilibrium Analytical Ultaracentrifugation studies.

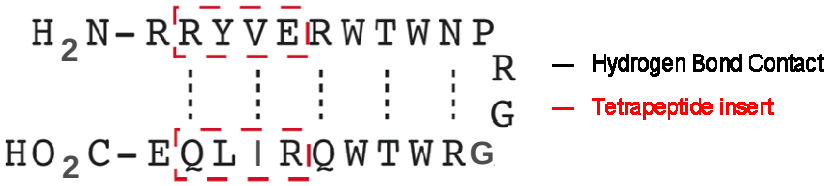


Fig. 6.7. Structure of the full-amide peptide **83**.

6.2.2.1 Sedimentation Equilibrium Analytical Ultracentrifugation of full-amide peptide

This part of the work was performed in collaboration with Jay Steinkruger.

The absence of intermolecular aggregation is crucial for the successful interpretation of data from the studied **β-hairpin** model system. The derived thermodynamic model assumes a monomeric species. Sedimentation Equilibrium Analytical Ultracentrifugation (**Fig. 6.8**) was carried out to determine the aggregation state of full-amide peptide **83** (**Fig. 6.6**) under the conditions used for the BTE experiments (full length thioester **84** (**Fig. 6.6**) could not be analyzed due to the significant thioester hydrolysis which occurs over the experimental timeframe). Samples were spun at a various rotor speeds (between 24 and 48 krpm) using peptide concentrations of 100 and 50 μM. The 97 resultant data fit well to a monomer model, indicating that no significant intermolecular association is occurring under the conditions of the BTE experiment. The full-amide peptide **83** was also analyzed at the concentration of 150

μM, where a loss of material was seen, observed as a decrease of optical density across the entire cell, which indicates aggregation. So, for the BTE experiments, it is important to implement the assay at or below 100 μM.

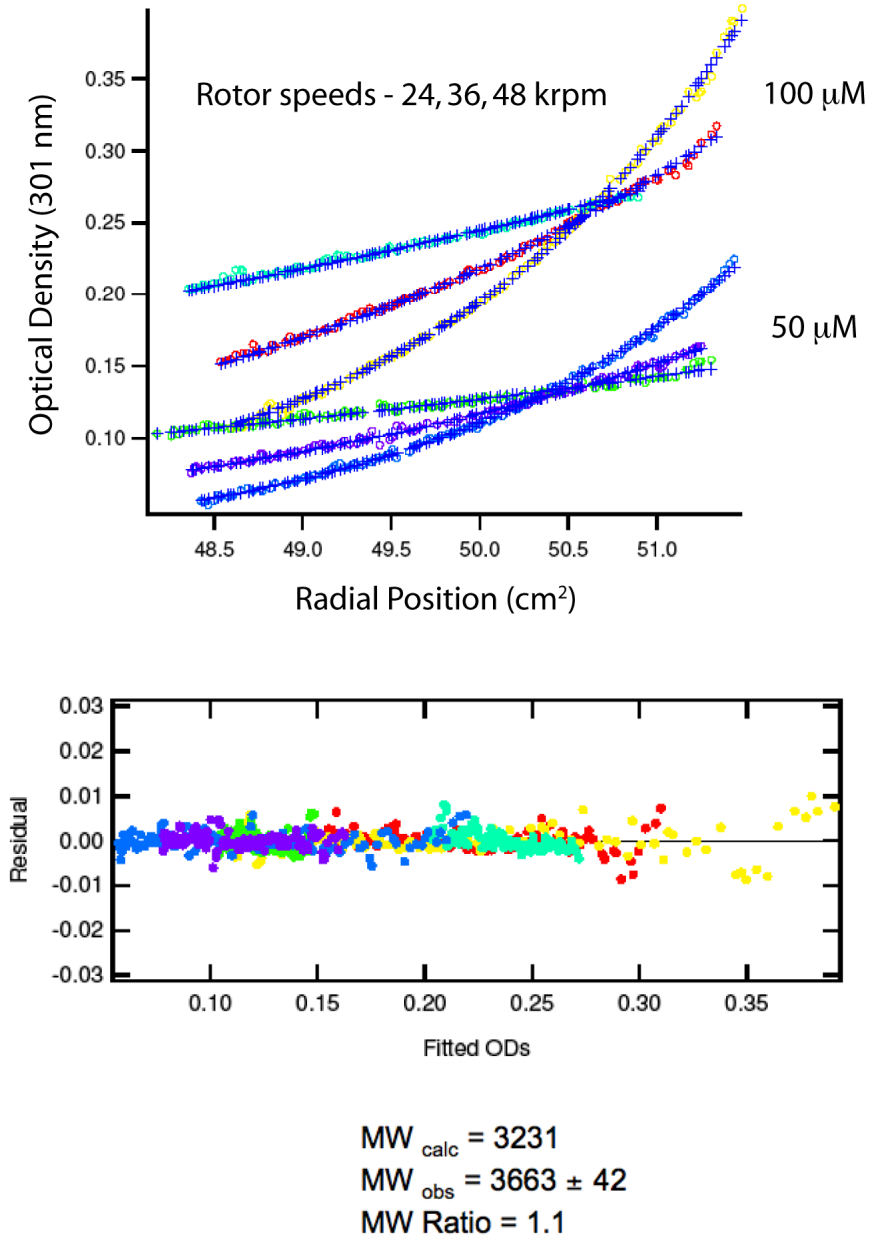


Fig. 6.8. Representative Sedimentation Equilibrium Analytical Ultracentrifugation data for full-amide peptide **83** at 100 and 50 μM (top to bottom), in the upper scheme (regression fit). In the lower scheme, in coloured circles, there are the collected data. The data fit well to a single species model of molecular weight 3663 g/mol as judged by the randomness of residuals. The expected mass of the monomer is 3231 g/mol.

6.2.2.2 NMR spectroscopic studies

This part of the work was performed by Dr. Felix Freire-Iribarne.

Full-amide peptide **83** and full length thioester **84** were studied by 1D and 2D NMR spectroscopy and every proton was assigned. The high-resolution conformational analysis of

these peptides allowed us to examine the folding of full length thioester **84** in more detail, by comparing the chemical shifts of the α proton ($\delta C_\alpha H$) of each residue in the two sequences. $\delta C_\alpha H$ values are influenced in characteristic ways by local secondary structure.[51,60] Relative to "random-coil" values, the $\delta C_\alpha H$ values for residues in a β-sheet secondary structure are shifted downfield (a shift in the opposite direction is observed for α-helical residues). **Fig. 6.9** compares $\Delta\delta C_\alpha H$ data [$\Delta\delta C_\alpha H$= $\delta C_\alpha H$ (observed)- $\Delta\delta C_\alpha H$ (random coil)]. From these studies the full-amide peptide **83** and the full length thioester are folded in aqueous solution and are consistent with a **β-hairpin** molecule. These data are very important because it is necessary to avoid deviations in the folded structure of the thiodepsipeptide from the full-amide peptide of interest, that could lead to the misinterpretation of BTE data.

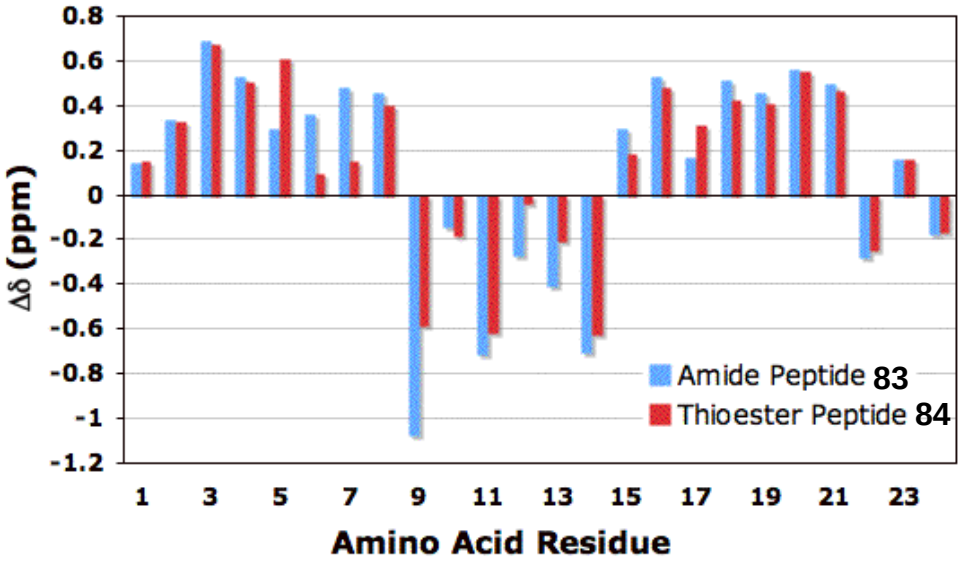


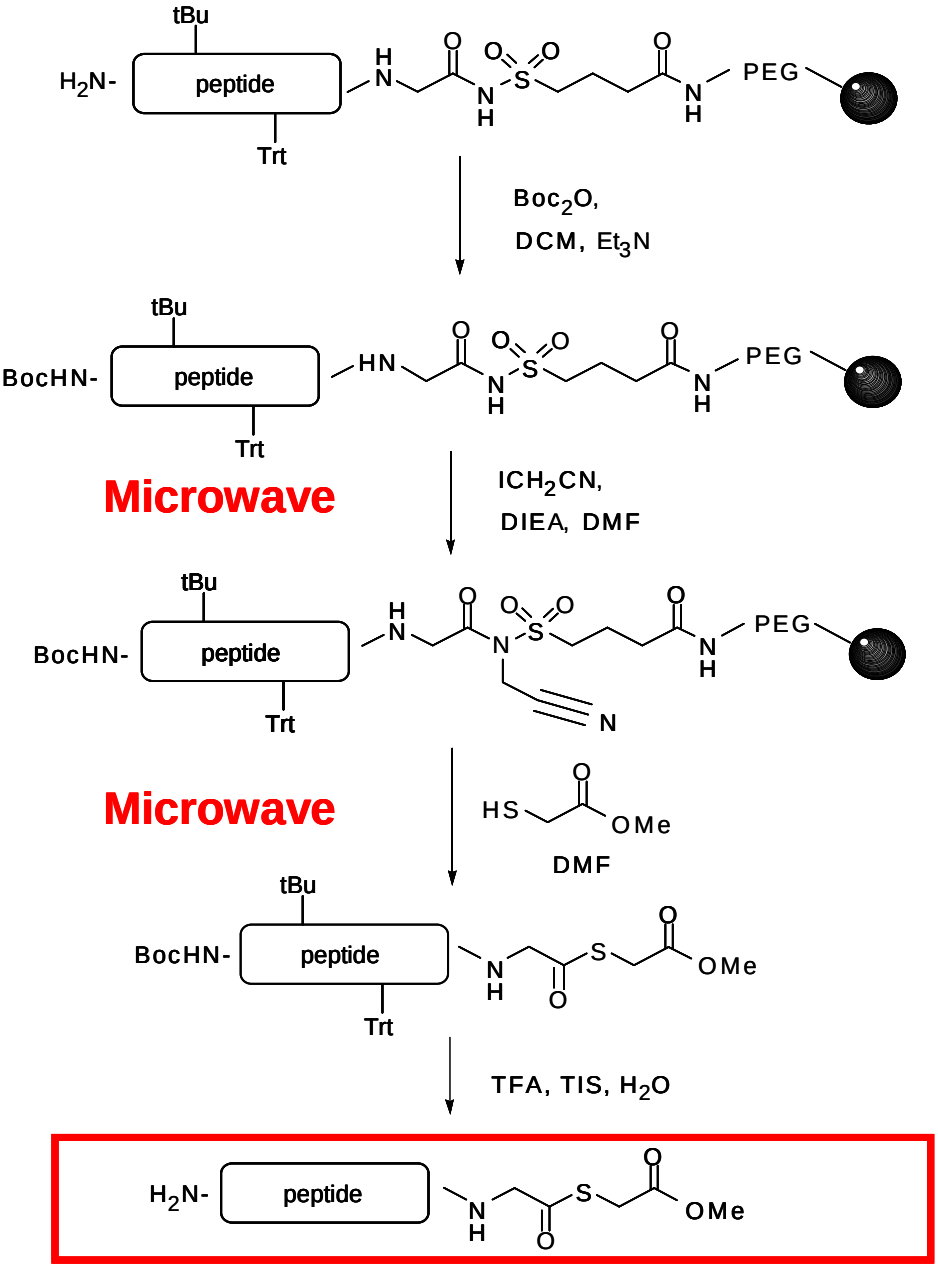
Fig. 6.9. Chemical shift deviation (CSD) for full-amide peptide **83** and full length thioester peptide **84**: the structures are folded similarly.

