Stability and Global Fold of the Mouse Prohormone Convertase 1 Pro-Domain†

Michael A. Tangrea, Patrick Alexander, Philip N. Bryan, Edward Eisenstein, John Toedt, and John Orban*

Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, Maryland 20850

Received November 16, 2000; Revised Manuscript Received March 9, 2001

ABSTRACT: We have purified the mouse prohormone convertase 1 (PC1) pro-domain expressed in *Escherichia coli* cells and demonstrated, using a number of biophysical methods, that this domain is an independent folding unit with a T_m of 39 °C at a protein concentration of 20 μM and pH 7.0. This differs significantly from similar pro-domains in bacteria and human furin, which are unfolded at 25 °C and require the catalytic domain in order to be structured [Bryan et al. (1995) *Biochemistry 34*, 10310−10318; Bhattacharjya et al. (2000) *J. Biomol. NMR 16*, 275−276]. Using heteronuclear NMR spectroscopy, we have determined the backbone ^{1}H, ^{13}C, and ^{15}N assignments for the pro-domain of PC1. On the basis of ^{1}H/^{13}C chemical shift indices, NOE analysis, and hydrogen exchange measurements, the pro-domain is shown to consist of a four-stranded β-sheet and two α-helices. The results presented here show that both the bacterial pro-domain in complex with subtilisin and the uncomplexed mouse PC1 pro-domain have very similar overall folds despite a lack of sequence homology. The structural data help to explain the location of the secondary processing sites in the pro-domains of the PC family, and a consensus sequence for binding to the catalytic domain is proposed.

The biosynthesis of many secreted proteases depends on an N-terminal pro-domain. Some examples are aspartic proteinases, such as carboxypeptidase Y (*1*), yapsin 1 (*2*), and protease A (*3*), cysteine proteinases such as cathepsin D and papain (*4*), and serine proteases such as *Bacillus* subtilisin (*5*), α-lytic protease (*6*), Kex proteases (*7*), and prohormone convertases (*8*). The prohormone convertases (PCs), a family of eukaryotic enzymes, have been shown to be homologous to bacterial subtilisin (*8*). To date, seven PCs have been identified in mammals (*9*). These serine proteases are responsible for processing multidomain proteins into biologically active peptide hormones (*10*). The PCs consist of five major domains: the N-terminal signal peptide, the pro-domain, the catalytic domain, the P domain (or homo-B-domain), and the C-terminal tail (*11*, *12*). The pro-domain is believed to serve numerous functions, such as promoting folding, stability, sorting, and transiently inhibiting the hormone processing activity (*13*).

The prohormone convertase PC1[1] (also known as PC3 or SPC1) is expressed in neuroendocrine tissue and is responsible for processing of prohormones in the dense core vesicles of the regulated secretory pathway (*9*, *14*, *15*). PC1 consists of 753 amino acids in total, including all five domains (*16*). The endoproteolytic activity of PC1 was first demonstrated with a series of experiments in which PC1 accurately cleaved the multifunctional precursor pro-opiomelanocortin (POMC) (*17*, *18*). PC1 cuts preferentially after the amino acid sequence RX(K/R)R, where X can be any amino acid, and is required for proper processing of proinsulin, proglucagon, and POMC (*19*). Although the catalytic domain of PC1 has about 25% sequence identity to the subtilisin catalytic domain, the pro-domains of PC1 (83 amino acids) and subtilisin (77 amino acids) share no detectable sequence homology (*20*).

While there is primary structural information for the prohormone convertases from various organisms, very little is known about the secondary or tertiary structures for any of the domains of the PC family. The sequence-specific NMR assignments for the inhibitory pro-domain of human furin have been determined, but this protein is globally unfolded in isolated form and therefore no structural data has been derived (*21*). Also, a C-terminal peptide fragment (residues 81−104) from the pro-domain of PC7 was recently found to be helical in 50% trifluoroethanol using nuclear magnetic resonance techniques (*22*).

In this paper, we describe the purification of mouse PC1 pro-domain expressed in *Escherichia coli* cells. Since the PC1 pro-domain is folded in the absence of the catalytic domain, we were able to characterize its structure and stability using a number of biophysical methods including circular dichroism (CD), analytical ultracentrifugation, and multidimensional heteronuclear NMR spectroscopy. The secondary and tertiary structural elements of the pro-domain obtained from NMR data indicate a folding topology similar to that of the bacterial pro-domain of subtilisin despite the lack of sequence homology. These structural data provide new insights into the positioning of the secondary processing site and shed light on the sequence requirements for binding the catalytic domain.

† Supported by a Life Technologies Fellowship (to M.A.T.), NIH Grant GM42560 (to P.N.B.), and NIST Grant COMM 60NANB6D0108 (to J.O.). M.A.T. is in the Molecular and Cell Biology graduate program at the University of Maryland, College Park.

* To whom correspondence should be addressed. Phone: 301-738-6221. Fax: 301-738-6255. E-mail: john@magpie.carb.nist.gov.

[1] Abbreviations: PC1, prohormone convertase 1; CD, circular dichroism; NMR, nuclear magnetic resonance; HSQC, heteronuclear single-quantum coherence; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy.

10.1021/bi0026472 CCC: $20.00
Published on Web 04/14/2001

MATERIALS AND METHODS

Cloning. The cloned mouse PC1/PC3 gene was graciously provided by the laboratory of Donald Steiner in plasmid pCMV5-PC3. Initially, the pro-region DNA was amplified in 30 cycles of a typical PCR reaction using pCMV5-PC3 as the template. Each cycle consisted of a 45 s 94 °C melt, a 45 s 50 °C anneal, and a 30 s 72 °C extension. The 5′-oligonucleotide encoded a *Nde*I site and the 3′-oligonucleotide a translation stop (TAA) and *Bam*HI site, which enabled digestion of the PCR product with restriction enzymes for cloning the fragment into plasmid pG5 (*23*). Correct clones were identified by restriction analysis. DNA sequences were verified by sequencing according to the Sequenase protocol (U.S. Biochemical) and found to be identical to the reported sequence (*24*). Due to nucleic acid contamination during protein purification, the gene was modified at the 5′-end to encode six histidine residues at the N-terminus of the protein, facilitating efficient purification on a nickel affinity resin. Uracil-containing single strand DNA template was prepared by the method of Kunkel (*25*). Oligonucleotide-directed mutagenesis was performed using standard molecular biology techniques. Correct clones were identified by DNA sequencing. The resulting pro-domain peptide has a seven amino acid extension at its N-terminus (MHHHHHH) prior to the wild-type N-terminal K1 at the putative pre-pro processing site and is 90 amino acids long, ending with R83 at the putative pro-mature processing site at its C-terminus (*24*).

