The Pro Region Required for Folding of Carboxypeptidase Y Is a Partially Folded Domain with Little Regular Structural Core†

Poul Sørensen,‡ Jakob R. Winther,§ Niels C. Kaarsholm,∥ and Flemming M. Poulsen*,‡

Department of Chemistry and Department of Yeast Genetics, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby, Copenhagen, Denmark, and Novo Research Institute, Novo Allé, DK-2880 Bagsvaerd, Copenhagen, Denmark

Received July 7, 1993⊗

ABSTRACT: The pro region of carboxypeptidase Y (CPY) from yeast is necessary for the correct folding of the enzyme [Winther, J. R., & Sørensen P. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9330–9334]. Using fluorescence, circular dichroism, and heteronuclear NMR analyses, it is demonstrated that the isolated pro region is a partially folded protein domain under the conditions where it is functional. It is characterized by a relatively high content of secondary structural elements but a very low content of defined tertiary structure. Although these characteristics are reminiscent of the compact denatured states that have been identified as intermediates in protein folding ("molten globules"), the pro region exhibits only very weak binding of the hydrophobic probe 1-anilino-8-naphthalenesulfonate, and it is resistant toward complete thermal unfolding. Altogether the data indicate an extremely flexible structure that has little regular structural core. It is proposed that the feature of a partially folded domain *per se* is important for the function of the pro region of CPY as a "co-translational chaperone".

Most prokaryotic and eukaryotic proteases are synthesized as inactive precursors that are normally activated to the mature and functional state upon cleavage of an N-terminal extension, the pro region (Neurath, 1989). The pro region may also specify the cellular targeting address of the cognate protease (Klionsky *et al.*, 1990). Recent *in vitro* studies show that the pro region can also be necessary for the correct refolding of the protease. In the presence of the pro region, a substantial activation barrier against both folding and unfolding of the protease is alleviated (Zhu *et al.*, 1989; Winther & Sørensen, 1991; Baker *et al.*, 1992). In some cases, this function is not restricted to the intact zymogen but can also be observed when pro regions are supplied independently of their protease partner in a bimolecular reaction. This phenomenon can be observed both *in vitro* (Ohta *et al.*, 1990; Baker *et al.*, 1992; Eder *et al.*, 1993) and *in vivo* (Silen & Agard 1989; Fabre *et al.*, 1992). In the absence of the pro region, the protease alone can be stable as an inactive intermediate when diluted from concentrated solutions of denaturant (Baker *et al.*, 1992; Eder *et al.*, 1993). Considering the intimate relationship in the cell between protein folding and transport, the fact that pro region dependent secretion *in vivo* has been reported for a number of otherwise unrelated proteases and prohormones [e.g., see Ikemura *et al.* (1987), Pohler *et al.* (1987), Silen *et al.* (1989), Gray and Mason (1990), Suter *et al.* (1991), and Nebes and Jones (1991)] suggests that the mechanism of pro region dependent folding may have evolved convergently.

In an earlier report, we showed that the pro region of carboxypeptidase Y (CPY)[1] from yeast is necessary for the correct and efficient folding of the precursor and for keeping the proenzyme inactive (Winther & Sørensen, 1991). In a further dissection of the folding pathway of CPY and its precursor, we have proceeded with structural studies of the pro region of CPY. The pro region of CPY has now been made available for high-resolution structural studies by high-level expression as an autonomous folding unit in *Escherichia coli*.

MATERIALS AND METHODS

Materials. Deoxynucleotides were synthesized on an Applied Biosystems Model 380A oligonucleotide synthesizer (USA). NAP-prepacked Sephadex G-25 gel filtration columns were from Pharmacia (Sweden); 99% ^{15}N-labeled $(NH_4)_2SO_4$ was from Cambridge Isotope Laboratories (USA). ANS was from Sigma (USA).

The following strains of *E. coli* were used: K38 (Tabor, 1990) was a gift from S. Tabor, and BL21(DE3) (Studier & Moffatt, 1986) was obtained from Novagen (USA). Plasmids pT7-7 and pGP1-2 (Tabor, 1990) were gifts from S. Tabor (Harvard Medical School), and pLysS (Studier, 1991) was from Novagen.

Recombinant DNA Techniques. Standard recombinant DNA techniques were carried out according to Sambrook *et al.* (1989).

Construction of a Pro Region Expression Cassette. For the expression of the pro region of CPY (CYPR) in *E. coli*, the T7 RNA polymerase/promoter system was chosen (Tabor, 1990). In this system, the gene of interest is placed on a plasmid under the control of a T7 RNA polymerase promoter, and the gene for T7 RNA polymerase either may be present on a second plasmid under the control of a heat-inducible λ

† This work was supported by the Danish Protein Engineering Research Centre.

* To whom correspondence should be addressed.
‡ Department of Chemistry, Carlsberg Laboratory.
§ Department of Yeast Genetics, Carlsberg Laboratory.
∥ Novo Research Institute.

⊗ Abstract published in *Advance ACS Abstracts,* October 15, 1993.

[1] Abbreviations: ANS, 1-anilino-8-naphthalenesulfonate; CD, circular dichroism; CYPR, pro region of CPY; CPY, carboxypeptidase Y; EDTA, ethylenediaminetetraacetic acid; FPLC, fast protein liquid chromatography; Gu-HCl, guanidine hydrochloride; HEPES, *N*-(2-hydroxyethyl)-piperazine-*N*′-2-ethanesulfonic acid; HSQC, ^1H–^{15}N heteronuclear single quantum correlation; IPTG, isopropyl β-D-thiogalactopyranoside; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; PCR, polymerase chain reaction; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; TFE, trifluoroethanol; TOCSY, total correlation spectroscopy.