Crucial to the success of BTE is the ability to efficiently synthesize and purify thioester-containing peptides. These peptides focused on the use of Ellman's 4-sulfamylbutiryl safety catch linker attached to an appropriate resin, such as AM NovaGel from Novabiochem.[57]

The thioester peptide **85** (**Fig. 6.10**) was synthesized on a Fmoc-Gly-4-sulfamylbutyryl AM NovaGel resin, according to the safety catch method (**Scheme 6.2** and **Experimental section**).

Fig. 6.10. Primary sequence of the thioester **85**.

The safety catch strategy relies on the alkylation of a sulfonamide nitrogen, after peptide has been assembled, with either iodoacetonitrile[57b,57c] or TMS-diazomethane,[57a,58] to activate the sulfonamide, which is subsequently displaced by a nucleophile of choice. If a thiol is chosen as the nucleophile, cleavage from the resin generates a peptide with a C-terminal thioester.[59] The microwave irradiation is used to speed up the activation/cleavage steps.



Scheme 6.2. Cartoon depicting the thioester **85** synthesis.

The thiols **86-100** (**Fig. 6.11**) were synthesized manually in parallel on a Wang resin, using standard Fmoc solid phase peptide synthesis, on a 25 μmol scale.

Fig. 6.11. Structure of the synthesized thiols **86-100**. α is the substitution position (α = 19 proteinogenic amino acids, except cysteine, because its side chain is a thiol).

The purification of the peptides by HPLC resulted quite difficult due to MeCN shortage. This forced the laboratory to find alternative organic solvents (B solvent). In **Fig. 6.12** an example of purification is shown; this study indicates that MeCN is the best solvent for obtaining higher yields and a better HPLC resolution. In fact, in MeOH and in EtOH the peaks are broader and less resolved than in MeCN. In this solvent on the left and on the right of the main principal desired peak there are two smaller peaks which probably are superimposed in the alcoholic solvents. Moreover MeOH and EtOH are more dense than MeCN, and consequently give problems of counter pressure in the HPLC instruments. The alcoholic solvents, in presence of TFA, fundamental component of the purification, form the correspondent methyl/ethyl TFA esters, changing the composition of the eluents, and consequently, their separation ability, during the purification. A percentage of water can be added to the alcoholic B solution to overcome this problem.

The alcoholic solvents had also to be removed by rotary evaporation, before the peptide liofilization. However with MeOH and MeOH:H_2O = 9:1 as B solvents the purification was possible, with longer time, higher cost and lower yield. The product to be purified corresponded always to the principal peak, indicated by the arrows. After purification, the collected fractions were evaporated and injected in the analytical column, in the following HPLC conditions: C18 analytical column (4.6×250 mm), flow rate 1 mL/min, gradient of solvent 10-60% B (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A (H_2O:CF_3CO_2H, 100:0.1, v/v) over 50 min. A purity >= 95% was considered acceptable.

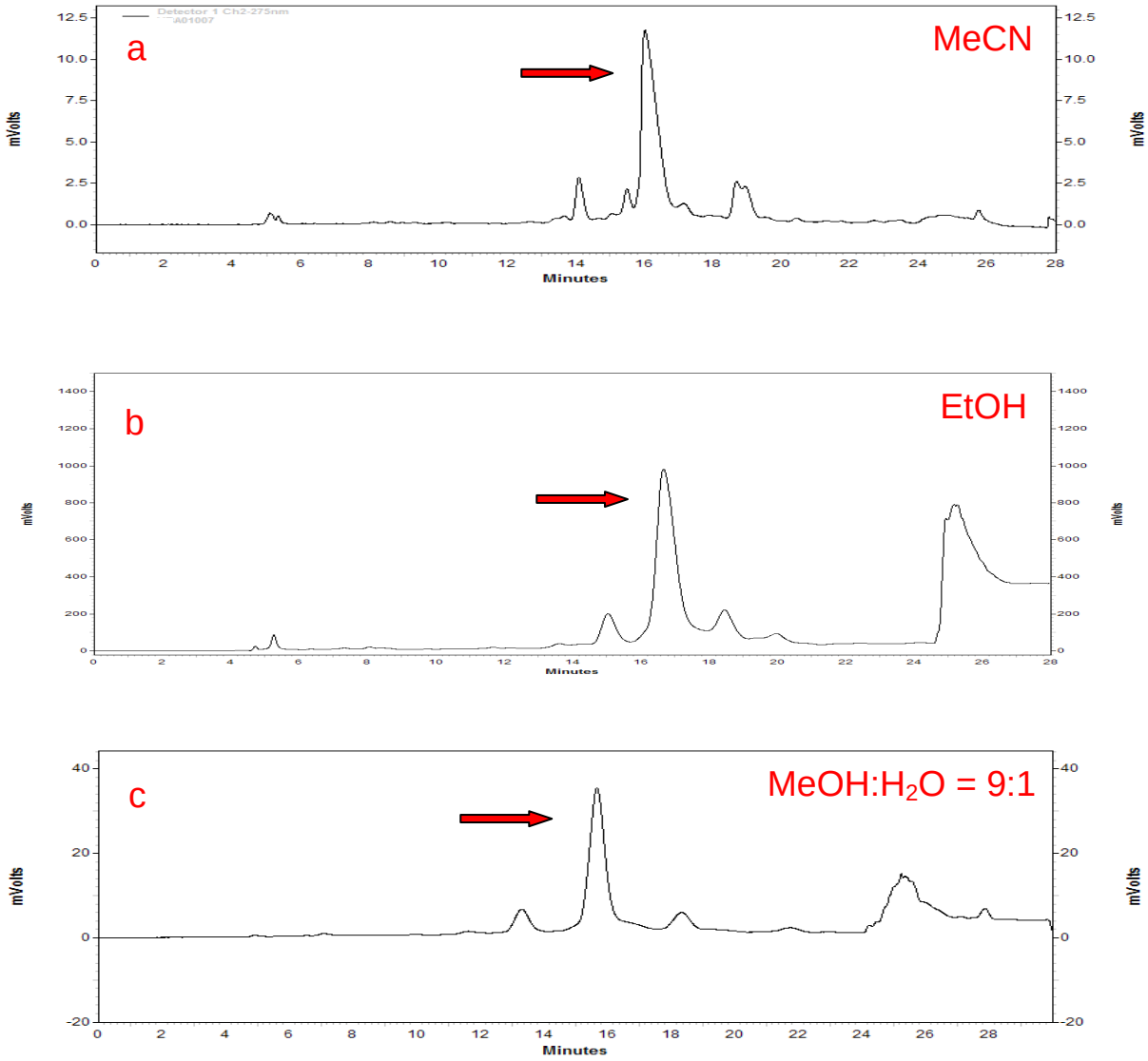


Fig. 6.12. Example of the thioester **85** purification, using different B solvents: a) MeCN, b) EtOH, c) MeOH:H_2O = 9:1.

In the next table (**Table 6.1**) there is a summary of the BTE data (see **Experimental section**), expressed as ΔG_{fold} (Kcal/mol).

Mutant	Average	STDEV
Val	-1.44	0.08
Ile	-1.41	0.06
Glu	-1.00	0.04
Thr	-0.78	0.07
Ser	-0.73	0.04
Ala	-0.60	0.10
Leu	-0.53	0.10
Asn	-0.22	0.11
Gly	-0.13	0.08
Arg	-0.06	0.08

Table 6.1. BTE summary of the ΔG_{fold} (Kcal/mol) obtained with 10 of 19 full-length thioester mutants.

The β-branched amino acids, such as valine and isoleucine, are known to favour β-sheets whereas glycine and asparagine are known as low **propensity** amino acids in forming β-sheets (see **Introduction**). From the **Table 6.1** it was confirmed that our scale is in good agreement with what is known in literature, for the amino acids which favour or disfavour β-sheet formation. Then, with these preliminary data, a comparison was started (**Fig. 6.13-6.18**) with different β-sheet **propensity** scales, to see if there is an agreement between our scale and others reported in literature and to study a possible **context dependence**. In the case of agreement, the data should lie on a straight line.

The Muñoz & Serrano[42a] (**Fig. 6.13**) correlation is the more interesting. If glutamic acid and arginine are left out, the correlation is quite good. This particular scale should be the most appropriate, since **context** effects should be minimized, and since their approach to statistical analysis of the protein structure database is more sophisticated than that of Chou & Fasman[36] (**Fig. 6.14**).

The glutamic acid mutant (E) really seems to be an anomaly in the BTE studied system. All the other scales ranked it quite low (<14th), however, in this one is ranked pretty high (3rd out of 10).

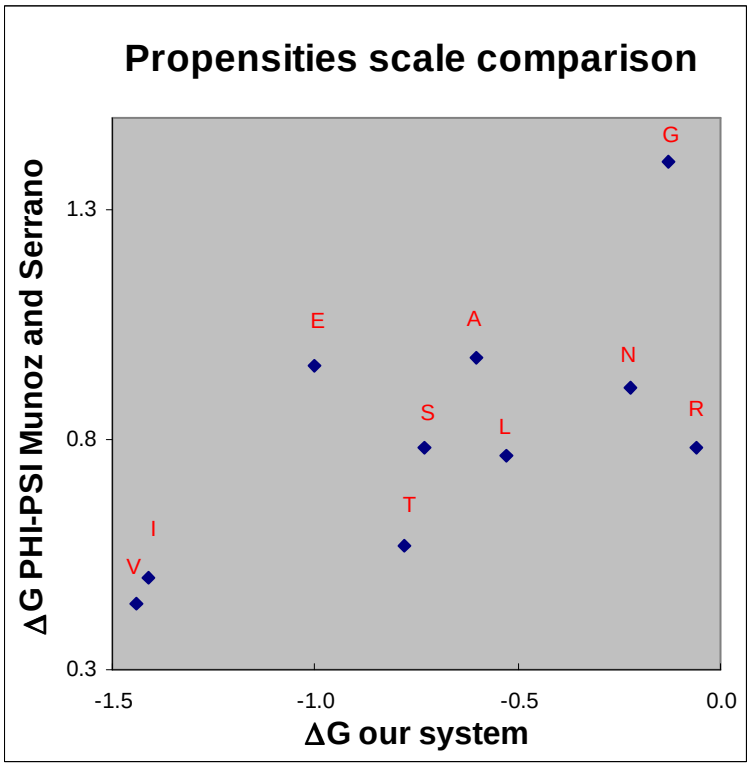


Fig. 6.13. **Propensity** scales comparison between the BTE scale and the L. Munoz & V. Serrano's scale.[42a]

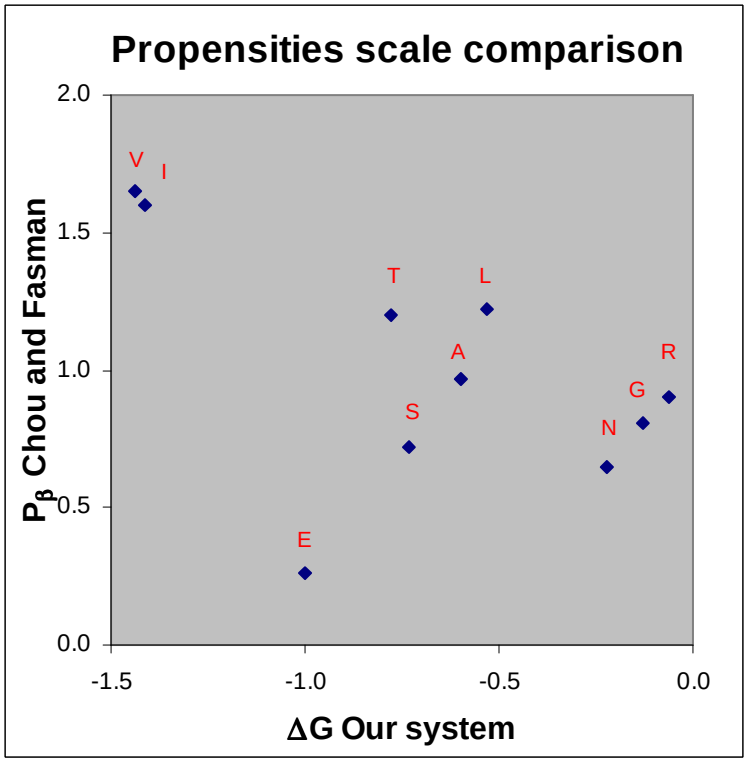


Fig. 6.14. **Propensity** scales comparison between the BTE scale and the P. Y. Chou & G. D. Fasman's scale.[36]