Expression and Purification. Competent *E. coli* BL21-(DE3) cells were transformed, and cell fermentation was carried out using the methods described in ref *23*. The harvested *E. coli* paste was suspended in 30 mL of phosphate-buffered saline (PBS), and phenylmethanesulfonyl fluoride (PMSF) was added to a final concentration of 1 mM. DNase I (1 mg) and lysozyme (3 mg) were also added to aid in DNA digestion and cell lysis, respectively. Further cell lysis was achieved by multiple passes through a French pressure cell press using an Aminco pressure cell subjecting the suspension to 1000 psi. The suspension was then centrifuged at 10000*g* for 25 min. The supernatant was removed and filtered through a 0.45 μm membrane and chromatographed on a 2.5 × 10 cm column containing 10 mL of His-Bind resin (Novagen). The target protein was eluted from the column with a 5−500 mM imidazole linear gradient in 50 mM potassium phosphate and 0.5 M NaCl, pH 7.9. Samples (15 μL) from the column fractions were checked by sodium dodecyl sulfate (SDS)−polyacrylamide gel electrophoresis (PAGE) on 10−20% running gels for purity. The purified fractions were pooled together and dialyzed extensively against deionized water using a 3500 MWCO membrane at 4 °C. The sample was then lyophilized. The mass of the purified protein was verified by MALDI on a Voyager-DE (PerSeptive Biosystems). The typical yield was about 35 mg of purified unlabeled protein from a 1.5 L fermentation. An $A_{280} = 1.45$ for a 1 mg/mL solution was used to determine concentration.

Uniformly labeled $^{13}C/^{15}N$ samples were prepared by expression in M9 minimal media with $^{15}NH_4Cl$ and [^{13}C]-glucose as the sole nitrogen and carbon sources, respectively. Labeled protein samples were purified as above with typical yields of 15 mg from a 1.5 L fermentation.

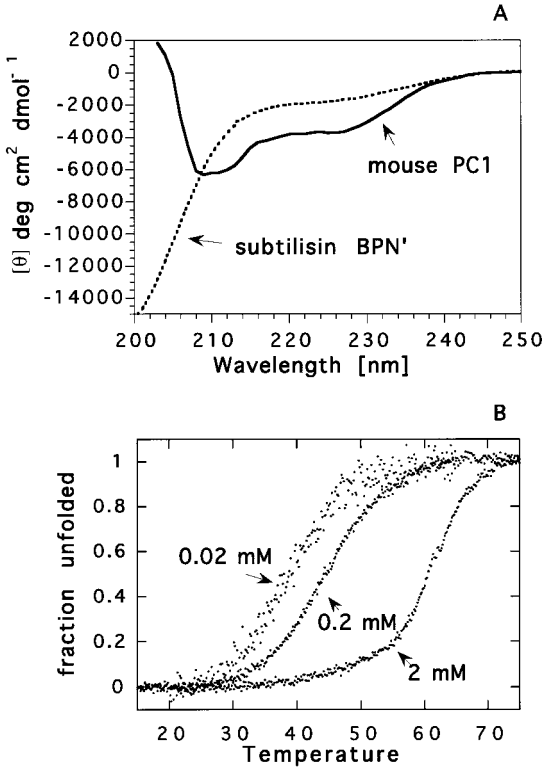

FIGURE 1: (A) CD profiles of the pro-domains of PC1 and subtilisin BPN′. Mean residue ellipticity (deg cm^2/dmol) is plotted versus wavelength. Spectra were measured in 0.1 M KP_i, pH 7.2, using a 1 cm cylindrical cuvette at 25 °C with a protein concentration of 20 μM. (B) Temperature unfolding profile of the PC1 pro-domain. Mean residue ellipticity (deg cm^2 dmol^{-1}) at 222 nm is plotted versus temperature. Solvent conditions were 0.1 M KP_i, pH 7.2.

Circular Dichroism. Circular dichroism measurements were performed with a Jasco spectropolarimeter, model J-720, using water-jacketed quartz cells with path lengths of 1 cm, 1 mm, or 0.1 mm depending on the pro-domain concentration which ranged from 20 μM to 2.0 mM. Far-UV wavelength scans were recorded at 25 °C from 250 to 200 nm. The average for five CD spectra is presented. The ellipticity results were expressed as mean residue ellipticity, [θ], in deg cm^2 dmol^{-1}. Temperature-induced unfolding of the pro-domain was performed in the temperature range between 15 and 95 °C in 1 cm, 1 mm, and 0.1 mm cuvettes. Ellipticities at 222 nm were monitored every 10 s at a scanning rate of 1 deg/min. Reversibility of the denaturations was confirmed by comparing the CD spectra at 25 °C before heating and after the heating and cooling cycle.