0006-2960/93/0432-12160$04.00/0

promoter, enabling harsh expression conditions (Tabor & Richardson, 1985), or may be integrated on the chromosome under the control of an IPTG-inducible lacUV5 promoter, enabling very mild expression conditions (Studier & Moffat, 1986). We have found that by choosing the appropriate strain/plasmid configuration and growth conditions this system can be used to investigate the properties of a given protein with regard to its *in vivo* folding, transport, and stability.[2] An expression cassette designed to express the CPY pro region (CYPR) as an independent protein domain was constructed by tailoring the sequence encoding proCPY (Valls *et al.*, 1987). Appropriate restriction enzyme sites for directional cloning into the pT7-7 expression vector (Tabor, 1990) as well as start and stop signals for protein synthesis were introduced. Utilizing the PCR reaction, a *Nde*I restriction site (containing the methionine start codon) was inserted at the 5′ end of the sequence, and an *Sal*I site and the TAA termination codon were inserted at the 3′ end of the sequence using the synthetic oligonucleotides 5′-GGGGCATATGATCTCATTGCAAA-GA-3′ (sense) and 5′-GGGGGTCGACTTAGTTGACAC-GAAGCTG-3′ (antisense). The resulting fragment was then inserted into the *Nde*I and *Sal*I sites of pT7-7, yielding the construct pT7-CYPR. The recombinant 92 amino acid residue translation product starts with the introduced methionine and is followed by the first amino acid in the pro region of proCPY.

Expression. For high-level expression of CYPR into insoluble inclusion bodies, the pT7-CYPR construct was transformed into K38 cells containing the pGPI-1 plasmid (Tabor, 1990). The cells were grown at 30 °C in Terrific Broth medium (Sambrook *et al.*, 1989) supplemented with 100 μg/L kanamycin and 100 μg/L ampicillin until a culture density with an OD_{600} of 4 was obtained. At this point, CYPR expression was induced by incubating the flasks in a 70 °C water bath with vigorous shaking until a temperature of 44 °C was reached. The flasks were then transferred back to the incubator set at 44 °C, and induction was continued for 4 h. For the purpose of obtaining soluble pro region, the pT7-CYPR expression plasmid was transformed into the BL(DE3) strain containing the pLysS plasmid (Studier, 1991). Shaking flask cultures of BL21(DE3)/pLysS/pT7-CYPR were grown at 37 °C in M9ZB (per liter: 1 g of NH_4Cl, 3 g of KH_2PO_4, 6 g of Na_2HPO_4, 5 g of NaCl, 1 mL of 1 M $MgSO_4$, 4 g of glucose, and 10 g of tryptone) supplemented with 100 μg/L ampicillin and 25 μg/L chloramphenicol. CYPR expression was induced at an OD_{600} of 0.8 by the addition of IPTG to 0.4 mM, and growth was continued for 4 h.

Purifications. The pro region was purified from the soluble cytoplasmic fraction. Pellets of cells from expression of pT7-CYPR in BL21(DE3)/pLysS were lysed by three rounds of freeze–lysis aided by the low level of constitutive expression of T7 lysozyme directed by the pLysS plasmid (Studier, 1991). DNase treatment was performed by adjusting the $MgCl_2$ concentration to 10 mM, and the $MnCl_2$ concentration was adjusted to 1 mM before addition of DNase I to a final concentration of 10 μg/mL and incubation for 20 min at room temperature. Subsequently, the cell debris was removed by centrifuging at 10000*g* for 20 min at 4 °C. The cell lysate was applied on a Mono Q anion-exchange column equilibrated in 50 mM HEPES/1 mM EDTA, pH 8.0, and FPLC was carried out by using a linear gradient from 0 to 0.5 M NaCl. CYPR-containing fractions were pooled, and contaminating high molecular weight proteins were removed by gel filtration on a Sephadex G-50 column. The peak containing pure CYPR was identified by SDS–PAGE and its identity verified by N-terminal sequencing and amino acid analysis.

Refolding of CPY by Addition of Isolated CYPR. The pro region mediated refolding of unfolded CPY was carried out by adding 2.9 μM CYPR to a molar equivalent of a stable folding intermediate of CPY in 5 mM HEPES/1 mM EDTA, pH 7.5, with or without the addition of salts at increasing concentrations. Following 16-h incubation at 25 °C, the extent of renaturation was determined by measuring the enzyme activity as previously described (Winther & Sørensen, 1991). The stable folding intermediate of CPY was generated by denaturation for 1 h at 25 °C in 6 M Gu-HCl, 10 mM HEPES, and 1 mM EDTA, pH 7.5, followed by removal of the denaturant by desalting twice on NAP column into 5 mM HEPES/1 mM EDTA, pH 7.5.

Fluorescence Experiments. Fluorescence spectra were recorded on a Perkin-Elmer LS 50 fluorometer at 25 °C. The slit width for both excitation and emission corresponded to 5 nm, and the scan rate was 240 nm/min.

Circular Dichroism. CD spectra were recorded on a Jobin Yvon Model V dichrograph calibrated with (+)-10-camphorsulfonic acid. All spectra were recorded at 25 °C. Spectra in the far-UV range (180–260 nm) were recorded with a cell path length of 0.01 cm. Near-UV CD samples were scanned between 260 and 350 nm using a 0.5-cm path length. Typical protein concentrations were about 0.5 mg/mL. The spectra were smoothed with a Fourier transform algorithm, and the appropriate background spectra were subtracted. The result, $\Delta\epsilon$, was based on the molar concentration of the peptide bond (Johnson, 1990).