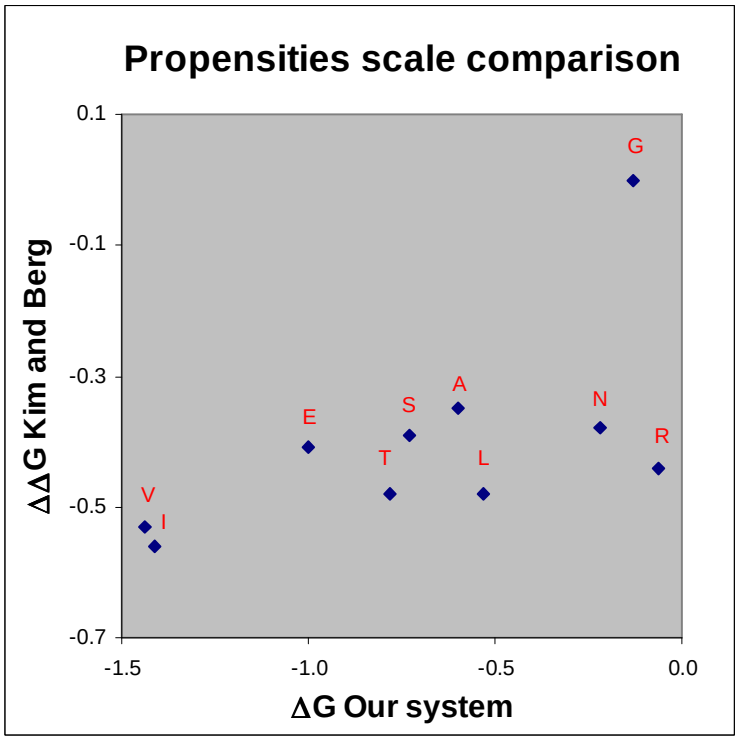


Fig. 6.15. **Propensity** scales comparison between the BTE scale and the C. W. A. Kim & J. M. Berg's scale.[39a]

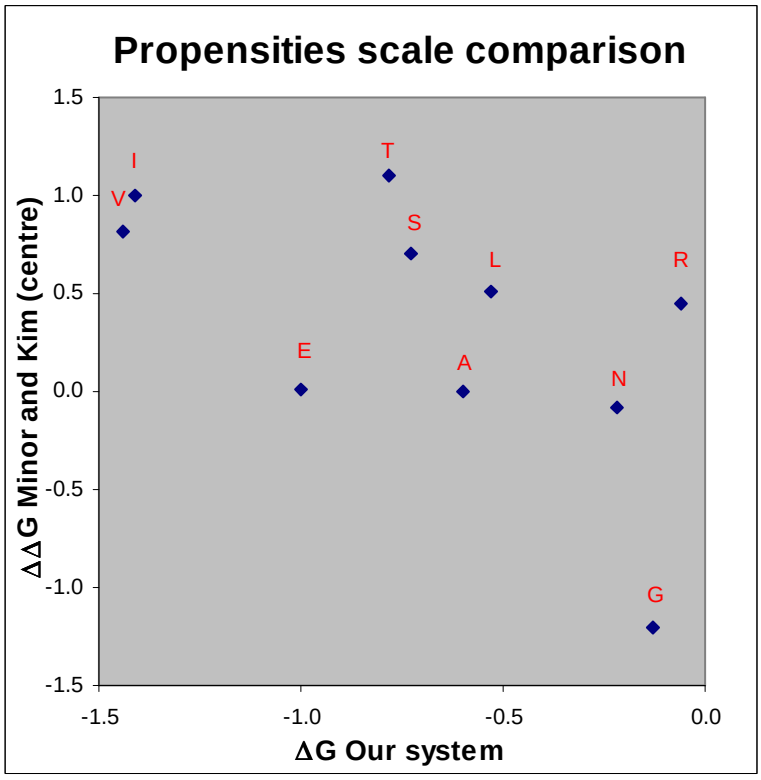


Fig. 6.16. **Propensity** scales comparison between the BTE scale and the D. L. Minor & P. S. Kim's scale.[41b]

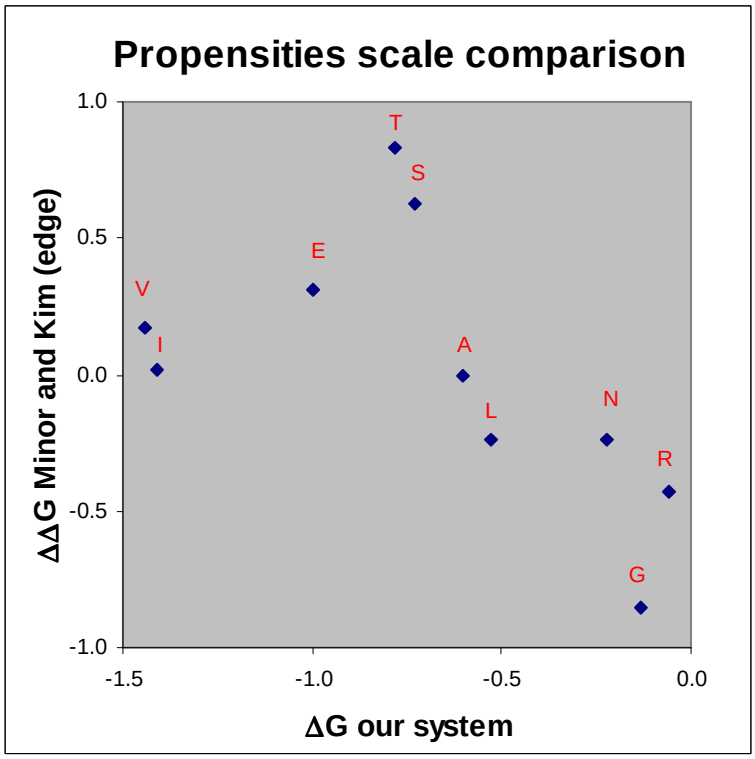


Fig. 6.17. **Propensity** scales comparison between the BTE scale and the D. L. Minor & P. S. Kim's scale.[43]

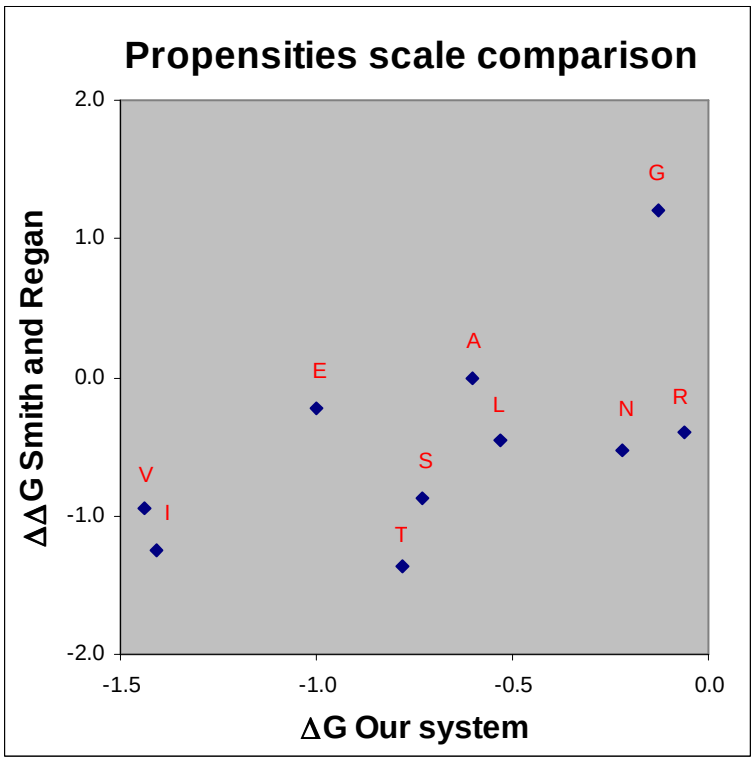


Fig. 6.18. **Propensity** scales comparison between the BTE scale and the C. K. Smith, J. M. Withka & L. A. Regan's scale.[41]

If the sequence diagram is considered (**Fig. 6.19**), the three cross strand neighbors of the mutated amino acid are tyrosine, valine and glutamic acid, so there is no immediate cross strand salt bridge, but there is an arginine on the other strand (diagonal + 1) that could be potentially involved and could justify these data, making our system **context dependent**.

Several of the better amino acids from the scale are missing, so it will be interesting to see how the comparison develops with the new BTE mutants.

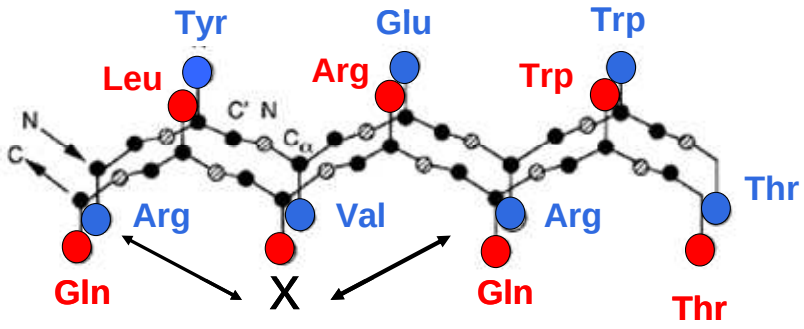


Fig. 6.19. Cartoon depicting the reciprocal position of the amino acids close to the mutants site in the studied **β-hairpin**, which could explain the ΔG_{fold} data and the propensities scales comparisons.

6.3 Conclusions

In conclusion, an antiparallel **β-hairpin** (**83**, **Fig. 6.6**) was designed, synthesized, purified, characterized by 1D and 2D NMR and Sedimentation Equilibrium Analytical Ultracentrifugation. Having shown that it is stable and is folded, and does not aggregate in aqueous solution below 100 μM, it was studied via backbone thioester exchange to determine an experimental β-sheet **propensity** scale for α-amino acid residues.

Up to now ΔG_{fold} data for 10 of the 19 amino acids (cysteine was not studied because its side chain is a thiol) were obtained, and a comparison was started with different **propensity** scales. These preliminary results are very promising because the trends observed are consistent with what is known in literature. The BTE assays have to be finished with the last nine mutants to complete the comparison with other **propensity** scales and to understand if there is a good agreement with them, and if and how the synthesized promising system is **context dependent**.

6.4 Experimental section

Materials

N-9-Fluorenylmethoxycarbonyl (Fmoc) L-α-amino acids were obtained from Novabiochem. All Fmoc-amino acids with side chains that required protection are listed with the protecting

groups in parentheses: Arg (Pbf), Asn (trityl), Asp (tert-butyl), Gln (trityl), Glu (tert-butyl), His (trityl), Lys (Boc), Ser (tert-butyl), Thr (tert-butyl), Trp (Boc), Tyr (tert-butyl). All solid phase resins were purchased from Novabiochem. N,N-diisopropylethylamine was distilled over CaH before use. CH_2Cl_2 and DMF were obtained from Aldrich Chemical Co. as "biotech grade, >99.9%" and used without further purification. HPLC grade MeCN, MeOH and EtOH were obtained from Burdick & Jackson, Aldrich or Fisher Scientific.

The H_2O used in HPLC separations was millipore grade. All other reagents were obtained from Aldrich Chemical Co. unless otherwise noted.

Procedures

All HPLC runs were carried out on Shimadzu systems equipped with UV/Vis detectors.

HPLC traces were collected at 220 and 275 nm. Several reverse phase columns were purchased from Vydac: C18 preparative (25×250 mm), C18 analytical (4.6×250 mm), C18 semipreparative (10×250 mm), and C18 semipreparative Everest^{TM} (10×250 mm). The columns denoted C18 are the standard polymeric resins. The two different versions of C18 columns provide slightly different retention patterns of peptides.

Mass spectra (MS) for peptides were obtained using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) on a Bruker spectrometer. The peptides were embedded in a matrix of α-cyano-4-hydroxycinnamic acid.

General Peptide Synthesis

Several protocols for solid phase peptide synthesis are described below.