Analytical Ultracentrifugation. Sedimentation equilibrium experiments were performed with a Beckman XL-A Optima analytical ultracentrifuge equipped with a four-hole, An-55 rotor. All experiments were performed at 25 °C at rotor speeds of 32 000 and 40 000 rpm. The concentration distribution of pro-domains at equilibrium was acquired in a step mode as an average of 20 measurements of absorbance at 280 nm at each radial position, with nominal spacing of 0.001 cm between radial positions. Scans separated by time intervals of 15 min for the first 4 h and then every 3 h were compared. A partial specific volume of 0.721 cm^3 g^{-1} was calculated on the basis of the amino acid composition. Equilibrium constants for the dissociation of the dimeric pro-

Table 1: Line Widths (Hz) for Resolved Methyl Proton Resonances versus Concentration of the PC1 Pro-Domain (mM)[a]

[proPC1]	L30 δCH_3 (δ 0.34 ppm)	I22 δCH_3 (δ 0.23 ppm)	I59 δCH_3 (δ −0.57 ppm)
0.05[b]	16	18	29
0.2	17	15	30
0.6	18	15	30
2.0	17	16	24

[a] Line width at half-height (±1 Hz) was measured for three upfield-shifted resonances directly from one-dimensional proton NMR spectra. [b] The 0.05 mM spectrum was collected at 500 MHz using a cryoprobe (K. Colson, Bruker Instruments, Billerica, MA).

domain to monomers were estimated by using a modified version of IGOR software (Wavemetrics, Lake Oswego, OR) (*26*).

NMR Spectroscopy. Both unlabeled and isotopically labeled NMR samples were prepared by dissolving the lyophilized protein powder in 450 μL of 100 mM potassium phosphate buffer, pH 7.0, containing 10% D_2O/90% H_2O. An additional sample was also prepared in >99.9% D_2O. NMR spectra were recorded on Bruker DRX-500 and DRX-600 spectrometers equipped with 3-axis gradient probes. Data were processed on a Dell Precision 410 workstation using NMRPipe (*27*). Two-dimensional ^{15}N HSQC spectra were recorded in States−TPPI mode using a gradient water flip-back scheme for solvent suppression (*28*). Homonuclear two-dimensional TOCSY (*29*) and NOESY (*30*, *31*) spectra were recorded in both H_2O and D_2O with mixing times of 55 and 150 ms, respectively.

Three-dimensional ^{15}N NOESY-HSQC and ^{15}N TOCSY-HSQC experiments (*32*, *33*) were recorded with mixing times of 150 ms (NOESY) and 55 and 70 ms (TOCSY), at 15 and 25 °C. For the ^{15}N NOESY-HSQC experiment, 64 t_1, 32 t_2, and 512 t_3 time points were utilized with sweep widths of 13.0, 33.0, and 13.0 ppm (F_1, F_2, F_3). For the ^{15}N TOCSY-HSQC experiment, 50 × 20 × 512 time points were collected in the $t_1 \times t_2 \times t_3$ dimensions using a DIPSI-2 mixing sequence (*34*, *35*).

Three-dimensional HNCACB (*36*) and CBCA(CO)NH (*37*) spectra were recorded at 25 °C with sweep widths of 68.4, 33.0, and 13.0 ppm (F_1, F_2, F_3). For the HNCACB experiment, 64 t_1 × 32 t_2 × 512 t_3 time points were collected, and the CBCA(CO)NH experiment consisted of 56 t_1, 25 t_2, and 512 t_3 time points. The carrier frequencies for both experiments were ^1H 4.70, ^{13}C 46.0, and ^{15}N 118.0 ppm. The three-dimensional HBHA(CBCACO)NH experiment (*37*) consisted of 60 t_1, 32 t_2, and 512 t_3 complex points at sweep widths of 8.0, 32.0, and 16.0 ppm, respectively. A three-dimensional HNCO experiment (*38*) was acquired with a ^{13}C carrier frequency of 177.0 ppm. The complex points