NMR Spectroscopy. In order to carry out NMR spectroscopy, the CYPR samples in HEPES buffer were exchanged with 0.1 M NH_4HCO_3 by gel filtration, freeze-dried, and dissolved in the appropriate buffer according to the experiment. The ^1H chemical shifts were measured in parts per million relative to the methyl group proton resonance of 3-(trimethylsilyl)propionic-d_4 acid, and the ^{15}N chemical shifts were measured in parts per million relative to the nitrogen resonance of NH_3. The spectra were recorded at 25 °C using a Bruker AMX600 spectrometer.

RESULTS

Construction and Expression of the Recombinant CPY Pro Region. In an attempt to facilitate the subsequent purification, we initially chose to express the recombinant pro region under conditions that usually favor formation of insoluble inclusion bodies as this represents a significant purification step in itself. However, even under extreme conditions (high protein expression under heat shock conditions) previously described to result in high amounts of inclusion bodies (Strandberg & Enfors, 1988), almost all CYPR was in the soluble cytoplasmic fraction, and only a few percent were in inclusion bodies (data not shown). Thus, *in vivo*, CYPR exhibits kinetic partitioning to a state that is soluble and stable toward digestion by endogenous proteases (Mitraki *et al.*, 1991).

Previously, it was demonstrated that the refolding of CPY in the absence of the pro region could be induced to a low level of activity by salts and the efficiency of this reaction (Winther & Sørensen, 1991) follows the Hofmeister series [see, e.g., von Hippel and Wong (1965)]. This dependency is also observed for the bimolecular reaction between the isolated pro region and the enzyme. The effect of CYPR is only manifested when added in the presence of salt (data not shown; Winther *et al.*, in preparation). The structural background

[2] P. Sørensen, unpublished data.

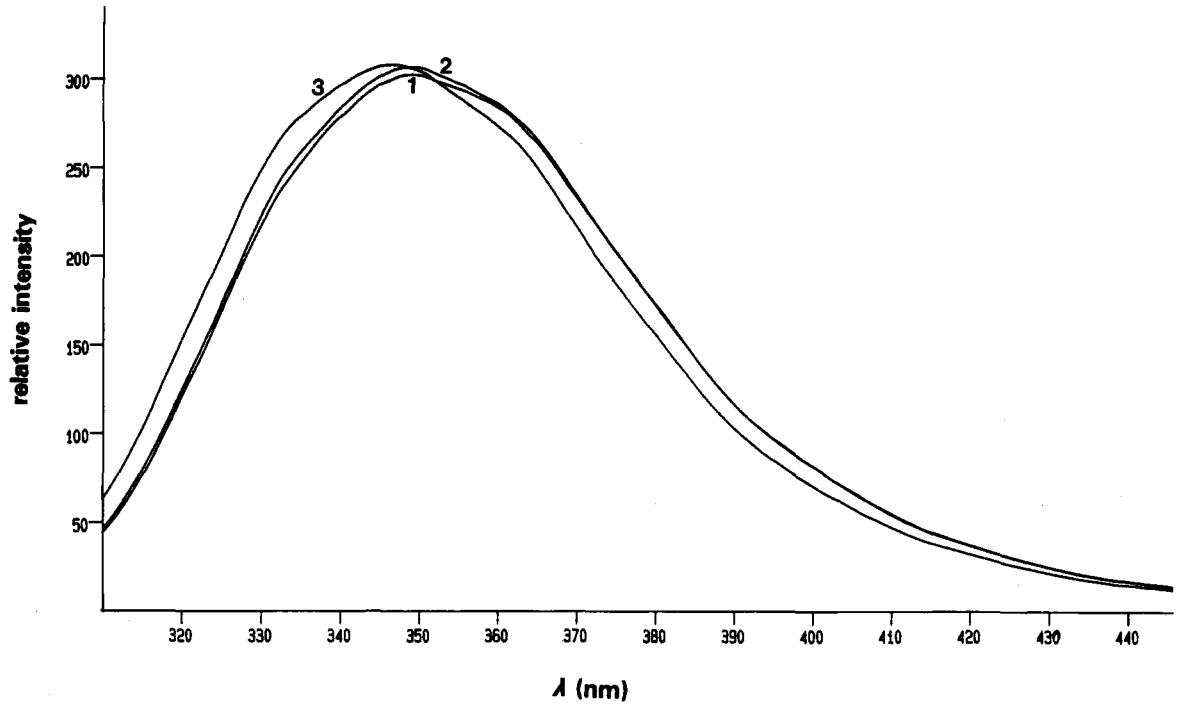

FIGURE 1: Fluorescence emission spectra of CYPR as a function of salt. CYPR at 5 μM was incubated in 10 mM HEPES/1 mM EDTA, pH 7.5, and the fluorescence was determined in the absence (trace 1) or in the presence of 0.15 M (trace 2) or 1.5 M NaCl (trace 3). Excitation was carried out at 295 nm, and emission was recorded between 300 and 500 nm.

for this observation was investigated by studying the fluorescence characteristics of CYPR in the presence or absence of salts. Increasing the ionic strength to 1.5 M NaCl resulted in a slight blue-shift of the λ_{max} from 349 to 346 nm (Figure 1) as an indication of the two tryptophan residues of CYPR changing positions from being nearly fully solvent-exposed, as in an unfolded protein, to being only partially buried in a more hydrophobic environment [see, e.g., Horowitz and Xu (1992)]. Thus, the salt effect on the bimolecular reaction is not due to a major structural change in CYPR.