Automated peptide synthesis was typically used in situations where peptides were needed promptly, such as thioester **85**, because automated synthesis is generally faster and more reliable than manual synthesis. The Symphony synthesizer was used. Manual solid phase peptide synthesis was used in situations where 3 or more related peptides could be synthesized in parallel (manual synthesis is more time-effective in a parallel mode), so for the thiols **86**-**100**, or if the sequence was difficult to get with the synthesizer, like the full-amide peptide **83**.

Solid Phase Synthesis Method 1: on Symphony Automated Synthesizer.

Thioester **85** was partially synthesized on solid phase using a Symphony automated synthesizer (Protein Technologies Inc.). This instrument can routinely synthesize peptides on a 15-50 μmol scale, and may be modified for synthesis of up to 150 μmol.

Synthetic cycles were completed with a standard coupling time of 30-60 min using Obenzotriazol-1-yl-N,N,N',N'-tetramethyluronium hexafluorophosphate (HBTU, 5 eq.), N-Methyl Morpholine (NMM, 20 eq.), with DMF as the solvent. Five eq. of Fmoc amino acid were used for each coupling cycle. Deprotection steps used 20% piperidine in DMF for 1×5 min + 1×15 min. C-terminal amide peptides were prepared on rink amide NOVAPEG resin.[61]

Solid Phase Synthesis Method 2: manual synthesis.

Peptide synthesis was carried out using Alltech filter tubes. Manual synthesis allows the scale to be conveniently adjusted from 25 μmol-50 μmol. Synthetic cycles were completed with a standard coupling time of 120 min using O-benzotriazol-1-yl-N,N,N',N'-tetramethyluronium hexafluorophosphate (HBTU, 3 eq.). DMF was the solvent. Four equivalents of Fmoc amino acid were used for each coupling cycle.

Deprotection steps used 20% piperidine in DMF for 15 min. C-terminal amide peptides were prepared on rink amide resin. N-terminal thiols were obtained by the coupling of S-tritylthioglycolic acid[62] as the last residue under standard conditions by adding the residue directly to the reaction vial at the appropriate step.

Full-amide and thiol peptide synthesis

The synthesis was done manually, on a 50 μmol scale for the full-amide peptide **83**, and on a 25 μmol scale for the thiols **86**-**100**, starting from a Wang resin for both.

Thioester 85 synthesis

After all the amino acid sequence was synthesized by automatic synthesizer on a 50 μmol scale, with Fmoc amino acids, the N-terminus was Boc protected by adding 14:5:1 Boc_2O:CH_2Cl_2:Et_3N w/v/v (ca. 2.5 mL) to the resin and rocking for 2 h. The resin was washed five times with CH_2Cl_2 and dried under vacuum. The dry resin was swollen in 1 mL DMF in a 4 mL fritted filter tube, and 0.5 M DIEA in DMF (0.2 mL) and iodoacetonitrile (32 μL, 442 μmol) were added to the resin. The iodoacetonitrile was filtered over basic alumina immediately prior to use. The filter tube was placed in a CEM Discover™ microwave system and heated 30 min at 50 °C. The reaction mixture was then filtered, and the resin was washed with CH_2Cl_2 (5×5 mL) and DMF (5×5 mL). DMF (1 mL) and methyl thioglycolate (91 μL, 1.0 mmol) were added to the resin. The Alltech tube was again placed in a glass microwave tube and microwaved 30 min at 50 °C. The liquid was filtered into a flask, and the resin was then washed three times with CH_2Cl_2 (5 mL), collecting the filtrate each time. The combined filtrates were condensed with a stream of N_2 and placed under vacuum prior to global deprotection.

Deprotection

Peptides were cleaved from resin and deprotected by stirring with (CF_3CO_2H:H_2O:triisopropylsilane, 90:5:5, v/v/v) (2 mL/ 25 μmol) for 3 h, followed by precipitation into cold diethyl ether. The precipitate was collected by centrifugation/decantation prior to purification. C-terminal thioester peptide (already cleaved from resin) was deprotected using the same cleavage procedure.

Purification and Characterization

*[Thiols **86-90** were synthesized, purified and characterized by Jay Steinkruger, thiols **97-100** were purified and analyzed together with Kelsey Mayer]*

All the peptides were purified by reverse-phase HPLC and characterized by analytical HPLC (**Table 6.2**), for the purity, and MALDI, for the identity. The estimated purity of the peptides is above 95%, based on the analytical HPLC.

Peptide	**Retention time (min)***	**$[M + H]^+$ calculated**	**$[M + H]^+$ observed****
83	26	3230.7	3231.2
85	23	1864.0	1864.4
86, **Ile**	29	1489.8	1489.7
87, **Val**	27	1475.8	1476.3
88, **Ala**	25	1447.8	1447.5
89, **Gly**	21	1433.8	1433.7
90, **Asn**	19	1490.8	1490.7
91, **Ser**	23	1463.8	1464.2
92, **Thr**	23	1477.8	1477.9
93, **Leu**	29	1489.8	1489.9
94, **Asp**	26	1491.8	1492.3
95, **Arg**	22	1532.8	1533.2
96, **Gln**	23	1504.8	1505.5
97, **Glu**	To be done	1505.8	1505.9
98, **Tyr**	To be done	1539.8	1539.8
99, **His**	26	1513.8	1514.0
100, **Pro**	25	1473.8	1474.2

Table 6.2. Characterization data for the peptides described in this chapter.

* HPLC Conditions: C18 analytical column (4.6×250 mm), flow rate 1 mL/min, gradient of solvent 10-60% B (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A (H_2O:CF_3CO_2H, 100:0.1, v/v) over 50 min. ** MALDI-TOF

Representative MALDI analysis follows:

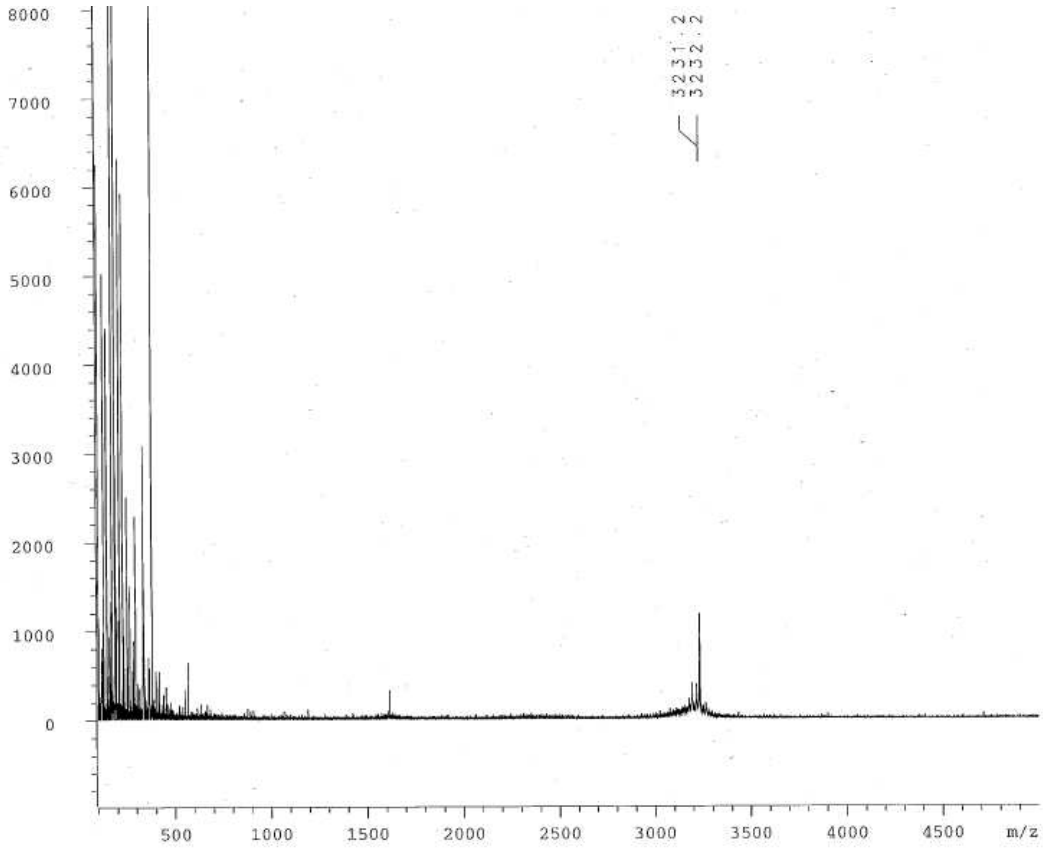


Fig. 6.20. MALDI of the full-amide peptide **83**.

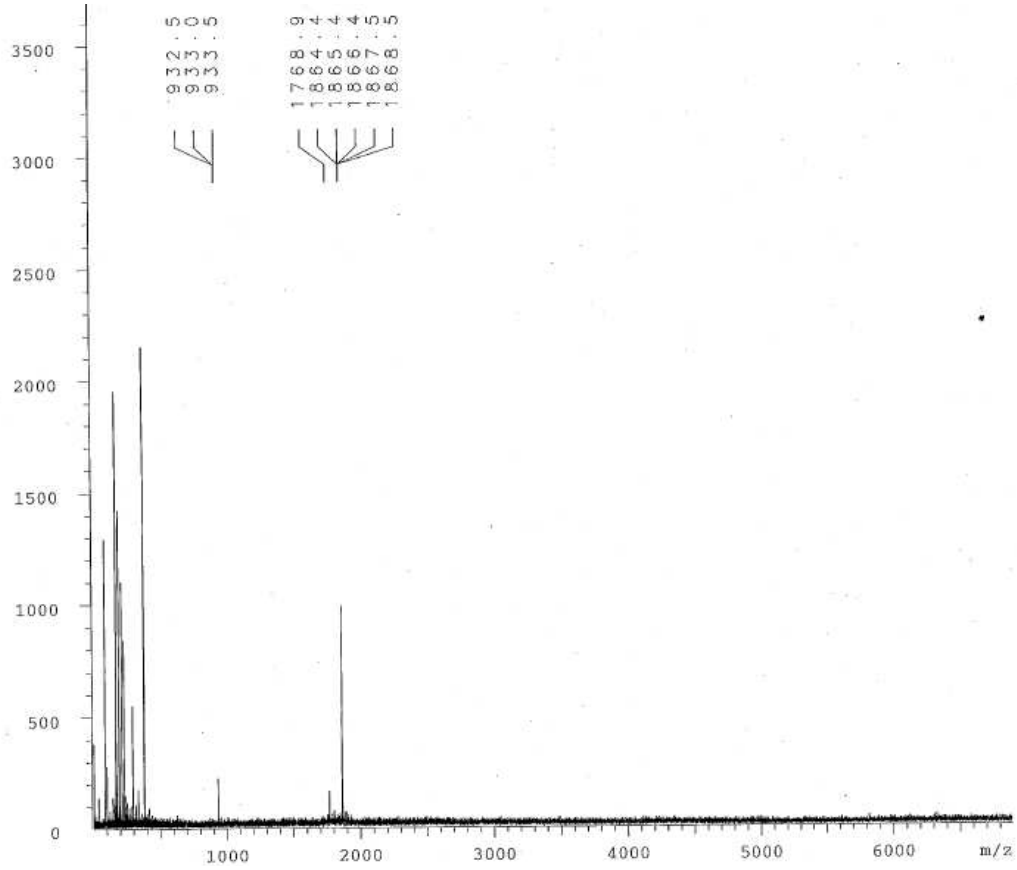
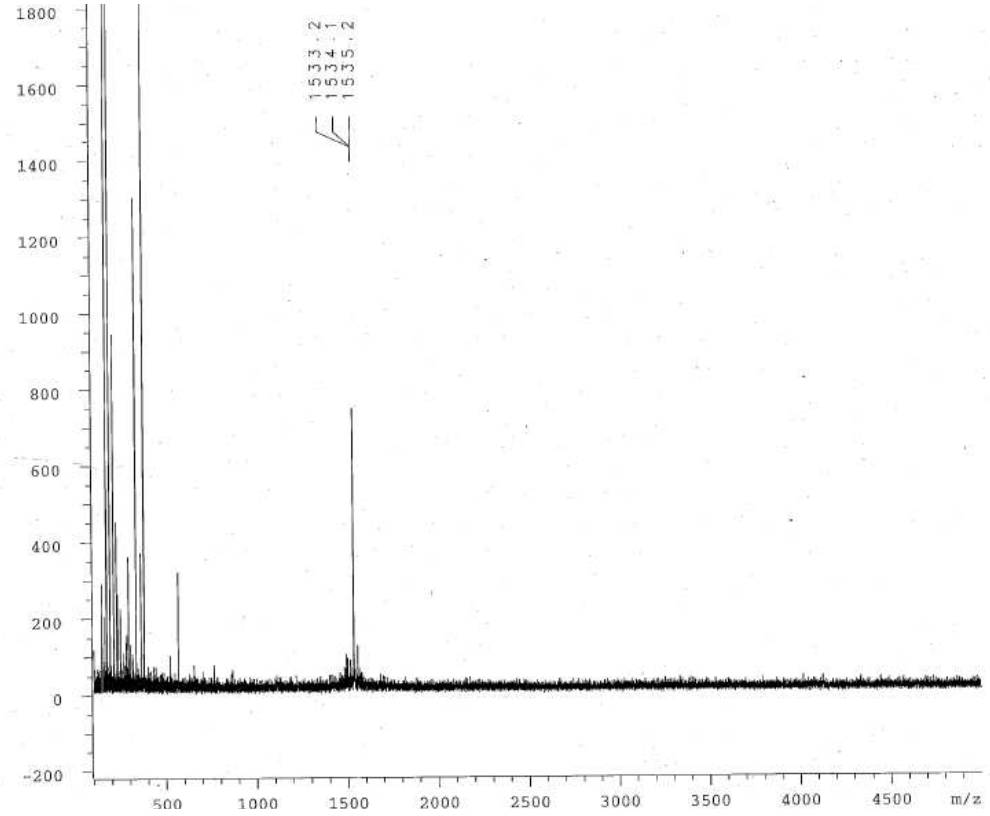


Fig. 6.21. MALDI of the thioester **85**.