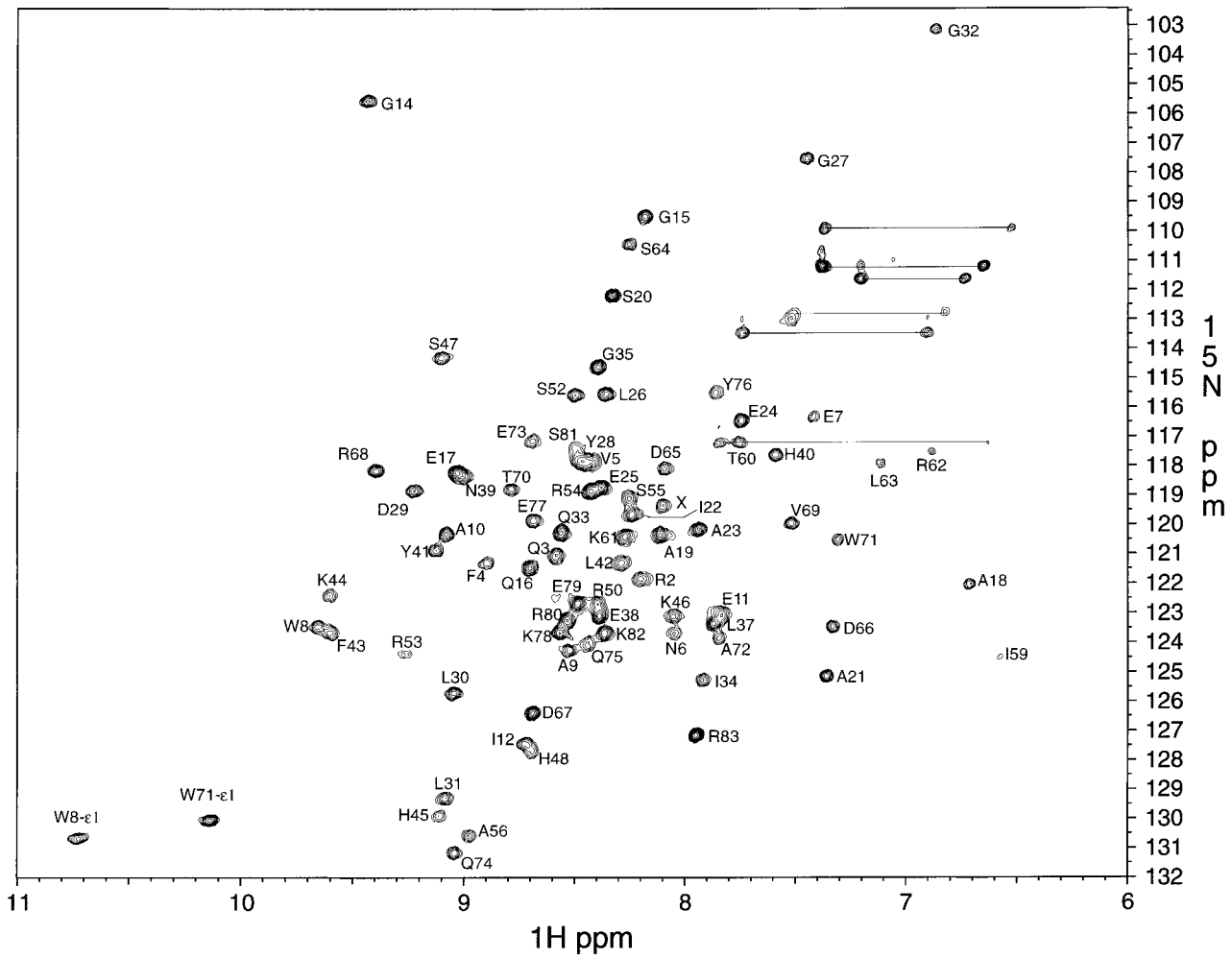

FIGURE 2: ^{15}N HSQC spectrum of the mouse PC1 pro-domain at 25 °C, pH 7.0. Horizontal lines connect cross-peaks due to side chain Asn/Gln amide protons. The peak labeled X is unassigned but may be due to one of the histidines in the N-terminal His_6 tag, based on triple resonance data.

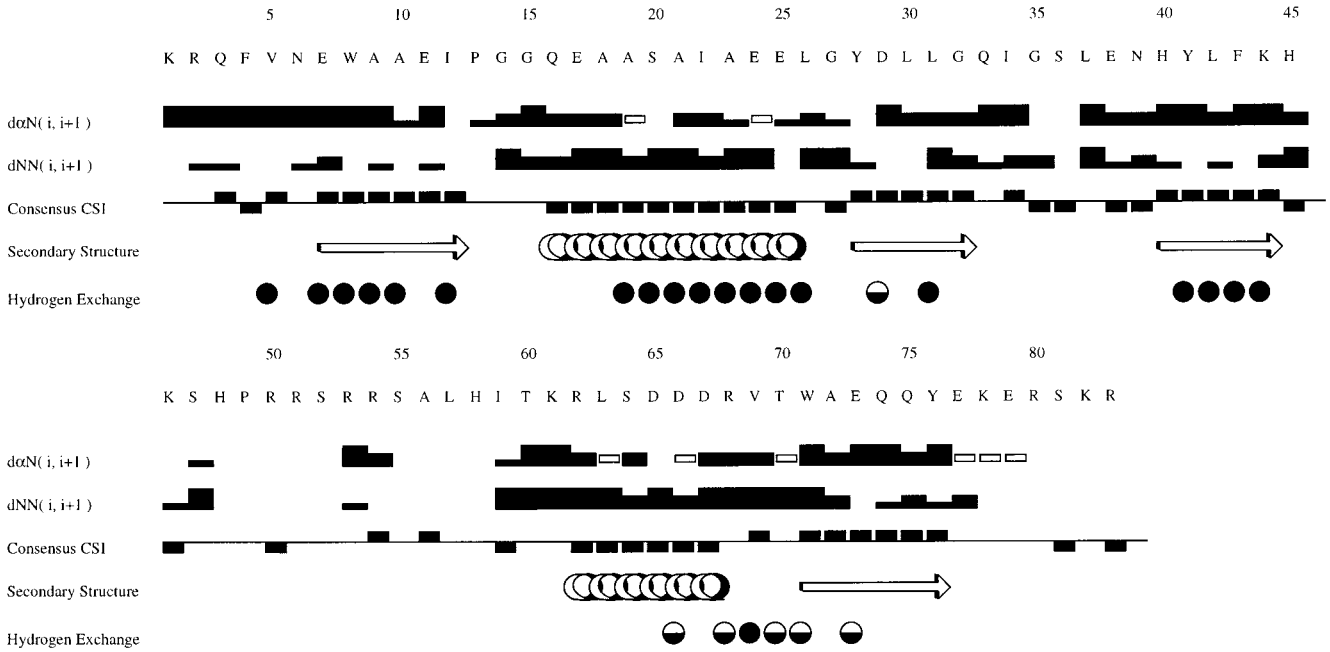

FIGURE 3: Summary of sequential NOEs, consensus chemical shift index (CSI), secondary structure, and hydrogen exchange data. The sequential NOE $d_{\alpha N}(i, i+1)$ is between the H_α of residue i and the H_N of residue $i+1$. The sequential NOE $d_{NN}(i, i+1)$ is between the H_N of residue i and the H_N of residue $i+1$. NOE intensity is proportional to the height of the solid bars. Open bars indicate positions where sequential connectivities could not be established due to peak overlap. Consensus CSI was determined using the method of Wishart and Sykes (*43*). Secondary structures are delineated on the basis of the consensus CSI. For H−D exchange data, solid circles represent amide protons with protection factors greater than 4000, and half-filled circles represent detectable amide protons with protection factors of less than 4000. The remaining amide protons were in fast exchange and could not be detected in D_2O. The figure was generated using VINCE (http://www.rowland.org/rnmrtk/vince.html).

were 64 t_1, 25 t_2, and 512 t_3 at sweep widths of 18.0, 32.0, and 13.0 ppm, respectively.

Hydrogen−deuterium (H−D) exchange experiments were carried out by adding D_2O to the lyophilized ^{15}N-labeled pro-domain and recording ^{15}N HSQC spectra at 25 °C over a 24 h period. Exchange rates were determined by fitting data to a first-order exponential equation of the form $A_t = A_0 \exp(-k_{ex}t)$, where A_t is the peak volume at time t and k_{ex} is the first-order rate constant. Random coil intrinsic exchange rates (k_{int}) for amide protons were corrected for sequence-dependent effects, pH, and temperature using the methods described by Bai et al. (*39*).