Specific binding of the fluorescent hydrophobic probe ANS (1-anilino-8-naphthalenesulfonate) is often used as an indication for the presence of protein folding intermediates, as ANS does not bind to the fully unfolded denatured state and normally not to native states either but has a high affinity for proteins in compact denatured states (Semisotnov *et al.*, 1991). This was confirmed for a stable folding intermediate of CPY which binds significant amounts of the probe (Figure 2). In the case of CYPR, however, only very little ANS binding was observed. This result is remarkable considering that CYPR contains 37% of aromatic and hydrophobic residues.

^1H and ^{15}N NMR spectra were recorded to obtain more specific information about the structure of CYPR under conditions that support folding activity. Addition of high amounts of salt leads to line broadening and a low signal-to-noise ratio. However, in accordance with the fluorescence data, the spectrum is not qualitatively changed. Thus, for convenience, the concentration of 0.15 M NaCl was chosen in the NMR experiments. The conformation of the peptide backbone was investigated with a ^1H–^{15}N heteronuclear single quantum correlation (HSQC) spectrum (Bodenhausen & Ruben, 1980) of the fully ^{15}N-labeled CYPR. The spectrum reveals resonances of protons bound to nitrogen, and as such, it is a sensitive probe for protein conformation. Thus, in proteins, peptide bond amides, side chain amides, and side chain guanidino nitrogen are typically observed. We used HSQC to identify the chemical shift dispersion of the amide hydrogen atoms (Figure 3). The spectrum shows a wider distribution of amide chemical shifts than is observed for random-coil peptides, most notably demonstrated by the H^N resonances observed in the regions both between 7.0 and 7.9 ppm and beyond 8.5 ppm (Figure 3). This indicates the presence of structured peptide bonds in CYPR (Wishart *et al.*, 1991; Neri *et al.*, 1992; Egan *et al.*, 1993).

The side chain conformation of CYPR was assessed with TOCSY and NOESY spectra. The TOCSY spectrum in the top panel of Figure 4 and the ^1H NMR spectrum in Figure 6 show that the ^1H resonances of the two tryptophan residues (W42 and W74) and one of the two tyrosine residues (Y49 or Y86) are shifted from random-coil chemical shift values (Wishart *et al.*, 1991). Although the ^1H NMR spectrum overall is similar to a random-coil spectrum, a closer examination of the spectrum reveals deviations from the random-coil chemical shifts for individual spin systems. In the aromatic region, six distinct spin systems can be identified: two of the benzenoic part of tryptophan, two of the phenolic part of tyrosine, and two of the imidazole part of histidine. There is no sequence-specific assignment of the spin systems, so we have designated the individual partners of the three pairs of residues A and B, respectively. The chemical shifts reported here are as follows: for Tyr^A, H^{δ} 7.18, H^{ϵ} 6.83; for Tyr^B, H^{δ} 7.19, H^{ϵ} 6.90; for Trp^A, $H^{\epsilon 3}$ 7.45, $H^{\zeta 2}$ 7.52, $H^{\zeta 3}$ 7.08, H^{η} 6.97; for Trp^B, $H^{\epsilon 3}$ 7.53, $H^{\zeta 2}$ 7.66, $H^{\zeta 3}$ 7.27, H^{η} 7.19; for His^A, $H^{\epsilon 1}$ 6.98, $H^{\delta 2}$ 7.83; for His^B, $H^{\epsilon 1}$ 7.07, $H^{\delta 2}$ 7.85. The resonances of the three phenylalanine residues are observed between 7.2 and 7.4 ppm.

The NOESY spectrum gives further evidence for a partially folded structure (Figure 4, bottom panel). On one hand, there are NOEs between Trp^A and Tyr^A resonances and resonances that could be from methyl groups, indicating the presence of tertiary structure. On the other hand, the NOESY spectrum is scarce in interresidue NOEs and backbone sequential effects. Given the size of CYPR and the relatively high content of secondary structure, the absence of strong NOEs indicates a high degree of dynamics in the peptide backbone with most side chains in flexible disordered positions. These two sets of