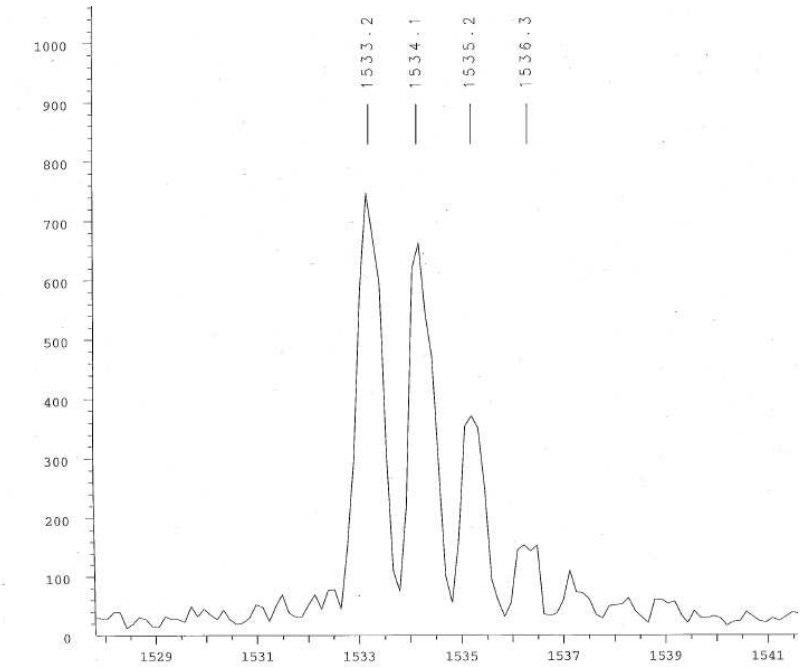


Fig. 6.22. MALDI of the arginine thiol **95**, full spectrum (upper image) and zoom on the isotopic pattern (lower image).

Representative characterization of the purity of the peptide follows:

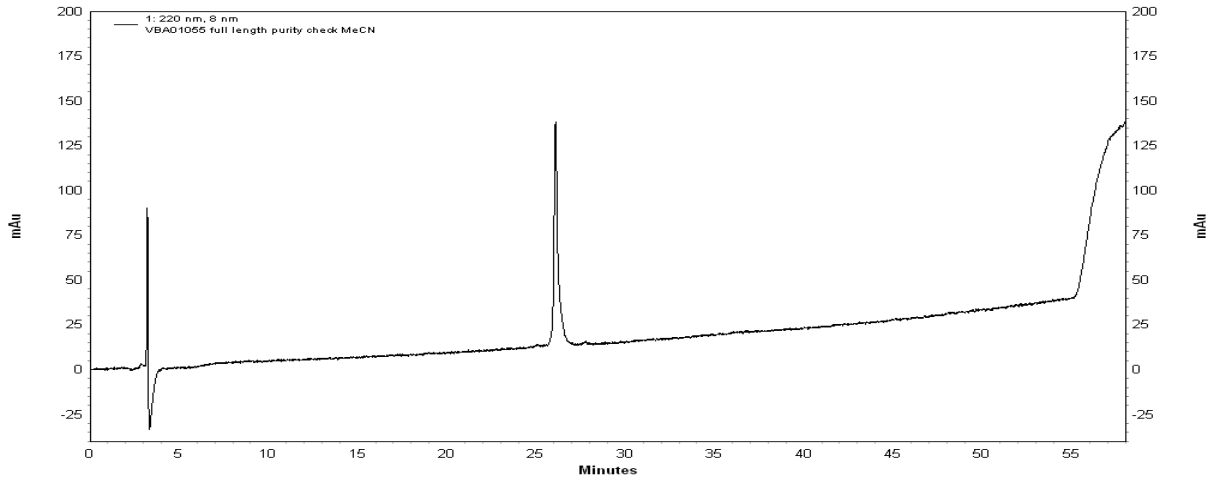


Fig. 6.23. Purity check of the full-amide peptide **83**.

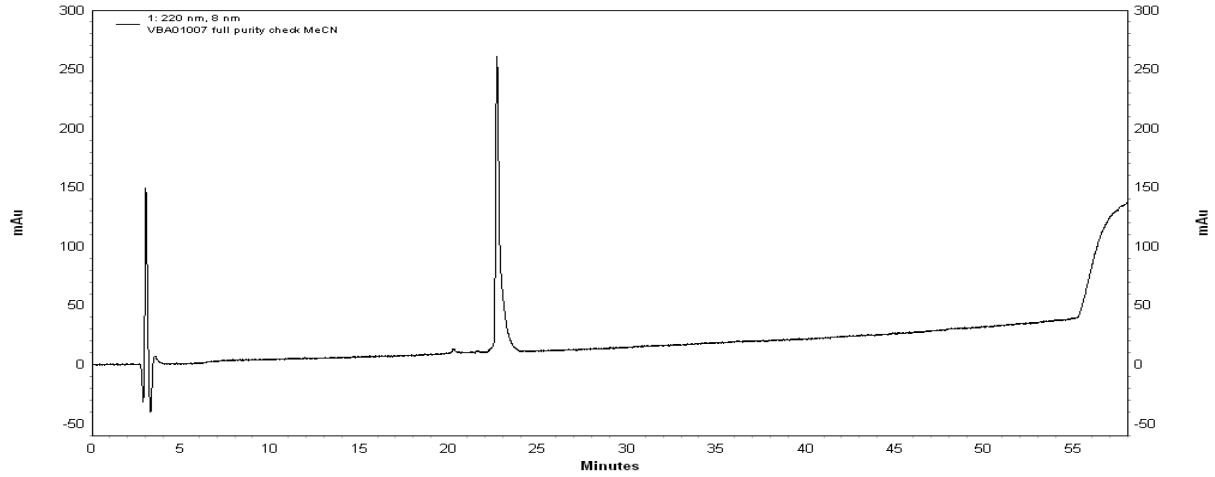


Fig. 6.24. Purity check of the thioester **85**.

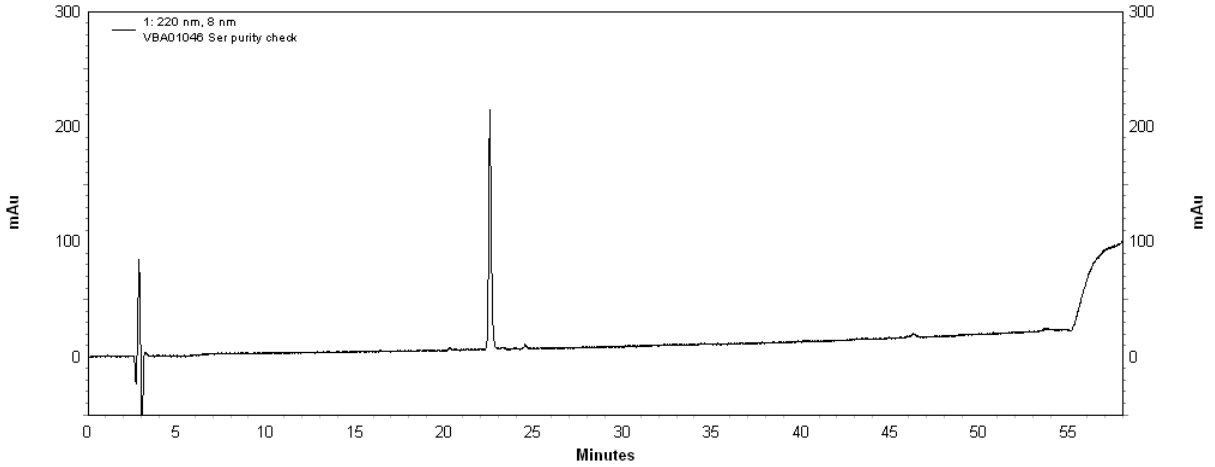


Fig. 6.25. Purity check of the serine thiol **91**.

Sedimentation Equilibrium Analytical Ultracentrifugation

[Analysis by Jay Steinkruger]

Sedimentation studies were conducted on a Beckman XLA ultracentrifuge at 25 ºC. Peptide solutions in 50 mM pH 7 HEPES buffer were loaded into 1.2 cm cells and absorbance was monitored at 301 nm. At each speed, data were collected with a 0.001 cm step size every 2 h until two consecutive spectra were identical (typical equilibration times were 10 h). Non-linear regression was performed in accordance with the expression $c_r = c_o \exp [M(1-\nu\rho)\omega^2(r^2 - r_o^2)/2RT] + \text{base}$, where c_r is the concentration (in absorbance units) at radial position r, c_o is the concentration at an arbitrary reference position r_o near the meniscus, ν is the partial specific volume, ρ is the solvent density, ω is the rotor speed, R is the gas constant, T is the temperature, and base is a baseline absorbance correction to account for non-sedimenting species. Molecular weight estimates were obtained from the parameter *M*. Global fits to the

data at all speeds and concentrations were judged to be adequate by randomness of residuals. A partial specific volume of 0.7 mL/g was calculated based on amino acid composition.

Thioester Exchange Assays

Stock solutions of 250 mM sodium phosphate buffer (pH 7.0 at 250 mM) and 20 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP) were produced by using Millipore H_2O. Typically, assays were initiated by mixing approximately equal portions of a C-terminal thiol peptide and the N-terminal thioester. Approximately a pinch of spatula of each dry peptide was placed into a 500 μL vial insert. Most assays were conducted on a 300 μL scale by addition of liquids into the vial insert, to give initial assay concentrations of maximum 0.1 mM peptides, 2.0 mM TCEP, and 50 mM buffer (TCEP was included in the assay solutions to prevent disulfide formation during the thioester exchange equilibration). H_2O was used for the excess volume. Typically 30-50 μL of assay solution were injected per HPLC run, the peaks area were quantified and the ΔG_{fold} was calculated as explained in the **Introduction**. Each assay was allowed to equilibrate ca. 2 h before HPLC injection. The HPLC column used in every case was either a C_{18} Everest™ or a C_{18} Vydac analytical column (4.6 mm i.d. 250 mm length) at a flow rate of 1.0 mL/min. A gradient of B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) was used for elution. Equilibrating species were identified by HPLC retention time and MALDI mass spectrometry.

BTE assay gradients

To establish in which time range the species are at the equilibrium, time equilibrium studies were performed, in which the mixture was injected every hour, and equilibrium was considered when no significative change in ΔG_{fold} was observed.

To study replicates of the BTE assays, different samples were prepared with the same thiol and thioester, which were injected consecutively.

To get the best separation between the species, some analytical columns were tried, and the best resulted C18 Everest™. Also different solvent gradients of MeCN:H_2O with 0.1% TFA were checked, because the species involved have very similar retention times and their separation resulted quite difficult.

Full-amide peptide 83 purification.

HPLC: C18 preparative column (25×250 mm), flow rate 15 mL/min, gradient of 40-50% B solvent (MeOH:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min.