RESULTS

Analysis of the Temperature Unfolding Profile for the PC1 Pro-Domain. Figure 1A shows the CD spectrum of the PC1 pro-domain. The mean residue ellipticity at 222 nm is −3800 deg cm^2 dmol^{-1}, consistent with the mixed α, β, and disordered structure content determined from NMR (see below). To determine the equilibrium constant for pro-domain folding, analysis of pro-domain unfolding versus temperature was carried out in 100 mM potassium phosphate, pH 7.2, at pro-domain concentrations of 20 μM, 200 μM, and 2.0 mM. Temperature unfolding profiles are presented in Figure 1B. The unfolding reactions are completely reversible after scanning to 65 °C. The fraction native is determined by subtracting unfolded baseline from the experimental CD signal and then dividing by the total CD difference between 100% folded and 0% folded at that temperature. As the concentration of the pro-domain is increased, the temperature of the melting transition increases. At 20 μM, the melting temperature is 39 °C, while at 2 mM (the concentration of the H−D exchange experiments), the melting temperature is 60 °C. Since concentration dependence of stability is often an indication of association and/or aggregation, sedimentation equilibrium experiments were carried out.

Sedimentation Equilibrium Studies. The sedimentation equilibrium method is widely used for accurate determination of molecular homogeneity and associations of proteins. We applied this method to the PC1 pro-domain in order to quantitatively characterize its tendency to associate. Four different initial concentrations of the pro-domain were centrifuged at rotor speeds of 32K and 40K rpm. The four concentrations ranging from 0.07 to 1.38 mg/mL were centrifuged to equilibrium. In all distributions, the pro-domain behaves as a homogeneous population of monomers with average molecular mass of 10 800 Da (calculated MW = 10 654.8) in 100 mM potassium phosphate buffer, pH 7.0. Differences between the fitted curves and the data were small (root-mean-square deviations ∼0.005) and were distributed almost randomly around zero.

Free Energy of Unfolding. The ΔG of unfolding for the pro-domain at 20 μM and 2 mM concentrations was determined from CD melting data. At 20 μM, the ΔG for the unfolding reaction is ∼2 kcal/mol at 25 °C. As the protein concentration is increased, the ΔG of unfolding also increases (Figure 1B). From both the CD melting data and from the rate of H−D exchange (see below), the ΔG for unfolding is ∼5 kcal/mol at 2 mM and 25 °C. Concentration-dependent effects of this type are frequently ascribed to oligomerization of proteins. However, the sedimentation results demonstrate that the pro-domain associates only weakly, if at all. The absence of significant amounts of dimer/oligomer is further supported by an analysis of proton NMR line widths as a

function of concentration. Table 1 shows that there is no appreciable change in line width from 0.05 to 2 mM pro-domain concentration. Moreover, the line widths for L30 δCH3 and I22 δCH3 are consistent with methyl group line widths obtained for other proteins in the same molecular mass range and under similar conditions. The broader line width for I59 δCH3 is most likely due to chemical exchange effects in the β3−α2 loop (see Discussion).

Given the absence of strong association of the folded pro-domain, it is difficult to explain the observed concentration dependence of stability (*40*). If the folded pro-domain dimerized with a K_d of 4 mM (the minimum value consistent with the sedimentation data), there would be little effect of pro-domain concentration on the observed equilibrium constant for folding in the millimolar range. For example, if K_d = 4 mM and [pro-domain] = 2 mM, the equilibrium constant for folding would only increase by 1.6-fold, relative to [pro-domain] = 20 μM. Since each kcal/mol gain in ΔG corresponds with a 5.4-fold increase in the equilibrium constant at 25 °C, the 3 kcal/mol increase which we observe for the pro-domain would correspond with a 150-fold increase in equilibrium constant. This means that the K_d would need to be less than 200 μM if the $\Delta\Delta G$ of 3 kcal/mol was due to dimerization. A K_d of this magnitude is clearly inconsistent with both sedimentation results and the analysis of NMR line widths.

One possible explanation for increased stability at high concentration may be the volume-excluding properties of the pro-domain itself. Effects of protein concentration on stability have been reported previously (*41*), although the magnitude of the effect seen with the pro-domain is large. In any case, we would argue that the $\Delta G_{unfolding}$ of ~2 kcal/mol observed at low protein concentration most accurately reflects the intrinsic conformational stability of the PC1 pro-domain.

Resonance Assignments. The HSQC spectrum of the PC1 pro-domain showed good signal dispersion characteristic of a folded protein (Figure 2). Sequence-specific resonance assignments were made for H_N, N_H, C_α, H_α, and C′ atoms as well as C_β and H_β atoms for most residues. These were obtained from a combination of the following standard three-dimensional double and triple resonance experiments using a $^{13}C/^{15}N$-labeled sample: HNCO (*38*), HNCACB (*36*), CBCA(CO)NH (*37*), HBHA(CBCACO)NH (*37*), and ^{15}N TOCSY-HSQC (*32*, *33*). No resonances were observed for the first eight residues at the N-terminus corresponding to the N-terminal methionine, the H_6 attachment, and K1. These residues are most likely disordered and are exchange broadened in NMR spectra. Backbone assignments were obtained for 78 of the other 82 residues. The NMR signals for S36, R51, L57, and H58 were not detected. S36 is located in the β2−β3 turn, and the other three unassigned residues are in the 17-residue loop between β3 and α2.

Chemical Shifts and NOEs. The chemical shift index was used to determine the secondary structure elements for the PC1 pro-domain (*42*, *43*) (Figure 3). The pro-domain consists of four β-strands and two α-helices, and these regions of secondary structure were further confirmed by sequential NOE data (Figure 3). Strong $d_{\alpha N}(i, i+1)$ NOEs were found in β-strands, and strong $d_{NN}(i, i+1)$ NOEs were found in α-helices. Furthermore, long-range NOEs (not shown) were detected between the β-strands which indicate that the fold of the mouse PC1 pro-domain is very similar to that of the bacterial subtilisin pro-domain. NOE analysis indicates a β-α-β-β-α-β fold with a four-stranded antiparallel β-sheet (Figure 4B).