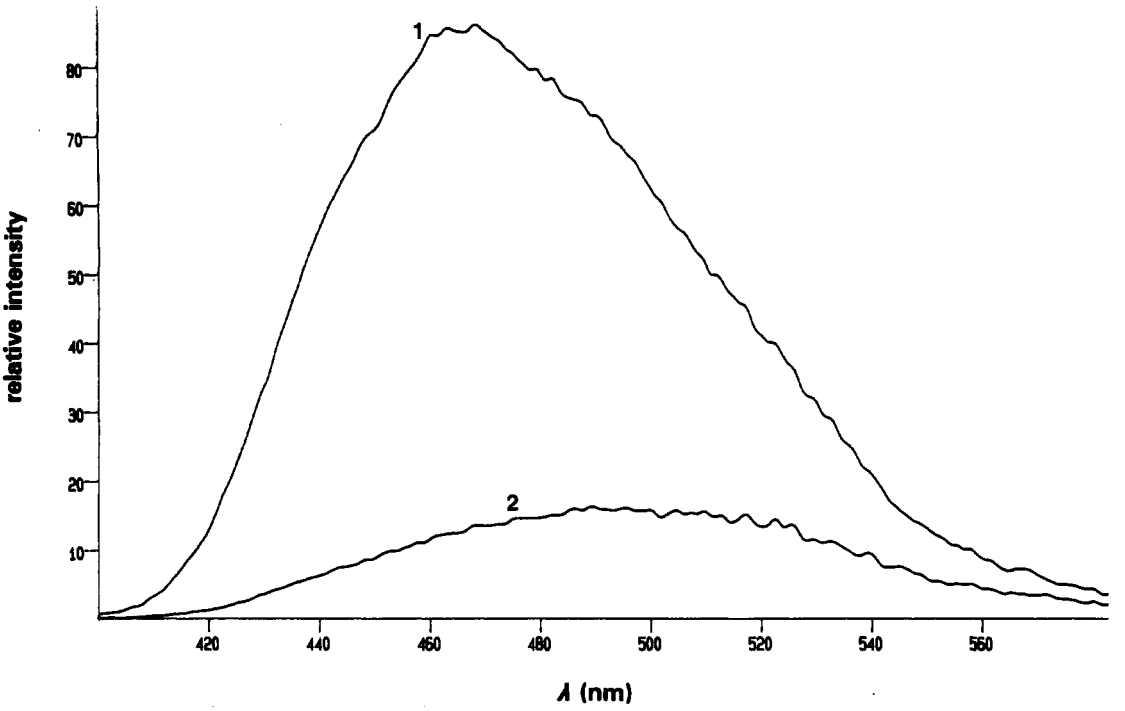

FIGURE 2: Fluorescence emission spectra of CYPR and CPY in the presence of the hydrophobic fluorescent probe ANS (1-anilino-8-naphthalenesulfonate). The proteins were incubated at 5 μM in 10 mM HEPES/1 mM EDTA, pH 7.5, and the fluorescence was determined. Trace 1, CYPR; trace 2, stable intermediate of CPY. The spectra are corrected for the background fluorescence of free ANS. The stable folding intermediate of CPY was generated as described under Materials and Methods. ANS was added to a final concentration of 250 μM, and the binding was determined. Excitation was carried out at 385 nm, and emission was recorded between 400 and 600 nm.

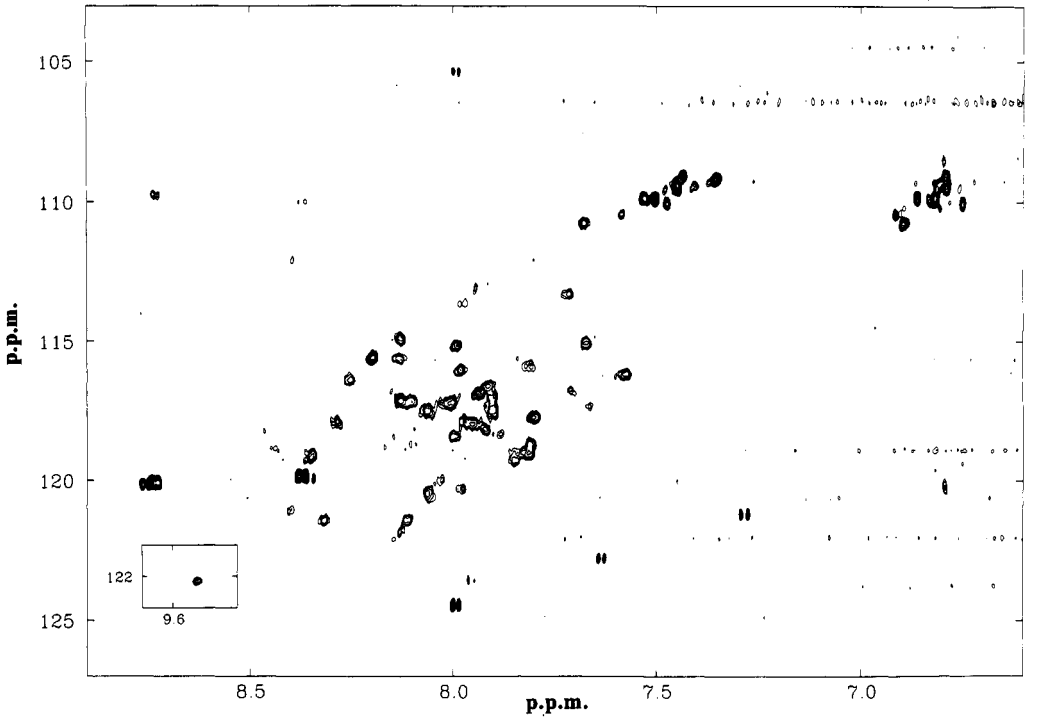

FIGURE 3: Heteronuclear single quantum correlation spectrum (HSQC) of the fully ^{15}N-labeled pro region of CPY. CYPR was uniformly labeled with ^{15}N by growing BL21(DE3)/pLysS cells on minimal medium with 99% ^{15}N-labeled $(NH_4)_2SO_4$ as the only nitrogen source. All other expression conditions were as described for obtaining soluble CYPR under Materials and Methods. The sample concentration was 0.3 mM in 20 mM phosphate buffer/0.15 M NaCl, pH 7.5.

information indicate that CYPR is a partially folded protein domain with only little rigid tertiary structure.

As a control, CD analysis was performed, and in agreement with the NMR data, far-UV CD analysis of CYPR predicted a relatively high secondary structure content even in the absence of any stabilizing salt (Figure 5a) while a near-UV CD spectrum showed little or no indications of folded tertiary structure (Figure 5b). Furthermore, the intermediate character of the structure was indicated by the fact that addition of 4 M Gu-HCl leads to a marked decrease of the far-UV CD spectrum in the 218–260-nm range and, conversely, addition of the structure-inducing agent trifluoroethanol leads to a significant increase in the far-UV CD signal (Figure 5a). A comparison with the near-UV CD spectra recorded under