Thioester 85 purification.

HPLC: C18 preparative column (25×250 mm), flow rate 15 mL/min, gradient of 19-29% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min and, a second time, gradient of 40-50% B solvent (MeOH:H_2O:CF_3CO_2H, 90:10:0.1, v/v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min.

Thiols purification.

Serine 91.

HPLC: C18 preparative column (25×250 mm), flow rate 15 mL/min, gradient of 0% for 5 min, 0-40.2 over 5 min, 40.2-50.2% B solvent (MeOH:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v).

Threonine 92.

HPLC: C18 preparative column (25×250 mm), flow rate 15 mL/min, gradient of 0% for 5 min, 0-40.2 over 5 min, 40.2-50.2% B solvent (MeOH:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v).

Leucine 93.

HPLC: C18 preparative column (25×250 mm), flow rate 15 mL/min, gradient of 46-56% B solvent (MeOH:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v).

HPLC: C18 semipreparative column (10×250 mm), flow rate 3 mL/min, gradient of 52-62% B solvent (MeOH:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v).

Aspartic acid 94.

HPLC: C18 preparative column (25×250 mm), flow rate 15 mL/min, gradient of 40-50% B solvent (MeOH:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min.

Arginine 95.

HPLC: C18 preparative column (25×250 mm), flow rate 15 mL/min, gradient of 34-44% B solvent (MeOH:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v).

Glutamine 96 (still to be purified).

Glutamic acid 97.

HPLC: C18 preparative column (25×250 mm), flow rate 15 mL/min, gradient of 38-48% B solvent (MeOH:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v).

Tyrosine 98.

HPLC: C18 preparative column (25×250 mm), flow rate 15 mL/min, gradient of 23-33% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min.

Histidine 99.

HPLC: C18 preparative column (25×250 mm), flow rate 15 mL/min, gradient of 19-29% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min.

Proline 100.

HPLC: C18 preparative column (25×250 mm), flow rate 15 mL/min, gradient of 21-31% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min.

BTE assays gradients.

Isoleucine and Leucine mutants

HPLC: C18 analytical column (4.6×250 mm), flow rate 1 mL/min, gradient of 8-32% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 10 min, then 32-42% over 10 min.

Valine mutant

HPLC: C18 analytical column (4.6×250 mm), flow rate 1 mL/min, gradient of 19-29% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 30 min.

Alanine mutant

HPLC: C18 analytical column (4.6×250 mm), flow rate 1 mL/min, gradient of 19-29% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min.

Glycine mutant

HPLC: C18 analytical column (4.6×250 mm), flow rate 1 mL/min, gradient of 16-26% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 40 min.

Asparagine mutant

HPLC: C18 analytical column (4.6×250 mm), flow rate 1 mL/min, gradient of 19-29% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min.

Serine mutant

HPLC: C18 analytical column (4.6×250 mm), flow rate 1 mL/min, gradient of 19-29% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min.

Threonine mutant

HPLC: C18 analytical column (4.6×250 mm), flow rate 1 mL/min, gradient of 25-35% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min.

Glutamic acid mutant

HPLC: C18 analytical column (4.6×250 mm), flow rate 1 mL/min, gradient of 25-35% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min.

Arginine mutant

HPLC: C18 analytical column (4.6×250 mm), flow rate 1 mL/min, gradient of 19-29% B solvent (CH_3CN:CF_3CO_2H, 100:0.1, v/v) in A solvent (H_2O:CF_3CO_2H, 100:0.1, v/v) over 20 min.

BTE assays

The HPLC traces for all the BTE assays studied up to now, after equilibrium had been attained, are shown below. In bold there are the identified species, by MALDI, corresponding to thioester hydrolysis (TE Hy), thioester (TE), full length thioester (FL) and thiol (TH).

The thioester hydrolisis was identified because from the MALDI a loss of 88 uma of was observed, which corresponds to a loss of methylthioglycolate and in addition a OH.

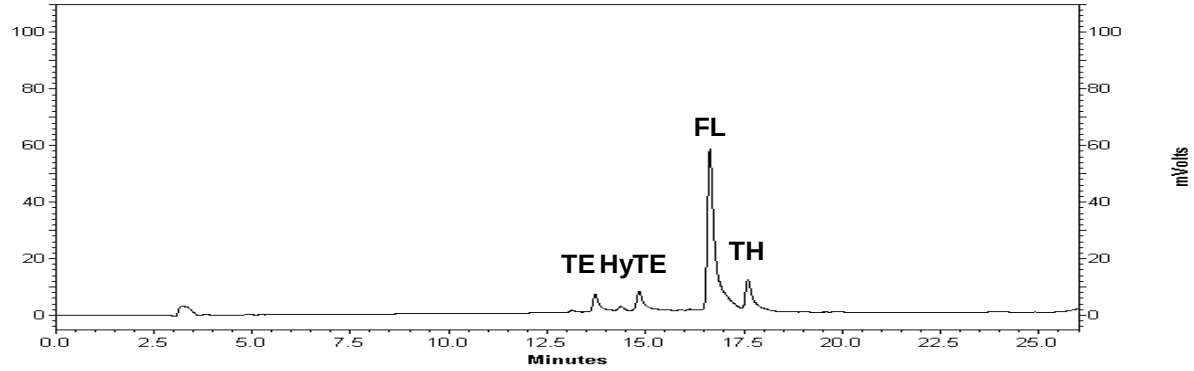


Fig. 6.26. BTE assay of the isoleucine mutant **86**.

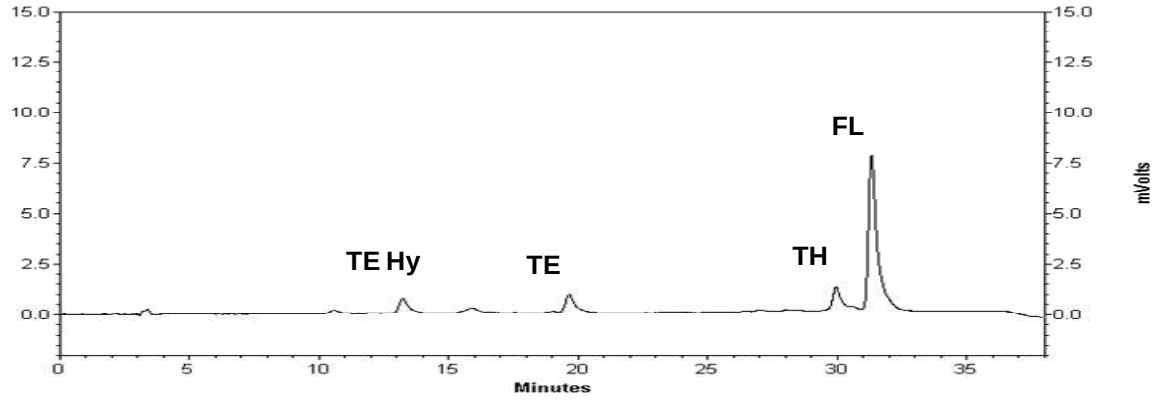


Fig. 6.27. BTE assay of the valine mutant **87**.

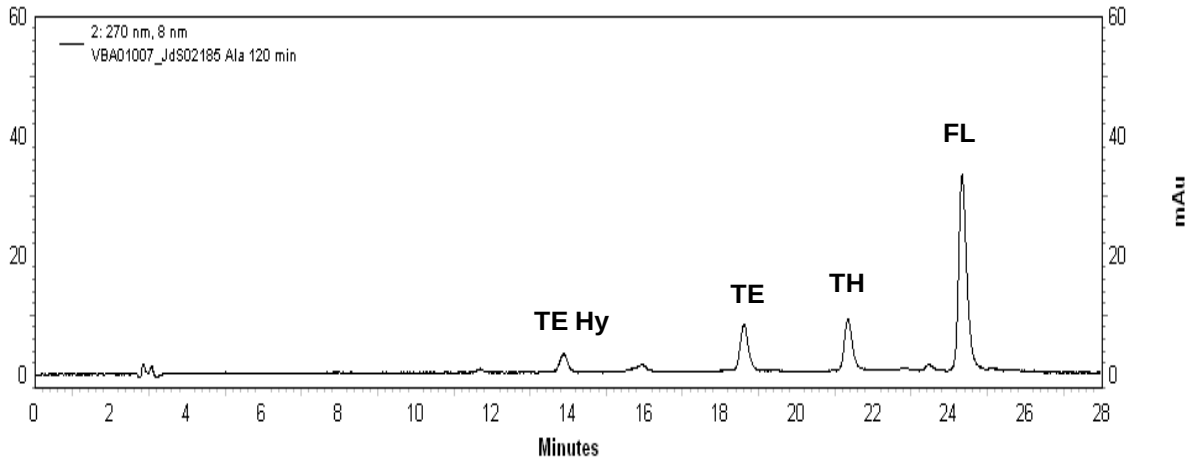


Fig. 6.28. BTE assay of the alanine mutant **88**.

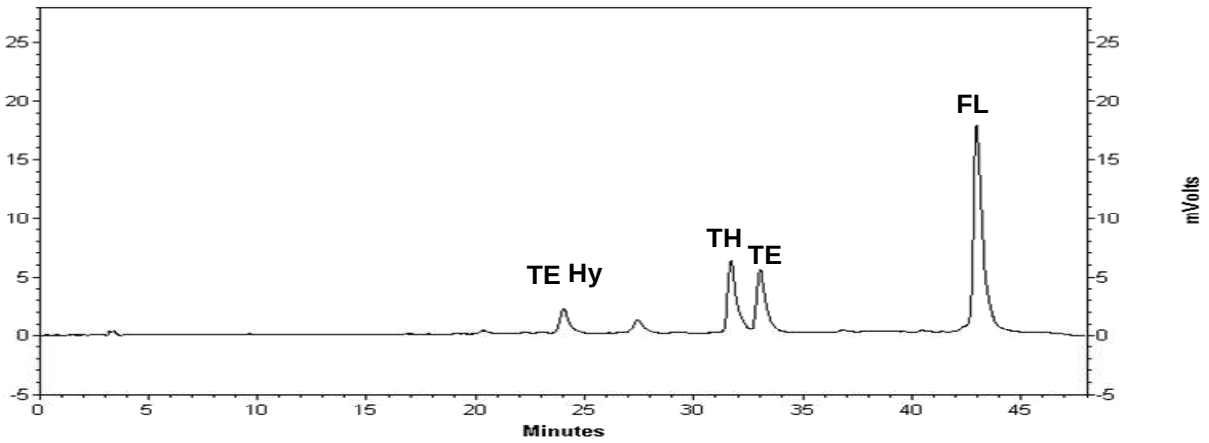


Fig. 6.29. BTE assay of the glycine mutant **89**.

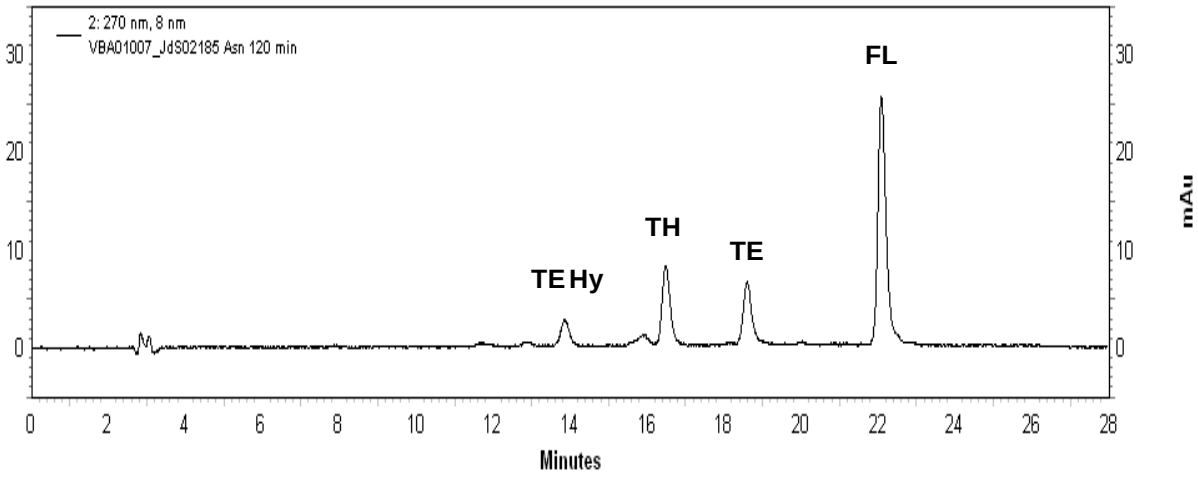


Fig. 6.30. BTE assay of the asparagine mutant **90**.