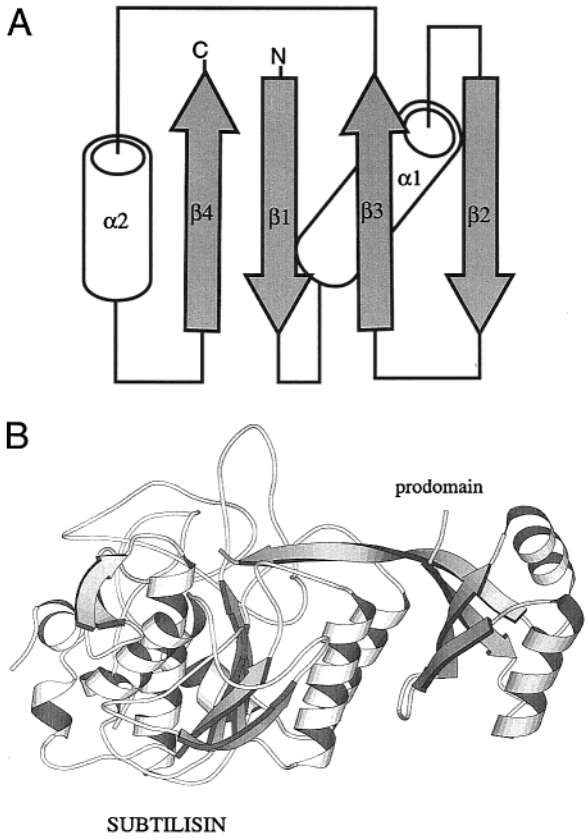

FIGURE 4: (A) Schematic diagram of the global fold for the mouse PC1 pro-domain from NMR data. (B) Structure of the bacterial pro-domain in complex with subtilisin BPN′ (PDB accession code 1SPB). The ribbon drawing depicts the α-carbon backbone of the bimolecular complex of subtilisin BPN′ (lighter shading) and the pro-domain (darker shading). The C-terminus of the pro-domain extends into the subtilisin active site. The figure was generated using Molscript (*50*).

Hydrogen−Deuterium Exchange. Protection from H−D exchange was observed for 26 main chain amide protons in the mouse PC1 pro-domain (Figure 3). Of the detectable amide protons in D_2O solution, 20 were classified as slow exchanging. These NHs have protection factors ($=k_{int}/k_{ex}$) in the range of 4200−21 800 and ΔG_{op} values [$=-RT$ ln-(k_{ex}/k_{int})] from 4.9 to 5.9 kcal/mol with an average ΔG_{op} of 5.3 ± 0.3 kcal/mol. Moreover, the H−D exchange results are consistent with the α/β folding topology determined from triple and double resonance NMR experiments.

DISCUSSION

Comparison of Stability between the Pro-Domains of Subtilisin BPN′ and Mouse PC1. The isolated bacterial pro-domain of subtilisin BPN′ from *Bacillus amyloliquefaciens* is largely unfolded even under optimal folding conditions (Figure 1) but has some propensity to fold on its own. The free energy of independent folding, although unfavorable, is small enough that a few percent of pro-domain molecules are folded under aqueous conditions at 25 °C (*44*). Similar results were observed for the pro-domain of human furin, which is unfolded at 30 °C (*21*). In contrast, the pro-domain of PC1 is fully folded at 25 °C, allowing structural studies,

A

```
                3                 13                23                33                43
BPN'            KSNGEKKYIVGFKQTMSTMSAAKKKDVISEKGGKVQKQFKYV--DAASAT
2°Structure          ββββββ           αααααααααα      βββββββ      βββββ
mPC1/3          KRQFVNEWAAEIPGG--------QEAASAIAEELG-YDLLGQIGSLENHYLFK
                1                 10                20                30                40

                53                                  63                73
BPN'            LNEK----------------------AVKELKKDPSVAYVEEDHVAHAY
2° Structure                        ααααααα     βββββ
mPC1/3          HKSHPRRSRR SALHITKRLSDDDRVT-WAEQQYEKERSKR
                     50                60                70                80
```

B

```
2°Structure                  β β β β β β                 αααααα ααααα    βββββ ββ          βββββ
                        1                 10                20                30                40
mPC1/3                KRQFVNEWAAEIP----------GGQEAASAIAEELGYDLLGQIGSLENHYLFK
hfurinI             QGQKVFTNTWAVRIP----------GGPAVANSVARKHGFLNLGQI----FGDYYHFW
hPC2                  ERPVFTNHFLVELHK----------GGEDKARQVAAEHGF-GVRKLPFAEGLYHFY
mPC4                  QAPIYVSSWAVRVT----------KGYQEAERLARKFGFVNLGQIFPDDQYFHLR
hPACE4          APPPRPVYTNHWAVQVL----------GGPAEADRVAAAHGYLNLGQIGNLEDYYHFY
mPC5/6                RTRVYTNHWAVKIA----------GGFAEADRIASKYGFINVGQIGALKDYYHFY
rPC7    LTEAGGLDTLGAGGLSWAVHLDSLEGERKEESLIQQANAVAQAAGLVNAGRIGELQGHYLFV
Conserved               v  t N  WAV           GG    A   vA     GF  NI GQI g  l    y YHF
                              *  *                                *  **  **           *  *
2° Structure                                  ααα ααααα   βββββ
                                   50                60                70                80
mPC1/3          ——HKSHPRRSRRSALHITKRLSDDDRVTWAEQQYEKERSKR
hfurinI         ------HRGVTKRSLSPHRPRHSRLQREPQVQWLEQQVAKRRTKR
hPC2            ------HNGLAKAKRRRSLHHKQQLERDPRVKMALQQEGFDRKKR
mPC4            ------HRGVAQQSLTPHWGHRLRLKKDPKVRWFEQQTLRRRVKR
hPACE4          ------HSKTFKRSTLSSRGPHTFLRMDPQVKWLQQQEVKRRVKR
mPC5/6          ------HSRTIKRSVLSSRGTHSFISMEPKVEWIQQQVVKKRTKR
rPC7            QPAGHGQAMEAEAMRQQAEAVLAKHEAVRWHSEQRLLKRAKR
Conserved         H  g   krS              L   dP V W   QQ   K  R  KR
                                                  *  *
```

FIGURE 5: (A) Alignment of amino acid sequences for the bacterial BPN′ and mouse PC1 pro-domains based on secondary structures. Secondary processing sites are underlined. (B) Sequence alignment between mammalian PC pro-domains with the secondary structure of mouse PC1 at the top of the figure. The numbering pertains to the PC1 sequence. Conserved residues are in bold type. In the bottom line, highly conserved residues (present in at least five out of seven sequences) are shown in upper case lettering while less conserved residues (present in four out of seven sequences) are shown in lower case lettering. Putative surface residues, predicted from alignment with the crystal structure of the subtilisin BPN′ pro-domain, are marked with an asterisk. The primary (C-terminal) and secondary processing sites are underlined.

but is nevertheless on the margin of stability at biologically relevant concentrations.