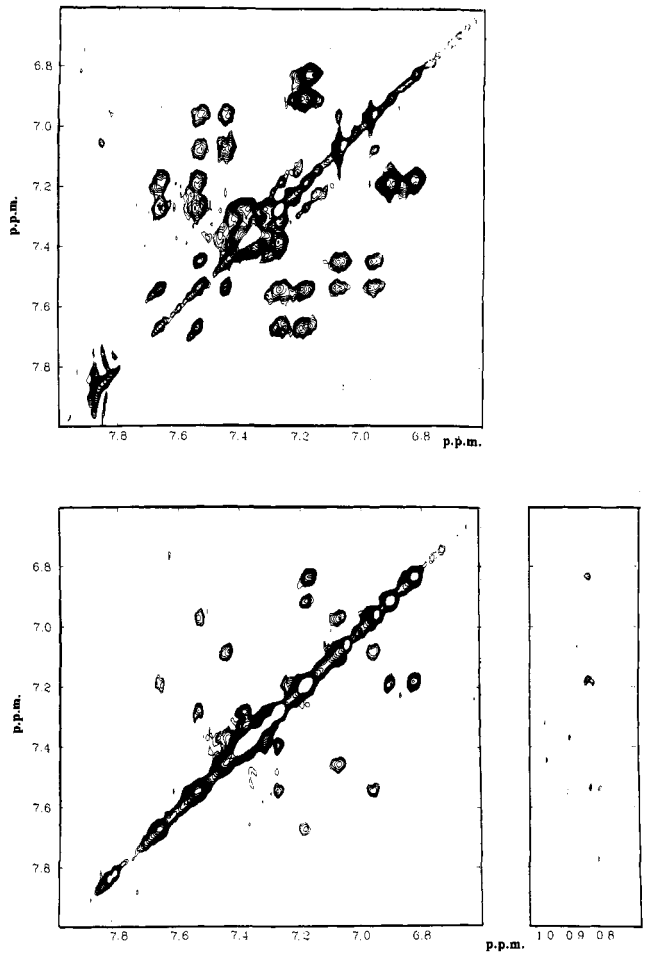

FIGURE 4: Aromatic region of contour plots of 1.4 mM CYPR in 20 mM phosphate buffer/0.15 M NaCl, pH 7.5. (Top) TOCSY spectrum. (Bottom) NOESY spectrum.

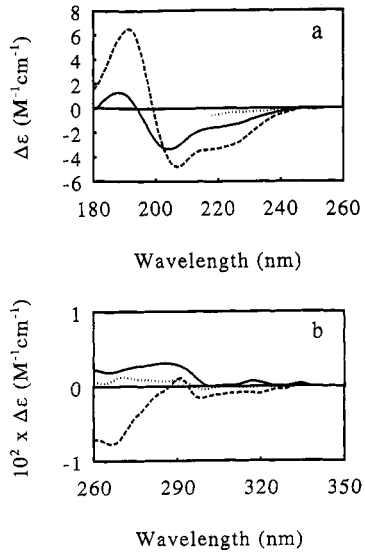

FIGURE 5: Far-UV (a) and near-UV (b) CD spectra of CYPR under different solvent conditions. Panels a and b: 0.51 mg/mL in 10 mM HEPES/1 mM EDTA, pH 7.5 (—); 0.53 mg/mL protein in buffer containing 50% (v/v) TFE (- - -); 0.50 mg/mL protein in buffer containing 4 M Gu-HCl (···).

these different conditions shows little indication of further formation of tertiary structure (Figure 5b).

The thermal stability of the partially structured pro region was examined by recording ^{1}H NMR spectra in a set of experiments where CYPR was gradually heated to 90 °C

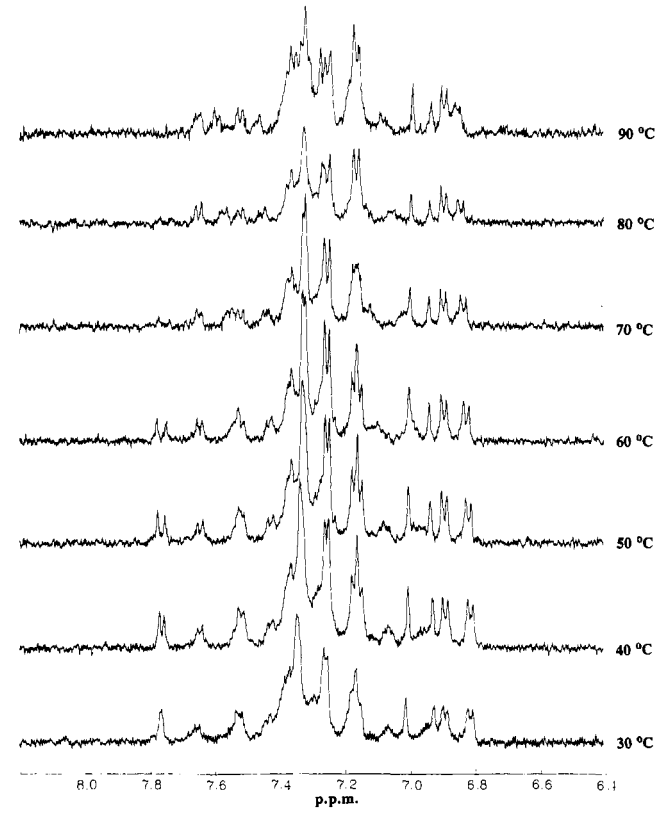

FIGURE 6: One-dimensional ^{1}H NMR spectra of 0.15 mM CYPR in 20 mM phosphate/0.15 M NaCl in D_2O, pH 8.0. The sample was gradually heated to 90 °C to follow the degree of unfolding. The indicated pHs are the direct pH meter readings not corrected for the presence of D_2O.

(Figure 6). The structures responsible for the asymmetric positions of the aromatic residues are remarkably resistant toward complete thermal denaturation, and even at 90 °C, CYPR is not fully unfolded (Figure 6). The spectra show that in particular the resonances of Trp^A and Tyr^A (see above) shift with increasing temperatures toward lower fields, whereas resonances of Tyr^B and Trp^B are only little perturbed by the increase in temperature. The change in chemical shift of the aromatic protons of Trp^A and Tyr^A over the entire temperature range suggests that the structure is in rapid chemical exchange with the equilibrium, shifting toward a random-coil conformation at increased temperatures. A spectrum recorded at 25 °C after heating to 90 °C showed that the local thermal unfolding was reversible (data not shown).