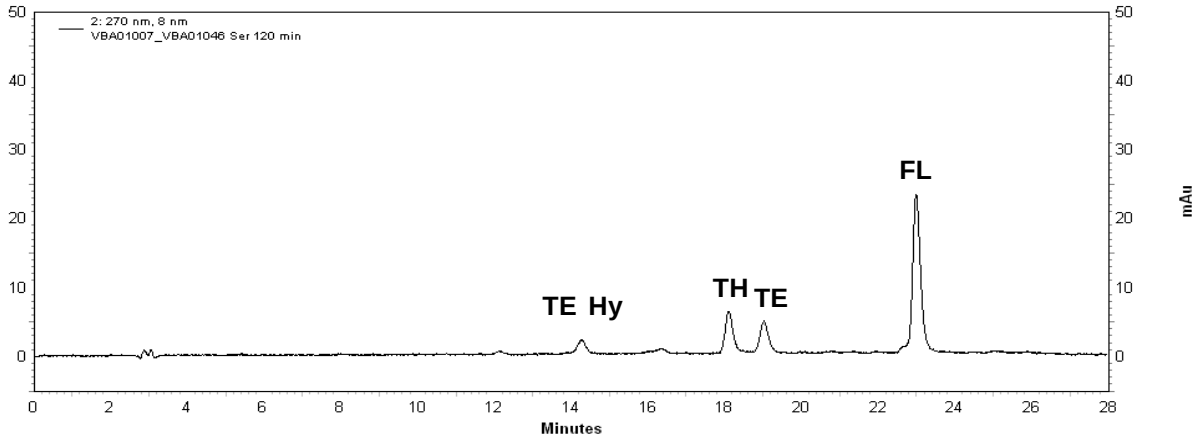


Fig. 6.31. BTE assay of the serine mutant **91**.

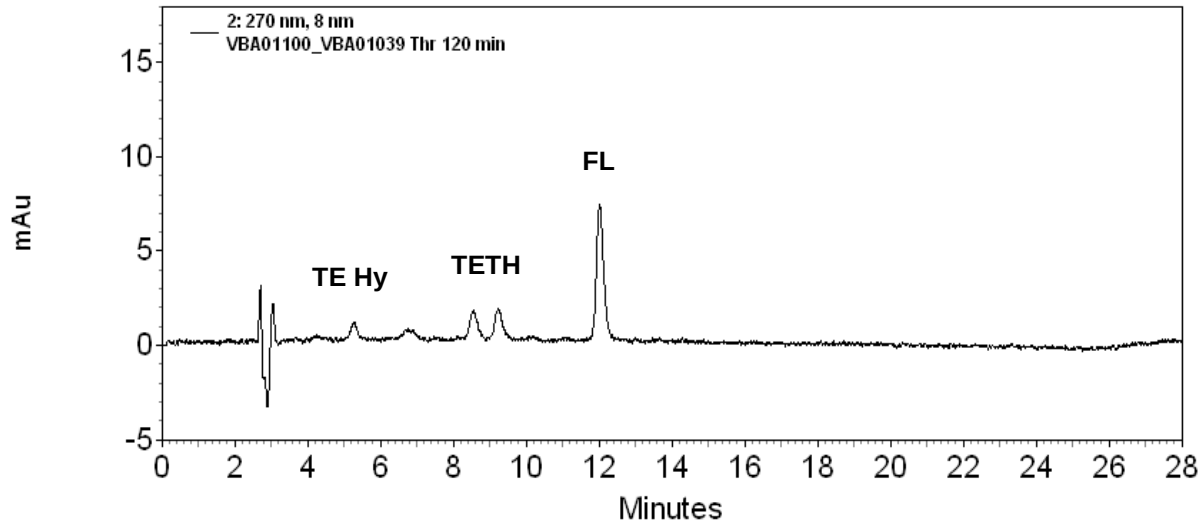


Fig. 6.32. BTE assay of the threonine mutant **92**.

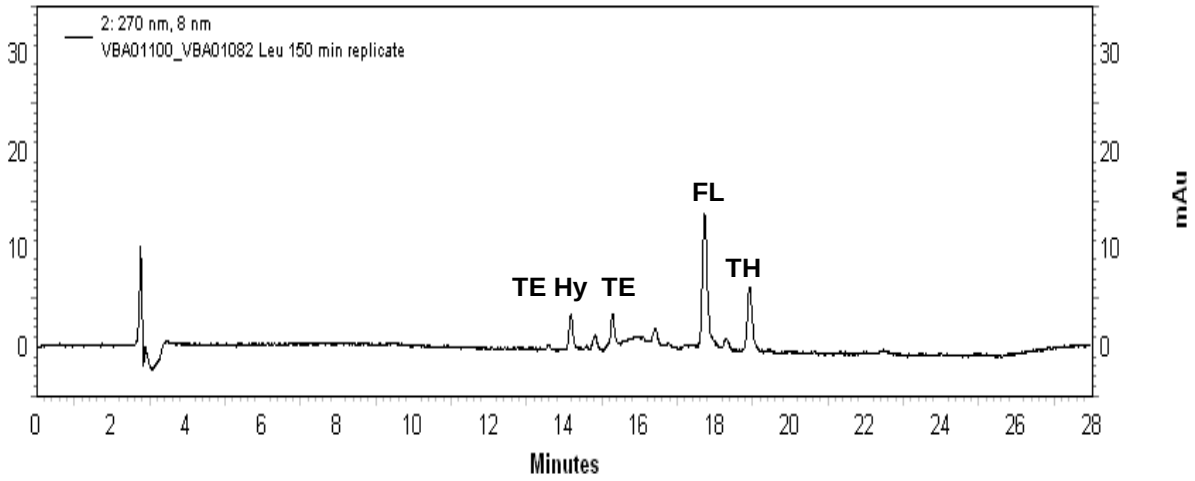


Fig. 6.33. BTE assay of the leucine mutant **93**.

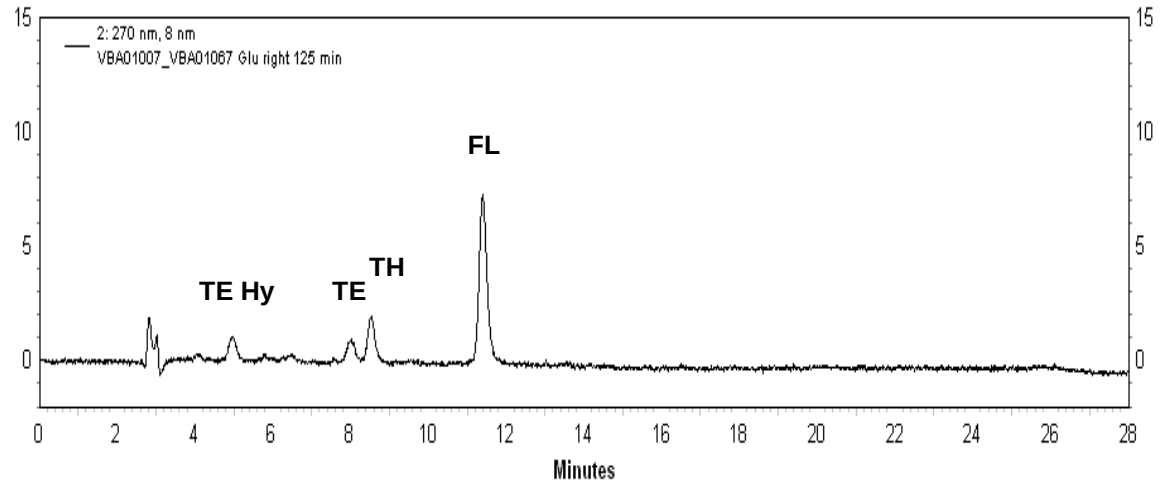


Fig. 6.34. BTE assay of the glutamic acid mutant **96**.

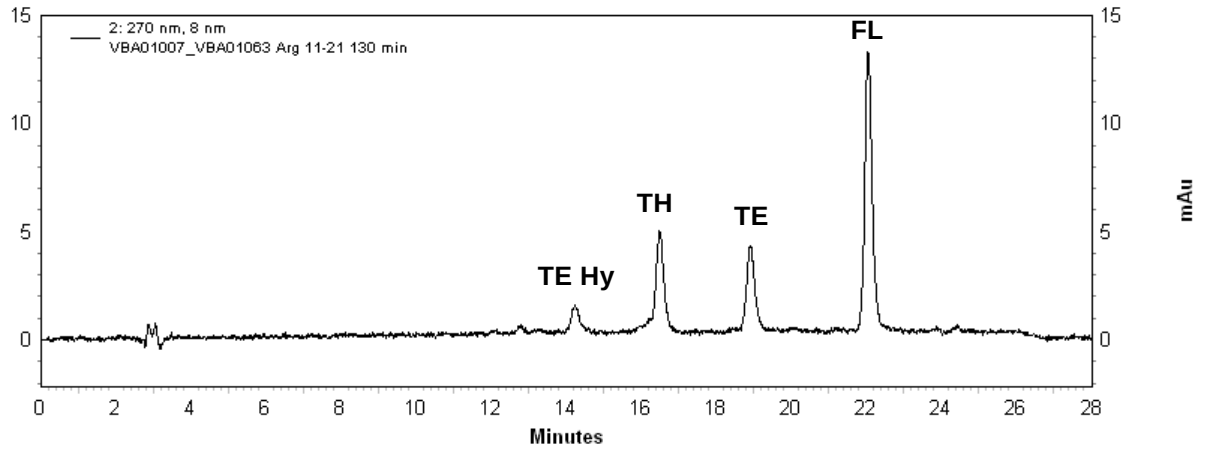


Fig. 6.35. BTE assay of the arginine mutant **95**.

There is a by-product forming from the thioester in buffer solution, which was analyzed by MALDI and HPLC after one day, and corresponds to a cyclic product. In fact a loss of 106 uma respect to the thioester was observed, which corresponds to the loss of methylthioglycolate and of a proton.

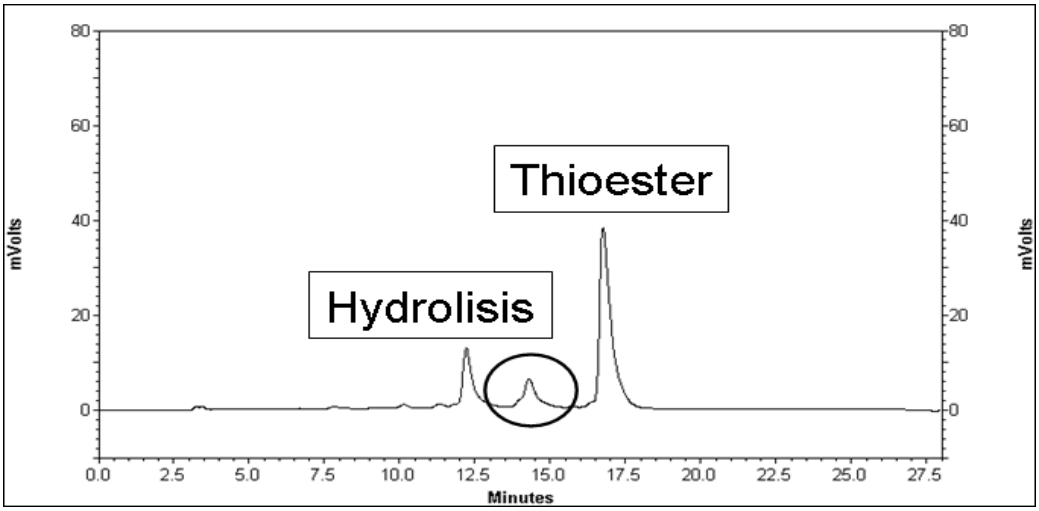


Fig. 6.36. HPLC trace of the thioester **85** dissolved in the buffer solution, injected the day after.

In summary ΔG data obained from the BTE assays are reported in **Table 6.3**.