Low independent stability of these pro-domains may have been selected for in the course of evolution. The pro-domain is required to promote folding of the catalytic domain and then is proteolytically degraded. In the case of subtilisin, pro-domain mutants which were engineered to be highly stable are strong inhibitors of the catalytic domain (*45*). Indeed, the stability of the pro-domain is directly linked to its binding affinity for the catalytic domain. For example, as the fraction of folded pro-domain approaches one, the observed association constant with subtilisin approaches its maximum (*46*). The final step in subtilisin biosynthesis is the release of the pro-domain to generate active subtilisin. With an independently folded pro-domain, this final step may become rate limiting. Thus the pro-domains may have evolved a stability which is adequate for folding the catalytic domains but not so stable that they are long-lived inhibitors.

Structural Comparison between Subtilisin BPN′ and PC1 Pro-Domains. Both the uncomplexed mouse PC1 pro-domain (Figure 4A) and the bacterial subtilisin pro-domain in complex with subtilisin (Figure 4B) (*44*, *47*) have very similar overall folds despite a lack of sequence homology. Their structures consist of a four-stranded antiparallel β-sheet and two α-helices. The lengths of the secondary structural elements are very similar for both proteins. However, some differences are also apparent. Alignment of the amino acid sequences based on their secondary structures reveals different sizes for the $\beta1-\alpha1$ and $\beta3-\alpha2$ loops in the bacterial subtilisin and mouse PC1 pro-domains (Figure 5A). These different loop lengths are most likely due to differences in the position of the secondary processing site between the bacterial and mammalian systems. In the bacterial pro-domain, the secondary processing site is located in the $\beta1-\alpha1$ loop (*47*) whereas in PC1 pro-domain it is located in the $\beta3-\alpha2$ loop (see below). Another difference between these two folds is that the C-terminal residues of the subtilisin pro-domain are in an extended β-strand conformation whereas residues 76−83 of the PC1 pro-domain have a random coil conformation. In the case of the subtilisin pro-domain, ordering of the C-terminal residues is a result of complex formation with the catalytic domain. The C-terminal residues 72−77 extend out from the central part of the pro-domain and bind in a substrate-like manner along subtilisin's active site cleft. Residues Y77, A76, H75, and A74 of the pro-domain occupy subsites S1 to S4 of subtilisin, respectively. In contrast, the PC1 structural studies were carried out on the uncomplexed pro-domain.

Sequence and Structural Homology in the PC Pro-Domains. Comparison of amino acid sequences between the various mammalian PC pro-domains shows an identity level of 30−40% (Figure 5B). This indicates that, in complex with their catalytic domains, these pro-domains will most likely have a similar fold to that of mouse PC1 pro-domain even though some may be unfolded as isolated domains. The high sequence identity (>80%) for PC1 pro-domains across other species such as rat, pig, and human indicates that these too will probably adopt the same fold as mouse PC1 pro-domain.

In light of the structural data, it is interesting to note that the sequence homology between mammalian PC pro-domains

is quite high in the β-strand regions (Figure 5B). On the basis of the alignment with the bacterial pro-domain structure, both buried and surface residues appear to be conserved for over 50% of the residues in the β-strand regions of the PC pro-domains. The relatively high degree of conservation within the β-strand elements may serve two purposes: (1) to maintain packing with the α-helices within the pro-domain and (2) to maintain a basal level of favorable contacts with the catalytic domain. This second point is based on structural homology with the bacterial pro-domain and the observation that most of the bacterial pro-domain residues involved in contacts with subtilisin are in the C-terminal tail and the β-sheet (Figure 4A). Therefore, in addition to the primary processing site at the C-terminus of the PC pro-domains, a β-sheet consensus sequence for binding to the catalytic domain is proposed (Figure 5B). Consensus residues appear in all four β-strands, but putative consensus surface residues have the highest concentration in the β2-strand, which is an outer strand of the β-sheet. Presumably, specificity arises from changes around the primary processing site as well as variations in other less conserved surface/contact residues in or nearby the β-sheet.

Relation between Secondary Processing Sites and Structure. It is instructive to note the positions of the secondary processing sites in light of the new structural data obtained here for the mouse PC1 pro-domain (Figure 5B). While the primary processing site is conserved across the different PC pro-domains, the secondary processing sites are less well conserved both in sequence and in the position of proteolytic cleavage on the polypeptide chain (*48*). With the exception of PC7, the secondary processing sites are contained within a large 17-residue loop between the β3-strand and α2-helix that is solvent accessible. Some backbone amide protons in the β3$-$$\alpha$2 loop of PC1 pro-domain (R51, L57, H58) are exchange broadened, suggesting that this loop may be flexible and may exist in multiple conformational states. Further NMR experiments will be necessary to confirm this. The PC pro-domains are able to accommodate a wide variety of sequences in this large loop since there is little sequence homology in this region. In addition, the position of the secondary processing site varies slightly but always remains within the loop and hence is readily accessible to the catalytic domain. In the case of PC7, the putative secondary processing site is located in the β1$-$$\alpha$1 loop, which is the same loop used by the bacterial subtilisin pro-domain (Figure 5). PC7 is believed to be evolutionarily divergent from the other mammalian PCs (*49*), and its pro-domain is about 30 residues longer than the other PC pro-domains. On the basis of the data presented here, the PC7 pro-domain most likely has the same overall fold as the other PC pro-domains but has longer β1$-$$\alpha$1 and β3$-$$\alpha$2 loops.