DISCUSSION

Carboxypeptidase Y from yeast is synthesized as a precursor of 532 amino acid residues containing a signal peptide of 20 residues and a pro region of 91 residues (Valls *et al.*, 1987). At least three important functions have been ascribed to the pro region of CPY: (1) it provides the targeting signal for directing the enzyme to the vacuole (Valls *et al.*, 1990), (2) it maintains the zymogen inactive until the vacuole has been reached; and (3) it is necessary for the correct and efficient folding of the enzyme *in vitro* (Winther & Sørensen, 1991) and *in vivo* (Ramos *et al.*, submitted for publication). Much attention has recently been focused on the latter kind of pro region function [see Fabre *et al.* (1992) and references cited therein]. As yet, however, no detailed structural characterization of a pro region involved in assisting protein folding has been presented. In this work, we show that CYPR (the recombinant pro region of CPY) is active in assisting refolding

of CPY provided that hydrophobic interactions are favored. Under these conditions, CYPR is a partially folded protein domain characterized by a relatively high content of secondary structure elements as judged by its HSQC and TOCSY NMR spectra and its far-UV CD spectrum [Figures 3, 4 (top), and 5a] but only a low content of rigid tertiary structure as judged by its intrinsic fluorescence, NOESY, and near-UV CD spectra [Figures 2, 4 (bottom), and 5b]. Although these are characteristics reminiscent of the compact intermediate collectively known as molten globules (Kuwajima, 1989; Baum *et al.*, 1989), CYPR deviates from the "typical" molten globule. First, CYPR only binds little of the hydrophobic probe ANS (Figure 3) although it is rich in hydrophobic residues. This suggests that a regular structural core either is not solvent-accessible or is absent. Although gel filtration experiments show that in the absence of CPY the pro region tends to self-associate at higher protein concentrations (data not shown), we favor the latter explanation based on the few interresidue NOEs observed in the NOESY experiment and the very low near-UV CD signal. This is also reflected in the polarity and extent of binding of ANS to CYPR compared to the stable folding intermediate of CPY (Figure 3). The latter exhibits a high binding of ANS with an emission maximum of about 470 nm, close to the value of free ANS in absolute ethanol or ANS bound to other equilibrium folding intermediates (Stryer, 1965; Chaffotte *et al.*, 1992). Compared to this, CYPR exhibits weak ANS binding and an emission maximum at about 490 nm, equal to a much less hydrophobic environment. This indicates that CYPR has only very little regular structural core. Second, particularly evident for some of its aromatic residues, the pro region is remarkably resistant toward complete thermal unfolding *in vitro* (even when exposed to temperatures as high as 90 °C, Figure 6). Whether this effect is due to an enhancement of hydrophobic interactions at higher temperatures and whether these residues form part of a core remain to be determined. Finally, almost all of CYPR is soluble *in vivo* in *E. coli* even when produced at high levels, and likewise it is selectively refolded when total solubilized inclusion body protein from *E. coli* is renatured by rapid dilution (data not shown). Under these conditions of high concentration of non-native structures (a situation probably not very different from the folding situation in the cell), the majority of the many hydrophobic residues in CYPR are rapidly stabilized before improper interactions leading to aggregation can take place. This is striking since it is believed that non-native protein species are particularly susceptible to aggregation and that only proteins with fast folding kinetics will partition to their native state (Mitraki *et al.*, 1991). With regard to the structural characteristics, similar observations has been made by Goldberg and co-workers for a proteolytic fragment of the *E. coli* tryptophan synthetase (Chafotte *et al.*, 1991, 1992). This fragment exhibits an extremely fast (less than 4 ms) nonspecific collapse to a state that has less nativelike secondary and supersecondary stucture than the typical molten globule. They proposed that it is a precursor to the molten globule and coined the term "pre-molten globule" to this structure (Chafotte *et al.*, 1991, 1992).

How can the feature of a partially folded pro region that is active in assisting protein folding be rationalized, and what are the implications for folding *in vivo*? In the cell, folding of the nascent polypeptide chain is believed to be a co-translational and vectorial process in an environment of high protein concentration (Rothmann, 1989; Gething & Sambrook, 1992). In this scheme, pro regions are envisaged to act as "co-translational chaperones" (Ellis & van der Vies, 1990) which guide the correct folding and transport of their cognate enzymes. It could be imagined that an N-terminal covalently linked chaperone should (i) exhibit extremely fast folding kinetics in order to be able to interact productively at an early stage with the growing nascent polypeptide chain before aggregation or proteolysis can occur, (ii) be able to interact with a large conformational space of nonnative structures on its target enzyme, and obviously (iii) be stable toward proteolysis and aggregation itself. All these demands can be satisfied by a partially folded structure as the one described here for CYPR if it is assumed that its rate of folding is kinetically competitive. We propose a model in which this competitiveness is accomplished through dynamic high-affinity interactions with structures on the enzyme. The stabilization of these interactions allows the enzyme to overcome its high activation energy barriers in folding. On the basis of the strong ANS binding of the stable intermediate of CPY (Figure 2) and the stimulating effect of high concentrations of salts on CPY refolding (Winther & Sørensen, 1991), we speculate that the interactions between CPY and its pro region are mainly of hydrophobic nature. Thus, it may be the partially folded protein structure *per se* that is important for the stability and function of the pro region. Interestingly, by employing a genetic approach, hydrophobic boxes were identified as essential for the function of the pro region of the pro region of the unrelated protease subtilisin (Kobayashi & Inouye, 1992). More generally, a similar model describing recognition via nonnative structures has been proposed for the action of the molecular chaperones SecB and hsc70 on their substrates (Randall, 1992; Park *et al.*, 1993). A functional requirement for flexibility and dynamics might explain why there is no sequence identity among pro regions, even where the cognate enzymes themselves are highly homologous (Jacobs *et al.*, 1985). Rather, this requirement would be expected to be sustained by evolutionary constraints at the level of secondary structure.