ΔG_{BTE} (kcal/mol)												
	Run # 1	Run # 2	Run # 3	Run # 4	Run # 5	Run # 6	Run # 7	Run # 8	Run # 9	Run # 10	Ave-rage	ST-DEV
85 + **86** (**Ile**)	-1.31	-1.43	-1.42	-1.39	-1.32	-1.47	-1.44				-1.41	0.05
85 + **87** (**Val**)	-1.54	-1.46	-1.39	-1.48	-1.33						-1.44	0.08
85 + **88** (**Ala**)	-0.47	-0.62	-0.53	-0.52	-0.56	-0.54	-0.68	-0.77	-0.73	-0.62	-0.60	0.10
85 + **89** (**Gly**)	-0.06	-0.06	-0.04	-0.24	-0.19	-0.18					-0.13	0.08
85 + **90** (**Asn**)	-0.05	-0.13	-0.25	-0.16	-0.35	-0.38	-0.22	-0.19			-0.22	0.11
85 + **91** (**Ser**)	-0.75	-0.73	-0.76	-0.68							-0.73	0.04
85 + **92** (**Thr**)	-0.73	-0.79	-0.88	-0.73							-0.78	0.07
85 + **93** (**Leu**)	-0.49	-0.38	-0.54	-0.62	-0.62						-0.53	0.10
85 + **93** (**Glu**)	-0.97	-1.05	-1.02								-1.00	0.04
85 + **95** (**Arg**)	-0.05	-0.15	-0.09	-0.04							-0.06	0.08

Table 6.3. Summary of BTE assays data.

6.5 References

1. Smith, C. K.; Regan, L. *Acc. Chem. Res.* **1997**, *30*, 153-161.

2. Gellman, S. H. *Current Opinion in Chemical Biology* **1998**, *2*, 717-725.

3. Somers, W. S.; Phillips, S. E. V. *Nature* **1992**, *359*, 387-393.

4. Puglisi, J. D.; Chen, L.; Blanchard, S.; Frankel, A. D. *Science* **1995**, *270*, 1200-1203.

5. Derrick, J. P.; Wigley, D. B. *Nature* **1992**, *359*, 752-754.

6. Tan, S. Y.; Pepys, M. B. *Histopathology* **1994**, *25*, 403-414.

7. Sibanda, B. L.; Thornton, J. M. *Nature* **1986**, *316*, 170-174.

8. Blanco F. J.; Jimenez, M. A.; Herranz, J.; Rico, M.; Santoro, J.; Nieto, J. L. *J. Am. Chem. Soc.* **1993**, *115*, 5887-5888.

9. Hecht, M. H. *PNAS* **1994**, *91*, 8729-8730.

10. Salemme, F. R. *Prog. Biophys. Mol. Biol.* **1983**, *42*, 95-133.

11. Mossing, M. C.; Sauer, R. T. *Science* **1990**, *250*, 1712-1715.

12. Albright, R. A.; Mossing, M. C.; Matthews, B. W. *Biochemistry* **1996**, *35*, 735-742.

13. a) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 584-590; b) Neira, J. L.; Fersht, A. R. *Folding Des.* **1996**, *1*, 231-241; c) Viguera, A. R.; Jimenez, M. A.; Rico, M.; Serrano, L. *J. Mol. Biol.* **1996**, *255*, 507-521; d) Searle, M. S.; Zerella, R.; Williams, D. H.; Packman, L. C. *Protein Eng.* **1996**, *9*, 559-565.

14. Searle, M. S. *J. Chem. Soc., Perkin Trans.* **2001**, *2*, 1011-1020.

15. a) Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. *J. Mol. Biol.* **1989**, *206*, 759-777; b) Hutchinson, E. G.; Thornton, J. M. *Protein Sci.* **1994**, *3*, 2207-2216.

16. Venkatachalam, C. M. *Biopolymers* **1968**, *6*, 425-1436.

17. Fasman G. D. Plenum Press: New York. **1989**, 193-316.

18. Haque, T. S.; Gellman, S. H. *J. Am. Chem. Soc.* **1997**, *119*, 2303-2304.

19. a) Ramirez-Alvarado, M.; Blanco, F. J.; Serrano, L. *Nat. Struct. Biol.* **1996**, *3*, 604-612; b) de Alba, E.; Jimenez, M. A.; Rico, M. *J. Am. Chem. Soc.* **1997**, *119*, 175-183; c) Griffiths-Jones, S. R.; Maynard, A. J.; Searle, M. S. *J. Mol. Biol.* **1999**, 292, 1051-1069; d) Ramirez-Alvarado, M.; Blanco, F. J.; Niemann, H.; Serrano, L. *J. Mol. Biol.* **1997**, 273, 898-912.

20. Stanger, H. E.; Faisal, A. S.; Espinosa, J. F.; Giriat, I.; Muir, T.; Gellman, S. H. *PNAS* **2001**, *98*, 12015-12020.

21. a) Scholtz, J. M.; Qian, H.; York, E. J.; Stewart, J. M; Baldwin, R. L. *Biopolymers* **1991**, *31***,** 1463-1470; b) Rohl, C. A.; Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L.

1992, *Biochemistry, 31*, 1263-1269; c) Zimm, B. H.; Doty, P.; Iso, K. *PNAS* **1959**, *45*, 1601-1607.

22. a) Qian, H.; Schellman, J. A. *J. Phys. Chem.* **1992**, *96*, 3987-3994; b) Chakrabartty, A.; Baldwin, R. L. *Adv. Protein Chem.* **1995**, *46*, 141-176.

23. Engel, J.; Schwarz, G. *Angew. Chem. Int. Ed.* **1970**, *9*, 389-400.

24. Wang, X.; Espinosa, J. F.; Gellman, S. H. *J. Am. Chem. Soc.* **2000**, *122*, 4821-4822.

25. Schenck, H. L.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4869-4870.

26. a) Rohl, C. A.; Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biochemistry* **1992**, *31*, 1263-1269; b) Zimm, B. H.; Doty, P.; Iso, K. *PNAS* **1959**, *45*, 1601-1607.

27. a) Avbelj, F.; Moult, J. *Biochemistry* **1995**, *34*, 755-764; b) Avbelj, F.; Fele, L. *J. Mol. Biol.*, **1998**, *279*, 665-684.

28. Searle, M. S.; Griffiths-Jones, S. R.; Skinner-Smith, H. *J. Am. Chem. Soc.* **1999**, *121*, 11615-11620.

29. de Alba, E.; Blanco, F. J.; Jimenez, M. A.; Rico, M.; Nieto, J. L. *Eur. J. Biochem.* **1995**, *233*, 283-292.

30. a) Wouters, M. A.; Curmi, P. M. G. *Proteins: Struct., Funct., Genet.* **1995**, *22*, 119-131; b) von Heijne, G.; Blomberg, C. *J. Mol. Biol.* **1977**, *117*, 821-824; c) Lifson, S.; Sander, C. *J. Mol. Biol.* **1980**, *139*, 627-639.

31. Yang, A.; Honig, B. *J. Mol. Biol.* **1995**, *252*, 366-376.

32. Hubbard, T. J. *Proceedings of the biotechnology computing track, protein structure prediction minitrack of the 27th HICSS*; IEEE Computer Society Press: Los Alamitos, CA, **1994**, 336-354.

33. Smith, C. K.; Regan, L. *Science* **1995**, *270*, 980-982.

34. Swindells, M. B.; Macarthur, M. W.; Thornton, J. M. *Nat. Struct. Biol.* **1995**, *2*, 596-603.

35. a) Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. *J. Phys. Chem.* **1996**, *100*, 2661-2666; b) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; Macarthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494-506.

36. Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13*, 211-222.

37. a) Kemp, D. S. *Tibtech* **1990**, *8*, 249-255; b) Mutter, M.; Altmann, K.-H. *Int. J. Pept. Protein Res.* **1985**, *26*, 373-380.

38. Scheraga, H. A. *Pure Appl. Chem.* **1978**, *50*, 315-324.

39. a) Kim, C. A.; Berg, J. M. *Nature* **1993**, *362*, 267-270; b) Kim, C. A.; Berg, J. M. *Nat. Struct. Biol.* **1996**, *3*, 940-945.

40. Carson, M. *J. Appl. Crystallogr.* **1990**, *24*, 958-961.

41. a) Smith, C. K.; Withka, J. M.; Regan, L. *Biochemistry* **1994**, *33*, 5510-5517; b) Minor, D. L. Jr.; Kim, P. S. *Nature* **1994**, *367*, 660-663; c) Regan, L. *Curr. Biol.* **1994**, *4*, 656-658.

42. a) Munoz, V.; Serrano, L. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 301-311; b) Finkelstein, A. V. *Protein Eng.* **1995**, *8*, 207-209.

43. Minor, D. L. Jr.; Kim, P. S. *Nature* **1994**, *371*, 264-267.

44. a) Garratt, R. C.; Thornton, J. M.; Taylor, W. R. *FEBS* **1991**, *280*, 141-146; b) Finkelstein, A. V.; Reva, B. A. *Nature* **1991**, *351*, 497-499.

45. a) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266-280; b) Pace, C. N.; Shaw, K. L. *Proteins, Suppl.* **2000**, *4*, 1-7.

46. a) Hvidt, A.; Nielsen, S. O. *Adv. Prot. Chem.* **1966**, *21*, 287-386; b) Woodward, C.; Carulla, N.; Barany, G. *Methods Enzymol.* **2004**, *380*, 379-400.

47. a) Rowan, S. J.; Cantrill, S. J.; Cousins, G. R. L.; Sanders, J. K. M.; Stoddart, J. F. *Angew. Chem. Int. Ed.* **2002**, *41*, 898-952; b) Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J. L.; Sanders, J. K. M.; Otto, S. *Chem. Rev.* **2006**, *106*, 3652-3711.

48. Woll, M. G.; Gellman, S. H. *J. Am. Chem. Soc.* **2004**, *126*, 11172-11174.

49. a) Zacharias, D. E.; Murrayrust, P.; Preston, R. M.; Glusker, J. P. *Arch. Biochem. Biophys.* **1983**, *222*, 22-34; b) Pawar, D. M.; Khalil, A. A.; Hooks, D. R.; Collins, K.; Elliott, T.; Stafford, J.; Smith, L.; Noe, E. A. *J. Am. Chem. Soc.* **1998**, *120*, 2108-2112.

50. Phillips, S. T.; Piersanti, G.; Bartlett, P. A. *PNAS* **2005**, *102*, 13737-13742.

51. Hadley, E. B.; Witek, A. M.; Freire, F.; Peoples, A. J.; Gellman, S. H. *Angew. Chem. Int. Ed.* **2007**, *46*, 7056-7059.

52. Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *PNAS* **2001**, *98*, 5578-5583.

53. a) Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *20*, 4236-4237; b) Stanger, H. E.; Syud, F. A.; Espinosa, J. F.; Giriatt, I.; Muir, T.; Gellman, S. H. *PNAS* **2001**, *98*, 12015-12020.

54. Good, N. E.; Winget, G. D.; Winter, W.; Connolly, T. N.; Izawa, S.; Singh, R. M. M. *Biochemistry* **1966**, *5*, 467-477.

55. Sharman, G. J.; Searle, M. S. *Chem. Comm.* **1997**, 1955-1956.

56. Fooks, H. M.; Martin, A. C. R.; Woolfson, D. N.; Sessions, R. B.; Hutchinson, E. G. *J. Mol. Biol.* **2006**, *356*, 32-44.

57. a) Kenner, G. W.; McDermot. J. R.; Sheppard, R. C. *J. Chem. Soc. Chem. Comm.* **1971**, 636-637; b) Backes, B. J.; Ellman, J. A. *J. Org. Chem.* **1999**, *64*, 2322-2330; c) Backes, B. J.; Virgilio, A. A.; Ellman, J. A. *J. Am. Chem. Soc.* **1996**, *118*, 3055-3056.

58. Ingenito, R.; Bianchi, E.; Fattori, D.; Pessi, A. *J. Am. Chem. Soc.* **1999**, *121*, 11369-11374.

59. Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121*, 11684-11689.

60. a) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. *J. Am. Chem. Soc.* **2004**, *126*, 7238-7243; b) Fesinmeyer, R. M.; Hudson, F. M.; White, G. W. N.; Olsen, K. A.; Euser, A.; Andersen, N. H. *J. Biomol. NMR* **2005**, *33*, 213-231; c) Olsen, K. A.; Fesinmeyer, R. M.; Stewart, J. M.; Andersen, N. H. *PNAS* **2005**, *102*, 15483-15487.

61. Price, J. L.; Hadley, E. B.; Steinkruger, J. D.; Gellman, S. H. *Angew. Chem.* **2010**, *122*, 378-381.

62. Woll, M. G.; Hadley, E. B.; Mecozzi, S.; Gellman, S. H. *J. Am. Chem. Soc.* **2006**, *128*, 15932-15933.