REFERENCES

1. Winther, J. R., and Sorensen, P. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9330−9934.
2. Cawley, N. X., Olsen, V., Zhang, C. F., Chen, H. C., Tan, M., and Loh, Y. P. (1998) *J. Biol. Chem. 273*, 584−591.
3. van den Hazel, H. B., Kielland-Brandt, M. C., and Winther, J. R. (1993) *J. Biol. Chem. 268*, 18002−18007.
4. Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Menard, R., Magny, M. C., Storer, A. C., and Thomas, D. Y. (1995) *J. Biol. Chem. 270*, 10838−10846.
5. Zhu, X., Ohta, Y., Jordan, F., and Inouye, M. (1989) *Nature 339*, 483−484.
6. Baker, D., Sohl, J., and Agard, D. A. (1992) *Nature 356*, 263−265.
7. Brenner, C., and Fuller, R. S. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 922−926.
8. Steiner, D., Smeekens, S. P., Ohagi, S., and Chan, S. J. (1992) *J. Mol. Biol. 267*, 23435−23438.
9. Zhou, A., Webb, G., Zhu, X., and Steiner, D. F. (1999) *J. Biol. Chem. 274*, 20745−20748.
10. Steiner, D. F. (1998) *Curr. Opin. Chem. Biol. 2*, 31−39.
11. Seidah, N. G., Chretien, M., and Day, R. (1994) *Biochimie 76*, 197−209.
12. Rouille, Y., Duguay, S. J., Lund, K., Furuta, M., Gong, Q., Lipkind, G., Oliva, A. A., Jr., Chan, S. J., and Steiner, D. F. (1995) *Front. Neuroendocrinol. 16*, 322−361.
13. Creemers, J. W., Jackson, R. S., and Hutton, J. C. (1998) *Semin. Cell Dev. Biol. 9*, 3−10.
14. Seidah, N. G., Marcinkiewicz, M., Benjannet, S., Gaspar, L., Beaubien, G., Mattei, M. G., Lazure, C., Mbikay, M., and Chretien, M. (1991) *Mol. Endocrinol. 5*, 111−122.
15. Smeekens, S. P., and Steiner, D. F. (1990) *J. Biol. Chem. 265*, 2997−3000.
16. Van de Ven, W. J., Roebroek, A. J., and Van Duijnhoven, H. L. (1993) *Crit. Rev. Oncog. 4*, 115−136.
17. Thomas, L., Leduc, R., Thorne, B. A., Smeekens, S. P., Steiner, D. F., and Thomas, G. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 5297−5301.
18. Benjannet, S., Rondeau, N., Day, R., Chretien, M., and Seidah, N. G. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 3564−3568.
19. Zhou, A., and Mains, R. E. (1994) *J. Biol. Chem. 269*, 17440−17447.
20. Siezen, R. J., and Leunissen, J. A. (1997) *Protein Sci. 6*, 501−523.
21. Bhattacharjya, S., Xu, P., and Ni, F. (2000) *J. Biomol. NMR 16*, 275−276.
22. Bhattacharjya, S., Xu, P., Zhong, M., Chretien, M., Seidah, N. G., and Ni, F. (2000) *Biochemistry 39*, 2868−2877.
23. Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (1992) *Biochemistry 31*, 3597−3603.
24. Smeekens, S. P., Avruch, A. S., LaMendola, J., Chan, S. J., and Steiner, D. F. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 340−344.
25. Kunkel, T., Bebenek, K., and McClary, J. (1991) *Methods Enzymol. 204*, 125−139.
26. Brooks, I. S., Soneson, K. K., and Hensley, P. (1993) *Biophys. J. 64*, A244.
27. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
28. Mori, S., Abeygunawardana, C., Johnson, M. O., and van Zijl, P. C. (1995) *J. Magn. Reson. B 108*, 94−98.
29. Braunschweiler, L., and Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521−528.
30. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.
31. Kumar, A., Ernst, R. R., and Wuthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1−6.
32. Fesik, S. W., and Zuiderweg, E. R. P. (1988) *J. Magn. Reson. 78*, 588−593.
33. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry 28*, 6150−6156.
34. Shaka, A. J., Lee, C. J., and Pines, A. (1988) *J. Magn. Reson. 77*, 274−293.
35. Rucker, S. P., and Shaka, A. J. (1989) *Mol. Phys. 68*, 509−517.
36. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson., Ser. B 101*, 201−205.
37. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.
38. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson. 96*, 432−440.
39. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins 17*, 75−86.

40. Schellman, J. A. (1975) *Biopolymers 14*, 999−1018.
41. Minton, A. P. (2000) *Biophys. J. 78*, 101−109.
42. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) *Biochemistry 31*, 1647−1651.
43. Wishart, D. S., and Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171−180.
44. Bryan, P., Wang, L., Hoskins, J., Ruvinov, S., Strausberg, S., Alexander, P., Almog, O., Gilliland, G., and Gallagher, T. D. (1995) *Biochemistry 34*, 10310−10318.
45. Ruvinov, S., Wang, L., Ruan, B., Almog, O., Gilliland, G., Eisenstein, E., and Bryan, P. (1997) *Biochemistry 36*, 10414−10421.
46. Ruan, B., Hoskins, J., and Bryan, P. N. (1999) *Biochemistry 38*, 8562−8571.
47. Gallagher, T. D., Gilliland, G., Wang, L., and Bryan, P. (1995) *Structure 3*, 907−914.
48. Muller, L., Cameron, A., Fortenberry, Y., Apletalina, E. V., and Lindberg, I. (2000) *J. Biol. Chem. 275*, 39213−39222.
49. Seidah, N. G., Hamelin, J., Mamarbachi, M., Dong, W., Tadros, H., Mbikay, M., Chretien, M., and Day, R. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 3388−3393.
50. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI0026472