In summary, we have investigated the structural characteristics of the pro region of carboxypeptidase Y, and the results are in agreement with the presence of a partially folded protein domain under the conditions where it is functional. Current investigation of the ^{13}C- and ^{15}N-labeled pro region with multidimensional NMR methods will provide more detailed structural information.

ACKNOWLEDGMENT

We thank Pia Skovgaard for technical assistance, Ib Svendsen for N-terminal sequencing and amino acid analysis, and Jens Chr. Madsen and Birthe B. Kragelund for acquiring the NMR spectra. Morten Kielland-Brandt is thanked for critical reading of the manuscript.

REFERENCES

Baker, D., Sohl, J. L., & Agard, D. A. (1992) *Nature 356*, 263–265.

Baum, J., Dobson, C. M., Evans, P. A., & Hanley, C. (1989) *Biochemistry 28*, 7–13.

Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185–189.

Chaffotte, A., Guillou, Y., Delepierre, M., Hinz, H.-J., & Goldberg, M. E. (1991) *Biochemistry 30*, 8067–8074.

Chaffotte, A., Cadieux, C., Guillou, Y., & Goldberg, M. E. (1992) *Biochemistry 31*, 4303–4314.

Eder, J., Rheinhecker, M., & Ferhst, A. R. (1993) *Biochemistry 32*, 18–26.

Egan, D. A., Logan, T. M., Liang, H., Matayoshi, E., Fesik, S. W., & Holzman, T. F. (1993) *Biochemistry 32*, 1920–1927.

Ellis, R. J., & van der Vies, S. M. (1991) *Annu. Rev. Biochem. 60*, 321–347.
Fabre, E., Tharaud, C., & Gaillardin, C. (1992) *J. Biol. Chem. 267*, 15049–15055.
Gething, M.-E., & Sambrook, J. (1992) *Nature 355*, 33–45.
Gray, A. M., & Mason, A. J. (1990) *Science 247*, 1328–1330.
Horowitz, P. M., & Xu, R. (1992) *J. Biol. Chem. 267*, 19464–19469.
Ikemura, H., Takagi, H., & Inouye, M. (1987) *J. Biol. Chem. 272*, 7859–7864.
Jacobs, M., Eliasson, M., Uhlén, M., & Flock, J.-I. (1985) *Nucleic Acids Res.*, 8913–8926.
Johnson, W. C. (1990) *Proteins: Struct., Funct., Genet. 7*, 205–214.
Klionsky, D. J., Herman, P. K., & Emr, S. D. (1990) *Microbiol. Rev. 54*, 266–292.
Kobayashi, T., & Inouye, M. (1992) *J. Mol. Biol. 226*, 931–933.
Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87–103.
Mitraki, A., Haase-Pettingell, C., & King, J. (1991) *ACS Symp. Ser. 470*, 35–49.
Nebes, V. L., & Jones, E. W. (1991) *J. Biol. Chem. 266*, 22851–22857.
Neri, D., Wider, G., & Wüthrich, K. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 4397–4401.
Neurath, H. (1989) *Trends Biochem. Sci. 14*, 268–271.
Park, K., Flynn, G. C., Rothman, J. E., & Fasman, G. D. (1993) *Protein Sci.* 325–330.
Pohlner, J., Halter, R., Beyreuther, K., & Meyer, T. F. (1987) *Nature 325*, 458–462.
Randall, L. (1992) *Science 257*, 241–245.
Rothman, J. E. (1989) *Cell 59*, 591–601.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., & Gilmanshin, R. I. (1991) *Biopolymers 31*, 119–128.
Silen, J. L., & Agard, D. A. (1989) *Nature 341*, 462–464.
Silen, J. L., Frank, D., Fujishige, A., Bone, R., & Agard, D. A. (1989) *J. Bacteriol. 171*, 1320–1325.
Strandberg, L., & Enfors, S.-O. (1991) *Appl. Environ. Microbiol. 57*, 1669–1674.
Stryer, L. (1965) *J. Mol. Biol. 13*, 482–495.
Studier, F. W. (1991) *J. Mol. Biol. 37*, 37–44.
Studier, F. W., & Moffat, B. A. (1986) *J. Mol. Biol. 189*, 113–130.
Suter, U., Heymach, J. V., & Shooter, E. M. (1991) *EMBO J. 10*, 2395–2400.
Tabor, S. (1990) *Current Protocols in Molecular Biology*, Suppl. 11, pp 16.2.1–16.2.11, Wiley and Sons, New York.
Tabor, S., & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1074–1078.
Valls, L. A., Hunter, C. P., Rothman, J. H., & Stevens, T. H. (1987) *Cell 48*, 887–897.
Valls, L. A., Winther, J. R., & Stevens, T. H. (1990) *J. Cell Biol. 111*, 361–368.
von Hippel, P. H., & Wong, K.-Y. (1965) *J. Biol. Chem. 240*, 3909–3923.
Winther, J. R., & Sørensen, P. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9330–9334.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311–333.
Zhu, X., Ohta, Y., Jordan, F., & Inouye, M. (1989) *Nature 339*, 483–484.